(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101826114 B
(45)授权公告日 2012.05.09

(21)申请号 201010182844.4
(22)申请日 2010.05.26
(73)专利权人 南京大学
地址 210093 江苏省南京市鼓楼区汉口路22号
(72)发明人 陈振宇 封煜佳 王浩然 刘嘉 吴一帆
(74)专利代理机构 南京天翼专利代理有限责任公司 32112
代理人 黄明哲
(51)Int.Cl.
*G06F 17/30*(2006.01)
*G06Q 30/02*(2012.01)

(56)对比文件
US 5640553 A,1997.06.17, 全文.
CN 101482884 A,2009.07.15, 全文.
CN 101320375 A,2008.12.10, 全文.

审查员 史江峰

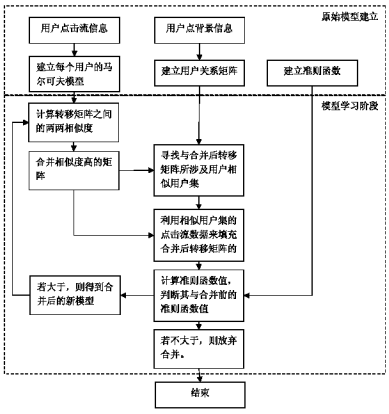
权利要求书 3 页 说明书 7 页 附图 1 页

(54)发明名称

一种基于多马尔可夫链的内容推荐方法

(57)摘要

一种基于多马尔可夫链的内容推荐方法，利用用户的点击流信息建立马尔可夫模型，同时利用用户的背景信息建立用户关系矩阵。然后对相似的马尔可夫模型进行合并，并根据用户关系矩阵得到的相似用户集合的点击流对合并后的马尔可夫模型的零行进行稀疏项的填充。本发明为一种网络上的个性化信息推荐技术，根据用户的兴趣特点，行为，以及个人资料向用户推荐感兴趣的商品和信息，在庞大的数据中为用户推荐其所感兴趣的信息和商品，减少浏览的时间，同时解决的了协同推荐中用户评分项相对较少，并且有很多稀疏项的问题，提高了推荐的精确度。




CN 101826114 B

1. 一种基于多马尔可夫链的内容推荐方法，其特征是通过网站获取用户点击流数据，以及用户背景信息，对其进行分析，并生成内容推荐模型；当一个用户产生新的点击流时，利用当前的点击流数据以及内容推荐的模型产生用户可能感兴趣的项目，并推荐给用户；包括以下步骤：

1)、原始模型建立：建立原始模型，包括每个用户的马尔可夫模型，用户关系矩阵以及用于评价聚类结果好坏的聚类准则函数；

2)、模型学习阶段：使模型进行学习，合并相似的马尔可夫模型，并利用背景相似用户的点击数据填充合并后马尔可夫模型的的零行，也就是缺省信息；

3)、用户推荐：利用用户当前的点击以及所处组别的模型，进行推荐；

具体为：

1)、原始模型建立：

1.1)、记录并提取每个用户的点击流数据，所述点击流信息是基于控件的点击流信息；

1.2)、利用点击流数据对每个用户建立马尔可夫模型，包括转移矩阵 A 和初始状态 λ，用户集合 G：

转移矩阵 A 中，每个页面 X 表示模型的一个状态，$X_t$ 表示当前状态，$X_{t-1}$ 则表示前一刻的状态，设 $P_{ij} = (X_t = x_j | X_{t-1} = x_i)$，$0 < i < n$，$0 < j < n$，n 为总用户数，即 $P_{ij}$ 表示由状态 $x_i$ 转移到状态 $x_j$ 的概率，当 A 所指向的用户没有点击过页面 $X_t$ 时，出现 $P_{t1}$，$P_{t2}$，……$P_{tn}$，这一行无法计算，设置为零行，

$$A=(p_{ij})=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix},$$

初始状态 $\lambda = (p_i) = (p_{i2}, p_{i2} \cdots P_{in})$；

当马尔可夫模型仅由用户 $u_1$ 的点击流数据建立时，用户集合即为 $G = \{u_1\}$；

1.3)、从网站注册用户的注册文件中获得用户背景信息，包括用户年龄、性别、学历、工作、地域，根据这些用户背景信息来建立用户关系矩阵，并利用用户背景信息确定用户之间的相似性；

1.4)、建立评价聚类结果好坏的聚类准则函数，得到初始准则函数值 Z；

2)、模型学习阶段：

2.1)、计算每个转移矩阵之间的两两相似度，然后确定所有马尔可夫模型之间的相似度；

2.2)、设定相似度阈值，合并相似度超过阈值的马尔可夫模型，并计算合并后的马尔可夫模型的转移矩阵以及初始状态，此时的用户集合 G 包含了合并的马尔可夫模型所代表的所有用户，同时删除被合并的马尔可夫模型；

