Journal of Language Research
ZABANPAZHUHI
Quarterly Scientific Journal of Language Research, Vol. 15, No. 49, Winter 2024
Alzahra University, http://jlr.alzahra.ac.ir
Article Type: Research
pp.31-59

Contextualized Text Representation Using Latent Topics for Classifying Scientific Papers[1]

Maryam Mousavian[2]
Masood Ghayoomi[3]

Received: 09/08/2023
Accepted: 19/11/2023

Abstract

Annually, researchers in various scientific fields publish their research results as technical reports or articles in proceedings or journals. The collocation of this type of data is used by search engines and digital libraries to search and access research publications, which usually retrieve related articles based on the query keywords instead of the article's subjects. Consequently, accurate classification of scientific articles can increase the quality of users' searches when seeking a scientific document in databases. The primary purpose of this paper is to provide a classification model to determine the scope of scientific articles. To this end, we proposed a model which uses the enriched contextualized knowledge of Persian articles through distributional semantics. Accordingly, identifying the specific field of each document and defining its domain by prominent enriched knowledge enhances the accuracy of scientific articles' classification. To reach the goal, we enriched the contextualized embedding models, either ParsBERT or XLM-RoBERTa, with the latent topics to train a multilayer perceptron model. According to the experimental results, overall performance of the ParsBERT-NMF-1HT was 72.37% (macro) and 75.21% (micro) according to F-measure, with a statistical significance compared to the baseline (p<0.05).

Keywords: Article Content Analysis, Contextualized Representation, Distributional Semantics, Neural Network, Scientific Article Classification, Topic Modeling

---

[1] DOI: 10.22051/jlr.2023.44640.2331
[2] Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran; maryam.mousavian@aut.ac.ir;
ORCID: https://orcid.org/0000-0002-5053-2377
[3] Faculty of Linguistics, Institute for Humanities and Cultural Studies, Tehran, Iran. (corresponding author); m.ghayoomi@ihcs.ac.ir
ORCID: https://orcid.org/0000-0001-6685-1332

1. Introduction

Nowadays, a large volume of scientific papers are published in print and in the electronic format in different countries. This highlights the need to pursue using information science for a pervasive effect on policy making organizations in science. Information science "investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information" (Borko, 1968). One of the tasks in information science is the classification of sciences to make it possible to draw the road map in the field. Since there are different fields of science, accurate classification will be a tough task, especially when dealing with interdisciplinary or multidisciplinary scientific papers. Classification of documents is one of the old tasks of librarians, but due to fast dissemination of research-based articles, this task cannot be done manually any longer. This fact will be severe when a repository of article archives, such as Scopus, contains a huge number of articles. To this end, machine learning methods can be beneficial and pave the ground to reach the goal.

This paper aims at proposing a content-based classification model that takes the advantage of contextualized text representation in a deep neural network model to classify Persian articles.

2. Research Background

Although the classification of scientific articles is, in general, a text classification task, not much research has been done in this domain, specifically to classify scientific papers.

Kim & Gil (2019) used the Term-Frequency and Inverse of the Document-Frequency (TF-IDF) vectorization method (Salton, 1971) and Latent Dirichlet Allocation (LDA) (Blei et al., 2003) for clustering scientific articles. Their proposed method benefited from three kinds of information: users' input, abstracts' keywords, and keywords of the topics extracted by LDA. The

extracted keywords were used to vectorize each paper with TF-IDF and to cluster the papers by the K-means clustering algorithm (MacQueen, 1967), where optimal value of K obtained by the Elbow scheme. The K-means clustering algorithm used in this paper calculates the center of the cluster that represents a group of papers with a specific topic and allocates a paper to a cluster with high similarity, based on the Euclidean distance between the TF-IDF value of the paper and the center value of each cluster. In this research, they utilized 3264 papers published in Future Generation Computer System journal from 1984 to 2017.

To classify scientific papers, Chowdhury & Schoen (2020) evaluated the performance of common machine learning classification algorithms, such as Support Vector Machine (SVM), Naive Bayes, decision tree, and K-Nearest Neighbor (KNN). Their dataset contained 107 abstracts, collected from the abstracts of the articles in science, business, and social sciences. The SVM algorithm performed the best with the 89.5% F1-score.

Rivest et al. (2021) classified 40 million scientific articles using character-based convolutional deep neural networks. They used additional features, such as title, keywords, and authors' affiliation, in the model. Based on the results, simple features, such as direct reference and bibliographic information, had the most significant impact on common classification algorithms and neural networks, respectively.

Mustafa et al. (2021) evaluated the proposed framework using two diversified datasets. One of them is based on research publications from the Journal of Universal Computer Science and another one contains research publications from the Association of Computing Machinery. In their proposed method, metadata, like title and keywords, was extracted from the documents. The metadata's Word2Vec representation (Mikolov et al., 2013) and similarity calculation were the base of their proposed model. Their research experiments resulted in the best average accuracy by 86%.

The above reviewed papers used English documents to build the models and to classify. There are a number of researches accomplished for classification of scientific articles in other languages, such as Persian. The

Persian language, which is mostly considered as a low-resource language, is not much considered in research in comparison to high resource languages, such as English. In the rest of this section, the articles focused on Persian scientific papers are reviewed.

EmamiAzadi & AlmasGanj (2006) used the Probabilistic Latent Semantic Analysis (PLSA) method (Hofmann, 1999) and evaluated the Persian scientific article classification by Farsdat dataset (Bijankhan et al., 1994), including 6 different subjects. They used the authors' specifications to improve the model. The proposed method has enhanced the PLSA model by eliminating inappropriate hidden variables during training.

Teymoorpoor et al. (2009) performed the classification of the indexed articles in International Scientific Indexing in the field of nanotechnology by using an unsupervised model based on information retrieval. The evaluation dataset consisted of 1990 articles from 2003 to 2009. In this research, the hierarchical classification of the nano tree was used. Each article was an observation and each node at a specific level of the nano tree was a label. First, the articles were initially assigned to their class using the nano tree. Then, articles, which were not classified in the initial classification, were categorized using the TF-IDF vector space model and the cosine similarity metric.

Karami et al. (2018) introduced a fuzzy model, named Fuzzy Latent Semantic Analysis, as an approach in topic modeling to estimate the number of topics. They used five different datasets on health and medical research in their experiments, namely MuchMore Springer Bilingual Corpus, nursing notes, Ohsumed collection, Twitter health news, and subsets of the Wall Street Journal dataset. In their experimental results, 69% to 75% of F-measure were obtained.

Rabiei et al. (2019) classified environmental research articles using the SVM algorithm. In this research, they introduced a new method for weighting when constructing vectors that can be used for discovering the representative terms of scientific domains. The data used in this research was 16,626 documents related to the environment field, which have been received from doctoral and master theses archived in Irandoc, the organization of managing scientific articles in Iran.

