ISSN 2047-3338

Latent Semantic Analysis Method Used for Mobile Rating and Review Summarization

Jayashri Khairnar[1] and Mayura Kinikar[2]

[1,2]Department of Computer Engineering, MIT Academy of Engineering, Pune University, Pune, India

***Abstract*– With the evolution of web technology, there is a huge amount of data present in the web for the internet users. These users not only use the available resources in the web, but also give their feedback, thus generating additional useful information. Opinion Mining is a process of automatic extraction of knowledge from the opinion of others about some particular topic or problem. Opinion Mining or Sentiment Analysis is a Natural Language Processing and Information. In this paper, design and develop a mobile rating and review summarization in mobile environment using latent semantic analysis (LSA) in support vector machines. Support vector machines and latent semantic analysis is implementing to develop rating and review summarization. Rating and review information is use sentiment analysis and feature based summarization. For classification use support vector machine (SVM), it performs the sentiment classification task also consider sentiment classification accuracy. Propose a Latent semantic analysis (LSA) to identify product features. Latent semantic analysis (LSA) is based on filtering mechanism; also find a way to reduce the size of summary based on the product feature which is obtained from Latent semantic analysis (LSA).**

***Index Terms*– Text Mining, Support Vector Machine (SVM), Sentiment Classification, Feature Extraction and Latent Semantic Analysis (LSA)**

I. INTRODUCTION

TEXT mining offers a way for individuals and corporations to exploit the vast amount of information available on the Internet. In current search engine people to search for other people's opinions from the Internet before purchasing a product or seeing a movie because practically, when we are not familiar with a specific product, we ask our trusted sources to recommend one [7].

Many website provide user rating and commenting services, and these reviews could reflect users' opinions about a product. With the propagation of reviews, ratings, recommendations, and other forms of online expression, online opinion could present essential information for businesses to market their products and manage their reputations. Current search engines can efficiently help users obtain a result set, which is relevant to user's query. However, the semantic orientation of the content, which is very important information in the reviews or opinions, is not provided in the current search engine.

Sentiment classification also goes under different names, among which opinion mining, sentiment analysis, sentiment extraction, or affective rating. Essentially, the task of determining whether a movie review is positive or negative is similar to the traditional binary-classification problem. Given a review, the classifier tries to classify the review into positive category or negative category. However, opinions in natural language are usually expressed in subtle and complex ways. Thus, the challenges may not be addressed by simple text-categorization approaches such as n-gram or keyword-identification approaches.

Approaches are used to classifying texts as positive or negative using Support Vector Machines (SVMs), a well-known and powerful tool for classification of vectors of real-valued features. Sentiment analysis is performed to determine the semantic orientation of the reviews and mobile-rating score is based on the sentiment analysis result. In addition to the accuracy of the classification, system response time is also taken into account in our system design. The same design can be applied to other domains such as restaurant, hotel, etc. Meanwhile, increasingly more cellular phones have begun using global positioning system (GPS) functionality, which can utilize user's current location to provide enhanced services and make cellular phones become context aware. Moreover, the opinion-mining result can be used by recommendation systems to identify which items a user will find worthwhile. In mobile environment, it is inappropriate to display detailed review due to the size of the screen. Hence, employ summarization technique to reduce the size of information.

The system will summarize the reviews (including positive reviews and negative reviews) and provide the user an overview about the reviews. Meanwhile, mobile-review summarization is similar to customer review that focuses on product feature [17]. This summarization task is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative.

In this paper, we employ feature-based summarization for mobile review. Product feature and opinion-word identification are essential to feature-based summarization. Propose a latent-semantic analysis (LSA) based product-feature-identification approach to identify product features. Moreover, extend the result to propose an LSA-based filtering mechanism, which can further reduce the size of the summarization according to the features.

II. RELATED WORK

Text mining offers a way for individuals and corporations to exploit the vast amount of information available on the Internet. As computer networks become the backbones of science and economy enormous quantities of machine readable documents become available. In this paper we used Sentiment Analysis, Feature-Based Summarization, Product-Feature Identification, and Opinion Word Identification modules.

