Research Paper

JIS

Journal of Information Science
1–14

DOI: 10.1177/0165551520962781
journals.sagepub.com/home/jis
SAGE

Deep Persian sentiment analysis: Cross-lingual training for low-resource languages

Rouzbeh Ghasemi
Computer Engineering Department, Amirkabir University of Technology, Iran

Seyed Arad Ashrafi Asli
Computer Engineering Department, Amirkabir University of Technology, Iran

Saeedeh Momtazi
Computer Engineering Department, Amirkabir University of Technology, Iran

Abstract
With the advent of deep neural models in natural language processing tasks, having a large amount of training data plays an essential role in achieving accurate models. Creating valid training data, however, is a challenging issue in many low-resource languages. This problem results in a significant difference between the accuracy of available natural language processing tools for low-resource languages compared with rich languages. To address this problem in the sentiment analysis task in the Persian language, we propose a cross-lingual deep learning framework to benefit from available training data of English. We deployed cross-lingual embedding to model sentiment analysis as a transfer learning model which transfers a model from a rich-resource language to low-resource ones. Our model is flexible to use any cross-lingual word embedding model and any deep architecture for text classification. Our experiments on English Amazon dataset and Persian Digikala dataset using two different embedding models and four different classification networks show the superiority of the proposed model compared with the state-of-the-art monolingual techniques. Based on our experiment, the performance of Persian sentiment analysis improves 22% in static embedding and 9% in dynamic embedding. Our proposed model is general and language-independent; that is, it can be used for any low-resource language, once a cross-lingual embedding is available for the source–target language pair. Moreover, by benefitting from word-aligned cross-lingual embedding, the only required data for a reliable cross-lingual embedding is a bilingual dictionary that is available between almost all languages and the English language, as a potential source language.

Keywords
Convolutional neural network; cross-lingual embedding; long short-term memory network; low-resource languages; sentiment analysis

1. Introduction

In recent years, there was a massive rise in the amount of raw data produced by online businesses, and especially by their users. Such data with no analysis, however, are invaluable for these business owners. By analysing their users' behaviour, they could obtain precious information about the users' needs, thoughts, reactions and ideas. Sentiment analysis is one of the major tasks in natural language processing applications which has been focused by researchers to benefit from this task in real-life scenarios of online businesses.

Many new businesses formed around sentiment analysis applications, such as social media monitoring and user review analysis. These businesses gain their capital by providing analytical reports to other companies. These kinds of analyses provide companies many insights and advantages to their opponents in the market, such as finding current trends and

Corresponding author:
Saeedeh Momtazi, Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran.
Email: momtazi@aut.ac.ir

users' opinions about a certain occasion. Hence, they can make well-informed decisions and strategies in the market to secure their profit. Another great application of sentiment analysis forms around customers' feedback. By building a sentiment analysis model with current data, companies will gain the power to predict users' reactions to changes or new ideas.

A variety of models have been examined in the sentiment analysis task in the past. Both supervised and unsupervised techniques have been widely used and have shown promising results. The preliminary studies showed that unsupervised models that worked based on sentiment lexicons, grammatical analysis and syntactic patterns with hand-engineered rules perform as well as traditional supervised approaches, such as support vector machines and naïve Bayes [1–3].

Recent researches in the sentiment analysis field, however, show that deep learning models are dominating other models in many domains [4]. These results indicate the importance of supervised models for sentiment analysis and the necessity of training data to achieve an accurate sentiment analysis model.

In less-resource languages, obtaining this goal is much more challenging. Although deep learning models have outperformed traditional approaches, they need a large set of labelled data to be adequately trained. Meanwhile, labelling raw data could be expensive and time-consuming. Approaches, such as transfer learning and active learning, have evolved to overcome these kinds of issues in less-resource languages [5].

The idea behind this research is to use the available resources in the rich-resource languages to improve the performance of sentiment analysis in Persian. To this aim, two different approaches have been examined in this research: (1) cross-lingual transfer which trains a model on a rich-resource language, English in our case, and then uses it for Persian sentiment analysis, and (2) bilingual model which trains a model on both Persian and English resources and uses it on Persian. It should be mentioned that in the second approach, our model is a bilingual model rather than a multilingual since we limit our experiments on using English and Persian data.

Our proposed model has been tested on two different embedding frameworks, namely BilBOWA [6] and VecMap [7], which are based on two different cross-lingual embedding approaches. While BilBOWA is a supervised embedding model that works based on sentence-aligned parallel corpora, VecMap benefits from word-aligned methods for embedding.

In other words, although different state-of-the-art deep neural models have been used for Persian sentiment analysis, despite their promising results in English, they did not achieve significant gain compared with traditional methods, due to the lack of enough training data in low-resource languages, such as Persian. The contribution of this article is to develop a new cross-lingual framework for supervised sentiment analysis of Persian to benefit from available data in other languages within a bilingual training scenario.

The remainder of the article is organised as follows: section 'Related Works' briefly introduces related research works. Section 'Cross-lingual Sentiment Analysis' presents our proposed model. In section 'Dataset', we present the dataset statistics. Then, results are discussed in depth in section 'Evaluation', and finally, we make our conclusion in section 'Conclusion'.

2. Related works

Sentiment analysis is a subfield of natural language processing which concentrates on text mining and opinion mining. Generally, there are plenty of works performed for sentiment analysis in English. In low-resource languages, such as Persian, however, there is a limited number of research in the field that will be reviewed in this section.

Similar to other languages, the main sentiment analysis researches for Persian are categorised into rule-based and machine learning techniques.

In 2012, Shams et al.[8] published the first article about Persian sentiment analysis. They proposed an unsupervised method based on linear discriminant analysis to classify user reviews on products/items from different domains, such as mobile phones, cameras and hotels. Lack of a large dataset prevents scalability, even if language-specific problems are not being considered. Consequently, their methods are not suited for Persian social texts [9,10].

Bagheri et al.[11] proposed a machine learning-based method for classifying a small manually chosen collection of cell phone reviews, and later, they improved this method using feature selection methods [12]. Similar to the proposed model by Shams et al.[8], this study also suffers from two main shortcomings: (1) small size and domain dependency of the dataset that was used for training [13], and (2) using only one single classification algorithm, namely naïve Bayes.

