UPGAT: Uncertainty-Aware Pseudo-neighbor Augmented Knowledge Graph Attention Network

Yen-Ching Tseng[1,2(✉)], Zu-Mu Chen[1], Mi-Yen Yeh[1], and Shou-De Lin[2]

[1] Institute of Information Science, Academia Sinica, Taipei, Taiwan
franklyn.chen@gmail.com, miyen@iis.sinica.edu.tw
[2] Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan
{r0822a10,sdlin}@csie.ntu.edu.tw

Abstract. The uncertain knowledge graph (UKG) generalizes the representation of entity-relation facts with a certain confidence score. Existing methods for UKG embedding view it as a regression problem and model different relation facts independently. We aim to generalize the graph attention network and use it to capture the local structural information. Yet, the uncertainty brings in excessive irrelevant neighbor relations and complicates the modeling of multi-hop relations. In response, we propose UPGAT, an uncertainty-aware graph attention mechanism to capture the probabilistic subgraph features while alleviating the irrelevant neighbor problem; introduce the pseudo-neighbor augmentation to extend the attention range to multi-hop. Experiments show that UPGAT outperforms the existing methods. Specifically, it has more than 50% Weighted Mean Rank improvement over the existing approaches on the NL27K dataset.

Keywords: uncertain knowledge graph · embedding · graph attention

1 Introduction

Knowledge Graph (KG) embedding has been widely studied in recent years, allowing machine learning models to leverage structural knowledge. As a generalized form, *Uncertain Knowledge Graphs* (UKG) no longer simply consider the existence of relational facts. Instead, they also express the corresponding plausibility. For example, the occurrence of protein interactions is probabilistic; if two words are synonymous is also probabilistic due to lexical ambiguity.

All existing uncertain knowledge graph embedding methods, including [2–4,9,13], model each relation fact independently. To be more concrete, the model samples one single relation fact, termed *triplet*, at a time, and predicts its confidence score. The simplicity of these methods helps avoid the caveat of overfitting. However, the probabilistic nature of UKG complicates the structural

This work was supported in part by National Science and Technology Council, Taiwan, R.O.C., under grants 110-2628-E-001-001 and 111-2628-E-001-001-MY2.

H. Kashima et al. (Eds.): PAKDD 2023, LNAI 13936, pp. 53–65, 2023.
https://doi.org/10.1007/978-3-031-33377-4_5

information and also makes the graph denser since there can be many extra uncertain relations between each entity pair. Therefore, we claim modeling subgraph information is indispensable for advancing UKG embedding quality.

Among all methods that incorporate subgraph information on embedding deterministic KG, graph attention network (GAT) [12] can aggregate neighboring triplets and use the attention mechanism to weigh each triplet based on their importance. This feature is highly desirable for uncertain and dense graphs, as it can filter out implausible or irrelevant neighbors on an as-needed basis.

Nonetheless, present GAT methods such as [14] are designed for the deterministic KG only. They cannot be directly applied to UKG due to the following challenges we aim to tackle in this study. First, it is challenging to retain the uncertainty information of the subgraph while filtering out unrelated ones. We presume that many neighboring triplets surrounding a given triplet can be uninformative for embedding such a triplet due to the uncertainty and high density of the uncertain KG. We show in the experiment section that, in reality, the plausibility does not imply relevance - it is groundless to focus on the more plausible neighbors trivially. Second, unlike deterministic KGs where the existence of a path is a binary question, it is difficult to determine the confidence score of a multi-hop relation without predefined inference rules.

To address these challenges, the conventional practice will devise a dedicated model structure that embeds confidence scores. In contrast, we take an innovative perspective to encode the uncertainty information implicitly and reach an even better performance proven in experiments. Our contributions are: (i) our proposed model, *UPGAT*, is the first to incorporate subgraph features and generalize GAT for the uncertain KG, (ii) rather than formulating the confidence score for neighbor triplets, the proposed graph attention with *the attention baseline mechanism* can catch the uncertainty while distinguishing unrelated information, and (iii) without using rule-based inferences, *the pseudo-neighbor augmented graph attention* overcomes the difficulty of identifying and leveraging multi-hop relations, where we add predicted n-hop neighbors, termed pseudo-neighbors, into the neighbor aggregation mechanism.

Experiments on three public datasets depict that our proposed method is superior to the existing technology on both the link prediction and confidence score prediction tasks. Specifically, UPGAT shows more than 50% Weighted Mean Rank (WMR) improvement on the NL27K dataset.

