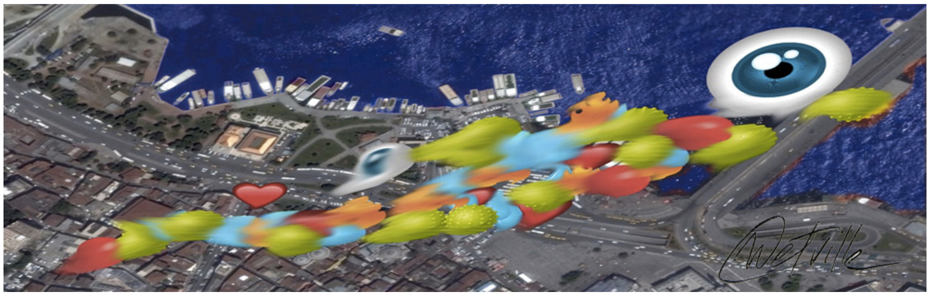
Proceedings



ESSEM 2015

2nd International Workshop on

Emotion and Sentiment in Social and Expressive Media

Opportunities and challenges for emotion-aware multiagent systems

Edited by

Cristina Bosco
Erik Cambria
Rossana Damiano
Viviana Patti
Paolo Rosso

May 5, 2015, Istanbul, Turkey
A workshop of AAMAS 2015

Preface

The 2nd International Workshop on Emotion and Sentiment in Social and Expressive Media (ESSEM 2015[1]) is taking place on May 5, 2015, in Istanbul as a workshop of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015). Emotions play a key role in the interactions that occur in a multi-agent system. Relevant perspectives include, on the one hand, research on architectures and cognitive models, which is concerned with the integration of emotional states into agents and the role of emotions in agent communication; on the other hand, research on techniques for sentiment analysis and opinion mining, devoted to automatic processing of affective information conveyed by spontaneous, multi-faceted user responses about shared contents. The main goal of the workshop is to attain cross-fertilization between the two perspectives. While the former relies mainly on cognitively inspired agent models, the latter relies on the use of statistical and learning techniques supported by resources such as corpora and linguistic datasets. By proposing ESSEM 2015 as an AAMAS workshop, we intend to stimulate a tighter integration of agent-based paradigms with techniques for sentiment and opinion mining, which have raised a growing interest in a social web and big data perspective. Since they typically involve interaction and feedback as part of their functioning, social and expressive media – ranging from online communities, blogs and fan-generated narratives to interactive art installations – will provide an effective testbed for:

- integrating complementary aspects: emotion generation and detection, affect expression and reception;
- studying the role of affect in the generation of social behavior, and communicative behavior in particular.

As the paradigmatic situation, we envisage a range of applicative contexts that involve a *performance* of some kind, targeted to some audience/user. The performance/reception paradigm, be it a digitally mediated co-creation session or the contribution of a user in a social forum, provides a conceptual framework for studying affect generation and detection in an integrated approach. Reception can take the form of immediate feedback, possibly regulated by a protocol, or of asynchronous response (e.g. tags and comments) conveyed through a plurality of media, against a background shaped by reputation and trust as a relevant part of the audience value systems.

Given the leading thread described above, we have proposed a special focus for ESSEM 2015: opportunities and challenges for emotion-aware multiagent systems. The final program includes four full papers (out of the eight submissions that were reviewed in the full paper category) and six short papers, which all in all cover different topics of the call for papers and represent an interesting variety of point of views on the ESSEM themes. The members of the

[1] `http://di.unito.it/essem15`

program committee did an exceptional job, managing to complete the review process in record time, by providing authors with three reviews per full/short paper, we would like to thank them for their accurate work. We thank our invited speakers *Ana Paiva* (INESC-ID, Portugal) and *Gülşen Eryiğit* (Istanbul Technical University, Turkey) for accepting to deliver the ESSEM keynote talks for this edition, crucial contributions towards a successful cross-fertilization of ideas between the agent and NLP communities about emotion and sentiment themes. We also thank *Catherine Pelachaud* (panel chair, CNRS, TELECOM ParisTech, France), *Chloé Clavel* (Telecom-ParisTech, France), *Emiliano Lorini* (IRIT-CNRS, Toulouse, France), *Munindar P. Singh* (North Carolina State University, USA), *Gualtiero Volpe* (University of Genova, Italy) which accepted to animate our final discussion panel. We would like to express our gratitude for the official endorsement we received from the CELI Torino, CIRMA and WIQ-EI (Web Information Quality Evaluation Initiative). Furthermore, we are grateful to all authors who submitted their works to ESSEM 2015.

We sincerely hope that ESSEM 2015 could be the occasion for merging the the agent community and the sentiment analysis and opinion mining community, both interested in emotions and sentiments in social and expressive media, although from complementary perspectives. Still many are the open questions for the two research communities. Does the progress in sentiment analysis and opinion mining also bring opportunities for modeling the affective component of agent behaviour? Are technologies for automatic sentiment analysis mature enough to allow for a tighter integration with emotions in modelling feedback in expressive media? Which affective frameworks should be considered in a perspective that integrates both the affective information extracted from user responses and the affective agent behavior to be generated? Are there any relationships between the analysis of emotion and sentiment in social and expressive media and the captology research field, e.g. the study of computers as persuasive technologies? Hopefully, some of these questions will have a tentative answer in the context of the ESSEM 2015 workshop.

April 2015

Cristina Bosco
Erik Cambria
Rossana Damiano
Viviana Patti
Paolo Rosso

Organization

Program Chairs and Organizers

Cristina Bosco	University of Turin, Italy
Erik Cambria	Nanyang Technological University, Singapore
Rossana Damiano	University of Turin, Italy
Viviana Patti	University of Turin, Italy
Paolo Rosso	Universitat Politècnica de València, Spain

Program Committee

Alexandra Balahur	European Commission Joint Research Centre, Italy
Cristina Battaglino	University of Turin, Italy
Paula Carvalho	INESC-ID and ISLA Campus Lisboa, Portugal
Marc Cavazza	Teesside University, UK
Chloé Clavel	Telecom-ParisTech, France
Morena Danieli	University of Trento, Italy
Dipankar Das	Jadavpur University, India
Mehdi Dastani	Utrecht University, the Netherlands
Berardina Nadja De Carolis	University of Bari Aldo Moro, Italy
Elisabetta Fersini	University of Milano-Bicocca, Italy
Giancarlo Fortino	University of Calabria, Italy
Carlos A. Iglesias	Universidad Politécnica de Madrid, Spain
Anup Kalia	North Carolina State University, Releigh, USA
Iolanda Leite	Technical University of Lisbon, Portugal
Emiliano Lorini	IRIT-CNRS, Toulouse, France
Viviana Mascardi	University of Genova, Italy
Alessandro Moschitti	University of Trento, Italy
Malvina Nissim	University of Groningen, the Netherlands
Nicole Novielli	University of Bari Aldo Moro, Italy
Antonio Origlia	University of Naples Federico II, Italy
Roberto Paredes	Technical University of Valéncia, Spain
Paolo Petta	Austrian Research Institute for AI, Austria
Daniele Radicioni	University of Turin, Italy
Francisco Rangel	Autoritas Consulting, Spain
Hassan Saif	The Open University, UK
Bjoern Schuller	Technical University of Munich, Germany
Giovanni Semeraro	University of Bari, Italy
Michael Thelwall	University of Wolverhampton, UK
Marko Tkalcic	Johannes Kepler University, Austria
Emilio Vivancos	Technical University of Valéncia, Spain
Gualtiero Volpe	University of Genova, Italy
Enrico Zovato	Nuance Communications, Italy

Publicity Chair

Cristina Battaglino	University of Turin, Italy

Endorsements

CIRMA, Università di Torino
CELI s.r.l Torino
WIQ-EI (Web Information Quality Evaluation Initiative)

Table of Contents

Workshop ESSEM 2015

Invited Talks

Full Papers

Short Papers

Agents and Robots that Can Listen to the Users Heart

Ana Paiva

University of Lisbon
GAIPS, INESC-ID and Instituto Superior Tecnico
Lisboa, Portugal
ana.paiva@inesc-id.pt

Abstract Empathy is often seen as the capacity to perceive, understand and experience others emotions. This notion is often seen as one of the major elements in social interactions between humans. As such, when creating social agents that are believable and able to engage with users in social interactions, empathy needs to be addressed. For the past few years, many researchers have been looking at this problem, not only in trying to find ways to perceive the users emotions, but also to adapt to them, and react in an empathic way. This talk will provide an overview of this new challenging area of research, by analyzing empathy in the social relations established between humans and social agents (virtual and robotic).

Short Biography Ana Maria Paiva is a Full Professor in the Department of Computer Science and Engineering (Departamento de Engenharia Informática) of Instituto Superior Técnico from the Technical University of Lisbon (Universidade Ténica de Lisboa). She is also the group leader of GAIPS, (Grupo de Agentes Inteligentes e Personagens Sintéticas), a research group on agents and synthetic characters at INESC-ID. Her main scientific interests lay in the area of Autonomous Agents, Embodied Conversational Agents and Robots and Multiagent Simulation Systems. Prof. Ana Paiva has been researching in the area of artificial intelligence for the past twenty years, having also taught at IST during that period

Parsing Web2.0 Language and Sentiment Analysis: The case of Turkish

Gülşen Eryiğit

Istanbul Technical University
Department of Computer Engineering
34469 Istanbul, Turkey
gulsen.cebiroglu@itu.edu.tr

Abstract With the increasing number of people using micro blogging sites (such as Facebook, Twitter, Youtube, Instagram, Foursquare or Google+), social media data became a highly attractive source for machine learning and natural language processing (NLP) research employed in many high-tier applications including social network analysis, information extraction, web mining, opinion mining and brand monitoring. The language used in social media differs severely from formally written texts, in that people do not feel forced to write grammatically correct words or sentences. In this talk, we are going to present our current research on Parsing Web2.0 Language and Sentiment Analysis for a morphologically very complex language: Turkish.

Short Biography Gülşen Eryiğit is an Assistant Professor in the department of Computer Engineering at Istanbul Technical University (ITU), Turkey. She is a founding member of the ITU Natural Language Processing Group and a member of the ITU Learning from Big Data Group. Her current research focuses on natural language processing of Turkish. In this field, she acted as a reviewer or co-author of various publications in prestigious journals and conferences. She was the Turkeys representative in CLARIN (EU 7th Framework Programme, CLARIN - Common language resources and technology infrastructure). Currently, she is the leader of two ongoing research projects: one EU Cost Action project "Parsing Web2.0 Sentences" funded by the Scientific and Technological Research Council of Turkey (TUBITAK) and one national project "Turkish Mobile Personal Assistant" funded by Turkeys Ministry of Science, Industry and Technology. She also serves as the NLP coordinator of one interdisciplinary research project "A Signing Avatar System for Turkish to Turkish Sign Language Machine Translation" funded by TUBITAK. She is serving as a consultant or technical advisor to several IT companies in Turkey mostly for brand monitoring and Turkish sentiment analysis projects. She has been a reviewer for many research and industrial proposals for European Commission (H2020 and Cost Action Programs), TUBITAK, Turkey's Ministry of Science, Industry and Technology and Swiss National Science Foundation.

Towards a Formal Specification of Moral Emotions

Alexander Pankov and Mehdi Dastani

Department of Information and Computing Sciences
Utrecht University, The Netherlands
a.pankov@students.uu.nl , m.m.dastani@uu.nl

Abstract. We propose a semi-formal specification of the elicitation conditions and prototypical coping strategies for three of the moral emotions: anger, contempt and disgust. We utilize existing psychological theories – appraisal theories of emotion and the CAD triad hypothesis – and incorporate them into a unified framework. Key features of the approach, such as its dynamic and epistemic natures, allow for modeling qualitative, quantitative and dynamic aspects of the moral emotions. We show that successful conceptualization is not only possible, but can shed light on the rationality behind moral emotions, as well as their importance to building socially aware agents.

Keywords: moral emotions, appraisal theories, coping strategies, formal specification

1 Introduction

Moral emotions respond to violations of internalized moral rules, and motivate morally congruent behavior [17,43]. According to Gewirth, the main characteristic of a moral rule is that it must bear on the interests or welfare either of society as a whole or of individuals other than the judge or agent [16]. Therefore, moral emotions are viewed as having two prototypical features: disinterested elicitation conditions (self having no direct stake in the triggering even) and pro-social action tendencies (benefiting others or the social order) [17]. According to the CAD triad hypothesis, and supported by experimental evidence [36], three moral emotions – *contempt*, *anger* and *disgust* – are typically elicited, across cultures, by violations of three specific categories of moral rules advocated by Richard Shweder: ethics of *community*, *autonomy* and *divinity* [38]. Furthermore, there are reasons to think that emotions in general, and moral emotions in particular, play important role in rational behavior [39], healthy mental life [44], and in maintaining social and moral norms [12,16,30,4] within societies.

Although there have been many efforts in the Artificial Intelligence (AI) community to provide a precise specification of emotions [10,9,23,24,42,41], there have not been, to our knowledge, a precise specification dedicated to these three moral emotions and their role in dealing with moral transgressions. The aim of our work is to propose a semi-formal specification of the appraisal and coping process involved in the *other-condemning* – about actions or character of others – moral emotions: anger, contempt and disgust. We focus on these three emotions due to their overtly social nature (being concerned with other agents), and, as a consequence, their potential to influence others' behavior. The choice of Shweder's ethics as the underlying moral theory is warranted by the convincing experimental evidence showing a one-to-one correspondence, across different cultures, between Shweder's ethics and the three emotions under discussion [36]. The proposed specification will, first, allow to operationalize and build emotionally aware software agents with applications ranging from improving education in virtual environments to social media analysis, and building believable video game characters. We note that social media as a unique public, and at the same time, virtual, environment, can be especially useful in analyzing the role emotions play in real human behavior. Second, the specification allows us to analyze how humans and other animate subjects may experience emotions, and how their mental structures change as a consequence. This second aspect enables researchers to disambiguate informal emotion theories and simulate hypothetical situations (morally impossible otherwise) and analyze complex psychological processes, such as aggression, depression and others that have been related to specifics in the appraisal and coping processes. Moreover, it is interesting to see if such formal model can shed light on the rationality and predominance of cooperative, morally congruent, behavior: it will be suggested that coping with moral emotions affects the adoption of goals promoting sanctioning of moral violations - a mechanism for maintaining and reinforcing the social status of moral rules. Last, but not least, the proposed specification is the first step towards a logical formalization of these emotions and can fuel future work by providing a framework in which other emotions can be analyzed.

The approach will be in the spirit of dynamic [14] and belief–desire–intention (BDI) [7,31] models, and, as a result, will provide a cognitive model of intelligent agents capable of experiencing and coping with socially-grounded emotions. The main theoretic and empirical support from cognitive psychology will be the appraisal and coping theories of emotion [22,15,28,21,37], as well as the CAD triad hypothesis [36,17]. Such a support cast – especially appraisal theories – have shown promise in explaining the relationship between social norms and emotions [40], and will now be applied to the domain of behavior triggered by moral emotions. According to these theories, the essential relationship between moral emotions and behavior is in the content of the agent's attitudes behind the emotion. Different categories of attitudes (such as those concerned with Shweder's ethics) lead to different emotions and behaviors. This matches perfectly with the BDI paradigm of modeling intelligent agents as entities possessing (uncertain) beliefs about the world, and aiming at desirable state of affairs by means of deliberation and action.

In what follows, we first present in Sect. 2 an overall mechanism for coping with the other-condemning moral emotions (i.e., anger, disgust, contempt). Then, in Sect. 3, 4 and 5 we provide a detailed description of each of the three emotions, together with a semi-formal specification of their elicitation and common coping strategies. Finally, Sect. 6 delivers concluding remarks on the results of this endeavor.

2 Mechanism for Coping with Moral Transgressions

At the outset, we asserted that the main trigger for an other-condemning moral emotion is a moral transgression. We now ask what is the psychological mechanism accounting for the individual's appraisal and behavior when dealing with moral transgressions. We believe that an answer to this question, and a general account of the similarities and differences between the moral emotions, can be given based on a theory of emotion elicitation and coping. Following the literature on moral emotions [36,17,21] and the relation emotions have to norms in human [12] and artificial [8] societies, we propose the following basic mechanism.

The other-condemning moral emotions get elicited by violations of internalized moral norms. Depending on the category (e.g., community, autonomy, divinity) of the violated moral norm, and thus the specific appraisals involved, different type of moral emotion, requiring different coping strategies, occurs. In most cases a sanction-oriented behavior is promoted, for it alleviates the negative emotion by dealing with concerns that triggered it. As a consequence of this behavior, the status of the violated norm may be reinforced.

Further clarifications are due in order to make the above picture complete. We need to, first, be more explicit in defining the conditions under which moral emotions occur: their general elicitation conditions, and the psychological appraisals behind Shweder's ethics [1]. Second, we need to describe the coping dynamics involved in the moral emotions in such a way that they actually make sense in light of the sanctioning behavior alluded to.

Let us, first, illustrate the proposed mechanism by means of a popular example from the domain of social media: *trolling*. Trolling is usually defined as a provocative behavior of posting inflammatory, offensive, or off-topic messages, and as quite similar to the concepts of flaming and cyberbullying. A troll, in that context, is the agent performing such behavior. There are several recent studies from the psychological literature that provide inside on the cognitive content behind trolling. First, a positive correlation between trolling behavior and personality traits such as sadism (strongest), psychopathy, and Machiavellianism have been shown [5]. Some of these traits have been associated with inability or unwillingness to follow social norms [6,18]. Second, a study have shown a strong correlation between the inflammatory (flaming) nature of trolling and unfairness, harm, and anger [20]. Finally, in popular culture trolling is said to "promote antipathetic emotions of disgust and outrage" [32]. From all this we conclude that trolling can serve as an interesting testbed for our model of the

[1] Here we adhere to Shweder's ethics; however, it should be clear that any such distinction based on the norm content will keep the overall coping mechanism more-or-less intact. What will change are the types of concerns (virtues) involved in the elicitation conditions

moral emotions. For example, imagine a participant in social media discussion posting a comment on a given topic, and receiving a trolling reply. In case the provocative comment is an offense aimed at the person who posted the original comment, then one would not be surprised if some of the participants react with *anger*, verbally attacking the offender or reporting him to the site administrators to be banned. Similarly, if the trolling comment simply uses foul language without attacking someone in particular, one would expect response of reporting or banning the *disgusting* offender; not trying to argue with him, as any such attempt might lead to more foulness. Finally, one can imagine a trolling comment that is not offending but simply off-topic. In such case, banning seems quite harsh and a more *contemptuous* reaction of ignoring the comment can be expected. In all cases, in accord with the proposed basic mechanism and the cited literature, trolling elicits in the participants an emotion condemning the behavior, and leads to behavior that promotes the agreed upon norm.

Couple of remarks are required before we proceed. Note that throughout we prefer using the term "coping strategies" [22] instead of "action tendencies" [15], although in most of the literature the two have been used interchangeably. The reason for this choice is the deliberative nature of the coping process, which gives it higher potential in modeling different behaviors. What is more important to our discussion, is that emotions in general, and moral emotions in particular, motivate behavior in a rational and predictable manner. Coping strategies capture, we think, successfully this quality of emotions, and give flexibility in explaining differences between moral emotions. Such flexibility comes mainly from the distinction between *belief-affecting*, *goal-affecting* and *intention-affecting* coping strategies (see [22] for the similar, but not crisp, distinction between problem-directed and emotion-directed coping). As the names suggest goal-affecting coping strategies modify directly the desires of the agent, whereas belief-affecting strategies work on the level of beliefs (still being able to subsequently change the goals and behavior of the agent). Intention-affecting coping function by modifying directly the intentions (planned actions) of the agent.

It is also important to stress here, that we stay agnostic about the essence of moral rules or the process of their internalization (we point, however, to [11] and [1] for a discussion on these topics). What is of interest to us, is their agreed upon pro-social nature [16] and categorization based on content [38,36], the rest remains out of scope for this work.

In the next three sections, for each emotion in the other-condemning family, we first review the psychological literature on its elicitation conditions and typical coping strategies, then we analyze its moral flavor by identifying the content of the moral norm category being violated. Finally, we provide detailed definitions of the three other-condemning emotions, and provide a semi-formal specification of their elicitation conditions and coping strategies.

3 Anger

The first to provide systematic treatment of anger, with surprisingly strong cognitive flavor, was Aristotle. In his *Rhetoric,* he writes: "Anger may be defined as a belief that we [...] have been unfairly slighted, which causes in us both painful feelings and a desire or impulse for revenge." His definition points out some key features: the negative nature of anger, its provocation by slight, and its motivational power for aggression.

Elicitation In recent literature on emotion, anger has been viewed as the main motivator of aggressive behavior, and as triggered by the frustration or thwarting of a goal commitment (for an overview see [21, pp. 218]). In our trolling example, this amounts to saying that the original poster's wish to present and discuss his opinion without being offended has been thwarted by an offensive comment. This broad view has been refined by appraisal theories according to which *any* negative emotion can arise from goal incongruence, therefore, it is important to specify what makes the provocation of anger different from other negatively-valanced emotional states, such as sadness, guilt, remorse. To address this question, most appraisal theorists incorporate the agent's attribution of *blame* to another person [21,15]. As a result, blame towards someone else becomes necessary for anger, for without the attribution of blame we can expect emotion such as sadness instead of anger; and with attribution of blame, but towards oneself, we can expect, for instance, guilt or remorse.

What does it mean, however, to blame someone for his deeds? According to [21], blame is an appraisal based on *accountability* and imputed *control.* To attribute accountability is to know who caused the relevant goal-frustrating event, and to attribute control is to belief that the accountable agent could have acted differently without, therefore, causing the goal-incongruence. Therefore, to blame, instead of simply hold someone responsible, is to think that the blameworthy agent could have acted otherwise. The difference is apparent in the case of trolling, where the person posting the offensive comment could, obviously, have refrained from commenting.

Obviously, attribution of blame is crucial to the elicitation of anger, but is it all there is to it? Lazarus argues that secondary appraisal processes can favor "maximizing the possibilities of success" in coping with the threatening situation, and therefore, influence which emotion gets elicited. According to him (1) if *coping potential* (evaluation of the possibility to actualize personal commitments) favors attack as viable, then anger is facilitated; and (2) if future expectancy is positive about the environmental response to attack, then anger is facilitated. Similarly, [37] writes about the coping ability of the agent in terms of an appraisal of power (availability of resources to act and anticipated effort) and adjustment ability (possibility/cost of changing/dropping goals). Both theorists seem to refer to the same mechanisms which we will group under the title of coping potential, a type of secondary appraisal, to use Lazarus' term.

Coping Most psychologists agree that the innate coping strategy in anger is *aggression* towards the blameworthy agent [2,3]. Frijda calls the action tendency (in his terms) underlying aggressive behavior "agonistic" [15, pp. 88]. Supposedly, such behavior includes *attack* and *threat* as actions, with the goal being the removal of the obstruction that caused anger. However, secondary appraisal influences the selection of strategies of attack, and they can differ greatly in content [21, pp. 227]. Furthermore, when planning an attack the agent chooses between types of attack (e.g., verbal versus physical, or punishment versus warning) based on coping potential. For instance, in our trolling example, the participant's decision to report the post to an administrator is based on the evaluation of his inability to argue with the offender: an estimate of his coping potential

From this we can conclude that in most cases of anger, the applied coping strategy aims at attacking the cause of goal-incongruence (intention-affecting coping) instead of re-appraisal (belief-affecting coping). The main reason for this seems to be the nature of anger: it gets promoted in cases when attack is viable and aggression needed [21, pp. 226, Table 6.1].

Moral anger Anger is usually viewed as an immoral emotion, but in many instances it is actually triggered by moral concerns. Of course, it does not mean that anger is always a moral emotion. For instance, consider a modified social media scenario where someone creates a post considered offensive by someone else. In this case, that someone else, can rightfully be angry because of the appraised offense, without any of his moral views being offended.

Moral anger, on the other hand, is a type of anger that arises when *harm* has been done to someone else and his rights have been violated [30, pp. 70]. The relationship between this definition and Shweder's ethics of autonomy has been demonstrated in [36] (as part of the CAD triad hypothesis). As already mentioned in our discussion on the psychological mechanisms behind the moral emotions, Shweder's autonomy norms are best seen as norms pertaining to harm against persons. [38, pp. 98] write: "The ethics of autonomy aims to [...] promote the exercise of individual will in the pursuit of personal preferences." Combining this aspect of moral anger with the elicitation conditions of core anger, allows us to define moral anger in psychological terms.

Elicitation (moral anger): *Displeasure from thwarting of a personal goal aimed at preserving the autonomy of agents, combined with attribution of blame for the goal-thwarting state of affairs to another agent, and an estimate of one's own coping potential as favoring punishment of the blameworthy agent.*

Coping (moral anger): *Intention-affecting strategies aimed at sanctioning the blameworthy agent by means of attack or threat.*

3.1 Anger: Appraisal Specification

Assuming φ as denoting a state of affairs, we use $Control_i(\varphi)$, which should be read as "agent i has control over φ", and define it as there exists an action such that φ will be false after agent i executes the action. In other words, "agent i

can prevent φ from being true". An instance of the $Control_i(\varphi)$ formula can be $Control_{troll}(discussNoOff)$, where $troll$ denotes the agent from our trolling example, and $discussNoOff$ denotes the state of affairs where discussion proceeds with no offenses.

Moreover, we use $Account_i(a, \varphi)$, which should be read as "agent i is accountable for (caused) φ by doing a", and define it as φ is true now and was not true before i performed a.[2] Again, $Account_{troll}(offComment, \neg discussNoOff)$ can be an instance of this formula. Control and accountability, as defined here, are not viewed as epistemological but as ontological concepts representing causal relationships between events. It is their appreciation by an agent that provides the necessary inside on the agent's epistemic state, including his attribution of blame. Although similar concepts have been previously analyzed from a logical perspective [25], here we only focus on their role in anger and contempt.

Therefore, we can now define $Blame_{i,j}(a, \varphi)$, which should be read as "agent i blames agent j for doing a and causing φ", as agent i believes that agent j is accountable for φ by doing a, and that before doing a, j had control over φ. Finally, before defining anger, we need a way of talking about the practical possibility of an agent to realize a state of affairs. In our example, this can be understood as a participant being able to restore the no-offense nature of the discussion, by say, reporting the offender and leading to the removal of the offensive comment. For this we use $Pos_i(\varphi)$, which should be read as "there is a practical possibility of agent i to make φ true", and define it as there exists an action a, such that if performed by i, φ will be true (e.g., $Pos_{obs}(discussNoOff)$).

We now introduce $Anger_{i,j}(a, \varphi)$, which should be read as "agent i is angry at agent j for doing a and preventing i from achieving φ as planned", and define it as agent i has an achievement goal φ, blames agent j for performing action a, thereby preventing i's plan from achieving φ, and believes there is still a practical possibility of achieving φ. For example, a participant in a social media discussion can be angry at the troll for posting an offensive comment and preventing the discussion (i.e., $Anger_{obs,troll}(offComment, discussNoOff)$).

In this specification, the achievement of goal φ captures the prototypical feature of all emotions, i.e., to be about a desired goal state. Thwarting this goal, as expected for a negatively-valanced emotions, is represented as the agent's belief not to be able to achieve his goal as planned, although agent i believes this was possible before action a was performed by agent j. The belief of i about the practical possibility for achieving φ by some other, not considered before, means highlights the positive evaluation by the agent of his coping potential–the type of secondary appraisal claimed to be an indispensable part of anger.

Proceeding to moral anger, we reassert that it is a flavor of anger with its content related to other agents and their autonomy. Autonomy was reduced to exercise of individual choice in the pursuit of personal preferences. We surmise that the concept of *harm* captures this meaning: preserving one's autonomy means not harming him. Although there are different types of harm distinguished in the literature [27,19], what they all have in common is the violation of per-

[2] We assume that only one agent acts at each moment in time.

sonal preferences by others. In case of physical harm, we can say the personal preference is for protecting one's own body. In case of psychological harm, the personal preference can be viewed as about (not) having certain types of beliefs.

We use $Harm_{i,j}(a, \varphi)$, which should be read as "agent i harmed agent j by doing a and preventing him from achieving φ", and define it as i is accountable by doing action a for j not being able to achieve his goal φ. For example, the troll harmed the original poster by posting an offensive comment and preventing him from discussing the topic without being offended (e.g., $Harm_{troll,poster}(offComment, discussNoOff)$. This definition is quite similar to the one for anger, for we can view anger as triggered by harm to oneself.

We now specify moral anger $MAnger_{i,j,k}(a, \varphi, \psi)$, which should be read as "agent i is morally angry at j for harming k by doing a, preventing k from achieving ψ and preventing i from following his moral norm φ", and define it as 1) $Anger_{i,j}(a, \varphi)$ (i.e., agent i is angry at agent j for doing a and thereby preventing him from achieving the moral norm φ), and 2) agent i believes $Harm_{j,k}(a, \psi)$ with $\varphi \rightarrow \psi$ (i.e., φ being the case requires ψ to be the case as well). Note that we refer to i's goal φ as a moral norm, for it implies no harm to k, therefore preserving k's autonomy, one of the moral categories according to Shweder. However, what matters for the elicitation of moral anger is φ's relation to the autonomy of agents. It is this relation with the autonomy of agents that gives a moral accent to φ, i.e., the preservation of agents' autonomy is considered as a moral rule.

We can see how the above definition captures our analysis of the concept of moral anger, namely as a type of anger with content related to harm done to someone else. Here the formula $Harm_{j,k}(a, \psi)$ represents the harm aspect of moral anger, whereas $\varphi \rightarrow \psi$ captures the logical relationship between the internalized moral rule φ and the violated personal preference ψ.

To illustrate, let us again take our social media example. In its first case, that of directly offending a participant of an online discussion, k from our definition could become the agent posting the original comment, j could be the troll and i can be the observing participant (experiencing the moral anger). Furthermore, for this scenario, ψ could be the original poster's wish to present and discuss his opinion without being offended, φ could represent the "no-offensive language" rule of conduct when posting comments, and the action a would be the actual act of posting an offensive comment. All to the effect of the following moral anger being elicited: $MAnger_{obs,troll,poster}(offComment, noOffLang, discussNoOff)$.

3.2 Anger: Coping Specification

The elicitation of anger – including moral anger – commonly leads to behavior targeted at resolving the psychological tension that triggered it. In our model this amounts to an intention-affecting coping strategy aimed at removing anger preconditions. The prototypical action is attack towards the blameworthy agent.

Furthermore, moral anger is elicited by violation of the autonomy of other agents. We reduced the concept of autonomy to that of harm. Therefore, we specify that coping with moral anger involves adopting the intention of performing an action a for which it is known to lead to $Harm_{j,k}(a, \psi)$ not being true.

This way successfully triggering the thus defined coping strategy removes the presence of moral anger – a property necessary for successful coping [22,44].

In our running example, this amounts to saying that in case of moral anger one should expected attacking behavior (banning, arguing) towards the trolling agent. This way the problem of harming the original poster will be mitigated by allowing the discussion to continue or defending the character of the poster.

4 Disgust

Disgust is an emotion that, from an evolutionary perspective, can be viewed as based on *distaste* - a term referring to the sensory-motor functions of smelling and tasting. Similar to anger, it has simpler (core disgust) and more complex (moral disgust) forms [35]. Research on disgust has gained popularity in the 1990s with some of the main contributors being Paul Rozin and his colleagues [33,34,35].

Elicitation Disgust is considered a response both to physical objects and to social violations [35,28,17]. Lazarus unites the physical and social aspects of disgust by defining it as "taking in or being too close to an indigestible object or idea (metaphorically speaking)"[21, pp. 260]. This and other definitions [28,33] focus on the mouth and dislike towards physical objects, and then suggest that some class of non-physical objects can cause a similar feeling. Furthermore, [35] argue that disgust grew out of a distaste response found also in other animals, which was then shaped by evolution to become a more generalized “guardian of the temple of the body”. Thus, getting coupled to, and triggered by, motivation to protect oneself from any sort of *contamination*, including of ideas. Contamination, in this discussion, will have one important property: an agent gets contaminated by coming into contact with another contaminated agent.

Coping All forms of disgust include a motivation to avoid, expel, or otherwise break off contact with the offending entity, often coupled to a motivation to purify, or otherwise remove residues of any physical contact that was made with the entity [35]. This motivation is clearly adaptive when dealing with potentially lethal contamination of food, but it appears to have made the transition into our moral and symbolic life as well [35]. Thus making moral disgust (see below) a powerful drive for action when dealing with norm violations.

As with anger, coping with disgust usually requires intention-affecting (action-directed) strategies to achieve the required result, purity. This does not mean that belief-affecting strategies are not possible, but that in most cases actions are required to deal with the feeling of disgust.

Moral disgust The variation of disgust, called moral disgust, is triggered by people who violate local social rules for how to use their bodies, particularly in domains of sex, drugs, and body modification [17]. Rozin and his colleagues have demonstrated that moral disgust derives from physical disgust by showing that it has the same bodily basis and the same logic of contamination: we do not like to have contact with objects that have touched a person we deem morally disgusting [35]. For example, we would not like to live in the former home of a condemned pedophile, or, following our running example, we would not like to argue with a person posting only comments containing foul language.

Furthermore, according to the CAD triad hypothesis [36], we can make a link between disgust and Shweder's ethics of divinity: social norms concerning the natural order. What follows is that disgust gets triggered by violations of such norms. In explaining the ethics of divinity, [38] write: "[T]he ethics of divinity protect the soul, the spirit, the spiritual aspects of the human agent and nature from degradation." Interestingly, none of the moral transgressions under the "divinity" label used in forming the CAD triad hypothesis [36], have to do with religious violations. Thus, we conclude that the name of this category should not be taken literally, instead, it should be understood as referring to purity and the natural order of things - with the divine being an instance of the natural order. Our methodology, then, requires us to combine this result with the standard appraisal theory account of the elicitation and coping with disgust, resulting in the following definition.

Elicitation (moral disgust): *Displeasure from the thwarting of a personal goal aimed at protecting the perceived natural order among agents, including protecting against contamination.*

Coping (moral disgust): *Intention-affecting strategies aimed at avoiding, expelling, or otherwise breaking off contact with the offending entity.*

4.1 Disgust: Appraisal Specification

Here we apply more-or-less the same strategy as with anger: use primitive concepts such as goals, beliefs and actions together with the more complex one, the appraisal of accountability. The difference will be in introducing the special atoms C_i, which should be read as "agent i is contaminated".

We use $Disgust_i(a, \varphi)$, which should be read as "agent i is disgusted from experiencing a which caused φ", and define it as agent i has an avoidance goal φ, believes a to have caused φ, and believes that φ leads to the contamination state C_i. Again, the avoidance goal φ captures the prototypical feature of any emotion: to be about a(n) (un)desired state φ, whereas a and C_i capture the property of disgust of being about a kind of contamination of the agent.

As was the case with anger and its moral flavor, moral disgust is actually a type of disgust, with the moral aspect coming from concerns about the actions of others. Therefore we use $MDisgust_{i,j}(a, \varphi)$, which should be read as "agent i is disgusted from agent j doing a which caused φ", and define it as agent i has an avoidance goal φ, believes j to have caused φ by doing a, and believes that φ leads to the contamination state C_i. Here, due to the generality of the definition, there is no need of specifying a third agent, as we did with moral anger, for the appraised contamination triggering disgust can be on any object, not necessarily an agent.

Applying the above definition to our running example should clarify. If the trolling comment from the example contained language considered foul (dirty) by some participant, he is expected to be disgusted by it. In our definition this amounts to saying that j is the troll, i is the participant reading the nasty comment, a is the action of posting a comment containing fault language, and φ expresses i's exposure to dirty language. Then, from assuming that i does not want to be exposed to dirty language, it directly follows that i would experience disgust towards the troll and his comment, which is expressed by the fact $MDisgust_{obs,troll}(foulComment, foulLang)$. In this case the contamination we talk about is purely one of contamination of ideas, but this, as we stated before, is to be expected for the moral flavor of disgust.

4.2 Disgust: Coping Specification

The prototypical coping strategy when dealing with disgust is an intention-affecting strategy to try and expel the source of contamination.

An agent i feeling disgust from doing a will try performing an action (e.g., expelling the source of contamination) if he thinks it will remove the contamination itself (i.e., C_i). As defined, this coping strategy trigger applies to core disgust. However, having in mind that moral disgust is a type of disgust after all, we see that such a coping strategy would work for the moral variant as well.

Finally, in our trolling example with foul language and elicited disgust, one should expect actions that somehow prevent further contamination. This includes reporting/banning the offending agent, but not arguing with him, for this will only cause further contamination.

5 Contempt

Contempt is one of the least discussed emotions in the psychological literature [17, Table 1]. If research on the facial expression of contempt is excluded, there is almost no other empirical research on contempt. In most discussions it falls in between anger and disgust, and is sometimes said to be a blend of the two [29], folded into the anger family [21], or else said to be part of anger [28]. Here, however, it is discussed separately because of its important role as the only moral emotion from the other-condemning family not having a core/immoral variant: all instances of contempt are triggered by violations of social - in most cases, moral - norms related to obeying social hierarchies.

Elicitation For our discussion we adopt the view that contempt is part of the reproach emotions family, and is elicited by disapproving of someone else's *blameworthy action* [28, pp. 145]. This is quite similar to what we said about the triggering conditions of anger. This is also the reason why [28] see anger's elicitation conditions as a blend between those of a reproach emotion (such as contempt) and a negative event-based emotion (such as distress). [28] emphasize, however, that anger is not a compound emotions, instead its elicitation conditions have an overlap with those of distress and contempt.

As stated in the introduction, there is evidence [36] for the relation between contempt and violations of Shweder's ethics of community [38]. Shweder writes [38, pp. 98]:

> The ethics of community [...] aims to protect the moral integrity of the various stations or roles that constitute a society or community

The main concepts discussed by [38] regarding the ethics of community are those of *hierarchy* and *duty*. Detailed account of hierarchy and duty in societies is not the aim of this work, however, we suggest these two concepts can be abstracted away in a meaningful way. Hierarchy we consider to be a set of roles, which define a special kind of relation between agents. We call it a *social significance* relation, and should be seen as a relation capturing the potential effects of one's actions on the wellbeing of others', or society as a whole. Violations of one's duties are then indicated by this relation for each possible situation. Such an abstraction, we think, covers the basic cognitive content behind roles and duties, and can serve us in conceptualizing contempt. For example, when participating in social media discussions, one can distinguish two roles: the poster of the original comment and the participant. Their relationship (it terms of social hierarchy and duties) can then be captured by a mechanism to indicate if each action performed is a violation of the duties (e.g., following the topic, writing in the same language) derived from the these two roles.

Coping Contempt motivates neither attack nor withdrawal; rather it seems to cause social-cognitive changes such that the object of contempt will be treated with less *warmth*, *respect*, and *consideration* in future interactions [26]. We are sure there is a lot one can say about these concepts, but we simplify the matter by stipulating that warmth, respect and consideration all supervene on the perceived social significance of the other agent. Thus, less (more) perceived social significance means less (more) warmth, respect and consideration in future interactions. As a result all belief changes for coping with contempt become bound to reduction of the level of belief in the "social significance" of the other agent. In our running example this would amount to saying that in response to off-topic comment by an agent, participants will change their appreciation of the importance that participant has to the discussion. His role, including his and others' duties, during the discussion will change.

Note that contempt offers the first example of a belief-affecting coping strategy among moral emotions. This makes contempt significantly different than moral anger and moral disgust. However, we argue that despite its "passive" nature, contempt is still capable of reinforcing the social status of moral norms by indirectly sanctioning moral violators. The corresponding mechanism goes much in the spirit of [13]: becoming aware of others' disapproval, can cause negative emotion (shame) in the subject. Therefore, coping with contempt can lead to epistemic changes that can stimulate the expression of disapproval, which can trigger negative feelings (e.g., shame) in the moral violator, which, on its own, can serve as a sanction for his behavior. Nevertheless, this shaming function, although important as a mechanism for reinforcing the status of social norms, will remain out of scope for our proposed framework. In what follows we will assume the following about contempt.

Elicitation (contempt): *Displeasure from the thwarting of a personal goal concerned with preserving the social hierarchy, combined with the attribution of blame for the goal-thwarting state of affairs.*

Coping (contempt): *Belief-affecting strategies for changing the level of the personal social significance of the blameworthy agent.*

5.1 Contempt: Appraisal Specification

Here we use the special atoms V_i and $Sig_{i,j}$ for talking about violations of duties by agent i, and social hierarchies (in this case agent j is significant to agent i), respectively. As stated above contempt is a negative emotion triggered by violation of a goal concerned with preserving the social hierarchy, together with the attribution of blame for the goal-thwarting state of affairs to someone else. The appraisal of blame has already been defined in previous sections and can be used directly. Preserving the social hierarchy will be modeled as an avoidance goal whose violation leads to breaking the social hierarchy by a significant other.

We now specify contempt $Contempt_{i,j}(a, \varphi)$, which should be read as "agent i is contemptuous towards agent j for doing a and making φ true", and define it as agent i has an avoidance goal φ, blames agent j for performing the physical action a, thus making φ true (i.e., $Blame_{i,j}(a, \varphi)$), believes j to be a significant other ($Sig_{i,j}$) and that φ violates a duty derived from the social hierarchy (i.e., $\varphi \rightarrow V_i$). The above definition captures several key components of contempt: goal-incongruence, violation of a norm concerned with preserving the social hierarchy and the attribution of blame. This attribution of blame is what contempt shares with anger and is why [28] have considered them similar.

