

## 《智能信息处理》课程考试

# 融合知识图谱的推荐算法概述

张贺

考核	到课 [10]	作业 [20]	考试 [70]	课程成绩 [100]
得分				

2020 年 12 月 9 日

# 融合知识图谱的推荐算法概述

张 贺

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

**摘 要** 当前的时代是信息爆炸的时代,信息爆炸使得信息的利用率降低,这种现象被称为信息超载。推荐系统作为一种信息过滤的重要手段,是当前解决信息超载问题的最有效的方法之一,是面向用户的互联网产品的核心技术。推荐系统已经广泛应用在实际生活中的很多场景,但是仍然面临着一些问题,如数据稀疏、冷启动等问题。近年来,利用知识图谱作为辅助信息来增强推荐已经引起了人们相当大的兴趣,这种方法可以缓解上述问题,进行更准确的个性化推荐。本文概述了几种最新的融合知识图谱的推荐算法。

**关键词** 推荐系统; 知识图谱; 数据稀疏; 冷启动

## Overview of recommendation algorithm combining knowledge graph

Zhang He

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)

**Abstract** The current era is the era of information explosion. Information explosion reduces the utilization rate of information. This phenomenon is called information overload. As an important means of information filtering, the recommendation system is one of the most effective methods to solve the problem of information overload. It is the core technology of user-oriented Internet products. The recommendation system has been widely used in many scenarios in real life, but it still faces some problems, such as data sparseness, cold start, etc. In recent years, the use of knowledge graphs as auxiliary information to enhance recommendations has attracted considerable interest. This method can alleviate the above-mentioned problems and make more accurate personalized recommendations. This article outlines several latest recommendation algorithms combining knowledge graphs.

**Keywords** Recommendation system; knowledge graph; data sparseness; cold start

## 1 引言

随着信息技术和互联网的高速发展,人们走入了信息过载的时代,在这个时代,作为信息消费者,从海量信息中找到自己感兴趣的信息是一件非常困难的事情。推荐系统可以代替用户对信息进行初步的评估,通过预测向用户呈现他们可能感兴趣的信息。

传统的推荐系统只使用用户和物品的历史交互信息作为输入,这会带来两个问题,一是在实际场景中,用户和物品的交互信息往往是非常稀疏的,即存在数据稀疏问题。例如,一个电影类APP可能包含了上万部电影,然而一个用户打过的电影可能平均只有几十部。使用如此少量的

已观测数据来预测大量的未知信息,会极大地增加算法的过拟合风险;另外,对于新加入的用户或者物品,由于系统没有其历史交互信息,因此无法进行准确地建模和推荐,这种情况也叫做冷启动问题。

解决稀疏性和冷启动问题的一个常见思路是在推荐算法中额外引入一些辅助信息作为输入。辅助信息可以丰富对用户和物品的描述、增强推荐算法的挖掘能力,从而有效地弥补交互信息的稀疏或缺失。

知识图谱作为一种结构化辅助信息,由于其储存了用户、物品、属性之间的关系,蕴含着丰富的语义,所以能提高推荐结果的准确性、多样性和可解释。近年来,融合知识图谱的推荐算法研究取得了显著成果。本文概述了几种最新发表

的融合知识图谱的推荐算法。

## 2 知识图谱



图 1. 知识图谱

近几年，知识图谱作为一种新兴类型的辅助信息引起了众多研究人员的关注。知识图谱旨在描述真实世界中存在的各种实体或概念。知识图谱<sup>[1]</sup>是一种语义网络，由节点和边组成的网状知识结构，每个节点表示现实世界中存在的“实体”，每条边为实体与实体之间的“关系”。一个知识图谱应用的基础是知识表达，知识表达由若干个（实体-关系-实体）三元组以及实体及其属性值对实现。

实体：现实世界的语义本体；

关系：对应本体间的关系，连接了不同类型的实体；

属性：描述一类实体的特性，实体被属性所标注。

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。知识图谱的引入让推荐结果更具精确性、多样性、可解释性。知识图谱为物品引入了更多的语义关系，通过语义关系可以更深层次地发现用户兴趣。

