The Method for a Summarization of Product Reviews Using the User's Opinion

Jung-Yeon Yang, Jaeseok Myung, Sang-goo Lee
Department of Computer Science and Engineering
Seoul National University
Seoul, Korea
{jyyang, jsmyung, sglee}@europa.snu.ac.kr

Abstract— **As the number of transactions in E-market places is growing, more and more product information and product reviews are posted on the Internet. Because customers want to purchase good products, product reviews became most important information. But, because of the massive volume of reviews, customers can't read all reviews. In order to solve this problem, a lot of research is being carried out in Opinion Mining. Through the Opinion Mining, we can know about contents of whole product reviews. Traditionally research on Natural Language Processing was applied to the Opinion Mining area in early stage. Recently, the computational statistics are applied to handle massive volume of reviews. In this research, we suggest a method for summarization of product reviews using the user's opinion, feature occurrences, and the rate of review in order to improve the performance of existing methods. With this method, we can handle massive volumes of reviews in a short time efficiently. We guarantee the correctness of the review summary by finding out the semantic meaning of reviews. Besides, we show these advantages through some experiments.**

Keywords-opinion mining; review summary; product review;

I. INTRODUCTION

As the e-business area became active through the Internet, a lot of product information appeared online and customers have more chances to purchase products. But, in the e-market place, people cannot see the concerned product. So, customers depend on the product review, because they regard reviews as the indirect experiences. This situation causes the flood of reviews. Consequently, customers want to see brief summary of reviews. In order to solve these problems, Opinion Mining researches are carried out[1][12][13].

Opinion Mining is the sub-area of mining research area. And it makes semantics of documents clear. In early stage of the Opinion Mining research, in order to find out the semantics of documents, Natural Language Processing (NLP) was applied[1][2]. But, in case of massive data processing, NLP has shortcomings. That is it needs very long time. In order to solve these problems, recently, computational statistics are applied that can handle massive data in a more practical time[7]. But, this approach also has some weak points that it uses only feature occurrences and rate of review. So, it's hard to make semantics of documents clear.

In this paper, we suggest a method for summarization of product reviews using user's opinion, feature occurrences, and the rate of review in order to improve the performance of existing methods. And we define some factors which are used in calculation of product feature score. For each factor, we describe the characteristics and suggest the application for scoring of product features. In order to show summary to customers, we score the product features like a numeric rate review system. Through the application of computational statistic, we can handle massive volume of product reviews. And we guarantee the correctness of the review summary by finding out the semantic meaning of review sentences.

In chapter II, we introduce existing methods for the summarization of product reviews and we describe our method for summary in chapter III. In chapter IV, we define factors which are used in calculation of product feature scores. And we explain the calculations case by case. In chapter V, we show the experiment results. In experiments, we compare the performance error between existing method and our method. In chapter VI, we make a conclusion and suggest future works and some issues.

II. RELATED WORK

A. Extraction of features and opinion words

Unlike the traditional document retrieval methods whose purpose is to extract important keywords, in case of feature extraction on the summarization of product reviews, we select and extract product features in which customers are interested. We use the part-of-speech (POS) tagging in order to extract features which are used in the summarization of product reviews. POS tag is a famous and useful approach. It is helpful to find out the semantics and role of words in sentences. The algorithms of Hu et al[1][4] and Popescu et al[3] apply POS tagging to identify nouns.

Hu et al apply an association-mining tool to identify frequent noun phrases. In order to extract features, Popescu et al use a Point-wise Mutual Information (PMI) value between nouns. Both approaches are based on feature occurrences in review sentences. In case of an opinion word extraction, many existing researches use pattern rules, linguistic model, distance between words, co-occurrence between words and n-gram[8][11][15].

978-0-7695-3531-9/09 $25.00 © 2009 IEEE
DOI 10.1109/eKNOW.2009.15

We also use feature occurrences in the feature extraction. And, in order to extract opinion words, we use pattern rules and n-gram. We have developed the extraction tool which can extract feature words and opinion words massively and automatically.

B. Sentiment Classification

In order to understand semantics of review sentences in the product review, algorithms which use NLP approaches are studied mainly[6][9][10][14]. Many algorithms are based on WordNet, PMI, and association rules. But these algorithms cannot handle massive amount of data in a short time. Moreover, they decide semantics of words with the general usage and meaning of words and depend on expert's manual pre-definitions.

On the other side, we use a probability distribution of opinion words with rates. We classify an opinion word into a positive one or a negative one automatically. As we use the rates of reviews, PMI, and the word cluster through the program, we can classify a sentiment of opinion words.

