Scientia Iranica (2024) 31(10), 764–773

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions D: Computer Science & Engineering and Electrical Engineering
https://scientiairanica.sharif.edu

Stance detection on social media, case study: Persian sentences using deep learning architecture

S.M. Mohammadi[a], S. Farzi[b,*], S.M. Alavi[a], and Gh. Heydary Joonaghany[a]

a. *Department of Mathematics and Computer Science, Arak Branch, Islamic Azad University, Arak, Iran.*
b. *Department of Computer, K. N. Toosi University of Technology, Tehran, Iran.*

Received 28 May 2023; received in revised form 17 December 2023; accepted 21 February 2024

KEYWORDS
Deep learning;
Persian sentence;
Stance detection;
Long short-term memory networks.

Abstract. In this paper, we aim to identify the stance of the Persian language in social networks. The dataset is used to detect the stance with Farsi content. In a new method, posts related to one or more target entities are expressed, and a Hybrid LSTM-CNN architecture is used. Unlike previous research, a rotational learning rate is employed, and a new method for processing data before entering the network is presented to improve results, which can stance Persian in the network. With social recognition, in addition to addressing problems related to lack of data, the Bidirectional Encoder Representations from Transformer (BERT) model was investigated to detect Persian stance. As a result, the tagged data from the Telegram social network in the field of business sports was analyzed for a limited time frame and achieved higher accuracy than competitors. More precisely, the proposed model improved results by 11.3% in terms of accuracy at the end of the training course. 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 and time by 12% compared to existing models.

1. Introduction

With the increasing demand for the expansion of social networks worldwide, there is a wide presence of social network users with diverse beliefs, views, and opinions, leading to the publication of numerous news and messages. Identifying the stance of text reflects the degree of acceptance, approval, and acceptability among users in the social network. The news referred to by writers on social networks is very important. For example, in sports, it helps detect the popularity or fans of a particular team or discipline. Hence, this research was conducted considering the need to identify the stance of the nation in the Persian language regarding posts and comments published in virtual space across different

*. *Corresponding author.*
E-mail addresses: m.mohammadi766@yahoo.com (S.M. Mohammadi); saeedfarzi@kntu.ac.ir (S. Farzi); sm.alavi@iau.ac.ir (S.M. Alavi); Gh.Heydari@iau.ac.ir (Gh. Heydary Joonaghany)

To cite this article:
S.M. Mohammadi, S. Farzi, S.M. Alavi, and Gh. Heydary Joonaghany "Stance detection on social media, CASE study: Persian sentences using deep learning architecture", *Scientia Iranica* (2024) **31**(10), pp. 764-773
DOI: 10.24200/sci.2024.62504.7876

social networks. This research specifically focuses on the Persian language due to a lack of research in this field, which is the common language of the country. In this regard, the primary obstacle was the lack of necessary research in the Persian language. After numerous follow-ups and referring to various collections, as well as various companies who had done work in this field, we managed to obtain ***1000 labeled data points from users of the Telegram social network in the field of sports for a limited timeframe, reflecting diverse stances on different topics.***

The presented model is implemented in the PyTorch framework, leveraging the tools of neural networks, machine learning, deep learning, and artificial intelligence to obtain promising results in this field. Stance detection in posts presents a 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 propose an attention-based neural ensemble model for detecting stances in target-specific posts and comments. Our integrated model efficiently learns contextual information, thereby improving stance detection performance and outperforming advanced deep learning-based methods on benchmark datasets for single and multi-target stance detection. Although 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, sentiment analysis aims to classify the sentiment polarity of a post based on its content, 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, a hybrid LSTM-CNN architecture was used, and unlike previous research, a rotational learning rate was utilized, 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. In addition to addressing the problems related to the lack of data, the Bidirectional Encoder Representations from Transformer (BERT) model was investigated for detecting Persian stances. Posts often 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 posts, Mohammad et al. [1] presented a post-stance detection task focused on a single target in SemEval-2016. Superior systems in this work proposed several approaches based on deep learning, using CNN [2], RNN [3], etc. Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) that can learn and remember long-term memories. LSTMs retain data over time; they are useful for 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. [4] used targeted embedding in an attention-based neural network, while Zhou et al. [5] proposed another approach.

