# Sentiment Detection from Bangla Text using Contextual Valency Analysis

K. M. Azharul Hasan
Department of Computer Science and Engineering
Khulna University of Engineering and Technology (KUET),
Bangladesh
az@cse.kuet.ac.bd

Mosiur Rahman, Badiuzzaman
Department of Computer Science and Engineering
Khulna University of Engineering and Technology (KUET),
Bangladesh
palashcse2k8@gmail.com, s.sayem@ymail.com

***Abstract*— Sentiment Analysis or opinion mining is an area of important research over the last decade. The basic task in sentiment analysis is classifying the polarity of a given text whether the expressed opinion in the text is positive, negative, or neutral. This paper presents an approach to sentiment assessment from Bangla text using contextual valence analysis. In linguistics valence of a verb is the number of satellite noun phrases with which a verb combines. We have used the WorldNet to get the senses of each word according to its parts of speech and SentiWordNet to get the prior valence (i.e. polarity) of each word. We calculate the total positivity, negativity and neutrality of sentence or document with respect to total sense. We developed our own methodology to calculate the sentiment from Bangla text using valency analysis. Sufficient examples and experiments are presented to describe the methodology.**

***Keywords*— Sentiment detection, opinion mining, Valency analysis, Bangla verb, SentiWordNet, Machine learning,**

## I. INTRODUCTION

Sentiment analysis or opinion mining has been quite popular and has led to building of better products, understanding user's opinion, executing and managing of business decisions[1]. It is an important application of natural processing, computational linguistics, and text analytics to identify and extract subjective information in source materials[2]. Bangla, one of the important Indo-Iranian languages, is the sixth-most popular in the world and spoken by a population that now exceeds 250 million. It is the primary language in Bangladesh and second language in India[3]-[5]. The increasing user-generated content on the blogs and news corpus for Bangla forces the research for sentiment analysis or opinion mining to get the user opinion of the products or services. The advances in sentiment analysis have opened up an important channel using which businesses can gauge the online trend about their products and services. The linguistic meaning of valence derives from the definition of valency in chemistry. It refers to combinatory potential of word and phrase[6]. That is whether a word or phrase can or cannot just combine to any other sign in any random way. In linguistics, verb valency or valence refers to the number of arguments controlled by a verbal predicate. i.e. a single word can be a noun, verb, adjective or adverb in a sentence according to their position and meaning. In this paper, based on the valency of a word, we detect the sentiment of a sentence or paragraph from Bangla text. We obtain prior valance of a word using SentiWordNet[7][8]. For each word in SentiWordNet there are three categories of values corresponding to its sense namely positive, objective and negative. Senses are different according to their parts of speech in corresponding sentence. The senses may have nonzero scores for all the three categories, which indicates the corresponding word has each of the three opinion related properties to a certain degree We mainly focus on prior valance of each of the word using three categories of opinion determined by valency analysis of SentiWordNet. And then we calculate the positivity, negativity and neutrality of the sentence. SentiWordNet is the result of the automatic annotation of all the synsets of WordNet[9] which is a database of English words containing 150,000 words organized in over 115,000 synsets for a total of 207,000 word-sense pairs. In this paper, we propose a frame work for sentiment detection based on valency analysis using SentiWordNet. That is to analyze whether the text expresses a favorable or unfavorable sentiment for a specific subject.

## II. RELATED WORKS

So far the authors know there is no work for sentiment detection or opinion mining from Bangla text using valency analysis. Although there are some works [10][11] on valency analysis of Bangla verbs. A methodology proposed in [10] for analysis of the requirements of Bangla verbs for its arguments and established the relationship of verb expectations for its arguments and compatibility of nouns for a specific a verb whereas [11] describes the ontology of type of Bangla Verbs by drawing its factual existence for Bangla discourse. Authors in [12] used valency-based verb for locating verb and parsing of Bangla language. However, there are some works addressed sentiment detection from Bangla texts are mentioned here. [2] proposed an automatic sentiment detection technique using machine learning formula for Bangla text. They also discover some important challenges for Bangla language processing. [13] proposed an emotion tracking system based on topic or event by employing sense based affect scoring techniques for

annotated news stories and blog corpora in Bangla using SentiWordNet[7] and WordNet Affect[8]. [14] recognizes the emotion of Bangla language using a pre-processing technique to retrieve and store the bloggers' comments on specific topics. [15] detect the opinion from Bangla text based on news corpus using support vector machine. A semi-supervised approach to sentiment classification is proposed in [16] to detect the ambiguous and unambiguous review via a combination of active learning, transductive learning, and ensemble learning. Hence they introduces a sentiment based classification or clustering of documents [17]. It incorporates the users feedback into the clustering algorithm that helps to detect the dimension on which the clustering to be done using sentiment detection dataset. In this paper we have introduced the valency based sentiment detection for Bangla language. Valency had been used for sentiment detection has been applied in literature such as [6][18] and we have introduced this for Bangla.

