



(12)发明专利

(10)授权公告号 CN 103778151 B

(45)授权公告日 2017.06.09

(21)申请号 201210407962.X

(22)申请日 2012.10.23

(65)同一申请的已公布的文献号

申请公布号 CN 103778151 A

(43)申请公布日 2014.05.07

(73)专利权人 阿里巴巴集团控股有限公司

地址 英属开曼群岛大开曼资本大厦一座四
层847号邮箱

(72)发明人 陈智强 邵纪东 林古立 刘水

顾海杰

(74)专利代理机构 北京润泽恒知识产权代理有

限公司 11319

代理人 苏培华

(51)Int.Cl.

G06F 17/30(2006.01)

(56)对比文件

CN 102446177 A,2012.05.09,
US 2006149674 A1,2006.07.06,
CN 100481076 C,2009.04.22,
CN 100465959 C,2009.03.04,
CN 100465959 C,2009.03.04,
CN 100465959 C,2009.03.04,
CN 100465959 C,2009.03.04,

审查员 李小敏

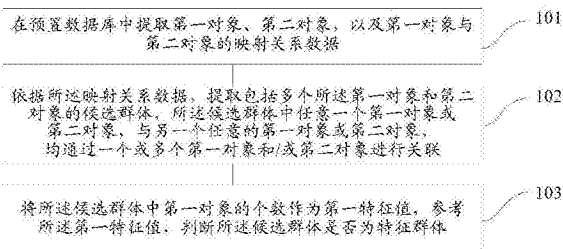
权利要求书2页 说明书10页 附图3页

(54)发明名称

一种识别特征群体的方法及装置和搜索方
法及装置

(57)摘要

本申请提供了一种识别特征群体的方法及装置和搜索方法及装置。搜索识别特征群体的方法包括：在预置数据库中提取第一对象、第二对象，以及第一对象与第二对象的映射关系数据；依据所述映射关系数据，提取包括多个所述第一对象和第二对象的候选群体，所述候选群体中任意一个第一对象或第二对象，与另一个任意的第一对象或第二对象，均通过一个或多个第一对象和/或第二对象进行关联；将所述候选群体中第一对象的个数作为第一特征值，参考所述第一特征值判断所述候选群体是否为特征群体。本申请可以识别出恶意注册的网站账号群体，避免买家参与到虚假交易，进而提高网络购物的交易安全。



1. 一种识别特征群体的方法,其特征在于,包括:

在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;

依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;

将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值判断所述候选群体是否为特征群体。

2. 如权利要求1所述的方法,其特征在于,还包括:

针对各第一对象,将关联同一第二对象的第一对象确定为直接关联对象;

依据各第一对象的直接关联对象的个数,生成候选群体对应的第二特征值。

3. 如权利要求2所述的方法,其特征在于,还包括:

在参考所述第一特征值判断所述候选群体是否为特征群体的同时,参考所述第二特征值。

4. 如权利要求1所述的方法,其特征在于,所述第一对象为用户账号,所述第二对象为用户标识;所述映射关系数据包括,各用户账号与其所关联的所有用户标识的映射关系列表,以及各用户标识与其所关联的所有用户账号的映射关系列表。

5. 如权利要求1所述的方法,其特征在于,所述第一对象和第二对象具备属性值,所述依据映射关系数据,提取包括多个第一对象和第二对象的候选群体的步骤包括:

针对各个第一对象或第二对象,依据所关联的第一对象或第二对象对应的属性值,更新对应的属性值;

提取属性值相同的第一对象和第二对象,组成候选群体。

6. 如权利要求5所述的方法,其特征在于,所述针对各个第一对象或第二对象,依据所关联的第一对象或第二对象对应的属性值,更新对应的属性值的步骤包括:

步骤1、针对各个第一对象或第二对象,依据所述映射关系数据,确定所关联的第一对象或第二对象;

步骤2、提取所述关联的第一对象或第二对象所对应的属性值;

步骤3、将提取的属性值中最小的值作为该第一对象或第二对象对应的修正值;

步骤4、判断各第一对象或第二对象的属性值,与对应的修正值是否相同,若是,则结束操作,若否,则以修正值替换属性值,并返回步骤1。

7. 如权利要求6所述的方法,其特征在于,所述候选群体包括一个或多个,所述第一对象的属性值的初始值为账号ID的字符串,所述第二对象的属性值的初始值与用户标识相同或是用户标识的字符串。

8. 如权利要求2所述的方法,其特征在于,所述针对各第一对象,确定其直接关联对象的步骤包括:

针对各第一对象,确定关联的第二对象,并确定各关联的第二对象所关联的第一对象;

将所有关联的第二对象所关联的第一对象进行合并,去重,将去重后的第一对象作为直接关联对象。

9. 如权利要求2所述的方法,其特征在于,所述第一特征值记为 n ,所述第二特征值为各第一对象的直接关联对象的个数加和后,再除以 $n(n-1)$ 的结果。

10. 如权利要求1所述的方法,其特征在于,所述参考第一特征值,判断候选群体是否为特征群体的步骤包括:

若所述第一特征值大于第一阈值,则所述候选群体为特征群体。

11. 如权利要求3所述的方法,其特征在于,参考第一特征值和第二特征值,判断候选群体是否为特征群体的步骤包括:

若所述第一特征值大于第一阈值,且所述第二特征值大于第二阈值,则所述候选群体为特征群体。

12. 一种搜索方法,其特征在于,包括:

接收特征群体的搜索请求;

通过权利要求1-11任一项所述的方法识别特征群体,并返回特征群体的属性信息。

13. 一种识别特征群体的装置,其特征在于,包括:

映射关系数据提取模块,用于在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;

候选群体提取模块,用于依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;

特征群体判断模块,用于将所述候选群体中第一对象的个数作为第一特征值,并按照所述第一特征值判断所述候选群体是否为特征群体。

14. 一种搜索装置,其特征在于,包括:

搜索请求接收模块,用于接收特征群体的搜索请求;

信息返回模块,用于通过权利要求12所述的装置识别特征群体,并返回特征群体的属性信息。

一种识别特征群体的方法及装置和搜索方法及装置

技术领域

[0001] 本申请涉及交易信息处理领域,特别是涉及一种识别特征群体的方法和装置,以及一种搜索方法和装置。

背景技术

[0002] 互联网上大部分网站都具有账号注册功能,一些大型网站甚至拥有上亿的用户。

[0003] 目前,一些购物网站的卖家为了宣传自己的商品,会人工地或者程序自动地在网站上注册大量无用的账号。网站上的大量恶意注册账号常常会给网站的正常运营造成巨大的破坏,例如,利用恶意注册账号购买某个商品,会产生大量的虚假销售记录,消费者会被虚假的销售记录欺骗而购买该商品。

[0004] 目前,针对网站恶意注册的行为,可以通过注册前预防和注册后处理两个环节来处理,注册前预防一般是通过在网站用户注册阶段增加验证码输入的步骤,这种方法能在一定程度上防止基于程序自动注册的恶意注册行为;注册后处理是针对预防无效情况下已产生的恶意注册账号,针对某一种用户标识,如Cookie,若在同一个Cookie下登录过的不同账号数大于预设值,则判断这些账号属于恶意注册账号。

[0005] 以上现有技术中存在的问题是,针对注册前预防的方式,目前已有具备验证码识别功能的自动注册程序可以破解验证码,并且该方式无法预防通过大量人工注册的恶意注册行为;注册后处理的方式只考虑了账号间的直接关联关系,只能发现少量的恶意注册账号,由于大部分恶意注册账号未必会共用一个用户标识。比如,以Cookie为例,同一个网站用户在不同时间不同机器上使用账号,其关联的Cookie很可能是不相同的,这种情况下简单的查看一个Cookie关联的账号数并不能有效的发现异常信息。

[0006] 因此,目前需要本领域技术人员解决的一个技术问题就是,提供一种识别特征群体的机制,以识别出恶意注册的网站账号群体,避免买家参与到虚假交易,进而提高网络购物的交易安全。

发明内容

[0007] 本申请所要解决的技术问题是提供一种识别特征群体的方法,以识别出恶意注册的网站账号群体,避免买家参与到虚假交易,进而提高网络购物的交易安全。

[0008] 本申请还提供了一种识别特征群体的装置,用以保证上述方法在实际中的应用及实现。

[0009] 为了解决上述问题,本申请公开了一种识别特征群体的方法,包括:

[0010] 在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;

[0011] 依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;

[0012] 将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值判断所述候选群体是否为特征群体。

[0013] 优选的,所述方法还包括:

[0014] 针对各第一对象,将关联同一第二对象的第一对象确定为直接关联对象;

[0015] 依据各第一对象的直接关联对象的个数,生成候选群体对应的第二特征值。

[0016] 优选的,所述方法还包括:

[0017] 在参考所述第一特征值判断所述候选群体是否为特征群体的同时,参考所述第二特征值。

[0018] 优选的,所述第一对象为用户账号,所述第二对象为用户标识;所述映射关系数据包括,各用户账号与其所关联的所有用户标识的映射关系列表,以及各用户标识与其所关联的所有用户账号的映射关系列表。

[0019] 优选的,所述第一对象和第二对象具备属性值,所述依据映射关系数据,提取包括多个第一对象和第二对象的候选群体的步骤包括:

[0020] 针对各个第一对象或第二对象,依据所关联的第一对象或第二对象对应的属性值,更新对应的属性值;

[0021] 提取属性值相同的第一对象和第二对象,组成候选群体。

[0022] 优选的,所述针对各个第一对象或第二对象,依据所关联的第一对象或第二对象对应的属性值,更新对应的属性值的步骤包括:

[0023] 步骤1、针对各个第一对象或第二对象,依据所述映射关系数据,确定所关联的第一对象或第二对象;

[0024] 步骤2、提取所述关联的第一对象或第二对象所对应的属性值;

[0025] 步骤3、将提取的属性值中最小的值作为该第一对象或第二对象对应的修正值;

[0026] 步骤4、判断各第一对象或第二对象的属性值,与对应的修正值是否相同,若是,则结束操作,若否,则以修正值替换属性值,并返回步骤1。

[0027] 优选的,所述候选群体包括一个或多个,所述第一对象的属性值的初始值为账号ID的字符串,所述第二对象的属性值的初始值与用户标识相同或是用户标识的字符串。

