Sentiment Classification of Hindi Language using Natural Language Processing Techniques

Dr. Meghna Jain*

*Assistant Professor, Gyan Ganga Institute of Technology & Sciences, Jabalpur, India.

Corresponding Email: *meghnajain13@gmail.com

Abstract: This paper has presented Hybrid Approach for determination of sentimental phrase or words from Hindi text automatically through use of Hindi sentiment's lexicon and classifying them into polarity i.e. Positive, Negative and Neutral.

Keywords: Hindi, Sentiment, Technique.

1. INTRODUCTION

Sentiment Classification is computational study for expressing online through posting by writers, people and users about particular products, services and topics. There is a use of "Natural Language Processing (NLP)" technique for classification of expressions into Positive, Negative or Neutral. In current scenario, Web 2.0 application (Blogs, Social media, Forums, Wikis, Chats and Review channels) are popular medium amongst citizens for expressing opinions or sentiments publically toward any subject. A larger amount of "User Generated Content (UGC)" has been stored in digital form on web. The information might be meaningful for Individuals, Government and Organizations for making decisions correctly and making them much effective. This has brought many challenges when these expressions are automatically analyzed. Classification of sentiments is known widely in business classification, market research, social status, policy making and various decision making systems.

Background
Research in the field of sentiment classification has been done in various Indian languages. Two researchers from IIT Hyderabad i.e. Bakliwal and Arora [1] have developed a Hindi Subjective Lexicon (HSL) of all the synonyms and antonyms which were possible and used machine learning technique as well as n-Gram Modeling technique for analyzing sentiments from text. In a similar manner, Joshi, Balamuraly and Bhattacharya from IIT Bombay [3], used Hindi-SentiWordNet (H-SWN) wherein all the sentimental words have been classified into Positive and Negative classes having fixed numerical score.

Problem Statement

Model for Sentiment Classification of various Indian Languages extracts only Sentimental Words like negation words, adjective, adverb etc. from a give text piece and this is then classified into positive, neutral or negative level. There are some parameters which are used for supporting Sentiment Classification which includes part of speech, terms, syntactic negation and dependencies. There is a requirement of Computational Model which has ability for identifying, understanding and interpreting sentimental words in an automatic manner and produce better result in classification of sentiments having greater accuracy.

Experiment Setup

A. Dataset after introduction of Unicode (UTF-8)

standards for Indian various Languages, there is a rapid growth in number of blogs, discussion forums, websites, etc. in Hindi, this has resulted in an increase in public sentiments on web. After collection of data from such resources, a scheme for manual annotation is performed and this creates a Dataset of 1000 sentences for testing and trending proposed "Sentiment Classification Model" for Hindi. Such sentences mainly arise through specific domains such as political and social discourse.

B. Resource Generation

Hindi can be considered as language which has scarce resources since its resources exist in an under developed phase. Hence there is a need that the resources which are essential for sentimental classification have to be generated. To generate resources like this, a small tagset for Hindi Part of Speech (POS) has been described below :

Table No.1**.** POS Tag for Hindi as a Language

| Category	Type	Hindi words
Noun (N)	Common Noun	Example – category like cat, dog, human being
	Proper Noun	Example – name of people, place, animal or thing
Nominal Modifier (J)	Adjective (ADJ)	Which tells the important features
Verb (V)	Main Verb (VM)	**Any particular verb which indicates the act of doing**
	Auxiliary Verb (VA)	Any particular auxiliary verb
Pronoun (P)	Pronominal (PN)	Word used instead of a noun
Adverb (A)	Manner (AMN)	Which tells additionally about the verb