A semantic-level attention mechanism in a bidirectional GRU-CNN structure is utilized for target-specific stance detection in tweets. More recently, Dey et al. [6] proposed a biphasic LSTM-based model with attention, and Wei et al. [7] presented a neural memory model focusing on target and post interactions. However, most works related to post-stance detection have explored traditional deep learning models in their methods. Experimental results from studies conducted on single and multi-objective benchmark stance detection datasets demonstrate the effectiveness of the new method compared to the advanced deep learning-based approaches discussed above. Additionally, this research is of interest and importance to several private companies. The objective of post-detection research is to automatically determine the stance of a post regarding a specific target, classifying it as in favor of, against, or neutral. On a different note, 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, based on the mechanism of self-attention [8].

Over contribution

In this article, a new architecture in the security network is presented, which is compatible with previous data. We have developed an end-to-end architecture prepared with a data set on Persian stance. This represents a very important achievement in the field, making sports-related data available to others or researchers for free. With a small volume of sports data and a large amount of other data, we were able to easily fix our model with the available data and get good results. The rest of this paper is divided into four sections: Section 2 reviews related work, Section 3 describes the data set used, 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, and Section 4 first presents a system compared to the base model and then describes its configuration. Section 5 presents the conclusion.

2. Related work background

It should be noted that research exactly like the one in question has not been conducted in the Persian language. However, similar research has been done on the

Persian language in social networks, especially about the identification of rumors. Nevertheless, extensive research has been conducted on the English language in worldwide and in Iran. Below are examples of such studies:

In a paper from 2017, Aker et al. applied simple classification methods along with several features on the PHEME dataset, and decision tree and random forest classifications were used to classify stances. The features in this article included text, user, and publishing network [9]. In a paper from 2019, Pamungkas et al. presented an article in which textual user and publication features were extracted from the SemEval2017 dataset, and an accuracy of 79% was achieved using support vector machine classification with a linear kernel [10]. In another paper from 2019, Ganem et al. presented an approach for detecting stances in fake news by combining textual features and word vectors. The proposed method, using neural networks of short-term memory, achieved a rate of 59.7% [11].

In a paper in 2018, Hanselowski et al. reached a 59% accuracy on the FNC dataset using textual features with a three-layer neural network model consisting of 600 neurons [12].

Chen et al. suggested examining a claim from diverse and comprehensive perspectives. They believed that different views may have varying intensities of approval or rejection. For each claim, they first obtained all relevant texts. Subsequently, each text was classified into its important parts that were effective in prediction, while the texts were graded in terms of confirming or rejecting the claim [13].

Sediq et al. stated in their paper that there are hidden meanings in humans natural language. They presented a method that can be used to analyze complex news texts, such as humor, sarcasm, or misleading content [14]. In 2019, Pouran et al. presented a paper where they considered the sequence of posts and their responses as a sequence of words input to a neural network, incorporating an attention mechanism approach. They used datasets from PHEME and Twitter15, reaching 77% accuracy with their method [15]. Baly et al. in a paper from 2018, collected Arabic news and created a dataset similar to FNC for identifying stance and rumors. They classified the stances into four classes: agreeing, opposing, discussed, and irrelevant. Using slope reinforcement with a series of textual features, they reached an accuracy of 56% [16]. Bourgonje et al. applied logical regression classification to the FNC dataset. First, related and unrelated classes are separated, and then within the related class, favorable and unfavorable stances are obtained using textual features. By combining these two stages of classification, the article reached an accuracy of 59.89% [17]. Kochkina et al., in a paper from 2018, proposed a sequence-based short-term memory neural network of 60 Kochkina. By modeling the conversational structure of tweets, they reached an accuracy of about 78%. Text features have also been used to improve the results of this article [18]. In 2019, Canforti et al. presented an article focusing on diagnosing stance and rumors. They removed all irrelevant samples from the data and classified the remaining data into three classes: agreeing, opposing, and discussing, using a bilateral short-term memory neural network. They obtained the stance of the headlines about the body of the newspaper and then used them to identify and confirm the authenticity of rumors [19]. Gorrell et al., in a paper from 2019, utilized a short-term memory neural network approach for stance detection that divided tweets into four classes: support, denial, inquiry, and opinion. In the confirmation section, they classified rumors into three classes: true, false, and confirmed. Tweets with a tree structure were given as input to the neural network, and the accuracy reached 7.57% [20]. In 2019, Islam et al. introduced an article presenting a multi-mode learning approach to simultaneously determine the user's stance and determine the truth of rumors.

