How to Cite:
Patil, V. (2022). Dual sentiment analysis with two sides of one review for polarity shift. *International Journal of Health Sciences*, *6*(S1), 10169–10177. https://doi.org/10.53730/ijhs.v6nS1.7390

Dual sentiment analysis with two sides of one review for polarity shift

Vaishali Patil
Asst Professor, MCA, Thakur Institute of Management Studies, Career Development & Research (TIMSCDR), Mumbai, India
Corresponding author email: patilvaishali0205@gmail.com

***Abstract*---**The Bag Of Words' performance is limited because to inherent flaws in its treatment of the polarity shift problem; the bag of words representation considers the two opposite texts to be quite similar. To address this issue, a model known as dual sentiment analysis (DSA) has been developed. Dual Sentiment Analysis expands sentiment categorization analysis from one side of a review to two sides of a single review. For both training and test reviews, a data expansion strategy is presented to solve sentiment reversed review. In the sentiment classifier, a dual training method is utilized to train the original and reversed reviews. The Dual Sentiment Analysis paradigm divides polarity categorization into three categories (positive, negative, and neutral) (positive, negative and neutral). A corpus-based method for constructing a pseudo-antonym dictionary is devised to remove dual sentiment analysis' reliance on an external antonym dictionary for review reversal. The findings show that DSA is effective in supervised sentiment analysis.

***Keywords*---**prediction, sentiment analysis, feedback, styling.

Introduction

With the rapid growth of social media content and E-Commerce businesses, individuals are more aware of the importance of expressing their opinions on products through blogs and reviews. Due to the massive volume of data produced every second, finding and monitoring people's opinions on the internet is tough. As a result, automatic sentiment analysis is required for analysing client feedback, opinion, sentiment assessment, and attitude from written language.

Sentiment classification is a fundamental problem in sentiment analysis, with the goal of classifying the sentiment of a given text (e.g., positive or negative). The Bag-of-words (BOW) model is generally used for text representation in sentiment classification, which follows the strategies employed in classical topic-based text classification.

International Journal of Health Sciences ISSN 2550-6978 E-ISSN 2550-696X © 2022.
Manuscript submitted: 27 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 1 May 2022

In the BOW model, a review text is represented by a vector of independent words. After that, statistical machine learning (such as naive Bayes, maximum entropy classifier, and support vector machines) approaches are used to train the sentiment classifier. The BOW method is easy and successful for topic-based text classification, but it is not well suited for sentiment classification because it disturbs word order, breaks syntactic structures, and discards some semantic data. As a result, many sentiment analysis projects have attempted to improve BOW by using language data. However, because of the inherent flaws of BOW, most of these initiatives had only a minor impact on classification accuracy. The polarity shift problem is one of the most well-known challenges. Polarity shift is a linguistic phenomenon in which the text's sentiment polarity can be reversed. The most common sort of polarity shift is negative. To address the polarity shift problem, several ways have been offered in the literature. However, the majority of these necessitated either extensive linguistic knowledge or additional human annotations. Because of their great reliance on external resources, the systems are challenging to implement in practise. With the lack of supplementary annotations and language understanding, some attempts were made to remedy the polarity shift problem.

Literature Survey

Document, sentence, entity, and aspect level sentiment analysis are the three layers of sentiment analysis. The document's level expresses the worldwide favourable or negative opinion of a single entity in the text. Each sentence in the paper is examined to see whether it is good, negative, or neutral. To construct an opinion poll, an aspect-based opinion polling system takes a set of textual reviews and some predetermined aspects as input and rectifies the polarity of each aspect from each review. It is based on a concept. A sentiment and a targeted object on which an opinion has been stated comprise feedback. For example, in a product review sentence, it determines the product properties the reviewer has praised and whether the comments are positive or negative. Two techniques used in sentiment classification.

Machine learning approach

Supervised

The Supervised approach is one of the types of Machine learning. To develop a prediction model, a machine learning approach needs training data. To make predictions on documents beyond the training set, predictive models such as decision trees, logistic regressions, and neural networks are used. This method has the benefit of being based on learning patterns that may be used to make automated and efficient predictions. In addition, the algorithms are capable of identifying previously unimagined and intricate patterns that would be impossible for a human to predict. However, it has downsides in that it requires a large amount of training data to develop the model, and verifying it is time consuming and difficult. Every document must have a rating, and if there are attributes of documents, each of these must have a rating. Another issue emerges when two separate reviewers assign two different sentiment ratings to the same document, which can lead to unanticipated inaccuracies in the model's building and measurement.

