(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国 际 局

(43) 国际公布日
2020 年 10 月 1 日 (01.10.2020)

WIPO | PCT

(10) 国际公布号
WO 2020/192184 A1

(51) 国际专利分类号:
G06Q 20/40 (2012.01)

(21) 国际申请号: PCT/CN2019/124807

(22) 国际申请日: 2019 年 12 月 12 日 (12.12.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910239821.3 2019年3月27日 (27.03.2019) CN

(71) 申请人: 北京三快在线科技有限公司 (**BEIJING SANKUAI ONLINE TECHNOLOGY CO., LTD**) [CN/CN]; 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。

(72) 发明人: 黄剑飞 (**HUANG, Jianfei**); 中国北京市海淀区北四环西路 9 号 2106-030, Beijing 100080 (CN)。 陈振 (**CHEN, Zhen**); 中国北京市海淀区北四环西路9号2106-030, Beijing 100080 (CN)。

(74) 代理人: 北京曼威知识产权代理有限公司 (**BEIJING MAVAIP INTELLECTUAL PROPERTY LAW FIRM**); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 411 室, Beijing 100085 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) **Title:** GANG FRAUD DETECTION BASED ON GRAPH MODEL

(54) 发明名称: 基于图模型检测团伙欺诈

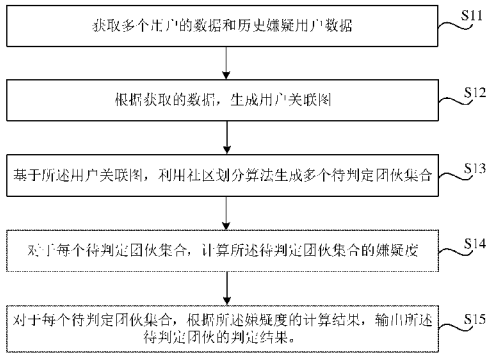

图 1

S11 Obtain the data of multiple users and historical suspected user data

S12 Generate a user association graph according to the obtained data

S13 Based on the user association graph, use a community division algorithm to generate multiple gang sets to be determined

S14 For each gang set to be determined, calculate the degree of suspicion of the gang set to be determined

S15 For each gang set to be determined, according to the calculation result of the degree of suspicion, output the determination result of a gang to be determined

(57) **Abstract:** A method and apparatus for detecting gang fraud based on a graph model, and a storage medium. The method for detecting gang fraud based on the graph model comprises: obtaining the data of multiple users and historical suspected user data (S11); generating a user association graph according to the obtained data (S12), wherein the nodes of the user association graph are user association subgraphs that are generated according to data characteristics, and the edge weight of the user association graph comprises the similarity of the nodes; based on the user association graph, using a community division algorithm to generate multiple gang sets to be determined (S13); for each gang set to be determined, calculating the degree of suspicion of the gang set to be determined (S14); and for each gang set to be determined, according to the calculation result of the degree of suspicion, outputting the determination result of a gang to be determined (S15).

(57) 摘要: 一种基于图模型检测团伙欺诈的方法和装置、存储介质。所述基于图模型检测团伙欺诈的方法包括: 获取多个用户的数据和历史嫌疑用户数据 (S11); 根据获取的数据, 生成用户关联图 (S12), 其中, 所述用户关联图的节点为根据数据特征生成的用户关联子图, 所述用户关联图的边权重包括节点的相似度; 基于所述用户关联图, 利用社区划分算法生成多个待判定团伙集合 (S13); 对于每个待判定团伙集合, 计算所述待判定团伙集合的嫌疑度 (S14); 对于每个待判定团伙集合, 根据所述嫌疑度的计算结果, 输出所述待判定团伙的判定结果 (S15)。

WO 2020/192184 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) **指定国**(除另有指明，要求每一种可提供的地区保护)：ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

基于图模型检测团伙欺诈

技术领域

[01]本公开涉及网络技术领域，具体地，涉及一种基于图模型检测团伙欺诈的方法和装置、存储介质。

背景技术

[02]金融领域对交易风险控制的要求较高，以保证资金交易的安全性。在实际应用中，可能会存在一些欺诈行为。比如，欺诈者诱骗很多的普通消费者向其转账，但是却不向这些消费者返还相应的回报，以此进行牟利。为了识别上述的欺诈行为，可以将高风险的欺诈者识别出来，以采取措施尽量避免消费者的资金损失。在一个例子中，可以利用交易模型来识别欺诈者，比如，将某个支付账户定性为欺诈者账户，将欺诈者账户进行的资金交易定性为风险交易。

发明内容

[03]本公开提供一种基于图模型检测团伙欺诈的方法和装置、存储介质，以解决相关技术中难以识别团伙欺诈的技术问题。

[04]为实现上述目的，本公开实施例的第一方面，提供一种基于图模型检测团伙欺诈的方法，所述方法包括:

[05]获取多个用户的数据和历史嫌疑用户数据；根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据数据特征生成的用户关联子图，所述用户关联图的边权重包括所述节点的相似度；基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合；对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度；对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

[06]可选地，生成所述用户关联图，包括：选取所述多个用户的所述数据和所述历史嫌疑用户数据中的特征组合和组数；基于所述特征组合和所述组数，利用特征一致性相等或模糊性相等方式生成用户关联子图；以所述用户关联子图为节点拼接生成用户无权重关联图；以所述用户无权重关联图中节点的相似度作为边权重，生成用户相似权重关联图作为所述用户关联图。

[07]可选地，利用所述社区划分算法生成所述多个待判定团伙集合，包括：基于所述用户关联图，利用所述社区划分算法生成 n 个团伙集合，n 为正整数；对于每个所述团伙集合，根据所述团伙集合的用户数量的大小进行调整，以得到多个新的团伙集合；将所述多个新的团伙集合确定为所述多个待判定团伙集合。

[08]可选地，根据所述团伙集合的所述用户数量的大小进行调整，包括：对用户数量大于极大阈值的团伙集合，调用所述社区划分算法进行划分，以使所述新的团伙集合中的用户数量小于或等于所述极大阈值；若用户数量小于极小阈值的团伙集合的数量大于预设阈值，调用层次聚类算法对所述用户数量小于极小阈值的所述团伙集合进行凝聚。

[09]可选地，所述社区划分算法包括图标签传播算法或 GN 算法；所述层次聚类算法包括凝聚算法或分裂算法。

[10]可选地，计算所述待判定团伙集合的所述嫌疑度，包括：从所述数据特征中选取目标数据特征；根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的所述嫌疑度。

[11]可选地，计算所述待判定团伙集合的所述嫌疑度，包括：抽取每个所述待判定团伙集合的团伙特征；将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的所述嫌疑度。

[12]可选地，计算所述待判定团伙集合的所述嫌疑度得分，包括：从所述数据特征中选取目标数据特征；根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的第一嫌疑度得分；抽取每个所述待判定团伙集合的团伙特征；将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的第二嫌疑度得分；根据所述第一嫌疑度得分以及所述第二嫌疑度得分，计算所述待判定团伙集合的综合嫌疑度得分。

[13]本公开实施例的第二方面，提供一种基于图模型检测团伙欺诈的装置，所述装置包括：

[14]获取模块，用于获取多个用户的数据和历史嫌疑用户数据；第一生成模块，用于根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据所述数据的特征生成的用户关联子图，所述用户关联图的边权重包括所述节点的相似度；第二生成模块，用于基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合；计算模块，用于对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度；输出模块，用于对于

每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

[15]可选地，所述第一生成模块包括：第一选取子模块，用于选取所述多个用户的数据和所述历史嫌疑用户数据中的特征组合和组数；第一生成子模块，用于基于所述特征组合和所述组数，利用特征一致性相等或模糊性相等方式生成用户关联子图，并以所述用户关联子图为节点拼接生成用户无权重关联图；第二生成子模块，用于以所述用户无权重关联图中节点的相似度作为边权重，生成用户相似权重关联图作为所述用户关联图。

