

基于混合增强智能的知识图谱推理技术研究

杨瑞达¹ 林欣¹ 杨燕^{1*} 贺樑¹ 窦亮²

¹(华东师范大学计算机科学与软件工程学院 上海 200062)

²(国家新闻出版署出版融合发展(华东师大社)重点实验室 上海 200062)

摘要 如今,知识图谱被广泛应用在各个领域,例如问答系统、推荐系统等。而基于知识图谱的应用表现很大程度上依赖于知识图谱本身的知识完备性与准确性。单纯通过人工补齐与审核的方式来构建知识图谱已无法满足超大规模知识图谱的需求。针对上述问题,提出一种基于混合增强智能的知识图谱推理框架,即同时利用机器模型与人的知识信息来完成知识图谱推理。该框架在基于知识图谱嵌入的向量空间中,利用混合增强智能模型来寻找到实体节点之间的有效路径。与现有方法不同的是,该方法在训练模型时,高效地利用人的知识信息来指导模型的优化。实验表明,该框架在公开数据集上的表现相较于现有方法有一定提升。

关键词 知识图谱 知识图谱推理 强化学习 混合增强智能

中图分类号 TP3 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2019.06.029

KNOWLEDGE GRAPH REASONING BASED ON HYBRID-AUGMENTED INTELLIGENCE

Yang Ruida¹ Lin Xin¹ Yang Yan^{1*} He Liang¹ Dou Liang²

¹(School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China)

²(SAPPRFT Key Laboratory of Publishing Integration Development, East China Normal University, Shanghai 200062, China)

Abstract Nowadays, knowledge graph is widely used in various fields, such as question answering system, recommendation system and so on. The performance of application based on knowledge graph depends on the completeness and accuracy of knowledge graph. It is impossible to construct knowledge graph by manual completion and auditing to meet the needs of large-scale knowledge graphs. To address these challenges, we proposed a knowledge map reasoning framework based on hybrid-augmented intelligence. It uses machine model and human knowledge information to complete knowledge map reasoning simultaneously. The framework used hybrid-augmented intelligent model to find effective paths between entities in vector space based on knowledge graph embedding for knowledge reasoning. Different from the existing methods, this method made efficient use of human knowledge information to guide the optimization of the training model. Experiments on public datasets demonstrated the improvement of the proposed framework compared to existing methods.

Keywords Knowledge graph Knowledge graph reasoning Reinforcement learning Hybrid-augmented intelligence

0 引言

近几年,知识图谱被广泛应用在各个领域,例如,

问答系统^[1]、推荐系统^[2-3]和精准医疗系统^[4]等。知识图谱本身的准确性和知识覆盖率将会对上层的应用造成很大的影响并可能成为制约应用表现的瓶颈。而单纯通过人工构建高准确度的知识图谱的方式已经无

法满足海量规模知识图谱的需求。因此,许多人使用知识图谱推理来补充缺失的知识,旨在通过自动化的方式来提升知识图谱的覆盖度。

目前,对于知识图谱推理的方法大体上可以分为两种类型,即基于知识图谱嵌入的方法与基于随机游走模型的方法。前者的代表有 TransE^[5],它将知识图谱中的实体和关系嵌入成为某个空间中的向量,再通过向量之间的运算来表示实体与关系之间的联系。此类方法在公开数据上有较好的表现,但当知识图谱中出现“一对多(某个实体通过某个关系链接到多个相关实体,例如‘美国,总统’将对应到‘特朗普’、‘奥巴马’等多个实体)”等复杂联系时存在较大的局限。另外,基于知识图谱嵌入的方法在可解释性上有所欠缺。路径排序算法 PRA(Path Ranking Algorithm)作为随机游走模型的代表,将两个实体之间存在的路径作为特征,再利用该特征判断两个实体之间是否存在某个关系。这种方法具有更好的可解释性,但它作用在离散的特征空间中,导致难以评估实体和关系之间的相似性。