Compared with the above-mentioned machine learning-based approach, Basiri et al.[14] and Basiri et al.[15] stated the first lexicon-based method for sentiment analysis in the Persian language. They described some difficulties of Persian text processing and compared their method with machine learning algorithms on polarity detection problems. Later, they showed that direct translation is not the right solution for the resource insufficiency difficulty of the Persian language [9]. These results motivated them to offer a hybrid approach and design a lexicon for Persian sentiment analysis [13]. Their approach showed a higher achievement than pure lexicon-based and supervised learning-based systems for both polarity detection and rating prediction.

To overcome the problem of lack of enough training data in the Persian language, different researchers worked on providing sentiment lexicon to enrich this type of lexical resource of Persian and use rule-based approaches rather than supervised learning models for the task.

Dehdarbehbahani et al.[16] proposed a semi-supervised random walk model. They exploited English WordNet to build a semantic network, which helped to identify Persian language vocabulary polarity.

Alimardani and Aghaei[17] created the Persian SentiWordNet using Persian WordNet and used this lexicon in the feature weighting phase. They reported that support vector machine and logistic regression algorithms have a higher performance than naïve Bayes in polarity detection.

In another study, Asgarian et al.[18] created a Persian sentiment lexicon based on Persian WordNet. They also proved that the quality of the sentiment lexicon directly affects the overall quality of the Persian sentiment detection system.

As mentioned, since the main issue in rule-based unsupervised methods is developing a sentiment lexicon for the language, research on developing different sentiment lexicons for Persian continued in recent years. Henegar [19], LexiPers [20], PerSent [21] and SentiFars [22] are the main recent lexicons that are provided to this aim.

Moreover, there have been researches on expanding lexical resources for Persian sentiment analysis. Akhoundzade and Devin[23] proposed an unsupervised approach for expanding the Persian sentiment lexicon.

As mentioned, although the results of rule-based models are comparable with traditional supervised learning methods, with the success story of deep learning techniques, the performance of supervised sentiment analysis increased significantly, and more efforts have been paid to supervised deep learning methods.

Roshanfekr et al.[24] studied sentiment analysis in Persian text using neural networks. They compared the naïve Bayes support vector machine with bidirectional long short-term memory (LSTM) as well as convolutional neural network (CNN). Basiri et al.[25] worked on the influence of gathering methods to classify sentiments in the field of Persian reviews. Bokaee Nezhad and Deihimi[26] examined various models for Persian sentiment analysis, including CNN, CNN–LSTM and GRU, and showed that CNN–LSTM performs the best based on precision, recall and f-score metrics. A similar approach which includes the combination of CNN, LSTM and BiLSTM networks has also been proposed by Zobeidi et al.[27] for fine-grained Persian sentiment analysis. Basiri and Kabiri[28] proposed a new hybrid system for opinion mining in the Persian language. Their proposed method consists of three layers for addressing language-related problems, utilising the advantages of the lexicon-based method and exploiting the benefits of machine learning. Dashtipour et al.[29] proposed a hybrid framework for Persian sentiment analysis. Their model benefits from a dependency parser and bag-of-concepts for the concept of dependency grammar-based rules as well as neural network architecture. Dastgheib et al.[30] proposed a hybrid method using transfer learning in the semi domain correctly. They used a variety of classifiers, such as neural networks, decision tree classifiers, as well as rule-based methods. The abundance of similar domains helped them to use a classifier, that was trained on one domain, to classify data on other domains with a high yield performance. Bagheri[31] proposed an unsupervised language-independent model for aspect-based sentiment analysis of review documents. The author used different approaches, such as sentiment-aspect detection model, joint sentiment-topic model and aspect-sentiment unification model. The proposed model has been evaluated on Persian sentiment analysis.

Even though the recently proposed techniques tried to improve the performance of Persian sentiment analysis compared with the traditional machine learning methods, they all suffer from the lack of training data and cannot achieve comparable results to rich-resource languages. To have an overview of the sentiment analysis datasets that have been used in the related papers, in Table 1, we provided a brief list of data resources that have been mentioned in the related papers. The main issue about all these datasets is that none of them are publicly available for further research in the field. In some of the papers, the source of data gathering has not been mentioned.

In the next section, we describe our proposed framework to keep using deep learning-based classifiers while benefitting from a much larger set of training data with the help of cross-lingual embedding approaches. Although monolingual embedding for Persian has been widely studied by Hadifar and Momtazi[32], we will show that by cross-lingual embedding, we benefit from larger data resources for the task

3. Cross-lingual sentiment analysis

Our proposed architecture for cross-lingual sentiment analysis includes a neural model for word embedding and a deep neural network for document classification. For each part of this framework, we propose using different architectures. It helps to prove that the proposed architecture is general enough to be used with any embedding or any deep classifier. Such general framework makes the system flexible to be used for sentiment analysis of any low-resource language. It

Table 1. Overview of related works on Persian sentiment analysis, their domains and their dataset.

Paper	Domains	No. of reviews	Publicly available	Classes type
Ashrafi Asli et al.[5]	Electronic products	50,000 positive/15,000 negative/29,000 neutral	No	3 classes
Dashtipour et al.[29]	Electronic products/hotel	3300 positive/3300 negative	No	Binary
Dastgheib et al.[30]	Electronic products	26,000	No	3 classes
Bagheri[31]	Electronic products	1200	No	Binary
Bokaee Nezhad and Deihimi[26]	Politics/electronic products	3872 positive/3872 negative/3872 neutral	No	3 classes
Zobeidi et al.[27]	Electronic products	151,229	No	5 classes and binary
Asgarian et al.[18]	Electronic products	31,730	No	Binary
Roshanfekr et al.[24]	Electronic products	200,761	No	3 classes
Basiri and Kabiri[9]	Electronic products	14,000	No	5 classes
Alimardani and Aghaei[17]	Hotel	4360 positive/1805 negative	No	Binary
Bagheri et al.[11]	Electronic products	511 positive/318 negative	No	Binary
Shams et al.[8]	Electronic products	200 positive/200 negative	No	Binary

also provides the possibility to compare different embeddings and architectures for the task. Figure 1 provides a comprehensive view of our proposed architecture.

For the embedding part, we use BilBOWA [6] and VecMap [7], which are two different cross-embedding models from two different perspectives. For the document classification part, we use CNN [33], LSTM and their combinations as CNN–LSTM [34] and LSTM–CNN due to their superior results in sentiment analysis in different languages, including Persian [5,24]. The detailed description of the embeddings and classifiers is described in the next subsections.

