ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Logistic Regression: Aggregating Reviews by User Preference Modeling

Vanashri N. Sawant, Prof. Supriya Sarkar

Department of Computer Engineering, Savitribai Phule Pune University, SKN Sinhgad Institute of Technology & Sciences, Lonavala, Pune, Maharashtra, India

Department of Computer Engineering, Savitribai Phule Pune University, SKN Sinhgad Institute of Technology & Sciences, Lonavala, Pune, Maharashtra, India

ABSTRACT: Although personalized search has been discovered for many years and many personalization strategies have been invented, it is still undecided whether personalization is consistently effective on different queries for different users, and under dissimilar search contexts. According to these user comments, rating can be done about any product launched in market, but it quite difficult. To assuage this problem, we propose a logistic regression algorithm which will gather round all the comments fetched by system witness and able to compute positive-negative probability, with the help of this probability we can rate movie, industrial company and education societies. This proposed system designed on the basis of comments left by user earlier. An actual movie, industrial company, education societies review data set achieved from a huge data portal, our system help to improve significance comparison performance. The proposed algorithm can accurately predict a user's preferences in their interest area. Using online user survey we can get true user interest preference about multimedia as well as web mining content.

KEYWORDS: Logistic Regression. Data mining, personalized search contents, user comments, Logistic regression, recommender system, Sentiment analysis

I. INTRODUCTION

The word sentiment originates from Latin sentire which means feel. In social and economic sense sentiment means a thought, view, or attitude, especially one based mainly on emotion instead of reason. Sentiment Analysis is natural language processing (NLP) and computational techniques to automate the extraction or classification of sentiment from typically unstructured text.World Wide Web has grown up in a huge manner which has contained a millions of documents with various type of information. This Information can be extracted from the web 2.0 as per user requirements. Whatever information collected or given to user is known Information Retrieval. It is nothing but giving more relevant data to respective user. It comes under Web mining which is an advanced approach to data mining for discovering various patterns from web. Web Mining can be classified into three parts: 1) Web usage mining 2) Web content mining 3) Web structure mining.

Sentiment analysis is nothing but the opinion mining. A lot of new contents invented on web or in multimedia data if we analyze the comments given by user after using that data, so it will help us for understand system user demands. Sentiment analysis is a study focused on the sentiments, evaluations, emotions towards any product, organization, services, movie data portal etc. Mostly Natural language processing (NLP) and text mining will work on sentiment analysis. Human related all activities extracted from their opinions because they are the central key point of influencers behaviors. At the time of fixing our self decision we have to consider opinions of others. Consumers or public opinions about any product or services have been always caught by respective organization or business as a guideline. For analysis of customer's demands a survey of distribution of product, special group can be conducted by organization or business. Individual user also wants to know opinions of past users so that he will easily decide either buy or reject that product. Also in political area opinion mining worked well, before voting any election candidate voter will ask other's opinions. In past before reaching any final decision individual person hears other review/thoughts.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Smart TVs are the example of connected TV, which expands the function of television sets by integrating internet and web 2.0 features [1]. TV user can have access to many traditional as well as internet services. User can watch any program on TV as per their wish and they will give reviews on those program. User can watch many choices on TV. According to user access we can analyze content search or recommendations that consider individual preferences.

Personalized web search concentrate on the right content delivery to right person in extracting information from web [1, 2]. When user searches data on web or in data warehouse through their clicked pages approach [3, 4, 5] we can made some conclusion about user profile. Numbers of techniques which can be adopted from Information retrieval include personalized web search. Accuracy is major challenge in personalized search [6, 7] for that precision and recall used. This leads me to think in such direction how we can detect the exact degree of that Opinions according to comment left by users. For this purpose, by referring state of art and finding out how much work has been done in this area, finally come to my Dissertation Topic.

II. MOTIVATION OF THE DISSERSTATION

There is no limit to the range of information conveyed by comments, often these short messages are used to share opinions and sentiments that people have about what is going on in the world around them. Working with these informal text genres presents challenges for natural language processing beyond those typically encountered when working with more traditional text genres, such as newswire data. Comments are short: a sentence or a headline rather than a document. In general, the term, “personalization”, means providing right contents to right users in accordance with their preferences. There are also two additional ways of utilizing user profiles for personalization: 1) query expansion by adding new terms to query fetched by user according to their preference as we call it re-weighting [8] and 2) re-ranking and filtering of the search results by means of consumer profiles [9].Recommendation is one of the active research areas where those of personalization techniques are used in content providing services. It aims at recommending items that users had not yet considered, but are likely to be preferred. As one of the most successful approaches to building recommender systems, collaborative filtering (CF) uses the known preferences of a group of users to make recommendations or predictions of the unknown preferences for other users[10,11,12]. Although content based filtering is simpler and easy to analyze for recommendation, collaborative filtering generally shows better performance than content-based filtering.

Outside of personalization, there are several studies using user comments as a main source for their experiments. Some studies show that the number of user comments posted on news [13] and blog posts [14] is an indicator of popularity. Recent studies also introduced several methods to identify useful comments [15, 16]. Utilizing user comments more directly for search, Yee et al. [17] have examined the potential impact of user comments on search accuracy in social Web sites [17].

Now days for reducing time complexity, space complexity, data redundancy & accessing complex queries we will apply logistic regression technique to overcome these drawbacks in previous work.

III. RELATED WORK

Our day-to-day life has always been predisposed by what people think. Our own opinions are exaggerated by others Ideas and opinions about same thing. The explosion of Web 2.0 has led to increased activity in Podcasting, Blogging, and Tagging, Contributing to RSS, Social Bookmarking, and Social Networking. As a result there has been a flare-up of interest in people to mine these commenting data for opinions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. Personalizing web search has numerous preceding attempts. One approach is to inquire users to stipulate general interests. The user interests help us to sort out search results according to checking content similarity between clicked web pages and user interests [1, 18]. User interest categories or term lists/vectors provide an idea about user profiles [4]. This document provides the purpose of the feasibility study, the background of the proposed project, the methodology used for performing the study, and any reference materials used in conducting the feasibility study for the dissertation titled “An optimized approach for Personalized Content Search using Logistic Regression”. To verify the feasibility of system, two methodologies are used: Surveying and Brain Storming. Literature survey by studying various IEEE papers and other related reference material is conducted. To determine projects capability, the feasibility study has been conducted. The results of this study will be used to make a decision whether or not to continue with the project.

There are two research area i.e. personalization and information retrieval (IR) can provide many clues about any user depending upon their opinions. Only few only a few studies focus on utilizing user comments for a personalization search system. In general, it means making conformation that correct contents supplying to correct user in considering to their preferences [2].There are 3 ways for guessing user profiles in personalized information retrieval:-1) Utilizing dynamic input by user [1], 2) Queries left by user in past [3] and clicked through data [4] [19], and 3) User's profile on social network [6]. In content providing services personalization techniques are used on the basis of recommendation system which is one of the active research areas.

A. Survey on Web personalization

Web personalization is the process of customizing a Web site to the needs of specific users, taking advantage of the knowledge acquired from the analysis of the user's navigational behavior (usage data) in correlation with other information collected in the Web context, namely structure, content and user profile data. Due to the explosive growth of the Web, the domain of Web personalization has gained great momentum both in the research and the commercial area. In this paper we present a survey of the use of Web mining for Web personalization. More specifically, the paper introduces the modules that comprise a Web personalization system, emphasizing on the Web usage mining module. A review of the most common methods that are used as well as technical issues that occur is given, along with a brief overview of the most popular tools and applications available from software vendors.

B. Survey on Recommender system

Recommendation as a social process plays an important role in many WWW applications. Recommender systems or recommendation systems are a subclass of information filtering system that seek to predict 'rating' or 'preference' that a user would given through their opinions. The paper presents an overview of the field of recommender systems along with the description of various approaches that are being used for generating recommendations. Recommendation techniques can be classified in to three major categories: Collaborative Filtering, Content Based and Hybrid Recommendations. The paper elaborates these approaches and discusses their limitations by describing the major problems suffered by recommendation methods. This paper focuses on the Cold Start, Collaborative Filtering, Content-Based Recommendation, Recommendation System, Sparsity Problem.

