

## A COMBINED DEEP LEARNING MODEL FOR PERSIAN SENTIMENT ANALYSIS

ZAHRA BOKAEE NEZHAD<sup>1</sup> AND MOHAMMAD ALI DEIHIMI<sup>2</sup>

<sup>1</sup>Department of Information and Computer Science, Zand university, Shiraz, Iran.

<sup>2</sup>Department of Electrical and Electronics Engineering, Bahonar University, Shiraz, Iran.

\*Corresponding author: [m.a.deihimi@gmail.com](mailto:m.a.deihimi@gmail.com)

(Received: 10<sup>th</sup> Dec 2018; Accepted: 17<sup>th</sup> April 2019; Published on-line: 1<sup>st</sup> June 2019)

<https://doi.org/10.31436/iiumej.v20i1.1036>

---

**ABSTRACT:** With increasing members in social media sites today, people tend to share their views about everything online. It is a convenient way to convey their messages to end users on a specific subject. Sentiment Analysis is a subfield of Natural Language Processing (NLP) that refers to the identification of users' opinions toward specific topics. It is used in several fields such as marketing, customer services, etc. However, limited works have been done on Persian Sentiment Analysis. On the other hand, deep learning has recently become popular because of its successful role in several Natural Language Processing tasks. The objective of this paper is to propose a novel hybrid deep learning architecture for Persian Sentiment Analysis. According to the proposed model, local features are extracted by Convolutional Neural Networks (CNN) and long-term dependencies are learned by Long Short Term Memory (LSTM). Therefore, the model can harness both CNN's and LSTM's abilities. Furthermore, Word2vec is used for word representation as an unsupervised learning step. To the best of our knowledge, this is the first attempt where a hybrid deep learning model is used for Persian Sentiment Analysis. We evaluate the model on a Persian dataset that is introduced in this study. The experimental results show the effectiveness of the proposed model with an accuracy of 85%.

**ABSTRAK:** Hari ini dengan ahli yang semakin meningkat di laman media sosial, orang cenderung untuk berkongsi pandangan mereka tentang segala-galanya dalam talian. Ini adalah cara mudah untuk menyampaikan mesej mereka kepada pengguna akhir mengenai subjek tertentu. Analisis Sentimen adalah subfield Pemprosesan Bahasa Semula Jadi yang merujuk kepada pengenalan pendapat pengguna ke arah topik tertentu. Ia digunakan dalam beberapa bidang seperti pemasaran, perkhidmatan pelanggan, dan sebagainya. Walau bagaimanapun, kerja-kerja terhad telah dilakukan ke atas Analisis Sentimen Parsi. Sebaliknya, pembelajaran mendalam baru menjadi popular kerana peranannya yang berjaya dalam beberapa tugas Pemprosesan Bahasa Asli (NLP). Objektif makalah ini adalah mencadangkan senibina pembelajaran hibrid yang baru dalam Analisis Sentimen Parsi. Menurut model yang dicadangkan, ciri-ciri tempatan ditangkap oleh Rangkaian Neural Convolutional (CNN) dan ketergantungan jangka panjang dipelajari oleh Long Short Term Memory (LSTM). Oleh itu, model boleh memanfaatkan kebolehan CNN dan LSTM. Selain itu, Word2vec digunakan untuk perwakilan perkataan sebagai langkah pembelajaran tanpa pengawasan. Untuk pengetahuan yang terbaik, ini adalah percubaan pertama di mana model pembelajaran mendalam hibrid digunakan untuk Analisis Sentimen Persia. Kami menilai model pada dataset Persia yang memperkenalkan dalam kajian ini. Keputusan eksperimen menunjukkan keberkesanan model yang dicadangkan dengan ketepatan 85%.

