Novateur Publication's
International Journal of Innovation in Engineering, Research and Technology [IJIERT]
ICITDCEME'15 Conference Proceedings
ISSN No - 2394-3696

DUAL SENTIMENT ANALYSIS

Farheen Pathan.
Department of computer, Vishwabharati Academy's College of engineering , Ahmednagar, India.

Anjali Phaltane.
Department of computer, Vishwabharati Academy's College of engineering, Ahmednagar, India.

Prof.Sarika joshi.
Department of computer, Vishwabharati Academy's College of engineering, Ahmednagar, India.

ABSTRACT

Sentiment analysis or opinion mining aims to use automated tools to detect subjective information such as opinions, attitudes, and feelings expressed in text. Ideally, an opinion mining tool would process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good). Then begin by identifying the unique properties of this problem and develop a method for automatically distinguishing between positive and negative reviews. The classifier draws on information retrieval techniques for feature extraction and scoring, and the results for various metrics and heuristics vary depending on the testing situation. Now a days the most popular way to model text in statistical machine learning approaches in sentiment Analysis is Bag-of-words (BOW).Determining the polarity of a sentiment bearing expression requires more than a simple bag-of-words approach. Sometimes the performance of BOW remains limited due to some fundamental deficiencies in handling the polarity shift problem. To address this problem for sentiment classification, a model is proposed called dual sentiment analysis (DSA).So that first a novel data expansion technique is proposed by creating a sentiment-reversed review for each training and test review. Basis of this propose a dual training algorithm is proposed to make use of original and reversed training reviews and a dual prediction algorithm is proposed to classify the test reviews by considering two sides of one review. Also extend the DSA framework from polarity (positive-negative) classification to 3-class (positive-negative-neutral) classification finally, for removing DSA's dependency on an external antonym dictionary for review reversion a corpus-based method is developed to construct a pseudo-antonym dictionary in this way two tasks, nine datasets, two antonym dictionaries, three classification algorithms, and two types of features are considered. At the end results shows the effectiveness of DSA in supervised sentiment classification.

INTRODUCTION-

Sentiment analysis is to extract the opinion of the user from the text document. Identifying the orientations of opinions from the text. This movie was awesome.[sentiment]This was boring.[sentiment][21]sentiment analysis and opinion mining, as a special text mining task for determining the subjective attitude (i.e., sentiment) expressed by the text, is becoming a hotspot in the field of data mining and natural language processing[2].Opinions and its related concepts such as sentiments, evaluations, attitudes and emotions are the subjects of study of sentiment analysis and opinion mining. The inception and rapid growth of the field coincide with those of the social media on the Web, e.g., reviews, forum discussions, blogs, micro blogs, Twitter, and social networks, because for the first time in human history, we have a huge volume of opinionated data recorded in digital forms. Since early 2000, sentiment analysis has grown to be one of the most active research areas in natural language processing. It is also widely studied in data mining, Web mining, and text mining.[13]

The semantic orientation of a review can be positive, negative, or neutral. We examine the effect of valence shifters on classifying the reviews. Three types of valence shifters are examined: negations, intensifiers and diminishers.Negations are used to reverse the semantic polarity of a particular term, while intensifiers and diminish are used to increase and decrease, respectively, the degree to which a term is positive or negative. Sentiment classification is a basic task in sentiment analysis, to classify the sentiment (e.g., positive or negative) of a given text. the bag-of-words (BOW) model is typically used for text representation. In the BOW model, a Review text is represented by a vector of independent words to train a sentiment classifier statistical machine learning algorithms (such as naïve Bayes, maximum entropy classifier, and support vector machines) are then employed. The BOW model is very simple and quite efficient

ISSN No - 2394-3696

in topic-based text classification, it is actually not very suitable for sentiment classification because it disrupts the word order, breaks the syntactic structures, and discards some semantic information. One of the most well-known difficulties is the polarity shift problem. Determining the polarity of a sentiment bearing expression requires more than a simple bag-of-words approach. Polarity shift is a type 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.

