IJICTR
International Journal of Information & Communication Technology Research

Volume 15- Number 1 – 2023 (56 -62)

Humor Detection in Persian: A Transformers-Based Approach

Fateme Najafi-Lapavandani
Faculty of Mathematics & Computer Science
Amirkabir University of Technology
Tehran, Iran
fatemenajafi135@gmail.com

Mohammad Hassan Shirali Shahreza*
Faculty of Mathematics & Computer Science
Amirkabir University of Technology
Tehran, Iran
hshirali@aut.ac.ir

Received: 5 May 2022 – Revised: 25 August 2022 - Accepted: 27 November 2022

Abstract—**Humor is a linguistic device that can make people laugh, and in the case of expressing opinions, it can transform a phrase's polarity. Humorous sentences presenting ideas and criticism, occasionally using informal forms, have made their way to social media platforms like Twitter in almost every domain. Persian speakers likewise express their opinions through humorous tweets on Twitter. As one of the early efforts for detecting humor in Persian, the current research proposes a model by fine-tuning a transformer-based language model on a Persian humor detection dataset. The proposed model has an accuracy of 84.7% on the test set. Moreover, This research introduced a dataset of 14,946 automatically-labeled tweets for humor detection in Persian.**

Keywords: Humor Detection; Sentiment Analysis; Natural Language Processing; Deep learning; Persian language

Article type: Research Article

Publisher: ICT Research Institute

[DOI: 10.61186/itrc.15.1.56]

I. INTRODUCTION

Using humor in speech or text is a way of expressing opinions that may have a different sense than what it seems because its creative characteristics prevent the correct understanding of the text.

Social networks and online stores have increased with the growth of using the Internet. And it becomes common to publish opinions on these networks, write reviewes about products on websites, or post ideas about different domains. Owner of these websites are interested to analyzing the opinions and feelings of users in these cases. Because these opinions are sometimes mixed with humor, irony, and sarcasm, the effectiveness of sentiment analysis is affected.

In the Cambridge Dictionary, humor is defined as “the ability to be amused by something seen, heard, or thought about, sometimes causing you to smile or laugh, or the quality in something that causes such amusement.” [1] Irony is defined as “the use of words that are the opposite of what you mean, as a way of being funny” and “a type of usually humorous expression in which you say the opposite of what you

* Corresponding Author

intend.” [2] Whereas these terms are not the same, the main focus of this research is on humor detection and distinguishing these two classes is not a goal of this research.

In recent years, Twitter has become a great source of expressing users’ opinions, sentiments, and criticism in almost all domains. This case attracts the attention of companies and researchers to Twitter, which works on opinion mining and sentiment analysis. Many tweets are humorous or ironic, and Persian-speaker users are no exception. Analyzes aware of these phenomena in the text may be able to predict the sentiment and polarity of the text with better accuracy. Many tweets are not expressed in the official form. Dealing with informal language in Natural Language Processing tasks is a critical challenge, especially in Persian, because written informal Persian has no standard rules. The problem of humor detection is a classification problem. The samples should be classified with one of the two labels: 1 and 0, indicating humorous text and non-humorous text in order.

The main contributions of this research for the humorous detection in Persian are:

- Introduction of a new Persian dataset for humor detection, automatically tagged
- Presenting pre-trained language models trained on the dataset
- Analysis and evaluation of the trained models

The next section includes related works on humor and irony detection. In the third section, we introduce the new Persian dataset used in this research. The fourth section introduces the proposed methods to solve the problem. The performed experiments and their results are analyzed in the fifth section. The last section discusses the future work and concludes the paper.

II. Related Works

In this section, we will review the humor detection task in general and take a glance at irony and sarcasm detection tasks, followed by attempts on Persian for detecting irony and sarcasm.

A. Humor Detection

In the machine learning approach, researchers consider a set of extracted features from the text and train models to classify the texts into humorous and non-humorous. Mihalcea and Strapparava [3] extracted content-based and humor-specific stylistic features from their datasets with 32000 samples and applied SVM and NB classifiers. While in [4], these algorithms were applied to different features, such as human-centeredness and negative polarity. Yang et al. [5] proposed a method for automatically extracting humor anchors. They use semantic structure-based features, such as ambiguity, incongruity, interpersonal effect, and phonetic style, and used a random forest classifier. Targeting comedian accounts on Twitter, Raz [6] classified tweets using syntactical, pattern-based, lexical, and morphological features. Paper [7] is another attempt on Twitter that uses phonetic, morpho-syntactic, lexico-semantic, and pragmatics features for a Gradient Boosted Regression Trees (GBRT) model.

