

文件编号: [TEAMNAME]-SWC2020-[TEAMNUMBER]

受控状态: ☒ 受控 ☐ 非受控

保密级别: ☐ 公司级 ☐ 部门级 ☒ 项目级 ☐ 普通级

采纳标准: CMMI DEV V1.2



迹意

TRACE MEANING

技术研究报告

Version 1.0.0

2019.11.16

Written by 新阳



All Rights Reserved

# 目录

|          |                           |          |
|----------|---------------------------|----------|
| <b>1</b> | <b>问题聚焦.....</b>          | <b>1</b> |
| 1.1      | 问题描述.....                 | 1        |
| 1.2      | 问题抽象.....                 | 1        |
| 1.3      | 问题定位.....                 | 1        |
| 1.4      | 问题评估.....                 | 1        |
| 1.5      | 问题分解.....                 | 2        |
| <b>2</b> | <b>相关工作.....</b>          | <b>3</b> |
| <b>3</b> | <b>技术方案.....</b>          | <b>4</b> |
| 3.1      | 技术方向.....                 | 4        |
| 3.2      | 技术选择.....                 | 4        |
| 3.3      | 结果期望.....                 | 4        |
| <b>4</b> | <b>技术实践.....</b>          | <b>5</b> |
| 4.1      | 使用的开发框架及依赖的 Library ..... | 5        |
| 4.2      | 技术实践过程.....               | 5        |
| <b>5</b> | <b>结果验证.....</b>          | <b>5</b> |

## 记录更改历史

[illegible]

# 1 问题聚焦

## 1.1 问题描述

旅游旺季,不少年轻人会选择出境游,尤其是自由行受越来越多人的青睐。自由行需要做攻略,但哪些地方是真正符合自己内心深处的那一隅呢。于是,此软件正是建立在这样一个前提下兴然而起。据了解,现在不少旅游网站及淘宝上都有出境游的地陪服务项目,以东南亚国家居多,提供这些服务的人有客栈老板、司机、兼职者或留学生等。但是他们的建议是否能真正满足自己想去的的地方?其中的风险也需注意。

## 1.2 问题抽象

### 1、在界面中显示信息

主要是数据库方面的技术问题。如果是作为本地程序使用,可以把 htm 存为 hta, 双击打开文件,就可使用本地数据库了。目前采用的云服务器是阿里云 CentOS7.3 版本的操作系统,云数据库 HBase 版(ApsaraDB for HBase)是基于 Hadoop 的一个分布式数据库,支持海量的 PB 级的大数据存储,适用于高吞吐的随机读写的场景。目前在阿里内部有数百个集群,10000 台左右规模的集群,服务数百个业务线,在订单存储、消息存储、物联网、轨迹、wifi、安全风控、搜索等领域有较多的在线应用。为了获取数据的方便与接下来进展的顺利,利用服务器搭建的服务器为 MYSQL。其中一个问题便是数据库的增删改查,利用 js 进行数据库的增删改查已经完毕。

### 2、地图显示

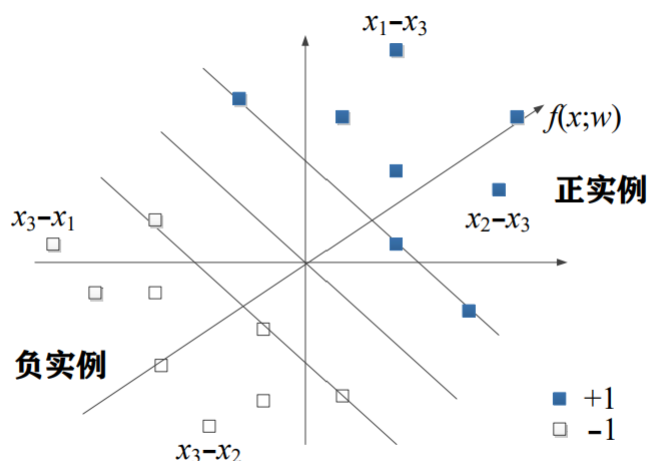
主要是 API 调用,调用高德地图(amap)的 web 服务 API,首先是进入网站: <http://lbs.amap.com/api/webservice/summary/>,然后申请 key,具体的 service 层:由于高德返回的是 json 格式的数据,我们接收数据是定义一个类来接收,这个类的属性与返回数据的 json 字段要相同。此时用到了用到了一个数据类型的转换器。得益于 Spring 的支持。

## 1.3 问题定位

基于排序学习的快应用平台推荐算法。

## 1.4 问题评估

技术性: 数据库的设计,以及增删改查,推荐算法的覆盖;



利用排序算法,对测试集进行排序获取物品的排序列表如图,该模型采用乘法隐含因素模型来描述用户-物品的二元效用函数,并定义排序模型为用户对已评分物品的打分与预测评分间的误差,并使其最小化,然后使用基于 MapReduce 的分布式随机梯度下降算法求解得到最优的排序模型。

