ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

Dynamic Feature Subsumption based Multiclass Sentiment Analyzer using Machine Learning Techniques

Jayanag.B[1], Dr. K.V.Sambasiva Rao[2]

Senior Assistant Professor, Department of CSE, V.R.Siddhartha Engineering College, Vijayawada, India[1]

Principal, M.V.R College of Engineering, Paritala, India[2]

Abstract: Sentiment classification on product reviews which has been studied earlier concluded that the subjective text consists of either positive and negative opinions but the user reviews can be classified in to broader level(multiclass) which gives deeper view of a user in more than two classes .Multiclass classification has been done for different domains using the overall ratings for a product given by the user, but nobody concentrated on classifying each user opinion in to multiple categories using the scoring function. We proposed a novel methodology to get multiclass sentiment labels for each textual comment considering each feature of the product that too for each sentence in the comment. Maxentropy parts of speech tagger proposed by Stanford University is used in our work to extract tagged features from the text which are used in identifying opinion of a user. Our work improved the original porter stemmer algorithm by adding new rules so as to improve overall performance. Results are evaluated on training and testing data when given to machine learning algorithms, our approach got high percentage of accuracy when is compared with other existing works.

Keywords: Sentiment classification, machine learning techniques, feature subsumption, Natural language processing(NLP), sentiwordnet, support vector machine(svm).

I. INTRODUCTION

Millions of people worldwide depend on online services to purchase a product from their homes which saves their effort and time. The increasing profits of e-commerce year after year shows that people had changed the habit of going to a shop for shopping. People started searching for useful information on user opinions before purchasing a product, because to purchase a product any buyer needs to know the quality of the product, this is possible when they evaluate each user review on that product. But the reviews are so bulk that manually cannot give summarized result for large set so it is required to have an automatic system which can analyze all reviews and can present the overall opinion of the product. Many works on sentiment analysis classified reviews on products as positive or negative without considering the context or targets on which opinions are expressed. Multiclass sentiment analysis gives deeper understanding on each review, i.e strength of positive as well as negative statements using multiple class categories. Sentiment analysis on complete document or for each sentence based on term frequency of words is common approach to classify opinions, but to consider multiple features of a product as targets in each sentence of each comment is a challenging task.

Section II describes the related work on sentiment analysis section III describes our proposed methodology with algorithms section IV shows results and section V concludes the paper.

II. RELATED WORK

Research works on opinion mining is done to classify reviews in to either positive or negative classes using machine learning as well as natural language processing techniques. Sentiment analysis of reviews from online are taken in many studies to classify people's opinions using some of the machine learning algorithms and also utilized feature extraction methods of NLP .Below gives idea of previous works on binary classification of product reviews who used different classification algorithms on their proposed approaches.

Changli zang[1] has worked on two class sentiment classification of Chinese reviews. They observed each review for appraisal phrases so as to extract subjective text fragments which decides the opinion of a user. Appraisal features are represented in word subsequences format also called string kernel but this representation is used for document similarities. Their approach did not concentrated on how to handle modifier dependency words like "not".

Alec go[2] analyzed the sentiments of twitter microblog on multiple domains. It is also a binary classification model which collects twiiter comments for each query term and divides training dataset in to positive and negative using the emoticons which are removed while post processing the data. Machine learning algorithms are trained for POS tagged unigrams and bigram features taken from training data but while test data which may have emoticons cannot be used to refer the polarity of a comment by the learning algorithm so this makes classification tough and decreases the accuracy.

Mathew[3] worked on classifying restaurant reviews in to either positive or negative. They used bag of sentiment words from each review and their frequencies to train ensembled algorithms. But to train large set of unigrams for each ensembled model is very difficult which decreases the accuracy.

Akshat[4]they concentrated on binary classification of reviews on multiple domains .They used simple n-gram

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

features which omitted preprocessing steps like stopword removal, stemming. Negation handling mechanism using POS tagged n-grams is very complex.

Trivedi[5] paper sentence classification using naïve bayesian algorithm has been proposed. They maintained tri ,bi and unigram words data which are labeled with polarity ,this is static method which increases the effort to analyze the overall product or a brand reviews. The words in database is matched when the order of them does not change while the natural languages are free form texts certainity on their order cannot be expected. So there should be loose proximity over the order.

In [6] they proposed binary sentence level classification using SCDDF approach. Each product review is classified based on term counting method i.e if number of positive words are more than negative their approach summarizes as positive with our checking target on which the words are used which decreased their classification performance. Dependency words handled without checking the context and the object on which it is used, so their predictions are not accurate.

R c balabantaray [7] proposed multiclass classification of user tweets without any particular topic of interest in twitter in to 6 categories of human emotions. The approach used by them is to manually annotate the tweets with emotions as class labels rather than automatically classifying and used different methods to select features, for each method given with fixed weights. These data is used to analyze the performance by svm classifier.