[16]可选地，所述第二生成模块包括：第三生成子模块，用于基于所述用户关联图，利用所述社区划分算法生成 n 个团伙集合，n 为正整数；调整子模块，用于对于每个所述团伙集合，根据所述团伙集合的用户数量的大小进行调整，以得到多个新的团伙集合；第三确认子模块，用于将所述多个新的团伙集合确定为所述多个待判定团伙集合。

[17]可选地，所述调整子模块还包括：划分单元，用于对用户数量大于极大阈值的团伙集合，调用所述社区划分算法进行划分，以使所述新的团伙集合中的用户数量小于或等于所述极大阈值；凝聚模块，用于若用户数量小于极小阈值的团伙集合的数量大于预设阈值，调用层次聚类算法对所述用户数量小于所述极小阈值的所述团伙集合进行凝聚。

[18]可选地，所述社区划分算法包括图标签传播算法或 GN 算法；所述层次聚类算法包括凝聚算法或分裂算法。

[19]可选地，所述计算模块包括：第二选取子模块，用于从所述数据特征中选取目标数据特征；第一计算子模块，用于根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的嫌疑度。

[20]可选地，所述计算模块包括：第一抽取子模块，用于抽取每个所述待判定团伙集合的团伙特征；第一输入子模块，用于将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的嫌疑度。

[21]可选地，所述计算模块包括：第三选取子模块，用于从所述数据特征中选取目标数据特征；第二计算子模块，用于根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的第一嫌疑度得分；第二抽取子模块，用于抽取每个所述待判定团伙集合的团伙特征；第二输入子模块，用于将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的第二嫌疑度得分；第三计算子模块，用于根据所述第一嫌疑度得分以及所述第二嫌疑度得分，计算所述待判定团伙集合的综合嫌疑度得分。

[22]本公开实施例的第三方面，提供一种非易失性计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时，促使所述处理器实现上述第一方面中任一项所述方法的步骤。

[23]本公开实施例的第四方面，提供一种基于图模型检测团伙欺诈的装置，包括：存储器，其上存储有计算机程序；以及处理器，用于执行所述存储器中的所述计算机程序，以实现上述第一方面中任一项所述方法的步骤。

[24]采用上述技术方案，至少能够达到如下技术效果：本公开根据获取的用户数据，生成用户关联图，并利用社区划分算法生成待判定团伙集合，通过计算待判定团伙集合的嫌疑度，即可以分辨出该待判定团伙集合是否属于欺诈团伙，提高了识别团伙欺诈的准确性。另外，本公开还使用社区划分算法和层次聚类算法，解决了团伙划分结果中团伙规模过大、较小的团伙规模数量很多的问题。并且，本公开通过相似索引的手段来提升图模型数据处理能力，同时利用子图组装、相似边权重可配置地方式生成用户相似权重关联图，这种方法更加灵活可并行，可以进一步提升欺诈场景下的大规模数据处理能力。

[25]本公开的其他特征和优点将在随后的具体实施方式部分予以详细说明。

附图说明

[26]附图是用来提供对本公开的进一步理解，并且构成说明书的一部分，与下面的具体实施方式一起用于解释本公开，但并不构成对本公开的限制。

[27]图 1 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的方法流程图。

[28]图 2 是根据一示例性实施例示出的图 1 中步骤 S12 的具体流程图。

[29]图 3 是根据一示例性实施例示出的图 1 中步骤 S13 的具体流程图。

[30]图 4 是根据一示例性实施例示出的图 1 中步骤 S14 的具体流程图。

[31]图 5 是根据另一示例性实施例示出的图 1 中步骤 S14 的具体流程图。

[32]图 6 是根据再一示例性实施例示出的图 1 中步骤 S14 的具体流程图。

[33]图 7 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置框图。

[34]图 8 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的第一生成模块框图。

[35]图 9 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的第二生

成模块框图。

[36]图 10 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的调整子模块框图。

[37]图 11 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的计算模块框图。

[38]图 12 是本公开另一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的计算模块框图。

[39]图 13 是本公开再一示例性实施例示出的一种基于图模型检测团伙欺诈的装置的计算模块框图。

[40]图 14 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的硬件装置框图。

具体实施方式

[41]以下结合附图对本公开的具体实施方式进行详细说明。应当理解的是，此处所描述的具体实施方式仅用于说明和解释本公开，并不用于限制本公开。

[42]为了应对无处不在的攻击，欺诈检测在当下显得至关重要。针对金融欺诈检测可以采用如下几种检测方法。

[43]在一个例子中，可以使用基于黑白名单、信誉库查找的检测方法。该方法需要不定期维护添加新的黑白名单或信誉库内容，这种维护方法成本比较高，如第三方有偿数据购买，且方法响应和覆盖性有限。

[44]在另一个例子中，可以使用基于规则引擎的检测方法。线上金融欺诈手段多变，当欺诈者改变欺诈手段后，基于规则引擎的检测方法往往就会失效，需要投入大量运营和财力资源去更新规则引擎。

[45]在又一个例子中，可以使用基于有监督机器学习的检测方法。有监督机器学习是欺诈检测中应用最广泛的学习方法。机器学习模型通过会运用如决策树、随机森林、支持向量机 SVM（Support Vector Machine）和朴素贝叶斯算法等，进行数百个变量（也即高维空间）的复杂计算，准确锁定欺诈行为。但有监督机器学习方法依赖于标注数据，标注数据在金融欺诈场景获取难度比较大、正负样本失衡，其中，正样本只有当欺诈发

生后打标才有，在金融欺诈场景欺诈手段多变且样本较少会导致打标注较难。若缺少足够欺诈标注数据，有监督机器学习的能力有限。

[46]在再一个例子中，可以使用基于无监督学习的检测方法。无监督学习是目前欺诈检测应用探索的一个分支，主要是基于聚类和图方法进行研究，当前无监督技术成熟度较低，难度比较大，没有现成的解决方案可以有效的将无监督机器学习用于欺诈检测。主要难度在于如何解决大规模数据处理、嫌疑判定量化等。

[47]图 1 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的方法流程图，以提高识别团伙欺诈的识别水平。如图 1 所示，该基于图模型检测团伙欺诈包括以下步骤。

[48]S11，获取多个用户的数据和历史嫌疑用户数据。

[49]S12，根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据数据特征生成的用户关联子图，所述用户关联图的边权重包括节点的相似度。

[50]S13，基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合。

[51]S14，对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度。

[52]S15，对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

[53]在步骤 S11 中，所述用户数据可以是用户申请的各种客户端账号的数据，比如申请美团账号时的用户数据、申请支付宝账号时的用户数据、申请微信账号时的用户数据等，所述账号可以关联用户申请的银行卡号，比如储蓄卡或者信用卡。所述用户数据也可以是利用支付平台进行支付的用户对应的数据，比如利用美团进行支付的用户数据、利用支付宝进行支付的用户数据、利用微信进行支付的用户数据等等。更进一步的，所述用户数据还包括用户的基础数据。用户的基础数据包括申请人填写申请书资料、人行报告查询信息、申请人授权的移动端行为数据、电商数据、以及社交数据。所述历史嫌疑用户数据可以包括黑白名单信息，黑白名单可以是网络中的任何实体类型，比如账户、地址、电话号码等。黑名单包括行内积累的欺诈、严重逾期、或者交换黑名单，白名单包括 vip 客户或者人工标记无风险的电话、地址等。

[54]在获取多个用户的数据和历史嫌疑用户数据后，执行步骤 S12，根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据数据特征生成的用户关联子图，所述用户关联图的边权重包括节点的相似度。

[55]参考图 2，根据获取的数据，生成用户关联图，可以包括以下步骤。

[56]S121，选取所述多个用户的数据和所述历史嫌疑用户数据中的特征组合和组数。

[57]S122，基于特征组合和组数，利用特征一致性相等或模糊性相等方式生成用户关联子图，并以所述用户关联子图为节点拼接生成用户无权重关联图。

[58]S123，以所述用户无权重关联图中节点的相似度作为边权重，生成用户相似权重关联图作为所述用户关联图。

[59]在步骤 S121 中，所述数据中的特征可以是设备 ID、IP 地址、IMSI（国际移动用户识别码）、IMEI（国际移动设备识别码）、地理信息、登录时间等特征。所述特征组合是从所述数据中的特征中选出至少一个特征作为一组，所述组数至少也是一组。

