

# 移动互联网基于用户网络结构的信息过滤机制

王景欣<sup>1</sup> 袁 坚<sup>1</sup> 耿军伟<sup>2</sup> 冯振明<sup>1</sup>

(清华大学电子工程系 北京 100084)<sup>1</sup> (北京市电力公司输电公司 北京 100068)<sup>2</sup>

**摘 要** 针对提高移动互联网信息过滤机制准确率及用户冷启动问题,提出了一种基于用户网络结构的信息过滤机制。该机制将用户网络结构作为信息过滤机制的一部分,帮助机制识别用户需求,实现准确信息过滤。具体包含:利用用户网络社交结构限制信息传播途径,提供协同过滤中用户相似性度量新维度;利用基于资源流行度的用户网络兴趣结构修正协同过滤中的用户相似性度量;利用用户网络兴趣簇指纹解决冷启动问题。数据分析表明该机制具备较好的性能。该工作对移动互联网中用户网络结构的应用问题研究具有启发作用。

**关键词** 协同过滤,网络结构,兴趣簇,资源流行度

## Information Filtering Mechanism Based on User Network Architecture for Mobile Internet

WANG Jing-xin<sup>1</sup> YUAN Jian<sup>1</sup> GENG Jun-wei<sup>2</sup> FENG Zhen-ming<sup>1</sup>

(Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)<sup>1</sup>

(Power Transmission Company, Beijing Power, Beijing 100068, China)<sup>2</sup>

**Abstract** To improve the accuracy of information filtering mechanism and resolve the cold start problem, an information filtering mechanism based on user network architecture was proposed. Here, the social network is treated as a part of the information sharing and dissemination system, like information processing nodes, to help us recognize mobile users' needs and identify more interesting services. The user interest network structure is used to improve the collaborative filtering. The social network is used to limit the transmission of information. The fingerprints are used to solve the cold start problem. Data analysis shows that this mechanism has better performance.

**Keywords** Collaborative filtering, Network structure, Interest cluster, Resource popularity

## 1 引言

信息过载是信息时代信息极大丰富的负面影响之一,由于用户处理信息的时间和能力有限,各种媒介提供的信息量远远超过了个人或系统所能接收、处理或有效利用的范围。大量冗余信息严重干扰了用户对相关有用信息的准确选择。用户在接受移动互联网服务时,迫切需要高效、及时、准确、个性化的信息和服务,因此,信息过滤机制在移动互联网中尤为重要。信息过滤机制即通过对用户需求与信息流进行匹配计算,从中抽取符合用户个性化需求的信息并将其传送给用户的系统化方法<sup>[1]</sup>。

已有信息过滤机制利用用户自身兴趣、用户间兴趣相似度以及相似用户评价进行信息过滤,这是将用户兴趣网络应用于信息过滤的初步尝试。但是,已有的过滤机制中存在如下 3 个问题有待进一步探讨:首先,鲜见用户网络社交结构的相关应用,相对于传统互联网,移动互联网的终端与用户绑定且随用户移动,这种特征为通过移动终端观测用户社交网络并使用其进行信息过滤提供了可能。其次,已有的协同过滤方法中所使用的用户兴趣网络未考虑资源流行度的因素,因此,其过滤性能存在提高的空间。再次,过滤机制的冷启动问题未得到妥善解决。

针对以上问题,本文提出一种基于用户网络结构的信息过滤机制:将基于资源流行度的用户网络兴趣结构、兴趣簇指纹和用户网络社交结构应用于信息过滤机制当中,提高信息过滤的准确性,解决过滤机制的冷启动问题。相对于已有协同过滤机制,其核心在于:利用用户网络社交结构限制信息传播途径,提供用户相似性度量的新维度;利用基于资源流行度的用户网络兴趣结构修正协同过滤中的用户相似性度量;利用用户网络兴趣簇指纹解决冷启动问题。通过实际数据分析证明,该机制有效地提高了协同过滤机制的推荐准确度。

