



(12)发明专利申请

(10)申请公布号 CN 106504084 A

(43)申请公布日 2017.03.15

(21)申请号 201611010102.7

(22)申请日 2016.11.16

(71)申请人 航天信息股份有限公司

地址 100195 北京市海淀区杏石口路甲18
号航天信息园

(72)发明人 齐超 胡懋地

(74)专利代理机构 北京工信联合知识产权代理
事务所(普通合伙) 11266

代理人 郭一斐

(51)Int.Cl.

G06Q 40/00(2012.01)

权利要求书1页 说明书4页 附图1页

(54)发明名称

一种用于识别供应链中核心企业的方法及
系统

(57)摘要

本发明提供了一种供应链核心企业的识别方法和系统,利用图论中特征向量中心性度量方法来计算企业在供应链中的重要程度,本发明把各企业之间的购销关系当成是一个无向图结构,通过简单可行的算法,有效地计算出每个企业的
重要性,有利于税务机关企业与行业中所有企业的关联。所述方法包括:构建购方识别号、销方识别号和贸易金额的三元组的集合;建立纳税人识别号的集合,并建立邻接矩阵和无向图结构;计算邻接矩阵中每个节点的相对中心性;对相对中心性最大的前k个节点所对应的纳税人识别号进行标记,以表示该纳税人识别号对应的企业是供应链核心企业。

100

建立购方识别号、销方识别号的二元
组的集合 101

取出发票中的贸易金额添加到所建立
二元组的集合中,形成三元组的集合 102

建立纳税人识别号的集合及邻接矩阵 103

构成无向图结构 104

计算节点y在特征向量X中的
相对中心性得分 105

计算最大特征值λ和相
对应的特征向量 106

对X进行归一化,然后分别对应每个节
点的值进行标准化,将特征向量中每
个节点的值转换为该节点在整个向量
中所占的百分比 107

取出相对中心性得分最高
的前k个节点编号,确定
对应的纳税人识别号 108

1. 一种用于识别供应链中核心企业的方法,包括:

获取发票中的购方识别号和销方识别号并建立购方识别号和销方识别号的二元组的集合,其中购方识别号和销方识别号均为纳税人识别号;

获取发票中的贸易金额添加并到所述二元组的集合中,从而形成三元组的集合;

建立纳税人识别号的集合,设纳税人识别号的个数为n,则将纳税人识别号与[1,n]之间的整数建立一一映射,建立一个n*n的邻接矩阵A,且每个企业对应一个纳税人识别号;

将企业之间的贸易金额作为邻接矩阵中相邻节点之间的权重并且添加至邻接矩阵中,以构成无向图结构;

通过 $AX = \lambda X$ 计算出特征值 λ ,以计算出的最大特征值作为 λ 的值,并计算出与 λ 对应的特征向量X的值,其中A为邻接矩阵;

通过 $x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$ 计算出节点v在特征向量X中的相对中心性得分 x_v ,其中v、t均为节点,x_t为节点t在特征向量X中的相对中心性得分, λ 为特征值,M(v)表示与节点v相连的点的集合,G=(E,V),其中E表示边的集合,V表示节点的集合;

对特征向量X进行归一化,然后分别对应每个节点的值进行标准化,将特征向量中每个节点的值转换为该节点在整个向量中所占的百分比;

取出相对中心性得分最高的前k个节点,并根据邻接矩阵中每个节点对应的纳税人识别号,对相对中心性得分最高的前k个纳税人识别号进行标记,以表示该纳税人识别号对应的企业是供应链中的核心企业。

2. 根据权利要求1所述的方法,其特征在于,若两个企业有多次贸易关系,则对贸易金额进行累加。

3. 根据权利要求1所述的方法,其特征在于,将企业之间的贸易金额作为邻接矩阵中相邻节点之间的权重时,按一定阈值倍数对所述贸易金额进行缩小。

4. 根据权利要求1所述的方法,其特征在于,所述k的值小于n的值。

5. 一种用于识别供应链中核心企业的系统,包括:

数据处理模块,用于构建购方识别号、销方识别号和贸易金额的三元组的集合;

图构造模块,建立纳税人识别号的集合,并建立邻接矩阵和无向图结构;

特征向量中心性度量计算模块,计算邻接矩阵中每个节点的相对中心性,即特征向量;以及

核心企业标记模块,对相对中心性最大的前k个节点所对应的纳税人识别号进行标记,以表示该纳税人识别号对应的企业是供应链核心企业。

一种用于识别供应链中核心企业的方法及系统

技术领域

[0001] 本发明涉及税务大数据领域，并且更具体地，涉及一种用于识别供应链中核心企业的方法及系统。

背景技术

[0002] 随着互联网和大数据技术的兴起，税务机关的税务数据规模不断增长，基于税务数据的企业画像技术应运而生。互联网时代，很多企业和政府机构也嗅到了大数据带来的机遇。由于征信双方信息不对称，信息平台的数据处理能力和数据挖掘能力不够等问题，税务机关往往不能对企业形成一个立体化的描述。大部分的企业通过使用数据参与经营决策，减低经营成本，获取经营收入，使得企业的涉税行为更加隐蔽和无形，税务机关通过现有的征管手段很难跟踪和获知这些情况。

