



(12) 发明专利申请

(10) 申请公布号 CN 117522566 A

(43) 申请公布日 2024. 02. 06

(21) 申请号 202311634725.1

(22) 申请日 2023.11.30

(71) 申请人 中银金融科技有限公司

地址 200120 上海市浦东新区(上海)自由贸易试验区银城中路200号4楼408室

(72) 发明人 魏思雨

(74) 专利代理机构 北京路浩知识产权代理有限公司 11002

专利代理师 李鑫

(51) Int. Cl.

G06Q 40/03 (2023.01)

G06F 18/23 (2023.01)

G06F 18/27 (2023.01)

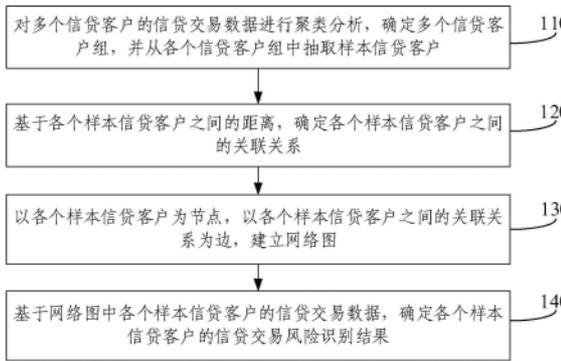
权利要求书2页 说明书12页 附图2页

(54) 发明名称

信贷交易风险识别方法、装置、电子设备和存储介质

(57) 摘要

本申请提供了一种信贷交易风险识别方法、装置、电子设备和存储介质,其中方法包括:对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。本申请提供的方法和装置,提高了信贷交易风险的识别效率,提高了信贷交易风险的识别准确性。



1. 一种信贷交易风险识别方法,其特征在于,包括:

对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;

基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;

以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;

基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

2. 根据权利要求1所述的信贷交易风险识别方法,其特征在于,所述基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系,包括:

确定预设距离阈值以及各个样本信贷客户之间的距离;

将各个样本信贷客户之间的距离与所述预设距离阈值进行比较;

在所述距离小于所述预设距离阈值的情况下,确定所述距离对应的两个样本信贷客户的关联关系数值为第一数值;

在所述距离大于或者等于所述预设距离阈值的情况下,确定所述距离对应的两个样本信贷客户的关联关系数值为第二数值;

基于各个样本信贷客户的关联关系数值构建邻接矩阵;所述邻接矩阵用于表示各个样本信贷客户的关联关系。

3. 根据权利要求2所述的信贷交易风险识别方法,其特征在于,所述基于各个样本信贷客户的关联关系数值构建邻接矩阵之后,所述方法还包括:

基于所述邻接矩阵中各个样本信贷客户对应的第一数值和第二数值的数量,确定所述网络图的网络粘性。

4. 根据权利要求3所述的信贷交易风险识别方法,其特征在于,所述基于所述邻接矩阵确定所述网络图的网络粘性之后,所述方法还包括:

基于各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的节点属性;

基于所述邻接矩阵和各个样本信贷客户的节点属性,建立网络自回归模型;所述网络自回归模型用于对所述网络图的网络结构进行预测。

5. 根据权利要求1所述的信贷交易风险识别方法,其特征在于,所述以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图之后,所述方法还包括:

基于力导向图、桑基图和树型图中的至少一种,对所述网络图进行可视化显示。

6. 根据权利要求1所述的信贷交易风险识别方法,其特征在于,所述基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果之后,所述方法还包括:

在所述样本信贷客户的信贷交易风险识别结果为异常的情况下,将所述样本信贷客户所属的信贷客户组中的各个信贷客户确定为候选异常信贷客户;

对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户。

7. 根据权利要求1所述的信贷交易风险识别方法,其特征在于,所述对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户,包括:

获取各个候选异常信贷客户的社交媒体数据；

将各个候选异常信贷客户的信贷交易数据和/或社交媒体数据输入信贷交易风险识别模型,得到所述信贷交易风险识别模型输出的各个候选异常信贷客户的信贷交易风险识别结果；

将所述信贷交易风险识别结果为异常的候选异常信贷客户,确定出为异常信贷客户；

其中,所述信贷交易风险识别模型是以神经网络模型为初始模型,基于多个样本异常信贷客户的信贷交易数据和/或社交媒体数据对所述初始模型进行训练后得到的。

8.一种信贷交易风险识别装置,其特征在于,包括:

聚类单元,用于对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户；

确定单元,用于基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系；

建图单元,用于以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图；

识别单元,用于基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

9.一种电子设备,包括存储器和处理器,其特征在于,所述存储器中存储有计算机程序,所述处理器被设置为通过所述计算机程序执行权利要求1至7任一项所述的信贷交易风险识别方法。