---

**KEYWORDS:** *sentiment analysis; natural language processing; deep learning; Persian; machine learning*

## 1. INTRODUCTION

In recent years, social media sites have gained a considerable number of members. These sites allow users to express their ideas, so there is a great opportunity for companies and organizations to be informed of users' ideas as public view [1]. Today many companies try to provide acceptable customer services to encourage people to buy their products. There is also a profound need for some organizations to know people's political views. On the other hand, with such massive opinionated text data, such as comments in social media sites, reviews in different websites, or even blog posts, companies and organizations must rely on tools to analyse this data. Sentiment Analysis technique has emerged for these purposes [2]. It is one of the most popular applications of NLP.

Sentiment Analysis can be applied at different levels of scope: document level, sentence level, and aspect level. Document level Sentiment Analysis is used to detect the whole document sentiment. Sentence level Sentiment Analysis is used to determine a sentence sentiment while Aspect level Sentiment Analysis is used to recognize every aspect of each entity which is mentioned in a sentence. For example in the sentence: "I really love Apple products, but the latest iPhone is not that great!" There is a positive sentiment about Apple products but a negative opinion about the latest iPhone [3] [4].

Although extracting features from documents plays an important role in classification tasks, most Persian Sentiment Analysis research uses traditional machine learning methods. These methods need to manually extract features from data. Some of the famous techniques are Naïve Bayes (NB), Support Vector Machine (SVM) and logistic regression [5]. On the other hand, text inputs should be fed into Neural Networks in a proper way. Bag of Words (BOW) is one famous method to represent each sentence to the proper form for machine learning. This method is quite simple and effective, but it ignores word order in the document [6], which can cause a significant problem in Sentiment Analysis tasks for two sentences with the same set of words and different sentiments. BOW also doesn't consider the semantic similarity between words [7].

There are several methods for representing a word properly to feed to machine learning algorithms such as N-gram and binary features. Currently, except scant research in Chinese and Indian languages, the majority of the research is performed on English text. Moreover, Persian language faces some challenging complexity such as a wide variety of suffixes, lack of enriched dataset, etc. [5].

Deep learning has become the main part of many new applications such as Speech Recognition, Object Detection, Text Mining, etc. In this study, a novel deep learning model for Persian Sentiment Analysis is presented. It combines two types of Artificial Neural Network: Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN). In recent years, CNN and RNN have been used successfully in Sentiment Analysis tasks.

CNN is well suited to extract local features from text [8]. RNN is used for processing sequential data [9]. LSTM is a special kind of RNN that is capable of learning long-term dependencies. It can also solve the problem of gradient diffusion and gradient explosion [10].

In this paper, Convolution and Recurrent Neural Network are combined for Persian Sentiment Analysis. We take advantage of automatic feature extraction by using CNN. Long-distance dependencies are also learned by LSTM. On the other hand, instead of representing each word by its index in the vocabulary, we use word2vec as an embedding

algorithm, which is key to the state-of-the-art results achieved by the proposed neural network architecture. Two datasets are introduced in this study. They have about 9000 and 3000 sentences respectively that are labelled as positive, negative and neutral.

The contents of this paper are divided into the following sections: section 2 talks about previous works. Section 3 consists of the proposed model. We describe our hybrid CNN and RNN model in detail in this section. Section 4 describes the dataset and presents experimental results. Finally, the paper is concluded in the last section.

## 2. RELATED WORK

Sentiment Analysis is a field of NLP that focuses on identification and extraction of people's opinions. Generally, the various challenges present in the Persian language have limited the quantity of works done on Persian Sentiment Analysis. However, Sentiment Analysis is an on-going field of research, and in this section, some studies on Sentiment Analysis are summarized.

Pang and Lee [11] used a Movie Reviews dataset and selected different words from their dataset as features. They examined features through different machine learning techniques such as Naive Bayes, Max-Entropy and SVM. In comparison to studies reported for standard topic-based categorization, they were not able to achieve desirable results. They also couldn't address the co-reference problem in a sentence. Wang and Manning [12] investigated a simple NB and SVM variant for sentiment classification. They showed that using word-bigram features created robust results. Furthermore, their study showed that SVM and NB were effective for full-length and short reviews respectively. For text classification, Bag of Words model was used to represent their document. One drawback of this model occurs when the size of the training set is small. BOW also ignores the grammatical structure and word order. However, Mass et al. [13] introduced a novel model that captured both semantic and sentiment similarities among words. To capture sentiment similarities, they used a supervised learning method. The semantic similarities were also learned by word2vec. Although their model was sensitive to seed words and missing key words could lead to poor performance, they achieved acceptable results.