As with the previous two emotions, let us see how this definition fairs with our running example. In terms of roles, it suffices to say again that there are two roles involved: poster and participant. Poster's duty is to start a topic by clearly stating a proposition, whereas the participant's duty is to contribute to that topic with his opinion or new information, but not to change it. Assuming this simplistic social structure, it becomes obvious how posting an off-topic (trolling) comment can trigger contempt: φ from the above definition becomes the norm of participants not changing the original topic and a the action of actually posting a comment that does: $Contempt_{obs,troll}(offComment, onTopic)$.

5.2 Contempt: Coping Specification

Contempt has the interesting characteristic of affecting one's appreciation of the other agent's social significance, without having direct influence on one's behavior [26]. We specify this prototypical coping strategy as agent i feeling contempt towards agent j will reduce his belief in the social significance of j (i.e., his belief in the formula $Sig_{i,j}$). Note that, although reduction in the social significance of the offending agent might also be possible when coping with anger or disgust, in our work we treat only prototypical coping mechanisms. Such reduction in the social significance is essential to coping with contempt, whereas it is not in the case of anger or disgust.

Again, by trying out this definition in our example, we see its immediate logic: dealing with off-topic comments (the trigger of contempt) involves ignoring them, instead of fighting them, which will only trigger some aggression and further pollute the discussion underway.

6 Conclusion

In this work we provide a semi-formal specification of the elicitation conditions and coping strategies of a set of socially-grounded emotions, dubbed moral. The specification is based on appraisal theories of emotion and the CAD triad hypothesis, and is grounded in a dynamic, multi-agent BDI framework. In this system, emotions are defined based on agents' actions and attitudes (including graded beliefs, goals and intentions). The moral aspect of the modeled emotions is based on Shweder's ethics, and is represented using concepts grounded in the agents' beliefs and goals. Coping strategies are represented as belonging to several categories depending on their effects on the attitudes of agents, and are applied using a triggering mechanism based on the elicitation conditions of the emotion, plus an estimates of their potential for alleviating the emotion that triggered them.

The result should be viewed as twofold. First, the current conceptualization contributes to building a precise ontology of emotions, by incorporating cognitive theories into existing intelligent agent models. Second, it paves the way towards building and analyzing emotionally and morally aware agents capable of coexisting in a dynamic multi-agent environment.

We consider this work as only the first step towards a complete formal specification and operationalization of the attitudes behind moral emotions. We intend to extend the set of emotions, as well as the variety of coping strategies in future work. Furthermore, we ignored some aspects of the coping process that may be important in implementing real-world scenarios. These include the concepts of coping power (availability of resources) and adjustment ability (possibility and cost of changing/dropping goals) found in the literature. An important point to be addressed in the future is a mechanism for triggering coping strategies using thresholds on the emotion intensity. A possible extension to the base formalism is the introduction of complex actions. In the present work moral rules have been modeled in a simplistic manner without representing their logical structure. Future work will address this by extending the base language with means of talking about norms and obligations.

References

1. Andrighetto, G., Villatoro, D., Conte, R.: Norm internalization in artificial societies. Ai Communications 23(4), 325–339 (2010)
2. Averill, J.R.: Anger and aggression: An essay on emotion (1982)
3. Averill, J.R.: Studies on anger and aggression: Implications for theories of emotion. American Psychologist, 38(1), 1145–1160 (1983)
4. Blackburn, S.: Ruling passions. Clarendon Press Oxford (1998)
5. Buckels, E.E., Trapnell, P.D., Paulhus, D.L.: Trolls just want to have fun. Personality and individual Differences 67, 97–102 (2014)
6. Cleckley, H.M.: The mask of sanity: An attempt to clarify some issues about the so called psychopathic personality. Aware Journalism (1964)
7. Cohen, P.R., Levesque, H.J.: Intention is choice with commitment. Artificial Intelligence 42(2–3), 213–261 (Mar 1990)
8. Conte, R., Castelfranchi, C.: Understanding the functions of norms in social groups through simulation. Artificial societies: The computer simulation of social life (1995)
9. Dastani, M., Lorini, E.: A logic of emotions: from appraisal to coping. Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 2 pp. 1133–1140 (2012)
10. Dastani, M., Meyer, J.J.C.: Programming agents with emotions pp. 215–219 (2006)
11. Dubreuil, B., Grégoire, J.F.: Are moral norms distinct from social norms?: A critical assessment of jon elster and cristina bicchieri. Theory and Decision 75(1), 137–152 (2013)
12. Elster, J.: Rationality, emotions, and social norms. Synthese 98(1), 21–49 (1994)
13. Elster, J.: Alchemies of the Mind. Cambridge Univ Press (1999)
14. Fischer, M.J., Ladner, R.E.: Propositional dynamic logic of regular programs. Journal of computer and system sciences 18(2), 194–211 (1979)
15. Frijda, N.H.: The emotions. Cambridge Univ Pr (1986)
16. Gewirth, A.: Reason and morality. University of Chicago Press (1981)
17. Haidt, J.: The moral emotions. Handbook of affective sciences pp. 852–870 (2003)
18. Hare, R.D., Hart, S.D.: chap. Psychopathy, mental disorder, and crime. Sage Publications, Inc (1993)
19. Helwig, C.C., Zelazo, P.D., Wilson, M.: Children's judgments of psychological harm in normal and noncanonical situations. Child Development 72(1), 66–81 (2001)
20. Johnson, N.A., Cooper, R.B., Chin, W.W.: Anger and flaming in computer-mediated negotiation among strangers. Decision Support Systems 46(3), 660–672 (2009)
21. Lazarus, R.S.: Emotion and adaptation. Oxford University Press, USA (1991)
22. Lazarus, R.S., Folkman, S.: Stress, appraisal, and coping. Springer Publishing Company (1984)
23. Lorini, E.: A dynamic logic of knowledge, graded beliefs and graded goals and its application to emotion modelling. Logic, Rationality, and Interaction pp. 165–178 (2011)
24. Lorini, E., Schwarzentruber, F.: A logic for reasoning about counterfactual emotions. Artificial Intelligence (2010)
25. Lorini, E., Longin, D., Mayor, E.: A logical analysis of responsibility attribution: emotions, individuals and collectives. Journal of Logic and Computation p. ext072 (2013)

26. Oatley, K., Johnson-Laird, P.N.: The communicative theory of emotions: Empirical tests, mental models, and implications for social interaction. Martin, Leonard L. (Ed); Tesser, Abraham (Ed), (1996). Striving and feeling: Interactions among goals, affect, and self-regulation. , (pp. 363-393). Hillsdale, NJ, England rtin, Leonard L(6), 363–393 (1996)
27. Ohbuchi, K.i., Kameda, M., Agarie, N.: Apology as aggression control: its role in mediating appraisal of and response to harm. Journal of personality and social psychology 56(2), 219 (1989)
28. Ortony, A., Clore, G.L., Collins, A.: The cognitive structure of emotions. Cambridge Univ Pr (1990)
29. Plutchik, R.: Emotion: A psychoevolutionary synthesis. Harper & Row New York (1980)
30. Prinz, J.: The emotional construction of morals. Oxford University Press (2007)
31. Rao, A.S., Georgeff, M.P.: Modeling rational agents within a bdi-architecture. KR 91, 473–484 (1991)
32. Redmond, S.: Celebrity and the media. Palgrave Macmillan (2014)
33. Rozin, P., Fallon, A.E.: A perspective on disgust. Psychological Review, 94(1), 23–41 (Jan 1987)
34. Rozin, P., Haidt, J., McCauley, C.R.: Disgust: The body and soul emotion. Dalgleish, Tim (Ed); Power, Mick J. (Ed), (1999). Handbook of cognition and emotion. , (pp. 429-445). New York, NY, US: John Wiley & Sons Ltd, xxi, 843 pp. doi lgleish, Tim(9), 429–445 (1999)
35. Rozin, P., Haidt, J., McCauley, C.R.: Disgust. Lewis, Michael (Ed); Haviland-Jones, Jeannette M. (Ed); Barrett, Lisa Feldman (Ed), (2008). Handbook of emotions (3rd ed.). , (pp. 757-776). New York, NY, US: Guilford Press, xvi, 848 p wis, Michael(8), 757–776 (2008)
36. Rozin, P., Lowery, L., Imada, S., Haidt, J., et al.: The cad triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). Journal of personality and social psychology 76, 574–586 (1999)
37. Scherer, K.R.: Appraisal considered as a process of multilevel sequential checking: A component process approach. Appraisal processes in emotion: Theory, methods, research 92, 120 (2001)
38. Shweder, R.A., Much, N.C., Mahapatra, M., Park, L.: The "big three" of morality (autonomy, community, divinity) and the "big three" explanations of suffering. Morality and health pp. 119–169 (1997)
39. Sloman, A., Croucher, M.: Why robots will have emotions. Proc 7th Int. Joint Conf. on AI (1981)
40. Staller, A., Petta, P.: Introducing emotions into the computational study of social norms: A first evaluation. Journal of artificial societies and social simulation 4(1), U27–U60 (2001)
41. Steunebrink, B., Dastani, M., Meyer, J.J.: A formal model of emotion-based action tendency for intelligent agents. Progress in Artificial Intelligence pp. 174–186 (2009)
42. Turrini, P., Meyer, J.J.C., Castelfranchi, C.: Coping with shame and sense of guilt: a dynamic logic account. Autonomous Agents and Multi-Agent Systems 20(3), 401–420 (2010)
43. Vélez García, A.E., Ostrosky-Solís, F.: From morality to moral emotions. International Journal of Psychology 41(5), 348–354 (2006)
44. Watkins, E.R.: Constructive and unconstructive repetitive thought. Psychological Bulletin 134(2), 163 (2008)

Towards an Empathic Social Robot for Ambient Assisted Living

B. De Carolis, S. Ferilli, G. Palestra, V. Carofiglio

Dipartimento di Informatica, Universita' di Bari "Aldo Moro", Bari, Italy
berardina.decarolis@uniba.it, stefano.ferilli@uniba.it,
giuseppe.palestra@uniba.it

Abstract. In the context of Ambient Assisted Living, assistance and care are delegated to the intelligence embedded in the environment that, in our opinion, should provide not only a task-oriented support but also an interface able to establish a social empathic relation with the user. This can be achieved, for instance, using a social assistive robot as interface towards the environment services. In the context of the NICA (Natural Interaction with a Caring Agent) project we developed the behavioral architecture of a social robot able to assist the user in the interaction with a smart home environment. In this paper we describe how this robot has been endowed with the capability of recognizing the user affective state from the combination of facial expressions and spoken utterances and to reason on in order to simulate an empathic behavior.

1 Introduction

One of the new trends in the context of Ambient Assisted Living (*AAL*) concerns the integration of new technologies with a social environment to support people in their daily activities increasing their quality of life [1,2]. In this view, an assistive home environment should provide not only a task-oriented support but also an interface able to establish a social empathic relation with the user. This is what we call a "caring home". Achieving this objective, in our opinion, requires developing:

- *Methods and models for defining and developing Ambient Intelligence (AmI) systems for Assisted Living* that are able to define environments that manage devices and services autonomously and proactively with respect to the needs of the users populating the environment.
- *Methods and models for analysis of the user behavior* with particular emphasis on affective aspects in order to achieve personalization, adaptation and proactivity that are typical of an AAL system.
- *Natural Interaction of the user* with the information and services offered by the system. Such an interface has two fundamental and interconnected objectives: being a means to interact with the environment and being, for the user, a friendly caring agent. For this reason it is important to understand not only the meaning of the communication but also the conveyed emotions and the user's attitude during the interaction. This requires the emotional analysis of the user's verbal and non-

verbal communicative acts (i.e. linguistic and prosodic aspects of the user's vocal input, facial expressions, postures and gestures).

In the context of the NICA (Natural Interaction with a Caring Agent) project we developed the behavioral architecture of a social robot able to assist the user in the interaction with a smart home environment [3]. In this paper we propose the use of a social empathic robot acting as a virtual caregiver. In particular we discuss how it has been endowed with the capability of recognizing the user's affective state from the combination of facial expressions and spoken utterances and of reasoning on it in order to simulate an empathic behavior.

The choice of a social assistive robot as an interaction metaphor is driven by the following considerations. If properly designed, social and conversational agents and robots may improve the naturalness and effectiveness of the interaction between users and systems [4]. They have the potential to involve users in human-like conversations using verbal and non-verbal signals for providing feedback, showing empathy and emotions in their behavior [5,6]. Indeed, several studies report successful results on how expressive conversational agents and robots can be employed as an interaction metaphor in the assisted-living domain and in other ones [7,8] where it is important to settle long-term relations with the user [9].

Empathy can be defined as "an affective response more appropriate to someone else's situation than to one's own" [10]. Then the expression of empathy aims at demonstrating that the other's feelings are understood or shared. Moreover, according to [11], empathy facilitates the creation of social relations. Empathic agents are perceived as more caring and trustworthy than neutral agents [12] and they can induce empathy in users [13]. In particular, the simulation of empathy in socially assistive robotics is supported by the findings of many psychologists showing that empathy plays a key role for therapeutic improvement and that empathy mediates pro-social behavior (e.g., [14,10]).

Taking these findings into account, we decided to endow a social assistive robot with the capability of recognizing the user affective state and attitude, reasoning on it and, consequently, deciding whether to trigger an empathic behavior toward the user. Moreover, in order to improve the long-term relation between the user and the robot, it keeps in its social memory information about which are the antecedents of emotions for the user, that is what triggers the emotions (events, situations, thoughts, etc.) in order to improve its empathic capability. These behaviors have being modeled according to the analysis of a corpus collected by human caregivers.

The paper is structured as follows: after providing an overview of the related work in Section 2, in Section 3 we show how the empathic behavior is simulated in the robot; in Section 4 a brief illustration of a case study is described; finally we conclude the paper with discussion and directions for future work.

2 Related Work

The main aim of Ambient Assisted Living (AAL) is to improve the life quality of elderly people who need special care and assistance by providing cognitive and physical support and access to the environment services [15]. Many of these projects besides developing technological platforms to monitor the health state and comfort of the user, provide natural and pleasant interfaces for interacting with the smart environment services. Several studies report successful results on how expressive conversational agents and robots can be employed as interaction metaphor in the assisted living domain. For instance, projects *ROBOCARE* [16], *Nursebot* [17], *Care-o-bot* [18], *CompaniAble* [19] and KSERA [20] aim at creating assistive intelligent environments in which robots offer support to the elderly at home, possibly having also a companion role. van Ruiten et al. [21] conducted a controlled study using *I-Cat* [22] in which they confirmed the results that, as shown in [23], elderly users like to interact with a social robot and to establish a relation with it. The reason of the success of socially intelligent agents and robots is due to the fact that interaction between human and machine has a fundamental social component [24]. Thus endowing social agents with user models that involve the consideration of both cognitive and affective components of the user state of mind is a key issue for enabling the adaptation of the agents behavior to both physical and emotional user's needs, as in the case of the simulation of the empathic behavior.

As far as simulating empathic behavior in social agents is concerned, there are several studies that aim at evaluating the impact of empathy on the interaction and in particular on settling a social relation between the agent and the user [29].

Paiva [25] defines empathic agents as "agents that respond emotionally to situations that are more congruent with the user's or another agent's situation, or as agents, that by their design, lead users to emotionally respond to the situation that is more congruent with the agent's situation than the user's one ". In this view, Klein et al. [26] describe an experimental study aimed at evaluating interfaces that implement strategies for affectively supporting users experience with negative moods and emotions by showing empathy and by actively supporting them. Results show how the affect-support was effective in relieving the user negative affective states when interacting with the computer. Along this perspective we find the work by Prendinger et al. [27] that developed an embodied agent in the scenario of job interviews that is able to recognize physiological data of users in real-time, to interpret this information as affective states, and to respond to affect by employing an animated agent. Sabourin et al. [28] present a study about designing pedagogical empathic virtual agents in a narrative-centered learning environment. They adopt a cognitive model, structured as a Bayesian network, which includes personal attributes of users (i.e. personality and goals of students), environment variables (i.e. dynamic attribute capturing a snapshot of the student's situation and activity) and physiological data about the user behavior (i.e. biofeedback parameters such as heart rate or galvanic skin response).

Recently several projects on AAL are endowing assistant robots with social capabilities. In [30] the possible role of empathy in socially assistive robotics is discussed. Leite et al. [29] propose a multimodal framework for modeling some of the user's affective states in order to personalize the learning environment by adapting a robot's empathic responses to the particular preferences of the child who is interacting with the robot.

Looking in more details to human-robot interaction, several EU-projects have addressed the modeling, definition, and implementation of social and cognitive skills in Social Assistive Robots (SARs) [44,45,20]. In particular, in order to enhance human robot interaction, emotional behavior recognition and generation have also been developed for social robots. In literature, two different approaches can be found to address this issue: social robots as agents able to generate emotions in human - robot interaction and robots able to recognize emotions of the human partners and to consequently adjust their behaviors. We reported here some examples of both approaches by considering only a mobile humanoid robot, the NAO by Aldebaran [37].

In their work, Cohen et al. [38] proposed two robots, the NAO and the i-Cat, able to express recognizable emotions and compared the recognition rates of the emotions in the two cases. For both robots, recognition rates for the expressions were relatively high but they focused their attention on NAO robot considering its body and colored eyes to express recognizable emotions. Tielman [39] proposes a model for adaptive emotion expression for the NAO. The robot communicates these emotions through its voice, eye colors, posture and gestures. An experiment with 18 children and two NAOs was carried on to test the effect of adaptive emotions on robot-child interaction. In the experiments, the children played a quiz with both an affective robot using the model for adaptive emotion expression and a non-affective robot. The experiment results confirmed that children responded more expressively to a robot that adaptively expressed itself than to a robot that did not.

Others studies present robots able to recognize and generate emotions. In the work of Zhang et al. [42], Facial Action Coding System has been incorporated in order to describe physical cues and facial behavior useful for the detection of six basic emotions plus neutral from real-time and posed facial expressions. The system was implemented on NAO humanoid. In Lim et al. [40] a developmental robot able to understand and express emotions in voice, gesture and gait using a model trained with voice data was presented. The recognized emotions were happiness, sadness, fear and neutral. In experiments, authors assumed an adult-infant simple interaction based on 4 Japanese words for ‘hello’, ‘look’, ‘no’, ‘bye bye’.

Another important field of application are robotic tutors developed with the ability to perceive emotions experienced by learners, and to incorporate these into pedagogical strategies. In a recent study, researchers addressed the problem of creating empathic robot tutors to support school students studying geography topics on a multi-touch table. The NAO robot tutor was equipped with a game-specific AI player that allowed it to play any of the different roles in the game. The next steps will be to use the AI to generate appropriate commentary feedback from the robot in a way that it can seem empathic to the users while still portraying its tutor role [41].

Most of the previous works with empathy in robotics focused on the perception and impact of empathy on participant attitude towards the robot.

3 Simulating Empathic Behavior

The concept of empathy is related to the understanding of what is happening to the other person. Therefore, according to [46], a model for simulating empathy in a robot should be able to i) recognize the affective cues and the affective state of the user and ii) interpret the motivations that triggered that emotion, iii) answer by expressing its emotions (as a consequence of the recognized state) by using different modalities (voice, facial expressions, and body movements and gestures) since the combination of verbal and non-verbal communication provide social cues that make robots appear more intuitive and natural. Our first attempts towards simulating empathy with a socially assistive robot are based on the understanding of the emotions of others (i.e., human users). We have developed a simple vision-based facial expression detection system capable of identifying a basic set of facial expressions including smiling, frowning, sadness, anger, etc. The list of facial expressions our system is capable of detecting is a subset of the Ekman's six basic emotions on human facial expression: joy, sadness, fear, anger, disgust and surprise [51]. The recognition of facial expressions is combined with the analysis of speech-based communication. In particular the speech prosody is analyzed in order to recognize its valence and arousal.

3.1 Collecting a behavioral information from human caregivers

To define and implement feasible behaviors of the robot, we integrated data collected from human caregivers with the guidelines that they follow in assistance of elderly people. In particular two human caregivers recorded their experience during the assistance of two elder women, both affected by a chronic disease, for a period of one month. These women lived alone and had a son/daughter which could intervene only in case of need and for solving relevant medical and logistic problems. Data have been collected using a paper-diary on which the caregiver had to annotate two kinds of entries: (i) the schedule of the daily tasks and (ii) the relevant events of the day, using a schema like the one reported in Table 1.

Table 1 - Some entries from the caregivers' paper-diary

Time	Event	Signs	Reason	Action	Communicative action	Recognized affect	Effect
10.00	…	…	medical visit	I remind Maria about the appointment with the doctor at 11.00.	Remind *Maria, I would like to remind you that today you have an appointment with the doctor at 11.00 a.m..*	…	…
10.30	…	…	medical visit	I ask and help Maria to dress up.	Ask_for *Today is a wonderful day. You can put on your beautiful dress that you like so much!*	…	Maria is dressed
10.40	…	…	medical visit	I send a reminder to Maria's daughter about the medical visit.	…	…	The daughter answered that she is coming.

10.45	Maria is worried	Sit down, Moaning “Oh my… Oh my”, Sad face	medical visit	I go toward Maria and try to console her.	Console *I’m sorry to see you so sad! You will see that everything will be all right .*	sadness	Maria is less sad

In particular, each row of the table represents a relevant event with the attributes for describing it and the action performed by the caregiver when this event occurred. For example, let’s consider the 4th row: at 10.45 (*time*) Maria is worried (*event*). The caregiver inferred Maria’s state since she was moaning, saying “Oh my, oh my” (*signs*) because she had to go to the doctor (*reason*). The caregiver recognized Maria’s sadness (*recognized affective state*). Hence, she went toward Maria (*action*) trying to encourage her by saying “Come on, don’t worry! You will not have any problem for sure.” (*communicative action*). After this action she noticed that Maria was less sad (*effect*).

From the collected data, we extracted the knowledge needed to build the reasoning strategies of the agent, so as to make its behavior believable. Overall, we collected a corpus of about 900 entries, which we used for: i) understanding which are the events and context conditions relevant to goal and action triggering; ii) understanding when considering affective and social factors is important during the interaction in real-life scenarios; iii) defining situation-oriented action plans and dialogue strategies; iv) collecting example dialogues between elderly people and human caregivers useful for testing the robot behavior.

3.2 An overview of NICA Architecture

As described in [5], the approach that we adopted in designing the architecture of NICA consists in interfacing the agent’s Body (for example Nao, Aibo, a conversational agent,…) with a Mind that, using several knowledge bases, reasons on which goal to pursue. NICA’s Mind has been modeled as a **BDI** (Belief, Desire, Intentions) agent, whose behavior is driven by persistent goals [38].

Briefly, the agent has a mission stated in the list of its persistent goals that have to be pursued during the agent lifecycle. At each stage of its life cycle, the agent evaluates whether there have been changes in the environment or in the user’s state that may threaten its persistent goals and cause a change in the planned behavior by triggering new goals and/or by modifying the scheduled actions.

At the present stage of development the agent considers a set of persistent goals related to the user’s wellbeing, the execution of necessary actions of the user’s daily routine, and so on. These goals correspond to the ones that human caregivers indicated as the most important ones in their daily assistance.

The agent implements a life cycle based on the following steps:

1. *Perception*: allows collecting data from sensors present in the environment and to handle the user input (speech, gestures, facial expressions or actions in the environment).
2. *Interpretation*: evaluates changes in the world and user state that are relevant to the agent’s reasoning and transforms them into a set of agent’s beliefs. In

particular it interprets the user's input.

3. *Goal Activation*: goals are triggered based on the current beliefs.
4. *Planning and Execution*: once a goal has been triggered it is achieved through the execution of a plan appropriate to the situation.

Although the agent can purse different persistent goals, since this paper focuses on how it reasons on the user affective state and how to trigger empathic behaviors in our examples we will consider the following goal as the most relevant one:

(BEL A NOT(Is(U, Negative(affective_state)) - "The Agent A has to belief that the user U is not in a negative affective state".

This means that NICA has to believe that the user is not in a negative affective state. As illustrated in Figure 1, in order to check whether this goal has been threatened the agent has to:

i) interpret the user's communicative actions expressed through speech and facial expressions;

ii) in case of expression of an emotion, recognize it and react emotionally to it;

iii) trigger a goal accordingly;

iii) achieve this goal through a communicative plan ("what to say") that can then be rendered as a combination of voice and animations of the agent's body ("how to say") [35];

iv) keep in its social memory information about which are the antecedents of emotions for the user, that is what triggers the emotions (events, situations, thoughts, etc.).

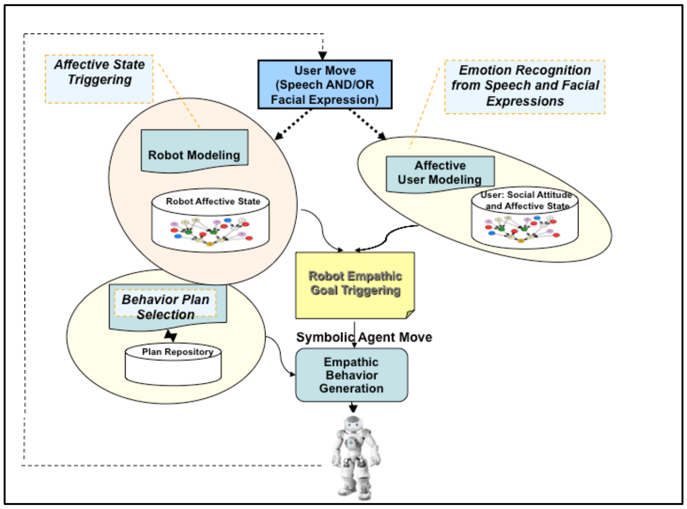

Affective State Triggering
User Move (Speech AND/OR Facial Expression)
Emotion Recognition from Speech and Facial Expressions
Robot Modeling
Robot Affective State
Affective User Modeling
User: Social Attitude and Affective State
Robot Empathic Goal Triggering
Behavior Plan Selection
Plan Repository
Symbolic Agent Move
Empathic Behavior Generation

Fig. 1. A schema illustrating the triggering of the empathic behavior in the robot.

The social memory is used to remember relations about events and the user's affective state. The importance of this piece of knowledge in the agent's mind is related to the need of establishing empathy with the elder person by remembering relevant data and this requirement was outlined several times by the human caregivers during the data collection phase.

As far as reasoning is concerned, in order to deal with the uncertainty typical of this domain (e.g. dealing with exceptional situations or with the smooth evolution of the user's affective state over time), we employ probabilistic models to reason on the user and to decide which behavior to adopt, that is the most appropriate set of actions to perform for satisfying the inferred user's goal.

At the present stage of the project, we simulate the interaction between the agent and the user by embodying NICA's Mind in the Nao robot. We adopted the approach proposed by Johnson et al. [43] to simulate emotion through Nao eyes by combining specific LED color patterns (Fig. 2).

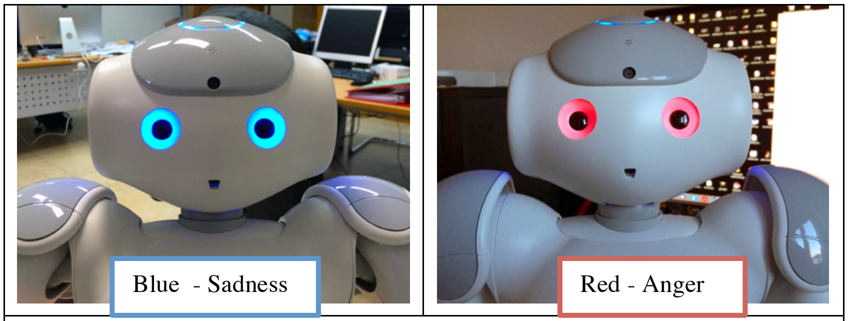


Fig. 2. A simulation of two emotion expressions with Nao.

3.3 Recognizing the User Affective State

In the current prototype of the system, we use a simple vision-based facial expression detection system capable of identifying a basic set of facial expressions in order to recognize the emotions of the human users. The list of facial expressions our system is able to detect is a subset of the Ekman's six basic emotions on human facial expression. The facial expression recognition system we adopted is fully automatic and, considering four-class expressions classification, the recognition rates we achieved were 82%, 76% and 95% using respectively Multi-SVM, k-Nearest Neighbors and Random Forest.

As far as spoken interaction is concerned, we employ VOCE (VOice Classifier for Emotions) a module that classifies the valence and arousal in the voice prosody. Our classifier follows an approach similar to [33,34]. In particular, the valence dimension is classified from positive to negative along a 4-point scale (from 1=very negative to 4=positive). Arousal is classified in a 3-point scale from high to low.

As far as the valence classification is concerned, the accuracy of the C4.5 algorithm is 83.12%, very close to the one of the K-NN that is 82.45%. As far as the arousal is concerned, C4.5 has an accuracy of 79.8% while the one of K-NN is 83.63 (validated using a *10 Fold Cross Validation* technique).

3.4 Reasoning on the User’s Affective State

In our model of empathy for a virtual caring robot we start from the *recognition of the user’s affective state for monitoring the belief* associated to this emotional state. In this way, during its lifecycle, the agent evaluates whether it is appropriate to trigger an affective communicative goal aimed at triggering the empathic behavior.

The robot’s beliefs about the user’s affective state are monitored with a dynamic model based on Belief Network (DBN) [32]. In fact, when modeling affective phenomena we must take into account the fact that affective state smoothly evolve during the interaction, from one step to the subsequent one and the state at every time of the interaction depends on the state it assumes in the previous turn. For this reason, the DBN formalism is particularly suitable for representing situations that gradually evolve from a dialog step to the next one. Moreover, Belief Networks are a well-known formalism to simulate probabilistic reasoning and deal with uncertainty in the relationships among the variables involved in inference process. The DBN model is shown in Figure 3. The model is employed to infer which is the most probable emotional state the user is experiencing at every step of the interaction by monitoring speech and facial expressions and it is also used to monitor the overall evolution of the user’s affective state (i.e. the belief of the agent about the positive or negative affective state of the user). In the model this is expressed by a temporal link between the *Bel(AffectiveState)Prev* and the *Bel(AffectiveState)* variables. At present we consider only a subset of the affective states that can be relevant for the generating an empathic response: sadness, happyness and anger.

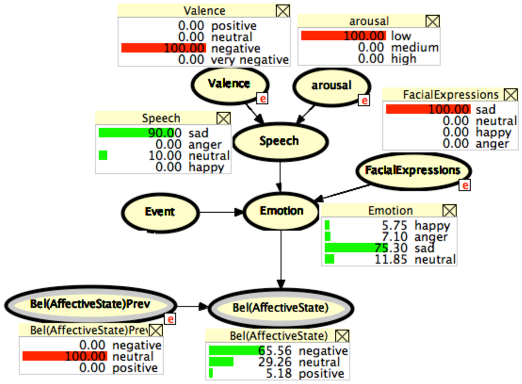


Fig. 3. The DBN model of the agent’s beliefs about the user affective state.

In particular, every time a new user move is entered, its acoustic features are analyzed and the resulting evidence are introduced and propagated in the network to recognize the user’s emotion and the overall polarity of her affective state. The same happens for the facial expression recognition module. The new probabilities of individual emotions are read and contribute to formulate the behavior of the agent; the probability of the dynamic variable (Bel(AffectiveState)) representing the valence of user's affective state is employed by the agent to check the consistency between its persistent goal of maintaining the user in a positive or neutral affective state and the actual emotional state the user is in at the time ***t***, thus causing the activation of the empathic goal.

3.5 Emotion Activation in the Robot's Mind

In order to activate an affective state in the robot for triggering affective goals we revised the emotion modeling method that we employed in another project [47]. The model is based on event-driven emotions according to Ortony, Clore and Collin's (*OCC*) theory [48]. In this theory, *positive* emotions (happy-for, hope, joy, etc.) are activated by *desirable* events while *negative* emotions (sorry-for, fear, distress, etc.) arise after *undesirable* events. In addition we considered also Oatley and Johnson-Laird's theory in which positive and negative emotions are activated (respectively) by the belief that some goal will be achieved or will be threatened [49]. In the context in which we employ the robot, we consider emotions in the *Well-being* category (joy, distress) and those concerning the *FortuneOfOthers* category (happy-for, sorryfor). Then, the cognitive model of emotions that is built on these two theories should represent the system of beliefs and goals behind emotion activation and endows the robot with the ability to *guess the reason why she feels a particular emotion and to justify it*.

The model of emotion activation is also represented with a DBN since we need to reason about the consequences of the observed event on the monitored goals in successive time slices. We calculate the intensity of emotions as a function of the *uncertainty* of the robot's beliefs that its goals will be achieved (or threatened) and of the *utility* assigned to achieving these goals. According to the utility theory, the two variables are combined to measure *the variation in the intensity of an emotion* as a product of the change in the probability to achieve a given goal, times the utility that achieving this goal takes to the robot.

Let us consider, for instance, the triggering of *Sorry-for* in the robot's model that is represented in Figure 4. This is a negative affective state and the goal that is involved, in this case, is *preserving others from bad*. In this figure R denotes the robot and U the user. The robot's belief about the probability that this goal will be threatened (Bel R (Thr-GoodOf U)) is influenced by his belief that some undesirable event E occurred to the user (Bel R (Occ E U)). According to Elliott and Siegle [50], the main variables influencing this probability are the desirability of the event (Bel R not(Desirable E)) and the probability that the robot attaches to the occurrence of this event (Bel R (Occ E U)). The user moves are interpreted as *observable* consequences of occurred events, that activate emotions through a model of the impact of this event on the robot's beliefs and goals. The user may say that a not desirable event occurred to him and may feel sadness or distress (Feel U(emotion)) that denotes that the event is undesirable. The probability of this node to be true depends on the emotion node in the network in Figure 3. This influences R's belief that U would not desire the event E to occur (Bel R Goal U ¬(Occ E U)) and (since R is in a *empathy* relationship with U, R adopts U's goals), its own desire that E does not occur (Goal R ¬(Occ E)). This way, they concur to increase the probability that the robot's goal of *preserving others from bad* will be threatened. Variation in the probability of this goal activates the emotion of *sorry-for* in R.

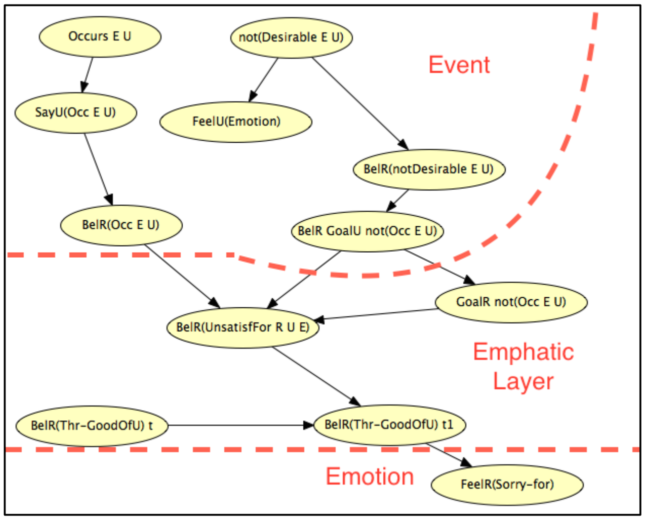


Fig. 4. A portion of the DBN representing the robot's mental state for the triggering of Sorry-For

The intensity of this emotion is the product of this variation times the *weight* the robot gives to the mentioned goal. The strength of the link between the goal-achievement (or threatening) nodes at two contiguous time instants defines the way the emotion, associated with that goal, decays, in absence of any event affecting it. By varying appropriately this strength, we simulate a more or less fast decay of emotion intensity. Different decays are attached to different emotion categories (positive vs. negative, FortuneOfOthers vs. Wellbeing and so on) and different temperaments are simulated, in which the *persistence* of emotions varies.

3.6 Triggering Empathic Behavior in the Robot

In order to decide how to behave as a consequence of the triggering of an emotion in the agent state of mind the agent triggers an affective goal. The list of empathic goals is inspired by the indications that human caregivers gave us during the data gathering phase at the beginning of the project, by the literature of empathy and pro-social behavior [29] and by the results of another study on the influence of empathic behaviors on people's perceptions of a social robot [21].

Currently, the empathic goals are the following:

- *console* by making the user feel loved and cuddled;
- *encourage* by providing comments or motivations like for example " don't be sad, I know you can make it!"
- *congratulate* by providing positive feedback on the user's behavior;
- *joke* by doing some humor in order to improve the user's attitude;

- *calm down* by providing comments and suggestion to make the user feel more relaxed.

For instance in case the *sorry-for* emotion is felt by the robot, the *console* goal should be triggered. Once a goal has been selected as the most appropriate to the emotion felt by the agent, the behavior planner module computes the agent behavior using plans represented as context-adapted recipes. Each plan is described by a set of *preconditions*, the conditions that have to be true to select the plan, the *effect* that the plan achieves and the *body*, the conditional actions that constitute the plan. After the execution of each action in the plan, the correspondent effect is used to update beliefs in the agent's mental state.

A sample of a portion of plan used to achieve the Console goal is the following:

```
<Plan name="Console">
<SelectCond> <Cond var="affective_goal" value=”console”/></SelectCond>
<Body>
        <Act name="Move" to=”U”/>
        <Cond var=”Feel(U,Sad)”>
        <Act name="Express" to=”U” var=”Sorry-for(R,U))”/>
        </Cond>
        <Cond var=”Know_Reason” value=”0” >
        <Act name="Ask" to=”U” var=”Why(U,Feel(U,Sadness))”/>
        </Cond>
        <Cond var=”Know_Reason” value=”1”>
        <Act name="Inform" to=”U” var=”Understand(R,U)”/>
        </Cond>
        <Act name="Express" to=”U” var=”Console(R,U)”/>
        … … …
</Body>
</Plan>
```

The tag <Cond> allows selecting actions on the basis of the current situation.

For instance, the action <Act name="Express" to=”U” var=” Sorry-for(R,U)/> is used to express the sorry attitude of the robot R and will be performed only if the user feel sad. In the same way, the action “Ask” about “Why the user is sad” will be performed only if the agent does not know why the user is in the current state. Moreover, if the action is complex, then it can be specified in a subplan describing elementary agent actions. Each communicative act in the plan is then rendered using simple template-based surface generation technique [35]. These templates are selected on the basis of the type of communicative act and its content and are expressed in metalanguage [36] that is then interpreted and executed by the agent's body. Plans and surface generation templates have been created and optimized combining actions on the basis of pragmatic rules that were derived from the corpus dataset.

4 A Case Study

In this section we show an example of an empathic behavior of the agent in a typical interaction scenario that we envisaged as a suitable one for testing our agent framework.

It's morning and Nicola, a 73 y.o. man, is at home alone. He doesn't feel very well since he has a cold and fever. Nicola is sitting on the bench in his living room that is equipped with sensors and effectors. According to the situation the smart environment selects a workflow and starts to execute scheduled tasks accordingly. The caring robot has to check Nicola's health state and recommend him to take some medicine. After a while Nicola starts whispering and says with a sad facial expression: "Oh My …oh poor me…". This is perceived by the robot that interprets it and activates the most appropriate behavior.

The voice classifier recognizes a *negative* valence with a low arousal from the prosody of the spoken utterance and the facial expression classifier recognizes the *sadness* emotion. These evidences are propagated in the DBN and the belief about the affective state of the user is in a negative affective state with the higher probability (65.56), as shown in Figure 3. Then, since the goal of keeping the user in a state of well-being is threatened, the DBNs modeling the robot's affective mind are executed to trigger the robot's affective goal (sorry-for in this case). As described in the previous section, the goal to pursue in this situation is the "console" one. Then, the correspondent plan is selected (see previous section) and the execution of its actions begins. The plan includes the following actions since the agent does not know why the user is sad and it will ask the user about it:

```
MoveTo(NAO,NICOLA)
Express(NAO,Sorry-
for(NAO,NICOLA))
Ask(NAO,NICOLA,Why(Feel(NICOLA,Sa
dness)))
Express(NAO,Console(NAO,NICOLA))
```

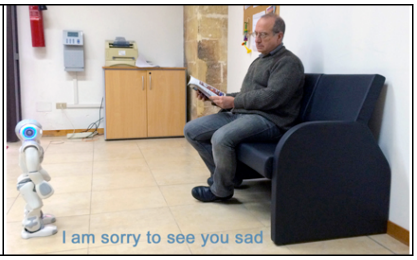


Fig. 5 A simulation of the scenario with an elderly person.

When a new belief about the event that occurred to the user related to a particular affective state is acquired by the robot during the interaction, it is stored in the agent Social Memory. In this way the robot will remember which event causes a particular affective state in the user, for instance the event "has_disease" is associated to the affective state "sadness". This information can be used by the agent in the dialogue with the user for preventing this state or for improve the relation between the user and the robot.

5 Discussion and Future Work

This paper presented issues concerning the importance of taking into account affective factors when modeling the user in social interaction with a caring agent. In our opinion, besides assisting the elderly user in performing tasks, the agent has to establish a social long-term relationship with the user so as to enforce trust and confidence.

The underlying idea of our work, in fact, is that endowing the robot with a social empathic behavior is fundamental when the devices of a smart home are integrated pervasively in everyday life environments. In this paper we illustrated how this capability has been designed and implemented in a caring assistant for elderly people.

Evaluating the efficacy of the empathic behavior of the social robot in a real-context at the moment is not feasible due to the lack of enough smart homes equipped with social robots. Therefore, we performed a quantitative evaluation of the decisions and plans executed by the agent compared to the behaviors of the human caregivers that we annotated in a previous phase of the project. To this aim we randomly split our corpus into 70/30 training/test partitions. For each item of the test set, we formalized the corresponding scenario in order to set the evidences in the simulation test. Then, we observed the robot's behavior in terms of selected communicative acts: the behavior of the robot was classified as 'correct' if it matched the choice of the human caregiver, as 'incorrect' vice versa. Results of the evaluation are encouraging and indicate that the system performance is quite good since the choices of the agent match the human actions in the dataset in the 79% of cases. We are aware that it is important to conduct an evaluation study with real elderly users. This kind of experiment should aim at assessing the impact of the use of a social robot vs. seamless interaction with the environment services in smart environments. Another important issue to be addressed in our future work concerns the interpretation of gestures and postures of the user.

Acknowledgements

This work fulfils the research objectives of the PON02_00563_3489339 project "PUGLIA@SERVICE - funded by the Italian Ministry of University and Research (MIUR).

References

1. Bierhoff, I. and van Berlo, A. More Intelligent Smart Houses for Better Care and Health, Global Telemedicine and eHealth Updates: Knowledge Resources",vol. 1, 322-325, 2008.
2. Steg, H. et al. Ambient Assisted Living – European overview report, September, 2005
3. De Carolis, B., Ferilli S., Greco D. Towards a Caring Home for Assisted Living.Workshop "The Challenge of Ageing Society: Technological Roles and Opportunities for Artificial Intelligence"co-located with the AI*IA 2013.
4. Brahnam, S. & De Angeli, A. (2008). Editorial - Special issue on the abuse and misuse of social agents. Interacting with computers, 20 (3).
5. R. Niewiadomski, M. Ochs and C. Pelachaud, Expressions of Empathy in ECAs, in Proceedings of the 8t Int. Conf. on IVA, LNAI, vol. 5208. Springer-Verlag, Berlin, Heidelberg, pp. 37-44, 2008.
6. W.S.N. Reilly. Believable Social and Emotional Agents. PhD thesis (1996)
7. Nijholt, A., de Ruyter, B., Heylen, D. and Privender, S. Social Interfaces for Ambient Intelligence Environments. Chapter 14 in: True Visions: The Emergence of Ambient Intelligence. Aarts, E. and Encarnaçao, J., Eds. Springer, New York, 2006, 275—289.
8. Ortiz, M. Del Puy Carretero, D. Oyarzun, J.J.Yanguas, C. Buiza, M. F. Gonzalez, and I. Etxeberria. Elderly users in ambient intelligence: does an avatar improve the interaction?.

In Proceedings of the 9th conference on User interfaces for all (ERCIM'06), Constantine Stephanidis and Michael Pieper (Eds.). Springer-Verlag, Berlin, Heidelberg, 99-114, 2006.

9. T. Bickmore and R. W. Picard, Establishing and maintaining long-term human-computer relationships, ACM Transactions on Computer Human Interaction, 12, 2, 293-327, 2005.
10. Hoffman, M. L. (1981). The development of empathy. In J. Rushton & R. Sorrentino (Eds.), Altruism and helping behavior: Social personality and developmental perspectives (pp. 41–63). Hillsdale, NJ: Erlbaum.
11. C. Anderson and D. Keltner (2002). The role of empathy in the formation and maintenance of social bonds. Behavioral and Brain Sciences, 25, pp 21-22.
12. S Brave, C Nass, and K Hutchinson. Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int. J. of Human Computer Studies 62(2):161–178, 2005.
13. A. Paiva, et. al. “Caring for Agents and Agents that Care: Building empathic relations with synthetic agents”, AAMAS 2004, ACM Press, 2004.
14. Eisenberg, N., & Miller, P. A. (1987). The relation of empathy to prosocial and related behaviors. Psychological Bulletin, 101, 91–119.
15. http://www.aal-europe.eu/wp-content/uploads/2012/04/AALCatalogue_onlineV4.pdf
16. A. Cesta, G. Cortellessa, F. Pecora and R. Rasconi, Supporting Interaction in the RoboCare Intelligent Assistive Environment, AAAI 2007 Spring Symposium, 2007.
17. J. Pineau, M. Montemerlo, M. Pollack, N. Roy and S. Thrun, Towards Robotic Assistants in Nursing Homes: Challenges and Results, Robotics and Autonomous Systems 42(3–4), pp. 271–281, 2003.
18. Graf B, Hans M, Schraft RD (2004) Care-O-bot II – development of a next generation robotics home assistant. Auton. Robots 16, 193–205.
19. CompanionAble project (2011) http://www.companionable.net/
20. Cuijpers, R.H. (2012). http://ksera.ieis.tue.nl/
21. A.M. van Ruiten, A. M., Haitas, D., Bingley, P., Hoonhout, H.C.M., Meerbeek, B.W. and Terken, J.M.B. Attitude of elderly towards a robotic game-and-train- buddy: evaluation of empathy and objective control. In R. Cowie and F. de Rosis (Eds.) Proceedings of the Doctoral consortium, in the scope of ACII2007 Conference, 2007.
22. J. N. van Breemen, (2004). iCat: a generic platform for studying personal robot applications. Paper presented at the IEEE SMC, Den Haag.
23. J. Broekens, M. Heerink, H. Rosendal. Assistive social robots in elderly care: a review. Gerontechnology 2009; 8(2):94-103; doi: 10.4017/gt.2009.08.02.002.00
24. Reeves and C. Nass, The media equation: how people treat computers, television, and new media like real people and places, Cambridge University Press, New York, NY, 1996.
25. A. Paiva et al. Empathy in Social Agents. International Journal of Virtual Reality, Vol. 10, No. 1, pg. 65-68, 2011.
26. Klein J, Moon Y, Picard R (2002) This computer responds to user frustration: Theory, design, and results. Interacting with Computers, 14:119–140.
27. Prendinger H, Mori J, Ishizuka M (2005) Recognizing, modeling, and responding to users affective states. In Proceedings of User Modeling 2005, Lecture Notes in Computer Science, Volume 3538/2005, 149, DOI: 10.1007/11527886_9 2005
28. Sabourin J, Mott B, Lester J (2011) Computational Models of Affect and Empathy for. Pedagogical Virtual Agents. Standards in Emotion Modeling, Lorentz Center International Center for workshops in the Sciences.
29. I. Leite and C. Martinho and A, Paiva Social Robots for Long-Term Interaction: A Survey. International Journal of Social Robotics, pg. 1--18, January, 2013.
30. Maja Mataric', Adriana Tapus, and David Feil-Seifer (2007) “Personalized Socially Assistive Robotics”, Workshop on Intelligent Systems for Assisted Cognition, Rochester, New York, USA, October, 2007.

31. C. Castelfranchi and F. Paglieri, The role of beliefs in goal dynamics: Prolegomena to a constructive theory of intentions, Synthese 155, pp. 237-263, 2007.
32. F.V. Jensen, Bayesian Networks and Decision Graphs, Statistics for Engineering and Information Science, Springer, 2001.
33. Vogt T, Andre' E, Bee N (2008) EmoVoice - A Framework for Online Recognition of Emotions from Voice. In Proceedings of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems: Perception in Multimodal Dialogue Systems (PIT '08), Springer-Verlag, Berlin, 188-199.
34. Sundberg, J., Patel, S., Björkner, E., & Scherer, K.R. (2011). Interdependencies among voice source parameters in emotional speech. IEEE Transactions on Affective Computing, 2(3).
35. Reiter E, Dale R (2000) Building Natural Language Generation Systems. Studies in natural language processing. Cambridge University Press, Cambridge, United Kingdom. ISBN 0-521- 62036-8.
36. De Carolis B, Pelachaud C, Poggi I, Steedman M (2004) APML, a Mark-up Language for Believable Behavior Generation, in H. Prendinger Ed, Life-like Characters, Tools, Affective Functions and Applications, Springer.
37. http://www.aldebaran-robotics.com/ (last visited Oct 3, 2012).
38. Cohen, I., Looijeand, R. & Neerincx, M. A. (2011). Child's recognition of emotions in robot's face and body. In Proceedings of the 6th international conference on human-robot interaction (pp. 123–124).
39. Tielman M. Adaptive emotional expression in robot-child interaction. Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14. New York, New York, USA: 2014. p. 407-414.
40. Lim, A.; Okuno, H.G., "The MEI Robot: Towards Using Motherese to Develop Multimodal Emotional Intelligence," Autonomous Mental Development, IEEE Transactions on , vol.6, no.2, pp.126,138, June 2014
41. Ribeiro, T., Pereira, A., Deshmukh, A., Aylett, R., & Paiva, A. (2014, May). I'm the mayor: a robot tutor in enercities-2. In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (pp. 1675-1676). International Foundation for Autonomous Agents and Multiagent Systems.
42. Zhang, L., Jiang, M., Farid, D., & Hossain, M. A. (2013). Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. Expert Systems with Applications, 40(13), 5160-5168.
43. Johnson, D. O., Cuijpers, R. H., & van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. International Journal of Social Robotics, 5(4), 503-513.
44. DOMEO (Domestic Robot for Elderly Assistance) (2012). www.aal-domeo.eu
45. MOVEMENT website (2012). http://www.is.tuwien.ac.at/fortec/reha.e/ projects/movement/
46. Feshbach, N. D. 1987. Parental empathy and child adjustment/maladjustment. In Eisenberg, N., and Strayer, J., eds., Empathy and its development. Cambridge University Press.
47. de Rosis, De Carolis, Carofiglio, Pizzutilo 2003. Shallow and inner forms of emotional intelligence in advisory dialog simulation. In H. Prendinger and M. Ishizuka (Eds.): "Life-Like Characters. Tools, Affective Functions and Applications". 271-294. Springer 2003.
48. Ortony, A., Clore, G.L. and Collins, A., 1988. The cognitive structure of emotions. Cambridge University Press.
49. Oatley, K. and Johnson-Laird, P.N., 1987. Towards a Cognitive Theory of Emotions. Cognition and Emotion, 29-50.
50. Elliott, C. and Siegle, G., 1993. Variables influencing the intensity of simulated affective states. In Proceedings of the AAAI Spring Symposium on Mental States.'93, 58-67.
51. Ekman, P. & Oster, H. (1979). Facial Expressions of Emotion. Annual Review of Psychology, 30, 527-554.

Social Mood Revealed

Bartosz Ziembiński

Polish Academy of Sciences, Warsaw, Poland,
b.ziembinski@phd.ipipan.waw.pl

Abstract. Social mood, the aggregated mood of a society, emerges from complex system of individual moods and their influences on each other. The real social networks consist of millions or even billions nodes constantly interacting with each other. Can such a complex system be modeled by a graph consisting of a small number of agents with simple interactions between them? Profile of Mood States, known and well-vetted psychometric instrument, distinguishes seven mood dimensions (Tension, Happiness, Calmness, Vigor, Fatigue, Confusion and Friendliness). If we apply them to a society at large, i.e. to social mood, is it possible to measure influences of one mood dimension on another? In addition to this, is it possible to both maintain good approximations of social mood changes and be able to observe such interactions at the same time? In this work we investigate these questions and propose a framework which can approximate or even, in some circumstances, be predictive of future social mood states. The framework consists of a model of social influence and an evolutionary algorithm learning proper network topology and model parameters.

Keywords: Social Mood, Collective Emotions, Social Networks, Social Influence, Agent-Based Modeling, Complex Systems, Social Simulation, Sentiment Analysis

1 From individual to social mood

From the psychological research it is known that the emotional state, as well as the amount of information, play the main role in human decision-making [11, 9]. Traditionally, in theoretical considerations, the second factor played a more important role. For instance, *rational choice theory*, economical perspective that perceives people as *rational actors*, explains decision-making process through the paradigm of utility maximization [13]. Agents base their actions on pragmatic calculations of their best interests.

However, emotions can profoundly affect human decision-making process as well, in many cases driving an individual to make a choice that seems "irrational" in the framework of said theory. For example, behavioral finance has provided proofs stating that financial decisions are significantly affected also by emotion and mood and not only by rational utility maximization [23]. Damasio states that personally beneficial decision making requires emotion as well as reason [9]. He also proposed the *somatic marker hypothesis*, that describes a mechanism by which emotional processes can guide (or bias) behavior [5]. Pfister and Böhm have developed a classification of how emotions function in decision-making, that conceptualizes an integral role for emotions, rather than simply influencing decision making [29].

Thus, emotions affect individual choices and decisions. Does this also apply to larger groups of people, i.e. can societies experience mood states that affect their collective decision making? Prechter's *socionomic hypothesis* suggests that the social mood drives various types of social action in the areas of cultural, political and financial behaviour [30]. However, assuming that the social mood affects the society behaviour analogously to the way in which one's emotions drive his individual actions is quite unreasonable. The society is a complex system with its emergent properties. The social mood, as a state of the whole system, is something different than just a simple sum of its parts [2]. Therefore, researchers attention has been focused on finding the relations between the social mood and the behaviour of societies [23, 26, 6, 31].

The first problem with such investigations is to actually find a way to measure the social mood. Large surveys of public mood are generally expensive and difficult to undertake. That is why there were proposed some ways to assess the social mood indirectly. For example: from the results of football games [12] and from weather conditions [18]. Recently though, researchers came up with other, low-cost and very efficient, way to measure the public mood. They were able to do it through sentiment analysis of social media content such as Twitter feed, discussion forums or blogs [27, 34, 33]. Social mood measured by means of Twitter turned out to be predictive of many social phenomena including stock market [7], political elections [15, 25], box-office revenues for movies [3] etc. If social mood can be related or even predictive of so many social matters, it is important to have better ways to analyze it and to understand its behaviour.

In this paper we propose a framework that can translate huge, highly complicated social network (of individual moods and their influences on each other) to a fairly simple and an order of magnitude smaller graph of agents. It behaves in a similar manner to the real network concerning dynamics and interactions of the mood dimensions. The translated network can be then more easily analyzed.

Such a framework enhances the state of the art of social sciences, offering a tool to measure and to interpret social mood and the interactions between the mood dimensions. Its novelty is provided by its data-driven approach. Most of the social influence models employ bottom-up methodology: begin with simple rules (of agents interactions), then observe the emergent behaviour of the system [14, 32, 24, 19, 16, 10]. The goal of such investigations is to examine how the model behaves given particular assumptions. The model, however, may or may not reflect a real-life social system. We believe that we propose more holistic approach. It consists of two stages. We begin with measuring the actual social mood by the means of the real-world data. Then we tune our highly customizable model of social influence to reflect these measurements. This way we provide not only the theoretical considerations, but also a model that can indeed approximate social mood changes, that are happening in a day-to-day reality. On the other hand, it is not just a numerical approximation - the construction of the model enables one to investigate the interactions between agents, representing different mood dimensions. To the best knowledge of this paper's authors, models of mood dimension interactions have not been proposed in the scientific literature so far. The same is for data-driven models of social influence. Therefore, the proposed framework might be of a great interest for social scientists.

The paper is structured as follows. In Section 2 we describe our framework. Firstly, we explain how we assess the social mood. Then, we describe our model of social mood. The description involves a model of social influence and an evolutionary algorithm aiming to find the best network topology and model parameters. Section 3 describes the empirical experiments that were conducted. In Section 4 we discuss the results of the experiments. We draw final conclusions in Section 5.

2 Social mood translation

In a real world people affect each others individual emotional states during communication. If we sum those individual emotional states up, we will receive a global measure called social mood. The question is, if we can replace the real social network of emotional influences with its model, say with a number of nodes two times smaller? And at the same time be able to maintain similar dynamics of mood influences and good approximation of a global mood state? Then, could we create a model four times smaller? How small could that model be? It is clear that the smaller it is, the easier it would be to analyze it and to understand its dynamics.

In this paper we propose a framework, which is able to translate huge complex social network of individuals to a simple graph with fixed, small number of nodes (not more than 50 nodes). In this graph each node is an agent which is a representation of a class of individuals in the original network. Every agent apart from its mood state, has its own level of *impressionability* and *influence*. Respectively, these are the measures of how much an agent is sensitive to influence of others and how influential it is. The edge between nodes denotes their ability to affect each other. The values of parameters and the topology of the graph is determined by data-driven evolutionary algorithm which approximates the social mood time series.

The next two subsections will describe the framework in detail. First, we will describe how we measure social mood and then, how model of its dynamics is constructed.

2.1 Assessing the social mood

We measure social mood by analyzing Twitter feed in terms of 7 mood dimensions. We list them here (with explanation of what does, respectively, the low and high score of each dimension mean):

1. Tension - relaxed or anxious,

2. Happiness - happy or depressed,
3. Calmness - calm or angry,
4. Vigor - apathetic or vital,
5. Fatigue - rested or tired,
6. Confusion - sure or confused,
7. Friendliness - aloof or kind.

We use similar mood dimensions and methodology of assessing the public mood to the one used in [7] (namely Profile of Mood States). The motivation behind this is that we believe we should measure social mood in more than just one classic dimension (positive vs. negative) to obtain some number of potentially different aspects of public mood. The efficiency of this sentiment tracking tool was cross-validated against big socio-cultural events like the U.S presidential election (November 4, 2008), Thanksgiving (November 27, 2008) etc. [7, 6]. In addition to this, in [7] authors find an accuracy of 87.6% in predicting the daily up and down changes in the closing values of the Dow Jones Industrial Average index, which indicates that the classification can have good practical applications.

Data We recorded a collection of public tweets which were posted during 14 days from July 7th to July 20th, 2014. We were interested only in tweets expressing author's mood state, thus we only tracked tweets containing words: ”feel” and ”feeling” (20,110,489 tweets). For each post, we obtained its date and time of submission, as well as the content of the message (which is a text limited to 140 characters).

One can have an impression that the dataset is particularly small (14 days), concerning the fact that in other papers datasets can span over several months. The difference, though, is in temporal resolution of the datasets. Whereas researchers usually measure social mood in terms of days, in this work we measure it every 5 minutes. We do it to be able to observe the intraday dynamics of collective emotions and to be able to track the microchanges in social mood. We believe that such investigations may be helpful, for instance, for financial intraday traders, for trading algorithms or for people responsible for communication and public relations. If we compare the sizes of the datasets, we will obtain $14d \times 24h \times 12 = 4032$ time intervals for our dataset. In this paper experiments were conducted for time periods between 09:30 and 16:00 EST from Monday till Friday, as these are the times when New York Stock Exchange is opened. This gives us $10d \times 6.5h \times 12 = 780$ time slots. On the contrary, if we take a daily resolution into consideration and, say, we will have a dataset of 9 months, it gives us around $9m \times 31d = 279$ time intervals.

Another fact is that the volume of tweets posted nowadays is much greater than it used to be in the past. In this paper we collected 20,110,489 tweets during 14 days and in [7] authors collected 9,853,498 tweets during over 9 months.

Generating social mood time series In order to obtain a mood score of a tweet we compare each word from a tweet against each word from a lexicon of so called *emotional* words. The lexicon is derived from an existing psychometric instrument, namely the Profile of Mood States (POMS) [22]. It is known and well-vetted psychometrical instrument used to measure one's emotional state. It consists of 65 adjectives describing the mood state which are linked with different emotional dimensions. The examined person has to refer to these adjectives on a five-point scale.

To create a computational version of the test, we expand the basic lexicon of 65 adjectives from POMS with similar words, which we obtain by analysing word co-occurrences in big collections of texts. The expanded lexicon consists of 965 associated terms which are collected in the following procedure. We use Bing search engine to query for phrases "*is* [*adj*] *and*" and "*was* [*adj*] *and*", where [*adj*] denotes a particular adjective from the original lexicon which we want to find similar words to[1]. For each of the queries, we download first 200 results. For each result, we extract the word after conjunction *and*. Then we sort extracted words by most frequent occurrences. From the most frequent words we choose similar adjectives by hand. The advantage of querying search engines is that they are a relatively simple way of searching over a large collection of documents. Moreover, it also enable us to retrieve similar words which actually are in use.

Having the lexicon of *emotional* words, the social mood of Twitter feed is measured in the following way. Tokenization is performed on each tweet and then each word from a tweet is compared with each adjective from the lexicon. If there is a match, the adjective from the lexicon is mapped back to its original POMS term and via the POMS scoring table to its respective POMS dimension. Then, a counter of corresponding dimension is incremented by one.

To obtain a social mood time series we split our collection of tweets into groups of messages sent in 5 minutes long time periods. For each hour H, we distinguish time intervals: $[H{:}00, H{:}05)$, $[H{:}05, H{:}10)$, ..., $[H{:}55, H+1{:}00)$. Then for tweets from each of such time intervals, we employ our mood measuring procedure. At the end, we obtain times series:

$$M = \{M_t : t \in T\} \tag{1}$$

where t corresponds to successive time intervals and

$$M'_t = [d'_{t,1}, d'_{t,2}, ..., d'_{t,7}] \tag{2}$$

where, $d'_{t,1}, d'_{t,2}, ..., d'_{t,7} \in \mathbb{N}$ are values of respective mood dimensions: Tension, Happiness, Calmness, Vigor, Fatigue, Confusion and Friendliness.

For our social mood time series not to be dependent on the volume of tweets in a given period of time, we then normalize the values of mood dimensions in the following way. For each mood dimension $d_{t,i}$, $i \in 1, 2, ..., 7$:

$$d_{t,i} = \frac{d'_{t,i}}{\sum_{j=1}^{7} d'_{t,j}} \tag{3}$$

[1] Bing search engine distributes a dedicated API. For the details visit http://www.bing.com/dev/en-us/dev-center.

Obtaining final elements of social mood time series:

$$M_t = [d_{t,1}, d_{t,2}, ..., d_{t,7}] \tag{4}$$

All mood times series in the rest of the paper are normalized in the same manner.

2.2 Model of social mood

Real-world social mood networks consist of big number of people, each of them having their own mood state. These people can interact with their acquaintances, affecting their moods, as well as being affected by them.

Therefore, if we want to translate such a network to a smaller graph, we need to find a way to model 3 things:

1. Collective mood state of individuals - mostly, we already have it done. We model it with 7-dimensional vector like in equation (4).
2. Topology of the social network - we need to find a way to translate the connections between nodes in a big social network to analogous connections in a small graph.
3. Social influence - we need to build a model of how agents are affecting each others mood states in a small graph.

We will start with approaching the topology issue, then we will describe our model of social influence and finally we will present the evolutionary algorithm which aims to find the best topology and influence parameters. All these components will, in the end, describe our framework.

Topology of the network We use evolutionary approach to find the best network topology (as well as other parameters of the model). This choice is made, because we want the algorithm:

- to be population-based - in order to be able to compare obtained solutions at any time of the algorithm run,
- to be anytime - meaning that it can return a valid solution, even if it is interrupted before it ends,
- not to make any assumptions about the topology and the parametrs of the model.

However, evolutionary algorithm needs initial population in which topologies are somehow constructed. To model the social network in the beginning stage of the evolution, we decided to use two classes of graphs.

First class are random graphs. We believe that they are the simplest and the most natural way to initialize network topologies, concerning the fact that we take advantage of an evolutionary approach. To construct the particular graph, first we draw $p \in (0, 1)$ from the uniform distribution. Then, every possible edge occurs independently with the probability p.

Second class consists of scale-free graphs, which are graphs whose degree distribution follows the power law. The motivation behind this choice is that many real-world social networks, as well as cyberspace networks, are conjectured to be scale-free [4, 17, 8]. To generate graphs, whose node degrees follow the power law distribution, we used Barabási-Albert model [1]. The algorithm uses a preferential attachment mechanism, which means that the more connected a node is, the more likely it is to receive new links. More formally, the probability p_i that the new added node is connected to the pre-existing node i is:

$$p_i = \frac{k_i}{\Sigma_j k_j} \tag{5}$$

where k_i is the degree of node i and the sum is made over all pre-existing nodes j.

Model of social influence There exists a multitude of social influence models in the sociophysics literature. They can be classified into discrete (including binary) models and continuous models depending on the representation of opinions that are being influenced.

The typical discrete models include Ising model [14], Sznajd model [32], social impact model [24], voter model [19], etc. These descriptions of social influence, sometimes called the *toy models*, are useful for simplifying the opinion dynamics explanations (e.g. using the *temperature* notion to introduce the stochastic behaviour [14] or proposing *United we Stand, Divided we Fall* rule to implement the phenomenon of social validation [32] etc.). However in our case, the drawback of these models is their discrete nature, because our measurments of the social mood have continous characteristic.

This fact brings our attention to the continuous models, that mainly include Hegelsmann-Krause model [16], Deffuant-Weisbuch model [10] and their numerous variants and extensions [28, 20, 21, 35]. These approaches, however, also possess some limitations, as far as our work is concerned. Some of them assume bounded confidence of agents, which means that the agent adjusts its opinion only towards the opinions that are not very distinct (that lay in the ϵ-interval around the agents' opinion) [10, 21, 35]. As we want to model interactions between mood dimensions, this approach is not suited for our case (for instance a state with high value of *Happiness* may affect a state with low value of *Friendliness*). Other drawback is that some of the models assume influence dynamics, that leads to a consensus [21, 16, 28]. Consensus is not a typical feature of many social situations, neither is it a typical state of the mood dimensions dynamics. Mood dimensions do not tend to average themselves and often tend to differentiate (e.g. low value of *Happiness* and high value of *Tension*). Our model need to have a way to describe this phenomena. Another feature that it should possess is the ability to describe the fact that agents may not always be easily influenced by others.

Therefore, we propose our own model of social influence, which is similar to Hegelsmann-Krause model, but also introduces some major differences. They enable us to model the characteristics of mood dimensions dynamics, which we just mentioned. It is defined as follows:

1. $\mathcal{A} = \{1, 2, ..., n\}$ is the set of agents.
2. Each agent i, at discrete moment in time t, has its own mood state:

$$M_{i,t} = [d_{i,t,1}, d_{i,t,2}, ..., d_{i,t,D}] \tag{6}$$

   where $d_{i,t,k} \in \mathbb{R}$ and D is a constant denoting the number of mood dimensions[2].
3. Each agent i, at discrete moment in time t, knows if each of its mood dimensions increased or decreased during the last time step:

$$\Delta_{i,t} = [\Delta_{i,t,1}, \Delta_{i,t,2}, ..., \Delta_{i,t,D}] \tag{7}$$

   where $\Delta_{i,t,k} \in \mathbb{R}$ and $\Delta_{i,t,k} = d_{i,t,k} - d_{i,t-1,k}$, for $t > 0$. For each agent i first element of the sequence Δ_i is specified at the beginning:

$$\Delta_{i,0} = [a_{i,0,1}, a_{i,0,2}, ..., a_{i,0,D}] \tag{8}$$

   where $a_{i,0,1}, a_{i,0,2}, ..., a_{i,0,D}$ are specified initial values.
4. Each agent i has its level of:
   - *influence* $\varphi_i \in [0, 1]$, which denotes how much it is affecting others,
   - *impressionability* $\delta_i \in [0, 1]$, which denotes how much it is being affected by others.
5. Agents are organized in the network $\mathcal{N} = (\mathcal{A}, E)$, where E is a set of connections or edges, which are 2-element subsets of the set $\mathcal{A}$.
6. Sequence of agent's mood states is specified as follows. For each agent i:
   - First element of the sequence is specified at the beginning:

$$M_{i,0} = [b_{i,0,1}, b_{i,0,2}, ..., b_{i,0,D}] \tag{9}$$

     where $b_{i,0,1}, b_{i,0,2}, ..., b_{i,0,D}$ are initial values.
   - Elements of the next time steps $t > 1$ are defined using a recursive rule. For each mood dimension $d_{i,t,k}$, $k \in 1, 2, ..., D$:

$$d_{i,t,k} = d_{i,t-1,k} + \delta_i d_{i,t-1,k} \Sigma_j sgn(\Delta_{j,t-1,k})\varphi_j \tag{10}$$

     where the sum is made over all agents j connected to the agent i (indicated by the set E).
7. The global social mood state, at each discrete moment in time t, is defined as a sum of agents' mood states:

$$M_t = \Sigma_{i \in \mathcal{A}} M_{i,t} \tag{11}$$

Thus, in the model in every discrete time step t part δ_i of agent's i mood can be affected by its neighbours. If their particular mood dimension went up in the previous time step, the neighbours will try, taking their influence parameters φ into consideration, to increase it. In other case, analogously, they will try to decrease it.

[2] In this paper $D = 7$.

How to construct the model from Twitter data? We only described aggregated social mood $M = \{M_t : t \in T\}$ acquired from Twitter data so far. However, to be able to use it in our model of social influence, we need split data. To achieve this, the idea is to split tweets into some kind of equivalence classes associated with mood dimensions. The agents then are not representatives of individuals, but representatives of mood dimensions. The easiest way to achieve this is to employ the mood measuring procedure for each tweet, identify its dominant mood dimension (the dimension with the highest score) and then classify the tweet as Tension, Happiness, Calmness, Vigor, Fatigue, Confusion or Friendliness representative. If there is more than one dimension with the highest score, classify the tweet randomly as a representative of one of its dominant dimensions.

Using this procedure, we can obtain decomposition of tweets into seven different groups associated with mood dimensions. We can then couple each group $i \in \{1, 2, ..., 7\}$ with different agent, obtaining corresponding mood time series:

$$M_i = \{M_{i,t} : t \in T\} \tag{12}$$

Having agents as representatives of mood dimensions, we can then apply our social influence model to observe how different dimensions are influencing each other. This way, we obtain a graphical representation of influence dynamics. In this approach the influence of one mood dimension on the other is not based on the actual Twitter social graph or other kind of individuals topology. The influence is measured on the macro level, the same way that in a society optimists have an influence on pessimists or electorate of one political party has an influence on the other electorate. The influence is measured as the change in aggregated sum of micro-interactions among the individuals.

On the other hand, conducting such a simulation (running the model), we are able to calculate mood estimators $\widehat{M_{i,t}}$ - set of vectors of mood scores of every agent i, in every time step t. Those estimators can be then summed to obtain global social mood estimator $\widehat{M_t}$. It is then easy to assess how good is our estimation (and all in all - simulation) calculating the mean absolute percentage error (MAPE):

$$MAPE = \frac{1}{|T|} \sum_{t}^{t \in T} \frac{|\widehat{M_t} - M_t|}{M_t} \tag{13}$$

We choose MAPE as a measure of performance of our simulations, because we need to compensate for two things:

1. scores of some mood dimensions are usually much greater than scores of other mood dimensions - therefore, we need percentage error to measure performance of all dimensions approximations,
2. the values of estimators may be greater or less than actual mood score - therefore, we need to measure error in absolute values.

Another thing is that, as the experiments showed, only seven agents in the model may not be a sufficient number to approximate the global social mood M

well. We may therefore want to have more than just one representative of each mood dimension. To achieve this, we introduce splitting parameter S. We then employ the same grouping procedure to tweets set as before, but after obtaining seven groups for seven different mood dimensions, we split each group into S smaller groups of the same size. In this paper $S \in \{1, 2, ..., 5\}$, therefore we conducted experiments for numbers of 7, 14, 21, ..., 35 agents in the model.

Evolutionary algorithm The main parameters that need to be adjusted in social mood model to reflect the real-world data are the network topology, $\mathit{\Delta}_0$, φ and δ parameters of the agents. In our approach, we start with a random graph or scale-free graph network topology (with the same probability), random values of $\mathit{\Delta}_0$ vector generated independently from [-1,1] interval and random values of φ and δ parameters generated independently from $[0, 1]$ interval. Then, we adjust these parameters using an evolutionary algorithm which is defined in the following way.

We start with population of $P = 100$ randomly generated models of social mood[3]. For each model we conduct the simulation and calculate the mean absolute percentage error (MAPE). $MAPE = [MAPE_1, MAPE_2, ..., MAPE_7]$ is also a vector, because there are 7 mood dimensions, so we calculate the mean value of this vector coordinates obtaining our final error Er.

$$Er = \frac{1}{7} \sum_{i=1}^{7} MAPE_i \tag{14}$$

We then sort our models ascending by the value of Er and build next generation of models in the following way. The m fittest models (where m is equal to 50% in our simulations) are retained in the next generation and the others are discarded. A single mutated copy is made of each remaining model so that the size of the population always remains constant. Mutations are applied to r agents from a particular model (where r is equal to 10%) and can take four forms with equal probability:

1. The agent receives new values of *influence* φ and *impressionability* δ parameters. They are generated independently and randomly from $[0, 1]$ interval.
2. A new link in the network is added between the agent and different, randomly chosen agent.
3. An existing, randomly chosen link of the agent is removed from the network.
4. The agent receives new values of $\mathit{\Delta}_0$ vector. They are generated independently and randomly from $[-1, 1]$ interval.

After G generations, we obtain the model with the least error Er which is the best fit to the data.

[3] Some parameters of the evolutionary algorithm are constrained (eg. P = 100, m = 50%, r = 10%). The values were handpicked to optimize the performance.

3 Experiments

In order to evaluate the framework, the experiments are conducted to see how well can it approximate the social mood changes and predict the future values of social mood.

3.1 Approximations of mood changes

To evaluate the quality of the framework's approximations of social mood changes, for each value of the splitting parameter S and for the number of generations $G = 300$, we test it on 60 one-hour-long time intervals. They span across 10 days in July 2014, from 7th till 11th and from 14th till 18th. The periods of time lay between 9:45 and 15:45 EST, as this is the time when New York Stock Exchange is opened (actually it is 9:30 - 16:00, but first and last quarters are the most unstable, that is why we do not want to include them). Much of the research on social mood and electronic sentiment is focused on finding financial applications, therefore we wanted to follow that trend. Each time period starts at $[H{:}45, H{:}50)$, which is the starting point, and then there are 11 time intervals that are approximated $[H{:}50, H{:}55)$, $[H{:}55, H + 1{:}00]$, ..., $[H + 1{:}40, H + 1{:}45)$. Thus, we test the approximation on $10d \times 6h \times 11 = 660$ time slots. The results can be seen in the Table 1. They were obtained against a benchmark of $G = 300$ generations in the evolutionary algorithm. These outcomes can be further improved if the computations are longer (e.g. for $S = 1$ we can achieve 2 percentage point better results if we set G to 600).

Table 1. Two tables present mean value, median and standard deviation of: approximation MAPEs (on the left-hand side) and prediction MAPEs (on the right-hand side), for each value of the splitting parameter S.

S	Mean	Median	SD
1	10.69%	9.45%	4.05
2	10.11%	8.99%	3.84
3	9.49%	8.72%	3.64
4	9.49%	8.74%	3.22
5	9.19%	8.49%	3.29

S	Mean	Median	SD
1	18.63%	17.55%	8.32
2	28.38%	22.90%	17.56
3	27.75%	21.85%	16.24
4	27.30%	25.63%	12.44
5	28.30%	23.70%	14.74

3.2 Predictions of mood changes

In order to evaluate the predictive power of the models, for each value of the splitting parameter S and for each model computed in previous subsection between 9:45 and 14:45 EST, we predict twelve following five-minutes-long time intervals. Therefore, on each of ten days, for five different starting hours, we predict twelve time intervals. Thus, in our experiment we predict $10d \times 5h \times 12 = 600$ time slots. The results can be seen in the Table 1.

4 Discussion

The comparison of social mood approximations for different values of the splitting parameter S confirms the intuitive anticipation that the larger the value is, the better are the approximations (in a matter of fact concerning all comparison indicators: mean, median and standard deviation). One could suspect this fact. In larger graphs there are more agents and more connections among them. Thus, it can be easier for the model to tune to the data. On the other hand, the person studying the graphical model would like to have as small network as possible. They are then easier to analyze and to understand. As the experiments show, the approximations of the models with smaller splitting parameters are worse by around one percentage point. In most cases, this should still be a satisfactory level of error, which one can accept for the sake of the clarity of the graphical model.

As far as the predictive power of models is concerned, the problem with social mood assessed by the means of Twitter is that this kind of system is not closed. In other words, external factors affect the social mood on Twitter, and not only users influence each other. Therefore, prediction power of a particular model is limited by the way of how the next time interval is similar to the previous one in terms of mood changes dynamics.

During our experiments the predictive power of models with the splitting parameter $S = 1$ turns out to be the best, even though they are not the best fit to the data. Most probably it is due to the overfitting of models with greater splitting parameter. In addition to this, in case of $S = 1$ there is no noise created by the interactions between the representatives of the same mood. In the Figure 1, MAPEs of ten predicted time intervals for different values of S are presented. One can notice said smaller amounts of noise for $S = 1$.

Topologies of the evolved networks are something that could be a topic of a separate investigation. From the models that we obtained during our experiments, we can state that they are different for different moments of time, concerning not only their shapes but also their parameters. These facts are not surprising and are probably due to the fact that in different moments of time people were exposed to different external factors. The question of what kind of social situation causes which kind of network topology would be an interesting issue for the future research.

From our conclusions about topologies, first notable fact is that, during the evolution, bigger networks lose their "scale-free property" (understood as a degree distribution following the power law in a graph which is not infinite). Some models' degree distributions look quite similar to the distributions following the power law, but still are disturbed. Rest of the networks turn into more random graphs.

Another fact concerning bigger networks produced by the algorithm (with $S > 2$) is that they are not really easy to read and analyze. Each mood dimension have a few representatives, but usually only some of them are connected to others. It is not clear how someone should interpret such a graph. We can say that only part of the people with particular dominant mood dimension is engaged

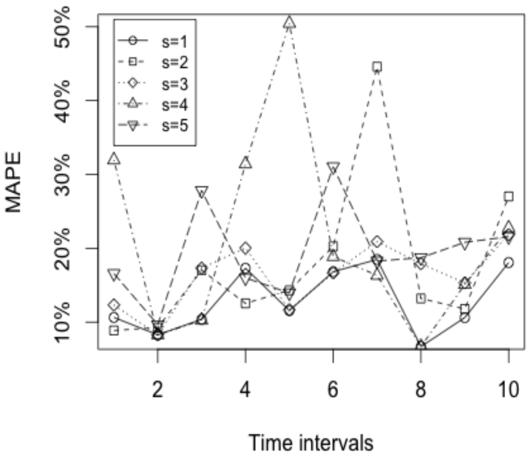


Fig. 1. Figure presents MAPEs of ten predicted time periods for different values of the splitting parameter S.

in interactions, but still conclusions are chaotic. Another issue is the interaction between representatives of the same mood dimension. The aim of the framework is to translate the complexity to something simple. It is not obvious if bigger networks can make that much of a simplification.

This brings our attention to smaller networks. Small graphs, with only one representative for each mood dimension ($S = 1$), do not possess the problems stated above. They also have bigger predictive power of social mood changes (only the approximation is a little bit worse, but as it was stated earlier - it is satisfactory). Thus, we may recommend them as a better source of information and a better tool to investigate social mood. We can see an example of such a network in Figure 2.

5 Conclusion

In this paper, we investigate whether a complex network of individual emotions influencing each other can be approximated by a small graph with similar properties. Our experiments show that small networks can indeed approximate social mood with reasonable mean absolute percentage errors ranging from 9.19% to 10.69%. These results can be further improved using longer computations. Our studies show also that if the following period of time is similar to the previous one, meaning that it is not affected by big amount of external factors, models can be even predictive of the future social mood states. The models with the

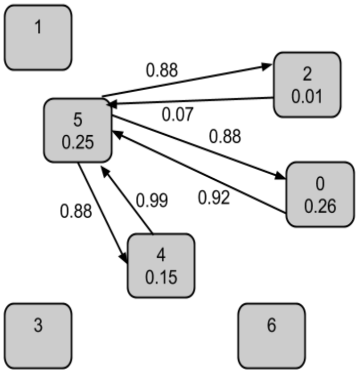


Fig. 2. Figure presents a simple graph (graphical model for case of $S = 1$) in which each node is an agent representing one of the mood dimensions: 0 - Tension, 1 - Happiness, 2 - Calmness, 3 - Vigor, 4 - Fatigue, 5 - Confusion and 6 - Friendliness. Values of *impressionability* parameters are presented inside of the nodes. *Influence* parameters are placed next to the edges.

splitting parameter $S = 1$ can predict following twelve intervals of time with mean MAPE = 18.63%.

In the literature, there are hardly any data-driven models of social mood dynamics, as well as data-driven models of social influence (there are models of these phenomena, but they are not data-driven). Thus, our attempt proposes quite complete framework of assessing and analyzing social mood based on the real-world data. One could argue, that other techniques could be used to approximate/predict social mood like Markov Models, Conditional Random Fields etc. Although, the benefit of our approach is that it creates "the map" of interactions between the mood dimensions. We are able to notice which dimension is connected to which one. We can therefore infer where the influences are and how strong they are (looking at the values of *influence* and *impressionability* parameters).

Presented framework may be useful in situations where quick information about emotion dynamics is needed. For instance, for people responsible for communication or public relations. Different situations may include, for example, financial intraday trading: algorithmic as well as conducted by human. On the other hand, when the temporal resolution of mood measurements is changed to longer periods of time, like days for instance, the framework can also be useful for long-term analysis of social mood dynamics. Such investigations might be of high interest for variety of institutions monitoring societies.

As far as future work in concerned, one can notice, that during the evolution most of the networks lose their scale-free property. Therefore, it is not clear if this is a good starting point of evolutionary algorithm. One could consider different initial topologies. Another interesting question is what types of graphs are being produced by the framework. Is it a single class or a group of them?

Apart from further work on the framework itself or studying types of networks that it produces, there are also issues more of a sociological nature. One could investigate whether different network topologies and model parameters are somehow related to the nature of real-world social events. We hypothesize that deeper studies of graphs' "shapes" and distributions of *influence/impressionability* measures can give interesting conclusions about society dynamics, as well as reflect its qualitative properties. Such correlations between social reality and topologies of small graphs could be of a great interest for social scientists.

