Sentiment classification of review data using sentence significance score optimisation

Ketan Kumar Todi, S.N. Muralikrishna* and B. Ashwath Rao

Department of CSE,
Manipal Institute of Technology,
Manipal Academy of Higher Education,
Manipal, Karnataka-576104, India
Email: todiketan1996@gmail.com
Email: murali.sn@manipal.edu
Email: ashwath.rao@manipal.edu
*Corresponding author

Abstract: A significant amount of work has been done in the field of sentiment analysis in textual data using the concepts and techniques of natural language processing (NLP). In this work, unlike the existing techniques, we present a novel method wherein we consider the significance of the sentences in formulating the opinion. Often in any review, the sentences in the review may correspond to different aspects which are often irrelevant in deciding whether the sentiment is positive or negative on a topic. Thus, we assign a sentence significance score to evaluate the overall sentiment of the review. We employ a clustering mechanism followed by the neural network approach to determine the optimal significance score for the review. The proposed supervised method shows a higher accuracy than the state-of-the-art techniques. We further determine the subjectivity of sentences and establish a relationship between subjectivity of sentences and the significance score. We experimentally show that the significance scores found in the proposed method correspond to identifying the subjective sentences and objective sentences in reviews. The sentences with low significance score corresponds to objective sentences and the sentences with high significance score corresponds to subjective sentences.

Keywords: aspect; sentiment classification; clustering; neural network; optimisation; significance score.

Reference to this paper should be made as follows: Todi, K.K., Muralikrishna, S.N. and Rao, B.A. (2021) 'Sentiment classification of review data using sentence significance score optimisation', *Int. J. Data Analysis Techniques and Strategies*, Vol. 13, Nos. 1/2, pp.59–71.

Biographical notes: Ketan Kumar Todi is a student pursuing his BTech in Computer Science and Engg at Manipal Institute of Technology, Manipal. His research interests include natural language processing and machine learning. Currently, he is working as an intern at NTU Singapore.

S.N. Muralikrishna is working as an Assistant Professor (Senior Scale) in the Department of Computer Science and Engg at Manipal Institute of Technology, Manipal. He received his MSc in Computer Science from Mangalore University and MTech in Computer Science and Technology from the University of Mysore. His research interests include natural language processing, machine learning and computer vision.

B. Ashwath Rao is working as an Assistant Professor (Selection Grade) in the Department of Computer Science and Engineering at Manipal Institute of Technology, Manipal. He has done graduation and post-graduation in Computer Science and Engineering. His research interests include natural language processing, image processing and high-performance computing.

1 Introduction

Sentiment analysis (SA) is a classification task in which we try to predict the opinion given in the text. Sentiment classification can be broadly classified into three categories-namely document level, sentence level and aspect based. Document-level analysis assigns a polarity of sentiment considering all sentences in the document. A sentence-level analysis is similar to document level analysis but the analysis is at a finer sentence level. The third type is the aspect based, in which we try to identify the importance of various attributes in determining the sentiment of a given entity.

Sentiment analysis has a broad range of applications. It is important in analysing the review data of product, hotel, food, poem, movie and others. All these reviews play an important role for various entities. It helps the producer in determining how his/her product is being perceived by the customers and provides clues to further make the products more customer friendly. For the customers, it plays a very large part, as we often get influenced by the reviews of others when we have choices in deciding a product. The aspect-based analysis will help in all types of reviews. For example in a sentence, “The actors did a commendable work. The locations were a sight to behold” in a movie review, give us an inclination that the author likes the actors and locations in the movie. Hence, they give no opinion about the movie. Existing aspect-based methods identify the different aspects present in the review to find the overall sentiment based on a predetermined aspect significance. In this paper, we present a method which uses the cluster-based method to categorise the sentences in the review instead of focusing on aspect extraction. A supervised neural network is used to assign a significance score in the review at the cluster level. In the earlier example, a movie review may contain different aspects like the producers’ previous works, the songs, the shooting locations, analysis of the acting, and of course about the story of the movie. Though the previous descriptions do help, they have a somewhat lesser significance in deciding the sentiment as compared to the later.

Many models often use term frequency-inverse document frequency (TF-IDF) method where the frequency of the word is multiplied by the inverse of the frequency of the document in which the words are present. Most of the models use either the frequency count or binary count as a feature set for their words. These models generally exploit the information conveyed by the words in the set of documents and transform textual data to the numeric form. However, there are methods which use a set of keywords and their polarity to assign a sentiment score. However, these techniques do not consider the different aspects of the review.

In Section 2, we give an overview of the related work. The motivation for the work is presented in Section 3 and the proposed methodology in Section 4. This is followed by the results in Section 5 and finally conclusion and future work in Section 6.

2 Related works

A lot of work has been reported in the field of automatic sentiment analysis or opinion mining. There are multiple dimensions to the problem of sentiment analysis. Some of the very relevant problems in sentiment analysis are subjectivity classification, sentiment classification, measuring the usefulness of a review, aspect and polarity identification, opinion spam detection, etc. (Ravi and Ravi, 2015). Subjectivity classification refers to the problem of identifying whether a sentence conveys an opinion or a fact. Identifying subjectivity and objectivity helps in improving the performance of the system. Several methods have been proposed in the literature to address this challenge (Chaturvedi et al., 2018). These methods use machine learning or lexicon-based approaches. Some of the methods in the literature review incorporate contextual cues to improve the accuracy of the system. Sentiment classification refers to the problem of identifying the polarity of the review. This may be done at the document level or at the sentence level. Several machine learning techniques have been used extensively in the sentiment classification. Most of the methods use conventional bag-of-word (BoW) model or TF-IDF to represent the features (Zhang et al., 2018). A number of classifiers and their fusion have been used in the past to address sentiment classification. The major limitation of the BoW model is that it does not consider the word order and also the semantics of the sentences. The improved methods use additional linguistic information like part-of-speech (POS), opinion lexicon, bag-of-ngrams, parse trees, etc. to capture additional information in the sentences. The work presented in Penalver-Martinez et al. (2014) is an example where the system uses natural language processing pipeline for sentiment classification. The sentence splitter, POS tagger, lemmatiser are used in the initial step. Ontology-based features of the word are identified for a given domain. This feature set is used to identify the polarity using a score function. Pang et al. (2002) carried out an extensive work on sentiment analysis. They carried out the experiments on the basis of both frequency and presence count, and by using unigram and bigram features. They also used POS tags and adjectives for classification and performed experiments with three different models naive Bayes, support vector machine (SVM), and maximum entropy, with SVM achieving the best results using the word and its position in the document as a feature. For calculating the value of position feature the document was divided into three parts. This paper inspired us to consider the fact that the sentiment of the review as a whole might be different from the analysis of the movie review.

Like Pang et al. (2002), GangLi Li and Liu (2010) also used part-of-speech POS tag, for classification but they used only adjectives and adverbs for predicting the sentiment of the document. A BoW model with frequency vector and presence vector for every word is used to obtain the clusters representing the positive and negative sentiments. They achieved an accuracy of 75.8% accuracy initially. Then TF-IDF was carried out to get a weight vector corresponding to the words, and the frequency and presence vectors were clustered again by assigning it a weight from the weight vector. The variance between the best and lowest accuracy was high. So they used WordNet to find terms that were closer to either 'good' or 'bad', and they also found out the distance of the terms from these two reference words. This was helpful in reducing the size of the feature vector. The frequency vectors of the words and these weights from WordNets are used for clustering and the rest of the neutral words are assigned a weight of 0, thereby achieving the best results.

The aspect-based methods identify various aspects of the review and find the polarity of the aspect. In a product review the different features of the products are the aspects. Aspect-based method helps in ranking the product based on the overall polarity score (Kumar and Abirami, 2018). Part-of-speech POS tags have been widely used for sentiment analysis. Singh et al. (2013) used adjectives, adverbs, and verbs as the important features for predicting the sentiment in movie reviews. In their first model, they use adjectives and adverbs for classifying the sentiment of the document. They use SentiWordNet for getting the scores of the words. If the adverb was affirmative the adjective score was enhanced and vice versa and the sum of all the scores was the score of the entire document. Then in their second model, they used adverb and verb combination in addition to adjective and adverb combination in a similar manner predicting a score, where the adverb-verb score was given less importance than adjective-adverb score. In the aspect-based model they identify different aspects in a movie review and scan each sentence to determine if it contains the keyword of an aspect. If the keyword is present then the above adjective-adverb and verb-adverb model was used to find a score, for that opinion. The aspect-based method just reflected the performance of the movie on various parameters. They made two different models for +ve and -ve classes and reported two different accuracies.