针对上述问题,本文提出一种基于混合增强智能的知识图谱推理方法,利用强化学习模型进行知识图谱推理,并高效地加入了人的判断信息来获得更加准确的知识图谱推理结果。首先,本文以 TransE 训练获得的知识图谱嵌入向量为基础,搭建强化学习模型。然后,训练强化学习智能体(agent),使其在每一对实体之间寻找可用作推理的有效路径(path)。本文不仅在强化学习的回报函数设计时考虑到了“路径是否可达”,“路径长度”等因素,还将人的知识信息加入到回报函数中。最后,为了评估混合增强学习方法的表现,本文在公开数据集 NELL(Never End Language Learning)上进行了对比实验。实验结果表明,本文提出的方法有利于提高知识图谱推理的精度。

本文的主要贡献如下:1) 本文将强化学习模型应用于知识图谱推理领域,相比于现有的模型框架,设计了更加有效、合理的回报函数;2) 本文将人的知识信息准确高效地加入到模型中。在强化学习模型训练的过程中,同时考虑“重要性”和“不确定性”两个因素,将部分机器决策结果抛出给人进行判断。3) 本文在公开数据集 NELL 上进行了实验。与多种类型的算法进行了对比,证明了本文的框架的有效性。

1 相关工作

目前,对于知识图谱的关系推理问题,一类方法是

采用将知识图谱嵌入到向量空间中来进行计算的方法^[5-10]。最早的基于平移的知识图谱嵌入方法是文献[5]提出的 TransE 方法,该方法将知识图谱中的实体和关系均嵌入到同一个向量空间中,并且通过训练使得它们之间存在如下关系:头实体向量加关系向量约等于尾实体向量。该方法在数据集上有较好表现,但无法准确处理“一对多”的情况。在此基础上,文献[6]提出了 TransD 的方法,将实体和关系分别投影到了不同的空间,并且由实体和关系共同来确定投影矩阵,由此更好地解决了“一对多”的问题。在知识图谱嵌入的基础上,有些研究的方法加入了其他信息的考虑^[11-13]。其中,文献[11]提出利用知识图谱中路径的信息结合知识图谱嵌入来进行知识图谱预测。在嵌入两个实体之间关系的基础上,将实体之间的路径嵌入为向量。

除此之外,另一类方法是基于发现路径来进行知识图谱的预测。文献[14]率先提出了这类方法,通过随机游走的方法来发现路径从而进行关系推理。此后,文献[15]提出了基于递归随机游走的方法,同时利用了知识图谱信息和文本信息来进行知识图谱的推理。但此类使用随机游走的方法会寻找到过量的推理路径而对推理结果造成影响。文献[16]将强化学习框架运用到了知识图谱推理当中,并且结合了知识图谱嵌入的信息,获得了更好的推理表现。

本文采用了强化学习为基础进行推理路径的寻找,与上述研究有所区别的是,本文重新设计了强化学习的回报函数。此外,本文提出将人的知识信息融入模型的训练,以此提高知识图谱推理的效果。

2 基于混合增强智能的知识图谱推理

2.1 基于强化学习的推理模型

本文采用文献[16]的底层框架,即马尔科夫决策过程来作为强化学习的基础模型。 $\langle S, A, P, R \rangle$ 可以用来表示马尔科夫决策过程的四个要素。其中 S 表示当前模型 Agent 所处的状态(State); A 表示模型进行的行为(action)的选择; P 表示从当前状态到下一个状态的概率矩阵; R 表示当前状态下,进行某一个或某一系列行为选择的回报函数。此外,本文使用一个策略网络 $\pi_{\theta}(s_t, a) = p(a|s; \theta)$ 来表示强化学习的 Agent。具体介绍如下:

状态(state):强化学习的重要组成部分,它表征了当前 Agent 所处的环境,并影响到行为的选择。在状态向量的设定中考虑到了知识图谱中隐藏的语义信

息,使用 TransE 模型将知识图谱中的实体嵌入到连续的向量空间中。在强化学习模型中,状态表征了当前模型在知识图谱中的实体位置。当选择一个行为之后,模型通过关系移动到了下一个实体位置上。给定一个当前位置 n 的实体 e_n 与一个目标实体 e_{target} , 状态的定义如下:

$$s_n = (e_n, e_{target})$$

式中: e_n 表示当前实体嵌入的向量, e_{target} 表示目标实体嵌入的向量。当表示起始实体 e_{source} 所处的状态时, $e_n = e_{source}$ 。