A) Sentiment Analysis

Sentiment analysis of natural language text is a large and growing field. Sentiment analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. Sentiment analysis is a Natural Language Processing and Information Extraction task that aims to obtain writer's feelings expressed in positive or negative comments, questions and requests, by analyzing a large numbers of documents. Converting a piece of text to a feature vector is the basic step in any data driven approach to Sentiment analysis.

Term frequency has always been considered essential in traditional Information Retrieval and Text Classification tasks. But Pang-Lee [2] found that term presence is more important to Sentiment analysis than term frequency. That is, binary-valued feature vectors in which the entries merely indicate whether a term occurs (value 1) or not (value 0).It also reported that unigrams outperform bigrams when classifying movie reviews by sentiment polarity. As a result, the sentiment analysis research from the determination of the semantic orientation of the terms. Determining semantic orientation of words Hatzivassiloglou and McKeown [14] hypothesize that adjectives separated by "and" have the same polarity, while those separated by "but" have opposite polarity.

Starting with small seed lists, this information is used to group adjectives into two clusters such that maximum constraints are satisfied. Sentiment classification is a recent sub discipline of text classification which is concerned not with the topic a document is about, but with the opinion it expresses. Functional to the extraction of opinions from text is the determination of the orientation of "subjective" terms contained in text, i.e. the determination of whether a term that carries opinionated content has a positive or a negative connotation [3].

Esuli and Sebastiani proposed new method for determining the orientation of subjective terms. The method is based on the quantitative analysis of the glosses of such terms, i.e. the definitions that these terms are given in online dictionaries, and on the use of the resulting term representations for semi-supervised term classification. Sentiment classification can be divided into several specific subtasks: determining subjectivity, determining orientation, determining the strength of orientation [3]. Esuli and Sebastiani [5] described SENTIWORDNET, which is a lexical resource in which each WordNet synset is associated with three numerical scores, i.e., Obj(s), Pos(s), and Neg(s), thus describing how objective, positive, and negative the terms contained in the synset.

Traditionally, sentiment classification can be regarded as a binary-classification task [2], [6]. Dave, Lawrence, Pennock [6] use structured reviews for testing and training, identifying appropriate features and scoring methods from information retrieval for determining whether reviews are positive or negative. These results perform as well as traditional machine learning method then use the classifier to identify and classify review sentences from the web, where classification is more difficult. Various supervised or data-driven techniques to Sentiment analysis like Naïve Byes, Maximum Entropy and SVM. Pang Lee [2] compared the performance of Naïve Bayes, Maximum Entropy and Support Vector Machines in Sentiment analysis on different features like considering only unigrams, bigrams, combination of both, incorporating parts of speech and position information, taking only adjectives etc.

It is observed from the results that:

i). Feature presence is more important than feature frequency.

ii). Using Bigrams the accuracy actually falls.

iii). Accuracy improves if all the frequently occurring words from all parts of speech are taken, not only Adjectives.

iv). Incorporating position information increases accuracy.

v). When the feature space is small, Naïve Bayes performs better than SVM. But SVM's perform better when feature space is increased.

According to their experiment, SVMs tended to do the best, and unigram with presence information turns out to be the most effective feature. In recent years, some researchers have extended sentiment analysis to the ranking problem, where the goal is to assess review polarity on a multipoint scale. Goldberg and Zhu [8] proposed a graph-based semi supervised learning algorithm to address the sentiment-analysis task of rating inference and their experiments showed that considering unlabeled reviews in the learning process can improve rating inference performance.

B) Feature Based Summarization

In product-review summarization, people are interested in the reasons why this product is worth buying rather than the principal meaning of the comment. To achieves summarization, numerous techniques using natural language processing (NLP), machine learning, and statistical approaches that can evaluate product features within a collection of review documents. To summarize evaluations for each product feature, sentiment analysis and feature scoring methods have been used. Feature based summarization is used in mobile review summarization. The summarization task [17] is different from traditional text summarization because we only mine the features of the product on which the customers have expressed their opinions and whether the opinions are positive or negative. We do not summarize the reviews by selecting a subset or rewrite some of the original sentences from the reviews to capture the main points as in the classic text summarization. Task is performed in three steps: (1) mining product features that have been commented on by customers; (2) identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative; (3) summarizing the results.

