This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Aspect-based sentiment analysis using a hybridized approach based on CNN and GA

ADNAN ISHAQ[1], SOHAIL ASGHAR[2], AND SAIRA ANDLEEB GILLANI.[3]

[1]Department of Computer Science, COMSATS University Islamabad, Islamabad PK 44000 (e-mail: sattidani121@gmail.com)
[2]Department of Computer Science, COMSATS University Islamabad, Islamabad PK 44000(e-mail: Sohail.Asg@gmail.com)
[3]Department of Computer Science, Bahira Univeristy Karachi Campus, Karachi, PK (e-mail: Sairagillani.bukc@bahria.edu.pk)

Corresponding author: Adnan Ishaq (e-mail: sattidani121@gmail.com).

ABSTRACT Sentiment analysis is a computational analysis of unstructured textual data, used to assess the person's attitude from a piece of text. Aspect-based sentimental analysis defines the relationship among opinion targets of a document and the polarity values corresponding to them. Since aspects are often implicit, it is an extremely challenging task to spot them and calculate their respective polarity. In recent years, several methods, strategies and improvements have been suggested to address these problems at various levels, including corpus or lexicon-based approaches, term frequency and reverse document frequency approaches. These strategies are quite effective when aspects are correlated with predefined groups and may struggle when low-frequency aspects are involved. In terms of accuracy, heuristic approaches are stronger than frequency and lexicon based approaches, however, they consume time due to different combinations of features. This article presents an effective method to analyze the sentiments by integrating three operations: (a) Mining semantic features (b) Transformation of extracted corpus using Word2vec (c) Implementation of CNN for the mining of opinion. The hyperparameters of CNN are tuned with Genetic Algorithm (GA). Experimental results revealed that the proposed technique gave better results than the state-of-the-art techniques with 95.5% accuracy rate, 94.3% precision rate, 91.1% recall and 96.0% f-measure rate.

INDEX TERMS Aspect-based Sentiment Analysis; Convolutional Neural Network; Genetic Algorithm;

I. INTRODUCTION

The Natural Language Processing (NLP) always has a great importance in Sentiment Analysis (SA). SA is a subfield of NLP, could also be referred to as opinion mining. Opinion mining aims to identify and extract emotions attached to the piece of text. Nowadays, SA become much popular in online communities for processing social media data (blogs, tweets, reviews, and forum discussion). SA evaluates the text, either positive, neutral, or negative. It can acquire the level of polarity (high, low, medium) of the text as well [1]. It explores the emotions, mindset and attitude from gathering feedback as reviews on numerous websites.

Opinion mining could be carried out using two methods; the first is a machine learning method in which SA centered on the rate of co-occurrence of words while the second approach is lexicon-based, which comprises of ontologies, lexicons, and semantic networks to find out the polarity attendant by the text. The technique of machine learning is expanded over two parts, supervised and unsupervised. In supervised learning, we train the machine by the mean of data that is well labeled i.e. already tagged with the correct answer. Then provide the machine with a new set of extended data, thus a supervised learning algorithm to analyze the training data and give the correct result. In unsupervised learning, the machine is trained through the information that is neither labeled nor classified, allows the algorithm to work without any guidance.

The development of network technology provides a new way to communicate over user-created content (blogs, forums, social networks, websites reviews, e-commerce websites, etc.). With this exponential growth, individuals and organizations have a keen interest in data mining technology that uses this subjective information resource. SA is one of the most frequently used research areas in computer science, aimed at identifying and extracting user opinions [2].

The SA is interested in extracting the emotions conveyed in the paragraphs of the text. Pioneering work on Aspect-Based Sentiment Analysis (ABSA) believes that SA research

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

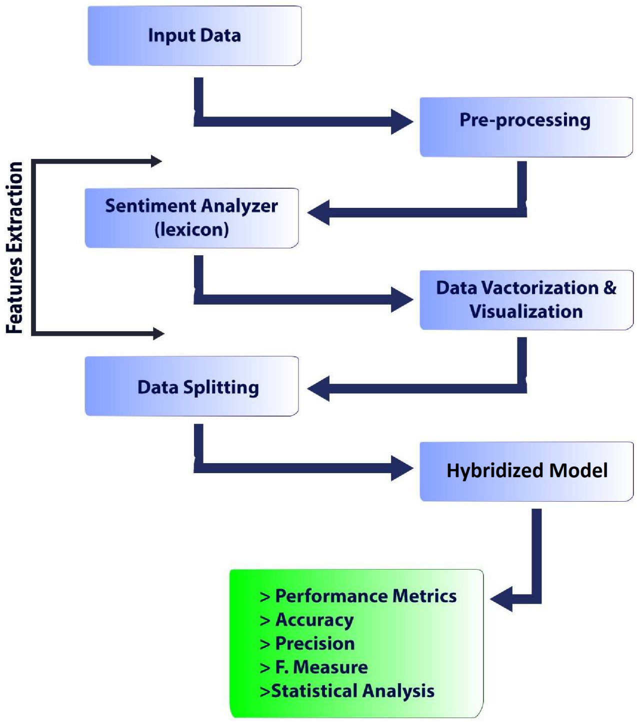

FIGURE 1: Proposed model

can be performed at three levels: document, sentence, word, or aspect. In document level, SA assumes how each document provides an opinion on a single entity. It is assumed that the document represents only one topic, but often this is not the case. In a sentencing level analysis, each sentence is treated as a different entity and it is considered that the sentence should hold merely one supposition. Subjective characterization and grouping estimation are two undertakings of the sentence-level analysis. Though, this is not very handy in all circumstances. For example, a comment about a cell phone can have contradictory emotions in different parts of the cell phone. The reviewers praise the focus and image quality of the camera and sometimes criticize its power consumption in the same sentence. To convey different opinions about different aspects of the same product, you need to act on the ABSA method. In comparison with traditional SA, ABSA presents some challenges: (a) Link every portion of the text to a side (b) Determine the part of the text that describes the same side (e.g., a reviewer may be concerned about a battery and another reviewer discusses power consumption life, both are pointing to the identical side) and (c) comparative sentences, etc. Though, most of the present work on ABSA is concentrated in English and few in other languages [3]. The main function of ABSA is to enable investigations with component-based aspects, It may have three operations, the first is to remove the questions that have been identified by the conclusion holder, The second is the extremeness of the views that determine the characteristics of different categories i.e. (affirmative, negative, and fair), and the final job is to identify words equivalent to the characteristics.

Over the last few decades, the amount of interactive content on the Internet has grown substantially worldwide, and these interactive content websites allow users to publish almost any content. The content you post contains your judgments and emotional state about something. The content is available on the web in the shape of reviews, blogs, and literature. This data is present in the shape of the structure and unstructured data. The structural data is one that is comprised of transactional data, numeric data, and data that is well organized. Such data is important for organizations to take some serious decisions while unstructured data exist in the shape of a text document, email, pdf file and reviews. Mining this data to find people and their views who are interested in a particular item or service is very significant to many stakeholders, including governments and businesses. Many of these tasks, such as big data and SA, pose many challenges, like the complexity of processing data and the complexity of processing unstructured text written in natural language. The language that is used on the Internet is different for each individual, therefore a systematic approach is needed for opinion containing data preprocessing and text classification.

ABSA take the reviews and study the aspect associated with it. Every single aspect has a score associated with it based on this score the polarity of the whole sentence could be measured. After that, the reviews were classified as trained and tested, and the model is now trained. The trained model is able to find polarity in upcoming reviews. The working design of the methodology is presented in Fig.1, which shows the training and testing units of the reviews that are passed to the machine learning algorithm, and sentiment analysis is carried out. The content available on the web in the shape of reviews, blogs and literature is growing on a daily basis. This data is present in the shape of structure and unstructured data. Structure data is one which is comprised transactional data, numeric data and data which is well organized. Such a data is important for organizations to take some serious decisions while unstructured data available in the form of text document, email, pdf file and reviews. Aspect based sentimental analysis take the reviews and study the aspect associated within it. Each single aspect has a score associated with it on the basis of this score the polarity of whole sentence could be measured. After that, the reviews were classified as trained and tested, and the model is now trained. The trained model is able to find polarity in upcoming reviews. The working design of the methodology is presented in Fig. 1, Showing the training and testing units of the study that are passed to the machine learning algorithm, sentimental analysis is carried out. The language that is used on the internet is different for each individual, therefore a systematic approach is needed for opinion containing data preprocessing and text classification.