2 Preliminaries

2.1 Problem Statement

Definition 1 (Uncertain knowledge graph). *Let $\mathcal{G}$ be an uncertain knowledge graph, such that $\mathcal{G} = \{(l, s_l)|h, t \in \mathcal{E}, r \in \mathcal{R}, s_l \in [0, 1]\}$, where $l = (h, r, t)$ is a triplet containing head and tail entities h and t with relation r; $\mathcal{E}$ is the entity set, $\mathcal{R}$ is the relation set, and s_l is the confidence score of triplet l.*

Definition 2 (Negative and positive samples). *An uncertain relation fact sample,* (l, s_l)*, is a negative sample if* $s_l = 0$ *and is a positive sample otherwise.*

Note that as existing datasets only include positive samples, most studies randomly draw out-of-dataset triplets as negative samples. If the true confidence score of such a sample is non-zero, it is termed as a *false-negative triplet* [4].

Given an uncertain knowledge graph $\mathcal{G}$, the *Uncertain Knowledge Graph Embedding* problem aims to encode each entity and relation in a low-dimensional space while preserving the probabilistic graph structure. Note that the deterministic KG is a special case of UKG, where the confidence scores are either one or zero. Hence, applying deterministic KG embedding methods to UKG can be incompatible and/or have degraded performance. Moreover, the inherent higher density of uncertain KG could lead to more false-negative triplets.

Most existing embedding methods for uncertain knowledge graphs [3,4,13] follow the paradigm of traditional KG embedding methods to estimate the confidence score of a single triplet, $l = (h, r, t)$, using a parameterized score function $S(\hat{h}, \hat{r}, \hat{t})$, where $\hat{h}$, $\hat{r}$, and $\hat{t}$ are the respective embedding vectors. As an example, UKGE [3] applies the DistMult [15] scoring function for S and uses Mean Square Error (MSE) loss to learn the confidence score s_l as a regression model: $\mathcal{J}^+ = ||S(\bar{h}, \bar{r}, \bar{t}) - s_l||^2$.

Definition 3 (Subgraph Features). *Given an uncertain relation fact,* (l, s_l)*, where* $l = (h, r, t)$*, let the subgraph feature be any connected subgraph* $G' \subset G$*, where* $(l, s_l) \in G'$ *and* $|G'| > |\{(l, s_l)\}|$.

The above mentioned methods for embedding uncertain KGs model different relation facts independently. Nonetheless, studies on deterministic KGs, such as [6,7,11,12,14], have shown the benefit for node classification and link prediction to aggregate the neighboring triplets around an entity. We believe the complex and continuous graph structure of uncertain KGs makes it even more essential to model subgraph features explicitly.

2.2 Motivations and Challenges

To exploit subgraph features, we found graph attention network (GAT)-based method is well-suitable yet has not been studied for UKGs. Concerning the probabilistic nature of UKGs, where a triplet can be surrounded by implausible neighboring relation facts, GAT can learn the importance of each neighboring relation fact and assigns different attention scores accordingly. Other ways to model multi-hop relationships are subject to various limitations, which we have more discussions about in Sect. 2.3.

KBGAT [14] is a variant of graph attention network (GAT). Different from GAT that only considers the node features in the graph, KBGAT can encode the edge (relation) features in the knowledge graph. Given an entity h_i and one of its neighbor entities h_j connected with relation r_k, let their embeddings be $\bar{h}_i$, $\bar{h}_j$,

and $\bar{g}_k$ respectively. The corresponding neighbor feature, $\bar{c}_{ijk}$, attention value, b_{ijk}, and the attention score, α_{ijk}, are as follows: $\bar{c}_{ijk} = \mathbf{W_1}[\bar{h}_i||\bar{h}_j||\bar{g}_k]$; $b_{ijk} = \text{LeakyReLU}(\mathbf{W_2}\bar{c}_{ijk})$; $\alpha_{ijk} = \text{softmax}_{jk}(b_{ijk})$. Finally, the new presentation of entity h_j is $\bar{h}'_i = \sigma(\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{R}_{ij}}\alpha_{ijk}\bar{c}_{ijk})$, where $\mathcal{N}_i$ denotes the neighbor entity set of entity h_i, $\mathcal{R}_{ij}$ denotes the set of relations connecting entities h_i and h_j, and σ represents any non-linear function.

However, components of KBGAT are not well-generalizable to UKG, resulting in worse performance, for which we have identified two major issues.