[0003] 供应链核心企业作为企业画像中重要的标识，通过标记供应链核心企业能够有效找到行业供应链中的重要企业，这样有利于税务机关企业与行业中所有企业的关联，查看企业对行业的影响。

发明内容

[0004] 本发明提供了一种用于识别供应链中核心企业的方法及系统，利用图论中 Eigenvector Centrality 特征向量中心性度量方法来计算企业在供应链中的重要程度，该方法把各企业之间的购销关系当成是一个无向图结构，计算方法简单可行，有效地计算出每个企业的重要性。供应链核心企业的识别和标注是进行企业画像的重要部分，方便税务机关通过查看供应链核心企业的税务情况了解整个行业的发展情况，对行业风向和投资决策和投资依据具有重大意义。

[0005] 根据本发明的一个方面，提供一种用于识别供应链中核心企业的方法，包括：

[0006] 获取发票中的购方识别号和销方识别号并建立购方识别号和销方识别号的二元组的集合，其中购方识别号和销方识别号均为纳税人识别号；

[0007] 获取发票中的贸易金额添加并到所述二元组的集合中，从而形成三元组的集合；

[0008] 建立纳税人识别号的集合，设纳税人识别号的个数为n，则将纳税人识别号与[1, n]之间的整数建立一一映射，建立一个n*n的邻接矩阵A，且每个企业对应一个纳税人识别号；

[0009] 将企业之间的贸易金额作为邻接矩阵中相邻节点之间的权重并且添加至邻接矩阵中，以构成无向图结构；

[0010] 通过 $AX = \lambda X$ 计算出特征值 λ ，以计算出的最大特征值作为 λ 的值，并计算出与 λ 对应的特征向量X的值，其中A为邻接矩阵；

[0011] 通过 $X_v = \frac{1}{\lambda} \sum_{t \in M(v)} X_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} X_t$ 计算出节点v在特征向量X中的相对中心性得分 x_v ，其中v、t均为节点， x_t 为节点t在特征向量X中的相对中心性得分， λ 为特征值，M

- (v) 表示与节点v相连的点的集合, $G = (E, V)$, 其中E表示边的集合, V表示节点的集合;
- [0012] 对特征向量X进行归一化, 然后分别对应每个节点的值进行标准化, 将特征向量中每个节点的值转换为该节点在整个向量中所占的百分比;
- [0013] 取出相对中心性得分最高的前k个节点, 并根据邻接矩阵中每个节点对应的纳税人识别号, 对相对中心性得分最高的前k个纳税人识别号进行标记, 以表示该纳税人识别号对应的企业是供应链中的核心企业。
- [0014] 优选地, 若两个企业有多次贸易关系, 则对贸易金额进行累加。
- [0015] 优选地, 将企业之间的贸易金额作为邻接矩阵中相邻节点之间的权重时, 按一定阈值倍数对所述贸易金额进行缩小。
- [0016] 优选地, 所述k的值小于n的值。
- [0017] 根据本发明的另一方面, 提供一种用于识别供应链中核心企业的系统, 包括:
- [0018] 数据处理模块, 用于构建购方识别号、销方识别号和贸易金额的三元组的集合;
- [0019] 图构造模块, 建立纳税人识别号的集合, 并建立邻接矩阵和无向图结构;
- [0020] 特征向量中心性度量计算模块, 计算邻接矩阵中每个节点的相对中心性, 即特征向量; 以及
- [0021] 核心企业标记模块, 对相对中心性最大的前k个节点所对应的纳税人识别号进行标记, 以表示该纳税人识别号对应的企业是供应链核心企业。

附图说明

- [0022] 通过参考下面的附图, 可以更为完整地理解本发明的示例性实施方式:
- [0023] 图1为根据本发明优选实施例的供应链核心企业的识别方法的流程图; 以及
- [0024] 图2为根据本发明优选实施例的供应链核心企业的识别系统的结构图。

具体实施方式

[0025] 现在参考附图介绍本发明的示例性实施方式, 然而, 本发明可以用许多不同的形式来实施, 并且不局限于此处描述的实施例, 提供这些实施例是为了详尽地且完全地公开本发明, 并且向所属技术领域的技术人员充分传达本发明的范围。对于表示在附图中的示例性实施方式中的术语并不是对本发明的限定。在附图中, 相同的单元/元件使用相同的附图标记。

