

Implementation of Sentiment Comments

Sharuti Jain¹, Er. Jyoti Arora²

¹M.Tech student, Department of Computer Science, Desh Bhagat University, Punjab

²Assistant professor, Department of Computer Science, Desh Bhagat University, Punjab

Abstract:- In this paper, we have studied emotional comments made by Indian Users in data mining and reviewed to concept-level sentiment analysis provide novel approaches to opinion mining. In this we extract the sentic patterns. A recent knowledge base technology uses sentic computing which rely on common-sense computing, and machine learning for enhance the accuracy of tasks. The Web has drastically changed the way that individuals express their perspectives and suppositions. They can now post surveys of items at vendor locales and express their perspectives on practically anything in Web gatherings, examination gatherings, and online journals, which are by and large called the client created content. This online word-of-mouth conduct speaks to new and quantifiable wellsprings of data with numerous useful applications. Sentiment analysis utilizes opinion terms defined as a part of the assessment vocabulary and estimation designs in the feeling example database, Other than descriptors, other substance words.

Keywords:- Sentiment analysis, Linguistic rules, POS tagging, Opinion mining, Machine learning

1. INTRODUCTION

Data mining is a division of computer science. It is the procedure of discovering patterns in large data sets. The aim of the data mining process is to abstract information from a data set and converts it into an understandable structure. Data extracted from social networking sites is unorganized and indistinct. In everyday life conversations are noticed on social networking sites, people do not care about the spellings and accurate grammatical rules of a sentence that leads to different types of uncertainties, such as lexical, syntactic, and semantic. So, analyzing and extracting information patterns from such data sets are more complex. A significant amount of research has already been carried out to categorize sentence into various categories of emotion. The emotions that have been

Knowledge Base

It is of two types:

Common knowledge base and common sense knowledge base.

worked upon are either positive or negative or finding the polarity. The relationship between emotion and text is also important. When mapping textual information to emotion space. The study of emotions in written text is conducted from two opposite points of view.

- 1) The first is the viewpoint of a writer.
- 2) The second is concerned with how a reader understands the emotion in a text.

In this paper, the second point of view is taken because we are interested in the way people understand emotions.

We are able to prepare bag of words for anger and fear especially effective for data obtained from Facebook which has posts in Hinglish. The results delivered by many online tools available for the categorization of data into various emotions are not effective. Most of them provide only 60% of the correct result, which are much less than the required amount. Also we need to be extra careful while categorizing a word into a particular category of emotions, because there are various words that represent different emotions in different context. So the most of the emphasis must be on obtaining a perfect bag of words, where we live in a "nothing is perfect" era.

In this paper we discussed to concept level sentiment analysis, knowledge base technology, common sense computing and machine learning.

Concept Level Sentiment Analysis

The analysis at concept-level is studied to understand the semantic and affective information similar with natural language opinions and, so, to allow a relative graceful feature based sentiment analysis. Alternatively gathering isolated opinions about a whole item, (e.g. Samsung LED) users are generally more interested in comparing different products according to their specific features (e.g. Samsung LED vs. Onida LED) or even sub features.

Common knowledge is command that is known by all. Common knowledge need not involve one specific subject. So, common knowledge can be concern with a number of subjects. On the other hand common

sense knowledge is the collection of facts and information. It is thing which is acquired by a person's day to day experience. So there is a difference between both types of knowledge. A machine does not contain such kind of knowledge as humans. Attempts to build both the knowledge bases are countless.

Machine Learning

The research of pattern identification and computational learning theory. It investigates the construction and research of algorithms that can absorb from and make projection on data. Such algorithms perform by building a model. It emphasis on prediction, based on *known* stuff learned from the data. It has strong connection to mathematical expansion, which distribute methods, theory and application domains to the sector. It is used to upgrade the accuracy of the tasks.

The research of emotions is the puzzled part. Now the emotions were explained using a list of tables. Such approaches are inadequate to explain overall emotions and multi word opinions.

