



(12) 发明专利申请

(10) 申请公布号 CN 102880691 A

(43) 申请公布日 2013. 01. 16

(21) 申请号 201210349196. 6

(22) 申请日 2012. 09. 19

(71) 申请人 北京航空航天大学深圳研究院

地址 518057 广东省深圳市南山区高新技术
南区虚拟大学园 A501 室

(72) 发明人 李超 余建 吕志强

(74) 专利代理机构 北京科迪生专利代理有限责
任公司 11251

代理人 李新华 成金玉

(51) Int. Cl.

G06F 17/30 (2006. 01)

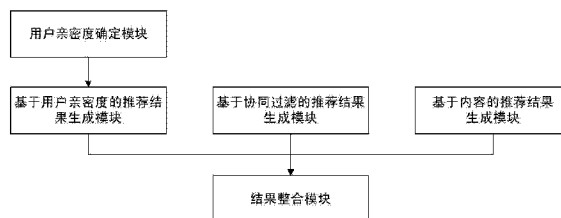
权利要求书 2 页 说明书 5 页 附图 5 页

(54) 发明名称

一种基于用户亲密度的混合推荐系统及方法

(57) 摘要

一种基于用户亲密度的混合推荐系统及方法,该系统及方法能向社交网站中的用户推荐其感兴趣的项目。该系统由 5 个模块组成:用户亲密度确定模块、基于用户亲密度的推荐结果生成模块、基于协同过滤的推荐结果生成模块、基于内容的推荐结果生成模块、结果整合模块。该方法通过确定用户亲密度、获得基于用户亲密度的推荐结果、获得基于协同过滤的推荐结果、获得基于内容的推荐结果、整合结果 5 个过程实现。该方法充分利用社交网站中的数据,弥补传统推荐系统的不足,使该系统应用具有实用性强、准确度高、以及实现方便的优点。



1. 一种基于用户亲密度的混合推荐系统,其特征在于包括:

用户亲密度确定模块:在生成对用户的推荐结果之前,扫描数据库,根据数据库中保存的用户数据,获取某用户与其他用户的互动行为,统计这些行为进行的次数,确定该用户对其他用户的亲密度,并将亲密度归一化,将结果存储在数据库中;所述用户亲密度是采用社交网站中两个用户之间发生的“@”、转发、评论操作的次数来度量,并且用户亲密度关系属于单向关系,以 $I(A \rightarrow B)$ 表示用户 A 对用户 B 的亲密度,有 $I(A \rightarrow B) \neq I(B \rightarrow A)$;

基于用户亲密度的推荐结果生成模块:根据用户亲密度确定模块的结果,从数据库中获取与该用户亲密度最高的前 N 个用户, N 由本系统的输入指定,合并这些用户所关注的项目作为推荐结果 E_1 ,将该结果保存至数据库,为后续操作提供相关数据;

基于协同过滤的推荐结果生成模块:扫描网站数据库,根据数据库中保存的用户所关注的项目情况,获得该用户与其他用户的相似度,将与该用户相似度最高的前 N 个用户所关注的项目作为推荐结果 E_2 , N 由本系统的输入指定,将该结果保存至数据库,为后续操作提供相关数据;

基于内容的推荐结果生成模块:扫描网站数据库,获得该用户与待推荐项目的描述关键词向量,将与该用户关键词向量相似度最高的前 N 个项目作为推荐结果 E_3 , N 由本系统的输入指定,将该结果保存至数据库,为后续操作提供相关数据;

结果整合模块:从数据库中提取前面三个模块的生成结果,合并成为待推荐项目集合 E,扫描该集合,确定集合中每个项目在 E_1 、 E_2 、 E_3 总共出现的次数,最后将出现次数高于指定次数的项目作为结果推荐给用户,在用户访问网站时进行显示,用户根据推荐的结果选择自己感兴趣的项目进行关注。

2. 根据权利要求 1 所述的一种基于用户亲密度的混合推荐系统及,其特征不在于:所述用户亲密度确定模块实现过程如下:

(1) 选择需要推荐的用户,扫描该用户与其他所有用户的互动记录,确定该用户与其他用户间发生的“@”、转发、评论操作的次数;

(2) 使用以下公式计算用户亲密度:

$$I(A \rightarrow B) = s \cdot s(s) + c \cdot s(c) + a \cdot s(a)$$

其中 $s(s)$ 、 $s(c)$ 、 $s(a)$ 分别代表每进行一次分享、评论、“@”操作对亲密度得分的贡献值,由系统输入确定; s 、 c 、 a 分别代表该用户对另一用户的分享、评论、“@”操作的次数,求得亲密度之后,将其保存下来,用作下一步的归一化处理;

