Neural Network Approach for Irony Detection from Arabic Text on Social Media

Ali Allaith[1], Muhammad Shahbaz[1], Mohammed Alkoli[2]

[1] University of Engineering and Technology, Lahore, Pakistan
[2] University of Mysore, Mysore, India
allaith.net@gmail.com, m.shahbaz@uet.edu.pk,
mohd.alkoli@gmail.com

Abstract. Irony plays an important part in human social interaction which is used to emphasize occurrences that deviate from the expected. Humans manipulate each other in a very negative way by writing the opposite of what they mean. However, irony detection is a complex task even for humans. In this research, we study the problem of irony detection as a classification problem and utilized the dataset offered by the IDAT workshop. We also propose a classification system for detecting irony in the Arabic tweets using neural networks[1].

Keywords: Irony Detection, Text Classification, Neural Networks, Arabic Language.

1 Introduction

The increasing of the information on the internet, especially in social media, leads to many natural language problems. People express their textual opinion in online resources like public forums and microblogging sites. One of the ways to express our opinion, which is the most interesting, is by using figurative languages such as irony and sarcasm. While there is a gap between the intended meaning and the literal meaning, irony detection becomes a very difficult task due to the ambiguous interpretations. Many researches and shared tasks related to the irony detection have been focused on different languages such as English, French, and Italian languages however, a little contribution has been done in Arabic language.

Irony detecting has its implications in sentiment analysis, opinion mining, and advertising as shown in [1,2,3] respectively. For example, the detection of irony before applying sentiment analysis is a big challenge where sometimes the presence of irony content may reverse the sentiment polarity of the text from positive to negative and vice versa. Therefore, the sentiment analysis systems which exploit the basic approaches of

sentiment detection based on the frequency of the positive and negative terms will fail to detect the exact sentiment class if there is an irony expression in the text.
Research in irony will not just improve the performance of sentiment analysis systems, but instead it can help in understanding the cognitive process involved and how human beings process and produce the kind of utterance. The irony is a broad concept which has a strong association with some other discipline such as linguistics and psychology. Many models have been developed for irony detection but the first model was developed by (Utsumi, 1996) in [4]. Other models have been proposed to address the irony detection among tweets using different features such as cue-words or hashtags [5, 6, 7]. In [8], they analyzed tweets based on the #irony and #humor hashtags in order to identify features to distinguish between these two classes. The corpus-based approach is another mechanism to identifying irony from text [7]. They collected a corpus that contains 40,000 tweets relying on the self-tagged approach. The corpus used to build a model to distinguish the ironic from non-ironic tweets using Naïve Bayes and decision tree classification algorithms. The lexicon-based approach has been used in [9] by proposing a model to detect irony using lexical features such as rare terms frequency, emoticons, positive and negative terms. Amazon reviews were used as a corpus for irony classification in [10]. The corpus was annotated using crowdsourcing. Different features have been used to build the model including n-grams, emoticons, review rates, punctuation marks, and interjections. They used a set of machine learning algorithms such as Naïve Bayes, decision tree, logistic regression, random forest, and support vector machine.
Irony detection for other languages has been proposed for French and Italian languages in [11, 12] respectively. On the other hand, few researchers have addressed the problem of irony detection. An Arabic research in irony detection has been proposed in [13]. The proposed system is called “Soukhria”. They used four groups of features: surface features, sentiment features, shifter features, and contextual features to build binary classification for irony detection in tweets.
The above mentioned techniques showed that there is no optimal model that could be considered as a baseline for the irony detection problem. This paper proposes a neural network model to classify ironic from non-ironic text in Arabic. The dataset (training set and testing set) was offered by the IDAT workshop. The training dataset comprised of 4000 tweets while the testing set is 1000 tweets. The training set tagged with two class either ironic or non-ironic.

2 Task Overview

In Ironic Detection in the Arabic Text (IDAT) overview [14], the task is to detect the irony in Arabic tweets from different political issues and events in the middle east from the period from 2011-2018. The dataset was collected from Twitter using a set of predefined keywords. Tweets were written in both modern standard Arabic and Arabic dialects. In this task, the system has to determine whether a tweet is ironic or non-ironic. It is a binary classification task. The system evaluation was performed according to

(accuracy, precision, recall, and F1-measure). The submission was ranked using F1-measure.

3 Proposed Model

In this part, we will describe the model which has been used to accomplish this task. The inspiration for using neural networks for word embedding is the fast processing of data in comparison with the traditional machine learning algorithms. We used an open-source, free, and lightweight library called "fastText". This library offered by Facebook AI Research Lab. It is a neural network-based algorithm that is used efficiently in text classification and word representation learning by obtaining vector representations for words. We used the training dataset with ironic and non-ironic labels to automatically recognize the irony from the testing set.

3.1 Dataset

The dataset was offered by IDAT to apply text classification of irony detection from Arabic tweets. The dataset was obtained from Twitter and is related to different events in the middle east during the years 2011 to 2018. It was collected using a set of predefined keywords and hashtags. The tweets in the dataset are written in modern standard Arabic and Arabic dialects. The training and testing dataset descriptions are shown in Table 1.

Table 1. Dataset Description.

Training Dataset	
Class Name	Number of Tweets
Ironic	2,100
Non-ironic	1,933
Total Tweets	4,033
Testing Dataset	
Total Tweets	1,006
Words Statistics	
Total Words	92,763
Unique Words	26,368
Average Words Per Tweet	18.6

3.2 Text Pre-processing

Looking at the text on the dataset, we observe that most of the tweets have punctuations, numbers, and emoticons. One of the first steps to improve the performance of the irony detection model is to apply some pre-processing tasks such as text normalization, stop word removal, and word segmentation.

Table 2. Text Pre-processing.

Raw Text	Pre-processed Text
"@Abdurrahmanezz "" تتهم "" الوطنية للتغيير"" الرئيس بالاستعانة بأشخاص محسوبين على نظام مبارك ومحاولة استرضاء ""اليمين المتطرف"" #مصر #خلافات"	الوطنية للتغيير تتهم الرئيس بالاستعانة باشخاص محسوبين على نظام مبارك ومحاولة استرضاء اليمين المتطرف مصر خلافات
مداد_نيوز|| استكمالاً لمحاولة التدخل في الشأن الليبي، # قطر تحاول قتل سيف الإسلام القذافي لهذه الأسباب. #ليبيا #نظام_الحمدين #قطر #سيف_الإسلام_القذافي #تميم_يدعم_الإرهاب #قطر_تدعم_الإرهاب	مداد نيوز استكمالا لمحاولة التدخل في الشان الليبي قطر تحاول قتل سيف الاسلام القذافي لهذه الاسباب ليبيا نظام الحمدين قطر سيف الاسلام القذافي تميم يدعم الارهاب قطر تدعم الارهاب
30#يونيو ..!! السؤال بقي: يا ترى هو هينزل رافع علم #مصر زينا ولا علم الجماعة..!!	يونيو السؤال بقي يا ترى هو هينزل رافع علم مصر زينا ولا علم الجماعة

3.3 Model Construction

As we mentioned that we used fastText, a neural network library, to build a model to detect irony from Arabic tweets. The algorithm is capable of training and prediction with millions of example text data within a few minutes over a multi-core CPU. We first prepare the dataset by applying Arabic text pre-processing and convert the dataset to the fastText format as follows:

__label__<1><Tweet> for ironic tweets
__label__<0><Tweet> for non-ironic tweets

We have adjusted the algorithm's parameters to build two models as follows in Table 3.

Table 3. FastText Parameters.

Parameters		
Parameter	Model1	Model2
Epochs	40	50
Learning Rate	0.2	0.1
N-grams	1	2

4 Experimental Results

We performed experiments with two different neural network models to classify the tweets in the dataset into ironic and non-ironic classes. These two classes were used to train the training set and predict the testing set offered by the IDAT. For both models, the training data contains 80% (4,023 tweets) and testing set 20% (1,006 tweets). Our system ranked 8 and 17 among 27 submission runs in the IDAT result of the irony detection of the Arabic tweets. Both models achieved 81.7% and 79.4% for model1 and

model2 respectively as described in the previous section. The main cause of misclassification is the imbalance of the dataset classes where the ironic samples are more than the non-ironic samples. Also using Arabic stemmer may improve the classification results where in our experiments we use the basic text preprocessing tasks as shown in 3.2. Fig. 1. and Fig. 2. below show the top 10 frequent words in both ironic and not-ironic tweets.

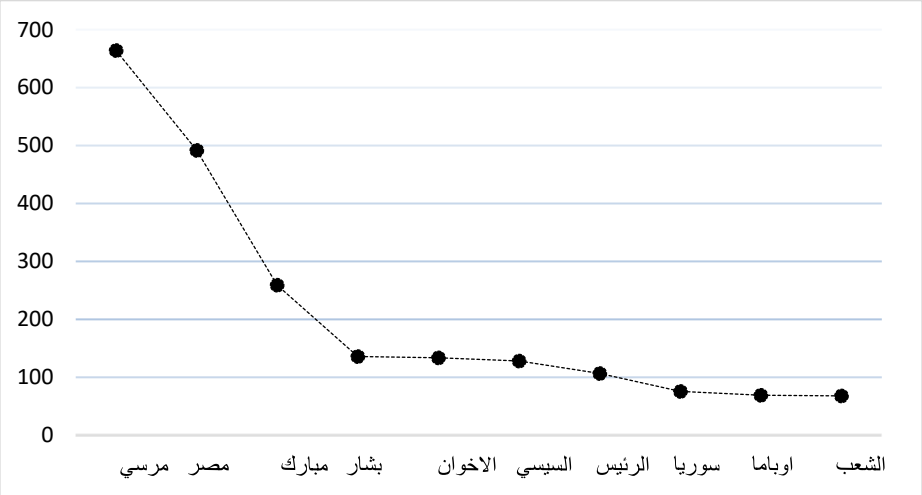

Fig. 1. Top 10 frequent words in ironic tweets from the highest to the lowest: Morsi, Egypt, Mubarak, Bashar, Brothers, Alsisi, President, Syria, Obama, and The Prople.

