

# 基于供求关系及协同过滤技术的推荐模型研究

姜雅倩,王直杰,张 珏

(东华大学 信息科学与技术学院,上海 201620)

**摘 要:**推荐系统已经成功地应用于电子商务、数字图书馆等方面。但随着近年来公共服务平台的发展,现存的推荐系统不能有效处理公共服务平台中不同企业类型之间供求关系的推荐问题,不能针对供求关系产业链做出准确、迅速的推荐。因此,根据公共服务平台的供求关系产业链并结合协同过滤技术,提出了一种新的个性化推荐模型,它基于网络平台中的企业分类、供求关系等来建立模型,并通过建立企业类用户群来缩小协同过滤时用户群体的数量,降低计算时属性空间的维度,从而提高推荐的效率。使用该模型进行推荐可以更好地帮助企业建立沟通渠道、获得服务信息,满足企业个性化的要求。

**关键词:**供求关系;协同过滤;个性化;推荐

**中图分类号:**TP18

**文献标识码:**A

**文章编号:**1673-629X(2007)06-0018-04

## Research on Recommendation Model Based on Supply and Demand Relation and Collaborative Filtering

JIANG Ya-qian, WANG Zhi-jie, ZHANG Jue

(College of Information Sciences and Technology, Donghua University, Shanghai 201620, China)

**Abstract:** Recommendation systems have gained successful applications in E-commerce, digital library and other domains. However, with the development of public service platform, existing recommendation systems can not effectively deal with recommendation issue of the supply and demand relation occurred in different kinds of enterprises and can not make proper and quick recommendations referring to supply and demand relation of industry chain. Thus, combining the feature of public service platform and collaborative filtering, presents a novel personalized recommendation model. The model is created on the basis of the enterprises categories and the supply and demand relation. By establishing enterprises user group, the model reduces the quantity of the user group when collaborative filtering, and eventually improves the efficiency of recommendation. Using this model for recommendation can make great help for the enterprises to establish communication channel and obtain the service information.

**Key words:** supply and demand; collaborative filtering; personalization; recommendation

### 0 引 言

随着网站信息量的增加,人们在感叹信息量丰富的同时,也被大量无用的信息搅得无所适从,毕竟每个人所关注的领域是不同的。因此,人们越来越希望网站可以针对每个人的情况随时进行调整,从而符合个性化的要求。个性化是智能信息服务的体现,通过学习用户信息和用户需求,自动地向用户提供信息。在过去几年里,个性化已经被广泛应用于电子商务领域,并且产生了巨大的产业冲击力。

推荐系统是个性化技术最普遍的应用之一,根据用户的喜好和行为,它能够自动地过滤掉用户不感兴趣的信息,预测感兴趣的信息。现有的推荐系统大都是采用协同过滤技术向用户推荐信息的<sup>[1]</sup>,然而随着系统用户资源的增多,计算的复杂度增加,推荐系统的效率和推荐质量也大大降低<sup>[2]</sup>。

公共服务平台是为中小型企业提供信息化服务的网络环境,它根据中小型企业规模小、数量多的特点,将一系列相互联系、相互作用、相互影响的资源整合在一起,通过产品供应链和信息资源库为中小型企业提供技术、市场、管理等各方面的服务,满足企业的需求,降低企业发展的成本,创造中小型企业发展的环境。

笔者基于公共服务平台的特点,提出了一种新的个性化推荐模型。对基于供求关系的推荐模型进行研究,该模型根据平台中企业之间的供求关系将企业进

收稿日期:2006-08-28

基金项目:上海市科委资助项目(04EZ11068)

作者简介:姜雅倩(1982-),女,山东莱阳人,硕士研究生,主要研究方向为Web数据挖掘、个性化服务;王直杰,教授,主要研究方向为企业信息化、神经网络理论与应用。

行分类,向当前某一类型企业用户推荐与该类企业具有供求关系的其它企业;研究如何利用协同过滤技术向当前用户提供服务等其它信息。将以上两种技术结合起来用于推荐,可以使网络平台中的企业、服务机构等信息更加紧密地联系在一起,提供良好的网络平台环境,更加方便地建立企业之间的沟通渠道,促进企业之间的合作关系。在上海市纺织面料公共服务平台上将文中提出的模型进行了实践应用,基本达到了预期的效果。

## 1 基于供求关系推荐模型的研究

### 1.1 模型的建立

将网络平台内的所有企业分类,如果某两个不同类型的企业之间存在着供求关系,则将这两类企业用箭头链接起来。遍历整个网站,将不同类型企业间的所有供求关系进行连接形成一个网,则这个网是由许多条的供求关系产业链结合而成,网内的任一节点与其相连的节点之间存在着一种供求关系。如图1所示。

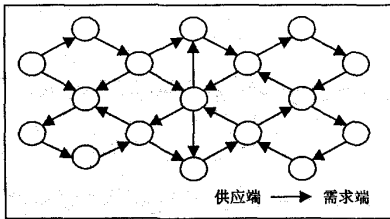


图1 网络平台上企业间的供求关系网

在图1所示的网状关系下建立模型,将每一类企业作为一个实体类(Entity)。对于任一实体类 $Entity_k$ ,设其需求参数为: $d_k = \{d_k^1, d_k^2, \dots, d_k^m\}$ ,供应参数设为: $s_k = \{s_k^1, s_k^2, \dots, s_k^n\}$ ,其中 $d_k^i (i = 1, 2, \dots, m)$ 为供求关系产业链中 $Entity_k$ 的上游企业,即能提供供给 $Entity_k$ 生产运作所需物品的企业; $s_k^j (j = 1, 2, \dots, n)$ 为供求关系产业链中 $Entity_k$ 的下游企业,即 $Entity_k$ 生产的产品所要销售到的企业。将 $Entity_1$ 的参数 $d_1^1, d_1^2, \dots, d_1^m$ 去匹配网络平台中的实体类 $Entity_2, Entity_3, \dots, Entity_p$ ,若存在 $k$ 使得 $d_1^i = Entity_k (i = 1, 2, \dots, m; k = 2, 3, \dots, p)$ 或 $s_1^j = Entity_k (j = 1, 2, \dots, n; k = 2, 3, \dots, p)$ ,则表明在网络平台中实体 $Entity_k$ 与实体 $Entity_1$ 存在着供求关系,根据这种供求关系可以建立一个连接,用于推荐。

### 1.2 算法描述

在.NET环境下,根据上述模型对实体 $Entity_1$ 进行个性化推荐,可设计算法为如下形式:

```
ArrayList dList = new ArrayList();
```

```
ArrayList sList = new ArrayList();
for(int k=2; k<K; k++)
{
    for(int i=1; i<m; i++)
    {
        if(d[1,i] == Entity[k])
        {
            dList.Add(Entity[k]);
            goto net;
        }
    }
    for(int j=1; j<n; j++)
    {
        if(s[1,j] == Entity[k])
        {
            sList.Add(Entity[k]);
            break;
        }
    }
    net: continue;
}
```

其中: $dList, sList$ 分别存放网络平台中 $Entity_1$ 的上游企业和下游企业; $K$ 为平台中的企业类型个数; $m, n$ 分别为 $Entity_1$ 的上、下游企业实体个数; $d[1, i]$ 为 $Entity_1$ 的第 $i$ 个上游企业实体; $s[1, j]$ 为 $Entity_1$ 的第 $j$ 个下游企业实体。

### 1.3 实例分析

下面以纺织面料公共服务平台中的部分网络为例进行说明,如图2所示。

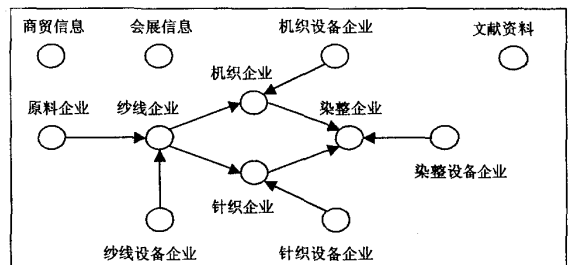


图2 纺织面料公共服务平台的部分网络

设当前用户为纱线企业用户,针对纱线企业用户进行基于供求关系的个性化推荐。根据图2所示的供求关系产业链图,设实体类 $Entity_1$  = “纱线企业”、 $Entity_2$  = “针织企业”、 $Entity_3$  = “机织企业”、 $Entity_4$  = “原料企业”、 $Entity_5$  = “纱线设备企业”……。在供求关系产业链中,纱线企业的上游企业为原料企业、纱线设备企业、软件企业、检测企业等;下游企业为机织企业、针

