Chapter 1
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 and word semantic orientation extracted from the lexical resource SentiWordNet. Since SentiWordNet is an English resource, we apply a word-translation from French to English before polarity extraction. We show moreover in this article the problems derived by such a translation and their consequences on the word semantic orientation.

1.1 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 by far considered as a classification problem: decide to which class a document is attributed; class of positive or negative polarity.

Most of the work in the field has been intensively applied on the English language. For this purpose, English resources and corpora (such as MPQL [WWH05], Movie Review Data [PLV02], Product Review [YNBN03], Book Review [GA05], SentiWordNet [ES06], WordNet-Affect [SV04, VSS04], the Whissell’s Dictionary...
of Affect Language [Whi89]) 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 but 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 and semantic orientation of selected POS tags. The latter feature 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 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.

1.2 Previous Work

So far, researchers have been used the same classification methodologies and techniques as topic-based categorization [PLV02] with special emphasis on linguistic features in order to increase the performance. As linguistic features, [Gam04, MTO05, NDA06] 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 [MPI07, WK09, TNKS09, ABM09].

Moreover, statistical approaches have been coupled with semantic approaches in order to achieve better results [KH04, PL04, WWH05, Osh08]. 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 [TTC09, AB06, ES05,
GA05, MTO05, TL03].

Only very few work [Den08, ACS08] have explored sentiment analysis in a mul-
tilingual 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.

1.3 Feature Design

Similarly to previous sentiment analysis studies, we have defined three categories
of features. These include lexical, morpho-syntactic and semantic (word polarity)
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.
In order to improve the relevance of unigram features, we propose below further
variants.

Stop words The French language contains a lot of stop words like "de", "du",
"à", "le" and "la". Generally, these words don’t hold polarity information so they
aren’t relevant for the classification. A stop list for removing those words may im-
prove the results.

Lemmatization Grouping all inflected forms of a word in a single term may be
usefull in sentiment analysis. For example, consider the words "aimé", "aimait",
"aimer", "aiment" and "aime", all these words share the same polarity but will be
considered as five separate features during the classification. When applying lema-
tization, we would obtain a unique feature. Features reduction would improve the
tuning of the training process.

Moreover the lemmatization is quite important for our experiments because
WordNet and SentiWordNet use lemmatized words in their dictionary.
1.3.2 Morpho-syntactic features

Definitely not all the words are relevant to the sentiment classification. Some studies showed that restricting features to adjectives would improve performance [HM97], for instance. Part-of-speech (POS) tags are used to add information to word unigrams features in order 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 is also important to aid word sense disambiguation before polarity extraction in SentiWordNet.

1.3.3 Semantic features

As it is shown in 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\(^1\) to extract word polarity and calculate the overall polarity score of the review for each POS tag. SentiWordNet is a corpus-based lexical resource 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.

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\(^2\) 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.

1.4 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\(^3\).

\(^1\) SentiWordNet 1.0.1
\(^2\) This choice is based on the assumption that reviewers spontaneously use an everyday language.
\(^3\) We extracted spectators’ reviews from http://www.allocine.com
Table 1.1 Performance of Different Feature Sets.

<table>
<thead>
<tr>
<th>Features</th>
<th># of features</th>
<th>Results [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) Unigrams</td>
<td>14635</td>
<td>92.00</td>
</tr>
<tr>
<td>(2) Unigrams + stop list</td>
<td>14270</td>
<td>92.00</td>
</tr>
<tr>
<td>(3) Unigrams + lemmatisation</td>
<td>10624</td>
<td>92.00</td>
</tr>
<tr>
<td>(4) Unigrams + lemma. + POS</td>
<td>12229</td>
<td>93.00</td>
</tr>
<tr>
<td>(5) Unigrams + lemma. + POS (N, V, ADJ, ADV)</td>
<td>10350</td>
<td>90.00</td>
</tr>
<tr>
<td>(6) Unigrams + lemma. + POS (ADJ)</td>
<td>2109</td>
<td>79.50</td>
</tr>
<tr>
<td>(7) Unigrams + lemma. + polarity</td>
<td>10632</td>
<td>93.00</td>
</tr>
<tr>
<td>(8) Unigrams + lemma. + POS + polarity</td>
<td>12237</td>
<td>92.00</td>
</tr>
</tbody>
</table>

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\(^4\). 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\(^5\), a French POS tagger and lemmatization tool.

We used Support Vector Machine (SVM) classification method to train and classify French movie reviews. We used SVM\(^{Light}\) [Joa98] classification tool with its standard configuration (linear kernel) to implement a series of experiments where each time we define a set of combined features and evaluate the accuracy of the approach.

\subsection*{1.4.1 Results and Discussion}

The results of the following experiments are summarized in Table 1.1 above. For each experiment labeled from (1) to (8), we present the number of used features and the accuracy measured on the test corpus.

\(^4\) 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.

