Expert Systems With Applications 141 (2020) 112948

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Review

A review: Knowledge reasoning over knowledge graph

Xiaojun Chen, Shengbin Jia, Yang Xiang*

College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 30 April 2019
Revised 8 September 2019
Accepted 10 September 2019
Available online 13 September 2019

Keywords:
Knowledge graph
Reasoning
Rule-based reasoning
Distributed representation-based reasoning
Neural network-based reasoning

ABSTRACT

Mining valuable hidden knowledge from large-scale data relies on the support of reasoning technology. Knowledge graphs, as a new type of knowledge representation, have gained much attention in natural language processing. Knowledge graphs can effectively organize and represent knowledge so that it can be efficiently utilized in advanced applications. Recently, reasoning over knowledge graphs has become a hot research topic, since it can obtain new knowledge and conclusions from existing data. Herein we review the basic concept and definitions of knowledge reasoning and the methods for reasoning over knowledge graphs. Specifically, we dissect the reasoning methods into three categories: rule-based reasoning, distributed representation-based reasoning and neural network-based reasoning. We also review the related applications of knowledge graph reasoning, such as knowledge graph completion, question answering, and recommender systems. Finally, we discuss the remaining challenges and research opportunities for knowledge graph reasoning.

Contents

* Corresponding author.
E-mail addresses: xiaojunchen@tongji.edu.cn (X. Chen), shengbinjia@tongji.edu.cn (S. Jia), tjdxxiangyang@gmail.com (Y. Xiang).

https://doi.org/10.1016/j.eswa.2019.112948

1. Introduction

Reasoning is one of the basic forms of simulated thinking, and a process of deducing new judgements (conclusions) from one or several existing judgements (premises). The reason why AlphaGo was able to win at Chinese chess is that it has super reasoning ability and artificial intelligence to provide a new interpretation from a small amount of data. Thus, reasoning ability is very important. The DeepMind points out that artificial intelligence algorithms must have reasoning capabilities and the process of reasoning must rely on prior knowledge and experience

In the era of knowledge engineering, a large number of knowledge graphs (KGs), such as YAGO (Suchanek, Kasneci, & Weikum, 2008), WordNet (Miller, 1995), and Freebase (Bollacker, Evans, Paritosh, Sturge, & Taylor, 2008) have been developed. KGs contain a large amount of prior knowledge but can also effectively organize data. They have been widely used for question-answering systems, search engines, and recommendation systems. Knowledge graphs are able to mine, organize, and effectively manage knowledge from large-scale data to improve the quality of information services and provide users with smarter services. All of these facets rely on the support of knowledge reasoning over knowledge graphs which is therefore one of the core technologies in the field of reasoning.

Knowledge reasoning over knowledge graphs aims to identify errors and infer new conclusions from existing data. New relations among entities can be derived through knowledge reasoning and can feed back to enrich the knowledge graphs, and then support the advanced applications. Considering the wide application foreground of knowledge graphs, the study of knowledge reasoning on large-scale knowledge graphs has become one research focus in natural language processing in the past few years.

Our main contributions are as follows: First, we complement existing work with 147 publications. Second, we identify problems faced by these methods. Finally, we discuss future directions of this research field. The rest of this paper is organized as follows. Section 2 states the methodology used to find and filter surveyed publications. Section 3 briefly introduces world's leading knowledge graphs and the definition of knowledge reasoning. Section 4 covers techniques that conduct reasoning based on rules. We describe the reasoning methods based on first order predicate logic rules, ontology and random walk algorithm. Section 5 discusses reasoning techniques that further utilize representation learning. Section 6 focuses on the techniques that perform reasoning based on neural network and reinforcement learning. Section 7 further explores the applications of such knowledge in downstream tasks, such as knowledge graph completion, question-answer systems, and recommendation systems. Section 8 discusses future research directions of knowledge graph reasoning. Finally, concluding remarks end the paper.

2. Methodology

This review follows a strict discovery methodology; inclusion and exclusion criteria that are used to search and restrict publications related to knowledge graph reasoning.

Inclusion criteria Candidate articles for inclusion in the survey need to be part of relevant conference proceedings or searchable via Google Scholar. The included papers from the publication search engine are the first 300 results that contain "'knowledge graph' AND ('reasoning' OR 'inference')" in the article including title, abstract and text body. Conference candidates are all publications from 2012 to 2019 in the proceedings of major natural language processing and artificial intelligence conferences, including ACL, EMNLP, NAACL, ISWC, CIKM, AAAI, NIPS, IJCAI, ICML, WWW, ICLR and COLING.

Exclusion criteria Works that are not related to knowledge graph reasoning are excluded, determined in a manual inspection in the following manner: First, proceeding tracks are excluded that clearly do not contain knowledge graph reasoning related publications. Next, publications both from proceedings and from Google Scholar are excluded based on their title and finally on their full text.

Result The inspection of the titles of the Google Scholar results by three authors of this article led to 164 publications, 66 of which selected after checking the content. The selected proceedings contain 414 publications. Based on their titles, 335 of them were selected and inspected, resulting in 81 publications that were categorized and listed in this survey. Table 1 shows the number of publications for each source. In total, 714 candidates were found using the inclusion criteria in Google Scholar and conference proceedings and then reduced according to exclusion criteria, resulting in 147 publications.

3. Introduction to knowledge reasoning

3.1. Definition of knowledge reasoning

Reasoning technique has a long history. As early as in ancient Greece period, the famous philosopher Aristotle proposed the syl-

Table 1
Sources of publication candidates along with the number of publications in total, after excluding based on the title (I), and finally based on the full text (selected). **Works that are found both in a conference's proceedings and in Google Scholar are only counted once, as selected for that conference.**.

Venue	All	I	Selected
Google Scholar Top 300	300	164	66
ACL	39	23	7
EMNLP	54	49	17
NAACL	46	27	5
ISWC	60	42	3
CIKM	45	29	4
AAAI	36	36	17
IJCAI	41	41	7
ICML	15	14	4
NIPS	34	34	6
WWW	30	26	6
ICLR	5	5	3
COLING	9	9	2
Conference	414	335	81
All	714	499	147

logism which is the basis of modern deductive reasoning. From the Lambda Calculus which defines computers to various intelligent computing platforms, and from expert system to large-scale knowledge graphs, all of which are inseparable from reasoning. With respect to the basic concepts of knowledge reasoning, academia has given different definitions. Zhang and Zhang (1992) pointed out that reasoning is the process of analyzing, synthesizing and making decisions on various things, starting from collecting the exist facts, discovering interrelationships between things, to developing new insight. In short, reasoning is the process of drawing conclusions from existing facts by the rules. Kompridis (2000) believed that reasoning is a collective term for a range of abilities, including capacity of understanding things, apply logic, and calibrate or validate architecture based on existing knowledge. Tari (2013) defined the concept of knowledge reasoning as the mechanism behind inferring new knowledge based on the existing facts and logic rules. In general, knowledge reasoning is the process of using known knowledge to infer new knowledge.

Early reasoning studies were carried out among scholars in the fields of logic and knowledge engineering. The scholars of logic advocated utilization of formalized methods to describe the objective world and it believed that all reasoning was based on existing logical knowledge, such as first-order logic and predicate logic (Wu, Han, Li, Zheng, & Chen, 2018). They always focused on how to draw correct conclusion from the known propositions and predicates. In order to lighten the rigidity of the reasoning process, methods such as non-monotonic reasoning (McCarthy, 1980) and fuzzy reasoning (Zadeh, 1965) were developed for the purpose of using it in the more complicated situations.

Unlike scholars from the Logic field who used propositions or first-order predicates to represent concepts in the objective world, the scholars from the knowledge engineering field used semantic networks to represent richer concepts and knowledge for describing the relationships between entities and attributes. Nevertheless, early knowledge graphs were totally relied on expert knowledge. The entities, attributes, and relationships in knowledge graph were entirely handcrafted by the experts in the fields, such as CyC (Lenat & Guha, 1989).

With the explosive growth of Internet data scale, traditional methods based on artificially built knowledge bases (KBs) cannot adapt to the need to mine a large amount of knowledge in the era of big data. For this reason, data-driven machine reasoning methods have gradually become the main stream of knowledge reasoning research.

Table 2
Examples of world's leading knowledge graphs and their statistics (Paulheim, 2017).

Knowledge graphs	#Entities	#Relations	#Facts
WordNet	0.15M	200,000	4.5M
Freebase	50M	38,000	3B
YAGO	17M	76	150M
DBpedia (En)	4.8M	2800	176M
Wikidata	16M	1673	66M
NELL	2M	425	120M

3.2. Introduction of leading knowledge graphs

In 2012, Google introduced its Knowledge Graph (Singhal, 2012) project and took advantage of it to improve query result relevancy and users' search experience. Due to the increasing amount of Web resources and release of linked open data (LOD) projects, many knowledge graphs have been constructed. In this section, we will present a brief introduction of the world's leading knowledge graphs. Table 2 shows examples of leading knowledge graphs and their statistics.

WordNet WordNet is a lexical database for the English language. WordNet was created by the Cognitive Science Laboratory of Princeton University in 1985. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations, like the IS-A relation between dog and mammal or the PART-WHOLE relation between car and engine. WordNet has been used for a number of purposes in information systems, including word-sense disambiguation, information retrieval, text classification, text summarization, machine translation, and even crossword puzzle generation. WordNet version 3.0 is the latest version available and contains more than 150,000 words and 200,000 semantic relations.

Freebase Freebase is a large collaborative knowledge base consisting of data composed mainly by its community members. It was constructed by Metaweb. Freebase contains data harvested from sources such as Wikipedia, NNDB, Fashion Model Directory, and MusicBrainz, as well as data contributed by its users. Freebase's subjects are called 'topics', and the data stores about them depended on their 'type', types themselves are grouped into 'domains'. Google's Knowledge Graph is powered in part by Freebase. There are about 3 billion triples currently in Freebase.

YAGO YAGO (Suchanek, Kasneci, & Weikum, 2007) is an open source knowledge base developed by the Max Planck Institute. The information in YAGO is extracted from Wikipedia (e.g., categories, redirects, infoboxes), WordNet (e.g., synsets, hyponymy), and GeoNames. YAGO combines the clean taxonomy of WordNet with the richness of the Wikipedia category system, assigning the entities to more than 350,000 classes. YAGO attaches a temporal dimension and a spatial dimension to many of its facts and entities. It extracts and combines entities and facts from 10 Wikipedias in different languages. Currently, YAGO has knowledge of more than 17 million entities (like persons, organizations, cities, etc.) and contains more than 150 million facts about these entities. YAGO has been used in the Watson artificial intelligence system.

DBpedia DBpedia is a cross-language project aiming to extract structured content from the information created in the Wikipedia project. There are more than 45 million interlinks between DBpedia and external datasets including Freebase, OpenCyc, etc. DBpedia uses the resource description framework (RDF) to represent extracted information. The entities of DBpedia are classified in a consistent ontology, including persons, places, music albums, films, video games, organizations, species, and diseases. DBpedia was used as one of the knowledge sources in IBM Watson's Jeopardy!

winning system and can be integrated into Amazon Web Services applications.

Wikidata Wikidata is a multilingual, open, linked, structured knowledge base that can be read and edited by both humans and machines. It supports more than 280 language versions of Wikipedia with common source of structured data. Wikidata inherits crowdsourcing collaboration mechanism from Wikipedia and also supports editing based on triples. It relies on the notions of item and statement. An item represents an entity. A statement is composed of one main property-value pair that encodes the fact like "taxon name is Pantera Leo" and optional qualifiers to add information about it like "taxon author is Carl Linnaeus" (Pellissier Tanon, Vrandečić, Schaffert, Steiner, & Pintscher, 2016).

NELL Never-Ending Language Learning system (NELL) is a semantic machine learning system that runs 24/7, forever, learning to read the web. It is developed by a research team at Carnegie Mellon University. The inputs to NELL include (1) an initial ontology defining hundreds of categories and relations that NELL is expected to read about, and (2) 10 to 15 seed examples of each category and relation. Given these inputs, NELL automatically extracts triple facts from the Web. So far, NELL has accumulated over 120 million candidate beliefs by reading the web, and it is considering these at different levels of confidence, along with hundreds of learned phrasings, morphological features, and web page structures that NELL uses to extract beliefs from the web.

3.3. Knowledge reasoning oriented knowledge graph

With the development of knowledge graphs, reasoning over knowledge graphs has also increased a general concern. Referring to the definition of reasoning, we give the definition of reasoning over knowledge graphs as follows:

Definition 1Knowledge reasoning over KGs:. Given a knowledge graph $KG = < E, R, T >$ and the relation path *P*, where *E*, *T* represent the set of entities, *R* denotes the set of relations, and the edges in *R* link two nodes to form a triple $(h, r, t) \in T$, generating a triplet that does not exist in the KG $G' = \{(h, r, t) | h \in E, r \in R, t \in T, (h, r, t) \notin G\}$.

Its goal is to use machine learning methods to infer potential relations between entity pairs and identify erroneous knowledge based on existing data automatically with the purpose of complementing KGs. For examples, if the KG contains a fact like (*Microsoft, IsBasedIn, Seattle*), (*Seattle, StateLocatedIn, Washington*) and (*Washington, CountryLocatedIn, USA*), then we obtain the missing link (*Microsoft, HeadquarterLocatedIn, USA*). The object of knowledge reasoning is not only the attributes and relations between entities, but also the attribute values of entities and the conceptual hierarchy of ontology. For example, if an entity's identity card number attribute is known, the entity's gender, age, and other attributes can be obtained through reasoning.

KG is basically a semantic network and a structured semantic knowledge base which can formally interpret concepts and their relations in the real world (Xu, Sheng, He, & Wang, 2016). There is no need for knowledge graph to adopt cumbersome structure such as framework (Minsky, 1988) and script (Norenzayan, Smith, Kim, & Nisbett, 2002) in structed expressions, instead of simple triples with more flexible forms. Therefore, reasoning over knowledge graph is not limited to traditional reasoning methods based on logic and rules, but also can be diverse. At the same time, knowledge graph consists of instances, which makes the reasoning methods more concrete.

In recent years, researchers have implemented many open information extraction (OIE) systems, such as TextRunner (Banko, Cafarella, Soderland, Broadhead, & Etzioni, 2007), WOE (Wu & Weld, 2010), which greatly expands the data source for knowledge graph construction Akbik and Löser (2012); Banko et al. (2007); Fader, Soderland, and Etzioni (2011); Kertkeidkachorn and Ichise (2017); Wu and Weld (2010) and Zhang et al. (2019a). So, the rich contents of knowledge bases provide new opportunities and challenges for the development of knowledge reasoning technology. With the popularity of knowledge representation learning, neural networks and other technologies, a series of new reasoning methods have been coming out.

4. Knowledge reasoning based on logic rules

Early knowledge reasoning approaches including ontology reasoning have received much attention and produced a series of reasoning methods. Furthermore, these methods including predicate logic reasoning, ontology reasoning, and random walk reasoning, can be applied for reasoning over knowledge graphs.

4.1. Knowledge reasoning based on first-order predicate logic rules

Reasoning mainly relies on first-order predicate logic rules in the early stage of statistical relational learning study. First-order predicate logic uses propositions as the basic unit for reasoning, while propositions contain individuals and predictions. Individuals that can exist independently correspond to entity objects in the knowledge base. They can be a concrete thing or an abstract concept. The predicate is used to describe the nature and things of the individual. For example, interpersonal relationships can be reasoned using first-order predicate logic by regarding relationships as predicates, characters as variables, and using logical operators to express interpersonal relationships, and then setting the logic and constraints of relational reasoning to perform simple reasoning. The process of reasoning using first-order predicate logic is given in the following formula,

(*YaoMing, wasBornIn, Shanghai*)∧(*Shanghai, locatedIn, China*)⇒(*YaoMing, nationality, China*)

First-Order Inductive Learner (FOIL) (Schoenmackers, Etzioni, Weld, & Davis, 2010) is a typical work of predicate logic, which aims to search all the relations in the KG and acquire the Horn clauses set of each relation as a feature pattern for predicting whether the correspondence exists. Finally, the relation discrimination model is obtained using the machine learning method. There are a large number of related works about FOIL. For example, nFOIL and tFOIL (Landwehr, Kersting, & Raedt, 2007) integrate the naïve Bayes learning scheme and a tree augmented naïve Bayes with FOIL respectively. nFOIL guides the structure search by the probabilistic score of naïve Bayes. tFOIL relaxes the naïve Bayes assumption to allow additional probabilistic dependencies between clauses. kFOIL (Landwehr, Passerini, De Raedt, & Frasconi, 2010) combines FOIL's rule learning algorithm and kernel methods to derive a set of features from a relational representation. So, FOIL searches relevant clauses that can be used as features in kernel methods. Nakashole, Sozio, Suchanek, and Theobald (2012) present a query-time first-order reasoning approach for uncertain RDF knowledge bases with a combination of soft deduction rules and hard rules. Soft rules are used for deriving new facts, while hard rules are used to enforce consistency constraints among both KG and inferred facts. Galárraga, Teflioudi, Hose, and Suchanek (2013) propose the AMIE system for mining Horn rules on a knowledge graph. By applying these rules to the KBs, new facts can be derived for complementing knowledge graphs and detecting errors.

