Research Article

Transfer Learning on Knowledge Graph Construction: A Case Study of Investigating Gas-Mining Risk Report

Chong Wang,1,2,3 Jiren Wang,1,2 Like Wei,3 Xiangqian Wang,4 Chaoyu Yang4 and Lifang Zhang5

1College of Safety Science and Engineering, Liaoning Technical University, Fuxin 123000, China
2Key Laboratory of Mine Thermodynamic Disasters and Control of Ministry of Education, Huludao 125105, Liaoning, China
3Information Research Institute, Ministry of Emergency Management, Beijing 100029, China
4School of Economics and Management, Anhui University of Science and Technology, Huainan 232033, China
5China Jiliang University, Hangzhou, Zhejiang 310000, China

Correspondence should be addressed to Chaoyu Yang; yangchy@aust.edu.cn

Received 17 December 2021; Revised 23 January 2022; Accepted 4 February 2022; Published 14 March 2022

Academic Editor: Tabasam Rashid

Copyright © 2022 Chong Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This study addressed the problem of automated Knowledge Graph (KG) construction from unstructured documents, with the assistance of transfer learning. Despite a large amount of effort made to discover KG, how to explore unknown KGs from existing knowledge remains a challenge. In this paper, we firstly formulate the KG detection process as a transfer-learning problem, which consists of two main steps. At first, we pretrain a backbone model using the source domain. Due to sufficient samples from the source domain, this backbone model can be trained better. Second, we migrate this model (from the known domain) to the target domain by fine-tuning key parameters. The fine-tuning operation only requires less computation, which is very efficient. As such, the backbone model can be successfully transferred into the target domain, even with limit training samples. Experimental evaluations using one real-world dataset of gas-mining reports demonstrate the advantages of utilizing the proposed algorithm to construct KG using transferable information.

1. Introduction

Knowledge Graph (KG) is emerging as a critical tool for containing, inferring, and connecting knowledge. In recent years, we have witnessed an increasing interest in KG related research due to its numerous applications, including question answering [1], link prediction [2], and user profiling [3, 4], to name a few.

A typical KG consists of several tuples, while one tuple can be further comprised of a head \( h \), a relation \( r \), and a tail \( t \), respectively \((h, r, t)\). In practice, a tuple of \((h, r, t)\) represents the head entity, \( h \), a relation \( r \), and the tail entity \( t \). Certainly, one head \( h \) could be associated with more than one relation. Similarly, one tail \( t \) can be also linked to different heads with different relations. When encountering the relation, we normally specify or allow only one relation that exists given a pair of \( h \) and \( t \). Since KG is able to represent knowledge in a straightforward and intuitive way, there is an increasing demand to build or construct KGs for different domains. Accordingly, the KG construction from acquired documents is of great practical interest to both academia and industry. However, expert assistance is usually required to manually identify heads, relations, and tails, which can be very labor expensive and time consuming. On the other hand, due to individual bias and preference, the resultant KGs could be significantly different from experts. As such, there is an increasing demand to develop robust (semi)automated techniques for Knowledge Graph construction. In the past decade, researchers have been actively investigating efficient KG construction approaches. Interested readers are referred to Section 2.

Despite the general interest of developing KG construction techniques, another problem comes from the domain obstacle. Due to the significant difference across
domains, there could exist different terminologies. Even with the same pair of terminologies, the relation could still vary for different context. As a result, it is difficult to develop a generic construction method that could be applicable to all domains. To address this issue, in this paper, we would like to introduce the concept of transfer learning to improve the construction performance by reusing existing KG models across different domains.

Transfer learning has attracted considerable interest in the past decade, which investigates that the application of existing model(s) trained from source domain on the target domain [5]. This transferring technique has achieved success in a variety of applications, such as relation extraction [6], facial landmark detection [7], and skin lesion classification [8]. A comprehensive literature review is provided in Section 2.

Following the work mentioned above, we propose a novel KG construction algorithm in this paper, where the construction performance is enhanced using the concept of transfer learning. That is, we propose the algorithm to integrate the transfer learning to enhance the existing model across different domains. As such, the trained model is able to balance the knowledge presentation from both the source and the target resources. As such, the proposed algorithm can help in reducing the labor cost in labeling the target domain by reusing a pretrained model. To our knowledge, it is the first attempt to conduct Knowledge Graph utilizing the concept of transfer learning. Empirically, a real-world data repository of gas-mining reports is employed. To investigate the potential factors and the relationship among risks, we apply the proposed algorithm to build the KG. The experimental results indicate that we achieve superior performances on the task, validating the effectiveness of the proposed model.

