KI^{2}TE: Knowledge-Infused InterpreTable Embeddings for COVID-19 Misinformation Detection

William Shiao
University of California Riverside
wshia002@ucr.edu

Evangelos E. Papalexakis
University of California Riverside
epapalex@cs.ucr.edu

ABSTRACT

As COVID-19 continues to spread across the world, concerns regarding the spread of misinformation about it are also growing. In this work, we propose a preliminary novel method to identify fake articles and claims by using information from the CORD-19 academic paper dataset. Our method uses the similarity between articles and reference manuscripts in a shared embedding space to classify the articles. This also provides an explanation for each classification decision that links a particular article or claim to a small number of research manuscripts that influence the decision. We collect 90K real articles and 20K fake articles about the coronavirus, as well as over 700 human-labelled claims from the Google FactCheck API, and evaluate its performance on these datasets. We also evaluate its performance on MM-COVID [13], a recent COVID-19 news dataset. We demonstrate the explainability of our model and discuss its limitations.

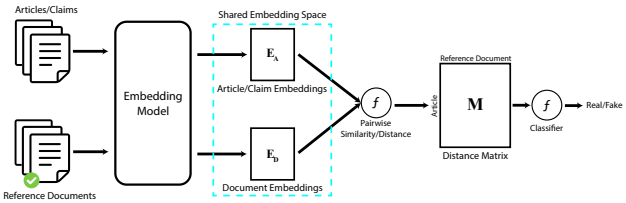

Figure 1: An illustration of the classification stages.

1 INTRODUCTION

The current time dictates an unprecedented outbreak of the novel coronavirus (SARS-CoV-2) in most countries across the world. With millions of people stuck at home and accessing information via social media platforms, there is an increasing concern about the spread of misinformation regarding the pandemic.

In the recent years, we have experienced the proliferation of websites and outlets that publish and perpetuate misinformation. However, with the pandemic and the US presidential elections in 2020, it has become a larger problem than ever. The most effective method to counter this is human fact-checking. However, this often requires domain expertise and can be prohibitively expensive. Domain expertise was an especially large issue during the early stages of the pandemic, when information about COVID-19 was limited and when conspiracy theories and snake oil "cures" propagated quickly.

Fake news has been a large issue even before the start of the pandemic. For example, misinformation was widespread over Twitter during events like Hurricane Sandy [10] and the Boston Marathon bombings [9]. Studies have also shown that humans are bad at detecting misinformation, the mean accuracy of 1,000 participants averaged over 100 runs being only 54% [20]. Furthermore, it has been shown that fake news spreads faster than real news [23], making it even more important that we combat its spread.

On top of this, the recent spread of misinformation about COVID-19 poses some new issues. Information about the virus has been sparse, especially during the start of the pandemic. This makes it harder for the average person to differentiate between true and false information. Information about the virus also evolves fairly quickly.

Many different approaches for fake news classification have been proposed. One class of approaches revolve around checking whether or not statements are likely to be connected in a knowledge graph [5–8]. The downside to this approach is that it requires the user to either create a new knowledge graph for the task or use an existing one. Creating a new knowlege graph is often difficult and are usually built with some human supervision [25]. However, Wang *et al.*[25] shows that deep language models like BERT [4] and the GPT models [2, 17] can be used to build knowledge graphs directly. This suggests the models retain a lot of the knowledge acquired from training on datasets.

Several recent state-of-the-art fake news detection models rely on a BERT architecture for processing text [12, 15, 27, 28]. While BERT tends to perform well for this task, a common issue is the lack of explainability in its classification decisions.

In this work, we present a preliminary model that uses S-BERT [19] embeddings to construct a similarity matrix against a set of reference documents. This allows us to explain classification decisions as a function of the article's similarity to specific documents if we train an interpretable classifier like a random forest or logistic regression.

While this model is relatively simple and can be further refined, we believe that this approach provides an interesting and useful step towards interpretable high-performance models.

An overview of our contributions are shown here:

- **Novel embedding scheme:** We propose KI^{2}TE, a novel embedding scheme built on top of other embedding models.
- **Dataset collection**: We gather over 100K news articles with coarse labels.
- **Extensive evaluation**: We evaluate the performance and explainability of KI^{2}TE on 3 different datasets.

