

文章编号: 1000-5641(2017)05-0154-08

## 基于复杂属性商品的混合协同过滤推荐模型

周兰凤, 麻双克, 付 正, 张 晴

(上海应用技术大学 计算机科学与信息工程学院, 上海 201418)

**摘要:** 协同过滤作为应用最广、研究最多的推荐算法, 但依旧面临数据稀疏性、冷启动、数据质量差等固有问题, 同时也鲜有研究者从实用角度基于商品性价比方面提高预测精确度. 为此, 本文综合考虑用户主观评分和商品客观评分, 并在此基础上结合情境预过滤、社会网络理论以及专家意见提出了一种混合协同过滤推荐模型, 在一定程度上缓解了上述缺点. 并通过真实网上汽车销售数据实验, 表明该模型相对传统协同过滤具有更高的预测精度, 更适用于具有复杂属性的商品.

**关键词:** 协同过滤; 情境; 复杂属性; 个性化推荐

**中图分类号:** TP399 **文献标志码:** A **DOI:** 10.3969/j.issn.1000-5641.2017.05.014

## A hybrid collaborative filtering recommendation model based on complex attribute of goods

ZHOU Lan-feng, MA Shuang-ke, FU Zheng, ZHANG Qing

(School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

**Abstract:** Collaborative filtering as the most widely used, the most recommendation algorithm, the shortcomings inherent in the data sparse, cold start poor data quality and others, and few studies based on commodity price to improve the prediction accuracy. At the same time, facing the full e-commerce market network Navy, the ratings and reviews also indirectly led to the predict a decline in accuracy. Therefore, this paper comprehensive consideration of the user subjective ratings and objective product score, and on this basis, combined with situation pre filtering, social network theory and expert opinions put forward a hybrid collaborative filtering recommendation model, to some extent alleviate the above shortcomings. And through experiment with real online car sales data, the model has higher forecast accuracy than the traditional collaborative filtering, and is more suitable for the commodity with complex attributes.

**Key words:** collaborative filtering; context; complex attribute; personalized recommendation

收稿日期: 2017-06-19

基金项目: 国家自然科学基金(41671402)

第一作者: 周兰凤, 女, 副教授, 研究方向为人工智能及大数据. E-mail: lfzhou@sit.edu.cn.

通信作者: 张 晴, 女, 讲师, 研究方向为图像处理. E-mail: zhangqing0329@gmail.com.

## 0 引言

电子商务的发展,使得协同过滤推荐迅速应用于信息过滤和信息推荐系统中,并成为该领域的研究热点,主要根据用户的相似性来产生推荐<sup>[1]</sup>.随着电子商务服务范围的扩大,项目评分的维度急剧增大且用户历史数据呈极端稀疏性.基于此,研究者提出了大量的解决方案,像基于关联规则、神经网络<sup>[2]</sup>、聚类<sup>[3]</sup>、贝叶斯网络<sup>[4]</sup>、时序<sup>[5]</sup>基于用户相似度的推荐;基于标签<sup>[6-7]</sup>、分类<sup>[8]</sup>,但是,这些研究均基于用户之间的相似度作推荐,没有考虑商品性价比问题,同时,也没有考虑用户恶意评分给推荐准确度带来的影响.因此,如何排除用户恶意评分影响,尽量向目标用户推荐符合其兴趣的高性价比商品成为电子商务市场亟待解决的问题之一.

基于此,本文提出一种基于复杂属性商品的混合协同过滤推荐模型,该模型通过情景分类解决用户数据的极度稀疏性,综合考虑用户对商品的属性评分和商品属性客观评分,在尽量体现用户偏好的基础上克服网络水军的恶意评分,同时模型通过社会网络理论和专家意见来克服传统计算用户相似度不能向用户推荐高性价比商品的弊端.

## 1 传统协同过滤推荐算法

### 1.1 传统协同过滤推荐算法描述

协同过滤推荐是目前应用比较广,研究比较多的算法之一,也被称为社会过滤,其基本思想认为:具有相似兴趣、爱好的用户,也具有相似的信息、商品需求.主要分为基于用户和基于项目两类.前者认为若两个用户的爱好相似,则其目标评分也应该相似,后者认为用户过去喜欢的商品和现在喜欢的商品类似.

