Mining Semantic Patterns for Sentiment Analysis of Product Reviews

Sang-Sang Tan(✉) and Jin-Cheon Na

Wee Kim Wee School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718, Singapore
{tans0348,tjcna}@ntu.edu.sg

Abstract. A central challenge in building sentiment classifiers using machine learning approach is the generation of discriminative features that allow sentiment to be implied. Researchers have made significant progress with various features such as n-grams, sentiment shifters, and lexicon features. However, the potential of semantics-based features in sentiment classification has not been fully explored. By integrating PropBank-based semantic parsing and class association rule (CAR) mining, this study aims to mine patterns of semantic labels from domain corpus for sentence-level sentiment analysis of product reviews. With the features generated from the semantic patterns, the F-score of the sentiment classifier was boosted to 82.31% at minimum confidence level of 0.75, which not only indicated a statistically significant improvement over the baseline classifier with unigram and negation features (F-score = 73.93%) but also surpassed the best performance obtained with other classifiers trained on generic lexicon features (F-score = 76.25%) and domain-specific lexicon features (F-score = 78.91%).

Keywords: Sentiment analysis · Semantic parsing · Pattern mining · Machine learning · Sentiment classification

1 Introduction

The preferences and decisions of internet users are increasingly influenced by peer opinions from online reviews, social networks, blogs, and other user-generated content on the web [1–3]. This growing reliance on user-generated content has triggered wide interest among stakeholders to capture these data and turn them into insightful information for various purposes including decision-making, target marketing, competitor analysis, etc. To this end, sentiment analysis has been extensively studied for gathering, extracting, and classifying users' sentiment expressed in textual content as a means to understand users' attitudes towards the targets of analysis. In the field of digital library, sentiment analysis can be employed in many ways—for example, as sentiment-based recommender system or sentiment-based searching and browsing features—for advanced retrieval of digital objects.

Sentiment analysis can be regarded as a subtask of natural language processing (NLP) that attempts to build, in machines, the abilities to imitate some cognitive abilities of human beings in interpreting human language for implying sentiment. Since a big

J. Kamps et al. (Eds.): TPDL 2017, LNCS 10450, pp. 382–393, 2017.
DOI: 10.1007/978-3-319-67008-9_30

part of the complexity of sentiment analysis lies in the processing of meanings in human language, we posit that the problem should be addressed in light of semantically sound approaches. As pointed out by researchers [4, 5], progress in sentiment analysis should be made towards "(a) enriching shallow representations with linguistically motivated, rich information, and (b) focusing different branches of research and combining resources and work forces to join hands with related work in NLP" [4] (pg. 66).

In the interest of exploring the potential of semantically motivated approach for sentiment classification, we propose the use of semantic parsing and pattern mining to derive a set of semantic patterns as features for training sentiment classifiers. Given a piece of text, the primary goal of semantic parsing is to detect the events described in the text and to identify the participants and their roles in the events, as a means to answer the question "Who did What to Whom, and How, When and Where?" [6] (pg. 1). Unlike syntactic parsing that focuses on grammatical relations between the components of a sentence, semantic parsing is an important step towards the understanding of meanings. In this study, we used a PropBank-based semantic parser [7] to transform sentences from word-level representations to their semantic representations. The semantically labelled data were then fed into class association rule (CAR) mining algorithm [8] to extract the semantic patterns that were regularly associated with the positive and negative sentiment. For instance, a semantic pattern like '*negation + buy.01 (predicate) + thing bought (argument)*' could be an evidence of negative sentiment. Since expressions that convey the similar meanings tend to share common semantic forms although they might differ in the use of words, word orders, and syntactic forms, such patterns would provide more generalizable features, resulting in a less sparse feature space for learning the classification model.

2 Related Work

Researchers have explored a wide range of features for sentiment classification. Word n-grams, which were popularized by Pang et al. [9], are among the most commonly used features that have produced acceptable performance on sentiment classification. Different types of data have their own unique characteristics that might be useful for inferring sentiment polarities. Tweets data, in particular, contain hashtags and emoticons that have shown to be closely related to the emotions expressed in tweet messages, making them well-suited as features for classifying sentiment in tweets [10, 11]. Other popular features for classifying sentiment include punctuation marks, part-of-speech tags, sentiment shifters like negators and other modifiers (e.g., very and barely), and stylistic features such as words per document and words per sentence.

