International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 12 | Dec -2016 www.irjet.net p-ISSN: 2395-0072

A Review on Analysis and Classification of Sentiments using Dual Sentiment Filtration

Ms. Pooja J. Biradar[1], Dr. K. V. Metre[2]

[1]ME Student, Dept. Of Computer Engineering, MET's BKC Institute of Engineering, Adgaon, Maharastra, India
[2] Professor, Dept. Of Computer Engineering, MET's BKC Institute of Engineering, Adgaon, Maharastra, India

---

Abstract - *In past few years there is rapid growth in internet content. People are interested to express their opinion on any topic. It is difficult to search opinions and notice them and get them analyzed on web as there is availability of large amount of content on web. Also there exists a problem of polarity shifting which changes statement orientation of given texts. For the sentiment analysis the text is to be modeled in statistical machine learning approaches for which one popular technique used is Bag-of-words (BOW). Bag-of-words is very simple and quite efficient in some type of text classification but because of some basic limitations which it has while handling the polarity shift problem, the performance of BOW sometimes remains limited. Then there was also a difficulty while handling more complex polarity shift patterns. For addressing the problem of text classification a model called Dual Sentiment Analysis (DSA) is proposed. In this, firstly a new technique for data expansion is proposed which is generated by reversing the sentiment review for each of the training and test review. On this basis, two algorithms are proposed called Dual Training Algorithm and Dual Prediction Algorithm. The Dual Training Algorithm is one which uses both the original and reversed training reviews in pairs for sentiment classifier's learning. The other, Dual Prediction algorithm classifies the test reviews by considering two sides of one review. The DSA framework is also extended from polarity classification i.e., positive-negative to the 3-class classification i.e., along with positive, negative the neutral class also. Neutral is also considered along with the existing two positive and negative. It is done by taking the neutral review in consideration. And then finally, for removing the DSA Framework's dependency on external dictionary for reversing reviews, one corpus-based method is proposed to construct the pseudo antonym dictionary. Along with DSA the proposed system is also going to work towards the use of syntactic construction as features for inconsistency classifier that can help to improve performance.*

Keywords: **Natural language processing, machine learning, sentiment analysis, opinion mining, BOW (Bag of Word)**

1. INTRODUCTION

In today's e-business world or in competitive structure of market, fine analyzed data is required for betterment of services, probability calculations, predictions, business decisions and summary of market reputation etc. This analysis is done through the detail summary of product reviews etc. To analyze such kind of data, opinion mining techniques and sentiment classifications are used. As more and more common users become comfortable with the Internet, an increasing number of people are writing reviews. It has become necessity to analyze these reviews. The reviews are the sentiments of the users, so Sentiment Analysis is becoming an important work in data mining field. Sentiment classification is a special task of text classification whose objective is to classify a text according to the sentimental polarities of opinions it contains e.g., *favorable* or *unfavorable*, *positive* or *negative*. This task has received considerable interests in the computational linguistic community due to its potential applications. That is, Sentiment classification is the major and basic task in Sentiment analysis for classifying the given text i.e., positive or negative. The techniques which are generally used in traditional topic-based text classification are also followed by the Sentiment classification. Then the statistical machine learning algorithms are employed to train a sentiment classifier. The statistical machine learning algorithms are naïve Bayes, maximum entropy classifier and support vector machines. The most popular text representation model in machine learning based sentiment classification is known as the bag-of-words (BOW) model, where a piece of text is represented by an unordered collection of words, based on which standard machine learning algorithms are employed as classifiers. Although the BOW model is simple and has achieved great successes in topic-based text classification, it disrupts word order, breaks the syntactic structures and discards some kinds of semantic information that are possibly very important for sentiment classification. Such disadvantages sometimes limit the performance of sentiment classification systems. There are several approaches that exist as a solution to the polarity shifting problem. Polarity shifting is the problem which affects the performance of sentiment classification. DSA i.e. Dual Sentiment Analysis can be effective technique to address the problem of polarity shifting. DSA makes the analysis of both sides (negative and positive) of single review. Bag of Words is the way to illustrate text in statistical machine learning approaches in sentiment analysis. But sometime due to fundamental insufficiency, the performance of BOW remains limited to handle polarity shifting problem. Polarity shifting is a type of linguistic exception which reverses the opinion of text. It seems that negation is the most important polarity