10.一种计算机可读的存储介质,其特征在于,所述计算机可读的存储介质包括存储的程序,其中,所述程序运行时执行权利要求1至7任一项所述的信贷交易风险识别方法。

## 信贷交易风险识别方法、装置、电子设备和存储介质

### 技术领域

[0001] 本申请涉及计算机技术领域,具体而言,涉及一种信贷交易风险识别方法、装置、电子设备和存储介质。

### 背景技术

[0002] 在银行办理客户提出的信贷申请(例如信用卡业务)时,通常需要对客户进行风险识别。现有技术中,通常是通过人工对庞大的信贷客户数据进行分析,信贷交易风险的识别效率低,识别准确性差。

[0003] 因此,如何提高信贷交易风险的识别效率和识别准确性成为业界亟待解决的技术问题。

### 发明内容

[0004] 本申请提供一种信贷交易风险识别方法、装置、电子设备和存储介质,用于解决如何提高信贷交易风险的识别效率和识别准确性的技术问题。

[0005] 本申请提供一种信贷交易风险识别方法,包括:

[0006] 对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;

[0007] 基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;

[0008] 以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;

[0009] 基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

[0010] 在一些实施例中,所述基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系,包括:

[0011] 确定预设距离阈值以及各个样本信贷客户之间的距离;

[0012] 将各个样本信贷客户之间的距离与所述预设距离阈值进行比较;

[0013] 在所述距离小于所述预设距离阈值的情况下,确定所述距离对应的两个样本信贷客户的关联关系数值为第一数值;

[0014] 在所述距离大于或者等于所述预设距离阈值的情况下,确定所述距离对应的两个样本信贷客户的关联关系数值为第二数值;

[0015] 基于各个样本信贷客户的关联关系数值构建邻接矩阵;所述邻接矩阵用于表示各个样本信贷客户的关联关系。

[0016] 在一些实施例中,所述基于各个样本信贷客户的关联关系数值构建邻接矩阵之后,所述方法还包括:

[0017] 基于所述邻接矩阵中各个样本信贷客户对应的第一数值和第二数值的数量,确定所述网络图的网络粘性。

[0018] 在一些实施例中,所述基于所述邻接矩阵确定所述网络图的网络粘性之后,所述方法还包括:

[0019] 基于各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的节点属性;

[0020] 基于所述邻接矩阵和各个样本信贷客户的节点属性,建立网络自回归模型;所述网络自回归模型用于对所述网络图的网络结构进行预测。

[0021] 在一些实施例中,所述以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图之后,所述方法还包括:

[0022] 基于力导向图、桑基图和树型图中的至少一种,对所述网络图进行可视化显示。

[0023] 在一些实施例中,所述基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果之后,所述方法还包括:

[0024] 在所述样本信贷客户的信贷交易风险识别结果为异常的情况下,将所述样本信贷客户所属的信贷客户组中的各个信贷客户确定为候选异常信贷客户;

[0025] 对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户。

[0026] 在一些实施例中,所述对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户,包括:

[0027] 获取各个候选异常信贷客户的社交媒体数据;

[0028] 将各个候选异常信贷客户的信贷交易数据和/或社交媒体数据输入信贷交易风险识别模型,得到所述信贷交易风险识别模型输出的各个候选异常信贷客户的信贷交易风险识别结果;

[0029] 将所述信贷交易风险识别结果为异常的候选异常信贷客户,确定出为异常信贷客户;

[0030] 其中,所述信贷交易风险识别模型是以神经网络模型为初始模型,基于多个样本异常信贷客户的信贷交易数据和/或社交媒体数据对所述初始模型进行训练后得到的。

[0031] 本申请提供一种信贷交易风险识别装置,包括:

[0032] 聚类单元,用于对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;

[0033] 确定单元,用于基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;

[0034] 建图单元,用于以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;

[0035] 识别单元,用于基于所述网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

[0036] 本申请提供一种电子设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器被设置为通过所述计算机程序执行所述的信贷交易风险识别方法。

[0037] 本申请提供一种计算机可读的存储介质,所述计算机可读的存储介质包括存储的程序,其中,所述程序运行时执行所述的信贷交易风险识别方法。

[0038] 本申请提供的信贷交易风险识别方法、装置、电子设备和存储介质,对多个信贷客

户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果;由于采用了聚类分析和分组抽取样本的方法,实现了对庞大的信贷客户数据进行分析,由于采用了构建网络图进行信贷交易分析的方法,使得可以从样本信贷客户之间的关联关系出发,挖掘样本信贷客户之间潜在的信贷交易风险,缩小了异常信贷客户的识别范围,提高了信贷交易风险的识别效率,提高了信贷交易风险的识别准确性。

### 附图说明

[0039] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本申请的实施例,并与说明书一起用于解释本申请的原理。

[0040] 为了更清楚地说明本申请或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图是本申请的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0041] 图1为本申请提供的信贷交易风险识别方法的流程示意图;

[0042] 图2为本申请提供的信贷交易风险识别装置的结构示意图;

