TSentiment: On Gamifying Twitter Sentiment Analysis

Marco Furini
Department of Communication and Economics
University of Modena and Reggio Emilia
Reggio Emilia, Italy
Email: marco.furini@unimore.it

Manuela Montangero
Department of Physics, Computer Science and Mathematics
University of Modena and Reggio Emilia
Modena, Italy
Email: manuela.montangero@unimore.it

***Abstract*—Social media platforms contain interesting information that can be used to directly measure people' feelings and, thanks to the use of communication technologies, also to geographically locate these feelings. Unfortunately, the understanding is not as easy as one may think. Indeed, the large volume of data makes the manual approach impractical and the diversity of language combined with the brevity of the texts makes the automatic approach quite complicated. In this paper, we consider the gamification approach to sentimentally classify tweets and we propose TSentiment, a game with a purpose that uses human beings to classify the polarity of tweets (e.g., positive, negative, neutral) and their sentiment (e.g., joy, surprise, sadness, etc.). We created a dataset of more than 65,000 tweets, we developed a Web-based game and we asked students to play the game. Obtained results showed that the game approach was well accepted and thus it can be useful in scenarios where the identification of people' feelings may bring benefits to decision making processes.**

***Keywords*-Gamification, GWAP, Sentiment Analysis, Sentiment Classification, Twitter Analysis.**

I. Introduction

Understanding the sentiments of people is becoming an important aspect in many different decision making process because it may be helpful in identifying others' problems and strategies strengths. Indeed, these information can be used to make more informed decisions that will likely end in better use of resources, better organization, better products/services, better citizen lifestyle, better human relations and, eventually, better society. For instance, enterprise managers may wish to track the impact that products, services and events have on people, whereas city administrators could improve the services offered to citizens and could address challenges of development and sustainability more efficiently based on what people feel [1].

While few years ago, people opinions and feelings were analyzed by opinion pool agencies through interviews, questionnaires and forms [2], [3], nowadays the focus is on social media platforms [4]. Indeed, in the last few years, we have witnessed an exponential growth of data publicly available in social media (e.g., blogs, forum, tweets) because, within social media platforms, people talk about everything: from private to public matters, from general to niche topics. Therefore, social media represent an interesting source of information to directly measure people' feelings. Furthermore, usually people access to these media from mobile environments and this creates additional contents in the form of hidden metadata. For instance, many applications attach, to the explicit user's contents, additional information like OS language, device type, capture time and geographical location [5], [6].

In the recent literature, many researchers focused on social media platforms to understand the citizens' feelings and the locations of these feelings. In this paper, we propose an alternative approach: motivated by the success of social games and applications [7], [8], we propose TSentiment, a game designed to involve people in the process of sentiment analysis. TSentiment falls in the area of Game With A Purpose (GWAP) [9], an area where games are used to exploit the computational power of humans to perform tasks that are easy for humans, but difficult for computers.

TSentiment focuses on the Twitter platform and is designed to identify the emotion embedded in Italian tweets. To implement this game, we first collected 65,514 tweets geolocated in the area of three different Italian cities (Modena, Bologna and Milano). Then, we involved students from our University (the University of Modena and Reggio Emilia): we sent the invitation to ca. 870 students, 470 of them played the game. We observed the game for 14 days and, despite the presence of some problems (*e.g.*, 25% of the participants evaluated less than 10 tweets), the game approach was well accepted as shown by observing the daytime playout, the average number (52) of evaluated tweets and the participants' engagement. Therefore, we think that the game approach could be useful in many different scenarios as it provides human evaluation for a large set of data. For instance, it could be used to help city administrators to make better decisions about public resources, to help managers to improve products/services, or to create a reference corpus based on the Italian language that would be helpful in detecting sentiments from social media platforms.

The paper is organized as follows. In Section II we present studies and proposals in the area of sentiment analysis applied to Twitter messages; in Section III we present details of our proposal; in Section IV we show the developed game. Possibile fields of applications are discussed in Section V, whereas conclusions are drawn in Section VI.

