Neural Computing and Applications (2022) 34:20213–20229
https://doi.org/10.1007/s00521-022-07531-8

ORIGINAL ARTICLE

Intelligent fake reviews detection based on aspect extraction and analysis using deep learning

Gourav Bathla[1] · Pardeep Singh[1] · Rahul Kumar Singh[1] · Erik Cambria[2] · Rajeev Tiwari[1]

Received: 12 August 2021 / Accepted: 9 June 2022 / Published online: 20 July 2022

Abstract
In the era of social networking and e-commerce sites, users provide their feedback and comments in the form of reviews for any product, topic, or organization. Due to high influence of reviews on users, spammers use fake reviews to promote their product/organization and to demote the competitors. It is estimated that approximately 14% of reviews on any platform are fake reviews. Several researchers have proposed various approaches to detect fake reviews. The limitation of existing approaches is that complete review text is analysed which increases computation time and degrades accuracy. In our proposed approach, aspects are extracted from reviews and only these aspects and respective sentiments are employed for fake reviews detection. Extracted aspects are fed into CNN for aspect replication learning. The replicated aspects are fed into LSTM for fake reviews detection. As per our knowledge, aspects extraction and replication are not applied for fake reviews detection which is our significant contribution due to optimization it offers. Ott and Yelp Filter datasets are used to compare performance with recent approaches. Experiment analysis proves that our proposed approach outperforms recent approaches. Our approach is also compared with traditional machine learning techniques to prove that deep neural networks perform complex computation better than traditional techniques.

Keywords Fake reviews · Machine learning · Neural networks · Deep learning

1 Introduction

In e-commerce and social networking sites, users write reviews about any product or topic [1]. Genuine users write reviews with clear intention so that they can share their good or bad experience about the product which can be helpful for other users. Competitors and users with bad intentions have misused this feature. These users write spam or fake reviews to demote competitor's products or promote their own products. The motivation for writing fake reviews is also due to fact that reviews influence purchase decisions [2, 3], reputation of the product [4, 5] and profit [6]. It is not certain how many spam opinions exist on e-commerce sites, but [7] have mentioned that approximately 8–15% are spam opinions. It is also stated in several research works that competitors hire spammers to write fake reviews. These fake reviews are highly complex to be understood by machines. Furthermore, fake reviews are not detected by recent algorithms efficiently as spammers invent new techniques to overcome these algorithms [8]. There are so many instances where online platforms have to optimize their filtering algorithm to remove fake reviews. These optimized algorithms analyse reviews structure so that further no spammers can post fake reviews on their platforms. Several research works have been carried out to detect spam reviews but the accuracy is not adequate. There is a need for an improved approach that

✉ Rahul Kumar Singh
rk.singh@ddn.upes.ac.in

Gourav Bathla
gbathla@ddn.upes.ac.in

Pardeep Singh
pardeep.singh@ddn.upes.ac.in

Erik Cambria
cambria@ntu.edu.sg

Rajeev Tiwari
rajeev.tiwari@ddn.upes.ac.in

[1] School of Computer Science, University of Petroleum & Energy Studies (UPES), Dehradun, India

[2] School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

can detect opinions with better accuracy. In this research work, reviews aspects and polarity are observed and accuracy is improved using CNN and LSTM hybrid model.

Fake review-based research works have focused on reviewer's behaviours and review text analysis [9]. Researchers have analysed reviewers' behaviours such as number of reviews, extreme ratings, timing of reviews, review length, and emotional strength in review. Spammers can overcome these detection techniques so the efficiency of behaviour-based techniques is not adequate. In review text-based approaches, sentiments and replication of text are used for fake reviews detection. It is stated by researchers that original reviews have contextual content and fake reviews have strong emotional content [10, 11]. In our proposed approach, the most essential part of sentence, i.e. aspects and sentiments, is used for fake reviews detection.

Aspect is the concept based on which the opinion is given for any product or topic [12–14]. Explicit and implicit aspects are specifically used in opinions. Explicit aspects are clearly reflected in opinions such as '***food in the restaurant***', '***screen of phone***', and '***battery of laptop***'. Implicit aspects are extracted from concepts such as laptop is affordable, i.e. price aspect is extracted which is implicit and hidden in opinion. Several techniques such as frequency-based, NLP-based, pointwise mutual information (PMI), word alignment, graph-based, and rule-based are proposed by researchers to extract aspects [15]. Commonsense knowledge with rule-based approaches and dependency tree are also used to extract aspects [16]. Noun/noun phrases are considered as aspects and adjectives/adverbs as opinions [17]. In [18], the authors proposed that all noun phrases are not potential aspects, if PMI is less, then these noun phrases are not considered as aspects. Aspects are employed for sentiment analysis in several research works [19] [20]. In addition, aspects are used in review summarization and opinions extraction by several researchers. However, to the best of our knowledge, aspects are not used for fake reviews detection which is our significant contribution for researchers and web-based organizations to detect spam opinions.

Aspect-level is fine-grained extraction of opinions about features [15]. SentiWordNet [21], SenticNet [22], and WordNet-Affect are used for extracting sentiments in aspect-level analysis. In this research work, SentiWordNet is used for calculating polarity of aspects. SentiWordNet provides synset in WordNet which depends on the context of lexicons. It is proved already by several researchers that spammers replicate existing reviews and replace some words with synonyms and change the polarity of sentence [10, 11]. In our proposed approach, aspect replication is used based on the assumption that spammers use aspects of reviews and change the polarity of aspect opinions. Existing approaches have focused on complete review replication and modification in the review text, polarity, and semantics which increases computational time and degrades accuracy. The advantage of our proposed approach is that it mainly focuses only on the essential components of reviews, i.e. aspects and sentiments. The example of genuine and fake reviews is as follows.

1.1 Genuine review

"Lovely clean and bright room. Superb views, very quiet, comfy bed etc. Excellent service all round and nice staff. For the money and quality it stacked up very well indeed. Great location too."

1.2 Fake review

"Untidy and dark room. Bad views, very noisy, discomforting bed etc. Bad service all round and unfriendly staff. For the money and quality it stacked up very bad. Unpleasant location too."

It is clear from the above review sample that spammers change the polarity of aspects such as room, bed, staff, and location. The sentiment of room is concluded from tags such as lovely, clean, and bright. The polarity of room is changed by using words such as untidy and dark in fake review. This sample is filtered from many lines reviews to better demonstrate our approach. It is evident from above example that there is no need to focus on complete text reviews of many lines and words which just increase computational time.

Spammers have little knowledge about the product, so some words are manipulated in deceptive reviews [23]. In [7], the authors have observed that fake reviews can be analysed using part of speech (POS) tagging. In our approach, POS tagging is used for assigning noun/noun phrases as aspects and adjectives/adverbs/verbs as sentiments. Deep learning techniques outperform traditional classifiers for analysing spam opinions [24]. The reason is that traditional classifiers reach the threshold level and cannot provide accuracy beyond a threshold. In our proposed approach, CNN and LSTM hybrid model is used for aspect replication and sentiment learning. Extracted aspects and respective polarity are fed into CNN model to find aspect replication and filtered aspect replication is fed into LSTM for training and evaluating performance.

The main contributions of this research work are as follows:

1. Efficient POS tagging-based aspect extraction techniques are applied to find aspect and polarity from reviews.

2. Aspect replication in spam reviews is computed instead of complete text replication which makes our approach better as compared to existing approaches.
3. Extracted aspects and sentiments are fed into deep learning model (CNN and LSTM) to analyse fake reviews.
4. Experiments are conducted extensively on Ott [25] and Yelp filter [26] datasets to analyse accuracy. Experiment analysis proves that our proposed approach provides better precision and accuracy as compared to existing approaches.

The rest of the paper is organized as follows. In Sect. 2, related work of aspect extraction and spam opinions techniques are surveyed. In Sect. 3, background and preliminaries are outlined to cover spam opinion detection, aspect extraction techniques, and CNN, LSTM architecture. The proposed approach is elaborated in Sect. 4. Experiment setup and analysis are discussed in Sect. 5. Finally, Sect. 6 concludes the paper with future directions.

2 Related work

Existing research works have focused on review text or reviewer behaviour to detect fake reviews. Sentiments and replication of review text are observed as the most common method used by spammers to write fake reviews. In this section, recent state-of-the-art approaches are surveyed to analyse the merits and demerits of existing research works which motivate us to propose an efficient approach to detect fake reviews with less computational time and improved accuracy.