Shokouhian et al. (2020) presented a hybrid supervised and unsupervised learning model to classify scientific articles thematically in the field of health. To conduct this research, they prepared scientific papers in health from the PubMed database, from 2009 to 2019. They clustered and labeled documents using LDA. Eventually, they utilized vectors obtained from the LDA model to train the SVM classifier.

Ghayoomi & Mousavian (2022) performed research on classifying Persian scientific papers. In addition to the basic classification machine learning algorithms, they proposed using perceptron and convolutional neural networks as well as static representation, Word2Vec (Mikolov et al., 2013), and dynamic representation, ParsBERT (Farahani et al., 2021), for classifying the scientific articles in humanities. The ParsBERT representation with the perceptron model obtained the best result.

3. Contextualized Text Representation

According to the distributional hypothesis, meaning is determined by context, and words that appear in a similar context tend to have similar meanings (Harris, 1954). Hence, contexts of a word have been introduced as an intermediate way to represent semantic meanings. Examples (1) to (4) show that similar contexts imply that the meaning of the words 'auto', 'automobile', 'car', and 'vehicle' are similar.

1. John drives an auto.
2. John drives an automobile.
3. John drives a car.
4. John drives a vehicle.
5. John has a car.
6. John fixed his car.

Vector representation is a semantic representation method (Jurafsky & Martin, 2000). In this method, the semantic properties of words are represented numerically in a vector space which has magnitude and direction. In addition, a vector space is characterized by dimension, and it is possible to use mathematical operations in linear algebra, such as addition or

multiplication, to add or multiply vectors.

The recent attempt by Mikolov et al. (2013) explored contextualized semantic properties proposed by Harris (1954) in a vector space, called word embedding. The result of this method was proposing a single, static vector representation of a word appeared in different contexts. Similar contexts in examples (1) to (4) cause to have similar vectors for the words 'auto', 'automobile', 'car', and 'vehicle'. Since contexts in examples (3), (5) and (6) are different, only one vector will be created.

In addition to this static word embedding method, contextualized word embeddings was proposed by Devlin et al. (2019) to represent different vectors for a polysemous word in a given local context; as a result, the word's meaning and the local context of the target word are reflected in the contextualized vector representation. In examples (7) and (8), we have two vectors for the 'bank'.

7. John walked in the bank.
8. John walked by the bank of the river.

This allows downstream tasks to model a natural language more realistically. Based on Harris theory, the languages of special domains have structures and regularities that can be observed by analyzing the corpora of these domains.

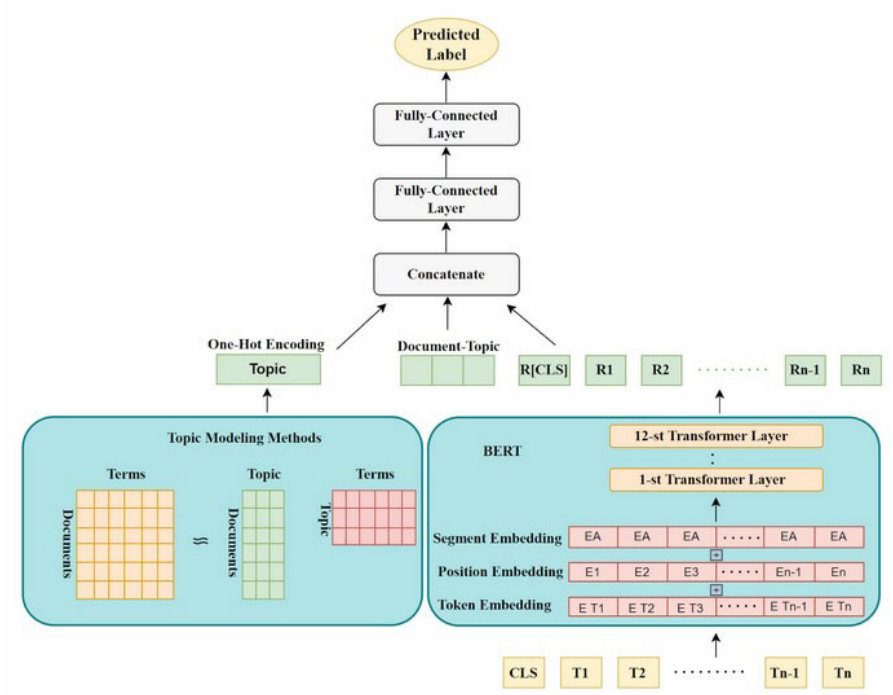
In the current research, we used contextualized representation to learn the structures and word meanings of articles in Humanities, and boosted the performance of the model with the semantic representation of topic modeling methods.

4. Proposed Model

This paper aims at assigning a label to the articles to identify their fields of study. The architecture of our proposed model is described in Figure 1. As it is shown in the figure, the model contains two main modules, namely the representation module and the classification module described in the following sections.

Figure 1

The architecture of the model



4.1. Representation Module

The purpose of developing and implementing this module is to access specific and hidden knowledge in each distinctive domain. The representation of each article in the proposed model contains three different types of representations: 1) contextualized representation of transformer-based language models; 2) probabilistic semantic distribution of articles using topic modeling methods; and 3) one-hot encoding vector of each article's subject. To determine which articles are more thematically similar, we use the topic number of each article as a one-hot vector, which is acquired from the topic modeling algorithms. Each of these three representations contains valuable and diverse information in the dataset. The integration of the concatenated information provides deeper knowledge for each article.

4.1.1. ParsBERT

Transformer-based language models are prevalent among the pre-trained language models since the models obtained the state-of-the-art results.

One of these models is the Bidirectional Encoder Representations from Transformers (BERT) model proposed by Devlin et al. (2019). This model, which is created based on the bidirectional transfer model, supports non-English languages in its multilingual model. Unfortunately, the multilingual BERT has been trained on a limited amount of data for each non-English language. For this reason, the ParsBERT model was proposed by Farahani et al. (2021) to overcome this limitation for the low-resource languages, including the Persian language. As Farahani et al. (2021) reported, the extended model achieved state-of-the-art results compared to other architectures and multilingual models for this language.

The ParsBERT model is based on the BERT model architecture, including a multi-layer bidirectional transformer. Farahani et al. (2021) used the original BERT BASE configuration (12 hidden layers, 12 attention heads, 768 hidden sizes), and trained it with a massive amount of crawled Persian documents.