(3) 使用以下公式对亲密度进行归一化:

$$N(A \rightarrow B) = (I(A \rightarrow B) - I_{\min}(A \rightarrow B)) / (I_{\max}(A \rightarrow B) - I_{\min}(A \rightarrow B))$$

其中 $N(A \rightarrow B)$ 是归一化处理结果, $I(A \rightarrow B)$ 是用户 A 对当前用户的亲密度得分, $I_{\min}(A \rightarrow B)$ 是该用户对其他用户亲密度计算中得分的最小值, $I_{\max}(A \rightarrow B)$ 该用户对其他用户亲密度计算中得分的最大值;亲密度归一化完毕后,将结果输出保存至网站数据库中。

3. 一种基于用户亲密度的混合推荐方法,其特征不在于步骤如下:

(1) 当需要向用户推荐项目时,扫描网站数据库,获得该用户与其他用户的分享、评论、“@”操作次数,然后根据操作次数以及设定的操作贡献值计算用户与其他用户的亲密度,并将获得的用户亲密度进行归一化,然后将归一化结果保存在数据库中;

(2) 从数据库选取亲密度最高的前 N 个用户, N 由本系统的输入指定,获得这些用户

所关注的项目,将这些项目合并成为没有重复项的集合,将该集合作为基于用户亲密度的推荐结果集保存于数据库中;

(3) 使用基于协同过滤的推荐结果生成获得基于协同过滤的推荐结果集,将其保存于数据库中;

(4) 使用基于内容的推荐结果生成模块获得基于内容的推荐结果集,将其保存于数据库中;

(5) 从数据库中获得前面三个结果集,合并成为没有重复项的待推荐结果集,扫描该结果集中的项目,计算出每个项目在三个集合中的出现总次数,将总次数高于指定次数的项目作为结果推荐给用户,在用户访问网站时进行显示,用户根据推荐的结果选择自己感兴趣的项目进行关注。

一种基于用户亲密度的混合推荐系统及方法

技术领域

[0001] 本发明一种基于用户亲密度的混合推荐系统及方法,属于计算机领域网络数据挖掘领域。

背景技术

[0002] 社交网站 (Social Network Sites, 即 SNS) 是一种线上的方便人与人之间进行社交的网站, SNS 的用户可以在线上发布照片、状态、日志等, 其他用户可以对这些进行评论、转发等操作, 增进人与人之间交流和沟通, 从而达到社交的目的。当前的 SNS 普遍都具有推荐功能, 推荐功能的目的是向用户推荐其感兴趣并可能接受的项目。一方面, 这提高了用户体验, 使用户能够更快找到自己感兴趣的项目, 另一方面, 从商业角度考虑这能达到 SNS 营销的目的。因此, 选择一种准确高效的推荐方法至关重要。然而, 由于 SNS 数据稀疏、散乱等特点, 单纯使用基于协同过滤的推荐或基于内容的推荐所获得的效果较差。根据 SNS 数据产生的一个特点, 即较大部分数据产生于用户互动的分享操作、评论操作、“@”操作的过程中。通过网站的数据库, 可以很容易的获得用户间产生的分享、评论、“@”操作的频率。用户互动的频率与程度反映了用户之间亲密度的不同, 依据用户间的亲密度不同, 可以在成对的用户间进行推荐。将基于协同过滤、基于内容以及基于用户亲密度的推荐结合成一个混合推荐方法, 更大程度地利用 SNS 数据向用户推荐更合适的内容。

发明内容

[0003] 本发明的技术解决问题: 克服现有技术的不足, 提供一种基于用户亲密度的混合推荐系统及方法, 充分利用社交网站中的数据, 通过对用户间互动操作的分析, 获得用户相互的亲密度, 通过亲密度关系来进行推荐, 并结合已有的基于协同过滤的推荐和基于内容的推荐, 使系统应用具有实用性强、准确度高、以及实现方便的优点。

[0004] 本发明的技术方案: 一种基于用户亲密度的混合推荐系统及方法, 其充分利用 SNS 的数据进行推荐, 如图 1 所示包括:

[0005] 用户亲密度确定模块: 在生成对用户的推荐结果之前, 扫描数据库, 根据数据库中保存的用户数据, 获取某用户与其他用户的互动行为, 统计这些行为进行的次数, 确定该用户对其他用户的亲密度, 并将亲密度归一化, 将结果存储在数据库中; 所述用户亲密度是采用社交网站中两个用户之间发生的“@”、转发、评论操作的次数来度量, 并且用户亲密度关系属于单向关系, 以 $I(A \rightarrow B)$ 表示用户 A 对用户 B 的亲密度, 有 $I(A \rightarrow B) \neq I(B \rightarrow A)$;