织企业等。即  $\text{Entity}_1$  的需求参数为  $d_1 = \{\text{“原料企业”}, \text{“纱线设备企业”}, \text{“软件企业”}, \text{“检测企业”}, \dots\}$ ; 供应参数为  $s_1 = \{\text{“机织企业”}, \text{“针织企业”}, \dots\}$ 。对实体  $\text{Entity}_1$  应用 1.2 节中的算法, 结果知在纺织面料公共服务平台中, 与  $\text{Entity}_1$  存在供应关系的上游企业有原料企业、纱线设备企业, 下游企业有机织企业、针织企业, 则将这四类企业的信息作为纱线企业的供求关系企业推荐给纱线企业用户。

## 2 协同过滤推荐模型的研究

### 2.1 协同过滤技术

在网络平台中, 除了企业之间的供求关系还有很多其它服务性等非供求关系机构(如纺织面料公共服务平台中的商贸信息、会展信息、资料文献等, 如图 2 所示), 企业用户在获得了与自身企业相关的供求信息之外还会需要某些服务类的信息, 由于这些服务机构与企业之间并不存在供求关系, 所以不能通过基于供求关系的推荐模型来实现, 因此笔者通过协同过滤技术加以补充, 最终达到理想状态下的个性化。

协同过滤技术是目前在个性化推荐系统中应用较为成功的技术, 它可以分为基于用户(user-based)的协同过滤技术和基于项目(item-based)的协同过滤技术<sup>[3]</sup>。基于用户的协同过滤技术是在访问模式上将当前进行访问的用户与以往的用户相匹配, 根据与当前用户行为相同或相近的已有用户的偏好, 向当前活动用户推荐信息。基于项目的协同过滤技术是依据产品、网页等项目之间的相似性向用户进行推荐的。

一般来说, 在基于用户的协同过滤推荐系统中, 找到与当前用户相似的用户群要在线实时进行运算, 并且随着用户群体和项目群体数量的增大, 计算量增大, 推荐系统的效率和推荐质量会降低<sup>[4]</sup>。而在基于项目的协同过滤推荐系统中, 繁重的计算量通常在离线时进行, 在线推荐部分只需在离线时计算结果的基础上进行少量的运算就可以完成<sup>[5]</sup>。因此, 与基于用户的协同过滤技术相比, 基于项目的协同过滤技术避免了在线繁重的计算量, 具有较高的预测准确性<sup>[6]</sup>。

文中采用基于项目的协同过滤方法, 根据用户-实体矩阵<sup>[5]</sup>, 发现实体  $\text{Entity}_i$  之间的相关度, 将相似的实体推荐给用户。如表 1 所示。

表 1 用户-实体矩阵

	$\text{Entity}_1$	$\text{Entity}_2$	...	$\text{Entity}_n$
$\text{User}_1$		◆	...	◆
$\text{User}_2$	◆		...	
...	...	...	...	...
$\text{User}_m$		◆	...	◆

推荐过程主要有三部分:

1) 确定当前用户所在的企业类用户群。

2) 相关性测定: 计算实体之间的相似度, 根据实体之间的相似度确定与目标实体  $\text{Entity}_i$  的最邻近集合。定义常数  $k$ , 取与目标实体最邻近的  $k$  个实体作为目标实体的最邻近实体。