Socher et al. [14] used a Sentiment Treebank and proposed a model which is called Recursive Neural Tensor Network (RNTN). It was too similar to the model that was proposed by Socher one year before [15] which is called Matrix-Vector RNN (MV-RNN). MV-RNN was similar to Recursive Neural Network but to represent words and phrases, it assigned both a vector and a matrix to each node in the parse tree. Their model could learn the meaning of each word. It could also learn how a word modifies its neighbours. However, one main problem with MV-RNN was that the number of parameters heavily depended on the size of the vocabulary and became very large. To address this problem, the authors proposed a RNTN model that used the same, tensor-based composition function for all nodes.

Le and Mikolov [6] proposed a Paragraph Vector, an unsupervised algorithm that learned continuous distributed vector representations. They used their model for Sentiment Analysis task and achieved state-of-the-art results. The novelty of their study was that the model could be applied to variable-length pieces of texts. Their model also represented each document by a dense vector.

Bagheri et al. [16] suggested a model based on a lemmatization approach for Persian Sentiment Analysis. They employed a Naive Bayes algorithm for classification. Their model, such as most of other Persian Sentiment Analysis studies used traditional machine

learning models (Alimardani et al. [17], Hajmohammadi et al. [18], Basiri et al. [19]). One drawback of these studies was that they relied on manual feature engineering while the proposed model tries to learn high-level features from data in an incremental manner using a deep learning model. Hence, the feature extraction step is not done by expertise. The proposed model can also be more accurate and precise.

Roshanfekar et al. [20] proposed a deep learning model for Persian Sentiment Analysis. Their model had two learning phases, using Skip-gram model for learning vector representation of words and utilizing two deep neural networks (Bidirectional LSTM and CNN) separately in a supervised way. In our study, two hybrid neural network architectures are employed to classify Persian reviews and comments. Concatenated CNN with LSTM and CNN with Gated Recurrent Unit (GRU) are used to classify sentiments into three classes as Positive, Negative, and Neutral. In the proposed model, local features can be captured by CNN and context can be learned by LSTM or GRU whereas in just another work with deep learning for Persian Sentiment Analysis, two single networks were used separately [20] and their model couldn't harness both CNN's and LSTM's abilities. However, the ordering of layers in our proposed model plays a significant role in how well it performs.

To obtain word embedding, Word2vec technique is used in the proposed model. Word2vec uses a neural network with a single hidden layer to train the model. This technique addresses some of the challenges in Persian such as different writing style, existence of inter words space, etc. while most of other word representing models such as term frequency, which was deployed in the previous Persian Sentiment Analysis studies such as Bagheri et al. [16] Alimardani et al. [17], couldn't solve these difficulties.

As mentioned before, there isn't enough research on Persian Sentiment Analysis with deep learning. Based on the author's knowledge, it is the first study on Persian Sentiment Analysis with a hybrid deep learning architecture. We propose the combination of CNN and RNN (LSTM, GRU) that gives a better performance than CNN and RNN separately.

### 3. PROPOSED METHODOLOGY

In this study, a hybrid deep learning model is presented to classify each Persian sentence as positive, negative, or neutral sentiment. Figure 1 shows the overall framework of the proposed model. According to Fig. 1, after creating a Persian dataset, for each sentence, the word embedding step is done by word2vec. Afterward, CNN extracts local features. Multiple kernel sizes are used in this step. A Rectified Linear Unit (Relu) is also applied as activation of CNN layer output. Kim [21] demonstrated that CNN can improve results in sentence classification tasks. After using a pooling layer to create higher order features, the low-level translation constant features which are learned by CNN passed to LSTM as the inputs. LSTM can learn the long term dependencies and process sequence inputs. Finally, after using a fully connected layer, the model tries to predict the sentiment of each input sentence as positive, negative, or neutral.

