



(12)发明专利

(10)授权公告号 CN 104268171 B

(45)授权公告日 2017.09.19

(21)申请号 201410462802.4

审查员 刘冰珣

(22)申请日 2014.09.11

(65)同一申请的已公布的文献号

申请公布号 CN 104268171 A

(43)申请公布日 2015.01.07

(73)专利权人 东北大学

地址 110819 辽宁省沈阳市和平区文化路3号巷11号

(72)发明人 于亚新 田宏增 隋鸣飞 续宗泽 王国仁

(74)专利代理机构 沈阳东大知识产权代理有限公司 21109

代理人 梁焱

(51)Int. Cl.

G06F 17/30(2006.01)

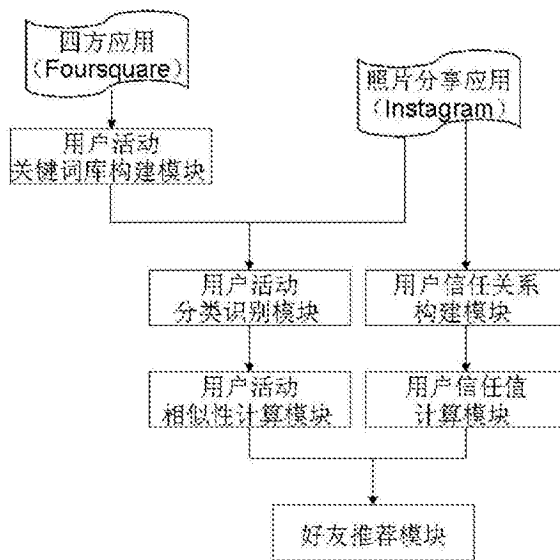
权利要求书3页 说明书15页 附图5页

(54)发明名称

基于活动相似和社交信任的社交网好友推荐系统及方法

(57)摘要

本发明一种基于活动相似和社交信任的社交网好友推荐系统及方法,属于信息推荐和数据挖掘领域,该方法主要是利用用户社交信任值和活动偏好相似性来实现基于位置社交网络中好友推荐,由于活动可以体现用户兴趣偏好,因此通过用户间活动相似性发现与其偏好相似的好友;由于社交信任能反映用户间交互紧密程度,因此根据不同程度信任关系进行好友推荐具有更合理的可解释性;实验证明,本发明推荐效果在准确性和合理解释性上均优于现存的好友推荐方法,实际应用价值很高,如果能得到推广,对企事业单位明确目标客户范围,提高广告服务的关联度和准确性,提高广告营销价值均有重要的指导和决策意义。



1. 一种采用基于活动相似和社交信任的社交网好友推荐系统进行的方法,该系统设置于计算机中,包括用户活动关键词库构建模块、用户活动分类识别模块、用户活动相似性计算模块、用户信任关系构建模块、用户信任值计算模块和好友推荐模块,其中,

用户活动关键词库构建模块:用于根据用户的活动类别和每个活动类别中所对应的词条构建词库;

用户活动分类识别模块:用于提取用户在第三方应用上所发布的信息,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取用户的活动类别集合;

用户活动相似性计算模块:用于根据某类活动对应词条在目标用户发布信息中出现的次数、该类活动对应词条在其他用户发布信息中出现的次数、用户总数和包含该类活动的用户数,获得该类活动相对于目标用户的重要程度,再根据每类活动相对于每个用户的重要程度,计算获得两个用户之间的相似度;

用户信任关系构建模块:用于从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系;若为关注关系,则进一步判断目标用户与其他用户之间为直接关注关系或间接关注关系,如果是直接关注,则确定两者为直接信任关系,如果是间接关注,则确定两者为间接信任关系;若为评价关系和点赞关系,则确定两者为辅助信任关系;

用户信任值计算模块:用于根据目标用户与其他用户之间的交互关系,确定目标用户与其他用户的直接信任值、间接信任值、辅助信任值、目标用户的全局信任值和局部信任值,进一步获得目标用户与其他用户之间的总体信任值;

好友推荐模块:用于对目标用户和其他用户进行好友评分,并将分数从大到小进行排序,取前几名用户作为目标用户的好友进行推荐;

其特征在于,包括以下步骤:

步骤1、确定用户的活动类别和每个活动类别中所对应的词条,即完成在计算机中用户活动关键词库构建模块的构建;

步骤2、采用计算机中的用户活动分类识别模块提取目标用户在第三方应用上所发布的信息,包括文本、图片标签、图片注释、视频标签和视频注释,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取该目标用户的活动类别集合;

步骤3、重复步骤1至步骤2,获取其他目标用户的活动类别集合;

步骤4、采用计算机中的用户活动相似性计算模块根据某类活动词条在目标用户的活动类别集合中出现次数,目标用户的活动类别集合中所有词条出现的次数、用户总数和包含该类活动的用户数,获得该类活动相对于目标用户的重要程度;

步骤5、重复步骤4,分别获得每类活动相对于目标用户的重要程度;

步骤6、重复步骤4至步骤5,获得每类活动相对于其他用户的重要程度;

步骤7、采用计算机中的用户活动相似性计算模块根据每类活动相对于每个用户的重要程度,计算获得两个用户之间活动的相似度;

步骤8、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系,若为关注关系,则执行步骤9,若为评价关系和点赞关系,则执行步骤10;

步骤9、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他

用户之间为直接关注关系或间接关注关系,若为直接关注,则确定两者为直接信任关系,若为间接关注,则确定两者为间接信任关系;

步骤10、确定两者为辅助信任关系;

步骤11、采用计算机中的用户信任值计算模块确定目标用户与其他用户之间的总体信任值,具体如下:

步骤11-1、根据目标用户关注的用户数和被关注的用户数,其他目标用户关注的用户数和被关注用户数,确定目标用户与其他用户的直接信任值;

计算公式如下:

$$dTru(u_i, u_j) = \sqrt{\frac{in(u_j)}{in(u_j) + out(u_i)}} \quad (1)$$

其中, $dTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的直接信任值; $in(u_j)$ 表示用户 u_j 的被关注用户数; $out(u_i)$ 表示用户 u_i 关注的用户数; u_i 为目标用户;

步骤11-2、根据用户两两之间的直接信任值,确定获得多条路径下,目标用户与其他用户的间接信任值;

目标用户与其他用户在第 p 条路径下的间接信任值计算公式如下:

$$iTru(u_i, u_j)_p^n = \prod_{m=i}^{j-1} dTru(u_m, u_{m+1}) \quad (2)$$

其中, $iTru(u_i, u_j)_p^n$ 表示第 p 条路径下,用户 u_i 与用户 u_j 的间接信任值; $dTru(u_m, u_{m+1})$ 表示用户 u_m 与用户 u_{m+1} 的直接信任值, $m = i, \dots, j-1$; n 表示路径长度, $n = 2$ 或 3 , p 表示第 p 条路径;

目标用户与其他用户获得多条路径下的间接信任值计算公式如下:

$$iTru(u_i, u_j) = \frac{1}{N} \sum_{p=1}^N iTru(u_i, u_j)_p^n \quad (3)$$

其中, $iTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的间接信任值; N 表示路径条数;

步骤11-3、根据用户两两之间的评价和点赞次数,及用户两两之间的评价和点赞次数最大值,确定目标用户与其他用户的辅助信任值;

计算公式如下:

$$sTru(u_i, u_j) = \frac{C_{ij}}{\max\{C\}} \quad (4)$$

其中, $sTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的辅助信任值; C_{ij} 表示用户 u_i 对用户 u_j 进行的评价和点赞次数; $\max\{C\}$ 表示用户两两之间的评价和点赞次数最大值;

步骤11-4、根据获得的目标用户与其他用户的直接信任值、间接信任值和辅助信任值,确定目标用户的局部信任值;

计算公式如下:

$$lTru(u_i, u_j) = \alpha * dTru(u_i, u_j) + \beta * iTru(u_i, u_j) + \gamma * sTru(u_i, u_j) \quad (5)$$

其中, $lTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的局部信任值; $\alpha = 0.5 \pm 0.2$, $\beta = 0.3 \pm 0.2$, $\gamma = 1 - \alpha - \beta$;

步骤11-5、根据目标用户 u_i 的全部所用用户中,被关注数的最大值和最小值,及某个所

用用户 u_j 的被关注数,确定目标用户某个所用用户 u_j 的全局信任值;

计算公式如下:

$$gTru(u_j) = \frac{in(u_j) - \min_{u_k \in U} \{in(u_k)\}}{\max_{u_k \in U} \{in(u_k)\} - \min_{u_k \in U} \{in(u_k)\}} \quad (6)$$

其中, $gTru(u_j)$ 表示目标用户 u_i 的某个所用用户 u_j 的全局信任值; $in(u_j)$ 表示用户 u_j 的被关注数, $\min_{u_k \in U} \{in(u_k)\}$ 表示被关注用户数最小值, $\max_{u_k \in U} \{in(u_k)\}$ 表示被关注用户数最大值;

步骤11-6、根据获得的目标用户的局部信任值和全局信任值,确定目标用户的总体信任值;

计算公式如下:

$$tTru(u_i, u_j) = \omega * lTru(u_i, u_j) + (1 - \omega) * gTru(u_j) \quad (7)$$

