



(12)发明专利申请

(10)申请公布号 CN 109993644 A

(43)申请公布日 2019.07.09

(21)申请号 201711476904.1

(22)申请日 2017.12.29

(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

(72)发明人 刘丹 郭宏 潘竞旭 田宜喜

谢宇 张玉魁 陈荣兴

(74)专利代理机构 北京同达信恒知识产权代理

有限公司 11291

代理人 黄志华

(51)Int.Cl.

G06Q 40/00(2012.01)

G06Q 10/06(2012.01)

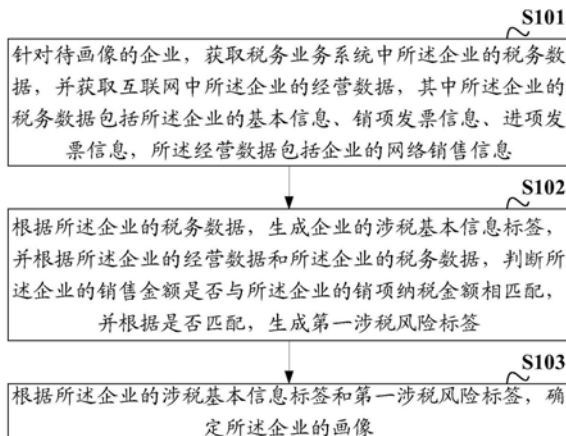
权利要求书3页 说明书12页 附图2页

(54)发明名称

一种画像确定方法、装置、电子设备及存储介质

(57)摘要

本发明公开了一种画像确定方法、装置、电子设备及存储介质,所述方法包括:针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。提供了一种能够真实、全面描述企业税务形象的方案。



1. 一种画像确定方法,其特征在于,所述方法包括:

针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;

根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;

根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。

2. 如权利要求1所述的方法,其特征在于,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像之前,所述方法还包括:

通过所述企业的税务数据,识别所述企业是否为小规模纳税人企业,如果所述企业为小规模纳税人企业,根据所述企业的销售金额是否大于预设的销售金额阈值,确定所述企业是否满足一般纳税人条件,并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

3. 如权利要求1或2所述的方法,其特征在于,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像之前,所述方法还包括:

根据所述企业的进项发票信息和销项发票信息,确定所述企业的发票异常信息,并根据所述企业的发票异常信息,确定所述企业的第二涉税风险标签;

所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像包括:

根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签,确定所述企业的画像。

4. 如权利要求3所述的方法,其特征在于,所述方法还包括:

根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型,确定所述企业的涉税风险等级。

5. 如权利要求4所述的方法,其特征在于,预先对所述涉税风险模型进行训练的过程包括:

针对训练集中的每个样本企业,根据该企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。

6. 如权利要求1所述的方法,其特征在于,所述针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据之后,根据所述企业的税务数据,生成企业的涉税基本信息标签之前,所述方法还包括:

对所述企业的税务数据和经营数据进行数据清洗和格式转换。

7. 如权利要求1所述的方法,其特征在于,所述方法还包括:

针对待画像的行业,获取税务系统中所述行业的税务数据,其中行业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;

根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速

增长行业、快速下降行业、稳定发展行业、周期性行业；

根据所述行业的基本信息标签和发展标签，确定所述行业的画像。

8. 如权利要求1所述的方法，其特征在于，所述方法还包括：

针对待画像的地区，获取税务系统中所述地区的税务数据，其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据；

根据所述地区的税务数据，确定所述地区的涉税基本信息标签，并根据所述地区中包含的企业的税务数据及趋势算法，确定所述地区的销项金额的发展趋势，并根据所述地区的销项金额的发展趋势，确定所述地区的地区发展标签，其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区；

根据所述地区的基本信息标签和发展标签，确定所述地区的画像。

9. 一种画像确定装置，其特征在于，所述装置包括：

获取模块，用于针对待画像的企业，获取税务业务系统中所述企业的税务数据，并获取互联网中所述企业的经营数据，其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息，所述经营数据包括企业的网络销售信息；

判断确定模块，用于根据所述企业的税务数据，生成企业的涉税基本信息标签，并根据所述企业的经营数据和所述企业的税务数据，判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配，并根据是否匹配，生成第一涉税风险标签；

第一画像模块，用于根据所述企业的涉税基本信息标签和第一涉税风险标签，确定所述企业的画像。

10. 如权利要求9所述的装置，其特征在于，所述判断确定模块，还用于通过所述企业的税务数据，识别所述企业是否为小规模纳税人企业，如果所述企业为小规模纳税人企业，根据所述企业的销售金额是否大于预设的销售金额阈值，确定所述企业是否满足一般纳税人条件，并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

11. 如权利要求9或10所述的装置，其特征在于，所述判断确定模块，还用于根据所述企业的进项发票信息和销项发票信息，确定所述企业的发票异常信息，并根据所述企业的发票异常信息，确定所述企业的第二涉税风险标签；

所述第一画像模块，具体用于根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签，确定所述企业的画像。

12. 如权利要求11所述的装置，其特征在于，所述装置还包括：

第一确定模块，用于根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型，确定所述企业的涉税风险等级。

13. 如权利要求12所述的装置，其特征在于，所述装置还包括：

训练模块，用于针对训练集中的每个样本企业，根据该企业的涉税风险等级，为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签，将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型，对涉税风险模型进行训练。