因此,其通过计算用户或商品评分之间的相似性,寻找最近邻居,然后综合邻居评分得到预测评分,进而为目标用户提供推荐,其推荐的个性化程度比较高.也正是因为如此,协同过滤从用户浏览、购买、评分等历史信息中分析用户的偏好,不需要用户的直接输入,减少了用户的负担.

其最近邻的选择主要有余弦相似度和修正的余弦相似度两类.设评分项目集合由  $n$  个项目组成,项目  $c$  是其中一个,  $R_{i,c}$  表示用户  $i$  对评分项目集合中的项目  $c$  的评分,  $I$ 、 $J$  分别表示第  $I$  个和第  $J$  个项目,  $\overline{R}_i$  表示用户  $i$  的所有评分的均值,则两种最近邻的选择公式为

$$\text{Sim}(i, j) = \cos(I, J) = \frac{I \times J}{\|I\| \times \|J\|} = \frac{\sum_{c=1}^n R_{i,c} R_{j,c}}{\sqrt{\sum_{c=1}^n R_{i,c}^2} \sqrt{\sum_{c=1}^n R_{j,c}^2}}, \quad (1)$$

$$\text{Sim}(i, j) = \frac{\sum_{c=1}^n (R_{i,c} - \overline{R}_i)(R_{j,c} - \overline{R}_j)}{\sqrt{\sum_{c=1}^n (R_{i,c} - \overline{R}_i)^2} \sqrt{\sum_{c=1}^n (R_{j,c} - \overline{R}_j)^2}}. \quad (2)$$

Lemire 和 Maclachlan 于 2005 年提出了一种新的算法<sup>[9]</sup>——SlopeOne, 该算法的优点是计算复杂度低、精确率比较高,不再基于相似性做预测,而是利用线性回归模型做预测.该

算法利用均值来代替未知项目之间的差异

$$\text{dev}_{i,j} = \sum_{u \in U_{i,j}} \frac{u_i - u_j}{|U_{i,j}|}, \quad (3)$$

其中,  $U_{i,j}$  表示同时对项目  $I_i$  和项目  $I_j$  评分的用户集合,  $u_i$  表示用户  $u$  对项目  $I_i$  的评分,  $u_j$  表示用户  $u$  对项目  $I_j$  的评分,  $|U_{i,j}|$  表示同时对项目  $I_i$  和项目  $I_j$  评分的用户个数.  $\text{dev}_{i,j}$  表示项目  $I_i$  和项目  $I_j$  之间的平均偏差.

## 1.2 传统协同过滤推荐存在的问题

在现实中, 商品属性繁多的商品往往是较为贵重的商品, 因此用户在购买之前往往会交流彼此的心得体会以及购物经验, 这种经验交流往往会对目标用户商品的选择, 产生直接或间接的影响<sup>[10]</sup>, 然而传统的协同过滤往往是过于强调用户间兴趣的相似性, 而忽视了用户特征以及商品属性对用户选择的影响.

为解决该问题, Yuan 等<sup>[11]</sup> 和 Ding 等<sup>[12]</sup> 认为社会网络可以用来描述用户间的交互行为; 邓晓懿等<sup>[13]</sup> 提出了基于情境聚类 and 用户评级的协同过滤推荐模型, 通过引入情境因素和用户评级有效地解决了数据稀疏性和评分数据高维化的问题; Ma 等<sup>[14]</sup> 提出了一种结合项目评分和用户评分的混合推荐算法来缓解矩阵的稀疏性.

尽管这些研究其结果性能均优于传统的协同过滤算法. 但是, 由于基于用户的协同过滤算法仅仅是根据用户相似度进行推荐的, 就比较容易产生“特殊化”的问题, 也就是说, 仅仅是考虑了用户间的关系, 没有考虑用户评分的准确性, 即没有考虑用户是不是随意写的评分, 或者是不是“网络水军”.