6 Acknowledgments

The paper is cofounded by the European Union from resources of the European Social Fund. Project PO KL "Information technologies: Research and their interdisciplinary applications".

References

1. R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
2. P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.
3. S. Asur and B. A. Huberman. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 492–499. IEEE, 2010.
4. A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3):590–614, 2002.
5. A. Bechara and A. R. Damasio. The somatic marker hypothesis: A neural theory of economic decision. *Games and economic behavior*, 52(2):336–372, 2005.
6. J. Bollen, H. Mao, and A. Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*, 2011.
7. J. Bollen, H. Mao, and X. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.
8. A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
9. A. Damasio. *Descartes' error: Emotion, reason and the human brain.* Random House, 2008.
10. G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3(01n04):87–98, 2000.
11. R. J. Dolan. Emotion, cognition, and behavior. *Science*, 298(5596):1191–1194, 2002.
12. A. Edmans, D. Garcia, and Ø. Norli. Sports sentiment and stock returns. *The Journal of Finance*, 62(4):1967–1998, 2007.
13. M. Friedman. *Essays in positive economics.* University of Chicago Press, 1953.
14. S. Galam, Y. Gefen, and Y. Shapir. Sociophysics: A new approach of sociological collective behaviour. *Journal of Mathematical Sociology*, 9(1):1–13, 1982.
15. D. Gayo-Avello, P. Metaxas, and E. Mustafaraj. Limits of electoral predictions using Twitter. *International AAAI Conference on Weblogs and Social Media*, 2011.

16. R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
17. D.-I. O. Hein, D.-W.-I. M. Schwind, and W. König. Scale-free networks. *Wirtschaftsinformatik*, 48(4):267–275, 2006.
18. D. Hirshleifer and T. Shumway. Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032, 2003.
19. R. A. Holley and T. M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, pages 643–663, 1975.
20. E. Kurmyshev, H. A. Juárez, and R. A. González-Silva. Dynamics of bounded confidence opinion in heterogeneous social networks: Concord against partial antagonism. *Physica A: Statistical Mechanics and its Applications*, 390(16):2945–2955, 2011.
21. J. Lorenz. Heterogeneous bounds of confidence: Meet, discuss and find consensus! *Complexity*, 15(4):43–52, 2010.
22. D. McNair, M. Lorr, and L. Droppleman. *Profile of mood states.* San Diego, CA: Educational and Industrial Testing Service., 1971.
23. J. R. Nofsinger. Social mood and financial economics. *The Journal of Behavioral Finance*, 6(3):144–160, 2005.
24. A. Nowak, J. Szamrej, and B. Latané. From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review*, 97(3):362, 1990.
25. B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *International AAAI Conference on Weblogs and Social Media*, 2010.
26. K. R. Olson. A literature review of social mood. *The Journal of Behavioral Finance*, 7(4):193–203, 2006.
27. A. Pak and P. Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, 2010.
28. Z. Pan. Trust, influence, and convergence of behavior in social networks. *Mathematical Social Sciences*, 60(1):69–78, 2010.
29. H.-R. Pfister and G. Böhm. The multiplicity of emotions: A framework of emotional functions in decision making. *Judgment and Decision Making*, 3(1):5–17, 2008.
30. R. R. Prechter Jr. Whats going on? *Pioneering studies in socionomics*, page 1, 1979.
31. W. M. Rahn, B. Kroeger, and C. M. Kite. A framework for the study of public mood. *Political Psychology*, pages 29–58, 1996.
32. K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *International Journal of Modern Physics C*, 11(06):1157–1165, 2000.
33. M. Thelwall, K. Buckley, and G. Paltoglou. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173, 2012.
34. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas. Sentiment strength detection in short informal text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.
35. D. Urbig, J. Lorenz, and H. Herzberg. Opinion dynamics: The effect of the number of peers met at once. *Journal of Artificial Societies and Social Simulation*, 11(2), 2008.

My MOoD, a Multimedia and Multilingual Ontology Driven MAS: Design and First Experiments in the Sentiment Analysis Domain

Maurizio Leotta, Silvio Beux, Viviana Mascardi, and Daniela Briola

DIBRIS, University of Genova, Italy
maurizio.leotta@unige.it, silviobeux@gmail.com,
viviana.mascardi@unige.it, daniela.briola@unige.it

Abstract. In this paper we introduce the architecture of a *M*ultimedia and *M*ultilingual *O*ntology *D*riven Multiagent System (*My MOoD*) for classifying documents consisting of audiovisual and textual elements, according to classes described in a domain ontology. My MOoD will integrate software components devoted to the analysis of images, videos, and sound, with the multilingual text classifier based on BabelNet presented in this paper. All the integrated components will be wrapped by agents and will perform their classification based on a common domain ontology, which is a parameter of the multiagent system. Wrapper agents will interact in order to share the classification of the document's elements and agree on a coherent classification of the document as a whole, exploiting their background knowledge and reasoning capability to resolve ambiguities. Changing the ontology (and tuning or substituting the classifiers for dealing with the domain of interest) will allow the multiagent system to classify heterogeneous multimedia documents in whatever domain and for many different purposes. In the My MOoD instance discussed in this paper, the ontology (*sentiHotel*) describes the accommodation domain and the classification mines the sentiment of hotel reviews written in five different languages.

Keywords: Multimedia, Multilingual, Multiagent, Ontology, BabelNet

1 Introduction and Motivation

When it was born, at the beginning of the new millennium, sentiment analysis was conceived as a research area addressing text only, written in only one language. Because of the lack of multimedia social networks which were limited, at that time, to Friendster (2002), MySpace, LinkedIn and Hi5 (2003), Flickr and Facebook (2004), and the hardness of managing multilingual and multimedia objects, it is no surprise that the seminal works by Turney [34] and Pang et al. [26] published in 2002 had monolingual textual documents as their sole target. The well known article by Pang and Lee dating back to 2008 [25] defines opinion mining and sentiment analysis as areas dealing with the computational treatment of opinion, sentiment, and subjectivity in text, and even the most recent surveys on the topic [20, 35] do not consider the possibility to extract sentiments from objects rather than text. While the problem of multilingualism was addressed starting from 2007 [1, 12, 22], researchers drove their interest towards multimedia contents including images and video as valuable sources of opinions only in the last five years. Starting from 2010, visual sentiment analysis [6, 7, 31, 38, 39] emerged as an area complementing that of textual sentiment analysis, aiming at extracting the polarity conveyed by visual content, including movies. Multimodal sentiment analysis, taking spoken content into account [16, 30], is an even more recent approach.

Being able to extract and analyze emotions and opinions from multimedia, multimodal and multilingual social objects such as news, tweets, blogs, etc, would of course give great advantages, including economic ones. Strengthening the polarity of a written opinion because of images or spoken sentences that support it, or, on the other hand, making a deeper analysis when texts and videos referring to the same event seem to express different emotions, would give more precise and reliable (and hence, more precious and valuable) results. However, the complexity of each individual task involved in the multimedia and multilingual sentiment analysis process makes it so challenging that only a "divide et impera" approach, dividing the burden of the challenge among many intelligent, autonomous and cooperating entities, can work.

An intelligent software agent is a software component which is situated (receives sensory input from its environment and can perform actions that change the environment in some way), autonomous (acts without the direct intervention of humans or other agents and has control over its own actions and internal state), responsive (perceives its environment and responds in a timely fashion to changes that occur in it), pro-active (exhibits opportunistic, goal-directed behaviour and takes the initiative where appropriate) and social (interacts, when appropriate, with other artificial agents and humans in order to complete its own problem solving and to help others in their activities) [17]. A multiagent system, or MAS for short, is a system designed and implemented as several interacting agents. Quoting [17] again, *"multiagent systems are ideally suited to representing problems that have multiple problem solving methods, multiple perspectives and/or multiple problem solving entities"*. The problem of classifying different elements of a complex multimedia object, each of which may be a piece of text expressed in different languages, a fragment of audio or video track, an image, a manual sketch, and then combining these classifications to provide a coherent and meaningful classification of the object as a whole, requires to involve multiple problem solving entities (the classifiers) and to coordinate their outcomes in a non trivial way. A MAS is thus an extremely suitable approach for facing such a complex problem.

In this paper we present the design of the *My MOoD MAS* (My MOoD in the sequel), a general purpose *M*ultimedia and *M*ultilingual *O*ntology *D*riven multiagent system, and the first experiments to mine the polarity of multilingual texts exploiting the *SentiHotel* ontology. Although My MOoD is still in its design stage, we are confident that, once implemented, it will ensure the modularity, flexibility and scalability required for tackling a challenging task like multimedia and multilingual sentiment analysis.

The paper's structure is the following: Section 2 discusses the state of the art. Section 3 introduces the architecture of My MOoD. Section 4 describes the Multilingual Text Classifier. Section 5 describes the SentiHotel ontology. Section 6 discusses the results of the experiments carried out with hotel reviews in five languages. Section 7 concludes and highlights some directions for the future work.

2 State of the Art

Multilingual sentiment analysis. In [22], Mihalcea et al. explore methods for generating subjectivity analysis – namely identifying when a private state is being expressed and identifying attributes of that private state including who is expressing the private state, the type(s) of attitude being expressed, about whom or what the private state is being expressed, the intensity of the private state, etc. [37] – in a target language L by exploiting tools and resources available in English. Given a bilingual dictionary or a parallel corpus acting as a bridge between English and the selected target language L, the methods can be used to create tools for subjectivity analysis in L. Experiments are carried out with Romanian. Ahmad et al. [1] classify sentiments within a multilingual framework (English, Arabic, and Chinese) following a local grammar approach. Domain-specific keywords are selected by comparing the distribution of words in a domain-specific document to the distribution of words in a general language corpus. Words less prolific in a general language corpus are considered to be keywords. Denecke [12] introduces a methodology based on lexical resources for sentiment analysis available in English (SentiWordNet, http://sentiwordnet.isti.cnr.it/) for determining polarity of text within a multilingual framework. The method is tested for German movie reviews selected from Amazon and is compared to a statistical polarity classifier based on n-grams. The paper by Boiy and Moens [5] describes machine learning experiments with regard to sentiment analysis in blog, review and forum texts found on the World Wide Web and written in English, Dutch and French. The proposed approach combines methods from information retrieval, natural language processing and machine learning. An automated sentiment analysis on multilingual user generated contents from various social media and e-mails is described in [33]. The sentiment analysis is based on a four-step approach including language identification for short texts, part-of-speech tagging, subjectivity detection and polarity detection techniques. The prototype has been tested on English and Dutch. More recently, the paper [3] presents an evaluation of the use of machine translation to obtain and employ data for training multilingual sentiment classifiers. The authors demonstrate that the use of multilingual data, including that obtained through machine translation, leads to improved results in sentiment classification and that the performance of the sentiment classifiers built on machine translated data can be improved using original data from the target language. The languages explored by the authors are Turkish, Italian, Spanish, German and French. Finally, the paper [14] describes the adoption of meta-learning techniques to combine and enrich existing approaches to single and cross-domain polarity classification based on bag of words, n-grams or lexical resources, adding also other knowledge-based features. The proposed system uses the BabelNet multilingual semantic network [24] to generate word sense disambiguation and vocabulary expansion-derived features. Being based on BabelNet, the system can cope with multilingual documents. By now its evaluation has been carried out on a monolingual dataset, the Multi-Domain Sentiment Dataset (version 2.0, http://www.cs.jhu.edu/ mdredze/datasets/sentiment/). Evaluating the polarity classification approach in other languages is part of the authors' future work.

Ontology driven sentiment analysis. One of the first papers on ontology-based sentiment classification is [29], where the ontology was used to classify and analyze online product reviews by providing lexical variations and synonyms of terms that could be met in the reviews. In [10], Chaves and Trojahn present Hontology, a multilingual ontology for the hotel domain. Hontology has been proposed in the context of a framework for ontology-driven mining of Social Web sites contents. Comments are annotated with concepts of Hontology, which are manually labeled in Portuguese, Spanish and French. Hontology reuses concepts of other vocabularies such as Dbpedia.org and Schema.org. The work on Hontology was further expanded in [11]. ArsEmotica [4] is a software application for associating the predominant emotions with artistic resources of a social tagging platform. A rich emotional semantics (i.e., not limited to a positive or a negative opinion) is extracted from tagged resources through an ontology driven approach. The ArsEmotica Ontology (AEO [27]) is based on Plutchik's model [28] and incorporates, in a unifying model, multiple ontologies which describe different aspects of the connections between media objects (e.g., the ArsMeteo artworks, http://www.arsmeteo.org/), persons and emotions. In particular, it includes an ontology of emotions which have been linked, via `owl:sameAs`, to the corresponding emotions in DBpedia. Furthermore, it incorporates an ontology of artifacts, derived from the alignment of a domain ontology obtained from the DB of the ArsMeteo on line portal, with the OMR (Ontology for Media Resources, http://www.w3.org/TR/mediaont-10/). The paper [19] proposes the deployment of original ontology-based techniques towards a more efficient sentiment analysis of Twitter posts. The novelty of the proposed approach is that posts are not simply characterized by a sentiment score, as is the case with machine learning-based classifiers, but instead receive a sentiment grade for each distinct notion in the post. The proposed architecture aims at providing a more detailed analysis of post opinions regarding a specific topic.

Multiagent systems for sentiment analysis. While we know many papers dealing with agents which show emotions and sentiments, we are aware of only two papers where agents are used to analyze sentiments of documents. In [2] a MAS exploiting machine learning classification for analyzing the sentiment of product features in different social media sources is presented. The MAS exploits different agents to deal with different kind of information from different social media networks. Agents communicate and interact with each other to learn new information. Kechaou et al. [18] describe a MAS based on a thorough linguistic analysis which enables to resolve the ambiguities and complexities of the natural evaluative language and to strengthen, as well as consolidate, the results achieved at the various analysis stages. We are not aware of other agent-based approaches to sentiment analysis.

Comparison. While exploiting an ontology for driving the sentiment analysis is far from being an original idea and the papers on this topic are much more than those that we mentioned in this section, exploiting a MAS for that purpose seems to have received little attention by the research community. The two MASs we are aware of address textual documents only, and written in only one language. From this point of view our proposal, albeit preliminary, seems to be an original one. With respect to the existing literature on multilingual sentiment analysis, our work is among the few ones that perform an evaluation involving five languages, hence demonstrating the actual multilingualism and flexibility of the approach. As far as the adopted tools are concerned, the work closer to our is [14] for the heavy exploitation of BabelNet.

3 MAS Architecture

IndianaMAS [21] is a project funded by the Italian Ministry for Education, University and Research, MIUR, spanning from March 2012 to February 2015. It integrates intelligent software agents, ontologies, multilingual natural language processing, sketch and image recognition techniques to develop a technology platform for the digital preservation of rock carvings. The IndianaMAS platform has been conceived as a general, scalable and flexible holonic MAS, namely a MAS consisting of components which are at the same time "part" of a bigger MAS (the MAS that contains them), as well as independent MASs [15]. Classification of texts, images and sketches is driven by an ontology [8] named *Indiana Ontology*, modeling information about Mount Bego's prehistoric rock art. For reaching the same objectives as the IndianaMAS project, but in a different domain, the ontology can be changed with any other ontology from any other domain, keeping the general MAS architecture almost unchanged: image and sketch recognition algorithms must of course be modified in order to recognize images and sketches in the domain of interest; classifiers for audio and video tracks must be added if the input documents contain elements of this kind; multilingual text classification, instead, requires limited or no tuning at all as the only assumption it makes is the existence of an ontology modeling those concepts according to which the classification must be performed.

Given the raising importance of sentiment analysis in social and expressive media, we investigated how to move from the IndianaMAS for the rock art domain to a more general MAS for classifying multimedia and multilingual documents consisting of text, sketches, drawing, images, but also video and audio tracks. The result of our investigation is the My MOoD MAS shown in Figure 1.

Our research is currently targeted to the ontology-driven classification of multilingual textual documents only: the Multilingual Text Classifier Agent MuTCA wrapping the Multilingual Text Classifier in Figure 1 is highlighted for this reason. Since emotions can be extracted from audiovisual content as well, as witnessed by the literature on visual and multimodal sentiment analysis, and since - although, to the best of our knowledge, not yet addressed by the research community - it should be possible to extract the polarity of manual sketches exploiting techniques similar to those described in [9], agents for classifying movies, audio tracks, images and manual sketches have been included in the My MOoD architecture as well.

The interaction among these different agents and holons will allow My MOoD to correctly interpret multimedia contents also in case of ambiguous classifications. Consider for example an image with a woman wearing an elegant long white dress, whose expression is clearly touched. If the agent devoted to image recognition can extract concepts like "woman", "elegant" "white dress", and "moved" from the picture, and the domain of interest deals with religious celebrations including "wedding" and "funeral", the correct classification could be both of them: in Hindu tradition, in fact, white is the standard color for funerals and the woman might be a related of the deceased, whereas in Western cultures a woman dressed in white attending a religious celebration is likely to be the bride. If the textual caption of the image says nothing about the event, but states that it took place in New Delhi, then intelligent agents able to reason about all the information extracted from the document, including geographical data, can agree that the picture shows a Hindu funeral.

The analyzed documents can be stored – temporarily or permanently – into an internal DataBase together with their classification. The DataBase can also store aggregated results. The user of My MOoD can perform queries on the stored data. Queries will be based on the ontology, which is the core of the system and the driver of the domain modeling and of the document classification.

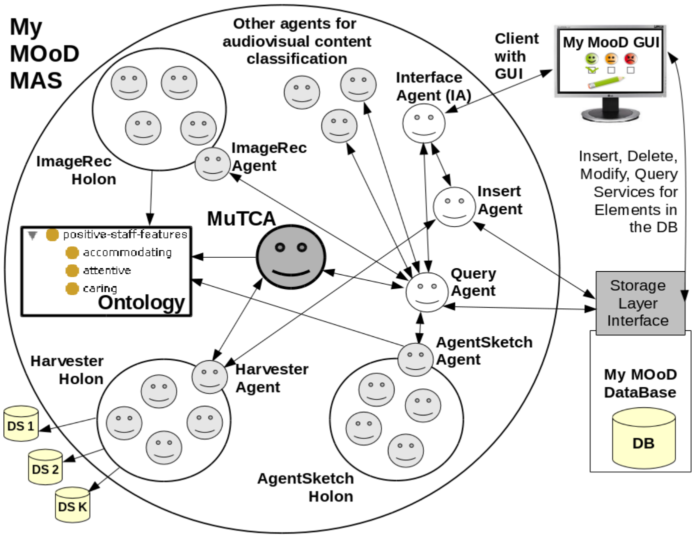


Fig. 1. My MOoD architecture (for sake of clarity, not all the arrows modeling control and data flow between components are shown).

The main components of My MOoD will be:

- The ontology, which structures the domain of interest of the project. In IndianaMAS the domain was that of Mount Bego rock art modeled by the Indiana ontology, while in this paper it is that of opinions about hotels modeled by the SentiHotel ontology.
- The Multilingual Text Classifier Agent MuTCA.
- The other components that we implemented and tested in IndianaMAS and that we will reuse, after the required tuning and integration, namely:
  - the AgentSketch holon, for interpreting manual drawings and sketches;
  - the ImageRec holon, for recognizing and classifying images;
  - a holon for searching the Web to retrieve documents that meet the user's needs and requirements;
  - the Interface, Insert and Query Manager Agents, for providing an interface between the MAS and the user and for offering operations over data in the DataBase.
- Additonal agents for classifying other kinds of digital objects.
- The internal DataBase, to store multimedia documents (or their references/URLs) that have been classified.
- The Web Interface, to let users perform operations on data (look for new data on the web, store retrieved data, analyze and query them).

4 Ontology-driven Multilingual Text Classifier

Classifying a document can be defined as the task of assigning it to one or more classes or categories. For instance, we might want to classify a text w.r.t. a set of geographical, historical, and topic classes (e.g., understanding whether a text is about the neolithic rock art in France, as we did in the IndianaMAS project). Our Multilingual Text Classifier (TEXTCLASS in the sequel), designed and implemented to face such a classification task, takes in input (1) an ontology whose classes model the domain of interest and whose names are expressed in any language from a predefined set[1] and (2) a document containing the text to classify written in any language from the above set. It returns a classification of the text w.r.t. the ontology taken in input. The classification performed by the TEXTCLASS is multilingual and exploits BabelNet and WordNet.

WordNet[2] [13, 23] is the main resource for lexical knowledge upon which BabelNet is based. WordNet groups English words into sets of synonyms called synsets. A label that indicates the part of speech (e.g., n means noun) and sense number is associated with each word in the synset. Words are assigned sense numbers based on frequency of use in semantically tagged corpora. Senses in WordNet are generally ordered from most to least frequently used, with the most common sense numbered 1 . Frequency of use is determined by the number of times a sense is tagged in the various semantic concordance texts. To make an example, a sysnset can be of the form:

$$\{play_{1n}, drama_{1n}, dramatic\_play_{1n}\}$$

WordNet also provides a textual definition (gloss) for each synset. The major weakness of WordNet is that it available for English only; BabelNet was born to overcome this limitation.

BabelNet[3] [24] is a very large multilingual semantic network, based on the automatic mapping of concepts onto WordNet and Wikipedia[4], the largest multilingual Web encyclopedia. The result is an "encyclopedic dictionary", in which words (Babel Senses) in different languages (BabelNet 3.0 supports 271 languages including all European languages, most Asian languages, and even Latin) are grouped into sets of synonyms called Babel Synsets. Each Babel Synset has different features like shorts definitions (glosses) in many languages harvested from both WordNet and Wikipedia, and many relations in the semantic network provided by WordNet (e.g., hypernymy and hyponymy, meronymy and holonymy, antonymy and synonymy, etc.).

Given an ontology o and a document d to classify, TEXTCLASS identifies the classes in o which d belongs to. For instance, in case of a geographic ontology, TEXTCLASS associates with each document (for example, a tourist guide) the geographical place(s) that it describes.

The strengths of TEXTCLASS are the following: (1) it is able to classify documents described in several languages (2) using ontologies in different languages; (3) the languages used in the documents and in the ontologies can be different; (4) there is no need to state in advance the languages of the ontologies and documents, as TEXTCLASS can automatically recognize them[5]; (5) the documents' format can be either plain text or pdf; and (6) documents can be classified w.r.t. different ontologies in a single step

[1] In theory we could cope with any of the languages supported by BabelNet; in practice, if we want to apply a stemming stage to words as we actually do for obtaining acceptable results, we can manage only those for which the Porter stemmer is implemented. If a stemmer is not available, stemming could even be avoided, but we would expect poor results without it.

[2] http://wordnet.princeton.edu/

[3] http://babelnet.org/

[4] https://www.wikipedia.org/

[5] The automatic language recognition feature is currently implemented for texts but not for ontologies; extending it to ontologies would be straightforward.

(provided that all the ontologies are described in the same language). In the following, to keep the description simpler, we describe the functioning of TEXTCLASS when only one ontology is used.

More in detail, TEXTCLASS (1) extracts the text T, i.e. a list of words, from d (if d is not on the computer, downloads the document from the URL); (2) detects the language l used in T; (3) translates each word $w \in T$ into the language of the ontology using BabelNet and WordNet; (4) classifies T w.r.t. o; and finally (5) returns the classification.

The current prototype of TEXTCLASS integrates one module for each step above and has been developed in Java on the Ubuntu 14.04.1 Linux platform.

Extracting Text from Document (Module 1). This module is devoted to extracting the text T (a list of words) contained in document d.

T = extractText(URL or FilePath of d)

Implementation Details: The document can be provided to TEXTCLASS in two ways: (1) it could be already saved in a local directory (e.g., /home/user/text/sample.pdf) or (2) it could be available online (e.g., http://site/sample.pdf). In the latter case, the file is downloaded in a temporary folder by using the copyURLToFile(...) method provided by the org.apache.commons.io.FileUtils library. Then, the file is read by using different methods depending by the file type. TEXTCLASS currently supports txt and pdf files.
In both cases the file is opened and its textual content loaded, cleaned (i.e., substituting all the occurrences of multiple white spaces or non visible characters such as tab and newline with a single white space) and assigned to a list of String T that is provided to Module 2.

Detecting the Language of the Text (Module 2). This module is devoted to recognizing the language L_T used in text T extracted from document d. This step is necessary because the following modules need to know the language of the document. TEXTCLASS adopts a naive Bayes with character n-gram for fast language detection.

L_T = detectLanguage(T)

Implementation Details: TEXTCLASS employs the Language-Detection library[6] that is able to detect, with a precision greater than 99%, 53 languages making use of naive Bayesian filters. In particular, TEXTCLASS analyses the text T provided by the previous module and, depending on its length, calls the language detector library using different profiles. Indeed, in case of very short texts (few words), it is recommended to use specific profiles rather than the standard ones. To speed up the language detection, TEXTCLASS avoids to provide the complete text of the document to the language detector, given that, potentially, TEXTCLASS could be required to classify documents long tens or hundreds pages. From our experiments, we noticed that using the first 100 words of the text (e.g., about 500-800 characters) provides very good results in terms of both precision and performance.

Translating Text (Module 3). The main goal of this module is to translate each word of the text T into the language L_o used to describe the ontology (L_o is an information associated with each ontology). For each word $w \in T$, two steps are performed.

[6] https://code.google.com/p/language-detection/

– First, all the synsets containing the word w are retrieved. Note that w is supposed to belong to the language L_T. Obviously w can appear in more than one synset.
  For instance, in case of an Italian text containing the word “pulito”, the BabelNet function getSynsets is called with the parameters L_T=Language.IT and w=“pulito”, and returns a set of synsets S. Indeed, “pulito” in Italian has different meanings[7], including for instance: (1) free from dirt or impurities[8], (2) characterized by freedom from troubling thoughts (especially guilt)[9].
– Second, given L_o the target language used in the ontology, all the words associated with each synset $s \in S$ in the language L_o are retrieved by means of the BabelNet function getSenses. In the case of the word w=“pulito” and L_o=Language.EN, we obtain several translations including: clear, clean, neat, uncontaminated, orderly, elegantly, untarnished, untainted, unstained, stainless, unsullied.

T' = translateText(T, L_T, L_O) T' is a set containing a list for each $w \in T$.

If no translation is found for w, the corresponding list contains only w, otherwise it contains all the computed translations. Each list is also associated with how many times the original word w was found in the text T.
Implementation Details: Since such operations are repeated for each word and are time consuming (the BabelNet indexes have a total size of about 30GB), we execute a preprocessing step that consists in (1) removing all the stop words[10] from text T[11] obtaining a cleaned text in L_T, and (2) searching the translations of each word only once even if it is repeated multiple times in the original text.

Assigning Weights to the Ontology Nodes (Module 4). In this phase, the ontology nodes are labeled with weights in order to consider the frequency of the corresponding terms in the text. In detail, each node in o is compared with all the elements (i.e., words) of all the lists in T'. Every time a match is found, the label containing the weight of the node is increased by the value associated with the list containing the matching word.

o_W = assignWeights(o) o_W is the weighted ontology.

Implementation Details: the ontology that drives the classification is expressed in OWL and is navigated and manipulated by means of the Jena Java framework[12]. To increase the probability of finding a match between the words translated by BabelNet and the words used for labeling the ontology nodes, we reduce the inflected words (both in the ontology and in the list) to their word stem. For this purpose we adopt the Snowball framework[13] by Martin Porter, that contains specific stemming algorithms for 16 languages.

Generating the Final Classification (Module 5). In this phase, all the nodes in o_W are visited and those with a weight greater than 0 are inserted into the result list. The

[7] http://BabelNet.org/search?word=pulito&lang=IT

[8] http://BabelNet.org/synset?word=bn:00099776a&details=1&orig=pulito&lang=IT

[9] http://BabelNet.org/synset?word=bn:00099807a&details=1&orig=pulito&lang=IT

[10] Stop words are words which are filtered out before or after processing of natural language data, http://en.wikipedia.org/wiki/Stop_words

[11] We used the lists of stop words included in the BabelNet API (24 languages supported). Each list typically includes from one to several hundreds of stop words.

[12] https://jena.apache.org/

[13] http://snowball.tartarus.org/texts/introduction.html

list is ordered in decreasing order of weight. For instance, when classifying texts using a geographic ontology, TEXTCLASS can return the following result [[Liguria, 25],[Italy, 12],[Nice, 4],[France, 2]]. This result could be interpreted as: the text T describes something located in Liguria (an Italian region) but also, to a lesser extent, something that concerns Nice, a French city near the border with Italy. We would obtain such a result if, for example, the text was centered around Monte Beigua, located in Liguria, whose name has the same root as Mount Bego, located in France, whose petroglyphs are studied by archaeologists working in Nice.

C = classification(o_W) C is the final classification list.

5 Modeling Opinions in the Accommodation Domain: the SentiHotel Ontology

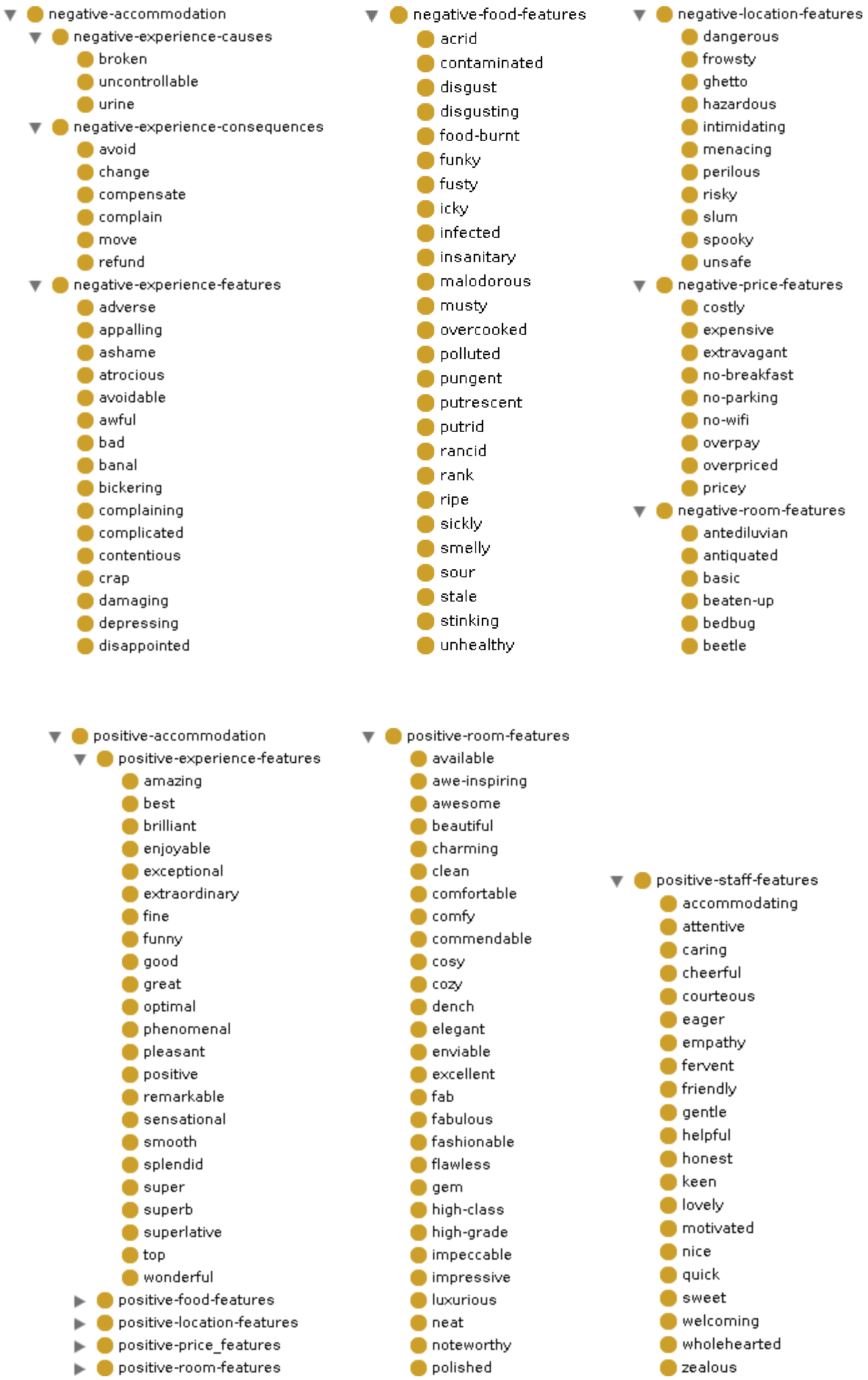
In order to test the behavior of TEXTCLASS with an ontology different from the *Indiana* one, we developed an ontology of opinion words in the accommodation domain that integrates the four emotional branches of *WordNet Affect* [32] (`positive-emotion`, `negative-emotion`, `neutral-emotion` and `ambiguous-emotion`) and added to them about 400 opinion words (as subclasses) based on 30 positive reviews and 30 negative reviews retrieved from [36] and carefully analyzed by the authors to devise the most frequent concepts expressing positive/negative feelings. The ontology was manually developed in OWL Lite using Protégé 3.4.8 (http://protege.stanford.edu/) and is publicly available from http://www.disi.unige.it/person/MascardiV/Download/sentiHotel.owl.

The dataset described in [36] is available to the community and contains reviews from *Tripadvisor* (and other sources, that we did not used because out of scope), which we used to create our ontology and to test the Multilingual Text Classifier, as described in Section 6.

The domain depended opinion words, each mapped into an OWL Class, are divided into a `negative-accommodation` branch divided into eight sub-trees (`negative-experience-causes`, `negative-experience-consequences`, `negative-experience-features`, `negative-food-features`, `negative-location-features`, `negative-price-features`, `negative-room-features`, `negative-staff-features`) and containing 270 classes, and a `positive-accommodation` branch divided into six sub-trees (`positive-experience-features`, `positive-food-features`, `positive-location-features`, `positive-price-features`, `positive-room-features`, `positive-staff-features`) and containing 100 classes. We created no branches for neutral and ambiguous words.

The negative branch is larger than the positive one because reviewers use many different terms to express negative emotions, including impolite and slang words, while they use almost the same terms (“splendid”, “wonderful”, “amazing”, ...) in the positive ones.

In the negative branch, we added two more sub-trees related to the experience, modeling the causes and the consequences of the bad experience: in these branches we added some terms, not essentially “emotions related”, that are often found in negative reviews (for example “refund”, or “broken”). In Figure 2 the reader can see the trees structure of positive and negative branches, with some examples of the terms under



- negative-accommodation
  - negative-experience-causes
    - broken
    - uncontrollable
    - urine
  - negative-experience-consequences
    - avoid
    - change
    - compensate
    - complain
    - move
    - refund
  - negative-experience-features
    - adverse
    - appalling
    - ashame
    - atrocious
    - avoidable
    - awful
    - bad
    - banal
    - bickering
    - complaining
    - complicated
    - contentious
    - crap
    - damaging
    - depressing
    - disappointed

- negative-food-features
  - acrid
  - contaminated
  - disgust
  - disgusting
  - food-burnt
  - funky
  - fusty
  - icky
  - infected
  - insanitary
  - malodorous
  - musty
  - overcooked
  - polluted
  - pungent
  - putrescent
  - putrid
  - rancid
  - rank
  - ripe
  - sickly
  - smelly
  - sour
  - stale
  - stinking
  - unhealthy

- negative-location-features
  - dangerous
  - frowsty
  - ghetto
  - hazardous
  - intimidating
  - menacing
  - perilous
  - risky
  - slum
  - spooky
  - unsafe
- negative-price-features
  - costly
  - expensive
  - extravagant
  - no-breakfast
  - no-parking
  - no-wifi
  - overpay
  - overpriced
  - pricey
- negative-room-features
  - antediluvian
  - antiquated
  - basic
  - beaten-up
  - bedbug
  - beetle

- positive-accommodation
  - positive-experience-features
    - amazing
    - best
    - brilliant
    - enjoyable
    - exceptional
    - extraordinary
    - fine
    - funny
    - good
    - great
    - optimal
    - phenomenal
    - pleasant
    - positive
    - remarkable
    - sensational
    - smooth
    - splendid
    - super
    - superb
    - superlative
    - top
    - wonderful
  - positive-food-features
  - positive-location-features
  - positive-price_features
  - positive-room-features

- positive-room-features
  - available
  - awe-inspiring
  - awesome
  - beautiful
  - charming
  - clean
  - comfortable
  - comfy
  - commendable
  - cosy
  - cozy
  - dench
  - elegant
  - enviable
  - excellent
  - fab
  - fabulous
  - fashionable
  - flawless
  - gem
  - high-class
  - high-grade
  - impeccable
  - impressive
  - luxurious
  - neat
  - noteworthy
  - polished

- positive-staff-features
  - accommodating
  - attentive
  - caring
  - cheerful
  - courteous
  - eager
  - empathy
  - fervent
  - friendly
  - gentle
  - helpful
  - honest
  - keen
  - lovely
  - motivated
  - nice
  - quick
  - sweet
  - welcoming
  - wholehearted
  - zealous

Fig. 2. My MOoD ontology (part of).

each sub-tree: due to space limitation, we cannot describe the complete ontology. The interested reader can retrieve it from the web.

6 Experiments with Hotel Reviews in Five Languages

The research question we tackled to evaluate the effectiveness of TEXTCLASS is:

> **RQ**: *Is* TEXTCLASS *able to classify documents in English, Italian, Spanish, French and German w.r.t. the opinion/sentiment they describe using the SentiHotel ontology?*

The metrics used to answer RQ is the number of documents correctly classified over the total number of documents.

Data set. We conducted our preliminary evaluation of TEXTCLASS over a sample of multilingual reviews from TripAdvisor[14]. In particular, we focused on classifying reviews in English, Italian, Spanish, French and German.

For English reviews we chose Wang TripAdvisor Data Set [36]; this Data Set is composed by more than 12000 Json files each of which contains about 10 TripAdvisor reviews with different information about them (e.g., review text, overall score, ID). From this dataset we randomly chose 455 English reviews with a balanced distribution of different overall scores (i.e., we have a similar number of positive and negative reviews).

For Italian, Spanish, French and German reviews, we randomly selected 25 reviews for each language, 5 for each value of the overall score (from 1 to 5), resulting in a total of 100 reviews.

Procedure. To answer our RQ we proceeded as follows:

- We selected the positive-review and negative-review sub-trees of the SentiHotel ontology. Such sub-trees play respectively the role of positive o_P and negative o_N ontologies during the classification performed by TEXTCLASS.
- For each review, we executed TEXTCLASS and recorded the classification w.r.t. the ontologies o_P and o_N. In particular, we recorded the number of different positive and negative elements in the ontologies o_P and o_N that match at least one word in the text of the review. For instance, for a review we can find that m_P=12 is the total number of matches in the positive ontology o_P while m_N=4 is the total number of matches in the negative ontology o_N.
- For each review, we computed the normalized classification C_{norm} in order to fit the range [1,5] (i.e., the same used by the TripAdvisor's reviews). The formula used is $C_{norm} = 5 - ((4 * m_N)/(m_P + m_N))$. In the previous example we obtain $C_{norm} = 4.00$. We have no cases in which m_P=m_N=0. In the other cases, the formula correctly returns 3 when m_P=m_N, 5 when m_N=0 and 1 when m_P=0.
- We classify each review as positive if $C_{norm} >= Tr$, negative otherwise. We initially set Tr to 3, which is the C_{norm} value returned when m_P=m_N, namely when there are as many negative opinion words as the positive ones. Higher values should indicate a positive polarity and lower values a negative one. As discussed below, the results obtained with Tr equal to 3 were not satisfactory, so we empirically devised another threshold, 3.4, giving better results.

[14] http://www.tripadvisor.com/

– For each review, we compared our classification (i.e., computed as shown in the previous step) with the overall score provided by the real user and recorded in the dataset together with the review. The classification is correct when: (1) we classified a review as positive and the user provided a score $>= 3$, (2) we classified a review as negative and the user provided a score < 3. In the other cases the classification is wrong.

Results. Table 1 reports the data used to answer RQ. For each dataset (i.e., set of reviews in a specific language) and for each overall score (i.e., the number [1,5] assigned by the users), it reports the number of correctly classified reviews and the corresponding percentage over the total number of reviews. In the last columns, we report aggregate results over all the five datasets.

Table 1. TEXTCLASS Classification Results (threshold = 3)

Overall Score	Reviews EN			Reviews IT			Reviews FR			Reviews ES			Reviews DE			Reviews		
	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total
	N	%		N	%		N	%		N	%		N	%		N	%	
5	74	91,4	81	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	93	92,1	101
4	145	90,1	161	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	164	90,6	181
3	77	82,8	93	5	100,0	5	4	80,0	5	4	80,0	5	5	100,0	5	95	84,1	113
2	16	31,4	51	3	60,0	5	2	40,0	5	2	40,0	5	0	0,0	5	23	32,4	71
1	39	56,5	69	2	40,0	5	4	80,0	5	4	80,0	5	4	80,0	5	53	59,6	89

Concerning the reviews with evaluation 5 (i.e., very good) or 4 (good), we can see that TEXTCLASS is able to provide, most of the times, a correct classification. In particular, in case of overall score = 5 and considering all the languages employed in the five datasets, TEXTCLASS correctly classifies the 92.1% of the reviews. In three cases, IT, FR, and DE the classification is perfect. Similarly, TEXTCLASS correctly classifies the 90.6% of the reviews with overall score = 4, and in the cases of IT, FR, and DE the classification is completely correct.

Conversely, TEXTCLASS is not able to classify correctly the reviews with evaluation 1 (i.e., very bad) or 2 (bad). Indeed, respectively only in the 59.6% and 32.4% of the cases it produce a correct results. From the data reported in Table 1, it is evident that the result of the classification is unbalanced, and tends to favor positive ratings.