First, KBGAT does not take confidence score information into account in its aggregation mechanism. It is presumable that due to the uncertainty and high density, irrelevant triplets are prevailing in the uncertain KG. The intuitive assumption is that there exist positive correlations between the confidence score and attention score. Therefore, a naive solution is to explicitly formulate a confidence score in the attention mechanism. For instance, let s_{ijk} be the confidence score of triplet (h_i, r_k, h_j), we have explored the following three setups: (1) concatenate the confidence score into the feature vector: redefine $\bar{c}_{ijk}$ as $\bar{c}_{ijk} = \mathbf{W_1}[\bar{h}_i||\bar{h}_j||\bar{g}_k||s_{ijk}]$, (2) linearly weight confidence score with the learned attention value: Redefine $\bar{h}'_i$ as $\bar{h}'_i = \sigma(\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{R}_{ij}} s_{ijk}\alpha_{ijk}\bar{c}_{ijk})$, and (3) let the attention value be the confidence score: $\bar{h}'_i = \sigma(\sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{R}_{ij}} s_{ijk}\bar{c}_{ijk})$. Analyses in Sect. 4.3 show none of them is perceivably better over KBGAT [14].

In fact, these naive assumptions are questionable. First, the embeddings of neighbor triplets should have contained their confidence score information and obviated the need to model it explicitly. We only formulated confidence scores in the loss function. Second, there is no clear evidence that weighting neighbors according to their plausibility helps the attention process. Low-confidence triplets are not necessarily irrelevant. Hence, we avoid adding constraints between the attention and the confidence scores. Rather, we aim to propose an uncertainty-aware attention mechanism to properly handles irrelevant neighbors.

Second, finding multi-hop relations is infeasible using path-searching algorithms due to uncertainty. To broaden the scope of attention beyond 1-hop neighbors, Xie et al. [14] propose the n-hops auxiliary neighbor mechanism. However, given uncertain relations, such n-hops operation is difficult to be realized without extra domain knowledge. For instance, Chen et al. [3] apply human-defined first-order logic to imply the plausibility of multi-hop relations. Please note this problem is not unique to attention-based models but to most existing methods.

With regard to the above issues, therefore, we propose a robust attention-based model to reduce the impact of irrelevant neighbors and extend its attention range beyond one hop, even if relation links are uncertain.

2.3 Related Work

Traditional embedding methods for deterministic KGs, such as [1,15], focus primarily on modeling single triplets. Subsequent studies explore more about subgraph modeling. For example, [7] exploits paths as subgraph features. There are also logic rule-based methods, e.g. [10,16], which can model more subgraph

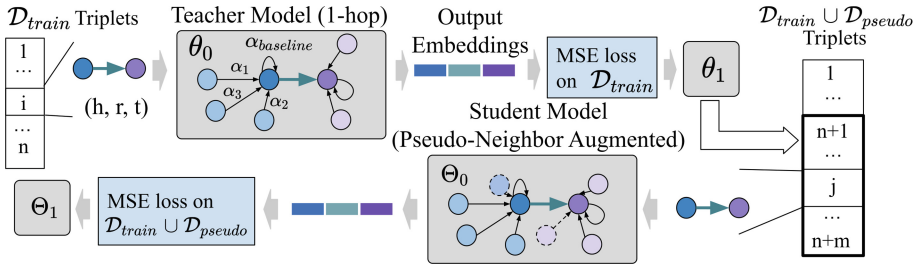

Fig. 1. The overview of UPGAT. α is the attention score. Dashed circles and arrows represent pseudo-neighbors.

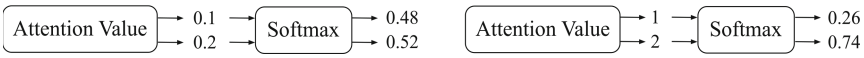

Fig. 2. The attention value on the left is lower than that on the right (0.1 vs 1.0), but, after softmax, the attention score on the left becomes significantly higher (0.48 vs 0.26)

patterns other than paths with well-designed rule templates. Yet, such templates rely on prior domain knowledge. Recently, neural-network-based methods with deeper and more complex artificial neural networks are proposed, e.g., [8,11,12,14]. Some of them can model larger subgraphs without extra human knowledge.

Existing works for uncertain knowledge graph embedding, to our knowledge, only focuses on modeling single triplets; none of them explore how to utilize subgraph features. UKGE [3], the first method for embedding uncertain knowledge graphs, applies probabilistic logic rules (PSL) to infer unseen triplets. PASSLEAF [4] aims to alleviate the false-negative problem. SUKE [13] proposes an evaluator-generator architecture. BEUrRE [2] models each entity as a box and relations as affine transforms to capture the uncertainty. FocusE [9] adds an extra layer after the scoring layer, balancing the high-confidence and uncertain graph structure.

3 Approach

3.1 Overview

The uncertainty-aware pseudo-neighbor augmented knowledge graph attention network (UPGAT) is a generalization of the graph attention network featuring the ability to incorporate uncertain subgraph features and exploit multi-hop relations. Particularly, for the issues discussed in Sect. 2.2, the attention baseline mechanism alleviates the negative impact of irrelevant neighbors; pseudo-neighbor augmentation overcomes the difficulty of identifying multi-hop paths.

