

DOI: 10.11992/tis.201805001

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20180702.1311.002.html>

知识图谱的推荐系统综述

常亮, 张伟涛, 古天龙, 孙文平, 宾辰忠

(桂林电子科技大学 广西可信软件重点实验室, 广西 桂林 541004)

摘要: 如何为用户提供个性化推荐并提高推荐的准确度和用户满意度, 是当前推荐系统研究面临的主要问题。知识图谱的出现为推荐系统的改进提供了新的途径。本文研究了知识图谱近年来在推荐系统中的应用情况, 从基于本体的推荐生成、基于开放链接数据的推荐生成以及基于图嵌入的推荐生成3个方面对研究现状进行了综述。在此基础上, 提出了基于知识图谱的推荐系统总体框架, 分析了其中涉及的关键技术, 并对目前存在的重点和难点问题进行了讨论, 指出了下一步需要开展的研究工作。

关键词: 知识图谱; 推荐系统; 本体; 开放链接数据库; 图嵌入; 网络表示学习; 相似度; 预测评分

中图分类号: TP301 **文献标志码:** A **文章编号:** 1673-4785(2019)02-0207-10

中文引用格式: 常亮, 张伟涛, 古天龙, 等. 知识图谱的推荐系统综述[J]. 智能系统学报, 2019, 14(2): 207-216.

英文引用格式: CHANG Liang, ZHANG Weitao, GU Tianlong, et al. Review of recommendation systems based on knowledge graph[J]. CAAI transactions on intelligent systems, 2019, 14(2): 207-216.

Review of recommendation systems based on knowledge graph

CHANG Liang, ZHANG Weitao, GU Tianlong, SUN Wenping, BIN Chenzhong

(Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China)

Abstract: In current research on recommendation systems, the provision of personalized recommendations to users and the improvement of the accuracy and user satisfaction of recommendations are main concerns. The emergence of knowledge graphs provides a new way to improve recommendation systems. The applications of knowledge graphs to recommendation systems in recent years are summarized in this paper, and the current status of the research is investigated in detail from three aspects: ontology-based recommendation generation, recommendation generation based on linked open data, and recommendation generation based on graph embedding. On this basis, this paper proposes the general framework of recommendation systems based on knowledge graph, analyzes the key technologies involved, discusses the existing key issues and difficulties, and indicates the further research work to be carried out.

Keywords: knowledge graph; recommendation system; ontology; linked open data; graph embedding; network representation learning; similarity; prediction score

移动互联网、人工智能等技术的迅速发展为人们的工作生活带来了许多便利, 但是同时也带来了信息过载问题^[1]。搜索引擎和推荐系统是解决信息过载的代表技术。传统的搜索引擎在本质上来讲是帮助用户过滤和筛选信息, 这种方式满足了大多数人的需求, 但没有提供个性化的服

务。相对于传统搜索引擎来说, 推荐系统可以兼顾个性化需求和解决信息过载问题。推荐系统是信息过滤系统的一个子集, 目的在于根据用户的喜好、习惯、个性化需求以及商品的特性来预测用户对商品的喜好, 为用户推荐最合适的商品, 帮助用户快速地做出决策, 提高用户满意度^[2]。推荐系统的价值在于能够提供尽量合适的选择或者是推荐而不需要用户明确提供他们所想要的内容。随着大数据时代的到来, 传统推荐系统在挖掘数据价值上存在的问题正在限制其性能发挥^[3]。

收稿日期: 2018-05-02. 网络出版日期: 2018-07-02.

基金项目: 国家自然科学基金项目(61572146, U1501252, U1711263); 广西创新驱动重大专项项目(AA17202024); 广西自然科学基金项目(2016GXNSFDA380006).

通信作者: 宾辰忠. E-mail: binchenzhong@guet.edu.cn.

知识图谱的出现为大数据环境下的推荐系统设计提供了一种有效途径。

2012年5月17日谷歌正式提出了“知识图谱”这个术语^[4]。知识图谱旨在描述真实世界中存在的各种实体或概念,以及他们之间的关联关系。其中,每个实体或概念用一个全局唯一确定的ID来标识,每个属性-值对用来刻画实体的内在属性,而关系用来连接两个实体,刻画它们之间的关联。知识图谱可以融合多种数据源丰富数据语义信息,并且可以结合推理得到的隐含信息为用户提供服务。随着信息检索、智慧城市^[5]等应用领域的发展要求,将知识图谱应用于这些领域以提高用户体验和系统性能,成为学术界和工业界关注的热点。

在推荐系统领域,人们关注的往往是用户与项目之间的联系,缺少对用户与用户、项目与项目等之间的相互联系的考虑。基于知识图谱的推荐系统通过将用户与用户、用户与项目以及项目与项目之间相互连接起来,增强数据的语义信息以进一步提高推荐准确度,具有重要的研究意义和实用价值,逐渐成为推荐系统研究领域最为活跃的分支之一。

1 基于知识图谱的推荐系统

随着互联网上信息的爆发式增长,对于多源异构数据的挖掘是推荐系统中的关键问题。知识图谱的提出为推荐系统中多源异构数据的整合处

理带来了新的契机,而基于知识图谱的推荐系统也作为一个新的研究领域被人们所关注。

知识图谱的提出最初是为了优化搜索引擎,将数据组织成相互链接并且对于其中的每个实体都富含属性标签,实体间存在多种关系,以此来实现搜索引擎的优化^[6]。在对其进行深入研究后,了解到知识图谱的应用不局限于搜索,在诸如智能问答、推荐系统以及金融诈骗领域均有涉猎^[7]。

基于知识图谱的推荐系统主要是利用知识图谱对多源异构数据的整合性,可以对大数据环境下互联网上的数据进行知识抽取^[8],得到更加细粒度的用户和项目的特征信息,从而更精准的计算用户与用户、用户与项目以及项目与项目之间的相关性,最后为用户做出推荐。

该推荐有3个组成要素:用户知识图谱、项目知识图谱、推荐方法。在此基础上,我们给出了基于知识图谱的一个推荐系统模型,如图1所示。该模型分为客户端和服务端,客户端收集用户的原始数据(包括浏览数据、上下文数据如时间、用户状态等)上传至服务器端进行处理。服务器端一方面从各个垂直网站及百科网站中进行相关数据搜集,构建项目知识图谱;另一方面,对用户偏好进行获取,建立用户知识图谱;此外,通过分析推理的方法获取上下文信息,利用GPS定位来获取当前用户地理位置信息等。最后,综合上述信息,通过推荐产生器向目标用户推送其感兴趣的项目,并基于用户反馈对推荐性能进行评估,进一步调整推荐产生器,以适应用户偏好。

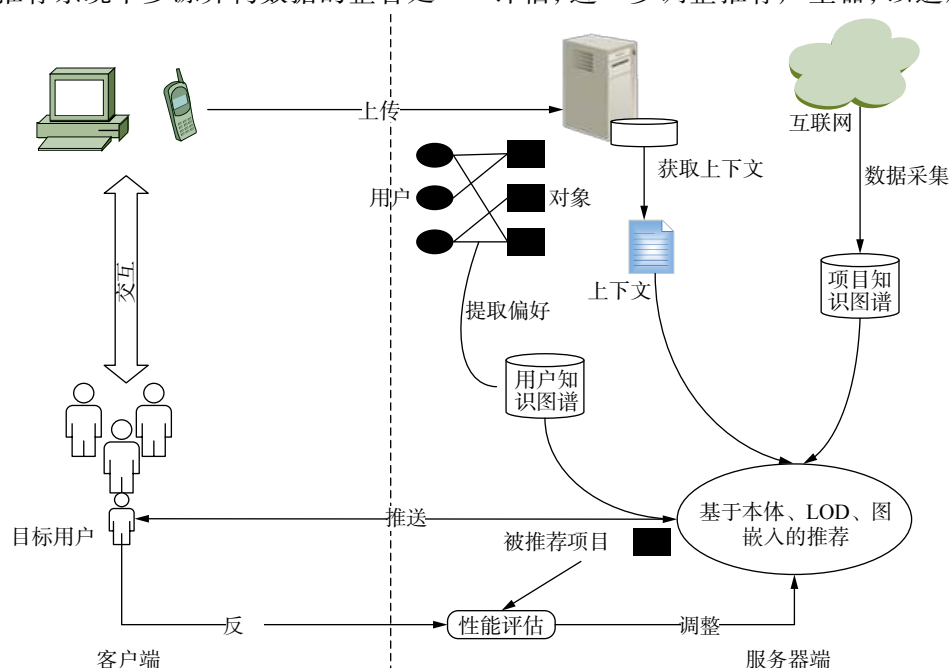


图1 基于知识图谱的推荐系统总体框架

Fig. 1 The overall framework of recommender system based on knowledge graph

对于已给出的基于知识图谱推荐系统整体框架,首先在文献中对该推荐系统进行较为全面的考察,然后围绕该框架对该推荐系统的关键技术进行分析。在对基于知识图谱的应用情况进行总结的基础上,对基于知识图谱的推荐系统面临的重点和难点问题进行讨论。

2 基于知识图谱的推荐系统

依据推荐过程中知识图谱表现形式,本文将文献中的基于知识图谱的推荐方法大致分为两大类:基于本体的推荐生成、基于开放链接数据(linked open data, LOD)的推荐生成。此外,鉴于近年来将知识图谱嵌入低维向量空间的方法在推荐领域扮演的角色越来越重,本节将基于图嵌入的推荐生成单独作为一类来考察。

2.1 基于本体的推荐生成

本体^[9]作为知识图谱的一种体现形式,强调的是概念之间的层级关系,有逻辑学和语义学的基础,支持简单的自动推理。基于本体的推荐生成的主要思路是:利用本体中细粒度的对概念层级关系的描述,通过对概念的细粒度的分类描述来更加精准的对实体特征进行表示,从而挖掘出事实中蕴含的深层次信息。融入现有的基于内容的推荐^[10]、基于协同过滤的推荐方法^[11]中。最后结合用户当前的上下文信息生成推荐^[12]。

Niaraki等^[13]较早地提出基于本体的推荐系统。以用户模型和上下文模型为基础构建本体知识库,基于本体中概念的层级关系采用多准则决策来扩展以用户为中心的路线规划建模技术,最后通过比较判断过程和一致性指标对相关概念进行加权,确定用户和上下文模型来满足用户的需求。将用户和环境结合起来,基于本体中概念的层级关系,提供以用户为中心的路线规划,克服了标准建模不足的缺点,提高了系统推荐路线的准确率。与此类似,Dodwad等^[14]也是基于本体的层次结构对概念进行加权,更加细粒度的分析用户的偏好。Saraswathib等^[15]则针对求职者在门户网站的海量数据面前无法快速的筛选工作,动态地构建本体,通过重排序和相似度算法为用户推荐职位。

除了上述基于本体的层次结构以及概念之间的逻辑关系进行推荐之外,还有研究人员基于传统的推荐算法结合本体完成推荐任务。

Moreno等^[16]完成基于本体和协同过滤的混合推荐系统 SigTur/E-Destination。首先,基于人口特征、旅游景点及路线信息构建旅游本体。在新

用户访问系统时,利用调查问卷中用户信息进行特征抽取,与旅游本体中人物特征关联进行用户建模。基于用户兴趣权重及用户特征信息计算用户的兴趣度(S)及置信度(CL),得到一个推荐列表。然后,基于用户评分、交互行为等,利用基于内容和协同过滤技术对信息进行过滤,通过 Pearson 回归算法进行相似度排序,进而得出推荐列表。最终结合 GIS 定位系统再次过滤,并使用 ICEfaces 框架实时更新推荐列表。为了验证推荐的准确率,作者将推荐结果与用户反馈进行对比,结果表明,90%以上用户认为该系统推荐较为准确。