Gillimanik[8] has worked on tweets written on famous people and also on general talk.They labeled tweets with multiple emotion classes with the help of workers who agreed to label emotions manually with their human intelligence rather than machine intelligence. Annotated tweets are subjected to machine algorithms to do that bag of words are used to represent a tweet.

III. PROPOSED APPROACH

In our approach we perform multi category feature wise classification of unannotated raw comments on product and their features. Our proposed methodology is extension of SCDDF approach[6]where we automize the multiclass classification task using calculated weighted score for each sentence in each document Working steps are shown in the architecture depicted below.

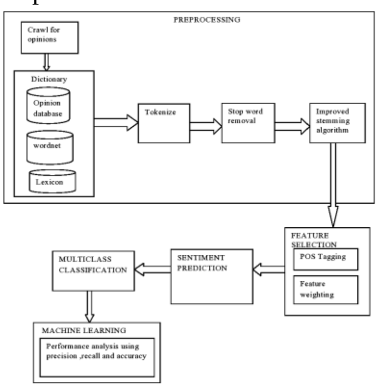

Figure 1: Extended SCDDF architecture

A. Data Preprocessing

User reviews can be retrieved using different web crawlers or web spider tools available online. We used Microsoft web access manger to retrieve product reviews online which we further refined by removing the nonrelevant data so that to have only opinions. Raw opinions still not ready for classification so data is preprocessed before assigning the labels. We included 2 steps in preprocessing firstly we tokenize each line of the review using line tokenizer. Upper case words are replaced with lower case letters. Removed the regular expressions like punctuations, alpha numerics and symbols, also filter stopwords such as connectors (“and”,”or”) indefinite articles(“a”,”an”,”the”) etc a total of 318 words list are removed from dataset. In the list we have not included the negation words like “not” as they are key factors in handling “negation phrases” in a sentence. Second step is to remove suffixes using improved stemmer algorithm and later features are identified by mapping each word with unigram lexicon dictionary. A sentiment lexicon is created with possible positive and negative words which are used as dictionary while mapping the sentences.

B. Improved Stemmer Algorithm

Porter stemmer algorithm is last step of preprocessing phase this method is used commonly in many information retrieval systems to strip suffix from each word. It is used to minimize the data set and also to find stem of the word, so as to identify the opinion of each sentence. We studied on improved version of this algorithm [10] and found some errors and come with solutions to them described below.

Error1: If a word ends with suffix “ing” the algorithm removes the suffix, but some words with “ing” forms where m=1 (m=vowel-consonant pair) does not get meaningful roots for example see “using” etc when stemmer is started it shows the root words “us” which gives different meaning for the same word and cant be used to determine sentiments or to find the features of a sentence. In the same way we found errors “ying” suffix words “dying” and ”buying” both give words dy and bu as the root words found.

Solution 1: This problem is resolved by us by adding “e” to a word that ends with “ing” and its m=1 then suffix is replaced by “e” whereas while m=0 where stem starts with vowel followed by the suffix then it is replaced by “ie” or a word ends with “ying” and found a vowel before a Y suffix the “ing” is removed.

(Dying->die, Buying->buy)

Error 2: if the suffix is y it is changed with I for example happy is changed to happi which makes difficult to identify features and also to identify their parts of speech happy-> happi

Solution 2: we did not touched or changed such words(Happy->happy) because there is no use of replacing ‘y’ with ‘i’ for such words.

C. Feature Selection

Until now we have processed the data to extract features efficiently and easily to predict sentiments. Feature engineering is necessary to get patterns of information so it will be easy to learn and classify the sentiments. We considered phrases showing different feelings through opinion words on particular product features. These phrases give us features sets for classification. To identify them we used pos tagging method of Stanford which tag each word with respective parts of speech. Instead of all words in each sentence we consider nouns, adjectives and adverbs as our features tagged by the tagger. We map each word with our lexicon words if they match they are taken in the feature set. Here we take unigrams of parts of speech features checking the presence of the word in the dictionary then values for them are given using score function.

D. Parts of Speech tagging

POS tagged words are source to identify the required features of our analysis so we used Stanford maximum entropy modeled tagger[11] which is developed and tested. Maxentropy pos tagger model is used to automatically tag each sentence of whole document. To tag a sentence with one of the parts of speech like any other classification problem needs features so here this model uses syntactic and semantic contextual features. To understand the context of a word in a sentence, left 3 words of current token to be tagged should be checked i.e it uses left 3 words architecture along with distribution similarity features. This Tagger is trained with wall street journal(WSJ) corpus data annotated with Penn tree bank tag set and tested them with other section of data. The taggers efficiency is high having an maximum error rate of only 3% because it's approach is to tag each word is a enhanced approach considering syntatic structure as well as context of each word in a sentence based on the previous tag assigned. Performance is 97% when WSJ data is tested by left3words tagger model.

