维普资讯 http://www.cqvip.com



# 基于 Web 社会网络的个性化 Web 信息推荐模型

陈 君 唐 雁
(西南大学计算机与信息科学学院 重庆 400715)

**摘 要** 随着Web信息量的快速增长,个性化的Web信息推荐系统扮演着越来越重要的角色。目前,大多数Web信息推荐系统存在着个性化程度不高,对用户历史数据依赖性高,系统不具备开放性,用户偏好"走样"概率高的问题。针对这四个方面的问题,本文提出了一种新的、采用语义Web技术、基于Web社会网络的个性化Web信息推荐模型,详细分析了用户偏好的获取,Web社会网络的生成以及待过滤Web信息的采集。并且,利用从实际中采集的真实数据进行了实验,证明了模型的可行性和有效性。

**关键词** 个性化,Web信息推荐,资源描述框架,Web社会网络

## A Personal Web Recommendation Model Based on Web Social Network

CHEN Jun TANG Yan
(Faculty of Computer and Information Science, Southwest China University, Chongqing 400715)

**Abstract** As fast growing of Web information, personal Web recommendation system would play more and more important role. Currently, however, most such recommendation systems mainly suffer from four problems. They separately are lacking personality, depending too much on users' history data, performing poor in decentralized environment, and sacrificing users' interest profiles. In this paper, we propose a new personal Web recommendation model based on Web socialnetwork, using semantic Web technology. User' s interest capturing, Web social network formation, and filtering-needed Web information collection are thoroughly studied. Model' s feasibility and effectiveness are proved by experiments.

**Keywords** Personalization, Web recommendation, Resource description framework, Web social network

## 1 引言

目前,Web已成为人们获取信息的一个重要途径。如何设计个性化的Web信息推荐系统高效地向用户推荐符合其兴趣偏好的Web信息已成为一个重要的研究课题。

实现个性化的Web推荐,需要解决三个问题:1)用户偏好的获取;2)待过滤的Web信息采集与预处理;3)匹配用户偏好与Web信息。目前,大多数Web推荐服务由用户浏览的站点提供。它们通过挖掘用户的历史数据获取用户偏好,比如挖掘用户的浏览日志;通过存储站点本身提供的Web信息实现待过滤信息的采集,并对这些信息进行预处理;然后对用户偏好与已分类的Web信息进行匹配计算,生成推荐结果。

这样的Web信息推荐系统存在着四个方面的问题。1)个性化程度不高。由于站点的访问量较大,系统通常对站点中的用户历史记录进行聚类,将用户分成几大类,并对同属一类的用户生成相同的偏好描述。因此,无法有效地获取和表示用户个性化的偏好。2)对用户历史数据依赖性高。无法为新用户或浏览日志数据稀疏的用户进行有效的推荐。3)系统不具备开放性。在采集待过滤Web信息时,系统不考虑用户偏好,直接抓取站点本身提供的Web信息,将用户的Web信息浏览体验局限在了特定的站点中。4)用户偏好"走样"的概率高。通常,"盲目"采集得到的待过滤Web信息集信息量大且内容繁杂。由于语言的二义性,很难对大量的Web信息进行有效的过滤,导致匹配计算时对用户偏好的"走样"。

针对目前Web推荐系统中存在的问题,本文提出了一种在用户所在的客户端进行偏好获取、基于Web社会网络的个性化Web信息推荐模型。该模型采用了语义Web技术,因此,本文第2节将简要介绍相关的语义Web技术和Web社会网络;第3节介绍相关工作;第4节给出基于Web社会网络的个性化Web信息推荐模型,并详细分析用户偏好的获取,Web社会网络的生成以及待过滤Web信息的采集。最后,利用采集自Web的真实数据进行实验,证明模型的可行性与有效性。

## 2 语义 Web(Semantic Web)

语义Web[7]的概念由Web创始人Tim Berners-Lee提出。语义Web以传统Web为基础。它的最终目标就是要使各类信息在Web这样一个动态性极强的开放式环境中能被无歧义地表示和交换,实现Web信息的机器可读,从而为更智能、更复杂的Web开放式应用创造条件。知识的表示是语义Web的基础,也是目前语义Web研究的重点之一。语义Web中的知识表示不仅要解决数据本身的表示,还要提供一种方法或技术对数据的推理规则进行描述,从而使机器能真正理解数据所隐含的语义。