[60]选取特征组合和组数后，利用特征一致性相等或模糊性相等方式，将不同的特征组合关联起来形成用户关联子图。比如，不同账号登录的设备 ID 相同，则可以利用特征一致性相等方式，将该两个账号关联起来；不同账号登录的 IP 地址部分相同，即同一个局域网下登录过不同账号，则可以利用特征模糊性相等方式将该两个账号关联起来。生成用户关联子图后，以所述用户关联子图为节点拼接生成用户无权重关联图。接着，以所述用户无权重关联图中各个节点的相似度作为边权重，生成用户相似权重关联图，相似度可以使用相似度衡量函数来计算。例如，可以用杰卡德距离（Jaccard similarity coefficient）来计算相似度。可以将两个节点之间的相似度大小作为边权重，基于边权重大小可选择剪枝优化生成用户相似权重关联图。其中，所述剪枝优化，其核心是设置一个边权重阈值，若两个节点之间的关联边的权重（相似度）小于边权重阈值，可以剪掉该关联边。

[61]生成用户关联图后，执行步骤 S13，基于用户关联图，利用社区划分算法生成多个待判定团伙集合。请参照图 3，利用社区划分算法生成多个待判定团伙集合，包括以下步骤。

[62]S131，基于所述用户关联图，利用社区划分算法生成 n 个团伙集合，n 为正整数。其中，所述社区划分算法包括图标签传播算法或 GN（Girvan-Newman）算法。

[63]S132，对于每个所述团伙，根据其用户数量的大小进行调整，以得到多个新的团伙集合。

[64]S133，将所述多个新的团伙集合确定为多个待判定团伙集合。

[65]对于步骤 S132，调整每个团伙集合的用户数量的大小，以使得每个所述团伙集合中用户数量小于或等于极大阈值，并且用户数量小于极小阈值的团伙集合的数量小于或等于预设阈值。其中，所述极大阈值大于所述极小阈值。

[66]当某个所述团伙集合中用户数量（比如不同的账号数量）大于极大阈值时，则继续使用社区划分算法进行划分以使所述团伙集合中用户数量小于或等于所述极大阈值。比如，假设极大阈值为 20，当一个团伙集合中不同的账号数量超过 20 个时，继续使用社区划分算法进行划分以使该团伙集合中用户数量小于或等于 20 个。若用户数量小于极小阈值的团伙集合的数量大于所述预设阈值，则调用层次聚类算法对用户数量小于极小阈值的团伙集合进行凝聚，这里层次聚类可选层次凝聚法或分裂法。比如，假设极小阈值为 3，预设阈值为 15，若不同的账号数量小于 3 个团伙集合数超过 15 个，则调用层次聚类算法对用户数量小于 3 的团伙集合进行凝聚，使得不同的账号数量小于 3 个团伙集合数不超过 15 个。

[67]仍以极大阈值为 20、极小阈值为 3、预设阈值为 15 为例进行说明。经过步骤 S131-132 后，生成 m 个新的团伙集合，这 m 个新的团伙集合中，每个团伙集合的用户数量都小于或等于 20。这 m 个新的团伙集合中，团伙集合的用户数量小于 3 的团伙集合不超过 15 个。将每个新的团伙集合作为一个待判定团伙集合，因此得到 m 个待判定团伙集合。

[68]生成待判定团伙集合后，执行步骤 S14，对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度。嫌疑度的计算方式包括但不限于以下三种方式。

[69]第一种计算方式，参考图 4，计算所述待判定团伙集合的嫌疑度，包括以下步骤。

[70]S141a，从所述数据特征中选取目标数据特征。

[71]S142a，根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的嫌疑度。

[72]目标数据特征是指核心要素特征。在一个实施例中，可以基于业务和场景，指定目标数据特征。目标数据特征可以是 UUID（Universally Unique Identifier，通用唯一标识码）、IP 地址、操作位置信息、操作时间节点信息、设备类型信息、系统版本信息中的一项或多项。举例来说，对于判断虚假注册这种场景，核心要素可以是 IP 地址、设备类型和注册时间，那么对于一个待判定团伙集合来说，那么所述目标数据特征可以是 IP 地址、设备类型和注册时间中的一个或多个。

[73]在另一个实施例中，简单起见，可以直接使用生成用户关联图中所有的特征字段为

目标数据特征。

[74]在再一个实施例中，可以计算某个数据特征在整体数据中的占比，以及该数据特征在待判定团伙集合中的占比。其中，整体数据是指所有用户的数据。若这两个占比之间的差异超过目标阈值，则将该数据特征作为目标数据特征。

[75]举例来讲，以某客户端的账号数量 100 个为例，其中使用虚拟手机号注册账号的数量为 8 个，则使用虚拟手机号注册的账号在账号中的分布比例为 8%。生成的某个待判定团伙集合中，账号数量是 10 个，其中有 7 个账号是使用虚拟手机号注册的，分布比例为 70%，70%对比 8%，差异性很大。则以虚拟手机号注册的账号为目标数据特征，所述目标数据特征在所述待判定团伙集合中的占比为 0.7，可以将该占比作为所述待判定团伙集合的嫌疑度。

[76]或者，以某客户端账号数量 100 个为例，其中历史嫌疑用户注册账号的数量为 8 个，则历史嫌疑用户注册的账号在账号中的分布比例为 8%。生成的某个待判定团伙集合中，账号数量是 10 个，其中有 8 个账号是历史嫌疑用户注册的，分布比例为 80%，80%对比 8%，差异性很大。则以历史嫌疑用户注册的账号为目标数据特征，所述目标数据特征在所述待判定团伙集合中的占比为 0.8，可以将该占比作为所述待判定团伙集合的嫌疑度。

[77]第二种计算方式，参考图 5，计算所述待判定团伙集合的嫌疑度，可以包括以下步骤。

[78]S141b，抽取每个所述待判定团伙集合的团伙特征。其中，所述团伙特征至少包括历史嫌疑用户占比特征，还可以包括团伙规模、共享设备账号数量占比等特征。

[79]S142b，将所述团伙特征输入训练好的回归模型中以使所述回归模型输出所述待判定团伙集合的嫌疑度。其中，所述回归模型可以是 GBDT(Gradient Boosting Decision Tree; 梯度提升决策树）模型。

[80]第三种计算方式，参考图 6，计算所述待判定团伙集合的嫌疑度，可以包括以下步骤。

[81]S141c，从所述数据特征中选取目标数据特征。

[82]S142c，根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的第一嫌疑度得分。

[83]S143c，抽取每个所述待判定团伙集合的团伙特征。

[84]S144c，将所述团伙特征输入训练好的回归模型中以使所述回归模型输出所述待判定团伙集合的第二嫌疑度得分。

[85]S145c，根据所述第一嫌疑度得分以及所述第二嫌疑度得分，计算所述待判定团伙集合的综合嫌疑度得分，作为待判定团伙集合的嫌疑度。

[86]接着根据计算结果，输出所述待判定团伙的判定结果。比如，当综合嫌疑度得分超过预设值时，则可以判定所述待判定团伙为欺诈团伙。

[87]举例来讲，某个待判定团伙的第一嫌疑度得分为 0.7，第二嫌疑度得分 0.8，则所述待判定团伙集合的综合嫌疑度得分可以取两个得分的平均值 0.75，超过预设值 0.6，则该待判定团伙为欺诈团伙。

[88]本公开根据获取的用户数据和历史嫌疑用户数据，生成用户关联图，并利用社区划分算法生成多个待判定团伙集合，对于每个待判定团伙集合，通过计算待判定团伙集合的嫌疑度，即可以分辨出该待判定团伙集合是否属于欺诈团伙，提供了识别团伙欺诈的准确性。另外，本公开还使用社区划分算法和层次聚类算法，解决了团伙划分结果中团伙规模过大、较小的团伙规模数量很多的问题。并且，本公开通过相似索引的手段来提升图模型数据处理能力，同时利用子图组装、相似边权重可配置地方式生成用户相似权重关联图作为用户关联图，这种方法更加灵活，可以进一步提升欺诈场景下的大规模数据处理能力。

