

# 基于知识图谱的稀疏数据协同过滤推荐算法

许雪晶<sup>1</sup>, 林辰玮<sup>2</sup>

(1. 莆田学院信息工程学院, 福建莆田 351100; 2. 湄洲湾职业技术学院, 福建莆田 351119)

**摘要:** 因缺乏足够的交互关系支撑导致推荐精度不佳, 对此, 提出基于知识图谱的稀疏数据协同过滤推荐算法。抽取用户与物品的交互关系, 构建知识图谱, 利用知识图谱中的实体关系对用户和物品进行扩展表示。结合卷积神经网络(CNN)将交互关系扩为复杂结构, 捕获上下文信息, 以欧氏距离算相似度。找到目标用户相似邻居集, 用用户协同过滤预测评分, 融合时间加权策略动态调整, 生成推荐列表。测试表明, 该算法归一化折损累计增益(NDCG)值高, 平均绝对误差(MAE)和均方根误差(RMSE)低, 推荐效果较理想。

**关键词:** 知识图谱; 稀疏数据; 推荐算法; 相似度; CNN网络; 推荐精度

**中图分类号:** TP391 **文献标志码:** A **文章编号:** 1671-1807(2025)06-0030-06

用户在享受丰富网络资源的同时, 也面临着“信息过载”的困扰。如何帮助用户在海量数据中迅速找到符合自己兴趣的信息, 成为当前信息科学领域研究的热点。推荐系统作为解决这一问题的有效工具, 已广泛应用于电子商务、社交媒体、视频流媒体等多个领域。然而, 传统的推荐算法, 如基于内容的推荐和基于行为的推荐, 在面对数据稀疏性和冷启动问题时, 往往表现不佳。因此, 探索新的推荐算法, 以提高推荐系统的准确性和效率, 具有重要的研究意义和实践价值。

近年来, 国内外学者已经从多种不同的角度对推荐算法进行了优化研究。例如, 韩志耕等<sup>[1]</sup>通过时间窗口机制筛选用户近期的行为数据, 然后基于用户-项目评分矩阵计算用户之间的相似度, 并引入近邻可信度因子来评估近邻用户的可靠性。在推荐过程中, 选择可信度高的近邻用户进行预测评分, 进而生成推荐列表。但由于数据稀疏性, 相似度计算的准确性受到限制, 可能导致部分用户无法得到有效的推荐。此外, 时效性的引入也增加了算法对数据更新的敏感性, 需要频繁更新数据以保持推荐的时效性。余学兵等<sup>[2]</sup>通过相关因子计算数据相似性, 利用云模型补全数据缺失项, 在上述数据处理的基础上, 通过长短期记忆网络构建稀疏数据过滤推荐模型, 基于该模型实现数据过滤推荐。该方法能够有效处理数据缺失问题, 提高数据过滤推荐的全面性, 但是相关因子的选

择可能存在局限性, 如果选择不当可能导致相似性计算的偏差, 影响整个数据处理流程的准确性, 进而影响推荐结果的准确性。欧朝荣和胡军<sup>[3]</sup>不仅考虑了用户的显式评分数据, 还充分利用了用户的隐式反馈(如点击、浏览等行为), 以更全面地捕捉用户的兴趣偏好。通过显隐式反馈的融合, 算法能够更准确地计算用户之间的相似度, 并为用户推荐更符合其兴趣的物品。但由于隐式反馈的获取和处理相对复杂, 且存在较大的噪声干扰, 可能导致推荐的精度受到一定影响。史加荣和何攀<sup>[4]</sup>利用矩阵补全技术对用户-物品评分矩阵进行填充, 以缓解数据稀疏性问题。然后采用宽度学习模型对用户和物品进行特征提取和表示, 以更深入地挖掘用户和物品之间的潜在关系。但矩阵补全技术的准确性受到多种因素的影响, 如补全算法的选择、补全数据的来源等, 若这些因素处理不当, 可能导致补全后的矩阵存在误差, 进而影响推荐的精度。

