

An Ontology-Enhanced Hybrid Approach to Aspect-Based Sentiment Analysis

Daan de Heij, Artiom Troyanovsky, Cynthia Yang, Milena Zychlinsky Scharff, Kim Schouten^(✉), and Flavius Frasinca

Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

{411991dh,414550at,405574cy,382501mz}@student.eur.nl,
{schouten,frasincar}@ese.eur.nl

Abstract. Numerous reviews are available online regarding a wide range of products and services. Aspect-Based Sentiment Analysis aims at extracting sentiment polarity per aspect instead of only the whole product or service. In this work, we use restaurant data from Task 5 of SemEval 2016 to investigate the potential of ontologies to improve the aspect sentiment classification produced by a support vector machine. We achieve this by combining a standard bag-of-words model with external dictionaries and an ontology. Our ontology-enhanced methods yield significantly better performance compared to the methods without ontology features: we obtain a significantly higher F_1 score and require less than 60% of the training data for equal performance.

1 Introduction

The large number of online reviews has brought a new challenge: quantifying the opinion expressed by individuals in these reviews. In this paper, we focus on Aspect-Based Sentiment Analysis (ABSA) of reviews. ABSA is useful for fine-grained sentiment analysis: polarities are connected to specific aspects expressed in the text [6]. To do so, we use the aspects defined in the annotations of the data. For each aspect we establish the target: the part of the sentence that explicitly mentions the aspect. Sometimes, the aspect is implicitly mentioned by the reviewer, and thus the target does not exist. The sentiment of the aforementioned aspects refers to either the explicitly mentioned aspect, as depicted by its target, or the implicitly mentioned one. In sentences with one aspect, the aspect will have the same polarity as the sentence. Some sentences have multiple aspects, as shown in Example 1, where both *ambiance* and *service* are mentioned as explicit aspects. In such a case, the two aspects are each assigned a polarity, here positive and negative, respectively.

“The *ambiance* was nice, but *service* was bad.” (1)

Currently, research in the field of ABSA focuses on machine learning, due to its high accuracy [6]. To reduce the reliance on large training data sets, we

consider a hybrid option, combining machine learning with a knowledge-driven approach. Specifically, we add an ontology, which is an explicit specification of a conceptualization of a domain [1]. Because support vector machines (SVMs) have been shown to work well when applied to text [4], we add an ontology for the domain in question to an SVM. Thus, we combine a machine learning approach with external knowledge. The ontology can provide relational patterns between certain words and concepts that the machine would otherwise have to learn from the training data. This lowers the dependence on training data.

The structure of this paper is as follows. Due to space, we only describe the previous work this research is built upon in Sect. 2. Then, in Sect. 3 we describe the data that we operate on, followed by the specification of our proposed methodology in Sect. 4. In Sect. 5 we evaluate the proposed approach. Last, Sect. 6 gives our conclusion and states future work directions.

2 Related Work

The work in [7] also uses the combination of an SVM model with an ontology and is used as a starting point for our research. The results from [7] show that the ontology improves the performance of the SVM significantly. The proposed method also outperforms more basic versions of the same algorithm, indicating that ontology features stay relevant even at small amounts of training data. However, the gap between the performance with and without ontology does not widen as the training data decreases, implying that ontology features do not substantially reduce the required size of training data for aspect sentiment analysis. The authors indicate that this may be because of the use of external information in the form of sentiment dictionaries, which already reduces the need for training data. Furthermore, the number of domain-specific sentiment expressions included in the ontology used is limited, and improving this could potentially lead to a more pronounced increase in performance on aspect sentiment analysis.

Even though our approach is similar to the one in [7], the major difference is the ontology design and implementation. First, the ontology was kept relatively small to keep it manageable. By using a structure with improved readability, we are able to expand the ontology in a practical way. This allows us to include more concepts that could be relevant and potentially reach higher accuracy. With an increased number of domain-specific sentiment expressions, we expect the ontology to reduce the required number of training examples. Second, besides including individual words as lexical representation for concepts, we also include multi-word expressions that are frequently found in reviews. We thus expand the knowledge of the ontology, without increasing the number of concepts.