There have been several attempts to use different types of classifiers. Some of them are naive Bayes, support vector machines (SVM), MaxEnt (maximum entropy), artificial neural networks. A combination of SVM and neural network was used by Moraes et al. (2013). They tried taking different numbers of words as features by computing information for the words. They achieved the best results with word features using a neural network with hidden layers. Fuzzy logic and concept analysis have been used by Li and Tsai (2013) to extract information at the word level. They used it to determine the relation between objects and their attributes. This helps to re-weight the feature on the basis of the concept it is connected to. Fuzzy logic helps in the better construction of the concept graph and handling of uncertainty. The attributes of objects are identified and an (object, attribute, relation) tuple is used for graph formation. They achieved an F-score of 88.59% using fuzzy logic with formal concept analysis (FCA).

The machine learning techniques provide various mechanisms to combine the information at various levels in the classification process. The fusion may take place at the feature level or at the classifier level. The voting-based classifiers fusion is one of the common ways to combine the classifiers. The work presented in Balazs and Velásquez (2016) discusses the various techniques of information fusion for opinion mining. The data that is acquired for opinion mining can be from various modalities and sources. Ontologies and lexicons can be combined in the applications to improve the performance of opinion mining system. In Soleymani et al. (2017), a theoretical framework to fuse information at various levels of sentiment classification is given. The importance of present-day research in the field of sentiment analysis is given in Mäntylä et al. (2018). This work presents a comprehensive review of sentiment analysis and the growth in the number of publications of related articles. Also, it discusses the types of data available in various application domains, the taxonomy of sentiment analysis and the influence of computer-assisted sentiment classification. In Serrano-Guerrero et al. (2015), authors have presented different web services available to perform sentiment analysis. The most popular web services available for sentiment classification are presented in this paper. The best result accuracy reported is 80.4% with the SentimentAnalyzer web-service on movie review dataset.

Recent works in sentiment analysis focus on detecting neutral sentiments. This is also called as the subjectivity classification. This helps in improving the accuracy of the overall system. Handling neutral reviews is critical and they are ignored in many cases. In Valdivia et al. (2018), different methods of sentiment analysis are applied to the data and then a consensus function is used to filter the results. A weighted aggregation of classifiers is used to predict the subjectivity.

3 Motivation

We have drawn inspiration from the work presented in Pang et al. (2002). Normally, any product review contains information about various other aspects in addition to the product. So the review contains information about other aspects which are not so relevant. For example, “This film should be brilliant. It sounds like a great plot.....I hate the Spice Girls..... The theatre is really, really bad.....”. In this example, a bag-of-feature model fails as there are words which are indicative of sentiments. However, a human would be able to determine the sentiment by looking at the focus of each sentence. The review is about the movie and the words in the sentence contributing to the sentiment of movie review have a higher significance than the words referring to other entities such as theatre or Spice Girls. So, we use an aspect-based model, in a broader sense, where each sentence is given a certain level of importance depending on the nature of the sentence. There are limited works done in this direction. Unlike any of the existing methods, we formulate this as a problem of finding an optimal significance score to each of the sentences in the review so as to get the correct sentiment label using a training set. This is detailed in the following section.

