

Detaining the Subjective Relevant Events Based on User's Likelihood by an Exotic Approach

R. Sridevi¹ Mrs. V. Aruna²

¹PG Scholar ²Assistant Professor

¹Department of Computer Science and Engineering ²Department of Information Technology

^{1,2}Sri Vidya College of Engineering & Technology Viruthunagar, TN, India

Abstract— Data Mining is the process of examining the different types of data and summarizing it to the useful information. The Opinion mining is the process of natural language processing, analysis, computational identification and extraction of subjective relevant information. Social media is an emerging trend for communication between people worldwide and it has attracted millions of users. A typical characteristic of such sites is that they allow anyone to post or like anything they like on any subject. Here the user's feedback and satisfaction is playing a vital role. In some cases the user may not give the feedback or reviews which they have viewed or downloaded directly. Instead of that the user is just searching the available things based on their interest. So there is no possibility for capturing the domain interest behavior of the user explicitly. This paper proposes to detain the subjective relevant events based on user likelihood by using the exotic approach. The user likelihood is analyzed and the domain is extracted. The Experimental results determine the viability for detaining the subjective relevant events.

Key words: Data Mining, Social Media, Search Engine, Recommendation System, Events

I. INTRODUCTION

Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. The information can be used to increase revenue, cuts costs and is called knowledge discovery. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. It also is a buzzword, and is frequently also applied to any form of large-scale data or information processing. Information processing is the process of collection, extraction, warehousing, analysis, and statistics for many application of computer decision support system, including artificial intelligence, machine learning, and intelligence. Sentiment analysis refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state, or the intended emotional communication. Existing approaches to sentiment analysis can be grouped into four main categories: keyword spotting, lexical affinity, statistical methods, and concept-level techniques. Keyword spotting classifies text by affect categories based on the presence of unambiguous affect words based on emotion. Lexical affinity not only detects words, it also assigns arbitrary words a probable affinity to particular emotions.

Statistical methods leverage on elements from machine learning such as latent semantic analysis, support vector machines, bag of words and Semantic Orientation with Point wise Mutual Information. More sophisticated methods try to detect the holder of a sentiment and the target. To mine the opinion in context and get the feature which has been opinionated, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. A social network is a social structure made up of a set of social actors and a set of the dyadic ties between these actors. The social network perspective provides a set of methods for analyzing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics. Social networks and the analysis of them is an inherently interdisciplinary academic field which emerged from social psychology, sociology, statistics, and graph theory. Social network analysis is now one of the major paradigms in contemporary sociology, and is also employed in a number of other social and formal sciences. Together with other complex networks, it forms part of the nascent field of network science.

A social networking service is a platform to build social networks or social relations among people who share interests, activities, backgrounds or real-life connections. A social network service consists of a representation of each user, his or her social links, and a variety of additional services. Social network sites are web-based services that allow individuals to create a public profile, to create a list of users with whom to share connections, and view the connections within the system. Most social network services are web-based and provide means for users to interact over the Internet, such as e-mail and instant messaging. Social network sites are varied and they incorporate new information and communication tools such as mobile connectivity, photo/video/sharing and blogging. Online community services are sometimes considered as a social network service, though in a broader sense, social network service usually means an individual-centered service whereas online community services are group-centered. The domains may be sports, motivation, nature, positive quotes, cooking, healthcare, gardening, general knowledge, job recruitment, pencil arts, shopping, comedies, books, school memory, data mining, networks, cloud computing, grid computing, image processing, communication networks, mobile computing, and mechanical engineering.

The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them. Suggestions for books on Amazon, or movies on Netflix, are real world examples of the operation of industry-strength recommender

systems. The design of such recommendation engines depends on the domain and the particular characteristics of the data available. For example, movie watchers on Netflix frequently provide ratings on a scale of 1 (disliked) to 5 (liked). Such a data source records the quality of interactions between users and items. Additionally, the system may have access to user-specific and item-specific profile attributes such as demographics and product descriptions respectively. Recommender systems differ in the way they analyze these data sources to develop notions of affinity between users and items which can be used to identify well-matched pairs. Collaborative Filtering systems analyze historical interactions alone, while Content-based Filtering systems are based on profile attributes; and Hybrid techniques attempt to combine both of these designs.

