Deep Learning Meets Knowledge Graphs: A Comprehensive Survey

Shuo Yu
School of Computer Science and Technology, Dalian University of Technology
Chengchuan Xu
School Software, Dalian University of Technology
Xiaomei Bai (xiaomeibai@outlook.com)
Computing Center, Anshan Normal University
Ranjith Kuncheerathodi
School of Computer Science and Engineering, Amrita Vishwa Vidyapeetham University
Selena Firmin
Institute of Innovation, Science and Sustainability, Federation University Australia
Feng Xia
Institute of Innovation, Science and Sustainability, Federation University Australia

---

Research Article

Keywords: knowledge graphs, deep learning, knowledge graph representation

Posted Date: September 6th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2021923/v1

Deep Learning Meets Knowledge Graphs: A Comprehensive Survey

Abstract

Knowledge Graphs (KGs) which can encode structural relations connecting two objects with one or multiple related attributes have become an increasingly popular research direction. Given the superiority of deep learning in representing complex data in continuous space, it is handy to represent KGs data, thus promoting KGs construction, representation, and application. This survey article provides a comprehensive overview of deep learning technologies and KGs by exploring research topics from diverse phases of the KGs lifecycle, such as construction, representation, and knowledge-aware application. We propose new taxonomies on these research topics for motivating cross-understanding between deep learning and KGs. Based on the above three phases, we classify the different tasks of KGs and task-related methods. Afterwards, we explain the principles of combing deep learning in various KGs steps like KGs embedding. We further discuss the contribution and advantages of deep learning applied to the different application scenarios. Finally, we summarize some critical challenges and open issues deep learning approaches face in KGs.

Keywords: knowledge graphs, deep learning, knowledge graph representation.

1 Introduction

There has been a rapid growth in the importance of Knowledge Graphs (KGs) along with their wide application to question-answering systems, search engines, and recommendation systems [1–3]. In the era of knowledge engineering, a large number of KGs, such as Freebase [4], YAGO [5], and NELL [6], have been developed gradually. KGs are large networks of real-world entities described in terms of their semantic types and their relations to each other, representing facts in the form of triples (head entity, relation, tail entity). Furthermore, KGs can mine, organize, and manage knowledge from large-scale data to improve the effectiveness of information usage and provide users with more innovative services.

According to popular mainstream research, we identify a series of problems in the lifecycle of knowledge graphs from construction to application. In the first phase of building KGs, extracting knowledge efficiently and reducing erroneous knowledge generated during this fact-extracting process have become significant issues [7]. After the KGs have been created, it is also a challenge to handle the dynamically growing facts and improve the scalability of KGs. In addition, although the current knowledge graph contains tens of millions of entities and relations, most suffer from incompleteness and noise [8]. How to correct the facts and complete KGs has currently become a popular research topic. Finally, researchers have to consider integrating advanced KGs related technologies with different real-life scenarios for successful applications.

In the light of the successful application of deep learning to graph learning areas, it can encode and represent graph data into vectors in continuous space to extract the desired features of a graph [9–11]. Recently, there has been an increasing interest in extending deep learning approaches for KGs motivated by convolutional neural networks (CNNs), recurrent neural networks (RNNs), and autoencoders from deep learning. Some new neural network architectures have been rapidly developed to handle the complex data of knowledge graphs over the past few years.

This article conducts a comprehensive survey of current literature on deep learning for KGs, proposing a new classification criterion, contributing significantly to the construction, representation, and application of knowledge graphs. Our main contributions are summarized as follows.

- **Comprehensive Overview:** We comprehensively overview each phase of the KGs lifecycle and the deep learning methods proposed. Besides, we have explained and compared the more insightful methods in detail. In addition, we analyze the contribution and advantages of deep learning methods to the different application scenarios.
- **New Taxonomies:** We propose a new classification criterion based on the process of KGs. Specifically, we refine the lifecycle of a knowledge graph into three phases: knowledge graph construction, knowledge representation, and knowledge-aware application. Then, based on the above three phases, we classify the different tasks of KGs and task-related methods. This new classification method is more similar to human cognition and can better help researchers become familiar with the KGs domain.
- **Broad Audience:** We provide comprehensive coverage on advanced research topics, which is suitable for most researchers, especially academics new to KGs. The knowledge-aware applications section covers extensive research topics from distinct disciplines, which provides insights for scholars from both computer science and other research domains.
- **Future Directions:** This survey summarizes the challenges and open issues of deep learning approaches in knowledge graphs, which sheds light on potentially valuable future research directions. Specifically, knowledge transfer in cross-domain, interactive learning and cognitive learning, scalability, dynamics, anomaly detection, and explainability are illustrated.

Table 1 Related Surveys

Survey	Aspects	Contribution
Hur et al. [12]	KGs construction	This paper critiques state-of-the-art automated techniques to produce knowledge graphs of near-human quality autonomously.
Lin et al. [13]	Knowledge representation learning (KRL)	This paper introduces the motivations for KRL, and an overview of the existing approaches with applications of KRL.
Zou et al. [14]	KGs application	This paper introduces KGs applications stemming from different domains, and points out advancements of applying KGs.
Abu et al. [15]	Domain-specific KGs	This paper offers a comprehensive definition of a domain-specific KGs and the state-of-the-art approaches drawn from academic works.
Ji et al. [16]	Knowledge acquisition, KRL, Temporal KGs, KGs application	This paper focuses on the classification of KGs related technologies and provides a comprehensive review of KGs.

The rest of this paper is organized as follows. Section 2 presents the definition of KGs and a brief introduction to deep learning on KGs. Deep learning in knowledge graph construction, representation, and application are respectively introduced in Section 3, Section 4, and Section 5. Challenges and open issues are come up with in Section 6. Finally, Section 7 concludes the paper.

2 An Overview of Knowledge Graphs

2.1 Related Surveys

As shown in Table 1, previous surveys on deep learning approaches for knowledge graphs have mainly unfolded against merely representative aspects of KGs, such as knowledge graph construction [12], knowledge representation learning [13], knowledge graph application [14, 17] or domain-specific knowledge graphs [15]. Based on the mentioned surveys, this survey focuses on introducing the knowledge graph systematically or comprehensively. Of course, previous researchers have also presented surveys that comprehensively introduce the knowledge graph. For example, Ji et al. [16] provides a more full-scaled view from fourfold and goes deeper into the flow of KRL. However, like previous reviews, its taxonomies focus too much on the elaboration and introduction of technologies, which is more suitable for researchers who have some insights into fields related to KGs. In contrast, this survey starts from the flow of the KGs lifecycle, focusing on the relevance of various phases from construction to application, and comprehensively introduces deep learning-related methods in each phase. In summary, compared with previous surveys, our survey has a more comprehensive introduction to deep learning and KGs.

2.2 Deep Learning for Knowledge Graphs

This survey follows a phase decomposition of the knowledge graphs lifecycle. It provides a comprehensive literature review of research related to deep learning at each phase, including knowledge graphs construction, knowledge representation, and knowledge-aware applications, where the taxonomy structures are illustrated in Fig. 1.

Knowledge Graphs Construction: Knowledge graph construction is the foundation for the following applications. The prerequisite for construction is the precise extraction of knowledge from structured and unstructured data sources. The current mainstream research directions for knowledge extraction mainly include relation extraction, entity extraction, and attribute extraction. Various deep neural network (DNN) based methods are widely used for relation extraction tasks, including graph neural network (GNN), CNN, RNN, etc [18]. Entity extraction is divided into entity discovery/recognition, entity disambiguation, and entity resolution. Attribute extraction is the task of adding auxiliary information to the knowledge graphs by extracting the attributes of entities, of which the forms include neural networks, multi-model, and others.

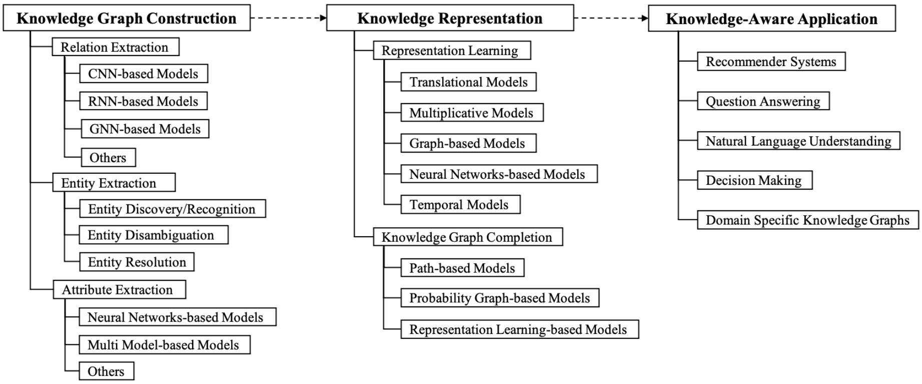


Fig. 1 Categorization of deep learning on knowledge graphs through lifecycle.

Knowledge Representation: Knowledge representation is essential for further utilization of the acquired knowledge that provides the computational conditions and theoretical possibilities for downstream applications [19–21]. We categorize deep learning for knowledge representation into knowledge graphs representation learning and knowledge graphs completion. The former is a critical research issue for capturing plenteous semantic information of elements in KGs, including translational, multiplicative, graph-based, neural networks, and temporal models. The latter is the task of predicting missing facts in KGs, and the relevant deep learning methods can be divided into three types: the first is traditional path-based reasoning method, the second type is the probability graph model, and finally, the model based on representation learning [22].

Knowledge-Aware Application: Deep learning in knowledge-aware applications includes recommender systems, question answering, natural language understanding, decision making, and domain-specific Knowledge Graphs (e.g., scholarly, biomedical, musical).

2.3 Notations and Definitions

The definition of a knowledge graph will change with its orientation, leading to a diverse definition of it. For example, the definition of a domain-specific knowledge graph will emphasize the role of ontologies in building the graph. However, for a domain-independent knowledge graph, its definition will stress the importance of multiple kinds of triples. In this survey, we follow the pervasive approach of defining the knowledge graph, KG $\mathcal{G}$, considered to contain a collection of triples $\{(h, r, t)\} \subseteq \mathcal{N} \times \mathcal{R} \times \mathcal{N}$, in which the entity set is denoted as $\mathcal{N}$ while the relation set is denoted as $\mathcal{R}$. Specific notations and their descriptions are listed in Table 2.

Table 2 Notations and Definitions

Notations	Definitions
$\mathcal{N}$	The set of entities in KG
$\mathcal{R}$	The set of relations in KG
$\mathcal{G}$	A knowledge graph $\mathcal{G} = \{\mathcal{N}, \mathcal{R}\}$
$\mathcal{T}$	The set of tasks
(h, r, t)	Triple of head, relation and tail
(h, r, t, τ)	Triple added by a temporal dimension
$(\mathbf{h}, \mathbf{r}, \mathbf{t})$	Embedding of head, relation and tail
$\mathbf{M}$	Transition matrix
$\mathbf{I}$	Identity matrix
$\mathbf{h}^\top$	Transpose vector
$\mathbf{h}_\perp$	Projected vector
$f_r(\cdot), g(\cdot)$	Score function
$\sigma(\cdot)$	Activation function
$\cdot_{l1/2}$	L1 or L2 norm
$\mathbb{R}^d$	d-dimensional Euclidean space
$\mathbb{C}^d$	d-dimensional complex space
$\mathbf{concat}(\cdot), [\mathbf{h};\ \mathbf{t}]$	Connection operator
$\langle h, r \rangle$	Hadmard product
$\mathbf{h} \circ \mathbf{r}$	Hadmard product
$\mathbf{h} \star \mathbf{t}$	Circular correlation
$\mathbf{diag}(\cdot)$	Connection operator
ω	Convolutional filters
$*$	Convolution operator

3 Deep Learning for Knowledge Graphs Construction

3.1 Relation Extraction

Relation extraction (RE) is the key subtask to construct large-scale KGs, which aims to extract entity pairs from natural language text and recognize the semantic relationships between them. Traditional methods [23–28] have been put forward to achieve relation extraction tasks. However, those methods were highly counting on the manual design features. They encountered problems such as high time complexity, which may be challenging to apply to large-scale relation extraction tasks. In recent years, with the widespread application of deep learning technology, it has been widely used automatically in extraction tasks and many new neural models have been set forth, many of which are listed in Table 3.

Deep neural networks, including CNNs, RNNs, and GNNs, have been diffusely applied in relation extraction and other knowledge graphs construction fields. There are apparent advantages of convolutional neural networks in feature extraction. Liu et al. [29] first applied a convolutional network to relation extraction by incorporating lexical features. Then based on the previous work, Zeng et al. [30] explored a CNN-based model with position features, which encoded the relative distances to entity pairs. However, this model only considers local features but ignores global attributes. The method proposed by Nguyen et al. [31] utilizes multiple window sizes filters for relation extraction while considering CNN as an encoder. Santos et al. [32] developed a novel loss function while ranking CNN to tackle the relation classification task. As shown in Fig. 2(a), PCNN [33] aims to capture structural information between two entities by devising a piecewise max-pooling layer. To capture sentence-level interaction information for more accurate rates of relation extraction, Jiang et al. [34] realized the information sharing between the different sentences and dealt with the multi-label characteristics of relation extraction.

