Deep Tweets: from Entity Linking to Sentiment Analysis

Pierpaolo Basile, Valerio Basile, Malvina Nissim, Nicole Novielli

▶ To cite this version:

Pierpaolo Basile, Valerio Basile, Malvina Nissim, Nicole Novielli. Deep Tweets: from Entity Linking to Sentiment Analysis. Proceedings of the Italian Computational Linguistics Conference (CLiC-it 2015), Dec 2015, Trento, Italy. hal-01342435

HAL Id: hal-01342435

https://hal.inria.fr/hal-01342435

Submitted on 6 Jul 2016

Deep Tweets: from Entity Linking to Sentiment Analysis

Pierpaolo Basile[1] and **Valerio Basile[2]** and **Malvina Nissim[2]** and **Nicole Novielli[1]**

[1]Department of Computer Science, University of Bari Aldo Moro

[2]Center for Language and Cognition Groningen, Rijksuniversiteit Groningen

[1]{pierpaolo.basile,nicole.novielli}@uniba.it, [2]{v.basile,m.nissim}@rug.nl

Abstract

English. The huge amount of information streaming from online social networking is increasingly attracting the interest of researchers on sentiment analysis on micro-blogging platforms. We provide an overview on the open challenges of sentiment analysis on Italian tweets. We discuss methodological issues as well as new directions for investigation with particular focus on sentiment analysis of tweets containing figurative language and entity-based sentiment analysis of micro-posts.

Italiano. *L'enorme quantità di informazione presente nei social media attira sempre più l'attenzione della ricerca in sentiment analysis su piattaforme di micro-blogging. In questo articolo si fornisce una panoramica sui problemi aperti riguardo l'analisi del sentimento di tweet in italiano. Si discute di problemi metodologici e nuove direzioni di ricerca, con particolare attenzione all'analisi della polarità di tweet contenenti linguaggio figurato e riguardo specifiche entità nel micro-testo.*

1 Introduction

Flourished in the last decade, sentiment analysis is the study of the subjectivity and polarity (positive vs. negative) of a text (Pang and Lee, 2008). Traditionally, sentiment analysis techniques have been successfully exploited for opinionated corpora, such as news (Wiebe et al., 2005) or reviews (Hu and Liu, 2004). With the worldwide diffusion of social media, sentiment analysis on micro-blogging (Pak and Paroubek, 2010) is now regarded as a powerful tool for modelling socio-economic phenomena (O'Connor et al., 2010; Jansen et al., 2009).

The success of the tasks of sentiment analysis on Twitter at SemEval since 2013 (Nakov et al., 2013; Rosenthal et al., 2014; Rosenthal et al., 2015) attests this growing trend (on average, 40 teams per year participated). In 2014, Evalita also successfully opens a track on sentiment analysis with SENTIPOLC, the task on sentiment and polarity classification of Italian Tweets (Basile et al., 2014). With 12 teams registered, SENTIPOLC was the most popular task at Evalita 2014, confirming the great interest of the NLP community in sentiment analysis on social media, also in Italy.

In a world where e-commerce is part of our everyday life and social media platforms are regarded as new channels for marketing and for fostering trust of potential customers, such great interest in opinion mining from Twitter isn't surprising. In this scenario, what is also rapidly gaining more and more attention is being able to mine opinions about *specific aspects* of objects. Indeed, interest in Aspect Based Sentiment Analysis (ABSA) is increasing, and SemEval dedicates now a full task to this problem, since 2014 (Pontiki et al., 2014). Given a target of interest (e.g., a product or a brand), ABSA traditionally aimed at summarizing the content of users' reviews in several commercial domains (Hu and Liu, 2004; Ganu et al., 2013; Thet et al., 2010). In the context of ABSA, an interesting task is represented by finer-grained assignment of sentiment to entities. To this aim, mining information from micro-blogging platforms also involves reliably identifying entities in tweets. Hence, entity linking on twitter is gaining attention, too (Guo et al., 2013).

