

时序知识图谱研究综述

石津妮

(大连海事大学 信息科学技术学院, 辽宁省大连市 中国 116026)

摘 要 知识图谱是以图的形式表现客观世界中的概念和实体及其之间关系的知识库, 是语义搜索、智能问答、决策支持等智能服务的基础技术之一。本文给出知识图谱的定义, 辨析其与本体等相关概念的关系。本体是知识图谱的模式层和逻辑基础, 知识图谱是本体的实例化; 本体研究成果可以作为知识图谱研究的基础, 促进知识图谱的更快发展和更广应用。当前知识图谱研究大多关注静态知识图谱, 不会随时间变化, 而时变的知识图谱探索较少。但是时序信息非常重要, 因为很多结构化的知识只在特定的时间范围内有效, 事实的变化遵循一个时间序列。近来的研究开始将时序信息融入到知识表示学习 (KRL) 和知识图谱补全 (KGC) 中, 称为时序知识图谱。本文从时间嵌入、实体动态、时间关系依赖、时间逻辑推理四个研究领域进行回顾。

关键词 知识图谱; 本体; 知识表示学习; 知识图谱补全; 时序知识图谱

中图法分类号 TP311

文献标识码 A

Context-Aware Temporal Knowledge Graph Embedding

SHI Jin Ni

(School of Information Science and Technology, Dalian maritime university, Liaoning Dalian 116026 China)

Abstract A knowledge graph is a knowledge base that represents objective concepts/entities and their relationships in the form of graph, which is one of the fundamental technologies for intelligent services such as semantic retrieval, intelligent answering, decision support, etc. This paper clarifies the concept of knowledge graph through differentiating it from related concepts such as ontology in that the ontology is the schema layer and the logical basis of a knowledge graph while the knowledge graph is the instantiation of an ontology. Research results of ontologies can be used as the foundation of knowledge graph research to promote its developments and applications. Current knowledge graph research mostly focuses on static knowledge graphs where facts are not changed with time, while the temporal dynamics of a knowledge graph is less explored. However, the temporal information is of great importance because the structured knowledge only holds within a specific period, and the evolution of facts follows a time sequence. Recent research begins to take temporal information into KRL and KGC, which is termed as temporal knowledge graph. This paper reviews the four research fields of time embedding, entity dynamics, temporal relation dependence and temporal logic reasoning.

Key words Knowledge graph; Ontology; KRL; KGC; temporal knowledge graph

1 引言

知识图谱 (knowledge graph) 是以图的形式表现客观世界中的实体 (概念、人、事物) 及其之间关系的知识库。2012 年, 谷歌提出了知识图谱的概念。自此, 知识图谱得到了广泛关注和应用研究, 现已发展成为语义搜索、

智能问答、决策支持等智能服务的基础技术之一。另一方面, 本体的研究起始时间更早, 已积累了较多的研究成果和应用经验。如能确定知识图谱与本体的相关关系, 就可将已有的本体研究成果作为知识图谱研究的基础, 促进知识图谱更

快发展。大规模知识图谱已被用于许多现实世界的应用当中,包括实体链^[1],关系抽取^[2]和问答^[3]。然而,这些KGs仍是很不完整的^[4],这限制了它们的有效性。为了解决这个问题,提出了一种称为知识图谱补全的新技术。KGC的目标是通过知识图谱嵌入(KGE)从现有知识图谱中获得新的事实,换句话说,通过将知识图的实体和关系转化为低维连续空间,并预测新的事实。一般来说,KGE的目标是学习一个得分函数来衡量一个事实的可信性。经典嵌入模型,如TransE^[5],DistMult^[6]和HolE^[7],仅使用KG中观察到的事实来学习得分函数。在最近的许多方法中也引入了附加信息,以进一步提高嵌入性能。

然而,现有方法将关系事实视为时间不变的,而忽略了相应的有效时间周期。事实上,许多关系是随着时间的推移而变化的,它们只在某个时期有效。因此,在知识图谱嵌入中加入时间信息是非常重要的。

2 知识图谱定义

知识图谱是以图的形式表现客观世界中的实体(概念)及其之间关系的知识库。知识图谱的研究起源于语义Web。在2000年的XML大会上,Tim Berners Lee提出了语义Web的理念,目标是为Web网页添加语义,支持机器自动处理,以提供诸如信息代理、搜索代理、信息过滤等语义服务。此后,互联网逐步从仅包含网页与网页之间超链接的文档万维网转变为包含大量描述各种实体和实体之间丰富关系的数据万维网。基于关键词的传统搜索引擎技术也逐渐开始添加语义搜索功能。2005年,美国Metaweb公司成立,致力于开发用于Web语义服务的开放共享的世界知识库。Metaweb基于诸如维基百科、美国证券交易委员会等的公开数据集,提取现实世界中的实体(人或事物)及其之间的关系,然后以图结构存储在计算机中。2010年谷歌收购了Metaweb,获得其语义搜索技术,并于2012年提出知识图谱的概念。

