基于服务关注点相似度推荐的信任访问控制模型*

李园园，张永胜，吴明峰

（山东师范大学 信息科学与工程学院，山东省分布式计算机软件新技术重点实验室，济南 250014）

摘　要： 针对服务计算环境下用户对其所使用服务的评分，依据其服务关注点的不同而不同，即使是同一个服务，不同用户的评价标准也不一样，推荐者的选取不仅与其所处环境上下文有关，还与推荐者对服务的关注点有关。为了使用户推荐更加可靠、有效，提出基于服务关注点相似度的推荐算法。该算法解决了用户盲目搜索推荐者的问题，使用聚类算法生成用户聚类簇，根据用户间的相似度在聚类簇内进行推荐者的搜索，既提高了推荐的可靠性，又提高了搜索的效率。实验显示，此算法比传统算法在推荐准确性与推荐搜索效率上存在明显优势。

关键词： 服务关注点；推荐；信任；聚类；相似度

中图分类号： TP393　　**文献标志码：** A　　**文章编号：** 1001-3695(2012)12-4644-04

doi:10.3969/j.issn.1001-3695.2012.12.061

Trust access control model based on service attention spot similarity recommendation

LI Yuan-yuan, ZHANG Yong-sheng, WU Ming-feng

(*Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, School of Information Science & Engineering, Shandong Normal University, Jinan* 250014, *China*)

Abstract: In the service oriented computing, the evaluation that the user for the service which its uses is different from the difference of service spots. Even if the identical service, the different user's evaluation criteria is dissimilar, the selection of recommender is related to not only context but also service spots. In order to enable recommendation more reliable and effective, this paper put forward recommendation algorithm based on service attention spot similarity. In order to solve the problem of bind searching, it generated the user clusters with clustering algorithm. It searched recommenders based on similarity between users in the cluster class. It not only improved the reliability of the recommendation, but also improved the efficiency of the searching. The experiments show that this algorithm has obvious advantages compared with traditional algorithm on the recommendation accuracy and searching efficiency.

Key words: service attention spot; recommendation; trust; cluster; similarity

0　引言

自 2001 年 IBM 提出将动态电子商务的理念转向 Web services，各种应用程序可以在不同的运行平台上进行交流整合，服务计算时代就已在不知不觉中来临了。2002 年 6 月，张良杰先生筹办的以 Web Services Computing 为名的学术专题研讨会首次将服务与计算联系在一起，向“服务计算”迈出了第一步。2003 年 11 月，在 IEEE 的推动下，把服务的概念进一步拓展，正式确认服务和计算联系在一起的一门新学科诞生。

近年来，服务计算在工业界得到了广泛的应用，与此同时，像欺骗、恶意推荐等安全威胁也日益严重。信任机制是解决这些问题的有效手段，能为解决服务计算环境中的安全问题提供新思路。

信任机制根据相关的信任信息与策略作出信任决策，信任信息来源于两方面：a）主体（服务提供者）与客体（服务请求者）间的直接信任；b）来自于其他节点的推荐信任。在服务计算环境下，如果节点之间没有直接交互或者直接交互的次数较少，不能根据主观判断了解对方的实际情况，那么，来自其他节点的推荐信任至关重要，推荐信任值的大小就决定了双方是否交互。因此，对于推荐信任的研究对信任机制来讲意义重大。

1　相关工作

近年来，很多专家学者运用不同的理论和方法对信任问题进行了研究，取得了阶段性的成果。

Blaze 等人[1]提出了信任管理的概念，指出开放系统中的安全信息并不完整，需要可信赖的第三方提供推荐信息来完善，以保障决策安全。Jøsnag 等人[2]认为信任是主观的、模糊的、不确定的，基于主观逻辑的信任评估模型利用证据空间和观念空间来度量信任关系，在信任关系中引入主观逻辑，利用主观逻辑运算子进行信任度的推导和计算。但是该模型没有

收稿日期： 2012-05-18；**修回日期：** 2012-06-27　　**基金项目：** 山东省自然科学基金资助项目（ZR2011FM019）；山东省研究生教育创新计划资助项目（2011No. 292）

作者简介： 李园园（1986-），女，硕士研究生，主要研究方向为访问控制安全模型（lizzy.01@163.com）；张永胜（1962-），男，教授，主要研究方向为软件工程环境、Internet/Intranet 工程、网络信息安全；吴明峰（1985-），男，硕士研究生，主要研究方向为 Web 服务安全.