本文第 2 节介绍相关工作;第 3 节分析可用的移动互联网用户网络结构;第 4 节基于用户网络社交结构、用户评价和兴趣簇结构,构建信息过滤机制;第 5 节通过数值分析证明其有效性;最后总结本文的工作。

## 2 相关工作

信息过滤机制研究用户的需求模板,并将其与海量信息进行匹配,持续为用户提供其感兴趣或具有潜在兴趣的信息。信息过滤机制分为内容过滤、协同过滤两种。本节给出以上两种过滤机制的相关研究进展。

基于内容的过滤,通过学习用户的历史兴趣来进行相关的预测。通过为用户构建兴趣模型来进行内容过滤,常应用

本文受 973 国家重点基础研究发展计划(2007CB307100,2007CB307105)资助。

王景欣(1982-),女,博士生,主要研究方向为移动互联网、社交网络、网络安全,E-mail:cindy1980s@gmail.com。

于新闻信息推送系统中。其主要缺点在于,由于仅仅只基于用户自身的历史兴趣,因此它只能推荐与用户自身历史兴趣有关的内容,其系统的灵活性差,当大量新型内容加入时,无法进行有效的推荐,因此该方法有局限性,不适合信息多变的大型动态系统。基于内容推荐的相关系统有 Letizia<sup>[2]</sup>、Syskill&Webert<sup>[3]</sup>、CiteSeer<sup>[4]</sup>、Web Personalizer<sup>[5]</sup> 和 if-Web<sup>[6]</sup>等。

协作过滤(Collaborative Filtering),又叫协同过滤,其核心思想为人类的兴趣可以通过分析与其具备相似行为的用户进行预测<sup>[1]</sup>。其基本思路为:首先利用系统内用户的评价过滤低质量的信息,然后利用用户间的相似性寻找相似用户群,最后,针对某一特定用户,利用与当前用户兴趣相似性高的用户群的偏好值预测该用户的兴趣。与基于内容过滤相比,协作过滤共享了其它用户的评价和经验,解决了内容过滤中针对项目的局限性,能够处理非结构化的推荐对象,并发掘用户尚未发现的新兴趣。

协作过滤自从1992年由Goldberg等<sup>[7]</sup>首次提出并应用在邮件推荐系统“Tapestry”中后,在个性化推荐系统中得到广泛的应用,如音乐推荐系统——Ringo<sup>[8]</sup>、新闻推荐系统——Grouplens<sup>[9]</sup>、网站推荐系统——PHOAKS<sup>[10]</sup>、电影推荐系统——Movielens<sup>[11]</sup>、笑话推荐系统——Jester<sup>[12]</sup>等。协作过滤更是被大多数的电子商务平台采用,如Amazon、当当、eBay、淘宝等。商家根据用户的购买习惯和用户间相似性分析,向用户推荐其可能感兴趣的货品,提高电子商务的销售能力。

虽然协同过滤的理论研究和实际应用都取得了良好的成果,但是随着用户数量的不断增多,协同过滤技术面临着过滤准确性、数据稀疏性、冷启动、超高维、安全性等问题。尤其是过滤准确性和冷启动问题,已成为国内外的研究热点<sup>[1]</sup>。其中,冷启动问题分为用户冷启动和项目冷启动两种。用户冷启动指当有新的用户加入时,用户的历史记录有限,系统难以分析得到其相似用户群,从而难以进行推荐。项目冷启动指系统中的项目从未被任何用户评价过时,由于系统过滤依赖用户对项目的评价进行推荐,因此系统将不会向任何用户推荐该项目。

本文的信息过滤机制即为协同过滤机制的一种,通过引入用户网络社交结构和兴趣结构,结合社会学相关知识,致力于提高信息过滤机制准确率并解决用户冷启动问题。

### 3 移动互联网用户网络结构分析

本节分析移动互联网用户网络结构特征与信息过滤机制的耦合点,为下面的机制设计做铺垫。

协同过滤的关键是默认具有相同或相似兴趣偏好的用户相互信任彼此的选择。这种信任机制是协同过滤的基础。先前研究的信任机制大多基于用户操作历史的相似性,如用户对电影、音乐、新闻的评分数据相似度,用户购买物品的相似度等。

但是,人的潜在的计算能力不仅仅包含用户对于数据、信息等的处理能力,也包含用户的各种社会关系网络。随着复杂网络及社会学相关研究的深入,用户网络中的某些结构化特征被挖掘出来,使其存在应用于用户相似性度量的可能,于是,可在用户间建立新的信任机制,寻找新的相似性高的用户

群,提高信息推荐的准确度。本文中,将用户网络结构分为社交结构和兴趣结构两种,分别将其应用于解决协同过滤中的相似性度量 and 用户冷启动问题。

#### 3.1 移动互联网用户网络社交结构

本文的移动互联网用户网络社交结构特指的是用户由于相遇而构成的用户社交网络结构。

举例说明基于用户移动的社交网络的形成过程<sup>[13]</sup>。在一个7个陌生人组成的派对中,起初大家彼此都不认识。派对开始后,客人们自发组成小组交谈。起初的小组之间是相互分隔的,只有同组的用户间产生社交连接(用实线表示),如图1(a)所示。随着派对的进行,有两个客人更换了小组,这两位客人又与新的组中的用户产生社交连接(虚线表示),如图1(b)所示。此时,有的客人虽然尚未交谈,但是,利用已有的社交连接,可以将网络中的任意两个客人联系起来,客人们之间形成了连通的社交网络,这就是一个由于用户移动而产生社交网络的全过程。用户不断地与新的用户相遇,产生社交连接,时间的推移使得网络中越来越多的用户彼此产生连接,一个庞大的用户社交网络就此形成。

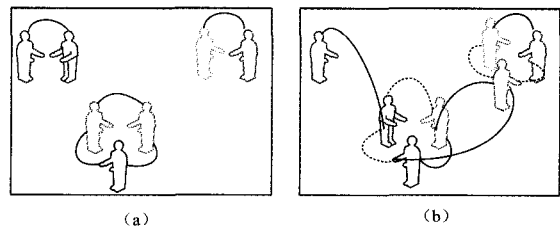


图1 基于用户移动的社交网络形成过程

本文提供一个计算用户相似度的新维度,即用户社交网络。该维度默认在该社交网络中联系越紧密,其信息需求也越相似。该默认规则基于以下实际观测:

1)处于同一个地理位置的用户可能有共同的信息需求,如具备地域有效性的信息(附近的餐馆、加油站、停车场等),95%的智能手机用户曾搜索过与地理位置相关的本地信息<sup>[14]</sup>。

2)社交关系紧密的用户存在同化现象。用户在社会交往中往往伴随着信息传递的需求。

3)社交关系紧密的用户拥有相似的信息需求深度。如“防火墙”一词的相关信息,普通消费者关注其性能和价格,而同一个课题组的研究者则倾向于其内部机制及相关设计等更深的信息。

基于以上分析,有理由相信,当信息沿着社交网络的方向进行优先传递时,更易抵达适当的用户群。

于此同时,移动互联网终端随用户移动,为寻找用户的社交网络提供了很好的基础。而且,移动互联网终端普遍搭载的短距离通信技术为用户社交网络的检测提供条件,如蓝牙技术、红外数据传输、超宽带技术和射频识别技术等<sup>[15]</sup>,对用户的地理距离有强依赖性,只能检测到一定距离范围内的设备。

因此,移动互联网的信息过滤机制中有必要也有条件加入此社交网络结构维度,来提高信息过滤的准确率。

#### 3.2 移动互联网用户网络兴趣结构

本文的移动互联网用户网络兴趣结构特指作者在文献<sup>[16]</sup>中提出的基于资源流行度的用户网络兴趣结构。

用户兴趣网络由于用户下载过相同的资源而创建。在用户兴趣网络中,以网络节点为用户,如果两个用户下载过同一个资源,则该两个用户间生成一条边<sup>[17]</sup>。利用文献[16]中的方法对用户兴趣网络进行分层聚簇,此处的用户兴趣簇等价于协同过滤算法中的相似用户群。