[0006] 基于用户亲密度的推荐结果生成模块: 根据用户亲密度确定模块的结果, 从数据库中获取与该用户亲密度最高的前 N 个用户, N 由本系统的输入指定, 合并这些用户所关注的项目作为推荐结果 E_1 , 将该结果保存至数据库, 为后续操作提供相关数据;

[0007] 基于协同过滤的推荐结果生成模块: 扫描网站数据库, 根据数据库中保存的用户所关注的项目情况, 获得该用户与其他用户的相似度, 将与该用户相似度最高的前 N 个用户所关注的项目作为推荐结果 E_2 , N 由本系统的输入指定, 将该结果保存至数据库, 为后续

操作提供相关数据；

[0008] 基于内容的推荐结果生成模块：扫描网站数据库，获得该用户与待推荐项目的描述关键词向量，将与该用户关键词向量相似度最高的前 N 个项目作为推荐结果 E_3 ，N 由本系统的输入指定，将该结果保存至数据库，为后续操作提供相关数据；

[0009] 结果整合模块：从数据库中提取前面三个模块的生成结果，合并成为待推荐项目集合 E，扫描该集合，确定集合中每个项目在 E_1 、 E_2 、 E_3 总共出现的次数，最后将出现次数高于指定次数的项目作为结果推荐给用户，在用户访问网站时进行显示，用户根据推荐的结果选择自己感兴趣的项目进行关注。

[0010] 所述用户亲密度确定模块实现过程如下：

[0011] (1) 选择需要推荐的用户，扫描该用户与其他所有用户的互动记录，确定该用户与其他用户间发生的“@”、转发、评论操作的次数；

[0012] (2) 使用以下公式计算用户亲密度：

[0013] $I(A \rightarrow B) = s \cdot s(s) + c \cdot s(c) + a \cdot s(a)$

[0014] 其中 $s(s)$ 、 $s(c)$ 、 $s(a)$ 分别代表每进行一次分享、评论、“@”操作对亲密度得分的贡献值，由系统输入确定； s 、 c 、 a 分别代表该用户对另一用户的分享、评论、“@”操作的次数，求得亲密度之后，将其保存下来，用作下一步的归一化处理；

[0015] (3) 使用以下公式对亲密度进行归一化：

[0016] $N(A \rightarrow B) = (I(A \rightarrow B) - I_{\min}(A \rightarrow B)) / (I_{\max}(A \rightarrow B) - I_{\min}(A \rightarrow B))$

[0017] 其中 $N(A \rightarrow B)$ 是归一化处理结果， $I(A \rightarrow B)$ 是用户 A 对当前用户的亲密度得分， $I_{\min}(A \rightarrow B)$ 是该用户对其他用户亲密度计算中得分的最小值， $I_{\max}(A \rightarrow B)$ 该用户对其他用户亲密度计算中得分的最大值；亲密度归一化完毕后，将结果输出保存至网站数据库中。

[0018] 一种基于用户亲密度的混合推荐方法，实现步骤如下：

[0019] (1) 当需要向用户推荐项目时，扫描网站数据库，获得该用户与其他用户的分享、评论、“@”操作次数，然后根据操作次数以及设定的操作贡献值计算用户与其他用户的亲密度，并将获得的用户亲密度进行归一化，然后将归一化结果保存在数据库中；

[0020] (2) 从数据库中选取亲密度最高的前 N 个用户，N 由本系统的输入指定，获得这些用户所关注的项目，将这些项目合并成为没有重复项的集合，将该集合作为基于用户亲密度的推荐结果集保存于数据库中；

[0021] (3) 使用基于协同过滤的推荐结果生成获得基于协同过滤的推荐结果集，将其保存于数据库中；

[0022] (4) 使用基于内容的推荐结果生成模块获得基于内容的推荐结果集，将其保存于数据库中；

[0023] (5) 从数据库中获得前面三个结果集，合并成为没有重复项的待推荐结果集，扫描该结果集中的项目，计算出每个项目在三个集合中的出现总次数，将总次数高于指定次数的项目作为结果推荐给用户，在用户访问网站时进行显示，用户根据推荐的结果选择自己感兴趣的项目进行关注。