2.1 Fake reviews analysis using review text

In [27], neural networks are explored to detect deceptive opinions using document-level representations. In this work, the authors have mentioned that previous works have focused on only discrete features based on linguistic views. In this research work, discourse and document-level semantics are considered. CNN model is deployed to learn representation from sentences and it is fed to recurrent neural network to analyse the discourse semantics. Experiments are conducted on three datasets and it is proved that accuracy and F1 are improved by using proposed approach. In [28], authors have stated that traditional classifiers do not provide adequate accuracy for deceptive opinions. Paragraph Vector Distributed Bag of Words (PV-BOW) and Denoising AutoEncoder (DAE)-based approach is proposed to improve accuracy. Dataset is pre-processed using tokenization and lemmatization. ReLU is used as an activation function for hidden layers and Sigmoid function is used as an activation function for output layers. Gold standard dataset is used for experiment analysis. Accuracy, F1-measure, Precision, and Recall are used for the evaluation of the proposed approach. In [29], a reputation score is assigned to reviewers to check whether reviewers are spam or genuine. The advantage of this approach is that there is no need to label large instances. Further k-centre clustering is applied based on time interval. It is stated that spammers post reviews with strong emotional tendencies. Reputation value is calculated by combining content features and reviewer behaviours. Music product review from Amazon is used as dataset for experiment analysis. Precision, Recall, and F-measure are used in this research. The performance of proposed approach is better as compared to existing approaches. In this paper, Amazon product review dataset is used and complete review text is used for experiment analysis. In [13], deep convolution neural network is employed to extract aspects. Authors have used seven-layer neural network to improve precision for aspect extraction. It is stated that there are limitations in conditional random fields and linguistic patterns which need to be improved. Google, Amazon embedding, and SemEval datasets are used for experiment analysis. Precision values are compared with linguistic patterns and it is proved that the proposed approach outperforms LP. In [30], several neural network architectures are discussed specifically for spam opinions. It is also stated in this research work that traditional machine learning techniques do not provide the semantic information of reviews which is necessary for deceptive spam opinions analysis. Experiments are conducted using various neural network architectures such as CNN, RNN, LSTM, and GRU. CNN outperforms other models in terms of accuracy. It is due to the fact that CNN can find complex and high-level features from opinions. N-Gram model is used to extract information from review text in [31]. The ensemble approach is used that combined N-gram and CNN. Experiments are conducted on Yelp dataset to analyse the performance of proposed approach. The limitation of this approach is that several features of reviews and reviewer behaviour are used for detection of fake reviews. This increases the computational complexity. In [24], various machine learning and deep learning models are used for detection of fake reviews. CNN, LSTM, SVM, and kNN models are used to evaluate the performance on Yelp and Ott datasets. The limitations of this approach is that single model is used for fake reviews detection and there is scope of improvements in word embedding and hyperparameter settings. Semi-supervised learning is used to detect fake reviews in [32]. It is stated by authors that labelled dataset can be an issue as accurate labels and large-scale data is required. The optimization is required in this approach to select features that can perform better in less computational time.

2.2 Fake reviews analysis using reviewer behaviour

A graph-based approach is proposed to detect opinion spam using entities in [33]. In this approach, weights are assigned to entities based on their importance. Review polarity strength, terms, rating deviation, reviews per day, etc., are spam features discussed in this research. Accuracy, Precision, Recall, and F-measure are used as an evaluation metric for validating the proposed approach. Accuracy is calculated using various features such as content-based, behaviour-based, relation-based, and proposed features in this research. Accuracy is improved using the proposed set of features. In [9], authors have proposed an approach based on singleton reviews. It is argued by authors that very few research works have focused on reviewers who have given only a single review. There can be the possibility that reviewer can change username and submit spam reviews. Text-based and semantic-based similarity is calculated between reviews. Experiment analysis is conducted on Yelp and Trustpilot datasets. Precision and F1 score are improved by using proposed approach. In [34], autoencoder and neural random forest are used to detect spam opinions. The reason for selecting these models is that autoencoder can use unsupervised representations in features and random forest is used to combine several decision trees. Reviewer behaviour and review contents are used in this research. The entropy of ratings, length of summary, entropy of ratings timing, etc., are used as features. Precision, Recall, and F1-measure are used as evaluation metric and Amazon review dataset is used for experiment analysis. It is proved that better accuracy is provided by the proposed approach as compared to existing approach. However, complete review text is used for analysing semantics which increases computational complexity. In [35], 133 features from content and behaviour-based features are extracted. It is mentioned that class imbalance exists in datasets. This is resolved by using random sampling. The accuracy is improved using sampling. However, multiple classifiers are not deployed in this approach and there is scope of improvement to achieve adequate accuracy on large datasets. The behaviours of reviewers are used in [36] to detect fake reviews. Several identification indicators such as Star User, Deviation Rate, Bias Rate, Review Similarity Rate, Review Relevancy Rate, Content Length, and Illustration are used. There is requirement of using advanced neural network-based models to improve the classification.

In Table 1, recent approaches of fake reviews detection are analysed based on novelty, pros, cons, evaluation metrics, and dataset. It is revealed from our analysis that in existing research works, complete text reviews are used which increases computational complexity. There is need to reduce computational complexity. Furthermore, shallow architectures are used for fake reviews detection, and there is need to optimize neural network setup by using dropout, effective feature selection, and hyperparameter tuning. In existing research works, aspects are not given much focus.

In our proposed approach, only relevant features are used which decreases computational complexity. The complexity analysis is elaborated in Sect. 4. Furthermore, traditional machine learning or deep learning models are employed in most of recent approaches. In our proposed approach, hybrid deep learning model is applied to utilize CNN and LSTM advantages which is described in Sect. 4.

3 Background and preliminaries

3.1 Spam opinion detection

The users are main target for any organization [40]. Spammers target these users to promote their products. Several research works have been carried out to detect spam opinions [32, 41]. The broad categories are review-based and reviewer behaviour-based. In review-based approaches, NLP techniques are applied on text to check fake reviews. Text replication, extreme polarity, rating deviation, etc., are used by researchers extensively to detect spam opinions. The difference in opinions and ratings, max number of reviews, location, number of times of reviews, etc., are employed by researchers. Graph-based techniques are also used for spam opinions [33, 42]. In this research work, several studies are surveyed extensively and concluded that aspects are not used for spam opinions detection.

Several researchers have proposed approaches to detect spam opinions using deep learning. CNN, RNN, LSTM models are deployed for spam opinions. Autoencoder is combined with Random Forest for spam opinions in [34]. In [24], CNN, LSTM, and Multilayer perceptron is employed on labelled as well as an unlabelled dataset.

3.2 Aspect extraction techniques

Various techniques are used for extracting aspects. N-gram based, POS tagging-based, and deep learning-based techniques are used for aspect extraction in existing approaches. It is stated by several researchers that nouns/noun phrases are used as aspects and adverbs/adjectives are used as opinions. Deep learning-based techniques are used for aspect extraction and it is analysed by researchers that it provides better accuracy. In [13], seven-layer deep convolutional network is used to find that word is aspect or non-aspect. Linguistic patterns are combined with neural

Table 1 Summary of Fake reviews detection techniques

Approaches	Technique used	Dataset	Evaluation metrics	Pros	Cons
[24]	CNN, Multilayer Perceptron, and LSTM are employed on labelled and unlabelled datasets for spam opinion	Yelp, Ott	Accuracy	Proposed approach outperforms traditional machine learning approaches	Hybrid models are not used. Training and testing of model is done based on complete review text
[33]	Multi-iterative graph-based opinion detection (MGSD) approach is proposed which considers various entities into consideration for spam opinions	Crowdsourced and Ott	Accuracy	Proposed approach achieved adequate accuracy for crowdsourced and Ott dataset	Traditional graph processing and machine learning techniques are used. Graph neural network and deep learning models can improve accuracy significantly. In content-based features, complete review text is used
[34]	Autoencoder and random forest model to detect spam opinion	Amazon reviews	Precision, F-measure, accuracy	Quality feature selection method combined with autoencoder and random forest yields better accuracy	Complete review text is used for semantics which increases computational complexity
[37]	User and product level semantic review	Mobile01 review, YelpChi, YelpNYC and YelpZip	Precision	Proposed approach achieved 10% improvement over existing approaches on 4 real datasets	Computational complexity is increased due to use of complete review text
[29]	Unsupervised learning – clustering, review content-based and behaviour-based characteristics	Amazon reviews	Precision, F-score	The reputation score of reviewer is computed by using quality of reviews. K-centre clustering technique is used to detect spammers	Amazon review dataset is used and text of complete reviews are analysed for quality
[38]	Statistical analysis for detecting manipulation in reviews	Amazon reviews	Coefficient, median	Ratings deviation and sentiments in content is analysed	Labelled dataset is not used and accuracy is not calculated. The sentiment of complete text review is used
[39]	Vertical ensemble tri-training and active learning for detection of fake reviews in unlabelled data	Ott, Amazon reviews	Precision, F-value	Proposed approach outperforms AdaBoost, Stacking Ensemble technique	Traditional machine learning techniques are compared. Deep learning techniques should be analysed and compared. Complete review text is used for fake reviews analysis

networks to further improve the approach. Experiment analysis proves that deep learning-based aspect extraction technique outperforms existing state-of-the-art approaches.