With the development of deep learning techniques for natural language processing tasks, recent researches use these techniques to detect humor in the text. Oliveira and Rodrigo [8], use bag-of-words and word vector features on the Yelp dataset with 1.6 million reviews. They applied Random Forests, Linear Discriminants, two types of deep feedforward networks, recurrent neural networks, and convolutional neural networks on the training set for classification.

A dataset of Spanish tweets with 20000 human-annotated samples was created in [9]. They extract stylistic, affective, structural, and content features and, using the word2vec algorithm as an embedding method, proposed a Bidirectional LSTM architecture for humor detection. Chen and Soo [10] also used convolutional neural networks for this task. Weller Seppi [11] detects humor using a transformer architecture with a dataset of 16000 labeled instances. Annamoradnejad [12] collected a dataset with 200,000 samples and, using BERT embedding, compared the BERT-based method with deep learning algorithms such as Convolutional Neural Networks (CNN), Attention-Based Recurrent Neural Networks and BiRNN-CNN.

B. Irony and Sarcasm Detection

As irony and sarcasm are two similar phenomena to humor, studying the approaches for solving these tasks can be helpful. In the first attempts, researchers used a set of extracted features from the text and trained machine learning algorithms to classify the texts into ironic or non-ironic [13, 14]. These solutions consider structural and sentimental features and do not use text.

Some researchers have considered the context of the text for Irony Detection [15-17]. They believe the irony of a tweet is related to the previous tweets of the author, and they extracted features from the author of the tweet or his previous tweets. This approach does not work for text only and needs tweets of active users.

Like many other tasks in NLP, deep learning is the recent approach to detecting irony in the text: using GloVe [18, 19] and Elmo [20] for vector representation, applying CNN followed by an LSTM [21], bidirectional GRU deep neural network [22], BiLSTM based on attention and CNN [19], and transfer learning [23]. Our proposed method will consider the limitations of these networks using state-of-the-art pre-trained language models instead of static or LSTM-based word embeddings.

SemEval [24] is an ongoing series of international research workshops in the field of natural language processing with a focus on semantic evaluation. One of the topics discussed in this series was Sarcasm detection in Arabic and English. More competitors in this research have presented their solutions using transformer-based language models [25].

C. Irony and Sarcasm Detection in Persian

For Irony and sarcasm Detection in Persian, we review two works. Bekainejad et al. [26] presented a model on a dataset of 2500 tweets. After pre-processing and preparing the data, they extracted the features of the text, like polarity and the use of punctuation marks. In evaluating the examined models, the model based on the Support Vector Machine had the highest F1 value

[DOI: 10.61186/itrc.15.1.56]

of 80%. The text is not used as a feature in the classification algorithms, and only a few limited features extracted from the text are used for the classification. However, in current research, we intend to use the tweet text in the algorithm.

Golazizian et al. [27] introduced a deep neural network for emoji prediction. They presented a deep neural network for detecting irony in Persian using transfer learning with the result of 83% for F1. FastText [20] was used to embed tweets. Considering that the first network is customized, in this research, we will examine the problem with a more general pre-trained model and use a more efficient technique to vectorize tweets instead of FastText.

III. DATASET

Twitter, pun of the day, Yelp reviews, and news headlines are usually used as a dataset for the humor detection task. We could not find any previous dataset for Persian humor detection. The two most similar Persian datasets are suitable for sarcasm [26] and irony [27] detection. Those datasets were manually-labeled tweets.

Our dataset consists of Persian tweets. Using hashtags to identify humor and irony is not common among Persian users, and crawling without hashtags needs an annotating step to create a labeled dataset. For manually labeling text, annotators decide on labels and classes according to a set of rules usually determined by linguistics. The most assigned label to each sample defines their class. Due to the creative nature of humor, linguists cannot establish general rules for its detection, which makes the problem of labeling more difficult; In addition, labeling itself is a time-consuming process.

