Aspect context aware sentiment classification of online consumer reviews

Barkha Bansal
The NorthCap University, Gurugram, India, and
Sangeet Srivastava
Wenzhou-Kean University, Wenzhou, China

Abstract

Purpose – Aspect based sentiment classification is valuable for providing deeper insight into online consumer reviews (OCR). However, the majority of the previous studies explicitly determine the orientation of aspect related sentiment bearing word and overlook the aspect-context. Therefore, this paper aims to propose an aspect-context aware sentiment classification of OCR for deeper and more accurate insights.

Design/methodology/approach – In the proposed methodology, first, aspect descriptions and sentiment bearing words are extracted. Then, the skip-gram model is used to extract the first set of features to capture contextual information. For the second category of features, cosine similarity is used between a pre-defined seed word list and aspects, to capture aspect context sensitive sentiments. The third set of features includes weighted word vectors using term frequency-inverse document frequency. After concatenating features, ensemble classifier is used using three base classifiers.

Findings – Experimental results on two real-world data sets with variable lengths, acquired from Amazon.com and TripAdvisor.com, show that the advised ensemble approach significantly outperforms sentiment classification accuracy of state-of-the-art and baseline methods.

Originality/value – This method is capable of capturing the correct sentiment of ambiguous words and other special words by extracting aspect-context using word vector similarity instead of expensive lexical resources, and hence, shows superior performance in terms of accuracy as compared to other methods.

Keywords Aspect context sentiment, Consumer opinion, Ensemble methods, Contextual information

Paper type Research paper

1. Introduction

Consumers often tend to share the positive, neutral or negative opinions about their purchased products and various related aspects of the product through different platforms such as online blogs, shopping websites, microblogs and WOM (word of mouth) (Lan *et al.*, 2018). This leads to generation of a vast amount of online consumer reviews (OCR) on all such platforms (Li *et al.*, 2019; Wang *et al.*, 2017). Valuable information can be extracted from these unstructured OCR on various important topics such as products and services, which is beneficial for both consumers and brands (Liu *et al.*, 2017). One popular way to extract information and gain insight into consumer opinion is sentiment analysis. Sentiment analysis has become increasingly popular because of open challenges, new application areas and outstanding benefits in marketing, political campaigns, election monitoring, financial predictions and other important tasks (Han *et al.*, 2019). It can be performed at various levels of granularities, one of the recently popularized methods being: aspect-based sentiment analysis.

Aspect based sentiment analysis emphasizes on detecting aspects (of product) and their corresponding sentiments. For instance, "New iPhone has a great battery life," has an aspect battery-life and related positive sentiment, depicted by sentiment bearing word Great. The task of detecting such aspect sentiment pairs in the unsupervised setting has been performed efficiently by many previously proposed methods. However, the unstructured nature of OCR makes it more challenging to extract implicit aspects and related sentiments. Specifically, the majority of OCR comes in the form of informal text, which includes noise, ambiguous or polysemous words, abbreviations, misspelled words and constantly generated new slangs. The issue of ambiguity and widely generated new words can be solved by capturing aspect-context. Though, only a few studies have focused on aspect-context aware sentiment classification methods. For instance, in the review "New phone's price is cheap," sentiment bearing word cheap expresses positive sentiment for the aspect Price, but when

The current issue and full text archive of this journal is available on Emerald Insight at: **https://www.emerald.com/insight/2398-6247.htm**

Information Discovery and Delivery
 [ISSN 2398-6247]
[DOI 10.1108/IDD-12-2019-0089]

Compliance with ethical standards

Funding: The authors received no specific funding for this work.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Received 27 December 2019
Revised 3 March 2020
17 March 2020
Accepted 19 March 2020

used in conjunction with the aspect screen, it will depict negative sentiment. Hence, for more precise sentiment classification, it is important to capture more implicit word relations and contextual information from the texts. For this purpose, we use state-of-the-art continuous space language models (Mikolov *et al.*, 2013b; Mikolov *et al.*, 2013c; Mikolov *et al.*, 2013a) to extract distributed word vectors. Specifically, we use the skip-gram shallow neural networks for capturing contextual relationships in the text. The advantage of this method over previous methods is that it exhibits less time complexity as compared to previous models (Mikolov *et al.*, 2013b).

Continuous space language models, pioneered by Deerwester *et al.* (1990), Rumelhart *et al.* (1986), Elman (1990), Elman (1991) and Hinton *et al.* (1986), are the language models where words are projected to a continuous space. In this paper, we use continuous vector representations (Schwenk, 2016) that have shown excellent results across various application areas including sentiment classification, both coarse (binary) and fine-grained (e.g. 5 way) (Le and Mikolov, 2014) . Distributed representations are based on the hypothesis that similar linguistic items (i.e. with similar meaning) are projected closer in the vector space and vice versa. For example, with properly trained distributed vector representations, we can have vector operations as follows: DV (France) - DV(Paris) + DV(Rome) $\cong$ DV(Italy); where DV is distributed vector (Yang *et al.*, 2018). Hence, distributed vector representations are capable of capturing semantics, contextual information and, are very important for similarity tasks. While non-distributed vectors are not capable of capturing semantics, word or document meaning, similarities, etc.

Moreover, to best extract information from OCR, we incorporate two more sets of features in our study. First, POS (part-of-speech) tagging is used and sentiment features are extracted using the cosine similarity measure to best capture aspect context aware sentiments (Li *et al.*, 2019; Poria *et al.*, 2016; Topal and Ozsoyoglu, 2017). Secondly, traditional non-distributed method term frequency-inverse document frequency (TF-IDF) is used for extracting the weighted word vectors. Finally, the ensemble of these features is followed by classification using an ensemble approach. Main contributions of this study can be summarised as follows:

- This paper proposes a novel method to capture aspect-context for solving the issue of ambiguous sentiment bearing words.
- Domain independent seed words are used instead of expensive lexicons. One of the disadvantages of these lexicons is that they are not comprehensive enough, especially, in the case of newly generated slangs, abbreviated words and other informal words.
- Finally, ensemble features are used to retain maximum information and ensemble classifiers result in superior classification.

The rest of the paper is structured as follows. Section 2 details the related work, Section 3 describes data and methodology, Section 4 analyses the results. Finally, Section 5 concludes our work and Section 6 shows limitations and future work.

2. Related work

Today, online users are keener to express their opinion or feedback on social media and e-commerce websites rather than traditional surveys and forms, hence, OCR mining techniques become imperative. Therefore, researchers, brand managers and decision-makers have realized the importance of intelligent systems for mining online consumer-generated data to gain valuable insights in a less expensive way. Many studies have been proposed to derive useful information from OCR, which can be beneficial for both the consumer and the brand. For example, consumer reviews in the form of tweets were explored by Ibrahim *et al.*, with the aim of deriving the most popular topics, specifically, issues, as discussed by users (Ibrahim and Wang, 2019). This may help serve the need of a consumer quickly and efficiently while improving a brand's image.

