

OPINION POLARITY DETECTION

Using Word Sense Disambiguation to Determine the Polarity of Opinions

Tamara Martín-Wanton, Aurora Pons-Porrata

Center for Pattern Recognition and Data Mining, Universidad de Oriente, Patricio Lumumba s/n, Santiago de Cuba, Cuba

Andrés Montoyo-Guijarro, Alexandra Balahur

Department of Software and Computing Systems, University of Alicante, Alicante, Spain

Keywords: Opinion Mining, Polarity Detection, Word Sense Disambiguation.

Abstract: In this paper, we present an unsupervised method for determining the polarity of opinions. It uses a word sense disambiguation algorithm to determine the correct sense of the words in the opinion. The method is also based on *SentiWordNet* and *General Inquirer* to determine the polarity of the senses. Due to the characteristics of these external resources, the proposed method does not depend on the knowledge domain and can be extended to other languages. In the evaluation carried out over the *SemEval Task No. 14: Affective Text* data our method outperforms both unsupervised and supervised systems presented in this task.

1 INTRODUCTION

Opinion Mining (also known as sentiment classification or subjectivity analysis) refers to a broad area of Natural Language Processing and Text Mining. It is concerned not with the topic a document is about, but with the opinion it expresses, that is, its aim is to determine the attitude (feelings, emotions and subjectivities) of a speaker or a writer with respect to some topic. A major task of Opinion Mining is the classification of the opinion's polarity, which consists in determine whether the opinion is positive, negative or neutral with respect to the entity to which it is referring (e.g., a person, a product, a movie, etc.).

Most existing approaches apply supervised learning techniques, including Support Vector Machines, Naïve Bayes, AdaBoost and others. On the other hand, unsupervised approaches are based on external resources such as *WordNet Affect* or *SentiWordNet*. Supervised techniques, even having better results, have several disadvantages: they are subject to overtraining and are highly dependent on the quality, size and domain of the training data.

In this paper, a new unsupervised method for determining the polarity of opinions is presented. It is based on the assumption that the same word in different contexts may not have the same polarity.

For example, the word “drug” can be positive or negative depending on the context where it appears (“she takes drugs for her heart”, “to be on drugs”). With this aim, we use a word sense disambiguation algorithm to get the correct sense of words in the opinion and the polarity of the senses is obtained from the annotations of *SentiWordNet* and *General Inquirer*. The proposed method also handles negations and other polarity shifters obtained from the *General Inquirer* dictionary. Due to the characteristics of the used resources, this method does not depend on neither the knowledge domain, nor the language. The method is evaluated over the *SemEval Task No. 14: Affective Text* data.

2 USED RESOURCES

The proposed method for determining the polarity of opinions uses the following resources: *WordNet*, *SentiWordNet* and a subset of the *General Inquirer*.

WordNet (Miller et al., 1993) is a lexical database based on psycholinguistic theories about the mental lexicon. In *WordNet* the words are grouped into sets of synonyms (synsets). Each synset is provided with a glossary and can be connected to other synsets by semantic relations (e.g., hypernymy, hyponymy, antonym, etc.). There

are versions for various languages. Each of these is interconnected with the version in English by an interlingual index. This fact allows the methods based on *WordNet* to be independent on the language.

SentiWordNet (Esuli and Sebastiani, 2006) is a lexical resource for opinion mining. Each synset in *WordNet* has assigned three values of sentiment: positive, negative and objective, whose sum is 1. It was semi-automatically built so all the results were not manually validated and some resulting classifications can appear incorrect. For example, FLU#1 (*an acute febrile highly contagious viral disease*), is annotated as Positive = 0.75, Negative = 0.0, Objective = 0.25, despite of having a lot of negative words in its gloss.

General Inquirer (GI) (Stone et al., 1966) is an English dictionary that contains information about the words. For the proposed method we use the words labelled as positives, negatives and negations (*Positiv*, *Negativ* and *Negate* categories in GI). From the *Positiv* and *Negativ* categories, we build a list of positive and negative words respectively. From the *Negate* category we obtain a list of polarity shifters terms (also known as valence shifters).