A. RESEARCH CONTRIBUTION

This paper puts forward an efficient procedure sentimental analysis.

- A structural technique for extraction of features from unsaturated data of hotel, automobiles and movie reviews has been defined
- Preprocessing of data.
- Semantic features have been extracted.
- Convolutional neural network (CNN) is being used to extract opinions
- The tuning of CNN parameters is done with a multi-objective function using Genetic Algorithm (GA).

The order of the paper is as follow Section: 2 include literature review, Section: 3 contains Problem statement, Section: 4 and Section: 5 presents the proposed methodology and proposed techniques, Section: 6 contain experimental design while the study of the results in Section: 7 and eventually Section: 8 concludes the paper.

II. LITERATURE REVIEW

This section reflects a comprehensive literature review of the research work carried out in the SA field. Optimizing data structures, optimizing query technology and optimizing parallelism can improves system's efficiency. As a built model, it performs well enough for users competing in power, response rate, and expansibility. Al-molsmi et al. [4] submitted a commentary on cross-domain SA, as SA has gained much consideration in current years. Therefore, technology, methods, detailed overview of methods can be used for cross-domain SA by providing a comprehensive literature introduction. Jianqiang et al. [5] also performed a comparative study to analyze the preprocessing methods used in the analysis of twitter. The accuracy and F1-measurement of the classifier for twitter classification can be enhanced with the use of a suitable preprocessing method.

Wei et al. [6] has developed an ontological system for the analysis of the product review. A sentiment ontology tree is designed to reflect the information of product functionality and the emotions associated with it in a hierarchical relationship. Human-labeled data analysis ensures greater accuracy. This model can extract features automatically in future. It may reduce manual effort, while it could compromise a bit of accuracy Karagoz, P., Kama, [7], presented a framework for SA of Turkish informal texts using Frequency Based Aspect Extraction with Sentiment Word Support (FBAE-SWS), and Web Search Based Aspect Extraction (WSBAE). This article emphasized on improving aspect extraction as an unsupervised method and detect the polarity depending on the aspect of the emotional word. It also provides a tool, including a Graphical User Interface (GUI) for implementing the proposed algorithm and visualizing the analysis results.

In [8], a new SA integration method based on POS and n-gram has been proposed. It also considered semantics, emotional cues, and the order between words called EnSWF. Some comments or opinions include idioms that help express emotions, but these idioms are not considered in the proposed method of analysis. Considering the document metaphor, satirical and ironic challenges the proposed method. In [9], a method based on deep learning has been proposed to categorize user opinions expressed in comments (called RNSA). It performs well with the provided dataset, however, execution failed while processing a large dataset. In [10], author used a deep feedforward neural network with a global mean pool and a long-term short-term memory model with dense layers to identify subjective information from text documents and solve emotional prediction problems at the sentence level. In [11], an RNN-based method is proposed that uses a word vector as the input acquired from the GloVe method. The model they suggest took into account the impact of market trends, and other factors on sentiment predictions.

Rezaeinia et al. [12], an improved word vector model are proposed by combining dictionary-based methods, part-of-speech tagging methods, word localization techniques, and word2vec / GloVe methods. To learn to embed emotion-specific words, [13] describes how to integrate emotional information in the text. They also propose how to develop a neural network model to handle fine SA. SA of Hindi comments, [14], learning model centered on CNN. They tried distinctive settings to modify input size and regularization technology to determine CNN parameters Output size, dropout rate, period, activation function, etc. To improve the accuracy of SA of Arabic data CNN and LSTM is proposed in [15]. This model uses soft voting in it, the predicted category probabilities of the data are averaged over the two CNNs. Then select the LSTM model and the category with the highest average as the final selection Integrated model prediction.

Lee et al. [16], assessed the impact of text quality Comments established on comment length, word count, and readability. Emotion analysis task is performed on movies dataset. Three models of deep learning family (simple CNN, LSTM, and RNN). The authors claim the dataset is short and easy to read higher accuracy compared to long and short length data sets readability. The CNN-based method proposed in [17], user behavior (personal characteristics and social activity) for emotion analysis. In [18], authors proposed a new model of neural network with two hidden layers. The first layer shows sentence vectors that indicate sentences in short-term and long-term memory networks and the second layer encodes sentence relationships into a document representation. It also suggests an improved way to first clean the dataset and remove the emotionally less polar sentences in the dataset. In [19], SA used a rule-based method with the help of SentiWordNet and SVM for feature extraction along with term frequency and inverse document frequency.

In [20], author suggested a way to determine the sentiment of reviews based on the hotel dataset. Hotel reviews are preprocessed into a list of terms. First, the potential Dirichlet Allocation (LDA) was determined Glossary; Semantic similarity then sorts the term list according to the topics generated by potential Dirichlet assignments. (LDA) integrated into the five sides of the hotel. Next, when computing the resemblance, the Frequency anti-clustering frequency (TF-ICF) method. Finally, classify consumer sentiment (satisfied or not) with word embedding and long-term memory (LSTM). Kontopoulos et al. [21] presented the execution of a real ontology-based deployment methodology for more effectual twitter's tweet SA. The suggested model utilizes the ontology methodology to analyze the twitter in a more effective way not just by giving a complete tweet score, but also by analyzing any hidden dimension of tweets, instead, provide scores for a certain tweet which makes the process of feature engineering more reliable to achieve a higher degree of accuracy. Ciric et al. [22] suggested a model for evaluating the tweet's sentiment. Several machine learning models are used in the proposed work to conduct opinion mining, and their assembly is carried out in order to achieve better outcomes. The gathered outcomes are equated with various approaches for thorough analysis. The tabular view of literature review is shown in Table 1.

Bahraini et al. [23] suggested a hybrid model for the study of twitter tweets by target-based sentiment. A hybrid solution is shown to outperform other strategies as well as demonstrated better results with different features and functions. Freitas et al. [24] recommended a model for extraction of features through ontology technique, and data sources are data sets for a movie review and hotel reviews. The analysis of sentiment is remarkably effective with higher efficiency and accuracy in their experiments. Zhao et al [25] the authors proposed a Chinese-language approach that achieved better results than previous models. It has acquired greater accuracy. In [26], author introduced a model in which the Iris General Electronics tweets were classified in 3-class sentiment. Using a supervised learning method and lexicon-based subjectivity score, 61.6 percent accuracy was achieved.

Sam et al. [27] simplified models has been developed to evaluate customer comments about electronic products on social networking sites. The keyword extraction and electronic product ontology creation is about understanding consumer behavior online. Ravi et al. [28] and Yadollahi et al. [29] tended to an investigation of the systems utilized for opinion mining. Clavel et al. [30] additionally utilized SA in human-agent interaction since conclusion digging was once in a while utilized for human-agent interaction prior. Dragoni et al. [31] creating a new ontology of common sense for an analysis of sentiments. Kim [32] used CNN to distribute exam progress. This course was trained on sentence-level classification tasks based on pretrained word vectors. A basic CNN technique with few hyperparameters adjustments and static vectors delivers unimaginable results on various benchmarks. This technique is applied using a single layer convolutional design.

Stojanvaski et al. [33] recommends a model having a profound convolutionary neural system. Pre-prepared word vector inserting is accomplished in this work by actualizing unaided learning in huge companies. The informational index utilized is a piece of SemEval 2015 and the outcome underpins the Twitter SemEval 2015 benchmark and no high-quality highlights have been utilized in this structure. The F1score surveyed was 64.85%, Future scope of this structure is that the entire framework would be recreated through utilizing twitter-based organizations that could make improvement in effectiveness of the model under exchange. Ouyang et al. In [34], an approach was proposed that included Word2Vec and CNN, and three sets of convolutional layer and max pooling layer were used in this project. This is the first hypothesis that word2vec and CNN use statements using a seven-layer structure model, a linear parameter rectifier with standardized functionality, and crimping techniques. Publicly accessible movie review data uses five different tags. Negative, some negative, neutral, some positive, and positive. System test accuracy is 45.4

Jindal et al. [35] built the image system, emotion recognition framework works with deep and complex neural network. This methodology has been pulled together with huge data for the identification and transition of objects. The data set utilized was the image called Flicker. In [36] differential neural system model has been proposed. It combines a convolutional neural network with treebank data to guide the conclusion checking task. The model provides various elements such as grammatical information and structural information data, and its effect is superior to any other single-aspect model. Authors used Static word embedding , and static word embedding can be designed by showing which terms work to express the rating in a better way. We can also provide prefix and suffix data for sensory extraction. Bouazizi et al. [37] built up a framework for a feeling examination of twitter tweets with a specific capacity to identify mocking explanations that improve the effectiveness of conclusion mining. The capacity to identify wry tweets results at a more significant level of exactness. Abbasi et al. [38] recommends a model using multilingual SA to share data over the internet. SA in the English and Arabic dialects is achieved by using style and syntax tools to capture specific element parts. The execution results show the accuracy is above 91%.