RELATED WORK

Tasks in sentiment analysis can be divied into four categorizations: document level sentence-level, phrase-level, and aspect-level sentiment analysis. Focusing on the phrase/subsentence- and aspect-level sentiment analysis, Wilson et al. [20] discussed effects of complex polarity shift. Choi and Cardie [4] further combined different kinds of negates with lexical polarity items. Nakagawa et al. [15] developed a semi-supervised model for sub sentential sentiment analysis Bing Liu [6] studied the problem of determining the semantic orientations (positive, negative or neutral) of opinions expressed on product features in reviews. In machine learning methods, sentiment classification is regarded as a statistical classification problem, where a text is represented by a bag-of words and the supervised machine learning algorithms are applied as classifier [17].One common way is to directly reverse the sentiment of polarity-shifted words, then sum up the sentiment score word by word [4], [9], [10],[18].Das and Chen [5] proposed a method by simply attaching "NOT" to words in the scope of negation. Na et al. [8] proposed to model negation by looking for specific part-of-speech tag patterns. Ikeda et al.[8] proposed a machine learning method based on a lexical dictionary extracted from General Inquirer1 to model polarity-shifters for both word-wise and sentence-wise sentiment classification. I and Huang [11] proposed a method first to classify each sentence in a text into a polarity-unsuited part and a polarity-shifted part according to certain rules, then to represent them as two bags-of-words for sentiment classification. I et al. [12] further proposed a method to separate the shifted and unsuited text based on training a binary detector.Orimaye et al. [16] proposed a sentence polarity shift algorithm to identify consistent sentiment polarity patterns and use only the sentiment-consistent sentences for sentiment classification. A preliminary version of this paper was published in [22]. In this paper, we extend our previous work in three major aspects. First, we strengthen the DSA algorithm by adding a selective data expansion procedure. Second, we extend the DSA framework from sentiment polarity classification to positive-negative-neutral sentiment classification. Third, we propose a corpus-based method to construct a pseudo-antonym dictionary that could remove DSA's dependency on an external antonym dictionary. The data expansion technique has been seen in the field of handwritten recognition [3], [19] where some synthetic training data is added to improve the performance of the handwriting recognition systems .Fujita and Fuji no [7] proposed a method that provides reliable training data using example sentences from an external dictionary.

PROPOSED SYSTEM

DUAL SENTIMENT ANALYSIS

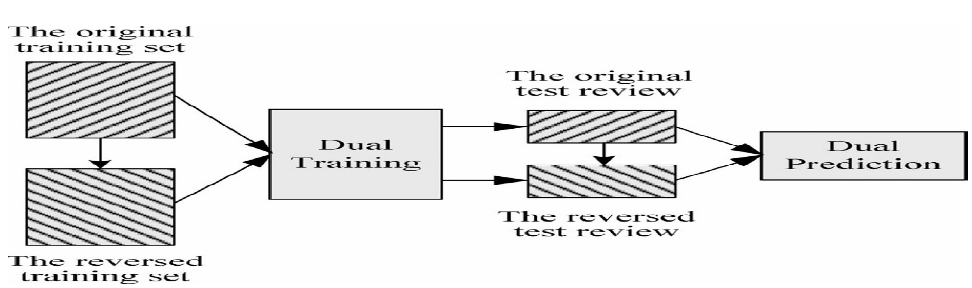

Fig1. The process of dual sentiment analysis.

In fig. The rectangle filled with slash denotes the original data, and the rectangle filled with backslash denotes the reversed data.DSA framework and its algorithm contains two main stages: 1) dual training and 2) dual prediction.

DUAL TRAINING

All of the original training samples are reversed to their opposites in the training stage.They are referred as "original training set" and "reversed training set" respectively. There is a one-to-one correspondence between the original and reversed reviews in data expansion technique. Then by maximizing a combination of the original and reversed training samples the classifier is trained. This process is called dual training. DT algorithm is also explained by using the logistic regression model as an example. The method can be easily adapted to the other classifiers such as naive Bayes and SVMs[1].

