

# 基于混杂社会网络的个性化 Web 服务推荐方法

杨 洁<sup>1</sup>, 朱咸军<sup>2</sup>, 周献中<sup>3</sup>, 柳 毅<sup>1</sup>

(1. 杭州电子科技大学信息管理学院, 浙江杭州 310018; 2. 金陵科技学院软件工程学院, 江苏南京 211169;

3. 南京大学控制与系统工程系, 江苏南京 210093)

**摘 要:** 为了缓解 Web 服务推荐中存在的冷启动和数据稀疏问题, 以及满足用户个性化的需求, 本文提出了基于混杂社会网络的 Web 服务推荐框架及算法. 该网络加入了服务提供者这一元素, 可提供更多的真实信息, 有助于缓解冷启动问题. 根据提出的服务推荐框架, 设计了用户-候选服务信任值预测算法 (Computing Trust Value, CTV), 以及服务推荐算法 (Recommend Queue, RQ). 在真实数据集上建立实验, 结果表明本文提出的方法在预测准确率 MAE、RMSE, 推荐准确率 MAP、NDCG, 以及填充率和覆盖率上都有所提高, 有助于提升个性化推荐的性能.

**关键词:** 混杂社会网络; Web 服务推荐; 个性化推荐; 服务提供者; 冷启动; 可信性

**中图分类号:** TP311 **文献标识码:** A **文章编号:** 0372-2112 (2020)02-0341-09

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.02.017

## Personalized Web Service Recommendation Based on Heterogeneous Social Network

YANG Jie<sup>1</sup>, ZHU Xian-jun<sup>2</sup>, ZHOU Xian-zhong<sup>3</sup>, LIU Yi<sup>1</sup>

(1. Department of Information Management, Hangzhou Dianzi University, Hangzhou, Zhejiang 310018, China;

2. School of Software Engineering, Jinling Institute of Technology, Nanjing, Jiangsu 211169, China;

3. Department of Control and System Engineering, Nanjing, Jiangsu 210093, China)

**Abstract:** To ease cold start and data sparsity problems in web service recommendation, and satisfy user's personalized requirements, we propose a web service recommendation framework and algorithms based on heterogeneous social network. Web service provider, as an element in the network, could provide more real information to help relieve cold start. Based on the proposed recommendation framework, we design CTV (Computing Trust Value) algorithm to predict trust value of user to candidate service and RQ (Recommend Queue) algorithm to recommend services. At last, we conduct a series of experiments based on real data set. The results indicate that the proposed method outperforms the existing methods at the metrics of MAE, RMSE for predictive accuracy, MAP, NDCG for recommendation accuracy, and filling ratio and coverage ratio, and it is helpful to improve personalized recommendation performance.

**Key words:** heterogeneous social network; web service recommendation; personalized recommendation; web service provider; cold start; trustworthiness

## 1 引言

随着面向服务的计算模式的广泛应用, 发布于互联网上的 Web 服务数量成指数级增长, 如何为用户推送满足个性化需求的服务成为了急需解决的问题. 为用户推送满足功能需求的服务已取得了重要的成果<sup>[1,2]</sup>. 部分研究的焦点转移到了满足用户非功能需求的推荐<sup>[3]</sup>. 在实现个性化推荐的同时, 解决常见的冷启

动、数据稀疏问题, 便是本文的研究主题.

随着社会网络的兴起, 用户间建立联系、分享信息、评价物品的使用体验, 逐渐形成了信任关系. 结合用户间信任和社会网络应用于 Web 服务推荐中<sup>[4,5]</sup>, 具有缓解冷启动和数据稀疏的作用<sup>[6,7]</sup>, 同时也能提高个性化推荐的性能. Chang 等<sup>[8]</sup>根据社会网络中用户之间的相似度提出了社会距离的概念, 确定用户之间的分离程度, 并结合用户的评分记录估算信任值. Jiang 等<sup>[9]</sup>介绍

了多种信任模型,结合社会网络构建形成信任网络图以应用于推荐系统中.俞春花等<sup>[10]</sup>将用户-用户信任与上下文相似度结合以实现个性化的移动服务推荐.Kalař等<sup>[11]</sup>基于可信用户的专业知识和时间因素,应用协同过滤(Collaborative Filtering, CF)推荐服务.

用户对服务的信任评价,被广泛应用于服务组合、QoS(Quality of Service)预测等研究中<sup>[12,13]</sup>.信任划分为直接信任与间接信任<sup>[14]</sup>,直接信任是指用户对服务使用的历史评价,间接信任是参照用户所属的用户群对服务的信任评价.杨震等<sup>[13]</sup>提出信任由基础信任和经验信任组合而成,应用 Bayes、Chebyshe 和 Bernstein 给出了基础信任的判定和经验信任的量化方法. Deng等<sup>[5]</sup>利用朋友的反馈和相似性建立自身对服务的信任.

将用户-服务信任和社会网络模型结合应用于服务推荐中,有助于提升个性化推荐的性能. Tang等<sup>[4]</sup>基于用户-用户和用户-服务信任,应用 CF 实现推荐以满足用户的偏好,未将冷启动问题作为研究的目标. Deng等<sup>[5]</sup>将用户-服务信任与社会网络结合,提出随机游走算法实现服务推荐.

不仅用户之间、用户-服务存在信任,用户对服务提供者(Web Service Provider, WSP)也有信任<sup>[15]</sup>. 马文龙等<sup>[3]</sup>提出将对 WSP 的信任划分为 0 到 1 之间的五个状态. Sun<sup>[16]</sup>建立基于 WSP 交互的社会网络,以定位协同合作者. 束束等<sup>[14]</sup>将 WSP 的声誉作为用户衡量资源信任值的重要组成成份.

WSP 给出服务的具体功能描述,发布符合服务等协议的服务;WSP 是易获取的,服务被发布时,它的 WSP 信息也随之公布;它可以为个性化推荐提供支持,避免对未评分的服务采取同等对待的做法.

本文提出了基于混杂社会网络的推荐框架及算法.将服务、WSP 融入到社会网络,用户-用户、用户-服务、用户-WSP 间的信任以及 WSP-服务提供联系构建形成混杂社会网络.由于用户间的信任联系以及 WSP-服务的提供联系包含了丰富的信息,为提升推荐性能和改善冷启动问题提供了可靠的依据.

