



(21) 申请号 202111560166.5

G06Q 50/00 (2024.01)

(22) 申请日 2021.12.20

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 111984832 A, 2020.11.24

申请公布号 CN 114265988 A

CN 112507245 A, 2021.03.16

(43) 申请公布日 2022.04.01

审查员 王国海

(73) 专利权人 人工智能与数字经济广东省实验室(广州)

地址 510220 广东省珠海市海珠区新港东路2429号首层自编051房

(72) 发明人 魏哲巍 王涵之

(74) 专利代理机构 北京兴智翔达知识产权代理有限公司 11768

专利代理师 张玉梅

(51) Int. Cl.

G06F 16/9536 (2019.01)

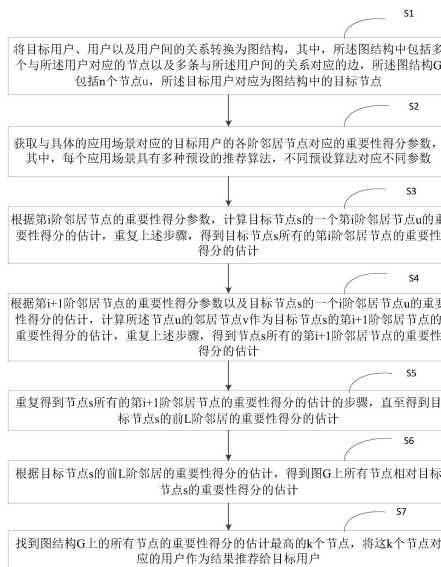
权利要求书3页 说明书11页 附图2页

(54) 发明名称

基于重要性得分的好友推荐方法

(57) 摘要

本发明公开了一种基于重要性得分的好友推荐方法,包括:将目标用户、用户以及用户间的关系转换为图结构;根据第*i*阶邻居节点的重要性得分参数,计算节点*s*所有的第*i*+1阶邻居节点的重要性得分的估计;得到目标节点*s*的前*L*阶邻居的重要性得分的估计;根据目标节点*s*的前*L*阶邻居的重要性得分的估计,得到图结构*G*上所有节点相对目标节点*s*的重要性得分的估计 $\hat{\pi}$;找到图结构*G*上的所有用户节点的重要性得分的估计 $\hat{\pi}$ 中最高k个用户节点,将这k个用户节点对应的社交用户作为结果推荐给目标用户。本实施例提供的基于重要性得分的好友推荐方法,实现基于各推荐场景的、通用的、可统一维护的好友推荐方法。



1. 一种基于重要性得分的好友推荐方法,其特征在于,包括:

将目标用户、用户以及用户间的关系转换为图结构,其中,所述图结构中包括多个与所述用户对应的节点以及多条与所述用户间的关系对应的边,所述图结构包括n个节点u,所述目标用户对应为图结构中的目标节点s;

获取与应用场景对应的目标用户的各阶邻居节点对应的重要性得分参数,其中,每个应用场景具有多种预设的推荐算法,不同预设算法对应不同参数;

根据第i阶邻居节点的重要性得分参数,计算目标节点s的一个第i阶邻居节点u的重要性得分的估计,重复上述步骤,得到目标节点s所有的第i阶邻居节点的重要性得分的估计;

根据第i+1阶邻居节点的重要性得分参数以及目标节点s的一个i阶邻居节点u的重要性得分的估计,计算所述节点u的邻居节点v作为目标节点s的第i+1阶邻居节点的重要性得分的估计,重复上述步骤,得到节点s所有的第i+1阶邻居节点的重要性得分的估计;

重复得到节点s所有的第i+1阶邻居节点的重要性得分的估计的步骤,直至得到目标节点s的前L阶邻居的重要性得分的估计,其中i为0-L的整数;

根据目标节点s的前L阶邻居的重要性得分的估计,得到图结构上所有节点相对目标节点s的重要性得分的估计 $\hat{\pi}$,其中, $\hat{\pi}$ 的第u维 $\hat{\pi}_u$ 存储的数值为图上节点u相对目标节点s的重要性得分的估计;

找到图结构上的所有用户节点的重要性得分的估计 $\hat{\pi}$ 中最高的k个节点,将这k个节点对应的用户作为结果推荐给目标用户。

2. 如权利要求1所述的好友推荐方法,其特征在于,根据第i阶邻居节点的重要性得分参数,计算目标节点s的一个第i阶邻居节点u的重要性得分的估计,重复上述步骤,得到目标节点s所有的第i阶邻居节点的重要性得分的估计包括:

获取目标节点s的一个第i阶邻居节点u,其中,所述目标节点s的第i阶邻居节点属于目标节点s的前L阶的邻居节点;

对目标节点s的第i阶邻居节点u设置残余值 $r_u^{(i)}$ 和保留值 $\hat{\pi}_u^{(i)}$,其中,残余值 $r_u^{(i)}$ 为目标节点s的第i阶邻居节点u的残余值,表示目标节点s的第i阶邻居节点u可能获得但还未获得的重要性得分,保留值 $\hat{\pi}_u^{(i)}$ 为目标节点s的第i阶邻居节点u的保留值,表示目标节点s的第i阶邻居节点u已获得的重要性得分;

使用公式三,根据节点u作为目标节点s的第i阶邻居的残余值 $r_u^{(i)}$,得到节点u作为目标节点s的第i阶邻居的保留值 $\hat{\pi}_u^{(i)}$,其中,节点u作为目标节点s的第i阶邻居的保留值 $\hat{\pi}_u^{(i)}$ 的初始值为0, $i=0,1,2,\dots,L$,节点u作为目标节点s的第i阶邻居的残余值 $r_u^{(i)}$ 的初始值为0, $i=0,1,2,\dots,L$,目标节点s作为目标节点s的第i阶邻居的残余值 $r_s^{(i)}$ 的初始值为0, $i=1,2,\dots,L$,目标节点s作为目标节点s的第0阶邻居的残余值 $r_s^{(0)}$ 的初始值为1;

所述公式三为: $\hat{\pi}_u^{(i)} = \frac{w_i}{Y_i} \cdot r_u^{(i)}$,

其中, w_i 为第i阶邻居节点的重要性得分参数中的权重参数, Y_i 为第i阶邻居节点的累积

权重参数, $Y_i = 1 - \sum_{x=0}^{i-1} w_x$;