Based on the above observations, we discuss open issues in Section 2. In Section 3, we propose an extension of the SENTIPOLC task for Evalita 2016 by also introducing entity-based sentiment analysis as well as polarity detection of messages containing figurative language. Finally, we discuss the feasibility of our proposal in Section 4.

2 Open Challenges

From an applicative perspective, microposts comprise an invaluable wealth of data, ready to be mined for training predictive models. Analysing the sentiment conveyed by microposts can yield a competitive advantage for businesses (Jansen et al., 2009) and mining opinions about specific aspects of entities being discussed is of paramount importance in this sense. Beyond the pure commercial application domain, analysis of microposts can serve to gain crucial insights about political sentiment and election results (Tumasjan et al., 2010), political movements (Starbird and Palen, 2012), and health issues (Michael J. Paul, 2011).

By including explicit reference to entities, ABSA could broaden its impact beyond its traditional application in the commercial domain. While classical ABSA focus on the sentiment/opinion with respect to a particular aspect, entity-based sentiment analysis (Batra and Rao, 2010) tackles the problem of identifying the sentiment about an entity, for example a politician, a celebrity or a location. Entity-based sentiment analysis is a topic which has been unexplored in evaluation campaigns for Italian, and which could gain the interest of the NLP community.

Another main concern is the correct polarity classification of tweets containing figurative language such as irony, metaphor, or sarcasm (Karoui et al., 2015). Irony has been explicitly addressed so far in both the Italian and the English (Ghosh et al., 2015) evaluation campaigns. In particular, in the SENTIPOLC irony detection task, participants were required to develop systems able to decide whether a given message was ironic or not. In a more general vein, the SemEval task invited participants to deal with different forms of figurative language and the goal of the task was to detect polarity of tweets using it. In both cases, participant submitted systems obtaining promising performance. Still, the complex relation between sentiment and figurative use of language needs to be further investigated. While, in fact, irony seems to mainly act as a polarity reverser, other linguistic devices might impact sentiment in different ways.

Traditional approaches to sentiment analysis treat the subjectivity and polarity detection mainly as text classification problems, exploiting machine-learning algorithms to train supervised classifiers on human-annotated corpora. Sentiment analysis on micro-blogging platforms poses new challenges due to the presence of slang, misspelled words, hashtags, and links, thus inducing researchers to define novel approaches that include consideration of micro-blogging features for the sentiment analysis of both Italian (Basile et al., 2014) and English (Rosenthal et al., 2015) tweets. Looking at the reports of the SemEval task since 2013 and of the Evalita challenge in 2014, we observe that almost all submitted systems relied on supervised learning.

Despite being in principle agnostic with respect to language and domain, supervised approaches are in practice highly domain-dependent, as systems are very likely to perform poorly outside the domain they are trained on (Gamon et al., 2005). In fact, when training classification models, it is very likely to include consideration of terms that associate with sentiment because of the context of use. It is the case, for example, of political debates, where names of countries afflicted by wars might be associated to negative sentiments; analogous problems might be observed for the technology domain, where killer features of devices referred in positive reviews by customers usually become obsolete in relatively short periods of time (Thelwall et al., 2012). While representing a promising answer to the cross-domain generalizability issue of sentiment classifiers in social web (Thelwall et al., 2012), unsupervised approaches have not been exhaustively investigated and represent an interesting direction for future research.

3 Task Description

Entity linking and sentiment analysis on Twitter are challenging, attractive, and timely tasks for the Italian NLP community. The previously organised task within Evalita which is closest to what we propose is SENTIPOLC 2014 (Basile et al., 2014). However, our proposal differs in two ways. First, sentiment should be assigned not only at the message level but also to entities found in the tweet. This also implies that entities must be first recognised in each single tweet, and we expect systems also to link them to a knowledge base. The second difference has to do with the additional irony layer that was introduced in SENTIPOLC. Rather than dealing with irony only, we propose a *figurative* layer, that encompasses irony and any other shifted sentiment.

