International Journal of Technical Research and Applications e-ISSN: 2320-8163,
www.ijtra.com Special Issue 41 (AVALON) (March 2016), PP. 31-33

SENTIMENT ANALYSIS OF MOVIE REVIEWS IN HINDI

Alok Gupta[#1], DipeekaSonavane[*2], Kajal Attarde[#3], Neelam Shelar[#4], Pramila Mate[#5]

[#]*Computer Department, Terna Engineering College*
Mumbai, India

[1]alok.gupta1994@yahoo.com
[2]sonavanedipeeka24@gmail.com
[3]kajal.attarde1102@gmail.com
[4]neelamshelar@ymail.com
[5]pramila.mate@gmail.com

Abstract*—In this modern age of digitization, everyone uses online services for various day-to-day activities provided by numerous E-commerce websites. Based on the service provided we write reviews which explain and help other users to make decision about service available. These reviews or collection of data is very useful to improve and determine the pros and cons of service. Sentiment analysis is a field of Data Mining where user reviews are collected, analyzed and categorized in positive, negative and neutral reviews. In India, there has been a need for regional languages interfaces for better understanding about the web content and make it more user-friendly to native users. With that there was need for a system to categorize the user content of various regional languages. Here we propose a system to categorize movie reviews which are in Hindi into positive, negative and neutral for better decision making about the movie. The system utilizes a dictionary of positive and negative word list and phrases to determine the polarity of review. Special cases such as Negation handling, overall review of the movie is successfully implemented and tested on a dataset containing 500 reviews giving a total precision: 85.15% and recall: 97.23%.

Keywords*— Sentiment analysis, Data Mining, movie reviews, Hindi, phrases.

I. INTRODUCTION

Data mining is a widely growing concept now-a-days. The data flows in and out over the high speed Internet around the world in seconds through blog, websites, social networks, etc. and users around the world interact with each other using various features available at such sources. The data collected can be used to make web content more user friendly by analysing it according to user needs. For this the concept of forms, comments, likes, reviews, sharing etc. was developed where user can read/write about a content on web which will give clear understanding about the content to other users as well.

This collection of user content helps to analyze the user needs and also to help other online web users to know about the content on web. Sentiment Analysis is a natural language processing task that mines information from various text forms such as reviews, news, and blogs and classify them on the basis of their polarity as positive, negative or neutral.

The diversity in India is unique having variety of languages according to regions. With increase in web content and Internet era all the information is easily available to the people on Internet, but the information available is mostly in English which is not understood by the native people of India. Hindi language is widely understood in India. Many Governmental websites, Online Newspapers, blogs, etc. in India provide Hindi content to the user for better understanding. Here the user can also comment/read other user's views. But developing a system to determine sentiment of reviews in Hindi is very difficult due to insufficient resources such as pre-annotated corpus of Hindi words and no specific structure of the words used in a sentence which is why it is also known as free language in contrast with English language which has a specific structure to user words inside a sentence.

Our proposed system takes a movie review as input and analyses these review and checks for any overall polarity defined by the user. If not then checks for negation and compares the review with stored pre-annotated corpus consisting of words and phrases and assigns polarity to the review. The pre-annotated corpus consist of SentiWordnet [3] and we added some words that improved the dictionary to give a total precision: 85.15% and recall: 97.23%.

The rest of the paper is organized as follows: Section II describes the existing work done until now, Section III explains the proposed algorithm, Section IV gives the experimentation results and last section concludes the study.

II. EXISTING WORK

A lot of research has been done in the field of sentiment analysis on English Language but work on Hindi language is very few. Some of the efficient ones were:

SentiWordnet which was obtained by applying word lexical transfer technique on Bi-lingual dictionary [3]. Cross Lingual approach [5] to determine the polarity of the reviews by training the machine in one language and which is used to determine the sentiment of reviews of other language using machine translation. Improved Hindi SentiWordnet with rules for negation and discourse integration handling [4]. Using pre-annotated corpus of most common words used in Hindi and creating an unsupervised dictionary that will determine the polarity of review and check for negations [6].

III. PROPOSED WORK

The proposed work is similar to [4] and [6] using a pre-annotated corpus and additional inclusion of phrases, checking for overall polarity of the review with negation handling.

Input is taken from static dataset which were collected from various online sources [9] [10] [11] [12]. Some reviews have been translated from English to Hindi [13] and given to the system.

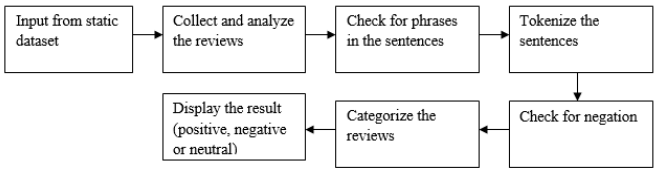

Figure 1. Block diagram of Overall working of system

After collection of reviews the phrases are checked in the sentences. Then sentences are tokenized into words and then the tokenized words are check for negation, if negate words are present then the preceding and succeeding word is checked for polarity. Comparison is done between number of positive and negative words depending on that the review is categorized into positive, negative or neutral and the result is displayed.

A. Proposed Algorithm

1. Start
2. Take input from static dataset.
3. Check if any overall polarity of movie is defined.
   If found then
   Directly assign the polarity
   Go to Step 7
   Else
   Continue
4. Compare all the phrases in positive and negative dictionary with review
   If match found
   Increment the respective counter
   Else
   Continue
5. For remaining words in the sentence
   Tokenize the sentence
   Check for negation
   If found
   Assign opposite polarity
   Else
   Continue
6. For all other remaining tokenized words
   Check in unigram positive and negative dictionary
   If found then
   Increment the respective counter
   Else
   Continue
7. Compare the positive and negative counter and display the result.
8. Stop.