行动(action): 在确定的某一关系下, 对于给定的一对实体对 (e_i, e_j) , 本文通过模型去寻找用于知识图谱的推理的有效路径。对于一个实体 e , 本文将强化学习的行为定义为选择一个“关系”, 使当前的实体链接到下一个实体。以此类推, 多个行为选择的关系即能构成某一条从头实体 e_{source} 到达尾实体 e_{target} 的路径。

回报函数(Reward): 回报函数是对于模型采取一系列行为的反馈值。本文在回报函数设计时, 考虑到了可达性、路径长度、路径区分度等因素。具体的定义如下:

1) 可达性: 在一个复杂的知识图谱中, 每一个实体通过多个关系与其他实体相连。因此, 从起始实体开始, 可选择的路径数量随着路径长度的增长而指数增加。在众多的路径中选择出可用于推理的有效路径是强化学习模型的任务。本文首先将可达性这一因素考虑到回报函数中, 即从头实体 e_{source} 出发通过该条路径能够达到目标实体 e_{target} 。可达性的具体定义如下:

$$R_{reachability} = \begin{cases} 1 & \text{到达目标实体 } e_{target} \\ -1 & \text{其他} \end{cases} \quad (1)$$

若当前的路径能够到达目标实体, 给该条路径的回报是 1, 否则为 -1。

2) 路径长度: 一个关系能够表示一对实体之间的信息联系。随着路径长度的增加, 即路径包含的关系数量的增加, 信息的联系也在逐渐减弱。在训练模型寻找有效路径的过程中, 本文更倾向寻找长度更短的路径来进行推理。因此, 在回报函数中加入了考量路径长度的项, 具体定义如下:

$$R_{length} = \frac{1}{l_{path}} \quad (2)$$

式中: l_{path} 表示当前路径的长度, 即其中包含的关系的数量。

3) 路径区分度: 对于某一个关系而言, 由于样本中头尾实体的表达均为固定向量, 模型在学习过程中, 将会倾向于寻找语义相似的路径。以上情况可能导致寻找得到的路径存在着一定程度的信息冗余。因此,

在回报函数中加入“路径独立性”的考虑, 训练模型寻找更多数量且彼此不同的路径。具体定义如下:

$$R_{diversity} = \frac{1}{num(P)} \sum_{p_i \in P} \left(\frac{1}{\cos(p, p_i) + 2} - \cos(p, p_i) \right) \quad (3)$$

式中: $\cos(p, p_i)$ 表示两条路径之间的余弦距离。路径向量 $p = r_1 + r_2 + \dots$ 表示组成该路径的关系向量之和。 P 表示已发现的所有推理路径的集合。

2.2 基于混合增强智能的推理模型

强化学习模型训练过程中, 回报函数引导了每一次模型进行行为选择。因此, 回报函数的设定对于强化学习模型的最终表现有非常重要的影响。在 2.1 节的基础模型中, 回报函数的设计考虑到了知识图谱结构上的特征, 但没有融入常识信息和语义信息特征。而常识信息和语义信息通常难以准确获取并加入到回报函数中。为改善这一问题, 本文提出了混合增强智能的推理模型, 即利用人的知识信息来优化模型。

考虑到人工的资源稀缺并且价格高昂, 将人的知识信息融入强化学习的模型中需要保证信息利用的高效性。具体来说, 混合智能模型需要考虑两个关键问题: 1) 如何将最需要进行人工判断的信息返回给人; 2) 人工的判断结果如何返回给机器进行处理。在本文的应用场景下, 由于模型进行行为选择的次数多, 并且人工难以直观判断每一步行为选择是否合理, 本文将路径为粒度, 将需要判断的路径抛出给人工进行处理。

