

## 基于上下文相似度和社交网络的 移动服务推荐方法

俞春花<sup>1</sup>, 刘学军<sup>1,2</sup>, 李 斌<sup>1</sup>, 章 玮<sup>3</sup>

(1. 南京工业大学计算机科学与技术学院, 江苏南京 211816; 2. 复旦大学, 上海 200433;  
3. 中国人民解放军 73677 部队, 江苏南京 210016)

**摘 要:** 针对传统的基于协同过滤的移动服务推荐方法存在的数据稀疏性和用户冷启动问题, 提出一种基于上下文相似度和社交网络的移动服务推荐方法 (Context-similarity and Social-network based Mobile Service Recommendation, CSMSR). 该方法将基于用户的上下文相似性引入个性化服务推荐过程, 并挖掘由移动用户虚拟交互构成的社会关系网络, 按照信任度选取信任用户; 然后结合基于用户评分相似性计算发现的近邻, 分别从相似用户和信任用户中选择相应的邻居用户, 对目标用户进行偏好预测和推荐. 实验表明, 与已有的服务推荐方法 TNCF、SRMTC 及 CF-DNC 相比, CSMSR 方法有效地缓解数据稀疏性并提高推荐准确率, 有利于发现用户感兴趣的服务, 提升用户个性化服务体验.

**关键词:** 移动服务推荐; 上下文; 相似性计算; 社交网络; 协同过滤; 稀疏性; 冷启动问题

**中图分类号:** TP301 **文献标识码:** A **文章编号:** 0372-2112 (2017)06-1530-07

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.06.036

## Mobile Service Recommendation Based on Context Similarity and Social Network

YU Chun-hua<sup>1</sup>, LIU Xue-jun<sup>1,2</sup>, LI Bin<sup>1</sup>, ZHANG Wei<sup>3</sup>

(1. College of Computer Science and Technology, Nanjing Tech University, Nanjing, Jiangsu 211816, China;  
2. Fudan University, Shanghai 200433, China; 3. 73677 PLA Troops, Nanjing, Jiangsu 210016, China)

**Abstract:** Concentrating on the data sparsity problem and the new user cold-start problem faced by traditional collaborative filtering algorithm in mobile recommender system, an approach named CSMSR (Context-similarity and Social-network based Mobile Service Recommendation) is proposed. The approach integrates mobile users' context information and social network information into collaborative filtering recommendation process. Firstly, it imports the user-based context similarities into the personalized service recommendation process. Secondly, it searches the corresponding nearest neighbors for each mobile user according to the given mobile users' ratings and the mining social network. Finally, it predicts unknown mobile users' preferences and generates recommendations. The experimental results show that CSMSR outperforms the existing service recommendation methods, such as TNCF, SRMTC and CF-DNC in terms of MAE (Mean Absolute Error) and P@N, and it performs well in finding out the interested services of users and enhancing the user experience.

**Key words:** mobile service recommendation; context; similarity measure; social network; collaborative filtering; data sparsity; cold-start problem

### 1 引言

随着移动通信技术的飞速发展, 移动设备已成为人们获取信息的主要平台之一. 移动设备拥有移动性强、上下文感知能力强、可携带性好、入网方便等特点,

让人们可以随时随地获取移动网络服务和信息内容. 然而移动互联网服务和信息传递颇受上下文信息、移动社会化网络的影响, 如何从浩瀚的移动信息海洋中发现用户感兴趣的服务, 提升用户个性化服务体验, 成为移动推荐系统亟待解决的难题<sup>[1]</sup>.

移动用户使用移动互联网时,大多与上下文环境(如时间、位置、周围人员、活动状态、网络条件等)密切相关.如有的移动用户喜欢“早上”而不是“下午”浏览新闻,有的喜欢“在公交车上”而不是“在办公室”玩游戏.只有将用户本身的动态上下文信息充分融入移动推荐系统,才能更好地生成移动推荐结果. Park 等人<sup>[2]</sup>通过贝叶斯网络模型构建地点、时间、天气、温度等上下文对移动用户的影响,使用 EM 算法学习在不同上下文中的条件概率,提高推荐准确度.高全力等人<sup>[3]</sup>分析各种上下文信息对用户认知行为的影响,提出了基于用户认知行为的上下文偏好获取算法,有效挖掘用户的认知行为与其偏好之间的关联关系,提高偏好获取准确度和推荐精度.

