一种基于后项不定长关联规则的 Web 个性化推荐方法*)

丁增喜　王菊英　王大玲　鲍玉斌　于　戈
(东北大学信息科学与工程学院　沈阳110004)

A Web Personalized Recommendation Method Based on Uncertain Consequent Association Rules

DING Zeng-Xi　WANG Ju-Ying　WANG Da-Ling　BAO Yu-Bin　YU Ge
(College of Information Science and Egineering, Northeast University, Shenyang 110004)

Abstract　Web usage mining plays an important part in supporting personalized recommendation on Web and association rule uncovers the interesting relations among items hidden in data. The paper gives an idea of association rule merging-deleting based on the analysis of association rule characteristics and implements it in the rule preparation before the Web personalized recommendation. Furthermore, based on the comparisons in precision, coverage and F_1 of recommendation system and the rule numbers used in three kinds of association rules, a Web personalized recommendation method based on uncertain consequent is put forward. After integrative analysis of several recommendation methods, the method given in the paper can be thought as a good selection. At last several page-weighted techniques are introduced in the paper.

Keywords　Association rule, Web usage mining, Personalized recommendation, Precision, Coverage

1　引言

随着 Internet 的发展，电子商务、远程教育等各种基于网络的服务也得到迅猛发展。实现网上信息个性化已经成为一种热点。由于 Web 使用挖掘[1,2]自身所具有的一些特性，其在个性化推荐中的应用已经得到快速的发展。这方面的研究有基于用户聚类和资源聚类技术的个性化推荐方法[3~5]，基于访问路径特性的推荐思想[6]。基于以上技术和方法的个性化推荐系统有很多，比如 WebPernerlizer 系统等[7]；而基于关联规则的个性化推荐方法也很多，其中一种方法是通过一个频繁项目图结构来在线地给出用户推荐信息[8]。

在以往基于关联规则的个性化推荐方法中使用的大多是后项长仅为1的关联规则[8,9]，而实际上，应用关联规则挖掘算法不仅生成后项长为1的规则，还生成许多后项长大于1的规则。这样会造成以下三个问题：①关联规则作为一种揭示频繁数据项间的一些内在联系的技术[10]，只通过后项长为1的规则还不能完全反映一些数据项间的关联关系，这样在生成推荐信息时就会造成推荐系统的准确率和覆盖率的下降。②在利用规则生成算法生成的规则中后项长度为1的规则数目可能比较多，这样当在线利用规则生成推荐时会造成系统性能的下降。③由于利用单支持度生成的关联规则常常造成稀有频繁项问题[11]。因此，从生成关联规则到利用关联规则生成推荐信息的整个过程中有必要采取一定的解决策略以提高推荐系统的准确率、覆盖率以及推荐系统的整体性能。

针对问题①、②，本文提出了一种改进的基于后项不定长规则的个性化推荐方法。在该方法中我们应用的关联规则不仅包含后项长度为1的部分，而且还包含后项长度大于1的部分，我们称之为后项不定长的关联规则。此外，我们针对个性化推荐方法的特性又提出了一种关联规则合并删减的思想，该思想可以帮助我们减少实际用于推荐的关联规则数量。针对问题③，我们在采用 Lin 提出的自适应支持度的思想[12]，通过设定最小可信度和生成关联规则数目范围来自动调整规则生成时的支持度，从而尽量减小稀有频繁项问题。

通过实际数据测试我们给出了三种个性化推荐方法在所使用的关联规则数量、推荐方法准确率、覆盖率以及综合测度上的比较。实验结果显示我们所提出的推荐方法在保证准确率不下降的前提下，在覆盖率和综合测度上要比以往基于后项长度仅为1的规则的推荐方法要好。此外我们的推荐方法使用了重复规则的合并删减思想，所以推荐所用规则数目要比原有规则集以及后项长仅为1的规则集的数目又要少。因此应用我们所提出的推荐方法可以在很大程度上提高推荐系统的性能。本文的最后又提出了几种为提高推荐系统准确率而采用的页面加权技术。

