(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
2021 年 9 月 10 日 (10.09.2021)

WIPO | PCT

(10) 国际公布号
WO 2021/174693 A1

(51) 国际专利分类号:
G06Q 10/06 (2012.01) *G06F 16/906* (2019.01)
G06F 16/901 (2019.01) *G06F 16/36* (2019.01)

(21) 国际申请号: PCT/CN2020/093201

(22) 国际申请日: 2020 年 5 月 29 日 (29.05.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010146003.1 2020年3月5日 (05.03.2020) CN

(71) 申请人: 平安科技（深圳）有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN]; 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 段洪云 (DUAN, Hongyun); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 汪伟 (WANG, Wei); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 彭琛 (PENG, Chen); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 北京英特普罗知识产权代理有限公司 (INTELLECPRO CHINA LIMITED); 中国北

(54) **Title:** DATA ANALYSIS METHOD AND APPARATUS, AND COMPUTER SYSTEM AND READABLE STORAGE MEDIUM

(54) **发明名称:** 一种数据分析方法、装置、计算机系统及可读存储介质

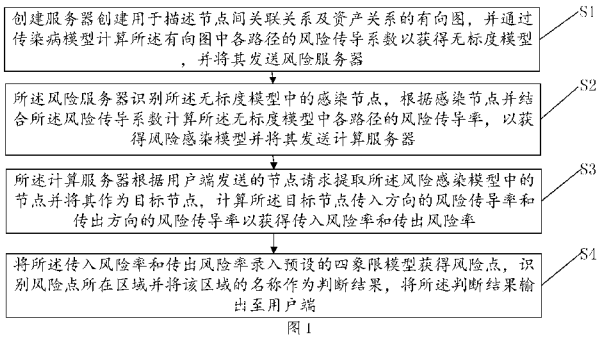

图 1

S1 A creation server creates a directed graph used for describing an association relationship and an asset relationship between nodes, calculates a risk conduction coefficient of each path in the directed graph by means of an infectious disease model, so as to obtain a scale-free model, and sends the scale-free model to a risk server

S2 The risk server identifies an infection node in the scale-free model, calculates the risk conductivity of each path in the scale-free model according to the infection node in combination with the risk conduction coefficient, so as to obtain a risk infection model, and sends the risk infection model to a calculation server

S3 The calculation server extracts a node in the risk infection model according to a node request sent by a user end, takes the node as a target node, and calculates the risk conductivity of the target node in an incoming direction and the risk conductivity thereof in an outgoing direction, so as to obtain an incoming risk rate and an outgoing risk rate

S4 Input the incoming risk rate and the outgoing risk rate into a preset four-quadrant model to obtain a risk point, identify a region where the risk point is located and take the name of the region as a determination result, and output the determination result to the user end

(57) **Abstract:** Disclosed are a data analysis method and apparatus, and a computer system and a readable storage medium, which are based on the field of big data. The method comprises: creating a directed graph used for describing an association relationship and an asset relationship between nodes, and calculating a risk conduction coefficient of each path in the directed graph by means of an infectious disease model, so as to obtain a scale-free model; identifying an infection node in the scale-free model, and calculating the risk conductivity of each path in the scale-free model according to the infection node in combination with the risk conduction coefficient, so as to obtain a risk infection model; and extracting a node in the risk infection model according to a node request sent by a user end, taking the node as a target node, and calculating the risk conductivity of the target node in an incoming direction and the risk conductivity thereof in an outgoing direction, so as to obtain an incoming risk rate and an outgoing risk rate. According to the present application, the technical effects of deep mining of information in a knowledge graph and providing deep and valuable information are achieved.

(57) **摘要**: 本申请公开了一种数据分析方法、装置、计算机系统及可读存储介质，基于大数据领域，包括：创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算有向图中各路径的风险传导系数以获得无标度模型；识别无标度模型中的感染节点，根据感染节点并结合风险传导系数计算无标度模型中各路径的风险传导率，以获得风险感染模型；根据用户端发送的节点请求提取风险感染模型中的节点并将其作为目标节点，计算目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率。本申请实现了对知识图谱中信息的深度挖掘，并提供深层次且有价值的信息的技术效果。

WO 2021/174693 A1

[见续页]

京市西城区车公庄大街9号五栋大楼C座11层, Beijing 100044 (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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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))。

一种数据分析方法、装置、计算机系统及可读存储介质

本申请要求于2020年3月5日提交中国专利局、申请号为CN 202010146003.1，发明名称为“一种数据分析方法、装置、计算机系统及可读存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及计算机技术领域，其涉及到大数据的数据挖掘技术，尤其涉及一种数据分析方法、装置、计算机系统及可读存储介质。

背景技术

当企业风险管理逐渐成为业内焦点，市面上出现了许多企业分析及风险监控类产品，如企查查、天眼查等，但是当前企业风险分析聚焦于企业自身各项风险，如工商信息、财务报表、法律诉讼、舆情风险等，然而当前的风险监控类产品却只能获知法律文件、账面数据以及风险评级等信息，并且这些信息通常采用知识图谱进行管理及展示。

然而当前的知识图谱仅说明了节点数据及各节点之间的关联关系，但是仅凭这些信息仅仅能够提供节点基本信息，以及节点之间的关联信息，而未对知识图谱中的信息进行深度挖掘，导致其仅能够提供简单的信息，而无法向使用者提供可直接使用的有价值的信息。

发明内容

本申请的目的是提供一种数据分析方法、装置、计算机系统及可读存储介质，用于解决现有技术存在的未对知识图谱中的信息进行深度挖掘，导致其仅能够提供简单的信息，而无法向使用者提供可直接使用的有价值的信息的技术问题。

为实现上述目的，本申请提供一种基于大数据的数据分析方法，包括：

创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

为实现上述目的，本申请还提供一种基于大数据的数据分析装置，包括：

创建服务器，用于创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

风险服务器，用于识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

计算服务器，用于根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

为实现上述目的，本申请还提供一种计算机系统，其包括多个计算机设备，各计算机

设备包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，所述多个计算机设备的处理器执行所述计算机程序时共同实现上述数据分析方法的步骤。

为实现上述目的，本申请还提供一种计算机可读存储介质，其包括多个存储介质，各存储介质上存储有计算机程序，所述多个存储介质存储的所述计算机程序被处理器执行时共同实现上述数据分析方法的步骤。

本申请提供的一种数据分析方法、装置、计算机系统及可读存储介质，通过创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，以描述各节点之间的关联程度；通过识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型；因此根据感染节点，并通过风险传导系数和平均传导率，计算无标度模型中各路径的风险传导率，用于表达风险从感染节点传导至其他节点的风险传导率，因该风险传导率是基于平均传导率和风险传导系数所获得，因此能够最真实的反应出相关的两个节点之间风险传导的概率。根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；实现了全面的获知目标节点的传出风险率和传入风险率，以全面评价目标节点的风险特征，以便于向用户全面传递目标节点的风险环境，有利于用户根据该环节进行判断。通过四象限模型计算所述传入风险率和传出风险率以获得判断结果，并将其输出至所述用户端，实现了采用四象限模型评价目标节点的风险特征，并将该特征以名称或图标的方式输出至用户端，以便于用户快速获知用于描述周围节点对目标节点影响，以及目标节点对周围节点的影响的风险环境，实现了对知识图谱中信息的深度挖掘，进而解决了现有技术中因未对知识图谱中的信息进行深度挖掘，导致其仅能够提供简单的信息，而无法向使用者提供可直接使用的有价值的信息的技术问题。