One way to gain insight into consumer reviews is to extract popular product aspects and related consumer sentiment. Further, deriving implicit aspects and contextually sensitive sentiment may produce more reliable and accurate results. Therefore, we use the continuous space language model to capture implicit relations. Recently, shallow neural networks for distributed word vector representations have been proposed, namely, continuous bag of words (CBOW) and continuous skip-grams (SG) (Mikolov *et al.*, 2013b; Mikolov *et al.*, 2013c; Mikolov *et al.*, 2013a). These methods are more efficient as compared to previously proposed deep learning methods in terms of less time complexity. CBOW uses continuous distributed representation as opposed to the BOW (bag of words: non-distributed) model and is similar to the feed-forward neural network language model (NNLM) (Bengio *et al.*, 2003). It predicts a word when surrounding words are given, whereas skip-gram predicts surrounding words (in a particular range) for a given input word. Increasing the range (neighbouring window size) can increase the accuracy but also computational complexity (Mikolov *et al.*, 2013b). Skip-gram can learn from small training sample and is able to accurately represent infrequent words and phrases in a corpus. CBOW is shown to learn quicker than skip-gram, and, is more accurate for data with recurrent words (Giatsoglou *et al.*, 2017).

Following the word vector representations, paragraph (or document) representations were introduced (Le and Mikolov, 2014). Paragraph vectors (PV) are fixed length representations of texts of variable-length. This unsupervised algorithm considers word order like an n-gram model (where *n* is large) but is better than the bag of n-grams where the feature dimensions are very high. In their paper, Le and Mikolov (2014), devise two paragraph-vector models: distributed memory (PV-DM) and distributed bag of words (PV-DBOW). According to the authors, PV-DM alone can achieve comparable accuracy to the combination of both, although the combination is superior for the task of classification (Le and Mikolov, 2014). The word vectors and paragraph vector in PV-DM can be averaged, summed or concatenated. However, distributed representation algorithms are computationally intensive and expensive as compared to non-distributed algorithms. Distributed vectorization methods are also known to work well with the large data sets, whereas non-distributed methods are capable of handling both small and large data sets. Nevertheless, this paper focuses on deriving aspect context sentiments, which cannot be achieved using only non-distributed methods.

Word vector representations have been recently used for the sentiment classification of sentences, documents, tweets and reviews but not widely applied across different domains (Abdelwahab and Elmaghraby, 2016; Giatsoglou *et al.*, 2017; Jiang *et al.*, 2016a, 2016b; Maas *et al.*, 2011; Zhang *et al.*, 2015). Moreover, to the best of our knowledge, different models of paragraph vectors have not been widely studied in comparison to word vector models for the task of classification. Some of the recent applications of these models in this direction are given as follows. Word vector representations have shown their capability for sentiment classification in languages other than English, hence this model can be easily extended to multi-lingual texts (Abburi *et al.*, 2016; Al-Amin *et al.*, 2017; Cerón-Guzmán and León-Guzmán, 2016; Giatsoglou, 2017; Sanguansat, 2016; Zhang *et al.*, 2015). In their paper Zhang *et al.* (2015), classified Chinese consumer comments using word2vec with SVM-perf. Giatsoglou *et al.* (2017), studied different models such as the hybrid of lexicon and word embeddings for classifying online user reviews in Greek and English languages. Their proposed hybrid approach was immediately followed by Chatzakou *et al.* (2017) to identify varying emotions in online activities. However, the hybrid feature vectors proposed by the above studies use expensive lexicons that may not comprise more informal words and newly generated special words. Liu (2017) performed citation sentiment classification using the word vector representations. In some instances, word and sentence vector representations do not yield any significant improvement in classification accuracy when compared to non-distributed vectorization methods (Bansal and Srivastava, 2018b). Further, Jiang *et al.* (2016b) observed only a slight improvement by using TF-IDF and IDF to combine distributed word vectors on Yelp and Trip Advisor data sets.

The distributive word representation has been further used in many studies for aspect sensitive sentiment analysis. Poria *et al.* (2016) proposed to capture semantic relation using distributed vectorization and other syntactic features using POS tags, finally concatenating to make an enlarged vector. Further, they proposed a deep learning approach for aspect classifications. Many studies have attempted to capture contextual information for the task of document classification (Han *et al.*, 2018), but in this paper, we have extended the task of capturing aspect context. Again, Araque *et al.* (2016) used word2vec model for extracting context sensitive features with other ensemble features such as POS tagging and lexical features. However, for sentiment features, they use lexicon which has few drawbacks, especially for mining OCR. As mentioned earlier, the online generated text contains slangs, abbreviations, multi-lingual words and other special words that may express emotions but are not included in pre-existing lexicons. Jiang (2016a) also showed the superiority of incorporating four types of features: linguistic, sentiment, topic and word2vec features. Abdi *et al.* (2018) combined multiple lexicons to build a high coverage lexicon, followed by the word2vec model to create a hybrid feature vector. In this paper, we attempt to eliminate the usage of expensive lexicons that may not comprise informal words.

Another class of models for aspect based sentiment analysis includes topic-sentiment models (Yang *et al.*, 2018b). Latent Dirichlet allocation (LDA) first proposed by Blei *et al.* (2003) is an unsupervised topic detecting model that has been extended to various semi-supervised and supervised models for the task of aspect based sentiment analysis. Again, LDA generates explicit topics from the corpus and has been widely used for aspect based sentiment classification on online consumer reviews (Bansal and Srivastava, 2018a). Some important advancements in LDA are discussed, especially for aspect based sentiment classification of online consumer reviews. Jo and Oh (2011) proposed a topic-sentiment model to automatically discover aspects in consumer reviews and corresponding sentiments. One downside to the proposed sentence-LDA (SLDA) model, is that it assumes all words in a single document depict only one aspect. They also proposed the aspect and sentiment unification model (ASUM), which is an extension of SLDA, used to detect aspects and corresponding sentiment pairs. However, their proposed model was outperformed by a more sophisticated approach joint multigrain topic sentiment (JMTS) (Alam *et al.*, 2016). JMTS can extract aspects automatically and does not require labelled data, rating based reviews or any other supervised setting. However, two small sets of aspect independent positive and negative word lists were required for sentiment analysis task. In this paper, we also borrow the same list for aspect context aware sentiment analysis and update it to make even more domain-independent. A weakly supervised learning LDA was proposed by Yang *et al.* (2018a) for identifying domain-specific aspects and sentiment representations for abstractive review summarization. However, this approach requires a manual setting of the number of aspect topics and other parameters. More recently, hybrid of two approaches (LDA and Word2vec) has been proposed by García-Pablos *et al.* (2018) in the form of W2vlda that performs aspect identification and classification, aspect and sentiment bearing word separation and finally, sentiment classification. One drawback to their approach is that they only retain sentences with a single aspect.

The traditional methods for sentiment classification include lexicons or using dictionaries for labelling the data sets, followed by supervised learning. The drawbacks of these old-fashioned techniques comprise labelling a full set of data, which is expensive (Han *et al.*, 2019). Moreover, the lexicons can only label limited words and cannot keep up with the new slangs, sentiment bearing words and other special words. Further, the motivation behind this work was to extend aspect based sentiment analysis to an aspect context approach. Specifically, we attempt to capture implicit word relations and contextual information for aspect based sentiment analysis as opposed to the majority of previous studies that explicitly capture aspects and corresponding sentiments. Our approach is beneficial in many ways, for instance, we do not assume that each review consists of a single aspect (Jo and Oh, 2011). Secondly, we attempt to extract aspect context sensitive sentiment features without relying on a sentiment lexicon but use a domain-independent small set of words. Thirdly, we use ensemble features and ensemble classifier for incorporating maximum information and superior classification accuracy. We finally compare state-of-the-art topic-sentiment, distributed and non-distributed text vectorization methods for showing the superiority of our approach.

3. Data and methodology

3.1 Data pre-processing

The Amazon mobile phones review data set[1] includes more than 400,000 reviews, whereas hotel reviews from Trip advisor contain more than 20,000 reviews, of variable lengths, where each review may consist of multiple sentences (Alam *et al.*, 2016), with a corresponding consumer rating on a scale of 1-5. A consumer rating of 1 shows the maximum level of consumer dissatisfaction and a rating of 5 shows the highest level of consumer satisfaction. Table I describes both data sets and Table II shows a few examples of consumer reviews in both data sets. As done in previous studies, we decide to label reviews with ratings above 2 as positive (1) and ratings below 3 as negative (0) (Table III) (Bansal and Srivastava 2018a; Bansal and Srivastava, 2019). We also use the balanced data sets in our study. To balance the data sets, we remove the comments with the corresponding rating of 3 and label the reviews with 4 and 5 ratings as positive, and 1 and 2 ratings as negative, taking the equal number of positive and negative reviews. Table III shows the consumer ratings and corresponding class labels for both balanced and unbalanced data sets. Table IV shows the number of reviews in each class for both balanced and unbalanced data sets. Balancing the data sets reduces the sizes of both data sets. In both unbalanced data sets, positive reviews are more than negative reviews. Pre-processing steps include converting to lower case, removing punctuations, stop words, numbers and white spaces.

Table I Data set description: consumer reviews from two different domains: hotel reviews from tripadvisor.com and mobile phone reviews from Amazon.com

Data set	No. of reviews	Sentences per review	Words per review
Mobile phone Reviews	413,840	3.1	40.5
Hotel reviews	20,491	12	104

Table II Example of consumer reviews from two data sets: hotel reviews from tripadvisor.com and mobile phone reviews from Amazon.com

Data set	Review	Ratings
Mobile phone reviews	I brought this phone as a replacement for my daughter, 5 who is very hard on cell phones. I must say it was a great purchase. The phone work wonderful. Thank you	5
Hotel reviews	Nice hotel expensive parking got good deal stay hotel anniversary 4 arrived late evening took advice previous reviews did valet parking check quick easy little disappointed non-existent view room clean nice size bed comfortable woke stiff neck high pillows not soundproof like heard music room night morning loud bangs doors opening closing hear people talking hallway maybe just noisy neighbours	4

Table III Class labels based on consumer ratings, for both balanced and unbalanced data sets

Consumer rating	Class label in balanced data set	Class label in unbalanced data set
1	0	0
2	0	0
3	–	1
4	1	1
5	1	1

3.2 Distributed word vectors: hyper parameters setting

Hyper-parameter settings used in this study are reported in Table V (word2vec and doc2vec for comparisons with the proposed approach). The vector dimension of a word representation or a document representation plays an important role in quality feature representations and improved classification accuracy. First, we increase vector dimensions from 100 to 500 (set at 100, 200, 300, 400 and 500) and find that increasing vector dimensions increases the accuracy of class prediction for all models. For word2vec, classification accuracies were comparable at 400 and 500 dimensions, thus we only report results for 400 dimensions. For doc2vec, vector dimensions are set to 500. Vector dimensions can go up to 1,000, but with the increase in computational time. Next,[2] training window size for CBOW is set to 5, skip-gram is set to 10 and for doc2vec window size is set to 5. A window represents the number of contextual words/co-occurring words around a given word to be considered to train the model. The minimum frequency of a word included in every model is set to 5 (Mikolov *et al.*, 2013c), that is we discard all the words that occur less than 5 times. Finally, we use negative sampling (1e-3 and 1e-5) instead of hierarchical softmax as the former works better for data consisting of recurrent words[2]. The weighting schemes used to combine word vector representations are also mentioned in Table V. Non distributed vectorization methods (TF-IDF) have been used using the Scikit-learn library (Pedregosa *et al.*, 2011). The Gensim library has been used to extract distributed vector representations (word2vec and doc2vec) (Rehurek and Sojka, 2011). Hyper-parameter settings can be modified according to the task using Gensim library.

3.3 Sentiment bearing words and aspects detection

Algorithm 1 shows the detailed steps of the proposed framework. First, phrase modelling is performed to detect bigrams. Then, tri-grams are detected and added to the text. For example, following extracted n-grams: happy_birthday,

Table IV Number of reviews in each class for both balanced and unbalanced cases, in Trip Advisor and Amazon data sets

	Balanced		Unbalanced	
	Negative	Positive	Negative	Positive
Mobile review Data set	97,078	97,078	97,078 (24%)	316,762 (76%)
Hotel review Data set	3,214	3,214	3,214 (15.7%)	17,277 (84.3%)

Table V Hyper-parameter settings for Word2Vec (continuous bag of words and skip-gram) and Doc2Vec (paragraph vector-distributed memory and paragraph vector-distributed bag of words) models

	Word2Vec		Doc2Vec	
Model architecture	**CBOW**	**Skip-gram**	**PV-DM**	**PV-DBOW**
Vector size	200-500	200-500	200-500	200-500
Window size	5	10	5	5
Sub-sampling Rate	1e-3	1e-3	1e-5	1e-5
Training algorithm	Negative sampling	Negative sampling	Negative sampling	Negative sampling
Min_count	5	5	5	5
Weighting Scheme	TFIDF, Sum	TFIDF, Sum	Sum, Concat	–

nice_touch, big_disappointment exhibit a neutral, aspect-specific positive and highly negative sentiments, respectively, that otherwise comprising unigrams could not have detected accurately.

Then, POS tagging is used to retain aspect depicting and sentiment bearing words. For handling negation terms (don't, not, didn't and no), not is joined at the end of the adjective (Asghar *et al.*, 2017). The polarity of this adjective will be reversed when extracting word polarity vectors for the ensemble. Further, we also retain adverbs for more reliable and accurate results. Nouns usually describe the aspects, whereas adjectives describe sentiments or opinions towards the aspect (Chen and Xu, 2017). Next, after analyzing the extracted tags, adjective-noun and noun-adjective pairs are retained for finding aspect context sensitive sentiment (Poria *et al.*, 2016). One traditional way to determine aspect based sentiments is aggregating the explicit sentiments of noun-adjective and adjective-noun pairs in a document. However, this approach overlooks the implicit word relations and contexts. Finally, verb-adverb pairs are also retained, wherein the pairs like updates_slowly will be incorporated for sentiment detection of a review and the verb will be treated as an aspect. Figure 1 shows the proposed framework in brief.