KnOD'21 Workshop - April 14, 2021

2 PROBLEM FORMULATION & PROPOSED METHOD

2.1 Problem Definition

Given
- a set **A** of labelled article/claims about COVID-19.
- a set **D** of credible reference documents.

Classify each article/claim $a \in \mathbf{A}$ as real or fake.
Explain the classification decision as a function of **D**.

2.2 Proposed Method

We first embed each article/claim in **A** and each document in **D** into a shared embedding space. We found using Sentence-BERT (SBERT) [19] for this step led to the best results, but we also evaluate the performance of our method using FastText [11]. We then calculate the pairwise similarity between each article/claim and each reference document, which gives us a distance matrix **M**. Each row of **M** can be thought of as a new embedding for the corresponding article in **A**. We then train a classifier on **M**. These steps are described as pseudocode in Algorithm 1 below.We evaluate our method using logistic regression and a random forest, both of which offer a good balance between performance and interpretability.

Algorithm 1 Given a set of articles and reference documents, returns KI^{2}TE embeddings.

```
procedure KI²TE(A, D)
    E_A ← ComputeEmbeddings(A)
    E_D ← ComputeEmbeddings(D)
    for a_i ∈ E_A do
        for d_j ∈ E_D do
            M_{i,j} ← dist(a_i, d_j)
        end for
    end for
    return M
end procedure
```

2.3 Model Explainability

When trained with a interpretable classifier, this approach allows us to explain classification decisions on an article with supporting documents **D**. We evaluated our approach using two models: logistic regression and a random forest.

Logistic regression trains a weight vector w and bias b such that the cross-entropy is minimized. The magnitude of a weight w_j corresponds to the importance of a feature $\mathbf{M}_{i,j}$ in article $\mathbf{A}_i$. We can find the importance of that feature in a classification decision with $w_j \times \mathbf{M}_{i,j}$. Since $\mathbf{M}_{i,j}$ corresponds to the distance to document $\mathbf{D}_i$, we can see how much each document contributes to the classification decision.

A random forest involves training a set of decision trees on random samples of the training dataset. The classification results is the mode of the classification results of each of the trees in the forest. The prediction function of a random forest can be written out in terms of the sum of feature contributions [21]. This allows us to see which documents led to a specific classification decision in a random forest.

2.4 Compared to KNN

At first glance, this approach may appear to be similar to a K-nearest-neighbors (KNN) classifier trained on the reference embedding matrix and used to classify articles. However, they are different and there are several key advantages of our approach:

(1) The reference data can be one-class data, like in our use case, where all of the CORD articles are considered to be accurate.
(2) In KNN, each of the k nearest neighbors are considered to be of equal importance, but each neighbor is assigned a different weight in our approach.
(3) In KNN, only the k nearest neighbors are considered for classification, but we consider all of the data points in our approach.

However, one advantage of KNN over our approach is that KNN scales better when there are more reference documents, especially if a database that supporting approximate nearest-neighbors is used. We talk more about this limitation in Section 3.5, as well as ways to reduce its impact.

3 EXPERIMENTAL EVALUATION

We evaluate the performance of our method along three aspects:

(1) The classification accuracy and F1 score on the Google FactCheck claims, the MM-COVID [13] dataset, and our gathered set of news articles.
(2) The explainability of our method.
(3) The sensitivity of our model with respect to the number of documents.

3.1 Classification Performance

We evaluate the accuracy and F1 score of our model, and similar baseline models on the 3 datasets described in Section 3.4. We also evaluate them on 3 different pieces of the news dataset, as described in Section 3.2.

3.2 Explanability

In Fig. 2 we show a four classification results, with the top contributors to each decision. Due to space and copyright considerations, we provide only the titles of the articles and manuscripts. However, results are taken from a model trained on subsets of the news articles focused on vaccine and transmission news.