[0024] 本发明与现有技术相比的优点在于：

[0025] (1) 本发明考虑了用户间关系对用户推荐的影响，使推荐结果更加具有代表性，增大了用户接受推荐结果的可能性，达到推荐的目的。

[0026] (2) 本发明充分利用 SNS 的数据,从度量上刻画了用户间的关系,丰富推荐结果,一定程度上弥补了 SNS 数据稀疏、离散的特点。

[0027] (3) 本发明降低网站的推荐系统的耦合度,三个推荐模块以松耦合的方式结合,模块间即可单独工作,也可协同工作。当某一模块需要维护时,其他模块仍然能支撑系统运转,不影响功能的连续性,提高了系统的可维护性和可用性。

附图说明

[0028] 图 1 为本发明系统的体系结构图;

[0029] 图 2 为本发明系统中的用户亲密度确定模块的获得用户亲密度得分的实现过程;

[0030] 图 3 为本发明系统中的用户亲密度确定模块的归一化实现过程;

[0031] 图 4 为本发明系统中的基于用户亲密度的推荐结果生成模块的实现过程;

[0032] 图 5 为本发明系统中的结果整合模块实现过程。

具体实施方式

[0033] 下面结合附图对本发明进行详细说明。

[0034] 如图 1 所示,本发明一种基于用户亲密度的混合推荐系统及方法由用户亲密度确定模块、基于用户亲密度的推荐结果生成模块、基于协同过滤的推荐结果生成模块、基于内容的推荐结果生成模块、结果整合模块构成。

[0035] 整个实现过程如下:

[0036] (1) 当需要向用户推荐项目时,扫描网站数据库,获得该用户与其他用户的分享、评论、“@”操作次数,然后根据操作次数以及设定的操作贡献值计算用户与其他用户的亲密度,并将获得的用户亲密度进行归一化,然后将归一化结果保存在数据库中。

[0037] (2) 从数据库中选取出亲密度最高的前 N 个的用户, N 由本系统的输入指定,获得这些用户所关注的项目,将这些项目合并成为没有重复项的集合,将该集合作为基于用户亲密度的推荐结果集保存于数据库中。

[0038] (3) 使用基于协同过滤的推荐结果生成模块获得基于协同过滤的推荐结果集,将其保存于数据库中。

[0039] (4) 使用基于内容的推荐结果生成模块获得基于内容的推荐结果集,将其保存于数据库中。

[0040] (5) 从数据库中获得前面三个结果集合并成为没有重复项的待推荐结果集,扫描该结果集中的项目,计算出每个项目在三个集合中的出现总次数。将总次数高于指定次数的项目作为结果推荐给用户,在用户访问网站时进行显示,用户可以根据推荐的结果选择自己感兴趣的项目进行关注。。

[0041] 上述各模块的具体实现过程如下:

[0042] 1. 用户亲密度确定模块

[0043] 该模块实现过程如图 2、3 所示

[0044] 当需要向用户 A 进行推荐时,设用户间每进行一次分享、评论、“@”的得分分别为 $s(r)=3$, $s(c)=1$, $s(a)=5$, 扫描数据库,根据数据库中保存的用户数据统计用户 A 与其他用户进行的操作次数,获得分享、评论、“@”进行的次数分别为 s、c、a,然后使用:

[0045] $I(A \rightarrow B) = s \cdot s(s) + c \cdot s(c) + a \cdot s(a)$

[0046] 计算用户 A 与其他用户的亲密度, 然后使用:

[0047] $N(A \rightarrow B) = (I(A \rightarrow B) - I_{\min}(A \rightarrow B)) / (I_{\max}(A \rightarrow B) - I_{\min}(A \rightarrow B))$

[0048] 其中 $N(A \rightarrow B)$ 是归一化处理结果, $I(A \rightarrow B)$ 是用户 A 对当前用户的亲密度, $I_{\min}(A \rightarrow B)$ 是用户 A 对其他用户亲密度计算中的最小值, $I_{\max}(A \rightarrow B)$ 用户 A 对其他用户亲密度计算中的最大值。表 1 显示了以上计算的结果, 其中用户分别使用其 ID 进行表示。

[0049] 表 1

[0050]

A	B	a	s	c	$I(A \rightarrow B)$	$N(A \rightarrow B)$
1000154	1000154	0	0	3	9	#
1000154	1016077	0	1	0	1	0
1000154	1514333	0	1	0	1	0
1000154	1675399	0	6	0	6	0.263158
1000154	1760320	0	1	0	1	0
1000154	1760654	0	6	0	6	0.263158
1000154	1760683	0	19	0	19	0.947368
1000154	1774520	3	5	0	20	1
1000154	1774587	0	1	0	1	0
1000154	1774714	0	2	0	2	0.052632
1000154	1774856	0	1	0	1	0
1000154	1774925	1	4	0	9	0.421053