普适性:多品种普适,其次是多周期普适。这样才初步算是开始。再看有没有过拟合的问题。很幸运的是,经过一段时间的摸索和实盘检验。目前日产品已经初步具备了这一点,在 tick 周期上可以做到多品种普适了。

热度:首先,有车族越来越多。随着生活水平的提升,普通消费者的收入水平相对以前有了很大的提升,这几年买车变得越来越容易,低首付、贴息买车等促销活动层出不穷,优惠力度之大,连原本从来没想过要买车的小白都心动不已了。这里利好信息直接的好处是让买车不再是一件难事了,越来越多的新车开始进入寻常百姓家庭。自驾游的自由度非常的高,自驾游旅游者通常不需要由旅行社安排旅游行程,在旅游目的地选择、到达与停留时间以及食宿安排上都有很大的自主性,区域性。对自驾车旅游者来说,旅游是一个学习与累积经验的过程。旅游的尝试首先从自驾车旅游者的居住周边开始,逐步向远辐射。我国旅游景点众多,自驾游爱好者的出行习惯会随其收入、教育程度、年龄、地域性、旅游目的等因素而有所不同。在旅游消费中,可能是时间、费用、品质与人群等多种因素的组合,不同的阶段不同的出行目的,完全灵活的按照自己的想法去制定合适的路线,和一起出行的人以及目的地。

## 1.5 问题分解

### (1) 位置信息定位服务;

能根据用户的位置生成准确的地理位置信息。

此项难度不大,调用地图 API 即可,放到最后。

### (2) 定时定点或当前位置时间地点语音留言发送、接收服务;

根据用户当前位置,定时向用户推送该地点的留言信息。

### (3) 当前条件下符合的留言推送;

可以对留言进行筛选,根据用户的喜爱偏好来推送留言信息。

需要 JS 连接数据库操作,从数据库中获取需要的信息。

### (4) 留言的编辑,包括删除、重新编辑、分享;

用户可以创建新的留言信息,也可以对已经发送的留言信息进行删除、修改、分享等操作。

作。

即是数据库的增删改查。

#### (5) 位置轨迹绘制。

用户每到达指定地点就会记录用户的位置，进而形成运动轨迹。

每天定点进行记录或者每隔多少分钟记录一次。

## 2 相关工作

### 基于排序学习的推荐算法

传统的推荐算法主要可以分为 3 大类:基于内容的推荐算法、协同过滤推荐算法以及混合推荐算法。这些传统推荐算法重点考虑用户和项目之间的二元关系,大都可以转化为评分预测问题,根据用户对项目的评分进行排序后产生推荐列表。近几年,研究人员发现:如果仅仅依据用户对项目的评分产生推荐结果并不能准确地体现用户的偏好。例如图 1 中,用户对物品 A 和物品 B 分别打分为 2 分和 3 分,那么使用不同的推荐算法进行预测将会得到的不同结果:一种预测结果为 A 物品 2.5 分和 B 物品 3.6 分;另一种预测结果为 A 物品 2.5 分和 B 物品 2.4 分。两种预测结果的平方误差均为  $(0.52 + 0.62)$ , 然而得到的物品 A 和物品 B 的排序却是相反的。由此可见,仅仅依赖于评分单一特征并不能非常准确地反映用户的偏好。为了解决传统推荐算法所存在的上述问题,研究人员考虑将排序学习技术融合进推荐算法的推荐过程之中,认为项目间的排序比传统推荐算法依据项目评分大小的顺序进行推荐更为重要。推荐算法是根据用户偏好需求模型来进行项目推荐,与用户偏好需求越匹配的项目则越倾向于推荐给该用户。因此,将排序学习融入推荐算法的主要思路是对用户的历史行为记录提取特征进行训练,学习得到项目的排序函数以最终对用户生成项目推荐列表。基于排序学习的推荐算法和传统的推荐算法有着本质的区别,其本质区别在于:传统的推荐算法如基于内容的推荐算法、协同过滤推荐算法等不需要训练阶段,直接通过计算用户之间的相似度和项目之间的相似度,预测用户对项目的兴趣度,以此来排序产生推荐结果;而基于排序学习的推荐算法结合了机器学习的特点,是一种监督性学习,需要通过训练数据集训练得到排序模型,并且调整排序模型的参数得到排序模型的最优解,然后对测试数据集使用该排序模型产生最终的推荐结果。近几年,基于排序学习的推荐算法得到越来越多的关注,目前已经成为推荐系统领域的研究热点之一。国外许多大学和研究机构对基于排序学习的推荐算法展开了深入研究。

