Symposium on Emotion Modelling and Detection in Social Media and Online Interaction

In conjunction with the 2018 Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour (AISB 2018)

5th April 2018

Preface

Emotions@AISB 2018: Symposium on Emotion Modeling and Detection in Social Media and Online Interaction

Message from the Chairs

Welcome to Emotions@AISB 2018, the Symposium on Emotion Modeling and Detection in Social Media and Online Interaction. The symposium is part of AISB 2018, the convention on Artificial Intelligence organized by the Society for the Study of Artificial Intelligence and Simulation of Behaviour. This symposium addresses the opportunities and challenges of emotion modeling and tools for online interaction. This interest is motivated by the worldwide diffusion of social media, which has profoundly changed the way we communicate and access information. Everyday, people interact with each other to share opinions about commercial products on dedicated platforms, report their personal experiences on microblogging and social networking sites, try to solve domain-specific problems through collaborative knowledge building and sharing in online question and answering.

On one hand, such a pervasive use of online social media in computer-mediated communication is opening new challenges for social sciences and human-computer studies. Indeed, one of the biggest drawbacks of communication through social media is to appropriately convey and recognize sentiment through text: while display rules for emotions exist and are widely accepted in traditional face-to-face interaction, people might not be prepared for effectively dealing with the barriers of social media to non-verbal communication. On the other hand, user-generated content comprises an invaluable wealth of data, ready to be mined for training predictive models. As such, microblogging and online interaction analysis are attracting the interest of researchers and practitioners in NLP, machine learning, big data analysis. Indeed, analyzing opinions and emotions conveyed by microposts can yield a competitive advantage for businesses, and can serve to gain crucial insights about political sentiment and election results or other social issues.

Emotions@AISB 2018 aims at fostering discussion around interdisciplinary research area at the intersection between cognitive sciences, computational linguistics, and social computing. We have invited three paper categories, full papers, short papers, and poster and demo papers, to encourage submissions of contributions describing different stages of research on affect recognition in social media, ethical concerns, and applications, with a special focus on education, entertainment, health, e-government, games, and hate speech monitoring. We are pleased to present a collection of five papers discussing theoretical models, empirical studies, and tools. All papers went through a thorough review process that involved at least three reviewers and were evaluated based on their originality, quality, and relevance to the workshop. Furthermore, we invited two key researchers with major contributions in this field to discuss their visions and share the state of their research with the community in form of a keynote: Diana Maynard (University of Sheffield, UK) delivering a talk on "Twits, Twats and Twaddle. Analysis of hate speech towards politicians in the GATE social media toolkit" and Pietro Cipresso (University of Milan Sacro Cuore', Italy) that will report on "The role of a virtual body in modeling emotions for Social Media and Online Interaction: The BodyPass project".

We thank the members of our program committee and additional reviewers for their constructive reviews: Alessandro Ansani, Ruth Aylett, Francesco Barbieri, Pierpaolo Basile, Valerio Basile, Erik Cambria, Chloé Clavel, Mihaela Cocea, Danilo Croce, Rossana Damiano, Celso De Melo, Anna Esposito, Valentina Franzoni, Marco Guerini, Delia Irazu Hernandez Farias, Simona Frenda, Emiliano Lorini, Saif Mohammad, Alessandro Moschitti, Marinella Paciello, Isabella Poggi, Paolo Rosso, Diana Santos, Björn Schuller, Mohammad Soleymani, Khiet Truong, Carlo

Strapparava, Enrico Zovato. In addition, we thank all authors for submitting interesting papers, AISB and the local organizers for hosting us and for their exceptional support. Last but not least, we heartily thank our invited speakers, Diana Maynard and Pietro Cipresso, for agreeing to share their expertise on key topics of Emotions@AISB2018.

March 27, 2018

Emotions@AISB2018 Co-Chairs
Francesca D'Errico, University of Roma-Tre, Italy
Floriana Grasso, University of Liverpool, UK
Malvina Nissim, University of Groningen, The Netherlands
Nicole Novielli, University of Bari Aldo Moro, Italy
Viviana Patti, University of Turin, Italy

Table of Contents

Music Emotion Capture: sonifying emotions in EEG data

George Langroudi and **Anna Jordanous** and **Ling Li**[1]

Abstract. People's emotions are not always obviously detectable, due to difficulties expressing emotions, or geographic distance (e.g. if people are communicating online). There are also many occasions where it would be useful for a computer to be able to detect users' emotions and respond to them appropriately. A person's brain activity gives vital clues as to emotions they are experiencing at any one time. The aim of this project is to detect, model and sonify people's emotions. To achieve this, there are two tasks: (1) to detect emotions based on current brain activity as measured by an EEG device; (2) to play appropriate music in real-time, representing the current emotional state of the user. Here we report a pilot study implementing the Music Emotion Capture system. In future work we plan to improve how this project performs emotion detection through EEG, and to generate new music based on emotion-based characteristics of music. Potential applications arise in collaborative/assistive software and brain-computer interfaces for non-verbal communication.

1 INTRODUCTION

It is not always straightforward to detect people's emotions; people may be unable to express their emotions in usual ways due to physical or other limitations, or emotions may be difficult to detect if people are geographically separated e.g. interacting online. There are also many occasions where it would be useful for a computer to be able to detect users' emotions, particularly where the computer is working collaboratively with a user. If a computer can accurately detect a user's emotional state, it may be able to respond appropriately, demonstrating some 'empathy' with its user - which is likely to lead to a more positive user experience overall.

The aim of this project is to detect, model and musically sonify people's emotions in real time. To achieve this, there are two steps: (1) to detect emotions based on current brain activity as measured by an EEG device; (2) to generate appropriately matching music in real-time, representing the current emotional state of the user.

The pilot study reported in this work-in-progress paper uses a version of Russell's circumplex model of emotions [14] updated by Scherer [16], alongside a music database tagged with relevant metadata, to implement a music player in C# that interacts with emotions experienced by someone wearing an Emotiv EEG headset. The Emotiv headset device simplifies EEG signal capture, making EEG data more accessible for experiments such as this. Section 2 describes the Emotiv headset, as well as previous research on modelling emotions and expressing emotions musically. Section 3 describes implementation of our proof-of-concept Music Emotion Capture system.

Evaluation of software which responds to people's emotions is a tricky issue that needs careful handling, for ethical reasons. Evaluation (Section 4) of this work in progress has currently been limited to system evaluation, as we would like to improve some aspects of the system implementation before we evaluate the functionality of Music Emotion Capture with users. Section 4 describes how, in future work, we plan to improve how this project performs emotion detection through EEG, and how new music can be generated based on emotion-related characteristics of music.

We see a large number of potential applications of the ideas in this work. Section 4.3 describes how this work could be applied in music therapy or interactive software, mainly centred around applications with brain-computer interfaces for enhancing non-verbal communication and interaction, e.g. for improving computer-human communication/empathy in collaboration, or for customising HCI to user experience, or enabling people in a vegetative state ('locked-in') with a new mechanism for expressing the emotions they are feeling.

2 BACKGROUND

2.1 Computational modelling of emotions

How can human emotions be modelled computationally? For computers to be able to use data about human emotions, a computational model needs to be able to represent emotions accurately in a machine-readable format amenable to reasoning and analysis.

There are two leading theories for representing emotion: Ekman's six basic emotions [2] and Russell's circumplex model [14].[2]

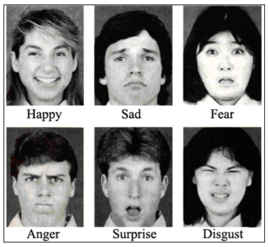

Figure 1. Ekman's basic emotions (from MIT Brain & Cognitive Sciences)

Ekman's model of six basic emotions Ekman has provided a simple set of six[3] dimensions for the study of emotions in psychology [2]: *happy, sad, fear, anger, surprise, disgust* (see Figure 1)[4].

Ekman's model is highly applicable for work focusing on detecting emotions through analysis of facial features. One issue is that, although now more widely interpreted, it was originally derived specifically around the study of emotions as revealed in facial expressions,

[1] School of Computing, University of Kent, Medway campus, UK, email: George.Langroudi@outlook.com ; [A.K.Jordanous / C.Li]@kent.ac.uk

[2] This is not to say that other models do not exist, e.g. [11].

[3] Though this model is widely treated as having six dimensions, we acknowledge that Ekman later argued for (*contempt*) to be added [10].

[4] Image source `https://ocw.mit.edu/courses/brain-and-cognitive-sciences/9-00sc-introduction-to-psychology-fall-2011/emotion-motivation/discussion-emotion/diss_img.jpg`.

as Figure 1 illustrates. It has been questioned[5] whether facial expressions (as analysed by Ekman) are fully representative of emotions, and whether Ekman's particular model is too focused on one culture's expression of emotions, a Western expression, without being applicable to other cultures [15]. Hence for this work's focus, on detection of emotions at the brain activity level, rather than through facial expression, alternative options should be considered too.

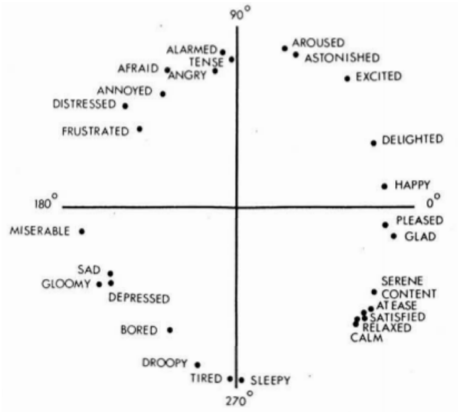

Figure 2. Russell's Circumplex model of affect [14]

Russell's circumplex model Predating Ekman's model, Russell's circumplex model [14] spatially represents emotions and their relations to each other. Emotions are plotted on a two-dimensional graph: see Figure 2. The two dimensions are *valence* and *arousal*:

- **valence**: the extent to which an emotion is considered positive or negative; e.g. a positive valence emotion on Russell's model is *HAPPY*, and a negative valence emotion is *MISERABLE*.
- **arousal** The extent to which an emotion represents a psychologically activated state; e.g. a high arousal emotion on Russell's model is *ALARMED*, and a low arousal emotion is *TIRED*.

Unlike Ekman's model, Russell's model acknowledges interrelations and dependencies between emotions; e.g., if one is feeling happy, there is an expectation that feelings of sadness are likely to be correspondingly low or non-existent. On the Russell model, similarity-based relationships between emotions can be analysed using distance. Similar emotions are plotted close to each other and dissimilar emotions far away from each other. For example, in Figure 2 the emotions *MISERABLE/SAD/GLOOMY*, all similar in sentiment, are positioned close to each other, but far away from a cluster of more positive emotions such as *HAPPY/PLEASED/GLAD*, which are opposed in meaning to the first cluster of emotions.

Scherer's update to the Russell model Russell's model gives emotions neatly arranged in a circular pattern, such that the only emotions expressed are those that fall in (x, y) co-ordinates in the circumplex graph that roughly satisfy the circle equation of $(xh)^2 + (yk)^2 = r^2$. No other emotions are represented in this Russell model. But what if we want to represent the emotions being experienced by someone at a state of, say, 0 arousal and with only a small positive level of valence?

[5] We note this comes from Russell, the author of another emotions model.

Russell's model was updated by Scherer [16] to represent a greater variety of emotions than those at the edge of the circle: see Figure 3.

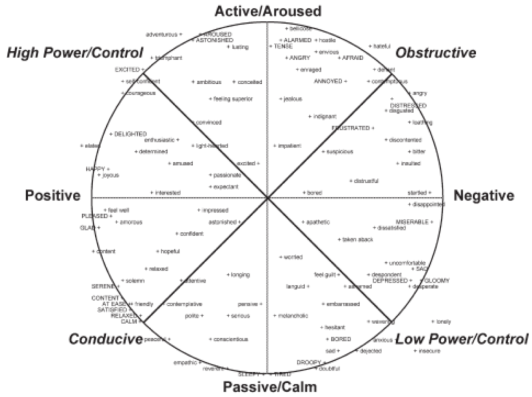

Figure 3. Scherer's updates to the Russell circumplex model [16]

Scherer's updated model was applied for measuring the emotional content of blog posts by [13]. During this application, the graphical model was converted into quantitative data and reported in [13]. This becomes very useful for computational implementation; we now have a quantified version of a widely-cited model of emotions that can be experienced, which has been successfully applied for modelling emotional content. The model is based on two simple dimensions of valence and arousal, and is representative of a wide range of emotions, across the full ranges of these dimensions. Hence we chose the Russell model (as updated by [16, 13]) for this work.

2.2 Detecting emotions through EEG

The Emotiv EEG headset detects and measures EEG signals from the scalp. It allows data acquisition of 14 channels at 128Hz per channel, which is sufficient to capture the brain activity below 64Hz. This frequency range covers the frequency bands alpha, beta, theta and delta; i.e. the majority of the EEG used in clinical practice. Unlike clinical EEG measuring equipment, the headset can be prepared and used fairly quickly for experimental purposes. This accessibility comes with a trade-off in accuracy: as it has fewer channels than some high spec EEG equipment, with less rich data (especially when the topological distribution patterns are of interest). This is, though, a more accessible way of obtaining EEG data.

Emotiv's SDK gives access to four measurements: *Engagement, Excitement, Meditation* and *Frustration*. Each measurement is given as a float between 0 and 1. The measurements offered by Emotiv are useful indicators to use for our pilot study, but there is some discussion concerning these detectors' validity and reliability [5]. We discuss alternative ways to detect emotion-related data in Section 4.

2.3 Musical expression of emotions

Work on expressing emotions through music has been a topic of study for decades, e.g. the seminal work by Meyer [12]. Juslin & Laukka [7] reviewed several studies covering the acoustic expression of emotions through music and through speech. They found evidence in the reviewed studies that music can be used to express emotions, in a manner similar to how we use vocal cues. For example, as shown in Figure 4, the expression of the emotion 'Anger' is often expressed via music with fast tempos,[6] high variability in volume (sound level) and in the range of musical notes (pitches) used, as well as other acoustic patterns. In contrast, musical expressions of the emotion ‘Sadness’ are typically within music with slow tempos and low variability in volume/sound level and in the range of notes/pitches, as well as having other musical features (see Figure 4).

