

DOI: 10.16652/j.issn.1004-373x.2019.23.039

基于改进协同过滤技术的个性化旅游线路推荐研究

艾静超

(沈阳工程学院, 辽宁 沈阳 110036)

摘 要: 为降低旅游路线制定的盲目性与随机性,解决海量旅游信息导致旅游路线选择困难的问题,提出基于改进协同过滤技术的个性化旅游线路推荐方法。根据旅游者拍摄的照片分析旅游者实际旅游足迹,得到旅游点热度,根据旅游点热度确定两个旅游者间的偏好一致度,得到旅游者近邻。根据近邻在各旅游点的浏览时间确定旅游点偏好程度,采用余弦计算方法确定近邻对旅游点偏好程度与旅游者对旅游点偏好程度的偏差值,构建基于用户的协同过滤模型。为防止基于用户的协同过滤模型中冷启动与数据稀疏性问题发生,将基于用户的协同过滤模型和基于地理位置的旅游路线推荐模型相结合,配合旅游者与旅游点地理位置信息,推荐满足旅游者偏好的个性化旅游路线。旅游路线推荐结果显示,所提方法在基于旅游者当前位置向旅游者推荐个性化旅游路线的同时,可确保线路中不存在路线交叉往返现象,降低行程花费时间1 h左右。

关键词: 个性化旅游; 线路推荐; 协同过滤; 偏好程度; 偏差确定; 偏好确定

中图分类号: TN911.1-34; TP391.3

文献标识码: A

文章编号: 1004-373X(2019)23-0182-05

Research on individualized travel route recommendation based on improved collaborative filtering technology

AI Jingchao

(Shenyang Institute of Engineering, Shenyang 110036, China)

Abstract: In order to reduce the blindness and randomness of tourist route formulation and solve the problem of tourist route selection difficulty caused by massive tourist information, a personalized tourist route recommendation method based on improved collaborative filtering technology is proposed. The tourist actual footprints are analyzed according to the photographs taken by tourists to get hot tourist spots. The preference consistency between two tourists is determined according to the tourist spots' hot degree to get the two tourists' approximation. The preference degree of tourist spots is determined according to the approximation in sightseeing time for each tourist spot. The cosine calculation method is used to determine deviation value of the approximation's preference degree to a tourist spot and tourists' preference degree to a tourist spot. The collaborative filtering model based on user is constructed. In order to prevent cold start and data sparsity in the user-based collaborative filtering model, the user-based collaborative filtering model and the location-based tourism route recommendation model are combined to realize the personalized tourist route recommendation that meet the preferences of tourists in combination with the geographic location information of tourists and tourist spots. The results of tourist route recommendation show that the proposed method can not only recommend personalized tourist routes to tourists based on their current location, but also ensure that there is no cross-trip phenomenon in the routes.

Keywords: personalized tourism; route recommendation; collaborative filtering; preference degree; deviation determination; preference determination

0 引 言

当前社会经济高速发展,导致旅游业竞争态势越来越激烈,降低旅游路线制定的盲目性与随机性^[1],提升旅

游路线个性化、为旅游者提供更多可选择的旅游路线成为当下相关学科及行业的研究方向。“自由行”是当下较为热门的旅游方式,但大多数旅游者在选择自由行时通常面临这样一个问题——海量旅游信息导致旅游路线选择困难,个性化旅游线路推荐方法是解决这一问题的有效方法^[2]。

以往在推荐旅游线路时多利用运筹学方法判断最优旅游路线^[3],该类方法忽略旅游者个性化喜好与需求、

收稿日期:2019-04-17

修回日期:2019-05-10

基金项目:沈阳工程院校内科研立项(RWYB-1506)

Project Supported by Research Project in Shenyang Institute of Engineering (RWYB-1506)