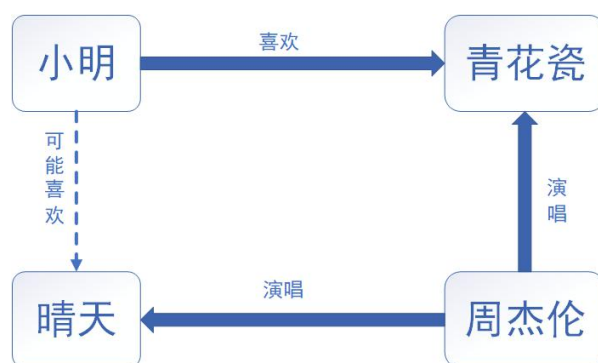


图 2. 发现用户潜在兴趣

知识图谱中同一实体可能存在不同的关系连接，这会使推荐结果更发散，避免推荐结果过于单一。

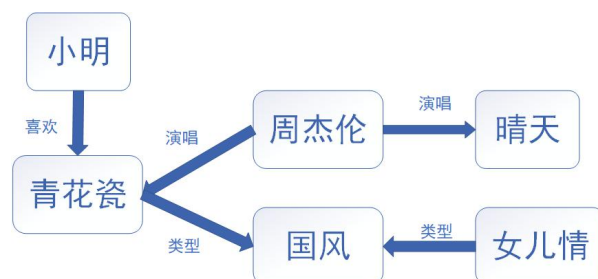


图 3. 知识图谱提供不同的关系连接种类

知识图谱可以连接用户的历史记录，给用户推荐其更满意的推荐结果，使用户更加信任推荐系统。依据知识图谱可以预测小明可能喜欢“女儿情”这首歌，因为它和“青花瓷”同属于国风类型；也可以预测小明可能喜欢“晴天”，因为它和“青花瓷”都是周杰伦的原唱歌曲。

## 3 融合知识图谱的推荐算法

本节概述了几种最新的融合知识图谱的推荐算法研究成果。它们都表明，知识图谱作为辅助信息可以很好的增强推荐效果。

### 3.1 融合知识图谱信息和短期偏好的推荐算法

由于知识图谱学习任务的目标是还原知识图谱中三元组的关系，其构建并非是以推荐任务为目标的，导致了知识图谱学习任务很难高效的帮助推荐任务提升推荐性能，另外，用户兴趣易被短期的环境和心情所影响，针对以上两点文献[2]

中提出了一种融合知识图谱信息和短期偏好的推荐模型。

该模型首先通过 RippleNet 算法提取用户和知识图谱实体关系对，然后将这些关系对按照知识图谱三元组的形式存储和参与训练；采用基于注意力机制的双向 GRU 网络从用户近期交互的物品序列中提取用户的短期偏好；之后采用交替训练的方法同时训练知识图谱学习模块和推荐模块，并得到用户和物品的特征表示；最后，通过这些特征表示和用户的短期偏好向用户综合推荐。

这种推荐算法考虑了用户短期偏好的影响，并且着重考虑了用户与知识图谱之间的联系，有效利用以上信息实现了更好的推荐。

### 3.2 融合知识图谱和贝叶斯网络的推荐算法

文献[3]通过引入知识图谱表示学习方法改进了基于贝叶斯网络的协同过滤推荐算法。这一改进弥补了贝叶斯网络在特征提取上的局限性，增加了语义信息，可以更好地发现用户新兴趣。

知识图谱表示学习对知识图谱中实体和关系进行语义表示学习，通过将富有语义信息知识图谱三元组嵌入  $n$  维语义空间并生成对应向量，从而实现了知识图谱的数值化。通过知识图谱表示学习，可以快速计算两个实体间的语义相似性，并方便地将知识图谱用到其他学习任务中。

该模型首先通过知识图谱嵌入表示得到低维表示向量，然后通过待预测项语义近邻的评分来预测待预测项的评分；从数据中手动提取特征及评分，训练贝叶斯网络，获得联合概率分布表，然后计算待预测数据的评分；最后按照融合一定的权重对两个模型的评分进行融合，根据最终评分进行推荐。

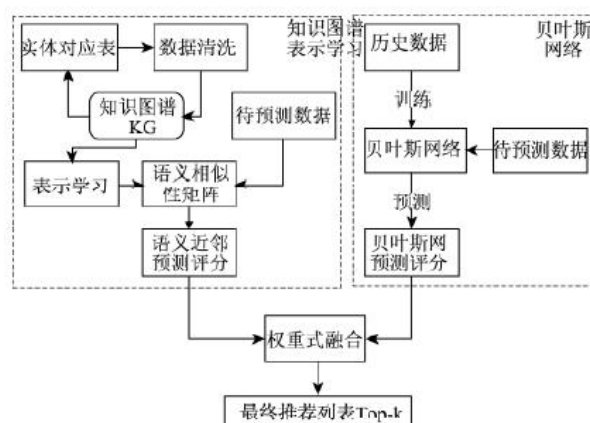


图 4. 权重式融合模型

实验表明，这一改进的推荐算法优于单一知识图谱或贝叶斯网络的推荐算法，将知识图谱表示向量与贝叶斯网络融合的方法是有效的。

### 3.3 融合知识图谱语义信息的推荐算法

文献[4]提出了一种结合知识图谱语义信息的用户画像构建方式，并且利用用户画像之间的相似性进行协同过滤推荐。其中以 DBpedia 中电影知识图谱为例提出了将自动编码器的网络结构与基于知识图谱的语义信息结合，赋予隐藏层中的神经元电影主题意义，从用户的观影历史中，得到每个用户对相关主题的偏好程度，完善用户画像的构建，运用协同过滤算法进行推荐。

模型中将自动编码器的网络结构看作知识图谱中实体之间的连接，从而使神经网络中的隐藏层有了具体的语义解释。具体而言，每个神经元表示所采用的知识图谱中的一个实体，神经元之间的连接则是实体之间的关系。根据用户过去评价的电影为每个用户构建不同的自动编码器模型进行训练，隐藏层的节点数量取决于所评价电影的主题信息数。由于每个隐藏层节点对应于知识图谱中电影的一个主题，因此隐藏层便有了其实际意义。

将知识图谱主题语义信息附加到自动编码器中，通过训练可以获得有意义、可理解的表示。相对于只利用用户电影评分信息的传统推荐算法而言，该推荐模型能够充分利用关联开放数据中丰富的语义信息，实现更好的推荐。

### 3.4 融合知识图谱和图注意力网络的推荐算法

现有的方法大多将知识图谱的网络结构划分为单独路径或仅利用了一阶邻居信息，造成无法建立整个图上的高阶连通性问题。为解决该问题，文献[5]提出了融合知识图谱和图注意力网络的推荐模型。该模型融合知识图谱，通过图注意力网络的递归传播过程可以捕获图中节点的高阶连通性。

该模型以电影项目为例，引入了电影项目的知识图谱，将其作为推荐系统的辅助信息，之后基于图注意力网络，设计了双线性信息采集器，结合线性信息采集器，对用户、项目和知识图谱中实体（如电影类型、导演、演员）的交互信息和高阶关系进行建模，以捕获用户的深层兴趣。

该模型通过融合知识图谱并挖掘推荐场景中用户到项目基于特征交互的高阶连通性,使得推荐效果明显改善。

## 4 结语

本文概述了四种融合知识图谱的推荐模型。知识图谱具有强大的语义处理能力和开放组织能力。通过融合知识图谱做辅助信息可以有效缓解推荐系统中的数据稀缺和冷启动问题,实现更好的推荐。目前越来越多的研究聚焦在深度学习与知识图谱在推荐算法中的应用,并取得了很好的成果。

## 参考文献

- [1]. A. S. Introducing the knowledge graph: things, not strings[J]. Official google blog, 2012, 16.
- [2]. 高仰, 刘渊. 融合知识图谱和短期偏好的推荐算法[J/OL]. 计算机科学与探索:1-14[2020-12-11]. <http://kns.cnki.net/kcms/detail/11.5602.TP.20201207.1435.015.html>.
- [3]. 侯位昭, 张欣海, 宋凯磊, 司佳, 齐幸辉, 张博. 融合知识图谱及贝叶斯网络的智能推荐方法[J]. 中国电子科学研究院学报, 2020, 15(05):488-494.
- [4]. 陈涛, 刘学军, 张伯君. 融合知识图谱语义信息的推荐方法[J]. 计算机工程与设计, 2020, 41(11): 3047 - 3052.
- [5]. 潘承瑞, 何灵敏, 胥智杰, 王修晖, 宋承文. 融合知识图谱的双线性图注意力网络推荐算法[J/OL]. 计算机工程与应用:1-11[2020-12-11]. <http://kns.cnki.net/kcms/detail/11.2127.TP.20200909.1131.004.html>.