The content of this section is divided into the following parts: word embedding, convolutional, pooling, dropout, RNN, and fully connected.

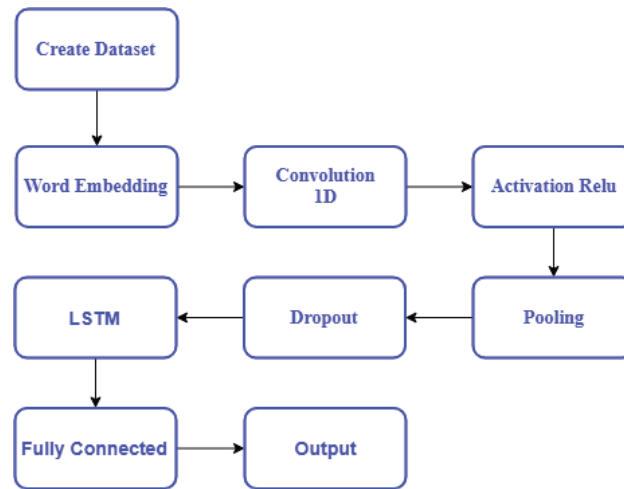


Fig. 1: System architecture of the Proposed Model.

### 3.1 Word Embedding

To perform Sentiment Analysis, the first step is choosing a proper representation of the target document. This step plays a significant role in deep learning. In this study, we use unsupervised learning of word-level embedding using a word2vec model [22]. This model consists of both skip-gram and continuous Bag of Words for computing vector representation. These vectors capture some useful semantic information about words and their relationship to each other.

Let  $S$  be the size of bag of words and  $Le$  be the length of a word-embedding, so the word embedding for vocabularies is encoded by column vectors in an embedding matrix  $Q \in R^{S \times Le}$ . A sentence can be represented in the way that is shown in Eq. (1):

$$Sen = \{w_1, w_2, w_3, \dots, w_l\}, w_i \in [1, S], i \in [1, l] \quad (1)$$

The sentence-level representation presents in the Eq. (2-3):

$$C_i = Q [w_i], C_i \in R^{Le} \quad (2)$$

$$C = [C_1, C_2, C_3, \dots, C_l], C \in R^{Le \times l} \quad (3)$$

### 3.2 Convolution

CNN is a specific type of neural network. CNN can work well with spatial data because it only uses the certain connection from the previous layer; it can understand the sentiment of inputs in an acceptable way [8]. The convolution operation goes across the whole input matrix that was mentioned in section 3.1 to extract sentiment features.

For each window of given input with the size of  $w$ , the same weight matrix is applied to extract local features vector of size  $l-w+1$  by utilizing the matrix over all words of the input. A Rectified Linear Unit (Relu) is also applied as activation of a CNN layer output that replaces negative outputs with zero. The output of this layer has the same shape as the input.

### 3.3 Max Pooling

We apply the max-pooling operation to the output of the convolutional layer, so the feature maps are passed on to this layer to take the maximum feature value. Let  $c_i$  be a feature that was generated in the previous section: max-pooling operation takes the maximum feature value  $c_{max}$  among one map  $c$  in the way that shows in Eq. (4).

$$C_{max} = \max\{c\} = \max \{c_1, \dots, c_l\} \quad (4)$$

### 3.4 Drop out

In this section, we use a dropout layer that randomly sets a portion of the input to zero. This layer prevents network overfitting. It also distributes the network so as to not concentrate on particular pieces of input.

### 3.5 RNN

A recurrent neural network (RNN) is a type of artificial neural network. It is used in several NLP studies. They are designed to recognize the characteristics of a sequence of data. [23] As mentioned before, features are captured by CNN. In this step, LSTM and GRU as two kinds of Recurrent Neural Networks are used to learn context. Therefore, the features are taken to Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU).