## 2 推荐模型的建立

针对以上问题, 本文在对情境因素进行分类的基础上, 提出了一个基于用户评级的协同过滤推荐算法, 模型基本框架如图 1 所示. 模型首先利用用户注册获得用户的爱好和部分情境信息, 构建用户爱好矩阵, 以解决协同过滤推荐算法所面临的冷启动问题; 其次利用情境信息和传统分类对用户爱好矩阵进行分类, 以减小计算量; 再次, 引入社会网络和建立评级模型以解决稀疏性问题并分析用户间的爱好关系, 找到相似性最高的  $N$  个用户及其评价的  $K$  个商品; 最后利用“专家意见”对  $K$  个商品进行筛选, 选出性能最好的商品. 这里的“爱好”或者“偏好”是指用户对已购买商品或者理想商品各个属性的评分.

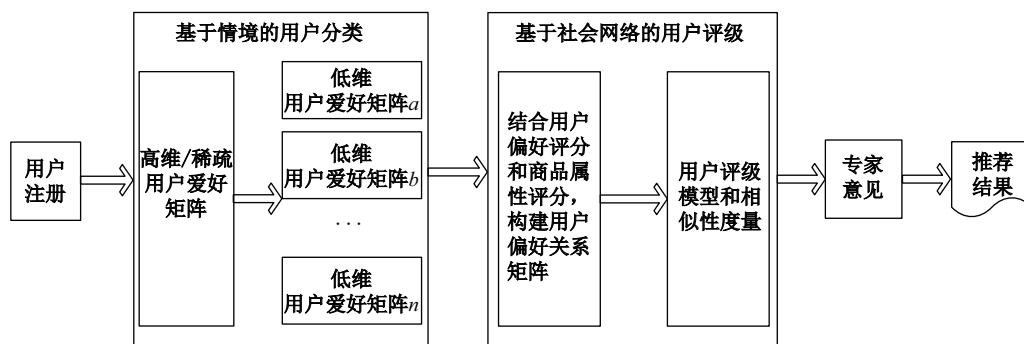


图1 模型基本框架

Fig. 1 The basic framework of the model

### 2.1 情境分类

本文将情境因素分为两种: 用户情境和环境情境. 用户情境包括用户的年龄、职业和性别, 用三元组表示, 如公式(4)所示, 其中, 年龄集合分为6个年龄段, 即20岁以下、21-25岁、26-30岁、30-40岁、40-50岁、50岁以上; 性别集合由0和1构成, 分别代表男性和女性; 职业分为20种不同类型, 如工人、推销员、分析师、学生等等. 具体公式为

$$\begin{cases} E_u = \langle \text{Age}, \text{Occupation}, \text{Sex} \rangle, & (4a) \\ \text{Age} \in (A = \{A_i | i = 1, 2, \dots, 6\}), & (4b) \\ \text{Gender} \in (G = \{0, 1\}), & (4c) \\ \text{Occupation} \in (O = \{O_i | i = 1, 2, 3, \dots, 20\}). & (4d) \end{cases}$$

环境情境则包括价格、目的、地点、时间等4种属性, 如公式(5)所示, 具体类别则需根据实际数据进行量化. 最后, 根据情境因素将用户划分为不同类别, 以使相同类别内用户具有较高相似度. 具体公式为

$$\begin{cases} E_c = \langle \text{Price}, \text{Des}, \text{Address}, \text{Time} \rangle, & (5a) \\ \text{Price} \in (P = \{P_i | i = 1, 2, \dots, 8\}), & (5b) \\ \text{Des} \in (D = \{D_i | i = 1, 2, \dots, 10\}), & (5c) \\ \text{Address} \in (G = \{1, 2, 3, \dots, 27\}), & (5d) \\ \text{Time} \in (T = \{T_i | i = 0, 1, 2, \dots, 5\}). & (5e) \end{cases}$$

## 2.2 基于社会网络的用户评级

首先, 根据用户评分数据和该商品客观评分数据构建用户、商品矩阵, 用以描述用户、商品之间的关系, 并作为数据预处理的基础矩阵; 其中, 客观评分来自于所有购买客户的综合评价所得, 能够比较真实地反映商品的各个属性.

