Aspect-Based Sentiment Analysis Using Lexico-Semantic Patterns

Kim Schouten, Frederique Baas, Olivier Bus, Alexander Osinga, Nikki van de Ven, Steffie van Loenhout, Lisanne Vrolijk, and Flavius Frasincar

Erasmus University Rotterdam
PO Box 1738, NL-3000 DR
Rotterdam, the Netherlands
{414224fb, 419325ob, 427832ao, 356202nv, 431068sl, 413096lv}@student.eur.nl
{schouten, frasincar}@ese.eur.nl

Abstract. With its ever growing amount of user-generated content, the Web has become a trove of consumer information. The free text format in which most of this content is written, however, prevents straightforward analysis. Instead, natural language processing techniques are required to quantify the textual information embedded within text. This research focuses on extracting the sentiment that can be found in consumer reviews. In particular, we focus on finding the sentiment associated with the various aspects of the product or service a consumer writes about. Using a standard Support Vector Machine for classification, we propose six different types of patterns: lexical, syntactical, synset, sentiment, hybrid, surface. We demonstrate that several of these lexico-syntactic patterns can be used to improve sentiment classification for aspects.

Keywords: Lexico-semantic patterns, Support Vector Machines, aspect-based sentiment analysis

1 Introduction

With its ever growing amount of user-generated content, the Web has become a trove of consumer information. Consumers everywhere are invited to share their experiences with products or services they bought and these experiences are in turn shared with prospective buyers to inform their decision making. In this, the Web has transformed the marketplace, putting the electronic word-of-mouth at the core of the decision making process. While reviews are marketed as being useful for prospective consumers, companies are even more interested in all of the expressed opinions toward their products and services. That information enables them to improve their products and optimize marketing strategies.

Unfortunately, the free text format of reviews prevents direct analysis of sentiment. Hence, data mining and natural language processing techniques are used to extract the highly valuable sentiment information. Before sentiment can be extracted, however, the sentiment scope has to be determined, since sentiment can be extracted for complete documents, sentences, or aspects. The advantages of the first two options are that they are easier to do. The disadvantage is that they can not cope with situations where within the unit of analysis (i.e., the document or the sentence), two or more things are discussed that have conflicting sentiment values. To deal with this, sentiment analysis has moved to the aspect level, where sentiment is associated with actual characteristics of the product or service under review. This naturally solves the problem of conflicting sentiment, but the process becomes more complex since the aspects themselves have to be found first. In our research, we focus on the sentiment analysis only, using the aspects that are already provided in the labeled data.

More specifically, we want to investigate the use of lexico-semantic patterns for sentiment analysis, based on the hypothesis that people tend to use similar linguistic structures to express sentiment. For this, we look at lexical patterns, Part-of-Speech (i.e., word types like nouns, verbs, etc.) patterns, and synset (i.e., a set of synonyms that have a single meaning) patterns, and, in addition, at combinations of these. For example, a pattern like 'low' followed by 'quality' denotes a different sentiment than 'low' followed by 'price'. This shows the difficulty of sentiment analysis and it forms the basis why we want to consider various combinations of attributes. We pose that an extended analysis of patterns will contribute to the existing sentiment analysis literature.

The paper is structured as follows. We start by discussing some of the related work in Sect. 2, followed by the description of the types of features we want to investigate in Sect. 3. We then describe our methodology and its evaluation in Sect. 4. We give our conclusions and possible directions for future work in Sect. 5.

2 Related Work

This work is a continuation of [7], which argues that patterns, either over adjacent words or over the grammatical structure of a text, can be employed together with a classifier to perform sentiment analysis. The scope in that work is still the sentence level, with all the advantages and disadvantages as discussed in the previous section. The features used are synset-based features, lexical features, and features that use the grammatical structure instead of word adjacency. We extend this research by first moving to the aspect level. Furthermore, we investigate n-grams up to n=4, including some hybrid patterns like a synset followed by a Part-of-Speech tag. However, we only use word adjacency for our patterns, so grammatical relations are not employed to create patterns of non-adjacent words.

Using n-grams instead of just unigrams has been shown to increase performance and it is straightforward to implement [2,4]. For example, [4] uses both unigrams and bigrams to estimate aspect sentiment. However, the unigram feature still proved to be the most important in the ablation experiment, where this feature was left out to measure the drop in performance compared to including it in the feature set.