UPGAT consists of two pipelined training stages depicted in Fig. 1. First, train a 1-hop graph attention model, θ, to generate the pseudo-neighbor triplets, indexed $n+1$ to $n+m$. Second, train a multi-hop model, Θ, based on the augmented knowledge graph to explore both one- and multi-hop relations.

3.2 1-Hop Attention Module with Attention Baseline Mechanism

This subsection explains the proposed attention model with the attention baseline mechanism, tailored to handle irrelevant neighbors and to increase robustness. We start with the case where only one-hop neighbor features are used.

For the following discussions, let $\bar{h}_i$ and $\bar{g}_k$ denote the embeddings of an entity h_i and a relation r_k, respectively. Let $\mathcal{R}_{ij}$ be the set of relations connecting entities h_i and h_j. Let $\mathcal{N}_i$ be the set of the 1-hop neighbor entities of entity h_i.

Given an entity h_i, let the neighbor entity-relation representation $\bar{c}_{ijk}$ be $\mathbf{W_1}[\bar{h}_j \circ \bar{g}_k]$, for a neighbor entity h_j and the connecting relation g_k, where $\mathbf{W_1} \in \mathbb{R}^{d \times d}$ and d is the embedding size. The corresponding attention value, i.e., the importance of the neighbor feature, is defined in Eq. (1), where $\mathbf{W_a} \in \mathbb{R}^{1 \times d}$.

$$a_{ijk} = \text{LeakyReLU}(\mathbf{W_a}[\bar{h}_i \circ \tanh(\bar{c}_{ijk})]). \tag{1}$$

We aggregate neighbor features with an element-wise product to preserve both the semantic interaction and confidence score information between the entity and the relation in the embedding vectors. We pass this information to the attention mechanism and let it decides what is important instead of explicitly modeling the confidence score. Compared to concatenating embedding vectors as in GAT [12] and KBGAT [14], it also reduces the feature dimension, enhancing the robustness against noises from uncertainty.

To mitigate the impact of many irrelevant uncertain neighbors, we introduce the attention baseline mechanism, before normalizing the attention values by softmax to get the attention score. The normalized attention score is essentially a convex combination among all neighboring features. Consequently, when all neighbors are irrelevant, the attention mechanism still has to "choose" from one of them, as shown in Fig. 2. Thus, the attention baseline mechanism serves as the "none of them" option. Each neighbor can get a high normalized attention score only when its attention value is higher than the baseline value.

Formally, for each entity h_i, add a baseline value a_i^{baseline} as follows where g_0 is the embedding for the special self-loop relation of the attention baseline:

$$\begin{aligned} \bar{c}_i^{\text{baseline}} &= \mathbf{W_1}[\bar{h}_i \circ \bar{g}_0], \\ a_i^{\text{baseline}} &= \text{LeakyReLU}(\mathbf{W_a}[\bar{h}_i \circ \tanh(\bar{c}_i^{\text{baseline}})]). \end{aligned} \tag{2}$$

For the attention score, normalize the baseline value and the attention value of other neighbors together as below in Eq. 3. Then, get the new embedding, $\bar{h_i}'$, of entity h_i as shown in Eq. (4), composed of the weighted neighbor and self representation.

$$\begin{aligned} \alpha_{ijk} &= \frac{\exp(a_{ijk})}{\exp(a_i^{\text{baseline}}) + \sum_{m \in \mathcal{N}_i} \sum_{n \in \mathcal{R}_{im}} \exp(a_{imn})}, \\ \alpha_i^{\text{baseline}} &= \frac{\exp(a_i^{\text{baseline}})}{\exp(a_i^{\text{baseline}}) + \sum_{m \in \mathcal{N}_i} \sum_{n \in \mathcal{R}_{im}} \exp(a_{imn})}. \end{aligned} \tag{3}$$

$$\bar{h_i}' = \sigma(\alpha_i^{\text{baseline}} \bar{c}_i^{\text{baseline}} + \sum_{j \in \mathcal{N}_i} \sum_{k \in \mathcal{R}_{ij}} \alpha_{ijk} \bar{c}_{ijk}). \tag{4}$$

The non-linear function σ we choose is ELU. [5].

$\alpha_i^{\text{baseline}}$ in Eq. (4) can be regarded as a gate that controls the amount of neighbor information. As the neighbor's attention values are lower, the new embedding $\bar{h_i}'$ contains less neighbor information, and vice versa. A similar concept is the self-attention mechanism proposed by [12]. However, it cannot be applied to multi-relational graphs, so this mechanism is removed in KBGAT and replaced by residual connections. Nevertheless, the residual connection cannot be used to control the intensity of neighbor information. Our method can be viewed as a generalization of self-attention for multi-relational graphs.

