Stance detection on Social Media, CASE STUDY: Persian sentences using Deep Learning architecture

S.M. Mohammadi[1], S.Farzi[*2], S.M.Alavi[1] , Gh. Heydari[1]

[1]Department of Mathematics and Computer science, Arak Branch, Islamic Azad University, Arak, Iran.
[2]Department of Computer, *K.N. Toosi* University of Technology, Iran.

Abstract

In this paper, we request to identify the stance of Persian language in social networks. The data set is used to detect the stance with Farsi content. A post related to one or more target entities was expressed in a new method, for the first time, and Hybrid LSTM-CNN architecture was used, unlike previous research, a rotational learning rate was used, and a new method for processing data before entering the network for is presented to improve the results, which can stance Persian in the network. With social recognition, in addition to solving the problems related to the lack of data, the BERT model was investigated to detect the Persian stance. As a result, the tagged data from Telegram social network in the field of business sports has been analyzed for a limited time frame, which has achieved higher accuracy than competitors. More precisely, at the end of the training course, the proposed model improved results results by 11.3% in terms of accuracy. As a result, the proposed model has been able to generalize the learned pattern from the training data to the validation data in terms of quality by 12% and in terms of time compared to the presented models.

Keywords: Deep Learning, Persian Sentence, Stance Detection, Long Short-Term Memory Networks.

[1]corresponding author :
*[1]E-mail address:m.mohammadi766@yahoo.com – Tel:09035027112 –Seyed Mahmoud Mohammadi
[2]Assistant Dr.S.Farzi, E-mailaddress: saeedfarzi@kntu.ac.ir
Tel:02184062450 – Saeed Farzi
[1]Assistant Dr.S.M.Alavi, E-mailaddress:sm.alavi@iau.ac.ir
Tel:09183631557- Seyed Majid Alavi
[1]Advisor Dr.Gh.Heydari.g, E-mail address: Gh.Heydari@iau.ac.ir
Tel:09188613211- Gholam Reza Heydari Joonaghani

1. Introduction

Nowadays, according to the increasing request countries to the expansion of social networks, the wide presence of social network users with a variety of beliefs, views and opinions and the publication of many news and messages, identifying the stance of text, in fact, the degree of acceptance, acceptability, the approval of users in the social network. The news that the writer is referring to in social networks is very important. For example, in sports, it detects the popularity or fans of a particular team or discipline. For this purpose and considering the request to identify the stance of the nation in the Persian language, regarding the posts and comments published in virtual space and different social networks. This research was conducted, because we chose the Persian language and felt a lack in this field that ~~a~~ research with technique had not been done in the Persian language, which is the common language of country, and there was a gap. In this regard, the existing obstacle and problem was the lack of a necessary Persian language research, which after numerous follow-ups and referring to various collections. As well as companies various who had done work in this field. Finally, we managed to obtain ***1000 labelled data on the numbers of the Telegram social network people in the field of sports for a limited time related to the stance of different***.

The presented model is implemented in the PyTorch framework. By working it and using the tools of neural networks machine learning deep learning and artificial intelligence, good results were obtained in this field. Stance detection on posts unique challenge for the research community because posts are short, informal user-generated texts that typically do not follow grammatical rules. In this paper, we proposed an attention neural ensemble model to detect target-specific posts and comments stances. The main contribution of our integrated model is the efficient learning of contextual information, which in turn improves the stance detection performance and outperforms advanced deep learning-based methods for single and multi-target stance detection benchmark datasets. Considering that this research has been conducted in the Persian language comprehensive expansion has not been done. We can consider it a subset of sentiment analysis. However, based on its content, sentiment analysis aims to classify the sentiment polarity of a post while stance identification depends on a specific goal. In this paper, we propose a neural ensemble method that combines attention-based, densely connected Bi-LSTM and nested LSTM models with multi-core convolution in a unified architecture. For this purpose, hybrid LSTM-CNN architecture was used and, unlike previous research, a rotational learning rate was used, and a new method for processing data before entering the social network is presented to improve the results, which can stance Persian in the social network. Identifying the social in addition, to solving the problems related to the lack of data, the BERT model was investigated to detect the Persian stance. In addition, posts contain a large number of special abbreviations as well as other post-specific syntaxes such as #hashtags and emoticons. To address the challenges of stance detection in the post, Mohammad et al. (2016) presented a post-stance detection task that focused on a single target in SemEval-2016. Superior systems in this work proposed several approaches based on deep learning using CNN (Wei et al., 2016), RNN (Zarla and Marsh, 2016), etc. Long short-term memory networks are a type of recurrent neural network RNN that can learn and remember long-term memories. LSTMs will retain data over time; they are very useful in forecasting time series because they remember previous inputs. LSTMs have a chain-like structure where four layers interact uniquely. In addition to long-term data memorization, LSTMs are used for speech identification, music composition, and drug development. Later, (Du et al., 2017) used targeted embedding in an attention-based neural network, while (Zhou et al., 2017) proposed.