[0043] 图3是本申请提供的电子设备的结构示意图。

### 具体实施方式

[0044] 为了使本技术领域的人员更好地理解本申请方案,下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本申请一部分的实施例,而不是全部的实施例。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,都应当属于本申请保护的范围。

[0045] 需要说明的是,本申请中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本申请的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元或模块的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元或模块,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元或模块。

[0046] 在本申请的技术方案中,所涉及的客户个人信息的收集、存储、使用、加工、传输、提供和公开等处理,均符合相关法律法规的规定,采取了必要保密措施,且不违背公序良俗。

[0047] 图1为本申请提供的信贷交易风险识别方法的流程示意图,如图1所示,该方法包括步骤110、步骤120、步骤130和步骤140。

[0048] 步骤110、对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,

并从各个信贷客户组中抽取样本信贷客户。

[0049] 具体地,本申请实施例提供的信贷交易风险识别方法的执行主体为信贷交易风险识别装置。该装置可以通过软件实现,例如运行在计算机中的信贷交易风险识别程序;也可以为执行信贷交易风险识别方法的装置,例如移动终端、平板电脑、台式计算机或者服务器等。

[0050] 信贷客户是指需要借款或者信用支持的个人或者企业。信贷交易数据是指包括贷款金额、贷款期限、贷款利率、还款记录等信息在内的与信贷交易相关的数据。这些数据通常记录在银行、金融机构等信贷机构的数据库中,用于评估客户信用风险、制定贷款策略和进行风险控制等。

[0051] 聚类分析是一种数据分析方法,它将数据集中的对象划分为若干个具有相似特征的组,每个组内的对象相互之间的相似度较高,而不同组之间的对象则具有较大的差异性。在金融领域中,聚类分析常常被用于对客户群体进行分类,以便更好地理解客户特征、行为和需求,从而制定针对性的营销策略、产品设计和风险管理措施。

[0052] 因此,可以采用聚类分析算法,对多个信贷客户的信贷交易数据进行处理,将多个信贷客户分为多个信贷客户组。每个信贷客户组中的各个信贷客户具有较高的相似度。信贷交易数据为混合型变量,即包括连续型变量,又包括名义型变量(按货币单位衡量的变量),因此可以聚类分析算法可以采用基于高尔(Gower)距离算法、基于围绕中心点的划分(Partitioning Around Medoid, PAM)聚类算法和基于轮廓系数(Silhouette Coefficient)算法中的至少一种。

[0053] 在每个信贷客户组中可以抽取至少一个样本作为样本信贷客户或者按照预设百分比抽取样本作为样本信贷客户,将其作为代表对每个信贷客户组进行分析。抽取的方法可以采用随机抽取,或者根据各个样本在信贷客户组中的位置进行抽取。例如,可以计算聚类后每个信贷客户组的聚类中心,计算每个信贷客户组中各个信贷客户距离聚类中心的距离,根据该距离对各个信贷客户进行升序排列,按照排列序号选择预设数量的信贷客户作为样本信贷客户。

[0054] 步骤120、基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系。

[0055] 具体地,在抽取样本信贷客户之后,可以计算各个样本信贷客户之间的距离,根据这些距离判断各个样本信贷客户之间的关联关系。此处的关联关系可以为信贷客户在信贷交易中存在的业务联系。如果两个样本信贷客户之间的距离越小,则表明这两个样本信贷客户之间越可能存在关联关系;如果两个样本信贷客户之间的距离越大,则表明这两个样本信贷客户之间越不可能存在关联关系。

[0056] 步骤130、以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图。

[0057] 具体地,图计算在大数据时代应运而生,它是研究事物之间的复杂关联关系,并对其描述、刻画、分析和计算的技术。网络图可以对万物进行关联关系展示,例如样本信贷客户之间的关联关系。

[0058] 网络图主要是由节点(Vertex)和边(Edge)构成的。节点表示事物,边表示事物之间的关系。可以以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建

立网络图。该网络图可以用于对信贷交易风险进行识别。

[0059] 步骤140、基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

[0060] 具体地,在网络图中可以根据相互紧密联系的样本信贷客户的信贷交易数据,对各个样本信贷客户的信贷交易风险进行识别,得到各个样本信贷客户的信贷交易风险识别结果。

[0061] 例如,可以在网络图中确定具有紧密连接联系的样本信贷客户,对这些样本信贷客户的信贷交易数据进行分析,如果任一样本信贷客户的信贷交易数据存在异常信贷交易,则可以认为该样本信贷客户的信贷交易风险识别结果为异常,同时将与该样本信贷客户紧密连接的各个样本信贷客户的信贷交易风险识别结果也确定为异常(该样本信贷客户为异常信贷客户)。可以通过人工核验的方式对信贷交易风险识别结果进行核验。