也未考虑旅游者旅游时间等条件,导致旅游路线推荐结果缺乏个性化与科学性。为解决这一问题,本文提出基于改进协同过滤技术的个性化旅游线路推荐方法,考虑旅游点热度和旅游者偏好程度,改进基于用户的协同过滤模型中冷启动与数据稀疏性问题,结合协同过滤推荐模型与基于地理位置的推荐模型,将科学、合理且具有个性化的旅游路线推荐给旅游者。

1 个性化旅游路线推荐方法

1.1 基于用户的协同过滤模型

基于用户的协同过滤推荐是基于全部用户对事物的偏好程度^[4],采用“K-邻居”的分类方法挖掘同当前用户偏好一致的“邻居”用户群,根据这K个邻居的历史偏好信息,为当前用户提供推荐服务。在个性化旅游路线推荐方法中,可将偏好一致度较高的用户集合推荐给旅游者。基于用户的协同过滤推荐具体过程如下。

1.1.1 构建旅游点关联图

当旅游者在陌生区域旅游时,一般情况下会先选取感兴趣的旅游点,然后依照旅游时间选取旅游路线。由于互联网的普及,越来越多的人将自己在旅游过程中拍摄的照片与旅游路线等信息上传至社交网络中与他人分享,分析大量该类信息,可得到旅游者实际旅游线路,确定旅游者对各类旅游点的兴趣和旅游点热度^[5]。

在离线状态下构建旅游点关联图时,将全部旅游者旅游序列内的旅游点作为关联图内节点,表示旅游点、旅游者在旅游序列内的游览路线可作为关联图的边。

Photo ID(照片编号)、User ID(游览者所在地)、Time(拍摄时间)、Longitude(拍摄地经度)、Latitude(拍摄地纬度)以及Category(类别)共同构成旅游者分享的照片数据结构^[6]。由此可知照片数据中涵盖旅游者准确时空位置信息,基于各照片的经纬度采用Haversine公式可得到旅游者分享的各照片同所游览区域内各旅游点的距离,当距离低于200 m时,可认定照片是在某旅游点拍摄的,据此获取旅游者的旅游点列表,用 $S_u = (p_1, p_2, \dots, p_n)$ 表示。

旅游点热度评价计算方法为:

$$\text{Popular}(p) = N(p)/N_{\max} \quad (1)$$

式中: $N(p)$ 和 N_{\max} 分别表示旅游者在旅游点 p 处拍摄的照片数量和全部同类旅游点内旅游者拍摄的最大照片数量。

1.1.2 构建邻接矩阵

判断两个旅游者偏好一致度时可采用邻接矩阵^[7],判断过程中,采用余弦相似度确定两个旅游者是否具有

一致的偏好类型,确定旅游者 u_1 和 u_2 间偏好一致度的表达式为:

$$\cos(u_1, u_2) = \frac{\sum_{i=1}^k \text{Popular}(p)_1 \cdot \text{Popular}(p)_2}{\sqrt{\sum_{i=1}^k \text{Popular}(p)_1^2 \cdot \text{Popular}(p)_2^2}} \quad (2)$$

通过式(2)能够得到两个旅游者旅游偏好之间的一致度。由于不同旅游者对同一旅游点的热度评价有所差异,例如,有些旅游者在某些旅游点未拍摄照片,则将该旅游者对此旅游点的热度评价定义成一个中间值。对比预先设定的阈值和偏好一致度计算结果,当偏好一致度高于设定阈值时,可将两个旅游者归为一类。

1.1.3 偏好程度计算

根据旅游者的历史旅游线路,获取基于时间的旅游者兴趣偏好^[8]。旅游者浏览一个旅游点时都会在该旅游点停留一定时间,在全部旅游者的历史旅游路线中依照游览过的旅游点、到达旅游点的时间和离开旅游点的时间可得到各旅游者在游览过的各旅游点的游览时间。据此得到任意旅游者在游览任意旅游点时所用的平均时间,用 $\bar{V}(p)$ 表示,表达式如下:

$$\bar{V}(p) = \frac{1}{n} \sum_{u=U} \sum_{p_x \in S_u} (t_{p_x}^d - t_{p_x}^a) \sigma(p_x = p), \quad \forall p \in P \quad (3)$$

