

Article An Intelligent Question-Answering Model Over Educational Knowledge Graph for Sustainable Urban Living

Yutong Fang ¹, Jianzhi Deng ^{1,2,*}, Fengming Zhang ¹ and Hongyan Wang ³

- ¹ College of Information Science and Engineering, Guilin University of Technology, Guilin 541006, China
- ² Guangxi Key Laboratory of Embedded Technology and Intelligent System, Guilin 541006, China
- ³ School of Software Engineering, University of Science and Technology of China, Hefei 230026, China
- * Correspondence: dengjzh@glut.edu.cn

Abstract: With the development of education informatization and the accumulation of massive educational resources and teaching data in urban environments, educational knowledge graphs that provide good conditions for developing data-driven intelligent education have been proposed. Based on such educational knowledge graphs, the question-answering method can provide students with immediate coaching and significantly increase their learning interest and productivity. However, there is little research on knowledge graph question-answering focused on the educational field. Students tend to consult complex questions that require reasoning; however, the existing QA system cannot satisfy their complex information needs. To help improve sustainable learning efficiency, we propose a novel intelligent question-answering model applied in smart cities, which can reason over the educational knowledge graph to locate the answers to given questions. Our approach uses a highly expressive bilinear graph neural network technology to perform forward reasoning, utilizing the contextual information between graph nodes to improve reasoning ability. On this basis, we propose two-teacher knowledge distillation. We construct two distinct teacher networks by combining forward and backward reasoning, then incorporate the intermediate supervision signals from the two networks to guide the reasoning process, thereby mitigating the phenomenon of spurious path reasoning. Extensive experiments on the MOOC Q&A dataset prove the effectiveness of our approach.

Keywords: sustainable learning; intelligent question answering; educational knowledge graph; bilinear graph neural network; smart cities

1. Introduction

In recent years, the rapid development of information technology has led data to grow exponentially, and the types of data have become more diverse. Obtaining accurate and valuable information is particularly urgent in a large-scale data environment. It has always been a wish of people to be able to quickly and accurately obtain helpful information. The emergence of intelligent question-answering makes up for this shortcoming and provides convenience for sustainable urban living, as user intentions can be understood and analyzed the through the user's natural language questions to provide the most direct answer. Question-answering over knowledge graphs (KGQA) is one of the essential branches of intelligent question-answering systems. It aims to seek answers to given questions from the knowledge graph.

The knowledge graph (KG) uses resource description framework (RDF) triples to store a substantial amount of entities and the deep semantic relationships that connect them. According to the rules of RDF, each piece of knowledge in a knowledge graph is expressed in the form of (h, r, t) triples, where h, r, and t represent the head entity, relation, and tail entity, respectively. This method offers a useful method of organizing enormous amounts of information, and as a result is gaining increasing attention. As education



Citation: Fang, Y.; Deng, J.; Zhang, F.; Wang, H. An Intelligent Question-Answering Model Over Educational Knowledge Graph for Sustainable Urban Living. *Sustainability* **2023**, *15*, 1139. https://doi.org/10.3390/su15021139

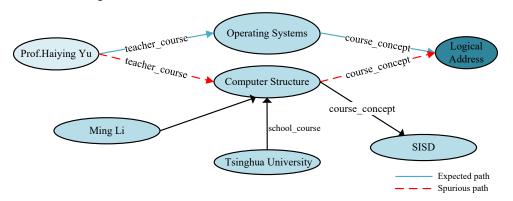
Academic Editor: Haoran Wei, Zhendong Wang, Yuchao Chang and Zhenghua Huang

Received: 30 November 2022 Revised: 25 December 2022 Accepted: 4 January 2023 Published: 7 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). develops, a wide variety of teaching materials and educational resources are created [1]. Educational knowledge graphs can compile dispersed and disorganized educational data into structured knowledge that is simple to retrieve, modify, and preserve, lowering users' usage costs and promoting quick cognitive improvement [2] in urban environments. For instance, Yu et al. from Tsinghua University [3] proposed MOOCCube, a massive open online course (MOOC) knowledge graph that includes entities such as concepts, courses, teacher and student behaviors, and information about their interactions to support the teaching and research requirements of numerous scenarios. Xu and Guo [4] proposed a knowledge graph of knowledge points applicable to K-12 education, in which the nodes mainly contain knowledge points, schools, teachers, etc. The results of their experiments demonstrated that employing this knowledge graph can significantly increase the accuracy of recommendations for educational resources.

The development of educational knowledge graphs serves as the foundation for intelligent question-answering. However, there are few studies on educational knowledge graph question-answering, which makes it impossible to meet the demands of developing educational platforms [5]. For instance, the WebQuestionsSP [6] dataset based on Freebase [7] comprises 513 relations, while the MOOC Q&A based on MOOCCube [3] only contains ten relation types. On the other hand, students prefer to consult complex questions that require multi-hop reasoning in the knowledge graph, and the existing QA system cannot satisfy their complex information needs. For example, the question from the MOOC Q&A [3] dataset shown in Figure 1 ("What concepts are covered in the Operating Systems course taught by Prof. Haiying Yu?") relies on (Prof. Haiying Yu, teacher_course, <Operating Systems>) and (<Operating Systems>, course_concept, Logical Address). Therefore, this paper aims to propose an intelligent question-answering model that can answer complex multi-hop problems in educational contexts.



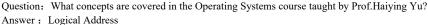


Figure 1. An example of multi-hop question answering from the MOOC Q&A dataset and MOOC-Cube knowledge graph. For the sake of simplicity, we only display the part of the subgraph that is relevant to the question.

However, there are limitations to the existing KGQA techniques. Because Graph Convolutional Networks (GCN) [8] have a remarkable ability to capture graph structure information, several studies [9,10] have used improved GCN to generate entity representations to predict answers. However, the traditional Graph Neural Network (GNN) models only utilize the feature information obtained by simple Linear Aggregation (LA) of neighboring nodes, which ignores the possible interactions between graph nodes. To find the correct answers to various complex questions that require multi-hop reasoning on large knowledge graphs, it is necessary that the KGQA model can both effectively utilize the context information between neighbors. Additionally, when using deep learning models for training, models only have access to weakly supervised signals (i.e., questions and

answers), as labeling the entire reasoning path is an expensive process. Because the intermediate reasoning paths are absent, the resulting models may be susceptible to spurious path reasoning. Figure 1 shows an example of this; while the final answer is correct, it may have been obtained based on spurious path reasoning (indicated by the blue lines). Existing research [11] has combined bidirectional reasoning to create teacher networks that offer supervision signals (i.e., intermediate entity distributions) to lead the reasoning process to address this problem. Essentially, this is the application of knowledge distillation [12]. By distilling the supervision signals generated by the teacher network and then transmitting them to the student network, the student network can achieve good performance. However, more than one teacher may be required to provide comprehensive supervision signals. In reality, a student does not learn everything from one teacher. Instead, they may receive advice on understanding a concept from many professors during class, after class, or even online.