区分直接信任与推荐信任,无法有效抑制恶意推荐的影响。沈昌祥等人[3]给出了关于可信计算的相关概念。李建欣等人[4]提出了一种面向网络的动态信任管理框架 DTM,该模型将信念引入信任,通过正则事件序列刻画信任建立及资源授权过程。Wang 等人[5]使用贝叶斯网络推理来提高推荐信任的可靠性,但该方法需要依赖高度可靠的专家经验,主观性较强,不能广泛地应用于服务计算环境。文献[6~8]提出了一些全局的协同推荐机制,其中假设网络中存在一些专门的服务中心来收集和整理各个用户对某一服务的反馈信息;通过一定的方法对反馈信息进行整合;对待选服务进行评估,选出评价优秀的服务作为服务提供者。然而服务计算环境中没有专门的集中认证中心收集各方面的信息来验证推荐信息的可靠性。Billhardt 等人[9]提出面向分布式用户的协同推荐机制,通过节点之间的信任关系组成信任网络,进而进行信息挖掘,将挖掘的信息进一步整合后加权计算。潘静等人[10]通过相关因子量化推荐信任关系,从而得到信任可传递空间,然后应用信任子网分割算法、信任迭代计算来确定声誉高的推荐者。但其缺点在于没有将对同一服务的各个方面的推荐进行分类,可能导致推荐信息的可用性不高。

针对目前信任问题的研究现状及存在的相关问题,本文对推荐信任问题进行研究,主要解决以下三个问题:

a)推荐信息的准确性。服务计算环境面向网络的所有用户,网络环境高度开放,开放的网络固然提供了更多的机会使用户发布更多的反馈信息,但同时,推荐信息受制于推荐者的主观偏好、合作态度,恶意的推荐者可能误导决策方向[10]。开放的网络环境允许网络中的实体随时加入和离开网络,如服务的发布和撤销、用户的注册与注销都会影响推荐信息的准确性,进而影响信任评估的准确度。

b)搜索推荐信息的效率。开放的网络环境存在相当数量的推荐信息,怎样在较短的时间内获得更有效的推荐信息至关重要。

c)相似推荐信息的选择。例如,同一个系统提供的服务,系统开发者 U_1 关注的是系统服务的吞吐量,普通的系统用户 U_2 只在乎服务的响应速度;对于同一个服务,不同的用户给出不同的评价,在用户间进行推荐时,可能会对服务给出较低评价;当用户 U_3 的服务关注点与 U_1 的相同却采纳了 U_2 的意见,则会对该服务的服务能力产生误解,造成服务的错选。因此,在选取推荐信息时,应将不同用户对该服务的关注点考虑在内,以提高推荐的可靠性。

2　理论基础

2.1　协同过滤原理

协同过滤是一种从可利用的信息源中提取形式化信息的过程,推荐信息的取得依据是用户间的相似度。

基于协同过滤的推荐原理,如图 1 所示。用户 A、B、C 对服务 s_1、s_2、s_3 有一定的兴趣,阴影区域表示三个用户对 s_1、s_2、s_3 都作出过评价,可以根据这个推断出他们间的相似度。若他们兴趣相似,则把服务 s_4、s_5 推荐给用户 B 也是合理的,因为用户 A 和 C 都喜欢服务 s_4、s_5。

2.2　传统的基于用户的协同过滤推荐算法

传统协同过滤算法的核心是一个评分矩阵,该矩阵中的行代表用户,列代表项目。例如,矩阵 $R_{m\times n}$ 表示系统中有 m 个用户、n 个项目;$R_{i,j}$ 表示用户 i 对项目 j 的评价,取值通常为 1~5,若用户还未使用该项目或者尚未对该项目评分,则取值为 0。