3.4 Aspects context sensitive ensemble features

Firstly, the TF-IDF (term document-inverse document frequency) scheme is used for extracting the first set of features. TF-IDF scheme is extensively used to allocate distinctive weights to words in the document according to their significance (Kotlerman *et al.*, 2018). Weight (Wi,j) of a term (ti) in a given document (dj) is defined as:

$$W_{i,j} = T.F_{i,j} * \log \frac{N}{D.F_i}$$

where T.F is the frequency of term i in document j, N is the total number of documents (reviews), D.F the is total frequency of documents containing term i.

The second set of features is obtained using the skip-gram shallow neural net to best capture contextual properties and semantic word relations in the text (Mikolov *et al.*, 2013b). First, context is defined as follows. In a sentence containing n number of words $\{w_1, w_2, \ldots w_n\}$, context $C(w_i)$ of the word w_i is defined as: $C(w_i) = w_{i-k}, \ldots, w_{i-1}, w_{i+1}, \ldots w_{i+k}$, where k is the window size.

Particularly, to capture the contextual information of words, the Skip-Gram model is used, where, the target term is at the input layer and contextual words are at the output layer. The output of contextually similar word vector is updated by the hidden layer. Firstly, each and every word w_i is represented by n-dimension vector v_{w_i} in the space, where semantically similar words are mapped closer to each other. Then, each vector v_{w_i} is trained to maximize the log probability of context $C(w_i)$ of word w_i, mathematically shown as follows (Shalaby *et al.*, 2019),

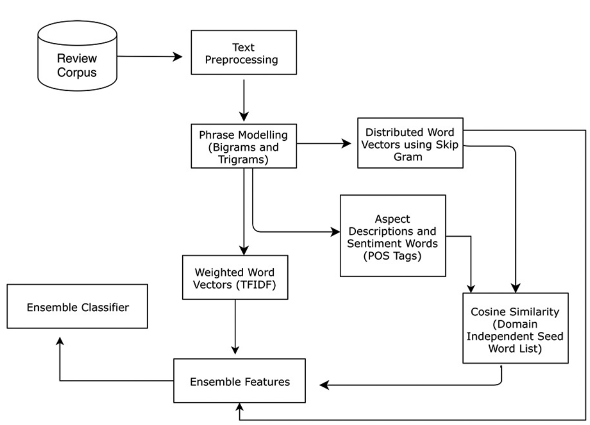
$$\frac{1}{N}\sum_{w_{i=1}}^{N} \sum_{c \in C(w_i)} \log p(c|w_i)$$

where $\log p(c|w_i)$ is the conditional probability and can be replaced by following the negative sampling objective (Mikolov *et al.*, 2013b).

$$\log \sigma(v_c.v_w) - \sum_{j=1}^{K} \log \sigma(-v_{c_j}.v_w)$$

where the first log term represents that the pair (w, c) came from the corpus, whereas the second log-term represents that k pairs (w, c_j) are drawn from randomly sampled negative examples. v_c and v_w represent vectors for the context word and the input word. The negative sampling method was proposed to reduce complexity without reducing the quality of vectors (Mikolov *et al.*, 2013b). As mentioned earlier, negative sampling works better for frequently occurring terms as compared to hierarchical softmax objective. This was also

Figure 1 Proposed framework in brief



confirmed by manual analysis of the quality of word vectors (similarity).

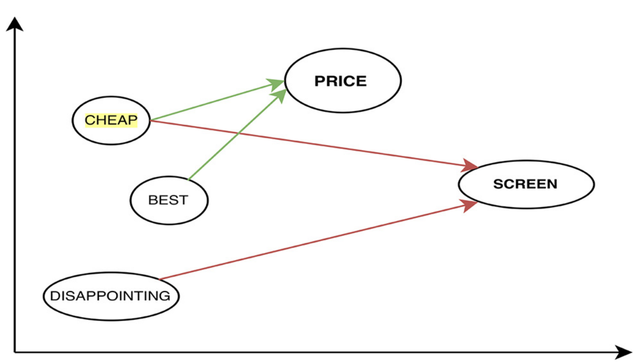
The last set of features includes sentiment information of sentiment bearing words extracted in Section 2.3. Online unstructured texts contain constantly generated new slangs, abbreviated words, multi-lingual words and ambiguous words that express the sentiment of the consumer. Therefore, traditional sentiment lexicons are not comprehensive and fail to determine the correct sentiment orientation of these words. To handle this, we use the cosine similarity measure and a small set of key words. The set of separate positive and negative seed words (Table VI) is borrowed from Lin and He (2009). These keywords would help to extract better sentiment carrying words and their respective polarities. The keyword list used by Lin and He was updated according to a small set of movie review corpus. Hence, we update the seed word list in our experiment and use it for both the mobile phone and hotel review data sets.

The sentiment polarity of positive seed words is taken as (+1) and negative seed words is taken as (−1). Previous studies have used various similarity measures to calculate the shortest average distance between new sentiment bearing word and only positive or negative words (Pablos *et al.*, 2015). This approach overlooks the aspect context of ambiguous words. For instance, a word cheap is positive in the context of aspect Price, but negative in the context of aspect screen (Figure 2). Thus, sentiment orientation of a target word (adjective and adverb) is detected by first, measuring average cosine distance between aspect and positive seed word list (A^+), second, the average distance between aspect and negative seed word list is measured (A^-) (Algorithm 1). Then, cosine similarity between target adjective and target aspect (B) is measured. Finally, the target adjective is classified as positive if $|A^+ - B| > |A^- - B|$, otherwise negative. That is, in the context of aspect PRICE, if S represents cosine similarity between two vectors, then S $(V_{PRICE}, V_{BEST}) \approx S(V_{PRICE}, V_{CHEAP})$, therefore CHEAP will be classified as positive. On the contrary, $S(V_{SCREEN}, V_{DISAPPOINTING}) \approx S(V_{SCREEN}, V_{CHEAP})$, CHEAP will be classified negative in the context of aspect SCREEN, hence classifying each adjective correctly and comprehensively.

Table VI Positive and negative seed word list Lin and He (2009)

Positive seed words	Dazzling brilliant phenomenal excellent fantastic mesmerizing riveting spectacular cool awesome exciting love wonderful best great superb beautiful
Negative Seed words	Sucks terrible awful hideous bad stupid slow worst waste unexcit rubbish tedious frustrated awkward disappointing

Figure 2 A graph illustrating the semantic word relations between aspect price, aspect screen and ambiguous term cheap



Cosine similarity measure (S) between two word vectors (V_i, V_j) of words (W_i, W_j), is defined as follows (Li *et al.*, 2019):

$$S(i, j) = \frac{Vi \cdot Vj}{||Vi|| \times ||Vj||}$$

The new set of features is generated by detecting the sentiment orientation of each adjective in the document. Finally, the three sets of features, namely, weighted word vectors, distributed word vectors to capture context and aspect context sensitive sentiment features are concatenated.

```
Algorithm 1: Aspect Context Aware Ensemble
             Feature Generation
Input: A corpus (C) including online consumer
       review documents (D = {d1, .,dm}) .
       Sentiment labelled Positive and
       negative seed words lists.
Output: Aspect Context Aware Ensemble
       Featuresinitialization;
//Phase 1: Text Pre-processing
for each document dm in corpus C do
       Step 1. convert to lower case
       Step 2. remove stop words
       Step 3. remove punctuations
       Step 4. remove numbers
       Step 5. eliminate white space
end//
Phase 2: Aspect word descriptions and opinion
         words extraction
for each document dm do
       detect and update bigrams
end
for each document dm do
       detect and update trigrams
end
for each term ti in document dm do
       Negation Handling: Add _NOT in the end of
       adjectives
       Extract noun-adjective, adjective-noun
       pairs
end
//Phase 3: Distributed Word Vectors
for each word w in original corpus C, do
       distributed word vector v_wi
       representations using Skip-gram Model
end
//Phase 4: Word Polarity Vectors
for each word w (adjectives) in document dm do
       associate sentiment score sw by:
       sw is positive (+1) if
       Step 1. A = Cosine_Distance (v1 Aspect,
               v1, (Seed word list))
       Step 2. B = Cosine_Distance (v_Aspect, v_w)
       Step 3. sw is positive if B - A+ > B - A-,
               otherwise, negative. (A+
               represents positive seed list)
end
for each document dm in corpus C do
       represent using respective word
       polarities
end
//Phase 5: Weighted Word Vectors
```

```
for each word w in document dm do
      Calculate word weights according to TF-IDF
      scheme
end
for each document dm in corpus C do
      represent using respective word weights end
//Phase 6: Aspect Context Aware Ensemble
      Features
for each document dm in corpus C do
      concatenation of distributed word vectors,
      weighted word vectors and word polarity
      vectors.
end
```

3.5 Ensemble classifiers

The ensemble of multiple classifiers has been used for improved classification accuracy (Han *et al.*, 2019). First, three individual classifiers, namely, support vector machine (SVM), logistic regression (LR) and random forest (RF) are trained by all the features as mentioned in the previous section. Then, the majority voting technique is used on the prediction results of these three classifiers, without any weighting scheme. That is, the ensemble classifier decides the final polarity class by majority voting results of the base classifiers. The individual prediction accuracy of the Naïve Bayes (NB) classifier was unacceptable and hence not included in the ensemble. Ensemble classification is done using the Scikit-learn library.

4. Results and discussion

Table VII shows a few examples of the similar terms and documents extracted by word2vec and doc2vec models. It can be easily observed from Table VI, that the word cheap, can carry both negative (Cheep, Plasticy, Flimsy, Knock-off) and positive sentiment (Inexpensive), and hence, is ambiguous in nature. The sentiment polarity of such words, thus, depend on the aspect context. Further, incorrect spelled words such as Cheep and words such as Plasticy express consumer opinion but may not be included in pre-existing lexicons but carry sentiments. To automatically capture the implicit relations between such sentiment bearing words and aspects, we generate aspect context ensemble features. The classification is performed using an ensemble classifier, trained on three base classifiers: LR, SVM and RF.

Experimental results obtained from the proposed approach are compared with different word embedding schemes including state-of-the-art distributed word vector representations and paragraph vector representations. The

Table VII Similar terms extracted by word2vec and similar documents extracted by doc2vec: Amazon mobile review data set

Key search word	Similar words extracted by word2vec: Skip-gram
Samsung	Sumsung, Galaxy, Samsun, Phone
Camera	Autofocus, Zeiss, Pictures, Optics, Mpixels
Cheap	Cheep, Plasticity, Inexpensive, Flimsy, Knock-off
Key search word	Similar document extracted by doc2vec: DM
Too small	Too small, two small, to small, too small
Too bad	Disappointing, is bad, very bad

word vectors are combined using average of word vectors and weighted combination of word vectors using the TF-IDF scheme to represent a document (review). TF-IDF scheme is used to better capture more important words as compared to simple averaging of vectors. With each document represented (vectorized) and labelled (negative and positive), we use four base machine learning (ML) algorithms with 10-folds cross-validation for testing the accuracy of each model. Four ML algorithms used are NB, SVM, RF and LR.

Similarly, for further comparison of our proposed method, distributed document vectors are also generated using PV-DM and PV-DBOW models on both data sets to represent documents (reviews). The whole document is represented instead of individual words and phrases. In PV-DM, paragraph vector and word vectors are summed and concatenated. Although computationally expensive, concatenation is applied in an attempt to better capture the semantic relationships as compared to simple summation. Again, with documents represented in vector forms and labelled with sentiment class labels, we use four different classification algorithms with 10-fold cross-validation to test the accuracy of models. We use the Scikit Learn library for all ML algorithms. Further, the grid search is used for optimizing all the hyper-parameters of classifiers to further produce more reliable comparison. For learning all vector representations 10 epochs are performed.

Tables VIII and IX show the point estimates of classification accuracies for unbalanced mobile phone review data set and hotel review data set respectively, with standard error (for interval estimate) of 10 fold cross-validations at 500 vector dimensions. Results show that the proposed model outperforms other distributed vector representations and base classifiers for the task of sentiment classification in both the data sets. Although NB took the least training time, the classification accuracy is unacceptable with the exception in paragraph vectors. In the case of mobile phone reviews, paragraph vectors showcase superior classification accuracy with NB. Nevertheless, the other three classification techniques do not exhibit expected results with paragraph vectors, although paragraph vectors took maximum computational time to learn feature vectors. As expected, the proposed method also showcased improved classification accuracies at lower vector dimensions of 200-400. Figure 3 shows different accuracies at different vector dimensions for the selected models depicting a marginal difference in Amazon mobile review data set. Also, it can be noted that, as the vector dimension increases, the improvement exhibited by our method also increases. This can be mainly due to the high-quality sentiment feature vector, that is, more accurate classification of words into positive or negative at higher vector dimensions. However, the overall accuracies at the lower dimensions (specifically 200 and 300) were unacceptable. Hence, we retained the vector dimension at 500 and proceeded further.

Although in the hotel review data set each review consisted of multiple sentences, the distributed document vector exhibits very low accuracy. We also compare the classification accuracies of the proposed method with other existing topic-sentiment models ASUM and JMTS, as shown in Table IX. The proposed approach outperforms the previous methods with large margins. Moreover, as seen in Table IV, balancing the data sets reduces the data size drastically. Hence, experimenting on balanced data sets reduced the classification accuracy, as distributed vector representations rely on large

