



SENTIMENT ANALYSIS FOR TWO SIDES OF REVIEW USING DUAL PREDICTION

V.S.Swarnalekha,
¹PG Scholar,CSE
E.G.S.Pillay Engineering College,
Nagapattinam, India

V.M.Suresh
Assistant Professor
E.G.S.Pillay Engineering College,
Nagapattinam, India
sulekhasundarvs@gmail.com

ABSTRACT

Dual sentimental analysis is an important current research area. The sentiment found within comments, feedback or critiques provide useful indicators form any different purposes. We propose a dual training algorithm to make use of original and reversed training reviews in match for learning a sentiment classifier, and a dual prediction algorithm to classify the test reviews by examine two sides of one review. We also extend the DSA framework from polarity (positive-negative) classification to 3-class (positive-negative-neutral) classification, by taking the neutral reviews into consideration. Dual Sentiment analysis provides companies with a means to evaluate the extent of product acceptance and to determine strategies to increase the product quality. It also facilitates policy makers or politicians to Analyze public sentiments with regard to policies, public services or political issues.

KEYWORDS:- *Sentimental analysis, feedback, Opinion Mining, Classification, Sentiment Identification.*

INTRODUCTION

Data Mining is one of the important step of the “Knowledge Discovery in Databases” processes or KDD, which relevance patterns from large datasets. Also, it includes the techniques and design of artificial intelligence, machine learning and statistics. Opinion mining or sentiment analysis is to evaluate the users’ opinions or thoughts which are in the form of unstructured data. To interpret and understand the person's views, emotions and understanding, the system must be made reliable and efficient. There are two techniques used in Sentiment Classification.

- Machine Learning Approach
- Lexicon Based Approach

Machine Learning techniques include supervised and unsupervised learning approaches. Supervised learning consists of some classifier such as Decision tree, Liner, Rule-based and Probabilistic classifiers.

Lexicon Based approaches are confidential into Dictionary based and Corpus-based methods. The corpus-based method further divided into the statistical and semantic approach.

BOW MODEL

Sentiment classification is a essential task in sentiment analysis, with its aim to classify the sentiment of a given text. The familiar practice in sentiment classification follows the techniques in conventional topic-based text classification, where the bag-of-words (BOW) standard is

typically used for text representation. In a review text is represented by a vector of independent words. A large number of researches in sentiment analysis aimed to appreciate BOW by consolidating linguistic knowledge [6], [10].

POLARITY SHIFT

Polarity shift is a kind of linguistic circumstance which can reverse the sentiment polarity of the text. Negation is the most essential type of polarity shift. For example, by adding a negation word “don’t” to a positive text “I like this book” since the word “like”, the sentiment of the text will be reversed from positive to negative. Anyhow, the two sentiment-opposite texts are acknowledged to be very similar by the BOW representation. The main reason why standard machine learning algorithms often decline under the circumstance of polarity shift. Several methods have been proposed in the literature to indicate the polarity shift problem [10], [11].

DUAL SENTIMENT ANALYSIS

Effective way to handle this is a simple yet enhanced model, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification. The original and reversed reviews are designed in a one-to-one correspondence. In DSA a dual training (DT) algorithm and a dual prediction (DP) algorithm appropriately, to make use of the original and reversed samples in pairs for training a statistical classifier and make predictions. Also DSA framework is implemented for 3-class (positive-negative-neutral) sentiment classification, by getting the neutral reviews into consideration in both dual training and dual prediction [1].

Different Classes of Sentiment Analysis

a. **Positive Sentiments:** These are the good words about the target in consideration. If the positive sentiments are raised, it is indicated to be good. In case of commodity reviews, if the positive reviews about the commodities are more, it is bought by many customers.

b. **Negative Sentiments:** These are the bad words about the target in deliberation. If the negative sentiments are increased, it is discarded from the optional list. In case of commodity reviews, if the negative reviews about the commodities are more, no one intend to buy it.

c. **Neutral Sentiments:** These are neither good nor bad words about the intention. Hence it is neither adopted nor depreciates.

II.RELATED WORKS

D. Rui Xia et al. [1] have performed the task of tackling the polarity shift problem. Here the polarity shift causes the negation of the statement. In Bag-of-words technique, two sentiments opposite texts are considered to be very similar which causes the polarity shift. Today most of the researchers use BOW way for sentiment analysis. They have proposed the dual sentiment analysis (DSA) model to solve the polarity shifting. The data is expanded by creating the reversed review for each training and test review.

Shulong Tan et al.[2.] have proposed LDA based two models to interpret the sentiment variations on twitter. Twitter sentiment analysis is an important research area for academic as well as business fields for decision making like for the seller to decide if the product should be produced in a large quantity as per the buyers feedback and for the students to decide if the study material to be referred or not. Distilled out the foreground topics effectively and removed the noisy data accurately. Found the exact reasons behind sentiment variations on twitter data using RCB-LDA model which is very useful for decision making.