They utilized textual information and user and distribution networks to aid their model. The user features used in this paper yielded better results than other papers in this field, achieving f1-66% on the data of 16 Twitter15 datasets [21]. Regarding the determination of the truth of rumors, they utilized features extracted from the components of the rumors and time-based features, using a recurrent neural network. The datasets used in this article were PHEME and SemEval2017. This article reached f1-59% [22]. In 2019, Wei et al. 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 error due to the imbalance of the four classes of stance detection through a ring graph network with textual features. Using a hierarchical method, they classified these classes, which are a combination of agreement, disagreement, and discussion, into related classes. By presenting a two-layer neural network, they tried to reduce the error in the FNC dataset and a general dataset similar to FNC, reaching an accuracy of about 89% [23]. In 2019, Svanera et al. proposed a method in which the fc7 features could be extracted from imaging data. They used an automatic procedure to extract stimulus representations and embedded high-dimensional neuroimaging data onto a space designed for visual object discrimination, resulting in a more manageable space from the dimensionality of attitude [24].

According to the paper by Zhang et al., 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 [25]. Mohtarami et al. presented a model that extracts the part of the text that determines the result of the prediction to predict the stance of the news about the claim. Their model works at the paragraph level and examines each one separately [26]. Ma et al. 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, while each job can have its characteristics [27]. In this paper, we survey the work on stance detection across these communities and present an exhaustive review of stance detection techniques on social media, including the task definition, different types of targets in stance detection, feature sets 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 [28].

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 [29]. In this paper, new baseline models were studied for fake news classification using CNN [30]. 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 [31]. In 2021, Schiller et al. analysis emphasizes the need of focus to robustness and debiasing strategies in multi-task learning approaches [32]. 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 [33]. This manuscript proposes a news standpoint discrimination method based on a social background information fusion heterogeneous network [34]. This paper sets out to analyze Data Augmentation (DA) methods in detail according to the above categories [35]. 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 [36]. In this paper, a new deep learning-based method fuses a back-translation method and a paraphrasing technique for data augmentation [37].

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 then 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 Twitter stance detection. First, we use multi-kernel filters to extract a sequence of higher-level features from the attached posts of the target.

These sequences of features are then 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 vectors (representations). For this purpose, we use a pre-trained BERT model to generate these representations. ***In cases where Farsi is used, we substitute BERT with IMBERT.***

BERT is a self-supervised transformer model based on a large set of English data. It is only pretrained on raw texts and trained with an automated process to generate inputs and labels from these texts without human annotation. As a result, it can use a lot of publicly available data. More precisely, it is pretrained with two objectives:

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

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

- Masked Language Modeling (MLM): The model randomly masks 15% of the input words in a sentence and then runs the entire masked sentence through the model to predict the masked words. Unlike traditional recurrent neural networks that process words one after the other or autoregressive models like GPT that hide future tokens internally, MLM allows the model to learn the two-way representation of the sentence.
- Next Sentence Prediction (NSP): Models associate two masked sentences as input during pre-training. These sentences may or may not correspond to sentences that were together in the original text. The model then predicts whether the experimental results of the two sentences follow each other or not.