其中, $tTru(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的总体信任值, ω 表示调整因子, $\omega = 0.7 \pm 0.2$;

步骤12、采用计算机中的好友推荐模块对目标用户和其他用户进行好友评分,并将分数从大到小进行排序,取前几名用户作为目标用户的好友进行推荐;

好友评分计算公式如下:

$$score(u_i, u_j) = \xi * simAct(u_i, u_j) + (1 - \xi) * tTru(u_i, u_j) \quad (8)$$

其中, $score(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的好友评分, ξ 表示调整因子, $\xi = 0.6 \pm 0.2$; $simAct(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度。

2. 根据权利要求1所述的方法,其特征在于,步骤4所述的获得该类活动相对于目标用户的重要程度,计算公式如下:

$$v_{ik} = \frac{\sum_m f_{ik_m}}{\sum_i f_i} \times \lg \frac{|u|}{|u_k|} \quad (9)$$

其中, v_{ik} 表示活动 A_k 对用户 u_i 重要程度, f_{ik_m} 表示活动 A_k 中第 m 个词条在用户 u_i 的活动类别集合中出现次数, f_i 表示用户 u_i 的活动类别集合中所有词条出现的次数, $|u|$ 表示用户的总数, $|u_k|$ 表示包含 k 类活动的用户数。

3. 根据权利要求1所述的方法,其特征在于,步骤7所述的计算获得两个用户之间的相似度,计算公式如下:

$$simAct(u_i, u_j) = \frac{\sum_{k=1}^{|A|} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{|A|} v_{ik}^2} \sqrt{\sum_{k=1}^{|A|} v_{jk}^2}} \quad (10)$$

其中, $simAct(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度, v_{ik} 表示活动 A_k 对用户 u_i 重要程度; v_{jk} 表示活动 A_k 对用户 u_j 重要程度, $|A|$ 表示活动数。

基于活动相似和社交信任的社交网好友推荐系统及方法

技术领域

[0001] 本发明属于信息推荐和数据挖掘领域,具体涉及一种基于活动相似和社交信任的社交网好友推荐系统及方法。

背景技术

[0002] 随着社交网的日益繁荣和发展,整个世界几乎已被社交网所覆盖,其中尤其以基于位置的社交网络服务LBSNs (Location-based Social Networks) 最为引入注目。LBSNs通过记录用户在何时、何地发生了何种行为,来加强社交网与地理位置的关联性,从而为企业提供各种营销决策服务。

[0003] 好友推荐是LBSNs众多服务之一,也是目前学术界和工业界的研究热点之一,但通过查阅相关资料得知,迄今为止,LBSNs中的好友推荐方法大都基于位置信息展开个性化推荐,一方面没有充分考虑在某个位置所发生的语义活动对推荐的影响,另一方面即便涉及一些语义信息但使用的也不够完整。事实上,LBSNs中有一部分应用包含用户活动语义信息,比如Instagram(照片分享)应用中产生的照片语义描述标签和注释,Foursquare(四方)应用中的语义签到信息等,这些语义信息在某种程度上体现出了用户行为活动偏好,如果充分利用此信息可以有效提高推荐质量,但这点却一直被忽略。此外,在进行好友推荐时,除了要考虑用户位间的活动相似性以外,用户间是否具有社交信任关系也是一个非常重要的推荐依据,但这一因素在LBSNs好友推荐中也未引起足够重视。

[0004] 传统LBSNs中,好友推荐主要是基于用户共同好友数量、年龄职业等基本属性相似方式实现,由于很少涉及对用户活动兴趣偏好和用户之间信任关系的考率,导致推荐质量不高。

发明内容

[0005] 针对目前好友推荐方法存在虚假推荐和未考虑用户行为偏好问题,本发明提出一种基于活动相似和社交信任的社交网好友推荐系统及方法,简称FRBTA (Friend Recommendation Based on Trust and Activity)方法,以达到提高实用价值、明确目标客户范围、提高广告服务的关联度和准确性、提高广告营销价值的目的。

[0006] 一种基于活动相似和社交信任的社交网好友推荐系统,该系统设置于计算机中,包括用户活动关键词库构建模块、用户活动分类识别模块、用户活动相似性计算模块、用户信任关系构建模块、用户信任值计算模块和好友推荐模块,其中,

[0007] 用户活动关键词库构建模块:用于根据用户的活动类别和每个活动类别中所对应的词条构建词库;

[0008] 用户活动分类识别模块:用于提取用户在第三方应用上所发布的信息,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取用户的活动类别集合;

[0009] 用户活动相似性计算模块:用于根据某类活动对应词条在目标用户发布信息中出现的次数、该类活动对应词条在其他用户发布信息中出现的次数、用户总数和包含该类活

动的用户数,获得该类活动相对于目标用户的重要程度,再根据每类活动相对于每个用户的重要程度,计算获得两个用户之间的相似度;

[0010] 用户信任关系构建模块:用于从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系;若为关注关系,则进一步判断目标用户与其他用户之间为直接关注关系或间接关注关系,如果是直接关注,则确定两者为直接信任关系,如果是间接关注,则确定两者为间接信任关系;若为评价关系和点赞关系,则确定两者为辅助信任关系;

[0011] 用户信任值计算模块:用于根据目标用户与其他用户之间的交互关系,确定目标用户与其他用户的直接信任值、间接信任值、辅助信任值、目标用户的全局信任值和局部信任值,进一步获得目标用户与其他用户之间的总体信任值;

[0012] 好友推荐模块:用于对目标用户和其他用户进行好友评分,并将分数从大到小进行排序,取前几名用户作为目标用户的好友进行推荐。

[0013] 采用基于活动相似和社交信任的社交网好友推荐系统进行的方法,包括以下步骤:

[0014] 步骤1、确定用户的活动类别和每个活动类别中所对应的词条,即完成在计算机中用户活动关键词库构建模块的构建;

[0015] 步骤2、采用计算机中的用户活动分类识别模块提取目标用户在第三方应用上所发布的信息,包括文本、图片标签、图片注释、视频标签和视频注释,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取该目标用户的活动类别集合;

[0016] 步骤3、重复步骤1至步骤2,获取其他目标用户的活动类别集合;

[0017] 步骤4、采用计算机中的用户活动相似性计算模块根据某类活动词条在目标用户的活动类别集合中出现次数,目标用户的活动类别集合中所有词条出现的次数、用户总数和包含该类活动的用户数,获得该类活动相对于目标用户的重要程度;

[0018] 步骤5、重复步骤4,分别获得每类活动相对于目标用户的重要程度;

[0019] 步骤6、重复步骤4至步骤5,获得每类活动相对于其他用户的重要程度;

[0020] 步骤7、采用计算机中的用户活动相似性计算模块根据每类活动相对于每个用户的重要程度,计算获得两个用户之间活动的相似度;

[0021] 步骤8、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系,若为关注关系,则执行步骤9,若为评价关系和点赞关系,则执行步骤10;

[0022] 步骤9、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他用户之间为直接关注关系或间接关注关系,若为直接关注,则确定两者为直接信任关系,若为间接关注,则确定两者为间接信任关系;

[0023] 步骤10、确定两者为辅助信任关系;

[0024] 步骤11、采用计算机中的用户信任值计算模块确定目标用户与其他用户之间的总体信任值,具体如下:

[0025] 步骤11-1、根据目标用户关注的用户数和被关注的用户数,其他目标用户关注的用户数和被关注用户数,确定目标用户与其他用户的直接信任值;

[0026] 计算公式如下:

$$[0027] \quad dTru(u_i, u_j) = \sqrt{\frac{in(u_j)}{in(u_j) + out(u_i)}} \quad (1)$$

[0028] 其中, $dTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的直接信任值; $in(u_j)$ 表示用户 u_j 的被关注用户数; $out(u_i)$ 表示用户 u_i 关注的用户数; u_i 为目标用户;

[0029] 步骤11-2、根据用户两两之间的直接信任值, 确定获得多条路径下, 目标用户与其他用户的间接信任值;

[0030] 目标用户与其他用户在第 p 条路径下的间接信任值计算公式如下:

$$[0031] \quad iTru(u_i, u_j)_p^n = \prod_{m=i}^{j-1} dTru(u_m, u_{m+1}) \quad (2)$$

[0032] 其中, $iTru(u_i, u_j)_p^n$ 表示第 p 条路径下, 用户 u_i 与用户 u_j 的间接信任值; $dTru(u_m, u_{m+1})$ 表示用户 u_m 与用户 u_{m+1} 的直接信任值, $m=i, \dots, j-1$; n 表示路径长度, $n=2$ 或 3 , p 表示第 p 条路径;

[0033] 目标用户与其他用户在获得多条路径下的间接信任值计算公式如下:

$$[0034] \quad iTru(u_i, u_j) = \frac{1}{N} \sum_{p=1}^N iTru(u_i, u_j)_p^n \quad (3)$$

[0035] 其中, $iTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的间接信任值; N 表示路径条数;

[0036] 步骤11-3、根据用户两两之间的评价和点赞次数, 及用户两两之间的评价和点赞次数最大值, 确定目标用户与其他用户的辅助信任值;

[0037] 计算公式如下:

$$[0038] \quad sTru(u_i, u_j) = \frac{C_{ij}}{\max\{C\}} \quad (4)$$

[0039] 其中, $sTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的辅助信任值; C_{ij} 表示用户 u_i 对用户 u_j 进行的评价和点赞次数; $\max\{C\}$ 表示用户两两之间的评价和点赞次数最大值;