II. Background and Related Work

Considerable attention has been recently given to the exploit of large sets composed by user-generated data coming from social media in order to study social behaviour and happiness. Some approaches (e.g., [10]) use linguistic analysis techniques to automatically identify citizens' feelings, some are based on manual approach (e.g., [11]), and some others are based on a semi-automatic approach (e.g., [12]). All these approaches have advantages and drawbacks. The manual approach may provide accurate results, but it is impracticable with large data volumes; on the other side, automatic analysis is difficult to implement due to the ambiguity of natural language and to the characteristics of the posted content (*e.g.*, tweets, largely used, are usually coupled with hashtags, emoticons and links, thus creating difficulties in determining the expressed sentiment [13]). Moreover, automatic techniques need to be trained, thus they require large datasets of annotated posts or lexical databases where affective words are associated with sentiment values. While these resources are available for the English language [12], [13], [14], their availability is very limited for other languages [11], [15].

In [13] authors wish to better understand how sentiment is conveyed in tweets and SMS messages. The millions of gathered tweets were filtered by means of SentiWordNet [16], while SMS were taken from the NUS SMS Corpus[1]. Dataset entries were then manually annotated on Amazon's Mechanical Turk service[2] upon monetary reward. In [17] authors examine the relationship between social behavior and neighborhood' sentiments of people living in Pittsburgh. The study uses a set of 231.302 geotagged tweets and sentiment analysis is performed automatically by means of the lexicon SentiSense [12] and using specifically designed metrics. In [10] authors study the correlation between social level of happiness and geographic location, both at state and urban level, across U.S.A. To measure happiness they use word frequency distributions, collected from a large corpus (10 millions) of geolocated tweets, with roughly 10,000 individual words scored for their happiness independently by users of Amazons Mechanical Turk service.

The majority of results refer to dataset containing text messages written in English, while very little has been done for different languages, Italian included. It happens that private companies develop their tools for commercial use, but they do not share or freely distribute their resources. For what concerns the Italian language, in [18] authors present a project to develop a corpora for sentiment analysis with special attention to irony detection in on-line political discussion through Twitter. In [11] an Italian dataset of 1500 tweets, intended to be used for Twitter sentiment analysis in both training and testing, has been manually annotated. The dataset has been used to validate the approach implemented in Felicittà[3], a Web platform designed to estimate the happiness in Italian cities by means of sentiment analysis over geotagged tweets.

Here, in this paper we exploit the GWAP strategy in order to involve human beings into the process of understanding the citizens' sentiments that arise in the area where they live. This crowdsourcing approach is a new phenomenon that emerged in the last decade: a proposer asks users of the Web to help accomplish a specific task, sometimes upon an explicit reward. This strategy is used by many organizations, companies and institutions for problem solving and decision making, anytime there is a large task that can be better accomplished by humans than computers, because it involves creativity, reasoning, or emotions (*e.g.*, [19], [20], [21]). There is a lot of in-going research (*e.g.* [22], [23], [24]) to understand what motivates Web users to participate to crowdsourcing activities, even without a tangible reward. As a matter of fact they do, nevertheless one way to involve people is GWAP, *i.e.*, invite them to play a game: people participate because they wish to be entertained and do not necessarily need to be interested in the global task to be accomplish [9]. Probably, the most famous GWAP game is ESP [9], a game designed to label pictures: as known, it is very difficult for computers to understand the content of an image, but this task is quite easy for humans. Therefore, ESP gamified this process by displaying the same picture to two players (unknown to each other) and by asking them to label the picture. If players submitted the same label for the picture, the label was considered appropriate.

III. Proposal

The motivation behind our proposal is to investigate whether the demanding task of using the huge amount of tweets to understand the sentiments of citizens living in a certain geographic area can be accomplished by involving people into playing a game. Hence, we designed TSentiment, an on-line game that asks players to classify the sentiment of tweets. We first collected tweets located in the area we decided to investigate and then asked people to map these tweets into emotions upon a reward that is "just" becoming (and staying) the top scoring player (called *Sentiment detector*). In the following, we present the emotional space we use, the rules of the game and how the game engine works.

