Sentiment Analysis of Bangla Song Review- A Lexicon Based Backtracking Approach

5 authors, including:

Tapasy Rabeya
Daffodil International University
9 PUBLICATIONS **72** CITATIONS

SEE PROFILE

Narayan Ranjan Chakraborty
Universitetet i Agder
39 PUBLICATIONS **270** CITATIONS

SEE PROFILE

Manoranjan Dash
Siksha O Anusandhan (Deemed to be University)
94 PUBLICATIONS **387** CITATIONS

SEE PROFILE

Ahmed Al Marouf
The University of Calgary
48 PUBLICATIONS **339** CITATIONS

SEE PROFILE

Sentiment Analysis of Bangla Song Review- A Lexicon Based Backtracking Approach

Tapasy Rabeya
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
tapasyrabeya.cse@diu.edu.bd

Narayan Ranjan Chakraborty
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
narayan@daffodilvarsity.edu.bd

Sanjida Ferdous
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
bintysanjida@gmail.com

Manoranjan Dash
Faculty of Management Sciences
Siksha O Anusandhan (Deemed to be University)
Bhubaneswar, India
manoranjandash@soa.ac.in

Ahmed Al Marouf
Department of Computer Science and Engineering
Daffodil International University
Dhaka, Bangladesh
marouf.cse@diu.edu.bd

Abstract: In the last decade, Sentiment analysis is a prospering experimental topic of research as because of a lot of opinionated data accessible on Blogs & social networking sites. It is the reference to the assignment of Natural Language Processing to decide whether text or content contains any subjective information like positive, negative or not. These social media and other online platforms are giving an immense stage to uncover human's gifts in a fast speed, and average citizens can likewise put their feeling through the remarks which emphatically show how they are accepting another potential. In this paper, we have presented a sentimental analysis of Bengali song reviews from a specific YouTube channel to analyse people acceptance rate of a new young star. For detecting the sentiments, we have used a backtracking algorithm, where the heart of this approach is a sentiment lexicon. And the research showed the backtracking algorithm performed more than 70% accuracy to detect actual public sentiment.

Index Terms: Opinion mining, sentiment analysis, backtracking approach, acceptance rate, sentiment lexicon.

I. Introduction

For amazing evaluation of latest technology, dynamic features of the internet, the availability of various platform people are becoming more habitual towards virtual world. It became an essential communication hub, and also people spent most of the leisure time here. Nowadays it is quite easy to gain popularity through social networking sites. People are posting their recorded songs, tutorials, motivational speeches etc. on YouTube and getting popularities as well by providing a massive amount of opinions on the comment section. Multiple comments make it tough to conclude to a specific aspect of the video as opinion differs from person to person. That produces a massive volume of data which can be used to generate a decision from the comments.

Sentiment analysis is mechanised mining of attitudes, emotions from the text, speech and

different sources though natural language. This paper shows the work on sentiment analysis of Bengali reviews from a specific YouTube channel by applying a lexicon based backtracking algorithm.

The rest of the paper is structured as follows. The second section presented a literature review, which shows that maximum works of sentimental analysis deal with people's review on various issues. In section three, we discussed our data sources. The methodology of our sentimental analysis and data analysis is discussed in section four. Finally conclution is delivered in section five.

II. Literature Review

In the present era, the researchers are showing their great interest and effort in the field of sentiment analysis, as it becomes one of the hottest research areas in computer science.

Nizar A. Ahmed has introduced a novel framework for detecting sentiment from Arabic tweets. The speciality of their work was a sentiment lexicon. SentiStrength English sentiment lexicon has an excellent contribution to their work as they translate the English lexicon into Arabic and after that expanded it using Arabic exchequer. The assessment process of their work has been going through by testing their introduced framework performance by 4400 Arabic tweets [2].

A comparative analysis was done on the sentiment of foreign tourists of Bangkok for their visiting purpose. They classified the reasons into five categories. Moreover, the result showed 71.93% travelling purpose followed by business and family tour. In order to see the possible result, they have used four approaches like- Decision Tree, Naïve Bayes, SVM and ANN. Among the four approaches, ANN has shown the highest accuracy that was 80.33% [3]. Shweta Rana has performed another comparative exploration on movie review by using Naive Bayes and Linear SVM. The result indicates the best accuracy from Linear SVM which is followed by the Synthetic words approach [8]. Zamahsyari wanted to present the performance of different classifier working on economic news from Indonesian

online media so that it could be beneficial for the government making an economic policy [11].