For data collection and labeling, [27] noted that the rate of ironic messages in the Telegram channel “OfficialTwitterFarsi” is higher than their collected daily tweets. This channel's messages are some of the daily tweets with a strong impression on Twitter.

The update of Telegram messenger on 2021, December 30th, allows Telegram users to record a reaction by choosing one of the “😁👍👎❤️😢” emojis for each message. We decided to collect tweets from this Telegram channel. The recorded reactions for each message can indicate the tags that Telegram users have assigned to each message. We can see reactions to each message process as an annotating task with thousands of volunteer annotators that select an emoji between a set of 5 for labeling.

Using a Telegram API, we received 129,536 unique messages from the “OfficialTwitterFarsi” channel. Still, due to the lack of recorded reactions for messages far away from the latest Telegram update, We consider only 31,983 messages sent between December 1st, 2021 to August 25th, 2022. Among these messages, we removed tweets with videos, images, or music files to focus on text-only tweets.

We assume “OfficialTwitterFarsi” channel members react to a message with “😁” when they notice the humor in each message and apply an automated label-assigning solution for the remaining 18,133 messages up to this point. Some messages are removed if they have less than ten reactions, have media beside them, or the message template is not the same as the tweet's template. There are some advertisement messages on the "OfficialTwitterFarsi” channel, and their template varies from tweet messages.

Each sample was carried out based on the two highest repetition reactions. We assigned the label “1” as a sign of humor and irony to the messages when their highest reaction was "😁." Suppose the highest reaction recorded for a sample was "👍” while the second highest reaction is “😁”; In that case, the boundary of distinguishing humorous and non-humorous is ambiguous, so we ignore this message category because their humor is unknown. Other messages get a "0" tag as non-humorous messages.

The final dataset contains 7,014 humorous and 7,932 non-humorous, nearly clean tweets. In Example tweets and their irony labels., examples of these data are presented. Details of the irony dataset provides the details of the created dataset .

We need a multilingual dataset for irony detection to implement the second solution. We gathered the irony and sarcasm detection corpora from Kaggle [28, 29] and SemEval datasets for the irony and sarcasm detection in English, Brazilian Portuguese, Arabic, and Hindi, including 94,428 unique tweets, of which 65,711 are ironic. Details of the multilingual irony dataset shows the number of each language sample.

TABLE I. EXAMPLE TWEETS AND THEIR IRONY LABELS.

Tweet	Label
پشت یه کامیونه نوشته بود: سلطان خیانت هیدروژن! هم پیوند کوالانسی میگیره هم هیدروژنی! فکر کنم رانندش لیسانس شیمی داشته🤦‍♂️😂😕	Humorous
آره مهاجرت خوبه ولی قشنگترش این بود که همینجا کنار خانواده و دوستامون به خواسته‌هایی که داشتیم برسیم):	Non-humorous
تاس کباب داشتیم بابام جفت شیش آورد همه‌شو خورد	Humorous
مدیون تاول های پامون تو راه اشتباه نباشیم! هر جا که فهمیدیم مسیر درست را انتخاب نکردیم،بدون تردید دور بزنیم و برگردیم!	Non-humorous

TABLE II. DETAILS OF THE IRONY DATASET

Property	Humorous	Non-humorous

[DOI: 10.61186/itrc.15.1.56]

No. of tweets	7,014	7,932
Avg. no. of tokens per tweet	30	45
Max. no. of tokens per tweet	260	430

TABLE III. DETAILS OF THE MULTILINGUAL IRONY DATASET

Language	Ironic	Non-ironic
English	52,733	24,474
Arabic	745	2,375
Portugal	12,351	2,476
Hindi	936	196
Total	65,711	28,717

IV. PROPOSED METHOD

Previous works on humor detection show the effectiveness of transformer-based language models[12]. Therefore, we examine solutions using different transformer-based language models for humor detection in Persian.