For our work, generation of music through this set of musical patterns would be highly interesting to study: a valid standalone research project in its own right (e.g. [3]). For faster progress in this pilot study, though, we also investigated if there were precomposed, existing sources of musical data that we could use, with annotations representing emotions associated with each piece of music.[7] The *Emotion in Music Database* (http://cvml.unige.ch/databases/emoMusic/) [18] consists of 744 musical extracts (45 second snippets of music from the Free Music Archive (http://freemusicarchive.org/). As described by [18], each extract is provided with metadata annotations of crowd-sourced ratings of both valence and arousal, gathered by people submitting ratings as they listened to the music and these ratings being averaged out over the full extract and over different raters, and normalised to a range of [-1, 1]. The approach of [18] is not without its flaws: notably, taking average ratings without use of accompanying measures of standard deviation/variance hides subtleties in variation of ratings over time. For our work, though: this is a dataset that has been deployed in previous research, is openly available and fits with our decision to use a valence- and arousal-based model of emotion.

Summary of Cross-Modal Patterns of Acoustic Cues for Discrete Emotions

Emotion	Acoustic cues (vocal expression/music performance)
Anger	Fast speech rate/tempo, high voice intensity/sound level, much voice intensity/sound level variability, much high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and microstructural irregularity
Fear	Fast speech rate/tempo, low voice intensity/sound level (except in panic fear), much voice intensity/sound level variability, little high-frequency energy, high F0/pitch level, little F0/pitch variability, rising F0/pitch contour, and a lot of microstructural irregularity
Happiness	Fast speech rate/tempo, medium–high voice intensity/sound level, medium high-frequency energy, high F0/pitch level, much F0/pitch variability, rising F0/pitch contour, fast voice onsets/tone attacks, and very little microstructural regularity
Sadness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contour, slow voice onsets/tone attacks, and microstructural irregularity
Tenderness	Slow speech rate/tempo, low voice intensity/sound level, little voice intensity/sound level variability, little high-frequency energy, low F0/pitch level, little F0/pitch variability, falling F0/pitch contours, slow voice onsets/tone attacks, and microstructural regularity

Note. F0 = fundamental frequency.

Figure 4. How music and voice express different emotions (from [7])

3 IMPLEMENTATION

The Music Emotion Capture software was written in C# using Emotiv’s Professional SDK. The interface was made using the OpenTK library for C# to make a basic OpenGL display.

3.1 Detecting valence and arousal

The Emotiv SDK translates the signals from the Emotiv EEG headset to give measurements of the extent to which each of the four basic emotions (Excitement, Engagement, Meditation, Frustration) is being experienced by the user at a given time. Using Russell’s circumplex model, (only) Valence and Arousal measurements are needed. We used Engagement as a proxy measure of arousal, and used Excitement as a proxy measure for Valence. We acknowledge the potential flaws from these choices of proxy; while not perfect translations, these compromises allow us to progress in our aims.

3.2 Mapping emotions to music

Emotiv will give us a constantly updated float between 0 and 1 for each of the four basic emotions. We mapped valence and arousal to the X and Y axis respectively of a graph (see Figure 5), plotting a point on the graph, and then used the emotion mapping from [13].

When the Music Emotion Capture software runs, the software continuously takes EEG measurements. The EEG measurements are continually translated into valence-arousal co-ordinates, to generate data about possible emotional states. In the terminal window in Figure 5 you can see this data: the emotions in an ordered list (the closest emotion at the top) as well as the ‘distance’ between that emotion and the user’s current emotion (see Section 2.1).

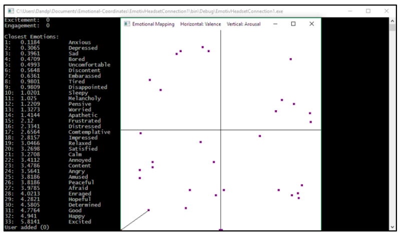

Figure 5. Output of the Music Emotion Capture pilot program, with the terminal showing emotions currently being detected by the system at a given point in time, ordered by relevance, and the graph plotting each of the Emotional Coordinates (x axis: valence, y axis: arousal)

Using the Emotion in Music Database described in Section 2.3 [18] we mapped each song to a location then checked for the song that matches the users emotion closest every 10 seconds

4 EVALUATION AND FUTURE WORK

Evaluation to date has been through white-box tests, e.g. testing with different settings/checking that the code functions as expected when different emotions are detected. As of yet, we have not yet run experiments with subjects; this will happen when the current pilot study is developed further. The evaluation methodology is crucial for this project, especially how emotions relate with music. At that point, we intend to evaluate the project by testing people with the headset, asking people whether they feel the music they hear matches their emotional state. We will also ask participants to describe their emotional state at various points in time during the study, so that we have data to evaluate whether self-reported emotion match what the headset was picking up at that time. Emotion databases are also available e.g. AAAC, RECOLA,[8] which we see as a foundation for data-driven evaluation, to augment user-reported emotion-music associations.

This pilot study represents work in progress; the above user evaluations form a key part of future work. Certain design choice compromises were made during this pilot study (as described above). We would like to carry out some of the following future work improving the system before we then conduct a formal user evaluation.

4.1 Improved Valence/Arousal detection

In Section 2.2 we noted that although Emotiv provide easy access to four emotional measurements via their headset, (1) questions have been raised regarding the validity of these measurements and (2) our

[6] The tempo of a piece of music is the speed at which a regular metrical beat or pulse occurs; e.g. 180 beats per minute (bpm) would be a fast tempo in Western Classical music, whereas 45 bpm would be a slow tempo.

[7] However we discuss in Section 4 our intentions to develop this pilot study with real-time generation of music based on the findings of [7].

[8] AAAC: http://emotion-research.net/wiki/Databases; RECOLA: http://diuf.unifr.ch/diva/recola/index.html.

choice of two of these as valence and arousal proxies is potentially flawed. Research has been done on improving Emotiv-based emotion recognition by fusing biometric and eye tracking technologies [9]. Also, other EEG based arousal/valence work indicated 82% accuracy for automatic classification of positive, negative and neutral valence based on film clip viewing, using frequency feature and its dynamics in EEG [1]. We envisage new feature extraction and pattern recognition methods, which have updated the state-of-the-art in many areas. Therefore we would like to investigate work calculating valence and arousal from raw EEG signals (e.g. [4]). Moreover, we will investigate different approaches for quantifying EEG patterns in emotions with time-frequency features, synchrony-based features [8] and perform pattern recognition considering recent work achieved with deep neural network architectures and generative models in identifying emotional states [19] and generating emotion-specific music.

4.2 Generating music using the user's emotions

As described in Section 2.3, Juslin & Laukka [7] review how acoustic cues in music (and in speech) can be used to express different emotions. They conclude from their review that: "emotion-specific patterns of acoustic cues used to communicate each emotion ... are generally consistent with K. R. Scherer's (1986) theoretical predictions." [the predictions by Scherer that formed the basis for [16]]. Given that the pilot study reported above is based on Scherer's development of the Russell circumplex model of emotions (Section 3), the guidelines from [7] are a promising direction for future work.

4.3 Potential applications

Many practical applications of this work emerge from the potential of **brain-computer interfaces (BCI) for enhancing non-verbal communication of emotion**. Several interesting paths of investigation emerge for Human-Computer Interaction (HCI) issues around how collaborative software can be **better customised to the user experience**. If the user is currently experiencing, say, high levels of frustration, the software can interact differently to if the user is experiencing high levels of happiness.[9] For example, a BCI could be used to enhance the quality and accuracy of computer-human communication during collaboration, e.g. with a Creativity Support Tool [17], or with software designed to be a creative partner in a *co-creative* scenario (a scenario which users sometimes find more troubling to navigate [6]).

Other applications in the domain of **assistive technologies** include enabling people in a vegetative state ('locked-in') to communicate emotions they are feeling, either directly to software or to other people via the sonification of their emotions. **Music therapy** applications can also be envisaged: playing music to people which is aimed at bringing their emotional state into the region of positive emotions.

There is also potential application to the task of **playlist creation**. Being able to detect a person's current mood may allow musical software to better recommend songs. For example, if the software detected that the user was relaxed it may want to recommend a similarly relaxing playlist. With the possibility of more portable, functional EEGs[10] one can imagine music players on a user's phone that adjust its suggestions to adapt to the user's current state.

[9] This idea stems from conversations with composer Oded Ben Tal.

[10] Developments look promising: see https://kokoon.io/ ; https://spectrum.ieee.org/the-human-os/biomedical/devices/wireless-earbuds-will-record-your-eeg-send-brainwave-data-to-your-phone .

5 CONCLUSION

This work-in-progress paper reports a pilot study which enables us to detect, model and musically sonify the emotions of people who interact with our EEG-based Music Emotion Capture software. An Emotiv headset lets us gather EEG data for translation into music representing the user's current emotions, building upon the Russell circumplex model of emotions. There are many exciting avenues for potential applications, following further work and full evaluation.

ACKNOWLEDGEMENTS

We would like to thank members of the Data Science group at the School of Computing, University of Kent, in particular Palani Ramaswamy, for their useful comments.

REFERENCES

[1] L. Brown, B. Grundlehner, and J. Penders, 'Towards wireless emotional valence detection from EEG', in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2188–2191, Boston, MA, (2011).

[2] P. Ekman, 'An argument for basic emotions', *Cognition & Emotion*, **6**(3), 169–200, (1992).

[3] S. Elisha and T. Weyde, 'Exploration of emotion-based cross-modal mappings for generating music for videos', in *2nd Conference on Computer Simulation of Musical Creativity*, Milton Keynes, UK, (2017).

[4] G. Garcia-Molina, T. Tsoneva, and A. Nijholt, 'Emotional brain–computer interfaces', *International Journal of Autonomous and Adaptive Communications Systems*, **6**(1), 9–25, (2013).

[5] T Harrison. The Emotiv mind: investigating the accuracy of the Emotiv EPOC in identifying emotions and its use in an intelligent tutoring system. Honours report, University of Canterbury, New Zealand, 2013.

[6] A. Jordanous, 'Co-creativity and Perceptions of Computational Agents in Co-creativity', in *Eighth International Conference on Computational Creativity (ICCC'17)*, pp. 159–166, Atlanta, GA, (2017). ACC.

[7] P. N. Juslin and P. Laukka, 'Communication of emotions in vocal expression and music performance: different channels, same code?', *Psychological Bulletin*, **129**(5), (2003).

[8] L. Li, A. Witon, S. Marcora, H. Bowman, and D. P. Mandic, 'EEG-based brain connectivity analysis of states of unawareness', in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pp. 1002–1005. IEEE, (2014).

[9] J-M. López-Gil, J. Virgili-Gomá, R. Gil, T. Guilera, I. Batalla, J. Soler-González, and R. García, 'Method for improving eeg based emotion recognition by combining it with synchronized biometric and eye tracking technologies in a non-invasive and low cost way', *Frontiers in Computational Neuroscience*, **10**, 85, (2016).

[10] David Matsumoto and Paul Ekman, 'The relationship among expressions, labels, and descriptions of contempt.', *Journal of personality and social psychology*, **87**(4), 529, (2004).

[11] Albert Mehrabian, 'Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament', *Current Psychology*, **14**(4), 261–292, (1996).

[12] L. B. Meyer, *Emotion and meaning in music*, University of Chicago Press, Chicago, 1956.

[13] G. Paltoglou and M. Thelwall, 'Seeing stars of valence and arousal in blog posts', *IEEE Trans. Affective Computing*, **4**(1), 116–123, (2013).

[14] J. A. Russell, 'A circumplex model of affect', *Journal of Personality and Social Psychology*, **39**, 1161–78, (1980).

[15] J. A. Russell. Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies., 1994.

[16] K. R. Scherer, 'What are emotions? and how can they be measured?', *Social Science Information*, **44**(4), 695–729, (2005).

[17] B Shneiderman, 'Creativity support tools: Accelerating discovery and innovation', *Commun. ACM*, **50**(12), 20–32, (December 2007).

[18] M. Soleymani, M. N. Caro, E. M. Schmidt, C-Y. Sha, and Y-H. Yang, '1000 songs for emotional analysis of music', in *Procs 2nd ACM international workshop on Crowdsourcing for multimedia*, (2013).

[19] A. Supratak, S. Schneider, H. Dong, L. Li, and Y. Guo, 'Towards desynchronization detection in biosignals', in *NIPS Time Series Workshop*, Long Beach, CA, (2017).

Associating Colours with Emotions Detected in Social Media Tweets

Robert Harvey[1], **Andrew Muncey**[1] **and Neil Vaughan**[1]

Abstract. This project involves two major areas of work, the detection of emotions in text from Twitter posts (tweets), and the association of that emotion with colour. Emotion mining is the field of natural language processing which is concerned with the detection and classification. It is a subfield of semantic analysis which contains both emotion and opinion mining. Both tasks depend on an emotion model to classify detected emotions and to associate a colour depending on the location of the emotion in the model. This research paper demonstrates preliminary results from classification of tweets to assign emotion labels. Also designs are presented for a prototype web interface for displaying the assigned colour and emotion associated with tweets.

1 INTRODUCTION

The overall aim of this project was to develop a colour coding system for real-time tweets to be used for marketing, or audience research. This incorporates three individual project aims of collecting tweets as they are posted, analysing the text, and applying a coloured label.

The motivation is that tweets could be colour coded according to the emotional sentiment content, enabling people to rapidly visualise an overview of sentiments within tweets.

The hypothesis is that a program can be created to apply a colour label to social media posts based on its emotion.

To achieve this aim, the system will complete the following objectives:

1. Identify current approaches to NLP processing used in text-to-scene applications
2. Assess suitability of python programming with machine learning for tweet classification.
3. Research methods of displaying the emotions using colour through Twitter python API.
4. Investigate the associations between colour and emotion for labelling.
5. Develop prototype system to display tweets with the coloured emotion label.

2 BACKGROUND

As early as the 1950s development into NLP began with what Lehnert & Ringle (1982) calls the "era of machine learning" during which, techniques focused on the extraction of single words and interpreting them separately for translating texts between languages. However, this technique lacked the ability to correctly understand words which can have opposite meaning depending on the surrounding words and structure. An example would be the word "accident" which can be good (happy accident) or bad (damaging accident) (Cambria & White, 2014).

Recent approaches to NLP for sentiment analysis involves training large neural networks with large knowledge bases of vocabulary. One such method in this approach is called 'skip-gram' which passes each keyword to another neural network, which then predicts the words either side to produce a binary tree which can be analysed (Witten, Frank, Hall, & Pal, 2016).

Recent SemEval winning methods have shown that word embedding is shown to perform best for sentiment analysis. The topic was well explored in the Computational Linguistics community, with machine learning (Strapparava & Mihalcea 2008), using a Lexicon to associate colour (Volkova et al., 2012), crowdsourcing Word-Emotion associations (Mohammad & Turney, 2013), Word–Colour associations (Mohammad, 2011), and color of text emotions (Strapparava & Ozbal 2010).

