THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Sentence Level Sentiment Analysis in Malayalam

Aleena Beegum
M. Tech Scholar, Department of CSE, Ilahia College of Engineering & Technology,
Muvattupuzha, India
Noorjahan V. A.
Assistant Professor, Department of CSE, Ilahia College of Engineering & Technology,
Muvattupuzha, India

Abstract:
In this paper, we propose a sentence level sentiment analysis for Malayalam language. This is to analyze the sentiment of Malayalam sentence. In the existing system sentiment of a sentence in English is done by reversing the sentence or split the sentence as words sequence. When the sentence is split into words sequence, polarity inconsistency occurs. We address this issue by segmenting the sentence as unigrams and bigrams then predicts sentence level polarity based on the segmentation results. Here we introduce a training phase and prediction phase. We extend the polarity framework (positive-negative) into three classes (positive-negative-neutral) by considering neutral reviews into consideration. Finally, we develop a corpus based method to get the polarity of words. We conduct experiments for sentiment analysis on different review datasets. Experimental results show that: (1) our method will perform well on different datasets of film reviews; (2) when we consider the neutral polarity then gets more accurate result than the binary polarity (positive-negative).

Keywords: Sentiment analysis, sentence-level sentiment analysis, natural language processing, document-level sentiment analysis

1. Introduction

In natural language processing sentiment analysis plays an important role. Sentiment analysis also known as opinion, it analyses and extract information from people's opinion/emotions etc. It also determines whether our sentence is positive-negative-neutral. Malayalam belongs to Dravidian family, a large family of languages of south and central India.

In proposed system we can classify the sentence into three polarity classes positive, negative and neutral by using polarity prediction. This paper discusses about two phases— training phase and prediction phase. In training phase, we segment the sentence into words and combination of words, and then calculate its score values. In prediction phase we predict the polarity of the sentence.

Sentiment analysis has two main approaches— sentence level sentiment analysis and document level sentiment analysis [3]. Sentence level sentiment analysis is used to determine the sentiment of a sentence, which have been commented on by an opinion holder and determine the polarity of the sentence. Document level sentiment analysis is used to classifying a document based on the overall sentiment of the document.

In proposed system we can classify the sentence into three positive, negative, and neutral sentences by using polarity prediction.

2. Related Work

Here we are discussing some papers on sentiment analysis.

Sentiment analysis is an active research in NLP, analyze people opinions, and sentiments. Sentiment analysis of Malayalam film review is discussed on this paper. In this paper sentiment analysis is done by using rule based approach. This paper will give the information about new films. The other machine learning or rule based approaches can be used for this system. This system has 82% accuracy [1].

Bag-of-words (BOW) is the most popular way to model the text in sentiment analysis. However, the performance of BOW is limited due to polarity shift problem. This paper proposes a dual sentiment analysis (DSA), to address this problem. The basic idea of DSA is to create reversed reviews that are opposite to the original reviews. It contains two algorithms (dual training and dual prediction) to classify and predict the sentiment of a sentence. It focused on creating reversed reviews to assist supervised sentiment classification. There are two different phases Dual training and Dual prediction. Training is used to segment the sentence by words and its combinations and calculate the score. In prediction predict the polarity of sentences [2].

The system develops a joint segmentation and classification framework for sentence level sentiment classification. This paper implements a candidate generation model and segmentation ranking model to avoid inconsistency in sentiment polarity. Experiment

on tweets datasets and a review dataset, this method outperforms baseline methods. The result of bag-of-word is 75% accuracy and this method has 85% result accuracy [3].

The system determines the sentiment polarity requires more than one simple bag-of-words approach. This paper experiments on simple heuristics based on compositional semantics, learning–based methods that do not incorporate computational semantics with 89% accuracy result. Learning with computational semantics has 90% accuracy. This system will find that context-word negators, are not work properly on previous works. Example This is not bad, Overall polarity of this sentence is positive because here two negative words (not, bad). Implement this concept to get polarity of sentence, we introduce neutral polarity to get accurate result [4].

Sentiment classification is to analyze the subjective information in text. Sentiment analysis is to extract information from the opinion and emotion of people. More people begin to express their views and perspectives of things on the internet. Sentiment analysis has an application on discussion forum, blogs, digital libraries, movie reviews etc. The sentiment classification is divided into two categories sentence level sentiment classification and document level [5].

In this paper construct a polarity shifted corpus for different kind of product reviews. To create corpus both sentiment words and shifting words are denoted. Sentiment classification is to classify the text document or sentences. Polarity shift is the main sentiment classification problem. The system will implement two steps for polarity shift in sentiment calcification, detect the polarity shift and designing specific classification algorithms. Some words are difficult to judge the polarity [6].