式中: u 表示旅游者; U 和 n 分别表示全部旅游者和全部旅游者内游览旅游点 p 的数量; $t_{p_x}^d$ 和 $t_{p_x}^a$ 分别表示到达旅游点 p_x 和离开旅游点 p_x 的时间; $\sigma(p_x = p) = \begin{cases} 1, & p_x = p \\ 0, & \text{其他} \end{cases}$ 。

由于旅游者在各旅游点的平均游览时间无法实际描述其对此类旅游点的兴趣偏好程度,因此采用基于时间的旅游者兴趣偏好实施描述^[9]。旅游者 u 对旅游点类别属性 c 的偏好程度的表达式为:

$$\text{Int}(u, c) = \sum_{p_x \in S_u} \frac{(t_{p_x}^d - t_{p_x}^a)}{\bar{V}(p_x)} \sigma(\text{Cat}_{p_x} = c), \quad \forall c \in C \quad (4)$$

式中: Cat_p 表示旅游点 p 的类别属性, $\sigma(\text{Cat}_{p_x} = c) = \begin{cases} 1, & \text{Cat}_{p_x} = c \\ 0, & \text{其他} \end{cases}$ 。

由式(4)可判断旅游者 u 对旅游点类别属性 c 的偏好程度,与全部旅游者在相同旅游点平均游览时间相比,由式(4)得到的偏好程度是根据旅游者在各存在类别属性 c 的旅游点所花费的时间获取的,能够体现一个旅游者对具有类别属性 c 的旅游点偏好程度。

1.1.4 旅游路线推荐

根据用户协同过滤的思想^[10],通过邻接矩阵确定与旅游者偏好一致的近邻,获取近邻对旅游点偏好程度,在确定旅游者对未知旅游点偏好程度时,需借鉴近邻对

此旅游点偏好程度。采用余弦计算方法确定近邻对旅游点偏好程度与旅游者对旅游点偏好程度的一致性^[11],并得到两者间的偏差值,基于得到的偏差值预测旅游者对推荐景点的偏好程度,则基于用户的协同过滤模型计算公式如下:

$$P(U, L)_1 = \lambda \cdot \frac{\sum_{u=1, c=1}^U (\text{Int}(u, c) - \overline{\text{Int}(u, c)}) \cdot \cos(U, U_v)}{\sum_{v=1}^n \cos(U, U_v)} + \overline{\text{Int}(u, c)} \quad (5)$$

式中: λ , $\overline{\text{Int}(u, c)}$ 和 $\cos(U, U_v)$ 分别表示旅游点的修正参数、旅游者对全部旅游点平均偏好程度和旅游者与近邻的相似度。当偏差值计算结果与设定阈值相比较小时,则说明该旅游点适合推荐给旅游者,循环此过程即可得到全部推荐给旅游者的旅游点集合。

上述基于用户的协同过滤模型中易存在冷启动与数据稀疏性问题^[12],这些问题将导致旅游路线推荐有效性降低,因此需要改进协同过滤模型,将位置信息考虑其中。

1.2 基于地理位置的旅游路线推荐模型

列举旅游者全部感兴趣的旅游点,同时利用最小直径圆法分别确定各旅游点同中心点的方位^[13],在保障此圆中涵盖全部旅游点的同时确保圆直径尽可能小,以不同旅游点同圆心距离的平均数作为旅游者活动半径。若一个旅游点同圆心间距离过大时,需删除该旅游点并重新确定圆心方位。

