Dual Sentimental Analysis On Product Reviews

Shilpa A B,Jisha G

Department of Information Technology, Rajagiri School of Engineering & Technology,Rajagiri Valley Road,
Rajagiri Valley, Chittethukara, Kakkanad, Kerala 682039,India
Corresponding Author: Shilpa A B

Abstract :*Online reviews available on the internet are increased in large amount in recent years. Sentiment analysis is a natural language processing that predict the mood of the public. It have a major role in marketing intelligence to analyses, visualize and predict the relations between products and customer reviews. It helps marketing enterprises to find the potential risks and to design new products and marketing strategies. Using this a customer can easily see how the existing customers feel about a product and product manufacturer can find what the customers thought about their product by investigating reviews, tweets, blogs etc. In Sentiment classification BOW model is the generally used text representation model that follow the techniques of traditional topic based text classification. The performance of BOW sometimes remains limited due to some fundamental deficiencies in handling the polarity shift problem. Due to polarity shift problem, accuracy of finding the sentiment of a statement decreases. Dual sentimental analysis is done by taking original and reversed training reviews for sentiment classification to avoid the shift polarity problem and to improve the accuracy to predict the sentiments. Dual sentimental analysis will predict the sentiment of product reviews efficiently. Marketing industries can use this technique to predict the costumer responses and their opinions regarding their product. This paper presents the techniques and methods in Dual sentiment analysis and how it can be used to predict the sentiment of product reviews.*
Keywords: *Dual Sentimental Analysis,DualTraining,DualPrediction,Classification,Accuracy*

I.   Introduction

Sentiment analysis have a major role in marketing intelligence to analyse, visualize and predict the relations between products from customer reviews. It helps marketing enterprises to find the potential risks and to design new products and marketing strategies. Using this a customer can easily see how the existing customers feel about a product. The product manufacturer can find what the customers thought about their product by investigating reviews, tweets, blogs etc. Sentiment classification is a basic task in sentiment analysis it classify the sentiment of a given text. BOW model is the generally used text representation model that follow the techniques of traditional topic based text classification. Due to polarity shift problem, accuracy of finding the sentiment of a statement decreases. Negation is the most important type of polarity shift [5] by adding a negation word to a positive text, the sentiment of the text will be reversed from positive to negative. The two sentiment-opposite text are considered to be very similar by the BOW representation. This is the main reason why standard machine learning algorithms often fail under the circumstance of polarity shift. Sentiment can be predicts with more accuracy by using dual data sets. In dual sentimental analysis system the classifier is trained by maximizing a combination of the original and reversed training samples. It classifies the test reviews by considering two sides of one review. Also extend the DSA framework from positive and negative polarity classification to positive, negative and neutral classification. Naïve bayes classification is consider for dual sentimental analysis and best products will be predicted from the product reviews using the best classifier algorithm. . This paper presents an idea to implement dual sentiment analysis and how it can be used to predict the sentiment of product.

II. Sentimental Analysis

The aim of Sentimental analysis is to identify the sentiments and then classify their polarity. In sentiment analysis two types of polarities are considered Positive polarity and Negative Polarity. Sentiment analysis is the measurement of positive language and negative language. An object which carrying the positive opinion comes under the positive polarity and which holds the negative opinion is negative polarity. Feature-based Sentiment classification find the opinions on features of certain objects. For example, in reviews related to mobile classifying the sentiments only on the basis camera quality.

III. Existing Systems

The existing systems contain one of the most well-known difficulties called polarity shift problem. Polarity shift is a kind of linguistic phenomenon which can reverse the sentiment polarity of the text. Negation is the most important type of polarity shift. For example, by adding a negation word "don't" to a positive text "I like this book" in front of the word "like", the sentiment of the text will be reversed from positive to negative. However, the two sentiment-opposite texts are considered to be very similar by the BOW representation. This is the main reason why standard machine learning algorithms often fail under the circumstance of polarity shift. Several approaches have been proposed in the literature to address the polarity shift problem. Also there is existing system that make use of dual sentimental analysis for market analysis.