The transformer models rely entirely on self-attention to compute their input and output representations without using sequence-aligned RNNs or convolution [30]. Based on transformers, a language representation model called BERT (Bidirectional Encoder Representations from Transformers) was pre-trained for two unsupervised tasks in English, MLM (Masked Language Model) and NSP (Next Sentence Prediction). BERT versions have an architecture of 12 and 24 layers of transformers with a hidden size of 768 and 1024. [31] The excellent performance of BERT was a motivation for pretraining other transformer-based language models for different languages and tasks.

As the first solution, we examine fine-tuning the different language models for the classification task. Among the existing models, we examine the ParsBERT model [32] and two versions of XLM-RoBERTa [33]. All three are transformers-based language models and are pre-trained on different corpora.

ParsBert's Architecture is like $BERT_{Base}$. It is pre-trained on a Persian corpus of formal language text in these domains: general, scientific, lifestyle, itinerary, digital magazine, storybooks, and news [32].

Figure 1. Two other language models, base and large versions of XLM-RoBERTa are multilingual. These two models have been pre-trained with $BERT_{Base}$ and $BERT_{Large}$ architectures, respectively, on a corpus of 100 languages, each of which included texts in formal and informal forms. While they are pre-trained on 2.5TB of data, the Persian part was 13.259 billion tokens and a total size of 111.6 GB. [34] Architecture of the transformer-based proposed models [35]

shows the architecture of these proposed methods. Humor Detection model: Fine-tuning the transformer-based language model on the Persian dataset demonstrates fine-tuning the language model on the Persian dataset. In Fine-tuning of the multilingual language model on the multilingual dataset, the diagram shows the fine-tuning of the multilingual language model on the multilingual dataset and The previous fine-tuning model on the Persian humor detection dataset. represents the previous fine-tuning model on the Persian humor detection dataset.

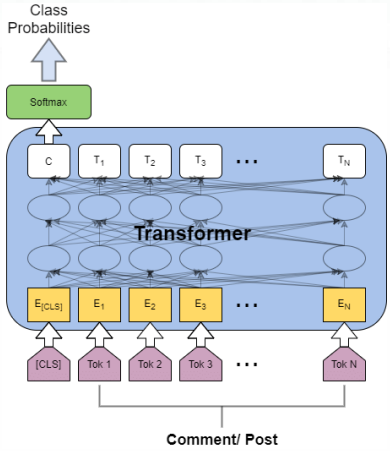

Figure 2. Architecture of the transformer-based proposed models [35]

Figure 3. Humor Detection model: Fine-tuning the transformer-based language model on the Persian dataset

Figure 4. Fine-tuning of the multilingual language model on the multilingual dataset

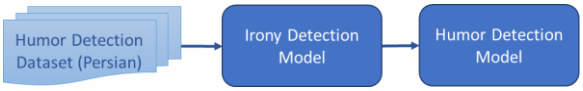

Figure 5. The previous fine-tuning model on the Persian humor detection dataset.

Wei et al.[36] show that fine-tuning a transformers-based language model for different tasks increases the performance of zero-shot learning. As the second solution for humor detection, we examine the impact of the data with samples from other languages. For this purpose, we used a corpus consisting of some sarcasm and irony detection datasets in four languages other than Persian to fine-tune the XLM-RoBERTa language model on it, and we examined the model results on the Persian humor dataset. This method is similar to zero-shot learning,. because in the first fine-tuning stage, the model does not see any sample of each class in Persian. After this, we continue examining the results by fine-tuning the model on the Persian data and comparing the results.

V. EXPERIMENTS

This section discusses the experiments that have been done to evaluate the proposed methods. To perform experiments, the multilingual corpus of Irony detection and the introduced Persian humor dataset, both of which were reviewed in the third section, is used. For the training and evaluation of the models, the Persian humor detection dataset is randomly splited into training and testing sets with a ratio of 8:2.

Accuracy, precision, recall, and F1 helped us to evaluate and compare models. Accuracy is the number of true predicted test samples divided by the number of all test samples. F1, as is shown in equation 1, is a measure of a test's accurateness in binary classification,

[DOI: 10.61186/itrc.15.1.56]

calculated from the test's recall and precision, while recall shows sensitivity with the number of true positive predicted divided by the number of all actual positive samples and precision indicates a positive predictive value with the number of true positive predicted samples divided by the number of all samples predicted as positive.

$$F1 = 2\frac{precision\ .recall}{precision+recall} \quad (1)$$

A. Baseline models

To compare the proposed methods, we need baseline models. We use machine learning algorithms for the baseline models. Naive Bayes, Support Vector Machine, and Multilayer Perceptron were chosen for classification, and the tf-idf(term frequency-inverse document frequency) method was used as sentence representation in 10,000 dimensions.

