Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 22 (2013) 142 – 151

Procedia Computer Science

17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013

Feature Based Summarization of Customers' Reviews of Online Products

Kushal Bafna, Durga Toshniwal

M. Tech. Computer Science, Electronics and Computer Engineering Dept., IIT Roorkee-247667, India
Associate Professor, Electronics and Computer Engineering Dept., IIT Roorkee-247667, India

Abstract

With the growing availability and popularity of opinion-rich resources such as review forums for the product sold online, choosing the right product from a large number of products have become difficult for the user. For trendy product, the number of customers' opinions available can be in the thousands. It becomes hard for the customers to read all the reviews and if he reads only a few of those reviews, then he may get a biased view about the product. Makers of the products may also feel difficult to maintain, keep track and understand the customers' views for the products. Several research works have been proposed in the past to address these issues, but they have certain limitations: The systems implemented are completely opaque, the reviews are not easier to perceive and are time consuming to analyze because of large irrelevant information apart from actual opinions about the features, the feature based summarization system that are implemented are more generic ones and static in nature. In this research, we proposed a dynamic system for feature based summarization of customers' opinions for online products, which works according to the domain of the product. We are extracting online reviews for a product on periodic bases, each time after extraction, we carry out the following work: Firstly, identification of features of a product from customers' opinions is done. Next, for each feature, its corresponding opinions' are extracted and their orientation or polarity (positive/negative) is detected. The final polarity of feature-opinions pairs is calculated. At last, feature based summarizations of the reviews are generated, by extracting the relevant excerpts with respect to each feature-opinions pair and placing it into their respective feature based cluster. These feature based excerpts can easily be digested by the user.

© 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of KES International

Keywords: Opinion Mining; Summarization; Sentiment Classification; Feature Based Clustering and Reviews;

1. Introduction

With the rapid expansion of e-commerce platform, online shopping on the products has enhanced drastically. Due to the vast variety of products and convenient shopping experience with attractive offers, these platforms have become popular for customers and even for manufacturers. But at the same time it becomes hard for the

1877-0509 © 2013 The Authors. Published by Elsevier B.V. Open access under CC BY-NC-ND license.
Selection and peer-review under responsibility of KES International
doi:10.1016/j.procs.2013.09.090

customers to get the help from the professional sales staff to buy the product. One approach that is being used to deal with these challenges is, merchant providing meta-data for the products sold online. The problem with the meta-data is that, the user finds difficult in deciding about the product by just knowing the features of the product, this lead to customer insecurity and which in turn have an adverse effect on online e-commerce revenue.

Online merchants have enabled forum which helps customers to get reviews and to express opinions about the product. But with hundreds or thousands of reviews available for a popular product, it had made difficult for the customer to read all the reviews and to make a well-versed decision on buying the product. And if the customer reads only some of the reviews, then biased view about the product may be generated.

The approach used to solve this problem is being addressed to a certain extent by the research community. For example, Amazon and Netflix use recommender system [1, 2]. The problem with this recommended system is that they are completely opaque, due to which it becomes hard for the customers to believe about the recommended items. This is due to the fact that there is no information available way a user should purchase an item. The user should blindly believe the algorithm underneath the recommended system.

In this research, we deal with the above challenges by generating feature based summarization of customers' opinions for product sold online, which works according to the domain of the product. For example, in the domain of Movie product reviews, features video and picture are domain synonyms and are grouped under the same feature group pictures, whereas in the domain of iPhone product reviews, video and picture represents two completely different features and are grouped under two separate feature group videos and photos respectively.

The research is carried out as follows: Initially, the set of customers' reviews for the product are extracted from the web. Those customer reviews are semantically analyzed as follows (1) Extraction of domain specific features of the product sold online from customer reviews are done. Here before extracting the features, all domain synonyms words and phrases, are grouped under the same feature group. (2) Opinion sentences are identified and corresponding opinion words are extracted. Positive or negative orientation of each opinion word is analyzed. As there can be more than one opinion words in a sentence, existing techniques have failed to deal with them well. We developed a new way to solve this issue by taking into consideration the distance of each opinion word with respect to product feature and then calculating the overall opinion of the sentence. This turns out to be highly useful. At last, summarization of the reviews is done by extracting the relevant excerpts with respect to each feature-opinions pair and placing it into their respective feature based cluster. These feature based excerpts can easily be digested by the user. Here the system implemented is dynamic i.e., after each period (daily or hourly basis) all the user reviews added during that span of the period are extracted from the web and an updated feature based summary is generated.