As far as sentiment classification features are concerned, our stance is that semantics-based features are more likely to have a significant impact on classifiers because the interpretation of sentiment relies on the understanding of meanings. Generally speaking, any approach that takes into consideration the denotations and connotations of words or phrases falls into the category of semantics-based approach. One such approach that has been proven superior is the sentiment lexicon approach because an important indicator of sentiment in textual content is the use of sentiment terms that express likes and dislikes.

It is well established that combining lexical knowledge tends to show promising improvement in sentiment classification (e.g., [12–14]). The study conducted by Mohammad et al. [11] highlighted the importance of sentiment lexicons through their experiments that compared the effects of various classification features by removing one feature set at a time in the classification process. Among the feature sets compared in the experiments (including n-grams, negations, part-of-speech tags, emoticons, punctuation marks, hashtags, among others), sentiment lexicons were found to produce the most influential features for sentiment classification, to the extent that the removal of such features dropped the F-scores of the classification by more than 8.5%.

A considerable amount of literature has been published on the construction of sentiment lexicons. Earlier studies in this line of research focused on building general-purpose sentiment lexicons that only include terms of which the prior sentiment scores can be assumed with minimal uncertainty based on denotations and connotations of the terms (e.g., [15, 16]). Words like '*generosity*' and '*admirer*' can be considered as inherently positive whereas words like '*betrayal*' and '*nauseating*' can be considered as inherently negative. Generic sentiment lexicons have substantial application and research values because they are highly reusable. However, researchers have recognized that domain-specific terms are extremely crucial for interpreting opinions that require domain knowledge, especially in certain domains such as medical and chemical domains. Therefore, a wide range of studies (e.g., [10, 13]) has devoted the efforts to build domain-specific sentiment lexicons from domain corpora. One of the challenges in this line of research is that the coverage of sentiment lexicons is always a concern due to the richness of human language. Even when a corpus of enormous volume is used to learn a sentiment lexicon, it is almost certain that there will be some terms in the unseen data that the lexicon fails to cover. When no match exists for a term, the lexicon fails to provide useful information for classification. With respect to the coverage issue in lexicon-based features, the present study suggests that semantic patterns obtained from mining semantically parsed data would constitute a more generalizable feature set, in the sense that the semantic patterns are likely to match more cases in the unseen data.

Many NLP tasks that require semantic interpretation and processing could benefit from semantic parsing. One of the earliest studies that exploited semantic parsing for sentiment analysis is the work undertaken by Kim and Hovy [17]. They applied frame-based semantic parsing to identify the opinion holders and topics expressed in online news media text. The goal of their study was to find out which semantic roles could be used to identify opinion holders and topics. Another study that also used frame-like schemas for detection of opinion holders and topics was carried out by Gangemi et al. [18]. Their study used VerbNet (https://verbs.colorado.edu/~mpalmer/projects/verbnet.html) to find verb classes and thematic roles of verb arguments that indicated the presence of opinion holders and topics. As far as we know, the integration of semantic parsing and pattern mining for feature generation in sentiment analysis has not been fully explored. It would seem, therefore, that further investigations are desirable to find out whether semantic patterns mined from PropBank's verb-oriented semantic labels would have a positive impact on the performance of sentiment classifiers.

3 Method

3.1 Semantic Parsing

The development of most semantic parsing tools relies on human-annotated resources that provide annotations for verbs and their arguments. Recent years have seen increasingly rapid advances in semantic parsing due to the continuing efforts devoted to the development and maintenance of high-quality resources like PropBank [19]. PropBank has known to be an extremely influential resource in semantic parsing. It was created by the Proposition Bank project to provide predicate-argument information on top of Penn Treebank's syntactic layer.

In the present study, Punyakanok et al.'s [7] PropBank-based semantic parser was used for semantic processing. This semantic parser is able to identify the following arguments of verb predicates:

- Core arguments (*A0–A5* and *AA*) which are labelled based on the semantics of verb predicates specified in PropBank. The numbers and types of arguments vary across predicates. For instance, the arguments of predicate '*break*' in PropBank are: *A0* - *breaker*, *A1* - *thing broken*, *A2* - *instrument*, *A3* - *pieces*, and *A4* - *argument 1 broken away from what*.
- Thirteen adjunct arguments which are labelled as *AM-adj* where *adj* is the type of adjunct. The adjunct types are listed in Table 1.