只考虑低流行度资源创建的用户兴趣网络即为基于资源流行度的用户兴趣网络<sup>[16]</sup>。作者在文献[16]中论证了基于低流行度资源形成的用户网络具备更加明晰的用户集群特性:分簇特征更为明显,且各簇内用户兴趣趋同性增强,不同簇间用户兴趣趋向差异增大,用户分簇准确性提高。因此,加入资源流行度参数的用户相似性度量将会更加准确地衡量用户间的相似性,即,基于资源流行度的用户网络兴趣结构应用于信息过滤机制会为其带来准确性的提高。

更进一步,用户兴趣簇特征的标示——簇指纹,可以准确地描述各个兴趣簇特征<sup>[16]</sup>。经过降维处理后的低维簇指纹规模小,准确率高,有潜力应用于协同过滤算法中解决冷启动问题。

定义 1(低维簇指纹  $\lambda$ ) 系统中簇内高流行度资源数目为  $n_s$ ,簇  $i$  的簇指纹  $\lambda_i$  为  $n_s$  维向量:

$$\lambda_i = \{t_{ij}, j \in [1, n_s]\}$$

式中,

$$t_{ij} = \begin{cases} 0, & \text{簇 } i \text{ 内无用户下载过资源 } j \\ 1, & \text{簇 } i \text{ 内有用户下载过资源 } j \end{cases} \quad (1)$$

#### 4 基于用户网络结构的信息过滤机制

本文将用户网络社交结构和兴趣簇结构引入信息协同过滤机制中来,提出一种用于移动互联网的新型的基于用户网络结构的信息过滤机制。该过滤机制由 3 个部分组成,如图 2 所示。

(1)基于用户网络社交结构的信息过滤:用户信息传递通过基于用户网络社交结构的机会网络,从信息传递的途径对信息进行第一遍过滤。

(2)基于用户评价的协同信息过滤:当新的信息到来时,用户的信息转发代理利用最近邻的评价值做出第二遍协同信息过滤,并基于资源流行度计算该信息对用户相似度的影响。用户在阅读该信息时对该信息做出评价,只有通过用户审核后的信息才能在网络中分发,这是信息的第三遍过滤。

(3)基于用户网络兴趣簇结构的冷启动解决方案:新用户加入时,根据簇指纹判断加入的初始簇,随后,新用户继承该簇的评价数据并将其作为自身最近邻初始值,相关信息存入用户的数据库中。

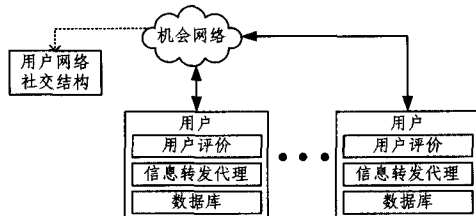


图 2 基于用户网络结构的信息过滤机制

##### 4.1 基于用户网络社交结构的信息过滤

该部分的信息过滤由用户社交网络形成的机会网络负责。基于用户网络社交结构的信息过滤指的是限定信息沿着用户社交网络分发。本文限制只有当用户相遇时,才能进行

信息的交换。随着用户的不断移动,信息将会在一定时间内在整个系统中传播。网络中的任何一个用户都有一定的概率得到该信息,在社交网络中与信息源用户联系紧密的用户将优先得到信息。基于社交网络中用户相对移动进行信息分发的网络又称为机会网络。

定义 2(机会网络)<sup>[18]</sup> 机会网络是一种利用节点移动所形成的相遇实现通信的自组织网络。

机会网络中源节点和目标节点间无需存在完整链路,节点采用“存储-等待-转发”的模式进行消息的分发。

##### 4.2 基于用户评价的协同信息过滤

基于用户评价的协同信息过滤分为两部分:用户评价和协同信息过滤,分别由用户自身和用户的信息转发代理完成。

用户评价由用户在阅读消息时完成,分为显式评分和隐式评分两类。显式评分指用户显式输入对信息的数值评分,用户参与度较大。隐式评分则不需要用户直接评分,而是系统根据用户浏览信息时的行为特征对评分进行预测,如停顿时间、删除、收藏等。