[0026] 除非另有说明, 此处使用的术语(包括科技术语)对所属技术领域的技术人员具有通常的理解含义。另外, 可以理解的是, 以通常使用的词典限定的术语, 应当被理解为与其相关领域的语境具有一致的含义, 而不应该被理解为理想化的或过于正式的意义。

[0027] 图1为根据本发明优选实施例的供应链核心企业的识别方法的流程图。如图1所示, 供应链核心企业的识别方法100从步骤101开始。在步骤101中, 从发票数据库中识别出发票上的购方识别号和销方识别号信息, 并建立购方识别号和销方识别号的二元组的集合, 其中, 购方识别号和销方识别号均为纳税人识别号。应当了解的是, 每一个企业均有且只有一个纳税人识别号, 只是因购销关系不同区分为购方识别号和销方识别号。

[0028] 在步骤102中, 取出发票中的贸易金额添加到步骤101形成的二元组的集合中, 从而形成三元组的集合, 若两个企业之间存在多次贸易关系, 则对贸易金额进行累加。应当了

解的是,所述两个企业之间存在贸易关系,是指两个企业之间互为购方和销方,并不仅限于某一个企业只作为购方,另一个企业只做销方。

[0029] 在步骤103中,建立纳税人识别号的集合,设纳税人识别号的个数为n,则将纳税人识别号与[1,n]之间的整数建立一一映射,建立一个n*n的邻接矩阵A。

[0030] 在步骤104中,将企业之间的贸易金额作为邻接矩阵中各节点之间的权重添加至邻接矩阵中,在进行权重添加时,按一定的阈值倍数对所述贸易金额进行缩小,其中,所述阈值可以为1000倍、10000倍或50000倍等,具体缩小的倍数根据实际情况选择。

[0031] 在步骤105中,通过 $AX=\lambda X$ 计算出特征值 λ ,由于会计算出多个特征值值,在本发明中,以计算出的最大特征值作为 λ 的值,并计算出与 λ 对应的特征向量X的值,其中A为邻接矩阵。

[0032] 在步骤106中,通过 $x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} x_t$ 计算出节点v在特征向量X中的相对中心性得分 x_v ,所述相对中心性得分,即为节点v在供应链中的重要程度。其中v、t均为节点,x_t为节点t在特征向量X中的相对中心性得分, λ 为特征值,M(v)表示与节点v相连的点的集合,G=(E,V),其中E表示边的集合,V表示节点的集合。

[0033] 在步骤107中,对步骤105中得出的特征向量进行归一化,即对特征向量中的每个节点的值进行相加求和,然后分别对应每个节点的值进行标准化,将特征向量中每个节点的值转换为该节点在整个向量中所占的百分比。

[0034] 在步骤108中,取出相对中心性得分最高的前k个节点编号,并根据邻接矩阵中每个节点对应的纳税人识别号,对相对中心性得分最高的前k个纳税人识别号进行标记,以表示该纳税人识别号对应的企业是供应链核心企业。其中,k的值小于纳税人识别号的个数n的值,并且在实际使用时,k的值推荐取5或者10。

[0035] 图2为根据本发明优选实施例的供应链核心企业的识别系统的结构图。如图2所示,供应链核心企业的识别系统200主要由数据处理模块201、图构造模块202、特征向量中心性度量计算模块203以及核心企业标记模块204组成。

[0036] 优选地,所述数据处理模块201用于构建购方识别号、销方识别号和贸易金额的三元组的集合。

[0037] 优选地,所述图构造模块202用于建立纳税人识别号的集合,并利用与纳税人识别号个数的映射建立邻接矩阵,并在邻接矩阵中添加权重构建无向图结构。

[0038] 优选地,所述特征向量中心性度量计算模块203,通过计算特征值和特征向量,计算邻接矩阵中每个节点的中心性,即企业在供应链中的重要程度。

[0039] 优选地,所述核心企业标记模块204通过对相对中心性得分最高的前k个节点所对应的纳税人识别号进行标记,以表示该纳税人识别号对应的企业是供应链核心企业。

[0040] 已经通过参考少量实施方式描述了本发明。然而,本领域技术人员所公知的,正如附带的专利权利要求所限定的,除了本发明以上公开的其他的实施例等同地落在本发明的范围内。

[0041] 通常地,在权利要求中使用的所有术语都根据他们在技术领域的通常含义被解释,除非在其中被另外明确地定义。所有的参考“一个/所述/该[装置、组件等]”都被开放地解释为所述装置、组件等中的至少一个实例,除非另外明确地说明。这里公开的任何方法的

步骤都没必要以公开的准确的顺序运行,除非明确地说明。



图1

图2