4.1.2. XLM-RoBERTa

The RoBERTa model (Liu et al., 2019), a multi-layer bidirectional transformer described in Vaswani et al. (2017), was proposed to improve the BERT model (Devlin et al., 2019). The differences between RoBERTa and BERT are the volume of the training data, the batch size, the length of train sequences, the masking pattern, and the Next Sentence Prediction (NSP) task during the pre-training step. They trained the model with a more extensive data set, batch size, and longer sequences. They modified the masking pattern, i.e., using dynamic masking versus static masking in the BERT model, and removed the NSP loss function during the pre-training step. The XLM-RoBERTa model was proposed by Conneau et al. (2020). Its structure was inspired by Cross-lingual MLM (XLM) and RoBERTa models (Liu et al., 2019). Using cross-lingual representation was yielded in XLM-RoBERTa to supply the possibility of transferring knowledge across languages to enhance the model performance.

4.1.3. Topic Modeling

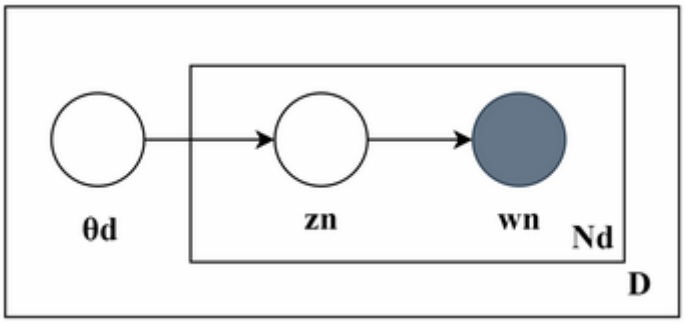
Statistical topic modeling is used in natural language processing to identify the abstract topics that exist in a collection of documents. Topic

modeling is often utilized in text mining to uncover hidden semantic structures. Utilizing topic modeling methods can provide a useful overview of a large collection, individual documents, and the relationship between them. We benefit from two methods for topic modeling, described in the following sections.

4.1.3.1. Latent Dirichlet allocation. Latent Dirichlet Allocation (LDA) proposed by Papadimitriou et al., (2000) is an unsupervised generative probabilistic model for the content analysis of texts. The documents are assumed to be random mixtures of latent topics, where a distribution of words defines a topic. LDA (Blei et al., 2003) is a popular method for topic modeling to illustrate topics by word probabilities.

Figure 2

Graphical model representation of LDA



As shown in Figure 2, given a corpus D consisting of M documents, with document d having N_d words ($d \in 1,..., M$), LDA models D, according to the following generative process (Jelodar et al., 2019):

- Choose a multinomial distribution $\phi(t)$ for topic t ($t \in \{1,..., T\}$) from a Dirichlet distribution with parameter β;
- Choose a multinomial distribution $\theta(d)$ for document d ($d \in \{1,...,M\}$) from a Dirichlet distribution with parameter α.
- For a word w_n ($n \in \{1,..., N_d\}$) in document d;
    - Select a topic z_n from $\theta(d)$.
    - Select a word w_n from $\phi(z_n)$.

The words in documents are the only observed variables in the above generative process, while other variables are latent variables (ϕ and θ) and hyperparameters (α and β). In order to infer the latent variables and hyperparameters, the probability of observed data D is computed and maximized as follows:

$$P(D|\alpha,\beta) = \prod_{0=1}^{M} \int p(\theta_d|\alpha)\left(\prod_{d_n=1}^{N} \sum_{z_{d_n}} p(Z_{d_n} \mid \theta_d)\ p(w_{d_n} \mid z_d,\beta)\right) d\theta_d$$

where α is the parameter of topic Dirichlet prior and the distribution of words over topics, drawn from the Dirichlet distribution given β; T is the number of topics; M is the number of documents; and N is the size of the vocabulary. (α, θ) is defined as the Dirichlet multinomial pair for the corpus-level topic distributions; and (β, ϕ) is defined as the Dirichlet-multinomial pair for topic-word distributions. The variable $\theta(d)$ is a document-level variable for sampling in a document; Z_{d_n} and w_{d_n} are the word-level variables to sample each word in each document.

4.1.3.2. Non-negative Matrix Factorization. Non-negative Matrix Factorization (NMF) was introduced by Févotte and Idier (2011) for text topic learning. According to the experimental results, Chen et al. (2019) proved that NMF performs much better than LDA for short text topic modeling, and NMF learns much higher-quality representative terms for the coherent topics than LDA. NMF was suggested for problems that, given a non-negative matrix V, find non-negative matrix factors W and H such that:

$$V \approx WH$$

The matrix factorization process is done in two steps such that in the first step a set of multivariate n-dimensional data vectors are placed in the columns of an $n \times m$ matrix V, where m is the number of examples in the data set. In the second step, this matrix is approximately factored into an $n \times m$ matrix W and an $r \times m$ matrix H. The factorization process results in a compact version of the original data matrix and locates a structure that is latent in the data. In other words, each data vector v is approximated by a linear combination of the columns of W, weighted by the components of h, where v

and *h* are the corresponding columns of *V* and *H*. In each iteration of the NMF algorithm, the new value of *W* or *H* is calculated by multiplying the current value by some factors that learn the approximation quality. They described cost functions that quantify the quality of the approximation. The cost function can be created using distance measures between two non-negative matrices, such as Euclidean distance. The algorithm's performance shows that the quality of the approximation improves monotonically by using these multiplicative update rules; and these update rules ensure covering the approximation to a locally optimal matrix factorization.

4.2. Classification Module

The MultiLayer Perceptron (MLP) neural network model is employed in our proposed model to classify scientific articles. MLP contains an input layer, one or more hidden layers, and an output layer of fully-connected neurons. This connection is between each neuron in one layer, with each neuron on the next layer. The input signals are guided forward from the inputs to the outputs through the hidden neurons. The number of neurons in the input layer is equal to the number of input variables of the dataset after the data preparation procedure. Equally the number of neurons in the output layer is similar to the number of classes in the dataset.

Feed-forward neural network often has one or more hidden layers followed by an output layer of linear neurons. The hidden layer with a non-linear activation function allows the network to learn non-linear and linear relations between input and output vectors. After fine-tuning the proposed model, the best result will be attained using two hidden layers with the rectified linear unit activation function, and 128 and 16 neurons. The output layer is constructed by 16 neurons and the softmax activation function.

5. Experiments

Our proposed model, shown in Figure 1, is motivated by BERT and the recent advances in transformers architecture, where the BERT pooler output is combined with document-topic distribution of topic modeling methods and

one-hot encoding of the topic number. The merged and enhanced vector feeds into MLP to discover each scientific article subject.