[0062] 又例如,社区为网络图中紧密连接的节点集合。社区内节点之间的联系比较强,而不同社区之间的节点的联系相对较弱。可以采用社区检测算法(Community Detection,又称社区发现算法)对网络图进行社区检测,用于将网络图分成多个社区。社区检测算法可以包括基于谱聚类进行社区检测的算法、基于层次聚类进行社区检测的算法,以及基于模块度(Modularity)进行社区检测的算法等。常用的社区检测算法有Girvan-Newman算法、Louvain算法、Clauset-Newman-Moore算法、标签传播算法等。可以在网络图中采用社区检测算法,将整个网络图分为多个社区,在同一个社区中的多个样本信贷客户具有更加紧密的连接关系,在各个社区内对各个样本信贷客户的信贷交易数据进行精确地分析,从而确定各个样本信贷客户的信贷交易风险识别结果。

[0063] 在确定各个样本信贷客户的信贷交易风险识别结果后,根据同一信贷客户组中各个信贷客户具有相似性,可以确定各个样本信贷客户所属信贷客户组中各个信贷客户的信贷交易风险识别结果。

[0064] 本申请实施例提供的信贷交易风险识别方法,对多个信贷客户的信贷交易数据进行分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果;由于采用了聚类分析和分组抽取样本的方法,实现了对庞大的信贷客户数据进行分析,由于采用了构建网络图进行信贷交易分析的方法,使得可以从样本信贷客户之间的关联关系出发,挖掘样本信贷客户之间潜在的信贷交易风险,缩小了异常信贷客户的识别范围,提高了信贷交易风险的识别效率,提高了信贷交易风险的识别准确性。

[0065] 需要说明的是,本申请每一个实施方式可以自由组合、调换顺序或者单独执行,并不需要依靠或依赖固定的执行顺序。

[0066] 在一些实施例中,步骤120包括:

[0067] 确定预设距离阈值以及各个样本信贷客户之间的距离;

[0068] 将各个样本信贷客户之间的距离与预设距离阈值进行比较;

[0069] 在距离小于预设距离阈值的情况下,确定距离对应的两个样本信贷客户的关联关系数值为第一数值;

[0070] 在距离大于或者等于预设距离阈值的情况下,确定距离对应的两个样本信贷客户的关联关系数值为第二数值;

[0071] 基于各个样本信贷客户的关联关系数值构建邻接矩阵;邻接矩阵用于表示各个样本信贷客户的关联关系。

[0072] 具体地,可以通过高尔距离算法或者其他距离算法计算各个样本信贷客户之间的距离,并根据实际情况设置预设距离阈值。

[0073] 将各个样本信贷客户之间的距离与预设距离阈值进行比较,如果距离小于预设距离阈值,则表明两个样本信贷客户之间可能存在关联关系。关联关系的强弱可以用关联关系数值来表示。例如第一数值可以用来表示具有较强的关联关系,第二数值可以用来表示具有较弱的关联关系。第一数值可以取“1”,第二数值可以取“0”。因此,两个样本信贷客户之间可能存在关联关系,关联关系数值可以用“1”来表示。如果距离大于或者等于预设距离阈值,则表明两个样本信贷客户之间可能不存在关联关系,关联关系数值可以用“0”来表示。

[0074] 如果用一个二维数组来表示各个样本信贷客户之间的关联关系,则关联关系数值则为该二维数组中的元素值,因此可以得到表示各个样本信贷客户的关联关系的邻接矩阵。

[0075] 预设距离阈值如果设置得较小,则邻接矩阵中的第一数值则较少,在后期生成网络图时,图中的边(连接)较少,这会使得网络图中样本信贷客户之间的关联关系较少,可以避免网络图出现视觉混乱,可以突出有密切关系的客户;预设距离阈值如果设置得较大,则邻接矩阵中的第一数值则较多,在后期生成网络图时,图中的边(连接)较多,这会使得网络图中样本信贷客户之间的关联关系较多,可能在网络图上出现视觉混乱,不能突出有密切关系的客户。因此,需要合理地设置预设距离阈值。

[0076] 此外,在信贷客户组中抽取样本的预设百分比,可以确定网络图中的节点数量,通过合理地设置预设百分比,可以使得后期生成的网络图大小合适,能够取得较好的可视化效果。

[0077] 本申请实施例提供的信贷交易风险识别方法,可以设置预设距离阈值,将各个样本信贷客户之间的距离与预设距离阈值进行比较,生成邻接矩阵用于表示各个样本信贷客户的关联关系,便于通过邻接矩阵进行网络图中各个节点的连接分析,提高了信贷交易风险的识别效率。

[0078] 在一些实施例中,基于各个样本信贷客户的关联关系数值构建邻接矩阵之后,方法还包括:

[0079] 基于邻接矩阵中各个样本信贷客户对应的第一数值和第二数值的数量,确定网络图的网络粘性。

[0080] 具体地,网络粘性是指在网络图中节点之间的联系紧密程度。

[0081] 根据邻接矩阵确定网络图的网络粘性的步骤可以包括:一、构建邻接矩阵;根据网络图的边列表或连接关系,构建一个邻接矩阵。矩阵的行和列分别代表网络中的节点,矩阵中的元素表示节点之间的连接强度;二、归一化邻接矩阵;可以对邻接矩阵进行归一化,使得每个元素都在0到1之间。常见的归一化方法包括将每一行除以该行的度数,或者将整个矩阵除以其最大元素;三、计算网络粘性;粘性可以通过不同的方式来衡量,常见的指标包