设用户集合为  $U = \{u_i | i = 1, 2, \dots, m\}$ , 用户对某一商品的各个属性的主观评分集合为  $I = \{ui_j | j = 1, 2, \dots, n\}$ , 商品各个属性客观评分集合为  $C = \{ci_j | j = 1, 2, \dots, n\}$ , 前8列代表用户对已购买商品8个属性的评分, 后8列代表该品牌商品8个属性的客观评分, 如表1所示.

其次, 通过一个权值  $\alpha$  来控制用户对商品各个属性的主观评分和汽车各个属性的客观评分的重要程度, 最终用户  $u_i$  对商品各个属性评分  $i_j$  的计算公式为

$$i_j = \alpha \times ui_j + (1 - \alpha) \times ci_j, \quad j = 1, 2, \dots, 8. \quad (6)$$

在实际购买商品时, 可能某一个单纯的偏好, 就能使用户决心购买某件产品, 其他的偏好作用并不大, 本文暂时称之为“孤立偏好”. 为此, 根据目标用户偏好评分对偏好进行排序, 排序后, 删除偏好顺序和目标用户不一致的用户, 排序后为矩阵  $U_1$ , 如表2所示. 然后, 为表示用户对该项评分是否满意, 将各个用户的偏好评分减去评分均值, 大于等于0的作为1, 代表用户对此偏好程度严重, 小于0的作为0, 表示偏好程度一般. 矩阵  $U_2$  如表3所示, 其中当  $i \neq j$  时,  $u_{ij} \neq 0$  表示用户  $u_i$  和用户  $u_j$  同时具有某些偏好,  $u_{ij} = 0$  表示用户  $u_i$  和用户  $u_j$  没有共同的偏好; 当  $i = j$  时, 默认  $u_{ij} = 0$ , 即不计算单个用户偏好的数量. 最后, 为表现用户之间拥有共同偏好的数量, 根据公式(7)建立用户偏好关系矩阵, 如表4所示, 其中,  $u_{13} = u_{31} = 3$  表示用户  $u_1$  和用户  $u_3$  拥有相同偏好3项( $i_2, i_6, i_7$ ), 但是用户  $u_1$  具有偏好5

项( $i_2, i_6, i_7, i_5, i_8$ ), 用户  $u_3$  具有偏好 3 项( $i_2, i_6, i_7$ ), 两者拥有的偏好数目不同; 相应地, 用户的等级应该也有所不同. 因此, 矩阵  $U_4$  仍然无法较好地反映用户偏好之间的真实关系, 需要进一步对矩阵  $U_4$  做相应处理, 使其不仅能反映出用户具有相同偏好的数目, 还能反映出用户偏好之间的等级差异.

$$U_4 = (u_{ij})_{3 \times 3} = U_3 \cdot U'_3, \quad u_{ij} = \begin{cases} u_{ji}, & \text{如果 } u_j \text{ 和 } u_i \text{ 有相同的偏好,} \\ 0, & \text{如果 } u_j \text{ 和 } u_i \text{ 没有相同的偏好.} \end{cases} \quad (7)$$

因此, 本文在矩阵  $U_4$  的基础上做了标准化处理: 在  $U_4$  中,  $u_{ij}$  表示被用户  $u_i$  和用户  $u_j$  所拥有的相同偏好的数量, 用户  $u_i$  的偏好数为  $\sum_{j=1}^m u_{ij}$ ,  $m$  为用户总数, 那么标准化用户偏好关系矩阵  $U_5$  可定义为

$$U_5 = (w_{ij})_{m \times m}, \quad w_{ij} = u_{ij} / \sum_{j=1}^m u_{ij}, \quad (8)$$