3.3 LSTM

Neural network-based approaches works in more intelligent manner as compared to traditional techniques [43]. Recurrent neural network (RNN) is used for time-series based and text where terms are dependent on previous terms. The drawback of RNN is that vanishing gradient descent occurs. LSTM is deployed to overcome this limitation. In LSTM, memory unit and forget gate exist which provide long-term dependencies. Experiment analysis proves that better accuracy is achieved using LSTM. Different components of LSTM are as follows.

(i) Cell State: In connected layers, content is saved in cells.

$$c_t = F_t * c_{t-1} + I_t * \overline{c_t} \tag{1}$$

where c_t is current cell state and c_{t-1} is previous cell state. F_t is forget state, and I_t is Input state.

(ii) Input Gate: Input to memory is filtered to remove irrelevant data

$$I_t = \sigma(W_{ix}x_t + W_{ia}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (2)$$

where W_{ix} is weight, x_t is input. h_{t-1} is hidden layer state and b_i is bias.

(iii) Output Gate: This gate forwards only relevant content

$$O_t = \sigma(W_{ox}x_t + W_{oa}h_{t-1} + W_{oc}c_t + b_o) \quad (3)$$

where W_{ax} is weight, x_t is input. h_{t-1} is hidden layer state and b_o is bias.

(iv) Forget Gate: It is used to forget or reset cell's memory.

$$F_t = \sigma\left(W_{fx}x_t + W_{fa}h_{t-1} + W_{fc}c_{t-1} + b_f\right) \quad (4)$$

where W_{fx} is weight, x_t is input. h_{t-1} is hidden layer state and b_f is bias.

3.4 CNN

Convolutional neural network (CNN) is used for extracting local structure. CNN is used by NLP researchers in [44, 45]. In CNN, weight sharing, local filters, and pooling are used [46, 47]. Convolution is combining two functions to produce a new function. In CNN, feature map is constructed by convolution on input data and filtering it. Stride size can be changed during filter move. Usually, stride size is 1. Padding is used to prevent feature map from shrinking. Max pooling is used with convolution layer to reduce the dimensions. This decreases the computational time required to process data. Various layers used in CNN are as follows:

(i) **Input Layer:** Review text is represented as

$$X_i = X_1 \varphi X_2 \varphi X_3 \ldots . \varphi X_n \quad (5)$$

(ii) **Convolutional Layer:** In this layer, features are generated using filter.

$$O_i = f\left(\sum_{i=0}^{n} W_i X_i + b_i\right) \quad (6)$$

where f is nonlinear activation function, W_i is weight, X_i are inputs, and b_i is bias.

(iii) **Max-Pooling Layer:** This layer selects the max features

$$\vec{O} = \max(o) \quad (7)$$

(iv) **Softmax Layer:** Feature vector z = [O_1,O_2,….. O_n] are fed using softmax layer as

$$y_i = softmax(WO + b) \quad (8)$$

4 Proposed approach

In various existing techniques, review text replication is used by spammers due to lack of knowledge of product or organization. In our proposed approach, aspect replication is used for detecting spam opinions instead of complete text replication. As per our knowledge from exhaustive literature review, aspect replication is not used for spam opinions detection. Experiment analysis validates that performance is improved significantly by our proposed approach as compared to existing approaches. Spammers do not have much experience about product or topic so spammers change sentiments of aspects in spam opinions as compared to original reviews. Aspect as a fine-grained specification is very significant for finding review structure. This is the reason that aspects are used for detecting spam opinions in our proposed approach.

This approach is proposed because aspects are the most significant part of review text. This significant part is ignored by researchers in the past. Aspects were only used for sentiment analysis but fine-grained analysis of aspects and respective polarity is not leveraged for detection of fake reviews. Further, our idea was to use hybrid neural network model that can manage complex computations. Our idea is superior to other methods because complete text reviews are used for fake reviews detection in other methods. Our idea is based on selecting the significant part of review text, i.e. aspects. Long-term dependencies in LSTM model also makes our idea superior to other methods.

Complete text is optimized to $\{(x_i,y_i), (x_j, y_j)\ldots.\}$ pair where x_i is used for aspects and y_i is used for sentiments of that particular aspect. These aspects and opinions create bipartite graph which can be analysed for original and fake reviews.

In Fig. 1, it is depicted that aspects and sentiments create bipartite graph. This graph can be used to observe the sentiments changed for particular aspects by spammers. Our proposed approach is divided into different phases as depicted in Fig. 2. The first phase includes pre-processing of datasets. stopwords removal, stemming, lemmatization, and POS tagging. Further, free and bounded morphemes are converted into lower case for better accuracy. In the second phase, aspects are extracted from text of reviews. POS tagging is used to find nouns and noun phrases. In the third phase, sentiments are extracted from reviews text only for extracted aspects. The motive of extracting sentiments is to find the polarity of these aspects. In the next phase, aspects and polarity are fed into CNN and LSTM model to learn aspects replication. The advantage of our approach is that complete review text is not used for feature selection. This saves a lot of computational time and moreover, the

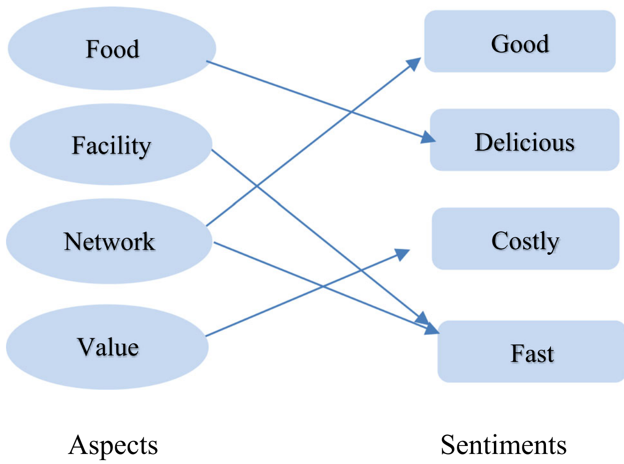


Fig. 1 Bipartite graph for aspects and sentiments

accuracy of deceptive opinion detection is better as compared to state-of-the-art approaches.

In Fig. 3, neural architecture setup used in this research work is depicted. Pre-processed data is fed for splitting data into training and testing. CNN and LSTM model is configured using dropout and Adam optimizer. Precision and accuracy are calculated using sklearn library and charts are plotted using matplotlib library.

$$O_i = \sum_{i=0}^{n} w_i x_i + b_i \tag{9}$$

Output is the combination of weight multiplied by input data and adding bias as mentioned in Eqs. 9. Different activation functions such as *Sigmoid*, *tanh,* and *ReLU* are formulated in Eqs. 10, 11, and 12, respectively.

$$Sigmoid : f(x) = \frac{1}{1 + e^{-x}} \tag{10}$$

$$tanh : f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{11}$$

$$ReLU : f(x) = max(0, x) \tag{12}$$

In our proposed neural network model, *ReLU* activation function is used. Sentiments are extracted based on specific aspects only. This makes our approach very efficient as compared to existing approaches in which the sentiments of all POS tags are used. These sentiments are compared with the copied reviews aspects sentiments.

In Fig. 4, the proposed approach using CNN and LSTM is depicted. The aspects and respective opinions are extracted from the complete review text. Extracted aspects and opinions are sent to CNN for filtering and pooling which provide aspects replication. The aspects replication learning is finalized in LSTM to detect spam opinions. Finally, fake reviews are analysed using different evaluation metrics.

In CNN model, 32 filters with kernel size 3×3, stride 1 and *ReLU* activation function are used in first layer. Max pooling layer with pool_size 2 and Dropout 0.25 is used. In next layer, 64 filters are used with kernel size 3×3 and *ReLU* activation function. After same dropout and max pooling layer, 64 filters are used with kernel size 2×2 and *ReLU* activation function. The replicated aspects are fed into LSTM model with 50 memory units and dropout 0.25. *Softmax* activation function is used in dense layer. Binary_crossentropy is used as loss function and Adam optimizer is used to improve accuracy.

In Algorithms 1 and 2, computational complexity is calculated for existing approaches and proposed approach respectively. The computational complexity of existing approaches is very high because during analysing review text replication, complete text of review R_i has to be compared with complete text of $R_{1.....k}$ except i. If there are n words in R_i and m words in R_j, total computational complexity is O (m*n) and this process is to be iterated for k number of reviews which increases computational complexity to $O(k{*}m{*}n)$. The computational complexity of our proposed approach is very less as only relevant words from text, i.e. aspects and respective sentiments, are to be compared with corresponding aspects and sentiments of other reviews. In our approach, computational complexity of R_i is $O(\log_2(n))$ and to check other reviews $R_{1.....k,}$ the computational complexity is $O(\log_2(n){*} \log_2(m))$ and this process is to be iterated for k number of review, and hence total computational complexity is $O(k{*} \log_2(n){*} \log_2(m))$. This is evident from computational complexity analysis that when a large number of reviews are to be analysed, our proposed approach saves a lot of computational time.