1) 不确定性: 文献[17]在判断两个实体之间是否存在某一个关系时, 给出属于 $[0, 1]$ 的概率值, 并且认为当概率值更接近两端时 (0 或 1), 给出的结论更加准确。在本文的问题场景下, 当强化学习中的模型面对行为的选择时, 给出的是选择各个行为的概率分布, 最终扩展成为选择路径的概率分布。受到文献[17]的启发, 本文认为模型在进行路径选择时, 其概率分布应当更加有差异性, 即机器对选择或不选择某一路径更加具有把握。反之, 对于各个路径的概率分布趋于平均的情况, 可以认为机器难以区分其中哪一条路径为更佳选择。本文通过计算熵的方式来评估路径选择的概率分布情况。在路径的概率分布越平均时, 熵的值越大, 反之则越小。本文中熵的计算公式定义如下:

$$H(T) = - \sum_{t_i \in T} p(t_i) \log(p(t_i)) \quad (4)$$

式中: $p(t_i)$ 表示选择路径 t_i 的概率。

2) 重要性: 除了不确定性之外, 本文将“重要性”也作为“是否将路径抛出给人判断”的衡量维度。对于某一个关系而言, 在模型的学习过程中, 某一条路径

可能会被选中多次,从而在关系预测时它将提供更大的推理权重。如果该条路径能够提供的推理信息量并不能和上述权重相匹配,则可能对最终的推理结果造成影响,而且这种影响将会比小权重的路径发生匹配错误的后果更加严重。因此,本文将此类重要路径抛出,交给人工判断,以保证它们能够提供更加准确的推理信息量。具体来说,给定一个关系 r ,通过模型获得多个推理路径 $t_i \in T$,重要性的评估方式为计算其中累积概率的最大值:

$$t_{\text{selected}} = \max(\sum p(t_i)) \quad (5)$$

最后,综合不确定性和重要性两个维度的考虑,本文选择抛出给人工判断的路径的规则为:对于每一个关系,若 $H(T) > c$,则选择该关系推理路径中的 t_{selected} 提供给人进行判断,其中 c 为常数。

上文中提到,在人工进行判断时,人更容易理解路径和关系之间的推理关系并且给出判断。因此,本文设计了用于路径展示和推理评分的人机交互的界面。例如,当推理的目标关系为“运动员的主场”时,给出路径(“运动员在某队服役”,“队伍的主场”)供人工判断以下推理是否成立:“运动员在某队服役”(A,B) \wedge “队伍的主场”(B,C) \Rightarrow “运动员的主场”(A,C)。具体的判断过程为,人对于当前的推理是否合理来给出1到5的评分。为保证人工判断的准确性,某一组待判断的推理路径可能分配给1~3个人进行判断,最终取其平均分并将结果反馈给机器模型。

在获得人工判断的评分结果后,本文将加入到该路径的回报函数中,并重新训练模型的参数。具体来说,对于人工判断完成的路径,本文在2.1节基础的回报函数上加入人工反馈项 R_{human} 。人工反馈项的定义如下:

$$R_{\text{human}} = (\text{score} - 3)^3 \quad (6)$$

式中: score 表示人对于该条推理路径的评分。对于人工判断完成的路径,其回报函数的定义如下:

$$R_{\text{sumH}} = R_{\text{sum}} + \lambda_4 R_{\text{human}} \quad (7)$$

式中: $R_{\text{sum}} = \lambda_1 R_{\text{reachability}} + \lambda_2 R_{\text{length}} + \lambda_3 R_{\text{diversity}}$ 表示除人工反馈以外的回报函数。若该条路径被模型重复选中,则不再抛出给人工进行判断,而直接使用前一次的人工判断结果。最后,本文计算该条路径综合的回报函数,并将其用于更新模型的参数。

2.3 推理模型的训练过程

在实际训练过程中,强化学习的模型面临选择的路径数量是巨大的,从而导致训练难以进行。这是由于常用的知识图谱结构复杂,每个实体都与多个实体

相连。在模型进行路径寻找时,路径的选择空间会随着路径长度的增加而呈指数型增长,而其中能够用作关系推理的路径数量只占极小的一部分。进而导致模型在训练过程中,只接收到大量的负反馈,使得训练难以继续。