IV. Proposed System

We propose a system that uses dual sentimental analysis to make prediction on product reviews. Dual sentimental analysis is done by dual training algorithm and dual prediction. It uses original and reversed samples to train the classifier .In Dual training process the classifier train maximum with original and reversed training data sets and by considering two sides of review prediction is made. In sentiment prediction stage user can search the opinion about a product and we get how good or bad the product is based on positive and negative confidence.

V. Dual Sentimental Analysis

Dual sentiment analysis (DSA) comes to address the problems like shift polarity problem and to improves the accuracy of sentiment classification. Dual sentimental analysis done with data expansion technique by creating a sentiment reversed review for each training and test review[4]. Using this we propose a dual training algorithm that uses   original and reversed training reviews in pairs for learning a sentiment classifier, and a dual prediction algorithm to classify the test reviews by considering two sides of one review.
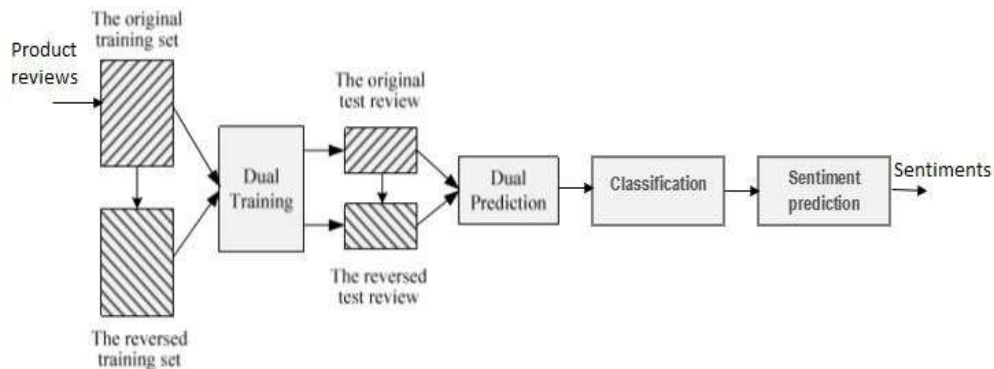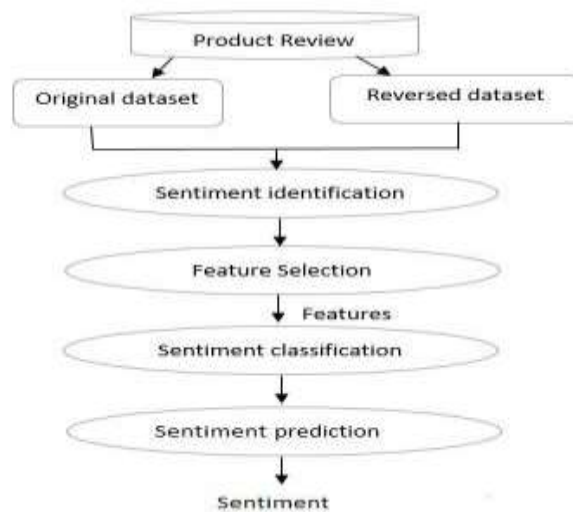

Fig 5.1. Architecture of dual sentimental analysis

fig 5.2. Dual Sentiment analysis process on product reviews

VI. Duel Training(DP)

Product reviews from world wide web, sites or web interfaces are given to training stage. Original and reversed training samples are created ie, original training set and reverse training set.For revers training set, data expansion technique is used .There is a one-to-one correspondence between the original and reversed reviews. The classifier is trained maximum by a combination of the likelihoods of the original and reversed training samples.ie, all of the original training samples are reversed to their opposites[1].
Process of Data expansion technique to create reverse review,
Text Reversion: All sentiment words out of the scope of negation are reversed to their antonyms. In the scope of negation, negation words (e.g., "no", "not", "don't", etc.) are removed, but the sentiment words are not reversed.
Label Reversion: For each of the training review, the class label is also reversed to its opposite (i.e., positive to negative, or vice versa), as the class label of the reversed review. Fig 5.3 showing how to creating the reversed training reviews,