2.3)、根据步骤 2.2) 得到的合并的马尔可夫模型所代表的用户，在步骤 1.3) 得到的用户关系矩阵中查找相似的用户，由相似的用户构成集合 GS；

2.4)、利用相似用户，即集合 GS 的用户的点击流信息来填充 2.2) 中得到的合并后的马尔可夫模型的转移矩阵的零行；

2.5)、计算合并后聚类的准则函数：对步骤 2.2) 中每一种可行的马尔可夫模型合并方案都合并，并计算准则函数值，选择其中最大的准则函数值 $Z_1$，与初始准则函数值 Z 比较，若 $Z_1 > Z$，则计算当前合并的马尔可夫模型的两两之间的相似度，回到步骤 2.2) 进行所有可行的合并，即二次合并，选取最大的二次合并的准则函数值 $Z_2$ 与 $Z_1$ 比较，若 $Z_2 > Z_1$ 则回到步骤 2.2) 进行三次合并，如此循环直至得到使准则函数值最大的合并，步骤 2.4) 得到的填充过的马尔可夫模型最终确定，进入步骤 2.6)；

2.6)、学习结束；

3)、利用模型进行用户推荐：

3.1)、用户产生新的点击流数据，记录该点击流数据用于下一次模型的学习；

3.2)、确定用户所处的马尔可夫模型，包括转移矩阵和初始状态；若用户为新用户，则根据用户关系矩阵，利用背景信息相似的用户生产马尔可夫模型；

3.3)、对用户当前的点击流数据以及相应的马尔可夫模型得到最热门的推荐，并显示给用户；

3.4)、结束；

步骤 1.3) 中所述的建立用户关系矩阵的步骤如下：

1.3.1)、根据用户背景信息数据，建立用户背景信息矩阵；

1.3.2)、根据用户背景信息矩阵计算两两用户之间的相似性；

1.3.3)、建立用户关系矩阵：

设当前共有 k 个用户，其中 $S_{ij}$ 表示用户 I 与用户 J 的相似度，

$$U=(US_{ij})=\begin{bmatrix} S_{11} & S_{12} & K & S_{1k} \\ S_{21} & S_{22} & & S_{2k} \\ & M & O & M \\ S_{k1} & S_{k2} & L & S_{kk} \end{bmatrix};$$

步骤 2.3) 所述的利用用户关系矩阵寻找背景相似用户，步骤如下：

2.3.1)、获取步骤 2.2) 得到的合并的马尔可夫模型所代表所有 q 个用户；

2.3.2)、对于 q 个用户，在用户关系矩阵中，检查每个用户的所在行，将所在行的相似度由大到小排序，选取前 m 个，得到每个用户各自相似度最大的 m 个用户；

2.3.3)、得到 q 组 m 个用户，将其合并后得到相似用户集合 SG；

步骤 2.4) 填充转移矩阵的零行，步骤如下：

2.4.1)、得到相似用户的点击流信息；

2.4.2)、获取合并后的马尔可夫模型的转移矩阵中由于数据缺失无法计算的项目；

2.4.3)、用相似用户的点击流信息填充转移矩阵中无法计算的项目，即零行。

2. 根据权利要求 1 所述的一种基于多马尔可夫链的内容推荐方法，其特征是步骤 3.2) 利用用户所属的马尔可夫模型进行推荐；当用户是新用户时：利用步骤 2.3) 与 2.4) 来计算当前用户的预测马尔可夫模型，并根据此模型进行推荐。

3. 根据权利要求 1 所述的一种基于多马尔可夫链的内容推荐方法，其特征是步骤 3.3) 根据当前点击数据以及相应的马尔可夫模型得到最热门的推荐，步骤如下：

3.3.1)、确定当前用户现在点击的页面为 $X_t = x_g$，以及所处的马尔可夫模型的转移矩阵为 $A_u = (p_{u-ij})$；

3.3.2)、取得 $A_u$ 中的 g 行，即状态 $x_g$ 的行，$(p_{u-gj}) = (p_{u-g1}, p_{u-g2}, \cdots\cdots p_{u-gn})$；

3.3.3)、对所有的 $p_{u-gj}$，$0 < i <= n$，进行降序排列为，$p_{u-gn1}, p_{u-gn2}, \cdots\cdots p_{u-gnn}$；