In light of the above challenges and in an effort to improve sustainable learning efficiency, we propose BGNN-TT, an intelligent question-answering model utilizing a bilinear graph neural network and two-teacher knowledge distillation. The main contributions of our paper are as follows:

- 1. We combine the highly expressive Bilinear Graph Neural Network [13] technology to perform multi-hop reasoning. More precisely, we introduce Linear Aggregation (LA) and Bilinear Aggregation (BA) in the reasoning process, which can reflect the contextual information between graph nodes and obtain comprehensive representation, leading to an overall improvement in reasoning ability.
- 2. Inspired by multi-view learning [14], we propose a two-teacher knowledge distillation approach to mitigate spurious path reasoning in educational knowledge graph question-answering. Based on bilinear neural network technology, we build two different teacher networks then incorporate the intermediate supervision signals (i.e., intermediate entity distributions) from the two teacher networks to direct the reasoning process, which can overcome the limitation of a single teacher and provide stronger supervision signals.
- 3. We apply our approach to the real MOOC Q&A [3] dataset and the large educational knowledge graph MOOCCube [3]. The results indicate that our proposed method outperforms other baseline models, verifying the efficacy of the method.

2. Related Works

2.1. Question-Answering over Knowledge Graph

Question-Answering over Knowledge Graph (KGQA) is a crucial technology that promotes urban sustainability. It is capable of understanding the user's purpose and returning the right response. Due to KGQA's rapid expansion, it is widely used in education, medical care, tourism, finance, and other fields of sustainable urban life. Until now, there have been two mainstream branches of KGQA: semantic parsing-based methods and retrieval-based methods. Traditional semantic parsing methods [15–17] use logical forms for querying. Recent studies [18–21] have used query graphs to reflect the question's semantic structure to improve performance. These SP-based algorithms achieve competitive performance at the expense of manually created features, patterns, language engineering, and data annotations, making it difficult for them to be extended to new domains.

Instead of constructing a structured representation, retrieval-based methods [22,23] retrieve the answer directly from KG according to the information conveyed in the question. Recently, multi-hop based KGQA has drawn the attention of researchers. A typical model in reasoning methods based on embedding representation knowledge graphs is the TransE model proposed by Bordes et al. [24], which predicts missing entities or relationships by performing translation operations on vectors in low-dimensional spaces. Because of TransE's strong generalization ability, it is widely used in KGQA. Hang et al. [25] proposed the KEQA model, using used TransE [24] to encode questions and entities to predict answers. Although the accuracy is increased in this way, the model cannot respond to questions involving multi-

hopping. Saxena et al. [26] proposed EmbedKGQA, which builds the KGQA model using ComplEx [27] for multi-hop questions.

As deep learning technology advances, neural networks are attracting much attention, and have been used to create better distributed representations of questions and entities. Miller et al. [28] proposed a Key-Value Memory Network. Because GCN [8] has a powerful ability to describe graph structure information, it is the basis for complex graph neural networks used for multi-hop reasoning methods. Sun et al. [9] proposed a GraftNet approach based on improved GCN. Based on the topic entity, it creates a subgraph targeted to the question, adds additional text to create a heterogeneous graph, iteratively updates the nodes of the heterogeneous graph using the enhanced GCN, and then reasons the answer to the question. However, GraftNet's method of extracting sub-graphs is heuristic-based, which introduces many irrelevant entities. Sun et al. [10] proposed the PullNet method, which improves on GraftNet's graph construction method. Meanwhile, a variety of graph neural networks have been developed, such as the Gated Graph Sequence Neural Networks [29], Graph Attention Network [30], and Bilinear Graph Neural Network [13]. To encode multiple graphs, Schlichtkrull et al. [31] proposed a Relation Graph Convolution Network (RGCN) and used it in the knowledge graph link prediction task. Cai et al. [32] proposed a Deep Cognitive Reasoning Network (DCRN) consisting of two phases, namely, the unconscious and conscious phases. He et al. [11] created a complex methodology based on the teacherstudent framework. Shi et al. [33] suggested a multi-hop QA approach that supports both label and text relations within a single transparent framework.

2.2. Educational Knowledge Graphs and QA Systems

Various online education platforms have emerged in sustainable urban living in recent years. For example, to assist students in overcoming academic challenges, the authors of [34] created a series of interactive robot-assisted teaching activities for classroom settings. Based on the attention mechanism and Long Short-Term Memory (LSTM) networks, the authors of [35] suggested a deep course recommendation model with multimodal feature extraction. Furthermore, with a variety of teaching materials and educational resources having been created in urban living, useful educational KGs have been launched to better organize this information. For example, Yu et al. from Tsinghua University [3] have proposed MOOCCube, a massive open online course (MOOC) knowledge graph that includes entities such as concepts, courses, teacher and student behaviors, and information about their interactions to support the teaching and research requirements of numerous scenarios. Xu and Guo [4] proposed a knowledge graph of knowledge points applicable to K-12 education in which the nodes mainly contain knowledge points, schools, teachers, etc. The results of their experiments demonstrated that employing this knowledge graph can significantly increase the accuracy of recommendations for educational resources. Lin et al. [36] used interactive data generated by university faculty in their research and teaching activities to build a university faculty knowledge graph containing faculty, research directions, research outcomes, and social adjuncts, which in turn provided qualitative and quantitative data support for faculty evaluation. Based on public KGs, a number of KGQA methods have been developed. Lin et al. [37] proposed a question-answering method over educational knowledge graph based on question-aware GCN. Yang et al. [38] designed an intelligent question-answering system based on a high school course knowledge graph. Based on the BiLSTM-CRF, Zhao et al. [39] created a question-answering system for ideological and political education. Even though these QA techniques were created for the educational domain, they fall short of offering the best results for the various questions that users ask.

3. Task Definition

In this section, we discuss the logic behind the answers to questions in natural language that are derived from the knowledge graph subgraphs that correspond to the queries.

Knowledge Graph (KG). A knowledge graph is denoted as $\mathcal{K} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$, where \mathcal{E} and \mathcal{R} denote an entity set and relation set, respectively; moreover, $\mathcal{T} = \{(\mathbf{h}, \mathbf{r}, \mathbf{t}) \mid \mathbf{h} \in \mathcal{E}, \mathbf{r} \in \mathcal{R}, \mathbf{t} \in \mathcal{E}\}$ denotes the triple set and \mathcal{G} is a subgraph related to question q, where $\mathcal{G} \in \mathcal{K}$.

Question-Answering over Knowledge Graph (KGQA). Given a natural language question *q* and subgraph \mathcal{G} associated with a question, the task of KGQA is to determine the answer entities denoted by the set \mathcal{A} , where \mathcal{A} is a subset of the set of entities \mathcal{E} .

4. Methodology

This study's research problems can be summarized as follows: (1) how to build an intelligent question-answering model over educational knowledge graph for sustainable urban living; (2) how to improve its multi-hop reasoning ability over the educational knowledge graph; and (3) how to mitigate the phenomenon of spurious path reasoning. To address the above problems, we propose an intelligent question-answering model using a bilinear graph neural network and two-teacher knowledge distillation (BGNN-TT). It is based on the following two hypotheses: (1) bilinear aggregation operation of the bilinear graph neural network can reflect the context information of the graph nodes, thereby improving the reasoning ability of the model; (2) integrating the supervision signals of different teachers can facilitate supervisory reasoning and alleviate spurious path reasoning.