协同信息过滤由用户的信息转发代理负责。用户的信息转发代理通过计算用户之间的相似度,并对特定信息的评分进行预测,来决定是否将此消息过滤。此过程无需用户参与。本文只考虑低流行度资源对用户相似度的影响,并采用 Pearson 相关系数衡量用户之间的相似程度,该系数便于计算,且可得到与余弦相似度相似的结果<sup>[19]</sup>:

$$\text{sim}(u, v) = \frac{\sum_{i \in I_{u,v}} (R_{u,i} - \bar{R}_u)(R_{v,i} - \bar{R}_v)}{\sqrt{\sum_{i \in I_{u,v}} (R_{u,i} - \bar{R}_u)^2} \sqrt{\sum_{i \in I_{u,v}} (R_{v,i} - \bar{R}_v)^2}} \quad (2)$$

式中,  $I_{u,v}$  代表用户  $u$  和用户  $v$  共同相关且低于资源流行度阈值的消息集合,  $R_{u,i}$  代表用户  $u$  对消息  $i$  的评价数值,  $\text{sim}(u, v)$  则表示用户  $u$  和用户  $v$  的相似度。

根据用户间的相似度可得用户  $u$  的最近邻集合  $U$ , 则针对特定消息  $i$ , 预测用户  $u$  的评价数值为:

$$P_{u,i} = \bar{R}_{u,I} + \alpha \times \frac{\sum_{u' \in U} (R_{u',i} - \bar{R}_{u',I}) \times \text{sim}(u, u')}{\sum_{u' \in U} |\text{sim}(u, u')|} \quad (3)$$

式中,  $\bar{R}_{u,I}$  为用户  $u$  对  $I$  类信息的平均评价数值,  $\text{sim}(u, u')$  为用户  $u$  和用户  $u'$  的相似度,  $R_{u',i}$  为用户  $u'$  对消息  $i$  的评价数值,  $\alpha$  为历史权重。

依照式(3)预测得  $P_{u,i}$ , 信息转发代理对照系统阈值设置决定是否将此消息过滤。只有通过信息转发代理过滤的消息才会呈献给用户。从而实现协同信息过滤。

##### 4.3 基于用户网络兴趣结构的冷启动解决方案

基于用户网络兴趣结构的冷启动解决方案的基本流程如下:

1)用户网络分层聚簇。通过搜集系统内用户对信息的评价数据,形成用户兴趣网络。如果用户评价过相同的数据,则用户间存在一条边,用户的相似度即为该边的度值。在用户兴趣网络的基础上,利用分层聚簇方法,将用户分为若干兴趣簇。

2)生成簇指纹。参照式(1),为每一个簇构建低维簇指纹,簇指纹由用户共同评价的数据和其评价数值组成。

3)当新用户加入时,首先根据其对簇指纹的判断选择初始簇属,赋予该簇的相关评价数据。其次,降低新用户信息转发代理的信息过滤阈值,使得新用户加入系统初期有机会

收到消息,给用户信息转发代理提供学习新用户兴趣趋向的条件。

## 5 信息过滤机制性能评估

本节利用实际数据集对本文提出的基于用户网络结构的信息过滤机制进行性能评估。针对本信息过滤机制的3个组成部分,性能评估分3个层面进行:

- (1)网络社交结构信息过滤的有效性;
- (2)基于资源流行度修正用户相似度的有效性;
- (3)簇指纹解决冷启动问题的有效性。

### 5.1 真实数据集

本文的数据来源于基于 BitTorrent (BT) 协议的北邮人 BT 对等网络共享系统。包含从 2009 年 10 月 20 日到 2009 年 10 月 31 日的用户下载记录,共 81531 条,涉及用户 10368 个,涉及资源 7376 个。用户下载记录的格式如表 1 所列。其中用户 ID 用 MD5 码表示。每条记录表示来自某 IP 的某用户在某时间下载了某资源。

