

An Unsupervised Aspect Detection Model for Sentiment Analysis of Reviews

Ayoub Bagheri¹, Mohamad Saraee², and Franciska de Jong³

¹ Intelligent Database, Data Mining and Bioinformatics Lab, Electrical and Computer Engineering Department, Isfahan University of Technology, Isfahan, Iran
a.bagheri@ec.iut.ac.ir

² School of Computing, Science and Engineering, University of Salford, Manchester, UK
m.saraee@salford.ac.uk

³ University of Twente, Human Media Interaction, P.O. Box 217, 7500 AE Enschede, The Netherlands
f.m.g.dejong@utwente.nl

Abstract. With the rapid growth of user-generated content on the internet, sentiment analysis of online reviews has become a hot research topic recently, but due to variety and wide range of products and services, the supervised and domain-specific models are often not practical. As the number of reviews expands, it is essential to develop an efficient sentiment analysis model that is capable of extracting product aspects and determining the sentiments for aspects. In this paper, we propose an unsupervised model for detecting aspects in reviews. In this model, first a generalized method is proposed to learn multi-word aspects. Second, a set of heuristic rules is employed to take into account the influence of an opinion word on detecting the aspect. Third a new metric based on mutual information and aspect frequency is proposed to score aspects with a new bootstrapping iterative algorithm. The presented bootstrapping algorithm works with an unsupervised seed set. Finally two pruning methods based on the relations between aspects in reviews are presented to remove incorrect aspects. The proposed model does not require labeled training data and can be applicable to other languages or domains. We demonstrate the effectiveness of our model on a collection of product reviews dataset, where it outperforms other techniques.

Keywords: sentiment analysis, opinion mining, aspect detection, review mining.

1 Introduction

In the past few years, with the rapid growth of user-generated content on the internet, sentiment analysis (or opinion mining) has attracted a great deal of attention from researchers of data mining and natural language processing. Sentiment analysis is a type of text analysis under the broad area of text mining and computational intelligence. Three fundamental problems in sentiment analysis are: aspect detection, opinion word detection and sentiment orientation identification [1-2].

Aspects are topics on which opinion are expressed. In the field of sentiment analysis, other names for aspect are: features, product features or opinion targets [1-5].

Aspects are important because without knowing them, the opinions expressed in a sentence or a review are of limited use. For example, in the review sentence “after using it, I found the size to be perfect for carrying in a pocket”, “size” is the aspect for which an opinion is expressed. Likewise aspect detection is critical to sentiment analysis, because its effectiveness dramatically affects the performance of opinion word detection and sentiment orientation identification. Therefore, in this study we concentrate on aspect detection for sentiment analysis.

Existing aspect detection methods can broadly be classified into two major approaches: supervised and unsupervised. Supervised aspect detection approaches require a set of pre-labeled training data. Although the supervised approaches can achieve reasonable effectiveness, building sufficient labeled data is often expensive and needs much human labor. Since unlabeled data are generally publicly available, it is desirable to develop a model that works with unlabeled data. Additionally due to variety and wide range of products and services being reviewed on the internet, supervised, domain-specific or language-dependent models are often not practical. Therefore the framework for the aspect detection must be robust and easily transferable between domains or languages.

In this paper, we present an unsupervised model which addresses the core tasks necessary to detect aspects from review sentences in a sentiment analysis system. In the proposed model we use a novel bootstrapping algorithm which needs an initial seed set of aspects. Our model requires no labeled training data or additional information, not even for the seed set. The model can easily be transform between domains or languages. In the remainder of this paper, detailed discussions of existing works on aspect detection will be given in section 2. Section 3 describes the proposed aspect detection model for sentiment analysis, including the overall process and specific designs. Subsequently we describe our empirical evaluation and discuss important experimental results in section 4. Finally we conclude with a summary and some future research directions in section 5.