DUAL PREDICTION-

In the prediction stage, for each test sample x, we create a reversed test sample ~x. Note that our aim is not to predict the class of ~x. But instead, we use ~x to assist the prediction of x. This process is called dual prediction.Let p(.|x) and p(.|~x) denote posterior probabilities of x and ~x respectively. In DP, predictions are made by considering two sides of one review:

- When we want to measure how positive a test review x is, we not only consider how positive the original test review is (i.e. p(+|x)), but also consider how negative the reversed test review is (i.e. . p(-|x),);
- Conversely, when we measure how negative a test review x is, we consider the probability of x being negative is (i.e. . p(-|x),), as well as the probability of ~x
 Being positive (i.e. p(+|x)).

And also a weighted combination of two component predictions is used as the dual prediction score.

DSA WITH SELECTIVE DATA EXPANSION-

In dual training procedure, all of the training reviews are used in data expansion. However, in many cases, not all of the reviews have such distinct sentiment polarity.
For example-

- Review (a). The book is very interesting, and the price is very cheap. I like it.
- Review (b). The book is somehow interesting, but the price is too expensive. I don't dislike it.

In both review, for review (a), the sentiment is very strong and the polarity shift rate is low. In this case, the original review and the reversed review will also be a good one. for review (b), the sentiment polarity is less distinct.In this case, the sentiment polarity of the reversed review is also not distinct and confident. Therefore, creating reversed review for review (b) is not that necessary in comparison with review (a).
A sentiment degree metric for selecting the most sentiment-distinct training reviews for data expansion. is proposed. The degree of sentiment polarity could be measured by

$$m(x)= | p(+|x)- p(-|x)|$$

Where p (+|x) and p(-|x) are the posterior probabilities predicted on the training review x[1].

Table 1: An Example of Data Expansion for Neutral Reviews

	Review Text	Class
Original review	The room is large. But it is not clean.	Neutral
Reversed review	The room is small. But it is clean.	Neutral

DSA FOR POSITIVE-NEGATIVE-NEUTRAL SENTIMENT CLASSIFICATION-

Polarity classification is the most classical sentiment analysis task. This task aims at classifying reviews into either positive or negative. In addition to the positive and negative reviews, there are many neutral reviews. In the DSA system, earlier it does not have the ability to classify the neutral reviews. In this paper ,it extend the DSA framework to the scenario of three-class (positive-neutral-negative) sentiment classification. It is call the DSA approach in three-class Sentiment classification DSA3.Naturally, neural review contains two main situations:1) Neither positive

nor negative (objective texts without expressing sentiment); 2) Mixed positive and negative (texts expressing mixed or conflicting sentiment).In DSA3, first conduct training data expansion by creating reversed reviews. For a ergative review, create a positive one; for a positive review, create a negative one; for a neutral review, create a neutral one. The selective data expansion procedure is still used in this case that is only the labeled data with high posterior probability will be used for data expansion. In the training stage, a multi-class machine learning models, such as multi-class logistic regression (also called softmax regression), is trained based on the expanded dual Training set. In the prediction stage, for each original test sample x, we
Create an reversed one ~x.

THE LEXICON-BASED ANTONYM DICTIONARY-
In the languages where lexical resources are abundant, a straightforward way is to get the antonym dictionary directly from the well-defined lexicons, such as WorldNet in English. The WorldNet antonym dictionary is simple and direct. Even if we can get an antonym dictionary, it is still hard to guarantee vocabularies in the dictionary are domain-consistent with our tasks. To solve this problem, we furthermore develop a corpus basedmethod to construct a pseudo-antonym dictionary.

THE CORPUS-BASED PSEUDO-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 information of two random variables is a quantity that measures the mutual dependence of the two random variables. MI is widely used as a feature selection method in text categorization and sentiment classification.