## III. Sentiment Detection using Valency Analysis for Bangla Text

In this section, we present the valancy determination of a Bangla word and then determine the sentiment of Bangla text. The methodology is based on the prior valence for each word for positivity, negativity or neutrality of a sentence or document with respect to total sense. We use SentiWordNet to set valence for each word and calculate the total sentiment of a sentence and hence the sentiment of a document. We are interested in identifying positive, negative or neutral sentiment from a Bangla text hence scope of the paper is limited to sentiment assessment only.

We split the sentence into words on the basis of space (" ") and coma (,). We have created a XML file to store the Bangla word and its corresponding POS. The tasks behind the XML are always the same, reading data from XML and writing data into it. The basic format of XML tag is <tag_name>value</tag_name>. We design a XML file for storing the Bangla words Bangla word as value and tag_name as its corresponding parts-of-speech (POS).

For example:

| | | |
|---|---|---|
| সে (se) | ভাল (valo) | ছেলে (sele) |
| Pronoun | Adjective | Noun |
| সে (se) | খারাপ (kharap) | না (na) |
| Pronoun | Adjective | adverb |
| পাখি (Pakhi) | আকাশে (Akase) | উড়ে (Ure) |
| Noun | Noun | Verb |

Determining POS follow some rule or pattern of sentence that indicates the option for a specific word. The word "ভাল" can be used as noun, adjective or adverb. In English the word "good" can be used only one constructional form namely "good". But in Bangla we can have the advantage in constructional form. For example,

"He is a good man"- "good" is used as adjective in this sentence or "We beat him good" -"good" is used as adverb. On the other hand for a Bangla sentence the word has a constructional difference that makes the Bangla word different for adjective or adverb. For example "সে ভাল ছেলে" (Se valo Sele) - "ভাল (valo)" is used as adjective and 'ভালমত পড়াশুনা কর (valomoto porasuna koro)"-"ভালমত" is used as adverb. So the word "ভাল" changes constructional form like "ভালই", "ভালভাবে", "ভালমত" when used as adverb. This gives the advantages that the POS of the word can easily be detected for each Bangla sentence. We translate the Bangla word into English word then using the sentiwordNet 3.0 we find the different sense and their polarity for different POS . We map the different forms of a sentimental word with single English word to get the polarity of the word. For example different forms of the word "ভাল" (valo) such as "ভালই" "ভালভাবে" "ভালমত" "ভালর" are translated into a single word "good" and the words "খারপ" "খারাপের" "খারাপভাবে" "খারাপমত" translated into English word "bad". And then use different sense and their polarity for different PoS from WordNet 3.0.

TABLE 1: WORDNET 3.0 OUTPUT FOR DIFFERENT WORDS

| Words | No. of sense as Noun | No. of sense as Adjective | No. of sense as Adverb | No. of sense as Verb |
|---|---|---|---|---|
| Good | 4 | 21 | 2 | 0 |
| Bad | 1 | 14 | 2 | 0 |
| Do | 3 | 0 | 0 | 13 |

A word can appear in different PoS form in the sentence. For each of the PoS of a word WordNet 3.0 generate different sense. Table 1 shows the some output of WordNet for different words. WordNet 3.0 give these senses from the contextual valence of each word. This is how we use the term contextual valence analysis. We will assign a prior valence(polarity) for each word from SentiWordNet 3.0.SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity. For our method we use English SentiWordNet. From SentiWordNet for a word we will get three values as positive, negative and objective(or neutral) from each sense of specific PoS. We calculate the total positivity, negativity and objectivity of a word for all of the sense corresponding to the PoS. And then use this value for assigning polarity to the word.