2　Web 页面个性化推荐系统的衡量参数

一般对一个个性化推荐系统来说，其性能的度量指标有两个：覆盖率(Coverage)和准确率(Precision)。覆盖率指的是在推荐的内容中用户喜欢的项占用户喜欢的所有项的百分比，准确率指的是在推荐的内容中用户喜欢的项所占推荐所有项的百分比[3]。若将 RS 看作是所有推荐页面的集合，US 看作是用户喜欢的所有页面的集合，那么覆盖率和准确率可以分别采用式(1)、(2)进行计算。

$$precision = |RS \cap US| / |RS| \quad (1)$$

$$coverage = |RS \cap US| / |US| \quad (2)$$

同时，为了将二者综合起来考虑从而使系统的性能总体上取得最优，Mobasher 等又提出一个将二者综合起来的度量指标 F_1[5]，F_1的计算可由式(3)得到。这几个指标界定了个性化推荐系统的性能和效率，而人们研究的重点也就放在了这几个方面。

$$F_1 = \frac{2 \times Precision \times Coverage}{Precision + Coverage} \quad (3)$$

*)国家自然科学基金资助项目(No. 60173051)。

3 基于关联规则的 Web 个性化推荐系统的实现过程

尽管 Web 个性化推荐系统的实现方法有很多种，但这些系统的整体结构一般分为两个部分：离线处理部分和在线处理部分[7]。前者主要应用数据挖掘技术从日志文件中发现各种规则模式；而在线处理部分则是根据用户访问信息，应用挖掘出的规则模式来实时给出推荐。图1给出了一个基于关联规则的 Web 个性化推荐系统结构图。在离线处理部分，通过对站点文件的预处理生成事务集[13]，再应用各种关联规则挖掘方法生成关联规则集。而在线处理部分中首先实时获取用户的当前访问序列，将访问序列同离线生成的关联规则集中每条规则的前项进行匹配，从而得到匹配关联规则集。通过计算匹配关联规则集中每条规则后项所含页面的推荐得分来得到候选页面集。如果设定了每次推荐给用户的页面个数为 k，则选取候选页面集中得分最高的前 k 个页面推荐给用户。

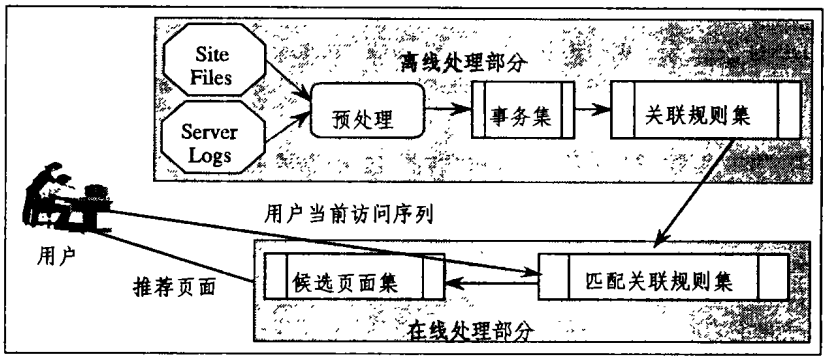

图1 Web 个性化推荐系统结构图

4 基于后项不定长关联规则的 Web 个性化推荐方法的实现及测试方法

4.1 关联规则的生成

关联规则挖掘作为数据挖掘的一个重要方法，揭示的是数据集中所有满足用户描述的最小支持度和最小可信度限制的项目关系。最小支持度决定着一条关联规则所能包含的数据项的最小数目，而一般常用关联规则挖掘算法都是通过预先设定规则最小支持度来生成关联规则。但是如果在整个数据集中只设定一个最小支持度，便预先认为数据集中的项目具有同种属性或者具有相同的频繁性。如果最小支持度设置太高，就会造成虽然出现不频繁但很具有挖掘价值项的丢失，这就是所谓的稀有项问题(Rare Item Problem)；如果最小支持度设置太低就会造成关联规则生成太多的问题。针对以上问题 Liu 等人提出了运用多个最小支持度挖掘关联规则的解决办法。但是该方法比较复杂且不适合个性化信息的推荐。为了选择合适的最小支持度，Lin 等人提出了自适应支持度的思想，通过预先设定最小可信度和关联规则生成数目范围来得到合适的关联规则。在生成过程中，根据生成的规则数是否符合设定规则数目范围来自动改变规则的最小支持数目(MinsupportCount)，从而动态地改变最小支持度的值。在该方法中，规则的生成可以采用以往各种常规挖掘算法来实现。这里我们使用 Han 等人提出的 FP-Growth 算法[14]。由于 FP-Growth 算法是一个频繁模式生成算法，所以在实际生成中我们增加了 FP-Growth 的从频繁模式到关联规则的生成功能，具体实现过程见文[15,16]。