2.1 知识图谱相关概念

本体(ontology)是共享概念模型的显式说明,描述概念与概念间的关系;是语义Web的关键技术,用于Web网页添加语义。语义Web理念中的本体与知识图谱,二者密切相关。本体描述概念及概念间的关系,是大多数知识图谱的模式层,是知识图谱的概念模型和逻辑基础。知识图谱与本体的相同之处在于:二者都通过定义元数据以支持语义服务。不同之处在于:知识图谱更灵活,支持通过添加自定义的标签划分事物的类别。本体侧重概念模型的说明,能对知识表示进行概括性、抽象性的描述,强调的是概念以及概念之间的关系。大部分本体不包含过多的实例,本体实例的填充通常是在本体构建完成以后进行的。知识图谱更侧重描述实体关系,在实体层面对本体进行大量的丰富与扩充。可以认为,本体是知识图谱的抽象表达,描述知识图谱的上层模式;知识图谱是本体的实例化,是基于本体的知识库。

知识图谱采用三元组描述事实,所使用的描述语言大多是已研发的本体语言,如RDFS^[8]、OWL^[9]等。知识图谱也可以通过RDFS或OWL定义规则用于知识推理。知识图谱的关键技术也与本体很相似,涉及:(1)知识图谱构建阶段的实体抽取、关系抽取、语义解析等机器学习和自然语言处理方法和算法,(2)用于知识图谱存储的知识表示、图数据库和知识融合等方法和技术,(3)知识图谱应用阶段的数据集成、知识推理等。

除了本体之外,与知识图谱相关的概念还有知识地图和科学知识图谱。知识地图(knowledge map)将特定组织内的知识索引通过“地图”的形式串联在一起,揭示相关知识资源的类型、特征以及相互关系。知识地图的主要功能在于实现知识的快速检索、共享和再重用,充分有效地利用知识资源^[10]。知识地图是关于知识的来源的知识^[11]。知识并非存储在知识地图中,而是存储在知识地图所指向的知识源中。知识地图指向的知识源包含数据库、文件

以及拥有丰富隐性知识的专家或员工。有的企业应用知识地图来揭示知识的结构,实现对知识及其相关知识的检索。另外,知识地图在文献学中也有应用,即科学知识图谱。科学知识图谱(mapping knowledge domain)是用来显示知识演化进程和知识结构的图形化与序列化的知识谱系。

2.2 知识图谱的构成

知识图谱由数据层(data layer)和模式层(schema layer)两部分构成^[12]。

模式层是知识图谱的概念模型和逻辑基础,对数据层进行规范约束。多采用本体作为知识图谱的模式层,借助本体定义的规则和公理约束知识图谱的数据层。也可将知识图谱视为实例化了的本体,知识图谱数据层是本体的实例。如果不需支持推理,则知识图谱(大多是自底向上构建的)可以只有数据层而没有模式层。在知识图谱的模式层,节点表示本体概念,边表示概念间的关系。

在数据层,事实以“实体-关系-实体”或“实体-属性-属性值”的三元组存储,形成一个图状知识库。其中,实体是知识图谱的基本元素,指具体的人名、组织机构名、地名、日期、时间等。关系是两个实体之间的语义关系,是模式层所定义关系的实例。属性是对实体的说明,是实体与属性值之间的映射关系。属性可视为实体与属性值之间的hasValue关系,从而也转化为以“实体-关系-实体”的三元组存储。在知识图谱的数据层,节点表示实体,边表示实体间关系或实体的属性。

3 时序知识图谱

知识图谱嵌入是将KG的组成部分(包括实体和关系)嵌入到连续的向量空间中,以简化操作,同时保留KG的固有结构。传统的知识库嵌入方法只对观察到的事实三元组进行编码,而忽略了时间敏感实体和事实之间的时间约束。然而,许多应用都涉及到动态图。这给学习和推理带来了重要的挑战,因为节点、属性和边会随着时间而变化。为了

解决这个问题,我们引入了时间感知知识库嵌入,它通过引入时间约束来约束任务。时间信息非常重要,因为结构化的知识只在特定的时期内成立,事实的演变遵循时间顺序。最近的研究开始将时间信息引入KRL和KGC,这被称为时间知识图,与以前的静态知识图不同。已经为同时学习时间嵌入和关系嵌入进行了研究工作。

3.1 时序信息嵌入

时间感知的嵌入主要融入了时序的信息,将三元组扩展到四元组 (h, r, t, τ) ,其中 τ 提供了关事实的额外的时间信息。Leblay and Chekol^[13]在带有时间标注的三元组上研究了时间范围预测,简单地扩展了已有的嵌入方法,基于向量的TransE方法TTransE,定义为:

$$f_{\tau}(h, r, t) = -||h + r + \tau - t||_{L1/2}$$

Ma等人^[14]扩展了现有的静态嵌入方法,并通过用时间戳嵌入替换Tucker的共享权重向量提出了ConT。时间范围4元组在原来3元组的基础上添加了时间范围 $[\tau_s, \tau_e]$,其中 τ_s, τ_e 分别表示三元组有效的起始和终止时间。当给定一个时间戳的时候,可以根据动态KG得到静态的子图 G_{τ} 。HyTE^[15]将时间戳看做超平面 w_{τ} ,将实体和关系的表征进行映射,头实体、尾实体和关系的表示投影分别为 $P_{\tau}(h) = h - (W_{\tau}^T h)w_{\tau}$, $P_{\tau}(t) = t - (W_{\tau}^T t)w_{\tau}$, $P_{\tau}(r) = r - (W_{\tau}^T r)w_{\tau}$ 。时间映射打分函数为:

$$f_{\tau}(h, r, t) = ||P_{\tau}(h) + P_{\tau}(r) - P_{\tau}(t)||_{L1/L2}$$

其中,映射转义公式为 $P_{\tau}(h) + P_{\tau}(r) \approx P_{\tau}(t)$ 。LSE4KGC^[16]将为词序列和时间词序列进行拼接,然后使用LSTM编码拼接后的时间感知的谓词序列。LSTM的最后的隐层作为时间感知的关系嵌入。扩展的TransE和DistMult对应的打分函数分别为 $||h + r_{temp} - t||_2, (h^{\circ}t)r_{temp}^T$ 。通过将实体的上下文定义为包含的事实集合,Liu^[17]提出上下文选择器来捕获有用的上下文,然后使用选择的上下文衡量时序一致性。特别是将目标事实分解成两部分,关系和实体,并分别测量每个部分的上下文的有用性。此外,还采用深度神经网络对上下

文进行编码,并对时间一致性进行评分。这种一致性与事实的合理性一起来模拟事实。Lacroix 等^[18]通过将时序知识图谱补全表述为4阶张量完备化,提出了扩展复分解的 TComplEx,并引入了加权正则化。

3.2 动态实体

现实中的时间会影响实体的状态以及对应的关系。

为了改善时间范围推理,CTPM^[19]将时间范围问题建模为状态变化检测,引入了上下文时间剖面,这是一种富含相关上下文的实体时间剖面,使用上下文学习状态到状态改变的向量。可以了解一个实体何时经历特定的状态变化。然后直接推断与状态变化相关的事实开始或结束时间。Goel 等人^[20]将实体和时间戳作为实体嵌入函数的输入,以保持实体在任何时间点的时间感知特性。通过为静态模型提供一个任意时间点实体特征的实体嵌入函数来构建新的时序知识图谱补全模型。这仅提供静态实体特征的现有时序知识图谱嵌入方法形成对比。所提出的嵌入函数与模型无关,并且可以潜在地与任何静态模型相结合。Know-evolve^[21],深度进化知识网络,研究了实体在KG中的演化以及它们演化之后的关系。其中的时态知识图谱是一组事实,每个事实都有一个时间戳关系。对于嵌入实体和带时间戳的关系,它们使用双线性模型,并使用深度递归神经网络来学习非线性进化的实体。学习阶段支持一个点过程,通过这个点过程,对t时刻的事实是否成立的估计是基于t1时刻的状态。为了捕获节点之间的交互,RENET^[22]通过基于RNN的事件编码器和邻居聚合器建模事件序列。具体来说,RNN用来捕获时间实体交互,通过邻居聚合器聚合并发的交互。基于这两个模块,可以按顺序推断未来的事实。

3.3 时序关系依赖

现有的时序依赖主要以遵循时间线的关系链形式存在,比如 wasBornIn --> graduateFrom --> workAt --> diedIn。Jiang^[23]提出时间感知嵌入,基于时序正则化的联合学习框架,来融合事件序列和一致性信息。定义时序打分函数为

$$f(< r_k, r_l >) = ||r_k T - r_l||_{L1/2}$$

$T \in R^{d \times d}$ 是一个不对称矩阵,对于一个时间序列关系对<r_k, r_l>编码关系的时间顺序。利用整数线性规划公式,进一步应用分离、有序和段落这三个时间一致性约束。

