Exploring Lexico-Semantic Patterns for Aspect-Based Sentiment Analysis

Frederique Baas
Erasmus University Rotterdam
Rotterdam, the Netherlands
414224fb@student.eur.nl

Olivier Bus
Erasmus University Rotterdam
Rotterdam, the Netherlands
419325ob@student.eur.nl

Alexander Osinga
Erasmus University Rotterdam
Rotterdam, the Netherlands
427832ao@student.eur.nl

Nikki van de Ven
Erasmus University Rotterdam
Rotterdam, the Netherlands
356202nv@student.eur.nl

Steffie van Loenhout
Erasmus University Rotterdam
Rotterdam, the Netherlands
431068sl@student.eur.nl

Lisanne Vrolijk
Erasmus University Rotterdam
Rotterdam, the Netherlands
413096lv@student.eur.nl

Kim Schouten
Erasmus University Rotterdam
Rotterdam, the Netherlands
schouten@ese.eur.nl

Flavius Frasincar
Erasmus University Rotterdam
Rotterdam, the Netherlands
frasincar@ese.eur.nl

ABSTRACT

Web 2.0 has caused a boom in user-generated content, which contains a lot of valuable information. Analysis of these natural language data requires advanced machine learning techniques. This research focuses on determining aspect-based sentiment in consumer reviews using lexico-semantic patterns. We propose a method using a Support Vector Machine with 6 different pattern classes: lexical, syntactical, semantic, sentiment, hybrid, and surface. We show that several of these patterns, including synset bigram, negator-POS bigram, and POS bigram, can be used to better determine the aspect-based sentiment, using two widely used real-world data sets on consumer reviews. Our approach achieves 69.0% and 73.1% F_1 score for the two data sets, respectively, an increase of 15.3% and 16.1% respectively compared to the considered baseline.

CCS CONCEPTS

• **Information systems** → **Sentiment analysis**; *Information extraction*; Web mining;

KEYWORDS

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

SAC '19, April 8–12, 2019, Limassol, Cyprus

ACM ISBN 978-1-4503-5933-7/19/04. . . $15.00
https://doi.org/10.1145/3297280.3297377

ACM Reference Format:
Frederique Baas, Olivier Bus, Alexander Osinga, Nikki van de Ven, Steffie van Loenhout, Lisanne Vrolijk, Kim Schouten, and Flavius Frasincar. 2019. Exploring Lexico-Semantic Patterns for Aspect-Based Sentiment Analysis. In *The 34th ACM/SIGAPP Symposium on Applied Computing (SAC '19), April 8–12, 2019, Limassol, Cyprus*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3297280.3297377

1 INTRODUCTION

In our current society, the World Wide Web is hard to avoid. Fed by the growing use of tablets and smartphones, the time spent online increases [17]. Much of the user-generated content contains opinions, such as reviews of products or services. Everybody can be a tradesman, providing advices on what mobile phone to buy or which restaurant is worth visiting. The shift in communication from one-to-one to one-to-many shapes a new marketplace in which the electronic word-of-mouth triumphs. Hence, the electronic word-of-mouth is now at the foundation of the decision-making process of today's customer [27]. For companies, the large-scale online interactions of users make Web-based opinionated content very valuable as it is important to know the sentiment of current and potential customers towards their products or services.

The relevant product or service data, verbalized in natural language, are readily available on the Web. These new sources of information are instrumental to optimize marketing strategies. Data mining and natural language processing techniques are needed to be able to process this large amount of data and extract valuable information. As this process of extracting actionable knowledge requires a deep understanding of the explicit and implicit, regular and irregular, and syntactical and semantic language rules, it is challenging to analyze unstructured Web data [17].

In addition, the level of analysis is important when designing the sentiment mining process. Determining the sentiment can be on document, sentence, or aspect level. A difficulty that occurs with both document and sentence level analysis is *confluence*, where one unit of analysis (i.e., a document or sentence) contains polar opinions. To illustrate this phenomenon, the following sentence from a data set containing restaurant reviews, is shown below:

"The food was good, but the service was bad."

In this review, the aspects 'food' and 'service' are mentioned with respectively a positive and a negative sentiment. As these polarities are conflicting, the sentiment of the sentence as a whole will be in the middle, in this case neutral. By labeling this sentence as neutral, it is suggested that there is no sentiment in this sentence, which is not the case.

The previously identified problem can be prevented by analyzing a sentence on aspect level. Moreover, as consumers generally prefer to judge a product or service on its various aspects (e.g., the battery of a phone), instead of the product (e.g., phone) as a whole, the analysis on aspect-level will provide the most interesting information. In addition, an analysis at aspect-level can enable a better accuracy at higher level analysis, as well, as the aspect scores can be aggregated to sentence or document level. Because of these reasons we will focus on sentiment analysis at aspect-level.