14. 如权利要求9所述的装置，其特征在于，所述装置还包括：

清洗模块，用于对所述企业的税务数据和经营数据进行数据清洗和格式转换。

15. 如权利要求9所述的装置，其特征在于，所述装置还包括：

第二画像模块，用于针对待画像的行业，获取税务系统中所述行业的税务数据，其中行

业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速增长行业、快速下降行业、稳定发展行业、周期性行业;根据所述行业的基本信息标签和发展标签,确定所述行业的画像。

16. 如权利要求9所述的装置,其特征在于,所述装置还包括:

第三画像模块,用于针对待画像的地区,获取税务系统中所述地区的税务数据,其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据;根据所述地区的税务数据,确定所述地区的涉税基本信息标签,并根据所述地区中包含的企业的税务数据及趋势算法,确定所述地区的销项金额的发展趋势,并根据所述地区的销项金额的发展趋势,确定所述地区的地区发展标签,其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区;根据所述地区的基本信息标签和发展标签,确定所述地区的画像。

17. 一种电子设备,其特征在于,包括:处理器、通信接口、存储器和通信总线,其中,处理器,通信接口,存储器通过通信总线完成相互间的通信;

存储器,用于存放计算机程序;

处理器,用于执行存储器上所存放的计算机程序时,实现权利要求1-8任一所述的方法步骤。

18. 一种计算机可读存储介质,其特征在于,所述计算机可读存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现权利要求1-8任一所述的方法步骤。

## 一种画像确定方法、装置、电子设备及存储介质

### 技术领域

[0001] 本发明涉及数据处理技术领域,尤其涉及一种画像确定方法、装置、电子设备及存储介质。

### 背景技术

[0002] 企业作为社会经济活动的主体,涉及到社会经济发展的方方面面,与人们的生活息息相关,但是随着企业的发展,企业业务不断的拓展,企业的涉税行为越来越广泛,也越来越隐蔽,由于征纳双方信息不对称,从单一的税务系统已经无法了解一个企业的真实形象。以一家特色餐厅为例,在互联网上,该餐厅星级高、好评如潮,每天就餐需预约排队,每月数万条品评留言。但是在税务系统中,该餐厅缴税额很低,与互联网上该餐厅的数据明显不匹配,这种现状充分说明,依靠单一的税务系统所描述的企业税务形象与企业的真实税务形象存在着很大的差异,并且描述的企业的税务形象也不全面。

[0003] 因此,急需一种能够全面、真实描述企业税务形象的方案,便于全面、真实的了解企业的税务形象。

### 发明内容

[0004] 本发明提供一种画像确定方法、装置、电子设备及存储介质,用以解决现有技术中存在对企业税务形象描述不真实、不全面的问题。

[0005] 本发明公开了一种画像确定方法,所述方法包括:

[0006] 针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;

[0007] 根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;

[0008] 根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。

[0009] 进一步地,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像之前,所述方法还包括:

[0010] 通过所述企业的税务数据,识别所述企业是否为小规模纳税人企业,如果所述企业为小规模纳税人企业,根据所述企业的销售金额是否大于预设的销售金额阈值,确定所述企业是否满足一般纳税人条件,并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

[0011] 进一步地,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像之前,所述方法还包括:

[0012] 根据所述企业的进项发票信息和销项发票信息,确定所述企业的发票异常信息,并根据所述企业的发票异常信息,确定所述企业的第二涉税风险标签;

[0013] 所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像包括:

[0014] 根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签,确定所述企业的画像。

[0015] 进一步地,所述方法还包括:

[0016] 根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型,确定所述企业的涉税风险等级。

[0017] 进一步地,预先对所述涉税风险模型进行训练的过程包括:

[0018] 针对训练集中的每个样本企业,根据该企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。

[0019] 进一步地,所述针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据之后,根据所述企业的税务数据,生成企业的涉税基本信息标签之前,所述方法还包括:

[0020] 对所述企业的税务数据和经营数据进行数据清洗和格式转换。

[0021] 进一步地,所述方法还包括:

[0022] 针对待画像的行业,获取税务系统中所述行业的税务数据,其中行业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;

[0023] 根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速增长行业、快速下降行业、稳定发展行业、周期性行业;

[0024] 根据所述行业的基本信息标签和发展标签,确定所述行业的画像。

[0025] 进一步地,所述方法还包括:

[0026] 针对待画像的地区,获取税务系统中所述地区的税务数据,其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据;

[0027] 根据所述地区的税务数据,确定所述地区的涉税基本信息标签,并根据所述地区中包含的企业的税务数据及趋势算法,确定所述地区的销项金额的发展趋势,并根据所述地区的销项金额的发展趋势,确定所述地区的地区发展标签,其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区;

[0028] 根据所述地区的基本信息标签和发展标签,确定所述地区的画像。

[0029] 本发明公开了一种画像确定装置,所述装置包括:

[0030] 获取模块,用于针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;

[0031] 判断确定模块,用于根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;

[0032] 第一画像模块,用于根据所述企业的涉税基本信息标签和第一涉税风险标签,确

定所述企业的画像。

[0033] 进一步地,所述判断确定模块,还用于通过所述企业的税务数据,识别所述企业是否为小规模纳税人企业,如果所述企业为小规模纳税人企业,根据所述企业的销售金额是否大于预设的销售金额阈值,确定所述企业是否满足一般纳税人条件,并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

[0034] 进一步地,所述判断确定模块,还用于根据所述企业的进项发票信息和销项发票信息,确定所述企业的发票异常信息,并根据所述企业的发票异常信息,确定所述企业的第二涉税风险标签;