Finally, let the output entity embedding be $\bar{h_i}^o = \mathbf{W_E}\bar{h}_i + \bar{h}_i'$ where $\mathbf{W_E} \in \mathbb{R}^{d\times d}$; $\bar{h_i}$ and $\bar{h_i}'$ are the initial and new entity embedding respectively. Similarly, let the output relation embedding be $\bar{g_k}^o = \mathbf{W_R}\bar{g}_k$ where $\mathbf{W_R} \in \mathbb{R}^{d\times d}$.

This is the proposed attention module for 1-hop neighbors. The attention module beyond 1-hop will be described in Sect. 3.4.

3.3 Confidence Score Prediction and Training Objective

Confidence Score Prediction. To estimate the confidence score of a given triplet $l = (h_i, r_k, h_j)$, firstly, use DistMult [15] as the scoring function to get the score $p(l) = \bar{h_i}^o \cdot (\bar{g}_k^o \circ \bar{h}_j^o)$. Then, map $p(l)$ to the [0,1] range to get the confidence score prediction: $S(l) = S(p(l)) = \text{Sigmoid}\Big(w \cdot p(l) + b\Big)$, where w and b are the weight and bias for the linear transformation.

Training Objective . Given a triplet $l = (h, r, t)$ and its confidence score s_l, where $h, t \in \mathcal{E}, r \in \mathcal{N}, s_l \in [0, 1]$, use the mean-square-error loss function to make $S(l)$ approximate s_l.

For a triplet l from the training triplet set $\mathcal{L}^+$ and its confidence score s_l, $\mathcal{J}^+ = \frac{1}{|\mathcal{L}^+|}\sum_{l\in\mathcal{L}^+} ||S(l) - s_l||^2$. For a triplet l from the randomly drawn negative triplet set $\mathcal{L}^-$, assume $s_l = 0$ such that $\mathcal{J}^- = \frac{1}{|\mathcal{L}^-|}\sum_{l\in\mathcal{L}^-} ||S(l)||^2$. Finally, sum the two terms up to get the total loss for the 1-hop model: $\mathcal{J} = \mathcal{J}^+ + \lambda_1\mathcal{J}^-$, where λ_1 is a hyper-parameter set to 1 by default.

3.4 Pseudo-neighbor Augmented Graph Attention Network

This subsection covers how to extend the base 1-hop attention module to n-hop and concludes the proposed UPGAT model. In deterministic graphs, multi-hop features can be represented as auxiliary relations, which are the summation of the relation embeddings in the path between any n-hop entity pair. However, in an uncertain graph, the confidence score of these relations is unknown without data-dependent inference rules. We use pseudo-neighbors to overcome the limitation.

After training the 1-hop model, build the *pseudo-neighbor augmented uncertain knowledge graph*, consisting of the original KG and the predicted false-negative triplets. Precisely, use the 1-hop model as the teacher model to predict false-negative tail entities for all $(head, relation)$ pairs in the training data. Then, select those with top-k predicted confidence scores to form the new augmented

dataset. These pseudo-labeled triplets are the missing neighbors that may span across one or multiple hops in the original graph. The rationale for using the top-k filtering is that it causes more damage to the model to mispredict triplets with higher confidence scores as negative.

Finally, train a student model from scratch on the pseudo-neighbor augmented uncertain knowledge graph as the final model of the UPGAT. Pseudo-neighbors enable the graph attention to extend its attention range beyond 1-hop, no longer limited by the uncertainty of the paths. Note that the pseudo-labeling process follows a two-staged teacher-student schema to improve training stability. However, other semi-supervised learning strategies can be applied as well.

Formally, for the attention module beyond 1-hop, modify Eqs. (3, 4) by replacing $\mathcal{N}_i$ and $\mathcal{R}_{im}$ in the summation with $\mathcal{N}'_i = \mathcal{N}_i \cup \mathcal{N}_i^{pseudo}$ and $\mathcal{R}'_{im} = \mathcal{R}_{im} \cup \mathcal{R}_{im}^{pseudo}$ where $\mathcal{N}_i^{pseudo}$ is the predicted false-negative neighbor set of entity h_i; $\mathcal{R}_{ij}^{pseudo}$ is the set of predicted relations bridging h_i and h_j, which can be both 1-hop and multi-hop in the original graph.

