



A Proposed Model for Persian Stance Detection on Social Media

M. Farhoodi^a, A. Toloie Eshlaghy^{*a}, M. R. Motadel^b

^a Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran

^b Central Tehran Branch, Islamic Azad University, Tehran, Iran

PAPER INFO

Paper history:

Received 09 January 2023

Received in revised form 16 January 2023

Accepted 17 January 2023

Keywords:

Stance Detection

Data Augmentation

Fake News

Machine Learning

Deep Learning

Transfer Learning

Multi-Classifer Fusion

ABSTRACT

Stance detection is a recent research topic that has become an emerging paradigm of the importance of opinion-mining. It is intended to determine the author's views toward a specific topic or claim. Stance detection has become an important module in numerous applications such as fake news detection, argument search, claim validation, and author profiling. Despite considerable progress made in this regard in languages like English, unfortunately, we have not made good progress in some languages such as Persian, where we are confronted with a lack of datasets in this area. 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. The results show that each of these solutions can not only enhance stance detection performance, but when both are combined, a very significant improvement in the results is achieved.

doi: 10.5829/ije.2023.36.06c.03

1. INTRODUCTION

Social media plays a significant role in the number of people accessing news over the last few decades. With increasing public access to social media, a lot of dubious and incorrect content is being produced and shared for beneficial purposes. Nowadays, people often use social media to express their opinions on published content [1]. These stances together can prepare valuable information to get an overview of some important news or rumors. Automatic stance detection is a strongly motivated mining operation on social media and networks [2]. Stance detection aims to determine the author's point of view (such as a favor, neutral or against) toward a post, a claim or a new one [3]. It has become a key component in many applications such as claims validation, fake news detection, argument search, and author profiling, etc. [4].

The stance detection requires a large amount of labeled data. Research shows that most papers have worked on stance detection in English language [5-7] and that many of the prepared datasets are also in the same language [8-10]. Nevertheless, for stance detection in low data resource languages, it is necessary to use

techniques that are data independent or that may increase the amount of data without generating new labeled data. These techniques are known as data augmentation.

As well, stance detection is an issue of classification. In general, classification is used in many applications such as news topic identification, author detection, etc. But the results of these classifiers are sometimes different, and the samples misclassified by different classifiers are usually not the same in a lot of experiments. This is due to a variety of reasons, such as the use of different training sets, the use of different features, or the use of different parameters to adjust the algorithm used in each of these individual classifiers. Multi-classifiers are typically used to solve this problem. In this paper, we examined the impact of data augmentation methods on the accuracy of detecting Persian stance on social media. In addition, after applying a variety of learning approaches including machine learning, deep learning, and transfer learning, we attempted to fuse their results to determine the final output.

This paper is organized as follows: Section 2 provides an overview of the research. Section

*Corresponding Author Email: toloie@gmail.com
(A. Toloie Eshlaghy)

3 gives a detailed explanation of our approach and the proposed model. Section 4 sets out the results achieved. Finally, Section 5 outlines the findings and future work.

2. RELATED WORKS

In recent studies, stance detection can be categorized into various types:

1) Target-specific stance detection: It tries to detect the stance expressed in a text toward a particular target (e.g., a person, a social movement, an organization, a product, or a policy) [11], which most research has focused on it [12, 13].

2) Multi-target stance detection: It aims to detect the opinions of social media users with respect to two or further targets [10, 14]. Conforti et al. [9] stated that because in numerous applications, there are many natural dependencies between targets, target-specific models are not effective and should be focused on multi-target stance detection.

3) Claim-based stance detection: It is an appropriate technique to investigate the veracity of the news. Its aim to detect the stance in a part of the text or comment toward a claim [15]. Therefore, claim-based stance detection has been extensively used for rumor detection [6, 16, 17].

In recent years, most research has concentrated on social media posts and tweets. In general, the stance detection approaches can be separated into four main kinds [3]: 1) feature-based machine learning approaches which often apply machine learning algorithms like decision tree, logistic regression, SVM¹, etc. for learning [18], 2) deep learning methods which usually use deep neural networks (such as RNN² or LSTM³) [19-23]. Some of the common features used in these approaches are vector representation of words, i.e. Word2Vec and Glove, phrase embedding, n-grams of words or letters, 3) transfer learning which has made significant progress in NLP technology due to the development of large language models using contextualized word embedding based on Transformer architecture [24] and applied by most research [24, 25], and 4) ensemble learning approaches that use more than one classifier to get the final result of the stance detection [13, 20].

In addition, despite the increasing popularity of the stance detection task, almost existing approaches are limited to using the textual features of social media posts, overlooking the social nature of the task. But a limited number of studies focused on contextual features [26, 27].

However, most researches have focused on the English language. Of course, in recent times, many studies have been conducted on other languages other

than English, like Russian [28], Indian [13], Italian [18], Zulu [29], and recently Persian [21, 25].

Since a significant amount of data is needed for automatic stance detection, therefore, in low data resources languages such as the Persian language in which there is not enough labeled data, it is necessary to use approaches to increase data. These techniques are called data augmentation.

Data augmentation methods increase the number of instances in the train data by generating different versions of actual datasets without explicitly collecting new data [30]. Data augmentation methods are designed to increase system efficiency in addition to increasing the data.

The strategy of data augmentation in natural language processing is a complex task because of the inherent complexities of language. We cannot substitute each word with a synonym, and even if we do, the context will be different. An increase in data can take place at various levels: letter level, word level, phrase level, and document level. On the other hand, data augmentation techniques are usually performed in different methods: from rule-based methods [31] to model-based methods [32], which can be very complex. Implementing rule-based methods is much easier, but may not lead to significant improvements. Model-based methods have important effects on performance, but are more challenging to develop and use. On the other hand, the distribution of the generated augmented data should not be too similar or differ too much from the original dataset. Because it can lead to overfitting or poor performance through effective data augmentation approaches that should aim for a balance.

A review of the studies shows that almost all research on stance detection has employed individual classifiers. Numerous studies in other fields have shown that the results of multiple classifiers can outperform better than the best individual classifier and improve the system performance [33]. In other words, when there is high variability among single classifiers, Multiple Classifier Systems (MCSs) can generally achieve greater classification accuracy than any individual classifier [34]. In recent years, many application areas have adopted several methods of merging classifiers, such as object tracking [35], human action recognition [36], risk analysis [37], fault diagnosis [38], face recognition [39] and so on.

3. PROPOSED APPROACH and IMPLEMENTATION DETAILS

In this section, a Persian stance detection approach will be presented. The methodology depicted in Figure 2.