A. The Emotional Space

TSentiment classifies tweets into emotions, making a difference between polarity and sentiment. The former gives just an insight on the general feeling that a tweet gives: positive, negative or neutral; the latter is a refinement of

[1]http://wing.comp.nus.edu.sg/SMSCorpus/

[2]http://www.mturk.com/mturk/

[3]http://www.felicitta.net

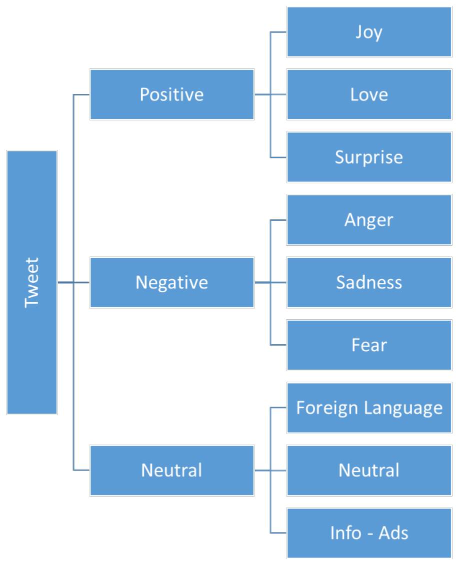

Figure 1. The adopted set of sentiments.

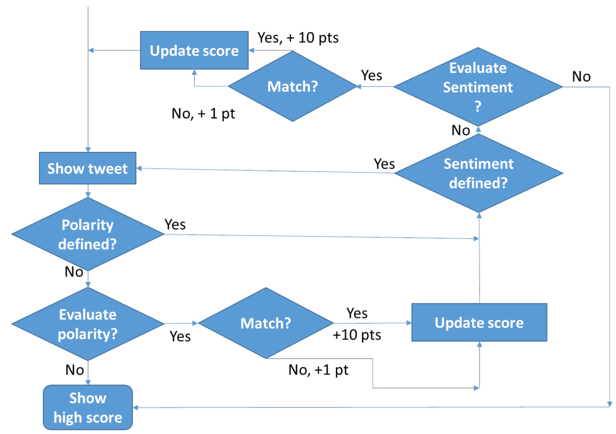

Figure 2. TSentiment Game Flow.

polarity, the player is asked to name his/her feeling by choosing among a list of given sentiments. Figure 1 shows the set of sentiments we adopted. It is driven from the emotional category model proposed in [25], where six primary emotions are defined: *anger, fear, sadness* (that we used to explicit negative polarity), and *joy, love, surprise* (used to explicit positive polarity). Other emotional category models exist with a larger number of sentiment categories (*e.g.*, [26], [27]), but we decided to limit the choice to the six emotions model in [25] to keep the game simple and fast to play. Tweets are short text messages, usually not containing difficult and articulate concepts, hence it could be difficult, for players, to classify tweets using a large set of emotions.

In addition to positive and negative polarities, we introduced a neutral polarity to classify all tweets that do not express a sentiment (*e.g.*, containing advertisements), that are meaningless (*e.g.*, containing only punctuation) or in a language unknown to the player. Indeed, we noticed that a large number of tweet falls into these cases, thus there is a real need to introduce this neutral polarity.

B. Rules of the Game

Object of the game: become the "Sentiment detector" by scoring more points than the other players.
Game setup: after entering the game, the game engine shows a tweet and asks the player to evaluate its polarity first, and then its sentiment (see, Figure 1).

If a player leaves the game, the game engine shows the top ten high-scores and asks the player for his/her name.

C. Game Engine

The game engine uses three different sets of data: $T = \{t_1, t_2, t_3, ..., t_n\}$ containing the tweets, $P = \{p_1, p_2, p_3, ..., p_n\}$ containing the tweets whose polarity has been judged by players, and $S = \{s_1, s_2, s_3, ..., s_n\}$ containing the tweets whose sentiment has been judged by players. In particular, each t_i is a triplet containing:

- *ID:* the tweet ID;
- *text*: the content of the tweet;
- *state*: defined or undefined.

At the beginning, T contains all tweets with "undefined" state. When two players agree on the same tweet classification, the state is turned to "defined".

Each p_i is a triplet containing:

- *ID*: the tweet ID;
- *state*: defined or undefined;
- *polarity*: a set of pairs $(player, polarity)$, where $polarity \in \{positive, negative, neutral\}$ is the call of $player$ on the tweet.

