International Journal of Information Science and Management
Vol. 18, No. 1, 2020, 1-15

The application of Deep Learning in Persian Documents Sentiment Analysis

Mohammad Bagher Dastgheib
Department of Designing & System Operation, Regional Information Center for Science and Technology, RICeST, IRAN.
Corresponding Author:
dastgheib@ricest.ac.ir

Sara Koleini
Department of Information & Communication Technology, Regional Information Center for Science and Technology, RICeST, IRAN.
koleini@ricest.ac.ir

Farzad Rasti
Department of Designing & System Operation
Regional Information Center for Science and Technology, RICeST, IRAN
Farzad.Rasti@gmail.com

Abstract

Nowadays the amount of textual information on the web is grown rapidly. The huge textual data needs more accurate classification algorithms. Sentiment analysis is a branch of text classification that is used to classify user opinions in case of market decisions, product evaluations or measuring consumer confidence. With the rise of the production rate of Persian text data in a commercial area, improvement of the efficiency of algorithms in Persian is a must. The structure of the Persian language such as word and sentence structures poses some challenges in this area. Deep learning algorithms are recently used in NLP and especially sentiment text classification for many dominant languages like Persian. The goal is to improve the performance of classification using deep learning issues. In this work, the authors proposed a hybrid method by a combination of structural correspondence learning (SCL) and convolutional neural network (CNN). The SCL method selects the most effective pivot features so the adaptation from one domain to similar ones cannot drop the efficiency drastically. The results showed that the proposed hybrid method that is learned from one domain can act efficiently in a similar domain. The result showed that applying a combination of SCL+CNN can improve the result of sentiment classification for two domains more than 10 percent.

Keywords: Deep learning, Persian Documents, Sentiment Analysis, Convolutional Neural Network (CNN), Structural Correspondence Learning (SCL).

Introduction

Due to the explosive growth of data on the web and increase of information in cyberspace in all aspects, such as business, news, and entertainment, the valuable source of user's opinions about products, news, services and goods can be obtained. Users and customers usually use the opinions of other users to decide on the use of services or the purchase of products, so the

analysis of emotions (opinions) is important and necessary. The extraction of information and its analysis requires pre-processing and fetching algorithms and information categorization.

Sentiment analysis is the process of recognizing, extract, evaluate, and study of affective states and subjective information by using different methods such as machine learning or statistical techniques. Analyzing people's opinions, sentiments, attitudes and emotions are used to identify and classify the opinions in the sentiment analysis field. Sentiment analysis provides a way to determine the user's perspective on a specific service or product[1].

Because of the importance of opinions in the entire of the human activities, sentiment analysis's systems are applied to the most of business and social domains. Sentiment analysis is one of the popular research areas in natural language processing (NLP), data mining, web mining and text mining (Pang & Lee, 2008). Sentiment classification is one of the most practical techniques that are used in sentiment analysis (Hu & Liu, 2004). In this method, the analysis of an automated sentiment classification categorizes the opinions into different categories. Sentiment classification considers the entire document as a basic information item; therefore, it can be called a document-level sentiment classification (Pang & Lee, 2008). Sentiment analysis is considered as the following three levels (Zhang, Wang & Liu, 2018):

1. Document level – this level classifies the document as expressing an overall positive or negative opinion. It considers the whole document as the basic information unit and assumes that the document contains sentiments about a single object.

2. Sentence level - in these level individual sentences in a document is classified. Each sentence cannot be assumed to be opinionated. One often first classifies a sentence as opinionated or not opinionated, which is called subjectivity classification. Then the resulting opinionated sentences are classified as expressing positive or negative opinions.

3. Aspect level – in this level the task is to extract and summarize people's opinions expressed on entities and aspects/features of the objects.

The sentiment analysis methods can be categorized into statistical / machine learning, knowledge /lexicon based and hybrid approaches as shown in Figure 1 (Shahnawaz & Astya, 2017).

Statistical methods are based on the corpus, lexicons, dictionaries and treasures. The results are obtained by statistical analysis of the text (Blitzer, Dredze & Pereira, 2007). Statistical methods by using the dictionaries (antonyms, synonyms, and word orientations) and calculating the use of vocabulary statistics in sentences, attempt to determine the orientation of the sentence (Hosseini Ramaki, Maleki, Anvari & Mirroshandel,2018).

Learning machine method is divided by supervised, semi-supervised and unsupervised. A large number of training documents are used in supervised learning methods as shown in Figure1. Table 1 shows the most frequently used algorithm classifiers in sentiment analysis (Medhat, Hassan & Korashy 2014).