## 2 结合信任的混杂社会网络

结合信任构建混杂社会网络(Heterogeneous Social Network, HSN)模型如图 1 所示,划分为两层:云层与用户层.该网络包含了三类结点:用户、Web 服务、WSP,两类联系:信任、提供.包含的联系为:(1) 用户-服务信任;(2) 用户-WSP 信任;(3) WSP-服务提供;(4) 用户-用户信任.两结点之间存在连接代表存在信任或提供联系.信任值的取值范围为 $[0,1]$ ,取值为 0 代表没有连接.提供的取值为 0 或 1.

社会网络:在用户层上,以用户为结点,用户之间的

信任联系为边,形成了基于信任的用户社会网络.具有以下特性:(1) 领域相关性.根据功能领域,用户对同一用户有不同的信任值;(2) 传递性.用户之间的信任关系是传递的.

HSN:结合社会网络和服务、WSP,形成了混杂社会网络.在保有社会网络特性的同时,还具有以下特性:(1) 异质性.网络中包含了不同类型的结点和联系,体现了网络结构的异质性;(2) 个性化.用户对服务、WSP 的信任值反映了用户的个性化认识;(3) 相互独立性.假设 WSP 之间是相互独立的,不会共享用户信息;(4) 真实性. WSP 提供服务的信息是真实的、易获取的,增强了推荐的可靠性.

### 2.1 网络构成

HSN 是一个有向图,可形式化为: $G = (S, P, U, E, W, t)$ .

$S, P, U$  分别表示图中结点的集合;

$S$  为发布于注册中心的服务集合,  $(s_1, s_2, \dots, s_m)$ ;

$P$  为提供了  $S$  的 WSP 的集合,  $(p_1, p_2, \dots, p_n)$ ;

$U$  为在注册中心所有用户的集合,  $(u_1, u_2, \dots, u_o)$ ;

$E$  为图中有向边的集合,  $E = \langle E_1, E_2 \rangle$ .  $E_1$  表示 WSP-服务提供联系,当  $p_i$  提供了服务  $s_j$ , 边的取值为 1, 当没有提供关系时,不存在边.  $E_2$  表示结点间构成的信任联系,  $E_2 = (T_1, T_2, T_3)$ , 分别为用户-用户信任  $T_1$ , 用户-服务信任  $T_2$ , 用户-WSP 信任  $T_3$ .

$W$ :表示 HSN 中的所有边的权值集合.

$t$ :标记一个时间窗.在一定的时间范围内,信任值是不变的.

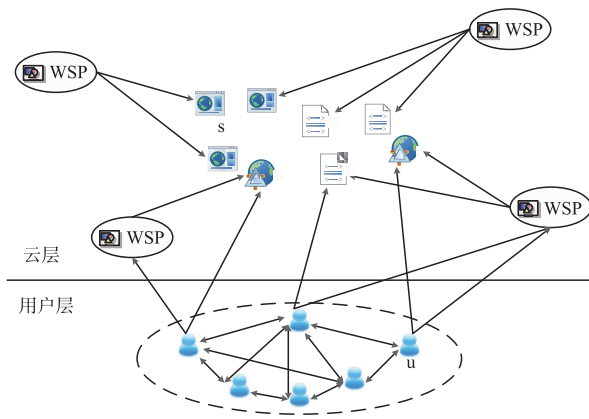


图1 结合信任的HSN

### 2.2 网络信任值计算

#### 2.2.1 用户-用户信任

为了向当前用户  $U$  推荐具有特定功能的服务, 认定为  $U$  信任的用户群的前提条件是:在与推荐任务相关的领域内,与  $U$  相似性高的用户群. 假设要为  $U$  推荐天气服务,查找调用过与天气相关的服务的用户,且该

用户与  $U$  相似性高,因而这个用户对天气领域的观点是可信的.  $U$  对可信任用户  $u_i$  的信任值表示为:

$$T_1(U, u_i) = \text{sim}(U, u_i) \quad (1)$$

其中,  $u_i \in U_i$ , 表示  $u_i$  属于可信任用户集  $U_i$ .

$U_i = \{u_i | \text{sim}(U, u_i) \geq \sigma, f(u_i, D(s)) = 1\}$ .  $\text{sim}(U, u_i) \geq \sigma$  表示相似性阈值设定为  $\sigma, \sigma > 0$ .  $f(u_i, D(s)) = 1$  表示  $u_i$  调用过与  $s$  相关领域的服务,  $D(s)$  表示与  $s$  相关的领域.

用户间相似性  $\text{sim}(u_i, u_j)$  采用皮尔逊相关系数计算:

$$\text{sim}(u_i, u_j) = \frac{\sum_{s \in s_c} (T_2(u_i, s) - \overline{T_{u_i}})(T_2(u_j, s) - \overline{T_{u_j}})}{\sqrt{\sum_{s \in s_c} (T_2(u_i, s) - \overline{T_{u_i}})^2} \cdot \sqrt{\sum_{s \in s_c} (T_2(u_j, s) - \overline{T_{u_j}})^2}} \quad (2)$$

$s_c$  为  $u_i$  与  $u_j$  调用过的服务的交集.  $\overline{T_{u_i}}, \overline{T_{u_j}}$  是  $u_i, u_j$  调用过的服务的信任值均值, 取值范围为  $[-1, 1]$ . 取负值时视为用户之间无边相连. 两用户之间的相似性满足对称性.

$$\text{sim}(u_i, u_j) = \text{sim}(u_j, u_i) \quad (3)$$

在获取指定的推荐任务后, 计算任意两个用户之间的相似性, 形成带有权重的用户社会网络, 获取  $U$  信任的用户群.

在遍历社会网络时, 如果两个用户之间没有边直接相连, 但存在路径可达, 可应用式(4)计算间接信任.

$$T_1(u_i, u_k) = T_1(u_i, u_j) \cdot T_1(u_j, u_k) \quad (4)$$

定义两用户之间途经的阶段不超过常数  $l$ . 根据社会网络的小世界特性<sup>[17]</sup>, 将  $l$  设为 6. 当存在多条路径时, 选择间接信任值最大的路径.

### 2.2.2 用户-服务信任

用户-服务信任值度量, 包括直接信任与间接信任. 直接信任是用户对服务的使用体验、反馈.