随着 Web 2.0 的发展,人们之间的虚拟交互变得愈发频繁,为了解决冷启动问题,学者们提出了社会化(social recommendation)推荐方法<sup>[4]</sup>,根据用户之间的社会关系信息构建用户之间的社会化关系网络.文献[5]将信任社会化网络与协同过滤算法结合,用户间的相似性不仅与用户对项目的评分相关,也与用户间的信任关系相关,将两者融合作为用户间相似的最终权重,避免传统协同过滤算法的稀疏性和冷启动问题,提高推荐准确性.文献[6~8]对用户间信任关系的传播现象进行建模,并给出了一种能够将社会关系信息进行整合的概率矩阵分解框架,提高推荐方法的性能.

针对以上背景,本文提出 CSMSR 方法,首先计算基于移动用户的上下文相似度,构造当前上下文的相似上下文集合,再采用相似上下文预过滤方法将“用户-服务-上下文”三维模型降维,得到“用户-服务”二维模型;然后利用移动用户的通信行为构建出的信任网络计算用户之间的信任度,按照信任度选取信任用户;再结合基于用户评分相似度计算发现的近邻,分别从相似用户和信任用户中选择相应的邻居用户,对目标用户进行偏好预测和推荐.

## 2 协同过滤算法简介

根据 Breese 等人<sup>[9]</sup>的分类方法,协同过滤推荐算法包括基于内存(memory-based algorithms)和基于模型(model-based methods)两种方法.其中,基于内存的方法又可分为基于用户(user-based)和基于项目(item-based)的算法.基于用户的协同过滤通过寻找邻居向用户推荐其可能感兴趣的物品;基于项目的协同过滤通过用户曾经评过分的物品寻找相似物品向用户推荐;为避免协同过滤技术中稀疏矩阵和处理大量数据时无法满足时效性等问题,提出了基于模型的协同过滤技术,该方法先用历史数据得到一个模型,模型的建立使

用贝叶斯网络技术<sup>[2]</sup>、聚类技术、潜在语义技术、概率模型<sup>[6~8]</sup>等机器学习方法,再用此模型进行预测,但其计算复杂性较高、模型训练时间较长.基于用户的推荐算法能够挖掘用户的潜在兴趣,简单易于实现,推荐个性化、自动化程度高,并能针对推荐结果进行合理解释,得到广泛应用,也是本文的研究对象.

基于用户的协同过滤推荐算法的基本思想是:(1)建立用户数据模型;(2)用户偏好相似度计算;(3)近邻用户选择;(4)产生预测.其主要实现过程包括相似度计算和用户偏好预测.相似性的度量方法主要包括 3 种:余弦相似性、Pearson 相关系数和修正的余弦相似性.其中 Pearson 相关系数使用最多,计算公式如下:

设用户  $a$  和用户  $b$  共同评分过的服务集合为  $S_{a,b}$ ,则用户  $a$  和用户  $b$  的相似度  $\text{sim}(a,b)$  为:

$$\text{sim}(a,b) = \begin{cases} \frac{\sum_{s \in S_{a,b}} (r_{a,s} - \bar{r}_a)(r_{b,s} - \bar{r}_b)}{\sqrt{\sum_{s \in S_{a,b}} (r_{a,s} - \bar{r}_a)^2} \sqrt{\sum_{s \in S_{a,b}} (r_{b,s} - \bar{r}_b)^2}}, & |S_{a,b}| \geq 2 \text{ 且 } H \neq 0 \\ 0, & |S_{a,b}| < 2 \text{ 或 } H = 0 \end{cases} \quad (1)$$

其中,  $H = \sqrt{\sum_{s \in S_{a,b}} (r_{a,s} - \bar{r}_a)^2} \sqrt{\sum_{s \in S_{a,b}} (r_{b,s} - \bar{r}_b)^2}$ ,  $r_{a,s}$  表示用户  $a$  对服务  $s$  的评分,  $\bar{r}_a$  和  $\bar{r}_b$  分别表示用户  $a$  和用户  $b$  对共同评分服务的平均评分.据式(1)可知,  $a$  和  $b$  的相似度满足  $-1 \leq \text{sim}(a,b) \leq 1$ ,  $-1$  表示两个用户完全负相关,  $0$  表示无关,  $1$  表示完全正相关.一般认为,  $\text{sim}(a,b)$  的值越大,两个用户的评分行为越相似.

根据用户间的相似度大小,选取相似度较大的前  $K$  个近邻用户作为目标用户的推荐近邻.考虑到不同用户有不同的评分尺度,采用式(2)进行评分预测:

$$R_{u,s} = \bar{r}_u + \frac{\sum_{p \in NB} \text{sim}(u,p) \times (r_{p,s} - \bar{r}_p)}{\sum_{p \in NB} \text{sim}(u,p)} \quad (2)$$

其中,  $R_{u,s}$  为目标用户  $u$  在目标服务  $s$  上的预测评分,  $NB$  为  $u$  的近邻集合,  $p$  为最近邻居集合中任一用户.