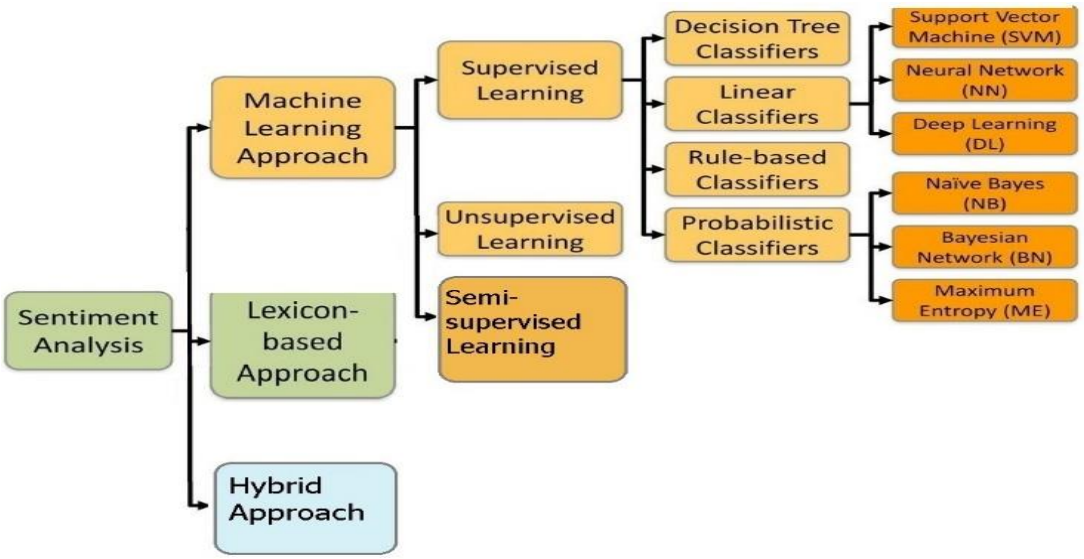

Figure 1. Sentiment Analysis methods
(Shahnawaz & Astya, 2017, Medhat et al., 2014, Zhang et al., 2018)

Table 1
Some of the main sentiment analysis's classifiers (Medhat et al., 2014)

Classifier	Description
Probabilistic classifiers	Using mixture models for classification, Each mixture component is a generative model that provides the probability of sampling a particular term for that component
Naıve Bayes Classifier	based on the distribution of the words in the Document, computes the subsequent probability of a class
Bayesian Network	The complete model for the variables and their relationships is considered. Therefore, a complete Joint Probability Distribution (JPD) over all the variables, is specified for a model.
Maximum Entropy Classifier	Converts labeled feature sets to vectors by encoding. This encoded vector is used to compute weights for each feature that can then be combined to determine the most likely label for a feature set.
Support Vector Machines Classifiers(SVM)	Determine linear separators in the search space and construct a nonlinear decision surface in the original feature space by mapping the data instances non-linearly to an inner product space where the classes can be separated linearly with a hyperplane.
Neural Network	Many neurons are used as input and output. The inputs to the neurons are denoted by the vector Xi which is the word frequencies in the ith document. There are a set of weights A which are associated with each neuron used in order to compute a function of its inputs. The outputs of the neurons in the earlier layers feed into the neurons in the later layers.
Decision tree classifiers	Using hierarchical decomposition of the training data space. To divide the data, a condition on the attribute value is used. This condition is the presence or absence of one or more words in the document. The division of the data space until the leaf nodes contain certain minimum numbers of records is done recursively

Classifier	Description
Linear classifiers	One of the main goal of statistical classification in machine learning is to use an object's characteristics to identify the appropriate class for the object. By using linear combination of the object's features, a linear classifier makes a classification decision.
rule based classifiers	In this method, the data space is modeled with a set of rules. The left hand side denotes a condition on the feature set expressed in disjunctive normal form and in the right hand side represents the class label. The conditions are on the term presence. Term absence infrequently used because it is not informative in sparse data**.**

In the conditions that are impossible or difficult to find training documents (such as the labor-intensive and time-consuming process), the unsupervised methods are applied. A dictio-based process can be referred as an unsupervised method. This technique considers a dictionary with sentiment words and phrases and their associated orientations and strength and incorporates intensification and negation to compute a sentiment score for each document (Taboada, Brooke, Tofiloski, Voll & Stede, 2011).Hybrid method uses statistical/machine learning and lexicon-based techniques to increase the performance and precision of sentiment analysis task as applied in this work.

Discriminative learning methods are widely used in natural language processing. These methods can achieve acceptable accuracy when their training and test data are drawn from the same distribution. However, we face new domains in which labeled data is scarce or non-existent for many NLP tasks. We seek to adapt existing models from a resource-rich source domain to a resource-poor target domain in such cases. In this work, we used structural correspondence learning to induce correspondence between features from different domains automatically. SCL is a general technique, which one can apply to feature based classifiers for any task. The key idea of SCL is to detect correspondences among features from different domains using a model to present correlations of other features with the pivot features. Pivot features are features which behave in the same way for discriminative learning in both domains. Non-pivot features from different domains which are correlated with many of the same pivot features are assumed to correspond, and we treat them similarly in a discriminative learner Blitzer, Dredze & Pereira, 2007; Zhang & Singh, 2014).

Structural learning models the correlations which are most useful for semi-supervised learning. It can be adapted for transfer learning too and consequently the structural part of structural correspondence learning is borrowed from it (Blitzer, McDonald & Pereira, 2006).

Deep learning is one of the most important recent approaches in Artificial Intelligent (AI) field that can be effectively applied for sentiment analysis.

