(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 111737588 B
(45)授权公告日 2021.01.08

(21)申请号 202010856930.2
(22)申请日 2020.08.24
(65)同一申请的已公布的文献号
申请公布号 CN 111737588 A
(43)申请公布日 2020.10.02
(73)专利权人 南京国睿信维软件有限公司
地址 210013 江苏省南京市鼓楼区古平岗4号院53号楼7楼
(72)发明人 曹保龙 彭天颖 王磊 卢浩然 周苏霞
(74)专利代理机构 南京苏创专利代理事务所(普通合伙) 32273
代理人 凤婷
(51)Int.Cl.
G06F 16/9535(2019.01)
G06F 17/16(2006.01)
G06K 9/62(2006.01)

审查员 赵小娟

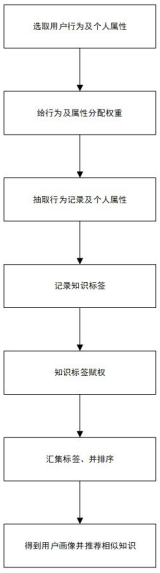
权利要求书2页 说明书4页 附图1页

(54)发明名称

用户画像知识相似度计算方法

(57)摘要

本发明公开了用户画像知识相似度计算方法，属于智能分析技术领域，首先，规定用户的重要行为及个人基础属性选取个数及以及具体行为操作和属性内容。其次，对每项用户行为及用户属性的进行参数赋权，决定用户行为和用户属性在整体知识相似度计算中的比重值。抽取用户的行为记录和属性，统计用户行为访问过知识关联的标签，并对标签一一赋权。最后，汇集用户知识标签的总体权重值，根据权重高低排序，得出用户知识画像。本发明基于用户行为操作及个人基础属性的参数权重计算知识相似度，通过用户的具体行为操作绘制用户画像，并相应地返回适配的相似知识。该方法的知识推荐准确度高，推荐比例可以动态调整，操作便捷。



CN 111737588 B

1.用户画像知识相似度计算方法，其特征在于：包括以下步骤：

步骤一：统计影响用户画像的标准行为及基础属性；

步骤二：给标准行为及基础属性分配权重，标准行为选取下载、提问、收藏、分享、评论5种行为操作，所述基础属性选取部门和岗位，合计7个参数，设定每个参数的权重，依次记为$\lambda_1,\lambda_2,\lambda_3,\lambda_4,\lambda_5,\lambda_6,\lambda_7$；

步骤三：抽取标准行为记录及基础属性标签，记录用户行为日志表，记录用户标准行为所关联的行为标签，记录用户基础属性关联的属性标签，行为标签和属性标签合并称为知识标签；

步骤四：根据知识标签，整理成标签矩阵，知识标签依次用大写字母表示，得到知识标签矩阵A_{nm}、B_{nm}、C_{nm}、D_{nm}、E_{nm}、F_{nm}、G_{nm}；其中n表示标准行为和基础属性对应的知识个数，m表示每个知识设定的选取关联标签的个数；

其中，标签矩阵A_{nm}的具体获得过程为：

用户的标准行为A的知识有n个，$0 \le n \le \infty$，每个知识有m个标签，得到一个知识相关联的标签矩阵：$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix}$，将其记为标签矩阵$A_{nm}$，其中，$a_{ij}$为标准行为A的第$i$个知识的第$j$个标签；

取每个知识最多关联5个标签，即$0 \le m \le 5$，当标准行为A的第i个知识的标签个数小于5时，则超过实际标签数的a_{ij}为空值；

步骤五：对标签矩阵内的每个标签赋权，得到加权矩阵$\lambda_1 A_{nm}$、$\lambda_2 B_{nm}$、$\lambda_3 C_{nm}$、$\lambda_4 D_{nm}$、$\lambda_5 E_{nm}$、$\lambda_6 F_{nm}$、$\lambda_7 G_{nm}$，

其中，$\lambda_1 A_{nm} = \lambda_1 \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} \lambda_1 a_{11} & \cdots & \lambda_1 a_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_1 a_{n1} & \cdots & \lambda_1 a_{nm} \end{bmatrix}$