	Review Text	Class
Original review	I like this movie. It is interesting	Positive
Reversed review	I don't like this movie. It is boring	Negative

Fig 5.3 Reversing the data set

For original training review 'I like this movie. It is interesting' is reversed by three steps,1) The sentiment word
"interesting" is reversed to its antonym "boring" 2) For the word like ,don't add to negate the word.3) Class label is reversed from positive to negative.

VII. Dual Prediction

In the prediction stage, for each test sample x, a reversed test sample($\sim$x) is created. our aim is not to predict the class of reversed test sample. But instead, we use $\sim$ x to assist the prediction of x[1]. This process is called dual prediction. In DP, predictions are made by considering two sides of a coin. When we want to measure how positive a test review x is, we not only consider how positive the original test review is but also consider how negative the reversed test review is. Also when we measure how negative a test review x is, we consider the probability of x being negative as well as the probability of $\sim$ x being positive. In traditional BOW, "like" will have a high positive score in predicting overall orientation of the test sample, despite of the negation structure "don't like". Hence test review will miss-classified as positive. This prediction error of original test sample is compensated by taking the dual prediction since it uses a weighted combination of original test sample and reversed test sample.

VIII. Features Selection

TF-IDF(Term Frequency–Inverse Document Frequency,), IG(Information Gain), MI(Mutual Information), Feature Vector , Count Vectorizer etc are various feature selection methods. TFIDF score is to be taken into consideration to balance most weighted and less weighted word. TF-IDF It is defined by multiplying value of frequency of word in review (TF) and frequency of word in whole corpus (IDF)[6]. Chi square method gives good result for both positive and negative class. Count Vector It is defined by number of occurrences of feature in review.

IX. Classification

Naïve bayes classification is used to predict the probability for a given tuple that belong to a particular class. Training and classifying steps of naïve bayes is easier when comparing with other classification algorithms. To create trained model, combined dataset (original and reversed data set) is given as input to train input set by naïve bayesclassifier.Trained model is then applied on test data to generate either positive or negative sentiment.

X. Sentiment prediction

Any queries or product related statements can be given as input to test the system and system predicts the chances of getting positive and negative feedback about the product.

XI. Experiment Result

In this section, we systematically evaluate dual sentimental analysis with sentimental analysis using naïve bayes classification fig 7.1 and fig 7.2.Feature extraction and Naive bayes classification model is implemented using Scikitlearn library, which is the most useful library for machine learning in Python. Accuracy of a model to make correct predictions from all predictions is classification accuracy [3]. It found that dual sentimental analysis showing high accuracy than sentimental analysis in terms of precision, recall, Fi-Score and support.

Precision: It is the ability of a classifier not to label as positive a sample that is actually negative TP/(TP+FP), TP is True positive and FP is false negative.
Recall : It is the ability of classifier to find all the positive samples TP/(TP+FN)
Fi score: Weighted average of precision and recall
Support: It is the number of occurrence of each class.

	precision	recall	f1-score	support
0	0.33	0.17	0.23	160
1	0.38	0.59	0.46	140
avg / total	0.35	0.37	0.34	300