Valdivia et al. [39] introduced a model for the presentation of a TripAdvisor SA matching between user emotions and automated emotion detection algorithms. Che et al. [40] propose an ABSA method that automatically compresses emotional sentences through a sentence compression random field. This significantly improves ABSA performance. Bui et al. [41] performed a cancer survivor's network study of temporal connection interpretation of changes in emotions. It is analyzed by presenting a new structure for the cancer survivors machine learning has been used to train sentiment classifiers across positions that are manually labeled with sentiment labels to identify sentiment post such as negative or positive. Wu et al. [42] have implemented a SA approach to decision-making in the online stock platform, which promotes investor decision-making and the understanding of equity companies. An analysis of the results shows that investor sentiment has a greater impact on value stocks. Comparative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

TABLE 1: Sentiment analysis comparison

Article	Authors	Approach	Accuracy
[7]	Karagoz, P., Kama, B., Ozturk, M., Toroslu,	FBAE-SWS,WSBAE	85.0%
[8]	Khan, J., Alam, A., Lee, Y. K.., & Hussain, (2019).	EnSWF	87.62% .
[9]	Abdi, A., Shamsuddin, S. M., Hasan,	RNSA	86.75% .
[11]	Spyromitros, E., Tsoumakas, G., and Vlahavas	k-nearest neighbor (kNN), BRkNN	68.8%
[14]	. Akhtar M., Gupta, D., Ekbal, C .,and Bhattacharyya, P	Conditional Random Field (CRF), Maximum Entropy (ME) and Support Vector Machine (SVM)	77.33 %
[16]	M., Deokar, A. V., Janze, C	Term frequency – Inverse document frequency (TF-IDF) matrix	80.80%
[18]	Madjarov, G., Džeroski, S., Gjorgjevikj, D., & Džeroski, S and	Predictive Clustering Trees (PCT) with Random forests (RF), and Hierarchy of multi-label classifiers (HOMER)	NA
[21]	Kontopoulos E., Dergiades T,. Berberidis C	Ontology based approach	NA
[6]	Wei W., Gulla JA HL	SOT (sentiment ontology tree) approach, Hierarchical Learning,	NA
[44]	Mouthami K., Nirmala D., & Murali B	Fuzzy context classification algorithm that includes parts of speech tag.	NA
[34]	Ouyang X., Pan Z., Hua C. L., & Liu L	For the evaluation of sentences with word2Vec a 7-level CNN model has been used	45.5%
[35]	Stuti Jindal and Sanjay Singh	A predictive system of image sentiments is generated with a pre-trained CNN based on large scaling data and object recognition with a pre-trained CNN based on large scaling data	53.5%
[36]	Yang T., Li Y., Pan Q.,& Lantian G	CNN using tree bank information	Binary classification accuracy 94.7% accuracy), Fine grained Accuracy 49.9%
[33]	Stojanovski D., Strezoski G., Madjarov I., &Dimitrovski I	CNN having multiples filters, two fully integrated layers as well as a softmax layer with different window sizes.	F1 score 64.85%

studies with direct evaluation approaches show the reliability of the method. In Table 2 shows a comparison of the latest technologies.

III. PROBLEM STATEMENT

ABSA aims to predict the polarity of a specific document for a particular aspect of an entity. The neural network architecture successfully predicted the overall polarity of the sentence, however, SA of certain aspects is an open question [14]. Although typical SA focus on predicting positive and negative polarities of a particular sentence. This task works when the specified text has only one aspect and polarity. A more common and complex task is to predict the aspects mentioned in a sentence and the emotions associated with each. In ABSA, this is similar to aspect and document, where the relationship between each word in the document is compared to the aspect vector. The relationship between aspect and language lacks the ability to learn expressions [18]. Most existing opinion mining methods are based on text-level analysis and can only detect well-expressed opinions. The goal of ABSA is to identify aspects of an entity and sentiments stated on each aspect. Extracting the aspect terms and presenting opinions from user-generated content is one of the most important tasks in ABSA [21].

IV. PROPOSED MODEL

The steps of our proposed work are data collection, preprocessing, semantic feature extraction, word2vec representation, and CNN implementation stages, as shown in Fig. 2.

A. DATA COLLECTION

Web scraping technique is used for review collection; hotel reviews collected from “https://webhose.io/ ”, automobiles reviews are collected from “https://www.cvedia.com/” while movie reviews are fetched from ” https://seedmelab.org/”. The raw view of dataset is shown in Fig. 3 The reviews are splinted into positive and negative categories. This simplifies the task and makes it easier to distinguish various aspects of positive and negative views. In the proposed model, three types of data set (hotel, automobiles, and movies) are taken using web scraping and then preprocessing is performed to get valuable data from the data set. Data is provided to the classifier, semantic features are obtained from specific domains, and the total score is determined based on the initial features. Word2vec is designed for corpora processed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

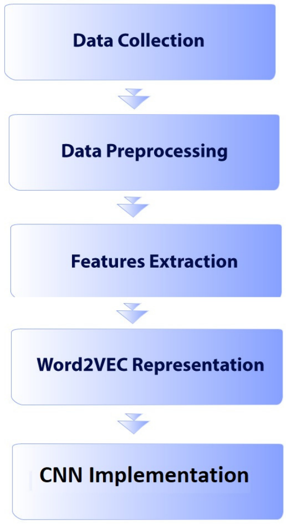

FIGURE 2: Phases of aspect based sentiment analysis

through unsupervised neural networks. Conclusively, a vector type of the preowned corpus is trained by the CNN alongside GA.

B. PREPROCESSING

In this step, raw data is transformed into structured information data. Empty rows, empty cells may present in the data collected through scrappy, panda library is used to clean the data and keep the useful data only as shown in Fig. 4. After analyzing the reviews manually, the provided reviews are used for labels during implementation of VADER. VADER is an analysis tool specifically adapted to extract emotions expressed in text. VADER offers a set of lexical features (e.g. words) that are generally classified as positive or negative subjective to their semantic orientation. VADER was found to be quite effective when it comes to social media texts, movies and product reviews. VADER is not bound only to calculate the positive and negative score, however, it also tells us about the intensity, i.e., how positive or negative a sentiment is. It works remarkably well on text type of social media, but is easily generalized to several domains. It does not include training data, but it is designed on a generalizable, high-valence, human created standard lexicon. It is reasonably fast that can be used online for streaming data and do not suffer seriously from a speed-performance compromise.

V. PROPOSED TECHNIQUE

The data used in our proposed model is collected through web scrapping using python pursed by its preprocessing. Thereafter, features are extracted from data and supplied to the classifier models. The semantics features are separated from the specified domain and total scored is determined based on initial features extracted using VADER. After that Word2vec is design to process the given corpus by utilizing unsupervised neural network and at the end vector form of processed corpus is used to train through CNN classifier. The classification is carried out and CNN hyperparameters were tuned using GA to achieve better model simplification. The proposed model for the sentiment analysis model is shown in the Fig. 1

A. FEATURE EXTRACTION

Words are converted into feature vectors in the first layer of the network to extract the semantics and morphological information about words V words represent word vocabulary $\mathrm{V}^{\mathrm{char}}$ represent character vocabulary. Given sentence contain n words (w1, w2, w3 ....wn) And changing Wn (every word) into vector, according to Eq. 1.

$$\mathrm{Vn} = \left\{\mathrm{r}^{word}, \mathrm{r}^{wchar}\right\} \tag{1}$$

While r^{word} words is word level embedding and r^{wchar} is character level embedding. To capture semantic and syntactic information word-level embedding is intended, while character-level embedding is intended to capture shape and morphological information.