3 TECHNOLOGY

Twitter is a social media platform that allows users to post 280-character posts called Tweets. Tweets can also contain a hashtag, which is a categorisation system that allows users to tweet about similar events, products, etc.

Machine learning algorithms for NLP can be developed in many languages; however, some languages will perform better and be easier to develop for than others. The most popular languages for machine learning and 'deep learning' are Python, R, and Java (Puget, 2016).

4 RELATED STUDIES

Machine learning has been used with various NLP techniques to enable them to be more accurate than the hard-coded alternatives. Techniques such as 'bag-of-words' which analyses each word separately without context can be used with machine learning (Cronin, Fabbri, Denny, Rosenbloom, & Jackson, 2017). Machine learning was used to perform emotion mining by Alm, Roth, & Sproat (2005) who used the emotion from text to change how words were spoken by a text-to-speech system.

NLP is only one aspect of the project, with the display of mood colours being the other. Mapping certain colours to moods will be different for each user so groups of colours will have to be assigned to each mood using the most common colours as a starting map (Moon, Kim, Lee, & Kim, 2013).

[1] Dept. of Computer Science, Univ. of Chester, CH2 4NU, UK. Email: {1501281, a.muncey, n.vaughan}@chester.ac.uk.

5 NATURAL LANGUAGE PROCESSING

The first NLP systems attempted to parse text using 'semantic information processing' which uses keywords in sentences to trigger actions (Lehnert & Ringle, 1982). These applications worked on the principle that language is structured and that there are varying probabilities that a word may appear in a sentence. Early programs used a statistical model for a specific domain and apply these to calculate the probability of the next word. This approach is called n-grams where n is the number of probabilities required to specify a statistical model (Shannon, 1948). One such program was called SHRDLU, which was able to determine actions specified by a user using natural language (Winograd, 1972).

From this syntax based approach, research into NLP split into using semantic analysis and machine learning techniques (Cambria & White, 2014). Semantic analysis aims to recognise the semantic structure of a sentence to understand its meaning. "… in order to understand a sentence, it is necessary to know its syntactic pattern." (Chomsky, 1957). This approach was popular in machine translation as the second language could be generated over the semantic structure of the original sentence (Schank, 2014).

Current approaches to NLP utilise neural networks for machine learning on top of adaptable knowledge bases. Current research into machine translation focuses on recurrent neural networks with enhanced 'long short-term memory' to better maintain information throughout a sentence (Hirschberg & Manning, 2015). This memory is achieved via backpropagation through time (BTT) algorithms which allow data to be propagated back through time and be remembered over multiple steps in the hidden layer of the network (Mikolov, Kombrink, Burget, Cernocky, & Khudanpur, 2011).

6 EMOTION MINING

Emotion mining is used to detect and understand, the emotions present in a passage of text. It is like opinion mining in that it is a form of sentiment analysis, however, opinion mining detects a person's emotion towards another entity, rather than their internal mood. Current emotion mining techniques work on the sentence level, classifying emotion for each sentence, and use annotated data models to calculate the expressed emotion (Yadollahi, Shahraki, & Zaiane, 2017). Polarity determination is also important for emotion mining as it determines whether a sentence is expressing positive or negative emotions (Ravi & Ravi, 2015). This is useful when applied to naïve approaches which look for key emotion words instead of analysing the entire sentence. Keyword searching may interpret the sentence "I am not happy" as happy as it is the only emotion word. However, polarity determination would recognise the not before the keyword and reverse happy to sad.

Research into emotion mining has focused on text generated by users on social media or for marketing purposes which aims to detect the emotion of the writer. However, for this project, the aim was to classify the emotion that the writer aims to invoke in the user. Pizzi. et al, (2007) expands this and used NLP to present each character in a story with their own internal emotion to better capture the emotions between characters. This can be used to calculate the overall mood of the section with weights given main characters.

Alm, et al., (2005) also used narrative text as the basis for their research. Their application of emotion mining is used to enhance a text-to-speech system for reading fairy tales. Machine learning was also utilised to detect the valence of basic emotions to change the pitch and speed of the output speech.

7 MODELLING MOOD, EMOTION AND COLOUR

To detect emotions, they must first be categorised in a model so that it can be analysed against the text. An emotional model is a system which uses either category of emotions, such as 'anger', or emotional dimensions, such as valence and arousal (Burkhardt & Stegmann, 2009).

In a study to display the mood of music via coloured lighting, Moon, et al., (2013) used Thayer's emotion model to determine the emotion. The Thayer emotional model is a 2D grid of emotions plotted against arousal and tension (Thayer, 1990). Because this model doesn't just use emotion adjectives, there is less ambiguity (Moon, Kim, Lee, & Kim, 2013).

From an emotion model, colours can be assigned to specific mood. Once an emotion has been determined, a system could generate a colour from a simple look-up table. However, the emotion a person associated with a colour can depend on many factors. Manav (2017) says that a person associates a colour with personal experience, memories, and cultural perceptions. Only cultural perceptions can be applied to a wide range of people and so multiple colour-emotion tables could be required. This could be made easier as, a study into emotion and colour preferences from Ou, et al., (2004) found that some emotions may be associated with the same colour across many countries.

8 IMPLEMENTATION

Following current approaches to NLP in literature, this project utilised machine learning techniques to perform emotion detection (Puget, 2016).

Tokenized Datasets

The first tokenized dataset contains tweets. Each tweet consists of a string, a category as string and an attribute label from the set {negative, neutral, positive}. The dataset was from Sentiment140 (2017) which is available online.

The second data used was from the Grimms' tale dataset. Each sentence has been labelled for emotion and mood by two separate annotators. The dataset contains 8 labels: {N, A, D, F, H, SA, SU+, SU-}, which relate to emotional classes: neutral, angry, disgusted, fearful, happy, sad, positively surprised, negatively surprised. A second version of the dataset was created using only 2 labels: {Neutral, Emotional}. The dataset was formatted to combine 20 labelled Grimms' tales into a single dataset, containing 2036 labelled sentences, only containing the emotion labels from the first annotator to improve consistency.

Method

The same method was applied to both datasets from tweets and Grimms combined dataset. This method was based on previous research by Kiritchenko et al. (2014) and the AffectiveTweets package for analyzing emotion and sentiment.

First a pre-processing filter was applied, converting the text string to Sparse Feature Vectors (SFV). The SFVs are calculated including word and character n-grams. This has been previously useful for filtering out infrequent features and setting the weighting approach. A support vector machine (SVM) was trained. For comparison, SVM training was completed twice for each dataset, once with SFV pre-processed data and once with raw data. Ten-fold cross validation was applied to assess the classification accuracy which has advantages over using a training/test data split.

Results

With the tweets dataset pre-processed into Sparse Feature Vectors (SFV), the trained SVM classified 74% of tweets correctly into one of the three labelled classes (Table. 1).

Without any pre-processing, SVM classification was not as successful, producing 36% correct classification.

The same method of pre-processing and SVM classification was applied to the Grimms' tales dataset. Sentences were correctly classified 66.6% of instances (Table. 2). When the dataset contains only two class labels {Neutral, Emotional} classification accuracy increased to 71%.

Correctly classified instances	37	74%
Incorrectly classified instances	13	26%

Table 1. Classification results: 74% correctly with the tweet data using pre-processed labelled data converted to Sparse Feature Vectors (SFV) and classified with a trained support vector machine (SVM).

1169	24	23	17	72	16	6	22	N
58	21	0	0	4	6	1	2	F
58	2	27	3	8	1	0	1	A
45	1	4	16	1	0	4	2	D
111	4	3	1	74	4	0	0	H
52	4	4	3	9	30	0	4	Sa
22	2	2	2	4	0	0	2	Su+
55	0	2	4	6	2	2	14	Su-
N	F	A	D	H	Sa	Su+	Su-	
Classified as								

Table 2. Confusion matrix - results correctly classifying 1351 (66%) from the 2036 sentence Grimms' tales dataset using a trained support vector machine (SVM).

Online web prototype design for tweets

Following on from this initial prototype of the tweet emotion classification methods, the next objective would be to display the tweets to the user along with a visual indication of its emotion. Designs were created for how this could be integrated into a web browser. In future work we plan to implement this feature for use in a web browser to display emotion for real-time tweets. The goal would be to use an individual web-page or use twitter itself via a web browser extension to add elements and styling. Figure 1. shows one of the proposed designs for a browser extension solution.

In the prototype, colours could be associated with the 8 emotional classes: N, A, D, F, H, SA, SU+, SU-}, which relate to emotional classes: neutral (N) - grey, angry (A) - red, disgusted (D) - green, fearful (F) - purple, happy (H) – yellow, sad (SA) - blue, positively surprised (SU+) - orange, negatively surprised (SU-) - turquoise.

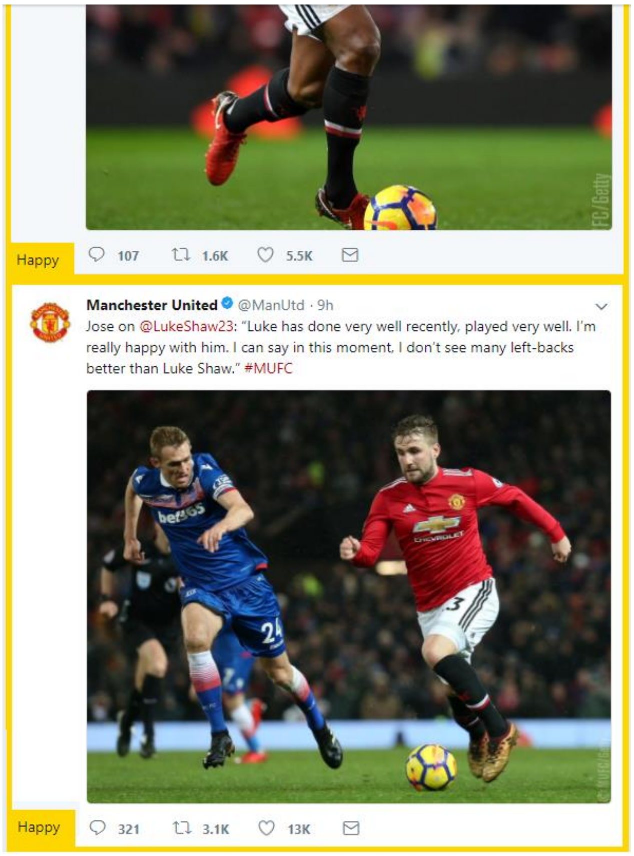

Figure 1. Emotion labels added to tweets inside the Twitter web browser.

9 CONCLUSIONS & FUTURE WORK

This paper builds on the existing work in emotion modelling. A proof-of-concept implementation was presented, which has been functionally tested and which will provide a test bed for further research into this area. We intend to perform a range of experiments to see how well this performs in various functions and to explore the advantages and disadvantages of several approaches for incorporating user satisfaction into the decision making.

Future work could include extending this to other platforms. A review into the current approaches to machine learning by Yadollahi, et al., (2017) finds that much of the research into emotion mining utilises machine learning with neural networks to classify the emotion of a complete sentence. On the topic of machine learning for emotion mining, Bantum, et al., (2017) says that its goal is to apply many features to detect and analyse text at the sentence level.

No single study found is the same as this project the most relevant sources for this project was Alm, et al., (2005) and Pizzi. et al, (2007) which aimed to detect emotion from a story first before completing their own output objective. Another notable study was Moon, et al., (2013) which displayed emotion via lights and so their work into matching colour and emotions was very important for the output task of the project.

REFERENCES

[1] Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing - HLT '05, 579-586.

[2] Bantum, E. O., Elhadad, N., Owen, J. E., Zhang, S., Golant, M., Buzaglo, J., . . . Giese-Davis, J. (2017). Machine Learning for Identifying Emotional Expression in Text: Improving the Accuracy of Established Methods. Journal of Technology in Behavioral Science, 2(1), 21-27.

[3] Burkhardt, F., & Stegmann, J. (2009). Emotional speech synthesis: Applications, history, and possible future. Proc. ESSV.

[4] Cambria, E., & White, B. (2014, May). Jumping NLP Curves: A Review of Natural Language Processing Research [Review Article]. IEEE Computational Intelligence Magazine, 9(2), 48-57.

[5] Chomsky, N. (1957). Logical Structures In Language. Journal of the Association for Information Science, 8(4), 284-2910.

[6] Hirschberg, J., & Manning, C. D. (2015). Advances in Natural Language Processing. Science, 349(6245), 261-266.

[7] Lehnert, W. G., & Ringle, M. H. (1982). Strategies for natural language processing. New York: Psychology Press.

[8] Manav, B. (2017). Color-emotion associations, designing color schemes for urban environment-architectural settings. Color research and application, 42(5), 631-640.

[9] Mikolov, T., Kombrink, S., Burget, L., Cernocky, J., & Khudanpur, S. (2011). Extensions of recurrent neural network language model. Acoustics, Speech and Signal Processing (ICASSP), 5528-5531.

[10] Moon, C. B., Kim, H., Lee, D. W., & Kim, B. M. (2013). Mood lighting system reflecting music mood. Color Research & Application, 40(2), 201-212.

[11] Ou, L.-C., Luo, R., Wookcock, A., & Wright, A. (2004). Ou, L. C., Luo, M. R., Woodcock, A., & Wright, A. (2004). A study of colour emotion and colour preference. Part I: Colour emotions for single colours. Color Research & Application, 39(3), 232-240.

[12] Pizzi, D., Charles, F., Lugrin, J.-L., & Cavazza, M. (2007). Interactive Storytelling with Literary Feelings. Affective Computing and Intelligent Interaction, 630-641.

[13] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. Knowledge-Based Systems, 14-46.

[14] Schank, R. C. (2014). Conceptual information processing. New York: American Elsevier Publishing Company.

[15] Shannon, C. E. (1948). A Mathematical Theory of Communication. Bell System Technical Journal, 27(4), 623-666

[16] Thayer, R. E. (1990). The Biopsychology of Mood and Arousal. Oxford University Press.

[17] Winograd, T. (1972). Understanding Natural Language. Cognitive Psychology, 3(1), 1-191.

[18] Yadollahi, A., Shahraki, A. G., & Zaiane, O. R. (2017, June). Current State of Text Sentiment Analysis from Opinion to Emotion Mining. ACM Computing Surveys, 50(2), 1-33.

[19] Sentiment140, (2017) http://help.sentiment140.com/, Available online. Accessed 19 Jan 2018.

[20] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723-762.

[21] Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. In Proceedings of the 2008 ACM symposium on Applied computing (pp. 1556-1560). ACM.

[22] Volkova, S., Dolan, W. B., & Wilson, T. (2012). CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 306-314). Association for Computational Linguistics.