The valence shifters are terms that can change the semantic orientation of another term (e.g., turning a negative into a positive term, "*This movie is not good*"). Examples of valence shifters are: *not*, *never*, *none* and *nobody*.

3 THE PROPOSED METHOD

The overall architecture of the polarity classifier is shown in Figure 1.

It consists of two basic components: word sense disambiguation and determination of polarity. The first, given an opinion, determines the correct senses of its terms and the second, for each word sense determines its polarity, and from them gets the polarity of the opinion.

Firstly, a pre-processing of the text is carried out including sentence recognizing, stop-word

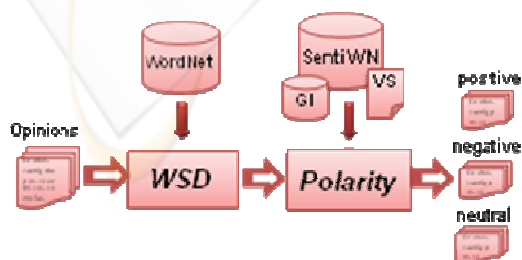


Figure 1: Overall architecture of the polarity classifier.

removing, part-of-speech tagging and word stemming by using the *TreeTagger* tool (Schmid, 1994).

Word Sense Disambiguation (WSD) consists on selecting the appropriate meaning of a word given the context in which it occurs. For the disambiguation of the words, we use the method proposed in (Anaya-Sánchez et al., 2006), which relies on clustering as a way of identifying semantically related word senses.

In this WSD method, the senses are represented as signatures built from the repository of concepts of *WordNet*. The disambiguation process starts from a clustering distribution of all possible senses of the ambiguous words by applying the Extended Star clustering algorithm (Gil-García et al., 2003). Such a clustering tries to identify cohesive groups of word senses, which are assumed to represent different meanings for the set of words. Then, clusters that match the best with the context are selected. If the selected clusters disambiguate all words, the process stops and the senses belonging to the selected clusters are interpreted as the disambiguating ones. Otherwise, the clustering are performed again (regarding the remaining senses) until a complete disambiguation is achieved.

Once the correct sense for each word on the opinion is obtained, the method determines its polarity regarding the sentiment values for this sense in *SentiWordNet* and the membership of the word to the *Positiv* and *Negativ* categories in GI. It is important to mention that the polarity of a word is forced into the opposite class if it is preceded by a valence shifter (obtained from the *Negate* category in GI).

Finally, the polarity of the opinion is determined from the scores of positive and negative words it contains. To sum up, for each word w and its correct sense s , the positive ($P(w)$) and negative ($N(w)$) scores are calculated as:

$$P(w) = \begin{cases} \text{positive value of } s \text{ in SentiWN} + 1 & \text{if } w \text{ belongs to the Positiv} \\ & \text{category in GI} \\ \text{positive value of } s \text{ in SentiWN} & \text{otherwise} \end{cases} \quad (1)$$

$$N(w) = \begin{cases} \text{negative value of } s \text{ in SentiWN} + 1 & \text{if } w \text{ belongs to the Negativ} \\ & \text{category in GI} \\ \text{negative value of } s \text{ in SentiWN} & \text{otherwise} \end{cases} \quad (2)$$

Finally, the global positive and negative scores (S_p, S_n) are calculated as:

$$S_p = \sum_{w: P(w) > N(w)} P(w) \quad S_n = \sum_{w: N(w) > P(w)} N(w) \quad (3)$$

If S_p is greater than S_n then the opinion is considered as positive. On the contrary, if S_p is less than S_n the opinion is negative. Finally, if S_p is equal to S_n the opinion is considered as neutral.

3.1 Example

The following example illustrates the method. Let us consider the headline 551 of *SemEval Task No.14 Affective Text* data: “Storms kill, knockout power, cancel flights.”