Table 1. Thirteen adjunct types and their descriptions

Label	Description	Label	Description
AM-ADV	Adverbial modification	*AM-NEG*	Negation
AM-DIR	Direction	*AM-PNC*	Proper noun component
AM-DIS	Discourse marker	*AM-PRD*	Secondary predicate
AM-EXT	Extent	*AM-PRP*	Purpose
AM-LOC	Location	*AM-REC*	Reciprocal
AM-MNR	Manner	*AM-TMP*	Temporal
AM-MOD	General modification		

- Continued arguments which extend other arguments (core or adjunct arguments). This type of arguments is labelled as *C-arg* where *arg* is the label of the argument for which the continuity needs to be indicated. For instance, *C-A1* and *C-AM-TMP* indicate that the current arguments are continued from core argument *A1* and adjunct argument *AM-TMP* respectively.
- Referential arguments which represent relative pronouns. Referential arguments are labelled as *R-arg* where *arg* is the label of the core argument or adjunct argument to which the relative pronoun refers.

Figure 1 shows the semantic labels generated by the semantic parser for the sentence '*Also, pieces of this liner would break off very easily while using it.*'. The semantic parser has identified two predicates—'*break*' and '*use*'—in the sentence, with each predicate and its arguments forming a sequence of semantic labels that would become the inputs to the association rule mining algorithm.

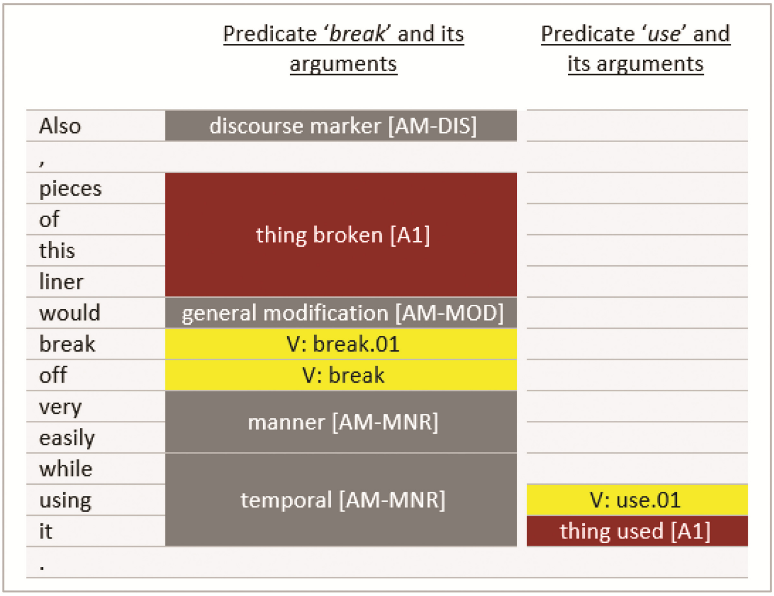

Fig. 1. Semantic labels generated by Punyakanok et al.'s [7] semantic parser for the sentence '*Also, pieces of this liner would break off very easily while using it.*'

3.2 Class Association Rule (CAR) Mining

Association rule mining is a machine learning method introduced by Agrawal et al. [20] for discovering interesting patterns of purchases in large-scale transaction data of supermarkets. The results of the analysis are a set of association rules or statements of regularities.

Let $T = \{t_1, t_2, t_3, \ldots, t_m\}$ be a set of transactions and $I = \{i_1, i_2, i_3, \ldots, i_n\}$ be a set of items such that each transaction t_k consists of one or more items from set I. An association rule discovered from the transaction data is a statement in the form $X \Rightarrow Y$, where X and Y are some items or itemsets that appear in the transactions, with $X \subset I$, $Y \subset I$, and $X \cap Y = \varnothing$. Various measures of significance and certainty can be applied to select the interesting rules, the most widely accepted measures being the support level and the confidence level. The goal of association rule mining is then to find the rules that satisfy the user-specified minimum support threshold (*minsup*) and minimum confidence threshold (*minconf*). The support level of a rule indicates how frequent the condition (i.e. the itemset X) appears in the transactions whereas the confidence level indicates how often the appearance of the condition actually leads to the consequence (i.e. the itemset Y). A confidence level of 1.0 indicates that Y always appears in the same transaction whenever X appears. The two measures are calculated as follows:

$$support(X) = \frac{|\{t \in T; X \subseteq t\}|}{|T|} \quad (1)$$

$$confidence(X \Rightarrow Y) = \frac{support(X \cup Y)}{support(X)} \quad (2)$$

The variant of association rule mining adopted in the present study is the one proposed by Liu et al. [8]. The original algorithm introduced by Agrawal et al. [20] does not impose any restriction to the targets or the consequences of rules. Liu et al.'s algorithm, on the other hand, integrated classification and association rule mining to acquire CARs for predetermined targets. Both algorithms have wide-ranging applications but the latter is better-suited for the purpose of pattern mining in the present study, i.e. to mine regular patterns of semantic labels that are commonly associated with the positive and negative sentiment. The resulted set of rules would be in the form X ⇒ positive and X ⇒ negative where X is the semantic pattern of which the appearance increases the probability of positive sentiment and negative sentiment respectively.

4 Experiments

4.1 Dataset

This study used the same cosmetic dataset we collected for our earlier study [21] that assessed the significance of multi-word sentiment terms in a lexicon-based, supervised sentiment classification task. The cosmetic dataset consists of 1100 positive sentences and 1100 negative sentences collected from MakeupAlley (www.makeupalley.com), a popular beauty website that provides consumers' reviews on beauty products. After the collected reviews were tokenized into sentences using Stanford Parser, sentences were labelled by two annotators; the first annotator labelled all the sentences whereas the second annotator labelled 300 sentences. Inter-rater reliability test using Cohen's kappa showed high agreement between the two annotators ($\kappa = .829$, $p < .0005$).

4.2 Feature Engineering, Classifier, and Validation

Evaluation of the semantic pattern features was performed by comparing the classification results produced by support vector machine (SVM) classifiers trained on different

Table 2. Feature sets for training the four groups of SVM classifiers (C1–C4)

Feature Sets	Classifiers			
	C1	C2	C3	C4
Baseline features (BL)	√	√	√	√
Generic lexicon features (GL)		√	√	√
Domain lexicon features (DL)			√	√
Semantic pattern features (SP)				√

feature sets, as shown in Table 2. Results were obtained using 10-fold cross validation, with each partition containing the same number of positive and negative sentences. SVMs with linear kernel and parameter C = 0.1 were used in all experiments.

The baseline features consist of binary features that indicated the presence or absence of unigrams and numeric features that indicated the number of negations in a sentence. Since lexicon features have shown promising results in sentiment classification [11], we compared the semantic pattern features to lexicon-related features generated using Hu and Liu's generic sentiment lexicon [16] and a set of domain-specific lexicons. In our earlier work [21], bigram domain lexicons were found to produce the best-performing lexicon features for sentiment classification. Therefore, we adopted the same approach in the present study to generate domain-specific lexicons consisting of bigram entries selected from nine of ten partitions (except the test partition) of the data using Pointwise Mutual Information (PMI) [10]. The following lexicon-related features were generated from the generic lexicon and the domain-specific lexicons:

- **Sum of Sentiment Scores**. For each sentence, sums of sentiment scores were obtained for all terms, the positive terms, and the negative terms. With Hu and Liu's sentiment lexicon, the positive terms were given prior scores of +1 whereas the negative terms were given prior scores of −1. The sentiment scores of terms in the domain-specific lexicons are positive and negative values calculated using PMI. For instance, the term '*stay power*' has a score of +1.435.
- **Count of Terms**. Three features were generated to indicate the number of positive, negative, and neutral terms in each sentence.

For each round of the 10-fold cross validation, semantic patterns were mined from nine of ten partitions, with the test partition excluded. The semantic patterns obtained from CAR mining formed a set of binary features, of which the values were determined based on the presence or absence of the patterns in each sentence. We considered the presence of a semantic pattern as the co-occurrence of all semantic labels of the pattern in a sentence, regardless of the order in which they occurred.