Let us use an example to illustrate a feature based summary. Assume that summarize the reviews of a particular digital camera, digital camera 1. The summary looks like the following: In Figure 1, picture quality and (camera) size are the product features. There are 253 customer reviews that express positive opinions about the picture quality, and only 6 that express negative opinions. The <individual review sentences> link points to the specific sentences and/or the whole reviews that give positive or negative comments about the feature.

Digital_camera_1:

Feature: picture quality

Positive: 253

<Individual review sentences>

Negative: 6

<Individual review sentences>

Feature: size

Positive: 134

<Individual review sentences>

Negative: 10

<Individual review sentences>

…

Fig. 1: An Example Summary

As a result, product features and opinion-word identification are essential in feature-based summarization. Practically, it is not easy to list all the product features and opinion words manually. Some researchers try to use a statistical approach to identify frequent feature words, because the product features may occur frequently in product reviews. However, the drawback of this approach is that it may miss infrequent features. During review summarization at the feature level, feature-opinion pairs are extracted from review documents using part-of-speech (POS) tagging. Hu and Liu [17] studied the problem of generating feature-based summaries of customer reviews of products sold online and proposed a method of word attributes, including occurrence frequency, part of speech (POS), and synset in Word- Net.

During review summarization at the feature level, feature-opinion pairs are extracted from review documents using part-of-speech (POS) tagging. Through the POS tagging, the most frequent noun words that describe the feature within the set of reviews are selected. Parts of Speech information is most commonly exploited in all NLP tasks. One of the most important reasons is that they provide a crude form of word sense disambiguation. Adjectives have been used most frequently as features amongst all parts of speech. A strong correlation between adjectives and subjectivity has been found.

Although all the parts of speech are important, people most commonly used adjectives to depict most of the sentiments and a high accuracy have been reported by all the works concentrating on only adjectives for feature generation. Pang Lee et al. (2002) [2] achieved as accuracy of around 82.8% in movie review domains using only adjectives in movie review domains. Zhuang et al. [10] proposed to make use of grammatical rules and keyword lists to seek for feature-opinion pairs and generate feature-based summarization. Lu et al. [11] utilized POS tagging and chunking function of the OpenNLP2 toolkit to identify phrases in the form of a pair of head term and modifiers. Their research focused on short comments; therefore, POS-tagging information can be employed to obtain the product features and opinion words .For example, the comment ”Fast ship and delivery” contains only one sentence; therefore, it is easier to obtain the head terms (i.e., noun or noun phrase) and modifiers (i.e., adjective) using POS-tagging information. Practically, this approach cannot be applied to other product-review applications.

First, most reviews contain many sentences rather than short comments. Second, most sentences in a review often contain many terms that are irrelevant to the product features or opinion words. Thus, we cannot identify the product features and opinion words in mobile reviews using the same approach. From survey, the problem of system response time in mobile environment, find a way to reduce the size of summary based on the product features obtained from LSA.Consider both sentiment classification accuracy using support vector machines and system response time to design the system in Mobile environment.

III. PROGRAMMER'S DESIGN

Support vector machines were introduced in [4] (Vapnik) and basically attempt to find the best possible surface to separate positive and negative training samples. Support Vector Machines (SVMs) are supervised learning methods used for classification. First module is sentiment analysis and Support vector machines perform sentiment classification task on review data. The goal of a Support Vector Machine (SVM) classifier is to find a linear hyperplane (decision boundary) that separates the data in such a way that the margin is maximized. Look at a two class separation problem in two dimensions like the one illustrated in Figure 2, observe that there are many possible boundary lines to separate the two classes. Each boundary has an associated margin. The rationale behind SVM's is that if we choose the one that maximizes the margin we are less likely to misclassify unknown items in the future.

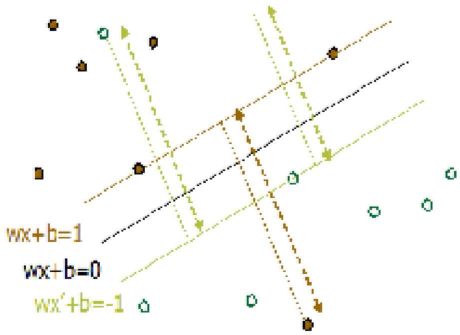

Fig. 2: Different boundary decisions are possible to separate two classes in two dimensions

LIBSVM is a well-known library for SVM that is developed by Chih-Chung Chang and Chih-Jen Lin. LIBSVM is a library for Support Vector Machines (SVMs).LIBSVM is an integrated software for sup-port vector classification, (C-SVC, nu-SVC), regression (epsilon-SVR, nu-SVR) and distribution estimation(one-class SVM) [16].