3.3.4)、设定推荐前 N 个内容为最热门内容，取前 N 个 $p_{u-gnj}$：为 $p_{u-gn1}, p_{u-gn2}, \cdots\cdots p_{u-gnN}$，那么所对应页面的 $X_{t+1} = x_{n1}, X_t = x_{n2}, \cdots\cdots X_t = x_{nN}$，即为所推荐的最热门内容。

# 一种基于多马尔可夫链的内容推荐方法

## 技术领域

[0001] 本发明涉及个性化推荐技术领域，根据用户的兴趣特点、行为以及个人资料向用户推荐感兴趣的商品和信息。个性化推荐基于海量的数据挖掘，常用于电子商务以及社会型网络应用，可以在庞大的数据中为用户推荐其所感兴趣的信息和商品，减少浏览的时间。本发明具体为一种基于马尔可夫链并结合用户背景信息的内容推荐方法。

## 背景技术

[0002] 个性化推荐技术是一个有巨大应用价值的技术。个性化推荐技术近年来不断的被各种电子商务型网站以及社会型网站所应用，为用户提供他们所感兴趣的信息和商品。个性化推荐技术最早是在 1995 年被提出来的。此后不断的被发展应用于电子商务领域，并为电子商务网站带来了巨大的利益，如亚马逊。近年来许多的社会型网络应用也不同程度的使用了推荐系统，比如豆瓣，用以为用户推荐感兴趣的信息。

[0003] 个性化推荐技术的方法主要包括以下三种 ：

[0004] 1) 基于关联规则的推荐算法 ；

[0005] 2) 基于内容的推荐算法 ；

[0006] 3) 协同过滤推荐算法。

[0007] 基于关联规则的推荐算法，首先挖掘关联规则形成规则库，然后为用户提供相应的推荐项目，但其可扩展性不能满足需求。

[0008] 基于内容的推荐算法是定义项目与项目之间的相似度，然后为用户推荐与其所浏览或感兴趣过的项目相似的项目。但是这样的算法在对音乐，电影等很难提取内容的项目时有非常大的难度。并且基于内容的推荐算法只能发现相似的项目，但是无法推荐用户可能有兴趣的其他类项目。

[0009] 协同过滤推荐算法则是在用户群中寻找相似的用户，然后综合这些相似用户对某一项目的评价来预测该用户对这个项目喜好程度。协同过滤算法是一项比较受欢迎的技术。它可以对比较复杂的项目比如音乐、电影进行推荐，同时也能够保证推荐的新颖性。但是用户的评价信息有时候非常稀疏，可能导致用户的相似性并不准确，从而使得所推荐的项目并不为用户所喜爱。同时协同过滤推荐算法的性能在用户以及项目数量大幅增加后可能会比较低。

## 发明内容

[0010] 本发明要解决的问题是 ：现有的个性化推荐技术的方法存在不同程度的不足，对于用户可能感兴趣的项目不能做到全面推荐，不能克服推荐算法中的可扩展性问题和稀疏性问题。

[0011] 本发明的技术方案为 ：一种基于多马尔可夫链的内容推荐方法，通过网站获取用户点击流数据，以及用户背景信息，对其进行分析，并生成内容推荐模型 ；当一个用户产生新的点击流时，利用当前的点击流数据以及内容推荐的模型产生用户可能感兴趣的项目，

并推荐给用户；包括以下步骤：

[0012] 1)、原始模型建立：建立原始模型，包括每个用户的马尔可夫模型，用户关系矩阵以及用于评价聚类结果好坏的聚类准则函数；

[0013] 2)、模型学习阶段：使模型进行学习，合并相似的马尔可夫模型，并利用背景相似用户的点击数据填充合并后马尔可夫模型的的零行，也就是缺省信息；

[0014] 3)、用户推荐：利用用户当前的点击以及所处组别的模型，进行推荐。

[0015] 本发明步骤具体为：

[0016] 1)、原始模型建立：

[0017] 1.1)、记录并提取每个用户的点击流数据，所述点击流信息是基于控件的点击流信息；

[0018] 1.2)、利用点击流数据对每个用户建立马尔可夫模型，包括转移矩阵 A 和初始状态 λ，用户集合 G：

[0019] 转移矩阵 A 中，每个页面 X 表示模型的一个状态，$X_t$ 表示当前状态，$X_{t-1}$ 则表示前一刻的状态，设 $P_{ij} = (X_t = x_j | X_{t-1} = x_i)$，$0 < i < n$，$0 < j < n$，n 为总用户数，即 $P_{ij}$ 表示由状态 $x_i$ 转移到状态 $x_j$ 的概率，当 A 所指向的用户没有点击过页面 $X_t$ 时，出现 $P_{t1}$，$P_{t2}$，…… $P_{tn}$，这一行无法计算，设置为零行，