The main goal of this paper is to determine whether and which lexico-semantic patterns are useful to detect sentiment of aspects in reviews, based on the hypothesis that people tend to use a fixed structure to express their opinion. For this purpose, we do not perform aspect detection, but instead focus on sentiment analysis in isolation, free from aspect detection errors that might otherwise be propagated and hence pollute the results. This means we compute the sentiment for predefined, manually annotated, aspects. The classifier predicts sentiment as being either positive, neutral, or negative.

When investigating this topic we consider lexico-semantic patterns to entail lexical patterns, POS patterns, and synset patterns as well as hybrid structures of these. An example of such a pattern is the combination of the word 'low' with different other words. The word 'low' in combination with the word 'price' will convey a positive sentiment, while the co-occurrence with 'quality' will result in a negative sentiment. This shows that it might be useful to determine patterns by considering various combinations of attributes (e.g., n-grams, Part-Of-Speech tags, synsets, negators etc.) and whether they regularly occur and mostly correspond to positive or negative aspect sentiment. With this paper we aim to contribute to the existing sentiment analysis literature by exploring lexico-semantic pattern occurrence in consumer reviews to accurately determine aspect sentiment. Hence, the main goal of this research is to provide insight and not to beat the state-of-the-art in terms of raw performance. Therefore, the evaluation includes an extensive analysis at the used patterns and not the traditional comparison with related work to show that this method is better.

A preliminary work on this topic has been published in [21] and this paper extends that version by providing a more detailed discussion of the used features, an in-depth look at the used methodology and data sets, and lastly, a more thorough evaluation that includes a failure analysis and a sensitivity analysis.

The paper is structured as follows. First, we discuss related work that inspired this research in Sect. 2. In Sect. 3 the proposed methodology is presented, including feature description, data analysis, and evaluation results. Last, in Sect. 4, the main conclusions and proposals for further work are expounded.

2 RELATED WORK

The detection and categorization of sentiment of user-generated content on the Web has sparked an exciting research field generating a great number of publications using a variety of approaches and techniques. As mentioned in Sect. 1, our aim is to improve natural language processing by increasing our understanding of the text's underlying structure. Natural language is a complex combination of vocabulary, grammar, but also external factors like intent, and interpretation. The latter, usually regarded as the pragmatic level of natural language processing is not covered in this work, where the main focus is on how the meaning of words is influenced by their grammatical function (e.g., noun, pronoun, verb, adverb, etc.). The structure of a sentence is governed by grammar, which means that words are ordered using specific, predefined patterns [13]. As people tend to use fixed patterns to express their opinion, both with respect to syntactical and semantic structure. These are useful properties when processing the unstructured data from consumer reviews. In [22], it is argued that patterns in adjacent occurrence of words and synsets, as well as patterns in syntactical structure can be employed for sentiment analysis of reviews at a sentence level. The authors use synset-based features, word-based features, and grammatical group-based features as input for a linear SVM. This section gives a comprehensive overview of the relevant literature on these three types of features and the different approaches in which these are used for sentiment analysis at an aspect-level. In our research we are going to investigate these existing patterns for aspect-based sentiment analysis as well as propose new features, most of them of a hybrid nature, that are not covered yet in existing literature.

2.1 Synset-based Features

Synset-based features can convey different sentiment values, depending on the context. Let us consider for example the synset 'small#JJ#1', where '#JJ#1' denotes the fact that we use the first sense of this word as an adjective. When this synset modifies the synset 'device#NN#1' ('NN' denotes a noun) it usually conveys a positive sentiment, while when it modifies the synset 'meal#NN#1' it tends to convey a negative sentiment. A synset-based feature can help to identify many of such patterns [15]. Moreover, synsets are generally considered helpful for sentiment analysis, as shown in [22], where it is always beneficial to add synsets to the feature set. However, the performance increases more if a more advanced way of handling semantic information is used. In [18] a combination of linguistics, common-sense computing, and machine learning is used to improve the accuracy of polarity detection. It allows sentiments to flow from synset to synset, based on the dependency relation of the input sentence. Through this, a polarity detection engine is obtained that outperforms state-of-the-art statistical methods.

2.2 Word-based Features

Often an n-gram feature is included in an SVM, as it increases the score considerably and it is relatively easy to incorporate [3, 10, 14]. In [10], for example, unigrams (single words) and bigrams (sequences of two words) are extracted from a term and its surface context to estimate aspect term polarity. The rationale behind this feature is that the pattern regarding certain occurrences of n-grams tend to correspond to either positive or negative aspects.

In [10], the authors also show that the unigram feature increases the accuracy score the most. An ablation experiment, meaning running the proposed algorithm with all features, except for the unigram feature, showed a 6.88%-point decrease in accuracy score for the used laptop data set and 3.03%-point decrease for the used restaurant data set.

