Content Based Fake News Detection Using Knowledge Graphs

Jeff Z. Pan[1⋆], Siyana Pavlova[1], Chenxi Li[1,2], Ningxi Li[1,2], Yangmei Li[1,2], and Jinshuo Liu[2⋆]

[1]University of Aberdeen, UK
[2]Wuhan University, China

Abstract. This paper addresses the problem of fake news detection. There are many works already in this space; however, most of them are for social media and not using news content for the decision making. In this paper, we propose some novel approaches, including the B-TransE model, to detecting fake news based on news content using knowledge graphs. In our solutions, we need to address a few technical challenges. Firstly, computational-oriented fact checking is not comprehensive enough to cover all the relations needed for fake news detection. Secondly, it is challenging to validate the correctness of the extracted triples from news articles. Our approaches are evaluated with the Kaggle's 'Getting Real about Fake News' dataset and some true articles from main stream media. The evaluations show that some of our approaches have over 0.80 F1-scores.

1 Introduction

With the widespread popularization of the Internet, it becomes easier and more convenient for people to get news from the Internet than other traditional media. Unfortunately, open Internet fuels the spread of a great many fake news without effective supervision. *Fake news* are news articles that are intentionally and verifiably false, and could mislead readers [AG17a]. With characteristics of low cost, easy access, and rapid dissemination, fake news can easily mislead public opinion, also disturb the social order, damage the credibility of social media, infringe the interests of the parties and cause the crisis of confidence [VRA18,SCV+17]. We all know how it has occurred and exerted an influence in the past 2016 US presidential elections [AG17b]. Hence, it is important and valuable to develop methods for detecting fake news.

Most existing works on fake news detection are based on styles, focusing on capturing the writing style of news content as features to classify news articles [GM17,Gil17,Wan17,JLY17]. Although they can be effective, these approaches cannot explain what is fake in the target news article. On the other hand, content based fake news detection, which is also known as fact checking [SSW+17], is more promising, as the detection is based on content rather

⋆ Corresponding authors

than style. Existing content based approaches focus on path reachability trying to find a path in an existing knowledge graph [PVGPW17,PCE+17] for a given triple [LCSR+15,SFMC17,SW16]. However, there are a few limitations of the existing content-based approaches, which lead to the following research questions:

RQ1: Can we use incomplete and imprecise knowledge graphs for fake news detection? All computational knowledge-based approaches mainly focus on simple common relations between entities, such as "*country*", "*child*", "*employerOf*". And the knowledge graphs they use are too incomplete and imprecise to cover the complex relations that appeared in fake news articles. For example, the triple *(Anthony Weiner, cooperate with, FBI)* extracted from a news article has the entities of "*Anthony Weiner*" and "*FBI*", and the relation of "*cooperate with*". The entities are easily found in open knowledge but the relation is not. In this paper, our idea is to make use of knowledge graph embedding for computing semantic similarities, so as to accommodate incomplete and imprecise knowledge graphs. As far as we know, this is the first work on this direction. We use a basic knowledge graph embedding model, namely TransE [BUGD+13] , to test the potential of knowledge graph embedding methods in content based fake news detection.

RQ2: What happens if we do not have a knowledge graph in the first place, but only have articles? For a fake news topic, it is likely that at the beginning we do not have the knowledge graph to rely on for fact-checking. Thus we can not only utilize the knowledge graphs based on news articles bases, we can also attempt to utilize related sub-graphs from open knowledge graphs. We could extract the sub-graph centered on the background topic of news articles from the open knowledge graph. Hence, we construct an external knowledge graph for these news articles based on facts related to the background topic in DBpedia dataset[1]. We can also construct two opposite but related knowledge graphs: one is based on only fake news articles provided by Kaggle[2]; and the other one is based on only true news articles from reliable news agencies[3,4,5]. "Related" means that they have the same background topic, and "opposite" means that they are based on fake and true news article base respectively. Fake news articles are available in online fake news web sites, such as 'the Onion', which often provide different categories of fake news articles.