[0020]

$$A=(p_{ij})=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix},$$

[0021] 初始状态 $\lambda = (p_i) = (p_{i2}, p_{i2} \cdots p_{in})$；

[0022] 当马尔可夫模型仅由用户 $u_1$ 的点击流数据建立时，用户集合即为 $G = \{u_1\}$；

[0023] 1.3)、从网站注册用户的注册文件中获得用户背景信息，包括用户年龄、性别、学历、工作、地域，根据这些用户背景信息来建立用户关系矩阵，并利用用户背景信息确定用户之间的相似性；

[0024] 1.4)、建立评价聚类结果好坏的聚类准则函数，得到初始准则函数值 Z；

[0025] 2)、模型学习阶段：

[0026] 2.1)、计算每个转移矩阵之间的两两相似度，然后确定所有马尔可夫模型之间的相似度；

[0027] 2.2)、设定相似度阈值，合并相似度超过阈值的马尔可夫模型，并计算合并后的马尔可夫模型的转移矩阵以及初始状态，此时的用户集合 G 包含了合并的马尔可夫模型所代表的所有用户，同时删除被合并的马尔可夫模型；

[0028] 2.3)、根据步骤 2.2) 得到的合并的马尔可夫模型所代表的用户，在步骤 1.3) 得到的用户关系矩阵中查找相似的用户，由相似的用户构成集合 GS；

[0029] 2.4)、利用相似用户，即集合 GS 的用户的点击流信息来填充 2.2) 中得到的合并后的马尔可夫模型的转移矩阵的零行；

[0030] 2.5)、计算合并后聚类的准则函数：对步骤 2.2) 中每一种可行的马尔可夫模型合并方案都合并，并计算准则函数值，选择其中最大的准则函数值 $Z_1$，与初始准则函数值 Z 比较，若 $Z_1 > Z$，则计算当前合并的马尔可夫模型的两两之间的相似度，回到步骤 2.2) 进行所

有可行的合并，即二次合并，选取最大的二次合并的准则函数值 $Z_2$ 与 $Z_1$ 比较，若 $Z_2 > Z_1$ 则回到步骤 2.2) 进行三次合并，如此循环直至得到使准则函数值最大的合并，步骤 2.4) 得到的填充过的马尔可夫模型最终确定，进入步骤 2.6)；

[0031] 2.6)、学习结束；

[0032] 3)、利用模型进行用户推荐：

[0033] 3.1)、用户产生新的点击流数据，记录该点击流数据用于下一次模型的学习；

[0034] 3.2)、确定用户所处的马尔可夫模型，包括转移矩阵和初始状态；若用户为新用户，则根据用户关系矩阵，利用背景信息相似的用户生产马尔可夫模型；

[0035] 3.3)、对用户当前的点击流数据以及相应的马尔可夫模型得到最热门的推荐，并显示给用户；

[0036] 3.4)、结束。

[0037] 步骤 1.3) 中所述的建立用户关系矩阵的步骤如下：

[0038] 1.3.1)、根据用户背景信息数据，建立用户背景信息矩阵；

[0039] 1.3.2)、根据用户背景信息矩阵计算两两用户之间的相似性；

[0040] 1.3.3)、建立用户关系矩阵：

[0041] 设当前共有 k 个用户，其中 $S_{ij}$ 表示用户 I 与用户 J 的相似度，

[0042] $U=(US_{ij})=\begin{bmatrix} S_{11} & S_{12} & K & S_{1k} \\ S_{21} & S_{22} & & S_{2k} \\ & M & O & M \\ S_{k1} & S_{k2} & L & S_{kk} \end{bmatrix}$。

[0043] 步骤 2.3) 所述的利用用户关系矩阵寻找背景相似用户，步骤如下：

[0044] 2.3.1)、获取步骤 2.2) 得到的合并的马尔可夫模型所代表所有 q 个用户；

[0045] 2.3.2)、对于 q 个用户，在用户关系矩阵中，检查每个用户的所在行，将所在行的相似度由大到小排序，选取前 m 个，得到每个用户各自相似度最大的 m 个用户；

[0046] 2.3.3)、得到 q 组 m 个用户，将其合并后得到相似用户集合 SG。

[0047] 进一步的，步骤 2.4) 填充转移矩阵的零行，步骤如下：

[0048] 2.4.1)、得到相似用户的点击流信息；

[0049] 2.4.2)、获取合并后的马尔可夫模型的转移矩阵中由于数据缺失无法计算的项目；

[0050] 2.4.3)、用相似用户的点击流信息填充转移矩阵中无法计算的项目，即零行。

[0051] 步骤 3.2) 利用用户所属的马尔可夫模型进行推荐；当用户是新用户时：利用步骤 2.3) 与 2.4) 来计算当前用户的预测马尔可夫模型，并根据此模型进行推荐。