The entity linking task and the entity-based polarity annotation subtask of the sentiment analysis
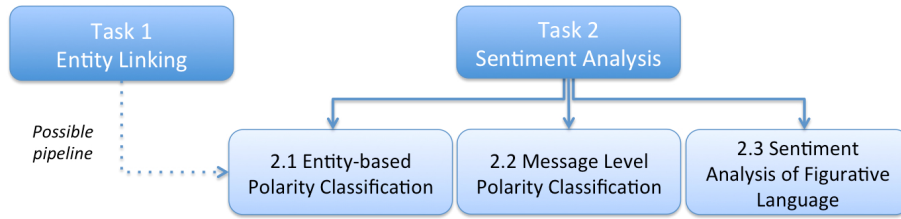
Figure 1: Task organization scheme

task can be seen as separate or as a pipeline, for those who want to try develop an end-to-end system, as depicted in Fig. 1.

3.1 Task 1 - Entity Detection and Linking

The goal of entity linking is to automatically extract entities from text and link them to the corresponding entries in taxonomies and/or knowledge bases as DBpedia or Freebase. Only very recently, entity linking in Twitter is becoming a popular tasks for evaluation campaigns (see Baldwin et al. (2015)).

Entity detection and linking tasks are typically organized in three stages: 1) identification and typing of entity mention in tweets; 2) linking of each mention to an entry in a knowledge-base representing the same real world entity, or NIL in case such entity does not exist; 3) cluster all NIL entities which refer to the same entity.

3.2 Task 2 - Message Level and Entity-Based Sentiment Analysis

The goal of the SENTIPOLC task at Evalita 2014 was the sentiment analysis at a message level on Italian tweets. SENTIPOLC was organized so as to include subjectivity and polarity classification as well as irony detection.

Besides the traditional task on message-level polarity classification, in the next edition of Evalita special focus should be given to entity-based sentiment analysis. Given a tweet containing a marked instance of an entity, the classification goal would be to determine whether positive, negative or neutral sentiment is attached to it.

As for the role of figurative language, the analysis of the performance of the systems participating in SENTIPOLC irony detection subtask shows the complexity of this issue. Thus, we believe that further investigation of the role of figurative language in sentiment analysis of tweets is needed, by also incorporating the lesson learnt from the task on figurative language at SemEval 2015 (Ghosh et

al., 2015). Participants would be required to predict the overall polarity of tweets containing figurative language, distinguishing between the literal meaning of the message and its figurative, intended meaning.

4 Feasibility

The annotated data for entity linking tasks (such as our proposed Task 1) typically include the start and end offsets of the entity mention in the tweet, the entity type belonging to one of the categories defined in the taxonomy, and the URI of the linked DBpedia resource or to a NIL reference. For example, given the tweet `@FabioClerici sono altri a dire che  un reato. E il "politometro" come lo chiama #Grillo vale per tutti.  Anche per chi fa #antipolitica`, two entities are annotated: `FabioClerici` (offsets 1-13) and `Grillo` (offsets 85-91). The first entity is linked as NIL since `Fabio Clerici` has not resource in DBpedia, while `Grillo` is linked with the respective URI: `http://dbpedia. org/resource/Beppe_Grillo`. Analysing similar tasks for English, we note that organizers provide both training and test data. Training data are generally used in the first stage, usually approached by supervised systems, while the linking stage is generally performed using unsupervised or knowledge based systems. As knowledge base, the Italian version of DBpedia could be adopted, while the entity taxonomy could consist of the following classes: Thing, Event, Character, Location, Organization, Person and Product.

As for Task 2, it is basically conceived as a follow-up of the SENTIPOLC task. In order to ensure continuity, it makes sense to carry out the annotation using a format compatibile with the existing dataset. The SENTIPOLC annotation scheme consists in four binary fields, indicating the presence of subjectivity, positive polarity, negative polarity, and irony. The fields are not mutually ex-

clusive, for instance both positive and negative polarity can be present, resulting in a *mixed* polarity message. However, not all possible combinations are allowed. Table 1 shows some examples of annotations from the SENTIPOLC dataset.