[0028] 优选的,所述针对各第一对象,确定其直接关联对象的步骤包括:

[0029] 针对各第一对象,确定关联的第二对象,并确定各关联的第二对象所关联的第一对象;

[0030] 将所有关联的第二对象所关联的第一对象进行合并,去重,将去重后的第一对象作为直接关联对象。

[0031] 优选的,所述第一特征值记为 n ,所述第二特征值为各第一对象的直接关联对象的个数加和后,再除以 $n(n-1)$ 的结果。

[0032] 优选的,所述参考第一特征值,判断候选群体是否为特征群体的步骤包括:

[0033] 若所述第一特征值大于第一阈值,则所述候选群体为特征群体。

[0034] 优选的,参考第一特征值和第二特征值,判断候选群体是否为特征群体的步骤包括:

[0035] 若所述第一特征值大于第一阈值,且所述第二特征值大于第二阈值,则所述候选群体为特征群体。

- [0036] 本申请还提供了一种搜索方法,包括:
- [0037] 接收特征群体的搜索请求;
- [0038] 通过上述识别特征群体的方法识别特征群体,并返回特征群体的属性信息。
- [0039] 本申请还提供了一种识别特征群体的装置,包括:
- [0040] 映射关系数据提取模块,用于在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;
- [0041] 候选群体提取模块,用于依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;
- [0042] 特征群体判断模块,用于将所述候选群体中第一对象的个数作为第一特征值,并按照所述第一特征值判断所述候选群体是否为特征群体。
- [0043] 本申请还提供了一种搜索装置,包括:
- [0044] 搜索请求接收模块,用于接收特征群体的搜索请求;
- [0045] 信息返回模块,用于通过权利要求上述识别特征群体的装置识别特征群体,并返回特征群体的属性信息。
- [0046] 与现有技术相比,本申请具有以下优点:
- [0047] 本申请依据用户账户和用户标识的对应关系,利用图论的方法,先找出有直接或间接关联的用户账户和用户标识作为一个候选群体,将具有间接关联关系的账户归于同一个账户群体,然后依据各用户账户直接关联的账户数计算候选群体的群体密度,依据账户数,或是结合候选群体密度和账户数来判断是否是恶意注册账户群体。相比于现有技术,本申请依据账户之间的间接关联关系可以找出潜在的恶意注册账号,发现恶意注册账号群体,从而可以避免买家参与到虚假交易,提高网络购物的交易安全。

附图说明

- [0048] 图1是本申请的一种识别特征群体的方法实施例1的流程图;
- [0049] 图2是本申请的一种识别特征群体的方法实施例2的流程图;
- [0050] 图3是本申请的一种的搜索方法实施例的流程图;
- [0051] 图4是本申请的一种识别特征群体的装置实施例1的结构框图;
- [0052] 图5是本申请的一种识别特征群体的装置实施例2的结构框图;
- [0053] 图6是本申请的一种搜索装置实施例的结构框图;
- [0054] 图7是本申请的邻接表对应的示意图
- [0055] 图8为本申请的候选群体的示意图。

具体实施方式

[0056] 为使本申请的上述目的、特征和优点能够更加明显易懂,下面结合附图和具体实施方式对本申请作进一步详细的说明。

[0057] 参考图1,示出了本申请的一种识别特征群体的方法实施例1的流程图,具体可以包括以下步骤:

[0058] 步骤101、在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的

映射关系数据。

[0059] 在具体的实现中,所述第一对象可以为用户账号,一个用户可以注册一个或多个账号;所述第二对象可以为用户标识,用户标识是账号在浏览网站的过程中留下的,能够确定用户身份的信息,例如,登录时产生的cookie、账户绑定的邮箱、手机号、或者是多种信息通过一定算法计算出来的一个值,一个账号可以对应多个用户标识,一个用户标识也可以用于多个账号。

[0060] 所述映射关系数据可以包括,各用户账号与其所关联的所有用户标识的映射关系列表,以及各用户标识与其所关联的所有用户账号的映射关系列表。用户账号与用户标识的关联关系可以存放在预置的数据库中,数据库中具体包括了,各用户账号所关联的所有用户标识的列表,以及,每个用户标识所关联的所有用户账号的列表。

[0061] 具体的,可以在预置数据库中提取一段时间内(例如最近3个月)的映射关系数据,这些映射关系数据可以视为图论中用于表示一个图的邻接表。例如, A_1 、 A_2 、 A_3 分别表示三个用户账号, B_1 、 B_2 、 B_3 分别表示三个用户标识,假设数据库中保存的映射关系可以表示为 (A_1, B_1) 、 (A_1, B_2) 、 (A_2, B_2) 、 (A_2, B_3) 、 (A_3, B_3) ,那么,该映射关系构成的邻接表对应的示意图如图7所示。

[0062] 该邻接表所表示的图有以下特点:图中的顶点为所有的用户账号或者用户标识,每一条边只连接一个用户账号或一个用户标识,表示某账号与某用户标识存在关联关系。