3) 推荐产生: 将  $k$  个最邻近实体和上文得到的目标实体  $\text{Entity}_i$  的上下游企业做为推荐的集合。

### 2.2 相关性测定

上面已经将网站内的企业按照供求关系进行了分类, 同一类型的企业在对于服务资源的要求方面具有一致性, 所以此处不是在所有用户群体中计算实体之间的相关度, 而是在当前用户所属于的企业类用户群中计算实体之间的相关度, 这样可以减小随用户群体数量增大而推荐效率降低的机率。对于实体类  $\text{Entity}_i$  的用户  $\text{User}_i$ , 浏览  $\text{Entity}_i$  的用户群  $\text{User}_k (k = 1, 2, \dots, m)$ , 通过计算实体  $\text{Entity}_i$  在实体类用户群空间的相似距离, 可以得到与实体类  $\text{Entity}_j (j = 1, 2, \dots, i - 1, i + 1, \dots, n)$  之间的相似程度。

$\text{Entity}_i$  和  $\text{Entity}_j$  两实体对象在  $m$  维属性集的值可作为一个矢量, 有两种具体的计算方法来测定两实体之间的相似性:

1) 用测试实体间相似的距离函数 Cosine Similarity<sup>[7]</sup>来测定: 用  $\text{Entity}_i, \text{Entity}_j$  两对象矢量的角度余弦值作为两对象的相关性测定。即

$$\text{sim}(\text{Entity}_i, \text{Entity}_j) = \cos(\vec{\text{Entity}}_i, \vec{\text{Entity}}_j) = \frac{\vec{\text{Entity}}_i \cdot \vec{\text{Entity}}_j}{\|\vec{\text{Entity}}_i\| \times \|\vec{\text{Entity}}_j\|}$$

2) 用 Pearson 相关系数方法来测定: 设  $i = \text{Entity}_i, j = \text{Entity}_j; U$  为企业用户群;  $\bar{R}_i, \bar{R}_j$  分别为  $\text{Entity}_i, \text{Entity}_j$  在所有属性集上的平均值;  $R_{u,i}, R_{u,j}$  分别为  $\text{Entity}_i, \text{Entity}_j$  在企业用户群空间, 即在  $m$  维属性集上的值。

$$\text{sim}(i, j) = \frac{\sum_{u \in U} (R_{u,i} - \bar{R}_i)(R_{u,j} - \bar{R}_j)}{\sqrt{\sum_{u \in U} (R_{u,i} - \bar{R}_i)^2 \sum_{u \in U} (R_{u,j} - \bar{R}_j)^2}}$$

实验证明, 在文中所提出的模型中采用第二种方法通常能取得较高的推荐准确性。

通过协同过滤技术可以为当前用户发现新的可能需要的信息, 使相似用户可以利用其他用户的结果来扩展所需信息的广度和精确度。

## 3 实验

文中提出的推荐系统正在上海市纺织面料公共服务平台上进行实践应用(网址为 <http://www.shfabric.com>)。

com/。根据平台中的供求关系网自动对不同类型的企业建立模型,这里选取了两个已经注册的企业用户(如图3所示),他们在供求关系产业链环节上分别属于针织企业和染整企业。当针织企业用户 sxzhenzhi 登陆到公共服务平台上时,推荐系统首先确定该企业在供求关系产业链中的位置,然后自动地为企业推荐上下游企业,推荐结果如图4所示(图4分别为针织企

选择用户名: \*   
 用户名只能由英文字母a-z(区分大小写)

贵公司名称: \*   
 贵公司名称只能由英文字母a-z(区分大小写)

供应链环节: \*

选择用户名: \*   
 用户名只能由英文字母a-z(区分大小写)

贵公司名称: \*   
 贵公司名称只能由英文字母a-z(区分大小写)

供应链环节: \*

图3 企业用户注册简况

当前用户: sxzhenzhi  
 进入会员中心

功能菜单: 企业信息, 产品信息, 设备信息

企业信息

搜索:  搜索

推荐公司列表 >>

企业名称	企业类型
杭州萧山坎山三家花式丝厂	纱线企业
远纺工业(无锡)有限公司	纱线企业
富丽达集团控股有限公司	纱线企业
常州市武进五洋纺织机械有限公司	针织设备企业
常州市润源经编机械有限公司	针织设备企业
德国迈斯西针织圆机有限公司	针织设备企业
舟山东方印染有限公司 曾用名: 舟山印染厂	染整企业
上海锦丰针织印染服饰有限公司	染整企业
合肥常春棉织厂	染整企业

