

## REVIEW SUMMARIZATION WITH MACHINE LEARNING LANGUAGE

Uma Bhavsar, Yogita Kajabe, Shubhangi Patil,  
Pratima Shinde, Bhushan S. Chaudhari  
Information Technology  
Sandip Institute of Technology and Research Centre, Nashik, India

### ABSTRACT

Now a days e-commerce becoming more popular, Seller selling product on the web often ask their customers for reviews about products which they have purchased and these reviews are used by other people to know features of the product. In this paper, we design and develop, review summarization system with Machine Learning Language. The goal of review summarization system is to take out the product features on which people write their opinions and then identifying whether the opinions are positive, negative or neutral. This system is based on the sentiment classification and portal algorithm. It differentiates the reviews into positive, negative and neutral. Latent Semantic Analysis (LSA) is used to identify features of the product. Portal algorithm is used to classify the reviews. In proposing system we display total percentages of reviews in graphical representation.

**KEYWORDS:** Sentiment classification, Latent Semantic Analysis (LSA), SVM, reviews.

### I. INTRODUCTION

Recently, the popularity of the internet is increasing day by day. So people give priority to online purchasing. Generally people are not aware about quality of product, they trusted on different sources to recommend one. Opinions provide essential information for business to improve quality of the product and maintain their reputation. In order to achieve customer's satisfaction and to full fill customer's requirements, reviews play an important role for sellers. Maximum numbers of people are comfortable with online purchasing. So, the numbers of reviews written by people are also increasing. As a result, the number opinions receive to particular product are increasing rapidly. Sometimes many reviews are very long, but it contains only a few sentences that represent opinion about the product feature. This is very difficult for user to read reviews and making decision about the product. Due to this type of reviews not only customers but also product manufactures face many problems. Product manufacturers cannot keep the track of customers opinions thus it is difficult to understand for manufacturing what changes will be requires in product. Product manufacturing process depends on the customer's requirements and satisfaction which is identified by the reviews written by customers. In the current system, many websites provide the people's reviews, but it doesn't provide the classification of reviews into the positive and negative reviews. Therefore, it is difficult for people to understand the quality of a particular product. Suppose, there is thousands of reviews are available for one product it is very time consuming to read that all reviews. Instead of that if the system provided positive and negative reviews differently and graphical presentation of the total percentages of the positive and negative reviews it makes easier for people to know about the features of the product. The system summarizes reviews into the positive and negative review and gives the overview about the reviews of products to the user.

We design the review summarization system with Machine Learning Language. We use sentiment classification, Latent Semantic Analysis (LSA) and a portal algorithm for classification of the reviews in this system. Sentiment classification is also known as opinion mining, sentiment analysis and

sentiment extraction. In this system we use the user's opinion about the product. This system involves three steps:

- (a) Enter url and Collect customers review.
- (b) Identifying review sentences that are positive or negative reviews.
- (c) Display graphical presentation of result.

For example, www.flipcart.com web site will return around the 1000 reviews for the particular product. So, any person wants to purchase an online product at that time he/she refer reviews which are available on the web site. There are 1000 reviews, out of which 80% are may be positive and 20% are negative. In the proposed system, we represent the classification of total percentages of positive, negative and neutral reviews in graphical form. Section II, describes related work. In section III, various machine learning approaches are described which are required for review summarization. In section IV, the conclusion is described.

## **II. RELEATED WORK**

Current search engine can effectively help people to obtain a result, which is related to their query. However, sentiment analysis is very important for review summarization, is not provided in current search engine. Due to this it is difficult for users to identify the quality of the product.

Hu and Liu's made the use of a priory algorithm to extract the frequent feature words. Opinion words are identified by NLP linguistic parser (tagger) to make the score list and from that list decide the review is positive, negative or neutral [2].

Pang et al. used naive Bays, support vector machines (SVMs) and entropy classification to perform sentiment classification on movie reviews. According to their demonstration SVM tends to do the best and unigram with presence information turns out to be producing intended results on features [3].

D.Laham used rank sentence for summarization. In this first create sentence matrix, in which each column contains weighted term frequency of sentence in the document and row represent the number of times that term occurred in a document. Then to derive the latent semantic structure the single value decomposition (SVD) is performed on the matrix. Then sentences with greatest weight are included in the summaries [7].

## **III. PROPOSED WORK**

The input to this system is the url which is selected by the user. The system will use the url of product and retrieve reviews about that product. Then different machine learning approaches are performing on reviews for classification of that reviews into positive, negative or neutral reviews. These processes will be described as follows.

### **3.1 Sentiment Classification**

Sentiment classification is use for review summarization. It reduces the size of text by removing unwanted characters. If there are any noisy sounds are present in the sentences which are written by the users, then it is removed by using sentiment classification.

### **3.2 Latent Semantic Analysis**

After removing unwanted characters Latent Semantic Analysis (LSA) is to be performed on the text. LSA is used for extracting and deducing the relationship of expected word in the text. LSA check any short form of word is present in the sentence, if short form is present in the text it will compare with the database. The database contains full form and meaning of short forms. This approach is used to count positive and negative term present in the text. As Positive terms are present more than negative terms in the review then it considered as positive review otherwise it's considered a negative review.