Deep Learning can be considered as a new field in machine learning which is capable to improve the computer systems by knowledge and data. Some skills are needed to obtain the best achievement in deep learning process. With increasing the training data, the amount of skill required will be decreased (Goodfellow, Bengio, Courville & Bengio, 2016).

Deep learning can be used for providing training documents in supervised, semi-supervised and unsupervised categories. It is very applicable in different areas such as NLP, computer vision and neural sciences. Therefore, it can be considered as a new technique for precise sentiment analysis field. Deep learning could be created by many networks such as CNN (Convolutional Neural Networks), RNN (Recurrent Neural Networks), Recursive Neural

Networks, DBN (Deep Belief Networks) and many more. Convolutional layers in CNN play the role of feature extractor, which extracts local features as they restrict the receptive fields of the hidden layers to be local. It means that CNN has a special spatially local correlation by enforcing a local connectivity pattern between neurons of adjacent layers. Such a characteristic is useful for classification in NLP, in which we expect to find strong local clues regarding class membership, but these clues can appear in different places in the input. For example, in a document classification task, a single key phrase (or an n-gram) can help in determining the topic of the document.

The following subjects can be referred as the main elements of Deep learning: enhanced software engineering, improved learning procedures, availability of power computing and data training. All of these elements are combined together to create Deep learning method (Ain et al., 2017).

Although many studies have been accomplished in sentiment analysis area for the English language, but the research in this field for the Persian language is ongoing (Hosseini et al., 2018). Therefore, due to the nature of the Persian language, it is essential to conduct research in this regard.

In this research, the sentiment classification process for the Persian language is performed by deep learning method. The Persian corpus for this study is SentiPers that includes more than 26K Persian sentences. We present an application of structural correspondence learning (SCL) to automatically persuade correspondences among features from different domains (Blitzer et al., 2006). We applied this method as a method for domain adaptation in NLP for adaptive learning sentiments in the Persian language.

In this paper, the related works in sentiment analysis and deep learning are presented in section 2. The proposed method is then reviewed in Section 3. The performed experiments and data description for this study with the obtained results are discussed in section 4. Finally, at section 5, the conclusion of this study is discussed.

Related Works

Neural Networks have the main role in many types of research that are accomplished in sentiment classification. For Instance, in one research project a DyCNN (Dynamic Convolutional Neural Network) which use a dynamic k-max pooling on linear sequences operation (Kalchbrenner, Grefenstette & Blunsom, 2014). A study proposed a neural network architecture called ConvLstm that consist of Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) on top of pre-trained word vectors. They showed that ConvLstm uses the LSTM as a replacement of pooling layer in CNN to decrease the lack of detailed local information and capture long-term dependencies in the sequence of sentences. Their model approved in two sentiment datasets IMDB, and Stanford Sentiment Treebank (SSTb). Experimental results show that ConvLstm attained comparable performances with fewer parameters in sentiment analysis tasks (Hassan and Mahmood, 2017). Kim (2014) implemented a CNN to learn sentiment-bearing sentence vectors. Additionally, Paragraph vector is suggested by Mikolov, Sutskever, Chen, Corrado & Dean (2013). Their method has a better sentiment analysis performance compare to bag-of-words model and ConvNets for learning the SSWE (sentiment specific word embedding). Tang, Wei, Qin, Liu & Zhou (2014) created a deep learning system called Coooolll for message-level Twitter sentiment classification. This system has two features as state-of-the-art hand-crafted properties and the sentiment-specific word

embedding (SSWE) features. The SSWE is trained by 10M tweets with positive and negative emotions, without any manual annotation. They claimed that the effectiveness of Coooolll has been verified in both positive/negative/neutral and positive/negative classification of tweets.

The semi-supervised learning algorithm called active deep network (ADN) is proposed in a study. This algorithm implemented by restricted Boltzmann machines (RBM) with unsupervised learning that used labeled reviews and abundant of unlabeled reviews. In order to identify reviews that should be identifying as training data, they applied active learning, after that they train ADN by using the selected labeled reviews and entire unlabeled reviews. They also combined ADN with information density to propose information ADN (IADN) method that can be applied the information density of all unlabeled reviews in the manual selected labeled reviews (Zhou, Chen & Wang, 2013). In another research, a Tweet corpus as part of their inaugural Sentiment Analysis in Twitter Task called SemEval is implemented (Rosenthal, Farra & Nakov 2019). It includes tweets and SMS with sentiment expressions commented by contextual phrase-level and message level polarity and LiveJournal sentences. Their test set consist of an in-domain Twitter dataset, an out-of-domain LiveJournal test set, and a dataset of tweets containing sarcastic content. They show that the performance on the LiveJournal test set is comparable to the in-domain Twitter test set, they also show that the performance decreases for sarcastic tweets. They discussed management of sarcastic language that has a potential to be one of the essential direction for the upcoming subject in Twitter sentiment analysis. Fernández-Gavilanes, Álvarez-López, Juncal-Martínez, Costa-Montenegro & González-Castaño (2016) created an approach to sentiment prediction in online textual messages such as tweets, SMS, and reviews. Their method is based on an unsupervised method. Their approach does not need to use the labeled text in pre-training phase. The base of their approach is to determine the dependencies between lemmatized labeled words with a sentiment Propagation algorithm that considered and distinguished between key linguistic occurrences.