A semantic-level attention mechanism in a bidirectional **GRU-CNN** structure to perform target-specific stance detection in tweets. Recently, (Dey et al., 2018) proposed a biphasic

LSTM-based model with attention, and (Wei et al., 2018b) presented a neural memory model through target and post interactions. However, most works related to post-stance detection have explored traditional deep-learning models in their methods. Experimental study results on single and multi-objective benchmark stance detection datasets demonstrate the effectiveness of the new method over the advanced deep learning-based methods discussed above. Of course, this research is also the interest and need of several private companies. The motive of the post-detection research is to automatically determine the stance of a post in favor of, against or none of them about a specific target. On the other hand, transformers have created a revolution in the artificial intelligence industry in recent years. Converter is a new architecture that aims to solve sequence-to-sequence tasks while handling long-term dependencies better than recurrent networks. Its basis is based on the mechanism of self-attention. (Siddiqua et al., 2019).

Over Contribution

In this article, the new architecture is presented in the security network, which is compatible with the previous data. We have developed an architecture End to End that has been prepared with a data set on the Persian stance. As a very important achievement in the field that has prepared sports, it will be available to others or searchers for free. With a small volume of sports data and a tone of data, we were able to easily fix our model with the available data and get a good result.The rest of this paper is divided into four sections. Section 2 reviews related work, Section 3 describes the data set used, and Section 4 explains our approach in detail and results.

2. Related work background

It should be stated that research exactly like the one in question has not been done in the Persian language. Of course; similar research had been done on the Persian language in social networks, especially about the diagnosis of rumors. However, there has been a history of research on the English language in the world and in Iran, which is referred to below as an example of these records and research: An example of research done: Based on a paper in 2017, Aker et al. applied simple classification methods along with several features on the **PHEME** data set, and decision tree and random forest classifications were used to classify stances. The features in this article are text, user and publishing network [1]. On behalf of a paper in 2019, Wahiu et al. presented an article in which the textual user and publication features were extracted from the **semeval2017** data set, and an accuracy of 79% was reached by support vector machine classification with a linear kernel [2]. According to a paper in 2019, Ganem et al. presented an approach that detects stances in fake news by combining textual features and word vectors. The proposed rate using neural networks of short-term memory reached 59.7% [3]. Based on a paper in 2018, on the **FNC** data set and using textual features, Hanslovsky et al. reached 59% with a three-layer neural network model of 600 neurons [4]. the paper using Chen et al. suggested examining a claim from diverse and comprehensive perspectives. They believed that different views may have different intensities of approval or rejection. For a claim, they first obtained all relevant texts. In the following, each text should be classified into its important parts that were effective in predicting, while the texts should be graded in terms of confirming or rejecting the claim [5]. Based on the paper by Sedik et al stated that there are hidden meanings in the natural language of humans, they presented a method that can also be used to analyze complex news texts. Such as humor, sarcasm or misleading content [6]. By presenting a paper in 2019, Puran and his colleagues considered the sequence of posts and their responses as a sequence of words and input to a neural network, with the approach of an attention mechanism. The data sets in **PHEME** and twitter15, and the method reached 77% accuracy on these data [7]. Based on a paper in 2018, Bali et al collected Arabic news and created a dataset similar to **FNC** that identifies the stance and rumor. They have classified the stances into four classes, agreeing, opposing, discussed and irrelevant, with the method of slope reinforcement and using a series of textual features, and in doing so, they have reached an accuracy of 56 per cent [8]. Bergonjeh et al. have applied logical regression classification to the **FNC** data set. First, related and unrelated classes are separated, and then to the related class, by using textual features, favorable and unfavorable stances have been obtained. By combining these two stages of classification, the article has reached an accuracy of 59.89 [9]. On behalf of Kokina et al., presented in 2011, proposed a sequence based on short-term memory neural network of 60 Kokina, which reached an accuracy of about 78% by modelling the conversational structure of tweets. Text features have also been used to improve the results of this article. [10]. Based on a paper in 2018, Kanforti et al. presented an article to diagnose the stance and rumor. They removed all the irrelevant samples from the data and classified the rest of the data into three classes agreeing, opposing and discussing with the bilateral short-term memory neural network. They obtain the stance of the headlines about the body of the newspaper and use them to identify and confirm the authenticity of rumors [11]. Based on a paper in 2019, Gorel et al. published a paper with the short-term memory neural network approach of horns in the stance detection section that divides the tweets into four classes of support, denial, inquiry and opinion, and in the confirmation section the truth of the rumor. This article classifies rumors into three classes: true, false and confirmed. Tweets with a tree structure are given as input to the neural network and the accuracy reaches 7.57% [12]. Based on a paper in 2019, Rihanul et al. presented an article with a multi-mode learning approach to determine the user's stance and determine the truth of the rumor simultaneously.

