Aspect Based Sentiment Analysis of Movie Reviews

Finding the polarity directing aspects

Viraj Parkhe
Dept. of Computer Science and Engineering
IIT(BHU), Varanasi
Varansi, India
viraj.parkhe.cse10@iitbhu.ac.in

Bhaskar Biswas
Dept. of Computer Science and Engineering
IIT(BHU), Varanasi
Varanasi, India
bhaskar.cse@iitbhu.ac.in

***Abstract*—Sentiment analysis of a movie review plays an important role in understanding the sentiment conveyed by the user towards the movie. In the current work we focus on aspect based sentiment analysis of movie reviews in order to find out the aspect specific driving factors. These factors are the score given to various movie aspects and generally aspects with high driving factors direct the polarity of the review the most. The experiment showed that by giving high driving factors to Movie, Acting and Plot aspects of a movie, we obtained the highest accuracy in the analysis of movie reviews.**

Keywords-Sentiment Analysis; Aspect Based Sentiment Analysis; Naive Bayes Classifier; Aspect importance

I. INTRODUCTION

Sentiment analysis[13] refers to finding whether the given document is either positive or negative in polarity using various language processing tools and/or machine learning approaches. In this work the sentiment analysis is applied on movie reviews to find whether the given review has positive or negative orientation. The sentiment analysis of a movie review can be at two levels: Document level sentiment analysis and Aspect level sentiment analysis [4]. At the first level, the entire document is classified as either positive or negative using some lexicon based scoring method or using machine learning approaches. Bo Pang et al[3] suggested a good method of sentiment classification using Naïve Bayes , SVM and Maximum Entropy classifiers. They experimented with different features like unigrams, unigrams and bigrams, adjectives, top unigrams etc. and compared their results. Hanhoon Kang et al[10] proposed a method for mitigating the error caused when the accuracies of the positive and negative classes are expressed as average values. For this they proposed an improved Naïve Bayes algorithm that reduced the accuracy gap. Sometimes the accuracy obtained by the machine learning algorithms is low, thus to address this problem Abd. Samad Hasan Basari et al [11] used Support Vector Machines coupled with Particle Swarm Optimization to increase the overall accuracy. In their study they increased the accuracy from 71.87% to 77%.

At the second level, the polarity of each aspect of a movie is determined. A movie has many different aspects such as Direction , screenplay, acting, story etc. and the reviewer may tend to give his/her opinion based on these aspects. In recent years the second approach has grown popular because we can better analyze the review if we take into account the polarities of individual aspects. In Document level analysis, we overlook the opinion expressed by the user over the different aspects of the movie. As reviewers tend to have different opinions about different aspects of the movie, aspect based analysis is the way to obtain a detailed analysis of the review.

Many researchers have worked on aspect based sentiment analysis. Tun Thura Thet et al [1] proposed a method for fine-grained analysis of sentiment orientation and sentiment strength of the reviewer towards the various aspects of the movie. It uses domain specific and generic opinion lexicons to score the words and with the help of dependency tree, it identifies various inter word dependencies , and helps in propagating the word score over the entire document. V.K. Singh et al [4] gave a new feature based heuristic for aspect level sentiment analysis. In their scheme they analyse the review text and assign sentiment label on each aspect of the review. Then each aspect text is scored using SentiWordNet [5] with feature selection comprising of adjective, adverbs, verbs and n-gram features. The overall document is then scored based on the aggregate score of each aspect. Jianxing Yu et al [12] proposed a method for identifying important aspects from online consumer reviews. They identified the important aspects based on the observations that such aspects are commented the most in a review and overall product opinion is greatly influenced by consumer's opinion on such important aspects. In their algorithm they formulate the aspect value distribution via a Multivariate Gaussian Distribution.

In this paper we propose a method to find out the aspects that dictate the sentiment score of the review the most. For this purpose we tend to use some "driving factors" which give weightage to different aspects of the movie. Thus the overall score of the document is the sum of individual aspect scores weighed by their driving factors.

Jianxing Yu et al [12] 's approach differs from our approach in the method by which they assign aspect values. They use a Multivariate Gaussian distribution while we use a randomized approach to assign values to the driving factors. Also we choose those driving factors that give the maximum

978-1-4673-6751-6/14 $31.00 © 2014 IEEE
DOI 10.1109/ISCMI.2014.16

accuracy, as the best driving factors. The rest part of the paper is organized as follows: Section II describes the proposed method, Section III gives the dataset, experimental results and performance, and Section IV gives the conclusion and future work and section V are references.