5.1. Dataset

A large number of scientific articles are published annually by researchers in Persian and they are archived either by governmental organizations, such as IranDoc, or institutes, such as the General Humanities Portal[1]. To conduct this research, we used the data prepared by Ghayoomi and Mousavian (2022). This dataset that is crawled from the General Humanities Portal contains 114,170 abstracts of Persian articles belonging to 16 fields of study in Humanities. In our experiments, the corpus is divided into training, validation, and test datasets. The statistical information of each dataset is reported in Table 1.

Table 1

Statistical information of the datasets used in the experiments

Set	Document
training	82202
validation	9134
test	22834

On the portal, the category of each article is defined. We put the effort into providing a uniform distribution of articles while dividing the data. Table 2 presents the statistical information of the dataset in detail.

[1] www.ensani.ir

Table 2

Statistical information of the documents in Humanities

Subject	Train	Validation	Test
Physical Education	3149	381	924
Literature	6357	698	1771
Library Science	1906	206	488
Philosophy and Logic	3514	420	929
Law	3781	430	1029
Art Science	2310	266	623
Geography	7481	824	2065
Social and Communication Sciences	6349	705	1818
History	3551	431	1022
Political Sciences and International Relations	4679	551	1366
Islamic Sciences	8219	914	2268
Economics	7134	839	1992
Women Studies	1498	173	406
Accounting and Management	10545	1104	2920
Psychology	9846	1018	2679
Linguistic	1883	174	534

5.2. Results

We evaluated our proposed model in different scenarios. First, we used the output of LDA as a vector representation to train normal machine learning classifiers, namely random forest, SVM, logistic regression, naive bayes, KNN, decision tree, and MLP. In the set of experiments, the contextualized representations, both ParsBERT and XLM-RoBERTa, were used.

5.2.1. Topic Modeling

We first ran the first set of experiments to find out which classifier with which number of topics can perform the best to do the further experiments. To begin our research, we used NMF and the common LDA topic modeling approaches. The document topic probabilistic distribution is fed to basic machine learning algorithms and SLP and MLP to predict the documents' topic.

The results of training the classifiers with 50, 100, and 200 topics are reported in Table 3. The results indicated that number of 200 topics with the KNN and decision tree, among the basic machine learning algorithms, performed the best and the worst respectively; and this number of topics with MLP performed the best among the entire series of experiments. According to the experimental results, the performance of MLP in comparison to KNN improved by over 3% on the micro F1 measure and over 4% on the macro F1 measure. It needs to be added that using MLP rather than SLP increases the performance by 1.5%. The NMF model rather than LDA has over 2.28% and 3.2% further improvement using SLP and MLP, respectively.

Table 3

Performance of classifiers trained with topic modeling approaches

Model	F1-Measure (micro)			F1-Measure (macro)		
	50	100	200	50	100	200
LDA-RandomForest	59.3	59.77	60.06	52.46	51.49	50.21
LDA-SVM	57.72	58.92	59.78	51.1	51.6	51.55
LDA-LogisticRegression	57.26	58.26	58.87	50.1	50.08	49.74
LDA-NaiveBayes	53.31	54.18	54.72	49.3	50.24	50.92
LDA-KNN	59.14	60.46	61.50	53.15	54.43	55.72
LDA-DecisionTree	40.11	38.14	38	35.53	33.58	33.49
LDA-SLP	58.01	60.34	63.15	51.85	54.44	58.13
NMF- SLP	60.44	62.49	62.78	52.39	55.34	54.48
LDA-MLP	59.52	61.81	64.56	54.78	57.58	60.72
NMF-MLP	62.72	64.57	65.62	56.14	59.16	59.89

5.2.2. Evaluating the Proposed Model

To perform further experiments and demonstrate the proposed model, we set up 8 learning scenarios described in the rest of this section -:

(a) XLM-RoBERTa-MLP (baseline): As a first experiment, the XLM-RoBERTa cross-lingual representation model is enriched with the MLP input. This model is comparable with the LDA-MLP model in

Table 5 since the classifier is the same and only the vectorization method is changed. According to the experimental results, the XLM-RoBERTa-MLP model obtained at least 10% better performance than the LDA-MLP model. We selected this model as the first baseline.

(b) XLM-RoBERTa-LDA-MLP: In this learning scenario, the LDA document topic distribution is concatenated with the XLM-RoBERTa model. The integrated vector feeds directly into fully-connected layers.

(c) XLM-RoBERTa-NMF-MLP: In this learning scenario, the XLM-RoBERTa model is enriched with the NMF document topic distribution. The results indicate that adding semantic distribution features using LDA did not improve the model's performance. Due to the fact that we classify articles based on their abstracts, it is expected to achieve better results with the NMF method than LDA.

(d) XLM-RoBERTa-NMF-1HT-MLP: In this learning scenario, the XLM-RoBERTa cross-lingual representation model is concatenated with the NMF document-topic distribution and one-hot encoding of document topics, thereafter called 1HT, as a feature. The NMF topic modeling method places articles into different clusters according to their subject fields. We add the topic number of each article as a one-hot vector to the model's features to determine which articles are more thematically similar.

As mentioned in the section about ParsBERT, transformer-based language models either do not support low-resource languages, such as Persian, or are limited to a small amount of data if a multilingual model is supplied. Accordingly, we arrange to replace the XLM-RoBERTa cross-lingual representation model with the ParsBERT contextualized representation model and repeat the experiments according to the four previous learning scenarios described as follows:

(e) ParsBERT-MLP: In this learning scenario, the ParsBERT contextualized representation model is enriched with the MLP input. The results of this model are comparable with XLM-RoBERTa-MLP

and LDA-MLP models. Since the classifier of the model is the same and the difference of the models is the vectorization type, we consider this model as the second baseline.

(f) ParsBERT-LDA-MLP: In this learning scenario, the ParsBERT contextualized representation model is concatenated with the LDA topic modeling distribution to evaluate the advantage of using LDA as a feature vector.

(g) ParsBERT-NMF-MLP: In this learning scenario, the ParsBERT model is enriched with the NMF document topic distribution.

(h) ParsBERT-NMF-1HT-MLP: This learning scenario is similar to scenario (d), except that the XLM-RoBERTa model is replaced with the ParsBERT model to utilize the large volume of Persian data used in the pre-training step of ParsBERT.

During the training phase of the model, the parameters of different models were set. Our proposed models with different settings in terms of the semantic representation mode as well as topic modeling are reported in Table 4.

Table 4

The parameters for training models

Parameter	Value
Maximum Sequence Length	128
Learning Rate	2e-6-2e-5
Epoch	3-100
Batch Size	100-500
Optimizer	Adam
Loss	Cross Entropy
MLP(Number of Layers)	5
LDA, NMF(Number of Topics)	200

According to the experimental results reported in Table 5, the BERT-based contextualized representation model performed better than the transformer-based cross-lingual representation model in general. This

improvement determines that the raw data volume used in the pre-training step has a positive impact on the performance of the classifier. Moreover, augmenting the BERT-based contextualized representation model with document topic distributions and 1HT vector performed better than the transformer model.