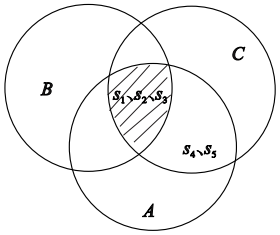

图1　协同过滤原理图

协同过滤算法中的关键步骤是目标用户邻居的形成,主要是通过衡量用户之间的相似程度,找出与目标用户具有相似爱好的最近邻居集合。相似度的度量方法主要有余弦相似性和 Pearson 相关系数[11]。

1)余弦相似性

将用户对项目的评分看做 n 维空间上的向量,通过向量的夹角余弦来度量用户间的相似性。例如,假设用户 u 和用户 v 在 n 维项目空间上的评分向量分别为 $u=(R_{u,1},R_{u,2},\cdots,R_{u,n})$ 和 $v=(R_{v,1},R_{v,2},\cdots,R_{v,n})$,用户间的相似性 $\mathrm{sim}(u,v)$ 可按式(1)计算:

$$\mathrm{sim}(u,v)=\cos(u,v)=\frac{u\cdot v}{\|u\|\cdot\|v\|} \tag{1}$$

其中:$u\cdot v$ 表示用户向量 u 和 v 的内积,$\|u\|$ 和 $\|v\|$ 分别是二者的模。用户向量间的夹角越小,即 $\mathrm{sim}(u,v)$ 的值越大,说明二者的相似度越高,反之二者的相似度越低。

2)Pearson 相关系数

在用户共同评分项目的基础上度量用户间的相似度,设 $I(u)\cap I(v)$ 为用户 u 和用户 v 共同评分的项目集合,二者间的相似性为

$$\mathrm{sim}(u,v)=\frac{\sum_{i\in I(u)\cap I(v)}(R_{u,i}-\overline{R_u})(R_{V,i}-\overline{R_V})}{\sqrt{\sum_{i\in I(u)\cap I(v)}(R_{u,i}-\overline{R_u})^2}\sqrt{\sum_{i\in I(u)\cap I(v)}(R_{v,i}-\overline{R_v})^2}} \tag{2}$$

其中:$R_{u,i}$、$R_{v,i}$ 分别表示用户 u 和用户 v 对项目 i 的评分值,$\overline{R_u}$、$\overline{R_v}$ 分别表示用户 u 和 v 的评分均值。大量实验显示,Pearson 相关系数在衡量用户或项目间相似度的问题上有很大的优势[12]。

3　基于服务关注点相似度的协同过滤推荐算法

3.1　数据模型

定义 1　使用向量模型[13]表示由 m 个服务关注点(如吞吐量、响应时间、延迟等)组成的属性集合 $\mathrm{Point}=\{P_1,P_2,\cdots,P_i,\cdots,P_m\}$。

定义 2　“用户—服务—关注点”三维模型记做 U-S-P 模型,它是一个三维的向量空间(user,service,point),三个维度分别是用户、服务、关注点。每个维度分别用各自的属性值组成的向量来表示[14]。

如图 2 所示,三维坐标中,横坐标 U 标注用户,纵坐标 S 标注服务,竖坐标 P 标注某一服务的关注点。$R(U_1,S_1,P_1)$ 为用户 U_1 对所用服务 S_1 的关注点 P_1 的评分,$R\in[1,5]$。

3.2　算法设计思想

本文研究的算法,初始数据需要用户输入对所调用服务的

某一方面的评价。然而,用户额外输入的要求给用户带来了不必要的麻烦,可能会让用户产生不满。本文需要用户输入 25 个对服务评分值,这对于大多数用户是可以接受的。随着用户访问服务数目不断增多,捕获到的用户对某一服务及其对于该服务的关注点的推荐信息也越来越准确。

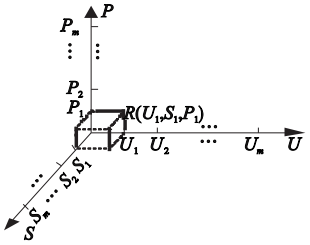

图2 数据模型

根据文献[15]中提到的上下文预过滤范式(在生成推荐结果之前,利用当前上下文过滤掉无关的用户偏好数据,从而构建当前上下文信息相关的推荐数据集)进行改进。在选取推荐者之前,考虑当前用户调用哪种服务,关注的是该服务的哪一具体属性,即服务的关注点。利用这些信息过滤掉与用户兴趣无关的数据,从而构建服务属性集。然后对筛选后的数据进行评分预测,并生成满足当前条件的推荐结果,其核心思想是将多维推荐问题转换为传统二维推荐问题求解。