In general, textual information, and user and distribution networks are used to help the article model. Because the user features used in this paper have better results than other papers in this field and it has reached f1-66% on the data of 16 Twitter15 datasets [13]. They are determining the truth of the rumor, which determines the truth of the rumor by using the features extracted from its component and time-based features using a recurrent neural network. The data sets of this article are **PHEME** and **SemEval2017**. This article has reached f1-59% [14]. In 2019, he and his colleagues presented an article in which a multi-state hierarchical learning framework was used to predict the status and validity of rumors. The first component in this article is the detection of the user's stance. In 2019, Zhang et al. presented an article with a hierarchical approach to reduce the error due to the imbalance of the four classes of stance detection through a ring graph network with textual features. With a hierarchical method, they classified these classes, which are a combination of agreement, disagreement, and discussion, in the form of related classes, and by presenting a two-layer neural network, they tried to reduce the amount of error in the **FNC** dataset and a general dataset similar to **FNC** has reached an accuracy of about 89% [15]. In 2019, Lin et. in this work the fc7 features could be revived from the imaging data. The extraction of stimuli representations Using an automatic procedure and the embedding of high-dimensional neuroimaging data onto a space designed for visual object discrimination, a more manageable space from the dimensionality of attitude [16]. According to the paper, Zhang et al. claimed that a text's stance on a claim may not fall into exactly one of the four classes of agree, disagree, debate, and unrelated. They considered the probability of stances for each text and claim and ranked the stance based on this probability [17]. Using Mesadi et al presented a model that, to predict the stance of the news about the claim, extracts the part of the text that determines the result of the prediction. Their model works at the paragraph level and examines each one separately [18]. As a result, they and colleagues considered the diagnosis of rumor and stance as a useful task. They proved a strong connection between the truth of the claim and the stated stance. In the deep network, they considered the common features of two tasks as common weights. This is while each job can have its characteristics [19]. In this paper, we survey the work on stance detection across those communities and present an exhaustive review of stance detection techniques on social media, including the task definition, different types of targets in stance detection, features set used, and various machine learning approaches applied. Our survey reports state-of-the-art results on the existing benchmark datasets on stance detection and discusses the most effective approaches[20].

In this paper, two solutions are used to address this issue: 1) the use of data augmentation and 2) the application of different learning approaches (machine learning, deep learning, and transfer learning) and a meaningful combination of their outcomes[21]. In this paper, new baseline models were studied for fake news classification using CNN [22]. This paper proposes a novel adaptive cost-sensitive loss function for learning imbalanced stance data using deep neural networks, which improves the performance of stance classifiers in rare classes[23]. Stance Detection Benchmark: How robust is your stance detection[24]. In this paper to address the cross-target stance detection in social media by leveraging the social nature of the task, we introduce CT-TN, a novel model that aggregates multimodal embedding derived from both textual and network features of the data[25]. This manuscript proposes a news stand point discrimination method based on a social back ground in formation fusion heterogeneous network[26]. In paper sets out to analyze data augmentation DA methods in detail according to the above categories [27]. This review paper provides an overview of data pre-processing in machine learning, focusing on all types of problems while building the machine learning problems[28]. In this paper a new deep learning-based method that fuses a back translation method, and a paraphrasing technique for data augmentation[29].

In this paper, our goal is to evaluate attention models and **BERT** a pre-trainined language representation in the Persian language. one of the versions of **BERT** is un-normalized multilingual model that contains 104 languages and Persian language is one of its languages. [30].

The following Figure 1 shows an overview of our proposed framework: First, we use convolution filters to extract a sequence of high-level features from the attached tweets of interest. These feature sequences are fed into nested **LSTMs** based on the attention mechanism to learn long-term dependencies. The final representations of these modules are concatenated and passed to the state prediction module to determine the stance label.

