

Extracting Product Features from Online Consumer Reviews

Completed Research Paper

Yin Kang

University of Maryland Baltimore County
Ky6@umbc.edu

Lina Zhou

University of Maryland Baltimore County
zhoul@umbc.edu

ABSTRACT

The exponential growth of user-generated content in online environment calls for techniques that can help to make sense of the content. Despite of a host of research on online consumer reviews, there is still a great demand for research to improve the techniques for feature extraction. To this end, we proposed extraction methods based on detailed categorization of review features. By taking into account of the characteristics and patterns of different types of features, the proposed methods not only identify new features but also filter irrelevant features. The results of an experiment demonstrate that our proposed methods outperform the state-of-the-art techniques.

Keywords

Features extraction, natural language processing, online consumer review

INTRODUCTION

As the Internet becomes more popular and powerful than ever before, online consumer reviews for a wide variety of products and services are playing an increasingly important role in e-commerce. Before deciding to buy a product, consumers tend to consult others' opinions or some recommendations on it. Online consumer reviews serve as a good supplement to expert reviews and product descriptions. They are used to reduce purchasing uncertainty (Dellarocas, 2003) and help consumers to make better decisions. Researchers have also found that consumer reviews have a positive effect on product sales (Chevalier and Mayzlin, 2006). As the volume of online reviews rapidly increases and the quality of reviews vary widely, many e-commerce websites have implemented the function that allows consumers to vote for the helpfulness of a review. However, these helpfulness votes could still be problematic for a variety of reasons. For instance, consumers may seek different types of information from an online review; newer reviews tend to receive fewer votes or no votes; and some spam reviews may have manipulated the helpfulness votes.

Extracting product features from online product reviews is fundamental to online review mining. There has been an extensive amount of work on product feature extraction, using linguistics-based methods, statistical methods or machine learning methods (e.g., Hu and Liu, 2004, Kobayashi, Inui and Matsumoto 2007, Popescu and Etzioni, 2007, Qiu, Liu, Bu and Chen 2009, Wong and Lam 2005, 2008). However, most of these researches are focused on explicit subjective product features. For example, in sentence, "The picture of this camera is amazing", "picture" is a feature that appears in the sentence and reviewer explicitly express opinion on it. This research aims to extract both explicit subjective and objective features from online consumer reviewers.

Despite of the extensive research on developing analytical techniques for online consumer reviews, this research makes several new contributions. First, we developed new criteria for classifying product features, which will guide our development of extraction methods. Second, we identified new patterns of product features from the sentence level such as prescriptive statements, comparison patterns and special words or structures. Third, we incorporated information from sources such as WordNet, sentence structure, and document frequency to prune irrelevant terms in feature extraction. The empirical results show that our proposed methods outperform the state-of-the-art techniques for feature extraction.

The remainder of this paper proceeds as follows. In the next section, we review prior literature as related to our research setting. After that, we propose a new method for feature extraction followed by our data collection methods and experiment design. In the following section, we report our empirical results. We discuss and conclude our findings in the last section.

RELATED WORK

We drew related work from two areas: ecommerce and text mining. Researchers in ecommerce focus primarily on how to apply extracted features to support other applications. While researchers in text mining primarily deal with the identification of product features and how to improve the performance of feature extraction.

Since our data set comes from online reviews and our work aims to generate more specific or personalized reviews by incorporating specific product features, it is necessary to introduce how product feature extraction related to ecommerce. Online consumer reviews can be defined as peer-generated product evaluations posted on company or third party websites (Mudambi and Schuff, 2010). There are two basic components in online customer review: rating (usually ranging from 1 to 5 stars) and open-ended customer-generated comments. An online review is regarded as an assessment of a single consumer's perceived quality of a product, which provides not only a way in which consumers can share their opinions, but also a valuable resource for potential customers to make purchase decisions (Liu, Huang, An and Yu 2008). Several researches in this area focus on how to predict the helpfulness of online reviews in term of how many product features it contains (Kim, Pantel, Chklovski and Pennacchiotti, 2006; Liu, Cao, Lin, Huang and Zhou, 2007). While other researchers aim to build a helpful recommender system based on review data (O'Mahony and Smyth, 2009) and make use of product features to analyze the product price (Archak, Ghose, G.Ipeirotis 2011).