## 3 本文算法

### 3.1 上下文相似度计算

在移动社交网络服务环境中,上下文环境动态变化,用户消费行为也会因环境中上下文信息的变化而改变.移动社交网络可以获取的用户上下文很丰富:用户位置信息(通过手机 GPS 或者移动网络定位获得)、用户状态信息(presence)(如在线、离线、开会等)、手机日历/日程信息、天气温度信息、好友评价信息等.为找出和当前上下文相关联的评价,需要对上下文之间的

相似性进行度量。

**定义 1** 使用向量模型表示由  $n$  个上下文属性组成的上下文  $\mathbf{C} = (\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_n)$ , 向量中的分量  $\mathbf{C}_i (i = 1, 2, \dots, n)$  表示一种类型的上下文, 如时间, 地点等. 一个上下文实例表示为  $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n)$ ,  $\mathbf{c}_i (i = 1, 2, \dots, n)$  为上下文属性  $\mathbf{C}_i$  的一个具体属性值, 比如  $\mathbf{C}_i$  为时间上下文,  $\mathbf{c}_i$  可能取值有上午、下午、晚上. 例如, 上下文向量  $\mathbf{C} = (\text{时间}, \text{位置}, \dots, \text{活动状态})$ , 上下文实例  $\mathbf{c} = (\text{上午}, \text{在家}, \dots, \text{听音乐})$  表示用户上午在家听音乐的时候选择了某一项移动服务。

**定义 2** “移动用户-移动服务-上下文”三维模型记作“ $\mathbf{U}\text{-}\mathbf{S}\text{-}\mathbf{C}$ ”模型, 它是一个三维的向量空间  $\{\mathbf{U}, \mathbf{S}, \mathbf{C}\}$ , 每个维度用各自属性值组成的向量表示. (移动用户, 移动服务, 上下文, 偏好值) 称为一条偏好记录. 偏好记录的集合叫做偏好数据集。

本文借鉴上下文预过滤思想, 找出移动用户当前上下文  $\mathbf{c}$  的相似上下文集合  $N_{(c)}$  代替  $\mathbf{c}$  作为预过滤条件, 缓解传统上下文预过滤方法导致的数据稀疏性问题。

文中使用 Pearson 相关系数计算上下文相似度, 构造  $\mathbf{c}$  的相似上下文集合  $N_{(c)}$ . 式(3)是基于用户的上下文相似度计算公式, 衡量对同一用户而言, 不同上下文之间的相似度, 其值根据该用户在不同上下文环境下对相同服务的评分计算而得。

$$\text{sim}_u(\mathbf{x}, \mathbf{y}) = \frac{\sum_{s \in S} (r_{(u,s,x)} - \bar{r}_{u,x})(r_{(u,s,y)} - \bar{r}_{u,y})}{\sqrt{\sum_{s \in S} (r_{(u,s,x)} - \bar{r}_{u,x})^2} \sqrt{\sum_{s \in S} (r_{(u,s,y)} - \bar{r}_{u,y})^2}} \quad (3)$$

其中  $S$  表示用户  $u$  在上下文  $\mathbf{x}$  和上下文  $\mathbf{y}$  下都存在评分值的移动服务集合;  $r_{(u,s,x)}$  表示用户  $u$  在上下文  $\mathbf{x}$  下对服务  $s$  的评分值;  $\bar{r}_{u,x}$  表示用户  $u$  在上下文  $\mathbf{x}$  下的平均评分值。

实际情况中, 若数据集中的移动用户对同一移动服务评分数很少或没有, 则无法根据式(3)计算上下文相似度. 此时可以利用上下文信息之间的关系衡量它们之间的相似度, 即计算上下文特征相似度. 比如, 当上下文为时间时, 则越接近的时间相似度越大. 文中使用式(4)计算上下文实例  $\mathbf{x}, \mathbf{y}$  的特征相似度:

$$\text{sim}_c(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \quad (4)$$

构造当前上下文的相似上下文集合后, 便可按照上下文预过滤思想, 将三维模型降为二维, 再进行偏好预测。

### 3.2 移动用户的社会网络

移动用户之间的信任关系包括直接和间接信任关系, 直接信任关系由用户之间直接通信关系形成, 间接

信任关系由用户之间信任传播形成, 它们共同组成移动用户之间的通信信任网络. 这里用信任值作为评价用户间信任关系紧密程度的标准. 社会信任网络中的好友可以直接成为推荐邻居, 从而避免计算移动用户之间的评分相似度时可能存在的稀疏性, 尤其是新用户的冷启动问题。