Proposed stance identifying framework: We describe the details of our proposed Neural Ensemble Model **PNEM** for tweeter stance detection. First, we use the complexity of multi-kernel filters to extract a sequence of higher-level features from the attached posts of the target.

These sequences of features are fed into attention-based densely connected **bi-LSTMs** and nested **LSTMs** to learn long-term dependencies. The final representations of these modules units-instances are connected and transferred to the state prediction module to determine the stance label.

3. Method

Previous works have investigated the importance of target information for stance detection. To integrate target information, we create a unified word vector matrix by concatenating target and post vector (representations). For this purpose, we use a pre-trained **BERT** model to generate these representations. ***In some cases that we use Farsi, change BERT to IMBERT.***

BERT is a self-supervised transformer model that is already based on a large set of English data. This means that it is only pre-trained on raw texts and trained with an automated process to generate inputs and labels from those texts, without humans labelling them, which is why. It can use a lot of publicly available data. More precisely, with two pre-trained objectives:

The cross entropy between two different distributions P and Q with the same support X is defined as:

$$H(P,Q) = -\sum[P(x) * LOG(Q(x))]$$

● Masked language modelling: Taking a sentence, the model randomly masks 15% of the input words and then runs the entire masked sentence through the model and should predict the masked words. This is different from traditional recurrent neural networks, which typically see words one after the other or from autoregressive models like **GPT**, which internally hides future tokens. This allows the model to learn the two-way representation of the sentence.

● Next sentence prediction**:** Models associate two masked sentences as input during pre-training. Sometimes they correspond to sentences that were together in the original text, sometimes not. Then the model should predict whether the experimental results of two sentences follow each other or not.

In this way, the model learns an internal representation of the English language that can then be used to extract useful features for other tasks. The dimensions of the matrix will be L × D, where the length of L is the sum of the length of the target and the length of the tweet, and D represents the dimension of the word representation vector. We use a pre-trained word embedding model to obtain vector representations of words. Then, we extract higher-level features in convolution. The input of this module is the tweet matrix to which the targets are attached. We apply several convolutions based on four different kernel sizes (filters of size **2**, **3**, **4**, and **5**). After convolution, each filter produces corresponding feature maps and then a max pool function is applied to them. Finally, the feature vectors generated from each kernel are concatenated to form a high-level feature vector.

Next, nested **LSTM** architectures **NLSTMs** create temporal hierarchies of memories which have achieved significant improvement over single-layer **LSTM** architectures for learning long-term dependencies.

In **NLSTMs**, **LSTM** memory cells have access to their internal memory, where they can selectively read and write long-term related information. Therefore, compared to **LSTM**, **NLSTM** internal memories operate on longer time scales and efficiently capture context information from input texts.

Recently, an attention mechanism has been introduced in neural network models to effectively model long-term dependencies by enabling the model to learn what it needs to learn based on the input text. Self-attention is an attention mechanism that connects different stances of a sequence.

This mechanism is very useful in machine reading, abstract summarization or image description generation.

To enhance the contribution of important elements in the final representation of the **NLSTM** module, we use a self-attention mechanism to collect all hidden states according to their relative importance weights. Finally, we connect the final post representation from the attention-based **NLSTM** module and send it to a fully connected soft max layer for stance detection. We use cross-entropy as a cost function and stochastic gradient descent based on Adam to learn the model parameters and train the model for 200 and 500 IPOUCH for the base model and the transformer model, respectively.

4. Experimental Study

In this section, we first propose a system compared to the basic model to express the configuration model.

4.1 Model configuration

The presented model is implemented in the PyTorch framework. The dimensions of the used **BERT (**Representations Encoder Bidirectional from Transformers model) are 768 by default since the utilized BERT model is pre-trained. For the convolution layers, we used 4 kernels with sizes 2, 3, 4 and 5, and the number of filters was set to 8. These hyper-parameters are set according to the original work done in as our baseline. Also, in our model, the **NLSTM** module including 1 layer with internal dimensions of 64 is used, which is the default value in [20]. The basic model after 200 IPAC and the presented model after 500 IPAC have reached convergence. Training with a batch size equal to 32 and an initial learning rate of 0.001 was done by the Adam optimizer.

4.2 DATA SET

The dataset used to evaluate the presented method is SemEval-2016 Task 6-A dataset. This dataset contains 3167 samples of posts related to 5 (targets). In this research, 10% of the data were used as (validation) data and 90% of the remaining data were used as training data. Each post corresponding to the tag "Favor" or "Against" is the desired target. The accuracy and loss diagram for the basic model is shown in figures 2, 3, 4 and 5. Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In proceeding of the 10th international workshop on Semantic Evaluation (SemEval), pages 31–41[1].

