Spy The Lie: Fraudulent Jobs Detection in Recruitment Domain using Knowledge Graphs

Nidhi Goyal[1], Niharika Sachdeva[2], and Ponnurangam Kumaraguru[3*]

[1] Indraprastha Institute of Information Technology, New Delhi, India
nidhig@iiitd.ac.in
[2] InfoEdge India Limited, Noida, India
niharika.sachdeva@infoedge.com
[3] International Institute of Information Technology, Hyderabad, India
pk.guru@iiit.ac.in

Abstract. Fraudulent jobs are an emerging threat over online recruitment platforms such as LinkedIn, Glassdoor. Fraudulent job postings affect the platform's trustworthiness and have a negative impact on user experience. Therefore, these platforms need to detect and remove these fraudulent jobs. Generally, fraudulent job postings contain untenable facts about domain-specific entities such as mismatch in skills, industries, offered compensation, etc. However, existing approaches focus on studying writing styles, linguistics, and context-based features, and ignore the relationships among domain-specific entities. To bridge this gap, we propose an approach based on the Knowledge Graph (KG) of domain-specific entities to detect fraudulent jobs. In this paper, we present a multi-tier novel end-to-end framework called FRaudulent Jobs Detection (FRJD) Engine, which considers a) fact validation module using KGs, b) contextual module using deep neural networks c) meta-data module to capture the semantics of job postings. We conduct our experiments using a fact validation dataset containing 4 million facts extracted from job postings. Extensive evaluation shows that FRJD yields a 0.96 F1-score on the curated dataset of 157,880 job postings. Finally, we provide insights on the performance of different fact-checking algorithms on recruitment domain datasets.

Keywords: Recruitment Domain · Fraudulent Jobs · Knowledge Graphs.

1 Introduction

Online recruitment platforms such as Glassdoor, Indeed.com, LinkedIn are of paramount importance for employers and candidates to connect, recruit, and find jobs. These platforms attract millions of job seekers per month. Unfortunately, candidates often come across fraudulent jobs offering more wages, flexible working hours, and appealing career growth opportunities. Federal Trade Commission (FTC) registered 101,917 fraud complaints[4] from job seekers over the

* Major part of this work was done while Ponnurangam Kumaraguru was a faculty at IIIT-Delhi.

[4] https://www.aarp.org/money/scams-fraud/info-2020/ftc-job-scams.html

period of 2014 to 2019. The proliferation of fraudulent jobs not only hamper candidate's experience [36] but also act as a repressing factor[5] in an enterprise's reputation. Therefore, it is desirable to detect and take off these fraudulent jobs. Fraudulent jobs[6] are dishonest, money seeking, intentionally and verifiably false that mislead job seekers.

Data Entry Clerks Position We have several openings available in this area earning $1000.00-$2500.00 per week. We are seeking only honest, self-motivated people with a desire to work in the home typing and data entry field, from the comfort of their own homes.The preferred applicants should be at least 18 years old with Internet access. No experience is needed. However the following skills are desirable: Basic computer and typing skills, ability to spell and print neatly, ability to follow directions. Earn as much as you can from the comfort of your home typing and doing data entry. You do NOT need any special skills to get started.	Data Entry Clerk Responsibilities include, but are not limited to: Review and process confidential and extremely time-sensitive applications. Identify objective data and enter (""key what you see"") at a high level of productivity and accuracy. Perform data entry task from a paper and/or document image. Utilize system functions to perform data look-up and validation. High volume sorting, analyzing, indexing, of insurance, legal and financial documents. Maintain high degree of quality control and validation of the completed work Identify, classify, and sort documents electronically.

Fig. 1. Examples of job postings a) fraudulent job on the left and b) legitimate at the right. These job postings are taken from publicly available dataset.

Existing approaches mainly focus on supervised machine learning (e.g., ANN, bagging ensemble methods, and random forests) based on handcrafted feature engineering to detect fraudulent jobs [1]. However, these methods are unable to perform and scale well for larger datasets. Thereafter, NLP researchers also proposed linguistic, string-based [38], writing styles [32], textual, and contextual features [22] of job postings. These methods ignore the factual information among domain-specific entities present in job postings, which are important to capture relationships [34]. Figure 1 shows that the [left] job is fraudulent, which mentions implausible facts such as {*'offering very high weekly salary-* $1000 − $2500*', 'No experience required', 'Earn as much as you can'*} for *Data Entry Clerks* position. In contrast, the [right] job is legitimate that covers genuine facts related to role and responsibilities such as {*'Review and process confidential time-sensitive applications', 'Identify, classify, and sort documents'*} for same position.