#### 3.2.1 计算用户直接信任度

在移动通信网中, 不同用户之间的互动关系主要有语音通信和短信通信两种方式, 两个用户之间通信次数及时长是衡量用户好友关系的关键指标。

**定义 3** 假设在一定时间周期内, 用户  $u \in U$  与所有用户的通信过程中, 总的主叫通话次数为  $T$ , 总的主叫通话时间为  $P$ , 发出短信息的总数为  $Q$ . 在这一段时间内, 用户  $u \in U$  对用户  $v \in U$  总的主叫通话次数为  $t_{uv}$ , 总的主叫通话时间为  $p_{uv}$ , 发送短信息数量为  $q_{uv}$ , 则用户  $u$  对用户  $v$  的信任值定义如下:

$$\text{dtr}_{u,v} = \frac{1}{3} \times \left( \frac{t_{uv}}{T} + \frac{p_{uv}}{P} + \frac{q_{uv}}{Q} \right) \quad (5)$$

同理, 可以计算用户  $v$  对用户  $u$  的信任值  $\text{dtr}_{v,u}$ . 由此可以依次计算所有用户相互之间的直接信任值. 使用加权平均是为了将信任值大小控制在 0 到 1 之间。

#### 3.2.2 计算用户间接信任度

由通信记录确定的直接好友关系较稀疏, 不利于推荐, 而且现实中并非所有好友间都有通信行为, 因而从通信记录中无法全面获取用户间的好友关系, 还需挖掘间接信任关系. 本文根据信任 TidalTrust 算法<sup>[10]</sup> 计算用户的间接信任度, 通过间接信任关系进行填充, 在一定程度上解决信任矩阵的稀疏性问题, 提高推荐的准确性。

TidalTrust 算法是由 Golbeck 提出的一种基于广度优先搜索的信任度计算方法, 它首先查找源用户  $u$  与目标用户  $v$  之间的所有最短路径 (路径中每条边的信任度值不得小于阈值), 然后通过信任度权重来计算用户  $u$  对不直接相连的目标用户  $v$  的间接信任度值  $\text{idtr}_{u,v}$ , 具体的公式如式(6)所示:

$$\text{idtr}_{u,v} = \frac{\sum_{\text{dtr}_{u,i} \in \text{TR}} \text{dtr}_{u,i} \text{dtr}_{i,v}}{\sum_{\text{dtr}_{u,i} \in \text{TR}} \text{dtr}_{u,i}} \quad (6)$$

其中,  $\text{TR} = \{\text{dtr}_{u,i} \mid \text{dtr}_{u,i} \neq \text{null}, \text{dtr}_{u,i} \geq \lambda\}$  表示用户  $u$  的信任邻居中信任度大于阈值  $\lambda$  的用户集合。

### 3.3 融合移动用户信任关系的评分预测

信任关系体现用户间的相似度, 是用户相似性的综合体现, 并不针对具体项目的偏好. 为了解决数据稀疏性及冷启动用户的问题, 将移动用户的好友关系作为一种特殊的近邻关系引入到推荐生成过程中. 本文将移动用户偏好的相似关系与信任关系进行融合, 缓

解因评分稀疏造成不能准确计算移动用户间相似度,从而不能准确预测用户偏好的问题. 本文选择混合方法融合用户相似度和信任度,根据邻居选择策略,从相似用户和信任用户中分别选择相应的相似用户和信任用户. 由于移动用户之间的相似度和信任度使用不同

的计算方法,他们之间的值并不能直接进行比较. 如移动用户之间的相似度为 0.7,移动用户之间的信任度为 0.2,在预测评分时并不表示相似用户的影响比信任用户的影响大. 因此本文将区分考虑两类邻居用户对预测评分的影响,其计算公式如下:

$$R_{u,s} = \begin{cases} \bar{r}_u + \alpha \times \frac{\sum_{p \in \text{NB}} \text{sim}(u,p) \times (r_{p,s} - \bar{r}_p)}{\sum_{p \in \text{NB}} \text{sim}(u,p)} + (1 - \alpha) \times \frac{\sum_{t \in \text{TR}} \text{tr}_{u,t} \times (r_{t,s} - \bar{r}_t)}{\sum_{t \in \text{TR}} \text{tr}_{u,t}}, & \text{NB} \neq \emptyset \text{ 且 } \text{TR} \neq \emptyset \\ \bar{r}_u + \frac{\sum_{p \in \text{NB}} \text{sim}(u,p) \times (r_{p,s} - \bar{r}_p)}{\sum_{p \in \text{NB}} \text{sim}(u,p)}, & \text{NB} \neq \emptyset \text{ 且 } \text{TR} = \emptyset \\ \bar{r}_u + \frac{\sum_{t \in \text{TR}} \text{tr}_{u,t} \times (r_{t,s} - \bar{r}_t)}{\sum_{t \in \text{TR}} \text{tr}_{u,t}}, & \text{TR} \neq \emptyset \text{ 且 } \text{NB} = \emptyset \end{cases} \quad (7)$$