\(^5\) TreeTagger was developed by Helmut Schmid at the Institute for Computational Linguistics of the University of Stuttgart. \url{http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/}
1.4.1.1 Lexical features

Similarly to [PLV02] we encoded all words features as binary features 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 (1) using the entire set of words is 91.50%. Comparing this results to [PLV02] reporting an accuracy of 82.90% on English movie reviews classification 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 size of our reviews (the collected French reviews are shorter).

When using a stop list (2) we increase the performance by approximately 0.25% up to 91.75%. Although the size of the feature set is reduce by approximately 2.40%, results are not significantly improved. Indeed, stop words seem to be already ignored by the classifier. Therefore a stop list is not useful.

However using the lemmatization (3), we increase the accuracy by 1.00% up to 92.50% and reduce the feature set by approximately 27.00%. These results are quite encouraging for the rest of our experiments because we need to work with lemmatized unigrams to query SentiWordNet.

In order to understand the misclassification of some reviews, we looked deep in their content and noted the following problems.

Neutral reviews Reviews manually interpreted as neutral such as "le film est visuellement réussi mais le scénario est d'une banalité affligeante" are randomly classified according to the dominant sentiment of contained words.

Ironic expressions We noticed that ironic expressions such as “trop fort les gars” that has a negative polarity although it is composed of positive words.

Negation Some reviews that use negation such as “il n’y a plus rien d’extraordinaire” are misclassified. However if we look at the corpus, there are many reviews containing negation and are well classified. For example among about 640 reviews which contain the regex “ne [a-z]* pas” there are only 7 misclassified ones. Generally this is a good result.

Prior classification Only one annotation error was found in the test set. Therefore we can consider that our approach for the prior classification works well.

Misspellings Misspellings 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 could be voluntary to express a kind of stress and intonation such as “ennnnorme”. The problem with such kind of words is that they are irregular in the corpus. For example, “ennnnorme” is highly positive but it is
not present in the feature set so it is not useful. Quite misspelled reviews tend to be misclassified.

**Out of scope span** Some reviews contain subjective sentences that describe other satellite subjects that do not concern the reviewer opinion about the movie. For example we could find a description of a particular scene that does not necessarily reflect the global opinion about the movie such as "Monsieur X est très gentil dans le film.". Such out of scope sentences may affect the classification.

### 1.4.1.2 Morpho-syntactic features

In further experiments, we appended POS tags to every lemmatized unigram so as to disambiguate same unigrams having different senses. However, the effect of this information seems to be irrelevant, as depicted on line (4) of Table 1.1, the accuracy is only increased by approximately 0.25% up to 92.75%.

When filtering unigrams to retain only nouns, adjectives, adverbs and verbs features, we intuitively expect a better classification since we might assume that polarity of a review is substantially hold by such POS categories. Results, shown on line (5) are disappointing since performance is decreased by 1.25% down to 91.25% comparing to the lemmatized unigrams (3). Furthermore, 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 (3) 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 one hand, we notice that negative reviews are better classified than positive ones. On the other hand, we found, in additional experiments, that negative reviews contain relatively an important number of positive adjectives (generally in the negative form). Since we didn’t take into account 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. This last experiment is in contradiction with the results of [HM97] but confirms the results of [PLV02].

### 1.4.1.3 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 1.1, results are improved by only 1.75% up to 93.25% compared to lemmatized unigrams experiment (3). The main reason of such an expectedly and 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 to be able to work with SentiWordNet. However, the quality of translation significantly affects the results of semantic polarity extraction and this is 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 two 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”; the French positive adjective “magique” is translated into the negative adjective “magic” as found in SentiWordNet.

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 “a-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 previously mentioned, the negation was not processed. In principle, the polarity of negated words should be inverted: a negative review which contains many positive words in the negation form should show an overall negative polarity and vice-versa.

Adverbs high negativity  Some adverbs like ne, pas, rien and plus have a very high negativity (about 0.75 over 1). Reviews which abundantly contains these adverbs tend to be classified as negative. We can distinguish two entailments:

- When these words are used in negative reviews, they accentuate the negativity of the document. In fact, this may be useful for negative reviews which relatively contain a large number of positive words in the negation form. Such high negativity may compensate the absence of negation processing and contributes to the improvement of the classification of negative reviews as we have seen in the current experiment.
- However, when these adverbs are used in positive reviews, they tend to inverse the global polarity, mostly for reviews which are weakly positive. In this experiment, some positive reviews are slightly affected by such a problem.

To sum up, we believe that it would be more relevant to eliminate the polarity processing of such adverbs and instead process the negation form.
1.5 Conclusions

In this paper, an unsupervised approach to sentiment analysis of French movie reviews in a bilingual framework was described. It has been 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 necessarily 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.

In future evaluations, the method will be analyzed within a larger training and test sets. Further linguistic analysis will be elaborated such as misspelling correction, negation, WSD and elimination of out of scope text spans from reviews, in addition to the improvement of the translation task.

References


