MORE SENSE: MOvie REviews SENtiment analysis boosted with SEmantics

Amna Dridi and Diego Reforgiato Recupero

University of Cagliari, Mathematics and Computer Science Department, Via Ospedale 72, 09124, Cagliari, Italy
{amna, diego.reforgiato}@unica.it

Abstract. Sentiment analysis is becoming one of the most active area in Natural Language Processing nowadays. Its importance coincides with the growth of social media and the open space they create for expressing opinions and emotions via reviews, forum discussions, microblogs, Twitter and social networks. Most of the existing approaches on sentiment analysis rely mainly on the presence of affect words that explicitly reflect sentiment. However, these approaches are semantically weak, that is, they do not take into account the semantics of words when detecting their sentiment in text. Only recently a few approaches (e.g. sentic computing) started investigating towards this direction. Following this trend, this paper investigates the role of semantics in sentiment analysis of movie reviews. To this end, frame semantics and lexical resources such as BabelNet are employed to extract semantic features from movie reviews that lead to more accurate sentiment analysis models. Experiments are conducted with different types of semantic information by assessing their impact in movie reviews dataset. A 10-fold cross-validation shows that F1 measure increases slightly when using semantics in sentiment analysis in social media. Results show that the proposed approach considering word's semantics for sentiment analysis is a promising direction.

Keywords: Sentiment analysis, Movie reviews, Frame semantics, BabelNet

1 Introduction

In the movie domain, there has been a huge number of review sites giving opinions about the performance of the movie. While the stars rating to a movie tell us about the success or failure of a movie quantitatively, reviews give us a deeper qualitative insight on different aspects of the movie.

Sentiment analysis of movie reviews aims to automatically infer the opinion of the reviewer with respect to various topics or the overall polarity of the review. Most of existing work on sentiment analysis of movie reviews use statistical methods to extract features from the review (bag of words, n-grams, word2vec,etc.) [10]. However, statistical methods suffer from the lack of handling meanings and semantics which are crucial for understanding reviewer's sentiment. Therefore, there is the need to shift from a word-level to a conceptual-level analysis of sentiments. This intuition has been the basis of a novel, multi-disciplinary approach to sentiment analysis, called *sentic computing*[1], which aims to include *semantic features* into sentiment analysis.

Following this trend, some works emerged and explored a new type of features called *semantic features* aiming to handle meaning and semantics. While they intend to extract semantics, the proposed approaches suffer from the lack of handling anaphora resolution and word sense disambiguation.

To overcome this problem, we propose, in this paper, a supervised approach for sentiment analysis using frame semantics and lexical resources such as BabelNet[2] [8] to extract semantic features from reviews. We experiment and evaluate our proposed approach with a movie reviews dataset. We perform a semantic incorporation through replacement and augmentation into Naïve Bayes (NB) model training. Our results show that combining our semantic features with unigrams slightly outperforms the baseline model trained from unigrams only.

2 Related Work

Several approaches have been proposed to solve polarity detection problem within movie reviews ranging from *supervised approaches* [10], *unsupervised* approaches [6, 7, 12, 14] to *hybrid* ones [5].

The *supervised approaches* use a wide range of features and labeled data for training sentiment classifiers. For instance, Pouransari and Ghili [10] applied the bag of by supervised ones. Hence, few hybrid approaches, that combine both supervised and unsupervised methods, emerged. For instance, Maas et al. [5] proposed a model that uses a mix of unsupervised and supervised techniques to learn word vectors capturing semantic term-document information as well as rich sentiment content in movie reviews domain.

All works mentioned above concentrate on the use of two types of features; n-grams features and lexicon-based features, for sentiment analysis. However, it has been argued that sentiment in text is not always associated with individual words, but instead, through relations and dependencies between words, which often formulate sentiment [13]. Therefore, a new type of features for sentiment analysis has been explored called *semantic features* aiming to handle meanings and semantics which are crucial for understanding sentiment. In this direction, some works start to emerge. Maas et al. [5], for instance, presented a model to capture both semantic and sentiment similarities among words. The semantic component of their model learns word vectors via an unsupervised probabilistic model of documents. In the same context, Mukherjee et al. [7] proposed to incorporate the word knowledge through Wikipedia to retrieve relevant opinionated text. To this end, the authors suggested a weakly supervised approach to sentiment classification of movie reviews. The weak supervision comes from the usage of resources like WordNet, POS-Tagger and sentiment lexicons.