这里

$$\sum_{j=1}^m w_{ij} = 1, \quad j = 1, 2, \dots, m. \quad (9)$$

表 1 用户、商品矩阵

Tab. 1 User-item matrix

	$ui_1$	$ui_2$	$ui_3$	$ui_4$	$ui_5$	$ui_6$	$ui_7$	$ui_8$	$ci_1$	$ci_2$	$ci_3$	$ci_4$	$ci_5$	$ci_6$	$ci_7$	$ci_8$
$u_1$	1	5	3	3	4	5	5	4	1	5	3	3	4	5	5	4
$u_2$	1	4	2	1	3	4	4	3	1	3	2	1	4	4	3	3
$u_3$	2	5	3	3	4	5	5	4	1	4	2	2	2	3	3	1
$u_4$	4	3	5	4	2	4	3	1	2	3	4	3	1	2	2	3
$u_5$	5	5	5	5	5	5	5	5	1	4	5	4	5	1	3	4

表 2 用户、偏好客观评分排序矩阵

Tab. 2 User, preference objective scoring matrix

	$i_2$	$i_6$	$i_7$	$i_5$	$i_8$	$i_3$	$i_4$	$i_1$
$u_1$	5	5	5	4	4	3	3	1
$u_2$	4	3.6	3.6	3.4	3	2	1	1
$u_3$	4.6	4.2	4.2	3.2	2.8	2.6	2.6	1.6

表 3 用户偏好关系矩阵

Tab. 3 User, preference objective scoring presort matrix

	$i_2$	$i_6$	$i_7$	$i_5$	$i_8$	$i_3$	$i_4$	$i_1$
$u_1$	1	1	1	1	1	0	0	0
$u_2$	1	1	1	1	1	0	0	0
$u_3$	1	1	1	0	0	0	0	0

表 4 用户偏好关系矩阵

Tab. 4 User, preferencerelation matrix

	$u_1$	$u_2$	$u_3$
$u_1$	0	5	3
$u_2$	5	0	3
$u_3$	3	3	0

如表5所示, 可以发现  $u_{13} = 3/8$ ,  $u_{31} = 3/6$ , 显然  $u_{13} \neq u_{31}$ , 其中, 分子表示用户之间拥有共同偏好的数量, 分母表示不同用户偏好之间的差异.

表5 标准化用户偏好关系矩阵

Tab. 5 Standarded user, preference relation matrix

	$u_1$	$u_2$	$u_3$
$u_1$	0	5/8	3/8
$u_2$	5/8	0	3/8
$u_3$	3/6	3/6	0

在得到标准化偏好用户关系矩阵  $U_5$  后, 为了对传统的相似性度量方法进行改进, 本文使用社会网络理论, 通过建立用户偏好评级模型来分析用户间的关系.

由于用户偏好关系图  $G$  可用于描述各个用户偏好之间的关系<sup>[15]</sup>, 在矩阵  $U_5$  基础上生成一个具有权重的有向图  $G = (U, L)$ , 如图2所示, 其中  $U = \{u_i | i = 1, 2, \dots, m\}$  代表用户偏好, 为所有节点的集合;  $L = \{I_i | i = 1, 2, \dots, n\}$  代表用户之间的偏好关系, 为所有有向边的集合.

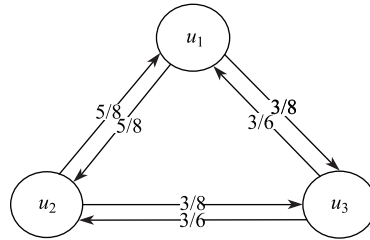


图2 用户偏好关系图

Fig. 2 The figure of user preference relation

社会网络理论认为, 不同用户彼此相互影响, 用户偏好具有一定的传递性<sup>[16]</sup>, 并且与PageRank模型中值的传递一样具有相同的衰减性<sup>[15]</sup>. 根据Arasu等对PageRank的定义<sup>[15]</sup>, 定义公式

$$UR(u_i) = d \times \sum_{u_j \in B(u_i)} UR(u_j) / \text{Outweight}(u_j) + (1 - d) / m. \quad (10)$$

设  $u_i$  和  $u_j$  分别为两个节点;  $\text{Outweight}(u_j)$  表示从节点  $u_j$  向外的链接数目;  $B(u_i)$  表示连接到节点  $u_i$  的节点集合;  $UR(u_i)$  表示节点  $u_i$  的PageRank值;  $d$  是一个衰减系数, 大于0且小于1的值, 一般为0.8左右, 表示这个节点PageRank值的80%分配到所连接的节点中, 另外的20%被分配到所有的节点中;  $m$  表示节点的总数.