4.2 个性化推荐中的关联规则合并删减思想

在应用关联规则挖掘算法生成的规则集中存在着大量对推荐系统来说具有重复性的关联规则。因此，在进行个性化推荐之前有必要对那些具有重复性的关联规则进行规则的合并删减，以减少推荐时所需要的规则数目。在实际应用关联规则进行推荐时我们是通过比较用户访问序列与规则前项间的匹配关系来生成匹配规则集的，此外在计算推荐页面的推荐得分时有可能需要考虑每个页面所在规则的可信度。所以我们可以将每条规则的前项、可信度看作是推荐相关属性，而其它的属性便可以看作是推荐非相关属性。对于推荐非相关属性，我们可以采取规则属性的合并(比如规则后项)或删除的方式进行处理。如果我们在测试推荐方法性能的时候采用规则的支持度作为这次测试的推荐阈值，则将规则的支持度也看作是推荐相关属性。这样我们便将那些具有相同推荐相关属性的规则看作是重复性关联规则。而对于这些重复性规则可以将它们合并成一条新的关联规则，合并的同时删除旧的重复性规则。合并后的新规则保留了重复性规则中的推荐相关属性，同时将原规则集中的对应非相关属性合并成新规则中的对应属性或是在新规则中被删除。图2给出了一个将 n－1条规则前项、规则支持度和可信度相同的重复性规则合并成一条新规则的示例。

重复性关联规则集：
规则1：AB⇒CD　sup＝50%　conf＝80%
规则2：AB⇒EF　sup＝50%　conf＝80%
……
规则 n－1：AB⇒X　sup＝50%　conf＝80%
操作步骤：
(1)合并结果
规则 n：AB⇒CDEF……X　sup＝50%　conf＝80%
(2)删除原重复性规则

图2 重复性规则的合并删减

通过对重复性关联规则的合并删减可以大规模减少推荐所用规则数目。如果在推荐时考虑的推荐相关属性越少，规则的合并程度就越大，在实时推荐时所用的规则数目就越少。如果将图2示例中考虑的支持度看作是非相关属性，那么对于原始规则集中那些虽然支持度不同但是规则前项和可信度相同的规则合并成一条新的规则同时删除原有的重复性规则。在生成的新规则中将不会包含支持度属性。

4.3 测试数据及测试方法

我们以美国 DePaul 大学 CTI Web 服务器(http://www.cs.depaul.edu)记载的日志文件作为源日志文件，经过清理之后的事务文件包含13745个会话和683个页面。为了便于分析，剔除那些出现次数小于0.5%或大于80%的页面和长度小于2的会话记录。我们随机抽取预处理后数据集的70%作

为训练集，用于关联规则集的挖掘，而剩余的30%作为测试集，用于推荐系统的性能测试。我们采用文[14,15]中的FP-Growth方法来生成关联规则集，剔除了结果中支持度大于90%的部分，这样便得到了比较和测试所用的原始关联规则集。原始关联规则集中有19730条关联规则，其中包含8027条后项长只为1的规则。在对原始规则集进行合并删除之后，我们得到的规则数则减少为6476条。图3是三类规则集数目的比较图。

从图3可以看出经过合并删减之后规则的数目不仅比原始规则数目有了很大程度的减少，而且也比后项长只为1的规则数目有一定的减少。如果不考虑关联规则的支持度属性，经过合并删减后的规则数目将会更少。