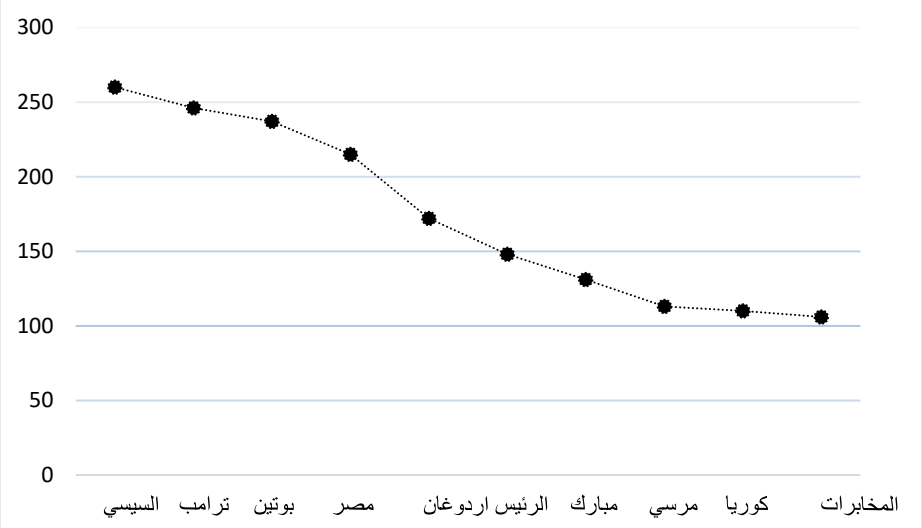

Fig. 2. Top 10 frequent words in non-ironic tweets from the highest to the lowers: Alsisi, Trump, Putin, Erdogan, President, Mubarak, Morsi, Korea, and Intelligence

References

1. Reyes, A., Rosso, P., & Buscaldi, D.: Humor in the blogosphere: First clues for a verbal humor taxonomy. Journal of Intelligent Systems, 18(4), 311-332 (2009).
2. Sarmento, L., Carvalho, P., Silva, M. J., & De Oliveira, E.: Automatic creation of a reference corpus for political opinion mining in user-generated content. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion (pp. 29-36) (2009).
3. Kreuz, R.:Using figurative language to increase advertising effectiveness. In Office of naval research military personnel research science workshop. Memphis, TN (2001).
4. Utsumi, A.: A unified theory of irony and its computational formalization. In Proceedings of the 16th conference on Computational linguistics-Volume 2 (pp. 962-967). Association for Computational Linguistics (1996).
5. Van Hee, C.: Can machines sense irony?: exploring automatic irony detection on social media (Doctoral dissertation, Ghent University) (2017).
6. Hernández-Farías, I., Benedí, J. M., & Rosso, P.: Applying basic features from sentiment analysis for automatic irony detection. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 337-344). Springer, Cham(2015).
7. Reyes, A., Rosso, P., & Veale, T.: A multidimensional approach for detecting irony in twitter. Language resources and evaluation, 47(1), 239-268 (2013).
8. Reyes, A., Rosso, P., & Buscaldi, D.: From humor recognition to irony detection: The figurative language of social media. Data & Knowledge Engineering, 74, 1-12 (2012).
9. Barbieri, F., & Saggion, H.: Modelling irony in twitter. In Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (pp. 56-64) (2014).
10. Buschmeier, K., Cimiano, P., & Klinger, R.: An impact analysis of features in a classification approach to irony detection in product reviews. In Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 42-49) (2014).
11. Benamara, F., Grouin, C., Karoui, J., Moriceau, V., Robba, I.: Introduction to the French irony Detetcion Shared Task (Analyse d'opinion et langage figuratif dans des tweets : présentation et résultats du Défi Fouille de Textes DEFT2017) DEFT@TALN2017. Orléans, France (2017).
12. Cignarella, A. T., Frenda, S., Basile, V., Bosco, C., Patti, V., & Rosso, P.: Overview of the evalita 2018 task on irony detection in italian tweets (ironita). In Sixth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian (EVALITA 2018) (Vol. 2263, pp. 1-6). CEUR-WS (2018).
13. Karoui, J., Zitoune, F. B., & Moriceau, V.: Soukhria: Towards an irony detection system for arabic in social media. Procedia Computer Science, 117, 161-168 (2017).
14. Bilal Ghanem, Jihen Karoui, Farah Benamara, Véronique Moriceau and Paolo Rosso. IDAT@FIRE2019: Overview of the Track on Irony Detection in Arabic Tweets. In: Metha P., Rosso P., Majumder P., Mitra M. (Eds.) Working Notes of the Forum for Information Retrieval Evaluation (FIRE 2019). CEUR Workshop Proceedings. CEUR-WS.org, Kolkata, India, December 12-15.