附图说明

图 1 为本申请数据分析方法实施例一的流程图；

图 2 为本申请数据分析方法实施例一中欧拉图谱的数据关联方式的有向图；

图 3 为本申请数据分析方法实施例一中欧拉图谱的数据关联方式的树状图；

图 4 为本申请数据分析方法实施例一中具有初级传导系数的有向图；

图 5 为本申请数据分析方法实施例一中将风险传导系数其加载在各路径上的具有初级传导系数的有向图；

图 6 为本申请数据分析方法实施例一中风险感染模型的有向图；

图 7 为本申请数据分析方法实施例一中四象限模型的示意图；

图 8 为本申请数据分析装置实施例二的程序模块示意图；

图 9 为本申请计算机系统实施例三中计算机设备的硬件结构示意图。

附图标记：

1、数据分析装置　　2、计算机设备　　11、创建服务器

12、风险服务器　13、计算服务器　14、判断服务器

21、存储器　　22、处理器

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本申请，并不用于限定本申请。基于本申请中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

本申请提供的一种数据分析方法、装置、计算机系统及可读存储介质，适用于计算机领域，为提供一种基于大数据创建模块、感染模型创建模块、风险计算模块和风险判断模块的数据分析方法。本申请通过创建用于描述节点间关联关系及资产关系的有向图，并通

过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型；识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型；根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；通过四象限模型计算所述传入风险率和传出风险率以获得判断结果，并将其输出至所述用户端。

实施例一

请参阅图 1，本实施例的一种基于大数据的数据分析方法，包括：

S1：创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

S2：所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

S3：所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

在一个优选的实施例中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率之后包括：

S4：将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端。

在示例性的实施例中，通过创建有向图以表达节点之间的关联关系和资产关系，其中，节点可以为企业，也可以为自然人，因此，可将企业或自然人的名称或姓名作为节点的名称；通过传染病模型计算有向图在均衡状态下（即：受到感染节点影响的节点数与摆脱感染节点影响的节点数相等）时的平均传导概率，根据所述平均传导概率计算有向图中各路径的风险传导概率，并将其加载于各路径上，以获得无标度模型；其中，所述无标度模型是以反应了各节点之间关联关系和资产关系的有向图为基础，通过传染病模型对有向图中各路径风险传导概率进行计算并加载于各路径上的数据模型。

需要说明的是，有向图在数学上是表示物件与物件之间的关系的方法，其由一些小圆点(称为顶点或结点)和连结这些圆点的直线或曲线(称为边)组成的；于本实施例中，所述圆点则对应本申请中的节点，所述连接这些圆点的直线或曲线，所述连接圆点的直线或曲线的箭头及加载于所述直线或曲线上的信息对应所述资产关系。传染病模型是一种用于计算节点影响力的标准测量模型，所述传染病模型中的节点有三个状态：易感者（susceptible），感染者（infected），恢复者（recovered）。易感者能够被感染者染病，感染者染病且有感染易感者的能力，恢复者由感染者恢复而成，不再具备感染能力同时也不会再被感染。传染病模型参数有恢复率 μ，传染概率 β，重复仿真次数 T，仿真时间 timespace。仿真过程如下：选择网络中一个节点 i 作为感染者节点，以传染概率 β 向 i 连接的邻居易感者节点传染病毒，被染病的节点继续以 β 的概率向他们的邻居易感者节点传染。同时每个染病节点在每阶段以 μ 的概率转变成恢复节点。

可采用黑名单制度识别所述无标度模型中的感染节点，即：若某一节点属于黑名单则判定该节点为感染节点；设定感染节点的风险为 100%，将与感染节点通过路径连接的节点作为直接连接的节点，根据该路径上的风险传导系数计算风险从感染节点传导至所述直接连接的节点的风险传导率；再识别与所述连接的节点相连的节点及其连接的路径，并将所述相连的节点设为间接连接的节点，通过所述连接的路径上的风险传导系数计算风险从感

染节点经直接连接的节点传导至所述间接连接的节点的风险传导率，依次类推，直至将所述无标度模型中所有与感染节点直接连接和间接连接的节点计算完成后，获得风险感染模型。

所述节点请求中包括有用于对应风险感染模型中节点的节点名称，将所述风险感染模型中具有所述节点名称的节点作为目标节点，识别风险感染模型中的路径，将指向目标节点的路径作为传入路径，将从目标节点指出的路径作为传出路径，根据所述传入路径的风险传导率获得传入方向的传入风险率，根据所述传出路径的风险传导率获得传出方向的传出风险率。

用户可在四象限模型的坐标系中设置区域，并对该区域赋以名称；根据传入风险率和传出风险率在四象限模型中获得风险点，将风险点所在的区域的名称作为判断结果，将所述判断结果输出至用户端。

在一个优选的实施例中，所述创建用于描述节点间关联关系及资产关系的有向图包括：

S101：从服务系统中获取节点数据以及各节点之间的关联关系，根据所述关联关系构建用于描述节点间关联关系的有向图。

示例性地，从服务系统中获取知识图谱，从所述知识图谱中获取节点数据以及各节点之间的关联关系，若两个节点数据之间具有关联关系，则在所述两个节点数据所对应的节点之间绘制路径，若节点 A 对节点 B 为投资关系，则节点 A 与节点 B 之间的路径为 A 指向 B；其中，因所述知识图谱中的路径具有箭头，故可从知识图谱中直接获知节点 A 与节点 B 之间的关系，故在本申请中不做赘述。

需要说明的是，所述服务系统是储存有大量企业和个人数据、并基于企业和个人数据建立有大量的企业关系数据、股权关系数据、诉讼关系数据等，所述服务系统可覆盖工商、财经公告、法律文书、社交媒体和海外舆情等各大维度；于本申请中所述服务系统采用的是欧拉图谱，其为一种用于储存企业和个人数据，并根据企业和个人数据构建包含关联人物、关联企业的企业关联网络，从股东持股关系、对外投资关系、供应链关系、股权出质关系、融资担保关系和企业高管六大方面的数据关系的知识图谱；本申请所要解决的技术问题是如何深度挖掘当前的知识图谱（本申请中的欧拉图谱）的知识关系以获取潜在应用价值的信息；而对于根据知识图谱（本申请中的欧拉图谱）获得节点数据及各节点之间的关联关系，本领域技术人员很容易通过提取的欧拉图谱获得，故在此不做赘述。因该系统为现有技术，故在此不做赘述。

其中，所述欧拉图谱的数据关联方式如图 2 和图 3 所示。

S102：计算所述有向图中相互关联的节点之间的资产关联比值，并将其加载于所述相互关联节点之间的路径上以描述所述有向图中节点间资产关系。

示例性地，提取相互关联的两个节点的节点数据，获取所述节点数据中的被投资数据和投资数据，其中，将所述两个节点中发出箭头（路径）的一方作为投资节点，将箭头（路径）指向的一方作为被投资节点，提取投资节点中的投资数据，以及被投资节点中与所述投资数据关联的被投资数据，将所述被投资数据与投资数据相除获得资产关联比值，并将所述资产关联比值加载在所述路径上。