The Hourglass of Emotions (Figure 1) is a biologically excellent and psychologically-inspired model goes above the precisely and dimensional approaches. It may explain human emotion in terms of four dimensions.

The four dimensions are as follows: Pleasantness, Attention, Sensitivity, and Aptitude. Each of these dimensions contains six sentic levels.

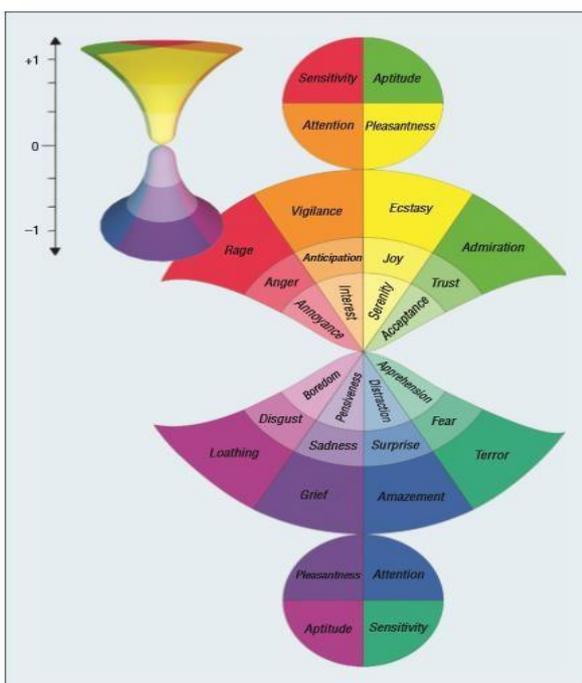


Fig 1: Hourglass of Emotions

2. REVIEW OF RELATED WORK

Poria, Soujanya et al. [1] has proposed the use of Sentic patterns for the purpose of the sentiment analysis from the social data. The authors proposed the use of dependency-based rules for concept-level sentiment analysis. In this work, the authors have introduced a novel paradigm to concept-level sentiment analysis, common-sense computing, and machine learning for improving the accuracy of tasks.

Yassine, Mohamed et al. [2] has worked on the development of a framework for emotion mining from text in online social networks. This paper presents a new perspective for studying friendship relations and emotions' expression in online social networks where it deals with the nature of these sites and the nature of the language used. It mainly uses the k-means clustering algorithm.

Narayanan, Vivek et al. [3] has worked on a fast and accurate sentiment classification using an Naive Bayes model. The authors have explored different methods of improving the accuracy of a Naive Bayes classifier for sentiment analysis. They have also observed that a combination of methods like effective negation handling, word n-grams and feature selection by mutual information results in a significant improvement in accuracy.

Cambria, Erik et al. [4] has implemented the semantic multidimensional scaling for open-domain sentiment analysis. In this work, the largest existing taxonomy of common knowledge is blended with a natural-language-based semantic network of common-sense knowledge, and multi-dimensional scaling is applied on the resulting knowledge base for open-domain opinion mining and sentiment analysis.

Gun Woo Park et al. [5] stated as Social Media are at the heart of our communications and are among the most visited places on the Web. Social network services like Facebook, Friendster, Myspace and Orkut have established themselves as very popular and powerful tools for making and finding friends for identifying other people who have similar interest. Search behavior of Web users reflects the interest of Web users and this also leads to similar profiles. Some research has also being carried out so as to identify people who are highly associated with similar interest of Web search

Federico et al. [6] of the researchers considered that the Web is a huge virtual space to express individual opinions and influence many aspect of life. Internet contains a treasure of data that can be mined to detect valuable opinions, with inferences even in the

political field. The snippets of text are a gold mine for companies and individuals that want to monitor their reputation and get timely feedback about their products and actions. Sentiment analysis offers these organizations the ability to monitor the different social media sites in real time and act accordingly. The revolution in information technology has helped in making Web Sentiment Analysis at less cost than ever before, even if an excess of information has always meant a no usable knowledge. The most valuable information is often encoded in thousands of textual pages, whose automatic treatment is inextricably linked to the automatic lexical analysis and semantic synthesis of contents.