Traditional FOIL algorithms achieve high inference accuracy on small-scale knowledge bases. In addition, experimental results show that the "entity-relation" association model has strong reasoning ability. However, it is difficult to exhaust all inference patterns due to the complexity and diversity of entities and relations in large-scale knowledge graphs. In addition, the high complexity and the low efficiency of exhaustive algorithms make original FOIL algorithm inappropriate for reasoning over large-scale graphs. To solve this problem, Galárraga, Teflioudi, Hose, and Suchanek (2015) extend AMIE to AMIE+ by a series of pruning and query rewriting techniques for mining even larger KBs. Additionally, AMIE+ increases the precision of the predictions by considering type information and using joint reasoning. Demeester, Rocktäschel, and Riedel (2016b) present a scalable method to incorporate first-order implications into relation representations to improve large scale KG inference. While AMIE+ mines a single rule at a time, Wang and Li (2015) propose a novel rule learning approach named RDF2Rules. RDF2Rules mines Frequent Predicate Cycles (FPCs) to parallelize this process. It is more efficient to deal with large-scale KBs than AMIE+ due to a proper pruning strategy.

To formalize the semantic web and inference efficiently, some researchers proposed a tractable language, called description logic (DL). Description logic is a crucial foundation for ontology reasoning that was developed on the basis of propositional logic and first-order predicate logic. The goal of description logic is to balance representation power and reasoning complexity. It can provide well-defined semantics and powerful reasoning tools for knowledge graphs and satisfy the needs of ontology construction, integration and evolution. Therefore, it is an ideal ontology language. A KB expressed using a DL is composed of terminological axioms (TBox) and assertional axioms (ABox) (Lee, Lewicki, Girolami, & Sejnowski, 1999). The TBox is composed of a collection of inclusion assertions stating general properties of concepts and roles. For instance, an assertion is the one that states that a concept denotes a specialization of another concept. The ABox consists of assertions on individual objects. The consistency of the knowledge base is the basic problem in knowledge graph reasoning. The complex entity or relation reasoning in a knowledge graphs can be transformed into a consistency detection problem through TBox and ABox, thus refining and realizing knowledge reasoning. Halaschek-Wiener, Parsia, Sirin, and Kalyanpur (2006) present an description logic reasoning algorithm for complementing knowledge graphs under both the addition and removal of ABoxes assertions. It provides a critical step towards reasoning over fluctuating/streaming data. Calvanese, De Giacomo, Lembo, Lenzerini, and Rosati (2006) propose the language EQL based on an epistemic first-order query language, which is able to reason about incompleteness for querying description logic knowledge graphs. A large number of fuzzy description logics are proposed to extend classical description logics with fuzzy capability. Li, Xu, Lu, and Kang (2006) propose a novel discrete tableau algorithm for satisfiability of $\mathcal{FSHI}$ knowledge bases with general TBoxes, which supports a new way to achieve reasoning with general TBoxes in fuzzy DLs. Furthermore, Stoilos, Stamou, Pan, Tzouvaras, and Horrocks (2007) extend DL with fuzzy set theory in order to represent knowledge and perform reasoning tasks. To equip description logics for dealing with metaknowledge, Krötzsch, Marx, Ozaki, and Thost (2018) enrich DL concepts and roles with finite sets of attribute-value pairs, called attributed description logics, for knowledge graph reasoning. Existing DL reasoners do not provide users explanation services. To address this problem, Bienvenu, Bourgaux, and Goasdoué (2019) develop a framework to equip reasoning system with explanation ability under inconsistency-tolerant semantics.

4.2. Knowledge reasoning based on rule

The basic idea of rule-based knowledge reasoning models is to reason over KG by applying simple rules or statistical features. The reasoning component of Never-Ending Language Learning system (NELLs) (Mitchell et al., 2015) learns the probability rule and then instantiates the rule after manual screening, finally inferring a new relationship instance from other learned relation instances. Spass-YAGO expands the knowledge graph by abstracting the triples into equivalent rule classes. Paulheim and Bizer (2014) propose SDType and SDValidate that exploit statistical distributions of properties and types for type completion and error detection. SDType uses the statistical distribution of types in the head entity and tail entity position of the property for predicting the entities' types. SDValidate computes the relative predicate frequency (RPF) for each statement, with a low RPF value meaning incorrect. Jang and Megawati (2015) present a new approach for evaluating the quality of knowledge graph. They choose the patterns appearing more frequently as the generated test patterns for evaluating the quality of knowledge graph after analyzing the data patterns. Wang, Mazaitis, and Cohen (2013) and Wang, Mazaitis, Lao, and Cohen (2015) propose Programming with Personalized PageRank (ProPPR) for reasoning over a knowledge graph. Reasoning for ProPPR is based on a personalized PageRank process over the proof constructed by SLD resolution theorem-prover. Catherine and Cohen (2016) have shown that ProPPR can be used for performing knowledge graph recommendations. They formulate the problem as a probabilistic inference and learning task. Cohen (2016) propose TensorLog, where inference uses a differentiable process. Inspired by TensorLog, Yang, Yang, and Cohen (2017) describe a framework, neural logic programming, in which the structure and parameter learning of logical rules are combined in an end-to-end differentiable model.

Rule-based reasoning methods can also combine manually defined logic rules with various probability graph models and then obtain new facts by performing knowledge reasoning based on the constructed logical network. For example, Jiang, Lowd, and Dou (2012) propose a Markov logic-based system for cleaning NELL. This allows knowledge bases to make use of joint probabilistic reasoning, or, applies Markov logic network (MLN) (Richardson & Domingos, 2006) to a web-scale problem. It uses only the ontological constraints and confidence scores of the initial system, and labelled data. Chen and Wang (2014) present a probabilistic knowledge base (ProbKB), which allows an efficient SQL-based inference algorithm for knowledge completion that applies MLN inference rules in batches. Kuželka and Davis (2019) theoretically study the suitability of learning the weights of a Markov logic network from a KB in the presence of missing data. After learning the weights, an MLN could be used to infer additional facts to complete knowledge graphs. However, it is difficult to introduce clause confidence into MLN, because the clause value in logic rules must be Boolean variables. Moreover, various combinations of Boolean variable assignments make learning and reasoning difficult to optimize. To solve this problem, probabilistic soft logic (PSL) (Kimmig, Bach, Broecheler, Huang, & Getoor, 2012) is proposed. PSL uses FOIL rules as a template language for graphical models over random variables with soft truth values ranging in the interval [0,1]. Reasoning in this setting is considered as a continuous optimization task, which can be handled efficiently. For this reason, Pujara, Miao, Getoor, and Cohen (2013a) use PSL to reason candidate facts and their relevant extraction confidences collectively, recognize co-referent entities, and incorporate ontological constraints. Furthermore, they propose a partitioning technique (Pujara, Miao, Getoor, & Cohen, 2013b) to reason over large-scale knowledge graph with considering balancing the reasoning speed and accuracy. The method first generates a knowledge graph where entities and relations are nodes, ontological constraints are edges. Then the edge min-cut, a clustering technique, is used to partition the relations and labels. Finally, it uses PSL to define a joint probability distribution over knowledge graphs to accomplish collective reasoning. Bach, Broecheler, Huang, and Getoor (2017) propose Hinge-Loss Markov Random Fields (HL-MRFs), which can capture relaxed, probabilistic inference with Boolean logic and exact, probabilistic inference with fuzzy logic, making them useful models for both discrete and continuous data. They also introduce PSL to make HL-MRFs easy to define and use for large KGs.

4.3. Knowledge reasoning based on ontology

Knowledge reasoning over knowledge graphs, which is intimately bound up with ontology languages such as Resource Description Framework Schema (RDFS) and Web Ontology Language (OWL), is closely related to ontology. A knowledge graph can be regarded as a data structure of knowledge storage. Although it does not have formal semantics, it can reason by applying RDFS or OWL rules to a KG. Pujara et al. (2013b) have proven that the ontology represented by OWL EL is suitable for being transformed into a KG and perform reasoning on it efficiently. The reasoning method based on ontology mainly uses the more abstract frequent patterns, constraints or paths to infer. When reasoning through the ontology conceptual layer, the concept is mainly described by the OWL. OWL is able to provide rich statements and is capable of knowledge representation.

Zou, Finin, and Chen (2004) propose an inference engine F-OWL which use a Frame-based system to reason with OWL ontologies. F-OWL supports consistency checking of the knowledge base, extracts hidden knowledge via resolution and supports further complex reasoning by importing rules. Sirin, Parsia, Grau, Kalyanpur, and Katz (2007) present the OWL-DL reasoner Pellet to support incremental reasoning against dynamic knowledge graphs through reusing the reasoning results from previous steps to update the process incrementally. Chen, Goldberg, Wang, and Johri (2016) propose the ontological pathfinding (OP) algorithm that generalizes to web-scale KBs through a range of optimization and parallelization technologies: a relational KB model to use reasoning rules in turn, a novel rule mining algorithm to divide the mining tasks into smaller sole child tasks, and a pruning strategy to remove noisy and resource-consuming rules before using them. Wei, Luo, and Xie (2016a) propose and implement a distributed knowledge graph reasoning system (KGRL) based on OWL2 RL inference rules. KGRL has a more powerful reasoning ability due to more expressive rules. It can eliminate redundant data and make the reasoning result more compact through optimization. In addition, it can also find the inconsistent data within a knowledge graph.

For reasoning methods based on ontology to be efficient, it is important that they are scalable to large-scale knowledge graphs. Zhou et al. (2006) present a storage and inference system Minerva for large-scale OWL ontologies. Minerva combines a DL reasoner and a rule engine for ontology inference to improve efficiency. In order to improve the scalability and performance of reasoning, Soma and Prasanna (2008) propose two methods to parallelize the inference process for OWL knowledge bases. In the data partitioning approach, knowledge graph is partitioned and the complete rule-base is applied to each subset of the KG. In the rule-base partitioning approach, the rule-base is partitioned and each node of a parallel system applies one subset of rules to the original KG. Chen, Chen, Zhang, Chen, and Wu (2013b) present an OWL reasoning framework for massive and complex biomedical knowledge graph, which takes advantage of MapReduce algorithm and OWL property chain reasoning method. Recently, Marx, Krötzsch, and Thost (2017) present a simpler, rule-based fragment of multi-attributed predicate logic that can be used for ontological reasoning on a large knowledge graph.

4.4. Knowledge reasoning based on random walk algorithm

A line of research has proven that incorporating path rules into knowledge reasoning can improve inference performance. Inspired by this, many researchers have injected path rules into knowledge reasoning tasks. The path ranking algorithm (PRA) (Lao & Cohen, 2010) is a general technique for performing reasoning in a graph. To learn an inference model for a particular edge type in a KB, PRA finds sequences of edge types that frequently link nodes that are instances of the edge type being predicted. PRA then use those types as features in a logistic regression model to predict missing edges in the graph. A typical PRA model is composed of three components: feature extraction, feature computation, and relation-specific classification. The first step is to find a set of latent valuable path types that link the entity pairs. To this end, PRA performs a path constraint random walk over the graph to record those starting from h and ending at t with limited lengths. The second step is to compute the values in the feature matrix by calculating random walk probabilities. Given a node pair (h, t), and a path π, PRA computes the feature value as a random walk probability $p(t|h, \pi)$, i.e., the likelihood of reaching t when given a random starting from h and following relations contained in π. It is calculated as follows:

$$p(t|h,\pi) = \sum_{e' \in range(\pi')} p\big(h, e'; \pi'\big) P\big(t|e'; r_l\big)$$

where $P\big(t|e'; r_l\big) = \frac{r_l(e',t)}{|r_l(e',t)|}$. Then, the probability of a specific relation r between an entity pair (h, t) is calculated. The last step is to train each relation and obtain the weight of path features using a logistic regression algorithm.

The PRA model not only has high accuracy but also significantly improves the computational efficiency, and provides an effective solution to solve the problem of reasoning over large-scale knowledge graphs. Lao, Mitchell, and Cohen (2011) have shown that a soft reasoning procedure based on a combination of constrained, weighted, random walks through the KG can be used to reliably predict new beliefs for the KB. They describe a data-driven pathfinding method, while the original PRA algorithm generates paths by enumeration. To make PRA applicable to reason on large-scale KGs, they modify the path generation procedure in PRA to only generate paths that are potentially useful for the task. Specifically, they demand that a path is contained in the PRA model only if it retrieves at least a target entity in the training set, as well as being of length less than l, because a small number of possible relation paths are beneficial for inference. Finally, the weighted probability and score of all paths between two entities is a measure of the likelihood that a relation exists between two entities. Furthermore, Lao, Subramanya, Pereira, and Cohen (2012) have also shown that path-constrained random walk models can effectively predict new beliefs when taking advantage of combining a large-scale parsed text corpus and background knowledge. Experimental results show that the model can infer new beliefs with high accuracy by combining syntactic patterns in parsed text and semantic patterns in the background knowledge.

Although the PRA method has good interpretability, one main problem of random walk inference is the feature sparsity. To address this problem, Gardner, Talukdar, Krishnamurthy, and Mitchell (2014) incorporate vector similarity into random walk inference over KGs, to reduce the feature sparsity inherent using surface text. Namely, when following a series of edge types in a random walk, they permit the walk to follow edges that are semantically harmonious to the given edge types, as defined by some vector space embedding of the edge types. This combines notions of distributional similarity and symbolic logical inference, resulting in reducing the sparsity of the feature space constructed by PRA. On the one hand, reasoning on the whole knowledge graph is time-consuming, and inference is usually related to local information, so inference can be performed locally on the KG. On the other hand, global information is coarser in size and,

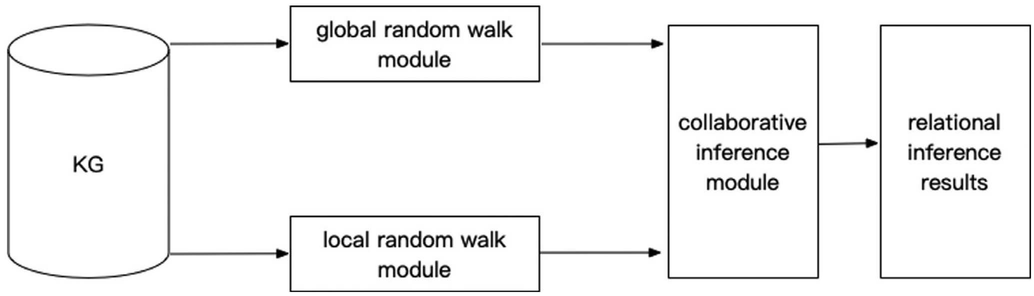

Fig. 1. Simple illustration of TRWA.

when combined with fine-grained locality information, can improve the accuracy of reasoning. Based on the above two reasons, Gardner and Mitchell (2015) define a simpler and more efficient algorithm called subgraph feature extraction (SFE). SFE does only the first step of PRA. They first perform a local search to characterize the graph around the entity node when some node pairs are given. Then, they run a set of feature extractors over these local subgraphs to obtain feature vectors for each entity pair. It greatly outperforms PRA, not only in time complexity but also in inference performance.