To address these issues, we construct a fact-validation dataset consisting of 4 million facts of job postings from a popular recruitment platform. We utilize the fact validation dataset and employ automatic fact-checking algorithms [5] to find missing facts and validating the triples present in job postings using the triple classification task. Towards this end, we propose a multi-tier novel unified framework to leverage the fact-checking module, the contextual module using deep neural network-based approaches, and a unique meta-data knowledge module to accomplish fraudulent job detection tasks. We demonstrate the efficacy of our

[5] https://hrdailyadvisor.blr.com/2015/01/19/what-is-recruitment-fraud-is-your-company-at-risk/

[6] https://www.consumer.ftc.gov/articles/0243-job-scams

proposed approach on an annotated (proprietary & public) dataset of 157,880 job postings and validate them on open-source datasets, thus demonstrating our solution's generalizability. We summarize the contributions as follows:

- We propose a multi-tier novel unified framework called FRJD, which employs a fact-checking module using knowledge graph representations, the contextual module using deep neural networks, and considers unique meta-data properties of job postings to accomplish fraudulent jobs detection task.
- We study the fact validation dataset that consists of 4 million facts in form of entities and relationships and utilize it for the triple classification task.
- Extensive experiments on real-world recruitment domain datasets demonstrate the promising performance of FRJD compared to state-of-the-art models.

The organization of the rest of the paper as follows. Section 2 reviews the related literature of fraudulent content detection in the recruitment domain as well as in general. Thereafter, In Section 3, we formulate our problem. Our proposed framework FRJD is described in Section 4. Section 5 demonstrates the experimental setup along with datasets, experimental settings, and comparison with different approaches. Section 6 demonstrates our evaluation results. Section 7 describes ablation study. We conclude this work and provide future work in Section 8.

2 Related Works

This section describes the related literature of fraud detection in domain-specific scenarios and in general.

Content-based approaches. Research explores the textual content using TF-IDF [7], stylometric [31], and RNN (recurrent neural networks) [6,21]. Some approaches [20] exploits the graph-based techniques for fake content detection while others [13,8] use contextual embedding models such as ELMO and BERT [33] to learn language-based features.

Fact checking using Knowledge Graphs. Existing research [30] suggests knowledge graph-based techniques for fact checking in the news domain. Knowledge graph representation methods [40,14] are used to predict the plausibility of the facts using external KGs (DBpedia [18], Freebase [4]).

Most of the fact-checking methods also rely on experts, such as journalists or scientists, to assess the content and the crowd's wisdom [17]. Another set of approaches finds streams in knowledge graphs to support fact checking [35]. Some works [15] leverage unstructured and structured sources [39] for automatic fact-checking.

Despite the popularity of knowledge graph-based approaches, these are still underexplored in the recruitment domain, and there is limited information available in external knowledge bases [18,4] for domain-specific scenarios. Additionally, expert-based methods are expensive as they need the hiring of experts, and also are limited in number and unable to treat all the content being produced.

Domain-specific Scenarios. Recent research [38,1] focuses on content-based approaches that use handcrafted features such as empirical ruleset (binary, categorical, string-based) and Bag-of-words to identify fraudulent jobs in the recruitment domain. Works [29,22] conducted the research using behavioral activity or binary features as context. Kim et al. [16] propose hierarchical clustering deep neural network to detect fraud in the work processes of job placement automatically. We compare the most relevant studies with our work in Table 1.
Our research uniquely inclined towards using a hybrid approach consisting of the knowledge graph, contextual, and meta-data features at the same time requiring no job seeker responses and providing different insights on fact-checking algorithms.

Table 1. Different kind of features used in related literature.