Considering the size of the dataset, the *minsup* threshold of CAR mining in the experiments was set to a very low value (0.0003), allowing any pattern that occurred more than twice to be considered as frequent itemset. Several thresholds of *minconf* in the range of 0.60 and 0.85 were tested in the experiments. Besides *minsup* and *minconf*, another crucial factor that might affect the outcomes of CAR mining is the form of semantic labels used in the mining process. The semantic patterns (X) that will be generated, the number of rules, as well as the support level and confidence level of each rule might vary according to the semantic labels used in the process. For instance, the labels for the core arguments of the verb predicates can take many forms:

- The more general form like *A0–A5* and *AA*. It is general in the sense that all predicates have the same set of labels for their core arguments.
- The specific form that ties an argument to its predicate, like [*A0: break.01*]. The predicate and its sense number (*break.01*) are appended to the core argument label (*A0*) so that the label will not be confused with the core argument labels of other predicates.

- The intermediate form that provides the definition of the argument per se (based on the definition in PropBank). This label is not strictly tied to the predicate because the same definition can describe the arguments of different but related predicates. For instance, the definition for argument 2 of predicate '*purchase*' is '*seller*', yet the same definition is also used to describe argument 2 of predicate '*buy*' as well as argument 1 of predicate '*sell*'.

Table 3 shows the different forms of semantic labels for the example given in Fig. 1. The labels of the core arguments are highlighted in bold. Different forms of labels would appear at different frequencies in the data, thus the choice of labels has a huge impact on the outcomes of the pattern mining algorithm. The intermediate form is neither too general nor too specific, so it was chosen as the approach in our experiments to generate inputs for CAR mining.

Table 3. Different forms of semantic labels for the sentence '*Also, pieces of this liner would break off very easily while using it.*'

Label Form	Semantic Labels	
General	t1	*[AM-DIS], [**A1**], [AM-MOD], [break.01], [break], [AM-MNR], [AM-TMP]*
	t2	*[use.01], [**A1**]*
Specific	t1	*[AM-DIS], [**A1: break.01**], [AM-MOD], [break.01], [break], [AM-MNR], [AM-TMP]*
	t2	*[use.01], [**A1: use.01**]*
Intermediate	t1	*[AM-DIS], [**thing broken**], [AM-MOD], [break.01], [break], [AM-MNR], [AM-TMP]*
	t2	*[use.01], [**thing used**]*

Semantic parsing provides the abstraction of meanings but the semantic labels might discard some information that could be quite distinctive in sentiment analysis. For example, '*really badly*' and '*really nicely*' can both be labelled as [*AM-MNR*] (manner), even though they are obviously related to different sentiment. To address this problem, for every word encompassed by a semantic label, the sum of polarity scores of the words was appended to the label so that each label is a two-tuple of [*semantic label:sum of polarity scores*]. For instance, the sum of polarity scores for the phrase '*really badly*' is -1, then the enhanced semantic label would be [*AM-MNR:* -1]. The polarity scores were obtained from Hu and Liu's generic sentiment lexicon [16]. Note that the sums of polarity scores are not shown in Table 3 but in the subsequent discussion, all semantic labels would be presented with the appended scores.

5 Results and Discussion

As can be seen from the results presented in Table 4, the sentiment classifiers have obviously benefited from the semantic pattern features. Results were obtained for five *minconf* thresholds. The size of the feature set (i.e. the number of semantic patterns) decreased as *minconf* increased. It has been observed that the performance of the sentiment classifiers increased marginally as *minconf* increased in the range of 0.60 and 0.75.

A reasonable explanation is that higher *minconf* values caused some spurious rules with low confidence levels to be excluded from the feature set. For example, these two ambiguous rules were excluded when *minconf* was set to 0.70:

- [*get.01:0*] + [*thing gotten:+ 1*] ⇒ negative (conf. level = 0.62)
- [*apply.02:0*] + [*applied to:0*] ⇒ negative (conf. level = 0.67)

Table 4. Classification results with p values indicated (* $p \leq .05$, ** $p \leq .01$, *** $p \leq .001$). BL: Baseline features. GL: Generic lexicon features. DL: Domain lexicon features. SP: Semantic pattern features. #SP: Number of semantic patterns.

