

# Opinion Summary Generation for Product Reviews

Dr. (Mrs.) Saruladha. K<sup>1</sup>,  
B.E., M.Tech., Ph.D.,  
Asst. Professor

Department of Computer Science and Engineering  
Pondicherry Engineering College

Banupriya. D<sup>2</sup>, Nargis Banu. J<sup>2</sup>  
Department of Computer Science and Engineering  
Pondicherry Engineering College

**Abstract** - The objective of this paper is to generate opinion summaries from crawled product reviews extracted automatically from Amazon.com and processed using feature-based sparse non negative factorization technique. The required features for generation of summaries are extracted from the online reviews. As the number of online reviews is large, the feature space is huge which would increase the computational complexity of the opinion summary generation algorithm. This necessitates the reduction of feature space which is done efficiently by using SentiWordNet. Feature-Based Sparse Non-Negative Matrix Factorization method (FS-NMF) is proposed to ensemble the reviews into feature relevant clusters which are based on non-negative matrix factorization (NMF) framework. Extensive case studies, experiments, and summary evaluation on ROUGE demonstrate the effectiveness of this system.

**Keywords:** Feature extraction, Opinion mining, feature-based summarization, Non-negative Matrix Factorization.

## I. INTRODUCTION

Vendor selling products on the e-commerce websites often inquire their customers to review the products that they have purchased and the associated services. Since most of the people prefer e-commerce for transactions of commodities, the number of customer reviews for each product grows rapidly which can be in hundreds or even thousands for popular products. This makes it difficult for a potential customer to look up those reviews and to make a clear decision on whether to purchase the product. It also makes it hard for the manufacturer of the product to monitor customer opinions.

This summarization technique is different from traditional summarization techniques such as Centroid, Graph, LSA, and NMF because they only generate random generation of summaries which fails to mine the opinions on product features. The opinion summary generation is carried out as in the following three major steps.

- Crawling reviews automatically from Amazon.com.
- Extracting product features from the reviews that have been commented by customers by constructing term-sentence matrix.
- Summarizing the reviews based on product features.

In this paper, a Feature-Based Sparse Non-Negative Matrix Factorization method (FS-NMF) is proposed to ensemble the reviews into feature relevant clusters. This method is based on Non-Negative Matrix Factorization (NMF) framework. NMF was proposed by Paatero and Tapper [12] but Lee and Seung [7] highly popularized this method by using simple multiplicative update algorithms to perform NMF. In general, the multiplicative update algorithms seem to converge very slowly for large-scale problems. Due to this drawback, there is a need for more efficient and fast algorithms for NMF. Many approaches have been proposed to handle these problems. One of them is to apply projected gradient (PG) update algorithms instead of multiplicative ones. The FS-NMF method is fast and quite efficient for large scale problems subject to non-negativity and sparsity constraints.

## II. RELATED WORK

### A. Mining and summarizing product reviews

Generally, mining and summarizing product reviews involve three tasks: feature extraction, opinion mining, and summarization. The main goal of feature extraction is to identify frequent product features; opinion mining is to mine the customer opinions on product features; and summarization aims to deliver compact results to users.

### B. Opinion Feature Extraction

Hu and Liu (2004) [5] proposed a mining based method to detect product features. First, they used NLP linguistic parser to parse each review and produce the part-of-speech (POS) tag for each word. Then, association miner is applied to the transactions of nouns and noun phrases to identify frequent features. Then, only nouns and noun phrases with nearby adjectives from the frequent features are ranked based on their frequencies and are considered as the opinion features.

Yi and Niblack (2005)[16] defined a set of feature extraction heuristics and select features from the noun phrases obtained from the customer reviews based on a given product. Popescu and Etzioni (2005) [13] proposed an unsupervised information extraction system to acquire frequent candidate nouns by setting a threshold frequency. The frequent candidates are then assessed by mutual information between a candidate and a product class.

A more recent work by Dingding Wang, Shenghuo Zhu and Tao Li (2013)[2] performed Part of speech (POS) analysis on

the sentences. Only nouns and noun phrases are considered to identify set of features. The term frequency-inverse sentence frequency (tf-isf) is then computed and top five features are treated as opinion features.

### C. Summary Generation

Recent works on summary generation usually list all the opinionated sentences. Hu and Liu (2004) [5] returned the number of positive and negative sentences for each extracted product feature. Meng and Wang, (2009) [10] generated the summary with most frequent terms or phrases of each product feature.