B. WORD LEVEL EMBEDDING

Word level embedding is encrypted in an embedding matrix $\mathrm{W}^{word}\varepsilon\mathrm{R}^{dword} * \mathrm{V}^{word}$ by using vectors of column. The word level embedding coincides to individual column Wi^{word} and R^{dword} through which i^{th} vocabulary word is produced. Word W is word level embedded by the product of matrix vectors as shown in Eq. 2.

$$\mathrm{R}^{word} = \mathrm{V}^{\mathrm{w}}\mathrm{W}^{word} \tag{2}$$

Where V^{w} represent the vector size of |v-word|. At index W have 1 value 0 at o their positions W^{word} matrix is a parameter for learning and dword is embedding word level size. In this paper, word level embedding is performed by word2vec.

C. WORD2VEC

Word2vec is part of open source tools that is provided by Google under the Apache License 2.0 in 2013. This is a very effective tool that can extract the features of a particular domain without human intervention. In addition, it works well whether for too small text size or a single word. By providing a huge corpus context and using word2vec, you can create words with the right sense and run faster on large datasets.

Word meaning is the ultimate perspective of deep learning, and using word2vec to categorize larger entities can fully satisfy the meaning of words. In the proposed method, the dataset is trained on vectors. This model contains 3 million words mixed with 3 million words and phrases. By running such a large corpus, it is possible to obtain accurate relationships. Fig. 5 shows a similar relationship between words using the word2vec transform and relationship between dissimilar words in the dataset. The width of edge indicates similarity, and nodes represent words. Words with the same emotion label have the same vector, so you can easily specify word similarity.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

Hotel_Address	Review_Date	Hotel_Name	Reviewer_Nationality	Review_To	Total_Num	Review	Review_To	Total_Num	Reviewer_S	Tags	days_since_
s Gravesandes	8/3/2017	Hotel Arena	Russia	397	1403	Only the park outside of the hotel was beautiful	11	7	2.9	[' Leisure tr	0 days
s Gravesandes	8/3/2017	Hotel Arena	Ireland	0	1403	No real complaints the hotel was great great locatic	105	7	7.5	[' Leisure tr	0 days
s Gravesandes	7/31/2017	Hotel Arena	Australia	42	1403	Location was good and staff were ok It is cute hotel	21	9	7.1	[' Leisure tr	3 days
s Gravesandes	7/31/2017	Hotel Arena	United Kingdom	210	1403	Great location in nice surroundings the bar and rest	26	1	3.8	[' Leisure tr	3 days
s Gravesandes	7/24/2017	Hotel Arena	New Zealand	140	1403	Amazing location and building Romantic setting	8	3	6.7	[' Leisure tr	10 days
s Gravesandes	7/24/2017	Hotel Arena	Poland	17	1403	Good restaurant with modern design great chill out	20	1	6.7	[' Leisure tr	10 days
s Gravesandes	7/17/2017	Hotel Arena	United Kingdom	33	1403	The room is spacious and bright The hotel is located	18	6	4.6	[' Leisure tr	17 days
s Gravesandes	7/17/2017	Hotel Arena	United Kingdom	11	1403	Good location Set in a lovely park friendly staff Fooc	19	1	10	[' Leisure tr	17 days
s Gravesandes	7/9/2017	Hotel Arena	Belgium	34	1403	No Positive	0	3	6.5	[' Leisure tr	25 days
s Gravesandes	7/8/2017	Hotel Arena	Norway	15	1403	The room was big enough and the bed is good The b	50	1	7.9	[' Leisure tr	26 days
s Gravesandes	7/7/2017	Hotel Arena	United Kingdom	5	1403	Rooms were stunningly decorated and really spaciou	101	2	10	[' Leisure tr	27 days
s Gravesandes	7/6/2017	Hotel Arena	France	75	1403	Style location rooms	4	12	5.8	[' Business	28 days
s Gravesandes	7/6/2017	Hotel Arena	United Kingdom	28	1403	Comfy bed good location	6	7	4.6	[' Leisure tr	28 days
s Gravesandes	7/4/2017	Hotel Arena	Italy	0	1403	This hotel is being renovated with great care and wi	59	6	9.2	[' Business	30 days

FIGURE 3: Unprocessed reviews

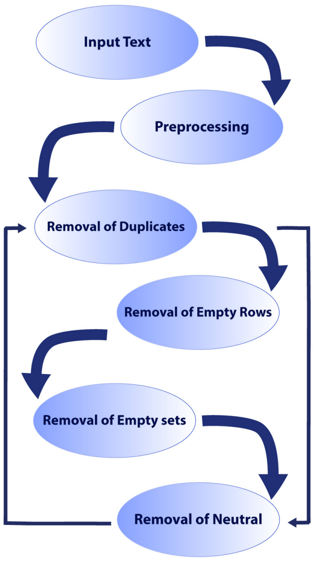

FIGURE 4: Data preprocessing

D. CHARACTER LEVEL EMBEDDING

All words employed important methods of extracting shape and morphological information from the words and thought that these words selected important features of the classification stage. For example, sentiment analysis of Twitter data uses various hashtags to display important information, such as # ihateit, and various adverb information, ending with a suffix, such as angry. Convolution generates local blur around the words around each character and connects them for character-level embedding using the max operation. The specified word is a composition of m characters.

$\{k_1, \quad k_2, \ldots, k_m\}$ every character km is converted in to an embedded character $V_m^{|char|}$, according to Eq. 3. Embedding metrics have embedded character encoded in column vectors. $W^{char} \varepsilon \left|v^{char}\right|$. Each character k

$$\mathrm{R}^{char} = \mathrm{v}^{\mathrm{k}} \mathrm{W}^{char} \tag{3}$$

V^k represent the size of $\left|v^{char}\right|$. v^k have value 1 and index at number 0 at other position. $\left\{r_1^{char}, r_2^{char} \ldots\ldots\ldots\ldots r_n^{char}\right\}$ is convolutional layer input. The convolutional apply matrix vector operation on every character k^{char} from the list of characters. $\mathrm{x}^{\mathrm{m}} \varepsilon \mathrm{r}^{char} \mathrm{R}^{dchar}$ as in Eq. 4. That is the embedding character contain its left neighbor $(r^{char} - 1)/2$ and right neighbors $(r^{char} - 1)/2$

$$X^n = \left(k^{char}_{n-(rchar-1)/2}\right) /2 \cdots k^{char}_{n+(rchar-1)/2}\Big)^{\top} \tag{4}$$

The vector k^{char} have J^{th} element calculated through convolutional layer in the CNN that is the w which is the embedding character level, according to Eq. 5.

$$\left[\mathrm{K}^{wchar}\right]_{\mathrm{j}} = \mathrm{max}_{1<\mathrm{n}<\mathrm{N}} \left[\mathrm{W}^0 \mathrm{X}_{\mathrm{n}} + \mathrm{a}^0\right]_{\mathrm{j}} \tag{5}$$

latterly, local features are extracted utilization same matrix around every window character in the given word.

E. ANALYSIS AT SENTENCE LEVEL

A sentence y contains m words $\left\{w^1, w^2, \ldots w^n\right\}$ which is then changed over to the word level joints and embedding level characters $\left\{v^1, v^2, \ldots v^n\right\}$ Next stage consist of representation of sentence level extraction while extracting features at sentence level we may have following issues.

1) Distinct size of sentence.
2) The sentence may hold significant information = at any position.

These problems are solved by using CNN to calculate the remaining feature vectors of wide sentences. CNN always works along lines when used for character-level feature extraction. Create local features for each word, use the max operation to create a fixed-size feature vector for the sentence, and then link. According to Eq. 6,Matrix vector operation performed on each word in the sequence $\left\{u^1, u^2, \ldots u^n\right\}$

$$\mathrm{X_n} = \{\mathrm{u_c} \ldots \mathrm{u_d}\} \tag{6}$$

where the $c = n - \frac{u^{nod}-1}{2}, d = n + \frac{u^{nod}-1}{2}$. convolutional layer computed the j th element whom vector $r^{st} \in R^{c^{1/4}}$ as in Eq. 7.

$$\left[r^{st}\right]_j = 1 < m < M \max \left[w^1_{x_m} + a^1\right]_j \tag{7}$$

$where \quad w^1 \in R^{c^1_u} \times \left(d^{\text{word}} + c^0_u\right) k^{word}$. At last r^{st}_x vector that contain feature vector global comparative with x sentence is handled through two layers of neural network. To

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

(a) Hotel features.