获取目标节点s的下一个第i阶邻居节点u,重复上述步骤,直至得到目标节点s所有的第i阶邻居节点的重要性得分的估计。

3.如权利要求2所述的好友推荐方法,其特征在于,根据第i+1阶邻居节点的重要性得分参数以及目标节点s的一个i阶邻居节点u的重要性得分的估计,计算所述节点u的邻居节点v作为目标节点s的第i+1阶邻居节点的重要性得分的估计,重复上述步骤,得到节点s所有的第i+1阶邻居节点的重要性得分的估计包括:

获取节点u的一个邻居节点v;

根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)'}$,重复上述步骤直至得到目标节点s的所有i+1阶邻居的残余值;

根据目标节点s的第i+1阶邻居节点更新后的残余值 $r_v^{(i+1)'}$ 以及公式三,计算节点v作为目标节点s的第i+1阶邻居更新后的保留值 $\hat{\pi}_v^{(i+1)'}$;

获取目标节点s的下一个第i+1阶邻居节点v,重复上述步骤,直至得到目标节点s所有的第i+1阶邻居节点的重要性得分的估计。

4.如权利要求3所述的好友推荐方法,其特征在于,所述将目标用户、用户以及用户间的关系转换为图结构包括:

对图结构上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列,其中,节点u的邻居节点v为图上与节点u有边直接相连的节点;

获取节点u的邻居节点v包括:按照节点u排列后的邻居节点的顺序获取邻居节点v。

5.如权利要求4所述的好友推荐方法,其特征在于,所述对图结构上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列包括:

分别形成每个节点u的邻接表,其中,所述节点u的邻接表存储有节点u的所有邻居节点v,并且所述邻居节点v按照其度数的增序排列,所述节点v的度数等于节点v具有的邻居节点的个数;

所述按照节点u排列后的邻居节点的顺序获取邻居节点v包括:

在节点u的邻接表中,按邻居节点v度数递增的顺序获取节点u的邻居节点v。

6.如权利要求3所述的好友推荐方法,其特征在于,所述第i阶邻居节点的重要性得分参数包括权重参数 w_i 以及拉普拉斯系数a、b,

所述根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)'}$,包括:

当获取到的节点u的度数满足第一预设条件时,通过公式一计算节点v作为目标节点s

的第 $i+1$ 阶邻居更新后的残余值 $r_v^{(i+1)'};$

其中,所述第一预设条件为当节点 u 的度数 $d_u \leq \left(\frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b}\right)^{\frac{1}{a}},$

所述公式一: $r_v^{(i+1)'} = r_v^{(i+1)} + \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_v^a \cdot d_u^b},$

其中,节点 u 是节点 s 的一个 i 阶邻居节点,且 $r_u^{(i)} \neq 0$,节点 v 是节点 u 的邻居节点, ε 为误差参数。

7.如权利要求6所述的好友推荐方法,其特征在于,所述根据节点 u 作为目标节点 s 的第 i 阶邻居节点的残余值 $r_u^{(i)}$ 、节点 u 、 v 的度数 d_u 、 d_v 和节点 u 的邻居节点 v 以作为目标节点 s 的第 $i+1$ 阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点 v 作为目标节点 s 的第 $i+1$ 阶邻居节点的残余值 $r_v^{(i+1)'}$ 还包括:

若节点 u 的度不满足第一预设条件,对部分邻居节点 v 进行子集采样处理,所述部分邻居节点 v 为未按照公式一更新残余值的邻居节点 v ,对采样到的节点,通过公式二计算 $r_v^{(i+1)'}$;

公式二: $r_v^{(i+1)'} = r_v^{(i+1)} + \varepsilon;$

其中, ε 为误差参数。

8.如权利要求7所述的好友推荐方法,其特征在于,所述子集采样处理包括:

对于目标节点的第 i 阶邻居节点 u ,根据公式五计算节点 u 的部分邻居节点 v 的采样概率 p_v ,所述部分邻居节点 v 为未按照公式一更新残余值的邻居节点 v ,

公式五为: $p_v = \frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b} \cdot \frac{1}{d_v^a};$

将部分邻居节点 v 按照度数 d_v 划分为 $\log n$ 个集合,其中第 t 个集合中的节点的度数满足 $2^t \leq d_v \leq 2^{t+1}-1;$

对于第 t 个集合中的节点,按照该集合中所有节点对应的采样概率的最大值 $p_v^*(t)$ 进行伯努利采样;

对于伯努利采样到的节点 v ,以 $1 - \frac{p_v}{p_v^*(t)}$ 的概率放弃采样该节点;

将未放弃采样的节点作为采样到的节点。

基于重要性得分的好友推荐方法

技术领域

[0001] 本发明是关于好友推荐,特别是关于一种基于重要性得分的好友推荐方法。

背景技术

[0002] 随着我国互联网产业的高速发展,社交媒体的规模迅速增大。

[0003] 基于社交软件的核心业务,我国主流的社交软件可大致分为以微信、钉钉等为代表的通讯类社交软件;以微博、知乎为代表的媒体类社交软件;和以探探、soul为代表的交友类社交软件。各类型社交软件业务多样,功能各异,但是都普遍具有好友推荐的功能,即根据用户之间的关联关系,寻找对于目标用户具有较高重要性的用户,将这些高重要性的用户推荐给目标用户。高质量的好友推荐可以优化用户的使用体验、扩大社交范围、增进社交网络中的用户交互力度,进而提高用户粘性。因此,好友推荐的效率与质量对于社交产业的蓬勃发展至关重要。

[0004] 目前,不同社交软件会根据自身的功能定位选取不同的用户重要性衡量指标,同一社交软件也会在各类型的好友推荐中,依据具体的功能选择特定的用户重要性衡量指标进行好友推荐,基于不同用户重要性衡量指标的好友推荐方法不具有通用性。

[0005] 基于此,本申请的发明人发现,现有好友推荐方法的不通用性,会使得各场景下的好友推荐方法无法互用、难以统一维护,这对各社交软件维护好友推荐功能的过程带来了额外的负担,造成了潜在的资源 and 收益损耗。

[0006] 公开于该背景技术部分的信息仅仅旨在增加对本发明的总体背景的理解,而不应当被视为承认或以任何形式暗示该信息构成已为本领域一般技术人员所公知的现有技术。