Algorithm 1. Computational complexity of existing approaches

```
Input: Text of reviews R_{1......k} and i:= test_review_number
Output: fake or genuine review
1.  text preprocessing
2.  initialize n:=0, j:=0, m:=0, y:=0, k:=count(reviews),threshold:=0,
    sum:=0
3.  for (y=1 to k) do
      n=count(no_of_words(R_y))
      sum+=n
    end for
4.  threshold=0.5*(sum/k)
5.  for (j=1 to k) do
      for n in R_i  do
         for m in R_j and (i !=j) do
              count replication in m and n words
         end for
      end for
    end for
6.  if count>threshold then
      return 'review is Fake'
    else
      return 'review is Genuine'
```

Algorithm 2. Computational complexity of proposed approach

```
Input: Aspects(x) and sentiments(y) of Text of reviews
R_{1......k} and i:=test_review_number
Output: fake or genuine review
  1   text preprocessing
  2   initialize n:=0, j:=0, m:=0, y:=0, k:=count(reviews),threshold:=0,
      sum:=0
  3   for (y=1 to k) do
              n=count(no_of_aspects(R_y))
              sum+=n
      end for
  4   threshold=0.5*(sum/k)
  5   for (j=1 to k) do
        for (x_i and y_i) in R_i  do
          for (x_j and y_j) and (i !=j) in R_j do
              count replication in aspects and update in sentiments
          end for
        end for
      end for
  5   if count>threshold then
              return 'review is Fake'
      else
              return 'review is Genuine'
```

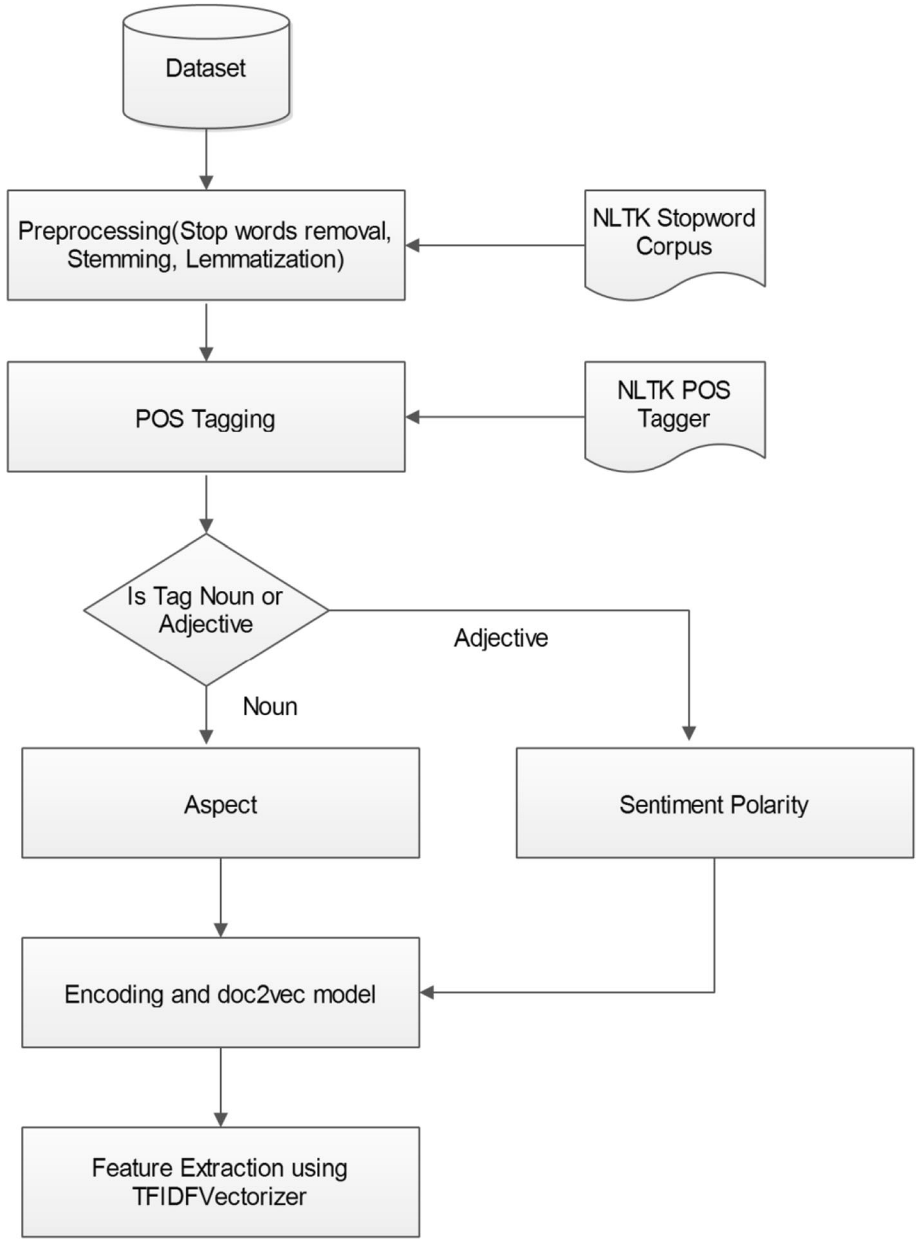


Fig. 2 Flow diagram of proposed approach

In Algorithm 1, complete review text $R_{1\ldots..k}$ is input and fake/genuine review is output. Threshold is calculated based on average of total terms in reviews. Threshold value is set to *50%* of average value. Text replication is calculated using n terms in review R_i and m words in R_j. The review is considered as fake if replication count is above threshold and genuine if count is less than threshold. In Algorithm 2, aspects and sentiments of reviews $R_{1\ldots..k}$ are input and fake/genuine review is output. Threshold is set to 50% of average of aspects in reviews. The aspects x_i and sentiments y_i of review R_i are compared with aspects x_j and sentiments y_j of review R_j. If aspect replication count is above threshold, it is considered as fake and if count is less than threshold, it is considered as genuine.

In this section, the proposed approach is elaborated with architecture, flow diagram, and computational complexity analysis. The main challenge of our proposed approach is that it is completely dependent on aspect replication and respective polarity. If spammers use advanced techniques such as multi-word expression, phrase structure, and reviews without any aspects, then this approach will not perform well. This issue will be resolved definitely in our future research work that will cover various sentiments and

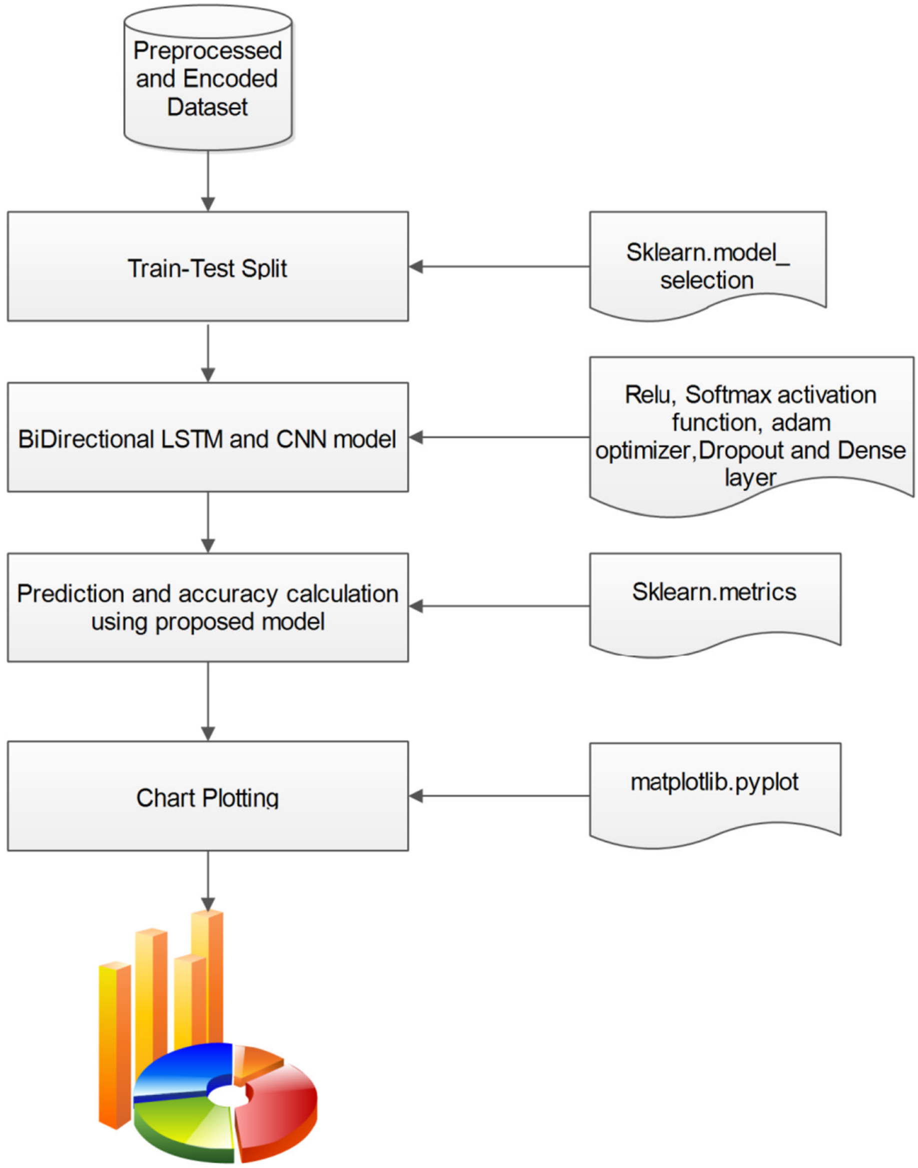


Fig. 3 Neural network architecture setup

word sense disambiguation approaches to deal with advanced techniques. The next challenge is that if spammers use code-mix languages such as English mixed with Hindi, French, and German, then there will be modifications required in our proposed approach. Further embedding and context can be improved by using BERT, GRU, and Graph-neural network approaches. In the next section, experiment setup and results are discussed.