[89]值得说明的是，对于图 1 所示的方法实施例，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，本公开并不受所描述的动作顺序的限制。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，所涉及的动作并不一定是本公开所必须的。

[90]图 7 是本公开一示例性实施例示出的一种基于图模型检测团伙欺诈的装置。如图 7 所示，所述基于图模型检测团伙欺诈的装置 300 包括以下模块。

[91]获取模块 310，用于获取多个用户的数据和历史嫌疑用户数据。

[92]第一生成模块 320，用于根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据数据特征生成的用户关联子图，所述用户关联图的边权重包括节点的相似度。

[93]第二生成模块 330，用于基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合。

[94]计算模块 340，用于对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度；

[95]输出模块 350，用于对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

[96]可选地，如图 8 所示，所述第一生成模块 320 包括：第一选取子模块 321，用于选取所述多个用户数据和所述历史嫌疑用户数据中的特征组合和组数；第一生成子模块 322，用于基于所述特征组合和所述组数，利用特征一致性相等或模糊性相等方式生成用户关联子图，并以所述用户关联子图为节点拼接生成用户无权重关联图；第二生成子模块 323，用于以所述用户无权重关联图中节点的相似度作为边权重生成用户相似权重关联图作为所述用户关联图。

[97]可选地，如图 9 所示，所述第二生成模块 330 包括：第三生成子模块 331，用于基于所述用户关联图，利用社区划分算法生成 n 个团伙集合，n 为正整数；调整子模块 332，用于对于每个所述团伙集合，根据所述团伙集合的用户数量的大小进行调整，以得到多个新的团伙集合；确认子模块 333，用于将所述多个新的团伙集合确定为多个待判定团伙集合。

[98]可选地，如图 10 所示，所述调整子模块 332 还包括：划分单元 3321，用于对用户数量大于极大阈值的团伙集合，调用社区划分算法进行划分，以使所述新的团伙集合中用户数量小于或等于所述极大阈值；凝聚单元 3322，用于若用户数量小于极小阈值的团伙集合的数量大于预设阈值，调用层次聚类算法对所述用户数量小于极小阈值的所述团伙集合进行凝聚。

[99]可选地，所述社区划分算法包括图标签传播算法或 GN 算法；所述层次聚类算法包括凝聚算法或分裂算法。

[100] 可选地，如图 11 所示，所述计算模块 340 包括：第二选取子模块 341a，用于从所述数据特征中选取目标数据特征；第一计算子模块 342a，用于根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的嫌疑度。

[101] 可选地，如图 12 所示，所述计算模块 340 包括：第一抽取子模块 341b，用于抽取每个所述待判定团伙集合的团伙特征；第一输入子模块 342b，用于将所述团伙特征输入训练好的回归模型中以使所述回归模型输出所述待判定团伙集合的嫌疑度。

[102] 可选地，如图 13 所示，所述计算模块 340 包括：第三选取子模块 341c，用于从所述数据特征中选取目标数据特征；第二计算子模块 342c，用于根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的第一嫌疑度得分；第二抽取子模块 343c，用于抽取每个所述待判定团伙集合的团伙特征；第二输入子模块 344c，用于将所述团伙特征输入训练好的回归模型中以使所述回归模型输出所述待判定团伙集合的第二嫌疑度得分；第三计算子模块 345c，用于根据所述第一嫌疑度得分以及所述第二嫌疑度得分，计算所述待判定团伙集合的综合嫌疑度得分。

[103] 关于上述实施例中的装置，其中各个模块执行操作的具体方式已经在有关该方法的实施例中进行了详细描述，此处将不做详细阐述说明。

[104] 本公开还提供一种非易失性计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时，促使所述处理器实现上述任一项可选实施例的基于图模型检测团伙欺诈的方法步骤。

[105] 本公开还提供一种基于图模型检测团伙欺诈的装置，包括：存储器，其上存储有计算机程序；以及处理器，用于执行所述存储器中的所述计算机程序，以实现上述任一项可选实施例的基于图模型检测团伙欺诈的方法步骤。

[106] 图 14 是根据一示例性实施例示出的一种基于图模型检测团伙欺诈的装置 400 的框图。如图 14 所示，该装置 400 可以包括：处理器 401，存储器 402，多媒体组件 403，输入/输出（Input/Output）接口 404，以及通信组件 405。

[107] 其中，处理器 401 用于控制该装置 400 的整体操作，以完成上述的基于图模型检测团伙欺诈的方法中的全部或部分步骤。存储器 402 用于存储各种类型的数据以支持在该装置 400 的操作，这些数据例如可以包括用于在该装置 400 上操作的任何应用程序或方法的指令，以及应用程序相关的数据。该存储器 402 可以由任何类型的易失性或非易失性存储设备或者它们的组合实现，例如静态随机存取存储器（Static Random Access Memory，简称 SRAM），电可擦除可编程只读存储器（Electrically Erasable Programmable Read-Only Memory，简称 EEPROM），可擦除可编程只读存储器（Erasable Programmable Read-Only Memory，简称 EPROM），可编程只读存储器（Programmable Read-Only Memory，简称 PROM），只读存储器（Read-Only Memory，简称 ROM），磁存储器，快闪存储器，磁盘或光盘。多媒体组件 403 可以包括屏幕和音频组件。其中屏幕例如可以是触摸屏，音频组件用于输出和/或输入音频信号。例如，音频组件可以包括一个麦克风，麦克风用于接收外部音频信号。所接收的音频信号可以被进一步存储在存储器 402

或通过通信组件 405 发送。音频组件还包括至少一个扬声器，用于输出音频信号。I/O 接口 404 为处理器 401 和其他接口模块之间提供接口，上述其他接口模块可以是键盘，鼠标，按钮等。这些按钮可以是虚拟按钮或者实体按钮。通信组件 405 用于该装置 400 与其他设备之间进行有线或无线通信。无线通信，例如 Wi-Fi，蓝牙，近场通信（Near Field Communication，简称 NFC），2G、3G 或 4G，或它们中的一种或几种的组合，因此相应的该通信组件 405 可以包括：Wi-Fi 模块，蓝牙模块，NFC 模块。

[108] 在一示例性实施例中，装置 400 可以被一个或多个应用专用集成电路（Application Specific Integrated Circuit，简称 ASIC）、数字信号处理器（Digital Signal Processor，简称 DSP）、数字信号处理设备（Digital Signal Processing Device，简称 DSPD）、可编程逻辑器件（Programmable Logic Device，简称 PLD）、现场可编程门阵列（Field Programmable Gate Array，简称 FPGA）、控制器、微控制器、微处理器或其他电子元件实现，用于执行上述的基于图模型检测团伙欺诈的方法。

[109] 在另一示例性实施例中，还提供了一种包括程序指令的计算机可读存储介质，例如包括程序指令的存储器 402，上述程序指令可由装置 400 的处理器 401 执行以完成上述的基于图模型检测团伙欺诈的方法。

[110] 以上结合附图详细描述了本公开的优选实施方式，但是，本公开并不限于上述实施方式中的具体细节，在本公开的技术构思范围内，可以对本公开的技术方案进行多种简单变型，这些简单变型均属于本公开的保护范围。

[111] 另外需要说明的是，在上述具体实施方式中所描述的各个具体技术特征，在不矛盾的情况下，可以通过任何合适的方式进行组合。为了避免不必要的重复，本公开对各种可能的组合方式不再另行说明。

[112] 此外，本公开的各种不同的实施方式之间也可以进行任意组合，只要其不违背本公开的思想，其同样应当视为本公开所公开的内容。

权利要求书

1.一种基于图模型检测团伙欺诈的方法，其特征在于，所述方法包括:

获取多个用户的数据和历史嫌疑用户数据;

根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据数据特征生成的用户关联子图，所述用户关联图的边权重包括所述节点的相似度;

基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合;

对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度;

对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

2.根据权利要求 1 所述的方法，其特征在于，生成所述用户关联图，包括:

选取所述多个用户的所述数据和所述历史嫌疑用户数据中的特征组合和组数;