括节点的平均度数、网络的平均连接强度、聚类系数等。本申请实施例中网络粘性可以根据各个样本信贷客户对应的第一数值和第二数值的数量比例确定。例如可以将第一数值“1”的数量占第一数值“1”和第二数值“0”的总数量的比例,作为网络粘性。

[0082] 此外,还可以计算各个样本信贷客户的关系密度,即图网络中平均每个样本信贷客户所认识其他信贷客户的客户数量。

[0083] 通过网络粘性和关系密度,可以对网络图进行分析和研究。例如可以假定网络图中的网格大小,以进行信贷客户网络关系的定量研究。在控制其他条件不变的情况下,研究网络大小对总体网络粘性及估计值大小的影响;在控制其他条件不变的情况下,研究个体之间关系密度大小对总体网络粘性及估计值大小的影响。

[0084] 在一些实施例中,基于邻接矩阵确定网络图的网络粘性之后,方法还包括:

[0085] 基于各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的节点属性;

[0086] 基于邻接矩阵和各个样本信贷客户的节点属性,建立网络自回归模型;网络自回归模型用于对网络图的网络结构进行预测。

[0087] 具体地,在对网络图进行分析时,通常需要考虑节点属性和节点之间的关联关系。节点属性为与节点相关联的特征或者属性,例如节点属性可以包括标识属性、类别属性和数值属性等。例如运用社会网络分析方法对网络图进行分析时,通常需要的数据包括“source”,“target”,“value”,即从第*i*个节点到第*j*个节点的值“*i*”和“*j*”,边的权重,“name”,“group”,“size”,即每个节点的名字,所在的类别序号,节点的大小。对于样本信贷客户对应的节点,节点属性可以为样本信贷客户所能贷款的总额。边的粗细代表信贷客户间联系的紧密程度。

[0088] 可以构建用于对网络图的网络结构进行预测的网络自回归模型。网络自回归模型(Network Autoregressive Model)是一种用于建模和预测动态网络演化的方法。它基于过去的网络结构和节点属性信息,利用自回归模型来预测未来的网络演化。可以将不同时间对应的邻接矩阵和各个样本信贷客户的节点属性输入网络自回归模型,由网络自回归模型通过学习过去的网络结构和节点属性信息,预测未来网络的演化趋势。此外,当网络数据丢失或不完整时,网络自回归模型可以利用已有的数据来重构和恢复缺失的网络结构。通过模型的预测能力,可以填补数据缺失的部分,从而实现对网络的重构和恢复。

[0089] 本申请实施例提供的信贷交易风险识别方法,通过构建建立网络自回归模型,对网络图的网络结构进行预测,使得在生成的网络图中能够挖掘样本信贷客户之间潜在的信贷交易风险,提高了信贷交易风险的识别效率,提高了信贷交易风险的识别准确性。

[0090] 在一些实施例中,步骤130之后还包括:

[0091] 基于力导向图、桑基图和树型图中的至少一种,对网络图进行可视化显示。

[0092] 具体地,力导向图(Force-directed graph)是一种网络图可视化方法。在力导向图中,每个节点表示网络中的一个实体或对象,每条边表示它们之间的连接或关系,可以用来显示复杂的网络划分关系。

[0093] 桑基图(Sankey diagram)是一种用于可视化流量、能量和资源等的图形表示方法。它通过图形化的方式展示了不同节点之间的数量关系和流动情况,从而方便更好地理解数据的分布和流向。在桑基图中,每个节点代表一个状态或过程,每条连线表示数据或物质在不同状态之间的转移和流动。

[0094] 树型图(Tree diagram)是一种用于表达层级结构信息的图形表示方法,它可以清晰地展示各个层级之间的父子关系和组成结构。在树型图中,整个结构类似于一棵树,每个节点代表一个层级或分类,而边则代表节点之间的关系。

[0095] 可以采用上述三种方式生成交互图,圆圈大小代表信贷客户的“size”,即“客户及家庭所能贷款总额”,边的粗细代表信贷客户间联系的紧密程度“value”,两种不同的圆圈颜色表示两种分类。同时通过参数(zoom)的设置,可以进行不复原的交互操作,拖动网络节点或边,可以自主变换布局,将这部分信贷客户间的联系展现得更明晰,更直观。同时通过参数(font size)的设置,将鼠标放在网络节点(圆圈)上,能显示出节点代表的信贷客户的标识,便于判断分析。

[0096] 本申请实施例提供的信贷交易风险识别方法,可以通过力导向图、桑基图和树型图中的至少一种对网络图进行可视化显示,便于直观地显示网络图,提高了信贷交易风险的识别效率。

