

Sentiment Analysis For Two Sides Of Reviews Using Dual Prediction Algorithm

G.Grace Ranjitham¹

M.E., Computer Science Student,
M.I.E.T Engineering College, Trichy,
grace.georgekumar@gmail.com

S.Mohana²

Assistant Professor, Department of CSE,
M.I.E.T Engineering College, Trichy
mohana.p3@gmail.com

B.Vinothini³

M.E., Computer Science Student,
M.I.E.T Engineering College, Trichy
vinobasker04@gmail.com

Abstract—The performance of the Bag Of Words are limited due to fundamental deficiencies in handling polarity shift problem, the two opposite texts are considered to be very similar by the bag of words representation .To overcome this problem, a model called dual sentiment analysis(DSA) is proposed. Dual Sentiment Analysis extends the analysing of sentiment classification from one side of review in to two sides of one review. A data expansion technique is proposed to address sentiment reversed review for both training and test review. Dual training algorithm is used to train the original and reversed review in sentiment classifier. Dual prediction algorithm works in addressing the polarity shift problem by classifying the test reviews of both sides. The Dual Sentiment Analysis framework extends the two types of polarity classification (positive and negative) in to three types of classification (positive, negative and neutral). To remove dual sentiment analysis dependency on an external antonym dictionary for review reversion, corpus-based method is developed to construct a pseudo-antonym dictionary. The results demonstrate the effectiveness of DSA in supervised sentiment analysis.

Keywords :- Polarity shift, opinion mining, Sentiment identification, sentiment analysis.

I. INTRODUCTION

The important clue behind the opinion mining is “What people think?”. In the part, when an individual needed to make a decision they commonly asked for opinions from friends and family. When an organization needs to find opinions of the public about its product and services, they will conduct a survey. With the explosive growth of the social media on the internet in the past few years, the world has been transformed E commerce sites, online sites, web logs, product rating sites, discussion groups are some of the sources on which people can express their views on almost anything in discussion.

Sentiment analysis is to extract the opinion of the user from the text document and also it identifies the orientations of the opinions from the text. Sentiment analysis and opinion mining is a special text mining task for obtaining the

subjective attitude (i.e., sentiment) expressed by the text, it is becoming hotspot in the field of data mining and natural language processing. It also used widely in online shopping.

The semantic classification of a review can be positive, negative, or neutral. To classify the sentiment (eg.positive, negative or neutral) for a given text, the bag of words (BOW) model is typically used for text representation. The text in the review is represented by a vector of independent words to train a sentiment classifier in the BOW. The BOW model is simple and quite efficient, but it is not suitable for sentiment classification, because it breaks the word order, disrupts the semantic structures and discards some syntactic information. One of the well-known difficulties is the polarity shift problem. Polarity shift is a type of linguistic sensation which provides the opposite result of the sentiment polarity of the text. Negation is one of the most important types of polarity shift. For instance, “I don’t like this mobile”, in which eliminate the negation word to a given negative text in front of the word “like”, the sentiment polarity of the given text will be reversed from negative to positive. In the BOW representation, the two sentiment reversed texts are considered to be very similar. This is the main reason why standard machine learning algorithm often fails under the circumstance of polarity shift.

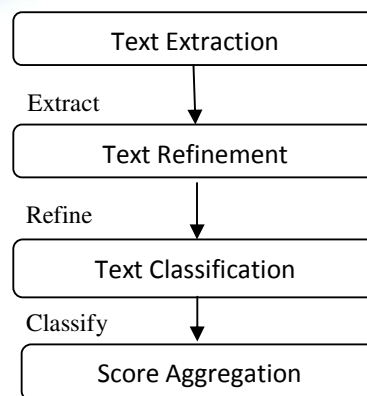


Fig .1.Sentiment analysis

A. Related Works

Let us discuss about the different classifier of sentiment analysis. The sentiment mining tasks are classified in to four categories such as document level sentiment analysis, sentences level sentiment analysis, phrase level sentiment analysis and aspect level sentiment analysis. 1. Document level sentiment analysis: The specific topic is considered for the single review in this document level classification. There are two types of method in document level classification, such as supervised learning method and unsupervised learning methods. The algorithms like support vector machine, naive bayes are used to train the system in the supervised learning method. 2. Sentence level sentiment analysis, in which the polarity of every sentence is calculated. It predicts the both subjective sentences and objective sentences. The opinion words in the subjective sentences helps to determine the sentiment about an entity. 3. Phrase level sentiment analysis: It determines whether an expression is inactive or polar and then predicate the polarity of the polar expressions. In this approach, the contextual polarity is identified for a large subset of sentiment expressions by using this system. 4. Aspect level sentiment analysis: It is also called as feature based sentiment classification. For example, let's consider the following "The camera clarity of MOTO mobile is highly developed", at this juncture the undreamt gesture out is "snap clarity". This sort of sentiment classification is added efficient.