其中，所述相互关联的两个节点的节点数据是描述其中一个节点向另一节点进行投资的信息，故所述两个节点的节点数据中，必然一个为投资节点的节点数据，另一个为被投资节点的节点数据。需要说明的是，所述资产关联比值是指被投资节点接收到的金额与投资节点对外投资总金额之比；所述与投资数据关联的被投资数据是指，因投资数据对被投资节点进行投资所生成的被投资数据；例如，投资节点的投资数据为 100 万，被投资节点的被投资数据为 50 万，而其中，被投资数据中 20 万为基于所述投资节点的投资数据所得，其他 30 万是基于其他投资节点的投资数据所得，那么所述投资节点和被投资节点之间的资产关联比值为 20%。

在一个优选的实施例中，所述通过传染病模型计算所述有向图中各路径的风险传导系

数以获得无标度模型包括：

S111：通过传染病模型计算所述有向图的模型指标以获得平均传导概率。

本步骤中，所述传染病模型中具有均衡条件设定，通过所述传染病模型计算所述有向图的模型指标在均衡条件下的平均传导率。

需要说明的是，传染病模型是一种用于计算节点影响力的标准测量模型， 基于传染病模型的原理，风险在有向图中的传递过程是易感者收到风险节点的感染会成为感染者，感染者会恢复并成为恢复者，而恢复者也会因再次收到风险节点的感染而再次成为感染者。

因此，通过传染病模型计算有向图的模型指标，获得所述有向图中易感者、感染者和恢复者达到均衡条件下时，有向图中所有节点的总体传染概率，这个总体传染概率是指能够保持有向图均衡条件的风险传导概率。

具体地，提取所述有向图的模型指标中的节点总数、节点度、平均度和平均密度，并将其录入所述传染病模型的目标函数中进行计算，以获得平均传导概率。

其中，所述目标函数如下所示：

$$\frac{\partial s_k(t)}{\partial t} = \rho_k(t-1) - \gamma_k \eta_k k \theta(\rho(t)) s_k(t)$$

该方法首先做出如下定义：

K 是指传染病模型中个体总数，对应于节点总数；风险个体记为 I，健康个体记为 S；S(t)为健康个体在时间 t 时的数量； ρ (t)是在时间 t 时的健康个体比例；k θ 是单位时间内一个风险个体能传染的数量与当时健康个体总数的比例，对应于所述节点度与平均度的乘积；η 是指传染病模型的模型平均密度，对应于所述平均密度；风险个体通过概率 γ 将风险传出，并且风险个体会恢复健康但不会免疫；

通过对上述偏微分方程的动态求解得到均衡状态下（即：当期新染病的人数与当期新恢复健康的人数相等，于本实施例中为受到感染节点影响的节点数与摆脱感染节点影响的节点数相等）的总体传染概率 γ，将所述总体传染概率 γ 设为平均传导概率。

需要说明的是，商业环境传染病模型的目标函数只有在均衡状态下，医学传染病模型即动态模拟人群中染病人数的变化，得到在均衡状态下群体的传染比例。

S112：将所述平均传导概率分别与所述有向图的路径上的资产关联比值相乘以获得初级传导系数，通过随机森林模型识别所述有向图中各相互关联节点之间的相关度，并将其加载于所述相互关联的节点的路径上；

示例性地，提取所述有向图各路径上的资产关联比值，并将其与平均传导概率相乘以获得初级传导系数，通过所述随机森林模型提取任一相互关联的两个节点的节点数据，并识别其之间的关联性；

若判断所述两个节点的节点数据相互关联，则在所述两个节点之间的路径赋以值为 1 的相关度；若判断所述两个节点的节点数据不关联，则在所述两个节点之间的路径赋以值为 0 的相关度；将所述风险传导系数加载于所述有向图的路径上。

需要说明的是，随机森林是一个包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定的模型；因随机森林里面有很多的决策树，且每一棵决策树之间是没有关联的，因此在得到森林之后，当有一个新的输入样本进入的时候，就让森林中的每一棵决策树分别进行一下判断，看看这个样本应该属于哪一类（对于分类算法），然后看看哪一类被选择最多，就预测这个样本为那一类。

所述随机森林通过以下方式训练获得：

1）准备 N 个样本，则有放回的随机选择 N 个样本(每次随机选择一个样本，然后返回继续选择)。这选择好了的 N 个样本用来训练一个决策树，作为决策树根节点处的样本。

2）当每个样本有 M 个属性时，在决策树的每个节点需要分裂时，随机从这 M 个属性中选取出 m 个属性，满足条件 m << M。然后从这 m 个属性中采用分类策略（比如说信息增益）

来选择 1 个属性作为该节点的分裂属性。

3）决策树形成过程中每个节点都要按照步骤 2 来分裂，一直到不能够再分裂为止。

4）按照步骤 1~3 建立大量的决策树，这样就构成了随机森林了。

于本实施例中，用户通过设定所述 m 个属性以及分类策略并通过上述方法来训练初始随机森林，以获得用于识别相互关联的两个节点的节点数据，是否具有关联性的随机森林模型；因随机森林的训练方法属于本领域技术人员的公职常识，因此通过设定 m 个属性及分类策略训练随机森林的训练过程在此不做赘述。

S113：将所述有向图的路径上的初级传导系数和相关度相乘以获得风险传导系数，将所述风险传导系数加载于所述有向图的路径上以获得无标度模型。

例如：所述具有初级传导系数的有向图中包括标号为 A—G 的节点，其布局如图 4 所示。

若通过随机森林模型识别所述有向图的结果是：除节点 C 与节点 E 的相关度为 0，其他的相互关联的节点之间的相关度均为 1；则将各路径上的初级传导系数及其相关度相乘以获得风险传导系数，并将其加载在各路径上，如图 5 所示。

综上，基于所述传染病模型在均衡状态下计算无标度模型整体的平均传导率，以模拟整个无标度模型在动态平衡的状态下风险的传导概率，用于向用户提供理想的，能够量化的全模型平均风险传导概率。同时，由于有些相互关联的节点之间会因为某些因素而不发生风险传导，例如，有限责任等，因此利用随机森林模型识别有向图中相互关联的节点之间的相关度，并根据该相关度判断该风险是否会从关联的两个节点之间传递，以真实准确的描述风险传递方式，其目标在于识别两个节点之间的相关度。

在一个优选的实施例中，所述识别所述无标度模型中的感染节点包括：

S201：判断所述无标度模型的节点是否属于预设的黑名单；

示例性地，创建黑名单，所述黑名单中具有单位名称；将无标度模型中各节点的名称与所述黑名单的进行比对，若节点的名称与所述黑名单中某一单位名称一致，则判定该节点属于黑名单。