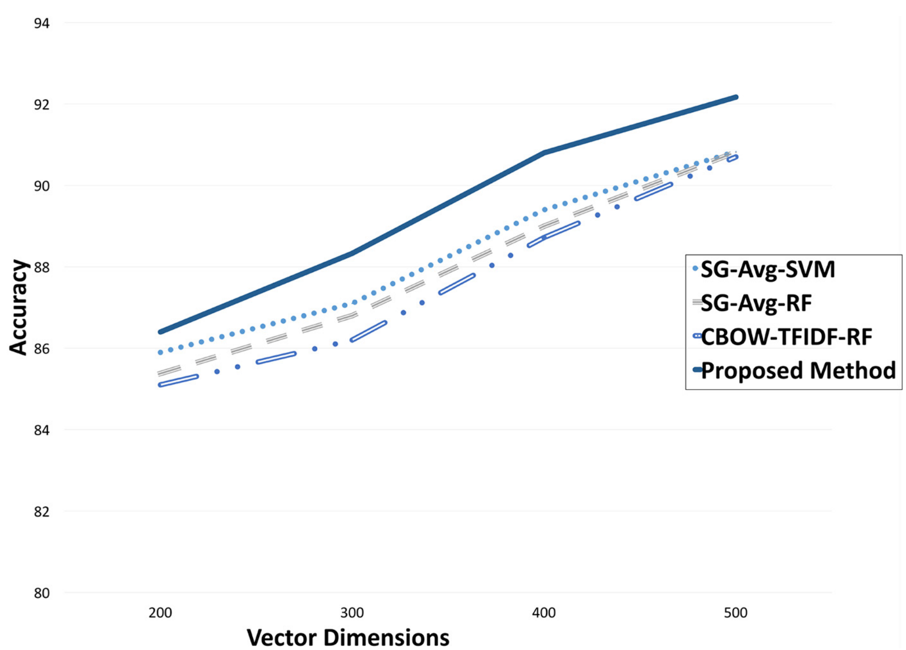
Table VIII Comparison of classification accuracies (at 500 vector dimensions) for mobile phone review data set (unbalanced), reported in (%) for different feature extraction methods (mean + 95% confidence interval)

	Feature extraction	NB	Logistic regression	SVM	Random forest
Word2Vec	SG Avg.	48.8 ± 0.11	90.2 ± 0.12	90.78 ± 0.3	90.18 ± 0.22
	SG + TF-IDF	47.9 ± 0.71	87.1 ± 0.15	89.7 ± 0.18	90.1 ± 0.17
	CBOW Avg.	53.38 ± 0.74	89.9 ± 0.16	90.3 ± 0.18	90.44 ± 0.2
	CBOW+TF-IDF	53.76 ± 0.22	88.6 ± 0.21	89.8 ± 0.25	90.38 ± 0.38
Doc2Vec	PV-DM Sum	68.49 ± 3.98	84 ± 0.9	83.7 ± 0.89	81 ± 4.4
	PV-DBOW	69.91 ± 4.01	89.1 ± 1.1	90 ± 1.5	88.8 ± 4.12
	PV-DM Concat	71.6 ± 3.55	86.1 ± 2.3	89.9 ± 0.27	86.5 ± 4.11
Proposed ensemble approach	Ensemble Features (SG + TF-IDF + Aspect Context Aware Sentiment + Concat.)	91.65 ± 0.1	(Ensemble Classifier)		

Table IX Comparison of classification accuracies (at 500 vector dimensions) for hotel review data set (unbalanced), reported in (%) for different feature extraction methods (mean + 95% confidence interval)

	Feature extraction	NB	Logistic regression	SVM	Random forest
Word2Vec	SG Avg.	51.3 ± 0.41	91.53 ± 0.5	90.9 ± 0.17	89.66 ± 1.1
	SG + TF-IDF	52.4 ± 1.18	91.8 ± 0.16	91.1 ± 0.3	90.27 ± 0.7
	CBOW Avg.	56.19 ± 1.6	90.59 ± 0.2	89.8 ± 0.35	88.4 ± 1.29
	CBOW+TF-IDF	57.4 ± 1.12	92 ± 1.11	90.1 ± 0.68	88.89 ± 1.31
Doc2Vec	PV-DM Sum	65.2 ± 3.51	80.2 ± 1.9	81.50 ± 2.01	79.3 ± 1.3
	PV-DBOW	67.19 ± 3.51	80.52 ± 1.9	80.9 ± 1.42	79.9 ± 1.21
	PV-DM Concat	67.7 ± 2.51	82.92 ± 1.57	83.97 ± 1.32	79.82 ± 1.58
Proposed ensemble approach	Ensemble Features (SG + TF-IDF + Aspect Context Aware Sentiment + Concat.)	*92.72 ± 0.31*	(Ensemble Classifier)	JMTS (Alam *et al.*, 2016) 84	ASUM (Jo and Oh, 2011) 84

Figure 3 Different classification accuracies exhibited at different vector dimensions for selected models in Amazon Mobile Review Data set reported in (%)



data sets for training purposes. The proposed model reached an accuracy of 83.7 per cent and 87.2 per cent in the balanced mobile review data set and balanced hotel review data set, outperforming other base models. Balanced data sets contained very small vocabulary, which is insufficient to generate quality vectors.

As seen in Tables VIII and IX, the difference between accuracies is marginal, especially with the standard error. Therefore, we further find AUC-ROC (area under receiver operating characteristic curve) for selected models depicting a marginal difference in respective data sets. AUC-ROC for both data sets (unbalanced) are shown in Table X. As seen in Table X, AUC-ROC are correlated to accuracies as expected.

Although 10-fold cross-validation and AUC-ROC results show that the proposed model performs better in the case of both the data sets, the difference between the classification of respective accuracies and AUC are marginal (especially considering the standard error) for few models. Therefore, to further compare the model performances we apply McNemar's test (Devos *et al.*, 2014). The null hypothesis (*H0*) states that the performances of the two classifiers are equal. The alternative hypothesis (*H1*) states that the performances of two classifiers differ significantly (Significance level = 0.05, 0.01, 0.001). The results of McNemar's test for selected models are given in Tables XI and XII. For the mobile review data set, all model performances are significantly different. For the hotel review data set, classification performances of skipgram + TFIDF + LR and the proposed approach are significantly different at a 95 per cent confidence level. Nevertheless, the proposed model has outperformed other text representation techniques for the task of classification (both balanced and unbalanced data sets).

The proposed model shows superior performance as opposed to the base models including distributed vector representations of words and documents. Hence, for the objective of sentiment classification, aspect context sensitive feature has shown improvement through our study. It is also notable that distributed document vectors have shown lower classification accuracies. One possibility is that not all order sensitive documents could be efficiently represented as distributed vectors, in the case of paragraph vectorization. Also, training word vectors and paragraph vectors with a larger vocabulary could improve the classification accuracy (but also the computational cost). Mainly, we incorporated TF-IDF, which is capable of assigning higher weights to more relevant and important words. Then, we extracted the aspect context sensitive features with the help of cosine similarity. It can also be noted that word vector representations outperform paragraph vectors. One possible explanation is that in the case of spelling mistakes, combining word vectors would represent a document better, and hence better classification accuracy. The possibility that word vector might perform better than paragraph vector in case of short texts is ruled out because our data set mostly contains paragraph level reviews. We believe that training these models with larger corpora harvested from different sources of social media and online websites, will further improve the classification accuracy of the proposed model.