¹ Support Vector Machine

² Recurrent Neural Network

³ Long Short-Term Memory

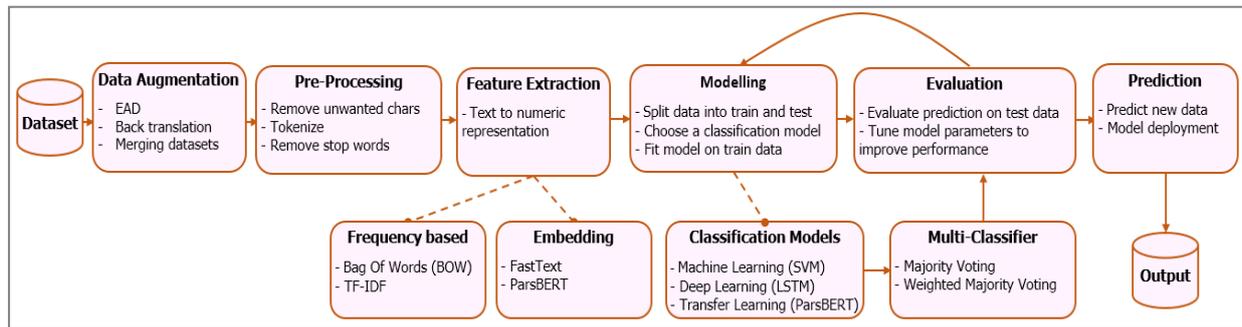


Figure 1. Flow diagram illustrating the methodology of this study

In our approach, data augmentation was used to the original dataset in various ways and the different augmented data were produced, which are illustrated in Figure 3 with letters A to E.

3. 1. Data Augmentation As mentioned above, a large amount of labeled data is needed for stance detection. In languages with low data resources such as Persian language, this amount of data does not exist. Thus, one way to increase the accuracy, enhancement the amount of data by using data augmentation methods.

Data augmentation techniques refer to strategies that enlarge the data in size or amount artificially without explicitly collecting new data [30]. Some of the data augmentation techniques are [40]:

1) Paraphrasing-based methods: these methods try to make minor changes in sentences without changing the semantics of the sentences and enter the changed sentences as new examples into the database; so, the augmented data transfer very similar information as the original sentences.

Back-translation is the most common method in these categories which consists of three steps: 1) each text sample in the dataset is translated to the default language,



Figure 2. An example of the back-translation method

2) translated samples are back-translated to the original language, and 3) duplicate samples are removed from the mixture of the two-source dataset and the created data. This method lets to the production of textual data of different rewording to the real text while keeping the original context and meaning [41]. Figure 2 shows an example of two steps 1 and 2 (in this example, the original data is in English).

3) Noising-based methods: This method focuses to add low noise to the data with a little change so that the meaning of the increased data is very similar to the source data [40]. One of the most common methods in this category is EDA¹, which consists of four simple but powerful operations. These operations include synonym replacement, random insertion, random swapping, and random deletion of words [19].

4) Sampling-based methods: Corresponding to the data distribution, we can add new samples. For example, it is possible to create a larger dataset by merging the original dataset and a similar dataset in another language.

3. 2. Pre-processing Texts published in cyberspace such as posts on social media or web contain a lot of noise. As the performance of machine learning models is dependent upon data quality in addition to the quantity and variety of the data [42, 43], therefore cleaning the data and normalizing them is necessary thing. In this process, after tokenizing the text, sequences such as punctuation marks, numbers, additional spaces, stop words, and undesirable characters were removed in the text.

3. 3. Feature Extraction Since machine learning or deep learning algorithms are only able to understand numerical data instead of textual data, it is necessary to make the text meaningfully for them. Therefore, they must be expressed numerically. For this purpose, some algorithms such as TF-IDF, Word2Vec, etc. enable words to be expressed numerically to solve such problems. So, we used the following two approaches for feature extraction:

¹ Easy Data Augmentation

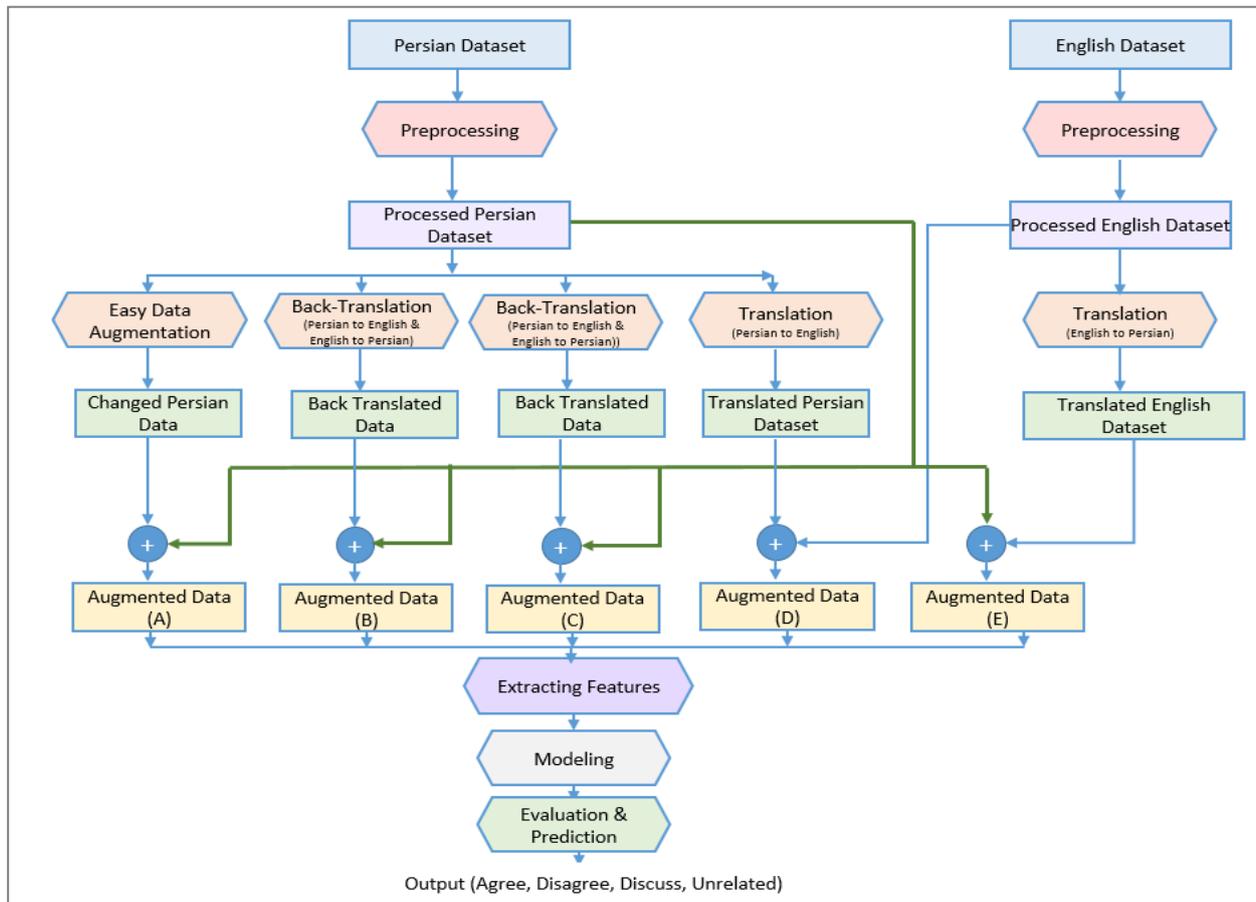


Figure 3. Proposed methodology for data augmentation

3.3.1. Frequency-based In this approach, each word in the text is represented by its frequency as follows:

- Bag of words (BoW): displays the number of occurrences of each bag that is created for each word without considering the word order within the text [44].
- TFIDF: is a statistical measure used to determine the mathematical significance of words in documents [44].

3.3.2. Embedding In word embedding, each word is represented in a continuous vector space. In this space, all words with semantic or syntactical similarity must be placed in the same area [45]. In this paper, we used two pre-trained embeddings as described below:

-FastText embedding: This method is based on the skip-gram model which learns to predict a target word near to the specific word and represents each word as a bag of character n-grams [46-48]. In this approach due to the use of n-gram word tokenization, for misspelled words, unusual words, or words that did not exist in the train data, an embedding is provided. This model is presented by Facebook and learned using Wikipedia 2017, UMBC web base corpus, and statmt.org news

datasets that contains 16 billion tokens. The embedding dimension is 300, the vocabulary is 1 million words [45].

- BERT embedding: Bidirectional Encoder Representations from Transformers (BERT) is another of the strongest document and word representations [47]. It is a transformer that includes an attention mechanism that learns the contextual relationships between words in a given sentence [48]. Under this approach, depending on the context, the same word may have a different embedding. Just like fastText, it is possible to embed rare words. In this paper, we use Pars-BERT that is a Persian language model based on BERT architecture and includes over 3.9 million documents, 73 million sentences, and 1.3 billion words with many writing styles on many topics [49].

3.4. Modeling At this step, the following tasks were carried out:

- Divide the data into train and test: in this regard, 80% of the data is considered to be train data and the rest as test data. As well, we used k-fold cross-validation and set k=10. On the other hand, since the samples are unbalanced, i.e., the amount of instances per class is not

equal, the stratifiedKfold library in Python has been used to shuffle the data in a balanced way.

- Select classification models: three different approaches were used for modeling: machine learning (such as SVM, Decision Tree, logistic regression, etc.), deep learning (LSTM), and transfer learning (ParsBERT).

- Fit models on training data: In this step, we trained our models on train data, that is, we passed the data into the model so that the model can update its internal mathematical variables and be prepared to predict. In other words, during fitting, we can pass various parameters like batch-size, epochs, learning rate, etc.

3. 5. Multi-classifier fusion

In this step, we suggested a multi-classifier fusion model for Persian stance detection. As mentioned previously, the results of individual classifiers, because of the algorithm, the results obtained from them are different. Multiple classifiers therefore combine several individual classifiers for better results. Multiple classifiers use different methods to fuse the results of each classifier, but majority voting is the most popular approach where each classifier "vote" for a particular class, and the class with the most votes is predicted by the multi-classifier system [50]. Under the majority voting approach, all individual classifiers have the same "authority" to classify correctly no matter how well they perform [51]. To solve this problem, weighting methods have been suggested, which are more appropriate to solve the problems that the member's classifiers perform the same task. In this approach, the output of each classifier is usually weighted according to its calculated accuracy on the train-data [52].

In this paper, we tried to use traditional majority and weighted majority voting methods to achieve the final results. Figure 4 shows the proposed model of the stance detector based on multi-classifier fusion. As described in this figure, first the texts in the dataset are preprocessed and then the desired features are extracted. In the next step, these features are used in the relevant classifier. As we know, in the machine learning approach, the SVM algorithm is used because it performs better (see 4.C in this article). In the deep learning approach, the LSTM is used and finally, in the transfer learning approach, the Pars-BERT transformer is used. In the next step, the results of these three classifiers are fused using both majority and weighted majority voting methods, and finally returns a specific stance as the final result.

3. 6. Evaluation For evaluating the performance of our approaches, we use Accuracy and F1-measure. Accuracy measures the proportion of the number of correct predictions relative to the total samples and F1-

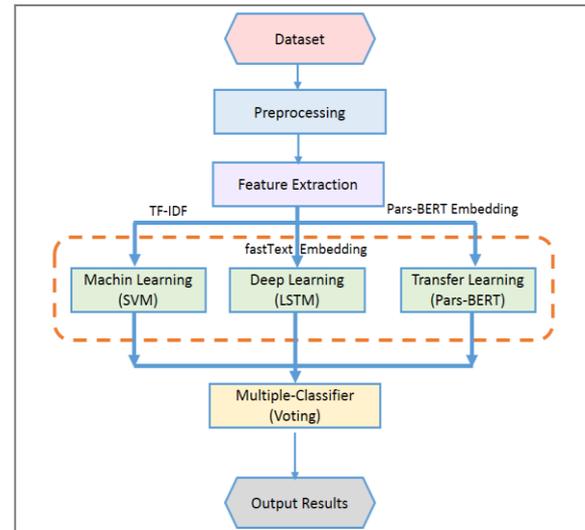


Figure 4. The Proposed model for multiple classifier fusion

measure determines the harmonic mean of precision and recall and is calculated as Equation (1):

$$F - Measure = \frac{2 * Precision * Recall}{Precision + Recall} \quad (1)$$

3. 7. Prediction

Once the model is trained on data, it should be capable of predicting new data. In this respect, we obtained 20% of the data set as test data and applied the relevant algorithms to it. Test results are given in the following section.

4. EXPERIMENTAL RESULTS

4. 1. Dataset

In the current research, we used two datasets in Persian and English languages. In the following we describe each of them:

4. 1. 1. Persian Dataset¹

This dataset includes 534 claims gathered from Shayeaat² and Fakenews³ and includes of two parts [21]: the first part consists of claims with news headlines and the second one includes claims with the body text of the articles. The labels of each news headline or article's body are:

- Agree: The article expresses that the claim is right, without any kind of cover and quotation.
- Disagree: The article expresses that the claim is wrong, without any kind of cover and quotation.
- Discuss: The article does not get any argument about the right or wrong of the claim.
- Unrelated: The claim is not reported in the article.

