

《智能信息处理》课程作业

基于知识图谱的推荐系统研究

邓杰进

作业	分数[70]
得分	

2020 年 12 月 6 日

基于知识图谱的推荐系统研究

邓杰进

(大连海事大学计算机科学与技术辽宁省大连市中国 116026)

摘要 推荐系统是一种信息过滤系统,用于预测用户对物品的“评分”或“偏好”。^[1]推荐系统为用户推荐个性化的在线商品或信息,广泛分布在 Web 场景中,在立足于大数据分析的基础之上,根据用户的喜好倾向进行智能推荐,在提升产品销量的同时,也能给用户带来良好的消费体验。在过去的几十年里,推荐系统得到了巨大的发展,推荐系统的性能也在不断提高。但是推荐系统中仍然面临一些关键问题,例如:数据稀疏、冷启动等问题。由于近几年来知识图谱的兴起,为推荐系统的改进提供了新的途径。基于知识图谱的推荐系统,可以发挥知识图谱辅助信息优势,收集相关信息数据,提供辅助信息来源,使推荐系统更加精准、可靠。

关键词 知识图谱;推荐系统;异质信息网络;协同过滤

中图法分类号 TP311.20 DOI 号 10.3969/j.issn.1001-3695.2014.01.030

Research on recommendation System based on Knowledge Map

Dengjiejin

(Computerscienceandtechnology, Dalianmaritimeuniversity,LiaoningDalian,116026,China)

Abstract A recommendation system is an information filtering system that predicts a user's "rating" or "preference" for an item. [1] The recommendation system recommends personalized online goods or information for users, which is widely distributed in Web scenarios. Based on big data analysis, intelligent recommendation is carried out according to users' preferences and tendencies, which can not only improve product sales, but also bring good consumption experience to users. In the past few decades, the recommendation system has been greatly developed, and the performance of the recommendation system is constantly improving. However, the recommendation system still faces some key problems, such as sparse data and cold startup. Due to the rise of knowledge map in recent years, it provides a new way to improve the recommendation system. The recommendation system based on knowledge map can give full play to the advantages of knowledge map auxiliary information, collect relevant information data and provide auxiliary information sources, so as to make the recommendation system more accurate and reliable.

Keywords knowledge map; Recommendation system; Heterogeneous information network; Collaborative filtering

1. 引言

随着互联网的飞速发展,我们正处于大数据时代,面对海量的数据信息,用户难以从中快速有效地汲取所需要的信息。随着推荐系统的不断发展,推荐系统能够通过对历史用户记录进行分析,从而判断用户当前的状态,从而为用户提供良好、准确、符合人性化的体检。特别是在结合知识图谱以后,

改善了冷启动和稀疏性问题,将辅助信息作为输入,为大数据下的产品推荐系统设计提供了有力的途径。基于知识图谱的商品推荐系统,收集、分析和处理信息,推测用户的喜好、习惯,针对性推荐,帮助用户选择合适的产品,快速决策,服务质量也将大大提升。

2. 知识图谱

知识图谱这一概念最初是由 Google 公司提出, 最初的目的是用来提高搜索引擎的搜索质量, 增强用户的搜索体验。知识图谱本质上是一种语义网络。其节点代表实体(entity)或者概念(concept), 边代表实体或概念之间的关系, 可以帮助形式化描述、理解现实世界的事物及其相互关系。随着互联网技术的快速发展, 知识图谱的应用也从搜索引擎逐步扩展到各个领域。例如推荐系统、智能问答和文本分析^[2]。知识图谱通过对错综复杂的文档数据进行有效的加工、处理、整合, 转化为简单、清晰的“实体, 关系, 实体”的三元组, 最后聚合大量知识, 从而实现知识的快速响应和推理。一个知识图谱由若干个三元组(h, r, t)组成, 其中 h 和 t 为实体代表一条关系的头节点和尾节点, r 代表关系。如图 1, 吴京和战狼即为实体, 导演为关系。

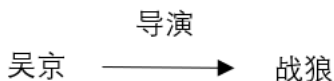


图 1

知识图谱有自顶向下和自底向上两种构建方式。所谓自顶向下构建试借助百度百科类网站等结构化数据源, 从高质量数据中提取本体和模式信息, 加入到知识库; 所谓自底向上构建, 则是借助一定的技术手段, 从公开采集的数据中提取出资源模式, 选择其中置信度较高的新模式, 经人工审核之后, 加入知识库中。