表 1 数据集用户下载记录格式

用户 ID (只显示 MD5 码后 20 位)	资源 ID	下载时间 (2009-10-20)	IP 地址
238d1902b62e99aec02b	36084	00:39:14	59.66.65.154
857acc19df08d9cb2982	33064	00:43:17	59.66.25.58
2b5b83a7eac3ba951e1f	36119	00:44:25	59.66.23.14
139393bbe134fe24f495	4385	00:45:37	59.66.64.2
076d50ad58c5615b1e89	3693	00:46:14	59.66.15.46

由于移动互联网用户群和传统互联网用户群有极大的重合性,因此,虽然该数据搜集于传统互联网用户,该分析结论同样适用于移动互联网用户。

### 5.2 网络社交结构信息过滤的有效性

此处的社交网络为用户相遇产生的机会网络。但是受限于数据集,我们做出以下假设:IP 地址相近的用户为经常相遇的用户。该假设一定程度上反映了用户网络社交结构,其下的分析结果有一定的借鉴意义。

实验设计如下:针对每一个用户,选取 100 个 IP 相近的用户组成社交用户群,同时随机选取 100 个用户组成随机用户群。针对该用户,参照式(2)计算两个用户群中的平均用户相似度。哪一种方法的平均用户相似度高,则说明该方法所得的相似用户群准确,消息过滤效果好。

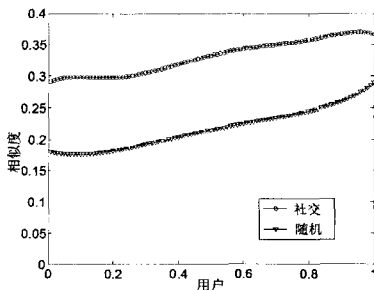


图 3 基于社交结构的相似用户群内平均用户相似度

图 3 给出两种方法下的用户相似度对比(为了方便比较,对原始数据进行了曲线拟合)。为了便于观察,将  $x$  轴中的用户依照活跃度倒序排列,并均分成 100 等份, $x$  轴的一个点代表网络中 1% 的用户。可见,社交用户群要好于随机用户群。说明,基于网络社交结构进行信息过滤具备一定的效果。同时也可看到,越不活跃的用户所得的用户群性能越好,这是由

于活跃度高的用户兴趣广泛,与其它用户差异相对较大。

### 5.3 基于资源流行度修正用户相似度的有效性

此处对比加入资源流行度前后,相似用户群计算结果的准确性差异,以此来衡量基于资源流行度修正用户相似度的有效性。

实验设计如下:针对每一个用户,对比原始资源生成的相似用户群和剔除网络中流行度占前 10% 的资源后生成的相似用户群,参照式(2)分别计算相似用户群中的平均用户相似度。哪一种方法的平均用户相似度高,则说明该方法所得的相似用户群准确,效果好。

图 4 给出两种方法下的用户相似度对比(为了方便比较,对原始数据进行了曲线拟合)。为了便于观察,将  $x$  轴中的用户依照活跃度倒序排列,并均分成 100 等份, $x$  轴的一个点代表网络中 1% 的用户。可见,剔除热门资源后生成的相似用户群要好于原始资源生成的相似用户群。说明,基于资源流行度修正用户相似度的效果较好。同时可以看到,越不活跃的用户所得的相似用户群性能越好,理由同上一小节所述,即活跃度高的用户兴趣广泛,与其它用户差异相对较大。

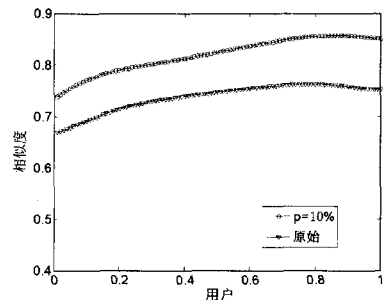


图 4 基于资源流行度的相似用户群内平均用户相似度

### 5.4 簇指纹解决冷启动问题的有效性

当使用簇指纹预测所得的用户初始簇属与用户的实际簇属相吻合时,说明该方法成功将用户归入合适的簇。即,簇指纹解决冷启动问题的有效性的指标为根据簇指纹成功预测用户簇属的比重。