对生成的关联规则集经过处理之后我们得到了三类规则集：未经合并删除处理的原始关联规则集、后项长为1的关联规则集以及经过合并删除之后的后项不定长关联规则集。下面我们将应用测试集对基于以上三种关联规则集在推荐阈值（这里采用规则的可信度作为推荐阈值）分别为0.1～0.9情况下进行推荐系统的准确率、覆盖率以及综合测度的比较分析。

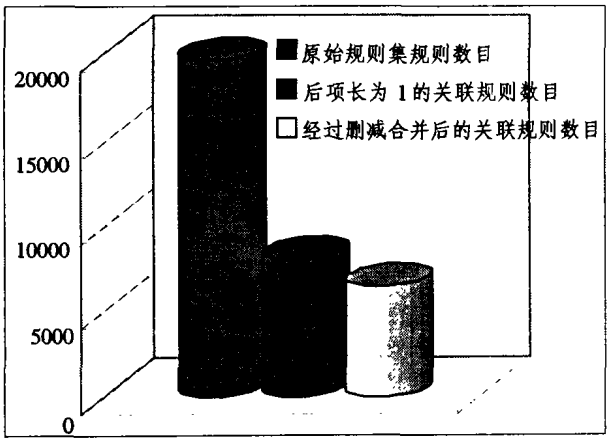

图3 关联规则数目比较

我们的测试方法如下：先设定活动用户滑动窗口的最大长度为 $n=5$，取测试集中第 i 条会话的前 k 项（$k\leqslant n$）构成的项目集作为活动用户的访问序列集，根据该访问序列集到关联规则集中找规则前项与访问序列集相匹配的规则，将得到的匹配规则后项所含项目集作为该次的推荐信息集 RS_{ik}。而测试集中第 i 条会话中剩余部分构成的项集就可以看作是用户真正喜欢的项集 US_{ik}。在利用关联规则获得推荐信息时，只有前项完全与活动用户的访问序列集相匹配的规则才能用于生成推荐信息。在进行推荐信息生成时，对于生成的重复推荐信息也看作为不同信息进行统计。那么在给定推荐阈值时，对测试集中每一条会话分别统计用户访问序列长为1～n 情况下的推荐页面数、用户所喜欢的页面数以及两者交集中的页面数。这样推荐系统的准确率和覆盖率计算公式则变为：

$$Precision=\frac{\sum_{i=1}^{m}\left(\sum_{k=1}^{n}\left(|RS_{ik}\cap US_{ki}|\right)\right)}{\sum_{i=1}^{m}\left(\sum_{k=1}^{n}|RS_{ik}|\right)} \tag{4}$$

$$Coverage=\frac{\sum_{i=1}^{m}\left(\sum_{k=1}^{n}\left(|RS_{ik}\cap US_{ki}|\right)\right)}{\sum_{i=1}^{m}\left(\sum_{k=1}^{n}|US_{ki}|\right)} \tag{5}$$

其中 m 为测试集中用户会话的个数，US_{ki} 表示当用户访问序列长为 k 时由第 i 条会话生成的用户喜欢的页面集；RS_{ki} 表示当用户访问序列长为 k 时由第 i 条会话生成的推荐页面集。

图4、图5分别表示当推荐阈值分别为0.1～0.9的条件下基于以上三类规则集的推荐系统的准确率、覆盖率比较图。这里我们将基于原始规则集也就是未经合并删减后的规则集的推荐方法看作为第一种方法；将基于后项为1的规则集的推荐方法看作为第二种方法，而基于后项不定长但合并删除后的规则集的推荐方法看作为第三种方法。

从图4中可以看出随着推荐阈值的增加基于三类规则集的推荐方法的准确率变化平稳但是总体趋势是上升的，这说明随着推荐阈值的增加，正确推荐的个数相对于用户喜欢的个数增加较快。而在这三种推荐方法中准确率变化曲线相互交错，但总的来看第二种推荐方法的准确率相对其他两种方法来说较差一些，而第一种方法的准确率最好。第三种方法，也就是基于后项不定长但经合并删除后的规则集的推荐方法的准确率居于二者之间。