II. PROPOSED SYSTEM

The user will give a mail ID and password in a login page of social media or search engine. It Checks whether the mail ID and password is correct then the page will move to the user's home page of the social media or search engine. In social media, user gets like, share and view the information on any subject. In search engine, user gets view, like, download the information on any subject. Now we consider only likes in social media or download in search engine. After that we calculate how many lies are there in different domain of social media or how many downloads are there in different domain of search engine. Find the maximum like in social media or downloads in search engine. We have to identify the maximum likes in social media or the maximum downloads in search engine based on the subject of user domain. Now recommendation system is used to recommend the related events based on the interest of the user. Then the domain related events will send to the user's mail ID in search engine.

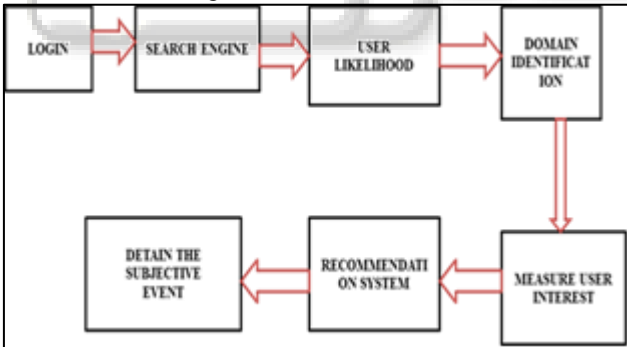


Fig. 1: Architecture Diagram

III. METHODOLOGY

A. Opinion Feature Identification:

Domain relevance et al [2], characterizes how much a term is related to a particular like (i.e., a domain) based on two kinds of statistics, namely, dispersion and deviation.

Dispersion quantifies how significantly a term is mentioned across all documents by measuring the distributional significance of the term across different documents in the entire like (horizontal significance).

Deviation reflects how frequently a term is mentioned in a particular document by measuring its distributional significance in the document (vertical significance).

Each like L_a has a like frequency (LF – IDF).

- LF-Like Frequency
- Inverse Domain Frequency

Each like L_a has a like frequency LF_{ab} in a domain D_b and global domain frequency DF_i

The weight W_{ab} of like L_a in domain in domain D_b the calculated as follows:

$$W_{ab} = \begin{cases} (1 + \log LF_{ab}) * \log(\frac{N}{DF_i}) & \text{if } LF_{ab} > 0 \\ 0 & \text{otherwise} \end{cases}$$

$a=1, \dots, M$ for a total number of M likes.

$b=1, \dots, N$ for a total number of N domain in the corpus.

The standard variance SV_a for like L_a is calculated as follows.

$$SV_a = \sqrt{(\sum_{b=1}^N (W_{ab} - wa')^2) / N}$$

The average weight W_a of like L_a across all domain is calculated by

$$W_a = 1/N (\sum_{b=1}^N w_{ab})$$

The dispers on disp of each like L in the corpus is defined as follows

$$Disp_a = W_a' / SV_a$$

The derivation devi of like L in domain D is given by

$$Devi_{ab} = W_{ab} - W_b'$$

Wr the average weight in the domain calculated over all M terms as follows.

$$W_b' = 1/M (\sum_{a=1}^M W_{ab})$$

The domain relevance dr for like L in the corpus is finally defined as follows.

$$Dr_a = disp_a * \sum_{b=1}^N devi_{ab}$$

Clearly, the domain relevance dv_a incorporates horizontal and vertical distributional significance of term LF_i in the likes.

B. Exotic Algorithm:

Input: A domain specific/independent corpus C

Output: Domain relevance scores (IDR or EDR)

For (candidate feature LF_i)

for (document D_1 in D)

find w_{ab} then

cal SV_a

cal $disp_a$

for (document D_j in C)

cal ($devi_{ab}$)

Return(list)

C. Opinion Features Algorithm:

Input: Domain review corpus R and domain-independent corpus D

Output: A validated list of opinion features of the user domain related events.

Login

search(S.M/S.E)

extract(info)

For (word)

cal Tf

count domain

find MAX

measure interest

If (info=match (events))

recommend;

Else
return

D. Web Crawler Algorithm:

Web crawler can copy all the pagers the pages they visit for later processing by a search engine that indexes the downloads pages so that users can search them much more quickly

Input: set of popular URLs S

Output: Repository of visited web pages R

```
L1: If (s==NULL)
    then(P->S & P==NULL)
    get (P*)
    if(P*==R) then
        return 1;
else
    add (P*->R)&&(P* !=R,S)&(P*->s)
loop L1
```

IV. BACKGROUND RESEARCH

A. Opinion Mining:

Opinions and sentiments[2] expressed in text reviews can be generally analyzed at the document, sentence, or even phrase (word) levels. The purpose of document-level (sentence-level) opinion mining is to classify the overall Subjectivity or sentiment expressed in an individual review document (sentence).

Hatzivassiloglou and Wiebe [3] studied the effects of dynamic adjectives, semantically oriented adjectives, and gradable adjectives on predicting subjectivity; they proposed a supervised classification method to predict sentence subjectivity. particular resource vary across social media. The user review has a high impact on both the social media and search engine. From the user review, the domain is extracted. The domain extraction is based on the weighted of keywords used in the user review and score is also calculated for the keywords. The domain [4] proposed three machine learning methods, naive Bayes, maximum entropy, and support vector machines, to classify the whole movie reviews into positive or negative sentiments.. They found that standard machine learning techniques produced good results in comparison to human-generated baselines. Moreover, machine learning methods did not perform as well on sentiment classification as on a traditional topic based categorization. To prevent a sentiment classifier from considering irrelevant or even potentially misleading text, Pang and Lee [5] proposed to first employ a sentence-level subjectivity detector to identify the sentences in a document as either subjective or objective, and subsequently discarding the objective ones. They then applied the sentiment classifier to the resulting subjectivity extract, with improved results.

McDonald et al. [15] investigated the use of a global structured model that learns to predict sentiments on different levels of granularity for a textual review. The primary advantage of the proposed model is that it allows classification decisions from one level in the text to influence decisions at another. A regression method based on the bag of opinions model was proposed for review rating prediction from sparse text patterns [6]. Review rating estimation is a much more complicated problem compared to the binary sentiment classification. Generally, sentiments

are expressed differently in different domains. The sentiment classification methods discussed above can be tuned to work very well in a given domain; however, they may fail in classifying sentiments in a different domain. Bollegala et al. [7] proposed a cross-domain sentiment classifier using an automatically extracted sentiment thesaurus. An unsupervised learning method was proposed to classify review documents as thumbs up (positive) or thumbs down (negative) in [8]. The sentiment of each review document is predicted by the average sentiment orientations of phrases in the review.

Domain-dependent Contextual information is also considered for better estimation of the phrase sentiments. One limitation of this work is its reliance on an external search engine. Zhang et al. [9] proposed a rule-based semantic analysis approach to classify sentiments for text reviews. They used ord dependence structures to classify the sentiment of a sentence, and predicted document-level sentiments via aggregating the sentence sentiments. Rule-based approaches like this typically suffer from poor coverage due to the lack of comprehensiveness in their rules. In addition, Maas et al. [10] presented an approach to document-level and sentence-level sentiment classification tasks, which uses a mix of unsupervised and supervised techniques to learn word vectors by capturing semantic term-document information as well as rich sentiment content. Differently, sentiment analysis at the phrase (word) level mainly focuses on classifying sentiment polarities of opinion phrases (words). Generally, the sentiment polarity of an opinion word is usually context-dependent as well as domain-specific. Wilson et al. [11] presented an approach to predicting contextual sentiments at the phrase level by applying machine learning techniques on a variety of feature factors.