3.1 Problem formulation

Let us consider a set of review documents used for training $D = \{D_1, D_2, D_3....D_N\}$ along with their labels $O_i \in \{+1, -1\}$, for $1 \leq i \leq N$, where $\{+1, -1\}$ indicate the positive and the negative sentiments of the document respectively. Each document D_j, $1 \leq j \leq N$, is composed of a set of sentences $\{S_{j1}, S_{j2}, ...S_{jn_j}\}$. Any sentence $S_{ij} \in D_i$, $1 \leq i \leq N$ and $1 \leq j \leq n_i$, will have a set of words denoted by $\{w_{ij1}, w_{ij2}, w_{ij3}...w_{ijm}\}$. This set of words in the sentence S_{ij} is encoded as a vector W_{ij} by averaging the individual word embeddings called the sentence embedding. As explained in the example, the sentences in the document may correspond to the sentiment of various aspects such as the movie, theatre, Spice Girls, etc. However, our interest is to compute the sentiment score for the movie and other aspects are of lesser significance. So in the proposed method, we cluster the embedded sentences into K partitions, $\{C_1, C_2, C_3, ...C_k\}$. The value of k is a user-defined parameter. Let $N_s = \sum_{i=1}^{N} n_i$ be the total number of sentences in D. The set of all sentences in the review $S = \{S_{ij}\}$ is represented as $S = \{W_{ij}\}$ of some predefined dimension L. S is partitioned into K partitions. In the experiments, hard clustering is applied so that a sentence belongs to only one of the clusters. In the proposed method, the idea is to give higher importance to the sentence which is necessarily contributing to the main aspect of the review and lesser importance to the sentences which are focusing on other aspects and are not that significant to the review. The partitioning helps us to associate each sentence to a cluster and we can assign each cluster with a value indicating the importance of the cluster in sentiment analysis. So, we use k weight values $\{v_1, v_2, ...v_k\}$ to indicate the importance of each sentence in the document. Now it can be seen that aspect-based sentiment analysis can be seen as a maximisation problem where we try to maximise the following objective function

$$\mathbf{argmax} \sum_{i=1}^{N} \sum_{j=1}^{n_j} v_k * X(S_{ij}, k), \quad \forall k \tag{1}$$

where the membership function X is defined as in equation (2).

$$X(t, k) = \begin{cases} 1, & t \in C_k \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

4 Methodology

The overall methodology followed in the proposed technique has two stages. The first stage is to identify the different aspects present in the review using clustering. The second stage is to find a significance score to each of the sentences based on to cluster they belong to and this depends on the aspect. This is shown in Figure 1.

