河北省乡村旅游推荐系统研究与设计

高万林1，张晓建1，于丽敏1，2

（1.中国农业大学信息与电气工程学院，北京 100083； 2.山东农业工程学院，山东 250100）

摘 要：近年来推荐技术在旅游领域中的应用受到很多科研工作者的青睐，推荐系统能够向游客推荐其潜在的喜好信息，帮助游客根据自己过去行为喜好进行选择。我国乡村旅游已经成为一种趋势，但是推荐技术在乡村旅游中的应用却较少有人研究。本文首先分析了乡村旅游在季节、旅游对象、地理空间分布和重游率方面具有季节性强、游客多为城镇人、游客出游阈值为距离城市20km、重游率高的特点。其次依据季节性和地理空间分布特征提出了一种融合时间和地理位置上下文的协同过滤算法，最后在已建立的河北省乡村旅游项目基础上设计了乡村旅游推荐系统。该系统包括信息展示交互、日志采集处理、推荐引擎、数据存储四个主要模块，并对每个模块的功能和业务流程进行了介绍。最终使得河北省乡村旅游系统能够积极推送给用户个性化的旅游信息，帮助用户做出快速准确的选择。

关键词：乡村旅游；推荐系统；个性化推荐；协同过滤；上下文

中图分类号：TP311.1；TP312 文献标识码：A

Research and Design of Hebei Rural Tourism Recommendation System

Gao Wanlin1, Zhang Xiaojian1，Yu Limin1,2

(1. *Department of Computer Science*, *China Agricultural University,* *Beijing* 100084*, China;*

2. *Department of Computer Science, Shandong Agricultural Engineering Institute, Shandong* 250100*, China;*)

Abstract: In recent years, the application of recommended technology in the field of tourism has been favored by many researchers. The recommendation system can recommend visitors to their potential preferences and help visitors to choose. China's rural tourism has become a trend, but the recommended technology in rural tourism applications are less research. In this paper，we first aiming at the characteristics of rural tourism in the season, tourists, geography spatial distribution and re - voyage rate. In the season, with seasonal characteristics. In tourists, most of them are urban people. In geography spatial, travel trip threshold is 20km from the city, which means distance from the city within 20 ~ 100km within the scope of the rural tourism is the frequent occurrence of the zone. Secondly, base on the characteristics of seasonal and geo-spatial distribution, a collaborative filtering algorithm combining time and geographical context is proposed. Finally, based on the established rural tourism project in Hebei province, the rural tourism recommendation system. The system includes information display interaction, log collection and processing, recommendation engine, data storage four main modules, and each module functions and business processes are introduced. Finally make it possible that Hebei rural tourism system can actively push to the user personalized tourism information. To help users make quick and accurate choice.

Key words: rural tourism；recommended system；Personalized recommendation；Collaborative filtering；Context

引言

[[1]](#footnote-1) 乡村旅游[1,2]是指以乡村地区为活动场所，利用乡村独特的自然环境、田园景观、生产经营形态、民俗文化风情、农耕文化、农舍村落等资源，为城市游客提供观光、休闲、体验、健身、娱乐、购物、度假的一种新的旅游经营活动。近年来我国乡村旅游得到了快速发展，据中国产业调研网发布的《2017-2020年中国休闲农业与乡村旅游市场现状研究分析与发展前景预测报告》[3]显示，截至2015年4月底，全国有9.5万个村开展休闲农业与乡村旅游活动，休闲农业与乡村旅游经营单位达193万家，其中农家乐达220万家，规模以上园区超过4.1万家，年接待游客接近8.4亿人次，年营业收入超过3200亿元。同时黄蓉[4]博士通过问卷调查并运用Logistic模型分析，说明了中国国内旅游的一大趋势是从城镇流向农村。因此如何帮助游客快速准确的找到自己感兴趣的景点，提高用户寻找信息的效率，从而促进乡村旅游的发展是非常值得研究的课题。