基于所述特征组合和所述组数，利用特征一致性相等或模糊性相等方式生成用户关联子图;

以所述用户关联子图为节点拼接生成用户无权重关联图;

以所述用户无权重关联图中节点的相似度作为边权重，生成用户相似权重关联图作为所述用户关联图。

3.根据权利要求 1 所述的方法，其特征在于，利用所述社区划分算法生成所述多个待判定团伙集合，包括:

基于所述用户关联图，利用所述社区划分算法生成 n 个团伙集合，n 为正整数;

对于每个所述团伙集合，根据所述团伙集合的用户数量的大小进行调整，以得到多个新的团伙集合;

将所述多个新的团伙集合确定为所述多个待判定团伙集合。

4.根据权利要求 3 所述的方法，其特征在于，根据所述团伙集合的所述用户数量的大小进行调整，包括:

对用户数量大于极大阈值的团伙集合，调用所述社区划分算法进行划分，以使所述新的团伙集合中的用户数量小于或等于所述极大阈值;

若用户数量小于极小阈值的团伙集合的数量大于预设阈值，调用层次聚类算法对所述用户数量小于极小阈值的所述团伙集合进行凝聚。

5.根据权利要求 4 所述的方法，其特征在于，

所述社区划分算法包括图标签传播算法或 GN 算法;

所述层次聚类算法包括凝聚算法或分裂算法。

6.根据权利要求 1 所述的方法，其特征在于，计算所述待判定团伙集合的所述嫌疑度，包括:

从所述数据特征中选取目标数据特征;

根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的所述嫌疑度。

7.根据权利要求 1 所述的方法，其特征在于，计算所述待判定团伙集合的所述嫌疑度，包括:

抽取每个所述待判定团伙集合的团伙特征;

将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的所述嫌疑度。

8.一种基于图模型检测团伙欺诈的装置，其特征在于，所述装置包括:

获取模块，用于获取多个用户的数据和历史嫌疑用户数据;

第一生成模块，用于根据获取的数据，生成用户关联图，其中，所述用户关联图的节点为根据所述数据的特征生成的用户关联子图，所述用户关联图的边权重包括所述节点的相似度;

第二生成模块，用于基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合;

计算模块，用于对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度;

输出模块，用于对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。

9.一种非易失性计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时，促使所述处理器实现权利要求 1 至 7 中任一项所述方法的步骤。

10.一种基于图模型检测团伙欺诈的装置，其特征在于，包括:

存储器，其上存储有计算机程序; 以及

处理器，用于执行所述存储器中的所述计算机程序，以实现权利要求 1 至 7 中任一项所述方法的步骤。

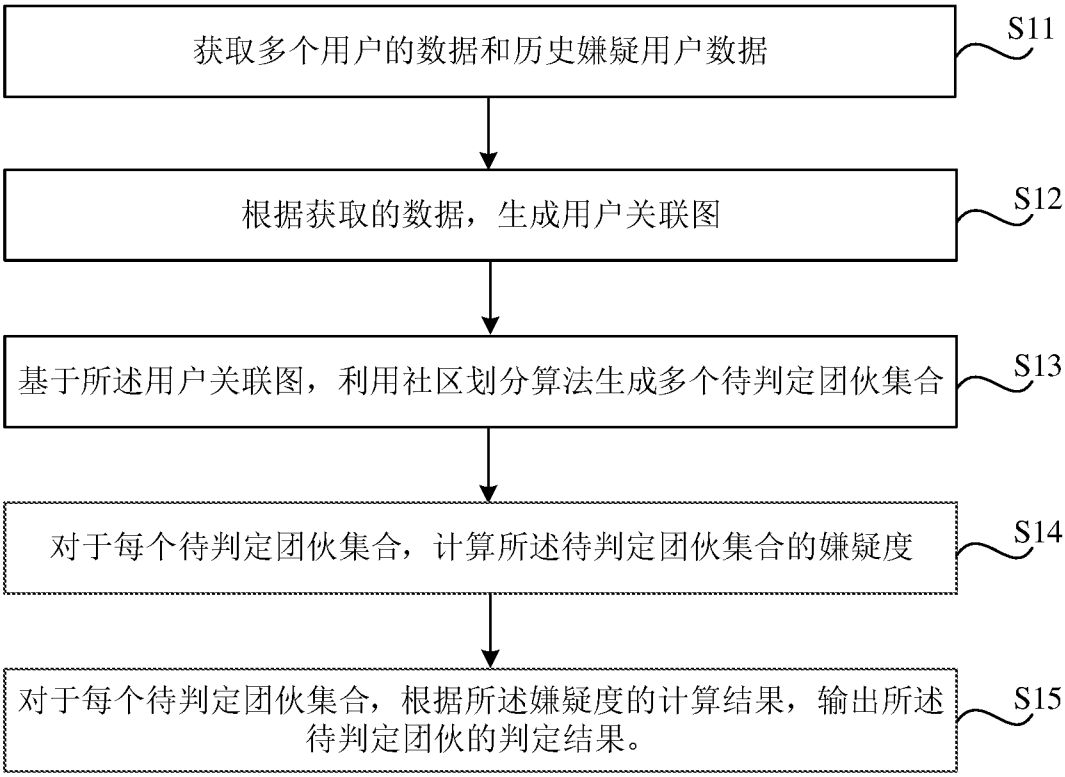
获取多个用户的数据和历史嫌疑用户数据
S11
根据获取的数据，生成用户关联图
S12
基于所述用户关联图，利用社区划分算法生成多个待判定团伙集合
S13
对于每个待判定团伙集合，计算所述待判定团伙集合的嫌疑度
S14
对于每个待判定团伙集合，根据所述嫌疑度的计算结果，输出所述待判定团伙的判定结果。
S15

图 1

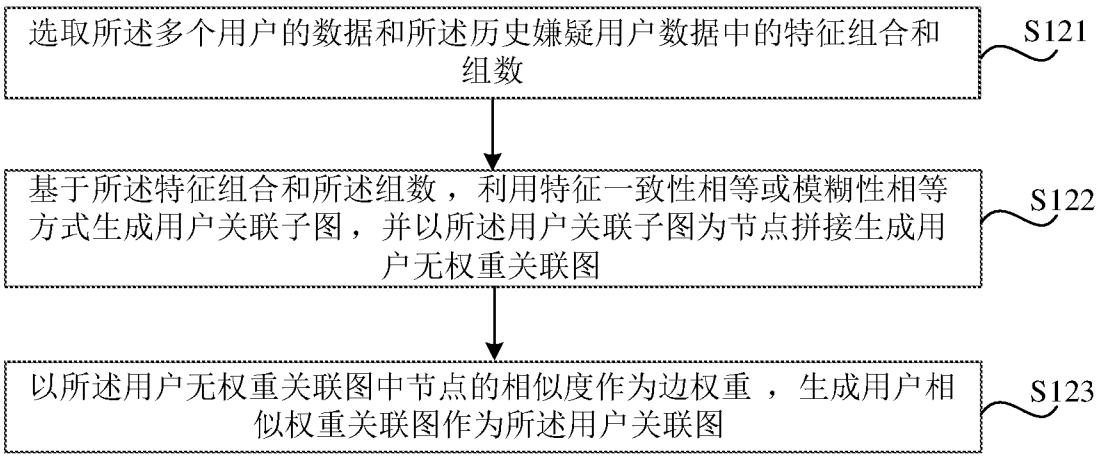
选取所述多个用户的数据和所述历史嫌疑用户数据中的特征组合和组数
S121
基于所述特征组合和所述组数，利用特征一致性相等或模糊性相等方式生成用户关联子图，并以所述用户关联子图为节点拼接生成用户无权重关联图
S122
以所述用户无权重关联图中节点的相似度作为边权重，生成用户相似权重关联图作为所述用户关联图
S123