REFERENCES

1]Rui Xia, Feng Xu, Chengqing Zong, Qianmu Li, Yong Qi, and Tao Li "Dual Sentiment Analysis: Considering Two Sides of One Review," IEEE Trans. Know.Data Eng,vol. 27, no. 8, Aug. 2015.
2]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.
3] J. Cano, J. Perez-Cortes, J. Arlandis, and R. Llobet, "Training set expansion in handwritten character recognition," in Proc. Struct.,Syntactic, Statistical Pattern Recognit., 2002, pp. 548–556.
4] Y. Choi and C. Cardie, "Learning with compositional semantics as structural inference for subsentential sentiment analysis," in Proc.Conf. EmpiricalMethodsNatural Language Process., 2008, pp. 793–801.
5] S. Das and M. Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards," in Proc. Asia Pacific Finance Assoc. Annu.Conf., 2001.
6] X. Ding and B. Liu, "The utility of linguistic rules in opinion mining," in Proc. 30th ACM SIGIR Conf. Res. Development Inf. Retrieval, 2007, pp. 811–812.
7] S. Fujita and A. Fujino, "Word sense disambiguation by combining labeled data expansion and semi-supervised learning method," in Proc. Int. Joint Conf. Natural Language Process., 2011, pp. 676–685.
8] D. Ikeda, H. Takamura, L. Ratinov, and M. Okumura, "Learning to shift the polarity of words for sentiment classification," in Proc. Int. Joint Conf. Natural Language Process., 2008.
9] S. Kim and E. Hovy, "Determining the sentiment of opinions," in Proc. Int. Conf. Comput. Linguistic, 2004, pp. 1367–1373.
10] A. Kennedy and D. Inkpen, "Sentiment classification of movie reviews using contextual valence shifters," Comput. Intell., vol. 22, pp. 110–125, 2006.
11] S. Li and C. Huang, "Sentiment classification considering negation and contrast transition," in Proc. Pacific Asia Conf. Lang., Inf. Comput., 2009, pp. 307–316.
12] S. Li, S. Lee, Y. Chen, C. Huang and G. Zhou, "Sentiment classification and polarity shifting," in Proc. Int. Conf. Comput. Linguistics, 2010, pp. 635–643.
13]B. Liu, "Sentiment analysis and opinion mining," in Synthesis Lectures on Human Language Technologies, vol. 5, no. 1. San Rafael, CA, USA: Morgan & Claypool, 2012, pp. 1–165.

14] 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.
15] T. Nakagawa, K. Inui, and S. Kurohashi. "Dependency tree-based sentiment classification using CRFs with hidden variables," in Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 786–794.
16] S. Orimaye, S. Alhashmi, and E. Siew, "Buy it—don't buy it: Sentiment classification on Amazon reviews using sentence polarity shift," in Proc. 12th Pacific Rim Int. Conf. Artif. Intell., 2012, pp. 386–399.
17] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: Sentiment classification using machine learning techniques," in Proc. Conf. Empirical Methods Natural Language Process., 2002, pp. 79–86.
18] L. Polanyi and A. Zaenen, "Contextual lexical valence shifters," in Proc. AAAI Spring Symp. Exploring Attitude Affect Text, 2004, pp. 1–10.
19] T. Varga and H. Bunke, "Generation of synthetic training data for an HMM-based handwriting recognition system," in Proc. IEEE Int. Conf. Document Anal. Recognit., 2003, pp. 618–622.
20] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," Comput. Linguistics, vol. 35, no. 3, pp. 399–433, 2009.
21] R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Inf. Sci., vol. 181, no. 6, pp. 1138–1152, 2011. R. Xia, C. Zong, and S. Li, "Ensemble of feature sets and classification algorithms for sentiment classification," Inf. Sci., vol. 181,no. 6, pp. 1138–1152, 2011.
22] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, "Dual training and dual prediction for polarity classification," in Proc. Annu. Meeting Assoc. Comput. Linguistics, 2013, pp. 521–525.