Domain specific feature-based summary process is shown below using an example. Let us take the example of Apple iPhone 4S. The summary generated for apple iPhone 4S looks like the following:

As shown in the Table 1, feature iPhone and size has more positive reviews with the count 203 and 135 positive reviews respectively. 70 positive user opinions about Ios, 122 positive user opinions about screen touch or any other feature can be viewed by just clicking onto their respective features. Similarly, System Portability of the iPhone and battery has most negative reviews with count 133 and 86 respectively as shown in Table 2.

Our work is different from traditional text based summarization [3, 4, 5] in following ways. A summary generated using traditional text based summarization is a free text based document when compared to ours where it is more structured. We generate structured summary based on features of the product and positive negative opinion about the product i.e., Opinion sentences and not the other irrelevant information that user had described in his reviews.

However, it is not a trivial task to extract excerpts of strong and weak points from reviews. In this research, we propose a system which automatically extracts feature-based opinion excerpts and generates a summary of positive and negative aspects about the features.

Table 1. Top 5 positive features based cluster for Apple iPhone 4S

IPhone	203+
Size	135+
Screen Touch	122+
Sound	89+
Ios	70+

Table 2. Top 5 Negative features based cluster for Apple iPhone 4S.

System Portability	133-
Battery	86-
IPhone	67-
Sound	30-
Bluetooth	27-

We proposed a number of techniques using the information present in natural language processing, artificial intelligence and opinion mining to generate feature based opinion summarization of customer reviews for the product. We will evaluate our proposed methodology using the customer reviews extracted from the Amazon product website. Our results show that these methodologies are highly effective.

2. Related Work

Our work is closely related to [6, 7], where the authors propose a sentence-based analysis, in which it uses association rule mining [8] to extract the most frequent features. The underlying principle of this method according to [6, 7] is that, product features will occur most frequently when compared to other words or word phrases in a user reviews of a product (co-occurrence based approach). But, association rule mining has some significant drawback and challenge: association rule mining generates [9] many features which doesn't actually represent feature of the product, but are just some frequently occurring noun phrases like e.g. "Comment", or "problem" [10]. Hence, we built a model which is the combination of association rule mining and probabilistic approach. The basic rationale of this approach is that each product field (e.g. The field of iPhone, having features like "Sound", or "Touch") has its particular language, i.e. Nouns or noun phrases representing features (features are basically nouns [12]) for a product have a higher probability of occurrence in the document belonging to that product field compared to document belonging to any other product field. Hence, we are able to remove all the common frequently descriptive phrases in the feature formation process using the above probabilistic model.

In [13], the author used information extraction system which is opinion orientation centric framework. Question answering kind of system was created using a summary representation of opinion. Author proposes opinion based “scenario templates” to act as summary representations of the opinions expressed in customer review. Our task is different. We do not use any template for summary generation. We try to extract domain specific product features and their corresponding opinions to automatically generate a summary.

Opinion mining is an old research area. Two important research areas in opinion are sentiment analysis and feature-based opinion mining. Sentiment classification analysis each review and classify them as positive or negative. [20, 21] uses a document level sentiment classification. Our work is different, we are only interested in reviews of product which contains feature and opinion pairs rather than the whole review and we perform classification at sentence level. Subjective sentence level classification is studied in [24], which finds whether a sentence is an objective sentence or a subjective one. Sentence level opinion or sentiment categorization is studied in [22, 24]. A review can have many features, and each can have their own opinion polarity. e.g., “the Sound System of this phone is awesome and so is the Touch, but the battery life is short.” “Sound System”, “Touch” and “Battery life” are featured. The opinion "awesome" on “Sound System”, “Touch” is positive, and the opinion "short" on "Battery Life” is negative.