shift. BOW representation considered two opposite words are very similar while making sentiment analysis. For example, " I like these flowers" is original statement whereas by adding negative word "don't" in front of "like" word such as, "I don't like these flowers" change the sentiment from positive to negative as they both are looking similar. Also there are two types of approaches for sentiment analysis first one is supervised and other is unsupervised approach. NLP is one of the unsupervised approaches which have the capability to extract meaning of natural language sentences. It makes article as well as sentiment patterns present in the text to understandable meaning. NLP is the rule based method in which developer is free to use their own knowledge for analysis purpose. NLP have some drawbacks such as, it requires help of human being to generate rules and it entirely depends on the domain of awareness. As primary work to the existing system is, make dual sentiment analysis of sentences to overcome polarity shift problem in NLP. To fulfill this approach, a corpus based pseudo antonym dictionary is developed. With this approach dependency of DSA on external antonym dictionary is reduced [1]. As secondary work, which will make use of syntactic constructions as features for the inconsistency classifier is done which will improve the performance. In addition to this, consideration for more complex polarity shift patterns is also done.

2. RELATED WORK

Abbasi, S. France, et al. [2], proposed FRN i.e. Feature Relation Network. It is a rule based multivariate text feature selection method. It is method intended to enhanced sentiment classification by enabling sets of heterogeneous n-gram features. There are several important phases which are included in data mining and which contains sentiment polarity and intensity assignment. In this study author proposed the use of higher set of n-gram feature spanning with multiple variable and fixed categories of n-gram. For opinion classification they have combined the extended feature set with the method of feature selection. The proposed technique, FRN includes semantic information that is inherited from lexical resources of n-gram features. Therefore, in this paper feature selection method shows improved classification performance than the existing feature selection methods.

Lin and Y. He et al [3], detects the sentiment and simultaneous topics from texts. For that they have used automated tools. In this paper, joint sentiment-topic (JST) model is introduced for the framework of probabilistic modeling. It based on latent Dirichlet allocation (LDA) and its main target is to detect topics from text simultaneously. The reverse series of sentiment and the process of topic generation are also implemented to obtain Reverse-JST. It is highly portable unlike the supervised approaches of sentiment classification. Author mainly focused on document-level sentiment classification for general domains in conjunction with topic detection and topic sentiment analysis, based on the proposed weakly supervised joint sentiment-topic (JST) model. Topics as well as topic sentiment identified by joint sentiment-topic (JST) model is more informative. In future work they have planned for incremental learning of JST.

K. Dave and S. Lawrence et al. [4], M. Gamon[5] et al. begins to structured reviews for training and testing. In this process they identified appropriate features as well as scoring methods. In feature selection process, training and testing of raw document they slab HTML tags and separate out the documents into sentences. Before splitting sentences into single-word tokens they are parsing through parser. In language modifications, to identify negating phrases such as "not" or "never" and mark all words following the phrase as negated, such as turning "not good or useful" into "NOT good NOT or NOTuseful." In this paper numbers of issues are identified such as, inconsistency in ratings, comparisons and uncertainty, scant data, alter distribution etc [4]. In paper [5], M. Gamon, represents the technique based on customer feedback data from the survey of web which is noisy and fragmentary. They have conducted experiments on sentiment classification using SVM filter. SVM is used for text classification. The standard supervised machine learning task consists of training and classification of sequential SVM.

A. Kennedy and D. Inkpen[6], examined three types of bearing shifters such as, negations, intensifiers and diminishers. In this paper, to identify positive and negative terms, as well as negation terms, intensifiers, and diminishers General Inquirer is utilized. A corpus based semantic orientation values of terms are computed using association scores with the small group of +ve and –ve terms. In this paper, there are two classification approaches are introduced such as, first count is introduced for positive and negative terms in the review. The positive and negative terms are taken from GI i.e. General Inquirer. The second method is to select the label of sense which estimated to frequent listed by GI. For improvement of the SVM result, the term-counting method integrates the documents score from SVM.

S. Li, Y. Chen et al [7], proposed machine learning algorithm to integrate polarity shifting information into the document level sentiment classification. In this primarily, feature selection method is used to automatically generate training data for classification of binary classifier on polarity shifting detection of sentences. Polarity shifting means the contradiction of sentence is distinct from the contradiction demonstrated by the sum of content words in the given sentences. In this paper, author referred such type of polarity shifting as polarity shifting structure. In polarity shifting, lack of relevant training data creating a large database of polarity shifting sentences which seem as time-consuming task.