We reported also the classification returned for the reviews with overall score = 3. They express a judgment that obviously is neither positive nor negative. Thus, a binary classification (i.e., positive vs negative) cannot be used for classifying such kind of reviews. But, for such reviews, we expect TEXTCLASS to behave as a classifier which assigns a review to one of the two classes (positive and negative) with a probability of 50%, while, by adopting the threshold Tr=3, this is not true (see the 84.1% reported in the table). Thus we searched for a threshold value that allows to obtain, for the overall score = 3, a results as close as possible to 50%. Such threshold value is 3.4.

Table 2 reports the results of the classification performed using Tr=3.4. Concerning the reviews with evaluation 5 (i.e., very good) or 1 (very bad), we can see that TEXTCLASS is able to provide, most of the times, a correct classification. In particular, in case of overall score = 5 and considering all the languages employed in the five datasets, TEXTCLASS correctly classifies the 83.2% of the reviews. In three cases, IT, FR, and DE the classification is perfect. Similarly, TEXTCLASS correctly classifies the

92.1% of the reviews with overall score = 1, and in the cases of IT, FR, and ES the classification is completely correct.

Table 2. TextClass Classification Results (threshold = 3.4)

	Reviews EN			Reviews IT			Reviews FR			Reviews ES			Reviews DE			Reviews		
Overall Score	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total	Correctly Classified		Total
	N	%		N	%		N	%		N	%		N	%		N	%	
5	65	80,2	81	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	84	83,2	101
4	116	72,0	161	5	100,0	5	5	100,0	5	4	80,0	5	5	100,0	5	135	74,6	181
3	49	52,7	93	1	20,0	5	2	40,0	5	3	60,0	5	1	20,0	5	56	49,6	113
2	36	70,6	51	4	80,0	5	5	100,0	5	3	60,0	5	1	20,0	5	49	69,0	71
1	63	91,3	69	5	100,0	5	5	100,0	5	5	100,0	5	4	80,0	5	82	92,1	89

As expected, in cases of reviews that does not express a sharp judgment (i.e., overall score 4 and 2) the correctness of the classification performed by TextClass is slightly worse. In particular, in case of overall score = 4 and 2, and considering all the languages employed in the five datasets, TextClass correctly classifies respectively the 74.6% and 69% of the reviews. Also in this cases, TextClass is able, for certain languages, to perform an exact classification (i.e., IT, FR, and DE when overall score = 4, and FR when overall score = 2). Only in the case of the reviews in DE, with score = 4, the TextClass is not able to carry out a good classification.

It is interesting to note that the classification of reviews written in French is always correct (obviously excluding overall score 3), and similar results are achieved for Italian reviews.

Obviously the choice of the threshold is subjective but its value has been selected in order to balance the results on the reviews with overall score = 3 and not for achieving the best possible classification. A deeper investigation will be necessary to understand if 3.4 is a good value for the threshold also for larger data sets involving more languages, or if further tuning is required.

To summarize, with respect to the research question RQ we can say that, using an appropriate threshold that we empirically set to 3.4, TextClass is able to classify correctly the majority of the reviews in all the five considered languages. The preliminary evaluation reported in this paper shows the feasibility of the approach implemented by TextClass, even if further investigations are required to refined and fine-tune both the approach and the tool.

7 Conclusions and Future Work

In this paper we have discussed the design of My MOoD and our first experiments with TextClass. Although My MOoD is not implemented yet, the implementation and the correct functioning of IndianaMAS, which shares with My MOoD the architecture and – to some extent – the purpose, makes us very confident in the possibility to actually build and use My MOoD for multilingual and multimedia sentiment analysis.

With respect to TextClass, our first implementation neglects many aspects that should improve its analysis capability. In particular, we do not deal with negation which is a well known threatening aspect for carrying out a correct analysis of a document's polarity. To make an example, in the first version of the SentiHotel ontology we included timely as a positive feature of the staff. However many negative reviews contained complaints about the time required to obtain some services. These reviews were tagged with timely which was considered as a positive feature, thus resulting into wrong classifications. We plan to add a document pre-processing stage during which negated words and sentences will be recognized and changed into their antonyms. This stage will be of course language-dependent and for this reason requires time and special care.

In our future work we also plan to investigate the reason behind the unbalanced classification obtained when using the standard threshold Tr=3. For instance, this could depend from the kinds of positive/negative words used in the reviews, from the coverage of the positive/negative words by BabelNet or from the inability of TEXTCLASS of managing negations. However, we plan to better validate the effectiveness of the approach implemented by TEXTCLASS through a cross-validation (aka, leave-one-out) procedure. To this aim we will use a leave-one out cross validation with k datasets. We will split the original datasets into $k-1$ datasets used for training the threshold Tr and one dataset used for testing the effectiveness of TEXTCLASS employing such threshold, with the testing dataset rotated so as to test TEXTCLASS on each of the k available datasets.

References

1. K. Ahmad, D. C. D, and Y. Almas. Multi-lingual sentiment analysis of financial news streams. In *Proc. of CAASL2*, pages 1–12. Linguistic Society of America, 2007.
2. M. Almashraee, D. M. Diaz, and R. Unland. Sentiment classification of on-line products based on machine learning techniques and multi-agent systems technologies. In *Proc. of ICDM 2012*, pages 128–136, 2012.
3. A. Balahur, M. Turchi, R. Steinberger, J. M. P. Ortega, G. Jacquet, D. Küçük, V. Zavarella, and A. E. Ghali. Resource creation and evaluation for multilingual sentiment analysis in social media texts. In *Proc. of LREC-2014*, pages 4265–4269. ELRA, 2014.
4. M. Baldoni, C. Baroglio, V. Patti, and P. Rena. From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, 6(1):41–54, 2012.
5. E. Boiy and M. Moens. A machine learning approach to sentiment analysis in multilingual web texts. *Inf. Retr.*, 12(5):526–558, 2009.
6. D. Borth, T. Chen, R. Ji, and S. Chang. Sentibank: large-scale ontology and classifiers for detecting sentiment and emotions in visual content. In *Proc. of ACM MM 2013*, pages 459–460, 2013.
7. D. Borth, R. Ji, T. Chen, T. M. Breuel, and S. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proc. of ACM MM 2013*, pages 223–232, 2013.
8. D. Briola, V. Deufemia, V. Mascardi, L. Paolino, and N. Bianchi. Ontology-driven processing and management of digital rock art objects in indianamas. In *Proc. of EuroMed 2014*, LNCS, pages 217–227. Springer, 2014.
9. G. Casella, V. Deufemia, V. Mascardi, G. Costagliola, and M. Martelli. An agent-based framework for sketched symbol interpretation. *J. Vis. Lang. Comput.*, 19(2):225–257, 2008.
10. M. Chaves and C. Trojahn. Towards a multilingual ontology for ontology-driven content mining in social web sites. In *Proc. of ISWC 2010, Volume I*, 2010.
11. M. S. Chaves, L. A. de Freitas, and R. Vieira. Hontology: A multilingual ontology for the accommodation sector in the tourism industry. In *Proc. of KEOD 2012*, pages 149–154, 2012.
12. K. Denecke. Using sentiwordnet for multilingual sentiment analysis. In *Proc. of ICDE 2008*, pages 507–512, 2008.
13. C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press, 1998.

14. M. Franco-Salvador, F. L. Cruz, J. A. Troyano, and P. Rosso. Cross-domain polarity classification using a knowledge-enhanced meta-classifier. Under review, 2015.
15. C. Gerber, J. H. Siekmann, and G. Vierke. Holonic multi-agent systems. Technical Report DFKI-RR-99-03, Deutsches Forschungszentrum für Künztliche Inteligenz, 1999.
16. H. Gunes and B. Schuller. Categorical and dimensional affect analysis in continuous input: Current trends and future directions. *Image Vision Comput.*, 31(2):120–136, 2013.
17. N. R. Jennings, K. P. Sycara, and M. Wooldridge. A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38, 1998.
18. Z. Kechaou, M. B. Ammar, and A. M. Alimi. A multi-agent based system for sentiment analysis of user-generated content. *IJAIT*, 22(2), 2013.
19. E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades. Ontology-based sentiment analysis of twitter posts. *Expert Syst. Appl.*, 40(10):4065–4074, 2013.
20. B. Liu and L. Zhang. A survey of opinion mining and sentiment analysis. In *Mining Text Data*, pages 415–463. Springer, 2012.
21. V. Mascardi, D. Briola, A. Locoro, V. Deufemia, L. Paolino, N. Bianchi, H. de Lumley, D. Grignani, D. Malafronte, and A. Ricciarelli. A holonic multi-agent system for sketch, image and text interpretation in the rock art domain. *IJICIC*, 10(1):81–99, 2014.
22. R. Mihalcea, C. Banea, and J. Wiebe. Learning multilingual subjective language via cross-lingual projections. In *Proc. of ACL 2007*, pages 976–983. ACL, 2007.
23. G. A. Miller. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41, 1995.
24. R. Navigli and S. P. Ponzetto. Babelnet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artif. Intell.*, 193:217–250, 2012.
25. B. Pang and L. Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, Jan. 2008.
26. B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *CoRR*, cs.CL/0205070, 2002.
27. V. Patti, F. Bertola, and A. Lieto. Arsemotica for arsmeteo.org: Emotion-driven exploration of online art collections. In *Proc. of FLAIRS 2015*, 2015.
28. R. Plutchik. The nature of emotions. *American Scientist*, 89(4), 2001.
29. J. Polpinij and A. K. Ghose. An ontology-based sentiment classification methodology for online consumer reviews. In *Proc. of IEEE/WIC/ACM WI-IAT'08*, pages 518–524, 2008.
30. B. Schuller, A. Batliner, S. Steidl, and D. Seppi. Recognising realistic emotions and affect in speech: State of the art and lessons learnt from the first challenge. *Speech Communication*, 53(9-10):1062–1087, 2011.
31. S. Siersdorfer, E. Minack, F. Deng, and J. S. Hare. Analyzing and predicting sentiment of images on the social web. In *Proc. of MM 2010*, pages 715–718, 2010.
32. C. Strapparava and A. Valitutti. Wordnet affect: an affective extension of wordnet. In *Proc. of LREC-2004*. ELRA, 2004. ACL Anthology Identifier: L04-1208.
33. E. Tromp and M. Pechenizkiy. Senticorr: Multilingual sentiment analysis of personal correspondence. In *Proc. of ICDMW '11*, pages 1247–1250. IEEE Computer Society, 2011.
34. P. D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proc. of ACL 2002*, pages 417–424, 2002.
35. G. Vinodhini and R. M. Chandrasekaran. Sentiment analysis and opinion mining: A survey. *IJARCSSE*, 2(6):282–292, June 2012.
36. H. Wang, Y. Lu, and C. Zhai. Latent aspect rating analysis without aspect keyword supervision. In *Proc. of ACM SIGKDD 2011*, KDD '11, pages 618–626, New York, NY, USA, 2011. ACM.
37. T. A. Wilson. Fine-grained subjectivity and sentiment analysis: Recognizing the intensity, polarity, and attitudes of private states. Doctoral Dissertation, University of Pittsburgh, 2008.
38. Q. You, J. Luo, H. Jin, and J. Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *Proc. of AAAI 2015*, page 10, 2015.
39. J. Yuan, S. Mcdonough, Q. You, and J. Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *Proc. of WISDOM 2013*, page 10, 2013.

Felt Emotions

Cristiano Castelfranchi

ISTC-CNR **TY**-*Theoretical Psychology Group*[1]
cristiano.castelfranchi@istc.cnr.it

Abstract What does it mean "feeling" something? How body activation and its perception is crucial in emotional experience? How it impact on the cognitive components of human emotions and their "appraisal" function, or is affected by them? Which are the different mental paths of emotional experiences?

Keywords: Emotions, Feeling, Mind and Body, Cognitive Architecture, Appraisal

1. Premise and claims

Do we need "true" emotions – not just their simulated expression - in our artificial partners (agents or robots) for a meaningful social interaction? Do (cognitive) agents or robots need "true" emotions for their own preference and decision processes, social empathy or hostility, be sensible to moral norms and social duties? In any case, not "true" emotion without feeling something. No meaningful *model* of emotions can be provided without accounting for the fact that emotions are 'felt', modeling what this means, and how is integrated with the other representational and motivational components in the emotional "architecture".
This is what we will try to do in this paper. Our claims are the following:

- No real emotions without a 'body' (goals, beliefs, stimulus, reactions, actions, ... are not enough);
- No real 'body' without 'feeling', that is, without the body being felt, and sends sensorial signals about its current state;
- The bodies can autonomously, automatically, and primarily react to external stimuli, without a high level (belief-based) evaluation and forecast, and its 'reaction' (*motion*) is perceived and interpreted by the control system. Contrary to Ortony's et al. model [1] this is enough for simple emotions also in humans (like *fright* and *start* due to a loud noise, even before really realizing what it was).
- No human emotions without high-level cognition [2], but in the sense that the bodily response should be received, interpreted and attributed.
- There are <u>different routes</u> to the emotional process; bottom-up and top-down routes, and also their parallel use: while my body reacts to the stimulus (for example in *fright*) I continue to process it and evaluate it at a higher level, eventually converging or diverging with the implicit emotional evaluation ("danger!") and confirming it or blocking it ("What a stupid! It was just the wind!").
- Several of those paths are optional, while some of them are always there; in particular *the elicitation of a 'somatic' response and its ascription to the perceived/conceived event*.

2. Our Model of Emotions

2.1 Aspects of Emotions

Emotion forms a complex, **hybrid** subjective state (*state of mind*[3]).

[1] This research was part of the European Projects Mind RACES - EC's 6th Framework Programme: Cognitive Systems & of HUMAINE network.

[2] Not only in Ortony's sense: always high level cognitive evaluation and a top-down input to the body reaction.

[3] Also including the active representation of the body.

A constituent element of each complex emotion is the "mental state", an integrated structure of assumptions and goals which is closely linked to its functional nature, determines its intensional nature and classification, and accounts for its activation and motivation.

- **Intension** (what emotion is about- see below)
- **Hedonistic valence**: the general thesis is that *the emotions indicating failure* (actual or threatened) *to achieve a goal are subjectively negative, i.e. unpleasant.*
- **Constituent elements of the emotions**
  The basic constituent elements of the emotions are **beliefs, evaluations, goals**, **arousal** and **feeling** - i.e. somatic activation and its proprioception - the "tendency towards action" or **conative** component, and the **expressive** component.
- **Feeling.** By "feeling/to feel" in this context we do not intend the broad family of affects, moods, emotions, drives, … . We mean a crucial *component* of emotions (but not only of emotions):
  a) sensations from the body (proprioception/enteroception) and *about* body states and motions (like "feeling cold" "feeling pain" "feeling shiver");
  b) associated with, or ascribed to (see below: causal attribution), or interpreted as responses/reactions to a given perceived stimulus, or a given endogenous mental representation. In such a way the "feeling" acquires "*intension*", "*aboutness*", and become an emotional experience not just a strange event in our body.

(b)-feeling - which includes (a) - is a necessary component of a real (that is *experienced*) "emotion". An emotion contains (b) that contains (a).

2.2 Emotional Paths

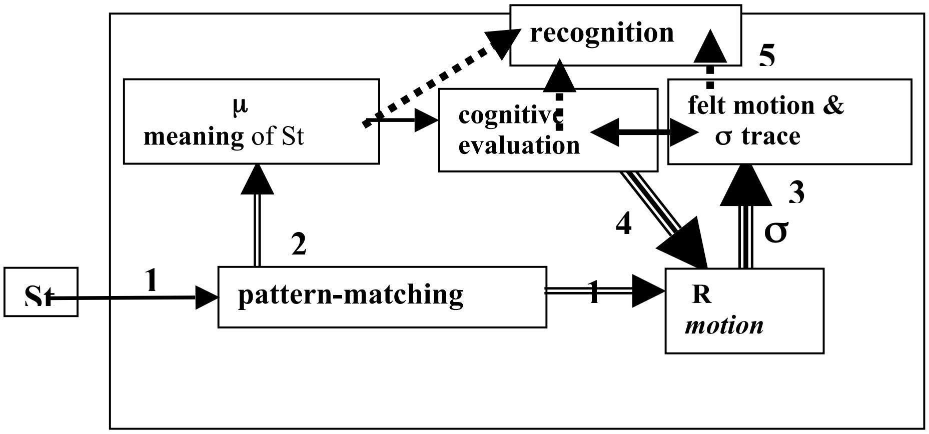


Fig.1 Emotional paths

It is very important not only identifying the links between the various *layers* (like in [1]), thus the *paths* of the possible processes, the flow of the emotional-information-processing, but to make explicit the *semantics* of those 'arrows' (this is a criticism to several other important models ex. [2] [3] [4] [5]), which is very rich and very diverse.

Let's sketch our coarse and incomplete but already complex model (Fig.1), in order to later disentangle different possible *emotional paths* (Figs. 2-3), and different 'functions' of the inter-layers links.

St is the *stimulus*, the perceived event (if any) eliciting the emotional reaction; **R** is the bodily *response* or related activation, i.e. the bodily 'motion'; **m** is the high level *interpretation* of St or an endogenous mental scenario; **σ** is the proprio-entero-ceptive *signal* to the control system ‘informing’ about the visceral-muscular reaction R, that is, *the felt sensations from and about the body motion*. Let's carefully consider those links.

The visceral route (from the ‘sensitive’ to the ‘rational’ soul):

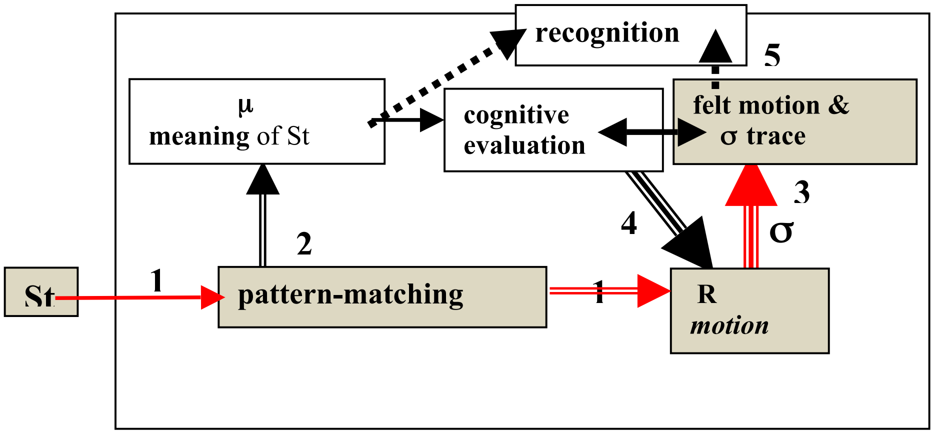


Fig.2 The visceral path

This path (1+1+3) characterizes what one could call the **‘visceral route’**; that is, an emotional automatic (stimulus-response) somatic reaction **R** to a perceived event **St**, plus:

- the sensation/signal **σ** of this bodily motion,
- its *memorization* and *association* to the stimulus (formation of the ‘somatic marker’), and
- the *implicit evaluation* of St in terms of the pleasant/unpleasant quality of **σ**.

In fact we assume that a positive or negative emotional response *does not presupposes an evaluation* of the event but **is** such an evaluation ([6] [7] [8]) This is similar to the "appraisal" of the stimulus postulated by many theories of emotions although in an ambiguous way; never clearly disentangling ‘explicit cognitive evaluation’ (beliefs, expectations, etc.) from implicit appraisal, positive or negative sensations and feelings about.

Moreover, the felt response can be *memorized* and *associated* with the stimulus. In such a way, also the he automatic activation from memory of this associated internal response (*evocation*) (in Damasio's terms, a "somatic marker"; [9]) *is* its appraisal.

The associated negative or positive emotion makes the situation bad or good, unpleasant or pleasant, and we dislike or we like it.

> *“Appraisal” consists of an automatic association (conscious or unconscious) of an internal affective response/state either pleasant or unpleasant, attractive or repulsive, etc., to the appraised stimulus or representation.*

It does not consists in a *judgment* of appropriateness or capability - possibly supported by additional justifications; on the contrary, it just consists in a subjective positive or negative experience/feeling associated with the stimulus or to the mental representation, usually previously conditioned to it in similar circumstances, and now retrieved. We consider these sub-symbolic, implicit forms of "evaluation" as evolutionary forerunners of cognitive evaluations.[4]

[4] We in fact distinguish between "appraisal" - that should be the *unconscious* or *automatic*, implicit, an *intuitive* orientation towards what is good an what is bad for the organism- and "evaluation": the cognitive judgments relative to what is good or bad for p (and why). We define an evaluation of an entity *x* as *a belief of an evaluating agent e about*

The top-down route (from 'rational' to 'sensitive' soul)

The path in Fig. 3 (1+2+4+3(+5)) characterizes what one could call the **'cognitive-appraisal based route'**; that is an emotional reaction to the evaluation of a perceived event St, plus the sensation **s** of this bodily motion, its memorization and association to the stimulus (formation of the 'somatic marker'), and implicit evaluation of St/**m** in terms of the pleasant/unpleasant quality of s. Consider that "Cognitive Evaluation" is a really simplified module: we put together two well-known kinds of appraisal: the evaluation of the event ("What happened? How to attribute it? Which predictions?") and the **'coping' evaluation**: how to deal with that event.

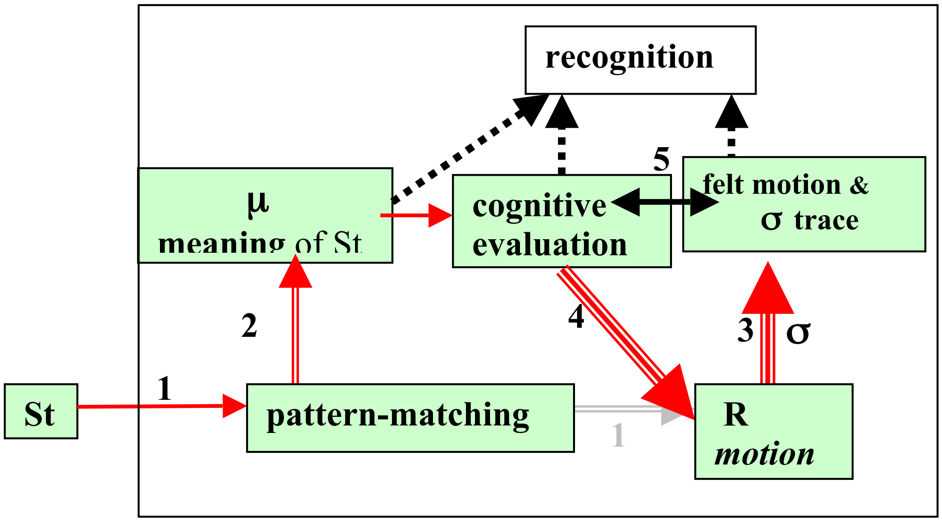


Fig. 3 Cognitive evaluation and body reaction

As we said the emotional reaction can just be due to a though or internally generated scenario or prospect (*imagination*), not to an external stimulus: route 1-1, 1-2 may not be there at all. Let's now provide the 'semantics' of the arrows in Figs. 1-3.

3. Semantics

Arrow 3 (*James' arrow*; the bottom-up one) is the most important passage in order to make what happens to and in our body, its *reaction* (to a stimulus (*Arrow 1*) or to a thought (*Arrow 4*)) a real 'emotion'. This represents in fact several things:

3i) ***Feeling***; the 'perception' of what is happening: sensation.

To feel consists in this path:

a) The run time *monitoring* of the internal (body) environment and of its dynamics (in response to external or internal stimuli) in order the brain and also the mind (beliefs, decisions, etc.) respond to it. Our control and government system (both as brain and as mind) is not only for dealing with the external environment chances, but also for managing internal events in an adaptive way. We both have a 'Foreign-office' and a 'Ministry of Interior'. Actions on and in the internal environment are usually not 'intentional' actions. More frequently they are unconditioned or conditioned responses, or proto-intentional actions.

x's usefulness with regard to a goal p. Evaluations are a special kind of *beliefs*, characterized by a strict relationship with action, by virtue of their link with *goals*. Evaluations not only imply goals, but also can *generate* them [5] [6].

b) However, 'feeling' does not necessarily and always implies an *actual bodily activation*. This can be just the *origin* of the 'sensation'. Feeling can simply be the retrieval of a memory trace of a previously felt body-experience associated (conditioned) with a given situation, mental representation or action: *evocation* of **σ** (*somatic* marker).

3ii) ***Attribution***; the mental attribution of the bodily reaction and sensation to a given event or mental content as the cause and releaser of it. [5] Without such a causal attribution or perceived link (between mental content and felt bodily reaction) (*arrow 5*) it seems that we do not experience real 'emotions' but just bodily alterations. The signals from the body must be interpreted and associated with a significant event (Schacter-Singer).

3iii) ***Recognition***; arrow 3 contributes to the complete bottom-up process that ends with the possible 'categorization' of the perceived state as a given 'emotion' on the basis of different visceral, postural, expressive sensations, of the St, and/or the associated specific thought contents. [6] Human beings feel *specific* emotions also thanks to their cultural categories and their learning to 'recognize/categorize' them by cognitively discriminating one from the other on the basis of content, context, and sensations. Thus to feel a specific emotion also means to be aware of it, to recognize what is happening to you. People can be confuse about their perceived emotions; they can feel depressed when they are just tired; but subjectively speaking they are depressed since they interpret their sensations in such a way and with some pertinent related content. As any other recognition / categorization process this is a 'constructive' activity.

The link emotions-feeling seems to mean that emotions are useful for **evaluating** and acting with priorities in the external world but on the basis of and in relation to the value of the events for our body, its internal life, and its long term ***interests*** (more than actual goals – [4]) like 'integrity'. Our body seems some sort of physical 'memory' of adaptive functions and successful or unsuccessful experiences. The body reacts with an alarm when the event threats or concerns some primary adaptive function of the animal. If that 'goal' is achieved or close to the achievement the feeling is pleasant (happiness, joy, satisfaction, etc.), if it is damaged or treated the feeling is unpleasant (ex. boredom, anxiety, fear, guilt, shame, ..)

Arrow 4 (the top-down) represents three different things:

4i) The **top-down somatic activation due to a mental content** (a mental imagery, a though, an inference and prediction, a memory, an evaluation, etc.)

The body reacts not to an external stimulus (that can be completely lacking) but to its mental interpretation and evaluation, or to an endogenously produced mental representation. In the last case there is no route on *arrow 2* and **m** (the endogenous mental representation) is the starting point.

In Ortony, Norman and Rivelle model (like in all the strongly 'cognitive appraisal' based models) arrow 4 is the necessary path for having a true human emotion: a bodily reaction to a stimulus is not enough; it should be elicited by a mental evaluation of it. We do agree that a mere 'visceral' bottom line path is not enough for a full emotion, but we disagree about the necessity of *arrow 4*; for us also *arrow 1 + arrow 3* is enough: a reactive bodily response to something just due to a primitive 'pattern recognition', or 'releaser' without any complex mental evaluation or interpretation or thinking, but this reaction is 'felt', attributed, and recognized at the central cognitive level. This happens for example in *frights*, *starts* of fear, strong *disorientation,* and is enough for a true emotion.

4ii) The **feedback** or **loop** on the body of a previous bottom-up flow (Arrow 3), i.e. of the subjective 'interpretation' of a felt bodily response.[7] *Arrows 3 and 4* can determine **a recursive loop**: this also holds in visceral emotion but requires this level of cognitive appraisal of the somatic input (thanks to *arrow 5*) (what in RET and Cognitive therapy they call the 'secondary evaluation/reaction'. This is for example the very well known case of panic crises due to the subject's interpretation of heart acceleration and his reaction to this interpretation and worries, and again and again (cit.). Figure 4.

[5] This kind of 'belief' is important in human emotion also for accounting for Schacter's emotional or non-emotional quality of bodily modifications.

[6] See our claim about this kind of 'belief' in human emotion in order to account for cultural differences, individual competence, and for more or less sophisticated and discriminative emotional systems (their might be culture with one, two or three species of hostile disposition and feelings, while Italian culture distinguishes between at least 15 different hostility affects).

[7] Probably it would be better to have two arrows 4. One for the (positive or negative) feedback loop from path 3; one for the impact on the body after 2.

4iii) There might be another meaning of this arrow, and more precisely of its *loop function*. The evaluation of the bodily activation (motion) and stimulus **s** and the reaction to it might be not that of fostering body motion, but – on the contrary – its inhibition: the attempt to remain cold and quite, and maintaining the 'self-control'.

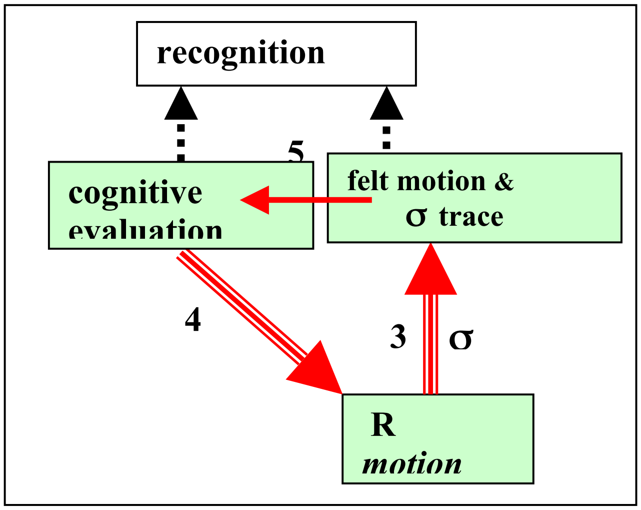


Fig. 4 A possible vicious circle

4. To feel with an unaffected body

As we said 'feeling' does not necessarily and always implies *an actual bodily activation*. It is possible to 'feel' something also in absence of a current specific signal/sensation from the body about some bodily reaction to an event or a thought. In this case a given thought or mental representation **m**, or a given stimulus, just activates the *central trace* of a previously associated (conditioned) affective/somatic response ('somatic marker') **s;** that is, they *evocate* previously memorized sensory experiences. We call this kind of feeling: "evocation-based feeling"; while the other involving actual sensations and signals from the body is called "bodily-activation-based feeling". The main merit of the (not so clear) Damasio's notion is precisely the clarification of the fact that a 'somatic' marking, that is the activation of the 'somatic marker' does not require the activation of the body, a current somatic signal; the activation of the *central memory trace* of that somatic response is enough. The other merit is the idea that one can have classical conditioning mechanisms not just on mere stimuli and responses, but upon mental scenarios and high-level representations used for/in reasoning and decision making.
This is very important for our model. It allows us to claim that:

- Always, when a subject says "I feel that" "I feel so…" she is really feeling something, that is she has some sensation, some 'somatic' signal, but not necessarily from her activated body; what she senses is the recalling of previous sensations and affective reactions: they are 'evoked', 'imagined', 'simulated' sensations.

We believe that for a good and general theory of 'feeling' this should be generalized also to cases like "I (do not) feel safe", "It is just a sensation. I do not feel confident", "I feel able to …", "I feel that everything is going worst", etc., that mainly are "evocation-based feelings" due to the unconscious re-evocation of previous affective or sensory-motor experiences.

5. On James' arrow again: implementing "affectus" and how feeling *affects* a reason-based mind

Let us now focus other very important meanings of James' arrow (arrow 3): its 'informative', epistemic-value function and its 'motivational' (conative-value) function.

For us this is the most important aspect of the notion of "affectus" (Spinoza), that is, how body "affects" mind. There are –it is true – several important impacts of bodily motion and signals on cognitive processes in strict sense: how mood affects memory retrieval; how attention is modified by emotional reactions; how emotional states favor some heuristics or others, or a given framing; how they can cut -or expand- decision time, or shortcut at all any decision process. However, in our view, "affectus" is the most significant "intrusion" of the perceived bodily activation within the intimate architecture of a reason-based (not necessarily 'rational') mind, by introducing a radically heterogeneous criterion within that symbolic "computation". The felt sensation from the body affects both the epistemic and the motivational-decisional aspects of cognitive processing [10].

5.1 The felt certainty of beliefs

The first point is known in the literature as *ex-consequentia-reasoning, affect as information* [11]. The idea is that the perceived bodily activation is used as evidence on which a belief *about the world* (the event) (not about the body or the mind) is based. For example we believe that a given situation is 'dangerous' just because we feel fear; or we believe that a person is sexy and perceive her/him as exciting (just) because we are sexually excited. Let us consider the classical example of the emotion of fear providing bases and evidence for the idea that there is some danger around. Arrows 3 is altering the normal cognitive process that ground beliefs and their credibility. The *credibility* of a given belief and its assumption depends on other beliefs that support it, and of its sources: (i) direct perception of the fact ("I saw it"); (ii) social communication ("They say that …"); (iii) inference and reasoning ("I conclude that…").

- *The many the different converging sources and supports, the stronger the belief (its 'certainty' or credibility);*
- *The more reliable/trusted the source the more credible the belief.*

These are the two principles founding beliefs formation and their strength. Now Arrow 3 affects this process and introduces a completely heterogeneous and independent principle:

- *The more intense the felt sensation (the motion) the greater the subjective certainty of the belief.* We have both possible schemes (Figure 5):
  - A belief, based on usual 'evidences' and 'sources' (direct perception, inference, communication) ("It is frozen! There is danger! I should be careful") activating an emotional response of fear or worry; and then a feedback of the felt motion (bodily response to this idea) on the belief 'credibility' and 'evidence' (low part of Figure 5); but also:
  - A belief just follower of and derived from a mere affective experience (motion) and *implicit appraisal* just automatically aroused by some low level stimulus (Figure 5 but also the path depicted in Figure 2).

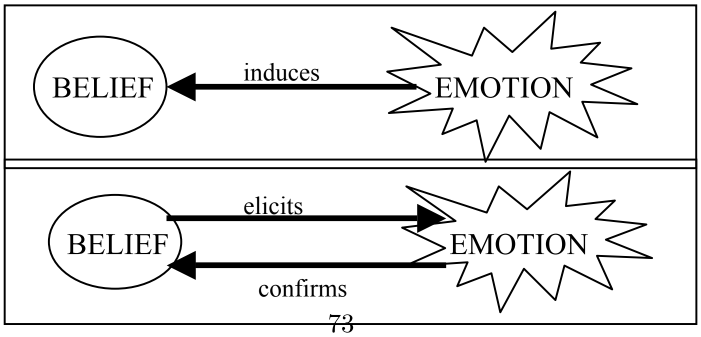


Fig. 5 Affect as Information

The wisdom of the body and Pascal's heart. This is a "non-rational" principle, more precisely a not "reason-based" principle for believing. One cannot justify (except in fact *post hoc* and in a self-deceptive way), cannot account for or support this belief. This – obviously - does not mean that it is necessarily non adaptive or advantageous; or that there are no other implicit and unknown "reasons" for believing so. As Pascal claimed "The heart has its own *reasons* that Reason cannot understand". That is, our implicit embodied "memory" of previous events and their value might provide a correct, automatic, unconscious, primary "evaluation" of the new event; even when we are not able to retrieve those previous affective experience that shape our current (not arguable) reaction [12]. This affective *pre*judice can be non adaptive when the association has been arbitrary, or the analogy is not well grounded and just superficial, when the response is too generalized, when the learning has been based on very few strong experiences, etc.

A new kind of "implicit beliefs": when "feeling that…" implicitly is "believing that..".

There is a family of implicit beliefs (in the sense of 'potential') that are *feeling-based*; actually consists in a feeling state (a mood about..), which can be also interpreted at a declarative symbolic level and can bring to an explicit belief.

As we saw, a given mood-about, feeling, can either confirm and support a given belief (eliciting it) or also can *produce* its related belief. In this case, when the feeling is active we might say that the related belief is already *implicit*; it is presupposed or entailed by that feeling but is not explicitly formulated – like in Figure 5, or not activated, although preexisting; only the feeling is active. For example, a sense [8] of impotence, of not being able; or a feeling of non-safety or threat, can *anticipate* its explicit beliefs and hold before and without its formulation. This phenomenon in some sense is analogous to the other theory of *implicit/potential* beliefs. When X believes/knows that p, and in fact p implies q, X implicitly and potentially also believes/knows that q; but not actually. For example, X cannot detect possible contradictions with other beliefs. This is because psychologically speaking we do not know all the logical consequences of what we know; we have to actually *derive* them, to 'write' them in some memory or data-base for effectively believing them

Analogously, a feeling "that p", a mood "about" something, implies the belief that p, although this might just be a 'candidate' belief, not so strong and certain to be *accepted* a belief:

(Feel-that X p) ➔ (Bel X p)[9]

Obviously the reverse relation is not true: (Bel X p) NOT implies (Feel X p).

5.2 The felt importance of goals

A very similar revolution is introduced by the felt bodily reaction or activation in the goal reasons-based processing. The *value* or *importance* of a goal, its motivating force, is normally derived from the means-end relations and reasoning: that is, from some consideration of plausible pros and cons.

- Given (in a given moment for a given person) the subjective value of the final aims or motives of the person, and of his active goals;

[8] Dictionary defines one meaning od "sense" as: *a feeling derived from multiple or subtle sense impressions.*

[9] The feeling that p implies the belief in two sense: either (Feel X p) Contains (Bel X p), like for example a "fear" mental state contains a prediction; or like the idea of "to kill" contains the idea of "to die"; or (Feel X p) potentially produces (Bel X p)

- given what she believes and takes into account about the possible effects and means-end relations of actions and sub-goals;
- the value of a given action A or sub-goal SG *is the sum of the values of all the foreseen positive consequences (realized aims) less the values of all the compromised and renounced aims (costs).*

This 'calculation' is based on *reasons* (beliefs about effects) and is arguable and questionable ("Did you consider this possible danger?"). Now, Arrow 3 as "affectus" again subverts this schema, introducing a completely heterogeneous and independent principle of goal value:

- *The more intense the felt sensation (the motion) the greater the subjective value of the goal, its priority.*

In the theory of 'felt needs' for example we explained in such a way why *needs are particularly "pushing" motives (*compared with 'desires' 'intentions' 'wishes' etc.) [13].

- First, they are conceived as <u>necessary</u> for the aim, not as useful but optional;
- Second, they are <u>conceived (framed) in negative terms</u>, in terms of losses rather than gains (if you don't have... you lose, you will not..) and we know that the avoidance of damages is more influencing than the perspective of gains (*prospect theory*);
- Third, they are related to some <u>pain</u> or disturbance, to a negative felt bodily sensation **s**, <u>which must be stopped or avoided</u>;
- Fourth, mental representation with <u>sensory motors components</u> have a stronger impact than very abstract, merely conceptual representations [14].

- *The stronger the pain (and the persuasion/belief that it depends on the lack of O and will stop taking O) the stronger the goal of having O.*

This goal is like an "impulse" since is sensation and affect driven. This is in fact general for the goals activated by an emotion, that we call "impulses":

- *The stronger the bodily-sensation (feeling) the stronger (important, cohercive, urgent) the impulse.*

Also in this case a non-rational mechanism replaces or alters the reason-based one. Feeling – through James' arrow – affects mind and replaces belief-value calculation (*credibility*) and goal-value calculation (*importance*).

6. Final Remarks

Those are some of the main functions of emotions. Without modeling the feeling component, that is the felt somatic response and activation, we cannot really understand and model emotions. This case is also instructive for understanding *what does it means to 'reincorporate' mind*. It does not means to eliminate cognitive 'mental representations' like *goals* (intentions, desires, projects, …) or *beliefs* (assumptions, evaluations, expectations, …); it means to understand the specific links between them and the body, in terms of both

- 'grounding' conceptual representations in sensory-motor intelligence (the "embodiment" approach in Cognitive Science), and
- relating cognitive processing (like believing, preferring) with the 'experience' of the body and of its reactions and felt internal states.

We have to "embody" mental representations and processes, but also to "mentalize" the body. If we want to build "real" emotions in artificial creatures (not just their imitation and depiction) we have to work on this kind of "architecture", by integrating body and mind, cognition and sensations.

We didn't answered in this paper to our initial questions: *Do we need "true" emotions – not just their simulated expression - in our artificial partners (agents or robots) for a meaningful social interaction? Do (cognitive) agents or robots need "true" emotions for their own preference and decision processes, social empathy or hostility, being sensible to moral norms and social duties?* However, this in fact was not our aim.

Our claim was that: if one *would* intend to model real emotional process she should (also) take into account, modeling, and reproduce the somatic processes of "feeling" something: "true" emotion are "felt".

References[10]

[1] Ortony, A., Norman, D., Revelle, W. (2004) Affect and Proto-affect in Effective Functioning. In J.M. Fellous & M.A. Arbid (Eds.) *Who Needs Emotions? The Brain Meets the Machine*. N.Y. Oxford University Press.

[2] Scherer, K. (1984) On the nature and function of emotion: a component process approach. In K.R. Scherer & P. Ekman (Eds.) *Approaches to emotion*. Hillsdale, N.J., Erlbaum.

[3] P. Petta, C. Pelachaud, R. Cowie (eds.) (2011) *Emotion-oriented systems. The HUMAINE handbook* (Cognitive Technologies). Berlin: Springer-Verlag, 2011

[4] Sloman, A. & Logan, B. (2000) Evolvable architectures for human-like minds. In G. Hatano, N. Okada & H. Tanable (Eds.) *Affective minds*. Amsterdam: Elsevier.