针对现有问题, 本文提出了一种基于知识图谱的稀疏数据协同过滤推荐算法, 通过将知识图谱的丰富知识结构和关系信息与传统的协同过滤推荐算法相结合, 不仅克服了传统协同过滤算法在处理稀疏数据时的局限性, 也有效提升了推荐结果的准确性和个性化程度。这种结合不仅丰富了推荐系统的数据源, 也增强了推荐系统的解释性和可理解性, 使得推荐结果更加符合用户的实际需求。

**收稿日期:** 2024-09-23

**基金项目:** 福建省中青年教师教育科研项目(科技类)(JAT210399, JAT241369)

**作者简介:** 许雪晶(1988—), 女, 福建莆田人, 硕士, 讲师, 研究方向为数据挖掘研究; 林辰玮(1988—), 男, 福建莆田人, 硕士, 助理研究员, 研究方向为信息化建设、大数据等。

## 1 基于知识图谱的稀疏数据协同过滤推荐算法

### 1.1 基于知识图谱的用户和物品的扩展表示

利用知识图谱技术,将图谱中的实体和关系嵌入到低维向量空间中,得到每个实体和关系的向量表示,实现用户和物品的扩展表示。利用知识图谱技术将图谱中的实体和关系嵌入低维向量空间,能够有效扩展用户和物品的表示,通过捕捉实体间的语义关系和上下文信息,增强对用户和物品特征的理解,从而在数据稀疏的情况下提高推荐算法的准确性和鲁棒性。

首先从电子商务网站的购买记录、用户评分系统以及用户浏览历史等多个渠道,对用户的行为数据进行收集。以此可以结合用户的购买行为或商品评价行为,对用户与物品的交互关系进行抽取,这些关系可以表示为〈用户,交互关系,物品〉形式的三元组。通过上述方式可以构建出一个标注数据集  $D = \{(t_1, e_1, r_1), (t_2, e_2, r_2), \dots, (t_n, e_n, r_n)\}$ , 其中,  $t_i$  为用户评价文本,  $e_i$  为文本中的实体集合,  $r_i$  为实体之间的关系集合<sup>[5]</sup>。在完成用户-物品关系抽取后,通过对实体以及关系进行整合,可以构建知识图谱。假设知识图谱中的三元组集合为  $T$ , 则整合过程可以表示为

$$T = f(E, R) \quad (1)$$

式中:  $f(\cdot)$  为整合函数;  $E$  为知识图谱中的实体集合;  $R$  为知识图谱中的关系集合。将构建的知识图谱与上文抽取得到的〈用户,交互关系,物品〉进行关联,可以得到关于用户和物品的学习表示。对于用户,除了考虑其与物品的交互关系,还需要结合知识图谱中的相关信息进行扩展表示<sup>[6]</sup>。考虑用户购买的物品所属的品牌、类别等,以及这些品牌、类别之间的关系,由此得到的用户  $u$  的扩展表示为

$$\text{User}(u) = \sum_{i \in I(u)} \text{Item}(i) + \sum_{b \in B(u)} \text{Brand}(b) + \sum_{c \in C(u)} \text{Category}(c) \quad (2)$$

式中:  $I(u)$  为用户  $u$  购买或交互过的物品集合;  $B(u)$  和  $C(u)$  分别为用户  $u$  购买或交互过的物品所属品牌和类别集合;  $\text{Item}(i)$  为物品  $i$  的扩展表示;  $\text{Brand}(b)$  和  $\text{Category}(c)$  分别为品牌  $b$  和类别  $c$  的扩展表示<sup>[7]</sup>。同理,对于物品,除了考虑其与用户的交互关系,还需要结合知识图谱中的相关实体和关系进行扩展表示。物品  $i$  的扩展表示为

$$\text{Item}(i) = E(i) + \sum_{a \in A(u)} \text{Attribute}(a) + \sum_{b \in B(u)} \text{Brand}(b) + \sum_{c \in C(u)} \text{Category}(c) \quad (3)$$