Despite the CNN-based approaches having been successful, researchers found that CNN cannot handle temporal sequence, particularly for the long-distance dependency between entity pairs. So recurrent neural network is introduced for its superiority in sequential data modeling. Zhang et al. [35] adopted bi-directional RNN (BiRNN) to model long-distance relation patterns. Like the bi-directional structure of BiRNN, bidirectional long short term memory (BLSTM) [36] is designed to deal with the relation extraction tasks by mining the sentence-level representation. Miwa and Bansal [37] proposed a novel end-to-end model using LSTMs, which was based on sequences and tree structures.

Graph convolutional network (GCN) is also constantly applied to relation extraction for its superiority in capturing contextual relationships and structural features. The syntactic dependency tree contains the dependency relationships between words in a sentence. Introducing GCN into the relation extraction task can mine deeper semantic information. C-GCN [38] pools information over arbitrary dependency structures efficiently in parallel by extending GCN. RESIDE [39] utilized GCN to encode grammatical information in the text. AGGCNs [40] take full dependency trees as inputs in an end-to-end manner by applying GCN. Due to the importance of long-tail relations, few long-tail imbalance data are available. Zhang et al. [41] applied GCN to learn relational knowledge. Recently, MrGCN [42] is designed to accomplish discourse-level RE tasks, along with a novel graph pooling approach.

Attention mechanisms and other deep learning techniques are widely integrated into relation extraction tasks. For example, to alleviate the wrong labeling problem, PCNN+ATT [43] is introduced to represent the semantics of sentences for RE, which is a kind of sentence-level attention-based model. At the same time, Zhou et al. [44] applied attention mechanisms and BLSTM networks to relation classification, aiming to capture the semantic information. Considering multi-lingual data for relation extraction, MNRE [45] utilized mono-lingual attention to select sentences within a language and a cross-language attention mechanism to take advantage of the consistency and complementarity of different languages. To deal with the wrong labeling problem mentioned before, Wang et al. [46] utilized the prior knowledge from KGs and supervised the learning process directly. Reinforcement learning is a branch of machine learning, and it emphasizes how to act based on the environment to maximize the expected benefits. Qin et al. [47] proposed a reinforcement learning-based model for distant supervision RE, which can deal with the false-positive problem. Due to noise labeling problems in distant supervision relation extraction, DSGAN [48] firstly considers adversarial learning. The model utilizes a generator and a discriminator that distinguish positive and negative data samples. In addition to the methods mentioned above, other deep learning techniques, for example, deep residual learning, can also be applied for relation extraction [49].

3.2 Entity Extraction

3.2.1 Entity Discovery/Recognition

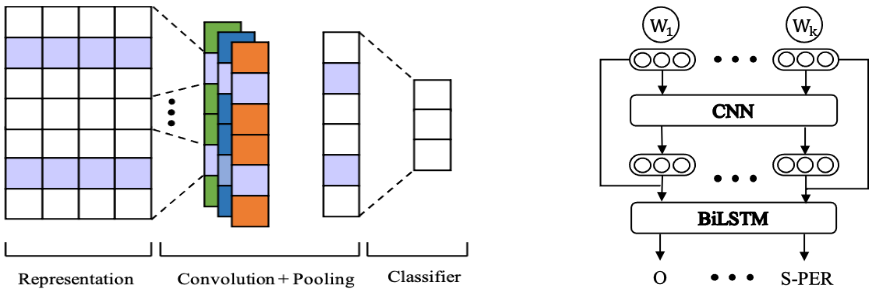
The entity is the essential component of knowledge graphs, and entity recognition, also known as named entity recognition (NER), plays a major role in knowledge graph construction. NER is the process of locating the boundaries of named entities in text and classifying them into a set of predefined types. Traditional methods can be divided into three categories, including supervised learning methods, semi-supervised learning methods, and other methods based on rules, dictionaries, and online knowledge bases. As the applicability of traditional methods declines, many other techniques based on deep learning are constantly proposed. For example, Lample et al. [50] proposed a neural architecture that does not rely on term resources or features but only on small-scale supervised training data and unannotated corpora. For decoupling the work of feature engineering and increasing the applicability and usability of the model, Jason and Eric [51] proposed a new neural network structure, as illustrated in Fig. 2(b). They used BiLSTM and CNN hybrid structure to generate and measure word-level and character-level features, thus eliminating the need for most feature engineering. Andrej et al. [52] developed a novel framework based on parallel recurrent neural networks for NER. The motivation lies in its ease of distribution and ability to reduce the total number of parameters. MGNER [53] is aimed to deal with multi-granularity named entity recognition. This task refers to situations where multiple entities in a sentence do not overlap or are entirely nested. Li et al. [54] designed a unified model for NER, which can deal with both flat and nested NER tasks. For domain-specific entity recognition tasks, K-BERT [55] can add domain knowledge into a model; unsurprisingly, it outperforms BERT [56]. Recently, Nie et al. [57] proposed a novel knowledge-aware model for NER, tackling the heterogeneity issue between NER and KB-type systems.

3.2.2 Entity Disambiguation

Entity disambiguation, also known as entity linking, is the task of matching the ambiguous entities to the corresponding entities in the knowledge graphs, which is the first step to making machines understand natural language. For instance, Jingjing Guo, a former member of China's diving team, won the gold medal at the 2004 Athens Olympics. The name Jingjing Guo should be linked to the entity of Jingjing Guo in the KG. The current entity disambiguation approaches are constantly being proposed through deep learning-based work

Table 3 A summary of relation extraction approaches

Category	Method	Architecture	Brief Problem or Contrubution
CNN-based	Liu et al. [29]	CNN + synonym coding	Semantic knowledge aggregation among words
	O-CNN [30]	CNN + max-pooling	Capturing lexical and sentence-level features
	Multi-CNN [31]	CNN + multiple window sizes + max-pooling	Unbalanced corpus
	CR-CNN [32]	CNN + max-pooling + ranking loss	Efficiency improvement of artificial classes
	PCNN [33]	CNN + piecewise max-pooling	Noise labeling problem
	MIMLCNN [34]	CNN + cross-sentence max-pooling	Multi-label nature of distant supervised relation extraction
RNN-based	RNN+PI [35]	RNN + max-pooling + position indicators	Long-distance pattern learning problem
	BLSTM [36]	Bi-LSTM + max-pooling	Sentence level representation
GCN-based	C-GCN [38]	GCN + path-centric pruning	Dependency tree optimization and efficiency improvement
	RESIDE [39]	GCN + additional supervision	Modeling syntactic information
	AGGCNs [40]	GCN + multi-head attention + dense layer	Irrelevant information from the dependency trees
	KATT [41]	GCN + CNN+ attention	Long-tail problem of relations
	MrGCN [42]	GCN + pooling	Learn larger receptive fields
Other	PCNN+ATT [43]	Attention + PCNN	Noise labeling problem
	Att-BLSTM [44]	Attention + BiLSTM	Capturing sentence-level semantic information
	MNRE [45]	Attention + CNN	Multi-lingual data for relation extraction
	DSGAN [48]	Attention + PCNN/CNN + GAN	Noise labeling problem
	ResCNN-x [49]	Residual connection + CNN-x	Distant supervised noisy relation extraction



(a) PCNN

(b) BiLSTM-CNN

Fig. 2 Illustrations of some knowledge graph construction tasks: (a) Relation extraction with PCNN; (b) Entity recognition with BiLSTM-CNN.

to do entity linking. For example, Huang et al. [58] came up with a fresh deep semantic relatedness model based on deep neural networks to measure entity semantic relatedness. Ganea et al. [59] put forward a novel deep learning model for joint document-level entity disambiguation, which leverages a neural attention mechanism over local context windows to learn the representation. To measure the semantic matching between the mention's context and target entity, Phan et al. [60] developed a deep learning model by utilizing the attention mechanism and LSTM for entity disambiguation. Cao et al. [61] proposed a new neural model named neural collective entity linking (NCEL) for encoding local contextual features and global coherence information by employing a GCN. Through thinking of connections as underlying variables, Le and Titov [62] induced relations by optimizing the entity-linking model in an end-to-end manner. Recently, Bootleg [63] proposed a self-supervised model for entity disambiguation tasks.

3.2.3 Entity Resolution

Entity resolution refers to identifying all mentions of the same real-world entity within a knowledge base or across multiple knowledge bases [64]. As the real world develops, KGs gradually expand, making semantic relationships between entities a greater challenge to overcome. Entity resolution can handle the above problems and reduce the complexity by duplicating and linking entities and proposing canonicalized references to entities. Meanwhile, entity resolution has been applied in various domains, such as finance and biology. Entity resolution may be named differently in different studies, including deduplication [65], linking discovery [66], record linkage [67–69]. Traditional methods of entity resolution were based on distance or similarity. They then used machine learning techniques to determine whether two entities are the same, for example, the method [70] proposed by Konda et al. However, deep learning-based models for entity resolution have been demonstrated successful in gaining better performance. Zhu et al. [64] modeled the entity resolution problem as a multi-type graph summarization problem. Meanwhile, the method was based on similarity measures, which were not only limited to simple distance-based metrics but also structural similarity. Deep entity resolution (DeepER) [71] switched each tuple to distributed representation vector by using recurrent neural network with long short term memory to capture similarities. Mudgal et al. [72] showed that deep learning solutions were significantly outperformed existing frameworks on unstructured textual data. Li et al. [73] developed a novel entity resolution model named graph entity resolution (GraphER) using graph convolutional neural networks, aiming to handle structured entity resolution problems in a token-centric manner.

3.3 Attribute Extraction

Attribute extraction is the task of extracting entity attributes of which the list is constructed by extracting the attribute name and value of the entity from the original data of different information sources. When forming a triple, the attributes of an entity can be thought of as a relationship between the entity and its attribute values, which can be represented by (entity, attribute, attribute value). Attribute extraction includes neural network-based attribute extraction and other forms of attribute extraction, for example, multi-modal attribute extraction. Whether the data is tagged or not, attribute extraction can be divided into supervised, unsupervised, and semi-supervised attribute extraction.

Alternatively, attribute extraction can be regarded as special relation extraction. Still different from general relation extraction, the difficulty of attribute extraction compared with relation extraction lies in recognizing the entity's attributes and attribute values. However, the structure of attribute value is uncertain, so most studies are based on rule or pattern extraction. ReNoun [74] uses a dependency parsing-based pattern discovery approach to achieve attribute extraction for long-tail entities by inductively extracting patterns from the training set through a remotely supervised approach. The parsing results, however, lose the rich context around the entities in the schema, and the process is costly for a large corpus. MetaPAD [75] aimed to mine a novel typed textual pattern structure, called meta pattern, which is extended to a frequent, informative, and precise subsequence pattern from a specific context of massive text corpora. In addition, attribute extraction based on neural networks can be transformed into sequence tagging or machine reading comprehension tasks. For example, Zhao et al. [76] used a sequence tagging model incorporating BERT to extract attributes from medical texts. Unlike most research about attribute extracting information from text data, it is worth mentioning that Robert et al. [77] introduced multi-modal attribute extraction.

4 Deep Learning for Knowledge Representation

A network is composed of a set of nodes and edges, commonly employed to represent data in the real world. The effective representation of network features can significantly improve the performance of downstream tasks [78]. Knowledge representation is an abstract representation of the real world, and storage is the foundation of knowledge graph construction, management, and application. The vast amount of knowledge in the knowledge graphs from the real world can only be processed by computers after it is appropriately represented. At first, the knowledge representation methods include Predicate Logic, Production Rule, Frame Systems, Probabilistic Graphical Model, and so on. With the development of science, standard knowledge representation languages have been constantly set forth, for example, RDF, XML, OWL, and so on. However, previous knowledge representation methods are based on symbolic logic, which can depict explicit and discrete knowledge but cannot represent a large amount of knowledge in the real world that is not easy to be explained by symbolic logic. Much deep learning-based knowledge representation methods are used to overcome the disadvantages of conventional methods and effectively mine and analyze the semantic relations between knowledge entities. This section reviews recent advances in deep learning in knowledge representation which is divided into knowledge graph representation learning and knowledge graph completion.