In the community of text mining, researchers focus on how to extract product features effectively and efficiently. In a pioneering work, Hu and Liu (2004) proposed a technique for product feature extraction using association rule mining based on the assumption that people often use the same words when they express their opinions. Popescu and Etzioni (2007) developed a similar algorithm to determine whether a noun/noun phrase is a feature by computing the point wise mutual information (PMI) score between the phrase and class specific discriminators. Later, to achieve a better performance, different researchers use different methods including machine learning, linguistic rules. Ghani, Probst, Liu, Crema and Fano (2006) proposed a method to extract attribute and value pairs from textual product descriptions. They viewed products as sets of attribute value pairs rather than as atomic entities. The method learns these attributes by applying supervised and semi-supervised learning techniques to the product descriptions available from retailers' web sites. Lee and Bradlow (2007) presented a method to support conjoint study design by automatically eliciting an initial set of attributes and levels from online customer reviews. Raju, Shiashtla and Varma (2009) proposed a novel solution to extracting features from a set of product descriptions. They classified product attributes into tangible and intangible (explicit and implicit) categories. They first constructed a graph from the text using word co-occurrence statistics, then computed word clusters and extract attributes from these clusters using graph based methods. Hai, Chang and Kim (2011) proposed a novel two-phase co-occurrence association rule mining approach to identifying implicit features. Double Propagation (Qiu, Liu, Bu and Chen, 2009) is a state-of-the-art unsupervised technique for extracting features. It mainly extracts noun features, and works well for medium-size corpora. However, for large corpora, this method can introduce a great deal of noise (low precision), and for small corpora, it can miss important features.

Most existing methods have their own limitations, Hu and Liu (2004)'s work doesn't work well on infrequent features; Double propagation works well for medium-size corpora. However, for large and small corpora, it can result in low precision and low recall (Zhang, Liu, Lim and O'Brien-Strain 2011), it has been shown in (Qiu et al., 2009) that many feature and opinion word pairs have long range dependencies. To deal with above problems, we design a new set of methods to extract features.

PROPOSED METHODS

In this section, we propose methods for extracting product features from online consumer reviews. To compare with our methods, we first review several existing classification criteria for product features, and then introduce our categorization criteria and propose extraction methods based on the type of the features.

Feature classification

One popular classification scheme for product features (Hu and Liu 2004) contains two types of features: explicit and implicit features. Explicit features are explicitly mentioned in review sentences, while implicit features refer to those features that do not directly appear in the reviews.

The second type of classification criteria is based on how features are expressed. Extant extraction methods are mostly based on syntactic analysis. Hu and Liu (2004) used adjectives to find near nouns which are candidates of features that reviewers commented on. Qiu et al (2009) proposed a domain sentiment word extraction approach based on the propagation of both known sentiment lexicon and extracted product features. The method exploits dependency relationship to capture the association both between features and sentiment words and between sentiment words and features themselves. Thus, the key task of these methods is to summarize how people express their opinions, what adjectives they usually use, what sentence structure they use. Experimental results show that these syntactic based methods are more effective and have better recall and precision than other alternative methods.

We propose a scheme that is based on but extends the previous work on feature objectivity. First, as mentioned above, double propagation mainly focuses on subjective statements that express opinions on some features, while ignoring objective statements that only describe the characteristic of some features without expressing any positive or negative opinions. Second, double propagation only considers dependency structures, missing other subjective patterns, in our study we incorporate

comparative structures and some other simple patterns. Third, based on our observation, there are some unique features that are either seldomly mentioned (e.g., ergonomical) or have their own structures, such as brand names or special models. Therefore, our proposed classification scheme is based on subjectivity of features, which contains subjective features and objective features. Subjective features are those features appear in subjective statements where reviewers express their opinions explicitly, objective features appear in objective statements which don't evolve reviewers' opinions.