[0040] 步骤11-4、根据获得的目标用户与其他用户的直接信任值、间接信任值和辅助信任值, 确定目标用户的局部信任值;

[0041] 计算公式如下:

$$[0042] \quad lTru(u_i, u_j) = \alpha * dTru(u_i, u_j) + \beta * iTru(u_i, u_j) + \gamma * sTru(u_i, u_j) \quad (5)$$

[0043] 其中, $lTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的局部信任值; $\alpha = 0.5 \pm 0.2$ 、 $\beta = 0.3 \pm 0.2$ 、 $\gamma = 1 - \alpha - \beta$;

[0044] 步骤11-5、根据目标用户 u_i 的全部所用用户中, 被关注数的最大值和最小值, 及某个所用用户 u_j 的被关注数, 确定目标用户某个所用用户 u_j 的全局信任值;

[0045] 计算公式如下:

$$[0046] \quad gTru(u_j) = \frac{in(u_j) - \min_{u_k \in U} \{in(u_k)\}}{\max_{u_k \in U} \{in(u_k)\} - \min_{u_k \in U} \{in(u_k)\}} \quad (6)$$

[0047] 其中, $gTru(u_j)$ 表示目标用户 u_i 的某个所用用户 u_j 的全局信任值; $in(u_j)$ 表示用户 u_j 的被关注数, $\min_{u_k \in U} \{in(u_k)\}$ 表示被关注用户数最小值, $\max_{u_k \in U} \{in(u_k)\}$ 表示被关注用户

数最大值；

[0048] 步骤11-6、根据获得的目标用户的局部信任值和全局信任值，确定目标用户的总体信任值；

[0049] 计算公式如下：

$$[0050] \quad tTru(u_i, u_j) = \omega * lTru(u_i, u_j) + (1 - \omega) * gTru(u_j) \quad (7)$$

[0051] 其中， $tTru(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的总体信任值， ω 表示调整因子， $\omega = 0.7 \pm 0.2$ ；

[0052] 步骤12、采用计算机中的好友推荐模块对目标用户和其他用户进行好友评分，并将分数从大到小进行排序，取前几名用户作为目标用户的好友进行推荐；

[0053] 好友评分计算公式如下：

$$[0054] \quad score(u_i, u_j) = \xi * simAct(u_i, u_j) + (1 - \xi) * tTru(u_i, u_j) \quad (8)$$

[0055] 其中， $score(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的好友评分， ξ 表示调整因子， $\xi = 0.6 \pm 0.2$ ； $simAct(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度。

[0056] 步骤4所述的获得该类活动相对于目标用户的重要程度，计算公式如下：

$$[0057] \quad v_{ik} = \frac{\sum_m f_{ik_m}}{\sum f_i} \times lg \frac{|u|}{|u_k|} \quad (9)$$

[0058] 其中， v_{ik} 表示活动 A_k 对用户 u_i 重要程度， f_{ik_m} 表示活动 A_k 中第 m 个词条在用户 u_i 的活动类别集合中出现次数， f_i 表示用户 u_i 的活动类别集合中所有词条出现的次数， $|u|$ 表示用户的总数， $|u_k|$ 表示包含 k 类活动的用户数。

[0059] 步骤7所述的计算获得两个用户之间的相似度，计算公式如下：

$$[0060] \quad simAct(u_i, u_j) = \frac{\sum_{k=1}^{|A|} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{|A|} v_{ik}^2} \sqrt{\sum_{k=1}^{|A|} v_{jk}^2}} \quad (10)$$

[0061] 其中， $simAct(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度， v_{ik} 表示活动 A_k 对用户 u_i 重要程度； v_{jk} 表示活动 A_k 对用户 u_j 重要程度， $|A|$ 表示活动数。

[0062] 本发明优点：

[0063] 本发明一种基于活动相似和社交信任的社交网好友推荐系统及方法，该方法主要是利用用户社交信任值和活动偏好相似性来实现基于位置社交网络中好友推荐，由于活动可以体现用户兴趣偏好，因此通过用户间活动相似性发现与其偏好相似的好友；由于社交信任能反映用户间交互紧密程度，因此根据不同程度信任关系进行好友推荐具有更合理的可解释性；实验证明，本发明推荐效果在准确性和合理解释性上均优于现存的好友推荐方法，实际应用价值很高，如果能得到推广，对企事业单位明确目标客户范围，提高广告服务的关联度和准确性，提高广告营销价值均有重要的指导和决策意义。

附图说明

[0064] 图1为本发明一种实施例的基于活动相似和社交信任的社交网好友推荐系统结构示意图；

- [0065] 图2为本发明一种实施例的基于活动相似和社交信任的社交网好友推荐方法流程图；
- [0066] 图3为本发明一种实施例的一个社交网用户间交互关系图；
- [0067] 图4为本发明一种实施例的对应图3实例的社交信任网示意图；
- [0068] 图5为本发明一种实施例的对应图3实例的社交网局部信任值示意图；
- [0069] 图6为本发明一种实施例的对应图3实例的社交网总信任值示意图；
- [0070] 图7为本发明一种实施例的测试参数 α 和 β 对实验结果影响示意图,其中图(a)为准确率变化曲面图;图(b)为召回率变化曲面图；
- [0071] 图8为本发明一种实施例的测试参数 ω 对实验结果影响示意图,其中,图(a)为准确率变化趋势;图(b)为准确率变化趋势图；
- [0072] 图9为本发明一种实施例的测试参数 ξ 对实验结果影响示意图,其中,图(a)为准确率变化趋势图;图(b)为准确率变化趋势图；
- [0073] 图10为本发明一种实施例的测试推荐个数对评测指标的影响示意图,其中,图(a)为准确率随推荐个数变化趋势图;图(b)为召回率随推荐个数变化趋势图。

具体实施方式

[0074] 下面结合附图对本发明一种实施例做进一步说明。

[0075] 本发明实施例中,如图1所示,基于活动相似和社交信任的社交网好友推荐系统,该系统设置于计算机中,包括用户活动关键词库构建模块、用户活动分类识别模块、用户活动相似性计算模块、用户信任关系构建模块、用户信任值计算模块和好友推荐模块,其中,

[0076] 用户活动关键词库构建模块:用于根据用户的活动类别和每个活动类别中所对应的词条构建词库;

[0077] 本发明实施例中,借助第三方应用Foursquare中的活动分类并根据实际需要构建六类活动类别,即Travel (旅行)&Transport (运输)、Shop (购物)&Service (服务)、Outdoors (户外运动)&Recreation (休闲)、Arts (艺术)&Entertainment (娱乐)、Nightlife Spot (夜生活、Food (饮食)等;事实上,可根据实际情况按需构建不同粒度的活动类别。每类活动由词条项(Terms)构成,包含这些词条项的文本主要源于与位置相关的信息,如位置名称、位置类别、位置建议(Tips)、位置评论(Comments)以及位置描述文本等,通过调用Foursquare中的公开API函数,并利用开源程序Lucene中的Standard Analyzer分词器对这些文本进行分词和去停用词等操作来提取用户活动词条项。

[0078] 用户活动分类识别模块:用于提取用户在第三方应用上所发布的信息,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取用户的活动类别集合;

[0079] 本发明实施例中,提取Instagram用户在不同位置上传照片的标签(Tag)和标题(Caption),利用Lucene中的Standard Analyzer分词器提取活动词条,并按(用户名,用户ID,位置ID,活动关键词列表集合)关系形式存储;其次,将Instagram中每一个用户在不同位置的活动关键词列表集合中的词条和活动库每一个活动类别下对应的关键词列表集合中的词条进行关键词匹配,匹配上的关键词个数与在该位置所提用户活动关键词个数之比即为匹配率。最后,由于本发明设定位置与用户活动是一对一映射关系,即一个位置对应一个活动类别,因此从六类活动中选择匹配率最高的一类活动作为用户在该位置的活动类

别,存储形式为(用户ID,位置ID,活动类别ID,活动类别名称)。

[0080] 用户活动相似性计算模块:用于根据某类活动对应词条在目标用户发布信息中出现的次数、该类活动对应词条在其他用户发布信息中出现的次数、用户总数和包含该类活动的用户数,获得该类活动相对于目标用户的重要程度,再根据每类活动相对于每个用户的重要程度,计算获得两个用户之间的相似度;

[0081] 用户信任关系构建模块:用于从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系;若为关注关系,则进一步判断目标用户与其他用户之间为直接关注关系或间接关注关系,如果是直接关注,则确定两者为直接信任关系,如果是间接关注,则确定两者为间接信任关系;若为评价关系和点赞关系,则确定两者为辅助信任关系;

[0082] 本发明实施例中提出了用户之间信任关系概念,主要包括直接信任关系dTru、间接信任关系iTru和辅助信任关系sTru三种。若用户 u_i follow(关注)用户 u_j ,则二者之间具有直接信任关系;如果用户 u_i follow用户 u_k ,而用户 u_k follow用户 u_j ,则 u_i 和 u_j 之间具有间接信任关系;如果 u_i 标识 u_j 为喜欢(点赞)或者 u_i 对 u_j 进行评论,则认为 u_i 对 u_j 有辅助信任关系;一对用户间的信任关系为局部信任关系;其他用户对当前用户的信任关系为全局信任。满足预定义最大长度的可达信任路径为最大信任长度路径,该路径长度用 n 表示。