4.1 Knowledge Graph Representation Learning

Over the past years, we have witnessed a rapid growth in knowledge graph representation learning (KGRL), also known as knowledge graph embedding (KGE). Network representation learning (NRL) is an effective graph analytics technique and promotes users to deeply understand the hidden characteristics of graph data [79]. Similar to the work of NRL on assigning low-dimensional representations to nodes, KGRL aims to map entities and relations into low-dimensional dense real-valued vectors while capturing their plenteous semantic information for completing downstream tasks and applications. Meanwhile, due to the similarity when handling features of graph-structured data, some classical approaches of NRL including DeepWalk [80], Node2vec [81], Line [82], SDNE [83] and GMM-based models such as GCN [84] and GraphSage [85], also provide some novel insight for KGRL. However, NRL focuses on retaining topological structure information in the representation space without learning relational representation [86]. KGRL emphasizes head-to-tail relation, learning node, and relation representations based on preserving structural information. A knowledge graph is a multi-relational graph since the representation of relationships in a knowledge graph is no longer a single relation representation but a multi-relation representation. The classification of KGRL models and some related algorithms are listed in Table 4.

Table 4 A summary of knowledge graph representation learning approaches

Category	Method	Entity embed.	Relation embed.	Scoring Function
Translational Models	TransE [87]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\lVert\mathbf{h}+\mathbf{r}-\mathbf{t}\rVert_{l_1/l_2}$
	TransH [88]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^d$	$\left\lVert\left(\mathbf{h}-\mathbf{w}_r^\top\mathbf{h}\mathbf{w}_r\right)+\mathbf{d}_r-\left(\mathbf{t}-\mathbf{w}_r^\top\mathbf{t}\mathbf{w}_r\right)\right\rVert_{l_1/l_2}$
	TransR [89]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^k,\mathbf{M}_r\in\mathbb{R}^{k\times d}$	$\lVert\mathbf{h}_r+\mathbf{r}-\mathbf{t}_r\rVert_{l_1/l_2}$
	TransD [90]	$\mathbf{h},\mathbf{t},\mathbf{w}_h,\mathbf{w}_t\in\mathbb{R}^d$	$\mathbf{r},\mathbf{w}_r\in\mathbb{R}^k$	$\lVert\mathbf{h}_\perp+\mathbf{r}-\mathbf{t}_\perp\rVert_2^2$
	TransAt [93]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$P_r\left(\sigma\left(\mathbf{r}_h\right)\mathbf{h}\right)+\mathbf{r}-P_r\left(\sigma\left(\mathbf{r}_t\right)\mathbf{t}\right)$
Multiplicative Models	RESCAL [98]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{h}^\top\mathbf{M}_r\mathbf{t}$
	DisMult [99]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathbf{h}^\top\operatorname{diag}\left(\mathbf{M}_r\right)\mathbf{t}$
	HOLE [100]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{r}^\top(\mathbf{h}\star\mathbf{t})$
	ComplEx [101]	$\mathbf{h},\mathbf{t}\in\mathbb{C}^d$	$\mathbf{r}\in\mathbb{C}^d$	$\mathrm{Re}\left(\mathbf{h}^\top\operatorname{diag}(\mathbf{r})\overline{\mathbf{t}}\right)$
	ANALOGY [102]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{h}^\top\mathbf{M}_r\mathbf{t}$
	SimplE [103]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{r}'\in\mathbb{R}^d$	$\frac{1}{2}\left(\mathbf{h}\circ\mathbf{r}\mathbf{t}+\mathbf{t}\circ\mathbf{r}^{-1}\mathbf{t}\right)$
	TuckER [104]	$\mathbf{h},\mathbf{t}\in\mathbb{R}_e^d$	$\mathbf{r}\in\mathbb{R}_r^d$	$\mathcal{W}\times_1\mathbf{h}\times_2\mathbf{w}_r\times_3\mathbf{t}$
Graph-based Models	RGCN [107]	$\mathbf{e_i}\in\mathbb{R}^d$	$R_r\in\mathbb{R}^{d\times d}$	$e_s^T R_r e_o$
	SACN [108]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$g(\mathrm{vec}(\mathbf{M}(\mathbf{h},\mathbf{r}))W)\mathbf{t}$
	KBGAT [109]	$\mathbf{e_i}\in\mathbb{R}^{N_e\times T}$	$\mathbf{r_k}\in\mathbb{R}^{N_e\times T}$	$\left(\prod_{m=1}^{\Omega}\mathrm{ReLU}\left(\left[\vec{h}_i,\vec{g}_k,\vec{h}_j\right]*\omega^m\right)\right)\cdot\mathbf{W}$
	RGHAT [111]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathrm{ReLU}(\mathrm{vec}(\mathrm{ReLU}([\overline{\mathbf{h}};\ \overline{\mathbf{r}}]*\omega))\mathbf{Q})\mathbf{t}$
Neural Networks-based Models	NTN [113]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r},\mathbf{b}_r\in\mathbb{R}^k,\widehat{\mathbf{M}}\in\mathbb{R}^{d\times d\times k}$ $\mathbf{M}_{r,1},\mathbf{M}_{r,2}\in\mathbb{R}^{k\times d}$	$\mathbf{r}^\top\sigma\left(\mathbf{h}^T\widehat{\mathbf{M}}\mathbf{t}+\mathbf{M}_{r,1}\mathbf{h}+\mathbf{M}_{r,2}\mathbf{t}+\mathbf{b}_r\right)$
	SME [114]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$G_{left}(\mathbf{h},\mathbf{r})^\top G_{right}(\mathbf{r},\mathbf{t})$
	ConvE [115]	$\mathbf{M}_h\in\mathbb{R}^{d_w\times d_h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d_w\times d_h}$	$\sigma\left(\mathrm{vec}\left(\sigma\left(\left[\mathbf{M}_h;\ \mathbf{M}_r\right]*\boldsymbol{\omega}\right)\right)\mathbf{W}\right)\mathbf{t}$
	ConvKB [116]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathrm{concat}(\sigma([\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}]*\boldsymbol{\omega}))\cdot\mathrm{w}$
	NAM [118]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\sigma\left(\mathbf{z}^{(L)}\cdot\mathbf{t}+\mathbf{B}^{(L+1)}\mathbf{r}\right)$
Temporal Models	Know-Evolve [120]	$\mathbf{v}^{\mathbf{e}^\mathrm{s}},\mathbf{v}^{\mathbf{e}^\mathrm{s}}\in\mathbb{R}^d$	$\mathbf{R}_\mathbf{r}\in\mathbb{R}^{d\times d}$	$g_r^{e^s,e^o}(t)=\mathbf{v}^{\mathbf{e}^\mathrm{s}}(t-)^T\cdot\mathbf{R}_\mathbf{r}\cdot\mathbf{v}^{\mathbf{e}^\mathrm{o}}(t-)$
	TTransE [121]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$-\lVert\mathbf{h}+\mathbf{r}+\tau-\mathbf{t}\rVert_{L_{1/2}}$
	HyTE [122]	$\mathbf{h},\mathbf{t}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\lVert P_\tau(\mathbf{h})+P_\tau(\mathbf{r})-P_\tau(\mathbf{t})\rVert_{L_1/L_2}$
	TA-DisMult [124]	$\mathbf{e}_s,\mathbf{e}_o\in\mathbb{R}^d$	$\mathbf{e}_{p_{seq}}\in\mathbb{R}^d$	$\left(\mathbf{e}_s\circ\mathbf{e}_o\right)\mathbf{e}_{p_{seq}}^T$

4.1.1 Translational Models

Translational models represent relations by interpreting them as the translation from head to tail entities, which are also used to measure the plausibility of triples. Several translation-based models have been set forth to model multi-relational data more effectively and easily. As shown in Fig. 3(a), TransE [87] emphasizes that the vector of the head entity plus the relation is closer to the vector of the tail entity in the representation space, namely $\mathbf{h}+\mathbf{r} \approx \mathbf{t}$, of which the scoring function is defined as

$$f_r(h,t) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{l_1/l_2} \tag{1}$$

where l_1 and l_2 represent norms. Inspired by translational ideas, many extensions have been constantly set forth. For instance, the strong assumption of $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ in the TransE model may lead to improper learning of representations in reflexive relations, one-to-many relations, and many-to-one relations. To solve the problems, Wang et al. proposed TransH [88] shown in Fig. 3(b), which emphasizes the projections of the head and tail entities being close to each other on the hyperplane corresponding to the relation $\mathbf{r}$, with a scoring function defined as

$$f_r(\mathbf{h},\mathbf{t}) = \|(\mathbf{h} - \mathbf{w}_r^\top \mathbf{h}\mathbf{w}_r) + \mathbf{d}_r - \left(\mathbf{t} - \mathbf{w}_r^\top \mathbf{t}\mathbf{w}_r\right)\|_{l_1/l_2} \tag{2}$$

where $\mathbf{w}_r$ is the normal vector. In both TransE and TransH models, entities and relations are represented in the same semantic space, which will limit the expressive ability of the model to some extent. Due to this limitation, TransR [89] is proposed, which maps head entity vector $\mathbf{h}$ and tail entity vector $\mathbf{t}$ to $\mathbf{h}_r$=$\mathbf{h}\mathbf{M}_r$ and $\mathbf{t}_r$=$\mathbf{t}\mathbf{M}_r$, with scoring function defined as

$$f_r(h,t) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_{l_1/l_2} \tag{3}$$

where $\mathbf{M}_r$ is the projection matrix. However, it doesn't make sense for TransR to use the same projection matrix while computing $\mathbf{h}_r$ and $\mathbf{t}_r$. As shown in Fig. 3(c), TransD [90] utilizes dynamic projection matrix, for the head entity $\mathbf{h}$ and tail entity $\mathbf{t}$, with mapping functions are respectively shown as

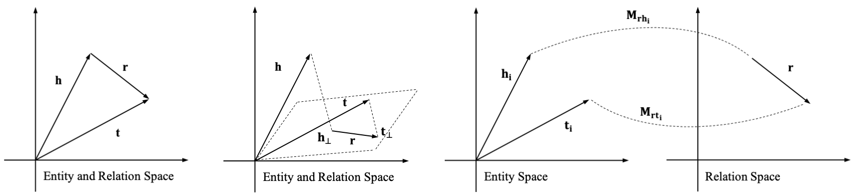
$$\mathbf{M}_{rh} = \mathbf{r}_p\mathbf{h}_p^\top + \mathbf{I}^{m\times n}, \mathbf{M}_{rt} = \mathbf{r}_p\mathbf{t}_p^\top + \mathbf{I}^{m\times n} \tag{4}$$

where $\mathbf{I}^{m\times n}$ is the identity matrix. After mapping, vectors of head entity and tail entity are respectively computed as

$$\mathbf{h}_\perp = \mathbf{M}_{rh}\mathbf{h}, \quad \mathbf{t}_\perp = \mathbf{M}_{rt}\mathbf{t} \tag{5}$$

with the scoring function is defined as

$$f_r(\mathbf{h},\mathbf{t}) = \|\mathbf{h}_\perp + \mathbf{r} - \mathbf{t}_\perp\|_2^2 \tag{6}$$



(a) TransE (b) TransH (c) TransD

Fig. 3 Illustrations of translational models: TransE, TransH and TransD.

In addition to the methods mentioned above, several variations have been proposed. Following the translational principles, TransG [91] is proposed to deal with multiple relation semantic problems through mining the multiple potential meanings of pairs of entities whose relations may be associated with corresponding triples in the knowledge graph. By improving the loss function of the previous translational method, TransA [92] can adaptively learn the embedding of entities and relations in a knowledge graph. Besides, the previous approaches ignored the hierarchical routine of human cognition, so TransAt [93] adopts an attention mechanism to handle the problems above. TransC [94] distinguishes concepts from instances by encoding instances, concepts, and relations in the same semantic space. Cui et al. proposed TransL [95] to learn the entity and relation embeddings by leveraging the local connection explicitly. TransRHS [96] seamlessly integrated hierarchical relation structure into the embeddings. Recently, TransROWL [97] aims to enhance the effectiveness of representation by injecting background knowledge into models.