Y.Hu et.al [3] has proposed a joint Bayesian model ET-LDA that is Event- Topic LDA which performs the task of topic modeling and event apportionment so as to carry out sentiment analysis significantly and qualitatively. Here consideration two large scale data sets from two

different domains combined with two events. The work done here is most effective for topic modeling because the topic may consists of many passages where the tweet may belong to specific event in a paragraph or general event in the topic

Abbasi et al. [4] Attributes for Sentiment Classification Using FRN is a rule-based multivariate text feature selection method that considers semantic information and also leverages the syntactic relationship between n-gram features. FRN was able to select attributes resulting in significantly greater classification accuracy regardless of the feature subset sizes. FRN's use of syntactic relation and semantic information about n-grams empowered it to achieve improved results over various univariate, multivariate and hybrid feature selection methods.

C. Lin, et al[5] proposes a model for weakly supervised sentiment analysis called Joint Sentiment-Topic which is probabilistic modeling framework for detection of sentiment from Text and Reverse-JST based on latent Dirichlet allocation which detects sentiment and topic simultaneously from text. Weakly supervised nature of JST makes it highly portable to other domains and JST model achieved either better or comparable performance.

Yong [6] proposed a Tweet Sentiment Analysis Model evaluation is done on political dataset. The proposed model provides fast and less expensive different approach over traditional polls for mining public views.

Hussain in [7] proposed a hybrid method of SVM and Particle Swarm Optimization. To solve the problem of dual optimization, hybrid method is proposed. For extraction of more features, from prepared datasets is used. The efficiency of SVM is affected by the hybridization of SVM-PSO is used. The accuracy of SVM is damaged by the hybridization of SVM-PSO and which is 77%.

Rajasri in [8] analyzed the Twitter posts about electronic commodity like mobiles, laptops, etc.

using Machine Learning approach. A new feature vector is refined for classifying the tweets as positive, negative and extracts peoples' opinion about products. Classifiers like Nave Bayes, SVM, Maximum Entropy and Ensemble classifiers are used for classification. This feature vector performs well for electronic products.

Danny in [9] proposed an aspect-based sentiment classification approach to analyzing opinions for tweets. Different experiments are implemented on the aspect-based sentiment classifier, to improve the performance of existing classifiers. From experimental results, it proposes that a layered classification approach that uses the aspect-based classifier as the first layer classification and the tweet-level classifier as the second layer classification is more efficient than a classifier trained using target-dependent features

III.PROPOSED DESIGN

Propose a simple yet efficient model, called dual sentiment analysis (DSA), to address the polarity shift problem in sentiment classification **Figure 3.** Show a System of proposed architecture. A data expansion technique by creating sentiment-reversed reviews. The original and reversed reviews are established in a one-to-one correspondence & dual training (DT) algorithm and a dual prediction (DP) algorithm correspondingly, to make use of the original and reversed specimen in pairs for training a statistical classifier and make predictions. DSA structure from polarity (positive-negative) classification to 3-class (positive, negative, neutral) sentiment classification. To curtail DSA's dependency on an external antonym dictionary, we finally develop a corpus-based method for construct pseudo-antonym dictionary. The pseudo antonym dictionary is language-independent and domain adaptive makes DSA model possible to be applied into a wide range of applications.

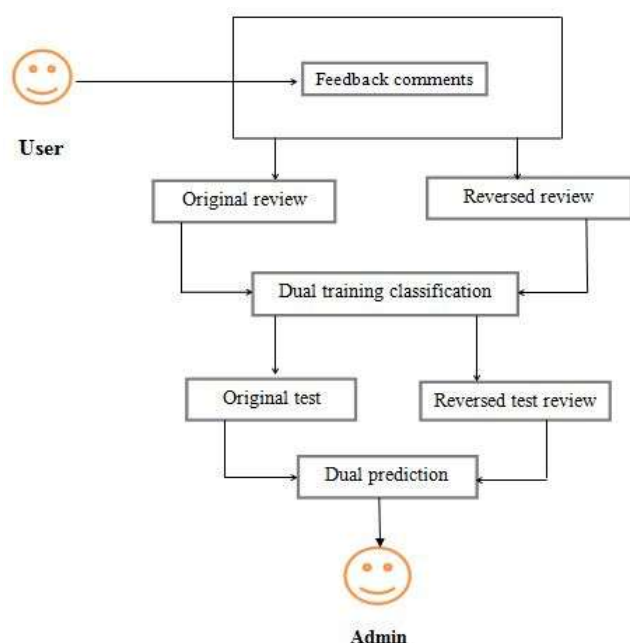


Fig 3.1 System Architecture

USER FEEDBACK COMMENTS

Our analysis of feedback comments on eBay and Amazon divulge that even if a buyer gives a positive rating for a transaction, s/he still consent comments of mixed opinions respecting different aspects of transactions in feedback comments. For example for comment c2, a purchaser gave a positive feedback classification for a transaction, but left the subsequent comment: “bad communication, will not purchase from repeatedly. super slow ship(ping). item as described.”. Apparently the buyer has negative opinion almost the communication and delivery aspects of the transaction, against an overall positive.

