

Sentiment Analysis System using Hybrid Word Embeddings with Convolutional Recurrent Neural Network

Fahd Alotaibi

Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia
fsalotaibi@kau.edu.sa

Vishal Gupta

Department of Computer Science and Engineering, University Institute of Engineering and Technology, Panjab University Chandigarh, India
vishal@pu.ac.in

Abstract: *There have been wide ranges of innovations in sentiment analysis in recent past, with most effective ones involving use of various word embeddings methods for analysis of sentiments. GloVe and Word2Vec are acclaimed to be two most frequently used. A common problem with simple pre-trained embedding methods is that these ignore information related to sentiments of input texts and further depend on large text corpus for training purpose and generation of relevant vectors which is hindrance to researches involving smaller sized corpuses. The aim of proposed study is to propose a novel methodology for sentiment analysis that uses hybrid embeddings with a target to enhance features of available pre-trained embedding. Proposed hybrid embeddings use Part of Speech (POS) tagging and word2position vector over fastText with varied assortments of attached vectors to the pre-trained embedding vectors. The resultant form of hybrid embeddings is fed to our ensemble network-Convolutional Recurrent Neural Network (CRNN). The methodology has been tested for accuracy via different Ensemble models of deep learning and standard sentiment dataset with accuracy value of 90.21 using Movie Review (MVR) Dataset V2. Results show that proposed methodology is effective for sentiment analysis and is capable of incorporating even more linguistic knowledge-based techniques to further improve results of sentiment analysis.*

Keywords: *Analysis of sentiments, convolutional neural networks, part of speech tagging, natural language processing, word2Vec, GloVe, fastText, hybrid embedding, recurrent neural networks.*

*Received January 13, 2021; accepted October 14, 2021
<https://doi.org/10.34028/iajit/19/3/6>*

1. Introduction

Sentiment analysis deals with classification of text or online reviews into classes of positive, negatives or neutral. Even though word embedding techniques can be designed to capture sentiment information of the text [3, 7, 20] but they inappropriately map terms with different polarity to close vectors causing improper classification of sentiments [24].

The goal of study is to propose a hybrid word embedding with Convolutional Recurrent Neural Network (CRNN) for increasing the power and accuracy of pre-trained word embeddings (word2Vec/fastText/GloVe) for the purpose of sentiment analysis. Analysing results of our methodology delineates that method is successful in improving efficiency of word embeddings and provides better classification results. We have made use of Internet Movie Database (IMDB) movie review dataset [16] for testing the proposed system.

This paper strives to propose a novel technique for sentiment analysis by making use of pre-trained embeddings of words to generate hybrid embeddings; enhanced by adding commonly used features. The resultant hybrid embeddings is fed into proposed

ensemble network-CRNN. Proposed method makes use of Part of Speech (POS) tagging and word position based features with fastText with varied assortments of attached vectors to the pre-trained embedding vectors.

Our paper is organized as: section 2 describes related work for sentiment analysis; section 3 describes details of proposed system for sentiment analysis system using hybrid word embeddings with convolutional recurrent neural network; section 4 describes experiment result and discussions along with their comparison with existing state of art approaches for sentiment analysis.

2. Related Work

Major techniques used for sentiment classification range from the traditional lexicon-based methods to the latest deep learning-based methods [13, 14, 21, 25]. Lexicon based techniques are simple and effective because of their dependence on dictionary of positive and negative polarity values assigned to terms [25]. However, lexicon-based method sometimes suffers from low accuracy [7] depending on human effort for annotated data of sentiment scores. Machine learning based Natural language Processing (NLP) approaches