The rest of the paper is organized as follows. Section 2 gives a brief account of Knowledge Graph and transfer learning, respectively. Section 3 details the proposed algorithm, in which the transfer learning is integrated to improve the standard Knowledge Graph construction model, via transferring existing model to another domain. Section 4 discusses the experimental implementation and evaluates the performance of the proposed method, followed by concluding remarks given in Section 5.

2. Related Work

In this section, we provide background information on the study area, focusing on the theory of Knowledge Graph and transfer learning.

2.1. Knowledge Graph. The Knowledge Graph (KG) is a structured, directed graph: \( G = (E, R) \), where \( E \) and \( R \) represent entities (real-world objects and abstract concepts) and semantic relationships, respectively. Within KG, the data is usually formulated as the form of a tuple, (head is \( h \), relation is \( r \), and tail is \( t \)), where \( h \in E, t \in E, \) and \( r \in R \). Due to the wide application of KGs for the downstream tasks, a large amount of research effort has been conducted on building or constructing KGs, while the main aim includes identify (head/tail) entities (see Section 2.1.1) and match relations (see Section 2.1.2) from unstructured text or documents.

2.1.1. Entity Identification. Entity identification is often known as named entity recognition (NER), which aims to discover important objects (entities) from the text. The NER task has a long research history, from which hand-crafted features and sophisticated machine learning algorithms (such as neural network [9] and BiLSTM [10]) are applied in many literatures.

In [11], a multi-feature adaptive fusion Chinese named entity recognition (MAF-CNER) model is proposed. In particular, this model employed the bidirectional long short-term memory (BiLSTM) neural network as the backbone model as the feature extraction. There were in total three different features taken into account, including characteristics of characters, strokes, and radicals. Later, a fusion strategy was implemented by a weighted concatenation operation to combine these three features, before forming the final classification. The evaluation experiments have been considered using one Chinese news dataset, while this model resulted in an improved F1 values compared with other methods. Similarly, the BiLSTM model has also been employed in [12]. However, instead of only focusing on identifying the correct entities, this paper also suggested an auxiliary task to provide additional information for forming the global feature representation. This task has been formulated as determining or discovering certain types of entities from the training sentence. As such, one multi-task learning architecture has been designed jointly with two aims, recognizing the entities and their types simultaneously. The work in [13] proposed the combination of utilizing character and sentence vectors trained by distributed memory model of paragraph vectors (PV-DM). Additionally, the BiLSTM model was implemented with an additional conditional random field (CRF) layer, while this CRF layer is used to tag the input sentence. Finally, experiments have been conducted using a set of Chinese judicial documents, and the results showed that enhanced BiLSTM method can boost the NER performance significantly.

2.1.2. Relation Matching. Relation matching is used to extract relational facts from given texts, which forms a significant part of the KG construction. Due to the lack of labeled relational data, weak supervision or distant supervision has been firstly employed using heuristic rules to automatically generate training samples. The basic assumption is that sentences containing the pair of related entities may form a certain relation. More recent work has been focusing on utilizing the deep networks to provide supervision learning. For instance, in [14], a joint identification algorithm has been proposed to extract entity and relation. Authors have introduced two main steps, that is, entity and approximation relation model. For the former, the entity model took the input text and predicted all entities. At the second step, the relation extraction model considered every pair of entities independently by
inserting different entity markers, before predicting the potential relation type. More importantly, authors further introduced a batch of computational strategy, enabling efficient pair calculation of the relation model. The work from [15] utilized the dependency trees for the relation matching purpose. More precisely, a novel attention based Graph Convolutional Network was implemented taking as input the full dependency trees. This model can estimate the importance or significance of the relevant sub-structures, and apply the attention mechanism to identify the underlying relation. Experimental results show that this model generated state-of-the-art results on a couple of relation extraction tasks. More recently, a multi-grained lattice framework was proposed for Chinese relation extraction [16]. Their method consisted of three components/steps. At first, the word-level information was integrated with the input character sequence to improve the segmentation accuracy. The second step is to utilize the external linguistic knowledge to alleviate polysemy ambiguity. At last, the dropout mechanism was applied to randomly remove feature detectors during the forward propagation. The experimental results show that their model significantly outperformed other methods on several Chinese datasets by identifying the correct relation.