It supports multi-class classification. LIBSVM involves two steps: first, training a data set to obtain a model and second, using the model to predict information of a testing data set. SVM procedure includes Transform data to the format of an SVM package, Conduct simple scaling on the data, Select model here use linear formula, Use cross-validation to find the best parameter, Use the best parameter to train the whole training set and Test. SVM take more time because it use not related features. To remove system response time issue use Latent semantic analysis method. LSA feature vector can be used as input features to train SVMs.

A) Latent Semantic Analysis Method

Latent semantic analysis is utilized to transform to a smaller feature space for classification with linear support vector machines. LSA is a fully automatic mathematical / statistical technique for extracting and inferring relations of expected contextual usage of words in passages of discourse. Propose a novel approach based on LSA to identify related product-feature terms. Essentially, LSA is a theory and method to analyze relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. That work was interested in addressing the issues of synonymy (there are many ways to refer to the same idea) and polysemy (most words have more than one distinct meaning). Supposing that a collection of documents $D = \{d_1 \dots d_n\}$ with terms from $W = \{w_1 \dots w_m\}$ are given, then the system can construct a co-occurrence matrix M, where its dimension is $n * m$ and each entry M_{ij} denotes the number of times the term w_j occurred in document di. Each document di is represented using a row vector, while each term w_j is represented using a column vector.

LSA use singular value decomposition Singular Value Decomposition (SVD) is a method that separates a matrix into three parts; left eigenvectors, singular values, and right eigenvectors. It can be used to decompose data such as images and text. Singular Value Decomposition is a powerful technique for dimensionality reduction. It is a particular realization of the Matrix Factorization approach.

LSA applies singular-value decomposition (SVD) to the term-document matrix M, and a low-rank approximation of the matrix M could be used to determine patterns in the relationships between the terms and concepts contained in the text.

$$M = U\Sigma V^T \qquad (1)$$

In Eq. (1) the matrices U (n*t) and V (t*m) have orthonormal columns and ∑ (t*t) is the diagonal matrix of singular values as shown in Fig. 3:

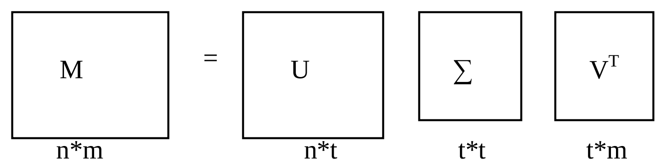

Fig. 3: SVD of term by document matrix M

The original term-document matrix could be approximated by reducing the dimensions of the term-document space, and this will allow the underlying latent relationships between terms and documents to be exploited during searching. Eq. (2) shows that the reduced matrix M is obtained by reducing the dimensionality, where the system truncates the singular-value matrix Σ to size k. It is dimensionality-reduction step.

In Eq. (2) the matrices U (n*k) and V (k*m) have orthonormal columns and ∑ (k*k) is the diagonal matrix of k singular values as shown in Fig. 4:

$$\tilde{M} = U\Sigma V^T \approx U\Sigma V^T = M. \qquad (2)$$

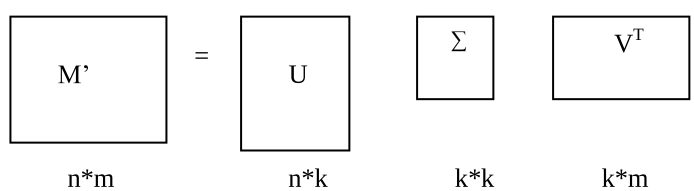

Fig. 4: SVD after dimensionality-reduction step

Fig. 5 shows the algorithm, where the inputs include a term-document matrix, several product-feature seeds, the reduced dimensionality in SVD operation, and the number of extracted features for each seed. Perform linear algebra SVD operation on the term-document matrix, after that compute the similarities between the seed product-feature vector and, pairwise, the other term vectors. The top ones will be collected as related product-feature terms for a specific product feature. Here we used getTermVectorFromTerm DocMatrix procedure to obtain the term-vector representation of a product feature. The seed is supposed to be one of the terms in the term-document matrix, and it is easy to obtain its corresponding document-vector representation. After that similarity list is used to store the similarities between the seed and the other terms then get after sorting in descendant order, get top ones and their corresponding feature names using getTopRelatedFeatures procedure.

