10.12871/clicit2014216

The *FICLIT+CS@UniBO* System at the EVALITA 2014 Sentiment Polarity Classification Task.

Pierluigi Di Gennaro
DISI - University of Bologna, Italy
pierluigi.digennaro@gmail.com

Arianna Rossi
FICLIT - University of Bologna, Italy
ar.ariannarossi@gmail.com

Fabio Tamburini
FICLIT - University of Bologna, Italy
fabio.tamburini@unibo.it

Abstract

English. This paper presents a work in progress on the design of a sentiment polarity classification system that participates in the EVALITA 2014 SENTIPOLC task. Although we have been working on the system implementation for only three months, the results are promising, as the system ranked 5th (out of 9) in the subjectivity detection task and 7th (out of 11) in the sentiment polarity classification task.

Italiano. *Questo contributo presenta la progettazione di un sistema automatico per la classificazione della* sentiment polarity *che ha partecipato al task SENTIPOLC della campagna di valutazione EVALITA 2014. Nonostante i soli tre mesi di sviluppo, i risultati parziali sono promettenti in quanto il sistema si è classificato 5° (su 9) nel task di identificazione della soggettività e 7° (su 11) nel task relativo all'identificazione della polarità.*

1 Introduction

We developed two different approaches to Sentiment Polarity detection for the EVALITA 2014 SENTIPOLC task: (a) we started from the seminal paper (Basile, Nissim, 2013) and applied the same algorithm that had been proposed, but on a different lexicon, that was specifically developed for this system, and (b) we tried to devise more complex syntactically-driven polarity combination techniques.

In section 2 we describe the development of the annotated lexicon, in section 3 we illustrate the procedures applied by the proposed system, in section 4 we describe the system for the Subjectivity Classification task and, lastly, in section 5, we discuss the overall results obtained in the EVALITA 2014 Sentiment Polarity Classification task.

2 Sentiment-lexicon generation

Our lexicon was created by collecting words from various sources and was annotated using a semi-automatic polarity classification procedure. Sentiment polarity shifters were also taken into account and inserted into the lexicon.

2.1 Adjectives and Adverbs

We started by considering all the adjectives and adverbs extracted from the De Mauro - Paravia Italian dictionary (2000). All the glosses connected to the different senses of each lemma were automatically classified by using the online Sentiment Analysis API provided by *Ai Applied*[1]. This automatic procedure assigned either a positive or a negative polarity score to each lemma/sense pair in the intervals [-1,-0.5], for negative polarity, and [0.5,1], for positive polarity.

2.2 Nouns and Verbs

Although adjectives and adverbs are widely considered to be a primary source of subjective content in a text (Taboada *et al.*, 2011), also some nouns and verbs have a polarity value. We extracted nouns and verbs from Sentix (Basile, Nissim, 2013), since we expected those lemmas to be a selected choice of sentiment words, and used the automatic procedure seen above to classify their polarity.

2.3 Manual check

The polarity lexicon annotated with the automatic procedure described above was then inspected

[1]http://ai-applied.nl/sentiment-analysis-api

manually to clean it up. When the API had assigned a wrong polarity score, a value of 1.01 or -1.01 was assigned to the word, in order to clearly discriminate the automatic from the manually assigned values for future work. In addition, all the lemmas that had an objective value were left out and were not considered in our system, assigning to them a conventional polarity value equal to 0.

2.4 Everyday language and abbreviations

Lastly, the specific features of the informal language of social media were taken into account and all those words that our system could not identify from the tweets' development set were then extracted. By doing so, we were able to collect several words used in everyday language, i.e. *cazzata* (bullshit), *coglione* (moron), and abbreviations, i.e. *tt, nn* (not translatable), that were not yet included in our lexicon and assign a polarity value to them.

2.5 Sentiment polarity shifters

There are several linguistic phenomena that can cause a shift of the polarity of a word from one pole to the other or intensify its semantic intensity (Taboada *et al.*, 2011). Only negators and shifters were considered in the current approach, but others will be taken into account in our future research.

1. **Negators**: words like *non* (not), *nessuno* (nobody), *niente* (nothing), *nulla* (nothing), *mai* (never), etc. reverse the polarity of sentiment words (Polanyi, Zaenen, 2006). A value of -1 was assigned to negators, so that, in a sentence like *Non si vede bene* (You can not see well), *non* negates *bene* and flip its polarity from + 0,76 to -0,76.

2. **Intensifiers**: they increase or decrease the semantic intensity of the lexical item(s) they accompany (Taboada *et al.*, 2011). A positive percentage was assigned to amplifiers, whereas a negative one was assigned to downtoners, as shown in Table 1. This percentual value multiplies the polarity score of the opinion word, so if, for example, *felice* (happy) has a positive score of 0.84, *molto felice* (very happy) will have a positive score of: 0.84 × (1 + 0.25)= 1.05. The same procedure was applied to words accompanied by downtoners, so if, for instance, *grave* (serious) as a negative value of 0.7, *poco grave* (not very serious) will have a value of: -0.7 × (1 - 0.25)= -0.52.

Intensifiers	Value
completamente	+0.75
drasticamente	+0.50
molto	+0.25
abbastanza	-0.15
poco	-0.25
leggermente	-0.50

Table 1: Percentages for some positive and negative intensifiers

2.6 Context-dependent words

A large set of words do not have a positive or negative value *per se*, but, on the contrary, they can take a different value depending on the context they happen (Liu, 2012). For example, in an expression like *maniere forti* (strong-arm methods), *forte* (strong) has a negative meaning, whereas in *forte legame* (strong link) it has a positive one. Moreover, some of these words are objective in most domains, but they can acquire a subjective value in others. The word *poeta* (poet), for instance, can be objective, as in *Dante è stato un poeta del XIII secolo* (Dante was a poet of the 13th century), but can also have a subjective metaphorical meaning, as in *Luca scrive delle lettere bellissime. È proprio un poeta!* (Luca writes wonderful letters. He's really a poet!). We decided not to consider context-dependent words in our system since they need a more sophisticated approach that involves word sense disambiguation and metaphor detection.

3 System implementation

As a first step for the development of our sentiment polarity classification system, we implemented the algorithm proposed in the seminal paper (Basile, Nissim, 2013). Starting from their corpus of Italian tweets called TWITA, they developed a simple system which assigns one out of three possible values – positive, neutral or negative – to a given tweet. In order to assign the values, the system extracts the information from a polarity lexicon that was specifically developed thanks to various general lexical resources, namely SentiWordNet (Esuli, Sebastiani, 2006; Baccianella *et al.*, 2010), Multi-WordNet (Pianta *et al.*, 2002) and WordNet (Fellbaum, 1998). We developed the same algo-

rithm that was proposed in (Basile, Nissim, 2013), but we used instead the lexicon described in section 2, considering it as the starting point, or baseline, for any further improvement.
We can summarize the process in the following steps:

1. The system calculates the polarity score of each entry in the lexicon as the mean of the different word senses' scores.

2. Given a tweet, the system assigns a polarity score to each of its tokens by matching them to the lexicon.

3. The system calculates the polarity score of a complete tweet as the sum of the different polarity scores of its tokens: a polarity score greater than 0 indicates a positive tweet, a polarity score lower than 0 indicates a negative tweet, a polarity score equal to 0 indicates a neutral tweet.

In view of the results and thanks to the experience obtained from this development, we also tried to devise more complex syntactically-driven polarity combination techniques.

3.1 Token processing

Before proceeding with the syntactic analysis, we applied some rules of substitution or elimination to all those textual parts that were irrelevant to the classification task or that could hinder POS-tagging, lemmatization and parsing. In particular:

- a generic label "*URL*" replaced URLs (http://abc.org);

- character # and @ were removed from hashtags (#abc) and mentions (@abc);

- a generic label "*EMOPOS*" replaced positive emoticons (see table 2)

- a generic label "*EMONEG*" replaced negative emoticons (see table 2)

We added the labels "EMOPOS" and "EMONEG" to the lexicon, and associated them to a polarity score of 1.0 and -1.0 respectively.