[23] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. Computational Intelligence, 29(3), 436-465.

[24] Mohammad, S. (2011). Colourful language: Measuring word-colour associations. In Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics (pp. 97-106). Association for Computational Linguistics.

[25] Strapparava, C., & Ozbal, G. (2010). The color of emotions in texts. In Proceedings of the 2nd Workshop on Cognitive Aspects of the Lexicon (pp. 28-32).

The lexicon of feeling offended

Francesca D'Errico[1], Isabella Poggi[1]

Abstract.
The work investigates a neglected but widespread emotion: feeling offended. After overviewing previous research on offense, which considers feeling offended mostly as a damage to a person's image, a survey study is presented aimed at deepening the definition of feeling offended and the description of its typical episodes. The definition emerging from the content analysis allows to code and group the more frequent words used in the corpus and to run a lexicometric analysis that describes the emotions connected to the feeling of offense, the individual traits of offender and victim, the causes of offense, and the types of judgment that most typically appear offensive. The applicability of the present work is related to the extraction of potential states of this emotion in other corpora where implicit or past intergroup conflicts are not completely solved, by also allowing a possible annotation of their seriousness in view of a potential process of reconciliation.

1 INTRODUCTION

Emotions are an alert device that wakens our attention any time an adaptively very important goal of ours is, or is likely to be, dramatically achieved or thwarted. The dimension of valence that characterises all emotions, their being pleasant or unpleasant ones, depends on their warning about either a goal achievement or thwarting, and all negative emotions are in some sense healthy in that the unpleasant feeling they cast over us, by warning that a relevant goal is thwarted, may induce us to care that goal in the future.
Among the most important goals of humans are those of positive image and self-image, the goals of eliciting positive evaluations about ourselves both by other people and oneselves, and the "image" and "self-image" emotions that monitor these relevant goals are those generally called "self-conscious" emotions (Lewis, 1986), like the positive one of pride and the negative ones of shame or humiliation.
This work deals with an emotion that has not been so often investigated so far: feeling offended. A very unpleasant emotion that we feel when our image and possibly our self-image are dramatically disrupted by some event that in some way reveals a low consideration of ourselves on the part of other people. After overviewing previous research on offense (Section 2), we present a survey study on the feeling of offense (Sect. 3), then by exploiting lexicometric techniques, by analysing the words used by our subjects in answering our survey we extract how the concepts displayed more deeply describe the emotions connected to the feeling of offense, the personal characters of offender and victim, the causes of offense, and the types of judgment that most typically appear offensive (Sect. 4).

[1] Dept. Philosophy Communication and Visual Arts, Univ. Roma Tre, Via Ostiense 234, Rome (Italy). Email: {Francesca.derrico}@uniroma3.it; {isabella.poggi}@uniroma3.it

2. Related work

Feeling offended is a complex emotional state that is often modulated by personal factors like gender and self-esteem on the basis of different expectations or causal attributions (internal vs external); but it also involves relational factors that affect the interpretation of the offense, since the person who is the cause of the offense can be a relative, a friend, an acquaintance, a colleague; and being offended by these people, with whom one has more or less involving social relationships, may imply heavy emotional costs. Within personal factors, self-esteem plays a crucial role in the feeling of offense, since it can affect self-relevant emotions like shame and pride (Brown and Marshall, 2001) – people with low self-esteem tend to feel shame more than others; while gender mediates the feeling of offense, especially in family contexts (Mosquera et al., 2002) – women feel offended more than men.
The multidimensional factors that characterize this feeling have been investigated in various psychological fields, from the dynamic approach to social psychology. According to Zander (1976), the feeling of offense is a deeply unpleasant emotion which goes through three phases: 1. identification of the cause, interpreted as an insult to an ideal value; 2. feeling of offense, with its relative intensity related to the 'expectations of recognition', and 3. reaction to the feeling of offense, modelled according to socio-historical variables.
Within the socio-cognitive framework, Mosquera et al. (2002) found that in the so-called "honour cultures" like Spain, as opposed, for example, to the Netherlands, the prototypical case of offence takes place in public and is referred to masculinity or to female sexual morality. According to Cohen et al. (1996), the higher the honour concern and the significance of the honourable person, the strongest the emotional response to insults. In cultures with a stronger code of honour, like Spain, the reactions to the offence are more embarrassed than elsewhere, mostly in relation to threats to family honour, so important for individual self-esteem that when someone is offended one's own self-esteem can be damaged too. Further gender differences are reported in emotional reactions of both shame and anger in case the insults undermine the sexual dimension, especially with Spanish women, for whom the sexual code of honor (sexual shame) is stronger than for Duch women.
Interpersonal and intergroup elements are also central to the feeling of offense in studies on forgiveness (Mc Cullough, 2000; Paleari & Regalia, 2005): those who feel offended may feel inferior in terms of perceived control (Baumeister et al.,2003)

and experience feelings of victimization or anger (McCullough, 2000), and this results in a need to restore their sense of power, by also increasing power-seeking behavior (Foster & Rusbult, 1999). Such needs are welcome within a possible socio-emotional reconciliation perspective in which the offender attempts to undertake a "cycle of apology-forgiveness" (Nadler, 2002): the responsible person puts one's own personal image in the hands of the other person, at the risk of not being forgiven. This presupposes a willingness to forgive as a result of a long-term reconciliation path, for example where the transgressor admits one's responsibilities.

Studies on forgiveness and reconciliation provide a complex framework where personality, ruminating tendencies, emotional stability, empathy towards the transgressor can lead the offended person to forgive, if offenses are explicit (e.g., betrayal, physical and emotional humiliation).

Three problems in these studies are, first, somewhat circularity, in that they define self-esteem as what is affected by offenses; second, that they mainly investigate the emotional responses triggered by explicit offenses, especially verbal insults, while neglecting less direct and less explicit offenses; third, that they almost exclusively focus on the offense as seen as a public blow to image before other people.

To fill in these gaps requires to investigate the feeling of offense in everyday contexts of interpersonal relationships, and to assess if it is triggered only or typically by explicit communicative actions like public discredit or insults, or whether even by more implicit and indirect actions.

This has been recently done by Poggi & D'Errico (2018), who in a survey study examined several autobiographic reports about the feeling of offense. Besides providing a formal definition of this emotion and finding out its most typical causes, they show that it is not triggered only by blatant aggressive communicative acts, by often by more indirect and implicit blows, like non-communicative actions and even non-actions. Further, the feeling of offense is not only felt in public contexts, but much more – and more importantly – in affective relationships: the offense tends to elicit high arousal emotions, such as anger, pride, and revenge, in males and people with high self-esteem, while it triggers low arousal emotions, such as bitterness, shame, and even fear, in people with low self-esteem. Such emotions tend to be felt by females more than males, both due to cultural pushes for women to lower self-esteem, and due to their higher tendency to care personal relationship, hence to suffer more for their breaking.

3. Feeling offended: a survey study

The goal of our work is to investigate the feeling of offense through direct self report of people. The focus of our analysis is the definition and description of this emotional state in terms of a socio-cognitive model (Castelfranchi, 2000; Miceli and Castelfranchi, 2014; D'Errico and Poggi, 2016) which views emotions as complex subjective states, entailing feelings, physiological, expressive and motivational states, triggered when a given cognitive configuration is represented in the mind of a person: the beliefs about an occurring or imagined event, attributions and evaluations about the self, as well as the goals at stake and the goals triggered by the emotional syndrome.

3.1. Research questions

On the basis of subjects' reports, we want to fulfil the following research goals:

1. to provide a definition of the feeling of offense in terms of its "mental ingredients", that is, the beliefs and goals represented in the mind of the person who is feeling that emotion;
2. to explore the associated lexicon of 'feeling offended' emotional states by means of a lexicometric analysis that will help is to describe the connected emotions, the individual traits of offender and victim, the causes of offense, and the types of judgment that most typically appear offensive.

3.2. Method

To investigate these issues, we submitted a semi-structured online survey to a sample of 129 participants, mainly Italians, balanced and composed by 61% women (n.79, vs. 50 males), age 31,2 (SD = 14,1), the majority with a high school bachelor (54%) or a University degree (26%).

The survey included 14 close and 11 open questions, asking how frequently the participants felt offended, for what reasons, who offended them, and in what life domains (work, family, friends…). To go deeply into the emotional experience of feeling offended, participants were asked to report one case in which they felt so, the specific reasons why they did, if they believed the other intended to offend, their relationship with the other before and after the offense, and what other emotions they connected to the feeling of offense. We also asked if they reminded of some case in which another wanted to offend them but they did not feel offended, and if so, why they did not; and conversely, if in some cases another person had felt offended by them, but should not have felt so, and why. Finally participants were asked to provide a definition of "feeling offended".

Two types of analysis were conducted on the open questions: first, a classical manual content analysis, then an automatic analysis through lexicometric techniques.

4. Content analysis

The content analysis of the open answers to the questionnaire – both the participants' definitions and their description of offensive events – first allowed us to outline a general definition of "feeling offended" and to extract its mental ingredients.

4.1. Definition

Feeling offended is a negative emotion felt by a person A, caused by either a communicative or a non-communicative act by another person B that results into an aggression to A's image, since it explicitly points at or implicitly entails a negative property of A: a property worth a negative evaluation of A by B with respect to an evaluation criterion relevant for the image which A wants to project, and shared with B. Being attributed this negative property is seen as a true wound to A's image, that in some way implies a lack of respect for A (lack of care for his/her image), and the aggression on the part of B is considered unjust by A, who thinks s/he does not really deserve to be attributed such a property. Actually A, though sharing the evaluation criterion with B, may not share the same factual knowledge: A and B share the value in terms of which facts can be judged, but not the really occurred facts. The problem with A – which definitely causes A to feel offended – is that B is relevant for A, in that A has/had the goal of keeping a positive social relationship with B. The whole fact results in subsequent negative social emotions of A towards B, such as disappointment and feeling betrayed by B, finally ending with a break in the social relationship of A with B, and at the same time, a loss of self-esteem for A.

4.2. Mental ingredients

Defining an emotion in some sense means to find out the necessary conditions for a person to feel that emotion; so the above definition of "feeling offended" can be translated into a set of conditions: the beliefs and goals that are necessary/sufficient for a person to feel offended. Among them, three types of conditions can be distinguished (somehow like in Searle's (1969) analysis of speech acts): a) preparatory conditions, b) essential conditions, and c) aggravating conditions.

PREPARATORY CONDITIONS:

1. A has the goal of a positive image before B
2. A has the goal of a positive image before third parties C
3. A has the goal of a positive self-image
4. A believes that property X is pertinent for his goal of image before B or before third parties C

ESSENTIAL CONDITIONS:

1. B perfoms an Action A
   Or else
2. B omits to perform Action A
3. A believes that this explicitly communicates or indirectly implies
4. That B attributes a flaw X to A
5. X thwarts the image that A wants to project of himself to B
   And/or to third parties C
   And/or to him/herself
6. A believes that X makes him/her inferior to B / C
   Or
7. to the category to which A wants others to believe he belongs to.
8. All of this causes A to feel
   a. a negative image emotion
      (sadness, displeasure, shame, humiliation)
   and/or
   b. a negative social emotion towards B
      (anger or rancour)
   c. a negative emotion of affiliation
      (inferiority, feeling of exclusion)
   d. a negative emotion of attachment towards B
      (disappointment about B)

AGGRAVATING CONDITIONS:

The negative emotion of A is as more dramatic as

1. The manifestation of A's flaw is public, i.e.,
   A believes that third parties C will come to know about As' flaw or inferiority
2. A believes that B's attack to A's image is deliberate
3. A has a low self-esteem
4. A's self-image is strongly dependent on the image that others (B and/or C) have of A
5. For A the goal of having a positive social
   (possibly affective) relationship to B is important
6. A esteems B

4.3. Close emotions

One of the questions in the survey inquired what emotions are closer to the feeling of offense, either in the sense of making part of the feeling itself, or of being felt at the same time.
Participants in their answers would mention humiliation (12 occurrences), and particularly often anger (50), both along with connected emotions like disappointment (29), and with antagonistic ones like sadness (25), while only one case shame is reported. Further mentioned emotions are bitterness (15) and rancour (6); and the latter is often seen, on the one side, as a final result of feeling offended, on the other as the reason why the offense may finally cause a break of the social relationship between A and B.

4.4. Causes of offense

Unlike the traditionally accepted view that people are most typically offended by insults or other overt communicative acts of discredit, the causes of offense resulting from our data show a much wider range of events (Table 1).

a. B's COMMUNICATIVE ACTION	b. B's ACTION	c. B's IMPLIED MENTAL STATE
criticism	omission	carelessness
accusation	attitude	
unjust accusation	behavior	
formal negative judgment	arrogance	
reproach	taking advantage	
insult	betrayal	preference
aggressiveness (including impoliteness)	incomprehensio n	
exclusion		distrust
negative prediction	deception, hypocrisy	
slander	injustice	
gossip		
mockery		
silence		
refusal (of an		

offer, an invitation, of listening)		

Table 1 | Causes of offense

Actually, people can feel offended not only by overt communicative actions, but also by non-communicative actions, and even by non-actions: in a word, by anything that in some way uncovers either a clearly stated or even an implied mental state from which a low opinion of B about A can leak out.

COMMUNICATIVE ACTIONS (col. a)

A person can be offended by a criticism, a slander, an unjust accusation; by gossip, insults, mockery, but also by a reproach, a formal negative judgment (like a bad score); or by a negative prediction, like in this example:

(Participant 107): *quando a venti anni mi dicevano che non avrei fatto molto nella vita*
(when at 20 people told me I would not do so much in my life)

Yet, the bulk of offensive action is exclusion

(29): *quando un professore mi ha cacciato da un esame orale*
(when a teacher sent me away of an oral examination)

And exclusion can simply be communicated by a refusal.

NON-COMMUNICATIVE ACTIONS (col. b).

Sometimes, what is offensive for A is not a particular communicative behavior, but a general attitude of B:

(35): *Quando frequentavo l'università, una mia insegna[n]te, nonché relatrice, spesso mi faceva sentire un'ignorante.*
(As I attended the university, a teacher of mine, and tutor, often made me feel an ignorant person)

One more offensive behavior is B's taking advantage of A: this makes A feel "used" like an object, not credited the dignity of a human person with her personal goals and desires.
Finally, injustice is offensive: as stated in this case.

(27): *A lavoro quando non mi è stato riconosciuto il merito di un compito svolto*
(At work when I was not acknowledged the merits of a task performed)

Being subject to injustice is offensive also for an underlying thought: how unworthy am I so as to be treated this way?