2 Related Work

Several methods have been proposed, mainly in the context of product review mining [1-14]. The earliest attempt on aspect detection was based on the classic information extraction approach of using frequently occurring noun phrases presented by Hu and Liu [3]. Their work can be considered as the initiator work on aspect extraction from reviews. They use association rule mining (ARM) based on the Apriori algorithm to extract frequent itemsets as explicit product features, only in the form of noun phrases. Their approach works well in detecting aspects that are strongly associated with a single noun, but are less useful when aspects encompass many low-frequency terms. The proposed model in our study works well with low-frequency terms and uses more POS patterns to extract the candidates for aspect. Wei et al. [4] proposed a semantic-based product aspect extraction (SPE) method. Their approach begins with preprocessing task, and then employs the association rule mining to identify candidate product aspects. Afterward, on the basis of the list of positive and negative opinion

words, the semantic-based refinement step identifies and then removes from the set of frequent aspects possible non-product aspects and opinion-irrelevant product aspects. The SPE approach relies primarily on frequency- and semantic-based extraction for the aspect detection, but in our study we use frequency-based and inter-connection information between the aspects and give more importance to multi-word aspects and aspects with an opinion word in the review sentence. Somprasertsri and Lalitrojwong's [8] proposed a supervised model for aspect detection by combining lexical and syntactic features with a maximum entropy technique. They extracted the learning features from an annotated corpus. Their approach uses a maximum entropy classifier for extracting aspects and includes the postprocessing step to discover the remaining aspects in the reviews by matching the list of extracted aspects against each word in the reviews. We use Somprasertsri and Lalitrojwong's work for a comparison to our proposed model, because the model in our study is completely unsupervised.

Our work on aspect detection designed to be as unsupervised as possible, so as to make it transferable through different types of domains, as well as across languages. The motivation is to build a model to work on the characteristics of the words in reviews and interrelation information between them, and to take into account the influence of an opinion word on detecting the aspect.

3 Aspect Detection Model for Sentiment Analysis

Figure 1 gives an overview of the proposed model used for detecting aspects in sentiment analysis. Below, we discuss each of the functions in aspect detection model in turn.

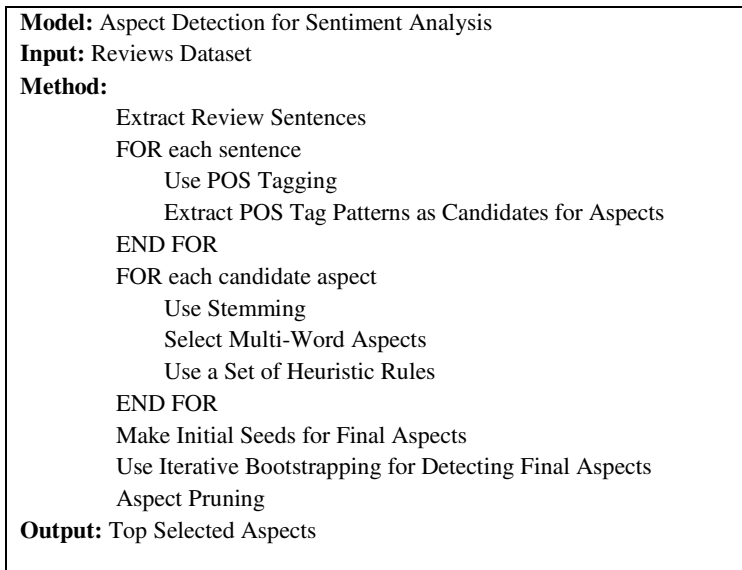


Fig. 1. The proposed model for aspect detection for sentiment analysis

3.1 Candidate Generation

In this paper we focus on five POS (Part-Of-Speech) tags: NN, JJ, DT, NNS and VBG, where they are the tags for nouns, adjectives, determiners, plural nouns and verb gerunds respectively [15]. Additionally stemming is used to select one single form of a word instead of different forms [16]. Based on the observation that aspects are nouns, we extract combination of noun phrases and adjectives from review sentences. We use several POS patterns introduced in table 1.