The reason for this is that only a small portion of the news articles contain information also present in CORD-19 documents. Below are the titles of 5 articles that have poor explanability in our model:

1) "Kevin Ferris: John Prine, thanks for the many blessings you shared through your life and music"
2) "San Bernardino County reports 4 more coronavirus deaths, 146 new cases"
3) "Trump indicates he no longer has the coronavirus, says he is 'immune'"
4) "Gary Neville slams EPL teams: Clubs are frightened"

	News		Claims	
Model	Acc.	F1	Acc.	F1
BERT + KI^{2}TE + RF	0.729 ± 0.003	0.786 ± 0.002	0.921 ± 0.02	**0.524** ± 0.479
BERT + RF	0.757 ± 0.004	0.809 ± 0.004	**0.926** ± 0.015	0.03 ± 0.074
BERT + KI^{2}TE + LR	0.742 ± 0.003	0.714 ± 0.054	0.913 ± 0.005	0.5 ± 0.498
BERT + LR	**0.791** ± 0.004	0.802 ± 0.036	0.904 ± 0.026	0.211 ± 0.064
FT + KI^{2}TE + RF	0.714 ± 0.004	0.607 ± 0.008	0.912 ± 0.014	0.499 ± 0.5
FT + RF	0.788 ± 0.004	0.804 ± 0.045	0.922 ± 0.012	0.051 ± 0.079
FT + KI^{2}TE + LR	0.773 ± 0.004	**0.810** ± 0.003	0.907 ± 0.016	0.474 ± 0.52
FT + LR	0.755 ± 0.004	0.724 ± 0.056	0.901 ± 0.02	0.0 ± 0.0
	MM-COVID		**Filtered News**	
Model	Acc.	F1	Acc.	F1
BERT + KI^{2}TE + RF	0.89 ± 0.011	0.778 ± 0.018	0.834 ± 0.01	0.906 ± 0.006
BERT + RF	0.922 ± 0.005	0.948 ± 0.003	0.839 ± 0.01	0.908 ± 0.006
BERT + KI^{2}TE + LR	0.918 ± 0.008	0.846 ± 0.017	0.843 ± 0.008	**0.91** ± 0.005
BERT + LR	**0.943** ± 0.002	**0.962** ± 0.001	**0.847** ± 0.01	0.909 ± 0.007
FT + KI^{2}TE + RF	0.853 ± 0.008	0.681 ± 0.021	0.828 ± 0.018	0.903 ± 0.011
FT + RF	0.901 ± 0.004	0.935 ± 0.003	0.831 ± 0.012	0.904 ± 0.008
FT + KI^{2}TE + LR	0.899 ± 0.01	0.806 ± 0.019	0.826 ± 0.008	0.903 ± 0.005
FT + LR	0.864 ± 0.003	0.912 ± 0.004	0.822 ± 0.009	0.902 ± 0.005
	Vaccine News		**Transmission News**	
Model	Acc.	F1	Acc.	F1
BERT + KI^{2}TE + RF	0.865 ± 0.002	0.925 ± 0.001	0.79 ± 0.006	0.865 ± 0.005
BERT + RF	0.866 ± 0.003	0.926 ± 0.002	0.805 ± 0.007	0.875 ± 0.006
BERT + KI^{2}TE + LR	0.869 ± 0.003	0.926 ± 0.002	0.797 ± 0.014	0.866 ± 0.009
BERT + LR	**0.886** ± 0.002	**0.934** ± 0.001	**0.833** ± 0.013	**0.889** ± 0.009
FT + KI^{2}TE + RF	0.86 ± 0.002	0.923 ± 0.001	0.766 ± 0.01	0.855 ± 0.007
FT + RF	0.876 ± 0.002	0.931 ± 0.001	0.807 ± 0.015	0.881 ± 0.01
FT + KI^{2}TE + LR	0.873 ± 0.004	0.929 ± 0.002	0.751 ± 0.006	0.847 ± 0.004
FT + LR	0.853 ± 0.005	0.919 ± 0.003	0.725 ± 0.017	0.839 ± 0.011

Table 1: Top: Results on the news, claims, and MM-COVID [14] datasets. Bottom: Results on samples of the news datasets. RF stands for random forest, LR stands for logistic regression, and FT stands for FastText [11].