[5] Bach J. (2009) *Principles of Synthetic Intelligence. PSI: An Architecture of Motivated Cognition*. Oxford University Press, 2009

[6] Castelfranchi C (2000) Affective appraisal vs cognitive evaluation in social emotions and interactions. In: Paiva A (eds) Affective interactions. Towards a new generation of computer interfaces. Springer, Berlin Heidelberg, New York, pp 76–106

[7] Castelfranchi, C., Miceli, M. (2009) The Cognitive-Motivational Compound of Emotional Experience. *Emotion Review*, 1, 3, 221-228.

[8] Fiorella de Rosis , Cristiano Castelfranchi, Peter Goldie and Valeria Carofiglio (2011). Cognitive Evaluations And Intuitive Appraisals: Can Emotion Models Handle Them Both? in P. Petta, C. Pelachaud, R. Cowie (eds.). *Emotion-oriented systems*. The HUMAINE handbook (Cognitive Technologies). Berlin: Springer-Verlag, 2011.

[9] Damasio, A. (1994) *Descartes' error*. Papermac.

[10] Castelfranchi, C., Giardini, F., Marzo, M. (2006). Relationships between rationality, human motives, and emotions. *Mind & Society, 5,* 173-197.

[11] Schwarz, N. (2010) Feelings-as-Information Theory. In P. Van Lange, A. Kruglanski, & E. T. Higgins (eds.), "Handbook of theories of social psychology". Sage.

[12] Bargh J.A. & Chartrand, T.L. (1999) The unbearable Automaticity of being. *American Psychologist*, 54, 462-479.

[13] Castelfranchi, C. (1998) To believe and to feel: The case of "needs". In D. Canamero (ed.) *Proceedings of AAAI Fall Symposium "Emotional and Intelligent: The Tangled Knot of Cognition"* 1998, AAAI Press, 55-60.

[14] Miceli, M. e Castelfranchi C. (1997). Basic principles of psychic suffering: A prelimirary account. *Theory & Psychology*, *7*, 769-798.

[10] We don't have enough room for a real bibliography; just main references.

Towards Formal Modeling of Affective Agents in a BDI Architecture

Bexy Alfonso, Emilio Vivancos, and Vicente Botti

Universitat Politècnica de València,
Departamento de Sistemas Informáticos y Computación, Spain
{balfonso,vivancos,vbotti}@dsic.upv.es

Abstract. When designing agents to simulate human behavior, the incorporation of personality, emotions and mood into the agent reasoning process provides the agent with a closer to human behavior. We have designed an Open Affective Agent Architecture (O3A) based on widely accepted psychological theories. O3A offers a flexible way of integrating the affective characteristics of agents into their logic. We extend the operational semantic of the BDI agent language AgentSpeak modifying the traditional BDI reasoning cycle to incorporate affective components. This informal semantic description allows an agent to have a representation of not only the current state of the environment, but also of the agent affective state.

Keywords: Agents, Emotions, Semantic, Formalization

1 Introduction

Researches on multi-agent systems have traditionally focused on the search of rational solutions that maximize the quality or utility of the result. However, when an agent needs to simulate humans behavior, this kind of approaches is not the most appropriate. Human decisions are influenced, at greater or lesser extent, by affective characteristics such as the personality, the emotions or the current mood of the individual. In environments where agents must act like humans, the incorporation of emotions into the agent-based reasoning process provides the agent with a closer to human behavior. Many representations and formalizations of affective agents are based on the cognitive perspective of emotions. These formalizations model the appraisal process, the emotions dynamics, or the influence of affective traits on the cognitive processes. However a global formalization of the interrelation between the cognitive and the affective reasoning process is still required.

We have designed O3A, an Open Affective Agent Architecture, which is based on widely accepted psychological and neurological theories. An informal presentation of the main components of O3A can be found in [2]. O3A is built over a traditional BDI architecture and offers components to represent affective traits like personality, emotions and mood. The interaction of the architecture components with the cognitive processes of the agent, produces behavior biased by the agent mood. Our final aim is to to offer a feasible and comprehensive way of building affective agents using the O3A features. We also extend the reasoning cycle of AgentSpeak [4, 24], with new steps to facilitate emotional-based reasoning. The new components that we are proposing are flexible enough to be adapted to any particular requirement of the agent emotional processing.

2 Background

Several authors have proposed mechanisms to incorporate affective components into intelligent agents [3, 8]. For example in [8] Dias *et al.* propose FAtiMA, a BDI architecture that uses emotions and personality to influence agents' behavior. In contrast to O3A, FAtiMA does not have an explicit model to address personality. The agent's personality is implicitly represented in different agent processes and internal structures. In O3A the setting and fitting of the agent personality can be performed in a centralized way according to the widely accepted model of personality FFM (Five Factor Model) [11]. A more detailed comparisons of O3A with other similar approaches can be found in [2].

On the other hand, some works have proposed to incorporate affective traits into agent-based systems in a formal way. Some leading works [16, 19] are considered the base for further approaches that have modeled and formalized the emotion logics. Oatley & Jenkins' model of emotions [16] has inspired works like [22], and [23]. These works extend the KARO framework [14, 15] in order to use this formalism to give a logical account of emotional agents. In [14] Meyer models the dynamics of emotions and the influence of emotions on how an agent deals with its goals and plans. This paper distinguishes four emotions: happiness, sadness, anger and fear. These four emotions are also analyzed in [22], where the authors formally specify the emotion influence on behavior following the OCC model [17]. In [21] Steunebrink et al. make a qualitative formalization of the OCC model offering a method for calculating emotions intensities. Other KARO based model is the one proposed in [23]. Steunebrink *et al.* formalize the triggers conditions for some of the OCC emotions in BDI agents with achievement goals, specifically the appraisal part of OCC.

Rao and Georgeff's BDI_{CTL} logic [19] has been the starting point for some works like [18]. Pereira *et al.* present an improved version of their $EBDI$ logic to model the role of fear, anxiety, and self-confidence in a emotional BDI agent.

Other authors have focused on modeling the eliciting conditions for a subset of emotions, or on the emotions influence on the agent cognitive process. Our aim is to reach a formalization of affective agents at a higher level, offering a flexible approach where the processing of the emotions can be easily adapted to the specific emotional characteristics of the problem to be solved. Therefore, the formalization of our architecture uses general components in order to integrate the affective components with the BDI agent cognitive process. Mood represents the agent emotional state at every moment, and this representation is based on a dimensional theory. The agent has also a personality and emotions that are automatically generated as a result of the agent internal reasoning process and the agent interaction with the environment. In this article we present an ongoing work to extend the operational semantics used in Jason [4, 24], a well know agent-oriented programming language grounded in a logical computable language (AgentSpeak)[4].

3 Extension of the Jason operational semantics

3.1 Considerations to formalize the O3A architecture

Psychological and neurological theories try to explain the influence of emotions in human decisions. Our proposal considers two kind of emotions: primary and secondary. Primary emotions are "infant like" fast reactions easily deductible, while secondary emotions are the result of a more complex reasoning based on expectations or previous experiences [6, 17]. The O3A architecture proposes a set of new components to be included into a traditional BDI architecture with the ultimate goal of offering a computational model of these psychological concepts.

O3A is based on some of the most relevant theories of emotions and personality [5, 6, 17, 20]. These theories helped to build a formal specification of O3A from a cognitive perspective based on the appraisal theory, where the emotional state of an agent responds to a dimensional approach of the mood.

O3A uses *Primary and Secondary Emotions*. Lets start defining how these primary and secondary emotions are derived. We assume that percepts from the environment are labeled with the most common reactive emotions that individuals can experience. This assumption is based on the idea that an event often cause similar emotions in different individuals. For example, when facing a hurricane, people generally will feel fear. The *emotion reactive component* of O3A is responsible of deriving these primary emotions from percepts. On the other hand, secondary emotions are the result of a more complex and internal reasoning. In our approach, secondary emotions can emerge when events of any nature (internal or external) appear. The *emotion deliberative component* of O3A considers some variables: desirability, likelihood, expectedness, causal attribution, and controllability[1][10]. These variables produce an appraisal pattern that is used to derive secondary emotions. While primary emotions are reactive and fast responses, secondary emotions are the result of a "though process" and a more complex reasoning process that considers expectation and experience [6]. We are currently considering that both primary and secondary emotions belong to the same set of possible emotions: the OCC model, but a more detailed study is required to select the set of primary emotions and the set of secondary emotions to be considered for any specific domain of application.

In our approach the mood is represented in a three-dimensional space where three values describe the agent mood: Pleasure, Arousal and Dominance (PAD) [13]. We map each emotion of the OCC model to three values representing a point in the PAD space according to [9]. These three-dimensional points can be synthesized in a tuple which represents the mood. In our approach this 'synthesis'

[1] A more detailed explanation of their meaning can be found in [1]

is made by averaging the values of all the points corresponding to the appraised emotions. Then the calculated mood and the previous mood are merged following the proposal of [9]. When the calculated mood is similar to the previous one, the intensity of the new mood will be increased. Although this change doesn't depend on other processes, we assume that the mood update will occur in each reasoning cycle of the agent. We also propose a function that modifies the value of the mood according to the current mood and the agent personality. The personality is represented in O3A using the Five Factor Model [11] which describes quite accurately individual traits through five dimensions: openness, conscientiousness, extraversion, agreeableness, and neuroticism. The initial mood of an agent is calculated using the personality. The mapping from the agent five dimensions of personality to the three dimensions of the PAD space is done according to Mehrabian's work [12].

In O3A the calculated mood influences the agent behavior. Mood helps to prioritize the agent intentions to affront a given internal or external change. According to [7] the trait Dominance of the PAD representation is a good indicator of how much risk can be taken in order to achieve a goal. If an O3A agent has more than one option to respond to an event, the option selected will be that whose value of risk[2] is closer to the value of the agent dominance.

O3A has a function to determine *when and how beliefs are affected by the agent current mood*. As a result of that function, a belief will be evaluated depending on the current agent mood and knowledge. From this evaluation the belief can be considered for example positive or negative. This evaluation is implicit when percepts are labeled with the most common emotions categories experienced after this percept. For instance, the emotion "joy" could indicate a "positive" percept according to agent interests. Nevertheless, if the percept has not emotional category labels, the current state of the agent and the mood will be used to assign primary emotions to this percept. For example the percept *human_shadow* normally wouldn't produce any specific reaction, nevertheless, if the agent is very scared, this percept may produce the primary emotion "fear", and consequently a reaction to this fear. In [1] we show examples of O3A agents where emotions are used to improve the solution to some classical problems in behavioral economics.

3.2 Extension of the BDI reasoning cycle

In order to offer an integral description of a emotional BDI agent, we have extended the operational semantic of Jason [4, 24] with affective traits. The agent configuration is defined by a tuple $\langle ag, C, M, T, Mo, P, s \rangle$ where ag is the agent program, which contains a set of beliefs (bs) and a set of plans (ps). C, M, T, and s represent respectively the agent circumstance, communication parameters, temporary information for a reasoning cycle, and the label of the current step in the reasoning cycle . In order to include the affective state in the agent configuration, two elements have been added to the agent configuration tuple:

[2] The risk value is a property of plans and it is set by the programmer.

the agent current mood (Mo) and the agent personality (P). In each reasoning cycle the agent current mood can be modified. Mood (Mo) is defined by a tuple $\langle mP, mA, mD \rangle$ according to the dimensional theory of A. Mehrabian [13]. We represent the agent personality (P) using a tuple $\langle pO, pC, pE, pA, pN \rangle$, following the Five Factor Model of personality [11]. The initial agent mood is determined by the agent personality (following the mapping offered in [12]). The personality is also used to define a "equilibrium mood" of the agent.

Primary and secondary emotions change in each cycle of the BDI algorithm, so we have added two new components to the tuple that represents the temporary information (T). The temporary information (T) is represented by the tuple $\langle R, Ap, \iota, \varepsilon, \rho, PEM, SEM \rangle$, where R and Ap are the sets of relevant and applicable plans respectively[3]. ι, ε, and ρ are used to record a particular intention, event, and applicable plan. PEM and SEM represent primary and secondary emotions. Both sets of emotions can contain any emotion defined in the OCC model. We have also added to the O3A architecture the "surprise" emotion that can be derived as a primary or secondary emotion.

O3A has added new steps to the Jason reasoning cycle and thus the corresponding transitions rules. The resulting reasoning cycle is shown in figure 1. New steps are colored while new and modified transitions are represented by arrows with dashed lines. Although the Jason formalization considers that an agent can perceive new information from the environment, to the best of our knowledge, there is no explicit step in the reasoning cycle for this task. We have decided to make this step explicit as a initial step (`Perceive`). The next step is `DerivePEM` step, in charge of deriving primary emotions. It is also a task of the `DerivePEM` step to endow beliefs (derived from percepts) of reactive emotions associated to the agent emotional and cognitive state (the function of the *Beliefs component* of the O3A architecture). The next two steps are `DeriveSEM`, in charge of deriving secondary emotions, and `UpMood` which updates the mood based on the new appraised emotions. `DeriveSEM` follows `RelPl` (which determines the relevant plans), since the derivation of secondary emotions considers the relevant plans for a triggering event. The current mood is updated using the new primary and secondary emotions. If no new emotions are apprised, the O3A reasoning cycle goes on with the `ApplPl` step, which determines the applicable plans. In O3A the agent current mood is also used to select the next applicable plan affecting the `SelAppl` step. Finally the `MoodDecay` step determines the mood tendency to return to its equilibrium state.

In the operational semantic of Jason [4, 24] the elements denoted by T_R, T_{Ap}, T_ι, T_ε, and T_ρ are used to represent the current set of relevant plans T_R, the current set of applicable plans T_{Ap}, the current intention T_ι, event T_ε, and applicable plan T_ρ. We have added to this notation the T_{PEM} and T_{SEM} elements which indicates the PEM (primary emotions) and SEM (secundary emotions) components of the of temporary information tuple (T).

[3] In Jason the relevant plans are those that are candidate to be executed as a consequence of the activation of an event, while applicable plans are those relevant plans whose condition regarding the state of the world is satisfied.

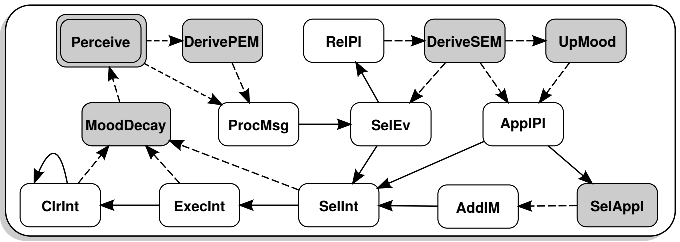


Fig. 1. Extension of the reasoning cycle of AgentSpeak

We have also defined new functions that are used by the agent interpreter on some steps of the reasoning cycle. These functions can be customized by the agent programmer: $IniMood\,(P)$ that is used to determine the agent "equilibrium mood"; $SEMDer\,(ag_{bs}, T_{\varepsilon}, T_R)$ is used to derive secondary emotions; $UpMo(Mo, T_{PEM}, T_{SEM})$ used to update the current mood; $MoodDc(Mo, P, \delta)$ determines how mood will decay in each reasoning cycle; $NewP(PercSet, ag_{bs})$ determines new percepts; and $RemP(PercSet, ag_{bs})$, which determines what percepts are not longer detected in the environment.

3.3 Transitions between the steps of the O3A reasoning cycle

In this section we present the description of the new and updated steps added to the Jason operational semantics. Note that, as described in the previous section, the initial state of the O3A reasoning cycle is denoted by the tuple $\langle ag, C, M, T, Mo, P, \texttt{Perceive}\rangle$. The mood Mo has a initial value (and a "equilibrium mood") determined by the function $IniMood\,(P)$, that rerceives the agent personality as a parameter. To the best of our knowledge the operational semantics for the perception process has not been defined yet, so we carefully provide a transition rule to this end.

Perception: This is the initial step (**Perceive** in Fig. 1). The agent checks the environment and updates its belief base. The function $NewP(PercSet, ag_{bs})$ is used to add the new perceived beliefs. When removing those percepts that are not longer detected in the environment, O3A uses the function $RemP(PercSet, ag_{bs})$. These functions start from a set of percepts ($PercSet$) that are detected in the environment where the agent is situated. In the "perceptions update rule", if no new percepts are acquired, then the reasoning cycle goes directly to the `ProcMsg` step without deriving primary emotions.

Derivation of primary emotions (DerivePEM): Primary emotions are the result of reactive emotional responses to the agent perceptions. These reactive responses are similar for most individuals, and therefore O3A assumes that percepts are labeled with the most common categories for reactive emotions that people usually experience. The structure as well as what should be these labels is up to the programmer. Primary emotions can also be inferred given the percepts' nature and the current mood of the agent. The function $PEMDer\,(ag_{bs}, Mo)$

performs the task of deriving primary emotions. In this O3A first approach only explicit labels for emotion categories are considered.

Relevant Plans (RelPl): In the original Jason reasoning cycle there were two transitions starting from the step `RelPl` because there were two cases: when there are relevant plans for the selected event, and when there aren't. In O3A there is only one transition since secondary emotions can be derived even if there are no relevant plans (which is the next step).

Derivation of Secondary Emotions (DeriveSEM): In this step secondary emotions are derived using the function $SEMDer\,(ag_{bs}, T_{\varepsilon}, T_R)$. There are three possible transitions from this step. The cycle can move towards `UpMood` if primary or secondary emotions were appraised (first condition). If no emotions were appraised but there are relevant plans (second condition), the cycle goes on with the `ApplPl` step. If none of these two previous conditions hold, then the next step will be `SelEv`. Although this is not an optional step in the BDI reasoning cycle, it is up to the programmer to decide when and how secondary emotions should be derived.

Update Mood (UpMood): After primary and secondary emotions are appraised, the current mood is updated. The function $UpMo(Mo, T_{PEM}, T_{SEM})$ represents this task.

Selection of an Applicable Plan (SelAppl): This transformation rule is similar to its Jason counterpart, except the S_{Ap} function (which selects one applicable plan) has an additional parameter: the current mood (Mo). The agent current mood influences the process of selecting the action to respond to one event or to reach a goal. This function can be customized by the programmer.

Mood Decay (MoodDecay): This step determines how mood decays. Information related to the agent personality is used by O3A in a function represented by $MoodDc(Mo, P, \delta)$, where δ is the decay rate that determines how mood will decay in each reasoning cycle.

4 Conclusions

In this work we offer an informal semantic description of O3A, an Open Affective Agent Architecture. This formalization integrates the affective characteristics of an O3A agent into the BDI reasoning cycle. We have extended the operational semantic used in Jason. Our architecture has its grounds on widely studied psychological and neurological theories, and it can be easily adapted to the particular emotional requirements of the problem to be solved. Agents that are created according to the proposed reasoning cycle will be able to have a representation of the current state of the world as well as of its own affective state. Therefore the O3A decision making process will be influenced by emotions, mood and personality. Different implementations of each of the proposed functions will modify the way in which the affective changes occur in the agent, so, although we offer a default implementation for the functions, they can also be fitted to particular requirements or theories.

This is a work in progress, and we are currently engaged in completing this formalization. Our immediate aim is to evaluate the O3A architecture and its formalization by enriching previous experiments [1] in order to reach agents' behaviors closer to humans behavior.

Acknowledgments. This work was supported by the Spanish government grant MINECO/FEDER TIN2012-36586-C03-01 and HUMBACE (Human and Social Behavior Models for Agent-Based Computational Economics) project. Thanks also to the Research and Development Support Programme of the Universidad Politécnica de Valencia PAID-06-11 Project.

References

1. B. Alfonso, E. Vivancos, V. Botti, and P. Hernández. Building Emotional Agents for Strategic Decision Making. In *Proceedings of the 7th ICAART*, pages 390–397, 2015.
2. B. Alfonso, E. Vivancos, and V. J. Botti. An Open Architecture for Affective Traits in a BDI Agent. In *In Proceedings of the 6th ECTA. Part of the 6th IJCCI*, pages 320–325, 2014.
3. C. Becker-Asano and I. Wachsmuth. Affective Computing With Primary and Secondary Emotions in a Virtual Human. *AAMAS '10*, 20(1):32–49, 2010.
4. R. H. Bordini and J. F. Hübner. Semantics for the Jason Variant of AgentSpeak (Plan Failure and Some Internal Actions). In *Proceedings of the 19th Conference ECAI 2010*, pages 635–640, Amsterdam, The Netherlands, 2010. IOS Press.
5. C. Castelfranchi. Affective Appraisal Versus Cognitive Evaluation in Social Emotions and Interactions. In *Affective Interactions*, pages 76–106. Springer, 2000.
6. A. R. Damásio. *Descartes' Error: Emotion, Reason, and the Human Brain.* Quill, 1994.
7. H. A. Demaree, M. A. DeDonno, K. J. Burns, P. Feldman, and D. E. Everhart. Trait dominance predicts risk-taking. *Personality and Individual Differences*, 47(5):419–422, 2009.
8. J. Dias, S. Mascarenhas, and A. Paiva. Fatima Modular: Towards an Agent Architecture With a Generic Appraisal Framework. In *Proceedings of the International Workshop on Standards for Emotion Modeling*, 2011.
9. P. Gebhard. ALMA: A Layered Model of Affect. In *Proceedings of the 4th IFAAMAS*, pages 29–36, NY, USA, 2005. ACM.
10. S. C. Marsella and J. Gratch. EMA: A Process Model of Appraisal Dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009.
11. R. R. McCrae and O. P. John. An Introduction to the Five-Factor Model and its Applications. *Journal of personality*, 60(2):175–215, 1992.
12. A. Mehrabian. Analysis of the Big-Five Personality Factors in Terms of the PAD Temperament Model. *Australian Journal of Psychology*, 48(2):86–92, 1996.
13. A. Mehrabian. Pleasure-Arousal-Dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. *Current Psychology*, 14(4):261–292, 1996.
14. J.-J. C. Meyer. Reasoning About Emotional Agents. *International Journal of Intelligent Systems*, 21(6):601–619, June 2006.

15. J.-J. C. Meyer, W. Van der Hoek, and B. Van Linder. A Logical Approach to the Dynamics of Commitments. *Artificial Intelligence*, 113(1):1–40, 1999.
16. K. Oatley and J. Jenkins. *Understanding Emotions*. John Wiley & Sons, 1996.
17. A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, July 1988.
18. D. Pereira, E. Oliveira, and N. Moreira. Formal Modelling of Emotions in BDI Agents. In F. Sadri and K. Satoh, editors, *Computational Logic in Multi-Agent Systems*, volume 5056 of *Lecture Notes in Computer Science*, pages 62–81. Springer, 2008.
19. A. S. Rao and M. P. Georgeff. Decision procedures for BDI logics. *Journal of logic and computation*, 8(3):293–343, 1998.
20. R. M. Ryckman. *Theories of Personality*. PSY 235 Theories of Personality Series. Thomson/Wadsworth, 2007.
21. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotions: Integrating Qualitative and Quantitative Aspects. In G. Mali, C. Spyropoulos, N. Fakotakis, and N. Avouris, editors, *Proceedings of the 18th ECAI'08*, pages 256–260, Greece/Amsterdam, 2008. Patras / IOS Press.
22. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotion-Based Action Tendency for Intelligent Agents. In *Proceedings of the 14th EPIA '09*, pages 174–186, 2009. Springer-Verlag.
23. B. R. Steunebrink, M. Dastani, and J.-J. C. Meyer. A Formal Model of Emotion Triggers: an Approach for BDI Agents. *Synthese*, 185:83–129, 2012.
24. R. Vieira, Á. F. Moreira, M. Wooldridge, and R. H. Bordini. On the Formal Semantics of Speech-Act Based Communication in an Agent-Oriented Programming Language. *J. Artif. Intell. Res.(JAIR)*, 29:221–267, 2007.

Rule based appraisal of emotions in drama

V. Lombardo[1], C. Battaglino[1], R. Damiano[1], and A. Pizzo[2]

[1] Department of Computer Science and Cirma, University of Torino
[2] Department of Humanities and Cirma, University of Torino
vincenzo|rossana|battagli@di.unito.it,antonio.pizzo@unito.it

Abstract. In stories, the emotional charge of the characters plays an important role in engaging the audience. The emotional states of the characters allow the audience to understand their motivations and to perceive their reactions to a dramatic situation. In this paper, relying on a semantic representation of the drama features, we present and evaluate an emotional rule system that generates the characters' emotions based on their representation of their mental states.

Keywords: emotion annotation, drama ontology

1 Introduction

Computing characters' emotions is relevant for a number of tasks ranging from retrieval to editing. Consider, for example, the following scenarios: a system, conceived for the general public, that searches a (multimedia) story bank (such as, e.g., [6], see below), through an effective tool that goes beyond mere editorial metadata (title, author, etc.), able to answer queries of the type "the novel where a woman drowns her husband with the help of her lover but eventually goes insane from remorse"; an environment for assisted drama editing (such as, e.g., Dramatica[3]), where the writer can visualize the course of characters' emotions along the plot and assess their timing and coherence.

Cognitive theories of emotions can provide a systematic account of characters' emotions in stories. According to cognitive theories [13], emotions stem from how a character *appraises* a given situation with respect to its own goals and moral standards: if it appraises some event as beneficial, it is happy; if it appraises some event as deleterious, it is worried or disgusted; etc. Since the notion of appraisal advocates an intentional account of agency, cognitive theories of emotions have been integrated into virtual characters by using the well known BDI model, which provides the required primitives for the appraisal process [2, 3].

In this paper, we leverage a computational model of emotions [1], based on the OCC theory of emotion appraisal [15], to create a set of rules that compute characters' emotions based on a description of their goals and values. We assume a BDI based description of the characters [11, 12], where characters are driven by their goals and respond to the violations of their values, engaging in conflicts

[3] http://dramatica.com

that are the input to their emotions. suitable to develop functionalities such as the search and editing functions mentioned above.

This paper is structured as follows: after surveying the related works about how appraisal theories are encoded in intelligent agents (Section 2), we illustrate the basic encoding of the drama facts through the Drammar ontology (Section 3). Section 4 presents the emotional rules system, while Section 5 presents the experiment.

2 Related Work

A varieties of recent projects have investigated the creation of story repositories with formal tools. Propp's work, in particular, has been the object of formalization with AI tools in fields that range from the creation of fictional story worlds [7] to narrative generation [8]. The DramaBank Project [6] is a repository of semantically encoded narratives, based on a formal annotation, oriented at the surface generation of different stories from shared nuclei [18]. The DramaBank annotation language accounts for causality and intentionality in stories with specific operators, such as *Attempt to cause*, but does not account for an emotional level in characters, since they are mostly concerned with the encoding of plots rather than character structures. The Narrative Knowledge Representation Language (NKRL) proposed by [19] also provides tools for the annotation of the narrative content, but it does not acknowledge the role of the characters and their emotions.

The integration of emotions into virtual characters' architectures has seen its first, pioneering approach in [5], emotional states are explained as a consequence of specific configurations of mental states (e.g., beliefs and goals), that are the output of a person's appraisal of the environment she/he is situated in. For example, a situation may be *desirable* with respect to the person's goals, or it may be appraised as *immoral* because it contains some immoral action with respect to the moral beliefs of the person. A number of computational models of emotions, including [5], rely on the appraisal theory proposed by Ortony, Clore and Collins [15] (OCC). A relevant feature of computational models of emotions is that the emotion appraisal process is carried out in a domain–independent fashion. In [17], the independence of the appraisal process from the domain is limited to the desirability of events, which is based on goal processing; the appraisal of actions as praiseworthy and blameworthy, on the contrary, is reduced to the principles such as "help my goals to succeed" or "do not cause my goals to fail". In [4], the appraisal of events is independent from the domain, and is carried out by processing the syntactic information encoded in the representation of plans (e.g., the probability of success) and goals (e.g., the success or failure conditions). The system, however, does not contain the necessary information for generating the appraisal variables, which are necessary to Attribution emotions. In [1], Attribution emotions, such as Pride or Shame, are derived from the evaluation of actions in a domain–independent way, based on the notion of moral values (such as 'honesty', 'freedom').

3 Drammar Ontology

Drammar[4] is a computational ontology for the representation of the elements of the drama (for details about the encoding, see [10]). For a description of the theoretical foundations for dramatic elements see [10].

Drammar representation of characters centers upon the notion of agents' intention (realized through a plan) and the goal achieved (or tried to achieve). A plan consists of the actions that are to be carried out in order to achieve some goal; plans are organized hierarchically, with high–level behaviors formulated as lower–level plans (called subplans). Goals originate from the values of the agents that are engaged by the plans, i.e., put at stake or balanced through the plan actions, given the beliefs (i.e., the knowledge) of the agents. The representation of dramatic characters is formalized through the rational agent paradigm, or BDI (Belief, Desire, Intention) paradigm [2] (which has already seen some applications in the computational storytelling community [14] [16]).

The scenes are the places for the interplay of the actions that are carried out by the agents to achieve their goals. The scene is built in order to orchestrate the conflicts (or, alternatively, the support relations) over the goals and to induce into the agents the emotions sought after by the author of the drama. The emotions felt by the agents are the dramatic qualities *par excellence* and are computed through the appraisal operation. The appraisal operation, encoded through SWRL rules, will be addressed in detail in the next section. The representation example in Fig. 1 refers to a scene taken from the "North by Northwest"

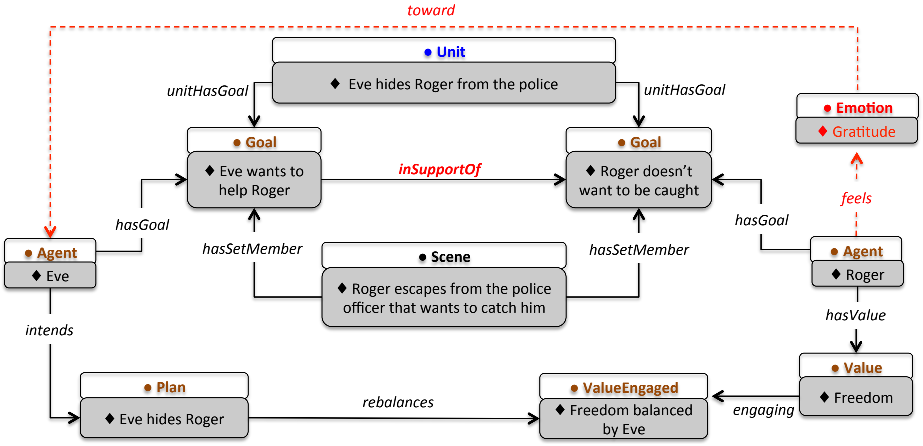


Fig. 1: The representation of the example: Eve helps Roger to hide from the police officers.The dotted lines indicate the annotation of emotion felt by Roger.

[4] An available version of the ontology, encoded in the OWL 2 RL language, can be downloaded at http://www.di. unito.it/ vincenzo/FTP_SWJ/

movie by Alfred Hitchcock, a tale of mistaken identity where the main character Roger tries to prove that he is not the 'double' George Kaplan. In the example, we model the scene in which Eve helps Roger to hide from the police officers who want to catch him, because they believe that he is an assassin. Eve is a spy of the USA government and knows that Roger is not an assassin, so she helps him. Roger feels Gratitude toward Eve, because her goal is in support of Roger's goal of not being caught and her plan re-balances Roger's *Freedom* value.

In Fig. 1, the incident described above is encoded in the `Unit` *"Eve hides Roger from the police"* (top). What motivate this unit are the following goals: 1) Roger's goal to not be caught by the police officers; 2) Eve's goal to help Roger; with the first goal being supported by the second one. The plan devised by Eve to achieve her goal engages Roger's value of Freedom. Goals and engaged values are handled through a scene structure depicted in the figure. In the example, emotions are represented by the properties `feels` and `toward` instantiated by the rules, so that Roger feels gratitude toward Eve.

4 Rule-based emotion generation

The automatic annotation of emotions is conducted via a set of rules, informed on a computational model of the emotional agent, namely the Moral Emotional Agent described in [1].

As anticipated in Section 2, in OCC theory emotions are activated as a consequence of a person's (here, an agent's) subjective appraisal of a given situation. The *appraisal process* encompasses the following elements: the appraising *agent*, the appraised *situation*, the *dimension* of appraisal. Depending on the configuration of these elements, different emotion types are generated. The OCC theory acknowledges three main dimensions of appraisal: the utilitarian dimension of *desirability* (or undesirability), that [1] map onto the achievement (or failure) of goals, following an established tradition in computational models of emotions (e.g., Joy or Distress); the moral dimension of *praiseworthiness* (or blameworthiness), that [1] map onto the compliance (or conflict) with moral values (e.g., Pride or Shame); the *affection* for an entity involved in the situation. The utilitarian dimension can be also appraised by the agent from the point of view of another agent, thus generated other agent-oriented emotions (e.g. Pity or Reproach).

The *target* of the emotion, then, varies depending on the appraisal of the situation as a mere event or as an intentional act: in the former case, the target of the emotion is the event itself and the relevant dimension of appraisal is the desirability of the event; in the latter case, the target is the agent who intentionally performed the act and the relevant appraisal dimension is the praiseworthiness of the action. A third case is the appraisal of a specific entity (e.g., an object or a person) involved in the situation according to an affective, subjective inclination (e.g., Love and Hate): here, we do not consider this case since the affection towards the target is intrinsic to the appraising agent and cannot be computed. If the appraised situation is still ongoing, a *prospect-based* emotion will be generated based on the agent's expectation about its outcome (e.g., Hope or Fear). Otherwise, the generated emotion type depends on the actual outcome of the event with respect to the dimensions of desirability and praiseworthiness (e.g., Relief).

In OCC, emotions are grouped into emotion families depending on the appraisal dimensions. When the appraisal dimension is desirability, *Well–being* emotions are generated; these can be *Prospect–based* if the refer to the prospective accomplishment of events. The appraisal of actions according to the moral dimension gives rise to *Attribution* emotions. The appraisal of situations from the perspective of other agents gives rise to *Fortune–of–Others* emotions.

In previous work, we chose SWRL rule language [9] as the formal tool for encoding the emotion annotation rules [10]. The SWRL rules augment the OWL–based representation with a rule layer built on top of it, adding the possibility to declare arbitrary Horn clauses expressed as *IF THEN* rules. Encoding emotion generation using *SWRL* rules enables the automatic generation of the emotions of the characters in a scene annotated in Drammar.

Translating the computational model of emotions into the emotion generation rules involves a mapping of the elements of the appraisal process (appraising agent, situation and dimension of appraisal) onto the primitives of the Drammar ontology. Basically, the rule antecedent represents a character's appraisal of a situation, and is based on the character's goals, values and plans (e.g., a goal achieved or not, a value put at stake, a plan the character is committed to). The rule consequent asserts what emotions the character feels as a consequence of the appraisal and what is the target of the emotion.

The appraised situation is mapped onto a scene of the drama and the appraising agent is mapped onto a character featured in the scene. Modelling the appraisal dimension requires a more complex mapping. The content of the scene is represented as a set of variables that correspond to goals (`Goal` in Drammar), achieved by plans (`Plan` in Drammar), and values (`Value` in Drammar), engaged by the execution of plans. Appraisal dimensions are represented as relations over this set of variables. The appraisal of an event as *desirable* (or *undesirable*) depends on the relation between a goal of the appraising agent and another's agent goal, achieved by the plan of the other agent in the scene. The relation is expressed through the properties `inConflictWith` or `inSupportOf`: an event is desirable if the goal it achieves is `inSupportOf` of the agent's goal, undesirable otherwise. Notice that, in this case, a plan is construed as an (intentional) event, in line with the OCC theory.

The appraisal of an action as *praiseworthy* (or *blameworthy*) depends on the relation between a character's value and a plan committed by another agent (or by the agent itself) as a way to achieve some goal. The relation between a value and a plan is expressed by the property `atStake` concerning one of the values of the character: if the value is put `atStake` as a consequence of the execution of a plan in the scene, the plan is blameworthy; otherwise, if a value is not at stake anymore after the execution of a plan, the plan is praiseworthy.

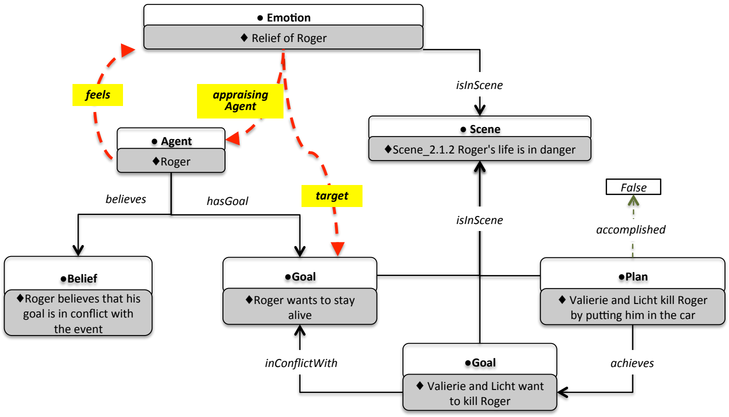
The temporal dynamics of the appraised situation, relevant for Prospect–based emotions, is grasped by a property describing the status of the plan execution in the agent's expectations. The status of a prospect event is expressed by the property `accomplished` of a plan, whose value is a string. A plan accomplishment can be *uncertain* (i.e., "uncertain") if the agent expects the plan to achieve its goal, *successful* (i.e., "true") if the plan has been successfully executed and has achieved its goal as expected, *failed* (i.e., "false") if the plan has not achieved its goal, differently from what expected. The Fig. **??** illustrates the rules for emotion generation. *Well-being* emotions, such as Distress and Joy, depend on the relation between a `Goal` *?G* and a `Goal` $?G_{SA}$ owned by an `Agent`. An event is desirable if it encompasses a plan that achieves a goal *?G* `inSupportOf` of the agent's goal $?G_{SA}$, undesirable if the goal *?G* is `inConflict` with the agent's goal.

Fortune-of-others emotions, such as Happy-for another agent, depends on the agent's emotions *Love/Hate* for another agent encoded in the representation and on the (un)desirability of an event for an other agent's *Goal* $?G_{OA}$. For example, if the `Agent` *?SA* loves another `Agent` *?OA* and the `Goal` *?G* is `inSupportOf` the `Goal` $?G_{OA}$ of the other `Agent` *?OA*, *?SA* feels *Happy-for* for the other agent *?OA*. Otherwise, the agent feels *Gloating* toward the other agent.

Attribution emotions arise when the agent appraises the consequences of an action with respect to its values. This happens when an `Agent` *?SA* owns a `Value` *?V* that is a *ValueEngaged ?VE* in the effects of the `Plan`. The `Agent` *?SA* appraises the `Plan` *?P* as *praiseworthy* if the value *?VE* is re–balanced by the plan (i.e., the data property `atStake` of *?V* is *false* as a consequence of the plan); the `Plan` *?P* is blameworthy if *?VE* is put at stake by the plan (i.e., the data property `atStake` of *?V* is *true* as a consequence of the plan). Attribution emotions can be self– or other–directed: the `Agent` *?SA* feels Pride or Shame if it intends the `Plan` *?P* and the plan is, respectively, praiseworthy or blameworthy. Otherwise, if another `Agent` *?OA* in the scene intends the `Plan` *?P*, *?SA* feels Admiration or Reproach.

Compound emotions arise when the agent feels Well-being emotions and Attribution emotions at the same time. Gratification (Remorse) emotion rule fires if the `Agent` *?SA* feels *Joy* (*Distress*) and *Pride* (*Shame)* in the `Scene` *?S*, *SWRL Gratification*(*SWRL Remorse*) rule fires and *?SA* also feels *Gratification (Remorse).* Gratitude (Anger) emotion rule fires if the `Agent` *?SA* feels *Joy* (*Distress*) and *Admiration* (*Reproach)* in the `Scene` *?S*, *SWRL Gratitude*(*SWRL Anger*) rule fires and *?SA* also feels *Gratitude (Anger).*

In the following (Fig. 2), we describe the activation of the SWRL rule for Relief for the agent Roger in the running example taken from the "North by Northwest" movie by Alfred Hitchcock. In particular, we focus on the scene in which two foreign spies, Valerie and Licht, believing that Roger is George Kaplan, try to kill him by forcing him to drink bourbon and by putting him into a moving car. Roger manages to exit from the car before it falls off a cliff. The `Scene` *"Scene_2.1.2 Roger's life is in danger"* has one `Agent`: the main character *"Roger"*. The emotional charge of the scene is usually described in the traditional *misè en scene* focusing on the conflict between the two goals: Valerie and Licht want to kill Roger; Roger wants to stay alive. Given the event represented by Valerie and Licht's goal and by their failed plan, the system succeeds in calculating the resulting characters' emotional charge. Following the SWRL rules, the system outputs Roger's relief as the emotions triggered in the scene that corresponds to the unit. In (Fig. 2), the event is represented by the `Plan` *"Valerie and Rick kill Roger by putting him in the car"* that achieves Valerie and Licht's `Goal` *"Valerie and Licht want to kill Roger"*. The plan has the data


• Emotion
◆ Relief of Roger
feels
appraising Agent
isInScene
• Scene
◆Scene_2.1.2 Roger's life is in danger
• Agent
◆Roger
target
False
believes
hasGoal
isInScene
accomplished
●Belief
◆Roger believes that his goal is in conflict with the event
●Goal
◆Roger wants to stay alive
●Plan
◆ Valierie and Licht kill Roger by putting him in the car
inConflictWith
●Goal
◆ Valierie and Licht want to kill Roger
achieves

EVENT: the Plan ``*Valerie and Licht kill Roker by putting him in the car''* that achieves the Goal ``*Valerie and Licht want to kill Roger''*

Fig. 2: The annotation of the scene for the `Agent` *"Roger"*. The property `target`, `feels` and `appraisingAgent` are inferred by the rule for Relief emotion.

property `accomplished` set to *false*, this means that the event is *discofirmed*. The `Agent` *Roger* has the `Goal` *"Roger wants to stay alive"* that is `inConflictWith` Valerie and Licht's goal and the agent believes that his goal is in conflict with the event. Thus, the `Agent` *Roger* appraises the event as an *undesirable disconfirmed* event that leads to the activation of the Relief SWRL rule. The Relief rule consequent asserts that the `Agent` *Roger* is the `appraisingAgent` that feels the `Emotion` *Relief of Roger*, with the `Goal` *"Roger wants to stay alive"* as target (property `target`).