总体来看,基于本体的推荐生成技术的优点在于:能够细粒度化概念之间的上下文关系,增强数据之间的关联性,可以更加细粒度的分析用户的偏好。缺点在于:本体的构建费时,因为现阶段主要还是通过手工构建,在某些特定领域需要专家的参与。基于本体的推荐的范围过窄,因为通用的大规模本体少,大多数本体只针对某个具体应用领域构建。此外在实际应用中不同本体之间映射、扩充与合并处理等操作复杂,现实的知识体系变化时先前构建的本体必须作相应的演化保持一致性,导致更新困难。新用户问题,当一个新用户访问系统时,由于没有他的行为数据,所以也无法根据他的历史行为预测其兴趣,从而无法借此给他做个性化推荐。针对上述问题,可以从以下方向进一步研究:应用深度学习等相关方法半自动甚至自动地构建本体,加快本体的构建过程。对具体的应用可适当地扩大通用的、适用范围更广的本体,丰富本体的语义,增强本体间的联系。还可以增加对特征的描述,同时对特征描述进行推理,从而得出事实中蕴含的结论并将其应用于推荐,以此来及时地更新推荐,使推荐更具多样化。新用户的问题可以利用用户注册时提供的年龄、性别等数据做进行粒度的个性化。借助用户的社交网络账号登录(需要用户授权),导入用户在社交网站上的好友信息,然后给用户推荐其好友喜欢的物品。此外还可以要求用户在登录时对一些物品进行反馈,收集用户对这些物品的兴趣信息,再给用户推荐那些和这些物品相似的物品。

2.2 基于 LOD 的推荐生成

基于 LOD 的推荐生成技术主要思路是:将链接数据库中丰富的语义信息融入到现有的方法中,着重考虑用户偏好、项目之间属性相似度。通过利用 LOD 中大量相互关联的数据,更加精细

化的衡量资源之间的相似性,挖掘用户的偏好,最后结合上下文信息生成推荐结果。因此在引入 LOD 之后,语义相似度的计算成为关键所在。目前,语义相似计算的方法主要包括链接数据语义距离(linked data semantic distance)、基于隐式语义反馈的路径算法(SPrank)等。

Passant 等^[17]较早地提出将 LOD 引入到推荐系统中。他认为用户的评分信息并不能充分地反映用户的偏好,还应该关注历史数据的语义信息,从而提出将链接开放数据库 DBpedia 引入到音乐推荐中,计算知识库中资源的语义距离来完成推荐任务。首先将已有的音乐数据集利用 SPARQL 语句映射到 DBpedia 中以获取音乐资源更为丰富的属性信息。然后基于 LDSO 过滤无用的属性-值及错误信息。最后利用 LDSO 计算出相关实体的语义距离并利用 SPARQL 语句查询呈现出来。系统的优点在于不需要依靠传统的推荐算法只需一组公开的数据集就可以实现推荐,但是对新用户的问题没有提及,只是对现有用户数据分析做出推荐。

Noia 等^[18]提出利用 LOD 语义丰富的特点来提高基于内容的电影推荐系统的准确率。首先利用 SPARQL 语句将 DBpedia、LinkedMDB、Freebase 中电影相关 RDF 数据抽取出来用于对知识库中资源进行更加详细的描述,并通过向量空间模型(VSM)来计算资源之间的相似性,项目 m_i 和 m_j 的相似性可以标识为

$$\text{sim}^p(m_i, m_j) = \frac{\sum_{n=1}^t w_{n,i,p} w_{n,j,p}}{\sqrt{\sum_{n=1}^t w_{n,i,p}^2} \sqrt{\sum_{n=1}^t w_{n,j,p}^2}}$$

式中: $w_{n,i,p}$ 表示根据 TF-IDF 所得到的权重, TF 为具有 p 属性的节点 n 的频率, IDF 为电影总数与链接到资源 n 的电影数量之间比率的对数。然后根据用户的评分记录为用户建模,用户 u 可以建模为

$$\text{profile}(u) = \{ \langle m_j, v_j \rangle \mid v_j = 1 \text{ if } u \text{ likes } m_j, \\ v_j = -1 \text{ otherwise} \}$$

最后通过公式计算得分为用户进行 top-N 推荐:

$$\tilde{r}(u, m_i) = \frac{\sum_{m_j \in \text{profile}(u)} v_j \cdot \frac{\sum_p \alpha_p \cdot \text{sim}^p(m_j, m_i)}{P}}{|\text{profile}(u)|}$$

Noia 等^[19]总结了 ESWC2014 年利用 LOD 实现书籍推荐系统的相关算法。其中推荐的任务主要包括 3 部分:冷启动环境下的评分预测、基于用户反馈做 top-N 推荐、基于内容的推荐系统的多样性。主要简述数据集的特征以及推荐效果最

佳的方法。数据集是利用 SPARQL 语句处理 DBpedia 获取实验所需信息,数据格式为 userID、itemID、score。推荐效果最佳的算法是 hello-World(一个混合策略的利用 LOD 的推荐系统),考虑数据集中书籍的类型、层级分类、作者信息等,利用基于内容的方法进行推荐,基于余弦相似度进行基于协同过滤的推荐,利用 Wikidata 中数据的链入链出进行基于流行度的推荐。最后通过准确率、召回率以及均方根误差的计算来评判推荐的准确性。

Lu 等^[20]将利用 DBpedia、Geonames 和 Wikidata 中的数据构建了关于全世界的旅游知识图谱。包含 1 200 多个景点类型,从而可以在很细的语义粒度上对景点进行刻画。通过 categories 来表示用户的兴趣,并且为用户在每类 category 上的兴趣度赋予一个得分值。一般采用两种方式: frequency-based 和 time-based。前者根据用户访问某类景点的频率/次数来计算得分值。后者根据用户在某类景点所待的时间长短来计算得分值。文献[20]在这两种情况的基础上,在计算得分值时,把各个景点在所位于的城市中的影响力也考虑进去。例如,如果 A 和 B 都是 casino 类型的景点; A 所位于的城市以 casino 类型的景点闻名于世,但 B 所位于的城市却不是以 casino 类型的景点著称; Mary 游览了 A , Paul 游览了 B ; Mary 和 Paul 游览 casino 类型景点的频率和所花的时间都一样多。则,在考虑了城市的影响力之后得到的结果是: Mary 对 casino 类型景点的喜好程度比 Paul 对 casino 类型景点的喜好程度更高。通过知识图谱中的信息,可以获得不同类型的景点在其所位于的城市中的影响力。综合上面两个方面的优势,与采用简单本体的方法相比,文献[20]的工作可以在更细的语义粒度上刻画景点类型,可以更好地捕获游客的兴趣,最终得到更好的推荐效果。