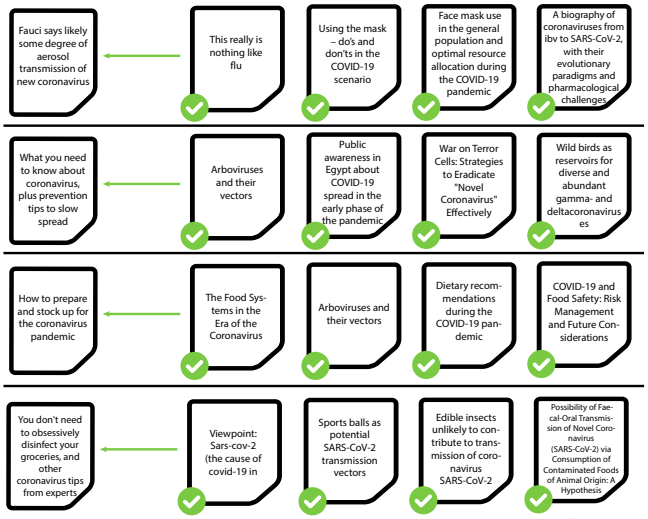

Figure 2: Four sample classification results of real articles (only titles shown) with the top contributors (only titles shown) to the decision in a random forest classifier. The model was trained on the vaccine and transmission subsets of the news articles, as described in Section 3.2.

5) "Coronavirus: Indian takeaway offering free toilet rolls with orders over £20"

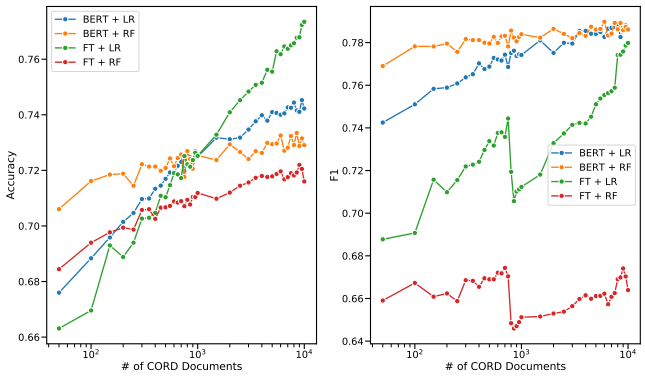

Figure 3: Accuracy (left) and F1 score (right) as the number of reference documents increases.

We can see that (1) is about the death of a celebrity from coronavirus, and it is unlikely that any CORD-19 document would have a reference to it. (2) is about relatively small area in the U.S. and would likely not have any references to it in CORD-19. (3) is political news and also likely does not have many CORD-19 references. (4) is primarily sports news and does not contain information about the virus itself. (5) is about a specific restaurant and will not have any related information in CORD-19.

However, KI^{2}TE still maintains similar accuracy to our baseline models. This is because the document distances also serve as a proxy to the raw embeddings, allowing it to maintain much of the information from the original BERT/FastText embeddings. However, the explainability of our model suffers in this case. To resolve this, we extract 3 versions from the news dataset.

We extract a filtered set, which has articles with sports teams and popular cities/countries removed, and refer to it as the "Filtered News" dataset. We also extract only articles that contain the word "vaccine" and call this the "Vaccine News" dataset. Finally, we extract only articles that contain the word "transmission" and name it the "Transmission News" dataset. The purpose of the last two datasets is to provide a smaller sample with articles focusing more on attributes of the virus, rather than on other topics (like those shown above). The performance of our models on these datasets are shown in Table 1 above.

3.3 Sensitivity to Number of Reference Manuscripts

We evaluate the accuracy and F1 score of KI^{2}TE as the number of reference documents increase, and the results can be seen in Fig. 3. Generally, we can see that as the number of reference documents increase, the accuracy and F1 score of KI^{2}TE increases. However, increasing the number of reference documents has a diminishing effect. Interestingly, the FastText-based models exhibit a large dip in F1 score after about 1,000 documents, but it recovers and continues to increase.

3.4 Datasets

In this section, we describe the steps involved in the data collection and filtering the news articles for analysis. We used five datasets for this work.

We chose to crawl our own news datasets because we were unable to find any up-to-date fake news datasets at the time of writing.

3.4.1 CORD-19. The first dataset we used was the COVID-19 Open Research Dataset (CORD-19) [26], which is a growing collection of scientific paper prepared by the White House in partnership with leading research groups characterizing the wide range of literature related to coronaviruses.