Sentiment analysis has two types which are supervised approach and unsupervised approach. Supervised approach is called machine learning. Christo Ananth et al. [3] proposed a system in which the cross-diamond search algorithm employs two diamond search patterns (a large and small) and a halfway-stop technique. It finds small motion vectors with fewer search points than the DS algorithm while maintaining similar or even better search quality. The efficient Three Step Search (E3SS) algorithm requires less computation and performs better in terms of PSNR. Modified objected block-base vector search algorithm (MOBS) fully utilizes the correlations existing in motion vectors to reduce the computations. Fast Objected - Base Efficient (FOBE) Three Step Search algorithm combines E3SS and MOBS. By combining these two existing algorithms CDS and MOBS, a new algorithm is proposed with reduced computational complexity without degradation in quality.

Natural language processing (NLP) is one of the unsupervised approaches. NLP is capable to extract the meaning automatically from natural language sentences. It is a field of Artificial Intelligence. It makes use of entities and semantic patterns in the text to understand its meaning. The combination of language dictionaries, linguistic constructs like part of speech and noun phrases are used in the NLP. The major advantage of rule based methods is that it provides independence for the rule developers to use their domain knowledge is for analysis purpose with the help of rules. It doesn't require any training data. The drawbacks of NLP approach is that they require a help from the human in

generating the rules and it entirely depend on the domain awareness of rule developers.

B. Contributions

In this paper, the previous works are extended in three major aspects. First, the DSA algorithms strengthen by adding a selective data expansion procedure. The hand writing recognition is also uses the data expansion technique, where some syntactic training data is included to increase the performance of the handwriting recognition systems. Second, it widens the DSA framework from sentiment polarity classification in to positive, negative, and neutral sentiment classification. Third, a corpus based method is anticipated to develop a pseudo antonym dictionary that could remove DSA dependency on an external antonym dictionary. This corpus based pseudo-antonym dictionary can be learnt using the labeled training data only. The basic idea is to first use mutual information(MI) to identify the most positive relevant and the most negative relevant features, rank them in two separate groups and pair the features that have the same level of sentiment strength as pair of antonym words. In information theory, the mutual data of two random variables is a measurement of the inter relationship of the two random variables. In text categorization and sentiment classification the MI is used as a feature selection method. DSA algorithm is used to address the problem of polarity shift. DSA consist of dual training algorithm and dual prediction algorithm. Dual training algorithm is proposed to make use of original and reversed training reviews in pairs of learning a sentiment classifier and a dual prediction algorithm is used to categorize the analysis reviews by taking in to considering two sides of one review.

II. SYSTEM MODEL

DSA is proposed to predict the sentiment of two sides of a review. The feedback/comments of the users about various products are recorded in the online web sites. The feedback comments are taken in to two stages called original review and reversed review. The dual training algorithm are classified in to two types such as original training sample and reversed training sample. Calculate feature weight using dual prediction algorithm. In the prediction stage, for each test sample x , a reversed test sample $\sim x$ is created. $\sim x$ is to assist the prediction of x . Finally, sentiment for the specified data set is predicted.

A. User Feedback Comments

The analysis of the feedback comments on the different e-commerce web sites reveals that even if a buyer gives a positive opinion for a product, but they still leaves the comments of mixed opinions regarding diverse aspects of transactions in feedback. For example , a buyer gives a positive feedback rating for a product, excluding the subsequent statement: "Poor transaction, I won't buy from again, awfully slow in delivering the item". By this comment

we can make sure that the buyer has negative opinion towards the communication and delivery facets of the transaction, although an overall positive feedback rating towards the product.