	Content	**Knowledge**	**Context**
Vidros et al. [38]	✔		
Mahbub et al. [22]			✔
Nindyati et al. [29]			✔
Alghadmi et al. [1]	✔		
This work (FRJD)	✔	✔	✔

3 Problem Formulation

Let $\mathcal{J}$={ $\mathcal{J}_1$, $\mathcal{J}_2$,........,$\mathcal{J}_N$} be the set of job postings and $\mathcal{Y} = \{y_1, y_2, ..., y_n\}$ be corresponding labels such that $y_i \in \{0, 1\}$. For every $\mathcal{J}_i$, we extracted a set of triples $\mathcal{T}^i$ where $\mathcal{T}^i$ ={t_1^i, t_2^i, t_3^i,. . . ...,t_k^i} and $k > 0$; using *OpenIE*. A triple t_j^i $\in \mathcal{T}^i$ is of the form (subject (s), predicate (p), object(o)) where (s, o) $\in \mathcal{E}$ and $p \in \mathcal{P}$. We further define $m^i \in \mathcal{M}$ and $c^i \in \mathcal{C}$ as meta features and contextual features extracted from $\mathcal{J}_i$ (See Section 4).
Given a job posting $\mathcal{J}_i$ and its corresponding extracted set of triples $\mathcal{T}^i$, contextual vector c^i, and meta vector representation m^i. Our objective is to learn function φ where φ: $\mathcal{F}$ ($\mathcal{KG}^A_{false}$ $(T)^i$, $\mathcal{KG}^A_{true}$ $(T)^i$, c^i, m^i) where $\mathcal{KG}^A_{true}$ $(\mathcal{T})^i$ is the scoring function, we learn from triple $t^i \in \mathcal{T}^i | y_i = 0$ of legitimate job postings and $\mathcal{KG}^A_{false}$ $(\mathcal{T})^i$ from triple $t^i \in \mathcal{T}^i | y_i = 1$ of fraudulent job postings. Here $\mathcal{KG}^A$ $\in$ { *TransE*, *TransR*, *TransH*, *TransD*, *DistMult*, *ComplEx*, *HolE*, *RotatE*} which are popular fact-checking algorithms from existing knowledge graph literature.

4 FRaudulent Jobs Detection Engine

This section describes our multi-tier novel framework- FRaudulent Job Detection Engine (FRJD) using knowledge graphs and deep neural networks. Figure 2 depicts the overall architecture for the detection of fraudulent job postings. This framework consists of three components a) *Fact-checking module*, b) *Contextual embedding generation*, c) *Meta-features generation*.

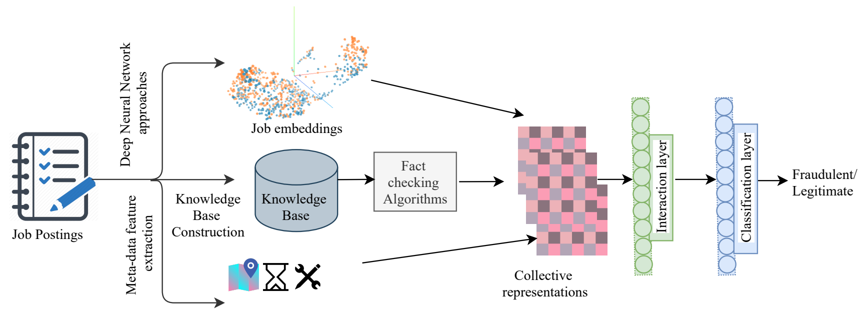

Fig. 2. An overview of our proposed framework- FRaudulent Jobs Detection Engine (FRJD).

Fact-checking module. This module identifies fraudulent job postings using fact-checking algorithms separately trained on legitimate and fraudulent jobs. We construct two domain-specific knowledge graphs ($\mathcal{KG}_{false}$, $\mathcal{KG}_{true}$) using triples extracted from legitimate ($\mathcal{T}^i|y_i = 0$) and fraudulent ($\mathcal{T}^i|y_i = 1$) jobs postings respectively. We pre-process the job postings and apply *OpenIE5* [24] to get triples[7] in the form of (s, p, o) as an output. We follow the similar methodology [11] to construct domain-specific knowledge graph. Thereafter, we obtain the low-dimensional knowledge graph representation for each entity and relation using various fact-checking algorithms [5,14,19,28,37]. Our objective is to obtain a separate score from both the knowledge graphs i.e. ($\mathcal{KG}_{true}$, $\mathcal{KG}_{false}$) for each triple t^i using scoring function $\mathcal{G}_p(s^i, o^i)$ introduced in [40]. We also provide the comparative analysis of these algorithms (see Table 3).

$$G_p(s^i, o^i) = \left\|(s - w_p^\top s w_p) + d_p - (o - w_p^\top o w_p\right\|_2^2 \tag{1}$$