LSTM is an artificial Recurrent Neural Network. It is capable of learning Long-Term dependencies so it can process the entire sequence of data. A common LSTM unit is composed of a cell, an input gate; a forget gate and an output gate. The input gate determines which information should enter the cell state. The forget gate determines which information must be removed from previous cell states to only keep relevant information. The output gate determines how much of the internal state should be exposed to the higher layers. Figure 2 shows how LSTM computes hidden state with Eq. (5-10) [23].

$$i_t = \sigma(x_t U^i + h_{t-1} W^i) \quad (5)$$

$$f_t = \sigma(x_t U^f + h_{t-1} W^f) \quad (6)$$

$$o_t = \sigma(x_t U^o + h_{t-1} W^o) \quad (7)$$

$$\hat{C}_t = \tanh(x_t U^g + h_{t-1} W^g) \quad (8)$$

$$C_t = \sigma(f_t \odot C_{t-1} + i_t \odot \hat{C}_t) \quad (9)$$

$$h_t = \tanh(C_t) \odot o_t \quad (10)$$

Where  $i, f, o$  are input, forget, and output gates, respectively.  $\sigma$  denotes the logistic sigmoid function,  $w$  is the recurrent connection at the previous and current hidden layer.  $U$  is the weight matrix that connects the inputs to the current hidden layer.  $x$  is the input vector.  $\hat{C}_t$  is a candidate hidden state.  $C_t$  is the internal memory of the unit.

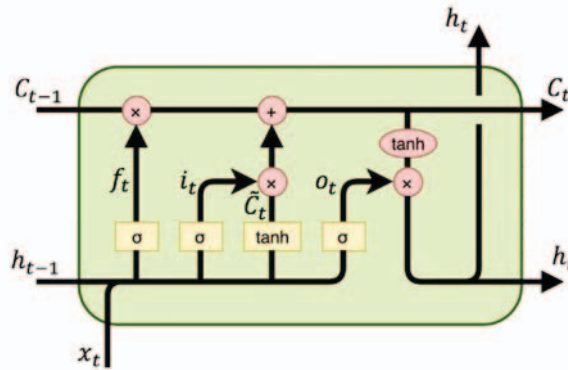


Fig. 2: LSTM unit.

GRU is a gating mechanism in RNN. It is like LSTM but has fewer parameters. It is also easy to train. A GRU unit has two gates: an update gate and a reset gate. The update gate

determines how much of the previous memory should be kept around. The reset gate determines how to combine the new input with the previous memory. Figure 3 shows how GRU computes the hidden state with Eq. (11-14) [23].

$$r_t = \sigma(x_t U^r + h_{t-1} W^r) \tag{11}$$

$$z_t = \sigma(x_t U^z + h_{t-1} W^z) \tag{12}$$

$$\hat{h}_t = \tanh(x_t U^h + (r_t \odot h_{t-1}) W^h) \tag{13}$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \hat{h}_t \tag{14}$$

Where  $r$  is a reset gate, and  $z$  is an update gate.

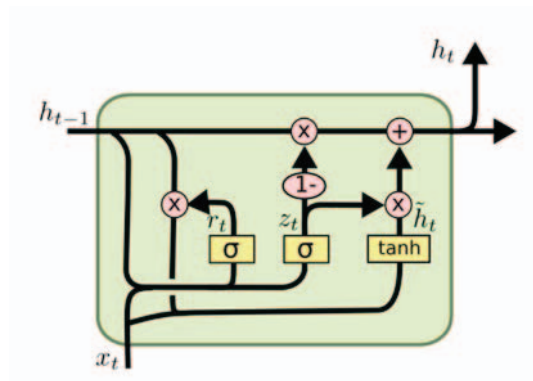


Fig. 3: GRU unit.

### 3.6 Fully Connected

The last layer is a fully connected layer that receives all the features generated before the final output. We utilize cross-entropy as the loss function, which measures the difference between the real and the output sentiment distribution. The output of dense layer is calculated by loss function. Adam optimizer is also used in this step. The fundamental architecture of the proposed model for a Persian sentence is shown in Fig. 4. According to Fig. 4, for each input sentence, word embedding step is done by Word2vec. Afterward, the model tries to predict each sentence sentiment with the combination of two Neural Networks: CNN and LSTM.

