

A Comparison between Sentiment Analysis of Student Feedback at Sentence Level and at Token Level

¹ Chandrika Chatterjee, ² Kunal Chakma

^{1,2} Computer Science and Engineering, National Institute of Technology-Agartala,
Agartala-799055, India

Abstract - Sentiment classification is a special case of text classification whose aim is classification of a given text according to the sentimental polarities of opinions it contains, that is *favorable* or *unfavorable*, *positive* or *negative*. Student feedback is collected as response to set of positive and negative questions. The idea is to identify and extract the relevant information from feedback questions in natural language texts in order to determine a set of best predictive attributes or features for classification of unlabelled opinionated text. Sentiment classification is used for training a binary classifier using feedback questions annotated for positive or negative sentiment and evaluates the corresponding feedback received. This paper compares the sentiment classification of student feedback questions at sentence level and at token level for different classifiers.

Keywords - *Sentiment Analysis, Tokens, Classification, Support Vector Machine, Decision Tree.*

1. Introduction

The motive of opinion mining or sentiment analysis is to classify the polarity of a given text at the document, sentence, and phrase or word level — characterize whether the overall viewpoint expressed is positive, negative or neutral. “One set of sentiment analysis problems share the following general character: given an opinionated piece of text, wherein it is assumed that the overall opinion in it is about one single issue or item, classify the opinion as falling under one of two opposing sentiment polarities, or locate its position on the continuum between these two polarities. A large portion of work in sentiment-related classification/regression/ranking falls within this category.” [1]. Feedback systems for course evaluation are necessary to improve teaching effectiveness and course quality. However student feedback evaluation system generates enormous quantity of data making the analysis difficult thus needs an automated analysis. To ensure conscious feedback the feedback form is a set of positive and negative questions mixed together .Since we

gather student feedback in form of responses to questions in a single sentence, we need sentiment analysis in sentence level. In sentiment classification, machine learning methods have been used to classify each question as positive or negative .We test our test data based on our training model which is classified using supervised learning algorithm. In the second step we evaluate the total responses for every question and determine the polarity of feedback received in context of the question. The evaluation of response is purely data driven and hence simple while the classification of questions in form of natural language texts involves sentiment analysis which is analyzed in this paper. To test our model we collected data from students who posted their views in online discussion forums dedicated for this feedback evaluation purpose. Here we show the sentence level classification based on classification of constituent tokens. The sentiment polarities of each sentence are assigned to its constituent tokens to train the classifier. We illustrate the comparison between the sentiment classifications at sentence level and at token level. The results for the different classifiers are analyzed for two test data sets.

The rest of the paper is described as follows. Section 2 highlights the related work in this area; our proposed work is discussed in Section 3, followed by the Results and Analysis in Section 4. The paper concludes in Section 5.

2. Related Works

Traditional student course evaluation feedback systems are pen and paper based where it generates a huge amount of data and hence makes the feedback analysis very difficult. We find in [2] opinion mining of students’ posts in Internet forums to classify their opinion about course and comparisons of accuracy of different classifiers. In [3] Song et. al. proposed automatic analysis of texts to improve E-Learning system.[4] analyzed the role of

semantic orientation and gradability of adjectives for prediction of subjectivity. Bo Pang [5] proposed different models for distinguishing between opinions and facts and between positive and negative pinions. Opinion was classified at sentence level which is much harder than document level sentiment analysis as words and relationship among words play a vital role in determining the polarity. Sentiment polarities were assigned to individual words and accuracy achieved was quite high. Reference [6] used word dependencies to show the contribution of phrase level sentiment analysis to evaluate sentiment at sentence level. In this paper we propose an automatic sentiment polarity classification of student feedback system. For token level classification the sentiment polarities of each sentence are assigned to each word of the sentence for training set. The accuracy for token level and sentence level classification of student feedback data is compared.

3. Proposed Work

This paper aims to show the different accuracy level of sentiment classification of student feedback questions when sentiment analysis is done at sentence level and sentiment analysis is done at token level. The outcome results using different classifiers are shown.

3.1 Sentence Level

Here the classifier assigns the class label to each sentence. Sentiment Analysis is done at sentence level.