Lu, Zhai, and Sundaresan (2009) [9] produced aspect ratings of short comments for each product. Dingding Wang, Shenghuo Zhu and Tao Li (2013)[2] proposed SumView system generated summaries by selecting sentences with highest probability for each product feature.

### D. Document summarization

Following are the related work on different document summarization methods.

#### i. Non-negative Matrix Factorization

Li & Ding, (2006) [8] conducted Non-negative Matrix Factorization (NMF) on the term-sentence matrix to extract sentences with the highest probability in each term. First, it decomposes the term-sentence matrix into term-topic and sentence-topic matrix with random initialization. Then, it chooses the most representative sentences from sentence-topic matrix to form the summary.

#### ii. Feature-based Non-negative Matrix Factorization

The feature-based non-negative matrix factorization method (FNMF) proposed by Dingding Wang, Shenghuo Zhu and Tao Li (2013) [2] is based on non-negative matrix factorization (NMF) framework. Instead of random initialization, this method initializes by taking product features into consideration. While decomposing the term-sentence matrix, multiplicative update algorithm is used to update the corresponding term-topic and sentence-topic matrices. Finally, sentence with highest probability for each topic is included for summarization.

### E. Limitations of Existing Work

Though existing summarization techniques work on feature extraction, opinion mining, summary generation, only smaller datasets are taken for experiments. Also, the generated summaries are not much focused on product features. The main challenging aspects in existing summarization techniques are

- Large feature space.
- Handling sparsity and negativity of data in feature space.
- Computational complexity is more because of huge feature space.
- Identification of relevant and important features.

## III. PROPOSED WORK

### A. System Framework

The high level system architecture of the proposed method is depicted in Fig.1. First of all, a crawler is developed to retrieve product reviews from Amazon.com. Once a product ID (which is the unique product number provided by Amazon.com for each product) is given, all the user reviews and comments for this product are downloaded. Synthetic datasets are also taken for experiments. POS tagger is used to tag each word with parts of speech (JJ for Adjective, NN for Noun, RB for Adverb, etc.). After pre-processing such as removing stop words and stemming, the feature-sentence matrix is constructed in the pre-processing step where each row represents a term and each column represents a sentence. After opinion mining and feature selection process, product features are automatically extracted and the top features are recommended to the users.

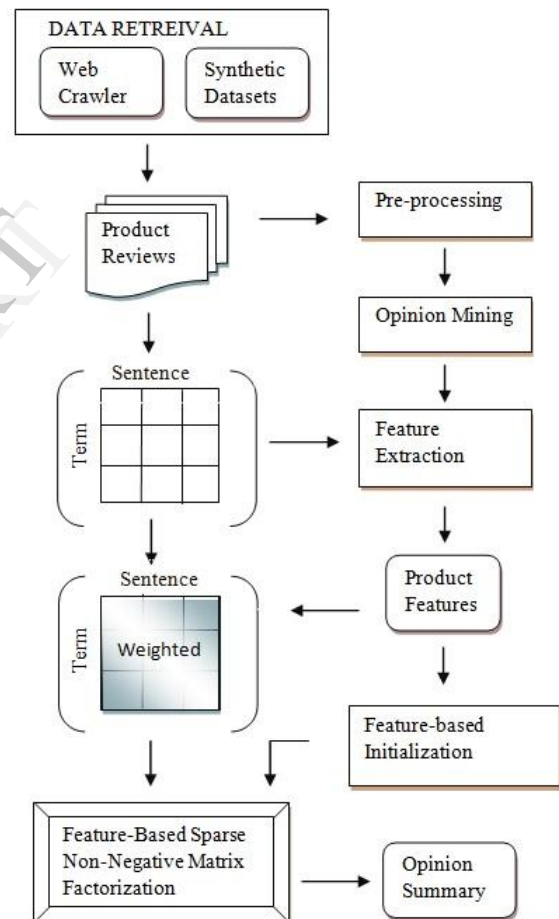


Fig.1. Proposed Architecture

Based on these features, weighted term-sentence matrix is constructed and the proposed Feature-Based Sparse Non-Negative Matrix Factorization is performed to group the review sentences into feature relevant clusters. Finally, the sentence with the highest relevance to each feature is selected as the summary for each feature.

