Further Experiments in Sentiment Analysis of French Movie Reviews

Hatem Ghorbel and David Jacot

Abstract In sentiment analysis of reviews we focus on classifying the polarity (positive, negative) of conveyed opinions from the perspective of textual evidence. Most of the work in the field has been intensively applied on the English language and only few experiments have explored other languages. In this paper, we present a supervised classification of French movie reviews where sentiment analysis is based on some shallow linguistic features such as POS tagging, chunking and simple negation forms. In order to improve classification, we extracted word semantic orientation from the lexical resource SentiWordNet. Since SentiWordNet is an English resource, we apply a word-translation from French to English before polarity extraction. Our approach is evaluated on French movie reviews, obtained results showed that shallow linguistic features has significantly improved the classification performance with respect to the bag of words baseline.

Key words: Sentiment analysis, Opinion Mining, Polarity classification, Supervised learning, Linguistic features.

1 Introduction

Sentiment analysis is an emerging discipline whose goal is to analyze textual content from the perspective of the opinions and viewpoints they hold. A large number of studies have focused on the task of defining the polarity of a document which is

Hatem Ghorbel
University of Applied Sciences Western Switzerland, Haute Ecole Arc Ingénierie, St-Imier, Switzerland e-mail: hatem.ghorbel@he-arc.ch

David Jacot
University of Applied Sciences Western Switzerland, Haute Ecole Arc Ingénierie, St-Imier, Switzerland e-mail: david.jacot@master.hes-so.ch

by far considered as a classification problem: decide to which class a document is attributed; class of positive or negative polarity [HR10, WK09, PLV02].

Most of the work in the field has been intensively applied on the English language [Tur02, WK10]. For this purpose, English corpora and resources (such as MPQL [WWH05], Movie Review Data [PLV02], SentiWordNet [ES06] and WordNet-Affect [SV04]) have been constructed to aid in the process of automatic supervised and unsupervised polarity classification of textual data. Nevertheless, still very few experiments are applied on other languages.

In this context, we address in this paper the issue of polarity classification applied on French movie reviews. We used a supervised learning approach where we trained the classifier on annotated data of French movie reviews extracted from the web. As classification features, beyond the word unigrams feature taken as the baseline in our experiments, we extracted further linguistic features including lemmatized unigrams, POS tags, simple negation forms and semantic orientation of selected POS tags. The latter is extracted from the English lexical resource SentiWordNet after applying a word-translation from the French to English.

The main goal of our experiments is firstly to confirm that the incorporation of shallow linguistic features into the polarity classification task could significantly improve the results. Secondly, to address the problem of loss of precision in defining the semantic orientation of word unigrams from English lexical resources, mainly due to the intermediate process of word-translation from French to English correlated with further issues such as sense disambiguation.

In the rest of the paper, we first shortly describe the previous work in the field of sentiment analysis and polarity classification. Then we describe the set of extracted features used in polarity classification of French movie reviews. Finally we provide and discuss the obtained experiment results and end up by drawing some conclusions and ideas for future work.

2 Previous Work

Classical approaches in text retrieval and categorization has so far focused on mining and analyzing factual information such as entities, events and their properties. They basically utilize Natural Language Processing methods and techniques in order to extract objective features aiding the classification and categorization of textual expressions [PLV02] with special emphasis on linguistic features in order to increase the performance. As linguistic features, [Gam04, MTO05] present syntactically motivated features, most of them based on dependency path information and modeled as high n-grams. Further linguistic features such as part of speech, negation, verbs modality, and semantic information (from WordNet for instance) are recently explored [WK09, TNKS09].

Much of the previous work focuses on defining the characteristics of conveyed opinions on the basis of textual data with processing granularity ranging from words, to expressions, sentences and documents. For this purpose, statistical approaches have been coupled with semantic approaches in order to detect opinions and sentiments in textual spans with different compositional structure [KH04, WWH05]. Semantic approaches aim at classifying sentiment polarity conveyed by textual data using commonsense, sentiment resources, as well as linguistic information. For instance, [HL04, ES05, NSS07, Den08] classify polarity using emotion words and semantic relations from WordNet, WordNet Gloss, WordNet-Affect and SentiWordNet respectively.

An important theoretical issue in the semantic approach is still how to define the semantic orientation of a word in its context. Some studies showed that restricting features to those adjectives would improve performance. [HM97] have focused on defining the polarity of adjectives using indirect information collected from a large corpus. However, more researches showed that most of the adjectives and adverbs, a small group of nouns and verbs possess semantic orientation [ES05, MTO05, TL03].