对 QoS 指标归一化处理, 根据用户的个性化特点为每项 QoS <sub>$i$</sub>  设置不同的权重  $\lambda_i$  形成直接信任值. 可表示为:

$$T_2(U, s) = \sum_{i=1}^q \lambda_i \cdot \text{QoS}_i^U \quad (5)$$

其中,  $\sum_{i=1}^q \lambda_i = 1, 0 \leq \lambda_i \leq 1$ .

QoS <sub>$i$</sub>  <sup>$U$</sup>  是  $U$  对第  $i$  项 QoS 的评价值, 共有  $q$  项指标. 不同用户对同一项 QoS 有不同的评价, 体现了用户的个性化特点.

随着用户调用服务次数的增加, 直接信任值是动态变化的. 其计算为:

$$T_2^d(U, s) = \frac{\sum_{i=1}^N T_2(U, s)}{N} \quad (6)$$

$N$  为用户调用服务的总次数,  $T_2^d(U, s)$  为用户调用多次后的信任值的均值.

间接信任值的计算参照  $U$  信任的用户群对服务  $s$  的评价, 为  $U, u_i$  间信任值与  $u_i$  对  $s$  的直接信任值乘积的均值, 如式(7)所示.

$$T_2^d(U, s) = \frac{\sum_{i=1}^n T_1(U, u_i) \cdot T_2^d(u_i, s)}{n} \quad (7)$$

其中,  $T_1(U, u_i) \geq \sigma, \sigma > 0$ , 表示  $u_i$  是  $U$  信任的用户. 在后续实验中,  $\sigma$  设为 0.3.  $n$  是  $U$  信任的用户数量.

用户与服务间不存在直接信任值时,  $U$  遍历社会网络. 当  $U$  与用户群之间存在路径可达, 且根据式(1)确定为可信任的用户群.

假设  $u'$  是可信任的用户,  $U$  到  $u'$  经过的结点依次为  $u_{i+1}, \dots, u_{i+n}$ , 信任值计算表示为:

$$T_{2,p}(U, s) = T_1(U, u_{i+1}) \cdots T_1(u_{i+n}, u') \cdot T_2(u', s) \quad (8)$$

其中,  $T_1(U, u_{i+1}) \cdots T_1(u_{i+n}, u') \geq \sigma, \sigma > 0$ , 表示  $U$  可信任的用户需满足的阈值.

当  $U$  到  $u'$  之间存在多条路径时, 用户-服务信任值定义为多条路径的信任值均值:

$$T_2(U, s) = \sum_{p=1}^P T_{2,p}(U, s) / P \quad (9)$$

共有  $P$  条路径.

计算信任值的路径选择依据是: 如果  $U$  与  $u'$  之间存在直接路径可达, 尽管两者之间存在其他更复杂的路径(如路径中存在回路), 信任值不受复杂路径的影响.

### 2.2.3 用户-WSP 信任

对于一个服务来说, WSP 是有效信息. 它作为服务的重要信息, 提出用户-WSP 信任值的计算方式, 为推荐提供依据.

WSP 的使用频率体现了用户选择服务的隐式偏好. 用户调用 WSP 的频率可作为 WSP 信任值计算的依据.

服务-WSP 数量关系服从幂律分布, 以数据集 WS-Dream-QoSDataSet2<sup>[18]</sup> 为例, 如图 2 所示.

基于服务-WSP 之间的数量关系, 假定 WSP 的使用频率服从幂律分布, 被频繁使用的 WSP 是少量的<sup>[19]</sup>.

用户  $u_i$  对  $p_j$  的使用频率为  $u_i$  使用  $p_j$  的次数与  $u_i$  使用所有 WSP 的次数之比, 表示为:

$$\text{Fre}(u_i, p_j) = q_{i,j} / \sum_{j=1}^N q_{i,j} \quad (10)$$

$N$  为 WSP 的数量,  $q_{i,j}$  为  $u_i$  使用  $p_j$  的次数.

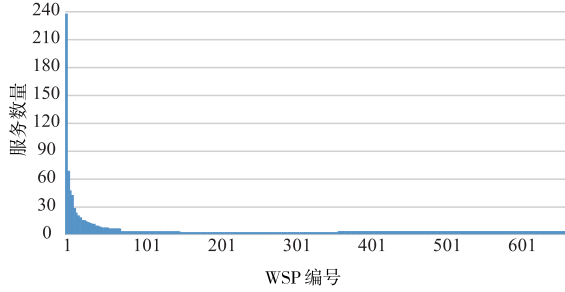


图2 服务-WSP数量的幂律分布

将用户-WSP 的使用频率转换为用户-WSP 的评分. 应用线性函数, 转换后评分的取值范围为  $[0, 5]$ . 对所有 WSP 的使用频率降序排序,  $u_i$  对排第  $j$  位的 WSP 的评分为:

$$r(u_i, p_j) = 5(1 - \sum_{j=1}^{j-1} \text{Fre}(u_i, p_j)) \quad (11)$$

对评分值归一化处理可获取用户对 WSP 的信任值:

$$T_3(u_i, p_j) = \frac{r(u_i, p_j) - \min_{i,j} r(u_i, p_j)}{\max_{i,j} r(u_i, p_j) - \min_{i,j} r(u_i, p_j)} \quad (12)$$

应用皮尔逊相关系数计算可得任意两个用户在 WSP 选择上的相似度, 从而为用户推荐 WSP.

$$c(u_i, u_j | \text{WSP}) =$$

$$\frac{\sum_{k=1}^t (r(u_i, p_k) - \overline{r(u_i, p)}) (r(u_j, p_k) - \overline{r(u_j, p)})}{\sqrt{\sum_{k=1}^t (r(u_i, p_k) - \overline{r(u_i, p)})^2 \sum_{k=1}^t (r(u_j, p_k) - \overline{r(u_j, p)})^2}} \quad (13)$$

$t$  为用户  $u_i, u_j$  共同调用的 WSP 的数量.

根据用户-WSP 信任预测用户-服务信任值:

$$T_2(U, s) = \frac{m_j T_3(U, p_j) + \sum_{i=1}^N T_2(U, s_i)}{m_j + N} \quad (14)$$

$N$  为  $U$  调用的  $p_j$  所提供的服务的数量,  $m_j$  为  $U$  对  $p_j$  的调用次数. 对 WSP 的调用次数越多, 表明用户对它的信任依赖越大, 对它提供的其它服务的信任也越大.

### 3 基于混杂社会网络的服务推荐

#### 3.1 服务推荐框架

基于  $U$  对服务的功能需求, 系统将返回符合用户个性化需求的服务列表 (TOP- $N$ ), 列表中的服务以信任值降序排序. 将服务推荐工作划分为两部分: (1)  $U$ -候选服务信任值计算及预测; (2) 对候选服务信任值排序形成可推荐序列. 服务推荐框架如图 3 所示.

由于用户数、服务数量庞大, 首先对用户集、服务集过滤. 服务过滤是查找符合  $U$  的功能需求的服务形成候选服务集. 用户过滤是查找曾使用过候选服务的用

户并形成候选用户集, 可缩小社会网络遍历的范围, 降低算法复杂度.

(1)  $U$ -候选服务信任值计算及预测步骤包括:

(a) 根据  $U$  的历史使用记录, 计算  $U$  对候选服务的直接信任值.

当  $U$  对候选服务没有使用经验时, 转向步骤 (b).

(b) 在候选用户集中计算用户-用户信任值. 遍历社会网络计算  $U$  对候选服务的间接信任值.

当初始用户-服务矩阵是高稀疏矩阵,  $U$  对候选服务不存在直接信任, 且  $U$  不存在可信任的用户时, 上述两种方式都会失效, 应用 WSP 预测  $U$ -候选服务信任值, 执行步骤 (c).

(c)  $U$ -WSP 信任计算. 查找候选服务的 WSP 信息, 计算  $U$ -WSP 的信任值, 并据此预测  $U$ -候选服务的信任值.

(2) 形成推荐列表: 根据第 (1) 步计算所得的  $U$ -候选服务信任值, 按降序排序形成推荐序列返回给  $U$ .

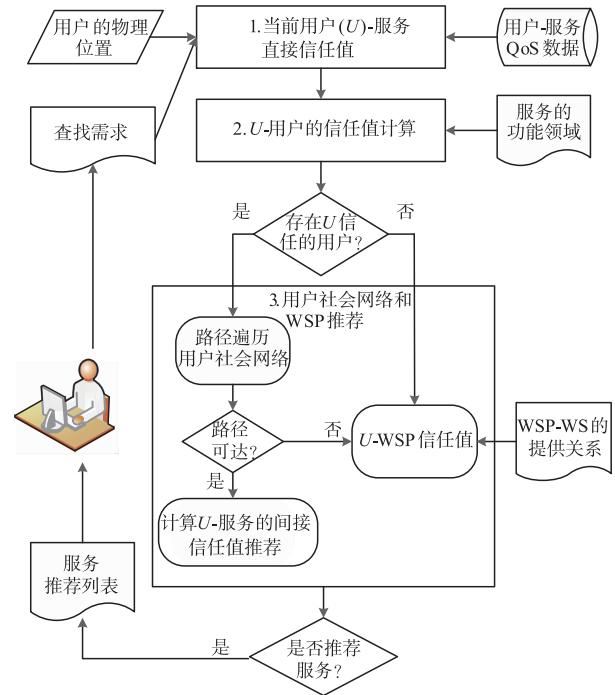


图3 服务推荐框架

#### 3.2 $U$ -候选服务信任值计算及预测

根据候选用户集与候选服务集, 计算、预测  $U$  对候选服务的信任值, 执行 CTV 算法.

##### 算法1 CTV (Computing Trust Value)

输入:  $U$ ; 候选服务  $s$

输出:  $U$ - $s$  信任值

1. 建立并初始化混杂社会网络 HSN

2. 创建用户-服务信任值矩阵  $T$ :

```

3. 根据式(5)计算  $T_2(u_i, s_j)$ 
4. 创建 WSP-服务的提供矩阵  $P: P[p_k, s_j]$ 
5. 创建用户-用户信任值矩阵  $U_T$ :
6. 根据式(1)和(4)计算  $T_1(u_i, u_j)$ 
7. IF  $T(U, s) > 0$ 
8.   RETURN  $T(U, s)$ 
9. ELSE
10.  IF  $T(U, s) = 0$ 
11.   FOR  $i$  FROM 1 TO  $m$ 
12.    IF  $T(u_i, s) > 0$ 
13.     THEN 标记  $u_i$  并存储到集合  $C(u)$ 
14.   ENDFOR
15.  ENDIF
16. ENDIF
17. PATH = TraversingSN( $U, C(u)$ )
18. IF PATH 不为空
19.  根据式(8)和式(9)计算信任值  $T_2(U, s)$ 
20. ELSE
21.  FOR  $k$  FROM 1 TO  $l$ 
22.   根据式(12)计算信任值  $T_3(U, p_k)$ 
23.  ENDFOR
24. 根据式(14)计算  $T_2(U, s)$ 
25. ENDIF

```

1 行建立并初始化混杂社会网络 HSN. 2、3 行创建用户-服务信任值矩阵  $T$ , 计算候选用户  $u_i$ -候选服务  $s_j$  的直接信任值. 4 行创建 WSP-服务的提供矩阵  $P$ , 提供者  $p_k$  与服务  $s_j$  的提供关系. 5、6 行创建用户-用户信任值矩阵  $U_T$ . 7、8 行判定矩阵  $T$  中  $U$  对  $s$  的信任值是否大于 0, 是则返回该值; 否则转向第 10 行. 11 ~ 14 行从候选用户集中查找对  $s$  的直接信任值大于 0 的用户, 并存储到集合  $C(u)$ . 17 行遍历社会网络, 查找  $U$  与  $C(u)$  之间的路径. 19 行计算  $U$  对  $s$  的信任值. 当路径为空时, 转 20 行. 21 ~ 23 行计算  $U$  对  $p_k$  的信任值. 24 行根据  $U$  对  $s$  的提供者  $p_k$  的信任值, 计算  $U$  对  $s$  的信任值.

由于用户-服务的直接信任值, 以及用户-WSP 信任值计算与遍历社会网络是顺序执行结构, 因而 CTV 算法的复杂度取决于遍历方法的复杂度. 假设  $U_T$  矩阵的稀疏性为  $\gamma$ , 候选用户集的大小为  $q$ , 以及小世界特性设定常数  $l$  为 6, 则算法的时间复杂度为  $O(((1-\gamma)q)^5 \cdot |C(u)|)$ .