Few studies have been done to analyze the emotions in the Persian language domain. One of the reasons for this boundary is difficult to access the linguistic resources and limitation in Persian language processing tools. One of the fundamental researches that are accomplished in this field is the production of a Sentipers (Hosseini et al., 2018), which is also used in this research as a test corpus. This corpus contains more than 26 thousand Persian sentences and includes of sentences in the Formal and colloquial language. In this corpus, the semantic for each sentence is computed and marked. The punctuation for each sentence is applied by a human, therefore it has a high degree of accuracy. The scope of each sentence is also labeled by keywords. Vaziripour, Giraud-Carrier & Zappala, (2016) explored the automatic classification of Persian tweets. In their research, the lack of Persian language resources in the processing of Persian texts for the analysis of emotion is revealed. They collected over one million Persian tweets in political contexts using Twitter's infrastructure functions and obtained a relative accuracy of 56% using a backup machine classifier.

Roshanfekr, Khadivi & Rahmati (2017) considered two deep neural network architectures for document classification based on sentiment polarity. They show that the deep learning method has the better performance such as F-score in compare to Naive Bayes *SVM (NBSVM) model.* One reason could be the use of word vector representations which is done in an unsupervised way. The learning step addresses most of the difficulties arise from different writing styles in Persian and since the learning is unsupervised it doesn't suffer from lack of the dataset. Cross-lingual adaptation is an example of using domain adaptation that denotes the

transfer of classification data between two languages. Prettenhofer and Stain (2011) designated an extension of Structural Correspondence Learning (SCL) for cross-lingual adaptation in the context of text classification. They used unlabeled documents from different languages, along with a word translation oracle, to persuade a cross-lingual representation that enables the transfer of classification data from the source to the target language.*In another study, the* Structural Correspondence Learning compared with different options of self-training for adaptation of a parse selection model to Wikipedia domains. The study indicates that none of the assessed self-training variants accomplishes an important progress over the baseline. The more indirect corruption of unlabeled data through SCL is more effective than solid self-training (Plank, 2009).

Proposed Method

Lack of NLP resources and tools like machine-readable Persian corpora, lexicon, and text analyzing tools can affect researches in the domain of Persian sentiment analysis. A common way to classify sentiment data is to learn a classifier and then use it to classify sentiment data. The open challenge in this way is that the learned classifier is domain-dependent. The learned classifier may have a good performance on one domain and low performance on other domains. Resource and data dependency can affect the performance of classifiers because the characteristics of objects can be changed over time.

The inadequacy of resources and sentiment data forced Persian domain researchers to use machine learning methods. This group of methods needs fewer resources than statistical methods. The commonly used classifiers like Naïve Bayes and SVM that have a good performance on the learned domain will not have a good performance on other domains (Blitzer et al., 2006; Tahmoresnezhad & Hashemi, 2017). It is important to declare a similarity measure for domains to compare them and select similar domains. In recent work, some similarity measures are introduced but there is no measure that all of the researchers accept it (Blitzer et al., 2007; Blitzer et al., 2006; Tahmoresnezhad & Hashemi, 2017).

In this work, we need to use domain adaptation in the scope of text and sentiment analysis. Structural Correspondence Learning (SCL) can automatically induce correspondences among features from different domains (Blitzer et al., 2006). This method is used for adaptive learning sentiments in English and Persian (Blitzer et al., 2006; Dastgheib, 2017).

The main idea of SCL is to detect and highlight correspondences among features from different domains by modeling their correlations with pivot features. Pivot features are features that behave in the same way for discriminative learning in both domains. Non-pivot features from different domains that are correlated with many of the same pivot features are assumed to correspond (Blitzer et al., 2006).

Some key features in similar domains are different. For example, the features of a good mobile phone cannot be applied to select a laptop (Dastgheib, 2017). The pivot features can be applied in both domains (Blitzer et al., 2007). For example features like excellent and unrivaled are common in both domains. But the features like the speed of hard disk or DVD drive are not applicable to the mobile domain. These uncorrelated independent features are correlated to pivot features in unlabeled data. So these independent features can be matched by this property (Blitzer et al., 2007; Dastgheib, 2017). After extracting these features, there are words in one domain that are tantamount to words in a similar domain (i.e. high-speed data received in the domain of mobile phones is equivalent to fast hard disk in the domain of laptop computers). So

the classifier that is learned in one domain can be used for similar domains with a good performance. To extract feature sets we follow the Blitzer method (Blitzer et al., 2006). Firstly *m* pivot features that have the high frequency in domains are extracted. Secondly, the covariance between these pivot features and non-pivot features must be estimated accurately. But these extracted features must be different enough to determine the nuance of the items. Finally, a classifier will be trained using these extracted data.

The features will be represented as binary features. Each bit will specify the existence of pivot feature in an instance. So *m* linear predictors are needed to solve the vector (X) for pivot features. These predictors use original feature space to calculate the vector X. Figure 2 shows the SCL algorithm for extracting features.