### 2.1 资源描述框架(Resource Description Framework)

RDF[4]是基于XML的资源描述框架,采用主语(subject)、谓语(predicate)和宾语(object)的三元组形式描述声明(statement)。其中,主语、谓语、宾语甚至声明都可以用URI(Uniform Resource Identifier)唯一定位。常用的URI有URL(Uniform Resource Locator)、用户Email地址的MD5码等,因此RDF不仅能采用XML对数据本身进行描述,还

维普资讯 http://www.cqvip.com

能利用三元组的形式描述数据推理所依赖的规则。因此，在语义 Web 中，可以用 RDF 对本体(Ontology)进行描述。

目前，Web 中最普遍的本体就是 FOAF(Friend-Of-A-Friend)。FOAF 项目由 Dan Brickley 和 Libby Miller 发起，主要研究如何采用 RDF 描述机器可读的用户信息。图 1 显示了一个典型的 FOAF 文件片段。

```
<rdf:RDF
    Xmlns: rdf = "http//www. w3. org/1999/02/22-rdf-syntax-ns
    #"
    xmlns:foaf="http//xmlns. com/foaf/0. 1/">
<foaf:Person>
  <foaf:name>Danny</foaf:name>
  <foaf:mbox rdf:resource="mailto:danny. ayers@gmail. com"/>
  ...
  <foaf:Knows>
    <foaf:Person>
      <foaf:name>Captsolo</foaf:name>
      <foaf:weblog rdf:resource="http://captsolo. net/info"/>
    </foaf:Person>
    <foaf:Person>
      <foaf:name>Jay</foaf:name>
      <foaf:weblog rdf:resource="http://icite. net/blog"/>
      <foaf: homepage rdf: resource = " http://icite. net/about/
      people. html"/>
    </foaf:Persom>
    ...
  </foaf:Knows>
  ...
</foaf:Person>
</rdf:RDF>
```

图 1

图 1 所示的 FOAF 采用 FOAF0. 1 版本[1]。该版本包含 Person、name、mbox、mbox-sha1sum、Knows、weblog、homepage 等用户信息描述词汇。图 1 文件描述了这样一些信息：①用户 Danny 的 Email 地址是 danny. ayers@gmail. com；②用户 Danny 认识两个用户，他们是 CaptSolo 和 Jay；③用户 CaptSolo 的 Weblog 地址是 http://captsolo. net/info；④用户 Jay 的 Weblog 地址是 http://icite. net/blog；⑤用户 Jay 的主页地址是 http://icite. net/about/people. html。

FOAF 文件可以由用户自己编写，也可以利用工具半自动或自动地生成 FOAF 文件。目前，已经有大量的工具可用，比如 Foafgen[2]，FOAF-a-Matic；Wordpress 或 MoveTable 等主流 Weblog 编辑工具的 FOAF 自动生成插件(Plug-in)。

### 2.2 Web 社会网络(Web Social Network)

现实生活中，人与人之间的联系能形成社会网络，在 Web 上，也可以利用用户间的关系形成每个用户对应的 Web 个人社会网络。多个不同用户的 Web 个人社会网络整合在一起可以形成 Web 社会网络。

近年，在 Web 上涌现了大量的社会网络服务[10](Social Network Service)，使 Web 用户可发布自己的个人信息；获取其他用户的个人信息；利用多个用户的个人信息形成 Web 社会网络；并通过所形成的 Web 社会网络进行诸如书签(Bookmark)、Weblog 等信息的交换或共享。Friendster[3]是社会网络服务最早的提供商之一，在 2003 年开始提供服务的短短几个月间就吸引了五百万之多的注册用户；随后 Google、Microsoft 也推出了相似的社会网络服务。但对用户而言，这类集中式的社会网络服务存在许多不足，比如：①用户不能完全控制自己的信息的存储与使用；②Web 社会网络中只包括在同一社会服务提供平台中进行了注册的用户。

语义 Web 技术的采用，使用户可以获得更具个性化的 Web 社会网络服务。用户可以用 FOAF 文件描述个人信息，并通过 Weblog、个人主页或论坛在开放式的环境中发布 FOAF 文件。由于采用 RDF 描述的 FOAF 文件具有机器可读性，其他用户的个人信息可以轻易地被获取，并整合为 Web 社会网络。在生成的 Web 社会网络中可以交换或共享 Web 上存在的各类信息，而不仅仅局限于交换某一个或几个信息提供商发布的信息。

## 3 相关工作

个性化 Web 信息推荐一直是研究的热点，随着语义 Web 的发展，掀起了基于语义 Web 技术的个性化 Web 信息推荐的研究热潮。研究主要集中在两个方面：1)利用本体优化 Web 信息分类，如文[11]；2)采用 RDF 描述用户偏好，提高匹配计算精度，如文[12]。这些研究所提出的模型能够优化 Web 信息推荐，但主要是针对本文 1 中所列举的一个或两个问题。

目前，国际上已经出现了大量针对语义 Web 中 Web 社会网络的研究，比如文[8]对 Web 社会网络本身进行分析、挖掘，对用户的被信任度进行预测或评估；文[10]研究了如何采用语义 Web 技术对 Web 中的信息进行整合，生成 Web 社会网络。而国内这方面的研究刚开始起步。

## 4 基于 Web 社会网络的推荐模型

针对现有 Web 推荐系统存在的四个问题：①个性化程度不高；②对用户历史数据依赖性过高；③系统开放性不强；④对用户偏好的"走样"。本文提出了一种新的、采用语义 Web 技术的、基于 Web 社会网络的个性化 Web 信息推荐模型：(1)在客户端进行用户偏好采集；(2)通过分析 Web 社会网络进行待过滤信息集采集；(3)在待过滤信息采集过程中，考虑用户的偏好，降低待过滤信息集中冗余信息对推荐结果的影响。图 2 是模型的整体框架。

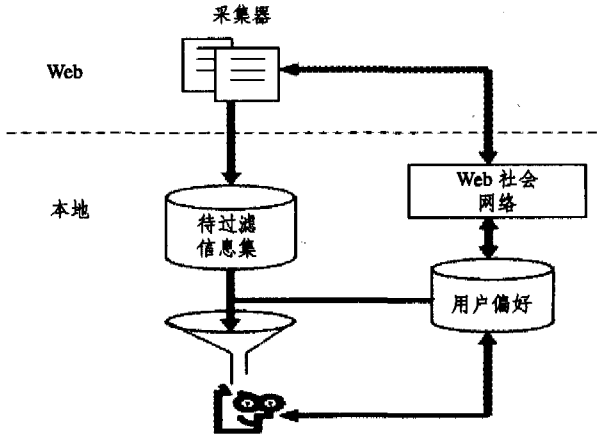


图 2

模型的工作流程如下：1)获取用户偏好；2)以用户偏好为基础，生成用户的 Web 个人社会网络；3)采集并整合其他用户的个人信息描述，如 FOAF 文件，生成用户的 Web 社会网络；4)从 Web 社会网络中寻找与用户偏好相似的用户，为待过滤信息集采集对象；5)采集待过滤信息；6)过滤信息，生成推荐结果。

### 4.1 获取用户偏好

由于用户偏好的获取在本地进行，可以充分利用客户端存储的用户信息。用户能够自己控制偏好的描述与存储，比如：用户可以及时删除暂时性的偏好；也可以随时加入新的偏好，使它们可以立即作用于推荐过程。用户还可以有选择性

维普资讯 http://www.cqvip.com

地，将不涉及泄露个人隐私的偏好生成 FOAF 文件并发布在 Web 上，使得用户通过其他机器上网时，也能利用自己的偏好信息。同时，也使得其他用户可以利用用户发布的 FOAF 文件。图 3 为记录了本文作者偏好的 FOAF 文件片段，它利用作者的 RSS 阅读器 Sage[5] 中的 Weblog 列表，由 Foafgen[2] 自动生成。

```
<foaf:Kmows>
  <foaf:Person>
    <foaf:name>Danny</foaf:name>
    <foaf:weblog rdf:resource="http://dannyayers.com/"/>
    <foaf:seeAlso rdf:resource="http://dannyayers.com/misc/foaf/foaf.rdf"/>
  </foaf:Person>
  <foaf:Person>
    <foaf:name>Paolo Massa</foaf:name>
    <foaf:weblog rdf:resource="http://moloko.itc.it/paolob-log/"/>
    <foaf:seeAlso rdf:resource="http://moloko.itc.it/trustmetricswiki/moim.cgi/MyProjects"/>
    <foaf:seeAlso rdf:resource="http://sra.itc.it/people/massa/paolofoaf.rdf"/>
  </foaf:Person>
</foaf:Knows>
```