通过分析 Foursquare 数据集内全部旅游者的活动半径得到^[14]旅游者偏好度较高的旅游点多存在区域性,通常情况下,未知旅游点同旅游者当前方位间距离若小于活动半径,旅游者会选择该旅游点。由此可知基于地理位置的旅游路线推荐模型是通过确定未知旅游点同活动中心间距离来判断旅游者对该旅游点的偏好程度,旅游点所处方位越远,旅游者游览意愿越低,则基于地理位置的旅游路线推荐模型描述为:

$$P(U, L)_2 = \frac{a}{x + b} \quad (6)$$

式中: $P(U, L)_2$ 和 x 分别表示旅游者游览意愿和推荐旅游点同旅游者活动中心点间距离; a, b 均为常数。

为防止基于用户的协同过滤模型中冷启动与数据稀疏性问题发生,将基于用户的协同过滤模型和基于地理位置的旅游路线推荐模型相结合,进行加权计算得到最终的个性化旅游线路推荐模型:

$$P(U, L) = \alpha P(U, L)_1 + \beta P(U, L)_2 \quad (7)$$

式中: α 和 β 均为加权值,两者之和为1,同时两者的值

也将随旅游者偏好而调整^[15]。当确定全部推荐旅游点后,利用加权dijkstra算法在全部推荐旅游点中确定最优旅游路线。

2 旅游路线推荐分析

2.1 数据来源

数据主要来源于 WAMDM 实验室的 Flickr 数据集,其中包含两方面内容:一方面是沈阳市与青岛市共 420 221 条照片数据,数据中包括照片编号、拍摄时间、拍摄经纬度等信息;另一方面是 23 002 名旅游者社交网站通讯录。

2.2 旅游路线推荐结果

实验为测试本文提出的基于协同过滤技术的个性化旅游路线推荐方法的推荐性能,以沈阳市和青岛市为实验对象,采用本文方法进行旅游路线推荐,结果如图1所示。



a) 沈阳市旅游路线推荐结果



b) 青岛市旅游路线推荐结果

图1 旅游路线推荐结果

Fig. 1 Recommended results of tourist routes

图1a)为沈阳市旅游路线推荐结果显示,以黎明广场作为出发地和结束地,由于旅游者当前所在位置位于沈阳市中心区域,推荐的旅游路线中包含北陵公园、怒江公园、丁香湖、铁西森林公园、浑河晚渡公园和长白岛森林公园,旅游路线中未出现路线交叉现象。由图1b)青岛市旅游路线推荐结果可知,以中山公园为出发地和结束地,由于旅游者当前所在位置位于青岛市南部边缘

位置,排除较远的旅游点,得到的推荐路线中包含青岛啤酒博物馆、贮水山儿童公园、北岭山森林公园、浮山森林公园、青岛大学和滨海风景区,旅游路线中同样未出现路线交叉现象。实验结果表明本文推荐方法可有效推荐旅游路线,并且两个城市旅游路线推荐结果中旅游点均为景观类,说明本文方法推荐的旅游路线均符合旅

游者的个性化选择。

2.3 游览时间对比

在出发地、出发时间与旅游点数量相同的条件下,对比采用本文方法获取的旅游路线推荐结果与采用运筹学的推荐方法得到的旅游路线推荐结果,对比结果如表 1,表 2 所示。

表 1 沈阳市旅游路线推荐结果

Table 1 Recommended results of tourism routes in Shenyang

路线	本文方法			对比方法		
	到达时间	游览时间 /min	结束时间	到达时间	游览时间 /min	结束时间
黎明广场	07:00	0	07:00	07:00	0	07:00
北陵公园	07:39	100	09:39	07:39	120	09:39
怒江公园	10:02	60	11:02	10:14	120	12:14
丁香湖	11:34	100	13:34	12:47	60	13:47
铁西森林公园	14:07	50	15:07	14:18	60	15:18
浑河晚渡公园	15:47	60	16:47	16:01	60	17:01
长白岛森林公园	17:01	60	18:01	17:26	60	18:26
黎明广场	18:40	0	18:40	19:28	0	19:28