[0051] 2. 基于用户亲密度的推荐结果生成模块

[0052] 该模块实现过程如图 2、3 所示

[0053] 指定系统选取亲密度最大的前三个用户作为用户 A 的亲密度用户, 则用户 A 的亲密度用户为 1774925、1760683、1774520。设用户 1774925 关注了(足球, 台球, 拳击)三个项目, 用户 1760683 关注了(游泳, 台球), 用户 1774520 关注了(运动, 足球), 那么将它们合并为由基于用户亲密度推荐产生的结果集 E_1 , 为(足球, 台球, 拳击, 运动, 游泳)。

[0054] 3. 基于协同过滤的推荐结果生成模块

[0055] 扫描网站数据库, 根据数据库中保存的用户所关注的项目情况, 使用协同过滤方法生成对用户 A 的推荐结果, 该方法属于本领域公知技术, 在此不再赘述。设: 基于协同过滤的推荐结果集 E_2 为(足球, 篮球, 运动, 游泳)

[0056] 4. 基于内容的推荐结果生成模块

[0057] 扫描网站数据库, 使用关键词匹配的方法生成对用户 A 的推荐结果, 该方法属于本领域公知技术, 在此不再赘述。设: 基于内容的推荐结果集 E_3 为(足球, 排球)。

[0058] 5. 结果整合模块

[0059] 从数据库中提取前面三个模块的生成结果 E_1 、 E_2 、 E_3 , 合并为待推荐集合 E, 就有(足球, 篮球, 运动, 游泳, 排球, 台球, 拳击)。扫描 E, 计算 E 中每个项目 I 出现的次数 Rw_I :

[0060] $Rw_I = f_1(I) + f_2(I) + f_3(I)$ 其中 $f_i(I) = \begin{cases} 1 & I \in E_i \\ 0 & I \notin E_i \end{cases} \quad i=1,2,3$