4.1.2 Multiplicative Models

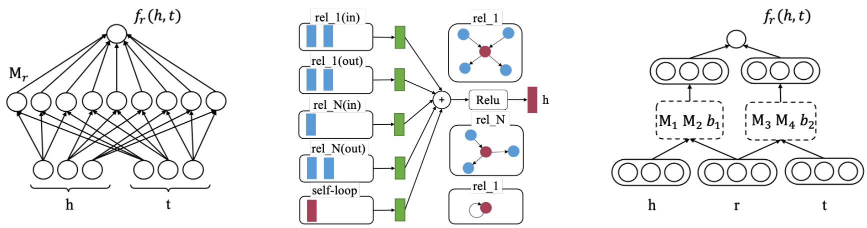
In multiplicative models, the likelihood of entity-relation-entity triple belonging to the KG is quantified by a multiplicative score function. As shown in Fig. 4(a), Nickel et al. developed RESCAL [98] which utilized tensor decomposition and intrinsic structure of multi-relational data. To be more precise, rank-r factorization was introduced, and the slice of $\mathcal{X}$ is factorized as

$$\mathcal{X}_k \approx A R_k A^T, \text{ for } k = 1, \ldots, m \tag{7}$$

where A is the representation of entities and R_k is an asymmetric matrix. DisMult [99] simplified RESCAL by restricting $\mathbf{M}_r$ to diagonal matrices, with the scoring function defined as

$$f_r(h,t) = \mathbf{h}^{\top} \operatorname{diag}(\mathbf{M}_r)\, \mathbf{t} \tag{8}$$

As a means to capture rich internal interactions in relational data while making computation and training easier, HOLE [100] combines the expressiveness of the tensor product with the efficiency and simplicity of TransE, using the cyclic correlation of vectors to represent entity pairs. ComplEx [101]



(a) RESCAL (b) RGCN (c) SME

Fig. 4 Illustrations of some representation learning models: RESCAL, RGCN and SME.

expands DisMult by utilizing the composition of complex embeddings to capture semantic information of asymmetric relations. Liu et al. [102] proposed ANALOGY, which models multi-relational data for optimizing the latent representations concerning the analogical properties of the embedded entities and relations, with the scoring function shown as

$$f_r(h,t) = \mathbf{h}^\top \mathbf{M}_r \mathbf{t} \tag{9}$$

where $\mathbf{h}, \mathbf{t} \in \mathbb{R}^d$ is the embeddings of entities, and $\mathbf{M}_r$ is the relation matrix limited the linear mapping. SimplE [103] introduces a simple improvement of canonical polyadic decomposition and is considered the inverse of relations, with the scoring function defined as

$$f_r(h,t) = \frac{1}{2}\left(\mathbf{h} \circ \mathbf{r}\mathbf{t} + \mathbf{t} \circ \mathbf{r}^{-1}\mathbf{t}\right) \tag{10}$$

Same as the model mentioned above, TuckER [104] was developed for link prediction by using tucker decomposition, and the scoring function is defined as

$$f_r(h,t) = \mathcal{W} \times_1 \mathbf{h} \times_2 \mathbf{w}_r \times_3 \mathbf{t} \tag{11}$$

where $\mathcal{W} \in \mathbb{R}^{d_e \times d_r \times d_e}$ is the core tensor and $\mathbf{w}_r \in \mathbb{R}^{d_r}$ is the relation embedding vector. Recently, Cao et al. introduced DualE [105] by integrating dual quaternions into knowledge graph representation, with scoring function defined as

$$f_r(h,t) = \langle a'_h, a_t \rangle + \langle b'_h, b_t \rangle + \langle c'_h, c_t \rangle + \langle d'_h, d_t \rangle \tag{12}$$

ConEx [106] considered a multiplicative composition of a 2D convolution with a hermitian inner product on complex-valued embeddings, with a score computed as

$$f_\tau(h,t) = \mathrm{Re}(\langle \mathrm{conv}\left(\mathbf{e}_h, \mathbf{e}_r\right), \mathbf{e}_h, \mathbf{e}_r, \overline{\mathbf{e}_t} \rangle) \tag{13}$$

where $\mathrm{conv}(\cdot,\cdot) : \mathbb{C}^{2d} \mapsto \mathbb{C}^d$ is defined as

$$\mathrm{conv}(\mathbf{e}_h, \mathbf{e}_r) = f(\mathrm{vec}(f([\mathbf{e}_h, \mathbf{e}_r] * \omega))\mathbf{W} + \mathbf{b}) \tag{14}$$

4.1.3 Graph-based Models

In recent years, with the advantages of graph neural networks in learning the representation of elements, we can better compute embeddings of entities and relations in knowledge graphs. As illustrated in Fig. 4(b) RGCN [107] presents relation-specific transformations and modeled multi-relational data by applying graph convolutional networks, and the forward-pass update of the entity is defined as

$$h_i^{(l+1)} = \sigma(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}) \tag{15}$$

where $h_i^{(l)} \in \mathbb{R}^{d^{(l)}}$ is the state of l-th layer, N_i^r are neighbors of node i specific to relation r, and $c_{i,r} =$ N_–i"^–r" is the normalization constant. RGCN adopts novel encoder-decoder architecture, in which the encoder generates entity embedding based on GCN [84] and the decoder predicts relations based on entity embedding by semantic matching model. Based on the same architecture, SACN [108] extends the RGCN to make it adaptable. Specifically, SACN introduced a weighted graph convolutional network (WGCN) as an encoder, which considered graph connectivity structure and node attributes in the computation. Moreover, the decoder of SACN was a convolutional network named Conv-TransE, with a scoring function defined as

$$f_r(h, t) = f\left(\operatorname{vec}\left(\mathbf{M}\left(\mathbf{h}, \mathbf{r}\right)\right) W\right) \mathbf{t} \tag{16}$$

KBGAT [109] utilises graph attention networks as encoder while using the multi-head mechanism to capture the semantic information from multi-hop neighbourhoods, along with ConvKB, which is introduced as a decoder for end-to-end link prediction. Ye et al. [110] proposed VR-GCN that combined convolutional and translational characteristics. Unlike GCN, in each layer of VR-GCN, the output is obtained by a nonlinear transformation after summing and averaging the representations of its neighbours and itself. CompGCN [18] jointly represents both entities and relations in the multi-relational graph by leveraging entity-relation composition operations, with updating function defined as

$$h_v = f(\sum_{(u,r) \in \mathcal{N}(v)} \boldsymbol{W}_{\lambda(r)} \phi\left(\boldsymbol{x}_u, \boldsymbol{z}_r\right)) \tag{17}$$

where h_v is the updated representation of node v, $\boldsymbol{x}_u, \boldsymbol{z}_r$ is the initial embedding of node u and relation r separately. $\boldsymbol{W}_{\lambda(r)} \in \mathbb{R}^{d_1 \times d_0}$ is the parameter specific to the type of the relation. RGHAT [111] is a new neighborhood-aware model and aimed to compute different weights for different neighboring relations and entities. Recently, KEGCN [112] was proposed to represent both entity and relation by using GCN and it utilizes several knowledge graph embedding methods into GCNs for modelling multi-relational graphs.

4.1.4 Neural Networks-based Models

In recent years, the research of neural networks encoding semantic information of multi-relational data has been widely concerned. Many knowledge representation methods based on neural networks are also emerging. For example, Socher et al. proposed NTN [113] which first projects entities to their vector embeddings in the input layer. Fig. 4(c) gives a simple illustration of SME [114], another neural network framework which models multi-relational graphs into a continuous representation space. SME defines several projection matrices to describe the intrinsic relationship between entities and relations, and creates two scoring functions for each triplet (h, r, t), where the linear form is defined as

$$f_r(h,t) = (\boldsymbol{M}_1\boldsymbol{l}_h + \boldsymbol{M}_2\boldsymbol{l}_r + \boldsymbol{b}_1)^{\mathrm{T}} \cdot (\boldsymbol{M}_3\boldsymbol{l}_t + \boldsymbol{M}_4\boldsymbol{l}_r + \boldsymbol{b}_2) \tag{18}$$

and bilinear form is defined as

$$f_r(h,t) = (\boldsymbol{M}_1\boldsymbol{l}_h \otimes \boldsymbol{M}_2\boldsymbol{l}_r + \boldsymbol{b}_1)^{\mathrm{T}} \cdot (\boldsymbol{M}_3\boldsymbol{l}_t \otimes \boldsymbol{M}_4\boldsymbol{l}_r + \boldsymbol{b}_2) \tag{19}$$

In addition to these, convolutional neural networks are also used for KGRL. For example, ConvE [115] applies 2D convolution over embeddings and multiple layers of nonlinear features to model the interactions in multi-relational data by reshaping and concatenating the resulting matrix as input. The scoring function is defined as follows:

$$f_r(h,t) = \sigma\left(\mathrm{vec}\left(\sigma\left([\mathbf{X}_h;\ \mathbf{X}_r] * \omega\right)\right)\mathbf{W}\right)\mathbf{t} \tag{20}$$

where f is nonlinear function and vec is the vectorization operation, $\mathbf{X}_h$, $\mathbf{X}_r$ denote the 2D reshaping of $\mathbf{h}$ and $\mathbf{r}$. ConvKB [116] also adopts CNNs to explore global relations and only uses one-dimensional convolution. The final scoring function of ConvKB is shown as

$$f(h,r,t) = \mathrm{concat}\left(g\left([\boldsymbol{v}_h, \boldsymbol{v}_r, \boldsymbol{v}_t] * \boldsymbol{\Omega}\right)\right) \cdot \mathbf{w} \tag{21}$$

Jiang et al. [117] proposed ConvR, an adaptive convolutional network specially designed to represent multi-relational data. In addition, NAM [118] also uses deep neural networks to encode semantic information, and the scoring function is defined as

$$f_r(h,t) = \sigma\left(\mathbf{w}^{\top}\sigma(\mathbf{W}[\mathbf{h},\mathbf{r},\mathbf{t}])\right) \tag{22}$$

where W is weight matrix.

4.1.5 Temporal Models

Previous knowledge graph representation learning has focused on static knowledge graphs where facts are not changed with time. However, the knowledge graph used in the real-life application is dynamic for the availability of large-scale event data with time stamps. The proposed temporal models offer knowledge graph representation learning methods to aggregate the rich temporal information in dynamic knowledge graphs for better knowledge graph representation. Generally, a novel temporal dimension is added to the triples, represented as: (h, r, t, τ). For example, t-TransE [119] takes temporal information into consideration. The temporal order score function is defined as

$$g(r_i, r_j) = \|\mathbf{r}_i\mathbf{M} - \mathbf{r}_j\|_1 \tag{23}$$

where $\mathbf{M} \in \mathbb{R}^{n\times n}$ is a transition matrix between pair-wise temporal ordering relation pair (r_i, r_j). Know-Evolve [120] learns non-linearly evolving entity representations at different times based on a novel deep evolutionary knowledge network, and it utilises the temporal point process to describe the impact of different time points. TTransE [121] extends TransE to encode time information in the same representation space as entities and relations, with the scoring function defined as

$$f_\tau(h, r, t) = -\|\mathbf{h} + \mathbf{r} + \tau - \mathbf{t}\|_{l_{1/2}} \tag{24}$$

HyTE [122], through a time-aware KGE approach, explicitly binds time to the entity-relation space by associating each timestamp with the corresponding hyperplane, of which the scoring function is shown as

$$f_\tau(h, r, t) = \|P_\tau(e_h) + P_\tau(e_r) - P_\tau(e_t)\|_{l_1/l_2} \tag{25}$$

The projected representation on w_τ is computed as

$$\begin{aligned} P_\tau(e_h) &= e_h - \left(w_\tau^\top e_h\right) w_\tau \\ P_\tau(e_t) &= e_t - \left(w_\tau^\top e_t\right) w_\tau \\ P_\tau(e_r) &= e_r - \left(w_\tau^\top e_r\right) w_\tau \end{aligned} \tag{26}$$

where the $\|w_\tau\|_2$=1. In addition, DE-SimplE [123], a novel temporal knowledge graph embedding approach, provides the entity's characteristics at any point in time by equipping the static model with a diachronic entity embedding function. TA-DistMult [124] applies RNNs to learn the embedding of relation types while considering the temporal information. Lacroix et al. proposed TComplEx [125] enlightened by the canonical decomposition of tensors, which was the extension of ComplEx [101]. RE-GCN [126] is aimed to learn the dynamic representation of entities and relations at each timestamp while capturing the structural dependencies within the knowledge graph. Wu et al. [127] developed TIE to address some challenges, such as catastrophic forgetting.

4.2 Knowledge Graph Completion

The advancement of current knowledge graphs automation technologies makes it easy to build large knowledge graphs containing millions of entities and relationships. Despite the tremendous scale of the knowledge graphs, they are still suffering from the problem of incompleteness and inadequacy. Research about knowledge graph completion (KGC) has been promoted to automatically complete the missing facts of a triplet, which can make the information in the knowledge graphs more comprehensive. According to the missing part, it can be divided into three sub-tasks:

1. *Head entity completion* Given the relation and tail entities, complete the missing head entities in the triplet, for example, (?, locatedIn, California).
2. *Tail entity completion* Given the relation and head entities, complete the missing tail entities in the triplet, for example, (Los Angeles, locatedIn, ?).
3. *Relation completion* Given the head and tail entities, complete the missing relation in the triplet, for example, (Los Angeles, ?, California).

Overall, classifying nodes and predicting edges are two essential tasks that many complex systems tried to undertake [128]. Put another way, the subtasks of KGC can also be divided into entity prediction, link prediction, and triple classification.

Knowledge graph completion algorithms can be divided into two types according to whether they can process new entities or new relations, namely, static knowledge graph completion (SKGC) and dynamic knowledge graph completion (DKGC). However, in this survey, we classify knowledge graph completion according to deep learning techniques. In recent years, relation completion has become a hot topic in academic circles, and much work has been focused on it. The relevant methods can be categorized into three types: the first type is the relation path-based reasoning method; the second type is the probability graph-based method; the final one is the method based on representation learning.

