|| Volume 2 ||Issue 1 ||AUGUST 2017||ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Sentiment Analysis Based on Interpretive Structure Modelling For Feature Selection and Extraction

Ms. Kanade Kalpana Tukaram
ME CSE, EES College of Engineering Aurangabad, Maharashtra, India
kanadek25@gmail.com

***Abstract*— Sentiment analysis and opinion mining is an emerging area of research for analysing web data and capturing the sentiment of the users. Sentiment analysis is the calculative analysis of views, sentiments, opinions and positivity or negativity of a text. This paper identifies the factors that are responsible for the different sentiments of a person regarding a particular entity. In this work, the objective is to categorize the factors that influence the system of sentiment analysis due to varied sentiments of an individual. The evaluation of the proposed model ISM, using the performance evaluation measures prove that sentiment expressed in a post should be considered as an important factor for identifying influential bloggers. The findings also indicate that ISM shows better results to find the influential bloggers as compared to the existing methods.**

Keywords: Sentiment Analysis, Level of Sentiment Analysis, Feature Extraction and Selection, ISM (Interpretive Structure Modelling)

I INTRODUCTION

Sentiment analysis is a process of extracting and understanding the sentiments defined in the text document. The explosion of data in the various social media channels like twitter, Facebook, and LinkedIn has given consumer new way of expressing their opinion on a particular product, person and places. The user opinion is always in the form of textual information. Per day millions of textual message data is sent over social media or online shopping website. Investigating and analysing the sentiment of the opinion is a very critical task to perform. The NLP with artificial intelligence capability and text analytics are used to determine whether the sentiment of the opinion is positive, negative and Neutral. The opinion mining and sentiment analysis is doesn't depend on any particular domain or platform. It spreads to all the social media networks, healthcare, management, economy and many more and also it is very useful for the growth of many companies and organizations.

Sentiment analysis is also providing a business intelligence which can be used to make good impactful decision. Sentiment analysis and sentiment classification are the two methodologies used in opinion mining. Whereas both have its own independent features, but sometimes it may be used interchangeably.

Sentiment classification indicates the sentiment orientation by assigning the class labels to the document or segment. Sentiment orientation is a kind of text classification that classifies text data based on the sentiment orientation of opinion. Sentiment orientation indicates the polarity of the opinion either true or false based on subjectivity [1]. Subjective analysis is a process of identifying whether the given text or reviews data is subjective or objective in nature.

Sentiment analysis is extremely useful in various situations. But it is very difficult process because of the complexity involved in the human language. It has got several variants like grammatical, cultural etc. Humans can easily interpret statements like "My order been delayed. Very good". But it may be difficult for the machine to understand. Similarly word "thin" may be taken as positive with respect to laptop but it may be negative when it comes to apartment wall. So to give the correct decision sentiment analysis must be sometime more business specific.

II LEVEL OF SENTIMENT ANALYSIS

In general the process of sentiment analysis is divided into 3 levels.

A. **Document level analysis***:* Document level sentiment analysis determines the overall opinion of the document. Apart from the sentiment orientation of the individual sentences it classifies the sentiment expressed by the whole document. The classification is expressed by either positive or negative sentiment [2]. This level analysis is helpful only if the document related to a single entity, because it expresses opinion on a single entity (E.g. Product, Person). Thus it is not applicable to the documents which contain the comparison of the multiple entities.

B. **Sentence level analysis***:* The sentence level sentiment analysis is also called as subjectivity classification [3]. It distinguishes subjective information from the objective information. It considers each sentence as a separate unit and it predicts that the sentence must contain only one opinion. It iterates each sentence and determines whether the sentiment orientation of the sentence is positive, negative or neutral.

C. **Entity and Aspect level analysis***:* Entity and aspect level sentiment analysis captures the mixture of emotion from the review sentence. It performs fine-grained sentiment analysis and it differentiates what actually user wants and do not-want. One main feature of aspect level sentiment analysis is that, it directly looks into the opinion instead of paragraph, sentences, phrases

and document. The goal is to find sentiments on entities and their aspects. Meant to say that aspect level sentiment analysis concentrate on sentiment (positive or negative) and a target (opinion).

A sentence or a document may contain a combination of positive and negative opinions.

1. Sentiment level opinion mining is performed by two tasks i.e. subjective and objective.

Objective: I purchased a new Android Smartphone few months ago.

Subjective: It is such a nice Gadget.