Input: An n*m term document matrix Reduced dimension k, the number of extracted features for each seed n, product feature seed set s.

Output: Array F, where each key represents a product feature seed f and f's related product feature.

Begin
Initialize associated array F
$U, \sum, V^t \leftarrow$ svd (M, k)
$M \leftarrow U^*\sum^* V^t$
for f € S do
    Wf $\leftarrow$ getTermVectorFromTerm DocMatrix(f ,M)
    Initialize similarity list sim
    i $\leftarrow$ 1
    for
        column of term w of from reduced matrix M
    sim[i] $\leftarrow W_{f\,*}W$
    i $\leftarrow$ i+1
    end
  sort(sim)
    getTopRelatedFeatures(sim,n,reduced matrix M)
    F[f] $\leftarrow$ relatedFeatureList
end
Return F
end

Fig. 5: LSA Algorithm

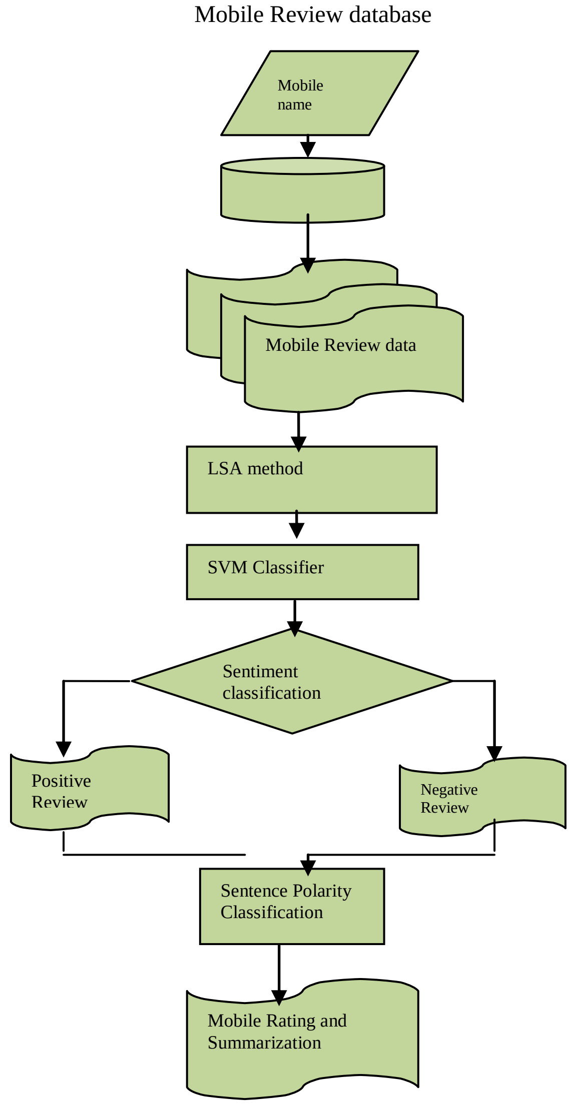

Fig. 6: Mobile review and summarization flow

Each product-feature seed can have its own semantically related term set. The advantage of this approach is that it could be applied to all the languages; it does not need any external dictionary, since LSA is language-independent, and it is based on linear algebra SVD operation. LSA feature vector can be used as input features to train SVMs.

B) Data Flow Architecture

Fig. 6 shows the data flow. The input is a mobile name and based on that mobile name retrieve reviews about this mobile from mobile-review database. These mobile reviews inputs of the LSA method for product feature identification. After that this is input to the train SVM sentiment classifier, which will classify the reviews into positive or negative classes. In addition to the sentiment classification of mobile review, further determine the polarity of a sentence using opinion words. Then, the system can provide both positive and negative summarization, regardless of the polarity of a review. The whole process includes LSA method, sentiment classification and feature-based summarization.

IV. SYSTEM MODULES

A) Dataset

Collect mobile reviews from Internet Blogs. Since the original data are a hypertext markup language (HTML) document, HTML-tag-removal process is required to extract the text information. Training data are necessary for SVM to train a classification model, and manual classification is performed to classify the training reviews into positive or negative reviews.