因为所有节点的PageRank的初始值都是未知的, 所以我们以均值作为其初始值, 即  $1/m$ , 利用上述公式进行计算, 然后将得到的值再次带入进行计算, 如此反复, 直到值最终收敛于一个比较固定的数, 即用户偏好级别.

### 2.3 专家意见

在上一节, 已经求出目标用户偏好的最近邻集合, 根据传统的协同过滤推荐算法, 需要利用相似度最高的用户的评分为目标用户为评分项目预测评分. 由于数据的稀疏性, 本算法不再对其进行预评分, 而是利用公式(3)求出用户之间的相似性, 继而选择相似度最高的前  $N$  个商品通过预定好的意见做进一步处理, 代替预评分; 这里的意见由人工指定, 目的在于寻找具有最高性价比产品.

## 3 实验结果及分析

### 3.1 实验设计

本文采用网上车市的网友评价及参数配置数据,数据集通过自制爬虫程序获得,包括 13 个级别及每个级别关注度排名前 10 的汽车品牌参数信息,以及 6 320 条网友信息数据. 网友信息包含两部分: 情境信息和用户评分信息. 情境信息包括购买车系、裸车购买价、购买时间、购买地点、购车目的,由于数据总量较少,本实验仅采用价格和购车目的作为情境因素. 用户评分信息和汽车客观评分信息一样,包括空间、动力、操作、油耗、舒适性、外观、内饰、性价比,用户评分分值为 1-5,分值越高,表示用户越满意.

专家意见主要考虑发动机性能和整车性能,发动机性能暂定为,排量相同情况下,油耗越低越好;整车质量根据厂家给的保修规则来判定,保修年限或里程越长,则认为其质量越可靠.

### 3.2 评判指标

算法性能从两方面进行检验,一方面将基于准确率进行评价,因为只有分类准确度可用于分类问题中;另一方面利用传统的平均绝对误差(Mean Absolute Error, MAE)来衡量用户预测评分和真实评分之间的差异.

分类准确度适用于 0-1 系统,即用户对该项目喜欢还是不喜欢、推荐结果是否正确,将预测过程假想为用户做决策的过程. 本文使用精确度(Precision)衡量,公式为

$$\text{Precision} = \frac{\sum_u R_u \cap T_u}{\sum_u |R_u|}, \quad (11)$$

其中,  $R_u$  是推荐结果集,  $T_u$  是真实结果集,结果集指的是汽车级别、车厢结构、驱动方式、变速箱形式等 4 种.

评分绝对误差用 MAE 计算,公式为

$$M_{\text{MAE}} = \frac{\sum_{i=1}^{\text{num}} \sum_{j=1}^k |p_{ij} - q_{ij}|}{\text{num} \times k}, \quad (12)$$

其中,  $p_{ij}$  表示第  $i$  个推荐车型第  $j$  个属性的预测评分,  $q_{ij}$  表示第  $i$  个用户对历史购买车型的  $j$  个属性的客观评分, num 表示测试集大小,  $k$  表示用户属性个数.

### 3.3 实验结果

作为对比,本文选取常用的基于用户的协同过滤推荐,即利用目标用户注册的偏好信息和其余用户偏好评分信息,一起构建用户项目评分矩阵,找目标用户最相似的车型,然后根据最终推荐车型信息和评分,将其最终结果和本算法推荐的结果分别计算 Precision 和 MAE. 如图 3、图 4 所示,结果通过计算 10 名目标用户的均值得出,横坐标表示用户数量的多少,竖坐标分别表示 Precision 和 MAE.