$\mathcal{J}^{pseudo}$ and $\mathcal{J}^{semi}$ shown below are the loss function of the predicted samples $\mathcal{L}^{pseudo}$ and the total loss for the semi-supervised learning of the student model respectively, where λ_2 is the hyper-parameter.

$$\begin{aligned} \mathcal{J}^{pseudo} &= \frac{1}{|\mathcal{L}^{pseudo}|} \sum_{l \in \mathcal{L}^{pseudo}} ||f(l) - s_l^{pseudo}||^2. \\ \mathcal{J}^{semi} &= \mathcal{J}^{+} + \lambda_1 \mathcal{J}^{-} + \lambda_2 \mathcal{J}^{pseudo}. \end{aligned} \tag{5}$$

4 Experiment

We evaluate our model with two tasks: confidence score prediction (CSP) and tail entity prediction (TEP). Sect. 4.2 compares UPGAT with existing works. The ablation studies in Sect. 4.3 intend to answer the following questions: (i) How well native solutions perform; (ii) If the attention baseline mechanism boosts the performance; (iii) If the pseudo-neighbor augmented graph attention successfully models multi-hop relations. Lastly, in Sect. 4.4, we verify if the proposed method works well on the deterministic KG.

4.1 Settings

We used the same training and testing datasets, CN15K/NL27K/PPI5K, and the same evaluation metrics as used by [4]. The CSP task is a regression problem to predict the confidence score given a triplet, so we take MSE as the metric. The TEP task is a ranking problem to predict the tail entity given a head entity and a relation so we choose weighted mean rank, weighted Hit@k, and NDCG as the metrics. Chen et al. [4] proposed new evaluation metrics such as WMR (Weighted Mean Rank) and WH@K, which are MR (mean rank) and Hit@K linearly weighted by confidence scores. The new metrics are claimed to be more suitable for the uncertain knowledge graph. We keep this setting for compatibility as these metrics show identical trends as the original ones, e.g., MR and H@K.

Table 1. Tail Entity Prediction & Confidence Score Prediction. (MSE in 0.01). Boldface indicates the best value for a metric.

		TEP				CSP	
datasets	models	WMR	WH@20	WH@40	NDCG	MSE+	MSE-
CN15K	UKGE$_{logi}$ [3]	1676.0	32.2%	38.5%	29.7%	28.2	**0.17**
	PASSLEAF [4]	1326.3	34.2%	41.3%	**30.4**%	23.8	0.36
	SUKE [13]	1849.5	32.3%	38.3%	29.8%	30.4	**0.05**
	UPGAT (ours)	**1098.7**	**36.0%**	**44.4%**	28.7%	**18.3**	0.27
NL27K	UKGE$_{logi}$ [3]	288.6	70.4%	76.8%	71.7%	7.9	0.32
	PASSLEAF [4]	242.3	71.8%	77.9%	**74.5**%	5.5	0.38
	SUKE [13]	268.7	71.5%	78.2%	73.8%	4.1	**0.03**
	UPGAT (ours)	**109.2**	**72.0%**	**78.4%**	73.3%	**2.6**	0.10
PPI5K	UKGE$_{logi}$ [3]	38.6	42.6%	68.8%	43.9%	0.76	0.28
	PASSLEAF [4]	34.9	45.1%	70.6%	44.5%	0.51	0.30
	SUKE [13]	37.0	45.9%	71.3%	**45.3**%	0.52	**0.17**
	UPGAT (ours)	**32.4**	**46.1%**	**72.2%**	44.6%	**0.34**	0.22

Our deterministic KGs settings follow those in [3] to binarize the uncertain KGs of CN15K/NL27K/PPI5K with the thresholds of 0.8/0.8/0.85. For evaluation metrics, we use the MR (Mean Rank) and Hit@K. The comparison of UKG embedding methods with DKG methods is unfair and beyond the scope of this paper, as they are optimized for the unique characteristics of DKG.

The UKGE$_{logi}$ model is UKGE [3] trained without PSL-enhanced data to get the initial entity and relation embeddings. For other hyper-parameters, we choose Adam optimizer with learning rate = (5e−4), batch size = 512, embeddings size = 512, $k = 20$ pseudo-labeled triplets, and negative sampling ratio = 10.

4.2 Results and Analysis

The experimental results of the TEP and CSP task are presented in Table 1. MSE+ and MSE− are the MSE on in-dataset positive samples and randomly drawn negative samples, respectively. We choose models with similar score function as the baselines, which, therefore, shares similar mathematical structures in the embedding space. Values for UKGE$_{logi}$ and PASSLEAF (with Distmult) are from [4]; values of SUKE are our reproduced results without PSL augmented data. FocusE [9] is not listed as it can be an add-on layer to most models.

Experimental results show that our method outperforms the existing methods in most of the evaluation metrics on the TEP task; there is over 50% WMR improvement on the NL27K dataset. Note that the NDCG score is dominated by a very small number of candidates with top confidence scores. The discounted factor of NDCG is a logarithmic function. Therefore, as the ranking goes up, the penalties for the candidates with lower confidence scores become significantly less. For example, the ratio of the discounted factor of the 3^{rd} to 1^{st} candidate is the same as the 99^{th} to 9^{th} candidate.