(b) Movie features.

FIGURE 5: Extracted features from hotel and movie dataset

extricate one or more representation level and every sentimental level t ε T score is then calculated, according to Eq. 8.

$$S(x) = w^3 h\left(w^2 r_x^{st} + a^2\right) a^3 \tag{8}$$

Where learning parameter is given by w$^2 \varepsilon R^{hu*c1u}$ and vectors are given by w$^3 \varepsilon R^{|T|*Hu}$

F. NEGATIVE TRAINING

The system is trained over a training set A, to limit the negative probability. A sentence is passed to the system contain parameter set θ and each sentiment label is computing $\tau \varepsilon$ T. These scores are converted into giving the label with conditional probability distribution plus network parameter set θ and then softmax operation is applied over $\tau \varepsilon$ T score, as in Eq. 9.

$$P(\tau|x,\theta) = \frac{e^{S_\theta(x)_i}}{\sum_{\forall i \in T} e^{S_\theta(x)_i}} \tag{9}$$

The stochastic gradient descent is applied to limit the negative log gradient as in Eq. 10.

$$\theta \longrightarrow \sum_{(u,v)\in A} -\log P(v|x,\theta) \tag{10}$$

(U, v) is the sentence compared to corpus training, and V suppresses the corresponding label.

G. CONVOLUTIONAL NEURAL NETWORK

The structure of CNN can be seen in Fig. 6, this consists of the separate layer stack which converts the volume of the input into target output by differential function. The constituent layers of CNN are

1) Convolution layer
2) Max-policy layer
3) Rely layer
4) Back propagation layer

The intention is to use sufficient filters (in this case, 128) to catch enough features in a particular sentence. In classification of images, different filters integrate different attributes for example edges, color density at various spots, transformation of colors etc. The problem of text classification stretches the similar idea to catch features for example "like" mean positive rather than similarity, "very much" communicate the

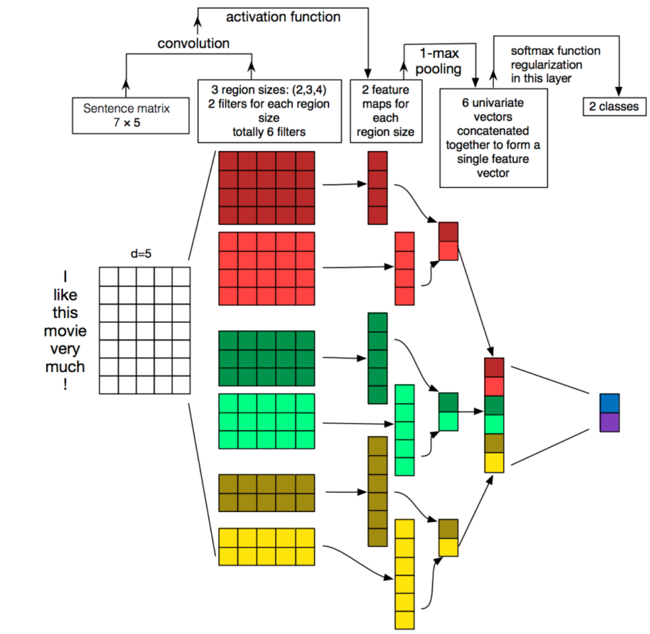

FIGURE 6: Convolutional neural network

degree of features utilizing filter with size 2. The fundamental aim is to provide sufficient range of features to grab all probable descriptor of the text. Maxpool shall have the highest output value of the vector on applying the filter. It picks a strongest expression aspect in the extracted function from the output and nothing to do with the length of the word. Every sample is represented as n*1 where n represents the length of dimension filter. This filter is practice as a drafting window, for instants 3*1 filter on a sentence. I like the car very much! would yield (I like this, this car very, car very much, very much!) sentence splits up to equal length ahead of embedding hence all the filter does not result in the identical dimensions outcome. The area size (2, 3, 4) is identical to 2, 3, 4 –G word and the first filter in this trigram will give different weight esteems to different words in trigrams. It suggest that higher weight are allotted to the first index (0-index) and lower to the second hence 128 filter will allocate respective weight that will be trained to optional weight value after sometimes for precise forecasting. For a particular sentence, a sentimental label is formed by scoring a sentence, by evaluating the order of words in a sentence as input and by passing through layers that extricate its feature with a high degree of complexity. Extraction of the features can be executed on lexical and character level. The distinctiveness of network design is to incorporate two convolutionary layers that allow to handle any size of phrases and words.

H. GENETIC ALGORITHM

GA is a technique to solve problems of optimization centered on natural selection. GA is quite useful to incorporate when you don't know the best ranges and dependencies of different CNN parameters. As CNN defines a hyperplane between the data as shown in Fig. 6. This technique reflects the procedure of natural selection, where the best regenerative individuals are chosen to create the next generation of offspring. The natural selection process initiates with selecting the best individual from the population. The offspring adopted the characteristics from their parents and delivered to next generation. If the parent is in good shape then there are more chances that offspring are better and can survive more than their parent. This process is continually repeated, eventually finding the optimal generation for the individual. The overall steps are shown in Fig. 7. This concept can be used to find an

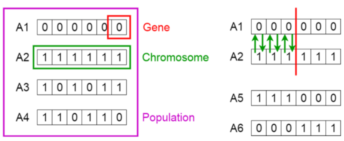

FIGURE 7: Genetic algorithm steps

appropriate solution for problem. GA selects a best solution from the set of solutions for specified problem. The genetic algorithm considers five stages.

1) Initial population
2) Fitness function
3) Selection
4) Crossover
5) Mutation

Initial Population: This process is initiated with the group of individuals called population. Everyone is the solution of the problem that you need to eliminate. Human characteristics are consist of parameters (variables) called genes. Gene interface with a string to frame a chromosome (solution). In genetic algorithms, a person's gene set is represented by a string and a letter as shown in Fig. 8. Binary values (strings of 1s and 0s) are typically used. It might be said that we encode a gene on a chromosome.

Fitness Function: Fitness functions determine a person's health (the capability of one to compete with others). It provides a fitness score to everyone. The probability is calculated by fitness score to select person for breeding. The raw fitness

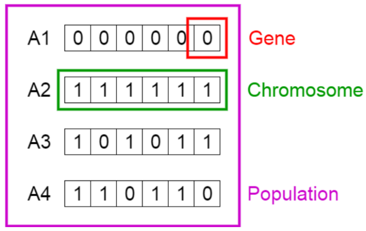

FIGURE 8: Initial population

score of objective function is converted into a value within selected range as in Eq. 11 is happened by fitness scaling.

$$F = \frac{G^{leam} + G^{validation}}{2} \tag{11}$$

Selection: This phase is used to select the most suitable entities and let them forward their genes to the next generation. It selects two sets of individuals (parents) based on their health score. The higher the physique, the more opportunities for breeding. This selection uses the scaled fitness values to select the next generation of parents, and typically applies higher selection probabilities to individuals with higher scaled values, as in Eq. 12.

$$\mathrm{g(x)} = \frac{\mathrm{f(x)}}{\sum_{\mathrm{x} \in \mathrm{p}} \mathrm{k}} \tag{12}$$

the number of copies of each feature ${}^t\mathrm{x}'$ in $\mathrm{p}^{(\mathrm{k})}$, that is chosen number of copies per function $t_{\mathrm{X}'}$ in $\mathrm{p}^{(\mathrm{k})}$, It's selected for $\mathrm{p}_\mathrm{n}^{(\mathrm{k})}$, is offered by the fractional part of the $\mathrm{e(x)}$ If the number of features selected in such a way is smaller than m (the usual case), then choose the leftover features for $\mathrm{p}_\mathrm{n}^{(\mathrm{k})}$ using fractional parts of $\mathrm{e(x)}$ from the highest values down. In general, this technique is designed to remove less adaptive features and repeat more adaptive features.

$$CR_{(y)} = n_y^+ / m_y \tag{13}$$

In Eq. 13, with n_y^+ denoting the number of patterns of class y that are classified as member of class y (correct) m_y denoting the number of all patterns belonging to class y.