Part-of-Speech information, or grammatical word categories, has been used in text classification for a long time. In [3], for example, Part-of-Speech is used to filter out certain words, as this research focuses on the sentiment orientation of adjectives. One of the main conclusions from this research is that adjectives that are linked to each other with a conjunction like ‘and’ often have the same or at least a similar sentiment value. The opposite is true when adjectives are linked with ‘but’. Furthermore, Part-of-Speech can to some extent be used to detect negated information. Negations are crucial for proper sentiment analysis. People are more likely to use negations with negative sentiment than with positive sentiment, so positive words are negated to become negative, but negative words are usually not negated to get positive words [5].

In [5], the authors investigated Part-of-Speech patterns for sentiment analysis on Twitter data. For example, sequences such as “I just”, “I seriously”, “I never”, etc., are all patterns of the form ‘Personal Pronoun followed by Adverb’. In their research, this pattern proved to be associated with negative sentiment. The top 100 best patterns, ranked by their Information Gain score, is included as features, which significantly improves the performance compared to only using unigram features.

3 Lexico-semantic Patterns

The various features we investigate in this research can be placed in six categories: synset, lexical, syntactical, sentiment, hybrid, and surface features. Synsets are a part of semantics, and are sets of synonyms that have a single meaning. Hence, synsets are more specific than the original words, since any ambiguity is eliminated. Both unigrams and bigrams are used here, with the caveat that for synset bigrams we ignore the order of the adjacent synsets. We believe this will make the features more robust, at only a small cost to accuracy. Hence, seeing synsets A and B is the same as seeing synsets B and A.

The lexical category consists of word patterns, where we use the lemma, or dictionary form, of each word in the patterns. We investigate unigrams through quadgrams, since n-grams with n larger than four are too sparse to be of practical use. The syntactical patterns are all sequences of Part-of-Speech (POS) labels. These labels match with any word in that particular word group (i.e., the ‘Noun’ label will match any noun). As such these patterns are more generic than lexical patterns, making them more robust, but less descriptive. We investigate POS patterns ranging from bigrams to quadgrams.

Furthermore, we look at negator-POS bigrams, which are in fact hybrids between syntactical and lexical features. The bigram consists of a negator, from a list of negator words like ‘not’ and ‘never’, followed by a Part-of-Speech label. It effectively splits the Part-of-Speech bigrams that start with an adverb into negating and non-negating bigrams, since words like ‘not’, ‘very’, and ‘highly’ all have the same Part-of-Speech label. We also look at hybrid patterns that combine a Part-of-Speech label and a synset in one bigram.

Since the task is sentiment classification, it makes sense to include sentiment related features as well. For that we use the SentiWordNet dictionary [1], where synsets are given a positive, negativity, and objectivity score that always sum up to one. We compute a sentiment score from those by subtracting the negativity score from the positivity score. This is denoted as a sentisynset. We also look at negator-sentisynset bigrams where a negator is followed by a sentisynset, since this will invert the influence it has on the sentiment classification.

The surface feature is actually not related to patterns, but instead it determines how much of the surrounding context in a sentence is taken into account when creating features for a given aspect. Whenever the exact location of an aspect within a sentence is provided in the annotated data, we use that to create a window of words around that aspect. The words within that window are the only source of information from which to create features for that specific aspect. This allows us to predict different sentiment classes for aspects that are in the same sentence. Unfortunately, for some aspects, the exact location within a sentence is not provided, in which case we cannot specify a specific window and are limited to use the whole sentence as a source for features. The window is defined as k words before the aspect and j words after the aspect, but bound to be within the same sentence.

4 Methodology

For the experiments, we use a linear multi-class Support Vector Machine (SVM). We perform a 10-fold cross-validation to ensure stable results, and from the 90% training data, we designate 20% as validation data. The latter is used to perform feature selection. The rest is used to train the SVM model itself. The final results, as reported in Table 5, are obtained by training on 80% of the training data, using 20% of the training data as a validation set, and evaluating on the official SemEval2015 test data for both data sets.