Each s_i is a triplet containing:

- *ID*: the tweet ID;
- *state*: defined or undefined;
- *sentiment*: a set of pairs $(player, sentiment)$, where $sentiment \in \{joy, love, surprise, anger, sadness, fear, foreign\,language, netural, Info/Ads\}$ is the call of $player$ on the tweet.

At the beginning, P and S are empty. New entries are initially set with "undefined" state and when two players agree on the same polarity or sentiment the state is turned to "defined".

When a player starts the game, the game server randomly selects t_i among the undefined ones. Then, the game proceeds in two steps, as shown in Figure 2:
Polarity evaluation: The game engine checks whether the ID of tweet t_i exists in P and if its status is set to defined. If so, the game skips the polarity investigation and moves to the sentiment evaluation. Otherwise, the player is asked to call the tweet polarity and, if not existing, new entries are created in P and S with the tweet ID. Then, the game engine

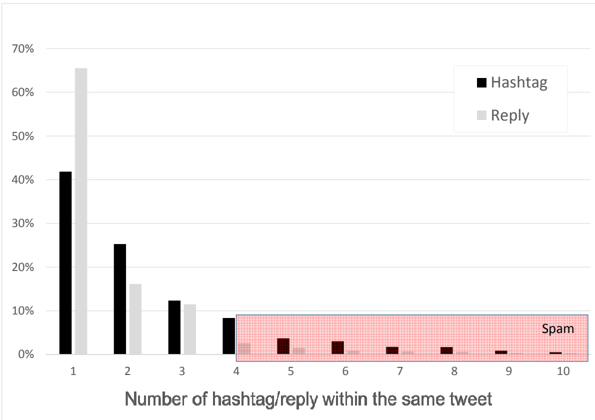

Figure 3. Percentage of tweets containing a different number of hashtags and of tweet addresses. Tweets with more than 4 hashtags or more than 3 tweet addresses are considered SPAM and excluded from the dataset.

checks whether the call matches the polarity of a previous call (if it exists) by checking the polarity filed in P. If so, players that provided the same call both gain 10 points, and the state of tweet t_i in P is set to defined. Otherwise, the player gains one point and the called polarity is stored in P.

Sentiment evaluation: The player is asked to call sentiment for t_i. Then, the game engine checks whether the call matches the sentiment of a previous call (if it exists) by checking the sentiment filed in S. If so, players that provided the same call both gain 10 points, and the state of tweet t_i in S and in T is set to defined. Otherwise, the player gains one point and the called sentiment is stored in S.

IV. Experimental Assessment

A. Data Collection and Filtering

We collected 65,514 tweets generated in the cities of Modena (population 184,000 ca), Bologna (population 386,000 ca) and Milano (population of 1,345,000 ca), then filtered to remove spam. Indeed, we observed that several tweets contained a considerable number of hashtags[4] (up to 15 hashtags in the same tweet) and/or a considerable number of tweet addresses (up to 13 tweet addresses in the same tweet). Looking at the different percentages (see Figure 3), we considered as "spam" all the tweets with more than 4 hashtags (3,843 tweets) and those containing more than 3 tweet addresses (2,306 tweets). As a result, the input dataset to the TSentiment game is composed of 59,446 tweets.

B. Results

As mentioned, we asked students to play the game. Indeed, we sent invitations, through the Department social media forum, to ca. 870 students. The developed Web interface is shown in Figure 4. During the first 14 days, we

[4]For instance, a tweet contained the message "#dance #house #housemusic #housedance #sport #fun #nike #health #sportive #sportswoman".

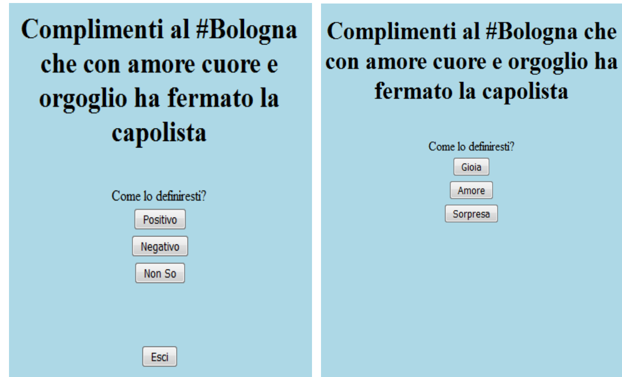

Figure 4. TSentiment interface: participants are asked to call the polarity of a tweet (left) and then to call its sentiment (right). (tweets are in Italian)