3. Proposed System
Here we are proposing a method to determine the polarity of sentences in Malayalam language. The words of the sentence are stored as numbers in database. First input the sentence, segment the sentence into words and combination of words. The combinations of words are classified as unigrams, bigrams, trigrams etc… For reducing the computation time, we usually use unigrams and bigrams. Figure shows the architecture of the system.
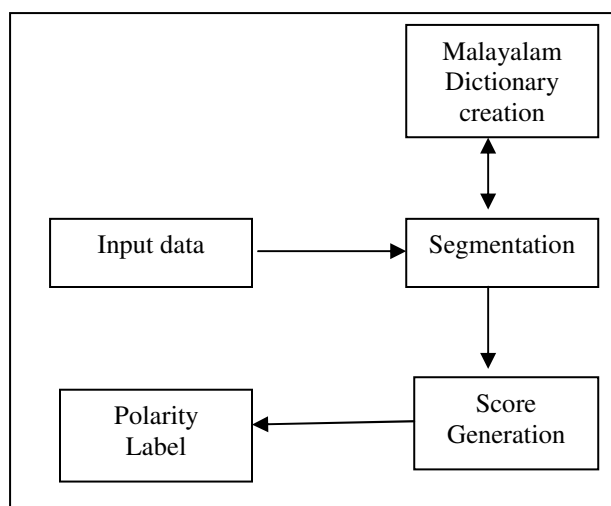

Figure 1: Base architecture of System

There are two modules ***training and prediction.*** In the ***training phase***, we train the dataset in all possible ways. For that first create a phase dictionary and a polarity dictionary. Then the comments are read and write into a file in a specified format. The segmented by single word and combination of words such as bigrams. Then calculate the score of the segmented words. Sum up the single word score and the combination words. And then find the top scored sentence.

In ***prediction phase*** we predict the polarity of sentence using phrase dictionary, polarity dictionary and the score values.

3.1. Phrase Dictionary
In phrase dictionary store the score value and domain of each single word and its combination.

3.2. Polarity Dictionary
Polarity dictionary contain the words, combination words and its polarity.

4. Solution Methodology

4.1. Algorithm
1. Input the Malayalam sentence.
2. Segment the sentence by single word and combination of words.
3. Calculate the Sum of score by using the equation score of single word= freq (w1) / total number of words score of combination of words = freq(w1*w2) / freq(w1) *freq(w2)

4. Check the sentence polarity using   function.

4.1 Set pos count, neg count and neu count =0.

4.2 If the pos count >neg count && pos count > neu count then set   the polarity as positive.

4.3 If the neg count >pos count && neg count >neu count then set the polarity as negative.

4.4 If the neu count >neg count && neu count> pos count then check if pos count, neg count!=0 and pos count = neg count then set the polarity as negative.

4.5 If the pos count > neg count then set the polarity as positive.

4.6 If the pos count =0 && neg count = 0 then set the polarity as neutral.

4.7 Neither these cases set the polarity as negative.

4.8 Out put the sentence polarity.

freq (w1) -Frequency of first word,

freq (w2) - Frequency of second word,

pos count- positive count,

neg count - negative count,

neu count - neutral count.

5. Experimental Result

5.1. Dataset

We conduct sentiment classification on a dataset from different domain. Run the Malayalam film review dataset as positive-negative-neutral polarities, sentence segmentation has a great influence on the positive-negative-neutral polarity in Malayalam sentence.

Malayalam sentences are inputted to the system then get the sentiment of the sentence. There is a corpus called phrase dictionary to get the single word and the combination words. There is a polarity dictionary to get the polarity of single word and combination of words. Table 1 shows the example of polarity dictionary.

Word	Polarity
ഡയറക്ഷനെ	Neutral
ശരിക്കുയും	Positive
ബാധിച്ചു	Negative
ശരിക്കുയും ബാധിച്ചു	Negative

Table1: Polarity Table

Example:

ഡയറക്ഷനെ ശരിക്കുയും ബാധിച്ചു.

After analyzing get the polarity of this sentence as negative.

In this proposed system mainly two matrices are used to calculate the performance of the system which are the Precision and Recall. Precision is the percentage of words tagged correctly from the total words addressed by a particular system. Recall is the percentage of words tagged perfectly from total words in a test set. When we take total words as 100 then the precision we get is the 90 % and the recall is the 88%.

Precision= 90/100

Recall   = 80/90

From this result performance of the system is higher compared to other system.

6. Discussion

We briefly discuss about the three classes of polarity from the previous one. In sentiment analysis the main problem is polarity shift problem [2]. For example, add negative word "don't" to a positive sentence "I like this camera" in front of "like", the sentence polarity is changed from positive to negative. We discuss about the accuracy of the reviews while considering the neutral polarity.

7. Conclusion

In this paper, we present a sentence level sentiment analysis in Malayalam. In existing system sentiment classification is based on segmentation result from Bag-of-words or separate segmentors, in proposed system generates segmentations and predict the sentence level sentiments. Classify the sentence into three classes positive, negative, and neutral.

Experiment is done in different datasets and the result on the classification prediction using positive - negative polarity and positive-negative-neutral.

Future work is to extend the classification of polarity frame work into very negative, negative, neutral, positive, very positive.

8. References
i.   Jisha P. Jayan, Deepu S. Nair, Elizabeth Sherly "A subjective feature extraction for sentiment analysis in malayalam language",vol.15.2015.
ii. Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi and Tao Li, "dual sentiment analysis:considering two sides of one reviews",vol 27.2015.
iii. Duyu Tang, Bing Qin, Furu Wei, Li Dong, Ting Liu and Ming Zhou , "A joint segmentation and classification frame work for sentence level sentiment classification " vol. 23.2015.
iv. Yejin Choi and Claire Cardie " Learning with compositional semantics as stuctural inference for sub sentential sentiment analysis" IEEE 2008.
v.   V.S.Jagtap and Karishma Pawar "Analysis of different approaches to sentence level sentiment classification," vol.2., pp.164-170, IEEE 2013.
vi. Xiaoqian Zhang,Shoushan Li and Hongxia Zhao, " Polarity shifting:Corpus construction and analysis",IEEE 2006.