实验设计如下:首先提取基于资源流行度的用户网络,并对其进行分簇,得到每个用户的实际簇属。然后针对每一个用户,提取基于资源流行度的用户网络,并对其进行分簇。随后参照式(1)构造低维簇指纹。参照该簇指纹预测该用户的初始簇属。对比预测所得的用户初始簇属与用户的实际簇属。

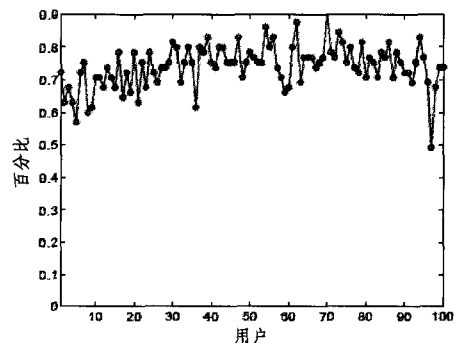


图 5 使用簇指纹预测用户簇属成功率

图 5 给出使用簇指纹预测用户簇属的成功率。为了便于观察,将  $x$  轴中的用户依照活跃度倒序排列,并均分成 100 等份, $x$  轴的一个点代表网络中 1% 的用户。可见,针对大多数

用户的簇属预测成功率在70%~80%,即,使用簇指纹在一定程度上解决了协同过滤中的冷启动问题,效果较好。

**结束语** 针对移动互联网信息过量的问题,本文提出了一种基于用户网络结构的信息过滤机制。该机制引入基于资源流行度的用户网络兴趣簇结构、簇指纹和用户网络社交结构,为协同过滤中的用户相似性度量提供新的维度,为信息过滤机制中的冷启动问题提供新的解决思路。数据分析显示,该机制对信息过滤的准确度有一定程度的提高。

该机制将用户网络结构作为信息过滤机制的一个信息过滤节点,协助识别用户的信息需求以提供更恰当的服务。可以看到,此时的用户已经不仅仅是服务的消费者和使用者,同时也是服务内容的提供者。用户的行为、计算能力和社会关系被开发以提高服务质量。用户网络结构作为信息过滤机制的一部分,帮助整合各种资源,精确地提取信息,识别移动互联网用户的需求并提供更多有趣的服务。不同于将机器网络和用户网络区分对待,而是研究它们之间的交互作用,这是一项有意义的尝试。

### 参 考 文 献

- [1] Su Xiao-yuan, Taghi M K. A survey of collaborative filtering Techniques[J]. *Advances in Artificial Intelligene*, 2009, 9:1-20
- [2] Lieberman H. Letizia: An Agent That Assists Web Browsing [C]//*Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence(IJCAI-95)*. San Mateo,CA; Morgan Kaufmann publishers Inc,1995:924-929
- [3] Michael J P, Jack M, Daniel B. Syskill&Webert: Identifying interesting Web sites[C]//*Proceedings of the 13th National Conference on Artificial Intelligence*. Portland,OR,US,1996:54-61
- [4] Giles C L, Bollacker K D, Lawrence S. Citeseer: An automatic citation indexing system[C]//*Proceeding of the 3rd ACM Conference on Digital Libraries*. New York; ACM Press,1998:89-98
- [5] Mobasher B, Cooley R, Jaideep S. Automatic personalization based on Web usage mining[J]. *Communications of the ACM*, 2000,43(8):142-151
- [6] Asnicar F, Tasso C. if Web: a prototype of user model-based intelligent agent for document filtering and navigation in the world wide Web[C]//*Proceeding of the 6th International Conference on User Modelling*. Chia Laguna, Sardinia, Italy, 1997: 3-11