In this way, the model learns an internal representation of the English language, which can then be

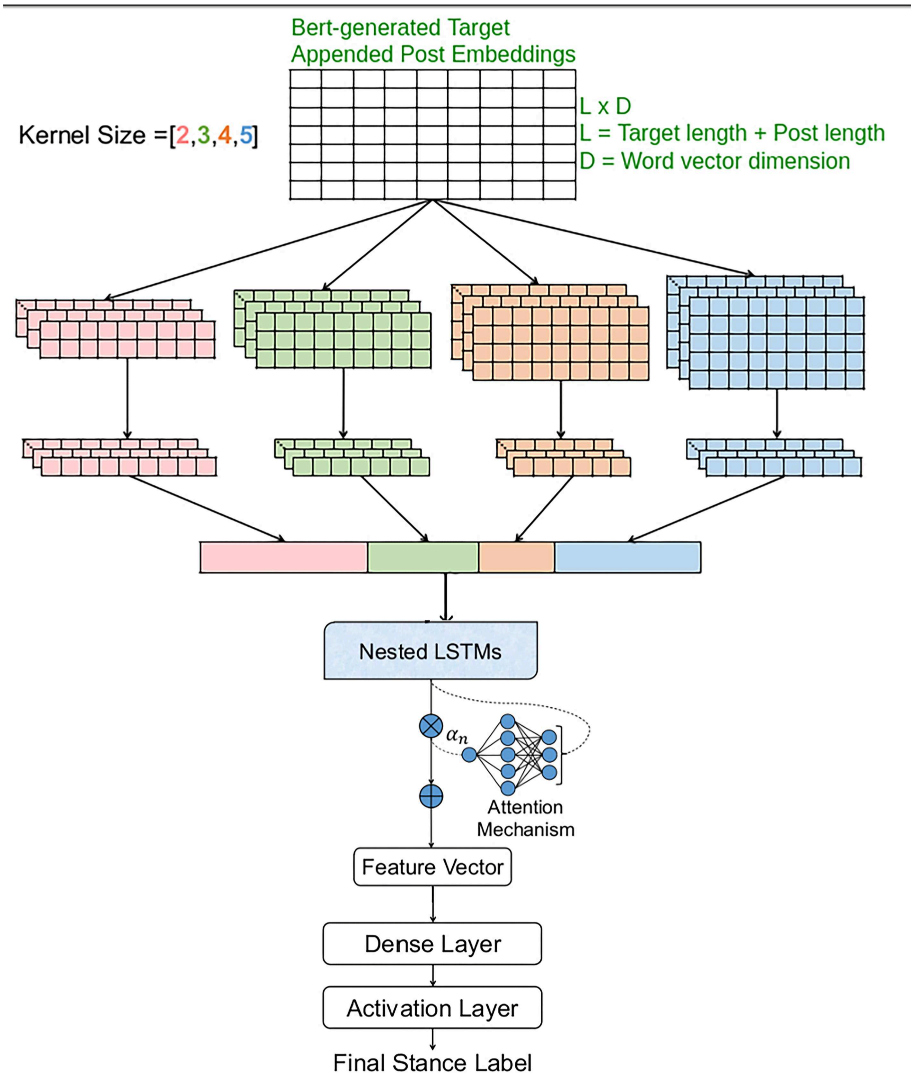

Figure 1. Proposed stance detection framework (Post).

used to extract useful features for other tasks. The dimensions of the matrix will be $L \times D$, 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 using 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 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 softmax layer for stance detection. We use cross-entropy as the cost function and stochastic gradient descent based on Adam to learn the model parameters and train the model for 200 epochs for the base model and 500 epochs for the transformer model.

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 BERT model used are 768 by default, as the BERT model is pre-trained. For the convolution layers, we utilized 4 kernels with sizes 2, 3, 4, and 5, and the number of filters was set to 8. These hyperparameters are set according to the original work done in our baseline. Also, in our model, the NLSTM module includes 1 layer with internal dimensions of 64, which is the default value in [20]. The basic model converged after 200 epochs, while the presented model converged after 500 epochs. Training was performed with a batch size of 32 and an initial learning rate of 0.001 using the Adam optimizer.

4.2. Data set

The dataset used to evaluate the presented method is the SemEval-2016 Task 6-A dataset, which contains 3167 samples of posts related to 5 targets. In this research, 10% of the data were used as validation data and the remaining 90% were used for training. Each post corresponding to the tags "Favor" or "Against" represents the desired target. The accuracy and loss diagrams for the basic model are shown in Figures 2–5.

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) [1].

