

Fine-Grained Opinion Mining as a Relation Classification Problem

Alexandru Lucian Gînscă

Alexandru Ioan Cuza University
Faculty of Computer Science, Iași
lucian.ginsca@info.uaic.ro

Abstract

The main focus of this paper is to investigate methods for opinion extraction at a more detailed level of granularity, retrieving not only the opinionated portion of text, but also the target of that expressed opinion. We describe a novel approach to fine-grained opinion mining that, after an initial lexicon based processing step, treats the problem of finding the opinion expressed towards an entity as a relation classification task. We detail a classification workflow that combines the initial lexicon based module with a broader classification part that involves two different models, one for relation classification and the other for sentiment polarity shift identification. We provided detailed descriptions of a series of classification experiments in which we use an original proximity based bag-of-words model. We also introduce a new use of syntactic features used together with a tree kernel for both the relation and sentiment polarity shift classification tasks.

1998 ACM Subject Classification I.2.7 Natural Language Processing

Keywords and phrases Opinion Mining, Opinion Target Identification, Syntactic Features

Digital Object Identifier 10.4230/OASICS.ICCSW.2012.56

1 Introduction

Opinion mining is one of the applications of natural language processing with the biggest growth in recent years concerning the number of publications and dedicated conferences as well as industry applications. Most of them refer to opinion mining as a text classification task in which a text fragment is labeled as either positive or negative. In this paper, we focus on a more detailed approach of identifying the opinion expressed towards a certain target in a text fragment.

One of the basic and most used approaches for opinion mining is lexicon-based opinion generation and it has been used for opinion retrieval as a standalone method but there are also many research works that combined both lexicon based and text classification techniques in opinion mining systems [1]. The authors of [11] mention probabilistic models as methods that have also been used to retrieve and classify opinions from documents [3]. The probabilistic approach relies on probabilistic assumptions based on frequency of query terms [4]. Another method described in [11] is the language model approach that has also been used for opinion retrieval [12]. Most language models imply word level processing, sentence level processing and paragraph level processing, but the core of language models is the bag-of-words representation. The state of the art statistical methods are based on the observation that similar opinion words frequently appear together in a corpus, as detailed in [14]. If two words frequently appear together within the same context, they are likely to share the same polarity.



© Alexandru Lucian Gînscă;
licensed under Creative Commons License NC-ND
2012 Imperial College Computing Student Workshop (ICCSW'12).
Editor: Andrew V. Jones; pp. 56–61

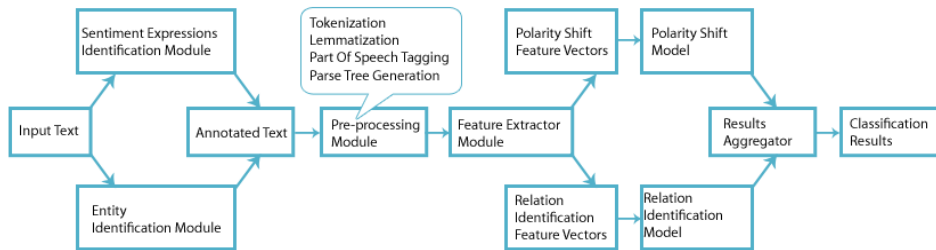


OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

2 Proposed approach

We present in this section a novel fine-grained opinion classification workflow that combines an initial lexicon based step with a broader classification part that involves two different models and we briefly describe a new proximity based bag-of-words model and the use of tree kernels for relation classification and polarity shift identification.

2.1 A novel mixed lexicon/machine learning classification workflow



■ **Figure 1** Classification workflow.

In Figure 1, we present an overview of our proposed classification workflow. The annotated text contains tags for entities and for sentiment bearing expressions. The actual entities and sentiments will be later replaced by an abstract token for the classification model as detailed in the following sections. The pre-processing module deals with the word level and sentence level text processing methods, such as lemmatization, part of speech tagging or generating a parse tree and depends on the features that will be later used in the workflow by the classifiers. The feature extraction module builds feature vectors both for the relation identification classifier as well as for the polarity shift identification one.

One of the key aspects of our approach is that we treat the fine-grained opinion identification problem as a relation identification one that is independent of the entity identification and sentiment extraction modules. This also represents one of the main advantages of the workflow that we propose, the fact that it can be easily adapted for different sentiment identification contexts, not only allowing different types of entities, but also different semantics of the expressed sentiments.

2.2 Relation identification with a novel proximity based bag-of-words model

The bag-of-words model is a common practice in text classification in which a document is represented by a vector of words. The vector is built from a dictionary that gathers all of the words from all the documents in the corpus. Three basic variations of the bag-of-words model can be identified: occurrence, where the values of the vector are 1 if the word appears in the document and 0, otherwise; appearances, where the values of the vector represent the number of times a word appear in that document and tf-idf, where the term frequency-inverse document frequency of the words of that document in respect with the whole corpus is used.