REVIEW CONVERSION

The data expansion technique of creating sentiment-reversed reviews based on an antonym dictionary, for each original review, the reversed review is established conferring to the following rules:

Text reversion: If there is a negation, first detect the purview of negation. All sentiment words out of the scope of negation are reversed to their antonyms. In the outlook of negation, negation

words are removed, but the sentiment words are not reversed.

Label reversion. For any of the training review, the class design is also reversed to its opposite (positive to negative, or vice versa), just as the class label of the reversed review.

DUAL TRAINING

The original training specimens are reversed to their opposites. Indicate to them as “original training set” and “reversed training set. In our data expansion technique, there is a one-to-one correspondence among the original and reversed reviews. The classifier is trained by maximizing a combination of the likelihoods of the original and reversed training samples. This process is called dual training. Note that our method can be easily adapted to the other classifiers such as naïve Bayes and SVMs.

DUAL PREDICTION

Dual prediction works in addressing the polarity shift problem. This time we think “I don’t like this book. It is boring” is an original test review, and “I like this book. It is interesting” is the reversed test review. Accordingly, it is very likely that the original test review will be misclassified as Positive. While in DP, due to the removal of negation in the reversed review, “like” this time the plays a positive role. Therefore, the probability that the reversed review being classified into Positive must be high. In DP, a weighted combination of two component predictions is used as the dual prediction output.

IV. EXPERIMENTAL ANALYSIS

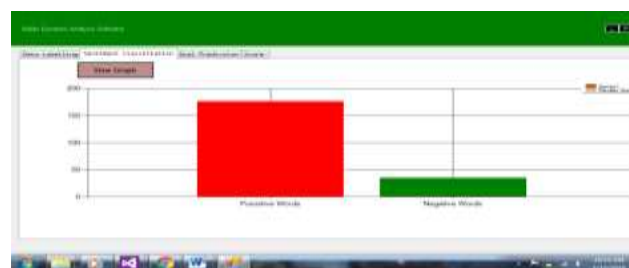


Fig 4.1 Sentiment classification



Fig 4.2 Dual prediction

V CONCLUSION

In this work, we propose a novel data expansion approach, to address the polarity shift problem in sentiment classification. The manner of using a pair of samples in training (dual training) and prediction (dual prediction). A DSA model is very effective for polarity classification and it significantly outperforms several alternative methods of considering polarity shift. DSA algorithm by developing a selective data expansion technique that chooses training reviews with higher sentiment degree for data expansion.

REFERENCES

- [1] R. Xia, F. Xu, C. Zong, Q. Li, Y. Qi and T. Li, "Dual sentiment analysis: Considering two sides of one review", in IEEE Trans. Knowl. Data Eng., vol. 27, no. 8, pp. 2120-2133, Aug. 2015.
- [2] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Interpreting the Public Sentiment Variations on Twitter, IEEE Transactions on Knowledge and Data Engineering, VOL. 26, NO.5, MAY 2014.
- [3] Y. Hu, A. John, F. Wang, and D. D. Seligmann, Et-lda: Joint topic modeling for aligning events and their twitter feedback, in Proc.26th AAAI Conf. Artif. Intell. Vancouver, BC, Canada, 2012.
- [4] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Trans. Knowl. Data Eng., vol. 23, no. 3, pp. 447-462, Mar. 2011.

- [5] C. Lin, Y. He, R. Everson, and S. Ruger, "Weakly supervised joint sentiment-topic detection from text," IEEE Trans. Knowl. Data Eng., vol. 24, no. 6, pp. 1134-1145, Jun. 2012.

- [6] Xujuan Zhou, Xiaohui Tao, Jianming Yong, "Sentiment Analysis on Tweets for Social Events", 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design, 978-1-4673-6085-2/13/\$31.00 ©2013 IEEE.

- [7] Abd. Samad Hasan Basaria, Burairah Hussina, I. Gede Pramudya Anantaa, Junta Zeniarjab, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", 1877-7058 © 2013 The Authors. Published by Elsevier Ltd.

- [8] Neethu M S, Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", 4th ICCCNT 2013, July 4 - 6, 2013, Tiruchengode, India.

- [9] Hsiang Hui Lek and Danny C.C. Poo, "Aspect-based Twitter Sentiment Classification", 2013 IEEE 25th International Conference on Tools with Artificial Intelligence.

- [10] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Comput. Intell., vol. 22, pp 110-125, 2006.

- [11] M. Gamon, "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis," in Proc. Int. Conf. Comput. Linguistics, 2004, pp. 841-847.