This section outlines the suggested model (BGNN-TT). As shown in Figure 2, the BGNN-TT consists of four main parts: the Question Embedding Module, the Bilinear Graph Neural Network Reasoning Module, the Answer Prediction Module, and the Two-teacher Knowledge Distillation Module. The question-embedding module uses a class of BiLSTM of Recurrent Neural Network to encode natural language questions and obtain the question representation. Then, the question representation and the knowledge graph are fed into the bilinear graph neural network. For each node, the bilinear graph neural network is used to obtain the context information between these nodes and update entity representation and distribution. At the same time, the two-teacher knowledge distillation module is used to incorporate the intermediate supervision signals (i.e., intermediate entity distributions) provided by several teacher networks to direct intermediate reasoning. Finally, we the sigmoid function [40] is applied to the last layer's entity representation to predict answers.

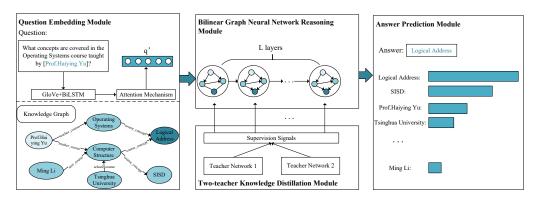


Figure 2. An overview of the proposed approach. Blue, light blue, and deep blue denote the topic entity, intermediate entities, and answer entity, respectively; longer rectangles denote a higher probability of being the correct answer.

4.1. Question Embedding Module

First, we utilize GloVe [41] and BiLSTM to embed the question and obtain the hidden state vector q_j for each word, where $q_j \in \mathbb{R}^d$ and j = 1, 2, ..., k. The last hidden state is considered to be the question representation, i.e., $q = q_k$. To further enhance the semantic

information of the question, we apply the attention mechanism to the question vector q to obtain a new question representation q':

$$\alpha_i = \operatorname{softmax}(W_{\alpha}(q \odot q_i) + b_{\alpha}) \tag{1}$$

$$v = \sum_{i=1}^{k} \alpha_j \cdot q_j \tag{2}$$

$$q' = W_q[q||v] + b_q \tag{3}$$

where $W_{\alpha} \in \mathbb{R}^{d \times d}$, $W_q \in \mathbb{R}^{d \times 2d}$, $b_{\alpha} \in \mathbb{R}^d$, and $b_q \in \mathbb{R}^d$ are parameters to be learned, \odot denotes the multiplication of the elements in the matrix, and " \parallel " indicates the concatenation operation. The newly introduced question representation includes both interactive information between the words and contextual information specific to a single question.

4.2. Bilinear Graph Neural Network Reasoning Module

GCN [8] is the foundation for complicated graph neural networks used in multi-hop reasoning techniques, and has a remarkable capacity to describe a graph's structure. A GCN generally follows a message-passing mechanism to update node representations, and the representation of layer l is usually obtained by graph convolution operations based on the representation of layer (l - 1):

$$h_v^l = \phi\left(h_v^{l-1}, \sum_{v' \in \mathcal{N}_r(v)} h_{v'}^{l-1}\right) \tag{4}$$

where $\mathcal{N}_r(v)$ denotes the neighboring node sets of node v and ϕ is a neural network layer.

However, the traditional Graph Neural Network (GNN) models only utilize the feature information obtained by simple Linear Aggregation (LA) of neighboring nodes, which ignores the possible interactions between them. In our approach, we adopt the Bilinear Graph Neural Network approach [13] to perform forward reasoning, which can utilize context information between a pair of neighboring nodes. It introduces bilinear aggregation as well as allows bidirectional feature propagation between each entity node and one of its neighboring nodes on a knowledge graph. By using both Linear Aggregation and Bilinear Aggregation, it is possible to improve multi-hop reasoning ability based on comprehensive entity representation.

According to their different calculation methods, Bilinear Aggregation can be divided into two categories: bilinear aggregation over all nodes (BA-A), and bilinear aggregation of the target node and neighboring nodes (BA-T), as shown in Figure 3. BA-A is a way to determine the structure information using the dot-product of the target node, each of its neighboring nodes, and two of their neighboring nodes. The equation is defined as follows:

$$(BA - A)^{l}(v) = FFN(\frac{1}{2}\sum_{v'}\sum_{v''}(h_{v'}^{(l-1)} \odot h_{v''}^{(l-1)}) + \sum_{v'}(h_{v}^{(l-1)} \odot h_{v'}^{(l-1)})),$$

$$(v', v'' \in N(v), v' \neq v'')$$
(5)

where N(v) denotes the neighbours of v, v' denotes the neighboring nodes that enter the node v, and v'' denotes the neighboring nodes that exit it.

BA-T only performs element-by-element multiplication of the target node and each of its neighboring nodes to find the structure information, as shown in Equation (6):

$$(BA - T)^{l}(v) = FFN\left(\sum_{v'} \left(h_{v}^{(l-1)} \odot h_{v'}^{(l-1)}\right)\right)$$
(6)

In our approach, we use BA-T to exploit the structure information between the target node and neighboring nodes. We will compare the reasoning ability of the model using BA-A and the model using BA-T in Section 5.4. As shown in Equation (7), the node representation h_v^l can be updated by combining LA and BA operations:

$$h_{v}^{l} = (1 - \beta) * LA^{l}(v) + \beta * BA^{l}(v)$$
(7)

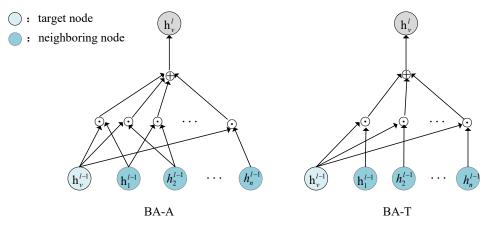
The value of β determines the applicable ratio of Linear Aggregation (LA) and Bilinear Aggregation (BA). LA can be calculated by

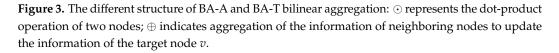
$$LA^{l}(v) = FFN\left(\left[h_{v}^{(l-1)} \|q'\|h_{N}^{(l-1)}\right]\right)$$
(8)

where $h_N^{(l-1)}$ denotes the the information generated by neighboring nodes. During the reasoning process, the feature information entering the target node is considered as the information corresponding to the reasoning path; conversely, the feature information leaving the target node is considered as information unrelated to the path. Consequently, while calculating the information of adjacent nodes generated by linear aggregation, information about neighboring nodes entering the target node should be added, while information about neighboring nodes leaving the target node should be subtracted:

$$h_N^{(l-1)} = FFN\left(\sum_r \sum_{v' \in N_r'(v)} h_{v'}^{(l-1)}\right) - FFN\left(\sum_r \sum_{v'' \in N_r''(v)} h_{v''}^{(l-1)}\right)$$
(9)

where N'_r denotes the set of neighboring nodes entering the target node and N''_r denotes the set of neighboring nodes leaving the target node.