In [14], it tries to extract domain specific feature by using the probabilistic approach [11] at document level. Our work differs from there in the following way: (1) we used the probabilistic approach at word level. (2) We used a combination of association mining and probabilistic approach to extract features. (3) Before extracting features initial processing is done to group domain synonym words under the same word groups using the semi-supervised learning problem in [23] to extract domain specific features. The results show that the accuracy of features generated using the proposed method are highly effective.

3. Proposed Technique

In this research, we will update the work of [6] who proposed a generic feature based opinion summarization system. We have greatly modified our process in order to make the system work according to the domain of the product. Here the system implemented is dynamic i.e., after each span of period (daily or hourly basis) all the user reviews updated during the period are extracted from the web and an updated feature based summary is generated. And also before the extraction of feature starts another algorithm is made to run on customers' reviews to form groups of all the domain synonyms nouns or noun phrases, using semi-supervised learning problem as described in [23]. The workflow of the propose technique is shown in the Fig. 1.

3.1. Feature Extraction

Here before finding the frequent features, pre-processing of all the user reviews are done as follows (1) Fuzzy matching to remove/replace misspelling words in the document. (2) Grouping of domain synonyms nouns or noun phrases i.e., the nouns or noun phrases which are domain synonyms, are grouped under the same noun group, using the semi-supervised learning problem in [23]. For example, in the domain of Movie product, features *video* and *picture* are domain synonyms and are grouped under the same feature group *pictures*, where as in the domain of iPhone product reviews, *video* and *picture* represents two completely different features and are grouped under two separate feature groups *videos* and *photos* respectively, which is not done by [6, 14].

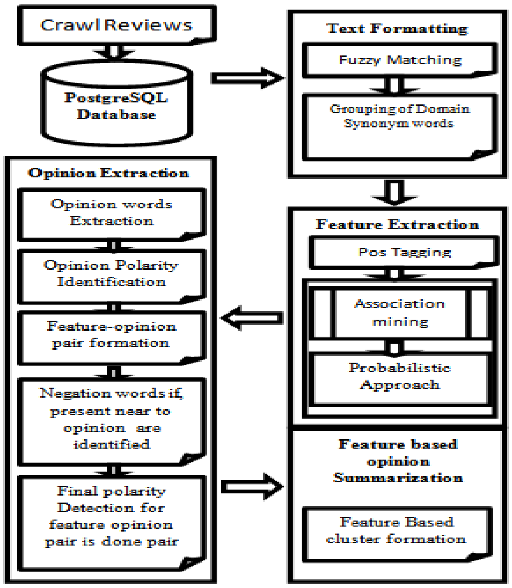

Fig. 1. Flow Diagram of Feature Based Summarization System.

Feature extraction consists of two phases: 1) Part of speech (POS) tagging all the documents representing user reviews of a product. 2) Domain specific feature extraction. In POS tagging, the word in the documents is tagged with their respective part of speech. For POS tagging the documents, we used Stanford NLP Parser [19], which POS tag the all the documents and generates POS tagged XML doc. As an output file.

The next task is to identify domain specific features. In this process not just generic features but also features which are specific to a product are also extracted. Earlier research on opinion indicates that the features of a product are usually represented using nouns [12]. Almost all features are represented using nouns, but all nouns may not be features with respect to product field. Example: my comments are always accurate with respect to iPhone reviews. Here comments represent nouns but they are not feature for the iPhone review.