RQ3: How can we use Knowledge Graph Embedding for fake news detection to reasonably utilize the complete and precise knowledge graph related to the fake news topic once we have it? We propose the approach of utilizing TransE [BUGD+13] to train a single model on each knowledge graph that we have constructed to compare their performance. The single model trained on the knowledge graph from fake news articles is regarded as a negative model,

---

1 http://wiki.dbpedia.org/
2 https://www.kaggle.com/mrisdal/fake-news
3 https://www.bbc.co.uk/
4 https://news.sky.com/
5 https://www.independent.co.uk/

and the other two are positive models. This approach takes the news articles as input, performs triple extraction on them and then detects whether they are fake or true on the single model.

RQ4: Is the single model based on one related knowledge graph enough? We propose a further approach to generate a binary TransE model (B-TransE) which combines a negative single model with a positive single model, as a single model may not be enough if it has been trained only by the knowledge graph from the fake article base or only by one of the true knowledge graphs. This approach is proposed to investigate whether the binary model performs better than the individual one.

Besides that, in order to improve the performance, we also propose a hybrid approach using a fusion strategy to combine the feature vectors produced by the models above.

Experiments on test datasets show that complete and precise knowledge graphs can play an important role in fake news detection, the binary model performs better than the individual one and the hybrid approach does improve the performance of fake news detection. Our major contributions of this paper are summarized as follows:

- To the best of our knowledge, we are the first to propose the approach of fake news detection based on the article content by constructing complete and precise enough knowledge graph to cover the fake news articles' topic.
- To the best of our knowledge, we are the first to propose the approach of fake news detection using positive and negative knowledge graph embedding models, and it performs well once the complete and precise knowledge graphs have been obtained.
- We find that a binary TransE model which combines positive and negative single models performs better than the individual one.
- We propose a hybrid approach to adopt a fusion strategy combined with trained TransE models to improve the performance of fake news detection.

2 Related Works

2.1 Fake News Detection

An effective approach is of prime importance for the success of fake news detection that has been a big challenge in recent years. Generally, those approaches can be categorized as knowledge-based and style-based.

Knowledge-based. The most straightforward way to detect fake news is to check the truthfulness of the statements claimed in news content. Knowledge-based approaches are also known as fact checking. The expert-oriented approaches, such as Snopes[6], mainly rely on human experts working in specific fields to help decision making. The crowdsourcing-oriented approaches, such as

[6] http://www.snopes.com/

Fiskkit[7] where normal people can annotate the accuracy of news content, utilize the wisdom of crowd to help check the accuracy of the news articles. The computational-oriented approaches can automatically check whether the given claims have reachable paths or could be inferred in existing knowledge graphs. Ciampaglia *et al.* [LCSR+15] take fact-checking as a problem of finding shortest paths between concepts in a knowledge graph; they propose a metric to assess the truth of a statement by analyzing path lengths between the concepts in question. Shiralkar *et al.* [SFMC17] propose a novel method called "Knowledge Stream(KS)" and a fact-checking algorithm called Relational Knowledge Linker that verifies a claim based on the single shortest, semantically related path in KG. Shi *et al.* [SW16] view fake news detection as a link prediction task, and present a discriminative path-based method that incorporates connectivity, type information and predicate interactions.

Style-based. Style-based approaches attempt to capture the writing style of news content. Mykhailo Granik *et al.* [GM17] find that there are some similarity between fake news and spam email, such as they often have a lot of grammatical mistakes, try to affect reader's opinion on some topics in manipulative way and use similar limited set of words. So they apply a simple approach for fake news detection using naive Bayes classifier due to those similarity. Shlok Gilda [Gil17] applies term frequency-inverse document frequency (TF-IDF) of bi-grams and probabilistic context free grammar (PCFG) detection and test the dataset on multiple classification algorithms. William Yang Wange [Wan17] investigates automatic fake news detection based on surface-level linguistic patterns and design a novel, hybrid convolutional neural network to integrate speaker related metadata with text. Jiang *et al.* [JLY17] find that some key words tend to appear frequently in the micro-blog rumor. They analyze the text syntactical structure features and presents a simple way of rumor detection based on LanguageTool.