5 Evaluation & Discussion

In this section, we describe an experiment that aims at evaluating the application of the emotional rules presented in Section 4 on the data obtained by the manual annotation of stories by experts.

Experimental Protocol. The annotated corpus included two Hollywood movies, the historical romance *Casablanca* (by Michael Curtiz) and the unlikely thriller *North by northwest* by Alfred Hitchcock, respectively; an opera, *Carmen* (George

Bizet, libretto of Henri Meilhac and Ludovic Halévy), and the Greek tragedy *Oedipus the King* (Sophocles). The characters whose emotions are annotated are: *Roger* (North by northwest movie), *Rick*, *Ilsa* and *Laszlo* (Casablanca movie), *Carmen*, *Don Josè*, and *Michaela* (Carmen opera), and finally *Oedipus* (Oedipus Greek tragedy).

Each drama in the corpus was segmented into units and analyzed by an annotator who identified the segment's main incidents and then annotated the main actional elements of the units and the OCC–classified emotion types felt by the main characters. The annotators were students of the Media and Arts program, trained in dramatic narration; each work was annotated by a different annotator, selected based on her/his familiarity with the work. Subsequently, for each segment identified by the annotator, a drama scholar annotated the goals, plans, and values involved in the segment in the formal language of *Drammar*. Then, the annotation was fed to a reasoner[5] for the application of the *SWRL* emotion rules presented in Section 4.

We compared the improvement brought about by the rule with the results of a preliminary experiment, described in [10]. With respect to previous work [10], the rule set presented in Section 4 contains a monotonically more fine–grained encoding of the agent's expectations about prospect events, and of the agent's appraisal of the relation between its goals and the goal achieved in the appraised situation. This improvement allows us to discriminate between Prospect-based emotions and Well-being emotions, thus avoiding conflicts in rule activation. For the comparison, we availed ourselves of the following measures: *Human Annotated Emotion Types Detection* and *Tokens Accuracy*. The *Human Annotated Emotion Types Detection* represents the capability of a system of detecting the set of emotions types (i.e., the emotional range) annotated by humans for each character in the corpus. It is calculated by computing precision and recall of the generated emotion types on the emotion types annotated by the human annotators. This measure is not dependent on the number of tokens of a specific emotion types. The *Tokens Accuracy* represents the accuracy of a system in generating the number of emotions tokens annotated by humans for a given character in the corpus. It is calculated by computing precision and recall on the emotions type tokens (i.e., the single instances of each emotion type). This measure takes into consideration the number of times that the human annotators or the systems generate a specific emotion type.

Results. Regarding the *Human Annotated Emotion Types Detection*, the average precision is 0.88 and the average recall is 0.94 (see Table 1). With respect to previous results [10], we obtained an higher average precision (0.88 against 0.71) and an higher average recall (0.94 vs 0.89). In particular, the improvement regards the precision for characters who felt emotions types that belong to Prospect-based emotions such as Roger (0.73 vs 0.69), Rick (0.71 vs 0.62), Ilsa (1 vs 0.5), Laszlo (1 vs 0.8) and Oedipus (1 vs 0.79) (see Table 1).

Regarding the *Tokens Accuracy* measure, the average precision is 0.72 while the average recall is 0.93 (see Table 2). With respect to previous results, that

[5] Pellet, www.clarkparsia.com/pellet

show an average precision and recall equal to 0.51 and 0.85, respectively, the improvement is more apparent when we computed the *Tokens Accuracy* measure because it considers also the number of times that a certain emotions type is annotated by humans and generated by the systems (see Table 2 - Roger (0.62 vs 0.32), Rick (0.52 vs 0.43), Laszlo (0.83 vs 0.5) and Oedipus (0.91 vs 0.62)).

	NbN		**Casablanca**			**Carmen**		**OedipusAll**	
	Roger	Rick	Ilsa	Laszlo	Carmen	D. Juan	Micaela	Oedipus	
Precision	0.73	0.71	1	1	0.8	0.79	1	1	0.88
Previous Precision	0.69	0.62	0.5	0.8	0.5	0.79	1	0.79	0.71
Recall	1	0.83	1	1	0.8	1	1	0.89	0.94
Previous Recall	1	0.83	0.75	1	0.8	1	1	0.78	0.89

Table 1: Detection of the emotions types (*Human Annotated Emotion Types*).

These improvements are due to the fact that our emotional rules with respect to those presented in [10] do not include the appraisal of Hope (Fear) as part of the appraisal of Disappointment (Relief) and Satisfaction (Fear-confirmed) and the appraisal of Joy (Distress) as part of the appraisal of Relief (Disappointment). For example, 0 tokens of the Fear emotion type are annotated by the human annotator for *Roger* in *North by Northwest*: while the previous rule system generated 6 emotion tokens of this emotion type, our rules discriminate the appraisal in a more efficient way and are in line with the human annotation.

	NbN		**Casablanca**			**Carmen**		**OedipusAll**	
	Roger	Rick	Ilsa	Laszlo	Carmen	D. Juan	Micaela	Oedipus	
Precision	0.62	0.52	0.75	0.83	0.57	0.55	1	0.91	0.72
Previous Precision	0.32	0.43	0.5	0.5	0.27	0.47	1	0.62	0.51
Recall	1	0.90	1	0.83	0.8	1	1	0.91	0.93
Previous Recall	1	0.90	0.66	0.83	0.6	1	1	0.83	0.85

Table 2: Precision and recall on emotions tokens (*Tokens Accuracy*).

6 Conclusion

In this paper, we described a system for the automatic generation of characters' emotions in stories, encoded in a set of SWRL rules. A rule based system alleviates the task of manual annotation of characters' emotions by providing a coherent and founded model for character emotion generation through a variety of media. We designed and ran an experiment where the emotions automatically generated by the rules were compared to the emotions assigned by human annotators to story characters on a corpus of stories ranging from traditional to new media. The experiments showed a good performance of the model with respect to the annotation provided by the humans.

References

1. C. Battaglino, R. Damiano, and L. Lesmo. Emotional range in value-sensitive deliberation. In *Proc. of the 12th Int. Conf. on Autonomous Agents and Multi-Agent Systems*, (AAMAS'13), pages 769–776, 2013.
2. M.E. Bratman. *Intention, Plans, and Practical Reason.* Harvard University Press, Cambridge (MA), 1987.
3. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
4. João Dias, Samuel Mascarenhas, and Ana Paiva. Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Workshop on Standards in Emotion Modeling*, Leiden, 2011.
5. C. D. Elliott. *The Affective Reasoner: A process model of emotions in a multi-agent system.* PhD thesis, Northwestern University, 1992.
6. D. Elson. Dramabank: Annotating agency in narrative discourse. In *LREC*, pages 2813–2819, 2012.
7. C. R. Fairclough and P. Cunningham. A multiplayer opiate. *Int. Journal of Intelligent Games & Simulation*, 3(2), 2004.
8. P. Gervás. Propp's morphology of the folk tale as a grammar for generation. In *Proc. of Int. Workshop CMN*, pages 106–122, 2013.
9. I. Horrocks, P. F. Patel-Schneider, and H. et al. Boley. Swrl: A semantic web rule language combining owl and ruleml. *W3C Member submission*, 21:79, 2004.
10. V. Lombardo, C. Battaglino, R. Damiano, A. Pizzo, and A. Lieto. Coupling conceptual modeling and rules for the annotation of dramatic media. *Semantic Web*, 00:1–32, 2014.
11. V. Lombardo and R. Damiano. Semantic annotation of narrative media objects. *Multimedia Tools and Applications*, 59(2):407–439, 2012.
12. V. Lombardo and A. Pizzo. Multimedia tool suite for the visualization of drama heritage metadata. *Multimedia Tools and Applications*, pages 1–32, 2014.
13. S. C. Marsella, J. Gratch, and P. Petta. Computational models of emotion. In *A blueprint for an affectively competent agent.* Oxford University Press, Oxford, 2010.
14. E. Norling and L. Sonenberg. Creating Interactive Characters with BDI Agents. In *Proc. of the Australian Workshop on Interactive Entertainment IE2004*, 2004.
15. A. Ortony, G. Clore, and A. Collins. *The Cognitive Structure of Emotions.* Cambrigde University Press, 1988.
16. F. Peinado, M. Cavazza, and D. Pizzi. Revisiting Character-based Affective Storytelling under a Narrative BDI Framework. In *Proc. of ICIDIS08*, Erfurt, Germany, 2008.
17. W. S.Scott Reilly. Believable social and emotional agents. Technical report, DTIC Document, 1996.
18. E. Rishes, S. M. Lukin, D. K. Elson, and M. A. Walker. Generating different story tellings from semantic representations of narrative. In *Interactive Storytelling*, pages 192–204. Springer, 2013.
19. G. P. Zarri. Conceptual and content-based annotation of (multimedia) documents. *Multimedia Tools and Applications*, 72(3):2359–2391, 2014.

Exploring Sentiment in Social Media and Official Statistics: a General Framework

Emlio Sulis[1], Mirko Lai[1], Manuela Vinai[2] and Manuela Sanguinetti[1]

[1]Università degli Studi di Torino, Dipartimento di Informatica
c.so Svizzera 185, I-10149 Torino (Italy)
{sulis,milai,msanguin}@di.unito.it
[2] Q.R.S. soc. coop.
V.le C. Battisti 15, 13900 Biella
vinai@qrsonline.it

Abstract. The integration between official statistics and social media data is a challenging topic. This contribution aims to present a recently-designed framework to compare sentiment analysis on social media content with social and economic data. Such framework - which has already been applied, in a preliminary fashion, to the Felicittà project - is meant to integrate official statistics and correlate it with online social media data. Its ultimate goal, in fact, namely consists in giving a contribution to the definition of a measure of subjective well-being that could fully benefit from both traditional, well-established social indicators and dynamic data obtained from the web.

Keywords: Subjective Well-Being, Sentiment Analysis, Official Statistics, Social Media

1 Introduction

The significant growth of user-generated content on the web, and in particular the increased availability of data from online social media, has fostered the development of automatic techniques for the extraction and processing of such content for different purposes. This development is reflected, among other things, by the spread of scientific contests whose main track is the Sentiment Analysis (henceforth SA) of texts in different languages (see eg. SemEval[1] for English, and SENTIPOLC@EVALITA2014[2] for Italian).

In turn, these achievements are encompassed into a broader and interdisciplinary debate related to the study and definition of measures that could be considered as reliable indicators of the well-being of a community. In fact, a growing debate has recently involved the measurement of social and individual well-being. New statistical measures have been proposed besides the bare Gross Domestic Product (GDP), traditionally seen as the best way to measure national

[1] http://alt.qcri.org/semeval2014/task9/

[2] http://www.di.unito.it/~tutreeb/sentipolc-evalita14

economic results. Among such measures are a large amount of indicators that, in several ways and from different points of view, attempt to assess the degree of "happiness" and life satisfaction, also designated with the expression *Subjective Well-Being*, or simply SWB (see Section 2). Such measures are usually provided by governmental institutions or entitled research organizations, and they generally include social indicators measuring life quality and concerning all major areas of citizens' lives. However, such data are static and their recovery may require much efforts in terms of time and resources. Moreover, the increasing success of Sentiment Analysis (SA) techniques on social media has made it possible to develop alternative tools and measures, with respect to the latter, to assess the degree of happiness and well-being. Social media and their content can thus be used to complement and corroborate the information gathered from traditional data sources as regards SWB detection.

The work presented here is just part of this research context. In particular, the purpose of this paper is to describe a framework whose entire definition and completion is still in progress, for the analysis and assessment of the degree of "happiness" of a given community in Italy, taking into account and combining together the information gathered from two main data sources: *a)* social media content, and Twitter in particular; *b)* the socio-demographic information made available by the main suppliers of official statistical data, such as the Italian National Institute of Statistics (ISTAT)[3]. The first point in particular has actually been explored and developed under a recent project called Felicittà[4] [1], i.e an online platform for estimating happiness in Italian cities that daily analyzes Twitter posts and exploits temporal and geo-spatial information related to tweets, in order to enable the summarization of SA outcomes.

The present work is both an extension and a comprehensive reference framework of that project. As a matter of fact, its aim is manifold and includes: 1) the use and further development of techniques for the visualization of SA outcomes in Italian texts; 2) the study of the correlations between official statistics and user-generated media content; 3) providing a contribution to the debate on what can be considered effective and reliable indicators of social well-being.

The remainder of the paper is structured as follows: Section 2 provides a brief introduction to the notion of subjective well-being, summing up the more recent work carried out on this matter while Section 3 describes the whole architecture of the system, as currently conceived. Final remarks in Section 4 close the paper.

2 Background and Related Work

The present contribution covers the debate on the Subjective Well-Being as a social indicator and sheds some light on happiness studies based on the sentiment analysis of social media.

[3] `http://www.istat.it/it/`

[4] `http://www.felicitta.net/`

Subjective Well-Being. As stated in Diener [5], SWB includes reflective cognitive evaluations about the quality of life, such as life and work satisfaction, interest and engagement, and affective reactions to life events, such as joy and sadness. The common measurements of SWB are self-report methods and surveys with questionnaires. Social indicators and life quality research is a specific field of study grown over the years as witnessed, for instance, by the birth of the review "Social Indicators Research" and underlined by the initiatives of the Organization for Economic Co-operation and Development (OECD) since the Nineties[5]. Namely the OECD recently proposed a survey-based measurement of SWB at national level [16], as alternative to purely economic measures.

The well-being of the population: from Easterlin to GNH. The Gross Domestic Product (GDP) is today the main measure of the nation's economic activity. However, since late 70's, a huge debate has grown over this measure [8]. Easterlin [7] first identified the paradox for which the increase of economic well-being in wealthier countries has no further increases in subjective well-being [13]. As alternative to GDP, new concepts have arisen as sustainable socio-economic development, governance, environmental conservation and so on. Besides the OECD, several organizations and countries take into account new measurements, basically focused on the concept of *happiness*: see, for instance, the World Happiness Report [10] in a recent United Nations initiative, or the Gross National Happiness (G.N.H.) index developed by Bhuthan. In Italy, an inter-institutional initiative proposes a set of indicators on "Equitable and Sustainable Well-Being(BES)"[6].

Social Media and Well-Being. The analysis of textual expressions in social media contents on a Big-Data scale would offers an opportunity to economists and sociologists in the measurement of social well-being. There's an open debate on the topic and several works already investigated this subject with contrasting results. Wang et al. [20] examine Facebook's Gross National Happiness (FGNH) indexes and Diener's Satisfaction with Life Scale (SWLS), and finally criticize the idea that a well-being index can be based on the contents of a specific online social networks. Quercia et al.[17] explore the relationship between sentiment expressed in Twitter messages and community socio-economic well-being and, on the contrary, they found interesting correlations between sentiment and general well-being. Kramer [11] proposed a metric to represent the overall emotional health of the nation as a model of "Gross National Happiness". Our work aims to improve these studies by the analysis of a set of more extensive official statistics, better detailed in 3.2. Social media analysis also suggests the prediction of stock market [2], and of collective mood state [15]. Emotions have been considered with respect to social media and their dynamics [14] [12], also with geographical concerns [6]. We attempt to enrich and extend these studies by focusing on a

[5] See e.g. the Better Life Index `http://www.oecdbetterlifeindex.org/`

[6] The national statistical institute ISTAT and the National Council for Economics and Labour (CNEL) propose the BES Index which analyzes the changes in quality of life in Italy focusing on 12 different areas `http://www.misuredelbenessere.it/`

finer-grained administrative territorial division; as a matter of fact, our data describe the situation not only at a national level, but also with respect to regions, provinces and municipalities.

The challenge of visualization. The lecture of patterns and trends from spreadsheets or lists of numbers is a difficult task when we have to deal with large amounts of data. An improvement is often obtained by the use of graphs. Shapes and lines immediately create meanings and significance from data. In this way, data visualization allows us to present trends, to discover what is often hidden [4] and simplify the identification of patterns not easily detectable [21]. Several different tasks can be spotted in the design of a visualization system [18]. Some interesting works already dealed with social media data, highlighting aspects of public sentiment in the web [19] or public interest information[7]. Such works inspired, in their main principles, the design of our visualization module within the framework.

3 Framework Description

The present framework aims to include several approaches and techniques in order to detect the well-being of a community under a broader perspective. The steps entailed in the design phase was: a) the definition of the whole pipeline; b) the selection of data from official statistics to be correlated with the analysis performed by the SA module; c) the presentation of the most promising patterns emerging from the comparison between social media data and official statistics. In this section, we describe the general framework architecture with an overview of its modules.

3.1 Architecture

The whole framework architecture, as shown in Figure 1, consists of 5 main parts: Providers, Data Gathering, Data Analysis, Data Exposure and Data Visualization.

Providers. Providers are the data sources: i.e Twitter, from which we retrieve the geolocated[8] Italian tweets using the Stream API, and the various socio-demographic data sources (detailed in Table 1), that return demographic and socio-economic variables of different Italian administrative divisions.

Data gathering. This module is further divided in submodules, each one tackling one particular task:

- the `Collect` submodule collects data from different providers;

[7] http://twitter.github.io/interactive/sotu2015/

[8] For details on the geolocalization methods used, see [1]

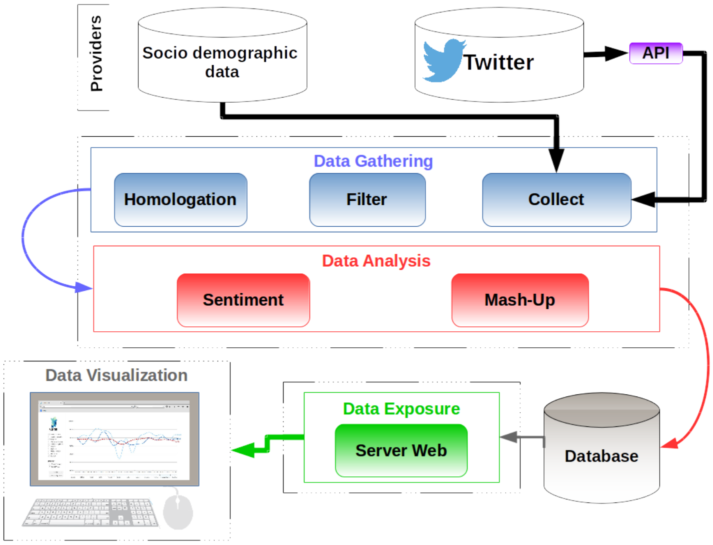


Fig. 1. Framework architecture.

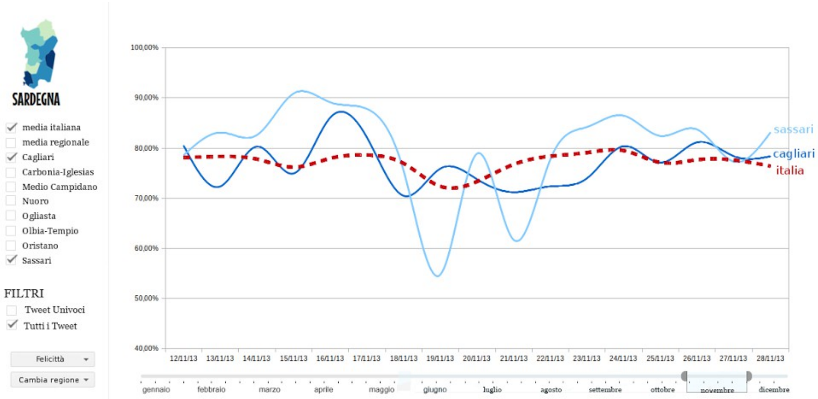
- the `Filter` submodule filters the collected data in order to remove all the possible noisy data, such as duplicate records, empty voices, characters instead of numbers and other formatting errors; as possible correlations have been observed between sentiment and time of the day or day of the week (weekdays or holidays), or between sentiment and geographical areas in a given time frame due to the occurrence of some special event, during this step, we also intend to add a further filter that leaves out all the tweets that bear such temporal or geographical bias[9], as already made in [3], in the creation of the validation corpus.
- the `Homologate` submodule is devoted to the proper organization of collected and filtered data into a unified format. For example, 1420070400, 01/01/2015 and Thu, 01 Jan 2015 00:00:00 GMT indicate the same date, and 058091, [41.53,12.28] and Roma indicate the same city. The Homologate submodule converts dates in YYYY/MM/DD format, and administrative divisions in the ISTAT code[10].

Data Analysis. First, the `Sentiment Analysis` submodule returns for each tweet a mood value (positive, negative, neutral); the SA engine is the one developed in Felicittà, as described in [1]. Then, the `Mash-Up` submodule aggregates Italian geolocated tweets by regions, provinces and municipalities. In this way, data about moods and social indicators can be grouped on the basis of the same

[9] Indeed, conventional expressions such as "*Happy New Year*", "*Merry Christmas*", and others, should not be considered as equally representative of, for example, joy.
[10] `http://www.istat.it/it/archivio/6789`

period and the same administrative level. The aggregate data are finally stored in a database. A correlation analysis across moods and, in turn, each statistic is performed, in order to quantify the strength of the relationship between the variables. As further detailed in 3.2, this is the most recent part of the project, that extends the one implemented in Felicittà.

Data Exposure. A web server exposes elaborated data by REST API. When a client runs a query, the server queries the database and returns the response.



☑ media italiana
□ media regionale
☑ Cagliari
□ Carbonia-Iglesias
□ Medio Campidano
□ Nuoro
□ Ogliasta
□ Olbia-Tempio
□ Oristano
☑ Sassari

FILTRI
□ Tweet Univoci
☑ Tutti i Tweet

Fig. 2. A query result in Felicittà that shows the degree of happiness in relation to an event in particular, that is the flood that hit the northern area of Sardinia in 2013. The graph shows that, based on the analysis of tweets from that area, i.e. the province of Sassari, at that time-frame (November, 19^{th}), a far lower degree of happiness is registered both with respect to other areas in Sardinia (such as the southern province of Cagliari) and the whole country.

Data Visualization. Finally, a web client presents the data obtained as response to the queries. For the time being, the visualization module allows to browse either the sentiment data (as in the example in Figure 2), or the sentiment data combined with demographic data, as shown in Figure 3. The part that shows socio-demographic statistics and correlations is yet to be completed.

3.2 Statistics

As a measure of the mood related to an area in a given period, we consider the percentage of positive tweets. In order to relate moods and numeric social indicators in different administration degrees, in Table 1 we summarized some social indicators that could provide an overview of the social well-being of a given community.

As data collection is not always an easy task and the Open Data is not yet widespread in Italian public administration, we realistically decided to focus

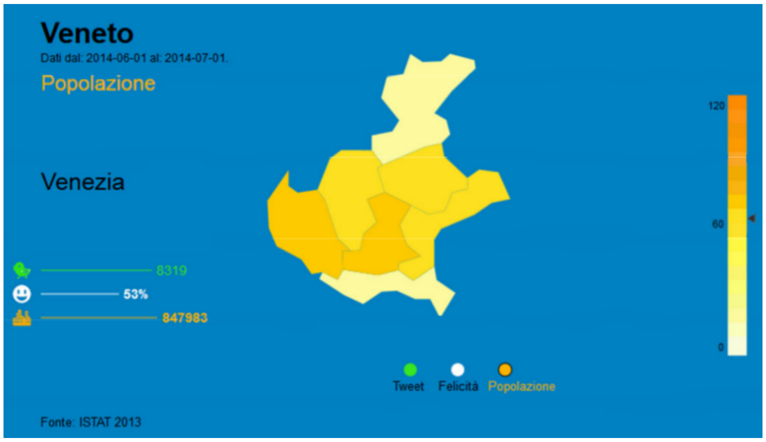


Fig. 3. Demographics and social media data in Felicittà. The provinces of each region (such as Veneto, in the picture) are coloured according to the number of inhabitants. Demographic data are combined with social media data (the number of tweets posted in Veneto in the interval of time) and mood (the percentage of positive tweets).

Description	**Source**	**Period**	**T.U.**
BES Measures	Istat	Y	R
Population by nationality, gender, marital status, and age	Istat	Y	M
Employed and unemployed by gender	Istat	M1	P
Workforce (Number of employees, artisans and so on)	INPS	M6	P
Retirements	INPS	M6	P
Companies registered and ceased by category	C.C.	Y	P
Exports / Imports	Istat	M3	R
Layoff	Istat	M1	P
Real estate market	A.E.	M3	P
Loans and bank deposits	B.I.	M6	P
Public debt of local governments	B.I.	Y	P

Table 1. Selection of Italian official statistics. Sources of data are Istat, Chamber of Commerce (C.C.), Italian Agency of Incomes (Agenzia delle Entrate, A.E.), Italian Welfare Institute (INPS) and Bank of Italy (Banca d'Italia, B.I.). Selected periods are Year (Y), month (M1), quarter (M3) or semester (M6) while Territorial Units (T.U.) are municipality (M), province (P), region (R).

our attention on data that could be easily accessed and retrieved from public administration web-sites. In order to detail different aspects of the society, we resort to different sources. In this way, we consider data from different fields and viewpoints, mainly demographic and economic.

As regards the demographic field, the main aspects considered are nationality, gender, age and marital status, since they are closely related to the perception of social well-being. We are interested, for example, in understanding whether and to what extent nationality may influence the sentiment expressed through social media, or whether married men are happier than singles.

Concerning the economic field, we consider both jobs data (e.g. the unemployement rate) and data about companies (e.g. enterprises demography). Our hypothesis is that people express negative sentiments more likely if they live in an area with significant unemployment rate or with a greater amount of cessations of business.

We also collected data about the real estate market, that we consider a typical indicator of the wealth of a territory. A correlation, in fact, is expected between this aspect and the overall mood detected in social media: the higher the prices (then the wealthier the area considered) and the greater the happiness may be. Similarly, we consider the amount of deposit and loan from the Bank of Italy as a measure of both individuals and public wealth. We selected this set of data as they are representative of different relevant social needs, with different time and granularity. A first integration between social media data and demographic data is shown in Figure 3.

Our current work then namely consists in exploring all the possible correlations between the indicators mentioned above and the output of the SA engine, and in improving the visualization module so as to better highlight such correlations and emerging patterns.

4 Conclusions and future work

In this paper we introduced an ongoing project on a framework for the analysis and assessment of the degree of “happiness” of a given Italian community, taking into account data from official statistics and SA data obtained from social media. We noticed at least two main problems: the representativeness of data and the role of ironic sentences. First, the diffusion of internet and the use of online social networks is not widespread in the same way over all kinds of population. Therefore, for instance, the sentiment of poorest people and elderly can be not represented or largely underrepresented. This is a classical problem of quite every sociological inquiries, mainly solved by representative sampling and qualitative research. A second issue is the presence of irony where the unintended meaning of words can often reverse the polarity of the message. We well know this problem and we state how exists a growing interest in this research subject, as we already investigate the role and the detection of irony and sarcasm[9]. As mentioned above, the work is still in progress and some issues limit the results, but there are also several expected positive impacts of the proposed approach.

First, we focus on the selection of data from official statistics that better correlate with social media data. An hypothesis is that a variation in the data on the labor market and, most of all, the youth employment situation in a given region entail a variation in the mood of the public opinion as expressed in online social media. Detecting the strength of the statistical relation between different variables could help in using social media as a tool for detection of social and economic trends. Another relevant concrete application of the present framework is the inclusion in the platform of Felicittà of the selected statistical data with the output emerging from the correlation analysis.

Acknowledgments. Part of the present contribution has been awarded in the 2014 Istat-Google Contest on "Producing official statistics with Big Data": `http://www.istat.it/it/archivio/144042`

References

1. L. Allisio, V. Mussa, C. Bosco, V. Patti, and G. Ruffo. Felicittà: Visualizing and estimating happiness in italian cities from geotagged tweets. In *Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI (ESSEM 2013)*, pages 95–106, 2013.
2. J. Bollen, H. Mao, and X.J. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011.
3. C. Bosco, L. Allisio, V. Mussa, V. Patti, G. Ruffo, M. Sanguinetti, and E. Sulis. Detecting happiness in Italian Tweets: Towards an evaluation dataset for sentiment analysis in Felicittà. In *Proceedings of the 5th International Workshop on EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA, (ESLOD 2014)*, pages 56–63, 2014.
4. A. Cairo. *The Functional Art: An introduction to information graphics and visualization.* New Riders, 2013.
5. E. Diener. Subjective well-being: The science of happiness and a proposal for a national index. *American psychologist*, 55(1):34, 2000.
6. P.S. Dodds, K.D. Harris, C.A. Kloumann, I.M.and Bliss, and C.M. Danforth. Temporal patterns of happiness and information in a global-scale social network: Hedonometrics and twitter. *PLoS ONE*, 2011.
7. R. Easterlin. Does money buy happiness? *The Public Interest*, 1973.
8. M. Forgeard, E. Jayawickreme, M. Kern, and M. Seligman. Doing the right thing: Measuring wellbeing for public policy. *International Journal of Wellbeing*, 2011.
9. D. Hernandez, E. Sulis, V. Patti, G. Ruffo, and C. Bosco. Valento: Sentiment analysis of figurative language tweets with irony and sarcasm. *Proc. Int. Workshop on Semantic Evaluation (SemEval-2015), Co-located with NAACL and *SEM. To appear.*, 2015.
10. J. Sachs J. Helliwell, R. Layard and Emirates Competitiveness Council. *World happiness report 2013.* Sustainable Development Solutions Network, 2013.
11. A. Kramer. An unobtrusive behavioral model of gross national happiness. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 287–290. ACM, 2010.

12. A. Kramer. The spread of emotion via facebook. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 767–770. ACM, 2012.
13. E.W.Dunn L.B.Aknin, M.Norton. From wealth to well-being? money matters, but less than people think. *The Journal of positive psychology*, 4(6):523–527, 2009.
14. L. Mitchell, M. Frank, K. Harris, P. Dodds, and M. Danforth. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLoS ONE*, 2013.
15. T. Nguyen, B. Dao, D. Phung, S. Venkatesh, and M. Berk. Online social capital: Mood, topical and psycholinguistic analysis. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, 2013.
16. OECD. *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, Paris, 2013.
17. D. Quercia, J. Ellis, L. Capra, and J. Crowcroft. Tracking gross community happiness from tweets. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*, 2012.
18. Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.
19. C.o Wang, Z. Xiao, Y. Liu, Y. Xu, A. Zhou, and K. Zhang. Sentiview: Sentiment analysis and visualization for internet popular topics. *Human-Machine Systems, IEEE Transactions on*, 43(6):620–630, 2013.
20. Kosinski M. Stillwell D.J. Wang, N. and J. Rust. Can well-being be measured using facebook status updates? validation of facebook's gross national happiness index. *Social indicators research*, pages 483–491, 2014.
21. N. Yau. *Visualize this: the FlowingData guide to design, visualization, and statistics.* Wiley Publishing, 2011.

Detecting events and sentiment on Twitter for improving Urban Mobility

A. Candelieri[1,2], F. Archetti[1,2]

[1] Department of Computer Science, Systems and Communication – University of Milano-Bicocca, Italy
[2] Consorzio Milano Ricerche, Italy
candelieri@milanoricerche.it, francesco.archetti@unimib.it,
archetti@milanoricerche.it

Abstract The streams of tweets from and to the Twitter account of urban transport operators have been considered. A computational module has been designed and developed in order to collect tweets and, on the fly, analyze them to detect some relevant event (e.g. accidents, sudden traffic jams, service interruption, etc.) and/or evaluate possible sentiments and opinions about the quality of service. Events are recognized through a simple word matching while sentiment analysis is performed via supervised learning (Support Vector Machine). The text mining solutions have been developed to work with Italian language; however they could be easily extended to other languages in the case tweets in other languages would be available. This approach has been tested for the urban transportation in Milan (Azienda Trasporti Milano, ATM) in the framework of the TAM-TAM project which has developed a technological platform for improving urban mobility by exploiting the large amount of information shared by the users of transportation services through Twitter. Events detected are used by other software modules of the TAM-TAM platform in order to support a more effective travel planning, while sentiment inferred may be used by the transport provider in order to tune the mobility supply to the commuter needs.

Keywords: smart urban mobility, sentiment analysis, crowdsourcing

1 Introduction

The relevance of "narrative aware design framework" in the design and implementation of smart urban environments has been already highlighted in [1][2]. The combined diffusion of smart mobile devices and social networks have been rapidly increasing the amount of contents generated by users, making crowdsourcing a huge source of potentially useful – usually unstructured – information to transform in actionable knowledge for services/products innovation as well improving urban quality of life. According to this vision, the Italian project TAM-TAM, co-funded by the Italian Ministry of Education, University and Research together with Regione Lombardia, has designed and developed a technological platform able to combine information from official data sources and the huge amount of unstructured information generated through crowdsourcing, even on the move, and related to transportation services in the city of Milan. Citizens, commuters and tourists already adopt socially awareness and collective intelligence to make more personalized and informed mobility decisions, mainly by reading and sharing short messages on Twitter. The aim of TAM-TAM is to close in the loop these streams of data and analyse them in order to provide users with added-value services. The benefits provided by the automatic analysis of tweets have been already investigated and proved in other domains, such as the automatic detection of anomalies related to power outage events during hurricane Irene on August 27, 2011 [3]. More recently the attention is focusing on terrorism, radicalization and hate-speech [4, 5]. With respect to these applications, where a first analysis is performed to discriminate between relevant and irrelevant tweets, limiting the collection of tweets to those posted *from* and *to* the Twitter account of the transportation company permits to consider all of them as relevant. The authors of this paper have been designing the tweets collection and analysis component of TAM-TAM, which is overall aimed at providing innovative services through:

- integration of data and information coming from different sources, both official and crowd-sourced (e.g., time-tables, on-line positioning data, traffic estimation, etc.);
- supporting intermodal and personalized transport options;
- computational modules for expressive-media contents analysis, based on sentiment and opinion mining techniques [6], for event detection and evaluation of the perceived quality of transport service;
- a travel planning software to provide users with information on costs, time, environmental impact and perceived quality of service with respect to the opinions of the other commuters;
- decision support functionalities to identify and address criticalities in the proposed urban transportation supply, enabling more effective and efficient plans according to variations in mobility users preferences.

The contribution of this paper consists in the development and validation of a computational module devoted to collect tweets, both *from* and *to* the Twitter account of the public transportation company in Milan, Azienda Trasporti Milano (ATM), and then analyse their content according to the following two goals: the automatic identification of events (e.g., accidents, sudden traffic jams, etc.), as posted by the users, the automatic detection of opinions about transport service (e.g., delays, inefficiencies, perceived security, dirt, etc.). Some preliminary results obtained during the first activities of the project have been initially reported in [7], where an initial design of the computational module is presented, further specialized in this paper.

2 TAM-TAM: general architecture

Figure 1 summarizes the overall architecture of the TAM-TAM platform, with a major focus on the component devoted to the analysis of tweets. The other relevant components and services of the platform are: *i)* the central database used to provide the different visualization layers related to official – structured – information, such as lines, time-tables, on-line positioning data, traffic estimation, etc.; *ii)* web and mobile apps for login/profiling and visualization.

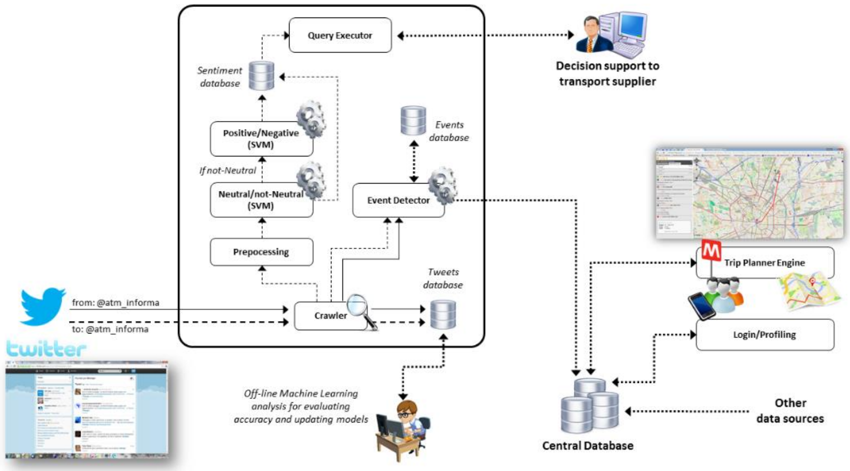


Fig. 1. Overall architecture of TAM-TAM with a focus on the tweets analyser component

Going more in detail, the main modules and functioning of the tweets analyser component are the following:

Crawler

Crawler is the module devoted to continuously collect tweets *from* (bold line) and *to* (dotted line) the Twitter account of the urban transportation company in Milan (*@atm_informa*). Moreover, Crawler is also devoted to store the acquired tweets, according to the data model provided by Twitter API, into a MySQL database (*Tweets database*) which is then used to perform further (off-line) analysis aimed at validating new machine learning algorithms and mine new models.

Event Detector

Event Detector implements a simple word-matching algorithm in order to identify, within tweets, keywords associated to relevant events. The set of keywords is based on the set of "standard" words generally used by ATM to inform customers about relevant events (e.g. strokes, accidents, interruptions, deviations, etc.) but it is completely customizable. The same set of keywords is also used to detect potential events communicated by the commuters. Finally, Event Detector search for other words, and their synonyms, referred to: type of transportation (i.e. bus, tram or underground), specific line and, where available, direction. All this information is well defined in the lists which could be retrieved from the web site of ATM.

While the events detected in tweets from *@atm_informa* are certain, the events detected in commuters-generated tweets have to be validated; this action is performed by considering the rate of tweets related to the same events in the last 15 minutes. Higher the rate higher the trustworthiness about the event; when the rate of an event becomes

0 the event is no more valid. This is very important because commuters are generally used to report and share information about events but not about the return to the normality, while transportation supplier communicates disruptive events as well as their rehabilitation.

Events are internally stored into the *Events database* – according to a structured format (i.e. type of event, type of transportation option, specific line, direction, timestamp, number of tweets in the last 15 minutes) – in order to perform all the rate-based considerations; subsequently, the Event Detector updates events within the TAM-TAM's *Central Database*, eventually modifying the number of related tweets in the last 15 minutes of a specific event or removing those which are no more "active" (rate=0). Continuously, the data in the *Central Database* are retrieved by other computational modules, in particular the trip planning applications in order to optimize trip according to the current situation on the urban transportation network (i.e. delays, events on a specific line, etc.)

Sentiment Analyser (pre-processing and neutral-positive-negative classification)

As the tweets published by the transportation company are only related to official communications and responses to requests by the commuters, they are not analysed for sentiment analysis. This is the reason why only dotted line goes through the corresponding computational modules (Figure 1).

The detection and further evaluation of possible sentiment in the tweets shared by commuters are performed through different sequential steps. First of all, some pre-processing is performed to transform the tweet in a vector of valued-features which can be analysed through Machine Learning algorithms. This pre-processing consists in removing stop-words (i.e. articles, prepositions and punctuations). Although the authors are conscious that emoticons may be used to enforce effectiveness of sentiment mining [8], in this first prototype of the tweets analyser they are not considered. Furthermore, the impact of applying – or not – stemming has also been considered, by using Snowball Stemmer (http://trimc-nlp.blogspot.it/2013/08/snowball-stemmer-for-java.html). The following pre-processing step consists in transforming the filtered tweets in a vector of valued-features. This procedure is better defined in section 3. It basically consists in a variation of a variation of the well-known TF-IDF (Term-Frequency – Inverse Document Frequency) weighting scheme, where features are computed differentially for each classification task (i.e. neutral vs not-neutral and positive vs negative). Similar tweets acquisition and analysis systems have been recently proposed, more specifically for English language, and for general purposes [9-13] as well for urban mobility [14].

After the current tweet has been pre-processed, only features related to TF-IDF for *Neutral vs Not-neutral* classification are given as input for a trained Support Vector Machine (SVM) classifier (details about the SVM learning are provided in section 3). The proposed classification output is then stored into the *Sentiment database*; in the case the output is "not-neutral" the values of the TF-IDF features for *Positive vs Negative*, of the current tweet, are given as input to a further SVM classifier, specifically trained. As in the previous step, the classification output is stored into the *Sentiment database* in order to enable, through the *Query Executor* module, the retrieval of useful information to support the transportation company in making decisions aimed at increasing commuters' satisfaction.

3 Materials and Methods

Design and development of the tweets analyser of the TAM-TAM platform initially required to collect a set of tweets to be used for the training and validation of sentiment mining classifiers. Collection was started on 12th June 2013 and is still in progress, for tweets posted both *from* and *to* the account of the public transport company in Milan (currently the collected tweets are around 45,000). A set of 1,332 collected tweets has been labelled by 3 different human supervisors according to the possible following three alternatives: *neutral* (570), *positive* (127) or *negative* (635). No specific training has been provided to the "labellers"; the set of tweets has been randomly given to each supervisor, separately, asking for a judgement about the sentiment. Mean Kappa statistics was 0.96, showing a high agreement among the labellers; final label of every tweet is the more frequent one ("neutral" is given in the case of 3 discordant labels).

