

# 一种改进的基于位置的推荐算法

刘建东

(吉首大学张家界学院 旅游与管理工程学院,湖南 张家界 427000)

**摘要:**基于位置的推荐算法在根据位置信息划分数据子集时会产生数据稀疏问题,对此提出了一种改进的推荐算法。该算法充分考虑了不同位置所产生的推荐结果的差异性,分别为相应的推荐结果设置不同权重,然后线性求和。改进算法既解决了数据的稀疏问题,又考虑了用户兴趣的本地化。实验结果表明,改进的算法提高了推荐准确性。

**关键词:**基于位置;推荐算法;数据稀疏

**DOI:**10.11907/rjdk.161555

**中图分类号:**TP312

**文献标识码:**A

**文章编号:**1672-7800(2016)009-0039-02

## 0 引言

随着在线社交网络与智能手机技术的快速发展,用户不仅仅满足于在社交平台用文字和图片来分享经历,还希望分享更丰富的信息,如在何时何地做了哪些事情等。现在大部分智能手机都能提供基于地点的签到服务,在线社交网络如 Foursquare、 Gowalla 等网站也支持基于地点的签到服务。因此,用户可利用智能手机在社交网络分享去过的地点。这些分享的地点构成了用户的运动轨迹。显然,经常去户外运动的用户与喜欢文艺的用户分享的地点是有差异的。因此,具有不同兴趣的用户分享的运动轨迹是不一样的。换句话说,用户兴趣不同,分享的地点也就不同,地点与用户兴趣之间产生了关联。因此,如何根据用户的签到地点来提供个性化服务是研究人员感兴趣的课题。

目前,基于地点的推荐算法有两个方向:①根据用户之间的距离进行推荐。如 Foursquare 通过计算用户与好友之间的距离推出探索功能,该功能可以为用户推荐附近的好友,LBS 软件也提供了推荐附近餐馆与商店的功能;②根据用户的签到地点为用户进行分类。因为用户的签到地点相同,用户兴趣可能相似,典型的是一个被称为 LARS(Location Aware Recommender System)的推荐系统。该系统由明尼苏达大学的研究人员提出,主要思想是将用户按兴趣相似度进行分类,为同类的用户推荐物品或者地点。评价兴趣相似度的标准有两个:①用户共同喜欢的物品越多,用户的兴趣越相似;②用户签到的地点越接近,用户的兴趣越相似。对于第①个标准而言,其实就是

基于物品的协同过滤算法,该算法是经典的推荐算法,具体过程可参看文献。本文主要讨论第②个标准。因为签到地点具有多级结构,所以 LARS 根据地点的范围从大到小逐级划分数据集,如签到地点具有国家、省、市的树状结构。LARS 首先根据国家把数据集划分成不同的子集  $set_1, set_2, \dots, set_i, \dots$ ;接着按省份将数据集  $set_i$  划分为  $(set_{i1}, set_{i2}, set_{i3}, \dots, set_{id})$  等子集,然后依次按市划分。一般来说,  $set_{i1}, set_{i2}, set_{i3}, \dots, set_{id}$  数据子集内用户兴趣的相似度比  $set_i$  内用户兴趣的相似度更高,因此,划分子集提高了推荐准确性。但是按 LARS 划分数据子集的用户数据量可能不够,进而影响推荐效果。子集内用户数据量少的问题称为数据稀疏问题,本文主要工作是针对可能存在的数据稀疏问题提出解决办法,改善推荐效果。

## 1 数据稀疏问题

用户在使用在线社交网络时喜欢分享自己的位置,从而会形成如下形式的数据记录:(用户,用户位置,物品,评分)。本文用  $(u_i, p_i, w_k, s_i)$  表示第  $i$  个用户在位置  $p_i$  评论第  $k$  个物品时所生成的数据。用户兴趣不同,分享的位置会有差异,所以 LARS 会根据用户签到位置对用户数据集进行划分。但是位置信息是一个树状结构,比如,国家、省、市、县的结构。因此,数据集也会被划分成一个树状结构,本文把该树状结构称为树状数据集。树状数据集的节点包含了所有同一个位置的用户行为数据。

如图 1 所示,用户  $u$  的评论数据信息是:  $(u, p_{22}, w_k, s_j)$ ,那么推荐算法就会根据位置信息依次找到“ $p$ ”,“ $p_2$ ”,“ $p_{22}$ ”这 3 个节点,在找到节点后把该数据信息依次划分

**作者简介:**刘建东(1978—),男,湖南城步人,吉首大学张家界学院旅游与管理工程学院讲师,研究方向为计算机应用技术。

到“ $p$ ”, “ $p_2$ ”, “ $p_{22}$ ”3个节点上。一般地, 在划分数据集过程中,  $p$  节点上的数据被划分为  $i$  个子集 ( $p_1, p_2, \dots, p_i$ ), 从概率上来说,  $p_i$  节点上的数据量只是  $p$  节点上的  $\frac{1}{i}$ 。而随着算法的进行,  $p_k$  节点上的数据被划分为  $j$  个子集 ( $p_{k1}, p_{k2}, \dots, p_{kj}$ )。同样的道理, 子集  $p_{kl}$  ( $1 \leq l \leq j$ ) 节点上的数据量只是  $p_k$  节点上的  $\frac{1}{j}$ 。更一般地, 假设第  $n$  层节点的分支数是  $N_n$ , 那么第  $n$  层上任一个节点的数据量只是总数据量上的  $\frac{1}{\prod_{c=1}^n N_c}$ 。其中  $c$  表示层次。显然, 层次越深的