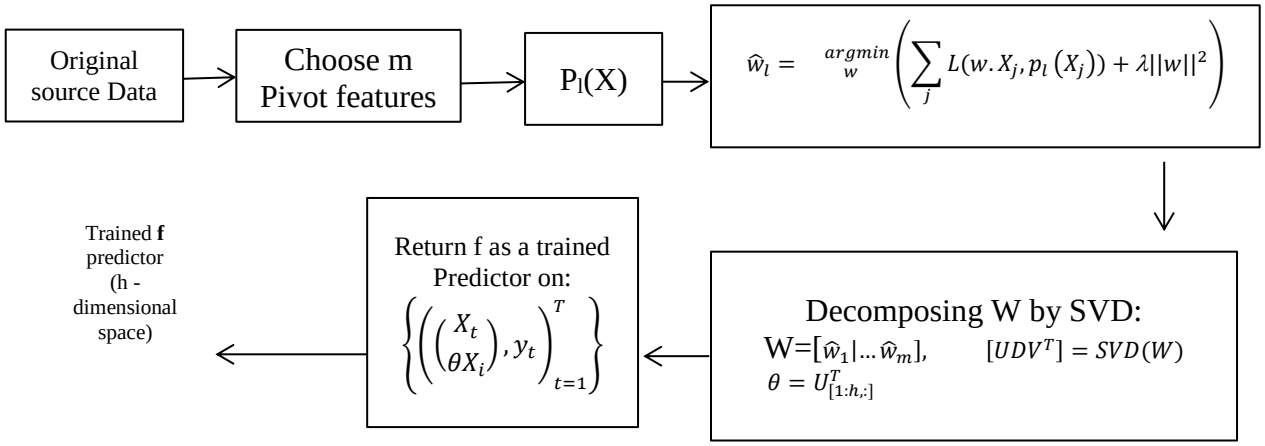

Figure 2. SCL algorithm for extracting feature vectors (Blitzer et al., 2006)

As shown in fig 2, the algorithm uses labeled source data (x_t,y_t) such that t=1..T. Also, the unlabeled data from both domains is shown by x_j. The Output of the SCL algorithm is trained predictor (f) on both domains. In the first step of the algorithm, *m* pivot features are selected from the original source data (p_l shows the l-th pivot feature). It is common to define penalty of estimation by a loss function. In this work, we use modified Huber loss function taken from (Ando & Zhang, 2005) and (Blitzer et al. 2006). As shown in step 4 of fig 2, L(p,y) is real-valued loss function for binary classifications. In Eq. (1) Weight vectors $\widehat{w}_l$ calculate the covariance pf non-pivot features with the pivot features (Blitzer et al., 2006) (Blitzer et al., 2006). If the weight given to i-th feature by l-th pivot predictor is positive, then feature z is positively correlated with pivot feature l (Blitzer et al., 2006).

$$\widehat{w}_l = \underset{w}{argmin}\left(\sum_j L(w.X_j, p_l(X_j)) + \lambda||w||^2\right) \quad (1)$$

Pivot features are frequent in both domains so non-pivot features in both domains are correlated with pivot features. Now we can use these positive correlations to estimate similar features. If two non-pivot features are correlated with many pivot features we can conclude that they have high similarity. So we have a linear projection ($\widehat{w}_l$) of original feature space onto $\Re$.

Each pivot feature is represented by *m* real-valued features in onto $\Re$. For computational and statistical reasons we estimate this new feature space by low rank approximation (Blitzer et al., 2006). As shown in step 5 in fig 2 and Eq. (2), if W be a column matrix of pivot features weight then it can be decomposed by singular value decomposition (SVD). In Eq. (2), UDV^T is the decomposition of W and θ is a row matrix of the top left singular vectors of W.

$$W=[\widehat{w}_1|\dots\widehat{w}_m], \qquad [UDV^T] = SVD(W) \quad \text{where} \quad \theta = U^T_{[1:h,:]} \tag{2}$$

The rows of θ are the principal Eigenvalues of pivot predictors in *h* dimensional space. This approximation is a projection from original features space onto $\Re^h$. at the last step of fig 2, θX_i is the mapping to the low dimensional space. After domain adaptation using SCL algorithm, original feature space transforms to reduced *h* dimensional space of pivot features. The next step is to train a classifier with this transformed reduced feature space. To do this we used Convolutional Neural Network (CNN). The CNN is configured to classify sentiment of sentences by features that are modeled by SCL. Each sentence is represented in *h* dimensional space by its pivot features. So a sentence represented by a matrix ($d \times h$), such that *d* denotes the number of pivot features in this sentence.