Table I
Summary of results

Polarity	Sentiment
30% Positive (2,842 tweets)	54% Joy 14% Surprise 32% Love
22% Negative (2,123 tweets)	36% Anger 30% Fear 34% Sadness
48% Neutral (4,478 tweets)	16% Foreign Language 25% Info/Ads 59% Neutral
Total classified tweets: 9,443	
Total evaluated tweets: 11,451	
Total tweets: 59,446	

had 470 students participating the game and they evaluated 11,451 different tweets. The game classified the polarity of 82% of them (i.e., 9443 tweets) as 30% positive, 22% negative and 48% neutral. Table I shows the classification of these 9,443 tweets by polarity and sentiment. It is to note that sometimes players quit after polarity classification, hence the total number of tweets classified by sentiment is slightly smaller than the number of tweets classified by polarity (9,380 vs. 9,443).

Figure 5 shows players playout daytime. It is interesting to observe that, with high probability, students started to play after checking their personal profile in the students' forum; *i.e.*, in the morning, while at home or when commuting; in the mid afternoon, when they finish their lectures; in the early evening when they arrive at home; after dinner and before going to bed. It is interesting to notice that there are peaks that roughly correspond to lectures breaks.

Figure 6 shows, day per day, the number of evaluations done during the 14 days. It can be observed that these numbers "go in waves": they tend to decrease, and then suddenly increase. This is likely due to the timing of the invitations we sent: at day #1 we sent the invitation to 250 students, at day #3 to other 70 students, at day #6 to other 70 students, at day #7 to other 200 students, at day #9 to

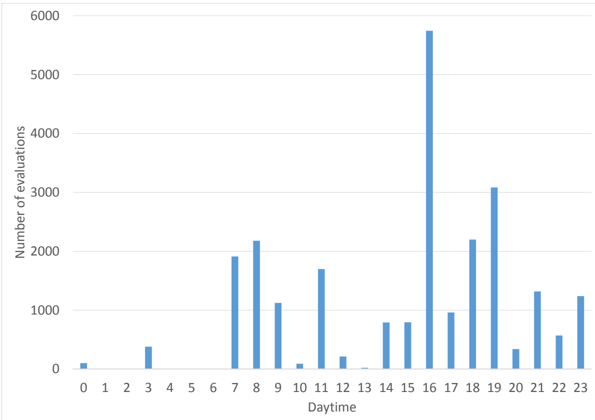

Figure 5. TSentiment playout daytime. According to the playout time, it is likely that participants played the game after receiving the invitation in their personal profile in the students' forum.

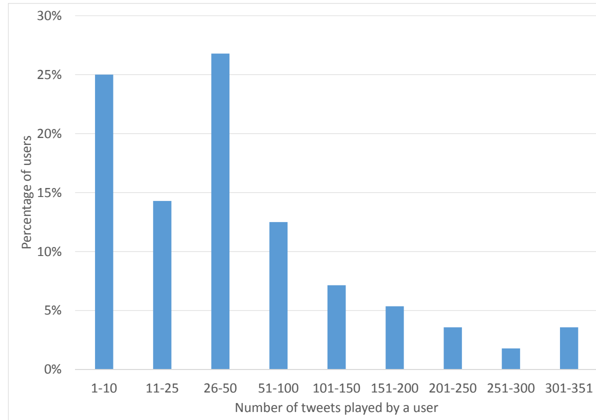

Figure 7. TSentiment analysis by users' activity: percentage of participants who played a certain number of tweets.

