



ISSN: 2350-0328

**International Journal of Advanced Research in Science,  
Engineering and Technology**

**Vol. 4, Issue 6 , June 2017**

# **Sentiment Analysis Model for Polarity Classification Based on User Generated Review**

**Harshada Deshmukh, Dr.P.L.Ramteke**

P.G. Student, Department of Computer Science and Information technology, H.V.P.M.College Of Engineering,  
Amravati, India

Associate Professor and Head of Computer Science and Information technology, H.V.P.M.College Of Engineering,  
Amravati, India,

**ABSTRACT:** Bag-of-words (BOW) is now the most popular way to model text in statistical machine learning approaches in sentiment analysis. However, the performance of BOW sometimes remains limited due to some fundamental deficiencies in handling the polarity shift problem. The propose model called dual sentiment analysis (DSA) address the problem for sentiment classification. Firstly propose a novel data expansion technique by creating a sentiment-reversed review for each training and test review. On this base, propose a dual training algorithm to make use of original and reversed training reviews in pairs for learning a sentiment classifier, and a dual prediction algorithm to classify the test reviews by considering two sides of one review. It extends the DSA framework from polarity (positive-negative) classification to 3-class (positive, negative, neutral) classification, by taking the neutral reviews into consideration. By dual sentiment analysis proposes approach is to analysis sentiment as well as its automatic rating count. This can be calculated by using user review on the basis of positive, negative and neutral response. Then calculate all review and display the result analysis.

**KEY WORDS:** Sentiment Analysis, Polarity Classification, Techniques, Movie Review.

## **I.INTRODUCTION**

In recent years, with the growing volume of online reviews available on the Internet, 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. Sentiment classification is a basic task in sentiment analysis, with its aim to classify the sentiment (e.g., positive or negative) of a given text. The general practice in sentiment classification follows the techniques in traditional topic-based text classification, where the Bag-of words (BOW) model is typically used for text representation. In the BOW model, review text is represented by a vector of independent words.

The statistical machine learning algorithms (such as naïve Bayes, maximum entropy classifier, and support vector machines) are then employed to train a sentiment classifier. Although the BOW model is very simple and quite efficient 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.

A propose simple yet efficient model, called dual sentiment analysis (DSA), is used to address the polarity shift problem in sentiment classification.

By using the property that sentiment classification has two opposite class labels (i.e., positive and negative), first propose a data expansion technique by creating sentiment reversed reviews. The original and reversed reviews are constructed in a one-to-one correspondence. On analysis of current dual sentiment analysis propose approach is to analysis sentiment as well as its automatic rating count. This can be calculated by using user review on the basis of positive, negative and neutral response. Than calculate all review and display the result analysis.

Thereafter, propose a dual training (DT) algorithm and dual prediction (DP) algorithm respectively to make use of the original samples in pairs for training a statistical classifier and make predictions. In DT, the classifier is learnt by

maximizing combination of likelihoods of the original training data set. In DP, predictions are made by considering two sides of one review. That is, not only measure how positive/negative the original review is, but also how negative/positive review in rating based.

## II. LITERATURE REVIEW

According to the levels of granularity, tasks in sentiment analysis can be divide into four categorizations: document-level, sentence-level, phrase-level, and aspect-level sentiment analysis. Focusing on the phrase/sub sentence- and aspect-level sentiment analysis, Wilson et al discussed effects of complex polarity shift. They began with a lexicon of words with established prior polarities and identify the contextual polarity of phrases based on some refined annotations. Choi and Cardie [4] further combined different kinds of negators with lexical polarity items though various compositional semantic models, both heuristic and machine learned, to improved sub sentential sentiment analysis. Nakagawa et al. developed a semi supervised model for sub sentential sentiment analysis that predicts polarity based on the interactions between nodes in dependency graphs, which potentially can induce the scope of negation. In aspect-level sentiment analysis, the polarity shift problem was considered in both corpus- and lexicon-based methods.

The data expansion technique has been seen in the field of handwritten recognition [3], where the performance of the handwriting recognition systems was significantly improved by adding some synthetic training data. In the field of natural language processing and text mining, Agirre and Martinez [2] proposed expanding the amount of labelled data through a Web search using monospermous synonyms or unique expressions in definitions from WorldNet for the task of word sense disambiguation.

## III. PROPOSED SYSTEM

In proposed system there are basically two modules are available .First module indicate dual sentiment analysis (DSA) framework in detail. And second module indicate the prediction user review rating on user previously review data. Fig.3 illustrates the process of a dual sentiment analysis. It contains two main stages: A) dual training (DT) and B) dual prediction (DP).

