International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017),e-ISSN:2455-2585
International Conference on
Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume-5, Special Issue-March, 2019.

Contextual Sentiment Analysis Using Latent Semantic Indexing

S.Vijaya

Assistant Professor
Department of Information Technology, KG College of Arts and Science, Saravanampatty.

Abstract:

Generally people share their opinions in all aspects. These opinions and ideas may be used for to enhance the business, improve a product, and so on. The upsurge of Web 2.0 has led maximum of these kind of opinion shared in Blogs, Twitters, Face book and so on. These data can be taken for analysis to find out the nature of the opinion like positive, negative or neutral. To analyze these opinions computational treatment of the opinions, sentiments and bias of text which is known as Sentiment Analysis need to be performed. Sentiment Analysis is also known as Opinion Mining. In general, keywords based variations are used to extract the sentiments. In case of the sentences bearing implicit sentiments, the efficiency of this method becomes very less as finding implicit sentiments are difficult in this kind of approach. Comparatively, Latent Sentiment Analysis with SVD improves the efficiency of Contextual based Sentiment Analysis in great manner.

Keywords: Opinions, Positive, Negative, Sentiment Analysis, Contextual Sentiment Analysis

I. INTRODUCTION

In the recent era Sentiment Analysis or Opinion Mining plays a vital role in the field of research as the opinions of other play major role in many aspects like enhancing business, improving a product, to have an idea about movie and so on[1] . The opinions may be expressed in terms of positive comments or negative comments or neutral. For example, in case of knowing about a product the comments posted in social medias can be analyzed to determine the product quality with respect to some topic or overall characteristics of a product. The main reason of taking the opinion from social media is the huge increase in the usage of Internet social medias and exchanging the opinion on the same driven the force behind Opinion analysis [2]. The web is a enormous repository of information in the form of structured and unstructured.

The document based sentiment analysis can also be done on these documents. For this purpose the document is summarized as positive labeled document, negative labeled document or neutral. The document is summarized based on the terms in the document bearing sentiments and then classified. The sentences in the document are classified based on the polarity. Among all the sentences the important thing is finding the entity towards which the sentiment is directed based on the polarity and the degree of the sentiment. The positive label sentiments may denote words like good, happy, nice, and awesome and so on and the negative sentence may contain the works such as not good, bad, dissipated, immoral and so on.

Feedbacks, suggestions, ideas and recommendations taken from the social media plays vital role in customer buying decisions and improving any kind of business. Especially in commerce , Word of Mouth is a way of conveying the attitudes of the customer , their opinions on their product and reactions about the products and services with others. As most of the information are communicated in social network online reviews on products, movies, books and authors so on has increased recently. For this reasons Sentiment Analysis gives a big hand. Exponentially growing availability of these opinions rich contents are available in the resources like online review sites , twitter, blogs , facebook and so on made this "opinion analysis" an easy and efficient Task. Liu[3] stated that , information available in textual format can be classified into two main things: Facts and Opinions. The method that classifies a given sentence into either opinionated sentence or non-opinionated sentence is known as subjectivity classification. The consequential sentences which are prejudiced are also devised into positive sentence or negative sentence, which is further known as sentence-level sentiment classification [4][5].

II. RELATED WORKS

Ashish Katrekar[6], explored the challenges and needs of Sentiment Analysis and proposed the importance of having better understanding of natural language sentiments and the importance of bridging the gap between the information in the form of unstructured by preprocessing and transforming into structured format for the efficient result of Sentiment Analysis.

Machine learning systems cannot differentiate the sarcasm from the implicit sentiment bearing sentence as well as these system cannot analyze the context sensitive meaning of that word. The acronyms present in the sentence such as FYKI,LOL are also pose major challenges in interpretation. Mixed opinion such as "I like the purse but the size is very big", is difficult to classify as either positive or negative.

A.B. Pawar et al.[7], presented a the importance of recognizing and identifying both explicit and implicit opinions with their respective features and classifying the opinions as positive, negative or neutral. They have also concluded that these sensible attention on implicit opinions as well as explicit will make helps the application to overcome its challenges and stay strong and active.

Bo Pang and Lillian Lee(2008)[8] presented a detailed survey on Sentiment Analysis, the demand and need for Sentiment Analysis , Applications, General Challenges, Classification, implications and publicly available resources in their paper. The paper provides the resources that can be used for effective Sentiment Analysis.

Kim Schouten, Flavius Frasincar[9], used implicit feature extraction in their work and suggested this method has given better result when compared to the co-occurrence based matrix. They have also proposed to test the power of word categories as Parts-of-Speech.

