

DOI: 10.3969/j.issn.2096-8248.2025.01.0008

引用格式: 张龙昌, 白静. 融合情景和社会网络的数字资源服务推荐算法[J]. 江苏海洋大学学报(自然科学版), 2025, 34(1): 62-70.

## 融合情景和社会网络的数字资源服务推荐算法

张龙昌<sup>1</sup>, 白静<sup>2</sup>

(1. 宿迁学院 信息工程学院, 江苏 宿迁 223800; 2. 东北财经大学 管理科学与工程学院, 辽宁 大连 116025)

**摘要:** 异构环境下, 为进一步提高基于用户的数字资源服务协同过滤推荐精准度和缓解冷启动、数据稀疏对推荐质量的影响, 提出融合读者情景和社会网络的数字资源服务协同过滤推荐方法。该方案设计融合评分相似度计算、情景相似度计算、邻居信任度计算的读者近邻相似度综合计算方法, 提出基于读者近邻相似度综合计算的评分预测和服务推荐方法, 基于迪杰斯特拉思想设计了信任度计算算法, 求解读者邻居信任度。通过在真实数据集上验证, 所提出的方法在一定程度上改善了基于用户协同过滤推荐的精准度, 并且一定程度上缓解了冷启动、数据稀疏对推荐质量的影响。

**关键词:** 情景; 社会网络; 数字资源服务; 协同过滤; 相似度计算

中图分类号: TP391 文献标志码: A 文章编号: 2096-8248 (2025) 01-0062-09

## Recommendation algorithms for digital resource services integrating context and social network

ZHANG Longchang<sup>1</sup>, BAI Jing<sup>2</sup>

(1. School of Information Engineering, Suqian University, Suqian 223800, China;

2. School of Management Science and Engineering, Dongbei University of Finance and Economics, Dalian 116025, China)

**Abstract:** In the heterogeneous environment, to enhance the accuracy of user-based collaborative filtering recommendation for digital resource services and mitigate the impact of cold start and data sparsity on recommendation quality, this study proposes a digital resource service collaborative filtering recommendation method that integrates reader context and social network. The proposed solution devises a comprehensive calculation method for determining reader neighbor similarity by integrating score similarity calculation, context similarity calculation, and neighbor trust degree calculation. Furthermore, it presents a rating prediction and service recommendation method based on the comprehensive calculation of reader neighbor similarity, drawing inspiration from Dijkstra, an algorithm is devised to calculate trust degrees in order to ascertain the reliability of reader neighbors. Experimental results obtained from real datasets validate that the proposed approach moderately enhances the accuracy of user-based collaborative filtering recommendations while also partially alleviating the impact of cold start and data sparsity on recommendation quality.

**Key words:** context; social network; digital resource service; collaborative filtering; similarity calculation

收稿日期: 2024-09-25; 修订日期: 2024-10-13

基金项目: 国家社会科学基金资助项目(19BTQ028)

作者简介: 张龙昌(1978—), 男, 教授, 博士, 研究方向为服务计算、个性化推荐, (E-mail) zlc\_041018@163.com。

通信作者: 白静(1987—), 女, 博士研究生, 研究方向为云计算、个性化推荐, (E-mail) bj490367659@126.com。

## 0 引言

在大数据背景下有效地向读者提供在线数字资源服务,需要有足够计算和存储能力的云计算平台存储大数据、有足够带宽的网络传输数据、有足够能力的终端支撑数字资源服务。现有信息技术已经能够有效支撑读者随时随地订阅数字资源服务。同时图书馆的数字化也随着信息技术的进步不断发展,数字资源的个性化服务不再局限于知识内容的提供,还包括一些增值服务,如检索推荐、在线阅读、参考咨询、教育培训等。本文的数字资源服务泛指数字图书馆通过信息技术向外提供的所有服务。数字图书馆既可以携其拥有的数字图书资源加入数字资源云平台向读者提供信息服务,也可以通过网络直接向读者提供数字图书服务。海量数字资源服务的选择已经成为读者面临的难题。在传统 Web 环境下,个性化推荐技术仍然存在不能准确、全面地向读者推荐数字资源问题。长期存在动态变化、异构的读者终端和网络环境(称其为异构环境,包含终端硬件环境、软件环境、网络环境的异构),以及读者在订阅数字资源服务过程中的不断移动,会导致其所处终端软硬件和网络环境不断变化,进一步加大了实现精准数字资源个性化服务的困难<sup>[1]</sup>。

传统数字资源服务个性化推荐,人们往往只关注“读者—数字资源服务”间的关联关系,而较少考虑它们所处的上下文环境(如时空信息、环境和实体、读者状态情况、网络条件、软硬件条件等)。而异构网络环境下,许多应用场景中仅考虑“读者—数字资源服务”二元关系并不能生成有效推荐。例如,某些应用场景中,有些读者更喜欢在有선宽带和大屏幕计算机环境下被推荐在线文本阅读资源,而有些读者则偏好在移动网络和智能手机环境下被推荐在线视频资源。还有些读者倾向于在“晚上”而非“早上”获得合适的英语故事推荐,或期望在“旅行”时获取相关的人文风景介绍等。此外,精力充沛时,部分读者更渴望获得专业前沿文献的推荐。因此,考虑不同情景对个性化服务需求的差异成为影响数字资源个性化服务精准度的关键因素之一。

现实中读者对数字资源的了解,往往是通过与同学、师长、朋友等交流获得,也可能通过广告、讲座、培训等渠道获得。如读者 A 从读者 B 和 C 处了解到数字资源 M 是关于数字图书馆的一本好电子书,而 B 和 C 的信息又是从其他读者处获得或者亲

身体验获得,这就形成了一个对数字资源 M 的交流社会网络,而这个网络在很大程度上影响了读者 A 对数字资源 M 的决定。在信息世界中数字资源信息的获得更加多样化,除了广告、讲座、培训可以通过数字形式展现外,同学、师长、朋友的交流可通过即时通讯(如微信、QQ)、公众号、微博等更多样化的方式进行。因此,读者的社会网络信息成为影响数字资源精准个性化服务的又一关键因素。