其中每一项标签都表现为$\lambda_1 a_{ij}$，1≤i≤n，1≤j≤m；

步骤六：汇集知识标签，分析并合并所有加权矩阵，统一标签计量维度，得到用户画像并推荐相似知识，具体过程为：

步骤a：记录标签a_{11}为$t_{1,1}$，记录标签a_{12}为$t_{1,2}$，……，记录标签a_{1m}为$t_{1,m}$，……，记录标签a_{nm}为$t_{1,nm}$，当遇到重复标签时，仅增加知识权重，不新增标签个数，

统计标签矩阵A_{nm}内的标签权重，标签矩阵A_{nm}内最多有nm个标签，其中标签$t_{1,1}$出现$k_{1,1}$次，标签$t_{1,2}$出现$k_{1,2}$次，……，标签$t_{1,nm}$出现$k_{1,nm}$次，推算出，标签$t_{1,1}$的权重为$\frac{k_{1,1}}{\sum_{i=1}^{nm} k_{1,i}} \times \lambda_1 = \lambda_{1,1}$，整个标签矩阵$A_{nm}$内所有标签的权重之和为$\left(\sum_{j=1}^{nm} \frac{k_{1,j}}{\sum_{i=1}^{nm} k_{1,i}}\right) \times \lambda_1 = \lambda_{1,1} + \lambda_{1,2} + \cdots + \lambda_{1,nm} = \lambda_1$，保证权重的总和始终是对该操作预设的权重值，标准行为A对应的标签数组$\{t_{1,1}, t_{1,2}, \ldots\ldots, t_{1,nm}\}$中的各个标签对应的权重为$\{\lambda_{1,1}, \lambda_{1,2}, \ldots\ldots, \lambda_{1,nm}\}$，这是一个1×$nm$的一维矩阵记为$\Lambda_{1,nm}$，其权重和为$\lambda_1$；

步骤b：计算所有知识标签的权重，得到7个1×nm的矩阵，将其组合成一个7×nm的权重矩阵$\Lambda_{7,nm}$；

步骤c：对照标签矩阵，整理重复标签，若a_{ij}代表的标签与b_{xy}代表的标签和d_{uv}代表的标签相同，那么标签a_{ij}的最终权重为标签a_{ij}的权重与b_{xy}的权重和d_{uv}的权重的总和，b_{xy}化为0，d_{uv}化为0；

步骤d：对矩阵$\Lambda_{1,nm}$进行矩阵化简，得到一个最简行阶梯型矩阵：$\Lambda_{7,nm}=\begin{bmatrix}\lambda_{1,1} & \lambda_{1,2} & \cdots & 0\\ \lambda_{2,1} & \lambda_{2,2} & & 0\\ \vdots & & \ddots & \vdots\\ 0 & & \cdots & 0\end{bmatrix}$，并排列大小，整理对应的标签；

步骤e：按照标签权重的大小，从高到低，梳理对应标签，组成权重与标签向量组$(\lambda_{p,q},t_{p,q})$，$(\lambda_{p+1,q+1},t_{p+1,q+1})$，……，其中，$\lambda_{p,q}>\lambda_{p+1,q+1}>\cdots\cdots$；

得到一系列按照权重高低排列的标签数组，标签数组就是用户行为及基础信息综合权重得到的用户画像标签。

用户画像知识相似度计算方法

技术领域

[0001] 本发明涉及一种用户画像知识相似度计算方法，属于智能分析技术领域。

背景技术

[0002] 当今是一个海量数据的时代，用户在各网站都能接收到大量信息。这其中有很大一部分都是无效、重复、或者用户不感兴趣的垃圾流量。用户在查询知识时需要花费大量的时间、精力来遍历搜索结果，或在茫茫知识库中打捞。为了向用户精准投放符合用户行为习惯及个人特征的相似知识，需要一种能够动态调整权重比例，并按照单篇知识的标签进行统计，得出用户知识画像和个人知识标签，进行动态知识相似推荐。

