

# Contextual Opinion mining in online Odia Text using ‘Support Vector Machine

Manoj Kumar Jena<sup>1</sup>, Dr. Sanghamitra Mohanty<sup>2</sup>

Dept of Computer Science & Application, Utkal University, Odisha India

**Abstract:** In recent years large amount of information on web make them as viable data source for analysis which are available in different domain such as policy issues, share market, products etc. Odia language is the Indian language which is spoken in the eastern part of India especially across Odisha state. Though the plenty of data source not available as like as in English language but some of the data source is available for study. The web content of different Odia news paper are analyzed using opinion mining techniques with the use of natural language processing. The paper proposes the algorithm for analyzing opinion for classifying them as positive or negative on different policy domain especially on recent issues like Nabakalabara (New idol creation) of the lord Jagannath of odisha.

**Keywords---** Opinion Mining, Web Content Mining, Sentiment Analysis, Support Vector Machine

## 1. INTRODUCTION

Searching for opinions as conveniently as general Web search from user generated content is an immerging phenomenon for computational linguistics research. However, the Opinion mining is now at crossroads as a discipline of information extraction and text mining. Upgrading current search engines from ‘searching for facts’ to ‘retrieving opinions’ as per the stakeholders need is hard as the opinion cannot be expressed with a few keywords.

Therefore, automatic processing of documents to detect opinion expressed therein, as a unitary body of research, has been denominated opinion mining. Most work on this area has been carried out on highly subjective text types such as articles in news paper, blogs or product reviews. Authors of such type of documents mostly express their opinions quite freely. In general, an opinion is a message expressing a belief about something, the expression of a belief that is held with confidence but not substantiated by positive knowledge or proof.

Pang and Lee (2008) captured different definitions about these terms based on applications done in this field. For example, Subjectivity Analysis is defined as the recognition of opinion-oriented language in order to distinguish it from objective language. Sentiment Analysis classifies reviews according to their polarity (positive or negative). Henceforth, all these terms refer

to the same field of study. Some tasks in opinion mining try to classify the detected opinion using different scales. In a number of cases, the purpose is to identify opinions in a text and classify them into positive, negative or neutral classes. In other occasions, the goal is to assign different rates, such as ‘‘very bad’’, ‘‘bad’’, ‘‘satisfactory’’, ‘‘good’’, ‘‘very good’’, or ‘‘excellent’’. The sentiments can be ranked into a range of one to five stars. Other systems use the ‘‘thumb up’’ or ‘‘thumb down’’ notation.

In this paper we have considered one issue on which Odia news papers published in Odisha have ventilated their views regarding Nabakalabera (New idol Creation) in Odisha. Nabakalebara means creation of wooden idols for the lord Jaganatth , Balavadra, Subhadra and Sudarsan which are the most for a period one and half months.

We have collected one hundred number of news from five online Odia daily regarding Nabakalebara and tried to analyze the views in form of news collected by using our new algorithm by applying different opinion mining techniques.

## 2. PROCESS OF OPINION MINING

Opinion mining is a sub-discipline of computational linguistics and state-of-the art technology that extracts people’s opinions and sentiments from complicated text from the web. Opinions are expressed on any object or entity, i.e., product, service, topic, an individual, organization, or event commented upon. And every object has a components set, and an attributes set. So an object based on the part-of relationship can be hierarchically decomposed. An opinion’s semantic orientation on a Context ‘C’ reveals whether it is positive, negative or neutral. A model for an object and opinions set on its Context is defined as a *Context-based opinion mining model*.