节点所表示的位置越准确, 但是所得到的数据量越少。

因为数据量降低的幅度很大, 所以随着子集的划分, 产生数据稀疏的可能性就越高。如果某个节点上的数据量太少, 那么基于物品的协同过滤算法在该节点就无法有效进行, 影响推荐的准确性, 因此要对数据稀疏问题提出解决办法。

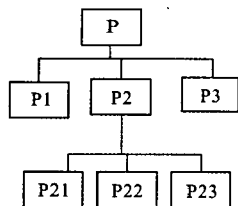


图1 节点

## 2 基于位置推荐的改进算法

### 2.1 算法改进思路

基于位置的推荐算法根据位置信息对数据集进行划分, 划分的层次越多, 同类中用户之间的位置越接近, 用户兴趣的相似度越高, 推荐效果越准确。但是另一方面, 层次越多, 划分的子集包含的数据量越少, 这样又会降低推荐的准确性。因此, 划分子集时既要考虑用户兴趣的相似度, 也要考虑子集的数量。基于位置的推荐算法问题在于仅计算深层次节点, 即仅考虑了兴趣相似度, 却忽略了数据量; 相反, 如果只考虑顶层节点, 也会存在问题。因为顶层节点的用户数据量大, 用户兴趣的相似度低, 这样会忽略用户兴趣的相似度。总而言之, 不能只考虑单独的节点。因为不同层次的节点在用户兴趣的相似度和用户数据量上有各自的优点和不足, 所以改进算法的思路是综合考虑不同层次的节点, 利用基于物品的协同过滤算法为每个节点的数据子集生成推荐列表, 然后分别为每个推荐列表设置不同的权重, 最后按设置的权重将这些推荐列表进行线性相加, 选择排名最高的前  $N$  条记录推荐。改进的基于位置的推荐算法流程如图2所示, 因为位置信息都是省一市一县的三层树状结构, 所以本算法也只考虑三层划分。

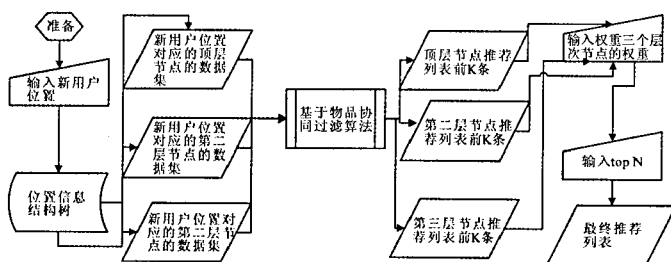


图2 改进的基于位置的推荐算法流程

### 2.2 改进算法具体步骤

#### 2.2.1 输入

输入用户  $u_k$  的位置  $p_k$  以及树状数据集  $Data\_Tree$ , 在使用基于物品的协同过滤算法时, 输入物品最相似的个数  $K$ , 然后输入前三层节点的数据集对应的推荐权重 ( $\lambda_1, \lambda_2, \lambda_3$ )。其中  $\lambda_i$  ( $1 \leq i \leq 3$ ) 代表数据集树中第  $i$  层结点上的用户行为产生的推荐列表权重, 最后输入推荐列表数目  $N$ 。

#### 2.2.2 逐步计算推荐列表

(1) 在  $Data\_Tree$  中根据用户位置  $p_k$  从根节点开始搜索, 相应找到路线上的节点  $N_1, N_2, N_3$  其中  $N_i$  ( $1 \leq i \leq 3$ ) 表示第  $i$  层上的节点。

(2) 得到节点  $N_i$  对应的用户数据子集  $UserSet(N_i)$ 。

(3) 利用基于物品的协同过滤算法得到  $UserSet(N_i)$  的推荐列表  $List(i)$ 。

$$w_{kl} = \frac{|W(k) \cap W(l)|}{\sqrt{|W(k)| |W(l)|}} \quad (1)$$

$$I_{uk} = \sum_{l \in N(u) \cap S(l, K)} w_{kl} r_{ul} \quad (2)$$

公式(1)和公式(2)是基于物品的协同过滤算法计算公式, 其中公式(1)是计算物品  $k$  与物品  $l$  之间的相似度,  $w_{kl}$  是用户数据子集中物品  $k$  与物品  $l$  的相似度,  $W(k)$  表示喜欢物品  $k$  的用户数,  $W(l)$  表示喜欢物品  $l$  的用户数。公式(2)是计算用户  $u$  对物品  $k$  的兴趣, 其中  $N(u)$  表示用户  $u$  喜欢物品的集合,  $S(l, K)$  表示和物品  $l$  最相似的  $K$  个物品的集合,  $w_{kl}$  是用户数据子集中物品  $k$  和物品  $l$  的相似度,  $r_{ul}$  是用户  $u$  对物品  $l$  的兴趣。