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

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, indicating the low generalizability of the base model. The accuracy and loss diagram for the presented model is also shown in the figure below. Notably, the higher accuracy of the test compared to the training

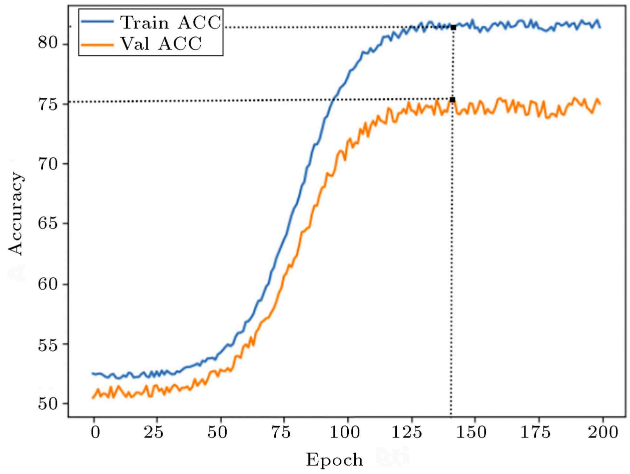

Figure 2. Basic model accuracy graph.

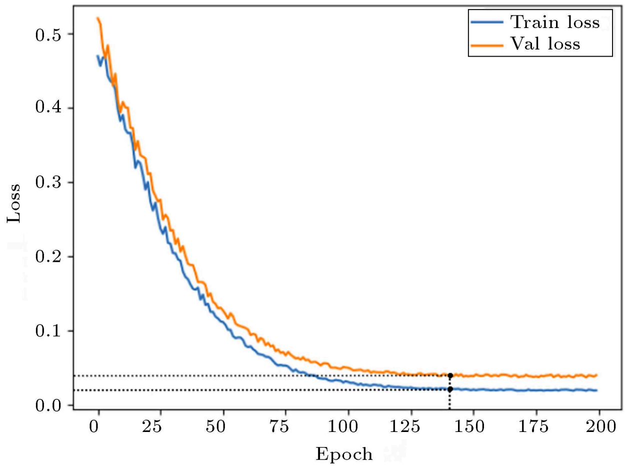

Figure 3. Basic model loss diagram.

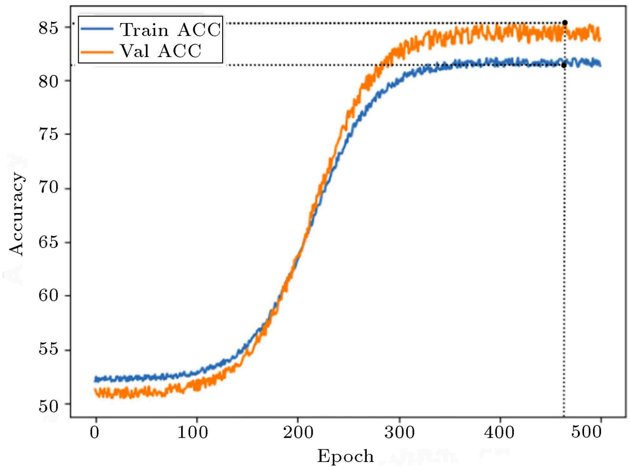

Figure 4. Accuracy graph of the proposed model.

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 smaller than in the basic model, which is due to the high generalizability of the learning process in the presented

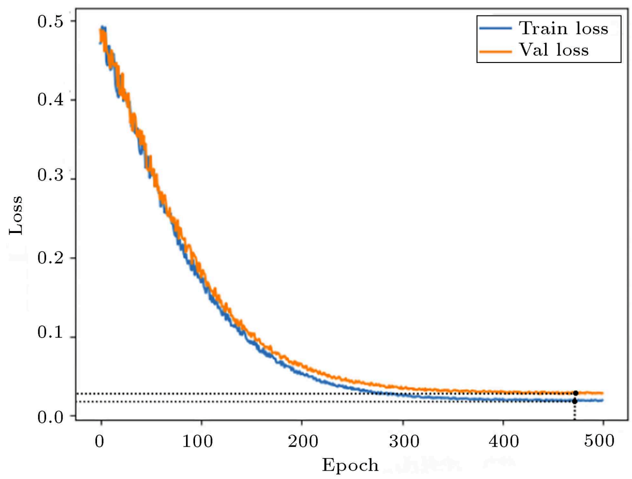

Figure 5. Loss diagram of the proposed model.