发明内容

[0007] 本发明的目的在于提供一种基于重要性得分的好友推荐方法,其能够适用于多种好友推荐方法。

[0008] 为实现上述目的,本发明提供了一种基于重要性得分的好友推荐方法,包括:

[0009] 将目标用户、用户以及用户间的关系转换为图结构,其中,所述图结构中包括多个与所述用户对应的节点以及多条与所述用户间的关系对应的边,所述图结构G包括n个节点u,所述目标用户对应为图结构中的目标节点s;

[0010] 获取与应用场景对应的目标用户的各阶邻居节点对应的重要性得分参数,其中,每个应用场景具有多种预设的推荐算法,不同预设算法对应不同参数;

[0011] 根据第i阶邻居节点的重要性得分参数,计算目标节点s的一个第i阶邻居节点u的重要性得分的估计,重复上述步骤,得到目标节点s所有的第i阶邻居节点的重要性得分的估计;

[0012] 根据第i+1阶邻居节点的重要性得分参数以及目标节点s的一个i阶邻居节点u的重要性得分的估计,计算所述节点u的邻居节点v作为目标节点s的第i+1阶邻居节点的重要性得分的估计,重复上述步骤,得到节点s所有的第i+1阶邻居节点的重要性得分的估计;

[0013] 重复得到节点s所有的第i+1阶邻居节点的重要性得分的估计的步骤,直至得到目标节点s的前L阶邻居的重要性得分的估计,其中i为0-L的整数;

[0014] 根据目标节点s的前L阶邻居的重要性得分的估计,得到图结构G上所有节点相对目标节点s的重要性得分的估计 $\hat{\pi}$,其中, $\hat{\pi}$ 的第u维 $\hat{\pi}_u$ 存储的数值为图上节点u相对目标节点s的重要性得分的估计;

[0015] 找到图结构G上所有用户节点的重要性得分的估计 $\hat{\pi}$ 中最高的k个节点,将这k个节点对应的用户作为结果推荐给目标用户。

[0016] 在一优选的实施方式中,根据第i阶邻居节点的重要性得分参数,计算目标节点s的一个第i阶邻居节点u的重要性得分的估计,重复上述步骤,得到目标节点s所有的第i阶邻居节点的重要性得分的估计包括:

[0017] 获取目标节点s的一个第i阶邻居节点u,其中,所述目标节点s的第i阶邻居节点属于目标节点s的前L阶的邻居节点;

[0018] 对目标节点s的第i阶邻居节点u设置残余值 $r_u^{(i)}$ 和保留值 $\hat{\pi}_u^{(i)}$,其中,残余值 $r_u^{(i)}$ 为目标节点s的第i阶邻居节点u的残余值,表示目标节点s的第i阶邻居节点u可能获得但还未获得的重要性得分,保留值 $\hat{\pi}_u^{(i)}$ 为目标节点s的第i阶邻居节点u的保留值,表示目标节点s的第i阶邻居节点u已获得的重要性得分;

[0019] 使用公式三,根据节点u作为目标节点s的第i阶邻居的残余值 $r_u^{(i)}$,得到节点u作为目标节点s的第i阶邻居的保留值 $\hat{\pi}_u^{(i)}$,其中,节点u作为目标节点s的第i阶邻居的保留值 $\hat{\pi}_u^{(i)}$ 的初始值为0($i=0,1,2,\dots,L$),节点u作为目标节点s的第i阶邻居的残余值 $r_u^{(i)}$ 的初始值为0($i=0,1,2,\dots,L$),目标节点s作为目标节点s的第i阶邻居的残余值 $r_s^{(i)}$ 的初始值为0($i=1,2,\dots,L$),目标节点s作为目标节点s的第0阶邻居的残余值 $r_s^{(0)}$ 的初始值为1;

[0020] 所述公式三为: $\hat{\pi}_u^{(i)} = \frac{w_i}{Y_i} \cdot r_u^{(i)}$;

[0021] 获取目标节点s的下一个第i阶邻居节点u,重复上述步骤,直至得到目标节点s所有的第i阶邻居节点的重要性得分的估计。

[0022] 在一优选的实施方式中,根据第i+1阶邻居节点的重要性得分参数以及目标节点s的一个i阶邻居节点u的重要性得分的估计,计算所述节点u的邻居节点v作为目标节点s的第i+1阶邻居节点的重要性得分的估计,重复上述步骤,得到节点s所有的第i+1阶邻居节点的重要性得分的估计包括:

[0023] 获取节点u的一个邻居节点v;

[0024] 根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,重复上述步骤直至得到目标节点s的所有i+1阶邻居的残余值;

[0025] 根据目标节点s的第i+1阶邻居节点更新后的残余值 $r_v^{(i+1)'}$ 以及公式三,计算节点v作为目标节点s的第i+1阶邻居更新后的保留值 $\hat{\pi}_v^{(i+1)'}$;

[0026] 获取目标节点s的下一个第i+1阶邻居节点v,重复上述步骤,直至得到目标节点s所有的第i+1阶邻居节点的重要性得分的估计。

[0027] 在一优选的实施方式中,所述将目标用户、用户以及用户间的关系转换为图结构包括:

[0028] 对图结构G上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列,其中,节点u的邻居节点v为图上与节点u有边直接相连的节点;

[0029] 获取节点u的邻居节点v包括:按照节点u排列后的邻居节点的顺序获取邻居节点v。

[0030] 在一优选的实施方式中,所述对图结构G上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列包括:

[0031] 分别形成每个节点u的邻接表,其中,所述节点u的邻接表存储有节点u的所有邻居节点v,并且所述邻居节点v按照其度数的增序排列,所述节点v的度数等于节点v具有的邻居节点的个数;

[0032] 所述按照节点u排列后的邻居节点的顺序获取邻居节点v包括:

[0033] 在节点u的邻接表中,按邻居节点v度数递增的顺序获取节点u的邻居节点v。

[0034] 在一优选的实施方式中,所述第i阶邻居节点的重要性得分参数包括权重参数 w_i 以及拉普拉斯系数a、b,

[0035] 所述根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)'}$, 包括:

[0036] 当获取到的节点u的度数满足第一预设条件时,通过公式一计算节点v作为目标节点s的第i+1阶邻居更新后的残余值 $r_v^{(i+1)'}$;