Table 2. Ablation Study - TEP & CSP (MSE in 0.01). Boldface indicate better value than existing methods on a metric.

		TEP				CSP	
dataset	models	WMR	WH@20	WH@40	NDCG	MSE+	MSE-
CN15K	1-hop w/o AB	**1215.6**	32.8%	**42.0%**	27.4%	**19.7**	0.29
	1-hop w/o AB + Naive	**1210.3**	32.9%	**42.1%**	27.7%	**19.7**	0.28
	1-hop	**1160.8**	**35.5%**	**43.3%**	28.3%	**19.1**	0.27
	1-hop+TS	**1111.9**	**35.8%**	**44.3%**	28.6%	**19.0**	0.26
	UPGAT (pseudo-neighbor+TS)	**1098.7**	**36.0%**	**44.4%**	28.7%	**18.3**	0.27
NL27K	1-hop w/o AB	**160.6**	70.9%	78.0%	71.1%	**2.9**	0.10
	1-hop w/o AB + Naive	**160.7**	70.8%	78.0%	71.1%	**2.9**	0.10
	1-hop	**151.5**	71.3%	78.2%	73.2%	**2.9**	0.10
	1-hop+TS	**119.3**	**71.8%**	**78.4%**	73.2%	**2.7**	0.10
	UPGAT (pseudo-neighbor+TS)	**109.2**	**72.1%**	**78.4%**	73.3%	**2.6**	0.10
PPI5K	1-hop w/o AB	35.1	44.2%	70.4%	43.1%	**0.42**	0.30
	1-hop w/o AB + Naive	**34.8**	45.2%	71.4%	43.0%	**0.41**	0.31
	1-hop	**33.2**	45.9%	70.9%	43.1%	**0.35**	0.24
	1-hop+TS	**32.9**	**46.0%**	**71.7%**	44.1%	**0.36**	0.27
	UPGAT (pseudo-neighbor+TS)	**32.4**	**46.1%**	**72.2%**	44.6%	**0.34**	0.22

For the CSP task, experimental results show that our method outstrips the existing methods in MSE+, reducing by up to 36% relative to the best model on NL27K. Contrarily, we incline not to emphasize the MSE of negative samples (MSE-) since there are no ground-truth negative labels in these three datasets. SUKE performs better on the negative MSE of the three datasets because it uses the evaluator to assign all low confidence scores to zero values, which we consider to be a post-processing method and does not imply better CSP accuracy.

4.3 Ablation Study

Table 2 is the result of the ablation study for the questions at the start of Sect. 4.

For the model name abbreviations, [1-hop] indicates the 1-hop attention model, depicted in Sect. 3.2; [1-hop w/o AB] refers to the 1-hop model without the attention baseline mechanism. Note that even the [1-hop w/o AB] model outperforms existing methods on most metrics, verifying our claimed advantage of incorporating subgraph features for UKG.

The naive solutions to model uncertainty: [1-hop w/o AB] vs [1-hop w/o AB +Naive]. [1-hop w/o AB + Naive] represents the three naive solutions discussed in Sect. 2.2 that explicitly model confidence score in the attention mechanism. As they have similar performance, only the one with the best WMR is shown. These three solutions achieve limited or no improvement, supporting the founding assumption that the plausibility of a neighbor cannot imply how relevant it is to the centered triplet.

The attention baseline: [1-hop] vs [1-hop w/o AB]. From the two models, it is evident that adding the attention baseline mechanism greatly improves

Table 3. TEP with the deterministic settings depicted in Sect. 4.1. (H@20 in %)

	CN15K		NL27K		PPI5K	
models	MR	H@20	MR	H@20	MR	H@20
UKGE$_{logi}$	3586.4	27.6	335.5	70.6	57.9	52.9
SUKE	3033.1	27.3	330.1	70.7	57.9	53.0
PASSLEAF	2402.0	27.7	312.6	71.5	55.2	54.8
UPGAT (ours)	**2368.5**	**28.0**	**288.1**	**72.1**	**53.3**	**55.6**

Table 4. Ablation Study for the Attention Baseline Mechanism in TPE. The (.) indicate the relative change w.r.t. (1-hop w/o AB).