Once the WSD method is applied, we obtain the following senses (for each sense we show the word, its part-of-speech: n–noun, v–verb, a–adjective, and the sense number in *WordNet*): *storm#n#1, kill#v#3, knockout#a#1, power#n#1, cancel#v#1, flight#n#9*.

Then, from the positive and negative values of the senses in *SentiWN* and the *Positiv* and *Negativ* categories of GI showed in Table 1, we obtain the positive and negative votes for the words (for example: $P(storm) = 0, N(storm) = 0.125+1, P(knockout) = 0.375, N(knockout) = 1$). Then, $S_p = 0$ and $S_n = 1.125+1.125+1+1 = 4.250$. Therefore, the headline is classified as negative.

Table 1: Annotations of the external resources used in the example.

Sense	SentiWN		GI	
	Positive value	Negative value	Positiv	Negativ
storm#n#1	0	0.125	no	yes
kill#v#3	0	0.125	no	yes
knockout#a#1	0.375	0	no	yes
power#n#1	0	0	no	no
cancel#v#1	0	0	no	yes
flight#n#9	0	0	no	no

4 EVALUATION

In order to evaluate the proposed method, we use the data from the *SemEval Task #14: Affective Text* (Strapparava and Mihalcea, 2007). The goal of this task is to annotate news headlines for emotions and for valence (positive, negative or neutral). In this paper, we only consider the valence annotation. A specific difficulty in this task is related to the small number of words present in news headlines.

The dataset consists of 1000 news headlines obtained from major newspapers. The corpus has 4279 words of which 3275 are ambiguous (more than one sense in *WordNet*); this represents a 76.54% of the corpus. The average number of senses in ambiguous words is 6.54, and for all words 5.24. Therefore, it is remarkable that the corpus is largely ambiguous.

We follow the coarse-grained evaluation, where Accuracy (Acc.), Precision (Prec.), Recall (Rec.) and F1 measures were used. The accuracy is calculated regarding to all possible classes (positive, negative and neutral), whereas the precision and recall do not take into account the neutral class. F1 is the harmonic mean of precision and recall.

The first experiment is focused on evaluating the impact of the word sense disambiguation. With this aim, we compare the proposed method against a method based only on GI and a method that uses the most frequent baseline to disambiguate the words (WSD-MFS) (see Table 2).

The GI-based method only takes into account the lists of positive and negative words of GI and handles valence shifters to determine the polarity of the headlines. Notice that, in this case, no disambiguation is carried out. The number of positive and negative words in the headline was calculated. If the number of positive words is greater than the number of negative words, then the headline is positive. On the contrary, if the number of positive words is less than that of negative words, the headline is negative. Finally, if there are neither positive nor negative words, then the headline is neutral.

The second method only differs from the proposed method in that it uses to disambiguate the MFS baseline. In *WordNet*, senses of a same word are ranked based on the frequency of occurrence of each sense in the *SemCor* corpus; the baseline is simply to assign as correct sense to each word its first sense in *WordNet*.

Table 2: The proposed method against the GI-based method.

	Acc.	Prec.	Rec.	F1
GI-based	31.2	31.18	66.38	42.43
WSD-MFS	42.8	36.73	71.22	48.46
Our Method	44.3	37.66	72.11	49.41

As we can see, the proposed method significantly outperforms the GI-based method and is slightly better than the MFS baseline. Notice that, previous *Senseval* evaluation exercises have shown that the MFS baseline is very hard to beat by unsupervised systems (Agirre and Soroa, 2007). This confirms our hypothesis that word sense disambiguation is useful

for determining the polarity of a word. The proposed method detects a higher number of positive and negative headlines (better recall), commits few mistakes (better precision) and detects more neutral headlines (better accuracy).

