

引文格式：

仇阿根, 张用川, 罗宁, 等. 结合用户特征的政务服务协同过滤推荐方法 [J]. 集成技术, 2023, 12(1): 42-55.

Qiu AG, Zhang YC, Luo N, et al. Government service collaborative filtering recommendation method based on user characteristics [J]. Journal of Integration Technology, 2023, 12(1): 42-55.

结合用户特征的政务服务协同过滤推荐方法

仇阿根¹ 张用川^{2*} 罗宁³ 郑莹莹³ 陆文³

¹(中国测绘科学研究院 北京 100830)

²(重庆交通大学 重庆 400074)

³(中国科学院软件所 北京 100190)

摘要 为推荐政务服务相关事项, 提高用户办事效率与政府服务水平, 该文提出一种推荐算法, 即结合用户特征的政务服务协同过滤推荐方法。该方法为解决传统协同过滤未考虑用户属性的问题, 将用户画像技术与其相结合。首先, 建立政务服务用户画像; 然后, 采用奇异值度量分析方法融合用户画像与基于用户的协同过滤算法, 使特征属性参与相似度计算, 改善用户之间的相似性, 并解决数据稀疏性的问题, 使推荐结果更具实际意义; 最后, 计算政务服务事项预测得分, 将得分最高的 TOP-N 推荐给用户。在实验部分, 该文利用某市企业法人的政务服务真实数据进行验证。结果显示, 该算法能够满足政务服务推荐的个性化要求, 预测准确性较高。

关键词 政务服务; 个性化推荐; 用户画像; 空间协同过滤; 地理位置信息

中图分类号 P 2 文献标志码 A doi: 10.12146/j.issn.2095-3135.20220715004

Government Service Collaborative Filtering Recommendation Method Based on User Characteristics

QIU Agen¹ ZHANG Yongchuan^{2*} LUO Ning³ ZHENG Yingying³ LU Wen³

¹(Chinese Academy of Surveying and Mapping, Beijing 100830, China)

²(Chongqing Jiaotong University, Chongqing 400074, China)

³(Institute of Software, Chinese Academy of Sciences, Beijing 100190, China)

*Corresponding Author: 532879324@qq.com

Abstract In order to recommend matters related to government services and improve user efficiency and government service level, a recommendation algorithm is proposed, that is, a collaborative filtering recommendation method for government services combined with user characteristics. Unlike traditional

收稿日期: 2022-07-15 修回日期: 2022-09-23

基金项目: 国家重点研发计划项目(2019YFB2102503)

作者简介: 仇阿根, 研究员, 研究方向为政务大数据服务和数据融合; 张用川(通讯作者), 研究员, 研究方向为地理信息系统和数据整合, E-mail: 532879324@qq.com; 罗宁, 副研究员, 研究方向为政府数据治理、金融信息安全和分布式计算; 郑莹莹, 中级工程师, 研究方向为软件测试和分布式系统; 陆文, 助理工程师, 研究方向为计算机软件与理论。

collaborative filtering which does not consider user attributes, this method combines user portrait technology with it. First, the method establishes a user portrait of government services, and then uses the singular value metric analysis method to integrate the user portrait and the user-based collaborative filtering algorithm, so that the feature attributes can participate in the similarity calculation, improve the similarity between users, and solve the problem of data sparsity. To make the results more practical, the method calculates the predicted government service score, and recommends the TOP-N with the highest score to the user. In the experimental part, the actual data of the government affairs service of a city's enterprise legal person is used for verification. The results show that the algorithm can meet the personalized requirements of the government affairs service recommendations and improve the prediction accuracy.

Keywords government service; personalized recommendation; user portrait; spatial collaborative filtering; geographic location information

Funding This work is supported by National Key Research and Development Program of China (2019YFB2102503)

1 引言

随着人类获取地理空间位置能力和精度的提升, 地理信息被广泛应用于各个领域。由于技术进步带来的数据增长, 在数据应用过程中, 出现了数据过载问题, 研究者们期望利用推荐系统解决该问题并取得了一定的进展。近年来, 党中央、国务院高度重视“互联网+政务服务”工作, 全国各地相继开展了智慧政府的建立及相应模式的创新, 不断提升自身的行政能力和服务能力^[1]。在实际生产场景中, 目前存在的问题有: (1) 政务服务事项繁杂、分类众多、覆盖面广, 给服务提供带来了困难; (2) 政务服务面向社会大众, 无论是企业还是个人, 因其自身差异, 都需要个性化的信息服务; (3) 随着大数据时代的发展, 用户难以在眼花缭乱的数据中找到自己所需要的信息^[2-3]。虽然政府系统、网站采用搜索引擎技术解决上述问题, 并取得了一定成效, 但传统的搜索引擎无法有效地整合政务服务中的个人特征, 政务服务信息同质化程度较高, 用户需要反复地人工过滤搜索的信息, 才能找到真正“合适”的信息^[4]。随着我国政务服务水平的不

断提升, 为公众提供个性化的政府资源推荐服务尤为重要。众多专家学者将目光投向用户画像领域——通过对用户信息的描述与提取, 将其作用于政务信息资源的个性化推荐。在大数据环境下, 用户画像技术能够全面细致地把握用户的兴趣和需求, 为用户提供个性、优质的政务服务信息资源^[5-6]。

在政府信息研究方面, 李娇等^[7]通过 Apriori 算法抽取用户访问页面之间的关联规则, 为个性化政务信息服务提供了有力的数据支持。胡海波^[8]构建了基于“LBS+SNS”的移动政务个性化信息服务模型, 分析了移动政务的信息服务模式, 并指出政务服务观念应从“人找服务”转变为“服务找人”。用户画像技术可以将用户诉求与设计方向相结合, 转化联结用户属性数据, 以勾画目标用户^[9]。RM Quintana 以用户的检索行为、兴趣偏好等个人信息为基础, 构建了 MOOC 用户画像模型^[10]。盛姝等^[11]利用医享网的患者数据, 基于用户角色属性、行为属性及文本特征, 构建了典型用户识别指标, 将用户群体分为 4 类, 构建画像概念模型; 再利用用户行为识别算法与主题聚类算法, 进一步挖掘用户处于不同角色关注的

主题内容,实现用户信息需求的精准分析^[11]。随着电子政务的深入发展和“互联网+政务服务”工作的推进,政务信息资源总量不断增长,用户需求越来越多样化、个人化。政府个性化信息服务应主动考虑用户需求、重新定位,以用户行为信息为服务导向,以用户需求为核心,主动收集用户属性及感兴趣的信息,预测用户的未来发展趋势,针对用户的信息需求,提供最贴切的信息服务。在不了解用户偏好的情况下,协同过滤推荐算法仅利用用户对项目的历史行为数据,就可预测用户对未知项目的评分,从而为用户进行推荐。杨峰提出的电子政务信息推荐服务框架则采用了信息协同过滤技术,向公众主动提供适合用户的信息组合,并能够较好地把握用户需求^[12-14]。

目前,国内政务信息资源个性化服务在推荐技术方面进行了一定研究,但以用户为中心进行导向型服务的相关研究较少。大部分研究针对信息资源处理方式进行改进和完善,且传统的个性化推荐方式以内容过滤或协同过滤为主,在推荐精度上仍存在偏差。由于政务服务的独特性,上述研究方法存在以下弊端^[15-16]:(1)在使用协同过滤算法时,未考虑用户体量大、存在热门信息等因素,其中,冷门信息将导致得分矩阵分布极端稀疏;(2)未综合考量用户属性与用户行为信息。为解决上述问题,本文提出了一种将用户画像与协同过滤算法进行融合的方法。具体步骤为:首先建立用户画像标签,然后信息量化用户画像,最后将量化信息填充到协同过滤算法的得分矩阵中,并参与推荐计算。本方法可以在考虑用户属性信息、兼顾用户行为信息的同时,解决得分矩阵分布极端稀疏的问题,提高推荐精度。

2 相关技术与理论

2.1 基于用户的协同过滤算法

协同过滤算法不需要了解用户偏好,仅通

过用户对项目的历史评分数据,就可预测用户对未知商品的评分,从而进行推荐。协同过滤算法简单、有效,在很多领域的推荐系统中得到了大量实际应用,其主要分为基于用户和基于项目的协同过滤算法。基于用户的协同过滤算法从用户对项目的兴趣出发,寻找与该用户较为相似的用户,然后根据相邻用户的行为数据,对项目做出选择。基于项目的协同过滤算法则是基于两个项目的相似程度与用户的行为数据,给推荐用户推荐可能需要的项目。在本文提及的政务服务个性化推荐方法中,建立用户画像是为了更精准地计算用户相似度,因此,本文采用了基于用户的协同过滤方法^[17]。基于用户的协同过滤算法的根本原则为:利用用户共同行为的相似程度大小,决定拟被推荐用户可能有意愿选择的项目。基于用户的协同过滤算法直接搜索相同的目标用户资料,以估计目标用户的选择偏好。具体地,根据被推荐用户和其他用户的相关性,在整个用户库中寻找具有关联的用户。本质上,基于用户的协同过滤算法就是利用目标用户间的相似程度进行项目选择,较为常见的相似度计算方法有 Pearson 相似度、Jaccard 相似度、余弦相似度和在此基础上进行改进的余弦相似度等。当相关数据不够密集时,改进的余弦相似度表现较好。因此,本研究选择改进的余弦相似度计算用户之间的相似度,改进的余弦相似度计算公式如公式(1)所示^[18]。

$$sim(i,i') = \frac{\sum_{u \in U} (r_{i,u} - \bar{r}_i)(r_{i',u} - \bar{r}_{i'})}{\sqrt{\sum_{u \in U} (r_{i,u} - \bar{r}_i)^2} \sqrt{\sum_{u \in U} (r_{i',u} - \bar{r}_{i'})^2}} \quad (1)$$

其中, $sim(i,i')$ 为用户*i*与*i'*之间的相似度,取值范围为[0,1],其值越大,用户*i*与*i'*越相近; U 为用户存在交集的项目集合; $r_{i,u}$ 为用户*i*对*U*的评分; $r_{i',u}$ 为用户*i'*对*U*的评分; \bar{r}_i 为用户*i*对其所有项目评分的平均值; $\bar{r}_{i'}$ 为用户*i'*对其所有项

目评分的平均值。

当最相似的邻居确定后, 利用公式(2)对用户 i 与用户 k 的并集项目中未评分项目 w 进行预测, 其中, k 为用户 i 相似度最大的前 k 个用户。

$$p(i,v)=\bar{r}_i+\frac{\sum_{w\in W}(r_{w,v}-\bar{r}_i)\times sim(i,w)}{\sum_{w\in W}sim(i,w)} \quad (2)$$

其中, W 为前 k 个相似用户的集合; $r_{w,v}$ 为 W 集合中用户 w 对项目 v 的评分; $sim(i,w)$ 为用户 i 与 w 的相似度。首先, 计算相似度加权的所有邻居偏好, 并在此基础上加上目标用户 i 的平均评分; 然后, 将评分进行排序, 将得分最高的几个项目推荐给目标用户。

2.2 用户画像技术

用户画像概念最早由交互设计之父 Alan Cooper 提出, 指通过建立真实用户信息的标签模型, 以实现用户信息的抽象化表达。该模型基于一系列的正式数据, 同时使用虚拟的方式代表用户数据^[19]。一般地, 用户画像的构建只需遵循两个基本原则: 首先是从具体的服务情景出发, 针对性地解决实际问题; 其次是按照用户的独有特点和特征进行设计。政务信息服务用户画像指收集用户使用政务信息服务过程中的数据, 数据收集应尽可能全面且具体, 包含用户的基本信息及其真实的动态数据情况等, 可分别从静态和动态的角度出发, 更好地掌握和收集用户的数据^[20]。在收集的数据基础上, 挖掘出表示用户特征的关键性标签, 并在这种共性指导下, 给予个性的特征以独特的呈现方式, 利用信息的行为规律, 实现政务信息资源的深层次个性化推荐服务^[21-22]。用户画像的构架流程, 可从以下 3 个方面进行阐述:

(1) 数据收集: 用户数据量大且来源多样, 可根据用户画像的不同特征属性, 将用户数据分成多个维度。目前, 用户画像数据收集手段主要包括社会调查、网络数据收集和平台数据库收集 3 种方式。具体指: 通过社会调查法中的访谈、

调查等方式收集数据; 利用网络爬虫等技术手段合法获取用户公共数据; 直接从平台数据库收集用户数据。例如, 通过各类管理系统、数字资源服务系统以及微博、微信等相关移动平台的用户数据库, 直接收集各类用户数据。

(2) 特征提取: 为确保用户画像数据的完整性, 避免脏数据影响模型构建, 需要对用户数据进行整理和分类, 并通过一定的数据挖掘方法提取用户特征及用户标签, 并构建用户画像。目前, 用户肖像研究主要采用人工提取和技术提取两种方法进行特征提取。

人工提取: 在相关理论、研究者的知识和经验的支持下, 通过文献综述、研究、访谈和专家建议, 对抽象用户属性进行描述, 从而提取用户特征, 构建用户画像。

技术提取: 利用机器学习算法(如决策树、逻辑回归和支持向量机等)提取用户特征, 通常这些算法适用于大数据背景下海量用户数据的研究场景, 如利用 LDA 模型文本挖掘用户感兴趣的微博主题, 获得用户偏好主题。

(3) 形成用户画像: 完成数据收集和特征提取后, 利用机器学习算法构建模型, 将处理后的结果转化为用户标签, 用各种直观、清晰的视觉图形呈现用户画像。目前, 常用的表现形式包括标签云、人物图像、统计图、直方图、雷达图等, 可根据实际需求创建。

本文从某市平台数据库中收集企业用户的基本信息与行为信息, 利用人工提取的方式, 对企业法人用户的用户画像进行构建, 具体画像信息如表 1 和图 1 所示。

3 方法改进

目前, 传统协同过滤方法在政务服务个性化推荐的研究中, 仅利用用户行为信息即用户得分矩阵计算用户间相似度, 没有考虑用户地理位置

表 1 用户画像信息

Table 1 User portrait information

| 用户 | | 画像信息 | | |
|------|------|------|--------|---------|
| 企业法人 | 基本信息 | 地理位置 | 生产经营信息 | 企业的评价信息 |
| 自然人 | 基本信息 | 地理位置 | 社会社交属性 | ... |

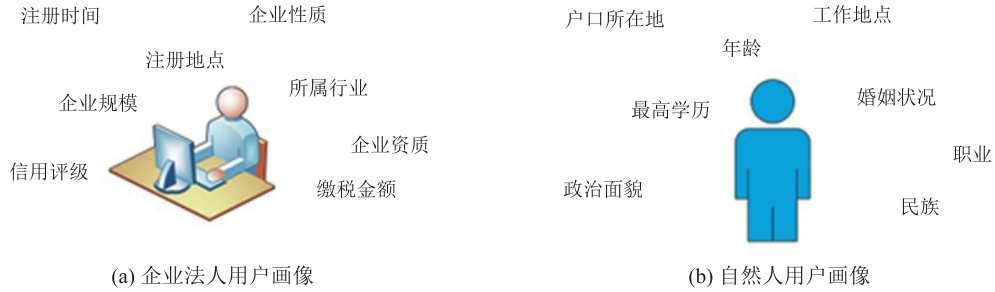


图 1 用户画像

Fig. 1 User portrait

等属性信息，推荐精度可能较差。此外，政务服务用户体量大，存在热门事项信息和冷门事项信息等特征，可能导致得分矩阵分布极端稀疏^[19]。

为解决上述问题，本文从两方面着手：一方面，根据推荐业务的特点结合地理位置信息，建立政务服务用户画像与自然人政务服务用户画像，获取用户标签，为后续推荐计算中结合用户地理位置等属性信息奠定基础；另一方面，从用户相似度计算的方面着手，改进传统协同过滤算法，将建立好的用户标签在得分矩阵中进行标记，量化用户地理位置等属性信息，同时对用户评分进行修正，提出利用用户行为信息、地理位置等属性信息，构建用户特征矩阵，用于计算用户间相似度，该矩阵的填充值即用户特征，是用户属性信息与用户得分融合计算的结果。改进原理如图 2 所示。

3.1 度量奇异值权重下的政务服务用户行为特征偏好

Bobadilla 等^[19]曾提出一种基于奇异性的相似度度量模型，以区分用户评分的相关性。其基本思想为：将评分较高的项目定义为正向项目，评分较低的为负向项目，若存在两用户对某项目的

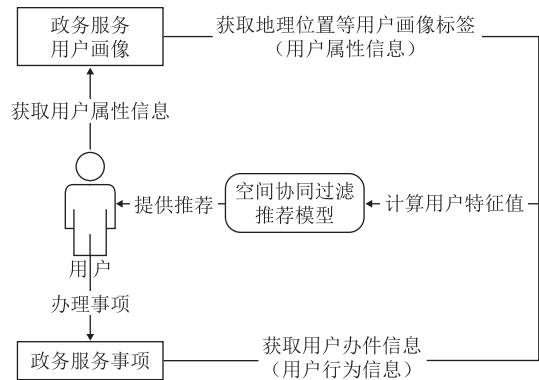


图 2 政务服务推荐算法流程

Fig. 2 Government service recommendation algorithm flow

评分都为正向，而其他大部分用户对其评分为负向，那么此项目对于两个用户的相似度计算具有较高的参考价值，将该项目的影响因素与传统协同过滤方法相结合，用于计算用户之间的相似度，在预测项目得分的同时，进行项目推荐。

在该方法中，针对不同用户对各项目的评分情况，分别为其标记“正向”标签与“负向”标签，以期实现用户行为的分类及数据表达。将用户属性与得分情况综合考量后，对办理事项进行标记，从而量化用户属性信息，以解决传统协同过滤方法中未考虑用户属性信息的问题，提升政

府服务事项推荐的准确率。

假设企业法人用户-办理事项评分矩阵如表 2 所示。通过构建用户画像, 得到 5 个企业法人用户的产业类型信息, 将此信息与企业法人用户-办理事项评分矩阵相结合, 把用户所属的产业类型标签标记在其评分矩阵的对应位置上, 得到如表 3 所示的企业法人用户-产业类型矩阵。

R 代表房地产业, F 代表金融业, 若某办理事项中 R 标签较多, 则表明该办理事项对于房地产业类型企业办理意愿更大, 反之则是金融业类型企业办理意愿更大。

如表 3 所示, 定义 R_i 为办理事项 i 的房地产业用户集合, F_i 为办理事项 i 的金融业用户集合。其中, $R_1=\{1,4\}$, $R_2=\{1\}$, $R_3=\{1,5\}$, $R_4=\{1,4\}$, $R_5=\{1,5\}$; $F_1=\{2,3\}$, $F_2=\{3\}$, $F_3=\{2\}$, $F_4=\{2,3\}$, $F_5=\{3\}$ 。

定义 I_R^i 为办理事项 i 的房地产业指数, I_F^i 为办理事项 i 的金融业指数, 计算公式如公式(3)与公式(4)所示。

$$I_R^i = \frac{N(R_i)}{N(U_i)} \quad (3)$$

其中, $N(R_i)$ 为办理事项 i 评分的房地产业类型用户数量; $N(U_i)$ 为办理事项 i 评分的所有用户数量。由公式(3)可计算得出: $I_R^1=0.5$, $I_R^2=0.5$, $I_R^3=0.667$, $I_R^4=0.5$, $I_R^5=0.667$ 。

$$I_F^i = \frac{N(F_i)}{N(U_i)} \quad (4)$$

其中, $N(F_i)$ 为对办理事项 i 评分的金融业类型用户的数量; $N(U_i)$ 为办理事项 i 评分的所有用户数量。由公式(4)可计算得出: $I_F^1=0.5$, $I_F^2=0.5$, $I_F^3=0.333$, $I_F^4=0.5$, $I_F^5=0.333$ 。

定义 M_R^k 为用户 k 办理房地产业相关事项的意愿, M_F^k 为用户 k 办理金融业相关事项的意愿。 M_R^k 与 M_F^k 的计算公式如公式(5)与公式(6)所示。

$$M_R^k = \frac{\sum_{i \in Item_k} r_{i,k} \times I_R^i}{N(Item_k)} \quad (5)$$

其中, M_R^k 为用户 k 的房地产业特征; $Item_k$ 为用户 k 办理的项目; $N(Item_k)$ 为用户 k 办理的项目总数。如公式(5)所示, M_R^k 可通过用户 k 每个办

表 2 企业法人用户-办理事项评分矩阵

Table 2 Enterprise legal person user-transaction score matrix

| 企业法人 | 用户评分 | | | | |
|------|--------|--------|--------|--------|--------|
| | 办理事项 1 | 办理事项 2 | 办理事项 3 | 办理事项 4 | 办理事项 5 |
| 用户 1 | 4 | 2 | 3 | 2 | 2 |
| 用户 2 | 5 | 0 | 5 | 4 | 0 |
| 用户 3 | 3 | 1 | 0 | 2 | 3 |
| 用户 4 | 4 | 0 | 0 | 2 | 0 |
| 用户 5 | 0 | 0 | 3 | 0 | 4 |

表 3 企业法人用户-产业类型矩阵

Table 3 Corporate user-industry type matrix

| 企业法人 | 办理事项 1 | 办理事项 2 | 办理事项 3 | 办理事项 4 | 办理事项 5 |
|------|--------|--------|--------|--------|--------|
| 用户 1 | R | R | R | R | R |
| 用户 2 | F | 0 | F | F | 0 |
| 用户 3 | F | F | 0 | F | F |
| 用户 4 | R | 0 | 0 | R | 0 |
| 用户 5 | 0 | 0 | R | 0 | R |

理项目 i 的评分与项目 i 的房地产业指数的乘积求和, 最后取平均值进行计算。以用户 4 为例,

$$M_R^4 = \frac{4 \times 0.5 + 2 \times 0.5}{2} = 1.5.$$

$$M_F^k = \frac{\sum_{i \in \text{Item}} r_{i,k} \times I_F^k}{N(\text{Item}_k)} \quad (6)$$

其中, M_F^k 为用户 k 的金融业特征。 M_F^k 可通过用户 k 每个办理项目 i 的评分与项目 i 房地产业指数的乘积求和, 最后取平均值进行计算。以用户 4 为

$$M_F^4 = \frac{4 \times 0.5 + 2 \times 0.5}{2} = 1.5.$$

公式(3)~(6)均基于企业法人用户产业属性中的两项特征展开说明, 对上述公式进行推广, 可得到办理事项指数公式与用户特征公式, 如公式(7)和公式(8)所示。

$$I_R^i = \frac{N(L_i)}{N(U_i)} \quad (7)$$

$$M_L^k = \frac{\sum_{i \in \text{Item}} r_{i,k} \times I_L^k}{N(\text{Item}_k)} \quad (8)$$

其中, L 为用户画像得到的所有用户属性的统称。

3.2 结合用户特征的政务服务协同过滤推荐方法

本文推荐算法融合了用户行为信息、地理位置等属性信息, 提高了推荐准确率, 改进了传统协同过滤算法关于用户相似度的计算方式, 在用户-得分矩阵的基础上, 利用用户-特征矩阵, 提出了空间协同过滤算法, 其具体算法步骤如图 3 所示。

本文提出的结合用户特征的政务服务协同过滤推荐算法在构建用户评分矩阵时, 使用了与传统协同过滤算法相同的方式。在用户相似度计算方面, 本文算法将用户评分属性信息相结合, 利用第 2.2 节提出的用户特征, 构建如表 4 所示的政务服务用户-特征矩阵。

空间特征值 M_g^k 可综合用户评分与地理位置信息得到, M_g^k 值越大, g 标签与用户 k 的地理关联性越高, 其计算公式如公式(9)所示。

$$M_g^k = \frac{\sum_{i \in \text{Item}} r_{kj} \times I_g^j}{N(\text{Item}_k)}, I_g^j = \frac{N(g_j)}{N(U_j)} \quad (9)$$

其中, g 为用户地理位置标签; r_{kj} 为用户 k 对办

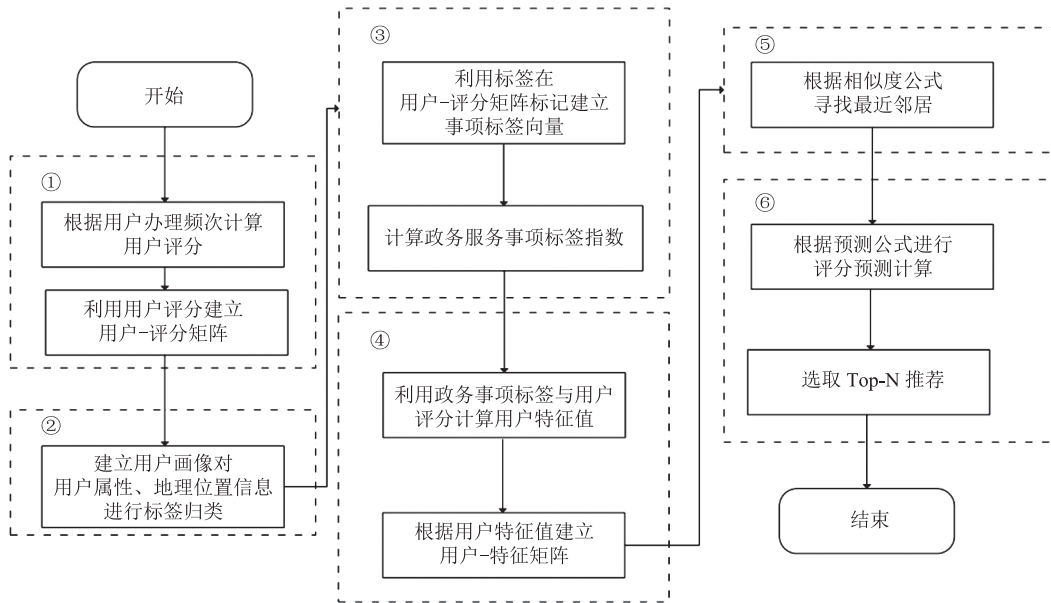


图 3 空间协同过滤算法流程

Fig. 3 Spatial collaborative filtering algorithm flow

表 4 企业法人用户-用户特征矩阵

Table 4 Corporate user-user feature matrix

| | 特征 1 | 特征 2 | 特征 3 | ... | 特征 n |
|--------|---------|---------|---------|-----|---------|
| 用户 1 | M_1^1 | M_2^1 | M_3^1 | ... | M_n^1 |
| 用户 2 | M_1^2 | M_2^2 | M_3^2 | ... | M_n^2 |
| 用户 3 | M_1^3 | M_2^3 | M_3^3 | ... | M_n^3 |
| ... | ... | ... | ... | ... | ... |
| 用户 m | M_1^m | M_2^m | M_3^m | ... | M_n^m |

理事项 j 的评分, 即办理事项 j 的办理次数与用户 k 所有事项办理次数的比值, 评分越高, 表明该政务服务事项的办理频次越高, 与用户的业务联系越紧密; I_g^j 为办理事项 j 的地理位置 g 标签的标签指数, 表示办理事项 j 与地理位置 g 标签的地理关联程度; $N(g_j)$ 为对办理事项 j 进行评分的用户中, 拥有 g 标签用户的数量; $N(U_j)$ 为对办理事项 i 进行评分的所有用户数量。

然后采用改进的余弦相似度方法计算政务服务用户相似度, 计算公式如公式(10)所示。

$$sim(i, i') = \frac{\sum_{L \in Lable} (M_L^k - \overline{M_L})(M_L^{k'} - \overline{M_L})}{\sqrt{\sum_{L \in Lable} (M_L^k - \overline{M_L})^2} \sqrt{\sum_{L \in Lable} (M_L^{k'} - \overline{M_L})^2}} \quad (10)$$

其中, $\overline{M_L}$ 为 L 特征标签的平均值; $Lable$ 为所有标签特征的集合。后续算法流程与传统协同过滤的预测评分计算相同, 即相似度排序后, 将排名靠前的几个用户代入公式(2)中计算预测得分, 选取 TOP-N 为用户推荐。

4 实验设计与分析

4.1 数据预处理

本文使用的数据均来源于某市政务服务平台与市场监督管理局, 由企业法人办件数据、企业法人用户信息数据、政务服务办理事项数据(政务服务事项实施清单)3 部分构成。根据实际情况对数据进行清洗处理。剔除重要字段缺失、错误

或其他可能影响实验结果的不良数据后, 选取办件数量排名前 10 000 的企业法人用户及其 55 000 余条办件数据, 构建企业法人用户-办理事项评分矩阵, 评分值区间为 $[0,1]$, 部分评分数据如表 5 所示。

表 5 企业法人用户-办理事项评分数据(部分)

Table 5 Score data of enterprise legal person user-matters

| handled (part) | | |
|----------------|---------------------------------|-------|
| 用户 ID | 事项实施编码 | 评分 |
| 1 | 11110000000021557L2000722008000 | 0.784 |
| 2 | 1211000040070983672112076007005 | 0.685 |
| 3 | TEXC11112222333013000715001014 | 0.674 |
| 4 | 1111000078250093063000715001014 | 0.631 |
| 5 | 11110000000021063W2110116057037 | 0.414 |

利用企业法人信息数据构建用户画像, 主要分为风险评级、产业类型、公司性质、注册资金、行政区划、注册时间 6 个大类, 部分标签内容如表 6 所示。根据用户画像标签, 可建立如表 7 所示的用户信息表。

为满足实验训练与检验算法的需要, 将数据分为训练集与测试集两部分, 其中 80% 为训练集, 20% 为测试集。利用企业法人用户-办理事项评分数据, 将训练集数据结合第 3.1 节提及的度量奇异值权重下的政务服务用户行为特征偏好方法, 构建企业法人用户-用户特征矩阵, 至此数据处理工作结束。

4.2 推荐算法评价指标

一方面, 本文采取平均绝对误差 (Mean Absolute Error, MAE) 作为实验结果的评价指标, 其主要根据实验所获取的预测评分与实际用户评分的差值大小来判定算法的优劣性。差值越小, 算法性能越优; 反之则算法性能越差。计算公式如公式(11)所示^[23]。

$$MAE = \frac{\sum_{i=1}^n |y_{ui} - \widehat{y}_{ui}|}{n} \quad (11)$$

其中, n 为用户 u 的评分项目数量; y_{ui} 为用户 u

表 6 企业法人用户标签数据(部分)

Table 6 Corporate user label data (part)

| 标签大类 | 具体属性 |
|------|-----------------------------------|
| 风险评级 | 优 |
| | 良 |
| | 中 |
| | 差 |
| | 不合格 |
| 产业类型 | 农、林、牧、渔业 |
| | 采矿业 |
| | 制造业 |
| | 电力、热力、燃气及水生产和供应业 |
| | 建筑业 |
| | 批发和零售业 |
| | 交通运输、仓储和邮政业 |
| 公司性质 | 住宿和餐饮业 |
| | ... |
| | 有限责任公司 |
| | 股份有限公司 |
| | 外商投资合伙企业 |
| 注册资金 | 联营 |
| | 集体所有制 |
| | ... |
| | <100 万元 |
| | 100 万元 |
| 行政区划 | 100 万元~500 万元(不包含 100 万元与 500 万元) |
| | 500 万元 |
| | 500 万元~1 000 万元(不包含 500 万元) |
| | ... |
| 注册时间 | 东城区 |
| | 朝阳区 |
| | 西城区 |
| | 海淀区 |
| | 丰台区 |
| 注册时间 | ... |
| | 2000 年之前 |
| | 2000—2005 年(不包含 2005 年) |
| | 2005—2010 年(不包含 2010 年) |
| 注册时间 | 2010—2015 年(不包含 2015 年) |
| | ... |
| | ... |

的实际评分; \hat{y}_{ui} 为用户 u 的预测评分。

另一方面, 由于推荐类算法可看作是一种二

元分类问题(推荐与不推荐), 所以也常采用准确率、召回率、 F_1 值作为评价模型优劣的指标^[17-18]。

其中, 准确率(Precision)是指在所有被推荐项目中, 用户偏爱项目所占的比率; 召回率(Recall)是指在所有被推荐项目中, 用户偏爱项目的总数与总项目集合中用户感兴趣项目总数的比率; F_1 值是模型准确率和召回率的调和平均, 其作为一种综合性的评价指标, 可以更加全面直观地评价算法的整体性能。

$$Precision = \frac{Item(rel)}{Item(rel) + Item(nrel)} \quad (12)$$

$$Recall = \frac{Item(rel)}{N(rel)} \quad (13)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

其中, $Item(rel)$ 为所有被推荐项目中, 用户感兴趣的项目总数; $Item(nrel)$ 为所有被推荐项目中, 用户不感兴趣的项目总数; $N(rel)$ 为总项目集合中用户感兴趣的项目总数。

为进一步比较基于用户空间协同过滤的政务服务推荐方法的优劣, 基于相同的政务数据和服务推荐流程, 研究采用传统协同过滤对面向企业法人的政务数据进行服务推荐。传统协同过滤算法通过计算相似度, 依据获取的预测评分实现个性化推荐。

4.3 实验结果

4.3.1 推荐结果分析

利用上述方法为企业法人进行政务服务推荐。推荐用户是某投资担保公司, 设立日期为 2013 年, 主营业务是信贷担保、票据承兑保证、外贸融资保证、工程投资保证, 其所属产业大类为金融业, 企业注册地为西城区。在参数设置中将邻近值 K 设置为 10, 并选取预测得分 TOP-10 进行推荐, 结果如表 8 所示。

由表 8 可知, 前 8 条事项与该公司的业务范

表 7 企业法人用户信息表(部分)

Table 7 Corporate user information table (part)

| 用户 ID | 风险评级 | 产业类型 | 公司性质 | 注册资金(万元) | 行政区划 | 注册时间 |
|-------|------|-------|---------|-----------|------|-------------|
| 1 | 优 | 批发零售业 | 有限责任公司 | <100 | 海淀区 | 2000 年之前 |
| 2 | 良 | 采矿业 | 股份有限公司 | 500~1 000 | 东城区 | 2010—2015 年 |
| 3 | 优 | 住宿餐饮业 | 股份有限分公司 | 500~1 000 | 丰台区 | 2015—2019 年 |
| 4 | 优 | 建筑业 | 股份有限公司 | <100 | 海淀区 | 2015—2019 年 |
| 5 | 中 | 教育业 | 股份有限公司 | 100~500 | 朝阳区 | 2010—2015 年 |

表 8 用户推荐事项(部分)

Table 8 Users recommend matters (part)

| 事项名称 | 事项实施编码 | 事项所属业务范围 |
|--------------|---------------------------------|----------|
| 抵押权首次登记 | 1111000077950327643000715001014 | 企业法人 |
| 抵押权注销登记 | 11110000660504319Y3000715001014 | 企业法人 |
| 抵押权更正登记 | TEXC111122223333013000715001015 | 企业法人 |
| 房屋等建筑物所有权登记 | TEXC111122223333013000715001007 | 企业法人 |
| 融资担保公司名称变更备案 | TE1234567890000233000155002000 | 企业法人 |
| 融资担保公司监事变更备案 | 1111010255142682563000155002000 | 企业法人 |
| 融资担保公司管理人员变更 | 1111010868355911XU3000155002000 | 企业法人 |
| 集体建设用地使用权登记 | TEDT111122223333013000715001007 | 企业法人 |
| 住房公积金开户缴存 | 1211000040070983672112076007005 | 企业法人 |
| 建设工程消防验收 | 11110101069580736N3000117051000 | 企业法人 |

围有较大关联, 后 2 条事项属于各企业法人的相关通用办理事项。因此, 该推荐结果具有一定的参考意义。

4.3.2 评价指标结果分析

将本文方法与传统协同过滤算法和基于内容的推荐算法进行对比, 根据 MAE 分析最优邻居值的大小, 并比较两种算法的优劣性。其中, 最大邻居值以 10 为间隔进行选取, 范围为 [10,80]。

由图 4 可知, 随着最大邻居值的不断增加, 三者的 MAE 值初始时均呈下降趋势, 当最邻近数为 60 时, MAE 值逐渐趋于平稳, 达到最优状态。在本实验选取的邻近值范围内, 与传统协同过滤算法和基于内容的推荐算法相比, 本文方法的 MAE 值较小, 降低了约 5.3%, 稳定程度更高, 说明在一定程度上, 本文提出的算法可提升协同过滤算法的预测评分质量。

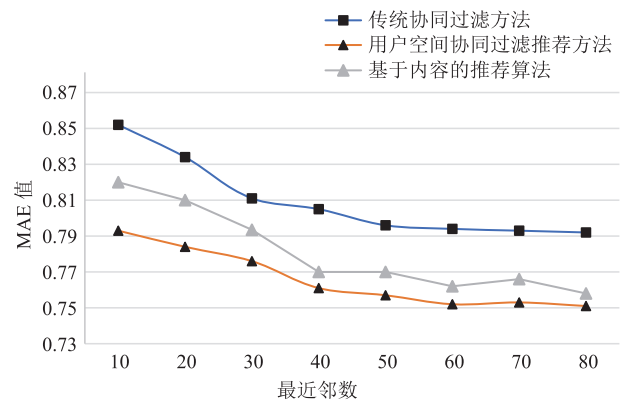


图 4 MAE 比较

Fig. 4 MAE comparison

由于准确率、召回率、 F_1 值为二元评分体系, 因此, 需要对实验结果进行评分转换, 以便分析本文实验数据。其中, 1~3 分为不相关项目, 4~5 分为相关项目。由图 4 可知, 设置最大邻居值 $K=60$, 在性能最优的状态下, 对不同算法的 3 个值进行分析, 用以比较两种算法的推荐

能力。

由图 5、图 6 和图 7 可知，与传统的协同过滤算法相比，本文提出的顾及位置与用户特征的政务服务协同过滤推荐方法的推荐效果较优。由

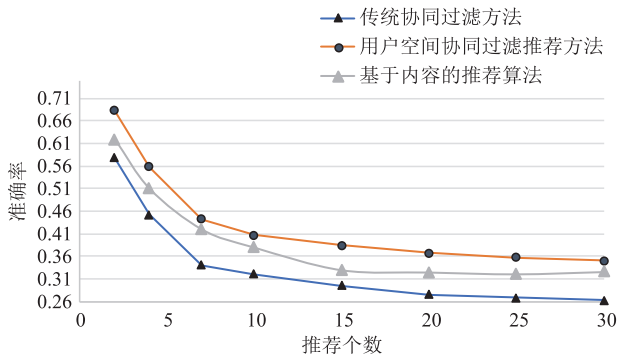


图 5 准确率比较

Fig. 5 Comparison of accuracy

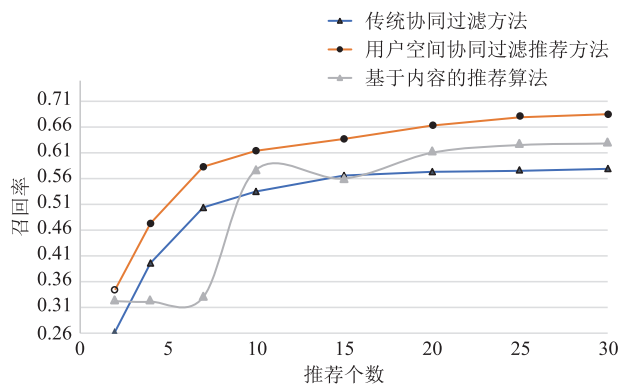


图 6 召回率比较

Fig. 6 Comparison of recall rates

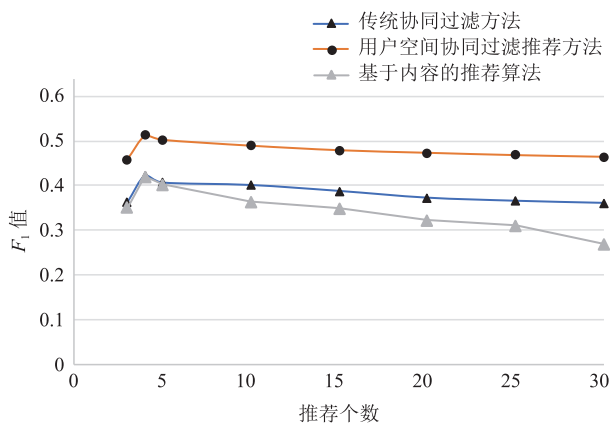


图 7 F_1 值比较

Fig. 7 Comparison of F_1 values

图 5 可知，随着推荐个数的增加，两种算法的准确率均不断降低，究其原因，当用户推荐个数不断增加时，参与计算的用户得分矩阵与用户特征矩阵的规模会随之扩大，导致了数据稀疏性的增加。虽然本文提出的算法在一定程度上可削弱稀疏性的影响，但并不能完全消除。两种算法的召回率与准确率呈相反趋势，随着推荐个数的增加，召回率不断上升，究其原因，随着实验推荐结果中政务服务事项的增加，其中含有用户感兴趣的事项也在增加，所以召回率呈现增加的趋势。随着推荐个数的增加，两种算法的 F_1 值均呈现先上升后下降的趋势，究其原因，是其变化趋势与准确率和召回率的变化速率有关。在传统的协同过滤中，少有空间信息参与计算。与无地理位置信息参与的推荐计算相比，有地理位置信息参与的推荐计算效果较优，这是由于政务服务事项推荐的业务与数据性质导致的。对于所有政务服务事项而言，由于地区不同、所属的行政级别不同，在政务服务业务中，完全相同的一项办理业务会被认为是不同的事项，拥有不同的统一事项实施编码，分条存储于政务服务事项库中。因此，在推荐计算中，若不考虑地理位置信息，就会为用户推荐多条相同但分属不同地区与行政级别的事项。因此通过增加用户以及事项的地理位置信息，可一定程度上减少该情况的发生，增加推荐精度。

综上所述，与传统协同过滤方法相比，本文方法在政务服务事项领域的性能更优，推荐效果更好。

4.3.3 地理位置信息因素影响

为验证地理位置信息对推荐效果的影响，本实验将处理完毕的数据进一步划分：一部分数据包包含用户的地理位置信息数据，另一部分不包含此类信息。同样利用准确率、召回率、 F_1 值 3 种二元评分指标，针对实验数据进行评分转换。其中，1~3 分为不相关项目，4~5 分为相关项

目, 设置最大邻近值 $K=60$ 开展实验, 并使用本文提出的一种结合地理位置与用户特征的政务服务协同过滤推荐方法, 以分析地理位置信息对推荐效果的影响。

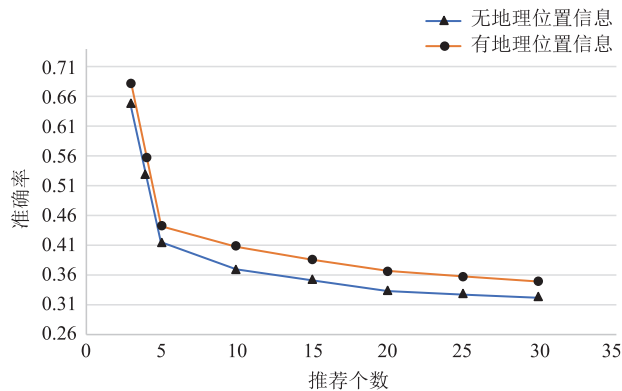


图 8 地理位置对准确率的影响

Fig. 8 The effect of location on accuracy

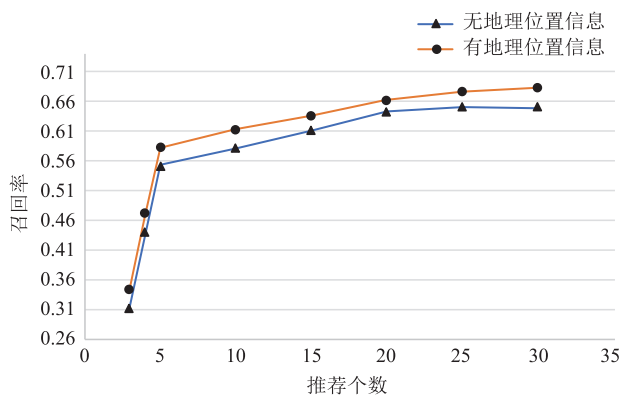


图 9 地理位置对召回率的影响

Fig. 9 Effect of location on recall ratio

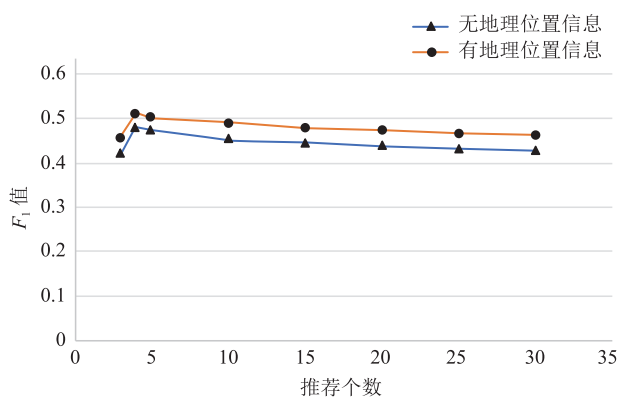


图 10 地理位置对 F_1 值的影响

Fig. 10 Effect of location on F_1 value

由准确率、召回率、 F_1 值 3 个评价指标的实验结果(如图 8、图 9 和图 10)可知, 与无地理位置信息参与推荐计算相比, 有地理位置信息参与的推荐计算效果较优, 这是政务服务事项推荐的业务与数据性质导致的。

5 结 语

政务信息资源的服务逐渐从大众化服务走向了个性化服务, 良好的政务推荐服务不仅让政府能够更好地服务公众, 而且公众能够更加及时地了解、清晰认识政府的方针政策, 二者良好交流关系的建立可促进社会的稳定发展。针对政务服务事项推荐, 本文提出一种结合用户特征的政务服务协同过滤推荐方法。该方法将用户画像技术与协同过滤技术相结合, 引入政务服务领域, 综合用户属性信息、用户位置信息、用户办件信息, 为用户推荐可能需要办理的政务服务事项。本文还构建了有地理位置信息的政务服务用户画像, 综合考虑用户属性与用户行为信息的同时, 解决了由于热门、冷门事项导致的数据稀疏性问题。实验结果显示, 与传统协同过滤方法相比, 基于用户空间协同过滤的政务服务推荐算法获得了较小的 MAE 值和较低的准确率、召回率、 F_1 值, 说明引入地理位置信息可以提升推荐精度。在后续工作中, 应进一步借鉴其他相关领域学科的经验, 提升画像构建的合理性; 其次, 本实验仅使用企业法人政务服务数据进行分析, 不足以构建较为完整的政务服务链条, 在对服务进行推荐时, 损失了一定的精度, 后续可利用多组不同来源以及类型数据提供更优质的服务, 构建完整的政务服务链条, 以期发现更具通用性与普适性的算法。

参 考 文 献

- [1] 王欣欣. 电子政务个性化信息服务研究 [J]. 图书

- 情报工作网刊, 2011, (5): 38-42.
- Wang XX. Research on e-government personalized information service [J]. *Journal of Library and Information Work*, 2011, (5): 38-42.
- [2] Zheng JF. Construction and application of music audio database based on collaborative filtering algorithm [J]. *Discrete Dynamics in Nature and Society*, 2022, 2022: 1756357.
- [3] 黄霞, 朱晓峰, 张琳. 个性化电子政务信息服务研究 [J]. *电子政务*, 2012, (2-3): 79-84.
- Huang X, Zhu XF, Zhang L. Research on personalized e-government information service [J]. *E-Government*, 2012, (2-3): 79-84.
- [4] Rafa T, Kechid S. Semantic representation of a geo-social user profile for a personalised information retrieval [J]. *Journal of Information & Knowledge Management*, 2021, 20(4): 1-26.
- [5] 宋羿弢. 基于用户情感画像的旅游推荐方法研究 [D]. 太原: 中北大学, 2021.
- Song YT. Research on travel recommendation method based on user emotional portrait [D]. Taiyuan: North Central University, 2021.
- [6] 武建新. 基于用户画像的可扩展协同过滤算法研究 [D]. 郑州: 郑州大学, 2021.
- Wu JX. Research on scalable collaborative filtering algorithm based on user portrait [D]. Zhengzhou: Zhengzhou University, 2021.
- [7] 李娇, 刘西林. 基于 Apriori 的电子政务个性化信息服务的实现 [J]. *情报理论与实践*, 2010, 33(1): 121-124.
- Li J, Liu XL. Realization of e-government personalized information service based on Apriori [J]. *Intelligence Theory and Practice*, 2010, 33(1): 121-124.
- [8] 胡海波. 基于 LBS 的移动政务个性化信息服务 [J]. *信息资源管理学报*, 2012, 2(2): 75-79.
- Hu HB. Study on personalized information services of mobile government based on LBS [J]. *Journal of Information Resource Management*, 2012, 2(2): 75-79.
- [9] 王凌霄, 沈卓, 李艳. 社会化问答社区用户画像构建 [J]. *情报理论与实践*, 2018, 41(1): 129-134.
- Wang LX, Shen Z, Li Y. User profiling of social Q&A community [J]. *Intelligence Theory and Practice*, 2018, 41(1): 129-134.
- [10] Quintana RM, Haley SR, Levick A, et al. The persona party: using personas to design for learning at scale [C] // *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2017: 933-941.
- [11] 盛姝, 黄奇, 郑姝雅, 等. 在线健康社区中用户画像及主题特征分布下信息需求研究——以医享网结肠癌圈数据为例 [J]. *情报学报*, 2021, 40(3): 308-320.
- Sheng S, Huang Q, Zheng SY, et al. Research on information needs under the distribution of user portraits and topic characteristics in online health communities: taking the colorectal cancer circle data of Medical Sharing Network as an example [J]. *Journal of Information*, 2021, 40(3): 308-320.
- [12] 潘锦丰, 叶东东, 谭北海, 等. 基于改进用户相似度的协同过滤算法 [J]. *现代计算机*, 2021, (21): 1-7.
- Pan JF, Ye DD, Tan BH, et al. Collaborative filtering algorithm based on improved user similarity [J]. *Modern Computer*, 2021, (21): 1-7.
- [13] 石磊, 李树青. 基于用户时点可见性的无趣项挖掘及协同过滤推荐方法 [J]. *数据分析与知识发现*, 2022, 6(5): 64-76.
- Shi L, Li SQ. Mining uninteresting items with visibility of user time points and collaborative filtering recommendation method [J]. *Data Analysis and Knowledge Discovery*, 2022, 6(5): 64-76.
- [14] 杨峰. 基于信息协同过滤的电子政务推荐服务系统研究 [J]. *情报探索*, 2014, (10): 79-81+86.
- Yang F. Research on e-government recommendation service system based on information collaborative filtering [J]. *Intelligence Exploration*, 2014, (10): 79-81+86.
- [15] 吴建东, 程永志, 杨帆. 基于 SOA 的电子政务个性化信息服务框架研究 [J]. *通信与信息技术*, 2018, (2): 24-26+23.
- Wu JD, Cheng YZ, Yang F. Research on e-government personalized information service framework based on SOA [J]. *Communication & Information Technology*, 2018, (2): 24-26+23.

- [16] 杨大寨. 基于本体理论的政务信息资源个性化应用研究 [D]. 合肥: 合肥工业大学, 2014.
Yang DZ. Research on personalized application of government information resources based on ontology theory [D]. Hefei: Hefei University of Technology, 2014.
- [17] 虞雅雯, 束静, 徐影. 基于用户的协同过滤算法对产品评分进行预测研究 [J]. 电脑编程技巧与维护, 2022, (5): 61-64.
Yu YW, Shu J, Xu Y. Research on product rating prediction based on a user-based collaborative filtering algorithm [J]. Computer Programming Skills and Maintenance, 2022, (5): 61-64.
- [18] 覃琼花. 基于协同过滤算法的个性化推荐系统研究 [J]. 科技资讯, 2022, 20(10): 4-6.
Qin QH. Research on personalized recommendation systems based on collaborative filtering algorithms [J]. Science and Technology Information, 2022, 20(10): 4-6.
- [19] Bobadilla J, Ortega F, Hernando A, et al. Recommender systems survey [J]. Knowledge-Based Systems, 2013, 46: 109-132.
- [20] Mohan KK, Shaik HUR, Srividya A, et al. White-box and black-box reliability modeling framework: integration through analytical model and user profile validation via deep learning—a practitioner’s approach [J]. International Journal of Reliability, Quality and Safety Engineering, 2021, 28(6): 2140007.
- [21] 章凌芳. 基于电子商务的用户画像推荐算法研究 [D]. 北京: 北京邮电大学, 2021.
Zhang LF. Research on user portrait recommendation algorithm based on e-commerce [D]. Beijing: Beijing University of Posts and Telecommunications, 2021.
- [22] 周妹璇. 基于深度神经网络的用户画像研究 [D]. 长沙: 湖南大学, 2018.
Zhou MX. Research on user portrait based on Deep Neural Network [D]. Changsha: Hunan University, 2018.
- [23] 陈一凡, 朱民耀, 朱晓强, 等. 基于时序特征的图卷积协同过滤推荐算法 [J]. 电子测量技术, 2022, 45(6): 79-85.
Chen YF, Zhu MY, Zhu XQ, et al. Graph convolution collaborative filtering recommendation algorithm based on time series features [J]. Electronic Measurement Technology, 2022, 45(6): 79-85.