Only very few work [Den08, ACS08] have explored sentiment analysis in a multilingual framework such as Arabic, Chinese, English, German and Japanese. Their methodology is based on standard translation from target language to English in order to reuse existing English corpora and resources for polarity classification.

3 Feature Design

We have defined three categories of features: lexical, morpho-syntactic and semantic features. Lexical and morpho-syntactic features have been formulated at the word level, whereas semantic features have been formulated at the review level[1].

3.1 Lexical features

This is the baseline of our experiments and is mainly composed of word unigrams. The global assumption in this choice is that we tend to find certain words in positive reviews and others in negative ones. Each unigram feature formulates a binary value indicating the presence or the absence of the corresponding word at the review level.

Lemmatization is argued to be relevant in sentiment analysis in order to group all inflected forms of a word in a single term feature, especially French is a inflected language. For example the words *aimé*, *aimait* and *aimer* share the same polarity but will be considered as five separate features during the classification. When applying lemmatization, we would obtain a unique feature. Features reduction would improve the tuning of the training process.

[1] In supervised learning, a training and test corpus is first annotated. A learning algorithm is then applied to induce the training model on the basis of the selected features. As it is shown in [ES05, PLV02] for instance, probabilistic algorithms (Bayes, maximum entropy) and linear discrimination (Support Vector Machine) are the most appropriate for the task of polarity classification.

3.2 Morpho-syntactic features

Some studies showed that restricting features to specific part-of-speech (POS) categories, for instance adjectives would improve performance [HM97]. In our approach, POS tags are proposed to be used to enrich unigrams features with morpho-syntactic information so as to disambiguate words that share the same spelling but not the same polarity. For example, it would distinguish the different usages of the word *négatif* that can either be a neutral noun *un négatif* or a negative adjective *un commentaire négatif*. Moreover POS tags are useful to handle negation and to aid word sense disambiguation before polarity extraction in SentiWordNet as it will be detailed hereafter.

Negation is handled at the shallow level of morpho-syntactic constituency of sentences avoiding the heavy processing of its deep syntactic structure. The detection of negated forms is performed by searching specific patterns formed from the abundantly utilized lexicalized forms of negation combined with particular n-grams of POS categories. We defined two simple patterns that cope with the negation form (1) at the verb level for example *le scénario* ***ne*** *brille* ***pas*** and (2) at the adjective and noun level for example ***sans*** *histoire originale*.

The scope of the negation is fixed with respect to a predefined context window of *n* POS categories within a textual span limited by a punctuation sign. We invert the polarity of the *n* verbs, nouns and adjectives within the context of each detected negation. We do not cope with other composed forms of negation such as conditional, double negation, the counter factual subordinates and modalities. The entailments of such a polarity inversion are first situated at the lexical level; unigrams features are inverted during features vector construction that is if we consider the previous example, in stead of having in the feature *original*, we would have a separate feature *!original* in the vector; second at the semantic level, polarity is inverted from positive to negative and vice-versa in the calculation of the overall polarity of a review as we will detail in the following section.

3.3 Semantic features

As it is shown is previous work [HL04, ES05, NSS07, Den08], the incorporation of corpus and dictionary based resources such as WordNetAffect, SentiWordNet and Whissell's Dictionary of Affect Language contributes in improving the sentiment classification. Based on such results, we use the lexical resource SentiWordNet[2] to extract word polarity and calculate the overall polarity score of the review for each POS tag. SentiWordNet is a corpus-based lexical ressource constructed from the perspective of WordNet. It focuses on describing sentiment attributes of lexical entries describe by their POS tag and assigns to each synset of WordNet three sentiment scores: positivity, negativity and objectivity.

---

[2] SentiWordNet 1.0.1

Since SentiWordNet describes English lexical resources, we go through a word-translation from French to English before polarity extraction. Words are lemmatized before being passed through the bilingual dictionary. We use POS information as well as the most frequently[3] used sense selection to disambiguate senses and predict the right synset. We only considered the positivity and the negativity features for the four POS tags noun, adjective, verb and adverb for this task.

More specifically, we added for each review and for each POS tag two features holding the scores of negativity and positivity as extracted from SentiWordNet. These two scores are calculated as the sum of polarities over all the words of the review respecting POS categorization. For example, for a given review, we obtain the following semantic features vector *(neg_adv, 6.38); (pos_adv, 1.25); (neg_noun, 0.12); (pos_noun, 0.50); (neg_adj, 0.62); (pos_adj, 0.12); (neg_verb, 0.12); (pos_verb, 0.38).*

4 Experiments

Since we didn't find any available sets of annotated data (already classified as negative or positive) of French movie reviews, we collected our own data from the web[4]. We extracted a corpus of 2000 French movie reviews, 1000 positive and 1000 negative, from 10 movies, 1600 were used for training and 400 for testing. We included reviews having a size between 500 and 1000 characters.