have been outperforming the traditional approaches for text classification. Upon analysing [19, 25], many authors have concluded that machine learning based approach has better accuracy than lexicon-based approach. Recently, the scope of natural language processing has increased immensely with deep learning techniques in which word vector representations using deep learning is a major breakthrough [3]. Severyn and Moschitti [23] had used Word2Vec algorithm for fifty million tweets and used pre-trained vectors for training the model. However, size of text corpus highly affects the efficiency of fastText [4], GloVe [17] and Word2vec. The standard Continuous Bag Of Words (CBOW) model used in Mikolov *et al.* [15] learns representation of terms with prediction of a term as per its context. For the purpose of learning richer word representations, Facebook [8, 15] released high-quality word vector representations which makes use of combination of known tricks to provide richer vectors. Recently, Lauren *et al.* [11] created word embeddings using a discriminant document embedding which implicitly uses skip-gram. Alqaraleh [2] proposed Sentiment Analysis for Turkish with use of ConvNet and this system has outperformed the existing state of art techniques for Turkish Sentiment Analysis. Arabic text classification was performed by Alghamdi and Assiri [1] using FastText word embeddings. Trend has surfaced recently by use of Convolution Neural Networks (CNNs) for performing Natural Language Processing with fastText embeddings [22]. Fu *et al.* [6] used Word2Vec algorithm for producing word embeddings of English and Chinese dataset of Wikipedia. Ren *et al.* [20] proposed classification technique using a new word representation on Twitter Dataset. Qin *et al.* [18] used pre-trained vectors in convolutional neural networks for data-driven tasks using CNNs.

3. Proposed System

The proposed sentiment analysis system is deep learning CNN and bi-directional LSTM inspired architecture that utilizes new form of hybrid embedding to add more power to fastText embeddings. We now explain the complete system in detail.

3.1. Improved Hybrid Embedding

The character level exploration of fastText makes it one of the most dependable embedding. But experiments suggest that its powers can further be enhanced by appending more feature vectors i.e., additional information to its vectors. We propose to extend fastText by use of Part of Speech Tag embedding and word position to vector embedding. POS tagging plays a fundamental and important step in natural language processing. It provides essential information about a word corresponding to its adjacent words along with the syntactic categorization of words.

POS tagging based vectors have been utilized to identify respective part of speech based tags. Here, each POS tag is converted to a new vector and appended with Word2Vec/fastText/GloVe vectors. To further enhance performance in computational linguistic tasks, position information for words could also be of utmost use. Purpose of Wordposition2vec embedding is to denote relative distances of a term to the two ends of a document. Incorporation of Word position as an embedding provides relative distances from both ends for word under consideration. For example, for a sentence “kid like to play in garden”, the relative distance of the words here, ‘like’ to ‘kid’ and ‘garden’ are -1 and 4, respectively however it will be different if we measure it from the right end. Outline of the algorithm is given in A1.

A1: Algorithm for producing hybrid embeddings

INPUTS

-PE (Position based Vector): {PE0....PEN}
 -d (dimension of POS tag Vector): {dT0.....dNNS}
 -W2Vec: Pre-Trained fastText Embedding

OUTPUT

-HE: Hybrid Embedding

1. Extract the text corpus and divide into sentence clusters as C1, C2 ...Cn.
2. Process for unique words, stop words and other entities such as numbers.
3. for each grouping Ci do
4. PEi = GeneratePositionVec(Ci)
5. dVec= <Di,Din>
6. di = GeneratePOSTag(Ci)
7. end for
8. for each Wi in Ci do
9. if Wi exist in W2Vec
10. WordVec= ExtractFromPreTrained(Wi)
11. else
12. WordVec= RandomVec(Wi)
13. ConvAppendTo(W2Vec)
14. end for
15. for each SentVec in Ci
16. Append(PEi, di, WV)
17. Normalize SentVec
18. ADD to HE
19. end for
20. Return HE to Model
21. Employ the HE layer in the ensemble model Architecture below

The algorithm takes as input pre-trained fastText embeddings (W2Vec), position-based vector and dimensions of our POS vector(d) and outputs the required HE i.e., hybrid embedding. The values from each of these computations are appended into the score value received from pre-trained embeddings. The enlarged vector is transformed to 300 dimensions to serve as hybrid embedding layer. The algorithm pre-processes the input text corpus against obtained embeddings to attain processed corpora. The vectors obtained will be passed through the proposed deep learning based network consisting of a convolutional layer, a dense layer and followed by a layer of

Bidirectional LSTM layer and an output layer. The next subsection details the complete approach.