[0063] 通过解析网站的行为日志可以获取映射关系数据,具体而言,记录登录用户在网站上的行为日志,行为日志中包含了用户账号以及对应的用户标识,解析行为日志即可获取用户账号与用户标识的对应关系,例如账号登录网站时,可以得到账号的cookie之间的关联,用户在填写自己的邮箱时,可以得到账号与邮箱之间的关联。解析出映射关系后,可以存放在预置的数据库中。

[0064] 步骤102、依据所述映射关系数据,提取包括所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联。

[0065] 依据映射关系数据可以初步提取出一个或多个恶意注册账号群体的候选群体。候选群体中,任意两个对象之间均是连通的。更具体而言,任意两个第一对象之间可以通过一个或多个第一对象和/或第二对象关联;任意两个第二对象可以通过一个或多个第一对象和/或第二对象关联;任意一个第一对象和任意一个第二对象之间也可以通过一个或多个第一对象和/或第二对象关联。

[0066] 如图8所示,给出了本申请所述的候选群体的示意图,其中, A_1 、 A_2 、 A_3 、 A_4 为用户账号, B_1 、 B_2 、 B_3 为用户标识,连线表示具备关联关系。由图可知,账号 A_1 与标识 B_1 和 B_2 关联,标识 B_1 分别与账号 A_1 、 A_2 和 A_3 关联,标识 B_2 分别与账号 A_1 、 A_3 和 A_4 关联,账号 A_4 分别与标识 B_2 和 B_3 关联;并且,任意一个用户账号或用户标识,与另一个任意的账号或用户标识,可以通过一个或多个用户账号或用户标识进行关联,如账号 A_1 与 A_2 可以通过标识 B_2 关联,账号 A_1 与标识 B_3 可以通过标识账号 A_1 和标识 B_2 关联。

[0067] 在本申请的一种优选实施例中,所述第一对象和所述第二对象均可以具备属性值,优选的,所述候选群体可以包括一个或多个,所述第一对象的属性值的初始值可以为账号ID的字符串,例如,第一对象用户账号名为amy,对应的账号ID为12345,对应的属性值为

“12345”；所述第二对象的属性值的初始值可以与用户标识相同或是用户标识的字符串，例如，第二对象用户标识为手机号21231441或邮箱amy@123.com，相应的，第二对象的属性值的初始值可以是“21231441”或“amy@123.com”。

[0068] 具体的，可以通过以下步骤提取候选群体：

[0069] 子步骤S11、针对各个第一对象或第二对象，依据所关联的第一对象或第二对象对应的属性值，更新对应的属性值。

[0070] 在具体的实现中，可以通过以下步骤更新属性值：

[0071] 子步骤S11-1、针对各个第一对象或第二对象，依据所述映射关系数据，确定所关联的第一对象或第二对象；

[0072] 子步骤S11-2、提取所述关联的第一对象或第二对象所对应的属性值；

[0073] 子步骤S11-3、将提取的属性值中最小的值作为该第一对象或第二对象对应的修正值；

[0074] 子步骤S11-4、判断各第一对象或第二对象的属性值，与对应的修正值是否相同，若是，则结束操作，若否，则以修正值替换属性值，并返回子步骤S11-1。

[0075] 以某个第一对象 A_1 为例，对应的属性值为 a_1 依据映射关系数据，可以确定其对应的第二对象 $B_1, B_2, B_3 \cdots B_n$ ，收集这些第二对象的属性值 $b_1, b_2, b_3 \cdots b_n$ ，比较 $b_1, b_2, b_3 \cdots b_n$ 的大小，取其中最小的值，例如 b_2 ，作为该第一对象对应的修正值 $a_1' = b_2$ ；同样的方法，生成各个第二对象对应的修正值。然后对比各第一对象或第二对象的属性值与其对应的修正值是否相同，例如对于第一对象 A_1 ，即判断 a_1 是否等于 b_2 。若存在不相同的情况，则以修正值替换相应的属性值，并且返回到子步骤S11-1中，重复执行上述步骤，直至各第一对象或第二对象的属性值与其对应的修正值相同为止。

[0076] 具体而言，对各个属性值的比较可以通过比较两个字符串的字典序来实现。例如，对于两个字符串 S_1 和 S_2 ，将 S_1 和 S_2 利用任意一种编码方式（例如使用UTF8编码）转成两个字节码数组 C_1 和 C_2 。比较 C_1 和 C_2 的第一个元素；如果两者不相等，则第一个元素值小的数组的字典序靠前，否则比较两个数组的第二个元素；如果第二个元素两者不相等，则值小数组的字典序靠前，否则依次类推；如果遍历过程中某个数组已被遍历，那么该数组的字典序靠前。

[0077] 例如，当 $S_1 = "abc"$ ， $S_2 = "123"$ 时。将 S_1 和 S_2 按ASCII码转成对应的字节码数组 $C_1 = \{97, 98, 99\}$ ， $C_2 = \{49, 50, 51\}$ ，由于 S_2 的第一个元素小于 S_1 ，那么 S_2 的字典序靠前。字典序靠前的字符串为比较小的字符串，由此对多个属性值的字符串进行比较，可以找到最小的属性值。

[0078] 子步骤S12、提取属性值相同的第一对象和第二对象，组成候选群体。

[0079] 通过步骤S11，各第一对象或第二对象的属性值均与修正值相同之后，可以将属性值相同的第一对象和第二对象，组成候选群体，即在候选群体中，包括多个第一对象和第二对象，无论是第一对象的属性值还是第二对象的属性值均相同。

[0080] 对于映射关系数据表示的邻接表，在对应的图中，将每个用户账户和用户标识作为一个顶点，以上提取候选群体的步骤可以理解为图论中利用标记扩散法(label propagation)寻找连通子图的过程。

[0081] Step 1、针对每个顶点对应有唯一的标记值（上面步骤中的属性值），对于每个顶点，确定与其相连的多个顶点；

[0082] Step 2、对于每个顶点,获取与之相连的其他顶点的标记值;

[0083] Step 3、若某个顶点的所关联的其他顶点的标记值中的最小值和该顶点的当前标记值不同,则以最小标记值取代当前标记值,并返回到Step 1,若相同,则提取具备相同标记的顶点,作为一个连通子图。

[0084] 此处得到的连通子图即为候选群体,连通子图中的每两个点都可以通过其他的点连接起来,即所述候选群体中任意两个第一对象或第二对象(任意两个第一对象,任意两个第二对象,任意两个第一对象和第二对象)均通过一个或多个第一对象或第二对象进行关联。

[0085] 其中,具备相同的第二对象的两个第一对象称为直接关联账号,通过多个第一对象或第二对象进行关联的两个第一对象称为间接关联账号,例如,对于账号A和账号B,如果能找出一系列账号 $X_1, X_2, \dots, X_n$ ($n \geq 1$),使得A与 X_1 直接关联, X_i 与 X_{i+1} ($1 \leq i \leq n-1$)直接关联, X_n 与B直接关联,则A与B是间接关联账号。通过连通子图的方法找到的候选群体,不仅找出了具备直接关联关系的账号,还把具备间接关联关系的多个账号集合在一起。

[0086] 步骤103、将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值,判断所述候选群体是否为特征群体。

[0087] 在本申请的一种优选实施例中,所述步骤103可以包括:

[0088] 子步骤S21、若所述第一特征值大于第一阈值,则所述候选群体为特征群体。

[0089] 针对各候选群体,以第一特征值来判断是否为特征群体,优选的,当候选群体所包含的账号数大于第一阈值,则认为该候选群体为恶意注册账号的特征群体。

[0090] 其中,第一阈值的具体取值与网站性质、规模以及恶意注册行为的具体标准相关,在具体的实现中可以根据数据进行必要的调试。优选的,第一阈值的取值可以为几百到几千。

[0091] 相比于现有技术,本申请依据账户之间的直接关联关系和间接关联关系确定候选群体,从而可以找出更多潜在的恶意注册账号,依据候选群体中账号的个数进一步确定恶意注册账号群体,从而可以避免买家参与到虚假交易,提高网络购物的交易安全。

[0092] 参考图2,示出了本申请的一种识别特征群体的方法实施例2的流程图,具体可以包括以下步骤:

[0093] 步骤201、在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据。

[0094] 步骤202、依据所述映射关系数据,提取包括所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联。

[0095] 步骤203、针对各第一对象,将关联同一第二对象的第一对象确定为直接关联对象。

[0096] 在本申请的一种优选实施例中,所述步骤203可以包括:

[0097] 子步骤S31、针对各第一对象,确定关联的第二对象,并确定各关联的第二对象所关联的第一对象;

[0098] 子步骤S32、将所有关联的第二对象所关联的第一对象进行合并,去重,将去重后的第一对象作为直接关联对象。

[0099] 第一对象的直接关联对象是指具有相同的第二对象的另一个第一对象,即具有相同用户标识的两个账号。

[0100] 具体的,针对某个用户账号,先找出与其关联的用户标识,以及各用户标识关联的多个用户账号;然后将各用户标识关联的多个用户账号进行合并,去重,即可得到第一对象的直接关联对象。例如,账号A与标识B和标识C关联,标识B与账号D和账号E关联,标识C与账号D和账号F关联,将标识B和标识C所关联的账号进行合并得到4个账号,分别为账号D、账号E、账号D和账号F,进行去重得到账号D、账号E和账号F,即为A的三个直接关联对象。

[0101] 步骤204、依据各第一对象的直接关联对象的个数,生成候选群体对应的第二特征值。

[0102] 在图论中,各第一对象的直接关联对象的个数即为第一对象的度数。本申请中,第二特征值即为图论中的群体密度,针对某个候选群体,将候选群体中的用户账号的个数记为 n ,所述第二特征值可以为各第一对象的度数加和后,再除以 $n(n-1)$ 的结果,在具体的实现中,群体密度的取值范围在0~1之间。

[0103] 步骤205、将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值和第二特征值,判断所述候选群体是否为特征群体。

[0104] 所述步骤205可以包括:

[0105] 子步骤S41、若所述第一特征值大于第一阈值,且所述第二特征值大于第二阈值,则所述候选群体为特征群体。

[0106] 与上个实施例不同的是,本实施例针对各候选群体,以第一特征值结合第二特征值来判断是否为特征群体,可以使得恶意群体的识别更为准确。优选的,当候选群体所包含的账号数大于第一阈值,群体密度大于第二阈值,则认为该候选群体为恶意注册账号的特征群体。