3.2 Syntactic analysis

Our system relies on the TULE parser (Lesmo, 2007) to analyze the syntactic structure of a single tweet. TULE includes a tokenizer, a morphological analyzer, a PoS-tagger and a dependency parser. It takes a natural language sentence as input and returns a dependency tree that describes its syntactic structure. For each token identified, the parser output includes its PoS-tag, the lemma and other morphological information about it.

<table>
<tr><th>Label</th><th>Emoticon</th></tr>
<tr><td>EMOPOS</td><td>(: :) :] [: :-) (-: [-: :-] (; ;) ;] [; ;-) (-; [-; ;-] :-D :D :-p :p (=: ;=D :=) :S @-) XD</td></tr>
<tr><td>EMONEG</td><td>:( ): :-( )-: ;( ); :-[ ]-: ;-( )-; :[:( ):]: :[:| :/ |: /: :=(:=| :=[xo :| D: O:</td></tr>
</table>

Table 2: Emoticons' list.

As one would expect, we found some difficulties in using TULE on certain tweets, thus we added a few pre-processing and filtering steps:

- *special characters*: special characters (i.e. *$*) were replaced by their equivalent Italian word (i.e. *dollaro*).

- *shortened URLs*: due to limited tweet length, Twitter can cut an URL; these were removed from the tweets.

Our system uses adjacency lists (based on Boost library) with only one root node to represent dependency parser trees. Each node represents a token and contains all the relevant information about it: POS-tag, lemma, lexicon category (negator or intensifier) and polarity score. The system assigns a polarity score to a token by matching its lemma to the lexicon. If the lemma can not be found, three options are taken into account:

- *The polarity score of the lemma is 0*: a polarity score equal to 0 is conventionally assigned to the token.

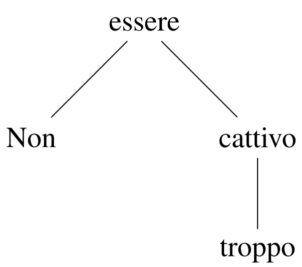
- *The lemma is a polarity shifter*: the polarity score equals the intensification value of the shifter;

- *The lemma is not a polarity shifter*: the polarity score corresponds to the mean of the different word senses' scores.

When the polarity score of each tree node (i.e. each word in the sentence) has been calculated, the system assigns a polarity score to the whole tweet by applying a set of polarity propagation rules to the dependency tree. The system can choose between two options:

- *All tokens in a given sentence are not polarity shifters*: the polarity score is the sum of the polarity scores of each token.
- *One or more tokens in a given sentence are polarity shifters*: polarity shifters increase, decrease or reverse the polarity score of the item(s) linked to it. Starting from the polarity shifter that is closest to the leaves of the parse tree, the system sums the polarity score of the nodes linked to it and then multiplies this value by the polarity shifter's value.

For example the polarity score (PS) of the sentence *Non essere troppo cattivo* (Do not be too bad) is obtained as follows:



[(PS(*cattivo*) × (PS(*troppo*) + 1)) + PS(*essere*)] × PS(*Non*)

A tweet can be composed by more than one sentence. In this case, its final polarity score is obtained by summing all the polarity scores of its sentences.

Lastly, the system classifies a complete tweet as:

- *positive* if its polarity score is higher than 0;
- *neutral* if its polarity score is equal to 0;
- *negative* if its polarity score is lower than 0.

4 Subjectivity classification Task

Starting from the assumption that sentiment polarity and subjectivity classification are closely related, we used the results of our system described in section 3 to define whether a tweet is subjective or objective. Thus, we did not to implement a different system for subjectivity classification, but instead we derive subjectivity classification from sentiment polarity.

Given a tweet, it is classified as objective if its polarity score is equal to 0, otherwise it is classified as subjective. We are conscious that this is a coarse-grain approximation. If neutral tweets can only be objective, positive and negative tweets can be subjective or objective. We postponed the development of a better subjectivity classification system for further developments.