Another pattern, considered in [3], [10], and [24], is related to the surface context of the aspect of interest. This is a basic feature that selects a window of n words surrounding the aspect term, as the distance between sentiment affecting words and the aspect term itself relates to the importance of words to detect the aspect polarity. In addition, weights can be assigned by dividing the sentiment value of each n-gram by the distance between that n-gram and the aspect of interest, to attribute more weight to n-grams that occur within the vicinity of the aspect, as described in [3] and [24].

2.3 Grammatical Group-based Features

Part-of-speech (POS) tagging is an often used method to exploit grammatical group patterns [8, 10, 11, 14]. POS tagging has been used in sentiment analysis in various ways. In [8], for example, a specific POS category is used as it focuses on learning the semantic orientation of adjectives. This research shows that adjectives that are connected by conjunctions, often have the same semantic orientation, except when adjectives were connected by "but" which often resulted in a reversed orientation, and thus negated sentiments. Negation may be local, but can also involve longer-distance dependencies [26]. It is less common for people to use negation words in expressing their positive impression. People are more likely to write "I like this phone" instead of "I don't hate this phone"[11].

More recently, the use of a sequence of POS tags is suggested in [11] to improve sentiment analysis. Working with twitter posts, [11] investigates patterns of word combinations that indicate a certain sentiment. For example, the following sequences are considered:

"I highly", "I seriously", "I never", "me crazy", "I just."

which are tagged in [11] as a pronoun and adverb combination (i.e., PRP-RB). This pattern is used more often when expressing negativity. The top-100 sequences of n-tags were selected based on Information Gain. This top-100 was then included as features, resulting in a significant improvement in the accuracy score of polarity classification compared to the model that used only word features.

In [5], an approach is presented were the weights of a feature matrix were adjusted using POS tags. The authors hypothesized that since nouns are entities and adjectives are qualifiers, more weight should be given to their combinations in order to determine positive or negative polarity of a sentence. Their best accuracy score was obtained by using POS bigrams and term frequencies resulting in a 76.6% accuracy, whereas only using bigrams and term frequencies resulted in an accuracy drop of over 20%.

3 METHODOLOGY

To facilitate this research, we have set up a basic linear Support Vector Machine (SVM) using the libSVM library [4]. This classifier, which is a supervised machine learning method, is robust in a high dimensional space, and performs well, even for only small samples [9]. Moreover, any feature can be used as input [20]. The SVM is setup to classify between positive, neutral, and negative as the three sentiment classes using the built-in multiclass support. When an aspect (given to us, and outside the scope of this research) is explicitly mentioned in the text, a word window around the target word is used to create the features from, whereas for aspects that do not have an explicit target, the whole sentence is used to generate features. The size of the word window is defined by two parameters (i.e., number of words before the target and number of words after the target) that are optimized during the training process. Since the feature selection and the size of the word window are not independent, they are ideally trained together. To keep the training feasible, some preliminary experiments were performed that indicated that a word window of 7 words before and 7 words after the target was ideal. This preliminary word window is used when performing the feature selection. Afterwards, the word window is optimized using the validation data.

For the experiments, we use 10-fold cross-validation, where, for each fold, we designate 20% of the 90% training data as validation data. These data are used to optimize the C parameter of the SVM, the size of the word window for explicit aspects, and to perform the feature selection.

3.1 Lexico-semantic Patterns

The features we investigate for this research are all aimed at capturing lexico-semantic patterns in natural language and can be divided five categories: semantic, lexical, syntactical, sentiment, and hybrid. The semantic features contain both *synset unigrams* and *synset bigrams*. The synset bigram feature is added as it is a logical combination of the already successful synset feature and the regular word bigram feature. One difference with a regular bigram is that we do not look at the order of the two adjacent synsets. This is to ensure a relatively high frequency of this type of feature, since not every word has an associated synset in the WordNet [6] semantic lexicon. This can be done since we mostly value the occurrence of two synsets in a sentence and care less about the order of occurrence. For disambiguating words, and thus producing synsets, we have applied the adapted Lesk algorithm [2], an extension of the Lesk algorithm [12].

The class of lexical features represents patterns in adjacent occurrence of certain words and whether this adjacent occurrence generally corresponds to either positive, neutral, or negative aspect sentiments. We consider a *word unigram* (single words), *word bigram* (2-word sequences), *word trigram* (3-word sequences) and *word quadgram* (4-word sequences). Larger n-grams are not considered due to their relative sparsity. All n-grams are constructed after lemmatization, and since these are regular n-grams, the order of the words is taken into account.

The syntactical features are based on POS-sequences. We use the Stanford POS tagger to identify the part of speech of each word, such as noun, verb, etc. Similar to the lexical-based features we have constructed a *POS bigram*, *POS trigram*, and a *POS quadgram*. These features are based on the principle that certain POS sequences are used more frequently to convey positive sentiment, while others are more often used to convey negative sentiment.