受文献[16]和文献[18]的启发,本文采用预训练方法来解决这一问题。具体来说,对于每一个关系,在知识图谱中获取到存在该关系的实体对集合。对于每一对实体对 (e_i, e_j) ,搜索获取 x 跳(即包含 x 个关系)内的所有路径 $T_{(e_i, e_j)}$ 。然后,获取各个实体对之间的共有路径。最后,使用这些路径来对强化学习模型进行预训练,来初步更新模型中的参数。

之后,本文对混合强化学习模型进行训练。训练过程使用蒙特卡洛策略梯度来更新参数。目标函数为模型累积行为选择获得的回报函数最大化,具体定义如下:

$$O(\theta) = \sum_t \sum_a \pi_\theta(s_t, a) R \quad (8)$$

式中: R 根据发现的路径所属的不同情况发生变化。给定某一个关系与实体对 $(e_{\text{source}}, e_{\text{target}})$,从头实体 e_{source} 开始,通过强化学习模型来寻找能够到达 e_{target} 的路径。由于模型在行为选择时可能选择到与当前实体无连接的“关系”,这种情况下使用回报函数给予惩罚。此外,本文还设置了最大路径阈值 l_{max} ,即模型在搜索路径时,若路径长度达到阈值则停止搜索。除上述情况外,更新策略如下:

1) 对于路径长度达到阈值而未到达 e_{target} 的路径,本文通过 $R_{\text{reachability}}$ 来对该路径进行惩罚;

2) 对于在长度阈值内到达 e_{target} 的路径,本文进行是否抛出给人的判断。若不必抛出给人工进行判断,则直接通过回报函数 R_{sum} 来更新模型参数。若需要抛出给人工判断,待人工判断结果返回后,对于该条路径通过回报函数 R_{sumH} 进行二次更新参数。

3 实验与分析

3.1 实验数据

为了验证混合增强智能方法的有效性,本文选择知识图谱推理中常用的两大任务进行评估,即链接预测和事实预测。链接预测任务为给定一组头实体与关系来预测尾实体。事实预测任务为判断“三元组事实”的正误。本文采用 NELL 的子数据集来进行实验评估。该数据集的基本信息如表1所示。

表1 数据集基本信息

名称	实体数量	关系数量	三元组数量
NELL-995	75 492	200	154 213

NELL 数据集是从无结构的网页中抽取到有效信息并构成结构化的全科知识图谱。NELL 的机器模型经历了数年时间的持续迭代,到如今已经积累了 2 810 379 个高可信度的三元组信息。在迭代的过程中,NELL 模型还采用了人工对于三元组的可信程度的评分信息,用于提高模型的表现和结构化知识的质量。

3.2 实验设置

本文选择了常用的 NELL-995 数据集进行实验,即从 NELL 模型的第 995 次迭代结果中获取数据。在知识图谱关系的选择上,本文选择了 Top 200 的关系作为实验的候选关系。此外,本文在两个任务上横向对比了现有的数个方法。这些方法是知识图谱嵌入类方法(TransE, TransR^[10])与强化学习类方法(DeepPath^[16])。本文方法简称为 HAI (Hybrid-Augmented Intelligence)。

混合增强智能模型的相关参数设置为: $H(T)$ 的阈值常数为 $c = 0.07$;知识图谱的嵌入维度为 100 维,状态的维度为 200 维;最大路径阈值 $l_{\max} = 20$; $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 四个参数分别为 0.05、0.45、0.1、0.4。

本文采用平均精度均值 MAP (Mean Average Precision) 来作为知识图谱推理的评价指标。在实际计算时,机器模型将输出排序的结果,再根据公式来评估 MAP 值。具体计算公式如下:

$$AP = \frac{1}{|\Omega_u|} \sum_{i \in \Omega_u} \frac{\sum_{j \in \Omega_u} h(p_{uj} < p_{ui}) + 1}{p_{ui}} \quad (9)$$