2. The polarity of the opinion is determined by using the subjective sentence.

Positive: It is such a nice Gadget.

Negative: It has poor radio signal reception quality.

In a document level sentiment analysis it determines the overall document and a sentence. The classes of the opinion are determined by polarity. i.e. either positive or negative. The binary classification method is used for this purpose. Most of the existing sentiment analysis algorithm uses the binary Classification methods. Mean to say that they assign review or opinions to bipolar classes such as positive or negative. Binary sentiment classification classifies reviews or opinions by using multi-point rating scale. Multi-point rating scale uses the rating inference. Using this rating inference class labels are assigned as scalar rating such as 1 to 5 "stars".

III RELATED WORK

The user obtained information in web will give lot of details in extracting the user opinions. It is difficult for the people to classify the opinions into positive and negative [2]. All product reviews to be summarize and sentiments are to be classified [3]. This paper analyses the sentiment for fine grained and classifies the sentiments using the sentences after extracting the relevant features [4]. Extracting the user's opinion from web includes getting review about the products [5]. At last for the both preferences the values are classified and improve the recommendation accuracy and hit ratio and overcome the misclassification [6]. In [7] the author proposed a method to the understand reviews feature additional efficiently.

Sentiment analysis are classified into three types: document, sentence and aspect level. The document level sentiment analysis aim is to check for the entire document as positive or negative [8] and opinion is determined as positive, negative or neutral during the sentiment analysis process [9]. The aspect level sentiment analysis finds the target of the opinion indicating that every opinion has a target [10]. The bag-of-words method for sentiment analysis in which the relationship between words was not at all considered and a document is represented as just a collection of words [11]. To determine the sentiment for the whole document, sentiments of every word was determined and those values are united with some aggregation functions. In this paper the feature extraction is done using naive bays algorithm**.**

The sentence level sentiment analysis finds that if each Sentence includes a positive, negative or neutral opinion, for a product or service. [13].

IV FEATURE EXTRACTION AND SELECTION

Feature extraction techniques [4] reduce the feature vector length by transforming all the features in lower dimensional feature vector. It maps the high-dimensional data on lower dimensional space.

A. Unigrams

Unigram features are bag-of-words (BoW) [4] features extracted by removing extra spaces and noisy characters between two words. E.g. "The movie was awesome." Here, words "The", "movie", "was", and "awesome" are all unigram features.

B. Bigrams

Bigrams are the features composed of every two consecutive words in the text. For the above example, "The movie", "movie was" and "was awesome" are the bigram features. These features have potential of including some contextual information.

C. Bi-tagged

Bi-tagged features are selectively extracted using part-of speech (POS) based fixed patterns [5]. Bigrams containing mostly adverbs and adjectives are considered more sentiment bearing.

Feature selection techniques [5] select the minimum notable features and represent the class attribute in the reduced feature space. Feature selection techniques can significantly improve the classification accuracy and also provide better insight into important class features, resulting in a better understanding of sentiments.

A. Term Frequency (TF): Specifies number of times a feature appears in document

B. Feature Presence (FP): Determines whether a feature appears in a document or not. It gives value in 0 or 1.Where, 1 indicates presence and 0 indicates absence of feature.

C. Term Frequency- Inverse document Frequency (TF-IDF): Signifies importance of a particular word in a document.

V PROPOSED METHOD

The primary aim of researchers is to provide an optimized solution. Majority of problems have high complexity due to the presence of many factors and the interaction among them. These factors affect the software system directly or indirectly. Various methodologies have been employed to obtain optimal solution and ISM (Interpretive Structural Modelling) is one of the technique that is vastly used. Figure 1 shows the proposed architecture. The various stages of opinion mining includes data pre-processing, feature extraction using

machine learning algorithm, aspect based sentiment analysis using opinion summary.

1) Activity Features: The activity module is based on two features: (1) Number of blog posts (2) Sentiment of a blog post. The feature (1) represents the total number of blog posts of a blogger. We used MYSQL platform to extract the blog posts of a blogger. The blog posts are given as input to Senti Strength to measure the sentiment score of each blog post.

2) Recognition Features: The recognition module is based on the following three features: (1) Number of in links to blog post (2) Number of comments on a blog post (3) Number of out links from a blog post to other blog sites.