Subjective statements: “This camera is awesome, with great lens ...” “I like the design of this phone, it’s amazing ...” “With this camera, I really haven’t taken a bad picture” “They didn’t correct the design flaw ...”

Objective statements: “This phone comes with a rechargeable battery ...” “It includes a rechargeable battery, a 4mb memory card ...” “This phone has a two-year warranty ...” “I got a while phone, and there are so many different colors to choose ...”

Extraction methods

Figure 1 shows an overall process flow of our extraction method. It consists of several key processes: preprocessing, feature extraction and pruning. The system takes product reviews as the input and produce product features as the output.

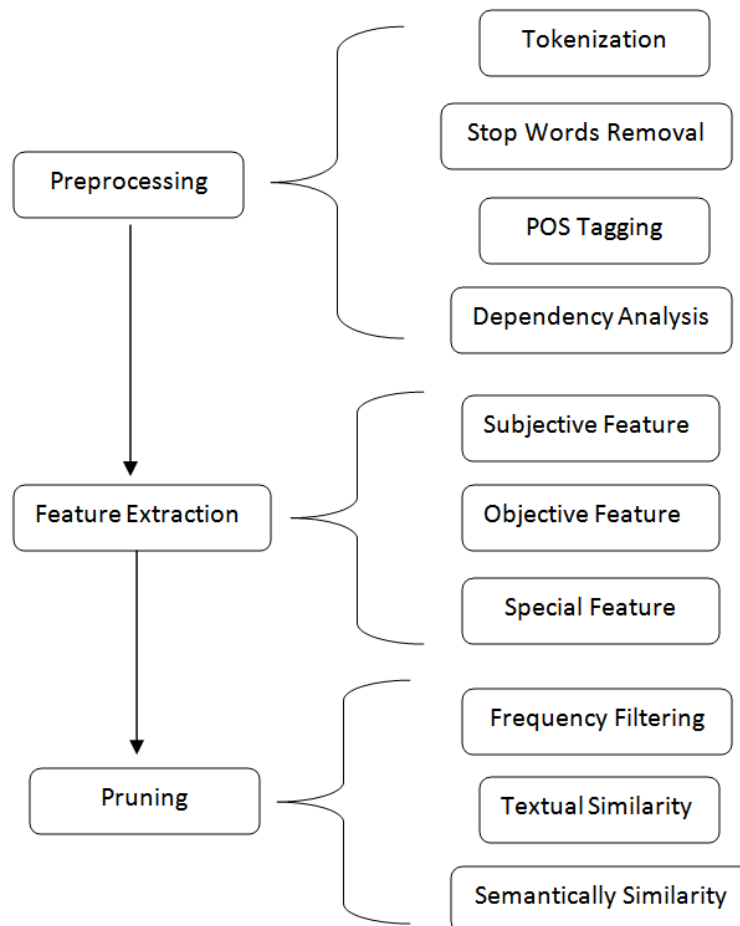


Figure 1. Main Components of the Feature Extraction Methods

Features are generally expressed as nouns or noun phrases with certain grammar patterns in online reviews (Liu, Hu and Cheng 2005). Typically, a noun or noun phrase acting as the object or the subject of a verb is a potential feature. To design extraction methods, we made the following assumption: different type of features has different grammatical patterns, but within one type, different features may share the same sentence structures. We will introduce extracting methods for different types of features separately.