The increase in accuracy on the training and test data shows that the model has completed the learning process, but due to the low generalizability of the learned patterns, the accuracy of the test is lower than the accuracy of the training.

The descending loss function clearly shows that the model is learning the data, but in the final steps, the amount of training loss is lower than the test loss, which is caused by the low generalizability of the base model. The accuracy and loss diagram for the presented model is also shown in the figure below .Considering that the accuracy of the test is higher than the accuracy of the training, it shows that the model has been able to generalize its learning well. The distance between the training loss function and the test in the presented model is lower than in the basic model, which is due to the high generalizability of the learning process in the presented model. As can be seen in the graphs, the presented model has obtained competitive performances in comparison to the basic model. More precisely, at the end of the training course, the presented model reached 86.8% accuracy and the basic model reached 75.5% accuracy. On the other hand, the distance between the training and validation loss curves in the presented model is smaller. This means that the learning process has been done better and the presented model has been able to generalize the pattern learned from the training data to the validation data. It is also worth noting that due to the use of the learning rate scheduler, the training process in both models is upward and as it approaches the optimal point, it continues on its path by reducing the learning rate and from over shoot. Avoids getting around the optimal point.

4.3 Evaluation results

The steep initial increase in accuracy is due to the model learning the most salient and easy-to-learn patterns in the data. As training progresses, the model starts to learn more subtle and complex patterns, which can be harder to capture accurately. This leads to a slower increase in accuracy and eventually to a plateau or slight decrease in accuracy. In some cases, the curve may exhibit a plateau followed by a sudden increase or decrease. This could indicate that the model has converted to a suboptimal solution and is unable to improve further. In such cases, it may be necessary to adjust the model architecture, hyper parameters, or data preprocessing to overcome the convergence issue.

The accuracy curve goes up slowly at first because the model has to initialize its parameters randomly and has no prior knowledge about the data it is trained on. During the early stages of training, the model is just starting to learn the patterns in the data, so its predictions may not be very accurate. As a result, the model may make many errors, and its accuracy may improve only gradually. Furthermore, the model's learning rate, which determines how much the parameters are updated during each training step, is typically set to a small value at the beginning of training. This helps to prevent the model from making large updates that could lead it to over shoot the optimal parameter values. However, this also means that the model updates its parameters slowly at first, resulting in a slow increase in accuracy. As the model continues to train and learn more about the data, its accuracy improves and it becomes better at making accurate predictions. The learning rate may also be increased to allow the model to make larger updates to its parameters, which can accelerate the rate of improvement in accuracy.

Overall, the conducted experiments show that utilizing general word embedding of the BERT model enhances the stance classification task, instead of learning the embedding specifically for stance detection.

This comes from the fact that learning embedding too much data to reach a well-generalized embedding and in cases with limited data availability using pre-trained embedding, such as BERT embedding, can benefit the classification model.

5. Conclusion

The purpose of stance recognition in the post is to extract the stances of users expressed in a tweet toward one or more target entities. To deal with this problem, most previous studies have investigated traditional deep learning models, for example,L ong Short Term Memory networks **LSTM** and **GRU**. However, compared to these traditional approaches, densely connected **Bi-LSTM** architectures and nested **LSTM** architectures have been proposed recently, which effectively over come the vanishing gradient and over fitting problems and also deal with long-term dependencies.

For this purpose, hybrid **LSTM-CNN** architecture was used and, unlike previous research, the rotational learning rate was used, and a new method for processing data before entering the network is presented to improve the results, which can stance Persian in the network. Identifying the social in addition, to solving the problems related to the lack of data, the **BERT** model was also investigated to detect the Persian stance.

The presented model has achieved higher accuracy than the basic model. More precisely, at the end of the training course, the presented model reached 86.8% accuracy and the basic model reached 75.5% accuracy. On the other hand, the distance between the training and validation loss curves in the presented model is smaller. This means that the learning process has been done better and the presented model has been able to generalize the pattern learned from the training data to the validation data. It is also worth mentioning that due to the use of the learning rate scheduler, the training process in both models is upward and the closer it gets to the optimal point, the lower the learning rate continues its path. It prevents over shoot around the optimal point. More precisely, at the end of the training course, the proposed model improved results by 11.3% in terms of accuracy.

As a result, the proposed model has been able to generalize the pattern learned from the training data to the validation data in terms of 12% quality and in terms of time compared to the presented models.