5. Conclusions

In this paper, we propose a hybrid approach consisting of ensemble features and ensemble classifiers, in an attempt to predict aspect context aware sentiments of OCR. This approach

Table X AUC-ROC results for selected models (at 500 vector dimensions) with similar accuracies: amazon mobile review and Trip Advisor hotel review data sets

Amazon mobile reviews	Proposed Model	SG-Avg-SVM	SG-Avg-RF	CBOW-TFIDF-RF
AUC-ROC	0.91	0.84	0.86	0.85
Trip Advisor hotel reviews	Proposed Model	SG-Avg-LR	SG-TFIDF-LR	CBOW-TFIDF-LR
AUC-ROC	0.92	0.89	0.86	0.89

Table XI McNemar's test results for selected model comparisons: mobile review data set (unbalanced)

Proposed approach	–	$p << 0.001^{***}$	$p << 0.001^{*}$	$p << 0.001^{***}$
SG-Avg-SVM		–	$p << 0.001^{***}$	$p << 0.001^{***}$
SG-Avg-RF			–	$p << 0.001^{***}$
CBOW-TFIDF-RF				–

Note: Significance level is shown as *0.05; **0.01; ***0.001

Table XII McNemar's test results for selected model comparisons: hotel review data set (unbalanced)

	Proposed approach	**SG-Avg-LR**	**SG-TFIDF-LR**	**CBOW-TFIDF-LR**
Proposed approach	–	$p << 0.001^{***}$	$p < 0.05^{*}$	$p << 0.001^{***}$
SG-Avg-LR		–	$p << 0.001^{***}$	$p << 0.001^{***}$
SG-TFIDF-LR			–	$p << 0.001^{***}$
CBOW-TFIDF-LR				–

Note: Significance level is shown as *0.05; **0.01; ***0.001

uses the skip-gram shallow neural network to represent words as distributed vectors to capture contextual information from the corpus. Next, aspect context aware sentiments are extracted using the cosine similarity measure and a small set of domain-independent seed words. Finally, weighted word vectors are obtained using TF-IDF and then, concatenated with previously extracted feature sets. The resultant ensemble of feature sets is used as input for ensemble classifiers, based on the majority voting method of three base classifiers, namely, SVM, RF and LR. We tested our model on two data sets with different domains and varying document lengths.

The classification accuracy of the proposed methodology exceeds state-of-the-art models for the task of sentiment classification in both the data sets. The proposed approach is also flexible and can be used for mining consumer reviews in other languages after appropriate customization. Moreover, another benefit of this method is that it does not use previous consumer ratings, lexical knowledge and pre-defined aspects for labelling purposes. Hence, it can be used to mine tweets, blogs and other real-world data sets.

In the future, we aim to perform experiments on very short length data sets such as tweets, to examine the efficiency of the proposed approach in case of the sparse data and to further improve its performance. Distributed word vector representations have shown their capability for superior sentiment classification in languages other than English. Hence, we will test the performance of our proposed method for the texts in other languages and multi-lingual texts.

6. Limitations and future work

Although it is cheaper to use a small set of seed words instead of lexicons, the latter has exhibited superior results in case of more fine-grained classification (more than two polarity categories) (Mohammad and Turney, 2010). Hence, immediate future work includes testing the performance of the proposed model for a 5-way or 3-way sentiment classification. Next, while analyzing the aspects extracted from the study, we observed that many aspects are very similar, for instance, battery and charger or Wi-Fi and internet. In the future, we want to combine these large numbers of similar aspects to generate attributes and simultaneously classify according to these attributes. Non-textual features such as emoji's and emoticons will be studied in the ensemble feature set (Asghar *et al.*, 2017). Finally, we will train different model architectures using a larger corpus of online reviews, as done by McAuley and Leskovec (2013) to determine the quality of features and accuracy of sentiment classification.

Availability of data

The data sets analysed during the current study, are available in the Zenodo repository: Amazon Mobile Review Data set [http://doi.org/10.5281/zenodo.1211639] and TripAdvisor Hotel Review Data set [http://doi.org/10.5281/zenodo.1219899]. The data sets were originally derived from Amazon Mobile Review[1] and TripAdvisor Hotel Review (Alam *et al.*, 2016), respectively.

Notes

1 kaggle.com/PromptCloudHQ/amazon-reviews-unlocked-mobile-phones/data (last accessed November 17)

2 code.google.com/archive/p/word2vec/ (last accessed January 2018).

References

Abburi, H. Akkireddy, E.S.A. Gangashetti, S. and Mamidi, R. (2016), "Multimodal sentiment analysis of Telugu songs", SAAIP@ IJCAI, pp. 48-52.

Abdelwahab, O. and Elmaghraby, A. (2016), "UofL at SemEval-2016 task 4: multi domain word2vec for twitter sentiment classification", *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 164-170.

Abdi, A., Shamsuddin, S.M., Hasan, S. and Piran, J. (2018), "Machine learning-based multi-documents sentiment-oriented summarization using linguistic treatment", *Expert Systems with Applications*, Vol. 109, pp. 66-85.

Al-Amin, M., Islam, M.S. and Uzzal, S.D. (2017), "Sentiment analysis of Bengali comments with Word2Vec and sentiment information of words", *International Conference on Electrical, Computer and Communication Engineering (ECCE)*, *IEEE*, pp. 186-190.

Alam, M.H., Ryu, W.J. and Lee, S. (2016), "Joint multi-grain topic sentiment: modelling semantic aspects for online reviews", *Information Sciences*, Vol. 339, pp. 206-223.

Araque, O., Zhu, G., García-Amado, M. and Iglesias, C.A. (2016), "Mining the opinionated web: classification and detection of aspect contexts for aspect based sentiment analysis", *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*, *IEEE*, pp. 900-907.

Asghar, M.Z., Khan, A., Bibi, A., Kundi, F.M. and Ahmad, H. (2017), "Sentence-level emotion detection framework using rule-based classification", *Cognitive Computation*, Vol. 9 No. 6, pp. 868-894.

Bansal, B. and Srivastava, S. (2018a), "Hybrid attribute based sentiment classification of online reviews for consumer intelligence", *Applied Intelligence*, Vol. 49 No. 1, pp. 1-13.

Bansal, B. and Srivastava, S. (2018b), "Sentiment classification of online consumer reviews using word vector representations", *Procedia Computer Science*, Vol. 132, pp. 1147-1153.

Bansal, B. and Srivastava, S. (2019), "Context-sensitive and attribute-based sentiment classification of online consumer-generated content", *Kybernetes*.

Bengio, Y., Ducharme, R., Vincent, P. and Jauvin, C. (2003), "A neural probabilistic language model", *Journal of Machine Learning Research*, Vol. 3, pp. 1137-1155.

Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003), "Latent Dirichlet allocation", *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022.

Cerón-Guzmán, J .A. and León-Guzmán, E. (2016), "A sentiment analysis system of Spanish tweets and its application in Colombia 2014 presidential election", *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*, *IEEE*, pp. 250-257.

Chatzakou, D., Vakali, A. and Kafetsios, K. (2017), "Detecting variation of emotions in online activities", *Expert Systems with Applications*, Vol. 89, pp. 318-332.

Chen, R. and Xu, W. (2017), "The determinants of online customer ratings: a combined domain ontology and topic text analytics approach", *Electronic Commerce Research*, Vol. 17 No. 1, pp. 31-50.

Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R. (1990), "Indexing by latent semantic analysis", *Journal of the American Society for Information Science*, Vol. 41 No. 6, pp. 391-407.

Devos, O., Downey, G. and Duponchel, L. (2014), "Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils", *Food Chemistry*, Vol. 148, pp. 124-130.