3. 推荐系统

推荐系统这一思想在 1994 年被提出, 并且逐步成为一个相对独立的研究方向, 现在商业方面的应用也非常成熟。推荐系统就是利用电子商务网站向客户提供商品信息和建议, 帮助用户解决应该购买什么商品。模拟销售人员帮助客户完成购买过程。个性化推荐是根据用户的兴趣特点和购买行为, 向用户推荐用户感兴趣的信息和商品。随着电子商务规模的不断扩大, 商品个数和种类快速增长, 顾客需要花费大量时间在海量信息中搜寻商品。为了解决这些问题, 个性化推荐系统应运而生。个性化推荐系统是建立在海量数据挖掘的基础之上, 为顾客提供个性化的决策支持和信息服务。在

真实场景中推荐系统也存在一些问题, 例如淘宝中每日的活跃用户上千万, 在线商品上亿, 产生的交互量也非常庞大, 但是对于观察到的交互矩阵仍然稀疏。另外推荐系统在冷启动方面也没有一个很好的解决方案。

目前的推荐系统主要分为 3 类, 分别为基于内容的推荐算法、基于协同过滤的推荐算法和混合推荐算法。

基于内容的推荐算法。基于内容的推荐根据用户选择的对象, 推荐其它类似的对象。一本通过信息检索方式, 来得到用户的偏好。一般分为三个步骤: 1) 为每个待推荐对象构建一个对象的属性资料; 2) 为每个用户构建一个用户偏好资料; 3) 计算用户偏好资料与待推荐对象属性资料的相似度。^[8]

基于协同过滤的推荐算法。基于协同过滤的推荐算法是推荐系统的重要组成部分, 在推荐系统中也是使用最多的。从运行方式上来看, 协同过滤主要分为基于近邻的协同过滤和基于模型的协同过滤。基于近邻的推荐是通过研究系统的评分直接对项目进行排序并做出决策, 而基于模型的方法可以根据评分记录学习一个模型。^[9]

混合推荐算法是指将多种推荐技术进行结合相互弥补缺点, 得到一个更好的推荐算法。最常见的混合推荐算法就是将协同过滤推荐算法与其它推荐算法相结合, 从而解决冷启动和稀疏性问题^[10]。虽然混合推荐理论上不会比任何一个单一算法的推荐差, 但是使用混合推荐算法的复杂度相应提高了。

4. 基于知识图谱的推荐

知识图谱包含了实体之间丰富的语义关联。为推荐系统提供了潜在的辅助信息来源。知识图谱在诸多推荐场景中都有应用的潜力, 例如: 电影、新闻、景点、餐馆、购物等。和其它种类的辅助信息相比, 知识图谱的引入可以然推荐结果更加具有以下特征:

精确性。知识图谱为物品引入了更多的语义关系, 可以深层次地发现用户的兴趣^[3]。如图 2。

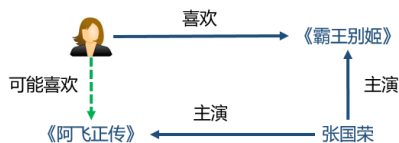


图 2

多样性。知识图谱提供了不同的关系连接种类,有利于推荐的发散。避免推荐结果限于单一类型^[3]。如图 3。



图 3

可解释性。知识图谱可以连接用户的历史记录和推荐结果,从而提高用户对推荐结果的满意度和接受度,增强用户对推荐系统的信任^[6]。由图二和图三可知,用户喜欢《霸王别姬》也有可能喜欢《阿飞正传》,因为它们都有相同的主演;用户也有可能喜欢《末代皇帝》,因为它们是相同题材的电影;用户也有可能喜欢《搜索》,因为它们的导演相同。

现在的基于知识图谱的推荐算法可以分为两类,

将知识图谱引入推荐系统,一般是基于路径的推荐和基于嵌入的推荐。基于路径的推荐,比较代表性的是 meta-path 和 meta-graph。该方法将知识图谱视为一个异质信息网络 (heterogeneous information network)^[5],然后构造物品之间的基于 meta-path 或 meta-graph 的特征。meta-path 是连接两个实体的一条特定路径,例如“演员→电影→导演→电影→演员”,这条 meta-path 可以连接两个演员,因此可以看作挖掘演员之间的潜在关系的方式。这种方法可以充分且直观地利用知识图谱中的网络结构。但是这种方法同样存在缺点,meta-path 或 meta-graph 需要手动设计,很难达到最优;同时该方法无法在实体不属于同一个场景的场景中应用,因为我们无法为这样的场景预定义 meta-path 或 meta-graph。基于嵌入的推荐,比较代表性的是基于异质信息网络的图嵌入。因为知识图谱中节点和边的不同类型,可以称为一种异质信息网络图,所以可以使用一些异质信息网络的嵌入方法对图上的实体和信息进行表示。主要方法有 Metapath2Vec、node2vec entity2vec。

5. 基于知识图谱的推荐的关键技术^[3]