III. SENTIMENT ANALYSIS – APPLICATIONS

As sentiment analysis plays a dynamic role in consumer market which can be used for getting reviews about the products, improve marketing based on consumer's attitudes and interest, to find general estimation about recent hot topics happening in town like movies, songs, new model cars etc. Hence, based on these benefits the applications of Sentiment analysis has the following categories[8].

i) Review-Related Websites

To provide wide-ranging review about movies, products and so on.

ii) Unsolicited mail detection

To detect incompatible , intense language in mails , spams and context sensitive facts detection

iii) Business

To know about customers attitudes, opinions , ideas and views about a product and enhance business based on the sentiments.

iv) Other Domains

Sentiment Analysis can also used in many domains for example to get general opinion about political leaders , the strength of a particular political party and so on, Sentiment Analysis can be applied to get effective results.

IV. CHALLENGES IN SENTIMENT ANALYSIS

Extracting sentiment information such as positive , negative and neutral cannot be possible from the sentences bearing implicit sentiments. Owing these kind of sentence analysis challenges, these challenges can be summarized as follows:

Implicit and Irony Sentiment

A sentence may bear implicit sentiment which represent either positive or negative sentiment.

For example the sentence "How can anyone sit through this movie?" gives negative comment implicitly and the sentence does not bear any explicit sentiment words. Here semantic plays major role than syntax.

Hence, to make the Contextual Sentiment Analysis efficient Latent Semantic Analysis is used with Singular Value Decomposition(SVD) technique. The following figures Fig.1.a and Fig.1.b. illustrates the existing and proposed techniques.

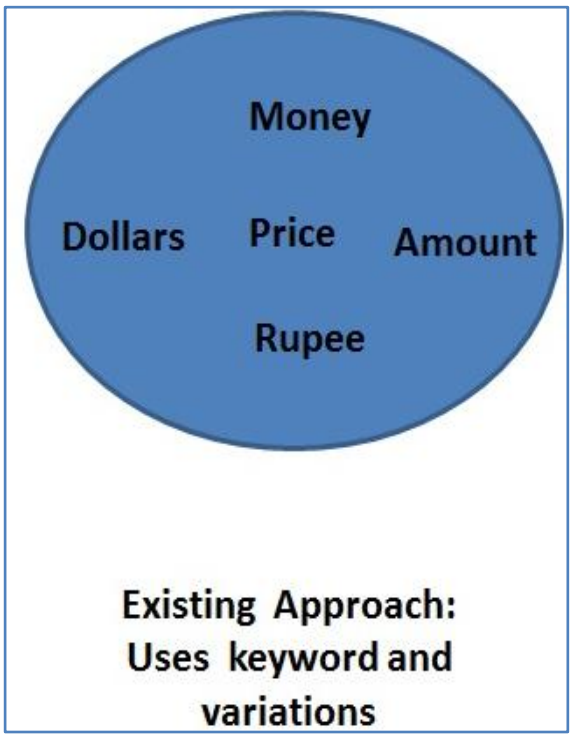

Fig.1. Existing Approach

The existing system takes direct meaning related to a word. For example to the word 'price' the most related words 'Dollars','Amount','Money','Rupee' are considered.

In the proposed Contextual Semantic Search the context based semantic meanings are taken to extend the analysis to enhance the efficiency of Sentiment Analysis.

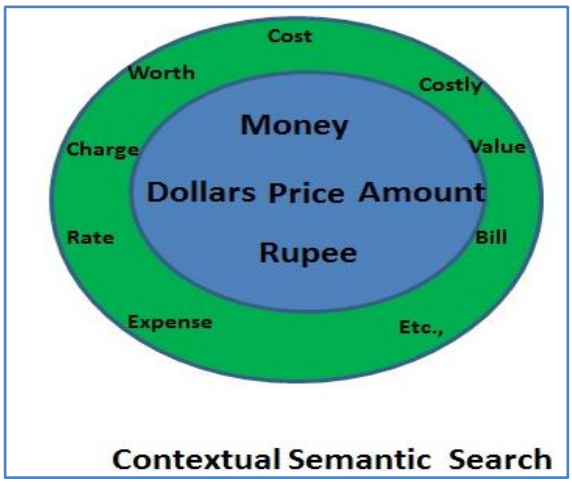

Fig.2. Proposed System(LSA with SVD)

V. PROPOSED SYSTEM

To enhance the efficiency of Sentiment Analysis Latent Semantic Analysis(LSA) technique is used. LSA is a technique which is widely used in Natural Language Processing(NLP) especially in distributional semantics to analyze the relationship between documents and terms using the concepts related to the documents. It assumes the contextual words occur in similar parts of text. A matrix is constructed in terms of paragraph and words as rows and columns respectively using Singular Value Decomposition(SVD) technique. SVD reduces the number of rows by preserving the similarity structure in the columns based on semantically related entities. Susan T. Dumais (2005)[10].

The complete SVD decomposition is an procedure that ingests much memory and time, mostly for big problems. Therefore, the endorsements prediction process can be driven very efficiently if the time consuming Singular Value Decomposition of LSI is exchanged by the partial symmetric Eigenproblem that can be resolved by making use of quick iterative solvers [11][12].