Acknowledgements

The authors would like to thank the committee of experts From the telegram social network and Dear professors.

References

[1] Aker, A. and Derczynski, L. and Bontcheva, K."Simple open stance classification for rumour analysis." Int. Conf. Recent Adv. Nat. Lang. Process. RANLP, vol. -Septe, pp. 31–39, (2017).

[2] Pamungkas, E. W. and Basile, V. and Patti , V. " Stance Classification for Rumour Analysis in Twitter": Exploiting Affective Information and Conversation Structure.”, (2019) .https://doi.org/10.48550/arXiv.1901.01911

[3] Ghanem, B. and Rosso, P. and Rangel, F."Stance Detection in Fake News a Combined Feature Representation.", pp. 66–71, (2019).

[4] Hanselowski, A. and PVS, A. and Schiller, B. and Caspelherr, F. and et al.\A Retrospective Analysis of the Fake News Challenge Stance Detection Task”, (2018). https://doi.org/10.48550/arXiv.1806.0 180

[5] Chen, S. and Khashabi, D. and Yin, Callison-Burch, W.Ch. and et al."Seeing Things from a Different Angle: Discovering Diverse Perspectives about Claims." NAACL, (2019).

[6] Sadiq, S. and Wagner, N. and Shyu, M. and et al.."High Dimensional Latent Space Variational Auto Encoders for Fake News Detection", USA, pp. 437-442, (2019).

[7] Pouran, A. and Veyseh, B. and Thai, M. T. and et al."Rumor detection in social networks via deep contextual modelling." USA, pp. 113–120, (2019).

[8] Baly, R. and Mohtarami, M. and Glass, J. and et al."Integrating Stance Detection and Fact Checking in a Unified Corpus.", pp. 21–27, (2018).

[9] Bourgonje, P. and Moreno Schneider, J. and Rehm , G.."From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles.", pp. 84– 89, (2018).

[10] Kochkina, E. Liakata, M. and Augenstein, I. " Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM", no. 2016, pp. 475–480, (2018).

[11] Conforti, C.and Pilehvar, M. T. and Collier N.\ "Towards Automatic Fake News Detection: Cross-Level Stance Detection in News Articles." , pp. 40–49, (2019).

[12] Gorrel, G.."RumourEval: Determining Rumour Veracity and Support for Rumours Genevieve.",Proc. 13th Int. Work. Semant. Eval, pp. 845–854, (2019) .

[13] Islam, M.and Muthiah, S. and Ramakrishnan , N."RumorSleuth: Joint Detection of Rumor Veracity and User Stance.", pp. 131–136, (2019).

[14] Wei, P. and Xu, N. and Mao, W.."Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity", pp. 4789–4800, (2019).

[15] Zhang, Q. and Lipani., A."From Stances’ Imbalance to Their Hierarchical Representation and Detection." no. 1, pp. 2323–2332.

[16] Svaner,M. and Savard, M. and Benini, S.A. et al."Transfer learning of deep neural network representations for fMRI decoding.",Journal of Neuroscience Methods,328, (2019).

[17] Zhang, Q. and Yilmaz,E. and Liang Sh."Ranking-based Method for News Stance Detection." In Companion Proceedings of the Web Conference, ACM Press, (2018).

[18] Mohtarami, M.R. and James Glass, B. andNakov, P. and et al."Automatic stance detection using end-to-end memory networks." NAACL-HLT ’18, New Orleans, LA, USA, (2018).

[19] Ma, J. andGao, W. and Wong, K.."Detect rumor and stance jointly by neural multi-task learning", In Companion Proceedings of the Web Conference, pages, 585–593, (2018).

[20] Aldayel, A. and Magdy, W."Stance Detection on Social Media: State of the Art and Trends", (2021).

[21] Farhoodi, M. and Toloie Eshlaghy A. andMotadel, M. R.."Proposed Model for Persian Stance Detection on Social.", (2023).

[22] Mottaghi,V. and Esmaeili, M. and AliBazaee, G. et al. "A Decision-Making System for Detecting Fake Persian News by Improving Deep Learning Algorithms– Case Study of Covid-19 News.", J. Appl. Res. Ind. Eng. Vol. 8, Spec Issue. 1-17, (2021),.

[23] Dutta,S. Caur,S. Chakrabarti,S. and Chakraborty T."Semi-supervised stance detection of tweets via distant network supervision" in Proceedings of the fifteenth ACM international conference on web search and data mining, 241-251, (2022).

[24]Schiller, B. and Daxenberger,B. J. and Gurevych,I."Stance detection benchmark: How robust is your stance detection?", **KI-Künstliche Intelligenz**, (2021).

