Deep Learning in Arabic Tweets Fake News Detection

Subjects: Computer Science, Artificial Intelligence Contributor: Shatha Alyoubi , Manal Kalkatawi , Felwa Abukhodair

Fake news has been around for a long time, but the rise of social networking applications has rapidly increased the growth of fake news among individuals. Fake news negatively impacts various aspects of life (economical, social, and political). Identifying fake news manually on these open platforms would be challenging as they allow anyone to build networks and publish the news in real time. Therefore, creating an automatic system for recognizing news credibility on social networks relying on artificial intelligence techniques, including machine learning and deep learning, has attracted the attention of researchers. Using deep learning methods has shown promising results in recognizing fake news written in English.

Arabic language Twitter fake news deep learning

1. Introduction

In recent years, the rapid growth of social networks has facilitated the exchange of news among users. Social networks can be utilized to inform society regarding the latest news, but they can also be a source of fake news. Twitter is considered one of the most widespread social networks in the Arabian area ^[1]. Posting news on Twitter is less costly in terms of both money and time than any other medium. Its simplicity and lack of content monitoring enable fake news to reach a wide range of users rapidly ^[2]. Fake news refers to false, misleading, and fabricated news delivered intentionally ^[3]. Fake news dissemination aims to deceive the audience for political, social, or financial gains. Consequently, fake news imposes significant risks on individuals, organizations, and governments ^[4]. Thus, there is an urgent need for efficient techniques to detect and eliminate fake news in social networks to prevent its negative impacts.

Many fact-checking websites, such as Anti-Rumors Authority and Misbar, have been implemented to check news veracity propagated on the internet in an early attempt to reduce the impact of fabricated news. These websites depend on human experts to manually confirm or reject the validity of the news ^[2]. It consumes time and effort to deal manually with a large volume of news and is not scalable. Most recently, the adoption methods of machine learning and deep learning have become popular in tackling anomaly detection problems such as fake news detection ^[5]. Two types of learning techniques are currently adopted in these methods to construct automated systems for false news detection on social networks, which are news content-based learning and social context-based learning. News content-based methodology mainly focuses on the writing style of the text content of news to discover syntactic or semantic patterns to classify news. The social context-based methodology primarily analyses

user behaviour and engagement in social media. Social context features can be explored from the user profile, discussions, and connected networks among users ^[4].

Informal writing styles, spelling errors, using diverse dialects, etc., makes processing Arabic content on social media more difficult. Other challenges that aggravate the processing complexity are the massive vocabulary and complex morphological patterns of the Arabic language. Moreover, the limited availability of Arabic datasets. These difficulties result in little research focused on detecting and eliminating fake news in Arabic. Few studies have proposed models using deep learning algorithms to identify fabricated news posted on the Twitter platform ^{[3][5][6]}. In general, these models were trained to target a specific topic of fake news posts and relied only on the textual content of the Tweets to produce a classification. Furthermore, the detection performance of the current models still requires improvement.

2. Deep Learning in Arabic Tweets Fake News Detection

Detecting fake news in Arabic is still in its infancy compared with other languages, such as English. The models are discussed below and summarized in **Table 1**.

Ref	Year	Dataset	Торіс	Classification Approach	Featur News Content	e Type Social Context	Textual Feature Representations	Result
[7]	2018	800 Tweets	General	NB, SVM, DT	✓	1	-	Accuracy 0.899
[<u>8</u>]	2019	177 Tweets	General	EM	1	s	-	F1-score 0.80
[<u>9]</u>	2019	268 labeled blog posts, 20,392 unlabeled blog posts	General	CNN	s		Word2vec(CBOW), char-level embeddings	F1-score 0.63
[<u>10</u>]	2019	9000 Tweets	General	RF, SVM, DT, NB	✓	1	-	F1-score 0.776
[<u>11</u>]	2019	1862 Tweets	Syrian crisis	LR, RF, DT, AdaBoost	✓	1	-	Accuracy 0.76
[<u>12]</u>	2020	4547 news	General	LSTM, mBERT	1		Word-level embeddings, char- level embeddings, mBERT	F1-score 0.643

Table 1. A summary of existing detection approaches for fake news in Arabic.

Ref	Year	Dataset	Торіс	Classification Approach	Featur News Content	e Type Social Context	Textual Feature Representations	Result
[<u>13</u>]	2020	AraNews (97,310 news), ATB (48,655 news), ANS (4547 news)	General	mBERT, AraBERT, XLM-RBase, XLM-RLarg	4		mBERT, AraBERT, XLM-RBase, XLM- RLarg	F1-score 0.70
[<u>14]</u>	2020	6895 news articles	Political	NB, XGBoost, CNN	1		BOW, TF-IDF, fastText	F1-score 0.984
[<u>1</u>]	2021	1862 Tweets	Syrian crisis	KNN, DT, NB, LR, LDA, SVM, RF, XGboost	1	\$	TF, TF-IDF, BoW	Accuracy 0.82
[<u>15</u>]	2021	37,000 Tweets	COVID- 19	NB, LR, SVM, MLP, RF, XGB	1		BOW, TF-IDF	F1-score 0.933
[<u>6]</u>	2021	10,828 Tweets	COVID- 19	AraBERT, mBERT, distilBERT- multi, mBERT COV19, AraBERT COV19	4		AraBERT, mBERT, distilBERT-multi, mBERT COV19, AraBERT COV19	F1-score 0.9578
[16]	2021	COVID-19- Fakes (70,959 Tweets), ArCOV19- Rumors (3032 Tweets), ANS (4091 news), AraNews (108,194 news)	COVID- 19, general	CNN, RNN, GRU, AraBERT v1, AraBERT v2, AraBERT v02, QARiB, Ar- Electra, Marbert, Arbert	4		Word2vec, fastText, doc2vec, glove, AraBERT v1, AraBERT v2, AraBERT v02, QARiB, Ar-Electra, MARBERT, Arbert	F1-score 0.95
[<u>3]</u>	2021	8786 Tweets	COVID- 19	XGB, RF, NB, SVM, SGD, CNN, RNN, CRNN	1		TF-IDF, word2vec, fastText	F1-score 0.54
[<u>17]</u>	2022	3157 Tweets	COVID- 19	LR, KNN, CART, SVM, NB, RF,	1	s	TF-IDF, glove	F1-score 0.935

Ref	Year	Dataset	Торіс	Classification Approach	Feature Type News Social ContentContext	Textual Feature Representations	Result
				AdaBoost, Bagging, ExtraTree			
[18]	2022	4299 Tweets	COVID- 19	RF, DT, XGBoost, SVM, KNN, NB, SGD, LR, RNN, BiRNN, GRU, BiGRU, LSTM, BiLSTM	1	N-Gram, TF-IDF, word2vec	Accuracy 0.81
[<u>19</u>]	2022	1098 news articles	Hajj	SVM, RF, NB	✓	-	F1-score 0.79

Detection models based on news content features make up the majority of current studies in news truth verification. Helwe et al. ^[9] developed an approach using two CNN models to assess Arabic weblog posts' credibility. Both models have similar layers, except the embedding layer. The first model used pre-trained word-level embeddings, while the second used character-level embeddings. Each model was trained on a labelled dataset in the first iteration, and then the predictions of unlabelled data for each model were picked to re-train the other model. In the experiment, they compared the proposed model to a support vector machine (SVM) trained with a TF-IDF feature representation, CNN-trained with character-level vector representation, CNN-trained with word-level vector representation, and a combined model based on Word-CNN and Char-CNN. Their proposed model scored the highest F1-score of 0.63. The fundamental limitation of the work is that the amount of labelled data is small.