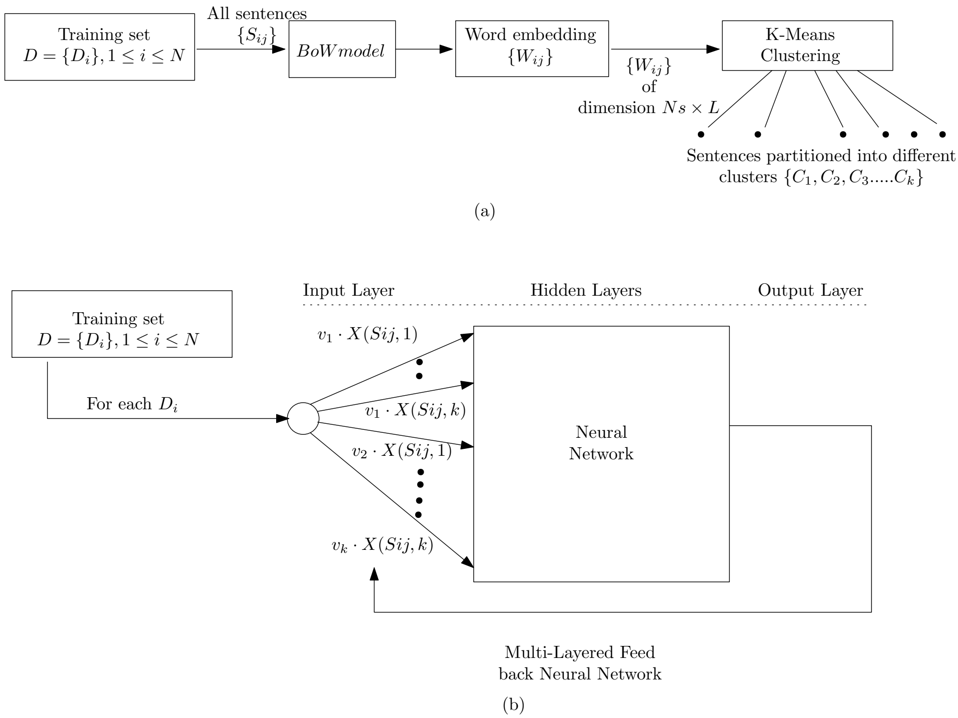
4.1 Binary weight assignment

In the first approach we used binary weight assignment for each of the clusters, i.e., after assigning a cluster label to a sentence in the review, the significance score 0 indicates that the sentence of that cluster doesn't play any role in the overall sentiment of the review. For example, let us assume we have $k = 4$ clusters, so the possible cluster weight assignment is 2^4 binary weight values. If the binary weight values are $\{1, 1, 0, 1\}$ then we will pick only those sentences which have a label either 0, 1 or 3 for predicting the sentiment of the document.

4.2 Optimally calculated weights

The second strategy is to use optimally computed weight by relaxing the condition of binary weight values. So instead of just using a sentence or completely ignoring it, we use a neural network to optimise the significance score to be assigned to the sentences in the review. A multi-layer feedback neural network as shown in Figure 1 is used. Here, we try to find the optimal values of v_i for $1 \leq i \leq k$ in equation (1). Only two epochs were used for training as multiple epochs caused overfitting. The dataset is divided into training set and test set.

Figure 1 Overview of the methodology, (a) representation of sentences using word embedding and clustering the sentences with K-means algorithm (b) a multi-layered neural network to find the optimal significance score v_i for every cluster



Note: The BoW model is not shown in the figure.

In the implementation, initially, the sentences in the training set are vectorised using BoW model. Then GloVe word embedding (Pennington et al., 2014) is used to obtain the sentence embedding by averaging the individual word embeddings. The sentences from each document of the training set are clustered using the K-means algorithm. After the clustering, a BoW model is used to represent the sentences in the documents, i.e., all the sentences belonging to a particular i^{th} cluster is represented using a BoW model. Depending upon the number of clusters k, one can get k vectors each having dimension same as vocabulary size. Finally, the vectorised representation of each cluster is passed to the neural network to obtain the significance score using the training set. The neural network finds the optimal values for v_i. Further, the same procedure mentioned above is used for the test set to measure the accuracy.

4.3 Significance score correspondence

In any review there can be sentences which are either subjective or objective in nature. The objective sentences represent the facts and subjective sentences represent the opinion. So in recent works there have been attempts to identify the objective sentences and ignore them in finding the overall sentiment. We claim that our method corresponds to identifying the subjective/objective sentences in a given review.

To show the correspondence in our method, we must show that a cluster with a high significance score has sentences that are subjective in nature. A low significance score to any cluster must have a larger number of objective sentences in it than subjective sentences. We use the method given in Liang and Zhang (2016) to experimentally show that our method corresponds to finding subjective sentence and objective sentence. The method given in Liang and Zhang (2016) uses a Bi-directional Recurrent Neural Network for subjectivity classification with an accuracy of over 90%. This neural network model uses the topology given in Augenstein et al. (2016). We define the conditional probabilities to show the evidence of our claim. We show that a subjective sentence being scored high is more obvious than an objective sentence being scored high. Similarly, an objective sentence being scored low is more obvious than a subjective sentence scored low. We use the conditional probability given in equations (3) and (4) as the verification. We define S as the event of a sentence being subjective and O the event of a sentence being objective. H is the event of a sentence being in a cluster with the highest score and L is the event of a sentence being in a cluster having the least score. The results of these experiments are in the following section.

$$P(H|S) \gg P(H|O) \tag{3}$$

$$P(L|O) \gg P(L|S) \tag{4}$$

5 Experiments and results

The experiments were conducted to obtain the optimal significance score using clustering and neural network-based optimisation. Further, the experiments were conducted to prove the correspondence between the proposed significance score.

5.1 Finding optimal significance score

We used *sklearn* k-means algorithm for the first stage. The number of clusters was set as $k = 4$, looking at the nature of the training data. The rest of the parameters were initialised with their default values - the maximum number of iterations as 300, the number of various initial seeds to be tried as 10, and the relative increment in results for convergence to have taken place as 0.0001. The clustering process was the same for both - the optimally assigned weights and the binary assigned weights.