Baumer *et al.* [7] also mentioned that social networking Websites create new ways for engaging people belonging to different communities and regions. Social networks facilitate their users to communicate with people exhibiting different moral and social values. These Websites provide a very powerful medium for communication among individuals that leads to mutual learning and sharing of valuable knowledge as studied by Sorensen. The most popular social networking Websites are those that allow people to communicate with each other by joining different communities and groups.

3. PROS AND CONS OF EXISTING APPROACHES

The existing scheme is entirely based upon the use of sentic pattern analysis of the purpose of emotion detection from the dataset collected from the social media data. The existing scheme evaluates various emotions like joy, serenity, sadness, grief, surprise, interest, rage, anger, fear, terror, trust, admiration, disgusting, etc.

The existing model is aimed at fetching the words containing the emotions and then calculates the message score on the basis of the points pre-stored in the word database.

The existing model is completely based upon the knowledge-based dataset.

The existing model does evaluate the points on the basis of the concept-level dependency based rules on the combination of the words.

The existing model is not capable of evaluating the three word combination, eg "pretty big fear", "extreme anger burst", etc.

Limitation of the model is that as the data sets contain the text that has posts/comments in different languages like Hindi, Urdu, Tamil and Punjabi. So the

model is unable to consider comments in such languages.

4. PROPOSED METHODOLOGY

To study the literature on the sentiment analysis, opinion mining and other similar techniques in order to study the shortcomings and strengths of the existing systems. To create and design the proposed solution for sentiment analysis to mitigate the shortcomings of the existing models. To implement the proposed model using the NETBEANS simulator. To evaluate the performance of the proposed model implementation using various types of social datasets. To obtain and conclude the results obtained for the final implementation of the proposed model.

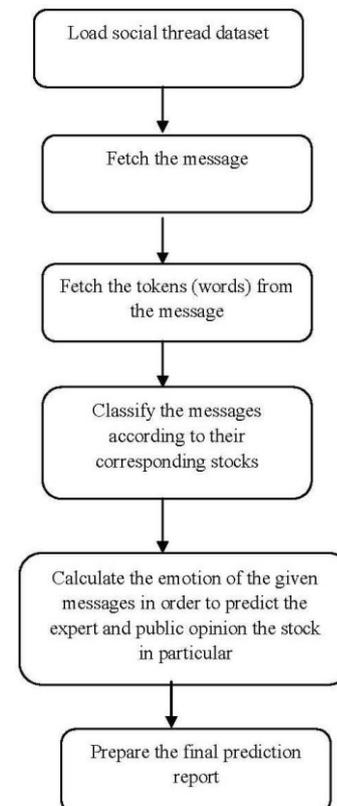


Fig: Work flow diagram of the Proposed model

5. RESULT

For Sentiment Analysis System, we used JAVA programming language to program Sentiment Analysis. The proposed work will be based upon the evaluation of the message based on the one, two or three word combinations. The proposed work will be using the large knowledge based database for the purpose of tokenization, which is the process to fetch the emoticon words out of the message. The emoticon words the processed for the score calculation based upon the single words and word combinations in the knowledge database. The model

will be able to learn the new word combination based upon the knowledge database for learning. The model will also offer the user feedback for the higher accuracy and a human-software intelligent interaction, which can help the machines to learn the new rules for the higher accuracy than the existing system.

We are proposing the use of a rich set of sentiment analysis features like positive, negative, fear, anger, etc. This feature selection method can improve opinion classification performance. This Feature Relation Network is rule-based sentiment classification method that finds the text features like anger and fear also. System. We used dictionary files for 4 emotion types: Happy (or Positive), Sad & Fear (or Negative) and Neutral. To cover up everything, we also defined a class of neutral Sentiment which is classified when the sentiment does not match any of the above sentiment.