[0052] 步骤 3.3) 根据当前点击数据以及相应的马尔可夫模型得到最热门的推荐，步骤如下：

[0053] 3.3.1)、确定当前用户现在点击的页面为 $X_t = x_g$，以及所处的马尔可夫模型的转移矩阵为 $A_u = (p_{u-ij})$；

[0054] 3.3.2)、取得 $A_u$ 中的 g 行，即状态 $x_g$ 的行，$(p_{u-gj}) = (p_{u-g1}, p_{u-g2}, \cdots\cdots p_{u-gn})$；

[0055] 3.3.3)、对所有的 $p_{u-gj}$，$0 < i <= n$，进行降序排列为，$p_{u-gn1}, p_{u-gn2}, \cdots\cdots p_{u-gnn}$；

[0056] 3.3.4)、设定推荐前 N 个内容为最热门内容，取前 N 个 $p_{u-gnj}$：为 $p_{u-gn1}, p_{u-gn2}, \cdots\cdots$

$p_{u-gnN}$，那么所对应页面的 $X_{t+1} = x_{n1}$，$X_t = x_{n2}$，……$X_t = x_{nN}$，即为所推荐的最热门内容。

[0057] 本发明利用点击流信息来进行内容推荐，并对其中的稀疏问题提供解决办法，也就是对用户的缺省信息通过相似用户进行填充。首先虽然用户对于项目的评分可能非常少，但是用户的浏览数据却可以显示用户对于项目的关注度。同时简单的点击次数会遗失很多的数据，所以利用点击流数据来作为用户推荐的基础信息。进一步考虑到光有点击流数据并不能完全解决稀疏性问题，在此基础上利用用户的背景信息再次填充含有零行的稀疏矩阵，以提高推荐系统的精确度。

[0058] 本发明的有益效果：

[0059] 1)、相比于基于关联规则的推荐算法，本发明提高了可扩展性。基于关联规则的推荐算法在每次有新的用户及数据加入时需要对所有的数据重新进行挖掘，可扩展性不强。而本发明可以为新增加的用户新建一类，然后利用相似用户的行为为其进行推荐；

[0060] 2)、相比于基于内容的推荐算法，本发明可以提供类似项目之外的用户可能感兴趣的项目；

[0061] 3)、相比协同过滤推荐算法，本发明利用用户的点击流信息来为用户进行项目推荐，能够在不同的点击后推荐不同的项目，提高了推荐的精度，同时解决了其中的稀疏性问题。

## 附图说明

[0062] 图 1 为本发明中的原始模型建立，以及模型学习阶段的流程示意图。

## 具体实施方式

[0063] 本发明的基于马尔可夫链并结合用户背景信息的内容推荐方法，可以应用到电子商务以及社会化网站中。可以为用户提供感兴趣的项目和链接，方便用户浏览网站，增加电子商务网站的下单率，提高应用的智能化程度。

[0064] 本发明中所涉及的术语解释：

[0065] 1) 马尔可夫模型：包括转移矩阵，初始状态，以及所代表用户集合三个部分。用三元组 MC(A，λ，G) 来表示。

[0066] 其中转移矩阵 A 中，每个页面 X 表示模型的一个状态，$X_t$ 表示当前状态，$X_{t-1}$ 则表示前一刻的状态，设 $P_{ij} = (X_t = x_j | X_{t-1} = x_i)$，$0 < i < n$，$0 < j < n$，n 为总用户数，即 $P_{ij}$ 表示由状态 $x_i$ 转移到状态 $x_j$ 的概率，当 A 所指向的用户没有点击过页面 Xt 时，出现 $P_{t1}$，$P_{t2}$，……$P_{tn}$，这一行无法计算，设置为零行，

[0067]

$$A=(p_{ij})=\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix},$$

[0068] 初始状态 $\lambda = (p_i) = (p_{i2}, p_{i2} \cdots p_{in})$；

[0069] 当马尔可夫模型仅由用户 $u_1$ 的点击流数据建立时，用户集合即为 $G = \{u_1\}$.