model. As shown in the graphs, the presented model has obtained competitive performance compared to the basic model. More precisely, at the end of the training course, the presented model reached 86.8% accuracy, whereas the basic model reached 75.5% accuracy. The smaller distance between the training and validation loss curves in the presented model suggests that the learning process has been more effective, and the presented model has been able to generalize the pattern learned from the training data to the validation data. Additionally, both models benefit from the use of the learning rate scheduler, ensuring a smoother training process. As the training approaches the optimal point, the learning rate is gradually reduced, preventing overshooting and avoiding 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 rises 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 prevent the model from making large updates that could cause it to overshoot 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 the general word embedding of the BERT model enhances the stance classification task, rather than learning the embedding specifically for stance detection.

This stems from the fact that learning embedding from too much data is required 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 posts is to extract users' stances expressed in a Tweet toward one or more target entities. To address this challenge, most previous studies have investigated traditional deep learning models, such as Long Short Term Memory networks (LSTM) and GRU. However, recent advancements have introduced densely connected Bi-LSTM architectures and nested LSTM architectures, which effectively overcome issues like vanishing gradients and overfitting, and also deal with long-term dependencies more efficiently.

For this purpose, a hybrid LSTM-CNN architecture was used, and unlike prior research, a rotational learning rate was employed. Additionally, a new method for processing data before entering the network was presented to improve results, which can stance Persian in the network. In In addition to addressing issues related to lack of data, the Bidirectional Encoder Representations from Transformer (BERT) model was also investigated for Persian stance detection.

The presented model has achieved higher accuracy than the basic model. Specifically, at the end of the training course, the presented model reached 86.8% accuracy, whereas the basic model reached 75.5% accuracy. Additionally, the presented model exhibits a smaller distance between the training and validation loss curves, indicating a more effective learning process and better generalization of learned patterns from the training data to the validation data. It is worth mentioning that due to the use of the learning rate scheduler, the training process in both models is upward, and the learning rate gradually decreases the closer it gets to the optimal point, preventing overshooting 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 with 12% higher quality and in less time compared to the other models.

Acknowledgements

The authors would like to thank the committee of experts from the Telegram social network and the esteemed professors.