S202：将属于所述黑名单的节点设为感染节点。

示例性地，将所述无标度模型中属于黑名单的节点设为传染节点。

在一个优选的实施例中，所述根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率包括：

S211：在所述无标度模型中识别与所述感染节点连续关联的连续路径，并以所述感染节点为起始依次对所述连续路径的子路径进行编号；其中，所述连续路径是指在所述无标度模型中以感染节点为输出源，依次与其串联的关联节点的整条路径，所述子路径为所述连续路径中相邻的两个节点之间的路径。

例如： 若节点 B 为感染节点，上述无标度模型中的连续路径包括：B-D-F 和 B-D-G；B-D-F 的子路径包括 B-D 和 D-F，B-D-G 的子路径包括 B-D 和 D-G；因该连续路径的感染节点为节点 B，因此以节点 B 为起始位置；在 B-D-F 连续路径中，将子路径 B-D 的编号为 01，将子路径 D-F 的编号为 02；在 B-D-G 连续路径中，将子路径 B-D 的编号设为 11，将子路径 D-G 的编号设为 12。

S212：将所述连续路径上的任一子路径设为待算路径，识别编号小于所述待算路径的子路径并提取其中的风险传导系数后汇总形成系数集，将所述系数集中的风险传导系数与所述待算路径的风险传导系数相乘获得真实传导系数，并将其加载于所述待算路径上。

例如，基于上述举例，若待算路径为子路径 B-D，其风险传导系数为 0.2 时，因该子路径在连续路径 B-D-F 和 B-D-G 中均不具有编号小于该待算路径的子路径，因此，将待算路径的真实传导系数设为 0.2 并将其加载于待算路径 B-D 上；

若待算路径为 D-F，其风险传导系数为 0.4 时，则在连续路径 B-D-F 中具有编号小于该待算路径的子路径 B-D，因此提取该子路径 B-D 中的风险传导系数并形成系数集（0.2），将系数集中的风险传导系数 0.2 与待算路径的风险传导系数 0.4 相乘，获得风险传导率

0.08，并将其加载于待算路径 D-F 上；

若待算路径为 D-G，其风险传导系数为 0.5 时，则在连续路径 B-D-G 中具有编号小于该待算路径的子路径 B-D，因此提取该子路径 B-D 中的风险传导系数并形成系数集（0.2），将系数集中的风险传导系数 0.2 与待算路径的风险传导系数 0.5 相乘，获得风险传导率 0.1，并将其加载于待算路径 D-G 上；

经上述方法所获得的风险感染模型如图 6 所示。

综上，根据感染节点，并通过风险传导系数和平均传导率，计算无标度模型中各路径的风险传导率，用于表达风险从感染节点传导至其他节点的风险传导率，因该风险传导率是基于平均传导率和风险传导系数所获得，因此能够最真实的反应出相关的两个节点之间风险传导的概率。

同时，采用黑名单制度识别感染节点，并将风险传导率加载于其对应的所述无标度模型的路径上，以获得风险感染模型，以便于快速识别风险感染模型中任意节点的传入风险和传出风险。

在一个优选的实施例中，所述根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点包括：

S301：接收用户端发送的具有节点名称的节点请求。

本步骤中的节点名称可为企业名称，也可为企业编号。

S302：将所述风险感染模型中所有节点的节点名称与所述节点请求进行对比，并提取与所述节点请求匹配的节点，并将所述节点作为目标节点。

例如，基于上述举例，用户端发送的节点请求具有节点名称 D，则将其与无标度模型中的节点 A-G 依次比对，获得节点 D 为与节点名称 D 相匹配的节点。

在一个优选的实施例中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率包括：

S311：将与所述节点请求匹配的节点设为目标节点，在所述无标度模型中提取与所述目标节点连接的路径；

S312：将指向所述目标节点的路径设为传入路径，将从所述目标节点指出的路径设为传出路径；

本步骤中的路径具有方向，将指向目标节点设为传入路径，用于描述风险从外界传入目标节点的情况；将从所述目标节点指出的路径设为传出路径，用于描述风险从目标节点传出的情况。

S313：通过加权调整公式计算所述目标节点的传入路径的风险传导率，以获得传入风险率，通过加权调整公式计算所述目标节点的传出路径的风险传导率，以获得传出风险率。

示例性地，加权调整公式为

$$T=\begin{cases} a*x+m & (1) \\ b*y+n & (2) \end{cases}$$

其中，T 为传入风险率或传出风险率，其中 T 中的公式（1）用于计算传入风险率，公式（2）用于计算传出风险率；

x 为传入路径的风险传导率，a 为传入系数，m 为传入调整值，传入系数和传入调整值可根据用户的需求自行调节；y 为传出路径的风险传导率，b 为传出系数，n 为传出调整值，传出系数和传出调整值可根据用户的需求自行调节。

例如，基于上述举例，在无标度模型中提取与节点 D 连接的路径，其分别为 B-D，D-F 和 D-G，且由风险感染模型的布局图可知：B-D 为传入路径，D-F 和 D-G 为传出路径；假设加权调整公式中传入系数和传出系数均为 1，传入调整值和传出调整值均为 0，则通过加权调整公式计算传入路径 B-D 的风险传导率，获得传入风险率 0.2，通过加权调整公式计算传出路径 D-F 和 D-G 的风险传导率，获得传出风险率 0.1+0.08=0.18。

综上，通过全面的获知目标节点的传出风险率和传入风险率，以全面评价目标节点的风险特征，以便于向用户全面传递目标节点的风险环境，有利于用户根据该环节进行判断。

在一个优选的实施例中，将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，包括：

S401：将所述传入风险率和传出风险率录入预设的四象限模型；

本步骤中，所述四象限模型是一种基于节点的传入风险率和传出风险率描述该节点风险特征的坐标系，该坐标系的原点为传入风险率和传出风险率分别为 0 的坐标点，所述传入风险率沿着上述坐标系纵轴的延伸方向增加，所述传出风险率沿着上述坐标系的横轴延伸方向增加，具体如图 7 所示。

示例性地，在四象限模型中，通过横坐标分界线和纵坐标分界线相互交叉将坐标系分为四个部分，分别为用于表达传入风险率和传出风险率均较低的“沉寂冰川”区域，用于表达传入风险率较高但传出风险率较低的“浩瀚海洋”区域，用于表达传入风险率较低但传出风险率较高的“活跃火山”区域，以及用于表达传入风险率和传出风险率均较高的“风暴中心”区域。进一步的，上述四个部分的横坐标分界线和纵坐标分界线可根据用户需求自行调节。

S402：在所述四象限模型中以所述传入风险率作为纵坐标，并以所述传出风险率作为横坐标获得风险点；

例如，基于上述举例，将所述传入风险率 0.2 和传出风险率 0.18 分别录入四象限模型的纵坐标和横坐标，以获得风险点。

S403：识别所述风险点所在的区域并将该区域的名称作为判断结果。

本步骤中，该风险点所在的区域，即为用于描述用户端发送的节点请求所对应节点的风险特征，将该区域的名称作为判断结果；例如，基于上述举例，横坐标分界线和纵坐标分界线的交点为（0.5,0.5），识别出所述风险点位于“沉寂冰川”区域，因此将“沉寂冰川”作为判断结果。