3 Specification of Data and Tasks

The data set used in this research is the restaurant review data from Task 5 of SemEval 2016 [5]. The training data consists of 350 reviews with 1992 individual sentences. The test data consists of 90 reviews with 676 individual sentences.

The data set is organized by review and by sentence, and each sentence is annotated with zero or more opinions. An opinion represents a combination of an aspect and the sentiment expressed on that aspect. The sentiment is either positive, negative, or neutral. In case an aspect is mentioned explicitly, the words that mention the aspect (i.e., the opinion target expression) are identified in the annotations as well. Additionally, each aspect is categorized into one of twelve given aspect categories.

The majority of the aspects are positive, while the neutral sentiment label only occurs in 3.9% and 5.1% of the instances, for the test and training data, respectively. Preliminary experiments show that predicting neutral besides positive and negative leads to a decline in accuracy. Therefore, we choose to treat neutral instances as positive when training the SVM and to only predict positive and negative sentiments when testing. Hence, all neutral instances are by definition incorrectly classified.

4 Method

In this section, we describe our proposed algorithm, including the structure of the ontology, the pre-processing of the data, the construction of the feature vector, and the selection of the meta-parameters of the SVM.

4.1 Ontology Design

The ontology can be divided into two parts, corresponding with two top-level classes: *Mention* and *Sentiment*. The class *Mention* has two subclasses, each with its own subclasses. The first subclass of *Mention* is *Entity*, with its domain-specific subclasses *Ambiance*, *Experience*, *Location*, *Person*, *Price*, *Restaurant*, *Service*, *Style Options*, and *Sustenance*. The domain-specific subclasses are annotated with the corresponding aspect categories, which is mostly a one-to-one mapping. Furthermore, within these classes, we group certain concepts together with the intention to improve precision. For example, *Warm Drink* and *Cold Drink* as subclasses of *Drink*.

The second subclass of *Mention* is *Property*. Its subclasses represent different properties of entities. These subclasses are constructed according to the *Entity* class and the sentiment they correspond to. Some properties have different sentiments when used in context with different concepts. To account for this, we create axioms where these concepts are connected to certain subclasses of *Entity* and then assigned a positive or negative polarity. One example is the Property *Cold*, as seen in Table 1. Cold is one of many properties for which the meaning depends on the context. The axioms within the ontology help to clarify the sentiment meaning of concepts such as *Big*, *Dry* and *Funny* amongst others.

Sentiment is the superclass of *Positive* and *Negative*. In turn, *Positive* and *Negative* are superclasses of classes such as *SustenancePositiveProperty* and *SustenanceNegativeProperty*, respectively, which are meant to link properties with

Table 1. Axioms involving the *Cold* class

Cold \sqcap Cold Drinks \sqsubseteq Positive
Cold \sqcap Warm Drinks \sqsubseteq Negative
Cold \sqcap Ambiance Inside \sqsubseteq Negative
Cold \sqcap Staff \sqsubseteq Negative

entities and their corresponding sentiment. This part of the ontology is particularly useful for sentences that have more than one aspect. The ontology allows us to find multiple aspects and corresponding properties, and mark them as positive or negative depending on their superclasses.

“The cheese was divine, however the room was very cramped.” (2)

This is illustrated in Example 2 above, where the aspect ‘food’ should have a positive sentiment whereas ‘ambiance’ should be negative. The ontology aids the SVM in this case, because “divine” refers to a subclass of *SustenancePositiveProperty* and “cramped” refers to a subclass of *AmbienceNegativeProperty*.

The analysis for the phrase “the cheese was divine” can be seen in Fig. 1. Note that each class is associated with multiple lexicalizations to account for different versions of spelling or for synonyms. Furthermore, the ontology is constructed manually to fit specifically with the domain of restaurant reviews, using information from the training data. To counteract possible over-fitting, the ontology is augmented with a list of commonly used concepts that are extracted from Yelp reviews of the best 10 and worst 10 rated restaurants in New York City.