[0083] 用户信任值计算模块:用于根据目标用户与其他用户之间的交互关系,确定目标用户与其他用户的直接信任值、间接信任值、辅助信任值、目标用户的全局信任值和局部信任值,进一步获得目标用户与其他用户之间的总体信任值;

[0084] 好友推荐模块:用于对目标用户和其他用户进行好友评分,并将分数从大到小进行排序,取前几名用户作为目标用户的好友进行推荐。

[0085] 采用基于活动相似和社交信任的社交网好友推荐系统进行的方法,方法流程图如图2所示,包括以下步骤:

[0086] 步骤1、确定用户的活动类别和每个活动类别中所对应的词条,即完成在计算机中用户活动关键词库构建模块的构建;

[0087] 本发明实施例中,构建的活动关键词库如表1所示(仅列出部分词条):

[0088] 表1

活动类别 ID	活动类别名称	部分词条
1	Travel & Transport (旅行和运输)	Airports (机场), Bike (自行车), Bus (公交车), Skate Parks (溜冰场), Parks (公园), ...
2	Shop & Service (购物和服务)	Baby Stores (婴儿商店), Bank (银行), Music Discs (音乐专辑), Parks (公园), ...
3	Outdoors & Recreation (户外运动和休闲)	Athletics (有氧运动), Badminton Courts (羽毛球场馆), Bike (自行车), ...
4	Arts & Entertainment (艺术和娱乐)	Art Galleries (艺术长廊), Dance (舞蹈), Music Discs (音乐专辑), ...
5	Nightlife Spot (夜生活)	Beer Gardens (啤酒屋), Nightclubs (夜总会), Parks (公园), ...
6	Food (饮食)	American Restaurants (美国餐馆), BBQ Joints (烤肉馆), ...

[0090] 步骤2、采用计算机中的用户活动分类识别模块提取目标用户在第三方应用上所发布的信息,包括文本、图片标签、图片注释、视频标签和视频注释,并根据提取的信息在用户活动关键词库构建模块中进行匹配,获取该目标用户的活动类别集合;

[0091] 步骤3、重复步骤1至步骤2,获取其他目标用户的活动类别集合;

[0092] 本发明实施例中,假定有15个用户 u_1 到 u_{15} (即 $|u|=15$),并识别出有6个用户(即 u_1 到 u_6)进行了5类活动(即 $|A|=5$, $A_1=Travel\&\ Transport$, $A_2=Shop\&\ Service$, $A_3=Outdoors\&\ Recreation$, $A_4=Arts\&\ Entertainment$, $A_5=Nightlife\ Spot$),且包含活动的用户数 $|u_j|=3$,而用户 u_1 - u_6 间有4个用户(即 u_1 到 u_4)存在社交网交互关系。

[0093] 假定 u_1 - u_6 每个用户的词条出现总数均为14(即 $\sum f_i=14$),则用户的活动类别集合分别如下,其中括号内数字为词条出现次数:

[0094] 从用户 u_1 文本中提取的词条集合 $T_{u1}=\{baby\ stores\ (4),\ music\ discs\ (3),\ parks\ (3),\ dance\ (2),\ nightclubs\ (2)\}$;从 T_{u1} 中识别出用户 u_1 的活动类别 $A_{u1}=\{A_2, A_4, A_5\}$;

[0095] $T_{u2}=\{bus\ (4),\ bike\ (3),\ parks\ (3),\ athletics\ (2),\ nightclubs\ (2)\}$, $A_{u2}=\{A_1, A_3, A_5\}$;

[0096] $T_{u3}=\{air\ ports\ (2),\ bike\ (3),\ badminton\ courts\ (4),\ parks\ (3),\ nightclubs\ (2)\}$, $A_{u3}=\{A_1, A_3, A_5\}$;

[0097] $T_{u4}=\{bike\ (10),\ money\ (2),\ price\ (2)\}$, $A_{u4}=\{A_1, A_3\}$;

[0098] $T_{u6}=\{music\ disc\ (10),\ quantity\ (2),\ year\ (2)\}$, $A_{u6}=\{A_2, A_4\}$;

[0099] $T_{u5}=\{music\ discs\ (10),\ money\ (2),\ price\ (2)\}$, $A_{u5}=\{A_2, A_4\}$;

[0100] 步骤4、采用计算机中的用户活动相似性计算模块根据某类活动词条在目标用户的活动类别集合中出现次数,目标用户的活动类别集合中所有词条出现的次数、用户总数和包含该类活动的用户数,获得该类活动相对于目标用户的重要程度;

[0101] 本发明实施例中,构造用户和各种活动类别间的用户-活动矩阵UAM (User

Activity Matrix) 如下:

$$[0102] \quad \begin{matrix} & A_1 & A_2 & \dots & A_j & \dots & A_K \\ \begin{matrix} u_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_n \end{matrix} & \begin{pmatrix} v_{11} & v_{12} & \dots & v_{1j} & \dots & v_{1K} \\ v_{21} & \dots & \dots & \dots & \dots & v_{2K} \\ \vdots & \dots & \dots & \dots & \dots & \dots \\ v_{i1} & \dots & \dots & v_{ij} & \dots & v_{iK} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{n1} & v_{n2} & \dots & v_{nj} & \dots & v_{nK} \end{pmatrix} \end{matrix}$$

[0103] 一个用户访问某个其余用户也经常访问的位置并不代表该用户对反映该位置的活动类别具有更大的偏好,有可能只是出于日常生活必需,比如用户访问餐馆的总次数肯定比访问博物馆的总次数要多,但是并不一定说明用户对饮食的偏好比对文化建筑的偏好要强烈;反之,如果用户访问某个其余用户很少访问的位置,则该位置所反映的活动类别能更好的代表该用户兴趣偏好,因此,此处使用活动对用户的重要性作为矩阵值。

[0104] 本发明实施例中,计算活动 A_k 对用户 u_i 的重要程度,可以借鉴TF-IDF的计算思想,将每个活动对应的关键词看成文档中的关键词,将某个活动对应的所有关键词出现的总次数看成该活动出现的次数,而每个用户可以看成是一个文档。

[0105] 所述的获得该类活动相对于目标用户的重要程度,计算公式如下:

$$[0106] \quad v_{ik} = \frac{\sum f_{ik_m}}{\sum f_i} \times \lg \frac{|u|}{|u_k|} \quad (9)$$

[0107] 其中, v_{ik} 表示活动 A_k 对用户 u_i 重要程度, f_{ik_m} 表示活动 A_k 中第 m 个词条在用户 u_i 的活动类别集合中出现次数, f_i 表示用户 u_i 的活动类别集合中所有词条出现的次数, $|u|$ 表示用户的总数, $|u_k|$ 表示包含 k 类活动的用户数。

[0108] 步骤5、重复步骤4,分别获得每类活动相对于目标用户的重要程度;

[0109] 步骤6、重复步骤4至步骤5,获得每类活动相对于其他用户的重要程度;

[0110] 本发明实施例中,每类活动相对于所有用户的重要程度计算结果如下:

$$[0111] \quad v_{12} = \frac{4+3+3}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0112] \quad v_{14} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0113] \quad v_{15} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0114] \quad v_{21} = \frac{4+3+3}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0115] \quad v_{23} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0116] \quad v_{25} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0117] \quad v_{31} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0118] \quad v_{33} = \frac{4+3+3}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0119] \quad v_{35} = \frac{3+2}{14} \cdot \lg \frac{15}{3} = \frac{5}{14} \cdot \lg 5 = \frac{1}{4}$$

$$[0120] \quad v_{41} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0121] \quad v_{43} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0122] \quad v_{52} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0123] \quad v_{54} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0124] \quad v_{62} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

$$[0125] \quad v_{64} = \frac{10}{14} \cdot \lg \frac{15}{3} = \frac{10}{14} \cdot \lg 5 = \frac{1}{2}$$

[0126] 步骤7、采用计算机中的用户活动相似性计算模块根据每类活动相对于每个用户的重要程度,计算获得两个用户之间活动的相似度;

[0127] 本发明实施例中,得到活动对用户的重要性之后,基于UAM矩阵,使用夹角余弦相似性公式计算不同用户之间的活动偏好相似性,如公式(10)所示,相似度值越大,表明两个用户之间活动偏好相似性越高,即两个用户之间活动偏好越相似。

[0128] 计算获得两个用户之间的相似度,计算公式如下:

$$[0129] \quad \text{simAct}(u_i, u_j) = \frac{\sum_{k=1}^{|A|} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{|A|} v_{ik}^2} \sqrt{\sum_{k=1}^{|A|} v_{jk}^2}} \quad (10)$$

[0130] 其中, $\text{simAct}(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度, v_{ik} 表示活动 A_k 对用户 u_i 重要程度; v_{jk} 表示活动 A_k 对用户 u_j 重要程度, $|A|$ 表示活动数。

[0131] 本发明实施例中,用户 u_1 到 u_6 对5类活动的重要程度值结果为:

$$[0132] \quad \begin{array}{c} A_1 \quad A_2 \quad A_3 \quad A_4 \quad A_5 \\ \begin{array}{c} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \end{array} \end{array} \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{4} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