B. Review conversion

The sentiment reversed reviews are created in the data expansion technique based on an antonym dictionary; the reversed review is created for each original review according

C. Dual training

Dual training algorithm is to make use of original and reversed training reviews in pairs for learning a sentiment classifier. The original training reviews are reversed to their opposites. It named as the original training set and reversed training set. Original training set consist of original set of data and the reversed training set consist of reversed set of data i.e., opposite sentiment is created for original data which is

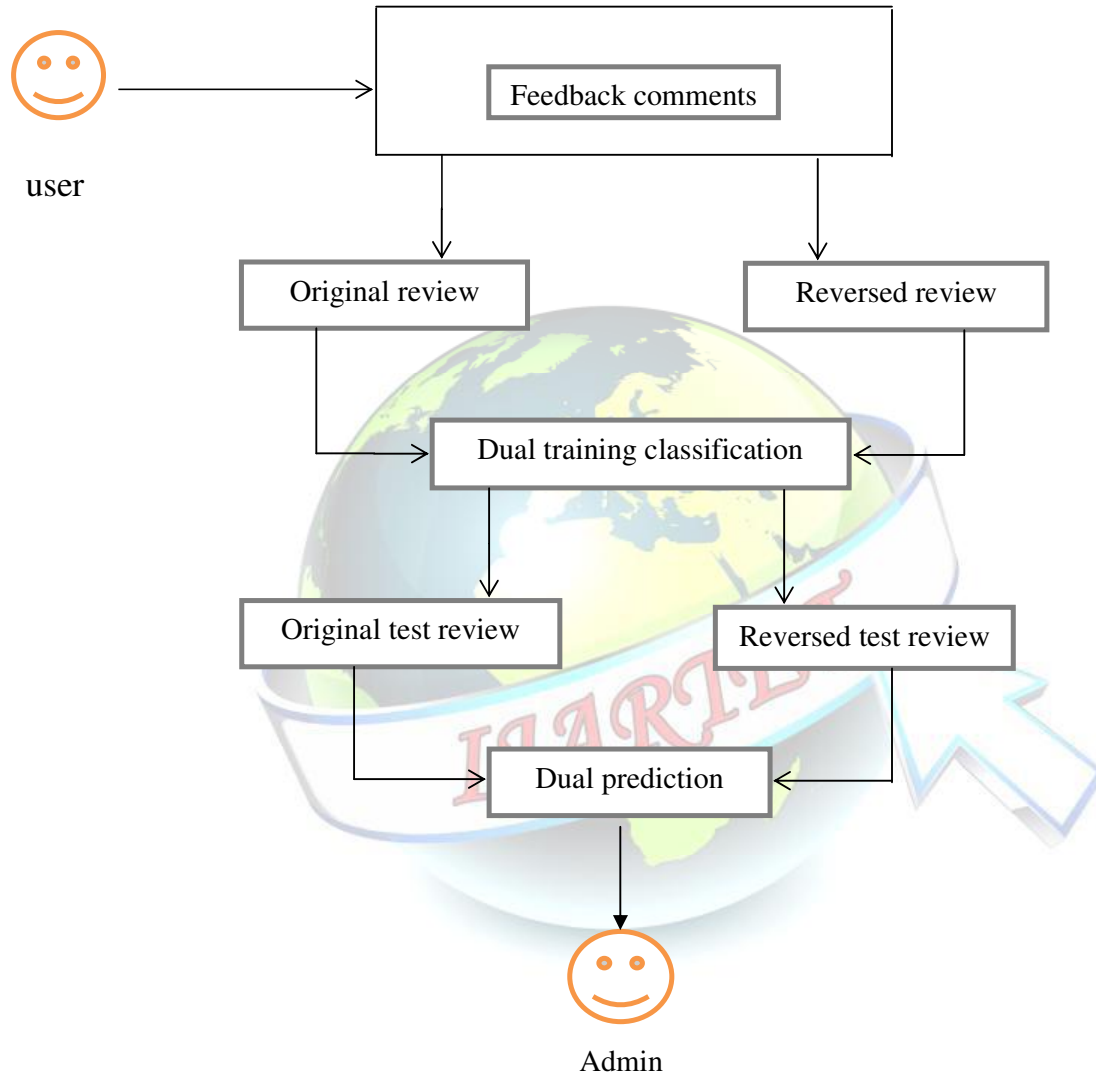


Fig.2.System architecture

to the following rules such as text reversion and label reversion 1.Text reversion: If there is a negation, first it detects the scope of negation. Except the negation sentences all the remaining words in the sentences are reversed to their antonyms. In text reversion the negation words like no, not, don't, etc. are removed, but the sentiment words are not reversed in the presence of negation in the sentence 2. Label reversion: For each and every training set of review, the class label is also reversed, i.e., positive to negative, or negative to positive. The label of the reversed review is removed, but the sentiment words are not reversed.

used to predict the sentiment more accurately. In our data expansion technique, there is a one-to-one correspondence between the original and reversed reviews. The classifier is trained by maximizing a combination of the likelihoods of the original to the removal of negation in the reversed review this process is called dual training. It not only measure how positive the training set is, it also determines how negative the training set sample is. For simplicity, in this paper we derive the DT algorithm by using the logistic regression model as an example. This method can also be used in other classifiers such as SVM and naive bayes.