[0037] 其中,所述第一预设条件为当节点u的度数 $d_u \leq \left(\frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b} \right)^{\frac{1}{a}}$,

[0038] 所述公式一: $r_v^{(i+1)'} = r_v^{(i+1)} + \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_v^a \cdot d_u^b}$,

[0039] 其中,节点u是节点s的一个i阶邻居节点,且 $r_u^{(i)} \neq 0$,节点v是节点u的邻居节点。

[0040] 在一优选的实施方式中,所述根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)'}$,还包括:

[0041] 若节点u的度不满足第一预设条件,对部分邻居节点v进行子集采样处理,所述部分邻居节点v为未按照公式一更新残余值的邻居节点v,对采样到的节点,通过公式二计算 $r_v^{(i+1)'};$

[0042] 公式二: $r_v^{(i+1)'} = r_v^{(i+1)} + \varepsilon;$

[0043] 其中, ε 为误差参数。

[0044] 在一优选的实施方式中,所述子集采样处理包括:

[0045] 对于目标节点的第i阶邻居节点u,根据公式五计算节点u的部分邻居节点v的采样概率 p_v ,所述部分邻居节点v为未按照公式一更新残余值的邻居节点v,

[0046] 公式五为: $p_v = \frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b} \cdot \frac{1}{d_v^a};$

[0047] 将部分邻居节点v按照度数 d_v 划分为 $\log n$ 个集合,其中第t个集合中的节点的度数满足 $2^t \leq d_v \leq 2^{t+1} - 1;$

[0048] 对于第t个集合中的节点,按照该集合中所有节点对应的采样概率的最大值 $p_v^*(t)$ 进行伯努利采样;

[0049] 对于伯努利采样到的节点v,以 $1 - \frac{p_v}{p_v^*(t)}$ 的概率放弃采样该节点;

[0050] 将未放弃采样的节点作为采样到的节点。

[0051] 与现有技术相比,通过本实施例提供的基于重要性衡量指标的好友推荐方法,可以通过将各阶邻居节点的重要性得分参数与主流的用户重要性衡量指标对应起来,使用目前主流的用户重要性衡量指标作为重要性得分参数,并且通过重要性得分参数的选取实现基于各推荐场景的、通用的、可统一维护的好友推荐方法。

附图说明

[0052] 图1是根据本发明一实施方式的基于重要性得分的好友推荐方法的流程图。

[0053] 图2是根据本发明一实施方式的节点示意图。

具体实施方式

[0054] 下面结合附图,对本发明的具体实施方式进行详细描述,但应当理解本发明的保护范围并不受具体实施方式的限制。

[0055] 除非另有其它明确表示,否则在整个说明书和权利要求书中,术语“包括”或其变换如“包含”或“包括有”等等将被理解为包括所陈述的元件或组成部分,而并未排除其它元件或其它组成部分。

[0056] 在社交网络中,用户重要性得分的衡量标准会受到推荐场景的影响,在不同的推荐场景下,好友推荐方法会选择不同的用户重要性衡量指标进行用户重要性得分的计算。现有的用户重要性衡量指标主要包括:个性化佩奇排名、热核排名、K步转移概率、卡茨指标等。但是,现有的好友推荐方法普遍只针对单一的用户重要性衡量指标,无法基于多种用户重要性衡量指标进行用户重要性得分的通用计算。一旦更改用户重要性衡量指标,例如从个性化佩奇排名指标修改为热核排名指标,现有的好友推荐方法则会失效,即基于个性化

佩奇排名的用户重要性得分计算方法无法直接用于计算基于热核排名的用户重要性得分。

[0057] 而具体选择哪一种用户重要性衡量指标,是需要根据实际场景去选择的。这里需要先明确“阶数”的概念,阶数用于衡量用户之间好友关系的亲疏程度。如果用户A和用户B在社交网络上存在好友关系,则用户A和B是1阶邻居。如果用户A和用户B之间没有直接的好友关系,但是用户A和用户C是1阶邻居,用户C和用户B又是1阶邻居,即如果用户A需要联系用户B,则需要先找到用户C,再从用户C找到B,我们称用户A和B是2阶邻居。不同的用户重要性衡量指标,对于各阶邻居的重要性衡量存在差异。例如,个性化佩奇排名更重视低阶邻居的重要性,而热核排名会比较轻视低阶邻居、重视中阶邻居、忽视高阶邻居。因此,具体的社交网络会根据其自身属性,选择恰当的用户重要性衡量指标,进行用户间相对重要性得分的计算,最终完成适合自身社交平台的好友推荐。例如,微信网络更侧重于日常沟通,关注近邻关系,所以适合使用基于个性化佩奇排名的用户重要性得分计算方法,而微博网络更侧重新鲜资讯获取,对近邻好友的看重程度有所降低,更注重距离稍远的好友关系,更适合使用基于热核排名的用户重要性得分计算方法。

[0058] 如图1所示,根据本发明优选实施方式的基于重要性得分的好友推荐方法的流程图,包括步骤S1-S6。

[0059] 步骤S1,将目标用户、用户以及用户间的关系转换为图结构,其中,所述图结构中包括多个与所述用户对应的节点以及多条与所述用户间的关系对应的边,所述图结构G包括n个节点v,所述目标用户对应为图结构中的目标节点s。

[0060] 将社交网络转化为图论中的图结构 $G=(V,E)$,其中,V、E分别表示图结构上的所有节点、边组成的集合,对应于实际社交网络中的用户群体和用户之间的好友关系。

[0061] 本实施例的用户指的是社交平台上的所有注册用户,用户间的关系可以是用户间的好友关系。

[0062] 具体的,我们将社交网络的用户对应到图结构上的节点,用户之间的好友关系对应到图结构上的边。例如在微信、QQ这类有好友关系的社交网络中,如果A用户和B用户之间存在好友关系,则图结构上需建立一条A用户节点和B用户节点之间的边。由此,将社交网络上的好友关系转化为图结构上节点之间的连边关系。

[0063] 步骤S2,获取与具体的应用场景对应的目标用户的各阶邻居节点对应的重要性得分参数,其中,每个应用场景具有多种预设的推荐算法,不同预设算法对应不同参数。