References

1. Mohammad, S., Kiritchenko, S., Sobhani, P., et al. "Detecting stance in tweets", In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*, pp. 31–41 (2016). DOI: 10.26615/978-954-452-049-6_005
2. Wei, W., Zhang, X., Liu, X., et al. "A specific convolutio Wang nal neural network system for effective stance detection", In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval)*", NAACL-HLT, pp. 384–388 (2016). DOI: 10.18653/v1/S16-1062
3. Zarrella, G. and Marsh, M. "Transfer learning for stance detection", In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*", pp. 458–463 (2016). DOI: 10.18653/v1/S16-1074
4. Du, J., Xu, R., He, Y., et al. "Stance classification with target specific neural attention networks", In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3988–3994 (2016). DOI: 10.24963/ijcai.2017/557
5. Zhou, Y., Cristea, A., and Shi, L. "Connecting targets to tweets: Semantic attention based model for target-specific stance detection", In *Proceedings of the 18th International Conference on Web Information Systems Engineering (WISE)*, pp. 18–32, Springer (2016). DOI: 10.1007/978-3-319-68783-4_2
6. Dey, K., Shrivastava, R., and Kaushik, S. "Topical stance detection for twitter: A two-phase lstm model using attention", In *Proceedings of the 40th European Conference on Information Re-trieval (ECIR)*, pp. 529–536, Springer (2018). DOI: 10.1007/978-3-319-76941-7_40
7. Wei, P., M., Mao, W., et al. "A target-guided neural memory model for stance detection in twitter", In *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, p. 18, IEEE (2018b). DOI: 10.1109/IJCNN.2018.8489665
8. Siddiqua, U.A., Chy, A.N., and Aono, M. "Tweet stance detection using an attention based neural ensemble model", In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, **1**, pp. 1868–1873 (2019). Minneapolis, Minnesota. Association for Computational Linguistics. https://aclanthology.org/N19-1185 DOI: 10.18653/v1/N19-1185
9. Aker, A., Derczynski, L., and Bontcheva, K. "Simple open stance classification for rumour analysis", *Int. Conf. Recent Adv. Nat. Lang. Process. RANLP*, Septe, pp. 31–39 (2017). DOI: 10.26615/978-954-452-049-6_005
10. Pamungkas, E.W., Basile, V., and Patti, V. "Stance classification for rumour analysis in Twitter: Exploiting affective information and conversation structure", p. 6 (2019). Publication at: https://www.researchgate.net/publication/330212371 DOI: 10.48550/ arXiv.1901.01911
11. Ghanem, B., Rosso, P., and Rangel, F. "Stance detection in fake news a combined feature representation", *Association for Computational Linguistics*, **W18-55**, pp. 66-71 (2019). DOI: 10.1371/journal.pone.0287298
12. Hanselowski, A., PVS, A., Schiller, B., et al. "A retrospective analysis of the fake news challenge stance detection task", *Proceedings of the 27th International Conference on Computational Linguistics*, Publisher: Association for Computational Linguistic, pp. 1859–1874 (2018). DOI: 10.48550/arXiv.1806.0180
13. Chen, S., Khashabi, D., Yin, W., et al. "Seeing things from a different angle: Discovering diverse perspectives about claims", NAACL, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, **1**(Long and Short Papers), Publisher: Association for Computational Linguistics, pp. 542–557 (June 2019). DOI: 10.48550/arXiv.1906.03538
14. Sadiq, S., Wagner, N., Shyu, M., et al. "High dimensional latent space variational auto encoders for fake news detection", USA, pp. 437–442 (2019). DOI: 10.1109/MIPR.2019.00088
15. Pouran, A., Veyseh, B., Thai, M.T., et al. "Rumor detection in social networks via deep contextual modelling", USA, pp. 113–120 (2019). DOI: 10.1093/comjnl/bxab118
16. Baly, R., Mohtarami, M., Glass, J., et al. "Integrating stance detection and fact checking in a unified gorpus", pp. 21–27 (2018). DOI: 10.48550/arXiv.1804.08012
17. Bourgonje, P., 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). DOI: 10.18653/v1/W17-4215