Naive Bayes is a generative model. We assume the two dataset class follows a Gaussian distribution and train a Gaussian Naive Bayes model.

We tried linear, polynomial, Radial Basis Function(RBF), and sigmoid as different kernels to train a Support Vector Machine algorithm. We consider only the support vector machine algorithm with RBF kernel because it performs better than others.

In the multilayer perceptron algorithm, a network with two hidden layers, each layer containing 50 nodes, was used. This network was activated with a hyperbolic tangent function and optimized with the Adam function; it iterates until convergence in the maximum number of 300.

The Support Vector Machine algorithm performs better among the three mentioned algorithms, with F1 equal to 79%.

B. Experiments on transformer-based language models

Figure 6. Pre-trained language models based on transformers were used for the first proposed method. The chosen transformers are a Persian language model, ParsBERT, and two versions of the XLM-RoBERTa as multilingual language models. These three transformers were fine-tuned separately on the training part of the Persian humor detection dataset for the classification task, as shown in Architecture of the transformer-based proposed models [35]

. Hyper-parameters' values shows the Hyper-parameters' values for implementing these methods.

Among these three models, fine-tuned on the cross-lingual language model, XLM-RoBERTa$_{Large}$, performs better, with an F1 of 84.6%.

TABLE IV. HYPER-PARAMETERS' VALUES

Hyper-parameter	Value
Embedding dimension	768
No. epochs	10
Batch size	16
Max. sequence lenght	128

TABLE V. COMPARISON OF THE PROPOSED MODEL

Model	Accuracy	Recall	Precision	F1
Baseline models				
Naïve Bayes (NB)	66%	68%	67%	65%
Support Vector Machine (SVM)	79%	78%	79%	79%
Multilayer Perceptron (MLP)	73%	73%	73%	73%
Fine-tuned transformer-based language models				
ParsBert Version 3	81.3%	81.4%	81.3%	81.3%
XLM-RoBERTa$_{Base}$	82.6%	82.8%	82.6%	82.5%
XLM-RoBERTa$_{Large}$	**84.7%**	**84.7%**	**84.6%**	**84.6%**
Fine-tuned on multilingual dataset				
Pre-trained on other languages	70.6%	70.7%	70.6%	70.6%
Fine-tuned the pre-trained on other languages model	84.5%	84.6%	84.4%	84.5%

C. Zero-shot learning approach

Another suggested method is using a corpus of irony detection in several languages other than Persian. To perform this experiment, we fine-tuned the large version of the XLM-RoBERTa language model on the second corpus. The model was evaluated on the test part of the introduced Persian humor detection dataset. A model that has not seen any samples of humorous and non-humorous from the Persian language has an F1 value of 70.6% with 70.6% accuracy.

The language model obtained at this stage was also fine-tuned on the introduced Persian humor detection dataset. The final model can detect humor on the test dataset with accuracy and F1 equal to 84.5%.

D. Comparing results

Details of each experiment are described in this section, and the results are in Comparison of the proposed model. In baseline algorithms experiments, the Support Vector Machine performs better with F1 equal to 79%. XLM-RoBERTa$_{Large}$ has the best performance with 84.6% for F1 among the three fine-tined models on the Persian humor dataset. The two models fine-tuned on the multilanguage irony detection dataset have 70.6% and 84.5% for F1 on the Persian test part before and after fine-tuning on the Persian humor detection dataset.

Among all the mentioned experiments, the fine-tuned cross-lingual language model, XLM-RoBERTa$_{Large}$, on the Persian dataset has the best performance with a value of 84.6% for the F1 in

[DOI: 10.61186/itrc.15.1.56]

detecting humor in the test dataset of the introduced dataset.

