User Emotion Detection via Taxonomy Management: An Innovative System

Alfredo Cuzzocrea[1], Giovanni Pilato[2], and Edoardo Fadda[3]

[1] University of Calabria, Rende, Italy
alfredo.cuzzocrea@unical.it
[2] ICAR-CNR, Palermo, Italy
giovanni.pilato@cnr.it
[3] Politecnico di Torino, Torino, Italy
edoardo.fadda@polito.it

Abstract. Catching the attention of a new acquaintance and empathize with her can improve the social skills of a robot. For this reason, we illustrate here the first step towards a system which can be used by a social robot in order to "break the ice" between a robot and a new acquaintance. After a training phase, the robot acquires a sub-symbolic coding of the main concepts being expressed in tweets about the IAB Tier-1 categories. Then this knowledge is used to catch the new acquaintance interests, which let arouse in her a joyful sentiment. The analysis process is done alongside a general small talk, and once the process is finished, the robot can propose to talk about something that catches the attention of the user, hopefully letting arise in him a mix of feelings which involve surprise and joy, triggering, therefore, an engagement between the user and the social robot.

1 Introduction

Engagement is one of the most basic and important phases in interactions between human beings. In the last years there has been a growing interest about this topic throughout the human-machine-interaction (HMI) and related fields [6]. Researchers have highlighted that engagement is a very complex phenomenon, including both cognitive and affective components: it should involve attention and enjoyment [3] [15].

We refer to this term as the "starting or intention to start an interaction". In particular, we focus our attention on the fact that, in making new acquaintances, the first impression is very important, and finding as soon as possible common interests to talk about, allows starting an empathetic interaction between two persons, with all that this implies.
In order to trigger both attention and enjoyment, given these premises, it would

be useful to design a social robotic system which tries to find the topic of interest of a new acquaintance while attempting to understand what might raise a sentiment of joy attempting to catch an empathetic attention of the user.
As a matter of fact, the knowledge of the topics of interest and the "joyful" subjects for the user can lead the first stages of a conversational interaction that allows the robot to facilitate the engagement of a friendly interaction, instead of a classical trivial interaction between a robot and an human user.
To reach this goal, the robot can access the social network data of the new acquaintance trying to coarsely profile her/his interests, catching useful information to engage a possibly interesting conversation for the user.
Social networks represent a great place, maybe the best, to gather information about people's opinions, since they are generally used to express personal thoughts and to discuss with other people about specific subjects [12][30]. These opinions are really useful to understand and classify the emotion of an event, a product, a person, etc. and analyze his trend [21][22][13].

In this paper we illustrate the design of a system which can be used by a social robot in order to "break the ice" between the robot and a new acquaintance.
First of all, the robot acquires a knowledge about the construction of prototypes describing each entry of the IAB Taxonomy.

The system needs a training phase where fundamental concepts, induced by a data driven construction of a conceptual space by using the Latent Semantic Analysis (LSA) procedure and a set of topics derived by the Latent Dirichlet Allocation (LDA) methodology, representing the Tier1 categories of the IAB v2.0 taxonomy are mapped in a semantic space.
A set of tweets is therefore retrieved for each word describing each entry of the IAB Taxonomy. A set of words describing the conceptual axes of two "conceptual spaces" induced from each set of tweets associated to a single IAB entry is built. Each conceptual axis is therefore described by a specific "bag of words" which constitute the axis description. Each axis is therefore coded as a vector in a semantic space built through LSA, and it is associated to the specific IAB entry. At the end of the procedure, each entry of the IAB taxonomy is associated to set of vectors, associated to the labels of each fundamental axis of the category, in the built semantic space.

On the other hand, a system which is able to detect a pattern of basic Eckman emotions [24], given a text, is trained too.
Once the system is trained, during a general conversation with a new acquaintance, the robot asks for the user Twitter ID, and while the conversation continues, it retrieves the most recent tweets of the user.
Each tweet is then encoded as a vector in a semantic space. The semantic similarity between each tweet and each vector representing each entry of the IAB taxonomy is computed, and the highest value of similarity is retained.
The above procedure allows to associate a tweet of the user to a pattern of IAB categories; furthermore, for each tweet a vector of Eckman fundamental emotions is computed. This leads to a selection of the Tier1 categories of the IAB taxonomy which are of interest of the user and that let arise in the user a specific

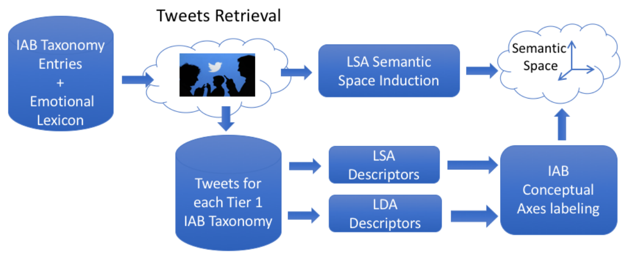

Fig. 1. *Training process*

emotion. In our case we have chosen to select the "joy" emotion, which is the most desirable when a person meets for the first time another human being.

The goal is to engage a conversation somehow polarizing it on topics that catch the attention of the user, trying to establish an empathetic relationship. Under some extensions, this approach has relationships with *adaptive metaphors*, like those developed in other scientific contexts (e.g., [5]).

2 The System

The proposed system is composed of a set of modules interacting in order to catch the attention of the user. The modularity of the proposed architecture makes it suitable to be implemented on top of Cloud infrastructure (e.g., [10]). The system has a training phase, shown in Fig.1, where a semantic space S is induced from Twitter data and a joyful-topic-detection process, illustrated in Fig. 2, which exploits the Twitter ID of the user in order to retrieve her posts and trying to catch the interests of the user that somehow let arise a "joy" emotion.

2.1 The IAB Taxonomy

The IAB (Interactive Advertising Bureau) Tech Lab Content Taxonomy is a concise taxonomy which is also an international standard to map contextual business categories [17] [18]. The latest release of the taxonomy, namely version 2.0, has been released on November 2017 and it accounts 698 entries distributed over 29 Tier-1 classes.
This taxonomy is particularly suited for being used by companies in the market, it is standardized and industry-neutral. These characteristics can be effectively exploited for profiling an user interests.

2.2 Tweets retrieval module

The dataset object of analysis is retrieved by using the Twitter APIs with the default access level. The default access level gives a random sample of the streaming

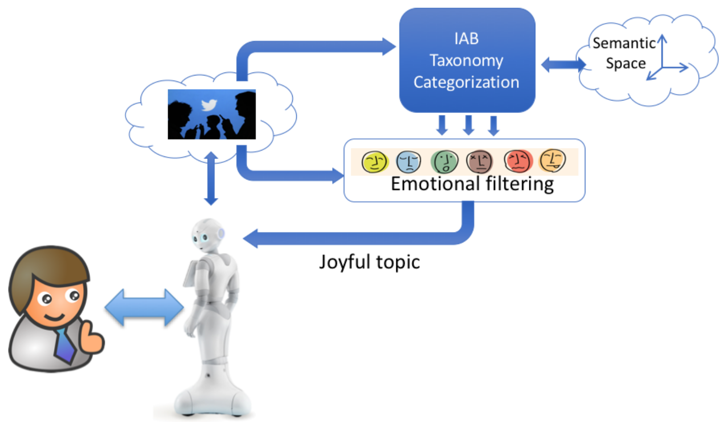

Fig. 2. *Joyful-topic-detection process*

of publicly available tweets. For our approach, we use only the tweet text content, which is preprocessed before being exploited to build a data-driven conceptual space. Stop-words are filtered out, and links are removed before processing the text since they often hide off-topic posts or even spam. Abnormal sequences of characters were discarded. The retrieval module can be used either for retrieving tweets satisfying a query composed of keywords or to download the last tweets of a given twitter user ID.

2.3 LSA-based Descriptors

The Latent Semantic Analysis (LSA) technique is a well-known methodology that is capable of giving a coarse sub-symbolic encoding of word semantics [20] and of simulating several human cognitive phenomena [19]. The LSA procedure is based on a term-document occurrence matrix $\mathbf{A}$, whose generic element represents the number of times a term is present in a document. Let K be the rank of $\mathbf{A}$. The factorization named Singular Value Decomposition (SVD) holds for the matrix $\mathbf{A}$:

$$\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \tag{1}$$

Let R be an integer > 0 with $R < N$, and let $\mathbf{U}_R$ be the $M \times R$ matrix obtained from $\mathbf{U}$ by suppressing the last $N - R$ columns, let $\mathbf{\Sigma}_R$ be the matrix obtained from $\mathbf{\Sigma}$ by suppressing the last $N - R$ rows and the last $N - R$ columns; let $\mathbf{V}_R$ be the $N \times R$ matrix obtained from $\mathbf{V}$ by suppressing the last $N - R$ columns. Then:

$$\mathbf{A}_R = \mathbf{U}_R\Sigma_R\mathbf{V}_R^T \tag{2}$$

$\mathbf{A}_R$ is a $M \times N$ matrix of rank R, and it is the best rank R approximation of $\mathbf{A}$ (among the $M \times N$ matrices) with respect to the Frobenius metric. The i-th row of the matrix $\mathbf{U}_R$ may be considered as representative of the i-th word.

The columns of the $\mathbf{U}_R$ matrix represent the R independent dimensions of the $\Re^{\Re}$ space S. Each j-th dimension is weighted by the corresponding value σ_j of $\mathbf{\Sigma}_R$. Furthermore, each j-th dimension can be tagged by considering the words having the highest module values of u_{ij}. This makes it possible to interpret the space S as a "conceptual" space, according to the procedure illustrated in [1][25].

2.4 LDA-based Descriptors

In the last years a Bayesian probabilistic model of text corpora, namely the Latent Dirichlet Allocation (LDA), has been proposed with the aim of finding topics in documents [2] by associating a set of words to each topic, obtaining a rough representation of a textual corpus.

One on the main advantages of LDA, like LSA, is the fact that the approach is completely unsupervisedThe only thing required by LDA to setup a priori is the number N of topics to extract. Latent topics are discovered through the identification of sets of words in the corpus that often occur together within documents. LDA is based on a generative process according to these two steps:

- For each topic $n = 1, 2, \cdots, N$, $\phi^{(n)} \sim Dirichlet(\beta)$ is a discrete probability distribution over a fixed Vocabulary constituting the $n{-}th$ topic distribution, and β is a hyperparameter for the symmetric Dirichlet distribution.
- For each document d_k of the document corpus, $\theta_{d_k} \sim Dirichlet(\alpha)$, which is a symmetric Dirichlet distribution for the specific document d_k over the available topics is computed. θ_{d_k} is a low dimensional coding of d_k in the topic space. For each word w_i belonging to the d_k document, $z_i \sim Discrete(\theta_{d_j})$ and $w_i \sim Discrete(\phi^{(z_i)})$ are being computed, where z_i is the topic index for w_i

The above process leads to the following distribution

$$p(\mathbf{w}, \mathbf{z}, \theta, \phi|\alpha, \beta) = p(\phi|\beta)p(\theta|\alpha)p(\mathbf{z}|\theta)p(\mathbf{w}|\phi_z) \tag{3}$$

where $\mathbf{z}, \theta, \phi$ are the latent variables of interest In LDA the posterior inference is given by:

$$p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\theta, \phi, \mathbf{z}|\mathbf{w}, \alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)} \tag{4}$$

which represents the learning of the latent variables given the observed data. The above formula is usually computed through variational inference and Gibbs sampling, as reported in literature [14][2][29]

2.5 Emotion Detection Module

This module deals with the detection of emotions in tweets. For the emotional labeling of tweets, we have considered the six Ekman basic emotions: *anger*, *disgust*, *fear*, *joy*, *sadness* and *surprise*, exploiting an emotions lexicon obtained from the Word-Net Affect Lexicon, as described in [26] [27] and adopting a

procedure that has been illustrated in [24], which we briefly recap below.
The methodology is based on LSA and starts from the fact that any text d can be mapped into a Data Driven "conceptual" space in the sense illustrated above, by computing a vector $\mathbf{d}$ whose i-th component is the number of times the i-th word of the vocabulary, corresponding to the i-th row of $\mathbf{U}_R$, appears in d. This leads to the mapping of the text as:

$$\mathbf{d}_R = \mathbf{d}^T \mathbf{U}_R \Sigma_R^{-1} \tag{5}$$

The emotional lexicon has been split into six lists, each one associated to one of the basic Ekman emotions {*anger, disgust, fear, joy, sadness, surprise*}. Fixed an emotion e, a set of 300 artificial sentences has been built by using five randomly selected words belonging to the list related to e. This procedure has been done for each list associated with a fundamental Ekman emotion, leading to a set of 1800 artificial sentences. Furthermore, all the 1542 words of the lexicon have been considered. Each one of the 3342 (i.e. 1542+1800) b texts associated with an emotion e has been mapped into the data driven "conceptual space" induced by TSVD according to the transformation in eq.(2). The above procedure leads to a cloud of 3342 (i.e. 1542+1800) vectors that have been used to map a tweet from the conceptual space to the emotional space. In particular, we have six sets $E_{anger}, E_{disgust}, \cdots, E_{surprise}$ of vectors constituting the sub-symbolic coding of the words belonging to the lexicon for a particular emotion together with their artifact sentences. The generic vector belonging to one of the sets will be denoted in the following as $\mathbf{b}_i^{(e)}$ where $e \in \{\text{"}anger\text{"}, \text{"}disgust\text{"}, \text{"}fear\text{"}, \text{"}joy\text{"}, \text{"}sadness\text{"}, \text{"}surprise\text{"}\}$ and i is the index that identifies the i-th $\mathbf{b}_i^{(e)}$ in the e set. Specifically, $\mathbf{b}_i^{(e)}$ is computed as:

$$\mathbf{b}_i^{(e)} = \mathbf{b}^T \mathbf{U}_R \Sigma_R^{-1} \tag{6}$$

where $\mathbf{b}$ is, time by time, the vector computed starting from one of the 3342 textual artifacts b according to the procedure illustrated at the beginning of this section.
Analogously, any textual content t of a tweet can be mapped into the Data Driven "conceptual" space by computing a vector $\mathbf{t}$ whose i-th component in the number of times the i-th word of the vocabulary, corresponding to the i-th row of $\mathbf{U}_R$, appears in t. This leads to the mapping of the tweet as:

$$\mathbf{t}_R = \mathbf{t}^T \mathbf{U}_R \Sigma_R^{-1} \tag{7}$$

Once the tweet t is mapped into the "conceptual" space as a vector $\mathbf{t}_R$, it is possible to compute its emotional fingerprint by exploiting the vectors $\mathbf{b}_i^{(e)}$, which act as "beacons" for the vector $\mathbf{t}_R$, helping in finding its position inside the conceptual space.
In particular, fixed $\mathbf{t}_R$, for each set E_e it is computed the weight:

$$w_e = \max cos(\mathbf{t}_R, \mathbf{b}_i^{(e)}) \tag{8}$$

once all the six w_e weights are computed, the vector $\mathbf{f}_t$, associated to the vector $\mathbf{t}_R$, and by consequence to the tweet t, is calculated as:

$$\mathbf{f}_t = \left[\frac{w_{(anger)}}{\sqrt{\sum_e w_e^2}}, \frac{w_{(fear)}}{\sqrt{\sum_e w_e^2}}, \dots, \frac{w_{(surprise)}}{\sqrt{\sum_e w_e^2}} \right] \tag{9}$$

The vector $\mathbf{f}_t$ finally constitutes the *emotional fingerprint* of the tweet t in the emotional space. The emotional space is therefore a six-dimensional hypersphere where all tweets can be mapped and grouped. We call the fingerprint $\mathbf{f}_t$ "emoxel", analogously as the *knoxel* in the conceptual space paradigm [4].

2.6 Conversational Engine

The conversational engine exploits a speech-recognition module which makes use of the Google speech recognition APIs; after that the speech-to-text task is performed, the recognized string is sent to a dialogue manager. A set of question-answer rules are set-up into the conversation engine in order to start a conversation that leads to the detection of the user interests by transparently invoking the most adequate procedures which analyze the social network posts of the new acquaintance.

The conversational agent engine allows for a natural human-robot interaction. The conversational module is based on a Rivescript engine, which is a a simple scripting language for realizing chatbots and other conversational entities.

We have chosen this kind of engine because of the following interesting features: it is plain text, line-based scripting language, simple to learn, quick to type, and easy to read and maintain [23]. The syntax required to build a Rivescript "knowledge base" is very simple: Question-Answers pairs are encoded in plain text; it is easy to write a set of rules that can be combined to build effective conversational agents; its core library is focused on rendering responses, and it is straightforward to make custom modules and scripts; last but not least is an Open Source tool released under the MIT license [28].

The choice of such an engine allows us to easily connect it to other kind of robots or other kind of services. As a matter of fact, the conversational engine is invoked through a REST service and the answer is delivered to the user after its processing.

A Rivescript knowledge base is made up of *Triggers/Replies* pairs. *Triggers* are identified by a "+" sign, while *Replies* are denoted by a "-" sign.

For example:

```
+ hi
- Hello there, my name is SocialRobot,
  please could you tell me your twitter ID?
```

the above pair makes it possible that whenever the user says "Hi", the conversational engine replies with "Hello there, my name is SocialRobot, please could you tell me your twitter ID? ".

At the beginning of the conversation, a specific Rivescript *Topic* is activated.

Topics are logical groupings of triggers. When the conversation is bound in a topic, what the user says can only match triggers that belong to the activated topic [23].
The topic is aimed at entertain a general conversation while the robot peeks the tweets of the user trying to roughly identify the subjects that interest the user and those that specifically trigger joyful emotions. Once the predominant subject has been identified, the robot activates another Rivescript Topic which is of particular interest for the user, trying to establish an empathetic engagement with the new acquaintance.
As an example, let us say that the system finds that, among the different higher level categories of the IAB taxonomy, the user is particularly interested in the "Automobiles" topic and that some of his tweets show the "joy" emotion for that topic, the conversation will be switched to the "Automobile" Topic and specific sentences will be said by the robot in order to catch the user attention and empathy, like "Great! with my superpowers I can see that you like automobiles. I like the *brand* automobiles! Which one do you prefer?"

3 Conclusions and future works

We have presented a preliminary work on a system that tries to catch the attention of a new acquaintance with the aim of establishing a first engagement with the user.
The system uses both LSA and LDA descriptors, as well as an emotion detection module to reach this goal. A conversational engine guides the initial process and continues with*t* the conversation.
Many issues have to be enhanced, starting from a more fine grained classification which should be also fast and reliable, the selection of specific entities that can catch in a more effective manner the attention of the user, as well as the automatic generation of conversational statements starting from the user tweets. Other lines of research shall consider privacy-preservation issues (e.g., [8, 11]), as well as complex web intelligent solutions (e.g., [7, 9]).

References

1. F. Agostaro, A. Augello, G. Pilato, G. Vassallo, and S. Gaglio, "A conversational agent based on a conceptual interpretation of a data driven semantic space.," Lecture Notes in Artificial Intelligence, vol. 3673, no. 2, pp. 381–392, 2005.
2. Blei, D., Ng, A., and Jordan, M. Latent Dirichlet allocation. Journal of Machine Learning Research,3:993–1022, January 2003.
3. Brethes L, Menezes P, Lerasle F, Hayet J (2004) Face tracking and hand gesture recognition for human–robot interaction. In: IEEE international conference on robotics and automation, vol 2. IEEE, pp 1901–1906
4. Chella, A., Frixione, M., and Gaglio, S. (2008) A cognitive architecture for robot self consciousness. Artificial intelligence in medicine, 44(2): 147–154
5. Cannataro M., Cuzzocrea A., and Pugliese A. (2001) A Probabilistic Approach to Model Adaptive Hypermedia Systems. In: 1st Int. Workshop on Web Dynamics, in conjunction on ICDT 2001
6. Corrigan Lee J., Peters C., Küster D. and Castellano G. (2016) Engagement Perception and Generation for Social Robots and Virtual Agents - Toward Robotic Socially Believable Behaving Systems - Volume I, Intelligent Systems Reference Library 105, pp 29-51, Springer

7. Cuzzocrea, A. (2006) Combining multidimensional user models and knowledge representation and management techniques for making web services knowledge-aware. Web Intell. Agent Syst. 4(3): 289-312
8. Cuzzocrea A., and Bertino, E. (2011) Privacy Preserving OLAP over Distributed XML Data: A Theoretically-Sound Secure-Multiparty-Computation Approach. J. Comput. Syst. Sci. 77(6): 965-987
9. Cuzzocrea A., De Maio C., Fenza G., Loia V., and Parente M. (2016) OLAP analysis of multidimensional tweet streams for supporting advanced analytics. In: AMC SAC 2016, pp. 992-999
10. Cuzzocrea A., Fortino G., and Rana O. (2013) Managing Data and Processes in Cloud-Enabled Large-Scale Sensor Networks: State-of-the-Art and Future Research Directions. In: 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2013, pp. 583–588
11. Cuzzocrea A., and Russo V. (2009) Privacy Preserving OLAP and OLAP Security. Encyclopedia of Data Warehousing and Mining 2009, pp. 1575-1581
12. D'Avanzo E., Pilato G. (2014) Mining social network users opinions' to aid buyers' shopping decisions. Computers in Human Behavior, Elsevier, Vol.51, pp 1284-1294
13. D'Avanzo E., Pilato G., and Lytras M. D.. Using twitter sentiment and emotions analysis of google trends for decisions making. In Program, pages Vol. 51, Issue 3. 2017.
14. Darling WM. A theoretical and practical implementation tutorial on topic modeling and gibbs sampling. Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies 2011 Dec 1 (pp. 642-647).
15. Delaherche E, Dumas G, Nadel J, Chetouani M (2014) Automatic measure of imitation during social interaction: a behavioral and hyperscanning-eeg benchmark. Pattern Recognit Lett.
16. Ekman P., Friesen W V. Constants across cultures in the face and emotion., Journal of personality and social psychology 17:124, 1971.
17. Interactive Advertising Bureau (IAB) Contextual Taxonomy, http://www.iab.net/ Retrieved December 2017
18. Kanagasabai, R., Veeramani, A., Ngan, L. D., Yap, G. E., Decraene, J., Nash, A. S. (2014). Using Semantic Technologies to Mine Customer Insights in Telecom Industry. In International Semantic Web Conference (Industry Track).
19. Landauer, T. K., and Dumais, S. T. "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge", In Psychological review, vol. 104(2), 1990, pp. 211-223.
20. Landauer, T. K., Foltz, P. W. and Laham, D. " An introduction to latent semantic analysis", In Discourse processes, vol. 25, 1998, pp. 259-284.
21. Liu, B. (2010). Sentiment Analysis and Subjectivity. In N. Indurkhya, & F. J.Damerau, Handbook of Natural Language Processing (pp. 627-665). CRC Press.
22. Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10 (pp. 79-86). Association for Computational Linguistics.
23. Petherbridge, N. "Artifical Intelligence Scripting Language - Rivescript.com," (online)
24. G. Pilato, E. D'Avanzo - Data-driven Social Mood Analysis through the Conceptualization of Emotional Fingerprints - Procedia Computer Science, 2018, (in press)
25. Santilli, S., Nota, L., Pilato, G.: The use of latent semantic analysis in the positive psychology: A comparison with twitter posts. In: Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. pp. 494–498. IEEE (2017)
26. C. Strapparava and R Mihalcea. Semeval-2007 task 14: Affective text. In Proceedings of the 4th International Workshop on Semantic Evaluations, pages 70–74. Association for Computational Linguistics, 2007.
27. C. Strapparava and R Mihalcea. Learning to identify emotions in text. In SAC '08 Proceedings of the 2008 ACM symposium on Applied computing. 2008.
28. Siddharth, G., Borkar, D., De Mello, C., Patil, S. "An E-Commerce Website based Chatbot" Proc of. (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (2) , 2015, 1483-1485
29. Teh, Y. W., Newman, D., Welling, M. A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation. In NIPS, Vol. 6, pp. 1378-1385, 2006
30. Terrana D., Augello A., Pilato (2014). Facebook users relationships analysis based on sentiment classification. Proc. of 2014 IEEE International Conference on Seman- tic Computing (ICSC), pp. 290-296