其中,NB 和 TR 分别表示用户  $u$  的相似用户集合和信任用户集合, $\text{sim}(u,p)$  和  $\text{tr}_{u,t}$  分别表示移动用户之间的相似度和信任度, $\alpha$  和  $1 - \alpha$  表示相似用户和信任用户在预测中所占的权重. 参数  $\alpha$  调整信任关系在用户评分预测中的权重.

#### 算法 1 CSMSR

输入:“用户-移动服务-上下文”评分数据集;目标用户  $u$  及当前上下文  $c$ ;移动用户的通信数据;待推荐的移动服务集合  $S$ .

输出:用户  $u$  在当前上下文  $c$  下,对  $S$  中偏好最大的 Top- $N$  项移动服务.

1. 取出目标用户  $u$  的所有偏好数据,构成基于  $u$  的“上下文-移动服务”二维偏好矩阵;
2. FOR EACH  $S_i$  IN  $S$
3. IF ( $u$  对  $S_i$  评分数  $< 2$ )
4. 利用式(4)构造  $c$  的相似上下文集合  $N_{(c)}$ ;
5. ELSE
6. 利用式(3)构造  $c$  的相似上下文集合  $N_{(c)}$ ; //过滤上下文信息后将“ $U-S-C$ ”三维模型降维成“ $U-S$ ”二维模型,并保证用户对一个移动服务有唯一偏好值
7. FOR EACH  $c_i$  IN  $N_{(c)}$ ;
8.  $w = c_i$
9.  $\max = \text{sim}_u(c_i, c)$
10. IF ( $\max < \text{sim}_u(c_i, c)$ )
11.  $w = c_i$ ;
12. END FOR
13. 用户  $u$  对移动服务  $S_i$  的唯一偏好记录  $= (u, S_i, w, r)$ ;
14. END FOR
15. 利用式(1)计算得出  $u$  的相似用户集合 NB;
16. 利用式(5)、(6)计算得出  $u$  的信任用户集合 TR;
17. 分别从相似用户集合 NB 和信任用户集合 TR 中各选择  $K$  个用户作为推荐近邻  $\text{NB}_k$  和  $\text{TR}_k$ ;
18. 根据  $\text{NB}_k$  和  $\text{TR}_k$  利用式(7)计算  $R_{u,s}$ ,并选取  $R_{u,s}$  值最高的  $N$  个作为最终推荐的 Top- $N$  项服务推荐给用户  $u$ .

### 3.4 性能分析

本算法的主要步骤是基于移动用户的上下文相似度计算、用户之间信任值预测以及对结果的排序算法. 一般情况下,Pearson 相关系数的计算复杂度为  $O(N^2)$ . 根据定义 3,移动用户之间的信任关系的计算复杂度为  $O(N)$ . 一般排序算法的计算复杂度为  $O(N^2)$ . 因此算法的整体复杂度为  $O(N^2)$ .

## 4 实验结果及分析

### 4.1 数据集

实验采用麻省理工学院多媒体实验室 MIT 收集的数据集<sup>[11]</sup>进行实验. 该数据集包括 106 个移动用户在 2004 年 7 月~2005 年 6 月共 12 个月移动用户的行为信息. 例如通话记录(通话和短信的时间、电话的主叫和被叫、短信的发送和接收)、位置、时间、手机状态、蓝牙设备、WLAN 设备访问信息等上下文. 数据集中包括通过问卷调查得出的 106 个移动用户之间的好友关系以及用户在实验室内和实验室外与其他人员的相处时间等数据.

为了验证本文所提方法的有效性,在 MIT 数据集上的实验,将通信数据的主体抽取出来,删除对所有网络服务评价记录为空的用户,这是为了删除一些广告及网外用户号码的无用数据. 选择 2004 年 7 月~2005 年 3 月这 9 个月内,106 个用户的网络服务使用数据作为训练集,将 2005 年 4 月~2005 年 6 月这 3 个月内的数据作为测试集.

### 4.2 评价指标

为了验证推荐的性能,使用平均绝对误差 (Mean Absolute Error, MAE) 和  $P@N$  评价指标.