It consists of over 200,000 documents, of which 100,000 have a PDF parse of their full text. Although not all of these documents have undergone peer review and includes preprints from sites like bioRxiv, we still consider this to be a relatively credible source of information about the virus.

3.4.2 Fake News Dataset. We crawled sites from NewsGuard's Misinformation Tracking Center[1] for our fake news dataset. NewsGuard is an organization that rates the trustworthiness of websites that share information online based on their credibility and transparency. We crawled on the sites based in the United States to ensure that we crawled only English language sites. We also only crawled the sites with sitemaps to ensure that all of the crawled pages were in fact news articles, not other pages, like store pages.

We also made the assumption that all articles on any of those sites were considered fake news. While this is a very strong assumption, we could not come up with a better method for labeling individual articles. We used the Newspaper3k[2] Python library library to extract article metadata and content.

We chose to scrape only COVID-19-related news articles by filtering the crawled articles by keywords like "COVID" or "coronavirus". We also removed duplicate lines (where a line is an HTML tag, not a sentence) from the plain text of the articles. This helps prevent pages with fixed headers or taglines appearing in the document text. Otherwise, articles with mentions of keywords in the header or footer would also be included in the crawl.

Certain properties of these sites made them difficult to crawl. Some sites mixed in abstracts of academic papers in with their articles to lend credibility. Other sites mixed in articles from Reuters or the Associated Press (AP), both of which we consider reliable sources. The Newspaper3k library also tended to perform worse at extracting the content of the articles, likely because the library was mainly tested on more mainstream news websites.

Many of these sites also had other purposes in addition to providing news articles. Some of them sold alternative medicinal products like colloidal silver. Others also had videos in addition to their text articles. We did our best to clean this data, but it is possible that some of these issues are still present in the data. After cleaning, we were left with around 20K fake news articles.

3.4.3 Real News Dataset. We used the list from the B.S. Detector Chrome extension[3] to pick the reputable sites. We then collected articles from all of the matching sites from the Common Crawl News archive [16]. After that, the HTML for each article was processed in the same manner as the fake news dataset. We gathered over 95K articles that mention the novel coronavirus, but we randomly subsample from this set of articles when training our models to reduce the class imbalance.

3.4.4 Google FactCheck Dataset. We also downloaded COVID-19-related claims from the Google FactCheck API[4]. These claims are gathered from a variety of fact-checking companies and are checked by humans. Each claim consists of a single sentence (or rarely, several sentences) and a rating from a fact checking agency. This rating does not necessary follow any particular format and can range from "Fake" to other, less clear ratings like "Needs Context" or "Missing Context". We chose to exclude those ambigious claims from the dataset. This led to a total of 739 claims, of which 97 are true, and 642 are false/misleading. While there is a heavily class imbalance and it is a small dataset, we chose to include this in our evaluation to test our model's performance on small datasets with accurate labels.

3.4.5 MM-COVID Dataset. We also used the Multilingual and Multidimensional COVID-19 Fake News Data Repository (MM-COVID) dataset [13], which contains fake news from 6 different languages. However, we only focus on the English portion of the dataset. The news articles are labelled by Snopes [5] and Poynter [6], both of which are fact-checking companies that use human fact-checkers. The MM-COVID dataset also includes tweets and replies to those tweets, but we only use the text of the articles in the dataset.

3.5 Limitations

One limitation of this method is that a new feature is added for each new reference document. This can significantly reduce the performance of the classifier and greatly increase the distance matrix time calculation when the number of reference documents is large. The simplest way to mitigate this would be to simply use a random sample of reference documents, but there may be very similar reference documents selected, which would not improve the performance or interpretability of the model. Another way to mitigate it would be to use standard feature selection methods (like Lasso [22]), but this still requires the calculation of the distance matrix across all reference documents.

One solution for this is to run `k-means++` [1] on $\mathbf{E_D}$ and set k to the number of reference documents we want to use. Then, we can select the nearest neighbor to each of the k centroids of the clusters. This leaves us with k reference documents, each of which theoretically represents a different part of the embedding space. This helps reduce the chance of similar documents being selected.