Features that we take for multiclass sentiment prediction are parts of speech tagged unigrams. As we need emotions of user we take adjectives and to find the topics on which user spoke we also use nouns. Taking all the nouns and adjectives increase the time of classification. We are interested in only subtopics of the main topic for example mobile is main topic and the features of mobile like battery(nouns) which are properties/characteristics of a product can be a subtopic of a product. To see the relativity of the feature towards the opinion we check each unigram whether it is a noun, if it is a noun then check it is a product feature or not next adjective(JJ) words are searched before or after the noun feature. Unlike n-grams model relatedness of the opinion towards the feature is checked without rigidity on the order of unigrams i.e opinion words can appear before or after a feature until we find a new feature that a user mentions. Generally we see in any languages qualities of an object are expressed before or after the object so we focus on how to extract such features. These unigram features are referenced from left to right in a sentence.

E. Feature Weighting

We considered only specific unigram features which follows above rules described .Features are weighted based on frequency of words in a document i.e term count method when unigram model is used. But instead we have given weights to opinion words of total 386 using sentiwordnet [9] online knowledge resource and created a positive words dictionary. Sentiwordnet give scores in 3 categories positive score, negative score and objective score in multiple sense forms for each part of speech. As we observe subjective words which express emotions in reviews, therefore only adjective and adverbs labeled positive score from sentiwordnet is taken with respective weights for all positive or negative words list we have created. Only positive signed score of the wordnet for both positive and negative is taken i.e high weights for positive sentiments and low weights for negative sentiments.

In our work we incorporated new method to weight only product features/specifications on which comments can be written i.e a list of dynamic features of a product which are nouns are scored within range [0.7-1] manually. Reason to introduce product feature weighting is till now classification is based on opinion word neglecting the topic of interest so we try to explore the sentiments of users on features or properties of a product. Secondly we distribute weights based on importance of the feature and how useful to the user.

Example: "Memory" is given higher importance than "color" for mobile as color is least important while using a mobile. So we rank each feature by weighting method.
In order to normalize the feature vector values and to improve the efficiency of classifier we choose (min=0.7,max=1) range of values which are used in weighting the features . Product of bigrams of nouns and adjective pairs with their respective weights gives us sentiment score of a particular feature.

F. Feature based multiclass sentiment predictor

The data we have taken is unannotated data so here we annotate the data automatically in to 5 categories that represent people emotions. Excellent, very good, good, average and worst which shows difference between strongly positive and weakly positive sentiments. So here we proposed predictor algorithm shown below which takes the pos tagged texts as input and verifies the sentiments of them based on rules defined on feature sets described in the section D. Product features that are nouns is the main component that is checked in each step of classification algorithm. The algorithm also checks each and every word of pos tagged text that satisfies the above defined rules to identify the sentiments on different product features. When the features are found algorithm tries to find weights for each word in feature set selected. Weights are searched from the opinion word corpus data as well as features list data which are sources to calculate feature score. Based on this feature score average we have classified the data in to multiple classes dynamically.

IV. DATA SETS

We used product review data on mobile phones collected online. In the data we consider mainly pros and cons section which has high and deep explanation on each feature of the product. We collected a data set of 100 on

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

mobile reviews from flipkart site and also tested on product review data on iphone product used in [2], camera reviews collected from multidomain data set used by mathew[3], ipod data used in paper[4] and also collected mobile reviews from[5]. We collected reviews without counting the positive and negative number values of variable sizes. We also tested the earlier studied data sets on product reviews on our approach and showed the results for them.

A. Result analysis using machine learning algorithms:

To understand the performance of multiclass feature based predictor we used machine learning approaches like support vector machines, neural networks(multilayer perceptron) and decision support algorithms. The feature vectors with their scores obtained from our proposed approach is divided with 80% for training a classifier and the remaining data for testing. These data sets are trained and tested using weka package available publically written in java for analyzing the accuracy of our work. We analyzed the performance using the count values of correctly classified instances of each and every data set. Accuracy of a test data is calculated by following formula.

$$\textbf{accuracy} = \frac{\textbf{number of correctly classified instances of test data}}{\textbf{total number of instances of test data}} \quad \textbf{(3)}$$

Our collected mobile data when tested using the LIBSVM multiclass (nu-svc) classifier package using linear kernel we got accuracy of 50%.When we trained the mobile data for SVM classification we have changed certain parameters like we added nu-svc classifier which is used to handle multiple labels. Our work performance is high when we use linear kernel type support vector with 80% of accuracy of correctly classified instances. When the same data is classified by Multilayer perceptron we got 75% of accuracy while the decision trees scored 87.5% of accuracy.