Fig7.1 Sentimental Analysis

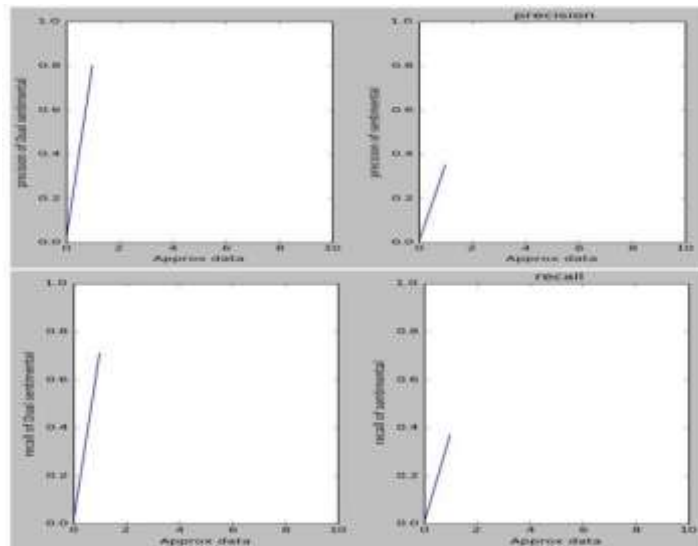|	precision	recall	f1-score	support
0	0.96	0.47	0.63	160
1	0.62	0.98	0.76	140
avg / total	0.80	0.71	0.69	300

Fig7.2 Dual sentimental analysis

Fig:7.3 Precision and recall comparison

From fig 7.3 dual sentimental analysis shows better result in terms of precision and recall .We get 80% precision and 71% recall for dual sentimental analysis for a test case .

A better model can also define in terms of loss value. The loss is calculated on training sets and validation and find out how well the model is doing for these two sets. Loss is the summation of the errors made for each example in training or validation sets. For a good system loss rate on each iteration will be less. Accuracy finds by fed the test samples to the model and number of wrong predictions are noted with respect to true predictions. Then loss rate is classified by taking percentage of misclassification. if the number of test

samples is 1000 and model classifies 952 of those correctly, then the model's accuracy is 95.2%.In fig7.4 shows that dual sentimental analysis shows less loss rate and improves accuracy
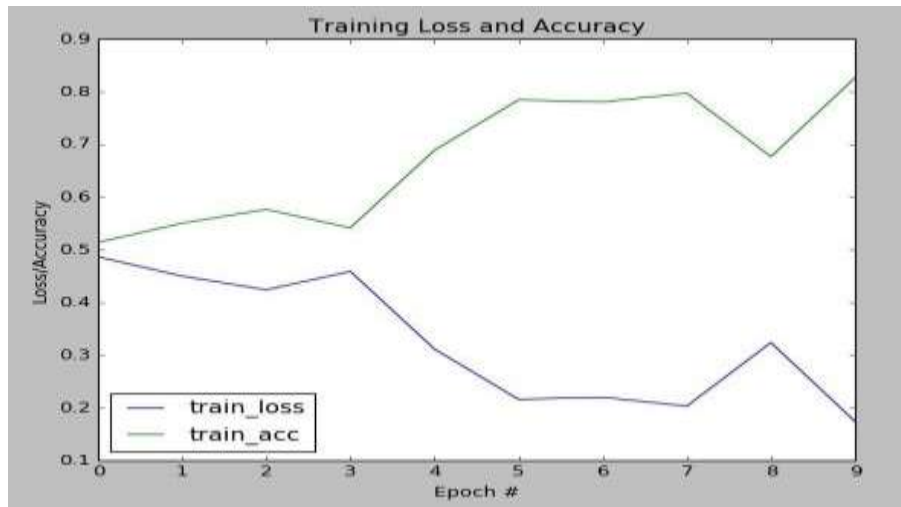.