C. Personalized Web Search System, Recommender System(Yoda)

C. Shahabi and Y.C. Chen.[2] proposed personalized search system and Yoda technique. During browsing and searching WWW, the information overload becomes major challenge due to dramatically increased in web pages. Personalization becomes a popular remedy to customize the Web environment towards a user's preference. Personalized search system works on query refinement & Personalized meta system.

D. Query Expansion and History Language Model

B. Tan, X. Shen, and C. Zhai [3] introduces history language model. Long-term search history contains rich information about a user's search preferences, which can be used as search context to improve retrieval performance. The paper introduces a statistical language modeling based methods to mine contextual information from long term search history and exploit it for a more accurate estimate of the query language model. But the paper has some drawbacks that it had used simpler model, can't provide any algorithm for contextual search on the client-side, also not beneficial for unstructured data.

E. Click based Personalization Strategies

Z. Dou, R. Song, and J.-R. Wen. [4] establish the concept of a large-scale personalized search evaluation framework based on query logs. In this framework, different personalized re-ranking strategies are simulated and the search accuracy is approximately evaluated by real user clicks recorded in query logs automatically. The framework enables us to evaluate personalization on a large scale. Two click-based personalized search strategies and three profile based personalized search strategies. Personalization brings significant search accuracy improvements on the queries with large click entropy, and has little effect on the queries with small click entropy. Paper works on both long-term and short-term contexts are very important in improving search performance for profile-based personalization. We use 12

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

days of MSN query logs to evaluate five personalized search strategies. The profile-based personalized search strategies proposed in this paper are not as stable as the click-based ones. They could improve the search accuracy on some queries, but they also harm many queries. The system fails in terms of repeated queries, can't give more specific result.

F. SaND(Social Network & Discovery) Tool

Z. Dou, R. Song, and J.-R. Wen.[6] investigates personalized social search based on the user's social relations – search results are re-ranked according to their relations with individuals in the user's social network. The work contains several social network types for personalization: Familiarity-based network, Similarity-based network, Overall network that provides both relationship types. Bookmarked based evaluation approached for off-line study. Investigation results showed that according to both evaluations, social network based personalization significantly outperforms non-personalized social search. In this work we simulated personal queries with tags used for bookmarking by the user, in the off-line study, and with tags the user was tagged with, in the user survey. In bothcases these types of personal queries are limited and do not cover the whole spectrum of possible personal queries. System fails for long run queries.

G. Collaborative Filtering

Kwon and K. Hong. [7] works on smart TV vastly which expands the function of television sets by integrating the Internet and Web 2.0 features into contemporary television sets and set-top boxes. TV users can now access a wide range of contents not only from traditional broadcasting services but also from the Internet through a single device. While the availability of numerous contents on a TV means more choices, it also poses a great challenge to its users as they have to decide what to watch out of an almost infinite number of competing choices, highlighting the importance of content searches or recommendations that consider each user's individual preferences. In the context of a recommender system, various studies have been conducted in an effort to recommend proper contents to connected TV users in accordance with their individual preferences. The context tagging-based user's preference prediction mechanism was used by extending the widely known recommender algorithm, collaborative filtering (CF) in order to increase the user's satisfaction about the recommender service. This method can't evaluate relevant results for large data set in consideration of scalability so the given technique needs an improvement in its scalability algorithm.

H. Query Log & Clicked through analysis

For web search personalization many researchers consumes query log and clicked through analysis. In [20], the authors combine a topic-sensitive version of Page Ranking [21] with the previous user clicked page history for focus on user personalization area. Joachims et al. [22] draw conclusion from clicks applicability, it supports to thought that users' clicks provide a reasonably accurate evidence of their preferences.