[0070] 2) 马尔可夫模型的相似度

[0071] 马尔可夫模型的动态特性由转移矩阵来描述，模型的相似性基于转移矩阵的相似

性。

[0072] 对任意两个马尔可夫模型 $MC_m$ 和 $MC_q$，有相应的转移矩阵 $A_m$ 和 $A_q$，它们的第 i 行分别为：$(P_{m-ij})$ 和 $(P_{q-ij})$，j = 1,2,...,n，这两行的相似度为：

[0073] $CE(P_{m-ij}, P_{q-ij}) = \sum_{j=1}^{n}(P_{m-ij}\log\frac{P_{m-ij}}{P_{q-ij}})$

[0074] 转移矩阵 $A_m$ 和 $A_q$ 的相似度为 $Sim(A_m, A_q,) = \sum_{i=1}^{n} CE(P_{m-ij}, P_{q-ij})/n$。

[0075] 则马尔可夫模型的相似度为：

[0076] $MCSim(MC_m, MC_q) = 2/(Sim(A_m, A_q,)+Sim(A_q, A_m,))$，且 $MCSim(MC_m, MC_q) = 2/0 = \infty$。

[0077] 3) 用户相似度

[0078] 用户的相似度计算为公知技术，如《结合用户背景信息的协同过滤推荐算法》，吴一帆，王浩然，这篇论文中所提的方法。

[0079] 4) 用户关系矩阵

[0080] 设共有 k 个用户，其中 $S_{ij}$ 表示用户 I 与用户 J 的相似度。

[0081] $U=(US_{ij}) = \begin{bmatrix} S_{11} & S_{12} & K & S_{1k} \\ S_{21} & S_{22} & & S_{2k} \\ & M & O & M \\ S_{k1} & S_{k2} & L & S_{kk} \end{bmatrix}$

[0082] 图 1 为本发明的原始模型建立以及模型学习阶段的流程示意图，包括以下步骤：

[0083] 1)、原始模型建立阶段

[0084] 1.1)、记录并提取每个用户的点击流数据；

[0085] 1.2)、利用数据对每个用户建立马尔可夫模型，包括转移矩阵和初始状态，设共有 n 个用户；$MC_k = (A_k, \lambda_k, G_k)$，$0 < k < n+1$，

[0086] 其中 $A_k=(p_{k-ij}) = \begin{bmatrix} p_{k-11} & p_{k-12} & \cdots & p_{k-1n} \\ p_{k-21} & p_{k-22} & & p_{k-2n} \\ \vdots & & \ddots & \vdots \\ p_{k-m1} & p_{k-m2} & \cdots & p_{k-mn} \end{bmatrix}$，$G_k = \{u_k\}$；

[0087] 初始状态下对每个用户建立一个马尔可夫模型，这个时候每个马尔可夫模型所代表的是只有一个用户的用户集合。在后面相似的马尔可夫模型合并后，每个马尔科夫模型将会代表多个用户，马尔可夫模型的用户集合中也对应包含多个用户。

[0088] 1.3)、建立用户关系矩阵，利用用户背景信息来确定用户之间的相似性：

[0089] $U=(USij) = \begin{bmatrix} S_{11} & S_{12} & K & S_{1k} \\ S_{21} & S_{22} & & S_{2k} \\ & M & O & M \\ S_{k1} & S_{k2} & L & S_{kk} \end{bmatrix}$

[0090] 1.4)、建立评价聚类结果好坏的准则函数。

[0091] 1.5)、结束；

[0092] 2)、模型的学习阶段

[0093] 2.1)、计算每个转移矩阵之间的两两相似度，并确定不同马尔可夫模型之间的相似度；即对于 $(A_1, A_1, ... A_k)$ 中的任意 $A_i$ 和 $A_j$ 计算相似度 $S_{ij}$；

[0094] 2.2)、设定相似度阈值，合并相似度超过阈值的马尔可夫模型，并计算合并后的马尔可夫模型的转移矩阵以及初始状态，此时的用户集合 G 包含了合并的马尔可夫模型所代表的所有用户，同时删除被合并的马尔可夫模型；这里相似度阈值设定越大，则可合并模型数量越少，用户集合 G 中的信息就比较少，会出现过多的零行，而阈值设定过低，用户信息的聚类就会受到影响，可能出现用户集合中信息过于混杂，无法准确按用户兴趣进行推荐，影响推荐结果。

[0095] 2.3)、根据步骤 2.2) 得到的合并的马尔可夫模型所代表的用户，在步骤 1.3) 得到的用户关系矩阵 U 中查找相似的用户，由相似的用户得到相似用户集合 $GS(u_1, u_2, \ldots u_{gsn})$，；相似用户集合 GS 与用户集合 G 在用户的背景信息上是相似的。在用户关系矩阵 U 中查找时，比如用户 j，检查用户关系矩阵的 j 行，找出这一行最高的三个值，这些值所对应的列为 i1，i2，i3，那么用户 j 在关系矩阵中的相似用户即为用户 i1，i2，i3。