Table 5

Performance of the proposed learning models

Scenario	Representation Model	Topic	Classifier	F1-Measure (micro)	F1-Measure (macro)
(a)	XLM-RoBERTa	-	MLP	74.02	71.51
(b)		LDA		73.91	70.81
(c)		NMF		73.84	70.68
(d)		NMF-1HT		74.11	71.59
(e)	ParsBERT	-	MLP	74.92	72.41
(f)		LDA		75.02	72.43
(g)		NMF		75.09	72.4
(h)		NMF-1HT		75.21	72.37

Enriching the BERT-based models with topic modeling obtained two different results. The transformer-based cross-lingual representation model was not able to make use of the LDA knowledge; while the contextualized representation model made use of it and it had a slight improvement compared to the baseline. Replacement of the LDA topic modeling model with the NMF model had an improvement on the transformer-based cross-lingual representation model; and it had a positive impact on the contextualized representation model based on the macro F1-measure. The results determined that the acquisition of distributional semantic information via topic modeling eases the challenge of identifying the article's category. Consequently, we put the effort into enriching the semantic information about the article by adding the 1HT feature to the model. According to the experimental results, a slight improvement on the contextualized representation model augmented with the NMF features was achieved. The result still highlights the importance of usability of semantic information to improve the classification task of scientific

articles. The differences in the results of the best model (ParsBERT-NMF-1HT) and baseline (XLM-RoBERTa-MLP) are statistically significant according to the two-tailed t-test ($p < 0.05$).

In Table 5, the best performance of the model was achieved with the ParsBERT-NMF-1HT learning scenario. We used this model to calculate the performance of the model for each topic separately. The results are reported in Table 6. As it can be seen, the Physical Education field obtained the highest and the Women Studies field obtained the lowest performance of the model. The content of the articles in Physical Education field is technical; therefore, it is expected to achieve a better performance. The articles on Women Studies are mostly interdisciplinary and this property of the articles misleads the classifier. Three fields, namely Geography, Psychology, Economics and Literature, obtained the performance between 80 to 90% according to F-measure which we can consider the result relatively high. Moreover, three fields, namely Accounting and Management, Library Science, and Law, obtained a performance between 70 to 80% according to F-measure. Six fields, namely Art Science, Linguistic, Political Sciences and International Relations, Islamic Sciences, Philosophy and Logic, and History, achieved a performance between 60 to 70% according to F-measure which is good enough; and the field of Social and Communication Sciences performs between 50 to 60% according to F-measure.

Table 6

Performance of the ParsBERT-NMF-1HT model in different subject fields

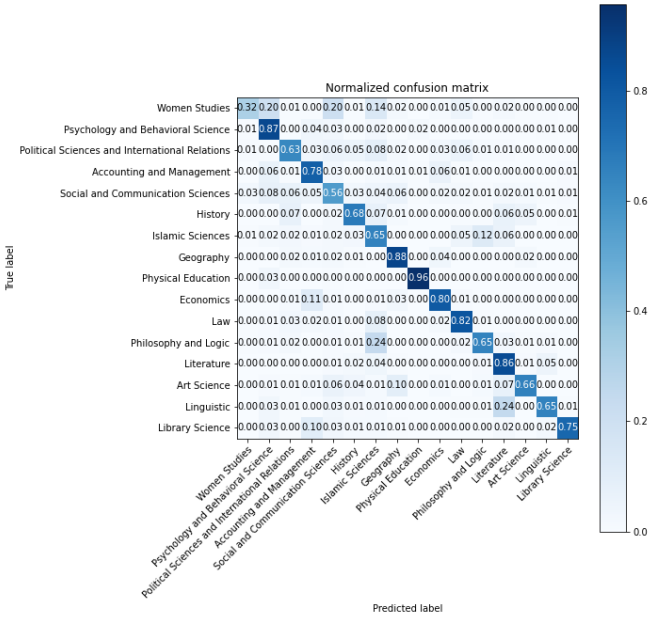
Subject	Range	Precision (per class)	Recall (per class)	F1-Measure (per class)
Physical Education	90<=x	91.37	95.12	93.21
Geography	80<x<90	81.52	89.53	85.34
Psychology and Behavioral Science		81.27	85.06	83.12
Economics		80.95	81.27	81.11
Literature		75.61	83.85	79.51
Accounting and Management	70<x<80	79.51	79.1	79.31
Library Science		77.86	79.3	78.57
Law		69.3	83.38	75.69
History	60<x<70	70.48	71.03	70.76

Subject	Range	Precision (per class)	Recall (per class)	F1-Measure (per class)
Art Science		70.39	68.69	69.53
Linguistic		69.18	66.85	67.99
Political Sciences and International Relations		71.71	62.37	66.71
Islamic Sciences		65.66	65.6	65.63
Philosophy and Logic		67.1	55.11	60.52
Social and Communication Sciences	50<x<60	65.15	55.22	59.77
Women Studies	30<x<40	46.83	32.75	38.55

The confusion matrix of the best performance of the model is represented in Figure 3. As can be seen, articles related to Women Studies are labeled with the lowest result, such as Psychology and Behavioral Science, Social Studies and Communication.

Figure 3

Normalized confusion matrix of ParsBERT-NMF-1HT model results



The articles in the field of Philosophy and Logic are labeled Islamic Science. One reason for misleading the classifier is the interdisciplinary feature of these two fields and content relatedness. The articles in the field of Linguistics are also misled by the classifier, and the wrong label of Literature is assigned.

6. Error Analysis and Discussion

In this section, several incorrectly classified data by the best model, i.e. the ParsBERT-NMF-1HT-MLP model, are examined from the test dataset and analyzed. The error analysis will help to improve future models by solving the deficiencies.

By analyzing and studying the samples in the mentioned table, we discovered that the main subject of interdisciplinary articles is not easily recognizable even by human resources. This issue has led to a reduction in the model's performance in predicting the subject of articles. The keywords in the articles' abstract are related to several topics; consequently, the model has trouble recognizing a single category for the article. Although the distributional semantics of documents and contextualized representation have facilitated this problem, its impact on model's performance is still apparent.

In Example (1), the relatedness of words, such as behavioral disorder, communication, socialization, and distraction to psychology, is undeniable. The bigram 'behavioral disorder' with nine repetitions creates more weight for Psychology rather than Social and Communication Sciences. Therefore, the model's misunderstanding to identify the correct label is not very surprising.