D. Dual prediction

Dual prediction works in addressing the polarity shift problem. This time we assume “I am not fond of this glossy magazine, it is tedious” is an unique test review, and “I am fond of this glossy magazine, it is interesting” is the reversed test review. In traditional BOW, “like” will contribute a high positive score in predicting overall orientation of the test sample, despite of the negation structure “don’t like”. Hence, 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” will be considered as positive role. Therefore, the probability that the reversed review being classified in to positive must be high, it measures both the positive and negative aspect of the original and reversed review. In DP, a weighted combination of two component predictions is used as the dual prediction output. In this manner, the prediction error of the original test sample can also be compensated by the prediction of the reversed test sample. Apparently, this can reduce some prediction errors caused by polarity shift and it rapidly increases the accuracy of the performance of the analysis of sentiment.

III . EXPERIMENTAL ANALYSIS

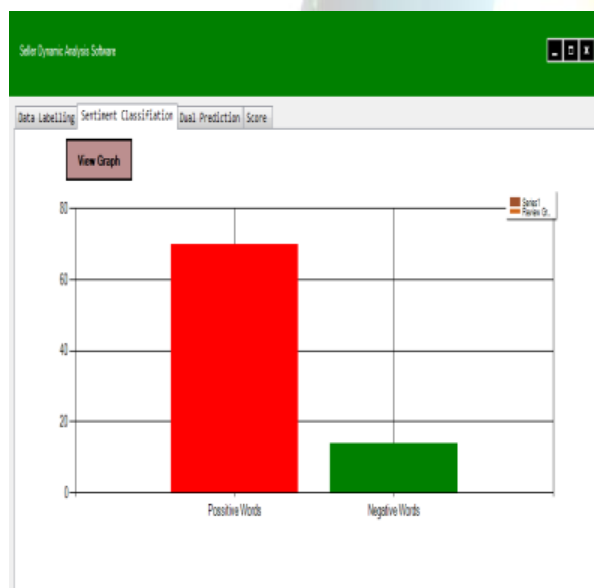


Fig 3. Sentiment Classification

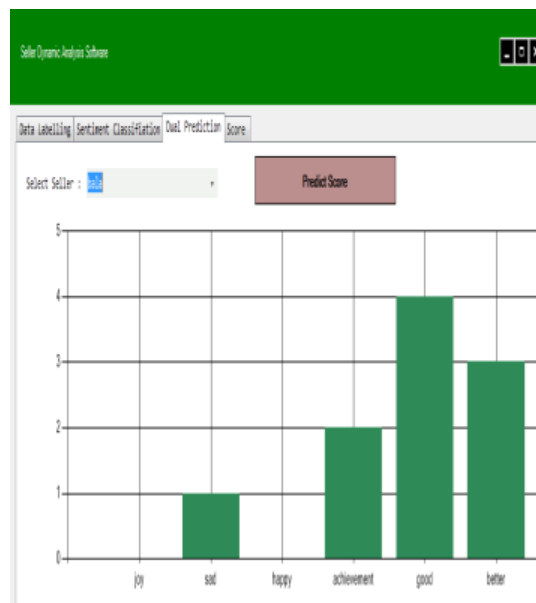


Fig.4. Dual prediction

III. PERFORMANCE EVALUATION

The performance of DSA is reported by selecting an increasing percentage of training reviews for data expansion. The prediction of sentiment on the multi-domain sentiment datasets by using logistic regression classifier and bigrams features. The predictive results are in terms of percentage. When the percentage is 0, no training samples are used for data expansion. In this case, DSA equals the standard baseline system. When the percentage is 1, all of the training samples are used for data expansion.

IV. CONCLUSION

Sentiment analysis is essential for anyone who is going to make a decision. Sentiment analysis is helpful in diverse field for measuring, recognizing and conveying sentiment. It is helpful for everyone when they want to buy a product and they can decide which product is best. Sentiment analysis is very important for Enterprises and helps them to know what customers think about their products. Therefore companies can take decisions about their products based on customer’s feedback. Thus companies can modify their products features and introduce new products according to customer’s opinions in a better and faster way. By using the DSA algorithm to make the sentiment analysis in more effectively.