Figure 2. Proposed Algorithm

B. Overall polarity detection

Sometimes user list some of the features of movie he liked and some which he disliked and provide an overall rating or polarity for the movie in his review. Proposed system will check for reviews which contain –कलमिलाकर", "सपर", "सिसत्"
and will check for positive or negative word present after that and assign the respective polarity.

Negative Sentence

वही पुरानी दिनचर्या कहानी। फिल्म में रोमांस काफी बेहतर था।

कुल मिलाकर यह फिल्म बेकार है ।

Figure 3. Example of Overall polarity detection

C. Negation Handling

Since Hindi sentences are unstructured, the negation can be a little tedious. Proposed system searches for –नहीं" and then compares the preceding and succeeding 3 words until a match is found in positive or negative pre-annotated dictionary.

Positive Sentence

यह फिल्म खराब नहीं है ।

Negative Sentence

फिल्म में दिलचस्प कुछ भी नहीं है ।

Figure 4. Example of Negation Handling

D. Polarity of Review

Here the review is checked for number of matches found in pre-annotated dictionary and compared. The majority opinion polarity is assigned to the review.

Positive Sentence

यह फिल्म शानदार है ।

फिल्म की कहानी नाटक अच्छी तरह से किया गया था।

Negative Sentence

यह फिल्म खराब है ।

संगीत और ज्यादा बेहतर हो सकता था ।

Figure 5. Polarity review with word and phrase detection

IV. EXPERIMENTATION RESULTS

Experiment is conducted on movie reviews using static dataset which contained Hindi reviews from various online websites. Input is applied to the system and output is given by the system by comparing the positive and negative matches with negation handling. In case of overall polarity if present in the input review then succeeding words are matched in pre-annotated dictionary to assign polarity.

Three evaluation measures are considered [6] based on which performance of system is calculated:

1. Precision: portion of true positive predicted instances against all positive predicted instances.

$$Precision = \frac{tp}{tp + fp}$$

2. Recall: portion of true positive predicted instances against all actual positive instances.

$$Recall = \frac{tp}{tp + fn}$$

3. Accuracy: portion of true predicted instances against all predicted instances.

$$Accuracy = \frac{tp + tn}{tp + tn + fp + fn}$$

Precision	Recall	Accuracy
85.15 %	97.23 %	85.42 %

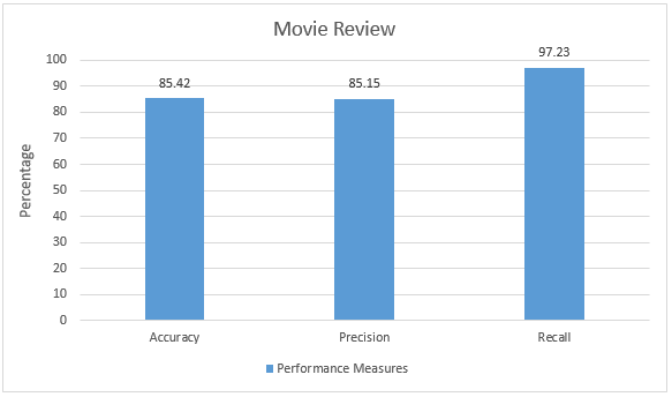

Figure 6. Performance of Proposed System

V. CONCLUSION

In this paper an algorithm is proposed for Sentiment analysis of Movie reviews in Hindi with the help of pre-annotated corpus consisting of words/phrases and negation handling of review. Also overall polarity if defined in the review then system recognizes the overall polarity of the system and assign that polarity to the review. Experimentation results indicate that the proposed algorithm is performing well in this domain and achieved the accuracy of 85.42%.

REFERENCES

[1] A. Das and B. Gambäck. Sentimantics: The Conceptual Spaces for Lexical Sentiment Polarity Representation with Contextuality, In the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA), ACL 2012, Pages 38–46, Jeju, South Korea.

[2] A. Das and S. Bandyopadhyay. Dr Sentiment Knows Everything! ACL/HLT 2011 Demo Session, Pages 50-55, June, Portland, Oregon, USA.

[3] A. Das and S. Bandyopadhyay. SentiWordNet for Indian Languages, In the 8th Workshop on Asian Language Resources (ALR), COLING 2010, Pages 56-63, August, Beijing, China.

[4] NamitaMittal,BasantAgarwal,GarvitChouhan,NitinBania,PrateekPareek,(2013), ‖Sentiment Analysis of Hindi Review based on Negation and Discourse Relation‖in proceedings of International Joint Conference on Natural Language Processing, pages 45–50,Nagoya, Japan, 14-18 .

[5] *Balamurali A R, Aditya Joshi, Pushpak Bhattacharyya*Cross-Lingual Sentiment Analysis for Indian Languages using Linked WordNets

[6] Richa Sharma, Shweta Nigam, Rekha Jain, (2014), ‖Polarity detection of Movie Reviews in Hindi Language‖ in International Journal on Computational Sciences and Applications (IJCSA) Vol.4 No.4.

[7] http://www.enchantedlearning.com/wordlist/positivewords.shtml

[8] http://www.enchantedlearning.com/wordlist/negativewords.shtml

[9] http://aajtak.intoday.in/film-review.html

[10] http://www.jagran.com/entertainment/reviews-news-hindi.html

[11] http://bollywood.bhaskar.com/reviews/movie-reviews/

[12] https://in.bookmyshow.com/movies/

[13] https://translate.google.co.in/