Table 1. Heuristic combination POS patterns for candidate generation

Description	Patterns
Combination of nouns	Unigram to four-gram of NN and NNS
Combination of nouns and adjectives	Bigram to four-gram of JJ, NN and NNS
Combination determiners and adjectives	Bigram of DT and JJ
Combination of nouns and verb gerunds (present participle)	Bigram to trigram of DT, NN, NNS and VBG

3.2 Multi-Word Aspects

In the review sentences, some aspects that people talk about have more than one single word, “battery life”, “signal quality” and “battery charging system” are examples. This step is to find useful multi-word aspects from the reviews. A multi-word aspect is represented by $a = a_1 a_2 \dots a_n$ where a_i represents a single-word contained in a , and n is the number of words contained in a . In this paper, we propose a generalized version of FLR method [17, 18] to rank the extracted multi-word aspects and select the importance ones. FLR is a word scoring method that uses internal structures and frequencies of candidates. The FLR for an aspect a is calculated as:

$$FLR(a) = f(a) * LR(a) \quad (1)$$

Where $f(a)$ is the frequency of aspect a , or in the other words it is number of the sentences in the corpus which a is appeared, and $LR(a)$ is LR score of aspect a which is defined as a geometric mean of the scores of subset single words as:

$$LR(a) = (lr(a_1) * lr(a_2) * \dots * lr(a_n))^{\frac{1}{n}} \quad (2)$$

The left score $l(a_i)$ of each word a_i of a target aspect is defined as the number of types of words appearing to the left of a_i , and the right score $r(a_i)$ is defined in the same manner. An LR score for single word a_i is defined as:

$$lr(a_i) = \sqrt{l(a_i) * r(a_i)} \quad (3)$$

The proposed generalization of the FLR method is on the definition of two parameters: $l(a_i)$ and $r(a_i)$. We change the definitions to give more importance to the

aspects with more containing words. In the new definition, in addition to the frequency we consider position of a_i in aspect a . For the score $l(a_i)$ of each word a_i of a target aspect, we not only consider a single word on the left of a_i , but we consider if there is more than one word on the left. We assign a weight for each position, that this weight is equal to one for the first word on the left, is two for the second word and so on. We define the $scorer(a_i)$ in the same manner. In addition, we apply the add-one smoothing to both of them to avoid the score being zero when a_i has no connected words.

3.3 Heuristic Rules

With finding the candidates, we need to move to the next level, aspect identification. For this matter we start with heuristic and experimentally extracted rules. Below, we present Rule #1 and Rule #2 for the aspect detection model.

Rule #1: Remove aspects which there are no opinion words with in a sentence.

Rule #2: Remove aspects that contain stop words.

3.4 Unsupervised Initial Seed Set

In this function we focus on selecting some aspects from the candidates as seed set information. We introduce a new metric named A-Score, which selects the seed set in an unsupervised manner. This metric is employed to learn a small list of top aspects with a complete precision.

3.5 A-Score Metric

Here we introduce a new metric, named A-score which uses inter-relation information between words to score them. We score each candidate aspect with A-score metric defined as:

$$A - Score(a) = f(a) * \sum_i \log_2 \left(\frac{f(a, b_i)}{f(a) * f(b_i)} * N + 1 \right) \quad (4)$$

Where a is the current aspect, $f(a)$ is the number of the sentences in the corpus which a is appeared, $f(a, b_i)$ is the frequency of co-occurrence of aspect a and b_i in each sentence. b_i is i th aspect in the list of seed aspects, and N is number of sentences in the corpus. The A-Score metric is based on mutual information between an aspect and a list of aspects, in addition it considers frequency of each aspect. We apply the add-one smoothing to the metric, so all co-frequencies be non-zero. This metric helps to extract more informative aspects and more co-related ones.

3.6 Iterative Bootstrapping Algorithm for Detecting Aspects

Iterative bootstrapping algorithm focuses on to learn final list of aspects from a small number of unsupervised seed set information. Bootstrapping can be viewed as an

iterative clustering technique which in each iteration, the most interesting and valuable candidate is chosen to adjust the current seed set. This technique continues until satisfying a stopping criterion like a predefined number of outputs. The important part in an iterative bootstrapping algorithm is how to measure the value score of each candidate in each iteration. The proposed iterative bootstrapping algorithm for detecting aspects is shown in figure 2. In this algorithm we use A-score metric to measure the value score of each candidate in each iteration.