Eliminating the nouns which do not represent features for a product field can be done using a combination of association mining and probabilistic approach. Initially, we use association rule mining [8] to extract the most frequent occurring features. The underlying principle of this according to [8] is that, product features will occur more frequently when compared to other words or word phrases in a user reviews for a product (co-occurrence based approach). Therefore, association rule mining is used to extract the frequent occurring nouns. However, drawback with the association rule mining generates many features which don't actually represent features of the product, but is just some frequently occurring noun phrases like e.g. "Comment", or "problem". To remove these commonly occurring descriptive phrase our probabilistic model is used in addition to the association mining approach (domain specific). The basic rationale of this approach is that each product field (e.g. The field of iPhone, having features like "Sound", or "Touch") has its particular language, i.e. Nouns or noun phrases representing features (features are basically nouns [12]) for a product have a higher probability of occurrence in the document belonging to that product field compared to document belonging to any other product field. In the above e.g., "comment" which do not represent a feature for any of the product, have almost similar probability of appearing in documents of all the product reviews and therefore will generate a characteristic power value which won't lesser than $\boldsymbol{\theta}$ satisfy equation 1 as shown below and will not be considered as a feature for iPhone product, even though its frequency of appearing in the document of product field is high. As in [6], the author was unable to resolve this type of issues. So we defined the Probabilistic or characteristic power equation to remove all the feature candidates which are no real features, but are just frequently occurring nouns or noun phrases from the features output generated using associative rule mining technique.

$$pr(cf \in C) - pr(cf \in G) >= \boldsymbol{\theta} \qquad (1)$$

$$pr(cf \in C) = \frac{|cf\ in\ C|}{|w\ in\ C|} \qquad (2)$$

$$pr(cf \in G) = \frac{|cf\ in\ G|}{|w\ in\ G|} \qquad (3)$$

cf – Represents, a frequently occurring noun, extracted from user reviews of a specific product (i.e., the product from which we are trying to extract the feature and generate feature-based summary) using association mining.

w – Represents, words present in the user reviews of a product.

C – Represents, users reviews corpus of a specific product.

G – Represents, users review corpus of Generic Product field.

$\boldsymbol{\theta}$ – Value is adjusted based on the results.

A generic product is chosen in such a way that the features of the generic product are in complete repulsion with the product field in question.

For example, the noun “Battery Life”, extracted using association mining rule from iPhone reviews, its characteristic power eq. 1 can be computed as the probability that an iPhone reviews corpus C containing the noun phrase “Battery Life” minus the probability that the generic product reviews G (example Body perfume) containing the noun phrase “Battery Life”. Probability in the equation 2 is defined as the frequency of occurrence of a noun in specific product reviews divided by the total number of words present in the specific product reviews. Probability in the equation 3 is defined as the frequency of occurrence of a noun in generic product reviews divided by the total number of words present in the generic product reviews. The characteristic power value of each and every noun extracted using association mining rule is calculated. All the nouns whose characteristic power value becomes greater than $\boldsymbol{\theta}$ are termed as domain Specific features of the Product.

In [14] author users only the characteristic power to extract the features specific to a domain. The drawbacks with [14] are as follows: firstly, for example in iPhone reviews using [14] the noun “cap” is selected as a feature of the iPhone even though noun “cap” is not a feature of the iPhone and the frequency of occurrence of the “cap” is less. Reason being the $\boldsymbol{\theta}$ in the characteristic power equation of [14] relatively depends upon both, noun probabilities in specific product as well as generic products Here the noun “cap” is not at all present in the generic documents, but due to relativity in equation 1 the characteristic power of “cap” become greater than $\boldsymbol{\theta}$ because, $\boldsymbol{\theta}$ value is adjusted relative to specific product and generic field. Our approach solves this problem by using the associative mining rule as in [6] in addition to probabilistic approach, thereby removing all very infrequently occurring nouns in the competition of feature formation for a product, which in turn makes our probabilistic approach perform better. Secondly, our probabilistic approach for finding the domain specific feature is more at word level than at document level as in [14]. Problem with [14] probabilistic approach is that, if a particular noun e.g. "size" is highly accumulated in only few of the documents of users reviews for specific product and absent in all most majority of documents of the specific product field, hence the probability of occurrence of a noun at document level for the specific product field will be really low relative to generic field thereby making the characteristic power value less than $\boldsymbol{\theta}$ even though noun "size" word level count(not document level count) is high enough in specific product field when compared to the word level count in generic product field to make it as a feature for a specific product. We solve the problem by making the characteristic power equation to work at word level and hence making the probabilistic value of "size" in the above example high enough in relative to generic field to overcome the threshold of characteristic power in equation 1. At last, we also grouped domain synonyms nouns or noun phrases i.e., the nouns or noun phrases which are domain synonyms, are grouped under the same noun group, using the semi-supervised learning problem in [23], which is not made by [14] . Results of our expriements shows that these techniques are highly effective in generation of features for a product.