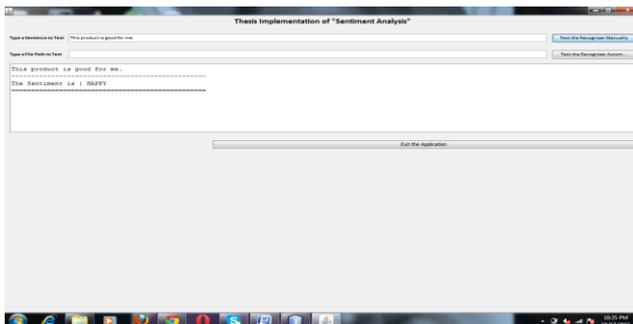


Fig: A running sample of the Code. Predicted as HAPPY (Positive)

This code runs and reads the sentence given to it. It then extracts the Nouns, Adjectives, Adverb and then uses the dictionary to predict the sentiment. Based on the weight values of each dictionary result, the sentiment is predicted. In the above case, the actual comment is a positive response and it does come out as positive.

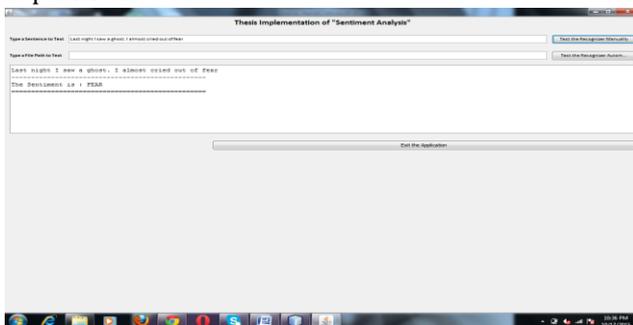


Fig: Another sample "Last night I saw a ghost. I almost cried out of fear". Predicted as FEAR

In the above example a sentiment FEAR is correctly predicted based on the analysis of the words and the dictionary weight value.

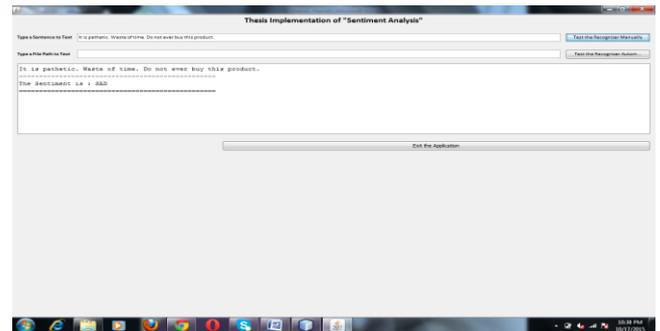


Fig: Another Success. "It is pathetic. Waste of time. Do not ever buy this product". Predicted as SAD (or negative)
Since, our system can only predicts between Happy or Positive, Sad or Negative, Fear and Neutral. So, anger or a bad review is always predicted as SAD.

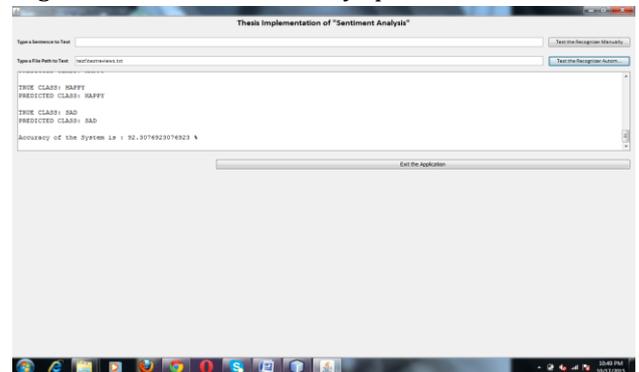


Fig: Accuracy of the system

This figure shows the accuracy of the system. In this we create the different texts are created in the text file and we gave the true class (sentiment) to the text and system gave the predicated class (sentiment) to the text.

The accuracy of the system is changed with different type of texts.