To transform a tweet in a vector of features, the authors had taken into account specific considerations about the properties of tweets with respect to other types of text contents. As tweets are short messages, usually unstructured and informally written, techniques like parsing, pattern matching, complex grammars are usually ineffective. In [15] the solution proposed to analyse the content is a representation where features are terms and each feature is valued by the frequency of each term, which could be a word or *n*-gram. More simply, in [16, 17] the features are terms and they are valued as Boolean (1 if the term is present in the text, 0 otherwise). Other approaches propose a representation based on some computation; in [18] words are weighted by their correspondent Inverse Document Frequency (IDF) score, that is the logarithm of the number of documents in the collection divided by the number of documents containing a specific word [19]. Alternatively, the score known as Term Frequency–Inverse Document Frequency (TF-IDF) may be adopted, that is the IDF score multiplied by the frequency of a specific word divided by the number of words in the document [19]. In a recent study, proposing Bayesian Ensemble Learning for sentiment analysis, these approaches for feature construction are compared [20].

In [21] an extension of the TF-IDF approach is proposed, consisting in weighting words by the difference of their TF-IDF scores (delta TF-IDF) with respect to the class associated to the text (i.e. positive or negative sentiment). The Support Vector Machine (SVM) classification learning technique [22, 23] has been used to identify a reliable model able to detect the polarity of a document with respect to the computed delta-TF-IDF. In particular, the proposed delta TF-IDF is defined as follows:

$$V_{t,d} = C_{t,d} * log_2\left(\frac{|P|}{P_t}\right) - C_{t,d} * log_2\left(\frac{|N|}{N_t}\right) = C_{t,d} * log_2\left(\frac{|P|}{P_t}\frac{N_t}{|N|}\right)$$

where $V_{t,d}$ is the value of the term (feature) t in document d, $C_{t,d}$ is the frequency of term t in document d, P_t is the number of positively labelled documents containing term t, $|P|$ is the number of the positively labelled documents, N_t is the number of negatively labelled documents containing term t, $|N|$ is the number of negatively labelled documents. This approach proved to be more accurate with respect to the other ones and is the core of the application presented in this paper. In particular, two different delta TF-

IDF representations are computed, one for *Neutral vs Not-neutral* and one for *Positive vs Negative* classification, respectively.

The dataset of the 1,332 labelled tweets has been first divided into two different datasets, one related to tweets having neutral and not-neutral labels and one related to tweets having positive and negative labels. Then, tweets in each one of these datasets have been pre-processed, accordingly to the procedure described in previous section 2, and delta TF-IDF has been computed for each term. Using and not using stemming has been considered, thus two different datasets have been generated from each of the previous ones, characterized by a different set of features.

Furthermore, in order to reduce dimensionality, features have been ranked according to the corresponding delta TF-IDF and only the first *n* relevant features (terms) have been selected for each class (where *n* has been experimentally set to 10 in the case stemming is not adopted and 15 in the case of using stemming). Taking into account this step, the number of initially labelled tweets is reduced because some tweets could contain no one of the selected features. Table 1 summarizes the figures of each one of the datasets built starting from the initial set of the 1,332 labelled tweets. In order to use all the available data, the two classification learning tasks have been performed separately, while the two steps classification is only performed on new coming tweets when the module is deployed within the platform.

Therefore, the number of "positive vs negative" tweets does not add-up to "not-neutral" due to the different filtering performed, for instance: when the original 1,332 tweets are filtered according to the 10 most relevant features for "neural vs not-neutral" (no-Stemming case), 554 tweets (1,332-778) are removed because they do not contain any of the selected words. Similarly, 559 tweets are selected, among the 1,332 having a not-neutral, when the filtering (no-Stemming) is applied.

Table 1. Number of pre-processed tweets

	No Stemming	Stemming
Neutral vs Not-neutral	778 tweets 327 neutral tweets 451 not-neutral tweets	864 tweets 371 neutral tweets 493 not-neutral tweets
Positive vs Negative	559 tweets 115 positive tweets 444 negative tweets	588 tweets 120 positive tweets 468 negative tweets

4 Results

As first result, the list of terms ranked according to delta TF-IDF values is reported in Table 2, with respect to the classification tasks, with and without stemming.

Table 2. List of terms ranked according to delta TF-IDF values

Neutral vs Not-neutral without stemming	Positive vs Negative without stemming	Neutral vs Not-neutral with stemming	Positive vs Negative with stemming
Neutral	***Positive***	***Neutral***	***Positive***
regolare (0.0586)	grazie (0.7732)	line (0.0815)	graz (0.7716)
linea (0.0533)	risposta (0.0504)	regol (0.0586)	buon (0.0787)
sciopero (0.0367)	lavoro (0.0394)	circol (0.0489)	rispost (0.0472)
tram (0.0337)	buon (0.0331)	sap (0.0375)	ottim (0.0425)
concerto (0.0321)	arrivato (0.0283)	regolar (0.0374)	info (0.0268)
domani (0.0313)	info (0.0268)	direzion (0.0367)	molt (0.0252)
direzione (0.0296)	1000 (0.0220)	tram (0.0355)	1000 (0.0220)
linee (0.0282)	molto (0.0189)	sar (0.0352)	compl (0.0173)
circolazione (0.0273)	complimenti (0.0173)	funzion (0.0344)	avre (0.0142)
sapere (0.0243)	avrei (0.0142)	concert (0.0321)	ripart (0.0126)
	point (0.0126)		qualcun (0.0110)
	qualcuno (0.0110)		arriv (0.0094)
	dovrebbe (0.0094)		serv (0.0079)
	piedi (0.0079)		attent (0.0063)
	attenti (0.0063)		almen (0.0047)
Not-neutral	***Negative***	***Not-neutral***	***Negative***
aria (-0.1246)	aria (-0.1559)	ari (-0.1246)	ari (-0.1559)
condizionata (-0.1120)	condizionata (-0.1386)	condizion (-0.1194)	condiz (-0.1496)
minuti (-0.0807)	minuti (-0.1386)	min (-0.081)	min (-0.1338)
estivo (-0.0476)	sono (-0.1118)	estiv (-0.0594)	metr (-0.1149)
attesa (-0.0453)	metro (-0.1071)	tren (-0.0579)	son (-0.1118)
orario (-0.0413)	linea (-0.0882)	attes (-0.0515)	tren (-0.0992)
sempre (-0.0363)	attesa (-0.0724)	lavor (-0.0463)	line (-0.0945)
senza (-0.0349)	estivo (-0.0646)	aspett (-0.0388)	orar (-0.0819)
metro (-0.0337)	orario (-0.0614)	metr (-0.0368)	estiv (-0.0693)
treni (-0.0327)	anche (-0.0504)	sempr (-0.0363)	aspett (-0.0677)
	treni (-0.0551)		perc (-0.0646)
	senza (-0.0504)		anche (-0.0598)
	come (-0.0457)		ferm (-0.0551)
	treno (-0.0441)		senz (-0.0504)
	ogni (-0.0362)		mezz (-0.0472)

With respect to the classification learning task, a combination between the SVM implementation provided by WEKA suite (Waikato Environment for Knowledge Analysis, http://www.cs.waikato.ac.nz/ml/index.html) and Genetic Algorithms – aimed to optimize SVM configuration (regularization C and γ of Radial Basis Function Kernel) – has been used [24-26].

As the classes are unbalanced, the Balanced Classification Accuracy and F-score have been used to select the best performing SVM classifier according to a 10 fold-cross validation procedure. Furthermore, SVM has been also compared to other classification learning algorithms offered by the WEKA suite, in particular the ZeroR classifier, which classify any instance as belonging to the most frequent class in the dataset (baseline), Artificial Neural Network (RBF-Network and Multi-Layer Perceptron, MLP) and Naïve Bayes. Table 3 summarizes the obtained results.

Table 3. Balanced Accuracy and F-scorethrough 10 fold-cross validation

	ZeroR	**SVM**	**RBF-Network**	**MLP**	**Naïve Bayes**
Neutral vs Not-neutral (without stemming)	50.00% / 0.425	79.07% / 0.795	78.33% / 0.789	73.84% / 0.748	78.08% / 0.782
Positive vs Negative (without stemming)	50.00% / 0.703	94.29% / 0.956	86.81% / 0.900	93.51% / 0.949	93.27% / 0.939
Neutral vs Not-neutral (with stemming)	50.00% / 0.415	78.53% / 0.783	76.42% / 0.775	75.71% / 0.765	77.56% / 0.778
Positive vs Negative (with stemming)	50.00% / 0.705	94.32% / 0.950	90.95% / 0.922	92.02% / 0.938	92.33% / 0.938

Balanced Accuracy and F-score are almost similar across the different classification learning algorithms and higher than baseline. According to the definition of BAC (i.e. average between sensitivity and specificity), its value is always 50% for the ZeroR and only F-score varies. SVM proved to be the most performing classification learning strategy, however, some differences resulted among the available datasets: in particular performances are higher in the case of *Positive vs Negative* classification than *Neutral vs Not-neutral* classification, while stemming does not make any difference in *Neutral vs Not-neutral* as well as *Positive vs Negative* classification.

As final decision, stemming has been adopted for "*Neutral vs Not-neutral*" classification but not for "*Positive vs Negative*" classification. Therefore, in the pre-processing step every tweet generates two different vectors: the first (stemmed) is the input of "*Neutral vs Not-neutral*" classification, while the second (not-stemmed) is the input of "*Positive vs Negative*" classification, if and only if it is classified as "*Not-neutral*" at the first step.

In the following Table 4 the SVM configurations associated to the performances in previous Table 2 are reported, along with the number percentage of overall instances used as Support Vectors (%SVs). This is another important index for evaluating the capability for any SVM classifier to correctly classify new instances not used for learning. It is easy to note that, according to both Balanced Accuracy and %SVs, the *Neutral vs Not-neutral* classification is more difficult than *Positive vs Negative* classification.

Table 4. Configuration of the SVM classifiers and number of Support Vectors (SVs) used

	C	γ	%SVs
Neutral vs Not-neutral (without stemming)	10	110.95	73%
Positive vs Negative (without stemming)	10	105.26	39%
Neutral vs Not-neutral (with stemming)	10	8.46	63%
Positive vs Negative (with stemming)	10	11.08	28%

5 Conclusions

The developed tweets analyser module, based on text mining and SVM classification and deployed into the prototype of the TAM-TAM platform, enabled innovative added-value services for commuters, aimed at improving urban mobility in the city of Milan. While event detection is used to optimize trip planning, sentiment analysis is currently more devoted to support transportation supplier in addressing commuters' needs and improve their satisfaction. On the other hand, the idea is to use the output of sentiment analysis according to a *collective intelligence* paradigm by providing also commuters with information about the perceived quality of transportation service, and specific mobility options, as spontaneously reported by the other commuters. This will allow users of the transportation service, citizens as well as tourists, to plan their trips by also considering some social indicators of satisfaction.

Currently the most relevant limitations of the work are two: the solution strictly depends on language as it has been currently validated only on Italian and the limited dataset of labelled tweets. While the first limitation is not yet so relevant, since almost all the tweets *from* and *to @atm_informa* are written in Italian, the second could be the reason of lower accuracy in the *Neutral vs Not-neutral* classification. *Gamification* based apps, aimed at enabling labelling by TAM-TAM users, have already been identified as effective solutions for increasing both the number of labelled tweets over time and labels objectivity according to the judgements provided by multiple users.

6 References

1. Srivastava, L., Vakali, A.: Towards a narrative aware design framework for smart urban environment. F. Álvarez et al. (Eds.): FIA 2012, LNCS 7281, 166–177 (2012).
2. Candelieri, A., Archetti, F., Giordani, I., Arosio, G., Sormani, R.: Smart cities management by integrating sensors, models and user generated contents. WIT Transactions on Ecology and the Environment, 179(1), 719-730 (2014).
3. Thom, D., Bosch, H., Koch, S., Worner, M., Ertl, T.: Spatiotemporal Anomaly Detection through Visual Analysis of Geolocated. IEEE Pacific Visualization Symposium (2012).
4. Burnap, P., Rana, O.F., Avis, N., Williams, M., Housley, W., Edwards, A., Morgan, J., Sloan, L.: Detecting tension in online communities with computational Twitter analysis. Technological Forecasting and Social Change (2013).
5. Burnap, P., Williams, M. L., Sloan, L., Rana, O. F., Housley, W., Edwards, A. M., Knight, V. A., Morgan, J., Procter, R., Voss, A.: Tweeting the terror: modelling the social media reaction to the Woolwich terrorist attack. Social Network Analysis and Mining (2014)
6. Pang, B., Lee, L.: Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1, 2),1–135 (2008).
7. Candelieri, A., Archetti, F.: Analyzing tweets to enable sustainable, multi-modal and personalized urban mobility: Approaches and results from the Italian project TAM-TAM. WIT Transactions on the Built Environment, 138, 373-379 (2014).
8. Pozzi, F.A., Maccagnola, D., Fersini, E., Messina, E.: Enhance user-level Sentiment Analysis on microblogs with approval relations. Proceeding of the 13th International Conference on Advances in Artificial Intelligence (2013).

9. Burnap, P., Rana, O., Williams, M.L., Housley, W., Edwards, A., Morgan, J., Sloan, L., Conejero, J.: COSMOS: Towards an integrated and scalable service for analysing social media on demand. Intern. Journal of Parallel, Emergent and Distributed Systems (2014)
10. Amati, G., Bianchi, M., Marcone, G.: Sentiment Estimation on Twitter (http://ceur-ws.org/Vol-1127/paper7.pdf)
11. Musto, C., Semeraro, G., Lops, P., de Gemmis, M., Narducci, F., Bordoni, L., Annunziato, M., Meloni, C., Orsucci, F. F., Paoloni, G.: Developing a Semantic Content Analyzer for L'Aquila Social Urban Network (http://ceur-ws.org/Vol-1127/paper6.pdf)
12. Amati, G., Angelini, S., Bianchi, M., Costantini, L., Marcone, G.: A scalable approach to near real-time sentiment analysis on social networks. In DART 2014, Information Filtering and Retrieval. Proceedings of the 8th International Workshop on Information Filtering and Retrieval, co-located with XIII AI*IA Symposium on Artificial Intelligence (AI*IA 2014), Pisa, Italy, December 10, 2014. CEUR Workshop Proceedings, 1314, 12-23 (2014).
13. Musto, C., Semeraro, G., Polignano, M.: A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. In DART 2014, Information Filtering and Retrieval. Proceedings of the 8th International Workshop on Information Filtering and Retrieval, co-located with XIII AI*IA Symposium on Artificial Intelligence (AI*IA 2014), Pisa, Italy, December 10, 2014. CEUR Workshop Proceedings, 1314, 59-68, (2014).
14. Chen, F., Krishnan, R.: Transportation Sentiment Analysis for Safety Enhancement, Final Project Report. Technologies for Safe and Efficient Transportation, Carnegie Mellon University (2013).
15. Joachims, T.: Text Categorization with Support Vector Machines: Learning with Many Relevant Features, Springer (1997).
16. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of EMNLP (2002).
17. Whitelaw, C., Garg, N., Argamon, S.: Using appraisal groups for sentiment analysis. In Proceedings of the 14th ACM International Conference on Information and Knowledge Management, 625-631 (2005).
18. Kim, S., Pantel, P., Chklovski, T., Pennacchiotti, M.: Automatically assessing review helpfulness. In Proceedings of EMNLP, 423-430 (2006).
19. Baeza-Yates, R.A..: Modern Information Retrieval. Addison-Wesley Longman Publishing Co. (1999).
20. Fersini, E., Messina, E., Pozzi, F.A.: Sentiment analysis: Bayesian Ensemble Learning. Decision Support Systems, 68, 26-38 (2014).
21. Martineau, J., Finin, T.: Delta TFIDF: An Improved Feature Space for Sentiment Analysis. In Proceedings of the Third International ICWSM Conference, 258-261 (2009).
22. Scholkopf, B., Smola, A. J.: Learning with kernels. Support Vector Machines, regularization, optimization and beyond. Massachussetts Institute of Technology, USA (2002).
23. Vapnik, V.: Statistical Learning Theory. New York, Wiley (1998).
24. Candelieri, A.: A hyper-solution framework for classification problems via metaheuristic approaches. 4OR, 9 (4), 425-428 (2011).
25. Candelieri A, Conforti D.: A Hyper-Solution Framework for SVM Classification: Application for Predicting Destabilizations in Chronic Heart Failure Patients. The Open Medical Informatics Journal, 4, 136-140 (2010).
26. Candelieri, A., Sormani, R., Arosio, G., Giordani, I., Archetti. F.: A Hyper-solution Framework for SVM Classification: Improving Damage Detection on Helicopter Fuselage Panels. ASRI 2013, Conf. on Intelligent Systems and Control. AASRI Procedia 4, 31–36 (2013)

When minorities' group discussions in social media become a resilient strategy.

Francesca D’Errico*, Isabella Poggi** and Rocco Corriero***

**Uninettuno University, Psychology Faculty, Italy*
***Roma Tre University, FilCoSpe Department, Italy*
****Altran Consulting, Rome*

Abstract. The paper presents a brief overview of how social media can influence minority interactions by focusing at their features. Starting from the notion of “active minority” (Moscovici, 1981) the aim of the study is to understand how participants of a social mediated community face a critical event by focusing at socio-psychological dimensions that affect their “on line resilience” and how they promote their “social empowerment”.
At this purpose was performed a quanti-qualitative analysis - lexicographic and content analysis by means of a software called *Taltac* (Bolasco, 2013). Results pointed out different socio-psychological processes (self-definition, trust, emotions and values) before and after critical situation emerging by on line discussions.

Keywords. Active minorities, resilience, political participation, social media

1 Introduction

The link between lack of power and personal resilience in everyday life has been explored in various social psychology perspectives [1]. The common determinants of a positive and participative approach to personal and social lack of power is what Freire defined “process of conscientisation” as development of critical thinking constructed by sharing common ideas, practice, and knowledge within a community. Belonging to a community, sharing knowledge and arguments – mostly in a context of lack of information – can give the opportunity to perceive a sense of control on the events and shared problems and thus it can be a source of personal resilience [2].

Within this framework the present work explores what means to be part of a minority group built through a social media (in particular, a Facebook group) and what discourses and argumentations are used to face a critical phase in which participants are involved.

The psico-social notions that contribute to understand this process are the construct of *active minority* [3] and its relation with *social trust* [4].

To be part of a minority group means to be part not only of a social minority in quantitative terms but also to have a marginal status and lack of power [5]; but a minority group can actually exert some influence on the majority, though in an indirect and slow way, because, just thanks to its marginal status, it can adopt an autonomous and divergent thinking useful to evaluate, judge and discuss the majority’s choices.

In the influence process minorities have to assume coherent, autonomous and egalitarian behavioural styles; but what are the discourses they construct to solve problems and at the same time to resist and to face critical events?

What are the characteristics of minority thinking? What arguments do people in a minority group use, and how do they construct a strategy to induce an opinion change in their own group or outside of it?

First of all we need to define and distinguish two different types of minority groups, one belonging as social category to the majority but having less power or a different opinion on a given topic (ingroup), and the other belonging to a different social category (outgroup).

This difference looks useful to understand what is the role of trust that members of minorities assume toward different institutions and what are the social organizations they trust in in while looking for possible strategies to face negative events.

2 Social mediated communities and their features

Social media and social networking seems to be a new way to construct meanings and to build new forms of relationships away of or in support to "real life". From this perspective social media can be used as the main or unique tool to be in touch by means of what we can name "social mediated communities": a virtual togetherness of members that produce discussions and proposals though not in physical presence or by regular contacts but sharing a social category [6] and/or place identity [7].

In this sense an interesting dynamics can develop when participants do not personally know each other but they share a common cause or make part of a same social category – like in the case study presented below, of a Facebook group of young university researchers. In this case researchers inform each other by sharing problems in their virtual group, because they acknowledge a competence to its participants, but maybe also because they don't feel confident to their respective academic real contexts.

Thus, social mediated communities in this case can be a good way to exchange information and support during a critical and problematic phase for their work-life that the researchers face everyday, as in the case of a national competition to access a university job (see below).

One important feature of group discussion is its structure: either formal or informal; formal when groups correspond to acknowledged institutions like political parties or formal associations; informal when outside the social network the group is not acknowledged as such. In this case the network can be the expression of a spontaneous majority or minority group joined by common interests or belonging to a real or virtual social group. Another feature is also relevant: an either vertical or horizontal structure. The group structure is vertical when it contains opinion leaders that regularly animate a discussion by links or posts, and horizontal when participants democratically and actively intervene. Lortie-Lussier [8] stated that a minority group with a leader is more influential, but this becomes a research question when groups live in a cyber-place.

3 A case study: The Roars (Return on academic research)

Within Italian institutions one of the most conflictual debates concerns University. In 2013, after a reform of the University proposed in 2011 by the right-wing Italian Ministry of Instruction Gelmini, professors and universities were evaluated by an independent agency called ANVUR (Evaluation National Agency of University and Research), passing from a local evaluation with single competitions between candidates to a professor role-crossing in a national evaluation process of "professor eligibility". The first round of evaluation started in February 2013, an after one year at least part of the evaluators' committees – in December 2013 – finished their work and published the lists of eligible and non-eligible candidates, with their corresponding evaluations.

The Ministry of Instruction and University decided that the eligibility would follow numeric criteria obtained by reference to the so called "medians" obtained considering the number of publications with the impact factor and citations by the scientific community. (Only later, when a new Minister came, Francesco Profumo, the more qualitative criteria were exploited).

In this confusing period – due to novelty and lack of information – within the very institution of the Anvur Agency an active group of researchers created a Facebook group and a website called "roars" (Return on academic research").

They do not know each other but their number grows up very fast and at the time of this work they count almost 5000 members (7 milions of views and 21.600 comments) who discuss together on the situation trying to find possible solutions together.

3.1. A textual and lexical analysis of Roars

Group discussions have been examined through two methods of analysis. On one side, they have gone through automatic lexicographic analysis by using TALTAC, a software for qualitative analysis. On the other side the lexicon obtained through quantitative measures was analyzed by means of "concordance analysis" that, like "traditional" discourse analysis, takes into account the text of all answers to draw recurrent concepts, topics or semantic areas from them.

The automatic quanti-qualitative analysis was performed on the subjects' discussions by *TalTac* ("Lexical and Textual Automatic Processing for Content Analysis": [9]), a software for textual data analysis based on a "lexicometric approach": an application of statistical principles to textual corpora. The "textual statistics" [10] aims to extract the semantic level in a text starting from the list of words obtained by statistical analysis; for example, in the specificities' analysis, the software extracts automatically a list of significant words obtained by a statistical comparison between subparts of text according to selected variables. From these statistically significant words we also extract also the list of the sentences where significant words appear in order to interpret and identify both contextualized meanings (looking also at "negations" or other unexpected sense of words) and relevant topics.

Here we describe topics and discussions that emerge from the Facebook group of "Roars" Italian researchers by focusing and detailing the used lexicon both in the first period (before the national evaluation results) and in the second period (after publication of first results) to identify the minority's discourses aimed at facing two critical periods, one characterized by lack of information and another by negative evaluations from commissions.

We extracted a corpus by means of Facebook Graph API (Application Program interface, for advanced users) and we obtained a *corpus* which counts 31948 (V) occurrences with 307682 (N) different words and a high lexical richness index [(V/N)*100], equal to 10,38%.

The lexical analysis includes some descriptive information, particularly interesting for the understanding of minority group discussions, like *adjective analysis* and *time analysis*.

3.2. Adjective analysis

We used the dictionary of positive and negative adjectives present in TalTac2 by analysing the negative index[1] to identify the polarization through a positive or negative lexicon. The index reveals that the corpus shows characteristics of negativity as high as 71%, that is higher than the reference value for Italia (which is 40% according to Bolasco [9]).

Looking in depth we see that negative adjectives are very often used in the first period concerning the Evaluation Process. We can recognize an adjectival use focalized on the first period on a sematic area of doubt and highly oriented toward a personalized criterion (*questionable discretion, objectionable, poor, questionable, unproductive, smoky, doubtful, bad*);

on the other hand the second period that corresponds to the first publication of concrete results focuses on the emotional use of adjectives revealing a sense of unfulfilled expectations, like *scandalous, disconcerting but also shameful, indecent, disgusting, detrimental, pernicious, laughable, mediocre*.

Moreover, the second period shows a very high frequency of the adjective "pertinent", that is used by the committees as the most typical "cause" of negative evaluations. (Actually, one of the most used motivations mentioned by evaluators to deny eligibility is that a candidate's publications, notwithstanding their high number and quality, show a lack of *pertinence* to the core of the discipline).

3.3. Text Imprinting: Time, mode and person analysis

Time analysis reveals a strong *orientation to the present,* because out of all verb frequencies our subjects express time information most frequently in present tense (82%, as opposed to 11% past and 7% future) mostly in the first period, while in the second, future time slightly increases (+2%): which in a certain sense can be seen as determination in planning. The most frequent mode is - as predictable - the indicative; but also a large use of conjunctive and conditional can be found (79% indicative vs 11% conditional and 10% conjunctive), which is quite high compared to general Italian use, with conditional and conjunctive around 4% and 7% respectively. This might

[1] The index is obtained by extracting all the positive and negative adjectives by means of Taltac dictionary of adjective and then by calculating the ratio between the total of the negative occurrences and the total of the positive ones (tot. Occ. Pos/tot. Occ. Pos*100).

be accounted for by participants' high level of education that allows the use of complex forms (i.e. hypothetical constructions), but it can be also an index of *uncertainty*; researchers in this context argue in search of strategies and possible solutions in a confused context.

Another characteristic is the conjugation extremely oriented toward the third person (69%) compared to first (26%) and second (5%); what we expected was – given the dialogical form of discussions – an overuse of the first and the second person; on the contrary the high frequency of third person might be due to a sort of *contraposition between "me-us" and "they"* (researchers vs evaluators), this tendency is present in the whole Roars's discussions but increases significantly in the second period.

3.4. The peculiar and characteristic lexicon of researchers as a minority group

Beside the absolute value of words, the *key words* or *peculiar lexicon* [9] are the over-represented words in the text created by a comparison between the corpus and the external lexicons of frequency, taken as reference model.[2] The measure of the variance from the reference lexicon is represented by the *square deviation*[3].

If we group the overrepresented vocabulary we find three main recurrent topics that we can phrase as follows:

1. *The evaluation process, its participants and its effects*
2. *The researchers' reference values: political and economic processes underlying the evaluation process*
3. *Participation and common proposal*

1. *The evaluation process, its participants and its effects*

This first topic spreads across all participant and all technical aspects of the evaluation:

(reported in a decreasing order, all up to 3,95; p<0.05) *qualified, certification, commissioners, universities, enabled, researchers, competitions, university, teaching, PhD, commissions, full professor, publications, evaluations, median, privacy, associate, professor, monographs, band, enabled, recruitment, candidate, monograph, bibliometrics , article, assessment, commission, Commission, rewarding, to enable, casual, relevant, candidate, round, quote, magazines, policy, judgment, publish, publishing, co-optation, meritocracy, curricula, quotes, results, areas, suitable rejected, suitable, economists, humanistic, accreditation, rankings, merit, discipline, parameters, indicators,*

[2] In this case we used the *stardard Italian*, resource in Taltac.

[3] *standard quadratic deviation*: considers significant words overused compared to the lexicon of reference, we then consider the forms with greater deviation of 3.84 which is the reference value of a chi square with one degree of freedom (p <0.05).

argue, excellence, merit, selection, progressions, bureaucracy, ministry, errors, extractions, judge, questionable criteria

Within this topic a very negative evaluation of evaluators emerges: in terms of types of discrediting criteria [11] they are attacked as to their competence (ignorant) and honesty (*clubs, nepotism, barons*), finally being described as a power out of the law (*illegitimate*) and out of control (*insanity, delirium, boycott*, *rubbish*): *recommended, shame, ridicule, crap, baron, distortions, sawn, nonsense, rubbish unlawful, scandalous, madness, delirium, clubs, fallacious, nepotism, trumpet, questionable, rag, decency, crap, boycott, shameful, embarrassing, ignorant.*

1. *The researchers' reference values: political and economic processes underlying the evaluation process*

The second topic covers the ***value dimension*,** with researchers widening their discourses to the political and economic field.

The term *politics* for example is a key term to understand the level of trust [4] in political and economic institutions. The term, that is overfrequent in the first period (110 vs 58) corresponds to "*bad politics*" (cattiva politica): a type of discredit that is mostly oriented on the competence and honesty dimension: "*ignorant and corrupted*", "*incapable, corrupted and ignorant*" [11]. Politicians seem to be the cause of the present economic crisis and researchers – as is reported in the concordances below - ask for a "political reinassance" (*a new political framework, political path change, to take universities back to zero, ethical path, political renovation)* as far as politicians did not invest economical resources in the academic context and expressed their willingness only as "an interested club" on the basis of "opportunistic reasons" .In the second period the debate on "*politics*" decreases (110 f. vs 58 f.), politics is only mentioned as a site in which to ask for explanations and economic resources in public context ("parliamentary explanation", "resources", "political and ethical choice") or becomes something to overcome by means of "justice"; moreover researchers now describe the executive side, represented by one more Minister of University, Maria Chiara Carrozza, as a powerless minister or sometimes as an absent interlocutor ("Someone has heard Maria Chiara Carrozza?") remarking a lack of dialogue between researchers, professors and government.

Selected Concordance of "Politica" (period 1: freq. 110)

- *always for opportunistic reasons , the* ***political*** *class began to ride . We must instead get through this*
- *This is not a virus, but a precise* ***political*** *will, a specific project shared cultural and stubbornly pursued by a club.*
- *but the measure in question is another example of* ***bad policy*** *and bad politics , we know , only combines disasters. . .*
- *The blame for this economic crisis , ethics and morality is an ignorant and corrupt*

***politicians** who shamelessly plundered the hopes of young people*

- *I expect, however, that will be conquered by force (see actions) rather than with **politics***
- *But I do not seem to remember any **political** force that has never explicitly raised the question of which model of nation we want*
- *I would like this question, which is not technical but **political** , came out from the corridors of becoming part of the political debate*
- *The problem is that when a question is considered "technical" **policy** stop to discuss*
- *only if there is a significant **policy** shift . When we become a decent country **politically** , I said decent , not perfect.*

Selected Concordance of " Carrozza " (period 2: freq . 39)

- *The answer is not there, **Carrozza** is showing as a great **political** with a small p .*
- ***Carrozza** is there by virtue of the Holy Spirit and knows neither the school nor the university 's problems*
- *From **Carrozza** never came a word. . . Poor things . . . And what figures are you talking about? Let's organize together*
- *Can you see the Minister **Carrozza** who " abolished " a procedure in the race ? :)*
- *I'm sorry, but I think the response of minister Carrozza is in many ways exemplary and necessary*
- ***Carrozza** must stop talking nonsenses*

The description of this more systemic and general discussion comes from the peculiar lexicon with the overrepresented words (reported below).

(reported in a decreasing order, all up to 3,95; p<0.05) : *Research, Euro, European funds, salary, incentives, budget, salary, economists, cuts, gross loans, Europeans, money, shots wage, virtuous, financing, Italy, culture, quality, university, history, respect, scientific merit, science, person, future development, the U.S., background, announcements, scientists, international, internationalization, standards, republic, government, accountability, justice, innovation, proposals, development, expertise, regulatory, rector, opportunities, needs, nobel, initiatives resources*

3. Participation and common proposal

The third topic is the "***researchers' participation and collective-cooperative activism***".

Recurring terms are linked to *petition, rights* (ricorso, diritti) and shared information to face a *common* problem or strategy and possible *future* solution taken together (we-ness).

Actually recurrent verbs are ones of doing, changing, working, (fattivi, volizione, potere) especially in the form of second plural person (*we can, we stay, we want, we must, we do*).

Within this topic we can include words close to the idea of things to do together in order to react and to be protected from injustice (up to 3,95 square deviation) like *possibly, hope, example, can, could, may, plan, possibility, partici-*

pate, can, I can, discussion, work, best, you, join, proposal, community, work, joint strength, our, us, network, compare, decide, change, change, future.

This area stands for the idea of resorting to a social resource to face a negative evaluation in a hierarchical and competitive context. In this connection, arguing on the web seems to be a way to give support and share information and at the same time it creates a secure place to discuss and find an exit strategy. Key words to understand this central topic are in this case rights, justice (as "super-partes" institution) and juridical words, seen as tools to defend themselves. (*law, human rights, justice, judiciary, self-defense, rules, extension, rules, terms, rules, just, recourse*)

Selected Concordance of "possiamo" (period 1: freq. 39)

- *we want and* ***we can*** *support our desire for an efficient country*
- *Then they win grants because they are good,* ***we can*** *do with their network. . . however, the theme of 'article linked towards resolving the dispute*
- *Then* ***we dissent*** *from his work as a minister (and I strongly disagree),*

Selected Concordance of "possiamo" (period 2: freq. 54)

- ***We can*** *no longer sit idly by and just complain amongst ourselves otherwise we will be accomplices in this disaster*
- *Yes, we need to move,* ***we can*** *not suffer a judgment that will have much influence on our lives. The important thing is to* ***participate.***
- ***We can****, however, recourse of and make use of our rights*
- *only* ***we can*** *find arguments to convince the 'public opinion' s investment in research is useful and far-sighted*
- *simply share what I figured out why I think so* ***we can*** *clarify each other's ideas.*

3.5 Verb lemmas

The characteristic analysis of verb lemmas (with square deviation starting from 60,00; while the reference point is 3,95), i.e. all verbs converted in their infinitive forms, highlights four types of verbs. The first one, that we can name verbs of **"*participation and social activism*",** goes from concrete actions to acts of group coordination and action planning. This type increases in the second period, after publication of the first results (s.d. > 3,95; agentic verbs: *power, succeed, find, finance, change, grow, participate, decide, earn, win, think, criticize, share, fix, work, show, get, take, raise, reward, predict, test, enable, risk, protest, overcome, invoke, invest, organize, enhance, reject, groped, challenge, boycott, create, propose, betting, solve, encourage, defend, prepare, rebalance, reform, excel, advertise, charge, rethink, take, fight, convince, fix, hold, secure, contribute, equate, try again, cooperate, prevent, respond to, influence, compel, start over, create, expand, promote, renew, resist, aspire, grow, to dream, to associate, to dare, enhance, fruit, flip, recover, trigger, coordinate, pursue, rebuild, program, address*).

Another interesting topic concerns verbs of on line group interaction, indicating that one of the main goals of the "ROARS" group is ***"reasoning"*** and arguing together:

Sometimes, for this minority, arguing is an instrument for the interpretation of acts or laws and for reasoning together on the future, on the basis, usually, of posted articles, links or news.

(s.d. > 3,95) *Understand, evaluate, discuss, click, refer, extrapolate, feel, add, to argue, to consider, recommend, center, disagree, implement, clarify, argue, protest, agree, infer, to reiterate, comment, confirm, restate, reason, verify, doubt, imply, specify, reflect, stress, indicate, infer, say, correspond, establish focus, to infer, to emerge, to intervene, to allow, to mention, simplify, differentiate, declare, issue, focus, excuse, introduce, monitor, communicate, examine, to assert, to aim.*

Another useful function of online interaction in this group, even if less represented compared to the other two, is the **"expression of emotions"**, especially negative emotions that are primarily acknowledgeable by the negative effects both on the economic and on the self-image side, and more in general on the researcher's life.

Quite representative of a first type of expressive verbs are those mentioning negative actions and their effects on life and emotions.

(with square deviation starting from 15,00; while the reference point is 3,95): *Displease, fear, err, block, complain, waste, scream, regret, shame, suicide, insult, scuttle, beat, abuse, spitting, damage, freeze, hit, brandish, point, demolish, slay, crazy, break, fumble, pick up, prejudice, cringe, break, subdue, aging, affect, scrounge up, derail, deter, insult, graze, vomiting, mock, live*

Positive emotions are less frequent; in the first period we find verbs like "hope", *auspicare*, in relation to positive expectations, but in the second period the coexistence of verbs like "*promote, enforce, continue to, thank, help, share, suggest, support, be grateful, agree, resolve*" highlights that positive emotion are only felt in presence of the potential support of the group composed by other active researchers in the same critical period.

This function of the Roars Group is evident when we extract "entities"[4] or complex expressions (marked in bold) in the corpus that are clear requests for help, usually followed by a very large variety of answers:

"Guys, I have a problem and I would like to **share** it here **hoping** that someone would **suggest** a way to **solve** it. (...) Has anyone had similar experiences and can **help** me? "

[4] Fragments Obtained through semantic concordance by means of so called *Entity research* (Taltac allows to extract phrases or fragments which present sequences or semi-sequences with two or more words that represent complex queries; Bolasco, 2013):

4 Conclusion

One aim of this study is highlighting how social media can be a tool for active minority groups to overcome some critical situations. Social networking is a place where people can share information, discuss on common problems and/or on political and economic situations, find shared strategies to adopt and finally develop group discourses and personal narratives. In minority groups arguing on line seems to correspond to a way to develop a group thinking by knowing the other member's point of view, by analysing together either macro-themes (political and economic situation) or micro-experience (personal experiences like rejection from the national evaluation, legal action). In this sense discourses can be seen as an active participation but also – beyond the institutional contexts where they live – they can facilitate the so called "process of conscientization" and critical consciousness [12] at an individual and group level. Social media – here, in "Roars", the social mediated communities - are a space where discussions are freely faced by constructing a "social knowledge" favoured by the fact that they can "de-traditionalise the public sphere" allowing the coexistence of different points of view and dissenting voices and opinions by means of horizontality and egalitarian access [1].

The work presented is based on the Italian group of researchers "Roars" before and after a national evaluation process aimed at their possible enrolment as professor within Italian University.

Roars, as a minority group lacking any consulting or decisional power, before the evaluation results discuss all possible scenarios in absence of institutional information, especially analysing the political and economic present and past situation. After the extraction of peculiar and characteristic lexicon [9] the concordance analysis points out that this political tendency of discussion becomes a request to be listen to by the Minister, who is felt as distant and absent ("someone has heard Carrozza?"), and a more general distrust emerges toward a political class described as "ignorant, unqualified and corrupt".

This lack of trust toward the political class and institutions [4] seems to be one motivation that lead Roars members to be active within the group by having peer discussions; in their "participation and common proposal" we can find a sense of trust in the belonging to a larger group, we find a "we" often opposed to "they". Within this topic the juridical part, with all the possible tools usually thought and constructed together, seems to be the "neutral" institution to turn to face personal and group injustices.

This exploratory analysis is a way to understand psychosocial processes within minority group, with a particular attention to social trust. It helps us to better understand trust dynamics by means of social media discussions: we have seen that in critical phases if there is a lack of trust toward "political" institutions, minorities discuss for a common solution or for a guarantee which helps them as super-partes institution (i.e. the case of justice). This can be a resilient strategy for a minority, but further studies can go more in depth and analyse further acts of "social commitment" [4] to overcome negative and uncertain phase.

From the use of social media some methodological issues also emerge concerning the analysis of argumentation on line. The retrieval of on line discussions give the researcher a very large material to analyse and to understand, sometimes in a real-time situation. So before analysing in depth this big data and material obtained from social media a possible route to better understand and to identify argumentations in discussions is to identify main topics by means of a lexicographic approach [9; 10] through *imprinting* analysis of lexicon (time, mode and person analysis by extraction and frequencies aggregation), *peculiar* and *characteristic analysis* supported by *concordances* and *entity research* in the text.

From this methodological point of view, this study is a way to connect a lexicographic approach to the specific need of studying argumentation on line. Future works could explore the minority groups' argumentations used by their members to "embrace" each other from an instrumental and emotional side in a critical situation and in a position of lack of power.

References

1. C. Campbell, and S. Jovchelovitch. Health, community and development : towards a social psychology of participation. *Journal of Community and Applied Social Psychology*, 10, 4, 2000.
2. N. Garmezy. Resilience in children's adaptation to negative life events and stressed environments. *Pediatrics*. (20):459–466, 1991.
3. S. Moscovici. *Psicologia delle minoranze attive*. Boringhieri, Torino, 1981.
4. C. Castelfranchi. La paradossale «sfiducia» degli italiani nelle istituzioni. *Sistemi Intelligenti*, n.1. pp. 113-121, 2013.
5. A. Mucchi Faina, *L'influenza sociale*. Il Mulino: Bologna, 2013.
6. H. Tajfel, , & J. Turner. An integrative theory of intergroup conflict. In: Austin W. G. & Worchel S. (eds.), *The social psychology of intergroup relations* (pp. 33-48). Monterey, CA: Brooks/Cole, 1979.
7. J. Dixon, & K. Durrheim, Displacing place identity: a discursive approach to locating self and other. *British journal of social psychology* 39 (1), 27-44, 2000.
8. M. Lortie-Lussier, S. Lemieux,, L. Godbout,. Reports of a Public Manifestation: Their Impact According to Minority Influence Theory. *The Journal of Social Psychology*, 1989, Vol.129 (3), p.285-295, 1989.
9. S. Bolasco *L'analisi automatica dei testi. Fare ricerca con il text mining*, Carocci, Roma, 2013.
10. L. Lebart, A. Salem, *Statistique textuelle* , Dunod, Paris, 1994.
11. F. D'Errico, I. Poggi, , L. Vincze. Discrediting signals. A model of social e-valuation to study discrediting moves in political debates. *Journal on Multimodal User Interfaces*. Vol.6 (3-4), pp.163-178, 2012.
12. P. Friere. *Pedagogy of the oppressed*. New York, NY: Continuum Press, 1990.
13. H. N. Zuniga Jung S. Valenzuela Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation. *Journal of Computer-Mediated Communication* 17 (2012) 319–336, 2012.

Author Index