The first part of this data set, which contains the pair (news headline, claim), includes 2029 examples, and the

¹ <https://github.com/majidzarharan/persian-stanceclassification>

² Shayeaat.ir

³ Fakenews.ir

second part, which contains the pair (article's body, claim), includes 1997 examples. Table 1 shows the distribution of labels in each section.

4. 1. 2. English Dataset This dataset is presented in SemEval-2017-Task8 and contains 297 rumors -which are collected around 8 events taken from the urgent news- along with 5271 response tweets, which is a total of 5568 pairs (tweets and tweet responses). This dataset is separated into two parts, training data, and test data. The tag set distributions in each of the parts of this data provided in Table 2.

This dataset used the tree-based conversation consisting of tweets that replied to the rumor tweet, directly or indirectly [53]. The labels of stances in this dataset are Support, Deny, Query, and Comment (SDQC). Therefore, this dataset aims to detect the stance of reply tweets toward a rumor tweet (that can be direct or nested responses). Figure 5 presents an example of the tree structure for tweets. In this figure, user1 and user3 directly respond to user0's tweet, but user2 has expressed his opinion in response to user1's post.

4. 2. Results of Applying Data Augmentation Techniques

First, we present experimental results on the Persian dataset (without increased data) in two parts (1, 2) in Table 3. Next, we look at the effects of the application of each of the data augmentation methods on the performance of the algorithm used.

4. 2. 1. Easy Data Augmentation (EDA) Karande et al. [26] proposed that all operations of the EDA technique were examined on the Persian dataset. It has shown that a combination of such operations would be more appropriate to these data. So, in this paper, we used the combination of these operations on the Persian dataset. The experimental findings are given in Table 4.

TABLE 1. Distribution of labels in the Persian dataset

	Agree	Disagree	Discuss	Unrelated	Total
Part 1 (Headline-Claim Stance)	405	164	802	658	2029
Part 2 (Article-Claim Stance)	137	206	1068	586	1997

TABLE 2. Distribution of labels in the English dataset

	Support	Deny	Query	Comment
Train	910	344	358	2907
Test	94	71	106	778

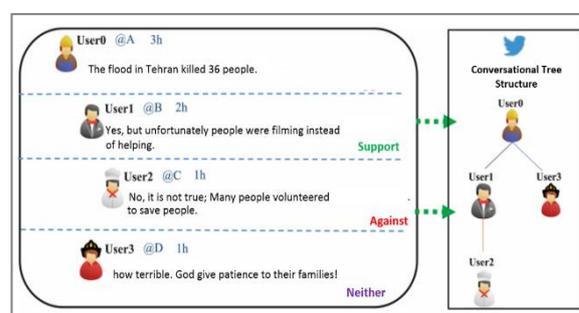


Figure 5. The tree structure of social media conversational

4. 2. 2. Back-translation For the back-translated data augmentation, in the current study, we have considered English as the target language. Table 5 provides the results of this approach.

4. 2. 3. Merging the Persian Dataset and English Dataset To accomplish this, we have implemented the following steps:

TABLE 3. Results on the original dataset

Data	Size	Feature	Accuracy	Precision	Recall	F-Measure
Part 1 (Headline-Claim)	2029	BOW	0.41	0.40	0.41	0.40
		TFIDF	0.53	0.53	0.53	0.53
Part 2 (Article-Claim)	1997	BOW	0.55	0.50	0.55	0.51
		TFIDF	0.83	0.83	0.83	0.83

TABLE 4. Results of EDA on the Persian dataset

Data	Size	Feature	Accuracy	Precision	Recall	F-Measure
Part 1 (Headline-Claim)	4058	BOW	0.52	0.52	0.51	0.52
		TFIDF	0.77	0.77	0.77	0.76
Part 2 (Article-Claim)	3994	BOW	0.63	0.61	0.63	0.61
		TFIDF	0.80	0.80	0.80	0.79

TABLE 5. Results of back-translation on the Persian dataset

Data	Size	Feature	Accuracy	Precision	Recall	F-Measure
Part 1 (Headline-Claim)	3998	BOW	0.49	0.48	0.49	0.48
		TFIDF	0.67	0.67	0.67	0.66
Part 2 (Article-Claim)	4058	BOW	0.46	0.45	0.46	0.45
		TFIDF	0.62	0.62	0.62	0.62

- First, we reviewed the stance detection datasets and chose one that was similar to Persian dataset. The sense of similarity here is that, first, both datasets were developed for the same purpose (here, to stance detection of a text, news, or a tweet reply toward to a claim or tweet). Second, the two datasets have the same tags or, if differentiated, the tags can be mapped to each other.

Some research was carried out on the preparation of dataset for stance detection in Persian. Tutek et al. [21] did the only work which was explained in part A of this section. One of the English datasets that may be regarded as equivalent is the one that was published in SemEval 2017-task 8. Its type is based on claims which its labels can be mapped to each other. Explanations of the English dataset can be found in part A.2 of this section.

- As the Persian dataset does not reflect the tweet tree conversation, we also tried to select only the top level of the English dataset tree structure. Thus, the size of our English dataset has been reduced to 3272 instances.

- At the next step, the English dataset labels were mapped to the Persian dataset as follows:

Agree \approx *Support*
Disagree \approx *Deny*
Discuss \approx *Query*
Unrelated \approx *Comment*

- The final step is to translate the English dataset into Farsi and add it to the Persian dataset. The experimental findings for augmented data are shown in Table 6.

4. 2. 4. Comparison of Data Augmentation Methods

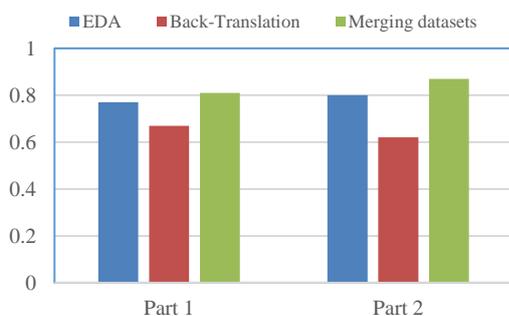
Figure 6 presents the results for each method of data augmentation on original dataset. The results indicate that the best method to increase the quality of stance detection performance is to merge the original dataset with a similar dataset in another language, which increases the diversity of data. If such a dataset is not found in other languages or is not accessible, the next best method is EDA method, which also shows a good improved algorithm performance. Back translation method though increased the accuracy, but compared to the other two methods, it does less improvement in the algorithm.