图 3

### 4.2 生成 Web 社会网络

Web 社会网络的生成步骤是：1)利用用户偏好生成该用户的 Web 个人社会网络；比如，对本文作者，模型可提取图 3 所示的 FOAF 文件中的用户 Danny 和 Paolo Massa，生成作者的 Web 个人社会网络；2)对已生成的 Web 个人社会网络中各节点，由采集器采集相应用户的 FOAF 文件，对 Web 个人社会网络进行扩展。其中步骤 2 可能需要重复进行，重复的次数由生成的 Web 社会网络规模决定。

整合多个 FOAF 文件，是生成 Web 社会网络的关键。它依赖于对各 FOAF 文件中所涉及用户的准确识别，否则 Web 社会网络中将存在重复节点，影响对 Web 社会网络的分析或挖掘。由于 FOAF 文件采用 RDF 进行描述，因此可利用 URI 对用户进行识别。FOAF 文件中可作为 URI 的有〈foaf:mbox〉、〈foaf:homepage〉、〈foaf:weblog〉所对应的值。但可能存在的问题是：(1)同一个用户可能有多个 Email 地址，尽管每一个都只是他自己使用。因此，若只以 Email 地址作为识别用户的 URI，可能将具有两个或两个以上邮箱的同一个用户识别为不同的用户；(2)同样，同一个用户也可能拥有两个或两个以上的主页或 Weblog。因此，一种可行的解决方案就是利用〈foaf:mbox〉、〈foaf:homepage〉、〈foaf:weblog〉的任意两个对用户进行识别。

若每次进行推荐时都重新执行一次 Web 社会网络的生成过程，将降低系统的执行效率。因此，在本模型中，利用 Web 社会网络对用户偏好进行更新，使得下一次推荐时，Web 社会网络的生成不需要从零开始。

### 4.3 采集待过滤 Web 信息

由于待过滤 Web 信息存储在客户端，因此需要尽量降低与用户偏好不相关的 Web 信息采集，减少对客户端机器存储空间的占用，并提高待过滤信息采集的速度。本模型提出的解决方案是：在采集待过滤信息的过程中就考虑用户的偏好，使得最终采集到的待过滤信息与用户偏好相关。

这一解决方案可以通过分析或挖掘已生成的 Web 社会网络实现。它的依据是[10]：用户间的联系将产生用户间的交互，而用户间的交互将促成用户间相似性的形成。因此，Web 社会网络中连接紧密的多个用户通常具有相似的偏好，他们所链接的 Web 信息可被归为同一类。分析或挖掘 Web 社会网络，就是发现连接紧密的节点。待过滤信息的采集将以这些节点为对象进行。具体的分析或挖掘算法可借用已有的大量社会网络分析方法。

### 4.4 过滤信息

虽然本模型在待过滤 Web 信息的采集过程中就考虑了用户的偏好，但仍然需要在生成最终的推荐结果前对这些 Web 信息进行过滤。4.3 节所述的待过滤信息采集过程确保了采集到的 Web 信息都与用户偏好相关，在此基础上进行基于内容过滤时可以避免对用户偏好的“走样”。

## 5 实验

在本文提出的模型中，Web 社会网络的生成是关键。因此，是否可以在 Web 中采集到数据，并生成 Web 社会网络决定着模型是否可行。待过滤信息的采集影响着 Web 信息的推荐结果，因此，采集到的信息集质量决定着模型是否有效。本文利用采集自 Web 的数据进行了相应的实验验证。

### 5.1 模型的可行性验证

在模型可行性验证中，本文采用了文[9]作者采集的 RDF 数据集。由于文[9]的目的是分析 FOAF 中所用词汇的情况，因此数据的采集具有普遍性。在原数据集中，共有 5958 个 RDF 文件，均以 N-Triple 三元组的形式记录。本实验利用 Seame[6] 对 RDF 进行检索，计算出该数据集中 FOAF 文件和用于描述个人用户信息的 FOAF 文件分别占的比率；分析用于描述个人用户信息的 FOAF 中可被用于生成 Web 社会网络的数据量，从而证明模型的可行性。表 1 给出了实验结果。