Experimentally, performing Bilinear Aggregation only on the target node and each of its neighboring nodes shows better performance compared to performing Bilinear Aggregation on all neighboring nodes. Consequently, in our method, BA-T is used to calculate BA:

$$BA^{l}(v) = FFN\left(\left[\begin{array}{c}h_{v}^{(l-1)} \parallel \\ q' \parallel \\ FFN\left(\sum_{v'} \left(h_{v}^{(l-1)} \odot h_{v'}^{(l-1)}\right)\right)\end{array}\right]\right)$$
(10)

where $h_{n'}^{(l-1)}$ is calculated by matching the question representation q' and its relation r:

$$h_{v'}^{(l-1)} = \sum_{r} \sum_{v' \in \mathcal{N}_r(v)} \sigma(\boldsymbol{W}_{\boldsymbol{R}} \boldsymbol{r} \odot \boldsymbol{q}') \boldsymbol{p}_{v'}^l$$
(11)

where $W_R \in \mathbb{R}^{d \times d}$ are the parameters to be learned, σ denotes the sigmoid function, FFN(·) is a feed-forward layer, $\mathcal{N}_r(v)$ denotes the neighboring node set of node v, and $p_{v'}^l$ is the assigned probability of neighboring node v' at step l, which is calculated by Equation (12).

The formula below can be used to compute the probability distribution over the intermediate entities obtained in step *l*:

$$\boldsymbol{p}_{v}^{l} = \sigma \left(\boldsymbol{W}^{T} \boldsymbol{h}_{v}^{l} + \boldsymbol{b} \right)$$
(12)

4.3. Answer Prediction Module

After propagating through L layers, each node's information is transmitted through L nodes. This allows the node to reason L steps and receive the related path information in L steps. As a result, we obtain the entity representation of the last layer, h_v^L , which contains the relation path information from the topic entity to the target entity. We use this entity representation to update the entity distribution:

$$\boldsymbol{p}_{v}^{L} = \sigma \left(\boldsymbol{W}^{T} \boldsymbol{h}_{v}^{L} + \boldsymbol{b} \right)$$
(13)

The entity with the greatest score is ultimately chosen as the answer.

4.4. Two-Teacher Knowledge Distillation Module

The phenomenon of spurious path reasoning exists in the existing intelligent questionanswering model, as shown in Figure 1. Therefore, we propose two-teacher knowledge distillation to improve further the intelligent question-answering model's accuracy and sustainable learning efficiency. This component can alleviate the phenomenon of spurious path reasoning to a certain extent, and can provide a guarantee for data-driven smart education in sustainable urban living.

Knowledge distillation is a common method of model compression [12]. It distills the "knowledge" learned in a complex high-performance teacher network and transmits it to a lightweight student network. More precisely, the teacher model's predictions are treated as "soft labels" and used to train the student models. Although knowledge distillation was first suggested for model compression, recent research [11] has discovered that using soft labels as the training objective might improve student network performance. However, using only a single teacher network may not provide comprehensive knowledge, and the supervision signals (i.e., intermediate entity distributions) may be limited. Inspired by multi-view learning [14], we propose a two-teacher knowledge distillation approach that incorporates the supervision signals of two different teacher networks to guide the reasoning process. As a result, different teacher networks can compensate for the limitations of a single teacher network, allowing the model to learn more effectively. Figure 4 depicts its structure. In this section, we first describe how the teacher network was built before reviewing the two-teacher distillation process.

We combine the Bilinear Graph Neural Network technology proposed in the previous section and build two teacher networks. Teacher Network 1 uses parallel reasoning, with forward reasoning and backward reasoning being isolated. We take correspondence restrictions into account in the intermediary entity distributions between them. Figure 5a depicts the architecture of Teacher Network 1.

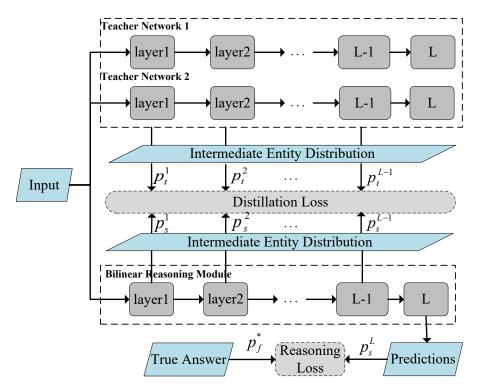


Figure 4. Framework of two-teacher knowledge distillation.

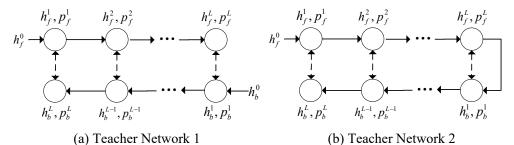


Figure 5. The architecture of the different teacher networks. For forward reasoning and backward reasoning we employ the subscripts f and b, respectively. The dotted arrows indicate that the distribution

of intermediate entities in the forward and backward reasoning processes should be consistent.

The loss of Teacher Network 1 includes the forward reasoning loss, backward reasoning loss, and consistency loss:

$$\mathcal{L}_{t1} = D_{KL} \left(\boldsymbol{p}_{f1}^{L}, \boldsymbol{p}_{f1}^{*} \right) + \lambda_{b} D_{KL} \left(\boldsymbol{p}_{b1}^{L}, \boldsymbol{p}_{b1}^{*} \right) + \lambda_{c} \sum_{l=1}^{L-1} D_{JS} \left(\boldsymbol{p}_{f1}^{l}, \boldsymbol{p}_{b1}^{L-l} \right)$$
(14)

where $p_{f1}^L(p_{b1}^L)$ denotes the final entity distribution for the forward (backward) reasoning process, $p_{f1}^*(p_{b1}^*)$ denotes the ground truth entity distribution, $D_{KL}(\cdot, \cdot)$ is the Kullback– Leibler divergence [42], $D_{JS}(\cdot, \cdot)$ is the Jensen–Shannon divergence [43], and $\lambda_b \in (0, 1), \lambda_c \in$ (0, 1) are adjustable parameters. The last item of Equation (14) is the consistency loss, which can be calculated by summing up the losses of each intermediate step; p_{f1}^l denotes the

entity distribution in step *l*, which corresponds to the distribution p_{b1}^{L-l} in step (L-l).

Teacher Network 2 utilizes hybrid reasoning. Unlike Teacher Network 1, Teacher Network 2 inputs the entity distribution and entity representation obtained in the last step of forward reasoning as initial values for backward reasoning. Figure 5b depicts the architecture of Teacher Network 2.

Similar to Teacher Network 1, the loss of Teacher Network 2 can be calculated as follows:

$$\mathcal{L}_{t2} = D_{KL} \left(\boldsymbol{p}_{f2}^{L}, \boldsymbol{p}_{f2}^{*} \right) + \lambda_{b} D_{KL} \left(\boldsymbol{p}_{b2}^{L}, \boldsymbol{p}_{b2}^{*} \right) + \lambda_{c} \sum_{l=1}^{L-1} D_{JS} \left(\boldsymbol{p}_{f2}^{l}, \boldsymbol{p}_{b2}^{L-l} \right)$$
(15)

After training the two teacher networks to converge, we can obtain the distribution of intermediate entities in the bidirectional reasoning process of the two teacher networks. Similar to He et al. [11], we take the average of the two distributions as the supervision signals:

$$\boldsymbol{p}_{t1}^{l} = \frac{1}{2} \left(\boldsymbol{p}_{f1}^{l} + \boldsymbol{p}_{b1}^{L-l} \right), l = 1, \dots, L-1$$
(16)

$$\boldsymbol{p}_{t2}^{l} = \frac{1}{2} \left(\boldsymbol{p}_{f2}^{l} + \boldsymbol{p}_{b2}^{L-l} \right), l = 1, \dots, L-1$$
(17)

where p_{t1}^l and p_{t2}^l are the intermediate supervision signals generated by Teacher Networks 1 and 2, respectively, and $p_{f1}^l(p_{b1}^{L-1})$ and $p_{f2}^l(p_{b2}^{L-1})$ denote the entity distributions for forward (backword) processes. Because BGNN-TT is forward-reasoning, the loss of BGNN-TT contains two parts, namely, the loss of forward reasoning and the loss of intermediate supervision signals in the two-teacher network.