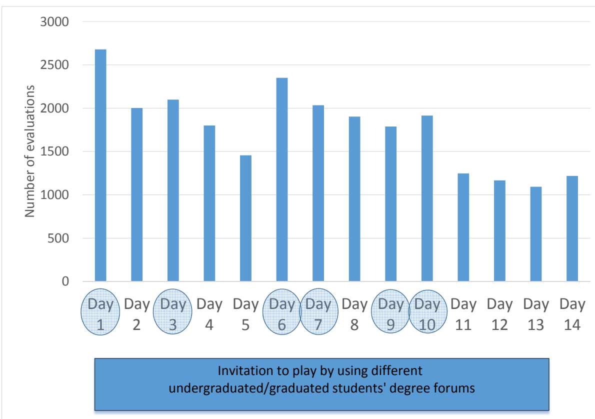

Figure 6. TSentiment analysis by days' activity: number of tweets played by participants during the 14 days. Day#5 and day #12 were Sundays.

other 200 students and at day #10 to other 80 students. In total, we sent the invitation to 870 students. Note that day#5 and day #12 were Sundays.

Figure 7 shows the percentage of users who played a certain number of tweets. It can be noted that a large number of them (i.e., 34%) played more than 51 tweets, and 27% played a number of tweets that varies between 26 and 50. Data also show a potential problem: 25% of the participants played a number of tweets not greater than 10. Likely, they did not find the game attractive but were nevertheless curios to see what it was about.

V. Fields of Application

There are many different fields of application where people feelings may be used. Here, we highlight a scenario where people feelings can be directly used and a scenario where people feelings are used as a mean to develop automatic sentiment analysis strategies.

Direct access to people feelings. Enterprise managers and city administrators might obtained benefits in understanding people feelings (e.g., better understanding of the brand/product reputation, better management of public resources). Furthermore, in these scenarios, the players' engagement may be guaranteed by placing small prizes (e.g., free parking or public transport tickets, free samples, etc.).

Reference corpus. We mentioned that the development of automatic sentiment analysis strategies requires large dataset of annotated posts and/or lexical databases. While these resources are available for the English language, for other languages they are not. TSentiment might be used to create a reference corpus for any language (in our case the Italian language) that, in turn, might be used in several frameworks for detecting sentiments from big data sources. The corpus might be used as a golden standard to test new strategies or to create a lexicon, useful to implement automatic sentiment analysis strategies for any language.

VI. Conclusions

In this paper, we proposed to use a gamification approach to sentimentally classify tweets. Indeed, we proposed TSentiment, a game with a purpose that uses human beings to classify the polarity (e.g., positive, negative, neutral) and the sentiment (e.g., joy, surprise, sadness, etc.) of tweets. We created a dataset of more than 65,000 tweets, we developed a Web-based game and we asked undergraduate/graduated students to play the game. The obtained results showed that players liked the game approach, but also showed two limitations that should be removed to improve the gaming experience: the filtering process should be improved to avoid meaningless tweets (e.g., we noticed a tweet composed of simple question marks "????") to be in the dataset and the players' engagement should go beyond the simple high-score (for example, by introducing notifications when the player's ranking changes). We plan to deepen these aspects in the near future. Given its simplicity, people may play TSentiment now and then (e.g., when they are idle like when waiting for public transport or while commuting or when they want to have a short break at work) and therefore it can be useful in scenarios where the identification of people' feelings may bring benefits to decision making processes.

VII. Acknowledgements

Authors wish to thank all students who played the game.

References

[1] B. A. Amaba, "Industrial and business systems for smart cities," in *Proc. of EMASC 2014*, pp. 21–22.

[2] J. W. Ainsworth, "Why does it take a village? the mediation of neighborhood effects on educational achievement," *Social Forces*, vol. 81, no. 1, pp. 117–152, 2002.

[3] G. Galster, "The mechanism(s) of neighbourhood effects: Theory, evidence and policy implications," in *Neighbourhood effects research: New perspectives*. 2012, pp. 23–56.

[4] M. Montangero and M. Furini, "Trank: Ranking twitter users according to specific topics," in *Proceedings of the 12th Annual IEEE Consumer Communications and Networking Conference (CCNC 2015)*, Jan 2015, pp. 767–772.

[5] M. Furini and V. Tamanini, "Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions," *Multimedia Tools and Applications*, vol. 74, no. 21, pp. 9795–9825, 2015. [Online]. Available: http://dx.doi.org/10.1007/s11042-014-2151-7

[6] M. Furini, "Users behavior in location-aware services: Digital natives vs digital immigrants," *Advances in Human-Computer Interaction*, vol. 2014, 2014. [Online]. Available: http://dx.doi.org/10.1155/2011/678165

[7] M. Furini, "Mobile games: What to expext in the near future," in *Proceedings of GAMEON Conference on Simulation and AI in Computer Games*. EuroSis Society, November 2007.