[25] Khiabani, P.J. and Zubiaga A. " Few-shot learning for cross-target stance detection by aggregating multimodal embedding", **arxivpreprint arXiv:2301.04535**, (2023).

[26] Ren, Y. andLiu, YJ. andGuo, X.And Shi, J. and Jia, M. "News stance discrimination based on a heterogeneous network of social background information fusion", **Entropy**, Vol. 25, No. 1, (2022).

[27] Hou, Li.B. and Y.and Che, W."Data augmentation approaches in natural language processing: A survey", **AI Open**, Vol. 3, (2022).

[28] Maharana, K. and Mondal, S. and Nemade B. "A review: Data pre-processing and data augmentation techniques", (2022).

[29] Beddiar, D.R. and Jahan, M.S.and Oussalah,M. "Data expansion using back translation and paraphrasing for hate speech detection", **Online Social Networks and Media**, ,Vol. 24, (2021).

[30] Karimi, S. and Sadat Shahrabadi, F. "Sentiment analysis using **BERT** (pre-training language representations) and Deep Learning on Persian texts) ", (2019).

Figure 1: Proposed stance detection framework (Post)

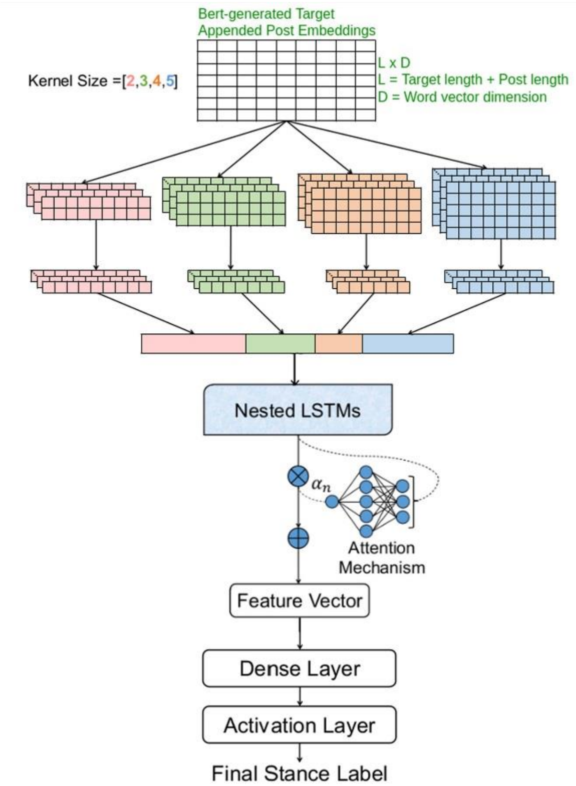


Figure 2: Basic model accuracy graph

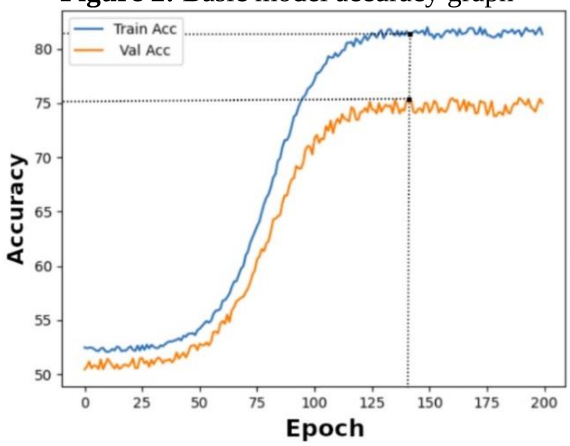


Figure 3: Basic model loss diagram

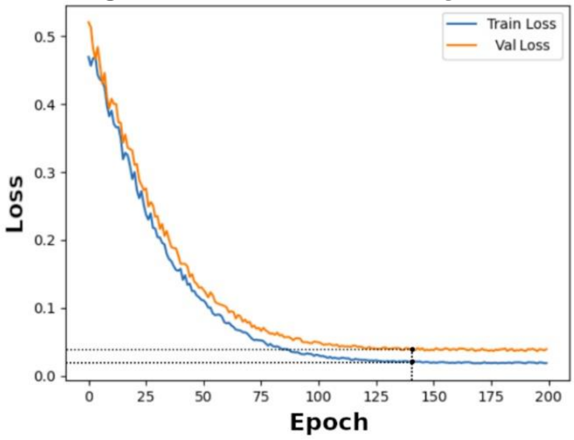


Figure 4: Accuracy graph of the proposed model

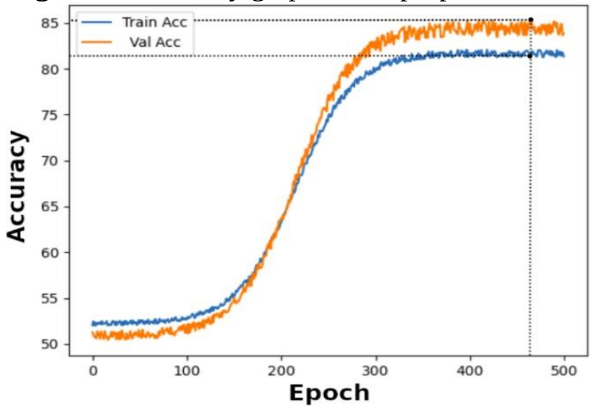
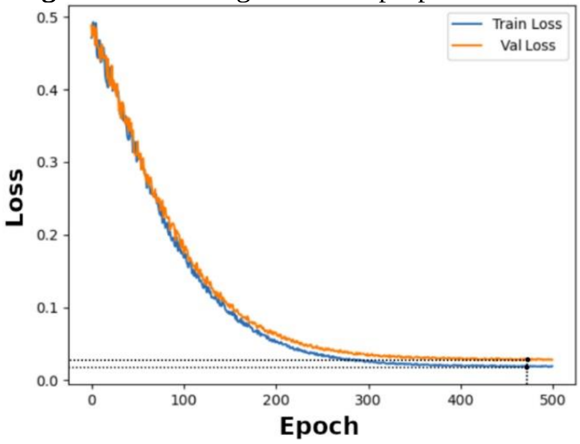


Figure 5: Loss diagram of the proposed model



Brief biographies of all authors:

Seyed Mahmoud Mohammadi is currently a PhD student under the supervision of Dr Farzin and Dr Seyed majid Alavi in the Department of Mathematics and Computer Science at Islamic Azad university Arak Branch, Iran. He obtained his BSc degree from University of Applied Science and Technology in 2014 and an MSc degree from Islamic Azad university in 2017 . His research interests include clustering and data mining and data analysis, industrial clusters, sustainability, uncertainty analysis.

Address: Department of Mathematics and Computer science, Arak Branch, Islamic Azad University, Arak, Iran.