6. CONCLUSION

In this thesis, everyday life conversations are noticed on social networking sites, people do not care about the spellings and accurate grammatical rules of a sentence that leads to different types of uncertainties, such as lexical, syntactic, and semantic. This online word of-mouth conduct speaks to new and quantifiable wellsprings of data with numerous useful applications. Presently if one needs to purchase an item, he/she is no more restricted to asking his/her loved ones in light of the fact that there are numerous item audits on the Web which give feelings of existing clients of the item. The emotions that have been worked upon are either positive or negative or finding the polarity. We have developed a Sentiment analysis System and we used Score based analysis to predict the sentiment of a user from his/her comments. The System is

quantitatively tested and the results are produced. The system has difficulty in predicting the sentiment if the sentence is very small or has a few words in it. With a fair size of 3 sentences text is predicted very accurately. In this we calculate the accuracy of the system with different types of comments.

REFERENCES

- [1] Poria, Soujanya, Erik Cambria, Grégoire Winterstein, and Guang-Bin Huang. "Sentic patterns: Dependency-based rules for concept-level sentiment analysis." *Knowledge-Based Systems* 69 (2014): 45-63.
- [2] Yassine, Mohamed, and Hazem Hajj. "A framework for emotion mining from text in online social networks." In *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, pp. 1136-1142. IEEE, 2010.
- [3] Narayanan, Vivek, Ishan Arora, and Arjun Bhatia. "Fast and accurate sentiment classification using an enhanced Naive Bayes model." In *Intelligent Data Engineering and Automated Learning-IDEAL 2013*, pp. 194-201. Springer Berlin Heidelberg, 2013.
- [4] Cambria, Erik, Yangqiu Song, Haixun Wang, and Newton Howard. "Semantic multidimensional scaling for open-domain sentiment analysis." *Intelligent Systems*, IEEE 29, no. 2 (2014): 44-51.
- [5] GunWooPark, Soojin Lee, SangHoon Lee, "To Enhance Web Search Based on Topic Sensitive Social Relationships Ranking Algorithm in Social Networks", *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops*, 2009.
- [6] Ronen Feldman, "Techniques and Applications for Sentiment Analysis", *Communications of the ACM*, Vol. 56 No. 4, Pages 82-89, 2013.
- [7] Federico Neri, Paolo Geraci & Furio Camillo, 'Monitoring the Web Sentiment, the Italian Prime Minister's case', *International Conference on Advances in Social Networks Analysis and Mining*, 2010 IEEE.
- [8] Baumer E. P. S., Sinclair J. & Tomlinson B., 'America is like Metamucil: Fostering critical and creative thinking about metaphor in political blogs', In *Proceedings of 28th International Conference on Human Factor in Computing Systems ACM*, Atlanta, GA, USA, pp. 34-45, 2010
- [9] Sorensen L., 'User managed trust in social networking comparing Facebook, myspace and linkdin', In *Proceedings of 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic System Technology, (Wireless VITAE 09)*, Denmark, pp. 427-431, 2009.
- [10] Evans B. M., Kairam S. & Pirolli P., 'Do your friends make you smarter: An analysis of social strategies inonline information seeking', *Information Processing and Management*, 46 (6), pp. 679-692, 2010.
- [11] Li J., Li Q., Khan S. U. & Ghani N., 'Community-based cloud for emergency management', In *Proceedings of the 6th IEEE International Conference on System of Systems Engineering (SOSE)*, Albuquerque, USA, pp. 55-60, 2011.
- [12] Li J. & Khan S. U., 'On how to construct a social network from a mobile ad hoc network', *Technical Report*, NDSU-CS-TR-09-009, 2009.
- [13] Lionel Martin, 'User Behavior under the Influence of Groups in Social Media', *EDIC Research Proposal*, 2009.
- [14] Divya Bohra, Sanjay Deshmukh, 'A Survey on Sentiment Analysis in NLP', *IJARCCCE*, Vol. 4, Issue 6, June 2015