Our work is compared with the results of earlier works on sentiment analysis by taking the results when the same linear SVM classifier, Maxentropy and naïve bayes classifiers are trained with 80 % of training and 20 % test iphone twitter data which is annotated with multiple sentiments dynamically we got maximum of 86.667% accuracy which is more than alec go work where he choosed different domain data there values are shown in table 1 which are taken when pos tagged features are used.

TABLE I: RESULTS OF IPHONE DATA

Work	Naïve bayes	Max ent	Svm (linear)
Unigram+pos(Alec go-2009)	79.9	79.9	81.9
Unigram+pos tagged SCDDF	86.667	86.6667	86.667

% of accuracies using the iphone testing and training datasets.

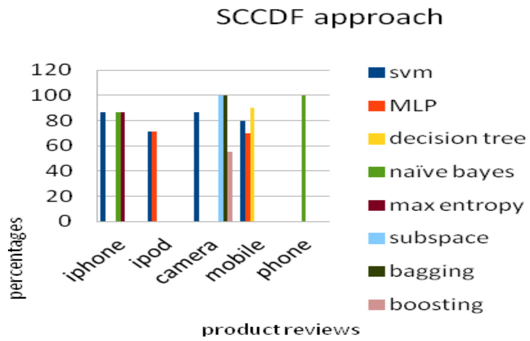

Figure 2: Comparision of product reviews with Classification algorithm.

V. CONCLUSION

This research work concentrated completely on analyzing product feature related reviews. Our system takes subjective phrases on product features and summarized the sentiments in to multiple different categories. The Proposed architecture improved the performance of stemmer algorithm by applying new rules to the original stemmer algorithm. Performance results of multiclass classification methodology are compared with other related research works which concentrated on product reviews and we achieved high percentage of accuracy. This work is useful for product development companies who intend to know the product features score given by users in each review for each feature and they can also view aggregated score values to analyze their performance.

VI. FUTURE WORK

This work can be further extended to analyze other than product domain data. Presently we took comments from product review sites like amazon.com so this work can be extended to do sentiment analysis of product reviews collected from social networking sites like facebook, twitter etc. Web crawling tools such as quick test professional (QTP) can be used to extract text efficiently from world wide web dynamically. Web crawling using QTP tool is a challenging task and our work unable to solve this problem which can be studied further. This work can be extended as a web service, which when requested can access different social networking sites for collecting reviews and summary can be displayed to the end user.

REFERENCES

[1] Changli Zhang,Wanli Zuo, Tao Peng, and Fengling He.Sentiment classification for chinese reviews using machine learning methods based on string kernel. In *Proceedings of the 2008 Third International Conference on Convergence and Hybrid Information Technology -Volume 02*, pages 909–914, 2008.

[2] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. Pages 1–6, 2009.

[3] Matthew Whitehead and Larry Yaeger Sentiment Mining Using Ensemble Classification Models Indiana University School of Informatics 2010.

[4] Akshat Bakliwal, Piyush Arora , Ankit Patil, Vasudeva Varma Towards Enhanced Opinion Classification using NLP Techniques Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP) 2011.

[5] Trivedi Khushboo N, Swati K. Vekariya Prof.Shailendra Mishra, Mining of Sentence Level Opinion Using Supervised Term

WeightedApproach of Naive Bayesian Algorithm Int.J.Computer Technology and Applications,Vol 3 (3), pages 987-991, 2012.

[6] B Jayanag, Dr. K.V.Sambasiva Rao.Sentiment Classification for Dynamic opinions using Natural Language Processing Published in International Journal of Engineering research & Technology vol 2 Issue12, December-2013.

[7] R C Balabantara,y Mudasir Mohammad, Nibha Sharma Multi-Class Twitter Emotion Classification: A New Approach. Published in International Journal of Applied Information Systems (IJAIS) – Foundation of Computer Science FCS, New York, USA Volume 4– No.1, September 2012 –48

[8] Zahan Malkani, Evelyn Gillie Supervised Multi-Class Classification of Tweets.

[9] http://sentiwordnet.isti.cnr.it

[10] Fadi Yamout, Rana Demachkieh, Ghalia Hamdan,Reem Sabra Further Enhancement to the Porter's Stemming Algorithm CandE American University I., Beirut, Lebanon 2004.

[11] http://nlp.stanford.edu/software/tagger.shtml

[12] Xiaowen Ding, Bing Liu, Philip S. Yu, A Holistic Lexicon-Based Approach to Opinion Mining University of Illinois at Chicago 2008.