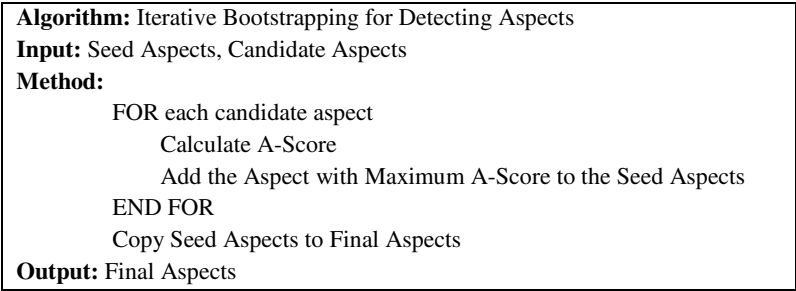


Fig. 2. The proposed iterative bootstrapping algorithm for detecting aspects

From figure 2, the task of the proposed iterative bootstrapping algorithm is to enlarge the initial seed set into a final list of aspects. In each iteration, the current version of the seed set and the list of candidate aspects are used to find the value score of A-Score metric for each candidate, resulting one more aspect for the seed set. Finally, the augmented seed set is the final aspect list and the output of the algorithm.

3.7 Aspect Pruning

After finalizing the list of aspects, there may exist redundant selected ones. For instances, “Suite” or “Free Speakerphone” are both redundant aspects, while “PC Suite” and “Speakerphone” are meaningful ones. Aspect pruning aims to remove these kinds of redundant aspects. For aspect pruning, we propose two kinds of pruning methods: Subset-Support Pruning and Superset-Support Pruning. We extracted these methods based on the experiment studies in our research.

Subset-Support Pruning

As we can see from table 1, two of the POS patterns are “JJ NN” and “JJ NN NN”. These patterns extract some useful and important aspects like “remote control” or “optical zoom”, but there are some redundant and meaningless aspects regarding to these patterns. Aspects like “free speakerphone” or “rental dvd player” are examples, while subset of them “speakerphone” or “dvd player” are useful aspects. This step checks multi-word aspects that start with an adjective (JJ POS pattern), and removes

those that are likely to be meaningless. In this step we remove the adjective part for aspects and then check a threshold if the second part is meaningful.

Superset-Support Pruning

In this step we remove redundant single word aspects. We filter single-word aspects which there is a superset ones of them. “Suite” or “life” are both examples of these redundant aspects which “PC Suite” or “battery life” are superset meaningful ones.

4 Experimental Results

In this section we discuss the experimental results for the proposed model and presented algorithms. We employed datasets of customer reviews for five products for our evaluation purpose (available at <http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html#datasets>). This dataset focus on electronic products: Apex AD2600 Progressive-scan DVD player, Canon G3, Creative Labs Nomad Jukebox Zen Xtra 40 GB, Nikon Coolpix 4300, and Nokia 6610. Table 2 shows the number of manually tagged product aspects and the number of reviews for each product in the dataset.

Table 2. Summary of customer review dataset

Dataset	Number of reviews	No. of manual aspects
Canon	45	100
Nikon	34	74
Nokia	41	109
Creative	95	180
Apex	99	110

4.1 Comparative Study

In our evaluation, after preprocessing and extracting the candidates, we score each multi-word aspect with the generalized FLR method and select those with the score higher than the average, and then we merge single-word and multi-word aspects in a list. Heuristic rules are then employed for the whole list of single and multi-word aspects to take into account the influence of an opinion word on detecting the aspect and remove useless aspects.

Finding an appropriate number of good seeds for bootstrapping algorithm is an important step. In our experiments we used A-score metric to extract automatically the seed set. We have experimented with different numbers of seeds (i.e., 5, 10, 15 and 20) for iterative bootstrapping, and found that the best number of the seeds is about 10 to 15. Therefore seeds were automatically chosen for iterative bootstrapping

algorithm, and the stopping criterion is defined when about 70 to 120 aspects have been learned. For the subset-support pruning method we set the threshold 0.5. In superset-support pruning step if an aspect has a frequency lower than three and its ratio to the superset aspect is less than experimentally threshold set one, it is pruned. Table 3 shows the experimental results of our model at three main steps described in section 3, Multi-word aspects and heuristic rules, Iterative bootstrapping with A-Score and Aspect pruning steps.