4.2 Feature Vector Construction

Before constructing the feature vector, we pre-process the raw SemEval data using the Stanford CoreNLP package [2]. The text is split into tokens, which are individual words, punctuation, or multi-word expressions. Tokens are combined into sentences and tagged with Parts-of-Speech tags denoting their grammatical types. Then, words are lemmatized and syntactic relations between words in each sentence are determined.

In order to construct a feature vector for an aspect, we derive features from the sentence the aspect appears in. The feature vector consists of three independent parts that together form one vector for each aspect. The first part is a bag-of-words model which we refer to as B. The second part of the feature vector, which we refer to as S, is constructed similarly but with sentiment scores instead of binary values.

The sentiment scores are derived from two external sources: the sentiment tool in the Stanford CoreNLP package [8] and the NRC Sentiment list [3]. These give sentiment values that are decimals between -1 and 1 and between -5 and 5 , respectively. If both sources are used for a word, the feature is assigned the average of the two sentiment values. Otherwise, the feature is assigned the value

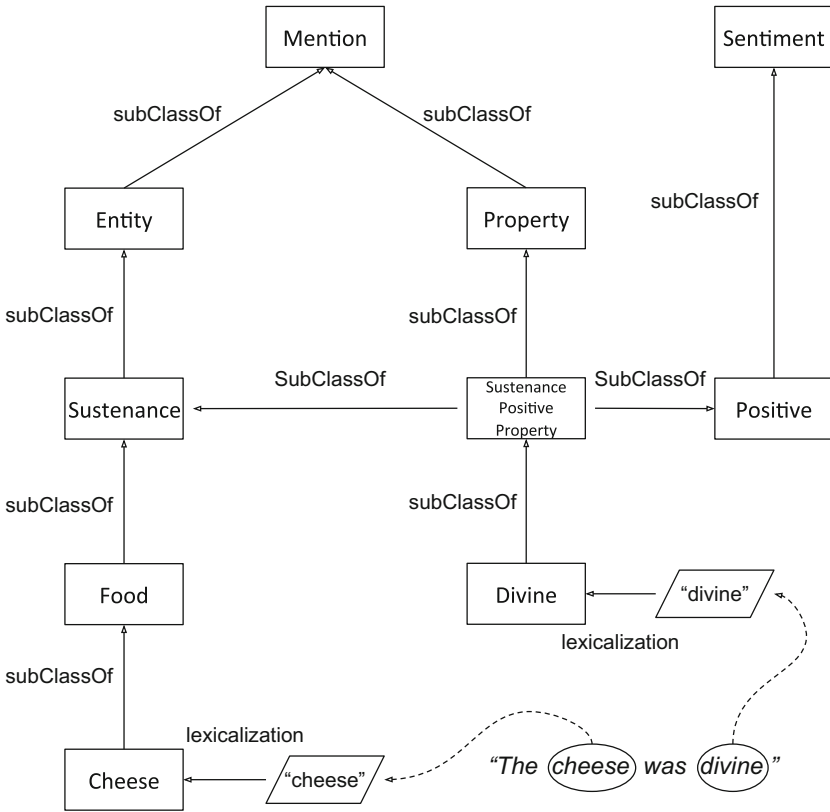


Fig. 1. Excerpt of the used ontology for the phrase “The cheese was divine”

extracted from the first source. In the third part, referred to as O, we create features using the ontology. We check each word in the *scope*, to identify whether it is linked to a concept in the ontology. If so, we check whether *Property* is a superclass of this concept, or if a superclass of this concept is annotated with the aspect category corresponding to the aspect. In this case, we assign a value of 1 to the feature representing this concept, as well as for each of its superclasses. In this way we construct features using only words that are descriptive, or that are directly related to the aspect category corresponding to the aspect. The *Positive* or *Negative* class could be a superclass at this point and be assigned a value of 1 too in that instance.