5 Dataset

In the proposed approach, datasets are required which contain truthful and deceptive reviews. Our proposed approach is compared with existing approaches by observing the difference between employing complete review text and aspects-sentiments from these reviews. Ott dataset[1] and Yelp dataset[2] are the most relevant dataset for experiment analysis. Therefore, Ott Dataset [25] is used which contains 1600 reviews with 800 truthful and 800 deceptive reviews. The statistics shows that Ott dataset is completely balanced. Experiments are also conducted on Yelp filter dataset. Yelp dataset [26] contains truthful and deceptive reviews of the restaurant. In Yelp dataset, 58517 total reviews, 50149 truthful reviews, and 8368 deceptive reviews are available. Only 14.3% reviews are deceptive. This clearly reflects that dataset is imbalanced. Only robust approach is able to perform accurate computation. Our

[1] https://www.kaggle.com/datasets/rtatman/deceptive-opinion-spam-corpus.

[2] https://www.yelp.com/.

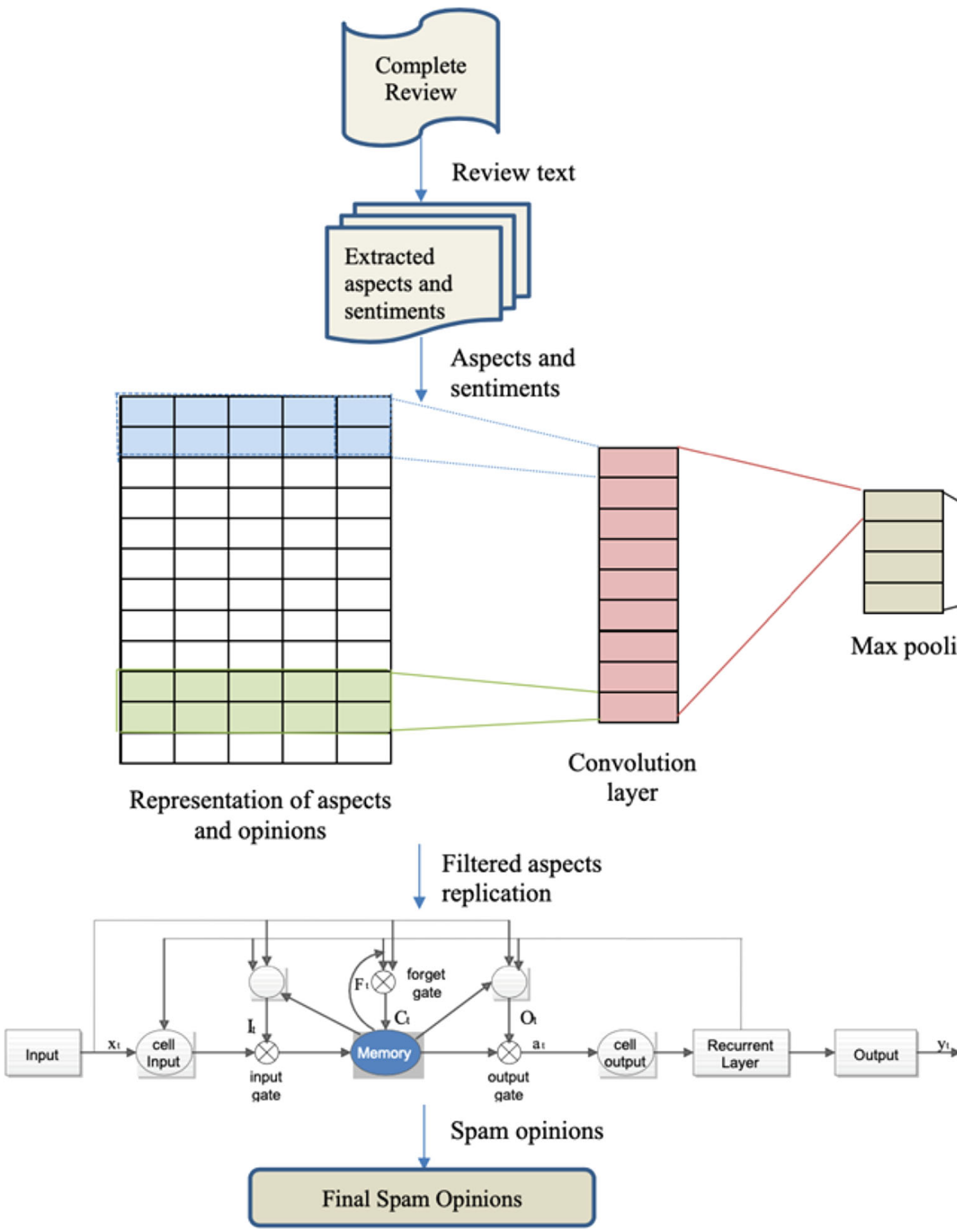


Fig. 4 Schematic diagram of proposed approach

proposed approach performs adequately on this dataset. In Tables 2 and 3, Ott and Yelp datasets statistics are elaborated. The data sample sizes of 20%, 40%, 60%, and 80% are used for evaluating performance for different training and testing data size. In each sample, dataset is divided into 80:20 for training and testing purpose. It is proved in Sect. 6.2 that as size increases for training and testing, performance improves. The reason for the improvement is that more number of features are available for training which enhances fake reviews detection accuracy.

Table 2 Ott dataset statistics

Reviews	Values
Total reviews	1600
Truthful reviews	800
Deceptive reviews	800

Table 3 Yelp filter restaurant dataset statistics

Reviews	Values
Total reviews	58,517
Truthful reviews	50,149
Deceptive reviews	8368
%Deceptive	14.3%

Table 4 Genuine and fake reviews evaluation matrix

Matrix for genuine and fake reviews	Predicted review	
	Fake	Genuine
Actual review		
Fake	True negative	False positive
Genuine	False negative	True positive

Table 5 Comparative analysis of accuracy for proposed approach and existing approaches on Ott dataset

Data sample size	(Chen, Zhao, & Yang, 2015) [48]	(Shahariar et al., 2019) [24]	Proposed approach
20	81.4	91.2	92.3
40	84.3	92.3	93.7
60	87.2	93	94.1
80	91.1	94.1	95.5

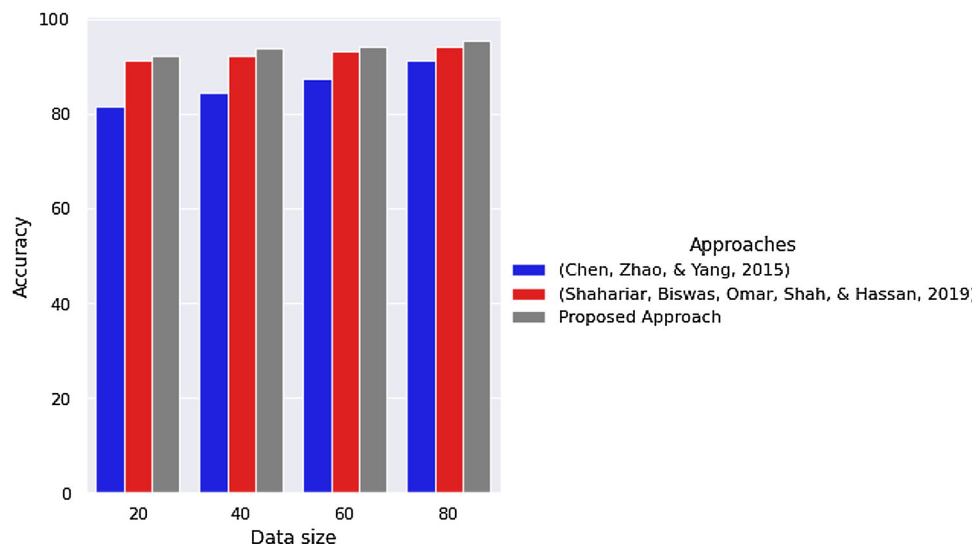


Fig. 5 Accuracy of proposed approach and existing approaches on Ott dataset

Table 6 Comparative analysis of accuracy for proposed approach and existing approaches on Yelp dataset

Data sample size	(Jia et al., 2018) [49]	(Ruan et al., 2020) [11]	Proposed approach
20	78.8	80.1	85.3
40	79.3	81.3	87.7
60	80.5	82.1	88.1
80	81.2	84.4	89.5

6 Experiment analysis

In experiment analysis, pre-processing of reviews such as stop words removal, stemming, lemmatization, tokenization, and POS tagging is implemented to filter data. NLTK library is used to implement pre-processing. Dataset statistics, experiment setup, and evaluation metrics are described in the following subsections.