[0096] 2.4)、利用相似用户，即集合 GS 的用户的点击流信息来填充 2.2) 中得到的合并后的马尔可夫模型的转移矩阵的零行。得到用户集合 G 的点击流信息，及由其产生的新的马尔可夫模型，也就是步骤 2.2) 合并得到的马尔可夫模型，记其中的转移矩阵为 $A_{Gij}$，同时删去原有的转移矩阵 $A_i$ 和 $A_j$ 以及它们所对应的初始状态。对于矩阵 $A_{Gij}$ 中出现的零行，利用用户集合 GS 的点击流数据来计算填入。

[0097] 2.5)、计算合并后聚类的准则函数：对步骤 2.2) 中每一种可行的马尔可夫模型合并方案都合并，并计算准则函数值，选择其中最大的准则函数值 $Z_1$，与初始准则函数值 Z 比较，若 $Z_1 > Z$，则计算当前合并的马尔可夫模型的两两之间的相似度，回到步骤 2.2) 进行所有可行的合并，即二次合并，选取最大的二次合并的准则函数值 $Z_2$ 与 $Z_1$ 比较，若 $Z_2 > Z_1$ 则回到步骤 2.2) 进行三次合并，如此循环直至得到使准则函数值最大的合并，步骤 2.4) 得到的填充过的马尔可夫模型最终确定，进入步骤 2.6)；

[0098] 根据某个相似度阀值来选择哪两个马尔可夫模型的合并时可能有多种选择。比如，ABC 是三个马尔可夫模型，相似度阀值设为 0.7，而 A 和 B 的相似度为 0.9，A 和 C 的相似度为 0.8，那么就可以选择 A 和 B 合并或者 A 和 C 合并。需要判断这两种合并哪种更加合理，准则函数就是判断合并是否合理的依据。所以需要预先将 A 和 B 合并，计算准则函数值 Sab。然后再将 A 和 C 合并，计算准则函数值 Sac。假设 S 是初始准则函数值，即未进行合并时的准则函数值。判断，Sab，Sac，S 这三个的大小。选取最大值所对应的方案。这就是合理的方案。步骤 2.5) 实际上是选取可行方案中最合理方案。

[0099] 2.6)、结束。

[0100] 当原始模型建立，并完成学习形成最终模型时，进入步骤 (3) 利用模型进行用户推荐。3)、利用模型进行用户推荐

[0101] 3.1)、用户产生新的点击数据，记录该点击数据用户下一次模型的学习。

[0102] 3.2)、确定该用户所在的类别，并得到相应的马尔可夫模型，包括转移矩阵和初始状态。若用户为新用户，则利用背景信息相似的用户生产马尔可夫模型；

[0103] 3.3)、根据用户当前的点击数据以及相应的马尔可夫模型得到 top-N 的推荐，并显示给用户；

[0104] 3.4)、结束。

[0105] 其中步骤 1.1) 中，记录并提取用户点击流信息，所述点击流是基于控件的点击

流。基于控件的点击流信息相比于基于页面的点击流可以更加精确的记录用户的行为以及兴趣。

[0106] 步骤 2.3) 所述的利用用户关系矩阵寻找背景相似用户，步骤如下：

[0107] 2.3.1)、得到使用当前马尔可夫模型的类别中的所有 n 个用户；

[0108] 2.3.2)、对于每个用户，在用户关系矩阵中寻找相似度最大的 m 个用户；

[0109] 2.3.3)、得到 n 组 m 个用户，将其合并后得到相似用户集合；

[0110] 2.3.4)、结束。

[0111] 步骤 3.2) 利用用户所在类别的马尔可夫模型进行推荐，当用户是新用户时：利用 2.3) 与 2.4) 来计算当前用户的预测马尔可夫模型，并根据此模型进行推荐。

[0112] 步骤 3.3) 根据当前点击数据以及相应的马尔可夫模型得到 top-N 的推荐，步骤如下：

[0113] 3.3.1)、确定当前用户现在点击的页面为 $X_t = x_g$，以及所处的马尔可夫模型的转移矩阵为 $A_u = (p_{u-ij})$；

[0114] 3.3.2)、取得 $A_u$ 中的 g 行，即状态 $x_g$ 的行，$(p_{u-gj}) = (p_{u-g1}, p_{u-g2}, \cdots\cdots p_{u-gn})$；

[0115] 3.3.3)、对所有的 $p_{u-gj}$，$0 < i <= n$，进行降序排列为，$p_{u-gn1}, p_{u-gn2}, \cdots\cdots p_{u-gnn}$；