3.2. Architecture Employed

The proposed system combines CNN and Recurrent Network (RNN) and is abbreviated as CRNN. RNN is capable of learning context and temporal features or otherwise sequential data with long term dependencies being maintained while CNN retains all the higher order potential features. Here, Convolutional layers are used for extracting features from multiple words horizontally that allows the network to extract even higher-level styles in learning. We employ Convolutional layer before the RNNs as RNN can input and output arbitrary lengths of data whereas CNNs use fixed length inputs and generate fixed length outputs which are inputted from the hybrid pre-trained embedding layer to first extract higher order semantic relations and then extract and retain long term dependencies. We experimented with various forms of RNNs such as gated recurrent unit, Long Short Term Memory (LSTM) unit and Bi-Directional LSTM and concluded that combination of convolution with Bi-directional recurrent networks resulted in best results Further, regularizes like addition of dropout was also tested for making sure that the model doesn't over fit. Figure 1 shows the proposed architecture employed in the study. The architecture comprises of Conv-Pool-Dense-BLSTM layers in sequence. The hybrid embedding obtained acts as input to the first 1D convolution layer comprising of kernel size of 5. Several other filter sizes were also tested and it was found that the size variably depends upon the size of the dataset and batch size. The vectors so obtained were further fed to Max pool layer and a dense layer. The dense layer has helped to highlight prominent features that could be utilized to extract temporal information. The vectors are further fed to a Bi-directional LSTM of size 10 that helps in understanding the context of the review and presence of a sigmoid layer expresses the final opinion of the sentiment. The methodology was tested on IMDB dataset and results have been recorded with different versions of the dataset available. Section 4 discusses the obtained results in details.

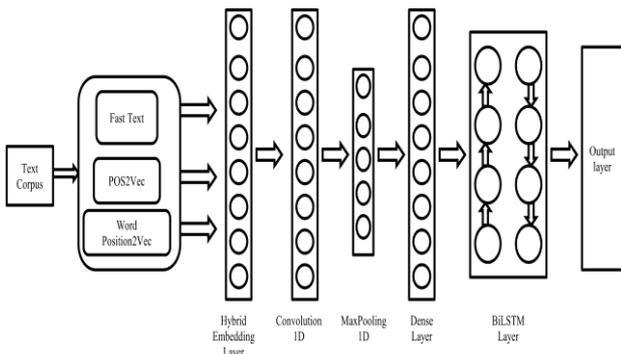


Figure 1. Proposed architecture.

4. Results and Discussions

For judging the efficiency of proposed embeddings, we decided to experiment with several variations of convolutional networks to find the optimal combination to be used with RNN. Table 1 summarizes architecture of the models used for experimented with their abbreviations in the last column. All these models were used for experimentation of different embeddings and finally paved the way for use of fastText and Model 7 as the final deployed model for sentiment analysis.

Table 1. Various naming conventions used in paper.

Architecture	Name
Hybrid Embedding with 2 Layer CNN	Model 1
Hybrid Embedding with 2 Layer CNN and 1 dense Layer	Model 2
Hybrid Embedding with Ensemble model (CNN+LSTM)	Model 3
Hybrid Embedding with Ensemble model (CNN+D+LSTM)	Model 4
Hybrid Embedding with Ensemble model (CNN+D+GRU)	Model 5
Hybrid Embedding with Ensemble model (CNN + BiLSTM)	Model 6
Hybrid Embedding with Ensemble model CRNN (CNN+ D + BiLSTM)	Model 7

The experiments concluded that the mixed architecture of CRNN with bidirectional LSTM delivered best results with hybrid embedding made using pre-trained hybrid embeddings. IMDB Movie Review Dataset was used for evaluation and testing of proposed methodology. The dataset is available in various versions and we had used both versions. Version V1(old variant) with 10,662 review out of which 5331 and positive and 5331 are negative as used in study of [20] because of suitability for comparison. There is another recent version of dataset available (V2) named as Large MVR Dataset consisting of 50,000 movie reviews.

In order to quantify the results, we evaluated efficiency of various models for Accuracy, Precision, Recall and F1. Table 2 proposes comparative results of effects of hybrid embeddings on various architectures employing fastText embeddings as initial feature vector. All architectures i.e., Model 1 through model 7 as illustrated in Tab1 were tested and it can be concluded from the table that mix of both Convolution and bidirectional RNN tends to outperform all the models.

Table 2. Evaluation metric results for HE obtained using fasttext.