C. Scoring product features

Some researches use a product scoring for a summarization of product reviews. Scaffidi et al shows a score of product features using the scoring algorithms which use the rate of reviews and feature occurrences[7]. Other algorithms use a set of vocabulary such as WordNet to assign score[1][4][5]. These algorithms use opinion words in order to score a product.

In real review documents, there are several opinions and many features are mentioned. Each feature has an own valuation and rate of review is estimated through the interaction among valuations of features. Rate cannot be the representative value of each product feature. In order to summarize reviews correctly, we have to find out the own valuation of each feature. In our method, we use the rate of review in order to calculate the score of product features. Additionally, we use semantics of opinion words which describes product features.

III. SUMMARIZATION OF PRODUCT REVIEWS

Summarization of product reviews is partitioned into several processes. Figure 1 shows all processes of our summarization method. In this chapter, we explain each process of these processes.

A. Word Extraction

We develop the PicAChoo[16] in order to extract product features and opinion words. PicAChoo is a tool that provides many functions in order to extract specific words. We can get POS tagged words using this tool. It provides several extraction algorithms and we can add other algorithms that we want to apply through the plug-in. In order to extract features and opinion words, we use feature occurrences, 2-gram neighbor and pattern-rules.

B. Sentiment Classification

After the word extraction, we classify the sentiment of opinion words which describe the product feature. In typical existing sentiment classification methods, experts define dictionaries which have positive words and negative words. But it's done manually. In practical cases, we don't know what words will be used in specific case. So it is hard to define dictionaries for every case of word usage.

We don't use dictionaries which are defined by experts manually. In order to find out a sentimental polarity of words, we have to consider the context of words. A word has a different meaning and a different sentimental polarity for different cases.

We have build a context-aware sentiment classification tool which can classify the sentimental polarity of a word in consideration of an own context. This tool uses two kinds of dictionaries in order to decide a sentimental polarity. These dictionaries are constructed automatically using the rate of reviews and review texts. The positive word dictionary is constructed using the review texts which are rated high. Oppositely, the negative one is constructed using the review texts which are rated low.

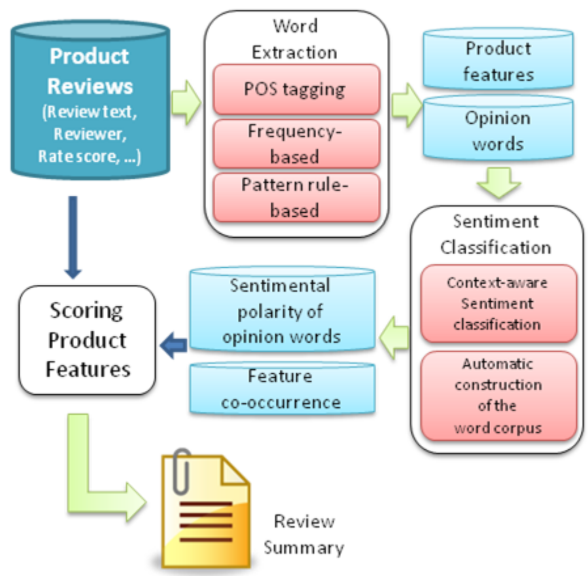

Figure 1. Summary process.

C. Scoring

In our method, the summarization of review is expressed by showing the score of each product feature. We use feature occurrences, rate of reviews, feature co-occurrence information and sentimental polarities of opinion words in order to calculate the score of each product feature. We define some factors which have an effect on the own score of features. And through the experiments, we show the performance of our method. Detailed explanations will be in the next chapter.

IV. METHODS FOR THE SCORING

A. *Scoring factors*

As stated in chapter II.C, in order to score the product feature, we had another viewpoint from the typical methods that uses only the rate of reviews and feature occurrences. We estimate the score of product features with the sentiment of features in the review. Accordingly, we define some factors for the scoring of product features. The score of the product feature is estimated through the combination of rate, feature occurrences, co-occurrence between features and a sentimental polarity of opinion words which describe the product feature. Following are details of each factor.