2.2. Transfer Learning. The procedure of transfer learning is to apply some learned knowledge from one known task to a similar but unknown task. We usually define the known task as the source domain, while the unknown task is defined as the target domain. As such, the technique of transfer learning is also known as the domain adaptation that investigates the common across different domains. The advantage of transfer learning can be summarized as the following:

(i) computational complexity: there is no need to train another model for the target domain from the scratch, as transfer learning aims to reuse the existing model (from the source domain). As a result, we only need to fine-tune existing model(s), which reduces the computational complexity;
(ii) data size: the target domain usually comes with a limited data size. That is, the sample number from the target domain is less compared with the source domain. As such, utilizing samples from the target domain might be insufficient for training purposes. With transfer learning, we can then train the model from the source domain (with a large number of training samples) before migrating the target one;
(iii) modeling performance: transfer learning has proven to achieve promising results in many areas with improved model accuracy. As the existing model has been trained using the source domain, the subsequent fine-tuning process (for the target domain) has a minimal impact on the previous model by avoiding the overfitting and improving the generalization capability.

Due to its wide application, transfer learning has consistently attracted a lot of attention from different topics. In [6], a novel relation extraction model has been proposed, which extends the adversarial based domain adaptation. More precisely, the authors addressed three main challenges when adopting a trained model to a new domain, including linguistic variation, imbalanced relation distribution, and partial adaptation. A relation-gated adversarial learning, accordingly, was introduced, while the experimental results revealed its superiority over many other methods. The work from [7] applied the transfer learning on the task of face recognition in the thermal infrared spectrum. Two particular problems were the low spatial resolution of limited availability of labeled thermal face samples. As a result, a visible-to-thermal based transfer learning algorithm has been proposed. In particular, a coupled convolutional network architecture is employed as the backbone model, and four types of parameter transfer strategies are conducted, such as Siamese (shared) layers, Linear Layer Regularization (LLR), Linear Kernel Regularization (LKR), and Residual Parameter Transformations (RPT). Similarly, a multi-view filtered transfer learning network has been introduced in [8] to solve the problem of skin lesion classification. The multi-view filtering method is particularly employed for determining the contributing samples from the source domain while neglecting the wrong ones. In addition, authors also considered to extract feature from various image views, and merge them into the final decision during the transfer learning process. Their experimental results in terms of skin lesion classification demonstrate the improved classification accuracy (approximately 91.8%) on Melanoma and Seborrheic Keratosis classification tasks.

Inspired by the superior performance from transfer learning, this study seeks to its application to the Knowledge Graph construction. Explicitly, we would like to explore to what extent transfer learning can overcome the issues of limited training samples from target domain, how well a trained model from source domain can be adopted to the target domain, and the impact from the number of training samples to the model performance.

3. Main Methodology

In this section, we present the proposed transfer learning based model for automated Knowledge Graph (KG) construction. Towards this end, we firstly formulate the KG construction process from the transfer-learning aspect. Secondly, the construction model for the target domain is enhanced using the existing model trained from the source domain. Overall, the overview of the proposed transfer learning based KG construction algorithm, termed as TLKG, is illustrated in Figure 1.

3.1. Problem Formulation. In the scenario of the KG construction, the input is the (un)structured texts or documents while the output will be the extracted KG. More precisely, the input document, denoted as \( D \), is a list of sentences. The KG is represented as a graph \( G = (E, R) \), where \( E \) and \( R \)
have (at least) two sets of domain documents: source and target. As a result, we formally introduce the concept of transfer learning to deal with this problem. Applying directly the trained model across different domains could reduce the model performance that is significantly different from each other. For instance, the training texts or documents might come from different domains, compared to that of testing samples. As such, the terminologies (entities) and relations could be significantly different from each other. Applying directly the trained model across different domains could reduce the model performance and generalization capability. As a result, we formally introduce the concept of transfer learning to deal with this problem.

In particular, when it comes to multiple domains, we will have (at least) two sets of domain documents: source ($D_s$) and target ($D_t$). Again, the former consists of the relatively large amount of training and test samples that is used to train the initial model. In contrast, the latter one has insufficient samples, which leads to a poor performance if the model has been trained directly on top. With the context of transfer learning, the problem has been formulated as estimating a transferring function (say $t$) that is able to migrate an existing model trained from the source domain to the target one, without compromising the performance. Assume that we have a construction model $M_s$ that has been trained using $D_s$, the proposed transfer learning method aims to minimize the following loss function:

$$\text{loss} = t(M_s, D_t),$$

where the transferring function $t$ takes the input of the existing model ($M_s$) and outputs the construction result from the target dataset ($D_t$).