[0133] 本发明实施例中,由于用户 u_7 到 u_{15} 对5类活动的 v_{ik} 均为0,因此为清楚简单起见,此处的UAM矩阵省略了用户 u_7 - u_{15} 的 v_{ik} 值。

[0134] 根据公式(10)计算各用户间的活动相似性如下:

$$[0135] \quad \text{simAct}(u_2, u_1) = \text{simAct}(u_1, u_2) = \frac{\frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + 0 \cdot \frac{1}{4} + \frac{1}{4} \cdot \frac{1}{4}}{\sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{4} + 0^2 + \frac{1^2}{4}} \cdot \sqrt{0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{4} + \frac{1^2}{4}}} = 0.17$$

$$[0136] \quad \text{simAct}(u_2, u_3) = \text{simAct}(u_3, u_2) = \frac{\frac{1}{2} \cdot \frac{1}{4} + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{4}}{\sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{4} + 0^2 + \frac{1^2}{4}} \cdot \sqrt{\frac{1^2}{4} + 0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{4}}} = 0.83$$

$$[0137] \quad \text{simAct}(u_2, u_4) = \text{simAct}(u_4, u_2) = \frac{\frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot 0}{\sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{4} + 0^2 + \frac{1^2}{4}} \cdot \sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{2} + 0^2 + 0^2}} = 0.61$$

$$[0138] \quad \text{simAct}(u_2, u_5) = \text{simAct}(u_5, u_2) = \frac{\frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{4} + 0^2 + \frac{1^2}{4}} \cdot \sqrt{0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{2} + 0^2}} = 0$$

$$[0139] \quad \text{simAct}(u_2, u_6) = \text{simAct}(u_6, u_2) = \frac{\frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{4} + 0^2 + \frac{1^2}{4}} \cdot \sqrt{0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{2} + 0^2}} = 0$$

$$[0140] \quad \text{simAct}(u_1, u_3) = \text{simAct}(u_3, u_1) = \frac{0 \cdot \frac{1}{4} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot \frac{1}{4}}{\sqrt{0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{4} + \frac{1^2}{4}} \cdot \sqrt{\frac{1^2}{4} + 0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{4}}} = 0.17$$

$$[0141] \quad \text{simAct}(u_1, u_4) = \text{simAct}(u_4, u_1) = \frac{0 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 0}{\sqrt{0^2 + \frac{1^2}{2} + 0^2 + \frac{1^2}{4} + \frac{1^2}{4}} \cdot \sqrt{\frac{1^2}{2} + 0^2 + \frac{1^2}{2} + 0^2 + 0^2}} = 0$$

$$[0142] \quad \text{simAct}(u_1, u_5) = \text{simAct}(u_5, u_1) = \frac{0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + \frac{1}{4}} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0.87$$

$$[0143] \quad \text{simAct}(u_1, u_6) = \text{simAct}(u_6, u_1) = \frac{0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + \frac{1}{4}} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0.87$$

$$[0144] \quad \text{simAct}(u_3, u_4) = \text{simAct}(u_4, u_3) = \frac{\frac{1}{4} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{4} \cdot 0}{\sqrt{\frac{1}{4} + 0^2 + \frac{1}{2} + 0^2 + \frac{1}{4}} \cdot \sqrt{\frac{1}{2} + 0^2 + \frac{1}{2} + 0^2 + 0^2}} = 0.87$$

$$[0145] \quad \text{simAct}(u_3, u_5) = \text{simAct}(u_5, u_3) = \frac{\frac{1}{4} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{\frac{1}{4} + 0^2 + \frac{1}{2} + 0^2 + \frac{1}{4}} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0$$

$$[0146] \quad \text{simAct}(u_3, u_6) = \text{simAct}(u_6, u_3) = \frac{\frac{1}{4} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{4} \cdot 0}{\sqrt{\frac{1}{4} + 0^2 + \frac{1}{2} + 0^2 + \frac{1}{4}} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0$$

$$[0147] \quad \text{simAct}(u_4, u_5) = \text{simAct}(u_5, u_4) = \frac{\frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + 0 \cdot 0}{\sqrt{\frac{1}{2} + 0^2 + \frac{1}{2} + 0^2 + 0^2} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0$$

$$[0148] \quad \text{simAct}(u_4, u_6) = \text{simAct}(u_6, u_4) = \frac{\frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 + 0 \cdot \frac{1}{2} + 0 \cdot 0}{\sqrt{\frac{1}{2} + 0^2 + \frac{1}{2} + 0^2 + 0^2} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 0$$

$$[0149] \quad \text{simAct}(u_5, u_6) = \text{simAct}(u_6, u_5) = \frac{0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} + 0 \cdot 0}{\sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2} \cdot \sqrt{0^2 + \frac{1}{2} + 0^2 + \frac{1}{2} + 0^2}} = 1$$

[0150] 步骤8、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他用户之间的交互关系,包括关注关系、评价关系和点赞关系,若为关注关系,则执行步骤9,若为评价关系和点赞关系,则执行步骤10;

[0151] 步骤9、采用计算机中的用户信任关系构建模块从第三方应用中判断目标用户与其他用户之间为直接关注关系或间接关注关系,若为直接关注,则确定两者为直接信任关系,若为间接关注,则确定两者为间接信任关系;

[0152] 本发明实施例中,构建用户之间follow关系矩阵UFM(User Follow Matrix),UFM为非对称矩阵,即用户之间的follow关系是非对称关系。矩阵值非0即1,若用户之间存在follow关系,则矩阵值为1,若不存在,则矩阵值为0。

[0153] 步骤10、确定两者为辅助信任关系;

[0154] 步骤11、采用计算机中的用户信任值计算模块确定目标用户与其他用户之间的总体信任值;

[0155] 本发明实施例中,假定图3给出了 u_1 、 u_2 、 u_3 和 u_4 间的一个社交网用户间交互关系,包括关注、评价及点赞关系;如图3所示,用户间关注关系用实线表示,评价及点赞关系则用虚线表示。

[0156] 图3对应的社交信任网如图4所示,其局部信任包括 $dTru$ 、 $iTru$ 和 $sTru$,分别用 \xrightarrow{D} 、 \xrightarrow{I} 和 \xrightarrow{S} 表示。图3中有两个 $iTru$,即 $iTru(u_2, u_1)$ 和 $iTru(u_2, u_4)$,因此为清楚起见, \xrightarrow{I} 用粗线表示。

[0157] 假定用户间交互次数,即Comment和Like的总次数为: $u_1 \xrightarrow{S} u_4 = 5$, $u_2 \xrightarrow{S} u_3 = 10$, $u_2 \xrightarrow{S} u_4 = 20$ 。求得被关注用户数最大值 $\max_{u_k \in U} \{in(u_k)\} = 2$,被关注用户数最小值 $\min_{u_k \in U} \{in(u_k)\} = 0$;

[0158] 各种具体信任值计算结果如表2所示,局部信任值和总信任值分别如图5和图6所示:

[0159] 表2

[0160]

$dTru$	$iTru$	$sTru$	$lTru$	$gTru$	$tTru$
$dTru(u_2, u_3) = 1/3$	$iTru(u_2, u_1) = 1/9$	$sTru(u_1, u_4) = 1/4$	$lTru(u_1, u_4) = \frac{1}{4}\gamma$	$gTru(u_1) = 1/2$	$tTru(u_1, u_4) = \frac{1}{4}\omega\gamma + 1 - \omega$
$dTru(u_2, u_4) = 1/2$	$iTru(u_2, u_4) = 1/6$	$sTru(u_2, u_3) = 1/2$	$lTru(u_2, u_1) = \frac{1}{9}\beta$	$gTru(u_2) = 0$	$tTru(u_2, u_4) = \frac{1}{9}\omega\beta + \frac{1}{2}(1 - \omega)$
$dTru(u_3, u_1) = 1/3$		$sTru(u_2, u_4) = 1$	$lTru(u_2, u_3) = \frac{1}{3}\alpha + \frac{1}{2}\gamma$	$gTru(u_3) = 1/2$	$tTru(u_2, u_3) = \frac{1}{3}\omega\alpha + \frac{1}{2}(1 - \omega)\gamma$
$dTru(u_3, u_4) =$			$lTru(u_2, u_4) =$	$gTru(u_4) =$	$tTru(u_2, u_4) =$

[0161]

$1/2$			$\frac{1}{2}\alpha + \frac{1}{6}\beta + \gamma$	1	$(\frac{1}{2}\alpha + \frac{1}{6}\beta + \gamma)\omega + 1 - \omega$
			$lTru(u_3, u_1) = \frac{1}{3}\alpha$		$tTru(u_3, u_1) = \frac{1}{3}\omega\alpha + \frac{1}{2}(1 - \omega)$
			$lTru(u_3, u_4) = \frac{1}{2}\alpha$		$tTru(u_3, u_4) = \frac{1}{2}\omega\alpha + 1 - \omega$

[0162] 根据实验测试, $\alpha = 0.5 \pm 0.2$ 、 $\beta = 0.3 \pm 0.2$ 时推荐准确率最佳,如图7(a)所示,其中A表示最优点; $\alpha = 0.5 \pm 0.2$ 、 $\beta = 0.3 \pm 0.2$ 时召回率最优,如图7(b)所示,为提高好友推荐准确度,本发明设置 $\alpha = 0.5$ 、 $\beta = 0.3$ 。