Furthermore we have implemented a hybrid feature, *negator-POS bigram*, which detects combinations of a negating word, drawn from the General Inquirer [7] word list, and the subsequent POS-tag. If a negation word like 'not' or 'never' is found in the text then it forms a *negator-POS bigram* with the POS-tag of the following word. This feature is a combination of a lexical and syntactical based feature. When you compare it with a *word bigram* and a *POS bigram* you notice two differences. Consider the following word sequences expressing how the food was: 'not good', 'not tasty', and 'very delicious'. First of all, compared to a *word bigram* you will find more instances with the *negator bigram*, since 'not good' and 'not tasty' will form two separate *word bigrams*, however, they correspond to the same *negator-POS bigram*. Since both combinations convey a negative sentiment, it could improve the sentiment polarity estimation, since more instances will be found using this hybrid pattern. Secondly, all three sequences will form the same *POS-bigram*, namely 'Adverb + Adjective'. However it is apparent that you would like to make a distinction between 'not tasty' and 'very delicious', which the *negator-POS bigram* feature is able to do so. Another hybrid feature implemented is the combination of a POS-tag and a synset (*synset POS-bigram*).

We also implemented sentiment patterns features using SentiWordNet. The features are *sentisynset unigram*, which is the sentiment score of a single WordNet synset, and *negator-sentisynset bigram*, which is similar to the *negator-POS bigram*, but with the sentiment score of a synset instead. This allows us to flip the sentiment score of the synset when preceded by a negator. To arrive at one sentiment score from SentiWordNet, we subtract the negativity score of a synset from the positivity score of that synset. This results in a number between -1 and $+1$. The rest of the features are encoded as a Boolean representing their presence or absence.

Table 1: Marginal effect of adding one feature versus majority baseline. Performance is measured as $\mathbf{F}_1$ on validation data

	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 fourgram*	0.500	0.637
+ *POS bigram*	0.599	0.634
+ *POS trigram*	0.602	0.640
+ *POS fourgram*	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

3.2 Data Analysis

We choose to train our algorithm on the SemEval 2015 data for the restaurants and laptop domains, retrieved from [16]. In this way we can build the model on the train data and evaluate on the test data. The data sets consist of reviews in English about restaurants and laptops. A review consists of multiple sentences: the restaurant data set contains 254 reviews with in total 1315 sentences, whereas the laptop data set contains 277 reviews with in total 1739 sentences. Each sentence can contain zero, one, or multiple aspects, and each aspect is labeled as either positive, neutral, or negative. The statistics for both data sets can be found in Fig. 1 and Fig. 2. When a sentence contains multiple aspects, it is possible that these aspects do not all have the same polarity.

"The screen is huge and colorful, but no LED back lighting."

This sentence contains two polarities, both about the display design features, but the first is positive, while the second is negative. The restaurant data set also distinguishes between aspects that are explicitly mentioned and aspects that are implied.

"Chow fun was dry; pork shu mai was more than usually greasy and had to share a table with loud and rude family."

This sentence contains two explicit features about food, the chow fun and the pork shu mai, both with a negative polarity. The literal expressions of aspects in the sentence are called targets. In addition, it contains the implicit feature about sharing the table with a loud and rude family, which is labeled as ambience. The laptop data set does not have targets, so for our research, they are all considered as implicit aspects. In Fig. 3, the distribution of target length is given for the restaurant data, where a length of zero means it is an implicit aspect, as well as an overview of how many sentences have conflicting polarities in the two data sets.

3.3 Evaluation

The evaluation consists of the following experiments:

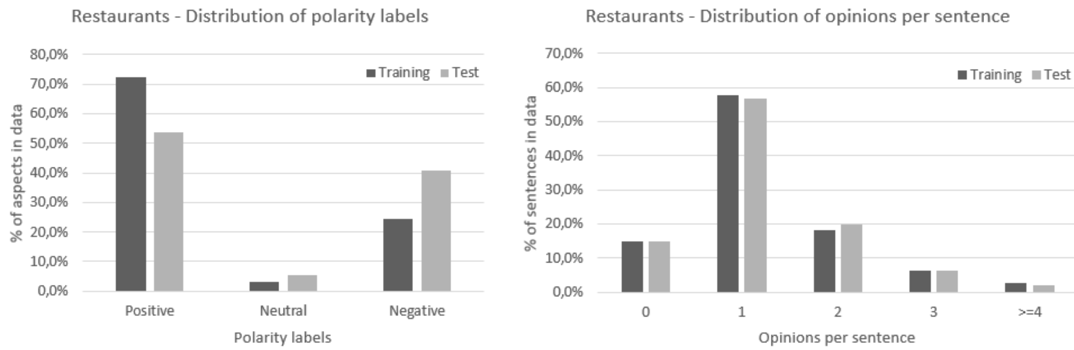

Figure 1: Opinion and polarity statistics for the restaurant data

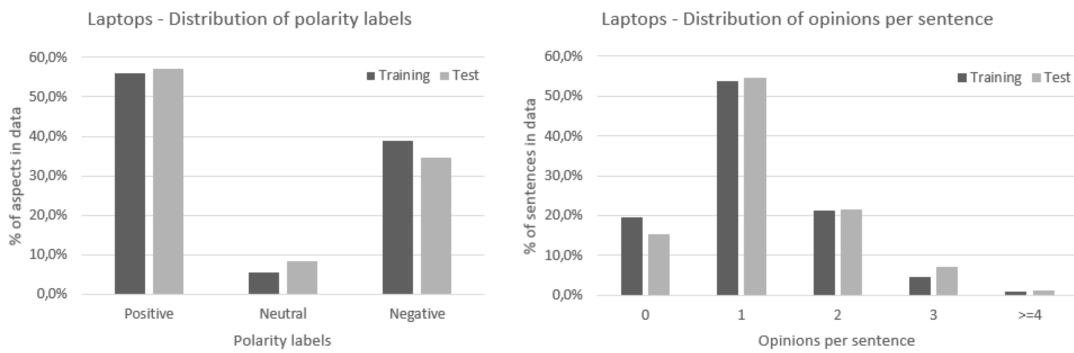

Figure 2: Opinion and polarity statistics for the laptops data

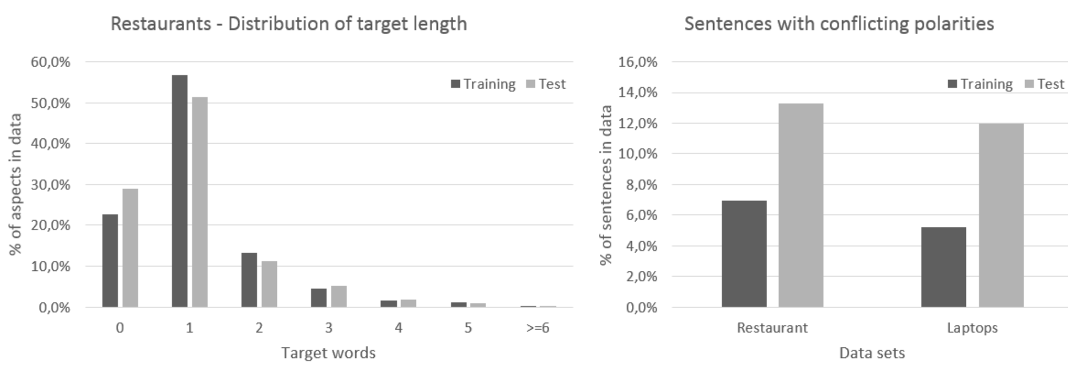

Figure 3: Size of aspect targets for restaurant data (laptop data do not include target information) and sentences with conflicting polarities for both data sets

(1) The marginal effect: what is the influence of adding one feature on the outcome, versus the baseline;
(2) The optimal combination of features: we use a forward feature selection algorithm that aims at finding the optimal combination of features;
(3) An ablation experiment: for the optimal combination of features, what is the impact of leaving one feature out;
(4) The optimal surface context: for which window of words around the aspect does the algorithm find the highest accuracy;
(5) Failure analysis: on what type of sentences does the algorithm fail and for what reason;
(6) Sensitivity of the algorithm to data differences by comparing in-sample, out-of-sample, and 10-fold cross validation results.

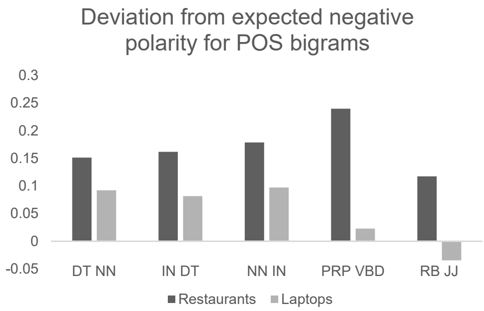

Figure 4: Deviation from the expected negative polarity for POS bigrams.

In Table 1 the marginal effect of adding one feature versus the baseline applying the majority rule (in our case all positive) for the restaurant and laptop domain 2015 test data sets is shown. To complete the feature selection, the best performing feature is selected, and then the process of measuring the marginal effect of adding another feature is repeated.