#### 2.2.3 线性相加计算结果

根据每个节点对应的推荐权重, 计算  $List = \sum_{i=1}^3 \lambda_i * List(i)$  得到综合推荐列表, 然后进行相似度排序, 选择前  $N$  个物品进行推荐。

## 3 实验及结果分析

实验采用 MovieLens 提供的测试数据集, MovieLens 数据集来自用户对电影的评价。该数据集包含了用户位置信息。实验将从数据集中提取 100 位用户对 800 部电影的评价信息, 选择其中 80% 的数据作训练集, 另外 20% 作测试集。

实验依次设置权重参数  $\lambda_1, \lambda_2, \lambda_3$  的值为 0.1、0.2、0.7, 然后以基于位置的推荐算法和改进的推荐算法做实

# 八数码问题解法效率比较及改进研究

付宏杰<sup>1</sup>,王雪莹<sup>2</sup>,周 健<sup>2</sup>,周孙静<sup>2</sup>,朱 珠<sup>2</sup>,张俊余<sup>2</sup>

(1. 吉林工程技术师范学院 信息工程学院,吉林 长春 130052;

2. 东北师范大学 信息与软件工程学院,吉林 长春 130117)

**摘 要:**八数码问题是人工智能中的一个典型问题,目前解决八数码问题的搜索求解策略主要有深度优先搜索、宽度优先搜索、启发式 A\* 算法。对这些算法进行研究,重点对 A\* 算法进行适当改进,使用曼哈顿距离对估价函数进行优化。对使用这些算法解决八数码问题的效率进行比较,从步数、时间、结点数、外显率等各参数,通过具体的实验数据分析,进一步验证各算法的特性。

**关键词:**八数码;深度优先搜索;宽度优先搜索;A\* 算法;曼哈顿距离

DOI:10.11907/rjdk.161867

中图分类号:TP312

文献标识码:A

文章编号:1672-7800(2016)009-0041-05

## 1 问题描述

八数码问题就是在一个大小为  $3 \times 3$  的九宫格上,放置 8 块编号为 1~8 的木块,九宫格中有一个空格,周围(上下左右)的木块可以和空格交换位置。对于问题,给定一个初始状态(如图 1(a)初始状态),目标状态(如图 1(b)目标状态)是期望达到 1~8 顺序排列的序列,并且空格在右下角,问题的实质就是寻找一个合法的移动序列。

验。实验结果发现,改进算法的推荐结果准确性与召回率提高 10% 左右。其中准确性和召回率的计算公式如公式(3)和公式(4)所示。

$$Precision@N = \frac{\sum_u |R(u, N) \cap T(u)|}{\sum_u R(u, N)} \quad (3)$$

$$Recall@N = \frac{\sum_u |R(u, N) \cap T(u)|}{\sum_u T(u)} \quad (4)$$

## 4 结语

本文针对基于位置的推荐算法在划分数据子集时产生的数据稀疏问题,提出了对不同节点的推荐列表进行线性加权的解决办法。实验结果表明,改进的推荐算法提高

## 2 问题分析和模型表示

### 2.1 模型建立

对于棋格,每个位置给定一个编号,从左上角开始顺序编号(见图 2),对于任意的状态,记为  $p = s[0]s[1]s[2]s[3]s[4]s[5]s[6]s[7]s[8]$ ,图 1 初始状态中的状态可以表示为  $p = 283164705$ ,图 2 目标状态中的状态可以表示为  $p' = 123456780$ 。

了推荐结果的准确性。

### 参考文献:

- [1] 项亮. 推荐系统实践[M]. 北京:人民邮电出版社,2012.
- [2] GOWALLA[EB/OL]. <http://mashable.com/category/gowalla/>.
- [3] 刘树栋,孟祥武. 一种基于移动用户位置的网络推荐方法[J]. 软件导刊,2014,25(11):56-74.
- [4] MAFENGWO[EB/OL]. <http://www.mafengwo.cn/>
- [5] CHAKPABORTY B. Integrating awareness in user oriented route recommendation system[C]. The International Joint Conference on Neural Networks, New Jersey:IEEE Press,2012. 1-5.
- [6] 徐雅斌,孙晓晨. 位置社交网络的个性化位置推荐[J]. 北京邮电大学学报,2007,38(5):118-122.
- [7] FOURSQUARE[EB/OL]. <https://foursquare.com/>
- [8] PLACES. GOOGLE[EB/OL]. <http://places.google.com/rate>

(责任编辑:杜能钢)

作者简介:付宏杰(1973-),女,吉林公主岭人,博士,吉林工程技术师范学院信息工程学院副教授,研究方向为人工智能、群体智能、机器学习。