B) Sentiment Classification

Many classification algorithms such as SVM [1], [4] decision trees, and neural networks [15] have been proposed and shown their capabilities in different domains. The goal of a Support Vector Machine (SVM) classifier is to find a linear hyper plane (decision boundary) that separates the data in such a way that the margin is maximized. In natural-language processing (NLP) and information retrieval (IR), bag-of-words model tries to use an unordered collection of words to represent a text, disregarding grammar and even word order. In other words, each word in the text contributes to a feature of the document. In this paper, employ similar approach to construct a feature vector of the document. Stop words are removed first and then each distinct word Wi in the document is used to represent a feature.

As a result, a document could be represented by a feature vector, and many machine-learning algorithms could be applied to perform classification tasks. SVM perform the classification and libsvm [16] package is used in the system.LIBSVM is a library for Support Vector Machines (SVMs).Support Vector Machines (SVMs) are a popular machine learning method for classification, regression, and other learning tasks. Since the year 2000, developing the package LIBSVM as a library for support vector machines. The Web address of the package is at http:// www. Csie.ntu.edu .tw/~cjlin/libsvm. LIBSVM is currently one of the most widely used SVM software.

C) Review Summarization

Product Feature Identification: Propose an LSA-based algorithm and system can obtain a semantically related feature set for each seed. Compared three product-feature-identification approaches, i.e., the LSA-based approach, frequency-based approach, and PLSA-based approaches [7].

Opinion Word Identification: This step is to identify sentences in the reviews that express opinions about these features. This involves distinguishing opinion words from factual words (subjectivity recognition). Opinion words about the product features are important in review summarization. Hu and Liu [17] summarization task is different from traditional text summarization. It perform mining product features that have been commented on by customers, identifying opinion sentences in each review and deciding whether each opinion sentence is positive or negative and summarizing the results. In addition to language sentence-structure characteristic. Zhuang et al. [10] used the dependency grammar graph to find out some relations between feature words and the corresponding opinion words in training data. They both rely on language sentence structure to extract opinion words. For opinion words first take into account POS-tagging information of the opinion words. Second, term frequency is taken into account; therefore, frequency of the opinion words should exceed a threshold value. Let AVG be the average of sum of square of frequency of all items as shown in Eq. (3). A $term_i$ will be selected only if its square of frequency is equal or larger than AVG. Positive opinion words and negative opinion words could be further obtained based on term frequency and POS tagging.

$$S_f = \sum_{i=1}^{n} \{\mathbf{Frequency}(\mathbf{term}_i)\}^2$$
$$\mathbf{AVG} = S_f / n. \qquad (3)$$

Feature Based Summarization: In product-review summarization Feature-based summarization is more appropriate. Feature-based summarization is based on product features and opinion words, but it is not easy to use compression ratio directly, since the sentence-selection criterion is based on the presence of product features. Use an LSA-based filtering approach to further select the content of the summary. LSA use to find out related feature terms of a specific product feature, and these related terms could be regarded as being semantically related to this product feature. For each given product feature f, LSA could discover related terms F that are semantically related to f. In general, F could be regarded as f's related terms, and the system can employ F to select summary sentences. When the user determines f, the system will generate a summary, which is related to product features F.

Practically, a positive mobile review may include negative comments about specific aspects and vice versa. In this paper, propose to analyze the polarity of a mobile review using SVM and analyze the polarity of a sentence using opinion words. In feature-based summarization, the system can utilize the polarity of opinion words to determine the polarity of sentences. Hence, the system can provide both positive- and negative-review summarization, regardless of the polarity of a review. With the proportion of positive and negative reviews, the system could provide the rating information to end users. The rating information combined with review summary could give end users the rating and summarization information about the mobile.

V. CONCLUSION

Design and implement a LSA method used in SVM for mobile rating and review summarization. Sentiment classification is applied to the mobile reviews, and used SVM for sentiment-classification. Product features and opinion words will be used as the basis for feature-based summarization. In feature-based summarization, product-feature identification plays an essential role, and LSA is use to identify related product features and identify GPS location. We will use frequency criterion to reduce the number of features, and it can takes less time to load the SVM model and classify the reviews. Furthermore, propose an LSA-based filtering approach to reduce the size of the summary.