Next, we obtain all words in the same *scope* that are syntactically related to the current word, and check for each related word whether it is linked to a concept in the ontology. If *Property* is a superclass of the related concept and *Entity* is a superclass of the concept linked to the current word, we create a new intersection class using the two concepts. This allows us to use axioms in the ontology. For the new class and for each of its superclasses, we construct a

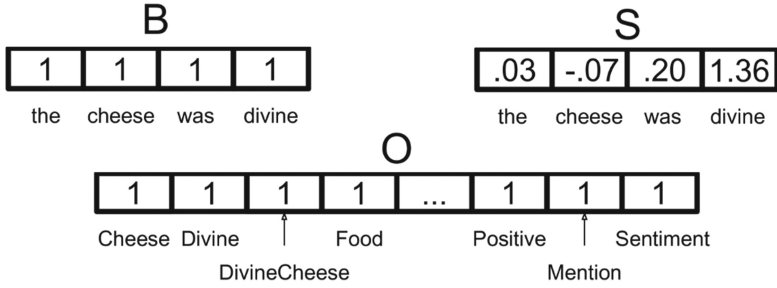


Fig. 2. Illustration of the feature vector. Note that the *Positive* feature is set to 1 since *Divine* is a subclass of *Positive* (cf. Fig. 1)

feature and we assign a value of 1 to these features. Only unique features are created and similarly, only unique new classes are created. In Fig. 2 we illustrate the feature vector with the construction of each part for the phrase “the cheese was divine”.

5 Evaluation

In this section, we evaluate the proposed methods and discuss the results, followed by a sensitivity analysis to determine the reliance on training data for each of the proposed methods.

To test the performance of the ontology for aspect sentiment classification, we evaluate several versions of the same algorithm. First, the standard bag-of-words model (B). Then, the features based on sentiment values (S). Next, the bag-of-words model combined with the binary ontology features (BO). Last, the sentiment features combined with binary ontology features (SO). The performances of the different versions are given in Table 2. The reported F_1 scores are averages from a randomized 10-fold cross-validation. The standard deviation is also reported, together with the p-values of the two-sided paired t-test to compare results statistically.

We find that replacing B with S does not result in a significant improvement in performance. At a 1% level, BO gives significantly better results than only B, and at a 20% level also significantly better results than S. SO instead gives results that are significantly better at a 1% level than both those with B and those with S. At a 20% level SO yields significantly better results than BO. These results imply that while the external dictionaries convey sentiment values, the sentiment score features do not significantly improve the performance for aspect sentiment classification, compared to the B features. However, the performance is significantly improved by using the O combined with either B or S features. This shows that the ontology has potential to improve the sentiment classification results, which is also in line with the out-of-sample F_1 scores.

To investigate whether including ontology features reduces the required size of training data, we analyze the sensitivity of the algorithm to data size by

Table 2. Performances of aspect sentiment classification

	avg. F_1	st. dev.	p-values of t-test			in-sample F_1 (training data)	out-of-sample F_1 (test data)
			B	S	BO		
B	0.7555	0.0412	-	-	-	0.8839	0.7835
S	0.7569	0.0305	0.8894	-	-	0.8448	0.7905
BO	0.7715	0.0405	0.0034	0.1018	-	0.8823	0.8079
SO	0.7813	0.0373	0.0039	0.0070	0.1364	0.8420	0.8068

training the SVM on a stepwise decreasing random part of the total available training data. The test data remains fixed, so the results can be compared for the different sizes of the training data. We perform this procedure for all four variants of the algorithm and for each variant and each size we obtain the average F_1 score over 5 runs. The results of the sensitivity analysis are shown in Fig. 3.