6.1 Experiment setup

NLTK library is used for stemming, tokenization, regular expression, stop words removal, and pre-processing. Pandas 1.2.3, Keras 2.3.0 and Numpy 1.20.1 libraries are used for creating deep learning model. Experiments are conducted on Google Colab GPU. WordNet, Stopwords, and SentiWordNet corpus are used for pre-processing. POS tagger is used to find tags in reviews. CountVectorizer, TF/IDF and Snowballstemmer are used. Glove text is used for embedding. 40 Epochs are used to validate the accuracy of proposed approach.

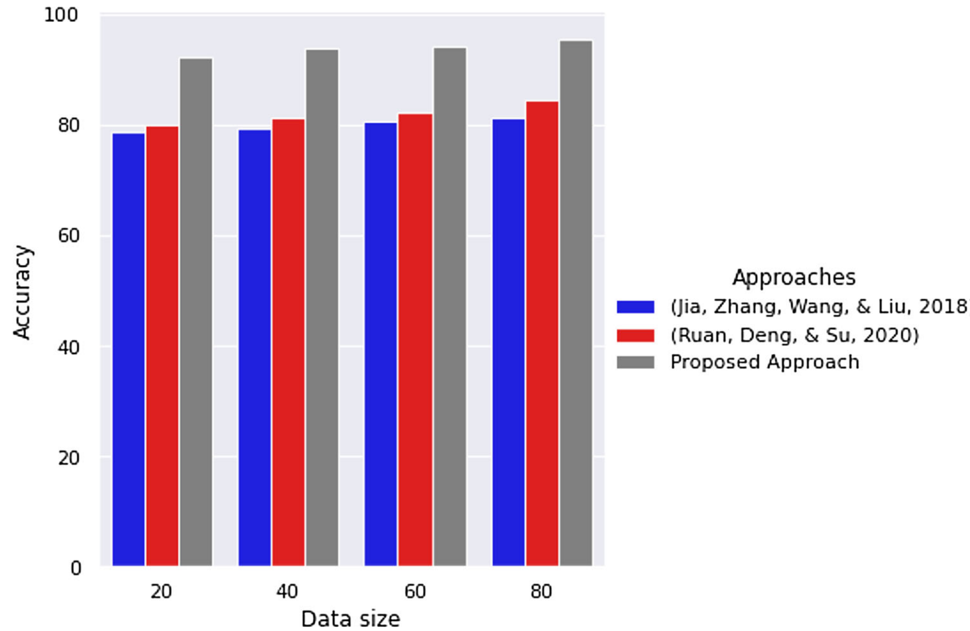


Fig. 6 Accuracy of proposed approach and existing approaches on Yelp filter dataset

Table 7 Comparative analysis of precision for proposed approach and existing approaches on Yelp dataset

Data sample size	MLP	NB	ELM	Proposed approach
20	73.2	78.8	82.5	86.3
40	73.7	79.3	83.9	87.2
60	74.2	80.5	84.5	87.9
80	74.7	81.2	84.8	88.7

Table 8 Comparative analysis of precision for proposed approach and existing approaches on Ott dataset

Data sample size	MLP	NB	ELM	Proposed approach
20	73.1	73.5	82.3	85.3
40	73.7	73.9	82.9	87.7
60	74.1	74.5	83.5	88.1
80	74.5	74.8	84.1	89.5

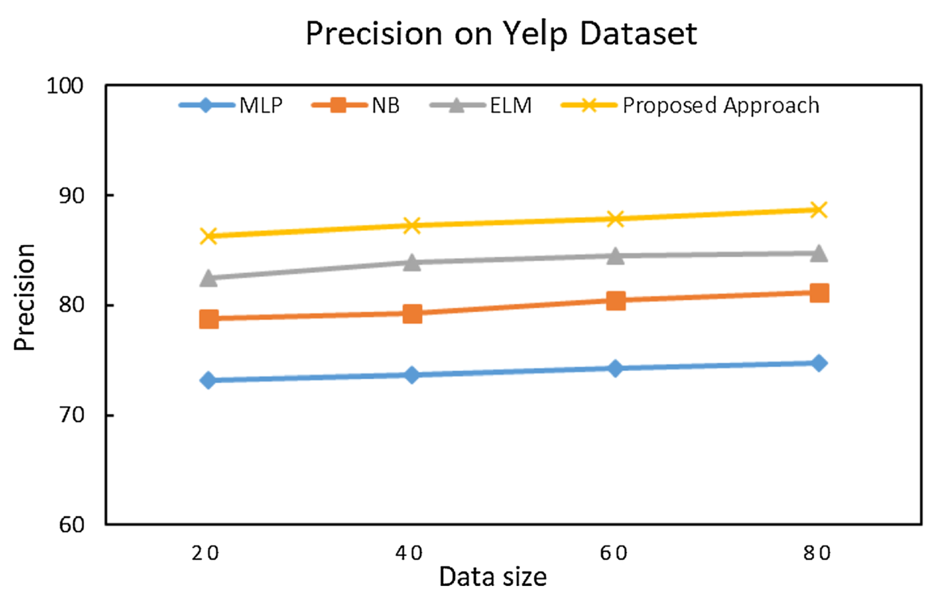


Fig. 7 Precision of proposed approach and existing approaches on Yelp filter dataset

Fig. 8 Precision of proposed approach and existing approaches on Ott dataset

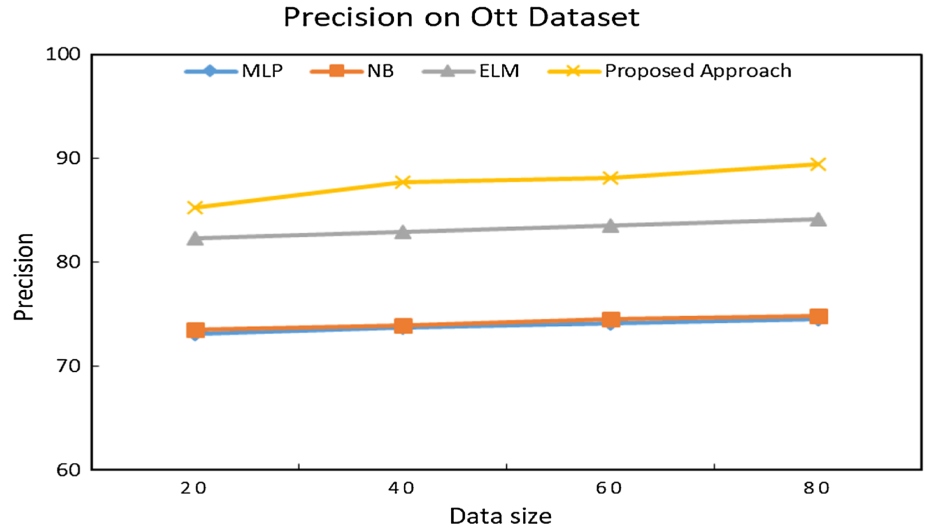


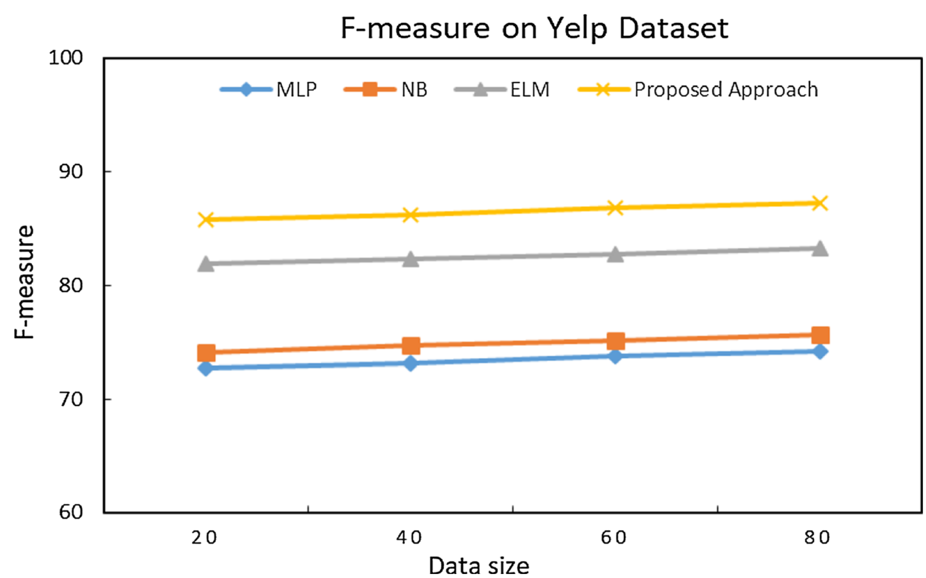
Table 9 Comparative analysis of F-measure for proposed approach and existing approaches on Yelp dataset