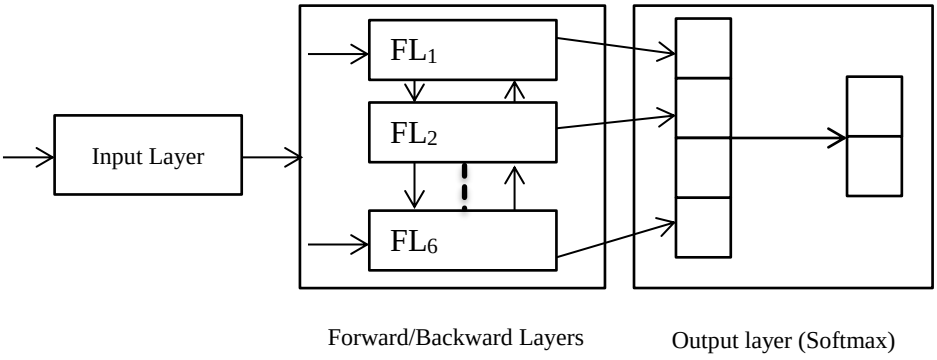

Figure 3. The CNN structure

As shown in Figure 3, the CNN structure used here has an input layer ($d \times h$), 6 hidden forward/backward layers and an output (Softmax) layer acts as classification layer. The output Softmax layer performs a probability distribution over the class labels. This CNN structure is modeled by Microsoft CNTK[2]. It is common to apply a dropout after hidden layers to prevent over fitting problem. This mean of regularization module blockages co-adaptation of hidden units (Roshanfekr, et al., 2017).

Experiments and Results

As mentioned in section 3, to classify sentiment of sentences we proposed a hybrid algorithm that uses domain adaptation by SCL and convolutional neural network for sentiment analysis classification.

Dataset used in this work is taken from SentiPers that developed by (Hosseini et al., 2018). This corpus contains more than 26K Persian sentences of users opinions from digital product domain and benefits from special characteristics such as quantifying the positiveness or negativity of an opinion through assigning a number within a specific range to any given

sentence. The sentences are labeled manually. These sentences are gathered from Digikala[3]. Each appliance in SentiPers contains a set of characteristics and sentences. The content of sentences is users opinion about electronic goods. The sentences are classified in positive, negative and neutral classes. A number in the interval [-2,0,+2] is assigned to each opinion that presents the score of the sentence in positive or negative classes. Table 2 shows an example of sentences with their sentiment scores.

Table 2
An example of SentiPers corpus sentences (iphone 4)

Score	Class	Sentence in Persian
0	Neutral	سلام من تست تمام گوشی های هوشمند برتر بازار رو خواندم.
+1	Positive	از لحاظ گرافیک و سرعت بی نظیره
-1	Negative	نسبت به قیمت و حافظه‌اش ارزش خرید ندارد.

In this work a subset of whole dataset is used in two domains (Mobile phones and laptop computers). This selected dataset has about 1K sentiment sentences on each domain. The sentences are labeled in three categories (for example, 742 positive, 181 negative and 81 nutral sentences). The average length of sentences in this dataset is 18.5 words and the dataset contains more than 500 distinct words.

To compute the sentiment score of a sentence, a pre-processing step is required to separate formal and informal sentences and normalize Arabic characters to the Persian equivallant. It is common in lexicon-based methods to estimate the polarity of the sentence using polarity of it's words. Also machine learning methods need some features that must be extracted from the sentences. These features also extracted from sentiment of desired words in the sentence. So LexiPres lexicon (Sabeti, Hosseini, Ghassemi-Sani, & Mirroshandel, 2016) is used here to indicate the polatity of the words. This lexicon has more than 25K words. Each word has a polarity score. LexiPers lexicon has more than 20K synsets in three categories (posive, nutral and negative).

To set up the experiment, we chose laptop and mobile phone categories that are similar and can be adapted by the hybrid proposed algorithm. For each category, we selected 10 items randomly. This set contains about 1K labeled sentences in two selected domains. This collection is divided into a training set and a test set. The training set has about 800 labeled sentences and in a similar manner, the test set has about 200 sentences.

Firstly, we apply the SCL algorithm to the collection as shown in fig 2. After choosing pivot features and computing singular value decomposition (SVD) the dimension of data will be reduced to h-dimensional space.. In this work, the calculation of SVD was done on a dual Xeon machine and took about three hours for the dataset. Secondly, the reduced features matrix will use to train CNN network (C.F. fig. 3). To implement CNN, Microsoft cognitive toolkit[4] is used. As shown in fig. 3, we used an input layer, 10 forward-backward layers, and an output (softmax) layer. The training process converged in 30 steps on a core i7 computer with GPU. Finally, the trained network used to determine the sentiment class of test set. To compare the obtained results with the other methods, we used traditional Naïve-Bayes classifier as a baseline. Also, the experiment is repeated without using SCL algorithm and using original data without reducing the dimension. Also, we test the results by training in one domain and test in

another domain. The results of these three experiments are shown in Figure 4, Table 3 and Table 4.

Table 3
Accuracy of models on SentiPres data

Method	M->L	L->M
Naïve-Bayes	42.5	40.5
CNN	54.3	51.7
CNN+SCL	64.5	63.1

Table 4
F-measure of models on SentiPers data

Method	M->L	L->M
Naïve-Bayes	0.55	0.53
CNN	0.59	0.56
CNN+SCL	0.74	0.71

In fig.4 and tables 3-4, L->M denotes that training set is selected from laptop computers domain and test is done on mobile phones domain. In a similar way, M->L is presented training the network on mobile phones domain and testing on laptop computers domain.
The first bar in fig. 4 shows the accuracy of the baseline (Naïve-Bayes classifier) in two domains. The second bar represents the accuracy of CNN without feature extraction and dimension reduction (SCL method) and finally, the last bar shows the accuracy of proposed method that is the combination of SCL and CNN.