The loss of forward reasoning can be calculated as

$$\mathcal{L}_1 = D_{KL} \left(\boldsymbol{p}_s^L, \boldsymbol{p}_f^* \right) \tag{18}$$

The loss of intermediate supervision signals in the two-teacher network can be calculated as

$$\mathcal{L}_{2} = \sum_{l=1}^{L-1} D_{KL} \left(\boldsymbol{p}_{s'}^{l} \, \boldsymbol{p}_{t}^{l} \right) \tag{19}$$

where p_s^l and p_t^l denote the entity distribution for forward reasoning and the teacher network in step *l*, respectively. The supervision signal p_t^l can be obtained by combining the supervision signals of the two teacher networks:

$$\boldsymbol{p}_{t}^{l} = \alpha \boldsymbol{p}_{t1}^{l} + (1 - \alpha) \boldsymbol{p}_{t2}^{l}$$

$$\tag{20}$$

where $\alpha \in (0, 1)$ is a parameter which controls the impact of the different teacher networks on the model.

Combining the above two losses, the total loss of the BGNN-TT model is defined as

$$\mathcal{L}_s = \mathcal{L}_1 + \lambda \mathcal{L}_2 \tag{21}$$

where $\lambda \in (0, 1)$ is a parameter to be learned. The full process is shown below in Algorithm 1. Algorithm 1: Pseudo-code for two-teacher knowledge distillation process.

Input: Training Dataset MOOC Q&A, parameters λ_c , λ_b , β , α , λ **Output:** BGNN-TT model

- 1 **for** *t* = 1,2 **do**
- 2 Create the teacher model t_i and initialize the parameters λ_c , λ_b , β .;
- 3 Training until model t_i converges;
- 4 Generate soft labels $p'_t(l = 1, ..., L 1)$ for the training data set MOOC Q&A;
- 5 end
- 6 repeat
- 7 Create the KGQA model BGNN-TT;
- 8 Initialize the parameters α , λ ;
- 9 Obtain the soft labels p_t^l using a weighted average calculation in Equation (20);
- 10 Utilize the loss function described in Equation (21) to optimize the model
 - BGNN-TT;
- 11 **until** the model BGNN-TT converges;

5. Experiment

This section primarily describes the experimental setup, including the datasets, parameter settings, comparison model, and evaluation metrics. Subsequently, we present the experimental findings and ablation study.

5.1. Data Set and Parameter Settings

We use the real MOOC Q&A [3] dataset to evaluate our approach. The answers to the questions are taken from MOOCCube [3], a sizable online educational knowledge graph created by Tsinghua University. Table 1 provides an example of the content of the dataset. The knowledge graph has a total of 52,195 triples, which are made up of ten types of relations such as concept_domain, course_video, and prerequisite_dependency, as well as seven types of entities, including 4723 users, 700 concepts, and 706 real online courses. Table 2 provides detailed information about MOOCCube. There are two different kinds of questions, one-hop questions and multi-hop questions, in the MOOC Q&A dataset. One-hop questions only include one head entity and one relation in the knowledge graph, whereas multi-hop questions might involve several entities and call for reasoning over several knowledge graph facts. The dataset contains 13,637 multi-hop questions and 5504 one-hop questions. When conducting the experiments, the training set, validation set, and test set were divided according to the ratios of 80%, 10%, and 10%, respectively. Table 3 displays the precise parameter settings for the BGNN-TT.

One-hop	Content				
Question	What are the papers related to instant messaging?				
Topic Entity	instant messaging				
	Session Initiation Protocol Extension for Instant Messaging				
Answer Entity	Teenage Communication in the Instant Messaging Era				
-	A model for Presence and Instant Messaging				
Multi-hop	Content				
Question	Which video from the 5G and AI course illustrates the digital communication system?				
Topic Entity	5G and AI				
Answer Entity	5G and AI, Chapter 3, Section 4				

Table 1. Examples of one-hop and multi-hop questions in the MOOC Q&A dataset.

Category	Types	Quantity
	concept	700
	course	706
	paper	5940
Entity	school	194
-	teacher	1933
	user	4723
	video	1405
	concept_field	44
	concept_paper	5927
	course_concept	10,346
	course_video	1591
Relation	school_course	703
Relation	school_teacher	2046
	teacher_course	2349
	user_course	24,933
	video_cncept	4040
	prerequisite_dependency	216

Table 2. Statistics of entities and relations in the MOOCCube knowledge graph.

Table 3. Parameter settings of BGNN-TT.

Parameters	Description	Values		
GloVe	The embedding dimension of GloVe.	300		
BiLSTM	The embedding dimension of BiLSTM.	50		
Drop out	The drop out rate of our method.	0.2		
Learning rate	The learning rate of our method.	0.0005		
α	The parameter in Equation (20)	0.5		
β	The parameter in Equation (7)	0.3		
λ	The parameter in Equation (21)	0.3		
λ_h	The parameter in Equation (14)	0.1		
λ_c	The parameter in Equation (14)	0.01		

5.2. Methods of Comparison and Evaluation Metrics

To validate the model, in this section we compare it with the mainstream benchmark models, including GraftNet [9], EmbedKGQA [26], and NSM [11].

GraftNet [9] is a model proposed by EMNLP 2018 for the KGQA task. It constructs a subgraph specific to the question based on the topic entity, introduces additional text to form a heterogeneous graph, iteratively updates the nodes of the heterogeneous graph using the improved GCN, and reasons the answer to the question.

NSM [11] is a model proposed by WSDM2021 for the KGQA task. It involves a sophisticated system which uses one single teacher to direct the reasoning process.

EmbedKGQA [26] is a model proposed by ACL2020 for the KGQA task. It conducts multi-hop reasoning through matching question embedding with pre-trained entity embedding obtained from ComplEx. We employ ComplEx [27], RotatE [44], DistMult [45], and ConvE [46] as score functions for answer prediction in our experiment.

We choose Hts@1, Hits@3, Hits@5, and mean reciprocal rank (MRR) as evaluation metrics. MRR represents the average inverse ranking of the entities in the prediction list, and Hits@K refers to the ratio of the actual answers when selecting the K highest scores in the predicted answers. These metrics are computed using the following formulas:

Hits@K =
$$\frac{1}{M} \sum_{q} I(p, \mathcal{E}_T)$$
 (22)

$$MRR = \frac{1}{M} \sum_{q} \left(\frac{1}{|\mathcal{E}_{T}|} \sum_{i \in \mathcal{E}_{\tau}} \frac{1}{r_{i}} \right)$$
(23)

where *M* denotes the number of questions, \mathcal{E}_T denotes the actual answer set of question *q*, *p* is the set of entities with the top K highest predicted score, the function *I* indicates whether the element is in the set, with range $\{0, 1\}$, and r_i denotes the ranking of entity *i* in the list of predicted answers.