4. 3. Results of Applying Different Learning Approaches

As discussed above, we used three learning approaches: 1) machine learning, 2) deep learning, and 3) transfer learning. The results of the implementation of each of these methods are given

TABLE 6. Results on the augmented dataset

Data	Size	Feature	Accuracy	Precision	Recall	F-Measure
Translated English dataset + part1	5301	BOW	0.77	0.79	0.77	0.78
		TFIDF	0.81	0.81	0.80	0.81
Translated English dataset + part2	5269	BOW	0.66	0.64	0.66	0.63
		TFIDF	0.87	0.88	0.87	0.87

**Figure 6.** Comparison of data augmentation methods

below.

4. 3. 1. Machine Learning Approach

In this step, we used different supervised algorithms for stance detection in two parts of the original dataset. In Table 7, the relevant algorithms are given along with the results obtained from each. As the results show, the SVM algorithm shows a higher accuracy than other algorithms.

For the augmented dataset (for example the D dataset), the same algorithms were implemented and the results in Table 8 demonstrated that, again, SVM performed best.

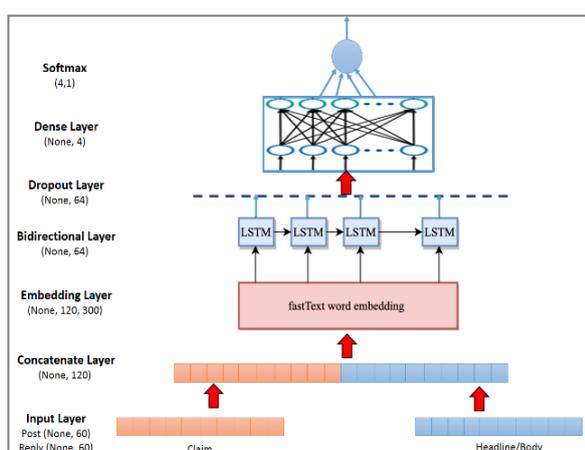
TABLE 7. Results of machine learning algorithms on original dataset

	Part 1 (Headline-Claim)				Part 2 (Article-Claim)			
	BOW		TFIDF		BOW		TFIDF	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Logistic Regression	0.40	0.39	0.53	0.52	0.51	0.47	0.64	0.61
Decision Tree	0.41	0.40	0.52	0.53	0.49	0.48	0.44	0.44
SVM	0.40	0.38	0.53	0.53	0.51	0.47	0.65	0.63
Random Forest	0.40	0.39	0.51	0.52	0.49	0.47	0.50	0.46
KNN	0.39	0.36	0.40	0.38	0.49	0.45	0.49	0.46
Ada-boost	0.43	0.38	0.50	0.46	0.52	0.39	0.52	0.42

TABLE 8. Results of machine learning algorithms on augmented dataset

	Part 1 (Headline-Claim)				Part 2 (Article-Claim)			
	BOW		TFIDF		BOW		TFIDF	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
Logistic Regression	0.77	0.78	0.80	0.80	0.74	0.70	0.85	0.84
Decision Tree	0.78	0.77	0.81	0.81	0.81	0.80	0.80	0.80
SVM	0.77	0.78	0.82	0.82	0.83	0.83	0.86	0.86
Random Forest	0.77	0.78	0.81	0.81	0.81	0.81	0.83	0.82
KNN	0.73	0.70	0.74	0.70	0.76	0.73	0.81	0.79
Ada-boost	0.60	0.61	0.72	0.67	0.72	0.69	0.69	0.68

4. 3. 2. Deep Learning Approach In this phase, LSTM deep neural network classifiers were used. LSTM is an RNN model which overcomes the vanishing gradient and is used to model sequential data tasks. It is capable of efficiently capturing long-range dependencies. The designed network architecture is presented in Figure 7.

**Figure 7.** The schematic of our deep learning model

Since in our dataset, a tag is assigned for each pair of (claim and news headline) or (claim and news body), therefore, the neural network here takes two inputs. In the second layer, the two inputs are concatenated. After that, it enters the embedding layer, which utilizes fastText embedding. Then there is the bi-directional layer in which LSTM is used. It is followed by a fully-connected network to map the outputs to the tag space. For optimizing the model, Adam optimizer [17] is used for 20 epochs. Batch size and embedding-dimension are respectively 16 and 300. We used the TensorFlow library [54] to implement this model. The hyper-parameters have been tuned by evaluation over the validation set to achieve the highest accuracy and F1-measurement. Table 9 provides the test results of applying this model to our primary and augmented dataset. As the results show, when the volume of the dataset increases, the algorithm can make predictions with higher accuracy.

4. 3. 3. Transfer Learning Approach In this paper, we used pre-trained BERT models to for applying transfer learning. The Pars-BERT model can be fine-tuned to a specific task. It involves matching the parameters of a pre-trained BERT model for a particular

TABLE 9. Results of deep learning model

	Part 1 (Headline-Claim)		Part 2 (Article-Claim)	
	Acc.	F1	Acc.	F1
Pars-BERT (Original dataset)	0.72	0.73	0.75	0.74
Pars-BERT (Augmented dataset)	0.95	0.94	0.97	0.96

task by using a low data resource [45]. Figure 8 presents the suggested network architecture.

Like the previous model, here we have two inputs for the neural network too. We used the base model of Pars-BERT and fine-tuned it using the stance detection corpus. Then it is followed by a fully-connected network to map the Pars-Bert's outputs to the tag space. The learning rate, batch size, and the number of epochs are set to $5e^{-05}$, 128 and 10, respectively. In addition, the epsilon is set to $1e^{-08}$. Adam was applied for optimizing the model.

Table 10 shows the test results of applying this model on our original and augmented dataset. As the results show, when the volume of the dataset increases, the algorithm can make predictions with higher accuracy.

4. 4. Results of Applying Multi Classifier Fusion

In this section, the results of the empirical test of the proposed model shown in Figure 3 are presented on the original dataset and augmented dataset. For a better comparison, Tables 11 and 12 show the test results of the individual classifiers first, and then the performance of the use of the multi-classifier fusion. Clearly, among individual classifiers, the use of transfer learning methods can produce good results. On the other hand, the findings show that the combination of classifiers may lead to a significant improvement. The amount of this improvement is higher when weighted majority voting is

TABLE 10. Results of the transfer learning model

	Part 1 (Headline-Claim)		Part 2 (Article-Claim)	
	Acc.	F1	Acc.	F1
LSTM (original dataset)	0.69	0.67	0.72	0.71
LSTM (augmented dataset)	0.86	0.87	0.94	0.94

TABLE 11. Comparison of results of multi-classifier versus individual classifier in the original dataset

Algorithm	Dataset	Part 1 (Headline-Claim)		Part 2 (Article-Claim)	
		Acc.	F1	Acc.	F1
Individual Classifier					
SVM		0.53	0.53	0.65	0.63
LSTM		0.69	0.67	0.72	0.71
Pars-BERT		0.72	0.73	0.75	0.74
Multiple-Classifer					
Majority Voting		0.76	0.77	0.78	0.77
Weighted Majority Voting		0.78	0.77	0.79	0.79