在用户共同评分项目的基础上度量用户间的相似度,设 $S(u)\cap S(v)$ 为用户 u 和用户 v 共同评分的项目集合,两者间的相似性为

$$\mathrm{sim}(u,v)=\frac{\sum_{s\in S(u)\cap S(v)}(R_{u,s,p_1}-\overline{R_{u,s}})(R_{v,s,p_2}-\overline{R_{v,s}})}{\sqrt{\sum_{s\in S(u)\cap S(v)}(R_{u,s,p_1}-\overline{R_{u,s}})^2}\sqrt{\sum_{s\in S(u)\cap S(v)}(R_{v,s,p_2}-\overline{R_{v,s}})^2}} \quad (3)$$

式(3)是对传统协同过滤算法中 Pearson 系数的改进,适用于三维空间。其中:s 表示用户 u 和 v 的共同评分的服务集合,R_{u,s,p_1} 表示用户 u 对服务 s 在服务关注点 p_1 下的评分值,$\overline{R_{u,s}}$ 表示用户 u 在服务 s 下的平均评分值,R_{v,s,p_2} 表示用户 v 对服务 s 在服务关注点 p_2 下的评分值,$\overline{R_{v,s}}$ 表示用户 v 在服务 s 下的平均评分值。

首先根据聚类规则将系统中的用户分成 $U(u\in 1\sim n)$ 个用户群,根据类内对每个用户的评价信息,将用户群的中心表示为 $U_i=(U_1,U_2,\cdots,U_n)$;然后根据式(3)相似度量方法为目标用户找到最相似的用户群;在用户群内再根据相似度量找出目标用户在这一群内的最近邻居。

基于聚类的协同过滤与传统的协同过滤相比多了用户或者服务聚类的过程,即先判断目标用户属于哪一类,再根据类内的邻居为目标用户产生推荐,属于一种局部最优的协同过滤。

3.3 算法过程描述

本节在上述研究的基础上,研究该算法的详细流程。系统中使用该算法进行定期更新。

输入:"用户—服务—关注点"三维模型;用户 U、服务 S、服务的某一关注点 P。

输出:服务 S 在当前关注点 P 下的相似度最大的 TOP-K 个用户的集合。

a)将三维空间内的点中任意选取 k 个对象,以这 k 个对象的评分向量作为初始的 k 个聚类中心,记为 $U_{\mathrm{Center}}=\{U_{C_1},U_{C_2},\cdots,U_{C_k}\}$

b)初始化用户聚类簇 $C=\{C_1,C_2,\cdots,C_k\}$

c) for each user $U_i\in U$ do

//将 U_i 加入到最相似的聚类簇中

d) for each cluster center $U_{C_j}\in U_{\mathrm{Center}}$ do

e) 利用式(3)计算 U_i 与 U_{C_j} 的相似性 $\mathrm{sim}(U_i,U_{C_j})$;

f) end for

g) 选择相似性最大的

$\mathrm{sim}(U_i,U_{C_j})=\max\{\mathrm{sim}(U_i,U_{C_1}),\mathrm{sim}(U_i,U_{C_2}),\cdots,\mathrm{sim}(U_i,U_{C_k})\}$;

h) if(U_i 只与一个聚类中心最相似) then

聚类簇 $C_m=C_m\cup U_i$

//当某用户处于一个聚类边缘时

i) else(U_i 与多个聚类中心最相似) then

j) 计算该用户对这几个聚类的信任度,从而进行聚类;

k) end if

l) end for

//调整聚类中心

m) for each cluster $C_i\in C$ do

n) for all user $U_j\in C_i$ do

o) 计算聚类簇 C_i 内所有用户对项目的平均值,生成新的聚类中心 U_{C_i}

p) end for

q) end for

r) 计算错误函数

s) until 不再明显地改变或者聚类的成员不再变化;