MAE 主要根据用户的预测评分和实际评分衡量评分预测的准确性,MAE 值越小,表明推荐质量越高. 假

设预测的用户评分集合为  $\{p_1, p_2, \dots, p_N\}$ , 对应的实际用户评分集合为  $\{q_1, q_2, \dots, q_N\}$ ,  $N$  表示预测的次数, 则

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (8)$$

$P@N$  根据预测的用户可能常用的 Top- $N$  项网络服务占用户实际常用的 Top- $N$  项网络服务的比值, 表明对用户常用 Top- $N$  项网络服务的预测正确率.

$$P@N = \frac{\text{\#relevant services in Top-}N \text{ services}}{N} \quad (9)$$

#### 4.3 上下文相似度预过滤与传统上下文预过滤方法比较

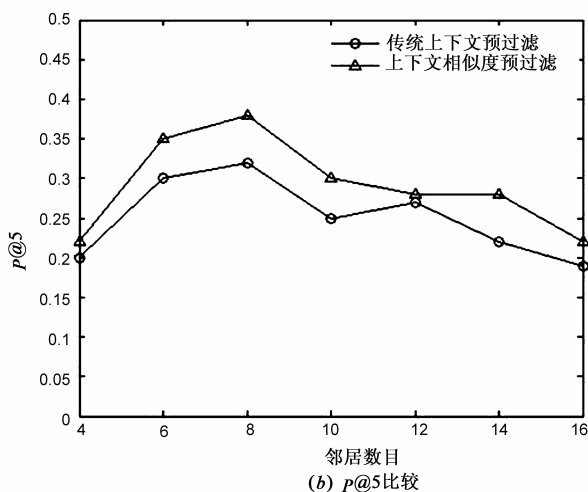
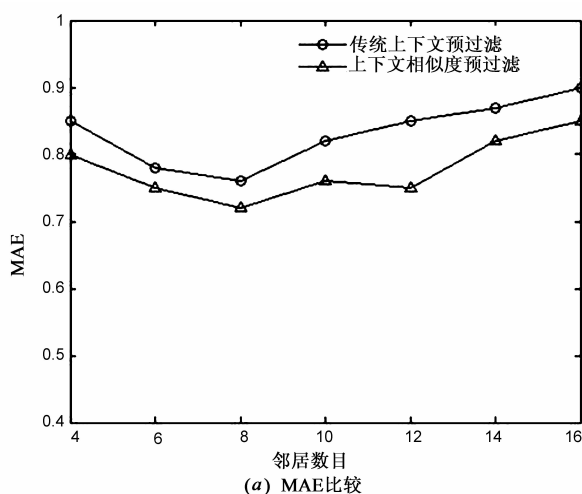


图1 两种方法比较

#### 4.4 对参数 $\alpha$ 的评估

参数  $\alpha$  调整相似用户和信任用户在评分预测中的权重, 以获得最佳推荐效果, 因此选择恰当的权重对 CSMSR 方法起至关重要的作用. 实验结果如图 2 所示.

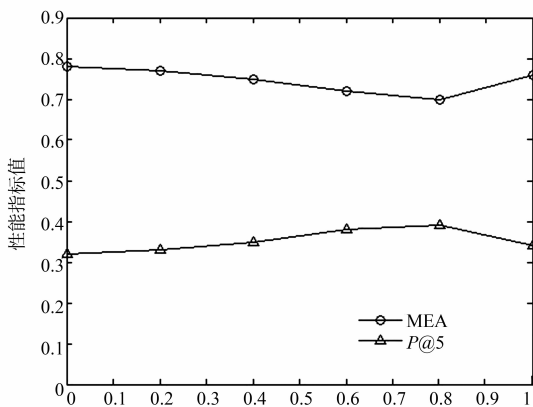


图2 不同 $\alpha$ 取值下MAE、 $P@5$ 的对比

从图中可以看出, 当  $\alpha$  取值为 0.7 到 0.8 之间时, MAE 值较小,  $P@5$  准确率较高, 推荐结果最优. 当  $\alpha = 0.1$  即较小时, MAE 值较大,  $P@5$  准确率较低; 随着  $\alpha$

该实验主要比较基于上下文相似度预过滤方法和传统上下文预过滤方法对推荐性能的影响. 当上下文表示为时间时, 传统的上下文预过滤方法使用目标用户所处的时间进行过滤, 而基于上下文相似度方法则利用与目标用户当前时间相似的时间段集合进行过滤. 实验结果见图 1.

从图中可看出, 选取不同邻居数目时, 上下文相似度预过滤方法比传统预过滤方法均具有更低 MAE 值, 更高  $P@5$  准确率. 这是因为利用相似上下文集合代替当前上下文有利于缓解数据稀疏性, 同时充分利用上下文信息提高推荐精度方面的作用.