[0035] 所述第一画像模块,具体用于根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签,确定所述企业的画像。

[0036] 进一步地,所述装置还包括:

[0037] 第一确定模块,用于根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型,确定所述企业的涉税风险等级。

[0038] 进一步地,所述装置还包括:

[0039] 训练模块,用于针对训练集中的每个样本企业,根据该企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。

[0040] 进一步地,所述装置还包括:

[0041] 清洗模块,用于对所述企业的税务数据和经营数据进行数据清洗和格式转换。

[0042] 进一步地,所述装置还包括:

[0043] 第二画像模块,用于针对待画像的行业,获取税务系统中所述行业的税务数据,其中行业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速增长行业、快速下降行业、稳定发展行业、周期性行业;根据所述行业的基本信息标签和发展标签,确定所述行业的画像。

[0044] 进一步地,所述装置还包括:

[0045] 第三画像模块,用于针对待画像的地区,获取税务系统中所述地区的税务数据,其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据;根据所述地区的税务数据,确定所述地区的涉税基本信息标签,并根据所述地区中包含的企业的税务数据及趋势算法,确定所述地区的销项金额的发展趋势,并根据所述地区的销项金额的发展趋势,确定所述地区的地区发展标签,其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区;根据所述地区的基本信息标签和发展标签,确定所述地区的画像。

[0046] 本发明公开了一种电子设备,包括:处理器、通信接口、存储器和通信总线,其中,处理器,通信接口,存储器通过通信总线完成相互间的通信;

[0047] 存储器,用于存放计算机程序;

[0048] 处理器,用于执行存储器上所存放的计算机程序时,实现上述任一所述的方法步骤。

[0049] 本发明公开了一种计算机可读存储介质,所述计算机可读存储介质内存储有计算机程序,所述计算机程序被处理器执行时实现上述任一所述的方法步骤。

[0050] 本发明公开了一种画像确定方法、装置、电子设备及存储介质,所述方法包括:针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。由于在本发明实施例中,针对待画像的企业,获取税务业务系统中企业的税务数据,并获取互联网中企业的经营数据,根据企业的税务数据,生成企业的涉税基本信息标签,并根据企业的经营数据和企业的税务数据,判断企业的销项金额是否与企业的销项纳税金额相匹配,生成第一涉税风险标签,根据企业的涉税基本信息标签和企业的的第一涉税风险标签,确定企业的画像,保证了企业的画像的真实性和全面性,提供了一种全面且真实反应企业税务形象的方案。

### 附图说明

[0051] 为了更清楚地说明本发明实施例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,下面描述中的附图仅仅是本发明的一些实施例,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。

[0052] 图1为本发明实施例提供的一种画像确定过程示意图;

[0053] 图2为本发明实施例提供的一种画像架构示意图;

[0054] 图3为本发明实施例提供的一种画像确定装置结构示意图;

[0055] 图4为本发明实施例提供的一种电子设备示意图。

### 具体实施方式

[0056] 为了使本发明的目的、技术方案和优点更加清楚,下面将结合附图本发明作进一步地详细描述,显然,所描述的实施例仅仅是本发明的一部分实施例,而不是全部的实施例。基于本发明中的实施例,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,都属于本发明保护的范围。

[0057] 实施例1:

[0058] 图1为本发明实施例提供的一种画像确定过程示意图,该过程包括:

[0059] S101:针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息。

[0060] 本发明实施例提供的企业画像确定方法应用于电子设备,该电子设备可以为手机、平板电脑、个人电脑(PC)、服务器等设备。

[0061] 具体的,电子设备使用大数据采集工具(Sqoop)首先是从税务业务系统获取税务数据,包括:电子底账、防伪税控、金税三期库,并将获取的税务业务系统中的税务数据,存



到数据仓库工具(hive)数据库的对应表中。

[0062] 此外,电子设备还可以通过网络爬虫获取互联中该企业的经营数据,较佳的,因为网络爬虫获取的数据大多都是非结构化的数据,电子设备可以将网络爬虫获取到的数据存储到分布式数据库(Hbase)的对应表中。在本发明实施例中,获取税务业务系统中企业的税务数据和互联网中企业的经营数据的过程是现有技术,不在进行赘述。

[0063] S102:根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签。

[0064] 具体的,获取待画像的企业的税务数据和经营数据后,根据所述企业的基本信息,确定所述企业的涉税基本信息标签,其中所述涉税基本信息标签包括:纳税人名称、纳税人识别号、所属税务机关、发票种类、行业名称、纳税人状态、是否新办企业、进项金额、进项税额、进项发票份数、进项物品名称、销项金额、销项税额、销项发票份数、销项物品名称、同比(税额)、同比(金额)、环比(税额)、环比(金额)、是否行业龙头企业、是否产业链核心企业、是否外贸型企业、是否快速增长企业、是否快速下降企业、是否稳定发展企业、是否周期性企业等;并根据所述企业的网络销售信息和销项发票信息,确定所述企业的销售金额和所述企业的销项纳税金额,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签,例如:如果匹配,生成销售金额与销项纳税金额匹配标签、如果不匹配,生成销售金额与销项纳税金额不匹配标签。

[0065] S103:根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。

[0066] 在电子设备中,还提供有供用户对企业的标签进行管理的企业标签管理页面,用户可以通过该页面进行标签的定义,如定义标签的具体含义、数据来源、标签生成的计算过程等,并可以通过该页面对标签进行增加、删除、修改等操作。用户还可以通过企业标签管理页面,设置每个标签所属的主体类别、标签之间的级别关系以及对标签的树状展现关系,电子设备可以根据用户设置的每个标签所属的主体类别、标签之间的级别关系以及对标签的树状展现关系,以树状存储待画像的企业的标签,进而确定该企业的画像。