B. *Using rate of reviews*

Like existing methods, we concentrate on the fact that the rate of a review is a value that shows the valuation of reviewer to the product. That is, the rate of a review is the only one value in which the reviewer expressed her own opinion explicitly. But, it is a valuation to the whole quality of the product. For each product feature, we can't know clearly how the reviewer feels. So, it is incorrect to use the rate as the score of each product feature. For example, let's assume that there is a review system which has 5-scaled rate. Some person gives a rating of 4 to the 'iPOD - Portable MP3 Player'. He likes the size of the player and the quality of the sound. But he thinks that it is a little expensive. So, he rates the products as 4 and writes his opinion like "this player is amazing one. Proper size~ fantastic sound!! . But it's a little expensive ☹". In this example, he reviewed about 3 features; size sound and cost.

When we summarize this review, if we use the rate of review as the score of each feature then 3 feature has the same score. This is not true. So, we have to distinguish features whether it is good features or not. In order to solve this problem, we have to consider other factors which effect the own score of product features.

C. *Using occurrence of features*

When we calculate the score of features, number of occurrence is one of the important factors. If a reviewer mentioned specific product feature in the review then it means that he think that feature is more important than others. So, we have to reflect the occurrence of features as a weight when the score is calculated. The weight is calculated through the following equation.

$$\mathrm{weight(feature)} = 2 - 2^{1-\mathrm{occurrence(feature)}} \quad (1)$$

If the product feature is mentioned, the scope of a weight is from 0 to 2. We use weight as the opinion strength in the scoring equation (5).

D. *Using co-occurrence among features*

With an occurrence of a feature, we concern the co-occurrence between features in order to calculate the score of product features. Additionally, rate of the review is used to estimate the score with co-occurrence information. In order to calculate the score of features, we compute an association value between a feature and a rate using the equation (2).

$$\mathrm{association}(f_i, k) = \sum_{j=1}^{NF} \left(P_{co}(f_i, f_j) * P_{score}(f_j, k) \right) \quad \begin{cases} 1 \le i \le NF \\ 1 \le k \le 5 \end{cases} \quad (2)$$

In the equation (2), f is a product feature and k is grade of rate. NF means the number of product features that we extract. $P_{co}(f_i, f_j)$ is a probability of co-occurrence between features, and $P_{score}(f_j, k)$ is a probability of feature occurrences when the rate of review is k. After we compute an association, we compute the score of product feature that is based on association values of the feature.

$$\mathrm{score}(f_j) = \frac{\sum_{k=1}^{5}(k * \mathrm{association}(f_j, k))}{\sum_{k=1}^{5} \mathrm{association}(f_j, k)} \quad (3)$$

Through above equations, we estimate the score of product features and express in 5-scale. We find out whether the co-occurrence information is a proper factor through experiments.

E. *Using opinion words*

Define Last factor is the usage of opinion words. In order to estimate the own score of product features, we have to consider the opinion words with other factors that are mentioned above. Prior methods use only opinion words using NLP. But those methods don't consider the union of NLP and other approaches.

In order to overcome weak points of NLP for the summarization of product reviews, we use opinion words to score product features with a rate, feature occurrence. At first, we classify the sentimental polarity of the opinion words which describes the product feature. And then, we use the distribution of the sentimental polarity and strength of the opinion. As stated in IV.C, we compute the strength of an opinion using the weight of the feature.

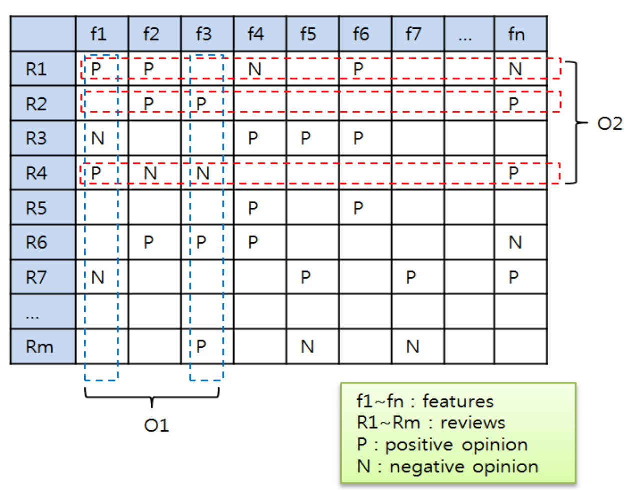

Figure 2. Scoring methods with opinions.

In Figure 2 we show two different cases, O1 and O2, in which we used opinion words. In the product review, reviewer says his opinion for each product feature. And there are many reviews on the same product. In all product reviews, many features are mentioned. But every feature is not mentioned in every product review. Figure 2 shows this situation. In the product reviews, the unit of the opinion is not a document or review, but each product feature. So, we can express the distribution of opinions like Figure 2 using the sentiment classification.