根据上述算法,可将离散的用户生成 k 个用户聚类簇,每个簇内有相应的簇中心,簇内成员节点的相似度最高,而簇间节点的相似度最低。

3.4 信任评估流程图

服务计算环境下,服务在服务注册中心注册,服务注册中心的服务在服务处理器中利用聚类算法根据其相似度将服务进行聚类,生成相应的聚类簇。用户请求服务,根据请求的条件找到相应的聚类簇,在聚类簇内搜索,计算信任度,进而实施访问决策。流程如图 3 所示。

4 模拟实验结果及其分析

通过模拟实验来比较此算法在以下两个方面的性能:

a)对推荐准确度的影响;

b)对推荐效率的影响。

准确性的评价标准是均方差(使用本文提出的算法对服务的评分与实际评分间的均方差)。推荐算法效率的评价指标是信任计算开销[16]。

实验运行环境为 Pentium® Duan-Core 2.1 GHz CPU,2 GB 内存,250 GB 硬盘和 Windows 7 旗舰版操作系统,实验数据分析软件环境为 MATALB 7.11。

4.1 算法准确性实验

将本文提出的算法分别与传统的余弦相似性(cosine)和 Pearson 相关系数方法进行比较,以均方差(MSE)进行度量,通过本文算法得到的 TOP-K 个推荐集的评分为 $f_1,f_2,\cdots,f_k$,相应的实际评分为 $a_1,a_2,\cdots,a_k$,均方差为

$$\mathrm{MSE}=\frac{1}{k}\sum_{i=1}^{k}(f_i-a_i) \quad (4)$$

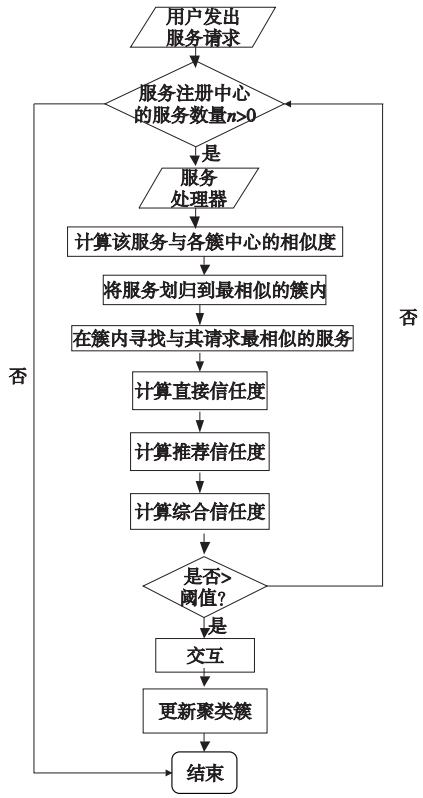

图3　信任评估流程

如图 4 所示:本文提出的改进算法与余弦相似性和 Pearson 相关系数相比推荐准确率高;但随着簇内推荐者数量的增多,三种算法的 MSE 降低的速率趋于缓和,这是由于随着推荐者的增多,原本可信度不高的推荐者也参与了推荐,在一定程度上影响了推荐的准确性。当推荐者过多,对推荐反而产生一定的负面影响,因而将邻居控制在 40 左右比较合适,此时 MSE 值最小。

4.2　算法效率实验

影响信任计算开销的主要因素是参与请求者直接信任与推荐信任计算的用户,即直接交互者与间接推荐者的搜索。算法的思想决定了参与信任计算所要查找的节点,查找节点的多少与信任计算开销成正比,信任计算开销用完成一次交互所需的时间来体现。因一次交互的时间较短不容易获得,因此这里取 50 次交互的平均时间作为完成一次交互所需的平均时间。

如图 5 所示,随着网络节点个数的增多,使用本文的算法完成一次交互的平均时间明显比基于 Bayesian 信任计算算法要短,效率高。这是由于本文算法使用推荐者二级搜索,首先寻找与其相似的簇头,以定位相似簇,然后在簇内寻找最近邻居,比盲目搜索具有一定优势,效率较高。