[0107] 其中,第一阈值和第二阈值的具体取值与网站性质、规模以及恶意注册行为的具体标准相关,在具体的实现中可以根据数据进行必要的调试。优选的,第一阈值的取值可以为几百到几千,第二阈值的取值可以为0.7到0.9之间。

[0108] 参考图3,示出了本申请的一种搜索方法实施例的流程图,具体可以包括以下步骤:

[0109] 步骤301、接收特征群体的搜索请求。

[0110] 步骤302、识别特征群体,并返回特征群体的属性信息。

[0111] 所述特征群体通过实施例1或2所述的方法识别出。用户提交搜索请求,请求搜索恶意注册的账号群体,服务器接收到搜索请求,会按照实施例1或2所述的方法确定特征群体,特征群体中包括了用户账户和用户标识,依据用户账户还可以搜索到特征群体相关用户的其他属性信息,例如,注册时间,信誉值,交易历史数据等,然后将特征群体以及相关的属性信息返回给用户。

[0112] 综上所述,本申请依据用户账户和用户标识的对应关系,利用图论的方法,先找出有直接或间接关联的用户账户和用户标识作为一个候选群体,将具有间接关联关系的账户归于同一个账户群体,然后依据各用户账户直接关联的账户数计算候选群体的群体密度,依据账户数,或是结合候选群体密度和账户数来判断是否是恶意注册账户群体。相比于现有技术,本申请依据账户之间的间接关联关系可以找出潜在的恶意注册账号,发现恶意注

册账号群体,从而可以避免买家参与到虚假交易,提高网络购物的交易安全。

[0113] 对于方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员应该知悉,本申请并不受所描述的动作顺序的限制,因为依据本申请,某些步骤可以采用其他顺序或者同时进行。其次,本领域技术人员也应该知悉,说明书中所描述的实施例均属于优选实施例,所涉及的动作和模块并不一定是本申请所必须的。

[0114] 参考图4,示出了本申请的一种识别特征群体的装置实施例1的结构框图,具体可以包括以下模块:

[0115] 映射关系数据提取模块401,用于在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;

[0116] 候选群体提取模块402,用于依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;

[0117] 第一判断模块403,用于将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值,判断所述候选群体是否为特征群体。

[0118] 在本申请的一种优选实施例中,所述第一对象可以为用户账号,所述第二对象可以为用户标识;

[0119] 所述映射关系数据可以包括,各用户账号与其所关联的所有用户标识的映射关系列表,以及各用户标识与其所关联的所有用户账号的映射关系列表。

[0120] 在本申请的一种优选实施例中,所述第一对象和所述第二对象均可以具备属性值,所述候选群体提取模块可以包括:

[0121] 属性值更新子模块,用于针对各个第一对象或第二对象,依据所关联的第一对象或第二对象对应的属性值,更新对应的属性值;

[0122] 候选群体组成子模块,用于提取属性值相同的第一对象和第二对象,组成候选群体。

[0123] 在本申请的一种优选实施例中,所述属性值更新子模块可以包括:

[0124] 子单元1,用于针对各个第一对象或第二对象,依据所述映射关系数据,确定所关联的第一对象或第二对象;

[0125] 子单元2、用于提取所述关联的第一对象或第二对象所对应的属性值;

[0126] 子单元3、用于将提取的属性值中最小的值作为该第一对象或第二对象对应的修正值;

[0127] 子单元4,用于判断各第一对象或第二对象的属性值,与对应的修正值是否相同,若是,则结束操作,若否,则以修正值替换属性值,并返回子单元1。

[0128] 在本申请的一种优选实施例中,所述候选群体可以包括一个或多个,所述第一对象的属性值的初始值可以为账号ID的字符串,所述第二对象的属性值的初始值可以与用户标识相同或是用户标识的字符串。

[0129] 在本申请的一种优选实施例中,所述第一判断模块可以包括:

[0130] 第一特征值判断子模块,用于若所述第一特征值大于第一阈值,则所述候选群体为特征群体。

[0131] 参考图5,示出了本申请的一种识别特征群体的装置实施例2的结构框图,具体可

以包括以下模块：

[0132] 映射关系数据提取模块501,用于在预置数据库中提取第一对象、第二对象,以及第一对象与第二对象的映射关系数据;

[0133] 候选群体提取模块502,用于依据所述映射关系数据,提取包括多个所述第一对象和第二对象的候选群体,所述候选群体中任意一个第一对象或第二对象,与另一个任意的第一对象或第二对象,均通过一个或多个第一对象和/或第二对象进行关联;

[0134] 直接关联对象确定模块503,用于针对各第一对象,将关联同一第二对象的第一对象确定为直接关联对象;

[0135] 特征值生成模块504,用于依据各第一对象的直接关联对象的个数,生成候选群体对应的第二特征值;

[0136] 第二判断模块505,用于将所述候选群体中第一对象的个数作为第一特征值,参考所述第一特征值和第二特征值,判断所述候选群体是否为特征群体。

[0137] 在本申请的一种优选实施例中,所述直接关联对象确定模块可以包括:

[0138] 第一对象确定子模块,用于针对各第一对象,确定关联的第二对象,并确定各关联的第二对象所关联的第一对象;