In O1 case, we just consider the distribution of a sentimental polarity of an opinion word on the same product feature. If the number of positive opinions is more than the number of negative opinions then the score of the product feature is high. We use the equation (4) for this case.

$$\mathrm{score(feature)} = (\mathrm{MaxR} - \mathrm{MinR}) * \mathrm{P_{pos}(feature)} + \mathrm{MinR} \tag{4}$$

MaxR and MinR mean the maximum rate and the minimal rate of the review system.

In case of O2 method, we assume that the rate of review is derived from the combination of several opinions about features. Because the rate of review is an explicit score which represents a reviewer's overall opinion, we can use this to score the product feature. For example, in the Figure 2, if the rate of R1 is 4 then it is derived from many positive features and negative features. f4 and fn has a bad influence upon the overall valuation of the product. On the other hands, f1, f2, and f6 may have a higher score than the rate of R1. That is, the positive opinion makes the rate of review high and the negative opinion makes the rate of review down. In order to reflect this phenomenon, we use the equation (5).

$$\begin{aligned}
&\text{if } \mathrm{SP}(\mathrm{f_i}, \mathrm{R_j}) \text{ is } \mathit{positive}, \\
&\mathrm{ownScore}(\mathrm{f_i}, \mathrm{R_j}) = \mathrm{rate}(\mathrm{R_j}) * \left(1 + \mathrm{P_{neg}}(\mathrm{R_j})\right) * \mathrm{weight}(\mathrm{f_i}) \\
&\text{if } \mathrm{SP}(\mathrm{f_i}, \mathrm{R_j}) \text{ is } \mathit{negative}, \\
&\mathrm{ownScore}(\mathrm{f_i}, \mathrm{R_j}) = \mathrm{rate}(\mathrm{R_j}) * \left(1 - \mathrm{P_{pos}}(\mathrm{R_j})\right) * \mathrm{weight}(\mathrm{f_i}) \\
&(1 \le \mathrm{i} \le \text{Num of features}, 1 \le \mathrm{j} \le \text{Num of reviews})
\end{aligned} \tag{5}$$

In the equation (5), $\mathrm{SP}(\mathrm{f_i}, \mathrm{R_j})$ means the sentimental polarity of the $\mathrm{i^{th}}$ feature in the $\mathrm{j^{th}}$ review. $\mathrm{P_{pos}}(\mathrm{R_j})$ and $\mathrm{P_{neg}}(\mathrm{R_j})$ is the percentage of positive opinions or negative opinions in the $\mathrm{j^{th}}$ review. We explained about the weight in chapter IV.C.

We use the percentage of the opposite sentiment in the equation for each sentimental polarity of opinion. Through the usage of this equation, we adjust the rate of the review. The weight of each feature makes adjustment value bigger, because the occurrence of a feature in a review highlights the strength of the opinion on the product feature. Because of the limitation that we use 5-scale rated reviews, the score which was derived from the equation (5) has a scope from 1 to 5. That is, if the score is lower than 1 through the equation then the score is 1.

After the computation of the score of product feature in all reviews, we estimate the representative score of the product feature using the following equation.

$$\mathrm{score}(\mathrm{f_i}) = \frac{\sum_{\mathrm{k}=1}^{\mathrm{NR}} \mathrm{ownScore}(\mathrm{f_i}, \mathrm{R_k})}{\mathrm{NR}} \tag{6}$$

In the above equation, NR means the number of reviews in which the $\mathrm{i^{th}}$ feature was mentioned and R is the review that has a comment about the $\mathrm{i^{th}}$ feature.

V. EXPERIMENTS

A. Data & Evaluation set

In this chapter, we show the performance of the each method through some experiments. We use practical review data from eOpinion.com for experiments. We randomly selected 100 products and 6,000 reviews from the 'Digital Camera' category. We applied each method to those reviews in order to observe the performance of both methods.

Especially, we used the 'Pros' field and the 'Cons' field in order to make the evaluation set. Unlike general review systems, the review data from the eOpinion.com has two important fields; 'Pros' and 'Cons'. In these fields, reviewer writes good points or bad points of the product explicitly. So, we can distinguish whether the product feature is good or not. Consequently, we make the evaluation set with these fields. In many experimental cases, we computed the error for each method using this evaluation set. The evaluation set also has a scope from 1 to 5.

B. Comparison of the performance

We compare the performance of each method in a size of error value. Like Table I, we did experiments in 5 cases. F is the method that uses the feature occurrence and the rate of review. This is the method from the prior research. In this case, opinion words were not used in scoring the product.