式中:  $E(i)$  为物品  $i$  在知识图谱中的嵌入向量;  $A(u)$  为物品的属性集合,  $\text{Attribute}(a)$  为属性  $a$  的扩展表示。

通过上述步骤即可完成对于用户和物品的扩展表示处理,具体通过从用户的行为数据中抽取用户与物品的交互关系,并对知识图谱进行构建,然后利用知识图谱中的实体关系对用户和物品进行扩展表示。

### 1.2 用户及物品相似度计算

考虑到在稀疏数据的情况下,直接计算用户-用户或物品-物品的相似度可能会受到数据稀疏性的影响,因此为实现协同过滤推荐效果,在得到用户和物品的嵌入向量后,采用卷积神经网络(convolutional neural network, CNN)来捕获嵌入向量的上下文信息,并使用这些上下文嵌入向量,对用户-用户与物品-物品之间的相似度进行计算<sup>[8]</sup>。CNN 能够有效地利用图结构中的节点关系,通过信息聚合和传递机制,增强节点(即用户和物品)的嵌入表示,从而更全面地反映其上下文特征。这种方法克服了传统方法在处理稀疏数据时的不足,能够更准确地计算用户-用户与物品-物品之间的相似度,进而提升协同过滤推荐的精度和效果。

首先,将用户-物品交互数据构建成一个图  $G = (V, E)$ 。在这个图中,用户和物品这两个实体可以作为节点,而用户与物品的交互关系,如购买行为或评价行为则可以作为边,其中,  $V$  和  $E$  分别为节点集合和边集合。然后采用随机生成的方式,为每个节点(用户和物品)初始化一个嵌入向量,然后在图上传播节点信息来更新节点嵌入<sup>[9]</sup>。在每一层 CNN 中,节点都会聚合其邻居节点的信息,并据此更新自己的嵌入。假设实体节点集合为  $N(v)$ , 则每一层的更新表达式为

$$h_v^{l+1} = \sigma \left( \sum_{u \in N(v)} \frac{1}{\sqrt{|N(v)|} \sqrt{|N(u)|}} W^l h_u^l + b^l \right) \quad (4)$$

式中:  $h_v^l$  为在第  $l$  层网络结构中实体节点的嵌入;  $N(u)$  为实体节点的邻居节点集合;  $W^l$  和  $b^l$  分别为该层对应的权值矩阵和偏置项;  $\sigma$  为激活函数<sup>[10]</sup>。采用上述更新方式,可以实现多层 CNN 传播,每个节点都会得到一个融合其邻居节点信息的嵌入向量,这个嵌入向量可以视为该节点的上下文嵌入。

上下文嵌入不仅包含节点自身的信息,还包含与之相连的其他节点(如历史交互、社交关系等)的信息<sup>[11]</sup>。基于上述提取得到的上下文嵌入,结合欧几里得距离对用户-用户和物品-物品的相似度进行计算,具体计算公式为

$$\text{sim}(u, v) = \sqrt{\sum_{i=1}^d (u_i - v_i)^2} \quad (5)$$

式中:  $d$  为嵌入向量的维度;  $u_i$  和  $v_i$  分别为用户  $u$  和用户  $v$  的分量;  $\text{sim}(u, v)$  为两个用户之间的相似度。

通过上述步骤即可完成对于用户及物品的相似度计算。通过结合卷积神经网络(CNN)对用户和物品的嵌入向量的上下文信息进行捕获,并结合欧几里得距离,对用户-用户和物品-物品的相似度进行计算。CNN 可以将用户和物品的交互关系扩展为一个更复杂的网络结构,从而有效规避数据稀疏性带来的协同推荐精度不佳的问题。

### 1.3 稀疏数据协同过滤推荐

在完成相似度计算后,选择基于用户进行协同过滤,从而得到预测评分,对未评价的物品进行排序,从而生成推荐列表,实现稀疏数据协同过滤推荐,这样能够充分利用用户间的相似性来填补评分数据的空白,提高推荐的针对性和准确性,同时在数据稀疏的情况下依然能够为用户提供个性化的推荐服务。

由于在稀疏数据环境中,许多用户-物品的评分存在缺失情况,因此为缓解这个问题,结合均值填充的方式,对稀疏数据进行处理<sup>[12]</sup>。在评分数据缺失严重的情况下,直接计算相似度或进行推荐可能导致结果偏差较大。均值填充通过用已有评分的平均值来填补缺失值,能够在一定程度上平滑数据分布,减少极端值的影响,从而提高后续相似度计算和推荐算法的稳定性与准确性。稀疏数据均值  $\bar{r}$  填充表达式为

$$\bar{r} = \frac{\sum_{u, i \in N_r} r_{ui}}{|N_r|} \quad (6)$$

式中:  $N_r$  为所有已知的用户评分集合;  $|N_r|$  为已知评分数量;  $r_{ui}$  为用户  $u$  对物品  $i$  的评分。对于目标用户  $u$  和未评价物品  $i$ , 通过找到与用户  $u$  最相似的  $K$  个用户邻居,然后利用这些邻居对物品  $i$  的评分对用户  $u$  对物品  $i$  的评分进行预测,从而实现基于用户的协同过滤,预测评分  $r'_{ui}$  的计算公式为

$$r'_{ui} = \frac{\sum_{u \in N(u)} \text{sim}(u, v) \cdot r_{vi}}{\sum_{u \in N(u)} |\text{sim}(u, v)|} \quad (7)$$

式中:  $r_{vi}$  为用户  $v$  对物品  $i$  的评分。考虑到用户兴趣以及物品流行度可能会随着时间变化,因此选择引入时间加权策略,来对预测评分进行动态调整<sup>[13]</sup>。假设  $t_{\text{current}}$  为当前时间;  $t_{ui}$  为用户  $u$  对物品  $i$  给出评分的时间,则时间间隔为  $\Delta t_{ui} = t_{\text{current}} - t_{ui}$ ,由此对时间衰减函数  $f(\Delta t_{ui})$  进行定义,具体表达式为

$$f(\Delta t_{ui}) = e^{-\alpha_{ui} \Delta t_{ui}} \quad (8)$$

式中:  $\alpha_{ui}$  为衰减因子,用于控制衰减速度,当衰减因子  $\alpha_{ui}$  较大时,评分随时间的衰减速度更快。结合上述构建的时间衰减函数,对式(7)进行优化,由此可以得到融合时间加权策略的预测评分表达式为

$$r'_{ui} = \frac{\sum_{u \in N(u)} \text{sim}(u, v) \cdot f(\Delta t_{ui}) \cdot r_{vi}}{\sum_{u \in N(u)} |\text{sim}(u, v)| \cdot f(\Delta t_{ui})} \quad (9)$$

通过上述步骤即可完成稀疏数据协同过滤推荐。首先结合均值填充法,对数据稀疏性进行缓解。然后通过找到与目标用户最相似的用户邻居集合,结合用户协同过滤对用户评分进行预测<sup>[14]</sup>。最后融合时间加权策略,对预测评分进行动态调整,并生成对应的推荐列表。至此,基于知识图谱的稀疏数据协同过滤推荐算法设计完成<sup>[15]</sup>。

## 2 实验论证

为了证明提出的基于知识图谱的稀疏数据协同过滤推荐算法在实际推荐效果方面优于常规的协同过滤推荐算法,在完成算法方面的设计后,通过构建实验对比环节,对所提方法的实际推荐效果进行检验。

### 2.1 实验说明

为验证所提的基于知识图谱的稀疏数据协同过滤推荐算法在实际推荐效果方面的优越性,实验选取两组常规的协同过滤推荐算法作为对比对象,分别为基于显隐式反馈的协同过滤推荐算法(常规方法 A),以及基于矩阵补全的协同过滤推荐算法(常规方法 B)。通过构建实验平台,采用三种方法对同一组稀疏数据进行模拟推荐,对比不同方法的实际推荐效果。

### 2.2 实验对象

实验所构建的用户商品评论数据集共包括两个部分,即用户评论信息和用户/商品属性信息。首先使用随机数生成器为每个用户分配一个唯一

的 ID(范围 1~10000),并使用另一个随机数生成器为每个商品分配一个唯一的 ID(范围 1~50000)。对于每个用户,根据其兴趣分布选择一定数量的商品进行评论。对于每个选定的商品,基于当前时间减去一个随机时间间隔,生成一个评论时间戳,然后分别基于商品属性以及用户评分分布,生成评论内容以及具体评分分值。由于用户通常不会对所有的商品进行评论,因此为保证数据集的稀疏性,实验所生成的评论总数为 800 000 条,远低于用户数×商品数。用户商品评论数据集中部分数据,如表 1 所示。

针对上述构建出的数据集,将评论内容转化为适合知识图谱构建的形式。将用户实体选为评论数据中的用户 ID,并将商品实体选为商品 ID,并直接建立用户与商品之间的交互关系。同时,在将实体和关系添加到知识图谱之前,进行唯一性检查。通过结合数据库的唯一性约束,来确保不会添加重复的实体。如果发现在不同评论中提到了相同但表达方式略有不同的实体(如同义词、缩写等),则需要进行实体合并,以确保它们在知识图谱中只有一个表示。

实验选取了 9 位用户作为推荐对象,分别基于这些用户的评论数据,采用不同的方法对其生成推荐列表。在采用所提方法进行模拟协同过滤推荐时,设定 CNN 网络层数为 2 层,节点嵌入维度为 64,批量大小为 32,设定衰减因子  $\alpha_{ui}$  为 0.08。待实验完成后,对不同方法的推荐结果进行记录,从而实现实验对比。

表 1 用户商品评论数据集部分数据

用户 ID	商品 ID	评论时间戳	评论内容	评分
1	234	1626883200	我很喜欢这款手机的外观设计,屏幕也很清晰,但电池续航有待提高	4
1	567	1626897600	买了这本书很久了,内容很实用,但有些地方读起来有些吃力	3
2	123	1626984000	这是我第二次购买这个品牌的运动鞋,质量一如既往的好,强烈推荐	5
3	456	1626811200	这件衣服的材质很好,穿着很舒服,而且价格也很实惠	4
5	321	1626710400	这款耳机的音质很棒,但佩戴时间长了耳朵会有些不舒服	4

### 2.3 推荐精度对比

所提方法所得到的商品 ID 推荐列表以及用户对于商品的预测评分,如表 2 所示。

通过表 2 实验结果可以看出,提出的推荐算法能够有效根据用户的评论数据,对用户的评分进行预测,从而生成对应的商品推荐列表,同时预测评分在实际评分区间之内。这是因为所提方法通过将用户和物品的交互关系在知识图谱中扩展为复杂的网络结构,并利用 CNN 对这些复杂关系进行建模,能够更深入地捕捉用户和物品嵌入向量的上下文信息。这种处理方式相比传统方法更能理解用户偏好和物品特征之间的复杂关系,提高了推荐的准确性。

为对不同方法的实际推荐精度进行对比分析,实验以不同推荐结果的归一化折损累计增益(normalized discounted cumulative gain, NDCG)作为对比指标,该值可以反映算法对结果的排序性能。实验考虑了不同用户群体的影响,由此得到的不同活跃度的用户群体对应的 NDCG 值对比结果,如图 1 所示。

表 2 所提方法下的商品推荐列表及预测评分

用户 ID	商品 ID/推荐列表	预测评分	实际评分
User1	[005, 010, 115]	4.2	[5.0, 4.0, 3.5]
User2	[003, 047, 089]	3.8	[4.5, 3.0, 4.2]
User3	[002, 082, 018]	4.5	[5.0, 4.8, 4.0]
User4	[001, 056, 511]	3.5	[3.0, 3.5, 4.0]
User5	[008, 213, 319]	4.0	[4.5, 3.8, 4.2]
User6	[004, 014, 020]	4.3	[4.5, 4.0, 4.5]
User7	[002, 005, 517]	4.7	[5.0, 4.5, 4.0]
User8	[001, 029, 016]	3.2	[3.0, 3.5, 3.0]
User9	[027, 511, 215]	4.1	[4.5, 4.0, 3.5]

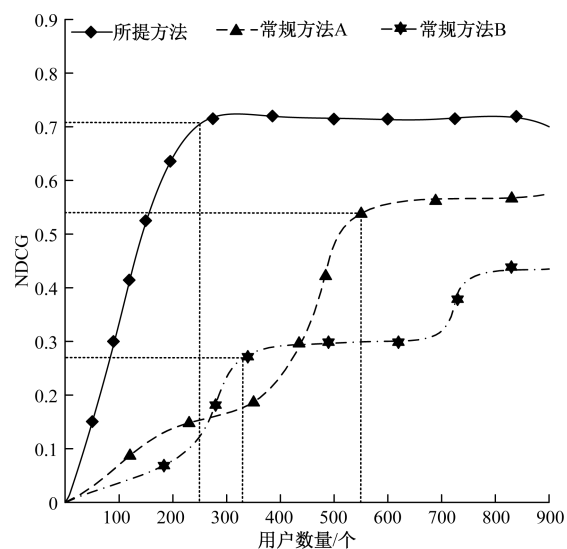


图 1 不同推荐方法下的 NDCG 值对比结果

通过上述实验结果可以看出, NDCG 值可能会受到用户活跃度的影响更大。对于活跃度较低的用户, 由于缺乏足够的用户行为数据, 算法可能难以准确地捕捉其兴趣, 导致 NDCG 值较低。通过对比 NDCG 值变化情况可以明显看出, 所提出的基于知识图谱的稀疏数据协同过滤推荐算法在实际推荐精度方面明显优于两种常规的推荐算法, 基于显隐式反馈的协同过滤推荐算法在处理稀疏数据时, 其局限性明显, 由于显隐式反馈数据本身可能就不丰富, 在数据稀疏情况下难以准确找到足够相似的用户或物品, 导致推荐准确性下降。基于矩阵补全的协同过滤推荐算法在实际应用中, 其对数据结构的假设可能与实际情况不符, 如果数据不满足低秩假设, 补全的效果就会大打折扣, 而且在处理高度稀疏的数据时, 由于可利用的已知元素过少, 也很难准确补全矩阵, 从而影响推荐的质量。而所提方法在预测用户评分时, 不仅基于找到的最相似用户邻居集合进行协同过滤, 还融入了时间加权策略。充分考虑到了用户兴趣随时间变化的可能性, 通过动态调整评分预测, 使得推荐结果更加符合用户的当前需求和偏好, 增强了推荐结果的个性化程度。

为了进一步验证所提方法的推荐效果, 选取平均绝对误差(mean absolute error, MAE)和均方根误差(root mean square error, RMSE)作为实验指标, MAE 通过计算预测评分与实际评分之间的差的绝对值来衡量推荐方法的精度。在稀疏数据场景下, MAE 能够直观地反映算法对用户偏好的预测准确性。RMSE 对预测误差进行了平方处理, 因此它对较大的误差更为敏感。在稀疏数据下, RMSE 能够揭示算法在预测极端偏好或极端评分时的稳定性。二者的计算公式为