Furthermore, we obtain b^i_{true} and b^i_{false} $\forall$ $\mathcal{T}^i$ of a job posting $\mathcal{J}_i$ from $\mathcal{KG}_{true}$ and $\mathcal{KG}_{false}$ respectively.

$$b^i \ = \sum_{\gamma=1}^{k} (G_p(s^i_\gamma, o^i_\gamma))/k \tag{2}$$

where $\gamma \in |T^i|$. Finally, we fuse both b^i to obtain a representation vector f^i.

$$f^i = b^i_{true} \bigoplus b^i_{false} \tag{3}$$

Contextual embedding generation. We employ a pre-trained deep neural network i.e., BERT [33] to generate contextual features for all job postings. Research works suggest that for real-time applications light weight model should

[7] Triples and facts are used interchangeably.

be preferred [25,9,10]. Hence, we use distilled version which requires fewer parameters, less space and time complexity while retaining 97% performance of BERT. Finally, we obtain s-dimensional (s=768) vector representation for each job posting $\mathcal{J}_i$ in the space $\in \mathcal{R}^{\omega}$ such that c^i

$$c^i = SBERT(\mathcal{J}_i) \tag{4}$$

Meta-features generation. In this module, we describe the meta-features of a job posting $\mathcal{J}_i$. We identify the domain-specific entities such as skills, companies, salary, locations, experience, and educational degree using state-of-the-art Named Entity Recognizer (NER) techniques [23] and rule-based heuristics. We consider the meta-information such as the number of skills mentioned, job length, educational degree, job location, telecommuting, and employment type. We perform normalization of these features to maintain the general distribution of the data. After extracting these features, we obtain a fused representation m^i such that m^i=[m^i_1, m^i_2, m^i_3, m^i_4,.....,m^i_k] where k is the number of meta-features extracted from job posting $\mathcal{J}_i$.
Finally, we concatenate together the factual, contextual, and meta representations to form $\mathcal{F}$ and pass them through the fully connected neural network layers.

$$\mathcal{F} \ = \{f^i \oplus c^i \oplus m^i\} \tag{5}$$

We use the Rectified Linear Units [26] as the non-linearity for faster training and drop out layers to avoid overfitting. We apply the sigmoid ($\sigma(.)$) layer and binary cross-entropy loss to classify the job postings into the legitimate and fraudulent. We use $ADAM$ as an optimizer [2] to handle sparse gradients.

5 Experimental Setup

In this section, we describe the dataset, experimental settings, and approaches that are used for comparison.

5.1 Dataset Description

Proprietary Dataset. We use the real-world job posting dataset from one of the largest online recruitment platforms in India. We curated a balanced dataset by sampling 70K Legitimate and 70K Fraudulent job postings from the legacy database annotated by domain-experts. We obtain 4 million triples from these job postings using OpenIE5 [24]. OpenIE results in noisy triples, therefore we improve the quality of these triples using Named Entity Recognizer (NER) to extract the important entities such as companies, institutes, skills, locations, qualifications, and designations to construct knowledge graph.
Public Dataset. The dataset of Employment Scam Aegean Dataset (EMSCAD)[8] contains 17,014 legitimate and 866 fraudulent jobs. For all experiments, we apply class balancing techniques by penalizing the class (legitimate) having more samples [38]. Table 2 reports the statistics of our dataset.

[8] http://emscad.samos.aegean.gr/

5.2 Experimental Settings

We use OpenKE [12] toolkit implementation to obtain knowledge graph representations. Given a set of triples (s, p, o) where entity pairs are (s, o) and p is the relation between them. We use these knowledge representations to map each entity to a v-dimensional vector and relation to a w-dimensional vector in the embedding space where v and w are hyperparameters. We use the best hyperparameter settings to train all fact checking algorithms such as *TransH*, *TransE*, *TransR*, *TransD*, *DistMult*, *ComplEx*, *HolE* [5,40,28,37]. To train FRJD, we use stratified sampling to split the train/test datasets into 70:30 and learning rate as 0.001.

Table 2. Statistics of Fraudulent and Legitimate jobs on proprietary dataset.

Statistic	Count
# of Fraudulent Jobs	70K
# of Legitimate Jobs	70K
Avg. words per Fraudulent job	70
Avg. words per Legitimate job	231
Avg. skills per Legitimate job	12
Avg. skills per Fraudulent job	9
# of entities	37.5K
# of relations	4.5K
# of triples	4M

Table 3. Results of triple prediction task on proprietary dataset.