[0061] 表 2 给出了计算结果：

项目	足球	篮球	运动	游泳	排球	台球	拳击
Rw_I	3	1	2	2	1	1	1

[0063] 设系统指定推荐出现次数大于等于 2 次的项目，那么则最后的推荐结果集合为（足球，运动，游泳）。

[0064] 本发明未详细描述的部分属于本领域公知技术。

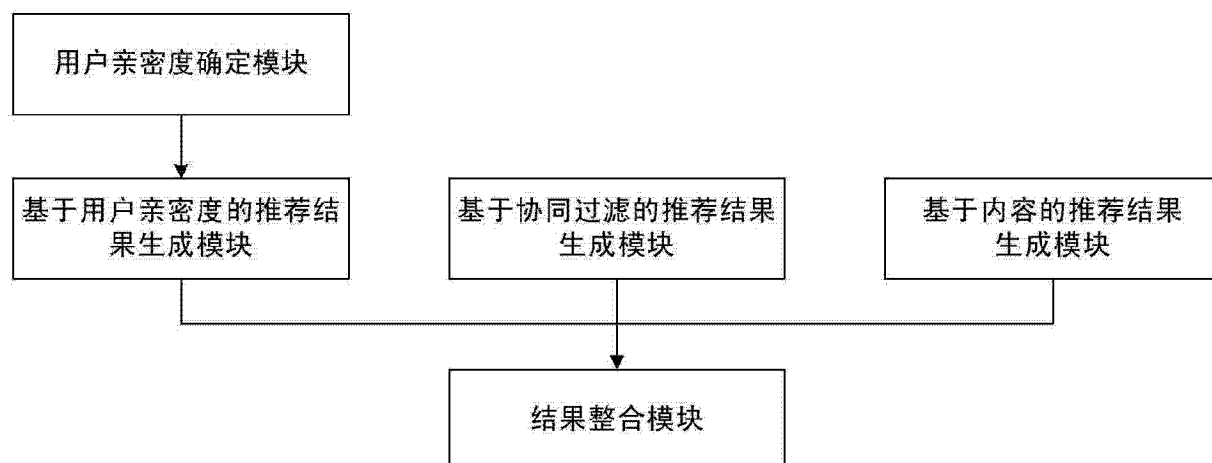


图 1

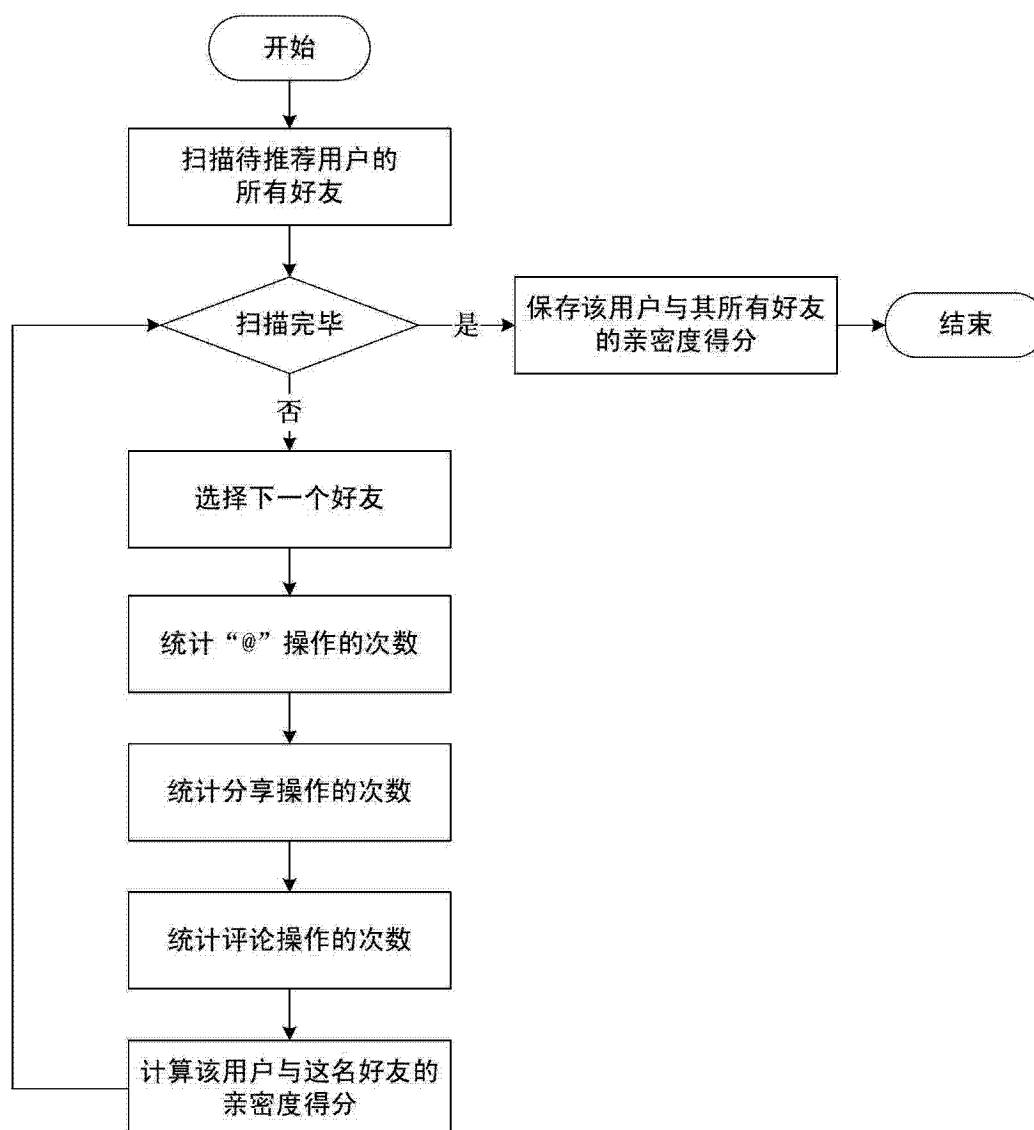


图 2