Table 2. Recall and precision at three main steps of the proposed model

Dataset	Multi-word aspects and heuristic rules	Iterative bootstrapping with A-Score	Aspect pruning
Precision			
Canon	26.7	75.0	83.1
Nikon	28.4	69.8	87.5
Nokia	23.9	73.5	79.0
Creative	14.8	79.2	88.9
Apex	19.3	78.8	82.0
Recall			
Canon	85.7	74.0	70.1
Nikon	82.4	72.5	68.6
Nokia	84.1	72.5	71.0
Creative	78.9	59.2	56.3
Apex	74.6	65.1	65.1

Table 3 gives all the precision and recall results at the main steps of the proposed model. In this table, column 1 lists each product. Each column gives the precision and recall for each product. Column 2 uses extracted single-word aspects and selected multi-word aspects based on generalized FLR approach and employing heuristic rules for each product. The results indicate that extracted aspects contain a lot of errors. Using this step alone gives poor results in precision. Column 3 shows the corresponding results after employing Iterative bootstrapping algorithm with A-Score metric. We can see that the precision is improved significantly by this step but the recall drops. Column 4 gives the results after pruning methods are performed. The results demonstrate the effectiveness of the pruning methods. The precision is improved dramatically, but the recall drops a few percent.

We evaluate the effectiveness of the proposed model compared with the benchmarked results by [4]. Wei et al. proposed a semantic-based product aspect extraction (SPE) method and compared the results of the SPE with the association rule mining approach (ARM) given in [3]. The SPE technique exploits a list of positive and negative adjectives defined in the General Inquirer to recognize opinion words semantically and subsequently extract product aspects expressed in customer reviews.

Table 3. Experiment results of comparative study

Product	ARM		SPE		Proposed model	
	Precision	Recall	Precision	Recall	Precision	Recall
Canon	51.1	63.0	48.7	75.0	83.1	70.1
Nikon	51.0	67.6	47.4	75.7	87.5	68.6
Nokia	49.5	57.8	56.5	72.5	79.0	71.0
Creative	37.0	56.1	44.0	65.0	88.9	56.3
Apex	51.0	60.0	52.4	70.0	82.0	65.1
Macro avg.	47.9	60.9	49.8	71.6	84.1	66.2
Micro avg.	46.1	59.9	48.6	70.5	83.6	66.2

Table 4 shows the experimental results of our model in comparison with SPE and ARM techniques (the values in this Table for ARM and SPE come from the results in [4]). Both the ARM and SPE techniques employ a minimum support threshold set at 1% in the frequent aspect identification step for finding aspects according to the association rule mining.

From Table 4, the macro-averaged precision and recall of the existing ARM technique are 47.9% and 60.9% respectively, whereas the macro-averaged for precision and recall of the SPE technique are 49.8% and 71.6% respectively. Thus the effectiveness of SPE is better than that of the ARM technique, recording improvements in macro-averaged precision and recall. However, our proposed model outperforms both benchmark techniques in precision, achieving a macro-averaged precision of 84.1%. Specifically, macro-averaged precision obtained by the proposed model is 34.3% and 36.2% higher than those reached by the existing ARM technique and SPE, respectively. The proposed model reaches to a macro-averaged recall at 66.2%, where improves the ARM by 5.3%, but it is about 5.4% less than SPE approach. When considering the micro average measures, we observe similar results to those we obtained by using macro average measures.

It is notable that we observe a more substantial improvement in precision than in recall with our proposed model and techniques. Observing from Table 4, our model makes significant improvements over others in all the datasets in precision, but in recall SPE has better performance. For example, our model records 36.2% and 34.3% improvements in terms of macro-averaged precision over the ARM and SPE techniques respectively, and 37.5% and 35% improvements in terms of micro-averaged precision. However, the proposed model achieves an averagely higher recall than the ARM technique but a slightly lower recall than the SPE technique. One reason is that for the iterative bootstrapping algorithm we limit number of output aspects between 70 and 120 aspects, therefore the precision for the output will be better than the recall, another reason for low recall is that of our model only works in detecting explicit aspects from review sentences.