[0163] 本发明实施例中,局部信任值和全局信任值对用户之间信任值的影响通过参数 ω 来反映,由图8中图(a)和图(b)可以看出,局部信任值对用户之间信任值的影响较大,且在 $\omega = 0.7 \pm 0.2$ 时,准确率和召回率达到最优效果,本发明最终设置 $\omega = 0.7$ 。

[0164] 好友推荐最终通过调整用户活动偏好相似性和用户信任关系两者之间的参数权重实现,准确率和召回率随参数 ξ 变化的趋势如图9中图(a)和图(b)所示。在 $\xi = 0.6 \pm 0.2$ 时,准确率和召回率达到最佳效果,本发明最终设置 $\xi = 0.6$ 。

[0165] 具体计算结果如下:

[0166] 步骤11-1、根据目标用户关注的用户数和被关注的用户数,其他目标用户关注的

用户数和被关注用户数,确定目标用户与其他用户的直接信任值;

[0167] 计算公式如下:

$$[0168] \quad dTru(u_i, u_j) = \sqrt{\frac{in(u_j)}{in(u_j) + out(u_i)}} \quad (1)$$

[0169] 其中, $dTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的直接信任值; $in(u_j)$ 表示用户 u_j 的被关注用户数; $out(u_i)$ 表示用户 u_i 关注的用户数; u_i 为目标用户;

[0170] 步骤11-2、根据用户两两之间的直接信任值,确定获得多条路径下,目标用户与其他用户的间接信任值;

[0171] 目标用户与其他用户在第 p 条路径下的间接信任值计算公式如下:

$$[0172] \quad iTru(u_i, u_j)_p^n = \prod_{m=i}^{j-1} dTru(u_m, u_{m+1}) \quad (2)$$

[0173] 其中, $iTru(u_i, u_j)_p^n$ 表示第 p 条路径下,用户 u_i 与用户 u_j 的间接信任值; $dTru(u_m, u_{m+1})$ 表示用户 u_m 与用户 u_{m+1} 的直接信任值, $m = i, \dots, j-1$; n 表示路径长度, $n = 2$ 或 3 , p 表示第 p 条路径;

[0174] 本发明实施例中,若用户 u_i follow 用户 u_q , 而用户 u_q follow 用户 u_j , 则 $n = 2$, 即 u_i 对 u_j 的间接信任值 $iTru(u_i, u_j)^2 = dTru(u_i, u_q) \cdot dTru(u_q, u_j)$ 。一方面,随着用户之间信任路径长度的不断增加,用户之间的间接信任值不断衰减,为降低计算复杂度,本发明规定用户之间最长信任路径长度 $n = 3$ 。另一方面,由于两个用户之间的间接信任路径具有不唯一性,因此, u_i 到 u_j 的总间接信任值计算方法如公式 (3) 所示:

$$[0175] \quad iTru(u_i, u_j) = \frac{1}{N} \sum_{p=1}^N iTru(u_i, u_j)_p^n \quad (3)$$

[0176] 其中, $iTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的间接信任值; N 表示两者间最长信任路径不超过 3 的路径条数,即用户间接信任值为不同间接信任路径信任值的平均值。

[0177] 步骤11-3、根据用户两两之间的评价和点赞次数,及用户两两之间的评价和点赞次数最大值,确定目标用户与其他用户的辅助信任值;

[0178] 计算公式如下:

$$[0179] \quad sTru(u_i, u_j) = \frac{C_{ij}}{\max\{C\}} \quad (4)$$

[0180] 其中, $sTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的辅助信任值, $sTru(u_i, u_j)$ 最终的取值范围介于 $[0, 1]$ 之间; C_{ij} 表示用户 u_i 对用户 u_i 进行的评价和点赞次数; $\max\{C\}$ 表示用户两两之间的评价和点赞次数最大值;

[0181] 步骤11-4、根据获得的目标用户与其他用户的直接信任值、间接信任值和辅助信任值,确定目标用户的局部信任值;

[0182] 局部信任为直接信任、间接信任和辅助信任三类信任的加权和,计算公式如下:

$$[0183] \quad lTru(u_i, u_j) = \alpha * dTru(u_i, u_j) + \beta * iTru(u_i, u_j) + \gamma * sTru(u_i, u_j) \quad (5)$$

[0184] 其中, $lTru(u_i, u_j)$ 表示用户 u_i 与用户 u_j 的局部信任值; $\alpha = 0.5 \pm 0.2$ 、 $\beta = 0.3 \pm 0.2$ 、 $\gamma = 1 - \alpha - \beta$;

[0185] 局部信任关系只反映了社交网络中一对用户间信任关系,纵观整个社交信任网,要衡量其中某个用户对于整个信任网的影响,还应该考虑用户全局信任值,全局信任关系反映了其余所有用户对某个用户信任程度。

[0186] 步骤11-5、根据目标用户 u_i 的全部所用用户中,被关注数的最大值和最小值,及某个所用用户 u_j 的被关注数,确定目标用户某个所用用户 u_j 的全局信任值;

[0187] 计算公式如下:

$$[0188] \quad gTru(u_j) = \frac{in(u_j) - \min_{u_k \in U} \{in(u_k)\}}{\max_{u_k \in U} \{in(u_k)\} - \min_{u_k \in U} \{in(u_k)\}} \quad (6)$$

[0189] 其中, $gTru(u_j)$ 表示目标用户 u_i 的某个所用用户 u_j 的全局信任值; $in(u_j)$ 表示用户 u_j 的被关注数, $\min_{u_k \in U} \{in(u_k)\}$ 表示被关注用户数最小值, $\max_{u_k \in U} \{in(u_k)\}$ 表示被关注用户数最大值;

[0190] 步骤11-6、根据获得的目标用户的局部信任值和全局信任值,确定目标用户的总体信任值;

[0191] 结合了用户之间局部信任关系和全局信任关系可以得到用户之间总体信任值,计算公式如下:

$$[0192] \quad tTru(u_i, u_j) = \omega * lTru(u_i, u_j) + (1 - \omega) * gTru(u_j) \quad (7)$$

[0193] 其中, $tTru(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的总体信任值, ω 表示调整因子, $\omega = 0.7 \pm 0.2$;

$$[0194] \quad tTru(u_j, u_i) = \omega * lTru(u_j, u_i) + (1 - \omega) * gTru(u_i) \quad (11)$$

[0195] u_i 对 u_j 的总信任值计算如公式(7)表示, u_j 对 u_i 的总信任值计算如公式(11)所示。由公式(7)和(11)可以看出,用户间的总信任关系并不是对称关系,即 $tTru(u_i, u_j) \neq tTru(u_j, u_i)$ 。

[0196] 本发明实施例中,为简单清晰起见,本发明假定推荐1位好友给目标用户,即 $k=1$ 。

$$[0197] \quad tTru(u_2, u_1) = \frac{1}{9} \omega \beta + \frac{1}{2} (1 - \omega) = \frac{1}{9} \times 0.7 \times 0.3 + \frac{1}{2} (1 - 0.7) = 0.023 + 0.15 = 0.173$$

$$[0198] \quad tTru(u_2, u_3) = \frac{1}{3} \alpha \alpha + \frac{1}{2} (1 - \omega) \gamma = \frac{1}{3} \times 0.7 \times 0.5 + \frac{1}{2} (1 - 0.7) (1 - 0.5 - 0.3) = 0.147$$

$$[0199] \quad tTru(u_2, u_4) = (\frac{1}{2} \alpha + \frac{1}{6} \beta + \gamma) \omega + 1 - \omega = (\frac{1}{2} \times 0.5 + \frac{1}{6} \times 0.3 + 0.2) \times 0.7 + 1 - 0.7 = 0.35 + 0.3 = 0.65$$

[0200] 步骤12、采用计算机中的好友推荐模块对目标用户和其他用户进行好友评分,并将分数从大到小进行排序,取前几名用户作为目标用户的好友进行推荐;

[0201] 好友评分计算公式如下:

$$[0202] \quad score(u_i, u_j) = \xi \times simAct(u_i, u_j) + (1 - \xi) \times tTru(u_i, u_j) \quad (8)$$

[0203] 其中, $score(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的好友评分, ξ 表示调整因子, $\xi = 0.6 \pm 0.2$; $simAct(u_i, u_j)$ 表示用户 u_i 对于用户 u_j 的相似度。

[0204] 本发明实施例中,将各参数最佳取值 $\alpha = 0.5$ 、 $\beta = 0.3$ 、 $\omega = 0.7$ 和 $\xi = 0.6$ 代入公式(8),计算目标用户 u_2 的候选好友分数,如下所示。

$$[0205] \quad score(u_2, u_1) = \xi \cdot simAct(u_2, u_1) + (1 - \xi) \cdot tTru(u_2, u_1) = 0.6 \times 0.17 + (1 - 0.6) \times$$

$$0.173=0.102+0.0692=0.1712$$

$$[0206] \quad \text{score}(u_2, u_3) = \xi \cdot \text{simAct}(u_2, u_3) + (1-\xi) \cdot \text{tTru}(u_2, u_3) = 0.6 \times 0.83 + (1-0.6) \times 0.147 = 0.5568$$

$$[0207] \quad \text{score}(u_2, u_4) = \xi \cdot \text{simAct}(u_2, u_4) + (1-\xi) \cdot \text{tTru}(u_2, u_4) = 0.6 \times 0.61 + (1-0.6) \times 0.65 = 0.626$$

$$[0208] \quad \text{score}(u_2, u_5) = \xi \cdot \text{simAct}(u_2, u_5) + (1-\xi) \cdot \text{tTru}(u_2, u_5) = 0.6 \times 0 + (1-0.6) \times 0 = 0$$

$$[0209] \quad \text{score}(u_2, u_6) = \xi \cdot \text{simAct}(u_2, u_6) + (1-\xi) \cdot \text{tTru}(u_2, u_6) = 0.6 \times 0 + (1-0.6) \times 0 = 0$$