表 2 青岛市旅游路线推荐结果

Table 2 Recommended results of tourism routes in Qingdao

路线	本文方法			对比方法		
	到达时间	游览时间 /min	结束时间	到达时间	游览时间 /min	结束时间
中山公园	07:00	0	07:00	07:00	0	07:00
青岛啤酒博物馆	07:16	30	07:46	07:16	30	07:46
贮水山儿童公园	08:00	50	09:00	08:00	60	09:00
北岭山森林公园	09:44	40	10:44	09:44	60	10:44
浮山森林公园	11:39	100	13:39	11:52	120	13:52
青岛大学	14:03	40	15:03	14:44	120	16:44
滨海风景区	15:30	100	17:30	17:13	60	18:13
中山公园	17:53	0	17:53	18:57	0	18:57

分析表 1 和表 2 得到,在出发地、出发时间与旅游点数量相同的条件下,根据本文方法得到的旅游路线推荐结果游览沈阳市和青岛市时,行程历时分别为 430 min 和 360 min,与采用运筹学的推荐方法相比分别节省 50 min 和 90 min。游览时间产生差异的主要原因在于本文方法推荐的旅游路线考虑当前旅游者所在位置,旅游路线中不存在路线交叉往返现象,而采用运筹学的推荐方法推荐的旅游路线均在不同程度上出现路线交叉往返现象。实验结果表明本文方法推荐的旅游路线可降低行程花费时间,且更符合人们的旅游习惯。

3 结 论

随着旅游业的蓬勃发展,旅游路线推荐问题成为相

关学科及行业的研究方向。本文提出基于改进协同过滤技术的个性化旅游路线推荐方法,综合考虑旅游点热度和旅游者兴趣偏好,为防止基于用户的协同过滤模型中冷启动与数据稀疏性问题发生,利用基于用户的协同过滤模型及基于地理位置的推荐模型向旅游者推荐最优旅游路线。旅游路线推荐结果显示,本文方法能够基于旅游者当前所在位置有效推荐个性化旅游路线,且推荐结果不存在路线交叉往返现象,降低行程花费时间。

参 考 文 献

[1] 常亮,曹玉婷,孙文平,等.旅游推荐系统研究综述[J].计算机科学,2017,44(10):1-6.