Figure 3 shows the F-score measures of different approaches using different product datasets. In all five datasets, our model achieves the highest F-score. This

indicates our unsupervised model is effective in extracting correct aspects. We can thus draw the conclusion that our model is superior to the existing techniques, and can be used in practical settings, in particular those where high precision is required.

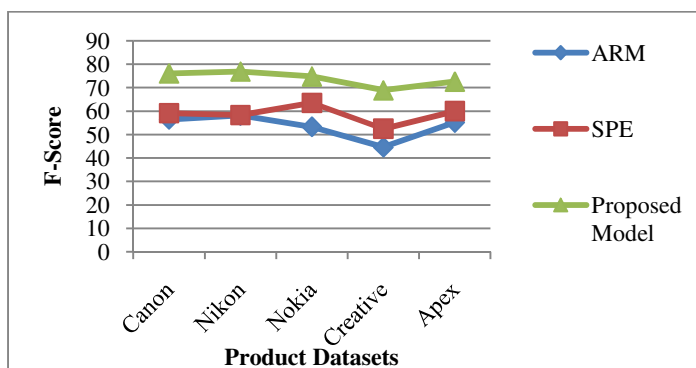


Fig. 3. F-scores of ARM, SPE, and the Proposed model for each dataset

This comparative evaluation suggests that the proposed model, which involves frequency-based and inter-connection information between the aspects and gives more importance to multi-word aspects and uses the influence of an opinion word in the review sentence, attains better effectiveness for product aspect extraction. The existing ARM technique depends on the frequencies of nouns or noun phrases for the aspect extraction, and SPE relies primarily on frequency- and semantic-based extraction of noun phrases for the aspect detection. For Example, our model is effective in detecting aspects such as “digital camera” or “battery charging system”, which both ARM and SPE are failed on extraction of these non-noun phrases. Additionally, we can tune the parameters in our model to extracts aspects with less or more words, for example aspect “canon power shot g3” can be finding by the model. Finally, the results show using a completely unsupervised approach for aspect detection in sentiment analysis could achieve promising performances.

As mentioned before, the proposed model is an unsupervised domain-independent model. We therefore empirically investigate the performance of using a supervised technique for aspect detection in comparison to the proposed model. We employ results of a supervised technique from Somprasertsri and Lalitrojwong’s work [8]. They proposed an approach for aspect detection by combining lexical and syntactic features with a maximum entropy model. Their approach uses the same data set collection of product reviews we experimented on. They extract the learning features from the annotated corpus of Canon G3 and Creative Labs Nomad Jukebox Zen Xtra 40 GB from customer review dataset. In their work, the set of data was split into a training set of 80% and a testing set of 20%. They employed the Maxent version 2.4.0 as the classification tool. Table 5 shows the micro-averaged precision, micro-averaged recall and micro-averaged F-score of their system output in comparison to our proposed model for the Canon and Creative datasets.

Table 4. Micro-averaged precision, recall and F-score for supervised maximum entropy and our unsupervised model

	Precision	Recall	F-score
Maximum entropy model	71.6	69.1	70.3
Proposed model	85.5	63.5	72.9

Table 5 shows that for the proposed model, the precision is improved dramatically by 13.9%, the recall is decreased by 5.6% and the F-score is increased by 2.6%. Therefore our proposed model and presented algorithms outperforms the Somprasertsri and Lalitrojwong's model. The significant difference between our model and theirs is that they use a fully supervised structure for aspect detection, but our proposed model is completely unsupervised and domain independent. Although in most applications the supervised techniques can achieve reasonable effectiveness, but preparing training dataset is time consuming and the effectiveness of the supervised techniques greatly depends on the representativeness of the training data. In contrast, unsupervised models automatically extract product aspects from customer reviews without involving training data. Moreover, the unsupervised models seem to be more flexible than the supervised ones for environments in which various and frequently expanding products get discussed in customer reviews.

5 Conclusions

This paper proposed a model for the task of identifying aspects in reviews. This model is able to deal with two major bottlenecks, domain dependency and the need for labeled data. We proposed a number of techniques for mining aspects from reviews. We used the inter-relation information between words in a review and the influence of an opinion word on detecting an aspect. Our experimental results indicate that our model is quite effective in performing the task. In our future work, we plan to further improve and refine our model. We plan to employ clustering methods in conjunction with the model to extract implicit and explicit aspects together to summarize output based on the opinions that have been expressed on them.