Prior classification of the corpus is elaborated according to user scoring: positive reviews are marked between 2.5 and 4 whereas negative reviews are marked between 0 and 1.5[5]. This prior classification is based on the assumption that the scoring is correlated to the sentiment of the review.

For our experiments, the data was preprocessed with the TreeTagger[Sch94], a French POS tagger and lemmatization tool. We applied Support Vector Machine (SVM) classification method and utilized SVM^{Light} [Joa98] classification tool with its standard configuration (inductive classification using linear kernel function)[6] to implement a series of experiments where each time we define a set of combined features and evaluate the accuracy of the approach. The simple validation method (data set division into traning and test corpora) has been applied to evaluate the approach.

[3] This choice is based on the assumption that reviewers spontaneously use an everyday language.

[4] We extracted spectators reviews from `http://www.allocine.com`

[5] Scores are bounded between 0 (for very bad) and 4 (excellent) with a step of 0.5. Reviews scored with 2 are not considered in the construction of our corpus since it is hard to manually classify them as positive or negative opinions.

[6] SVMLight software and detailed descriptions of all its parameters are available at http://svmlight.joachims.org.

Table 1 Performance of most relevant feature sets.

	Features	# of features	Results [%] Pos.	Neg.	Global
(1)	Unigrams	14635	92.00	91.00	91.50
(2)	Unigrams + lemmatization	10624	92.00	93.00	92.50
(3)	**Unigrams + lemmatization + negation**	**12002**	**92.50**	**94.00**	**93.25**
(4)	Unigrams + lemmtization. + POS	12229	93.00	92.50	92.75
(5)	Unigrams + lemmatization + POS + negation	13625	92.50	93.50	93.00
(6)	Unigrams + lemmatization + POS (ADJ)	2109	79.50	92.00	85.75
(7)	Unigrams + lemmatization + POS (ADJ) + negation	2492	80.00	91.00	85.50
(8)	**Unigrams + lemmatization + polarity**	**10632**	**93.00**	**93.50**	**93.25**
(9)	Unigrams + lemmatization + negation + polarity	12010	93.00	92.50	92.75

4.1 Results and Discussion

The results of the following experiments are summarized in Table 1 above. For each experiment labeled from (1) to (9), we present the number of used features and the accuracy measured on the test corpus.

4.1.1 Lexical features

Similarly to [PLV02] we encoded all words features as binary values indicating the presence or the absence of a word in a review. As a first step, we included the entire set of words without applying any specific filtration method.

The accuracy in experiment (1) using the entire set of words is found 91.50%; when comparing this result to Pang et al. [PLV02] who reported an accuracy of 82.90% on English movie reviews using similar features, we find that our results are approximately 10% higher. We believe that this gap is due to the nature of our corpus and the size of our reviews (the collected French reviews are shorter). Moreover, the incorporation of the lemmatization process (2) increases the accuracy by 1.00% up to 92.50%. This was quite expected since French is an inflected language. In experiment (3) we find that negation, although it is processed in a simple form, improves the results to reach 93.25%. Moreover, we notice that the classification of the negative reviews is being improved by the negation processing (from 93% up to 94%) which noticeably means that negation is relatively efficient at this lexical level.

After looking deeply through the reviews, we found that misclassification is mainly due to the following difficulties.

Misspellings Misspelled words are not standard unigrams and hence could not regularly be present in the training data. Reviews containing a large number of misspellings would have their features significantly reduced and so provide very poor information for the classification. We noted that isolated and common misspellings don't affect much the classification but reviews which contain relatively many misspellings tend to be misclassified. Sometimes misspellings are hard to be automatically corrected, especially those made voluntary in order to express a kind of stress and emphasis such as *énnnnorme*. The problem with such kind of words is that they are irregular in the corpus. For example, *énnnnorme* is highly positive but it is not present in the feature set so it is not useful. Quite misspelled reviews tend to be misclassified.