We used the Udacity IMDB movie review dataset (https://github.com/udacity/deep-learning/blob/master/sentiment-network/reviews.txt) consisting of 25k reviews consisting of an equal number of positive and negative reviews. K-fold cross-validation was used consisting of only 1k test sentences and the rest of them were used for training. In the work presented in Pang et al. (2002) authors have used 2K reviews in total, but here we are using a larger review dataset. Generally, neural network models use a sufficiently large training set and we have used a similar setup in our experiments. All the sentences of a single cluster were used to predict the sentiment of the sentences in that cluster, and then an optimal weight for each cluster was found using the neural network.

The words of all the documents were taken and arranged in a non-increasing order on the basis of the frequency count. The frequency of the words was used to form the vocabulary, for the bag of words model, which is used as the feature set for predicting the sentiment. In the binary weight assignment approach, the vocabulary of size 20,000 words was considered. Each sentence in the BoW representation was embedded to a 50-dimensional vector using the GloVe word embedding. In the optimised weight assignment, we considered the vocabulary size from 10,000 words to 30,000 words with an interval of 5,000 words. We followed a mini-batch training approach in which the entire dataset was divided into small fixed sized parts, and back-propagation was carried out after each part instead of being carried out after each example, as it helps in faster convergence. The experiment was conducted with batch sizes of 16, 32 and 64. The results of the experiments are presented in Tables 1 and 2.

Table 1 Accuracy of the sentiment analysis using binary weight assignment scheme (word embedding size L = 50)

Clusters taken	*Accuracy*	*Clusters taken*	*Accuracy in %*
0	81.25	1, 3	78.12
1	75.30	2, 3	77.72
2	73.89	0, 1, 2	89.52
3	63.51	0, 1, 3	87.90
0, 1	63.51	0, 2, 3	87.40
0, 2	85.48	1, 2, 3	84.68
0, 3	87.20	0, 1, 2, 3	88.21
1, 2	82.66		
Average accuracy in %			*80.42333333*

Table 2 Accuracy of the sentiment analysis using neural network model with different batch sizes

Vocabulary size in BoW model (in thousands)	*Accuracy in % (word embedding size L = 300)*		
	Batch size = 16	*Batch size = 32*	*Batch size = 64*
10	88.91	89.21	88.87
15	86.19	88.41	90.45
20	90.02	88.41	89
25	88	88.41	88
30	88.81	88	*90.93*
Averages accuracy	88.386	88.488	89.45
Overall average			*88.77467*

We compare the results with the SVM model proposed by Pang et al. (2002) which used the presence or absence of the unigram words as the features for their SVM model. The experiment was conducted using same amount of 25k IMDB dataset used in this paper. The SVM model achieved an accuracy of 88.07%. The results were better than the ones mentioned in the paper due to the model being trained on a much larger dataset of 25k reviews. The Joachim SVM as mentioned in Pang et al., with default parameter values were used for training and testing. The proposed model performs better with an average accuracy of 88.77% and a maximum accuracy of 90.93%.

Table 3 A summary of the performance of some of the existing techniques

Reference	*Description*	*F-measure accuracy (in%)*	*Dataset*
Bai (2011)	Predicting consumer sentiments from online text and Bayesian network and Markov model	78.00	IMDB movie review and others
Li and Liu (2010)	A clustering-based approach on sentiment analysis. Clustering based TF-IDF method	78.33	IMDB movie review
Li and Tsai (2013)	A fuzzy conceptualization model for text mining with application in opinion polarity classification and fuzzy formal concept analysis	88.59	Reuters news stories and IMDB movie review
Moraes et al. (2013)	Document-level sentiment classification: an empirical comparison between SVM and ANN	86.5	IMDB movie review
Pang et al. (2002)	Thumbs up (base paper) and SVM based methods	80	IMDB movie review
Singh et al. (2013)	Sentiment analysis of movie reviews and POS tagging based methods	78	IMDB Movie review