TABLE I. CASE OF EXPERIMENTS

Case Symbol	Method for scoring
F (existing method)	In order to calculate the own score of features, they use the feature occurrence & rate.
O1	Use the distribution of sentimental polarities on the same feature over whole reviews
O2	Use the distribution of sentimental polarities in the same review. Calculate the own score of features through the adjustment.
O3	Mean value of O1 and O2
CO	Use the feature co-occurrence

In case of using opinion words, we divide the case in detail. The O1 use the distribution of sentimental polarities on the same feature over all reviews. Only sentimental polarities of the feature were considered. Equation (4) is used for this method. O2 also uses the distribution of sentimental polarities on the feature. But, as stated in IV.E, other factors were considered in order to estimate the score of the product feature. Meanwhile, we observed the combination of the O1 and the O2 in order to find out whether the combination causes a synergy. The CO is the method in which we used the co-occurrence between features in the review. We use the equation (2) and (3) for this method.

As we can see from the results in Figure 3, F method has many errors. Most of product features were estimated in high score. This phenomenon is caused from the facts that the percentage of high-scored reviews much bigger than low-scored ones. Practically, in the review system, high-scored reviews are 5~10 times low as to low-scored ones. So, through this method, it's hard to find out bad features.

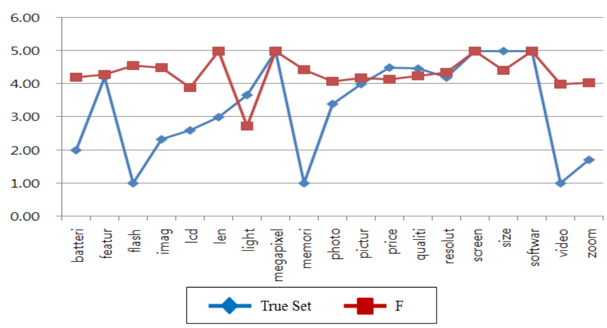

Figure 3. Score distribution of the F method.

In order to compare errors of each method, we made summation of errors per product. It means that the performance is good if the summation of errors is low.

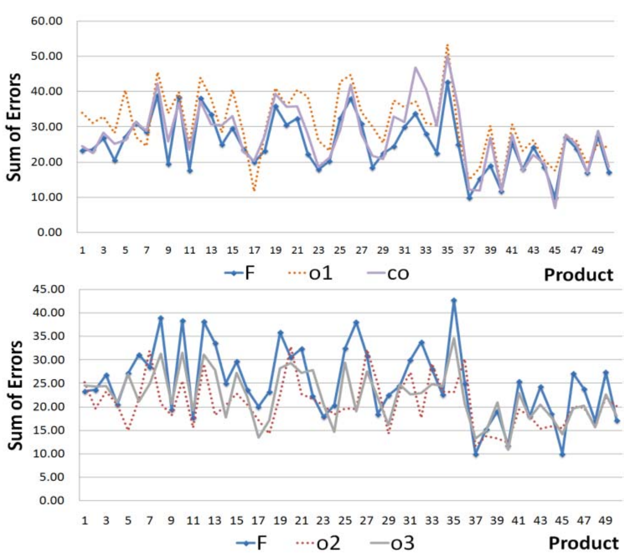

Figure 4. Comparisons with a summation of errors.

As we can see in Figure 4, O1 and CO has more errors than F. On the other hand, O2 and O3 show good performances than F. That is, Only O2 has lower errors than F over the all products.

In order to look upon the result more easily, as shown in Figure 5, we express the error difference between F and others in bar graphs. If the bar is on the positive area, it means the target method has low error rate. That is, O1 and CO has bad performance than F. In case of O2, most bars are in the positive area. But, we still observe that there are also some negative cases in O2.

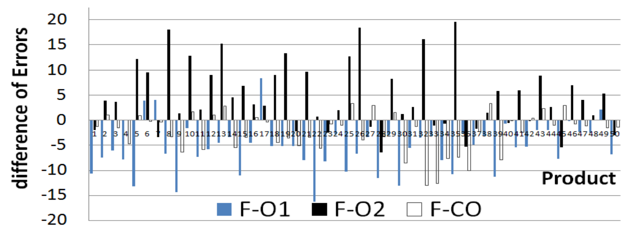

Figure 5. Error difference between two methods.