Oramas 等^[21]研究基于知识图谱的声音和音乐的推荐,主要描述如何创建和利用知识图谱并结合混合推荐引擎来完成推荐任务。文献[21]中构建了声音和音乐两个知识图谱,首先分别从 Free-sound.org、Songfacts.com 以及 Last.fm 中获取标签、文本描述信息以及用户收听和下载的信息,遵循基于本体的方法,利用 Babelify(对于给定的单词返回相关的 wordNet 以及维基百科页面)一种用于实体链接和同义词消歧的最先进的工具,通过将它们与在线数据集中定义的实体关联,丰富从文本描述中提取的标签和关键词。其中识别的命名实体链接到 DBpedia 资源,消歧的单词链

接到 WordNet synsets。最后基于声音和音乐知识图谱利用基于实体的邻居项目匹配和基于路径的邻居项目匹配为用户进行推荐。由于系统通过与外部的知识库关联在很大程度上补充了知识库中的语义信息,使得推荐更加准确、覆盖面更广。

Heitmann 等^[22]提出利用 LOD 构建一个开放的协同过滤推荐系统。传统的推荐算法由于其对数据的需求门槛比较高,这使得新服务提供商难以与领域中的现有推荐服务进行竞争。文献^[22]中主要利用 LOD 来解决获取构建推荐系统数据的问题。实现步骤是:将来自不同来源的用户项目连接的数据集成到通用的词汇表、将数据的表示从 RDF 图转换为用户项目矩阵、在用户项目矩阵中应用特定的协同过滤算法。最后通过一个真实的协同过滤音乐推荐系统,利用 LOD 丰富数据语义信息,在只使用原有数据时平均精度为 2%,平均召回率为 7%,在加入链接数据库之后平均精度为 14%,平均召回率为 33%,证明其准确高效性。

Ostuni 等^[23]更进一步融合 LOD 中隐含的语义反馈信息,提出基于隐式语义反馈的路径算法 (SPrank)。基于路径的特征对数据集进行挖掘,以捕获项目之间的复杂关系。SPrank 的主要思想是探索语义图中的路径,以便找到与用户感兴趣的项目相关的项目。从这些路径的分析中,提取基于路径的特征,并利用随机森林与渐变增强回归树相结合的学习算法来生成推荐列表。

Ristoski 等^[24]提出一种基于 LOD 的混合多策略推荐系统。将基础推荐结果与通用的推荐结果相结合形成最终的推荐结果。基础推荐利用 LOD 中丰富的数据信息结合基于协同过滤和基于内容的推荐,通用推荐是针对全球流行度得分进行推荐,对于所有的用户都是一样的。对于基础推荐使用的是基于内容、项目的协同过滤以及利用从 DBpedia 中获取的丰富的数据信息进行基于内容的推荐,通用推荐使用不同的推荐器进行组合。然后利用 K.M.Ting 等^[25]提出的堆叠模型将基础与通用推荐器结合起来。最后通过计算 F-Score 的值来衡量推荐的准确度,结果表明,本策略在一定程度上提高了推荐的准确率。

总体来看,基于 LOD 的推荐生成技术优点在于:数据关联表达很强、逻辑性强并具备一定的推理能力,可以自动发现一些隐含的语义信息,能够提高推荐准确度。但是,这种方式也存在缺点:由于推荐的过程依赖于外部知识库,因此外部知识库的完备与否对推荐结果的影响巨大,推荐多样性问题以及计算量大等问题有待解决。

2.3 基于图嵌入的推荐生成

随着 Deepwalk^[25]的提出,图嵌入技术开始受到学术界以及工业界的关注。其基本原理是基于随机游走等算法对图中节点进行采样生成节点序列,然后通过神经网络或者是其他机器学习算法将节点序列以及边映射到底维向量空间。图嵌入作为一种更为广泛的图表示学习方法,已经被应用到推荐领域,并且取得了较好的效果。

László 等^[26]较早将图嵌入技术应用于推荐领域。将 MovieLens 中电影与用户嵌入到同一个向量空间,进而计算用户与电影的空间距离,生成推荐列表。Wang 等^[27]做了同样的工作,将医学知识图谱、疾病-患者二部图、疾病-药物二部图分别嵌入低维向量空间,为病患推荐更为安全的药物治疗方式。通过加权平均将知识图谱与二部图结合生成包含更加细粒度属性信息的患者和药物向量。最终生成对给定患者的药物 top-k 列表。

Palumbo 等^[28]针对基于知识图谱的推荐生成提出一种通用的向量化模型 entity2rec。利用一种神经网络模型将知识图谱中的节点依据相连接的属性值训练为向量的模式,通过余弦相似度计算用户项目相关性,进而利用机器学习排序函数进行推荐列表的生成。首先利用 SPARQL 语句将知识图谱分解为属性子图,针对每个属性子图设置目标函数:

$$\max_{x_p} \sum_{e \in K_p} (-\log Z_e + \sum_{n_i \in N(e)} x_p(n_i) \cdot x_p(e))$$

生成特定属性的用户和项目的特征向量,其中:

$$Z_e = \sum_{v \in K_p} \exp(x_p(e) \cdot x_p(v))$$

然后通过:

$$\rho_p(u, i) = \begin{cases} s(x_p(u), x_p(i)), & p = \text{feedback} \\ \frac{1}{|R_+(u)|} \sum_{i' \in R_+(u)} s(x_p(i), x_p(i')), & \text{其他} \end{cases}$$

计算用户项目之间的相关性。最后分别通过机器学习算法 Adrank 和 LambdaMart 生成推荐列表。

3 基于知识图谱的推荐系统关键技术

基于知识图谱的推荐系统综合了知识图谱和传统推荐系统的特点,成为了一个新的研究主题,目前正处于初始发展阶段。在第2章考察相关研究工作的基础上,本节主要对第1章中提出的基于知识图谱的框架中的关键技术和研究现状进行进一步的分析。