### 3.3 服务推荐算法

推荐算法是计算用户集对候选服务集的信任值, 并对信任值按降序排序, 给每个用户提供个性化的服务序列.

#### 算法 2 RQ (Recommend Queue)

输入: TOP- $N$ ; 用户集  $U_C = \{U_1, \dots, U_m\}$ ; 候选服务集  $S_C = \{s_1, \dots, s_n\}$ ;

输出: 为每个用户推荐的服务序列

```

1. FOR  $U_i$  FROM  $U_1$  TO  $U_m$  IN  $U_C$ 
2.  FOR  $s_j$  FROM  $s_1$  TO  $s_n$  IN  $S_C$ 
3.    $T_2(U_i, s_j) = \text{CTV}(U_i, s_j)$  //调用 CTV 算法
4.  ENDFOR
5.  Sort( $T_2(U_i, s_j)$ ) //对信任值降序排序
6.  Output 为  $U_i$  推荐的前  $N$  的服务序列
7. ENDFOR

```

假设用户数量为  $m$ , 候选服务集的大小为  $n$ , 排序算法的时间复杂度为  $O(n^2)$ , 则 RQ 的时间复杂度为:  $O(m(nO(\text{CTV}) + O(n^2)))$ . 由于  $O(n^2) < O(\text{CTV})$ , 所以 RQ 的时间复杂度为  $O(mnO(\text{CTV}))$ .

## 4 实验

### 4.1 实验环境与配置

本文的实验环境与配置为: ThinkPad laptop, CPU 为 Intel Core i5, 1.8GHz, 内存为 4.00GB, 编程语言 Java 1.7, 实验平台为 Eclipse 4.3. 采用真实数据集 WS-Dream-QoSDataSet2, 去除只提供了一个服务的 WSP 形成数据集. 包含用户数为 339, 服务数为 4000, 以及 WSP 数量为 883.

服务过滤: 查找功能为查询的服务, 利用关键字搜索包含“search”的服务, 总数为 138, 形成候选服务集.

用户过滤: 由于数据集中, 每个用户都调用过候选服务, 因而候选用户集的大小为 339.

建立用户-服务信任矩阵: 对 QoS 值归一化处理. 响应时间、吞吐量的权重设为 0.5 和 0.5, 形成  $339 \times 138$  的矩阵  $T'$ .

建立 WSP-服务提供矩阵: 提取数据集中的信息, 形成  $883 \times 4000$  的 0-1 矩阵  $P$ .

建立用户-用户信任矩阵: 根据用户-服务信任值, 形成  $339 \times 339$  矩阵  $U_T'$ .

建立用户-WSP 信任矩阵: 根据 WSP 被使用次数呈现幂律分布的特性, 模拟生成 WSP 的使用频次. 根据式(11)和(12)计算用户-WSP 的评分、信任值, 形成矩阵  $U_p$ .

随机地选取  $T'$  中 80% ~ 99% 的元素置为空值, 模拟初始矩阵  $T$  的稀疏度变化范围. 根据不同稀疏度下的  $T$  矩阵计算生成用户-用户信任矩阵  $U_T$ . 矩阵  $P$  不会根据  $T$  稀疏度的变化而变化. 在不同的稀疏矩阵上, 运行 CTV 和 RQ 算法预测用户-服务信任值, 形成推荐列表.

### 4.2 衡量标准

#### 4.2.1 用户-服务信任值预测准确率

用户-服务信任值的预测准确率的衡量采用绝对平均误差 (Mean Absolute Error, MAE) 和均方根误差 (Root

Mean Square Error, RMSE), 分别表示用户-服务信任预测值与观察值之间的误差均值和方差.

$$MAE = \frac{\sum_{i,j} |r_{i,j} - \hat{r}_{i,j}|}{M} \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_{i,j} (r_{i,j} - \hat{r}_{i,j})^2}{M}} \quad (16)$$

其中,  $r_{i,j}$  为用户  $u_i$  对服务  $s_j$  的信任观察值,  $\hat{r}_{i,j}$  为预测值,  $M$  为预测值的数量.

#### 4.2.2 推荐算法的 FR 和 CR

采用填充率 (Filling Ratio, FR) 与覆盖率 (Coverage Ratio, CR) 衡量推荐算法在缓解冷启动问题上的效果. 假设用户-服务信任矩阵稀疏度为  $x$ , 用户数为  $m$ , 服务数为  $n$ , 则信任值的已知数量  $a = mn(1-x)$ , 未知数量为  $mn-a$ . 假设经过预测后由未知转换成已知的信任值数量为  $b$ , FR 和 CR 的计算如式 (17)、(18).

$$FR = \frac{b}{mn-a} \quad (17)$$

$$CR = \frac{a+b}{mn} \quad (18)$$

FR 表示在所有未知的信任值中可预测到的信任值比率. CR 表示最终可获取的用户-服务信任值比率. FR、CR 的值越大, 代表推荐算法在缓解冷启动和矩阵稀疏问题上的效果越好.

#### 4.2.3 推荐准确率

为了衡量推荐的排序结果采用评价指标: 平均准确率均值 (Mean Average Precision, MAP)、归一化折损累积增益 (Normalized Discounted Cumulative Gain, NDCG).

$$MAP(N) = \frac{\sum_{i=1}^N p(i)}{N} \quad (19)$$

$N$  表示推荐序列取前  $N$  个服务.  $p(i)$  是与理想排序相比, 排在第  $i$  位的准确率.

NDCG 值表示推荐列表排序与理想排序之间的比例. NDCG 值越大, 代表推荐质量越好. 当 NDCG 值为 1 时, 表示推荐列表是理想排序.

$$NDCG(N) = \frac{DCG(N)}{IDCG(N)} \quad (20)$$

$$DCG(N) = r_1 + \sum_{i=2}^N \frac{r_i}{\log_2 i} \quad (21)$$

$N$  表示取前  $N$  个服务.  $r_i$  是对排在第  $i$  位的评分. IDCG ( $N$ ) 是对理想排序计算 DCG ( $N$ ).

### 4.3 实验结果分析

#### 4.3.1 MAE、RMSE 分析

从候选用户集中随机地选取 10 个观察用户, 在初

始  $T$  矩阵的不同稀疏度下, 应用 CTV 算法预测每个用户对候选服务的信任值, 并计算 MAE、RMSE. 比较 CTV 算法、与其它三种预测方法的 MAE、RMSE 值, 如图 4 和图 5 所示.