3.2. Opinion Extraction

Opinion extraction is carried out in few steps. (1) Extraction of all the opinion word as well as detection of polarity (positive/negative) of each opinion word (2) Formation of feature-opinion pair by assigning the opinion word to nearest feature. (3) Finally, determining negation word near to each opinion word and detection of final polarity for each feature-opinion pair is done.

Prior research indicates that the opinions are generally represented using adjective [6]. POS tagged xml document generated by Stanford NLP Parser is taken as an input and all the adjectives (opinions) with respective to product reviews are extracted.Their semantic polarities (positive /negative) are detected. While the method in [6] uses seed lists for detection of polarity, we used opinion lexicon [15]. Opinion lexicon is an online dictionary which contains a large collection positive negative and netural adjectives (0, 1 or -1). However, user reviews of a product can contain certain non-familiar adjectives (e.g. “quasi-comedic”, “Trail-blazing”, “psychotic”). If opinion Lexicon is unable to detect the adjective's polarity, then it is given as an input to another online service called Sentiword-Net [16, 17] and at last delegated to the administrator of Feature Based Summarization system.

Opinion word identified in the above process is assigned to the nearest feature in the sentence. The basic rationale of assigning opinion word to the nearest feature is that, opinion word for the feature will always be the most closest one around it.

The assignment of the opinions to the feature is achieved by computing the distance of each opinion word to detected features in a sentence and then assigning an opinion to the feature to which it is most nearest. If there are two or more features with the same distance, then the opinion is assigned to the product feature mentioned first.

As there can be more than one opinion words in a sentence, existing techniques have failed to deal with them well. We developed a new way to solve this issue by taking into consideration the distance of each opinion word with respect to product feature and then calculating the overall opinion of the sentence. This turns out to be highly useful than the one in [14].

Next, we try to find negation word in the neibourhood of each adjective. If any negative word is found, then the orientation of feature-opinion pair is reversed and a final Polarity of each feature-opinions pairs are generated. Here for each feature of customers' reviews of a given product: all its positive, negative polarities are added up independently. As soon as the polarity profile of each feature is computed.

3.3. Summarization using Feature Based Clustering

In this phase, 2n clusters are formed for n features extracted using the above feature extraction process, where each feature will have two clusters, positive and negative. Positive cluster will store all positive reviews of the feature and the negative cluster will store all negative reviews. Next, extraction of relevant excerpts from the user reviews with respect to each feature opinion pair is done and are placed into their respective feature based cluster. Placing of excerpts into the clusters are based on the product feature present in the excerpt and the polarity of feature opinion pair (Detected in the above opinion extraction process) i.e., whether the excerpt represents positive opinions about the feature or a negative one. Opinion excerpts present in each feature based cluster allows for a very quick and efficient skimming through the relevant opinions on the feature.

4. Experimentation

A system, called dynamic FBS (Feature-Based Summarization) which work according to the domain of the product have been implemented in Java. We conducted experiments on customer reviews for 3 products: Cannon G3 Camera, iPhone 4s, Mp3 player. The customer reviews for products were crawled from Amazon website [18]. IDE used is Eclipse with Maven Plug-in.

The procedure is as follows: for each Product, its corresponding features are extracted. Next, for each features, their corresponding customer reviews containing the opinion word are selected. Then, the text excerpts from to each selected feature reviews are extracted (It is basically the sentence containing the feature opinion pair). Finally the excerpts are placed into their respective positive or negative feature based cluster. The top-5 excerpts of each feature based positive or negative clusters are inspected manually to check the accuracy of the feature based opinion summarization system. Results shows that, 80% of all excerpts have been classified correctly for the iPhone, while the other 20% where wrongly classified. Whereas using the method of [14] only 78% of classification of excerpts for correctly done. Firstly, the enhancement is due to a better way of assigning multiple conflicting opinion in the product reviews to each feature. Secondly, it is due to the extraction of domain specific feature compared to [14]. Feature extracted using the proposed method where manually checked for its accuracy and it was seen that the feature extracted using this method were 92% accurate when compared 88% of [14] and 80% of [6]. The enhancement in the accuracy is due to (1) we used the probabilistic approach at word level. (2) We used a combination of association mining and probabilistic approach [11] to extract domain specific feature. (3) Before extracting features, initial processing was done to group nouns or noun phrases which are domain synonyms to be grouped under the same noun group using the semi-supervised learning problem in [23]. The results for Cannon Camera, iPhone 4s, Mp3 player on the feature extraction accuracy and opinion sentence polarity detection accuracy of the system are shown below in Table 3 and Table 4 respectively. Here the system implemented is dynamic i.e., after each span of period (daily or hourly basis) all the new users' reviews added during that period are extracted from the Amazon website and an updated feature based summary is generated. Fig. 2 shows changes in the accuracy of our feature extraction process with daily updates of users' reviews. Fig. 3 shows changes in our opinion sentence polarity detection accuracy with daily updates of users' reviews. From, Fig. 2 and Fig. 3 we can say that our FBS system is robust enough to handle periodic update of users' reviews.

Table 3. Feature Extraction Accuracy of the System

Product Name	Feature Extraction Accuracy (Association Mining)	Feature Extraction Accuracy (Probabilistic)	Feature Extraction Accuracy (Association Mining and Probabilistic Approach)
iPhone 4s	80%	88%	92%
Cannon Camera	78%	85%	90%
Mp3	84%	89%	94%

Table 4. Opinion Sentence Polarity Detection Accuracy of the System

Product Name	Opinion sentence polarity detection accuracy of [14]	Opinion sentence polarity detection accuracy
iPhone 4s	78%	80%
Cannon Camera	75%	76%
Mp3	80%	82%

User satisfaction with respect to FBS system was measured using the online survey conducted by [14]. The ultimate aim was to compare the effectiveness of reading whole customer reviews with the summary generated using Feature based summarization system. Our feature based summary generation approach turn out to be highly effective. Overall, the accuracy of our feature extraction and opinion sentence polarity detection system is satisfying.

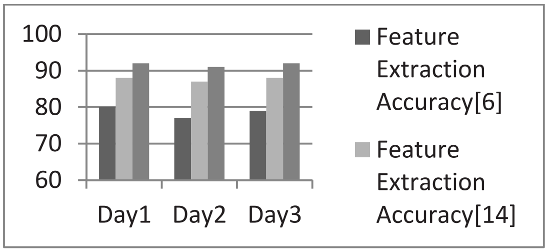

Fig. 2. Change in accuracy of Feature extraction process with daily updates of users' reviews

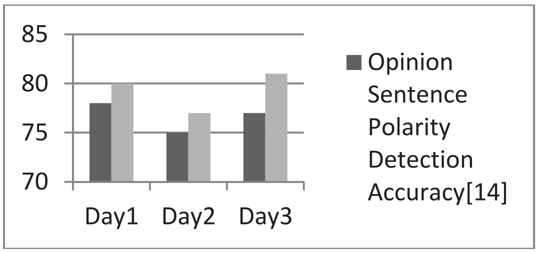

Fig. 3. Change in our opinion sentence polarity detection accuracy with daily updates of users' reviews

5. Conclusion

In this research, we proposed a novel technique for dynamic feature based summarization of customers' reviews which works according to the domain of the product. It was based on natural language processing and opinion mining. Results indicate that the proposed methods are highly effective and efficient in performing their tasks. Now digesting the information contained in large numbers of products review corpus has become much easier for users by allowing users for a quick and efficient way of skimming through the product reviews.

In future work, we will aim at improving the accuracy of our opinion polarity detection and feature extraction algorithms. And also work on discovering, opinions expressed with adverbs, verbs and nouns. The evaluations performed in this paper provide a good baseline for developing even more efficient feature based summarization system.