式中: Ω_u 为实际正确的结果集合。 p_{ui} 表示某结果所处的排序中的位置。 $h(p_{uj} < p_{ui})$ 表示在排序列表中,排在 p_{ui} 前的 p_{uj} 的个数。对于不同的预测任务计算得到 AP 值后,再计算 AP 的均值即可获得 MAP 值。

3.3 实验结果

本文在链接预测和事实预测两个任务下进行了对比实验。表 2 展示了在链接预测任务中几种方法对于不同关系的预测结果。从实验结果来看,基于强化学习的方法整体上能够优于基于知识图谱嵌入的方法。这是由于基于强化学习的方法在模型搭建时利用到了知识图谱嵌入的向量,其中蕴含了隐藏的语义信息。在此基础上通过强化学习的训练,能够更加精准地寻找路径从而提升推理的效果。

表2 链接预测任务结果对比

关系	TransE	TransR	Deep Path	HAI
athlethomestadium	0.718	0.722	0.890	0.852
athleteplaysforteam	0.627	0.673	0.750	0.757
athleteplayssport	0.876	0.963	0.957	0.957
organizationheadquarter edincity	0.620	0.657	0.790	0.790
organizationhireperson	0.719	0.737	0.742	0.746
personleadsorganization	0.751	0.772	0.795	0.794
teamploysinleague	0.891	0.882	0.881	0.866
worksfor	0.677	0.692	0.711	0.700
⋮	⋮	⋮	⋮	⋮
平均	0.737	0.781	0.796	0.795

对于部分关系的预测结果,混合增强智能模型与强化学习模型方法的实验结果相差较小。这是由于在返回结果供人工进行判断时,人工的评分接近中性(3分),因此未对模型的训练过程造成较大的影响。对于个别关系,基于混合增强智能模型的结果相较于 DeepPath 方法的 MAP 值更低,可能造成的原因是:人工评分路径的数量有限,在降低某一个路径的权重后,另一个低质量路径的权重占比相对提升,从而导致对推理结果造成影响。在单一关系下,混合增强智能模型相对强化学习模型最高提升约 8%。这是由于混合增强智能模型在强化学习模型的基础上额外融入了人的知识信息,使得模型更好地寻找推理路径。具体分析发现,在该关系下,人工评分对于某一高权重的低质量路径进行了降权,从而优化了推理结果。

此外,本文还在事实预测任务中进行了实验,结果如表 3 所示。事实预测任务直接评估某一个关系下,正三元组样本到负样本的降分排序。混合增强智能方法在融入人的知识信息后,相比于其他方法表现出更高的精度。

表3 事实预测任务结果对比

关系	TransE	TransR	Deep Path	HAI
agentbelongstoorganization	0.349	0.374	0.335	0.330
athlethomestadium	0.381	0.410	0.729	0.756
athleteplayssport	0.522	0.389	0.413	0.440
organization headquarter edincity	0.257	0.320	0.642	0.642
organizationhireperson	0.307	0.340	0.508	0.475
personborninlocation	0.265	0.308	0.289	0.398
teamploysport	0.356	0.441	0.398	0.398
worksfor	0.232	0.281	0.461	0.440
⋮	⋮	⋮	⋮	⋮
平均	0.383	0.389	0.492	0.495

表4展示了人工对于部分推理路径的评分。例如,“A实体和组织B实体合作”无法直接推理出“A实体属于组织实体B”,因此给以2分的评分。“A实体服役于队伍实体B且队伍实体B的主场是地点实体C”可以推理出“A实体的主场是地点实体C”,因此给以高分反馈。

表4 人工对于部分推理路径的评分

关系	推理路径	人工评分
agent belongs to organization	agent collaborates with agent	2
athlethomestadium	athlete plays for team -> team homestadium	5
athlete plays sport	athlete plays for team -> team plays sport	5
person leads organization	worksfor	2