B. Pang and L. Lee et al. [8], studied the problem of document classification not by topic but by overall sentiment. They specified that the examining factors of the sentiment classification make the classification more challenging problem. They were concentrating on only refining between positive and negative sentiments. There are three standard algorithm proposed in this paper for experimental analysis, they are namely, naïve Bayes classification, ME classification and SVM etc.

J. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou [9] stated that Research in *Automatic Text Classification* seeks to develop models for assigning category labels to new documents. It is based on a training set of documents that have been preclusive by domain experts. There have lot of studies of automatic text classification have concentrated on "topical classification". In topic classification documents are classified as per various subjects. This paper focused on area of 'Sentiment Classification', it is automatically classifying documents according to the overall sentiment expressed in them. There are two machine learning methods are introduced for classification of reviews into two phases such as, positive sentiment (recommended) and negative sentiment(not recommended).

S. Li and C. Huang [10] proposed two kinds of linguistic phenomena which are able to reverse the sentiment polarity. Their main focus is on sentiment shifting. It includes negation and contrast transition, hence they identify type of shifting even fully reverses the sentiment polarity. BoW modeling approach is used in this paper for complete reverse sentiment polarity. Specifically, the terms are possibly words, word n-grams, or even phrases extracted from the training data, with *N* being the number of terms. The weights are statistic information of these terms, e.g., *tf*, *idf* . Then the text *T* is represented as a vector $X(T)$ = <sta(t1), sta(t2), ….., sta(tN)>. The output label y has a value of 1 or -1 representing a positive or negative sentiment polarity. Consider the following two sentences:

a1. *This is not a good movie and I hate it.*

a2. *This is such a good movie and I do not hate it at all.*

Because they are represented as almost the same bag-of-words, their classification results would be the same when applying machine learning with one-bag-of-words modeling. But their sentiment polarities are obviously different from each other. Therefore, traditional bag-of-words modeling is not appropriate for sentiment classification to some extent.

Z. Hai, K. Chang, J. Kim, and C. C. Yang, [11] proposed IEDR method to identify opinion features from online reviews by exploiting the difference in opinion feature statistics across two corpora. IEDR is inter-corpus demography approach of opinion feature extraction. It is the feature-filtering principle which utilized dissimilarity in distributional characteristics of features across two corpora in one independent domain. In proposed method, they extract list of candidate opinion features from the context review corpus by defining the synthetic rules set. Then intrinsic-domain relevance and extrinsic-domain relevance score is estimated for each extracted candidate feature.

3. SYSTEM FLOW

The proposed method can be implemented as follows:

Figure 1 represents the process of dual sentiment analysis**:**

- Data expansion technique based on antonym dictionary is used to reverse reviews of sentences. Using text reversion and label reversion rules original reviews can be reversed.
- The DT (Dual Training) algorithm is derived by using the logistic regression model.