Subjective features

Subjectivity refers to aspects of language used to express opinions (Wiebe, Wilson, Bruce, Bell and Martin, 2004). Opinion that appears in text comes in two flavors: explicit where the subjective sentence directly expresses an opinion (“It’s a beautiful day”), and implicit where the text implies an opinion (“The earphone broke in two days”) (Hu and Liu, 2006). Most of the work

done so far focuses on the explicit sentiment (Hu and Liu, 2004, 2006; Kim et al 2006; Liu et al 2007; Popescu and Etzioni, 2007; Wong and Lam 2005, 2008), since it is easier to analyze. Hu and Liu (2004, 2006) proposed a grammar pattern based method. They found that infrequent features and frequent features sometimes share the same expression pattern. For instance, both pictures and software are features in the following two sentences: “The pictures are absolutely amazing”, “The software that comes with it is amazing”. Despite that picture is a frequent feature and software is an infrequent one, both features are expressed with the same grammatical patterns. Based on the observation, the authors extracted the noun words or phrases from sentences that contain certain adjectives. To improve the performance of the naïve method, Hu and Liu (2006) proposed a supervised sequential rule mining method, where they abstracted specific adjectives into general patterns, and also took order into consideration. However, this method is mainly focused on phrase reviews rather than complete sentence reviews. In addition, subjectivity sentences can be further divided based on its polarity – positive and negative, and a distinction can be further made between the polarity of sentiment and of its strength. In view that online consumer reviews are mostly in form of complete sentences, we intend to focus on explicit subjective sentences, and apply grammatical rule based method to extract subjective features. We further enhanced the method with a new pruning method to improve its precision.

Double propagation is a well-known method, which essentially extends grammar-based method by using dependency parser to detect the opinion word and features. It makes use of dependency grammars to generate eight rules based on dependency relations between opinion word and target word, and repeatedly extract opinion word and target word alternatively. In our work, to extract subjective features, we first applied double propagation, and then used comparison patterns to extracted more features.

Besides dependency, comparison is also an important pattern that expresses subjectivity in corpora. For example, in the following expression, “battery life is over 4.5 hours, compared to about 2.5 for G2”, there is a comparison between two different phones. Since “over 4.5 hours” is not a typical subjective structure, so this feature can not be extracted by double propagation, while apparently it is a positive opinion about battery. This kind of pattern actually is very useful for feature extraction, and is widely used to express opinions. Thus, we incorporated the comparison patterns into the extraction of product features in our proposed method. Jindal and Liu (2006a,b) proposed a data mining method called labeled sequential rule to identify comparative sentences and generate comparative relations. Because their method require a lot of labeling work, and they reported that indicating words would lead to a recall of about 90%, we decided to just use their indicating words (words end with ‘er/est’ and a set of manually collect words) to identify comparative sentences, deferring the introduction of pruning methods to the pruning section.

Objective features

An objective review may cover multiple aspects of a product, which is intended to be unbiased or impersonal. Objective reviews tend to state some facts about a product or its properties. Given that the information is concrete bits of facts, the review is usually expressed in concrete language in terms of precise numbers or quantities, weights and measures, and so on. In the context of product reviews, there could be many reviews describing some features of one product without using any opinion word. For instance, “It comes with a rechargeable battery...”, “There are three colors to choose from ...”, “It combines auto mode and manu mode ...” In these cases, double propagation no longer works, neither does the extraction of frequent terms because some features may just appear only once or a few times.

Objective patterns occur frequently in text and are expressed by a variety of lexico-syntactic structures such as “part-whole” relation (Girju, Badulescu and Moldovan 2006; Popescu and Etzioni, 2007) and “contain/include”, “come with”, which are also widely-used patterns to describe facts. We identified the following types of patterns of objective features:

NP+Verb+NP: NP is the noun word or noun phrase representing product or product feature. In this pattern, different verb choices may represent similar relationship between the two NPs. For example, we can infer from the expression of “the camera contains a lens” that memory card is an accessory of camera. Likewise, in the expression “This phone has a color screen”, screen is a feature of phone. Both of the above examples represent the part-whole relation. We followed Zhang et al (2011)’s method to extract these features using verbs such as “have” “include” “contain” “consist”, and “comprise”. Besides, there are other verbs expressing the relationship between the feature and the product other than the part-whole type. For example, “This camera comes with a 4mb memory card”, and “The phone is made of plastic”. In these examples, the feature actually is a property or accessory of the product. We propose to address such features by compiling verbal phrases such as “made of” and “come with”.