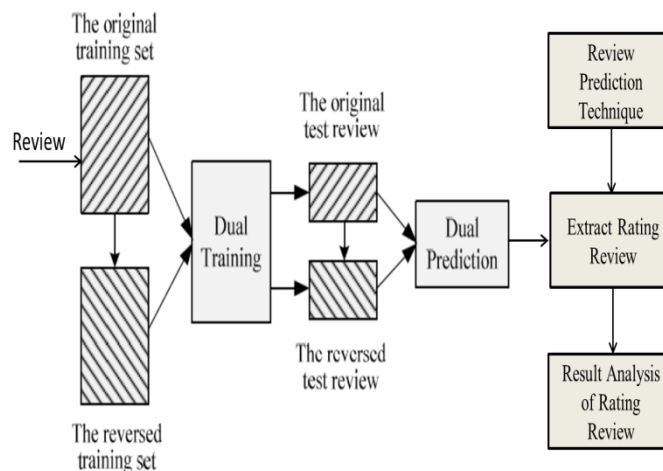


Fig1. The process of dual sentiment analysis. The rectangle filled with slash denotes the original data, and the rectangle filled with backslash denotes the reversed data.

### A. 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.

**B. 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.

**C. Review Prediction technique**

In review prediction technique, users predict the reviews like positive and negative. This collection of reviews used for various opinion analysis tasks. Users give the reviews on movies online. The admin then checks reviews, and displays an online rating for every movie.

**D. Extract Rating Review**

After gives the user online reviews, system extract the reviews like positive, negative and neutral .Then it displays an online rating reviews and extracting from the positive, negative and neutral reviews.

**E. Result Analysis Of Rating Review**

In the result analysis, all the reviews are gathers for a particular movie and calculate an average rating to score it.This score is generated for every movie in the system. This system also sorts and displays top rating movies as per analysis and calculate the result. This provides an automated movie rating system based on sentiment analysis.

**IV. ALGORITHM FOR SOCIAL SENTIMENT ANALYSIS**

This algorithm takes an English sentence and assigns sentiment ratings of "positive", "negative" and "neutral".

Identify and extract sentiment in given English string. Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials.

**A. Input:**

(Required): String sentence\* or A list of strings\*\*

**B. Output:**

Sentiment\* of given sentence(s).

\*Note: 4 sentiment types are returned: Positive, negative, neutral & compound. The first three sentiments scale from 0 to 1. Compound sentiment is the overall sentiment, where it scales between -1 to 1, negative to positive respectively.

**Ex:****Input:**

```
{  
  "sentenceList": [  
    "I like double cheese pizza",  
    "I love black coffee and donuts",  
    "I don't want to have diabetes"]  
}
```

**Output :**

```
[
  {
    "positive": 0.455,
    "negative": 0,
    "sentence": "I like double cheese pizza",
    "neutral": 0.545,
    "compound": 0.3612
  },
  {
    "positive": 0.512,
    "negative": 0,
    "sentence": "I love black coffee and donuts",
    "neutral": 0.488,
    "compound": 0.6369
  },
  {
    "positive": 0,
    "negative": 0.234,
    "sentence": "I don't want to have diabetes",
    "neutral": 0.766,
    "compound": -0.0572
  }
]
```

**V. SENTIMENT CLASSIFICATION**

Sentiment classifications are based on polarity, which may become positive, negative, or neutral. That's mean opinions may be classified into positive, negative, or neutral. Moreover, there is a forth type which is a constructive opinion which obtains suggestion to make the product better. Opinions are classified into three categories: the first one is direct opinions which opinion holder directly attack to target. Second one of opinion is comparative opinions which are opinion holder compare among entity. The third one is indirect opinions, which are implied as in idioms or expressed in a reverse way as in sarcasm. Researchers have studied sentiment analysis into three levels:

**A. Document Level Sentiment classification**

Document level sentiment classification aims to classify the entire document as positive or negative. There is much actual work use one of the two types of classification techniques which are a Supervised method and Unsupervised method to build level document sentiment.

**• Supervised method**

Sentiment classification is performed at document level sentiment. Sentiment classification can be used as a supervised classification problem with three classes positive, negative and neutral. Also, supervised request machine-learning algorithms like SVM Support Vector Machines to conclude the relationships between the opinions that expressed and text segment. A lot of researchers found that supervised learning techniques can perform well in SVM and Naïve Bayes.

**• Unsupervised method**

Unsupervised classification is performed at the sentence level. There are two types of unsupervised classification, which are lexicon-based and syntactic-pattern based. Sentence and aspect level sentiment classification for the lexicon-based can be used.