图 3 中,可以看出,随着用户偏好评分数量的增多,两种推荐算法精确度都在逐渐上升,但是基于汽车销售的个性化推荐,算法始终表现出明显优势. 图 4 中,也可看出,基于汽车销售的个性化推荐,算法在评分上相比传统的协同过滤也要精确很多,具有一定的优势.

## 4 结束语

针对传统协同过滤算法应用于复杂属性商品推荐存在的问题,提出一种基于复杂属性商品的混合协同过滤推荐模型,通过结合用户主观评分和商品客观评分,在尽量反映用户偏好的基础上得出一个混合评分,并以此混合评分作为依据,进行数据处理;在提高推荐准确度方面,通过专家意见提高商品性价比,以此来提高用户的满意度. 实验结果表明,与传统协同过滤推荐相比,该模型在平均绝对误差(MAE)和精确度(Precision)均具有较高优势. 因此,本模型相对传统协同

过滤有效地提高了预测精度, 更适合于复杂商品属性推荐. 未来的研究工作将结合社交网络进一步提高预测精度.

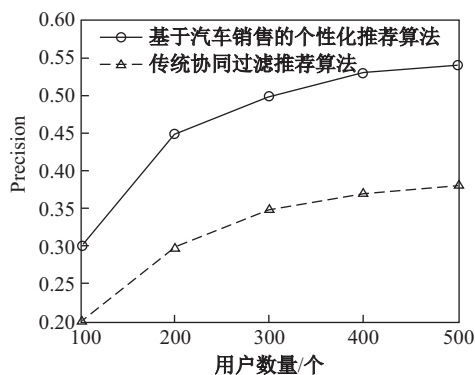


图 3 精确度

Fig. 3 Precision

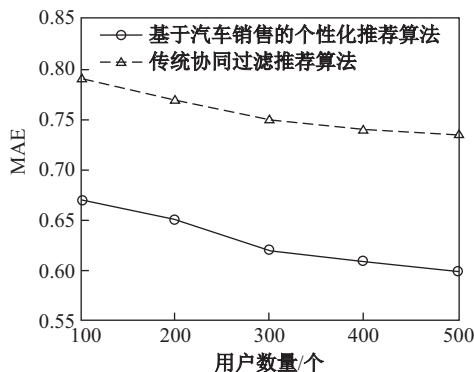


图 4 平均绝对误差

Fig. 4 Mean absolute error

### [参 考 文 献]

- [1] 许海玲, 吴潇, 李晓东. 互联网推荐系统比较研究[J]. 软件学报, 2009, 20(2): 350-362.
- [2] 张锋, 常会友. 使用 BP 神经网络缓解协同过滤推荐算法的稀疏性问题[J]. 计算机研究与发展, 2006, 43(4): 667-672.
- [3] 王辉, 高利军, 王听忠. 个性化服务中基于用户聚类的协同过滤推荐[J]. 计算机应用, 2007, 27(5): 1225-1227.
- [4] ZIEGLER C N, MCNEE S, KONSTAN J, et al. Improving recommendation lists through topic diversification[C]//Proceedings of the 14th International World Wide Web Conference. 2005: 22-32.
- [5] 陈曦, 成韵姿. 一种优化组合相似度的协同过滤推荐算法[J]. 计算机工程与科学, 2017, 39(1): 180-187.
- [6] 郭彩云, 王会进. 改进的基于标签的协同过滤算法[J]. 计算机工程与应用, 2016, 52(8): 56-61.
- [7] 宋伟伟, 杨德刚, 郑敏. 基于时间加权标签的协同过滤推荐算法研究[J]. 重庆师范大学学报(自然科学版), 2016, 33(5): 113-120.
- [8] 孙楠军, 刘天时. 基于项目分类和用户群体兴趣的协同过滤算法[J]. 计算机工程与应用, 2015, 51(10): 128-131.
- [9] LEMIRE D, MACLACHLAN A. Slope one predictors for online rating-based collaborative filtering[C]// Proceedings of the SIAM Data Mining (SDM'05). 2005: 21-23.
- [10] LIU F, LEE H J. Use of social network information to enhance collaborative filtering performance[J]. Expert Systems with Applications, 2010, 37(7): 4772-4778.