当前用户: wlrnliao  
 进入会员中心

功能菜单: 企业信息, 产品信息, 设备信息

企业信息

搜索:  搜索

推荐公司列表 >>

企业名称	企业类型
邢台天润劳动布有限公司	机织企业
天津市天旭布业有限公司	机织企业
宁波中经丝绸服饰有限公司	机织企业
常州市永魁染整有限公司	染整设备企业
海城市永魁染整有限公司	染整设备企业
无锡市致业机械制造有限公司	染整设备企业
泉州海元针织服装有限公司	针织面料企业
扬州润诚纺织制品有限公司	针织面料企业
上海黑川纺织有限公司	针织面料企业
吴江明月纺织制品有限公司	针织服装企业
常州吴氏针织有限公司	针织服装企业
绍兴市浩哥针纺织有限公司	针织服装企业

图4 推荐结果

业用户 sxzhenzhi 和染整企业用户 wlrnliao 的推荐企业列表)。从图中可以看出,系统对于当前不同类型的企业用户,根据其所在供应链环节向其推荐不同的上下游企业。

## 4 结论

协同过滤技术是目前应用较为广泛的个性化推荐技术。笔者针对公共服务平台的特殊特点,首次提出了基于供求关系产业链的个性化推荐模型,并根据企业类用户群的特殊性,将目标实体所在的属性空间维度降低,当用户数目和实体数目指数级增长时,仍能较高保证推荐算法的实时性和准确性要求。

基于企业间的供求关系及企业类用户群进行推荐有助于提高系统在预测和推荐时的性能。系统的输入端数据由企业分类数据库、企业数据库和用户数据库组成;输出端数据是由个性化的企业推荐列表组成。整个过程被划分成以下3个过程:根据企业分类建立

模型、用户信息获取、基于供求关系及相关性规则生成推荐。通过推荐列表,网络平台中的企业用户可以有效地找到与其存在供求关系的企业和能够提供服务的机构。其中,对企业进行正确的分类是很重要的,这需要领域专家在企业供求关系的基础上,根据特定的领域知识实施一定的分类规则。分类的结果会直接影响分析的结果,决定着推荐质量的好坏。

下一步要开展的工作包括研究推荐系统的实时性和推荐质量之间的平衡问题、多种推荐技术推荐结构的集成以及推荐系统中的隐私保护问题,使所提出的个性化推荐系统能够进一步完善。

## 参考文献:

- [1] Guo Xuetao, Lu Jie. Recommending Trade Exhibitions by Integrating 'Semantic Information' with Collaborative Filtering[C]//Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI'05). Washington D. C.: IEEE Computer Society, 2005: 747-750.
- [2] Ji Junzhong, Sha Zhiqiang, Liu Chunlian, et al. Online Recommendation Based on Customer Shopping Model in E-Commerce[C]//Proceedings of the IEEE/WIC International Confer-

(下转第25页)

的更新报文(BU)。在更新报文中,将 MN 的转交地址作为源 IP 地址,将 CN 的地址作为目的 IP 地址,MN 的家乡地址填写在家乡地址选项中,发送给 CN,使之更新 MN 的位置为直接和 MN 通信做好准备。这样一个路由优化模式完成了,MN 和 CN 可以进行直接通信了。

#### 4.3 优化方案的信令和时延分析

在文献[4]的方案中,MN 发送绑定更新(BU)来更新 CN 的绑定缓存表项(Binding Cache Entry)执行路由优化模式,总共需要 6 条信令。MN 与 CN 直接进行交换的信令 CoTI&CoT,通过 MN 的 HA 与 CN 交换的信令 HoTI&HoT。如果这 4 条信令成功交换后,MN 再发送绑定更新 BU 至 CN 更新绑定缓存表项,成功后 CN 发送绑定确认(BA)给 MN 对 BU 进行确认。其中 HoTI 和 CoTI 信令同时发送,响应信令 HoT 和 CoT 也几乎同时返回。这 4 条信令就是迂回路由过程(RRP),交换信令的通信过程需要 1 个往返时延。CN 的绑定更新过程交换的信令 BU 和 BA 需要 1 个往返时延。这样在路由优化模式中就需要 2 个 MN 到 CN 的往返时延,这对于一些对时延敏感的多媒体实时应用有着严重的影响。