Classifiers	Accuracy	Precision	Recall	F-Measure
C1: BL	73.95	74.04	73.94	73.93
C2: BL + GL	76.27	76.41	76.27	76.25
C3: BL + GL + DL	78.91 **	78.95 **	78.91 **	78.91 **
C4: BL + GL + DL + SP				
minconf = 0.60, #SP = 2913	81.59 ***	81.68 ***	81.60 ***	81.58 ***
minconf = 0.65, #SP = 2663	81.73 ***	81.82 ***	81.73 ***	81.71 ***
minconf = 0.70, #SP = 2136	82.14 ***	82.20 ***	82.15 ***	82.14 ***
minconf = 0.75, #SP = 2032	82.32 ***	82.38 ***	82.32 ***	82.31 ***
minconf = 0.85, #SP = 1579	81.91 ***	82.00 ***	81.92 ***	81.91 ***

Nevertheless, as demonstrated by the classifier's performance at *minconf* = 0.85, further increments of the *minconf* values would likely backfire due to the elimination of potentially useful rules. Based on this observation, a value around 0.75 seems to be a reasonable threshold for *minconf*. Despite the effects of the *minconf* values on the classifiers' performance, independent t-tests performed on the accuracy, precision, recall, and F-measure of the classification results showed that the performance of the C4 classifiers at all *minconf* thresholds was statistically significantly better than the performance of the baseline classifier (C1). Furthermore, with the inclusion of semantic pattern features, the C4 classifiers also outperformed classifier C2 and classifier C3, which were trained only on the baseline features and the lexicon-related features.

At higher confidence levels, the pattern mining algorithm was able to derive some interesting and meaningful semantic patterns, some of which have revealed implicitly expressed sentiment that is usually not detectable or distinguishable using other features. It is generally agreed that the classification of implicitly expressed sentiment is more challenging due to the absence of prominent evidence that signifies the manifestation of sentiment. As shown in the sentences that matched semantic pattern (1) and semantic pattern (2) in Table 5, without the use of sentiment-laden words, reviewers might express their sentiment implicitly by sharing experiences, describing scenarios, giving advice, and so forth. Unlike explicitly expressed sentiment that can be more easily detected from certain sentiment keywords, the classification of implicitly expressed sentiment often requires the overall meanings of the sentences to be taken into consideration. Sometimes, the presence of sentiment keywords might also mislead the sentiment classifier. For instance, the sentences that matched semantic pattern (3) both contain positive sentiment words like '*well*' and '*love*', and thus can easily be mistaken as positive sentences. However, the semantic patterns were able to recognize the two sentences as expressing negative sentiment. This finding suggests that semantic patterns might take us a step closer to perfecting sentiment analysis by improving the classification of implicit sentiment, which is expressed in a subtle manner.

Table 5. Semantic patterns with high confidence levels

(1) [*thing remaining:0*] + [*AM-NEG:0*] + [*stay.01:0*] ⇒ negative (conf. level = 1.00)
• *I have oily eyelids and the liner would not stay on after two to four hours.* • *It does not stay in the waterline at all!*
(2) [*AM-NEG:0*] + [*need.01:0*] + [*thing needed:0*] ⇒ positive (conf. level = 1.00)
• *I don't need to keep reapplying coats to get color.* • *The pots I have now will last me a very long time (as I said, you really do not need to use a lot).*
(3) [*want.01:0*] + [*thing wanted:+1*] ⇒ negative (conf. level = 0.93)
• *I really wanted this to work* ***well****.* • *Believe me, I wanted to* ***love*** *this product.*

6 Conclusion

In this paper, we explored the potential of semantic patterns in sentiment classification. The proposed method used PropBank-based semantic parsing and class association rule (CAR) mining to detect discriminative semantic patterns that constituted the features for building sentiment classifiers. Compared to other features, the semantic pattern features were able to improve the classifiers' performance to a greater extent. However, our experiments also revealed several issues in using PropBank-based semantic parsing to extract the sentiment-related semantic patterns. First, PropBank-based semantic parsing is verb-oriented so no semantic label was generated in the absence of verbs. Second, arguments of verb predicates might encompass large chunks of text, causing the details within the arguments to be completely omitted in the labelling process. As described earlier, this issue was partially solved in this study by appending the sum of polarity scores to each label. Such approach provided a pragmatic solution but might not be optimal. Despite the limitations, the proposed method has discovered some interesting semantic patterns that allowed subtly expressed sentiment to be recognized. This study thus suggests a potentially rewarding research direction for tackling the implicit sentiment problem that has known to be one of the highly challenging problems in sentiment analysis.