**B. Sentence Level Sentiment classification**

In this level, the task is to determine each sentence in the document as positive or negative opinions. Sentence level sentiment analysis has classified the polarity. This level is close to document level but here it accomplished by every sentence. However, there may be complex sentences in the text which make the sentence level is not helpful. There are two phases in level sentence sentiment done in every single sentence: First, each sentence classified, as subjective or objective, and the second one is the polarity of subjective sentence.

**C. Aspect Level Sentiment classification**

It supposes that a document has a hold opinion on many entities and their aspects. Aspect level classification needs discovery of these entities, aspects, and sentiments for each of them.

**VI. SENTIMENT ANALYSIS TECHNIQUE**

There are two techniques used in Sentiment Classification, Machine learning and Lexicon based. But in this paper, used only Lexicon based technique.

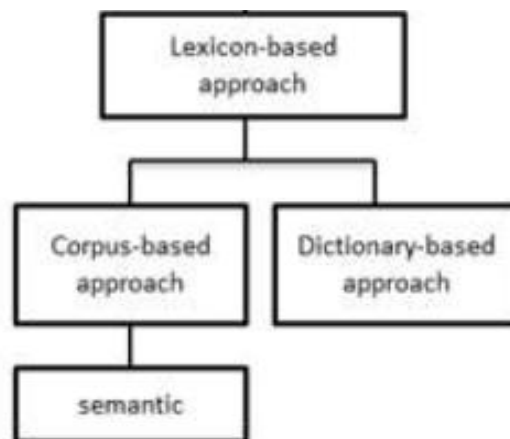
**Lexicon Based Approach**

Fig2. Lexicon based approach

**The Lexicon Based approach** involves calculating sentiment polarity for a review using the semantic orientation of words or sentences in the review.

Sentiment words are used in many sentiment classification tasks. Positive and negative sentiment words are used to express some desired and undesired states respectively. Lexicon based approach deals with searching the axioms such as adjective, adverb, noun etc. form the sentence and comparing with seed words or training data set with its corresponding polarity in the database of words.

There are three main approaches to collect the sentiment word list. Manual approach, it is very time consuming and automated approach. The two types of automated approaches are presented in the following subsections:-

**A. Dictionary-based approach**

Sentiment words are collected manually to form a small list, which is later developed by searching more words from a known corpora wordnet. Wordnet is a corpora the produces synonyms and antonyms for a word. The new words found excluding the seed words are included to the list. The process continues until no new words are found from the corpora. The major drawback of dictionary-based approach its inability of finding sentiment words with respect to a domain or a feature.

**B. Corpus-based approach**

The Corpus-based approach helps to overcome the drawback of dictionary based approach in finding sentiment words with feature specific orientations. It depends on certain patterns that occur together along with a seed list of sentiment words to find similar sentiment words in a large corpus.

Corpus-based approach has a disadvantage that corpus-based alone is not as effective as the dictionary based approach because it is hard to cover all English words, but this approach has a major advantage that it can help to find domain and feature specific sentiment words using a domain corpus. The corpus-based approach is performed using statistical approach or semantic approach as illustrated in the following subsections:

- **Semantic approach**

This method of finding co-occurrence seed sentiment words using semantic technique is done by deriving polarities using the co-occurrence of axioms as adjectives, adverbs in a corpus.

The Semantic approach gives sentiment values directly and gives similar sentiment values to semantically close words. These Semantically close words can be obtained by getting the list of sentiment words and iteratively expanding the initial set with synonyms and antonyms and then determining the sentiment polarity for an unknown word by the relative count of positive and negative synonyms of this word.

**VII. RESULT AND DISCUSSION**

In this system, user gives the reviews like positive, negative and neutral. This collection of reviews used for various opinion analysis tasks. Users give the reviews on movies online. The admin then checks reviews, and displays an online rating for every movie. Usually come across movie rating websites where users are allowed to review on movies online. These rating are provided as input to the website admin. The admin then checks reviews and displays an online rating for every movie. Online system that automatically allows users to post reviews and store them.

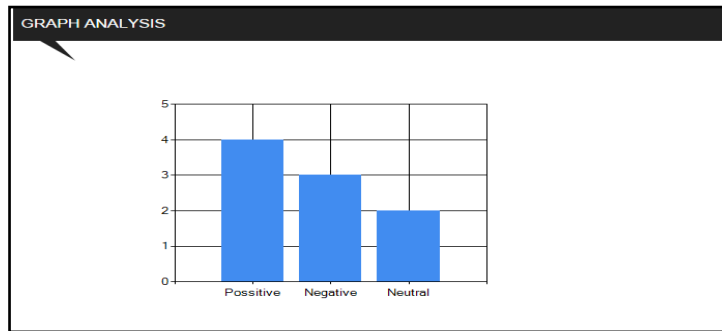
In the result analysis, first calculate the number of positive-negative-neutral count then on the basis of counting, it represents the graphically. The system now analyses this data to check for user sentiments associated with each review. The system now gathers all review for a particular movie and then calculates the number of count for the positive-negative-neutral. This provides an automated movie rating system based on sentiment analysis.