另一方面,路由优化模式要求 MN 和 CN 计算一个共享的密钥(Kbm),而每计算一次需要交换 4 个信令信息(CoTI、CoT、HoTI、HoT)。如果在信令交换的通信过程中 4 个信令信息丢失任意一个,MN 则需要和 CN 交换 2 个以上的信令信息,进一步增加了信令开销和通信时延。

由以上分析可以看出迂回路由过程使用了过多的信令,这些信令的交互带来至少 2 个 MN 到 CN 的往返时延,而在实际的应用中,MN 和 CN 之间的距离一般都比较远,所以信令交互带来的通信时延难以满足实际的应用,特别是对一些多媒体实时业务(VoIP,视频等)。由于文献[4]在路由优化模式中采用了迂回路由过程,切换性能难有突破性的进展,整体效果仍然不

是很理想。

优化方案利用 AAA 机制取代了迂回路由过程,降低了信令开销和交互这些信令的通信时延,切换性能有了显著提高。

## 5 总 结

在下一代以移动为特征的互联网环境中,AAA 如何与移动 IPv6 合理部署以及 AAA 协议和移动 IPv6 协议的融合是解决移动节点跨域移动问题的关键,也是研究的难点。文中讨论了三种融合的方案,并在信令开销和切换时延方面进行了分析比较,并详细描述了 AAA 认证和 MIPv6 注册融合时信令交互的过程。

从信令开销和切换时延上看,新的优化融合方案更符合需求。下一步的研究工作将集中在完善优化方案中的报文的格式,并将该方案应用到 FMIPv6 或者 HMIPv6 中,进一步加强移动 IPv6 的性能。

#### 参考文献:

- [1] Johnson D, Perkins C, Arkko J. Mobility Support in IPv6. RFC 3775[S]. IETF, 2004.
- [2] Guy C. 移动 IPv6 介绍[EB/OL]. 2004-09-01. <http://www.microsoft.com/china/technet/community/columns/cableguy/cg0904.msp>.
- [3] DuPont F, Laurent - Maknavicius M, Bournelle J. AAA for Mobile IPv6. Draft draft - dupont - mipv6 - aaa - 01. txt [S]. IETF, 2002.
- [4] Wang R C, Chen R Y, Chao Han - Chieh. AAA architecture For Mobile IPv6 based on WLAN[J]. Int. J. Network Mgmt, 2004, 14(5): 305 - 313.
- [5] Le F, Patil B, Perkins C, et al. Diameter Mobile IPv6 Application. Draft draft - le - AAA - diameter - mobileip6 - 04. txt[S]. IETF, 2005.
- [6] Ryn S, Mun Y. An Enhanced Mobile IPv6 Handover for Roaming between Administrative Domains Based on AAA. Draft - mun - mipshop - emip6 - aaa - 00. txt[S]. IETF, 2006.

(上接第 21 页)

- ence on Web Intelligence (WI'03). Washington D. C. : IEEE Computer Society, 2003: 68 - 74.
- [3] Deshpande M, Karypis G. Item - Based Top - N Recommendation Algorithms[J]. ACM Transactions on Information Systems, 2004, 22(1): 143 - 177.
- [4] 熊 馨, 王卫平, 叶跃祥. 基于概念分层的个性化推荐算法[J]. 计算机应用, 2005, 25(5): 1006 - 1008.
- [5] Kim Dong - Ho, Im I, Atluri V. A Clickstream - based Collaborative Filtering Recommendation Model for E - Commerce [C]//Proceedings of the Seventh IEEE International Conference on E - Commerce Technology (CEC'05). Washington D. C. : IEEE Computer Society, 2005: 84 - 91.
- [6] Sarwar B, Karypis G, Konstan J, et al. Item - based collaborative filtering recommendation algorithms[C]//Proceedings of the 10th International Conference on World Wide Web. New York: ACM Press, 2001: 285 - 295.
- [7] Shardanand 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: ACM Press/Addison - Wesley Publishing Co., 1995: 210 - 217.