A. Proposed Model

These factors affect the software system directly or indirectly. Various methodologies have been employed to obtain optimal solution and ISM (Interpretive Structural Modelling) is one of the technique that is vastly used. Attri et al [11] defined ISM as a learning process which is interactive and frames an all-inclusive model comprising of directly and indirectly related factors. The first step in this technique is to identify the variables, which are significant to the problem. The next step in the sequence is to choose a relation that is contextually relevant. After the relation is decided, a structural self-interaction matrix (SSIM) is prepared on the basis of pair wise evaluation of elements. Subsequently a Reachability Matrix (RM) is obtained from the SSIM ensuring transitivity.

All the above steps then result into a matrix model. Subsequently, the elements are partitioned and ISM is derived. In this paper we are using ISM methodology to spot out the appropriate range of identified elements/variables of the sentiment based system. According to [5,6], ISM comprises of diversified elements which are directly or indirectly related and are ordered and arranged into a comprehensive systematic model. This resulted model incorporates the configuration of an intricate issue or concern in a framed structure comprising graphs and words [12, 13]. It is a stepwise process which comprises of seven steps. The steps are enlisted here. Description of all the steps and the result obtained from them has been explained in the next section.

1. Identification of Factors using Surveys
2. Contextual Relationship Development
3. Framing of Structural Self Interaction Matrix (SSIM)
4. Verification of Transitivity
5. Framing of Reachability Matrix
6. Formation of Digraph
7. Development of ISM Model

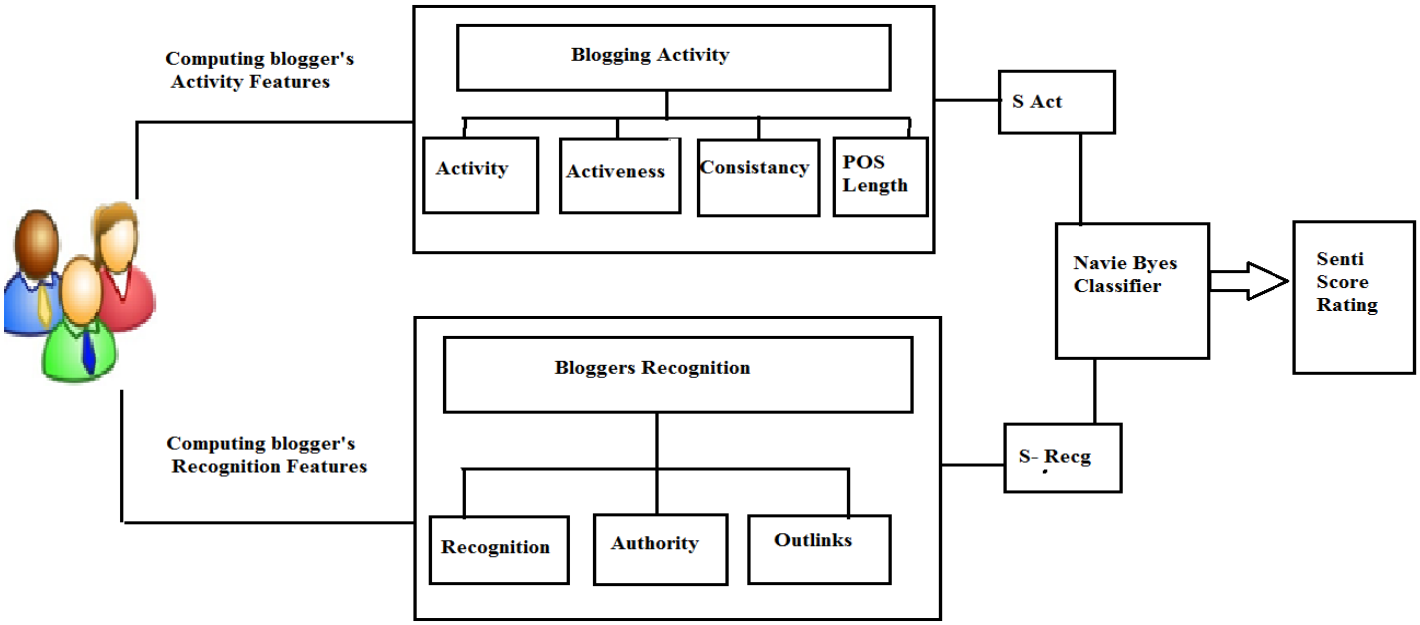


Figure1. Proposed Model for Feature Dependency using Sentiment Analysis

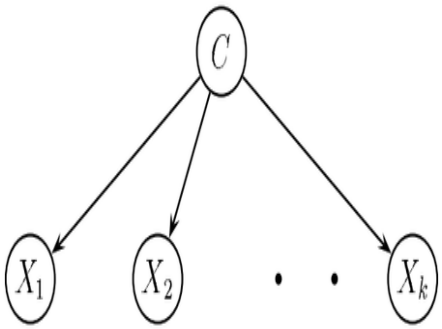
B. Proposed Algorithm

A Naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 3" in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple,

regardless of the presence or absence of the other features. For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers [(Zhang, 2004)]. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests [(Caruana & Niculescu–Mizil, 2006)].