C. *Combined Method*

Through the experiment, we show that the method O2 is the best among several methods. Moreover we prove that we have to consider many factors which affect on the features. But, O2 has also some cases that show a bad performance. In order to solve this problem, we use F as well as the O2. We estimate the score of product features with the mean-value of scores which are derived from F and O2. After computing the score using the mean-value, we compare the error summation of the F case and F&O2 case. The result is shown in Figure 6.

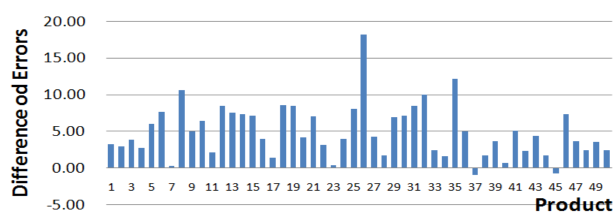

Figure 6. Improvement of combined method

As shown in Figure 6, performance shows the stability in comparison with other results. The volume of bars means reduction in errors in comparison with F. The combined method improves the performance about 20% in comparison with F.

VI. Conclusion and Future Work

In this paper, we suggest a method for summarization using user's opinions in practical cases. Especially, we score the product feature in order to show the quality of the feature to customers. We developed programs for the extraction, the sentiment classification and the scoring product features. We define several factors which affect on the score and find out the optimal method. We cover the shortcomings of prior methods and take good points of them. Through the experiment, we prove the performance of our method.

We will need to use the synonyms for summaries, and find out other factors which affect on the scoring of product features. Additionally, we'll need to analysis and summarize comparative sentences in reviews in order to provide useful information to customers.

ACKNOWLEDGMENT

This research was supported by the Ministry of Knowledge Economy, Korea, under the Information Technology Research Center support program supervised by the Institute of Information Technology Advancement. (grant number IITA-2008-C1090-0801-0031)

REFERENCES

[1] M Hu , B Liu, "Mining and summarizing customer reviews", Proceedings of the tenth ACM SIGKDD 04, August 22-25, 2004.

[2] B Liu , M Hu , J Cheng, "Opinion observer: analyzing and comparing opinions on the Web", Proceedings of the 14th international conference on WWW, May 10-14, 2005.

[3] A Popescu , O Etzioni, "Extracting product features and opinions from reviews", Proceedings of the conference on HLT/EMNLP, p.339-346, October 06-08, 2005.

[4] H Nakagawa , T Mori, "A simple but powerful automatic term extraction method", COLING-02 on COMPUTERM 2002, p.1-7, August 31, 2002.

[5] S Sista, S Srinivasan. "Polarized Lexicon for Review Classification", Proc. Intl. Conf. on Machine Learning, Models, Technologies & Applications, 2004.

[6] K Dave , S Lawrence , D Pennock, "Mining the peanut gallery: opinion extraction and semantic classification of product reviews", Proceedings of the 12th international conference on WWW, May 20-24, 2003.

[7] C Scaffidi, K Bierhoff, E Chang, M Felker, H Ng, C Jin, "Red Opal: Product-Feature Scoring from Reviews", Proceedings of the 8th ACM conference on Electronic commerce, June 11-15, 2007.

[8] X Ding, B Liu, "The utility of linguistic rules in opinion mining", Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, July 23-27, 2007.

[9] J Liu, G Wu, J Yao, "Opinion Searching in Multi-product Reviews", proceedings of the sixth IEEE International Conference on Computer and Information Technology, 2006.

[10] S Patwardhan, S Banerjee, T Pedersen, "UMND1: Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness", proceedings of the 4th International Workshop on Semantic Evaluations, p.390-393, June 2007.

[11] Y Matsuo, M Ishizuka, "Keyword Extraction from a Single Document Using Word Co-Occurrence Statistical Information", Proceedings of 16th International. Florida AI Research Society, p.392-396, 2003.

[12] M Gamon, A Aue, S Corston-Oliver, E Ringger, "Pulse: Mining Customer Opinions from Free Text", international symposium on intelligent data analysis. 2005.

[13] N Archak, A Ghose, P G. Ipeirotis, "Show me the money!: deriving the pricing power of product features by mining consumer reviews", Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining, P.56-65, 2007.

[14] G Mishne, "Experiments with mood classification in blog posts", Stylistic Analysis Of Text For Information Access, 2005.

[15] R Ghani, K Probst, Y Liu, M Krema, A Fano, "Text Mining for Product Attribute Extraction", ACM SIGKDD Explorations Newsletter, vol.8, issue.1, June 2006.

[16] http://ids.snu.ac.kr/ /wiki/PicAChoo_%28Pick_And_Choose%29