The process of iteratively adding one feature to the set of optimal features results in the combination of these four features for the laptop domain: *word unigram*, *synset bigram*, *sentisynset unigram*, and *synset unigram*. In the restaurant domain, the set of optimal features consists of *word unigram*, *synset bigram*, *sentisynset unigram*, *POS bigram*, and *negator-POS bigram*. Of interest is the fact that the POS bigram is selected for the restaurant data, but not for the laptop data. This can be traced back to the prevalence of certain POS bigrams when describing negative aspects. The same patterns, however, are not prevalent in the laptops domain. This is illustrated in Fig. 4, where the expected negative polarity for a number of POS bigrams are shown. When a data set, where, for instance, 30% of the aspects has a certain feature, the expected relative frequency of a certain POS bigram in a positive context would be approximately 30% as well. The graph shows the deviation from this expectation. For instance, the POS bigram DT-NN appears 15%-points more often in a negative sentence than expected for the restaurant data set. The acronyms indicate the following: DT indicates a determiner, NN indicates a noun, IN indicates a preposition, PRP indicates a personal pronoun, VBD indicates a verb in past tense, RB indicates an adverb, and JJ indicates an adjective. For example, a sentence that contains a personal pronoun followed by a verb in the past tense is strong evidence for a negative polarity when reviewing restaurants, but is not a particularly strong indicator for negativity when writing about laptops.

Subsequently, we have analyzed the optimal surface context to take into account words around the aspect of interest. This is only important for the restaurant data set. From all combinations we find that the optimal window contains 8 words before the aspect term and 8 words after the term. However, adjusting the window from $k = j = 7$ to $k = j = 8$ only increases the accuracy score on the 10-fold cross-validation with 1.27%. Hence, our preliminary experiments that pointed towards setting $k = j = 7$ were indeed close to the optimal values already.

Table 2: Ablation experiments for the laptop and restaurant domain 2015 training data set, with '-' denoting set difference.

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

In Table 2 we see the marginal contribution of all selected features on laptop and restaurant training data sets. The results were obtained by conducting an ablation experiment, i.e., running the algorithm with all features, except for the feature of interest. We can see that the *word unigram* feature has the strongest contribution for the laptop data set. Furthermore, we see that the *synset bigram* feature has a strong contribution in both the laptop and restaurant domain. Removing this feature leads to a drop in accuracy score of more than 2% for both domains.

To see which features and words have the most impact in predicting the polarities, the features that the linear SVM assigns the largest weights to are reported in Table 3. For this experiment, the SVM is run with optimal set of features and optimized parameters. To make interpretation easier, we have removed the 'neutral' polarity class for just this experiment and performed a binary classification. Note that for some words the word unigram feature and the (senti)synset unigram feature are perfectly collinear and therefore have the same weight. Furthermore, we observe some domain-specific terms in the largest weights, for the restaurant domain the word 'soggy' and for the laptop domain the word 'Dell'. Last, as expected from the relation between POS and sentiment discussed above, we find that there is a large difference in polarity for the use of the verb 'be' in the present tense compared to its use in the past tense.

Using the feed forward feature selection method, the set of optimal features types is determined and with these features, the SVM is trained. For both the restaurant and the laptops domain, the SVM is trained on the official SemEval 2015 training data, and tested with the official SemEval 2015 test data. The latter is not used prior to this step. The results can be found in Table 4.

To investigate why the algorithm has incorrectly predicted the sentiment of some of the aspects, we will examine a few of the wrongly predicted reviews in more detail. We start by considering the following review from the data set concerning the laptop domain:

Table 3: The most important features according to the absolute weight given by the SVM. The feature types are denoted as folows: W is word unigram, SS is synset unigram, and SSS is sentisynset unigram. A single * denotes present tense, and ** denotes past tense

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 4: Overview of classifications on the SemEval 2015 restaurants and laptops test data using the algorithm with forward feature selection on *ALL* 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%
Majority baseline			53.7%			57.0%
SVM with BoW			63.6%			70.0%

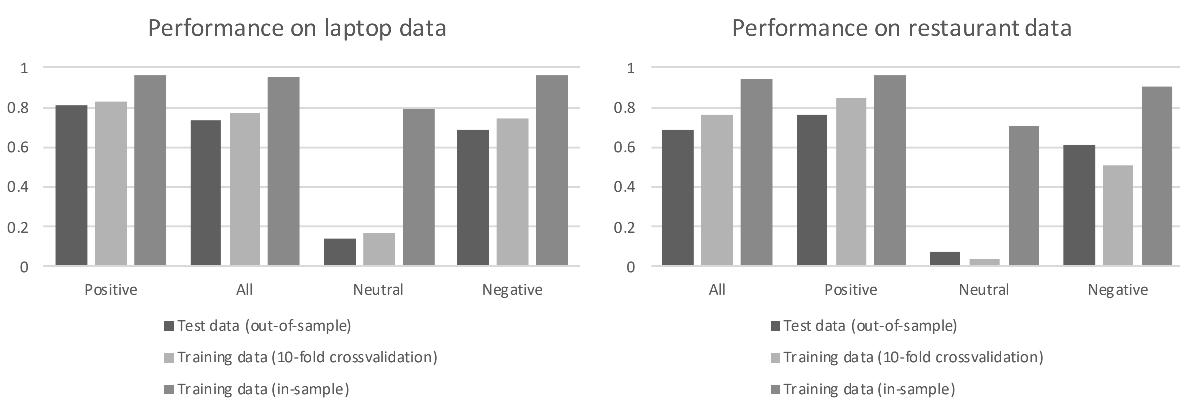

Figure 5: Overview of F_1-scores for the various sentiment classes on the test data, on the training data with 10-fold-cross-validation, and in-sample on the training data for the SemEval 2015 restaurant and laptop data