The word "good" has three PoS characteristics as noun, adjective and adverb. So the polarity of this sentence will be three types in three categories of PoS. Table 2 shows the polarity of the word "good" for different POS. We calculate sentence level sentiment first then merge the sentiment of each sentence to find the document level sentiment. From the prior valence(polarity) of each word we determine the sentence level sentiment as follows:

We find total sense (tw), positive sense (pw), negative sense (nw) and neutral sense (ow) for each word of sentence. We translate the Bangla sentence into its equivalent English

sentence then the polarity from SentiWordNet 3.0. Then we find the percentage of positivity, negativity and neutrality based on the valance (polarity).

TABLE 2: SENTIWORDNET 3.0 OUTPUT FOR WORD "GOOD" ACCORDING TO PARTS OF SPEECH

| Word | PoS | Total Sense | Total Positivity | Total Objectivity | Total Negativity |
|---|---|---|---|---|---|
| Good | Noun | 4 | 2 | 2 | 0 |
| Good | Adjective | 21 | 13 | 7.875 | 0.125 |
| Good | Adverb | 2 | 0.375 | 1.625 | 0 |

We follow the following rules for estimating sentiment in sentence level (if the sentence has m words):

Total sense(ts)=Sum of the senses of all the words in the sentence, $ts = \sum_{i=1}^{m} tw_i \ (1 < i < m)$

Total positivity(ps)=Sum of the positivity of all the words in sentence, $ps = \sum_{i=1}^{m} pw_i \ (1 < i < m)$

Total neutrality(os)=Sum of the neutrality of all the words in sentence, $os = \sum_{i=1}^{m} ow_i \ (1 < i < m)$

Total negativity(ns)=Sum of the negativity of all the words in sentence, $ns = \sum_{i=1}^{m} nw_i \ (1 < i < m)$

Hence the percentages of senses are found

Percentage of positivity $pcs = \frac{ps}{ts} \times 100\%$

Percentage of negativity $ncs = \frac{ns}{ts} \times 100\%$

Percentage of neutrality $ocs = \frac{os}{ts} \times 100\%$

Hence sentiment of sentence is $\max(pcs, ncs, ocs)$.

Example 1: Consider the Bangla sentence সে (Se) ভাল (valo) ছেলে (sele)

সে (He) + ভাল (Good) + ছেলে (Boy)

PoS : Noun + Adjective + Noun

Values from SentiWordNet 3.0.: সে (He) {tw=2, pw=0, ow=2, nw=0}, ভাল (good){tw=21, pw=13, ow=7.875, nw=0.125} and ছেলে (boy) {tw=4, pw= 0.25, ow=3.625, nw=0.125}

Total sense(ts) = (2+21+4) = 27

Percentage of positivity(pcs) = $((0+13+.25)/27) \times 100\% = 49\%$

Total neutrality(ocs)=$((2+7.875+3.625)/27) \times 100\% = 50\%$

Total negativity(ncs)= $((0+.125+.125)/27) \times 100\% = 1\%$

## IV. EXPERIMENTAL RESULTS

To determine the final sentiment of a sentence or document using valency we apply expert analysis. For a single input sentences/paragraph of Bangla sentences, we take the opinion of 20-30 peoples about the sentiment of the paragraph. Then we get the average of the opinions. We also implemented a prototype system using our methodology. Table 3 shows the summery of the experimental analysis and expert opinion. Table 3 shows the pcs, ncs and ocs as described in Section 3.0. From table we see that the sentiment decision is different only for paragraph 4. This is because paragraph has some exceptional sentences that need to handle specially. For example "সে ভাল না।". if there is a negative word then it alters the sentiment. and also "এটা বলা যাবে না যে, সে খারাপ ।এটা বলা যাবে না যে, সে ভাল না।" The double negative diminishes each other makes the sentences positive, i.e. This was not considred in the prototype system.

TABLE 3: EXPERT ANALYSIS FOR VARIOUS SENTENCE.

| | Positive (pcs) | Negative (ncs) | Neutral (ocs) | Sentiment decision by prototype system | Expert opinion (%) |
|---|---|---|---|---|---|
| Paragraph -1 | 50% | 10% | 40% | Positive | Positive (90%) |
| Paragraph -2 | 20% | 50% | 30% | Negative | Negative (80%) |
| Paragraph -3 | 20% | 25% | 55% | Neutral | Neutral (60%) |
| Paragraph -4 | 40% | 10% | 50% | Neutral | Positive (75%) |
| Paragraph -5 | 60% | 10% | 30% | Positive | Positive (65%) |