Acknowledgments. We would like to thank Professor Dr. Dirk Heylen and his group for giving us the opportunity to work with the Human Media Interaction (HMI) group from university of Twente.

References

1. Qiu, G., Liu, B., Bu, J., Chen, C.: Opinion word expansion and target extraction through double propagation. *Computational Linguistics* 37(1), 9–27 (2011)
2. Thet, T.T., Na, J.C., Khoo, C.S.G.: Aspect-Based Sentiment Analysis of Movie Reviews on Discussion Boards. *Journal of Information Science* 36(6), 823–848 (2010)

3. Hu, M., Liu, B.: Mining opinion features in customer reviews. In: American Association for Artificial Intelligence (AAAI) Conference, pp. 755–760 (2004)
4. Wei, C.P., Chen, Y.M., Yang, C.S., Yang, C.C.: Understanding what concerns consumers: A semantic approach to product feature extraction from consumer reviews. *Information Systems and E-Business Management* 8(2), 149–167 (2010)
5. Brody, S., Elhadad, N.: An unsupervised aspect-sentiment model for online reviews. In: 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, pp. 804–812 (2010)
6. Popescu, A., Etzioni, O.: Extracting product features and opinions from reviews. In: Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Vancouver, pp. 339–346 (2005)
7. Yi, J., Nasukawa, T., Bunesco, R., Niblack, W.: Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In: 3rd IEEE International Conference on Data Mining (ICDM 2003), Melbourne, FL, pp. 427–434 (2003)
8. Somprasertsri, G., Lalitrojwong, P.: Automatic product feature extraction from online product reviews using maximum entropy with lexical and syntactic features. In: IEEE International Conference on Information Reuse and Integration, pp. 250–255 (2008)
9. Zhu, J., Wang, H., Zhu, M., Tsou, B.K.: Aspect-based opinion polling from customer reviews. *IEEE Transactions on Affective Computing* 2(1), 37–49 (2011)
10. Zhai, Z., Liu, B., Xu, H., Jia, P.: Constrained LDA for Grouping Product Features in Opinion Mining. In: Huang, J.Z., Cao, L., Srivastava, J. (eds.) PAKDD 2011, Part I. LNCS, vol. 6634, pp. 448–459. Springer, Heidelberg (2011)
11. Su, Q., Xu, X., Guo, H., Guo, Z., Wu, X., Zhang, X., Su, Z.: Hidden sentiment association in chinese web opinion mining. In: 17th International Conference on World Wide Web, Beijing, China, pp. 959–968 (2008)
12. Moghaddam, S., Ester, M.: ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews. In: 34th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 665–674. ACM (2011)
13. Fu, X., Liu, G., Guo, Y., Wang, Z.: Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowledge-Based Systems* 37, 186–195 (2013)
14. Lin, C., He, Y., Everson, R., Ruger, S.: Weakly supervised joint sentiment-topic detection from text. *IEEE Transaction on Knowledge & Data Engineering* 24(6), 1134–1145 (2012)
15. Marcus, M., Santorini, B., Marcinkiewicz, M.A.: Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics* 19(2), 313–330 (1993)
16. Toutanova, K., Klein, D., Manning, C., Singer, Y.: Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network. In: Proceedings of HLT-NAACL, pp. 252–259 (2003)
17. Nakagawa, H., Mori, T.: Automatic Term Recognition based on Statistics of Compound Nouns and their Components. *Terminology* 9(2), 201–219 (2003)
18. Yoshida, M., Nakagawa, H.: Automatic Term Extraction Based on Perplexity of Compound Words. In: Dale, R., Wong, K.-F., Su, J., Kwong, O.Y. (eds.) IJCNLP 2005. LNCS (LNAI), vol. 3651, pp. 269–279. Springer, Heidelberg (2005)
19. Yang, Y.: An evaluation of statistical approaches to text categorization. *Inf. Retr.* 1(1-2), 69–90 (1999)