Working Procedure
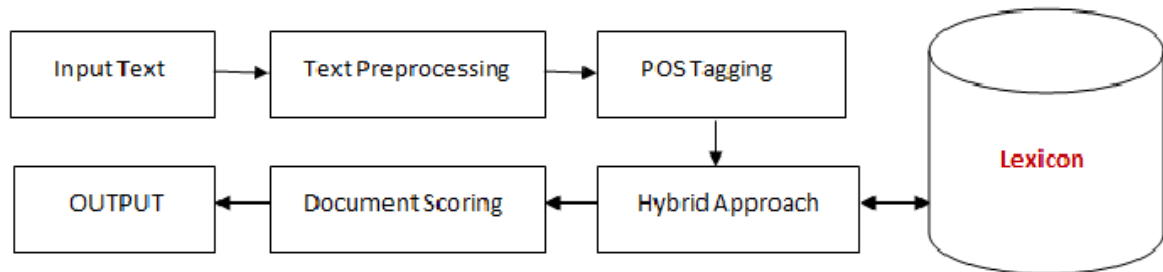
Figure 1. *Working Procedure*

Source: Malakar et al (2015)

STEP-1: The Input Phase
The first step when input is entered into the system. The entire text is broken into various sentences as proposed model for Sentiment Classification works on the sentence level.

STEP-2: Phase for Text Preprocessing
For achieving higher accurate result, certain pre-processing operations need to be applied over given sentence or text as there is some noise in text which does not returns any meaning. Since there is presence of noises, the correct results of sentiment classification could not be estimated.

STEP-3: Phase of POS Tagging
All words in given sentences can be of use in Sentiment Classification, since a few words could hold sentiments. So after activity of POS tagging, a group of words or words which are sentiment bearing are characterized by tag associated with them.

STEP-4: (Hybrid Approach Phase)
Two models are there in Hybrid Approach. The models considered under this approach are rule based model and statistical based mode. In Rule based model, there are set of rule which are used for handling negotiation since in certain conditions, polarity in a sentence can be changed to negative from positive and also at times vice versa. Negation handling can be considered as the biggest problem for performing Sentiment Classification.
In statistical based model, for achieving accuracy, there is a development of sentimental words and a database is formed having statistical score is later developed in this model. Positive value is assigned to words having positive sentiments and the value ranges between 0 - 1 and negative value is given which ranges between -1 - 0.

STEP-5: Phase for Output Representation
In this system, each sentimental word exist in sentences and the sum of them is calculated. The sum of the sentence is checked whether it is negative or positive. When sum is positive, it means it is a positive sentence, if sum is negative the sentence is negative and when the sum is 0, the sentence is said to be neutral.

2. RESULT EVALUATION

For testing "Sentiment Classification Model", a dataset having 1000 sentences was taken, out of which 50% of the sentences were tested to be positive and 50% of them came out to be negative. Such sentences have been considered as an input for Sentiment's Classification. Once the classification is carried out, the system which has been generated by us has given 70% correct results.

3. CONCLUSION

The "Sentiment Classification method" which has been proposed is based on Hybrid Approach that provides efficient results and better accuracy in their results as compared to previous researches. This study has explored a new dimension since Multi-Word Expressions have been included for improving quality in sentiment's classification for Hindi Language.

4. REFERENCES

1. Arora, P (2013). Sentiment Classification for Hindi Language (MS Thesis), IIT Hyderabad.
   Das, D. and Bandyopadhyay, S. (2010). Labeling emotion in bengali blog corpus a fine grained tagging at sentence level. In Proceedings of the Eighth Workshop on Asian Language Resouces, pg. 47–55, Beijing, China.
2. Joshi, A., Balamuraly, A. R. and Bhattacharya, Pushpak (2010). A Fall-back Strategy for
   Sentiment Classification in Hindi: A Case Study. Proceedings of ICON: 8th International
   Conference on Natural Language Processing, Macmillan Publishers, India.
3. Pang B., and Lillian L. (2008). Opinion Mining and Sentiment Classification. Foundations and Trends in Information Retrieval 2(1-2) : pg. 1–135.
4. Sharma, R. et al (2013). Opinion Mining In Hindi Language : A Survey. IJFCST, Vol.4, No.2
5. Malakar P. K. et al (2015). Sentiment Classification of Hindi Language using Natural Language Processing Techniques. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE). August 2015, PP 39-42