Yessenalina and Cardie [12] presented a compositional matrix-space model for phrase-level sentiment analysis. One of the benefits of the proposed approach is that by learning matrices for words, the model can handle unseen word compositions (e.g., unseen bigrams) as long as the component unigrams have been learned. A two-level affective reasoning method was proposed to mimic the integration of conscious and unconscious reasoning to address word-level sentiment analysis tasks [13]. Note that opinion mining at the document, sentence, or phrase (word) level does not discover what exactly people liked and disliked in reviews. In other words, it fails to associate the identified sentiments to the corresponding features commented on in the reviews. Clearly, an extracted opinion without the corresponding feature (opinionated target) is of limited value in reality [1]. Next, we survey existing work on extracting opinion features.3. Recommendations in social networks there are findings in the sociological and psychological disciplines that point to the relevance of a person's social network in determining their tastes, preferences, and activities. The principle of Homophile, for instance, is well established in the Social Networks field. McPherson et al. reported how "similarity breeds connection". They discovered that "people's personal networks are homogeneous with regard to many sociodemographic, behavioral, and intrapersonal characteristics". In other words, we share many attributes with the people close to us.

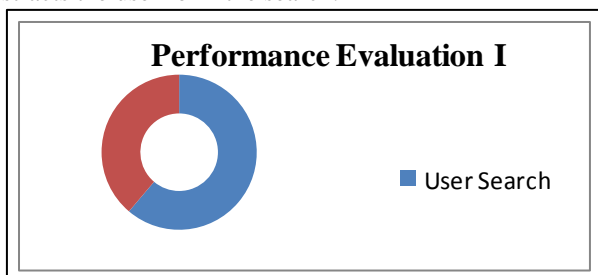
Reversing this principle suggests that, if we have information about the connections in a person's network, we can infer some of the person's attributes. It is possible that at least some of the similarities within a network are caused by the influence and interactions of the people in the network. People tend to remember information that was concretely given to them (that is, in personal interactions) better than abstract information (like statistical base rates). For example, Hogarth states that when considering to buy a certain car model we will likely give more thought to the direct advice of a friend than to each of the 100 respondents to a survey in a specialized magazine. More specifically, Leskovec et al. discuss the phenomenon of information cascades, in which individuals adopt a new action or idea due to influence by others. In the most extreme cases, knowledge about a full network's behavior determines the behavior of its members –making a “top hits” list available in a music downloading website affects the popularity of the songs, and several different networks, kept in isolation of each other, prefer completely different songs, to the point that it is impossible to predict which will be the most popular songs for a network without observing the behavior of the users in the network.

V. EXPERIMENT DESIGN

The experimental design is done using the .Net framework and Microsoft SQL server. The application is developed using the Microsoft Visual Studio and the SQL server is used as the database. In social media, user gets like, share, view the information on any subject. In search engine, user gets view, like, download the information on any subject. Now we consider only likes in social media or downloads in search engine. After that we calculate the how many likes are there different domains in social media or how many downloads are in different domain in search engine. Find the maximum like in social media or downloads in search engine. We have to the maximum like in social media or the maximum download in search engine on the subject is domain of the user. Now recommendation system is used. The domain related events are recommended to the user in social media The algorithm used are the opinion mining, exotic approach, pattern matching, recommendation system, and web crawler.

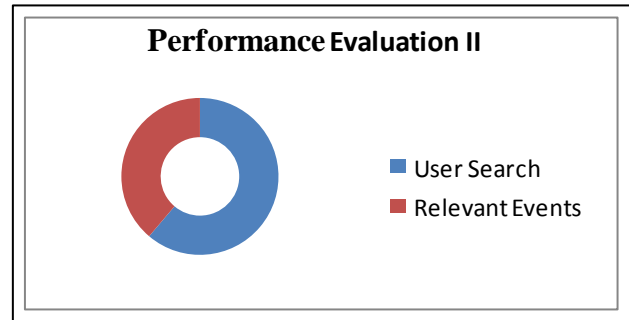
VI. PERFORMANCE EVALUATION

We conducted various experiments to comprehensively evaluate the EDR performance on more hours. We compared the proposed result and existing result. The advertisements in the existing system are high and it distracts the user form the search.