Table 1: Proposal for an annotation scheme that distinguishes between literal polarity (`pos`, `neg`) and overall polarity (`opos`, `oneg`).

subj	pos	neg	iro	**opos**	**oneg**	description
0	0	0	0	**0**	**0**	objective tweet
1	1	0	0	**1**	**0**	subjective tweet positive polarity no irony
1	0	0	0	**0**	**0**	subjective tweet neutral polarity no irony
1	0	1	0	**0**	**1**	subjective tweet negative polarity no irony
1	0	1	1	**1**	**0**	subjective tweet negative literal polarity positive overall polarity
1	1	0	1	**0**	**1**	subjective tweet positive literal polarity negative overall polarity

With respect to the annotation adopted in SENTIPOLC, two additional fields are reported to reflect the task organization scheme we propose in this paper, including the sentiment analysis of tweet containing figurative language. These fields, highlighted in bold face, encode respectively the presence of positive and negative polarity *considering the eventual polarity inversion due to the use of figurative language*, thus the existing `pos` and `neg` fields refer to *literal* polarity of the tweet. For the annotation of the gold standard dataset for SENTIPOLC, the polarity of ironic messages has been annotated according to the intended meaning of the tweets, so for the new task the literal polarity will have to be manually annotated in order to complete the gold standard. Annotation of items in the figurative language dataset could be the same as in the message-level polarity detection task of Evalita, but tweets would be opportunistically selected only if they contain figurative language, so as to reflect the goal of the task.

For the entity-based sentiment analysis subtask, the boundaries for the marked instance will be also provided by indicating the offsets of the entity for which the polarity is annotated, as it was done for SemEval (Pontiki et al., 2014; Pontiki et al., 2015). Participants who want to attempt entity-based sentiment analysis only can use the data that contains the gold output of Task 1, while those who want to perform entity detection and linking only, without doing sentiment analysis, are free to stop there. Participants who want to attempt both tasks can treat them in sequence, and evaluation can be performed for the whole system as well as for each of the two tasks (for the second one over gold input), as it will be done for teams that participate in one task only.

For both tasks, the annotation procedure could follow the consolidated methodology from the previous tasks, such as SENTIPOLC. Experts label manually each item, then agreement is checked and disagreements are resolved by discussion.

Finally, so far little investigation was performed about unsupervised methods and the possiblity they offer to overcome domain-dependence of approaches based on machine learning. In a challenge perspective, supervised systems are always promising since they guarantee a better performance. A possible way to encourage teams to explore original approaches could be to allow submission of two separate runs for supervised and unsupervised settings, respectively. Ranking could be calculated separately, as already done for the constrained and unconstrained runs in SENTIPOLC. To promote a fair evaluation and comparison of supervised and unsupervised systems, corpora from different domains could be provided as train and test sets. To this aim, it could be possible to exploit the *topic* field in the annotation of tweets used in the SENTIPOLC dataset, where a flag indicates whether the tweets refer to the political domain or not. Hence, the train set could be built by merging political tweets from both the train and the test set used in SENTIPOLC. A new test set would be created by annotating tweets in one or more different domains.

To conclude, we presented the entity linking and sentiment analysis tasks as related to one another, as shown in the pipeline in Figure 1, specifying that participants will be able to choose which portions of the tasks they want to concentrate on. Additionally, we would like to stress that this proposal could also be conceived as two entirely separate tasks: one on sentiment analysis at the entity level, including entity detection and linking, and one on sentiment analysis at the message level, including the detection of figurative readings, as a more direct follow-up of SENTIPOLC.

Acknowledgements

References

Timothy Baldwin, Marie Catherine de Marneffe, Bo Han, Young-Bum Kim, Alan Ritter, and Wei Xu. 2015. Shared tasks of the 2015 workshop on noisy user-generated text: Twitter lexical normalization and named entity recognition. In *Association for Computational Linguistics (ACL)*. ACL, Association for Computational Linguistics, August.

Valerio Basile, Andrea Bolioli, Malvina Nissim, Viviana Patti, and Paolo Rosso. 2014. Overview of the Evalita 2014 SENTIment POLarity Classification Task. In *Proc. of EVALITA 2014*, Pisa, Italy.

Siddharth Batra and Deepak Rao. 2010. Entity based sentiment analysis on twitter. *Science*, 9(4):1–12.

Michael Gamon, Anthony Aue, Simon Corston-Oliver, and Eric Ringger. 2005. Pulse: Mining customer opinions from free text. In *Proceedings of the 6th International Conference on Advances in Intelligent Data Analysis*, IDA'05, pages 121–132, Berlin, Heidelberg. Springer-Verlag.

Gayatree Ganu, Yogesh Kakodkar, and AméLie Marian. 2013. Improving the quality of predictions using textual information in online user reviews. *Inf. Syst.*, 38(1):1–15.

AAniruddha Ghosh, Guofu Li, Tony Veale, Paolo Rosso, Ekaterina Shutova, Antonio Reyes, and Jhon Barnden. 2015. Semeval-2015 task 11: Sentiment analysis of figurative language in twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 470–475, Denver, Colorado, USA. Association for Computational Linguistics.

Stephen Guo, Ming-Wei Chang, and Emre Kiciman. 2013. To link or not to link? a study on end-to-end tweet entity linking. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1020–1030, Atlanta, Georgia, June. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177.

Bernard J. Jansen, Mimi Zhang, Kate Sobel, and Abdur Chowdury. 2009. Twitter power: Tweets as electronic word of mouth. *J. Am. Soc. Inf. Sci. Technol.*, 60(11):2169–2188.

Jihen Karoui, Farah Benamara, Véronique Moriceau, Nathalie Aussenac-Gilles, and Lamia Hadrich Belguith. 2015. Towards a contextual pragmatic model to detect irony in tweets. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, ACL 2015, July 26-31, 2015, Beijing, China, Volume 2: Short Papers*, pages 644–650.

Mark Dredze Michael J. Paul. 2011. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 265–272.

Preslav Nakov, Sara Rosenthal, Zornitsa Kozareva, Veselin Stoyanov, Alan Ritter, and Theresa Wilson. 2013. Semeval-2013 task 2: Sentiment analysis in twitter. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pages 312–320, Atlanta, Georgia, USA, June. Association for Computational Linguistics.

Brendan O'Connor, Ramnath Balasubramanyan, Bryan Routledge, and Noah Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. In *Intl AAAI Conf. on Weblogs and Social Media (ICWSM)*, volume 11, pages 122–129.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proc. of the Seventh Intl Conf. on Language Resources and Evaluation (LREC'10)*.

Bo Pang and Lillian Lee. 2008. Opinion Mining and Sentiment Analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135, January.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August. Association for Computational Linguistics and Dublin City University.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion . 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 486–495, Denver, Colorado, June. Association for Computational Linguistics.

Sara Rosenthal, Alan Ritter, Preslav Nakov, and Veselin Stoyanov. 2014. SemEval-2014 Task 9: Sentiment Analysis in Twitter. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 73–80, Dublin, Ireland, August.

Sara Rosenthal, Preslav Nakov, Svetlana Kiritchenko, Saif M Mohammad, Alan Ritter, and Veselin Stoyanov. 2015. SemEval-2015 Task 10: Sentiment Analysis in Twitter. In *Proceedings of the 9th International Workshop on Semantic Evaluation*, SemEval '2015, Denver, Colorado, June.

Kate Starbird and Leysia Palen. 2012. (how) will the revolution be retweeted?: Information diffusion and the 2011 egyptian uprising. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 7–16, New York, NY, USA. ACM.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2012. Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology*, 63(1):163–173.

Tun Thura Thet, Jin-Cheon Na, and Christopher S.G. Khoo. 2010. Aspect-based sentiment analysis of movie reviews on discussion boards. *J. Inf. Sci.*, 36(6):823–848.

Andranik Tumasjan, Timm Sprenger, Philipp Sandner, and Isabell Welpe. 2010. Predicting elections with twitter: What 140 characters reveal about political sentiment.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 1(2).