PRP/Ex+Verb+NP: PRP/Ex refers to pronoun, usually used to represent a product, and NP indicates a feature. In the following sentences, “It plays mpeg video”, “It comes with a tripod and a memory card”, it refers to a product, and the object of verbs are features. Although this statement contains “comes with”, it cannot be directly treated as an example of the previous pattern (NP+verb+NP) without pronoun resolution, because it could lead to the extraction of wrong patterns. In addition, previous

methods are incapable of dealing with statements like “There are auto mode and manu mode”, where “auto mode” and “manu mode” should be considered as features, which will be addressed in this research. To extract aforementioned patterns, we first matched POS patterns and then substituted ‘verb’ with specific words, after that we extracted the object of each verb as candidate features.

Special features

The above patterns are not inclusive. We also include several special patterns such as “with/without” pattern, “no/not” pattern and unique expressions; they belong to this type because it is difficult and not necessary to determine their polarity.

“With” pattern describes sentences beginning with “with/without”, “by using”, etc. For example, “without auto mode, it is difficult to ...”, “with external flash on my Nikon SLR ...”, This is a specific pattern of reviews and posts in an online environment. Likewise, such a pattern either doesn’t contain any opinion word, or even if an opinion word presents, it doesn’t directly modify the features. We label this pattern as “with/without + noun/noun phrases”, and used it in feature extraction.

“No” pattern is in the form of “no” word followed by noun/noun phrase, which was first introduced by Zhang et al (2011) such as “no noise” and “no indentation”. The pattern also takes care of the some fixed “no” expressions such as “no problem” “no offense”. We extended the pattern as “no/not” pattern to address cases like “not a deal” as well as “no accessories” and incorporated them into our method for feature extraction.

To extract above two short patterns, we applied a naive match method (e.g. regular expression) to extract those patterns and then extracted the noun words as candidate features. Additionally, we manually filtered several idioms such as “no problem”.

The third type is unique expressions such as brand names and models, we summarized and categorized those special cases, and developed dictionary based method and some heuristic rules for them. For instance, models are usually expressed in the form of “Letter+Number” or “Brand+Number”, we first use regular expression to match those terms as candidates, and then use a manually compiled brand dictionary to extract those real brand features.

Pruning methods

Integrating several methods in extracting product features is expected to improve the recall, meanwhile it also is likely to reduce the precision. To filter those non-feature terms, the simplest way would be to remove those non-frequent terms.

Unlike traditional frequency based methods, we used the document frequency instead of Term Frequency * Inverse Document Frequency (TFIDF). The method was motivated by the following observation. After removing stop words, it becomes almost impossible for two reviews mentioning the same term to be unrelated to the product. Thus, for all candidates extracted by above methods, we calculated the DF of each term, and remove those terms whose DF is lower than a minimum threshold (manually set up as 2). In addition, we further pruned the terms to deal with two types of features: lexical redundant features and semantic non-relevant features. Lexical redundant features contain those features that share the same meaning but are represented in different formats, such as wifi and wi-fi or misspelling. To this end, we introduce textual feature similarity function that compares the edit distance between a candidate feature and all the other features. If the distance is less than one, the two features would be considered as similar to each other. To detect semantically irrelevant words, we introduced another function that calculates the semantic similarity between a candidate feature and a class word. We used WordNet’s path similarity to measure the similarity between two words. For example, within the type of camera, we calculated the similarity between one candidate with class word ‘camera’, and compared this value with the mean of similarities between a predefined set of features and the class word, those predefined features are collected from Hu and Liu (2004)’s result. Furthermore, if the distance between the candidate and class word is greater than a minimum threshold (set as the std-mean of the similarities between the class word and all those predefined features), it is extracted as a feature.

EXPERIMENTS

Data set

We conducted experiments to empirically test the proposed methods for product feature extraction. We employed the dataset of online product reviews from Hu and Liu (2004), which was originally collected from two websites: Amazon.com and Cnet.com. This data set consists of five electronics products: two digital cameras, one DVD player, one mp3 player, and one cellular phone.