	MRR		Hits @		
Model	**Raw**	**Filter**	**1**	**3**	**10**
TransH	**0.52**	**0.69**	**0.63**	**0.73**	**0.82**
TransD	0.50	0.67	0.62	0.69	0.80
TransR	0.20	0.60	0.55	0.64	0.73
TransE	0.51	0.60	0.56	0.62	0.68
HolE	0.22	0.48	0.34	0.49	0.71
ComplEx	0.29	0.34	0.25	0.35	0.52
DisMult	0.30	0.40	0.30	0.40	0.50
RotatE	0.28	0.41	0.39	0.40	0.43

5.3 Competing methods

We compare our method against several baselines for classification of fraudulent job postings.

- **Random Forest Classifier.** The approach [38] consists of ruleset-based binary classifier which consists of three categories: linguistic, contextual, and metadata. We report the results of model trained on the empirical ruleset against the complete imbalanced dataset of 17,880 job postings in public dataset (reported in published work). On other dataset, we use the same rulesets to report the results.
- **Logistic Regression.** Logistic regression is a statistical model that is popular for classification as well as regression tasks. We model the textual features using count-vectorizer and perform classification using LR.
- **Support Vector Machines.** SVM is a supervised machine-learning algorithm that is widely used for binary classification tasks. We use a spacy tokenizer to clean the text and utilize the textual features present in job postings to train SVM.

6 Evaluation Results

In this section, we evaluate and provide insights on various fact checking algorithms on our datasets. Table 3 reports the results on triple prediction using OpenKE [12]. We report the raw and filtered *MRR* (*Mean Reciprocal Rank*) and Hits@ (1, 3, 10) for all the models. Hits metrics are filtered (removal of the triples from test list which appeared while evaluation in the dataset). *TransH* and *TransD* achieve significant performances on these metrics, i.e., filtered *MRR* (0.69 and 0.67) and on hits@10 (0.82 and 0.80). Table 3 shows that *TransH* [40] outperforms *TransE* [5] on our dataset. According to [40] it better utilizes the one-to-many and many-to-many properties of the relation. Similarly, other algorithms such as *RotatE* are unable to perform well due to large number of many-to-many relations in our knowledge graph. We utilize *TransH* as fact checking algorithm in our fact-checking module in FRJD. We report the Precision (P), Recall (R) and F1-score (F1) of FRJD on proprietary and public datasets in Table 4. Our results show the efficacy of FRJD approach as compared to baseline methods. Therefore, it is noted that fact-checking contributes in our framework and help in identification of facts missed by content-based approaches. Our contextual module shows transformer models better capture the context in comparison to traditional approaches such as SVM and random forests which fail to perform well for fraudulent class. We further, demonstrate that incorporating the contextual, factual, and meta features together provide an average performance of 0.96.

Table 4. Performance of different models on proprietary and public datasets where M1, M2, M3 are contextual, factual, and meta features.

Dataset	**Approaches**	**Metrics**					
		Fraudulent			**Legitimate**		
		P	R	F1	P	R	F1
Proprietary Dataset	LR	0.71	0.64	0.67	0.94	0.31	0.47
	SVM	0.95	0.53	0.68	0.83	0.15	0.25
	RF	0.84	0.75	0.79	0.96	0.48	0.64
	FRJD (M1)	0.88	0.52	0.65	0.92	0.25	0.39
	FRJD (M2)	0.84	0.82	0.83	0.65	0.49	0.55
	FRJD (M3)	0.49	0.79	0.61	0.21	0.32	0.25
	FRJD (M1+M2+M3)	0.98	0.99	0.98	0.97	0.96	0.96
Public Dataset	LR	0.90	0.90	0.90	0.86	0.60	0.70
	SVM	0.57	0.83	0.67	0.98	0.97	0.97
	RF	0.28	0.75	0.41	0.98	0.90	0.94
	FRJD (M1)	0.40	0.20	0.27	0.60	0.80	0.69
	FRJD (M2)	0.94	0.90	0.92	0.90	0.73	0.80
	FRJD (M3)	0.91	0.61	0.73	0.81	0.22	0.34
	FRJD (M1+M2+M3)	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**