发明内容

[0003] 为了解决上述技术问题，本发明提供一种用户画像知识相似度计算方法，其具体技术方案如下：

[0004] 用户画像知识相似度计算方法，其特征在于：包括以下步骤

[0005] 步骤一：统计影响用户画像的标准行为及基础属性，所述标准行为包括查看、点赞、分享、收藏、下载、评论和提问，所述基础属性包括部门、岗位、角色和专业；

[0006] 步骤二：给标准行为及基础属性分配权重，设定每个参数的权重λ，并记为$\{\lambda\}_n=\{\lambda_1,\lambda_2,\ldots\ldots,\lambda_n\}$，具体为：

[0007] 所述标准行为选取下载、提问、收藏、分享、评论5种行为操作，所述基础属性选取部门和岗位，合计7个参数；

[0008] 每个参数的权重$\{\lambda\}_n=\{\lambda_1,\lambda_2,\ldots\ldots,\lambda_n\}$可表示为$\{\lambda\}_7=\{\lambda_1,\lambda_2,\ldots\ldots,\lambda_7\}$；

[0009] 标签矩阵具体可表示为：A_{nm}、B_{nm}、C_{nm}、D_{nm}、E_{nm}、F_{nm}、G_{nm}；

[0010] 加权矩阵具体为：λ_1A_{nm}，λ_2B_{nm}，λ_3C_{nm}，λ_4D_{nm}，λ_5E_{nm}，λ_6F_{nm}，λ_7G_{nm}；

[0011] 步骤三：抽取标准行为记录及基础属性标签，记录用户行为日志表，记录用户标准行为所关联的行为标签，记录用户基础属性关联的属性标签，行为标签和属性标签合并称为知识标签；

[0012] 步骤四：根据知识标签，整理成标签矩阵，标准行为依次用大写字母表示，得到标签矩阵A_{nm}、B_{nm}、C_{nm}、$D_{nm}\cdots\cdots$，其中n表示标准行为的对象个数，m表示设定的选取关联标签的个数；

[0013] 标签矩阵A_{nm}的具体获得过程为：

[0014] 用户的标准行为A的对象有n个，每个对象有m个标签，得到一个知识相关联的标签矩阵：$\begin{bmatrix}a_{11} & \cdots & a_{1m}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nm}\end{bmatrix}$，将其记为矩阵$A_{nm}(0\le n\le\infty,0\le m\le\infty)$，其中，$a_{ij}$为标准行为A的第$i$个对象的第$j$个标签；

[0015] 取每个对象最多关联5个标签，即$0\le m\le 5$，当对象i的标签数小于5时，则超过对

象i标签数的a_{ij}为空值；

[0016] 步骤五：对标签矩阵内的每个标签赋权，得到加权矩阵$\lambda_1 A_{nm}$、$\lambda_2 B_{nm}$、$\lambda_3 C_{nm}$、$\lambda_4 D_{nm}\cdots\cdots$，

[0017] $$\lambda_1 A_{nm} = \lambda_1 \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} \lambda_1 a_{11} & \cdots & \lambda_1 a_{1m} \\ \vdots & \ddots & \vdots \\ \lambda_1 a_{n1} & \cdots & \lambda_1 a_{nm} \end{bmatrix}$$

[0018] 其中每一项标签都可以表现为$\lambda_1 a_{ij}$；

[0019] 步骤六：汇集知识标签，分析并合并所有加权矩阵，统一标签计量维度，得到用户画像并推荐相似知识，具体过程为：

[0020] 步骤a：记录标签a_{11}为对象t_1，记录标签a_{12}为对象t_2，……，记录标签a_{1n}为对象t_n，当遇到重复标签时，仅增加对象权重，不新增标签对象，