2.2 Knowledge Graph Embedding

Antoine Bordes *et al.* [BUGD+13] propose a method, named TransE, which models relationships by interpreting them as translations operating on the low-dimensional embeddings of the entities. TransE is very efficient while achieving state-of-the-art predictive performance, but it does not perform well in interpret such properties as reflexive, one-to-many, many-to-one, and many-to-many. So, Zhen Wang *et al.* [WZFC14] propose TransH which models a relation as a hyperplane together with a translation operation on it. Yankai Lin *et al.* [LLS+15] propose TransR to build entity and relation embeddings in separate entity space and relation spaces. TransR learns embeddings by first projecting entities from entity space to corresponding relation space and then building translations between projected entities. Guoliang Ji *et al.* [JHX+15] propose a model named TransD, which uses two vectors to represent a named symbol object (entity and relation), and the first one represents the meaning of a(n) entity (relation), the other one is used to construct mapping matrix dynamically.

[7] http://fiskkit.com

3 Basic Notions

In this section we introduce some basic notions related to content-based classification of news articles with external knowledge.

A *knowledge graph* KG describes entities and the relations between them. It can be formalised as $KG = \{E, R, S\}$, where E denotes the set of entities, R the set of relations and S the triple set. An *article base* AB is a set of news articles for each of which we have a title, a full content text and an annotation of *true* or *fake*. A knowledge graph may be a readily available for fact checking, such as DBpedia, or one needs to construct one from an article base.

The task of *fact checking* is to check if a target triple (h, r, t) is true based on a given knowledge graph. The task of *content based fake news detection* (or simply fake news detection), is to check if a target news article is true based on its title and content, as well as some related knowledge graph.

4 Our Approach

4.1 Framework Overview

To detect whether a news article is true or not, and to answer our research questions as outlined in section 1, we propose a solution which uses, a tool to produce knowledge graphs (KG), a single B-TransE model, a binary TransE model and finally hybrid approaches. Firstly, we generate background knowledge by producing three different KG. This part addresses RQ1 and RQ2. Then we use a B-TransE model to build entity and relation embedding in low-dimensional vector space and detect whether the news article is true or not. We test a single TransE model and a binary TransE model and thus answer RQ3 and RQ4. Finally, we use some hybrid approaches to improve detection performance.

For the task of background knowledge generation, we need to produce three KG: one is based on fake news article base; one is based on open KG, such as DBpedia, a crowd-sourced community effort to extract structured information from Wikipedia; one is based on true news article base from reliable news agencies.

The external KG extracted from open knowledge graph includes two parts: $KG_1 = \{E_1, R_1, S_1\}$ based on entities from fake article base and $KG_2 = \{E_2, R_2, S_2\}$ centered on the the topic of news articles. These are further described in section 5.2.

External KGs such as DBpedia are excellent for general knowledge facts, such as *(Barack Obama, birthPlace, Hawaii)*. However, they are incomplete and imprecise as such KGs do not contain enough relations to represent current events, as the latter are generated daily. An example of such a relation is *(Anthony Weiner, cooperate with, FBI)*, which is not contained in DBpedia. Despite this, in section 5, we show that an incomplete and imprecise external open KG can perform well on the task of fake news detection.

The entities and the relation from the example above, however, can easily be extracted from an article on the topic. In order to be able to assess news items as true or fake, we propose an approach which uses external knowledge generated from real world news articles. We propose using a set of true and a set of fake articles to generate two models: $\mathbf{M}$ and $\mathbf{M'}$ as described in Section 3. We summarize these articles, as using the full article text causes redundancies and increases runtime. We further explore the performance of our approach, using only external knowledge from article bases, in order to answer the question what happens when we do not have a KG in the first place, but only news articles. In Figure 1 we outline the methods used to generate a KG from an article base.

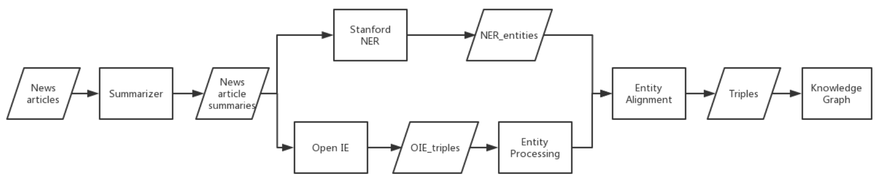

Fig. 1. Triple extraction from an article base

To construct KG from news articles, we start with a set of news articles and use OpenIE[8] to extract triples first. However, OpenIE does not perform well in triple extraction of news, so we propose some methods to improve the quality of the triples, including Stanford NER[9] and others which are further discussed in section 5.2. We then perform entity alignment and obtain the triples which constitute our article based KG.