References

1. Constantinides, E., Fountain, S.J.: Web 2.0: Conceptual foundations and marketing issues. J. Direct Data Digit. Mark. Prac. **9**, 231–244 (2008). doi:10.1057/palgrave.dddmp.4350098
2. Ye, Q., Law, R., Gu, B., Chen, W.: The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Comput. Hum. Behav. **27**, 634–639 (2011). doi:10.1016/j.chb.2010.04.014
3. Zhang, Z., Ye, Q., Law, R., Li, Y.: The impact of e-word-of-mouth on the online popularity of restaurants: a comparison of consumer reviews and editor reviews. Int. J. Hosp. Manage. **29**, 694–700 (2010). doi:10.1016/j.ijhm.2010.02.002
4. Ruppenhofer, J.: Extending FrameNet for sentiment analysis. Revista Veredas **17**, 66–81 (2013)
5. Ruppenhofer, J., Rehbein, I.: Semantic frames as an anchor representation for sentiment analysis. In: 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis, pp. 104–109. ACL, Stroudsburg (2012)
6. Palmer, M., Gildea, D., Xue, N.: Semantic role labeling. Synth. Lect. Hum. Lang. Technol. **3**, 1–103 (2010). doi:10.2200/S00239ED1V01Y200912HLT006
7. Punyakanok, V., Roth, D., Yih, W.T.: The importance of syntactic parsing and inference in semantic role labeling. Comput. Linguist. **34**, 257–287 (2008). doi:10.1162/coli.2008.34.2.257
8. Liu, B., Hsu, W., Ma, Y.: Integrating classification and association rule mining. In: Fourth International Conference on Knowledge Discovery and Data Mining, pp. 80–86. AAAI Press, New York (1998)
9. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up?: sentiment classification using machine learning techniques. In: ACL-02 Conference on Empirical Methods in Natural Language Processing, pp. 79–86. ACL, Stroudsburg (2002). doi:10.3115/1118693.1118704
10. Kiritchenko, S., Zhu, X., Mohammad, S.M.: Sentiment analysis of short informal texts. J. Artif. Intell. Res. **50**, 723–762 (2014). doi:10.1613/jair.4272
11. Mohammad, S.M., Kiritchenko, S., Zhu, X.: NRC-Canada: building the state-of-the-art in sentiment analysis of tweets. In: Second Joint Conference on Lexical and Computational Semantics, pp. 321–327. ACL, Stroudsburg (2013)
12. Cambria, E., Olsher, D., Rajagopal, D.: SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. In: Twenty-Eighth AAAI Conference on Artificial Intelligence, pp. 1515–1521. AAAI Press, New York (2014)
13. Ding, X., Liu, B., Yu, P.S.: A holistic lexicon-based approach to opinion mining. In: International Conference on Web Search and Data Mining, pp. 231–240. ACM, New York (2008). doi:10.1145/1341531.1341561
14. Kaji, N., Kitsuregawa, M.: Building lexicon for sentiment analysis from massive collection of HTML documents. In: Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 1075–1083. ACL, Stroudsburg (2007)
15. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347–354. ACL, Stroudsburg (2005). doi:10.3115/1220575.1220619
16. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 168–177. ACM, New York (2004). doi:10.1145/1014052.1014073

17. Kim, S.M., Hovy, E.: Extracting opinions, opinion holders, and topics expressed in online news media text. In: Workshop on Sentiment and Subjectivity in Text, pp. 1–8. ACL, Stroudsburg (2006)
18. Gangemi, A., Presutti, V., Recupero, D.R.: Frame-based detection of opinion holders and topics: a model and a tool. IEEE Comput. Intell. Mag. **9**, 20–30 (2014). doi:10.1109/MCI.2013.2291688
19. Palmer, M., Gildea, D., Kingsbury, P.: The proposition bank: an annotated corpus of semantic roles. Comput. Linguist. **31**, 71–106 (2005). doi:10.1162/0891201053630264
20. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM, New York (1993). doi:10.1145/170036.170072
21. Tan, S.-S., Na, J.-C.: Expanding sentiment lexicon with multi-word terms for domain-specific sentiment analysis. In: Morishima, A., Rauber, A., Liew, C.L. (eds.) ICADL 2016. LNCS, vol. 10075, pp. 285–296. Springer, Cham (2016). doi:10.1007/978-3-319-49304-6_34