3.1 数据采集与预处理

基于知识图谱的推荐系统的实现首先是数据采集,包括用户信息、交互信息、项目信息、上下文信息等。其中,用户信息主要是用户注册信息、交互信息从日志中获取、上下文信息由用户自主录入或者通过传感器获取、项目信息从各大垂直网站及百科网站中获取。在现有文献中,Gradgyenge等^[29]利用 SPARQL 语句从 DBpedia 中抽取音乐相关数据;Dodwad等^[30]从工作门户网站中抽取数据。

3.2 知识图谱的构建

知识图谱构建过程主要包括四个阶段:信息抽取、知识融合、知识加工、知识更新。信息抽取阶段旨在从异构数据源中自动抽取信息得到候选知识单元,相关技术有实体抽取、关系抽取、属性抽取。如Liu等^[31]利用K-最近邻算法和条件随机模型,实现了对Twitter文本数据中实体的识别。知识融合阶段的目的是消除概念的歧义,剔除冗余和错误概念,从而保证知识的质量,它包括实体链接和知识合并。比如苹果既可以代表水果也可以代表苹果公司,必须增加更多的描述信息来加以区别,用户的点击操作可以采用二进制量化表示。文献^[21]利用Babelify,对给定的单词返回相关的wordNet^[32]以及相关的维基百科页面。想要得到结构化、网络化的知识体系,还需要经历知识加工的过程。它包括三方面内容:本体构建、知识推理和质量评估。如Wang等^[33]利用跨语言知识链接方法来构建中英文本体库,使二者相互确认,从而提高了中文与英文关系的准确度。知识图谱的内容也需要与时俱进,因此需要不断的更新迭代,但目前的技术更新起来十分困难^[34]。

3.3 用户偏好获取

推荐系统用户偏好的获取不仅受到项目本身影响还会受到上下文影响。目前关于上下文的定义,还没有一个统一的说法,这里引用王立才等^[12]给出的定义“上下文就是指任何可以用来描述实体情况的信息,其中实体是指用户与应用程序交互相关的客体,也包括用户和应用程序本身”。

用户偏好的获取是为了有针对性地为用户提供服务,为用户提供量身打造的个性化推荐。一般用户偏好的获取分为显式偏好获取和隐式偏好获取。显式偏好获取需要用户提供自身的属性信息或者偏好信息,其精确性与用户的反馈信息和准确度有关。隐式偏好获取是指收集和记录用户历史行为,并采用聚类技术^[35]、神经网络技术^[36]、

关联规则技术^[37]、贝叶斯网络技术^[38]等自动构建用户模型,对用户偏好进行计算。由于用户的兴趣是随着时间不断变化的,偏好也会发生转移,为了获取用户最新的偏好以保证推荐的精确性,需要实时对用户偏好进行更新,主要用到的方法包括动态偏好建模技术、自适应技术等。

3.4 基于知识图谱的推荐

该部分是整个基于知识图谱推荐系统的核心。传统的推荐系统中,例如协同过滤、基于内容的推荐等基本算法同样也可以与基于知识图谱的推荐系统结合。在该过程中,主要是把知识图谱作为一个语义丰富、逻辑推理能力强的数据集融入到推荐生成过程中。本文第2节已经从基于本体、基于LOD以及基于图嵌入的推荐生成3个方面对基于知识图谱的推荐进行了考察。其中不仅涉及到传统的推荐算法,还包括图算法、相似性度量^[39]、贝叶斯网络等相关机器学习算法的应用。尤其是,文献^[24]中结合最短路径和K步马尔科夫链完成推荐任务;文献^[18]中结合基于内容的推荐算法为用户进行推荐;文献^[40]通过结合贝叶斯网络生成推荐列表。

3.5 推荐效用评价

效用评价在整个推荐流程中非常重要。朱郁筱等^[41]对推荐系统的评价指标进行了综述,从准确度、基于排序加权、覆盖率以及多样性指标等4个方面来对系统进行评价。

在基于知识图谱的环境下,推荐系统的多样性受到更多关注。Vargas等^[42]基于多样性提出了3个指标:选择、发现和相关性。

4 基于知识图谱的推荐系统应用进展

基于知识图谱的推荐系统具有两个鲜明的特征:语义丰富、个性化,从而使其在工业领域具有广阔的应用前景。本节就基于知识图谱的推荐系统的应用进展进行总结分析,表1中列举分析了一些典型案例。

1) 旅游推荐

旅游推荐近年来获得越来越多的关注。旅游推荐不仅包括旅游景点的推荐,还涉及周边服务。传统的推荐系统主要以列表的形式向用户展示推荐的内容,而在旅游推荐中,为了使用户更清楚地了解推荐项目的信息,可以选择基于地图的方式来展示推荐内容。典型的案例有:文献^[13]将本体与传统的推荐算法结合为用户做出推荐,首先根据注册信息构建用户画像计算用户兴趣度

以及置信度,然后依据交互行为结合兴趣度和置信度为用户做出推荐,最后结合GIS定位系统利用ICEfaces框架实时更新推荐列表。

表1 基于知识图谱进行推荐的典型案例

Table 1 Typical case for recommendation based on knowledge graph

应用领域	典型的推荐系统	知识图谱	关键技术
旅游推荐	TARes ^[20]	本体	K-means、基于内容的推荐、基于协同过滤的推荐技术。
	SMRes ^[21]	LOD	TF-IDF、相似度算法
电影/音乐	CoLRes ^[18]	LOD	基于实体的邻居项目匹配、基于路径的邻居项目匹配
	ERRes ^[29]	LOD	基于内容的推荐技术、遗传算法、VSM(向量空间模型)
电子商务	CBDORes ^[22]	LOD	传播激活、网络科学
职位推荐	SigTur/E-Destination ^[16]	本体	重排序、相似度算法