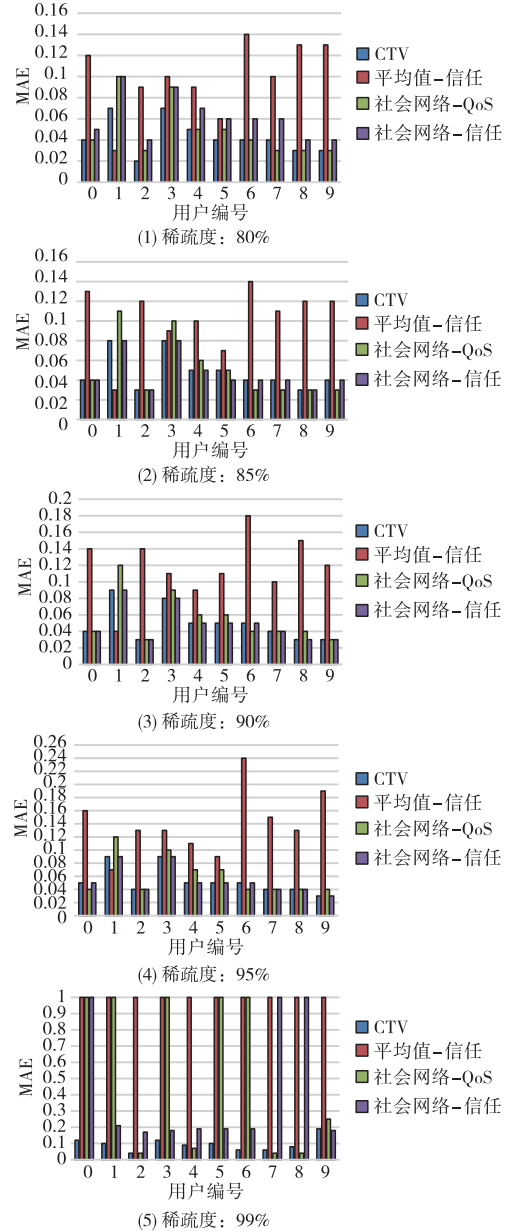


图4 不同矩阵稀疏度下各预测方法的MAE值

这三种方法分别为: (1) 平均值-信任, 根据  $T$  矩阵计算用户对已知服务的信任值均值, 优点是效率高, 缺点是无法提供个性化推荐; (2) 社会网络-QoS, 初始矩阵是用户-服务 QoS 矩阵, 遍历网络寻找与  $U$  相似的用户并预测, 优点是提高了预测 QoS 值的可靠性, 缺点是没有综合考虑用户对 QoS 指标的个性化偏好; (3) 社会网络-信任, 搜索基于信任的社会网络, 应用 CF 预测信任值, 优点是反映了用户的个性化需求, 缺点是用户间



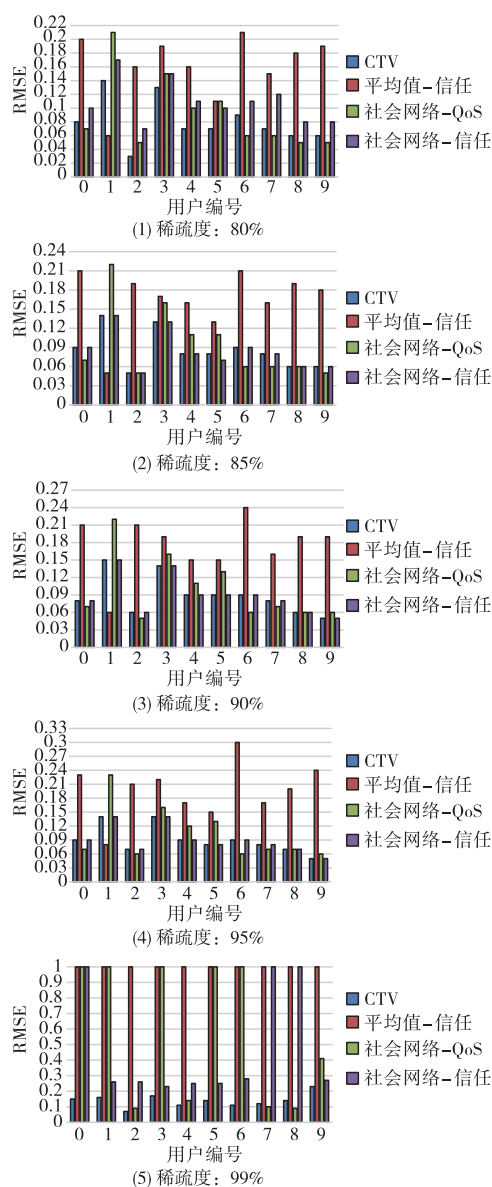


图5 不同矩阵稀疏度下各预测方法的RMSE值

信任值是领域无关的,导致预测准确率不高.以上三种方法都没有应用 WSP.

从图 4 和图 5 可看出,CTV 算法的 MAE、RMSE 值基本低于其它三种方法.当稀疏度为 80% 时,CTV 算法每个用户的 MAE 值都不大于 0.07,基本小于另三种预测方法.由于观察用户选取的随机性,编号为 1 的用户对候选服务信任观察值的取值集中于平均值附近,因而 MAE、RMSE 值在采用平均值-信任方法时较小.平均值-信任方法在其他 9 个观察用户上的 MAE、RMSE 值大于或等于其它三种方法.当稀疏度大于 95% 时,四种方法的预测准确率都有所降低.当用户对每个候选服务的信任值都无法预测时,将 MAE、RMSE 值标示为 1.当稀疏度为 99% 时会出现此情形,如图 4(5)和图 5(5)

所示.但 CTV 算法对候选服务的信任值还是可预测的,因而其可预测性较高.

#### 4.3.2 FR、CR 分析

通过 FR 和 CR 分析 RQ 算法在解决冷启动和矩阵稀疏问题上的优越性.