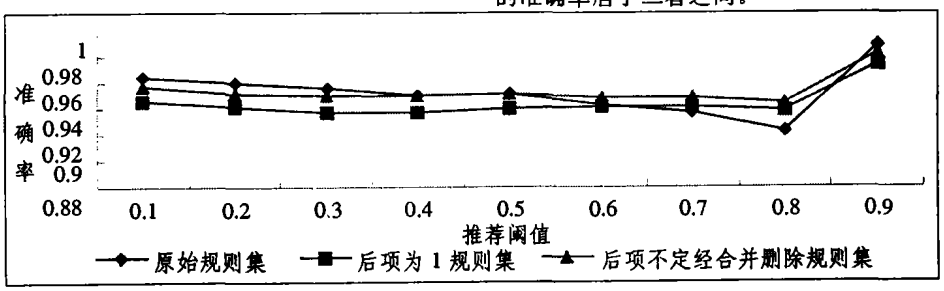

图4 基于三类规则集合推荐方法准确率比较图

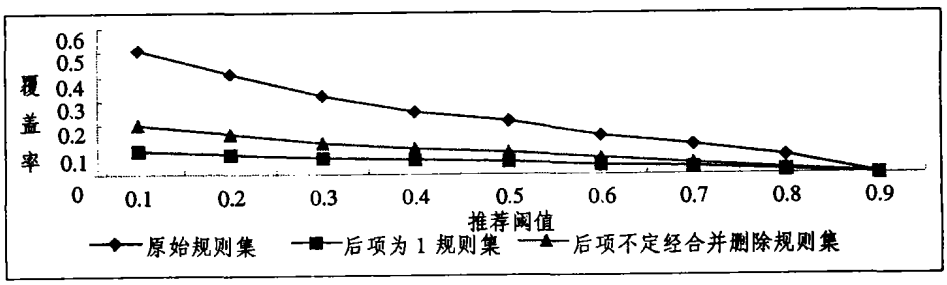

图5 基于三类规则集合推荐方法覆盖率比较图

从图5可以看出三种推荐方法的覆盖率随着推荐阈值的增加而逐步下降。其中第二种方法的覆盖率相对于其它两种推荐方法来说较差，而第三种方法的覆盖率处在两者之间。

以上两图表明第三种方法的准确率和覆盖率都要比第二种方法要好。这同我们分析的结果相同。虽然在基于原始规则集的推荐方法在覆盖率和准确率上要比基于合并删除后规则集的推荐方法好，但图3表明基于原始规则集的推荐方法所用的规则集数目较多，这样就不利于推荐系统推荐信息的实时产生，从而影响系统的响应速度。所以在实际生成推荐信息时不适宜采用原始规则集。

为了综合准确率和覆盖率获得整体最优，我们给出了三种推荐方法的综合测度 F_1（F_1的计算见式(3)）的比较图，如图6所示。

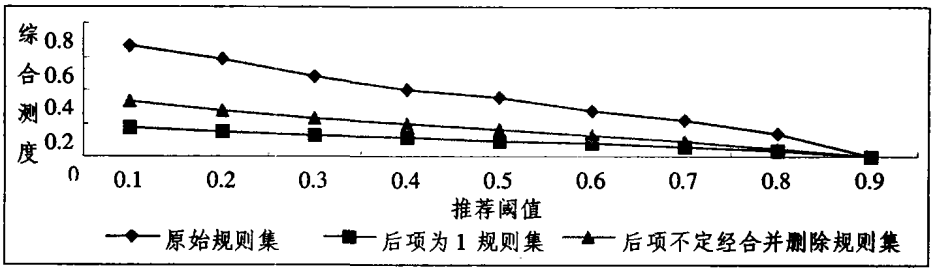

图6 基于三类规则集推荐方法综合测度比较图

从图6可以看出，就整体性能上分析，基于后项不定但经合并删除规则集的推荐方法综合测度相对于基于后项为1规则集推荐方法的综合测度要好。

5 对Web个性化推荐中页面加权方法的几点考虑

在应用Web使用挖掘实现Web个性化推荐时，为了提高个性化推荐的准确率和覆盖率或是出于其它目的考虑，一般要对推荐信息进行加权处理。这里将介绍几种在个性化推荐中所使用的页面加权技术。