3.1.1 Corpus

The performance of a data mining classifier relies heavily on the quality of the database used for training and testing and its similarity to real world samples (generalization). The required data are collected from on-line feedback survey. The training data consists of equal number of questions classified as positive, negative. Students can respond by marking the question as agree, do not agree or cannot say. We collected 300 questions out of which we choose 120 suitable questions as training data and 2 test sets with 25 different questions in each test dataset. The training set and the two set questions are all mutually exclusive questions.

3.1.2 Feature Extraction

For creating the machine learning model we use RAPIDMINER 5.3.015. RapidMiner is a world-leading open- source system for data mining [7]. We convert our corpus to feature vector.

3.1.2.1 Tokenization

Tokenization is the process of splitting a piece of text into its constituent meaningful elements called tokens. The list of tokens becomes input for further processing such as parsing or text mining.

3.1.2.2 Stemming

Stemming is the normalization of natural texts where words are conflated to stem. A number of so-called stemming Algorithms, or stemmers, have been developed, which attempt to reduce a word to its stem or root form. Thus, the key terms of a query or document are represented by stems rather than by the original words. Here we use Porter stemming algorithm.

3.1.2.3 tf-idf

tf-idf or term frequency-inverse document frequency, is a numerical statistic that is concerned for the purpose of reflecting how important a word is to a document tf in a collection or corpus. The tf-idf value is calculated by the ratio of number of occurrences of a word in a document to the total number of words the document contains. It addresses the fact that some words in general occur more commonly than others.

3.1.2.4 Generate n-grams

This operator creates term n-Grams of tokens in a document. A term n-Gram is defined as n visible consecutive tokens of a text. The term n-Grams consist of all series of consecutive tokens of length n. n is the maximal length of the ngrams. We use n as its default value: 2.

3.1.3 Classification

The supervised procedure of labeling an opinionated document as expressing either a positive or negative opinion by learning from already labeled instances in train data set is called sentiment polarity classification or polarity classification. The training set, consists of records each having multiple features (bigram, stem, *tf-idf* in this case) and are labeled. Classifier assigns a class label to a set of unclassified cases. A naive Bayes classifier assumes that the value of a particular feature is unrelated to the presence or absence of any other feature. Support Vector Machine gives the optimal hyper plane which gives the minimum distance to training samples. A decision plane separates between objects that belong to different classes.

3.2 Token Level

Fig. 1 shows the workflow diagram of sentiment analysis at token level.

3.2.1 Corpus

The same set of data (training set and test sets) that we used for sentence level sentiment analysis will be used.

3.2.2 Feature Extraction

For creating the machine learning model we use WEKA 3.6.10 API in our java code [8].

3.2.2.1 Tokenization

We use Stanford Tokenizer to a break the stream of question up into its constituent words, phrases, symbols called tokens. We create a dictionary of tokens in the training and test set separately.

3.2.2.2 POS Tag

POS Tagging is the process of tagging up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition, as well as its context. We use Stanford POS Tagger, v3.4 to tag all the tokens of test questions and create a POS dictionary for the training set and test set respectively.

3.2.2.3 Stemming

We use Porter stemming algorithm for stemming of tokens. It is used to remove the commoner morphological and inflexional endings from words in English.

3.2.2.4 Generate n-grams

A bigram is window of every two adjacent elements in a string of tokens, which can be characters or words; they are n-grams for $n=2$. So here we add the previous token and next token as a feature of any token.

3.2.3 Classification

We use J48 for classification of tokens as it gives the best performance. J48 is an open source Java implementation of the C4.5 algorithm in the weka data mining tool. C4.5 is an extension of Quinlan's earlier ID3 algorithm which is based on a top-down, greedy search through the space of possible branches without any backtracking. Here the classifier assigns the