The previous study ^[9] focused on examining the credibility of news blogs. Here, some works ^[12](13)(19)</sup> assessed news verification models using fabricated news generated from real news stories by modifying their semantics. Jude Khouja ^[12] proposed a system for claim verification based on the textual information of news. The work introduced a publicly available Arabic News Stance (ANS) dataset to determine the claims' veracity. The author acquired a subset of news titles from the Arabic news texts (ANT) dataset. The news titles were modified to generate fake claims. Two approaches have been used to train and test a generated dataset for claim classification: long short-term memory (LSTM) and pre-trained BERT model. The study reported that LSTM achieved the highest result for false claims recognition with an F1-score of 0.643. Nagoudi et al. ^[13] developed a method for automatically manipulating real multi-topic news to generate a fake news dataset, AraNews. They used transformer-based pre-trained models to detect manipulated Arabic news. Furthermore, they experimented with various modelling settings to examine the impact of their generated data on fake news verification models compared to a human-created fake news dataset. The authors reported that automatically generated news positively affects the fake news detection task. Compared to previous work ^[12], it achieved a better improvement with an F1-score of 0.0576. The ANS and AraNews datasets were utilized by another work ^[16]. The work intended

to examine the performance of language models such as AraBERT, QARiB, and AraGPT2 when applied to the Arabic fake news detection task. Each model was trained and evaluated using the ANT and AraNews datasets. The results showed that AraBERT and QARiB revealed some ability to identify false news with a similar accuracy of 0.80. In both experiments, AraGPT2 achieved the lowest accuracy. Himdi et al. ^[19] proposed a machine learning model to assess the veracity of Arabic news articles. They gathered factual news articles related to a single domain, which is the Hajj. After this, the acquired dataset was utilized to construct fake news articles relying on crowdsourcing. They extracted a set of linguistic features, including emotional, syntactical, polarity, and part of speech. The extracted features were used to train three classifiers, Naïve Bayes (NB), Random Forest (RF), and SVM, to detect Arabic false news. The results demonstrated that the extracted linguistic features could effectively detect fake news, and the best classifier was RF, which had a 0.79 accuracy rate.

Another approach exploited textual features defined in ^[14] to automatically identified satire news. They released a dataset that was collected from a variety of news websites. They analysed the linguistic properties of news and concluded that false news involves highly positive and negative keywords and tends to be written in a more subjective tone. Machine learning and deep learning models have been trained to identify satirical news. A CNN with pre-trained word embeddings achieved the highest performance with an accuracy of 0.98.

During the COVID-19 pandemic, a lot of false information was disseminated through various social networking applications. The effect of misinformation is not confined to individual lives but also includes society and the economy. Several studies ^{[3][6][15][16][17][18]} have concentrated on assessing the credibility of information related to the spread of COVID-19 in Arabic communities via social media platforms such as Twitter.

Combining information from both news content and social context sources may result in a better detection rate. Incorporating social context features, such as user behaviour, user profile, etc., with other news content features is uncommon in deep learning-based studies. Several efforts have been made to propose models for detecting misinformation using traditional machine learning algorithms utilizing both news content and social context aspects [1][7][8][10][11].

The detection of Arabic Tweets containing fabricated news is still in its early stages. As a result, few Arabic datasets for detecting fake news Tweets are publicly available to the research community. There have been few studies conducted to address the Arabic Tweets' veracity using deep learning techniques, and the majority of them focus on detecting fake news relating to a certain topic, such as COVID-19. Additionally, they relied only on Tweets text to produce a classification. A major challenge for these identification systems is that the underlying textual characteristics vary under different fake news. For this reason, models that used only textual content of news Tweets may have a generalizability issue. In contrast, machine learning models commonly use news content and social context features to more accurately identify various types of fake news. However, the existing models' detection performance still needs to be improved.

References

- Thaher, T.; Saheb, M.; Turabieh, H.; Chantar, H. Intelligent detection of false information in arabic tweets utilizing hybrid harris hawks based feature selection and machine learning models. Symmetry 2021, 13, 556.
- 2. Liu, Y.; Wu, Y.F.B. Fned: A deep network for fake news early detection on social media. ACM Trans. Inf. Syst. (TOIS) 2020, 38, 1–33.
- 3. Alqurashi, S.; Hamoui, B.; Alashaikh, A.; Alhindi, A.; Alanazi, E. Eating garlic prevents COVID-19 infection: Detecting misinformation on the Arabic content of Twitter. arXiv 2021, arXiv:2101.05626.
- 4. Kaliyar, R.K.; Goswami, A.; Narang, P. DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. J. Supercomput. 2021, 77, 1015–1037.
- 5. Al-Sarem, M.; Alsaeedi, A.; Saeed, F.; Boulila, W.; AmeerBakhsh, O. A novel hybrid deep learning model for detecting COVID-19-related rumors on social media based on LSTM and concatenated parallel CNNs. Appl. Sci. 2021, 11, 7940.
- 6. Ameur, M.S.H.; Aliane, H. Aracovid19-mfh: Arabic COVID-19 multi-label fake news & hate speech detection dataset. Procedia Comput. Sci. 2021, 189, 232–241.
- 7. Sabbeh, S.F.; Baatwah, S.Y. Arabic News Credibility on Twitter: An Enhanced Model Using Hybrid Features. J. Theor. Appl. Inf. Technol. 2018, 96, 2327–2338.
- 8. Alzanin, S.M.; Azmi, A.M. Rumor detection in Arabic tweets using semi-supervised and unsupervised expectation–maximization. Knowl.-Based Syst. 2019, 185, 104945.
- 9. Helwe, C.; Elbassuoni, S.; Al Zaatari, A.; El-Hajj, W. Assessing arabic weblog credibility via deep co-learning. In Proceedings of the Fourth Arabic Natural Language Processing Workshop, Florence, Italy, August 2019; pp. 130–136.
- Mouty, R.; Gazdar, A. The effect of the similarity between the two names of twitter users on the credibility of their publications. In Proceedings of the 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 30 May–2 June 2019; pp. 196–201.
- Jardaneh, G.; Abdelhaq, H.; Buzz, M.; Johnson, D. Classifying Arabic tweets based on credibility using content and user features. In Proceedings of the 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT), Amman, Jordan, 9–11 April 2019; pp. 596–601.
- 12. Khouja, J. Stance prediction and claim verification: An Arabic perspective. arXiv 2020, arXiv:2005.10410.
- 13. Nagoudi, E.M.B.; Elmadany, A.; Abdul-Mageed, M.; Alhindi, T.; Cavusoglu, H. Machine generation and detection of Arabic manipulated and fake news. arXiv 2020, arXiv:2011.03092.

- 14. Saadany, H.; Mohamed, E.; Orasan, C. Fake or real? A study of Arabic satirical fake news. arXiv 2020, arXiv:2011.00452.
- 15. Mahlous, A.R.; Al-Laith, A. Fake news detection in Arabic tweets during the COVID-19 pandemic. Int. J. Adv. Comput. Sci. Appl. 2021, 12, 778–788.
- Al-Yahya, M.; Al-Khalifa, H.; Al-Baity, H.; AlSaeed, D.; Essam, A. Arabic fake news detection: Comparative study of neural networks and transformer-based approaches. Complexity 2021, 2021, 5516945.
- 17. Qasem, S.N.; Al-Sarem, M.; Saeed, F. An ensemble learning based approach for detecting and tracking COVID19 rumors. Comput. Mater. Contin. 2021, 70, 1721–1747.
- 18. Amoudi, G.; Albalawi, R.; Baothman, F.; Jamal, A.; Alghamdi, H.; Alhothali, A. Arabic rumor detection: A comparative study. Alex. Eng. J. 2022, 61, 12511–12523.
- 19. Himdi, H.; Weir, G.; Assiri, F.; Al-Barhamtoshy, H. Arabic fake news detection based on textual analysis. Arab. J. Sci. Eng. 2022, 47, 10453–10469.

Retrieved from https://encyclopedia.pub/entry/history/show/108516