Unsupervised

Natural language processing is an unsupervised approach. Natural language processing (NLP) is an artificial intelligence field concerned with extracting meaning from natural language material automatically. It deciphers the meaning of the text by looking at entities and syntactic patterns. It also employs a variety of dictionaries, linguistic constructions like as pans of speech and noun phrases, as well as a variety of operators. The key benefit of rule-based techniques is that they allow rule developers to utilize their domain knowledge to create rules for analysis purposes. Rule-based approaches do not require any training data and are totally unsupervised. This is a significant benefit in real-world applications with little training data. It also has the ability to fine-tune the rules over time based on feedback from analysts or subject-matter experts, allowing the models to be adjusted. The fundamental disadvantage of the NLP technique is that it necessitates a significant amount of human engagement in the rule development process and is fully reliant on the domain knowledge of rule developers.

Lexicon based

Lexicons are collection of words where each word has a score indicating the positive, negative, neutral and objective nature of text. This method involves aggregating subjective word scores for a particular piece of text, with positive, negative, neutral, and objective word scores summed separately. Finally, we have our results. The highest score indicates the text's overall polarity.

Dictionary based approach

In this approach a set of opinion words are manually collected and a seed list is prepared. Then se search for dictionaries and thesaurus to find synonyms and antonyms of text. The newly found synonyms are added to the seed list. This process continues until no new words are found. Disadvantage: difficulty in finding context or domain oriented opinion words

Corpus based approach

A corpus is a collection of texts, usually on a specific topic. In this method, a seed list is created and then extended using corpus text. Thus is solves the problem of limited domain oriented text. It can be done in two ways

- Statistical approach: This method is used to locate terms that occur frequently in the corpus. The idea is that if a term appears frequently in positive literature, it has a positive polarity. Its polarity is negative if it is primarily seen in negative content.
- Semantic approach: This method uses the principle of word similarity to calculate sentiment values. This can be accomplished through the use of WorldNet. This can be used to find synonyms and antonyms for a particular word, as well as assess sentiment value.

Polarity Shift

Polarity shift is linguistic kind of phenomenon 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 movie" since the word "like", the sentiment of the text will be reversed from positive to negative. By the BOW representation the two sentiment-opposite texts are acknowledged to be very similar.

Classes of Sentiments

- Positive Sentiments: These are the good words about the target in consideration. If the positive sentiments are raised, then it’s being good. If the positive reviews about the commodities are more, then it is bought by many customers.
- Negative Sentiments: In deliberation, these are the negative words about the target. If negative sentiments become more common, it gets removed from the optional list. When it comes to commodity reviews, if there are more negative evaluations, no one wants to buy it.
- Neutral Sentiments: These are the words about the target that are not able to categories into neither positive nor negative sentiments. For example- Normally, I get hungry at noon. Then I order pizza.

Dual Training Analysis

The proposed method can be given as follows:

- 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

The Data Expansion Technique, which creates a reversed review of the original review, is the primary responsibility of DSA. The statistical classifier is trained using both original and reversed reviews. This is referred to as dual training (DT). Testing is carried out using Dual Training, also known as Dual Prediction. (DP).Figure 1 shows the system architecture of the proposed system. Steps involved in this method are

1. Data Expansion
2. Dual Training
3. Dual Prediction

Data Expansion

Data Expansion is the process of using an antonym dictionary to create a sentiment reversed review of the original review. Table 1 shows example of creating reversed review. Data expansion involves:

Sentiment word reversion

All sentiment words are inverted to their antonyms outside of the scope of negation. The reach of the negation word extends all the way to the end of the sub-sentence.

Handing negation

If a negative phrase occurs, define the scope of negation first, then eliminate negation words (such as no, not, and don't). The sentiment words are not flipped in the context of denial.

Label Reversion

The class label is also reversed (i.e. Positive to Negative and vice versa)

Table-1
Example for creating reversed review

	Review Test	**Class**
Original Review	**“I don’t like this movie. It is boring”**	**Negative**
Reversed Review	**“I like this movie. It is interesting”**	**Positive**

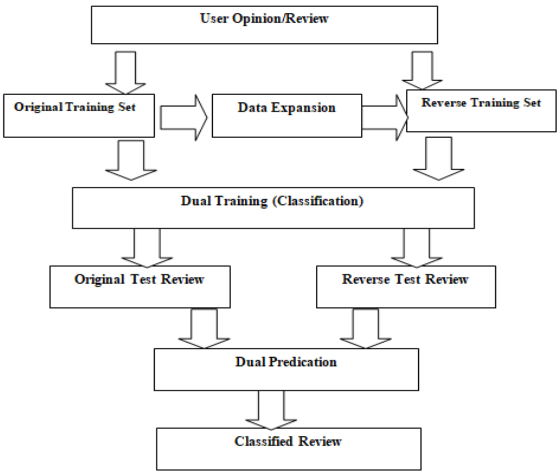

Fig-1 Process of dual sentiment analysis

Dual Training

All of the original training samples are reversed to their opposites throughout the training stage, and they are referred to as "original training set" and "reversed training set," respectively. In terms of data growth, the original and inverted reviews have a one-to-one association. The classifier is trained by maximizing the combination of feasibilities of the original and reversed training data. This method is known as dual training.