The relation path ranking algorithm predicts the relation between two entities by using the path linking them as a feature. The relation path ranking algorithm, implying complex graph structures and message transfer processes, is highly interpretable and automatically discovers association rules from the data. This tends to be more accurate than representation-based learning methods. This includes random walk, breadth-first search, and depth-first search. For example, PRA [129] learns a weighted combination of path-constrained random walkers and could utilize the complex path features of relational data. Due to the shortcomings of PRA, which ignores the relevance of relations, CPRA [130] couples the path ranking of multiple relations by introducing a common-path-based similarity measure, which shows more significant interpretability and predictive accuracy than PRA. Also, in dealing with the path of multiple relations, Das et al. proposed a method [131] that integrated multiple paths by using neural attention modeling with RNN, including a novel pooling function that does soft attention during gradient step and finding it to work better.

Probabilistic graphs use nodes to represent variables, namely candidate facts in the knowledge graph. The edges in the graph represent relations between candidate facts [132]. Each node is attached with a certain probability, and the missing relation is found through probabilistic reasoning. Probabilistic graphical models for KGC utilise the Bayesian network and Markov logic network [133]. Markov logic network is a probability distribution model with Markovian random variables combining the Markov network and first-order logic, which has substantial application prospects in natural language processing, complex network, information extraction and other fields. Qu et al. [134] defined the joint distribution of all possible triples by using a Markov logic network with first-order logic, which can be efficiently optimized with the variational EM algorithm. A Bayesian network is a directed acyclic graph in form, which takes the network structure and node attribute information into consideration to reflect the states of a part of the world being modelled and describe how these states relate to probabilities. GaussianPath [135] adopts a trainable Bayesian neural network to approximate Q-function, aiming to capture the uncertainty of a multi-hop reasoning path for knowledge graph completion.

The representation learning-based completion method represents the entities and relations in the knowledge graph in a low-dimensional space. It then defines a scoring function based on triples on each knowledge item. The final scores of all candidate entities are calculated through a scoring function for ranking them. Compared with the two methods mentioned above, the method based on representation learning is more general and efficient for knowledge graph completion. Besides, many previous knowledge representation learning methods can be added directly as a module to the process of knowledge graph completion tasks. For example, in the translational models, such as TransE [87], TransH [88], TransR [89], TransD [90], etc, can be used to complete the knowledge graph. Graph-based models, for instance, SACN [108], comprised of weighted GCNs and Conv-TransE, take the entity attribute and internal structure of the knowledge graph into consideration, and the effectiveness of SACN performs well in knowledge graph completion task. Moreover, ConMask [136] utilises external text to lock missing relevant entities. It leverages semantic averaging and CNNs to fully extract relation-dependent representation from the textual features of entities and relations in the KGs. TuckER [104] is based on the Tucker decomposition of a binary tensor of known facts for knowledge graph completion.

5 Deep Learning for Knowledge-Aware Applications

Deep learning techniques have received extensive research and attention in recent years. This paper systematically introduces the practical applications of deep learning techniques on knowledge graphs. According to our survey, the main focus of knowledge-aware applications based on deep learning is on recommender systems, question and answer systems, natural language understanding, and decision making, as described in Section 5.1 to Section 5.4 of this paper. Applications on specific domains are presented in Section 5.5.

5.1 Recommender Systems

Recommendation systems are used to solve the information overload problem faced by users and have tremendous applications in various fields. Traditional recommendation methods such as content filtering, collaborative filtering, and hybrid methods [137] cannot discover potential connections between entities denoting the latent interaction messages [138]. They also face problems such as cold start and poor scalability. Traditional methods can easily lead to poor accuracy or poorly interpretable recommendations for scenarios with sparse data, making it difficult for users to understand why they have been recommended irrelevant items [139]. The successful application of deep learning techniques on knowledge graphs has enabled the practical problems faced by traditional recommendation methods to be solved [140]. The existing KG-based recommendation methods can be categorized into embedding-based methods and path-based methods [14].

5.1.1 Embedding-based

The recommendation system based on knowledge graph representation learning uses the KGE algorithm to transform KG to map users, items, and the relationship between them into a continuous vector space. Further, it applies the learned information such as entity similarity to complete the recommendation task, improving the computation efficiency and providing semantic support for the recommendation system. Since introducing deep learning-based embedding methods in knowledge graphs, combining with some classical embedding models such as TransE to implement recommendation systems has achieved good results. For Example, Gourab Chowdhury et al. [141] introduced TransE to learn embeddings of entities and relations to propose a neural factorization model for recommendation tasks. Paula Gómez Duran et al. [142] applied graph convolutional networks (GCNs) to an existing collaborative filtering model to learn contextual information about interactions through a graph convolutional embedding layer to enrich recommendations. Traditional recommendation methods face several problems, as mentioned above, and for the same, appropriate embedding models are proposed. Entity2vec [143] is a recommended method that computes subgraph embeddings of specific attributes to learn correlations between users and items, achieving excellent performance despite sparse datasets and can better solve the problem of sparse data compared to traditional recommendation methods. Star-GCN [144] masks the input node embeddings and reconstructs the input node embeddings through two components: graph decoding and graph encoding. The model can learn entity embeddings that do not appear during the training process, thus alleviating the cold-start problem.

5.1.2 Path-based

A knowledge graph is a heterogeneous network in which the entities representing users, items, and the interactions between them may have different types. Different paths can visually reflect the different connections between users and items, so combining path reasoning with recommendation systems can make the recommendation process more rational [145]. Lin et al. [146]. used a gated attention mechanism to capture user preferences from specific types of paths and applied this information to path reasoning. Considering the dynamic nature of some recommender systems, TPRec [147] adds temporal information to knowledge graph path reasoning. Specifically, TPRec treats user-item interactions as temporal-aware relationships and constructs a knowledge graph of temporal-aware interactions, making the recommendation process more interpretable.

5.2 Question Answering

A knowledge graph-based question and answer system (KBQA) is a technology that extracts entities and relations from natural language questions and uses knowledge such as reasoning rules to obtain answers in a knowledge graph. Traditional KBQA methods include semantic analysis-based QA, information retrieval-based QA, etc [148]. Semantic analysis-based QA analyses the semantic information in natural language questions and converts it into a logical form that can be understood by the knowledge base, thus reasoning through knowledge in the knowledge-base to obtain the final answer [149]. Information retrieval-based question and answer extracts entities from the question, express the entities as question features based on specific templates and rules, and retrieves candidate answer subgraphs in the knowledge base for filtering based on the features [150].

However, traditional KBQA methods involve manual feature extraction and human-defined templates and rules. There are problems such as the difficult manual design of rules and tedious feature extraction, which make the methods very difficult to apply, and cannot handle complex semantic parsing scenarios, and the accuracy of the results is not high enough. Therefore, combining deep learning with KBQA has become a mainstream research trend in recent years. Luo et al. [151], proposed a single-relational question answering model based on BERT, which applied the attention mechanism to the semantic association construction of natural language questions and KG facts and optimized the representation of candidate objects with better results. Hao et al. [152], provided an end-to-end cross-attentive model for different aspects of a test taker's response, where correctness is judged based on a weighted sum of all aspects of the answer's score. Daniil Sorokin et al. [153], combined gated graph neural networks with semantic parsing, using neural networks to automatically encode the graph structure for semantic parsing without involving the manual extraction of features. Compared with traditional semantic parsing methods, the model introducing a graph neural network can handle more complex semantic parsing scenarios. Question answering is often accompanied by reasoning. Cai et al. [154], provided a deep cognitive reasoning network (DCRN) that encodes and later decodes questions to obtain reasoning path information. Finally, it receives answers through two stages: unconscious rough recognition of candidate entities and conscious precise recognition of candidate entities. Unlike specific rule-based reasoning question-and-answer systems, this method can accomplish reasoning tasks through graph structure information, free from the limitations of traditional methods.

5.3 Natural Language Understanding

Utilizing the rich background knowledge in the knowledge base during natural language understanding can enhance the learning ability of the model. Knowledge-driven NLU tasks have made good progress in recent years, and knowledge-aware NLU tasks, such as entity linking tasks and relationship extraction tasks, have emerged. The application of graph neural networks in natural language processing is summarised in detail by Wu et al. [155]. Specifically, the deep learning-based knowledge-aware NLU task can be understood as the following process: firstly, the prior knowledge in the external knowledge base is mapped into low-dimensional vectors by graph representation learning techniques, and then the text representation learning in the NLU task is enriched by using the knowledge base based on embedding, so that the text has higher-level semantic understanding, thus solving the problem of lack of background knowledge support in the traditional NLU model. Logeswaran et al. [156] proposed a new Zero-Shot entity linking model in which entities are described by only one paragraph of text and the model is not applied to a specific domain with robust domain application. KLMo [157] is a KG-enhanced pre-trained language framework that proposes a knowledge aggregator to model the interaction of entities and relations in KG to enhance the expressive power of the language model. KCL-TEN [158] is a textual entailment network where external knowledge and textual context interact, which introduces prior knowledge from an external knowledge base in the model and uses graph attention networks to learn graph-level representations of text. Peng et al. [159] proposed an event-based heterogeneous information network by incorporating knowledge from external knowledge bases. Based on this heterogeneous information network, they discovered an overall similarity between social events and used a graph convolutional neural network to accomplish the classification task of events. Dun et al. [160] provided a knowledge-aware attention network (KAN) to detect fake news, where they extract entities from the news content, and the corresponding entities and their contexts in the knowledge base are used as external supplementary knowledge. Entities and their contexts are encoded by a transformer encoder and are finally fed into the classifier via a multi-headed attention mechanism to determine the truthfulness of the news.

5.4 Decision Making

Knowledge graphs have become integral to many supported decision-making systems for automatic reasoning and deep relationship discovery capabilities. The knowledge-aware decision model consists of two main components: knowledge modeling and representation and reasoning on knowledge graphs [161].

Traditional knowledge graph-based decision models only associate simple knowledge representations, which may ignore the latent semantic information and fail to accommodate uncertain scenarios that further make the wrong decisions.

The introduction of deep learning-based reasoning techniques enhances the decision-making capability of knowledge graphs in complex scenarios. Lv et al. [162] focused on common-sense question answering tasks. They used graph representation learning techniques to relearn word representations in natural language utterances, using graph neural networks and graph attention mechanisms to aggregate evidence to predict final answers. Liu et al. [163] focused on the text generation task for common sense reasoning and decision. They proposed a knowledge graph-based pre-trained text generation model KG-BART, which integrates the knowledge graph with the text decoding and encoding process to generate higher quality, more logical utterances. Re-GCN [164] recursively captures the structural dependencies of KGs in different timestamp states to learn evolutionary representations of entities and relations. Experiments show that Re-GCN significantly improves the accuracy of future event prediction and decision-making tasks. However, deep learning methods improve the problems, such as low computational efficiency in traditional methods. They still have poor interpretability, which gives the theoretical foundation for authentic decision-making. Recent studies have begun to combine the advantages of both of these methods. Moon et al. [165] provided the DialKG Walker framework, which converts conversations into paths traversed on a knowledge graph, making conversational reasoning more interpretable. UniKER [166] combines deterministic horn rules with knowledge graph embeddings, making logical reasoning and KGE complementary. They enhance both the interpretability and computational efficiency of reasoning.

5.5 Domain-specific Applications

With The professional independence of certain fields, a domain-specific knowledge graph is built as the primary means of tackling many real-life problems in specific domains. However, the domain-specific knowledge graph approach has different data distribution and characteristics compared to general knowledge graphs. Many deep learning models for domain-specific applications have also emerged. For example, medical domain knowledge graphs, such as a clinical aid decision tool, a search engine for medical drug information, and an intelligent question and answer system for medical knowledge. Knowledge-aware models in medical applications are expected to lead to more efficient and accurate medical services. MedBERT [167] is a BERT-based model for automatic classification of medical queries, which incorporates medical domain knowledge as side information in the model, and finally, BERT encoding to complete the classification task of medical problems. In terms of disease-assisted detection, the knowledge graph can be used to assist in determining information such as physical examinations and test results. Chai et al. [168] developed a long short-term memory network based on biomedical knowledge mapping for the auxiliary diagnosis of thyroid diseases. Lu et al. [169] extracted patients' discharge information from their electronic health records and used external medical knowledge graphs to represent the textual information of the discharge, and finally, convolutional networks were set to predict the readmission rate of patients. In pharmacology, the prediction of drug-drug interactions (DDI) is an important issue. Lin et al. [170] proposed an end-to-end knowledge graph neural network. This framework efficiently learns information about drugs and their neighbourhoods and outperforms previous state-of-the-art models in predicting DDI. In addition to the above applications, knowledge graphs can also be applied in the cognitive domain. Rao et al. [171] focused on detecting psychiatric disorders on social media, and they integrated a bidirectional gated recursive unit with an attention mechanism. The model achieved better results on two mental illness detection tasks. In social computing, Yu et al. proposed an efficient algorithm for finding outlier motifs by exploring the user's query and constrained conditions [172].