3.4 时序逻辑推理

研究使用逻辑规则进行时序推理。Chekol^[24]探索马尔科夫逻辑网络和概率软逻辑,在不确定时序KG上进行推理。RLvLR-Stream^[25]通过探索谓词和参数的嵌入空间来挖掘封闭路径规则的空间。考虑时间相邻路径规则,从知识图谱流中学习规则的结构用于推理。

参考文献

- [1] Hua, W., Zheng, K., Zhou, X.: Microblog entity linking with social temporal context. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, pp. 1761–1775. ACM (2015).
- [2] Zeng, D., Liu, K., Chen, Y., Zhao, J.: Distant supervision for relation extraction via piecewise convolutional neural networks. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 1753–1762 (2015).
- [3] Luo, K., Lin, F., Luo, X., Zhu, K.: Knowledge base question answering via encoding of complex query graphs. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2185–2194 (2018).
- [4] Shi, B., Wenginger, T.: Open-world knowledge graph completion. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018).
- [5] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: Advances in Neural Information Processing Systems, pp. 2787–2795 (2013).
- [6] Yang, B., Yih, W.t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and

- inference in knowledge bases. arXiv preprint arXiv:1412.6575 (2014).
- [7] Nickel, M., Rosasco, L., Poggio, T.: Holographic embeddings of knowledge graphs. In: Thirtieth AAAI Conference on Artificial Intelligence (2016).
- [8] Pan JZ, Horrocks I. RDFS(FA):Connecting RDF(S) and OWL DL. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(2): 192-206.
- [9] McGuinness DL, Harmelen F. OWL Web ontology language overview. W3C Recommendation, 2004, 63(45): 990-996.
- [10] 陈强, 廖开际, 奚建清. 知识地图研究现状与展望. 情报杂志, 2006, 25(5): 43-46.
- [11] Zins C. Knowledge map of information science. Journal of the Association for Information Science and Technology, 2007, 58(4): 526 - 535.
- [12] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述. 电子科技大学学报, 2016, 45(4): 589 - 606.
- [13] J. Leblay and M. W. Chekol, "Deriving validity time in knowledge graph," in WWW, 2018, pp. 1771-1776.
- [14] Y. Ma, V. Tresp, and E. A. Daxberger, "Embedding models for episodic knowledge graphs," Journal of Web Semantics, vol. 59, p. 100490, 2019.
- [15] S. S. Dasgupta, S. N. Ray, and P. Talukdar, "Hyte: Hyperplane-based temporally aware knowledge graph embedding," in EMNLP, 2018, pp. 2001-2011.
- [16] A. Garcí'a-Durán, S. Dumančić, and M. Niepert, "Learning sequence encoders for temporal knowledge graph completion," in EMNLP, 2018, pp. 4816-4821.
- [17] Y. Liu, W. Hua, K. Xin, and X. Zhou, "Context-aware temporal knowledge graph embedding," in WISE, 2019, pp. 583-598.
- [18] T. Lacroix, G. Obozinski, and N. Usunier, "Tensor decompositions for temporal knowledge base completion," in ICLR, 2020, pp. 1-12.
- [19] D. T. Wijaya, N. Nakashole, and T. M. Mitchell, "CTPs: Contextual temporal profiles for time scoping facts using state change detection," in EMNLP, 2014, pp. 1930-1936.
- [20] R. Goel, S. M. Kazemi, M. Brubaker, and P. Poupart, "Diachronic embedding for temporal knowledge graph completion," in AAAI, 2020, pp. 3988-3995.
- [21] R. Trivedi, H. Dai, Y. Wang, and L. Song, "Know-evolve: Deep temporal reasoning for dynamic knowledge graphs," in ICML, 2017, pp. 3462-3471.
- [22] W. Jin, C. Zhang, P. Szekely, and X. Ren, "Recurrent event network for reasoning over temporal knowledge graphs," in ICLR RLGM Workshop, 2019.
- [23] T. Jiang, T. Liu, T. Ge, L. Sha, B. Chang, S. Li, and Z. Sui, "Towards time-aware knowledge graph completion," in COLING, 2016, pp. 1715-1724.
- [24] M. W. Chekol, G. Pirro, J. Schoenfish, and H. Stuckenschmidt, "Marrying uncertainty and time in knowledge graphs," in AAAI, 2017, pp. 88-94.
- [25] P. G. Omran, K. Wang, and Z. Wang, "An embedding-based approach to rule learning in knowledge graphs," IEEE TKDE, pp. 1-12, 2019.