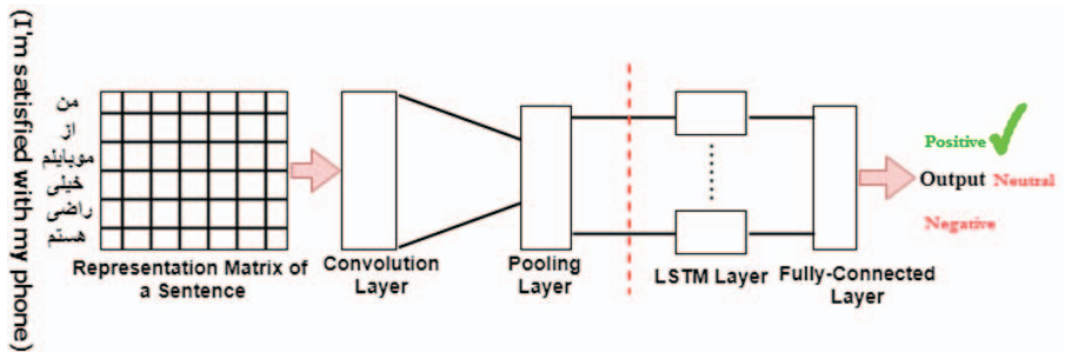


Fig. 4: Proposed Model Diagram.

## 4. EXPERIMENT AND RESULTS

The experimental setup consists of a window operating system running a python virtual environment on python 3.6. The setup was done on Pycharm, which is an open-source application. This section is divided into the following parts: data, convolutional and activation, max-pooling and dropout, RNN and baseline.

### 4.1 Data

In Persian, one problem of NLP tasks is that there is no standard dataset for Sentiment Analysis. In this section, we used a crawler to collect our data from [www.digikala.com](http://www.digikala.com) which is a famous Persian site for electronic products. We named this dataset “PE” which stands for Persian Electronic dataset. We also used twitter streaming API for creating our second dataset about the political issue in Persian. We named this dataset “PP”, which stands for Persian Political dataset. “PE” as the basic dataset, has 9066 sentences and “PP” has 2550 sentences. We manually classified each sentence as positive, negative, or neutral using ten different computer scientists and native speakers to decide about each sentiment. Table 1 illustrates all details about each dataset. In the table (Nu, N+, N-) are the number of neutral, positive, and negative sentences respectively.

Table 1: Dataset

Dataset	N+	N-	Nu	Total
PE	3022	3022	3022	9066
PP	850	850	850	2550

The following sections determine the optimal parameters drawn by the experiment.

### 4.2 Convolutional Layer and Activation

The optimal 1- dimensional kernel sizes were 3, 5, and 7. These kernel sizes produced higher accuracy. The optimal number of the filter was 128. The Relu activation layer was also used in this section

### 4.3 Maxpooling and Dropout

Regarding our experiment for max-pooling, the optimal kernel size was 2. The dropout layer was set at the rate of 0.5 to reduce overfitting. It conducted the model for understanding the data in a better way and could also be the cause of a higher accuracy.

### 4.4 RNN

LSTM and GRU have 128 and 256 units respectively. Regarding the experiment, increasing or decreasing, the number of units causes overfitting. It can also reduce accuracy.

### 4.5 Results

We compared our results with four well-known Sentiment Analysis studies. Table 2 reports the experimental results of each model on our Persian dataset.

According to the table, model 1 is one of the best traditional approaches called NBSVM [12]. Model 2 and 3 are two studies on Persian Sentiment Analysis. They deployed Naive Bayes and Logistic Regression respectively [19] [17]. Model 4 and 5 are suggested as other deep learning studies in Persian [20]. They utilized Bidirectional-LSTM and CNN separately on their proposed models. The results show the hybrid deep learning model outperforms other models since it can harness both CNN's and LSTM's abilities. Furthermore, the proposed model tries to learn high-level features from data in an incremental manner. The proposed model also uses word2vec as an unsupervised learning



step for word representation. Word2vec uses semantic information from the corpus. It helps the model to perform better and overcomes most challenges in Persian such as the different writing style, etc. As shown in the table, CNN-LSTM works almost better in comparison to CNN-GRU, so CNN-LSTM is chosen as the final proposed architecture.