图 2

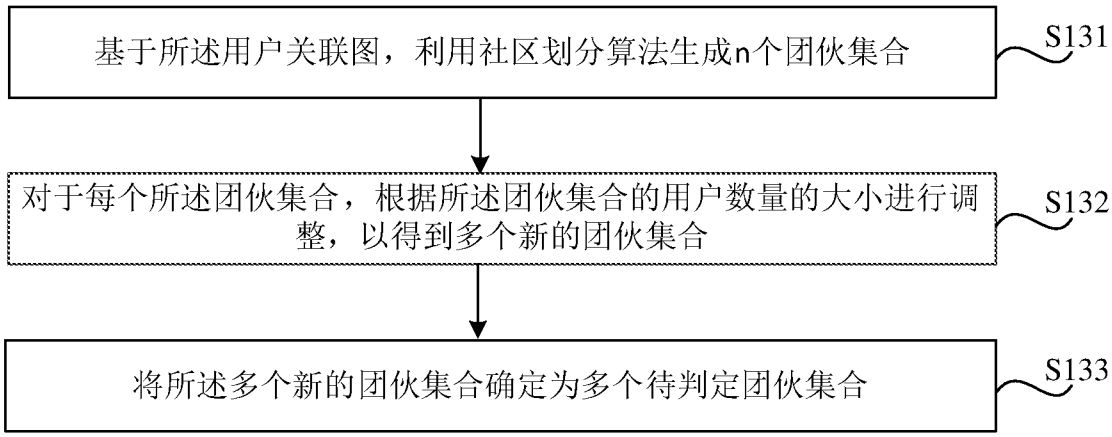
基于所述用户关联图，利用社区划分算法生成n个团伙集合
S131
对于每个所述团伙集合，根据所述团伙集合的用户数量的大小进行调整，以得到多个新的团伙集合
S132
将所述多个新的团伙集合确定为多个待判定团伙集合
S133

图 3

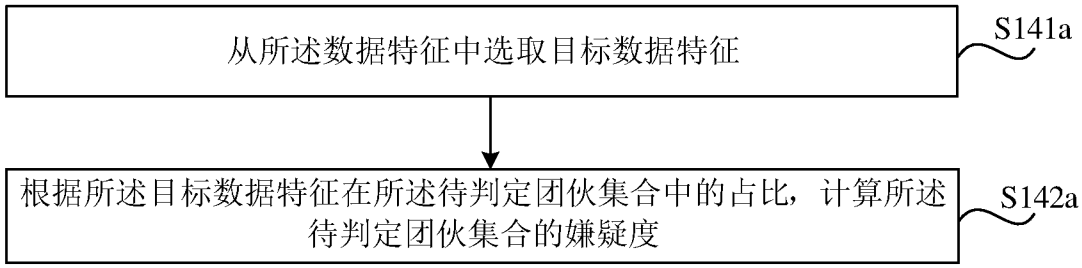
从所述数据特征中选取目标数据特征
S141a
根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的嫌疑度
S142a

图 4

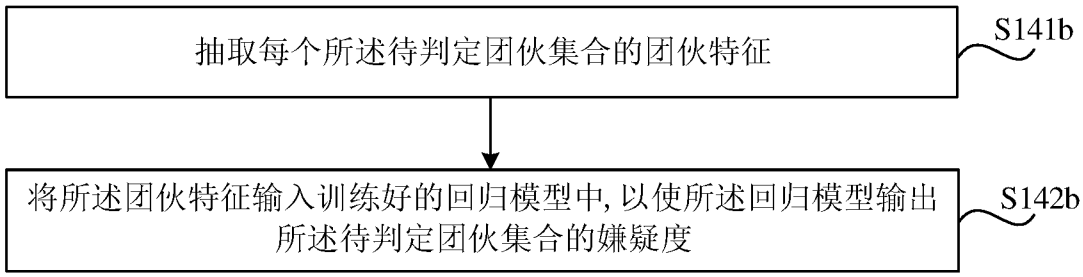
抽取每个所述待判定团伙集合的团伙特征
S141b
将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的嫌疑度
S142b

图 5

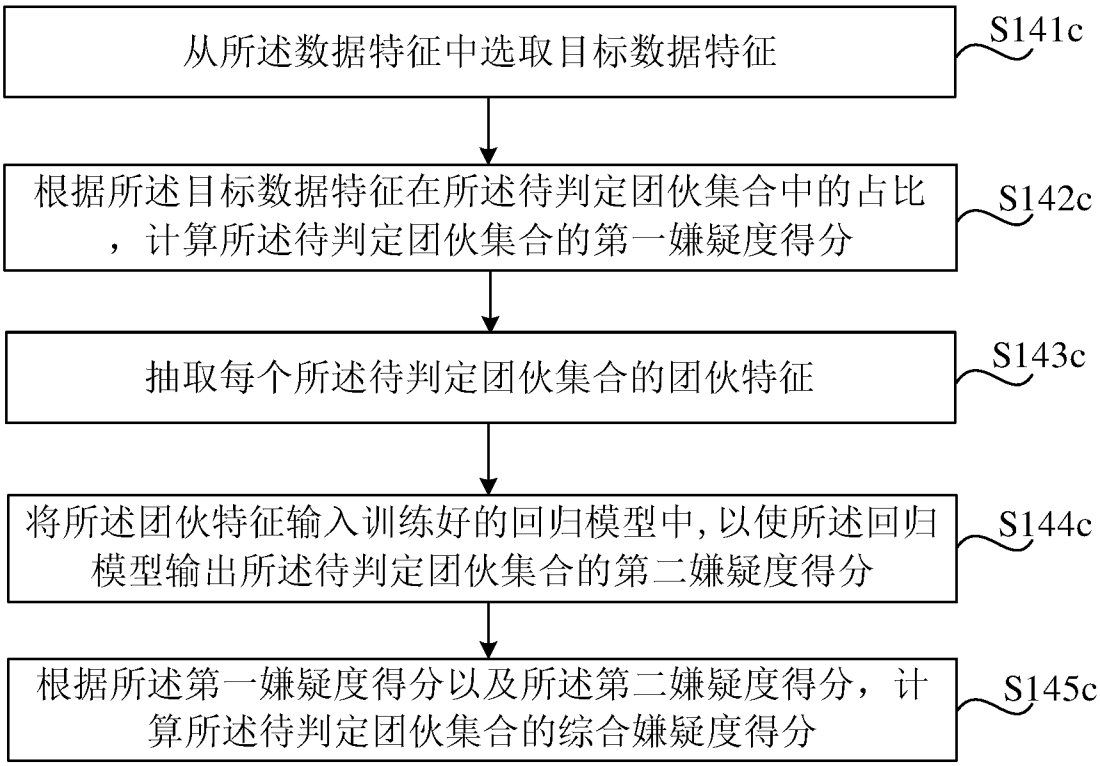
从所述数据特征中选取目标数据特征
S141c
根据所述目标数据特征在所述待判定团伙集合中的占比，计算所述待判定团伙集合的第一嫌疑度得分
S142c
抽取每个所述待判定团伙集合的团伙特征
S143c
将所述团伙特征输入训练好的回归模型中，以使所述回归模型输出所述待判定团伙集合的第二嫌疑度得分
S144c
根据所述第一嫌疑度得分以及所述第二嫌疑度得分，计算所述待判定团伙集合的综合嫌疑度得分
S145c

图 6

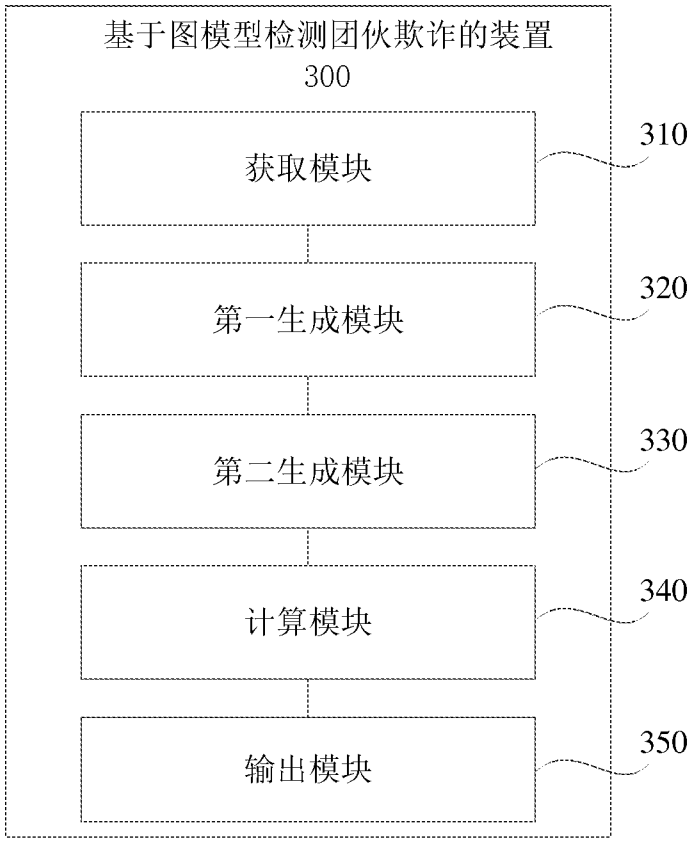
基于图模型检测团伙欺诈的装置
300
获取模块
310
第一生成模块
320
第二生成模块
330
计算模块
340
输出模块
350