To represent text as a vector of independent words Bag-of-words model is used, but the Polarity shift is the major problem in the Bag-of-words model. It is a kind of sentiment classification problem. Polarity Shift Detection Approaches help to improve the performance of machine learning classification algorithms [1].

Prerna Mishra analysed twitter data of Narendra Modi's (Indian Prime Minister) Digital India Campaign having 50% positive opinion by using a dictionary-based approach [5]. Another work has been done on movie review from twitter data. To find the polarity of the reviews they have introduced a Senti-lexicon algorithm. Analysing with 300 tweet review, they finally came up with 70% accuracy [6].

A review has been executed on Techniques Used for Sentiment Analysis on Twitter Data. They said the Machine Learning Method and Lexicon Based Approach are dominating the Sentiment Analysis and Opinion mining fields. Keep updating the training datasets can be a source of accurate results [7].

An aspect-based approach has been introduced while analysing Malayalam movie and product reviews. The author said that regular sentimental analysis is not enough, an aspect of the user is commenting is significant regarding analysing. This sentence level aspect-based sentiment analysis showed 84.7% accuracy [9].

A lexicon based backtracking approach has been introduced in order to detect emotion (happiness and sadness) from Bengali text. In this model, we have dealt with detecting emotion from different platforms text. Our focus was to see how people generally express their emotions while talking, wring etc. In order to find the solution after generating expression from a given text, we have decided to track the expression from right to left and found 77.16 per cent accuracy with an assumption that people generally express their emotion in the last part of a sentence [12].

Table 1: Sentimental Analysis using different approaches

Author	Reference	Methodology	Type of methodology	Data used
Duwairi et al., 2015	[2]	Lexicon based	Introduced new lexicon based framework for the Arabic language	Arabic tweet
Kuhamanee et al., 2017	[3]	Decision Tree, Naïve Bayes, SVM and ANN	Existing approach	The opinion of foreign tourists of Bangkok
Liu et al., 2016	[4]	the framework of opinion shift detection	New framework	online reviews from Yelp
Mishra et al., 2016	[5]	Dictionary Based approach	Existing Approach	twitter data on Modi ji's Digital India Campaign
Mumtaz and Ahuja, 2016	[6]	Senti-lexical algorithm	New algorithm	Movie review from twitter
Rana and Singh, 2016	[8]	SVM and Naïve Bayes Techniques	Existing Approach	Movie review
Thulasi and Usha, 2016	[9]	aspect based approach	New algorithm	Movie and Product Reviews in Malayalam
Wankhede and Thakare, 2017	[10]	Random Forest classification	Existing Approach	movie review
Nurwidyantoro, 2016	[11]	such as decision tree, random forests, and support vector machine, majority vote classifier	Existing Approach	economic news from Indonesian online media
Rabeya et al., 2017	[12]	Lexicon based backtracking approach	New algorithm	Sentences from facebook status, news, textbook, normal speech.

III. Data source:

The last execution of work is corresponding to the test data quality. For opinion mining Different websites, blogs, micro-blogs, and social networking sites offering a substantial amount of data. These online platforms are playing an important rule for creating and exposing new potentials. By analysing the comments posted on specific performance, the social acceptance rate of new talents can be presented. In our work, we have taken Bengali comments from a YouTube channel owned by Bengali new young star Mahtim Sakib.

IV. Methodology:

In this work for opinion mining of users comment we have used a backtracking algorithm. This algorithm was proposed for detecting emotion from Bengali text by sentimental analysis. Here we have used this algorithm with a bit modification.

This methodology can be presented in five steps. After generating expression from the given input sentence, the lexicon based backtracking

algorithm will be applied to detect the sentiment of each sentence.

A) Input sentence: As we are intended to analyse Bengali sentences, we have taken 201 Bengali comments from a specific channel of YouTube comment section. Both positive and negative comments were included there. We also faced some positive comments where no positive keywords were present.

Figure 1: sample input sentence

B) Corpus and Generating expression: For generating expressions we have used three kinds of words: positive, negative and negation. In our previous work for emotion detection, we had four kinds of words. However, regarding analysing song review we usually never face words call socially castoff. Example; খুন,জুয়া,ধর্ষণ... That's why we have omitted these words.