为了解决传统推荐算法所存在的上述问题,研究人员考虑将排序学习技术融合进推荐算法的推荐过程之中,认为项目间的排序比传统推荐算法依据项目评分大小的顺序进行推荐更为重要。推荐算法是根据用户偏好需求模型来进行项目推荐,与用户偏好需求越匹配的项目则越倾向于推荐给该用户。因此,将排序学习融入推荐算法的主要思路是对用户的历史行为记录提取特征进行训练,学习得到项目的排序函数以最终对用户生成项目推荐列表。基于排序学习的推荐算法和传统的推荐算法有着本质的区别,其本质区别在于:传统的推荐算法如基于内容的推荐算法、协同过滤推荐算法等不需要训练阶段,直接通过计算用户之间的相似度和项目之间的相似度,预测用户对项目的兴趣度,以此来排序产生推荐

结果;而基于排序学习的推荐算法结合了机器学习的特点,是一种监督性学习,需要通过训练数据集训练得到排序模型,并且调整排序模型的参数得到排序模型的最优解,然后对测试数据集使用该排序模型产生最终的推荐结果.近几年,基于排序学习的推荐算法得到越来越多的关注,目前已经成为推荐系统领域的研究热点之一.国外许多大学和研究机构对基于排序学习的推荐算法展开了深入研究

## 3 技术方案

### 3.1 技术方向

基于排序学习的推荐算法。

### 3.2 技术选择

近年来,移动推荐系统随着移动互联网的蓬勃发展也越发活跃起来.移动推荐系统中用户处于移动网络环境,移动性强,而且移动设备的处理能力差、屏幕小、输入能力较差、无线网络的带宽弱等因素使其对实时性和准确性的要求较高,从而原先适合传统互联网用户的推荐方法并不能直接应用到移动推荐中.

从用户的 GPS 历史数据中提取信息构建“用户-地理位置-活动”三维评分矩阵,并提出了 3 种推荐方法:第 1 种方法是基于矩阵的协同过滤模型,目标在于对用户的“地理位置-活动”二维矩阵进行因式分解,使得平方误差损失函数最小化;第 2 种对“用户-地理位置-活动”三维评分矩阵进行张量分解,同样使得平方误差损失函数最小化;第 3 种方法同样是对“用户-地理位置-活动”三维评分矩阵进行张量分解,但直接优化排序评价准则以学习最优排序函数.Shaw 等人使用签到数据对用户和地点建立空间概率模型和时间概率模型,空间概率模型表示在特定的地理位置坐标签到的概率,时间概率模型则表示以时间范围划分在地点签到的概率,然后采用线性回归构造损失函数,并采用坐标上升算法训练损失函数中的权重参数以优化损失函数.另外,作者利用 LambdaMART 排序学习算法构建决策树,非线性地训练输入的用户签到的时间和空间特征.Zhuang 等人讨论了用户使用移动设备时,基于上下文感知需要某种实体类别或者具体实体的概率模型,并通过随机游走算法来更新优化该概率模型,最终对用户生成一个实体排序列表.把融合多种上下文信息进行个性化推荐的问题建模为排序学习的问题,建立用户偏好模型、空间偏好模型以及时间偏好模型,利用基于地理建模内嵌的矩阵分解模型进行地理位置推荐和预测.

基于排序学习的推荐算法是一个新兴的研究领域,它仍有许多值得深入探索和亟待解决的问题.是我们今后需要解决完善的研究方向.

### 3.3 结果期望

近些年来,随着物联网、云计算和社交网络等技术的迅猛发展,大数据越发显现 4V(volume, velocity, variety, veracity)的特性.机器学习通过训练数据来训练模型,数据样本越大,

机器学习就可以训练出越复杂的模型,进而做出更准确的判别与预测.但是,大数据环境同时给机器学习造成了计算上的困难,主要表现为 CPU 计算量和内存存储量的限制.通常,我们可以采用并行计算、集成学习、近似估计这 3 种方法.并行计算是指可以采用 MapReduce 框架进行分布式计算,目前已有部分推荐算法采用分布式计算模式,但是仍然有很大的研究和应用空间.集成学习则是每次抽取少量样本数据并对其进行建模,多次迭代获得多个模型进行集成,得到最终的模型.Ueda 等人验证了这种方法可以有效处理大规模数据,并训练得到合成模型.近似估计是指提出一个近似算法,其复杂度低于原有算法的复杂度而不以准确度为代价都是在损失函数中引入指数函数等具有平滑性质的连续函数,然后根据不等式性质找到损失函数的边界作为损失函数的近似函数进行优化更新.该方法可以降低原有损失函数的复杂度而不损失其准确度.从这一点说明,技术是可行的。

## 4 技术实践

### 4.1 使用的开发框架及依赖的 Library

### 4.2 技术实践过程

## 5 结果验证