5.1 数据采集和预处理技术

知识图谱下的推荐系统,在具体运行中,第一个环节就是采集数据信息,涵盖范围较广,包括上下文信息、用户信息、项目信息和交互信息等。

5.2 知识图谱构建

构建知识图谱的过程,需要四个步骤,即抽取信息、融合知识、加工知识和更新知识。在此基础上,结合用户喜好和偏向提供更具有个性化的服务,使推荐的针对性更强。

5.3 知识图谱的推荐

以往推荐系统中的内容推荐和协同过滤算法,可以与知识图谱下的推荐系统有机结合。进一步丰富知识图谱的语义信息将更多逻辑性较强的数据集整合到推荐系统中。

5.4 推荐的有效评价

在推荐系统中,评价指标非常重要,对于一个推荐系统是否有用、效果是否好,主要通过各项指标进行评价。通常依据覆盖率、召回率、排序加权、准确性和多样性全面评价推荐系统。

6. 知识图谱在推荐系统的应用

知识图谱与推荐系统相结合,具体可以有以下三种形式。

依次学习。知识图谱特征学习可以获得关系向量与实体向量,引入到推荐系统中,然后可以逐步学习用和产品向量和物品向量。如图 4。

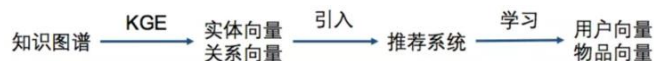


图 4. 依次学习

联合学习。将知识图谱特征学习与推荐系统有机得结合在一起,然后通过端到端的形式联合学习,从实体向量、关系向量、用户向量、物品向量等。如图 5。

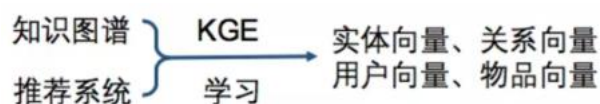


图 5. 联合学习

交替学习。知识图谱特征学习和推荐系统作为分离、相关的任务,立足于多任务学习的基础上实现交替学习的目的。知识图谱与推荐系统本质上是分离的,但同时也是存在关联的任务,基于多任务学习框架可以实现二者信息相互补充。如图 6。

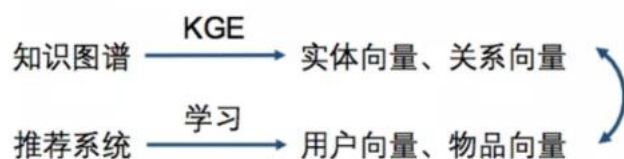


图 6. 交替学习

7. 结束语

将知识图谱引入推荐系统中, 可以看到将知识图谱应用于推荐系统所带来的优势, 可以充分利用辅助信息, 将辅助信息作为输入, 丰富输入信息, 解决交互数据稀疏性的问题, 从而改善推荐的效果。并且一定程度上解决冷启动等问题, 从而提高推荐的准确性和多样性, 给用户带来更好的体验效果。

8. 参考文献

- [1] Facebook, Pandora Lead Rise of Recommendation Engines - TIME. *TIME.com*. 27 May 2010 [1 June 2015].
- [2] 秦川, 祝恒书, 庄福振, 郭庆宇, 张琦, 张乐, 王超, 陈恩红, 熊辉. 基于知识图谱的推荐系统研究综述[J]. 中国科学: 信息科学, 2020, 50 (07): 937-956.
- [3] 韩慧. 基于知识图谱的商品推荐系统[J]. 信息通信, 2020 (06): 200-201.
- [4] 常亮, 张伟涛, 古天龙, 孙文平, 宾辰忠. 知识图谱的推荐系统综述[J]. 智能系统学报, 2019, 14 (02): 207-216.
- [5] 石//, 王睿嘉, 王啸. 异质信息网络分析与应用综述.
- [6] 张志明, 张一帆. 基于 B2C 电子商务网站的个性化商品推荐系统研究[J]. 厦门科技, 2014 (05): 45-47.
- [7] 王鸿伟. 基于网络特征学习的个性化推荐系统[D]. 上海交通大学, 2018.
- [8] 骆亮. 基于内容推荐算法和余弦相似度算法的领导决策辅助信息系统[J]. 广西科学院学报, 2018, 34 (02): 143-150.
- [9] 纪平, 胡学友, 杨文娟, 刘学亮. 基于矩阵分解的协同过滤推荐算法[J]. 合肥学院学报 (综合版), 2020, 37 (05): 10-18.
- [10] Bellogín A, Cantador I, Díez F, et al. An empirical comparison of social, collaborative filtering, and hybrid recommenders. *ACM Trans Intell Syst Technol*, 2013, 4: 1-29