Graph 1: Advertisement Versus User search

The existing system has high advertisements and the proposed system has relevant events compared to the existing system. The proposed system does not distract the user from their search.



Graph 2: Relevant Events Versus User search

VII. CONCLUSIONS

Social media is an emerging trend for communication between people worldwide and it has attracted millions of users. A typical characteristic of such sites is that they allow anyone to post or like anything they like on any subject. Here the user's feedback and satisfaction is playing a vital role. In some cases the user may not give the feedback or reviews which they have viewed or downloaded directly. Instead of that the user is just searching the available things based on their interest. So there is no possibility of capturing the domain interest behavior of the user explicitly. This paper proposes to detain the subjective relevant events based on user likelihood by using the exotic approach. Experimental results demonstrate that the EDR but also outperform main stream method, namely, web crawler algorithm.

REFERENCES

- [1] B. Liu, "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2003
- [2] Identifying Features in Opinion Mining via Intrinsic and Exotic Domain Relevance, Zhen Hai, Kuiyu Chang, Jung-Jae Kim, and Christopher C. Yang.
- [3] V. Hatzivassiloglou and J.M. Wiebe, "Effects of Adjective Orientation and Gradability on Sentence Subjectivity," Proc. 18th Conf. Computational Linguistics, pp. 299-305, 2000.
- [4] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment Classification Using Machine Learning Techniques," Proc. Conf. Empirical Methods in Natural Language Processing, pp. 79-86, 2002.
- [5] B. Pang and L. Lee, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts," Proc. 42nd Ann. Meeting on Assoc. for Computational Linguistics, 2004.
- [6] L. Qu, G. Ifrim, and G. Weikum, "The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns," Proc. 23rd Int'l Conf. Computational Linguistics, pp. 913-921, 2010.

- [7] D. Bollegala, D. Weir, and J. Carroll, "Cross-Domain Sentiment Classification Using a Sentiment Sensitive Thesaurus," *IEEE Trans. Knowledge and Data Eng.*, vol. 25, no. 8, pp. 1719-1731, Aug. 2013.
- [8] P.D. Turney, "Thumbs Up or Thumbs Down?: Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Ann. Meeting on Assoc. for Computational Linguistics*, pp. 417-424, 2002.
- [9] C. Zhang, D. Zeng, J. Li, F.-Y. Wang, and W. Zuo, "Sentiment Analysis of Chinese Documents: From Sentence to Document Level," *J. Am. Soc. Information Science and Technology*, vol. 60, no. 3, pp. 2474-2487, Dec. 2009.
- [10] A.L. Maas, R.E. Daly, P.T. Pham, D. Huang, A.Y. Ng, and C. Potts, "Learning Word Vectors for Sentiment Analysis," *Proc. 49th Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies*, pp. 142-150, 2011.
- [11] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis," *Proc. Conf. Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 347-354, 2005.
- [12] A. Yessenalina and C. Cardie, "Compositional Matrix-Space Models for Sentiment Analysis," *Proc. Conf. Empirical Methods in Natural Language Processing*, pp. 172-182, 2011.
- [13] E. Cambria, D. Olsher, and K. Kwok, "Sentic Activation: A Two-Level Affective Common Sense Reasoning Framework," *Proc. 26th AAAI Conf. Artificial Intelligence*, pp. 186-192, 203.
- [14] F. Li, C. Han, M. Huang, X. Zhu, Y.-J. Xia, S. Zhang, and H. Yu, "Structure-Aware Review Mining and Summarization," *Proc. 23rd Int'l Conf. Computational Linguistics*, pp. 653-661, 2010.
- [15] S.J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.