## V. CONCLUSION

With rapidly increasing technology, the early approach of word-of-mouth has been shifted towards the mass opinion what the people like and appreciate in majority. People rely and make decisions based on reviews and opinions. The rise in user-generated content for Bangla language across various genres news, culture, arts, sports etc in the web has open the data to be explored and mined effectively, to provide better services and facilities to the consumers. The system described in this paper proposes a method to recognize the sentiment or opinion from Bangla text using valency analysis. The system first performs simple parsing to identify the parts of speech and then applies rules to assign contextual valence (polarity) to the linguistic components in order to obtain sentence level

sentiment valence. The major limitation of the paper is that we have used the WordNet and SentiWordNet which is basically designed for English language. There is no problem to use WordNet and SentiWordNet for Bangla text if the combinatory potential of word and phrase are same for Bangla and English. But sometimes it differs. Hence it is important and necessary to develop WordNet and SentiWordNet for Bangla language.

## REFERENCES


[1] Pang, Bo and Lillian Lee, "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval", Vol. 2(1-2), pp. 1–135, 2008.

[2] K. M. Azharul Hasan, Md Sajidul Islam, G. M. Mashrur-E-Elahi, Mohammad Navid Izhar, "Sentiment Recognition from Bangla Text", Technical Challenges and Design Issues in Bangla Language Processing, 2013.

[3] K. M. Azharul Hasan, Al-Mahmud, Amit Mondal, Amit Saha, "Recognizing Bangla Grammar using Predictive Parser", International Journal of Computer Science & Information Technology (IJCSIT), 3: 6. pp. 61-73, 2011.

[4] K.M.A Hasan, A.Mondal, A.Saha "A context free grammar and its predictive parser for bangla grammar recognition" 13th International Conference Computer and Information Technology (ICCIT), pp. 87-91, 2010.

[5] Md. Asfaqul Islam, K. M. Azharul Hasan, Md. Mizanur Rahman, "Basic HPSG Structure for Bangla Grammar", Proceedings of the 15th ICCIT, pp. 185-189, 2012.

[6] Wilson, T., Wiebe, J., and Hoffmann, "Recognizing contextual polarity in phrase level sentiment analysis", In Proceedings of HLT/EMNLP, pp. 347–354, ACL, 2005.

[7] Esuli, A., and Sebastiani, F, "SENTIWORDNET: A publicly available lexical resource for opinion mining" In Proceedings of LREC-06. LREC, 2006.

[8] Strapparava, C., and Valitutti, A., "Wordnetaffect: An affective extension of WordNet" In 4th LREC, pp. 1083-1086, LREC, 2004.

[9] Fellbaum, C. (ed.): WordNet: An Electronic Lexical Databases. MIT Press, Cambridge,Massachusetts 1999.

[10] Subhash Chandra and Pampa Bhattacharyya, " Valency Analyzer of Verb Arguments for Bangla" proceeding of 13th Oriental COCOSDA Workshop, 2011.

[11] Debaprasad Bandyopadhyay, Valency of Bangla Verb and Problem of Compounding , 1998, Available at SSRN: http://ssrn.com/abstract=2029308

[12] Manas Kundu, "Machine Understanding of Bangla Sentences", Technical Paper, Interdisciplinary Series-3, Indian Statistical Institute, Kolkata

[13] Das, A., and Bandyopadhyay, S., "Subjectivity detection in English and Bengali: A CRF-based approach", In Proceeding of ICON. Macmillan Publishers, 2009.

[14] Das, D., Roy, S., and Bandyopadhyay, S., "Emotion tracking on blogs - A case study for Bengali", In Jiang, H. et al. (Eds.), IEA/AIE 2012, (LNAI) ol. 7345, pp. 447–456). Springer, 2010.

[15] Das, A., & Bandyopadhyay, S. "Phrase level polarity identification for Bengali", International Journal of Computational Linguistics and Applications, 1(2), pp. 169–181, 2010.

[16] Dasgupta, S., & Ng, V., "Topic-wise, sentiment-wise, or otherwise? Identifying the hidden dimension for unsupervised text classification", In Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 580—589, EMNLP, 2009.

[17] Dasgupta, S., & Nga, V., " Mine the easy, classify the hard: A semi-supervised approach to automatic sentiment classification", In Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP), pp. 701-709, ACL, 2009.

[18] Mostafa Al Masum Shaikh, Helmut Prendinger, and Ishizuka Mitsuru, "Assessing Sentiment of Text by Semantic Dependency and Contextual Valence Analysis", Second International Conference, ACII 2007 Lisbon, Portugal, September 12-14, 2007 Proceedings, pp. 191-202, 2007.