HE (FastText)	F1	Accuracy	Precision	Recall
Model 1	0.743	0.745	0.722	0.764
Model 2	0.752	0.761	0.750	0.752
Model 3	0.880	0.887	0.873	0.882
Model 4	0.882	0.889	0.884	0.881
Model 5	0.883	0.892	0.891	0.872
Model 6	0.886	0.891	0.875	0.891
Model 7	0.886	0.902	0.887	0.886

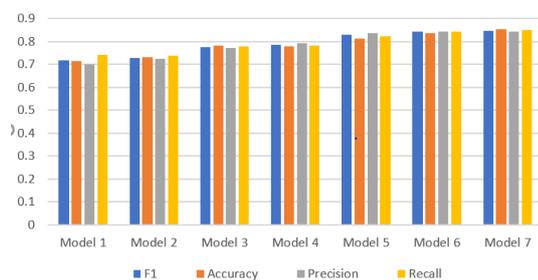
To further strengthen the suitability of proposed architecture, we had also experimented with use of Glove as feature vectors instead of fastText and compared it on various architectures. The hybrid embeddings were generated keeping GloVe as initial

feature vectors and appending POS and position based vectors with it. Table 3 depicts the generation of the hybrid embeddings using GloVe pre-trained embedding. The results suggest that the proposed architecture i.e., model 7 (Hybrid Embedding with Ensemble model CRNN (CNN+BiLSTM)), performed better than all the models which justify its employability for sentiment classification task.

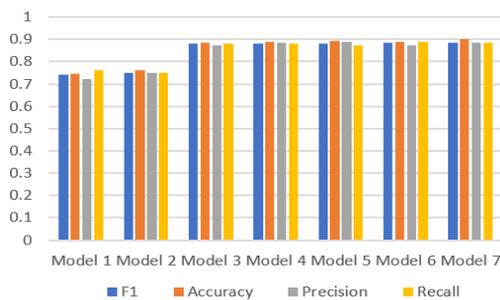
Table 3. Evaluation Metric Results for HE obtained using GloVe.

HE (GloVe)	F1	Accuracy	Precision	Recall
Model 1	0.719	0.715	0.701	0.740
Model 2	0.729	0.730	0.723	0.737
Model 3	0.775	0.782	0.772	0.780
Model 4	0.787	0.780	0.792	0.783
Model 5	0.830	0.812	0.837	0.824
Model 6	0.843	0.836	0.842	0.844
Model 7	0.847	0.854	0.844	0.851

Further, comparing Tables 2 and 3 leads to conclusion that use of fastText with hybrid embeddings is much desired than use of GloVe. Clearly fastText based hybrid embeddings with Convolution and bi-directional RNN is the best methodology for performing sentiment analysis since higher values were obtained for parameters employed in the study. Figure 2 demonstrates the same in graphical manner. The graphs also indicate higher values for the proposed model.



a) Graphical results with GloVe.



b) Graphical results with fastText.

Figure 2. Performance comparison of models using different word embeddings.

In order to further justify the use of fastText rather than employing any other embedding, the performance of proposed hybrid embedding was also tested and compared with pre-trained embeddings (fastText/Glove/word2vec) and employing model 7 as base architecture. Table 4 depicts the comparative performance of all these embeddings. It is clear from the table that fastText outperforms GloVe as well as Word2Vec when used with the CRNN. It can be

concluded from the results that fastText because of its character level approach has out-performed the rest of the pre-trained embeddings as better values are observed for almost all parameters.

Table 4. Comparison of different pre-trained embeddings with Model 7 as base architecture.

Model 7	F1	Accuracy	Precision	Recall
GloVe	0.884	0.888	0.881	0.888
fastText	0.886	0.902	0.887	0.886
word2Vec	0.883	0.886	0.879	0.889

We further compare the proposed approach with state of art approaches in literature; Table 5 depicts final results of proposed techniques and their performance comparison with those techniques of sentiment analysis. It is observed that proposed system has outstanding performance as compared to other state of art approaches. Results clearly state the supremacy of the proposed system over these.

Table 5. Performance comparison with existing systems.