Each of the reviews contains a textual review and a title. Before extracting features from each of the reviews, we first analyzed the data using natural language processing, including tokenization, stop-words removal (127 stop words from nltk, performed after syntactic analyses), POS tagging, entity recognition, chunking and dependency relation analysis. Some descriptive statistics of the dataset is reported in Table 1.

Product Name	No. Reviews	No. Features
Camera (Canon G3)	45	79
Camera (Nikon)	34	96
Cell Phone (Nokia 6610)	41	67
MP3 Player (Creative)	95	57
DVD Player (Apex)	99	49

Table 1. Statistical Description of Data set

Evaluation

We selected precision and recall as the evaluation metrics, which are commonly used in information retrieval and document classification research. Precision is defined as the ratio of the number of correctly classified items to the total number of items that were classified. Recall is defined as the ratio of the number of correctly classified items to the total number of items that were classified as the same category in the gold standard. We chosen Hu and Liu (2004)'s method and a newer method (Qiu et al 2009) as the baselines for comparison.

RESULTS

The performances of Hu and Liu (2004)'s method and the double propagation method (Qiu et al 2009) are reported in Table 2. The performance of our method is reported in Table 3.

Product	Hu and Liu (2004)		Double Propagation	
	Recall	Precision	Recall	Precision
Camera 1	0.82	0.75	0.81	0.87
Camera 2	0.79	0.71	0.81	0.90
Cell phone	0.76	0.72	0.86	0.90
MP3	0.82	0.69	0.84	0.81
DVD	0.80	0.74	0.86	0.92
Avg	0.80	0.72	0.83	0.88

Table 2. Precision and recall of Hu and Liu (2004) and Double Propagation

Product	Recall	Precision
Camera 1	0.83	0.81
Camera 2	0.82	0.88
Cell phone	0.90	0.85
MP3	0.87	0.82
DVD	0.88	0.83
Avg	0.86	0.83

Table 3. Precision and recall of our methods

As shown in table 2, the performance of double propagation is better than Hu and Liu (2004)'s work in terms of recall and precision. A comparison of the results reported in Table 2 and Table 3 shows that our methods outperformed double propagation in terms of recall for all the selected products. Comparing to Double propagation, our methods achieved a higher recall due to the fact that we considered both subjective and objective features, and to extract subjective features we incorporated new patterns beyond double propagation. Nevertheless, the increase in recall is very moderate partly because some features appear both in subjective statements and objective statements and the size of the dataset. Besides, the precision of

our method is better than Hu and Liu (2004)'s method, but worse than Double propagation (Qiu et al 2009). The results suggest that the proposed pruning methods are effective but should benefit from additional pruning strategies from (Qiu et al 2009).

The results also show that developing extraction methods based on feature types is promising. We conclude that only considering frequency of term is not inclusive, taking dependency relation and grammar rules into consideration is also important. In addition, it is necessary and useful to incorporate more grammars based on how features are expressed.

CONCLUSION

In this paper, we proposed new methods for extracting product features from online consumer reviews based on natural language processing and machine learning techniques. Our experimental results indicate that the proposed techniques are effective in extracting both subjective features and objective features.

In our future work, we plan to improve these techniques in the following ways. First, consider more structural patterns of both subjective and objective features to increase recall, design new experiments to prune features to improve precision; second, evaluate our methods to online reviews of other types of products such as movies and books; and third, apply the extracted features to improve the prediction of review helpfulness and review recommendations.