Neutral and mixed reviews Reviews manually interpreted as neutral such as *le film est visuellement réussi mais le scénario est d'une banalité affligente* are randomly classified according to the dominant sentiment of contained words. As a matter of fact, reviewers tend to argue their opinion by posting simultaneously positive and negative arguments organized in a concession or a contrast rhetorical form. Lexical classification shows its limits when the abundant polarity of text spans is not coherent to the final retained opinion, typically the case of a reviewer who starts by verbosely listing the film drawbacks and ends by confessing his admiration and concisely posting his favorable judgment. A further difficulty concerns ironic expressions such as *trop fort les gars* that has a negative polarity although it is composed of positive words. In addition, the classical issue of idiomatic expressions, proverbs and sayings could in some cases have a polarity that doesn't follow the polarity of its composed words, and hence affect negatively the classification.

4.1.2 Morpho-syntactic features

In further experiments, we appended POS tags to every lemmatized unigram so as to disambiguate same unigrams having different syntactic roles. However, the effect of this information seems to be not quite relevant, as depicted on line (4) of Table 1, the accuracy is only increased by approximately 0.25% up to 92.75%. When applying the negation processing (5) to the same experiment, results were slightly improved (up to 93.00) but still not higher than experiment (3) where no POS tags were used. This entails that ambiguity at the morpho-syntactic level of the reviews does not much effect on the polarity classification. Thus, we eliminate this feature from our next experiments.

When restricting unigrams features to only adjectives (6), the performance is getting worse; accuracy is decreased by 6.75% down to 85.75% comparing to (2) and the feature set is reduced by approximately 80%. In order to understand such inconsistency, we look deeper at the accuracy of positive and negative reviews separately. On a one hand, we notice that negative reviews are better classified than positive ones. On the other hand, we have found, in additional experiments, that negative reviews contain relatively an important number of positive adjectives (generally in the negative form). In the first experiment (6) and before processing the negation, these positive adjectives are assumed to negative features in the training model, which induces a further difficulty when classifying positive reviews containing these positive adjectives. However, in the second experiment (7), even after negation processing, the results didn't improve which obviously entails that the scope of processed negation didn't capture the adjectives and was mostly local to the verbal phrase. This last experiment is in contradiction with the results of [HM97] but confirms the results of [PLV02].

As we have already described the negation processing in the previous section, the most used form of negation is that detected at the verbal phrase level. Since deep syntactic dependency analysis of reviews is a quite costly task, it is difficult in this case to capture the adjectives related to the negated verb. Heuristic rules discussed previously such defining a context of a bag of words after the negated verb is not likely to give satisfactory results.

4.1.3 Semantic features

A part from the lexical and the POS features, we extend in our experiments the features set to words polarity extracted from SentiWordNet and formulated as a score representing the overall negativity and positivity of words in the reviews. As shown on the table 1 experience (8), results are improved by 1.75% up to 93.25% compared to lemmatized unigrams experiment (2). The main reason of such an barely perceptible improvement is the failure of extracting polarity information of words from SentiWordNet: among 2000 adjectives, we got the polarity information of only 800 entries in SentiWordNet (40% of success). This extraction problem is mainly due to the following problems.

Translation errors We translate words from French to English so as to cope with SentiWordNet interface. However, the quality of translation significantly affects the results of semantic polarity extraction; this is mainly due to the following reasons.

- The bilingual translator doesn't preserve the POS of words. For example, the *noun méchant* is translated into *wicked* which is implicitly an *adjective* and not a *noun*. Since the translator does not reveal information about the POS change after translation, *wicked* is assumed to be a *noun*. However, the *noun wicked* doesn't exist in SentiWordNet.
- Moreover, even if the translation is correct, it happens that the parallel words do not share the same semantic orientation across both languages due to a difference in common usage, for instance the French *positive* adjective *féériques* is translated into the *negative* English adjective *magical*

Lemmatization and POS tagging errors Misspellings are not standard unigrams and hence could not be found in SentiWordNet. Reviews containing a large number of misspellings would have their overall polarity uncorrect. In addition, misspellings and other lexical errors (for example punctuation, use of parenthesis *permanente(c'est* and composed words *as-tu-vu*) could significantly affect the results of lemmatization and POS tagging tasks elaborated by TreeTagger. In fact, TreeTagger is not implemented to cope with everyday French language as found in spontaneous movie reviews.

Negation As shown in the last experiment (9), the integration of negation processing didn't improve the results. For reminder, the negated verbs, nouns and adjectives would have their extracted polarity score from SentiWordNet inverted. We explain such an outcome by two reasons (i) negation didn't capture properly adjectives (considered as the most subjective lexicon) (ii) bilingual translation of subjective lexicon was not very precise.