[0139] 合并子模块,用于将所有关联的第二对象所关联的第一对象进行合并,去重,将去重后的第一对象作为直接关联对象。

[0140] 在本申请的一种优选实施例中,所述第一特征值可以记为 n ,所述第二特征值可以为各第一对象的直接关联对象的个数加和后,再除以 $n(n-1)$ 的结果。

[0141] 在本申请的一种优选实施例中,所述第二判断模块可以包括:

[0142] 第二特征值判断子模块,用于若所述第一特征值大于第一阈值,且所述第二特征值大于第二阈值,则所述候选群体为特征群体。

[0143] 参考图6,示出了本申请的一种搜索装置实施例的结构框图,具体可以包括以下模块:

[0144] 搜索请求接收模块601,用于接收特征群体的搜索请求;

[0145] 信息返回模块602,用于通过实施例4或5所述的装置识别特征群体,并返回特征群体的属性信息。

[0146] 由于所述装置实施例基本相应于前述图1-图3所示的方法实施例,故本实施例的描述中未详尽之处,可以参见前述实施例中的相关说明,在此就不赘述了。

[0147] 本申请可用于众多通用或专用的计算机系统环境或配置中。例如:个人计算机、服务器计算机、手持设备或便携式设备、平板型设备、多处理器系统、基于微处理器的系统、置顶盒、可编程的消费电子设备、网络PC、小型计算机、大型计算机、包括以上任何系统或设备的分布式计算环境等等。

[0148] 本申请可以在由计算机执行的计算机可执行指令的一般上下文中描述,例如程序模块。一般地,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本申请,在这些分布式计算环境中,由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中,程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[0149] 在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从

而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

[0150] 以上对本申请所提供的一种识别特征群体的方法和一种搜索方法,以及,一种识别特征群体的方法和一种搜索装置进行了详细介绍,本文中应用了具体个例对本申请的原理及实施方式进行了阐述,以上实施例的说明只是用于帮助理解本申请的方法及其核心思想;同时,对于本领域的一般技术人员,依据本申请的思想,在具体实施方式及应用范围上均会有改变之处,综上所述,本说明书内容不应理解为对本申请的限制。