Once we have generated our three external KG, we use TransE to train a single model on each of them and compare their performance. Since all the translation-based models aim to represent entities and relations in a vector space and there is no great difference between these models on our dataset, we choose the most basic model TransE. The single model is further described in section 4.2 and an outline of its usage can be seen in Figure 2. Our results, presented in section 5, show that the external open KG has the best performance.

Then, we explore what happens when we combine a negative single model and a positive single model. The binary TransE (B-TransE) model is further described in section 4.3 and an outline of its usage can be seen in Figure 3. In section 5 we then show that binary models perform better than single ones.

Finally, we use a hybrid approach using an early fusion strategy that combines the feature vectors produces by the models above in order to improve detection performance. Further details of this approach are in section 4.4.

[8] https://nlp.stanford.edu/software/openie.html
[9] https://nlp.stanford.edu/software/CRF-NER.html

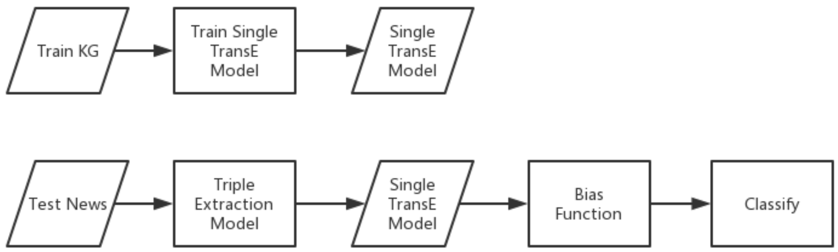

Fig. 2. Single TransE model

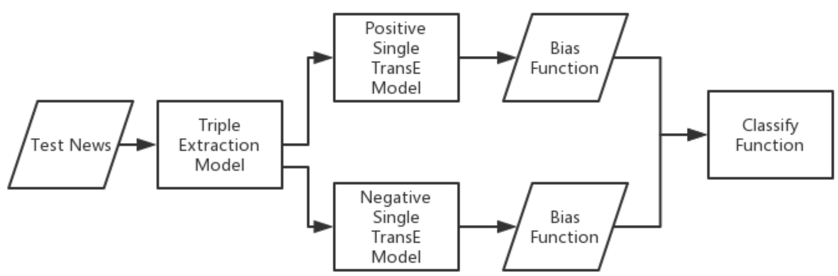

Fig. 3. Binary TransE model

4.2 Single TransE Model

To judge whether a given news article is true or fake through a knowledge graph, we extract triples from the news article and represent the triples in vector space, so that we can judge whether the news article is true or fake by the vectors. We use a Knowledge Graph to train a TransE model, which can represent triples as a vector, and we name our method Single TransE Model.

In the Single TransE Model, we define TransE model as $\mathbf{M}$, and a triple based on $\mathbf{M}$ as $(\mathbf{h}, \mathbf{t}, \mathbf{r})$. We denote the triples extracted from one news item as $\mathbf{TS}$, so each triple is defined as $triple_i = (\mathbf{h}_i, \mathbf{t}_i, \mathbf{r}_i)$, where i means the index of the triple in $\mathbf{TS}$. We represent one news item as $\mathbf{N} = \{\mathbf{TS}, \mathbf{M}\}$.

To classify one news item, we calculate the bias of each triple in $\mathbf{TS}$. The bias of $triple_i$ is defined as

$$f_b(triple_i) = ||\mathbf{h}_i + \mathbf{r}_i - \mathbf{t}_i||_2^2 \tag{1}$$

Then we use these biases to classify the news item through a classifier. There are two ways we use these biases to do classification.

Avg Bias Classification For the first one, we use average bias of a triple set to classify the news item through a classifier and name it Avg Bias Classification. The average bias of a triple set is defined as

$$f_{avgB}(TS) = \frac{\sum_{i=1}^{n} f_b(triple_i)}{|\mathbf{TS}|} \tag{2}$$

Where the $|\mathbf{TS}|$ refers to the size of the triple set.