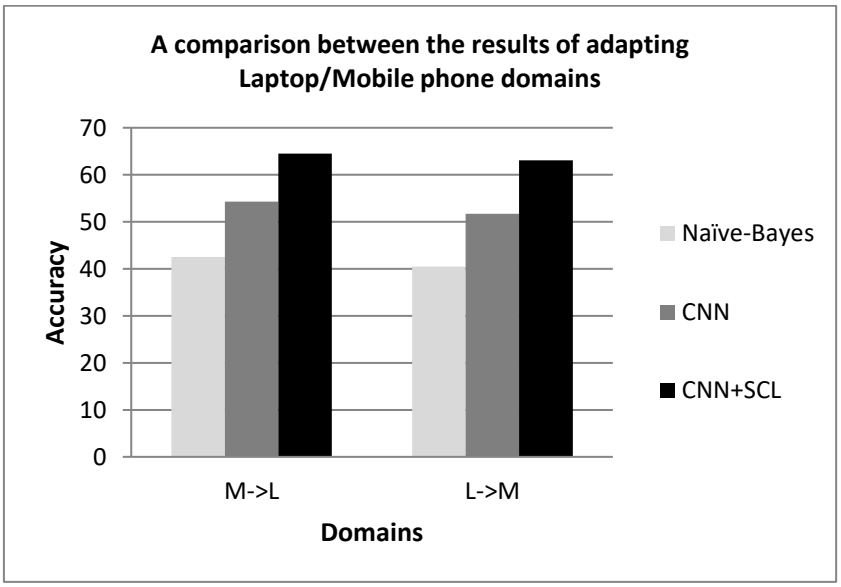

Figure 4. The results of adapting laptop/mobile domains in sentiment analysis

As shown in table 4, the proposed hybrid method (CNN+SCL) has the maximum F-measure score. F-measure considers both precision and recall in its calculations. Higher F-measure score is obtained in the event that precision and recall are both high. As mentioned in the results, the F-meaure point of CNN+SCL in both adapted domains are dominant. This score points to higher performance of sentiment classification.

The dataset used in this work is unbalanced. The negative samples are less than positive samples in the SentiPers dataset. This situation explains the low performance of traditional methods in the negative class. The deep learning can reach better performance because it can do better generalization on negative samples. The confusion matrix of used models is presented in tables 5-7. From these experiments, we conclude that the deep learning method (CNN) had better performance than the traditional naïve Bayes classifier. Classification of negative samples is a challenge for the model. One of the important reasons for the low performance in the classification of negative samples is unbalanced data in the training step. But despite these criteria, CNN can perform more generalization and had better performance.

Table 5
Confusion matric for Naïve Bayes (M->L)

Naïve Bayes	True Positive	TrueNegative
Predicted Positive	73	72
Predicted Negative	43	12

Table 6
Confusion matric for CNN (M->L)

CNN	True Positive	TrueNegative
Predicted Positive	84	58
Predicted Negative	32	23

Table 7
Confusion matric for CNN+SCL (M->L)

CNN+SCL	True Positive	TrueNegative
Predicted Positive	98	46
Predicted Negative	24	29

Also, table 8 gives an example of extracted pivot features. Similarly, some non-pivot features are also presented in table 9 from two domains. In our proposed method the top 400 meaningful words are considered as a set of pivot features. These pivot features are seen in two adapted domains. Using these pivot features can reduce the noise and will increase the accuracy of the classifier (CNN). The comparison between traditional CNN and proposed CNN+SCL method showed that reducing features using pivot features in domains by SCL is effective and can increase the accuracy and more generalization of negative samples.

Table 8
An example for extracted pivot features

Extracted pivot feature	Meaning in English
عالی	Excelent
وضوح بالا	high resolution
کیفیت	Quality

Table 9
An example for extracted non-pivot features (two domains)

Extracted pivot feature	Meaning in English
تاچ خازنی	Capacitive touch
صفحه گلس	Glass screen
دو سیم کارت	Dual-sim support

Training model with extracted pivot features creates an adaptable model that can be used in similar domains (EXP: mobile phone to laptop computers) with good performance and accuracy. These pivot features are more stable when we change from one domain to another. This stability of pivot features makes an adaptable model that can learn from one domain data and can be used in another domain. If the most of selected pivot features in source domain were found in destination domain, the model can work with minimum fault. One problem of using unbalanced training data is that the accuracy of the classifier is low. This happens due to many negative samples in the training dataset. Deep learning can achieve better performance on unbalanced data because it can produce a model with better generalization from negative samples (Roshanfekr et al., 2017). One of the problems in research in the field of Persian NLP is the lack of sufficient resources. So it is inevitable to use trained model for more than one domain. In such area, adaptation using SCL is a good solution for multi-domain Persian sentiment analysis.

Conclusion

In this work, we proposed a hybrid method using transfer learning from one domain to similar ones efficiently. Due to a variety of similar domains, it is acceptable to have a classifier that has been learned in one domain and can act in a similar domain with good performance. The SCL algorithms try to investigate pivot features that are constant in similar domains. This method used to extract pivot features that are constant in the adapted domain. Using these features can boost up the performance of the classifier in a multi-domain area. Imbalanced data on negative samples causes a generalization problem for the model. Using a deep learning method can tackle the problem by better generalization using minimum training samples. The CNN model is trained using extracted features by the SCL algorithm. Results showed that the proposed hybrid method that uses extracted pivot features can work batter than traditional Naïve Bayes and also CNN method. The accuracy and F-score will be dropped in the traditional method but using pivot features the performance of the classifier preserved in a domain adaption process. This method for dominant languages like Persian that suffer from lack of labeled data (training dataset) is helpful. CNN model that has better generalization on negative samples, better f-score, and accuracy. For future work, the authors propose that the hybrid model will be tested on more data-sets. Another task is to use this model to produce a large Persian sentiment corpus.

Endnotes

1. From Wikipedia data
2. https://docs.microsoft.com/en-us/cognitive-toolkit
3. Digikala is the biggest ecommerce startup in Iran
4. https://www.microsoft.com/en-us/cognitive-toolkit/

References

Ain, Q. T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., & Rehman, A. (2017). Sentiment analysis using deep learning techniques: a review. *International Journal of Advanced Computer Science and Applications*, *8*(6), 424.

Ando, R. K., & Zhang, T. (2005). A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, *6*(Nov), 1817-1853.

Blitzer, J., Dredze, M., & Pereira, F. (2007). Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 440-447).

Blitzer, J., McDonald, R., & Pereira, F. (2006). Domain adaptation with structural correspondence learning. In *Proceedings of the 2006 conference on empirical methods in natural language processing* (pp. 120-128). Association for Computational Linguistics.

Dastgheib M.B. (2017). Persian sentiment analysis using domain adaptation by Structural Correspondence Learning. In *Proceedings of the second national conference on computer and electronics*, Basir university of Tehran, Tehran, Iran. [in Persian]

Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications*, *58*, 57-75.

Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning*. Vol. 1. Cambridge: MIT press.

Hassan, A., & Mahmood, A. (2017). Deep Learning approach for sentiment analysis of short texts. In *Control, Automation and Robotics (ICCAR), 2017 3rd International Conference on* (pp. 705-710). IEEE.

Hosseini, P., Ramaki, A. A., Maleki, H., Anvari, M., & Mirroshandel, S. A. (2018). SentiPers: A sentiment analysis corpus for Persian. *arXiv preprint arXiv:1801.07737*.

Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.

Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, *5*(4), 1093-1113.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1-135.

Plank, B. (2009). A comparison of structural correspondence learning and self-training for discriminative parse selection. In *Proceedings of the NAACL HLT 2009 Workshop on Semi-supervised Learning for Natural Language Processing* (pp. 37-42). Association for Computational Linguistics.

Prettenhofer, P., & Stein, B. (2011). Cross-lingual adaptation using structural correspondence learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *3*(1), 13.

Rosenthal, S., Farra, N., & Nakov, P. (2019). SemEval-2017 task 4: Sentiment analysis in Twitter. *arXiv preprint arXiv:1912.00741*.

Roshanfekr, B., Khadivi, S., & Rahmati, M. (2017). Sentiment analysis using deep learning on Persian texts. In *Iranian Conference on Electrical Engineering (ICEE) 2017* (pp. 1503-1508). IEEE.

Sabeti, B., Hosseini, P., Ghassem-Sani, G., & Mirroshandel, S. A. (2016). LexiPers: An ontology based sentiment lexicon for Persian. In *2nd Global Conference on Artificial Intelligence (GCAI).* Vol. 41 (pp. 329-339). Berlin, Germany

Shahnawaz, & Astya, P. (2017). Sentiment analysis: Approaches and open issues. *2017 International Conference on Computing, Communication and Automation (ICCCA),* (pp.154-158) IEEE.

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, *37*(2), 267-307.

Tahmoresnezhad, J., & Hashemi, S. (2017). Visual domain adaptation via transfer feature learning. *Knowledge and Information Systems*, *50*(2), 585-605.

Tang, D., Wei, F., Qin, B., Liu, T., & Zhou, M. (2014). Coooolll: A deep learning system for twitter sentiment classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (pp. 208-212).

Vaziripour, E., Giraud-Carrier, C. G., & Zappala, D. (2016). Analyzing the Political Sentiment of Tweets in Farsi. In *Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)* (pp. 699-702).

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), e1253.

Zhang, Z., & Singh, M. P. (2014). Renew: A semi-supervised framework for generating domain-specific lexicons and sentiment analysis. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 542-551).

Zhou, S., Chen, Q., & Wang, X. (2013). Active deep learning method for semi-supervised sentiment classification. *Neurocomputing*, *120*, 536-546.