In future extend it other product-review domains. The same design can be applied to other domains such as restaurant, hotel, and etc. In future our work can extend to achieve greater efficiency in LSA method.

REFERENCES

[1] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques,"inProc.ACL-02Conf.Empirical Methods Natural Lang. Process., 2002, pp. 79–86.

[2] P. D. Turney, "Thumbs up or thumbs down?: Semantic orientation applied to unsupervised classification of reviews," in Proc. 40th Annu. Meeting Assoc. Comput. Linguist., 2002, pp. 417–424.

[3] A. Esuli and F. Sebastiani, "Determining the semantic orientation of terms through gloss classification," in Proc. 14th ACM Int. Conf. Inf. Knowl. Manage., 2005, pp. 617–624.

[4] V. N. Vapnik, The Nature of Statistical Learning Theory.New York: Springer-Verlag, 1995.

[5] A. Esuli and F. Sebastiani, “SENTIWORDNET: A publicly available lexical resource for opinion mining,” in Proc. 5th Conf. Lang. Res. Eval.,2006, pp. 417–422.

[6] K. Dave, S. Lawrence, and D. M. Pennock, “Mining the peanut gallery:opinion extraction and semantic classification of product reviews,” in Proc. 12th Int. Conf. World Wide Web, New York: ACM, 2003, pp. 519–528.

[7] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, “Movie Rating and Review Summarization in Mobile Environment”,IEEE VOL. 42, NO. 3, MAY 2012

[8] A. B. Goldberg and X. Zhu, “Seeing stars when there aren’t many stars:Graph-based semi-supervised learning for sentiment categorization,” in Proc. TextGraphs: First Workshop Graph Based Methods Nat. Lang. Process, Morristown, NJ: Assoc. Comput. Linguist. 2006, pp. 45–52.

[9] B. Snyder and R. Barzilay, “Multiple aspect ranking using the good grief algorithm,” in Proc. HLT-NAACL, 2007, pp. 300–307.

[10] L. Zhuang, F. Jing, and X.-Y. Zhu, “Movie review mining and summarization,” in Proc. 15th ACM Int. Conf. Inf. Knowl. Manage. 2006, pp. 43–50.

[11] Y. Lu, C. Zhai, and N. Sundaresan, “Rated aspect summarization of short comments,” in Proc. 18th Int. Conf. World Wide Web, New York: ACM,2009, pp. 131–140.

[12] T.K.Landauer, P.W.Foltz, andD.Laham, “Introduction to latent semantic analysis,” Discourse Processes, vol. 25, pp. 259–284, 1998.

[13] .T. Joachims, Learning to Classify Text Using Support Vector Machines:Methods, Theory and Algorithms.Norwell, MA: Kluwer, 2002.

[14] V. Hatzivassiloglou and K. R. McKeown, “Predicting the semantic orientation of adjectives,” in Proc. 8th Conf. Eur. Chap. Assoc. Comput.Linguist., Morristown, NJ: Assoc. Comput. Linguist., 1997, pp. 174–181.

[15] G. P. Zhang, “Neural networks for classification: A survey,” IEEE Trans.Syst., Man, Cybern. C, Appl. Rev., vol. 30, no. 4, pp. 451–462, Nov. 2000.

[16] (2001). LIBSVM: A library for support vector machines [Online]. Available: http://www.csie.ntu.edu.tw/ cjlin/libsvm.

[17] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proc.10th ACM SIGKDDInt. Conf. Knowl. Discov. Data Mining, 2004, pp.168–177.

Jayashri Khairnar received her Bachelor’s degree in Information Technology. Now; she is pursuing her M.E degree in Computer Engineering from MIT Academy of Engineering, Pune University, Pune, India. Her research areas are Data mining and Text mining, (Email: jaynit15@gmail.com)

Prof. Mayura Kinikar, B.E., M.E. Computer was educated at Doctor Babasaheb Ambedkar Marathwada University. Now, she is pursuing her PhD. She has worked in various capacities in academic institutions. Now she is Assistant Prof in MIT Academy of Engineering, Alandi, Pune. Her areas of interest include Data mining, text mining, web mining and warehousing, (Email: mukinikar@comp.maepune.ac.in)