图 7

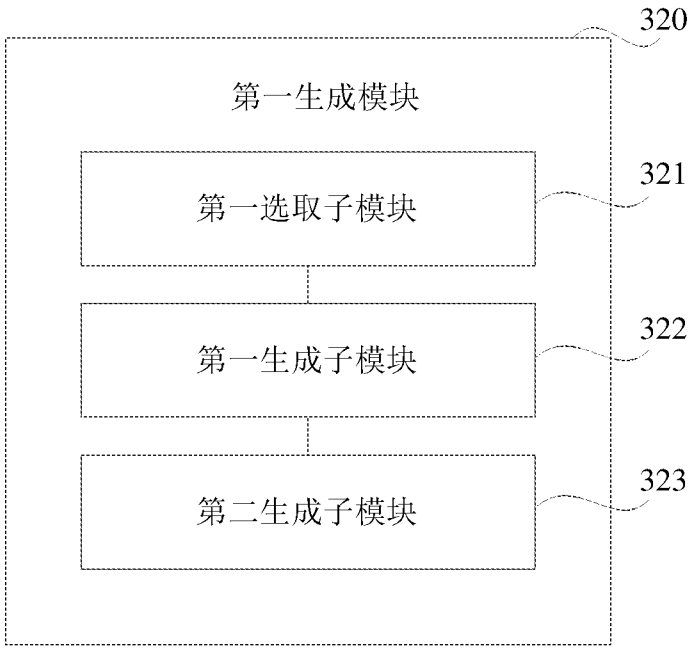
320
第一生成模块
第一选取子模块
321
第一生成子模块
322
第二生成子模块
323

图 8

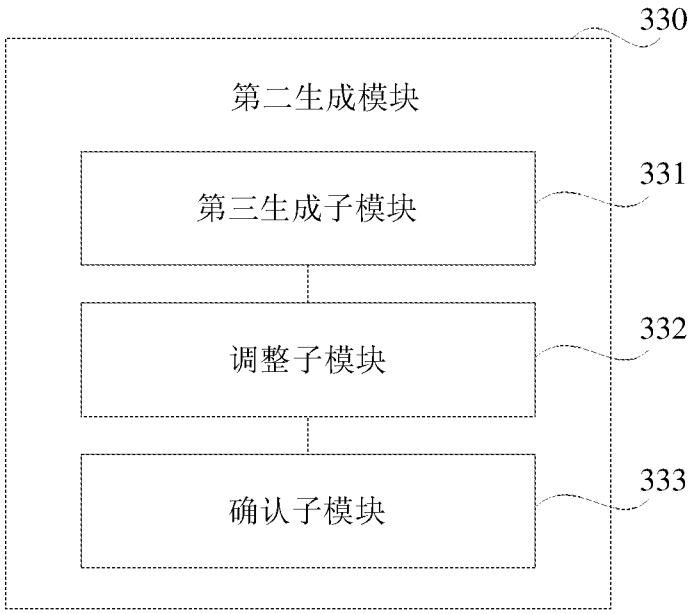
330
第二生成模块
331
第三生成子模块
332
调整子模块
333
确认子模块