4 RELATED WORK

There has been a lot of work in the area of fake news detection and models use a variety of different methods. These methods can generally be grouped into four categories: knowledge-based, style-based, propagation-based, and source-based models [29].

Knowledge-based models often attempt to compare the claims in a news article against facts stored in a knowledge base (KB) or knowledge graph (KG). Knowledge graphs are commonly represented as a set of subject-predicate-object triples, where the subject

[1] https://www.newsguardtech.com/coronavirus-misinformation-tracking-center/
[2] https://github.com/codelucas/newspaper
[3] https://gitlab.com/bs-detector/bs-detector
[4] https://toolbox.google.com/factcheck/apis
[5] https://www.snopes.com/
[6] https://www.poynter.org/

and object map to entities, which are typically represented as nodes. These models often predict the probability of triples existing in this graph and use that to determine the accuracy of a statement [5–7]. These knowledge graphs can be single-source, which uses a knowledge graph from a single source, or open-source, where the knowledge graph is created by merging data from multiple sources [29]. The downside of a knowledge-graph-based approach is that it requires a knowledge graph, which is non-trivial to construct. Many existing public knowledge graphs, like Wikidata [24], YAGO [18], and NELL [3], required some level of human supervision to construct.

Style-based models attempt to look at the style with which the article was written to assess the intentions of the author, with the assumption that fake news articles are written differently than authentic news articles. Propagation-based models look at how a news article spreads and works on a news cascade [29] or graph representation of that. Source-based models look focus on the author and publisher of new articles, with the assumption that many fake news items tend to come from the same sources. While we use the term article, all of these methods are applicable to, and been applied to other mediums, like social media posts.

5 CONCLUSION

In this work, we propose a similarity matrix-based embedding method: KI^{2}TE, which allows us to interpret the decisions of embedding-based models by linking them to a set of reference documents. We gather a coarsely-labelled dataset of news articles and human-labelled claims. We also evaluate our model on the MM-COVID [14] dataset. We show that our model has similar performance to baseline methods, with the added benefit of explainability on some classification decisions.

6 ACKNOWLEDGEMENTS

The authors would like to thank Rutuja Gurav, Pravallika Devineni, and Sara Abdali for their valuable help and feedback. Research was partially supported by the National Science Foundation Grant no. 1901379 and a UCR Regents Faculty Fellowship. This research was partially sponsored by the U.S. Army Combat Capabilities Development Command Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-13-2-0045 (ARL Cyber Security CRA). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Combat Capabilities Development Command Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] David Arthur and Sergei Vassilvitskii. 2007. K-means++: The advantages of careful seeding. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, Vol. 07-09-January-2007.

[2] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. *arXiv* (5 2020). http://arxiv.org/abs/2005.14165

[3] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. [n.d.]. *Toward an Architecture for Never-Ending Language Learning*. Technical Report. www.aaai.org

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. (10 2018). http://arxiv.org/abs/1810.04805

[5] Xin Luna Dong, Christos Faloutsos, Xian Li, Subhabrata Mukherjee, and Prashant Shiralkar. 2018. 3-FactCheckingGraph - Google Slides. https://docs.google.com/presentation/d/1JudymfQC14vpGdQY6nOodIcmOpec1vSnMa38FnSZiVY/edit#slide=id.g3fc8173fac_2_79

[6] Valeria Fionda and Giuseppe Pirrò. 2017. *Fact Checking via Evidence Patterns*. Technical Report.

[7] Mohamed H Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. 2019. Tracy: Tracing Facts over Knowledge Graphs and Text. (2019). https://doi.org/10.1145/nnnnnnn.nnnnnnn

[8] Matthew Gardner, Tom Mitchell, William Cohen, Christos Faloutsos, and Antoine Bordes. [n.d.]. *Reading and Reasoning with Knowledge Graphs*. Technical Report. www.lti.cs.cmu.edu

[9] Aditi Gupta, Hemank Lamba, and Ponnurangam Kumaraguru. 2013. $1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on twitter. In *eCrime Researchers Summit, eCrime*. IEEE Computer Society. https://doi.org/10.1109/eCRS.2013.6805772

[10] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. [n.d.]. *Faking Sandy: Characterizing and Identifying Fake Images on Twitter during Hurricane Sandy*. http://www.guardian.co.uk/world/us-news-

[11] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2017. Bag of Tricks for Efficient Text Classification. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. Association for Computational Linguistics, 427–431.