[0021] 统计矩阵A_{nm}内的标签权重，矩阵A_{nm}内最多有nm个标签，其中对象t_1出现k_{11}次，对象t_2出现k_{12}次，……，对象t_n出现k_{1n}次，推算出对象t_1的权重为$\frac{k_{11}}{\sum_{i=1}^{nm} k_i} \times \lambda_1$，整个矩阵$A_{nm}$内所有标签的权重之和为$\sum_{j=1}^{nm} \frac{k_{1j}}{\sum_{i=1}^{nm} k_{1i}} \times \lambda_1 = \lambda_1$，保证权重的总和始终是对该操作预设的权重值，标准行为A对应的标签数组$\{t_1,\ t_2,\ \ldots\ldots,\ t_{nm}\}$对应权重$\{k_{11},\ k_{12},\ \ldots\ldots,\ k_{1,nm}\}$，这是一个$1\times nm$的矩阵，其和为$\lambda_1$；

[0022] 步骤b：计算所有标准行为下，所有标签的权重，得到7个$1\times nm$的矩阵，将其组合成一个$7\times nm$的权重矩阵$K_{7,nm}$；

[0023] 步骤c：对照标签矩阵，整理重复标签，若a_{ij}代表的标签与b_{xy}和d_{pq}相同，被记为对象t_p，那么该标签的总权重就为$k_{ij}+k_{xy}+k_{pq}$，原k_{xy}化为0，k_{pq}也化为0，空标签不记录；

[0024] 步骤d：对权重矩阵$K_{7,nm}$进行矩阵化简，得到一个最简行阶梯型矩阵：$\begin{pmatrix} k_{11} & k_{12} & \cdots & 0 \\ k_{21} & k_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & 0 \end{pmatrix}$，并排列大小，整理对应的标签；

[0025] 步骤e：按照k_{ij}的大小，从高到低，梳理对应标签，组成向量组：$(k_{pq}, t_{pq}),\ (k_{p+1,q+1}, t_{k_{p+1,q+1}}), \cdots (k_{pq} > k_{p+1,q+1} > k_{p+2,q+2} \ldots)$，得到一系列按照权重高低排列的标签数组，标签数组就是用户行为及基础信息综合权重得到的用户画像标签。

[0026] 本发明的有益效果是：

[0027] 本发明通过对用户的行为操作及个人属性进行分析，记录操作相关知识的标签，并动态赋权，得到实时更新的用户知识画像。通过用户画像知识相似度的计算方法，量化用户对知识的关注度，明确用户的知识标签占比，为精准投放符合用户行为习惯及个人特征的相似知识提供了方法论。

附图说明

[0028] 图1是本发明的流程图。

具体实施方式

[0029] 如图1所示，本发明的用户画像知识相似度计算方法，包括以下步骤：

[0030] 步骤一：统计影响用户画像的标准行为及基础属性，标准行为包括查看、点赞、分享、收藏、下载、评论、提问，所述基础属性包括部门、岗位、角色、专业；本发明选取了用户的下载、提问、收藏、分享、评论5种标准行为(行为操作)及部门、岗位两种用户基础属性，合计7个参数。

[0031] 步骤二：给标准行为及基础属性分配权重，设定每个参数的权重λ，并记为$\{\lambda\}_n=\{\lambda_1,\lambda_2,\ldots\ldots,\lambda_n\}$，本发明共有7项计算值，取$\{\lambda_1,\lambda_2,\ldots\ldots,\lambda_7\}$。

[0032] 步骤三：抽取标准行为记录及基础属性标签，记录用户行为日志表，记录用户标准行为所关联的行为标签，记录用户基础属性关联的属性标签，行为标签和属性标签合并称为知识标签；

[0033] 步骤四：根据知识标签，整理成标签矩阵，标准行为依次用大写字母表示，得到标签矩阵A_{nm}、B_{nm}、C_{nm}、$D_{nm}\cdots\cdots$，其中n表示标准行为的对象个数，m表示设定的选取关联标签的个数。

[0034] 以用户的标准行为(行为操作)A为例(下载操作)，用户行为操作A的对象有n个(下载n篇知识)，每篇知识有m个标签，那么可以得到一个知识相关联的标签矩阵：$\begin{bmatrix}a_{11} & \cdots & a_{1m}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nm}\end{bmatrix}$，将其记为矩阵$A_{nm}$ $(0\le n\le\infty,0\le m\le 5)$(每项知识最多关联5个标签)。其中，$a_{ij}$为操作A的第$i$篇知识的第$j$个标签。如果知识$i$只有3个标签，则$a_{i4}$，$a_{i5}$都为空值。

[0035] 步骤五：对矩阵内的每个标签赋权，得到加权矩阵λ_1A_{nm}，

[0036] $$\lambda_1A_{nm}=\lambda_1\begin{bmatrix}a_{11} & \cdots & a_{1m}\\ \vdots & \ddots & \vdots\\ a_{n1} & \cdots & a_{nm}\end{bmatrix}=\begin{bmatrix}\lambda_1a_{11} & \cdots & \lambda_1a_{1m}\\ \vdots & \ddots & \vdots\\ \lambda_1a_{n1} & \cdots & \lambda_1a_{nm}\end{bmatrix}$$

[0037] 其中每一项标签都可以表现为λ_1a_{ij}。

[0038] 对余下的4个操作及两种基础信息做同样的赋权工作，一共得到7个加权矩阵：λ_1A_{nm}，λ_2B_{nm}，λ_3C_{nm}，λ_4D_{nm}，λ_5E_{nm}，λ_6F_{nm}，λ_7G_{nm}。

[0039] 步骤六：汇集知识标签，分析并合并所有加权矩阵，统一标签计量维度，得到用户画像并推荐相似知识：

[0040] 步骤a：记录标签a_{11}为对象t_1，记录标签a_{12}对象t_2，以此类推。当遇到重复标签时，仅增加对象权重，不新增标签对象。首先统计矩阵A_{nm}内的标签权重。矩阵A_{nm}内最多有nm个标签，其中标签t_1出现k_{11}次，标签t_2出现k_{12}次……我们可以推算出，标签t_1的权重为$\frac{k_{11}}{\sum_{i=1}^{nm}k_i}\times\lambda_1$。整个矩阵$A_{nm}$内所有标签的权重之和为：

[0041] $\sum_{j=1}^{nm}\frac{k_{1j}}{\sum_{i=1}^{nm}k_{1i}}\times\lambda_1=\lambda_1$。保证权重的总和始终是对该操作预设的权重值。操作A对应的标签数组$\{t_1,\ t_2,\ \ldots\ldots,\ t_{nm}\}$对应权重$\{k_{11},\ k_{12},\ \ldots\ldots,\ k_{1,nm}\}$，这是一个$1\times nm$的矩阵，其和为$\lambda_1$。

[0042] 步骤b:计算所有操作下,所有标签的权重,得到7个1×nm的矩阵。将其组合成一个7×nm的权重矩阵$K_{7,nm}$。

[0043] 步骤c:对照标签矩阵,整理重复标签。例如:a_{ij}代表的标签与b_{xy}和d_{pq}相同,被记为对象t_p,那么该标签的总权重就为$k_{ij}+k_{xy}+k_{pq}$。原k_{xy}化为0,k_{pq}也化为0。空标签不记录。

[0044] 步骤d:对矩阵$K_{7,nm}$进行矩阵化简,得到一个最简行阶梯型矩阵:$\begin{pmatrix} k_{11} & k_{12} & \cdots & 0 \\ k_{21} & k_{22} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & 0 \end{pmatrix}$,并排列大小,整理对应的标签。

[0045] 步骤e:按照k_{ij}的大小,从高到低,梳理对应标签,组成向量组:(k_{pq},t_{pq}),$\left(k_{p+1,q+1},t_{k_{p+1,q+1}}\right),\cdots(k_{pq}>k_{p+1,q+1}>k_{p+2,q+2}\cdots)$,得到一系列按照权重高低排列的标签数组。

[0046] 这组标签数组就是用户行为及基础信息综合权重得到的用户画像标签。

[0047] 下面以用户张三为例:

[0048] 用户张三在管理员设定的30天里。

[0049] 下载了文档中含有标签A_1 5次,A_2 6次,A_3 2次,A_4 2次,A_5 1次,$A_6$1次。下载的权重设为5,

[0050] 那么下载相关标签中,A_1的权重为:$\frac{5}{(5+6+2+2+1+1)}\times 5=\frac{25}{17}$,$A_2$的权重为$\frac{30}{17}$,$A_3$的权重为$\frac{10}{17}$,$A_4$的权重为$\frac{10}{17}$,$A_5$的权重为$\frac{5}{17}$,$A_6$的权重为$\frac{5}{17}$。