图 9

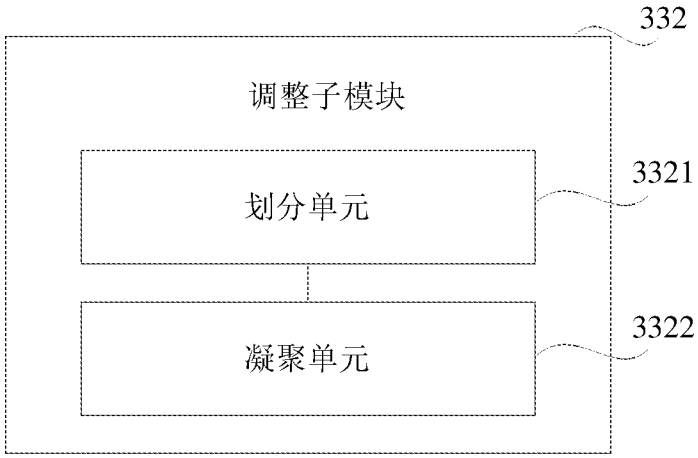
332
调整子模块
3321
划分单元
3322
凝聚单元

图 10

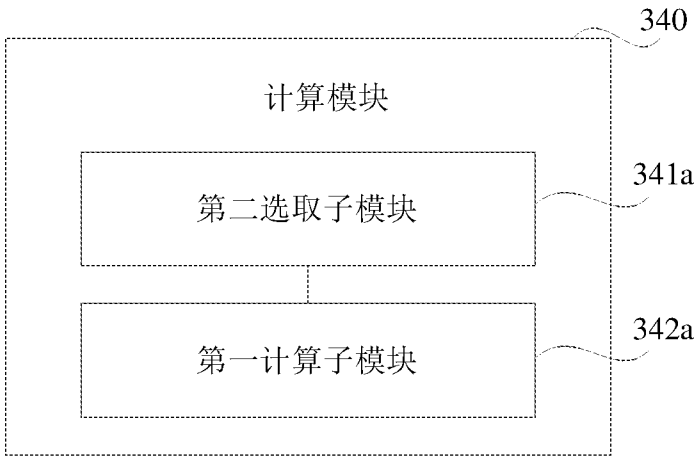
340
计算模块
341a
第二选取子模块
342a
第一计算子模块

图 11

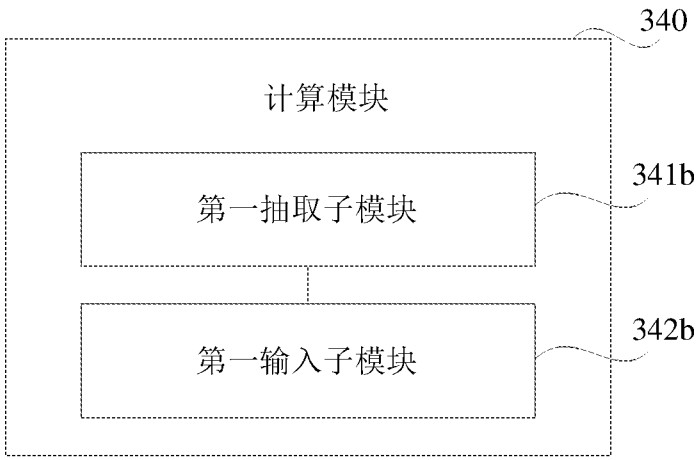
340
计算模块
341b
第一抽取子模块
342b
第一输入子模块

图 12

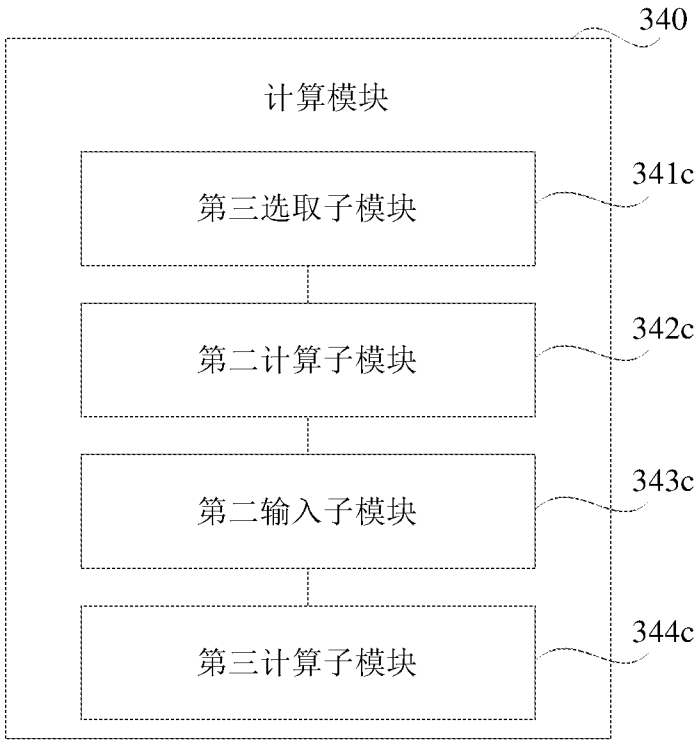
340
计算模块
341c
第三选取子模块
342c
第二计算子模块
343c
第二输入子模块
344c
第三计算子模块

图 13

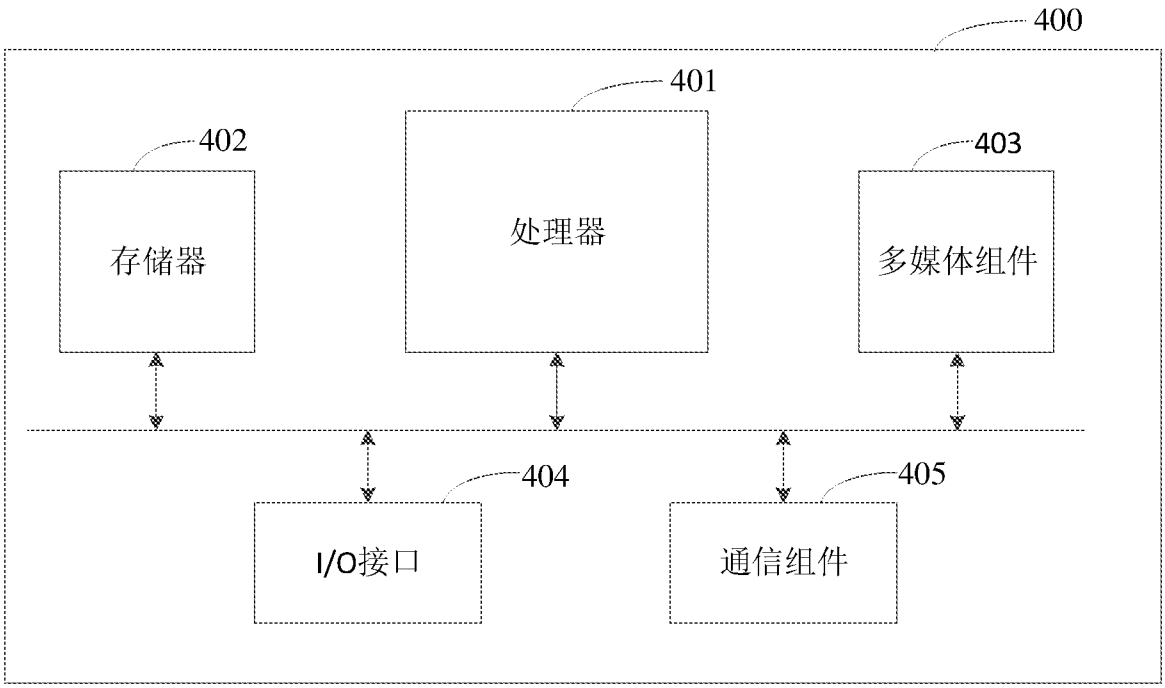
400
401
402
存储器
处理器
403
多媒体组件
404
I/O接口
405
通信组件

图 14

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2019/124807

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 20/40(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, CNPAT, CNKI: 图, 模型, 团伙, 欺诈, 账号, 账户, 相似度, 社区, 社团, 网络, 节点, graph, model, fraud, gang, group, account, ID, similarity, community, network, node

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 110070364 A (BEIJING SCIENCE AND TECHNOLOGY CO., THREE FAST ONLINE) 30 July 2019 (2019-07-30) description, paragraphs [0007]-[0043]	1-10
Y	CN 107194623 A (SHENZHEN FENQILE NETWORK TECH CO., LTD.) 22 September 2017 (2017-09-22) description, paragraphs [0064]-[0083]	1-10
Y	CN 105812195 A (ALIBABA GROUP HOLDING LIMITED) 27 July 2016 (2016-07-27) description, paragraphs [0067]-[0092] and [0099]-[0102]	1-10
A	CN 108681936 A (ZHEJIANG BANGSUN TECH CO., LTD.) 19 October 2018 (2018-10-19) entire document	1-10
A	US 2014067873 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 06 March 2014 (2014-03-06) entire document	1-10

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
02 March 2020	**12 March 2020**
Name and mailing address of the ISA/CN **China National Intellectual Property Administration (ISA/CN)** **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/124807

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 110070364 A	30 July 2019	None	
CN 107194623 A	22 September 2017	None	
CN 105812195 A	27 July 2016	None	
CN 108681936 A	19 October 2018	None	
US 2014067873 A1	06 March 2014	US 2013346467 A1	26 December 2013

国际检索报告

国际申请号
PCT/CN2019/124807

A. 主题的分类

G06Q 20/40(2012.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06Q

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称，和使用的检索词（如使用）)

WPI, EPODOC, CNPAT, CNKI: 图, 模型, 团伙, 欺诈, 账号, 账户, 相似度, 社区, 社团, 网络, 节点, graph, model, fraud, gang, group, account, ID, similarity, community, network, node

C. 相关文件

类 型*	引用文件，必要时，指明相关段落	相关的权利要求
PX	CN 110070364 A (北京三快在线科技有限公司) 2019年 7月 30日 (2019-07-30) 说明书第[0007]-[0043]段	1-10
Y	CN 107194623 A (深圳市分期乐网络科技有限公司) 2017年 9月 22日 (2017-09-22) 说明书第[0064]-[0083]段	1-10
Y	CN 105812195 A (阿里巴巴集团控股有限公司) 2016年 7月 27日 (2016-07-27) 说明书第[0067]-[0092]、[0099]-[0102]段	1-10
A	CN 108681936 A (浙江邦盛科技有限公司) 2018年 10月 19日 (2018-10-19) 全文	1-10
A	US 2014067873 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 2014年 3月 6日 (2014-03-06) 全文	1-10

☐ 其余文件在C栏的续页中列出。

☑ 见同族专利附件。

* 引用文件的具体类型：

"A" 认为不特别相关的表示了现有技术一般状态的文件

"E" 在国际申请日的当天或之后公布的在先申请或专利

"L" 可能对优先权要求构成怀疑的文件，或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件（如具体说明的）

"O" 涉及口头公开、使用、展览或其他方式公开的文件

"P" 公布日先于国际申请日但迟于所要求的优先权日的文件

"T" 在申请日或优先权日之后公布，与申请不相抵触，但为了理解发明之理论或原理的在后文件

"X" 特别相关的文件，单独考虑该文件，认定要求保护的发明不是新颖的或不具有创造性

"Y" 特别相关的文件，当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时，要求保护的发明不具有创造性

"&" 同族专利的文件

国际检索实际完成的日期	国际检索报告邮寄日期
2020年 3月 2日	2020年 3月 12日
ISA/CN的名称和邮寄地址 中国国家知识产权局(ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451	受权官员 刘芬 电话号码 86-(10)-53961434

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2019/124807

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	110070364	A	2019年 7月 30日	无			
CN	107194623	A	2017年 9月 22日	无			
CN	105812195	A	2016年 7月 27日	无			
CN	108681936	A	2018年 10月 19日	无			
US	2014067873	A1	2014年 3月 6日	US	2013346467	A1	2013年 12月 26日