II. PROPOSED METHOD

The method aims at developing a technique for aspect based sentiment analysis of movie reviews. Fig.1 describes the flow of the proposed method. The review of the movie from different sources are collected and pre-processed to make it suitable for applying in the method. The pre-processing step includes the formatting of the different reviews so that they can be aligned in a required format. For this the HTML tags and other tags were removed. In our case we pre-processed the reviews into simple text format. These pre-processed reviews are then passed through the aspect based text separator and the separated review text was obtained. The various movie aspects that we used are screenplay, music, acting, plot, movie, and direction. The functionality of the separator was to separate the review aspect wise. The separator used an aspect specific lexicon for the purpose of text separation. Table I. shows some of the words used to separate the sentences [1].Each word in the lexicon were associated with the part of speech of that word. While searching the sentence to match the lexicon word, we first tagged the sentence with the Stanford Part –Of-Speech tagger [9], and then we matched the lexicon word within the sentence having the same part of speech.

These aspect based separated sentences were given as an input to the classifiers meant for each aspects. A Naive Bayes classifier [3] was used for this purpose. It calculates the probability of a word or albeit a sentence, belonging to positive or a negative class of reviews. The traditional method of training and testing the classifier is applied. The output of the classifier is either 1 or -1 denoting that the input text was of positive or negative orientation respectively. Instead of NB we can use any classifier like SVM etc. that is able to clearly classify the text in two classes. However we must carry out proper processing of input data so that it meets the proper data format requirement for each classifier. Based on the weightage of the driving factors of the movie, the aspect based output is multiplied with the respective driving factor.

TABLE I. LEXICON USED FOR ASPECT BASED TEXT SEPARATOR

Aspect	Aspect Words
Screenplay	scene, scenery, animation, violence, screenplay,action etc
Music	music, score, lyric, sound, audio, musical, title track, etc
Acting	acting, role playing, act, actress, actor, role, portray, character, villain, performance etc
Plot	plot,story,storyline,tale,romance,dialog,script,storyteller ,ending,storytelling,revenge,betrayal,writing etc
Movie	movie, film, picture, moving picture,, motion picture, show, picture show, pic, flick, romantic comedy etc
Direction	directing,direct,direction,director,filmed,filming,film making,filmmaker,cinematic,edition,cinematography etc

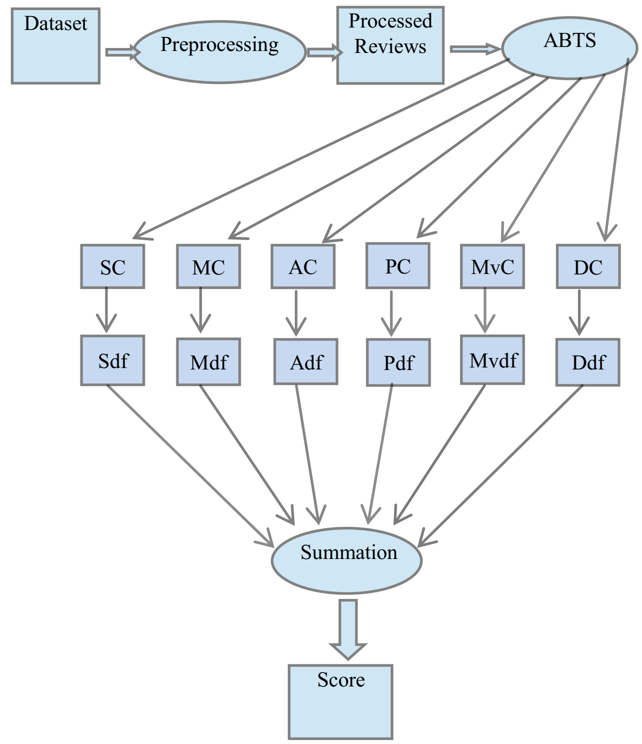

Figure 1. Diagram for the proposed method

ABTS= Aspect Based text Separator SC = Screenplay Classifier, MC = Music Classifier AC = Acting Classifier, PC = Plot Classifier, MvC = Movie Classifier, DC = Direction Classifier. Sdf= Screenplay driving factor multiplication, Mdf= Music driving factor multiplication, Adf = Acting driving factor multiplication, Pdf= Plot driving factor multiplication, Mvdf= Movie driving factor multiplication, Ddf= Direction driving factor multiplication.

Driving factors denote the importance of an aspect in the review. The more the value of driving factor for an aspect, the more is its importance in the review.

The driving factors follow the relationship:

$$\Sigma\alpha_i = 1 \tag{1}$$

Where α_i is the i th driving factor. The net output obtained is the sum of all the classifier outputs obtained multiplied with their respective driving factors. The output is:

$$O(d) = \Sigma\alpha_i X_i \quad , X_i = [-1,1]. \tag{2}$$

Where α_i is the driving factor of i th aspect and X_i is the output of i th classifier and d is the document under consideration. Now if:

O(d)<=0, d is classified as negatively oriented review.

O(d)>0, d is classified as positively oriented review.

Thus we have used a threshold score for the classification of the document.

III. DATASET, EXPERIMENTAL RESULTS AND PERFORMANCE

The dataset was acquired from the Large movie review dataset site of Stanford AI Lab [7][8]. The dataset was collected from IMDB and contains around 50,000 reviews, out of which 25000 are positive and 25000 are negative. Though there is no specific time span for review collection from IMDB, but it was ensured that no more than 30 reviews from a single movie get included in the final dataset. Because of even number of positive and negative reviews, so the minimum accuracy that we can obtain from the experiment is 50%. The dataset contains only highly positive and highly negative reviews. The authors of the dataset included a negative review only if it scored ≤ 4 out of 10 and included a positive review if it scored ≥ 7 out of 10 on a benchmark set by them [7]. Neutral reviews were omitted.It was seen that since the size of the reviews was varying and also since each aspect was commented on, in unequal number of sentences, the aspect based text separator separated the review into various aspects having unequal text distribution. It was also observed that in some of the reviews, all the aspects were not commented on. The score of the classifier for these cases were made 0. As mentioned in the previous section a Naïve Bayes classifier was used for classifying the separated aspect based text. The aspect based text given as input to individual classifiers was divided into training and testing data with a ratio of 70:30. The experiment was conducted with 1000 iterations, and during each iteration the driving factors were assigned random values following the constraints of (1). The driving factors which gave the highest accuracy were chosen as the best driving factors for the particular dataset under consideration.

The experiment conducted gave results as depicted in TABLE II given at the end of this paper.

The results in TABLE II depict the relationship between accuracy and the various driving factors used. We can see that the highest accuracy obtained was 0.79372 i.e. 79.372%. The factors corresponding to this value are Screenplay-0.07877, Music-0.11756, Acting- 0.28147, Plot-0.16390 Movie-0.31225, and Direction-0.108133.

Thus we can see from the factors that if we use the mentioned driving factors, we get an accuracy of 79.372%. This is the highest accuracy obtained using this method. Also it's worth noting that giving equal importance to all factors i.e. giving each a value of 0.165 has resulted in a lower accuracy of 78.268% than the highest accuracy obtained by unequal distribution of factors. Thus by changing the importance of that aspect, we can see its effect in the accuracy of the overall classification of the review. In the above case of 79.372% accuracy we have given most importance to the Movie, Acting and Plot aspects. Thus we can interpret from the results that in the reviews used from the dataset, the user has given more importance to these factors while writing the review. It also means that if the user tends to give a positive review towards these aspects then, due to their increased importance, the overall review will tend to be positive even if the user gives a negative feedback towards the other aspects. Giving more importance to certain factors also has an added advantage, it tends to suppress the users opinion about other factors. Suppose we have a review X and it contains user's opinion about two factors F1 and F2. Also the overall orientation of the review is positive in nature. The user has given a positive review about F1 and a negative about F2. Also the amount of text in the review for F1 aspect is less as compared to the F2 aspect. Now if we use any non-aspect based sentiment analysis method then since text size of F2 is greater than text size of F1 and also since F2 is negative in orientation, the overall document score will tend to reduce and skew towards negativity. On the other hand if driving factors are used and F1 is given more importance the document score will better reflect the positivity of the review. Since each aspect of a movie is analysed separately in this method, we can track the effect each aspect has towards the overall score of the document. This individual aspect based tracking can be used in a fined grained aspect based recommendation system, which recommends movies based on its various aspects instead of the overall rating of the movie. Also this method can be applied on a product review dataset thus enabling us to see what opinion each user has on the various aspects of the product, thus helping in the development of proper product placement strategy. It is very difficult to acquire such in-depth knowledge from the dataset using non-aspect based methods.

The various performance measures used were [4]:

Accuracy = (Total correctly classified documents / Total number of documents)

Precision = tp / (tp +fp)

Specificity = (tn / Total number of negatively oriented documents in the dataset)

Recall = (tp / Total number of positively oriented documents in the dataset)

Where tp, fp and tn are the true positives, false positives and true negatives obtained during the classification.

Table III. shows the data obtained by applying the various performance measures.

TABLE I. PERFORMANCE MEASURES

Sr. No	**Accuracy**	**Recall**	**Specificity**	**Precision**
1	0.79372	0.76568	0.82176	0.81117
2	0.78956	0.75888	0.82024	0.80848
3	0.78268	0.70512	0.86024	0.8345
4	0.77996	0.75176	0.80816	0.79669
5	0.76912	0.75192	0.78362	0.7787
6	0.7598	0.73184	0.78776	0.7751
7	0.74692	0.72312	0.77072	0.75926

8	0.7358	0.7156	0.756	0.74572
9	0.7254	0.6872	0.76368	0.74410
10	0.71812	0.68672	0.74952	0.73273

Some figures that denote the relationship between accuracy and various driving factors are given at the end of this paper.

IV. CONCLUSION AND FUTURE WORK

The experiment was conducted to find which movie aspects influence the orientation of the review using driving factors. It concluded with Movie, Acting and Plot aspects getting overall high driving factors and resulting in an accuracy of 79.372% for the current dataset in consideration. The importance of these aspects may or may not change, but since the experiments were conducted on a large dataset, it is quite unlikely that it will.

In this experiment we have found the importance of aspects of general sort of reviews. The future work can include finding driving factors for movie genre specific reviews. Reviews with different movie genre require unique analysis methods as Action movie, generally, demand good plot, screenplay, Adventure genre movies demand good screenplay, Music etc. Thus for scoring a document with more accuracy and for precise reflection of review orientation in the movie score, we intend to use different driving factors for different movie genre.

Also the current method used for classifying the text is Naïve Bayes Classifier which uses a bag-of-word approach. This approach doesn't consider the inter word meaning dependencies and also the context in which the word was used ie genre. For this purpose we tend to develop scoring method using context specific lexicon. Each word in the lexicon will have a different positive and negative score based on the context (genre) in which it was used. Also to incorporate the inter word dependencies we tend to use clause based scoring of a sentence. It scores each clause of a sentence individually and thus the overall sentence score is the sum of individual clause scores. Thus by coupling the above improved method with the use of genre specific driving factors we tend to obtain more refined scores for the movie reviews.

V. REFERENCES

[1] Tun Thura Thet, Jin-Cheon Na and Christopher S.G. Khoo , "Aspect-based sentiment analysis of movie reviews on discussion boards"., in Journal of Information Science,2010 36:823.

[2] Maite Taboada, Julian Brook, and Manfred Stede ,"Genre-Based Paragraph Classification for Sentiment Analysis" ,Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, pages 62–70,Queen Mary University of London, September 2009.

[3] Bo Pang and Lillian Lee, Shivakumar Vaithyanathan, "Thumbs up? Sentiment Classification using Machine LearningTechniques", Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, July 2002, pp. 79-86.

[4] V.K. Singh, R. Piryani, A. Uddin and P. Waila, "Sentiment Analysis of Movie Reviews A new Feature-based Heuristic for Aspect-level Sentiment Classification", Proceedings of the 2013 International Muli-Conference on Automation, Communication, Computing, Control and Compressed Sensing, Kerala-India, March 2013, IEEE Xplore, pp. 712-717.

[5]SentiWordNet, lexical resource for opinion mining,(http://sentiwordnet.isti.cnr.it/).

[6] J¨urgen Broß and Heiko Ehrig", Generating a Context-Aware Sentiment Lexicon for Aspect-Based Product Review Mining", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2010.

[7] Maas, Andrew L. and Daly, Raymond E. and Pham, Peter T. and Huang, Dan and Ng, Andrew Y. and Potts, Christopher, "Learning Word Vectors for Sentiment Analysis", in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, June 2011.

[8] Large Movie Review Dataset, Acquired from Stanford AI Lab: (http://ai.stanford.edu/~amaas/data/sentiment).

[9] Stanford Part-Of-Speech tagger, Stanford Natural Language Processing Group: (http://nlp.stanford.edu/software/tagger.shtml).

[10] Hanhoon Kang, Seong Joon Yoo, Dongil Han, "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews", Expert Systems with Applications 39, 2012.

[11] Abd. Samad Hasan Basaria, Burairah Hussina, I. Gede Pramudya Anantaa, Junta Zeniarjab, "Opinion Mining of Movie Review using Hybrid Method of Support Vector Machine and Particle Swarm Optimization", in Malaysian Technical Universities Conference on Engineering & Technology 2012, MUCET 2012 Part 4 - Information And Communication Technology.