(1)

بازی‌های رایانه‌ای اختلالات رفتاری آموزان پسر دبستانی

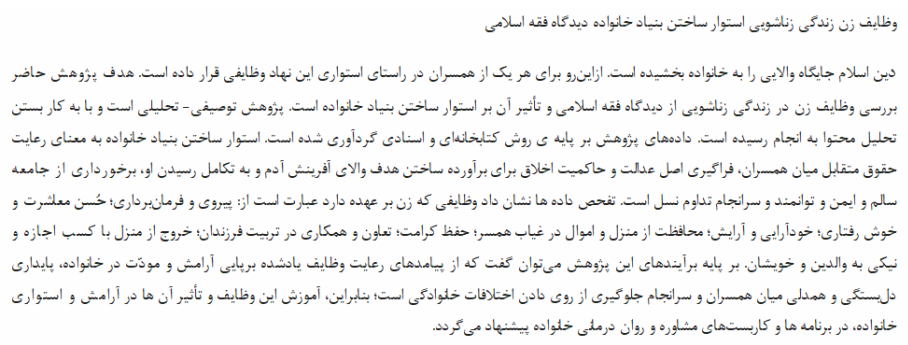
اختلالات رفتاری، اختلالات بسیار شایع و ناتوان‌کننده‌ای هستند که بر تنوعی از عملکردها، به ویژه بر عملکرد مدرسه کودکان اثر می‌گذارند و مشکلات بسیاری را برای معلمان، والدین و خود کودکان و نوجوانان ایجاد می‌کنند و آثار منفی بر یادگیری، ارتباط و کارآیی اجتماعی آنان می‌گذارند. پژوهش ها نشان می‌دهند که شیوع اختلالات رفتاری در کودکان رو به افزایش است و این احتمال داده می‌شود که گسترش روزافزون بازی‌های رایانه‌ای یکی از دلایل عمده آن است. مطالعه حاضر با هدف تعیین رابطه بین میزان استفاده از بازی‌های رایانه‌ای و بروز اختلالات رفتاری دانش آموزان پسر مقطع ابتدایی صورت گرفته است. این پژوهش با روش توصیفی و در قالب یک طرح پیمایشی و مقطعی انجام شد. داده‌های تحقیق، از ۳۱۴ نفر از دانش آموزان پسر که در سال تحصیلی ۱۳۹۲-۱۳۹۱، در پایه‌های سوم تا ششم ابتدایی شهر یزد به تحصیل اشتغال داشتند، جمع آوری شد. این افراد با استفاده از روش نمونه گیری خوشه‌ای چندمرحله ای، از بین ۱۰ مدرسه منطقه یک و دو یزد انتخاب شدند. ابزار سنجش تحقیق آزمون ارزیابی اختلالات رفتاری و پرسشنامه انجام بازی‌های رایانه‌ای بود. داده ها با استفاده از تحلیل واریانس چند متغیری تحلیل شدند. تفاوت معناداری بین دانش آموزان با سطوح متفاوت انجام بازی‌های رایانه ای، از نظر اختلال‌های رفتاری به طور کلی و نیز از نظر سه شکل از اختلال‌های رفتاری؛ یعنی اختلال سلوک، بی قراری و حواس پرتی وجود داشت. استفاده از بازی‌های رایانه ای، عاملی مهم و موثر در ابتلای دانش آموزان به اختلالات رفتاری است و خطر ابتلا به اختلالات رفتاری چون اختلالات سلوک، بی قراری و اختلال حواس پرتی را در بین دانش‌آموزان پسر دبستانی افزایش می‌دهد؛ بنابراین برای کاهش اثرات منفی استفاده بیش از حد از بازی‌های رایانه‌ای و ابتلا به اختلالات رفتاری دانش آموزان دبستانی لازم است تا نظارت بیشتری از سوی والدین بر فرزندانشان در زمینه میزان استفاده از بازی‌های رایانه‌ای صورت پذیرد.

Gold Label: Social and Communication Sciences

Predicted Label: Psychology

The content of the article in Example (2) can be attributed to two topics: Women Studies and Psychology because, on the one hand, the article negotiates with the position and duties of women in the family and, on the other hand, it deals with the relationship between spouses, especially the woman's relationship with her husband. Consequently, true label prediction for the model is problematic.

(2)



وظایف زن زندگی زناشویی استوار ساختن بنیاد خانواده دیدگاه فقه اسلامی

دین اسلام جایگاه والایی را به خانواده بخشیده است. ازاین‌رو برای هر یک از همسران در راستای استواری این نهاد وظایفی قرار داده است. هدف پژوهش حاضر بررسی وظایف زن در زندگی زناشویی از دیدگاه فقه اسلامی و تأثیر آن بر استوار ساختن بنیاد خانواده است. پژوهش توصیفی- تحلیلی است و با به کار بستن تحلیل محتوا به انجام رسیده است. داده‌های پژوهش بر پایه ی روش کتابخانه‌ای و اسنادی گردآوری شده است. استوار ساختن بنیاد خانواده به معنای رعایت حقوق متقابل میان همسران، فراگیری اصل عدالت و حاکمیت اخلاق برای برآورده ساختن هدف والای آفرینش آدم و به تکامل رسیدن او، برخورداری از جامعه سالم و ایمن و توانمند و سرانجام تداوم نسل است. تفحص داده ها نشان داد وظایفی که زن بر عهده دارد عبارت است از: پیروی و فرمان‌برداری؛ حُسن معاشرت و خوش رفتاری؛ خودآرایی و آرایش؛ محافظت از منزل و اموال در غیاب همسر؛ حفظ کرامت؛ تعاون و همکاری در تربیت فرزندان؛ خروج از منزل با کسب اجازه و نیکی به والدین و خویشان. بر پایه برآیندهای این پژوهش می‌توان گفت که از پیامدهای رعایت وظایف یادشده برپایی آرامش و مودّت در خانواده، پایداری دل‌بستگی و همدلی میان همسران و سرانجام جلوگیری از روی دادن اختلافات خانوادگی است؛ بنابراین، آموزش این وظایف و تأثیر آن ها در آرامش و استواری خانواده، در برنامه ها و کاربست‌های مشاوره و روان درمانی خانواده پیشنهاد می‌گردد.

Gold Label: Psychology

Predicted Label: Women Studies

Example (3) discusses women's employment, and contains words from both Economics and Women Studies fields, such as /?ešteqāl=e zanān/ 'women's job'. Such an interdisciplinary topic is challenging to identify the category not only for the human but also for the machine.