## B. Methodology

### i. Web Crawler

A web crawler is developed to retrieve product reviews from Amazon.com. Once a product ID (which is the unique product number provided by Amazon.com for each product) is given, all the user reviews, review titles, review IDs, customer IDs, real name and purchase verification and ratings for this product are downloaded. Instead of decomposing each review, it is considered as a sentence.

Real name field can be used to check whether those customer reviews are given by true customers. Only those who purchased the product will be able to comment the product relevant to its feature due to their experience in using that product. For this purpose, purchase verification field is considered. Finally, reviews from true customers are only considered for experiments filtering the fake reviews.

### ii. Opinion Mining

Lucene Analyzer is implemented for stop words removal and stemming. This analyzer is used for pre-processing since this analyzes sentences with high-speed. By using Stanford POS tagger, the nouns and nearby adjectives are extracted from each sentence. SentiWordNet is applied to assign scores for each adjective-noun pair. Based on the scores, frequent product features are extracted which are then included in candidate set.

### iii. Term-Sentence Matrix Construction

#### Hadoop MapReduce Implementation:

After pre-processing step, Hadoop MapReduce is implemented to construct a term-sentence matrix since it is specifically designed for storing and analyzing huge amounts of unstructured data. Initially, Hadoop cluster is started using Cygwin terminal to develop a Hadoop MapReduce environment on Eclipse platform. Then, the term-sentence matrix is built in three phases as described below:

#### PHASE 1:

Mapper: Map phase converts an input document into a set of (SentId: term) => occurrence pairs.

Reducer: Sums up the occurrences of (SentId: term) occurrences. Terms which occur below minimum term frequency are removed from consideration.

#### PHASE 2:

Mapper: Extract the term from the key of the previous output, and emit them. This will be used for mapping term and its corresponding position.

Reducer: Aggregates the term count.

#### PHASE 3:

Mapper: Reads the output of step 1 and emits SentId => (term: occurrence) pairs.

Reducer: Converts the output of step 2 into a map of term and position. Finally, term sentence matrix is constructed.

Additionally, in first phase, unimportant sentences are automatically omitted based on minimum term frequency.

### iv. Feature Extraction

The term frequency-inverse sentence frequency (tf-isf) for each term in the candidate set are computed based on the term-sentence matrix to measure its relevant importance in the sentence and top candidates with highest tf-isf scores are kept in the candidate set. They are considered as frequent opinion features.

### v. Feature-Based Sparse Non-Negative Matrix Factorization (FS-NMF)

Feature-Based Sparse Non-Negative Matrix Factorization method (FS-NMF) is proposed to generate opinion summaries. This method is based on Non-Negative Matrix Factorization (NMF) framework with the following additional features:

- It takes product feature into consideration.
- Data are sufficiently sparse.
- Projected gradient update algorithm.

As a result, each topic is related to a feature in factorization results. One of the most useful properties of using sparsity constraints in FS-NMF is that the resulting factorization results are often intuitive and easy to interpret. The computation is also fast and well-fitted for large-scale problems since projected gradient update algorithm is used for updating the factorized matrices. FS-NMF method automatically groups the sentences into feature relevant clusters, and selects the most significant customer opinions in the reviews based on selected product features.

#### ITERATIVE ALGORITHM OF FS-NMF

Input:  $M$ : weighted term-sentence matrix

Output:  $A$ : term-feature matrix

$X$ : feature-sentence matrix

1. initialize  $X$  by selecting the rows in  $M$  where each row corresponds to the selected feature.
  2. initialize  $A := (M(X^T X)^{-1})X^T$   
repeat
  3. Update  $X$ 
    - a. Set  $X$  value based on the difference found in  $gradient(X)$  with each value scaled by the  $learning\_rate(X)$
  4. Update  $A$ 
    - a. Set  $A$  value based on the difference found in  $gradient(A)$  with each value scaled by the  $learning\_rate(A)$
- until convergence

The term-sentence matrix constructed above is used to build a weighted term-sentence matrix. In weighted term-sentence matrix, each element  $a_{ij}$  is computed as follows:

$$a_{ij} = L(i, j) * G(i)$$

where,  $L(i, j)$  is the local weight of term  $i$  in sentence  $j$ ,  $G(i)$  is the global weight of term  $i$  in the document.

The weighting scheme below uses a binary local weight and an entropy-based global weight:

$L(i, j) = 1$ , if term  $i$  corresponds to the selected feature and appears in sentence  $j$  as per minimum term frequency,  
 $L(i, j) = 0$ , otherwise.

$$G(i) = 1 - \sum_j \frac{p_{ij} \log p_{ij}}{\log n}, \quad p_{ij} = \frac{t_{ij}}{g_i}$$

where,  $n$  is the number of sentences in the document,  $t_{ij}$  is the frequency of term  $i$  in sentence  $j$ ,  $g_i$  is the total number of times that term  $i$  appears in the whole document.

In general, FS-NMF is composed of the following steps:

#### a. Feature relevant initialization

Instead of random initializing  $A$  and  $X$  in the NMF algorithm where  $M = AX$ , the decomposed feature-sentence matrix  $X$  is initialized by selecting the rows in  $M$  where each row corresponds to the selected feature, and the term-feature matrix  $A$  is consequently initialized by  $(M(X^T X)^{-1})X^T$ .

#### b. Non-negative matrix factorization

The NMF algorithm is performed on the weighted term-sentence matrix with the feature relevant initialization. Since larger datasets is used, projected gradient update algorithm is used to do the fast computation of updating matrices. This method performs gradient-descent minimization.

While updating term-feature and feature-sentence matrices, projection is done on the resultant matrix to replace all the negative entries by a small positive number to avoid numerical instabilities. It is typically more efficient to choose the learning rates  $\eta^{(X)}$  and  $\eta^{(A)}$  so as to preserve non-negativity of the solutions.

#### vi. Feature-based Opinion Summary Generation

After convergence of the FS-NMF algorithm, the sentence with the highest probability for each feature is extracted from the feature sentence matrix to form the final summary. As a result, this summarization delivers overall opinion about each product feature.

### IV. CASE STUDY

A case study is demonstrated for review summaries generated by different summarization systems. In this case study, 710 reviews of a rice cooker from real customers have been crawled from Amazon.com (product ID: B000G30ESY). This product is chosen because:

- Rice cooker is a good example of a home appliance that people use in their daily life;
- The selected product is one of the best sellers.

The top five product features extracted by the feature identification method are “timer”, “size”, “quality”, “price” and “design”. Table 1 shows the summaries generated from different summarization methods.

From the comparison of the generated summaries, it can be observed that:

- NMF clustering based method performs better than Graph-based method. However, the first sentence is still not closely related to the “quality” feature.
- SumView (FNMF) generated summaries are based on “size”, “quality” and “price” features. However, all the three sentences do not reflect the majority of user opinions on these features. This method considered sentences with maximum feature occurrence for summary generation. This lacks to detect the overall opinion about that feature. Also, the opinions on “size” and “price” in the results of SumView are not focused and representative.
- Comparing with the summarization results using NMF which uses random initialization, it can be seen that the opinions on most of the features in the results of the proposed method are more focused and representative, which indicates the effectiveness of the feature-based initialization in proposed algorithm.
- Comparing with the results of feature-based initialization NMF in SumView, proposed FS-NMF detects the overall opinion about each feature. Because, initially, overall polarity for each feature is identified and based on that, sentences are weighted.

Method	Summary
NMF	<ul style="list-style-type: none"> <li>• What’s more, the rice comes out perfectly and evenly cooked.</li> <li>• You may think the price is a bit steep for a rice cooker, but do not be deterred, it is worth the price.</li> <li>• I love the small size, about the size of a toaster, very easy to clean, just the right size for up to four people with reasonable appetites.</li> </ul>
FNMF	<ul style="list-style-type: none"> <li>• In general, the quality of the cooker is very good.</li> <li>• You may think the price is a bit steep for a rice cooker, but do not be deterred, it is worth the price.</li> <li>• I love the small size, about the size of a toaster, very easy to clean, just the right size for up to four people with reasonable appetites.</li> </ul>
FS-NMF	<ul style="list-style-type: none"> <li>• The timer is convenient, since the cooker doesn’t tell you when the rice will be done if you don’t use the timer.</li> <li>• It is perfectly sized for singles or couples.</li> <li>• It is built from very high quality parts, nothing about it looks cheap, it screams quality.</li> <li>• And with this small one, great value for price.</li> <li>• This product has an excellent design which prevents water from dripping on top of the rice like some rice makers allow.</li> </ul>

Table 1: Comparison of different summarization methods

From this case study, it can be clearly observed that the proposed FS-NMF algorithm can effectively utilize the feature



information and performs better than existing summarization techniques for opinion summary generation.