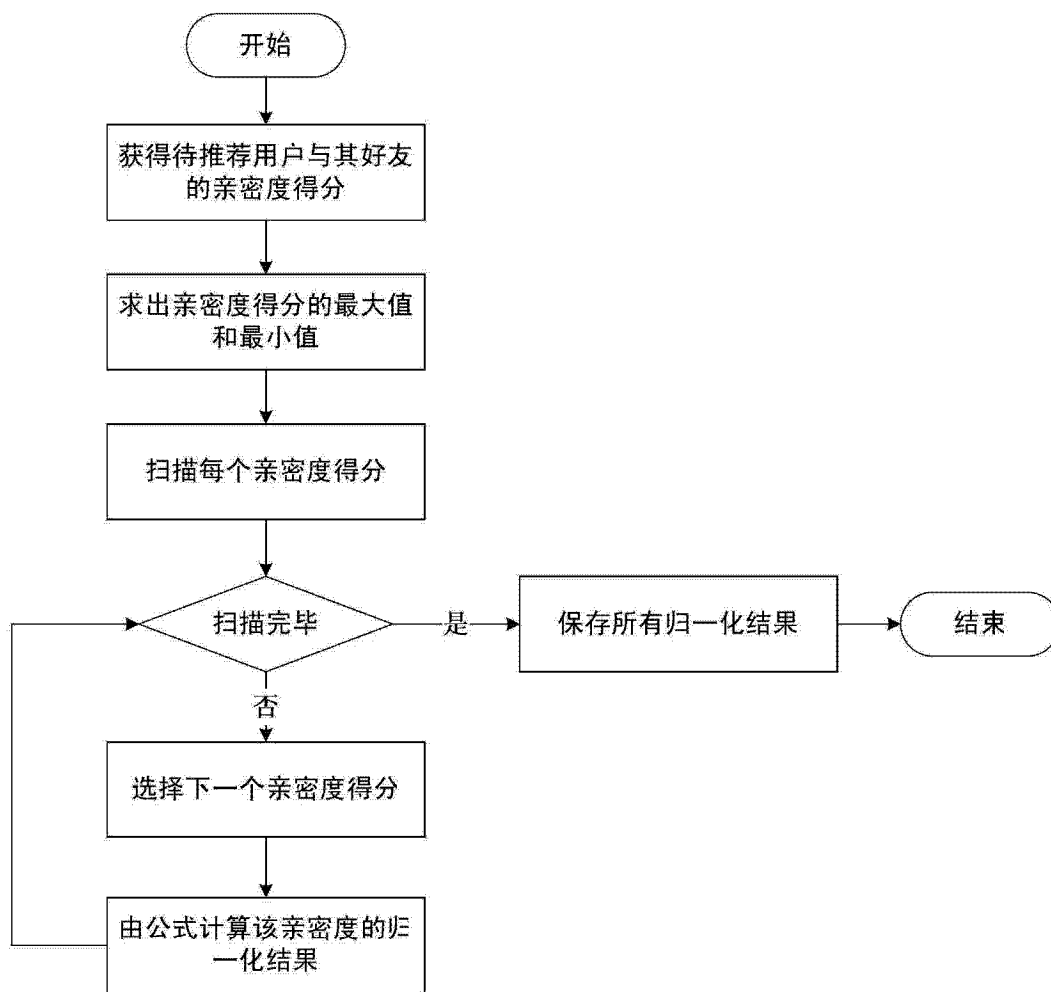


图 3

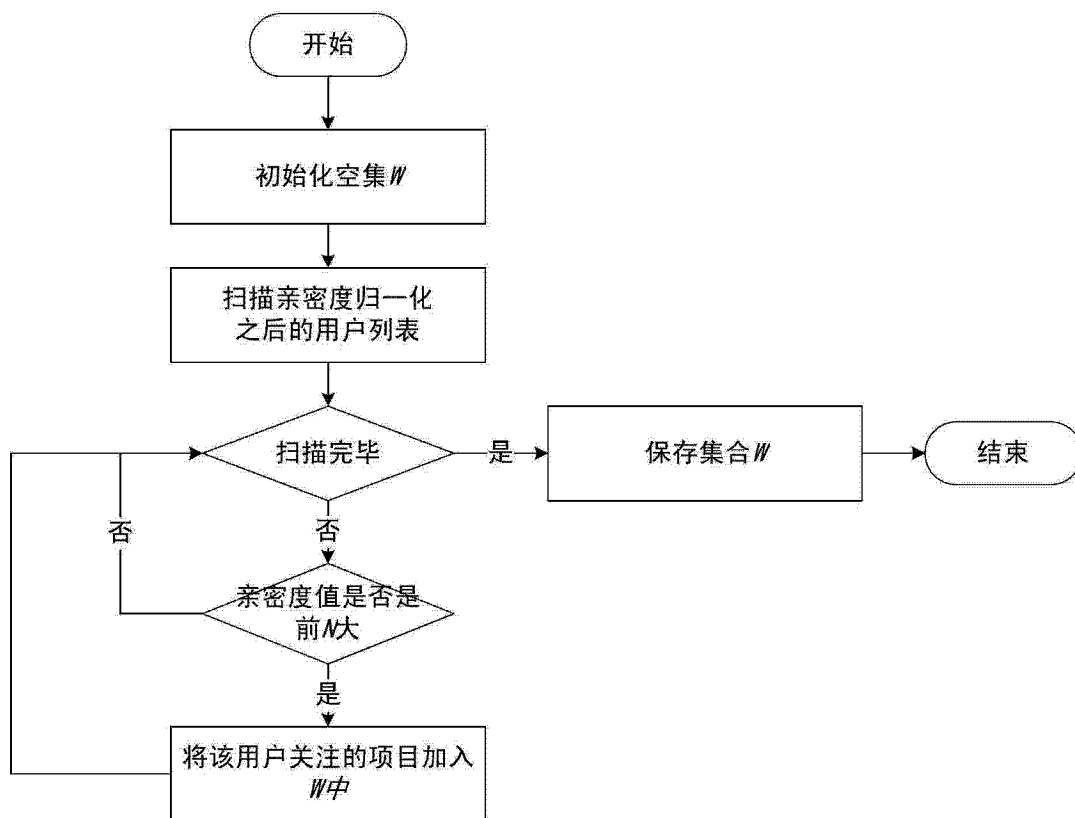


图 4

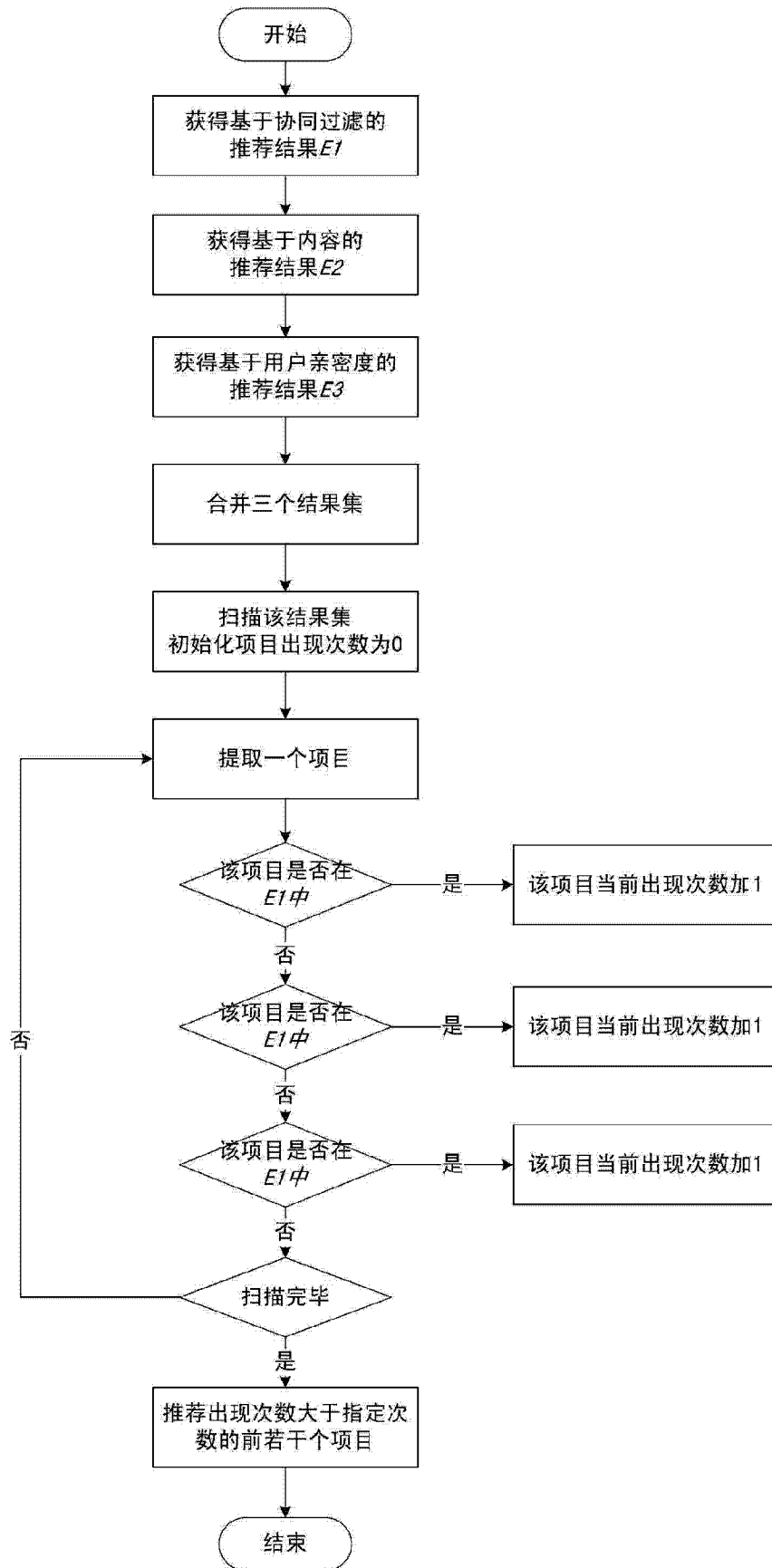


图 5