IMPLIED MENTAL STATE (col. c)

Sometimes A is offended not by what B does or does not do, but by an implicit mental state of B that can be indirectly inferred from B's communicative or non-communicative behavior (col. c). See this example:

(16): *Quando ho dato dei consigli a dei familiari ma non mi hanno ascoltato e si [sono] fidati di altri, i quali hanno fornito le mie stesse opinioni.*
(when I gave advice to relatives but they did not listen to me and trusted others, who provided the same opinions)

Here what is offensive for A is a substantive distrust of B for A that is made explicit by B's not following A's advice.

One more offensive mental state, generally implied by an omission, is the other's carelessness:

(52): *Una mia cugina non ha mantenuto la sua promessa di venirmi a trovare, e non mi ha più cercato*
(A cousin of mine did not keep her promise to come visit me, nor did look for me anymore)

That the other disregards her own promise means that you are not important for her, she does not care you and your feelings: something highly upsetting. Further, if your low importance for the other is a bad hit to your self-esteem, even more so is the comparison between how important you are for the other as opposed to other people. Thus when the other prefers someone else over you, you feel betrayed: and betrayal is not only offensive per se but mostly because A finally loses in the comparison between him/her and the rival, who is preferred by B. Like in this example:

(112): *mia sorella si è sposata e non mi ha voluto come testimone dopo che me l'aveva già chiesto*
(my sister got married and did not want me as her wedding-witness, after she had asked me to)

In the last two cases B's preference uncovers A's relative unimportance. Yet, the extreme case of this is not being acknowledged at all as a person, for example, when the other does not greet you when meeting you.

4.5. Offensive judgements

One more result emerging from the content analysis of open questions is an overview of what kinds of judgments are considered offensive. A judgment can be defined as an evaluation, that is, a belief about how we match to some standard, to some criterion of evaluation – in terms of our model, how adequate we are with respect to some goal, that is, how much power we have to achieve them – out of all the possible evaluations we elicit from others, the ones that are offensive for us are only those concerning those standards, those evaluation criteria, that we deem as relevant to the image we want to present to others and/or ourselves. But there are some criteria of evaluation that are generally important for everybody, and negative evaluations with respect to them are considered offensive. Based on previous works on the discrediting acts in political communication (Poggi et al., 2011; D'Errico & Poggi, 2012) we classified the offensive evaluations reported by our participants as four kinds of inadequacy (Table 2): a. physical (aesthetic or functional) inadequacy; b. competence (cognitive skill like knowledgebility, planning and reasoning); c. dominance (power, decisional effectiveness social influence skills); and d. benevolence (ethical qualities, altruism, honesty, morality).

Criterion of evaluation	Negative judgement
PHYSICAL	aesthetic inadequacy (physical appearance: ugly, dirty, ridiculous)
	functional inadequacy (disability

	and sexual inadequacy)
COMPETENCE	stupidity
	ignorance
	lack of skills
	lack of social skills (e.g., choosing a wrong boy-friend, wrong education of kids)
DOMINANCE	impotence
	affective irrelevance (e.g., not being respected)
	unusefulness
	dependence (e.g. a personality trait: subjective)
	inferiority (by status: objective)
BENEVOLENCE	immorality
	injustice
	cheating
	unreliability
	negligence
	egoism

Table 2 | Offensive judgements

A case in which people typically feel offended is when being teased or criticized for their PHYSICAL APPEARANCE, based on an aesthetic criterion: *fat* or *clumsy* are frequent offensive epithets, while a more "creative" one is *penguin* – an indirect way to name the same properties, reminded by a participant. Even more than aesthetic inadequacy, targeting the functional properties of a person's physical arrangement is offensive: the stygma of handicap humiliates people, hence being offensive, even when the intention of B is not to offend but, for instance, to help: pity is offensive and humiliating. Yet, no examples of this kind of inadequacy are mentioned in our corpus.
Vis à vis the criterion of COMPETENCE, participants mention stupidity, but also lack of social skills (like being told you are not able to educate your children).
Concerning DOMINANCE, the other's carelessness is offensive since it tells they consider you irrelevant. You feel totally inconsequential when you are considered or explicitly accused to be useless, but also when people ignore you, or they omit those simple acts that credit you dignity and consider you worth of respect. Further offensive attacks to a person's DOMINANCE are when others consider them inferior.
Coming to BENEVOLENCE, people are offended by accusations of immorality, of being a liar or an unreliable person, of negligence or non-compliance with one's duty, and finally of selfishness.

5. Lexicometric analysis

The automatic quanti-qualitative analysis was performed on the subjects' answers by TalTac (Trattamento Automatico Lessicale e Testuale per l'Analisi del Contenuto, i.e."Lexical and Textual Automatic Processing for Content Analysis": Bolasco, 2016), a software for textual data analysis based on a "lexicometric approach": an application of statistical principles to textual corpora. The "textual statistics" aims to extract the semantic level in a text starting from the list of words obtained by statistical analysis; for example, in the analysis of specificities, the software extracts a list of significant words obtained by a statistical comparison between sub-parts of text according to selected variables.
TalTac allows to extract the 'peculiar lexicon' (Bolasco, 2016) that is, the set of the words over-represented or under-represented in the text under analysis, by comparing the corpus to an external frequency lexicon, taken as a reference model (in our case, Standard Italian). The measure of the variance from the reference lexicon is represented by the standard deviation, which is the deviation between the form frequencies in the analysed text and in the frequency lexicon.
The corpus extracted from all answers in our survey is particularly feasible to lexicometric analysis since it allows to extract cognitive elements of emotions through the words used by people (Poggi & D'Errico, 2010; D'Errico & Poggi, 2014). This present corpus counts 14061 (N) occurrences with 3003 (V) different words and a medium lexical richness index [(V/N)*100], equal to 21,35%.
To go more into the categories emerging from our content analysis, we run a lexicometric analysis on the whole corpus of our participants' answers. This allowed us to single out all the words fitting the three categories above, with their relative frequencies: Emotions (emotional states), Actions and Judgements. In addition another interesting category emerged: one of Stable traits that may define either the Offender (O) or the Victim, i.e., the Offended Person (V), and that may be the cause or the condition of the feeling of offense or of the other emotional state.
We coded a total of 492 words by also reading at the real context of use (the actual sentences in which they occur and then grouped considering the semantic similarity) and we found that in our corpus the great part of this lexicon is related to the description of the emotional states (48%), followed by the actions that caused them (29%) the judgments (15%) and finally the stable traits of Offender or Victim that may have caused or allowed the emotional states.

ACTIONS	TRAITS	EMOTIONS	JUDGMENTS
29%	8%	48%	15%

The frequency configuration of the most frequent words coded as Emotions is summarized in Figure 1.

Figure 1. Emotions caused by the offense

From Figure 1 it emerges how feeling offended can be a negative state associated mainly with low arousal emotions like *delusa* (disappointed), *triste* (sad), *ferita* (wounded), *rimanerci male* (take it bad), *amareggiata* (embittered), *rancore* (rancor), *umiliato* (humiliated), *dispiaciuta* (sorry), *scoraggiata* (discouraged), *mortificata* (mortified), *imbarazzata* (embarrassed), *sofferenza* (suffering); but it also looks connected, although to a lesser extent, with words mentioning high arousal emotional states, such as *arrabbiata* (angry), *tensione* (tension), *frustrate* (frustrated). These opposite reactions to the feeling of offense can be linked to different levels of individuals' self-esteem, in that a passive reaction to an offence is mainly linked to low levels of self-esteem (Poggi & D'Errico, 2018).

As to the actions mentioned in the corpus, it emerges (Fig. 2) that the offense can be mainly caused by negative judgments expressed in different ways, ranging from *accusato* (accused), *giudicato* (judged), *rimproverata* (reproached), *rimprovero* (reproach), *accusata* (accused), *critica* (criticism), *denigrato* (denigrated), *deriso* (teased), *insult* (insults), to different judgments implied by provocation or jokes: *ridere* (laugh), *scherzava, scherzare* (joke), *battuta* (humorous joke), *frecciatina* (sideswipe), that can be a more indirect evaluation. Words that are linked to causes of feeling offended can be related to a lack of consideration, in this case people feel offended because they are *lasciato* (forsaken), *ignorato*, *ignorandola* (ignored), *sminuito* (diminished), *allontanato* (distanced), *isolato* (isolated). A less common cause of offense are relational problems like *tradito* (betrayed), *incomprensione* (incomprehension), *frainteso* (misunderstood), or aggressive actions such as harsh and impolite manners: *scortesi* (impolite), *urlato* (shouted), *strillato* (screamed), *in faccia* (in your face). Yet, as is clear from the high frequency showing in Fig.2, a word recurring very often is *tradito* (betrayed): a very serious and heavy action. This again is consistent with the above observation about the word *amicizia* (friendship): what most typically enhances the suffering of feeling offended is that a person to whom we feel bounded unexpectedly violated our expectation about us, betraying what we feel as a previous affective commitment s/he had with us.

Besides emotions, in the corpus *amicizia* (friendship) has a good frequency among the words used by participants; this can be accounted for by the fact that, as clearly demonstrated in the whole work (Poggi & D'Errico, 2018), the most serious cause of offense is when the offensive action comes from persons with whom the Victim has an important affective relationship: we are more offended by friends than by strangers.

Figure 2. Actions causing the offense

The category of Judgments (Fig. 3) collects all the negative evaluations mentioned in the corpus that include lack of benevolence – *falso* (false), *cattiveria* (badness), *bugiarda* (liar), *vigliacca* (coward), *cattive* (bad), *opportunismo* (opportunism), *egoismo* (selfishness) – lack of dominance – *sciocchezze* (nonsense), *debole* (weak), *inutile* (useless), *cazzata* (crap), *inadatto* (unfit), *mediocrità* (mediocrity), *passiva* (passive), *incapace* (unable), – negative physical features – *peso* (weight), *aspetto* (look), *brutte, brutto* (ugly), *fisicità* (physical), *difetti* (flaws), *complessate* (full of complexes), *bassezza* (shortness), *ingrassata* (fatter), *grassa* (fat), *estetico* (aesthetic), *robusta* (stout) – and to a lesser extent lack of competence – *stupida* (stupid), *incompetente* (incompetent).

Figure 3. Offensive judgments

A last category, not considered in the preceding work, that pops up from lexical frequencies, includes adjectives or nouns mentioning mental properties of person, that may be either the Victim or the Offender. Examples of the first kind are *suscettibilità* (hypersensitiveness), *permalosa* (touchy), *vulnerabili* (vulnerable) and *orgogliosa* (proud). As resulted from the content analysis by Poggi and D'Errico (2018), a particularly touchiness of the Victim, possibly favored by a low self-esteem, makes him or her more vulnerable to feeling offended. At the same time, one who is *orgogliosa* (proud) considers the goal of being appreciated and well judged by others particularly important, hence lowering the threshold of feeling offended. On the other hand, words like *superior* (superior) *superiorità* (superiority) and *presuntuosa* (presumptuous) describe the Offender, who tends to offend other people because displaying or boasting one's superiority, so much so as to be seen (by the Victim) as presumptuous.

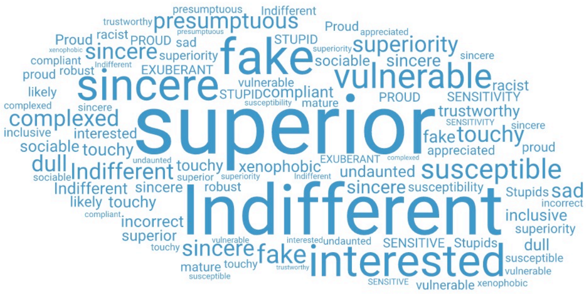

Figure 4. Offensive Traits

3.6. Discussion and conclusion

This work was aimed at investigating a neglected but widespread emotion: feeling offended. After overviewing previous research on offense, who considered feeling offended mostly as a blow to the person's public damage, we have presented a survey study on the feeling of offense, aimed at deepening the very definition and the description of feeling offended. The definition emerging from the content analysis allowed us to code and group the more frequent words used in the corpus and, though a lexicometric analysis, to describe the emotions connected to the feeling of offense, the individual traits of offender and victim, the causes of offense, and the types of judgment that most typically appear offensive. Beside working as a description of this emotion, the present work might be further applied to detect the feeling of offense in other corpora, and thus to discover implicit or past conflicts not completely solved, by allowing also a possible annotation of their seriousness. The automatic extraction of a generalized online feelings of offence in intergroup conflicts, for example, can be used to project a more emotionally focused reconciliation where 'offended group' need to be empowered by restabilising their psychological sense of control (Nadler, 2002). In this sense within social media (Carr, 2017) the emotional analysis can contribute to turn different forms of opinions polarization in a deep understanding of offended people's needs.

REFERENCES

[1] Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles?. Psychological science in the public interest, 4(1), 1-44.

[2] Bolasco S. (2016). L'analisi multidimensionale dei testi. Carocci Roma

[3] Brown, J. D., & Marshall, M. A. (2001). Self-esteem and emotion: Some thoughts about feelings. Personality and Social Psychology Bulletin, 27(5), 575-584.

[4] Carr, C. (2017). Social Media and Intergroup Communication. Oxford Research Encyclopedia of Communication.

[5] Castelfranchi, C. (2000). Affective Appraisal versus Cognitive Evaluation in Social Emotions and Interactions. In Paiva, A. (ed.), Affective Interactions. Towards a New Generation of Computer Interfaces. Berlin: Springer, pp.76-106.

[6] Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: An" experimental ethnography.". Journal of personality and social psychology, 70(5), 945.

[7] D'Errico, F., & Poggi, I. (2012). Blame the opponent! Effects of multimodal discrediting moves in public debates. Cognitive Computation, 4(4), 460-476.

[8] D'Errico, F., & Poggi, I. (2014). Acidity. The hidden face of conflictual and stressful situations. Cognitive Computation, 6(4), 661-676.

[9] D'Errico, F., & Poggi, I. (2016). Social emotions. A challenge for sentiment analysis and user models. In Emotions and Personality in Personalized Services (pp. 13-34). Springer International Publishing.

[10] Foster, C. A., & Rusbult, C. E. (1999). Injustice and powerseeking. Personality and Social Psychology Bulletin, 25(7), 834-849.

[11] Lewis, M. (2008). Self-conscious emotions: Embarrassment, pride, shame, and guilt. In M. Lewis, J. Haviland-Jones, & L. Feldman Barrett (Eds.), Handbook of emotions, 3rd ed. (pp. 742-756). New York:Guilford Press.