6 Challenges and Open Issues

6.1 Knowledge Transfer in Cross-Domain KGs

Knowledge graphs have been used as the primary approach to tackling many real-life problems in various domains. However, as cross-discipline becomes necessary across multiple research fields, specific-knowledge graphs can no longer solve cross-domain issues. DOZEN [173] learns the relations between entities across different domains from an existing ontology of external knowledge and a set of analogies linking entities and domains. The cross-graph knowledge transfer network [174] utilizes the graph structure to transfer knowledge across domains, which helps it explicitly model intra-domain and cross-domain interactions. But the current approaches are limited to only a few domains. As the independence of each subject is broken, there is no doubt that the transfer of knowledge from more domains to achieve multiple applications may become a critical issue.

6.2 Interactive Learning and Cognitive Learning

The contribution of deep learning to knowledge graphs and the next generation of artificial intelligence is confirmed by an increasing number of current studies. Improving the self-learning capability of knowledge graphs and designing knowledge graphs that enable lifelong learning has become a critical topic. In particular, interactive learning and cognitive learning have become one of the mainstream research directions to realize the next generation of AI through knowledge graphs. Cognitive graph [175] is built by iteratively coordinating retrieval and reasoning. The structural information offered by the cognitive graph enables our model to aggregate pieces of evidence from multiple reasoning paths and explain the reasoning process graphically. DCRN [154] is proposed under the direction of cognitive science to learn that humans can reason over a large-capacity memory to find answers. It is very instructive to make deep learning more intelligent and can continuously learn through interactive and cognitive learning.

6.3 Scalability

Over the past decade, deep learning for KG has successfully modeled complex data. Currently, the size and number of knowledge graphs are growing exponentially, making it critical to properly improve the scalability of deep learning applications. Gao et al. [176] proposes novel graph-augmented learning to rank models by subgraph segmentation of large-scale graphs and then the exact search of subgraphs to increase the efficiency of querying, which combines a novel subgraph matching network based on GGNNs and an enhanced BiMPM model. Nevertheless, there still has a long way to go to deal with cumbersome deep architectures and the increasingly growing knowledge graphs.

6.4 Dynamics

Most of the current deep learning approaches for knowledge graphs are limited to dealing with static data. Nevertheless, the actual knowledge graph is dynamic, with many facts holding for only a specific period. Data scarcity becomes challenging when dealing with dynamic knowledge graphs due to the emergence of new, previously unseen relations. Due to the dynamic nature of knowledge graphs, models may need to be retrained frequently, which requires time and memory to maintain a large training dataset. Dynamic deep learning approaches can handle new relations and entities but do not require retraining. Current methods use continuous learning frameworks to avoid overfitting and catastrophic forgetting of models. We still need many efforts to make the technology useful for large-scale real-world applications.

6.5 Anomaly Detection

The result will never be perfect regardless of the approach to constructing a knowledge graph. Due to the diversity of data sources and limitations of present knowledge extraction methods, anomalies such as redundant, inconsistent, contradictory, and inaccurate facts in a knowledge graph are unavoidable. To facilitate wide adoption and advanced usage, anomaly detection in knowledge graphs has become an essential task. Researchers have conducted preliminary explorations for the popular anomaly detection tasks, such as outlier detection, novelty detection, contextual anomaly detection, and collective anomaly detection. However, previous approaches are either domain-dependent, not scalable to large-scale graphs, or may require substantial human intervention [177]. There are still severe gaps between the present state of anomaly detection techniques and applications of high-quality KGs.

6.6 Explainability

For knowledge graphs and deep learning on knowledge graphs, explainability is not only a desirable property but a specific requirement for these techniques to be applied in reality. KPRN [178] models the association paths in the knowledge graph for user-item pairs to provide interpretable recommendations to users [179]. T-GAP [180] discusses the relationship between time displacement and query types using attention distributions with different timestamps relations. A case study of the inference process for a given query demonstrates the interpretability of the model's relational inference process. The current interpretable approach still has some limitations. Different from traditional interpretable solutions, rethinking the interpretability of deep learning in knowledge graphs in a unique perspective on knowledge graphs is also of high significance.

7 Conclusion

In this survey, we have conducted a comprehensive overview of deep learning on knowledge graphs. We provide a taxonomy for core components of deep learning on the following phases of the knowledge graph lifecycle, such as knowledge graph construction, knowledge representation, and knowledge graph application. According to the taxonomy, we categorize deep learning in knowledge graph construction into three groups; relation extraction, entity extraction, and attribute extraction. For the phase of knowledge representation, we divide deep learning methods into two categories: knowledge graph representation learning and knowledge graph completion. Then, we provide a comprehensive review of deep learning in knowledge-aware application on various domains like recommendation systems, question answering, natural language understanding, domain-specific knowledge graphs, and other state-of-the-art applications. We provide a thorough review, comparisons, and summarisation of these systems within or between categories. Finally, we summarise the challenges and open issues faced in the current study, along with representative research efforts, which are suggested to be future research directions in this rapidly growing field.

References

[1] Saffari, A., Oliya, A., Sen, P., Ayoola, T.: End-to-end entity resolution and question answering using differentiable knowledge graphs. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 4193–4200 (2021)

[2] Zhao, X., Chen, H., Xing, Z., Miao, C.: Brain-inspired search engine assistant based on knowledge graph. IEEE Transactions on Neural Networks and Learning Systems, 1–15 (2021)

[3] Huang, Y., Zhao, F., Gui, X., Jin, H.: Path-enhanced explainable recommendation with knowledge graphs. Proceedings of the 30th International Conference on World Wide Web **24**(5), 1769–1789 (2021)

[4] 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, pp. 1247–1250 (2008)

[5] Fabian, M., Gjergji, K., Gerhard, W., *et al.*: Yago: A core of semantic knowledge unifying wordnet and wikipedia. In: Proceedings of the 16th International Conference on World Wide Web, pp. 697–706 (2007)

[6] Carlson, A., Betteridge, J., Wang, R.C., Hruschka Jr, E.R., Mitchell, T.M.: Coupled semi-supervised learning for information extraction. In: Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, pp. 101–110 (2010)

[7] Paulheim, H.: Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web **8**(3), 489–508 (2017)

[8] Church, K., Bian, Y.: Data collection vs. knowledge graph completion: What is needed to improve coverage? In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 6210–6215 (2021)

[9] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S., Liu, H.: Graph learning: A survey. IEEE Transactions on Artificial Intelligence **2**(2), 109–127 (2021)

[10] Kong, X., Wang, K., Hou, M., Xia, F., Karmakar, G., Li, J.: Exploring human mobility for multi-pattern passenger prediction: A graph learning framework. IEEE Transactions on Intelligent Transportation Systems (2022). https://doi.org/10.1109/TITS.2022.3148116

[11] Liu, J., Xia, F., Feng, X., Ren, J., Liu, H.: Deep graph learning for anomalous citation detection. IEEE Transactions on Neural Networks and Learning Systems **33**(6), 2543–2557 (2022). https://doi.org/10.1109/TNNLS.2022.3145092

[12] Hur, A., Janjua, N., Ahmed, M.: A survey on state-of-the-art techniques for knowledge graphs construction and challenges ahead. In: Proceedings of the 4th IEEE International Conference on Artificial Intelligence and Knowledge Engineering (2021)

[13] Lin, Y., Han, X., Xie, R., Liu, Z., Sun, M.: Knowledge representation learning: A quantitative review. arXiv preprint arXiv:1812.10901 (2018)

[14] Zou, X.: A survey on application of knowledge graph. In: Journal of Physics: Conference Series, vol. 1487, p. 012016 (2020)

[15] Abu-Salih, B.: Domain-specific knowledge graphs: A survey. Journal of Network and Computer Applications **185**, 103076 (2021)

[16] Ji, S., Pan, S., Cambria, E., Marttinen, P., Philip, S.Y.: A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems **33**(2), 494–514 (2021)

[17] Liu, J., Ren, J., Zheng, W., Chi, L., Lee, I., Xia, F.: Web of scholars: A scholar knowledge graph. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2153–2156 (2020)

[18] Vashishth, S., Sanyal, S., Nitin, V., Talukdar, P.P.: Composition-based multi-relational graph convolutional networks. In: Proceedings of the 8th International Conference on Learning Representations (2020)

[19] Liu, J., Xia, F., Wang, L., Xu, B., Kong, X., Tong, H., King, I.: Shifu2: A network representation learning based model for advisor-advisee relationship mining. IEEE Transactions on Knowledge and Data Engineering **33**(4), 1763–1777 (2021). https://doi.org/10.1109/TKDE.2019.2946825

[20] Wang, W., Xia, F., Nie, H., Chen, Z., Gong, Z., Kong, X., Wei, W.: Vehicle trajectory clustering based on dynamic representation learning of internet of vehicles. IEEE Transactions on Intelligent Transportation Systems **22**(6), 3567–3576 (2020)

[21] Yu, S., Xia, F., Xu, J., Chen, Z., Lee, I.: Offer: A motif dimensional framework for network representation learning. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 3349–3352 (2020)

[22] Zhang, K., Yu, S., Wan, L., Li, J., Xia, F.: Predictive representation learning in motif-based graph networks. In: Australasian Joint Conference on Artificial Intelligence, pp. 177–188 (2019). Springer

[23] Hearst, M.A.: Automatic acquisition of hyponyms from large text corpora. In: Proceedings of the 14th International Conference on Computational Linguistics (1992)

[24] Brin, S.: Extracting patterns and relations from the world wide web. In: The 1998 World Wide Web and Databases, vol. 1590, pp. 172–183 (1998)

[25] Agichtein, E., Gravano, L.: Snowball: Extracting relations from large

plain-text collections. In: Proceedings of the 5th ACM Conference on Digital Libraries, pp. 85–94 (2000)

[26] Hasegawa, T., Sekine, S., Grishman, R.: Discovering relations among named entities from large corpora. In: Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), pp. 415–422 (2004)

[27] Mintz, M., Bills, S., Snow, R., Jurafsky, D.: Distant supervision for relation extraction without labeled data. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, pp. 1003–1011 (2009)

[28] Hendrickx, I., Kim, S.N., Kozareva, Z., Nakov, P., Ó Séaghdha, D., Padó, S., Pennacchiotti, M., Romano, L., Szpakowicz, S.: SemEval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: Proceedings of the 5th International Workshop on Semantic Evaluation, pp. 33–38 (2010)

[29] Liu, C., Sun, W., Chao, W., Che, W.: Convolution neural network for relation extraction. In: Proceedings of the 9th International Conference on Advanced Data Mining and Applications, pp. 231–242 (2013)

[30] Zeng, D., Liu, K., Lai, S., Zhou, G., Zhao, J.: Relation classification via convolutional deep neural network. In: Proceedings of the 25th International Conference on Computational Linguistics: Technical Papers, pp. 2335–2344 (2014)

[31] Nguyen, T.H., Grishman, R.: Relation extraction: Perspective from convolutional neural networks. In: Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing, pp. 39–48 (2015)

[32] dos Santos, C., Xiang, B., Zhou, B.: Classifying relations by ranking with convolutional neural networks. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 626–634 (2015)

[33] Zeng, D., Liu, K., Chen, Y., Zhao, J.: Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1753–1762 (2015)

[34] Jiang, X., Wang, Q., Li, P., Wang, B.: Relation extraction with multi-instance multi-label convolutional neural networks. In: Proceedings

of the 26th International Conference on Computational Linguistics: Technical Papers, pp. 1471–1480 (2016)

[35] Zhang, D., Wang, D.: Relation classification via recurrent neural network. arXiv preprint arXiv:1508.01006 (2015)

[36] Zhang, S., Zheng, D., Hu, X., Yang, M.: Bidirectional long short-term memory networks for relation classification. In: Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, pp. 73–78 (2015)

[37] Miwa, M., Bansal, M.: End-to-end relation extraction using lstms on sequences and tree structures. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1105–1116 (2016)

[38] Zhang, Y., Qi, P., Manning, C.D.: Graph convolution over pruned dependency trees improves relation extraction. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2205–2215 (2018)

[39] Vashishth, S., Joshi, R., Prayaga, S.S., Bhattacharyya, C., Talukdar, P.: RESIDE: Improving distantly-supervised neural relation extraction using side information. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 1257–1266 (2018)

[40] Guo, Z., Zhang, Y., Lu, W.: Attention guided graph convolutional networks for relation extraction. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 241–251 (2019)

[41] Zhang, N., Deng, S., Sun, Z., Wang, G., Chen, X., Zhang, W., Chen, H.: Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 3016–3025 (2019)

[42] Guo, X., Hsu, I., AbdAlmageed, W., Natarajan, P., Peng, N.: Mrgcn: Mirror graph convolution network for relation extraction with long-term dependencies. CoRR (2021)

[43] Lin, Y., Shen, S., Liu, Z., Luan, H., Sun, M.: Neural relation extraction with selective attention over instances. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2124–2133 (2016)

[44] Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., Xu, B.: Attention-based bidirectional long short-term memory networks for relation classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (volume 2: Short Papers), pp. 207–212 (2016)