[0210] 根据公式(6)计算候选推荐用户 $u_j (u_j \in C_i)$ 与目标用户 u_i 之间的友好评分,其中,参数 ξ 是调整因子,旨在调节活动偏好和社交信任间的权重。将评分结果按照从大到小顺序降序排列,选取前K个用户作为目标用户 u_i 的Top-K(排序靠前的前K个结果即为top-K)好友推荐结果。

[0211] 从计算结果可以看出,应选 u_4 推荐给 u_2 ;如果只考虑传统推荐算法中的行为偏好相似,应选 u_3 作为 u_2 好友,因为 u_3 和 u_2 的行为相似度值为0.83,大于 u_4 和 u_2 的行为相似度值0.61;但由于本发明考虑了社交信任关系,在社交网中, u_2 相比于 u_3 更倾向于信任 u_4 ,因此导致在综合评分后,使得 $\text{score}(u_2, u_3) = 0.5568 < \text{score}(u_2, u_4) = 0.626$,于是最终选取 u_4 作为 u_2 好友加以推荐。

[0212] 本发明中,通过实验测试提出的社交网好友推荐方法的应用效果。由图10中图(a)和图(b)可以看出,随着推荐列表长度的增加,准确率呈下降趋势,原因是相对于分母(即推荐)的增长速度,分子(即命中)的增长速度要慢;召回率呈上升趋势,原因是分母(即测试)不变,分子(即命中)增大。同时可以看出,相对于CFR(Common Friend Recommendation)和PSR(Preference Similarity Recommendation)方法,本发明提出的FRBTA算法在准确率和召回率上均取得了更好的推荐效果。