(3)

رشد اثر درآمد جنسیتی نیروی ایران ماتریس حسابداری اجتماعی

با توجه به طبیعت متفاوت ساختار تولید در بخش‌های مختلف اقتصاد، رشد این بخش ها اثر متفاوتی بر اشتغال دارد. از سوی دیگر با توجه به تجزیه بازار کار بر حسب جنسیت، اثر رشد بخش‌های مختلف اقتصاد بر اشتغال زنان و مردان نیز متفاوت است. در این پژوهش ساختار شغلی اشتغال زنان در بخش‌های مختلف اقتصادی در ایران و اثر رشد بخشی بر درآمد جنسیتی نیروی کار (مردان و زنان) در چارچوب مدل ماتریس حسابداری اجتماعی بررسی شد. از آنجایی که در ماتریس‌های حسابداری اجتماعی موجود در ایران حساب عوامل تولید خصوصاً نیروی کار بر حسب جنسیتی تفکیک نشده است، در همین راستا هدف اولیه در این پژوهش ارائه روشی به منظور تهیه ماتریس حسابداری اجتماعی که در آن درآمد نیروی کار بر حسب جنسیت تفکیک شده است. سپس بر اساس آن مدل ماتریس حسابداری اجتماعی ارائه و اثر رشد بخشی بر کل درآمد نیروی کار، درآمد نیروی کار زنان، و درآمد مردان به تفکیک مورد بررسی قرار گرفته است. نتایج نشان می‌دهند با رشد بخشی اثرات یکسانی بر درآمد شاغلین زن و مرد ندارد و درآمد نیروی کار مردان با اختلاف زیادی بیش از درآمد نیروی کار زنان افزایش می‌یابد.

Gold Label: Women Studies
Predicted Label: Economics

In the last example, Example (4), words, such as /še?ri/ 'poetic', /šā?erān/ 'poets', /sabk/ 'style', /še?r/ 'poem', and /pārādoks/ 'paradox', represent a literary article. The share of words that are particularly associated with linguistics in this abstract is very diminutive; therefore, the error of the model in this instance is predictable.

(4)

دلایل افزونی بسامد پارادوکس سبک هندی

سبک هندی متمایزترین سبک شعری کلاسیک فارسی است که با سبک‌های قبل و بعد خود تمایزی آشکار دارد و اصول و موازین خاص خود را داراست؛ از جمله این عنصر تمایز بخش بسامد فراوان تصاویر پارادوکسی در کنار صنایعی چون: حسامیزی، تجرید، اسلوب معادله و... است. توجه ویژه شاعران این سبک به پارادوکس و نقشی که این شگرد ادبی در ساختار شعر آن ها دارد، سبب گردید تا نگارندگان این جستار، به بررسی دلایل افزونی بسامد پارادوکس در شعر آن ها بپردازند.بر اساس یافته‌های این تحقیق، تصاویر پارادوکسی دارای کارکردهایی چون: ابهام انگیزی و دشواری، هنجارگریزی، ایجاز، دو بعدی بودن، تازگی و ایجاد شگفتی و لذت هستند که با معیارهای شعری و زیبایی شناختی سبک هندی تناسب و همخوانی فراوانی دارد و همین امر سبب استقبال و استفاده فراوان شاعران این سبک، از این شگرد ادبی شده است.

Gold Label: Linguistics
Predicted Label: Literature

In Tables 7 and 8, we report the F-measures results of all experiments for each class separately. According to the results, the latent topics in most of the subjects are useful for the classifier. This table shows that although the NMF topic modeling performed better than the LDA topic modeling, in general, according to the results in Table 5, the augmented LDA model performed better than the NMF model in the same field with the ParsBERT representation. Four subjects, including Accounting and Management, Linguistic, Political Sciences and International Relations and Women Studies, obtained the best performance with NMF; while in seven subjects, including Physical Education, Geography, Psychology and Behavioral Science, Library Science, Law, Art Science, and Islamic Sciences, obtained the best performance with LDA. Two subjects, namely Literature and Philosophy and Logic, obtained the best performance when no augmented topic modeling is used. However, the augmented NMF and 1HT obtained the best performance for the subjects Economics and Social and Communication Sciences. For History, using either the NMF model solely or the augmented NMF and 1HT model achieved equal results.

Table 7

Experimental results of XLM-RoBERTa for F1-Measure (per class)

Subject	XLM-RoBERTa			
	-	LDA	NMF	NMF–1HT
Physical Education	91.5	93.71	93.52	93.45
Geography	85.19	84.53	84.98	83.81
Psychology and Behavioral Science	82.28	82.81	82.85	83.28
Economics	80.05	78.7	79.22	79.34
Literature	78.87	78.94	78.08	79.74
Accounting and Management	77.93	77.58	78.48	78.45
Library Science	77.94	77.56	77.63	77.32
Law	75.53	75.04	74.42	74.6
History	69.3	68.97	70.2	68.38

Art Science	67.3	65.32	67.16	65.75
Linguistic	66.34	66.85	65.59	63.54
Political Sciences and International Relations	65.32	65.46	66.74	66.08
Islamic Sciences	63.4	62.48	64.64	62.06
Philosophy and Logic	58.86	62.5	62.1	62.73
Social and Communication Sciences	59.73	55.93	58.89	58.56
Women Studies	41.87	41.62	44.06	45.24

Table 8

Experimental results of ParsBERT for F1-Measure (per class)

Subject	ParsBERT			
	-	LDA	NMF	NMF–1HT
Physical Education	93.38	93.68	93.07	93.21
Geography	84.87	86.23	85.92	85.35
Psychology and Behavioral Science	83.2	83.49	82.69	83.13
Economics	80.9	81.04	80.56	81.11
Literature	80.06	80.03	79.98	79.52
Accounting and Management	79.32	78.92	79.35	79.31
Library Science	78.41	79.59	78.17	78.58
Law	75.28	76.67	75.56	75.69
History	69.15	70.61	70.76	70.76
Art Science	69.77	70.55	69.13	69.54
Linguistic	65.93	65.98	68.16	68.00
Political Sciences and International Relations	66.83	66.5	67.12	66.72
Islamic Sciences	63.44	66.69	65.34	65.64
Philosophy and Logic	64.06	58.34	60.64	60.52
Social and Communication Sciences	58.68	59.08	59.54	59.78
Women Studies	41.78	39.58	42.89	38.55

The performance of our presented models is compared with the model proposed by Ghayoomi & Mousavian (2022). To this end, we used the same training and test data as used in the experiments of Ghayoomi & Mousavian (2022). According to the experimental results in Table 9, the ParsBERT-NMF-1HT model as our proposed model improved the previous work by 0.5% according to F1-measure(micro). This determines that enriching the contextualized representation with semantic information about the article, using both LDA and NMF, and adding the 1HT feature have positive impact in comparison to the previous research.