Rank	Combined F-score	F-score (0)	F-score (1)
1	0.7140	0.6005	0.8275
2	0.6871	0.5819	0.7923
3	0.6706	0.5344	0.8067
4	0.6497	0.4868	0.8127
-	<u>0.6134</u>	<u>0.4514</u>	<u>0.7755</u>
5	**0.5972**	**0.4480**	**0.7464**
6	0.5901	0.5031	0.6770
7	0.5825	0.4200	0.7451
8	0.5593	0.4424	0.6761
9	0.5224	0.3237	0.7211
10	*0.4005*	*0.0000*	*0.8010*

Table 3: Task 1 results – Constrained run, Subjectivity detection. In bold face the official results from the proposed system, underlined the results obtained using only the lexicon and in italics the baseline.

5 Results and discussion

Tables 3 and 4 present the results of the proposed system in the Subjectivity and Polarity Detection tasks respectively.

Although we have worked on the system implementation for only three months, the results are promising, as it ranked 5th (out of 9) in the subjectivity detection task and 7th (out of 11) in the sentiment polarity classification task. We did not participate in the irony detection task.

As we can see from Tables 3 and 4, our official results, produced by combining the new annotated lexicon with the complex algorithm for propagating lexical polarity values across dependency trees, do not exceed the unofficial results obtained by using only the lexicon.

The polarity propagation process is not problem-free and in the future we will consistently improve it, in order to obtain more reliable results. Also the lexicon must be improved: more lemmas must be inserted and the annotation schema can be enhanced by rethinking some of its features.

Rank	Combined F-score	Pos. Pol. F-score	Neg. Pol. F-score
1	0.6771	0.6752	0.6789
2	0.6347	0.6196	0.6498
3	0.6312	0.6352	0.6271
4	0.6299	0.6277	0.6321
-	<u>0.6062</u>	<u>0.5941</u>	<u>0.6184</u>
5	0.6049	0.6079	0.6019
6	0.6026	0.6153	0.5899
7	**0.5980**	**0.5940**	**0.6019**
8	0.5626	0.5556	0.5695
9	0.5342	0.5293	0.5390
10	0.5181	0.5021	0.5341
11	0.5086	0.5159	0.5013
12	*0.3718*	*0.3977*	*0.3459*

Table 4: Task 2 results – Constrained run, Polarity detection. In bold face the official results from the proposed system, underlined the results obtained using only the lexicon and in italics the baseline.

References

Baccianella S., Esuli A. and Sebastiani F. 2010. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of LREC 2010*, Valletta, Malta, 2200–2204.

Basile V. and Nissim M. 2013. Sentiment Analysis on Italian Tweets. *Proceedings of the 4th ACL Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Atlanta, 100–107.

De Mauro T.. 2000. *Il dizionario della lingua italiana*. Paravia.

Esuli A. and Sebastiani F. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. *Proceedings of LREC 2006*, Genova, 417–422.

Fellbaum C., editor. 2000. *WordNet. An Electronic Lexical Database*. The MIT Press.

Lesmo L. 1983. The Rule-Based Parser of the NLP Group of the University of Torino. *Intelligenza Artificiale*, 2(4):46–47.

Liu B. 2012 Sentiment Analyses and Opinion Mining Synthesis Lectures on Human Language Technologies, 5,1 (2012): 1-167.

Pianta E., Bentivogli L. and Girardi C. 2002. MultiWordNet: developing an aligned multilingual database. *Proceedings of the First International Conference on Global WordNet*, Mysore, India, 21–25.

Polanyi L. and Zaenen A. 2006 Contextual Valence Shifters *Computing attitude and affect in text: Theory and applications*, Springer Netherlands, 1-10.

Taboada M., Brooke J., Tofiloski M., Voll K. and Stede M. 2011 Lexicon-Based Methods for Sentiment Analyses *Computational Linguistics*, 37(2):267–307.

Wiegand, M., Balahur, A., Roth, B., Klakow, D. and Montoyo, A. 2010 A survey on the role of negation in sentiment analysis *Proceedings of the ACL workshop on negation and speculation in natural language processing*, 60-68.