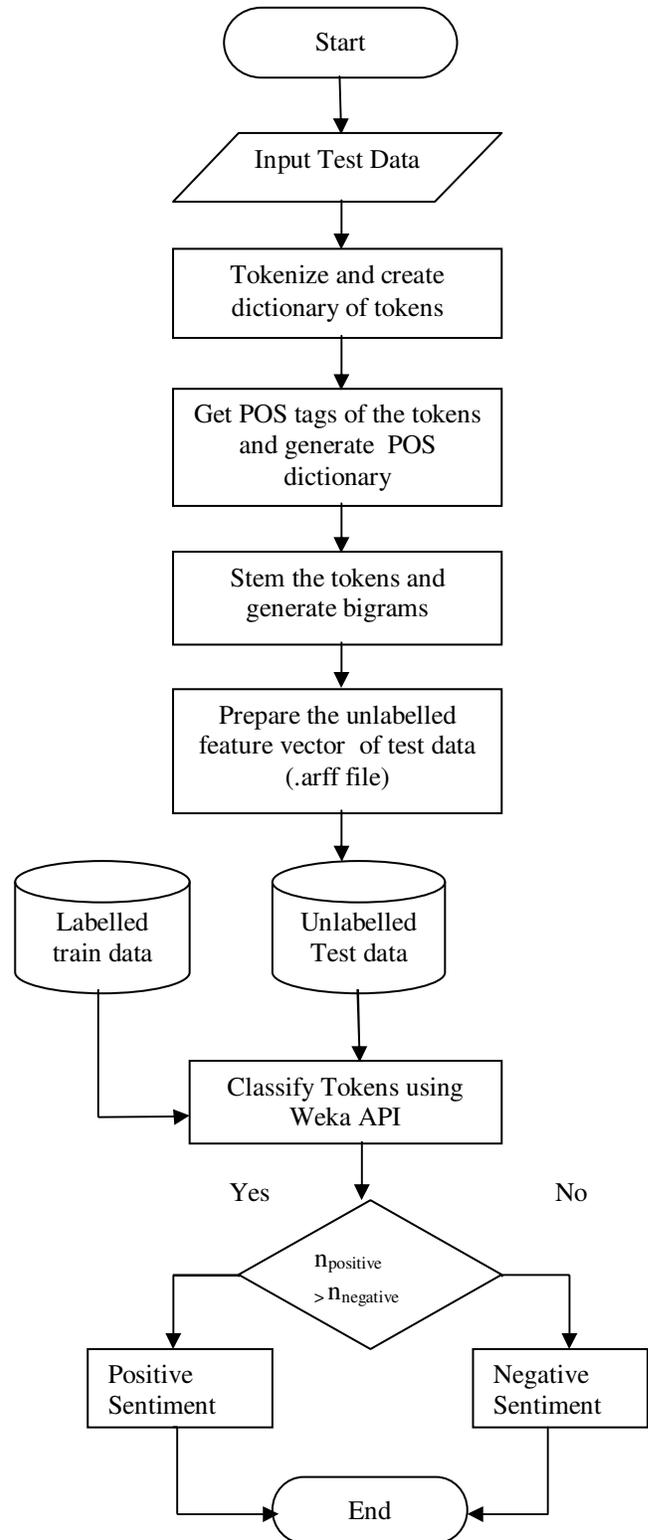


Fig. 1 Token Level Sentiment Analysis

class label to each token. Net class assigned to each sentence is $\max(n_{\text{positive}}, n_{\text{negative}})$ where n_{positive} is the number of tokens in the sentence that are classified as positive and n_{negative} is the number of tokens in the sentence that are classified as negative. Sentiment Analysis is done at token level.

4. Result and Discussion

From Table 1 and Table 2, accuracy of two test sets indicate that the token level opinion classification gives improved accuracy (4-5%) when any decision tree based classifier is used. Precision defines the fraction of selected items that are correct. Recall defines the fraction of correct items that are selected. In Table 3 we learn that for both the test sets negative class recall is comparatively much higher than positive class. Precision is fairly good for test set 1 but quite poor for test set 2. Traditional F-measure is the harmonic mean of precision and Recall. Its value lies between 0 and 1. Table 3 shows fair F-measure value of negative class in both test sets and poor F-measure value for test set 2.