[0067] 由于在本发明实施例中,针对待画像的企业,获取税务业务系统中企业的税务数据,并获取互联网中企业的经营数据,根据企业的税务数据,生成企业的涉税基本信息标签,并根据企业的经营数据和企业的税务数据,判断企业的销项金额是否与企业的销项纳税金额相匹配,生成第一涉税风险标签,根据企业的涉税基本信息标签和企业的涉税风险标签,确定企业的画像,保证了企业的画像的真实性和全面性,提供了一种全面且真实反应企业税务形象的方案。

[0068] 实施例2:

[0069] 此外,在上述实施例的基础上,在本发明实施例中,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像之前,所述方法还包括:

[0070] 通过所述企业的税务数据,识别所述企业是否为小规模纳税人企业,如果所述企业为小规模纳税人企业,根据所述企业的销售金额是否大于预设的销售金额阈值,确定所述企业是否满足一般纳税人条件,并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

[0071] 此外,企业的税务数据中还包括企业是否为小规模纳税人企业的信息,如果电子设备识别到所述企业为小规模纳税人企业,电子设备还可以根据所述企业的销售金额是否大于预设的销售金额阈值;其中所述预设的销售金额阈值是根据一般纳税人需要满足的销售金额确定的,确定所述企业是否满足一般纳税人条件,并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

[0072] 电子设备根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像,其中第一涉税风险标签,包含企业是否满足一般纳税人条件和企业的销售金额是否与所述企业的销项纳税金额相匹配。

[0073] 实施例3:

[0074] 为了保证对企业画像的全面性,在上述实施例的基础上,在本发明实施例中,所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像之前,所述方法还包括:

[0075] 根据所述企业的进项发票信息和销项发票信息,确定所述企业的发票异常信息,并根据所述企业的发票异常信息,确定所述企业的第二涉税风险标签;

[0076] 所述根据所述企业的涉税基本信息标签和第一涉税风险标签,确定企业的画像包括:

[0077] 根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签,确定所述企业的画像。

[0078] 具体的,在本发明实施例中,电子设备还可以对企业的进项发票信息和销项发票信息进行分析,确定企业的发票异常信息,并将企业存在的发票异常信息,确定为企业的第二涉税风险标签,其中所述发票异常信息包括:大宗商品购销不匹配、日用商品购销不匹配、潜在虚开、当月领用并开具、频繁购票、非正常时间开具、顶额开具、两头在外、省外金额突增、异常凭证抵扣、销货清单异常、红字发票异常、发票集中作废异常、异常作废、失控发票、简易征收优惠虚开、即征即退虚开、差额征税虚开、出口企业虚开)、(申报行为的异常信息,如:是否期初申报、是否期末申报、是否未及时申报、是否申报异常、是否税负正常、是否税负偏大、是否税负偏小)(纳税信用的信息,如:纳税人税务登记是否健全、纳税申报是否及时、账簿凭证管理是否规范、税款缴纳是否及时、是否违反税收法律、是否受到行政法规行为处理等。在本发明实施例中根据进项发票信息和销项发票信息,确定发票异常信息是现有技术,不在进行赘述。

[0079] 实施例4:

[0080] 为了便于提高画像确定的效率与准确,在上述各实施例的基础上,在本发明实施例中,所述针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据之后,根据所述企业的税务数据,生成企业的涉税基本信息标签之前,所述方法还包括:

[0081] 对所述企业的税务数据和经营数据进行数据清洗和格式转换。

[0082] 具体的,电子设备获取待画像的企业的税务数据、经营数据后,对税务数据、经营数据进行数据仓库(ELT)清洗操作,去除冗杂及与标签无关的数据,并将非结构化的数据转换为结构化的数据,存储到Hive对应的ELT表中,便于电子设备对税务数据、经营数据的获取和分析。在本发明实施例中,对税务数据和经营数据进行数据清洗和格式转换是现有技

术,不在进行赘述。

[0083] 较佳的,电子设备对待画像的企业的税务数据和经营数据进行数据清洗和格式转换之后,还可以对该企业的税务数据和经营数据进行汇总和插入操作,例如:对企业的进项和销项发票数据进行聚合计算,计算在一段时间内的总销售额和购进额,计算某一物品的销售额或购进额等;对企业的信息进行汇总生成企业的维度表和事实表,其中所述维度表包括行业、税务机关、纳税人基础信息、商品编码、地域、时间等维度;事实表用来存储企业交易事实以及维度的量值,包括进项发票事实表、销项发票事实表、发票销售事实表、发票票证事实表、作废发票事实表、红字发票事实表等。

[0084] 实施例5:

[0085] 为了便于税务部门对企业涉税风险进行监管,在上述各实施例的基础上,在本发明实施例中,所述方法还包括:

[0086] 根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型,确定所述企业的涉税风险等级。

[0087] 在本发明实施例中,预先对所述涉税风险模型进行训练的过程包括:

[0088] 针对训练集中的每个样本企业,根据该企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。

[0089] 其中,所述企业的涉税风险标签包括企业的第一涉税风险标签和第二涉税风险标签。

[0090] 在本发明实施例中,在训练集中保存有大量已知涉税风险等级的样本企业的涉税风险标签,根据训练集中的每个样本企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。在本发明实施例中,对涉税风险模型进行训练的算法可以采用聚类分类(Mahout)算法、极限学习机(Extreme Learning Machine,ELM)算法等实现。具体的,对涉税风险模型进行训练的过程属于现有技术,在本发明实施例中,不在进行赘述。