综上，实现了通过采用四象限模型评价目标节点的风险特征，并将该特征以名称或图标的方式输出至用户端，以便于用户快速获知用于描述周围节点对目标节点影响，以及目标节点对周围节点的影响的风险环境的技术效果。

实施例二

请参阅图 8，本实施例的一种基于大数据的数据分析装置 1，包括：

创建服务器 11，用于创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

风险服务器 12，用于识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

计算服务器 13，用于根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

可选的，所述数据分析装置 1 还包括：

判断服务器 14，用于将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端。

实施例三

为实现上述目的，本申请还提供一种计算机系统，该计算机系统包括多个计算机设备 5，

实施例二的数据分析装置 1 的组成部分可分散于不同的计算机设备中，计算机设备可以是执行程序的智能手机、平板电脑、笔记本电脑、台式计算机、机架式服务器、刀片式服务器、塔式服务器或机柜式服务器（包括独立的服务器，或者多个服务器所组成的服务器集群）等。本实施例的计算机设备至少包括但不限于：可通过系统总线相互通信连接的存储器 21、处理器 22，如图 9 所示。需要指出的是，图 9 仅示出了具有组件-的计算机设备，但是应理解的是，并不要求实施所有示出的组件，可以替代的实施更多或者更少的组件。

本实施例中，存储器 21（即可读存储介质）包括闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）、随机访问存储器（RAM）、静态随机访问存储器（SRAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、可编程只读存储器（PROM）、磁性存储器、磁盘、光盘等。在一些实施例中，存储器 21 可以是计算机设备的内部存储单元，例如该计算机设备的硬盘或内存。在另一些实施例中，存储器 21 也可以是计算机设备的外部存储设备，例如该计算机设备上配备的插接式硬盘，智能存储卡（Smart Media Card, SMC），安全数字（Secure Digital, SD）卡，闪存卡（Flash Card）等。当然，存储器 21 还可以既包括计算机设备的内部存储单元也包括其外部存储设备。本实施例中，存储器 21 通常用于存储安装于计算机设备的操作系统和各类应用软件，例如实施例一的数据分析装置的程序代码等。此外，存储器 21 还可以用于暂时地存储已经输出或者将要输出的各类数据。

处理器 22 在一些实施例中可以是中央处理器（Central Processing Unit，CPU）、控制器、微控制器、微处理器、或其他数据处理芯片。该处理器 22 通常用于控制计算机设备的总体操作。本实施例中，处理器 22 用于运行存储器 21 中存储的程序代码或者处理数据，例如运行数据分析装置，以实现实施例一的数据分析方法。

实施例四

为实现上述目的，本申请还提供一种计算机可读存储系统，其包括多个存储介质，所述存储介质可以是非易失性，也可以是易失性，如闪存、硬盘、多媒体卡、卡型存储器（例如，SD 或 DX 存储器等）、随机访问存储器（RAM）、静态随机访问存储器（SRAM）、只读存储器（ROM）、电可擦除可编程只读存储器（EEPROM）、可编程只读存储器（PROM）、磁性存储器、磁盘、光盘、服务器、App 应用商城等等，其上存储有计算机程序，程序被处理器 22 执行时实现相应功能。本实施例的计算机可读存储介质用于存储数据分析装置，被处理器 22 执行时实现实施例一的数据分析方法。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。

以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

1. 一种基于大数据的数据分析方法，其中，包括：

创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

2. 根据权利要求 1 所述的数据分析方法，其中，所述创建用于描述节点间关联关系及资产关系的有向图包括：

从服务系统中获取节点数据以及各节点之间的关联关系，根据所述关联关系构建用于描述节点间关联关系的有向图；

计算所述有向图中相互关联的节点之间的资产关联比值，并将其加载于所述相互关联节点之间的路径上以描述所述有向图中节点间资产关系。

3. 根据权利要求 1 所述的数据分析方法，其中，所述通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型包括：

通过传染病模型计算所述有向图的模型指标以获得平均传导概率；

将所述平均传导概率分别与所述有向图的路径上的资产关联比值相乘以获得初级传导系数，通过随机森林模型识别所述有向图中各相互关联节点之间的相关度，并将其加载于所述相互关联的节点的路径上；

将所述有向图的路径上的初级传导系数和相关度相乘以获得风险传导系数，将所述风险传导系数加载于所述有向图的路径上以获得无标度模型。

4. 根据权利要求 1 所述的数据分析方法，其中，所述识别所述无标度模型中的感染节点包括：

判断所述无标度模型的节点是否属于预设的黑名单；

将属于所述黑名单的节点设为感染节点。

5. 根据权利要求 1 所述的数据分析方法，其中，所述根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率包括：

在所述无标度模型中识别与所述感染节点连续关联的连续路径，并以所述感染节点为起始依次对所述连续路径的子路径进行编号；其中，所述连续路径是指在所述无标度模型中以感染节点为输出源，依次与其串联的关联节点的整条路径，所述子路径为所述连续路径中相邻的两个节点之间的路径；

将所述连续路径上的任一子路径设为待算路径，识别编号小于所述待算路径的子路径并提取其中的风险传导系数后汇总形成系数集，将所述系数集中的风险传导系数与所述待算路径的风险传导系数相乘获得真实传导系数，并将其加载于所述待算路径上。

6. 根据权利要求 1 所述的数据分析方法，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率包括：

将与所述节点请求匹配的节点设为目标节点，在所述无标度模型中提取与所述目标节点连接的路径；

将指向所述目标节点的路径设为传入路径，将从所述目标节点指出的路径设为传

出路径；

通过加权调整公式计算所述目标节点的传入路径的风险传导率，以获得传入风险率，通过加权调整公式计算所述目标节点的传出路径的风险传导率，以获得传出风险率。

7. 根据权利要求 1 所述的数据分析方法，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率之后包括：

将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端。

8. 一种基于大数据的数据分析装置，其中，包括：

创建服务器，用于创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

风险服务器，用于识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

计算服务器，用于根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

9. 一种计算机系统，其包括多个计算机设备，各计算机设备包括存储器、处理器以及存储在存储器上并可在处理器上运行的计算机程序，其中，所述多个计算机设备的处理器执行所述计算机程序时共同实现以下步骤：

创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

10. 根据权利要求 9 所述的计算机系统，其中，所述创建用于描述节点间关联关系及资产关系的有向图包括：

从服务系统中获取节点数据以及各节点之间的关联关系，根据所述关联关系构建用于描述节点间关联关系的有向图；

计算所述有向图中相互关联的节点之间的资产关联比值，并将其加载于所述相互关联节点之间的路径上以描述所述有向图中节点间资产关系。

11. 根据权利要求 9 所述的计算机系统，其中，所述通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型包括：

通过传染病模型计算所述有向图的模型指标以获得平均传导概率；

将所述平均传导概率分别与所述有向图的路径上的资产关联比值相乘以获得初级传导系数，通过随机森林模型识别所述有向图中各相互关联节点之间的相关度，并将其加载于所述相互关联的节点的路径上；