Experiments	Dataset	Accuracy
HE Mixed (Model 7)	MVR Dataset V2	90.21
HE Mixed (Model 7)	MVR Dataset V1	85.60
Improved Word Vector [20]	MVR Dataset V1	82.00
Khan <i>et al.</i> [9] (Min-Max-NormalizedSentiCS with MOMS)	MR Dataset	86.09
Kim [10] (IWV(356))	Movie Reviews Dataset	79.8
Ma <i>et al.</i> [13] (Sentic LSTM +TA+SA)	SemEval -2015 Dataset	76.47
Li <i>et al.</i> [12] (DJLT2,SVM)	SST and Movie Review Dataset	68.39
Deriu <i>et al.</i> [5] (IWV(356))	Movie Reviews Dataset	79.6

From Table 5, we can analyse that the proposed system has outstanding performance as compared to other state of art existing approaches because proposed system had used fastText based hybrid embeddings with Convolution and bi-directional LSTM. The hybrid embeddings helped in generation of new features which could be suitably utilized by the combination of CRNN.

The main advantage of fastText based word embedding is that instead of providing individual words as input to the CRNN, it separates words into n-grams and it not only provided word embeddings for n-grams in training set but was also able to provide word embeddings for rare words. FastText embedding is further improved by use of POS tagging and word position-based features. In the proposed deep architecture, bidirectional LSTM is capable of learning context and temporal features on otherwise sequential data with long term dependencies being maintained from both the directions while CNN retains all the higher order potential features.

5. Conclusions

The study has proposed a new approach for performing

sentiment analysis which successfully improves accuracy of popular pre-trained embeddings of words by deriving hybrid embeddings and further using the same with model of convolution and bi-directional RNN. Experimental results prove that hybrid embedding generated using word position approach, POS tagging approach, and pre-trained fastText resulted in improvement in accuracy of sentiment analysis with accuracy value of 90.21 using MVR Dataset V2. We have tested the same with different pre-trained embeddings of word2Vec/fastText/GloVe on several deep learning-based models using IMDB datasets. The results from the experiments also depict that the proposed methodology increases the accuracy of sentiment classification task with 90.21 accuracy value. The proposed methodology can serve as the basis for deep learning approaches devised for sentiment analysis.

Acknowledgement

This Project was funded by Deanship of Scientific Research (DSR), King Abdulaziz University Jeddah. The authors therefore acknowledge with thanks DSR technical and financial support.