Data sample size	MLP	NB	ELM	Proposed approach
20	72.7	74.1	81.9	85.8
40	73.1	74.7	82.3	86.2
60	73.8	75.1	82.8	86.8
80	74.2	75.7	83.3	87.2

Table 10 Comparative analysis of F-measure for proposed approach and existing approaches on Ott dataset

Data sample size	MLP	NB	ELM	Proposed approach
20	72.2	73.9	81.5	85.3
40	72.8	74.5	82.0	85.9
60	73.1	74.9	82.5	86.5
80	73.7	75.5	82.9	87.0

Fig. 9 F-measure of proposed approach and existing approaches on Yelp filter dataset



In Table 4, genuine and fake reviews evaluation matrix is elaborated. If actual review is 'Fake' and predicted review after the proposed approach is also 'Fake', it is assigned as 'True Negative' and if it is predicted as 'Genuine', it is assigned as 'False Positive'. If the actual review is 'Genuine' and predicted review after proposed approach is 'Fake', it is assigned as 'False Negative' and if it is predicted as 'Genuine', it is assigned as 'True Positive'.

The proposed approach is validated using various evaluation metrics such as Precision, F-measure, and Accuracy which are described as follows.

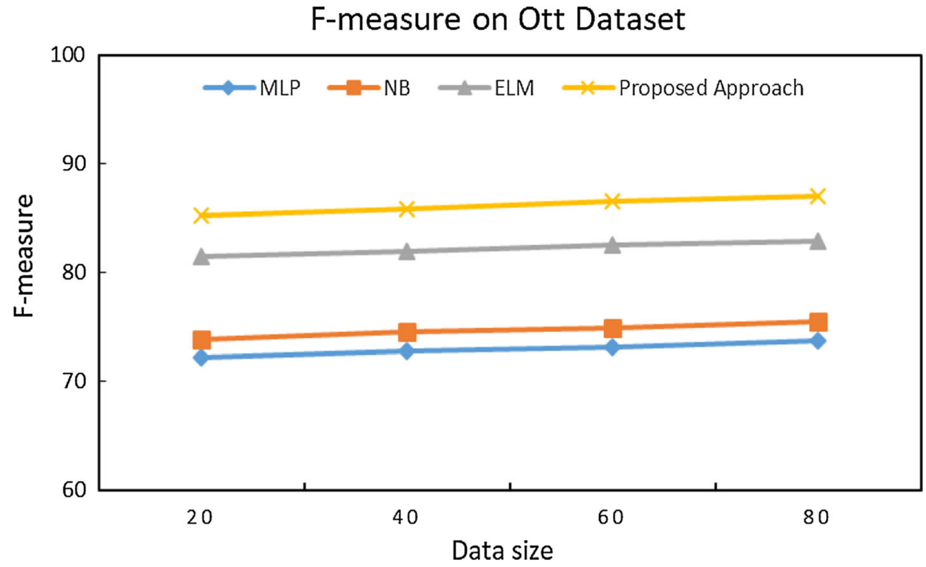


Fig. 10 F-measure of proposed approach and existing approaches on Ott dataset

Precision: It measures the number of positive labelled predicted which are in actual positive class.

$$Precision = \frac{True\,positive\,(TP)}{True\,positive\,(TP) + False\,positive\,(FP)} \tag{13}$$

F-Measure: Harmonic mean of precision and recall is calculated in F-measure. It is used to balance values of precision and recall.

$$F - measure = \frac{2 * precision * recall}{Precision + Recall} \tag{14}$$

Accuracy: Accuracy is the ratio of the number of predictions that are correct to total predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{15}$$

6.2 Results and discussion

Our proposed approach is compared with existing approaches to evaluate Precision, F-measure, and Accuracy. In [48], deep-level linguistic features are used which are derived from parser. Deep learning models are used in [24] for spam opinion. In Table 5, the proposed approach is compared with existing approaches on Ott dataset. It is clear that accuracy is better as compared to existing approaches. The reason for better performance of the proposed approach is that hybrid model is efficiently utilized and the most significant part of review, i.e. aspect and sentiments, is used effectively.

In Fig. 5, the accuracy of proposed approach and existing approaches is depicted for better understanding of readers.

Our proposed approach is also compared with existing approaches using Yelp Filter dataset. In [11], geolocation is used for fake reviews analysis. LSTM is combined with AdaBoost model for fake reviews detection in this approach. In [49], LDA and multi-layer perceptron are combined to evaluate accuracy on Yelp filter dataset. In Table 6, the accuracy of proposed approach is compared with existing approaches.

In Fig.6, accuracy on Yelp filter dataset is depicted. It is clear that our proposed approach outperforms existing approaches.

In experiment analysis, Precision and F-measure is also compared with existing approaches. In [50], ensemble learning module (ELM) is used for selected features extracted from datasets. Furthermore, our approach is compared with multi-layer perceptron (MLP) and Naïve Bayes (NB). Precision and F-measure of our proposed approach is better than MLP and NB as single classifier provide inconsistent performance on low-level features. Further, our approach also outperforms ELM as traditional classifiers such as J48, support vector machine (SVM), and linear regression are used in ELM which have limited capability in complex feature selection and extraction.

In Table 7 and Fig. 7, it is clearly observed that precision of proposed approach is better as compared to existing approaches on Yelp dataset.

The Precision of proposed approach is better as depicted in Table 8 and Fig. 8. Experiment analysis using various evaluation metrics has proved that our proposed approach based on aspect extraction and replication outperforms existing approaches. F-measure is also calculated to validate the effectiveness of our proposed approach.

In Table 9 and Fig. 9, F-measure of our proposed approach is compared with recent approaches on Yelp dataset. In Table 10 and Fig. 10, F-measure is compared on Ott dataset. It is revealed from experiment analysis that our aspect extraction and replication-based approach outperforms recent approaches.

7 Conclusion and future directions

In this proposed work, an efficient approach is designed and evaluated for fake review detection. It uses aspects instead of complete detailed review text for analysis. Spammer replicates the aspects and changes the sentiments of aspects in their fake reviews. Therefore, sentiments of aspects are computed using POS tagging and SentiWordNet. Thus such extracted aspects are fed into CNN and LSTM hybrid model for aspect replication and fake review detection. This approach of using essential part of reviews, i.e. aspects to train CNN and LSTM hybrid model, saves the computational time and offers better accuracy than peer compared approaches. Ott and Yelp filter datasets are used to compute precision and accuracy. The computational efficiency is found to be of order of $O(k^* \log_2(n)^* \log_2(m))$, which is better than the peer compared approaches. Further, proposed approach is compared with MLP, NB, and ELM techniques using Precision and F-measure. Experiment analysis validates that the proposed approach outperforms peer competing techniques. The proposed approach achieves accuracy *92.3%, 93.7%, 94.1%, and 95.5%* using sample size *20, 40, 60, and 80,* respectively, on Ott dataset. Further, accuracy *85.3%, 87.7%, 88.1%, and 89.5%* is achieved using sample size *20, 40, 60, and 80,* respectively, on Yelp dataset. Precision *86.3, 87.2, 87.9, and 88.7* is achieved using sample size *20, 40, 60, and 80,* respectively, on Yelp dataset. F-measure *85.8, 86.2, 86.8, and 87.2* on Yelp dataset, and *85.3, 85.9, 86.5, and 87.0* on Ott dataset is achieved using sample size *20, 40, 60, and 80,* respectively. However, if code mix, i.e. use of country-specific languages such as Hindi, French, German, Russian, and Arabic, are used in any fake review, then to detect such fake reviews some more robust datasets are need of future to improve on accuracy. In future, this approach can be applied on BERT or graph neural network to compare accuracy.