推荐系统被定义成一种信息过滤系统，用于预测用户对物品的“评分”或“偏好”[5,6]。目前推荐系统在旅游中的应用没有电商网站如亚马逊网[7,8]，DVD租赁网站Netflix[9,10]，淘宝网[11]等应用广泛和成熟，处于研究和发展阶段。具体应用如Triplehop公司开发的TripMatcher（由[www.ski-europe.com](http://www.ski-europe.com)使用）和VacationCoach公司开发的expert advice platform（由[www.travelcity.com](http://www.travelcity.com)使用），模拟现实中游客与旅行社间的咨询过程，根据用户当前需求，偏好和提出的约束条件给用户提供旅游方案。去哪儿网根据用户输入的目的地、出发日期、旅游天数等信息给用户推荐旅游景点或旅行包。Werthner[12]等人研究发现随着信息和通信技术（ICT）的发展，旅行者正在减少对旅游机构的依赖，旅游信息重点来源于网络。但随着互联网的发展，旅游信息已经出现过载现象。Park和Jang[13]的研究证明当旅游选择过多的时候，效果往往是相反的。加上人们时间碎片化，因此旅游平台需要借助推荐技术来帮助游客做出快速而准确的旅游建议，从而节省游客时间和提高平台效益。尽管目前我国乡村旅游已经成为一种趋势，但据调研发现推荐技术在乡村旅游中还没有得到研究和应用。由于乡村旅游相对于城市旅游在季节、旅游对象、地理空间分布和重游率方面有着自己的特性，无法直接使用城市旅游的推荐技术和推荐系统，因此结合乡村旅游自身的特点研究出一种适合乡村旅游的推荐系统具有重要的意义。

推荐系统产生推荐列表的算法目前主要分为基于协同过滤推荐（Collaboration filtering Recommendation），基于知识推荐（Knowledge-based Recommendation）,基于内容推荐（Content-based Recommendation）和混合推荐技术（Hybrid recommendation）。基于协同过滤推荐技术最早被提出，同时也是研究最多，实际应用最广的一种推荐技术。在旅游推荐中也被广泛研究和应用[14-16]。协同过滤算法又可以分为两类方法[17-19]基于近邻方法（neighborhood-based）和基于模型方法（model-based）。由于基于模型的方法是利用用户过往行为学习并建立预测模型，需要经过大量真实历史数据进行模型训练，如果数据量不够或者模型选择不好，则效果将远低于基于近邻的方法。

基于近邻方法又可以分为基于用户（user-based CF）和基于物品（item-based CF）的推荐。user-based CF给用户推荐那些和他有共同兴趣爱好的用户喜欢的物品，推荐结果着重反应和用户兴趣相似的小群体热点，是比较古老的算法，早在1992年已经在电子邮件的个性化推荐系统Tapestry中得到了应用。item-based CF给用户推荐那些和他之前喜欢的物品类似的物品，推荐结果着重于维护用户的历史兴趣，是相对比较新的算法，在著名的电子商务网站亚马逊[8]和DVD租赁网站Netflix[9]中得到了广泛应用。由于item-based CF适用于用户数量远大于物品数量的情况，并且可以提供令人信服的推荐解释。故本系统采用item-based CF作为基本算法。Lamsfus等人的研究指出[20]上下文情境对旅游选择有着重要作用，这也是最近旅游领域研究的热点，因此本文提出了一种结合河北省乡村旅游特征，运用基于物品的协同过滤算法，同时融合用户时间和地理位置上下文信息的乡村旅游推荐算法，并对河北省乡村旅游推荐系统进行了设计。

1 河北省乡村旅游特征分析

研究乡村旅游推荐系统，必须根据乡村旅游本身的特点进行推荐系统设计。根据前人对乡村旅游的研究总结出河北省乡村旅游在季节性、旅游对象、地理空间分布和重游率方面相对于城市旅游都有鲜明的特征。具体特征总结为如下四点：

1）季节性强[21]，虽然城市旅游也存在季节性，但乡村旅游季节性尤为明显。河北省位于我国偏北部春、夏、秋、冬四季分明，乡村旅游主要以自然风光、民俗风情、农业特色为主，不同季节旅游特色和主题都会有区别，凸显出严重的淡旺季现象。例如:对于喜欢采摘的游客，不同季节可采摘的种类大不相同，冬季采摘的淡季，夏季和秋季是采摘的旺季。对于热爱自然风景的游客，不同季节风景也会大不相同。

2）游客多为城镇人，城市快节奏生活，导致了越来越多的城市病，这些现象促使了城镇居民和工作者驱车去乡村远离城市，亲近自然，寻找清净。这种需求极大地促进了乡村旅游的发展，因此乡村旅游主要是针对城镇人度假休闲。