A very basic step of opinion mining and sentiment analysis is feature extraction. Figure 1 shows the process of opinion mining and sentiment analysis

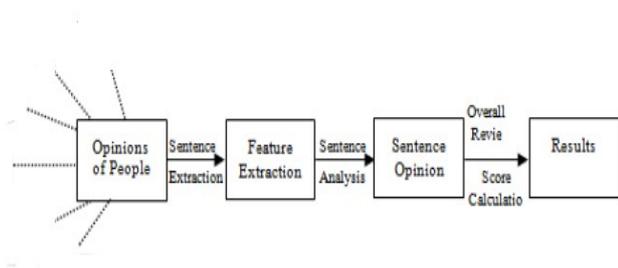


Figure 1. Process of Opinion Mining

### 3.ODIA LANGUAGE TEXT AND EXPRESSION OF OPINIONS

As receiving a classical status in Indo-European linguistic family, ‘Odia’ has a relatively free phrase order structure as compared to English. As an agglutinative language, ‘Odia’ language always demands a sophisticated morphological analyzer and/or synthesizer for any natural language generation system. Odia is a highly inflectional language is very dependent on the feature and syllable structure of the root respectively.

Opinion mining in online Odia text that deals with finding orientation of sentiment in a piece of text with respect to a topic. It mines the information from various text forms such as reviews, news, and blogs and classifies them on the basis of their polarity as positive, negative or neutral. It focuses on categorizing the text at the level of subjective and objective nature. Subjectivity of a text indicates that the text contains/bears opinion content whereas Objectivity of a text indicates that the text is without opinion content.

Subjective text: In Odia sentence, 'ନବକଳେବର ସମୟରେ ପ୍ରଭୁ ଶ୍ରୀ ଜଗନ୍ନାଥଙ୍କ ନବନୌବନ ଦର୍ଶନ ଅତ୍ୟନ୍ତ ଆନନ୍ଦଦାୟକ', (this sentence has an opinion, it talks about the nabakaleber of Lord Jagannath at Odisha and the writer's feelings about "ଆନନ୍ଦଦାୟକ" and hence it's subjective). Again the subjective text can be further categorized into 3 broad categories based on the sentiments expressed in the text.

1. Positive- ପ୍ରଭୁ ଶ୍ରୀ ଜଗନ୍ନାଥଙ୍କ ନବନୌବନ ଦର୍ଶନ ଅତ୍ୟନ୍ତ ଆନନ୍ଦଦାୟକ.
2. Negative- ପୁରୀର ଗ୍ରୀଷ୍ମ ବ୍ୟତୀତ କୃତ ଖରାପ.
3. Neutral- ଅଧିକ ଜନଗହନ ହେତୁ ମୋତେ ଅପରାଧକୁ ଅସୁବୁଧ ଲାଗିଲା. (Here user expresses his own feelings irrespective of any positive or negative polarity so it is neutral)

Objective text: ନବକଳେବର ସମୟରେ ପ୍ରଭୁ ଜଗନ୍ନାଥଙ୍କ ଦର୍ଶନ ଭକ୍ତମାନଙ୍କୁ ମିଳେ. (The sentence is a statement of fact. So it is objective).

Opinion of Odia text mainly depends upon three components, i.e. (1) Opinion holder that holds the

opinion, (2) Opinion Object on which the opinion holder is expressing the opinion and (3) Opinion orientation of an opinion on an object determines whether the opinion of an opinion holder about an object is positive, negative or neutral. In the above mentioned example, the blogger is the Opinion Holder, whereas the word '□□□□□' is the opinion object and the word '□□□□□□□□□□' is positively oriented opinion of Opinion holder.

### 4.OPINIONS MINING SYSTEM AND ITS STAKEHOLDERS

There are many techniques through which opinions can be analyzed from news paper articles. Currently, it has become a practice for websites, to facilitate the expression of opinions by guests and visitors on particular issues, products marketed or on a particular topic. Also, the expansion of web and social networking, are facilitating users posting opinions online. Thus, the content of reviews has increased rapidly, making the big e-commerce sites, or recommendations of products and services sites, to contain hundreds to tens of thousands of reviews per item. The large number of reviews promotes access to useful and relevant information to visitors. They can be used, for example to compare offers from different competitors on the market and make an informed decision about buying a certain offer. It is very difficult for a visitor to read all of them and to form an opinion on the subject or product because: - in some cases these reviews can be very long and only a few sentences may express opinions or may not contain opinions at all. Navigating only part of the may create a false impression about the topic; the user is not familiar with the various metrics used in comparing offers in a certain specialized field.