Max Bias Classification The second one, the Max Bias Classification uses the max bias of a triple set to judge whether a news item is true or fake. The max bias of a triple set is defined as

$$f_{maxB}(TS) = f_b(triple_{max}) \tag{3}$$

Where max refers to the index of the triple whose bias is the maximum.

4.3 B-TransE Model

We believe using just a single TransE model to do classification is not enough, because there are some true triples whose biases are large on both the true single model and the fake single model, so that these news items would be incorrectly classified as fake news if we use just a single true TransE model.

To solve this problem, we train two models, one model is trained based on the triples extracted from fake news and another is trained based on the triples extracted from true news, so that we can do classification by comparing the biases of the true model and the biases on the fake model. We name it B-TransE model:

- the model based on true news is defined as $\mathbf{M}$, and a triple based on $\mathbf{M}$ is defined as $(\mathbf{h}, \mathbf{t}, \mathbf{r})$

– the model based on fake news is defined as $\mathbf{M}'$, and a triple based on $\mathbf{M}'$ is defined as $(\mathbf{h}', \mathbf{t}', \mathbf{r}')$

In the B-TransE Model, we represent one news item as $\mathbf{N} = \{\mathbf{TS}, \mathbf{TS}', \mathbf{M}, \mathbf{M}'\}$, $\mathbf{TS}$ refers to triple set extracted from the news based on $\mathbf{M}$ and each triple is defined as $triple_i = (\mathbf{h}_i, \mathbf{t}_i, \mathbf{r}_i)$, and $\mathbf{TS}'$ refers to triple set based on $\mathbf{M}'$ and each triple is $triple'_i = (\mathbf{h}'_i, \mathbf{t}'_i, \mathbf{r}'_i)$, where i refers to the index of the triple in each triple set.

We define the bias of $triple_i$ and $triple'_i$ as

$$f_b(triple_i) = ||\mathbf{h}_i + \mathbf{r}_i - \mathbf{t}_i||_2^2 \tag{4}$$

$$f_b(triple'_i) = ||\mathbf{h}'_i + \mathbf{r}'_i - \mathbf{t}'_i||_2^2 \tag{5}$$

To judge whether a news item is true or fake, we propose two classify functions and do some experiments to verify the efficiency of each method. **Max Bias Classify** The first way, we use max bias on true single model and max bias on fake single model to do classification. And the Max Bias Classify function is defined as

$$f_{mc}(N) = 0, iff_b(triple_{max}) < f_b(triple'_{max}) \tag{6}$$

$$f_{mc}(N) = 1, otherwise \tag{7}$$

where $f_{mc}(N) = 0$ means the news item is true, and $f_{mc}(N) = 1$ means it is fake. **Avg Bias Classify** The another way, we use average bias on true single model and average bias on fake single model to do classification. And the Avg Bias Classify function is defined as

$$f_{ac}(N) = 0, iff_{avgB}(TS) < f_{avgB}(TS') \tag{8}$$

$$f_{ac}(N) = 1, otherwise \tag{9}$$

where $f_{ac}(N) = 0$ means the news item is true, and $f_{ac}(N) = 1$ means it is fake.

4.4 Hybrid Approaches

To improve the detection performance, we need a fusion strategy to combine the feature vectors from different models. The fusion strategy we use is known as early (feature-level) fusion, which means integrating different features first and using those integrated-features do classification.

In this part, we use the bias vector of the triple, whose bias is the maximum, rather than bias to do classifiction. The bias vector is defined as

$$\mathbf{v}_i = \mathbf{h}_i + \mathbf{r}_i - \mathbf{t}_i \tag{10}$$

The max bias vector is defined as Vec_{max}. We use two different feature vectors:

1.max bias vectors from the model based on true news is defined as Vec_{max}, 2.max bias vectors from the model based on fake news is defined as Vec'_{max}. The integrated vector is defined as V, so that:

$$V = (Vec_{max}, Vec'_{max}) \tag{11}$$

which means we concatenate two different max bias vectors to get an integrated vector, and we use this vector to do classification.