3）在地理空间分布上，由于乡村相对交通不发达，住宿餐饮等设施基础不完善，因此乡村旅游游客多选择自驾游。这就使得乡村旅游在地理空间分布上具有局限性。某些距离城镇较远的景点将会被很少选择。根据卢小丽[22]等人对国内外乡村旅游近20年的研究热点总结出，游客乡村旅游出游阈值为距离城市 20km，即距离城市20~100km之内的空间范围是乡村旅游的频繁发生地带。

4）重游率高,这也是区别于其他旅游的一大特点。乡村旅游大多是被作为休闲娱乐、缓解压力、亲近自然的选择，相对重游率非常高。

2 关键算法与技术

2.1 基于物品的协同过滤技术（Item-CF）

基于物品的协同过滤算法的核心思想是通过分析用户的行为记录计算物品之间的相似度。主要有两个步骤，一是计算物品之间的相似度，二是根据物品之间的相似度和用户的历史行为给用户生成推荐列表。

本文以余弦值计算物品间相似度，假设系统中用户个数为n，物品数量为m，则用户对物品的评分矩阵是个n\*m的矩阵R。记rp,q表示用户up对物品iq的评分，则每个物品都能表示成所有用户对它的评分向量，即iq=<r1,q,r2,q,……rn,q>,则物品间相似度可以通过如下公式进行计算：

计算好物品间相似度，下一步是结合用户历史行为，即用户对历史物品的评分，给用户生成初始推荐列表。本文是根据预测用户对物品的评分，从而推断用户对物品的喜爱程度。item-based CF算法是根据该物品的邻居的评分来预测该物品的评分。公式如下：

其中，Nk(ip)表示物品的邻居集合，即与物品ip最相似的k个物品集。K取25。

2.2 融合上下文技术

上下文又称为推荐系统中的情境，情境一般在推荐系统中起到过滤作用，应用时采取的方式有情境预过滤、情境后过滤和情境建模三种方式[24]。近年来上下文在旅游推荐系统中的应用越来越受到研究者的重视。如：Trattner[25]等人将天气条件作为影响POI的上下文，结果显示明显的提高了Rank-GeoFM算法。Kaya和 Bridge[26]提出了一种推荐拍摄位置的方法来改进现有算法。另有Vakeel 和 Ray[27]将动机定义为可用于向用户解释传递的推荐的语境因素。 他们区分不同POI签入用户的七种动机，作为后置过滤方法的依据。本文将采用情境后过滤的方式，基于位置和时间情境过滤掉无关的推荐和调整列表里推荐的排名。

位置上下文，根据前面所提到的乡村旅游的特点，距离城市20~100km之内的空间范围是乡村旅游的频繁发生地带。因此当用户所在位置位于城镇时，将把距离城镇20-100km的景点优先推荐给用户。当用户已经在乡村旅游过程中，则将距离近的景点优先推荐给用户。

时间上下文，根据前面所提到的乡村旅游的特点，季节性强，本文将首先获取到用户的系统时间，然后将时间泛化成季节来对推荐列表进行过滤。

3 乡村旅游推荐系统设计

3.1 河北省乡村旅游平台介绍

河北省乡村旅游平台主要是针对河北省乡村旅游发展现状及特色优势，为乡村旅游者提供信息化的乡村旅游服务，并基于数据挖掘和推荐技术，给用户提供个性化的服务。目标是拓展乡村旅游服务渠道，从而带动河北省乡村旅游产业的发展。平台架构设计图如下：

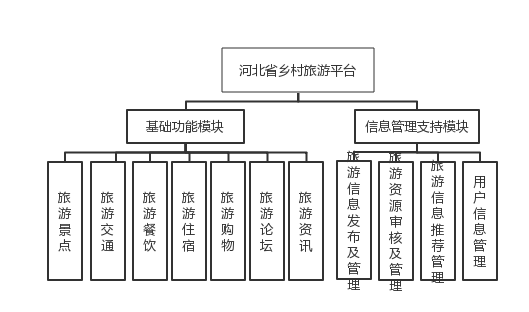


图1河北省乡村旅游平台架构

Fig.1 The Structure of Rural Tourism Platform in Hebei Province

编程语言：Java，开发工具：eclipse，服务器：tomcat7.0，数据库：MySQL，采用SpringMVC+hibernate技术架构开发。

乡村旅游推荐系统是基于河北省乡村旅游平台进行设计和开发的。使用Java语言，用户的行为日志收集，分析和存储，以及推荐算法的设计都写入河北省乡村旅游平台中，即将推荐系统功能嵌入到乡村旅游平台中。推荐系统的设计在下节具体介绍。

3.2 推荐系统设计

乡村旅游推荐系统对用户过往的旅游信息进行分析，对用户的选择给出建议，帮助用户克服信息过载和选择性难题。图1描述了乡村旅游推荐系统的框架。该系统由信息展示交互模块、日志采集处理模块、推荐引擎模块、数据存储模块四大主要部分组成。接下来介绍每部分的主要功能和系统业务流程描述。

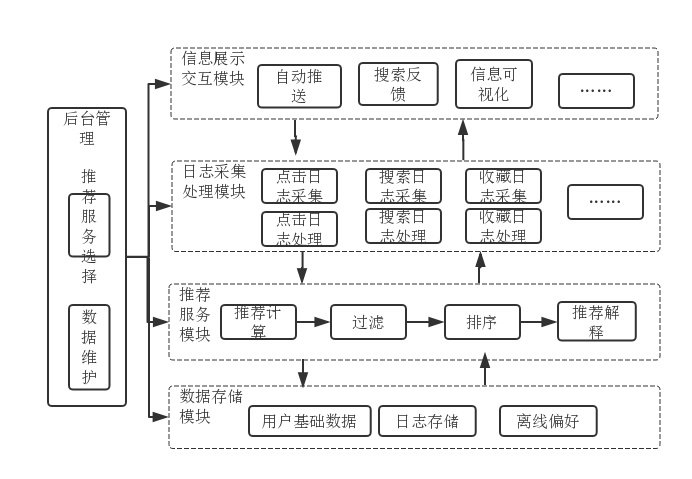


图2 乡村旅游推荐系统架构图

Fig.2 Architecture diagram of rural tourism recommendation system

3.2.1 信息展示交互模块

信息展示交互模块主要用于以图形化界面的形式完成系统与用户之间信息的交互和数据传递。交互的方式主要有系统自动推送和用户主动搜索反馈。

系统自动推送包括两种情况，一是用户登录后自动将离线计算的推荐列表推荐给用户，这是在用户不经过任何操作的情况下。另一种是用户登录后在浏览过程中对一些推荐信息进行点击、浏览、收藏行为，这里假设用户点击、浏览、收藏行为代表用户对该对象感兴趣，这些行为会生成行为日志，被日志采集模块进行采集分析处理，并作为下一步推荐列表的依据。

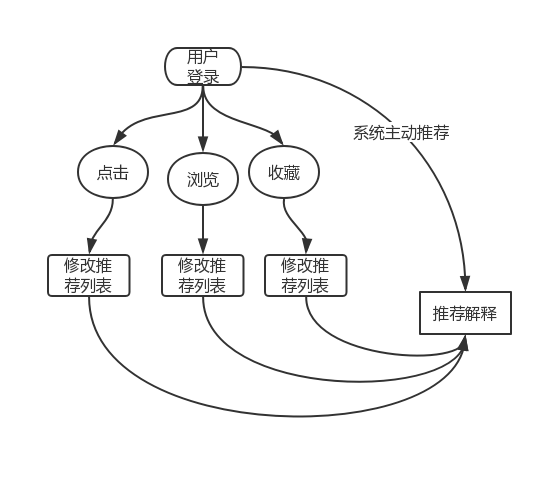


图3 系统主动推送流程图

Fig.3 The system actively pushes the flow chart

用户主动搜索反馈，是指用户带有目的性的去搜索，系统根据用户搜索关键字并结合推荐列表给用户做出反馈。

推荐信息的可视化本系统采用2种形式，第一种是传统的列表展示形式，即直接根据推荐列表的排序顺序进行展示；第二种是展示在地图上，调用百度地图API，这里同时会计算出景点到用户当前位置的距离，最后将景点及景点到用户当前位置的路径显示在地图上。

3.2.2 日志采集处理模块

用户行为的记录有一大部分依赖日志信息，因此日志信息采集和处理扮演着重要角色，此功能主要对用户点击日志、搜索日志、收藏、购买日志进行采集。日志采集处理模块采用flume框架，主要分为三层：Agent层、Collector层和Store层。其中Agent层负责日志收集工作；Collector层部署在中心服务器上，用于接收Agent层发送的日志数据，并经过简单处理后写到相应的Store层中；Store层负责提供永久或者临时的日志存储服务，并将日志传送给推荐引擎模块作为推荐依据。

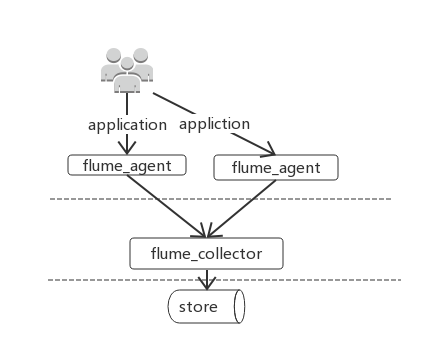


图4 信息采集原理流程图

Fig.4 Information collection principle flow chart

3.3.3 推荐引擎模块

推荐引擎是推荐系统功能的核心，负责推荐信息的计算，过滤和排序。借鉴项亮[23]的《推荐系统实践》上推荐引擎架构图，本系统中推荐引擎架构原理如下：

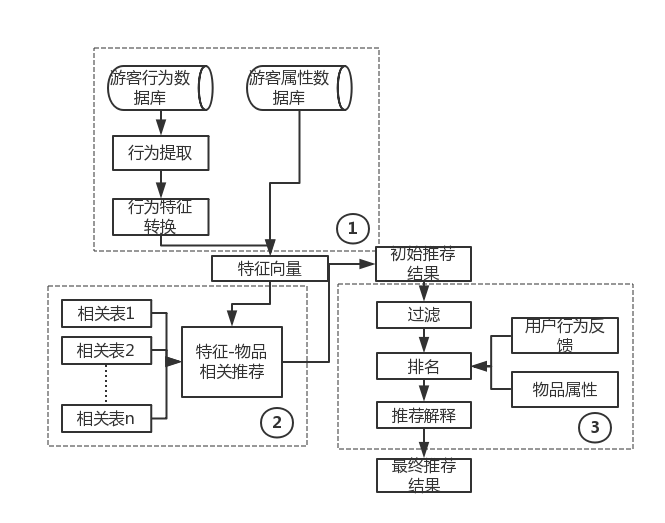


图5 推荐引擎架构图

Fig.5Recommended engine architecture

其中部分①负责从数据库和缓存中拿到用户行为数据，通过分析不同的行为生成当前用户的特征向量。此模块最终输出的是用户的特征向量。部分②负责将用户的特征向量结合相关表转化为初始推荐物品列表。部分③负责对初始推荐列表进行过滤、排名处理，同时给出推荐解释，生成最终推荐列表。

3.3.4 数据存储模块

数据存储包括基本的用户数据和物品数据存储，同时也包括离线推荐列表的存储，日志存储等，本文同河北省乡村旅游平台共同使用MySQL数据库进行数据存储。MySQL数据库使用标准的 SQL数据语言形式，具有体积小、速度快，支持Java语言，多线程，并且支持GIS空间扩展。可以满足系统的需求。

4 结束语

针对推荐技术在旅游中的应用逐渐成熟，但目前缺乏在乡村旅游中的应用的情况，本文提出了结合乡村旅游自身特征，将推荐技术应用到乡村旅游领域中的思想。主要包括：

1）研究了乡村旅游的特点，即季节性强，游客多为城镇人，地理空间分布规律，重游率高的特点。并将季节性强和地理空间分布规律作为上下文来提高推荐的精度。

2）做出了乡村旅游推荐系统的设计，主要包括了信息展示交互模块、日志采集处理模块、推荐引擎模块、数据存储模块四大模块，并对设计的模块功能进行了介绍。

3）介绍了主要用到的推荐技术——基于物品的协同过滤技术和融合上下文技术。并对技术原理进行了介绍。

本研究是对乡村旅游推荐技术和系统的初步研究，本文考虑还有两点不足之处，一是由于本研究主要考虑了用户旅游时间和用户当前地理位置的影响，影响用户旅游选择的因素还有很多，例如：用户年龄、天气等的影响，本文计划下一步进行更多因素的研究。二是本文设计的系统主要是针对河北省乡村旅游的特点设计，如果要推广到全国乡村旅游，在不同地区需要考虑的因素就会有差别，本文计划下一步研究不同地区乡村旅游的特征区别，从而将乡村旅游推荐系统应用到全国性的乡村旅游平台。

参考文献

1 郭焕成,韩非.中国乡村旅游发展综述[J].地理科学进展,2010,(12):1597-1605.

2 张祖群.当前国内外乡村旅游研究展望[J].中国农学通报,2014,(08):307-314.

3 中国产业调研网. 2017-2020年中国休闲农业与乡村旅游市场现状研究分析与发展前景预测报告. 2016[M]. 中国产业调研网, 2016.

4.黄蓉, 中国城镇居民的国内旅游需求研究, 2015, 华中科技大学.

5 Grossman L. How Computers Know What We Want—Before We Do[J]. TIME magazine, 2010, 27.

6 Lu J, Wu D, Mao M, et al. Recommender system application developments: a survey[J]. Decision Support Systems, 2015, 74: 12-32.

7 Linden G, Smith B, York J. Amazon. com recommendations: Item-to-item collaborative filtering[J]. IEEE Internet computing, 2003, 7(1): 76-80.

8 Smith B, Linden G. Two Decades of Recommender Systems at Amazon. com[J]. IEEE Internet Computing, 2017, 21(3): 12-18.

9 Gomez-Uribe C A, Hunt N. The netflix recommender system: Algorithms, business value, and innovation[J]. ACM Transactions on Management Information Systems (TMIS), 2016, 6(4): 13.

10 Hallinan B, Striphas T. Recommended for you: The Netflix Prize and the production of algorithmic culture[J]. New Media & Society, 2016, 18(1): 117-137.

11 Gu X, Fu Q, Zu Z, et al. Recommending content information based on user behavior: U.S. Patent 9,400,995[P]. 2016-7-26.

12 Werthner H, Alzua-Sorzabal A, Cantoni L, et al. Future research issues in IT and tourism[J]. Information Technology & Tourism, 2015, 15(1): 1-15.

13 Park J Y, Jang S C S. Confused by too many choices? Choice overload in tourism[J]. Tourism Management, 2013, 35: 1-12.

14Nilashi M, bin Ibrahim O, Ithnin N, et al. A multi-criteria collaborative filtering recommender system for the tourism domain using Expectation Maximization (EM) and PCA–ANFIS[J]. Electronic Commerce Research and Applications, 2015, 14(6): 542-562.

15 Ortega F, Hernando A, Bobadilla J, et al. Recommending items to group of users using Matrix Factorization based Collaborative Filtering[J]. Information Sciences, 2016, 345: 313-324.

16 Koohi H, Kiani K. A new method to find neighbor users that improves the performance of Collaborative Filtering[J]. Expert Systems with Applications, 2017, 83: 30-39.

17 Koren, Yehuda, and Robert Bell. "Advances in collaborative filtering." *Recommender systems handbook*. Springer US, 2015. 77-118.

18 Shi Y, Larson M, Hanjalic A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges[J]. ACM Computing Surveys (CSUR), 2014, 47(1): 3.

19 Hernando A, Bobadilla J, Ortega F. A non negative matrix factorization for collaborative filtering recommender systems based on a Bayesian probabilistic model[J]. Knowledge-Based Systems, 2016, 97: 188-202..

20 Lamsfus C, Wang D, Alzua-Sorzabal A, et al. Going mobile: Defining context for on-the-go travelers[J]. Journal of Travel Research, 2015, 54(6): 691-701.

21 张欣然, 破解乡村旅游季节性困境的对策研究. 农业经济, 2017(02): 第58-59页.]

22 卢小丽, 成宇行与王立伟, 国内外乡村旅游研究热点-近20年文献回顾. 资源科学, 2014(01): 第200-205页.

23 项亮. 推荐系统实践[M]. 人民邮电出版社, 2012.

24 Adomavicius G, Tuzhilin A. Context-aware recommender systems[M]//Recommender systems handbook. Springer US, 2015: 191-226.

25 Trattner C, Oberegger A, Eberhard L, et al. Understanding the Impact of Weather for POI Recommendations[C]//RecTour@ RecSys. 2016: 16-23.

26 Kaya M, Bridge D. Improved recommendation of photo-taking locations using virtual ratings[J]. RecTour 2016, 2016.

27 Vakeel K A, Ray S. A Motivation-Aware Approach for Point of Interest Recommendations[J]. Proceedings of RecTour, 2016.

1. 基金项目：中央高校基本科研业务费专项（2017XS002）;

   作者简介：高万林（1965—），男，教授，博士生导师，主要从事计算机网络与信息安全研究，E-mail：gaowlin@cau.edu.cn [↑](#footnote-ref-1)