Table 2: Comparative results of different methods on Persian dataset

Model	F-Score	Precision	Recall
Model1:NBSVM	44%	70.7%	31.9%
Model 2: Naïve Bayes	62.5%	65.4%	61.7%
Model 3: logistic regression	67.1%	68.9%	65.4%
Model 4: Bidirectional-LSTM	53.8%	54.3%	55%
Model 5: CNN	67%	66%	67%
<i>Proposed Model: CNN-LSTM</i>	85%	86.5%	86%
CNN-GRU	85.8%	85%	82%
GRU	74%	77.5%	75.6%

Figure 5 is the illustration of the ROC curve for the proposed model. The Roc curve analysis is a fundamental tool for evaluating the performance of different models. It compares True Positive with False Positive rates for all possible thresholds. Figure 5 shows ROC evaluation for CNN-LSTM model on the first dataset (PE). According to the curve, Macro and Micro evaluation for three classes (Class 0: Positive, Class 1: Negative and Class 2: Neutral) of this study are 86% and 83% respectively. A Macro-Average computes the metric independently for each class and then takes the average while A Micro-Average aggregates the contributions of all classes to compute the average metric. The ROC curve also shows the evaluation of each class separately.

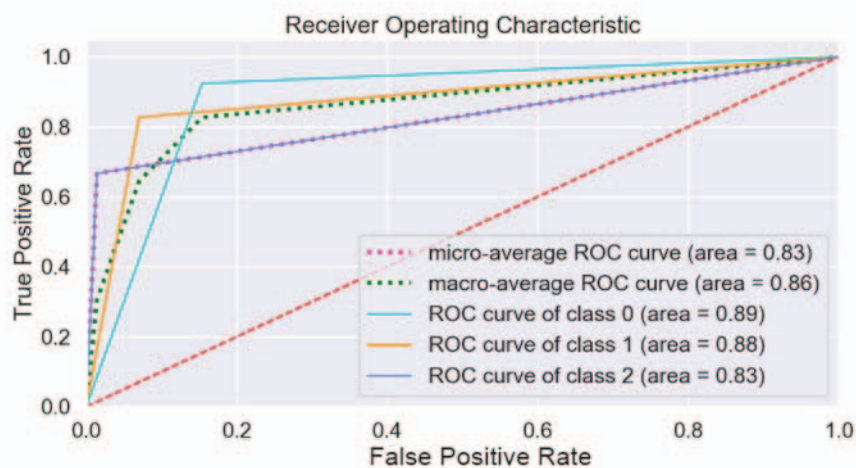


Fig. 5: ROC curve.

## 5. CONCLUSION

Sentiment Analysis is an on-going field of study. However, limited studies have been conducted in Persian language. Based on the author's knowledge, this is the first study on

Persian Sentiment Analysis that uses hybrid deep learning architecture CNN-LSTM. The CNN-LSTM architecture involves CNN for feature extraction and LSTM for learning the long term dependencies. Therefore, the model can harness both CNN's and LSTM's abilities and gets better results with an accuracy of 85%. Additionally, we collected two different Persian datasets from Twitter and digikala as the most famous digital products site. We labelled these sentences as positive, negative, or neutral. They can be useful datasets for Persian language.