[12] Jianxing Yu, Zheng-Jun Zha, MengWang, Tat-Seng Chua, "Aspect Ranking: Identifying Important Product Aspects from Online Consumer Reviews". Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 1496–1505,Portland, Oregon, June 19-24, 2011.

[13] Sentiment Analysis, Wikipedia – The Free Encyclopedia (http://en.wikipedia.org/wiki/Sentiment_analysis)

TABLE II. VARIATION OF ACCURACY WITH CHANGING DRIVING FACTORS

Accuracy	S-Df	M-Df	A-DF	P-DF	Mv-DF	D-DF
0.79372	0.07877495	0.1175615	0.2184798	0.1639036	0.3122561	0.1081334
0.78956	0.047883925	0.0046672	0.2291204	0.2001975	0.3032012	0.2136368
0.78268	0.165	0.165	0.165	0.165	0.165	0.165
0.77996	0.004460782	0.0132076	0.4391811	0.17665	0.3274761	0.0389702
0.76912	0.07476745	0.5480689	0.095835	0.1124063	0.1001353	0.0250248
0.7598	4.44E-05	0.0018357	0.4903199	8.41E-05	0.0429565	0.463557
0.74692	0.002677993	1.86E-04	0.9551522	0.0253268	0.0105148	0.0060463
0.7358	0.001611796	0.4900693	0.0181712	0.0030279	0.0108634	0.4731608
0.7254	0.003411332	0.0009225	0.0011156	0.490385	3.75E-04	0.5032535
0.71812	0.24075805	0.1980292	0.0024001	0.0037538	2.46E-04	0.5519468

Here S-DF = Screenplay Driving Factor, M-DF=Music Driving Factor, A-DF = Acting Driving Factor, P-DF = Plot Driving Factor, Mv-DF =Movie Driving Factor, and D-DF =Direction Driving Factor

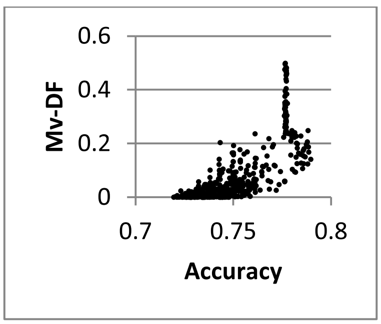

Figure 2. Accuracy Vs Mv-DF

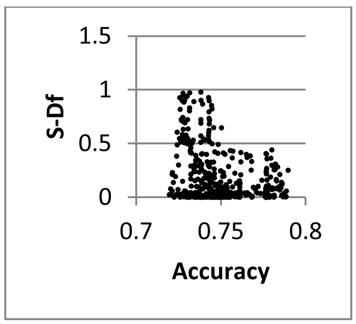

Figure 3. Accuracy Vs S-Df

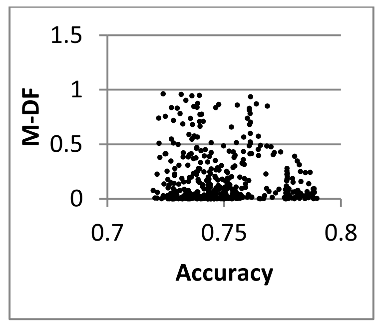

Figure 4. Accuracy Vs M-DF

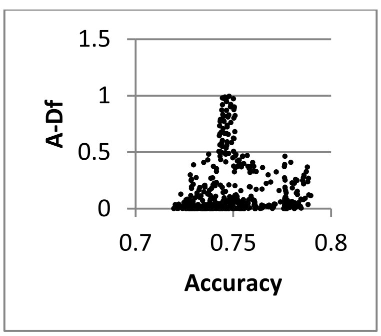

Figure 5. Accuracy Vs A-DF

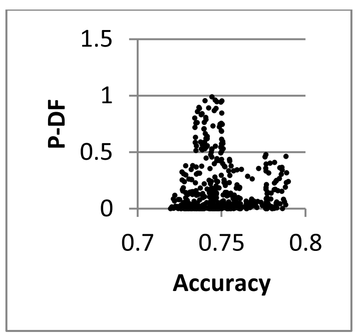

Figure 6. Accuracy Vs P-Df

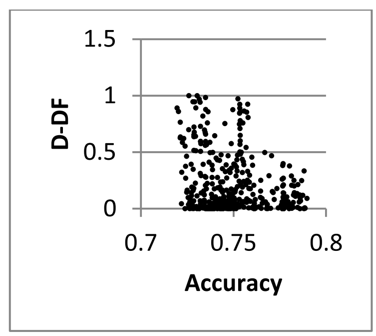

Figure 7. Accuracy Vs D-DF