Finally, we compare our method with the systems participating in *SemEval 2007 Task 14* (see Table 3). The results obtained by the unsupervised systems CLaC and UPAR7, have very low recall and high precision and, therefore, a very low value of F1, indicating that few headlines (about 35 of 410) are classified as positive and negative. Most headlines are classified as neutral; therefore, the accuracy is artificially high due to the imbalance of classes in the data (155 Positives, 255 Negatives and 590 Neutrals).

On the other hand, the supervised systems (except the SWAT that obtains very bad results) show a different behavior with respect to unsupervised systems, they have high recall but low precision. These systems detect a greater number of positive and negative headlines, but many neutral ones are misclassified. Hence, they achieve a low accuracy.

Table 3: Results of the valence annotation.

	Acc.	Prec.	Rec.	F1
Unsupervised methods				
CLaC	55.10	61.42	9.20	16.00
UPAR7	55.00	57.54	8.78	15.24
Our method	44.3	37.66	72.11	49.41
Supervised methods				
SWAT	53.20	45.71	3.42	6.36
CLaC-NB	31.20	31.18	66.38	42.43
SICS	29.00	28.41	60.17	38.60

As we can observe, the proposed method outperforms both supervised and unsupervised systems. Notice that it obtains the best F1 score and recall while achieving acceptable values of precision and accuracy. Therefore, we can conclude that our method presents a more balanced behaviour, that is, it performs well in the three classes: positive, negative and neutral.

5 CONCLUSIONS

In this paper, a new unsupervised method to opinion polarity detection has been introduced. Its most important novelty is the use of word sense disambiguation together with standard external resources for determining the polarity of the opinions. These resources allow the method to be extended to other languages and be independent of the knowledge domain.

The experiments carried out over the data of *SemEval Task No. 14* validate the usefulness of word sense disambiguation for determining the polarity of opinions. We have also shown that the proposed method outperforms both unsupervised and supervised systems participating in the competition.

Future work includes testing alternative resources for polarity detection. We believe that in many cases our approach fails because the wrong annotations of *SentiWordNet*. We also plan to evaluate the proposed method in other test collections of different knowledge domain.

REFERENCES

- Agirre, E., Soroa, A., (2007). Semeval-2007 task 02: Evaluating word sense induction and discrimination systems. *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, 7-12.
- Anaya-Sánchez, H., Pons-Porrata, A., Berlanga-Llavori, R. (2006). Word Sense Disambiguation based on Word Sense Clustering. In J. Simão; H. Coelho and S. Oliveira (Eds.), *Lecture Notes in Artificial Intelligence: Vol. 4140. IBERAMIA-SBIA* (pp. 472-481). Ribeirão Preto, Brazil: Springer.
- Esuli, A., Sebastiani, F. (2006). SentiWN: A Publicly Available Lexical Resource for Opinion Mining. *Proceedings of the Fifth international conference on Language Resources and Evaluation (LREC 2006)*, 417-422.
- Gil-García, R., Badía-Contelles, J. M., Pons-Porrata, A. (2003). Extended Star Clustering Algorithm. In A. Sanfeliu and J. Ruiz-Shulcloper (Eds.), *Lecture Notes in Computer Sciences: Vol. 2905. 8th Iberoamerican Congress on Pattern Recognition (CIARP)* (pp. 480-487). Berlin, Heidelberg: Springer-Verlag.
- Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. (1993). Introduction to WordNet: An Online Lexical Database. *International Journal of Lexicography*, 3(4), 235-244.
- Schmid, H. (1994). Probabilistic Part-of-speech Tagging Using Decision Trees. *Proceeding of the Conference on New Methods in Language Processing*, 44-49.
- Stone, P. J., Dunphy, D. C., Smith, M. S., Ogilvie, D. M. (1966). The General Inquirer: A Computer Approach to Content Analysis. *The American Journal of Sociology*, 73(5), 634-635.
- Strapparava, C., and Mihalcea, R. (2007). SemEval-2007 Task 14: Affective Text. *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval 2007)*, 70-74.