IV. PROPOSED SYSTEM

A. ***System Architecture:***

Our propose system consists of following sequence that shown in figure below: The application has been divided into four major categories-

1. Text Mining Process
2. Sentiment Analysis
3. Classification
4. Result Analysis

a) Text Mining Process

In practical term, the classification task requires a pre-classified database sample, called training set, which is either used to generate a classifier or to compare with new unlabeled data to be classified. This is important because the classifier accuracy is highly dependent upon such training set [1]. Before explaining each of these problems in detail, let's review a general architecture of a generic text mining analysis system.

- ***Comment pre-processing***

This is the most computational expensive phase, because it requires the processing of unstructured data [1]. This step can be divided in to sub steps.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

1. Tokenization: Is used to identify all noun words in a given text. These words called tokens or terms, are basic units if the documents.

2. Stop word removal: Is used to eliminate that word s that occurs frequently such as article, prepositions, conjunction and adverbs. These stopwaords depends on language of the text in questions [1].

3. Stemming: It makes linguistic standardization of token in the text, in which variants form of this token reduces in to common form called stem.

4. Document representation: After all the previous steps have been implemented, the set of document was initially unstructured becomes closer to being structured. The most common model of representation is bag- of word, which uses word document as features.

5. Feature selection: Final step of text mining process Its aim at finding a reduced set of attributes that provides a suitable representation of this database given a certain analysis to be performed.

- ***Comment Analysis***

The analysis step is usually considered the core of text mining, because this is when some type of useful, nontrivial knowledge is extracted from the text. The analysis can be classified in to two categories.

A. Descriptive analysis: Characterizes the general properties of the data by means of a characterization. It promotes data summarization. And discrimination provides descriptive comparisons among the database.

B. Predictive analysis: Makes inferences about the database, in order to make prediction [1].

- ***Validation***

In order to validate, the analysis is performed. It is necessary to user quantitative measures and qualitative measures. After such a validation it may be necessary to return to one or more of the previous step so as to perform modification and try alternatives. The input to the system is comments given by reviewer in any textual format with expressing their emotions.

b) Sentiment Analysis

It consists only one type of sentiment Analysis in this project.1.Sentence level sentiment Analysis. Sentence level sentiment analysis. This is the simplest form of sentiment analysis and it is assumed that the comment contains an opinion on one main object expressed by the adminof a reviewer system. Numerous papers have been written on this topic. The recommemdation system assumes that there is a finite set of classes into which the comment should be classified and personalized search done as per user opinions[2 web]. The simplest case is when there are two classes: positive and negative. More advanced representations utilize TFIDF, POS (Part of Speech) information

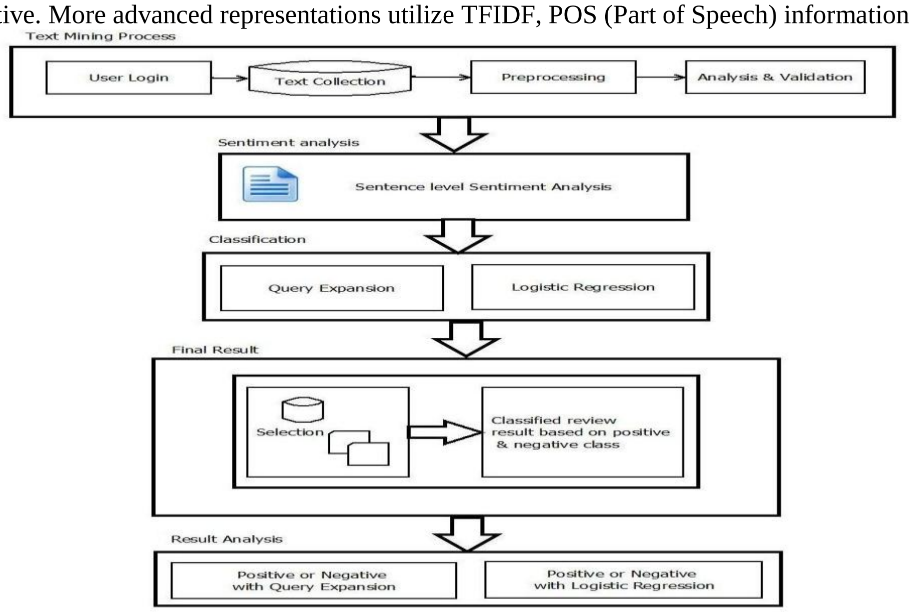

Fig. 1: System Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

c) Classification:

Classification is done using Logistic Regression algorithm. In these few types of Logistic regression approaches assumes 1.General model of Logistic Regression, 2. Linear model Logistic Regression and 3. Binary model Logistic Regression. For our system we assume Binary model Logistic Regression approach. The Logistic Regression approach assumes that there is a finite set of classes into which the commented sentence should be classified and training data is available for each class. The simplest case is when there are two classes: positive and negative. Simple extensions can also add some discrete numeric values into which the comment should be placed. So that proposed system can easialy focus on user interest area.

This proposed system consists of three modules using this modules we can classify the system.

1. Support Counting Module : This module is responsible for checking the percentage of tweets that contain at least one emoticon from the set word. That means decide the minimum thresholds for experiment.

2. Database Selection Module : The data set is consider for opinion mining is training data set. The training set contains comments given by reviewer offline or online, that will be classified submitted comments into two classes as positive and negative automatically.

3. Classification Module: This module is responsible for classifying the comments whose labels are unknown. Among the many two classification algorithms as Query expansion available in the literature and other is Logistic regression, we choose to use the Logistic regression for two reasons: First it is a time consuming while calculating result, in the sense no training is required, only the storage of pre-classified samples; second, because it has broadly used to classify opinions into positive and negative class.

d) Final result

We classify our sentiments in 2 categories:

1. Positive Sentiments.
2. Negative Sentiments.

In result we check frequency of each emotion of sentiment words. After that we will calculate positive and negative score for each sentiment by using both methods as Query Expansion and Logistic Regression. On basis of that result comparision we can see accuracy of Logistic Regression is more than Query Expansion

B. ***Description of the Proposed Algorithm***

Input: User Comments

Output: Movie Score.

1. Start
2. User submits comments on a movie Mi
3. Find Pi and Ni using logistic regression.
4. Calculate rank vector Rifrom Pi and Ni
5. Calculate Score Si for movie Mi
6. Stop.

C. ***Description of the Mathematical Model***

a) *Set Theory:*

1. Let S={} be as a Personalized Movie search system.
2. Verify user$U$
   $U=\{uid, pass\}$
   Where uid=unique id for specific user
   pass= password phrase
   $S=\{U\}$
3. Verify adminA
   $A=\{uid, pass\}$
   Where uid =unique id for specific user
   Pass= password phrase
   $S=\{U, A\}$
4. Approve user $u_n$by admin
   $U_{DB}=\{u1, u2, \ldots\ldots\ldots , un\}$

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Where *u1* is a user of system
U_{DB}= *user database*
$u1 \in U_{DB}$
$S = \{U, A, U_{DB}\}$

5. Obtain $MD_{DB}$is movie description database
$MD_{DB}= \{m1, m2, \ldots\ldots, mn\}$
Where *m1*is movie description of individual movie
$S=\{U, A, U_{DB}, MD_{DB}\}$
6. Collect user comments U_{CMT}
$U_{CMT} = \{ucmt1, ucmt2, \ldots\ldots, ucmtn\}$
Where *ucmt1*comments about movie from user
$S=\{U, A, U_{DB}, MD_{DB}, U_{CMT}\}$
7. Calculate $M_{RNK}$from U_{CMT}
$M_{RNK}=\{mrnk\}$
Where *mrnk*is a final rank of movie
$S=\{U, A, U_{DB}, MD_{DB}, U_{CMT}, M_{RNK}\}$
8. Final Set$S = \{U, A, U_{DB}, MD_{DB}, U_{CMT}, M_{RNK}\}$

b) *Mathematical model for proposed system:*

1. Identify user commentsU_{CMT}={}
2. Calculate rank vector Ri

$$Ri=\frac{Pi}{Pi+Ni}=\frac{Pi}{Ti}$$

Where Pi= positive comments
Ni=negative comments
Ti=total comments (Pi+Ni)

3. Movie rank obtained by equation

$$Si =\frac{1}{Ti}\sum_i AiRi$$

Where Si is a score of specific Movie,
Ni<Ai<Pi = user comments which isn't belongs from both sets.

