iJRASET
International Journal For Research in Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 8 **Issue:** XII **Month of publication:** December 2020

**DOI:** https://doi.org/10.22214/ijraset.2020.32545

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429*
*Volume 8 Issue XII Dec 2020- Available at www.ijraset.com*

# Fake News Detection using Knowledge Graphs

Pramita Kastha
*Dept. of Electronics and Communication Engineering, Indian Institute of Technology (IIT) Kharagpur*

***Abstract:** This paper looks into the problem of detecting fake news through efficient methods. There are already many highly researched works on this subject, but most of them focus on detecting fake news on social media and do not use the comprehensive content of the news for decision making. In this paper, we discuss different approaches to detect fake news, i.e content based and style based, and why the former one is better. Apart from this, we also try to highlight where fake news detection comes into use and also address the different challenges faced in building the model.*
***Keywords:** Knowledge Graph, Fake News, TransE, OpenIE, StanfordNER, SimeE-ER, Knowledge Graph Embedding, Avg Bias Classification, Max Bias Classification.*

## I. INTRODUCTION

With the advancement and increase of popularity of the Internet, it has become more and more easier for the common man to get news from the Internet rather than conventional media. However, it also kindles to the spread of a lot of fake news due to lack of effective supervision. Fake News can be defined as news articles that are purposefully written and false and could possibly misguide readers. With fake news costing less, being easily accessible and spreading like wildfire, it can easily create confusion among the public, disturb the social harmony, hamper the social media credibility, and creates lack of confidence towards one another; one of the primary examples being the 2016 United States presidential elections. Therefore, it is imperative to find out efficient methods for detecting fake news. There are broadly two ways to detect false content, viz. content-based and style-based. These are discussed in the next section.

## II. FAKE NEWS DETECTION - CONTENT-BASED VS STYLE-BASED

Style-based methods focus on capturing the writing style of the news content and try to classify articles on the basis of it. Most of the works [1],[2],[3],[4] were based on the fact that there were a number of similarities between the fake news and spam emails like both having a huge number of grammatical mistakes, trying to manipulate the reader's opinions and using a somewhat similar word set. Although these approaches can be very effectives at times, but they fail to explain what exactly is fake in the news content.
While on the other hand, content-based methods for detecting fake news, also known as fact checking [5], takes in use of the content of fake news rather than the style. It's the most straightforward way and aims on checking the truth value of the claims in the news content. The contemporary approaches majorly focus on finding a path in an existing knowledge graph [6],[7] for a certain triple. We will focus more on content-based approaches in this paper, since it looks more promising in the current status quo.

## III.DEVELOPING THE MODERN SOLUTION APPROACH

### *A. Limitations of Previous Existing Models*

The current proposed model focuses on removing the follow limitations that existed in the previous approaches and lead to research questions:
Can incomplete and vague knowledge graphs be used for detecting the ever-increasing fake news?
What do we do in case we do not have a knowledge graph and only articles?
How can a Knowledge Graph Embedding be used for detecting fake news once we have a complete and precise knowledge graph?
Is a single model based on one related knowledge graph good enough to put faith on?
These limitations were taken care of by the newest model proposed in 2018 by University of Aberdeen, U.K and Wuhan University, China which is discussed in the next section.

### *B. The Current Model*

To detect whether the news is fake or not, and to answer the above asked questions, the current model proposes a solution that uses a tool to create knowledge graphs, a binary TransE model [8] and finally some hybrid methods. Firstly, it focuses on generating background knowledge by creating three different knowledge graphs. The first one is based on fake news articles, the second on open-sourced knowledge graphs like DBpedia, Wikidata etc. Here it uses DBpedia. The last one is based on news from trusted sources. This addresses the first two questions.

Further, a binary TransE model is used to create an embedding in a low-dimensional vector space and find out the truth value of the news article. Before finalising the model with binary TransE model, it also checks with a single TransE model thus answering the next two questions. Finally, some hybrid methods are also carried out for detection performance improvement.

In order to build the knowledge graph, it extracts triples from news articles published by trustable news sources. This is carried out in many steps. First, the news article is summarized, since using the full article can cause unwanted redundancies and increase total runtime. After this, OpenIE and StanfordNER are used to extract the triples for the three knowledge graphs respectively.

Once the three knowledge graphs are generated, a TransE model is used to train each one of them and compare their respective performances. It uses Avg. Bias Classification and Max Bias Classification to classify the news articles as true or false. Using a single TransE model is not enough, because there always remain some triples whose biases are large on both the true and false single models and there exists a high chance of them getting classified under false if only a true TransE model is used. Hence, another model which uses the fake news articles as a base is also trained, so that classification can be done by comparing the biases of the true as well as the false model, thus using the binary TransE (BTransE) model.