5 Conclusions

In this paper, a supervised approach to sentiment analysis of French movie reviews in a bilingual framework was described. The simple validation method (data set division into traning and test corpora) has been applied to evaluate the approach. Preliminary results have shown that the combination of lexical, morpho-syntactic and semantic features achieves relatively good performance in classifying French movie reviews according to their sentiment polarity (positive, negative). Several problems having an effect upon the results of the classification were highlighted and potential solutions were discussed.

In order to extract the semantic orientation of words from SentiWordNet, we went through a standard word-translation process. Although translation does not necessary preserve the semantic orientation of words due to the variation of language common usage especially when it comes to spontaneous reviews on the web, and in spite of all its side effects, it has been argued that dictionary-based approach could contribute to achieve better results. Even if our first experiments showed little significance, further improvements have been proposed accordingly, particularly concerning negation processing.

In future work, the method will be analyzed within a larger training and test sets and evaluated using more sound methods such us cross validation. Further shallow linguistic analysis will be elaborated such as misspelling correction, more elaborated negation processing, WSD and elimination of out of scope text spans from reviews, in addition to the improvement of the translation task using French-English EuroWordNet.

References

[ACS08] A. Abbasi, H. Chen, and A. Salem. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 26:No. 3, Article 12, June 2008.

[Den08] K. Denecke. Using sentiwordnet for multilingual sentiment analysis. *In proceedings of the IEEE International Conference on Data Engineering (ICDE2008)*, pages 507–512, 2008.

[ES05] A. Esuli and F. Sebastiani. Determining the semantic orientation of terms through gloss classification. *In Proceedings of CIKM05*, pages 617–624, 2005.

[ES06] A. Esuli and F. Sebastiani. Sentiwordnet: a publicly available lexical resource for opinion mining. *In Proceedings of the 5th Conference on Language Resources and Evaluation LREC*, 6, 2006.

[Gam04] M. Gamon. Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis. *In Proceedings of the 20th International Conference on Computational Linguistics*, pages 611–617, August 2004.

[HL04] M. Hu and B. Liu. Mining and summarizing customer reviews. *In Proceedings of Knowledge Discovery and Data Mining (KDD'04)*, 2004.

[HM97] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. *In Proceedings of the 8th conference on European Chapter of the Association for Computational Linguistics*, pages 174–181, 1997.

[HR10] A. Hassan and D. Radev. Identifying text polarity using random walks. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 395–403, 2010.

[Joa98] T. Joachims. Making large-scale svm learning practical. *ACM Transactions on Information Systems (TOIS)*, 1998.

[KH04] S.-M. Kim and E. Hovy. Determining the sentiment of opinions. *In Proceedings of the 20th international conference on computational linguistics (COLING 2004)*, pages 1367–1373, August 2004.

[MTO05] S. Matsumoto, H. Takamura, and M. Okumura. Sentiment classification using word sub-sequences and dependency-trees. *Lecture notes in computer science*, 3518:301–311, 2005.

[NSS07] V. Nastase, M. Sokolova, and J.S. Shirabad. Do happy words sound happy? a study of the relation between form and meaning for english words expressing emotions. *In Proceedings of Recent Advances in Natural Language Processing (RANLP'07)*, pages 406–410, 2007.

[PLV02] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? sentiment classification using machine learning techniques. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 79–86, July 2002.

[Sch94] H. Schmid. Probabilistic part-of-speech tagging using decision trees. *In Proceedings of the International Conference on New Methods in Language Processing*, pages 44–49, 1994.

[SV04] C. Strapparava and A. Valitutti. Wordnet-affect: an affective extension of wordnet. *In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pages 1083–1086, May 2004.

[TL03] P.D. Turney and M.L. Littman. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems (TOIS)*, pages 15–346, 2003.

[TNKS09] T.T Thet, J-C. Na, C. Khoo, and S. Shakthikumar. Sentiment analysis of movie reviews on discussion boards using a linguistic approach. *In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion measurement*, 2009.

[Tur02] P.D. Turney. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL02)*, pages 417–424, 2002.

[WK09] M. Wiegand and D. Klakow. The role of knowledge-based features in polarity classification at sentence level. *In Proceedings of the Florida Artificial Intelligence Research Society Conference (FLAIRS Conference 2009)*, 2009.

[WK10] M. Wiegand and D. Klakow. Bootstrapping supervised machine-learning polarity classifiers with rule-based classification. *In Proceedings 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA)*, 2010.

[WWH05] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. *In Proceedings of the conference on empirical methods in natural language processing (EMNLP 2005)*, pages 347–354, October 2005.