逐步增加至 0.8 时, MAE 值逐步减小,  $P@5$  准确率逐渐增大; 权重  $\alpha$  继续逐步递增, 并越来越趋向于 1, MAE 值又缓慢递增,  $P@5$  的值逐渐减小. 在接下来的实验中, 将权重  $\alpha$  设置为 0.8 进行实验.

#### 4.5 性能及效率分析

对 CSMSR 方法, 使用 Top- $K$  邻居选取策略选择用户最近邻居数, 分别从相似用户和信任用户中选择 8 个最相似和最信任的用户组成邻居集合. 在计算用户之间的间接信任度值时, 为确保间接信任好友与目标用户之间的信任关系, 选取信任度阈值  $\lambda$  不低于 0.5 的用户集合作为信任邻居, 该值通过交叉验证实验选取.

本文选取现有的有代表性的服务推荐方法与本文提出的 CSMSR 方法进行比较, 包括王玉祥<sup>[5]</sup>等人提出的在上下文环境下基于信任网络和协作过滤算法 TNCF、王海艳<sup>[12]</sup>提出的基于可信联盟的服务推荐方法 SRMTC 及贾冬艳<sup>[13]</sup>提出的基于双重邻居选取策略的 CF-DNC 协同过滤推荐算法.

**实验 1** 分析这些服务推荐方法性能上的差异. 实验结果如图 3 所示.

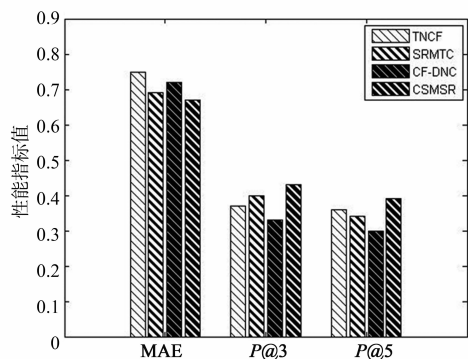


图3 CSMSR与对比算法推荐效果比较

由图可知,本文提出的 CSMSR 方法在  $P@3$  和  $P@5$  指标上的准确率显著高于其他三种算法,即对用户常用 Top- $N$  项服务的预测正确率较高。这在移动服务推荐中尤为重要,可以增加移动用户对推荐系统的信心。在 MAE 评价指标上,由于 SRMTC 方法既考虑了服务的推荐属性对用户相似度计算的影响,又加入了用户间的信任关系,排除了拥有较低信任度的用户推荐对建立用户联盟产生的不良影响,因此服务推荐的精确度与 CSMSR 不相上下,但是 SRMTC 方法的  $P@N$  准确率较低且执行效率低下(见实验 2)。另外两种方法则稍逊。CSMSR 方法通过把信任用户与相似用户融合在一起,形成最终的近邻用户集,一定程度上降低近邻相似矩阵的稀疏性,同时帮助缓解可能存在的冷启动用户问题,推荐效果最优。

**实验 2** 这些服务推荐方法效率比较。实验结果如图 4 所示。

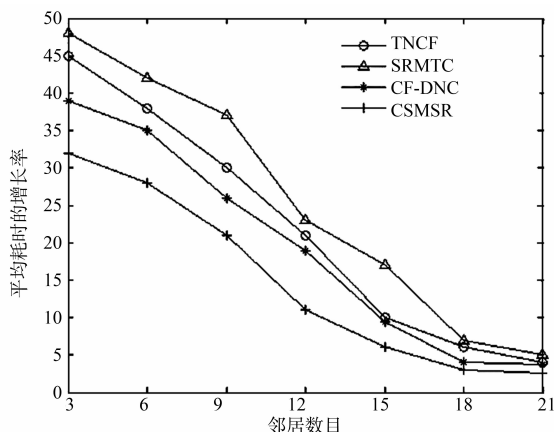


图4 耗时的平均增长速率

在近邻用户数稳步增长的条件,为每位用户提供网络服务推荐时增加了推荐候选的搜索范围,肯定消耗更多时间。正如图 4 所示,近邻用户数每增加 3 个,耗时的平均增长率都要大于 2.5%。但当近邻用户数达到 15 个以上时,耗时的平均增长率逐渐趋于稳定。也就是说,在近邻用户数增加到 18 个以后,在此基础上,近

邻用户数的增加量对平均耗时增长率的影响趋于稳定。这也表明此时近邻用户数量的增加不再是导致耗时增长的重要原因。此外,用户之间相似度及信任值的计算完全可以在线下完成的情况下,从平均耗时效率上考虑,说明该算法对增加用户数量的敏感程度不是很高。在耗时的增长率上,其他三种方法要明显高于本文提出的方法,这在实时性要求较高的移动网络服务环境下是不可忽视的优点。