The main motivation for a different type of a bag-of-words model is the intrinsic nature of the classical model that does not take into consideration the position of the words in the sentence. Models that try to solve this problem by using n-grams (usually up to 5-grams) instead of unigrams have the problem of an exponential increase in feature space. This is

why we propose a different type of a bag-of-words model designed specifically for the binary relation identification problem that uses the proximity measured in number of tokens between words.

The model is built as follows: For each word in the dictionary that is found in the sentence, we first compute the number of tokens (words, punctuation) between the word and the SENTIMENT token and then the number of tokens between the word and the TARGET token. If the word appears in the sentence after the SENTIMENT token, the value that is put in the feature vector is the number of tokens between the word and the SENTIMENT token multiplied by -1. The same applies for the case in which the word is situated after the TARGET tag.

2.3 Relation identification with a tree kernel based model

The tree kernel is a function $K(T1, T2)$ that returns a normalized similarity score in the range (0,1) for two trees T1 and T2 [2]. Details regarding the formal definition and in depth descriptions of tree kernels can be found in [16].

For the task of relation identification in the context of fine-grained opinion mining, we used Alessandro Moschitti's implementation of tree kernels that is described in [10] and [9] and is based on the SVM-Light library [5]. The SVM-Light implementation takes as input a parse tree with the binary label, but it also allows a combination of parse trees and numerical feature vectors for which the RBF or polynomial kernels can be used. It also allows the user to explicitly specify the way in which the results from each kernel are combined (addition or multiplication) and what weight is given to each kernel.

2.4 Opinion polarity shift identification with a tree kernel based model

Besides correctly identifying which sentiment bearing expression influences which target in a sentence, we are also interested to find out when a polarity shift for a sentiment expression that influences an entity takes place. A polarity shift is usually associated with negation and it represents the case in which the context changes a positive sentiment expression into a negative one and vice-versa. For the problem of polarity shift identification, we used a similar approach as for the relation identification one. For this task, we consider a positive instance, the case in which a polarity shift does not occur and a negative one, the case in which a polarity shift takes place.

3 Experiments and results

3.1 Evaluation corpus

Although the MPQA [15] corpus has been used in fine-grained opinion mining experiments, such as those presented in [8] and [13], most of them are directed to opinion holder and opinion expression identification and the targets identified in the MPQA corpus are less structured and can vary from named entities to abstract concepts described in a larger text span. For these reasons, we chose the JDPA [6] corpus as an evaluation benchmark for our classification experiments. The creators of the corpus provide details about it in [6].

From the JDPA corpus, we extracted the sentiment expression and their targets. To respect our proposed workflow described in the previous section, we replaced the actual sentiment expressions and targets with abstract tokens, "SENTIMENT" and "TARGET", respectively. Due to the high number of annotated sentiments and entities, we used for our test the "camera" set of files from the JDPA corpus. For the polarity shift identification task,

the extraction of the positive and negative instances is done by using the negation indicators from the JDPA corpus and replacing any sentiment expression with the SENTIMENT token. We replaced the negation identifier with the NEGATION token, whereas for the relation identification task we replaced the target expression with the TARGET token.

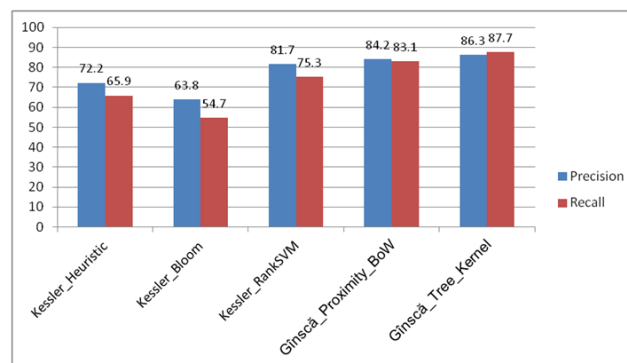
3.2 Relation classification results

In Table 1, we provide an overview of the best result for each method that we described in the previous section. For the tree kernel experiments, T represents the parse tree, V1 represents a one dimensional feature vector consisting of the number of tokens between the SENTIMENT and the TARGET tokens and V2 a two dimensional vector that also contains the number of punctuation marks between the SENTIMENT and the TARGET tokens. As it can be observed, the SVM with the tree kernel together with the two distance features provide the best results for the accuracy, precision and recall.

■ **Table 1** Overview of the best result for each method.

Base Model	Variation	Accuracy	Precision	Recall
Classic Bag-of-Words	Naïve Bayes	79.82	80.2	79.8
	SVM + RBF	76.2	79.8	76.2
	SVM + Poly.	78.4	79.6	78.4
Proximity Bag-of-Words	Naïve Bayes	82.5	84.2	82.6
	SVM + RBF	83.09	83.5	83.1
	SVM + Poly.	78.27	81.6	78.3
SVM + Tree Kernel	T	83.896	83.684	85.488
	T + V1	86.182	86.034	87.534
	T + V2	86.442	86.332	87.708

In our experiments, we used the occurrence bag-of-words model because we dealt with small sentences and the other two types brought little new information for the classifier. For the feature dictionary generation, we used the lemma of the words that appeared in all of the sentences.