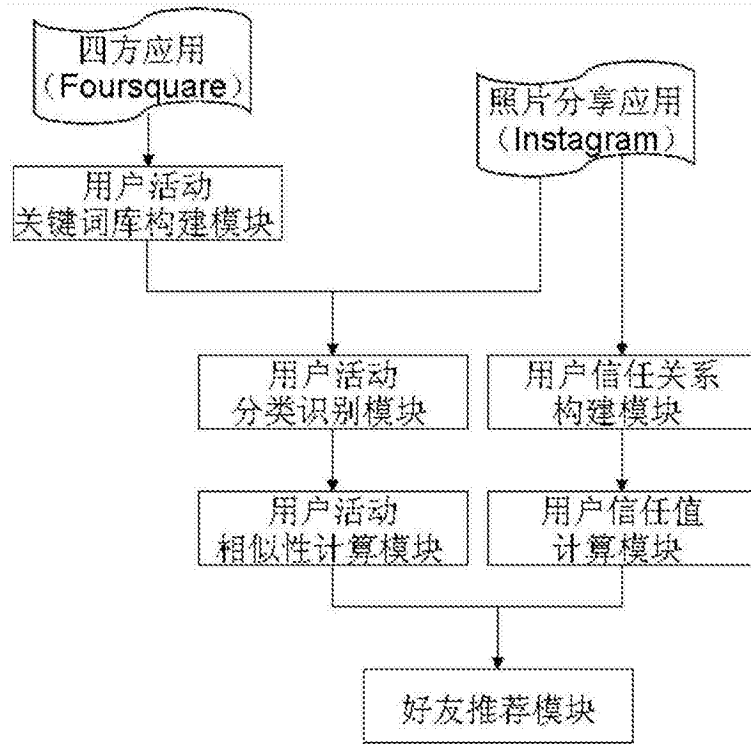


图1

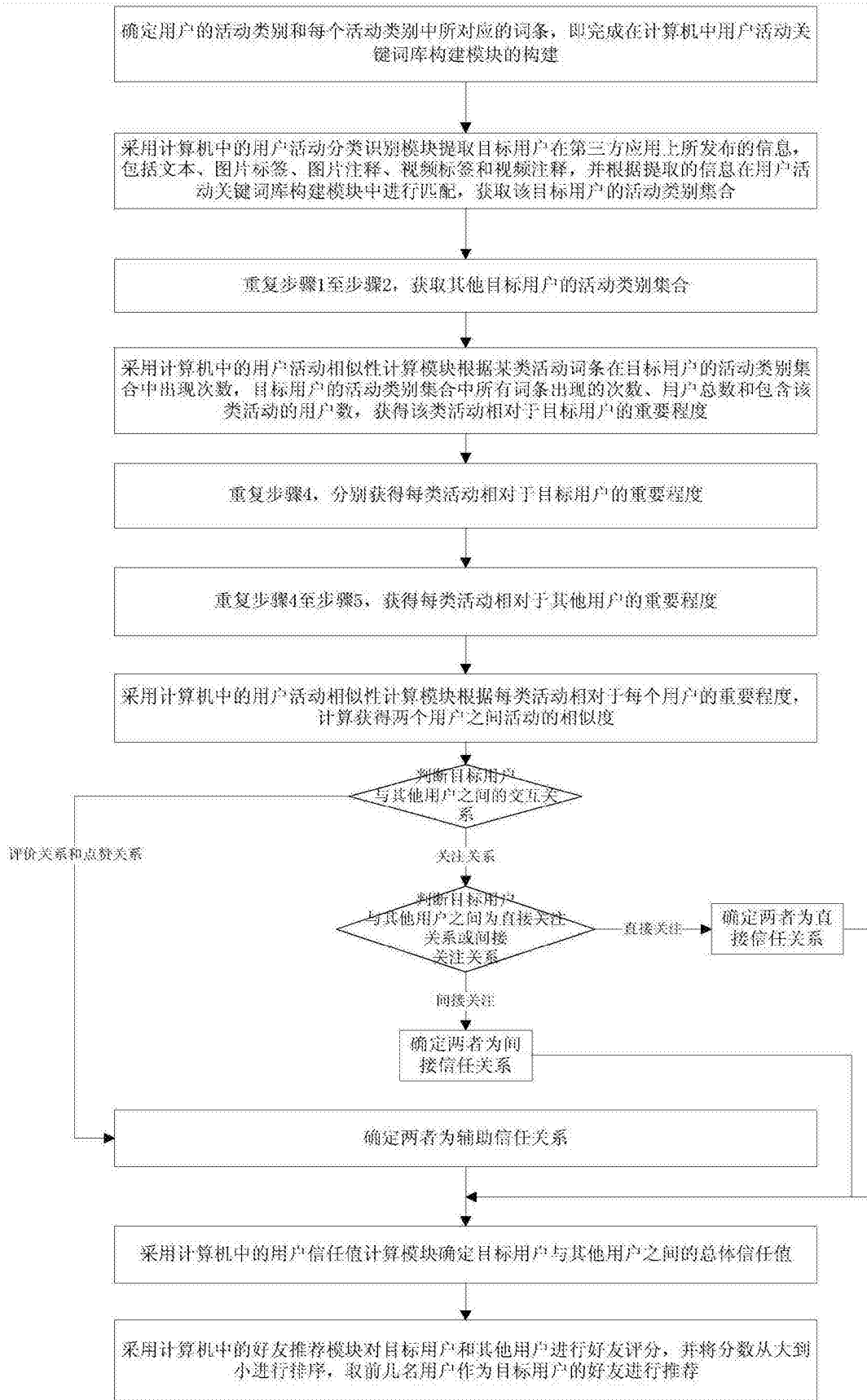


图2

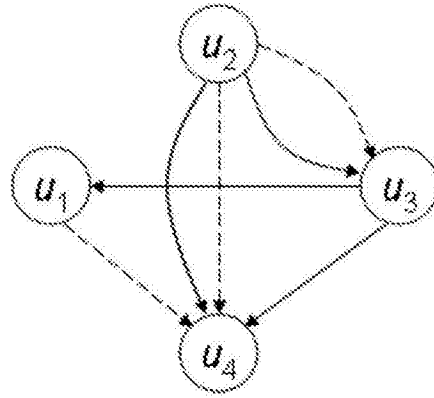


图3

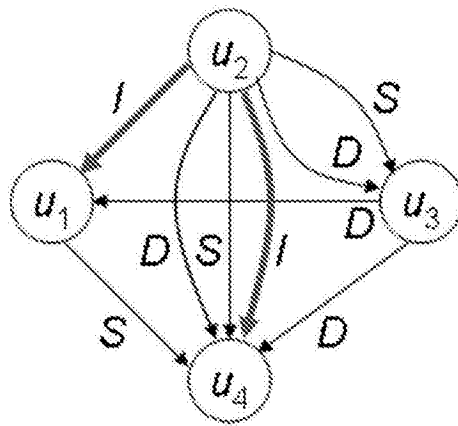


图4

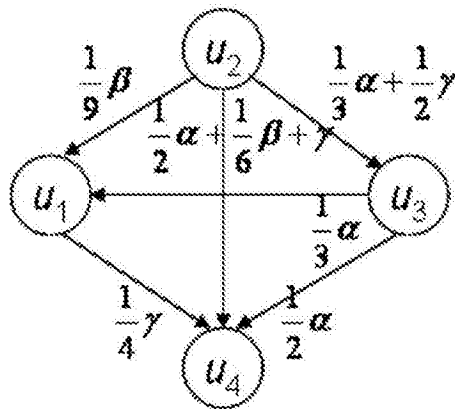


图5

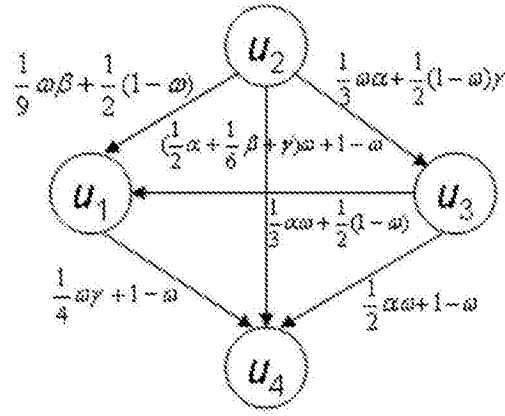


图6

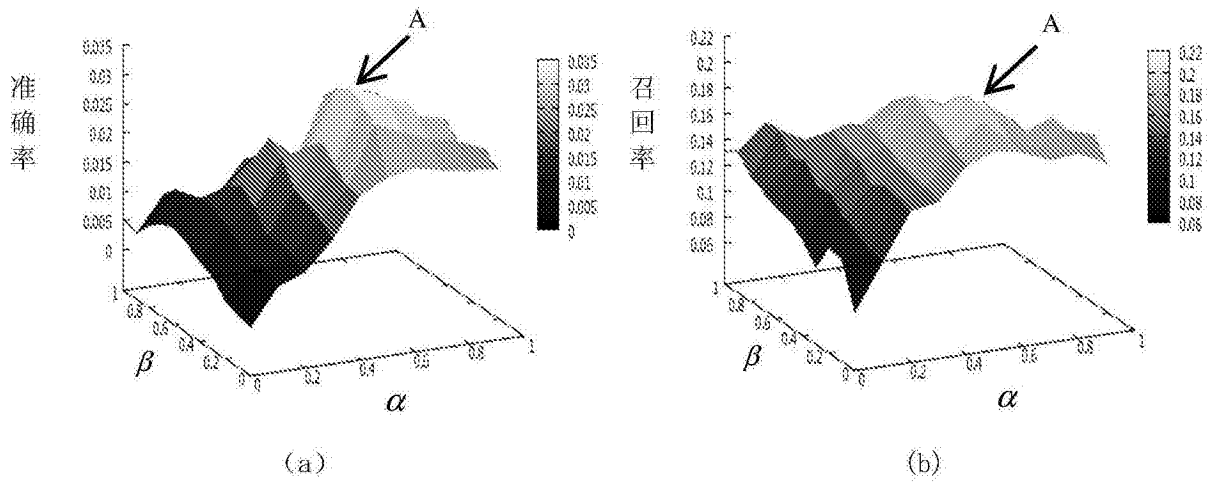


图7

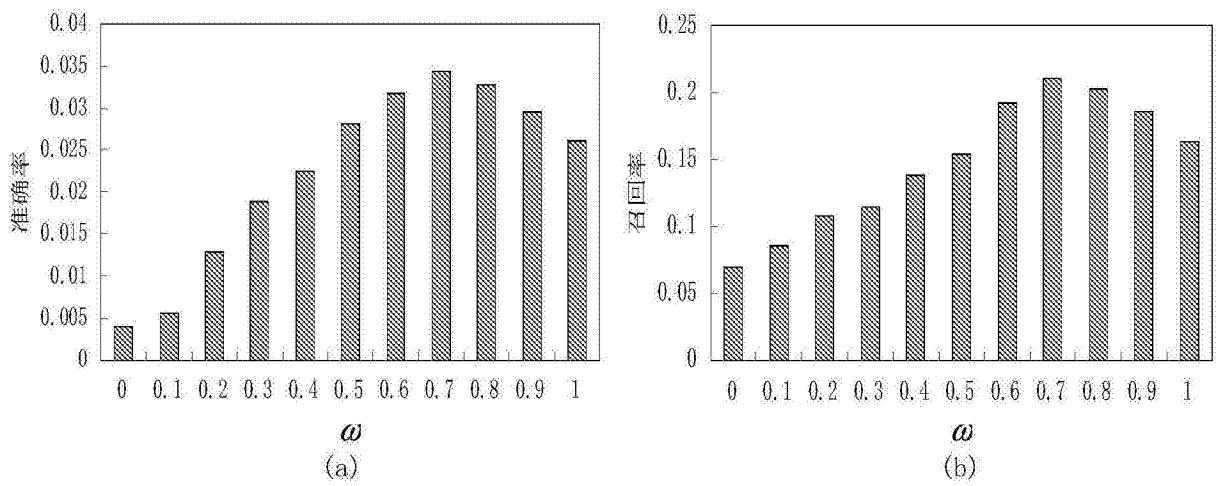


图8

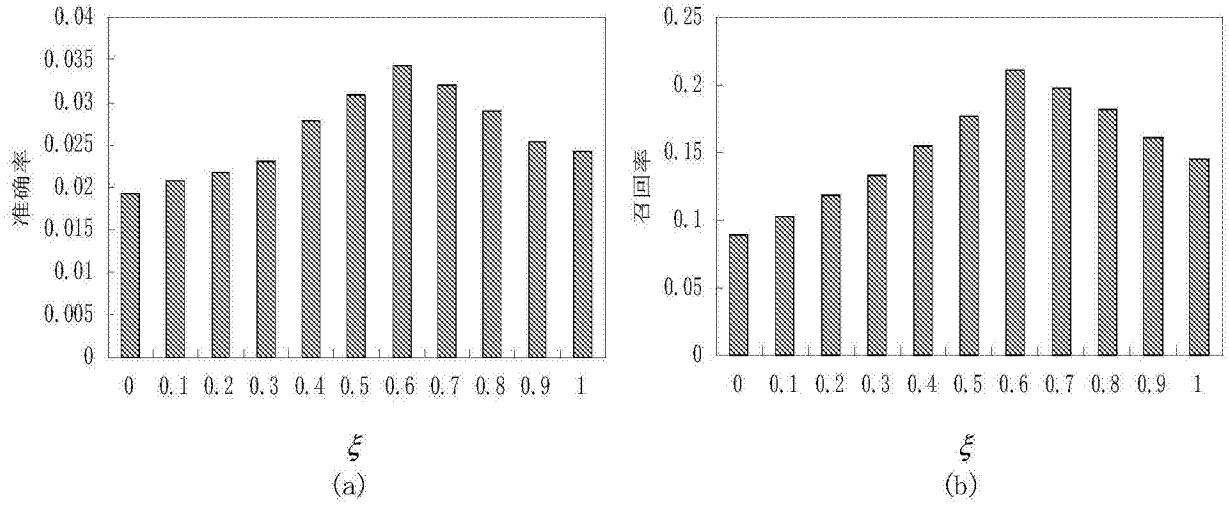


图9

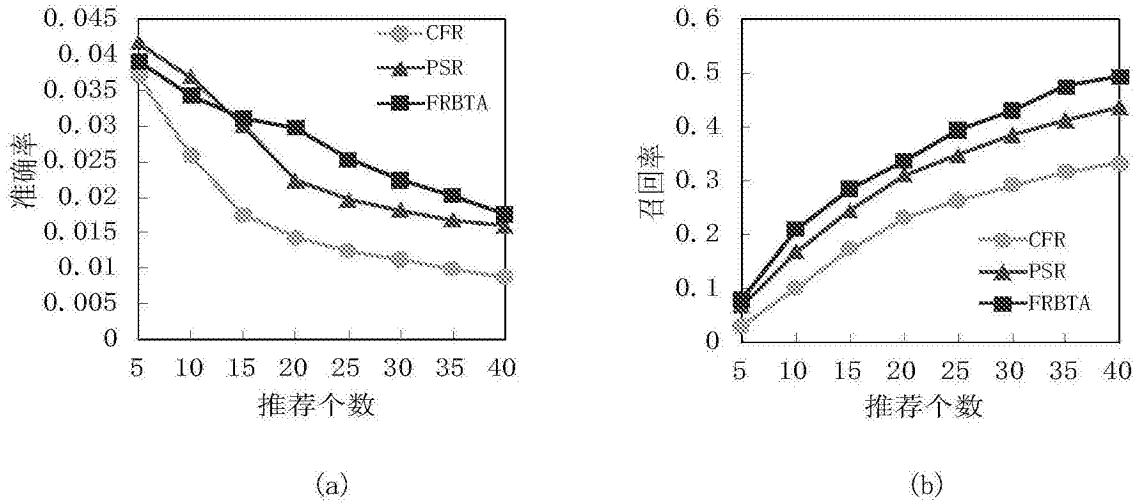


图10