用户-用户信任矩阵  $U_T$  的计算依据是  $T$  矩阵,因而  $T$  的稀疏度决定了  $U_T$  的稀疏度,从而影响 RQ 算法的 FR、CR 值.  $T$  矩阵的稀疏度取值范围是  $\{80\%, 85\%, 90\%, 95\%, 99\%\}$ ,对应的  $U_T$  的稀疏度是  $\{1.00\%, 5.45\%, 29.72\%, 73.12\%, 98.51\%\}$ .当  $T$  的稀疏度为 99% 时,  $U_T$  为 98.51%,对信任值预测是极大的挑战.

为了验证 HSN 中融入 WSP 的有效性,分析 RQ 算法的 FR 和 CR 时,比较两种方法:有 WSP,即 RQ 算法;没有 WSP,即基于信任社会网络的推荐.如表 1 所示.

随着  $T$  稀疏度的增加,FR 和 CR 递减.当  $T$  的稀疏度为 99% 时,RQ 的 FR、CR 值远高于无 WSP 时的预测结果,可得出 RQ 推荐方法在缓解冷启动问题时的优越性.

表 1  $T$  在不同稀疏度下推荐的 FR 和 CR,分为:无 WSP 和有 WSP

$T$ 的 稀疏度	FR (无 WSP)	CR (无 WSP)	FR (有 WSP)	CR (有 WSP)
80%	97.32%	97.83%	98.93%	99.13%
85%	97.48%	97.83%	98.57%	98.77%
90%	97.62%	97.83%	98.57%	98.70%
95%	96.24%	96.38%	96.47%	96.59%
99%	31.82%	32.32%	41.17%	41.59%

#### 4.3.3 MAP、NDCG 分析

为了判断推荐准确率,比较各种推荐方法的 MAP、NDCG 值.对 10 个用户的 MAP、NDCG 值分别取均值、最大值和最小值,如表 2 和表 3 所示.表 2 和表 3 中的 Null 表示该方法无法推荐服务,由于矩阵稀疏性高,没有已知的直接信任值,以及由此导致的服务信任值的不可预测.取值为 0.0 表示,推荐的服务序列与理想序列没有一个服务是匹配的.

从表 2 和表 3 可看出,随着稀疏度的增加,MAP、NDCG 值减小.可获取的信息越少,为用户提供准确推荐的挑战越大.在稀疏度为 99% 时,MAP、NDCG 值比稀疏度低的情形小.根据 4.3.2 节得出的结论,RQ 推荐的 FR、CR 较高,可预测的服务数量多,因而它的 MAP、NDCG 值稍大于其它方法.

与其它方法相比,社会网络-QoS 推荐的 MAP、NDCG 平均值较小的原因在于,理想排序是根据信任值大小排列,它所推荐的服务序列是根据 QoS 预测值排序,与信任值相差较远,因而 MAP、NDCG 值较小.

表 2 各推荐方法在不同矩阵稀疏度下 MAP 值

推荐方法 稀疏度	RQ			平均值-信任			社会网络-QoS			社会网络-信任		
	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值
80%	0.66	0.87	0.3	0.58	1.0	0.0	0.4	1.0	0.0	0.61	0.98	0.23
85%	0.60	0.9	0.25	0.43	0.9	0.0	0.37	1.0	0.0	0.62	0.93	0.25
90%	0.58	0.97	0.25	0.28	0.85	0.0	0.58	1.0	0.0	0.60	0.97	0.25
95%	0.60	0.82	0.26	0.16	0.8	0.0	0.356	1.0	0.0	0.60	0.82	0.26
99%	0.27	0.62	0.0	Null	Null	Null	0.17	1.0	0.0	0.24	0.48	0.0

表 3 各推荐方法在不同矩阵稀疏度下 NDCG 值

推荐方法 稀疏度	RQ			平均值-信任			社会网络-QoS			社会网络-信任		
	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值	平均值	最大值	最小值
80%	0.43	0.62	0.15	0.16	0.48	0.0	0.06	0.29	0.0	0.39	0.61	0.1
85%	0.42	0.62	0.15	0.12	0.3	0.0	0.03	0.11	0.0	0.43	0.63	0.15
90%	0.42	0.63	0.14	0.09	0.27	0.0	0.02	0.11	0.0	0.42	0.63	0.14
95%	0.42	0.62	0.15	0.03	0.3	0.0	0.07	0.18	0.0	0.42	0.62	0.15
99%	0.19	0.42	0.0	Null	Null	Null	0.03	0.12	0.0	0.12	0.39	0.0

#### 4.3.4 TOP-N 对推荐准确率的影响

为用户推荐 TOP-N 的服务序列,分析  $N$  的取值对 RQ 算法推荐准确率的影响. 当取值分别为 5、10 和 15 时,不同矩阵稀疏度下,MAP、NDCG 值如图 6 所示.

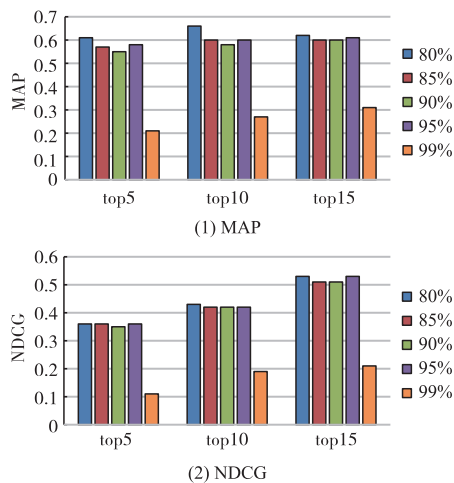


图6 TOP-N对推荐结果的影响

随着  $N$  的增大,特别是在高稀疏度的情形下,MAP 和 NDCG 值有增长的趋势. 表明  $N$  取值越大,推荐准确率越高.

## 5 结论

由于信任能反映用户个性化选择服务的特点,与 HSN 相结合可以缓解推荐中常见的冷启动及矩阵稀疏问题. 本文建立了基于 HSN 的 Web 服务推荐框架,提出了用户-服务信任值预测 CTV 算法及服务推荐 RQ 算

法,并计算预测准确率 MAE、RMSE 及推荐准确率 MAP、NDCG. 在 HSN 中融入了 WSP,以改善冷启动、矩阵稀疏性问题,并通过 FR、CR 衡量其优越性.

还存在值得继续深入研究的内容,构建社会网络和 Web 服务社会网络,以社区结构为中心,为用户推荐服务;分析 WSP 之间的博弈、共享等关系并应用到推荐中;提供者是普遍存在的,将本文提出的方法扩展到一般物品的推荐研究中.