5.1 基于用户当前访问页面与预推荐页面间距离的加权

一般来说网站都可以看作是一个具有根（主页）的网状树，其中各个节点——页面之间都有一定的最短连通距离，即使二者不连通，它们之间的距离也可以通过它们与根的距离的绝对值相加或其它计算方法来得到。那么从个性化推荐生成的角度考虑，某页面同当前用户的距离越远，当前用户下一步要访问该页的可能性就越小，对于推荐系统来说该页所具有的推荐价值就越高。因此，在进行推荐时，可以在综合规则可信度和页面间相对距离加权的基础上向用户推荐得分较高的那些页面。

定义1 页面权值函数 $W(p,m)$，用于计算各页面的加权值。其中 p 表示用户访问当前页面，m 表示候选推荐页面，其值可以通过对两个页面间距离进行归一化计算得到。

定义2 推荐得分函数 $S(p,m)$，用于计算各页面的推荐得分，其计算公式为：

$$S(p,m)=Confidence(r)+a*W(p,m) \tag{6}$$

其中 m 为候选推荐页面，p 为用户当前访问页面，$Confidence(r)$ 为匹配关联规则 r 的可信度，a 是一个归一化调节因子。

5.2 基于页面访问时间的加权

Web日志对用户访问行为的记录不仅包括用户访问的内容，而且还包括用户对该页面的起始访问时间，通过计算同一会话中相邻页面间访问时间的差就可以得到用户在一个页面上的停留时间。一般地，用户在某个页面上的停留时间越长，说明该页面要比其它的页面更具有吸引力，在推荐的时候也就应该优先推荐给访问用户。因此，对于一个Web个性化推荐系统而言，如果能将用户在页面上的停留时间应用到页面推荐得分的计算中来，就能进一步提高个性化推荐系统的准确率和覆盖率，从而提高系统的推荐效率，增强服务的竞争力。

假设对Web日志文件数据预处理后我们得到含有 n 条会话的事务，其中第 k 条含有 m 个页面的会话为：

$$\langle pageview_{1,k},stime_{1,k}\rangle\cdots\langle pageview_{i,k},stime_{i,k}\rangle,$$
$$\langle pageview_{i+1,k},stime_{i+1,k}\rangle\cdots\langle pageview_{m,k},stime_{m,k}\rangle$$

其中 $pagevie_{i,k}$ 表示第 k 条会话中的第 i 个访问页面（经过净化后对用户会话有用的页面），$stime_{i,k}$ 表示用户访问第 k 条会话的 $pageview_{i,k}$ 页面的起始时间，则

$$\Delta t_{i,k}=stime_{i+1,k}-stime_{i,k} \tag{7}$$

表示第 k 个会话中用户在第 i 个页面上的停留时间（最后一个访问页面除外，它的访问时间可以用该条会话的页面访问时间平均值来得到），这样通过计算便可得到第 k 条会话中用户在每个访问页面上的停留时间。这样第 i 个访问页面在第 k 条会话中的访问时间权重可按式(8)进行计算：

$$TW_k(Pageview_{i,k})=\Delta t_{i,k}/\sum_{i=1}^{m}\Delta t_{i,k} \tag{8}$$

通过以上计算得到的页面时间权值消除了由于网络拥塞等所带来的访问时间延长等问题对计算所造成的影响。这样访问页面 $pageview_i$ 在该事务中的访问时间权重则为：

$$TW(pageview_i)=\sum_{k=1}^{s}TW_k(pageview_i)/s \tag{9}$$

其中 s 为给定事务中出现页面 $pageview_i$ 的会话个数，该访问权重是一个大于0小于1的值。将该时间权值替换式(6)中的距离权值函数便可以计算出基于时间加权技术的页面推荐得分。

应用页面间距离的加权技术和页面访问时间的加权技术计算得到的推荐得分可以比较客观的评价一个页面对访问用户的重要程度。

结论与展望 个性化推荐是Web个性化技术的一个重要组成部分，应用Web使用挖掘方法获得用户的访问模式从而为其生成推荐页面是一种切实可行的方法，对此本文做了以下几方面的工作：