TABLE 12. Comparison of results of multi-classifier versus individual classifier in augmented data

Algorithm	Dataset	Part 1 (Headline-Claim)		Part 2 (Article-Claim)	
		Acc.	F1	Acc.	F1
Individual Classifier					
SVM		0.81	0.81	0.87	0.87
LSTM		0.86	0.87	0.94	0.94
Pars-BERT		0.95	0.94	0.97	0.96
Multiple-Classifer					
Majority Voting		0.95	0.95	0.98	0.97
Weighted Majority Voting		0.97	0.96	0.99	0.99

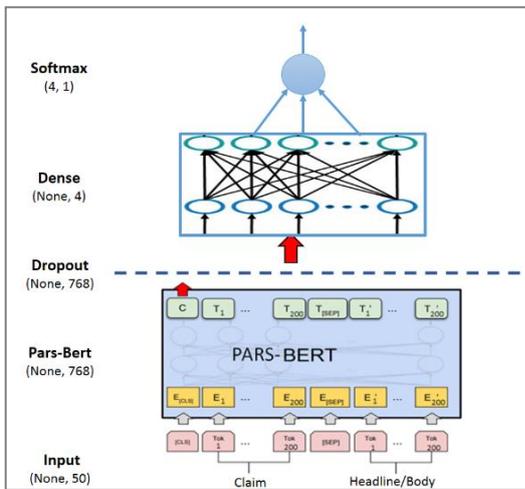


Figure 8. The schematic of our transfer leaning model

used. because in this case, it considers more weight to the classifier that has greater accuracy and increases the likelihood of its effectiveness in making the final decision.

As the results show, the use of transfer learning methods on the original data returns relatively good results. When the amount of data increases, the use of deep and transfer learning methods can make predictions with good accuracy and seek satisfaction. But as the above tables show, the use of multi-classifiers can perform better than the mentioned methods and provide quite acceptable accuracy. Therefore, according to the conditions, the following points can be considered:

- To improve the position detection accuracy, it is better to use similar data in another language, otherwise the EDA method is suitable.

- If similar data is not found in another language, it is better to use transfer learning methods.

- In any case, the use of multi-classifiers can lead to a good improvement.

4. 5. Comparison and Discussion As mentioned earlier by Tutek et al. [21] and Vaswani et al. [24] performed similar work in the field of Persian stance detection. Tutek et al. [21] used LSTM and Vaswani et al. [24] applied transfer learning and data augmentation on the dataset are discussed in this paper.

Table 13 compares the proposed model to the best models. As can be seen, multi-classifier fusion has a considerable impact on improving stance detection tasks

There are various advantages to the proposed model. Some of them are discussed below:

- It detects the stance by having only the content of a post and without extracting more features. This saves time and reduces computational costs.

-The size of the dataset can be increased without collecting more data and labelling them, which requires time, money and human resources.

- Combining the results of different classifier optimizes the use of the individual abilities and creates synergy.

-The proposed model is flexible in that other classifications may be used according to the subject and the intended application.

Unfortunately, no model is without disadvantages. Given below are some of the disadvantages of the proposed model:

- May not always have a dataset similar to the original dataset in another language. Of course, in this case, other methods of data augmentation should be used.

- It only uses content features to find the position. When interfacing with social networks, it is important to know which account answered the tweet. For example, the person who answers may be against the author of the tweet and thus reject the claims of that person, or it may be a bot, so user profiling can be a better help to identifying the stance.

TABLE 13. Comparison of the proposed model to the latest models presented for Persian stance detection

Model	Headline-Claim		Article-Claim	
	Accuracy	F1	Accuracy	F1
[21]	0.67	0.67	0.72	0.71
[24]	0.75	0.75	0.76	0.76
The proposed model (original dataset)	0.78	0.77	0.79	0.79
The proposed model (augmented data)	0.97	0.96	0.99	0.99

5. CONCLUSION AND FUTURE WORKS

In this paper first, we attempted to apply several data augmentation techniques to overcome the problem of the absence of data in stance detection in low data resource languages and to analyze the impact of each in improving the performance of algorithm. In this regard, we investigated Persian claimed-based stance detection and used various methods of data augmentation (including EDA, back translation, and merging similar datasets). The test results showed that if we can merge the source dataset with the similar dataset into other languages and create a bigger dataset, we achieve a significant improvement without spending time, money, and human resource in collecting data and labeling them. If such a dataset is unavailable or does not exist, a good improvement can be obtained by using the EDA technique.

Also, we proposed a model based on multi-classifier fusion for Persian stance detection, where, in addition to using different approaches such as machine learning, deep learning, and transfer learning to detection, the fusion of these classifiers' results is used to make the final decision. For this purpose, majority voting has been used and the results have shown that multi-classifier fusion can yield better results than the best classifier and improve the performance. Consequently, it may be concluded that multiple classification systems (MCS) are a good approach for stance detection. If we consider the accuracy of the classifiers as an efficient factor in the final decision to detect the Persian stance, we can achieve a significant improvement in the results.

So, our medal is innovative from two perspectives:

1. Use of the data augmentation method by combining the primary dataset and similar dataset in English and creating a larger dataset

2. Fusion of the individual classifier results through multi-classifiers.

Finally, the proposed model was also compared with the latest models presented in the field of Persian stance detection. While the proposed model has a significant improvement over related works, it also has limitations, which we will mention in the following and the future work that can be done to solve them.

As mentioned earlier, one of the limitations of this model is the use of content-only features, while contextual features such as user features and so on can also be used. Therefore, user profiling is one of the important tasks that can be addressed going forward.

Another thing that can be done to improve the model is to analyze the sentiment contained in the reply text of a tweet. For example, if someone answers "I'm sorry" to the tweet "Sanctions against Iran are increasing daily", analyzing the sentiment of the reply text can help identify the position.

In addition, one other thing that can be done to

improve the model is the detection of irony. Sometimes in social media spaces, people reply to a tweet with sarcasm, as its stance cannot be easily recognized. For example, if for the tweet "inflation has gone down by 20% this year", someone answers "don't get tired" It is unclear whether the stance of replier is to agree or disagree with that tweet. Therefore, the sarcasm detection module may also contribute to stance detection and be one of the tasks to focus on in the future.