Table 9

Comparing the performance of the proposed learning models with the previous work

Model	Topic	Classifier	F1-Measure (micro)	F1-Measure (macro)
Ghayoomi & Mousavian (2022)	-	SLP	74.71	72.55
Our Experiments	-	MLP	74.92	72.41
	LDA		75.02	72.43
	NMF		75.09	72.4
	NMF-1HT		75.21	72.37

7. Conclusion

This paper implemented the augmented BERT-based models with topic modeling for classifying Persian scientific articles in Humanities. We used the LDA topic modeling method, machine learning algorithms, and multilayer perceptron neural network for the initial experiments. In the initial scenarios, the LDA semantic distribution was fed to MLP neural networks and machine learning algorithms as a representation of abstracts. It became apparent that the MLP neural network model had remarkably better performance than machine learning algorithms. In the second step, knowledge from topic modeling methods, namely LDA and NMF, and transformer-based language models, including ParsBERT and XLM-RoBERTa, were combined and fed into the MLP neural networks to make the model more robust. In this series of

experiments, the NMF topic modeling method and ParsBERT language model had better performance than LDA and XLM-RoBERTa, respectively. Based on the better results from NMF on articles' abstracts as short texts, we designed a new scenario and developed the proposed model. We enriched semantic knowledge of articles by using 1HT feature to determine which articles were more thematically similar. The paper concludes that the contextual information is important and effective in the models to find the relations between the words. Moreover, we found that enriching the models with the topics, extracted from texts, determines that latent semantic properties in texts have a positive impact on the article classification task.

In this paper, we used the data to train the models which had 16 labels for 16 subject fields in Humanities. One direction of this research for the future work is using zero-shot or few-shot machine learning method to extend the number of the fields and the labels as well. The advantage of these machine learning methods is that they require no or very few samples in the training data.

References

Bijankhan, M., Sheikhzadegan, J., & Samareh, M. R. Y. (1994). FARSDAT - The Speech Database of Farsi Spoken Language. *Proceedings of the 5th Internationa Conference on Speech Science and Technology*, 2, 826–831.
https://www.researchgate.net/publication/292798168_The_speech_database_of_Farsi_spoken_language

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022. https://dl.acm.org/doi/10.5555/944919.944937

Borko, H. (1968). Information science: What is it? *American Documentation*, 19(1), 3–5. https://doi.org/10.1002/asi.5090190103

Chen, Y., Zhang, H., Liu, R., Ye, Z., & Lin, J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, 163, 1–13. https://doi.org/10.1016/j.knosys.2018.08.011

Chowdhury, S., & Schoen, M. P. (2020). Research paper classification using supervised machine learning techniques. *2020 Intermountain Engineering, Technology and Computing (IETC)*, 1–6. https://doi.org/10.1109/IETC47856.2020.9249211

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8440–8451. https://doi.org/10.18653/v1/2020.acl-main.747

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Volume 1 (Long and Short Papers), 4171–4186. https://doi.org/10.18653/v1/N19-1423

EmamiAzadi, T., & AlmasGanj, F. (2006). Topic classification of Persian texts based on the improved probabilistic latent semantic analysis. *The 12th Conference of Iran's Computer Society*, Tehran. https://civilica.com/doc/44669/

Farahani, M., Gharachorloo, M., Farahani, M., & Manthouri, M. (2021). Parsbert: Transformer-based model for Persian language understanding. *Neural Processing Letters*, 53(6), 3831–3847. https://doi.org/10.1007/s11063-021-10528-4

Févotte, C., & Idier, J. (2011). Algorithms for nonnegative matrix factorization with the $β$-divergence. *Neural Computing*, 23(9), 2421–2456. https://doi.org/10.1162/NECO_a_00168

Ghayoomi, M., & Mousavian, M. (2022). Application of the neural network-based machine learning method to classify scientific articles. *Iranian Journal of Information Processing & Management*, 37(4), 1217-1244. https://doi.org/10.35050/JIPM010.2022.008

Harris, Z. S. (1954). Distributional structure. *Word*, 10(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., & Zhao, L. (2019). Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11), 15169–15211. https://doi.org/10.1007/s11042-018-6894-4

Jurafsky, D., & Martin, J. H. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall: Upper Saddle River, New Jersey.

Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2018). Fuzzy approach topic discovery in health and medical corpora. *International Journal of Fuzzy Systems*, 20(4), 1334–1345. https://doi.org/10.1007/s40815-017-0327-9

Kim, S.-W., & Gil, J.-M. (2019). Research paper classification systems based on TF-IDF and LDA schemes. *Human-Centric Computing and Information Sciences*, 9(1), 1–21. https://doi.org/10.1186/s13673-019-0192-7

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *Roberta: A robustly optimized BERT pretraining approach*. ArXiv Preprint ArXiv:1907.11692. https://arxiv.org/abs/1907.11692

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281–297. https://projecteuclid.org/Proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/proceedings-of-the-fifth-berkeley-symposium-on-mathematical-statistics-and/toc/bsmsp/1200512974

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.), *Proceedings of the 26th International Conference on Neural Information Processing Systems* (pp. 3111–3119). Curran Associates, Inc.
https://proceedings.neurips.cc/paper_files/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf

Mustafa, G., Usman, M., Yu, L., Sulaiman, M., & Shahid, A. (2021). Multi-label classification of research articles using Word2Vec and identification of similarity threshold. *Scientific Reports*, 11(1), 1–20. https://doi.org/10.1038/s41598-021-01460-7

Papadimitriou, C. H., Raghavan, P., Tamaki, H., & Vempala, S. (2000). Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences*, 61(2), 217–235. https://doi.org/10.1006/jcss.2000.1711

Rabiei, M., HosseiniMotlagh, S. M., & MinaeiBidgoli, B. (2019). Using One-Class SVM for Scientific Documents Classification Case study: Iranian Environmental Thesis. *Iranian Journal of Information Processing and Management*, 34(3), 1211–1234. https://doi.org/10.35050/JIPM010.2019.036

Rivest, M., Vignola-Gagné, E., & Archambault, É. (2021). level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. *PloS One*, 16(5), e0251493. https://doi.org/10.1371/journal.pone.0251493

Salton, G. (1971). *The SMART Retrieval System — Experiments in Automatic Document Processing*. Prentice-Hall, Inc.

Shokouhian, M., Asemi, A., Shabani, A., & Cheshmesohrabi, M. (2020). Presenting a

Thematic Model of Health Scientific Productions Using Text-Mining Methods. *Iranian Journal of Information Processing and Management*, 35(2), https://doi.org/553-574. 10.35050/JIPM010.2020.061

Teymoorpoor, B., Sepehri, M.-M., & Pezeshk, L. (2009). A new method for topic classification of scientific texts (case study on the articles of the nanotechnology of Iranian specialists). *Policy of Science and Technology*, 2(2), 1–15. https://doi.org/20.1001.1.20080840.1388.2.2.2.7

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30. https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf