State of the Art Models for Fake News Detection Tasks

Wissam Antoun*, Fady Baly*, Rim Achour†, Amir Hussein* and Hazem Hajj*
*Electrical and Computer Engineering Department,
†Graduate Program in Computational Science,
American University of Beirut
Beirut, Lebanon
Emails: {wfa07, fgb06, rga34, anh21, hh63}@aub.edu.lb

***Abstract*—This paper presents state of the art methods for addressing three important challenges in automated fake news detection: fake news detection, domain identification, and bot identification in tweets. The proposed solutions achieved first place in a recent international competition on fake news. For fake news detection, we present two models. The winning model in the competition combines similarity between the embedding of each article's title and the embedding of the top five corresponding google search results. The new model relies on advances in Natural Language Understanding (NLU) end to end deep learning models to identify stylistic differences between legitimate and fake news articles. This second model was developed after the competition and outperforms the winning approach. For news domain detection, the winning model is a hybrid approach composed of named entity features concatenated with semantic embeddings derived from end to end models. For twitter bot detection, we propose to use the following features: duration between account creation and tweet date, presence of a tweet's link, presence of user's location, other tweet's features, and the tweets' metadata. Experiments include insights into the importance of the different features and the results indicate the superior performances of all proposed models.**

***Index Terms*—Fake News, Topic Identification, News Domain Detection, Twitter Bot detection**

I. Introduction

Fake news articles are typically created with the goal of deceiving or misleading readers [1]. As an example, earlier cases for writing fake news articles were used to increase profit by directing web-traffic with "Clickbait" content[1], and were often designed and written to go viral by targeting controversial topics.

Nowadays, the majority of people rely on social media as their news source. In the USA, 62% of American adults get some of their news from social networking sites [2]. Generally, non-expert users can't infer the validity of news they read, and humans have 70% success rate in fake news detection [3]. As a result, social media networks have become fertile grounds for spreading fake news, which is an emerging type of cybersecurity threat. Accordingly, it has become important to provide tools for automated detection of fake news.

There have been numerous research and industry efforts to automate and highlight fake news. Facebook is among the platforms with a large share of false news articles such as those preceding the 2016 US election [4]. To address the risk for fake news spreading, Facebook started adding tags to stories that can be flagged as false by fact-checkers. For the upcoming US election, Facebook is planning to label posts as "False Information", but this new policy will exclude Facebook ads placed by politicians [5]. In 2017, another effort was spearheaded for fact checking as the first step in fake news detection. Fake News Challenge (FNC-1) [6] was organized to develop new advances in intelligent systems for stance detection to predict one of four stance labels when comparing a document to its headline: Agree, disagree, discuss, and unrelated. The top ranked system used a weighted average model of a Convolutional Neural Network (CNN) and a gradient-boosting decision tree model [7]. The top three ranked models were further evaluated [8], and it was concluded that even the best performing features were not able to fully resolve the hard cases where even humans confused the agree, disagree and discuss labels. Another work on fake news detection focused on linguistic features such as Ngrams, punctuation, syntax, readability metrics and Psycholinguistic features (using the Linguistic Inquiry and Word Count (LIWC) tool) [3]. These features show that the difference in the style of writing can be exploited to detect the legitimacy of the content.

Despite the progress in fake news detection, accuracy performances remain limited in practical systems. To advance the field from this perspective, a contest was recently held as part of the Qatar International Cybersecurity Contest (QICC) [9]. The contest aimed at providing systems that have the best performances in practical applications with three challenging tracks:

- The Fake News Detection task aimed at detecting if an article is fake news or legitimate news.
- The News Domain Detection task aimed at detecting the news domain of an article: Politics, Business, Sports, Entertainment, Technology, or Education. The motivation is that different domains may require different approaches for the detection of fake news.
- The Twitter Bot Detection task aimed at detecting if a Twitter account belongs to a Human or is a Bot account.