6. REFERENCES

- Sobhani, P., "Stance detection and analysis in social media", Universite d'Ottawa/University of Ottawa, (2017),
- 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., (2022), 241-251.
- Küçük, D. and Can, F., "A tutorial on stance detection", in Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining., (2022), 1626-1628.
- Schiller, B., Daxenberger, J. and Gurevych, I., "Stance detection benchmark: How robust is your stance detection?", *KI-Künstliche Intelligenz*, (2021), 1-13. doi: 10.1007/s13218-021-00714-w.
- Kochkina, E., Liakata, M. and Augenstein, I., "Turing at semeval-2017 task 8: Sequential approach to rumour stance classification with branch-1stm", *arXiv preprint arXiv:1704.07221*, (2017). doi: 10.18653/v1/S17-2083.
- Yuan, C., Qian, W., Ma, Q., Zhou, W. and Hu, S., "Srlf: A stance-aware reinforcement learning framework for content-based rumor detection on social media", in 2021 International Joint Conference on Neural Networks (IJCNN), IEEE., (2021), 1-8.
- Zubiaga, A., Liakata, M., Procter, R., Wong Sak Hoi, G. and Tolmie, P., "Analysing how people orient to and spread rumours in social media by looking at conversational threads", *PLoS one*, Vol. 11, No. 3, (2016), e0150989. doi: 10.1371/journal.pone.0150989.
- Mohammad, S., Kiritchenko, S., Sobhani, P., Zhu, X. and Cherry, C., "Semeval-2016 task 6: Detecting stance in tweets", in Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), (2016), 31-41.
- Conforti, C., Berndt, J., Pilehvar, M.T., Giannitsarou, C., Toxvaerd, F. and Collier, N., "Will-they-won't-they: A very large dataset for stance detection on twitter", *arXiv preprint arXiv:2005.00388*, (2020). doi: 10.18653/v1/2020.acl-main.157.
- Sobhani, P., Inkpen, D. and Zhu, X., "A dataset for multi-target stance detection", in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers., (2017), 551-557.
- Du, J., Xu, R., He, Y. and Gui, L., "Stance classification with target-specific neural attention networks", International Joint Conferences on Artificial Intelligence., (2017).
- Lai, M., Cignarella, A.T., Fariás, D.I.H., Bosco, C., Patti, V. and Rosso, P., "Multilingual stance detection in social media political debates", *Computer Speech & Language*, Vol. 63, (2020), 101075. doi: 10.1016/j.csl.2020.101075.
- Zotova, E., Agerri, R., Nuñez, M. and Rigau, G., "Multilingual stance detection in tweets: The catalonia independence corpus", in Proceedings of the Twelfth Language Resources and Evaluation Conference., (2020), 1368-1375.
- Swami, S., Khandelwal, A., Singh, V., Akhtar, S.S. and Shrivastava, M., "An english-hindi code-mixed corpus: Stance annotation and baseline system", *arXiv preprint arXiv:1805.11868*, (2018). doi: arXiv:1805.11868v1
- Darwish, K., Magdy, W. and Zanouada, T., "Trump vs. Hillary: What went viral during the 2016 us presidential election", in Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part I 9, Springer. (2017), 143-161.
- Li, Y., He, H., Wang, S., Lau, F. and Song, Y., "Improved target-specific stance detection on social media platforms by delving into conversation threads", *arXiv preprint arXiv:2211.03061*, (2022). doi: arXiv:2211.03061.
- Derczynski, L., Bontcheva, K., Liakata, M., Procter, R., Hoi, G.W.S. and Zubiaga, A., "Semeval-2017 task 8: Rumoureal: Determining rumour veracity and support for rumours", *arXiv preprint arXiv:1704.05972*, (2017). doi: 10.18653/v1/S17-2006.
- Bar-Haim, R., Bhattacharya, I., Dinuzzo, F., Saha, A. and Slonim, N., "Stance classification of context-dependent claims", in Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers., (2017), 251-261.
- Cignarella, A.T., Lai, M., Bosco, C., Patti, V. and Paolo, R., "Sardistance@ evalita2020: Overview of the task on stance detection in italian tweets", in Proceedings of the Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2020), Ceur., (2020), 1-10.
- Wei, P., Lin, J. and Mao, W., "Multi-target stance detection via a dynamic memory-augmented network", in The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval., (2018), 1229-1232.
- Tutek, M., Sekulić, I., Gombar, P., Paljak, I., Čulinović, F., Boltužić, F., Karan, M., Alagić, D. and Šnajder, J., "Takelab at semeval-2016 task 6: Stance classification in tweets using a genetic algorithm based ensemble", in Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016), (2016), 464-468.
- Zarharan, M., Ahangar, S., Rezvaninejad, F.S., Bidhendi, M.L., Pilevar, M.T., Minaei, B. and Eetemadi, S., "Persian stance classification data set", in TTO., (2019).
- Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S. and On, B.-W., "Fake news stance detection using deep learning architecture (cnn-1stm)", *IEEE Access*, Vol. 8, (2020), 156695-156706. doi: 10.1109/ACCESS.2020.3019735.
- Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A.N., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N. and Parmar, N., "Tensor2tensor for neural machine translation", *arXiv preprint arXiv:1803.07416*, (2018). doi: arXiv:1803.07416.
- Nasiri, H. and Analoui, M., "Persian stance detection with transfer learning and data augmentation", in 2022 27th International Computer Conference, Computer Society of Iran (CSICC), IEEE. (2022), 1-5.
- Karande, H., Walambe, R., Benjamin, V., Kotecha, K. and Raghu, T., "Stance detection with bert embeddings for credibility analysis of information on social media", *PeerJ Computer Science*, Vol. 7, (2021), e467. doi: 10.7717/peerj-cs.467.
- Khiabani, P.J. and Zubiaga, A., "Few-shot learning for cross-target stance detection by aggregating multimodal embeddings", *arXiv preprint arXiv:2301.04535*, (2023). doi: arXiv:2301.04535.
- Ren, Y., Liu, Y., Chen, J., Guo, X., Shi, J. and Jia, M., "News stance discrimination based on a heterogeneous network of social background information fusion", *Entropy*, Vol. 25, No. 1, (2022), 78. doi: 10.3390/e25010078.
- Lozhnikov, N., Derczynski, L. and Mazzara, M., "Stance prediction for russian: Data and analysis", in Proceedings of 6th