[0091] 对涉税风险模型训练完成后,训练完成的涉税风险模型可以根据输入的企业存在的涉税风险标签,针对该企业输出相应的结果,即该企业的涉税风险等级。税务部门可以根据该企业的涉税风险等级对涉税风险等级高的企业进行监控,并对企业的走逃进行预测。

[0092] 实施例6:

[0093] 为了便于对行业的发展状况进行获知,在上述各实施例的基础上,在本发明实施例中,所述方法还包括:

[0094] 针对待画像的行业,获取税务系统中所述行业的税务数据,其中行业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;

[0095] 根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速增长行业、快速下降行业、稳定发展行业、周期性行业;

[0096] 根据所述行业的基本信息标签和发展标签,确定所述行业的画像。

[0097] 具体的,针对待画像的行业,电子设备可以根据该行业中包含的所有企业的税务数据,确定该行业中包含的每个企业的销项金额,进而确定该行业的销项金额,根据该行业的销项金额及趋势算法,确定该行业的销项金额的发展趋势,确定该行业是否为周期性行业,并根据该行业的销项金额的发展趋势,确定该行业是否为快速增长行业或快速下降行业或稳定发展行业从而确定所述行业的行业发展标签,例如:该行业的销项金额平均增长率大于百分之三,确定该行业为快速增长行业,该行业的销项金额平均增长率小于负百分之三,确定该行业为快速下降行业,该行业的销项金额平均增长率,在百分之三和负百分之三之间,确定该行业为稳定发展行业。此外如果该行业的销项金额呈现周期性波动,例如该行业在每年第三季度销项金额的较第二季度均上升超过百分三十,第四季度销项金额较第三季度均下降百分之三十,确定该行业为周期性行业。

[0098] 此外,针对待画像的行业,电子设备还可以根据该行业的税务数据,确定该行业的涉税基本信息标签,其中所述行业的涉税基本信息标签包括:行业名称、行业代码、行业门类、行业大类、是否重点监控行业、主要上游行业、主要下游行业、纳税人总户数;行业进项金额、行业进项税额、行业进项发票份数、行业销项金额、行业销项税额、行业销项发票份数;行业进项金额占比、行业进项金额排名、行业销项金额占比、行业销项金额排名;进项金额同比、进项金额环比、进项税额同比、进项税额占比;销项金额同比、销项金额环比、销项税额同比、销项税额占比等。并根据该行业的基本信息标签和发展标签,确定该行业的画像。在本发明实施例中,根据行业的税务数据确定行业的基础信息标签是现有技术,不再进行赘述。

[0099] 税务部门可以根据行业的画像,了解行业的发展状况和发展前景,对快速增长行业加大投入力度,对快速下降行业提供政策方面的扶持,还可以根据行业的画像,用于对宏观经济政策的事前、事中、事后的评估。

[0100] 实施例7:

[0101] 为了便于对地区的发展状况进行获知,在上述各实施例的基础上,在本发明实施例中,所述方法还包括:

[0102] 针对待画像的地区,获取税务系统中所述地区的税务数据,其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据;

[0103] 根据所述地区的税务数据,确定所述地区的涉税基本信息标签,并根据所述地区中包含的企业的税务数据及趋势算法,确定所述地区的销项金额的发展趋势,并根据所述地区的销项金额的发展趋势,确定所述地区的地区发展标签,其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区;

[0104] 根据所述地区的基本信息标签和发展标签,确定所述地区的画像。

[0105] 具体的,针对待画像的地区,电子设备可以根据该地区中包含的所有企业的税务数据,确定该地区中包含的每个企业的销项金额,进而确定该地区的销项金额,根据该地区的销项金额及趋势算法,确定该地区的销项金额的发展趋势,并根据该地区的销项金额的发展趋势,确定该地区是否为快速增长地区或快速下降地区或稳定发展地区,从而确定所述地区的地区发展标签。

[0106] 此外,针对待画像的地区,电子设备还可以根据该地区的税务数据,确定该地区的涉税基本信息标签,其中所述行业的涉税基本信息标签包括:地域名称、地域代码、所属税

务机关、税务机关代码、纳税人总户数、地域进项金额、地域进项税额、地域进项发票份数、地域销项金额、地域销项税额、地域销项发票份数；地域进项金额占比、地域进项金额排名、地域销项金额占比、地域销项金额排名；进项金额同比、进项金额环比、进项税额同比、进项税额占比；销项金额同比、销项金额环比、销项税额同比、销项税额占比等。并根据该地区的基本信息标签和发展标签，确定该地区的画像。在本发明实施例中，根据地区的税务数据确定地区的基础信息标签是现有技术，不再进行赘述。

[0107] 税务部门可以根据地区的画像，了解地区的发展状况和发展前景，对快速增长地区加大投入力度，对快速下降地区提供政策方面的扶持，还可以根据地区的画像，用于对宏观经济政策的事前、事中、事后的评估。

[0108] 此外，在本发明实施例中，还可以使用基于jQuery的UI插件集合体技术(EasyUI)对企业的画像、行业的画像、地区的画像及企业对应的涉税风险等级、行业 and 地区的发展趋势进行可视化展示。