E-mail: m.mohammadi766@yahoo.com
Tel: 09035027112
Phone number: +98-86-34132451
Post code: 3813145537

received the Ph.D. degree in computer engineering from Tehran University, Tehran, Iran, in 2016. He joined the Artificial Intelligence Department, K. N. Toosi University of Technology, Tehran, in 2017. His research interests include machine learning, information retrieval, and social network analysis

Address :Mirdamad Blvd, No. 470, Tehran,Iran
E-mail: saeedfarzi@kntu.ac.ir
Tel: 09183325834
Phone number: +98-21-8462450-401
Post code: 1969764499

Seyed Majid Alavi is currently an Assistant Professor at the Department of Mathematics and Computer Science, Faculty of Science, Islamic Azad University, Arak Branch, Iran. He obtained his PhD and MSc degrees from Islamic Azad University, Science and Research Branch and Ferdowsi University of Mashhad, in 2006, 1997, respectively, and his BSc degree from Shahid Bahonar University in 1994, all in Applied Mathematics. His research interests are in the areas of Numerical Analysis, Fuzzy systems, Data mining.

Address: Islamic Azad University Arak Branch, 3rd km of Khomein road, Imam Khomeini Square , Arak, Iran.
E-mail: sm. alavi@iau.ac.ir
Tel: 09183631557
Phone number: +98-86-34132451
Post code: 3813145537

Gholamreza Heydary Joonaghany is currently an Assistant Professor at the Department of Mathematics and Computer Science, Faculty of Science, Islamic Azad University, Arak Branch, Iran. He obtained his PhD from Islamic Azad University, Science and Research Branch, Tehran, Iran in 2018 and his MSC degrees from Isfahan University of Isfahan, Iran in 1998, and his BSC from Isfahan University of Isfahan, Iran in 1994, all in Pur Mathematics. His research interests are in the areas of Functional Analysis, Optimization and Fixed Point Methods.

Address: Islamic Azad University Arak Branch, 3rd km of Khomein road, Imam Khomeini Square , Arak, Iran.
E-mail: Gh.Heydari@iau.ac.ir
Tel: 09188613211
Phone number: +98-86-34132451
Post code: 3813145537