5.3. Main Results

Table 4 displays a comparison of our model's performance with major methods on the MOOC Q&A dataset. Overall, it is clear that our method obtains excellent performance on both one-hop and multi-hop questions. Every statistic exceeds the comparison model except for multi-hop MRR, which is marginally behind (0.012) NSM. The most significant improvement is seen compared to RotatE, with the eight metrics improving by 0.124, 0.109, 0.104, 0.088, 0.228, 0.183, and 0.198, respectively. Similarly, in one-hop questions the Hits@1 and MRR improved by 0.004 and 0.019, respectively, compared to the optimal EmbedKGQA using ComplEx. Compared to the optimal NSM in multi-hop questions, Hits@1, Hits@3, and Hits@5 achieved improvements of 0.05, 0.052, and 0.043, respectively. After analysis, we found that the NSM model only uses the supervision signals from one teacher to guide the model, which is not comprehensive. However, by incorporating various teachers' knowledge (i.e., supervision signals), our models can gain more comprehensive knowledge and achieve better performance. Compared to GraftNet, for multi-hop questions the Hits@1, Hits@3, Hits@5, and MRR were improved by 0.052, 0.067, 0.053, and 0.014, respectively. After analysis, we found that the GraftNet model only utilizes Linear Aggregation to capture the graph structure information, which cannot reflect the context information between graph nodes, meaning that the effect is not good. However, by introducing Bilinear Aggregation, our model allows bidirectional feature propagation between each entity node and one of its neighboring nodes on a knowledge graph, which can obtain comprehensive entity representation and improve the reasoning ability.

Models	Score Function	One-Hop				Multi-Hop			
		Hits@1	Hits@3	Hits@5	MRR	Hits@1	Hits@3	Hits@5	MRR
EmbedKGQA [26]	ComplEx	0.947	0.957	0.968	0.693	0.793	0.812	0.826	0.531
	RotatE	0.832	0.854	0.869	0.624	0.623	0.639	0.689	0.413
	DistMult	0.913	0.928	0.937	0.647	0.764	0.779	0.793	0.469
	ConvE	0.929	0.937	0.941	0.651	0.773	0.782	0.796	0.481
GraftNet [9]	-	0.911	0.926	0.938	0.663	0.799	0.802	0.819	0.597
NSM [11]	-	0.938	0.949	0.958	0.688	0.801	0.817	0.829	0.623
BGNN-TT	-	0.956	0.963	0.973	0.712	0.851	0.869	0.872	0.611

Table 4. Performance comparison of different approaches on the MOOC Q&A dataset.

5.4. Ablation Study

We conducted relative ablation research to verify the viability of the two-teacher knowledge distillation approach proposed in this paper. Table 5 displays the experimental outcomes.

Table 5. Experimental comparison results under different teacher guidance approaches.

Models		One	e-Hop		Multi-Hop				
widdels	Hits@1	Hits@3	Hits@5	MRR	Hits@1	Hits@3	Hits@5	MRR	
BGNN-TT (two-teacher)	0.956	0.963	0.973	0.712	0.851	0.869	0.872	0.611	
BGNN-TT (only teacher 1)	0.949	0.957	0.963	0.701	0.842	0.857	0.876	0.614	
BGNN-TT (only teacher 2)	0.938	0.949	0.958	0.691	0.826	0.839	0.847	0.609	
BGNN-TT (without teacher)	0.912	0.924	0.938	0.674	0.814	0.825	0.841	0.593	

It is clear that our model obtains better performance when directed by Teacher Network 1, indicating that Teacher Network 1's supervision signal is stronger than that of Teacher Network 2. When incorporating the two teacher network supervision signals to direct the model, the best performance is obtained, demonstrating that two teachers can overcome the limitation of one teacher and provide stronger supervision signals. Similarly, compared to the model without teacher guidance, the eight metrics improved by 0.044, 0.039, 0.035, 0.038, 0.037, 0.031, and 0.018, respectively.

To analyze the effect of LA, BA-A, and BA-T on experimental performance, we compared the performance when using only Linear Aggregation (LA), only Bilinear Aggregation (BA-A or BA-T), and both. Table 6 shows that BGNN-TT with both BA-T and LA performs best across the board and utilizing only LA performs better than using only BA. When using both BA-T and LA, the eight metrics perform around 0.039, 0.039, 0.034, 0.035, 0.034, 0.045, 0.043, and 0.02 better, respectively, than when using only LA. Following this investigation, we discovered that adding bilinear aggregation permits bidirectional feature propagation between each entity node and one of its nearby nodes on a knowledge graph. Thus, it is possible to enhance multi-hop reasoning ability based on comprehensive entity representation by employing both Linear Aggregation and Bilinear Aggregation.

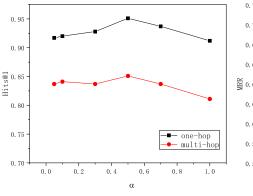
NG - 1-1-	One-Hop				Multi-Hop			
Models	Hits@1	Hits@3	Hits@5	MRR	Hits@1	Hits@3	Hits@5	MRR
BGNN-TT (both BA-T and LA)	0.956	0.963	0.973	0.712	0.851	0.869	0.872	0.611
BGNN-TT (both BA-A and LA)	0.939	0.947	0.957	0.703	0.837	0.841	0.859	0.607
BGNN-TT (only LA)	0.917	0.924	0.939	0.677	0.817	0.824	0.829	0.591
BGNN-TT (only BA-T)	0.609	0.924	0.938	0.674	0.574	0.597	0.601	0.447
BGNN-TT (only BA-A)	0.314	0.329	0.481	0.017	0.241	0.254	0.269	0.008

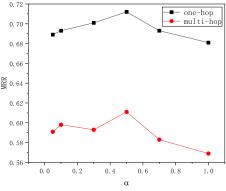
Table 6. Experimental comparison results using only BA, only LA, and both.

5.5. Impact of Different α and β Values

To examine the impact of various teacher network weights on the experimental performance, i.e., the values of the parameters in Equation (14), we tuned α among {0.05, 0.1, 0.3, 0.5, 0.7, 1.0}. As shown in Figure 6a,b, $\alpha = 0.5$ is optimal for Hits@1, $\alpha > 0.5$ results in a decline in Hits@1, and an increase in α somewhat enhance model performance when $\alpha < 0.5$. This phenomenon is seen with MRR as well. Therefore, $\alpha = 0.5$ is a good choice for our approach.

The ablation results shown in Table 6 show that the BGNN-TT model works best when both LA and BA-T are used. To investigate the effect of different weights of LA and BA-T, i.e., the values of the parameters in Equation (7), we tuned β among {0.1, 0.3, 0.5, 0.7, 0.9}. As shown in Figure 6c,d, $\beta = 0.3$ is good for both one-hop and multi-hop questions; the Hits@1 and MRR of both types of questions dramatically decline when $\beta > 0.3$. Overall, $\beta = 0.3$ is a good choice for our approach.





(a) Impact of varying α on Hits@1

(**b**) Impact of varying α on MRR

Figure 6. Cont.