[0064] 步骤S3,根据第i阶邻居节点的重要性得分参数,计算目标节点s的一个第i阶邻居节点u的重要性得分的估计,重复上述步骤,得到目标节点s所有的第i阶邻居节点的重要性得分的估计。

[0065] 步骤S4,根据第i+1阶邻居节点的重要性得分参数以及目标节点s的一个i阶邻居节点u的重要性得分的估计,计算所述节点u的邻居节点v作为目标节点s的第i+1阶邻居节点的重要性得分的估计,重复上述步骤,得到节点s所有的第i+1阶邻居节点的重要性得分的估计。

[0066] 步骤S5,重复得到节点s所有的第i+1阶邻居节点的重要性得分的估计的步骤,直至得到目标节点s的前L阶邻居的重要性得分的估计,其中i为0-L的整数。即重复步骤S3,S4总共L次。

[0067] 需要说明的是,步骤S3-S5是更新完所有i阶邻居的保留值(根据公式三),再更新

所有 $i+1$ 阶邻居的残余值(满足第一预设条件的按照公式一更新,其他的进行子集采样,采样到的按照公式二更新,未采样到的不更新),再更新所有 $i+1$ 阶邻居的保留值,再更新所有 $i+2$ 阶邻居的残余值,以此类推直至更新完 L 阶邻居。

[0068] 步骤S6,根据目标节点 s 的前 L 阶邻居的重要性得分的估计,得到图结构 G 上所有节点相对目标节点 s 的重要性得分的估计 $\hat{\pi}$,其中, $\hat{\pi}$ 的第 u 维 $\hat{\pi}_u$ 存储的数值为图上节点 u 相对目标节点 s 的重要性得分的估计。

[0069] 步骤S7,找到图结构 G 上的所有节点的重要性得分的估计 $\hat{\pi}$ 中最高的 k 个节点,将这 k 个节点对应的用户作为结果推荐给目标用户。

[0070] 由此,通过本实施例提供的基于重要性衡量指标的好友推荐方法,可以通过将各阶邻居节点的重要性得分参数与主流的用户重要性衡量指标对应起来,使用目前主流的用户重要性衡量指标作为重要性得分参数,并且通过重要性得分参数的选取实现基于各推荐场景的、通用的、可统一维护的好友推荐方法。

[0071] 其中,不同的用户重要性衡量指标可体现为不同的邻居节点的重要性得分参数。

[0072] 在一种实现方式中,图1中的步骤S3-S5可以包括:

[0073] 步骤S301,获取目标节点 s 的一个第 i 阶邻居节点 u 。

[0074] 其中,所述目标节点 s 的第 i 阶邻居节点属于目标节点 s 前 L 阶的邻居节点。

[0075] 需要说明的是,节点 u 虽然限定的是除目标节点外的用户节点,但是在具体实现时,并不限于此种方式。例如,节点 u 为 s 的 i 阶邻居节点,当 $i=0$ 时,其可以是节点 s 本身。

[0076] 步骤S302,对目标节点 s 的第 i 阶邻居节点 u 设置残余值 $r_u^{(i)}$ 和保留值 $\hat{\pi}_u^{(i)}$ 。其中,残余值 $r_u^{(i)}$ 为目标节点 s 的第 i 阶邻居节点 u 的残余值,表示目标节点 s 的第 i 阶邻居节点 u 可能获得但还未获得的重要性得分,保留值 $\hat{\pi}_u^{(i)}$ 为目标节点 s 的第 i 阶邻居节点 u 的保留值,表示目标节点 s 的第 i 阶邻居节点 u 已获得的重要性得分。具体的,将 $\hat{\pi}_u^{(i)}$ 当作节点 u 作为目标节点 s 的第 i 阶邻居时的重要性得分。

[0077] 步骤S303,使用公式三,根据第 i 阶邻居节点的重要性得分参数和节点 u 作为目标节点 s 的第 i 阶邻居的残余值 $r_u^{(i)}$,得到节点 u 作为目标节点 s 的第 i 阶邻居的保留值 $\hat{\pi}_u^{(i)}$ 。

[0078] 其中,节点 u 作为目标节点 s 的第 i 阶邻居的保留值 $\hat{\pi}_u^{(i)}$ 的初始值为0($i=0,1,2,\dots,L$),节点 u 作为目标节点 s 的第 i 阶邻居的残余值 $r_u^{(i)}$ 的初始值为0($i=1,2,\dots,L$),目标节点 s 作为目标节点 s 的第 i 阶邻居的残余值 $r_s^{(i)}$ 初始值为0($i=1,2,\dots,L$),目标节点 s 作为目标节点 s 的第0阶邻居的残余值 $r_s^{(0)}$ 初始值为1。

[0079] 公式三: $\hat{\pi}_u^{(i)} = \frac{w_i}{Y_i} \cdot r_u^{(i)}$,

[0080] 其中, w_i 为第 i 阶邻居节点的重要性得分参数中的权重参数, Y_i 为累积权重参数, $Y_i = 1 - \sum_{x=0}^{i-1} w_x$ 。

[0081] 需要说明的是,起始状态下,给定目标节点s,对于图结构上的所有用户节点u,其作为目标节点的各阶邻居 $r_u^{(i)}$ ($i=0,1,2,\dots,L$)的初始值都被设为0,除了 $r_s^{(0)}=1$ 。这表示在起始状态下,只有目标节点可能获得重要性得分,其他节点都暂无获得重要性得分的可能性。之后,重复地对图上各节点的保留值、残余值进行更新,每次更新中,都会将一部分残留值分配给保留值,即分配一部分节点u可能获得但还未获得的重要性得分。(对应步骤S303,S402,S403)

[0082] 公式三表示,在一次重要性得分的计算中,根据节点u作为目标节点s的第i阶邻居可能获得但还未获得的重要性得分,更新其已获得的重要性得分。具体而言,节点u作为目标节点s的第i阶邻居,其可能获得但暂未获得的重要性得分(记录在 $r_u^{(i)}$ 中)有一部分($\frac{w_i}{Y_i} \cdot r_u^{(i)}$)被确定地分配给 $\hat{r}_u^{(i)}$,即 $\hat{r}_u^{(i)}$ 在一次重要性得分计算中,可以被分配到的重要性得分为: $\frac{w_i}{Y_i} \cdot r_u^{(i)}$ 。