5 Experiments and Analysis

5.1 Data

Fake and True News Article Bases We use two article bases for our experiments: one with fake news and one with news that we regard as true. We use Kaggles Getting Real about Fake News dataset[2], which contains news articles on the 2016 US Election, and we select 1,400 of this dataset as our Fake News Article Base (FAB). These articles have been manually labeled as *Bias*, *Conspiracy*, *Fake*, *Bull Shit*, which we regard as *fake*. Our True News Article Base (TAB) was produced by using the BBC News[3], Sky News[4] and The Independent websites[5] to scrape 1,400 news articles whose topic was US Election and were published between 1st January and 31st December 2016. These articles have not been manually labeled, however, for the purposes of our experiments, we regard them as *true*. The statistics of two article bases are shown in Table 1. We divide each article base into two parts, 1,000 are for training a model and 400 are for testing.

Article Base	Label	Source	Quantity
FAB	*fake*	Kaggles Getting Real about Fake News	1,400
TAB	*true*	BBC, Sky, Independent news	1,400

Table 1. Statistics of fake and true news article bases.

Knowledge Graphs We produce three knowledge graphs for our experiments: one named FKG from FAB, one named D4 (DBpedia 4-hop) from DBpedia, and one named NKG from TAB.

FKG. We produce FKG= $\{E_0, R_0, S_0\}$ using the training set of FAB. FKG has the following characteristics: $|E_0|$ =4K entities, $|S_0|$ =8K triples.

D4. To build our KG from DBpedia with 4 hops, we use SPARQL query endpoint interface[10] to interview DBpedia dataset online. There is a public SPARQL endpoint over the DBpedia dataset[11]. DB4 includes two parts, they are KG_1 and KG_2. $KG_1 = \{E_1, R_1, S_1\}$ based on entities from FAB. It has the following

[10] https://rdflib.github.io/sparqlwrapper/

[11] http://dbpedia.org/sparql

characteristics: $|E_1|$ =215K entities, $|S_1|$ =760K triples. $KG_2 = \{E_2, R_2, S_2\}$ centered on 2016 US election. We take the entity "United States presidential election 2016" as h_0, extract all triples within four hops. It has the following characteristics: $|E_2|$ =132K entities, $|S_2|$ =312K triples. The reason we extract 4-hop subgraph is that one more hop produces lots of repetitive triples, and most appear in the 4-hop one. We just need to make sure that we get triples related to the topic even some are not related tightly, which also makes the KG construction easier and general.

NKG. We produce NKG= $\{E_3, R_3, S_3\}$ using the training set of TAB. NKG has the following characteristics: $|E_3|$ =15K entities, $|S_3|$ =19k triples.

5.2 Experiment Setup

Article Summarization. We use the titles and the first two sentences of each article to produce the summaries.

Knowledge Extraction. We use an extraction model to extract train triples from 1k fake news, which is used to train FML, and extract train triples from 1k true news, which is used to train FML. Simultaneously, we use an extraction model to extract test triple sets from 400 fake news and 400 true news, which means translating each news item into a triple set with a fake or true label. We use OpenIE to perform triple extraction. However, OpenIE does not perform very well on triple extraction from news articles so we use four methods to improve the quality of the entities and relations in the triples extracted:

- We disambiguate pronouns so that a text such as “The man woke up. He took a shower.” would be transformed to “The man woke up. The man took a shower”. We use Neuralcoref to do this.
- We use NLTKs WordNetLemmatizer to transform any verbs in the triples to their present tense.
- We shorten the length of the entities, which is extracted though OpenIE and is named OpenIEEntity. We find out the word which is real entity in the entity extracted though OpenIE and remove other words. Such as "western mainstream media like John Kerry" is shortened to "western mainstream media".
- We use Stanford NER to extract entities from news, which is named NER-Entity. Then align the OpenIEEntities to NEREntities.

To produce the two parts of the external KG from an open knowledge graph, as outlined in section 4.1, we use the following steps:

1. $KG_1 = \{E_1, R_1, S_1\}$ based on entities from a fake article base. Firstly, to obtain the set of entities E_1 from triples in fake article base. And then, to extract all triples S_1 from the open knowledge graph with these entities as subjects and objects respectively.

2. $KG_2 = \{E_2, R_2, S_2\}$ centered on the the topic of news articles. This sub-KG reflects true statements about the news topic in the real world. We take the entity h_0 that is the most related to the topic as the center, and extract all triples S_2 within a certain number of hops. As shown in Figure 4, it is a simplified three-hop sub-graph example. Supposing the node "0" to be h_0, firstly, to extract all triples denoted as T_1 that has the formula as (h_0, r, t). Secondly, to extract all triples denoted as T_2 that has the formula as (h_1, r, t), where h_1 refers to an entity in T_1, also one of the nodes "1" in the figure. And the rest can be done by analogy.