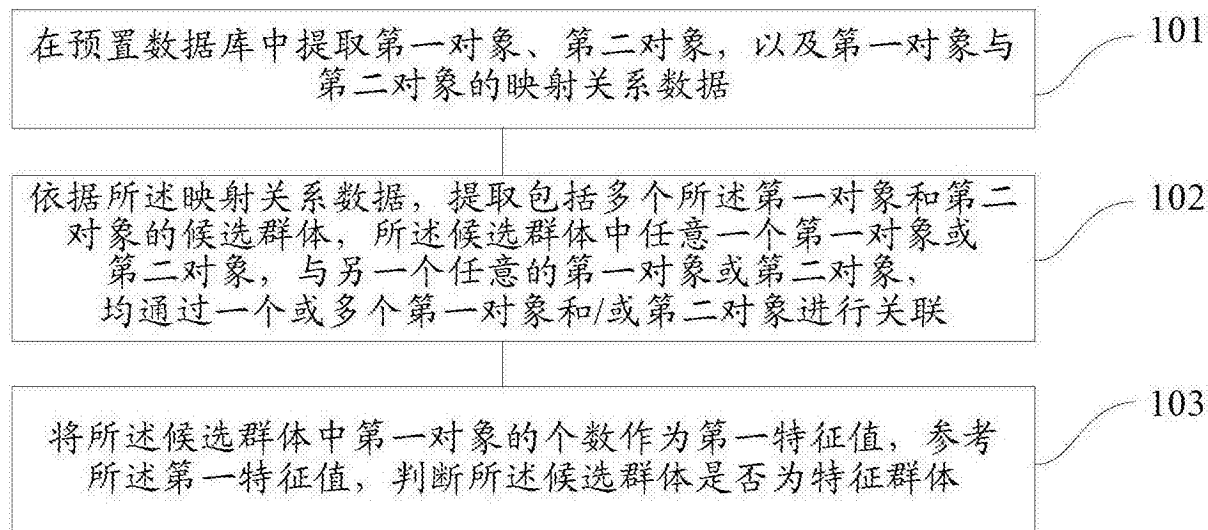


图1

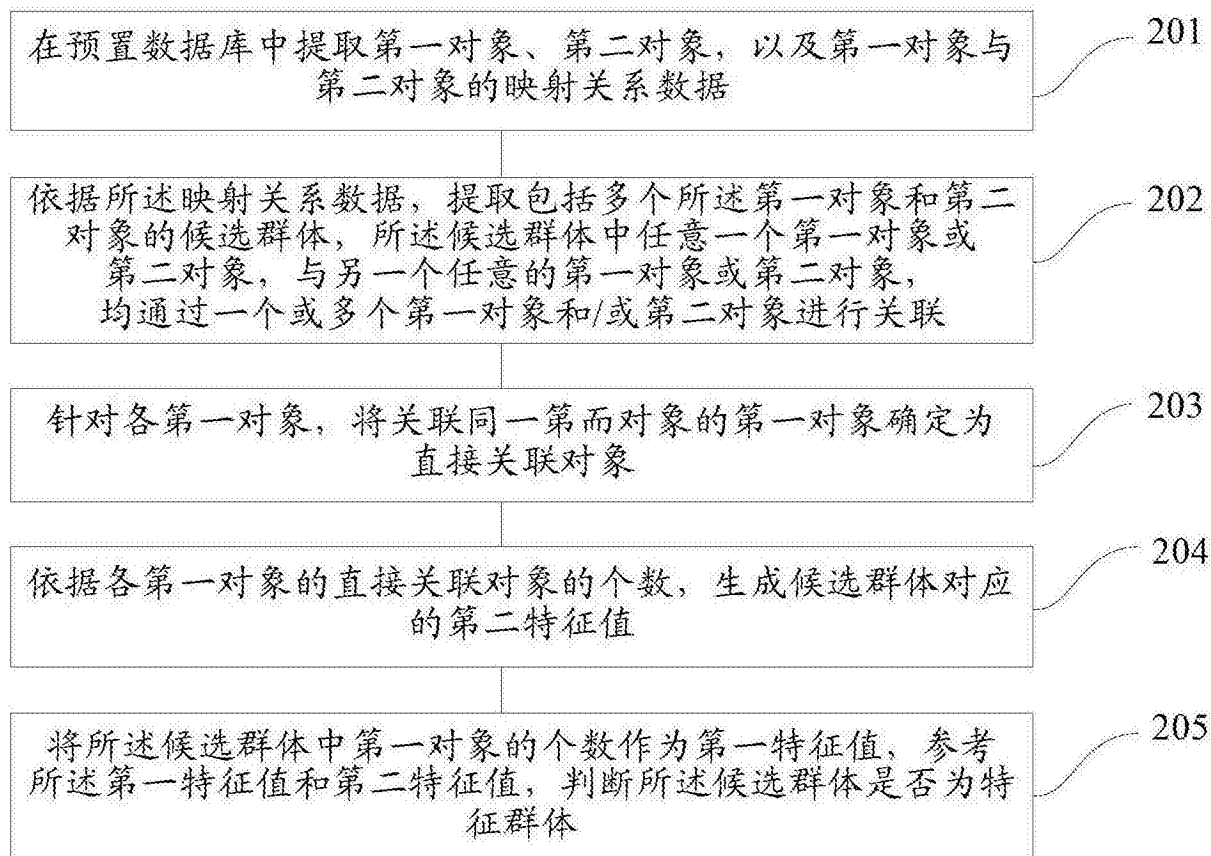


图2

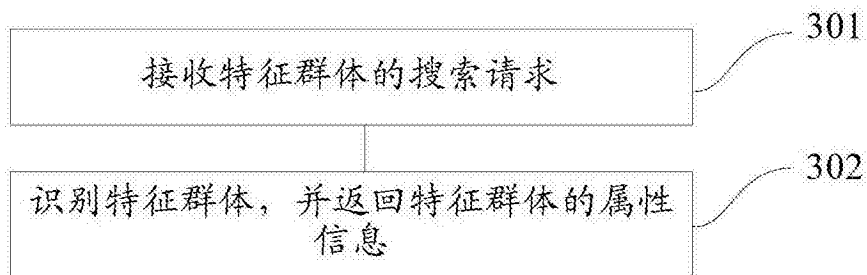


图3



图4

图5

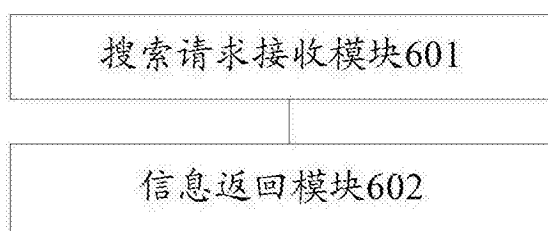


图6

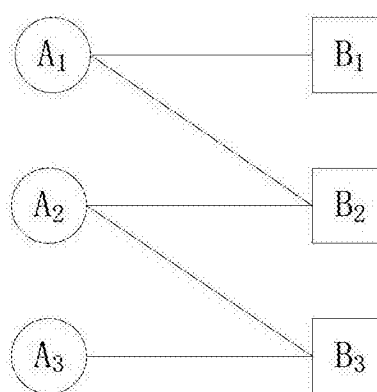


图7

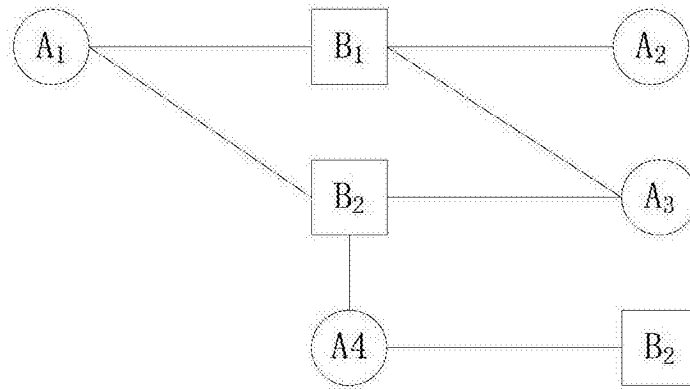


图8