### **3.3 Classification**

In this step classification of reviews is done in three categories as positive, negative and neutral. Portal algorithm is used for classification of reviews in which each word in the sentence is compared with the database. The Database contains that only that words which shows the positive or negative orientation of word. As, the word is found positive after comparison, then it increases the counter in

positive column and if the word is negative then counter in negative column is increased. In such case that the word is not positive neither negative, then it considers as neutral. The suffix stripping process is also performing by Portal algorithm. It will reduce the total number of terms in the review, and hence reduce the size and complexity of the text, which is always advantageous. After that clustering is performed to put reviews in a corresponding block, then the graph is generated according to the result which is obtained after completion of whole processes.

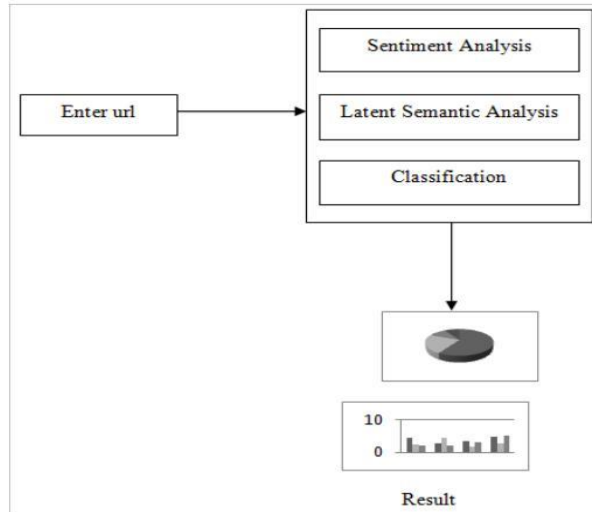


Fig: Block diagram for Review Summarization system.

Let us consider an example to illustrate a review summarization. Assume that the review of the movie is as, “\*Movie\_is\_gd\*!”. Then identifying that this is positive, negative or neutral review as:

**A) Sentiment Classification**

In Sentiment classification it checks for noisy sounds and if such a noisy sounds are found they are removed. In above sentence there are noisy sounds like ‘\_’, ‘\*’, ‘!’ are present, so that are removed in the first step. The result after performing sentiment classification is as “Movie is gd”. Means it reduces the size of text.

**B) Latent Semantic Analysis**

After completion of sentiment classification LSA is performed. In LSA it checks whether any short form of word is present in a sentence, if short form is found it checks to the database. The database contains the full form of the word and its meaning. As considering above example in this step ‘gd’ is found as short form so it checks in the database and replace it. LSA gives the result after completion as “Movie is good”. Means gd is replaced by good.

**C) Classification**

In this step every word in a sentence is compared with a database and according to that identified whether the review is positive, negative or neutral. In above example first word is ‘Movie’ so it compares with the database but the database, not contains that word so it takes another word. In this way all words in the sentence are compared with a database. During comparison the word ‘good’ is found in the positive column of the database so it identified that the sentence is positive, so the positive counter is increased. The result of this step is as.

Positive	Negative	Neutral
0	0	0
+1=1	0=0	0=0

**IV. CONCLUSION**

We design and implement a product review-summarization system. With this system we identify that the reviews are positive, negative or neutral and also display the graphical presentation of result.

Product based summarization is done by using machine learning approaches like Sentiment classification, Latent semantic analysis and portal algorithm.

## REFERENCES

- [1] C. Liu, W. Hsaio, C. Lee, G. Lu & E. Jou, “ Movie Rating and Review Summarization in Mobile Environment ”, in IEEE VOL. 42, May 2012, pp. 397-407.
- [2] M. Hu & B. Liu, “ Mining & Summarizing Customer Reviews ”, in Proc. 10<sup>th</sup> ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2004, pp. 168-177.
- [3] B. Pang, L. Lee & S. Vaithyanathan, “ Thumbs up? Sentiment Classification using Machine Learning Techniques ”, in Proc. ACL-02 Conf. Empirical Methods Natural Lang. Process, 2002, pp. 79-86.
- [4] P. Turney, “Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews ”, in Proc. 40<sup>th</sup> Annu. Meeting Assoc. Comput. Linguist, 2002, pp 417-424.
- [5] T. Joachims, “ Learning to Classify Text Using Support Vector Machines: Methods, Theory & Algorithms ”, in Norwell, MA : Kluwer, May 2002, pp. 224.
- [6] D. Shaw, “ Opinion Mining Of Movie Reviews ”.
- [7] T. Landauer, P. Foltz & D. Laham, “ Introduction to Latent Sematic Analysis ”, in Discourse Processes, Vol. 25, 1998, pp. 259-284.
- [8] T. Mullen & N. Collier, “ Sentiment Analysis using Support Vector Machines with diverse information sources ”, in Proc. EMNLP, 2004, pp. 412-418.
- [9] M. Hearst, S. Dumais, E. Osman, J. Platt & B. Scholkopf, “ Support Vector Machines, Intelligent System & their Applications ”, in IEEE, 13(4), Jul/Aug, pp. 18-28.
- [10] S. Basheer & S. Farook, “ Movie Review Classification & Feature based Summarization of Movie Reviews ”, in International Journal of Computer Trends & Technology (IJCTT), Vol. 4, Issue 7- July 2013, pp. 2235-2339.

## AUTHORS

**Bhawsar Uma C.** currently pursuing the Bachelor of technology from university of Pune and in the fields and area of interest in software testing.



**Kajabe Yogita B.** currently pursuing the Bachelor of technology from university of Pune and in the fields and area of interest in software testing.



**Patil Shubhangi S.** currently pursuing the Bachelor of technology from university of Pune and in the fields and area of interest in software testing.



**Shinde Pratima S.** currently pursuing the Bachelor of technology from university of Pune and in the fields and area of interest in software testing.