[0116] 3.3.4)、设定推荐前 N 个内容为最热门内容，取前 N 个 $p_{u-gnj}$：为 $p_{u-gn1}, p_{u-gn2}, \cdots\cdots p_{u-gnN}$，那么所对应页面的 $X_{t+1} = x_{n1}, X_t = x_{n2}, \cdots\cdots X_t = x_{nN}$，即为所推荐的最热门内容。

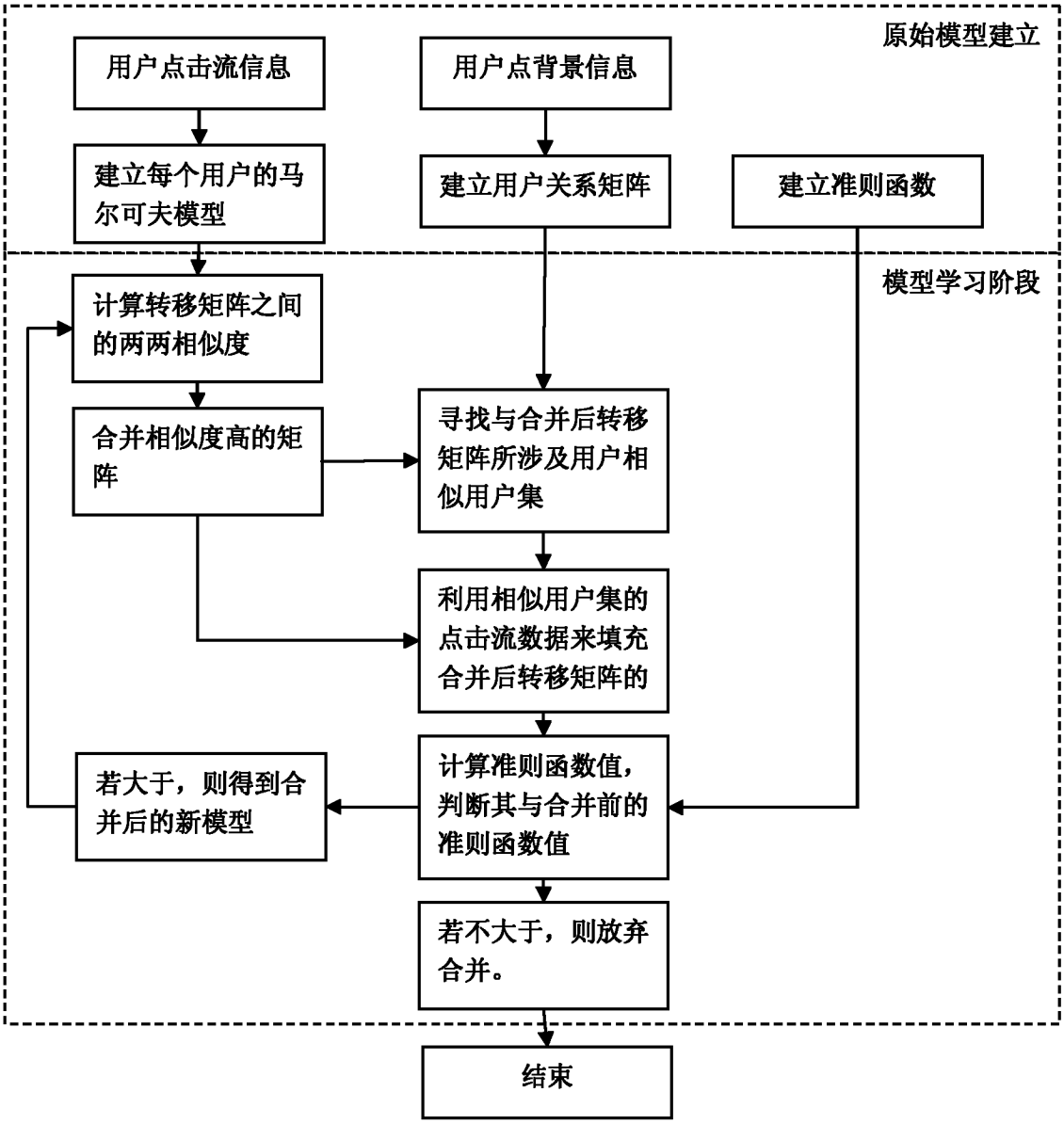

原始模型建立
用户点击流信息
用户点背景信息
建立每个用户的马尔可夫模型
建立用户关系矩阵
建立准则函数
模型学习阶段
计算转移矩阵之间的两两相似度
合并相似度高的矩阵
寻找与合并后转移矩阵所涉及用户相似用户集
利用相似用户集的点击流数据来填充合并后转移矩阵的
计算准则函数值，判断其与合并前的准则函数值
若大于，则得到合并后的新模型
若不大于，则放弃合并。
结束

图 1