## 5 结语

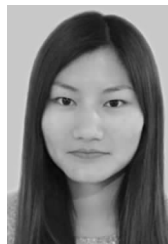
本文基于启发式的方法将上下文相似度和移动用户信任关系与协同过滤算法融合,为移动用户提供个性化网络服务推荐,既缓解了稀疏性和冷启动性,又考虑了上下文的动态性,更能接近现实世界并且具有更高的预测精确度。

## 参考文献

- [1] Jonna H, Albrecht S, Jani M, et al. Context-aware mobile media and social networks[A]. Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services[C]. Bonn, Germany: IEEE, 2009. 1-3.
- [2] Park M H, Hong J H, Cho S B. Location-Based recommendation system using Bayesian user's preference model in mobile devices[A]. Proceedings of the Ubiquitous Intelligence and Computing 2007[C]. Berlin: Springer-Verlag, 2007. 1130-1139.
- [3] 高全力,高岭,等.上下文感知推荐系统中基于用户认知行为的偏好获取方法[J]. 计算机学报, 2015, 38(9): 1767-1776.  
Gao Quan-li, Gao Ling, et al. A preference elicitation method based on users' cognitive behavior for context-aware recommender system[J]. Chinese Journal of Computers, 2015, 38(9): 1767-1776. (in Chinese)
- [4] Wang Z, Sun L, Zhu W, et al. Joint social and content recommendation for user-generated videos in online social network[J]. IEEE Trans on Multimedia, 2013, 15(3): 698-710.
- [5] 王玉祥,乔秀全,等.上下文感知的移动社交网络服务选择机制研究[J]. 计算机学报, 2010, 33(11): 2126-2135.  
Wang Yu-xiang, Qiao Xiu-quan, et al. Research on context-awareness mobile SNS service selection mechanism[J]. Chinese Journal of Computers, 2010, 33(11): 2126-2135. (in Chinese)
- [6] Jamali M, Ester M. A matrix factorization technique with trust propagation for recommendation in social networks[A]. Proceedings of the ACM Recommender Systems

- Conf[C]. New York: ACM Press, 2010. 135 – 142.
- [7] Ma H, Zhou D, et al. Recommender systems with social regularization[A]. Proceedings of the 4th ACM International Conference on Web Search and Data Mining[C]. Hong Kong, China: ACM Press, 2011. 287 – 296.
- [8] 印桂生, 张亚楠, 等. 基于受限信任关系和概率分解矩阵的推荐[J]. 电子学报, 2014, 42(5): 904 – 911.  
YIN Gui-sheng, ZHANG Ya-nan, et al. A constrained trust recommendation using probabilistic matrix factorization [J]. Acta Electronica Sinica, 2014, 42(5): 904 – 911. (in Chinese)
- [9] Breese J S, Heckerman D, et al. Empirical analysis of predictive algorithms for collaborative filtering[A]. Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence [C]. San Francisco: Morgan Kaufmann Publishers Inc, 1998. 43 – 52.
- [10] Golbeck J. Computing and Applying Trust in Web-based Social Networks [D]. Washington: University of Maryland, College Park, 2005.
- [11] Eagle N, Pentland A, et al. Inferring social network structure using mobile phone data[J]. Proceedings of the National Academy of Sciences (PNAS), 2009, 106(36): 15274 – 15278.
- [12] 王海艳, 杨文彬, 等. 基于可信联盟的服务推荐方法[J]. 计算机学报, 2014, 37(2): 301 – 311.  
Wang Hai-yan, YANG Wen-bin, et al. A service recommendation method based on trustworthy community[J]. Chinese Journal of Computers, 2014, 37(2): 301 – 311. (in Chinese)
- [13] 贾冬艳, 张付志. 基于双重邻居选取策略的协同过滤算法[J]. 计算机研究与发展, 2013, 50(5): 1076 – 1084.  
Jia Dong-yan, Zhang Fu-zhi. A collaborative filtering recommendation algorithm based on double neighbor choosing strategy[J]. Journal of Computer Research and Development, 2013, 50(5): 1076 – 1084. (in Chinese)

#### 作者简介



**俞春花** 女, 1991 年 1 月出生于江苏省泰州市. 现为南京工业大学计算机科学与技术学院硕士研究生. 主要研究方向为数据挖掘和个性化推荐.

E-mail: chunhuayu1991@163.com



**刘学军** 男, 1970 年 6 月出生于吉林省长春市. 博士, 现为南京工业大学教授、硕士生导师, CCF 高级会员. 主要研究方向为数据库、数据挖掘、传感器网络、隐私保护等.

E-mail: lxj\_njgd@163.com