Crossover: Crossover is very important stage of genetic algorithm. A random exchange point is selected from the gene for each parental pair that is mated. The crossover helps the genetic algorithm to derive and recombine the most suitable features from different individuals to a potentially superior child, as shown in Eq. 14.

$$child = parent2 + R_{cros} \times (parent1 - parent2) \tag{14}$$

For example, consider an crossover point at 3, as shown in Fig. 9. Offspring are created by exchanging genes between parents until the crossover point is reached as in Fig. 10. New offspring are added to the population as shown in Fig. 11.

Mutation: In the formation of some new offspring, some of those genes can be mutated with a low probability. It means that few bits in the bit string can be flipped as shown in Fig. 12. In addition to mating offspring, genetic algorithms

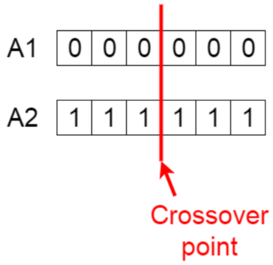

FIGURE 9: Selection point for crossover

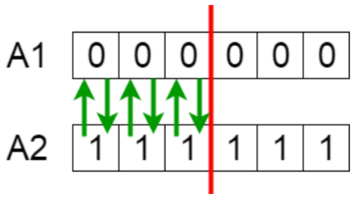

FIGURE 10: Exchange of genes between parents

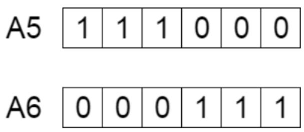

FIGURE 11: New offspring

also generate mutant offspring by randomly mutating a single parent of the current generation. Because mutations contribute to population diversity, the algorithm is more likely to generate fitness values that are better suited to individuals.

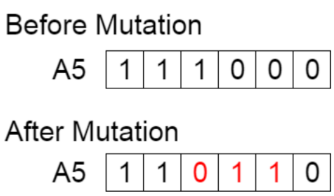

FIGURE 12: Mutation

Mutations occur to maintain diversity within a population and prevent premature fusion.

Termination: If the population converges, the algorithm terminates (no descendants are significantly different from the previous generation). GA are said to provide a set of solutions to our problem.

VI. EXPERIMENTAL DESIGN

Implementation of the proposed framework is done on the system having specifications; Intel Core i3-4005 CPU @1.70GHZ and 4GB RAM.The simulation runs on software called Anaconda Spyder and a python language environment. Software packages used are Keras v2.3.1, Tensorflow v2.0, pandas v1.0.3 And numpy v1. The various indicators for model validation are Precision, Accuracy, Recall, and F1-Score as defined in Table 2. For the validation and generalization of the proposed model, 20 iterations are considered in the CNN classification process. Besides, various models are implemented and outcomes are compared to validate the strength of the recommended model.

1) True positive (TP): the values of class that is correctly categorized.
2) True negative (TN): correctly categorized as uninteresting.
3) False negative (FN): misclassified as a category of no interest.
4) False positive (FP): class of interest is wrongly classified.

The formulas of performance evaluation metrics are shown in Table 2.

VII. RESULTS AND DISCUSSION

It is observed in experimental observations that the extraction of semantic features produces exceptionally refined data. Information gathered by examining the particular domain reviews decreases the false negative and false positive rates that produce better accuracy. Table 6 shows the comparison of precision, accuracy, and recall of different classifiers such as SVM, maximum entropy, random forest, stabilized discriminant analysis, decision tree and generalized linear model. CNN has been found to produce significant improvement of 91.6% , 93.4% , 92.2% , 89.23% in precision, accuracy, recall and f1 measure, respectively.

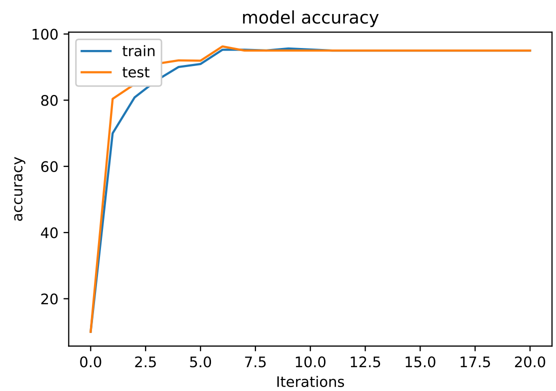

FIGURE 13: Accuracy vs iterations

The suggested approach has been contrasted with the SVM, leading to high accuracy rate and f-score value for positive as well as negative labels as shown in Table 4. The accuracy and loss rate of the suggested method are shown in Fig. 13 and Fig. 14. The graph shows that as the number of training iterations and tests increases, the percentage of accuracy increases, and as the number of iterations decreases, the percentage of loss decreases.

Fig. 15 graphically illustrates the performance of the integrated method on a hotel dataset. The accuracy of ensemble approach for hotel datasets is higher than the comparison techniques. The CNN-GA hybrid model's overall result is

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

TABLE 2: Performance Metrics Formula’s

Metric	Definition
Accuracy	(TruePositve+TrueNegative)/(TruePositve+TrueNegative+FalsePositive+FalseNegative)
Precision	TruePositve/(TruePositve+FalsePostive)
Recall	TruePositve/(TruePositve+FalseNegative)
F1-Score	(precision x 2 x recall) / (precision + recall)

TABLE 3: Detailed results on Automobiles Dataset

Parameters	SVM	DecTree	LDA	LinMod	RF	LogReg	Proposed
Accuracy	85.5	67	65	78.5	83	80	93
95% CI	(0.84, 0.87)	(0.65, 0.68)	(0.64, 0.66)	(0.75, 0.80)	(0.82, 0.85)	(0.78, 0.82)	(0.90, 0.95)
Information rate	0.6	0.62	0.6257	0.635	0.624	0.625	0.629
P0-value	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16	2.20E-16
Kappa	0.7112	0.6828	0.4429	0.637	0.6712	0.7179	0.7321
Mcnemar’s	0.4107	0.002628	0.00992	0.7273	0.0953	0.0013	0.5203
Pos Pred value	82.8	70.6	84.8	74	80.8	85.6	88.4

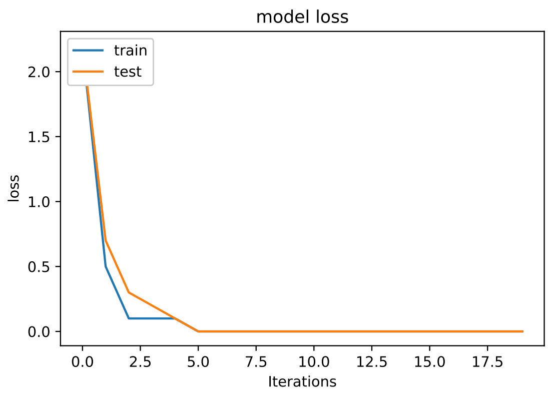

FIGURE 14: Loss vs iterations

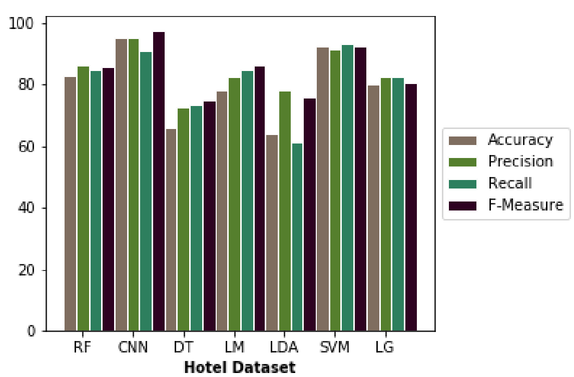

FIGURE 15: Performance on hotel dataset

TABLE 4: Performance Analysis on Hotel Dataset

Models	Precision	Accuracy	Recall	Fmeasure
Proposed	95.5	94.3	91	96.6
DT	72.71	67.0	74.52	75.44
SVM	91.5	92.3	90.2	92.34
LDA	78.53	65.0	62.43	76.75
LM	82.36	78.5	84.85	86.10
RF	86.3	83.0	84.8	85.69
LG	82.36	80.5	82.85	81.10

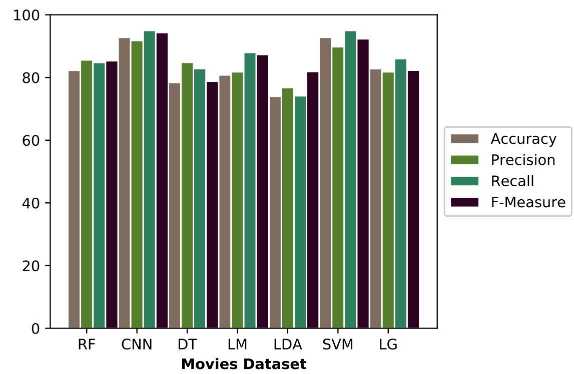

FIGURE 16: Performance on movies dataset

stronger than other state-of-the-art techniques, can be seen in Table. 4.