Liu, Han, Jiang, Liu, and Geng (2017b) study the two potential problems of the basic assumptions adopted by the existing random walk models. First, the algorithm extracts the relation path features through random sampling, which improves the computational efficiency while sacrificing the utilization of existing information in the KG. Second, using the supervised learning method to establish the relational inference model, the effectiveness of the model depends on the training data, especially those affected by data sparsity. Accordingly, the bidirectional semantics hypothesis and the inferential of relational-specific graph hypothesis were proposed, and the two-tier random walk algorithm (TRWA) was designed and implemented. The model is shown in Fig. 1. The main idea of TRWA is to combine two different feature modeling methods, subdivide the topological structure of the KG into global graph and local subgraph, and perform feature extraction separately. Finally, weighting and merging the global module and the local module to obtain complete logic rule inference algorithm.

A pure random walk without guidence has poor efficiency when finding useful formulas, and may even mislead inference due to introduced noise. Although some heuristic rules have been proposed to guide random walks, they still do not perform well because of the variety of formulas. To solve this problem, Wei, Zhao, and Liu (2016b) propose a novel goal-oriented inference algorithm that employs the specific inference target as the direction at each step in the process of random walk. Specifically, to accomplish such a goal-guided mechanism, the algorithm dynamically estimates the potentials for each neighbour at each step of random walk. Therefore, the algorithm is more inclined to traverse structures that are helpful in inferring the target and preventing transfer to noisy structures. Previous works on PRA usually neglect meaningful associations among certain relations, and cannot obtain enough training data for less frequent relations. Wang, Liu, Luo, Wang, and Lin (2016) propose a new multi-task learning framework for PRA, called coupled PRA (CPRA). CPRA performs inference using a multi-task mechanism. It consists of two modules: relation clustering and relation coupling. The former is used to discover highly correlated relations automatically, and the latter is used for coupling the learning of these relations. Through further coupling these relations, CPRA significantly outperforms PRA in terms of inference performance.

In general, the trend of knowledge reasoning based on logic rules is to abandon the manual rules gradually and then use pattern recognition to mine rules or features automatically for training models with machine learning methods. This type of model represents the knowledge graph as a complex heterogeneous network, so the reasoning tasks can be completed by the transfer probability, shortest path, and breadth-first search algorithms. However, this representation method has defects yet. First, the computational complexity of logic rule-based reasoning methods is still high, and their scalability is poor. Second, the nodes in the knowledge graph tend to obey the long-tailed distribution, that is to say, only a few entities and relations have a higher frequency of occurrence, and most of the entities and relations appear less frequently. Therefore, sparsity seriously affects the inference performance. In addition, how to handle multi-hop reasoning problem remains a greater challenge for logical models. Consequently, Lin et al. (2015a) and Das, Neelakantan, Belanger, and McCallum (2017) restrict the length of paths to 3-steps at most, so that it can reflect the logical connection between different objects. Therefore, scholars mainly focus on the reasoning methods based on distributed representation, which is not sensitive to data sparsity and is more expandable.

5. Knowledge reasoning based on distributed representation

Previous works to mine and discover unknown knowledge have relied on logic rules and random walk over the graph for lack of parallel corpora. Recently, embedding-based approaches have gained much attention in natural language processing. As is shown in Fig. 2, these models project the entities, relations, and attributes in the semantic network into continuous vector space to get distributed representation. Researchers have proposed a large number of reasoning methods based on distributed representation, including tensor decomposition, distance, and semantic matching models.

5.1. Knowledge reasoning based on tensor factorization

In the inference process, KG is often represented as a tensor and then is used for inferring unknown facts by tensor decomposition. Tensor decomposition is the process of decomposing high-dimensional arrays into multiple low-dimensional matrices. A three-way tensor $\mathcal{X}$ in which two nodes are identically formed by the concatenated entities of the domain and the third mode holds the relations is employed. A tensor entry $\mathcal{X}_{ijk} = 1$ represents that the fact (ith entity, kth predicate, jth entity) exists. If not, for unknown and unseen relations, the entry is set to zero. Then, the triplet score is calculated by the vector obtained through factorization, and the candidate with the high score is selected as the inference result.

The RESCAL model (Nickel, Tresp, & Kriegel, 2011) is a representative method of the tensor factorization model. Fig. 3 provides an illustration of this method. RESCAL decomposes high-dimensional and multi-relational data into a third-order tensor, which reduces

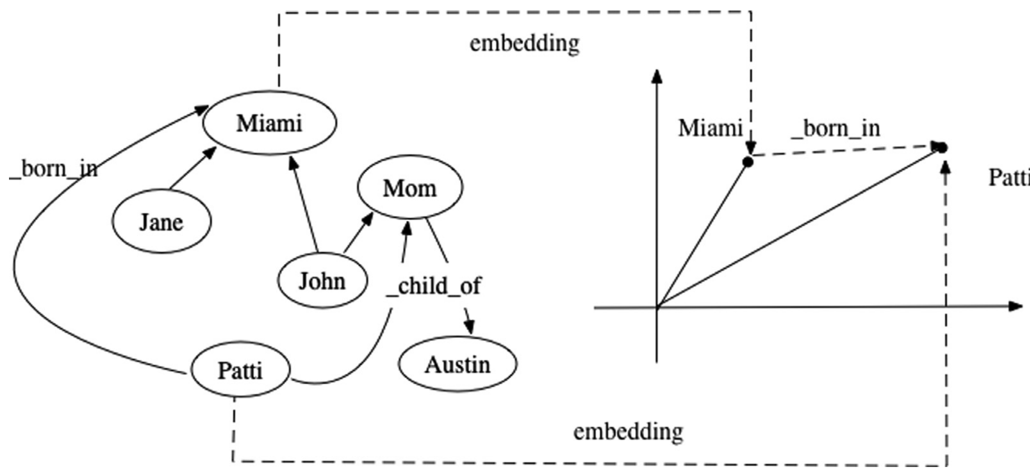

Fig. 2. Translations operating on the low-dimensional embeddings of the entities from knowledge graph.

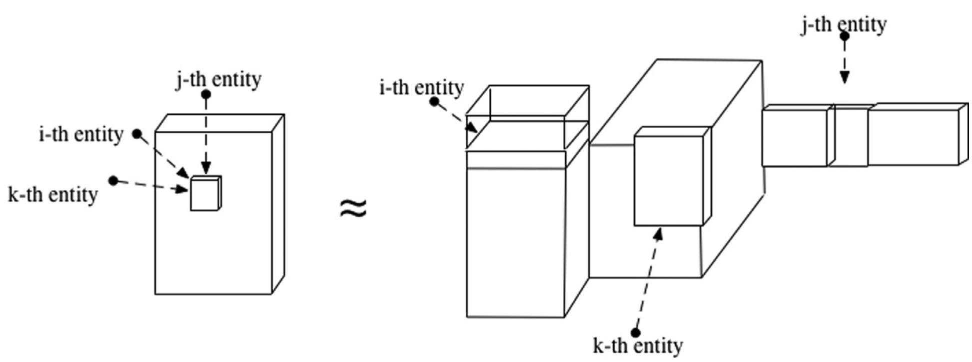

Fig. 3. Simple illustration of RESCAL.

the data dimension and retains the characteristics of the original data. It can be used for reasoning over knowledge graphs and achieve better results. Nickel, Tresp, and Kriegel (2012) demonstrate that tensor decomposition in the form of RESCAL factorization is a fit reasoning method for the binary relational data of the semantic web and demonstrate that the factorization is capable of successfully reasoning unseen triples on YAGO. Chang, Yih, Yang, and Meek (2014) propose a new knowledge inference model, TRESCAL, which is highly efficient and scalable. They promote the tensor decomposition model with two innovations. First, they remove the triples that do not satisfy the relation constraints from the loss. Second, they introduce a mathematical technique that significantly reduces the time computational complexity and space computational complexity. Nickel and Tresp (2013) also extend the RESCAL tensor factorization based on logistic regression. RESCAL-Logit uses different optimization strategies to improve inference accuracy. In Wu, Zhu, Liao, Zhang, and Lin (2017), PRESCAL based on paths of tensor factorization is proposed. It employs PRA to find all paths connecting the source and target nodes. Then, these paths are decomposed by tensor factorization for reasoning. Jainet al. (2017) develop a novel combination of matrix factorization (MF) and tensor factorization (TF) for knowledge base inference. It shows that the inference algorithm works robustly across diverse data, and model combination can gain better inference performance.

5.2. Knowledge reasoning based on distance model

TransE (Bordes, Usunier, Garcia-Duran, Weston, & Yakhnenko, 2013) is a commonly used embedding models and is the motivating base model. Since the time this model was proposed, a great deal of work has promoted it due to its simplicity and efficiency. Structured embedding (SE) method (Bordes, Weston, Collobert, & Bengio, 2011) is a simple version of TransE. SE uses two separate matrices to project head and tail entity for each relation and uses topology information of KG to model entities and relations. Since SE models relations with two separate matrices, there is a problem of poor coordination between entities. In addition, SE performs poorly on large-scale KGs. Therefore, Bordes et al. propose a more simplified model called TransE. The model is inspired by the results in Mikolov, Chen, Corrado, and Dean, in which the model learns distributed word representations such as $King - Man \approx Queen - Woman$. TransE model translates the potential feature representations by a relation-specific offset instead of transforming them through matrix multiplication. In particular, the score function of TransE is defined as:

$$f(h, r, t) = \|\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}\|_{l_1/l_2}$$

where $\|\cdot\|$ is the l_1 or l_2 norm of the difference vector. When reasoning is performed, the candidate entity or relation with a small score is the inference result.

Despite its simplicity and efficiency, TransE cannot deal with One-to-N, N-to-One, and N-to-N relations effectively. For example, given an N-to-One relation, e.g., *PresidentOf*, TransE might learn indistinguishable representations for *Trump* and *Obama*, who have both been president of the United States, although they are completely different entities. There are similar problems in N-to-One and N-to-N relations. To overcome the disadvantage of TransE in dealing with complex relations, a useful idea is to allow an entity to have different representations when involved in different relations.

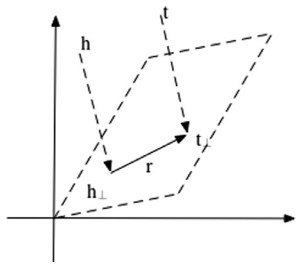

Fig. 4. Simple illustration of TransH (Wang et al., 2014c).

TransH (Wang, Zhang, Feng, & Chen, 2014c) follows this thought by introducing a relation-specific hyperplane. As is shown in Fig. 4, TransH projects entities on a hyperplane with $\boldsymbol{w_r}$ as the normal vector. Given a triple (*h, r, t*), the entity vectors $\boldsymbol{h}$ and $\boldsymbol{t}$ are projected onto the hyperplane, resulting in:

$$\boldsymbol{h_r} = \boldsymbol{h} - \boldsymbol{w_r^T h w}, \qquad \boldsymbol{t_r} = \boldsymbol{t} - \boldsymbol{w_r^T t w}$$

The projections are then translated by $\boldsymbol{r}$ on the hyperplane. If (*h, r, t*) holds, then $\boldsymbol{h_r} + \boldsymbol{r} \approx \boldsymbol{t_r}$. The scoring function is accordingly defined as:

$$f_r(h, r, t) = \|\boldsymbol{h_r} + \boldsymbol{r} - \boldsymbol{t_r}\|_{l_1/l_2}$$

TransH enables distinct representations of an entity in different relations by projecting an entity to a relation-specific hyperplane. Wen, Li, Mao, Chen, and Zhang (2016) present a canonical representation of KBs containing multi-fold relations. Using this representation, TransH is generalized to a new model, m-TransH. In addition, Fan, Zhou, Chang, and Zheng (2014) propose TransM which leverages the structure of the KG by pre-calculating the different weights for each training sample in respect of its relational mapping property. TransR (Lin, Liu, Sun, Liu, & Zhu, 2015b) shares a similar idea with TransH. But it introduces relation-specific spaces, instead of hyperplanes. In TransR, entities and relations are projected into different vector spaces. Fig. 5 shows a simple illustration of TransR. Given a triple (*h, r, t*), TransR first projects entity vectors $\boldsymbol{h}$ and $\boldsymbol{t}$ using the space-specific matrix $\boldsymbol{M_r}$, i.e., $\boldsymbol{h_r} = \boldsymbol{M_r h}$, $\boldsymbol{t_r} = \boldsymbol{M_r t}$, and then $\boldsymbol{h_r} + \boldsymbol{r} \approx \boldsymbol{t_r}$. The space-specific matrix makes entities close with each other if they hold the relation and keeps apart from those that do not hold the relation. TransR learns a unique vector for every relation, which may be under-representative to fit all entity pairs with this relation. Moreover, they extend TransR through clustering various entity pairs into groups and learning different relation embeddings for each group, called cluster-based TransR (CTransR). Previous methods, including TransE, TransH, and TransR, consider only direct links in knowledge graph reasoning. Lin et al. (2015a) think that multiple-hop paths also contain a large number of inference patterns between entities and propose path-based TransE (PTransE). In PTransE, path embeddings are obtained via composition of relation embeddings, and inference patterns are used to infer relations of entity pairs. RTransE (García-Durán, Bordes, & Usunier, 2015) also takes relation paths into account.

In methods such as TransE, TransH, and TransR, each relation has only one semantics, but in reality, *r* may have different meanings. As is shown in Fig. 6, for relation *location*, it not only represents the mountain-state relation but also represents the regional-country relation. Ji, He, Xu, Liu, and Zhao (2015) propose TransD which is based on a dynamic matrix to solve this problem. For a triplet (*h, r, t*), TransD utilizes two vectors to represent an entity or relation. The first one represents the meaning of entity, and the second one is used for constructing a mapping matrix. Therefore, mapping matrices are determined by both entities and relations. Compared with TransR/CTransR, TransD is less complicated and has no matrix-vector multiplication operations, which makes it train faster and can be applied on large-scale knowledge graphs. The heterogeneity (some relations connect many entities while others do not) and the imbalance (the number of head entities and that of tail entities in a relation may be different) of KGs are the two issues that affect inference performance. TranSparse (Ji, Liu, He, & Zhao, 2016), which consists of TranSparse (share) model and TranSparse (separate) model, is proposed as a way to address these problems. To overcome the heterogeneity, in TranSparse (share), the sparse degrees of transfer matrices are determined by the count of entity pairs connected by relations, and the two sides of relations share the same transfer matrices. To handle the issue of imbalance of relations, in TranSparse (separate), each relation has two separate sparse transfer matrixces, one for head and the other for tail. Xiao, Huang, Hao, and Zhu (2015) think that current translation-based methods suffer from the oversimplified loss metric, and treat each dimension identically. To address this issue, they propose TransA, an adaptive metric approach for embedding, which takes advantage of adaptive Mahalanobis metric and elliptical equipotential surfaces to provide a more flexible reasoning method. TransE utilizes inflexible Euclidean distance as a metric and has a limitation in dealing with complex relations. To solve these flaws simultaneously, Fang, Zhao, Tan, Yang, and Xiao (2018) extend TransA to TransAH, a revised translation-based method for knowledge graph inference. It replaces the Euclidean distance with weighted Euclidean distance by adding a diagonal weight matrix which assigns different weights to each feature dimension, and introduces the relation-oriented hyperplane. Finally, empirical experiments on large-scale knowledge graphs verify that TransAH is suitable for inference. Notably, Xiao, Huang, and Zhu (2016) propose a generative Bayesian non-parametric infinite mixture, called TransG, to address the issue of multiple relation semantics. Instead of assigning only one translation vector for one relation, they leverage a Gaussian distribution to handle multiple relation semantics by producing multiple translation components for a relation. Thus, TransG avoids mixing up semantic components of *r*, and different semantics are characterized by different components in TransG, which promotes the inference performance. The score function of TransG is defined as:

$$f(h, r, t) = \sum_{m=1}^{M_r} \pi_{r,m} e^{-\frac{\|v_h + v_{r,m} - v_t\|_2^2}{\sigma_h^2 + \sigma_t^2}}$$

Entities and relations may contain uncertainties that are often ignored in previous models. However, it is important to incorporate uncertainty information into knowledge reasoning because uncertainty can enhance precision of inference. Therefore, He, Liu, Ji, and Zhao (2015a) propose KG2E for modelling the certainty of entities and relations in the space of multi-dimensional Gaussian distributions. In KG2E, each entity or relation is represented by a Gaussian distribution, where the mean represents its position and the covariance denotes its certainty. Experimental results show that KG2E can effectively model the uncertainties of entities and relations in the process of knowledge inference. Chen, Chen, Shi, Sun, and Zaniolo (2019) propose a novel uncertain KG reasoning model UGKE, which preserves the uncertainty information. They also introduce probabilistic soft logic to infer confidence scores for triples out of KG during training.