Table 1 Accuracy of Test set 1 (in %)

Classifier	Sentence level	Token level
Support vector machine	72	52
Naïve Bayes	72	52
J48	72	76
Decision table	72	76

Table 2 Accuracy of Test set 2 (in %)

Classifier	Sentence level	Token level
Support vector machine	72.22	50
Naïve Bayes	72.22	50
J48	72.22	77.77
Decision table	72.22	77.77

Table 3 Precision, Recall for Token level classification J48

	Test set 1		Test set 2	
	positive	negative	positive	negative
Precision	0.875	0.706	0.357	0.692
Recall	0.583	0.923	0.555	1.00
F-measure	0.700	0.800	0.434	0.815

5. Conclusion

In the comparison, we see that token level sentiment analysis for decision tree based classifier gives improved results. Decision tree builds classification models or regression models in form of tree structure. It decomposes a dataset into smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. Decision tree based classifiers are very flexible and addresses every possible failure, however failure for a system concerns small set of symptoms that are relevant to the system at hand. However SVM and Naïve Bayes classifier gives poor results in token level classification. Changing the feature set we can analyze the different accuracy levels of sentence level sentiment analysis and token level sentiment analysis. Although the carefully chosen corpus in training set for training the classifier contained fifty percent positive and negative classes yet we observe a very good recall of negative class but very poor recall of positive class. This biased classification should be analyzed.

6. Future Scope

As our corpus is small and defined set of feedback questions, the differences in classification is not so prominent. Moreover any special suggestions from students are not entertained. The system developed assumes that feedback questions fed as input are a standardized set of questions without any explicit sarcasm and follows Standard English language.

The corpus was restricted to student feedback questions. The comparison in accuracy we showed here between token level and sentence level occurs in context to this domain. For other domain this comparison should be evaluated. Handling relation between tokens is also an option to be explored in sentiment analysis at token level.

Acknowledgment

The authors are thankful to the administration of my college National Institute of Technology, Agartala for providing sufficient computing facilities. A lot many thanks to my guide Mr Kunal Chakma and other Faculty members of this institute for their support. Special thanks also go to my friends for helping me in the due course of my experiment.

References

- [1] Bo Pang and Lillian Lee "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* Volume 2, Nos. 1–2 (2008) 1–135.
- [2] Alaa El-Halees "Mining Opinions in User-Generated Contents to Improve Course Evaluation" in *The Second International Conference on Software Engineering and Computer Systems*, 2011. Part II, CCIS 180, pp 107-115.
- [3] Song, D., Lin, H., Yang, Z.:" Opinion Mining in e-Learning" In: *IFIP International Conference on Network and Parallel Computing Workshops* (2007)
- [4] Vasileios Hatzivassiloglou ,Janyce M. Wiebe "Effects of adjective orientation and gradability on sentence subjectivity," in *Proceedings of the 18th Conference on Computational linguistics - Volume 1*, pp 299-305.
- [5] Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. "Thumps Up?Sentiment classification using machine learning techniques," In *Proceedings of the 2002*

Conference on Empirical Methods in Natural Language Processing (EMNLP-02) Philadelphia, Pennsylvania pp79 -86.

- [6] Arun Meena ,T.V. Prabhakar "Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis," in *Advances in Information Retrieval Lecture Notes in Computer Science* Volume 4425, 2007, pp 573-580.
- [7] Sebastian Land, Simon Fischer"RapidMiner 5," *Rapid-I Gmbh*, 2012.
- [8] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); *The WEKA Data Mining Software: An Update*; SIGKDD Explorations, Volume 11, Issue 1.

Authors

Chandrika Chatterjee is pursuing Masters of Technology in Computer Science & Engg. Department, From National Institute of Technology, Agartala (Tripura). She has completed her Bachelor of Engineering in Computer Science and Engg. Department, from Gurunanak Institute of Technology, Kolkata (West Bengal) under West Bengal University of Technology. Her field of interest is focused on data mining, Natural Language Processing, Machine Learning and Evolutionary Computing.

Kunal Chakma is working as Assistant Professor in Computer Science & Engg. Department, National Institute of Technology, Agartala (Tripura). He has completed Masters of Technology in Computer Science & Engg. Department, from National Institute of Technology, Agartala (Tripura). His field of interest is focused on Data mining, Natural Language Processing, Systems Software, Operating Systems.