将所述有向图的路径上的初级传导系数和相关度相乘以获得风险传导系数，将所述风险传导系数加载于所述有向图的路径上以获得无标度模型。

12. 根据权利要求 9 所述的计算机系统，其中，所述识别所述无标度模型中的感染节点包括：

判断所述无标度模型的节点是否属于预设的黑名单；

将属于所述黑名单的节点设为感染节点；

所述根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率包括：

在所述无标度模型中识别与所述感染节点连续关联的连续路径，并以所述感染节点为起始依次对所述连续路径的子路径进行编号；其中，所述连续路径是指在所述无标度模型中以感染节点为输出源，依次与其串联的关联节点的整条路径，所述子路径为所述连续路径中相邻的两个节点之间的路径；

将所述连续路径上的任一子路径设为待算路径，识别编号小于所述待算路径的子路径并提取其中的风险传导系数后汇总形成系数集，将所述系数集中的风险传导系数与所述待算路径的风险传导系数相乘获得真实传导系数，并将其加载于所述待算路径上。

13. 根据权利要求 9 所述的计算机系统，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率包括：

将与所述节点请求匹配的节点设为目标节点，在所述无标度模型中提取与所述目标节点连接的路径；

将指向所述目标节点的路径设为传入路径，将从所述目标节点指出的路径设为传出路径；

通过加权调整公式计算所述目标节点的传入路径的风险传导率，以获得传入风险率，通过加权调整公式计算所述目标节点的传出路径的风险传导率，以获得传出风险率。

14. 根据权利要求 9 所述的计算机系统，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率之后包括：

将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端。

15. 一种计算机可读存储介质，其包括多个存储介质，各存储介质上存储有计算机程序，其中，所述多个存储介质存储的所述计算机程序被处理器执行时共同实现以下步骤：

创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器；其中，所述节点是指信息所有者，所述关联关系用于反应各信息所有者之间的牵连和影响，所述资产关系用于反应各信息所有者之间的资产关联比值；

所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器；

所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率；其中，所述节点请求包括有与风险感染模型中节点对应的节点名称，其用于提取所述风险感染模型中的节点。

16. 根据权利要求 15 所述的计算机可读存储介质，其中，所述创建用于描述节点间关联关系及资产关系的有向图包括：

从服务系统中获取节点数据以及各节点之间的关联关系，根据所述关联关系构建

用于描述节点间关联关系的有向图；

计算所述有向图中相互关联的节点之间的资产关联比值，并将其加载于所述相互关联节点之间的路径上以描述所述有向图中节点间资产关系。

17. 根据权利要求 15 所述的计算机可读存储介质，其中，所述通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型包括：

通过传染病模型计算所述有向图的模型指标以获得平均传导概率；

将所述平均传导概率分别与所述有向图的路径上的资产关联比值相乘以获得初级传导系数，通过随机森林模型识别所述有向图中各相互关联节点之间的相关度，并将其加载于所述相互关联的节点的路径上；

将所述有向图的路径上的初级传导系数和相关度相乘以获得风险传导系数，将所述风险传导系数加载于所述有向图的路径上以获得无标度模型。

18. 根据权利要求 15 所述的计算机可读存储介质，其中，所述识别所述无标度模型中的感染节点包括：

判断所述无标度模型的节点是否属于预设的黑名单；

将属于所述黑名单的节点设为感染节点；

所述根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率包括：

在所述无标度模型中识别与所述感染节点连续关联的连续路径，并以所述感染节点为起始依次对所述连续路径的子路径进行编号；其中，所述连续路径是指在所述无标度模型中以感染节点为输出源，依次与其串联的关联节点的整条路径，所述子路径为所述连续路径中相邻的两个节点之间的路径；

将所述连续路径上的任一子路径设为待算路径，识别编号小于所述待算路径的子路径并提取其中的风险传导系数后汇总形成系数集，将所述系数集中的风险传导系数与所述待算路径的风险传导系数相乘获得真实传导系数，并将其加载于所述待算路径上。

19. 根据权利要求 15 所述的计算机可读存储介质，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率包括：

将与所述节点请求匹配的节点设为目标节点，在所述无标度模型中提取与所述目标节点连接的路径；

将指向所述目标节点的路径设为传入路径，将从所述目标节点指出的路径设为传出路径；

通过加权调整公式计算所述目标节点的传入路径的风险传导率，以获得传入风险率，通过加权调整公式计算所述目标节点的传出路径的风险传导率，以获得传出风险率。

20. 根据权利要求 15 所述的计算机可读存储介质，其中，所述计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率之后包括：

将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别所述风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端。

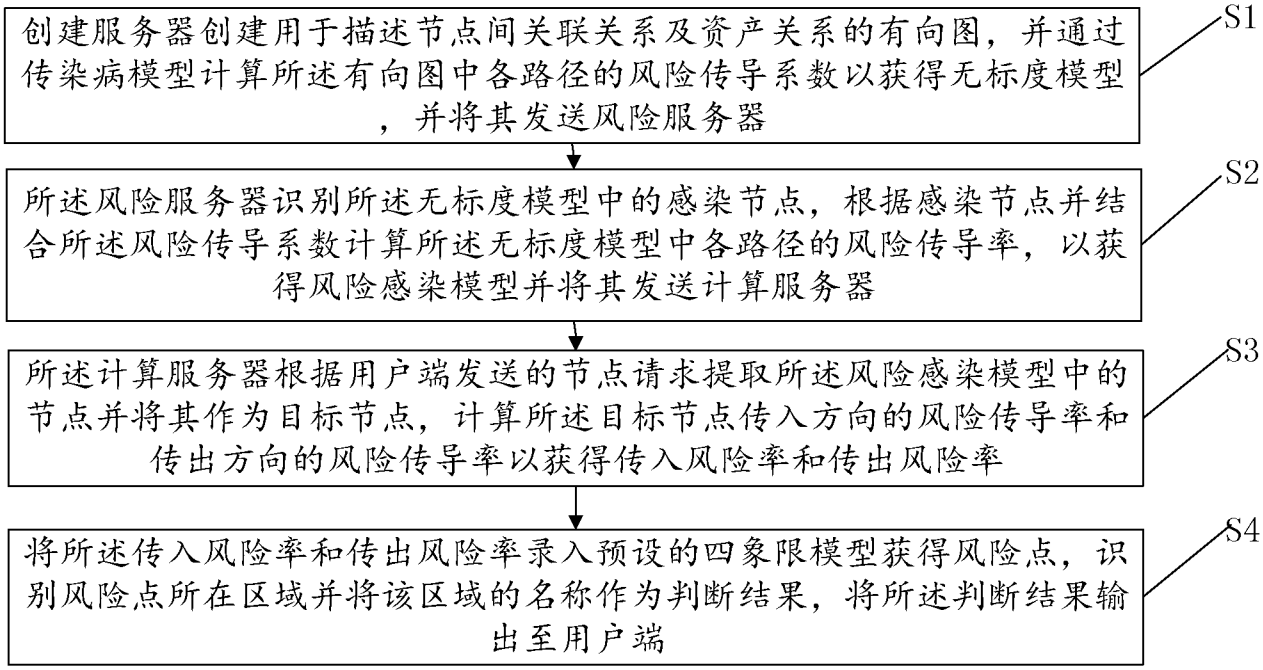
创建服务器创建用于描述节点间关联关系及资产关系的有向图，并通过传染病模型计算所述有向图中各路径的风险传导系数以获得无标度模型，并将其发送风险服务器
S1
所述风险服务器识别所述无标度模型中的感染节点，根据感染节点并结合所述风险传导系数计算所述无标度模型中各路径的风险传导率，以获得风险感染模型并将其发送计算服务器
S2
所述计算服务器根据用户端发送的节点请求提取所述风险感染模型中的节点并将其作为目标节点，计算所述目标节点传入方向的风险传导率和传出方向的风险传导率以获得传入风险率和传出风险率
S3
将所述传入风险率和传出风险率录入预设的四象限模型获得风险点，识别风险点所在区域并将该区域的名称作为判断结果，将所述判断结果输出至用户端
S4