Figure 2: Lexicons used in two analysis

We have added English positive words written in the Bengali language and negative English words written in the Bengali language in the word corpus for making our analysis more effective. As nowadays the people become more used to writing the mixed language. See the following comment.

ভাইয়া আপনার কণ্ঠস্বর অনেক সুইট।
Here the word "সুইট" reffers as "মিষ্টি" that has a positive meaning.

Table1: Positive and Negative words

হ্যাঁ বোধক		না বোধক	
ভাল	গুড	খারাপ	বেড
সুন্দর	নাইস	বাজে	বোরিং
মিষ্টি	সুইট	অসহ্য	ডিজগাস্টিং
অসাধারন	অওসাম	ঘৃণা	হেইট
অতুলনীয়	লাভ	অবজ্ঞা	ইগনর

Additionally, we have used a list of word phrase as because of facing such sentences with no emotional keywords, but still having a strong positive or negative aspect as following comments..

দোয়া করি তুমি অনেক বড় হউ।
গান থামিও না, এগিয়ে যাও।
Though these two sentences do not contain any strong positive word which will help to find the sentiment of those sentences. However, "অনেক বড় হউ" and "এগিয়ে যাও" are indicating strong positive meaning. Thats why we have stored word phrases those are used very frequently.

Table2: Word Phrase

শব্দ গুচ্ছ	অর্থ
মন তোমাকে দিলাম	হ্যাঁ বোধক
অনেক বড় হউ	হ্যাঁ বোধক
অনেক বড় হবে	হ্যাঁ বোধক
চালিয়ে যাও	হ্যাঁ বোধক
হাফ লেডিস	না বোধক

মনের মত	হ্যাঁ বোধক
সাড়া জাগাবে	হ্যাঁ বোধক
সাড়া জাগাচ্ছে	হ্যাঁ বোধক
এগিয়ে যাও	হ্যাঁ বোধক

Ones our corpus introduced with all the hash values of different words, the next step is to generate expression from the input sentence, we have denoted three numbers to three different types of words.

Table 3: symbols for representing different lexicons

হ্যাঁ বোধক	1
না বোধক	2
না বাচক	3

The generated expression will be like this

দারুণ, তুমি খুব সুইট ভাই, গলাটাও মিষ্টি।

1　　　　1　　　　1

111= দারুণ সুইট মিষ্টি =হ্যাঁ বোধক

C) Backtracking Algorithm:

After generating expression from input sentences, the backtracking algorithm has been applied to the generated expression in order to find the sentiment.

We have trained our algorithm with 288 unique expressions with their genuine sentiment.

Table 4: Stored expressions for analysis

2113	Negative
223	Negative
123	Negative
2213	Negative
21	Positive
131	Positive
1313	Positive
231	Positive

The backtracking formulas for detecting sentiment….

$$P = \sum_{i=1}^{n} \frac{var}{total\_P} base^{\Sigma position} \quad …………(i)$$

$$N = \sum_{i=1}^{n} \frac{var}{total\_N} base^{\Sigma position} \quad ………….(ii)$$

D) Experimental Result:

This experimental work is done on Bengali review text from a Bangladeshi young star's YouTube channel. We have carried out three tests by using twenty different unique test data set and found the satisfactory outcome, shown in figure 2.
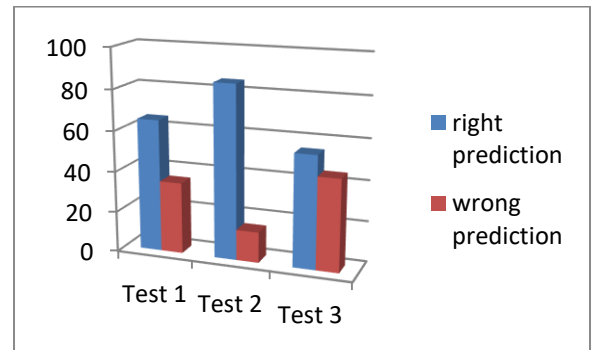

Figure 3: Bar chart for the test's outcome.

We have calculated the F-score regarding getting the accuracy of the tests by using precision and recall.

Table 5: precision, recall and F-score

Test no	Precision	Recall	F-score
Test 1	0.667	0.833	0.740
Test 2	0.813	0.923	0.865
Test 3	0.600	0.545	0.571

V. Conclusion:

In this experimental research, we have done an opinion mining on song review data from a specific YouTube channel. For analysis, we have gathered 288 comments from the comment prompt. We did not apply any preprocessing to gathered comments. We convert the input sentences on hash values and then generate expressions from them. Then the backtracking algorithm has been applied to the generated expression, and the algorithm shows 71.23% accuracy. The backtracking algorithm has performed well in order to detect sentiment from the comments. However, failed to detect exact sentiment from such sentences…

ভাইয়া তোমার গান শুনে তো পুরাই পাগল হয়ে গেলাম।

Here "পুরাই পাগল হয়ে গেলাম" indicating a positive meaning. But in some instance, it is also referred to as a negative one. Understanding to deal with such sentiments is still a challenge.

VI. References:

1. Zirpe, Sayali, and Bela Joglekar. "Polarity shift detection approaches in sentiment analysis: A survey." In *Inventive Systems and Control (ICISC), 2017 International Conference on*, pp. 1-5. IEEE, 2017.

2. Duwairi, R. M., Nizar A. Ahmed, and Saleh Y. Al-Rifai. "Detecting sentiment embedded in Arabic social media–a lexicon-based approach." *Journal of Intelligent & Fuzzy Systems* 29, no. 1 (2015): 107-117.

3. Kuhamanee, Taweesak, Nattaphon Talmongkol, Krit Chaisuriyakul, Wimol San-Um, Noppadon Pongpisuttinun, and Surapong Pongyupinpanich. "Sentiment analysis of foreign tourists to Bangkok using data mining through the online social network." In *Industrial Informatics (INDIN), 2017 IEEE 15th International Conference on*, pp. 1068-1073. IEEE, 2017.

4. Liu, Zihang, Yu-Ru Lin, Maoquan Wang, and Zhao Lu. "Discovering Opinion Changes in Online Reviews via Learning Fine-Grained Sentiments." In *Collaboration and Internet Computing (CIC), 2016 IEEE 2nd International Conference on*, pp. 1-10. IEEE, 2016.

5. Mishra, Prerna, Ranjana Rajnish, and Pankaj Kumar. "Sentiment analysis of Twitter data: Case study on digital India." In *Information Technology (InCITe)-The Next Generation IT Summit on the Theme-Internet of Things: Connect your Worlds, International Conference on*, pp. 148-153. IEEE, 2016.

6. Mumtaz, Deebha, and Bindiya Ahuja. "Sentiment analysis of movie review data using Senti-lexicon algorithm." In *Applied and Theoretical Computing and Communication Technology (iCATccT), 2016 2nd International Conference on*, pp. 592-597. IEEE, 2016.

7. Jain, Alind, Shreem Kapoor, and Anshu Malhotra. "A Review of Techniques Used for Sentiment Analysis on Twitter Data."

8. Rana, Shweta, and Archana Singh. "Comparative analysis of sentiment orientation using SVM and Naive Bayes techniques." In *Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on*, pp. 106-111. IEEE, 2016.

9. Thulasi, P. K., and K. Usha. "Aspect polarity recognition of movie and product reviews in Malayalam." In *Next Generation Intelligent Systems (ICNGIS), International Conference on*, pp. 1-5. IEEE, 2016.

10. Wankhede, Rasika, and A. N. Thakare. "Design approach for accuracy in movies reviews using sentiment analysis." In Electronics, Communication and Aerospace Technology

(ICECA), 2017 International conference of, vol. 1, pp. 6-11. IEEE, 2017.

11. Nurwidyantoro, Arif. "Sentiment analysis of economic news in Bahasa Indonesia using majority vote classifier." In *Data and Software Engineering (ICoDSE), 2016 International Conference on*, pp. 1-6. IEEE, 2016.

12. Rabeya, Tapasy, Sanjida Ferdous, Himel Suhita Ali, and Narayan Ranjan Chakraborty. "A survey on emotion detection: A lexicon based backtracking approach for detecting emotion from Bengali text." In *Computer and Information Technology (ICCIT), 2017 20th International Conference of*, pp. 1-7. IEEE, 2017.