表 1

| | |
|---|---|
| FOAF/RDF 的比率 | 98.4% |
| 个人用户的 FOAF/FOAF 的比率 | 59.9% |
| 个人用户 FOAF 中含〈foaf:knows〉的比率 | 34% |

试验证明存在一定数量用于描述个人用户信息的 FOAF 文件，并且这些 FOAF 文件中含有可用于生成 Web 社会网络的数据。

### 5.2 模型的有效性验证

描述本文作者偏好的 FOAF 由 Foafgen[2] 自动生成，它依据的是作者的 RSS 阅读器 Sage[5] 中的 Weblog 列表。该 FOAF 的片段如图 3 所示。结合作者 Weblog 的浏览频率，其 FOAF 文件中所含的用户 Danny 被用于生成作者的 Web 个人社会网络，如图 4(a)。采集 Danny 的 FOAF 文件对作者的 Web 个人社会网络进行扩展，结果如图 4(b)所示。

由于对图 4(b)所示的 Web 个人社会网络无法发现联系紧密的节点，因此继续采集 Dan，Marc，Jay，Captsolo，Morten，Jibber 的 FOAF，并整合这些 FOAF 文件。结果，Web 个人社会网络被扩展为含 37 个节点，58 条有向边的 Web 社会网络。其中，Danny，Dan，Marc，Jay，Captsolo，Mortern，Jibber 这七个用户间的有向边为 20 条，如图 4(c)所示。

(下转第 193 页)

维普资讯 http://www.cqvip.com

据进行共享，通过统一的 XML 数据标准将历史遗留数据和现有数据有机地整合在一起。此外，各 Agent 之间的消息传送也以 XML 格式进行规范化。这样就将异构的子系统有机地集成在一起，同时可以适应未来新技术的变化需求。

**结论** 从整体上看，智能交通系统是由一系列相对独立运行的子系统所构成，这些相对独立的子系统及 ITS 整体效益的发挥，在很大程度上依赖于是否能够充分做好 ITS 的协调和整合。国内外相关的研究与实践表明：共用信息平台是实现 ITS 系统协调、整合的重要技术手段，是 ITS 面向全社会提供信息服务的基本信息平台，是智能交通系统建设不可缺少的重要组成部分。

本文首先对 ITS 共用信息平台的特征和要求进行了分析，简要阐述了 Agent 技术用于共用信息平台建设的技术优势，然后提出了基于 Agent 技术的 ITS 共用信息平台模型 AITSM，最后基于该模型提出了一种公用信息平台架构，我们已经基于该架构实现了某大型 ITS 系统，实际应用表明了本文提出的模型是有效的。

虽然我们实现了该模型，但是还需要在以下几个方面进行深入的研究：(1)协作机制，每项任务的完成需要各 Agent 的联合行动，因而其协作机制中的协作结构、协作方案的研究有非常重要的意义；(2)移动 Agent 的安全问题和定位问题；(3)进一步考虑在基于 Agent 构建的共用信息平台框架上结合 CORBA 技术，这样就可屏蔽 Agent 运行平台和底层通信机制，能够更加有效地解决共用信息平台中 Agent 之间的通信问题[8]。

## 参 考 文 献


1 交通部公路科学研究所. 中国智能运输系统体系框架研究总报告[R]. 北京：交通部公路科学研究所，2001

2 YANG Dong-Yuan. Shared Information Platform for Expressway Management Information System[J]. Journal of TongJI University, 2002, 28(6):664～669

3 GUAN ji-zhen. System Framework and Integration of ITS Common Information Platforms[J]. Journal of Transportation Systems Engineering and Information Technology, 2002 (11):11～16

4 SHI Mei - Lin. An Agent - based Framework for Cross - domain Cooperation of Virtual Enterprise[J]. Computer Integrated Manufacturing Systems, 2003, 9(12):1072～1077