VI. Conclusions and Future Works

Besides introducing a new dataset for humor detection in Persian, this research proposed two methods for humor detection in Persian. These methods provide models for detecting humor in Persian by fine-tuning transformers-based language models based on the introduced dataset in Persian and the collected corpus in four different languages. The fine-tuned XLM-RoBERTa$_{Large}$ language on the introduced Persian humor detection dataset performs best among the proposed models. This model can detect humor in the text with an accuracy of 84.7% and an F1 of 84.6% on the test dataset.

One of the goals of humor detection in texts is to increase the performance of sentiment analysis models. Future works can analyze the effect of a feature that indicates whether the tweet is humorous or not, along with the text and other features that are used to detect the polarity of the text. Out work is pioneering work for humor detection in Persian. Another future work can be collecting a detailed Persian dataset that suits a contextual approach. A contextual dataset for Persian humor detection should contain the interaction of other users through the tweets (e.g., the number of likes, retweets, and replies), information about the author's account (e.g., joint year, the number of followers and followings), and a list of previous tweets.

References

[1] "Humor | English meaning." Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/humor (last accessed February 2023).

[2] "Irony | English meaning." Cambridge Dictionary. https://dictionary.cambridge.org/dictionary/english/irony (last accessed February 2023).

[3] R. Mihalcea and C. Strapparava, "Making computers laugh: Investigations in automatic humor recognition," in *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, 2005, pp. 531-538.

[4] R. Mihalcea and S. Pulman, "Characterizing humour: An exploration of features in humorous texts," in *International Conference on Intelligent Text Processing and Computational Linguistics*, 2007: Springer, pp. 337-347.

[5] D. Yang, A. Lavie, C. Dyer, and E. Hovy, "Humor recognition and humor anchor extraction," in *Proceedings of the 2015 conference on empirical methods in natural language processing*, 2015, pp. 2367-2376.

[6] Y. Raz, "Automatic humor classification on Twitter," in *Proceedings of the NAACL HLT 2012 student research workshop*, 2012, pp. 66-70.

[7] R. Zhang and N. Liu, "Recognizing humor on twitter," in *Proceedings of the 23rd ACM international conference on conference on information and knowledge management*, 2014, pp. 889-898.

[8] L. De Oliveira and A. L. Rodrigo, "Humor detection in yelp reviews," *Retrieved on December*, vol. 15, p. 2019, 2015.

[9] R. Ortega-Bueno, C. E. Muniz-Cuza, J. E. M. Pagola, and P. Rosso, "UO UPV: Deep linguistic humor detection in Spanish social media," in *Proceedings of the third workshop on evaluation of human language technologies for Iberian languages (IberEval 2018) co-located with 34th conference of the Spanish society for natural language processing (SEPLN 2018)*, 2018, pp. 204-213.

[10] P.-Y. Chen and V.-W. Soo, "Humor recognition using deep learning," in *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)*, 2018, pp. 113-117.

[11] O. Weller and K. Seppi, "Humor detection: A transformer gets the last laugh," *arXiv preprint arXiv:1909.00252*, 2019.

[12] I. Annamoradnejad and G. Zoghi, "Colbert: Using bert sentence embedding for humor detection," *arXiv preprint arXiv:2004.12765*, 2020.

[13] A. Reyes, P. Rosso, and T. Veale, "A multidimensional approach for detecting irony in twitter," *Language resources and evaluation*, vol. 47, no. 1, pp. 239-268, 2013.

[14] F. Barbieri, H. Saggion, and F. Ronzano, "Modelling sarcasm in twitter, a novel approach," in *proceedings of the 5th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2014, pp. 50-58.

[15] A. Rajadesingan, R. Zafarani, and H. Liu, "Sarcasm detection on twitter: A behavioral modeling approach," in *Proceedings of the eighth ACM international conference on web search and data mining*, 2015, pp. 97-106.

[16] D. Bamman and N. Smith, "Contextualized sarcasm detection on twitter," in *proceedings of the international AAAI conference on web and social media*, 2015, vol. 9, no. 1, pp. 574-577.

[17] B. C. Wallace and E. Charniak, "Sparse, contextually informed models for irony detection: Exploiting user communities, entities and sentiment," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 2015, pp. 1035-1044.

[18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532-1543.