本文贡献如下:① 比较共同评分读者集为并集和交集时,采用余弦相似和皮尔逊系数计算近邻对推荐结果的影响;② 提出新的读者情景相似计算,并与基于余弦相似和皮尔逊系数的情景相似计算进行比较;③ 提出读者邻居信任度计算方法及算法,并比较引入读者邻居信任度与未引入读者邻居信任度对推荐结果的影响;④ 提出了融合情景和社会网络的目标读者评分预测方法和协同过滤推荐算法。

## 1 相关工作

协同过滤是个性化推荐系统的主流技术,大多数协同过滤的改进算法是基于项目和基于用户。基于用户的协同过滤通过寻找兴趣相似邻居,并基于相似邻居的兴趣爱好向目标用户推荐其可能感兴趣的项目。该算法能够有效挖掘评分数据中用户的兴趣,算法简单易实现,个性化效果好,具有可解释性等优点,故得到了广泛应用。在数字图书个性化信息服务推荐领域,传统基于用户的协同过滤推荐考虑“读者—信息服务”二维空间,能较好地适应单一的 Web 环境;然而读者的个性化需求具有复杂性、多维性、动态性以及易变性等特点,研究者们开始将视角转向有时间、地点等情景信息(与上下文含义相同)参与的协同过滤推荐<sup>[2-4]</sup>。目前基于“读者—信息服务—情景信息”的三维推荐处于探索期,一些研究成果将情景信息应用到数字图书信息服务的协同过滤推荐中,如文献[5-10]。文献[5]用皮尔逊系数计算特定情景下目标读者和读者的评分相似度计算近邻,同样在特定情景下取前  $N$  个最相似读者预测目标读者评分,最后将预测评分乘以基于条件熵的情景权重。文献[6]同样采用皮尔逊系数计算特定情景下目标读者和读者的评分相似度,然后进行目标读者评分预测。因读者借阅行为轨迹易发生变迁,文献[7]提出一个非线性的时间衰减模型对兴趣评分模型生成的“用户—图书—评分”矩阵进行优化,将读者近期借阅的图书赋予更

高的评分值,同时在评分值上对读者较久远借阅的图书进行惩罚。文献[8]基于加权贝叶斯算法计算用户的行为相似度,然后进行协同推荐;再计算情景属性对所推荐资源的影响的权值,对预测评分加权。文献[9]基于皮尔逊系数计算特定情景下读者的近邻,并且强调数据稀疏下该方法无法正常获得读者近邻,因此又提出采用余弦相似计算读者近邻。

冷启动、评分数据稀疏长期以来制约基于协同过滤的信息服务推荐的精准度,尤其在大数据环境下显得更突出。针对冷启动问题,可以采用读者和服务聚类<sup>[10]</sup>、读者行为生成评分<sup>[11]</sup>、不同领域间的跨域推荐<sup>[12]</sup>、社会网络<sup>[9]</sup>等方法。这里对社会网络法进行简述。文献[9]基于通话、短信构建移动用户间的信任网络关系,两用户间有直接通讯的即构成直接信任关系,两用户间没有直接通讯关系但是可以通过多用户间接连通的构成间接信任关系,再基于通讯量计算直接信任度和间接信任度,进而基于信任进行协同推荐。针对社会化标签—评价矩阵方法可能面临的“冷开始”“新用户”“新资源”等问题,文献[13]提出综合运用基于社会化标签网络的内容推荐和基于知识互动型社会网络的协作过滤推荐来协同解决,由此构建了基于组合策略的个性化知识推荐架构。文献[14]利用社会化媒体平台中社

会网络的特征挖掘用户的兴趣偏好,社会网络方法能准确地得到用户与资源之间以及用户与用户之间的关联关系,得到每个用户对于发布资源的兴趣程度,提高推荐的准确性。文献[15]在协同过滤中融入用户综合信任度,由用户间通信信任、传递信任、相似度信任及信任等级加权后获得。

上述研究存在下述问题:①情景相似计算中,余弦相似计算两个向量的夹角,皮尔逊系数是度量两个向量的相关性,两者都受数值大小影响;而情景属性值体现读者某个状态,与数值大小无关,如职业有学术/教育家、艺术家、文员/行政、大专/研究生、客户服务;需要设计新的情景相似计算方法。②读者邻居的信任度计算中,不同应用场景下社会网络的构建方式和信任度计算方法有较大不同,需要在异构网络环境下构建新的读者社会网络和信任度计算方法。③单独考虑评分相似度、情景相似度(如前过滤或者后过滤<sup>[16]</sup>)、读者邻居信任度都会产生推荐信息的丢失,需要设计融合3种相似度的综合邻居相似度计算方法和评分预测方法。为此在基于用户的协同过滤推荐基础上,本文设计了融合情景和社会网络的数字资源服务协同过滤推荐模型(见图1)。其中相似度计算和目标读者评分预测是其主要实现过程。

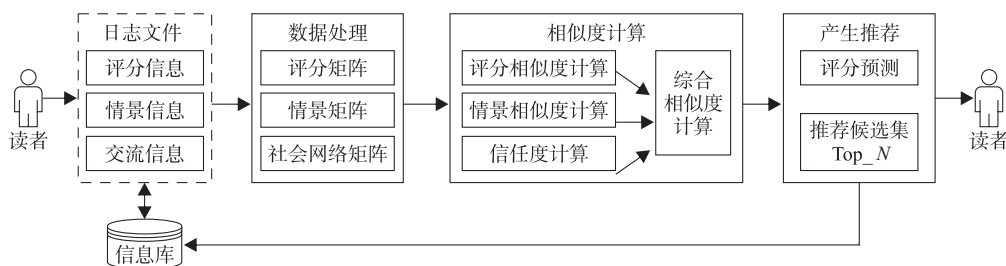


图1 融合情景和社会网络的数字资源服务推荐模型

Fig.1 Digital resource service recommendation model integrating context and social network

## 2 读者评分相似度

基于读者评分的邻居相似度计算主要采用Pearson系数(式(1))、余弦相似度(式(2))和修正余弦相似度。修正余弦是余弦的改进,因此这里只考虑前两种相似度计算。

$$\text{sim}_r(a, b) = \frac{\sum_{i \in I_{ab}} (r_{a,i} - \bar{r}_a)(r_{b,i} - \bar{r}_b)}{\sqrt{\sum_{i \in I_{ab}} (r_{a,i} - \bar{r}_a)^2} \sqrt{\sum_{i \in I_{ab}} (r_{b,i} - \bar{r}_b)^2}}, \quad (1)$$

$$\text{sim}_r(a, b) = \cos \theta =$$

$$\frac{a \cdot b}{\|a\| \|b\|} = \frac{\sum_{i \in I_{ab}} r_{a,i} \times r_{b,i}}{\sqrt{\sum_{i \in I_{ab}} (r_{a,i})^2} \sqrt{\sum_{i \in I_{ab}} (r_{b,i})^2}}. \quad (2)$$

1) 式(1)的共同评分集取交集(记作UCFR\_DRI)时,  $I_{ab} = I_a \cap I_b$  为读者  $a$  和  $b$  共同评分资源集合,  $I_a$  和  $I_b$  分别为读者  $a$  和  $b$  的评分集合;  $r_{a,i}$  和  $r_{b,i}$  分别表示读者  $a$  和  $b$  对资源  $i$  的评分,  $\bar{r}_a$  和  $\bar{r}_b$  分别表示读者  $a$  和  $b$  对共同评分资源的平均评分。由式(1)得,  $\text{sim}_r(a, b)$  取值在  $-1$  和  $1$  之间, 其取

值越大, 两用户的评分越相似。当  $|I_{ab}| < 2$  或者式 (1) 的分母为 0 时, 设  $\text{sim}_r(a, b) = 0$ 。

2) 式 (1) 的共同评分集取并集, 其中  $I_{ab} = I_a \cup I_b$  为读者  $a$  和  $b$  评分资源集合的并集,  $I_a$  和  $I_b$  分别为读者  $a$  和  $b$  的评分集合;  $r_{a,i}$  和  $r_{b,i}$  分别表示读者  $a$  和  $b$  对资源  $i$  的评分 (如果读者  $a$  对资源  $i$  没有评分, 则  $r_{a,i} = 0$ ;  $r_{b,i}$  同理),  $\bar{r}_a$  和  $\bar{r}_b$  分别表示读者  $a$  和  $b$  的资源平均评分。 $\bar{r}_a$  和  $\bar{r}_b$  的计算公式有两种形式 ( $|I_a \cup I_b|$ ,  $|I_a|$ ,  $|I_b|$  为集合中元素个数), CF\_CSN (1) 和 CF\_CSN (2) 分别代表式 (3) 和式 (4) 读者评分相似度计算。

同理, 对于余弦相似计算也存在共同评分集取

交集还是取并集的情况; 式 (2) 是计算两个向量的余弦夹角, 与共同评分取交集还是并集无关, 因此余弦相似计算记为 UCFR\_DRC。

$$\begin{cases} \bar{r}_a = (\sum_{i=1}^{|I_a \cup I_b|} r_{a,i}) / |I_a \cup I_b|, \\ \bar{r}_b = (\sum_{i=1}^{|I_a \cup I_b|} r_{b,i}) / |I_a \cup I_b|. \end{cases} \quad (3)$$

$$\begin{cases} \bar{r}_a = (\sum_{i=1}^{|I_a|} r_{a,i}) / |I_a|, \\ \bar{r}_b = (\sum_{i=1}^{|I_b|} r_{b,i}) / |I_b|. \end{cases} \quad (4)$$

在表 1 中, UserID 为 “90” 和 “91” 两个用户对 MovieID 为 “1 193”、“661”、“914”、“3 408”、“2 355” 的 5 部电影的评分, 采用 5 种相似度计算获得的两用户的相似度。采用不同的方法获得的用户相似度存在较大差距, 从而影响推荐结果和评分的预测。

表 1 读者评分实例及相似计算

Table 1 Examples of reader ratings and similarity calculation

UserID	1 193	661	914	3 408	2 355	UCFR_DRI	CF_CSN (1)	CF_CSN (2)	UCFR_DRC
90	5	3	3	4	0	1.0	0.549 7	0.272 1	0.999 4
91	0	4	4	5	0				

### 3 读者情景相似度

随着数字资源个性化服务应用和研究的深入, 其推荐精准度成为提升读者满意度的重要指标。异构网络环境下 (包含终端硬件环境、软件环境、网络环境的异构), 情景对读者数字资源的需求影响更加明显。表 2 描述了 3 位读者在不同情景下的数字资源服务需求情况。在前期工作中<sup>[1]</sup>, 本文对异构环境下读者上下文 (即情景) 及建模进行了深入研究, 在此基础上提出读者情景相似度计算方法, 进一步提升数字资源服务的推荐精准度。

**定义 1 (读者集合)**  $U = \{u_i | i \in [1, N]\}$ , 定义为数字资源服务面向的读者的集合, 其中  $N$  代表读者数。

**定义 2 (数字资源服务集合)**  $S = \{s_j | j \in [1, M]\}$ , 定义为数字资源服务集合, 其中  $M$  代表服务数; 任意服务  $s_j = \{s_{jp} | p \in [1, M_s]\}$ , 定义为服务属性 (尤其是服务质量属性即 QoS) 的集合,  $M_s$  为属性数目; 一个服务实例表示为  $s_j = (s_{j1}, s_{j2}, \dots, s_{jM_s})$ , 其中  $s_{jp}$  为服务  $s_j$  的一个具体属性值。若  $s_{jp}$  为响应时间 (QoS 指标之一), 其可能取值为秒数。

