

Does William Shakespeare REALLY Write Hamlet? Knowledge Representation Learning with Confidence

Ruobing Xie^{1,2}, Zhiyuan Liu^{1*}, Fen Lin², Leyu Lin²

¹ Department of Computer Science and Technology,

State Key Lab on Intelligent Technology and Systems,

National Lab for Information Science and Technology, Tsinghua University, China

² Search Product Center, WeChat Search Application Department, Tencent, China.

Abstract

Knowledge graphs (KGs), which could provide essential relational information between entities, have been widely utilized in various knowledge-driven applications. Since the overall human knowledge is innumerable that still grows explosively and changes frequently, knowledge construction and update inevitably involve automatic mechanisms with less human supervision, which usually bring in plenty of noises and conflicts to KGs. However, most conventional knowledge representation learning methods assume that all triple facts in existing KGs share the same significance without any noises. To address this problem, we propose a novel confidence-aware knowledge representation learning framework (CKRL), which detects possible noises in KGs while learning knowledge representations with confidence simultaneously. Specifically, we introduce the triple confidence to conventional translation-based methods for knowledge representation learning. To make triple confidence more flexible and universal, we only utilize the internal structural information in KGs, and propose three kinds of triple confidences considering both local and global structural information. In experiments, We evaluate our models on knowledge graph noise detection, knowledge graph completion and triple classification. Experimental results demonstrate that our confidence-aware models achieve significant and consistent improvements on all tasks, which confirms the capability of CKRL modeling confidence with structural information in both KG noise detection and knowledge representation learning.

Introduction

Recent years have witnessed the great thrive in artificial intelligence that has broad impacts on our daily lives. In tasks like information retrieval and question answering, people are not satisfied with merely semantic matching, but expect AI agents to have knowledge for understanding, reasoning and solving. Knowledge graphs (KGs), which provide effective well-structured relational information between entities, are essential supporters for knowledge-based AI agents. A typical KG usually stores knowledge with triple facts in the form of (*head entity*, *relation*, *tail entity*), which is also abridged as (*h*, *r*, *t*).

There are existing amounts of widely-utilized large-scale knowledge graphs such as Freebase (Bollacker et al. 2008), DBpedia (Auer et al. 2007) and other domain-specific KGs. However, these knowledge graphs are still far from complete to describe the infinite real-world facts, in which some kinds of knowledge may even change frequently. Therefore, knowledge construction and timely update are significant for knowledge-driven applications. Most conventional knowledge graph construction methods usually involve huge human supervision or expert annotation, which are extremely labor-intensive and time-consuming. Nowadays, automatic mechanism and crowdsourcing take larger parts in knowledge construction, while these methods may suffer from possible noises and conflicts due to limited human supervision. For instance, recent neural relation extraction model on benchmark achieves only around 60% precision when the recall is 20% (Lin et al. 2016). Moreover, (Heindorf et al. 2016) focuses on vandalism detection in Wikidata, which also verifies the existence and problems of noises in KGs.

In this paper, we concentrate on how to deal with noises in knowledge representation learning (KRL), which provides an effective and flexible way for using knowledge. KRL represents entities and relations with distributed representations mainly according to triple facts in KGs. Therefore, it is crucial to consider noises in knowledge representation learning and knowledge-driven tasks.

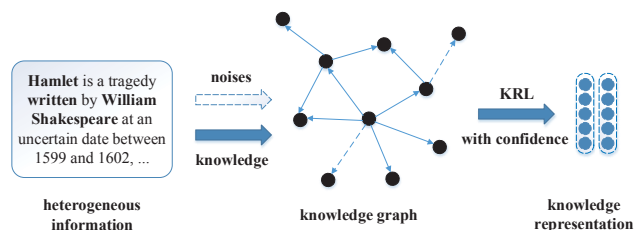


Figure 1: Noises in KGs and confidence-aware KRL.

We attempt to detect possible noises and conflicts located in existing knowledge graphs, while constructing noise-free knowledge representations simultaneously. However, most conventional KRL methods assume that all triple facts in existing KGs are completely correct. To solve this problem, we propose a novel confidence-aware knowledge representation learning framework (CKRL).

*Corresponding author: Z. Liu (liuzy@tsinghua.edu.cn)

tation learning (CKRL) framework taking possible noises into consideration. The notions of confidence and trust have been widely studied in fields such as cognitive science. Fig. 1 demonstrates a brief illustration of our confidence-aware KRL framework, where the knowledge graph has both knowledge and noises extracted from heterogeneous sources. These noises are expected to be detected via our confidence-aware model and to be ignored in knowledge representation learning.

Specifically, CKRL follows the translation-based framework proposed by (Bordes et al. 2013), and learns knowledge representations with triple confidences. We propose three triple confidences considering both local triple and global path information, and knowledge representations are learned to fit the global consistency under translation assumption weighted by those dynamic triple confidences. To make the triple confidence more universal and flexible, we only consider the internal structural information in KGs, which correspondingly makes noise detection much more challenging due to the limited information.

In experiments, we evaluate our models on three tasks including knowledge graph noise detection, knowledge graph completion and triple classification. The results demonstrate that our models achieve the best performances on all tasks, which confirm the capability of CKRL in noise detection and knowledge representation learning. The main contributions of this work are concluded as follows:

- We propose a novel confidence-aware KRL framework for knowledge graph noise detection and knowledge representation learning simultaneously, which only uses internal structural information in KGs.
- We evaluate our CKRL models on several datasets with different noise rates extended from a real-world dataset, and achieve promising performances on all tasks.
- The idea of triple confidence in CKRL could be utilized not only in knowledge representation learning, but also in knowledge construction.

Related Work

Knowledge Graph Noise Detection

It seems to be inevitable that noises do exist in KGs, which can strongly affect knowledge acquisition (Manago and Kodratoff 1987). Moreover, a novel task named Wikidata vandalism, which aims to combat with deliberate destructions in knowledge graphs, has attracted wide attention (Heindorf et al. 2015). Therefore, noise detection is essential in knowledge construction and knowledge application. Most knowledge graph noise detection works happen when constructing knowledge graphs. For instance, YAGO2 extracts knowledge from Wikipedia with human supervision that human judges are presented with selected facts for which they have to assess the correctness (Hoffart et al. 2013). Wikidata also relies on a crowd-sourced human curation software in which contributors can reject or approve a statement (Pelissier Tanon et al. 2016). DBpedia creates its mappings to Wikipedia infoboxes via a worldwide crowd-sourcing effort (Lehmann et al. 2015). These noise detections in large-scale

KGs are usually involved with huge human efforts, which are extremely labor-intensive and time-consuming.

Recently, there are also lots of researches focusing on automatic KG noise detection (Nickel et al. 2016). However, most existing methods mainly concentrate on feature selection from contents, users, items and revisions (Heindorf et al. 2016), and thus are constrained by the completeness of external information. There are also some efforts working on judging importance in graphs for nodes (Gyöngyi, Garcia-Molina, and Pedersen 2004) or for edges (De Meo et al. 2012), but few works concentrating on the confidence of each triple. Knowledge Vault (Dong et al. 2014) proposes a joint approach with both web content (e.g., texts, tabular data and human annotations) and prior knowledge derived from existing KGs to judge triple qualities, whose performance strongly depends on the external information. In this paper, we introduce three triple confidences to KG noise detection and knowledge representation learning, which only focus on the internal structural information in KGs.

Translation-based KRL Methods

Recent years many efforts concentrate on learning distributed representations for knowledge graphs, among which the translation-based methods are both straightforward and effective with the state-of-the-art performances. TransE (Bordes et al. 2013) projects both entities and relations into a continuous low-dimensional vector space, interpreting relations as translating operations between head and tail entities. The translation assumption in TransE implies the equation that $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$. The energy function is defined as follows:

$$E(h, r, t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|. \quad (1)$$

TransE can well balance both effectiveness and efficiency compared to traditional methods, while the over-simplified translation assumption constrains the performance when dealing with complicated relations. Some enhanced KRL methods based on TransE attempt to solve this problem with translations on relation-specific hyperplanes (Wang et al. 2014), relation-specific entity projection (Lin et al. 2015b) and type-specific entity projection (Xie, Liu, and Sun 2016). Moreover, the translation assumption only focuses on the local information in triples, which may fail to make full use of global graph information in KGs. (Lin et al. 2015a) extends TransE by encoding multi-step relation path information into knowledge representation learning. However, most conventional KRL methods assume that all triples in KG share the same confidence, which is inappropriate especially for those KGs constructed automatically with less human supervision. To the best of our knowledge, our model is the first embedding method to consider triple confidences of existing KGs in KRL. In this paper, we extend TransE to learn knowledge representations from noisy KGs, and it is not difficult for other enhanced translation-based methods to utilize our confidence-aware KRL framework.

Methodology

We first give the notations used in this paper. Given a triple fact (h, r, t) , we consider the head and tail entities $h, t \in E$

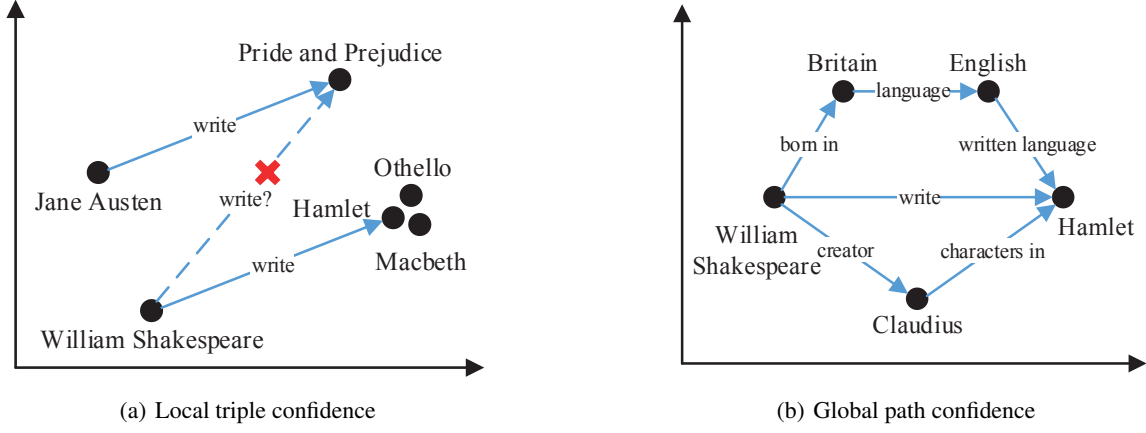


Figure 2: Effective mechanism of local triple confidence and global path confidence.

and the relation $r \in R$, where E and R stand for the sets of entities and relations. T represents the overall training triple facts including possible conflicts and noises.

To detect possible noises in knowledge graphs and learn better knowledge representations, we introduce a novel concept **triple confidence** for each triple fact. Triple confidence describes the correctness and significance of a triple, which could be measured with the favor of both internal structural information and external heterogeneous information.

Confidence-aware KRL Framework

We attempt to detect noises and learn better knowledge representations with triple confidence taken into consideration, concentrating more on those triples with high confidences. Following the translation-based framework, we design our **confidence-aware KRL energy function** as follows:

$$E(T) = \sum_{(h,r,t) \in T} E(h,r,t) \cdot C(h,r,t). \quad (2)$$

The confidence-aware energy function can be divided into two parts: $E(h,r,t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||$ stands for the **dissimilarity score under translation assumption**, which is the same as that of TransE. A lower $E(h,r,t)$ indicates that the entity and relation representations of this triple fit better with our translation framework. While differing from conventional methods, we also introduce the **triple confidence $C(h,r,t)$** as the second part of our energy function. A higher triple confidence implies that the corresponding triple is more credible, and thus should be more considered.

Triple confidence can be calculated both during and after knowledge graph construction, from various aspects including internal information such as **graphic evidence** and external information such as **textual evidence**. To make our triple confidence more universal and practical, we only consider the **internal structural information** after KG construction in our model, and propose both local and global triple confidences which are iteratively optimized during training. In CKRL, we bring in the confidences of different triples to learn more about those significant triples, and thus could get better knowledge representations.

Objective Formalization

We introduce the detailed training objective of our model in this section. Following TransE (Bordes et al. 2013), we formalize a **margin-based score function with negative sampling as objective for training**. This pair-wise score function attempts to make the scores of **positive triples to be higher than those of negative triples**. We have:

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \max(0, \gamma + E(h,r,t) - E(h',r',t')) \cdot C(h,r,t), \quad (3)$$

where $E(h,r,t)$ is the dissimilarity score of positive triple and $E(h',r',t')$ is that of negative triple. $\gamma > 0$ is the hyper-parameter of margin, and T' represents the negative triple set. Here the triple confidence $C(h,r,t)$ instructs our model to **pay more attention on those more convincing facts**.

For pair-wise training, since there are no explicit negative triples in knowledge graphs, we sample negative triples complying with the following rules:

$$T' = \{(h',r,t)|h' \in E\} \cup \{(h,r,t')|t' \in E\} \cup \{(h,r',t)|r' \in R\}, \quad (h,r,t) \in T. \quad (4)$$

It means that one entity or relation in a positive triple is randomly replaced by another entity or relation in the overall set. Note that differing from TransE, we also add relation replacements for better performances in relation prediction. We also discard all triples already in T from T' to make sure our generated negative triples are truly negative.

Local Triple Confidence

We first come up with the local triple confidence (LT) which only **concentrates on the inside of a triple**. Since our CKRL framework follows the translation assumption that $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$, it is straightforward to directly **utilize this dissimilarity function to judge triple confidences**. Moreover, the promising results of conventional translation-based methods on triple classification also confirm that positive triples should comply with the translation assumption well. Fig. 2(a) demonstrates the effective mechanism of local triple confidence.

Since knowledge representations learned under our translation framework should follow the global consistency with all triples in KG, we can infer that *William Shakespeare* is more likely to write *Hamlet* rather than *Pride and Prejudice*, even though there are noises in training set.

We assume that the more a triple fits the translation assumption, the more convincing this triple should be considered. To measure the local triple confidence during training, we first judge the current conformity of each triple with translation assumption. Inspired by the margin-based training strategy, we directly use the same pair-wise function to calculate the triple quality $Q(h, r, t)$ as follows:

$$Q(h, r, t) = -(\gamma + E(h, r, t) - E(h', r', t')). \quad (5)$$

A higher $Q(h, r, t)$ usually indicates a better triple judged by the translation assumption. At the beginning of training, we suppose all triples are correct, and initialize the local triple confidence $LT(h, r, t)$ for all triples as 1. Since both entity and relation embeddings will change during training, the current local triple confidence for each triple should be also updated according to how much this triple fit the translation assumption. Formally, the local triple confidence $LT(h, r, t)$ changes with its triple quality $Q(h, r, t)$ as follows:

$$LT(h, r, t) = \begin{cases} \alpha LT(h, r, t), & Q(h, r, t) \leq 0. \\ LT(h, r, t) + \beta, & Q(h, r, t) > 0. \end{cases} \quad (6)$$

$Q(h, r, t) \leq 0$ implies the current triple doesn't comply with the translation rule, and thus the corresponding local triple confidence should decrease. On the contrary, $Q(h, r, t) > 0$ encourages the triple to have a larger local triple confidence. Here, $\alpha \in (0, 1)$ and $\beta > 0$ are hyper-parameters that control the ascend or descend pace of local triple confidence, with the assurance that $LT(h, r, t) \in (0, 1]$. Note that the local triple confidence will decrease at a geometric rate while increase with a constant addition. It is because that we urge to punish the violations of translation rule, for those triples are more likely to be noises or conflicts, and thus should have smaller confidences.

Global Path Confidence

The local triple confidence is straightforward and effective, while simply concentrating on the inside of triples will fail to use rich global structural information in knowledge graphs. Moreover, local triple confidence won't work well with high noise rate. Therefore, we propose the global path confidence to take multi-step relation paths into consideration. For instance, in Fig. 2(b), there are two multi-step relation paths from *William Shakespeare* to *Hamlet*. The lower path provides strong evidence to infer the relation *write*, while the upper path just provides weaker evidence. These paths could also help us to judge triple qualities.

A triple considered to have high global path confidence should follow two conditions: (1) it has more reliable paths from its head to tail entity, and (2) these reasoning paths are semantically closer to the corresponding relation. In the following sub-sections, we first introduce how to quantify the relation path reliability for each triple, and then propose two strategies using prior co-occurrence information and learned

knowledge representations to measure the semantic similarity between paths and relations.

Relation Path Reliability We assume that a relation path should be considered more important if it carries more information flow from the head to tail entity. Specifically, we follow the path-constraint resource allocation (PCRA) (Lin et al. 2015a) to measure the relation path reliability. The key idea of PCRA is inspired by resource allocation (Zhou et al. 2007), which supposes there are certain resources associated with head entity h , and will flow throughout the whole knowledge graph via all relation paths. The resource amount that eventually flows to the tail entity t via a certain path p will be considered as the relation path reliability of p given the entity pair (h, t) .

Formally, given a path $p = (r_1, \dots, r_l)$ and entity pair (h, t) , the resource in h will flow to t through l steps. Since there are probably multiple tails given head and relation, the path is represented as $E_0 \xrightarrow{r_1} \dots \xrightarrow{r_l} E_l$, where E_i represents the entity set at the i -th step, $E_0 = \{h\}$ and $t \in E_l$. For entity $e \in E_i$, the resource $R_p(e)$ will be calculated as follows:

$$R_p(e) = \sum_{e' \in E_{i-1}(\cdot, e)} \frac{R_p(e')}{|E_i(e', \cdot)|}, \quad (7)$$

in which $E_{i-1}(\cdot, e)$ represents the direct predecessors of e via r_i , and $E_i(e', \cdot)$ represents the direct successors of e' via r_i . All entities will be initialized with the same resource amount, and finally after l steps from h to t , the resource amount $R_p(t)$ is regarded as the relation path reliability $R(h, p, t)$ of p with the given entity pair (h, t) .

Prior Path Confidence We first introduce prior path confidence (PP), which utilize the co-occurrence of relation and path to represent their dissimilarity. We suppose that the more a relation occurs with a path, the more they are likely to represent similar semantic meanings. Formally, given a triple (h, r, t) and its path set $S_{(h, t)}$ containing all paths between h and t , the quality of the i -th relation-path pair (r, p_i) is written as follows:

$$Q_{PP}(r, p_i) = \epsilon + (1 - \epsilon) \frac{P(r, p_i)}{P(p_i)}, \quad (8)$$

where $P(r, p_i)$ represents the prior probability of r and p_i co-occurrence, and $P(p_i)$ represents the prior probability of p_i in KG. ϵ is a hyper-parameter for smoothing. Therefore, the prior path confidence (PP) is designed as follows:

$$PP(h, r, t) = \sum_{p_i \in S_{(h, t)}} Q_{PP}(r, p_i) \cdot R(h, p_i, t). \quad (9)$$

It indicates that the prior path confidence of (h, r, t) depends on both relation-path similarities of all paths in $S_{(h, t)}$ and their corresponding relation path reliabilities. Note that since we merely consider the prior probabilities of paths and relations, prior path confidences are fixed during training.

Adaptive Path Confidence The prior path confidence stays static during training, which is inflexible and may be strongly constrained by existing noises and conflicts in

KGs. To address this problem, we propose the adaptive path confidence (AP) that could flexibly learn relation-path qualities according to their learned embeddings. Formally, given r and $p_i = \{r_{i1}, \dots, r_{ik}\}$, we directly represent the path embedding \mathbf{p}_i with the sum of its relation embeddings $\mathbf{r}_{i1} + \dots + \mathbf{r}_{ik}$ under the translation assumption. The relation-path quality function of AP is defined as follows:

$$Q_{AP}(r, p_i) = \|\mathbf{r} - \mathbf{p}_i\| = \|\mathbf{r} - (\mathbf{r}_{i1} + \dots + \mathbf{r}_{ik})\|. \quad (10)$$

Since we assume that the relation embedding should be similar as the path embedding, a lower $Q_{AP}(r, p_i)$ implies a more convincing relation-path pair. The adaptive path confidence is then written as follows:

$$AP(h, r, t) = \sigma\left(\sum_{p_i \in S(h, t)} \frac{R(h, p_i, t)}{Q_{AP}(r, p_i)}\right), \quad (11)$$

in which $\sigma(\cdot)$ stands for the sigmoid function. Adaptive path confidence can describe triple confidences dynamically with evolutionary relation embeddings during training, making our triple confidences more flexible and precise.

The overall triple confidence combines with all three kinds of confidences stated above. We have:

$$C(h, r, t) = \lambda_1 \cdot LT(h, r, t) + \lambda_2 \cdot PP(h, r, t) + \lambda_3 \cdot AP(h, r, t), \quad (12)$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyper-parameters.

Optimization and Implementation Details

We utilize mini-batch stochastic gradient descent (SGD) to optimize our model. In training, all entity and relation embeddings could be either initialized randomly or pre-trained with TransE. For those entity pairs that don't have paths, we directly set their path-based confidences as 0.

Path selection is essential in our model that will have significant impacts on the performances. Since the number of all paths grows exponentially with the increase of maximum path length, it is impractical to enumerate all paths in KG. Moreover, the path-based inference will be much weaker when the logical chain goes too far. Considering both effectiveness and efficiency, we limit the maximum length of paths to at most 2-steps to prevent possible error propagation. Since relations in KGs are directed edges, we also consider those reverse relations when we detect relation paths.

Experiment

Datasets

In this paper, we evaluate our CKRL model based on FB15K (Bordes et al. 2013), which is a typical benchmark knowledge graph extracted from Freebase (Bollacker et al. 2008). However, there are no explicit labelled noises or conflicts in FB15K. Therefore, we generate new datasets with different noise rates based on FB15K to simulate the real-world knowledge graphs constructed automatically with less human annotation.

Most noises and conflicts in real-world knowledge graphs derive from the misunderstanding between similar entities. It indicates that the noise (*Jane Austen*, *write*, *Hamlet*)

is more likely to occur in real-world KG rather than (*Soccer*, *write*, *Hamlet*), in which the latter could be easily detected via entity type constraints. Inspired by the pre-processing in the evaluation task named triple classification, we construct negative triples (i.e., noises) following the same setting in (Socher et al. 2013). Specifically, given a positive triple (h, r, t) in KG, we randomly switch one of head or tail entities to form a negative triple (h', r, t) or (h, r, t') . The generation of negative triples is constrained that h' (or t') should have appeared in the head (or tail) position with the same relation r in dataset, which means that the head entity of relation *write* in negative triples should also be a writer. This constraint focuses on generating harder and more confusing cases, for those negative triples with mistype entities could be easily detected. We can directly utilize entity type information in Freebase or follow the local closed-world assumption (Krompaß, Baier, and Tresp 2015) to collect type constraint information. Following this protocol, we construct three KGs based on FB15K with negative triples to be 10%, 20% and 40% of positive triples, and then discard a small number of negative triples that violate type constraints. All three noisy datasets share the same entities, relations, validation and test sets with FB15K, with all generated negative triples fused into the original training set of FB15K. The statistics are listed in Table 1.

Table 1: Statistics of datasets

Dataset	#Rel	#Ent	#Train	#Valid	#Test
FB15K	1,345	14,951	483,142	50,000	59,071
Datasets	FB15K-N1	FB15K-N2	FB15K-N3		
#Neg triple	46,408	93,782	187,925		

Experimental Settings

In experiments, we evaluate our CKRL models with three different confidence combination strategies. CKRL (LT) represents the strategy which only considers local triple confidence, CKRL (LT+PP) considers both local triple confidence and prior path confidence, while CKRL (LT+PP+AP) considers all three kinds of triple confidences. We implement TransE (Bordes et al. 2013) as baseline for the CKRL learning framework is based on TransE, and it is not difficult for our confidence-aware framework to be utilized in other enhanced translation-based methods.

We train our CKRL model using mini-batch SGD with the margin γ empirically set as 1.0. We select the overall learning rate δ among $\{0.0005, 0.001, 0.002\}$, which is fixed during training. For local triple confidence, we select the descend controller α among $\{0.5, 0.7, 0.9\}$ and the ascend controller β among $\{0.0001, 0.0005, 0.001\}$. For prior path confidence, the smoothing ϵ is empirically set as 0.01. The optimal configurations of our models are: $\delta = 0.001$, $\alpha = 0.9$, $\beta = 0.0001$, which are optimized on the validation set. We also evaluate various combination weights λ_i when we calculate the overall triple confidence based on the three proposed methods. We select a unified weighting

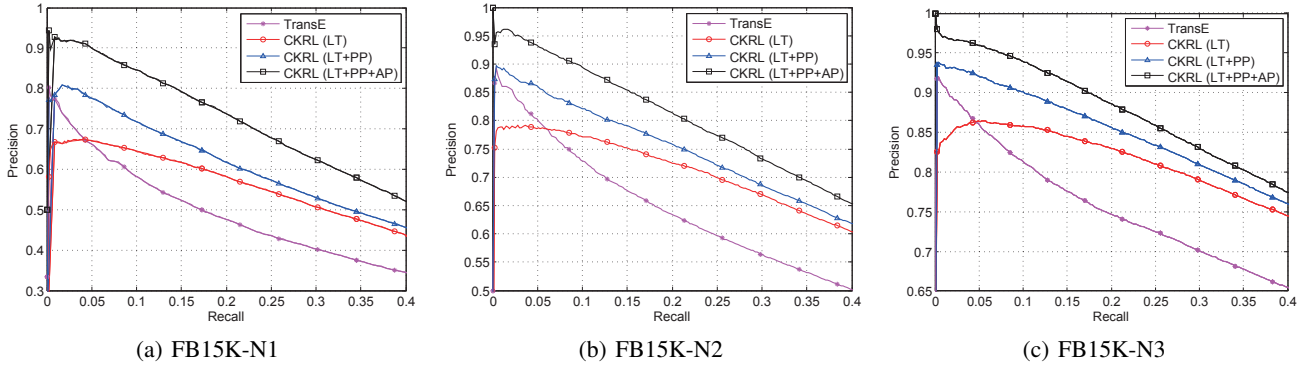


Figure 3: Evaluation results on knowledge graph noise detection.

strategy for different evaluation tasks and datasets according to their overall performances to show the robustness of our CKRL models. Specifically, for CKRL (LT+PP), we select $\lambda_1 = 0.9$ and $\lambda_2 = 0.1$, while for CKRL (LT+PP+AP), we select $\lambda_1 = 1.5$, $\lambda_2 = 0.1$ and $\lambda_3 = 0.4$. For fair comparisons, the dimensions of both entity and relation embeddings in all models are equally set to be 50.

Knowledge Graph Noise Detection

To verify the capability of our CKRL models in distinguishing noises and conflicts in knowledge graphs, we propose a novel evaluation task named knowledge graph noise detection. This task aims to detect possible noises in knowledge graphs according to their triple scores.

Evaluation Protocol Inspired by the evaluation metric of triple classification in (Socher et al. 2013), we consider the energy function scores $E(h, r, t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||$ as our triple scores, and then rank all triples in training set according to these scores. Those triples with higher scores are first considered to be noises. We utilize precision/recall curves to demonstrate the performances.

Experimental Results Fig. 3 demonstrates the results of knowledge graph noise detection, from which we can observe that: (1) our confidence-aware KRL models achieve the best performances on all three datasets with different noise proportions. It confirms the capability of our CKRL models in modeling triple confidence and detecting noises and conflicts in knowledge graphs. (2) CKRL (LT+PP+AP) has significant and consistent improvements in noise detection compared to other confidence-aware strategies. It indicates that the adaptive path confidence could provide more flexible and credible evidence for noise detection. Moreover, CKRL (LT+PP+AP) achieves impressively 84 ~ 94% in precision with different noise proportions when the recall is 10%, which implies that our models could truly help in real-world KG noise detection. (3) CKRL (LT+PP) performs better than CKRL (LT) especially at the beginning of PR curves, which matters more in real-world KG noise detection systems. It implies that even though the local triple confidence is capable of capturing KG global consistency

via learned knowledge representations, the global path information could still be a qualified supplement via multi-step path reasoning. (4) For further comparisons, we also evaluate PTransE (Lin et al. 2015a) which considers multi-step paths in KRL on this task with its energy function, while the results are surprisingly much worse than TransE. We find that PTransE can not detect noises in KGs well, for its path-based energy function scores only work when comparing positive and negative pairs. We don't show the results of PTransE due to the limited space.

Knowledge Graph Completion

Knowledge graph completion is a classical evaluation task that concentrates on the quality of knowledge representations (Bordes et al. 2012). This task aims to complete a triple when one of head, tail or relation is missing, which can be viewed as a simple question answering task.

Evaluation Protocol In this paper, we mainly focus on entity prediction, which is determined by the translation assumption that $\mathbf{h} + \mathbf{r} \simeq \mathbf{t}$. Following the same settings in (Bordes et al. 2013), we conduct two measures as our evaluation metrics: (1) Mean Rank of correct entities, and (2) Hits@10 that indicates the proportion of correct answers ranked in top 10. We also follow the different evaluation settings of "Raw" and "Filter" utilized in (Bordes et al. 2013).

Experimental Results In Table 2 we demonstrate the results of entity prediction with different noise rates, from which we can observe that: (1) all confidence-aware KRL models consistently and significantly outperform baseline on all noisy datasets with all evaluation metrics. It confirms the quality of learned knowledge representations, for they could not only detect noises in knowledge graphs, but also perform well in knowledge graph completion. (2) Comparing with evaluation results between different datasets, we find that the improvements introduced by our confidence-aware methods become more significant as the noise rate in KGs goes higher. It indicates that noises are harmful to entity prediction, and on the other hand reaffirms that considering triple confidence in knowledge representation learning is essential. (3) It seems that the global path confidence has few contributions on entity prediction. It may be partially

Table 2: Evaluation results on entity prediction

Datasets	FB15K-N1				FB15K-N2				FB15K-N3			
	Mean Rank		Hits@10(%)		Mean Rank		Hits@10(%)		Mean Rank		Hits@10(%)	
Metric	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
TransE	240	144	44.9	59.8	250	155	42.8	56.3	265	171	40.2	51.8
CKRL (LT)	237	140	45.5	61.8	243	146	44.3	59.5	244	148	42.7	56.9
CKRL (LT+PP)	236	139	45.3	61.6	241	144	44.2	59.4	245	149	42.8	56.8
CKRL (LT+PP+AP)	236	138	45.3	61.6	240	144	44.2	59.3	245	150	42.8	56.6

caused by the uncertainty and incompleteness in path information due to possible error propagation and limited path selection. In parameter analysis, we find that a higher weight of global path confidence within a reasonable range will improve the performances of entity prediction while harming those of KG noise detection. Taking longer relation paths into consideration with prior knowledge or patterns to assure high path qualities will partially solve this problem. (4) For a more comprehensive comparison, we evaluate TransE on the original FB15K dataset (without any noises) with the same parameter settings, evaluation protocols and test set. The result of Hits@10 (Filter) is 64.7% and Mean Rank (Filter) is 134, which can be viewed as the upper bound of our models. Compared to these results, the improvements of CKRL seem to be good as well.

Triple Classification

Triple classification aims to predict whether a triple in test set is correct or not according to the dissimilarity function, which could be viewed as a binary classification task. Triple classification could also be regarded as a simpler knowledge graph noise detection task in test set, for the noises in training set will influence the construction of knowledge representations, while the negative triples generated in test set will not.

Evaluation Protocol Since there are no explicit negative triples in existing knowledge graphs, we construct negative triples in validation and test set following the same protocol in (Socher et al. 2013). We also assure that the number of generated negative triples should be equal to that of positive triples. The classification is conducted as follows: we first learn different thresholds δ_r for each relation, which are optimized by maximizing the classification accuracies on validation set. In classification, if the energy function $||\mathbf{h} + \mathbf{r} - \mathbf{t}|| < \delta_r$, the triple will be classified to be positive, and otherwise to be negative.

Table 3: Evaluation results on triple classification

Datasets	FB15K-N1	FB15K-N2	FB15K-N3
TransE	81.3	79.4	76.9
CKRL (LT)	81.8	80.2	78.3
CKRL (LT+PP)	81.9	80.1	78.4
CKRL (LT+PP+AP)	81.7	80.2	78.3

Experimental Results Table 3 demonstrates the results of triple classification. We can find that: (1) The CKRL models outperform baseline on all datasets, and the improvements become more significant with higher noise rates. It confirms that learning knowledge representations with triple confidence could also help for triple classification. (2) The advantages confidence-aware models have over baseline in this task seem to be smaller than those in KG noise detection. It is because that the CKRL models concentrate more on calculating confidences for triples in training set, but not for negative triples generated in test set. The possible improvements for CKRL in triple classification can only derive from better-learned knowledge representations, which are less straightforward compared to that in KG noise detection. Although CKRL models learn better knowledge representations, conventional models without confidence may also achieve comparable results.

Conclusion and Future Work

In this paper, we propose a novel CKRL model which aims to detect noises in knowledge graphs and learn robust knowledge representations simultaneously. To make our models more flexible and universal, we only consider the internal structural information in KGs to define the local triple confidence and the global path confidence. We evaluate our models on KG noise detection, KG completion and triple classification. Experimental results indicate that CKRL can well capture both local and global structural information to measure triple confidences, which is essential when detecting noises in KGs and learning better knowledge representations. The utilization of triple confidence could also inspire the noise detection in real-world knowledge construction. The source code and dataset of this paper can be obtained from <https://github.com/thunlp/CKRL>.