This improvement concentrates on how the matrices S and V are expressed without the complete SVD decomposition. Supposing the popular association between the SVD of the matrix A and the symmetric Eigenproblem of the symmetric square matrices $A^T A$:

$$A = USV^T \qquad (1)$$

$$A^T(USV^T)^T = VS^TU^T \qquad (2)$$

$$A^TA = VS^T(U^TU)SV^T = VS^TSV^T \qquad (3)$$

Where A indicates the $m \times n$ ratings matrix, $S \in R^{m \times n}$ refers to a diagonal matrix having nonnegative diagonal elements known as the singular values, $U \in R^{m \times n}$ and $V \in R^{m \times n}$ stand for orthogonal matrices. The columns of matrices U and V are known as the left singular vectors and the right singular vectors, correspondingly. The decomposition can be calculated such that the singular values are ordered in descending sequence.

Since the fact that the matrix V is orthogonal, the below matrix identity is true:

$$AV = US \qquad (4)$$

At last, the matrix U can be expressed in the following manner:

$$AVS^+ \approx U$$

Where, S^+ represents the Moore--Penrose pseudo inverse (pinv). Evaluating the LSI procedure keenly, it can be seen that the matrix U does not need to be explicitly calculated and then stored in memory during the LSI. The usage of this observation brings the extra acceleration to the speed and reduction of the memory needs of the LSI.

VI. EXPERIMENTAL RESULTS

The reviews about different products were taken as samples. The proposed system was implemented using Python. The experimental results were compared with existing Sentimental Analysis using Keyword variation . The following table Table1 shows the results obtained by both the methods and it is clearly seen that the proposed method obtained better results.

Table1- Comparison of Proposed system LSI results (Precision ,Recall values) with existing system.

Product Name	***Keyword Variation Sentiment Analysis***		***Contextual based Sentiment Analysis***	
	Recall	***Precision***	***Recall***	***Precision***
Movie1	***0.720***	***0.642***	***0.920***	***0.824***
Car	***0.632***	***0.553***	***0.821***	***0.732***
Laptop	***0.674***	***0.814***	***0.853***	***0.924***
Digital Camera	***0.784***	***0.589***	***0.952***	***0.795***
DVD Player	***0.653***	***0.607***	***0.863***	***0.872***

VII. CONCLUSION

Sentiment Analysis plays a vital role in recent days. It is a tedious task and the complexity increases when the opinions are expressed as implicit. Hence, analyzing implicit opinions bearing sentences are being a huge challenge in this Sentiment Analysis. The existing methods which used keyword variation based Sentiment Analysis lack in its efficiency when applied on these implicit bearing sentences. The new approach Latent Sentiment Analysis overcomes the problem efficiently when compared to this existing method.

REFERENCES

1. Pang, Bo and Lee, Lillian and Vaithyanathan, Shivakumar, Thumbs up?: sentimentclassification using machine learning techniques, In Proceedings of the ACL-02 conference on Empirical Methods in Natural Language, 2002.
2. Subhabrata Mukherjee, "Sentiment Analysis A literature survey", Indian Institute of Technology, Bombay Department of Computer Science and Engineering.
3. Liu,B, "Sentiment Analysis and subjectivity", In:Handbook of Natural Language Processing,2nd edn. Pp.1-38(2012).
4. Dalal.M.K, Zave.M.A,"Automatic Text Classification : a technical review", Int. J.Comput. Appl.(0975-8887) 28(2),37-40(2011).
5. Tang.,H, Tan.S.,,Cheng X,"A survey on Sentiment Detection of reviews",In: Science Direct, Expert Systems with Applications,pp.10760-10773(2009).
6. Ashish Katrekar, "An Introduction to Sentiment Analysis",AVP, Big Data Analytics, GlobalLogic Inc., www.globallogic.com.
7. A.B. Pawar, M.A. Jawale and D.N. Kyatanavar," Fundamentals of Sentiment Analysis:Concepts and Methodology" , © Springer International Publishing Switzerland 2016
8. W. Pedrycz and S.-M. Chen (eds.), Sentiment Analysis and Ontology Engineering,Studies in Computational Intelligence 639, DOI 10.1007/978-3-319-30319-2_2.
9. Bo Pang and Lillian Lee," Opinion mining and sentiment analysis"**,** Foundations and Trends in Information RetrievalVol. 2, No 1-2 (2008) 1–135c2008 .
10. Kim Schouten, Flavius Frasincar, " Implicit Feature Detection for Sentiment Analysis", Erasmus University Rotterdam,PO Box 1738, NL-3000 DR Rotterdam, the Netherlands.
11. Susan T. Dumais (2005). "Latent Semantic Analysis". Annual Review of Information Science and Technology. **38**: 188–230. doi:10.1002/aris.1440380105
12. Moreira.A, M. Y. Santos and S. Carneiro, Density-based clustering algorithms-DBSCAN and SNN, July 2005