5.2 Correspondence to subjectivity identification

The result of the clustering and the significance score was used to establish the relationship between subjectivity/objectivity and the significance score values. The sentences in the cluster with the highest significance score and with the least score were subjected to the subjectivity classification method given in Liang and Zhang (2016). The result of the experiment is given in Tables 4 and 5. The cluster with the highest significance score had the majority of the sentences from subjective class. Similarly the cluster with the least significance score had majority of the sentences from objective class. We can observe this in the results. This is a strong evidence to our claim that the highest significance score contributes to the overall sentiment and the majority of them are subjective in nature. Also, for the cluster with the lowest significance score the sentences are objective in nature and the contribution of them is least in the overall sentiment. The result of verification is given in Table 6. As explained in Section 4, we used Bayes' conditional probability to show that, given a sentence, the probability of getting a high score to a subjective sentence S denoted by $P(H|S)$ is more when compared to getting a high score to an objective sentence O denoted by $P(H|O)$. Similarly, the conditional probability of getting a low score to an objective sentence $P(L|O)$ is more when compared to a subjective sentence getting a low score denoted by $P(L|S)$.

Table 4 The table showing the number of subjective sentences and number of objective sentences in the cluster with the highest significance score

Trial no.	*No. of objective sentences*	*Percentage of objective sentences (%)*	*Percentage of subjective sentences (%)*	*The percentage of sentences belonging to the cluster with most significant score out of total sentences (%)*
1	5,100	7.95	92.05	21.39
2	5,097	7.94	92.06	21.42
3	5,199	8.10	91.90	21.42
4	5,109	7.97	92.03	21.38
5	8,003	9.78	90.22	27.31
6	5,099	7.94	92.06	21.42
7	8,171	9.98	90.02	27.32
8	8,025	9.80	90.20	27.31
		8.99	*91.01*	*24.61*

Notes: The total number of subjective sentences in the dataset is 71.55% and objective sentences is 28.45%. Total number of sentences considered is $\approx 300K$.
Iterations = 11, epochs = 2.

Table 5 The table showing the number of subjective sentences and number of objective sentences in the cluster with the least significance score

Trial. no	*No. of objective sentences*	*Percentage of objective sentences (%)*	*Percentage of subjective sentences (%)*	*The percentage of sentences belonging to the cluster with least significant score out of total sentences(%)*
1	33,703	56.88	43.12	19.77
2	33,787	56.98	43.02	19.78
3	33,799	57.01	42.99	19.78
4	33,749	56.88	43.12	19.80
5	33,750	56.89	43.11	19.79
6	38,358	40.58	59.42	31.54
7	33,577	56.90	43.10	19.69
8	38,320	40.60	59.40	31.49
		52.84	*47.16*	*22.35*

Notes: The total number of subjective sentences in the dataset is 71.55% and objective sentences is 28.45%. Total number of sentences considered is $\approx 300K$.
Iterations = 11, epochs = 2.

Table 6 Results of verification using equations (3) and (4)

Conditional probability	*Bayes' rule*	*Value*	*Conditional probability*	*Bayes' rule*	*Value*
$P(H\|S)$	$\frac{P(S\|H) * P(H)}{P(S)}$	*0.31*	$P(H\|O)$	$\frac{P(O\|H) * P(H)}{P(O)}$	*0.07*
$P(L\|O)$	$\frac{P(O\|L) * P(L)}{P(O)}$	*0.4151*	$P(L\|S)$	$\frac{P(S\|L) * P(L)}{P(S)}$	*0.1473*

6 Conclusions and future work

The sentence significance score method helps to improve the accuracy of the system. The average accuracy of the proposed method is 88.77% and a maximum accuracy of 90.93%. This clearly outperforms the methods listed in Table 3. We experimentally showed that the majority of the sentences with low significance score are objective in nature and the high significance score corresponds to subjective sentences. The correspondence is evident from the results, however, further investigations are required to derive a conclusive proof of the complete correspondence using statistical measures. The results of the experiment conducted using the binary weight assignment indicates that certainly, the different aspects in the review play an important role in the overall sentiment. This can be seen looking at the maximum accuracy using only the sentences from clusters 0, 1 and 2. Forming a bag-of-word representation of the sentences in a cluster might have lead to the loss of sequential information. So using semantic representation may improve the result, with rest of the architecture being same. Another possible improvement is to identify the number of different aspects of the review. Instead of fixing the number of clusters one can find the optimal number of clusters using the available techniques. However, finding the true sentiment in a review is always challenging due to other factors such as reviewers credibility, mood, etc. This poses a few more additional challenges in this field.