In this paper, we present our approaches that tackled these

[1]Clickbait: is a kind of deceptive or misleading false advertisement that exploits users curiosity to attract attention to follow a link - wikipedia.org/wiki/Clickbait

978-1-7281-4821-2/20/$31.00 ©2020 IEEE

challenges. For fake news detection, our approach was motivated by the assumption that fake news would show stylistic differences hidden in the writing. The method consists of a classifier that uses state-of-the-art transformer-based language models to extract features from the text with the objective of detecting deep semantic differences in the writing. This method was inspired by the recent release of the 1.5B parameter GPT-2 natural language generation model along with a RoBERTa based detector that helps detect the output of the GPT-2 models [10]. For news domain identification, we also employ transformer-based models since these models can better distinguish between different domains by relying on their language comprehension gained from pretraining. We also compare with an approach which relied on building a domain-labeled dictionary combined with contextualized word embeddings. Finally, for the Twitter bot detection, we extract features for the tweets' text and combine them with the tweets' metadata.

The key contributions presented in this paper are the introduction of state-of-the-art approaches for fake news detection, news domain identification, and twitter bot detection. Additionally, we provide insights into feature importance for fake new detection.

The rest of the paper is organized as follows: In section II, we present a brief overview of the literature for the three challenge areas of fake news detection. Section III describes the proposed methods and section IV covers the experiments and results. Finally, the conclusion is presented in section V.

II. Related Work

A. Fake News Detection

Current research on fake news detection can be divided into three types of approaches: propagation based, source analysis, and content based. Propagation-based research suggests that the spread of fake news behaves differently than reliable news. These dissemination patterns can be used to flag news as false or true based on the propagation map [11]. Source analysis approaches depend completely on analysing the source of the news piece and its behavior. This allows for early detection and for a more robust way to contain the spread of false news [12]. Content based techniques focuses on extracting linguistic features, both lexical or syntactic. It assumes that fake news articles are written using deceptive language and syntactic styles [3], [13]–[16]. A new approach for stance detection was suggested by [17] combining multi-layer perceptron (MLP) representation with hand crafted features from the FNC-1 dataset. Skip-thought vectors are used to encode the headline and the body of each article. The hand crafted features include n-grams, char-grams, weighted TF-IDF score between body and heading of each article. Following the work of [17] on stance detection, [18] proposed to use bi-directional Recurrent Neural Networks (RNNs), together with neural attention, for encoding the headline of a news article, the first two sentences of a news article, and the entire news article. These representations are then combined with hand crafted features as used in [17].

B. News Domain Detection

Models for topic detection [19]–[23] can be divided into two main categories: deterministic models and probabilistic models. Deterministic models treat topics and texts as points in space through Vector Space Models (VSM) [24]. In [25], named entity recognition was used for new topic detection [20]. Some researchers usually use a tokenizer for word segmentation to model key topic detection [26]–[28]. However, new words and nonstandard writings make these models ineffective [20]. Probabilistic models treat topics and texts as probability distributions, which can be represented by Latent Dirichlet Allocation (LDA) [29], Author-LDA [30], Labeled LDA [31], TweetLDA [32], and other statistical representations [20]. Recent research uses deep learning algorithms for topic classification such as random multimodel deep learning (RMDL) introduced in [33], Trigger-aware Lattice Neural Network (TLNN) introduced in [34] and Dual-CNN [35].

C. Twitter Bot Detection

Digital bots are programs created with the intention of manipulating users for commercial, political, or social goals. These ”malicious” bots aim to spread false information such as targeting hate speech, or attacks in political campaigns. Early research to counter these bots used ”honeypots” to attract and profile the features or characteristics of such bot accounts [36]. Recent research focuses [37] on extracting features such as the number of likes, retweets, number of replies, mentions, followers to friend ratios, and URLs. In [37], it was concluded that bot accounts focused more on retweets or URL sharing compared to humans, who rely more on creativity in the tweet' content. For twitter bot identification, [38] collected a twitter bot corpus and showed that features like formality, structure, and temporal features can help in achieving high accuracy.

III. Systems Description

This section describes the proposed methods to advance the accuracy of fake news detection in three different aspects, which were also included in the QICC competition [9]: Fake News Detection, News Domain Identification, and Twitter Bot Detection.

A. Fake News Detection

The objective for this task is to develop an approach that can accurately identify whether a news article is fake or legitimate. We present two models: The winning model at the contest and a more recent deep learning approach that outperformed the winning model at the contest.

1) Contest Winning Feature-based Model: The contest winning feature-based approach consisted of separating the title from the body of each article, then extracting different similarity and lexicon features and finally the features are passed to an XGboost model.

Preprocessing: For the feature-based approach, the document heading is first separated from the body of each article. Each document body is then pre-processed [17] to remove:

extra empty spaces, newlines, non-alphanumeric characters, character words, and URLs.

Features: For the feature-based approach, the proposed features, shown in Table I, include similarity features [n-grams, char-grams, and common words tokens] between the heading and the body of each article [18], [39] and cosine similarity between the embedding of each heading and the embedding of the top five google search results of those headings. In addition, we include the frequency distribution for types of words in each article based on the lexicon capture for word types that includes: assertive verbs, factive verbs, hedges, implicatives, report verbs, bias, and subjectivity from each article [39].

Model Architecture: For the feature-based approach, several classifier models were tested with the above features including: support vector machine (SVM), naïve bayes (NB), random forest (RF) and XGboost.

TABLE I
DESCRIPTION OF THE FEATURES EXTRACTED FOR FAKE NEWS DETECTION

Features	Description
Common word tokens (0-17)	Counting the common words between the headlines and the body of the text
N-grams and Char-grams (18-43)	Cosine similarity of N-grams and Char-grams between the headlines and the body of the text
Google search similarity (44)	Cosine similarity between the embedding of the article heading and the embedding of google search results. Google-News word2vec embeddings were used.
Lexicon based features (45-54)	The following are the definitions of lexicon features as stated in [39] - Assertive verbs: represent the degree of certainty to which a proposition holds. - Factive verbs: verbs that assume the truth of the proposition. - Hedges:softens the degree of commitment to a proposition - Implicatives: generates pre-suppositions in a word. - Report verbs: capture the attitude towards the source - Subjectivity and bias: list of negative and positive words indicating opinions to detect subjective clues.

B. State of the Art Deep Learning Model

For the deep learning approach and in order for a model to detect the stylistic differences in the writing, the model has to have a deep understanding of the English language and its underlying semantics. Hence for the classifier choice, we chose to experiment with the state-of-the-art transformer-based language model and in particular XLNET [40] and RoBERTa [41]. We also compare against BERT [42]. Since the difference in the writing style can be hidden in every word, no pre-processing is needed for this model. We intentionally leave the articles as they are, and let the respective tokenizer of each language model handle the cleaning and tokenization. The deep learning classification models used are all based on pre-trained language models with a classification layer on top that will be fine-tuned for the fake news task.

C. News Domain Detection

The objective for this task is to develop a model that can accurately identify the topic domain of a news article into one of the following categories: Politics, Business, Sports, Education, Entertainment, or Technology. The proposed model is shown in Fig. 1.

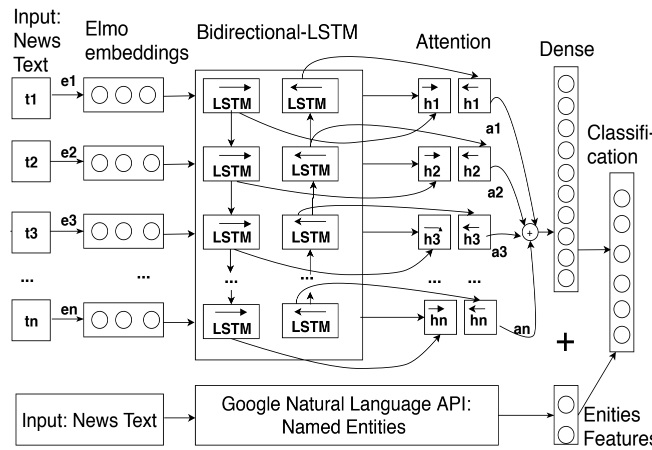

Fig. 1. News Domain Detection Model: ELMO word-embeddings with Bi-LSTM and attention concatenated with exracted named entities.

The model employs the pre-trained word embeddings of the articles followed by another feature extraction model which will be trained and fine-tuned to the news detection task. We then concatenate the resulting feature vector with frequency distribution of extracted named entities.

Features: The considered features include hand-crafted features with ELMO word-embeddings [43]. The hand-crafted features are named entities that were extracted using the Google Cloud Natural Language API and then grouped by topic. The distribution of entities is very indicative towards domains, and it is considered an important factor in the classification process.

Model Architecture: We first extract contextualized word embeddings using the pre-trained ELMO model [43]. The word embedding features are then fed into a Bidirectional Long Short Term Memory (Bi-LSTM) with an attention layer for better feature extraction [44]. The output features from this model are then concatenated with the hand-crafted features and fed into a softmax fully-connected layer. Several other models are also considered including: N-gram and N-char TF-IDF with SVM, NB, RF, XGBoost, Deep Neural Network (DNN), stacked Convolutional Neural Network (CNN), LSTM, Gated Recurrent Unit (GRU), Bi-LSTM, 3 Concatenated CNN (3CCNN), BERT, XLNET, RoBERTa and RMDL which is a combination of TF-IDF with DNNs and word-to-vector embedding using Glove [45] with RNNs and CNNs. We also present experiment with pre-trained language models (BERT, XLNET, RoBERTa) with a classification layer.

D. Twitter Bot Detection

The objective for this task is to develop a model that can accurately determine whether the source of the tweet is a bot. The approach relies on the tweets' metadata in addition to features extracted from the tweets' text.

Features: The input features of the classifier model are as follows:

- Duration between account creation and tweet date
- Number of Retweets
- Number of Favorites
- Followers to Follow Ratio
- Existence of media
- If the user is verified
- Location availability
- Existence of a bio
- Presence of a website in the user's metadata
- The number of hashtags in the tweets
- The existence of a direct quote (the tweet contains a ":")
- The number if emojis in the tweet
- If the tweet is a re-tweet
- Existence of a link in the tweet

Model Architecture The proposed model is presented in Fig. 2 and it composed of a voting classifier on the output of three ensemble classifiers: RFC, AdaBoost [46] and XGBoost [47] classifiers.

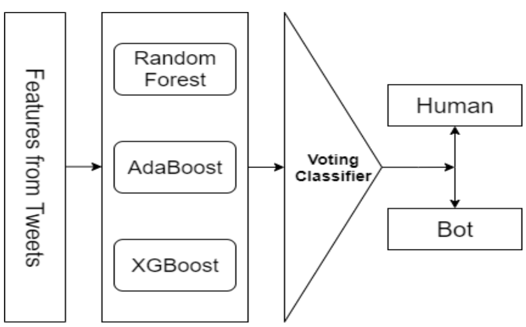

Fig. 2. Voting Classifier for ensemble tree methods.

IV. Experiments and Results

In this section, we show the results of the different approaches used in each task and highlight the insights behind the results. We also present an analysis of the different features from the fake news task.

A. Datasets

The dataset, provided by QICC, consisted of 384 articles for training (192 fake and 192 legitimate) the models for fake news detection and domain identification (Tracks 1A.1 and 1A.2). Additionally, 48 articles (24 fake and 24 legitimate) were provided as the development set. Most of the articles consisted of heading and bodies however some of them were without heading. Moreover, the text was given in English. The News Domain dataset consisted of the same articles from the fake news track and are split into 6 categories: Politics, Business, Sports, Entertainment, Technology, and Education. There were 64 articles per category for training and 8 articles per category for development. The final submission for evaluation consisted of 45 articles. As for the Twitter bot detection dataset, it consisted of 18384 tweet for training and 1615 as a development set. The metadata information included: tweet text, screen name, full name, time of tweet, user account creation date, number of followers, follows, listed, location, biographical details, verification status, links, bio and platform. To note, the text of the tweets is mostly in Arabic.

B. Fake News Detection

All experiments with transformer-based models were run on Google's Colab GPU enviroment, while the other experiments were conducted on local machines. Hyper-parameters' details for each model were obtained through randomized grid search and are available in our Github repository[2]. The precision, recall and F1-score were used for model evaluation and the results of each classifier are summarized in Table II.

TABLE II
Fake news detection results.

Model	Precision (%)	Recall (%)	F1-score (%)
NB	72	62	58
SVM	85	85	85
RF	88	88	87
Xgboost	90	90	90
mBERT-base	92	**100**	96
XLNET-base	**98**	98	**98**
RoBERTa	92	**100**	96

The results show that, while the feature-based ML achieved good results, the pre-trained deep learning language models classifiers (BERT, XLNet, RoBERTa) achieved the best performance. These results indicate that deep understanding of the language is necessary to detect the subtle stylistics differences in the writing of the fake articles. It was also noted that during the fine-tuning process, the pre-trained language models required only one epoch to learn the objective.

Feature Ranking for Fake News: In order to determine the performance of each feature we used RF feature importance classifier. The results are shown in Fig. 3.

It was found that the most important feature is the cosine similarity feature between google search and heading of each article the similarity with google search heading results. This was expected since the major factor used in detecting fake news in the real world is by referring to the source of these articles. It was found that by training XGBoost only on similarity with google search features we obtained F1-score of 81%. This shows that the rest of the features contributed only in the remaining 9% of the final result. It is also noticeable from Fig 3 that the lexicon based features were not significant for the QICC contest dataset although, in previous work, it played a vital role in examining the stylistic indicators of news credibility [39].

[2]https://github.com/aub-mind/fake-news-detection

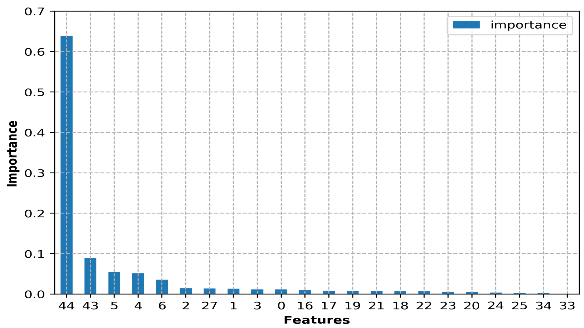

Fig. 3. Ranking feature importance for Fake News: Google search similarity feature (44) found to be the most important. The figure shows ranking of all other features: common word tokens (0-17), N-grams and Char-grams (18-43), and lexicon based features(45-54)

C. News Domain Detection

The hyper-parameters used for this task are detailed in our Github repository. The results of this task are shown in Table III.

TABLE III
NEWS DOMAIN DETECTION RESULTS.

Model	Precision(%)	Recall(%)	F1-score(%)
TF-IDF with NB	82	81	80
TF-IDF with SVM	73	67	68
TF-IDF with RF	70	69	68
TF-IDF with Xgboost	76	75	75
CNN*	83	82	82
LSTM*	83	79	79
GRU*	80	77	76
Bi-LSTM*	84	81	80
3CCNN*	81	81	80
Bi-LSTM/Attention*	86	85	85
Bi-LSTM/Attention**	**95**	93	**94**
mBERT-base	91	90	90
XLNET-base	93	90	89
RoBERTa	94	**94**	**94**
RMDL†	-	-	89

* Pre-trained word Embeddings (ELMO) only
** Pre-trained word Embeddings (ELMO) and exctracted entity features
† 30 RMDLs (10DNN, 10RNN, 10CNN)

Although the pre-trained language models achieved good results, the Bi-LSTM with attention in combination with the hand-crafted features were able to match them in performance. Also, the results show that the named entity features provide a significant performance improvement of 10% for the Bi-LSTM with attention model.

Furthermore, when analyzing the results of the best models, we notice that, out of the 4 classification errors, 3 articles were consistently being mislabeled. As it turns out the reason for the mislabeling, can be attributed to the article being part of two domains.

D. Twitter Bot Detection

All the evaluation results ranged between 96% to 99% with the right subset of features. The best features subset was found to be: Duration between account creation and tweet date, existence of a link in the tweet, if the user is verified, the existence of a direct quote, and the number of favorites. We noticed that, when the feature ”duration between account creation and tweet date” is used alone, a Random Forest classifier was able to score an accuracy score of 96%. Additionally, tweets with direct quotes showed a high correlation with the account being flagged as bot.

V. CONCLUSION

In this paper, we presented the state of the art models that achieved first place in an international fake news competition, while tackling three challenges: Fake News Detection, Domain Identification, and Twitter Bot Detection. For fake news detection, we concluded with state-of-the-art approach based on XLNET. For news domain detection, the winning model is based on a hybrid approach composed of frequency distribution of named entities concatenated with word embeddings from ELMO language model. For twitter bot detection system, the system is composed of features extracted from news tweets combined with the news tweets' metadata.

Insights from the experiments showed high impact features such as cosine similarity between the title's embedding and corresponding google searches. Named entities were also important for news topic identification. The experiments also showed the superiority of advances in language models that can provide a deep understanding of the language.

For future work, we suggest improving fake news performance by adding features from fact-checking websites in addition to Google searches.

REFERENCES

[1] S. Tavernise, "As fake news spreads lies, more readers shrug at the truth," Dec 2016. [Online]. Available: https://www.nytimes.com/2016/12/06/us/fake-news-partisan-republican-democrat.html
[2] J. Gottfried and E. Shearer, "News use across social media platforms 2016," Dec 2017. [Online]. Available: https://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/
[3] V. Pérez-Rosas, B. Kleinberg, A. Lefevre, and R. Mihalcea, "Automatic detection of fake news," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 3391–3401.
[4] C. Silverman, "This analysis shows how viral fake election news stories outperformed real news on facebook," Nov 2016. [Online]. Available: https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook
[5] L. H. OWEN, "Facebook is just gonna come out and start calling fake news fake (well, "false")." [Online]. Available: https://www.niemanlab.org/2019/10/facebook-is-just-gonna-come-out-and-start-calling-fake-news-fake-well-false/
[6] D. Pomerleau and D. Rao. Fake news challenge stage 1 (fnc-i): Stance detection. [Online]. Available: http://www.fakenewschallenge.org/
[7] B. Sean, S. Doug, and P. Yuxi, "Talos targets disinformation with fake news challenge victory," 2017. [Online]. Available: http://blog.talosintelligence.com/2017/06/talos-fake-news-challenge.html
[8] A. Hanselowski, A. PVS, B. Schiller, F. Caspelherr, D. Chaudhuri, C. M. Meyer, and I. Gurevych, "A retrospective analysis of the fake news challenge stance detection task," *arXiv preprint arXiv:1806.05180*, 2018.
[9] (2019) Qatar international fake news detection and annotation contest. [Online]. Available: https://sites.google.com/view/fakenews-contest
[10] I. Solaiman, "Gpt-2: 1.5b release," Nov 2019. [Online]. Available: https://openai.com/blog/gpt-2-1-5b-release/
[11] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science*, vol. 359, no. 6380, pp. 1146–1151, 2018.

[12] R. Baly, G. Karadzhov, D. Alexandrov, J. Glass, and P. Nakov, "Predicting factuality of reporting and bias of news media sources," *arXiv preprint arXiv:1810.01765*, 2018.
[13] S. Afroz, M. Brennan, and R. Greenstadt, "Detecting hoaxes, frauds, and deception in writing style online," in *2012 IEEE Symposium on Security and Privacy*. IEEE, 2012, pp. 461–475.
[14] V. Rubin, N. Conroy, Y. Chen, and S. Cornwell, "Fake news or truth? using satirical cues to detect potentially misleading news," in *Proceedings of the second workshop on computational approaches to deception detection*, 2016, pp. 7–17.
[15] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2931–2937.
[16] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *arXiv preprint arXiv:1702.05638*, 2017.
[17] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, "Combining neural, statistical and external features for fake news stance identification," in *Companion Proceedings of the The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 2018, pp. 1353–1357.
[18] L. Borges, B. Martins, and P. Calado, "Combining similarity features and deep representation learning for stance detection in the context of checking fake news," *Journal of Data and Information Quality (JDIQ)*, vol. 11, no. 3, p. 14, 2019.
[19] J. G. Fiscus and G. R. Doddington, "Topic detection and tracking evaluation overview," in *Topic detection and tracking*. Springer, 2002, pp. 17–31.
[20] P. Han and N. Zhou, "A framework for detecting key topics in social networks," in *Proceedings of the 2nd International Conference on Big Data Technologies*. ACM, 2019, pp. 235–239.
[21] Y. Cha and J. Cho, "Social-network analysis using topic models," in *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2012, pp. 565–574.
[22] T. N. Rubin, A. Chambers, P. Smyth, and M. Steyvers, "Statistical topic models for multi-label document classification," *Machine learning*, vol. 88, no. 1-2, pp. 157–208, 2012.
[23] R. Ibrahim, A. Elbagoury, M. S. Kamel, and F. Karray, "Tools and approaches for topic detection from twitter streams: survey," *Knowledge and Information Systems*, vol. 54, no. 3, pp. 511–539, 2018.
[24] J. M. Schultz and M. Y. Liberman, "Towards a "universal dictionary" for multi-language information retrieval applications," in *Topic detection and tracking*. Springer, 2002, pp. 225–241.
[25] G. Kumaran and J. Allan, "Text classification and named entities for new event detection," in *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2004, pp. 297–304.
[26] S. I. Nikolenko, S. Koltcov, and O. Koltsova, "Topic modelling for qualitative studies," *Journal of Information Science*, vol. 43, no. 1, pp. 88–102, 2017.
[27] G. Fuentes-Pineda and I. V. Meza-Ruiz, "Topic discovery in massive text corpora based on min-hashing," *Expert Systems with Applications*, 2019.
[28] H.-J. Choi and C. H. Park, "Emerging topic detection in twitter stream based on high utility pattern mining," *Expert Systems with Applications*, vol. 115, pp. 27–36, 2019.
[29] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.
[30] M. Rosen-Zvi, T. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*. AUAI Press, 2004, pp. 487–494.
[31] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 2009, pp. 248–256.
[32] D. Quercia, H. Askham, and J. Crowcroft, "Tweetlda: supervised topic classification and link prediction in twitter," in *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 2012, pp. 247–250.
[33] K. Kowsari, M. Heidarysafa, D. E. Brown, K. J. Meimandi, and L. E. Barnes, "Rmdl: Random multimodel deep learning for classification," in *Proceedings of the 2nd International Conference on Information System and Data Mining*. ACM, 2018, pp. 19–28.
[34] N. Ding, Z. Li, Z. Liu, H. Zheng, and Z. Lin, "Event detection with trigger-aware lattice neural network," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 347–356.
[35] G. Burel, H. Saif, M. Fernandez, and H. Alani, "On semantics and deep learning for event detection in crisis situations," 2017.
[36] K. Lee, J. Caverlee, and S. Webb, "Uncovering social spammers: social honeypots+ machine learning," in *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2010, pp. 435–442.
[37] Z. Gilani, R. Farahbakhsh, G. Tyson, L. Wang, and J. Crowcroft, "Of bots and humans (on twitter)," in *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM, 2017, pp. 349–354.
[38] H. Almerekhi and T. Elsayed, "Detecting automatically-generated arabic tweets," in *AIRS*. Springer, 2015, pp. 123–134.
[39] S. Mukherjee and G. Weikum, "Leveraging joint interactions for credibility analysis in news communities," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 2015, pp. 353–362.
[40] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding," *arXiv preprint arXiv:1906.08237*, 2019.
[41] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
[42] 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.
[43] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.
[44] Y. Wang, M. Huang, L. Zhao *et al.*, "Attention-based lstm for aspect-level sentiment classification," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 606–615.
[45] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Doha, Qatar: Association for Computational Linguistics, Oct. 2014, pp. 1532–1543. [Online]. Available: https://www.aclweb.org/anthology/D14-1162
[46] J.-J. Lee, P.-H. Lee, S.-W. Lee, A. Yuille, and C. Koch, "Adaboost for text detection in natural scene," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 429–434.
[47] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, "Xgboost: extreme gradient boosting," *R package version 0.4-2*, pp. 1–4, 2015.