7 Ablation Study

In this section, we study the effect of each component in detail. Table 4 shows the performance of all of the components of our proposed framework, FRJD. We use contextual features (M1), factual (M2), metadata features (M3) separately as sub-models 'M1, 'M2', and 'M3' in Table 4. The ablation study reveals that component M1 captures context with a precision of 0.88 for fraudulent job postings. Furthermore, the component M2 gives a Precision of 0.84 for fraudulent job postings but yields Precision of 0.65 for legitimate job postings. The possible reason could be the similar facts present in both the knowledge graphs. Finally, we test the M3 component, which reveals that the meta-features such as number of skills, qualifications, job length are rudimentary. Additionally, we also verified the significant reasons to mark these job postings as fraudulent include seek money, use of legitimate employer names, advertise paid training-based courses, share multiple accounts for promotion, etc. We identified some facts where the model fails to distinguish between true and false facts. These facts are demanding visa fees, that were common to both legitimate and fraudulent job postings for some job titles.

8 Conclusion and Future Work

We proposed a multi-tier novel end-to-end framework called FRaudulent Jobs Detection (FRJD), which jointly considers a) fact validation module using knowledge graphs, b) contextual module using deep neural networks c) meta-data inclusion to capture the semantics of job postings. We conducted our study on a fact validation dataset containing 4 million facts extracted from job postings. We compared and performed an extensive evaluation of 157,880 job postings. Finally, we provided various insights on fact-checking algorithms for our dataset. We believe that our framework is generalizable to other datasets in the recruitment domain. We intend to study the time complexity of FRJD and compare it with other approaches. In future, we plan to apply and test our approach for hierarchy-based [3], neural network-based [27], and path-based [35] fact-checking algorithms. We wish to compare different algorithms for learning heterogeneous documents such as CVs to build an integrated framework and explore user features in future studies.

Acknowledgements.

We would like to acknowledge the support from SERB, InfoEdge India Limited, and FICCI. We are grateful to PreCog Research Group and Dr. Siddartha Asthana for critically reviewing the manuscript and stimulating discussions.

References

1. Alghamdi, B., Alharby, F.: An intelligent model for online recruitment fraud detection. Journal of Information Security **10**(3), 155–176 (2019)

2. Bello, I., Zoph, B., Vasudevan, V., Le, Q.V.: Neural optimizer search with reinforcement learning. arXiv preprint arXiv:1709.07417 (2017)
3. Bianchi, F., Rossiello, G., Costabello, L., Palmonari, M., Minervini, P.: Knowledge graph embeddings and explainable ai. arXiv preprint arXiv:2004.14843 (2020)
4. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data. AcM (2008)
5. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in neural information processing systems. pp. 2787–2795 (2013)
6. Bourgonje, P., Schneider, J.M., Rehm, G.: From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles. In: Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism. pp. 84–89 (2017)
7. Chen, L.C., Hsu, C.L., Lo, N.W., Yeh, K.H., Lin, P.H.: Fraud analysis and detection for real-time messaging communications on social networks. IEICE TRANSACTIONS on Information and Systems **100**(10), 2267–2274 (2017)
8. Cruz, J.C.B., Tan, J.A., Cheng, C.: Localization of fake news detection via multitask transfer learning. arXiv preprint arXiv:1910.09295 (2019)
9. Gai, K., Qiu, M.: Reinforcement learning-based content-centric services in mobile sensing. IEEE Network **32**(4), 34–39 (2018)
10. Gai, K., Qiu, M., Zhao, H., Sun, X.: Resource management in sustainable cyber-physical systems using heterogeneous cloud computing. IEEE Transactions on Sustainable Computing **3**(2), 60–72 (2017)
11. Goyal, N., Sachdeva, N., Choudhary, V., Kar, R., Kumaraguru, P., Rajput, N.: Con2kg-a large-scale domain-specific knowledge graph. In: Proceedings of the 30th ACM Conference on Hypertext and Social Media. p. 287–288. HT '19, Association for Computing Machinery, New York, NY, USA (2019)
12. Han, X., Cao, S., Lv, X., Lin, Y., Liu, Z., Sun, M., Li, J.: Openke: An open toolkit for knowledge embedding. In: Proceedings of the 2018 conference on empirical methods in natural language processing: system demonstrations. pp. 139–144 (2018)
13. Huang, G.K.W., Lee, J.C.: Hyperpartisan news and articles detection using bert and elmo. In: 2019 International Conference on Computer and Drone Applications (IConDA). pp. 29–32. IEEE (2019)
14. Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: 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). pp. 687–696 (2015)
15. Khandelwal, S., Kumar, D.: Computational fact validation from knowledge graph using structured and unstructured information. In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. p. 204–208. CoDS COMAD 2020, Association for Computing Machinery, New York, NY, USA (2020)
16. Kim, J., Kim, H.J., Kim, H.: Fraud detection for job placement using hierarchical clusters-based deep neural networks. Applied Intelligence **49**(8), 2842–2861 (2019)
17. Kim, J., Tabibian, B., Oh, A., Schölkopf, B., Gomez-Rodriguez, M.: Leveraging the crowd to detect and reduce the spread of fake news and misinformation. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. pp. 324–332 (2018)

18. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al.: Dbpedia–a large-scale, multilingual knowledge base extracted from wikipedia. Semantic Web (2015)
19. Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Twenty-ninth AAAI conference on artificial intelligence (2015)
20. Lu, Y.J., Li, C.T.: Gcan: Graph-aware co-attention networks for explainable fake news detection on social media. arXiv preprint arXiv:2004.11648 (2020)
21. Ma, J., Gao, W., Mitra, P., Kwon, S., Jansen, B.J., Wong, K.F., Cha, M.: Detecting rumors from microblogs with recurrent neural networks. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. p. 3818–3824. IJCAI'16, AAAI Press (2016)
22. Mahbub, S., Pardede, E.: Using contextual features for online recruitment fraud detection. In: International Conference on Information Systems Development 2018 (08 2018)
23. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The stanford corenlp natural language processing toolkit. In: Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations. pp. 55–60 (2014)
24. Mausam, M.: Open information extraction systems and downstream applications. In: Proceedings of the twenty-fifth international joint conference on artificial intelligence. pp. 4074–4077 (2016)
25. Mozafari, J., Fatemi, A., Moradi, P.: A method for answer selection using distilbert and important words. In: 2020 6th International Conference on Web Research (ICWR). pp. 72–76. IEEE (2020)
26. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: Icml (2010)
27. Nguyen, D.Q., Nguyen, D.Q., Nguyen, T.D., Phung, D.: A convolutional neural network-based model for knowledge base completion and its application to search personalization. Semantic Web **10**(5), 947–960 (2019)
28. Nickel, M., Rosasco, L., Poggio, T.A., et al.: Holographic embeddings of knowledge graphs. In: AAAI (2016)
29. Nindyati, O., Nugraha, I.G.B.B.: Detecting scam in online job vacancy using behavioral features extraction. In: 2019 International Conference on ICT for Smart Society (ICISS). vol. 7, pp. 1–4. IEEE (2019)
30. Pan, J.Z., Pavlova, S., Li, C., Li, N., Li, Y., Liu, J.: Content based fake news detection using knowledge graphs. In: International Semantic Web Conference. pp. 669–683. Springer (2018)
31. Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B.: A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638 (2017)
32. Reddy, M.N., Mamatha, T., Balaram, A.: Analysis of e-recruitment systems and detecting e-recruitment fraud. In: International Conference on Communications and Cyber Physical Engineering 2018. pp. 411–417. Springer (2018)
33. Sanh, V., Debut, L., Chaumond, J., Wolf, T.: Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. arXiv preprint arXiv:1910.01108 (2019)
34. Shaar, S., Babulkov, N., Da San Martino, G., Nakov, P.: That is a known lie: Detecting previously fact-checked claims. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 3607–3618. Association for Computational Linguistics, Online (Jul 2020)

35. Shiralkar, P., Flammini, A., Menczer, F., Ciampaglia, G.L.: Finding streams in knowledge graphs to support fact checking. In: 2017 IEEE International Conference on Data Mining (ICDM). pp. 859–864. IEEE (2017)
36. Thakkar, D., Kumar, N., Sambasivan, N.: Towards an ai-powered future that works for vocational workers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. pp. 1–13 (2020)
37. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: International Conference on Machine Learning (2016)
38. Vidros, S., Kolias, C., Kambourakis, G., Akoglu, L.: Automatic detection of online recruitment frauds: Characteristics, methods, and a public dataset. Future Internet **9**(1), 6 (2017)
39. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. Communications of the ACM **57**(10), 78–85 (2014)
40. Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: AAAI (2014)