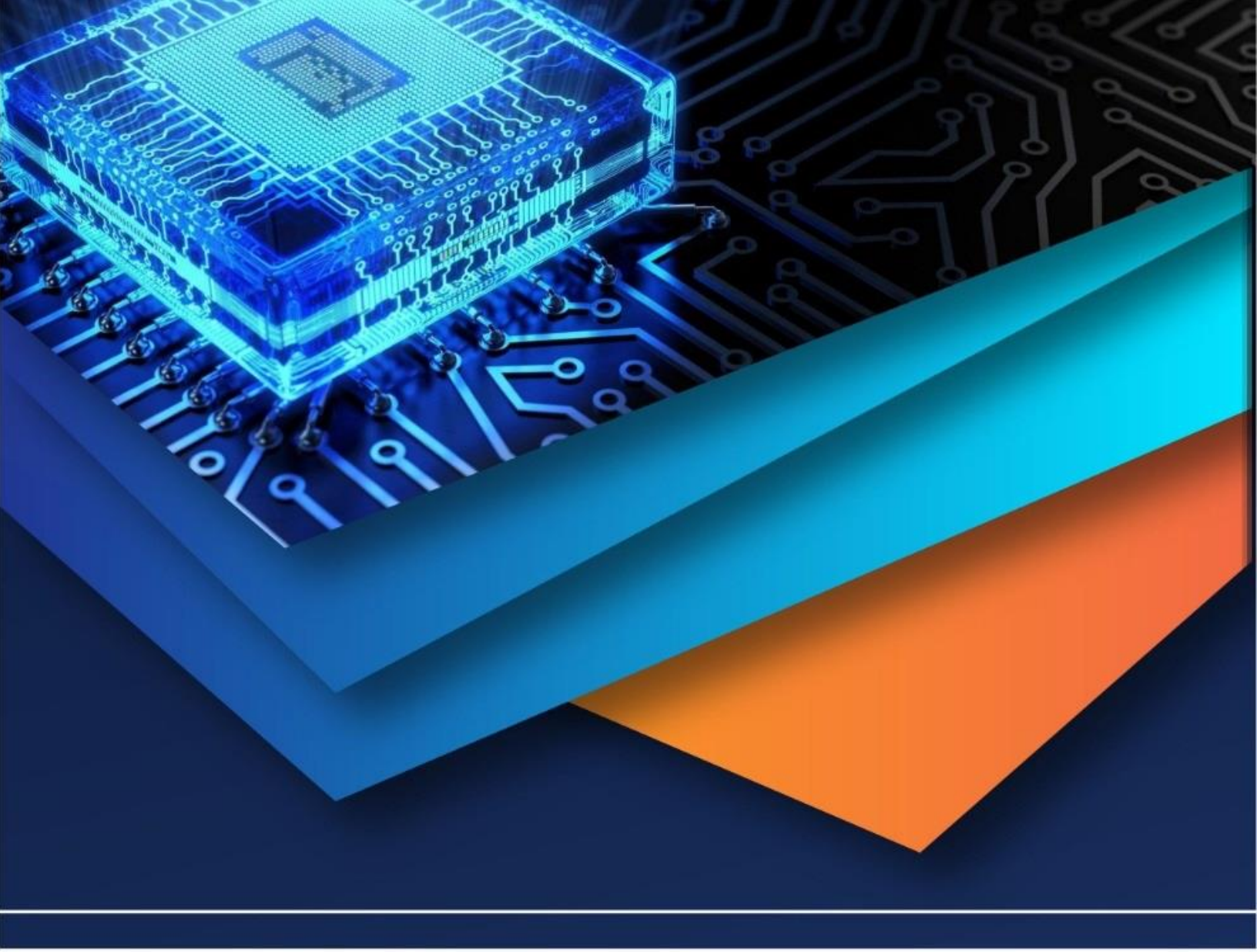
Finally, in order to improve detection performance, we use early fusion (a fusion strategy) to integrate the features first and then further use them to do classification.

## IV. CONCLUSIONS

In this paper we discussed the problem of ever-increasing fake news, why is it necessary to stop it, what are the different types of approaches to detect fake news, limitations to the pre-existing approaches and the modern solution approach, i.e fake news detection using incomplete and vague knowledge graphs, based the already existing TransE and the proposed binary TransE model and using hybrid approaches to increase their performance. This concludes that even incomplete and vague knowledge graphs can also be used in detecting fake news, which solves a big problem.

## REFERENCES

[1] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on, pages 900–903. IEEE, 2017.
[2] Shlok Gilda. Evaluating machine learning algorithms for fake news detection. In Research and Development (SCOReD), 2017 IEEE 15th Student Conference on, pages 110–115. IEEE, 2017.
[3] William Yang Wang. “liar, liar pants on fire”: A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648, 2017
[4] Ying Jiang, Yujie Liu, and Yalin Yang. Language tool-based university rumor detection on sina weibo. In Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on, pages 453–454. IEEE, 2017.
[5] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1):22–36, 2017.
[6] Jeff Z Pan, Guido Vetere, Jose Manuel Gomez-Perez, and Honghan Wu. Exploiting Linked Data and Knowledge Graphs in Large Organisations. Springer, 2017.
[7] Jeff Z Pan, Diego Calvanese, Thomas Eiter, Ian Horrocks, Michael Kifer, Fangzhen Lin, and Yuting Zhao. Reasoning Web: Logical Foundation of Knowledge Graph Construction and Query Answering: 12th International Summer School 2016, Aberdeen, UK, September 5-9, 2016, Tutorial Lectures, volume 9885. Springer, 2017
[8] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Advances in neural information processing systems, pages 2787–2795, 2013.
[9] Zhen Tan, Xiang Zhao, Yang Fang, Bin Ge, and Weidong Xiao. Knowledge Graph Representation via Similarity-Based Embedding, Science and Technology on Information Systems Engineering Laboratory, National University of Defense Technology, Hunan, China, 15 July, 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY