# Embedding Uncertain Knowledge Graphs

**Xuelu Chen, Muhao Chen, Weijia Shi, Yizhou Sun, Carlo Zaniolo**
University of California, Los Angeles
Los Angeles, CA, 90095, USA
{shirleychen, muhaochen, swj0419, yzsun, zaniolo}@cs.ucla.edu

## Abstract

Embedding models for deterministic Knowledge Graphs (KG) have been extensively studied, with the purpose of capturing latent semantic relations between entities and incorporating the structured knowledge they contain into machine learning. However, there are many KGs that model uncertain knowledge, which typically model the inherent uncertainty of relations facts with a confidence score, and embedding such uncertain knowledge represents an unresolved challenge. The capturing of uncertain knowledge will benefit many knowledge-driven applications such as question answering and semantic search by providing more natural characterization of the knowledge. In this paper, we propose a novel uncertain KG embedding model UKGE, which aims to preserve both structural and uncertainty information of relation facts in the embedding space. Unlike previous models that characterize relation facts with binary classification techniques, UKGE learns embeddings according to the confidence scores of uncertain relation facts. To further enhance the precision of UKGE, we also introduce probabilistic soft logic to infer confidence scores for unseen relation facts during training. We propose and evaluate two variants of UKGE based on different confidence score modeling strategies. Experiments are conducted on three real-world uncertain KGs via three tasks, i.e. confidence prediction, relation fact ranking, and relation fact classification. UKGE shows effectiveness in capturing uncertain knowledge by achieving promising results, and it consistently outperforms baselines on these tasks.

## 1 Introduction

Knowledge Graphs (KGs) provide structured representations of real-world entities and relations, which are categorized into the following two types: (i) *Deterministic KGs*, such as YAGO (Rebele et al. 2016) and FreeBase (Bollacker et al. 2008), consist of deterministic relation facts that describe semantic relations between entities; (ii) *Uncertain KGs* including ProBase (Wu et al. 2012), ConceptNet (Speer, Chin, and Havasi 2017) and NELL (Mitchell et al. 2018) associate every relation fact with a confidence score that represents the likelihood of the relation fact to be true.

KG embedding models are essential tools for incorporating the structured knowledge representations in KGs into machine learning. These models encode entities as low-dimensional vectors and relations as algebraic operations among entity vectors. They accurately capture the similarity of entities and preserve the structure of KGs in the embedding space. Hence, they have been the crucial feature models that benefit numerous knowledge-driven tasks (Bordes, Weston, and Usunier 2014; He et al. 2017; Das et al. 2018). Recently, extensive efforts have been devoted into embedding deterministic KGs. Translational models, e.g., TransE (Bordes et al. 2013) and TransH (Wang et al. 2014)), and bilinear models, e.g. DistMult (Yang et al. 2015) and ComplEx (Trouillon et al. 2016), have achieved promising performance in many tasks, such as link prediction (Yang et al. 2015; Trouillon et al. 2016), relation extraction (Weston, Bordes, and others 2013), relational learning (Nickel, Rosasco, and Poggio 2016), and ontology population (Chen et al. 2018).

While current embedding models focus on capturing deterministic knowledge, it is critical to incorporate uncertainty information into knowledge sources for several reasons. First, uncertainty is the nature of many forms of knowledge. An example of naturally uncertain knowledge is the interactions between proteins. Since molecular reactions are random processes, biologists label the protein interactions with their probabilities of occurrence and present them as uncertain KGs called Protein-Protein Interaction (PPI) Networks. Second, uncertainty enhances inference in knowledge-driven applications. For example, short text understanding often entails interpreting real-world concepts that are ambiguous or intrinsically vague. The probabilistic KG Probase (Wu et al. 2012) provides a prior probability distribution of concepts behind a term that has critically supported short text understanding tasks involving disambiguation (Wang et al. 2015; Wang and Wang 2016). Furthermore, uncertain knowledge representations have largely benefited various applications, such as question answering (Yih et al. 2013) and named entity recognition (Ratinov and Roth 2009).

Capturing the uncertainty information with KG embeddings remains an unresolved problem. This is a non-trivial task for several reasons. First, compared to deterministic KG embeddings, uncertain KG embeddings need to encode additional confidence information to preserve uncertainty. Second, current KG embedding models cannot capture the sub-

tle uncertainty of unseen relation facts, as they assume that all the unseen relation facts are false beliefs and minimize the plausibility measures of relation facts. One major challenge of learning embeddings for uncertain KGs is to properly estimate the uncertainty of unseen relation facts.

To address the above issues, we propose a new embedding model UKGE (Uncertain Knowledge Graph Embeddings), which aims to preserve both structural and uncertainty information of relation facts in the embedding space. Embeddings of entities and relations on uncertain KGs are learned according to confidence scores. Unlike previous models that characterize relation facts with binary classification techniques, UKGE learns embeddings according to the confidence scores of uncertain relation facts. To further enhance the precision of UKGE, we also introduce probabilistic soft logic to infer the confidence score for unseen relation facts during training. We propose two variants of UKGE based on different embedding-based confidence functions. We conducted extensive experiments using three real-world uncertain KGs on three tasks: (i) *confidence prediction*, which seeks to predict confidence scores of unseen relation facts; (ii) *relation fact ranking*, which focuses on retrieving tail entities for the query $(h, r, \underline{?t})$ and ranking these retrieved tails in the right order; and (iii) *relation fact classification*, which decides whether or not a given relation fact is a *strong* relation fact. Our models consistently outperform the baseline models in these experiments.

The rest of the paper is organized as follows. We first review the related work in Section 2, then provide the problem definition and propose our model UKGE in the two sections that follow. In section 5, we present our experiments. Then we conclude the paper in Section 6.

## 2 Related Work

To the best of our knowledge, there has been no previous work on learning embeddings for uncertain KGs. We hereby discuss the following three lines of work that are closely related to this topic.

**Deterministic Knowledge Graph Embeddings** Deterministic KG embeddings have been extensively explored by recent work. These models encode entities as low-dimensional vectors and relations as algebraic operations among entity vectors. There are two representative families of models, i.e. translational models and bilinear models.

Translational models share a common principle $\boldsymbol{h}_r + \boldsymbol{r} \approx \boldsymbol{t}_r$, where $\boldsymbol{h}_r$, $\boldsymbol{t}_r$ are the entity embeddings projected in a relation-specific space. The forerunner of this family, TransE (Bordes et al. 2013), lays $\boldsymbol{h}_r$ and $\boldsymbol{t}_r$ in a common space as $\boldsymbol{h}$ and $\boldsymbol{t}$ with regard to any relation $r$. Variants of TransE, such as TransH (Wang et al. 2014), TransR (Lin et al. 2015), TransD (Ji et al. 2015), and TransA, (Jia et al. 2016) differentiate the translations of entity embeddings in different language-specific embedded spaces based on different forms of relation-specific projections. Despite its simplicity, translational models achieve promising performance on knowledge completion and relation extraction tasks.

Bilinear models (Jenatton et al. 2012) model relations as the second-order correlations between entities, using the scoring function $f(h, r, t) = \boldsymbol{h}^T \boldsymbol{W_r} \boldsymbol{t}$. This function is first adopted by RESCAL (Nickel, Tresp, and Kriegel 2011), a collective matrix factorization model. DistMult (Yang et al. 2015) constrains $\boldsymbol{W_r}$ as a diagonal matrix which reduces the computing cost and also enhances the performance. ComplEx adjusts the corresponding scoring mechanism to a complex conjugation in a complex embedding space.

There are also other models for deterministic KG embedding, such as neural models like Neural Tensor Network (NTN) (Socher et al. 2013) and ConvE (Dettmers et al. 2018), and the circular-correlation-based model HolE (Nickel, Rosasco, and Poggio 2016).

**Uncertain Knowledge Graphs** Uncertain KGs provide a confidence score along with every relation fact. The development of relation extraction and crowdsourcing in recent years enabled the construction of large-scale uncertain knowledge bases. ConceptNet (Speer, Chin, and Havasi 2017) is a multilingual uncertain KG for commonsense knowledge that is collected via crowdsourcing. The confidence scores in ConceptNet mainly come from the co-occurrence frequency of the labels in crowdsourced task results. Probase (Wu et al. 2012) is a universal probabilistic taxonomy built by relation extraction. Every fact in Probase is associated with a joint probability $P_{isA}(x, y)$. NELL (Mitchell et al. 2018) collects relation facts by reading web pages and learns their confidence scores from semi-supervised learning with the Expectation-Maximum (EM) algorithm. Aforementioned uncertain KGs have enabled numerous knowledge-driven applications. For example, Wang and Wang (2016) utilize Probase to help understand short texts.

One recent work has proposed a matrix-factorization-based approach to embed uncertain networks (Hu et al. 2017). However, it cannot be generalized to embed uncertain KGs because this model only considers the node proximity in the networks with no explicit relations and only generates node embeddings. As far as we know, we are among the first to study the uncertain KG embedding problem.

**Probabilistic Soft Logic** Probabilistic soft logic (PSL) (Kimmig et al. 2012) is a framework for probabilistic reasoning. A PSL program consists of a set of first-order logic rules with conjunctive bodies and single literal heads. PSL takes the confidence from interval $[0, 1]$ as the *soft truth values* for every atom. It uses *Lukasiewics t-norm* (Lukasiewicz and Straccia 2008) to determine to which degree a rule is satisfied. In combination with Hinge-Loss Markov Random Field (HL-MRF), PSL is widely used in probabilistic reasoning tasks, such as social-trust prediction and preference prediction (Bach et al. 2013; 2017). In this paper, we adopt PSL to enhance the embedding model performance on the unseen relation facts.

## 3 Problem Definition

We define the uncertain KG embedding problem in this section by first providing the definition of uncertain KGs.

**Definition 1. Uncertain Knowledge Graph.** An uncertain KG represents knowledge as a set of relations ($\mathcal{R}$) defined

over a set of entities ($\mathcal{E}$). It consists of a set of weighted triples $\mathcal{G} = \{(l, s_l)\}$. For each pair $(l, s_l)$, $l = (h, r, t)$ is a triple representing a relation fact where $h, t \in \mathcal{E}$ (the set of entities) and $r \in \mathcal{R}$ (the set of relations), and $s_l \in [0, 1]$ represents the confidence score for this relation fact to be true.

Note that we assume the confidence score $s_l \in [0, 1]$ and interpret it as a probability to leverage probabilistic soft logic-based inference. The range of original confidence scores for some uncertain KG (e.g., ConceptNet) may not fall in $[0, 1]$, and normalization will be needed in these cases. Some examples of weighted triples are listed below.

**Example 3.1. Weighted triples.**

1. (choir, `relatedto`, sing): 1.00
2. (college, `synonym`, university): 0.99
3. (university, `synonym`, institute): 0.86
4. (fork, `atlocation`, kitchen): 0.4

**Definition 2. Uncertain Knowledge Graph Embedding Problem.** Given an uncertain KG $\mathcal{G}$, the embedding model aims to encode each entity and relation in a low-dimensional space in which structure information and confidence scores of relation facts are preserved.

Notation wise, boldfaced $\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}$ are used to represent the embedding vectors for head $h$, relation $r$ and tail $t$ respectively. $\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}$ are assumed lie in $\mathbb{R}^k$.

# 4 Modeling

In this section, we propose our model for uncertain KG embeddings. The proposed model `UKGE` encodes the KG structure according to the confidence scores for both *observed* and *unseen* relation facts, such that the embeddings of relation facts with higher confidence scores receive higher plausibility values.

We first design relation fact confidence score modeling based on embeddings of entities and relations, then introduce how probabilistic soft logic can be used to infer confidence scores for unseen relations, and lastly describe the joint model `UKGE` and its two variants.

## 4.1 Embedding-based Confidence Score Modeling for Relation Facts

Unlike deterministic KG embedding models, uncertain KG embedding models need to explicitly model the confidence score for each triple and compare the prediction with the true score. We hereby first define and model the plausibility of triples, which can be considered as a unnormalized confidence score.

**Definition 3. Plausibility.** Given a relation fact triple $l$, the plausibility $g(l) \in R$ measures how likely this relation fact holds. The higher plausibility value corresponds to the higher confidence score $s$.

Given a triple $l = (h, r, t)$ and their embeddings $\boldsymbol{h}, \boldsymbol{r}, \boldsymbol{t}$, we model the plausibility of $(h, r, t)$ by the following function:

$$g(l) = \boldsymbol{r} \cdot (\boldsymbol{h} \circ \boldsymbol{t}) \tag{1}$$

where $\circ$ is the element-wise product, and $\cdot$ is the inner product. This function captures the relatedness between embeddings $\boldsymbol{h}$ and $\boldsymbol{t}$ under the condition of relation $r$ and is first adopted by DistMult (Yang et al. 2015). We employ this triple modeling technique for three reasons: (i) This technique has represented the state-of-the-art performance for modeling deterministic KGs (Kadlec, Bajgar, and Kleindienst 2017), (ii) It agrees with the nature of our model to quantify the confidence of an uncertain relation fact by comparing the relation embeddings with the pair of head and tail embeddings, (iii) It does not introduce additional parameter complexity to the model like other techniques, such as TransH (Wang et al. 2014), TransR (Lin et al. 2015), ConvE (Dettmers et al. 2018) and ProjE (Shi and Weninger 2017). Nevertheless, this scoring function can be further explored in future work.

**From plausibility to confidence scores** In order to transform plausibility scores to confidence scores, we consider two different mapping functions and test them in the experimental section. Formally, let a triple be $l$ and its plausibility score be $g(l)$, a transformation function $\phi(\cdot)$ maps $g(l)$ to a confidence score $f(l)$.

$$f(l) = \phi(g(l)), \phi : \mathbb{R} \to [0, 1] \tag{2}$$

Two choices of mapping $\phi$ are listed below.
**Logistic function.** One way to map plausibility values to confidence score is a logistic function as follows:

$$\phi(x) = \frac{1}{1 + e^{-(\mathbf{w}x + \mathbf{b})}} \tag{3}$$

**Bounded rectifier.** Another mapping is a bounded rectifier (Chen et al. 2015):

$$\phi(x) = \min(\max(\mathbf{w}x + \mathbf{b}, 0), 1) \tag{4}$$

where $\mathbf{w}$ is a weight $\mathbf{b}$ is a bias.

## 4.2 PSL-based Confidence Score Reasoning for Unseen Relation Facts

In order to better estimate confidence scores, both observed and unseen relation facts in KGs should be utilized. Deterministic KG embedding methods assume that all unseen relation facts are false beliefs, and use negative sampling to add some of these false relations into training. One major challenge of learning embeddings for uncertain KGs, however, is to properly estimate the uncertainty of unseen triples, as simply treating their confidence score as 0 can no longer capture the subtle uncertainty. For example, it is common that a Protein-Protein Interaction Network KG may have no interaction records for two proteins that can be potentially binded. Ignoring such possibility will result in information loss.

We thus introduce probabilistic soft logic (PSL) (Kimmig et al. 2012) to infer confidence scores for these unseen relation facts to further enhance the embedding performance. PSL is a framework for confidence reasoning that propagates confidence of existing knowledge to unseen triples using soft logic.

**Probabilistic Soft Logic**   A PSL program consists of a set of first order logic rules that describe logical dependencies between facts (atoms). One example of logical rule is shown below:

**Example 4.1. A Logical Rule on Transitivity of Synonym Relation.**
(<u>A</u>,synonym,<u>B</u>) $\wedge$ (<u>B</u>,synonym,<u>C</u>) $\rightarrow$ (<u>A</u>,synonym,<u>C</u>)

This logical rule describes the transitivity of the relation synonym. In this logical rule, <u>A</u>, <u>B</u> and <u>C</u> are placeholders for entities, synonym is the predicate that corresponds to the relation in uncertain KGs, (<u>A</u>,synonym,<u>B</u>) $\wedge$ (<u>B</u>,synonym,<u>C</u>) is the body of the rule, and (<u>A</u>,synonym,<u>C</u>) is the head of the rule.

A logical rule serves as a template rule. By replacing the placeholders in a logical rule with concrete entities and relations, we can get rule instances, which are called *ground rules*. Considering Example 4.1 and uncertain relation facts from Example 3.1, we can have the following ground rule by replacing the placeholders with real relation facts in KG.

**Example 4.2. A Ground Rule on Transitivity of Synonym.**
(college, synonym, university) $\wedge$ (university, synonym, institute) $\rightarrow$ (college, synonym, institute)

Different from Boolean logic, PSL associates every atom, i.e., a triple $l$, with a *soft truth value* from the interval $[0, 1]$, which corresponds to the confidence score in our context and enables fuzzy reasoning. The assignment process of soft truth values is called an *interpretation*. We denote the soft truth value of an atom $l$ assigned by the interpretation $I$ as $I(l)$. Naturally, for observed relation facts, their observed confidence scores are used for assignment; and for unseen triples, the embedding-based estimated confidence scores will be assigned to them:

$$I(l) = s_l, l \in \mathcal{L}^+$$
$$I(l) = f(l), l \in \mathcal{L}^- \quad (5)$$

where $\mathcal{L}^+$ denotes the observed triple set, $\mathcal{L}^-$ denotes the unseen triples, $s_l$ denotes the confidence score for observed triple $l$, and $f(l)$ denotes the embedding-based confidence score function for $l$.

In PSL, Lukasiewicz t-norm is used to define the basic logical operations, including logical conjunction ($\wedge$), disjunction ($\vee$), and negation ($\neg$), as follows:

$$l_1 \wedge l_2 = \max\{0, I(l_1) + I(l_2) - 1\} \quad (6)$$
$$l_1 \vee l_2 = \min\{1, I(l_1) + I(l_2)\} \quad (7)$$
$$\neg l_1 = 1 - I(l_1) \quad (8)$$

For example, according to Eq. (6) and (7), $0.8 \wedge 0.3 = 0.1$ and $0.8 \vee 0.3 = 1$. For a rule $\gamma \equiv \gamma_{body} \rightarrow \gamma_{head}$, as it can be written as $\neg\gamma_{body} \vee \gamma_{head}$, its value $p_\gamma$ can be computed as

$$p_{\gamma_{body} \rightarrow \gamma_{head}} = \min\{1, 1 - I(\gamma_{body}) + I(\gamma_{head})\} \quad (9)$$

PSL regards a rule $\gamma$ as *satisfied* when the truth value of its head $I(\gamma_{head})$ is the same or higher than its body $I(\gamma_{body})$, i.e., when its value is greater than or equal to 1.

$$d_\gamma = 1 - p_\gamma = \max\{0, I(\gamma_{body}) - I(\gamma_{head})\} \quad (10)$$

Consider Example 4.2. Let (college, synonym, university) be $l_1$, (university, synonym, college) be $l_2$, and (college, synonym, institute) be $l_3$. Assuming that $l_1$ and $l_2$ are observed triples in KG, and $l_3$ is unseen, according to Equation (5), (6), and (9), the distance to satisfaction of this ground rule is calculated as below:

$$d_\gamma = \max\{0, I(l_1 \wedge l_2) - I(l_3)\}$$
$$= \max\{0, s_{l_1} + s_{l_2} - 1 - f(l_3)\}$$
$$= \max\{0, 0.85 - f(l_3)\}$$

where $s_{l_1}$ and $s_{l_2}$ are the ground truth confidence scores of corresponding relation facts in the uncertain KG.

This equation indicates that the ground rule in Example 4.2 is completely satisfied when $f(l_3)$, the estimated confidence score of (college, synonym institute), is above 0.85. When $f(l_3)$ is under 0.85, the smaller $f(l_3)$ is, the larger loss we have. In other words, a bigger confidence score is preferable. In the above example, we can see that the embedding-based confidence score for this unseen relation fact, $f(l_3)$, will affect the loss function, and it is desirable to learn embeddings that minimize these losses. Note that if we simply treat the unseen relation $l_3$ as false and use MSE (Mean Squared Error) as the loss, the loss would be $f(l_3)^2$, which is in favor of a lower confidence score mistakenly.

Moreover, we add a rule to penalize the predicted confidence scores of all unseen relation facts, which can be considered as a prior knowledge, i.e., any unseen relation fact has a low probability to be true. Formally, for an unseen relation fact $l = (h, r, t) \in \mathcal{L}^-$, we have a ground rule $\gamma_0$:

$$\gamma_0 : \neg l \quad (11)$$

According to Eq. (8) and (10), its distance to satisfaction $d_{\gamma_0}$ is derived as:

$$d_{\gamma_0} = f(l) \quad (12)$$

### 4.3   Embedding Uncertain KGs

In this subsection, we present the objective function of uncertain KG embeddings.

**Loss on observed relation facts**   Let $\mathcal{L}^+$ be the set of observed relation facts, the goal is to minimize the mean squared error (MSE) between the ground truth confidence score $s_l$ and our prediction $f(l)$ for each relation $l \in \mathcal{L}^+$:

$$\mathcal{J}^+ = \sum_{l \in \mathcal{L}^+} |f(l) - s_l|^2 \quad (13)$$

**Loss on unseen relation facts**   Let $\mathcal{L}^-$ be the sampled set of unseen relations, and $\Gamma_l$ be the set of ground rules with $l$ as the rule head, the goal is to minimize the distance to rule satisfaction for each triple $l$. In particular, we choose to use the square of the distance as the following loss (Bach et al. 2013):

$$\mathcal{J}^- = \sum_{l \in \mathcal{L}^-} \sum_{\gamma \in \Gamma_l} |\psi_\gamma(f(l))|^2 \quad (14)$$

where $\psi_\gamma(f(l))$ denotes the weighted distance to satisfaction $w_\gamma d_\gamma$ of the rule $\gamma$ as a function of $f(l)$ where $w_\gamma$ is a hand-crafted weight for the rule $\gamma$.

Note that when $l$ is only covered by $\gamma_0 : \neg l$, we have $\sum_{\gamma \in \Gamma_l} |\psi_\gamma(f(l))|^2 = |f(l)|^2$, which is essentially the MSE loss by treating unseen relation facts as false.

| Dataset | #Ent. | #Rel. | #Rel. Facts | Avg($s$) | Std($s$) |
|---|---|---|---|---|---|
| CN15k | 15,000 | 36 | 241,158 | 0.629 | 0.232 |
| NL27k | 27,221 | 404 | 175,412 | 0.797 | 0.242 |
| PPI5k | 4,999 | 7 | 271,666 | 0.415 | 0.213 |

Table 1: Statistics of the extracted datasets used in this paper. *Ent.* denotes entities and *Rel.* stands for relations. Avg($s$) and Std($s$) are the average and standard deviation of the confidence scores.

**The Joint Objective Function** Combining Eq. (13) and (14), we obtain the following joint objective function:

$$\mathcal{J} = \sum_{l \in \mathcal{L}^+} |f(l) - s_l|^2 + \sum_{l \in \mathcal{L}^-} \sum_{\gamma \in \Gamma_l} |\psi_\gamma(f(l))|^2 \quad (15)$$

Similar to deterministic KG embedding algorithms, we sample unseen relations by corrupting the head and the tail for observed relation facts to generate $\mathcal{L}^-$ during training.

We give two model variants that differ in the choice of $f(l)$. We refer to the variant that adopts Equation (3) as UKGE$_{logi}$ and name the one using Equation (4) as UKGE$_{rect}$.

# 5 Experiments

In this section, we evaluate our models on three tasks: confidence prediction, relation fact ranking, and relation fact classification.

## 5.1 Datasets

The evaluation is conducted on three datasets named as CN15k, NL27k, and PPI5k, which are extracted from ConceptNet, NELL, and the Protein-Protein Interaction Knowledge Base STRING (Szklarczyk et al. 2016) respectively. CN15k matches the number of nodes with FB15k (Bordes et al. 2013) - the widely used benchmark dataset for deterministic KG embeddings (Bordes et al. 2013; Wang et al. 2014; Yang et al. 2015), while NL27k is a larger dataset. PPI5k is a denser graph with fewer entities but more relation facts than the other two. Table 1 gives the statistics of the datasets, and more details are introduced below.

**CN15k** CN15k is a subgraph of the commonsense KG ConceptNet. This subgraph contains 15,000 entities and 241,158 uncertain relation facts in English. The original scores in ConceptNet vary from 0.1 to 22, where 99.6% are less than or equal to 3.0. For normalization, we first bound confidence scores to $x \in [0.1, 3.0]$, and then applied the min-max normalization on $\log x$ to map them into $[0.1, 1.0]$.

**NL27** NL27k is extracted from NELL (Mitchell et al. 2018), an uncertain KG obtained from webpage reading. NL27k contains 27,221 entities, 404 relations, and 175,412 uncertain relation facts. In the process of min-max normalization, we search for the lower boundary from 0.1 to 0.9. We have found out that normalizing the confidence score to interval $[0.1, 1]$ yields best results.

**PPI5k** The Protein-Protein Interaction Knowledge Base STRING labels the interactions between proteins with the probabilities of occurrence. PPI5k is a subset of STRING

that contains 271,666 uncertain relation facts for 4,999 proteins and 7 interactions.

In an uncertain KG, a relation fact is considered *strong* if its confidence score $s_l$ is above a KG-specific threshold $\tau$. Here we set $\tau = 0.85$ for both CN15k and NL27k. We follow the instructions from (Szklarczyk et al. 2016) and set $\tau = 0.70$ for PPI5k. Under this setting, 20.4% of relation facts in CN15k, 20.1% of those in NL27k, and 12.4% of those in PPI5k are considered strong.

## 5.2 Experimental Setup

We split each dataset into three parts: 85% for training, 7% for validation, and 8% for testing. To test if our model can correctly interpret negative links, we add the same amount of negative links as existing relation facts into the test sets.

We use Adam optimizer (Kingma and Ba 2014) for training, for which we set the exponential decay rates $\beta_1 = 0.9$ and $\beta_2 = 0.99$. We report results for all models respectively based on their best hyperparameter settings. For each model, the setting is identified based on the validation set performance. We select among the following sets of hyperparameter values: learning rate $lr \in \{0.001, 0.005, 0.01\}$, dimensionality $k \in \{64, 128, 256, 512\}$, batch size $b \in \{128, 256, 512, 1024\}$, The $L_2$ regularization coefficient $\lambda$ is fixed as 0.005. Training was stopped using early stopping based on MSE on the validation set, computed every 10 epochs. The best hyper-parameter combinations on CN15k and NL27k are $\{lr = 0.001, k = 128\}$ and $b = 128$ for UKGE$_{rect}$, $b = 512$ for UKGE$_{logi}$. On PPI5k they are $\{lr = 0.001, k = 128, b = 256\}$ for both variants.

## 5.3 Logical Rule Generation

Our model requires additional input as logical rules for PSL reasoning. We heuristically create candidate logical rules by considering length-2 paths (i.e., $(E_1, R_1, E_2) \wedge (E_2, R_2, E_3) \rightarrow (E_1, R_3, E_3)$) and validate them by *hit ratio,* i.e. the proportion of relation facts implied by the rule to be truly existent in the KG. The higher ratio implies that the rule is more convincing. When grounding out the logical rules, to guarantee the quality of the ground rules, we only adopt observed strong relation facts in our rule body. We eventually create 3 logical rules for CN15k, 4 for NL27k, and 1 for PPI5k. Table 2 gives some examples of the logical rules and their hit ratios. How to systematically create more promising logical rules will be considered as future work.

## 5.4 Baselines

Three types of baselines are considered in our comparison, which include (i) deterministic KG embedding models TransE (Bordes et al. 2013), DistMult (Yang et al. 2015) and ComplEx (Trouillon et al. 2016), (ii) an uncertain graph embedding model URGE (Hu et al. 2017), and (iii) UKGE$_{n-}$ and UKGE$_{p-}$ that are two simplified versions of our model.

- Deterministic KG Embedding Models. TransE, DistMult, and ComplEx have demonstrated high performance on deterministic KGs. Only the high-confidence relation facts from KGs are used for training. For each KG, we

| Dataset | Logical Rules | Hit Ratio |
|---|---|---|
| CN15k | (A, relatedTo, B)∧(B, relatedTo, C)→(A, relatedTo, C) | 37.0% |
| | (A, causes, B)∧(B, causes, C)→(A, causes, C) | 35.6% |
| NL27k | (A, atheletePlaysForTeam,B) ∧ (A, athletePlaysSport, C) →(B, teamPlaysSport, C) | 42.9% |
| PPI5k | (A, binding, B)∧(B, binding, C)→(A, binding, C) | 80.8% |

Table 2: Examples of logical rules. *Hit ratio* means the proportion of relation facts that have already existed in the KG

| Dataset | CN15k | | NL27k | | PPI5k | |
|---|---|---|---|---|---|---|
| Metrics | MSE | MAE | MSE | MAE | MSE | MAE |
| URGE | 10.32 | 22.72 | 7.48 | 11.35 | 1.44 | 6.00 |
| $\text{UKGE}_{n-}$ | 23.96 | 30.38 | 24.86 | 36.67 | 7.46 | 19.32 |
| $\text{UKGE}_{p-}$ | 9.02 | 20.05 | 2.67 | 7.03 | 0.96 | 4.09 |
| $\text{UKGE}_{rect}$ | **8.61** | **19.90** | **2.36** | **6.90** | **0.95** | **3.79** |
| $\text{UKGE}_{logi}$ | 9.86 | 20.74 | 3.43 | 7.93 | 0.96 | 4.07 |

Table 3: Mean squared error (MSE) and mean absolute error (MAE) of relation fact confidence prediction ($\times 10^{-2}$).

| metrics | CN15K | | NL27k | | PPI5k | |
|---|---|---|---|---|---|---|
| Dataset | linear | exp. | linear | exp. | linear | exp. |
| TransE | 0.601 | 0.591 | 0.730 | 0.722 | 0.710 | 0.700 |
| DistMult | 0.689 | 0.677 | 0.911 | 0.897 | 0.894 | 0.880 |
| ComplEx | 0.723 | 0.712 | 0.921 | 0.913 | 0.896 | 0.881 |
| URGE | 0.572 | 0.570 | 0.593 | 0.593 | 0.726 | 0.723 |
| $\text{UKGE}_{n-}$ | 0.236 | 0.232 | 0.245 | 0.245 | 0.514 | 0.517 |
| $\text{UKGE}_{p-}$ | 0.769 | 0.768 | 0.933 | 0.929 | 0.940 | 0.944 |
| $\text{UKGE}_{rect}$ | 0.773 | 0.775 | 0.939 | 0.942 | 0.946 | 0.946 |
| $\text{UKGE}_{logi}$ | **0.789** | **0.788** | **0.955** | **0.956** | **0.970** | **0.969** |

Table 4: Mean normalized DCG for global ranking task. Here *linear* stands for linear gain, and *exp.* stands for exponential gain.

have a KG-specific confidence score threshold $\tau$ to distinguish the high-confidence relation facts from the low-confidence ones, which will be discussed later in Section 5.7. These models cannot predict confidence scores. We compare our methods to them only on the ranking and the classification tasks. For the same reason, the early stopping is based on mean reciprocal rank (MRR) on the validation set. We adopt the implementation given by (Trouillon et al. 2016) and choose the best hyper-parameters following the same grid search procedure. This implementation uses (Duchi, Hazan, and others 2011) for optimization. The best hyper-parameter combinations on CN15k and NL27k are $b = 1024$, $\{lr = 0.01, k = 128\}$ for TransE and $\{lr = 0.05, k = 256\}$ for DistMult and ComplEx. On PPI5k they are $lr = 0.1$, $\{k = 128, b = 512\}$ for DistMult and $\{k = 256, b = 1024\}$ for TransE and ComplEx.

- Uncertain Graph Embedding Model. URGE is proposed very recently to embed uncertain graphs. However, it cannot deal with multiple types of relations in KGs, and it only produces node embeddings. We simply ignore relation types when applying URGE to our datasets. We adopt its first-order proximity version as our tasks focus on the edge relations between nodes.

- Two Simplified Versions of Our Model. To justify the

use of negative links and PSL reasoning in our model, we propose two simplified versions of $\text{UKGE}_{rect}$, called $\text{UKGE}_{n-}$ and $\text{UKGE}_{p-}$. In $\text{UKGE}_{n-}$, we only keep the observed relation facts and remove negative sampling, and in $\text{UKGE}_{p-}$, we remove PSL reasoning and use the MSE loss for unseen relation facts.

### 5.5 Confidence Prediction

The objective of this task is to predict confidence scores of unseen relation facts.

**Evaluation protocol** For each uncertain relation fact $(l, s_l)$ in the test set, we predict the confidence score of $l$ and report the mean squared error (MSE) and mean absolute error (MAE).

**Results** Results are reported in Table 3. Both our variants $\text{UKGE}_{rect}$ and $\text{UKGE}_{logi}$ outperform the baselines URGE, $\text{UKGE}_{n-}$, and $\text{UKGE}_{p-}$, since URGE only takes node proximity information and cannot model the rich relations between entities, and $\text{UKGE}_{n-}$ does not adopt negative sampling and cannot recognize negative links. The better results of $\text{UKGE}_{rect}$ than $\text{UKGE}_{p-}$ demonstrate that introducing PSL into embedding learning can enhance the model performance. Between the two model variants, $\text{UKGE}_{rect}$ results in smaller MSE and MAE than $\text{UKGE}_{logi}$. We notice that all the models achieve much smaller MSE on PPI5k than CN15k and NL27k. We hypothesize that this is because the much higher density of PPI5k facilitates embedding learning (Pujara, Augustine, and Getoor 2017).

### 5.6 Relation Fact Ranking

The next task focuses on ranking tail entities in the right order for the query $(h, r, \underline{?t})$.

**Evaluation protocol** For a query $(h, r, \underline{?t})$, we rank all the entities in the vocabulary as tail candidates and evaluate the ranking performance using the normalized Discounted Cumulative G ain (nDCG) (Li, Liu, and Zhai 2009). We define the gain in retrieving a relevant tail $t_0$ as the ground truth confidence score $s_{(h,r,t_0)}$. We take the mean nDCG over the test query set as our ranking metric. We report the two versions of nDCG that use linear gain and exponential gain respectively. The exponential gain version puts stronger emphasis on highly relevant results.

**Results** Table 4 shows the mean nDCG over all test queries for all compared methods. Though TransE, DistMult, and ComplEx do not encode the confidence score information, they maximize the plausibility of all observed relation facts and therefore rank these existing relation facts

| Dataset | head | relation | true tail | confidence | predicted tail | predicted confidence | true confidence |
|---|---|---|---|---|---|---|---|
| CN15k | rush | relatedto | fast | 0.968 | fast | 0.703 | 0.968 |
| | | | motion | 0.709 | move | 0.623 | 0.557 |
| | | | rapid | 0.709 | hour | 0.603 | 0.654 |
| | | | urgency | 0.709 | time | 0.601 | 0.105 |
| | hotel | usedfor | sleeping | 1.0 | relaxing | 0.858 | N/A |
| | | | rest | 0.984 | sleeping | 0.849 | 1.0 |
| | | | bed away from home | 0.709 | rest | 0.827 | 0.984 |
| | | | stay overnight | 0.709 | hotel room | 0.797 | N/A |
| NL27k | Toyota | competeswith | Honda | 1.0 | Honda | 0.942 | 1.0 |
| | | | Ford | 1.0 | Hyundai | 0.910 | 0.719 |
| | | | BMW | 0.964 | Chrysler | 0.908 | N/A |
| | | | General Motors | 0.930 | Nissan | 0.896 | 0.859 |

Table 5: Examples of relation fact ranking (global) results using $\mathtt{UKGE}_{logi}$. Top 4 results are shown. N/A denotes relation facts that are not observed in KG.

| Metrics | CN15k | | NL27k | | PPI5k | |
|---|---|---|---|---|---|---|
| Dataset | F-1 | Accu. | F-1 | Accu. | F-1 | Accu. |
| TransE | 23.4 | 67.9 | 65.1 | 53.4 | 83.2 | 98.5 |
| DistMult | 27.9 | 71.1 | 72.1 | 70.1 | 86.9 | 97.1 |
| ComplEx | 18.9 | 73.2 | 63.3 | 53.4 | 83.2 | 98.9 |
| URGE | 21.2 | 86.0 | 83.6 | 88.7 | 85.2 | 98.6 |
| $\mathtt{UKGE}_{n-}$ | 23.6 | 86.1 | 64.4 | 65.5 | 92.7 | 99.3 |
| $\mathtt{UKGE}_{p-}$ | 26.2 | 88.7 | 89.7 | 93.4 | 94.2 | 99.3 |
| $\mathtt{UKGE}_{rect}$ | **28.8** | **90.4** | **92.3** | **95.2** | **95.1** | 99.4 |
| $\mathtt{UKGE}_{logi}$ | 25.9 | 90.1 | 88.4 | 93.0 | 94.5 | **99.5** |

Table 6: F-1 scores (%) and accuracies (%) of relation fact classification

high. We observe that DistMult and ComplEx have considerably better performance than TransE, as TransE does not handle `1-to-N` relations well. ComplEx embeds entities and relations in the complex domain and handles asymmetric relations better than DistMult. It achieves the best results among the deterministic KG embedding models on this task. As $\mathtt{UKGE}_{n-}$ removes negative sampling from the loss function, it cannot distinguish the negative links from existing relation facts and results in the worst performance. $\mathtt{UKGE}_{p-}$ yields slightly worse performance than $\mathtt{UKGE}_{rect}$. Besides ranking the existing relation facts highly, our models also preserve the order of the observed relation facts and thus achieve higher nDCG scores. Both $\mathtt{UKGE}_{rect}$ and $\mathtt{UKGE}_{logi}$ outperform all the baselines under all settings, while $\mathtt{UKGE}_{logi}$ yields higher nDCG on all three datasets than $\mathtt{UKGE}_{rect}$. Considering the confidence prediction results of $\mathtt{UKGE}_{logi}$ in Section 5.5, we hypothesize that the easy saturation of logistic function allows $\mathtt{UKGE}_{logi}$ to better distinguish negative links from true relation facts, while this feature compromises its ability to fit confidence scores more precisely.

**Case study** Table 5 gives some examples of relation fact ranking results by $\mathtt{UKGE}_{logi}$. Given a query $(h, r, \underline{?t})$, the top 4 predicted tails and true tails are given, sorted by their scores in descending order. The predictions are consistent with our common-sense. It is worth noting that some quite reasonable unseen relation facts such as *hotel is used for re-*

*laxing*, can be predicted correctly. In other words, our proposed approach can be potentially used to infer new knowledge from the observed ones with reasonable confidence scores, which may shed light on another line of future study.

### 5.7 Relation Fact Classification

This last task is a binary classification task to decide whether a given relation fact $l$ is a *strong* relation fact or not. A relation fact is considered *strong* if its confidence score $s_l$ is above a KG-specific threshold $\tau$. The embedding models need to distinguish relation facts in the KG from negative links and high-confidence relation facts from low-confidence ones.

**Evaluation protocol** We follow a procedure that is similar to (Wang et al. 2014). Our test set consists of relation facts from the KG and randomly sampled negative links equally. We divide the test cases into two groups, strong and weak/false, by their ground truth confidence scores. A test relation fact $l$ is strong when $l$ is in the KG and $s_l > \tau$, otherwise weak/false. We fit a logistic regression classifier as a downstream classifier on the predicted confidence scores.

**Results** F-1 scores and accuracies are reported in Table 6. These results show that our two model variants consistently outperform all baseline models. The deterministic KG models can distinguish the existing relation facts from negative links, but they do not leverage the confidence information and cannot recognize the high-confidence ones. URGE does not encode the rich relations. Although $\mathtt{UKGE}_{n-}$ fits confidence scores in the KG, it cannot correctly interpret negative links as false. Consistent with the previous two tasks, the performance of $\mathtt{UKGE}_{p-}$ is worse than $\mathtt{UKGE}_{rect}$.

## 6 Conclusion and Future Work

To the best of our knowledge, this paper is the first work on embedding uncertain knowledge graphs. Our model $\mathtt{UKGE}$ effectively preserves both the relation facts and uncertainty information in the embedding space of KG. We propose two variants of our model and conduct extensive experiments on relation fact confidence score prediction, relation fact ranking, and relation fact classification. The results are very

promising. For future work, we will study how to systematically generate reasonable logical rules and test its impact on embedding quality. We are interested in extending `UKGE` for uncertain knowledge extraction from text.

## Acknowledgements

# References

Bach, S.; Huang, B.; London, B.; and Getoor, L. 2013. Hinge-loss markov random fields: Convex inference for structured prediction. *Uncertainty in Artificial Intelligence*.

Bach, S. H.; Broecheler, M.; Huang, B.; and Getoor, L. 2017. Hinge-loss markov random fields and probabilistic soft logic. *JMLR*.

Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In *SIGMOD*.

Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. Translating embeddings for modeling multi-relational data. In *NIPS*.

Bordes, A.; Weston, J.; and Usunier, N. 2014. Open question answering with weakly supervised embedding models. In *ECML-PKDD*. Springer.

Chen, M.; Weinberger, K. Q.; Xu, Z.; and Sha, F. 2015. Marginalizing stacked linear denoising autoencoders. *JMLR*.

Chen, M.; Tian, Y.; Chen, X.; Xue, Z.; and Zaniolo, C. 2018. On2vec: Embedding-based relation prediction for ontology population. In *SDM*.

Das, R.; Dhuliawala, S.; Zaheer, M.; Vilnis, L.; Durugkar, I.; Krishnamurthy, A.; Smola, A.; and McCallum, A. 2018. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *ICLR*.

Dettmers, T.; Minervini, P.; Stenetorp, P.; and Riedel, S. 2018. Convolutional 2d knowledge graph embeddings. In *AAAI*.

Duchi, J.; Hazan, E.; et al. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *JMLR* 12(Jul).

He, H.; Balakrishnan, A.; Eric, M.; and Liang, P. 2017. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings. In *ACL*.

Hu, J.; Cheng, R.; Huang, Z.; Fang, Y.; and Luo, S. 2017. On embedding uncertain graphs. In *CIKM*.

Jenatton, R.; Roux, N. L.; Bordes, A.; and Obozinski, G. R. 2012. A latent factor model for highly multi-relational data. In *NIPS*.

Ji, G.; He, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. Knowledge graph embedding via dynamic mapping matrix. In *ACL*.

Jia, Y.; Wang, Y.; Lin, H.; Jin, X.; and Cheng, X. 2016. Locally adaptive translation for knowledge graph embedding. In *AAAI*, 992–998.

Kadlec, R.; Bajgar, O.; and Kleindienst, J. 2017. Knowledge base completion: Baselines strike back. In *ACL*.

Kimmig, A.; Bach, S.; Broecheler, M.; Huang, B.; and Getoor, L. 2012. A short introduction to probabilistic soft logic. In *Proceedings of the NIPS Workshop on Probabilistic Programming: Foundations and Applications*, 1–4.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *ICLR*.

Li, H.; Liu, T.-Y.; and Zhai, C. 2009. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval*.

Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*.

Lukasiewicz, T., and Straccia, U. 2008. Managing uncertainty and vagueness in description logics for the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*.

Mitchell, T.; Cohen, W.; Hruschka, E.; Talukdar, P.; Yang, B.; Betteridge, J.; Carlson, A.; Dalvi, B.; Gardner, M.; Kisiel, B.; et al. 2018. Never-ending learning. *Communications of the ACM*.

Nickel, M.; Rosasco, L.; and Poggio, T. 2016. Holographic embeddings of knowledge graphs. In *AAAI*.

Nickel, M.; Tresp, V.; and Kriegel, H.-P. 2011. A three-way model for collective learning on multi-relational data. In *ICML*.

Pujara, J.; Augustine, E.; and Getoor, L. 2017. Sparsity and noise: Where knowledge graph embeddings fall short. In *EMNLP*.

Ratinov, L., and Roth, D. 2009. Design challenges and misconceptions in named entity recognition. In *CoNLL*.

Rebele, T.; Suchanek, F.; Hoffart, J.; Biega, J.; Kuzey, E.; and Weikum, G. 2016. Yago: A multilingual knowledge base from wikipedia, wordnet, and geonames. In *ISWC*.

Shi, B., and Weninger, T. 2017. Proje: Embedding projection for knowledge graph completion. In *AAAI*.

Socher, R.; Chen, D.; Manning, C. D.; and Ng, A. 2013. Reasoning with neural tensor networks for knowledge base completion. In *NIPS*.

Speer, R.; Chin, J.; and Havasi, C. 2017. Conceptnet 5.5: An open multilingual graph of general knowledge. In *AAAI*.

Szklarczyk, D.; Morris, J. H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N. T.; Roth, A.; Bork, P.; et al. 2016. The string database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic acids research*.

Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, E.; and Bouchard, G. 2016. Complex embeddings for simple link prediction. In *ICML*.

Wang, Z., and Wang, H. 2016. Understanding short texts. In *ACL*.

Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. Knowledge graph embedding by translating on hyperplanes. In *AAAI*.

Wang, Z.; Wang, H.; Wen, J.-R.; and Xiao, Y. 2015. An inference approach to basic level of categorization. In *CIKM*.

Weston, J.; Bordes, A.; et al. 2013. Connecting language and knowledge bases with embedding models for relation extraction. In *EMNLP*.

Wu, W.; Li, H.; Wang, H.; and Zhu, K. Q. 2012. Probase: A probabilistic taxonomy for text understanding. In *SIGMOD*.

Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2015. Embedding entities and relations for learning and inference in knowledge bases. *ICLR*.

Yih, W.-t.; Chang, M.-W.; Meek, C.; and Pastusiak, A. 2013. Question answering using enhanced lexical semantic models. In *ACL*.