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_i - q_i| \quad (10)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - q_i)^2} \quad (11)$$

式中:  $p_i$  为预测评分;  $q_i$  为实际评分。

通过式(10)和式(11)计算所提方法、常规方法 A 和常规方法 B 的 MAE 和 RMSE, 如图 2 和图 3 所示。

由图 2 和图 3 可知, 所提方法的 MAE 始终低于 0.5, RMSE 始终低于 0.4, 且明显低于基于显隐式反馈的协同过滤推荐算法和基于矩阵补全的协同过滤推荐算法, 说明所提方法的推荐结果精度更高。常规推荐算法在面临数据稀疏性问题时, 推荐

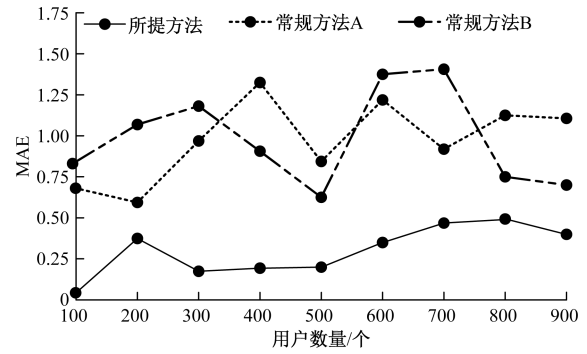


图 2 不同方法的 MAE 值对比结果

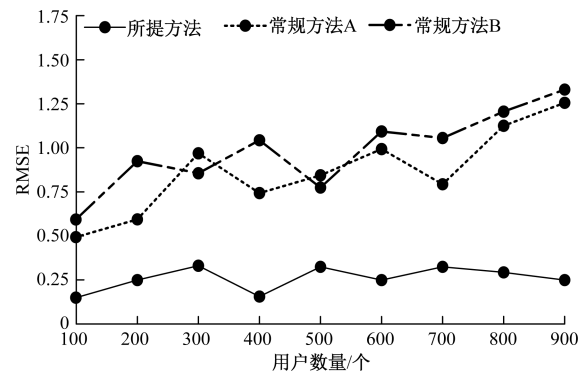


图 3 不同方法的 RMSE 值对比结果

精度常受影响。而所提方法创新性地引入了知识图谱技术, 通过构建用户与物品之间的复杂关系网络, 有效缓解了数据稀疏性问题。知识图谱不仅包含了直接的用户-物品交互信息, 还蕴含了丰富的实体关系, 为推荐算法提供了额外的上下文信息, 从而提高了推荐结果精度。

### 3 结论

综上所述, 提出的基于知识图谱的稀疏数据协同过滤推荐算法, 通过深入挖掘用户与物品的复杂交互关系, 并利用知识图谱的丰富语义和结构化特性, 有效缓解了数据稀疏性问题, 显著提升了推荐系统的精度和性能。实验结果表明, 采用该方法进行商品推荐时, 能够显著提高 NDCG 值, 这直接反映了推荐结果与用户实际需求的匹配程度较高, 证明了该方法在提升推荐精度和效果方面的有效性。同时该方法在降低预测误差(MAE 和 RMSE 较低)方面也展现显著优势, 说明其有效提高了推荐的准确性。这不仅验证了算法的有效性, 也为未来在更复杂、更稀疏的数据环境下实现高质量推荐提供了新思路。随着知识图谱技术和深度学习方法的不断进步, 相信该方法将在实际应用中展现更加广阔的前景, 为用户带来更加个性化、精准的推荐体验。

未来的研究重点将聚焦于如何更有效地利用辅助信息与深度学习技术相结合,以更精确地建模用户与物品之间的复杂关系,从而进一步提升在稀疏数据环境下的推荐效果与用户体验。

### 参考文献

- [1] 韩志耕, 范远哲, 陈耿, 等. 一种基于时效近邻可信选取策略的协同过滤推荐方法[J]. 计算机科学, 2023, 50(S2): 518-528.
- [2] 余学兵, 熊蕾, 黄丽, 等. 基于长短期记忆的稀疏数据过滤推荐算法[J]. 计算机仿真, 2023, 40(2): 395-398.
- [3] 欧朝荣, 胡军. 融合显隐式反馈的协同过滤推荐模型[J]. 控制与决策, 2024, 39(3): 1048-1056.
- [4] 史加荣, 何攀. 结合矩阵补全的宽度协同过滤推荐算法[J]. 智能系统学报, 2024, 19(2): 299-306.
- [5] 苏湛, 杨昊川, 艾均. 基于模糊偏好标签向量的推荐算法[J]. 应用科学学报, 2024, 42(3): 525-539.
- [6] 李驰, 游小钰, 张溢. 基于解耦图卷积网络的协同过滤推荐模型[J]. 中文信息学报, 2023, 37(11): 131-141.
- [7] 黄树添, 胡诗琳, 卜祥智, 等. 融入用户风险偏好的三支协同过滤推荐模型[J]. 南京大学学报(自然科学), 2023, 59(5): 777-789.
- [8] 张茂宇, 李海明. 融合双重语义相似度的混合协同过滤推荐算法[J]. 计算机工程与设计, 2023, 44(9): 2671-2678.
- [9] 张宝鑫, 杨丹, 聂铁铮, 等. 基于自监督的多视图图协同过滤推荐方法[J]. 计算机工程, 2024, 50(5): 100-110.
- [10] 向俊伟, 李玲娟. 融合 PCA 降维和均值漂移聚类的协同过滤推荐算法[J]. 南京邮电大学学报(自然科学版), 2023, 43(3): 90-95.
- [11] 杨本臣, 李依泽. 融合显隐式特征的深度协同过滤推荐算法[J]. 辽宁工程技术大学学报(自然科学版), 2023, 42(3): 354-361.
- [12] 成凌飞, 杨向峰, 陈天源, 等. 基于增强注意力机制的三通道神经协同过滤改进算法[J]. 传感器与微系统, 2023, 42(6): 137-140.
- [13] 景楠, 周正茜, 袁戟. 基于宽度学习和协同过滤的在线信息推荐算法[J]. 计算机应用与软件, 2023, 40(5): 279-287.
- [14] 张建华, 郭启迪, 曹子傲, 等. 基于匹配度和谱聚类的知识推荐研究[J]. 郑州大学学报(理学版), 2023, 55(4): 1-7.
- [15] 邓泓, 吴祎, 于程远, 等. 基于可信预测值的协同过滤推荐算法[J]. 江西师范大学学报(自然科学版), 2022, 46(6): 642-648.

## Collaborative Filtering Recommendation Algorithm Based on Knowledge Graph for Sparse Data

XU Xuejing<sup>1</sup>, LIN Chenwei<sup>2</sup>

(1. School of Information Engineering, Putian University, Putian 351100, Fujian, China;

2. Meizhouwan Vocational Technology College, Putian 351119, Fujian, China)

**Abstract:** Due to the lack of sufficient interaction support, the recommendation accuracy is poor. To address this, a sparse data collaborative filtering recommendation algorithm based on knowledge graph was proposed. Extract the interaction relationship between users and items, a knowledge graph was constructed, and the entity relationships in the knowledge graph was used to extend the representation of users and items. Combining CNN networks, interactive relationships was expanded into complex structures, contextual information was captured, and similarity using Euclidean distance was calculate. A set of similar neighbors was found for the target user, user collaboration filtering was used to predict ratings, the fusion time weighting strategy was dynamically adjusted, and a recommendation list was generated. Tests have shown that the algorithm has high NDCG values, low MAE and RMSE values, and ideal recommendation performance.

**Keywords:** knowledge graph; sparse data; recommendation algorithm; similarity; CNN; recommendation accuracy