		Deterministic		Uncertain	
	models	MR	H@20	WMR	WH@20
CN15K	1-hop w/o AB	2389.0	27.6%	1215.6	32.8%
	1-hop	2386.6 (−0.1%)	27.7% (+0.4%)	1160.8 (−4.5%)	35.5% (+8.2%)
NL27K	1-hop w/o AB	312.7	71.4%	160.6	70.9%
	1-hop	307.9 (−1.5%)	71.6% (+0.2%)	151.5 (−5.6%)	71. 3%(+0.6%)
PPI5K	1-hop w/o AB	55.3	54.7%	35.1	44.2%
	1-hop	55.3 (−0.1%)	54.8% (+0.2%)	33.2 (−5.2%)	45.9% (+3.8%)

its performance in most metrics. We attribute this to the mitigation of the irrelevant neighbor problem. The attention baseline mechanism strengthens the ability to extract the relevant information, outperforming the baseline methods that explicitly model the confidence score.

Pseudo-neighbor augmented graph attention: [UPGAT] vs [1-hop + TS] vs [1-hop]. We break down the contributions of the teacher-student semi-supervised learning (TS) and pseudo-neighbors. [1-hop + TS] generates pseudo-neighbors as UPGAT does, but the data are not used in the neighbor aggregation. Results show that applying graph attention over pseudo-neighbor augmented graph can further advance in all metrics, given that TS has already brought notable improvement over the [1-hop] model. Such results suggest that it is viable and effective to model multi-hop relations with the pseudo-neighbors in a data-driven manner.

4.4 Deterministic Settings

To verify if our proposed method is compatible with the deterministic KG (DKG), we compare UPGAT with existing UKG methods on DKGs in the TEP task. The result is presented in Table 3. UPGAT outperforms other methods on all metrics. As the number of ground truth labels is only 20% of the original uncertain settings, only H@20 is shown.

The ablation study shown in Table 4 further compares the attention baseline mechanism on the deterministic and uncertain KG. The attention baseline improves UKGs but has limited benefits for DKGs. For example, this mechanism

brings 3.8% WH@20 improvement for the uncertain setting and only 0.2% H@20 improvement for the deterministic setting in the PPI5k dataset. This agrees with our argument that the attention baseline is effective for extending the graph attention to UKGs where uncertainty complicates the neighboring features.

5 Conclusion and Future Work

The proposed Uncertainty-Aware Pseudo-neighbor Augmented Knowledge Graph Attention Network (UPGAT) is the first work to model the subgraph feature on uncertain KGs. The attention baseline mechanism generalizes the self-attention model for multi-relational graphs with uncertainty; The pseudo-neighbors successfully model multi-hop relations. Our model gets promising improvements over existing works on both uncertain and deterministic KGs. While this paper focuses on a mechanism of fixed attention weight for each entity, we believe the weightings of different relation queries are also important. Namely, the attention mechanism must be "relation query-aware", which we leave for future studies.

References

1. Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NIPS, pp. 2787–2795 (2013)
2. Chen, X., Boratko, M., Chen, M., Dasgupta, S.S., Li, X.L., McCallum, A.: Probabilistic box embeddings for uncertain knowledge graph reasoning. In: NAACL'21 (2021)
3. Chen, X., Chen, M., Shi, W., Sun, Y., Zaniolo, C.: Embedding uncertain knowledge graphs. In: AAAI (2019)
4. Chen, Z.M., Yeh, M.Y., Kuo, T.W.: Passleaf: a pool-based semi-supervised learning framework for uncertain knowledge graph embedding. AAAI **35**(5) (2021)
5. Clevert, D.A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus) (2016)
6. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR (2017)
7. Lin, Y., Liu, Z., Luan, H., Sun, M., Rao, S., Liu, S.: Modeling relation paths for representation learning of knowledge bases. In: EMNLP (2015)
8. Nguyen, D.Q., Nguyen, T.D., Nguyen, D.Q., Phung, D.: A novel embedding model for knowledge base completion based on convolutional neural network. In: NAACL (Short Papers), vol. 2, pp. 327–333 (2018)
9. Pai, S., Costabello, L.: Learning embeddings from knowledge graphs with numeric edge attributes. In: IJCAI, pp. 2869–2875 (2021)
10. Qu, M., Tang, J.: Probabilistic Logic Neural Networks for Reasoning. Curran Associates Inc. (2019)
11. Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: ESWC (2018)
12. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: ICLR (2018)

13. Wang, J., Nie, K., Chen, X., Lei, J.: Suke: embedding model for prediction in uncertain knowledge graph. IEEE Access **9**, 3871–3879 (2021)
14. Xie, Z., Zhu, R., Zhao, K., Liu, J., Zhou, G., Huang, J.X.: Dual gated graph attention networks with dynamic iterative training for cross-lingual entity alignment. ACM Trans. Inf. Syst. **40**(3) (2021)
15. Yang, B., tau Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases (2015)
16. Zhang, Y., et al.: Efficient probabilistic logic reasoning with graph neural networks. In: ICLR (2020)