图 1

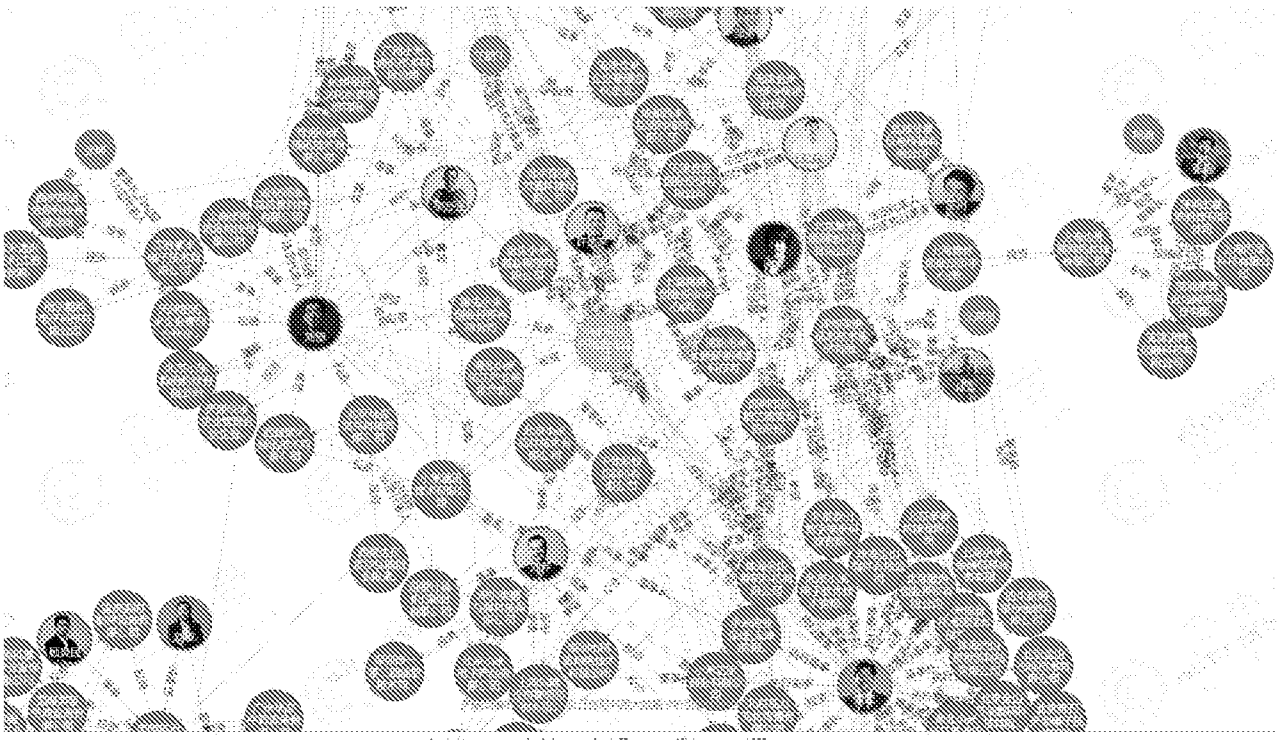

图 2

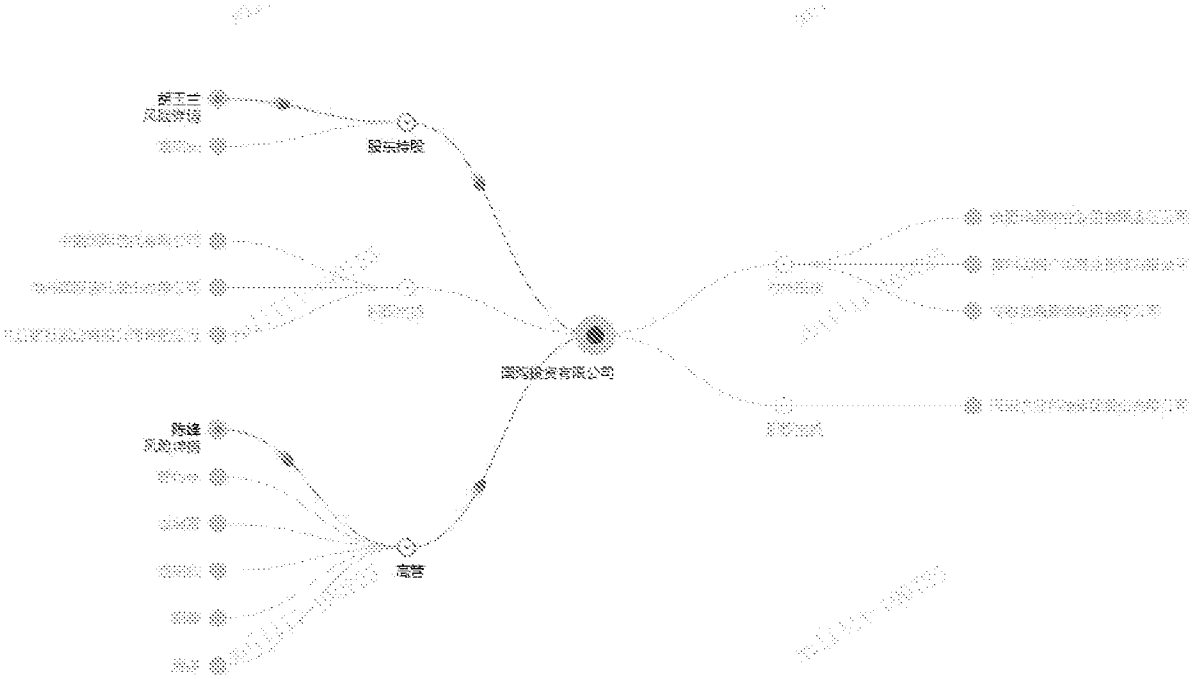

图 3

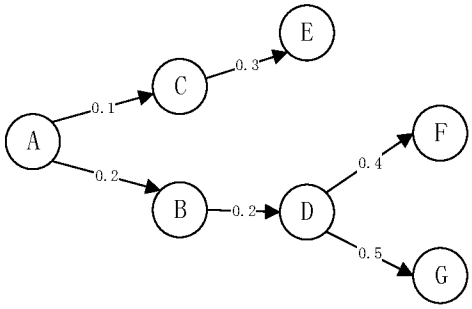
A
B
C
D
E
F
G
0.1
0.2
0.3
0.2
0.4
0.5

图 4

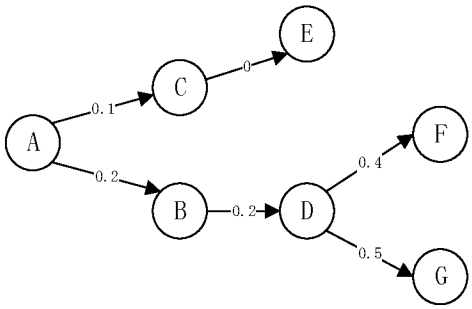
A
B
C
D
E
F
G
0.1
0.2
0
0.2
0.4
0.5

图 5

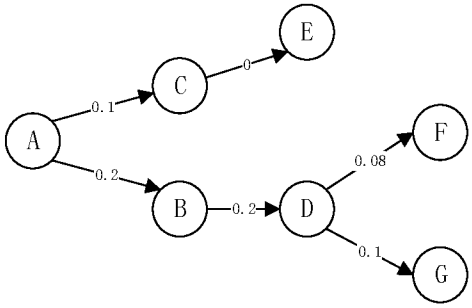
A
B
C
D
E
F
G
0.1
0.2
0
0.2
0.08
0.1

图 6


Ola 风险传导生态系统
高
周边风险传入
低
公司风险传出
高
活跃火山

图 7


创建服务器11
风险服务器12
计算服务器13
判断服务器14
风险识别装置1

图 8

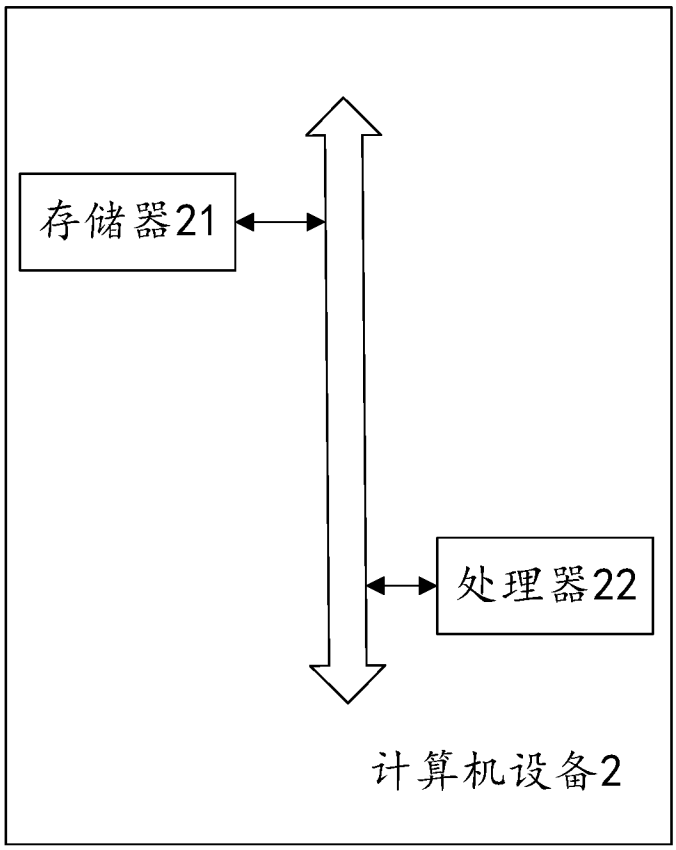

图 9

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/093201

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)i; G06F 16/901(2019.01)i; G06F 16/906(2019.01)i; G06F 16/36(2019.01)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; G06F

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)

DWPI; SIPOABS; CNABS; CNKI; CNTXT: 传染病 模型 风险 企业 公司 随机森林 SIS SIR RANDOM FOREST INFECTIOUS DISEASE RISK+ ENTERPRISE? COMPAN+ MODEL

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CN 109472485 A (CHENGDU BUSINESS BIG DATA TECHNOLOGY CO., LTD.) 15 March 2019 (2019-03-15) description, paragraphs [0030]-[0070], and figures 1-6	1-20
Y	CN 108090709 A (CHONGQING YUCUN BIG DATA TECHNOLOGY CO., LTD.) 29 May 2018 (2018-05-29) description, paragraphs [0022]-[0039]	1-20
A	CN 106204264 A (TIANYUN RONGCHUANG DATA TECHNOLOGY (BEIJING) CO., LTD.) 07 December 2016 (2016-12-07) entire document	1-20
A	CN 107563645 A (HANGZHOU YUNSUAN XINDA DATA TECHNOLOGY CO., LTD.) 09 January 2018 (2018-01-09) entire document	1-20
A	CN 109949164 A (SUN YAT-SEN UNIVERSITY) 28 June 2019 (2019-06-28) entire document	1-20
A	US 2010198631 A1 (BANK OF AMERICA CORP) 05 August 2010 (2010-08-05) entire document	1-20

☐ 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
24 October 2020	**29 October 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/CN2020/093201