V. EXPERIMENTAL SETUP AND RESULT

System user interface was used to capture comments and a search script was written in source format to make queries to get total 24 hours of comments captured, input of the system is the sample dataset collected from System user interface. Using Logistic Regression algorithm, classification is done for positive and negative level output. A result is collected as train data set and apply accuracy parameter to evaluate accurate result viz. Precision call (Pr), Recall (Re), F-measure.

- ***Sentiment score****:*It can be calculated using Logistic Regression and Query Expansion methods for movie review of a single movie.

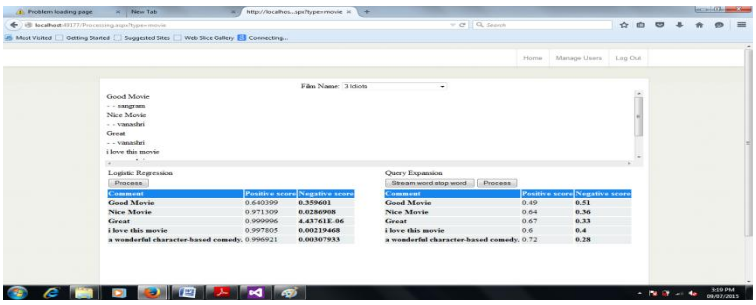

Fig..2. Sentiment score of Movie Review

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

- **Accuracy Graph Analysis:**
  Accuracy of sentiment score obtained by Logistic Regression is better than Query Expansion for each field.
  - Movie review data:

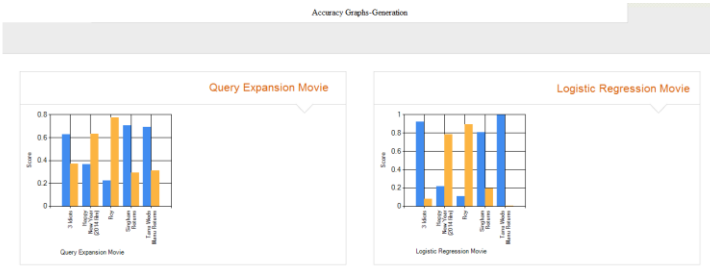

Fig. 3. Accuracy graph analysis for Movie review data

- **Time Complexity Graph Analysis:**

Time complexity of sentiment score obtained by Logistic Regression is better than Query Expansion for each field.
  - Movie review data:

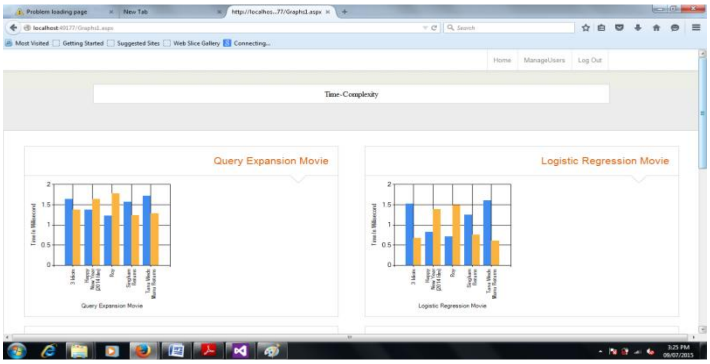

Fig. 4. Time complexity graph analysis for Movie review data

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network.As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

VI. CONCLUSION AND FUTURE SCOPE

Sentiment analysis is information extraction task. The proposed system introduces the concept of deducing the polarities of words based on the polarities of other words. Experimental results show that the number of new words with polarities deduced is approximately original sentimental word dictionary. Providing satisfactory solutions to these challenges will make the area of sentiment analysis far more widespread. In the proposed system presented a method for an automatic collection of a corpus user comments that can be used to train a sentiment classifier. Logistic Regression calculates the positive and negative sentiment score about given comments. It improves the efficiency and accuracy of the system. In the further improvement, we can expanding existing techniques to handle more general writings and crossing domains is an exciting opportunity for both academia and businesses. Also try to collect a multilingual corpus of Comment data and compare the characteristics of the corpus across different languages.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