3.1. Cross-lingual embedding

Cross-lingual embedding has been studied based on the different levels of cross-lingual resources, such as word level, sentence level and document level. Cross-lingual embedding models based on sentence-aligned parallel corpora [6,35] train word embeddings of two languages simultaneously by sharing parallel corpora. Cross-lingual models based on word-level alignment [36], however, use pre-trained embeddings and use a mapping technique to rotate embeddings to a common vector space. Different mapping approaches have been proposed to this aim, including regression methods, canonical methods, orthogonal methods and margin methods [7].

In this research, we use both embedding models. We select BilBOWA [6] as a sentence-aligned supervised approach and VecMap [7] as an orthogonal-based word-aligned approach.

The idea behind Bilingual Bag-of-Words without Word Alignments (BilBOWA) embedding is to train two monolingual skip-grams in two different languages together, in one shared skip-gram. Therefore, it needs a parallel corpora. BilBOWA uses an objective function that not only assigns similar words of each language with similar embeddings but also similar words from different languages will have similar embeddings. Thus, besides a monolingual objective function, a cross-lingual objective function is employed, which contains an additional factor to place the words with similar meanings and especially the translation pairs in two selected languages closer to each other in their shared words vector space [6]. This means that, in BilBOWA's skip-gram model, in addition to predicting a word from its context in monolingual text, a word should be predicted from the aligned sentence in the parallel corpus.

To train BilBOWA on Persian and English, we used Amirkabir Bilingual Farsi-English Corpus (AFEC), which is a parallel Persian–English parallel corpora with alignments at the sentence level [37].

VecMap embedding methodology is among the word-aligned models. First, two separate embeddings are built. Then, one or both of the embeddings are mapped with the orthogonal transformation to maximise their similarity to the other one [7]. To attain this target, a bilingual dictionary is needed. For our experiments, we used the English pre-trained Glove embedding and Persian pre-trained FastText embedding. The cross-lingual embedding is then built using the dictionary provided by the VecMap project for English–Persian mapping [38].[1] The dictionary includes 5000 words.

3.2. Document classification

As mentioned, CNN achieved promising results in sentiment analysis [24]. The model has also been expanded using LSTM layers in the architecture. In this article, we use these two models for our cross-lingual sentiment analysis.

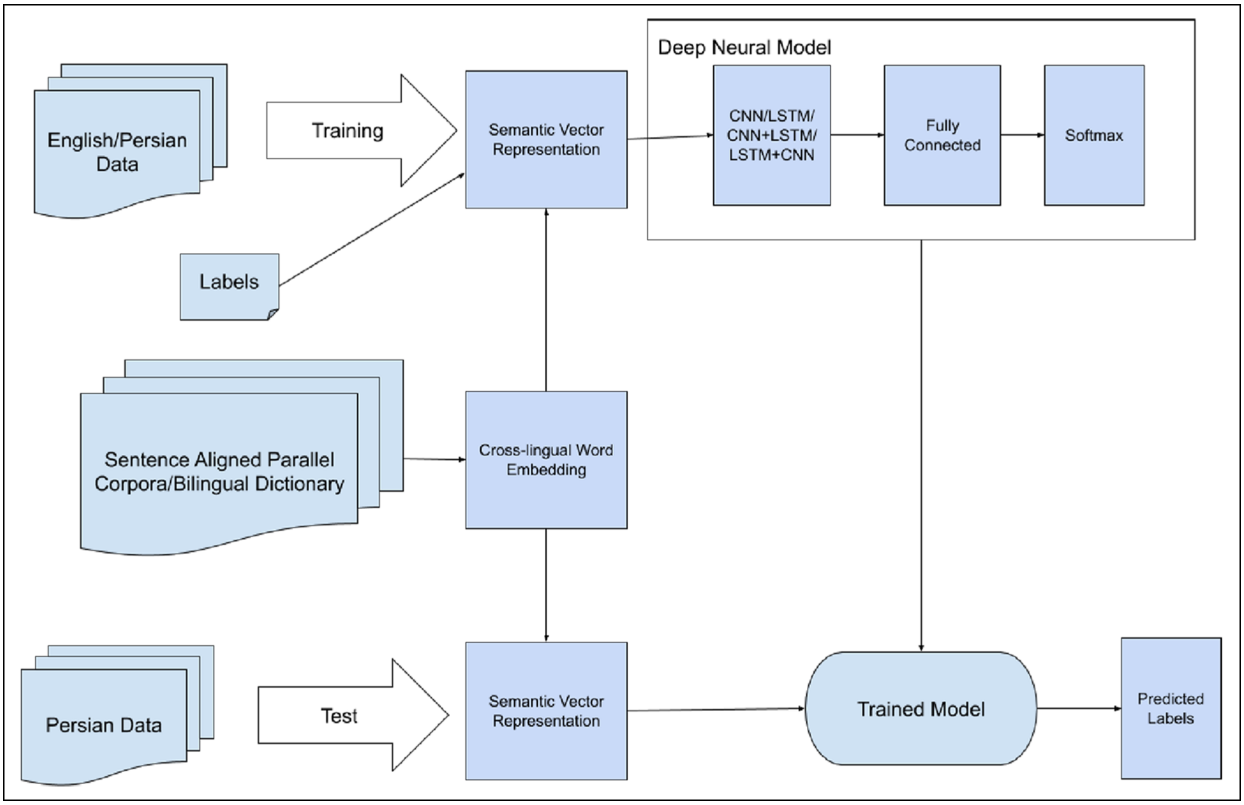

Figure 1. The architecture of our proposed model for cross-lingual sentiment analysis.

CNN is a particular type of deep neural network that can be used for models with spatial data. CNNs have been widely used for image processing tasks, such as classification or segmentation [39]. The success story of CNN in image processing motivated researchers to use this model in text processing tasks as well.

Convolutional layers only use special connections from the previous layer; in particular, local neurons are connected to the next layer neurons. In this research, the architecture of the CNN network has been inspired by the model described by Zhang and Wallace [40], as presented in Figure 2. The model has been widely used in different sentiment analysis frameworks and showed promising results [4,33,41].

In this architecture, a tokenized sentence is first converted to a sentencing matrix, where each row represents a word vector. For our case, these embeddings came from the cross-lingual embedding, which was described in the previous section.

The embedding matrix is then passed through a convolution layer, where different region filtering are applied to the matrix. The output of this filtering is feature maps that are extracted for each region size filter. In the next step, a max-pooling layer is used, and finally, a softmax function is applied to the output to receive a binary output from the model indicating positive or negative label of the input text.

In our proposed framework, we also used an LSTM model. LSTM is one of the popular architectures to deal with sentiment classification. The basic architecture of LSTM has been proposed by Hochreiter and Schmidhuber [42]. It is a typical recurrent neural networks (RNNs), which are established to deal with sequential data by sharing their inner weights beyond the sequence.

RNNs help to extract temporal features of the input. This feature in RNNs is an important issue in processing texts due to the sequential behaviour of text. RNNs, however, suffer from a shortcoming, which is the problem of vanishing or exploding gradients. LSTMs are proposed to relieve this problem.

The advantage of LSTM compared with traditional RNN is that in LSTM, inside each cell three gates are used which determine which information of the content should be updated or exposed.

The architecture of the LSTM model is represented in Figure 3. As it is shown in this architecture, two LSTM layers, forward and backward, are used to encode the input text; the output of these two layers are then concatenated and passed

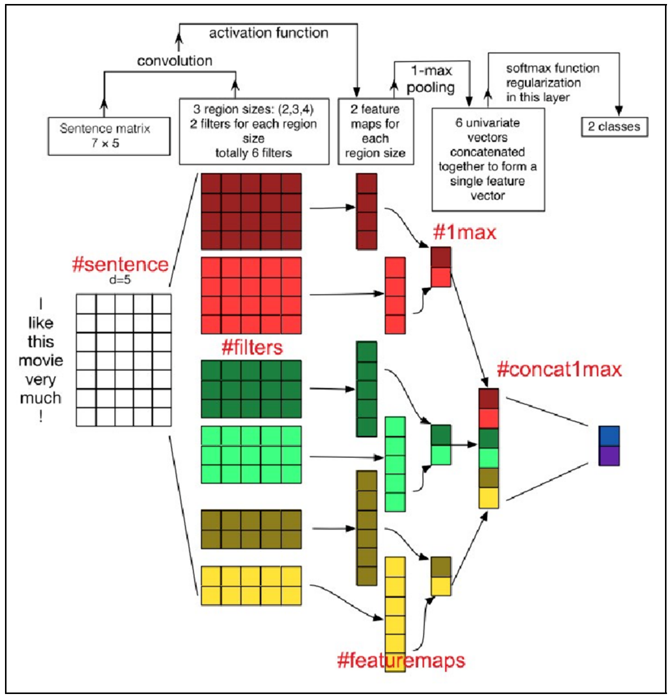

Figure 2. The architecture of the CNN model employed in a sentiment analysis task [40].

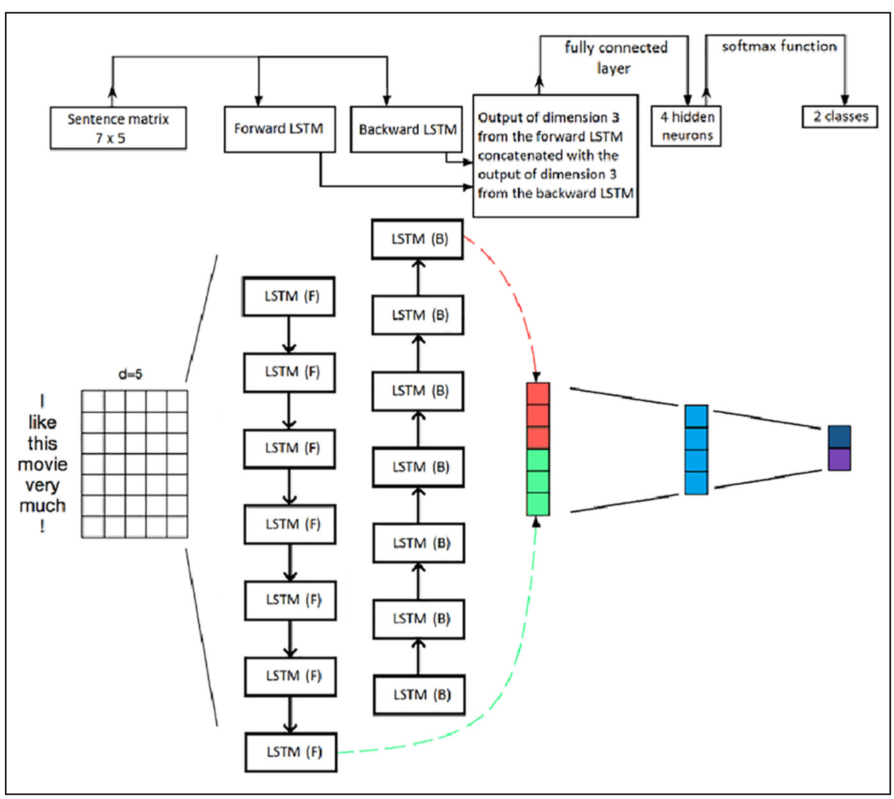

Figure 3. The architecture of the LSTM model [43].

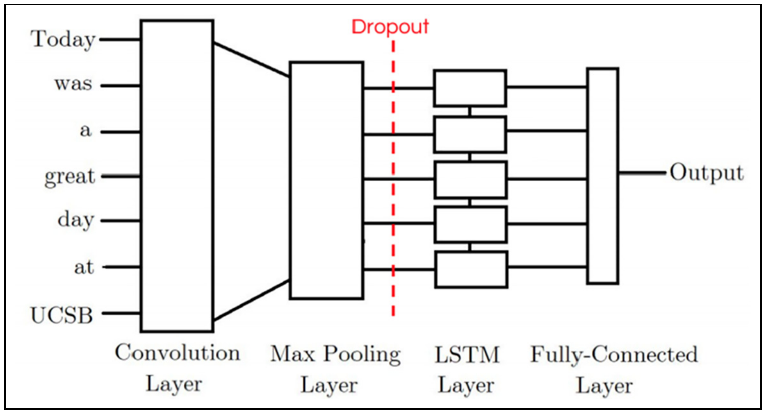

Figure 4. The architecture of the CNN–LSTM model [44].