Fig 7.4 Loss and accuracy rate of dual sentimental analysis

Dual sentimental analysis helps the marketing enterprises to find the potential risks and to design new products and marketing strategies. They can search their products and find what people think about their product by searching product and sentiments of a product is find by dual sentimental analysis. System predicts and states how much positive and negative the product is. Fig 7.5 and Fig 7.6 shows a test case, where we search for a product and our system predicts how good or bad the product is based on customers review. From fig 7.6 it shows that the product we search have a change to get 79% positive reviews and 21 % negative reviews
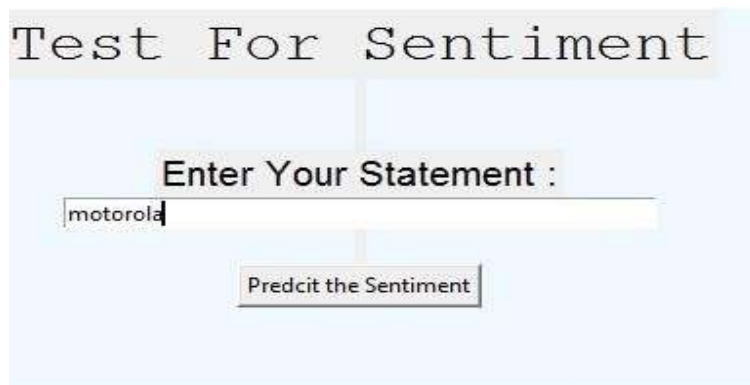

Fig 7.5 Testing DSA model

Fig 7.6 Sentiment classification result

XII. Conclusion

Dual Sentiment analysis aims to determine the opinion of people towards a particular product or about a company. Dual sentiment analysis address the polarity shift problem in sentiment classification. The basic idea of dual sentimental analysis is to create reversed reviews that are sentiment-opposite to the original reviews, and make use of the original and reversed reviews in pairs to train a sentiment classifier and make predictions. It is highlighted by the technique of one-toone correspondence data expansion and the manner of using a pair of

samples in training (dual training) and prediction (dual prediction).Accuracy and loss rate of dual sentimental analysis is calculated Performance of dual sentimental analysis is compared and dual sentimental analysis shows better results than sentimental analysis . Also product related statements can be given as input and system successfully predicts the chances of getting positive and negative feedback about the product.

References

[1]. X.Rui, X.Feng , Z.Chengqing, L.Qianmu ,Q. Yong,L.Tao,"Dual Sentiment Analysis: Considering Two Sides of One Review,"IEEE Transactions on Knowledge and Data Engineering, Volume: 27,Issue8,Aug. 2015.
[2]. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks,"IEEE Transactions on Knowledge and Data Engineering , vol. 23, no. 3, pp. 447-462, 2011.
[3]. M.K Fazal ,KAurangzeb,A. Shakeel, M Z Asghar,"Lexicon-Based Sentiment Analysis in the Social Web," Journal of Basic and Applied Scientific Research,May.2014.
[4]. R.J Bhagyashri, M. Manjushri," Review of Dual Sentiment Analysis," International Journal of Science and Research, Volume 4 Issue 11, Nov. 2015.
[5]. Y.R Mukta Y, A.K Mayura" Polarity shift in opinion mining," IEEE International Conference onAdvances in Electronics, Communication and Computer Technology ,Dec.2016.
[6]. B.Tang,K. Steven,H. Haibo,"Toward Optimal Feature Selection in Naive Bayes for Text Categorization",IEEE Transactions on Knowledge and Data Engineering, Vol.28, 2016.
[7]. http://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html
[8]. https://www.quora.com/What-are-the-best-supervised-learning-algorithms-forsentiment-analysis-in-text-Iwantto-do-text-analysis-to-determine-whether-anews-item-is-positive-or-negative-for-a-given-subject
[9]. https://www.quora.com/What-are-the-best-supervised-learning-algorithms-forsentiment-analysis-in-text-Iwantto-do-text-analysis-to-determine-whether-anews-item-is-positive-or-negative-for-a-given-subject
[10]. http://blog.aylien.com/naive-bayes-for-dummies-a-simple-explanation/
[11]. http://scikitlearn.org/stable/modules/generated/sklearn.linear_model.LogisticReg ression.html#sklearn.linear_model.LogisticRegression