REFERENCES

1. Hyung W. Kim, Keejun Han, Mun Y. Yi, Joonmyun Cho, and Jinwoo Hong, "Movie Mine: Personalized Movie Contents Search by Utilizing User Comments, IEEE Trans. On Consumer Electronics, Vol. 58, No.4, November 2012.
2. C. Shahabi and Y.C. Chen. "Web information personalization: Challenges and approaches," In Proceedings of 3rd Workshop on Databases in Networked Information Systems (DNIS), 2003.
3. B. Tan, X. Shen, and C. Zhai, "Mining long-term search history to improve search accuracy," In Proc. of ACM KDD, pp 718–723, 2006.
4. Z. Dou, R. Song, and J.-R. Wen. "A large-scale evaluation and analysis of personalized search strategies," In Proceedings of WWW, pp 581–590, ACM, 2007.
5. X. Shen, B. Tan, and C. Zhai., "Context-sensitive information retrieval using implicit feedback", In Proceedings of SIGIR 2005, pages 43–50, 2005.
6. D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'el, I. Ronen, E. Uziel, S. Yogev and S. Chernov, "Personalized social search based on the user's social network," In Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM'09),2009.
7. H. Kwon and K. Hong, "Personalized smart TV program recommender based on collaborative filtering and a novel similarity method", IEEE Trans. Consumer Electron., vol. 57, no. 3, pp. 1416-1423, August 2011.
8. P. A. Chirita, C. S. Firan, and W. Nejdl, "Personalized query expansion for the web," In Proceedings of ACM SIGIR, pp. 7–14, 2007.
9. X. Shen, B. Tan, and C. Zhai, "Implicit user modeling for personalized search," In Proceedings of ACM CIKM, pp 824–831, 2005.
10. J.A. Konstan,, et al., "GroupLens: Applying collaborative filtering to usenet news," Communications of the ACM vol. 40, no. 3, 1997, pp 77-87.
11. J.A. Konstan,, et al., "GroupLens: Applying collaborative filtering to usenet news," Communications of the ACM vol. 40, no. 3, 1997, pp 77-87.
12. G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item-to-item collaborative filtering," IEEE Internet Computing, vol. 7, no. 1, pp. 76-69, Jan./Feb.2003.
13. B.N. Miller, et al., "MovieLens unplugged: experiences with an occasionally connected recommender systems," In Proceedings of. IUI, 2003.
14. M. Tsagkias, et al., "News comments: Exploring, modeling, and online prediction," In Proceedings of ECIR, 2010.
15. T. Yano, and N.A. Smith,, "What's worthy of comment? content and comment volume in political blogs," In Proceedings of AAAI, 2010.
16. J. Chen et al., "Short and tweet: Experiments on recommending content from information streams," In Proceedings of CHI, 2010.
17. S. Faridani, et al., "Opinion space: A scalable tool for browsing online Comments," In Proceedings of CHI, 2010.
18. Wai Gen Yee, Andrew Yates, Shizhu Liu, and Ophir Frieder, "Are web user comments userful for search?", In Proceedings of 7th Workshop on Large-Scale Distrubutued Systems for Information Retrieval (LSDS-IR),pp. 63-70, 2009.
19. P. A. Chirita, W. Nejdl, R. Paiu, and C. Kohlsch¨utter. Using odp metadata to personalize search. In Proceedings of SIGIR '05, pages 178–185, 2005.
20. T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay,―Accurately interpreting clickthrough data as implicit feedback,‖ In Proceedings of ACM SIGIR, pp 154–161, 2005.
21. Qiu and J. Cho. Automatic identification of user interest for personalized search. In Proceedings of SIGIR, pages 727–736. ACM Press, 2006
22. T. H. Haveliwala. Topic-sensitive pagerank. In Proceedings of WWW, pages 517–526. ACM, 2002.
23. T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In Proceedings of SIGIR, pages 154–161, 2005.