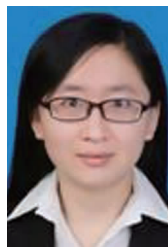
## 参考文献

- [1] CHENG B, LI C, ZHAO S, et al. Semantics mining & indexing-based rapid web services discovery framework [J/OL]. IEEE Transactions on Services Computing, 2018. DOI: 10.1109/TSC.2018.2831678
- [2] CHEN F, LU C, WU H, et al. A semantic similarity measure integrating multiple conceptual relationships for web service discovery [J]. Expert Systems with Applications, 2017, 67: 19–31.
- [3] 马文龙, 朱李楠, 王万良. 云制造环境下基于 QoS 感知的云服务选择模型 [J]. 计算机集成制造系统, 2014, 20(5): 1246–1254.  
MA Wen-long, ZHU Li-nan, WANG Wan-liang. Cloud service selection model based on qos-aware in cloud manufacturing environment [J]. Computer Integrated Manufacturing Systems, 2014, 20(5): 1246–1254. (in Chinese)
- [4] TANG M, XU Y, LIU J, et al. Trust-aware service recommendation via exploiting social networks [A]. Proceedings of IEEE International Conference on Services Computing



- [C]. Washington, DC, USA: IEEE Computer Society, 2013. 376 – 383.
- [5] DENG S, HUANG L, XU G. Social network-based service recommendation with trust enhancement[J]. Expert Systems with Applications, 2014, 41(18): 8075 – 8084.
- [6] QIAN X, FENG H, ZHAO G, et al. Personalized recommendation combining user interest and social circle[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(7): 1763 – 1777.
- [7] JIANG M, CUI P, LIU R, et al. Social contextual recommendation[A]. Proceedings of the 21st ACM International Conference on Information and Knowledge Management [C]. New York: ACM, 2012. 45 – 54.
- [8] CHANG W L, DIAZ A N, HUNG P C K. Estimating trust value: A social network perspective[J]. Information Systems Frontiers, 2015, 17(6): 1381 – 1400.
- [9] JIANG W, WANG G, BHUIYAN M Z A, et al. Understanding graph-based trust evaluation in online social networks[J]. ACM Computing Surveys, 2016, 49(1): 1 – 35.
- [10] 俞春花, 刘学军, 李斌, 等. 基于上下文相似度和社交网络的移动服务推荐方法[J]. 电子学报, 2017, 45(6): 1530 – 1536.
- YU Chun-hua, LIU Xue-jun, LI Bin, et al. Mobile service recommendation based on context similarity and social network[J]. Acta Electronica Sinica, 2017, 45(6): 1530 – 1536.
- [11] KALAÏ A, ZAYANI C A, AMOUS I, et al. Social collaborative service recommendation approach based on user's trust and domain-specific expertise[J]. Future Generation Computer Systems, 2018, 80: 355 – 367.
- [12] SU K, XIAO B, LIU B, et al. TAP: a personalized trust-aware QoS prediction approach for web service recommendation[J]. Knowledge-Based Systems, 2017, 115: 55 – 65.
- [13] 杨震, 杨甜甜, 范科峰, 等. 基于信任合成的云服务动态组合机制研究[J]. 电子学报, 2018, 46(3): 614 – 620.
- YANG Zhen, YANG Tian-tian, FAN Ke-feng, et al. Cloud service composition based on trust combination[J]. Acta Electronica Sinica, 2018, 46(3): 614 – 620.
- [14] 束束, 梁昌勇, 徐健. 基于信任的云服务系统多目标任务分配模型[J]. 计算机研究与发展, 2018, 55(6): 1167 – 1179.
- SHU Jian, LIANG Chan-yong, XU Jian. Trust-based multi-objectives task assignment model in cloud service system[J]. Journal of Computer Research and Development, 2018, 55(6): 1167 – 1179.
- [15] MEI L J, CHAN W K, TSE T H. An adaptive service selection approach to service composition[A]. Proceedings of IEEE International Conference on Web Services[C]. Washington, DC, USA: IEEE Computer Society, 2008. 70 – 77.
- [16] SUN Y. A new method to identify collaborative partners in social service provider networks[J]. Information Systems Frontiers, 2015, 18(3): 565 – 578.
- [17] NEWMAN M E J. Models of the small world[J]. Journal of Statistical Physics, 2000, 101(2-4): 819 – 841.
- [18] ZHENG Z B, ZHANG Y L, LYU M R. Exploring latent features for memory-based qos prediction in cloud computing[A]. Proceedings of the 30th IEEE Symposium on Reliable Distributed Systems[C]. Washington, DC, USA: IEEE Computer Society, 2011. 1 – 10.
- [19] SÁNCHEZ-MORENO D, GONZÁLEZ A B G, VICENTE M D M, et al. A collaborative filtering method for music recommendation using playing coefficients for artists and users[J]. Expert Systems with Applications, 2016, 66: 234 – 244.

#### 作者简介



**杨 洁(通信作者)** 女, 1986 年 12 月出生于山西平遥. 2014 年毕业于南京大学管理科学与工程专业, 获管理学博士学位. 现为杭州电子科技大学信息管理系讲师, 在读博士后. 主要研究方向为个性化推荐、Web 服务、云计算等.  
E-mail: yangjie.19861225@163.com



**朱咸军** 男, 1977 年 1 月出生于江苏盐城. 2016 年毕业于南京大学管理科学与工程专业, 获管理学博士学位. 现为金陵科技学院软件工程学院讲师, 在读博士后. 主要研究方向为协同计算、软件工程、大数据、智能信息处理与智能系统等.  
E-mail: mymailzxj@126.com



**周献中** 男, 1962 年 11 月出生于江苏泰兴. 现为南京大学工程管理学院控制科学与工程学科教授, 管理科学与工程学科博士生导师. 主要研究领域为人-信息-物理-系统(HCPS)理论与技术、智能系统与智能工程(人件技术、智能人机交互).  
E-mail: zhouxz@nju.edu.cn



**柳 毅** 男, 1978 年 7 月出生于辽宁沈阳. 2007 年毕业于上海理工大学管理科学与工程专业, 获管理学博士学位. 现为杭州电子科技大学信息管理系副教授, 硕士生导师. 主要研究方向为智能优化算法、信息管理与商务智能等.  
E-mail: liuyi@hdu.edu.cn