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 109472485 A	15 March 2019	None	
CN 108090709 A	29 May 2018	None	
CN 106204264 A	07 December 2016	None	
CN 107563645 A	09 January 2018	None	
CN 109949164 A	28 June 2019	None	
US 2010198631 A1	05 August 2010	US 8185430 B2	22 May 2012
		EP 2226750 A1	08 September 2010
		HK 1147831 A0	19 August 2011

国际检索报告

国际申请号
PCT/CN2020/093201

A. 主题的分类

G06Q 10/06(2012.01)i; G06F 16/901(2019.01)i; G06F 16/906(2019.01)i; G06F 16/36(2019.01)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; G06F

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））

DWPI; SIPOABS; CNABS; CNKI; CNTXT: 传染病 模型 风险 企业 公司 随机森林 SIS SIR RANDOM FOREST INFECTIOUS DISEASE RISK+ ENTERPRISE? COMPAN+ MODEL

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
Y	CN 109472485 A (成都数联铭品科技有限公司) 2019年 3月 15日 (2019-03-15) 说明书第[0030]-[0070]段，以及图1-6	1-20
Y	CN 108090709 A (重庆誉存大数据科技有限公司) 2018年 5月 29日 (2018-05-29) 说明书第[0022]-[0039]段	1-20
A	CN 106204264 A (天云融创数据科技北京有限公司) 2016年 12月 7日 (2016-12-07) 全文	1-20
A	CN 107563645 A (杭州云算信达数据技术有限公司) 2018年 1月 9日 (2018-01-09) 全文	1-20
A	CN 109949164 A (中山大学) 2019年 6月 28日 (2019-06-28) 全文	1-20
A	US 2010198631 A1 (BANK OF AMERICA CORP) 2010年 8月 5日 (2010-08-05) 全文	1-20

☐ 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 10月 24日	2020年 10月 29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 胡燕 电话号码 62089101

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2020/093201

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	109472485	A	2019年 3月 15日	无			
CN	108090709	A	2018年 5月 29日	无			
CN	106204264	A	2016年 12月 7日	无			
CN	107563645	A	2018年 1月 9日	无			
CN	109949164	A	2019年 6月 28日	无			
US	2010198631	A1	2010年 8月 5日	US	8185430	B2	2012年 5月 22日
				EP	2226750	A1	2010年 9月 8日
				HK	1147831	A0	2011年 8月 19日