References

Augenstein, I., Rocktäschel, T., Vlachos, A. and Bontcheva, K. (2016) *Stance Detection with Bidirectional Conditional Encoding*, arXiv preprint arXiv:1606.05464.

Bai, X. (2011) 'Predicting consumer sentiments from online text', *Decision Support Systems*, Vol. 50, No. 4, pp.732–742.

Balazs, J.A. and Velásquez, J.D. (2016) 'Opinion mining and information fusion: a survey', *Information Fusion*, January, Vol. 27, No. C, pp.95–110.

Chaturvedi, I., Cambria, E., Welsch, R.E. and Herrera, F. (2018) 'Distinguishing between facts and opinions for sentiment analysis: survey and challenges', *Information Fusion*, November, Vol. 44, pp.65–77.

Kumar, A. and Abirami, S. (2018) 'Aspect-based opinion ranking framework for product reviews using a spearman's rank correlation coefficient method', *Information Sciences*, September, Vol. 460, pp.23–41.

Li, G. and Liu, F. (2010) 'A clustering-based approach on sentiment analysis', *2010 International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp.331–337. IEEE.

Li, S-T. and Tsai, F-C. (2013) 'A fuzzy conceptualization model for text mining with application in opinion polarity classification', *Knowledge-Based Systems*, February, Vol. 39, pp.23–33.

Liang, D. and Zhang, Y. (2016) *AC-BLSTM: Asymmetric Convolutional Bidirectional LSTM Networks for Text Classification*, arXiv preprint arXiv:1611.01884.

Mäntylä, M.V., Graziotin, D. and Kuutila, M. (2018) 'The evolution of sentiment analysis – a review of research topics, venues, and top cited papers', *Computer Science Review*, February, Vol. 27, pp.16–32.

Moraes, R., Valiati, J.F. and Neto, W.P.G. (2013) 'Document-level sentiment classification: an empirical comparison between SVM and ANN', *Expert Systems with Applications*, Vol. 40, No. 2, pp.621–633.

Pang, B., Lee, L. and Vaithyanathan, S. (2002) 'Thumbs up?: sentiment classification using machine learning techniques', *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Vol. 10, pp.79–86, Association for Computational Linguistics.

Penalver-Martinez, I., Garcia-Sanchez, F., Valencia-Garcia, R., Rodriguez-Garcia, M.A., Moreno, V., Fraga, A. and Sanchez-Cervantes, J.L. (2014) 'Feature-based opinion mining through ontologies', *Expert Systems with Applications*, Vol. 41, No. 13, pp.5995–6008.

Pennington, J., Pennington, J., Socher, R. and Manning, C.D. (2014) 'Glove: Global vectors for word representation', *EMNLP*.

Ravi, K. and Ravi, V. (2015) 'A survey on opinion mining and sentiment analysis: tasks, approaches and applications', *Knowledge-Based Systems*, November, Vol. 89, pp.14–46.

Serrano-Guerrero, J., Olivas, J.A., Romero, F.P. and Herrera-Viedma, E. (2015) 'Sentiment analysis: a review and comparative analysis of web services', *Information Sciences*, August, Vol. 311, pp.18–38.

Singh, V.K., Piryani, R., Uddin, A. and Waila, P. (2013) 'Sentiment analysis of movie reviews: a new feature-based heuristic for aspect-level sentiment classification', *2013 International Multi-Conference on Automation, Computing, Communication, Control and Compressed Sensing (iMac4s)*, pp.712–717, IEEE.

Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S-F. and Pantic, M. (2017) 'A survey of multimodal sentiment analysis', *Image and Vision Computing*, September, Vol. 65, pp.3–14.

Valdivia, A., Luzón, M.V., Cambria, E. and Herrera, F. (2018) 'Consensus vote models for detecting and filtering neutrality in sentiment analysis', *Information Fusion*, November, Vol. 44, pp.126–135.

Zhang, L., Wang, S. and Liu, B. (2018) 'Deep learning for sentiment analysis: a survey', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, July–August, Vol. 8, No. 4, p.e1253.