[1] `http://sentic.net/`
[2] `http://babelnet.org/`

While they intend to extract semantics, all the proposed approaches suffer from the lack of handling anaphora resolution and word sense disambiguation. Usage of simple lexicon at the final stage for polarity detection also decreases its accuracy. To overcome this problem, Recupero et al. [4, 11] developed *Sentilo* an unsupervised domain-independent system based on sentic computing [2] and performs sentiment analysis by hybridizing natural language processing techniques and semantic web technologies. Following this trend, our work tends to be placed where both semantic frames and lexical resources such as BabelNet will be employed to extract semantic features from microblogs that lead to more accurate sentiment analysis models.

3 Semantic Sentiment Analysis of Movie Reviews

3.1 Data Description

We use a long text corpora refers to movie reviews. The movie reviews dataset is a subset of the IMDB movie reviews corpus [9] which is publicly available on Kaggle[3]. The labeled dataset is selected for sentiment analysis task. The sentiment of reviews is binary, meaning the IMDB rating < 5 results in a sentiment score of 0 (negative), and rating $>= 7$ have a sentiment score of 1 (positive). The binarity is given with the reviews in term of labels; *"positive"* and *"negative"*. No individual movie has more than 30 reviews. The selected subset consists of 3750 reviews collected from the IMBD movie review site and polarity labeled at the document level, 1820 for positive class and 1920 for negative class. This forms our gold standard dataset for movie reviews sentiment analysis.

3.2 Data Preprocessing

For our dataset, we performed a preprocessing step to clean up the data. Data preprocessing consists of three steps:

1. tokenization; we segment text by splitting it by spaces and punctuation marks, and form a bag of words.
2. removal of unnecessary punctuation like '!', '?', *etc.* as they do not provide any substantial information;
3. removal of HTML tags like '
' by using simple regular expressions matching to remove these HTML tags from the text.
4. removing stop words from the text using Stanford NLP stop word list[4].

[3] `https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset`
[4] `https://github.com/stanfordnlp/CoreNLP/blob/master/data/edu/stanford/nlp/patterns/surface/stopwords.txt`

3.3 Feature Extraction

We use a variety of features for our classification experiments. For the baseline, we use *n-grams features*. However, in our approach, we focus on identifying new sets of features to be added to the trained model for sentiment classification. Therefore, we investigate a novel set of features derived from word's semantics, expressed in term of *semantic frames* and lexical resources such as *BabelNet*. To this end, we leveraged *Framester*[5] a wide coverage hub of linguistic linked data standardized using *frame semantics* [3].

n-gram features *n-gram model* is a typical way to numerically represent texts. To identify a set of useful n-grams, we first remove stop words. Then, we tokenize text while assuring that short forms such as "*don't*", "*I'll*", "*she'd*", ... will remain as one word. Afterward, we calculate the total word counts for each word across all reviews. As the total number of words in the dictionary was huge (more than 160.000), we took only the 50.000 most frequent words according to their occurrence. That ensured that we remove most of the misspelled words. Also, words which occurred only once in the dataset would contribute nothing to the classifier.

TF.IDF features While *n-gram model* concentrates more on higher frequency parts of the review, it completely ignores the portions which might be less frequent but have more significance for the overall polarity of the review. To overcome this shortcoming of *n-gram model*, we feature representation of words using TF.IDF. The feature representation for this model is similar to the unigram model except that we use TF.IDF values for each word instead of their frequency counts.

Semantic features The semantic features that we extracted correspond to the *semantic frames* and the *BabelNet synsets* returned by Framester for each movie review.

- *BabelNet synsets* are sets of synonyms in different languages grouped by *BabelNet* which is an encyclopedic dictionary that provides concepts and named entities lexicalized in many languages and connected with large amounts of semantic relations, automatically created by linking Wikipedia[6] to WordNet[7] [8].
- *Semantic Frames* are a collection of facts that specify *"characteristic features, attributes, and functions of a denotatum, and its characteristic interactions with things necessarily or typically associated with it [1]."*

We propose two different methods to incorporate semantic features into the classifier.