References

- [1] Alghamdi N. and Assiri F., "A Comparison of fastText Implementations Using Arabic Text Classification," in *Proceedings of SAI Intelligent Systems Conference*, London, pp. 306-311, 2019.
- [2] Alqaraleh S., "Novel Turkish Sentiment Analysis System using ConvNet," *The International Arab Journal of Information Technology*, vol. 18, no. 4, pp. 554-561, 2021.
- [3] Araque O., Corcuera-Platas I., Sánchez-Rada J., and Iglesias C., "Enhancing Deep Learning Sentiment Analysis with Ensemble Techniques in Social Applications," *Expert Systems with Applications*, vol. 77, pp. 236-246, 2017.
- [4] Bojanowski P., Grave E., Joulin A., and Mikolov T., "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135-146, 2017.
- [5] Deriu J., Lucchi A., De Luca V., Severyn A., Müller S., Cieliebak M., Hofmann T., and Jaggi M., "Leveraging Large Amounts of Weakly Supervised Data for Multi-Language Sentiment Classification," in *Proceedings of of the 26th International Conference on World Wide Web*, pp. 1045-1052, 2017.
- [6] Fu X., Liu W., Xu Y., and Cui L., "Combine HowNet Lexicon to Train Phrase Recursive Autoencoder for Sentence-Level Sentiment Anal," *Neurocomputing*, vol. 241, pp. 18-27, 2017.
- [7] Giatsoglou M., Vozalis M., Diamantaras K., Vakali A., Sarigiannidis G., and Chatzisavvas K., "Sentiment Analysis Leveraging Emotions and Word Embeddings," *Expert Systems with Applications*, vol. 69, pp. 214-224, 2017.
- [8] Joulin A., Grave E., Bojanowski P., Douze M., Jégou H., and Mikolov T., "Fasttext. Zip: Compressing Text Classification Models," *arXiv preprint arXiv:1612.03651*, 2016.
- [9] Khan F., Qamar U., and Bashir S., "Multi-Objective Model Selection (MOMS)-based Semi-Supervised Framework for Sentiment Analysis," *Cognitive Computation*, vol. 8, no. 4, pp. 614-628, 2016.
- [10] Kim Y., "Convolutional Neural Networks for Sentence Classification," *arXiv preprint arXiv:1408.5882*, 2014.
- [11] Lauren P., Qu G., Zhang F., and Lendasse A., "Discriminant Document Embeddings with An Extreme Learning Machine for Classifying Clinical Narratives," *Neurocomputing*, vol. 277, pp. 129-138, 2018.
- [12] Li Y., Pan Q., Yang T., Wang S., Tang J., and Cambria E., "Learning Word Representations for Sentiment Analysis," *Cognitive Computation*, vol. 9, no. 6, pp. 843-851, 2017.
- [13] Ma Y., Peng H., Khan T., Cambria E., and Hussain A., "Sentic LSTM: A Hybrid Network for Targeted Aspect-Based Sentiment Analysis," *Cognitive Computation*, vol. 10, pp. 639-650, 2018.
- [14] Medhat W., Hassan A., and Korashy H., "Sentiment Analysis Algorithms and Applications: A Survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014.
- [15] Mikolov T., Grave E., Bojanowski P., Puhersch C., and Joulin A., "Advances in Pre-Training Distributed Word Representations," in *Proceedings of the 11th International Conference on Language Resources and Evaluation*, Miyazaki, pp. 52-55, 2018.
- [16] Movie Review Dataset, <https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews>. Last Visited, 2020.
- [17] Pennington J., Socher R., and Manning C., "Glove: Global Vectors for Word Representation," in *Proceedings of the Conference Empirical Methods in Natural Language Processing*, Doha, pp. 1532-1543, 2014.
- [18] Qin P., Xu W., and Guo J., "An Empirical Convolutional Neural Network Approach for Semantic Relation Classification," *Neurocomputing*, vol. 190, pp. 1-9, 2016.
- [19] Ravi K. and Ravi V., "A Survey on Opinion Mining and Sentiment Analysis: Tasks,

- Approaches and Applications,” *Knowledge-Based Systems*, vol. 89, pp. 14-46, 2015.
- [20] Ren Y., Zhang Y., Zhang M., and Ji D., “Improving Twitter Sentiment Classification Using Topic-Enriched Multi-Prototype Word Embeddings,” in *Proceedings of 30th AAAI Conference on Artificial Intelligence*, Phoenix, pp. 3038-3044, 2016.
- [21] Rezaeinia S., Rahmani R., Ghodsi A., and Veisi H., “Sentiment Analysis based on Improved Pre-Trained Word Embeddings,” *Expert Systems with Applications*, vol. 117, pp. 139-147, 2019.
- [22] Santos I., Nedjah N., and de Macedo Mourelle L., “Sentiment Analysis using Convolutional Neural Network with Fasttext Embeddings,” in *Proceedings of IEEE Latin American Conference on Computational Intelligence*, Arequipa, pp. 1-5, 2017.
- [23] Severyn A. and Moschitti A., “Twitter Sentiment Analysis with Deep Convolutional Neural Networks,” in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, New York, pp. 959-962, 2015.
- [24] Tang D., Wei F., Yang N., Zhou M., Liu T., and Qin B., “Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification,” in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, pp. 1555-1565, 2014.
- [25] Zhang H., Gan W., and Jiang B., “Machine Learning and Lexicon based Methods for Sentiment Classification: A Survey,” in *Proceedings of the 11th Web Information System and Application Conference*, Tianjin, pp. 262-265, 2014.



Fahd Alotaibi is an Associate Professor in the Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Saudi Arabia. His research interests are Artificial Intelligence, Natural Language Processing, Machine Learning, and Information Retrieval. He has written many research papers in reputed international and national journals.



Vishal Gupta is serving as Associate Professor in CSE at UIET, P.U. Chandigarh. He has published more than 90 research papers in top international journals and reputed conferences. His research area is natural language processing. He was selected in world top 2% Scientists ranking list-2019 released by Stanford University in Computer Science.