[45] Lin, Y., Liu, Z., Sun, M.: Neural relation extraction with multi-lingual attention. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 34–43 (2017)

[46] Wang, G., Zhang, W., Wang, R., Zhou, Y., Chen, X., Zhang, W., Zhu, H., Chen, H.: Label-free distant supervision for relation extraction via knowledge graph embedding. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2246–2255 (2018)

[47] Qin, P., Xu, W., Wang, W.Y.: Robust distant supervision relation extraction via deep reinforcement learning. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2137–2147 (2018)

[48] Qin, P., Xu, W., Wang, W.Y.: DSGAN: Generative adversarial training for distant supervision relation extraction. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 496–505 (2018)

[49] Huang, Y.Y., Wang, W.Y.: Deep residual learning for weakly-supervised relation extraction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1803–1807 (2017)

[50] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 260–270 (2016)

[51] Chiu, J.P., Nichols, E.: Named entity recognition with bidirectional lstm-cnns. Transactions of the Association for Computational Linguistics **4**, 357–370 (2016)

[52] Žukov-Gregorič, A., Bachrach, Y., Coope, S.: Named entity recognition with parallel recurrent neural networks. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 69–74 (2018)

[53] Xia, C., Zhang, C., Yang, T., Li, Y., Du, N., Wu, X., Fan, W., Ma, F.,

Yu, P.: Multi-grained named entity recognition. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, pp. 1430–1440 (2019)

[54] Li, X., Feng, J., Meng, Y., Han, Q., Wu, F., Li, J.: A unified MRC framework for named entity recognition. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 5849–5859 (2020)

[55] Liu, W., Zhou, P., Zhao, Z., Wang, Z., Ju, Q., Deng, H., Wang, P.: K-bert: Enabling language representation with knowledge graph. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, vol. 34, pp. 2901–2908 (2020)

[56] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 4171–4186 (2019)

[57] Nie, B., Ding, R., Xie, P., Huang, F., Qian, C., Si, L.: Knowledge-aware named entity recognition with alleviating heterogeneity. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, vol. 35, pp. 13595–13603 (2021)

[58] Huang, H., Heck, L., Ji, H.: Leveraging deep neural networks and knowledge graphs for entity disambiguation. arXiv preprint arXiv:1504.07678 (2015)

[59] Ganea, O.-E., Hofmann, T.: Deep joint entity disambiguation with local neural attention. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2619–2629 (2017)

[60] Phan, M.C., Sun, A., Tay, Y., Han, J., Li, C.: Neupl: Attention-based semantic matching and pair-linking for entity disambiguation. In: Proceedings of the 26th ACM on Conference on Information and Knowledge Management, pp. 1667–1676 (2017)

[61] Cao, Y., Hou, L., Li, J., Liu, Z.: Neural collective entity linking. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 675–686 (2018)

[62] Le, P., Titov, I.: Improving entity linking by modeling latent relations between mentions. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1595–1604 (2018)

[63] Orr, L.J., Leszczynski, M., Guha, N., Wu, S., Arora, S., Ling, X., Ré, C.: Bootleg: Chasing the tail with self-supervised named entity disambiguation. In: Proceedings of 11th Conference on Innovative Data Systems Research (2021)

[64] Zhu, L., Ghasemi-Gol, M., Szekely, P., Galstyan, A., Knoblock, C.A.: Unsupervised entity resolution on multi-type graphs. In: Proceedings of the 15th International Semantic Web Conference, pp. 649–667 (2016)

[65] Sarawagi, S., Bhamidipaty, A.: Interactive deduplication using active learning. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 269–278 (2002)

[66] Sherif, M.A., Ngomo, A.-C.N., Lehmann, J.: Wombat–a generalization approach for automatic link discovery. In: Proceedings of the 14th European Semantic Web Conference, pp. 103–119 (2017)

[67] Kum, H.-C., Krishnamurthy, A., Machanavajjhala, A., Reiter, M.K., Ahalt, S.: Privacy preserving interactive record linkage (ppirl). Journal of the American Medical Informatics Association **21**(2), 212–220 (2014)

[68] Ragan, E.D., Kum, H.-C., Ilangovan, G., Wang, H.: Balancing privacy and information disclosure in interactive record linkage with visual masking. In: Proceedings of the 2018 Conference on Human Factors in Computing Systems, pp. 1–12 (2018)

[69] Karim, M., Ramezani, M., Sunbury, T., Ohsfeldt, R., Kum, H.-C.: View: a framework for organization level interactive record linkage to support reproducible data science. arXiv preprint arXiv:2102.08273 (2021)

[70] Konda, P., Das, S., Doan, A., Ardalan, A., Ballard, J.R., Li, H., Panahi, F., Zhang, H., Naughton, J., Prasad, S., *et al.*: Magellan: toward building entity matching management systems over data science stacks. Proceedings of the 42nd International Conference on Very Large Data Bases **9**(13), 1581–1584 (2016)

[71] Ebraheem, M., Thirumuruganathan, S., Joty, S., Ouzzani, M., Tang, N.: Deeper–deep entity resolution. arXiv preprint arXiv:1710.00597 (2017)

[72] Mudgal, S., Li, H., Rekatsinas, T., Doan, A., Park, Y., Krishnan, G., Deep, R., Arcaute, E., Raghavendra, V.: Deep learning for entity matching: A design space exploration. In: Proceedings of the 2018 International Conference on Management of Data, pp. 19–34 (2018)

[73] Li, B., Wang, W., Sun, Y., Zhang, L., Ali, M.A., Wang, Y.: Grapher: Token-centric entity resolution with graph convolutional neural networks. In: Proceedings of the 34th AAAI Conference on Artificial

Intelligence, vol. 34, pp. 8172–8179 (2020)

[74] Yahya, M., Whang, S., Gupta, R., Halevy, A.: Renoun: Fact extraction for nominal attributes. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, pp. 325–335 (2014)

[75] Jiang, M., Shang, J., Cassidy, T., Ren, X., Kaplan, L.M., Hanratty, T.P., Han, J.: Metapad: Meta pattern discovery from massive text corpora. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 877–886 (2017)

[76] Zhao, G., Zhang, T., Wang, C., Lv, P., Wu, J.: Applications of bert based sequence tagging models on chinese medical text attributes extraction. arXiv preprint arXiv:2008.09740 (2020)

[77] IV, R.L.L., Humeau, S., Singh, S.: Multimodal attribute extraction. In: Proceedings of the 6th Workshop on Automated Knowledge Base Construction (2017)

[78] Sun, K., Liu, J., Yu, S., Xu, B., Xia, F.: Graph force learning. In: Proceedings of the 2020 IEEE International Conference on Big Data, pp. 2987–2994 (2020)

[79] Sun, K., Wang, L., Xu, B., Zhao, W., Teng, S.W., Xia, F.: Network representation learning: From traditional feature learning to deep learning. IEEE Access **8**, 205600–205617 (2020)

[80] Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: Online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710 (2014)

[81] Grover, A., Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 855–864 (2016)

[82] Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1067–1077 (2015)

[83] Wang, D., Cui, P., Zhu, W.: Structural deep network embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1225–1234 (2016)

[84] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: Proceedings of the 5th International Conference

on Learning Representations (2017)

[85] Hamilton, W.L., Ying, R., Leskovec, J.: Inductive representation learning on large graphs. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 1025–1035 (2017)

[86] Xia, F., Wang, L., Tang, T., Chen, X., Kong, X., Oatley, G., King, I.: CenGCN: Centralized convolutional networks with vertex imbalance for scale-free graphs. IEEE Transactions on Knowledge and Data Engineering (2022)

[87] Bordes, A., Usunier, N., Garcia-Durán, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2, pp. 2787–2795 (2013)

[88] Wang, Z., Zhang, J., Feng, J., Chen, Z.: Knowledge graph embedding by translating on hyperplanes. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence, vol. 28 (2014)

[89] Lin, Y., Liu, Z., Sun, M., Liu, Y., Zhu, X.: Learning entity and relation embeddings for knowledge graph completion. In: Proceedings of the 29th AAAI Conference on Artificial Intelligence, pp. 2215–2218 (2015)

[90] Ji, G., He, S., Xu, L., Liu, K., Zhao, J.: Knowledge graph embedding via dynamic mapping matrix. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pp. 687–696 (2015)

[91] Xiao, H., Huang, M., Zhu, X.: Transg: A generative model for knowledge graph embedding, 2316–2325 (2016)

[92] Jia, Y., Wang, Y., Lin, H., Jin, X., Cheng, X.: Locally adaptive translation for knowledge graph embedding. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, pp. 1–7 (2016)

[93] Qian, W., Fu, C., Zhu, Y., Cai, D., He, X.: Translating embeddings for knowledge graph completion with relation attention mechanism. In: Proceedings of the 27th International Joint Conference on Artificial Intelligence, pp. 4286–4292 (2018)

[94] Lv, X., Hou, L., Li, J., Liu, Z.: Differentiating concepts and instances for knowledge graph embedding. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 1971–1979 (2018)

[95] Cui, Z., Liu, S., Pan, L., He, Q.: Translating embedding with local connection for knowledge graph completion. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, pp. 1825–1827 (2020)

[96] Zhang, F., Wang, X., Li, Z., Li, J.: Transrhs: a representation learning method for knowledge graphs with relation hierarchical structure. In: Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence, pp. 2987–2993 (2021)

[97] d'Amato, C., Quatraro, N.F., Fanizzi, N.: Injecting background knowledge into embedding models for predictive tasks on knowledge graphs. In: Proceedings of the 18th International Conference on European Semantic Web Conference, pp. 441–457 (2021)

[98] Nickel, M., Tresp, V., Kriegel, H.-P.: A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on Machine Learning, pp. 809–816 (2011)

[99] Yang, B., Yih, W., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of the 3rd International Conference on Learning Representations

[100] Nickel, M., Rosasco, L., Poggio, T.: Holographic embeddings of knowledge graphs. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, vol. 30 (2016)

[101] Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: Proceedings of the 33rd International Conference on Machine Learning, pp. 2071–2080 (2016)

[102] Liu, H., Wu, Y., Yang, Y.: Analogical inference for multi-relational embeddings. In: Proceedings of the 34th International Conference on Machine Learning, pp. 2168–2178 (2017)

[103] Kazemi, S.M., Poole, D.: Simple embedding for link prediction in knowledge graphs. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 4289–4300 (2018)

[104] Balazevic, I., Allen, C., Hospedales, T.: TuckER: Tensor factorization for knowledge graph completion. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, pp. 5185–5194 (2019)

[105] Cao, Z., Xu, Q., Yang, Z., Cao, X., Huang, Q.: Dual quaternion knowledge graph embeddings. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, vol. 35, pp. 6894–6902 (2021)

[106] Demir, C., Ngomo, A.-C.N.: Convolutional complex knowledge graph embeddings. In: Proceedings of the 18th European Semantic Web Conference, pp. 409–424 (2021)

[107] Schlichtkrull, M., Kipf, T.N., Bloem, P., Van Den Berg, R., Titov, I., Welling, M.: Modeling relational data with graph convolutional networks. In: Proceedings of the 15th European Semantic Web Conference, pp. 593–607 (2018)

[108] Shang, C., Tang, Y., Huang, J., Bi, J., He, X., Zhou, B.: End-to-end structure-aware convolutional networks for knowledge base completion. In: Proceedings of the 33th AAAI Conference on Artificial Intelligence, vol. 33, pp. 3060–3067 (2019)

[109] Nathani, D., Chauhan, J., Sharma, C., Kaul, M.: Learning attention-based embeddings for relation prediction in knowledge graphs. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 4710–4723 (2019)

[110] Ye, R., Li, X., Fang, Y., Zang, H., Wang, M.: A vectorized relational graph convolutional network for multi-relational network alignment. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence, pp. 4135–4141 (2019)

[111] Zhang, Z., Zhuang, F., Zhu, H., Shi, Z., Xiong, H., He, Q.: Relational graph neural network with hierarchical attention for knowledge graph completion. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, vol. 34, pp. 9612–9619 (2020)

[112] Yu, D., Yang, Y., Zhang, R., Wu, Y.: Knowledge embedding based graph convolutional network. In: Proceedings of the 30th International Conference on World Wide Web, pp. 1619–1628 (2021)

[113] Socher, R., Chen, D., Manning, C.D., Ng, A.: Reasoning with neural tensor networks for knowledge base completion. In: Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems, pp. 926–934 (2013)

[114] Bordes, A., Glorot, X., Weston, J., Bengio, Y.: A semantic matching energy function for learning with multi-relational data. Machine Learning **94**(2), 233–259 (2014)

[115] Dettmers, T., Minervini, P., Stenetorp, P., Riedel, S.: Convolutional 2D knowledge graph embeddings. In: Proceedings of the 32ed AAAI Conference on Artificial Intelligence, pp. 1811–1818 (2018)

[116] 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: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), New Orleans, Louisiana, pp. 327–333 (2018)