References

[1] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," IEEE Internet Com-puting, vol. 7, Jan. 2003, pp. 76-80.

[2] R.M. Bell, Y. Koren, and C. Volinsky, "All together now: A perspective on the NETFLIX PRIZE," CHANCE, vol. 23, Apr. 2010, pp. 24-24.

[3] Goldstein, J., Kantrowitz, M., Mittal, V., and Carbonell, J. 1999. Summarizing Text Documents: Sentence Selection and Evaluation Metrics. SIGIR'99.

[4] Salton, G. Singhal, A. Buckley, C. and Mitra, M. 1996. Automatic Text Decomposition using Text Segments and Text Themes. ACM Conference on Hypertext.

[5] Tait, J. 1983. Automatic Summarizing of English Texts. Ph.D. Dissertation, University of Cambridge.

[6] M. Hu and B. Liu, "Mining and summarizing customer reviews," ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD), Seattle, USA: 2004.

[7] B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," Foundations and Trends in Information Retrieval, vol. 2, Jan. 2008, pp. 1-135.

[8] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Asso-ciation Rules," 20th Int. Conf. on Very Large Data Bases (VLDB), Santiago de Chile, Chile: 1994.

[9] S. Morinaga, K. Yamanishi, K. Tateishi, and T. Fukushima, "Mining product reputations on the Web," 8th ACM Int. Conf. on Knowledge Discovery and Data Mining (KDD), Edmonton, Can-ada: 2002.

[10] C. Scaffidi, K. Bierhoff, E. Chang, M. Felker, H. Ng, and C. Jin, "Red Opal: product-feature scoring from reviews," 8th ACM Conf. on Electronic Commerce (EC), San Diego, California: 2007.

[11] K. Kageura and B. Umino, "Methods of automatic term recogni-tion: A review," Terminology, vol. 3, Jan. 1996, pp. 259-289.

[12] H. Nakagawa and T. Mori, "A simple but powerful automatic term extraction method," Int. Workshop on Computational Ter-minology, Morristown, NJ, USA: 2002.

[13] Cardie, C., Wiebe, J., Wilson, T. and Litman, D. 2003. Combining Low-Level and Summary Representations of Opinions for Multi-Perspective Question Answering. 2003 AAAI Spring Symposium on New Directions in Question Answering.

[14] Silviu Homoceanu, Michael Loster, ChristophLofi and Wolf-Tilo Balke, "Will I like it? – Providing Product Overviews based on Opinion Excerpts", IEEE Conference on Commerce and Enterprise Computing (CEC),Luxembourg,2011.

[15]http://aliasi.com/lingpipe/demos/tutorial/sentiment/readme.html, LingPipe - tool kit for processing text using computational linguistics.

[16] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," 5th Conf on In-ternational Language Resources and Evaluation (LREC), Genoa, Italy: 2006.

[17] S. Baccianella, A. Esuli, and F. Sebastiani, "SENTIWORD NET 3.0 : An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining," 7th Conf on International Language Resources and Evaluation (LREC), Marrakech, Morocco: European Lan-guage Resources Association (ELRA), 2008.

[18] http://www.amazon.com/ , Online Shopping for customers.

[19] http://nlp.stanford.edu/software/tagger.shtml , The Stanford Natural Language Processing Group.

[20] E. Riloff and J. Wiebe. 2003. Learning extraction patterns for subjective expressions. EMNLP'2003, 2003.

[21] M. Gamon, A. Aue, S. Corston-Oliver, and E. K. Ringger. Pulse: Mining customer opinions from free text. IDA'2005.

[22] A-M. Popescu and O. Etzioni. Extracting Product Features and Opinions from Reviews. EMNLP-05, 2005.

[23] Zhongwu Zhai, Bing Liu, Hua Xu, Peifa Jia, "Clustering Product Features for Opinion Mining", In: Proc. Of the 4th ACM International WSDM Conf., pp. 347-354. ACM. 2011.

[24]V. Hatzivassiloglou and J. Wiebe. Effects of adjective orientation and gradability on sentence subjectivity. COLING'00, 2000