> "It is not a small, compact laptop, but I won't be traveling via air often, so its size and weight is not a problem for me." (1)

In this review, 'laptop design' is the implicit feature. Its sentiment is annotated as positive, yet predicted as negative. This is probably due to the complex semantics of the sentence, where first something negative is mentioned, but afterwards that negative point is itself rendered irrelevant with another negation. Our current set of features is not able to grasp this intricate expression and wrongly assumes that all the negations convey a negative sentiment. One could also argue that this sentence expresses a neutral polarity, showing the high level of subjectivity that needs to be taken into account in the field of natural language processing and evaluation.

Another example of an incorrectly classified aspect sentiment can be found in the following sentence.

> "kinda to light and plastic feeling." (2)

Here the 'laptop design' aspect is again implied. Its polarity is annotated as negative but it is predicted as positive by our algorithm. Note that there is a spelling error in this sentence, since 'to' should be spelled as 'too' here. While this spelling mistake is obvious for humans, the proposed algorithm is not able to pick that up. With 'light' being a positive trait for laptops, missing the 'too' means a wrong prediction in this case.

When looking at both reviews (1) and (2), we can conclude that the design features are user-specific. In review (1) a negative sentiment is assigned to the weight of the laptop as it is probably too heavy (negative), yet corresponds to a positive review since this specific user does not have to travel often. In contrast, review (2) assigns a negative sentiment to a laptop that is too light and prefers a more heavy one. These user specific preferences make it difficult to predict the right sentiment, even when taking the domain specific context into account. Lexico-semantic patterns will not help to catch all these intricate nuances of expressing sentiment.

The following sentence illustrates another such misspecification:

"The biggie though is the fact that it disconnects from the internet whenever if feels like it, even when the strength bar is filled."

This review about laptop connectivity is predicted positive, but annotated as negative. The specific writing style in this example is likely the cause of misclassification. Since there is a large variety in writing-styles, lexico-semantic patterns are able to capture these only to a limited extent.

The relatively low score on the restaurants data set can be explained partially by the distribution of polarities, as shown in Fig. 1. The restaurant training data set contained relatively more positive occurrences than the test data set. This is also supported by evidence from Fig. 5. While the accuracy on the test data is similar to the accuracy using 10-fold cross validation on the training data set for the laptops domain, the same comparison for the restaurants domain shows large differences. We also see that in general the algorithm is less accurate in predicting negative polarities than positive ones. The fact that the restaurant test data set contains relatively more negative polarities negatively influences the accuracy.

4 CONCLUSION

In our proposed method, we constructed a classifier that predicts the aspect-based sentiments in consumer reviews with the help of lexico-semantic patterns. We demonstrate that several of these lexico-semantic patterns can be used to improve the sentiment classification of an aspect. In the laptop domain we selected *word unigram*, *synset bigram*, *sentisynset unigram* and *synset unigram* as our best feature subset. Here we can conclude that lexico-semantic patterns using synsets contribute more to sentiment analysis than syntactical-based features such as bigrams and POS-tags. For the restaurant domain we selected *word unigram*, *synset bigram*, *sentisynset unigram*, *POS bigram*, and *negator-POS bigram*. Again, the *synset bigram* feature showed its added value. It is interesting to note that other patterns captured by the *negator-POS bigram* and the *POS bigram* features showed significant improvements in this domain, too. However, this increase is not seen in the laptop domain. A reason for this could be that the prevalence of certain patterns is more common or detectable in some domains than others. Overall, using lexico-semantic features in combination with a SVM shows to be a powerful method, as we correctly predicted 69.0% and 73.1% of the aspect-sentiment in, respectively, the restaurant and laptop domains. As future work we will like to extend our investigation to recently proposed features coming from ontological forms learned on-the-fly from the analysed text [1, 19] or based on domain expert knowledge [23, 25].

5 ACKNOWLEDGMENTS

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

REFERENCES

[1] Mattia Atzeni, Amna Dridi, and Diego Reforgiato Recupero. 2018. Using Frame-Based Resources for Sentiment Analysis within the Financial Domain. *Progress in Artificial Intelligence* 7 (08 2018).

[2] Satanjeev Banerjee and Ted Pedersen. 2002. An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In *Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2002) (Lecture Notes in Computer Science)*, Vol. 2276. Springer, 136–145.

[3] Tomáš Brychcín, Michal Konkol, and Josef Steinberger. 2014. UWB: Machine Learning Approach to Aspect-Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, 817–822.