[0083] 在一种实现方式中,步骤S1还可以包括:对图结构G上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列,其中,节点u的邻居节点v为图上与节点u有边直接相连的节点。

[0084] 步骤S4可以包括:

[0085] 步骤S401,获取节点u的一个邻居节点v。其中,节点v为节点u的邻居节点,因为节点u为目标节点的第i阶的邻居节点,故节点u的邻居节点v为目标节点s的第i+1阶的邻居节点。

[0086] 在一种实现方式中,获取节点u的邻居节点v指按照节点u的排序后的邻居节点的顺序获取邻居节点v。

[0087] 在一种实现方式中,所述对图结构G上的每个用户节点u的邻居节点v,按照节点u的邻居节点v的度数的增序排列包括:分别形成每个节点u的邻接表,其中,所述节点u的邻接表存储有节点u的所有邻居节点v,并且所述邻居节点v按照其度数的增序排列。

[0088] 具体的,将实际的社交网络结构转化为图结构时,将图上所有节点的邻居节点按照度数递增的顺序排序。例如,对于图上的节点u,与其有直接连边关系的节点称为节点u的邻居节点v,节点u的邻居节点的个数称为节点u的度数,用 d_u 表示。将节点u的所有邻居节点记录在一个表结构中,称为节点u的邻接表 N_u ,邻接表 N_u 中共记录了 d_u 个节点,将邻居节点按照度数递增的顺序排序即保证这 d_u 个节点是按照其度数的增序排列的。

[0089] 相应的,所述按照节点u排列后的邻居节点的顺序获取节点v可以包括:在节点u的邻接表中,按邻居节点v度数递增的顺序获取节点u的邻居节点v。

[0090] 步骤S402,根据节点u作为目标节点s的第i阶邻居节点的残余值 $r_u^{(i)}$ 、节点u、v的度数 d_u 、 d_v 和节点u的邻居节点v以作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,得到更新后的节点v作为目标节点s的第i+1阶邻居节点的残余值 $r_v^{(i+1)}$,重复上述步骤直至得到目标节点s的所有i+1阶邻居的残余值;

[0091] 步骤S403,根据目标节点s的第i+1阶邻居节点更新后的残余值 $r_v^{(i+1)'}$ 以及公式三,计算节点v作为目标节点s的第i+1阶邻居更新后的保留值 $\hat{\pi}_v^{(i+1)'}$;

[0092] 具体的,可以通过将目标节点s的一个第i+1阶邻居节点v作为公式三中的节点u,根据节点v作为目标节点s的第i+1阶邻居的残余值 $r_v^{(i+1)}$,得到节点v作为目标节点s的第i+1阶邻居的保留值 $\hat{\pi}_v^{(i+1)}$;

[0093] 步骤S404,获取目标节点s的下一个第i+1阶邻居节点v,重复上述步骤,直至得到目标节点s所有的第i+1阶邻居节点的重要性得分的估计。

[0094] 需要说明的是,步骤S301,302只在初始计算时执行,后续重复S303-S404,得到目标节点s的前L阶邻居节点的残余值和保留值。

[0095] 具体的,步骤402可以包括:

[0096] 步骤4021,当获取到的节点u的邻居节点v满足第一预设条件时,通过公式一计算节点v作为目标节点s的第i+1阶邻居更新后的残余值 $r_v^{(i+1)'}$;

[0097] 其中,所述第一预设条件为当节点u的度 $d_u \leq \left(\frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b} \right)^{\frac{1}{a}}$,

[0098] 所述公式一: $r_v^{(i+1)'} = r_v^{(i+1)} + \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_v^a \cdot d_u^b}$,

[0099] 其中,节点u是节点s的一个i阶邻居节点,且 $r_u^{(i)} \neq 0$,节点v是节点u的邻居节点,第i阶邻居节点的重要性得分参数包括权重参数 w_i 以及拉普拉斯系数a、b。

[0100] 其中, Y_i 为累积权重参数, $Y_i = 1 - \sum_{x=0}^{i-1} w_x$, ε 是一个误差参数, $\varepsilon = \frac{200\delta}{\log^2 \frac{1}{\delta}}$ 为本实施例使用者根据实际情况选择的重要性得分的误差参数阈值。

[0101] 步骤4022,若不满足第一预设条件,对部分邻居节点v进行子集采样处理,所述部分邻居节点v为未按照公式一更新残余值的邻居节点v,对采样到的节点,通过公式二计算 $r_v^{(i+1)'}$;

[0102] 公式二: $r_v^{(i+1)'} = r_v^{(i+1)} + \varepsilon$, ε 为误差参数。

[0103] 所述子集采样处理包括:

[0104] 对于目标节点的第i阶邻居节点u,根据公式五计算节点u的部分邻居节点v的采样概率 p_v ,计算式为公式五,所述部分邻居节点v为未按照公式一更新残余值的邻居节点v。

[0105] 公式五: $p_v = \frac{1}{\varepsilon} \cdot \frac{Y_{i+1}}{Y_i} \cdot \frac{r_u^{(i)}}{d_u^b} \cdot \frac{1}{d_v^a}$ 。

[0106] 将部分邻居节点v按照度数 d_v 划分为 $\log n$ 个集合,其中第t个集合中的节点满足 $2^t \leq d_v \leq 2^{t+1}-1$;

[0107] 对于第t个集合中的节点,按照该集合中所有节点对应的采样概率的最大值 $p_v^*(t)$

进行伯努利采样；

[0108] 对于伯努利采样到的节点 v ,以 $1 - \frac{p_v}{p_v^*(t)}$ 的概率放弃采样该节点；

[0109] 将未放弃采样的节点作为采样到的节点。

[0110] 以下给出更新残余值和保留值的实例。

[0111] 参阅图2,其为本发明一实施方式的节点示意图。节点 u_1 、 u_2 、 u_3 是节点 s 的第1阶邻居,节点 v_1 、 v_2 、 v_3 是节点 s 的第2阶邻居。

[0112] 第1轮更新:

[0113] $r_s^{(0)} = 1$ (初始值设定),根据公式三计算 $\hat{\pi}_s^{(0)} = \frac{w_0}{Y_0} \cdot r_s^{(0)}$;若节点 u_1 、 u_2 的度数满足第一预设条件、节点 u_3 不满足,则根据公式一更新 $r_{u_1}^{(1)}$ 、 $r_{u_2}^{(1)}$,并对节点 u_3 进行子集采样处理,再判断 u_3 是否被采样到,若被采样到则根据公式二更新 $r_{u_3}^{(1)}$,否则不更新 $r_{u_3}^{(1)}$ 。

[0114] 第2轮更新:

[0115] 根据公式三,计算 $\hat{\pi}_{u_1}^{(1)}$ 、 $\hat{\pi}_{u_2}^{(1)}$ 、 $\hat{\pi}_{u_3}^{(1)}$;若 $r_{u_1}^{(1)} > 0$,根据公式一更新节点 u_1 的满足第一预设条件的邻居 v 的残余值 $r_v^{(2)}$;否则,根据公式二更新采样到的邻居 v 的残余值 $r_v^{(2)}$ 。若 $r_{u_2}^{(1)} > 0$,根据公式一更新节点 u_2 的满足第一预设条件的邻居 v 的残余值 $r_v^{(2)}$;否则,根据公式二更新采样到的邻居 v 的残余值 $r_v^{(2)}$ 。若 $r_{u_3}^{(1)} > 0$,根据公式一更新节点 u_3 的满足第一预设条件的邻居 v 的残余值 $r_v^{(2)}$;否则,根据公式二更新采样到的邻居 v 的残余值 $r_v^{(2)}$ 。

[0116] 第3轮更新:

[0117] 根据公式三,计算 $\hat{\pi}_{v_1}^{(2)}$ 、 $\hat{\pi}_{v_2}^{(2)}$ 、 $\hat{\pi}_{v_3}^{(2)}$;对所有节点 v ,若残余值 $r_v^{(2)} > 0$,根据公式一更新节点 v 的满足第一预设条件的邻居 z 的残余值 $r_z^{(3)}$;否则,根据公式二更新采样到的邻居 z 的残余值 $r_z^{(3)}$ 。

[0118] 步骤S6,可以通过公式四进行实现,公式四为:

$$[0119] \quad \hat{\pi} = \sum_{i=0}^L \hat{\pi}^{(i)},$$

[0120] 其中 $\hat{\pi}$ 是一个 n 维向量, n 为图上的节点数(即为社交网络中的用户数), $\hat{\pi}$ 的第 u 维 $\hat{\pi}_u$ 记录的数值为图上节点 u 相对目标节点 s 的重要性得分的估计。 $\hat{\pi}^{(i)}$ 也是一个 n 维向量, $\hat{\pi}^{(i)}$ 的第 u 维记录的数值 $\hat{\pi}_u^{(i)}$ 为图上节点 u 作为目标节点 s 的第 i 阶邻居的保留值, $\hat{\pi}_u^{(i)}$ 被视作节点 u 作为目标节点 s 的第 i 阶邻居时的重要性得分的估计。

[0121] i 是一个递增变量,取值为0到 L 之间的所有整数。 $\hat{\pi}^{(0)}$ 是已知的,我们会根据 $\hat{\pi}^{(0)}$ 计算出 $\hat{\pi}^{(1)}$,再计算出 $\hat{\pi}^{(2)}$,最终计算出 $\hat{\pi}^{(L)}$ 。最后根据前 $L+1$ 个($i=0,1,\dots,L$) $\hat{\pi}^{(i)}$ 计算得到 $\hat{\pi}$ 。

[0122] δ 是一个误差阈值,需要本实施例的使用者根据实际情况选择 δ 的具体取值。本发明可以保证,对于社交网络上的任意用户 u (即对应图上的任意节点 u),按照本发明提供的计算步骤,计算得到的节点 u 关于目标节点 s 的重要性得分估计值 $\hat{\pi}_u$ 与节点 u 关于目标节点 s 的重要性得分准确值 π_u 之间,满足下述关系:

[0123] 当 $\vec{\pi}_u > \delta$ 时, $|\vec{\pi}_u - \hat{\pi}_u| > \frac{1}{10} \vec{\pi}_u$ 出现的可能性不超过1%。即在99%的可能性下,本发明计算得到的节点 u 关于目标节点 s 的重要性得分估计值 $\hat{\pi}_u$ 与真实值 π_u 之间的相对误差不超过 $\frac{1}{10}$ 。

[0124] L 为推荐好友的最大允许阶数, L 的取值由误差参数 δ 完全确定。具体地, $L = \log \frac{1}{\delta}$ 。 ε 是为了简化本发明中的计算过程而设置的一个中间误差参数,其可由误差阈值 δ 完全决定。具体的, $\varepsilon = \frac{200\delta}{\log^2 \frac{1}{\delta}}$ 。

[0125] 以下给出目前主流的用户重要性衡量指标对应的重要性得分参数的设置方式:

[0126] 基于个性化佩奇排名指标计算重要性得分时,重要性得分参数的设置方式为: $w_i = \alpha(1-\alpha)^i$, $a=0$, $b=1$, 其中, α 是参数,取值范围为 $\alpha \in (0, 1)$, 个性化佩奇排名认为目标节点 s 的越高阶的邻居节点相对目标节点 s 的重要性得分越低;基于热核排名指标计算重要性

得分时,重要性得分参数的设置方式为: $w_i = e^{-t} \cdot \frac{t^i}{i!}$, $a=0$, $b=1$, 其中, e 是自然对数, t 是参数,热核排名认为第 t 阶邻居的重要性最高;基于1步转移概率计算重要性得分时,重要性得分参数的设置方式为 $w_1=1, w_i=0 (i=0, 1, \dots, l-1)$, $a=0$, $b=1$, 其中, l 为参数, i 是一个递增变量,取值为0到 $l-1$ 之间的所有整数,1步转移概率认为只有第1阶邻居重要,其他各阶邻居均不重要;基于卡茨指标计算重要性得分时,重要性得分参数的设置方式为: $w_i = \beta^i$, $a=0$, $b=0$, 其中, β 是参数,保证最后重要性得分是唯一的、是有限数值的。

[0127] 用户需要根据实际的好友推荐场景,确定通用重要性得分计算范式中参数 w_i 、 a 、 b 的取值。例如,如果好友推荐场景中低阶邻居用户的重要性更高,则选择基于个性化佩奇排名指标的重要性得分计算公式更适合,即设置通用重要性衡量范式中的重要性得分参数为 $w_i = \alpha(1-\alpha)^i$, $a=0$, $b=1$ 。