Also, the large number of reviews makes it difficult for producers to follow reactions of potential customers. They face additional difficulties in pursuing wide range of products, traded on a variety of web sites [7]. So, it is useful to make a system to detect indicators of performance of a product, and domain specific metrics, to summarize the opinions obtained from the large amount of reviews, in several positive and negative aspects.

### 5. PROPOSED APPROACHES TO BE FOLLOWED

To do the opinion mining to Odia data, we have considered the supervised learning method. For the said

purpose we adopted the Support Vector Machine (SVM) as the machine learning classifier and are trying to divide the collected data into two parts, i.e. training data and test set. Where our training set trained the classifier, a test set test the classifier by giving the test data.

In this set of experiments, we match all the words in a sentence with the adjectives and adverbs lexicon. As the performance of POS tagger for Odia language isn't that good, so to minimize the error caused by POS tagger we are matching all the words in a sentence/review. On the reviews data, we performed unigram presence and simple scoring method classification. In unigram presence method, we count unigrams of positive, negative and objective polarity in a sentence/review and assigned the polarity for which the count is highest. In a simple scoring method, we summed the positive, negative and objective scores of each word in a sentence/review and assigned the polarity of the dominant score.

Algorithm:

1. Input the indexed news {a1,a2...an}
2. Extract the sentences from ai (i 1 to n) as {s1,s2,...sn}
3. Apply the classifier Support vector machine
4. Classify the sentence by tagging a score
5. Calculate the opinion from all the sentences by averaging the score.
6. If the score is greater than 0.5 then +ve or else -ve.
7. Calculate the term frequency -inverse Document.

Before classification of data, the system go through the transformation process, where the weight of each word in the corpus is to be calculated using TF-IDF so that it is easy to determine the words in the corpus of documents might be more favorable to use in further processing. As per the transformation process calculation –

$$wd = fw,d * \log(|D|/fw,D)$$

Where, D is collection of documents; w represents words; d is individual document belongs to D; |D| is size of corpus; fw,d is number of times w appears in d; fw,D is number of documents in which 'w' occurs in D.

**Support Vector Machine:** Support Vector Machine model is associated with a learning algorithm that

analyzes the data and identifies the pattern for classification. Support Vector Machine Classifier based upon 'decision plane' constructs N-dimensional hyper plane represented by vector which separates data into two categories. SVM takes the input data and for each input it predicts the class. SVM can be seen as a constrained optimization problem, in which class  $C_j \{1, -1\}$  corresponds to either positive or negative class that belongs to document  $D_j$ , the solution can be written as in equation (1)

$$\vec{\omega} = \sum_j \alpha_j c_j \vec{d}_j, \alpha_j \geq 0$$

Where  $\vec{\omega}$  is a vector,  $c_j$  is a class and  $d_j$  is a document [1].

SVM module is very good performance on experimental results and it is low dependency on data set dimensionality. However in case of categorical or missing value, it needs pre-processed.

## 6. CHALLENGES OF OPINION MINING IN ODISIA LANGUAGE

One major concern in the analysis of user-generated content in online applications is, to determine the polarity of opinions, by extracting the subject whom opinions are addressed and the arguments are based on. Analyzing existing techniques in opinion mining, we can summarize the following challenges while dealing/working with Odia language:

1. Phrase Order- Phrase/word arrangement in a sentence plays an important role in identifying the subjective nature of the text. Odia is a free phrase order language i.e. the subject, object and verb can come in any order whereas English is a fixed order language i.e. subject followed by a verb and followed by an object. Phrase order plays a vital role in deciding the polarity of a text, in the text same set of words with slight variations and changes in the word order affect the polarity aspect.
2. Morphological Variations- Handling the morphological variations is also a big challenge for Odia language. As Odia language is morphologically rich and agglutinating in nature, there are a lot of information is fused in the words as compared to the English language. In same 'sandhi and samasa' are also plays a

great role for information storing in Odia language.

3. Lack of online resources- the lack of sufficient online resources, tools and annotated corpora also adds to the challenges while addressing the problem of opinion mining especially when we are dealing with Odia like Indian languages.
4. Opinion mining process is centered on a domain, so, a solution determined for a given area (i.e. movie reviews analysis), will not work on another (e.g. social issues). The way of expressing feelings varies from one domain to another, the developed model requiring adaptation.

## 7.RESULTS

We have experimented with Resource-based methods of opinion mining with 100 news sentences as from our availed test case. However, looking to the input text our manual linguistic observation is supplying a higher percent of opinions in comparison to Resource based methods.

Table 1: Accuracy of Different Methods analysis of News

Resource-based Observation	60.5
Manual Observation	78.14

## 8.CONCLUSION

Most opinion mining systems as a natural language generation system pipeline pragmatic, semantic, lexical and syntactic decisions (Reiter, 1994). With the right formalism, constructing pragmatics, semantics and syntax simultaneously for Odia language Opinion mining and concluding the decision very fast and nearer to user is easier and better. The support vector machine approach elegantly captures the interaction between pragmatic and syntactic constraints on descriptions in a sentence, and the inferential interactions between multiple descriptions in a sentence. At the same time, it exploits linguistically motivated, declarative specifications of syntactic constructions to make contextually appropriate sentiment choices. Every method has some benefits and limitations and one can use these methods according to the situation for feature and text extraction. Based on the survey we can find the accuracy of different methods. We have tested using only limited number of news on a particular topic. The system is to be tested on more number of opinions in Odia language.

### References

1] Ion SMEUREANU, Cristian BUCUR, Applying Supervised Opinion Mining techniques on Online User Reviews, Informatica Economică vol. 16, no. 2/2012.

[2] Bo Pang and Lillian Lee, "Opinion Mining and Sentiment Analysis", foundations and TrendsR\_ in Information Retrieval Vol. 2, Nos. 1–2 (2008).

[3] Abbasi, "Affect intensity analysis of dark web forums," in Proceedings of Intelligence and Security Informatics (ISI), pp. 282–288, 2007.

[4] K. Dave, S. Lawrence & D. Pennock. \Mining the Peanut Gallery: Opinion Extraction and Semantic Classi\_cation of Product Reviews." Proceedings of the 12th International Conference on World WideWeb, pp. 519-528, 2003.

[5] B. Liu. \Web Data Mining: Exploring hyperlinks, contents, and usage data," Opinion Mining. Springer, 2007.

[6] B. Pang & L. Lee, \Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales." Proceedings of the Association for Computational Linguistics (ACL), pp. 15124, 2005.

[7] L. L. B. Pang, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no 1-2, pp. 1-135, 2008.

[8] A. Moran. (2012, Mar.) Sentiment Analysis: How does sentiment analysis work? [Online]. <http://www.quora.com/Sentiment-Analysis/How-does-sentiment-analysis-work#ans606524>

[9] M Caraciolo. (2012, Mar.) Working on sentiment analysis on Twitter with Portuguese language. [Online]. <http://aimotion.blogspot.com/2010/07/working-on-sentiment-analysis-on.html>

[10] L. L. B. Pang, "*Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales.*," in Proceedings of ACL, 2005, pp. 115–124. (2012, Mar.) Wikipedia. [Online]. <http://en.wikipedia.org>

[11] P. Raghavan, H. S. Christopher and D. Manning, *Introduction to Information Retrieval*. USA, New York: Cambridge University Press, 2008.

[12] H. Cui, V. Mittal, and M. Datar, (2006)"Comparative experiments on sentiment classification for online product reviews," presented at the proceedings of the 21st national conference on Artificial intelligence - Volume 2, Boston, Massachusetts.