

基于人性化用户特征的在线酒店推荐技术研究

王婧虹 李银胜

(复旦大学软件学院 上海 201203)

摘要 近年来,人们对于酒店的需求与日俱增。但面对形形色色的酒店服务,如何选择的问题日益突出,精准合理的个性化酒店推荐拥有极大市场。已有的传统推荐技术及其推荐系统,没有从内生性参数分析和解释用户行为,在用户兴趣点发掘的精确性以及数据稀疏情况下的推荐存在局限。对此提出一种面向酒店的人性化用户建模与推荐方法。根据所在实验室提出的数字灵魂模型,将其映射到酒店领域,得到人性化酒店用户模型。计算用户偏好与酒店特征间的相似度,并结合基于协同过滤的推荐技术,得出推荐候选集。以“酒店管理营运博弈沙盘”为实验平台,该沙盘是由所在实验室研发并已被数十家酒店管理院校使用的电子学习软件。实验结果表明,相比于单纯的特征参数匹配和普通的协同过滤,该方法拥有更高的准确率、召回率和运行效率,并在一定程度上解决冷启动和数据稀疏的问题。

关键词 酒店推荐 数字灵魂模型 人性化特征 协同过滤

中图分类号 TP3

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.02.020

A HUMANIZED USER FEATURES APPROACH FOR HOTEL RECOMMENDATION

Wang Jinghong Li Yinsheng

(School of Software, Fudan University, Shanghai 201203, China)

Abstract In recent years, people's demand for hotels is increasing rapidly. However, in the face of various hotel services, the problem of how to choose has become increasingly prominent, and accurate and reasonable personalized hotel recommendation has a huge market. The existing traditional recommendation technology and its recommendation system do not analyze and explain user behavior from endogenous parameters, and there are limitations in the accuracy of user interest point discovery and recommendation in the case of sparse data. This paper proposed a humanized user modeling and recommendation method for hotels. According to the Cyber-Anima model proposed by the laboratory, it was mapped to the hotel field, and a humanized hotel user model was obtained. We calculated the similarity between user preferences and hotel features, and combined the recommendation technology based on collaborative filtering to get the recommendation candidate set. Based on the e-learning software developed by our laboratory and used by tens of hotel management schools, we did the experiments. The experimental results show that this method has higher accuracy, recall and operation efficiency than the simple feature parameter matching and common collaborative filtering, and solves the problem of cold start and data sparseness to a certain extent.

Keywords Hotel recommendation Cyber-Anima model Humanized features Collaborative filtering

0 引言

我国的酒店市场空间巨大,尤其是随着经济水平和生活质量的不断提升,人们出行频率大幅提高,对酒店服务的需求扩大,酒店种类变得丰富且数量激增。

因此,为用户提供精准有效的酒店推荐服务尤为重要。

目前国内外均提出并实施了很多推荐系统,以方便用户挑选酒店。但已有的推荐系统大多存在数据稀疏、冷启动以及对用户行为缺乏理解的缺点,没有挖掘用户人性化数据,确定其兴趣所在,推荐其真正感兴趣的酒店资源。以下结合用户场景介绍目前酒店推荐存

在的问题。

总所周知,李彦宏、马云、马化腾分别为 BAT 创始人,均为中国著名男性企业家,如表 1 所示,因此对酒店的硬性需求基本一致。

表 1 共同特征

姓名	性别	职业	领域	出行类别
李彦宏	男	企业家	互联网	公务出差
马云	男	企业家	互联网	公务出差
马化腾	男	企业家	互联网	公务出差

若三位到达上海,需入住酒店时,按传统推荐方法,均会推荐上海万豪虹桥大酒店,因其离虹桥机场近、地理位置优越、商务风格、环境安静、设施齐全,较好满足三位基本要求。酒店特征描述如表 2 所示。

表 2 酒店特征

酒店	描述
上海西郊宾馆	设施齐全,地理优越,静谧,公园,草坪
上海万豪虹桥大酒店	设施齐全,地理优越,经典商务酒店
上海虹桥新华联索菲特大酒店	设施齐全,地理优越,周边“潮汕食府”

针对李彦宏,专注保守(性格),崇尚极简主义(观念),为其推荐经典的商务酒店即可,上海万豪虹桥大酒店乃不二之选;但针对马云,不像其他二位本身家庭条件优渥(环境)且拥有 IT 技术背景(知识),他创业历程艰辛坎坷(经历),且相对年龄较大(生理),年轻时不懈拼搏如今事业有成的他相当注重养生,喜爱太极(观念)。因此在满足基本需求的前提下,结合人性化特征,给马云推荐环境静谧,拥有可散步健身的大块草坪的上海西郊宾馆更为合适,符合其养生之道;针对马化腾,典型的广东潮汕人(环境),钟爱广式美食(观念),对于率性的他(性格),身处外地能很方便地品尝到家乡味道,定是人生幸事。因此在满足基本需求的前提下,结合人性化特征,给马化腾推荐周边就有“潮汕食府”的上海虹桥新华联索菲特大酒店更为合适,符合其饮食文化。上述三位人性化特征如表 3 所示。传统推荐与人性化推荐结果对比情况如表 4 所示。

表 3 人性化特征

特征	李彦宏	马云	马化腾
生理/年龄	49	53	46
性格	专注保守	开放善辩	率性务实
知识/专业学历	计算机硕士	外语本科	计算机本科
观念	崇尚极简主义	注重养生 喜爱太极	喜爱广式美食

续表 3

经历/创业	十来年 IT 工作后 创建百度	教书-开翻译社- 销售-几经波折后 创建阿里	十多年 IT 工作后 创建腾讯
环境/家庭条件	优渥	一般	优渥

表 4 推荐对比

姓名	传统推荐	人性化推荐
李彦宏	上海万豪虹桥大酒店	上海万豪虹桥大酒店
马云	上海万豪虹桥大酒店	上海西郊宾馆
马化腾	上海万豪虹桥大酒店	上海虹桥新华联索菲特大酒店

由此可见,客户选择酒店,其实是选择一个短期的生活圈^[10],仅仅满足其对硬件设施的需求是远远不够的^[13],更需从多方面考虑其人性化特性,让客户体会到被推荐的酒店处处都在为其着想,宾至如归^[12]。

因此,本文提出了一种基于人性化用户特征的在线酒店推荐方法。剖析、获取用户人性化特征形成酒店用户模型,结合特征匹配与协同过滤方法为用户推荐其感兴趣的酒店。并于课题项目“酒店管理营运博弈沙盘”进行实验,得出准确率、召回率、运行效率等综合表现最佳的推荐方式,且在一定程度上解决冷启动和数据稀疏的问题。

1 相关工作

目前大部分酒店推荐系统基于用户浏览和购买历史进行兴趣度分析,着力于采集电子商务网站中的大量注册、交易、评价等用户数据,和 Web 服务器中的日志数据、用户购物信息,但用户的人性化特性鲜为关注。

文献[1]采集用户已访问的酒店资源集合以及在线提交的资源要求等信息。根据用户于酒店网站上的在线行为,抽象出浏览用户特征并构建信息库,形成用户模块。分析用户兴趣特征和产品特征的相似性,利用聚类方法,挖掘出与用户兴趣相关的酒店产品。量化用户兴趣特征和产品特征,匹配算法简单明了,但产品特征需一一列举,且不易量化的特性均被忽略。文献[2-3]采集用户酒店评论信息,得到用户偏好和酒店评价数据。根据用户输入对酒店的要求(如城市,出游类型等信息,及对房间、食物、位置、服务、设备等的偏好),来构建用户模型。文献[2]认为群体用户的整体评价往往会忽略推荐用户某方面的特定需求。因此先为评论提取偏好,根据有相同或相似偏好的用户评价来形成推荐列表。协同过滤的思想,存在冷启动、

数据稀疏问题。文献[3]提出线下评估方法,线下提取酒店用户评论并进行处理得到项目特征矩阵,结合协同过滤,加快推荐速度,一定程度上优化冷启动问题但无法彻底解决,且需要不断更新离线资源。文献[4]采集用户内部数据(如国籍、性别、年龄等)和外部数据(如设备、社交媒体、时间、环境等)。通过对用户内外部信息数据进行语义特征提取等处理,并进行情感分析、行为分析、偏好分析,来构建用户模型。文献[5]采集用户基本信息和网站操作信息数据,根据用户基本信息(性别、年龄、职业等)、用户偏好信息及用户操作行为(学习记录、收藏夹等)来构建模型。文献[6]采集用户浏览网站时的眼动追踪数据和鼠标追踪数据,来得到用户偏好,再做基于内容的推荐。

综上,目前已有的推荐系统大多存在数据稀疏、冷启动以及对用户行为缺乏理解的问题。几乎没有挖掘用户人性化数据,确定其兴趣所在,推荐其真正感兴趣的酒店资源。因此,本文结合用户人性化特征,更准确、高效地为用户推荐酒店。

2 预备知识

2.1 用户建模与推荐技术

国内外相关研究机构与学者在用户建模以及推荐技术领域取得了大量的研究成果:目前广泛采用用户行为分析观察法来获取用户喜好信息,通过获取到的用户喜好信息建立用户模型^[8]。一般通过基于知识的推荐技术来直接启发用户主动提供偏好和需求,进而通过用户偏好结构以及关联规则和样例以及评论属性来建立用户轮廓模型^[9];基于内容的推荐技术主要从用户表示喜爱或者正面反馈的内容来获取用户偏好,进而通过基于关键词的向量空间模型、基于内容的标签网络以及潜在的主题模型等技术来构建用户轮廓模型;基于协同过滤的推荐技术通过获取到的用户偏好模型,运用余弦相似度、皮尔逊相似度等方法找到其相似用户^[11],再为其推荐相似用户喜爱的产品,以取得较高的推荐成效。然而这些方法所采取的用户数据相对较为单一,往往只是用户某个维度的体现,从而导致在推荐上的人性化效果不足。

2.2 人性化用户模型

人性化用户模型由作者所在课题组李银胜^[22]、沈剑平^[19]等人提出,亦称数字灵魂模型 Cyber-Anima。Cyber-Anima 模型旨在反映个体的内生特征概念,这些内生特征会对个体行为产生重大影响。该模型对用户

人性化特征进行数字处理,为探索用户欲望和推理用户意图提供了全新的视角。个体内生特征可从用户输入、互联网资源、应用系统、便携式设备等处获取的信息进行关联、相互转换,不断更新、完善。因此, Cyber-Anima 对于身处网络世界的个体来说是与生俱来的,在使用过程中无需再收集各种用户相关数据。所以,若结合人性化用户模型,可解决传统协同过滤方法的冷启动和数据稀疏问题^[22]。

人性化用户模型由六个维度构成,包括生理维度、性格维度、观念维度、知识维度、经历维度和环境维度,用 $A ::= \langle AID, D^I, D^E \rangle$ 表示^[19]。其中 AID 表示唯一标识, D^I 表示内生维度, $D^I ::= \langle D^I_{PH}, D^I_{BE}, D^I_{CH}, D^I_{EX}, D^I_{KN} \rangle$, 分别对应前5个维度要素; D^E 表示外生维度,即环境要素。各维度包含的具体内容如表5所示。

表5 六大维度内容

维度	内容
生理	基因, 身体素质, 疾病等
性格	温柔, 开放, 活泼, 害羞, 直白, 勇敢等
观念	信念, 信仰, 世界观, 人生观, 价值观等
知识	教育背景, 生活常识等
经历	学习, 工作, 旅行, 情感等
环境	家庭条件, 文化, 物理环境等

人性化模型采用基于本体的表示法,将六个维度中的属性作为本体的概念对和关系,从而结构化、语义化地描述用户^[14]。在构建过程中,通过确定本体的顶层概念,即用户模型的六个维度,然后按照六个维度的具体定义,将其子维度表示成相应的子概念。用户模型维度可以表示为 $D ::= C \times R$, 其中 C 表示概念集, R 表示关系集。其中关系是概念之间的联系,本文在构建人性化个体模型本体时主要运用的几种关系类型^[14]如表6所示。

表6 概念间关系类型

关系	定义	示例
分类	/r/isA	上位概念“isA”下位概念,如“单人间”是一种“房型”
部分	/r/partOf	下位概念“partOf”上位概念,如“工作经历”是“经历”的一部分
实例	/r/instanceOf	下位概念“instanceOf”上位概念,如“300元”是“房价”的一个实例
属性	/r/attributeOf	下位概念“attributeOf”上位概念,如“职业”是“我”的一个属性

2.3 语义特征提取

构建用户模型时涉及到语义特征提取技术,用户信息中往往包含大量文本信息。文本的特征项的选取即为从文本中抽取具有代表性的词来表示文本信息的特征^[19]。文本信息通常使用向量空间模型(VSM)来表示,每个文档表示为一个特征向量 $\mathbf{V} = (t_1w_1, t_2w_2, \dots, t_nw_n)$,其中 t_i 为词条项, w_i 为权值。

如果某个词或短语在一个文本中出现的频率 TF (term-frequency) 高,在其他文本中出现的频率低,则认为该词或短语具有良好的区别能力^[21]。IDF (Inverse Document Frequency) 指逆向文件频率,如果包含词条 t 的文档越少, IDF 越大,则说明词条 t 具有很好的类别区分能力。

结合 TF 与 IDF, TF-IDF 计算方法如下所示:

$$TF\text{-}IDF(w) = \frac{tf_i \log\left(\frac{N}{n_i} + 0.1\right)}{\sqrt{\sum_{i=1}^n (tf_i)^2 \times \log^2\left(\frac{N}{n_i} + 0.1\right)}} \quad (1)$$

式中: tf_i 为词 i 在训练集中的词频, N 为训练集文本数, n_i 未出现词 i 的文本数目。

3 基于人性化用户特征的酒店推荐

3.1 整体架构

本文依托于所在实验室提出的私有云浏览器“彩云阁”,在这里每个个体都拥有自己的 Cyber-Anima 模型^[22], 本文将应用到酒店推荐领域,提取出与酒店选择相关的用户特征,形成人性化酒店用户模型。综合用户特征与酒店特征匹配结果和协同过滤方法所得结果,可得酒店推荐候选集。如图 1 所示。

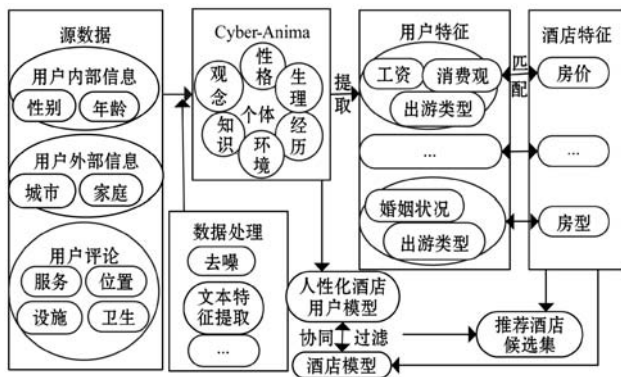


图1 整体架构

3.2 酒店模型

针对酒店可量化的参数,如星级、房型、房价、设施、设备、服务等,进行归一化处理,可得酒店特征参数表,如表 7 所示。

表 7 酒店特征参数

序号	酒店特征	数据处理值
1	星级	整型,取值为 1,2,3,4,5
2	房型	整型,1 - 单人房,2 - 双人房,3 - 家庭房,4 - 套房
3	房价	整型
4	停车场 - 酒店设施	布尔型,0 - 酒店无停车场,1 - 酒店有停车场
5	地铁站 - 周边设施	布尔型,0 - 酒店周边无地铁站,1 - 酒店周边有地铁站
⋮	⋮	⋮
k - 1	浴缸 - 房间设备	布尔型,0 - 房间内无浴缸,1 - 房间内有浴缸
k	接机 - 服务	布尔型,0 - 酒店不提供接机服务,1 - 酒店提供接机服务

酒店特征矩阵如下所示:

$$H(f) = (H_{11}, H_{12}, \dots, H_{1j}, H_{21}, \dots, H_{2j}, \dots, H_{ij}) \quad (2)$$

式中: H_{ij} 表示第 i 家酒店的第 j 个特征。两个酒店间相关性可通过余弦相似度得到,如下所示:

$$\text{sim}_h(i, j) = \frac{H_i \cdot H_j}{\|H_i\| \|H_j\|} \quad (3)$$

3.3 用户模型

与酒店特征一一对应,结合用户基本信息、输入信息及人性化信息可得用户偏好矩阵 $U(f)$,如下所示:

$$U(f) = (U_1, U_2, \dots, U_j) \quad (4)$$

式中: U_j 表示用户 u 对第 j 个特征的偏好。

为了更好地发掘用户之间的潜在联系,更进一步解决数据稀疏性问题,本文采用用户-用户矩阵对用户进行建模,并通过余弦相似度计算它们之间的相似度,如下式所示:

$$\text{sim}_u(u, v) = \frac{|U(u) \cap U(v)|}{\sqrt{|U(u)| \times |U(v)|}} \quad (5)$$

紧接着,用户 u 就可以通过 n 个相似的用户 (u_1, u_2, \dots, u_n) 和他们间的相似度 ($\text{sim}_{u_1}, \text{sim}_{u_2}, \dots, \text{sim}_{u_n}$) 表示。

3.4 推荐算法

将会影响酒店选择人性化特征引入到用户建模过程中,得到人性化酒店用户模型,再通过协同过滤的思想进行推荐。因此,称其为基于人性化特征的协同过滤 (Collaborative Filtering Based on Humanized Features), 简称 HF-CF。

在为酒店和用户分别建模后,可计算用户 u 对酒店 i 的期望评分 $pr(u, i)$,如下式所示:

$$pr(u, i) = r(v, j) \times \text{sim}_u(u, v) \times \text{sim}_h(i, j) \quad (6)$$

式中: $r(v, j)$ 表示用户 v 对酒店 j 的评分, $sim_u(u, v)$ 为用户 u 和用户 v 之间的相似度, $sim_h(i, j)$ 为酒店 i 和酒店 j 之间的相似度。

最后,根据期望评分可筛选出 Top-N 家酒店作为推荐候选集。

4 实验

4.1 评价指标

本文采用针对推荐方法常用的准确率、召回率和 F1-Score 三个指标来分析推荐效果^[20]。其中,准确率是指被推荐的列表中用户真正感兴趣的项目所占比例,反映了准确推荐的能力;召回率是指用户感兴趣的列表中推荐项目的所占比例,反映了全面推荐的能力;F1-Score 综合准确率和召回率,反映推荐服务的综合评价。推荐结果的列联表如表 8 所示。

表 8 列联表

推荐结果	推荐	不推荐
感兴趣	N_{tp}	N_{tm}
不感兴趣	N_{fp}	N_{fn}

因此,评价指标准确率、召回率、F1-Score 可通过以下公式计算:

$$P_{\text{recision}} = \frac{N_{tp}}{N_{tp} + N_{fp}} \quad (7)$$

$$R_{\text{ecall}} = \frac{N_{tp}}{N_{tp} + N_{tm}} \quad (8)$$

$$F1_Score = 2 \times \frac{P_{\text{recision}} \times R_{\text{ecall}}}{P_{\text{recision}} + R_{\text{ecall}}} \quad (9)$$

以上三大指标的值范围均 0-1,值越大表示推荐效果越好。但准确率和召回率相互影响,一般来说,若要达到高准确率,必须舍弃一些召回率,反之亦然。因此,准确率和召回率需要平衡,即让 F1_Score 值达到最佳。

4.2 实验平台及数据

本文以“酒店管理营运博弈沙盘”为实验平台,该沙盘是由所在实验室研发并已被数十家酒店管理院校使用的电子学习软件,旨在帮助学生更形象生动、直观快速地了解并掌握酒店管理营运知识。根据用户需求与偏好分配酒店是该软件核心功能之一,因此,一个准确高效的建模和推荐方法非常重要。

沙盘中酒店数据和用户数据,初期从各大酒店预订网站历史订单、评价中抓取、处理、分析后所得。在投入使用后,不断地产生新的订单数据,且各院校师生

会注入当地酒店数据与自身信息及偏好数据。整个数据库可供所有用户共享,真实性、有效性较高,并不断更新、维护。

本文获取沙盘数据库中 1 382 条用户数据、967 条酒店数据和 15 377 条订单数据进行实验。

4.3 基线算法

个性化推荐系统的推荐算法主要有基于内容的推荐、基于协同过滤的推荐、基于知识的推荐、基于效用的推荐和混合推荐等^[18]。在酒店推荐领域,由于酒店各项指标十分明确,有不少直接使用特征参数匹配的方法。本文采用特征参数匹配方法与基于协同过滤的推荐方法作为基线算法。

特征参数匹配方法 FM (Features Matching): 将服务项特征与用户偏好进行一一匹配,往往可以达到较高的准确率。但前期服务项数据与用户偏好信息的获取相当复杂,且数据量达到一定范围时,计算效率急剧下降,扩展性较差。

基于协同过滤的推荐方法 CF (Collaborative Filtering): 主要是通过对未评分项进行评分预测来实现,又可分为基于用户的协同过滤与基于项目的协同过滤^[15]。它只依赖用户行为,不需要对内容进行深入了解,适用范围广,可以用来推荐复杂项目,能起到意想不到的推荐结果^[16]。但前期需要大量的用户行为数据,即冷启动问题^[17];且用户量与项目量常常较为庞大,两者间存在用户行为关系的较为有限,即数据稀疏问题。本文运用人性化用户模型,可较好地解决这两个问题。

4.4 实验结果与分析

通过相关数据采集、处理,酒店建模,用户建模,基于协同过滤计算,可得出推荐结果。其中,用户模型中用 n 个相似用户及相似度表示用户 u 。当用户模型参数 n 取不同值时,推荐效果如表 9 所示。

表 9 用户模型参数 n 不同时推荐效果

n 取值	准确率	召回率	F1-Score
3	0.63	0.64	0.63
4	0.63	0.66	0.64
5	0.65	0.69	0.67
6	0.67	0.70	0.68
7	0.67	0.69	0.68
8	0.66	0.70	0.68
9	0.67	0.70	0.68

当 n 小于或等于 6 时,随着数值增大,各项指标都逐渐升高; n 大于 6 后,各项指标上升空间极为有限。

结合推荐效率,取 n 为 6 时的推荐效果与基线算法对比。根据三大指标统计结果如表 10 所示。

表 10 实验结果比较

指标	HF-CF	FM	CF
准确率	0.67	0.71	0.67
召回率	0.70	0.64	0.68
F1-Score	0.68	0.67	0.65

可以看出,本文提出的方法整体推荐效果良好,召回率和 F1-Score 值均比两种基线算法的高。尽管准确率相比特征参数匹配方法较低一点,但可扩展性和推荐效率更为客观。

5 结 语

本文分析了现有酒店推荐系统的现状,其主要缺点在于只对用户表象信息进行提取,缺乏全方面深入挖掘用户的人性化特征,如知识、性格、观念等。因此,本文提出一种基于人性化特征的酒店推荐方法,更加精准地分析用户的真实意图,提高个性化推荐的效果。以基于协同过滤的推荐算法作为基线算法,结合用户人性化特征,计算出推荐结果,并通过实验验证了该方法的可行性和有效性。

接下来仍有诸多工作需要继续,更多与酒店相关的业务知识和市场情况有待深入学习并分析,让研究更符合现实;更多人性化特征需要不断探索发掘,来完善人性化用户模型,让推理和推荐逻辑更严谨;此外,可引入更先进的推荐算法,提高准确率和推荐效率。

参 考 文 献

- [1] Xiong Y N, Geng L X. Research on personalized recommendation model based on dynamic interestingness for online hotel reservation [C]//International Conference on Internet Technology and Applications. IEEE, 2010:1-5.
- [2] Sharma Y, Bhatt J, Magon R. A multi-criteria review-based hotel recommendation system[C]//IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing. IEEE, 2015:687-691.
- [3] Takuma K, Yamamoto J, Kamei S, et al. A hotel recommendation system based on reviews: what do you attach importance to? [C]//Fourth International Symposium on Computing and NETWORKING. IEEE, 2017:710-712.
- [4] Tanjim-Al-Akib M, Ashik L K, Hosne-Al-Walid, et al. User-modeling and recommendation based on mouse-tracking for e-commerce websites [C]//International Conference on Computer and Information Technology. IEEE, 2017:517-523.
- [5] Zhao J, Chen L, Guo J. Research and construction of personalized user interest model for distance learning [C]//IEEE, Advanced Information Technology, Electronic and Automation Control Conference. IEEE, 2017:2324-2327.
- [6] Abdar M, Yen N Y. Design of a universal user model for dynamic crowd preference sensing and decision-making behavior analysis[J]. IEEE Access, 2017, 5(99):24842-24852.
- [7] Glushkova T. Personalization and user modeling in adaptive e-learning systems for schools [M]//E-Learning—Instructional Design, Organizational Strategy and Management. Intech, 2015:127-147.
- [8] Fasahte U, Gambhir D, Merulingkar M, et al. Hotel recommendation system[J]. Imperial Journal of Interdisciplinary Research, 2017, 3(11):317-324.
- [9] Zhang K, Wang K, Wang X, et al. Hotel recommendation based on user preference analysis [C]//IEEE International Conference on Data Engineering Workshops. IEEE, 2015:134-138.
- [10] Lin K P, Lai C Y, Chen P C, et al. Personalized hotel recommendation using text mining and mobile browsing tracking [C]//IEEE International Conference on Systems, Man, and Cybernetics. IEEE, 2016:191-196.
- [11] Chang Z, Arefin M S, Morimoto Y. Hotel Recommendation Based on Surrounding Environments [C]//Iai International Conference on Advanced Applied Informatics. IEEE, 2013.
- [12] Zulkefli N A B M, Baharudin B B. Hotel travel recommendation based on blog information [C]//International Symposium on Mathematical Sciences and Computing Research. IEEE, 2016:243-248.
- [13] Jannach D, Gedikli F, Karakaya Z, et al. Recommending hotels based on multi-dimensional customer ratings [M]//Information and Communication Technologies in Tourism 2012. Springer Vienna, 2012:320-331.
- [14] Zhang M, Li Y, Zhou A, et al. A recommendation system for travel services based on cyber-anima [C]//IEEE, International Conference on E-Business Engineering. IEEE Computer Society, 2015:114-118.
- [15] Davidson J, Liebald B, Liu J, et al. The YouTube video recommendation system [C]//ACM Conference on Recommender Systems. ACM, 2010:293-296.
- [16] Smith B, Linden G. Two decades of recommender systems at Amazon.com [J]. IEEE Internet Computing, 2017, 21(3):12-18.
- [17] Gomez-Urbe C A, Hunt N. The netflix recommender system: algorithms, business value, and innovation [J]. ACM Transactions on Management Information Systems (TMIS), 2016, 6(4):1-19.

- [18] Berkovsky S, Freyne J. Web personalization and recommender systems[C]//ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 2307-2308.
- [19] 沈剑平. 数字灵魂模型及其在智能推荐中的应用研究[D]. 上海: 复旦大学, 2015.
- [20] Niemann K, Wolpers M. A new collaborative filtering approach for increasing the aggregate diversity of recommender systems [C]//Acm Sigkdd International Conference on Knowledge Discovery & Data Mining. ACM, 2013.
- [21] Alexandrou D, Xenikoudakis F, Mentzas G. SEMPAT: adapting clinical pathways by utilizing semantic technologies [C]//2009 13th Panhellenic Conference on Informatics. IEEE Computer Society, 2009.
- [22] Li Y, Wu B, Zhong Z, et al. Cyber-A(nima): Modelling, reasoning and applications[C]//IEEE Internet of People and Smart City Innovation, 2017.

(上接第 11 页)

即使线性方法由于缺乏灵活性而不再能够利用和适应额外数据,附加标签信息仍然对这些方法有益。玻璃和外壳数据的典型例子如图 2 所示。

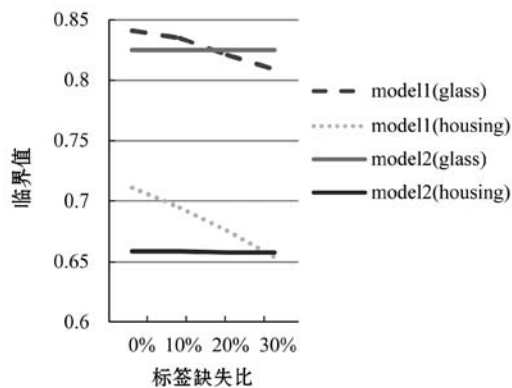


图 2 排名表现

4 结语

本文提出了一种将广义线性模型和 P-L 模型结合作为底层数据生成过程的新模型方法,并利用实验对比了其与传统的广义线性模型方法及基于实例的 Mallows 模型的优劣,实验结果显示本文构建的新模型方法在学习排序问题研究方面具有一定的优越性。尤其是在不完全的训练数据情况下,新的模型方法在计算与性能上均有更好的表现。此外,本实验也证明了广义线性模型在处理不同类别标签分类上对 P-L 模型也提供了一定的补充。

同时,本文的概率模型 1 采用最大似然估计的方法作为标准拟合模型的补充,减少了模型估计中存在

的偏差,一定程度上提高了模型的准确度,更符合模型的前提假设。此外,通过最大似然估计的估计方法也允许实验者通过添加不同类型的统计信息来补充实验预测,使实验更加精准可靠。

虽然本文提出的新模型在一定程度上为研究学习排序提供了新的研究思路与方法,但仍存在一些问题需要解决。例如本文提出的解决多分类标签排序的方法建立在第一部分广义线性模型对标签进行了正确分类的基础上,如果广义线性模型确定的分类结果存在偏差,则整个模型方法的可靠性与精准性均会受到一定程度的影响。针对该问题,拟将本文提出的模型方法进行更深层次地合并,通过类似局部线性回归等方法,放宽原模型中较为严格的假设,从而保证实验的准确与可靠。

参 考 文 献

- [1] Cao Z, Qin T, Liu T Y, et al. On text clustering with side information[C]//Proceedings of International Conference on Machine Learning(ICML). 2007: 129-136.
- [2] Har-Peled S, Dan R, Zimak D. Constraint classification for multiclass classification and ranking[C]//Proceedings of the 15th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2002: 809-816.
- [3] Hüllermeier E, Fürnkranz J, Cheng W, et al. Label ranking by learning pairwise preferences[J]. Artificial Intelligence, 2008, 172(16):1897-1916.
- [4] Hüllermeier E, Fürnkranz J. On loss functions in label ranking and risk minimization by pairwise learning[J]. Journal of Computer and System Sciences, 2010: 49-62.
- [5] Cheng W, Hühn J, Hüllermeier E. Decision tree and instance-based learning for label ranking[C]//Proceedings of the 26th Annual International Conference on Machine Learning. New York:ACM, 2009: 161-168.
- [6] Cheng W, Dembczynski K, Hüllermeier E. Label ranking methods based on the plackett-luce model[C]//International Conference on Machine Learning. DBLP, 2010: 215-222.
- [7] 王惠文, 黄乐乐, 王思洋. 基于函数型数据的广义线性回归模型[J]. 北京航空航天大学学报, 2016, 42(1): 8-12.
- [8] Hunter D R. MM algorithms for generalized bradley-terry models[J]. Annals of Statistics, 2004, 32(1):384-406.
- [9] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016:199-201.
- [10] Demšar J. Statistical comparisons of classifiers over multiple data sets[J]. Journal of Machine Learning Research, 2006, 7(1):1-30.
- [11] 周志华. 机器学习[M]. 北京:清华大学出版社, 2016:42-44.