[12] Heejung Jwa, Dongsuk Oh, Kinam Park, Jang Kang, and Hueiseok Lim. 2019. exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT). *Applied Sciences* 9, 19 (9 2019), 4062. https://doi.org/10.3390/app9194062

[13] Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. [n.d.]. *MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation*. Technical Report. www.newsguardtech.com

[14] Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. 2020. MM-COVID: A Multilingual and Multimodal Data Repository for Combating COVID-19 Disinformation. (11 2020). http://arxiv.org/abs/2011.04088

[15] Chao Liu, Xinghua Wu, Min Yu, Gang Li, Jianguo Jiang, Weiqing Huang, and Xiang Lu. 2019. A Two-Stage Model Based on BERT for Short Fake News Detection. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 11776 LNAI. Springer, 172–183. https://doi.org/10.1007/978-3-030-29563-9{_}17

[16] Joel Mackenzie, Rodger Benham, Matthias Petri, Johanne R. Trippas, J. Shane Culpepper, and Alistair Moffat. 2020. CC-News-En: A Large English News Corpus. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management* (Virtual Event, Ireland) *(CIKM '20)*. Association for Computing Machinery, New York, NY, USA, 3077–3084. https://doi.org/10.1145/3340531.3412762

[17] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. [n.d.]. *Language Models are Unsupervised Multitask Learners*. Technical Report. https://github.com/codelucas/newspaper

[18] Thomas Rebele, Fabian Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. YAGO: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 9982 LNCS. Springer Verlag, 177–185. https://doi.org/10.1007/978-3-319-46547-0{_}19

[19] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference* (8 2019), 3982–3992. http://arxiv.org/abs/1908.10084

[20] Victoria L. Rubin. 2010. On deception and deception detection: Content analysis of computer-mediated stated beliefs. In *Proceedings of the ASIST Annual Meeting*, Vol. 47. https://doi.org/10.1002/meet.14504701124

[21] Ando Saabas. 2014. Interpreting random forests | Diving into data. http://blog.datadive.net/interpreting-random-forests/

[22] Robert Tibshirani. 1996. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58, 1 (1996). https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[23] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (3 2018), 1146–1151. https://doi.org/10.1126/science.aap9559

[24] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: A free collaborative knowledgebase. *Commun. ACM* 57, 10 (9 2014), 78–85. https://doi.org/10.1145/

2629489
[25] Chenguang Wang, Xiao Liu, and Dawn Song. 2020. Language Models are Open Knowledge Graphs. (10 2020). http://arxiv.org/abs/2010.11967
[26] Lucy Lu Wang, Kyle Lo, Yoganand Chandrasekhar, Russell Reas, Jiangjiang Yang, Darrin Eide, Kathryn Funk, Rodney Kinney, Ziyang Liu, William Merrill, Paul Mooney, Dewey Murdick, Devvret Rishi, Jerry Sheehan, Zhihong Shen, Brandon Stilson, Alex D. Wade, Kuansan Wang, Chris Wilhelm, Boya Xie, Douglas Raymond, Daniel S. Weld, Oren Etzioni, and Sebastian Kohlmeier. 2020. CORD-19: The Covid-19 Open Research Dataset. arXiv:2004.10706 [cs.DL]
[27] Kai-Chou Yang, Timothy Niven, and Hung-Yu Kao. 2019. Fake News Detection as Natural Language Inference. *arXiv* (7 2019). http://arxiv.org/abs/1907.07347
[28] Tong Zhang, Di Wang, Huanhuan Chen, Zhiwei Zeng, Wei Guo, Chunyan Miao, and Lizhen Cui. 2020. BDANN: BERT-Based Domain Adaptation Neural Network for Multi-Modal Fake News Detection. In *Proceedings of the International Joint Conference on Neural Networks*. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IJCNN48605.2020.9206973
[29] Xinyi Zhou and Reza Zafarani. 2018. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *Comput. Surveys* 53, 5 (12 2018). https://doi.org/10.1145/3395046