[12] McCullough, M. E. (2000). Forgiveness as human strength: Theory, measurement, and links to well-being. Journal of Social and Clinical Psychology, 19(1), 43-55.

[13] Miceli, M., & Castelfranchi, C. (2014). Expectancy and Emotions. Oxford: Oxford University Press.

[14] Mosquera, P. M. R., Manstead, A. S., & Fischer, A. H. (2002). The role of honour concerns in emotional reactions to offences. Cognition & Emotion, 16(1), 143-163.

[15] Nadler, A. (2002). Social-psychological analysis of reconciliation: Instrumental and socio-emotional routes to reconciliation. Peace education worldwide: The concept, underlying principles, the research, 127-143.

[16] Paleari, F. G., Regalia, C., & Fincham, F. (2005). Marital quality, forgiveness, empathy, and rumination: A longitudinal analysis. Personality and Social Psychology Bulletin, 31(3), 368-378.

[17] Poggi, I., & D'Errico, F. (2010). The mental ingredients of Bitterness. Journal on multimodal user interfaces, 3(1), 79-86.

[18] Poggi, I., D'Errico, F., & Vincze, L. (2011). Discrediting moves in political debate. In Proceedings of Second International Workshop on User Models for Motivational Systems: the affective and the rational routes to persuasion (UMMS 2011) (Girona) Springer LNCS (pp. 84-99).

[19] Poggi, I., & D'Errico, F. (2018). Feeling offended. A blow to our image and our social relationship. Vol. 8, p.2221, *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2017.02221

[20] Zander, W. (1976). Taking offence as seen in depth-psychology. Psychotherapie und Medizinische Psychologie.

An experiment with an off-the-shelf tool to identify emotions in students' self-reported accounts

Lubna Alharbi, Floriana Grasso, and Phil Jimmieson
Department of Computer Science, University of Liverpool

Abstract. Identify student emotions is considered a critical tool for effective learning. Many studies analyse emotions in the learning process, though often they employ sophisticated techniques such as facial recognition and blood pressure measurement. We focus on the more practical scenario of identifying emotions from text based interactions between student and lecturer, as common in many online-learning environments. We present a preliminary discussion on the problem, starting from an experiment in using an off the shelf tool, Synesketch, to identify emotions in text describing educational situations, from the ISEAR dataset.

1 INTRODUCTION

The main aim of our research is to explore the use of emotion analysis to help enhance both students' and teachers' experience in an on-line learning environment. The idea is to complement the role of the on-line instructor, by providing information on behaviour patterns and emotional states of the students, in order to eventually suggest strategies to adjust the communication, or to reach out, and ultimately motivate students for better learning.

Currently many sophisticated systems utilize keyboard, cameras, and microphones on the students' computers to recognize student's emotions [9]. Other go as far as analysing student's emotions expressed in posts on social networks. Whilst these are all interesting avenues, we want to focus on a scenario in which the communication happens in an online, text based classroom. We therefore focus on recognizing the emotion from the text messages interchanged between student and teacher. Commonly students communicate with the lecturer asking questions, complaining, or seeking pieces of advice. This happens over time, during the course of an entire module, and possibly a programme of study. Therefore, considering emotions in these interchanged messages might help develop an emotional profile for students which could in turn help identify not only whether the student is manifesting a specific distress or emotional engagement at a particular moment in time, but also monitor the situation over time, and identify changes in patterns.

This paper report work in progress on this project, and we start by framing the context of interest. Then we discuss the studies related to the context of this research. Next, we will discuss the performance of a prototypical tool with a sample dataset extracted from the ISEAR database, and we conclude with some discussion.

2 MOTIVATION

The context in which we frame our discussion is the one of a suite of online MSc degree programmes, described more fully in [15]. Whilst they are not by any means unique in the landscape of online education, a description of the environment is useful here to identify the features that we want to focus on. Within these programmes, eight 15 UK credits modules are taught to classes of no more than 20 students, who interact asynchronously with each other and their instructors in an online learning environment (OLE). Each module is divided into "seminars", which take place over the period of one week. The pedagogical model is heavily based on a constructivist approach, with a good proportion of the graded components being based on discussion around an open question, moderated by the instructor. At the end of the taught modules, the students engage into a research project, which lasts 40 weeks. The first weeks of this project happen in an online classroom, in the same structure of other modules, and is facilitated by a "general dissertation advisor", with the aim of getting to the production of a suitable project proposal. Then, students move to the dissertation proper, under the guidance of their dissertation advisor. For this latter phase, the students experience a drastic change in the learning environment, moving from a highly structured online interaction with peers, with weekly deadlines, to a more self-paced interaction, over a long period of time, with their dissertation advisor only, and with only a few interim milestones in between. Some students are affected by this change more dramatically than others, and at times their engagement fades.

In addition to instructors, outside the online classroom, student progress and support is the responsibility of a *student support manager (SSM)* who takes on most of the roles of a personal/academic tutor (although again, of course, the role is effected in an online mode). The instructors will report problems and absences to the SSM for further action.

All communication, in the classroom between students and module instructor, or as part of the virtual discussion among peers, or in the later dissertation stage, and out of the classroom, between students and their SSM, and between instructors and SSMs, is logged in the OLE.

Therefore, we are in the presence a very rich environment in which there is a great number of people/roles around a student, each with a limited view of the student's situation, both academic and personal, but each crucial in contributing to the student's success. It is in this context that we hope to be able to provide support by building a mechanism to integrate all these communication snippets, and identify issues related to student motivation and emotional engagement. This support is not necessarily aimed at a direct intervention in the classroom, which could be counterproductive for the teachers [11], but rather at identifying long term issues which may span several modules, and therefore months, or, conversely, to spot a change in behaviour which would not be identified when looking at one class or one interaction only.

3 RELATED WORK

Text exchanged between lecturer and student may exhibit a number of emotions [1-3] that can potentially help identify student's engagement issues. The use of emotion in distance learning is not new, there are a number of sophisticated software for distance learning, such as FILTWAM [4], which is equipped with an intelligence system to read students emotions and correspondingly adjust the system interface to motivate students. However, often systems have, as a main focus, a strategy of direct intervention on the student's learning, and do not consider the ways in which lecturers can be put in the picture. Our point of departure is that lecturers and advisors should be kept at the core of education process, next to the students, as they have the most effective role to motivate students. This research therefore seeks to provide academics with mechanisms to track students' emotions, and give them the opportunity to intervene, for instance by adjusting the content dynamically or changing their teaching approach, or for advisors to reach out in different ways. This contrasts to solutions where the intelligent tutoring system intervenes directly, perhaps with tips or suggestions on tasks to perform, to alleviate the student's issues.

Ferguson, et al. [5] highlighted the importance of understanding the capabilities of students through many tools, such as activity-Based Assessment (ABA), where teachers keep monitoring students and assess them through series of questionnaires. Similarly, Otus learning system [6] help lecturers and students build interest profiles, which keep track of the interests of students (in courses, activities, books, and sports). This profile, which is manually created by lecturers and students and then updated electronically and manually, helps teachers develop most effective teaching programmes based on students' interests. Other work by Munezero and Montero [11] concentrates on the diaries of students to extract their emotions. However, diaries are only written for personal use, and may not consider, nor being addressed to, lecturers.

The procedure followed by Ganotice, et al. [7] to create students' emotion profile was through questionnaire; which is cumbersome and will be costly to be collected time and again. This approach is also more suitable to face-to-face learning than online learning.

In this study, we discuss the production of a system that can identify and monitor emotions of the student in an e-learning environment. This study builds on the work by Shen, et al. [10] who integrated the Shanghai e-Learning platform, with learners' emotions. Their model is guided by Affective Learning Model. The upper part of the model is a cognitive appraisal model for emotions, and the lower part of the model is a physiology recognition of emotions where these two models converge. Bayesian Networks were employed to model the relations between the emotional states and their causal variables [10].

4 USING AN EMOTION CAPTURING TOOL ON THE ISEAR DATASET

As our objective is not a direct intervention by an intelligent tutoring system, but rather a profiling tool, we seek practical solutions to the analysis and integration of several instances of interaction, among all different roles involved with the student's experience, including peers.

In this section we discuss the use of Synesketch [7] as a tool for emotion capturing. The advantages of this tool is that it is free, open source, and can be integrated with exchange message systems such as Skype. The tool has been used with many projects related to emotion extraction [1, 2, 9] and therefore constitutes a valid benchmark. To recognize the emotions in the text, a hybrid combination of a keyword-spotting method and a rule-based method is used [2]. Keyword-spotting approach is based on the use of a lexicon of words and expressions related to emotions. A survey-based word lexicon is utilized to automatically search WordNet for all semantic relatives of the initial word set. Common abbreviations and informal language common in Netspeak such as emoticons or acronyms like "ROFL" are included. These are useful as, while the interaction with teachers tend to be formal (with less emoticons etc) we also need to capture other types of interaction that are more likely to exhibit a personal/less formal style. Each lexicon (word or emoticon) is attached with six emotional weights that correspond to six basic emotional categories defined by Ekman [12]: happiness, sadness, anger, fear, disgust, and surprise. Those emotions are commonly considered in educational literature [13].

In order to assess the suitability of this tool, we used ISEAR (http://www.affective-sciences.org/researchmaterial). This is a dataset collected in a project led by Klaus R. Scherer and Harald Wallbott [12]. The data were collected from students who had been asked to recall situations which aroused one of seven major emotions (joy, fear, anger, sadness, disgust, shame, and guilt), by also reporting on how they appraised the situation and how they reacted. The dataset comprises these recollections from around 3000 respondents, all over the world. We chose this dataset as it was likely to include, given that the respondents were students, sentences related to an educational situation. For our study, we manually extracted 500 records related to education scenarios, such as records that described students' emotions to their educational performance, classmate, lecturers, university, and exams. From these, we wanted to focus on subsets which were self-labelled with the same emotion. For this paper, we show the results of an analysis of the "anger" emotion, as we think it provides interesting features that raise some more general considerations. 50 extracted records tagged with the "anger" emotion were analysed through Synesketch, for their emotional type (happiness, sadness, anger, fear, disgust, and surprise), their weight, or intensity of the emotion, and their valence (a positive or negative emotion).

A sample of those records and results of Synesketch are presented in table 1.

Table 1: sample of records from ISEAR, that are tagged with anger emotion

Entity	include	Val	Anger	Disgust	Fear	Happyness	Sadness	Surprise	Text
Classmate	No-keyword	-1	0.04	0	0.06	0	0.06	0	A classmate told me I must have bribed the class leader to let me go to your English lecture.
Classmate	No-keyword	0	0	0	0	0	0	0	My class leader told me I am not chosen for your English lectures.
Teacher	keyword	-1	0.40	1.0	0	0	0	0	When I talked with a teacher yesterday who, to say the least, was rude and unwise and had irrelevant opinions about a friend.
Classmate	Keyword	-1	1.0	0.05	0.05	.12	0.09	0.12	I felt anger against a colleague of mine during a rehearsal in acting. He hadn't learnt the text of an opera act in the course of several months and thus making difficulties for the rest of my college
Teacher	No-keyword	1	0	0	0.09	1.0	0.33	0	In a course I thought that I deserved good marks but I only got ordinary marks with no justification as to why I was given these marks
Oneself	No-keyword	-1	0.20	0.20	0.20	0.27	0.40	0	We were just about to go into the Exam room and I didn't see all my writing materials plus the identification card from the place where I left them
Oneself	keyword	-1	0.14	0.11	0.16	0.13	0.35	0	When I left after the examination to enter the University, and even though I had studied the whole year I made a bad exam
Oneself	keyword	-1	100	0.15	0.15	0.10	1.0	8	In the University we were convoked to a stroll like a freshman, I was still innocent about the manipulation. On seeing the goal of the exaltation to poor character people, I felt anger on feeling myself mass of maneuver.
Teacher	keyword	-1	0.13	0	0.06	0.11	0.13	0.06	When I felt being treated unjustly by a teacher.
Oneself	keyword	-1	0	0.05	0	0	0.24	0	When I got low marks in B.Sc final.
Oneself	No-keyword	-1	0.14	0.11	0.16	0.14	0.35	4	As I usually do not start learning until a short time before an examination, I once made up my mind to try to work for a longer time. But once again I did not do it - and got a bad mark
Teacher	No-keyword	-1	0	0	0	0.13	0.20	0.11	I remember that when I was in school I saw a case of partiality - one of the teachers gave private coaching classes to some girls and she would give these girls extra attention in class and would ignore the others.

University	No-keyword	-1	0	0.13	0	0.27	0.27	0	I was not approved to continue my studies at Moscow University, no matter that I had the highest marks and in general the best records of all the candidate
Teacher	keyword	1	0	0	0	0.13	0	0	I had made an error in planning a program and had publicly accepted the mistake, despite repeated requests not to bring the matter for discussion a professor kept constantly passing remarks. I reacted angrily.
University	No-keyword	0	0.08	0.08	0.08	0.17	0.17	0	When I was punished in school for no serious mistake of mine.
Exam	No-keyword	-1	0.06	0.05	0	0.09	0.24	0	When I got a low grade in an administration course.
Classmate	No-keyword	-1	0.10	0	0	0.10	0.03	0	At school, a couple of years ago, a so-called acquaintance told lies about me to a teacher I was told about it by friends who overheard the conversation
Teacher	keyword	-1	1.0	0.27	0	0.16	0.20	0	I had studied for almost one week for my physics-examination. With difficulty, I passed the exam. I was angry about the teacher and also about myself because I had not remembered enough during the exam and because the time that I spent studying was wasted.
Teacher	keyword	-1	1.0	0.27	0	0	0.05	0	I was furious when the teacher pointed out in class that I was the quietest girl in class.

We focused on five "entities" or topics that the students refer to, which can be identified in the records, these are in Table 2.

Table 2: type of entities within students' text

Entity	Frequency	%
classmate	14	28.0
lecturer	12	24.0
oneself	11	22.0
university	7	14.0
exam	6	12.0
Total	50	100.0

The valence (or sentiment) generated by Synesketch is summarised into three categories (-1 for negative, 0 for neutral, 1 for positive). In table 3, it could be seen that, Synesketch managed to recognize 40 records as negative, which presents 80% accuracy as in ISEAR they are all tagged as negative. Of the 20% of the records which were not classified, 12% were actually not identified, and 8% did not show any negative emotions. For instance; records "*My class leader told me I am not chosen for your English lectures*" and "*When the university withdrew the accommodation that it had given me*" do not include keywords to help Synesketch classify them with negative emotions.