[0109] 图2为本发明实施例提供的一种画像架构示意图，数据采集单元从税务系统和互联网中获取企业的税务数据和企业的经营数据，数据清洗单元和数据仓库单元对获取的企业的税务数据和经营数据进行清洗和存储，标签生成单元根据标签管理单元制定的标签生成规则，及企业的税务数据、网络销售信息，预测企业和地区的发展趋势，确定行业 and 地区的发展标签，企业画像单元、行业画像单元、地区画像单元，确定企业的画像、行业的画像、地区的画像，涉税风险预测单元预设企业的涉税风险等级，可视化展示单元，用于对企业的画像、行业的画像、地区的画像、企业的涉税风险等级、行业 and 地区的发展趋势进行可视化展示。

[0110] 实施例8：

[0111] 图3为本发明实施例提供的一种画像确定装置结构示意图，该装置包括：

[0112] 获取模块31，用于针对待画像的企业，获取税务业务系统中所述企业的税务数据，并获取互联网中所述企业的经营数据，其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息，所述经营数据包括企业的网络销售信息；

[0113] 判断确定模块32，用于根据所述企业的税务数据，生成企业的涉税基本信息标签，并根据所述企业的经营数据和所述企业的税务数据，判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配，并根据是否匹配，生成第一涉税风险标签；

[0114] 第一画像模块33，用于根据所述企业的涉税基本信息标签和第一涉税风险标签，确定所述企业的画像。

[0115] 所述判断确定模块32，还用于通过所述企业的税务数据，识别所述企业是否为小规模纳税人企业，如果所述企业为小规模纳税人企业，根据所述企业的销售金额是否大于预设的销售金额阈值，确定所述企业是否满足一般纳税人条件，并将所述企业是否满足一般纳税人条件确定为第一涉税风险标签。

[0116] 其中，所述判断确定模块32，还用于根据所述企业的进项发票信息和销项发票信息，确定所述企业的发票异常信息，并根据所述企业的发票异常信息，确定所述企业的第二涉税风险标签；

[0117] 所述第一画像模块33，具体用于根据所述企业的涉税基本信息标签、第一涉税风险标签和第二涉税风险标签，确定所述企业的画像。

[0118] 所述装置还包括:

[0119] 第一确定模块34,用于根据所述企业存在的涉税风险标签及预先训练完成的涉税风险模型,确定所述企业的涉税风险等级。

[0120] 所述装置还包括:

[0121] 训练模块35,用于针对训练集中的每个样本企业,根据该企业的涉税风险等级,为该样本企业存在的涉税风险标签添加对应的涉税风险等级标签,将添加涉税风险等级标签后的涉税风险标签输入到涉税风险模型,对涉税风险模型进行训练。

[0122] 所述装置还包括:

[0123] 清洗模块36,用于对所述企业的税务数据和经营数据进行数据清洗和格式转换。

[0124] 所述装置还包括:

[0125] 第二画像模块37,用于针对待画像的行业,获取税务系统中所述行业的税务数据,其中行业的税务数据包括行业的基本信息和行业中包含的企业的税务数据;根据所述行业的税务数据,确定所述行业的涉税基本信息标签,并根据所述行业中包含的企业的税务数据及趋势算法,确定所述行业的销项金额的发展趋势,并根据所述行业的销项金额的发展趋势,确定所述行业的行业发展标签,其中所述行业发展标签包括快速增长行业、快速下降行业、稳定发展行业、周期性行业;根据所述行业的基本信息标签和发展标签,确定所述行业的画像。

[0126] 所述装置还包括:

[0127] 第三画像模块38,用于针对待画像的地区,获取税务系统中所述地区的税务数据,其中地区的税务数据包括地区的基本信息和地区中包含的企业的税务数据;根据所述地区的税务数据,确定所述地区的涉税基本信息标签,并根据所述地区中包含的企业的税务数据及趋势算法,确定所述地区的销项金额的发展趋势,并根据所述地区的销项金额的发展趋势,确定所述地区的地区发展标签,其中所述地区发展标签包括快速增长地区、快速下降地区、稳定发展地区;根据所述地区的基本信息标签和发展标签,确定所述地区的画像。

[0128] 实施例9:

[0129] 在上述各实施例的基础上,本发明实施例还提供了一种电子设备,如图4所示,包括处理器401、通信接口402、存储器403和通信总线404,其中,处理器401,通信接口402,存储器403通过通信总线404完成相互间的通信;

[0130] 存储器403,用于存放计算机程序;

[0131] 处理器401,用于执行存储器403上所存放的计算机程序时,实现如下步骤:

[0132] 针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;

[0133] 根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;

[0134] 根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。

[0135] 上述雾计算设备提到的通信总线404可以是外设部件互连标准(Peripheral Component Interconnect,PCI)总线或扩展工业标准结构(Extended Industry Standard

Architecture, EISA) 总线等。该通信总线404可以分为地址总线、数据总线、控制总线等。为便于标识,图中仅用一条通信接口用于上述雾计算设备与其他设备之间的通信。

[0136] 存储器403可以包括随机存取存储器(Random Access Memory, RAM),也可以包括非易失性存储器(Non-Volatile Memory, NVM),例如至少一个磁盘存储器。可选的,存储器还可以是至少一个位于远离前述处理器的存储装置。

[0137] 上述的处理器401可以是通用处理器,包括中央处理器(Central Processing Unit, CPU)、互联网处理器(Network Processor, NP)等;还可以是数字信号处理器(Digital Signal Processing, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。

[0138] 实施例10:

[0139] 在上述各实施例的基础上,本发明实施例中还提供了一种计算机可读存储介质,所述计算机可读存储介质内存储有计算机程序,所述计算机程序被处理器执行如下步骤:

[0140] 针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;

[0141] 根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;

[0142] 根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。

[0143] 本发明公开了一种画像确定方法、装置、电子设备及存储介质,所述方法包括:针对待画像的企业,获取税务业务系统中所述企业的税务数据,并获取互联网中所述企业的经营数据,其中所述企业的税务数据包括所述企业的基本信息、销项发票信息、进项发票信息,所述经营数据包括企业的网络销售信息;根据所述企业的税务数据,生成企业的涉税基本信息标签,并根据所述企业的经营数据和所述企业的税务数据,判断所述企业的销售金额是否与所述企业的销项纳税金额相匹配,并根据是否匹配,生成第一涉税风险标签;根据所述企业的涉税基本信息标签和第一涉税风险标签,确定所述企业的画像。由于在本发明实施例中,针对待画像的企业,获取税务业务系统中企业的税务数据,并获取互联网中企业的经营数据,根据企业的税务数据,生成企业的涉税基本信息标签,并根据企业的经营数据和企业的税务数据,判断企业的销项金额是否与企业的销项纳税金额相匹配,生成第一涉税风险标签,根据企业的涉税基本信息标签和企业的的第一涉税风险标签,确定企业的画像,保证了企业的画像的真实性和全面性,提供了一种全面且真实反应企业税务形象的方案。

[0144] 对于系统/装置实施例而言,由于其基本相似于方法实施例,所以描述的比较简单,相关之处参见方法实施例的部分说明即可。

[0145] 本领域内的技术人员应明白,本申请的实施例可提供为方法、系统、或计算机程序产品。因此,本申请可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本申请可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0146] 本申请是参照根据本申请实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0147] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0148] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0149] 尽管已描述了本申请的优选实施例,但本领域内的技术人员一旦得知了基本创造性概念,则可对这些实施例做出另外的变更和修改。所以,所附权利要求意欲解释为包括优选实施例以及落入本申请范围的所有变更和修改。

[0150] 显然,本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围。这样,倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内,则本发明也意图包含这些改动和变型在内。