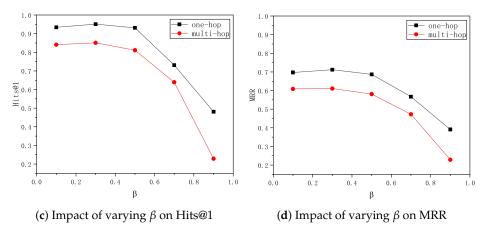


Figure 6. Impact of α and β on Hits@1 and MRR.

5.6. Case Study

In this section, we present a case study illustrating how the two teacher networks contribute to the reasoning process.

Considering the question "What concepts are covered in the Computer Networks course taught by Prof. Hong Wang?", the correct reasoning path is "Prof. Hong Wang" (teacher) \rightarrow teacher_course \rightarrow "Computer Networks" (course) \rightarrow course_concept \rightarrow TCP (concept). Figure 7 contrasts how the model performs with and without guidance from Teacher Network 1, Teacher Network 2, and both networks together.

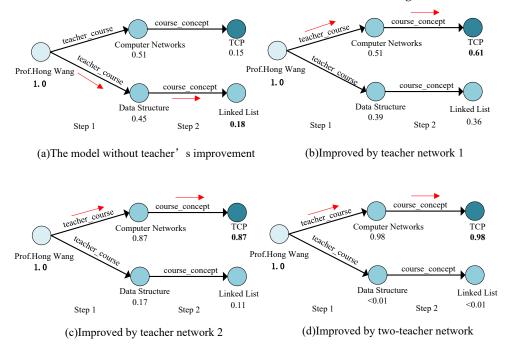


Figure 7. Case study from the MOOC Q&A dataset. Blue, light blue, and deep blue denote the topic entity, intermediate entities, and answer entity, respectively; the actual reasoning paths for various approaches are indicated by red arrows. For the sake of simplicity, we display only those entities with probabilities greater than or equal to 0.01. Bold numbers represent the highest probability of the entity at step 2.

As can be seen in Figure 7a, the initial model chooses the incorrect path, which leads to an irrelevant entity (Linked List). After improvement by Teacher Network 1 or 2, it assigns a probability of 0.51(0.87) to the entity "Computer Networks", and the model is able to find the correct answer entity "TCP" with a probability of 0.61(0.87) (Figure 7b,c).

In comparison, the two teacher networks together (Figure 7d) combine to enhance the intermediate entity distributions. As can be seen, when incorporating supervision signals from two teacher networks, the model assigns a very high probability of 0.98 to the entity "Computer Networks" in the first step. Therefore, it accurately locates the answer entity "TCP" with a high probability of 0.98 (Figure 7d).

This example demonstrates that incorporating the supervision signals from multiple teachers does in fact offer extremely helpful supervision signals in intermediate steps to improve the reasoning process.

6. Conclusions and Future Work

Recently, the exponential proliferation of online information has made it difficult for users to obtain information quickly and precisely. The emergence of intelligent questionanswering makes up for this shortcoming and provides convenience for sustainable urban living. Intelligent question-answering systems based on educational knowledge graphs can offer students a satisfying interactive experience and precise and knowledgeable tutoring services, improving sustainable learning efficiency.

In this paper, based on the intelligent question-answering scene in the education field in sustainable urban living, we propose a novel intelligent question-answering model over educational knowledge graph. Our approach uses a highly expressive bilinear graph neural network (BGNN) to capture graph structure information and perform forward reasoning. This approach can achieve more comprehensive entity representation and improve multihop reasoning ability using linear and bilinear aggregation in combination. In addition, to mitigate the spurious path reasoning phenomenon, we propose two-teacher knowledge distillation. We build two different teacher networks by combining forward and backward reasoning, then incorporate the supervision signals generated by both to guide our model. As a result, the teachers can compensate for a single teacher's limitations, which helps the model to acquire more comprehensive knowledge. Extensive experiments on the MOOC Q&A dataset demonstrate the effectiveness of our approach.

With respect to limitations and future work, although our model achieves good performance, it has flaws that need to be improved and perfected. It cannot handle questions requiring temporal and numerical comparisons. For instance, when asked "What is the newest literature on simulated annealing algorithm?", the model can locate literature related to the simulated annealing algorithm in the KG; however, it cannot determine the most recent literature. Therefore, future work should consider further improving the method for complex questions requiring temporal comparison capability by introducing a timeaware mechanism. In addition, the professionalism and size of the knowledge graph itself significantly impact the effectiveness of the graph neural network's reasoning. The more comprehensive the entity relationship contained in the knowledge graph, the better the Q&A effect based on the graph neural network. In addition to paying attention to the construction of relevant knowledge graphs, future intelligent education question-answering should try to apply emerging methods in deep learning, such as multi-task learning strategies and comparative learning. We believe that with the improvement of knowledge graph question-answering technology, the quality of question-answering texts in urban living can be effectively controlled, providing students with an intelligent instant tutoring service and promoting sustainability in urban living.

Author Contributions: Conceptualization, Y.F. and J.D.; methodology, Y.F.; software, J.D.; validation, Y.F., J.D. and H.W.; formal analysis, Y.F.; investigation, J.D.; resources, H.W., F.Z.; data curation, H.W., F.Z.; writing—original draft preparation, Y.F.; writing—review and editing, J.D.; visualization, Y.F., F.Z.; supervision, H.W., F.Z.; project administration, H.W., F.Z.; funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 81660031.

Institutional Review Board Statement: Not applicable.