Table 3: sentiments found with Synesketch

Valence		Frequency	%
Valid	negative	40	80.0
	neutral	5	10.0
	positive	5	10.0
	Total	50	100.0

Among the 50 records, it was found that 52% do not include an explicit keyword that can reveal the emotion (table 4). However, Synesketch managed to recognize correctly 80.76% of emotions from these records.

Table 4: samples with keywords

Keyword		Frequency	%
Valid	No-keyword	26	52.0
	keyword	24	48.0
	Total	50	100.0

Few instances showed Synesketch failed to classify emotions. For instance, this anger-tagged-record "*In a course I thought that I deserved good marks but I only got ordinary marks with no justification as to why I was given these marks*" has been classified by the tool as "happy" with score 1.0, and "sad" with score 0.33. This is because of the emotional word "good", which would be difficult to classify otherwise due to the lack of other keywords. However, the negated word "no justification" helped to identify the negative emotion.

In another instance, Synesketch gives the same score, of .27, for "happy" and "sad" to the record "*I was not approved to continue my studies at Moscow University, no matter that I had the highest marks and in general the best records of all the candidate*". The same score for both emotions might be due to the presence of "highest, and best" as indicator of happiness, and "not approved" as indicator of sadness. Similarly, the emotional word "bad" was categorized as sadness.

Or, Synesketch classified the emotion in record "*I had made an error in planning a program and had publicly accepted the mistake, despite repeated requests not to bring the matter for discussion a professor kept constantly passing remarks. I reacted angrily*" as positive, more precisely "happy", despite the record has many negative keywords such as "error", "mistake", and "angrily".

A crucial issue here is aggregation: whenever there are many sentences with different emotions, the tool aggregates the score and gives results that show high score for many emotions. For instance, for record "*we were taking our mock exams and someone else (a friend) was making noise when the examiner picked on me and said that he would tear up my answer sheet. I got very angry with the girls involved*", the scores were (anger = 1, disgust = 0.7, fear = 0.9, happiness = 0.8, sadness = 0.8, and surprise = 0). However, the sentiment of this record was negative, which is consistent with the tag for this record in ISEAR.

It is therefore clear that we need to consider models that use a wider context, both in terms of the actual textual context, as suggested by Altrabsheh, et al. [14], but also, more crucially we believe, on a longitudinal data and learning analytics for a student or population of students.

6 CONCLUSIONS AND FURTHER WORK

In this paper, we reported on a preliminary experiment on using an off-the-shelf emotion recognition tool, Synesketch, for education related messages. The tool could be used to classify the text into positive, negative or neutral. However, the tool was less reliable in the classification of emotions. One important issue was the aggregation strategy of the tool that suggests that better results could be achieved by looking at the passage or message level, rather than sentence level, and this is recognized by other research.

However, what is, we believe, less explored is the role of profiling in this task. We plan to build a richer model that could support emotion recognition on the basis of longitudinal data from the same student, and from the same module, to be able to capture changes in the interaction style or tone, and therefore identify more relevant and specific issues. In order to progress in our investigation and development, more relevant datasets need to be collected, especially coming from the scenario we are basing our framework on.

We believe that for realistic and complex online learning environments, being able to respond to a student's emotion when it is manifested with a short term remedial action is not enough. What would be impactful is identifying changes in the student's behaviour, which are not due to the spur of the moment, but are signs of deeper, longer term issues. This can only be achieved by looking at the student's interaction overtime, and with different interlocutors. We ultimately seek to produce an integrated model that can put together academic and personal interchanges towards the construction of an emotional/motivational student profile.

Incidentally, and to conclude, this model would not be necessarily limited to supporting students: burn-out among online teachers is a well recognised issue [16], and we can picture this approach to be useful in that context too, thus contributing to a better experience for all stakeholders in online learning.

REFERENCES

[1] J. Gonzalez-Sanchez, M. Baydogan, M. Chavez-Echeagaray, Elena, R. Atkinson, K, and W. Burleson, "Affect Measurement: A Roadmap Through Approaches, Technologies, and Data Analysis," in *Emotions and Affect in Human Factors and Human–Computer Interaction*: Academic Press, 2017, pp. 255–288.

[2] J. Sánchez-Rada, Fernando and C. Iglesias, A, "Onyx: A Linked Data approach to emotion representation," *Information Processing and Management,* no. 52, pp. 99–114, 2016.

[3] F. Tian *et al.*, "Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems," *Knowledge-Based Systems,* no. 55, pp. 148–164, 2014.

[4] K. Bahreini, R. Nadolski, and W. Westera, "Towards multimodal emotion recognition in e-learning environments," *Interactive Learning Environments,* 2014.

[5] D. Ferguson *et al.*, *Designing Personalized Learning for Every Student*. Alexandria, VA: ASCD, 2001.

[6] K. Khaleel. (2017). *Learning Profile Captures Student Interests*. Available: https://otus.com/learning-profile-captures-student-interests/

[7] F. Ganotice, J. Datu, Alfonso, and R. King, "Which emotional profiles exhibit the best learning outcomes? A person-centered analysis of students' academic emotions," *School Psychology International,* vol. 37, no. 5, pp. 498-518, 2016.

[8] U. Krcadinac, P. Pasquier, J. Jovanovic, and V. Devedzic, "Synesketch: An Open Source Library for Sentence-Based Emotion Recognition," *IEEE Transactions on Affective Computing,* vol. 4, no. 3, pp. 312 - 325, 2013.

[9] A. Alberdi, A. Aztiria, and A. Basarab, "Towards an automatic early stress recognition system for office environments based on multimodal measurements: A review," *Journal of Biomedical Informatics,* no. 59, pp. 49–75, 2016.

[10] L. Shen, M. Wang, and R. Shen, "Affective e-Learning: Using "Emotional" Data to Improve Learning in Pervasive Learning Environment," *Educational Technology & Society,* vol. 12, no. 2, pp. 176–189, 2009.

[11] M. Munezero and C. Montero, Suero, "Exploiting Sentiment Analysis to Track Emotions in Students' Learning Diaries," presented at the 13th Koli Calling International Conference on Computing Education Research, 2013.

[12] P. Ekman, "Facial Expression and Emotion," *Am. Psychologist,* vol. 48, pp. 384-392, 1993.

[13] F. Tian, P. Gao, L. Li, W. Zhang, H. Liang, and Y. Qian, "Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems," *Knowledge-Based Systems* no. 55, pp. 148–164, 2014.

[14] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Predicting Students' Emotions Using Machine Learning Techniques," in *AIED 2015, LNAI 9112*, 2015, pp. 537–540.

[15] Grasso, F., Leng, P.: Issues of quality in on-line degree programmes. In Ghaoui, C., ed.: Usability Evaluation of Online Learning Programs. Idea Group Publishing (2003) 198–217

[16] Hogan, R. Lance, and Mark A. McKnight. "Exploring burnout among university online instructors: An initial investigation." The Internet and Higher Education 10.2 (2007): 117-124.

Prosocial words in social media discussions on hosting immigrants. Insights for psychological and computational field.

Francesca D'Errico[1], **Marinella Paciello**[2], **Matteo Amadei**[3]

Abstract. The present work is aimed at exploring prosocial social media stances by means of real words used in on line discussions. In particular within online dynamics, the aim of the present study was twofold: 1) investigation of prosocial orientations emerging from a particular case of supportive communication on immigration promoted by a public figure (Morandi, a famous Italian singer). The methodology used is based on a machine learning approach combined with a psycho-lexical analysis of online discussions; more than 5 thousand prosocial comments were identified allowing to extract the peculiar words of people who stand for hosting immigrants. Results showed peculiar 'prosocial' moves expressed by means of identified group of words such the disconfirmation of stereotypes, the reflexive move, the recalling of universal values and the processing of causes and possible problems of immigration. The used lexicographic approach allows to develop a prosocial lexicon aimed at extracting the possible prosocial orientation also in other contexts exploring the peculiarity, and to develop a general model for user generated content in social network.

1 INTRODUCTION

Many researchers have recently focused on the so-called 'dark side' of social media by pointing out negative online behaviours (Hutchens, 2014; D'Errico et al., 2014), such as flaming and cyberbullying, by neglecting the understanding of dialectical nature of on line discussions that includes negative but also positive stance toward a possible moral topic. Within online dynamics, the aim of the present study was to investigate the prosocial orientation expressed by the used words starting from a particular case of supportive communication on immigration promoted by a public figure (Gianni Morandi, a famous Italian singer). The 'Morandi case' involved a public communicator, who, following a serious accident at sea that caused the death of more than 700 people, tried, by means of a public message on his Facebook page, to promote a prosocial orientation of his followers towards hosting immigrants. This post caused a popular debate, where people decided to agree or disagree with the immigrants hosting. The words used by people to argued their positive stance toward hosting immigrants are potentially linked a willingness to provide help, that in moral psychology can be acknowledged as the domain of care (Graham et al., 2001; Haidt and Joseph, 2007).

More specifically, the aim of the present work is to explore the context of social media discussions on hosting immigrants by extracting prosocial (i.e. pro hosting immigrants) words used to express users' potential ethical orientation. To this end we can refer to the theoretical field of moral psychology which includes central notions like 'moral agency' (Bandura, 1991; 2015) and 'moral agent' (Conte & Dignum, 2001), as representing the capability to recognise and adhere to acknowledged standards, focusing on the online ethical orientation of accepting and helping immigrants.

In particular according to Bandura's definition of moral agency (Bandura, 1991; 2015) people are active agents and, through self-regulatory processes, are able to control their moral behaviour – namely, to align moral conduct with moral and ethical principles. More specifically, the moral self-regulative process operates by means of the *human agentic functions* such as self-monitoring, self-evaluation and self-reflection, and it can take a proactive or inhibitive form. The proactive process fosters moral action by adjusting it with personal and social standards and by anticipating the positive emotional self-evaluative reactions: for example, when people believe it is important to promote the respect of human rights, they are proactively engaged in a plan of action aimed at protecting and defending this principle, and, in so doing, they anticipate positive emotions such as pride and satisfaction, derived from the possible advantageous consequences of their choices and actions. The inhibitive process instead hinders the engagement in immoral behaviours by considering them ethically and socially sanctionable and by anticipating the negative emotional self-evaluative reactions. For example, considering the previous case of human rights, the people refrained from action by not being in line with this principle, and so they avoided the possible negative emotions such as guilt and shame that may have resulted from the negative effects of their unethical action (Bandura, 1991). In other terms, moral self-regulation operates to promote ethical conduct, such as helping and caring for others in need, and to prevent unethical behaviour, such as harmful conduct, allowing people to remain anchored to the norms and principles that guide their moral behaviour.

In line with this theoretical framework we believe that in the case of prosocial position toward hosting immigrants people should be able to regulate their moral position in online settings through executive functions. Specifically we expected that they are able to recognize and recall moral standards and principles usually guiding moral decision, to evaluate the actual effect of hosting and to anticipate the negative consequences of unethical choices, and finally to take the victim perspective by recognizing him/her human qualities and him/her states of need. Moreover, it

[1] Dept. Philosophy Communication and Visual Art, Univ. Roma Tre, Via Ostiense 234, Rome (Italy). Email: {Francesca.derrico}@uniroma3.it.

[2] Psychology Faculty, Unienttuno University, Corso Vittorio Emanuele II, Email: {m.paciello }@uninettunouniversity.net.

[3] Dept of Engineering Artificial Intelligence Laboratory, Roma Tre University, Via della Vasca Navale 79, Rome, Italy.
Email : {matteo.amadei}@uniroma3.it

is likely that the moral regulation occurs with emotional regulation (D'Errico & Paciello, 2018) that allow to manage negative emotions that could emerge during online heated discussions, where diverging positions are strongly defended. The occurrence of moral and emotion regulation should guarantee emotional closeness and empathy towards those in a state of need. On the contrary, as attested in previous study (D'Errico & Paciello, 2018), when moral thought is wavering, moral disengagement mechanism emerges and creates a "divorce" between moral actions and moral principles that should regulate decision making and conduct. In this case it is probable to find a strong negative emotional activation, more difficult to regulate, and several justifications and argumentations for unethical position. A recent work (D'Errico & Paciello, 2018) has pointed out that in online settings moral disengagement mechanisms focus on victims (dehumanization and attribution of guilty) are very frequent, mostly when people express high level of hostile emotions.

Within these contexts the perception to share or not a common 'destiny', during the conversation, become central in the activation of 'moral control' and thus on prosocial messaging. D'Errico (2016) tested that within social networks disadvantages groups when feel to be a part of a common group can share normative and prosocial argumentation in order to really face common real problems; on the contrary thematic pages where people do not know each other and thus express opinions as part of (potential) ideological and differentiated groups can activate process of polarization between groups (Sia et al., 2002), since they feel part of opposed groups (Hogg et al, 1990).

These considerations led us to deep and identify the possible prosocial moves also in online discussion regarding hosting immigrants, where people do not know each other, and they are in a distributed online context.

2. METHOD

2. 1 Online communicative scenario

The communicative scenario identified concerned a message posted by Gianni Morandi, a popular Italian singer, on April 21st, 2015. The post was aimed at fostering closeness by means of social media (Facebook) between Italians and immigrants after a serious shipwreck in a channel of Sicily between the night of the 18th and 19th of April, causing the death of more than 700 migrants. (https://www.theguardian.com/world/2015/apr/19/700-migrants-feared-dead-mediterranean-shipwreck-worst-yet) The message was composed of text and an image. Both text and image had the goal of generating feelings of closeness and similarity through the comparison between the Italian emigration of the early twentieth century and the plight of the immigrants coming from Africa in recent years:

'About migrants and emigrants, we must never forget that thousands and thousands of Italians, in the last century, have left their homeland to America, Germany, Australia, Canada ... with the hope of finding work, a better future for their children, because in their country they could not get it with the humiliation, harassment, abuse of power and violence, who have had to endure! It was not that long ...'

2.2 Data and Procedure

A total of 12,583 comments made in response to Gianni Morandi's post, distributed from the 21st to the 27th of April 2015, were extracted through Facebook API. For encoding, comments directed at Morandi, both positive and negative, the simple expression of agreement or disagreement (Poggi et al., 2013) and more than 400 comments containing links to videos or images, were manually filtered. In addition to this coding, posters who simply denied the potential comparison between the two emigrations or self-celebration of one's own national identity were excluded, corresponding to about 2,805 posts.

Data cleaning carried out in this way led to the identification of two main categories: *prosocial and antisocial comments by excluding mere and personal vents*. They were coded by two judges reaching a very goog agreement (K Cohen= 0.86). In the present study we will focus only on prosocial comments that generally are featured of arguments in support of immigrant hosting (e.g., 'We are all humans!' 'Siamo tutti umani!').