[0097] 在一些实施例中,步骤140之后还包括:

[0098] 在样本信贷客户的信贷交易风险识别结果为异常的情况下,将样本信贷客户所属的信贷客户组中的各个信贷客户确定为候选异常信贷客户;

[0099] 对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户。

[0100] 具体地,在样本信贷客户的信贷交易风险识别结果为异常的情况下,可以将样本信贷客户所属的信贷客户组中的各个信贷客户确定为候选异常信贷客户。候选异常信贷客户不一定是真实的异常信贷客户,因此需要进行进一步地核验。

[0101] 可以采用人工核验或者计算机核验的方法,对各个候选异常信贷客户的信贷交易数据进行核验。例如可以对候选异常信贷客户进行回访等实现人工核验;也可以收集候选异常信贷客户的社交媒体数据进行分析,实现人工核验或者计算机核验。

[0102] 如果候选异常信贷客户的信贷交易数据核验结果为通过,则该候选异常信贷客户为正常信贷客户;如果候选异常信贷客户的信贷交易数据核验结果为不通过,则该候选异常信贷客户为异常信贷客户。

[0103] 本申请实施例提供的信贷交易风险识别方法,对各个候选异常信贷客户的信贷交易数据进行核验,提高了信贷交易风险的识别准确性。

[0104] 在一些实施例中,对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户,包括:

[0105] 获取各个候选异常信贷客户的社交媒体数据;

[0106] 将各个候选异常信贷客户的信贷交易数据和/或社交媒体数据输入信贷交易风险识别模型,得到信贷交易风险识别模型输出的各个候选异常信贷客户的信贷交易风险识别结果;

[0107] 将信贷交易风险识别结果为异常的候选异常信贷客户,确定出为异常信贷客户;

[0108] 其中,信贷交易风险识别模型是以神经网络模型为初始模型,基于多个样本异常信贷客户的信贷交易数据和/或社交媒体数据对初始模型进行训练后得到的。

[0109] 具体地,社交媒体数据是指信贷客户在社交媒体平台上生成的各种数字内容,包

括文本、图像、视频和语音等数据。通过对信贷客户的社交媒体数据进行分析,可以得到客户在社交媒体平台上的行为和兴趣。如果客户存在异常交易行为或者对异常交易行为感兴趣,则信贷客户进行异常信贷交易的风险较高。

[0110] 可以以神经网络模型为初始模型,训练后得到信贷交易风险识别模型,用于对各个候选异常信贷客户的信贷交易风险进行识别,得到候选异常信贷客户的信贷交易风险识别结果。

[0111] 信贷交易风险识别模型的预训练过程如下:

[0112] 首先,收集大量的信贷客户的信贷交易数据和/或社交媒体数据。这些信贷客户的样本标签为异常(存在异常信贷交易)。也就是说,这些信贷客户为样本异常信贷客户。其次,以神经网络模型为初始模型,将多个样本异常信贷客户的信贷交易数据和/或社交媒体数据输入至初始模型,由初始模型输出信贷交易风险预测结果。以信贷交易风险预测结果为预测值,以样本标签为实际值,根据预测值与实际值之间的差值对初始模型的参数进行调整,提高初始模型的预测能力,最终得到信贷交易风险识别模型。

[0113] 神经网络模型可以选择前馈神经网络、卷积神经网络和时序神经网络等,本申请实施例对神经网络模型的种类不做具体限定。

[0114] 本申请实施例提供的信贷交易风险识别方法,通过采用神经网络的方法对候选异常信贷客户的信贷交易风险进行识别,提高了信贷交易风险的识别准确性。

[0115] 在上述实施例的基础上,本申请实施例提供一种信贷交易风险识别方法,该方法主要用于识别银行信用卡业务中可能存在的交易风险,包括:

[0116] 步骤一、生成网络邻接矩阵

[0117] 通过分析信贷数据特点,对庞大的信贷数据进行聚类分析和分组抽取样本,计算各个样本之间的距离,根据计算所得的距离来判断信贷客户的连接关系,并生成邻接矩阵。

[0118] 步骤二、建立网络图

[0119] 以各个样本为节点,直接导入节点的属性。有了前述两个数据(节点属性和邻接矩阵),可以直接建立网络图。

[0120] 步骤三、建立网络自回归模型判断客户间的粘性大小

[0121] 通过邻接矩阵计算网络粘性以及生成网络自回归模型分析数据组的网络关系。

[0122] 步骤四、可视化显示网络图

[0123] 可以绘制可交互的网络图,进行可视化显示。

[0124] 下面对本申请实施例提供的装置进行描述,下文描述的装置与上文描述的方法可相互对应参照。

[0125] 图2为本申请提供的信贷交易风险识别装置的结构示意图,如图2所示,该装置包括:

[0126] 聚类单元210,用于对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;

[0127] 确定单元220,用于基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;

[0128] 建图单元230,用于以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;