2) 电影、音乐推荐

电影、音乐推荐在基于知识图谱的推荐系统中是一个热点,主要的原因在于传统的推荐系统数据集是以音乐以及电影为主。例如,CoLRes^[18]较早的将知识图谱用于音乐推荐领域,重点是结合外部链接数据库丰富语义信息。SMRes^[21]则将音乐推荐分为两部分歌曲音乐家的推荐以及音乐。制作人的推荐通过在Songfacts3和Last.fm4收集用户反馈信息以及歌曲数据集,通过基于实体的邻居项目匹配、基于路径的邻居项目匹配为用户生成推荐结果。

3) 电子商务

用户在网上购买书籍、服务、食品、衣服等项目时,没有足够的精力去检索、了解项目的详细信息,并且随着季节、天气等条件的改变,用户的需求也会有相应的改变。基于知识图谱的推荐系统可以融合多源信息,更加充分的挖掘用户和商品之间的潜在关联,将为用户的最终决策提供更有效的支持,这使得电子商务成为基于知识图谱的推荐系统的主要应用场合。目前国内外相关企业(如亚马逊、阿里巴巴等)已经构建知识图谱并应用于各自的电子商务推荐系统中。Gradgyenge等^[29]尝试将知识图谱应用于电子邮件销售系统。

4) 职位推荐

随着互联网上数据的爆炸式增长,网络上的信息多而繁杂。对于求职者筛选自己心仪的职位

来说比较困难。推荐系统作为处理信息爆炸问题的手段之一被用于职位推荐中,可以快速定位用户心仪的职位,帮助用户节省时间。CBDORes^[15]中利用基于知识图谱的推荐方法为求职者根据自身情况生成职位推荐。目前基于知识图谱的推荐系统面临许多问题和挑战,使其与具体的商业应用接轨还存在距离。近年来,人工智能、深度学习、智慧旅游等新技术和应用需求的快速涌现,为基于知识图谱的推荐系统的应用研究与实践提供了良好的机会。以深度学习为例,它能够依据大数据学习项目以及用户特征,将基于知识图谱的推荐系统与深度学习结合,将提高推荐准确度。此外,广告仍然是互联网企业目前主要的盈利模式之一,为相应的用户展示不同的广告,具有鲜明的商业价值,极有可能会成为基于知识图谱的推荐系统的应用实践方向之一。

5 基于知识图谱推荐的重点和难点

1) 对用户偏好的精准建模问题。知识图谱可以融合多源异构信息,包括丰富的用户信息及用户与项目的交互信息;如何全面考虑用户与用户、用户与项目之间的联系,进一步提高推荐的准确度,是未来的研究重点。

2) 对大数据的高效处理问题。由于数据生成速度快,而且多源数据融合过程中引入了更多的噪声和冗余,使得传统推荐算法中采用的精确计算方式越来越难以应对,概率模型统计方法将可能发挥更为重要的作用^[43]。

3) 推荐系统的效用评价问题。对传统推荐系统的评价指标主要是准确度和效率,通常采用的方法有MAE、RMSE、关联度等。这些指标当然也可以用于基于知识图谱的推荐系统,例如文献^[16]采用了准确率和召回率来表征推荐系统的准确度。然而,由于不同领域的推荐系统面临的实际问题不同,相应地使用了不同的数据集,使得难以形成统一的效用评价标准^[44]。

4) 如何通过深度挖掘和相关性知识发现提高推荐效果。由于知识图谱本身的特点,相对于因果关系挖掘,其更适合于相关关系挖掘。通过从知识图谱中深度挖掘项目之间、用户之间、以及项目与用户间的深层次关系,获得更多的相关性结果,有助于为用户进行个性化推荐,在提高推荐结果多样性的同时保持较高的推荐准确性^[45]。

5) 推荐系统的可扩展性。可扩展性一直是推荐系统应用中面临的难点问题^[46]。在传统推荐系统中,随着用户与项目数量的增加,会使得计算

量显著增大。目前主要的解决方法有聚类、数据集缩减、降维等。在知识图谱中,用户之间、用户与项目之间随时随地的信息互动会产生大规模的数据集;将这些数据集融入知识图谱后会急剧增加计算复杂度。如何将用户之间产生的数据动态、快速地融入知识图谱并提高推荐系统的可扩展性,是未来需要解决的难点问题。

6) 社交网络^[47]信息的充分利用问题。传统的推荐系统假设用户之间是独立同分布的,忽略了用户之间的社会交互及联系。知识图谱中用户与用户、用户与项目、项目与项目之间都存在一定的联系,用户不再是单一的用户需要进一步挖掘彼此之间的联系以发现更多的隐含信息。因此结合社交网络关系^[48]也是今后的一个研究重点。

6 结束语

本文对基于知识图谱的推荐系统研究现状进行了较为全面的综述。对于基于本体的推荐生成、基于 LOD 的推荐生成以及基于图嵌入的推荐生成 3 类方法,都取得了较好的推荐效果。由于每种方法各有优缺点,将多种方法综合起来使用是一条有效途径。本文在综合考察这些工作的基础上,给出了基于知识图谱的推荐系统总体框架。

由于知识图谱只有在达到一定规模时才能发挥较好的效果,因此在利用知识图谱进行推荐之前还存在许多难点问题有待解决。典型的问题包括多源信息融合、知识图谱中用户特征的挖掘及潜在需求的分析、多维度推荐、与时间空间推理的结合、用户隐私保护等。此外,近年来兴起的深度学习^[49]、表示学习^[50]、位置感知等也为基于知识图谱的推荐系统提供了许多可能的改进方向,值得进一步研究。

参考文献:

- [1] XIONG Haitao, LIU Zhengbin. A situation information integrated personalized travel package recommendation approach based on TD-LDA model[C]//2015 International Conference on Behavioral, Economic and Socio-cultural Computing (BESC). Nanjing, China: IEEE, 2015: 32–37.
- [2] 常亮, 曹玉婷, 孙文平, 等. 旅游推荐系统研究综述[J]. 计算机学报, 2017, 44(10): 1–6.
CHANG Liang, CAO Yuting, SUN Wenping, et al. Review of tourism recommendation system[J]. *Computer science*, 2017, 44(10): 1–6.
- [3] 孟祥武, 纪威宇, 张玉杰. 大数据环境下的推荐系统[J]. 北京邮电大学学报, 2015, 38(2): 1–15.
MENG Xiangwu, JI Weiyu, ZHANG Yujie. A survey of recommendation systems in big data[J]. *Journal of Beijing university of posts and telecommunications*, 2015, 38(2): 1–15.
- [4] 刘屹, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582–600.
LIU Yu, LI Yang, DUAN Hong, et al. Knowledge graph construction techniques[J]. *Journal of computer research and development*, 2016, 53(3): 582–600.
- [5] LUBERG A, TAMMET T, JÄRV P. Smart city: A rule-based tourist recommendation system[M]//LAW R, FUCHS M, RICCI F. *Information and Communication Technologies in Tourism 2011*. Vienna, Austria: Springer, 2011: 51–62.
- [6] TONG Rong, XUE Lijuan, WANG Haofen, et al. Building and exploring an enterprise knowledge graph for investment analysis[M]//GROTH P, SIMPERL E, GRAY A, et al. *The Semantic Web-ISWC 2016*. Cham, Germany: Springer International Publishing, 2016: 418–436.
- [7] 漆桂林, 高桓, 吴天星. 知识图谱研究进展[J]. 情报工程, 2017, 3(1): 4–25.
QI Guilin, GAO Huan, WU Tianxing. The research advances of knowledge graph[J]. *Technology information engineering*, 2017, 3(1): 4–25.
- [8] SZEKELY P, KNOBLOCK C A, SLEPICKA J, et al. Building and using a knowledge graph to combat human trafficking[C]//2015 International Semantic Web Conference on. New York, USA: Springer-Verlag, 2015: 205–221.
- [9] 唐晓波, 魏巍. 基于本体的推荐系统研究综述[J]. 图书馆学研究, 2016(18): 7–12, 58.
TANG Xiaobo, WEI Wei. A survey of ontology-based recommendation systems[J]. *Library science studies*, 2016(18): 7–12, 58.
- [10] LI Dongsheng, LV Qin, XIE Xing, et al. Interest-based real-time content recommendation in online social communities[J]. *Knowledge-based systems*, 2012, 28: 1–12.
- [11] 印鉴, 王智圣, 李琪, 等. 基于大规模隐式反馈的个性化推荐[J]. 软件学报, 2014, 25(9): 1953–1966.
YIN Jian, WANG Zhisheng, LI Qi, et al. Personalized recommendation based on large-scale implicit feedback[J]. *Journal of software*, 2014, 25(9): 1953–1966.
- [12] 王立才, 孟祥武, 张玉洁. 上下文感知推荐系统[J]. 软件学报, 2012, 23(1): 1–20.
WANG Licai, MENG Xiangwu, ZHANG Yujie. Context-aware recommendation systems[J]. *Journal of software*, 2012, 23(1): 1–20.
- [13] NIARAKI A S, KIM K. Ontology based personalized route planning system using a multi-criteria decision making approach[J]. *Expert systems with applications*, 2009,

- 36(2): 2250–2259.
- [14] DODWAD P R, LOBO L. A context-aware recommender system using ontology based approach for travel applications[J]. *International journal of advanced engineering and nano technology*, 2014, 1(10): 8–12.
- [15] KETHAVARAPU U P K, SARASWATHI S. Concept based dynamic ontology creation for job recommendation system[J]. *Procedia computer science*, 2016, 85: 915–921.
- [16] MORENO A, VALLS A, ISERN D, et al. SigTur/E-destination: Ontology-based personalized recommendation of tourism and leisure activities[J]. *Engineering applications of artificial intelligence*, 2013, 26(1): 633–651.
- [17] PASSANT A. dbrec-music recommendations using DBpedia[M]//PATEL-SCHNEIDER P F, PAN Yue, HITZLER P, et al. *The Semantic Web-ISWC 2010*. Berlin Heidelberg, Germany: Springer, 2010: 209–224.
- [18] DI NOIA T, MIRIZZI R, OSTUNI V C, et al. Linked open data to support content-based recommender systems[C]//*Proceedings of the 8th International Conference on Semantic Systems*. Graz, Austria: ACM, 2012: 1–8.
- [19] DI NOIA T, CANTADOR I, OSTUNI V C. Linked open data-enabled recommender systems: ESWC 2014 challenge on book recommendation[M]//PRESUTTI V, STANKOVIC M, CAMBRIA E, et al. *Semantic Web Evaluation Challenge*. Cham, Germany: Springer International Publishing, 2014: 129–143.
- [20] LU Chun, LAUBLET P, STANKOVIC M. Travel attractions recommendation with knowledge graphs[C]//*European Knowledge Acquisition Workshop*. Bologna, Italy: Springer, 2016.
- [21] ORAMAS S, OSTUNI V C, DI NOIA T, et al. Sound and music recommendation with knowledge graphs[J]. *ACM transactions on intelligent systems & technology*, 2017, 8(2): 21.
- [22] HEITMANN B, HAYES C. Using linked data to build open, collaborative recommender systems[C]//*Linked Data Meets Artificial Intelligence*. Stanford, California, USA: AAAI, 2010.
- [23] OSTUNI V C, DI NOIA T, DI SCIASCIO E, et al. Top-N recommendations from implicit feedback leveraging linked open data[C]//*Proceedings of the 7th ACM Conference on Recommender Systems*. Hong Kong, China: ACM, 2013: 85–92.
- [24] RISTOSKI P, MENCIA E L, PAULHEIM H. A hybrid multi-strategy recommender system using linked open data[C]//*Semantic Web Evaluation Challenge*. Cham, Germany: Springer, 2014, 475: 150–156.
- [25] TING K M, WITTEN I H. Issues in stacked generalization[J]. *Artificial intelligence research*, 1999, 10(1).
- [26] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: On-line learning of social representations[C]//ACM, 2014, 701–710.
- [27] GRAD-GYENGE L, KISS A, FILZMOSER P. Graph embedding based recommendation techniques on the knowledge graph[C]//*Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*. Bratislava, Slovakia: ACM, 2017: 354–359.
- [28] WANG Meng, LIU Mengyue, LIU Jun, et al. Safe medicine recommendation via medical knowledge graph embedding[EB/OL]. arXiv: 1710.05980, 2017.
- [29] 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*. Como, Italy: ACM, 2017: 32–36.
- [30] GRAD-GYENGE L, FILZMOSER P. Recommendation Techniques on a Knowledge Graph for Email Remarketing[C]//*eKNOW 2016 The Eighth International Conference on Information, Process, and Knowledge Management*. Venice, Italy: IARIA, 2016.
- [31] LIU Xiaohua, ZHANG Shaodian, WEI Furu, et al. Recognizing named entities in tweets[C]//*49th Annual Meeting of the Association for Computational Linguistics; Human Language Technologies*. Stroudsburg, PA, USA: ACL, 2011: 359–367.
- [32] FELLBAUM C. WordNet[M]//*The Encyclopedia of Applied Linguistics*. Blackwell: Blackwell Publishing Ltd, 2012: 231–243.
- [33] WANG Zhigang, LI Juanzi, LI Shuanjie, et al. Cross-lingual knowledge validation based taxonomy derivation from heterogeneous online wikis[C]//*28th Conference on Artificial Intelligence*. Menko Park, USA: AAAI, 2014: 180–186.
- [34] DESHPANDE O, LAMBA D S, TOURN M, et al. Building, maintaining, and using knowledge bases: A report from the trenches[C]//*the 32nd ACM SIGMOD International Conference on Management of Data*. New York, USA: ACM, 2013: 1209–1220.
- [35] ZHANG Daqiang, HSU C H, CHEM Min, et al. Cold-start recommendation using bi-clustering and fusion for large-scale social recommender systems[J]. *IEEE transactions on emerging topics in computing*, 2014, 2(2): 239–250.
- [36] HSIEH K L. Employing a recommendation expert system based on mental accounting and artificial neural networks into mining business intelligence for study abroad's P/S recommendations[J]. *Expert systems with applications*, 2011, 38(12): 14376–14381.