17 of 18

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank for Y.Z. for critically reviewing the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Liu, Q.; Chen, E.H.; Zhu, T.Y.; Huang, Z.Y.; Wu, R.Z.; Su, Y.; Hu, G.P. Research on educational data mining for online intelligent learning. *Pattern Recognit. Artif. Intell.* **2018**, *31*, 77–90.
- Zheng, Q.H.; Dong, B.; Qian, B.; Tian, F.; Wei, B.; Zhang, W.; Liu, J. The state of the art and future tendency of smart education. J. Comput. Res. Dev. 2019, 56, 213–228.
- Yu, J.F.; Luo, G.; Xiao, T.; Zhong, Q.Y.; Wang, Y.Q.; Feng, W.Z.; Luo, J.Y.; Wang, C.Y.; Hou, L.; Li, J.Z. et al. MOOCCube: A large-scale data repository for NLP applications in MOOCs. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, PA, USA, 5–10 July 2020; pp. 3135–3142.
- 4. Xu, Y.J.; Guo, J. Recommendation of personalized learning resources on K12 learning platform. *Comput. Syst. Appl.* **2020**, 29, 217–221.
- Agarwal, A.; Sachdeva, N.; Yadav, R.K.; Udandarao, V.; Mittal, V.; Gupta, A.; Mathur, A. EDUQA: Educational domain question answering system using conceptual network mapping. In Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 12–17 May 2019; pp. 8137–8141.
- Yih, W.T.; Richardson, M.; Meek, C.; Chang, M.W.; Suh, J. The value of semantic parse labeling for knowledge graph question answering. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; pp. 201–206.
- Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data, Vancouver, BC, USA, 10–12 June 2008; pp. 1247–1250.
- 8. Kipf, T.N.; Welling, M. Semi-supervision classification with graph convolutional networks. arXiv 2016, arXiv:1609.02907.
- Sun, H.; Dhingra, B.; Zaheer, M.; Mazaitis, K.; Salakhutdinov, R.; Cohen, W.W. Open domain question answering using early fusion of knowledge bases and text. arXiv 2018, arXiv:1809.00782.
- 10. Sun, H.; Bedrax-Weiss, T.; Cohen, W.W. Pullnet: Open domain question answering with iterative retrieval on knowledge bases and text. *arXiv* **2019**, arXiv:1904.09537.
- He, G.; Lan, Y.; Jiang, J.; Zhao, W.X.; Wen, J.R. Improving multi-hop knowledge base question answering by learning intermediate supervision signals. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining, Jerusalem, Israel, 8–12 March 2021; pp. 553–561.
- 12. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. arXiv 2015, arXiv:1503.02531.
- 13. Zhu, H.; Feng, F.; He, X.; Wang, X.; Li, Y.; Zheng, K.; Zhang, Y. Bilinear graph neural network with neighbor interactions. *arXiv* **2020**, arXiv:2002.03575.
- 14. Xu, C.; Tao, D.; Xu, C. Multi-view intact space learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 2015, 37, 2531–2544 [CrossRef]
- 15. Berant, J.; Chou, A.; Frostig, R.; Liang, P. Semantic parsing on freebase from question-answer pairs. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1533–1544.
- 16. Cai, Q.; Yates, A. Large-scale semantic parsing via schema matching and lexicon extension. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Sofia, Bulgaria, 4–9 August 2013; Volume 1: Long Papers, pp. 423–433.
- 17. Yih, W.; He, X.; Meek, C. Semantic parsing for single-relation question answering. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, MD, USA, 22–27 June 2014; Volume 2: Short Papers, pp. 643–648.
- Hu, S.; Zou, L.; Zhang, X. A state-transition framework to answer complex questions over knowledge base. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2098–2108.
- Luo, K.; Lin, F.; Luo, X.; Zhu, K. Knowledge base question answering via encoding of complex query graphs. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 2185–2194.
- Bao, J.; Duan, N.; Yan, Z.; Zhou, M.; Zhao, T. Constraintbased question answering with knowledge graph. In Proceedings of the 26th International Conference on Computational Linguistics, Osaka, Japan, 11–16 December 2016; pp. 2503–2514.
- Yih, W.-t.; Chang, M.-W.; He, X.; Gao, J. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing, China, 26–31 July 2015; Volume 1: Long Papers, pp. 1321–1331.
- 22. Bordes, A.; Chopra, S.; Weston, J. Question Answering with Subgraph Embeddings. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014; pp. 615–620.

- Bordes, A.; Weston, J.; Usunier, N. Open question answering with weakly supervision embedding models. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Nancy, France, 15–19 September 2014; pp. 165–180.
- 24. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating embeddings for modeling multi-relational data. *Adv. Neural Inf. Process. Syst.* 2013, 26.
- 25. Huang, X.; Zhang, J.; Li, D.; Li, P. Knowledge graph embedding based question answering. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, Melbourne, Australia, 11–15 February 2019; pp. 105–113.
- Saxena, A.; Tripathi, A.; Talukdar, P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 4498–4507.
- Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, E.; Bouchard, G. Complex embeddings for simple link prediction. In Proceedings of the International Conference on Machine Learning, New York City, NY, USA, 19–24 June 2016; pp. 2071–2080.
- 28. Miller, A.; Fisch, A.; Dodge, J.; Karimi, A.H.; Bordes, A.; Weston, J. Key-value memory networks for directly reading documents. *arXiv* **2016**, arXiv:1606.03126.
- 29. Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R. Gated graph sequence neural networks. arXiv 2015, arXiv:1511.05493.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; Bengio, Y. Graph attention networks. *arXiv* 2017, arXiv:1710.10903.
 Schlichtkrull, M.; Kipf, T.N.; Bloem, P.; Berg, R.v.d.; Titov, I.; Welling, M. Modeling relational data with graph convolutional
- networks. In Proceedings of the European Semantic Web Conference, Monterey, CA, USA, 8–12 October 2018; pp. 593–607.
 32. Cai, J.; Zhang, Z.; Wu, F.; Wang, J. Deep Cognitive Reasoning Network for Multi-hop Question Answering over Knowledge Graphs. In Proceedings of the The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), Bangkok, Thailand, 1–6 August 2021, pp. 219–229.
- Shi, J.; Cao, S.; Hou, L.; Li, J.; Zhang, H. TransferNet: An effective and transparent framework for multi-hop question answering over relation graph. arXiv 2021, arXiv:2104.07302.
- Hsieh, Y.Z.; Lin, S.S.; Luo, Y.C.; Jeng, Y.L.; Tan, S.W.; Chen, C.R.; Chiang, P.Y. ARCS-assisted teaching robots based on anticipatory computing and emotional big data for improving sustainable learning efficiency and motivation. *Sustainability* 2020, 12, 5605. [CrossRef]
- 35. Ren, X.; Yang, W.; Jiang, X.; Jin, G.; Yu, Y. A Deep Learning Framework for Multimodal Course Recommendation Based on LSTM+ Attention. *Sustainability* **2022**, *14*, 2907. [CrossRef]
- 36. Lin, Q.K.; Zhu, Y.; Lu, H.; Shi, K.; Niu, Z. Improving university faculty evaluations via multi-view knowledge graph. *Future Gen. Comput. Syst.* **2021**, *117*, 181–192. [CrossRef]
- 37. Lin, Q.; Zhang, L.; Liu, J.; Zhao, T. Question-aware Graph Convolutional Network for Educational Knowledge Base Question Answering. J. Front. Comput. Sci. Technol. 2021, 15, 1880–1887.
- Yang, Z.; Wang, Y.; Gan, J.; Li, H.; Lei, N. Design and research of intelligent question-answering (Q&A) system based on high school course knowledge graph. *Mobile Netw. Appl.* 2021, 26, 1884–1890.
- 39. Zhao, W.; Liu, J. Application of Knowledge Map Based on BiLSTM-CRF Algorithm Model in Ideological and Political Education Question Answering System. *Mobile Inf. Syst.* 2022, 2022, 4139323. [CrossRef]
- 40. Han, J.; Moraga, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In Proceedings of the International Workshop on Artificial Neural Networks, Perth, Australia, 27 November–1 December 1995; pp. 195–201.
- 41. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 26–28 October 2014; pp. 1532–1543.
- 42. Kullback, S.; Leibler, R.A. On information and sufficiency. Ann. Math. Stat. 1951, 22, 79–86. [CrossRef]
- 43. Fuglede, B.; Topsøe, F. Jensen-Shannon divergence and Hilbert space embedding. In Proceedings of the International Symposium on Information Theory, Chicago, IL, USA, 27 June–2 July 2004.
- 44. Sun, Z.Q.; Deng, Z.H.; Nie, J.Y.; Tang, J. RotatE: Knowledge graph embedding by relational rotation in complex space. In Proceedings of the 7th International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
- 45. Yang, B.; Yih, W.; He, X.D.; Gao, J.F.; Deng, L. Embedding entities and relations for learning and inference in knowledge bases. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- 46. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2d knowledge graph embeddings. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.