3.2. Supervised TLKG. The procedure of identifying the function $t$ in (1) is simple but effective, and we propose the following two steps:

(i) **Pretraining**: we initialize the model ($M_s$) on the source domain ($D_s$) and train $M_s$ until the convergence;

(ii) **Fine-tuning**: the trained model $M_s$ is adjusted using the target domain ($D_t$), which involves the fine-tuning technique for hyperparameters.

From the aforementioned steps, the first one is also called the source task, and the second one is referred to as the target task. With the KG construction context, the source task contains a collection of documents, that is well labeled with the correct entities and relations (the KG tuples). Again, the source task contains abundant training samples, which is sufficient to train the original model ($M_s$). With the second step, the target task has also a collection of documents, although their topic or content might be from another domain compared with the source task. Another common part is that the target task also contains training samples. However, the number is (much) less than that of source task. The lack of training samples leads to a poor construction performance if we directly apply the model on $D_t$. Towards this end, in our context, both the source and target datasets provide the correct labels (of the underlying Knowledge Graph) during pretraining and fine-tuning steps. In other words, the model of $M_s$ is optimized by the objective function in a supervised manner in both stages.

Overall, the proposed algorithm, termed TLKG, is summarized in Algorithm 1. Firstly, TLKG pretrains a construction model during the source task, in which the model is trained to discover correctly KG tuples using the training sample from the source domain ($D_s$). Secondly, the TLKG technique is to fine-tune the existing model for the domain adaption purpose. That usually requires training the existing model based on the target domain ($D_t$) for few echoes. Note that, the effectiveness of TLKG is mainly evaluated based on the performance from the target task.

Additionally, one may also decide the backbone model of the KG construction for $D_s$. This model is mainly used to discover potential entities from the given documents and match the pair of two entities (the head and tail) to form the relation. There is a great number of existing work in this regard. As the searching for the optimal backbone model is beyond the scope of this paper, we leave the investigation to future work.

4. Experimental Analysis

To investigate the efficiency of the proposed algorithm, we present the experimental results using a real-world dataset. The dataset statistics and the evaluation criterion are presented in Section 4.1. The effect of key parameters is evaluated in Section 4.2. The comparison results between the proposed algorithm and other methods are then presented in Section 4.3.

4.1. Setup and Configuration. To evaluate the proposed algorithm, we firstly need to decide the source and target domain, respectively. In the following, we will detail two datasets which serve this purpose.

The first dataset come from Wikipedia (Chinese), that plays the role of source domain. To train a model that can identify underlying KGs, we need a large amount of labeled training
samples. Towards this end, we collect paragraphs from Wikipedia, in particular, the Chinese version. Then, we use distant supervision to label the collected data samples. More precisely, we match Wikipedia sentences with one of the well-developed open Chinese KGs, ownthink (see https://github.com/ownthink/KnowledgeGraphData). To begin with, we match the mentioned entities from a sentence to the corresponding entities of ownthink. Note that there will be some pronouns within the collected paragraphs, while keeping them may cause the problem of the co-reference resolution. As a result, we replace the identified co-references with the implicit entity name. After this step, we are able to identify the mentioned entities from the given sentences according to ownthink. Second, we apply a corpus based paraphrase detection to discover the relations between extracted entities in the sentence. This paraphrase corpus has been established by populating predicate paraphrases from ownthink. In addition, the exact-string matching method is then employed to link entities with relations from the given sentence. The resultant Wikipedia (Chinese) dataset contains 5,654 tuples from 15,231 entities with relations from the given sentence. K. he resultant the exact-string matching method is then employed to link populating predicate paraphrases from ownthink. In addition, the overall statistics of the employed corpus is shown in Table 1.

The second dataset is about the gas-mining risk reports, which plays the role of target domain. In particular, we collected 10 real-world reports, concluding gas leaking accidents. Each report has been provided with several paragraphs, while we mainly retain (filter) those with the accident process, reason, and potential solution. As such, the length from filtered reports is approximately 100 sentences and the average length of the sentence is 15 Chinese words, leading to an overall 1000 samples. We further split the target domain into a training set (40%) and testing set (60%). Additionally, those sentences include details of accident temporal and spatial information, such as the accident location/time, the involving staff, and the direct and indirect reason. The overall statistics of the employed corpus is shown in Table 1.