2.3. Machine Learning Approach

Sentiment analysis can be defined as the computational analysis of sentiments and opinions of an entity in a text (Liu, 2010).

The task of identifying prosocial and antisocial comments is a typical task of sentiment analysis. Since this is a particular task of binary classification we could identify two classes: *antisocial* (that we denote with Y=+1) and *prosocial* (Y=-1).

In this work, the aim of the machine learning phase is to provide the best features for each class (prosocial and antisocial) that could be used for a psycho-lexical analysis. User generated contents (UGC) in social network are highly unstructured. Each post should be pre-processed before applying machine learning algorithms. During this phase, we have applied classical pre-processing techniques (as stop-words and rare words removal and filtering words with minimum frequency of 5 in the entire corpus) through standard Python libraries, as NLTK, Pandas and scikit-learn to obtain the words that can better represent the semantic features of the corpus. Hereafter we will identify this set as textual features. In addition, we used the writeprints model, described in [Zheng et al., 2006] to extract lexical and syntactic features. In literature these features were used mainly for author identification, but could be helpful also for sentiment analysis in social media. We have counted the frequency of different part of speech (such as nouns, adjectives, verbs, etc..) in each post, using a part of speech (POS) tagger (Treetagger wrapper that supports Italian language). We defined a set of features F ={f1, f2, f3, f4} where f1, f2, f3, f4 were used to denote lexical, syntactic, part of speech and textual features. Lexical features (F1) in a post could be divided in two groups: character-based features, such as total number of characters, total number of upper-case characters and frequency of each letter, and word-based features, such as total number of words, average word length, total number of characters in a word. Syntactic Features(F2) could be defined as the frequency of punctuations and function words (as defined by [Zheng et al., 2006]. Pos-Tagger features (F3), are defined as the frequencies of different part of speech, such as noun, adjectives, verbs inside a post. Textual features (F4) are the words obtained through text pre-processing, as described above. We combined these set of features iteratively, in order to improve the accuracy of the model. Many classification techniques were compared to accomplish the sentiment analysis task. In this preliminary work

we decided to compare some popular classifiers: AdaBoost, LinearSVM, Random Forest, Extra-Trees classifier, Multinomial Naive Bayes, SGD Classifier (Linear classifier with Stochastic gradient descent). We have tuned each algorithms' parameters, with the aim of obtaining the best performances from each classifier. In this study, we choose as data mining tool scikit-learn. The entire dataset (about 12K posts) was splitted into training set (80% of the examples) and testing set (the remaining 20%). We evaluated the results, obtained from different combination of the features sets in F, as shown in Tab1. We have trained the model with different combinations of the features and different classifiers. We have measured the output in terms of accuracy obtained on test data. The table shows that the use of textual features significantly improves model's accuracy. The combination of 1, f2, f3 with f4, as showed in tab1, improves the accuracy of the model.
The best model was obtained with LinearSVM classifier (C=0,41). We have selected the textual features that can better represent each class to be used for the psycho-lexical analysis.

	LinearSVM	ExtraTrees	RF	ADABoost	NB	SGD
f1	65,58%	**69,21%**	68,92%	68,59%	65,61%	43,22%
f1+f2	72,27%	**73,76%**	70,08%	70,70%	70,66%	59,58%
f1+f2+f3	62,80%	**74,71%**	71,11%	73,09%	58,01%	57,97%
f1+f2+f3+f4	**83,05%**	82,97%	81,94%	74,17%	82,31%	82,27%
f4	**82,25%**	82,15%	80,12%	71,25%	82,10%	80,11%

Table 1. Accuracy for different feature sets and different techniques

Linear SVM creates a hyperplane that uses support vectors to maximising the distance between the prosocial and antisocial classes. In this phase, we focused only on textual features. Each textual feature in the model has a coefficient (positive for proscial or negative for antisocial) that indicates its strength for the predicted class. The value of the coefficients can then be used to determine the weight (w) of the words for the sentiment analysis task; the weight of the single word was done by the comparison of the prosocial and antisocial used words, then we will report negative values for prosocial and positive for antisocial. In the present study, we will focus only on the prosocial side.

2.4. Results.

We report the more frequent words within the prosocial comments grouped considering their usage within the entire sentences.

words	w	words	w
mafia	-2,14	rapes	-0,22
capone	-1,68	safety	-0,22
mafiosi	-1,07	stink	-0,21
criminals	-0,54	theft	-0,21
crime	-0,27		

Table 2. Disconfermation of stereotypes

The first words that features the prosocial comments are 'mafia', 'mafiosi' and 'al-capone'. This fact was very unexpected, and thus while we extract complete sentences we meet comments where people try to disconfirm the negative immigrant stereotypes by means of the words that defined Italian immigrants in the first year of the 1900. In several comments people use words like 'rape 'theft', 'stink', 'criminal' that can be included in a very frequent comment reported below where people remember an historical report from 1912 written by the US Congress Immigration Inspectorate on Italian immigrants:

"They do not like water, many of them ***stink*** *because they hold the same dress for many weeks. (...) Many children are used to* ***beg*** *(..) They say they are committed to* ***theft*** *and, if hindered,* ***violent****. the voice of some* ***rape*** *consumed after* ***ambushes*** *in suburban streets has spread when women come back from work. (..) enter our country tolive by* ***gimmicks*** *or even* ***criminal*** *activities. (..) Our safety must be the first concern*

'For everyone ... remember that Italians in the US have mainly exported organized ***crime*** *... the* ***mafia****! Not for nothing the biggest* ***criminals*** *of the years 20,30 and 40 were Italians ... Giovanni Dioguardi, Johnny Torrio, Lucky Luciano, Al Capone*!'

A second possible moves that emerged from the prosocial part is the 'cognitive and metacognitive dimension' composed by a very large amount of 'constructive' words like 'memory', remember, forgotten, past, history, forget, think, believe, understand, thought, reflect; this move can be also 'evaluative' when the prosocial commenter express the invitation to remember by insulting ('ignorant', ignorance). The move in this case consist in revealing how historical dimension can help people to remember when 'we' where immigrants and thus promoting a possible humanization of the present immigrants.

words	w	words	w
memory	-1,66	past	-0,57
ignorant	-1,43	reason	-0,54
ignorance	-1,32	to think	-0,5
to remember	-1,26	ignorant	-0,45
Forgot	-1,15	thought	-0,42
you know	-1,03	reflect	-0,41
forget	-0,95	forget	-0,39
to read	-0,85	to understand	-0,39
forgets	-0,75	word	-0,37
bed	-0,75	I remember	-0,33
think	-0,74	I got it	-0,3
(memory) short	-0,74	history	-0,29
I see	-0,73	known	-0,27
answer	-0,73	know	-0,24
to believe	-0,72	I think	-0,22
I read	-0,66		

Table 3. The cognitive dimension: words of reflection and memory

"I regret to ***remember*** *that some of our "migrants" exported the* ***Mafia*** *and the Camorra to the United States of America and never adapted to the customs of the host country until after generations. Whoever is raving that the "" our "" were martyrs and those who now come from North Africa are the bad ones or sins of* ***memory*** *or coherence. "*

The prosocial comments' can also argue being based on a recalling to universal value, by enlarging the belonging categories (*Humanity, humans, brothers, men, people*) or by eliciting a normative dimension (*egoism, solidarity, duty, life, heart, peace, security, rights*), sometimes by putting shame on racist comments like in this case.

"These people do not take anything from us, the reality of those who inform themselves, those who try to look beyond their own little, miserable little garden. It is not goodness, it is ***humanity***.*"*

words	w	words	w
humanity	-1,71	to die	-0,63
selfishness	-1,2	well being	-0,63
human	-1,17	peace	-0,57
death	-1,15	reason	-0,54
compliments	-1,1	duty	-0,51
solidarity	-1,04	we should	-0,51
world	-1,03	education	-0,46
brothers	-0,99	human	-0,41
human	-0,97	truth	-0,36
people	-0,96	honest	-0,31
patience	-0,83	moral	-0,27
heart	-0,81	Pope	-0,22
men	-0,81	safety	-0,22
life	-0,68	fair	-0,2
right	-0,64	civil	-0,19

Table 4. Recall to universal values, the normative dimension

When people stand in favour of immigrants they write comments can be also merely emotional, the users can with his words try to create an empathic stance toward immigrants by emphasizing their *courage, steem, hope, patience, desperation, pietas* or they express their negative emotions that range from *sadness to anger*, fear, guilt, but also positive emotions toward Gianni Morandi, the supportive communicator (thank, admiration).

words	w	words	w
courage	-1,11	despair	-0,6
I thank	-1,08	disgust	-0,59
estimate	-1,07	desperate	-0,52
hope	-1	I admire	-0,48
sad	-0,97	guilt	-0,47
anger	-0,93	caution	-0,31
sadness	-0,91	pity	-0,25
I respect	-0,86	poor people	-0,25
patience	-0,83	fear	-0,22
hate	-0,81		

Table 5. The helper and helpee Emotional dimension

The emotional closeness and empathy can also be reached by humanizing the potential victims and thus by recalling of dangers and problems that immigrants have to face from their starting condition (*death, wars, hunger, bread, poverty, misery, expediences, causes*), or by considering the difficulties during the trip (boat, price, travel, tragedies, tragedy, they dead, sea, mediterraneo). The used verbs are the following '*run away, they flee from, they try*.

words	w	words	w
born	-1,67	poverty	-0,4
fleeing	-1,44	tragedy	-0,37
scow	-1,26	skin	-0,37
death	-1,15	family	-0,36
war	-0,97	lands	-0,34
dilapidated	-0,89	Mediterranean	-0,34
to escape	-0,88	continue	-0,34
price	-0,81	front	-0,34
people	-0,75	origin	-0,32
fleeing	-0,69	entrances	-0,32
came	-0,68	to leave	-0,3
travel	-0,68	destination	-0,29
area	-0,65	crime	-0,27
try	-0,63	to survive	-0,26
poor	-0,59	leaves	-0,26
cause	-0,53	sicily	-0,25
they go	-0,48	poor people	-0,25
non-EU	-0,46	hard	-0,25
tragedies	-0,46	misery	-0,25
hunger	-0,45	gimmicks	-0,25
escape	-0,44	poor	-0,24
bread	-0,44	sea	-0,24
Libya	-0,44	slaves	-0,23
reasons	-0,42	they die	-0,19

Table 6. Processing causes, dangers and potential problems of immigrants

"Before we talk about those who come here to steal our jobs, we try to understand why they **flee** their country, because they prefer to go to **sea** and maybe **die** just to try and survive."

3. DISCUSSION

This qualitative study showed possible online prosocial orientation in discussion on hosting immigrants. When people stand for immigrant hosting can argue their position by rely on 1) past and then remembering past injustices, on 2) present by focusing on the immigrants' state of need (their difficulties, dangers) or 3) by invoking the moral law as the kantian starry sky above. In particular, **according to theory of moral agency (Bandura, 2015)** the prosocial comments are characterized by some linguistic markers that suggest the activation of cognitive functions implied in moral regulation. People indeed make a self-reflection on actual behaviors of their in-group member, recall the historical errors that have damaged Italians reputations, and operate a falsification of the prejudices on immigrants to avoid the repetition of the same mistakes. Therefore, words refer to the rational reconstruction process of memories at support of hosting. Concurrently the words attesting the human ability to understand others and their difficulties occur: immigrants are understood in their serious hardship and suffering. The words as "struggle, run away, war" echoing their critical and dangerous situations. Thirdly people mention the moral principles that should guide action and moral decision. In particular in these prosocial comments they recall principles of universalism and benevolence (Schwartz, 1992), according to which all human beings have equal dignity and should to be helped and supported in a state of need. All these linguistic manoeuvrings suggest the presence of executive cognitive functions implied in moral regulation process.

Parallel to these words that highlight a certain type of moral functioning, there are also words that attested the presence of emotion linked to the morality as suggested by literature (Haidt and Joseph, 2007). In particular, it is possible to find, on one hand, emotions highlighting empathic concern toward immigrants in a state of need (piety, despair, sadness), on the other hand, emotions usually associated with the violation of moral standards (guilt, anger, disgust) that probably have emotionally connoted the reactive comments in response to those against the immigrants hosting. Future works will explore more in depth the role played by the emotional states and the possible role that the prosocial comments play in the online discussion in order to understand which moves can be effective to lowering hostile tones. From a methodological point of view the mixed methodology has allowed to improve the classifier accuracy by merging the information retrieved through a machine learning algorithm, based on a classifier, with psychological models and analysis. Furthermore, this model could be used in different contexts, in order to develop a general model for user generated content in social networks.

REFERENCES

[1] Bandura, A. (2015). Moral disengagement: How people do harm and live with themselves. Macmillan Higher Education.

[2] Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996) “Mechanisms of moral disengagement in the exercise of moral agency.” Journal of personality and social psychology, Vol. 71(2), pp.364.

[3] Conte, R., & Dignum, F. (2001). From social monitoring to normative influence. Journal of Artificial Societies and Social Simulation, 4(2), 7.

[4] D’Errico F., Paciello M. (2018). Online Moral Disengagement and Hostile Emotions in Discussions on Hosting Immigrants. Internet Research, 4

[5] D'Errico, F. (2016). With different words: The arguments that can empower an e-minority. Computers in Human Behavior, 61, 205-212.

[6] D’Errico F and Poggi I., Corriero R. (2014) “Aggressive language and insults in digital political participation.” In Proceedings of MCCSIS, pp.105-114.

[7] Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). “Mapping the moral domain”. Journal of Personality and Social Psychology, 101(2), 366.

[8] Haidt, J., Graham, J. (2007) “When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize.” Social Justice Research, 20, 98-116.

[9] Hogg, M.A.; Turner, J.C.; Davidson, B. (1990). "Polarized norms and social frames of reference: A test of the self-categorization theory of group polarization". Basic and Applied Social Psychology. 11: 77–100.

[10] Hutchens, M. J., Cicchirillo, V. J., & Hmielowski, J. D. (2015). “How could you think that?!?!: Understanding intentions to engage in political flaming”. New media & society, 17(8), 1201-1219.

[11] Liu, B. (2010) Sentiment analysis and subjectivity. Handbook of Natural Language Processing. Taylor and Francis Group.

[12] Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. Advances in experimental social psychology, 25, 1-65.

[13] Sia, C. L; Tan, B; Wei, K. K. (2002). "Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity". Information Systems Research. 13: 70–90.

[14] Zheng, Rong, Li, Jiexun, Chen, Hsinchun and Huang, Zan. "A framework for authorship identification of online messages: Writing-style features and classification techniques." Journal of the American Society for Information Science and Technology 57 , no. 3 (2006): 378-393.