[4] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A Library for Support Vector Machines. *ACM Transactions on Intelligent Systems and Technology* 2 (2011), 27:1–27:27. Issue 3. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[5] Oaindrila Das and Rakesh Chandra Balabantaray. 2014. Sentiment Analysis of Movie Reviews using POS tags and Term Frequencies. *International Journal of Computer Applications* 96, 25 (2014), 36–41.

[6] Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

[7] General Inquirer 2016. General Inquirer Home Page. http://www.wjh.harvard.edu/~inquirer/.

[8] Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. 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)*. Morgan Kaufman Publishers and Association for Computational Linguistics, 174–181.

[9] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. 2016. A Practical Guide to Support Vector Classification. https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.

[10] Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, 437–442.

[11] F. Koto and M. Adriani. 2015. 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)*. IEEE, 547–551.

[12] M. Lesk. 1986. Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In *Proceedings of the Fifth Annual International Conference on Systems Documentation (SIGDOC 1986)*. ACM, 24–26.

[13] R. Miarka and M. Žáček. 2011. Knowledge Patterns for Conversion of Sentences in Natural Language into RDF Graph Language. In *Proceedings of the Federated Conference on Computer Science and Information Systems 2011 (FedCSIS 2011)*. IEEE, 63–68.

[14] Naveen Nandan, Daniel Dahlmeier, Akriti Vij, and Nishtha Malhotra. 2014. SAP-RI: A Constrained and Supervised Approach for Aspect-Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, 517–521.

[15] Nir Ofek and Lior Rokach. 2014. Unsupervised Fine-Grained Sentiment Analysis Using Lexicons and Concepts. In *Proceedings of the Semantic Web Evaluation Challenge at ESWC 2014 (Communications in Computer and Information Science)*, Vol. 475. Springer International Publishing, 28–33.

[16] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, 486–495.

[17] Soujanya Poria, Erik Cambria, Alexander F. Gelbukh, Federica Bisio, and Amir Hussain. 2015. Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns. *IEEE Computational Intelligence Magazine* 10, 4 (2015), 26–36.

[18] Soujanya Poria, Erik Cambria, Grégoire Winterstein, and Guang-Bin Huang. 2014. Sentic Patterns: Dependency-Based Rules for Concept-Level Sentiment Analysis. *Knowledge-Based Systems* 69 (2014), 45 – 63.

[19] Diego Reforgiato Recupero, Valentina Presutti, Sergio Consoli, Aldo Gangemi, and Andrea Giovanni Nuzzolese. 2015. Sentilo: Frame-Based Sentiment Analysis. *Cognitive Computation* 7, 2 (2015), 211–225.

[20] M. Rushdi Saleh, M.T. Martín-Valdivia, A. Montejo-Ráez, and L.A. Ure na López. 2011. Experiments with SVM to Classify Opinions in Different Domains. *Expert Systems with Applications* 38, 12 (2011), 14799 – 14804.

[21] Kim Schouten, Frederique Baas, Olivier Bus, Alexander Osinga, Nikki van de Ven, Steffie van Loenhout, Lisanne Vrolijk, and Flavius Frasincar. 2016. Aspect-Based Sentiment Analysis Using Lexico-Semantic Patterns. In *Proceedings of the 17th International Conference on Web Information Systems Engineering (WISE 2016) (Lecture Notes in Computer Science)*, Vol. 10042. Springer International Publishing, 35–42.

[22] Kim Schouten and Flavius Frasincar. 2015. 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. Springer International Publishing, 223–233.

[23] Kim Schouten and Flavius Frasincar. 2018. Ontology-Driven Sentiment Analysis of Product and Service Aspects. In *18th Extended Semantic Web Conference (ESWC 2018) (Lecture Notes in Computer Science)*, Vol. 10845. Springer, 608–623.

[24] Kim Schouten, Flavius Frasincar, and Franciska de Jong. 2014. COMMIT-P1WP3: A Co-occurrence Based Approach to Aspect-Level Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, 203–207.

[25] Kim Schouten, Flavius Frasincar, and Franciska de Jong. 2017. Ontology-Enhanced Aspect-Based Sentiment Analysis. In *17th International Conference on Web Engineering (ICWE 2017) (Lecture Notes in Computer Science)*, Vol. 10360. Springer, 302–320.

[26] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In *Proceedings of the Conference on Human Language Technology and Conference on Empirical Methods in Natural Language Processing 2005 (HLT/EMNLP 2005)*. Association for Computational Linguistics, 347–354.

[27] Chul Woo Yoo, Yong Jin Kim, and G. Lawrence Sanders. 2015. The Impact of Interactivity of Electronic Word of Mouth Systems and E-Quality on Decision Support in the Context of the E-Marketplace. *Information and Management* 52, 4 (2015), 496 – 505.