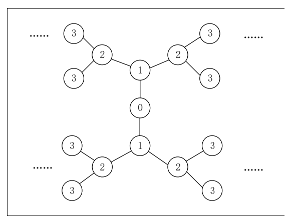

Fig. 4. A simplified three-hop sub-graph example

Model Generation. We generate three single trained models based on TransE for our experiments: the first model FML using the negative knowledge graph FKG; the second model TML-D4 using the positive knowledge graph D4; the third model TML-NKG using the positive knowledge graph NKG.

FML. The TransE model gets the input of S_0 and automatically produces the trained model FML.

TML-D4. The TransE model gets the input of $S_1 + S_2$ and automatically produces the trained model TML-D4.

TML-NKG. The TransE algorithm gets the input of S_3 and automatically produces the trained model TML-D4.

5.3 Fake News Detection

Using Single Models The results of the single TransE model with different bias function are shown in Table 2. From Table 2, we can conclude that: (1)TML-D4 model performs well for the fake news detection task. It means using incomplete knowledge graph is effective for fake news detection task. (2)FML and TML-NKG model also perform well. So, using imprecise knowledge graph is also effective for fake news detection. And if we dont have knowledge graph in the first place, but only have articles, contracting a knowledge graph from articles is a effective method. (3)Max Bias significantly outperforms than Avg Bias in terms of F Score. Maybe there are a few true triples in the triple set of one true news, so that the average bias of the triple set becomes smaller. Not all the triples extracted from one fake news is false, so max bias is more useful in fake news detection task. (4)What' more, TML-D4 performs a little better than TML-NKG and FML. The results may correlate with the training data of the TransE model: There are 1K training news of TML-NKG and FML, but there are 132K entities and 312K triples of the training data set of TML-D4.

Models	Bias Function	Precision	Recall	F1 score
FML	max bias	0.75	0.78	0.77
	avg bias	**0.80**	0.65	0.72
TML-D4	max bias	0.73	**0.86**	**0.79**
	avg bias	0.77	0.68	0.72
TML-NKG	max bias	0.69	**0.86**	0.77
	avg bias	0.79	0.71	0.75

Table 2. Performance of Single TransE Model.

Using B-TransE Model The results of B-TransE Model with different bias function are shown in Table 3. From the Table 3, we observe that:B-TransE Model is better than Single TransE Model. So, the approach based on one related knowledge graph is not enough, and combining related knowledge graph with external knowledge graphs is better than one for fake news detection.

Models	Bias Function	Precision	Recall	F1 score
FML + TML-D4	max bias	**0.85**	**0.80**	**0.83**
	avg bias	0.80	0.78	0.79
FML + TML-NKG	max bias	0.75	0.79	0.77
	avg bias	0.81	0.72	0.76

Table 3. Performance of Different Models.

Hybrid Approaches In this section, we do experiments on the test sets using the hybrid approach described in Section 4.4. Experimental results of combining different models are shown in Table 4. We use vectors from a single TransE model and integrated vectors from a B-TransE Model. The classification we use is SVM [Joa98,SS02], and we choose 'poly', 'linear' and 'rbf' as kernel functions. From Table 4, we can draw a conclusion that: the hybrid approach performs well for the fake news detection task, and it can somehow improve the detection performance of using one single model.

Knowledge Stream Finally, We test Knowledge Stream approach [SFMC17] on the 400 true articles and 400 fake articles required that a file exists for each article which contains all of the triples extracted from the given article and with

Approaches	Kernel	Precision	Recall	Accuracy
FML	poly	0.22	**0.90**	0.63
TML-D4	poly	0.82	0.87	0.85
FML + TML-D4	poly	**0.83**	0.88	**0.89**
FML	linear	0.61	0.91	0.75
TML-D4	linear	0.79	0.88	0.86
FML + TML-D4	linear	**0.81**	**0.92**	**0.87**
FML	rbf	0.74	**0.79**	0.81
TML-D4	rbf	0.94	0.77	0.80
FML + TML-D4	rbf	**0.95**	0.74	**0.81**

Table 4. Performance of Different Models.