CHANG Liang, CAO Yuting, SUN Wenping, et al. Review on

- tourism recommendation system [J]. Computer science, 2017, 44(10): 1-6.
- [2] 姜信景, 齐小刚, 刘立芳. 个性化信息推荐方法研究[J]. 智能系统学报, 2018, 13(2): 189-195.
JIANG Xinjing, QI Xiaogang, LIU Lifang. Research on the recommendation method of personalized information [J]. CAAI transactions on intelligent systems, 2018, 13(2): 189-195.
- [3] 刘凯, 张立民, 周立军. 深度学习在信息推荐系统的应用综述[J]. 小型微型计算机系统, 2019, 40(4): 738-743.
LIU Kai, ZHANG Limin, ZHOU Lijun. Survey of deep learning applied in information recommendation system [J]. Journal of Chinese computer systems, 2019, 40(4): 738-743.
- [4] 吴彦文, 齐旻, 杨锐. 一种基于改进型协同过滤算法的新闻推荐系统[J]. 计算机工程与科学, 2017, 39(6): 1179-1185.
WU Yanwen, QI Min, YANG Rui. A news recommendation system based on an improved collaborative filtering algorithm [J]. Computer engineering and science, 2017, 39(6): 1179-1185.
- [5] 吴清霞, 周娅, 文锦尧, 等. 基于用户兴趣和兴趣点流行度的个性化旅游路线推荐[J]. 计算机应用, 2016, 36(6): 1762-1766.
WU Qingxia, ZHOU Ya, WEN Diyao, et al. Personalized trip itinerary recommendation based on user interests and points of interest popularity [J]. Journal of computer applications, 2016, 36(6): 1762-1766.
- [6] 李渊, 丁燕杰, 王德. 旅游者时间约束和空间行为特征的景区旅游线路设计方法研究[J]. 旅游学刊, 2016, 31(9): 50-60.
LI Yuan, DING Yanjie, WANG De. A new approach for designing tourist routes by considering travel time constraints and spatial behavior characteristics of tourists [J]. Tourism tribune, 2016, 31(9): 50-60.
- [7] 肖晓丽, 钱娅丽, 李旦江, 等. 基于用户兴趣和社交信任的聚类推荐算法[J]. 计算机应用, 2016, 36(5): 1273-1278.
XIAO Xiaoli, QIAN Yali, LI Danjiang, et al. Clustering recommendation algorithm based on user interest and social trust [J]. Journal of computer applications, 2016, 36(5): 1273-1278.
- [8] 龚敏, 邓珍荣, 黄文明. 基于用户聚类与Slope One填充的协同推荐算法[J]. 计算机工程与应用, 2018, 54(22): 144-148.
GONG Min, DENG Zhenrong, HUANG Wenming. Collaborative recommendation algorithm based on user clustering and Slope One filling [J]. Computer engineering and applications, 2018, 54(22): 144-148.
- [9] 钟足峰, 段尧清, 杨曼. 可提高多样性的基于重排序图书推荐算法研究[J]. 现代情报, 2017, 37(12): 61-65.
ZHONG Zufeng, DUAN Yaoqing, YANG Man. Improving aggregate recommendation diversity of books using ranking-based techniques [J]. Modern information, 2017, 37(12): 61-65.
- [10] 付永平, 邱玉辉. 一种基于贝叶斯网络的个性化协同过滤推荐方法研究[J]. 计算机科学, 2016, 43(9): 266-268.
FU Yongping, QIU Yuhui. Method of personalized collaboration filter recommendation based on Bayesian network [J]. Computer science, 2016, 43(9): 266-268.
- [11] 李昆仑, 万品哲, 张德智. 基于改进用户相似性度量和评分预测的协同过滤推荐算法[J]. 小型微型计算机系统, 2018, 39(3): 567-571.
LI Kunlun, WAN Pinzhe, ZHANG Dezhi. Collaborative filtering recommendation algorithm based on improved user similarity measure and scoring forecast [J]. Journal of Chinese computer systems, 2018, 39(3): 567-571.
- [12] 李倩, 许伟, 蒋洪迅. 考虑用户口碑的旅游计划个性化推荐方法研究[J]. 管理评论, 2016, 28(6): 113-118.
LI Qian, XU Wei, JIANG Hongxun. Personalized recommendation for traveling planning based on online word-of-mouth [J]. Management review, 2016, 28(6): 113-118.
- [13] 李雅美, 王昌栋. 基于标签的个性化旅游推荐[J]. 中国科学技术大学学报, 2017, 47(7): 547-555.
LI Yamei, WANG Changdong. Tag-based personalized travel recommendation [J]. Journal of University of Science and Technology of China, 2017, 47(7): 547-555.
- [14] 孙彦鹏, 古天龙, 宾辰忠, 等. 基于多重隐语义表示模型的旅游路线挖掘[J]. 模式识别与人工智能, 2018, 31(5): 462-469.
SUN Yanpeng, GU Tianlong, BIN Chenzhong, et al. Travel routing mining based on multiple latent semantic representation model [J]. Pattern recognition and artificial intelligence, 2018, 31(5): 462-469.
- [15] 侯乐, 杨辉华. 基于ILS-CS优化算法的个性化旅游线路研究[J]. 计算机科学与探索, 2016, 10(1): 142-150.
HOU Le, YANG Huihua. Research on personalized trip itinerary based on ILS-CS optimization [J]. Journal of frontiers of computer science & technology, 2016, 10(1): 142-150.

作者简介: 艾静超(1979—), 女, 辽宁锦州人, 硕士, 讲师, 主要从事教育教学管理及旅游研究。

欢迎订阅 2020 年度《物联网技术》(月刊)

邮发代号: 52-253

定价: 20 元/册

全年定价: 240 元

电话: 029-85241792-8625

传真: 029-85241792-8618