An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix. We can visualize a naive Bayes graphically as follows:



This Bayesian network, predictive attributes Xi are conditionally independent given the class C.

VI CONCLUSION

In machine learning, the most informative features help in bringing the effectiveness to the model and increasing the learning capability of the classifiers. People are free to express their views and opinions publicly and these opinions are sentiments of peoples that can be analysed to understand the

Polarity of the reviews. In the start our main aim was to identify a predictive model to capture and classify the sentiment of the users for review. For doing this a comparative analysis between feature selection with the help of Bayesian classifier have been performed. After analysis of the opinion of the different users with the help of different models our aim has been successfully achieved.

ACKNOWLEDMENT

It is my great pleasure in expressing sincere and deep gratitude towards my guide **Prof. B. K. Patil** for providing me various resources, valuable and firm suggestion, guidance and constant support throughout this work.

REFERENCES

[1] Bing Liu," Exploring User Opinions in Recommender Systems", Proceeding of the second KDD workshop on Large Scale Recommender System and the Netflix Prize Competition", April 2012, LasVegas, USA.

[2] AntonioMoreno- Ortiz, Javier Fernández-Cruz," Identifying polarity in financial texts for sentiment analysis: a corpus-based approach", 7th International Conference on Corpus Linguistics: Current Work in Corpus Linguistics: Working with Traditionally conceived Corpora and Beyond (CILC 2015)

[3] Zhang Wenhao, Hua Xu, Wan Wei. Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. Expert Syst Appl 2012.

[4] E.Haddi, X.Liu and Y.Shi,"The Role of Text Pre-processing in Sentiment Analysis," Procedia Computer Science, 2013, vol. 17, pp. 26-32.

[5] B. Agarwal and N. Mittal, "Machine Learning Approaches for Sentiment Analysis," In Springer International Publishing Switzerland , 2016, pp.193-208.

[6] Sage A.P., Interpretive structural modelling: Methodology for large scale systems, New York, NY: McGraw-Hill (1977)

[7] Xing Fang ,Justin Zhan," Sentiment Analysis using product review data", Springer: Journal of Big data", 2015, North Carolina A&T State university, Greensboro, NC,USA.

[8] Subhabrata Mukherjee, Pushpak Bhattacharyya, "Feature Specific Sentiment Analysis for product Reviews", IET,2015, IIT Bombay.

[9] Himabindu Lakkaraju, Chiranjib Bhattacharyya, Indrajit Bhattacharyya and Srujana Merugu," Exploiting Coherence for the simultaneous discovery of latent facts and associated sentiments", SIAM International Conference on Data Mining (SDM), April2011.

[10] Mining Hu and Bing Liu," Miming and Summarizing customer reviews", KDD 04: proceedings of the tenth ACM SIGKDD international Conference on knowledge discovery and data mining.

[11] Warfield J.W., Developing interconnected matrices in structural modelling, IEEE Transactions on Systems Men and Cybernetics, 4(1), 51- 81 (1974)

[12] Raj T., Attri R. and Jain V., Modelling the factor affecting flexibility in FMS, International Journal of Industrial and System Engineering, 11(4), 350-374 (2012).

[13] Ravi, V. & Shankar, R. (2005). Analysis of interactions among the barriers of reverse logistics. Technological Forecasting and Social Change, 72(8), pp.1011-1029.

[14] Android Based Health Care Monitoring System Devashri Deshmukh, Ulhas B. Shinde || Volume 2 ||Issue 7 ||JAN 2017||ISSN (Online) 2456-0774 International Journal Of Advance Scientific Research And Engineering Trends

[15] Singh, M.D.Shankar, R., Narain, R., &Agarwal, A. (2003). An interpretive structural modelling of knowledge management in engineering industries. Journal of Advances in Management Research, 1(1), 28-40.90