To determine which types of features perform the best, a forward feature selection is performed. In each round the effect of adding just an isolated feature type is measured. The feature type that gives the highest increase in performance is added to the selected set of features. Again, all remaining types of features are tested, until no increase in performance is measured. The baseline score is simply the majority class. For our data, the 'positive' sentiment class is the most prevalent, as can be seen in Table 1.

The two datasets that are used in our experiments are the English restaurant review data set and the English laptop review data set from SemEval 2015 [6].

The first part of the evaluation is dedicated to the feature selection, showing the effect of each type of feature on performance. First, starting with no features at all, the baseline always predicts positive (the majority class). Every type of feature is added in isolation and the performance is measured. This is the first step in the forward feature selection and the results of this step are presented in Table 2. As expected, word unigrams and word bigrams are the two strongest types of features in this setup. Interestingly, the various feature types perform

TABLE 1: Sentiment value distributions for the two used data sets.

	Positive	Neutral	Negative	Total
Restaurants	1198	53	403	1654
Laptops	1103	106	765	1974

differently on the two data sets. Features that are useful for the laptop data are not beneficial for the restaurant data and the other way around. This shows how domain dependent sentiment analysis is.

Carrying out the forward feature selection procedure results in an optimal set of *word unigram*, *synset bigram*, *sentisynset unigram*, and *synset unigram* for the laptop domain and an optimal set of *word unigram*, *synset bigram*, *sentisynset unigram*, *POS bigram*, and *negator-POS bigram* for the restaurant domain.

Reversing the above process is known as an ablation experiment. Here we start with the optimal set of features, and record the effect of removing one of the feature types. These results are shown in Table 3. Of interest is that while

TABLE 2: The effect of using an additional particular feature type versus the majority baseline for both data sets.

	Laptops	Restaurants
Baseline	0.497	0.637
+ word unigram	**0.754**	0.694
+ word bigram	0.738	**0.713**
+ word trigram	0.572	0.637
+ word quadgram	0.500	0.637
+ POS bigram	0.599	0.634
+ POS trigram	0.602	0.640
+ POS quadgram	0.525	0.637
+ synset unigram	0.696	0.669
+ synset bigram	0.597	0.672
+ synset-POS bigram	0.663	0.675
+ negator-POS bigram	0.555	0.637
+ sentisynset unigram	0.580	0.637
+ negator-sentisynset bigram	0.497	0.637

TABLE 3: Results of the ablation experiments for both data sets. The '-' in the first column denotes set difference.

	Laptops Accuracy	Restaurants Accuracy
Using optimal feature set	76.80%	73.18%
- *word unigram*	-9.95%	-0.99%
- *synset bigram*	-2.49%	-2.20%
- *sentisynset unigram*	-1.94%	-1.58%
- *synset unigram*	-0.29%	not selected
- *POS-bigram*	not selected	-2.21%
- *negator bigram*	not selected	-0.95%

the word unigram features are very important for laptops, this is less true for restaurants, where synset bigrams and POS bigrams are the most important. In contrast, the sentisynset unigram feature is about as equally important for both domains.

Subsequently, the optimal window size is computed that limits the words from which features are extracted for a given aspect. This is only of interest for the restaurant data, since only there exact aspect locations are provided for many of the aspects. We find that the optimal window size is 8 words before and 8 words after the aspect (but always limited by sentence bounds). However, with $k = j = 7$ and $k = j = 9$, roughly the same performance is achieved, losing only 1.27% in accuracy.

To go one level deeper, we looked at the weight of individual features as assigned in the trained SVM model. To make interpretation of these weights easier, we removed the 'neutral' class, resulting in a binary classifier (i.e. positive and negative only). Note that some words only appear with or even have just a single meaning. In that case, the (senti)synset feature has the same weight as the lexical feature of the same word (e.g., 'amazing' in the first column). Of interest are the domain specific words that appear with high weights, such as 'soggy' which is obviously negative for the restaurant domain, but is not used in the laptops domain, and 'Dell' which for this data set is an indicator of negative sentiment for laptops, but of course irrelevant for restaurants.

The scores of the best performing feature sets for each data set are reported in Table 5. These use the optimal window size as discussed above. Overall, we obtain an F_1-score of 69% for restaurant reviews and 73.1% for laptops reviews. Looking at the precision and recall values for the different sentiment values, we can see that on the restaurant data, the SVM tends to classify too many aspects as positive, since both the precision for positive and the recall for negative is relatively low. This seems less the case for the laptops data, resulting in a higher overall score.