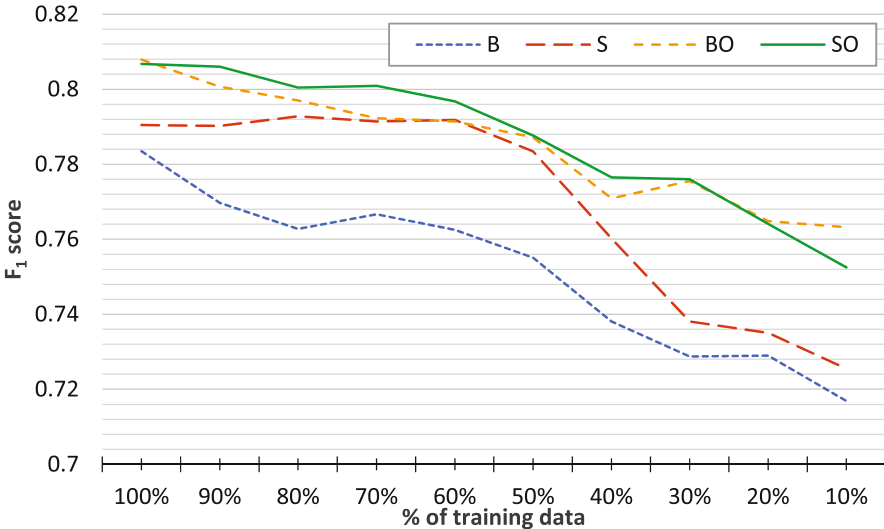


Fig. 3. The data size sensitivity (note that the y-axis does not start at 0 to improve readability)

One can see that B drops the fastest in performance with little training data. The F_1 scores of S initially remain stable as the proportion size drops. However, with less than 60% of the training data, performance drops substantially and the gap between BO and SO widens. The ontology-enhanced methods are clearly the most robust in this regard. Even at 10% of the original training data, the drop in performance is less than 6%. Moreover, the ontology-enhanced methods require less than 50% and 60%, respectively of the training data to achieve equal

performance with the bag-of-words features and the sentiment score features respectively at 100% of the training data. This implies that the ontology reduces reliance on training data.

6 Conclusion

In this paper we presented an ontology-enhanced hybrid approach for aspect-based sentiment analysis. The ontology is constructed specifically for the domain in question. It improves the performance of the SVM for classification of aspect sentiments and reduces the reliance on training data. This implies that while the external dictionaries already convey sentiment values, the added value of the ontology is substantial. Overall, the results lead us to conclude that the ontology is useful for aspect-based sentiment classification, both in combination with the standard bag-of-words and with sentiment scores from external dictionaries. In terms of future work, we suggest taking negations into consideration, as this could aid in correctly classifying certain aspects. Another option is to further augment the ontology in an automatic fashion or completely populate the ontology automatically. This should increase the coverage of the ontology as the ontology remains unused when no concept can be found in a sentence.

References

1. Gruber, T.R., et al.: A translation approach to portable ontology specifications. *Knowl. Acquisition* **5**(2), 199–220 (1993)
2. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S.J., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60. Association for Computational Linguistics (2014)
3. Mohammad, S.M., Turney, P.D.: Crowdsourcing a word-emotion association lexicon. *Comput. Intell.* **29**(3), 436–465 (2013)
4. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: Sentiment classification using machine learning techniques. In: *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp. 79–86. Association for Computational Linguistics (2002)
5. Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., De Clercq, O., et al.: Semeval-2016 task 5: aspect based sentiment analysis. In: *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016)*, pp. 19–30. Association for Computational Linguistics (2016)
6. Schouten, K., Frasinca, F.: Survey on aspect-level sentiment analysis. *IEEE Trans. Knowl. Data Eng.* **28**(3), 813–830 (2016)
7. Schouten, K., Frasinca, F., de Jong, F.: Ontology-enhanced aspect-based sentiment analysis. In: Cabot, J., Virgilio, R., Torlone, R. (eds.) *ICWE 2017*. LNCS, vol. 10360, pp. 302–320. Springer, Cham (2017). doi:[10.1007/978-3-319-60131-1_17](https://doi.org/10.1007/978-3-319-60131-1_17)
8. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.P.: Recursive deep models for semantic compositionality over a sentiment treebank. In: *Proceedings of the 2013 Conference on Empirical Methods on Natural Language Processing (EMNLP 2013)*, pp. 1631–1642. Association for Computational Linguistics (2013)