Existing models solely learn from time-unknown fact triples but neglect the temporal information in the knowledge graph. However, facts in the knowledge base always change dynamically over time. Recently, a line of research has incorporated time information into the reasoning procedure. t-TransE (Jiang et al., 2016b) learns time-aware embedding by learning relation ordering jointly with TransE. They make an effort to impose temporal order on time-sensitive relations, e.g., *wasBornIn* → *workAt* → *diedIn*. To better model knowledge evolution, TAE-TransE (Jiang et al., 2016a)

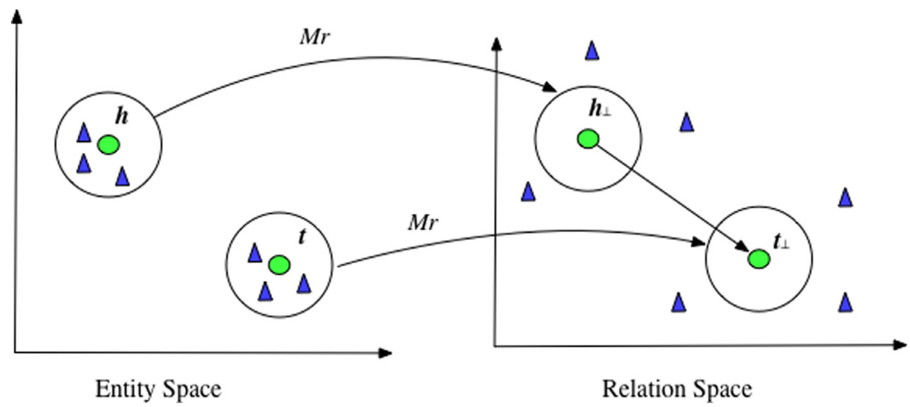

Fig. 5. Simple illustration of TransR (Lin et al., 2015b).

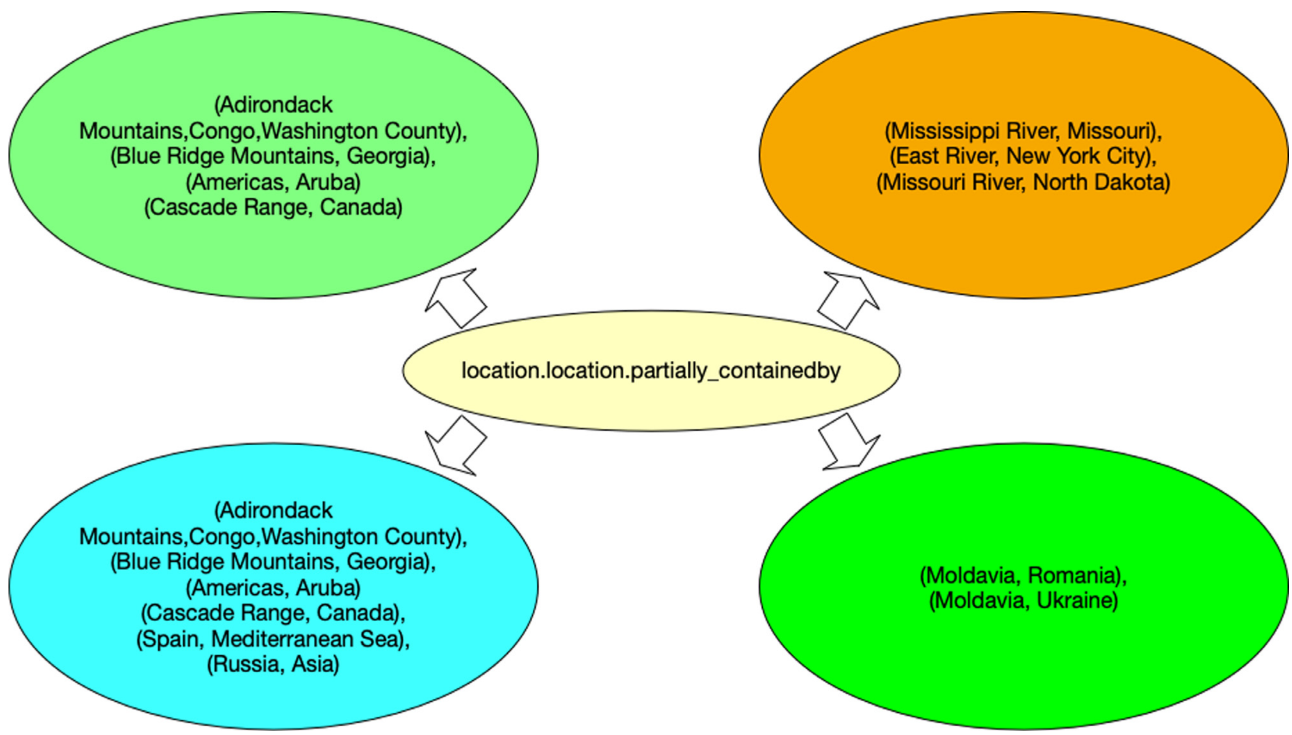

Fig. 6. Multiple types of entities of relation location (the relation *HasPart* has at least two latent semantics: composition related as (*Table, HasPart, Leg*) and location related as (*Atlantic, HasPart, NewYorkBay*)) (Ji et al., 2015) .

assumes that temporal ordering relations are relevant to each other and evolve dynamically. Know-Evolve (Trivedi, Dai, Wang, & Song, 2017) models the non-linear temporal evolution of knowledge components by using a bilinear embedding learning approach. They use a deep recurrent architecture to capture dynamical characteristics of the entities. Chekol, Pirrò, Schoenfisch, and Stuckenschmidt (2017) present an MLN-based approach for reasoning over uncertain temporal knowledge graphs. Leblay and Chekol (2018) try to use side information from the atemporal part of the graph for learning temporal embedding. HyTE (Dasgupta, Ray, & Talukdar, 2018) directly encodes time information to learn the temporally aware embedding.

5.3. Knowledge reasoning based on semantic matching model

SE uses two separate matrices to project head and tail entities for each relation r, which cannot effectively represent the semantic connection between entities and relations. Semantic matching energy (SME) (Bordes, Glorot, Weston, & Bengio, 2012; 2014) first represents entities and relations with vectors respectively, and then models correlations between entities and relations as semantic matching energy functions. SME defines a linear form for semantic matching energy functions and also a bilinear form. Latent factor model (Jenatton, Roux, Bordes, & Obozinski, 2012) captures various orders of interaction of the data using a bilinear structure. DistMult (Yang, Yih, He, Gao, & Deng, 2015) simplifies RESCAL by restricting M_r to be a diagonal matrix, which reduces the number of parameters and shows good reasoning ability and scalability in terms of validating unseen facts on the existing KB. Nickel, Rosasco, and Poggio (2016b) propose holographic embeddings (HoLE) to learn compositional vector space representation of knowledge graphs. HoLE applies circular correlation to generate compositional representations. Through using correlation as the compositional operator, HoLE can capture rich interactions but remains efficient to reason and easy to train at the same time. The major problem of current representation-based relational inference models is that they often ignore the semantical diversity of entities and relations, which will constrain the reasoning ability. Liu, Han,

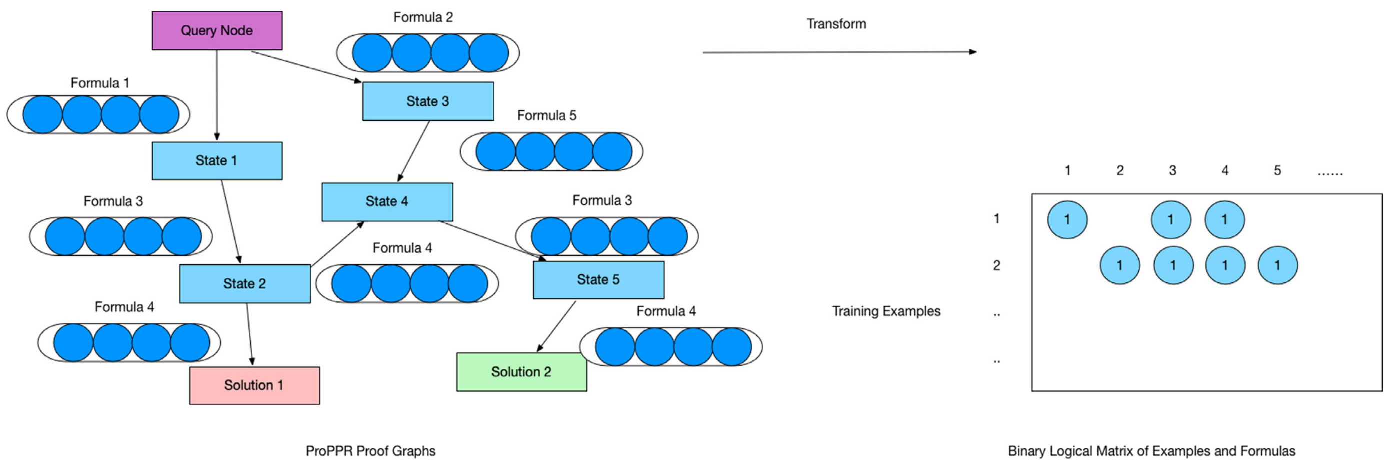

Fig. 7. The matrix factorization framework for learning first-order logic embeddings (Wang & Cohen, 2016).

Yang, Liu, and Wu (2017c) propose a new assumption for relation reasoning in knowledge graphs, which claims that each relations reflects the semantical connection of some specific attention aspects of the corresponding entities and could be modelled by selectively weighting on the constituent of the embeddings to help alleviate the semantic resolution problem. Accordingly, a semantic aspect-aware relation inference algorithm is proposed that can effectively improve the accuracy of relation inference on knowledge graphs.

Liu, Wu, and Yang (2017a) study the solutions of knowledge inference from the perspective of analogical inference. They formulate analogical structures and leverage them in a scoring function for optimizing the latent representations of entities and relations. In order to handle a large variety of dyadic relations, including symmetric and antisymmetric relations, Trouillon et al. (2017); Trouillon, Welbl, Riedel, Gaussier, and Bouchard (2016) propose ComplEx, based on complex embeddings. In ComplEx, each entity and relation is represented by a complex vector, and the scoring function is:

$$\begin{aligned}\varphi(r, s, o; \boldsymbol{\theta}) &= \mathrm{Re}\,(< w_r, e_s, \bar{e}_o >) \\ &= \mathrm{Re}\left(\sum_{k=1}^{K} w_{r,k} e_{sk} \bar{e}_{ok}\right) \\ &= < \mathrm{Re}\,(w_r), \mathrm{Re}\,(e_s), \mathrm{Re}\,(e_o) > \\ &\quad + < \mathrm{Re}\,(w_r), \mathrm{Im}\,(e_s), \mathrm{Im}\,(e_o) > \\ &\quad + < \mathrm{Im}\,(w_r), \mathrm{Re}\,(e_s), \mathrm{Im}\,(e_o) > \\ &\quad - < \mathrm{Im}\,(w_r), \mathrm{Im}\,(e_s), \mathrm{Re}\,(e_o) >\end{aligned}$$

where $w_r \in \mathcal{C}^K$ is a complex vector, $Re(x)$ means taking the real part of x and $Im(x)$ means taking the imaginary part of x. The score function represents the product of conjugate vector of $\boldsymbol{r}, \boldsymbol{s}, \boldsymbol{o}$, and then retains the real part of the final result. It uses only the Hermitian dot product, as it involves the conjugate-transpose of one of the two vectors. As a consequence, facts about antisymmetric relations can be handled well.

5.4. Knowledge reasoning based on multi-source information

Various auxiliary information, e.g., logical rules, textual descriptions and entity types can be combined to further enhance the performance. In this section, we discuss how such information can be integrated.

Logic rules can capture the rich semantics of natural language and support complex reasoning but often do worse in reasoning over large-scale knowledge graphs due to their dependence on logical background knowledge. In contrast, distributional representations are efficient and enable generalization. Therefore, injecting logic rules into embeddings for inference has received wide attention.

Rocktäschel, Bošnjak, Singh, and Riedel (2014) use first-order logic rule to guide entities and relations learning and then perform logic reasoning. Furthermore, they propose a paradigm for learning embeddings of entity pairs and relations that combine the strengths of matrix factorization and first-order logic domain knowledge (Rocktäschel, Singh, & Riedel, 2015). Two techniques, pre-factorization inference and joint optimization, for injecting logical background knowledge are presented. For pre-factorization inference, they first perform logical inference on the training data and infer facts as additional data. They propose a joint objective that rewards predictions that satisfy given logical knowledge, thus learning embeddings that do not require logical inference at test time. Demeester, Rocktäschel, and Riedel (2016a) present a highly efficient method based on matrix factorization for incorporating implication rules into distributed representations for KB inference. In the model, external commonsense knowledge is used for relation inference. Wang and Cohen (2016) propose a matrix factorization method to learn first-order logic embeddings. An overview of the framework is shown in Fig. 7. In detail, they first use ProPPR's structural gradient method (Wang, Mazaitis, & Cohen, 2014a) to generate a set of inference formulas from knowledge graphs. Then, they use this set of formulas, background graphs, and training examples to generate ProPPR proof graphs. To perform reasoning on the formulas, they map the training examples into the rows of a two-dimensional matrix, and inference formulas into the columns. Finally, these learned embeddings are transformed into parameters for the formulas, which makes first-order logic infer with learned formula embeddings. Guo, Wang, Wang, Wang, and Guo (2016) propose KALE, a novel method that learns entity and relation for reasoning by jointly modelling knowledge and logic. KALE consists of three key components: triple modelling module, rule modelling module, and joint learning module. For triple modelling, they follow TransE to model triples. To model rules, they use t-norm fuzzy logics (Hájek, 2013), which defines the truth value of a complex formula as a composition of the truth values of its constituents through specific t-norm based on logical connectives. After unifying triplets and rules as atomic and complex formulas, KALE minimizes a global loss to learn entity and relation embeddings. The larger the truth value is, the better the ground rules are satisfied. Embedding in this way can predict new facts that cannot even be directly inferred by pure logical inference. Recently, Ho, Stepanova, Gad-Elrab, Kharlamov, and Weikum (2018) have proposed an end-to-end rule learning system guided by external sources. It can learn high-quality rules with embedding support. pLogicNet (Qu & Tang, 2019) is proposed to combine existing rule-based methods and knowledge graph embedding methods. It models the distribution of all possible triplets with a Markov logic network, which is efficiently optimized with the variational EM algorithm. In the E-step, a knowledge graph embedding model is used to infer the hidden triplets, whereas in the M-step, the weights of rules are updated based on the observed and inferred triplets. Zhang et al. (2019b) propose IterE that learns embeddings and rules iteratively at the same time for knowledge graph reasoning.

Multi-source information like textual information and type information, considered as supplements for the structured information embedded in triples, is significant for inference in KGs. Wang, Zhang, Feng, and Chen (2014b) introduce a novel method of jointly embedding knowledge graphs and a text corpus so that entities and words or phrases are represented in the same vector space. Specifically, they define a coherent probabilistic TransE model (pTransE), which consists of three components: the knowledge model, the text model, and the alignment model. The knowledge model is used for fact modelling, and the alignment model guarantees that the embeddings of entities and words or phrases lie in the same space and impels two models to enhance each other. Experimental results show that the proposed method is very effective in reasoning new facts and capable of analogical reasoning. Furthermore, Wang and Li (2016) propose a new text-enhanced knowledge embedding (TEKE) method by making use of rich context information in a text corpus. The rich textual information is incorporated to expand the semantic structure of the knowledge graph to better support reasoning. In TEKE, they first annotate the entities in the corpus and construct a co-occurrence network composed of entities and words to bridge the knowledge graph and text information together. Based on the co-occurrence network, they define the textual contexts for entities and relations and incorporate the contexts into the knowledge graph structure. Finally, a normal translation-based optimization procedure is used for knowledge inference. Experiments on multiple datasets show that TEKE successfully solves the issue of structure sparseness that limits knowledge inference. He, Feng, Zou, and Zhao (2015b) integrate different knowledge graphs to infer new facts simultaneously. They present two improvements to the quality of reasoning over knowledge graphs. First, to reduce the data sparsity, they utilize the type consistency constraints between relations and entities to initialize negative data in the matrix. Second, they incorporate the similarity of relations between different knowledge bases into a matrix factorization model to make use of the complementarity of diverse knowledge bases. Xie, Liu, Jia, Luan, and Sun (2016) propose a novel method TKRL to take advantage of rich information located in hierarchical entity types. They use recursive hierarchical encoder and weighted hierarchical encoder to construct type-specific projection matrices for entities. Experimental results show that type information is significant in both predictive tasks. Tang, Chen, Cui, and Wei (2019) further propose a novel model named MKRL to predict potential triples, which integrate multi-source information, including entity descriptions, hierarchical types, and textual relations.

Generally, representation learning develops rapidly, and it has shown great potential in knowledge representation and reasoning over large-scale knowledge graphs. Knowledge representation learning can effectively solve the issue of data sparseness, and the efficiency in knowledge reasoning and semantic computing is higher than that of logic-based model. Based on TransE model, a number of improved knowledge graph inference methods have been proposed. However, the interpretability of these methods is poor (Xie, Ma, Dai, & Hovy, 2017). Specifically, the values of entity and relation vectors lack clear physical meaning. Therefore, there is still a long way for reasoning methods based on distributed representation to go.

6. Knowledge reasoning based on neural network

As an important machine learning algorithm, neural network basically imitates the human brain for perception and cognition. It has been widely used in the fields of natural language processing and has achieved remarkable results. The neural network has a strong ability to capture features. It can transform the feature distribution of input data from the original space into another feature space through nonlinear transformation and automatically learn the feature representation. Therefore, it is suitable for abstract tasks, such as knowledge reasoning.

Neural network has been used for knowledge graph inference for a long time (Nickel, Murphy, Tresp, & Gabrilovich, 2016a). In SE model, the parameters of the two entity vectors do not interact with each other. To alleviate the problems of the distance model, Socher, Chen, Manning, and Ng (2013) introduce a single layer model (SLM) which connects the entity vectors implicitly through the nonlinearity of a standard, single layer neural network. SLM can be used for reasoning relations between two entities. However, the non-linearity provides only a weak interaction between entity vectors. To this end, Socher et al. (2013) introduce an expressive neural tensor network (NTN) for reasoning that is illustrated in Fig. 8. The NTN model replaces a standard linear neural network layer with a bilinear tensor layer that directly relates the two entity vectors across multiple dimensions. NTN initializes the representation of each entity by averaging the word vectors, which results in improving performance. Chen, Socher, Manning, and Ng (2013a) improve NTN by initializing entity representations with word vectors learned in an unsupervised manner from text, and when doing this, existing relations can even be queried for entities that are unseen in the knowledge graphs. The increasing size of knowledge graph and complex feature space make the parameter size of reasoning methods extremely large. (Shi & Weninger, 2017b) present a shared variable neural network model called ProjE, and through a simple change in the architecture, achieves a smaller parameter size. Liu et al. (2016a) propose a new deep learning approach, called neural association model (NAM), for probabilistic reasoning in artificial intelligence. They investigate two NAM structures, namely, deep neural network (DNN) and relation-modulated neural network (RMNN). In the NAM framework, all symbolic events are represented in low-dimensional vector space to solve the problem of insufficient representation ability faced by existing methods. Experiments on several reasoning tasks have demonstrated that both DNN and RMNN can outperform conventional methods.

6.1. Knowledge reasoning based on convolutional neural networks

With the rise of deep learning, attempts are being made to introduce deep learning technology into the field of knowledge reasoning (Collobert et al., 2011). Xie et al. (2016) assert that most existing translation-based inference methods concentrate only on the structural information between entities, regardless of rich information encoded in entity description. For example, the phrase *Yao Ming is a famous basketball player in China* that contains the nationality information and occupational information of the entity *Yao Ming* simultaneously, and these multi-source heterogeneous information can be used for handling the problem of data sparsity effectively and enhancing the ability of distinguishing between entities and relation. Accordingly, they propose a novel method

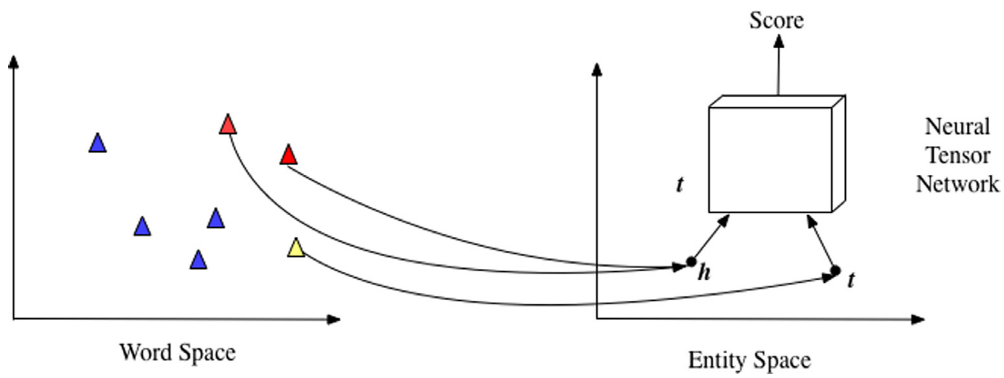

Fig. 8. Visualization of NTN (Socher et al., 2013).

for knowledge inference, named description-embodied knowledge representation learning (DKRL), which is able to make use of both fact triples and entity description. DKRL uses two encoders to represent semantics of entity descriptions, including a continuous bag-of-words (CBOW) model and a deep convolutional neural model, which can reduce the effect of data sparsity on the performance of inference models. DKRL also takes the zero-shot scenario into consideration, in which knowledge graphs contain some novel entities with only descriptions. It can learn representation for these novel entities automatically from their descriptions. The experiments in the zero-shot scenario show that the DKRL model can still achieve favourable results on the reasoning tasks. In order to infer new entities out of knowledge graph, (Shi & Weninger, 2017a) further propose a new open-world KGC task and introduce a model called ConMask to solve this task. It uses a relationship-dependent content masking to highlight words that are relevant to the task and then trains a fully convolutional neural network (FCN) for target fusion. Experiments on both open-world datasets and closed-world datasets show that ConMask can achieve good performance. However, these reasoning methods ignore the rich attribute information in the knowledge graph, such as age and gender can characterize entities in the knowledge graph. To this end, (Tay, Tuan, Phan, & Hui, 2017) first propose a novel multi-task neural network (MT-KGNN), which learns representations of entities, relations and attributes by encoding attribute information in the process of reasoning. MT-KGNN consists of RelNet and AttrNet. RelNet models the structure and relation of knowledge graph, while AttrNet models entities and corresponding properties. Notably, it is necessary to predefine relation and attribute. Otherwise, a large number of invalid calculations will occur and seriously affect the inference accuracy.

Annervaz, Chowdhury, and Dukkipati (2018) introduce a convolution-based model for knowledge inference. First, they use the DKRL encoding scheme, as it emphasizes the semantic description of the text. Afterward, entity and relation vectors are computed by the weighted sum with the attention mechanism. Experiments show significant improvement in performance on the natural language inference (SNLI) dataset. Dettmers, Minervini, Stenetorp, and Riedel (2018) propose ConvE, a multi-layer convolutional network model for knowledge inference, which can scale to large knowledge graphs. The architecture of ConvE is illustrated in Fig. 9. In ConvE, embedding representation of (s, r) pair is converted into a matrix and is regarded as a picture for extracting features with a convolution kernel. Unlike other inference methods, they use 1-N scoring to increase convergence speed. Ravishankar, Talukdar et al. (2017) observe that using a predefined scoring function, as in ConvE, might not perform well across all datasets. They define a simple neural network based score function ER-MLP-2d to fit different datasets. ER-MLP-2d, a variant of ER-MLP (Schlichtkrull et al., 2018), translates concatenated head and tail embeddings using the relation embedding of size 2d, which can achieve competitive performance on different datasets.

6.2. Knowledge reasoning based on recurrent neural network

Knowledge reasoning techniques that fuse relation paths and neural networks are also worth exploring. Neelakantan, Roth, and McCallum (2015) propose an approach composing the implications of a path using a recurrent neural network (RNN) called Path-RNN that reasons about conjunctions of multi-hop relations non-atomically. Path-RNN uses PRA to find distinct paths for each relation type and then takes embeddings of binary relation in the path as inputs vector. It outputs a vector in the semantic neighbourhood of the relation between the first and last entity of the path. For example, as shown in Fig. 10, after consuming the relation vectors along the path *Microsoft* → *Seattle* → *Washington* → *USA*, Path-RNN produces a vector semantically close to the relation *CountryofHeadquarters*. Shen, Huang, Chang, and Gao (2016) propose Implicit ReasoNets (IRNs) that learns to traverse knowledge graphs in vector space and infer missing triples. Rather than using human-designed relation paths in symbolic space and training a model separately, they propose to learn relation paths in vector space jointly with model training without using any additional information. Implicit ReasoNets also provides ways to understand the inference process. Das et al. (2017) note that the Path-RNN model has three defects: (1) It reasons about chains of relations, but not the entities that make up the nodes of the path. (2) It takes only a single path as evidence in predicting new predictions. (3) Path-RNN makes it impractical to be used in downstream tasks, since it requires training and maintaining a model for each relation type. Therefore, they present Single-Model which shares the relation type representation and the composition matrices of the recurrent neural network across all target relations, enabling the same training data to be represented by a reduced number of parameters. The Single-Model significantly increases the accuracy and practicality of RNN-based reasoning on Horn clause chains in large-scale KBs. Wang, Li, Zeng, and Chen (2018c) introduce an attention mechanism for the multi-hop reasoning problem. After finding reasoning paths between entities, they aggregate these paths' embeddings into one according to their attentions, and infer the relation based on the combined embedding.

Triples are not natural language. They model the complex structure with a fixed expression (h, r, t). Such short sequences may be under-representative to provide enough information for inference. Meanwhile, it is costly and difficult to construct useful long sequences from massive paths. It is inappropriate to treat them as the same type. To solve the above problems, Guo, Zhang, Ge, Hu, and Qu (2018) propose DSKG that employs respective multi-layer

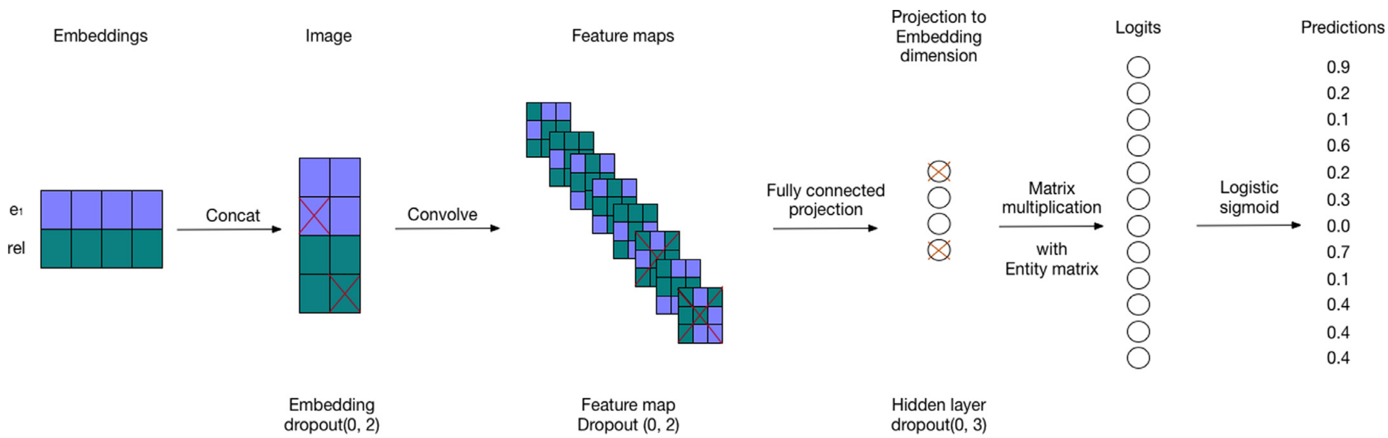

Fig. 9. Simple illustration of ConvE (Dettmers et al., 2018).

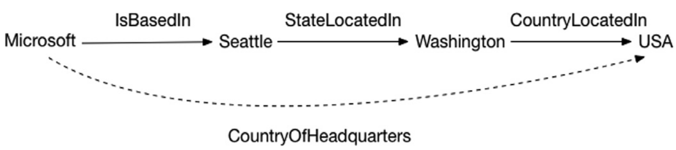

Fig. 10. Inference instance of Path-RNN.

RNN to process entities and relations. More specifically, DSKG uses independent RNN cells for the entity layer and the relation layer because this KG-specific architecture is capable of achieving better performance when relations are diverse and complex. Additionally, the DSGE model not only predicts entities but also has the ability to infer the triples.

6.3. Knowledge reasoning based on reinforcement learning

When using knowledge bases to promote question-answering system, we sometimes use only one of the triplets in the knowledge base to answer the question. However, when the question is complicated and the knowledge base is incomplete, it is necessary for the question-answering system to be able to infer unknown answer with existing triples. The process of inferring answers can be modelled as a serialized decision problem; thus, it can be solved with reinforcement learning. Xiong, Hoang, and Wang (2017) propose a novel reinforcement learning framework, DeepPath, for reasoning over a knowledge graph, which is the first to use reinforcement learning methods to solve multi-hop reasoning problems. In Deep-Path, the RL environment is modelled as a Markov decision process (MDP). A tuple $<S, A, P, R>$ is used for representing the MDP. DeepPath makes use of TransE and TransH to get representation vectors of the entities and relations. All the relations in the knowledge graph define the action space. The different embeddings of two entities define the state vector. Several factors contribute to the quality of the paths found by the RL agent. Thus, the reward function includes the following scoring criteria:

$$r_{\text{GLOBAL}} = \begin{cases} +1, & \text{if the path reaches } e_{\text{target}} \\ -1, & \text{otherwise} \end{cases}$$

$$r_{\text{ EFFICIENCY}} = \frac{1}{\text{length}}$$

$$r_{\text{ DIVERSITY}} = -\frac{1}{|F|}\sum_{i=1}^{F} \cos(P, P_i)$$

One main problem of reasoning methods based on reinforcement learning is that its action space is relatively large. Additionally, these methods cannot be applied to more complex tasks directly where the second entity is unknown and must be acquired by inferring. This problem also exists in path-based models (Neelakantan et al., 2015; Toutanova, Lin, Yih, Poon, & Quirk, 2016). Das et al. (2018) present a neural reinforcement learning approach, MINERVA, which learns how to guide the graph depending on the input query to find predictive paths. MINERVA avoids such modelling requirements. It represents the environment as a deterministic partially observed Markov decision process (POMDP) to reduce the action space. MINERVA encodes the history of decisions it has taken in the past using long short-term memory networks (LSTMs). Experiments show that MINERVA can learn long chains-of-reasoning. MINERVA treats relation selection and entity selection jointly as relation-entity pair selection via a single policy network, which underestimates the importance of entity selection when a 1-to-N or N-to-N relation appears during the path selection process. However, even if the relation path is correct, the different entities filled in the path may lead to different entities. To solve this problem, Li, Jin, Guan, Wang, and Cheng (2018) propose a multi-agent and reinforcement learning based method for path reasoning (MARLPaR). More specifically, they train two agents jointly, one for relation selection and another for entity selection. The relation selection agent is used to find common logistic paths for specific query relation. The entity selection agent is used to choose the most suitable entity from the tail entity set of the relations to find the candidate entity accurately. Godin, Kumar, and Mittal (2018) address the limitation of current approaches for reasoning over knowledge graphs that use reinforcement learning. Instead of simply returning a correct or incorrect answer, they allow the model to not answer a question by introducing a ternary reward structure in which a positive reward is given to a correct answer, a neutral reward for not answering a question, and a negative reward for an incorrect answer when the candidate entity cannot be reached in the boundary number of steps taken by an agent.

Although MINERVA has many advantages, it also has one major drawback: it assumes that there is an inference path while being agnostic to the scenario that no inference path exists. Deep-Path is trained to search more efficiently for paths between two entities while being agnostic to whether the entity pairs are positive or negative, whereas MINERVA learns to arrive at target nodes given an entity-query pair while agnostic to the quality of the found path. To address the previously mentioned problems, Chen, Xiong, Yan, and Wang (2018b) propose using variation inference to cope with complex reasoning. In order to increase the robustness of existing knowledge graph reasoning models and handle noisy environments, they combine "path-finding" and "path-reasoning" together as a whole from the perspective of the latent variable graph model. The graphic model regards the paths as discrete latent variables and relation as the observed variables with a given entity pair as the condition, thus, the path-finding module can be viewed as a prior distribution to predict the potential links in the knowledge graphs. In contrast, the path-reasoning module can be regarded as the likelihood distribution, which categorizes potential links into multiple classes. With this assumption, they introduce an approximate posterior and design a variational auto-encoder (Kingma & Welling, 2014) algorithm to maximize the evidence lower-bound. This variational framework unifies two modules into a unified framework and jointly train them. By active co-operations and interactions, the path finder can take the value of searched path into account and resort to more useful paths. Meanwhile, the path reasoner module can get more various paths from the path finder and generalizes better to unseen scenarios. Lin, Socher, and Xiong (2018) propose two modelling improvements for RL-based knowledge graph reasoning: reward shaping and action dropout. Reward shaping combines capability in modelling the semantics of triples with the symbolic reasoning capability of the path-based approach. Hard action dropout is more effective in encouraging the policy to sample various paths.

Reasoning methods based neural network attempt to use the powerful learning ability of neural network to represent the triples in knowledge graphs and thereby obtain better reasoning ability. However, model interpretability of neural network still exists in the area of knowledge graph reasoning, and how to explain the reasoning ability of neural network is worth studying. To date, there has been little research on reasoning methods based on neural networks. However, its powerful representation ability and outstanding performance showing in other fields promise broad prospects. In the future, it is worth exploring how to extend existing neural network methods to the filed of knowledge graph reasoning.

7. Application of knowledge graph reasoning

Knowledge graph reasoning methods infer unknown relations from existing triples, which not only provides efficient correlation discovery ability for resources in large-scale heterogeneous knowledge graphs but also completes knowledge graphs. Techniques such as consistency inference ensure the consistency and integrity of the knowledge graph. Inference techniques can perform domain knowledge reasoning through modelling domain knowledge and rules, which can support automatic decision making, data mining and link prediction. Due to the powerful intelligent reasoning ability, knowledge graphs can be widely used in many downstream tasks. In this section, we categorize these tasks into In-KG applications and Out-of-KG applications, described as follows.

7.1. In-KG applications

7.1.1. KG Completion

Constructing a large-scale knowledge graph requires constant updating relations between entities. However, despite their seemingly immense size, these knowledge bases are missing substantial amounts of information. For example, over 70% of people included in Freebase have no known place of birth, and 99% have no known ethnicity (West et al., 2014). One way to fill in missing facts in a knowledge base is to infer unknown facts based on existing triples, which is called knowledge graph completion, also known as link prediction (Liu, Sun, Lin, & Xie, 2016b).

Due to the noisy data source and the inaccuracy of the extraction process, noisy knowledge and knowledge contradictions phenomena in the knowledge graph also exist (Dong et al., 2014). A major problem of NELL is that the accuracy of the knowledge it acquires gradually decreases as it continues to operate. After the first month, NELL has an estimated precision of 0.9; after two months, precision has fallen to 0.71. The underlying reason is that the extraction patterns are not perfectly reliable, so false instances are sometimes extracted. The false instances will be used to extract increasing numbers of unreliable extraction patterns and false instances and finally dominate the knowledge base. NELL uses periodic human supervision to alleviate incorrect triples. However, human supervision is very expensive. Thus, knowledge graph reasoning methods are required to clean a noisy knowledge base automatically.

7.1.2. Entity classification

Entity classification aims to determine the categories (e.g., person, location) of a certain entity, e.g., *BarackObama* is a person, and *Hawaii* is a location. It can be treated as a special entity prediction task for the reason that the relation encoding entity types (denoted as IsA) is contained in the KG and has already been included into the embedding process. Thus, entity classification is obviously a KG completion problem.

7.2. Out-of-KG applications

7.2.1. Medical domain

At present, the medical domain has become a domain where knowledge graphs are actively used, and it is also a research focus in the artificial intelligence. When applied to medical knowledge graphs, knowledge reasoning methods can help doctors to collect health data, diagnose disease, and control errors (Yuan et al., 2018). For example, Kumar, Singh, and Sanyal (2009) propose a hybrid method based on case-based reasoning and rule-based reasoning to build a clinical decision support system for an intensive care unit (ICU). García-Crespo, Rodríguez, Mencke, Gómez-Berbís, and Colomo-Palacios (2010) design an ontology-driven differential diagnosis system (ODDIN), which is based on logical inference and probabilistic refinements. Martínez-Romero et al. (2013) build an ontology-based system for intelligent supervision and treatment of critical patients with acute cardiac disorders, where the expert's knowledge is represented by OWL ontology and a set of SWRL rules. On the basis of this knowledge, the inference engine executes the reasoning process and provides a recommendation about the patient's treatment for the doctor. Ruan, Sun, Wang, Fang, and Yin (2016) convert the data stored in traditional Chinese medicine knowledge graph into inference rules and then combine them with patient data for ancillary prescriptions inferred based on the knowledge graph.

Even for the same disease, the doctor may make different diagnoses according to the patient's condition because of the medical domain's dependence on subjective judgment. Thus, medical knowledge graphs must address a large amount of repetitive contradictory information, which increases the complexity of the medical reasoning model. Although traditional knowledge reasoning methods promote the automatic medical diagnosis process, they also have the defects of insufficient learning ability and low data utilization rates. In the face of increasing medical data, it is inevitable that some information will be missing and the diagnosis will be too time consuming. In order to solve the above problems, we need to explore and study efficient medical reasoning models.

7.2.2. Internet finance

Finance is also one active area in which knowledge graphs have been used. The investment relationship and the employment relationship in the knowledge graph can be used to identify stakeholder groups through a clustering algorithm. When some of the nodes have changed or large events occur, associations between changed entities can be inferred by path sorting and subgraph discovery methods. In the finance industry, anti-fraud is an important task. Through knowledge inference, people can verify the consistency of information to identify fraud in advance (Kapetanakis, Samakovitis, Gunasekera, & Petridis, 2012). In addition, knowledge inference also plays an important role in the field of securities investment (He, Ni, Cao, & Ma, 2016). For example, Ding, Zhang, Liu, and Duan (2016) propose a joint model to combine knowledge graph information and event embedding for stock prediction. However, this work doesn't capture the structural information in the text, and these information is very important for affecting stock to increase or decrease. Therefore, Liu, Zeng, Yang, and Carrio (2018) propose a joint learning model of tuple and texts using the TransE model and a convolution neural network to capture structured information in event tuple. The predictive results can support business decisions and improve investment planning.

Knowledge graph reasoning has improved the efficiency of resource allocation in the finance industry, strengthened the abilities of risk management and control, and effectively promoted the development of the financial industry. However, current data analysis and reasoning methods are difficult to meet the requirements of large-scale data analysis due to low standardization of finance industry data and its dispersion in multiple data systems. In response to this problem, external knowledge bases should be introduced to achieve reasoning over cross-domain large-scale knowledge graphs.

7.2.3. Intelligent question answering system

KB-based question and answering (KBQA) analyzes query question and then finds the answer from the knowledge base. However, KBQA also needs the support of reasoning techniques because the knowledge graph is incomplete. For example, Watson defeated humanity in Jeopardy, in which knowledge reasoning plays an important role. The questions of Jeopardy cover various areas and require candidates to analyze and reason entailment, irony and riddles. Intelligent question-answering systems, such as Apple's Siri, Microsoft's Cortana, and Amazon's Alexa, all require the support of knowledge graph inference.

The development of knowledge reasoning technology has laid a technical foundation for the development of intelligent question-answering systems. For example, Jain (2016) present factual memory network, which answers questions by extracting and reasoning over relevant facts from Freebase. It represents questions and triples in the same vector space, generates candidate facts, then finds out the answer using multi-hop reasoning. Zhang, Dai, Kozareva, Smola, and Song (2018a) propose an end-to-end variational reasoning network (VRN) for question answering with knowledge graph. VRN first recognizes the topic entity. Given the topic entity, the answer to the question can be retrieved through multi-hop reasoning on the knowledge graph. Narasimhan, Lazebnik, and Schwing (2018) ropose an algorithm based on graph convolution net (GCN) (Kipf & Welling, 2016) for reasoning in visual question answering. When answering questions, they combine the visualized situation with general knowledge encoded in the form of a knowledge base. However, there are still some problems to be solved in the intelligent question-answering systems. First, KBQA mainly focuses on single-fact questions. Specifically, answering the question requires only one triple in the KG. Meanwhile, for the complex problems that require multi-step reasoning, for example, when answering "What's the name of Yao Ming's wife's daughter?", KBQA performs poorly. Recently, Zhang, Dai, Toraman, and Song (2018b) imitate human brain to solve the problem. Second, current knowledge bases are composed of factual knowledge and lack of common sense. However, common sense plays an important role in the process of human brain reasoning, and common sense knowledge is difficult to standardize. Therefore, incorporating common-sense knowledge into KBQA for reasoning is a key issue in intelligent question answering.

7.2.4. Recommendation systems

The recommendation systems based on knowledge graph connect user and items, which can integrate multiple data sources to enrich semantic information. Implicit information can be obtained through reasoning techniques to improve recommendation accuracy. There are several typical cases for recommendation based on knowledge graph reasoning methods, such as shopping recommendation, movie recommendation and music recommendation. Wang et al. (2018a) propose knowledge-aware path recurrent network (KPRN), which not only generates representations for paths by accounting for both entities and relations but also performs reasoning based on paths to infer user preference. Unlike existing approaches that focus only on leveraging knowledge graphs for more accurate recommendation, Xian, Fu, Muthukrishnan, de Melo, and Zhang (2019) propose a policy-guided path reasoning (PGPR) method, which can reason over knowledge graph for recommendation with interpretation. PGPR is a flexible graph reasoning framework and can be extended to many other graph-based tasks such as product search and social recommendation.

With the help of reasoning techniques, it is possible to use multi-source heterogeneous data in recommendation systems. However, it is still in the initial development stage, and faces many challenges. In the future, how to solve the cold start issues and explicit reasoning over knowledge for recommendation systems are worth exploring.

7.2.5. Other applications

Knowledge reasoning techniques also play an important role in some other intelligent scenarios. For example, knowledge reasoning technology can be used to understand the user's query intent in search engines. In addition, it can be used for other computational linguistics tasks such as plagiarism detection, sentiment analysis, document categorization, spoken dialogue systems. Specifically, Franco-Salvador, Gupta, Rosso, and Banchs (2016a); Franco-Salvador, Rosso, and Montes-y Gómez (2016b) studied hybrid models that combine knowledge graph reasoning approach and continuous representation methods for the task of cross-language plagiarism detection. Cambria, Olsher, and Rajagopal (2014) show that how use SenticNet 3 and COGBASE to infer the polarity of a sentence. Franco-Salvador, Cruz, Troyano, and Rosso (2015) propose the use of meta-learning to combine and enrich current approaches by adding knowledge-based features obtained through inference to solve single and cross-domain polarity classification tasks. Franco-Salvador, Rosso, and Navigli (2014) leverage a multilingual knowledge graph, i.e., BabelNet, to obtain language-independent knowledge representation for documents to solve two tasks: comparable document retrieval and cross-language text categorization. Ma, Crook, Sarikaya, and Fosler-Lussier (2015) propose Inference Knowledge Graph to form part of a spoken dialogue system. Wang et al. (2018b) propose a graph reasoning model (GRM) to reason about the relationship of two persons from an image based on a social knowledge graph. As deep neural networks are widely used in natural language processing tasks, knowledge inference will usher in broader prospects.

8. Discussion and research opportunities

With the development of KG, knowledge graph reasoning has been widely explored and utilized in multiple knowledge-driven tasks, which significantly improves their performances. In this section, we first give a brief summary of these methods to identify the gap, and then propose research opportunities of knowledge graph reasoning.

Table 3
Summary of knowledge reasoning models.

Methods	Advantages	Disadvantages	Representative work	Applications
rule-based inference	capture hidden semantic information in KGs, improve the accuracy of knowledge reasoning significantly; simulate human reasoning ability, which makes it possible to incorporate priori knowledge to assist in reasoning	rules are not easy to obtain; rules with noise can mislead reasoning	MLNs FOIL PSL PRA	knowledge graph completion; diagnosis system; clinical decision support system
distributed representation-based inference	make full use of structural information existing in KGs; simple, easy to transfer to large-scale KGs	only consider the constraints that satisfy the KG facts, and the deeper compositional information is not considered, which limits the reasoning ability	RESCAL SE TransE TransH TransR TransG	knowledge graph completion; stock prediction; plagiarism
neural network-based inference	model triples directly, strong reasoning ability	high complexity and poor interpretability	SLM NTN	knowledge graph completion; question-answering system; recommendation system

8.1. Summary

In this paper, we provide a broad overview of currently available techniques, including rule-based reasoning methods, distributed representation-based reasoning methods, and neural network-based reasoning methods. We give a summary of advantages, disadvantages, representative works and applications of each type of models, which is shown in Table 3.

To sum up, there are differences and parallels between these three classes of reasoning methods, and they are complimentary in inference tasks. The relevance lies in the fact that all of them abstract the knowledge graph into topology and then use the topological relations between entities to model features and learn parameters. The main difference is that knowledge inference models based on neural network integrate CNN or RNN into the representation learning model or the logic rule model, extract features through the self-learning ability of deep learning model, and then utilize its memory reasoning ability to establish an entity relation prediction model. The representation learning model projects entities and relations into a low-dimensional vector space and performs reasoning based on semantic expression. The advantage is that the structural information in KG can be fully utilized when generating knowledge representation vectors. The disadvantage is that prior knowledge cannot be introduced to achieve inference when modelling. The logic rule model uses the abstract or concrete Horn clause for reasoning model, which is essentially rule-based reasoning. Its advantage is that it can simulate human logical reasoning behaviour, and introduce human prior knowledge to assist in reasoning. The disadvantage is that it has not solved a series of problems, including dependence on domain experts, high computational complexity, and poor generalization ability.

8.2. Research opportunities

Although existing models have already shown their powers in reasoning over KGs, there are still many possible improvements of them to be explored of. In this section, we will discuss the challenges of knowledge graph reasoning and give potential research opportunities.

8.2.1. Dynamical knowledge reasoning

Existing knowledge graph reasoning approaches mainly focus on static multi-relational data but neglect the useful time information contained in knowledge graphs. However, knowledge is not static and will evolve with time. We note that KG facts are not universally true, as they tend to be valid only in a specific time scope. For instance, (*BarackObama, PresidentOf, USA*) was true only from 2009 to 2016. Therefore, it is quite conceivable that taking temporal information into accountant during reasoning. Only a few works address this problem, but their efforts are still preliminary and reasoning methods for dynamical knowledge graph still need to be further explored.

8.2.2. Zero-shot reasoning

Existing knowledge graph reasoning models often require a large number of high-quality samples for training and learning, while it would consume considerable time and manpower. Recently, zero-shot learning has attracted much attention in many fields such as computer vision, natural language processing and so on. Zero-shot learning can learn from an unseen class or a class with only a few instances. In the reasoning process, the practical problem is that a large number of training samples cannot be obtained, resulting in many knowledge reasoning models being ineffective. It is natural that additional information such as text description and multi-modal information can help to deal with the zero-shot scenario. Besides, it's necessary to design a new framework which is more suitable for reasoning entities out of KGs.

8.2.3. Multi-source information reasoning

With the rapid development of mobile communication technology, people can upload and share multimedia contents including text, audio, images, and videos on the Web anytime. How to efficiently and effectively utilize these rich information is becoming a critical and challenging problem. And multi-source information has shown its potential to help reason over KGs while existing methods of utilizing such information are still preliminary. We could design more effective and elegant models to utilize these kinds of information better.

8.2.4. Multi-lingual knowledge graph reasoning

There are many KGs, such as Freebase, DBpedia have constructed multilingual versions by extracting structured information from Wikipedia. Multilingual KGs play important roles in many applications such as machine translation, cross-lingual plagiarism detection, and information extraction. However, to the best of our knowledge, only a few works have been done for reasoning over multilingual KGs. For example, (Abouenour, Nasri, Bouzoubaa, Kabbaj, & Rosso, 2014) construct an Arabic question-answering system to support semantic reasoning, and (Chen, Tian, Chang, Skiena, & Zaniolo, 2018a) present a cross-lingual inference method for KG completion based on French and German KG. Therefore, multilingual knowledge graph reasoning is also a significative but challenging work to be studied.

9. Conclusions

KG reasoning, which aims to infer new knowledge from existing triplets, has played an important role in many tasks and attracted much attention. In this paper, we give a broad overview of existing approaches with a particular focus on three types of reasoning methods, i.e., rule-based methods, distributed representation-based methods and neural network-based methods. Methods that conduct reasoning using logic rules were first introduced. We described the model details as well as advantages and disadvantages of such methods. After that, we discuss some more advanced approaches that perform KG reasoning with other information. The investigation on using reinforcement learning has just started and might receive increasing attention in the near future. Finally, we discuss the remaining challenges of knowledge graph reasoning and its application, and then give an oulook of the further study of knowledge graph reasoning. We hope that this review will provide new insights for further study of KG reasoning.

Declaration of Competing Interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from tjdxxiangyang@gmail.com.

Credit authorship contribution statement

Xiaojun Chen: Conceptualization, Formal analysis, Investigation, Methodology, Writing - original draft, Writing - review & editing. **Shengbin Jia:** Formal analysis, Supervision. **Yang Xiang:** Funding acquisition, Supervision.

Acknowledgements

We would like to thank all the reviewers for their insightful and valuable comments, which significantly improve the quality of this paper. This work is supported by the National Natural Science Foundation of China under Grant no. 71571136 and the Project of Science and Technology Commission of Shanghai Municipality under Grant no. 16JC1403000.

References

Abouenour, L., Nasri, M., Bouzoubaa, K., Kabbaj, A., & Rosso, P. (2014). Construction of an ontology for intelligent Arabic QAsystems leveraging the conceptual graphs representation. *Journal of Intelligent & Fuzzy Systems, 27*(6), 2869–2881.

Akbik, A., & Löser, A. (2012). Kraken: N-ary facts in open information extraction. In *Proceedings of the joint workshop on automatic knowledge base construction and web-scale knowledge extraction* (pp. 52–56). Association for Computational Linguistics.

Annervaz, K., Chowdhury, S. B. R., & Dukkipati, A. (2018). Learning beyond datasets: Knowledge graph augmented neural networks for natural language processing. In *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers)* (pp. 313–322).

Bach, S. H., Broecheler, M., Huang, B., & Getoor, L. (2017). Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research, 18*, 1–67.

Banko, M., Cafarella, M. J., Soderland, S., Broadhead, M., & Etzioni, O. (2007). Open information extraction from the web.. In *IJCAI: 7* (pp. 2670–2676).

Bienvenu, M., Bourgaux, C., & Goasdoué, F. (2019). Computing and explaining query answers over inconsistent dl-lite knowledge bases. *Journal of Artificial Intelligence Research, 64*, 563–644.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). 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). ACM.

Bordes, A., Glorot, X., Weston, J., & Bengio, Y. (2012). Joint learning of words and meaning representations for open-text semantic parsing. In *Artificial intelligence and statistics* (pp. 127–135).

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

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems* (pp. 2787–2795).

Bordes, A., Weston, J., Collobert, R., & Bengio, Y. (2011). Learning structured embeddings of knowledge bases. *Twenty-fifth AAAI conference on artificial intelligence*.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., & Rosati, R. (2006). Epistemic first-order queries over description logic knowledge bases. In *2006 international workshop on description logics dl'06* (p. 51). Citeseer.

Cambria, E., Olsher, D., & Rajagopal, D. (2014). Senticnet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. *Twenty-eighth aaai conference on artificial intelligence*.

Catherine, R., & Cohen, W. (2016). Personalized recommendations using knowledge graphs: A probabilistic logic programming approach. In *Proceedings of the 10th ACM conference on recommender systems* (pp. 325–332). ACM.

Chang, K.-W., Yih, W.-t., Yang, B., & Meek, C. (2014). Typed tensor decomposition of knowledge bases for relation extraction. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1568–1579).

Chekol, M. W., Pirrò, G., Schoenfisch, J., & Stuckenschmidt, H. (2017). Marrying uncertainty and time in knowledge graphs. *Thirty-first AAAI conference on artificial intelligence*.

Chen, D., Socher, R., Manning, C. D., & Ng, A. Y. (2013a). Learning new facts from knowledge bases with neural tensor networks and semantic word vectors. *International conference on learning representations (ICLR)*.

Chen, M., Tian, Y., Chang, K.-W., Skiena, S., & Zaniolo, C. (2018a). Co-training embeddings of knowledge graphs and entity descriptions for cross-lingual entity alignment. In *Proceedings of the 27th international joint conference on artificial intelligence* (pp. 3998–4004). AAAI Press.

Chen, W., Xiong, W., Yan, X., & Wang, W. Y. (2018b). Variational knowledge graph reasoning. In *Proceedings of NAACL-HLT* (pp. 1823–1832).

Chen, X., Chen, H., Zhang, N., Chen, J., & Wu, Z. (2013b). Owl reasoning over big biomedical data. In *2013 IEEE international conference on big data* (pp. 29–36). IEEE.

Chen, X., Chen, M., Shi, W., Sun, Y., & Zaniolo, C. (2019). Embedding uncertain knowledge graphs. In *Proceedings of the thirty-third AAAIconference on artificial intelligence (AAAI)*.

Chen, Y., Goldberg, S., Wang, D. Z., & Johri, S. S. (2016). Ontological pathfinding. In *Proceedings of the 2016 international conference on management of data* (pp. 835–846). ACM.

Chen, Y., & Wang, D. Z. (2014). Knowledge expansion over probabilistic knowledge bases. In *Proceedings of the 2014 ACM sigmod international conference on management of data* (pp. 649–660). ACM.

Cohen, W. W. (2016). Tensorlog: A differentiable deductive database. arXiv:1605.06523

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of Machine Learning Research, 12*(Aug), 2493–2537.

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

Das, R., Neelakantan, A., Belanger, D., & McCallum, A. (2017). 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* (pp. 132–141).

Dasgupta, S. S., Ray, S. N., & Talukdar, P. (2018). Hyte: Hyperplane-based temporally aware knowledge graph embedding. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 2001–2011).

Demeester, T., Rocktäschel, T., & Riedel, S. (2016a). Lifted rule injection for relation embeddings. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1389–1399).

Demeester, T., Rocktäschel, T., & Riedel, S. (2016b). Regularizing relation representations by first-order implications. In *Akbc2016, the workshop on automated base construction* (pp. 1–6).

Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. *Thirty-second AAAI conference on artificial intelligence*.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2016). Knowledge-driven event embedding for stock prediction. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 2133–2142).

Dong, X. L., Gabrilovich, E., Heitz, G., Horn, W., Murphy, K., Sun, S., et al. (2014). From data fusion to knowledge fusion. *Proceedings of the VLDB Endowment, 7*(10), 881–892.

Fader, A., Soderland, S., & Etzioni, O. (2011). Identifying relations for open information extraction. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 1535–1545). Association for Computational Linguistics.

Fan, M., Zhou, Q., Chang, E., & Zheng, T. F. (2014). Transition-based knowledge graph embedding with relational mapping properties. In *Proceedings of the 28th Pacific Asia conference on language, information and computing*.

Fang, Y., Zhao, X., Tan, Z., Yang, S., & Xiao, W. (2018). A revised translation-based method for knowledge graph representation. *Journal of Computer Research and Development, 55*(1), 139–150.

Franco-Salvador, M., Cruz, F. L., Troyano, J. A., & Rosso, P. (2015). Cross-domain polarity classification using a knowledge-enhanced meta-classifier. *Knowledge-Based Systems, 86*, 46–56.

Franco-Salvador, M., Gupta, P., Rosso, P., & Banchs, R. E. (2016a). Cross-language plagiarism detection over continuous-space-and knowledge graph-based representations of language. *Knowledge-based Systems, 111*, 87–99.

Franco-Salvador, M., Rosso, P., & Montes-y Gómez, M. (2016b). A systematic study of knowledge graph analysis for cross-language plagiarism detection. *Information Processing & Management, 52*(4), 550–570.

Franco-Salvador, M., Rosso, P., & Navigli, R. (2014). A knowledge-based representation for cross-language document retrieval and categorization. In *Proceedings of the 14th conference of the European chapter of the association for computational linguistics* (pp. 414–423).

Galárraga, L., Teflioudi, C., Hose, K., & Suchanek, F. M. (2015). Fast rule mining in ontological knowledge bases with AMIE+. *The VLDB Journal-The International Journal on Very Large Data Bases, 24*(6), 707–730.

Galárraga, L. A., Teflioudi, C., Hose, K., & Suchanek, F. (2013). Amie: Association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on world wide web* (pp. 413–422). ACM.

García-Crespo, Á., Rodríguez, A., Mencke, M., Gómez-Berbís, J. M., & Colomo-Palacios, R. (2010). Oddin: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements. *Expert Systems with Applications, 37*(3), 2621–2628.

García-Durán, A., Bordes, A., & Usunier, N. (2015). Composing relationships with translations. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 286–290).

Gardner, M., & Mitchell, T. (2015). Efficient and expressive knowledge base completion using subgraph feature extraction. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1488–1498).

Gardner, M., Talukdar, P., Krishnamurthy, J., & Mitchell, T. (2014). Incorporating vector space similarity in random walk inference over knowledge bases. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 397–406).

Godin, F., Kumar, A., & Mittal, A. (2018). Using ternary rewards to reason over knowledge graphs with deep reinforcement learning. *NIPS workshop on relational representation learning*.

Guo, L., Zhang, Q., Ge, W., Hu, W., & Qu, Y. (2018). DSKG: A deep sequential model for knowledge graph completion. In *China conference on knowledge graph and semantic computing* (pp. 65–77). Springer.

Guo, S., Wang, Q., Wang, L., Wang, B., & Guo, L. (2016). Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 192–202).

Hájek, P. (2013). *Metamathematics of fuzzy logic*: 4. Springer Science & Business Media.

Halaschek-Wiener, C., Parsia, B., Sirin, E., & Kalyanpur, A. (2006). Description logic reasoning for dynamic aboxes. In *2006 international workshop on description logics dl'06* (p. 200).

He, S., Liu, K., Ji, G., & Zhao, J. (2015a). Learning to represent knowledge graphs with gaussian embedding. In *Proceedings of the 24th ACM international on conference on information and knowledge management* (pp. 623–632). ACM.

He, W., Feng, Y., Zou, L., & Zhao, D. (2015b). Knowledge base completion using matrix factorization. In *Asia-pacific web conference* (pp. 256–267). Springer.

He, Y., Ni, L., Cao, L., & Ma, C. (2016). Research on case based reasoning system of stock theme events based on ontology. *Computer Technology and Development*, (1), 33–38.

Ho, V. T., Stepanova, D., Gad-Elrab, M. H., Kharlamov, E., & Weikum, G. (2018). Rule learning from knowledge graphs guided by embedding models. In *International semantic web conference* (pp. 72–90). Springer.

Jain, P., Murty, S., Chakrabarti, S. et al. (2017). Joint matrix-tensor factorization for knowledge base inference. arXiv:1706.00637

Jain, S. (2016). Question answering over knowledge base using factual memory networks. In *Proceedings of the NAACCL student research workshop* (pp. 109–115).

Jang, S., & Megawati, J. C. (2015). Semi-automatic quality assessment of linked data without requiring ontology.. In *Proceedings of the international semantic web conference (ISWC)* (pp. 45–55). Springer.

Jenatton, R., Roux, N. L., Bordes, A., & Obozinski, G. R. (2012). A latent factor model for highly multi-relational data. In *Advances in neural information processing systems* (pp. 3167–3175).

Ji, G., He, S., Xu, L., Liu, K., & Zhao, J. (2015). 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): 1* (pp. 687–696).

Ji, G., Liu, K., He, S., & Zhao, J. (2016). Knowledge graph completion with adaptive sparse transfer matrix. *Thirtieth AAAI conference on artificial intelligence*.

Jiang, S., Lowd, D., & Dou, D. (2012). Learning to refine an automatically extracted knowledge base using Markov logic. In *2012 IEEE 12th international conference on data mining* (pp. 912–917). IEEE.

Jiang, T., Liu, T., Ge, T., Sha, L., Chang, B., Li, S., et al. (2016a). Towards time-aware knowledge graph completion. In *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers* (pp. 1715–1724).

Jiang, T., Liu, T., Ge, T., Sha, L., Li, S., Chang, B., et al. (2016b). Encoding temporal information for time-aware link prediction. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 2350–2354).

Kapetanakis, S., Samakovitis, G., Gunasekera, P. B. D., & Petridis, M. (2012). *Monitoring financial transaction fraud with the use of case-based reasoning*.

Kertkeidkachorn, N., & Ichise, R. (2017). T2kg: An end-to-end system for creating knowledge graph from unstructured text. *Workshops at the thirty-first AAAI conference on artificial intelligence*.

Kimmig, A., Bach, S. H., Broecheler, M., Huang, B., & Getoor, L. (2012). A short introduction to probabilistic soft logic. In *Nips workshop on probabilistic programming: Foundations and applications: 1* (p. 3).

Kingma, D. P., & Welling, M. (2014). Auto-encoding variational Bayes. *Stat, 1050*, 1.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv:1609.02907

Kompridis, N. (2000). So we need something else for reason to mean. *International Journal of Philosophical Studies, 8*(3), 271–295.

Krötzsch, M., Marx, M., Ozaki, A., & Thost, V. (2018). Attributed description logics: Reasoning on knowledge graphs.. In *IJCAI* (pp. 5309–5313).

Kumar, K. A., Singh, Y., & Sanyal, S. (2009). Hybrid approach using case-based reasoning and rule-based reasoning for domain independent clinical decision support in ICU. *Expert Systems with Applications, 36*(1), 65–71.

Kuželka, O., & Davis, J. (2019). Markov logic networks for knowledge base completion: A theoretical analysis under the MCAR assumption. In *Proceedings of the thirty-fifth conference on uncertainty in artificial intelligence, UAI: 2019*.

Landwehr, N., Kersting, K., & Raedt, L. D. (2007). Integrating naive Bayes and foil. *Journal of Machine Learning Research, 8*(Mar), 481–507.

Landwehr, N., Passerini, A., De Raedt, L., & Frasconi, P. (2010). Fast learning of relational kernels. *Machine Learning, 78*(3), 305–342.

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

Lao, N., Mitchell, T., & Cohen, W. W. (2011). Random walk inference and learning in a large scale knowledge base. In *Proceedings of the conference on empirical methods in natural language processing* (pp. 529–539). Association for Computational Linguistics.

Lao, N., Subramanya, A., Pereira, F., & Cohen, W. W. (2012). Reading the web with learned syntactic-semantic inference rules. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning* (pp. 1017–1026). Association for Computational Linguistics.

Leblay, J., & Chekol, M. W. (2018). Deriving validity time in knowledge graph. In *Companion of the web conference 2018 on the web conference 2018* (pp. 1771–1776). International World Wide Web Conferences Steering Committee.

Lee, T.-W., Lewicki, M. S., Girolami, M., & Sejnowski, T. J. (1999). Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Signal Processing letters, 6*(4), 87–90.

Lenat, D. B., & Guha, R. V. (1989). *Building large knowledge-based systems; representation and inference in the Cyc project*. Addison-Wesley Longman Publishing Co., Inc..

Li, Y., Xu, B., Lu, J., & Kang, D. (2006). Discrete tableau algorithms for FSHI. In *Proceedings of the international workshop on description logics (dl)*.

Li, Z., Jin, X., Guan, S., Wang, Y., & Cheng, X. (2018). Path reasoning over knowledge graph: A multi-agent and reinforcement learning based method. In *2018 ieee international conference on data mining workshops (ICDMW)* (pp. 929–936). IEEE.

Lin, X. V., Socher, R., & Xiong, C. (2018). Multi-hop knowledge graph reasoning with reward shaping. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 3243–3253).

Lin, Y., Liu, Z., Luan, H., Sun, M., Rao, S., & Liu, S. (2015a). Modeling relation paths for representation learning of knowledge bases. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 705–714).

Lin, Y., Liu, Z., Sun, M., Liu, Y., & Zhu, X. (2015b). Learning entity and relation embeddings for knowledge graph completion. *Twenty-ninth AAAI conference on artificial intelligence*.

Liu, H., Wu, Y., & Yang, Y. (2017a). Analogical inference for multi-relational embeddings. In *Proceedings of the 34th international conference on machine learning-volume 70* (pp. 2168–2178). JMLR. org.

Liu, Q., Han, M., Jiang, L., Liu, Y., & Geng, J. (2017b). Two-tier random walk based relational inference algorithm. *Chineses Journal of Computers, 40*(6), 1275–1290.

Liu, Q., Han, M., Yang, X., Liu, Y., & Wu, Z. (2017c). Representation learning based relational inference algorithm with semantical aspect awareness. *Journal of Computer Reseach and Develoment, 54*(8), 1682–1692.

Liu, Q., Jiang, H., Evdokimov, A., Ling, Z.-H., Zhu, X., Wei, S., et al. (2016a). Probabilistic reasoning via deep learning: Neural association models. arXiv:1603.07704
Liu, Y., Zeng, Q., Yang, H., & Carrio, A. (2018). Stock price movement prediction from financial news with deep learning and knowledge graph embedding. In *Pacific rim knowledge acquisition workshop* (pp. 102–113). Springer.
Liu, Z., Sun, M., Lin, Y., & Xie, R. (2016b). Knowledge representation learning: A review. *Journal of Computer Reseach and Develoment, 53*(2), 247–261.
Ma, Y., Crook, P. A., Sarikaya, R., & Fosler-Lussier, E. (2015). Knowledge graph inference for spoken dialog systems. In *2015 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 5346–5350). IEEE.
Martínez-Romero, M., Vázquez-Naya, J. M., Pereira, J., Pereira, M., Pazos, A., & Baños, G. (2013). The iosc3 system: using ontologies and swrl rules for intelligent supervision and care of patients with acute cardiac disorders. *Computational and mathematical methods in medicine, 2013*.
Marx, M., Krötzsch, M., & Thost, V. (2017). Logic on mars: ontologies for generalised property graphs. In *Proceedings of the 26th international joint conference on artificial intelligence* (pp. 1188–1194). AAAI Press.
McCarthy, J. (1980). Circumscription-a form of non-monotonic reasoning. *Artificial intelligence, 13*(1–2), 27–39.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vector space. arXiv:1301.3781
Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM, 38*(11), 39–41.
Minsky, M. (1988). A framework for representing knowledge. *Readings in Cognitive Science, 20*(3), 156–189.
Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Betteridge, J., Carlson, A., et al. (2015). Never-ending learning. In *Proceedings of the twenty-ninth aaai conference on artificial intelligence* (pp. 2302–2310). AAAI Press.
Nakashole, N., Sozio, M., Suchanek, F. M., & Theobald, M. (2012). Query-time reasoning in uncertain RDF knowledge bases with soft and hard rules.. *VLDS, 884*, 15–20.
Narasimhan, M., Lazebnik, S., & Schwing, A. (2018). Out of the box: Reasoning with graph convolution nets for factual visual question answering. In *Advances in neural information processing systems* (pp. 2654–2665).
Neelakantan, A., Roth, B., & McCallum, A. (2015). Compositional vector space models for knowledge base inference. *2015 aaai spring symposium series*.
Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016a). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE, 104*(1), 11–33.
Nickel, M., Rosasco, L., & Poggio, T. (2016b). Holographic embeddings of knowledge graphs. *Thirtieth aaai conference on artificial intelligence*.
Nickel, M., & Tresp, V. (2013). Logistic tensor factorization for multi-relational data. In *Proceedings of the 30th international conference on machine learning-volume 28*. JMLR. org.
Nickel, M., Tresp, V., & Kriegel, H.-P. (2011). A three-way model for collective learning on multi-relational data.. In *Icml: 11* (pp. 809–816).
Nickel, M., Tresp, V., & Kriegel, H.-P. (2012). Factorizing yago: scalable machine learning for linked data. In *Proceedings of the 21st international conference on world wide web* (pp. 271–280). ACM.
Norenzayan, A., Smith, E. E., Kim, B. J., & Nisbett, R. E. (2002). Cultural preferences for formal versus intuitive reasoning. *Cognitive science, 26*(5), 653–684.
Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic web, 8*(3), 489–508.
Paulheim, H., & Bizer, C. (2014). Improving the quality of linked data using statistical distributions. *International Journal on Semantic Web and Information Systems (IJSWIS), 10*(2), 63–86.
Pellissier Tanon, T., Vrandečić, D., Schaffert, S., Steiner, T., & Pintscher, L. (2016). From freebase to wikidata: The great migration. In *Proceedings of the 25th international conference on world wide web* (pp. 1419–1428). International World Wide Web Conferences Steering Committee.
Pujara, J., Miao, H., Getoor, L., & Cohen, W. (2013a). Knowledge graph identification. In *International semantic web conference* (pp. 542–557). Springer.
Pujara, J., Miao, H., Getoor, L., & Cohen, W. W. (2013b). Ontology-aware partitioning for knowledge graph identification. In *Proceedings of the 2013 workshop on automated knowledge base construction* (pp. 19–24). ACM.
Qu, M., & Tang, J. (2019). Probabilistic logic neural networks for reasoning. arXiv:1906.08495
Ravishankar, S., Talukdar, P. P. et al. (2017). Revisiting simple neural networks for learning representations of knowledge graphs. arXiv:1711.05401
Richardson, M., & Domingos, P. (2006). Markov logic networks. *Machine learning, 62*(1–2), 107–136.
Rocktäschel, T., Bošnjak, M., Singh, S., & Riedel, S. (2014). Low-dimensional embeddings of logic. In *Proceedings of the acl 2014 workshop on semantic parsing* (pp. 45–49).
Rocktäschel, T., Singh, S., & Riedel, S. (2015). Injecting logical background knowledge into embeddings for relation extraction. In *Proceedings of the 2015 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 1119–1129).
Ruan, T., Sun, C., Wang, H., Fang, Z., & Yin, Y. (2016). Construction of traditional chinese medicine knowledge graph and its application. *Journal of Medical Informatics, 37*(4), 8–13.
Schlichtkrull, M., Kipf, T. N., Bloem, P., Van Den Berg, R., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In *European semantic web conference* (pp. 593–607). Springer.
Schoenmackers, S., Etzioni, O., Weld, D. S., & Davis, J. (2010). Learning first-order horn clauses from web text. In *Proceedings of the 2010 conference on empirical methods in natural language processing* (pp. 1088–1098). Association for Computational Linguistics.
Shen, Y., Huang, P.-S., Chang, M.-W., & Gao, J. (2016). Traversing knowledge graph in vector space without symbolic space guidance. arXiv:1611.04642
Shi, B., & Weninger, T. (2017a). Open-world knowledge graph completion. arXiv:1711.03438
Shi, B., & Weninger, T. (2017b). Proje: Embedding projection for knowledge graph completion. *Thirty-first aaai conference on artificial intelligence*.
Singhal, A. (2012). Introducing the knowledge graph: things, not strings. *Official Google Blog, 5*.
Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web, 5*(2), 51–53.
Socher, R., Chen, D., Manning, C. D., & Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems* (pp. 926–934).
Soma, R., & Prasanna, V. K. (2008). Parallel inferencing for owl knowledge bases. In *2008 37th international conference on parallel processing* (pp. 75–82). IEEE.
Stoilos, G., Stamou, G., Pan, J. Z., Tzouvaras, V., & Horrocks, I. (2007). Reasoning with very expressive fuzzy description logics. *Journal of Artificial Intelligence Research, 30*, 273–320.
Suchanek, F. M., Kasneci, G., & Weikum, G. (2007). Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on world wide web* (pp. 697–706). ACM.
Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). Yago: A large ontology from wikipedia and wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web, 6*(3), 203–217.
Tang, X., Chen, L., Cui, J., & Wei, B. (2019). Knowledge representation learning with entity descriptions, hierarchical types, and textual relations. *Information Processing & Management, 56*(3), 809–822.
Tari, L. (2013). Knowledge inference. *Encyclopedia of Systems Biology*, 1074–1078.
Tay, Y., Tuan, L. A., Phan, M. C., & Hui, S. C. (2017). Multi-task neural network for non-discrete attribute prediction in knowledge graphs. In *Proceedings of the 2017 acm on conference on information and knowledge management* (pp. 1029–1038). ACM.
Toutanova, K., Lin, V., Yih, W.-t., Poon, H., & Quirk, C. (2016). Compositional learning of embeddings for relation paths in knowledge base and text. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers): 1* (pp. 1434–1444).
Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *Proceedings of the 34th international conference on machine learning-volume 70* (pp. 3462–3471). JMLR. org.
Trouillon, T., Dance, C. R., Gaussier, É., Welbl, J., Riedel, S., & Bouchard, G. (2017). Knowledge graph completion via complex tensor factorization. *The Journal of Machine Learning Research, 18*(1), 4735–4772.
Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., & Bouchard, G. (2016). Complex embeddings for simple link prediction. In *International conference on machine learning* (pp. 2071–2080).
Wang, Q., Liu, J., Luo, Y., Wang, B., & Lin, C.-Y. (2016). Knowledge base completion via coupled path ranking. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers): 1* (pp. 1308–1318).
Wang, W. Y., & Cohen, W. W. (2016). Learning first-order logic embeddings via matrix factorization.. In *Ijcai* (pp. 2132–2138).
Wang, W. Y., Mazaitis, K., & Cohen, W. W. (2013). Programming with personalized pagerank: a locally groundable first-order probabilistic logic. In *Proceedings of the 22nd acm international conference on information & knowledge management* (pp. 2129–2138). ACM.
Wang, W. Y., Mazaitis, K., & Cohen, W. W. (2014a). Structure learning via parameter learning. In *Proceedings of the 23rd acm international conference on conference on information and knowledge management* (pp. 1199–1208). ACM.
Wang, W. Y., Mazaitis, K., Lao, N., & Cohen, W. W. (2015). Efficient inference and learning in a large knowledge base. *Machine Learning, 100*(1), 101–126.
Wang, X., Wang, D., Xu, C., He, X., Cao, Y., & Chua, T.-S. (2018a). Explainable reasoning over knowledge graphs for recommendation. arXiv:1811.04540
Wang, Z., Chen, T., Ren, J., Yu, W., Cheng, H., & Lin, L. (2018b). Deep reasoning with knowledge graph for social relationship understanding. In *Proceedings of the 27th international joint conference on artificial intelligence* (pp. 1021–1028). AAAI Press.
Wang, Z., & Li, J. (2015). Rdf2rules: learning rules from rdf knowledge bases by mining frequent predicate cycles. arXiv:1512.07734
Wang, Z., & Li, J.-Z. (2016). Text-enhanced representation learning for knowledge graph.. In *Ijcai* (pp. 1293–1299).
Wang, Z., Li, L., Zeng, D. D., & Chen, Y. (2018c). Attention-based multi-hop reasoning for knowledge graph. In *2018 ieee international conference on intelligence and security informatics (isi)* (pp. 211–213). IEEE.
Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014b). Knowledge graph and text jointly embedding. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 1591–1601).
Wang, Z., Zhang, J., Feng, J., & Chen, Z. (2014c). Knowledge graph embedding by translating on hyperplanes. *Twenty-eighth aaai conference on artificial intelligence*.
Wei, Y., Luo, J., & Xie, H. (2016a). Kgrl: an owl2 rl reasoning system for large scale knowledge graph. In *2016 12th international conference on semantics, knowledge and grids (skg)* (pp. 83–89). IEEE.

Wei, Z., Zhao, J., & Liu, K. (2016b). Mining inference formulas by goal-directed random walks. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1379–1388).

Wen, J., Li, J., Mao, Y., Chen, S., & Zhang, R. (2016). On the representation and embedding of knowledge bases beyond binary relations. In *Proceedings of the twenty-fifth international joint conference on artificial intelligence* (pp. 1300–1307). AAAI Press.

West, R., Gabrilovich, E., Murphy, K., Sun, S., Gupta, R., & Lin, D. (2014). Knowledge base completion via search-based question answering. In *Proceedings of the 23rd international conference on world wide web* (pp. 515–526). ACM.

Wu, F., Han, Y., Li, x., Zheng, Q., & Chen, X. (2018). Reasoning in artificial intelligence:advances and challenges. *Bulletin of National Natural Science Foundation of China, 32*(3), 262–265.

Wu, F., & Weld, D. S. (2010). Open information extraction using wikipedia. In *Proceedings of the 48th annual meeting of the association for computational linguistics* (pp. 118–127). Association for Computational Linguistics.

Wu, Y., Zhu, D., Liao, X., Zhang, D., & Lin, K. (2017). Knowledge graph reasoning based on paths of tensor factorization. *Pattern Recognition and Artificial Intelligence, 30*(5), 473–480.

Xian, Y., Fu, Z., Muthukrishnan, S., de Melo, G., & Zhang, Y. (2019). Reinforcement knowledge graph reasoning for explainable recommendation. arXiv:1906.05237

Xiao, H., Huang, M., Hao, Y., & Zhu, X. (2015). Transa: an adaptive approach for knowledge graph embedding. arXiv:1509.05490

Xiao, H., Huang, M., & Zhu, X. (2016). Transg: A generative model for knowledge graph embedding. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers): 1* (pp. 2316–2325).

Xie, Q., Ma, X., Dai, Z., & Hovy, E. (2017). An interpretable knowledge transfer model for knowledge base completion. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 950–962).

Xie, R., Liu, Z., Jia, J., Luan, H., & Sun, M. (2016). Representation learning of knowledge graphs with entity descriptions. *Thirtieth aaai conference on artificial intelligence*.

Xiong, W., Hoang, T., & Wang, W. Y. (2017). Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 564–573).

Xu, Z., Sheng, Y., He, L., & Wang, Y. (2016). Review on knowledge graph techniques. *Journal of University of Electronic Science and Technology of China, 45*(4), 589–606.

Yang, B., Yih, W.-t., He, X., Gao, J., & Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. *3rd international conference on learning representations, ICLR 2015, san diego, ca, usa, may 7–9, 2015, conference track proceedings*.

Yang, F., Yang, Z., & Cohen, W. W. (2017). Differentiable learning of logical rules for knowledge base reasoning. In *Advances in neural information processing systems* (pp. 2319–2328).

Yuan, K., Deng, Y., Chen, D., Zhang, B., Lei, K., & Shen, Y. (2018). Construction techniques and research development of medical knowledge graph. *Application Research of Computers, 35*(7), 1–12.

Zadeh, L. A. (1965). Fuzzy sets. *Information and control, 8*(3), 338–353.

Zhang, B., & Zhang, L. (1992). *Theory and applications of problem solving*: 9. North-Holland Amsterdam.

Zhang, N., Deng, S., Sun, Z., Wang, G., Chen, X., Zhang, W., et al. (2019a). 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).

Zhang, W., Paudel, B., Wang, L., Chen, J., Zhu, H., Zhang, W., et al. (2019b). Iteratively learning embeddings and rules for knowledge graph reasoning. In *The world wide web conference* (pp. 2366–2377). ACM.

Zhang, Y., Dai, H., Kozareva, Z., Smola, A. J., & Song, L. (2018a). Variational reasoning for question answering with knowledge graph. *Thirty-second aaai conference on artificial intelligence*.

Zhang, Y., Dai, H., Toraman, K., & Song, L. (2018b). Kg^2: Learning to reason science exam questions with contextual knowledge graph embeddings. arXiv:1805.12393

Zhou, J., Ma, L., Liu, Q., Zhang, L., Yu, Y., & Pan, Y. (2006). Minerva: A scalable owl ontology storage and inference system. In *Asian semantic web conference* (pp. 429–443). Springer.

Zou, Y., Finin, T., & Chen, H. (2004). F-owl: An inference engine for semantic web. In *International workshop on formal approaches to agent-based systems* (pp. 238–248). Springer.