## REFERENCES

- [1] Rajadesingan, A., R. Zafarani and H. Liu (2015). Sarcasm Detection on Twitter: A Behavioural Modeling Approach. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, (2015): 97-106.
- [2] Jandail, R. R. S., P. Sharma and C. Agrawal (2014). A Survey on Sentiment Analysis and Opinion Mining: A need for an Organization and Requirement of a customer. In International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering, India, (2016):17-24.
- [3] Feldman, R. (2013). Techniques and applications for sentiment analysis. *Communications of the ACM*, 56: 82-89.
- [4] Mir, J. and M. Usman (2015). An effective model for aspect based opinion mining for social reviews. In Tenth International Conference on Digital Information Management (ICDIM), South Korea, (2015): 49-56.
- [5] Bagheri, A., M. Saraee and F. d. Jong (2013). Sentiment classification in Persian: Introducing a mutual information-based. In 21st Iranian Conference on Electrical Engineering (ICEE), Iran, (2013):1-6.
- [6] Le, Q. and T. Mikolov (2014). Distributed Representations of Sentences and Documents. In Proceedings of the 31 st International Conference on Machine Learning, China, (2014): 1-9.
- [7] Socher, R., J. Pennington, E. H. Huang, A. Y. Ng and C. D. Manning (2011). Semi-supervised recursive auto encoders for predicting sentiment distributions. In EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing, United Kingdom, (2011):151-161.
- [8] Lecun, Y., L. Bottou, Y. Bengio and P. Haffner (1998). Gradient-based learning applied to document recognition. In Proceedings of the IEEE, 86: 2278 – 2324.
- [9] Mikolov, T., M. Karafi'at, L. Burget, J. H. Cernocky and S. Khudanpur (2010). Recurrent neural network based language model. In INTERSPEECH, Japan, (2010): 1045-1048.
- [10] Hochreiter, H. and J. Schmidhuber (1997). LONG SHORT-TERM MEMORY. *Neural computation*, 9(8): 1735-1780.
- [11] Pang, B., L. Lee and S. Vaithyanathan (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. In Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, (2002): 79-86.
- [12] Wang, S. and C. D. Manning (2012). Baselines and Bigrams: Simple, Good Sentiment and Topic Classification. In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, Republic of Korea, (2012): 90-94.
- [13] Maas, A. L., R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng and C. Potts (2011). Learning Word Vectors for Sentiment Analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Oregon, (2011): 1-9.
- [14] Socher, R., A. Perelygin, J. Y. Wu, J. Chuang, C. D. Mannin, A. Y. Ng and C. Potts (2013). Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In Proceeding of the conference on empirical methods in natural language processing (EMNLP), (2013): 1631–1642.
- [15] Socher, R., B. Huval, C. D. Manning and A. Y. Ng (2013). Semantic compositionality through recursive matrix-vector spaces. In Proceedings of the 2012 Joint Conference on Empirical

- Methods in Natural Language Processing and Computational Natural Language Learning, Korea, (2013): 1201–1211.
- [16] Bagheri, A., M. Saraee and d. J. Franciska (2013). Sentiment classification in Persian: Introducing a mutual information-based method for feature selection. In 21st Iranian Conference on Electrical Engineering (ICEE), Mashhad, (2013): 1-6.
- [17] Alimardani, S. and A. Aghaei (2015). Opinion Mining in Persian Language Using Supervised Algorithms. *Journal of Information Systems and Telecommunication*, 3:135-141.
- [18] Hajmohammadi, M. S. and R. Ibrahim (2013). A SVM-Based Method for Sentiment Analysis in Persian Language. In *International Conference on Graphic and Image Processing*, Singapore.
- [19] Basiri, M. E. and A. Kabiri (2017). Sentence-Level Sentiment Analysis in Persian. In 3rd International Conference on Pattern Recognition and Image Analysis (IPRIA 2017), Iran, (2017): 84-89.
- [20] Roshanfekar, B., S. Khadivi and M. Rahmati (2017). Sentiment analysis using Deep learning on Persian Texts. In 25th Iranian Conference on Electrical Engineering (ICEE2017), Iran, (2017):1503-1508.
- [21] Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Qatar, (2014): 1746–1751.
- [22] Mikolov, T., K. Chen, G. Corrado and J. Dean (2013). Efficient estimation of word representation in vector space, In *arXiv preprint arXiv*.
- [23] Chung, J., C. Gulcehre, K. Cho and Y. Bengio (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *arXiv preprint arXiv*.