Dual Prediction

The polarity shift problem can be solved via dual prediction. In the prediction stage, for each test sample x, a reversed test sample x is created. Here, is not aim to predict the class of x . But instead, it used to assist the prediction of x. This process is called dual prediction (DP).

Let us use the example in Table I again to explain why dual prediction works in addressing the polarity shift problem. This time we assume "/don't like this movie. It is boring" is an original test review. And "1 like this movie. 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 mis-classified 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. In this manner, the prediction error of the original test sample can also be compensated by the prediction of the reversed test sample. This can reduce some prediction errors caused by polarity shift.

Positive-Negative-Neutral Sentiment Classification

Let us use the example in Table I again to explain why dual prediction works in addressing the polarity shift problem. This time we assume "/don't like this movie. It is boring" is an original test review. and "1 like this movie. 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 mis-classified 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. In this manner, the prediction error of the original test sample can also be compensated by the prediction of the reversed test sample. This can reduce some prediction errors caused by polarity shift.

Table-2
Example for creating neutral review

	Review Test	Class
Original Review	“The house is larage.But it is not clean”	Neutral
Reversed Review	“The house is small .But it is clean	Neutral

The Antonym Dictionary for review reversion

The Lexicon based Antonym Dictionary

In the languages where lexical resources are abundant, a straight forward way is to get the antonym dictionary directly from the well-defined lexicons, such as WordNet in English. WordNet is a lexical database which groups English words into sets of synonyms, provides short, general definitions, and records the various semantic relations between these synonym sets. Using the antonym dictionary, it is possible to obtain the words and their opposites. The WordNet antonym dictionary is simple and direct. Syntax, location, appropriate domain, regulations, user's profile, process, task, and goal. They may draw on multiple sources of information, including both structured and unstructured digital information, as well as sensory inputs (visual, gestural, auditory, or sensor-provided). Even if we can get an antonym dictionary, it is still hard to guarantee Vocabularies in the dictionary are domain consistent with our tasks.

The Corpus Based antonym dictionary

In information theory, the mutual information of two random variables is an amount with the purpose of measures the mutual measurement between variables. MI is extensively used as a feature selection procedure in sentiment categorization. Primary, select each and every one adjectives, adverbs and verbs in the training text documents as candidate features, and make use of the MI metric in the direction of determine the significance of each candidate feature wi in the direction of the positive (13) and negative (-) class, correspondingly. It is significant in the direction of notice with the purpose of rather than a common-sense antonym dictionary, it is a —pseudo! Antonym dictionary, here, —pseudo! Means a pair of antonym words are not actually semantic-opposite, however contain opposite sentiment strong point just must in the direction of maintain the level of sentiment strength in review reversion. Actually, the MI gives a good evaluate of the appropriate sentiment strength. Consequently, the form of the similar level sentiment strength is able to be essential by means of pairing the positive and negative-relevant words among the same ranking positions. Furthermore, since the pseudo-antonym dictionary is learnt from the training text documents, it has a high-quality property: language-independent and domain-adaptive. This property formulates the DSA model potential in the direction of be applied into a various range, particularly when the lexical antonym dictionary is not obtainable across varied languages and domains.

Scope of dual sentiment analysis

Summary of viewpoints
For retailers, customer satisfaction is now a decision element. A great number of opinions/reviews are generated with the use of the internet, which must be managed at all times. Opinions/reviews can be adequately assessed utilizing sentiment polarity, degree of connected event, and sentiment polarity.

Identifing Influence
By checking the polarity of sentiment we can able to find out influence of customers.

Interative data visualization
Data visualization is a simple procedure that can provide additional weight to the analysis of the present status of the product if we have correct information about customer influence.

Acceptance of the Product
Companies can use dual sentiment analysis to determine the amount of product acceptance and develop strategies to improve product quality.

Case Study

Domain Sentiment datasets are used here. They contain product reviews taken from Amazon.com including four different domains: Apparel, DVD, Electronics and Kitchell. Each of the reviews is rated by the customers from Star-1 to Star-5. The reviews with Star-1 and Star-2 are labeled as Negative, and those with Star-4 and Star-5 are labeled as Positive. Each of the four datasets contains 1,000 positive and 1,000 negative reviews. Reviews in each category are randomly split up into 60%-40% (with 60% serving as training data and the remaining 40% serving as test data). Following four models are taken for evaluation:

- Normal Sentiment Analysis (NSA): Direct sentiment analysis that considers the original review only
- Dual Sentiment Analysis with unigrams (DSA-UN): Here DSA is performed considering only unigrams as feature.
- Dual Sentiment Analysis with unigrams and bigrams (DSA-U13): In this case, DSA is performed considering both unigrams and bigrams as features.

From the above three systems, better performance is obtained for DSA with unigrams and bigrams (DSA-UB) compared to other two systems. The results of the experiment performed using apparel, DVD, electronics and kitchen datasets are as shown in the table 3 below.

Table-3
Result Analysis

DATA SET	NSA	DSA-UN	DSA-UB
APPAREL	82.207	83.558	84.234
DVD	84.234	84.684	85.135
ELECTRONICS	84.685	86.909	88.10
AVERAGE	83.558	84.846	85.651

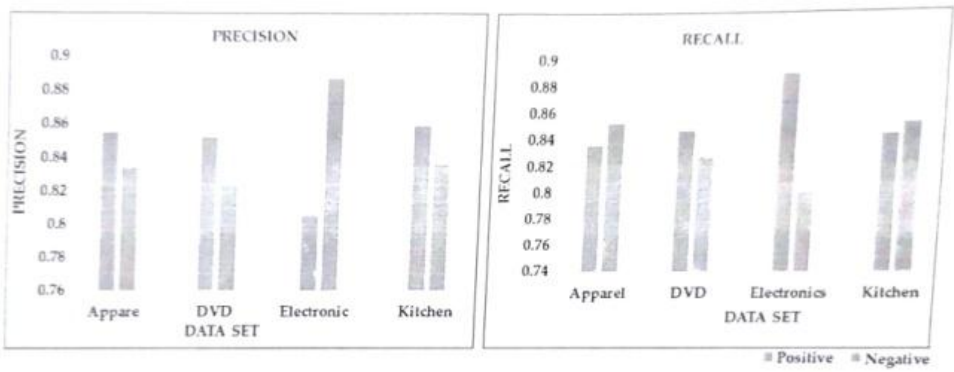

Fig-2 Precision and Recall of DSA-UB

Existing system used DVD, electronics and Kitchen datasets. In order to compare with the existing system, experiments are conducted using same datasets. Figure 3 shows the accuracy comparison between existing DSA and proposed system and from the figure, it is clear that proposed system outperforms the existing system.

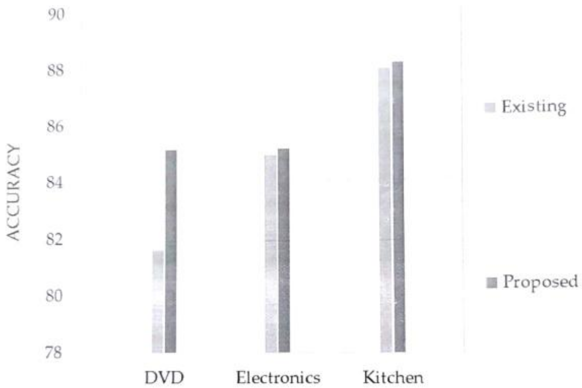

Fig 3- accuracy Comparison-Existing Vs Proposed

Conclusion

Dual sentiment analysis aids in the reduction of polarity shift. Data expansion techniques are used in dual sentiment analysis to obtain a reverse review of the original one. . When working with limited lexical resources and domain expertise, using a corpus-based pseudo-antonym dictionary implies a fairly broad practical approach. Dual training is carried out using a large number of original and reversed training samples to train classifiers. On neutral review, dual sentiment analysis (DSA) does not work. For document level sentiment analysis and complex polarity problems, DSA needs to be improved. Dual sentiment analysis can be used by businesses to measure the level of product acceptability and design methods to improve product quality.

References

[1] Rui Xia, Peng Xu, Chengqing 'Long, Qianmu Li, Yong Qi, and Tao Li ," Dual Sentiment Analysis: Considering Two Sides of One Review",IEEE transactions on knowledge and data engineering, vol. 27, no. 8, august 2015. K. Elissa, “Title of paper if known,” unpublished.

[2] Adhangle Sahyadri,Bhavani Matnta G,Kadam Sneha S..- Dual Sentiment Analysis",IJMTER ,vol. 04,1ssue 03, March-2017.

[3] Harpreet Kaur,Veenu Mangat,Nidhi,"A Survey of Sentiment Analysis Techniques",IEEE International Conference on I-SMAC,2017.

[4] Bajeela P. V,Swaradha P,"Effective sentiment classification using Dual sentiment analysis and random forest classifer-IJIRCCE,Vol.04,1ssue 04,April 2017.