[19] A. Kumar, S. R. Sangwan, A. Arora, A. Nayyar, and M. Abdel-Basset, "Sarcasm detection using soft attention-based bidirectional long short-term memory model with convolution network," *IEEE access*, vol. 7, pp. 23319-23328, 2019.

[20] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the association for computational linguistics*, vol. 5, pp. 135-146, 2017.

[21] A. Ghosh and T. Veale, "Fracking sarcasm using neural network," in *Proceedings of the 7th workshop on computational approaches to subjectivity, sentiment and social media analysis*, 2016, pp. 161-169.

[22] M. Zhang, Y. Zhang, and G. Fu, "Tweet sarcasm detection using deep neural network," in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: technical papers*, 2016, pp. 2449-2460.

[23] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann, "Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm," *arXiv preprint arXiv:1708.00524*, 2017.

[24] "SemEval." https://semeval.github.io/ (last accessed February 2023).

[25] I. A. Farha, S. V. Oprea, S. Wilson, and W. Magdy, "SemEval-2022 Task 6: iSarcasmEval, Intended Sarcasm Detection in English and Arabic," in *Proceedings of the 16th International Workshop on Semantic Evaluation (SemEval-2022)*, 2022, pp. 802-814.

[26] Z. B. Nezhad and M. A. Deihimi, "Sarcasm detection in Persian," *Journal of Information and Communication Technology*, vol. 20, no. 1, pp. 1-20, 2021.

[27] P. Golazizian, B. Sabeti, S. A. A. Asli, Z. Majdabadi, O. Momenzadeh, and R. Fahmi, "Irony detection in Persian language: A transfer learning approach using emoji prediction," in *Proceedings of The 12th Language Resources and Evaluation Conference*, 2020, pp. 2839-2845.

[28] "Kaggle." https://www.kaggle.com/ (last accessed February 2023) .

[29] SemEval. "iSarcasmEval: Intended Sarcasm Detection In English and Arabic." https://sites.google.com/view/semeval2022-isarcasmeval (last accessed February 2023).

[30] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.

[DOI: 10.61186/itrc.15.1.56]

[31] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
[32] M. Farahani, M. Gharachorloo, M. Farahani, and M. Manthouri, "Parsbert: Transformer-based model for persian language understanding," *Neural Processing Letters*, vol. 53, no. 6, pp. 3831-3847, 2021.
[33] A. Conneau *et al.*, "Unsupervised cross-lingual representation learning at scale," *arXiv preprint arXiv:1911.02116*, 2019.
[34] G. Lample and A. Conneau, "Cross-lingual language model pretraining," *arXiv preprint arXiv:1901.07291*, 2019.
[35] T. Ranasinghe, S. Gupte, M. Zampieri, and I. Nwogu, "Wlv-rit at hasoc-dravidian-codemix-fire2020: Offensive language identification in code-switched youtube comments," *arXiv preprint arXiv:2011.00559*, 2020.
[36] J. Wei *et al.*, "Finetuned language models are zero-shot learners," *arXiv preprint arXiv:2109.01652*, 2021.

Fateme Najafi-Lapavandani received the B.Sc. degree in Computer Science from Kharazmi University of Tehran, Tehran, in 2019 and the M.Sc. degree in Computer Science from Amirkabir University of Technology, Tehran, Iran, in 2022. Her research interests include Natural Language Processing (NLP), Generative Language Models, Neural Networks, and Machine Learning.

Mohammad Hassan Shirali-Shahreza is an Assistant Professor in the Faculty of Mathematics and Computer Science at Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. He received his B.Sc. degree in Computer Hardware Engineering from Isfahan University of Technology, Isfahan, Iran, in 1986, M.Sc. dgree in Computer Hardware Engineering from Sharif University of Technology, Tehran, Iran, in 1988, and Ph.D. degree in Ph.D. Computer Hardware Engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 1996. In 1994 he was a visiting research scholar in the Electrical Engineering Department, Southern Methodist University (SMU), Dallas, Texas, USA. His interesting fields are Human Interactive Proofs (HIP), Neural Networks, OCR (Optical Character Recognition), and Pattern Recognition. He is a member of Iranian Computer Society, and Iranian Society of Cryptology.

[DOI: 10.61186/itrc.15.1.56]