## V. SUMMARIZATION EVALUATION

In this set of experiments, different summarization methods with FS-NMF are compared using Rouge evaluation toolkit (Lin & Hovy, 2003). Since there is no benchmark data and automatic evaluation tool for the task of review summarization, in order to obtain a subjective score-based performance evaluation, topic relevant multi-document summarization evaluation is used to compare the performance of proposed feature-based initialization FS-NMF with other summarization methods.

### A. Rouge Evaluation

ROUGE toolkit (version 1.5.5) is used to measure the proposed method for performance evaluation. To evaluate the performance of summarization system, two types of summaries are considered. One is the system generated summaries that is referred to as 'system summaries' and the other is reference summaries which is known as 'model summaries'. Generally, model summaries are written by humans and usage of multiple model summaries yields more accurate ROUGE scores than using just one model summary. Since, ROUGE can handle any number of model summaries, many model summaries are considered.

It computes the quality of a generated summary by counting the unit overlaps between the system summary and a set of model summaries. Each of evaluation methods in ROUGE can generate three scores namely, recall, precision and F-measure.

ROUGE-N is an n-gram recall computed as follows

$$ROUGE - N = \frac{\sum_{S \in \{ref\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ref\}} \sum_{gram_n \in S} Count(gram_n)} \quad (1)$$

where,  $n$  is the length of the n-gram,  $ref$  is the model summaries,  $Count_{match}(gram_n)$  is the maximum number of n-grams cooccurring in a system summary and the model summaries, and  $Count(gram_n)$  is the number of n-grams in the model summaries.

### ROUGE-1

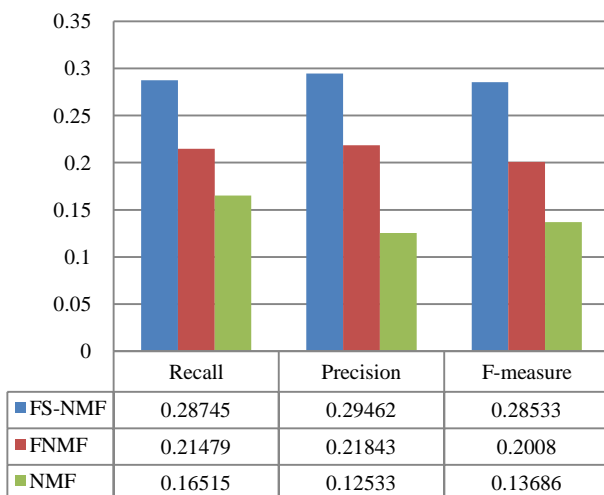


Fig2. Rouge 1

### ROUGE-2

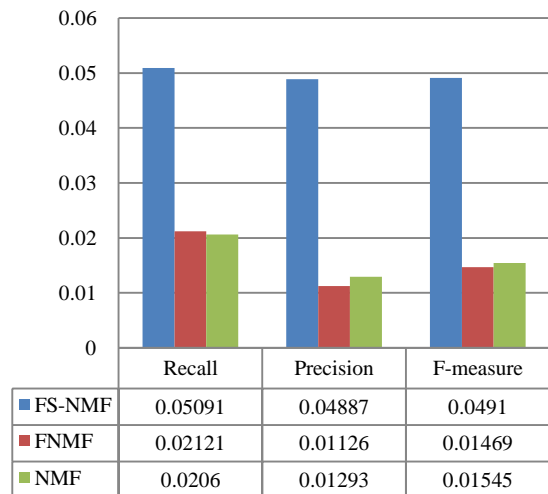


Fig.3. Rouge-2

### B. Experimental Results

Fig.2 and Fig.3 demonstrate the results by different summarization methods. From the experimental results, it can be observed that the feature-based initialization in FS-NMF incorporates the features (keywords) into opinion summarization effectively, so that the performance of FS-NMF outperforms traditional summarization methods and NMF algorithm.

The proposed method is compared with the popular summarization methods instead of some customized query-based summarization methods because those methods put more efforts on semantic analysis on query sentences which deviates the purpose of experimental design. Thus, only widely used general summarization methods is compared with this proposed method to demonstrate the potential of summarization approach and the improvement of the feature-based initialization for the NMF algorithm.