[0051] 提问中含有标签$B_1$3次,提问的权重为4,

[0052] 那么提问相关标签中,B_1的权重为$\frac{3}{3}\times 4=4$,

[0053] …

[0054] 部门标签中,权重为10,

[0055] 含有标签F_1、F_2,标签F_1权重为5,F_2权重为5,

[0056] 下载重点标签A_2和提问的标签B_1重复,计算为一个标签,总权重为$\frac{98}{17}$;部门标签中F_1与A_5重复,总权重为$\frac{90}{17}$;部门标签中F_2与A_6重复,总权重为5。

[0057] 用户张三的个人标签,权重由高到低排列为:$\left(A_2,\ \frac{98}{17}\right),\left(A_5,\ \frac{90}{17}\right),(A_6,\ 5),\left(A_1,\ \frac{25}{17}\right),\left(A_4,\ \frac{10}{17}\right),\left(A_3,\ \frac{10}{17}\right),\left(A_7,\ \frac{5}{17}\right),\cdots\cdots$

[0058] 根据得出的标签权重,系统进行赋权搜索,得出基于用户画像的知识相似推荐。

[0059] 以上述依据本发明的理想实施例为启示,通过上述的说明内容,相关工作人员完全可以在不偏离本项发明技术思想的范围内,进行多样的变更以及修改。本项发明的技术性范围并不局限于说明书上的内容,必须要根据权利要求范围来确定其技术性范围。

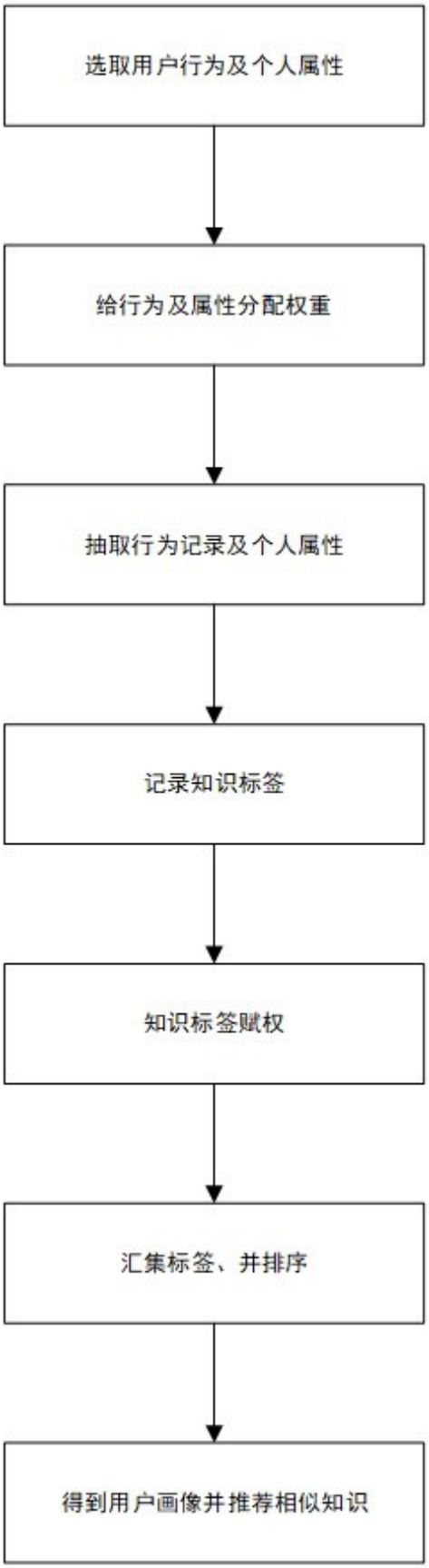

选取用户行为及个人属性
给行为及属性分配权重
抽取行为记录及个人属性
记录知识标签
知识标签赋权
汇集标签、并排序
得到用户画像并推荐相似知识

图1