[117] Jiang, X., Wang, Q., Wang, B.: Adaptive convolution for multi-relational learning. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp. 978–987 (2019)

[118] Liu, Q., Jiang, H., Evdokimov, A., Ling, Z.-H., Zhu, X., Wei, S., Hu, Y.: Probabilistic reasoning via deep learning: Neural association models. arXiv preprint arXiv:1603.07704, 1–12 (2016)

[119] Jiang, T., Liu, T., Ge, T., Sha, L., Li, S., Chang, B., Sui, Z.: Encoding temporal information for time-aware link prediction. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2350–2354 (2016)

[120] Trivedi, R., Dai, H., Wang, Y., Song, L.: Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In: Proceedings of the 34th International Conference on Machine Learning, pp. 3462–3471 (2017)

[121] Leblay, J., Chekol, M.W.: Deriving validity time in knowledge graph. In: Companion Proceedings of the 27th International Conference on World Wide Web, pp. 1771–1776 (2018)

[122] Dasgupta, S.S., Ray, S.N., Talukdar, P.: Hyte: Hyperplane-based temporally aware knowledge graph embedding. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2001–2011 (2018)

[123] Goel, R., Kazemi, S.M., Brubaker, M., Poupart, P.: Diachronic embedding for temporal knowledge graph completion. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, vol. 34, pp. 3988–3995 (2020)

[124] García-Durán, A., Dumančić, S., Niepert, M.: Learning sequence encoders for temporal knowledge graph completion. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 4816–4821 (2018)

[125] Lacroix, T., Obozinski, G., Usunier, N.: Tensor decompositions for

temporal knowledge base completion. In: Proceedings of the 8th International Conference on Learning Representations, pp. 1–12 (2020)

[126] Li, Z., Jin, X., Li, W., Guan, S., Guo, J., Shen, H., Wang, Y., Cheng, X.: Temporal knowledge graph reasoning based on evolutional representation learning. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 408–417 (2021)

[127] Wu, J., Xu, Y., Zhang, Y., Ma, C., Coates, M., Cheung, J.C.K.: Tie: A framework for embedding-based incremental temporal knowledge graph completion. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 428–437 (2021)

[128] Yu, S., Feng, Y., Zhang, D., Bedru, H.D., Xu, B., Xia, F.: Motif discovery in networks: a survey. Computer Science Review **37**, 100267 (2020)

[129] Lao, N., Cohen, W.W.: Relational retrieval using a combination of path-constrained random walks. Machine Learning **81**(1), 53–67 (2010)

[130] Wang, Q., Liu, J., Luo, Y., Wang, B., Lin, C.-Y.: Knowledge base completion via coupled path ranking. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 1308–1318 (2016)

[131] Das, R., Neelakantan, A., Belanger, D., McCallum, A.: Chains of reasoning over entities, relations, and text using recurrent neural networks. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Valencia, Spain, pp. 132–141 (2017)

[132] Chen, Z., Wang, Y., Zhao, B., Cheng, J., Zhao, X., Duan, Z.: Knowledge graph completion: A review. IEEE Access **8**, 192435–192456 (2020)

[133] Richardson, M., Domingos, P.: Markov logic networks. Machine Learning **62**(1-2), 107–136 (2006)

[134] Qu, M., Tang, J.: Probabilistic logic neural networks for reasoning. Advances in Neural Information Processing Systems **32**, 1–11 (2019)

[135] Wan, G., Du, B.: Gaussianpath: A bayesian multi-hop reasoning framework for knowledge graph reasoning. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, vol. 35, pp. 4393–4401 (2021)

[136] Shi, B., Weninger, T.: Open-world knowledge graph completion. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, pp.

1957–1964 (2018)

[137] Guo, Q., Zhuang, F., Qin, C., Zhu, H., Xie, X., Xiong, H., He, Q.: A survey on knowledge graph-based recommender systems. IEEE Transactions on Knowledge and Data Engineering **34**(8), 3549–3568 (2020)

[138] Da'u, A., Salim, N.: Recommendation system based on deep learning methods: a systematic review and new directions. Artif. Intell. Rev. **53**(4), 2709–2748 (2020). https://doi.org/10.1007/s10462-019-09744-1

[139] Batmaz, Z., Yürekli, A., Bilge, A., Kaleli, C.: A review on deep learning for recommender systems: challenges and remedies. Artif. Intell. Rev. **52**(1), 1–37 (2019). https://doi.org/10.1007/s10462-018-9654-y

[140] Gao, Y., Li, Y.-F., Lin, Y., Gao, H., Khan, L.: Deep learning on knowledge graph for recommender system: A survey. arXiv preprint arXiv:2004.00387 (2020)

[141] Chowdhury, G., Srilakshmi, M., Chain, M., Sarkar, S.: Neural factorization for offer recommendation using knowledge graph embeddings. In: Proceedings of the SIGIR 2019 Workshop on eCommerce, Co-located with the 42st International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1–6 (2019)

[142] Duran, P.G., Karatzoglou, A., Vitrià, J., Xin, X., Arapakis, I.: Graph convolutional embeddings for recommender systems. IEEE Access **9**, 100173–100184 (2021)

[143] Palumbo, E., Monti, D., Rizzo, G., Troncy, R., Baralis, E.: entity2rec: Property-specific knowledge graph embeddings for item recommendation. Expert Systems with Applications **151**, 113235 (2020)

[144] Zhang, J., Shi, X., Zhao, S., King, I.: STAR-GCN: Stacked and reconstructed graph convolutional networks for recommender systems. In: Proceedings of the 28th International Joint Conference on Artificial Intelligence

[145] Zhang, J., Gao, M., Yu, J., Yang, L., Wang, Z., Xiong, Q.: Path-based reasoning over heterogeneous networks for recommendation via bidirectional modeling. Neurocomputing **461**, 438–449 (2021)

[146] Lin, Y., Xu, B., Feng, J., Lin, H., Xu, K.: Knowledge-enhanced recommendation using item embedding and path attention. Knowledge-Based Systems **233**, 107484 (2021)

[147] Zhao, Y., Wang, X., Chen, J., Wang, Y., Tang, W., He, X., Xie, H.:

Time-aware path reasoning on knowledge graph for recommendation. ACM Transactions on Information Systems (TOIS), 1–27

[148] Yusuf, A.A., Chong, F., Xianling, M.: An analysis of graph convolutional networks and recent datasets for visual question answering. Artificial Intelligence Review (2022). https://doi.org/10.1007/s10462-022-10151-2

[149] Yih, S.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 Joint Conference of the 53rd Annual Meeting of the ACL and the 7th International Joint Conference on Natural Language Processing of the AFNLP (2015)

[150] Yao, X., Van Durme, B.: Information extraction over structured data: Question answering with freebase. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 956–966 (2014)

[151] Luo, D., Su, J., Yu, S.: A bert-based approach with relation-aware attention for knowledge base question answering. In: Proceedings of the 2020 International Joint Conference on Neural Networks, pp. 1–8 (2020)

[152] Hao, Y., Zhang, Y., Liu, K., He, S., Liu, Z., Wu, H., Zhao, J.: An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 221–231 (2017)

[153] Sorokin, D., Gurevych, I.: Modeling semantics with gated graph neural networks for knowledge base question answering. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 3306–3317 (2018)

[154] Cai, J., Zhang, Z., Wu, F., Wang, J.: Deep cognitive reasoning network for multi-hop question answering over knowledge graphs. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pp. 219–229 (2021)

[155] Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., Pei, J., Long, B.: Graph neural networks for natural language processing: A survey. arXiv preprint arXiv:2106.06090 (2021)

[156] Logeswaran, L., Chang, M.-W., Lee, K., Toutanova, K., Devlin, J., Lee, H.: Zero-shot entity linking by reading entity descriptions. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 3449–3460 (2019)

[157] He, L., Zheng, S., Yang, T., Zhang, F.: KLMo: Knowledge graph enhanced pretrained language model with fine-grained relationships. In: Findings of the Association for Computational Linguistics: EMNLP 2021, pp. 4536–4542 (2021)

[158] Chen, D., Li, Y., Yang, M., Zheng, H.-T., Shen, Y.: Knowledge-aware textual entailment with graph attention network. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 2145–2148 (2019)

[159] Peng, H., Li, J., Gong, Q., Song, Y., Ning, Y., Lai, K., Yu, P.S.: Fine-grained event categorization with heterogeneous graph convolutional networks. Proceedings of the 28th International Joint Conference on Artificial Intelligence (2019)

[160] Dun, Y., Tu, K., Chen, C., Hou, C., Yuan, X.: Kan: Knowledge-aware attention network for fake news detection. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence, vol. 35, pp. 81–89 (2021)

[161] Liu, S., Zaraté, P.: Knowledge based decision support systems: A survey on technologies and application domains. In: Proceedings of the 14th Joint International Conference on Group Decision and Negotiation, pp. 62–72 (2014)

[162] Lv, S., Guo, D., Xu, J., Tang, D., Duan, N., Gong, M., Shou, L., Jiang, D., Cao, G., Hu, S.: Graph-based reasoning over heterogeneous external knowledge for commonsense question answering. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, vol. 34, pp. 8449–8456 (2020)

[163] Liu, Y., Wan, Y., He, L., Peng, H., Yu, P.S.: Kg-bart: Knowledge graph-augmented bart for generative commonsense reasoning. Proceedings of the 35th AAAI Conference on Artificial Intelligence (2020)

[164] Li, Z., Jin, X., Li, W., Guan, S., Guo, J., Shen, H., Wang, Y., Cheng, X.: Temporal knowledge graph reasoning based on evolutional representation learning. Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (2021)

[165] Moon, S., Shah, P., Kumar, A., Subba, R.: Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 845–854 (2019)

[166] Cheng, K., Yang, Z., Zhang, M., Sun, Y.: Uniker: A unified framework for combining embedding and definite horn rule reasoning for knowledge

graph inference. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 9753–9771 (2021)

[167] Roy, S., Chakraborty, S., Mandal, A., Balde, G., Sharma, P., Natarajan, A., Khosla, M., Sural, S., Ganguly, N.: Knowledge-aware neural networks for medical forum question classification. In: Proceedings of the 30th ACM International Conference on Information and Knowledge Management, pp. 3398–3402 (2021)

[168] Chai, X.: Diagnosis method of thyroid disease combining knowledge graph and deep learning. IEEE Access **8**, 149787–149795 (2020)

[169] Lu, Q., Nguyen, T.H., Dou, D.: Predicting patient readmission risk from medical text via knowledge graph enhanced multiview graph convolution. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1990–1994 (2021)

[170] Lin, X., Quan, Z., Wang, Z.-J., Ma, T., Zeng, X.: Kgnn: Knowledge graph neural network for drug-drug interaction prediction. In: Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence, vol. 380, pp. 2739–2745 (2020)

[171] Rao, G., Peng, C., Zhang, L., Wang, X., Feng, Z.: A knowledge enhanced ensemble learning model for mental disorder detection on social media. In: Proceedings of the 13th International Conference on Knowledge Science, Engineering and Management, pp. 181–192 (2020)

[172] Yu, S., Xia, F., Sun, Y., Tang, T., Yan, X., Lee, I.: Detecting outlier patterns with query-based artificially generated searching conditions. IEEE Transactions on Computational Social Systems **8**(1), 134–147 (2020)

[173] Nguyen, H.-V., Gelli, F., Poria, S.: Dozen: Cross-domain zero shot named entity recognition with knowledge graph. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1642–1646 (2021)

[174] Ouyang, Y., Guo, B., Wang, Q., Yu, Z.: Cross-domain recommendation with cross-graph knowledge transfer network. In: Proceedings of the 10th IEEE International Conference on Communications, pp. 1–6 (2021)

[175] Du, Z., Zhou, C., Ding, M., Yang, H., Tang, J.: Cognitive knowledge graph reasoning for one-shot relational learning. arXiv preprint arXiv:1906.05489 (2019)

[176] Gao, H., Wu, L., Hu, P., Wei, Z., Xu, F., Long, B.: Graph-augmented learning to rank for querying large-scale knowledge graph. arXiv preprint

arXiv:2111.10541 (2021)

[177] Senaratne, A., Omran, P.G., Williams, G., Christen, P.: Unsupervised anomaly detection in knowledge graphs. In: Proceedings of the 10th International Joint Conference on Knowledge Graphs, pp. 161–165 (2021)

[178] Wang, X., Wang, D., Xu, C., He, X., Cao, Y., Chua, T.-S.: Explainable reasoning over knowledge graphs for recommendation. In: Proceedings of the 33th AAAI Conference on Artificial Intelligence, vol. 33, pp. 5329–5336 (2019)

[179] Raza, S., Ding, C.: News recommender system: a review of recent progress, challenges, and opportunities. Artif. Intell. Rev. **55**(1), 749–800 (2022). https://doi.org/10.1007/s10462-021-10043-x

[180] Jung, J., Jung, J., Kang, U.: Learning to walk across time for interpretable temporal knowledge graph completion. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 786–795 (2021)