4 结 语

本文提出了一种基于混合增强智能的知识图谱推理方法,该方法在强化学习的基础模型上加入了混合智能模型,并利用了知识图谱嵌入信息。其特色在于:一方面将人的知识信息高效地融入到机器模型的训练中,其中包含了机器难以获取的丰富常识信息,这些信息能够帮助机器更加准确地寻找推理的路径。另一方面,强化学习模型的搭建利用到了知识图谱嵌入的知识,在一定程度上能够挖掘知识图谱上的语义信息。实验结果表明本文提出的基于混合增强智能的知识图谱推理方法在公开数据集上的效果,与现有的知识图谱嵌入方法、强化学习方法相比有一定程度的提升。

在接下来的工作中,将考虑在强化学习中设置更加全面的回报函数。同时,在考虑“何时抛出需要人工判断的路径”的问题时,设计更加准确的模型来执行判断。

参 考 文 献

[1] Cui W, Wang H, Wang H, et al. KBQA: learning question answering over QA corpora and knowledge bases[J]. Proceedings of the Vldb Endowment, 2017, 10(5):565-576.

[2] Catherine R, Cohen W. Personalized Recommendations using Knowledge Graphs: A Probabilistic Logic Programming Approach[C]//Acm Conference on Recommender Systems. ACM, 2016.

[3] Palumbo E, Rizzo G, Troncy R. entity2rec: Learning User-Item Relatedness from Knowledge Graphs for Top-N Item Recommendation[C]//Eleventh Acm Conference on Recommender Systems. 2017.

[4] Rotmensch M, Halpern Y, Tlimat A, et al. Learning a

Health Knowledge Graph from Electronic Medical Records [J]. Scientific Reports, 2017, 7:5994.

[5] Bordes A, Usunier N, Garcia-Duran A, et al. Translating Embeddings for Modeling Multi-relational Data[C]//International Conference on Neural Information Processing Systems. Curran Associates Inc. 2013:2787-2795.

[6] Ji G, He S, Xu L, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]//Meeting of the Association for Computational Linguistics and the, International Joint Conference on Natural Language Processing. 2015:687-696.

[7] Fan M, Zhou Q, Chang E, et al. Transition-based knowledge graph embedding with relational mapping properties [C]//Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing. 2014.

[8] Xiao H, Huang M, Hao Y, et al. TransA: An Adaptive Approach for Knowledge Graph Embedding[EB]. arXiv:1509.05490, 2015.

[9] Wang Z, Zhang J, Feng J, et al. Knowledge Graph Embedding by Translating on Hyperplanes[C]//Twenty-eighth Aaii Conference on Artificial Intelligence. AAAI Press, 2014.

[10] Lin Y, Liu Z, Zhu X, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Twenty-ninth Aaii Conference on Artificial Intelligence. 2015.

[11] Lin Y, Liu Z, Luan H, et al. Modeling Relation Paths for Representation Learning of Knowledge Bases[EB]. arXiv:1506.00379, 2015.

[12] Wang Z, Li J. Text-enhanced representation learning for knowledge graph[C]//International Joint Conference on Artificial Intelligence. AAAI Press, 2016.

[13] Xiao H, Huang M, Hao Y, et al. TransG: A Generative Mixture Model for Knowledge Graph Embedding[EB]. arXiv:1509.05488, 2016.

[14] Lao N, Mitchell T, Cohen W W. Random walk inference and learning in a large scale knowledge base[C]//Conference on Empirical Methods in Natural Language Processing. 2011.

[15] Wang W Y, Cohen W W. Joint Information Extraction and Reasoning: A Scalable Statistical Relational Learning Approach [C]//Meeting of the Association for Computational Linguistics & the International Joint Conference on Natural Language Processing. 2015.

[16] Xiong W, Hoang T, Wang W Y. DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning [EB]. arXiv:1707.06690, 2017.

[17] Lin X, Peng Y, Choi B, et al. Human-Powered Data Cleaning for Probabilistic Reachability Queries on Uncertain Graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2017,29(7):1452-1465.

[18] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587):484-489.