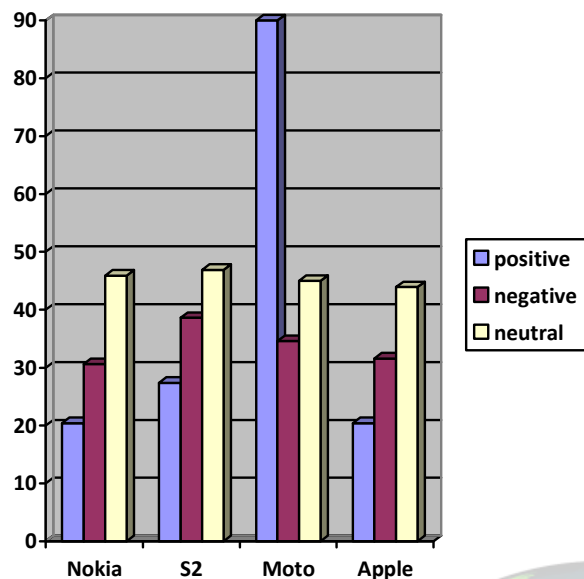


Fig.5. Sentiment classification using DSA

V. FUTURE ENHANCEMENT

In future, extends the DSA algorithm to a wider range of sentiment analysis tasks. It considers more complex polarity shift patterns such as transitional, sub-junctive and sentiment inconsistent sentences in creating reversed reviews. Interference effects on various storage devices.

REFERENCES

- [1] C. Yang, Z. Hai, K. Chang and J. Kim (2014), "Latent Aspect Rating Analysis without Aspect Keyword Supervision", IEEE Trans. Knowl. Data Eng., vol. 26, no. 3, pp. 447–462.
- [2] C. Huang S. Li, Y. Lee, Y. Chen and G. Zhou (2015), "Sentiment classification on customer feedback data: Noisy data, large feature vectors, and the role of linguistic analysis ", in Proc. Annu. Meeting Assoc. Comput. Linguistics.
- [3] Christo Ananth, A.Sujitha Nandhini, A.Subha Shree, S.V.Ramyaa, J.Princess, "Fobe Algorithm for Video Processing", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering (IJAREEIE), Vol. 3, Issue 3, March 2014 , pp 7569-7574
- [4] D. Inkpen and A. Kennedy (2010), "Sentiment classification of movie reviews using contextual valence shifters" Comput. Intell., vol. 22, pp. 110–125.
- [5] D. Moldovon and R. Mihalcea (2011), "Automatic detection of opinion bearing words and sentences", in Proc. Annu. Meeting Assoc. Comput. Linguistics.
- [6] S. Orimaye, A. Alhashmi and E. Siew. (2015), "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", Comput. Intell., vol. 24, pp. 110–125.
- [7] X. Yu, Y. Liu, X. Huang and A. An (2012), "Mining online reviews for predicting sales performance: A case study in the movie domain", IEEE Trans. Knowl. Data Eng., vol.24, no.4, pp.720–734.
- [8] X. Zhu, L. Zhuang and F. Jing (2014), "Movie review mining and summarization", in proc.15th ACM Inf. Knowl. Manage., pp. 43-50.
- [9] Y. Yang and X. Liu, "A re-examination of text categorization methods", in Proc. ACM SIGIR Conf. Res. Development Inf. Retrieval, 2007, pp. 42-49.
- [10] Y. Chen, S. Li, S. Lee, C. Huang and G. Zhou, "Sentiment classification and polarity shifting", in Proc. Int. Conf. Comput. Linguistics, 2010, pp.635–643.
- [11] Z. Hai, K. Chang, J. Kim, and C. C. Yang, "Identifying features in opinion mining via intrinsic and extrinsic domain relevance", IEEE Trans. Knowl. Data Eng., Vol. 26, no. 3, pp. 447-462, Mar. 2014

AUTHORS



G. Grace Ranjitham, she has received B.Tech Information technology degree from Anjalai ammal mahalingam engineering college, kovilvenni in 2013 and currently pursuing ME Computer Science and engineering degree in M.I.E.T engineering college, Trichy. Email: grace.georgekumar@gmail.com



S. Mohana, she has received B.E Computer science and engineering degree from SRMeaswari engineering college, Chennai in 2001 and she received ME Computer science and engineering degree in PABCET, Trichy in 2008. Currently pursuing Ph.D in Anna university. Email: mohana.p3@gmail.com



B. Vinodhini, she has received B.E Computer Science and Engineering degree from Anna university, Trichy in 2014 and currently pursuing ME Computer Science and Engineering degree in M.I.E.T engineering college, Trichy. Email: vinobaskar04@gmail.com