## VI. CONCLUSION

In this paper, the proposed web-based summarization system focuses on opinion summaries generation for all types of opinions that attempts to mine the implied opinions from various reviews and finding the opinion words related to user selected features. This system integrates reviews crawling, opinion mining, product feature identification, and feature-based initialization for efficient feature extraction. The final summary represents the overall customer opinions about various extracted product features.

Also, feature-based summarization method is improved from the existing systems based on proposed feature-based projected gradient sparse non-negative matrix factorization (FS-NMF) algorithm. Since this method handles large datasets, projected gradient update algorithm is used to do the fast computation of updating matrices.

Comprehensive experiments and a case study demonstrate that the proposed system is effective for generating opinion summaries. From the experimental results, it can be observed that the feature-based initialization in FS-

NMF incorporates the features (keywords) analysis into document summarization effectively so that the performance of FS-NMF outperforms traditional summarization methods and the other NMF algorithms. Performance metrics evaluation conducted by using ROUGE resembles the effectiveness of the proposed summarization system.

## VII. FUTURE WORK

This system can be further enhanced by crawling reviews from other e-commerce websites and by allowing user to input the additional desired feature for the product. In addition, if the generated opinion summaries are then analyzed for sentiment classification using Natural Language Processing problem to determine the polarity of opinionated text, the user can still be benefited. So far, some basic ROUGE evaluation techniques (ROUGE-N) have been explored. Hence, further exploration of more advanced ROUGE-S, ROUGE-L, ROUGE-W, and ROUGE-SU techniques using Document Understanding Conference (DUC) benchmark data can be done for topic-relevant summarization evaluation and to demonstrate the accuracy of this method.

## VIII. REFERENCES

- [1] Andrea Esuli and Fabrizio Sebastiani (2006), "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining".
- [2] Dingding Wang, Shenghuo Zhu and Tao Li(2013), "SumView: A Web-based engine for summarizing product reviews and customer opinions", *Journal: Expert System With Application* 40, 23-37.
- [3] Erkan, G., & Radev, D. (2004). "Lexpagerank: Prestige in multi-document text summarization", In *Proceedings of EMNLP*.
- [4] Gong, Y., & Liu, X. (2001), "Generic text summarization using relevance measure and latent semantic analysis", In *Proceedings of SIGIR* (pp. 75–95).
- [5] Hu, M., & Liu, B. (2004), "Mining and summarizing customer reviews", In *Proceedings of SIGKDD* (pp. 168–177).
- [6] K. NATHIYA, and Dr. N. K. Sakthivel (2013), "Development of an Enhanced Efficient Parallel Opinion Mining for Predicting the Performance of Various Products", *International Journal of Innovative Research in Computer and Communication Engineering* Vol. 1, Issue 2.
- [7] Lee, D. & Seung, H. (2001), "Algorithms for non-negative matrix factorization", In *NIPS*.
- [8] Li, T., & Ding, C. (2006), "The relationships among various nonnegative matrix factorization methods for clustering". In *Proceedings of IEEE international conference on data mining* (pp. 362–371).
- [9] Lu, Y., Zhai, C., & Sundaresan, N. (2009), "Rated aspect summarization of short comments", In *Proceedings of the 18th international conference on World Wide Web: WWW '09* (pp. 131–140).
- [10] Meng, X., & Wang, H. (2009), "Mining user reviews: From specification to summarization", In *Proceedings of ACL-IJCNLP*.
- [11] Mihalcea, R., & Tarau, P. (2005), "A language independent algorithm for single and multiple document summarization", In *Proceedings of IJCNLP 2005*.
- [12] Pentti Paatero and Unto Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error", *Environ metrics*, 5:111{126, 1994}.
- [13] Popescu, A.-M., Nguyen, B., Etzioni, O. (2005), "Opine: Extracting product features and opinions from reviews", In *Proceedings of HLT/EMNLP on interactive demonstrations*.
- [14] Rafal Zdunek and Andrzej Cichoki (2008), "Fast Nonnegative Matrix Factorization Algorithms Using Projected Gradient Approaches for Large-Scale Problems", *Hindawi Publishing Corporation*.
- [15] Wan, X., & Yang, J. (2008), "Multi-document summarization using cluster-based link analysis", In *Proceedings of the thirty-first annual international SIGIR conference*.
- [16] Yi, J. & Niblack, W. (2005), "Sentiment mining in web fountain", In *Proceedings of ICDE*.