Table 4 clearly shows the performance of the integrated approach for hotel data. Compared with the prior art, the proposed method has the highest precision, recall, accuracy and f-measure. The performance of integrated method is 4% better than SVM, 23% better than decision trees, 17% better than LDA, and 10% better than RF, and 12% better than LG.

Fig. 17 graphically illustrates the performance of the integrated method on automobiles dataset. Table 6 clearly demonstrates the performance of the suggested method for automobile data. Evaluated with the prior art, the proposed method has the highest precision, recall, accuracy and f-metric. The performance of the integrated method is 6% better than SVM, 21% better than decision trees, 15% better than LDA, and 11% better than RF, and 10% better than LG.

The movie review dataset is used as the standard benchmark dataset for IMDB. In Fig. 16, it can be observed that the outcomes of the integrated method (GA with CNN) are much better than the benchmark algorithm (RF, DT, LM, LDA, SVM, and LG). GA based integrated methods provide superior performance in terms of accuracy, precision, recall, and f measurements. The results are shown graphically in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2020.3011802, IEEE Access

TABLE 5: Performance evaluation on Movies Dataset

Models	Precision	Accuracy	Recall	Fmeasure
Proposed	91.5	95.2	92.3	94.0
DT	75	79	85	82
SVM	89.4	92.5	88.8	90.3
LDA	89	80	75	85
LM	87	86	88	90
RF	90.8	87.5	84.8	88.0
LG	87	88	86	85

Fig. 16 and in tabular in Table 5. In Table 5, from an accuracy perspective, the performance of the integrated method is 2% better than SVM and 6% better than RF. The accuracy of the integrated method is 91.5%, while the accuracy of RF and SVM is 90.8% and 89.4%. In addition, the performance of integrated approach is significantly better to other existing technologies in recall and f-measure.

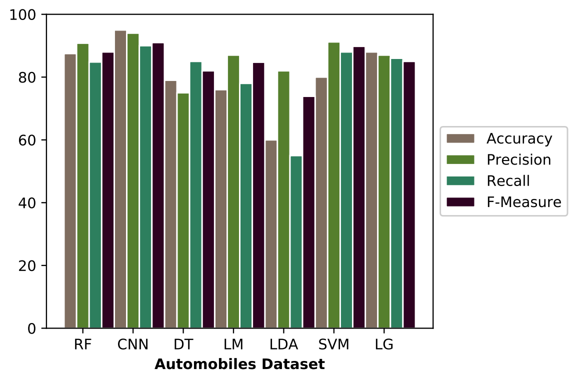

FIGURE 17: Performance on Automobiles Dataset

TABLE 6: Performance Evaluation on Automobiles Dataset

Models	Precision	Accuracy	Recall	Fmeasure
Proposed	91.6%	93.4%	92.2%	89.23%
DT	70%	70.6%	81.20%	77.59%
SVM	85.8%	84.8%	89.9%	87.33%
LDA	76.75%	74.0%	74.15%	81.90%
LM	81.8%	80.8%	88%	87.33%
RF	82.3%	85.6%	84.8%	85.33%
LG	81.8%	82.8%	86%	82.33%

Fig. 17 shows the performance outcomes of the latest technologies and integration methods for automobiles datasets. The proposed method has excellent performance on automobiles dataset in terms of accuracy, precision, recall, and f-measure. The results are described in detail in Table 6. Table 7 shows the performance index values in percentage format, showing that the proposed method performs well with all metrics as compared to SVM. Table 8 shows the detailed results for the different iteration levels and the different iterations for accuracy, loss, progress, and Base Learning Rate. For different parameters, we compare the outcomes of the recommended classifier with all six other classifiers in Table 3 and conclude that the proposed approach outperforms nearly all the provided classifiers for all test parameters. It can therefore be concluded that the recommended approach is 93.0% more accurate than all six comparator classifiers. The result comparison table supports the Ensemble methods given. The comparison results clearly indicate that the suggested approach produces better results in terms of accuracy, precision and f-measure. For certain datasets based on positive and negative ratings based on experimental settings, SVM has the highest accuracy. .

TABLE 7: Comparison of proposed model with SVM model

Parameters	Proposed Model	SVM
Labels	2	2
Data points	200	200
Accuracy	88.4	84.8
Label	NEG	NEG
F-score	89.23% TP: 185 FP: 6 TN: 9 FN: 0	87.69% TP: 168 FP: 4 TN: 23 FN: 5
Label	POS	POS
F-score	78% TP: 185 FP: 6 TN: 9 FN: 0	64% TP: 166 FP: 6 TN: 25 FN: 3

TABLE 8: Performance Evaluation w.r.t iterations on Movie and Hotel Datasets

Iteration	loss	Accuracy	Time Elapsed	Base Learning Rate
0	9.5881	38.91%	1s 1ms	1.00E-04
1	0.4045	80.78%	0s 703us	1.00E-04
2	0.0755	92.34%	0s 586us	1.00E-04
3	0.0833	92.34%	0s 622us	1.00E-04
4	0.0851	91.56%	0s 563us	1.00E-04
5	0.0692	92.19%	0s 547us	1.00E-04
6	0.0702	92.03%	0s 531us	1.00E-04
7	0.0662	92.34%	0s 531us	1.00E-04
8	0.0654	92.03%	0s 531us	1.00E-04
9	0.0657	92.19%	0s 539us	1.00E-04
10	0.0649	92.19%	0s 531us	1.00E-04

VIII. CONCLUSION

This article presents an effective classification method for sentiment analysis using Convolutional neural network and Genetic Algorithm. Semantic features are mined, and then, several models are trained together with proposed CNN based ensembler, SVM, maximum entropy, random forest, stabilized discriminant analysis, decision tree, generalized linear model. Using data collected through CNN execution and analysis, domain-specific reviews reduce false positives and false positives and improve accuracy. we adjust the CNN hyperparameters using a genetic algorithm to get optimal values. The experimental results show that the recommended method outperforms all other recent methods, with 95.5%, 94.3%, 91.1%, and 96.6% for accuracy, precision, recall, and f-measurement, respectively. Future strategies are to integrate parallel computing to speed up computation and explore metaheuristic-related features. A preferred work is to have a web-based ontology framework automation to incorporate sentiment analysis on social sites.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest: Their is no conflict of interest involved in this research.

REFERENCES

[1] Kumar, Ravindra, Husanbir Singh Pannu, and Avleen Kaur Malhi. "Aspect-based sentiment analysis using deep networks and stochastic optimization." Neural Computing and Applications (2019): 1-15

[2] Cambria, Erik, et al. "Sentiment analysis is a big suitcase." IEEE Intelligent Systems 32.6 (2017): 74-80

[3] Al-Smadi, Mohammad, et al. "Human annotated arabic dataset of book reviews for aspect based sentiment analysis." 2015 3rd International Conference on Future Internet of Things and Cloud. IEEE, 2015.

[4] Al-Moslmi T, Omar N, Abdullah S, Albared M (2017) Approaches to cross-domain sentiment analysis: a systematic literature review. IEEE Access 5:16173–16192

[5] Jianqiang Z, Xiaolin G (2017) Comparison research on text preprocessing methods on twitter sentiment analysis. IEEE Access 5:2870–2879.

[6] Wei W, Gulla JA (2010) Sentiment learning on product reviews via sentiment ontology tree. In: Proceedings of the 48th annual meeting of the association for computational linguistics. Association for Computational Linguistics, pp 404–413.

[7] Karagoz, P., Kama, B., Ozturk, M., Toroslu, I. H., & Canturk, D. (2019). A framework for aspect based sentiment analysis on turkish informal texts. Journal of Intelligent Information Systems, 53(3), 431-451.

[8] Khan, J., Alam, A., Hussain, J., & Lee, Y. K. (2019). EnSWF: effective features extraction and selection in conjunction with ensemble learning methods for document sentiment classification. Applied Intelligence, 49(8), 3123-3145.

[9] Abdi, A., Shamsuddin, S. M., Hasan, S., & Piran, J. (2019). Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. Information Processing and Management, 56(4), 1245-1259.

[10] S. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi, "Sentiment analysis based on improved pre-trained word embeddings", Expert Systems with Applications, Vol.117, No.44, pp.139-147, 2019.

[11] Souma, W., Vodenska, I., & Aoyama, H. Enhanced news sentiment analysis using deep learning methods. Journal of Computational Social Science, 1–14 (2019).

[12] Rezaeinia, S. M., Rahmani, R., Ghodsi, A., & Veisi, H. Sentiment analysis based on improved pre-trained word embeddings. Expert Systems with Applications, 117, 139–147 (2019).

[13] Tang, D., & Zhang, M. Deep Learning in Sentiment Analysis. In Deep Learning in Natural Language Processing (pp. 219–253). Springer (2018).

[14] Rani, S., & Kumar, P. Deep Learning Based Sentiment Analysis Using Convolution Neural Network. Arabian Journal for Science and Engineering, 1–10 (2018).

[15] Heikal, M., Torki, M., & El-Makky, N. Sentiment Analysis of Arabic Tweets using Deep Learning. Procedia Computer Science, 142, 114– 122 (2018).

[16] Li, L., Goh, T. -T., & Jin, D. How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis. Neural Computing and Applications, 1–29 (2018).

[17] Alharbi, A. S. M., & de Doncker, E. Twitter sentiment analysis with adeep neural network: An enhanced approach using user behavioral information. Cognitive Systems Research, 54, 50–61 (2019).

[18] W. Huang, G. Rao, Z. Feng, and Q. Cong, "LSTM with sentence representations for Document-level Sentiment Classification", Neurocomputing, Vol.308, No.45, pp.49-57, 2018.

[19] M. Fikri and R. Sarno, "A Comparative Study of Sentiment Analysis using SVM and SentiWordNet", Indonesian Journal of Electrical Engineering and Computer Science, Vol.13, No.3, pp.902-909, 2019].

[20] Priyantina, R. A., & Sarno, R. (2019). Sentiment Analysis of Hotel Reviews Using Latent Dirichlet Allocation, Semantic Similarity and LSTM. International Journal of Intelligent Engineering and Systems, 12(4), 142-155.

[21] Kontopoulos E, Berberidis C, Dergiades T, Bassiliades N (2013) Ontology-based sentiment analysis of twitter posts. Expert Syst Appl 40(10):4065–4074.

[22] Ciric M, Stanimirovic A, Petrovic N, Stoimenov L (2013) Comparison of different algorithms for sentiment classification. In: 2013 11th international conference on telecommunication in modern satellite, cable and broadcasting services (TELSIKS), vol 2. IEEE, pp 567–570 .

[23] Bahrainian S-A, Dengel A (2013) Sentiment analysis and summarization of twitter data. In: 2013 IEEE 16th international conference on computational science and engineering (CSE). IEEE, pp 227–234.

[24] Freitas LA, Vieira R (2013) Ontology based feature level opinion mining for Portuguese reviews. In: Proceedings of the 22nd international conference on World Wide Web. ACM, pp 367–370.

[25] Zhao Y, Niu K, He Z, Lin J, Wang X (2013) Text sentiment analysis algorithm optimization and platform development in social network. In: 2013 sixth international symposium on computational intelligence and design (ISCID), vol 1. IEEE, pp 410–413.

[26] Bakliwal A, Foster J, van der Puil J, O'Brien R, Tounsi L, Hughes M (2013) Sentiment analysis of political tweets: towards an accurate classifier. Association for Computational Linguistics, pp 49–58.

[27] Sam KM, Chatwin C (2013) Ontology-based sentiment analysis model of customer reviews for electronic products. Int J e-Educ e-Bus e-Manag e-Learn 3(6):477.

[28] Ravi K, Ravi V (2015) A survey on opinion mining and sentiment analysis: tasks, approaches and applications. Knowl Based Syst 89:14–46.

[29] Yadollahi A, Shahraki AG, Zaiane OR (2017) Current state of text sentiment analysis from opinion to emotion mining. ACM Comput Surv: CSUR 50(2):25

[30] Clavel C, Callejas Z (2016) Sentiment analysis: from opinion mining to human-agent interaction. IEEE Trans Affect Comput 7(1):74–93.

[31] Dragoni M, Poria S, Cambria E (2018) Ontosenticnet: a commonsense ontology for sentiment analysis. IEEE Intell Syst 33(3):77–85.

[32] Kim Y (2014) Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882

[33] Stojanovski D, Strezoski G, Madjarov G, Dimitrovski I (2015) Twitter sentiment analysis using deep convolutional neural network. In: International conference on hybrid artificial intelligence systems. Springer, pp 726–737.

[34] Ouyang X, Zhou P, Li CH, Liu L (2015) Sentiment analysis using convolutional neural network. In: 2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing (CIT/ IUCC/DASC/PICOM). IEEE, pp 2359–2364.

[35] Jindal S, Singh S (2015) Image sentiment analysis using deep convolutional neural networks with domain specific fine tuning. In: 2015 international conference on information processing (ICIP). IEEE, pp 447–451.

[36]) Yang T, Li Y, Pan Q, Guo L (2016) Tb-CNN: joint tree-bank information for sentiment analysis using CNN. In: 2016 35th Chinese control conference (CCC). IEEE, pp 7042–7044.

[37] Bouazizi M, Ohtsuki T (2016) Sentiment analysis: from binary to multiclass classification: a pattern-based approach for multiclass sentiment analysis in twitter. In: 2016 IEEE international conference on communications (ICC). IEEE, pp 1–6.

[38] Abbasi A, Chen H, Salem A (2008) Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. ACM Trans Inf Syst: TOIS 26(3):12

[39] Valdivia A, Luzo´n MV, Herrera F (2017) Sentiment analysis in tripadvisor. IEEE Intell Syst 32(4):72–77

[40] Che W, Zhao Y, Guo H, Su Z, Liu T (2015) Sentence compression for aspect-based sentiment analysis. IEEE/ACM Trans Audio Speech Lang Process 23(12):2111–2124.

[41] Bui N, Yen J, Honavar V (2016) Temporal causality analysis of sentiment change in a cancer survivor network. IEEE Trans Comput Soc Syst 3(2):75–87.

[42] Wu DD, Zheng L, Olson DL (2014) A decision support approach for online stock forum sentiment analysis. IEEE Trans Syst Man Cybern Syst 44(8):1077–1087.

[43] Rani S, Kumar P (2017) A sentiment analysis system to improve teaching and learning. Computer 50(5):36–43.

[44] Mouthami K, Devi K.N, Bhaskaran VM (2013) Sentiment analysis and classification based on textual reviews. In: 2013 international conference on information communication and embedded systems (ICICES). IEEE, pp 271–276

[45] Zhang Y, Wallace B (2015) A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820

ADNAN ISHAQ is student of MS Software Engineering at COMSATS University Islamabad, Islamabad. Currently, he is working on NLP and aspect base sentiment analysis.

SOHAIL ASGHAR is working as Professor & Chairman of Computer Science at COMSATS University Islamabad. In 2011 he joined as a Director of the University Institute of Information Technology, PMAS-Arid Agriculture University, Rawalpindi. Previously, In 1994, he graduated with honors in Computer Science from the University of Wales, United Kingdom. He then recieved his PhD from Faculty of Information Technology at Monash University, Melbourne Australia in 2006. Professor Sohail has taught and researched in Data Mining (including structural Learning, Classification, and Privacy Presevation in Data Ming, Text and Web Mining), Big Data Analytics, Data Science and Information Technology areas, and he has published extensively (More than 150 publications) in international journals as well as conference proceedings. He has also consulted widely on information Technology matters, especially in the framework of Data Mining and Data Science. In 2004 he acquired the Australian Postgraduate Award for Industry. Professor Sohail is a member of, the Australian Computer Society (ACS), and IEEE and also Higher Education Commission Approved Supervisor. Professor Sohail is in the Editorial Team of well reputed Scientific Journals.He has also served as Program Committee member of numerous International Conferences and regularly speaks at international conferences, seminars and workshops.

SAIRA ANDLEEB GILLANI is Senior Assistant Professor at Department of Computer Science, Bahria University Karachi Campus. She has received her PhD Degree from Corvinus University of Budapest, Hungary in 2016. She is working in NLP, text mining and data mining domain.