- [37] 薛福亮, 马莉. 利用动态产品分类树改进的关联规则推荐方法[J]. *计算机工程与应用*, 2016, 52(4): 135–141.
XUE Fuliang, MA Li. Improved association rule recommendation method based on dynamic product taxonomy[J]. *Computer engineering and applications*, 2016, 52(4): 135–141.
- [38] LU E H C, FANG S H, TSENG V S. Integrating tourist packages and tourist attractions for personalized trip planning based on travel constraints[J]. *GeoInformatica*, 2016, 20(4): 741–763.
- [39] LI Ting, LIU Anfeng, HUANG Changqin. A similarity scenario-based recommendation model with small disturbances for unknown items in social networks[J]. *IEEE Access*, 2017, 4: 9251–9272.
- [40] GRÜN C, NEIDHARDT J, WERTHNER H. Ontology-based matchmaking to provide personalized recommendations for tourists[M]//SCHEGG R, STANGL B. *Information and Communication Technologies in Tourism 2017*. Cham: Springer International Publishing, 2017.
- [41] 朱郁筱, 吕琳媛. 推荐系统评价指标综述[J]. *电子科技大学学报*, 2012, 41(2): 163–175.
ZHU Yuxi, LÜ Linyuan. Evaluation metrics for recommender systems[J]. *Journal of university of electronic science and technology*, 2012, 41(2): 163–175.
- [42] VARGAS S, CASTELLS P. Rank and relevance in novelty and diversity metrics for recommender systems[C]//5th ACM Conference on Recommender Systems. Chicago, USA: ACM, 2011: 109–116.
- [43] CATHERINE R, COHEN W. Personalized recommendations using knowledge graphs: A probabilistic logic programming approach[C]//10th ACM Conference on Recommender Systems. Boston, USA: ACM, 2016: 325–332.
- [44] LI Haoyang, WU Yuanxu, XIA Wei. Review on state-of-the-art technologies and algorithms on recommendation system[C]//2016 International Conference on Mechatronics Engineering and Information Technology. 2016.
- [45] MAYER-SCHNBERGER V, CUKIER K. Big data: A revolution that will transform how we live, work, and think[M]. Eamon Dolan/Houghton Mifflin Harcourt, 2014.
- [46] KETHAVARAPU U P K, SARASWATHI S. Concept based dynamic ontology creation for job recommendation system[J]. *Procedia computer science*, 2016, 85: 915–921.
- [47] OSTUNI V C, DI NOIA T, MIRIZZI R, et al. A linked data recommender system using a neighborhood-based graph kernel[M]//International Conference on Electronic Commerce and Web Technologies. Cham: Springer International Publishing, 2014: 89–100.
- [48] FENG Xiaodong, SHARMA A, SRIVASTAVA J, et al. Social network regularized sparse linear model for Top-N recommendation[J]. *Engineering applications of artificial intelligence*, 2016, 51: 5–15.
- [49] LIU Juntao, WU Caihua. Deep learning based recommendation: a survey[M]//International Conference on Information Science and Applications 2017. Singapore: Springer, 2017.
- [50] WANG Quan, MAO Zhendong, WANG Bin, et al. Knowledge graph embedding: A survey of approaches and applications[J]. *IEEE transactions on knowledge and data engineering*, 2017, 29(12): 2724–2743.

作者简介:



常亮, 男, 1980 年生, 教授, 博士, 中国计算机学会高级会员, 主要研究方向为数据与知识工程、形式化方法、智能系统。主持并完成多项科研项目, 其中国家自然科学基金项目 1 项、广西自然科学基金项目 1 项。发表学术论文 70 余篇, 被 SCI、EI 收录 60 余篇。



张伟涛, 男, 1993 年生, 硕士研究生, 主要研究方向为机器学习、推荐系统。



古天龙, 男, 1964 年生, 教授, 博士生导师, 博士, 主要研究方向为形式化方法、知识工程与符号推理、协议工程与移动计算、可信泛在网络、嵌入式系统。主持国家 863 计划项目、国家自然科学基金、国防预研重点项目、国防预研基金等 30 余项。发表学术论文 130 余篇, 被 SCI、EI 收录 60 余篇, 出版学术著作 3 部。