In result analysis, first show the Review analysis table and then graphical analysis with the help of particular movie.

- **Movie 1: Bahubali**

Table 3. Review Analysis

REVIEW ANALYSIS	
Positive Count	4
Negative Count	3
Neutral Count	2



Graph 3. Review Analysis of Graph

- Negative World Matching

```
string[] narr = { "not", "dislike", "never", "none", "no" };
foreach (string ne in narr)
{
    string[] wordsToMatch;
    wordsToMatch = new string[1];
    wordsToMatch[0] = ne;
```

- Find sentences that contain all the terms in the words To Match array.
- The number of terms to match is not specified at compile time.

```
var sentenceQuery = from sentence in sentences
let w = sentence.Split(new char[] { ',', '?', '!', ' ', ';', ':', '!', '!' },
StringSplitOptions.RemoveEmptyEntries)
Where
w.Distinct().Intersect(wordsToMatch).Count()==wordsToMatch.Count()
select sentence;
```

```
// Response.Write(sentenceQuery);
foreach (string c in sentenceQuery)
{
    nc += 1;
}
}
```

- Positive world match...

```
pc = pc + sentences.Length - 1 - nc;
// int s = sentences.Length - 1;
// Response. Write ("T="+s+" NC=" + nc+" Pc= "+pc);

tnc = tnc + nc;
tpc = tpc + pc;
if (nc == pc)
    tnc = tnc + 1;
if (nc > pc && pc != 0)
    tnc = tnc + 1;
```



```
if (nc < pc && nc != 0)
    tnc = tnc + 1;
}
```

```
txtPositive.Text = tpc.ToString();
txtNegativeCount.Text = tnc.ToString();
txtNeutralCount.Text = tnc.ToString();
con.Close();
```

### VIII. CONCLUSION AND FUTURE WORK

Sentiment Analysis is very important research because Sentiment Analysis help in summarizing opinion and reviews of public. They consider as research filed. However, Sentiment Analysis still needs to improve and progress. Moreover, there are many challenges like the polarity in a complex sentence.

In this dissertation, focus on creating reversed reviews to assist supervised sentiment classification. The basic idea of DSA is to create reversed reviews that are sentiment-opposite to the original reviews, and make use of the original and reversed reviews in pairs to train a sentiment classifier and make predictions. DSA is highlighted by the technique of one-to-one correspondence data expansion and the manner of using a pair of samples in training (dual training) and prediction (dual prediction). Then highlights the basic ideas about Sentiment Analysis and then explains in details the Sentiment Classification and Technique.

In the future scope, can generalize the DSA algorithm to a wider range of sentiment analysis tasks. And consider also a plan for more complex polarity shift patterns such as transitional, subjunctive and sentiment-inconsistent sentences in creating reversed reviews.

### 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] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, "Dual Training and Dual Prediction for Polarity Classification," *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 521-525, 2013.
- [3] A. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 23, no. 3, pp. 447-462, 2011.
- [4] T. Varga and H. Bunke, "Generation of synthetic training data for an HMM-based handwriting recognition system," *Proceedings of the IEEE International Conference on Document Analysis and Recognition (ICDAR)*, 2003.
- [5] X. Ding and B. Liu, "The utility of linguistic rules in opinion mining," Proceedings of the 30th ACM SIGIR conference on research and development in information retrieval (SIGIR), 2007.
- [6] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," Proceedings of the 18th ACM conference on Information and Knowledge Management (CIKM), 2009.
- [7] C. Lin, Y. He, R. Everson, and S. Ruder, "Weakly supervised joint sentiment-topic detection from text," IEEE Transactions on Knowledge and Data Engineering (TKDE), vol. 24, no. 6, pp. 1134-1145, 2012.
- [8] B. Liu, "Sentiment Analysis and Opinion Mining," *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool, pp. vol. 5, no. 1, pp. 1-165, 2012.
- [9] S. Li and C. Huang, "Sentiment classification considering negation and contrast transition," *Proceedings of the Pacific Asia Conference on Language, Information and Computation (PACLIC)*, 2009.
- [10] S. Li, S. Lee, Y. Chen, C. Huang and G. Zhou, "Sentiment Classification and Polarity Shifting," *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2010.
- [11] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, "Dual Training and Dual Prediction for Polarity Classification," *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 521-525, 2013.
- [12] X. Yu, Y. Liu, X. Huang, and A. An, "Mining online reviews for predicting sales performance: A case study in the movie domain," *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, vol. 24, no. 4, pp. 720-734, 2012.