· 提出重复关联规则的合并删减思想。通过合并删减推荐相关属性相同的关联规则可以大量减少实时推荐所用关联规则数目，从而有助于提高系统响应性能。

· 在应用重复关联规则的合并删减思想的基础上，提出了基于后项不定规则的Web个性化推荐方法。并将该方法同基于后项长只为1关联规则集推荐方法以及同基于原始规则

（下转第88页）

$=\sum_{i=P(k-1)+1}^{Pk}|R_i|$表示一个划分$R_{P(k-1)+1},\cdots,R_{Pk}$中所有关系元组数之和，则对其相应辅助视图$V_k$进行更新所需的连接运算费用为$T_{Vk_join}=\sum_{i=P(k-1)+1}^{Pk}(f_1*(|Rsum_pk|-|R_i|))$，可见当在总的费用中连接运算所占的比重较大时，应当在划分时在分配有较大f_i值的源关系R_i的组中尽量放进那些元组数较少的其他源关系，以求得较小的连接费用。改进后算法的形式化描述如下：

```
void Freq_Size_Partition(R[],n,K); //R[]存放源关系,n 为源关系个数
{ k=1; //划分为 K 组,k 为当前组号
  Fk=0; //第 k 组中关系的更新频率之和
  Gk=Φ; //存放第 k 组关系的集合
  M =⌊∑(i=1..n) fi/K⌋; //每一组中关系的更新频率和的期望值
  Sort_on_freq(R[],f[],n);//n 个源关系按更新频率从大到小排序,f[]记录排序编号
  Sort_on_size(R[],s[],n);//n 个源关系按元组数从大到小排序,s[]记录排序编号
  for (k=1;k<=K;k++)
  { for (i=1;i<=n,f[i]==0;i++) {}; //找到当前未被选中的更新频率值最大的关系
    Gk= Gk∪{R[f[i]]}; Fk:= Fk+f[i]; f[i]:=0;
    j=n;
    while (Fk<M && j>=1)
    { if (s[j]!=0) { Gk= Gk∪{R[s[j]]}; Fk:= Fk+f[j]; s[i]:=0;}
      j--;
    }
}
```

上述算法在保证划分后的各组具有相同的更新频率和的基础上，将包含元组数较少的关系分到更新频率值较高的关系所在的组中，以获得较低的连接运算费用。如果在更新某一辅助视图时给其他辅助视图加锁，则上述算法可以保证完全一致性。

·如果将源关系的关键属性都加进视图定义中，对视图进行增量更新的时间还会大大减少，其代价就是需要在数据仓库中占用更多的空间。那么如果源关系按照更新频率进行了分组，每一组都对应了一个辅助视图。对于那些包含着高更新频率源关系的"特殊组"所对应的辅助视图，我们在其所基于的各个源关系的关键属性上建立索引，可以提高辅助视图的更新效率。

总结　本文讨论了对于在高速局域网或计算网格环境下定义在多个数据源上的物化视图，利用数据仓库中的额外空间建立辅助视图，提出了根据各个源关系的规模（所含元组数）和不同的更新频率因素进行划分的策略和相应算法，最大可能地减小连接计算费用和网上传输费用的有效更新算法并进行了相关的分析。

参考文献

1 Agrawal D, Abbadi A E, Singh A, Yurek T. Efficient View Maintenance at Data
2 Y. Zhuge et al. View Maintenance in a Warehousing Environment. In: Proc. of ACM SIGMOD Conf. 1995. 316～327
3 Zhuge Y, et al. The Strobe Algorithm for Multi-source Warehouse Consistency. In:Proc. of Intl. Conf. On Parallel and Distributed Information Systems, Miami Beach,FL, USA 1996
4 Wang Hui, et al. Efficient Refreshment of Materialized Views with Multi Sources. In Proc. of CIKM'99, Kansas City, MO, USA,1999. 375～382
5 Bakgaard L, Roussopoulos N. Efficient Refreshment of Data Warehouse Views:[TR3642]. DCS,Univ. of Maryland at College Park,Dec. 1996
6 Hull R, Zhou G. Towards the Study of Performance Trade-offs Between Materialized and Virtual Integrated Views. In:Proc. of Workshop on Materialized Views: Techniques and Applications Montreal,1996. 91～102
7 Labrinidis A, Roussopoulos N. Reduction of Materialized View staleness Using Online Updates: [TR3878]. DCS, Univ. of Maryland at College Park,Feb. 1998
8 Quass D, Widom J. Online Warehouse View Maintenance. In: Proc. of ACM SIGMOD Conf. Tucson, Arizona,1997. 393～404
9 Hammer J,Garcia-Molina H, Widom J,Labio W, Zhuge Y. The Stanford Data Warehousing Project. IEEE Bulletin of the Technical Committee on Data Engineering,1995, 18(2):41～48