REFERENCES

1. Archak, N., Ghose, A., & Ipeiritos, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, 57(8), 1485-1509.
2. Chevalier, A. and Mayzlin, D. (2006) The Effect of Word of Mouth on Sales: Online Book Reviews, *Journal of Marketing Research*, 43 (August), 345–54.
3. Dellarocas, C. (2003) The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms, *Management Science*, 49 (10), 1401–1424
4. Ghani, R., Probst, K., Liu, Y., Krema, M. and Fano, A. (2006) Text mining for product attribute extraction, *Proceedings of the SIGKDD Explorations Newsletter*, vol. 8, pp. 41–48, 2006.
5. Girju, R., Badulescu, A. and Moldovan, D. (2006) Automatic Discovery of Part-Whole Relations, *Computational Linguistics* ,32(1):83-135
6. Hai, Z., Chang, K. and Kim, J. (2011) Implicit Feature Identification via Co-occurrence Association Rule Mining, *Proceedings of the Computational Linguistics and Intelligent Text Processing - 12th International Conference*, Tokyo, Japan
7. Hu, M. and Liu, B. (2004) Mining and summarizing customer reviews, *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* Aug 22-25 2004, Seattle, Washington, USA
8. Hu, M., Liu, B. (2006) Opinion feature extraction using class sequential rules, *Proceedings of the Spring Symposium on Computational Approaches to Analyzing Weblogs* Palo Alto, USA
9. Jindal, N., & Liu, B. (2006a). Identifying comparative sentences in text documents. *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 244-251). ACM.
10. Jindal, N., & Liu, B. (2006b). Mining comparative sentences and relations. *Proceedings of the National Conference on Artificial Intelligence* (Vol. 21, No. 2, p. 1331). Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.
11. Kim, S., Pantel, P., Chklovski, T. and Pennacchiotti, M. (2006) Automatically Assessing Review Helpfulness, *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing* July, Sydney, 423–430
12. Kobayashi, N., Inui, K. and Matsumoto, Y. (2007) Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, June, Prague
13. Lee, T. Y. and Bradlow, E. T. (2007) Automatic construction of conjoint attributes and levels from online customer reviews, *Working paper*, The Wharton School, University of Pennsylvania
14. Liu, B., Hu, M., Cheng, J. (2005) Opinion observer: Analyzing and comparing opinions on the web, *Proceedings of the WWW* Chiba, Japan
15. Liu, J. Cao, Y. Lin, C. Huang, Y. and Zhou, M. (2007) Low-quality product review detection in opinion summarization, *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, June, Prague, 334–342.

16. Liu, Y., Huang, X., An, A. and Yu, X. (2008) Modeling and Predicting the Helpfulness of Online Reviews. *Proceedings of the Eighth IEEE Int'l Conf. Data Mining (ICDM '08)*, Pisa, Italy, 443-452
17. Mudambi, S. M., and Schuff, D. (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Quarterly*, 34(1), 185-200.
18. O'Mahony, M. P., and Smyth, B. (2010). A classification-based review recommender. *Knowledge-Based Systems*, 23(4), 323-329.
19. Popescu, A. M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. In *Natural language processing and text mining* (pp. 9-28). Springer London.
20. Qiu, G., Liu, B., Bu, J., & Chen, C. (2009). Expanding domain sentiment lexicon through double propagation. *Proceedings of the 21st international joint conference on Artificial intelligence* (pp. 1199-1204).
21. Raju, S. Shishtla, P. Varma, V. (2009) A Graph Clustering Approach to Product Attribute Extraction. *Proceedings of the 4th Indian International Conference on Artificial Intelligence* Tumkur, India
22. Wiebe, J. Wilson, T. Bruce, R. Bell, M and Martin, M. (2004) Learning Subjective Language. *Computational Linguistics*, 30(3), 277-308.
23. Wong, T.L. and Lam, W. (2005) Hot item mining and summarization from multiple auction Web sites, *Proceedings of the fifth IEEE international conference on data mining (ICDM'05)*, Houston
24. Wong, T. L., & Lam, W. (2008). Learning to extract and summarize hot item features from multiple auction web sites. *Knowledge and Information Systems*, 14(2), 143-160.
25. Zhang, L., Liu, B., Lim, S. H., and O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. *Proceedings of the 23rd International Conference on Computational Linguistics: Posters* (pp. 1462-1470). August 23-27th, Beijing, China