**定义 3 (读者情景集合)**  $C = \{c_k | k \in [1, O]\}$ , 定义为读者在使用数字资源期间的所有状态下的情景集合, 其中  $O$  代表情景数; 读者在状态  $k$  下的情景用向量表示为  $c_k = (c_{k1}, c_{k2}, \dots, c_{kO_c})$ , 其中  $c_{kp}$  ( $p \in [1, O_c]$ ) 为情景  $c_k$  的一个具体属性值, 情景实

例须与读者实例关联。若  $c_{kp}$  为终端设备, 其可能取值为智能手机、平板电脑、移动笔记本、台式电脑。如表 2 中的 3 个向量  $c_t, c_1, c_2$  可以描述读者  $u_1$  在 3 种状态下的情景。

情景属性值通常为离散类型的概念, 如地点可以是宿舍、教室、餐厅、图书馆。相似情景计算可采用皮尔逊<sup>[5]</sup> (式 (4))、余弦<sup>[9]</sup> (式 (5))、欧式距离 (式 (6)) 等相似度计算方法。这些方法首先需要情景属性值进行归一化处理, 归一化公式为  $c_{kp} =$

$$\frac{c_{kp}}{\max_j c_{kj}}, (j \in [1, O_c])$$

$$\text{sim}_c(c_t, c_y) = \frac{\sum_{p=1}^{O_c} (c_{tp} - \bar{c}_t)(c_{yp} - \bar{c}_y)}{\sqrt{\sum_{p=1}^{O_c} (c_{tp} - \bar{c}_t)^2} \sqrt{\sum_{p=1}^{O_c} (c_{yp} - \bar{c}_y)^2}}, \quad (4)$$

$$\text{sim}_c(c_t, c_y) = \frac{c_t \cdot c_y}{\|c_t\| \|c_y\|} = \frac{\sum_{p=1}^{O_c} c_{tp} \times c_{yp}}{\sqrt{\sum_{p=1}^{O_c} (c_{tp})^2} \times \sqrt{\sum_{p=1}^{O_c} (c_{yp})^2}}, \quad (5)$$

$$\text{sim}_c(c_t, c_y) = \sqrt{\sum_{p=1}^{O_c} (c_{tp} - c_{yp})^2}. \quad (6)$$

上述 3 种方法的情景相似度计算与数值大小相关, 如地点的取值可以是 “0: 宿舍; 1: 教室; 2: 餐厅; 3: 图书馆”, 在计算时 3 对相似度影响高于 0, 而事实上 0, 1, 2, 3 对相似度的影响应为同等重要。基于上述分析, 提出采用式 (7) 计算情景相似度, 即

$$\text{sim}_c(c_t, c_y) = \frac{\sum_{p=1}^{O_c} (c_{tp} \wedge c_{yp})}{|c_t \cup c_y|} \quad (7)$$

其中  $c_{tp} \wedge c_{yp} = \begin{cases} 0, & c_{tp} \neq c_{yp} \\ 1, & c_{tp} = c_{yp} \end{cases}$ ,  $|c_t \cup c_y|$  为情景  $c_t$

属性与情景  $c_y$  属性并的个数。在表 2 中按照式(7)计算  $c_1, c_2$  与  $c_t$  的相似度都是 0.5; 很显然相对于  $c_1, c_2$  与  $c_t$  的相似更加稳定并且可信。为了消除这个影响, 设置式(7)中的  $|c_t \cup c_y|$  为  $\max_i |c_t \cup c_y| (i \in [1, O])$ 。

表 2 情景  $c_1, c_2$  与  $c_t$  的相似计算  
Table 2 Similarity calculation of context  $c_1, c_2$  and  $c_t$

情景	职业	年龄	时间	日期特性	地点	社区	终端设备	阅读软件	相似度	项目
$c_t$	学生		上午		教室		平板电脑			学术资源
$c_1$	学生		上午						0.5	学术资源
$c_2$	学生	20	上午	周末	教室	学校	平板电脑	爱奇艺	0.5	在线电影

表 2 中, 不同读者在选择数字资源项目时, 情景属性职业、时间均没有变化, 其重要程度显然小。如果按照与其他情景属性具有同样的权重进行计算就会影响相似计算的精准度。下面提出基于信息熵的情景属性权重计算方法。信息熵是度量信息源各可能事件发生的不确定性, 是各事件信息量的期望。对于随件事件  $X = \{x_1, x_2, \dots, x_n\}$ ,  $P(x_i)$  为  $x_i$  发生的概率, 则事件的信息熵  $H(X) = -\sum_{i=1}^n (P(x_i) \times \log_2 P(x_i))$ , 其取值范围为  $[0, \log_2 n]$ 。信息熵越大表明情景属性取值越分散(该属性权重越大), 信息熵越小则情景属性的取值越集中(该属性权重越小)。根据上面的分析, 情景属性权重定义为

$$W(p) = \frac{H(p)}{\sum_{p=1}^{O_c} H(p)} \quad (8)$$

基于上述分析, 情景的相似度计算式(7)更改为

$$\text{sim}_c(c_t, c_y) = \frac{\sum_{p=1}^{O_c} W(p) \times (c_{tp} \wedge c_{yp})}{|c_t \cup c_y|} \quad (9)$$

### 4 读者邻居信任度

读者在主动选择数字资源时, 总是会通过向朋友咨询、浏览介绍(编写介绍者可能是数字资源作者、读者)等方式确定。在网络环境下, 读者的社会网络建立可通过获取读者在微信、QQ、微博等的咨询、个人网页、公众号的浏览记录, 以及在线调查等方式获得读者的数字资源交流社会网络(实例如图 2 所示)。

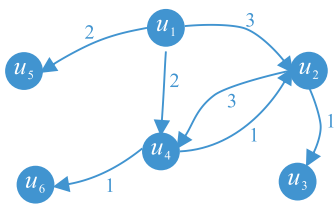


图 2 读者社会网络实例  
Fig.2 Examples of reader social network

在读者的数字资源交流社会网络中, 读者定义社会网络图的顶点, 各顶点之间的交流定义为弧, 箭头表示咨询方向, 各顶点相互沟通就构成读者的社会网络。图 2 中  $u_i$  表示一位读者个体, 顶点  $u_1$  到  $u_2$  有 1 条弧, 其权值为 3, 表示读者  $u_1$  向  $u_2$  咨询了 3 次;  $u_1$  到  $u_2$  有 1 条弧, 而  $u_2$  到  $u_1$  没有弧则其权值设为 0; 顶点  $u_2$  与  $u_4$  之间是两个互反方向弧, 分别表示  $u_2$  向  $u_4$  咨询了 3 次,  $u_4$  向  $u_2$  咨询了 1 次, 以此类推。设  $A$  为读者的社会网络矩阵, 图 2 可用下述矩阵描述。

$$A = \begin{matrix} & \begin{matrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 \end{matrix} \\ \begin{matrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{matrix} & \begin{bmatrix} 0 & 3 & \infty & 2 & 2 & \infty \\ 0 & 0 & 0 & 3 & \infty & \infty \\ \infty & 1 & 0 & \infty & \infty & \infty \\ 0 & 1 & \infty & 0 & \infty & 1 \\ 0 & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & 0 & \infty & 0 \end{bmatrix} \end{matrix} \quad (10)$$

从图 2 中可以得出, 读者之间是直接信任关系和间接信任关系。读者之间的直接咨询关系形成了直接信任关系, 读者之间的知识传播形成了间接信任关系, 进而构成了读者信任网络。在个性化推荐中, 尤其是新读者的冷启动和数据稀疏条件下, 可以将读者社会网络中信任度高的读者作为推荐邻居。

#### 4.1 计算读者直接信任度

在读者的社会网络中, 顶点有入度和出度。出度越大, 说明该读者越主动向网络中其他读者咨询、请教获取知识, 被请教和咨询的读者为其信任者; 入度越大, 说明该读者被其他读者咨询的次数越多, 其影响力越大。因为读者间的信任度会产生变化, 读者信任度的计算需要周期地进行。下面首先定义读者的直接信任度计算方法。

**定义 4(读者直接信任度)** 假设在一个周期内, 读者  $u \in U$  与网络中所有读者的交流中, 总的咨询次数为  $T$ , 在该周期内, 读者  $u$  对读者  $v \in U$  总的咨询

次数为  $T_{uv}$ , 则读者  $u$  对读者  $v$  的信任度  $P_{uv}$  定义为

$$P_{uv} = \frac{T_{uv}}{T} \quad (10)$$

同理, 可以计算读者  $v$  对读者  $u$  的信任度  $P_{vu}$ ,  $P_{uv}$  与  $P_{vu}$  不一定相等, 表明  $u$  信任  $v$ , 而  $v$  不一定信任  $u$ 。

#### 4.2 计算读者间接信任度

通过计算读者间接信任度来确定读者间接邻居, 通过间接邻居的填充, 在一定程度上解决直接邻居的稀疏问题。下面定义读者间接信任度计算方法。

**定义 5 (读者间接信任度)** 在计算读者间接信任度时, 如果读者  $u$  和  $v$  之间没有弧直接相连, 但存在一条有向的路径  $(u, u_1, \dots, u_k, v)$ , 则读者  $u$  对读者  $v$  的间接信任度  $P_{uv}$  定义为式 (11), 基于社会网络的小世界特性,  $k$  的取值最大为 6。

$$P_{uv} = P_{uu_1} \times P_{u_1u_2} \times \dots \times P_{u_kv} \quad (11)$$

读者  $u$  对  $v$  读者的间接信任度可能存在多个值, 即从读者  $u$  到读者  $v$  之间存在多个有向路径, 多个值中取最大值作为读者  $u$  对读者  $v$  的间接信任度。

#### 4.3 信任度计算算法

从节点  $u$  出发到任意节点  $v \in U$  可能会有多条路径, 每条路径可获得一个信任度, 取信任度最大值作为  $u$  对  $v$  的信任度, 以矩阵  $A$  为数据结构, 求解  $u$  到任意  $v \in U$  的信任度的算法可通过改进迪杰斯特拉算法 (Dijkstra) 实现。

##### 算法 1 DIJKSTRA\_MODI

输入: 读者社会网络矩阵  $A$ , 读者  $u$ ;

输出: 读者  $u$  对  $A$  中各节点的信任度  $\text{distan}$ ;

1. 将  $\text{vis}[]$  初始化为 0, 将  $\text{distan}[]$  初始化为 INF; //  $\text{vis}[i]$  为 0 表示节点  $i$  未被访问,  $\text{distan}[i]$  是  $i$  点到初始点的距离

2.  $\text{distan}[u] = 0$ ; // 将起始点到自身的距离设为 0

3. 转换矩阵  $A$  为  $A_{ij} = \frac{A_{ij}}{\sum_{j=1}^n A_{ij}}$ ; // 按照公式 (10) 将矩阵转  $A$  换为直接信任度矩阵

4. FOR  $i = 0$  TO  $n$

5.  $w = -1$ ,  $\text{MAX} = 0$ ; // 初始化  $w$  和  $\text{MAX}$ ,  $w$  用于记录查找到  $\text{MAX}$  的位置 (节点序号)

6. FOR  $j = 0$  TO  $n$

7. IF (节点  $j$  未被访问 &&  $\text{distan}[j]$  储存的值大于  $\text{MAX}$ )

8. 记录该位置并更新  $\text{MAX}$

9. IF (未找到新的  $\text{MAX}$  值) RETURN; // 可说明剩余点不连通, 退出算法

10.  $\text{vis}[w] = \text{true}$ ; // 记  $w$  被访问

11. FOR  $v = 0$  TO  $n$  // 开始对路径进行优化

12. IF (  $v$  点未被访问 && 图上  $w$  和  $v$  连通 &&  $\text{distan}[w] * d[w][v] > \text{distan}[v]$  ) { \* 最后一个判断条件表明从新添加的节点到  $v$  比  $\text{distan}[v]$  记录的信任度更高, 其中  $\text{distan}[w] * d[w][v]$  为公式 (11) 的程序实现 \* }

13. 更新  $\text{distan}[v]$  为  $\text{distan}[w] * d[w][v]$ ;

14. END FOR

15. END FOR

## 5 数字资源推荐

相同评分下, 两个读者所处的情景越相似, 互相推荐的结果越有参考价值。读者社会网络中的信任关系体现读者对哪些朋友给予的建议更加信任, 受其影响也会更大, 并且采纳的可能性也会更高, 但其并不是针对具体数字资源服务的偏好。为解决数据稀疏、冷启动读者和推荐准确度问题, 将读者的情景相似和信任网络引入到数字资源服务推荐生成过程中, 读者相似度计算见式 (12)。在式 (13) 的基础上引入读者情景相似度和邻居信任度, 预测目标读者对某数字资源服务的评分, 见式 (14)。

综合评分相似度、情景相似度和邻居信任度后的读者  $u$  和  $p$  的相似度计算见式 (12), 即

$$\text{sim}(u, p) = \text{sim}_r(u, p) \times \text{sim}_c(c_u, c_p) \times p_{u,p} \quad (12)$$

基本的目标读者评分预测见式 (13), 其中  $R_{u,i}$  为目标读者对资源  $i$  的预测评分,  $U_s$  为  $u$  的近邻集合。

$$R_{u,i} = \bar{r}_u + \frac{\sum_{p \in U_s} \text{sim}_r(u, p) \times (r_{p,i} - \bar{r}_p)}{\sum_{p \in U_s} \text{sim}_r(u, p)} \quad (13)$$

$$R_{u,s} = \bar{r}_u + \frac{\sum_{p \in U_s} \text{sim}_r(u, p) \times \text{sim}_c(c_u, c_p) \times p_{u,p} \times (r_{p,s} - \bar{r}_p)}{\sum_{p \in U_s} \text{sim}_r(u, p) \times \text{sim}_c(c_u, c_p) \times p_{u,p}} \quad (14)$$

其中  $R_{u,s}$  为目标读者  $u$  对数字资源服务  $s$  的预测评分,  $U_s$  为  $u$  的近邻集合。有些数据集中不存在读者情景和信任网络数据, 此时式 (14) 中的  $\text{sim}_c(c_u, c_p)$  和  $p_{u,p}$  设置为 1。现实中, 一位读者来到一个陌生环境获取数字资源, 其一般会事先采取咨询、浏览简介、听取建议等方式进行了解, 因此读者的社会关系网络在推荐数字资源前就会存在; 读者在登录推荐系统时, 需要采集读者的基本信息, 因此读者情景信息一般情况下也是存在的。基于评分相似度、情景相似度、读者邻居信任度和目标读者评分

预测设计融合情景和社会网络的协同过滤推荐算法(collaborative filtering recommendations integrating context and social networks, CF\_CSN)。

#### 算法 2 CF\_CSN

输入:“读者—情景—数字资源服务—评分”评分数据集  $R$ ; 目标读者  $u$  及其当前情景  $c_u$ ; 读者的咨询社会网络矩阵  $A$ ;

输出: 读者  $u$  在情景  $c_u$  下, 对集合  $S$  中最感兴趣的 Top- $N$  个数字资源服务

1. 从  $R$  中提取出读者集  $U$ 、数字资源服务集合  $S$  和评分数据集  $\bar{R}$  (行数据格式为(读者 ID, 评分)); 如果有相同读者对一个资源多次评分, 取其平均);

2.  $\text{trust} = \text{Call DIJKSTRA\_MODI}(A, u)$  // 调用函数计算  $u$  的信任邻居

3. FOR EACH  $v$  IN  $U$  // 评分相似度计算和情景相似度计算

// 评分相似度计算

4. 在评分集  $\bar{R}$  上, 利用式(3)或(4)计算评分相似度  $\text{sim}(u, v)$

5. 在  $u$  的相似评分邻居矩阵  $N_r$  中增加一行, 行数据格式为(评分相似度, 读者 ID);

// 情景相似度计算

6. 在评分集  $R$  上, 利用式(9)计算情景相似度  $\text{sim}_c(c_u, c_v)$ , 取情景相似度最大值;

7. 在  $u$  的相似情景邻居矩阵  $N_c$  中增加一行, 行数据格式为(情景相似度, 读者 ID);

// 计算目标读者  $u$  的相似读者矩阵  $N$ , 行数据格式为(相似度, 读者 ID)

8.  $u$  和  $v$  的相似度 = 评分相似度  $\text{sim}(u, v)$  \* 情景相似度  $\text{sim}_c(c_u, c_v)$  \* 信任度  $\text{trust}(v)$ ;

9. 在  $u$  的邻居矩阵  $N$  中增加一行, 行数据格式为(相似度, 读者 ID);

10. END FOR

11. 按照相似度从大到小对  $N$  进行倒序排列, 从低地址到高地址取  $k$  个邻居作为推荐近邻  $U_s$ ;

12. FOR EACH  $s$  IN  $S$  // 计算数字资源的预测评分

13. 在  $U_s$  集合上, 利用公式(8)计算  $R_{u, s}$ ;

14. IF ( $R_{u, s}$  大于 Top- $N$  服务集中的任何一个服务评分)

15. 将 Top- $N$  服务集中评分最小的服务替换为  $s$ ;

16. END FOR

## 6 实验结果及分析

### 6.1 数据集

MovieLens 数据集<sup>[17]</sup> 包含大约 6 040 个用户、3 900 部在线电影、18 个类别和 1 000 209 个

评分样本。评分包含用户 ID、电影 ID 和分数, 用户情景包含用户 ID (UserID)、性别 (Gender)、年龄 (Age)、职业 (Occupation)、邮政编码 (Zip-code), 这些情景的取值用数字表示。选取 UserID=“100”的用户进行比较分析, 邻居数选取 20 个 (20 个邻居到 30 个邻居之间预测精准度相对最好), “100”号用户对 76 部在线电影打了分, 选前 40 个评分作为训练集, 后 36 个和评分作为预测集。通过浏览 “100” 号用户的评论、建议、其他用户的追问情况, 为 “100” 号用户建立 20 个邻居信任网络。

### 6.2 评价指标

为了验证推荐算法的性能, 本文采用平均绝对误差 (mean absolute error, MAE) 和  $P@N$  作为评价算法好坏的度量标准。MAE 根据预测评分和实际评分的偏差来衡量预测算法的准确度, 偏差越小, 评分预测的准确度越高, 推荐质量就越高。设预测的读者评分集合为  $\{p_1, p_2, \dots, p_N\}$ , 实际读者评分集合为  $\{q_1, q_2, \dots, q_N\}$ ,  $N$  为预测的数字资源服务个数, 则

$$\text{MAE} = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (15)$$

$P@N$  根据预测的 Top- $N$  个数字资源服务列表中占读者实际需求的 Top- $N$  个数字资源服务的比值来衡量预测算法的优劣, 比值越大, 预测的正确度越高, 推荐质量就越高。

$$P@N = \frac{\# \text{relevant services in Top-}N \text{ digital resource services}}{N} \quad (16)$$

### 6.3 评分相似计算比较

通过评分计算读者 “100” 的相似邻居, 文中总结 2 类 5 种情况。基于皮尔逊系数的有 UCFR\_DRI, CF\_CSN (1) 和 CF\_CSN (2), 基于余弦相似的有 UCFR\_DRC。这里不考虑读者情景和社会网络对评分预测的影响 (即式 (14) 中的  $\text{sim}_c(c_u, c_p)$  和  $p_{u, p}$  取 1), 从而退化为式 (13)), 目标读者评分预测采用基本计算方法, 即式 (13), 实验结果见图 3。

从实验结果可以得出: ① 两种算法的推荐指标 MAE 和都偏低, 原因是评分数据过于稀疏; ② 如果关注推荐结果而不关注评分的预测, 那么 CF\_CSN (1) 有更好的推荐效果; ③ 如果关注评分预测而不关注推荐结果, 那么 CF\_CSN (2) 有较好的推荐效果; ④ 整体来看 UCFR\_DRC 推荐效果不佳。

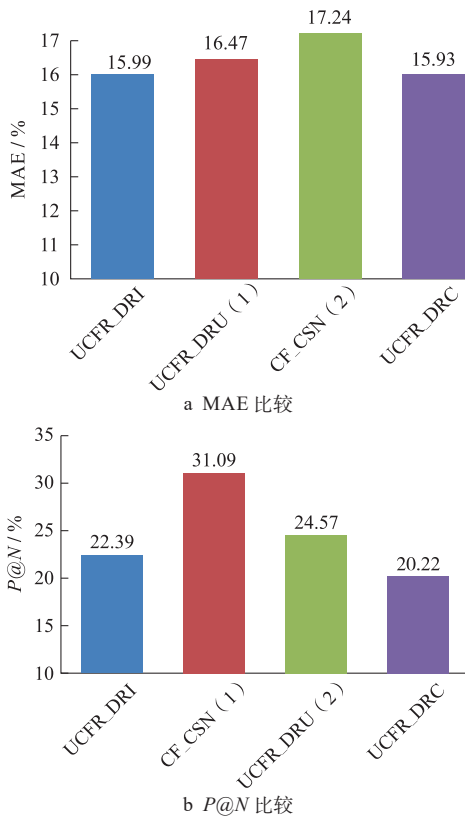


图 3 评分相似计算比较  
Fig.3 Comparison of score similarity calculation

### 6.4 情景相似计算比较

文献 [5] 和 [9] 将情景与评分直接关联, 即读者对电影的每一个评分都有一个具体情景。但是文献 [9] 也强调在数据稀疏情况下, 上述方法是不能计算出情景相似度的; 据此, 文献 [9] 提出在数据稀疏情况下采用余弦相似计算读者情景的相似度, 即式 (5) 的计算方法, 记为 UCFR\_CO (本实验的评分数据为稀疏数据集)。本文提出的方法, 即式 (9) 记为 CF\_CSN; 皮尔逊方法 (文献 [18] 采用的方法), 即式 (4) 记为 UCFR\_PE; 欧式距离法, 即式 (6) 记为 UCFR\_EU。目标读者评分预测采用式 (14) (其中不考虑读者社会网络对评分的影响, 即式 (14) 中  $p_{u,p}$  取 1), 实验结果见图 4。

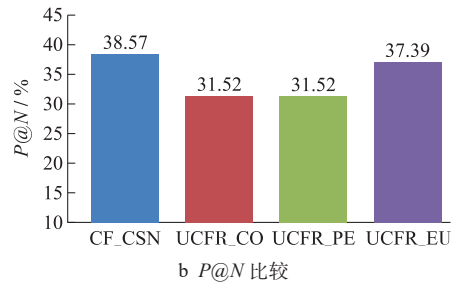
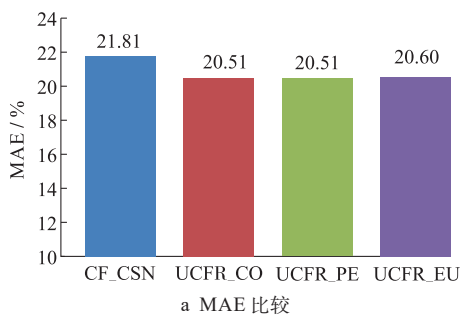


图 4 情景相似计算比较  
Fig.4 Comparison of context similarity calculation

从图 4 中可以看出, 本文提出的方法 CF\_CSN 相对于其他方法, 在 MAE 和  $P@N$  两个指标上表现都比较优越, 其次是基于欧式距离的 UCFR\_EU 方法。

### 6.5 读者邻居信任度比较

文献 [9] 采用通话次数、通话时间和短信数量计算邻居信任度, 不适用数字资源推荐中的邻居信任度计算。因此, 本文实验将读者邻居信任度计算引入到经典的基于余弦相似和基于皮尔逊系数 (即本文的方法, 见式 (4) 和式 (5)) 的协同过滤推荐中 (本实验中不考虑读者情景的影响, 即式 (14) 中的  $\sin_c(c_u, c_p)$  取 1), 然后对引入和未引入后的推荐结果进行对比。

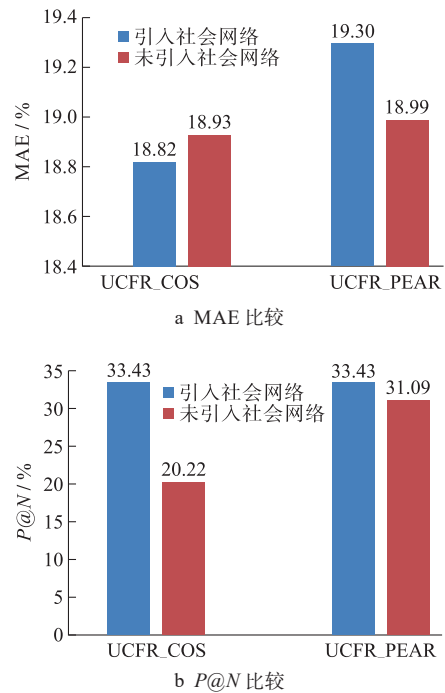


图 5 读者邻居信任度计算比较  
Fig.5 Comparison of reader neighbor trust calculation

从图 5 中可以看出, 本文提出的方法 CF\_CSN

在引入社会网络后,在 MAE 和  $P@N$  两个指标上表现都比较好,同时基于皮尔逊系数的推荐算法整体好于基于余弦相似的推荐算法。

## 7 结语

精准度、冷启动、数据稀疏等问题长期困扰着协同过滤推荐技术在数字资源服务推荐中的应用,在大数据和异构环境下显得更加突出。为进一步改善数字资源服务推荐质量,本文在基于用户的协同过滤推荐框架下,设计了一种综合考虑评分相似度、读者情景相似度和读者信任度的读者相似度计算方法去获取读者最近邻集合,融合 3 种相似度预测目标读者的评分,在一定程度上提高了推荐精准度和缓解了冷启动、数据稀疏问题。

仍存在下述问题需要进一步研究:① 挖掘日志中的隐式评分、融合隐式和显式评分,获取更充实更完整的评分数据集;② 引入外部辅助数据集,充分考虑读者情景,将辅助数据集矩阵和目标数据集矩阵进行联合分解,填补目标数据集中缺失数据,能够适应异构环境;③ 服务推荐集中的服务对读者环境的要求可能是必选项(如在线阅读服务,对于 PDF 类型文件需要终端安装 PDF 阅读器,对于 DOC 文件需要终端安装 Word),建立一种筛选机制过滤掉不符合读者环境的服务;④ 数字资源服务可能同时由多个数字服务提供商提供,严重重复,研究读者情景感知的服务质量最优的数字资源服务选择方法。

### 参考文献:

- [1] 张龙昌.异构环境下数字资源个性化服务读者上下文模型研究[J].新世纪图书馆,2022(7):51-58.
- [2] 周朴雄,陶梦莹.移动网络环境下情景敏感的个性化信息推荐系统研究[J].图书情报工作,2012,56(19):80-84.
- [3] 金红,陈礼珂,游兰,等.基于地理时空关联和社会影响的兴趣点推荐[JOL].计算机科学,1-14[2024-09-17].<http://kns.cnki.net/kcms/detail/50.1075.TP.20240824.1301.011.html>.
- [4] UNGER M, TUZHILIN A. Hierarchical latent context representation for context-aware recommendations[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(7): 3322-3334.
- [5] 曾子明,陈贝贝.移动环境下基于情境感知的个性化阅读推荐研究[J].情报理论与实践,2015,38(12):31-36.
- [6] 杨君,吴菊华,艾丹祥.一种基于情景相似度的多维信息推荐新方法研究[J].情报学报,2013,32(3):262-269.
- [7] 梁思怡,彭星亮,秦斌,等.时间上下文优化的协同过滤图书推荐[J].图书馆论坛,2021,41(3):113-121.
- [8] 程秀峰,范晓莹,杨金庆.一种融合了基于朴素贝叶斯算法与情境感知的协同推荐系统——以大学图书馆实体图书推荐为例[J].现代情报,2019,39(2):57-65.
- [9] 俞春花,刘学军,李斌,等.基于上下文相似度和社交网络的移动服务推荐方法[J].电子学报,2017,45(6):1530-1536.
- [10] 凌霄娥,周兵,李克潮.面向新读者和新图书的数字图书馆个性化推荐冷启动问题研究[J].情报理论与实践,2014,37(8):100-104.
- [11] 宋楚平.一种改进的协同过滤方法在高校图书馆图书推荐中的应用[J].图书情报工作,2016,60(24):86-91.
- [12] 李宇航,夏绍模,程华亮.基于跨域协同的移动图书馆个性化推荐模型研究[J].情报科学,2017(3):82-86.
- [13] 易明,邓卫华,徐佳.社会化标签系统中基于组合策略的个性化知识推荐研究[J].情报科学,2011,29(7):1093-1097.
- [14] 房小可.一种融合情境因素的社会化信息推荐新方法[J].图书情报工作,2015,59(22):105-111.
- [15] 郑孝遥,鲍煜,孙忠宝,等.一种基于信任的协同过滤推荐模型[J].计算机工程与应用,2016,52(5):50-54.
- [16] 王立才,孟祥武,张玉洁.上下文感知推荐系统[J].软件学报,2012,23(1):1-20.
- [17] MAXWELL HARPER F, KONSTAN JOSEPH A. The movielens datasets: history and context[J]. ACM Transactions on Interactive Intelligent Systems (TiiS), 2016, 5(4): 1-19.
- [18] ANWAR T, UMA V, SRIVASTAVA G. CDRc-CAS: cross-domain recommendation using context-aware sequences[J]. IEEE Transactions on Computational Social Systems, 2024, 11(4): 4934-4943.