Elman, J.L. (1990), "Finding structure in time", *Cognitive Science*, Vol. 14 No. 2, pp. 179-211.

Elman, J.L. (1991), "Distributed representations, simple recurrent networks, and grammatical structure", *Machine Learning*, Vol. 7 No. 2-3, pp. 195-225.

García-Pablos, A., Cuadros, M. and Rigau, G. (2018), "W2vlda: almost unsupervised system for aspect based sentiment analysis", *Expert Systems with Applications*, Vol. 91, pp. 127-137.

Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G. and Chatzisavvas, K.C. (2017), "Sentiment analysis leveraging emotions and word embeddings", *Expert Systems with Applications*, Vol. 69, pp. 214-224.

Han, H., Bai, X. and Li, P. (2018), "Augmented sentiment representation by learning context information", *Neural Computing and Applications*, Vol. 31 No. 12, pp. 1-8.

Han, Y., Liu, Y. and Jin, Z. (2019), "Sentiment analysis via semi-supervised learning: a model based on dynamic threshold and multi-classifiers", *Neural Computing and Applications*, pp. 1-13.

Hinton, G.E. McClelland, J.L. and Rumelhart, D.E. (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1*, MIT Press, Cambridge, MA, USA. chapter Distributed Representations, pp. 77-109, available at: http://dl.acm.org/citation.cfm?id=104279.104287.

Ibrahim, N.F. and Wang, X. (2019), "A text analytics approach for online retailing service improvement: evidence from Twitter", *Decision Support Systems*, Vol. 121, pp. 37-50.

Jiang, M., Zhang, Z. and Lan, M. (2016a), "Ecnu at SemEval-2016 task 5: extracting effective features from relevant fragments in sentence for aspect-based sentiment analysis in reviews", *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pp. 361-366.

Jiang, S., Lewris, J., Voltmer, M. and Wang, H. (2016b), "Integrating rich document representations for text classification", *2016 IEEE (Systems and Information Engineering Design Symposium (SIEDS)*, *IEEE*, pp. 303-308.

Jo, Y. and Oh, A.H. (2011), February). "Aspect and sentiment unification model for online review analysis", *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, *ACM*, pp. 815-824.

Kotlerman, L., Dagan, I. and Kurland, O. (2018), "Clustering small-sized collections of short texts", *Information Retrieval Journal*, Vol. 21 No. 4, pp. 273-306.

Lan, Q., Ma, H. and Li, G. (2018), "Characters-based sentiment identification method for short and informal Chinese text", *Information Discovery and Delivery*, Vol. 46 No. 1, pp. 57-66.

Le, Q. and Mikolov, T. (2014), "Distributed representations of sentences and documents", *International Conference on Machine Learning*, pp. 1188-1196.

Li, J., Dou, Z., Zhu, Y., Zuo, X. and Wen, J.R. (2019), "Deep cross-platform product matching in e-commerce", *Information Retrieval Journal*, pp. 1-23.

Lin, C. and He, Y. (2009), "Joint sentiment/topic model for sentiment analysis", *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pp. 375-384.

Liu, H. (2017), "Sentiment analysis of citations using word2vec. arXiv preprint arXiv:1704.00177",

Liu, M., Fang, Y., Choulos, A.G., Park, D.H. and Hu, X. (2017), "Product review summarization through question retrieval and diversification", *Information Retrieval Journal*, Vol. 20 No. 6, pp. 575-605.

McAuley, J. and Leskovec, J. (2013), "Hidden factors and hidden topics: understanding rating dimensions with review text", *Proceedings of the 7th ACM Conference on Recommender Systems*, *ACM*, pp. 165-172.

Maas, A.L., Daly, R.E., Pham, P.T., Huang, D., Ng, A.Y. and Potts, C. (2011), "Learning word vectors for sentiment analysis", *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*, *Association for Computational Linguistics*, pp. 142-150.

Mikolov, T., Yih, W.T. and Zweig, G. (2013a), "Linguistic regularities in continuous space word representations", *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 746-751.

Mikolov, T. Chen, K. Corrado, G. and Dean, J. (2013b), "Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781".

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S. and Dean, J. (2013c), "Distributed representations of words and phrases and their compositionality", *Advances in Neural Information Processing Systems*, pp. 3111-3119.

Mohammad, S.M. and Turney, P.D. (2010), "Emotions evoked by common words and phrases: using mechanical Turk to create an emotion lexicon", *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, *Association for Computational Linguistics*, pp. 26-34.

Pablos, A.G., Cuadros, M. and Rigau, G. (2015), "V3: unsupervised aspect based sentiment analysis for semeval2015 task 12", *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 714-718.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Vanderplas, J. (2011), "Scikit-learn: machine learning in python", *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830.

Poria, S., Cambria, E. and Gelbukh, A. (2016), "Aspect extraction for opinion mining with a deep convolutional neural network", *Knowledge-Based Systems*, Vol. 108, pp. 42-49.
Rehurek, R. and Sojka, P. (2011), *Gensim–Python Framework for Vector Space Modelling*, NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic.
Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986), "Learning representations by back-propagating errors", *Nature*, Vol. 323 No. 6088, p. 533.
Sanguansat, P. (2016), "Paragraph2vec-based sentiment analysis on social media for business in Thailand", *2016 8th International Conference on Knowledge and Smart Technology (KST)*, *IEEE*. pp. 175-178.
Yu, J. and Jiang, J. (2016), "Learning sentence embeddings with auxiliary tasks for cross-domain sentiment classification", *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 236-246.
Shalaby, W., Zadrozny, W. and Jin, H. (2019), "Beyond word embeddings: learning entity and concept representations from large scale knowledge bases", *Information Retrieval Journal*, Vol. 22 No. 6, pp. 525-542.
Topal, K. and Ozsoyoglu, G. (2017), "Emotional classification and visualization of movies based on their IMDb reviews", *Information Discovery and Delivery*, Vol. 45 No. 3, pp. 149-158.
Wang, X., Zhao, D., Yang, M., Duan, L., Xiang, M.M. and Guo, Q. (2017), "Public opinion dissemination on mobile internet-a case of Ebola", *Information Discovery and Delivery*, Vol. 45 No. 2, pp. 87-100.
Yang, X., Macdonald, C. and Ounis, I. (2018a), "Using word embeddings in twitter election classification", *Information Retrieval Journal*, Vol. 21 Nos 2/3, pp. 183-207.
Yang, M., Qu, Q., Shen, Y., Lei, K. and Zhu, J. (2018b), "Cross-domain aspect/sentiment-aware abstractive review summarization by combining topic modeling and deep reinforcement learning", *Neural Computing and Applications*, pp. 1-13.
Zhang, D., Xu, H., Su, Z. and Xu, Y. (2015), "Chinese comments sentiment classification based on word2vec and SVMPERF", *Expert Systems with Applications*, Vol. 42 No. 4, pp. 1857-1863.

Further reading

Schwenk, H. (2007), "Continuous space language models", *Computer Speech & Language*, Vol. 21, pp. 492-518.

Corresponding author

Sangeet Srivastava can be contacted at: srivassa@kean.edu