

知识图谱研究综述

李涓子,侯磊*

(清华大学 计算机科学与技术系,北京 100084)

摘要:知识图谱以结构化的方式描述客观世界中概念、实体及其间的关系,将互联网的信息表达成更接近人类认知世界的形式,提供了一种更好地组织、管理和理解互联网海量信息的能力。知识图谱与大数据和深度学习一起,正在成为推动人工智能发展的核心驱动力。文章概述了当前已有的重要知识资源,对知识图谱关键技术——知识表示、知识图谱构建和知识图谱应用进行了综述,并对知识图谱未来发展的挑战和趋势进行了总结展望。

关键词:知识图谱;知识表示;知识获取;语义集成;知识应用

中图分类号:TP391

文献标志码:A

文章编号:0253-2395(2017)03-0454-06

Reviews on Knowledge Graph Research

LI Juanzi, HOU Lei*

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: Knowledge graph describes the real-world concepts, entities and their relationships in a structured way, expresses the Internet information in the form that is much closer to the human cognition, and provides a better ability to organize, manage and understand the mass information on the Internet. Knowledge graph, together with big data and deep learning, are becoming the core driving forces to promote the development of artificial intelligence. The paper outlines the existing important knowledge resources, reviews the key techniques about knowledge graph, namely, knowledge representation, construction and application, and finally summarizes the challenges and trends of the future development of knowledge graph.

Key words: knowledge graph; knowledge representation; knowledge acquisition; semantic integration; knowledge application

0 引言

知识图谱(Knowledge Graph, KG)旨在描述客观世界的概念、实体、事件及其间的关系。其中,概念是指人们在认识世界过程中形成的对客观事物的概念化表示,如人、动物、组织机构等;实体是客观世界中的具体事物,如篮球运动员姚明、互联网公司腾讯等;事件是客观世界的活动,如地震、买卖行为等。关系描述概念、实体、事件之间客观存在的关联,如毕业院校描述了个人与其所在院校的关系,运动员和篮球运动员之间概念和子概念的关系等。谷歌于2012年5月推出谷歌知识图谱,增强其搜索引擎的搜索结果,标志着大规模知识在互联网语义搜索中的成功应用。

收稿日期:2017-06-12;修回日期:2017-06-15

基金项目:国家重点基础研究发展计划(973计划 No. 2014CB340504);国家自然科学基金(No. 61533018);教育部在线教育研究基金(全通教育)重点课题(No. 2016ZD102)

作者简介:李涓子(1964—),女,教授,博导,主要研究方向为语义 Web,新闻与社会挖掘。

* 通信作者:侯磊(HOU Lei),E-mail:greener2009@gmail.com

知识图谱将互联网的信息表达成更接近人类认知世界的形式,提供了一种更好地组织、管理和理解互联网海量信息的能力。知识图谱给互联网语义搜索带来了活力,同时也在智能问答、大数据分析与决策中显示出强大威力,已经成为互联网基于知识的智能服务的基础设施。知识图谱与大数据和深度学习一起,成为推动人工智能发展的核心驱动力之一。

知识图谱技术是指在建立知识图谱中使用的技术,是融合认知计算、知识表示与推理、信息检索与抽取、自然语言处理与语义 Web、数据挖掘与机器学习等技术的交叉研究。知识图谱研究,一方面探索从互联网语言资源中获取知识的理论和方法;另一方面促进知识驱动的语言理解研究。随着大数据时代的到来,研究从大数据中挖掘隐含的知识理论与方法,将大数据转化为知识,增强对互联网资源的内容理解,将促进当代信息处理技术从信息服务向知识服务转变。具体地,知识图谱技术包括知识表示、知识图谱构建和知识图谱应用三方面的研究内容。

- 知识表示研究客观世界知识的建模,以方便机器识别和理解,既要考虑知识的表示与存储,又要考虑知识的使用和计算;

- 知识图谱构建解决如何建立计算机算法从客观世界或者互联网的各种数据资源中获取客观世界的知识,主要研究使用何种数据和方法抽取何种知识;

- 知识图谱应用主要研究如何利用知识图谱建立基于知识的智能服务系统,更好地解决实际应用问题。

1 现有知识图谱资源

知识图谱经历了由人工和群体智慧构建,到面向互联网利用机器学习和信息抽取技术自动获取的过程。根据信息来源和获取方式的不同,目前的知识图谱分为以下几类。

依靠人工构建的知识资源:早期知识资源通过人工添加和合作编辑获得,如英文 WordNet^[1]和 Cyc 项目^[2]以及中文的 HowNet。Cyc 是一个通用的世界知识库,始建于 1984 年,其目的是将上百万条知识编码为机器可处理形式,并在此基础上实现知识推理等人工智能相关任务,共包含了五十万实体,近三万个关系以及五百万事实。

基于群体智能的知识图谱:维基百科是至今利用群体智能建立的互联网上最大的知识资源,因此出现了很多使用维基百科构建知识库的项目,如 DBpedia^[3-4]、YAGO^[5-6]和 Freebase^[7]等。DBpedia 以构建本体的形式对知识条目进行组织。YAGO 融合 WordNet 良好的概念层次结构和维基百科中的大量实体数据。Freebase 是基于维基百科,使用群体智能方法建立的包含 6 800 万实体的结构化数据的知识图谱。清华大学和上海交通大学通过利用互动百科、百度百科,建立大规模知识图谱 XLORE^[8]和 Zhishi.me^[9]。XLORE 以英文维基百科为桥梁,通过跨语言链接技术,建立融合了四大中英文百科数据的跨语言知识库。

基于互联网链接数据构建的知识资源:国际万维网组织 W3C 于 2007 年发起的开放互联数据项目 (Linked Open Data, LOD),为实现网络环境下的知识发布、互联、共享和服务提供了创新技术,为智能搜索、知识问答和语义集成提供了创新源动力。Sean Bechhofer^[10]等人在科学领域自建了一个近似于 Linked Data 的语义数据资源,该资源包含更准确的学术用语,并能很好地反映研究者的影响力。

基于机器学习和信息抽取构建的知识图谱:从互联网数据自动获取知识是建立可持续发展知识图谱的发展趋势。这类知识图谱构建的特点是面向互联网的大规模、开放、异构环境,利用机器学习和信息抽取技术自动获取 Web 上的信息构建知识库。如华盛顿大学图灵中心的 KnowItAll^[11]和 TextRunner^[12]、卡内基梅隆大学的“永不停歇的语言学习者”(Never-Ending Language Learner, NELL)^[13]都是这种类型的知识库。

2 知识表示

知识表示技术可以分成符号主义和联结主义。符号主义的基础是纽威尔和西蒙提出的物理符号系统假设^[14],认为人类认知和思维的基本单元是符号,而认知过程就是在符号表示上的运算。联结主义认为,人的认知就是相互联系的具有一定活性值的神经元所形成网络的整体活动,知识信息不存在于特定的点,而是在神经网络的联结或者权重中。具体的表示方法可以分为三类。

2.1 基于符号逻辑的知识表示

基于符号逻辑进行知识表示和推理,主要包括逻辑表示法(如一阶逻辑、描述逻辑)、产生式表示法和框架表示等。逻辑表示与人类的自然语言比较接近,是最早使用的一种知识表示方法。基于符号逻辑的知识表示技术虽然可以很好地描述逻辑推理,然而机器生成推理规则的能力很弱,往往需要大量的人力,而且传统方法对数据的质量要求较高。因此,在目前大规模数据时代,基于符号逻辑的方法已经不能很好地解决知识表示的问题。

2.2 万维网内容的知识表示

Tim Berners-Lee 在其著作《Waving the Web》^[15]中提出了语义网(Semantic Web)的概念。在语义网中,网络内容均有确定的意义,而且可以很容易地被计算机理解、获取和集成。互联网信息的描述主要包括基于标签的半结构置标语言 XML、基于万维网资源语义元数据描述框架 RDF 和基于描述逻辑的本体描述语言 OWL 等。XML 通过为内容置标,便于数据交换;RDF 通过三元组(主体,谓词,客体)描述互联网资源之间的语义关系;OWL 构建在 RDF 之上,是具有更强表达及解释能力的语言。这些技术使我们可以将机器理解和处理的语义信息表示在万维网上,当前在工业界大规模应用的是基于 RDF 三元组的表示方法。

2.3 表示学习

表示学习的目标是通过机器学习或深度学习,将研究对象的语义信息表示为稠密低维的实值向量。对不同粒度知识单元进行隐式的向量化表示,以支持大数据环境下知识的快速计算,主要包括张量重构^[16-17]和势能函数^[18-21]的方法。张量重构综合整个知识库的信息,但在大数据环境下张量维度很高,重构的计算量较大。势能函数方法认为,关系是头实体到尾实体一种平移变换,Bordes 等人提出的 TransE 模型^[19]是平移模型的代表。之后有大量的工作对 TransE 进行扩展和应用,如通过优化向量化表示模型^[22]、结合文本等外部信息^[23-25]、应用逻辑推理规则^[26]等方法,提升表示学习效果,以表示更复杂的关系^[27-28]。相比传统方法,知识表示学习方法可以显著提升计算效率,有效缓解数据稀疏性,更容易实现不同来源的异质信息融合。

3 知识图谱构建技术

知识图谱中知识的来源有两类,一类是互联网上分布、异构的海量资源;一类是已有的结构化异构语义资源。从第一类资源中构建知识图谱的方法根据获取知识的类型分为概念层次学习、事实学习等,而针对第二类资源进行的工作是异构资源的语义集成。

3.1 概念层次学习

概念是人们理解客观世界的线索,是人们对客观世界中的事物在不同层次上的概念化描述,概念层次是知识图谱的“骨骼”。概念层次学习就是通过合理的技术,抽取知识表示中的概念,并确定其上下位关系。概念层次学习多采用基于启发式规则的方法,其基本思路是根据上下位概念的陈述模式,从大规模资源中找出可能具有上下位关系的概念对,并对上下位关系进行归纳。另一类是基于统计的概念层次学习方法^[29],假设相同概念出现的上下文也相似,利用词语或实体分布的相似性,通过定义计算特征学习概率模型来得到概念结构。

3.2 事实学习

知识图谱中事实以三元组的形式表示,事实数量决定了知识图谱的丰富程度。按照知识图谱构建时采用的机器学习方法可以分为有监督、半监督及无监督的知识图谱构建方法。

有监督的事实知识获取方法需要有已标注文档作为训练集,可以分为基于规则学习、基于分类标注和基于序列标注方法等。基于规则学习的语义标注方法从带语义标注的语料中自动学习标注规则,利用规则对数据资源进行语义标注,适合比较规范资源上的知识获取;基于分类的知识获取方法将知识获取转化为分类问题,根据确定的标注特征从标注语料中学习标注模型;基于序列模式标注的方法同时考虑多个语义标注之间的关系,可以提高标注的准确率。另外还包括其他(如考虑层次关系的)语义标注的方法等。

半监督的知识获取方法主要包括自扩展方法和弱监督方法。自扩展方法^[30-31]需要初始的种子实体对,根据这些种子实体对,发现新的语义模板,再对语料进行迭代抽取以发现新的实体对,其主要问题是语义漂移;弱监督方法使用知识库中的关系启发式地标注文本,其主要问题在于训练实例中本身带有大量噪音。近

年来,基于深度学习的知识获取越来越受到研究者的青睐,涌现出一批优秀的研究工作^[32-33]。

无监督的知识获取方法主要是开放信息抽取,使用自然语言处理方法,无须预先给定要抽取的关系类别,自动将自然语言句子转换为命题,这种方法在处理复杂句子时效果会受到影响。代表性的系统有之前提到的 KnowItAll,该系统具有领域无关特性,可以使用自扩展的方式从大规模互联网信息中抽取语义信息,同时会自动地对抽取信息进行评估。

3.3 语义集成

知识库间的异构性阻碍了知识在整个语义网上的共享。语义集成,就是在异构知识库之间,发现实体间的等价关系,从而实现知识共享。由于知识库多以本体的形式描述,因此语义集成中的主要环节是本体映射。主要方法包括:

- 基于文本的方法主要利用本体中实体的文本信息,例如实体的标签和摘要。通过计算两个实体字符串之间的相似度来确定实体之间是否具有匹配关系。
- 基于结构的方法主要利用本体的图结构信息对本体进行匹配。其中较为代表性的方法有 SimRank^[34]和相似度传播^[35],这些方法利用本体的图结构,对实体间的相似度进行传播,从而提高对齐的效果。
- 基于背景知识的方法一般使用 DBpedia 或 WordNet 等已有的大规模领域无关知识库作为背景知识来提高匹配效果。
- 基于机器学习的方法将本体匹配问题视为机器学习中的分类或优化问题,从而采取机器学习方法获得匹配结果。例如,Niepert 等人将本体匹配转换为马尔可夫逻辑网络问题,将本体中的各种信息转化为各种约束条件,并求出最优解^[36]。

4 知识图谱应用

Google 最初提出知识图谱是为了增强搜索结果,改善用户搜索体验,但知识图谱的应用远不止这些,基于知识图谱的服务和应用是当前的一大研究热点。按照应用方式可以分为语义搜索、知识问答,以及基于知识的大数据分析与决策等。

语义搜索:利用知识图谱所具有的良好定义的结构形式,以有向图的方式提供满足用户需求的结构化语义内容,主要包括 RDF 和 OWL 的语义搜索引擎和基于链接数据的搜索等^[37]。语义搜索利用建立大规模知识库对搜索关键词和文档内容进行语义标注,改善搜索结果^[38],如谷歌、百度和搜狗在搜索结果中嵌入知识图谱,包括实体的结构化信息和相关实体的描述。

知识问答:基于知识库的问答^[39-41]通过对问句的语义分析,将非结构化问句解析成结构化的查询,在已有结构化的知识库上获取答案。基于知识的问答依赖于语义解析器的性能,在面对大规模、开放域知识库时性能较差。近年来很多研究者开始研究基于深度学习的知识库问答方法^[42-46],这类方法更具鲁棒性。

知识驱动的大数据分析与决策:利用知识图谱可以辅助行业和领域的大数据分析和决策。美国 Netflix 公司利用基于其订阅用户的注册信息和观看行为构建的知识图谱,分析了解到用户很喜欢 Fincher, Spacey 主演的作品表现都不错,以及英剧版的《纸牌屋》很受欢迎,因此决定拍摄了美剧《纸牌屋》,在美国及 40 多个国家成为热门的在线剧集。

5 结束语

知识图谱技术是人工智能知识表示和知识库在互联网环境下的大规模应用,显示出知识在智能系统中的重要性,是实现智能系统的基础知识资源。纵观知识图谱研究发展的相关研究现状,以下研究将成为未来知识图谱必须应对的挑战:1)研究知识表示和获取的新理论和方法,使知识既具有显式的语义定义,又便于大数据下的知识计算;2)随着信息技术从信息服务向知识服务的转变,研究建立知识图谱构建的平台,以服务不同的行业和应用;3)知识图谱虽然已经在语义搜索和知识问答等应用中展示出一定的威力,但是基于知识图谱的应用研究远不止这些,如何进一步推进知识驱动的智能信息处理应用是十分有价值的研究。

参考文献:

- [1] Fellbaum C. WordNet[M]. Hoboken:Blackwell Publishing Ltd,1998.

- [2] Lenat D B. CYC: A Large-scale Investment in Knowledge Infrastructure[J]. *Communications of the ACM*, 1995, **38**(11): 33-38. DOI:10. 1145/219717. 219745.
- [3] Bizer C, Lehmann J, Kobilarov G, *et al.* DBpedia-A Crystallization Point for the Web of Data[J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2009, **7**(3): 154-165. DOI:10. 1016/j. websem. 2009. 07. 002.
- [4] Auer S, Bizer C, Kobilarov G, *et al.* DBpedia: A Nucleus for a Web of Open Data[J]. *The Semantic Web*, 2007, **4825**: 722-735. DOI:10. 1007/978-3-540-76298-0_ 52.
- [5] Suchanek F M, Kasneci G, Weikum G. Yago: A Core of Semantic Knowledge[C]// Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 697-706. DOI:10. 1145/1242572. 1242667.
- [6] Suchanek F M, Kasneci G, Weikum G. Yago: A Large Ontology from Wikipedia and Wordnet[J]. *Web Semantics: Science, Services and Agents on the World Wide Web*, 2008, **6**(3): 203-217. DOI:10. 1016/j. websem. 2008. 06. 001.
- [7] Bollacker K, Evans C, Paritosh P, *et al.* Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge[C]// Proceedings of the 2008 ACM SIGMOD international conference on Management of data. ACM, 2008: 1247-1250. DOI:10. 1145/1376616. 1376746.
- [8] Wang Z, Li J, Wang Z, *et al.* Xlore: A Large-scale English-Chinese Bilingual Knowledge Graph[C]// Proceedings of the 2013th International Conference on Posters & Demonstrations Track-Volume 1035. 2013: 121-124.
- [9] Niu X, Sun X, Wang H, *et al.* Zhishi, me-weaving Chinese Linking Open Data[C]// International Semantic Web Conference. Springer Berlin Heidelberg, 2011, **7032**: 205-220. DOI:10. 1007/978-3-642-25093-4_ 14.
- [10] Bechhofer S, Buchan I, De Roure D, *et al.* Why Linked Data is Not Enough for Scientists[J]. *Future Generation Computer Systems*, 2013, **29**(2): 599-611. DOI:10. 1016/j. future. 2011. 08. 004.
- [11] Etzioni O, Cafarella M, Downey D, *et al.* Web-scale Information Extraction in Knowitall: (Preliminary Results)[C]// Proceedings of the 13th international conference on World Wide Web. ACM, 2004: 100-110. DOI: 10. 1145/988672. 988687.
- [12] Yates A, Cafarella M, Banko M, *et al.* Textrunner: Open Information Extraction on the Web[C]// Proceedings of Human Language Technologies: The Annual Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Association for Computational Linguistics, 2007: 25-26.
- [13] Carlson A, Betteridge J, Kisiel B, *et al.* Toward an Architecture for Never-Ending Language Learning[C]// Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence. 2010, **5**: 1306-1313.
- [14] Newell A, Simon H A. Computer Science as Empirical Inquiry: Symbols and Search[J]. *Communications of the ACM*, 1976, **19**(3): 113-126. DOI:10. 1145/360018. 360022.
- [15] Berners-Lee T, Fischetti M, Foreword By-Dertouzos M L. Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor[M]. San Francisco: HarperInformation, 1999.
- [16] Nickel M, Tresp V, Kriegel H P. A Three-way Model for Collective Learning on Multi-relational Data[C]// Proceedings of the 28th international conference on machine learning. 2011: 809-816.
- [17] Nickel M, Tresp V, Kriegel H P. Factorizing Yago: Scalable Machine Learning for Linked Data[C]// Proceedings of the 21st International Conference on World Wide Web. ACM, 2012: 271-280. DOI:10. 1145/2187836. 2187874.
- [18] Bordes A, Weston J, Collobert R, *et al.* Learning Structured Embeddings of Knowledge Bases[C]// Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011: 301-306.
- [19] Bordes A, Usunier N, Garcia-Duran A, *et al.* Translating Embeddings for Modeling Multi-relational Data[C]// Advances in Neural Information Processing Systems. 2013: 2787-2795.
- [20] Bordes A, Glorot X, Weston J, *et al.* A Semantic Matching Energy Function for Learning with Multi-relational Data[J]. *Machine Learning*, 2014, **94**(2): 233-259. DOI:10. 1007/s10994-013-5363-6.
- [21] Socher R, Chen D, Manning C D, *et al.* Reasoning with Neural Tensor Networks for Knowledge Base Completion[C]// Advances in Neural Information Processing Systems, 2013: 926-934.
- [22] Lin Y, Liu Z, Sun M, *et al.* Learning Entity and Relation Embeddings for Knowledge Graph Completion[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015: 2181-2187.
- [23] Wang Zhen, Zhang Jianwen, Feng Jianlin, *et al.* Knowledge Graph and Text Jointly Embedding[C]// Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing, 2014: 1591-1601.
- [24] Wang Z, Li J Z. Text-Enhanced Representation Learning for Knowledge Graph[C]// Proceeding of the 25th International Joint Conference on Artificial Intelligence, 2016: 1293-1299.
- [25] Xie R, Liu Z, Jia J, *et al.* Representation Learning of Knowledge Graphs with Entity Descriptions[C]// Proceedings of the

- 30th AAAI Conference on Artificial Intelligence, 2016:2659-2665.
- [26] Lin Y, Liu Z, Luan H, *et al.* Modeling Relation Paths for Representation Learning of Knowledge Bases[C]//Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing, 2015:705-714.
- [27] Xie R, Liu Z, Sun M. Representation Learning of Knowledge Graphs with Hierarchical Types[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016:2965-2971.
- [28] Lin Y, Liu Z, Sun M. Knowledge Representation Learning with Entities, Attributes and Relations[C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, 2016:2866-2872.
- [29] Wu W, Li H, Wang H, *et al.* Probbase: A Probabilistic Taxonomy for Text Understanding[C]//Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data. ACM, 2012: 481-492. DOI: 10.1145/2213836.2213891.
- [30] Carlson A, Betteridge J, Wang R C, *et al.* Coupled Semi-supervised Learning for Information Extraction[C]//Proceedings of the third ACM international conference on Web search and data mining. ACM, 2010: 101-110. DOI: 10.1145/1718487.1718501.
- [31] Shi B, Zhang Z, Sun L, *et al.* A Probabilistic Co-Bootstrapping Method for Entity Set Expansion[C]//Proceedings of the 25th International Conference on Computational Linguistics, 2014:2280-2290.
- [32] Lin Y, Shen S, Liu Z, *et al.* Neural Relation Extraction with Selective Attention Over Instances[C]//Proceedings of the 54th Annual Meeting of Association for Computational Linguistics, 2016:2124-2133.
- [33] Ji G, Liu K, He S, *et al.* Distant Supervision for Relation Extraction with Sentence-Level Attention and Entity Descriptions[C]//Proceeding of the Thirty-First AAAI Conference on Artificial Intelligence, 2017:3060-3066.
- [34] Jeh G, Widom J. SimRank: A Measure of Structural-context Similarity[C]//Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2002: 538-543. DOI: 10.1145/775047.775126.
- [35] Melnik S, Garcia-Molina H, Rahm E. Similarity Flooding: A Versatile Graph Matching Algorithm and Its Application to Schema Matching[C]//Proceedings of the 18th International Conference on Data Engineering. IEEE, 2002:117-128.
- [36] Niepert M, Meilicke C, Stuckenschmidt H. A Probabilistic-logical Framework for Ontology Matching[C]//Proceedings of the Twenty-fourth AAAI Conference on Artificial Intelligence, 2010:1413-1418.
- [37] Qu Y, Cheng G. Falcons Concept Search: A Practical Search Engine for Web Ontologies[J]. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2011, **41**(4): 810-816.
- [38] Hua W, Wang Z, Wang H, *et al.* Understand Short Texts by Harvesting and Analyzing Semantic Knowledge[J]. *IEEE transactions on Knowledge and data Engineering*, 2017, **29**(3): 499-512.
- [39] Bron M, Balog K, De Rijke M. Example Based Entity Search in the Web of Data[C]//European Conference on Information Retrieval. Springer Berlin Heidelberg, 2013:392-403. DOI: 10.1007/978-3-642-36973-5_33.
- [40] Cai Q, Yates A. Large-scale Semantic Parsing via Schema Matching and Lexicon Extension[C]//Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, 2013:423-433.
- [41] Zettlemoyer L S, Collins M. Learning to Map Sentences to Logical form: Structured Classification with Probabilistic Categorical Grammars[C]//Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence, 2015.
- [42] Bordes A, Weston J, Usunier N. Open Question Answering with Weakly Supervised Embedding Models[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2014:165-180. DOI:10.1007/978-3-662-44848-9_11.
- [43] Bordes A, Chopra S, Weston J. Question Answering with Subgraph Embeddings[C]//Proceedings of the 2014 Conference on Empirical Methods on Natural Language Processing. 2014:615-620.
- [44] Dong L, Wei F, Zhou M, *et al.* Question Answering over Freebase with Multi-Column Convolutional Neural Networks [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015, **1**:260-269.
- [45] Liu K, Zhao J, He S, *et al.* Question Answering over Knowledge Bases[J]. *IEEE Intelligent Systems*, 2015, **30**(5):26-35.
- [46] Zhang Y, He S, Liu K, *et al.* A Joint Model for Question Answering over Multiple Knowledge Bases[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016: 3094-3100.