IDs for each entity and relation accordingly. Once these files existed, they were run in Knowledge Stream to produce scores for each triple in each file. Table 5 shows the results of the comparison of the performance of the TransE FML and that of Knowledge Stream. From the table we observe that while Knowledge Stream has a very high recall value, TransE outperforms it significantly. Therefore, we conclude that: TransE is better than Knowledge Stream on the task of fake news detection when the background knowledge graph is constructed from real news articles.

Method	Function	Precision	Recall	F1 score
Knowledge Stream	max	0.50	0.99	0.66
	avg	0.47	**1.0**	0.64
TransE FML	max bias	0.75	0.78	**0.77**
	avg bias	**0.80**	0.65	0.72

Table 5. Performance of Different Models.

6 Conclusion and Future Work

In this paper, we tackle the problem of content based fake news detection. We have proposed some novel approaches of fake news detection based on incomplete and imprecise knowledge graphs, based on the existing TransE model and our B-TransE model. Our findings suggest that even incomplete and imprecise knowledge graph can help detect fake news.

As for future work, we will explore the following directions: (1) To combine our content based approaches with style-based approaches. (2) To provide explanations for the results fake news detection, even with incomplete and imprecise knowledge graphs. (3) To explore the use of the schema of knowledge graphs as well as approximate reasoning [PRZ16] and uncertain reasoning [PTRT12,SFP+13,JGC15] in fake news detection.

Acknowledgements. The work is supported by the Aberdeen-Wuhan Joint Research Institute, the EU K-Drive project (286348) and the China NDFC 61672393 project.

References

AG17a. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–236, 2017.

AG17b. Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36, 2017.

BUGD+13. 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.

Gil17. 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.

GM17. 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.

JGC15. Achille Fokoue Katia Sycara Yuqing Tang Jhonatan Garcia, Jeff Z. Pan and Federico Cerutti. Handling uncertainty: An extension of DL-Lite with Subjective Logic. In *Proc. of 28th International Workshop on Description Logics (DL 2015).*, 2015.

JHX+15. Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, volume 1, pages 687–696, 2015.

JLY17. Ying Jiang, Yujie Liu, and Yalin Yang. Languagetool based university rumor detection on sina weibo. In *Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on*, pages 453–454. IEEE, 2017.

Joa98. Thorsten Joachims. Making large-scale svm learning practical. Technical report, Technical report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund, 1998.

LCSR+15. Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. Computational fact checking from knowledge networks. 10, 01 2015.

LLS+15. Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, volume 15, pages 2181–2187, 2015.

PCE+17. 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.

PRZ16. Jeff Z Pan, Yuan Ren, and Yuting Zhao. Tractable approximate deduction for OWL. *Artificial Intelligence*, 235:95–155, 2016.

PTRT12. Jeff Z. Pan, Edward Thomas, Yuan Ren, and Stuart Taylor. Tractable Fuzzy and Crisp Reasoning in Ontology Applications. In *IEEE Computational Intelligence Magazine*, 2012.

PVGPW17. Jeff Z Pan, Guido Vetere, Jose Manuel Gomez-Perez, and Honghan Wu. *Exploiting Linked Data and Knowledge Graphs in Large Organisations.* Springer, 2017.

SCV+17. Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Alessandro Flammini, and Filippo Menczer. The spread of fake news by social bots. *arXiv preprint arXiv:1707.07592*, 2017.

SFMC17. Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. Finding streams in knowledge graphs to support fact checking. *arXiv preprint arXiv:1708.07239*, 2017.

SFP+13. Murat Sensoy, Achille Fokoue, Jeff Z. Pan, Timothy Norman, Yuqing Tang, Nir Oren, and Katia Sycara. Reasoning about Uncertain Information and Conflict Resolution through Trust Revision. In *Proc. of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS2013).*, 2013.

SS02. Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond.* MIT press, 2002.

SSW+17. 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.

SW16. Baoxu Shi and Tim Weninger. Fact checking in heterogeneous information networks. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 101–102. International World Wide Web Conferences Steering Committee, 2016.

VRA18. Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

Wan17. William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

WZFC14. Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, pages 1112–1119, 2014.