（上接第72页）

集基础上的推荐方法在准确率、覆盖率、综合测度以及所使用的规则数目上进行了比较。测试结果表明该方法比较适用于个性化推荐。

·为了进一步提高个性化推荐的准确率我们给出了几种个性化推荐中的页面加权技术。

在将来的研究中我们将把以上几种页面加权技术应用到基于实际数据的个性化推荐中，根据测试结果以便选取一种更好的方法。

参考文献

1 Cooler R,Mobasher B,Srivastav J. Web Mining: Information and Pattern Discovery on the World Wide Web. (ICTAI1997)
2 Srivastava J,Cooley R,Deshpande M,Tan P. Web Usage Mining: Discovery and Application of Usage Patterns from Web Data. ACM SIGKDD, Jan. 2000
3 Mobasher B,Dai H,Luo T,et al. Discovery of aggregate usage profiles for Web personalization. In:Proc. of the WebKDD 2000 Workshop at the ACM SIGKDD2000, boston,Aug. 2000
4 Mobasher B. Mining Web Usage Data for Automatic Site Personalization. To appear in Gaul, W. ,& Ritter,G. (Eds.), Classification , Automation, and New Media, Springer. 2001
5 Mobasher B,Dai H,Luo T, et al. Improving the Effectiveness of Collaborative Filtering on Anonymous Web Usage Data. In:Proc. of The IJCAI 2001 Workshop on Intelligent Techniques for Web Personalization(ITWP01)
6 Gaul W, Schmidt-Thieme L. Recommender Systems Based on Navigation Path Features. In: Proc. of the WEBKDD 2001 Workshop on Mining Web Log Data Across All Customer Touch Points. The Seventh ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining ,San Franscisco,CA,2001
7 Mobasher B. WebPersonalizer: A Server-Side Recommender System Based on Web Usage Mining: [Technical Report TR99-110]. Department of Computer Science, Depaul University
8 Mobasher B,Dai H,Luo T,et al. Effective Personalization Based on Association Rule Discover from Web Usage Data. In:Proc. of The 3rd ACM Workshop on Web Information and Data Management
9 Fu X, Budzik J, Hammond K. Mining navigation history for recommendation. In: Proc. 2000 int. conf. intelligent user interfaces,New Orleans,LA,ACM,Jan. 2000. 106～112
10 Agrawal R, Srikant R. Fast algorithms for mining association rules. In: Proc. Int. Conf. Very Large DataBase, 487-499, Santiago,Chile
11 Liu B,Hsu W,Ma Y. Mining Association Rules with Multiple Minimum Supports. KDD-99 San Diego CA USA,1999
12 Yang W,Sergio L,Alvarez A,et al. Efficient Adaptive-Support Association Rule Mining for Recommender Systems. Data Mining and Knowledge Discovery,2002,6:83～105
13 Cooler R,Mobasher B,Srivastava J. Data Preparation for Mining World Web Browsing Patterns. Knowledge and Information System(1999)
14 Han J,Pei J,Yin Y. Mining Frequent Patterns without Candidate Generation. In: Proc. 2000 ACM-SIGMOD Int. Conf. on Management of Data(SIGMOD'00)
15 丁增喜，刘洋，江志纲，等. 基于关联规则发现和页面加权的个性化推荐技术的研究. 小型微型计算机系统 sup. 2002,23
16 丁增喜. 支持 E-Services 的关联规则挖掘方法的研究与实现:[东北大学毕业论文]. 2003