- International Conference in Software Engineering for Defence Applications: SEDA 2018 6, Springer., (2020), 176-186.
30. Dlamini, G., Bekkouch, I.E.I., Khan, A. and Derczynski, L., "Bridging the domain gap for stance detection for the zulu language", in Intelligent Systems and Applications: Proceedings of the 2022 Intelligent Systems Conference (IntelliSys) Volume 1, Springer., (2022), 312-325.
 31. Feng, S.Y., Gangal, V., Wei, J., Chandar, S., Vosoughi, S., Mitamura, T. and Hovy, E., "A survey of data augmentation approaches for nlp", *arXiv preprint arXiv:2105.03075*, (2021). doi: 10.18653/v1/2021.findings-acl.84.
 32. Yin, F., Yao, Z. and Liu, J., "Character-level attention convolutional neural networks for short-text classification", in Human Centered Computing: 5th International Conference, HCC 2019, Čačak, Serbia, August 5–7, 2019, Revised Selected Papers 5, Springer., (2019), 560-567.
 33. Liu, R., Xu, G., Jia, C., Ma, W., Wang, L. and Vosoughi, S., "Data boost: Text data augmentation through reinforcement learning guided conditional generation", *arXiv preprint arXiv:2012.02952*, (2020). doi: 10.18653/v1/2020.emnlp-main.726.
 34. Mi, A., Wang, L. and Qi, J., "A multiple classifier fusion algorithm using weighted decision templates", *Scientific Programming*, Vol. 2016, (2016). doi: 10.1155/2016/3943859.
 35. Liang, S., Han, D. and Han, C., "A novel diversity measure based on geometric relationship and its application to design of multiple classifier systems", *Acta Automatica Sinica*, Vol. 40, No. 3, (2014), 449-458. doi: 10.3724/SP.J.1004.2014.00449.
 36. Chen, S., Lu, X., Chen, X., Chen, M., Chen, J., Wang, D. and Zhu, S., "Object tracking with multi-classifier fusion based on compressive sensing and multiple instance learning", *Mathematical Problems in Engineering*, Vol. 2020, (2020), 1-17. doi: 10.1155/2020/1574054.
 37. Bagheri, M.A., Hu, G., Gao, Q. and Escalera, S., "A framework of multi-classifier fusion for human action recognition", in 2014 22nd International Conference on Pattern Recognition, IEEE., (2014), 1260-1265.
 38. Pan, Y., Zhang, L., Wu, X. and Skibniewski, M.J., "Multi-classifier information fusion in risk analysis", *Information Fusion*, Vol. 60, (2020), 121-136. doi: 10.1109/ICPR.2014.226.
 39. Huang, J.-T., Wang, M.-H., Li, W.-J. and Gu, B., "Multiple classifier fault diagnosis system based on dynamic weight", *Acta Electronica Sinica*, Vol. 40, No. 4, (2012), 734. doi: 10.3969/j.issn.0372-2112.2012.04.018.
 40. Yu, Z., Nam, M.Y., Sedai, S. and Rhee, P.K., "Evolutionary fusion of a multi-classifier system for efficient face recognition", *International Journal of Control, Automation and Systems*, Vol. 7, (2009), 33-40. doi: 10.1007/s12555-009-0105-z.
 41. Li, B., Hou, Y. and Che, W., "Data augmentation approaches in natural language processing: A survey", *AI Open*, Vol. 3, (2022), 71-90. doi: 10.1016/j.aiopen.2022.03.001.
 42. 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*, Vol. 24, (2021), 100153. doi: 10.1016/j.osnem.2021.100153.
 43. Maharana, K., Mondal, S. and Nemade, B., "A review: Data pre-processing and data augmentation techniques", *Global Transitions Proceedings*, (2022). doi: 10.1016/j.gltp.2022.04.020.
 44. Qader, W.A., Ameen, M.M. and Ahmed, B.I., "An overview of bag of words; importance, implementation, applications, and challenges", in 2019 international engineering conference (IEC), IEEE., (2019), 200-204.
 45. Aizawa, A., "An information-theoretic perspective of tf-idf measures", *Information Processing & Management*, Vol. 39, No. 1, (2003), 45-65. doi: 10.1016/S0306-4573(02)00021-3.
 46. d'Sa, A.G., Illina, I. and Fohr, D., "Bert and fasttext embeddings for automatic detection of toxic speech", in 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies"(OCTA), IEEE., (2020), 1-5.
 47. Joulin, A., Grave, E., Bojanowski, P., Douze, M., Jégou, H. and Mikolov, T., "Fasttext. Zip: Compressing text classification models", *arXiv preprint arXiv:1612.03651*, (2016). doi: arXiv:1612.03651.
 48. Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K., "Bert: Pre-training of deep bidirectional transformers for language understanding", *arXiv preprint arXiv:1810.04805*, (2018). doi: arXiv:1810.04805.
 49. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., "Attention is all you need", *Advances in Neural Information Processing Systems*, Vol. 30, (2017). doi: 10.48550/arXiv.1706.03762.
 50. Farahani, M., Gharachorloo, M., Farahani, M. and Manthouri, M., "Parsbert: Transformer-based model for persian language understanding", *Neural Processing Letters*, Vol. 53, (2021), 3831-3847. doi: 10.1007/s11063-021-10528-4.
 51. Kuncheva, L.I., "Combining pattern classifiers: Methods and algorithms, John Wiley & Sons, (2014).
 52. Moreno-Seco, F., Inesta, J.M., De León, P.J.P. and Micó, L., "Comparison of classifier fusion methods for classification in pattern recognition tasks", in Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2006 and SPR 2006, Hong Kong, China, August 17-19, 2006. Proceedings, Springer., (2006), 705-713.
 53. Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y. and Liu, S., "Multiple classifier system for remote sensing image classification: A review", *Sensors*, Vol. 12, No. 4, (2012), 4764-4792. doi: 10.3390/s120404764.
 54. Kingma, D.P. and Ba, J., "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, (2014). doi: arXiv:1412.6980.

Persian Abstract

چکیده

تشخیص موضوع یک موضوع تحقیقاتی اخیر است که به یک پارادایم نوظهور نظیر کاوی با اهمیت تبدیل شده است که هدف آن تعیین دیدگاه نویسنده نسبت به یک موضوع یا ادعای خاص است. تشخیص موضوع به یک مؤلفه کلیدی در برنامه های مانند تشخیص اخبار جعلی، اعتبار سنجی ادعا، جستجوی استدلال، پروفایل نویسنده تبدیل شده است. با وجود پیشرفت قابل توجهی که در این زمینه در زبان های مانند انگلیسی صورت گرفته است، متأسفانه در برخی از زبان ها مانند فارسی که با کمبود دیتاست در این زمینه مواجه هستیم، پیشرفت خوبی نداشته ایم. در این مقاله از دو راه حل برای حل این مشکل استفاده شده است: (۱) استفاده از افزایش داده ها و (۲) اعمال رویکردهای مختلف یادگیری (مانند یادگیری ماشینی، یادگیری عمیق و یادگیری انتقال) و ترکیب معنی دار نتایج آنها. نتایج نشان می دهد که هر یک از این راه حل ها نه تنها می توانند به تنهایی منجر به بهبود عملکرد تشخیص موضوع شوند، بلکه زمانی که هر دو با هم ترکیب شوند، بهبود بسیار قابل توجهی در نتایج حاصل می شود.