■ **Figure 2** Comparison with Kessler's top 3 results.

We compare our best results to those reported by the authors of the JDPA corpus in their 2009 paper [7]. We retained the results from the best 3 methods that they have used: Heuristic, Bloom and Rank SVM.

The results presented in Figure 2 show that our two novel approaches to sentiment target identification, the proximity bag-of-words model and a tree kernel together with a feature vector composed of 2 elements outperform the top 3 approaches presented in [7].

3.3 Opinion polarity shift identification results

Given the fact that the tree kernel experiments provided the best results, we used for the polarity shift identification problem the same classification configurations as in section 4.3, for the relation identification task. In Table 2, we show the accuracy, precision and recall results for the tree kernel polarity shift identification experiments.

■ **Table 2** 10 Fold cross validation results for tree kernel polarity shift identification.

Window Size	Features	Accuracy	Precision	Recall
2	T	84.39	85.94	84.5
	T + V1	87.28	88.45	86.68
	T + V2	87.64	87.25	86.2
1	T	85.67	85.40	85.62
	T + V1	89.8	90.25	88.72
	T + V2	89.4	89.92	88.25
0	T	85.05	87.32	86.48
	T + V1	86.48	87.48	86.95
	T + V2	86.5	87.05	87.25

Because the negation identifier is regularly closer to the sentiment expression than the target is to the sentiment expression and the words before the negation and those after the sentiment expression have less influence on these, we chose to test a window size of maximum 2. The windows size represents the number of tokens before the first appearance and the number of tokens after the last appearance of the SENTIMENT or the TARGET tokens that are taken into consideration for classification from the whole sentence.

4 Conclusion and future work

We described in this paper a novel approach to fine-grained opinion mining that, after an initial step that involves the use of lexical resources, treats the problem of finding the opinion expressed towards an entity as a relation classification task. We detailed our classification workflow, a novel proximity bag-of-words model and we presented how tree kernels can be successfully used for relation classification, as well as for polarity shift identification. We included an overview of the best result obtained when using each method and we showed that both of our two novel approaches to the detection of sentiments expressed towards a certain target outperformed the methods proposed by the authors of the evaluation corpus.

Due to the fact that the best results were obtained when we used a tree kernel together with feature vectors, we plan to investigate the impact of using other features than those presented in this paper. So far, we focused our research on sentiment target identification but the same methods we used for this task can be used for another aspect of fine-grained opinion mining, opinion holder identification. This is a direction worth pursuing.

References

- 1 J. Bollen, A. Pepe, and H. Mao. Modeling public mood and emotion:twitter sentiment and socioeconomic phenomena. *Proceeding of the WWW Conference*, 2010.
- 2 A. Culotta and J. Sorensen. Dependency tree kernels for relation extraction. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 2004.
- 3 S. Gerani, M. Carman, and F. Crestani. Proximity-based opinion retrieval. *Proceeding of the Special Interest Group on Information Retrieval Conference (SIGIR)*, 2010.
- 4 X. Huang and W. Croft. A unified relevance model for opinion retrieval. *Proceedings of the ACM Conference on Information and Knowledge Management*, 2009.
- 5 T. Joachims. Making large-scale svm learning practical. *Advances in Kernel Methods - Support Vector Learning*, 1999.
- 6 J. Kessler, M. Jason, L. Clark, and N. Nicolov. The 2010 icwsm jdpa sentiment corpus for the automotive domain. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media Data Workshop Challenge (ICWSM-DWC 2010)*, 2010.
- 7 J. Kessler and N. Nicolov. Targeting sentiment expressions through supervised ranking of linguistic configurations. *Proceedings of the 3rd Int'l AAAI Conference on Weblogs and Social Media*, 2009.
- 8 S.-M. Kim and E. Hovy. Automatic detection of opinion bearing words and sentences. *Companion Volume to the Proceedings of IJCNLP-05, the Second International Joint Conference on Natural Language Processing*, pages 61–66, 2005.
- 9 A. Moschitti. A study on convolution kernels for shallow semantic parsing. *Proceedings of the 42-th Conference on Association for Computational Linguistic (ACL-2004)*, 2004.
- 10 A. Moschitti. Making tree kernels practical for natural language learning. *EACL 2006, 11st Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, 2006.
- 11 S. O. Orimaye. Sentence-level contextual opinion retrieval. *Proceedings of the 20th international conference companion on World Wide Web*, 2011.
- 12 B. Pang and L. Lee. Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135, 2008.
- 13 V. Stoyanov and C. Cardie. Topic identification for fine-grained opinion analysis. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING)*, pages 817–824, 2008.
- 14 M. Tsytsarau and T. Palpanas. Survey on mining subjective data on theweb. *Data Mining and Knowledge Discovery Special Issue*, 2011.
- 15 J. Wiebe and C. Cardie. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165–210, 2005.
- 16 D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. *Journal of Machine Learning Research*, pages 1083–1106, 2003.