With the context of transfer learning, the study is to firstly pretrain a backbone model using the source domain (the Wikipedia Chinese data) and transfer this model to construct the Knowledge Graph in the target domain (the gas-mining risk documents). In our paper, we are utilizing the model from [16] as our backbone model, and the relevant hyperparameters are listed in Table 2.

In terms of the result measurement, we use the following metric to evaluate the accuracy performance of KG construction:

$$P_{acc} = \frac{TP + TN}{TP + TN + FP + FN}$$  \hspace{1cm} (2)

where TP is the positive tuple predicted by the model to be positive; TN is the negative tuple predicted to be negative by the model; FP is the negative tuple predicted to be positive by the model; and FN is the positive tuple predicted to be negative by the model.

4.2. Parameter Validation. In this section, the impact from the number of training samples (K) based on the source dataset is assessed. Note that a bigger value of K could lead to a well-trained construction model. However, it might also result in an intensive training computation. On the other hand, a smaller K might suffer from the insufficient information. With the context of transfer learning, we also want to know a suitable value of K to avoid the overfitting on the source domain (a bigger value of K). Toward this end, experiments are conducted by considering different settings of K (K ∈ [30%, 50%, 70%, 100%]) to get an overall impact on the performance. Experiments are performed by randomly selecting K samples from the source domain (for the model training purpose). The model is later fine-tuned by retraining the model for 5 iterations using the training sample from the target domain, and the comparison result (in term of the accuracy) over 10 runs is shown in Figure 2.

From the average results, it is clear to see that with a more number of training samples from the source domain (larger K), a better training performance is obtained, compared with a smaller K. The reason is that a smaller number of training samples produce a poor-trained model, resulting in the poor performance in the target domain as well. With the increase of K, the construction accuracy for the target domain is also improved, as more information has been provided to train the backbone model in the source domain. However, we also notice that the testing accuracy reaches the peak when K = 70%, which means more samples from the source domain might not be useful to improve the generalization capability in the target domain. The reason could be that an overfitted backbone model is difficult to transfer to another domain. As such, in the following comparison, we are utilizing K = 70%
samples from the source domain to pretrain the backbone model before adopting the target domain.

4.3. Comparison with Other Works. In this section, the proposed algorithm is compared with three conventional algorithms, namely, the multi-grained lattice framework (MGL) [16], joint identification (JI) [14], and the dependency tree-based work (DT) from [15]. We have briefly introduced those algorithms in Section 2, and for hyperparameters for those baseline methods, you can refer to the original papers.

Figure 3 presents the average training and test accuracy obtained from different methods (based on the target domain, gas-risk reports). Notably, as mentioned before, the target domain contains 1000 samples, while the training samples occupy 40%. To directly apply the backbone model on the target domain (without transfer learning) results in a training performance 72.6% and testing performance 65.6%. The proposed algorithm outperforms existing methods via achieving an improvement in terms of the accuracy. For instance, DT [15] only achieves less than 67% testing accuracy, compared with that of 73.7% (the proposed). One main reason could be due to the lack of sufficient training samples in the target domain, which leads to the poorly-trained model. On the other hand, by utilizing the transfer learning, the proposed method has accumulated existing knowledge from the source domain before fine-tuning within the target domain, which helps in enhancing both the training and testing accuracy. In conclusion, it can be empirically confirmed that the proposed algorithm outperforms existing state-of-the-art approaches.

5. Conclusion

The automated construction for Knowledge Graph (KG) is an active area of research due to its ability to benefit the downstream tasks. In this study, we propose an improved algorithm based on the concept of transfer learning. More precisely, we firstly pretrain a backbone model using the samples from a source domain. Compared with the target domain, the source domain has more sufficient samples, which could lead to a well-trained backbone model. Then, this model has been transferred to the target domain, which is further fine-tuned for the domain adoption. Our work is then evaluated using a real-world dataset (for the gas-mining risk reports), and experimental results show that the proposed algorithm yields competitive detection accuracy compared to traditional approaches.

Data Availability

The data in the paper can be obtained through https://github.com/ownthink/KnowledgeGraphData.
Conflicts of Interest
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgments
This work was partially supported by the National Natural Science Foundation of China under Grant Nos. 61873004 and 51474007.

References