Table 4: The most influential features, according to the weight (positive or negative) assigned by the SVM. The feature types are denoted as folows: W is word unigram, SS is synset unigram, and SSS is sentisynset unigram. The SVM is run using the optimal set of feature types.

Restaurants				Laptops			
Positive		Negative		Positive		Negative	
Best (SSS)	0.348	Be (SSS)	-0.639	Be (SS)	0.893	Not (W)	-0.621
Be (SSS)	0.317	Not (SSS)	-0.562	Love (W)	0.696	Be (SS)	-0.593
Amazing (W)	0.31	Soggy (W)	-0.473	Amazing (W)	0.564	Worst (W)	-0.503
Amazing (SSS)	0.31	Worst (W)	-0.408	Great (W)	0.516	Worst (SS)	-0.503
Love (W)	0.304	Worst (SSS)	-0.408	Love (SS)	0.508	Dell (W)	-0.458

Table 5: Overview of classifications on the SemEval 2015 restaurants and laptops test data using the optimal features.

	Restaurants			Laptops		
	Precision	Recall	F_1-score	Precision	Recall	F_1-score
Positive	68.1%	87.4%	76.6%	76.5%	86.7%	81.3%
Neutral	33.3%	4.4%	7.8%	22.2%	10.1%	13.9%
Negative	72.7%	53.2%	61.4%	72.6%	66.0%	69.1%
All	69.0%	69.0%	69.0%	73.1%	73.1%	73.1%

5 Conclusion

In this work we employ and investigate lexico-semantic patterns for aspect-based sentiment analysis. We show that some of the investigated patterns improve the sentiment classification. For laptops the combination of word unigrams, synset unigram, synset bigrams, and sentisynset unigrams prove to be the best performing from amongst the feature types included in our experiments. It is interesting to see that semantical features such as synsets are preferred over other types of features such as the more syntactical Part-of-Speech (POS) bigrams. For restaurants, the best performing combination of feature types is word unigrams, synset bigrams, sentisynset unigram, POS-bigram, and negator-POS bigram. Again, the synset bigram is included, but additionally, the POS bigram and negator-POS bigram are included as well. Evidently, in the restaurant reviews, sentiment is expressed using more consistent syntactical patterns. This points to a difference in language use for reviews about laptops compared to reviews about restaurants. Exactly what these differences entail and why this phenomenon occurs is an interesting avenue for future research.

Another option for future work is to include even more feature types, as there are types of features that, as of yet, were not included in our experiments. Examples of these include additional lexicons and features based on grammatical relations. In conclusion, lexico-semantic patterns prove to be powerful predictors for sentiment analysis, as shown by the 69.0% and 73.1% F_1-score for restaurant and laptop reviews, respectively, but more research is needed to provide definitive answers.

Acknowledgments

The authors are partially supported by the Dutch national program COMMIT.

References

1. Baccianella, S., Esuli, A., Sebastiani, F.: SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In: Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010). vol. 10, pp. 2200–2204 (2010)
2. Brychcín, T., Konkol, M., Steinberger, J.: UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 817–822. Association for Computational Linguistics and Dublin City University (2014)
3. Hatzivassiloglou, V., McKeown, K.R.: Predicting the Semantic Orientation of Adjectives. In: Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics (ACL 1997). pp. 174–181. Morgan Kaufman Publishers and Association for Computational Linguistics (1997)
4. Kiritchenko, S., Zhu, X., Cherry, C., Mohammad, S.: NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 437–442. Association for Computational Linguistics and Dublin City University (2014)
5. Koto, F., Adriani, M.: The Use of POS Sequence for Analyzing Sentence Pattern in Twitter Sentiment Analysis. In: Proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA 2015). pp. 547–551. IEEE (2015)
6. Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., Androutsopoulos, I.: SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In: Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015). pp. 486–495. Association for Computational Linguistics (2015)
7. Schouten, K., Frasincar, F.: The Benefit of Concept-Based Features for Sentiment Analysis. In: Proceedings of the Second Semantic Web Evaluation Challenge at ESWC 2015. Communications in Computer and Information Science, vol. 548, pp. 223–233. Springer International Publishing (2015)