[8] B. Kirman, S. Björk, S. Deterding, J. Paavilainen, and V. Rao, "Social game studies at CHI 2011," 2011, pp. 17–20. [Online]. Available: http://doi.acm.org/10.1145/1979742.1979590

[9] L. von Ahn and L. Dabbish, "Designing games with a purpose," *Commun. ACM*, vol. 51, no. 8, pp. 58–67, Aug. 2008.

[10] L. Mitchell, M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth, "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place," *PLoS ONE*, vol. 8, no. 5, 2013.

[11] C. Bosco, L. Allisio, V. Mussa, V. Patti, G. Ruffo, M. Sanguinetti, and E. Sulis, "Detecting happiness in italian tweets: Towards an evaluation dataset for sentiment analysis in felicittá," in *Proc. of the 5th Workshop on Emotion, Social, Signals, Sentiment & Linked Open Data*, May 2014.

[12] J. Carrillo de Albornoz, L. Plaza, and P. Gervás, "Sentisense: An easily scalable concept-based affective lexicon for sentiment analysis," in *The 8th International Conference on Language Resources and Evaluation (LREC 2012)*, 2012.

[13] P. Nakov, S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson, "Semeval-2013 task 2: Sentiment analysis in twitter," in *Proc. of the Workshop on Semantic Evaluation*. Association for Computational Linguistics, June 2013.

[14] Y. H. Hassan Saif, Miriam Fernandez and H. Alani, "Evaluation datasets for twitter sentiment analysis: A survey and a new dataset, the sts-gold," *first ESSEM workshop*, 2013.

[15] E. T. K. Sang, "Using tweets for assigning sentiments to regions," in *Proc. of the International Workshop on Emotion, Social Signal, Sentiment & Linked Open Data*, May 2014.

[16] S. Baccianella, A. Esuli, and F. Sebastiani, "Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining." in *LREC*, vol. 10, 2010, pp. 2200–2204.

[17] Y.-R. Lin, "Assessing sentiment segregation in urban communities," in *Proceedings of the 2014 International Conference on Social Computing*. 2014, pp. 9:1–9:8.

[18] C. Bosco, V. Patti, and A. Bolioli, "Developing corpora for sentiment analysis: The case of irony and senti-tut," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 55–63, Mar. 2013.

[19] M. Furini, "Vimood: using social emotions to improve video indexing," in *Proceedings of the 12th International IEEE Consumer Communications and Networking Conference (CCNC2015)*, Jan 2015.

[20] M. Furini, "On gamifying the transcription of digital video lectures," *Entertainment Computing*, vol. 14, pp. 23 – 31, 2016. Elsevier, Science Direct.

[21] C. E. Palazzi, M. Roccetti, and G. Marfia, "Realizing the unexploited potential of games on serious challenges," *Comput. Entertain.*, vol. 8, no. 4, pp. 23:1–23:4, Dec. 2010. [Online]. Available: http://doi.acm.org/10.1145/1921141.1921143

[22] N. Kaufmann, T. Schulze, and D. J. Veit, "More than fun and money. worker motivation in crowdsourcing - a study on mechanical turk," in *Proc. of the 2011 Americas Conference on Information Systems*.

[23] H. Zheng, D. Li, and W. Hou, "Task design, motivation, and participation in crowdsourcing contests," *Int. Journal of Electronic Commerce*, vol. 15, no. 4, pp. 57–88, 2011.

[24] M. Hossain, "Users' motivation to participate in online crowdsourcing platforms," in *Innovation Management and Technology Research (ICIMTR), 2012 International Conference on*. IEEE, 2012, pp. 310–315.

[25] W. G. Parrott, *Emotions in social psychology: Essential readings*. Psychology Press, 2001.

[26] M. B. Arnold, "Emotion and personality." 1960.

[27] R. Plutchik, *Emotion: A Psychoevolutionary Synthesis*. New York: Harper and Row, 1980.