[5] http://lipn.univ-paris13.fr/framester/en/wfd_html/
[6] http://www.wikipedia.org/
[7] http://wordnetweb.princeton.edu/perl/webwn

– **Semantic replacement:** In this method, we replace all n-grams in movie reviews with their corresponding BabelNet synsets or semantic frames.

– **Semantic augmentation:** This method augments the original n-grams feature space with the semantic features as additional features for the classifier training in three different ways: i) augment the original n-grams with semantic frames, ii) augment with BabelNet synsets, and iii) augment with both semantic frames and BabelNet synsets . The size of the vocabulary in this case is enlarged by the introduced semantic features.

3.4 Binary Classification Polarity

The overall task in this paper is for binary classification polarity of movie reviews as negative or positive. Therefore, for this task we build a sentiment classifier using *Naive Bayes* (NB) method.

To implement this BN classifier, we use the following standard bag-of-features model. Let $\{f_1, \ldots, f_k\}$ be a predefined set of k features that can appear in a movie review. Each feature f_i could be expressed in term of frequency or TF.IDF. Let $w_i(r)$ be the function representing how f_i occurs in the movie review r. Then, each review is represented by the following review vector: $\overrightarrow{r} := (w_1(r), w_2(r), \ldots, w_k(r))$.

Our task of polarity detection is to assign to a given review m the sentiment $s^* = argmax_s P(s \setminus r)$. We derive NB classifier by first observing that by *Bayes' rule*:

$$P(s \setminus r) = \frac{P(s)P(r \setminus s)}{P(r)}$$

where $P(r)$ plays no role in selecting s^*. To estimate the term $P(s \setminus r)$, Naive Bayes decomposes it by assuming the f_i's are conditionally independent given r's class:

$$P_{NB}(s \setminus r) = \frac{P(s)(\prod_{i=1}^{k} P(f_i \setminus s)^{n_i(r)})}{P(r)}$$

We implement our BN classifier in *JAVA* using an open source code[8].

4 Evaluation Results

In this section, we evaluate the integration of semantic features on binary polarity detection and present the obtained results on our *movie reviews dataset* and their derivations that led to 7 datasets in total. We then compare these results with those obtained using unigram features (the baseline features).

Our goal for these experiments is then two-fold. First, we aim to evaluate whether our training data with labels derived from frame semantics and lexical resource BabelNet is useful for training sentiment classifiers for social media.

[8] `https://github.com/ptnplanet/Java-Naive-Bayes-Classifier/tree/master/src/main/java/de/daslaboratorium/machinelearning/classifier`

Second, we want to evaluate the effectiveness of the semantic features for sentiment analysis in user-generated data. How useful are the semantic features on movie reviews texts? How much gain do we get from these features?

For our set of experiments, we use NB trained from word unigrams as a baseline model, for each derived dataset. Then, we incorporate the semantic features into NB by either replacing the original bag-f-words feature space or augmenting into it. For each dataset, we train two Bayes classifiers, which use different features: *unigram features* and *TF.IDF features*.

For each dataset, we perform a 10-*cross validation* and report results averaged over 10 runs using $F1$ measure.

Table 1 shows $F1$ measures with bag-of-word (BOW) model (*unigram model*) and TF.IDF model, of our sentiment classification using unigrams as baseline and the other combinations of semantic features with replacement (*i) BabelNet synsets (BNS), (ii) semantic frames (SF), (iii) BabelNet synsets and semantic frames (BNS+SF)* and augmentation (*i) unigrams and BabelNet synsets, ii) unigrams and semantic frames, iii) unigrams, BabelNet synsets and semantic frames*).

Features	TF.IDF	BOW
unigrams	79.56	79.97
BabelNet synsets	79.27	**80.64**
Semantic Frames	78.97	78.30
BNS+SF	79.48	79.81
unigrams+BNS	79.57	79.69
unigrams+SF	79.57	79.79
unigrams+BNS+SF	**79.65**	79.65

Table 1. 10-cross validation results of IMDB dataset and their representations using several combinations of semantic features.

According to the results shown in Table 1, the incorporation of semantic features slightly outperforms the unigrams baselines in the two training model. The gain is small and that something we expected as this dataset is of a specific domain and the extracted semantic features are not able to well represent each object.

5 Conclusion

In this paper, we presented a supervised approach for sentiment polarity detection in movie reviews using semantic features. We defined semantic features in term of semantic frames and BabelNet synsets extracted with the linguistic linked data *Framester*. We explored two different approaches for incorporating them into sentiment analysis; with replacement and augmentation. Then, we trained a sentiment classifier that is able to determine positive and negative sentiments of reviews. The classifier is based on the multinomial Naive Bayes classifier that uses n-grams and TF.IDF models.

As future work, we plan to model the problem of sentiment analysis in social media as multi-class classification problem where we classify the sentiment in more than binary classes like *"Happy"*, *"Bored"*, *"Afraid"*, etc. Furthermore, we plan to remodel this problem as a regression problem where we can predict the degree of affinity for the review instead of a simple negative/positive class. As well as, we plan to perform more experiments on different type of user-generated data other than movie reviews, namely tweets, and blogs.

Acknowledgments

This work has been supported by Sardinia Regional Government (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2014-2020 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

References

1. Alan, K: Natural Language Semantics, Blackwell Publishers Ltd, Oxford, ISBN 0-631-19296-4, pp.251, 2001.
2. Cambria, E., Grassi, M., Hussain, A.and Havasi, C.: Sentic computing for social media marketing. Journal of Multimed Tools Appl, 59(2), pp.55777, 2012.
3. Gangemi, A., Alam, M., Asprino, L., Presutti, V.and Recupero, D.R.: Framester: A Wide Coverage Linguistic Linked Data Hub. In Proceedings of EKAW 2016, pp. 239-254, 2016.
4. Gangemi, A., Presutti, V. and Recupero, D.R.: Frame-Based Detection of Opinion Holders and Topics: A Model and a Tool. Journal IEEE Comp. Int. Mag, 9(1), pp.20-30, 2014.
5. Maas, A.L., and Daly, R.E., Peter, P. T., Huang, D., Y. Ng, A. and Potts, Ch.: Learning Word Vectors for Sentiment Analysis. In proceeding of ACL. pp. 142-150, 2011.
6. Momtazi, S.: Fine-grained German Sentiment Analysis on Social Media. In Proceedings of the Eight International Conference on Language Resources and Evaluation, LREC'12, pp. 23-25, 2012.
7. Mukherjee, S. and Bhattacharyya, P.: WikiSent: Weakly Supervised Sentiment Analysis through Extractive Summarization with Wikipedia. In proceedings of ECML/PKDD 2012, pp.774-793, 2012.
8. Navigli,, R.and Ponzetto, S.P.: BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Journal of Artificial Intelligence 193 (2012), pp. 217250, 2012.
9. Pang, B., Lee, L. and Vaithyanathan, S.: Thumbs up? Sentiment Classification using Machine Learning Techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-86.Association for Computational Linguistics, 2002.

10. Pouransari, H. and Ghili, S.: Deep Learning for Sentiment Analysis of Movie Reviews. From https://cs224d.stanford.edu/reports/PouransariHadi.pdf, 2014.
11. Recupero, D.R., Presutti, V., Consoli, S., Gangemi, A., Giovanni, A. and Nuzzolese, G.: Sentilo: Frame-Based Sentiment Analysis. Journal of Cognitive Computation, Volume 7, number 2, pp. 211-225, 2015.
12. Rothfels,J. and Tibshirani, J.: Unsupervised sentiment classification of English movie reviews using automatic selection of positive and negative sentiment items. Technical report, Stanford University, 2010.
13. Saif, H., M., He, Fernández, Y. and Alani, H.: Contextual semantics for sentiment analysis of Twitter. In proceedings of Journal of Information Processing and Management, volume 52(1), pp.5-19, 2016.
14. Zagibalov, T. and Carroll, J.: Automatic seed word selection for unsupervised sentiment classification of Chinese text. In Proceedings of the 22nd International Conference on Computational Linguistics, Volume 1, pp. 1073-1080. Association for Computational Linguistics, 2008.