- [7] David G, David N, Oki Brian M, et al. Using collaborative filtering to weave an information tapestry [J]. *Communication ACM*, 1992,35(12):61-70
- [8] Shardan U, Maes P. Social information filtering: Algorithms for automating "word of mouth"[C]//*Proceedings of the SIGCHI conference on Human factors in computing systems*. New York, USA, 1995:210-217
- [9] Joseph A K, Bradley N M, David M, et al. Grouplens: applying collaborative filtering to usenet news[J]. *Communications of the ACM*, 1997,40(3):77-87
- [10] Loren T, Will H, Brian A, et al. Phoaks: A system for sharing recommendations[J]. *Communications of the ACM*, 1997, 40(3):59-62
- [11] Dahlen B J, Konstan J A, Herlocker J L, et al. Jump-starting movielens: User benefits of starting a collaborative filtering system with "dead data"[R]. TR 98-017 Minnesota. USA; University of Minnesota, 1998
- [12] Goldberg K, Roeder T, Gupta D, et al. Eigentaste: A constant time collaborative filtering algorithm[J]. *Information Retrieval*, 2001,4:133-151
- [13] Barabasi A L. Linked: how everything is connected to everything else and what it means for business, science, and everyday life [M]. Plume, 2003
- [14] 艾瑞咨询. 中国搜索引擎行业年度监测报告[R]. 上海: 艾瑞咨询, 2011
- [15] Puers R, Thone J. Bio-Medical CMOS ICs [M]. Springer US, 2011:219-277
- [16] 王景欣, 王钺, 李一鹏, 等. 基于用户网络结构的信息协同过滤机制[J]. *物理学报*. 已录用
- [17] Li Yi-peng, Ren Yong, Yuan Jian, et al. Network statistical analysis in peer to peer application[J]. *Acta Physica Sinica*, 2011, 60(5):050000
- [18] 熊永平, 孙利民, 牛建伟, 等. 机会网络[J]. *软件学报*, 2009, 21(1):124-137
- [19] Badrul M S, Joseph A K, Borchers A l, et al. Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system [C] // *Proceedings of Computer supported cooperative work*. New York, NY, USA; ACM, 1998: 345-354

(上接第127页)

- [5] 王玮, 刘晓洁, 李涛, 等. 一种异地灾难恢复系统的设计与实现[J]. *计算机应用研究*, 2007, 24(9)
- [6] 朱立平, 李战怀, 温宗臣. RCDP: 基于冗余的连续数据保护架构[J]. *微处理机*, 2009, 30(4)
- [7] Maohua L, Chiueh T-C. File versioning for block-level continuous data protection[Z]. Montreal, QC, Canada; Institute of Electrical and Electronics Engineers Inc, 2009
- [8] 牟虹燕, 李战怀, 王泽林. 远程数据复制系统的性能优化策略[J]. *微处理机*, 2009, 30(3)
- [9] Flouris M D, Bilas A. Clotho: Transparent Data Versioning at the Block I/O Level [C] // *NASA/IEEE Conference on Mass Storage Systems and Technologies*. CollegePark, USA, 2004
- [10] Ta-Shma P, et al. Virtual machine time travel using continuous data protection and checkpointing [C] // *Broadway, 17th Floor*.

- [11] Wang Li-na, et al. Building Hot Snapshot Copy based on Windows File System[J]. *Wuhan University Journal of Natural Sciences*, 2006, 11(6):1503-1506
- [12] 李旭, 谢长生, 杨靖, 等. 一种改进的块级连续数据保护机制[J]. *计算机研究与发展*, 2009(05)
- [13] Sheng Y, et al. An efficient consistent point inserting method for block level continuous data protection [J]. *Gaojishu Tongxin/Chinese High Technology Letters*, 20(11):1101-1107
- [14] Sheng Y, et al. TH-CDP: An efficient block level continuous data protection system [Z]. Zhang Jia-jie, Hunan, China; IEEE Computer Society, 2009
- [15] 佟强, 陈珉, 王云. 基于 Mobile Agent 的分布式空间数据服务模式[J]. *计算机工程*, 2004, 30(1)