[0129] 识别单元240,用于基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

[0130] 本申请实施例提供的信贷交易风险识别装置,对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果;由于采用了聚类分析和分组抽取样本的方法,实现了对庞大的信贷客户数据进行分析,由于采用了构建网络图进行信贷交易分析的方法,使得可以从样本信贷客户之间的关联关系出发,挖掘样本信贷客户之间潜在的信贷交易风险,缩小了异常信贷客户的识别范围,提高了信贷交易风险的识别效率,提高了信贷交易风险的识别准确性。

[0131] 在一些实施例中,确定单元用于:

[0132] 确定预设距离阈值以及各个样本信贷客户之间的距离;

[0133] 将各个样本信贷客户之间的距离与预设距离阈值进行比较;

[0134] 在距离小于预设距离阈值的情况下,确定距离对应的两个样本信贷客户的关联关系数值为第一数值;

[0135] 在距离大于或者等于预设距离阈值的情况下,确定距离对应的两个样本信贷客户的关联关系数值为第二数值;

[0136] 基于各个样本信贷客户的关联关系数值构建邻接矩阵;邻接矩阵用于表示各个样本信贷客户的关联关系。

[0137] 在一些实施例中,确定单元用于:

[0138] 基于邻接矩阵中各个样本信贷客户对应的第一数值和第二数值的数量,确定网络图的网络粘性。

[0139] 在一些实施例中,确定单元用于:

[0140] 基于各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的节点属性;

[0141] 基于邻接矩阵和各个样本信贷客户的节点属性,建立网络自回归模型;网络自回归模型用于对网络图的网络结构进行预测。

[0142] 在一些实施例中,还包括:

[0143] 显示单元,用于基于力导向图、桑基图和树型图中的至少一种,对网络图进行可视化显示。

[0144] 在一些实施例中,识别单元用于:

[0145] 在样本信贷客户的信贷交易风险识别结果为异常的情况下,将样本信贷客户所属的信贷客户组中的各个信贷客户确定为候选异常信贷客户;

[0146] 对各个候选异常信贷客户的信贷交易数据进行核验,基于各个候选异常信贷客户的信贷交易数据核验结果,在各个候选异常信贷客户中确定异常信贷客户。

[0147] 在一些实施例中,识别单元用于:

[0148] 获取各个候选异常信贷客户的社交媒体数据;

[0149] 将各个候选异常信贷客户的信贷交易数据和/或社交媒体数据输入信贷交易风险识别模型,得到信贷交易风险识别模型输出的各个候选异常信贷客户的信贷交易风险识别

结果;

[0150] 将信贷交易风险识别结果为异常的候选异常信贷客户,确定出为异常信贷客户;

[0151] 其中,信贷交易风险识别模型是以神经网络模型为初始模型,基于多个样本异常信贷客户的信贷交易数据和/或社交媒体数据对初始模型进行训练后得到的。

[0152] 图3是本申请提供的电子设备的结构示意图,如图3所示,该电子设备可以包括:处理器 (Processor) 310、通信接口 (Communications Interface) 320、存储器 (Memory) 330和通信总线 (Communications Bus) 340,其中,处理器310,通信接口320和存储器330通过通信总线340完成相互间的通信。处理器310可以调用存储器330中的逻辑命令,以执行如下方法:

[0153] 对多个信贷客户的信贷交易数据进行聚类分析,确定多个信贷客户组,并从各个信贷客户组中抽取样本信贷客户;基于各个样本信贷客户之间的距离,确定各个样本信贷客户之间的关联关系;以各个样本信贷客户为节点,以各个样本信贷客户之间的关联关系为边,建立网络图;基于网络图中各个样本信贷客户的信贷交易数据,确定各个样本信贷客户的信贷交易风险识别结果。

[0154] 此外,上述的存储器中的逻辑命令可以通过软件功能单元的形式实现并作为独立的产品销售或使用,可以存储在一个计算机可读取存储介质中。基于这样的理解,本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的部分可以以软件产品的形式体现出来,该计算机软件产品存储在一个存储介质中,包括若干命令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0155] 本申请实施例提供的电子设备中的处理器可以调用存储器中的逻辑指令,实现上述方法,其具体的实施方式与前述方法实施方式一致,且可以达到相同的有益效果,此处不再赘述。

[0156] 本申请实施例还提供一种计算机可读的存储介质,其上存储有计算机程序,该计算机程序被处理器执行时实现以执行上述各实施例提供的方法。

[0157] 其具体的实施方式与前述方法实施方式一致,且可以达到相同的有益效果,此处不再赘述。

[0158] 本申请实施例提供一种计算机程序产品,包括计算机程序,计算机程序被处理器执行时实现如上述方法。

[0159] 以上所描述的装置实施例仅仅是示意性的,其中所述作为分离部件说明的单元可以是或者也可以不是物理上分开的,作为单元显示的部件可以是或者也可以不是物理单元,即可以位于一个地方,或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。本领域普通技术人员在不付出创造性的劳动的情况下,即可以理解并实施。