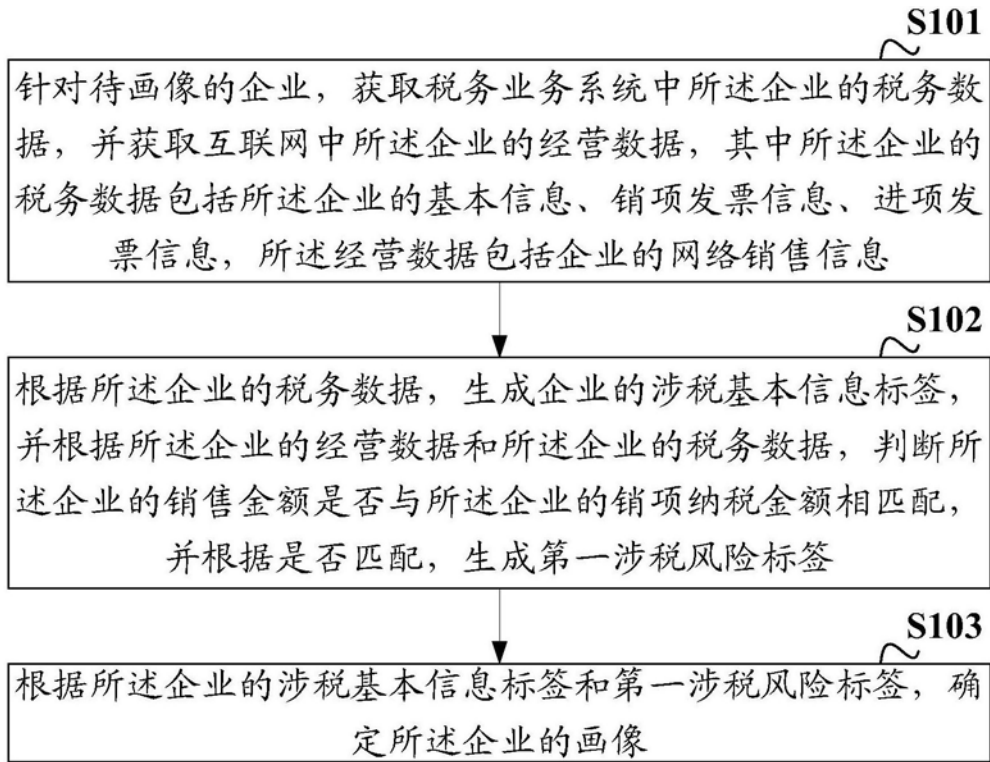


图1

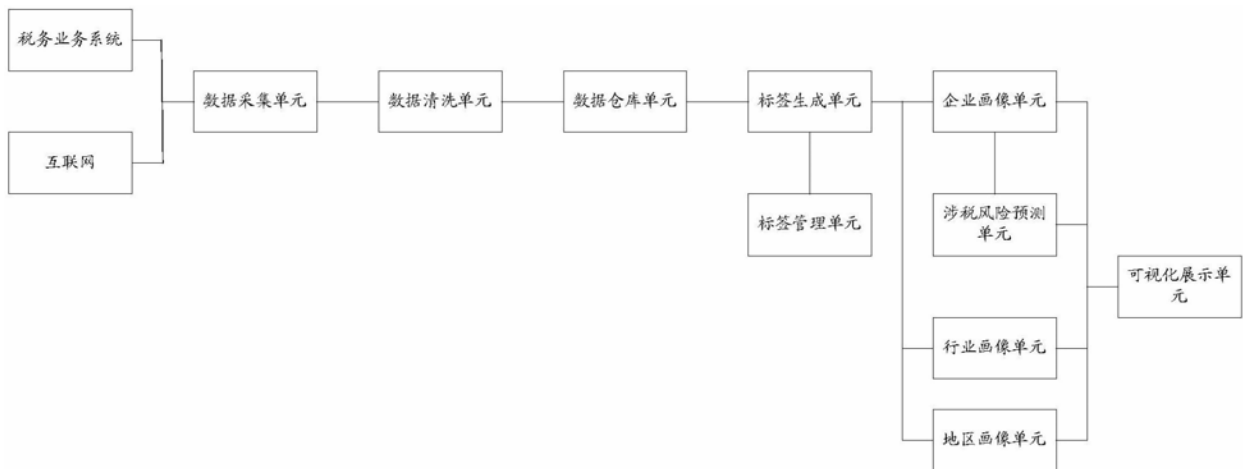


图2

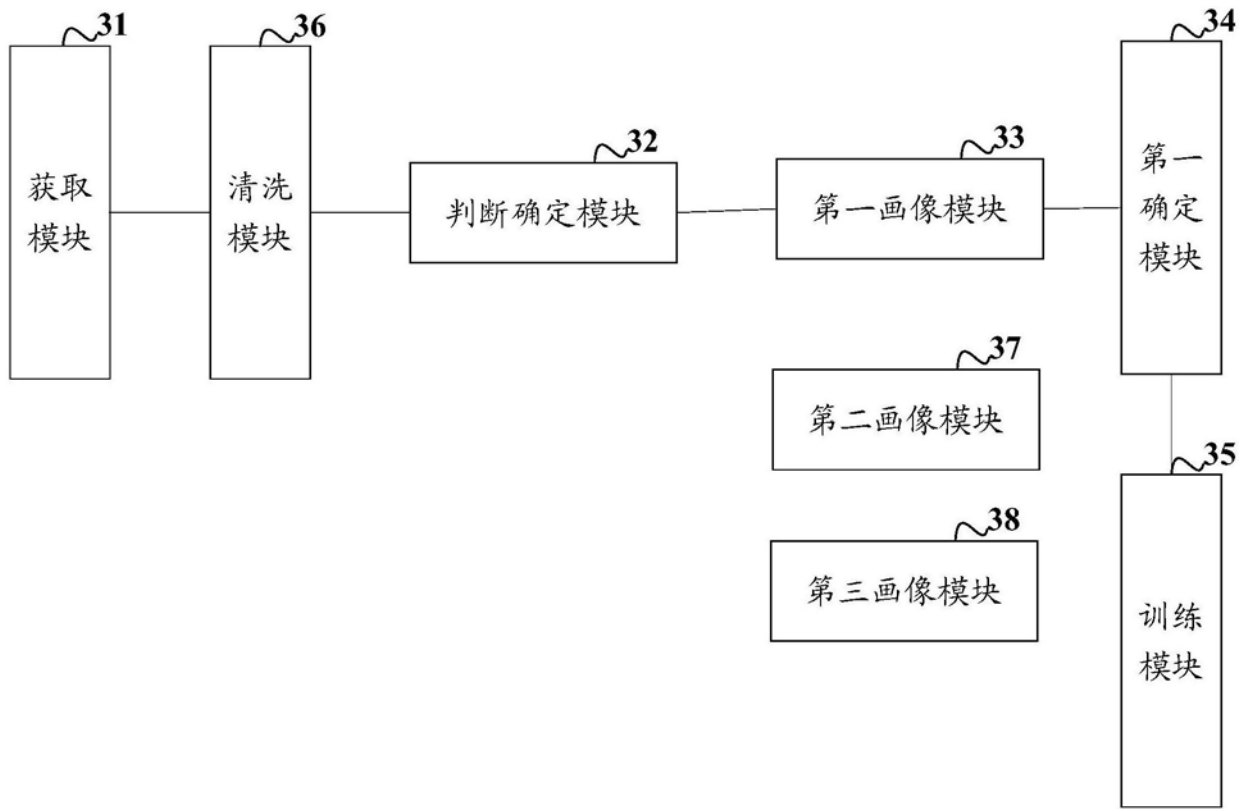


图3

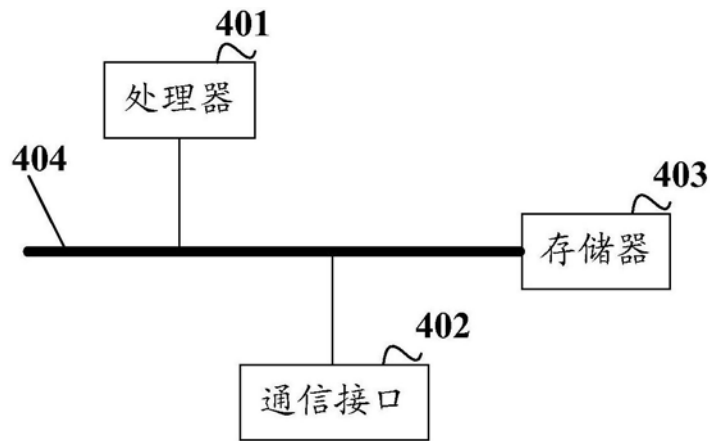


图4