(下转第 185 页)



- [9] ATZORI M, ATZORI M, SAYGIN Y. Towards trajectory anonymization: A generalization-based approach[C]// Sigspatial ACM Gis 2008 International Workshop on Security and Privacy in Gis and Lbs. New York: ACM, 2008: 52-61.
- [10] SWEENEY L. K- anonymity: A model for protecting privacy[J]. International Journal on Uncertainty, Fuzziness and Knowledge- Based Systems, 2002, 10(5): 557-570.
- [11] MACHANAVAJJHALA A, KIFER D, GEHRKE J. L -diversity: Privacy beyond k -anonymity[J]. Acm Transactions on Knowledge Discovery from Data, 2007, 1(1): 3.
- [12] LI N H, LI T C, VENKATASUBRAMANIAN S. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity[C]// IEEE, International Conference on Data Engineering. [S.l.]: IEEE, 2007: 106-115.
- [13] MAO J, SONG Q, JIN C, et al. TScluWin: Trajectory Stream Clustering over Sliding Window[M]// Database Systems for Advanced Applications. US: Springer, 2016.
- [14] ZHANG Z, WANG Y, MAO J, et al. DT-KST: Distributed top-k similarity query on big trajectory streams[J]. 2017: 199-214.
- [15] WU W, XIAO Y, WANG W, et al. k-symmetry model for identity anonymization in social networks[C]// EDBT 2010, International Conference on Extending Database Technology. Switzerland: DBLP, 2010: 111-122.
- [16] DWORK C. Differential privacy[J]. Lecture Notes in Computer Science, 2006, 4052(2): 1-12.
- [17] KELLARIS G, PAPADOPOULOS S, XIAO X, et al. Differentially private event sequences over infinite streams[J]. Proceedings of the Vldb Endowment, 2014, 7(12): 1155-1166.
- [18] CHEN R, FUNG B C M, DESAI B C, et al. Differentially private transit data publication: a case study on the montreal transportation system[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2012: 213-221.
- [19] CAO Y, YOSHIKAWA M. Differentially private real-time data release over infinite trajectory streams[C]// IEEE International Conference on Mobile Data Management. [S.l.]: IEEE, 2015: 68-73.
- [20] MIGUEL E ANDRÉS, NICOLAS E BORDENABE, LONSTANTINOS Chatzikokolakis, et al. Geo-indistinguishability: Differential privacy for location-based systems[C]// Proceedings of the 2013 ACM SIGSAC Conference on Computer and Communications security. New York: ACM, 2013: 901-914.

(责任编辑: 张 晶)

(上接第161页)

- [11] YUAN W, GUAN D, LEE Y, et al. Improved trust-aware recommender system using small worldness of trust networks[J]. Knowledge Based Systems, 2010, 23(3): 232-238.
- [12] DING L, STEIL D, DIXON B, et al. A relation context oriented approach to identify strong ties in social networks[J]. Knowledge-Based Systems, 2011, 24(8): 1187-1195.
- [13] 邓晓懿, 金淳, 韩庆平, 等. 基于情境聚类 and 用户评级的协同过滤推荐模型[J]. 系统工程理论与实践, 2013, 33(11): 2945-2953.
- [14] MA H, KING I, LYU M R. Effective missing data prediction for collaborative filtering[C]// Sigir Proceedings of Annual International ACM Sigir Conference on Research & Development 2007: 39-46.
- [15] LI Y M, LAI C Y, CHEN C W. Discovering influencers for marketing in the blogosphere[J]. Information Sciences, 2011, 181(23): 5143-5157.
- [16] YUAN W W, GUAN D H, LEE Y K, et al. Improved trust-aware recommender system using small-worldness of trust networks[J]. Knowledge Based Systems, 2010, 23(3): 232-238.
- [17] ARASU A, CHO J, GARCIA-MOLINA H, et al. Searching the Web[J]. ACM Transactions on Internet Technology, 2001(1): 2-43.

(责任编辑: 李 艺)