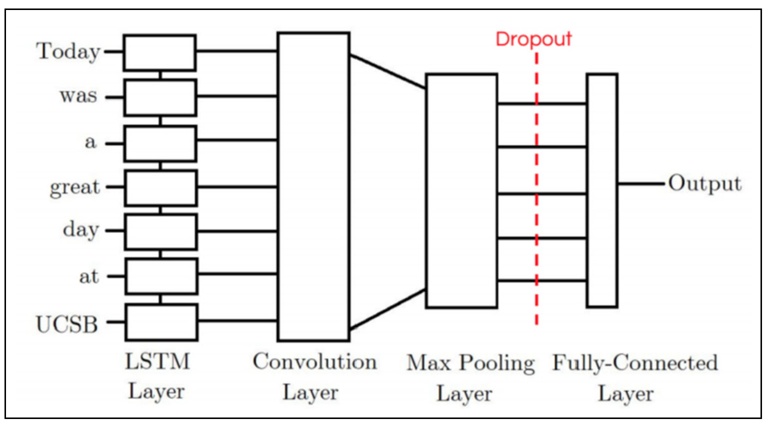

Figure 5. The architecture of the LSTM–CNN model [44].

to a fully connected layer. At the end of this architecture, a softmax layer is used to produce a binary output as the sentiment label of the input text [43].

In addition to the CNN and LSTM models, we also use the CNN–LSTM architecture proposed by Alayba et al.[34]. As the architecture of the model is represented in Figure 4, it includes CNN layers on the front end followed by LSTM layers and a dense layer on the output. Accordingly, it is essential to analyse this architecture as two submodules: the CNN part which extract local features, and the LSTM part for interpreting the features across the sequence of words in the text.

The CNN–LSTM model combination consists of an initial convolution layer, which is identical to the CNN model described above and takes cross-lingual word embeddings as an input. The output of the CNN part is then transformed into a smaller dimension, which is passed to the LSTM layer. The LSTM layer is able to learn the sequential features of words based on their order in the input text. This additional layer makes the model more powerful than the CNN model.

In the last configuration of our document classifier, we use the LSTM–CNN architecture. This model, which is presented in Figure 5, is comparative to the CNN–LSTM. It is comprised a beginning LSTM layer, which receives word embeddings for each token as inputs. The advantage of this layer is that the output of each token stores the information of the current token as well as the previous tokens. In other words, the LSTM layer encodes the original input text to a new vector space which captures sequential behaviour of the text. The output of the LSTM layer is then utilised by the convolution layer, which extracts local features. At the end of this architecture, the convolution layer's output will be pooled to a single layer perceptron and eventually produce output as either a positive or negative label.

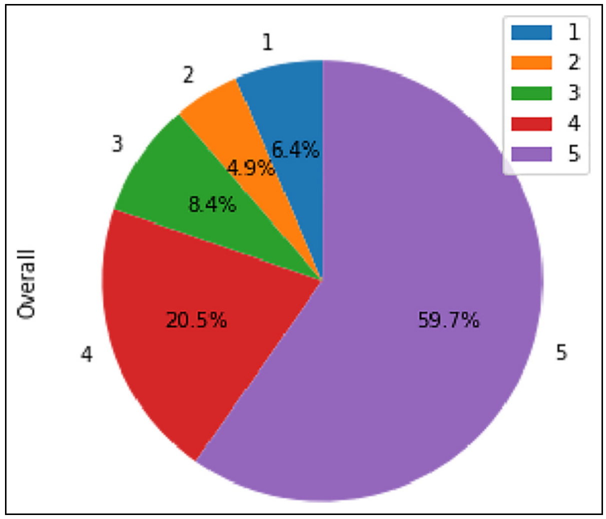

Figure 6. Distribution of the class labels in the Amazon dataset.

4. Dataset

The main challenge in Persian sentiment analysis is the lack of well-defined datasets. For English data, however, we have more reliable datasets that have been widely used for research purposes. In this article, we evaluate our proposed model on two datasets from English and Persian which are in the same domain. For English, the electronic domain of the Amazon dataset, which is from the user reviews in Amazon website,[2] is used. For Persian, the Digikala dataset, which is from the Digikala website,[3] an Iranian online shopping website, is used. For all experiments, we focus on two polarity classes: negative and positive sentiments.

4.1. Amazon electronic products data[4]

Amazon electronic products dataset is a well-known dataset for sentiment analysis tasks. The dataset consists of user reviews, including their rating for products. The user rates are categorised into five classes, 1 star to 5 stars, with an unbalanced distribution as presented in Figure 6. The total number of instances in this dataset is 1,689,188.

According to the user rating beside each review, the data are labelled as positive if the user rates are 5 stars, and the data with user rates of 1 or 2 stars are labelled as negative.

4.2. Digikala electronic products[5]

Digikala electronic products dataset is one of the limited numbers of datasets available for Persian sentiment analysis. The total number of user reviews in this dataset is 23,574 and the distribution of labels in this dataset is presented in Figure 7. Similar to the Amazon dataset, this dataset also has more bias towards positive data. As mentioned, the total number of samples in this dataset is about 23,000, which is much smaller than the available data for English (about 1.6 m). This shows the problem of lack of data in low-resource languages, such as Persian, and indicates the necessity of benefitting from datasets of other languages with the proposed cross-lingual framework.

Since an essential part of the modern human life is the process of decision-making, which is based on systems that rely on data, it is very important to have a reliable dataset. Therefore, if the dataset is biased, we will see big mistakes. As indicated in the data distribution of both the English and Persian dataset, the data ratio of positive samples is much higher than the negative instances. Therefore, to avoid any bias in data, we used random undersampling and did our experiments on a subset of these datasets with an equal number of positive and negative samples. The amount of train, validation and test instances for each dataset is summarised in Table 2. Since the test phase of all experiments is done on Persian, we need no test data from the Amazon dataset.

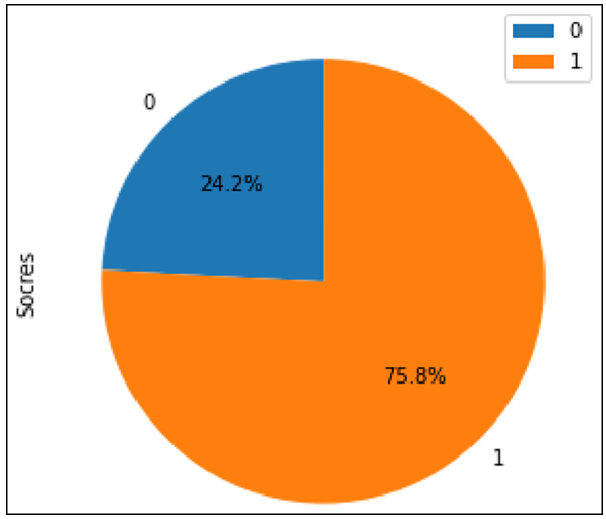

Figure 7. Distribution of the class labels in the Digikala dataset.

Table 2. Data distribution.

Dataset	Train	Validation	Test	Overall
Amazon	180,000	20,000	–	200,000
Digikala	6000	1000	4000	11,000

Table 3. Parameter settings used in the CNN model.

Model	Parameters	Values
CNN	Strides	10
	Activation	ReLU
	Filter region size	(2, 3, 4, 5)
	Feature maps	100
	Dropout rate	0.3
	Pooling layer	1-max pooling
	Optimizer	Adam
	Sequence length	128
	Learning rate	0.0001

CNN: convolutional neural network.

5. Evaluation

5.1. Setup of experiments

In the implementation of this research, we used different Python libraries, including Keras, Pandas, Gensim and Matplotlib.

The parameters used in our model are summarised in Tables 3 and 4. In terms of the embeddings, we use 300-dimentional vectors for both BilBOWA and VecMap.

5.2. Results

As mentioned, in our experiments, two different cross-lingual embeddings, namely BilBOWA and VecMap, and two different classifiers, namely CNN and CNN–LSTM, are used. Moreover, since the cross-lingual embeddings are pre-trained

Table 4. Parameter settings used in the LSTM model.

Model	Parameters	Values
LSTM	Epoch	10
	Batch size	128
	Filter	32
	Pool size	2
	Dropout rate	0.3
	Optimizer	Adam
	Activation	Sigmoid

LSTM: long short-term memory.

with no knowledge from the training data of the sentiment analysis task, we did our experiments in two different modes: static embedding that directly uses the pre-trained vectors and dynamic embedding that fine-tunes the embeddings with the training data.

For all experiments, the baseline is a monolingual sentiment analysis framework that trains the model with the Persian data and tests it on the same language. This model, which limits itself to the available Persian data and does not benefit from other resources, is the common model on all the state-of-the-art Persian sentiment analysis researches.

The results of all experiments are reported based on precision, recall and F-measure metrics.

Table 5 shows the result of our model and the baseline model within the CNN architecture. It should be mentioned that in all experiments, the 4000 test data from the Digikala dataset are used for testing the model, the train data, however, varies; English training means using 180,000 samples of Amazon, Persian training means using 6000 sample of Digikala and multilingual training means the combination of both English and Persian training samples.

Comparing the results of different training models and different cross-lingual embeddings, the following observations are notable:

- In all experiments, our proposed model on using English data for training the model outperforms the state-of-the-art baseline models than the use of the Persian data for training.
- Comparing the results of English training and Bilingual training shows that although the proposed cross-lingual sentiment analysis significantly outperforms monolingual models, we can further improve the model by benefitting from both English and Persian data in a Bilingual training step. The Bilingual training scenario helps to benefit from a large amount of data that are available for English and also learns the specific linguistics language-dependent features that are available in Persian data.
- In all experiments, VecMap outperforms BilBOWA, but the differences are not very significant. One potential reason is that since BilBOWA requires parallel data, we have a limitation on the corpora that can be used for training the model, while in VecMap, we have no limitation in this respect.
- Comparing static and dynamic embeddings shows that with dynamic embedding, which provides the possibility of retraining the embedding models, we can further improve the model by capturing semantic features that are available in the content of training data.

In the next step of our evaluation, we repeated all the experiments on the LSTM, CNN–LSTM and LSTM–CNN networks as our classifiers. The results of these experiments are available in Tables 6–8. As can be seen in the tabulated results, all the observations on the CNN results are also valid on the other three models; that is, in all models, the proposed cross-lingual training outperforms the baseline monolingual model, and the results of both BilBOWA and VecMap embeddings are comparable.

Comparing all classifiers, as can been seen in Figure 8, the best results are achieved using hybrid models. This indicates that LSTM–CNN and CNN–LSTM architectures which benefits from both models can capture both sequential and spacial information from the text to have a more accurate prediction. Moreover, the superior performance of LSTM–CNN compared with CNN–LSTM shows that capturing sequential data at the beginning of the architecture is more powerful because no information is missing. However, in the CNN–LSTM model, by running a CNN model at the beginning, there is a possibility of missing some sequential information. Overall, the best result is achieved by LSTM–CNN and VecMap, such that the F-measure reaches 95.04.

Table 5. Comparing the results of proposed cross-lingual sentiment analysis model with the baseline monolingual models with CNN as the classifier.

Train data	Static/dynamic	BilBOWA			VecMap		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Persian	Static embedding	60.28	58.35	59.30	67.04	66.20	66.62
	Dynamic embedding	68.29	65.98	67.12	73.92	72.80	73.36
English	Static embedding	70.31	68.53	69.41	72.68	71.70	72.19
	Dynamic embedding	79.72	78.19	78.95	81.49	80.00	80.74
Bilingual	Static embedding	73.28	71.63	72.45	80.11	78.32	79.21
	Dynamic embedding	77.72	76.15	76.93	85.32	83.28	84.29

Table 6. Comparing the results of proposed cross-lingual sentiment analysis model with the baseline monolingual models with LSTM as the classifier.

Train data	Static/dynamic	BilBOWA			VecMap		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Persian	Static embedding	63.74	62.12	62.92	69.97	68.50	69.23
	Dynamic embedding	70.02	68.71	69.36	76.39	75.17	75.78
English	Static embedding	70.98	69.24	70.10	75.12	73.67	74.39
	Dynamic embedding	78.32	76.22	77.26	82.03	80.81	81.42
Bilingual	Static embedding	77.23	75.54	76.38	80.91	79.44	80.17
	Dynamic embedding	80.62	79.29	79.95	87.23	86.05	86.64

Table 7. Comparing the results of proposed cross-lingual sentiment analysis model with the baseline monolingual models with CNN–LSTM as the classifier.

Train data	Static/dynamic	BilBOWA			VecMap		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Persian	Static embedding	65.84	64.02	64.92	72.30	70.95	71.62
	Dynamic embedding	71.76	70.45	71.10	78.08	76.73	77.40
English	Static embedding	70.71	68.95	69.82	77.30	76.26	76.83
	Dynamic embedding	78.27	76.90	77.58	84.92	83.47	84.19
Bilingual	Static embedding	76.82	75.27	76.04	84.25	83.25	83.75
	Dynamic embedding	85.02	83.59	84.30	89.68	88.13	88.90

Table 8. Comparing the results of proposed cross-lingual sentiment analysis model with the baseline monolingual models with LSTM–CNN as the classifier.

Train data	Static/dynamic	BilBOWA			VecMap		
		Precision	Recall	F-measure	Precision	Recall	F-measure
Persian	Static embedding	67.60	65.88	66.73	75.32	73.61	74.46
	Dynamic embedding	74.14	76.06	76.60	79.42	78.68	79.05
English	Static embedding	75.98	74.63	75.30	79.13	77.62	78.73
	Dynamic embedding	84.26	83.12	83.69	86.77	85.06	85.91
Bilingual	Static embedding	83.42	82.07	82.74	85.00	83.53	84.26
	Dynamic embedding	87.13	86.01	86.57	92.44	91.20	91.82

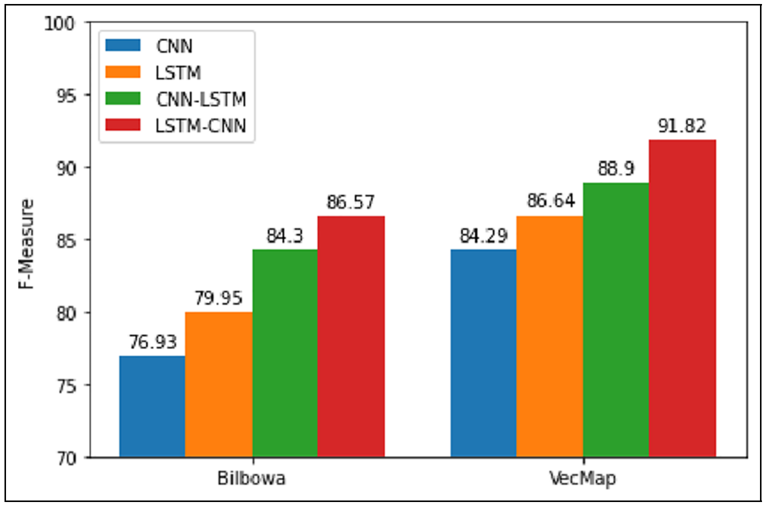

Figure 8. Comparing the results of different classifiers.

6. Conclusion

In this article, we used various methods to cover the weakness of low-resource languages in the sentiment analysis task. We deployed cross-lingual embedding to model sentiment analysis as a transfer learning model. Through learning methods, such as CNN and LSTM, and combinations of them, namely CNN–LSTM and LSTM–CNN, which have been employed as our architectures for this task, we found that combining training data of a low-resource language, such as Persian, with a rich-resource language, such as English, achieves excellent performance on sentiment classification. These results were achieved with the help of cross-lingual word embedding, which can be created by different approached, including sentence-aligned and word-aligned, models with comparable performance.

In future work, we study cross-lingual word embedding in more detail to produce better pre-training features as the input of the network, and meanwhile, try other classification approaches, such as hierarchical attention network.

We also aim to provide a multilingual framework to cover several languages and different domains in the future. This comprehensive framework provides the opportunity to show how using the advantages of cross-lingual word embedding can help enhance the sentiment analysis task of different languages.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Saeedeh Momtazi https://orcid.org/0000-0002-8110-1342

Notes

1. https://github.com/artetxem/vecmap
2. https://www.amazon.com/
3. https://www.digikala.com/
4. link: http://jmcauley.ucsd.edu/data/amazon/
5. https://github.com/rajabzz/digikala-crawler

References

1. Thelwall M, Buckley K, Paltoglou G et al. Sentiment strength detection in short informal text. *J Am Soc Inf Sci Technol* 2010; 61(12): 2544–2558.
2. Thelwall M, Buckley K and Paltoglou G. Sentiment strength detection for the social web. *J Am Soc Inf Sci Technol* 2012; 63(1): 163–173.
3. Momtazi S. Fine-grained german sentiment analysis on social media. In: *Proceedings of the eighth international conference on language resources and evaluation (LREC'12)*, Istanbul, Turkey, 23–25 May 2012, pp. 1215–1220. Paris: European Language Resources Association.
4. Zhang Y, Zhang Z, Miao D et al. Three-way enhanced convolutional neural networks for sentence-level sentiment classification. *Inf Sci* 2019; 477: 55–64.
5. Asli SAA, Sabeti B, Majdabadi Z et al. Optimizing annotation effort using active learning strategies: a sentiment analysis case study in persian. In: *Proceedings of the 12th language resources and evaluation conference*, Marseille, 11–16 May 2020, pp. 2855–2861. Paris: European Language Resources Association.
6. Gouws S, Bengio Y and Corrado G. BilBOWA: fast bilingual distributed representations without word alignments. In: *Proceedings of International Conference on Machine Learning, Lille*, 6–11 July 2015.
7. Artetxe M, Labaka G and Agirre E. Generalizing and improving bilingual word embedding mappings with a multi-step framework of linear transformations. In: *Proceedings of the thirty-second AAAI conference on artificial intelligence (AAAI-18)*, 2–7 February 2018, New Orleans, LA, pp. 5012–5019. Menlo Park, CA: AAAI.
8. Shams M, Shakery A and Faili H. A non-parametric LDA-based induction method for sentiment analysis. In: *The 16th CSI international symposium on artificial intelligence and signal processing (AISP 2012)*, Shiraz, Iran, 2–3 May 2012, pp. 216–221. New York: IEEE.
9. Basiri ME and Kabiri A. Translation is not enough: comparing lexicon-based methods for sentiment analysis in Persian. In: *2017 International symposium on computer Science and software engineering conference (CSSE)*, Shiraz, Iran, 25–27 October 2017, pp. 36–41. New York: IEEE.
10. Basiri M, Nilchi AN and Ghasem-Aghaee N. LSAP: a new lexicon-based method for sentiment analysis in Persia. In: *1st International conference on new research achievements in electrical and computer engineering (ICNRAECE)*, Tehran, Iran, 13 May 2016.
11. Bagheri A, Saraee M and de Jong F. Sentiment classification in Persian: introducing a mutual information-based method for feature selection. In: *2013 21st Iranian conference on electrical engineering (ICEE)*, Mashhad, Iran, 14–16 May 2013, pp. 1–6. New York: IEEE.
12. Saraee M and Bagheri A. Feature selection methods in Persian sentiment analysis. In: *International conference on application of natural language to information systems*, Salford, 19–21 June 2013, pp. 303–308. Berlin: Springer.
13. Basiri ME and Kabiri A. Words are important: improving sentiment analysis in the Persian language by lexicon refining. *ACM Trans Asian Low Resour Lang Inf Process* 2018; 17(4): 1–18.
14. Basiri ME, Naghsh-Nilchi AR and Ghasem-Aghaee N. Sentiment prediction based on dempster-shafer theory of evidence. *Math Probl Eng* 2014; 2014: 361201.
15. Basiri ME, Ghasem-Aghaee N and Reza A. Lexicon-based sentiment analysis in persian. *Curr Futur Dev Artif Intell* 2017; 1: 154–183. https://www.researchgate.net/publication/319038208_Current_and_Future_Developments_in_Artificial_Intelligence_-_ Intelligent_Computational_Systems_A_Multi-Disciplinary_Perspective_-_Volume_1
16. Dehdarbehbahani I, Shakery A and Faili H. Semi-supervised word polarity identification in resource-lean languages. *Neural Netw* 2014; 58: 50–59.
17. Alimardani S and Aghaei A. Opinion mining in Persian language using supervised algorithms. *J Inf Syst Telecommun* 2015; 3: 135–141.
18. Asgarian E, Kahani M and Sharifi S. The impact of sentiment features on the sentiment polarity classification in persian reviews. *Cognit Comput* 2018; 10(1): 117–135.
19. Asgarian E, Kahani M and Sharifi S. Hesnegar: Persian sentiment wordnet. *Signal Data Process* 2018; 15(1): 71–86.
20. Sabeti B, Hosseini P, Ghassem-Sani G et al. Lexipers: an ontology based sentiment lexicon for persian. arXiv preprint arXiv: 191105263, 2019.
21. Dashtipour K, Raza A, Gelbukh A et al. Persent 2.0: Persian sentiment lexicon enriched with domain-specific words. In: *International conference on brain inspired cognitive systems*, Guangzhou, China, 13–14 July, 2019, pp. 497–509. Berlin: Springer.
22. Dehkharghani R. Sentifars: a Persian polarity lexicon for sentiment analysis. *ACM Trans Asian Low Resour Lang Inf Process* 2019; 19(2): 1–12.
23. Akhoundzade R and Devin KH. Persian sentiment lexicon expansion using unsupervised learning methods. In: *2019 9th international conference on computer and knowledge engineering (ICCKE)*, Mashhad, Iran, 24–25 October 2019, pp. 461–465. New York: IEEE.
24. Roshanfekr B, Khadivi S and Rahmati M. Sentiment analysis using deep learning on Persian texts. In: *2017 Iranian conference on electrical engineering (ICEE)*, Tehran, Iran, 2–4 May 2017, pp. 1503–1508. New York: IEEE.

25. Basiri ME, Kabiri A, Abdar M et al. The effect of aggregation methods on sentiment classification in Persian reviews. *Enterp Inf Syst* 2019: 1–28.
26. Nezhad ZB and Deihimi MA. A combined deep learning model for persian sentiment analysis. *IIUM Eng J* 2019; 20(1): 129–139.
27. Zobeidi S, Naderan M and Alavi SE. Opinion mining in Persian language using a hybrid feature extraction approach based on convolutional neural network. *Multimed Tools Appl* 2019; 78(22): 32357–32378.
28. Basiri ME and Kabiri A. Homper: a new hybrid system for opinion mining in the Persian language. *J Inf Sci* 2020; 46(1): 101–117.
29. Dashtipour K, Gogate M, Li J et al. A hybrid Persian sentiment analysis framework: integrating dependency grammar based rules and deep neural networks. *Neurocomputing* 2020; 380: 1–10.
30. Dastgheib MB, Koleini S and Rasti F. The application of deep learning in Persian documents sentiment analysis. *Int J Inf Sci Manag* 2020; 18(1): 1–15.
31. Bagheri A. Integrating word status for joint detection of sentiment and aspect in reviews. *J Inf Sci* 2019; 45(6): 736–755.
32. Hadifar A and Momtazi S. The impact of corpus domain on word representation: a study on Persian word embeddings. *J Lang Resour Eval* 2018; 52(4): 997–1019.
33. Kim H and Jeong YS. Sentiment classification using convolutional neural networks. *Appl Sci* 2019; 9(11): 2347.
34. Alayba AM, Palade V, England M et al. A combined CNN and LSTM model for arabic sentiment analysis. In: *International cross-domain conference for machine learning and knowledge extraction*, Hamburg, 27–30 August 2018, pp. 179–191. Berlin: Springer.
35. Luong MT, Pham H and Manning CD. Bilingual word representations with monolingual quality in mind. In: *Proceedings of the 1st workshop on vector space modeling for natural language processing*, 2015, Denver, CO, 5 June 2015, pp. 151–159. Stroudsburg, PA: Association for Computational Linguistics.
36. Artetxe M, Labaka G and Agirre E. Learning principled bilingual mappings of word embeddings while preserving monolingual invariance. In: *Proceedings of the 2016 conference on empirical methods in natural language processing*, Austin, TX, 1–5 November 2016, pp. 2289–2294. Stroudsburg, PA: Association for Computational Linguistics.
37. Jabbari F and Ziabary SMM. Developing an open-domain english-farsi translation system using AFEC: Amirkabir bilingual farsi-english corpus. In: *The fourth workshop on computational approaches to Arabic script-based languages*, San Diego, CA, 19 June 2012, p. 17. Stroudsburg, PA: Association for Computational Linguistics.
38. Aliramezani M, Bokaei MH and Sameti H. Improving Persian word embeddings using cross-lingual joint space. In: *Information and knowledge technology conference*. 1–2 January 2020. Tehran, Iran.
39. Yenter A and Verma A. Deep CNN-LSTM with combined kernels from multiple branches for imdb review sentiment analysis. In: *2017 IEEE 8th annual ubiquitous computing, electronics and mobile communication conference (UEMCON)*, New York, NY, 19–21 October 2017, pp. 540–546. New York: IEEE.
40. Zhang Y and and Wallace B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:151003820, 2015.
41. Kim Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:14085882, 2014.
42. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9(8): 1735–1780.
43. Cliche M. Bb_twtr at SemEval-2017 task 4: Twitter sentiment analysis with CNNs and LSTMs. arXiv preprint arXiv:170406125, 2017.
44. Sosa PM. Twitter sentiment analysis using combined LSTM-CNN models. *Zugriff Am* 2017; 10: 2019.