[0128] 由此,本实施例提供的方法可得到任意误差参数阈值对应的重要性得分,需要消耗的时间为 $O\left(\frac{L}{\delta} \cdot \sum_{i=1}^L \|Y_i \cdot (D^{-a}AD^{-b})^i \cdot \vec{e}_s\|_1\right)$ 。其中, D 为由实际网络转化后得到的图结构的度数矩阵, D 的对角线上记录图结构上各节点的度数, D 非对角线上的值为0。具体地,矩阵 D 对角线上第 u 个值对应图上节点 u 的度数 d_u 。 A 为由实际网络转化后得到的图结构的邻接矩阵, A 的第 u 行第 v 列的值若为1,则表示节点 u 、 v 之间有边相连,否则 A 的第 u 行第 v 列的值为0。 $\vec{e}_s$ 是一个与目标用户 s 对应的独热向量,该向量第 s 维的值为1,其他维度的值均为0。

[0129] 这一时间与用户重要性得分计算的输出结果大小处于同一级别,因此本实施例的好友推荐方法的计算时间是近似最优的。

[0130] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0131] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0132] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0133] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0134] 前述对本发明的具体示例性实施方案的描述是为了说明和例证的目的。这些描述并非想将本发明限定为所公开的精确形式,并且很显然,根据上述教导,可以进行很多改变和变化。对示例性实施例进行选择 and 描述的目的在于解释本发明的特定原理及其实际应用,从而使得本领域的技术人员能够实现并利用本发明的各种不同的示例性实施方案以及各种不同的选择和改变。本发明的范围意在由权利要求书及其等同形式所限定。

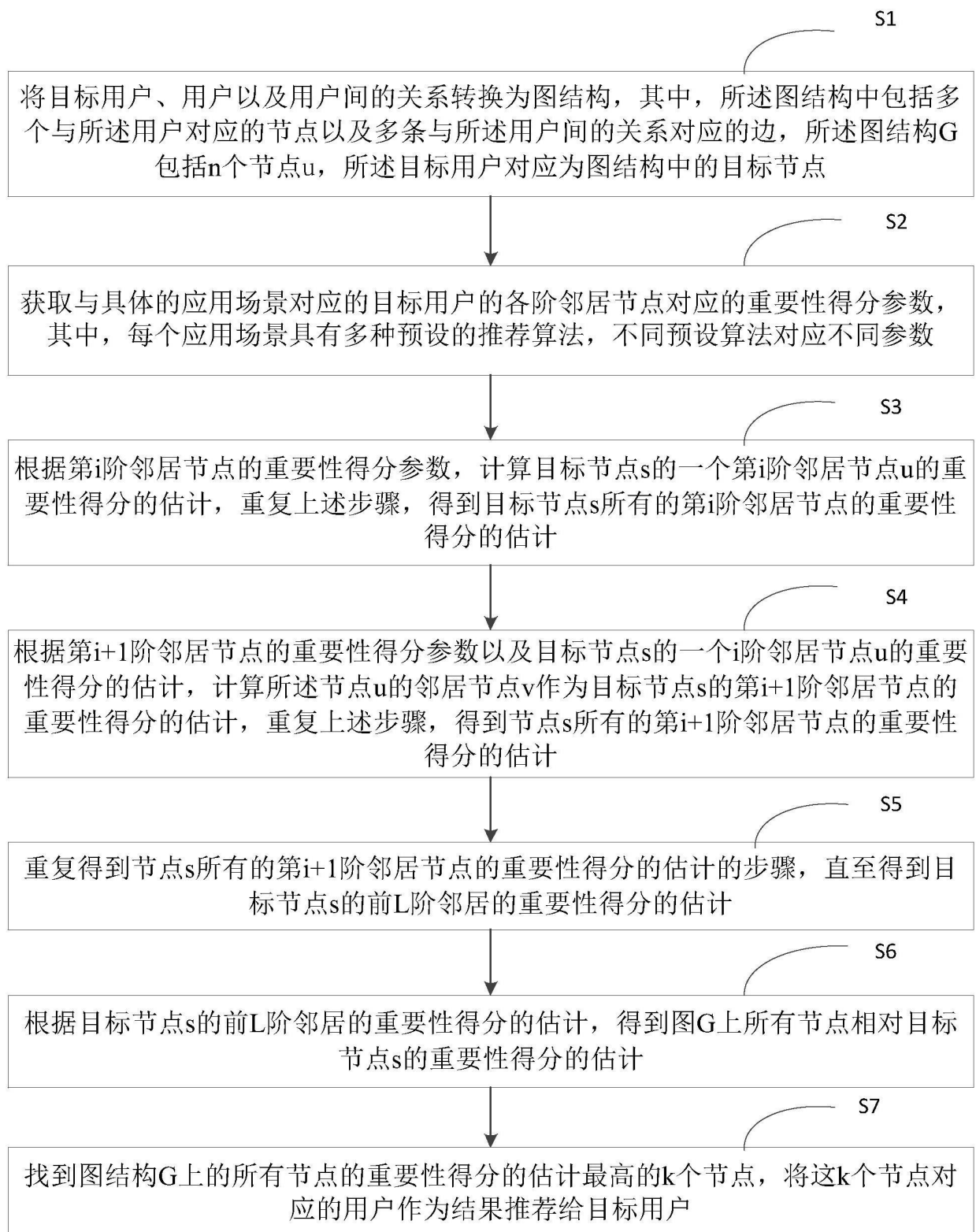


图1

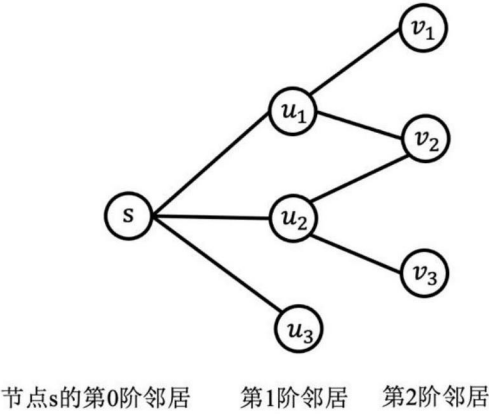


图2