Declarations

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

References

1. Cambria E, Schuller B, Liu B, Wang H, Havasi C (2013) Statistical approaches to concept-level sentiment analysis. IEEE Intell Syst 28(3):6–9
2. Bhuvaneshwari P, Nagaraja Rao A, Harold Robinson Y (2021) Spam review detection using self attention based CNN and bi-directional LSTM. Multimedia Tools Appl 1–18
3. Heydari A, Ma T, Salim N, Heydari Z (2015) Detection of review spam. Exp Syst Appl Int J 42(7):3634–3642
4. Li J, Pin L, Xiao W, Yang L, Zhang P (2021) Exploring groups of opinion spam using sentiment analysis guided by nominated topics. Expert Syst Appl 171
5. Filieri R, Alguezaui S, McLeay F (2015) Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Manag 51(C):174–185
6. Dellarocas C (2006) Strategic manipulation of internet opinion forums: implications for consumers and firms. Manag Sci 52(10):1577–1593
7. Ott M, Cardie C, Hancock J (2012) Estimating the prevalence of deception in online review communities. In: Proceedings of the 21st international conference on world wide web, Lyon, France
8. Wu Y, Ngai EW, Wu P, Wu C (2020) Fake online reviews: literature review, synthesis, and directions for future research. Decis Support Syst 132:113280
9. Sandulescu V, Ester M (2015) Detecting singleton review spammers using semantic similarity. In: WWW'15 Companion: proceedings of the 24th international conference on World Wide Web, Florence, Italy
10. Jindal N, Liu B (2007) Review spam detection. In: Proceedings of the 16th international conference on World Wide Web, Alberta, Canada
11. Ruan N, Deng R, Su C (2020) GADM: manual fake review detection for O2O commercial platforms. Comput Secur 88:101657
12. Ma Y, Peng H, Khan T, Cambria E, Hussain A (2018) Sentic LSTM: a hybrid network for targeted aspect-based sentiment analysis. Cogn Comput 10(4):639–650
13. Poria S, Cambria E, Gelbukh A (2016) Aspect extraction for opinion mining with a deep convolutional neural network. Knowl-Based Syst 108:42–49
14. Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowledge-Based Syst 235:107643
15. Rana T, Cheah Y (2016) Aspect extraction in sentiment analysis: comparative analysis and survey. Artif Intell Rev 46(4):459–483
16. Poria S, Cambria E, Ku L-W, Gui C, Gelbukh A (2014) A rule-based approach to aspect extraction from product reviews. In: Proceedings of the 2nd workshop on natural language processing for social media (SocialNLP), Dublin, Ireland
17. Liu K, Xu L, Zhao J (2012) Opinion target extraction using word-based translation model. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, Jeju Island, Korea
18. Popescu A-M, Etzioni O (2007) Extracting product features and opinions from reviews. In: Kao A, Poteet SR (eds) Natural language processing and text mining. Springer, London, pp 9–28
19. Do HH, Prasad PW, Maag A, Alsadoon A (2019) Deep learning for aspect-based sentiment analysis: a comparative review. Expert Syst Appl 118:272–299
20. Bathla G, Singh P, Kumar S, Verma M, Garg D, Kotecha K (2021) Recop: fine-grained opinions and sentiments-based recommender system for industry 5.0. Soft Comput 1–10

21. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. Lrec 10(2010)
22. Cambria E, Liu Q, Decherchi S, Xing F, Kwok K (2022) SenticNet 7: a commonsense-based neurosymbolic AI framework for explainable sentiment analysis. LREC
23. Li J, Cardie C, Li S (2013) Topicspam: a topic-model based approach for spam detection. In: Proceedings of the 51st annual meeting of the association for computational linguistics, vol 2: Short Papers, Sofia, Bulgaria
24. Shahariar GM, Biswas S, Omar F, Shah FM, Hassan SB (2019) Spam review detection using deep learning. In: 10th Annual information technology, electronics and mobile communication conference (IEMCON), Vancouver, BC, Canada
25. Ott M, Choi Y, Cardie C, Hancock J (2011) Finding deceptive opinion spam by any stretch of the imagination. In: Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, Portland, Oregon, USA
26. Mukherjee A, Venkataraman V, Liu B, Glance N (2013) What yelp fake review filter might be doing. In: Proceedings of the international AAAI conference on weblogs and social media (ICWSM-2013), Massachusetts, USA
27. Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: an empirical study. Inf Sci 385–386:213–224
28. Fahfouh A, Riffi J, Mahraz MA, Yahyaouy A, Tairi H (2020) PV-DAE: a hybrid model for deceptive opinion spam based on neural network architectures. Expert Syst Appl 157:113517
29. Zhong M, Tan L, Qu X (2020) Identification of opinion spammers using reviewer reputation and clustering analysis. Int J Comput Commun Control 14(6):759–772
30. Anass F, Jamal R, Mahraz MA, Ali Y, Tairi H (2020) Deceptive opinion spam based on deep learning. In: 4th International conference on intelligent computing in data sciences (ICDS), Morocco
31. Javed MS, Majeed H, Mujtaba H, Beg MO (2021) Fake reviews classification using deep learning ensemble of shallow convolutions. J Comput Soc Sci 4(2):883–902
32. Tian Y, Mirzabagheri M, Tirandazi P, Mojtaba Hosseini Bamakan S (2020) A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM. Inf Process Manag 57(6)
33. Noekhah S, Salim N, Zakaria NH (2020) Opinion spam detection: using multi-iterative graph-based model. Inf Process Manag 57(1):102140
34. Dong M, Yao L, Wang X, Benatallah B, Huang C, Ning X (2020) Opinion fraud detection via neural autoencoder decision forest. Pattern Recogn Lett 132:21–29
35. Budhi GS, Chiong R, Wang Z, Dhakal S (2021) Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. Electron Commer Res Appl 47:101048
36. Liu P, Zhenning X, Jun A, Wang F (2017) Identifying indicators of fake reviews based on spammer's behavior features. In: IEEE international conference on software quality, reliability and security
37. Yuan C, Zhou W, Ma Q, Lv S, Han J, Hu S (2019) Learning review representations from user and product level information for spam detection. In: International conference on data mining (ICDM), Beijing, 2019.
38. Hu N, Bose I, Koh NS, Liu L (2012) Manipulation of online reviews: an analysis of ratings, readability, and sentiments. Decis Support Syst 52(3):674–684
39. Yin C, Cuan H, Zhu Y, Yin Z (2021) Improved fake reviews detection model based on vertical ensemble tri-training and active learning. ACM Trans. Intell. Syst. Technol. (TIST) 12(4):1–19
40. Singh RK, Singh P, Bathla G (2020) User-review oriented social recommender system for event planning. Ingénierie des Systèmes d'Information 25(5):669–675
41. Gong M, Gao Y, Xie Y, Qin AK (2020) An attention-based unsupervised adversarial model for movie review spam detection. IEEE Trans Multimedia.
42. Wang G, Xie S, Liu B, Yu PS (2011) Review graph based online store review spammer detection. In: 11th international conference on data mining, Vancouver, Canada
43. Panda A, Yadlapalli B, Zhou Z (2021) Credit card fraud detection through machine learning algorithm. Big Data Comput Vis 1(3):140–145. https://doi.org/10.22105/bdcv.2021.142231
44. Collobert R, Weston J, Bottou L, Karlen MK, Kavukcuoglu KP (2011) Natural language processing (Almost) from scratch. J Mach Learn Res 12(1):2493–2537
45. Kim Y (2014) Convolutional neural networks for sentence classification. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), Doha, Qatar
46. Khodaverdian Z, Sadr H, Edalatpanah SA, Solimandarabi MN (2021) Combination of convolutional neural network and gated recurrent unit for energy aware resource allocation. arXiv preprint arXiv:2106.12178
47. Khodaverdian Z, Sadr H, Edalatpanah SA (2021) A shallow deep neural network for selection of migration candidate virtual machines to reduce energy consumption. In: 7th international conference on web research (ICWR)
48. Chen C, Zhao H, Yang Y (2015) Deceptive opinion spam detection using deep level linguistic features. In: Natural language processing and Chinese computing, Nanchang, China
49. Jia S, Zhang X, Wang X, Liu Y (2018) Fake reviews detection based on LDA. In: 4th international conference on information management (ICIM), Oxford, UK
50. Khurshid F, Zhu Y, Xu Z, Ahmad M, Ahmad M (2019) Enactment of ensemble learning for review spam detection on selected features. Int J Comput Intell Syst 12(1):387–394
51. Singh RK, Sachan MK, Patel RB (2021) 360 degree view of cross-domain opinion classification: a survey. Artif Intell Rev 54(2):1385–1506
52. Rasheed F, Wahid A (2021) Learning style detection in E-learning systems using machine learning techniques. Expert Syst Appl 174:114774
53. Salminen J, Kandpal C, Kamel AM, Jung SG, Jansen BJ (2022) Creating and detecting fake reviews of online products. J Retail Consum Serv 64:102771
54. Shahariar GM, Biswas S, Omar F, Shah FM, Hassan SB (2019) Spam review detection using deep learning, In: IEEE 10th annual information technology, electronics and mobile communication conference (IEMCON), pp 0027–0033