We will explore the following research directions in future: (1) external information such as entity attributes and entity descriptions could provide supplementary information to judge triple confidence. We will explore to combine external heterogeneous information with internal structural information to better understand entities and relations. (2) We observe that the experimental results on knowledge graph noise detection are promising even with high noise rate datasets. In future, we will extend our confidence-aware framework to noise detection in knowledge construction, which could protect knowledge graphs away from noises and conflicts via global structural information in KGs.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (NSFC No. 61572273, 61532010, 61661146007), Tsinghua University Initiative Scientific Research Program (20151080406), and the research fund of Tsinghua University - Tencent Joint Laboratory for Internet Innovation Technology. We also thank Yankai Lin, Lixin Zhang and the THUNLP team for their constructive advices.

References

- [Auer et al. 2007] Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and Ives, Z. 2007. Dbpedia: A nucleus for a web of open data. In *The semantic web*. 722–735.
- [Bollacker et al. 2008] Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of KDD*, 1247–1250.
- [Bordes et al. 2012] Bordes, A.; Glorot, X.; Weston, J.; and Bengio, Y. 2012. Joint learning of words and meaning representations for open-text semantic parsing. In *Proceedings of AISTATS*, 127–135.
- [Bordes et al. 2013] Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of NIPS*, 2787–2795.
- [De Meo et al. 2012] De Meo, P.; Ferrara, E.; Fiumara, G.; and Ricciardello, A. 2012. A novel measure of edge centrality in social networks. *Knowledge-based systems* 30:136–150.
- [Dong et al. 2014] Dong, X.; Gabrilovich, E.; Heitz, G.; Horn, W.; Lao, N.; Murphy, K.; Strohmman, T.; Sun, S.; and Zhang, W. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of KDD*, 601–610.
- [Gyöngyi, Garcia-Molina, and Pedersen 2004] Gyöngyi, Z.; Garcia-Molina, H.; and Pedersen, J. 2004. Combating web spam with trustrank. In *Proceedings of VLDB*, 576–587.
- [Heindorf et al. 2015] Heindorf, S.; Potthast, M.; Stein, B.; and Engels, G. 2015. Towards vandalism detection in knowledge bases: Corpus construction and analysis. In *Proceedings of SIGIR*, 831–834.
- [Heindorf et al. 2016] Heindorf, S.; Potthast, M.; Stein, B.; and Engels, G. 2016. Vandalism detection in wikidata. In *Proceedings of CIKM*, 327–336.
- [Hoffart et al. 2013] Hoffart, J.; Suchanek, F. M.; Berberich, K.; and Weikum, G. 2013. Yago2: A spatially and temporally enhanced knowledge base from wikipedia. *Artificial Intelligence* 194:28–61.
- [Krompaß, Baier, and Tresp 2015] Krompaß, D.; Baier, S.; and Tresp, V. 2015. Type-constrained representation learning in knowledge graphs. In *Proceedings of ISWC*. 640–655.
- [Lehmann et al. 2015] Lehmann, J.; Isele, R.; Jakob, M.; Jentzsch, A.; Kontokostas, D.; Mendes, P. N.; Hellmann, S.; Morsey, M.; Van Kleef, P.; Auer, S.; et al. 2015. Dbpedia—a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web* 6(2):167–195.
- [Lin et al. 2015a] Lin, Y.; Liu, Z.; Luan, H.; Sun, M.; Rao, S.; and Liu, S. 2015a. Modeling relation paths for representation learning of knowledge bases. In *Proceedings of EMNLP*, 705–714.
- [Lin et al. 2015b] Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of AAAI*.
- [Lin et al. 2016] Lin, Y.; Shen, S.; Liu, Z.; Luan, H.; and Sun, M. 2016. Neural relation extraction with selective attention over instances. In *Proceedings of ACL*, volume 1, 2124–2133.
- [Manago and Kodratoff 1987] Manago, M., and Kodratoff, Y. 1987. Noise and knowledge acquisition. In *Proceedings of IJCAI*, 348–354.
- [Nickel et al. 2016] Nickel, M.; Murphy, K.; Tresp, V.; and Gabrilovich, E. 2016. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE* 104(1):11–33.
- [Pellissier Tanon et al. 2016] Pellissier Tanon, T.; Vrandečić, D.; Schaffert, S.; Steiner, T.; and Pintscher, L. 2016. From freebase to wikidata: The great migration. In *Proceedings of WWW*, 1419–1428.
- [Socher et al. 2013] Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. 2013. Reasoning with neural tensor networks for knowledge base completion. In *Proceedings of NIPS*, 926–934.
- [Wang et al. 2014] Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of AAAI*, 1112–1119.
- [Xie, Liu, and Sun 2016] Xie, R.; Liu, Z.; and Sun, M. 2016. Representation learning of knowledge graphs with hierarchical types. In *Proceedings of IJCAI*.
- [Zhou et al. 2007] Zhou, T.; Ren, J.; Medo, M.; and Zhang, Y.-C. 2007. Bipartite network projection and personal recommendation. *Physical Review E* 76(4):046115.