

基于图数据库的用户目标信息个性化推荐方法

林楠¹

LIN Nan

摘要

用户目标信息个性化推荐系统在用户和商品交互数据稀少或新用户与商品缺乏历史数据时,会出现数据稀疏性和冷启动问题。这使传统方法通过简单的矩阵形式难以描述用户与商品关系,无法有效挖掘能够反映用户偏好的深层次特征,最终导致推荐信息与用户需求适配度低。为此,提出一种基于图数据库的用户目标信息个性化推荐方法。该方法利用图数据库的结构化特点,深入捕捉和分析用户与商品间的关联,提取并标注用户目标信息。从用户历史行为数据中提取出能够反映用户偏好的深层次特征,引入协同过滤模型,基于目标信息中挖掘出的用户偏好模式构建用户偏好模型。运用自然语言处理和机器学习技术,从信息资源中提取标签,并结合之前构建的用户偏好模型,设计个性化的目标信息资源推荐。实验结果表明:该方法不仅能显著提升推荐信息与用户需求的适配程度,还能增加个性化信息的推荐量,从而更好地满足用户的个性化需求。

关键词

图数据库; 用户目标信息; 个性化推荐; 协同过滤; 用户偏好模型

doi: 10.3969/j.issn.1672-9528.2025.02.037

0 引言

用户在面对庞大的信息源时,往往难以迅速找到真正符合自己需求和目标的内容。因此,深入研究用户目标信息个性化推荐方法具有重要的理论和实践意义。这不仅有助于提升用户体验,还能为信息传播、电子商务等众多领域的发展提供有力的支持。

目前,已有诸多学者在这一领域进行了积极探索,例如,赵莉苹^[1]通过采集历史数据,筛选关键特征并利用K-中心聚类算法深入挖掘用户行为特征,建立了用户行为特征与资源特征的关联规则,并结合偏好因子,为用户生成了个性化的资源推荐列表。在该方法中,K-中心聚类算法通常需要足够的数据点来形成有效的聚类。在数据稀疏的情况下,聚类结果会不准确,影响对用户行为特征的深入挖掘,使得难以从聚类结果中提取出能够反映用户偏好的深层次特征,导致最终生成的个性化资源推荐列表与用户需求适配度较低。李智彪等人^[2]通过多维度算法,即利用Jaccard相似系数、夹角余弦相似度和潜在狄利克雷分布(LDA)主题模型相似度算法,计算实验方剂与煎煮数据库中方剂间的相似度,进而为经典名方提供个性化的煎煮参数推荐。Jaccard相似系数是

基于集合的交集和并集来计算相似度的。当数据稀疏时,所形成的集合元素有限,会导致计算出的交集和并集不能准确反映方剂之间的真实相似关系,影响最终的推荐适配度。

针对现有推荐方法存在的局限性,本文将引入图数据库技术,设计并研究一种用户目标信息个性化推荐的新方法。

1 基于图数据库的用户目标信息提取与标注行为处理

个性化推荐系统中,用户与商品关系多维度、多层次且动态变化(因用户兴趣需求随时间改变),传统数据处理方式难以应对大规模多维度数据,导致推荐不精准、用户满意度低^[3]。利用图数据库的结构化特点可解决此问题,通过导入用户行为数据和商品信息构建用户-商品图,该图能展示直接联系、揭示潜在关联路径和社群结构,计算用户相似度可挖掘潜在兴趣偏好,实现用户目标信息提取与标注行为处理^[4]。

在此过程中,首先将用户行为数据(如浏览、购买、评价等)和商品信息(如属性、标签等)导入图数据库,构建出用户节点、商品节点以及节点之间的关系边,形成用户-商品图。利用图数据库中的节点和关系数据,可以计算用户之间的相似度。用户相似度的计算公式为:

$$T(u_i, u_j) = \frac{\vec{u}_i \cdot \vec{u}_j}{\|\vec{u}_i\| \cdot \|\vec{u}_j\|} \quad (1)$$

式中: T 表示用户 u_i 与 u_j 之间相似度; $\vec{u}_i$ 、 $\vec{u}_j$ 表示 u_i 、 u_j 的

1. 吉林农业大学 吉林长春 130118

[基金项目] 吉林省高教学会高教研重点课题“人工智能赋能高校数字化教学资源平台建设研究”(JGJX2023B12)

征向量。

基于用户兴趣偏好、购买习惯、社交关系等信息、行为数据及相似度计算结果构建用户画像；借助图数据库关系边识别用户标注行为（如评分、评论、点赞等）^[5]，并综合分析其频率、分布、趋势，深入了解用户偏好变化与市场需求；利用机器学习算法和关系数据预测用户未来标注行为。预测公式为：

$$P(A|B) = \frac{P(B|A)}{P(B)} \cdot \frac{1}{T(u_i, u_j)} \quad (2)$$

式中： P 表示用户标注行为的预测结果； A 表示用户的未来标注行为； B 表示已知的用户行为或特征。

对于高概率的标注行为，可以提前进行推荐或提醒，而对于低概率的标注行为，则可以适当减少不必要的推送，避免对用户造成干扰^[6]。

图数据库能够高效地存储和查询用户与商品之间的复杂关系，使得在交互数据稀少或新用户与商品缺乏历史数据的情况下，也能通过图结构中的关联信息来推断用户的潜在偏好。这种深入的捕捉和分析能力，为后续的用户偏好建模和个性化推荐打下了坚实的基础。

2 用户偏好建模

在用户目标信息个性化推荐过程中，利用图数据库的结构化特点对用户偏好进行深入建模。首先，通过分析用户与商品的关联关系以及用户的具体行为数据，能够提取出能够深刻反映用户偏好的深层次特征。随后，基于这些特征，引入了协同过滤模型来构建用户偏好模型。通过代入基于目标信息中挖掘出的用户偏好模式生成针对某个用户的偏好模型，并遍历所有用户完成综合模型的构建^[7]。该模型能够根据用户的历史行为数据，预测用户未来可能感兴趣的物品，从而实现个性化推荐。

协同过滤模型训练公式为：

$$M(u_i) = \frac{\sum_j S(i, j)}{P(A|B) \cdot I(m)} \quad (3)$$

式中： S 表示用户历史上明确表示过喜好的信息集合； I 表示与用户喜欢过的信息相似的其他信息集合； m 表示训练协同过滤模型的过程。在此基础上，根据训练的数据，建立用户偏好向量，计算公式为：

$$p(u) = \sum w(i, j) \cdot \vec{F}_i \cdot M \quad (4)$$

式中： p 表示用户偏好向量； w 表示重要性权重； $\vec{F}$ 表示某信息的权重向量。利用用户偏好向量信息，建立用户偏好模型，计算公式为：

$$f = \min \sum (r_i - \hat{r}(\theta))^2 + p(u) \cdot \lambda \|\theta\|^2 \quad (5)$$

式中： f 表示用户偏好模型； r 表示用户对某信息的评分； $\hat{r}$ 表示 r 的预测值； θ 表示训练数据集； λ 表示正则化系数。

在实际操作中，将相关数据代入上述模型，生成针对某个用户的个性化偏好模型。然后，遍历所有用户，按照相同的方式生成用户的偏好模型，从而完成整个综合模型的构建。此时，该模型就能够根据用户的历史行为数据，预测用户未来可能感兴趣的物品，进而实现个性化推荐^[8]。

3 基于综合标签的目标信息资源个性化推荐

运用自然语言处理和机器学习技术，本文设计了一套个性化的目标信息资源推荐方案。先基于目标信息资源特性构建含多维度的标签体系，自动抽取标签；再依据用户历史和标注行为，结合用户偏好模型构建用户-标签偏好矩阵，用隐语义分析技术挖掘潜在关联形成用户标签偏好向量^[9]；最后根据信息资源标签集合与用户标签偏好向量计算匹配度，用推荐算法生成个性化推荐列表，实现基于综合标签和用户偏好模型的个性化推荐，兼顾信息覆盖与用户需求。目标信息资源个性化推荐方案的流程图，如图1所示。

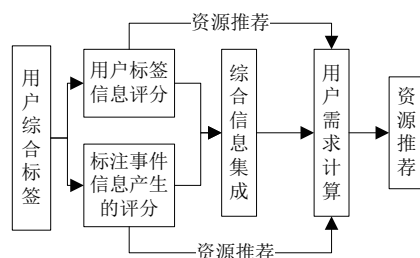


图1 目标信息资源个性化推荐

根据用户的历史行为数据和标注行为，构建用户-标签偏好矩阵。矩阵中的元素表示用户对某个标签的偏好程度。用户-标签偏好矩阵元素计算公式为：

$$C = \frac{f(i, t)}{|u(\chi)|} \quad (6)$$

式中： C 表示用户-标签偏好矩阵； t 表示矩阵元素； χ 表示信息资源集合。使用隐语义分析，挖掘用户与标签之间的潜在关系，对于每个用户，计算其对所有标签的偏好权重，形成用户的标签偏好向量，考虑标签的热门程度和用户对该标签的独特偏好，对标签权重进行归一化和调整。处理过程用公式表示为：

$$W = \frac{C}{\sqrt{L^2 + (\beta + 1)^2}} \quad (7)$$

式中： W 表示标签权重处理； L 表示标签字节长度； β 表示所有标签的集合。根据目标信息资源的标签集合和基于用户偏好模型生成的标签偏好向量，计算信息资源与用户之间的

匹配度，使用推荐算法生成个性化的推荐列表^[10]。推荐过程公式为：

$$Z=\frac{1}{l\cdot\gamma^2+h\cdot W}$$

(8)

式中：Z表示个性化推荐列表；l表示个性化推荐信息数量；γ表示信息资源的覆盖程度；h表示用户需求量。

4 对比实验

4.1 实验准备

本次实验选取了一家大型互联网公司作为研究对象，其用户网络信息样本如表 1 所示。

表 1 用户网络信息样本

用户 ID	性别	网络接入方式	上网时长 (h/周)	访问频次 (次/周)
U1	男	Wi-Fi	20	50
U2	女	4G	15	40
U3	男	Wi-Fi	30	70
U4	女	5G	10	30
U5	男	Wi-Fi	25	60
U6	女	4G	18	45
U7	男	3G	8	20

在实验中，收集了用户行为数据，包括用户浏览、购买、评价及收藏等各项活动，并同步收集了商品数据，详细记录了商品信息、属性及标签等内容。安装并成功配置了图数据库环境，以确保各项技术参数满足实验需求，具体参数如表 2 所示。

表 2 图数据库环境的技术参数

序号	项目	参数
1	数据库类型	Neo4j
2	安装版本	Neo4j 4.5.3
3	存储空间	10 TB
4	内存配置	64 GB
5	CPU 配置	Intel Xeon E5-2680 v4, 28 核心
6	网络带宽	1 Gbit/s
7	并发连接数	1000

基于这一环境，利用数据导入工具将用户行为数据、商品数据及用户画像数据导入图数据库中。在此过程中，构建了用户、商品、标签等节点，并明确了用户与商品、商品与标签等关系边界。根据数据的具体特点，设计了图数据库的模式，包括节点属性、关系类型等。根据图分析算法的结果，生成了一个个性化推荐列表。

完成上述设计后，引入了文献 [1] 中提出的基于用户行为特征的推荐方法，以及文献 [2] 中提出的基于融合多

维度相似算法的推荐方法，并将这两种方法作为对照组。同时，采用上述 3 种方法共同进行用户目标信息的个性化推荐。

4.2 实验结果与分析

在对比实验中，将 3 种不同方法推荐的信息与用户偏好程度进行匹配，并分析了用户对推荐信息的偏好程度，以此作为实验结果的关键指标。同时，还统计了个性化推荐信息的具体条数，具体如图 2~4 所示。

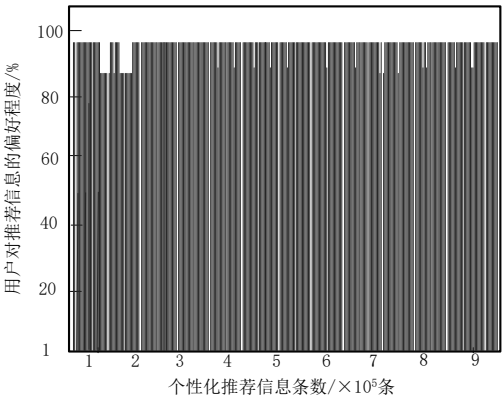


图 2 本文方法推荐后用户对推荐信息的偏好程度

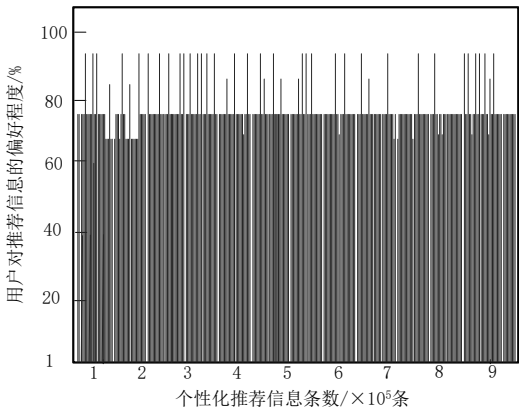


图 3 文献 [1] 方法推荐后用户对推荐信息的偏好程度

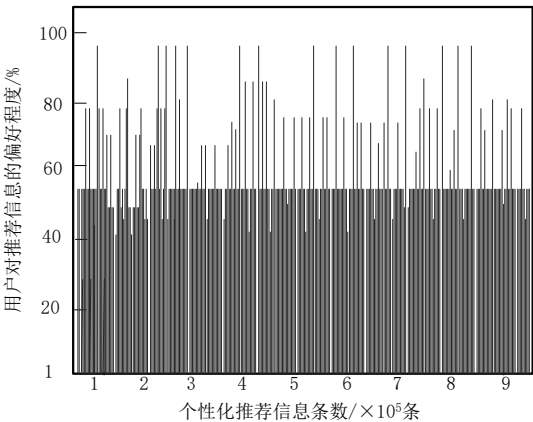


图 4 文献 [2] 方法推荐后用户对推荐信息的偏好程度

从图 2 可以看出，本文方法进行个性化推荐后，用户对

推荐信息的偏好程度在 80%~100% 之间，这显示出用户对推荐信息高度认可，也进一步证明该方法推荐的信息与用户需求信息相似度较高。应用文献 [1] 方法进行信息推荐后，用户偏好程度在 60%~80% 之间，相较于本文方法偏好程度有所降低，说明其推荐信息与用户需求的匹配程度相对一般。而从图 4 结果来看，用户对大部分推荐信息的偏好程度主要集中在 40%~60% 之间，表明这些推荐信息与用户需求的匹配程度较低，难以满足用户对信息的个性化需求。

在此基础上，对 3 种方法为用户推荐的个性化信息量进行了汇总，具体结果如表 3 所示。

表 3 不同方法用户推荐的信息量对比

用户编号	推荐的信息量 (×10 ⁵ 条)		
	本文方法	文献 [1] 方法	文献 [2] 方法
1	6.5	5.1	4.8
2	5.2	4.6	4.5
3	6.1	5.9	6.0
4	3.6	3.2	3.4
5	4.5	3.9	4.3
6	3.6	3.5	3.2
7	3.7	2.9	3.5
8	3.5	3.4	3.4
9	6.9	6.8	6.2
10	4.2	4.0	4.1

从表 3 所示的实验结果可以看出，在为用户进行目标信息的个性化推荐时，本文方法推荐的个性化信息量最多，而文献 [1] 和文献 [2] 方法推荐的个性化信息量相对较少。这一成果验证了本文方法的有效性。具体而言，该方法不仅显著提升了推荐信息与用户需求的匹配度，还大幅度增加了个性化信息的推荐数量，从而更好地满足了用户对网络信息多元化的需求。本文方法的优势核心在于充分利用了图数据库的结构化存储和高效查询能力，深入挖掘了用户与商品之间的潜在关联，并成功提取出了有价值的目标信息。这一关键步骤直接奠定了后续推荐过程有效性和个性化信息推荐量的坚实基础，使得最终推荐的个性化信息量达到了最高水平，进一步验证了本文方法的卓越性能和实际应用价值。

5 结语

个性化推荐系统的核心在于深度挖掘与分析用户数据，以便精准捕捉用户偏好。然而，现有的推荐方法在用户和商品交互数据稀缺或者面临新用户与商品冷启动的情况下，存在推荐信息与用户需求适配度低的问题。为此，本文提出一种基于图数据库的用户目标信息个性化推荐方

法。实验结果表明，该方法不仅有效解决了数据稀疏性和冷启动问题，还显著提升了推荐的精准度与多样性。展望未来，将继续深化这一领域的研究，不断探索图数据库技术在个性化推荐中的新应用与潜力，以期为用户带来更加丰富、精准、个性化的服务体验，进一步推动互联网个性化服务行业的蓬勃发展。

参考文献：

[1] 赵莉苹. 基于用户行为特征的网络教学云平台资源个性化推荐方法 [J]. 无线互联科技, 2024, 21 (18): 116-118.

[2] 李智彪, 江民财, 赵化勇, 等. 融合多维度相似算法的中药复方个性化煎煮推荐研究 [J]. 世界中医药, 2024, 19 (16): 2392-2398.

[3] 冼晓茵. 大数据新闻个性化推荐算法优化策略与用户体验影响研究 [J]. 新闻研究导刊, 2024, 15 (18): 11-15.

[4] 黄英辉, 刘辉, 王伟军, 等. 心理图式在应对推荐系统用户过度特化问题中的作用: 基于机器学习的实证研究 [J]. 图书与情报, 2024 (4): 118-130.

[5] 徐远宏. 基于协同过滤的农业平台学习资源个性化推荐方法探究 [J]. 科技资讯, 2024, 22 (16): 174-176.

[6] 郭羽宁, 景莉莉. 基于大数据分析的个性化推荐在吉林省乡村旅游直播中的运用研究 [J]. 产业科技创新, 2024, 6 (4): 27-30.

[7] 张峰涛. 基于用户画像的智慧图书馆文献资源个性化推荐方法研究 [J]. 数字通信世界, 2024(7): 34-36.

[8] 王鹤琴, 朱珍元. 国内个性化推荐技术及其算法与数据风险的研究现状和趋势 [J]. 安徽警官职业学院学报, 2024, 23 (4): 101-106.

[9] 宋徐春, 周济春, 吕旭东. 基于深度学习的乳腺癌术后脂质体多柔比星与表柔比星个性化治疗推荐研究 [J]. 复旦学报 (医学版), 2024, 51 (4): 443-454.

[10] 汪莉, 范博韬, 张志敏, 等. 基于客户特征提取的保险需求挖掘与产品个性化推荐方法 [J]. 数字技术与应用, 2024, 42 (7): 128-130.

【作者简介】

林楠 (1980—), 女, 吉林长春人, 硕士, 副教授, 研究方向: 大数据和人工智能。

(收稿日期: 2024-11-04)