[0160] 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各实施方式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。基于这样的理解,上述技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来,该

计算机软件产品可以存储在计算机可读存储介质中,如ROM/RAM、磁碟、光盘等,包括若干指令用以使得一台计算机设备(可以是个人计算机,服务器,或者网络设备等)执行各个实施例或者实施例的某些部分所述的方法。

[0161] 最后应说明的是:以上实施例仅用以说明本申请的技术方案,而非对其限制;尽管参照前述实施例对本申请进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本申请各实施例技术方案的精神和范围。

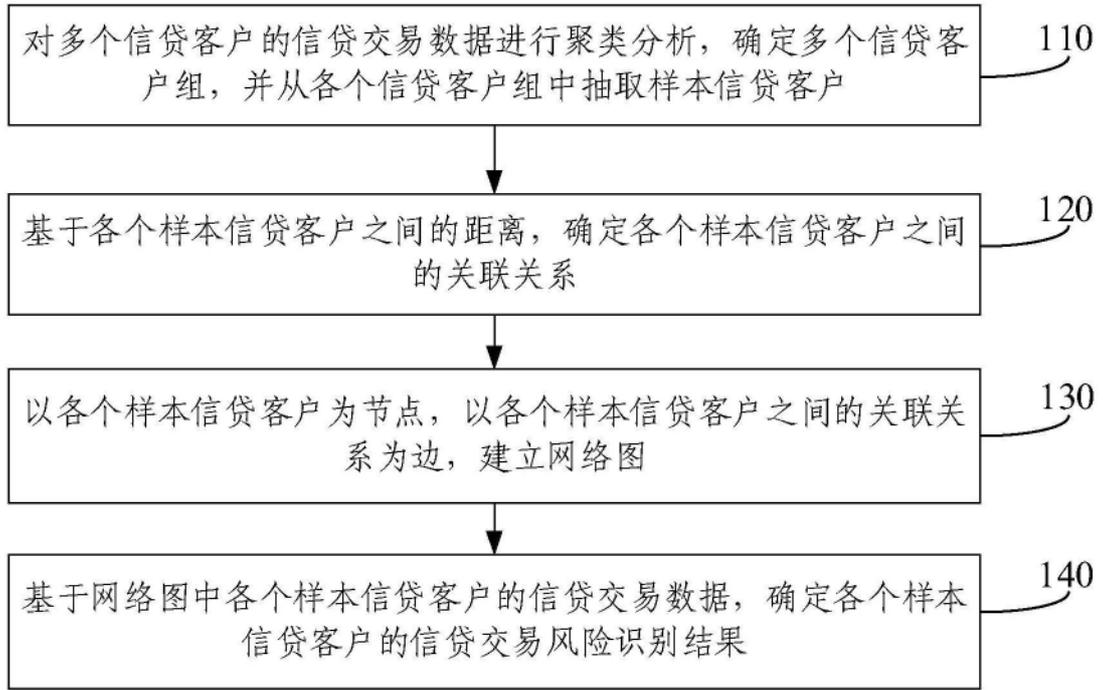


图1



图2

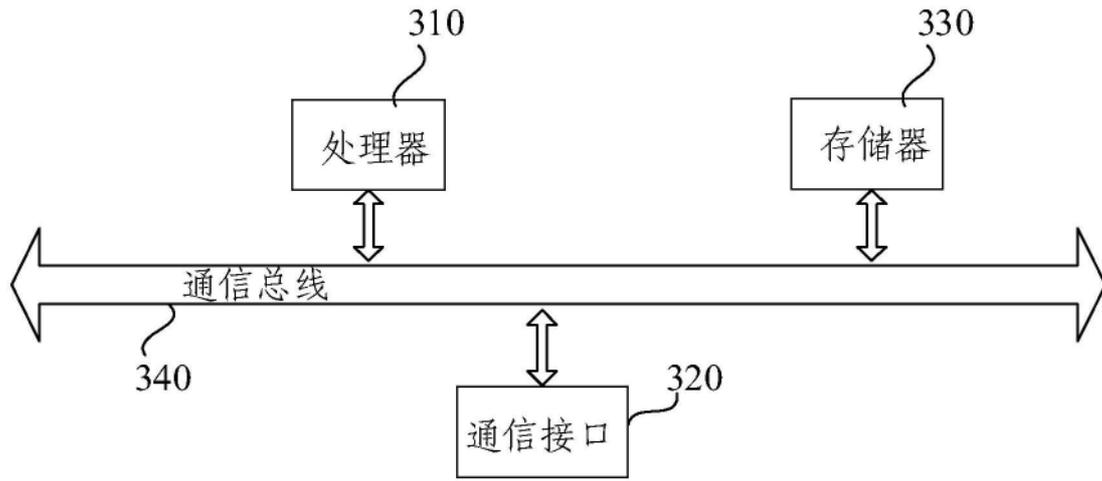


图3