5 Zhang Yun-yong. Mobile Agent Based B2B Collaborative Electronic Commerce System [J]. MINI- MICRO SYSTEM, 2002, 23 (8):935～939

6 Burmeister B, Haddadi A, Matylis G. Applications of multi-agent Systems in traffic and transportation [J]. IEEE Transaction on Software Engineering, 1997, 144(1):51～ 60

7 ZHAO Jian-you. Urban traffic flow control prototype system based on multi-agent [J]. Journal of Traffic and Transportation Engineering, 2003, 3(13):101～105

8 Bella vista P, Conrradi A, Stefanelli C. Middleware services for Interoperability in open mobile agent systems [Z]. [s. n. ], 2001


(上接第 187 页)

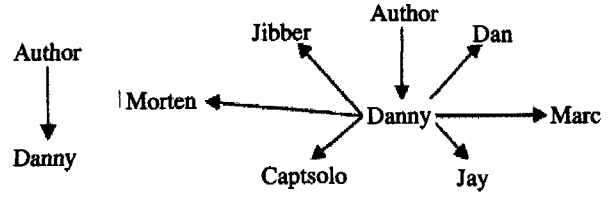


图 4

表 2

| |
|---|
| Danny:httP://dannyayers. com/ |
| Captsolo:http://captsolo. net/info |
| Jay:http://icite. net/blog |
| Morten:http://www. wasab. dk/morten/blog/ |
| Jibber:http://jibbering. com/ |
| Dan:http://www. foaf—project. org/2004/media |
| Marc:http://blogs. it/0100198 |

对图 4(c)中除 Danny 的 6 个用户节点的出入度进行分析，将满足如下两条规则的节点归为待过滤 Web 信息的采集对象。

**规则 1**：Dan, Marc, Jay, Captsolo, Mortern, Jibber 中与 Danny 互有有向边的节点。

**规则 2**：Dan, Marc, Jay, Captsolo, Mortern, Jibber 中入度大于 3 的节点。

根据规则，用户集{Danny, Jay, Morten, Captsolo, Jibber}的 Weblog 或主页被选为待过滤 Web 信息采集对象。

Danny, Dan, Marc, Jay, Captsolo, Mortern, Jibber 这七个用户对应的待过滤 Web 信息采集地址如表 2 所示。通过对 Web 信息内容的分析，可以看出他们有相似的偏好。试验选取与用户具有相同偏好的{Danny, Jay, Morten, Captsolo, Jibber}作为待过滤 Web 信息采集对象，采集结果能产生符合用户偏好的推荐，证明了模型的有效性。

**结论** Web 信息量的快速增长使个性化的 Web 信息推荐系统越来越重要。本文针对目前大多数 Web 信息推荐系统存在的个性化程度不高、对用户历史数据依赖性过高、系统开放性不强、对用户偏好的"走样"等问题，提出了一种新的、采用语义 Web 技术的、基于 Web 社会网络的个性化 Web 信息推荐模型。并且，利用实际数据进行实验，证明了模型的可行性和有效性。

## 参 考 文 献


1 FOAF0. 1. http://xmlns. com/foaf/0. 1/

2 Foafgen. http://www. toxi. co. uk/foafgen/foafgen_v0. 3. 0. zip

3 Friendster. http://www. friendster. com

4 RDF. http://www. w3. org/RDF/

5 Sage. http://sage. mozdev. org/

6 Sesame. http://www. openrdf. org/

7 Berners-Lee T, Hendler J, Lassila O. The Semantic Web. Scientific American, 2001, 284(5):34～43

8 Goldbeck J, Hendler J. Inferring reputation on the semantic web. In: Proc. 13th WWW Conference, 2004

9 Ding L, Zhou L, Finin T, et al. How the Semantic Web is Being Used: An Analysis of FOAF. In: Proc. of the 38th Intl. Conf. on System Sciences, Digital Documents Track (The Semantic Web: The Goal of Web Intelligence), 2005

10 Peter M. Flink: Using Semantic Web Technology for the Presentation and Analysis of Online Social Networks. Journal of Web Semantics, 2005

11 梁邦勇，李涓子，王克宏. 基于语义 Web 的网面推荐模型. 清华大学学报(自然科学版)，2004. 44～49

12 朱征宇. 基于 RDF 的个性化服务模型：[硕士论文]. 2002