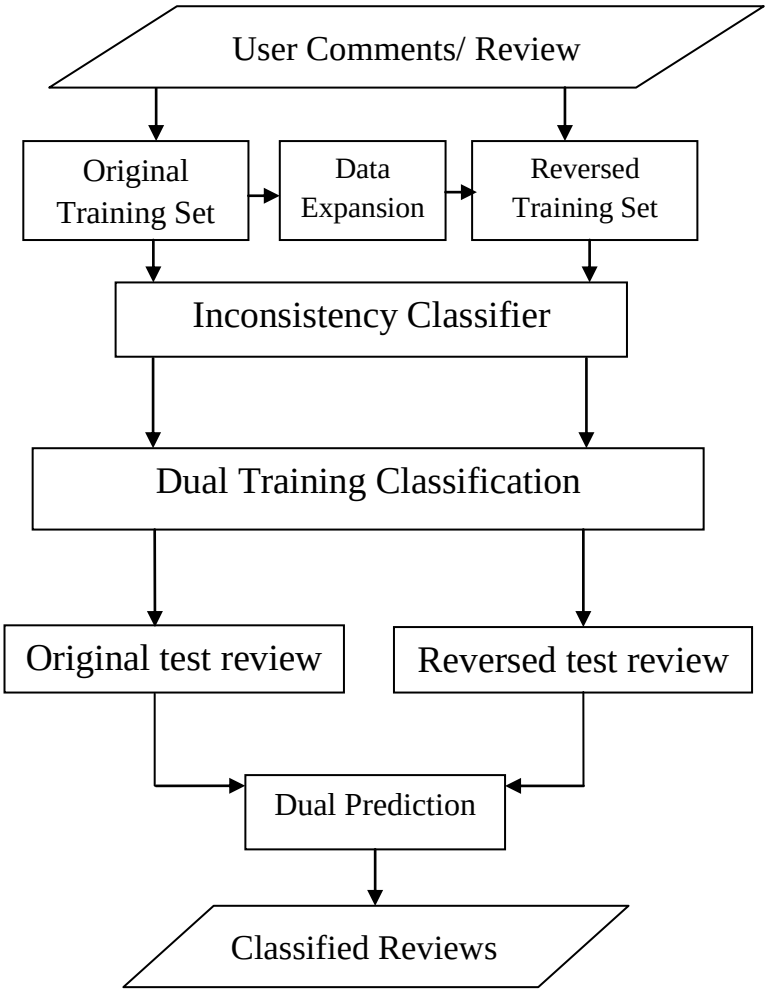

Fig. 1: System Architecture

There are following different modules listed as below:

1. User comment
2. Conversion of Review
3. Inconsistency Classifier
4. Dual Training
5. Dual Prediction

4. ALGORITHMIC STRATERGY

Two algorithms are basically developed for sentiment analysis Dual Training and Dual Prediction for carrying the sentiment analysis and classification process.

Input:

- Customer's Reviews

Output:

- Class of the Reviews/ Classified Reviews

Processing:

1. User Comment
   Obtain comments / reviews provided by the user for processing

2. Conversion of Review
   i. Determine the user speaking product using dictionary approaches
   ii. Generate dictionary for sentiment words
   iii. Data Expansion is done either by applying Text Reversion or Label Reversion technique.
   iv. A joint prediction is based on observations is made
   v. Representation of original and reversed reviews is made in BOW model

3. Inconsistency Classifier
   For training inconsistency classifier proceeds for BOW i.e. Bag of words classifier

4. Dual Training
   Then BOW identifies S_{incons} (W) with three contexts words i.e positive, negative and neutral to the left and right

5. Dual Prediction
   The class of the reviews is specified from the original and reversed test sets.

5. CONCLUSIONS

In this review paper we have studied some existing techniques of opinion mining and sentiment analysis such as, BOW, IEDR, JST, LDA etc. From all these technique, BoW is the popular technique of text mining in sentiment analysis. But its performance remains limited or restricted due to insufficiency in managing polarity shift problem. Polarity shift problem mainly arises in document level classification and it affects on the performance of statistical machine learning sentiment analysis system. Also other techniques have their own benefits and limitations which we have discussed in above section II. We analyzed that the idea of DSA will be better efficient technique to overcome the problem of polarity shifting.

ACKNOWLEDGEMENT

We are thankful to MET's Institute of Engineering Bhujbal Knowledge City Nashik, HOD of computer department, guide, parents and friends for their blessing, valuable guidance, support and motivation behind this work.

REFERENCES

[1] Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li, "Dual Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 8, August 2015.

[2] 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.

[3] Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," in Proc. 18th ACM Conf. Inf. Knowl. Manage., 2009, pp. 375–384.

[4] K. Dave, S. Lawrence and D. Pen-nock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in Proc. Int. World Wide Web Conf., 2003, pp. 519–528

[5] A.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.

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

[7] S. Li, R. Xia, C. Zong and C. Huang, "A framework of feature selection methods for text categorization," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2009, pp. 692–700.

[8] Pang and L. Lee, "Opinion Mining and Sentiment analysis," Found. Trends Inf. Retrieval, vol. 2, no. 1/2, pp. 1–135, 2008.

[9] J. Na, H. Sui, C. Khoo, S. Chan, and Y. Zhou, "Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews," in Proc. Conf. Int. Soc. Knowl. Org., 2004, pp. 49–54.

[10] S. Li and C. Huang, "Sentiment classification considering negation and contrast transition," in Proc. Pacific Asia Conf. Lang., Inf. Comput., 2009, pp. 307–316.

[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.