18. Kochkina, E., Liakata, M., and Augenstein, I. "Turing at SemEval-2017 task 8: Sequential approach to rumour stance classification with rbanch-LSTM", **2016**, pp. 475–480 (2018). DOI: 10.18653/v1/S17-2083
19. Conforti, C., Pilehvar, M.T., and Collier, N. "Towards automatic fake news detection: Cross-level stance detection in news articles", pp. 40–49 (2019). DOI: 10.18653/v1/W18-5507
20. Gorrell, G., Kochkina, E., Liakata, M., et al. "RumourEval: Determining rumour veracity and support for rumours genevieve", *Proc. 13th Int. Work. Semant. Eval*, pp. 845–854 (2019). DOI: 10.18653/v1/S19-2147
21. Islam, M.R., Muthiah, S., and Ramakrishnan, N. "RumorSleuth: Joint detection of rumor veracity and user stance", pp. 131–136 (2019). DOI: 10.1145/3341161.3342916
22. Wei, P., Xu, N., and Mao, W. "Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity", pp. 4789–4800 (2019). DOI: 10.18653/v1/D19-1485
23. Zhang, Q., Liang, S., Lipani, A., et al. "From stances' imbalance to their hierarchical representation and detection", **1**, pp. 2323–2332 (2019). DOI: 10.1145/3308558.3313724
24. Svanera, M., Savardi, M., Benini, S., et al. "Transfer learning of deep neural network representations for fMRI decoding", *Journal of Neuroscience Methods*, **328** (2019). DOI: 10.1016/j.jneumeth.2019.108319
25. Zhang, Q., Yilmaz, E., and Liang, Sh. "Ranking-based method for news stance detection", In *Companion Proceedings of the Web Conference*, ACM Press, pp. 41–42 (2018). DOI: 10.1145/3184558.3186919
26. Mohtarami, M., Baly, R., Glass, J.B., et al. "Automatic stance detection using end-to-end memory networks", NAACL-HLT '18, New Orleans, LA, USA, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long Papers), Publisher: Association for Computational Linguistics, pp. 767–776, (June 2018). DOI: 10.18653/v1/N18-1070
27. Ma, J., Gao, W., and Wong, K. "Detect rumor and stance jointly by neural multi-task learning", In *Companion Proceedings of the Web Conference*, pp. 585–593 (2018). DOI: 10.1145/3184558.3188729
28. Aldayel, A. and Magdy, W. "Stance detection on social media: State of the art and trends", **58**(4), 102–597 (July 2021). DOI: 10.48550/arXiv.2006.03644
29. Farhoodi, M., Toloie Eshlaghy, A., and Motadel, M.R. "Proposed model for persian stance detection on social", pp. 1048-1059 (June 2023). DOI: 10.5829/IJE.2023.36.06C.03
30. Mottaghi, V., Esmaeili, M., AliBazaee, Gh., 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.*, **8**, Spec Issue. pp. 1–17 (2021). DOI: 10.22105/jarie.2021.281257.1299
31. Dutta, S., Caur, S., Chakrabarti, S., et al. "Semi-supervised stance detection of tweets via distant network supervision", in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pp. 241–251 (2022). DOI: 10.1145/3488560.3498511
32. Schiller, B., Daxenberger, J., and Gurevych, I. "Stance detection benchmark: How robust is your stance detection?", *KI-Künstliche Intelligenz*, pp. 1–13 (2021). DOI: 10.48550/arXiv.2001.01565
33. Khiabani, P.J. and Zubiaga, A. "Few-shot learning for cross-target stance detection by aggregating multimodal embedding", arxivpreprint arXiv:2301.04535, pp. 2081–2090 (2023). DOI: 10.48550/arXiv.2301.04535
34. Ren, Y., Liu, YJ., Guo, X., et al. "News stance discrimination based on a heterogeneous network of social background information fusion", *Entropy*, **25**(1), p. 78 (2022). DOI: 10.3390/e25010078
35. Li, B., Hou, Y., and Che, W. "Data augmentation approaches in natural language processing: A survey", *AI Open*, **3**, pp. 71–90 (2022). DOI: 10.48550/arXiv.2110.01852
36. Maharana, K., Mondal, S., and Nemade, B. "A review: Data pre-processing and data augmentation techniques", *Global Transitions Proceedings*, **3**, pp. 91–99 (2022). DOI: 10.1016/j.gltp.2022.04.020
37. Beddiar, D.R., Jahan, M.S., and Oussalah, M. "Data expansion using back translation and paraphrasing for hate speech detection", *Online Social Networks and Media*, **24**, 1001543 (2021). DOI: 10.1016/j.osnem.2021.100153

Biographies

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 earned his BSc degree from the University of Applied Science and Technology in 2014 and an MSc degree from Islamic Azad University in 2017 . His research interests include clustering, data mining, data analysis, industrial clusters, sustainability, and uncertainty analysis.

Saeed Farzi earned his PhD degree in Computer Engineering from Tehran University, Tehran, Iran, in 2016. He joined the Artificial Intelligence Department at K. N. Toosi University of Technology, Tehran, in 2017. His research interests include machine learning, information retrieval, and social network analysis.

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 received his PhD and MSc degrees from Islamic Azad University, Science and Research Branch, and Ferdowsi University of Mashhad, in 2006 and 1997, respectively, and his BSc degree from Shahid Bahonar University in 1994, all in Applied Mathematics. His research interests include numerical analysis, fuzzy systems, and data mining.

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, his MSc degree from Isfahan University, Isfahan, Iran, in 1998, and his BSc from Isfahan University, Isfahan, Iran, in 1994, all in Pure Mathematics. His research interests include functional analysis, optimization, and fixed point methods.