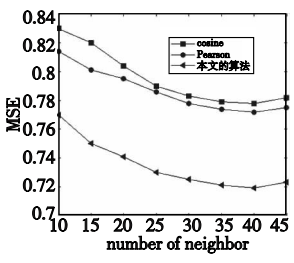

图4　三种推荐算法的MSE比较

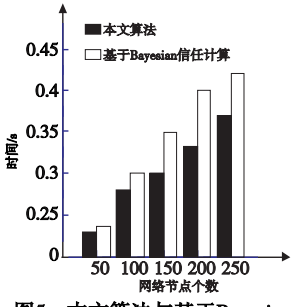

图5　本文算法与基于Bayesian信任计算的效率比较

5　结束语

本文提出的基于服务关注点相似度推荐的信任访问控制模型将与信任相关的服务关注点因素考虑在内,改进了 Pearson 相关系数,使其运用于三维空间向量;使用聚类算法将相似度较高的用户聚类成簇,在簇内再一次利用 Pearson 相关系数寻找最近邻居,解决了信任推荐中存在的一些问题。仿真实验表明,该模型在推荐准确率和搜索效率上具有优势。

参考文献:

[1] BLAZE M, FEIGENBAUM J, LACY J. Decentralized trust management[C]//Proc of the 17th Symposium on Security and Privacy. [S. l.]: IEEE Computer Society Press, 1996: 164-173.

[2] J∅SANG A. A subjective metric of authentication[C]//Proc of the 5th European Symposium on Research in Computer Security. London: Springer-Verlag, 1998: 329-444.

[3] 沈昌祥,张焕国,冯登国,等. 信息安全综述[J]. 中国科学, 2007, 37(2): 129-150.

[4] 李建欣,怀进鹏,李先贤. DTM:一种面向网络计算的动态信任管理模型[J]. 计算机学报,2009,32(3): 493-505.

[5] WANG Yao, VASSILEVA J. Bayesian network-based trust model [C]//Proc of IEEE Computer Society WIC International Conference on Web Intelligence. Washington DC: IEEE Computer Society, 2003: 372-378.

[6] VU L H, HAUSWIRTH M, ABERER K. QoS-based service selection and ranking with trust and reputation management [C]//Proc of OTM. 2005: 466-483.

[7] ALI A S, LUDWIG S A, RANA O F. A cognitive trust-based approach for Web service discovery and selection[C]//Proc of the 3rd European Conference on Web Service. 2005: 38-49.

[8] DAY J, DETERS R. Selecting the best Web service[C]//Proc of the 14th Annual IBM Centers for Advanced Studies Conference. 2004: 293-307.

[9] BILLHARDT H, HERMOSO R, OSSOWSKI S, *et al*. Trust-based service provider selection in open environments [C]//Proc of the 22nd Annual ACM Symposium on Applied Computing. New York: ACM Press, 2007: 1375-1380.

[10] 潘静,徐锋,吕建. 面向可心服务选取的基于声誉的推荐者发现方法[J]. 软件学报,2010,21(2): 388-400.

[11] HERLOCKEEY J L, KONSTAN J A, BORCHERS A, *et al*. An algorithmic framework for performing collaborative filtering[C]//Proc of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 1999: 230-237.

[12] WANG Jun, DEVRIES A P, REINDERS M J T. Unifying user-based and item-based collaborative filtering approaches by similarity fusion [C]//Proc of the 29th Annual International ACM SIGIIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2006: 501-508.

[13] 李蕊,李仁发. 上下文感知计算及系统框架综述[J]. 计算机研究与发展,2007,44(2): 269-276.

[14] 徐凤苓,孟祥武,王立才. 基于移动用户上下文相似度的协同过滤推荐算法[J]. 电子与信息学报,2011,33 (11): 2785-2789.

[15] ADOMAVICIUS G, TUZHILIN A. Context-aware recommender systems(book chapter) [M]. London: Springer Press, 2011: 217-253.

[16] 田俊峰,鲁玉臻,李宁. 基于推荐的信任链管理模型[J]. 通信学报,2011,32(10): 1-9.