



(12) 发明专利申请

(10) 申请公布号 CN 105630785 A

(43) 申请公布日 2016. 06. 01

(21) 申请号 201410583856. 6

(22) 申请日 2014. 10. 27

(71) 申请人 航天信息股份有限公司

地址 100097 北京市海淀区杏石口路甲 18 号

(72) 发明人 李立宁 黄新华 张玉魁 尹春天
许芳函 范立波 杨为琛

(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 李勇

(51) Int. Cl.

G06F 17/30(2006. 01)

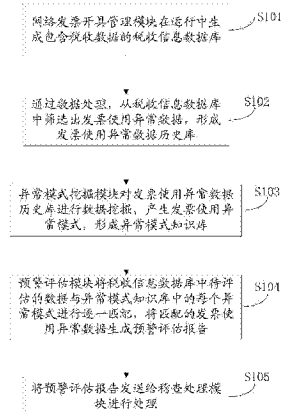
权利要求书2页 说明书12页 附图4页

(54) 发明名称

一种发票使用异常预警方法和系统

(57) 摘要

本发明公开了一种发票使用异常预警方法和系统,属于发票管理领域,包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,网络发票开具管理模块生成税收信息数据库;通过数据处理,筛选出发票使用异常数据,形成发票使用异常数据历史库;异常模式挖掘模块通过数据挖掘产生发票使用异常模式,形成异常模式知识库;预警评估模块将税收信息数据库中待评估的数据与异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告;将预警评估报告发送给稽查处理模块进行处理。本发明挖掘得到发票异常使用模式知识库,对发票使用异常数据实时预警和评估,使发票管理手段智能化,提高发票管理效率。



1. 一种发票使用异常预警方法,包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,其特征在于,包括以下步骤:

所述网络发票开具管理模块在运行中生成包含税收数据的税收信息数据库;

通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库;

所述异常模式挖掘模块对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库;

所述预警评估模块将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告;

将所述预警评估报告发送给所述稽查处理模块进行处理。

2. 根据权利要求1所述的发票使用异常预警方法,其特征在于:所述网络发票开具管理模块包括发票票源领购单元、发票开具单元和发票查验单元,所述发票查验单元用于发票的查询验证和举报异常发票。

3. 根据权利要求1或2所述的发票使用异常预警方法,其特征在于:所述形成发票使用异常数据历史库的步骤中,具体包括:

从所述税收信息数据库中的税收数据中筛选出发票使用异常数据;

对所述发票使用异常数据进行预处理;

将所述预处理后的发票使用异常数据进行转换;

形成所述发票使用异常数据历史库。

4. 根据权利要求1至3任一所述的发票使用异常预警方法,其特征在于:所述数据挖掘包括序列模式挖掘、时间序列分析、关联规则、聚类分析算法。

5. 根据权利要求1至4任一所述的发票使用异常预警方法,其特征在于:所述形成异常模式知识库的步骤还包括:

通过异常模式添加单元对所述数据挖掘产生的发票使用异常模式进行判断,如果确认为异常模式,则添加到所述异常模式知识库中;

如果判断为非异常模式,则通过异常模式删除单元将其删除。

6. 根据权利要求1至5任一所述的发票使用异常预警方法,其特征在于:所述预警评估模块包括数据检索单元、预警评估报告生成单元和预警报告发送单元。

7. 根据权利要求6所述的发票使用异常预警方法,其特征在于还包括:所述数据检索单元对检索匹配满足多个异常模式的发票使用异常数据进行筛查分析。

8. 一种发票使用异常预警系统,其特征在于:包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,其中:

所述网络发票开具管理模块适用于在运行中生成包含税收数据的税收信息数据库,通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库;

所述异常模式挖掘模块适用于对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库;

所述预警评估模块适用于将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告,将

所述预警评估报告发送给所述稽查处理模块进行处理；

所述稽查处理模块适用于根据所述预警评估报告，将发票使用异常数据中的发票和纳税人数据信息进行跟踪处理和反馈。

9. 根据权利要求 8 所述的发票使用异常预警系统，其特征在于：所述网络发票开具管理模块包括发票票源领购单元、发票开具单元和发票查验单元，所述发票查验单元用于发票的查询验证和举报异常发票。

10. 根据权利要求 8 或 9 所述的发票使用异常预警系统，其特征在于：所述异常模式挖掘模块包括发票使用异常模式产生单元、发票使用异常模式添加单元和发票使用异常模式删除单元。

11. 根据权利要求 8 至 10 任一所述的发票使用异常预警系统，其特征在于：所述数据挖掘包括序列模式挖掘、时间序列分析、关联规则、聚类分析算法。

12. 根据权利要求 8 至 11 任一所述的发票使用异常预警系统，其特征在于：所述预警评估模块包括数据检索单元、预警评估报告生成单元和预警报告发送单元。

一种发票使用异常预警方法和系统

技术领域

[0001] 本发明涉及发票管理技术领域,具体涉及一种发票使用异常预警方法和系统。

背景技术

[0002] 发票管理难题一直困扰着基层税务部门,在发票开具和使用过程中,存在大量制假售假、虚开代开、多征少征、以税谋私等各种违法违规行。这些问题已经影响到税收的公平公正。如何解决这些问题,全面、准确记录纳税人经营活动,有效的掌握税源、监控发票使用、做好发票管理是税收征管的重要环节。

发明内容

[0003] 鉴于上述问题,提出了本发明以便提供一种克服上述问题或者至少部分地解决上述问题的发票使用异常预警方法和系统,可有效发现并解决纳税人领购发票、发票开具、发票流通过程中存在的异常用票问题,对各类发票异常数据提供预警。

[0004] 依据本发明的一个方面,提供了一种发票使用异常预警方法,包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,包括以下步骤:

[0005] 所述网络发票开具管理模块在运行中生成包含税收数据的税收信息数据库;

[0006] 通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库;

[0007] 所述异常模式挖掘模块对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库;

[0008] 所述预警评估模块将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告;

[0009] 将所述预警评估报告发送给所述稽查处理模块进行处理。

[0010] 进一步地,所述网络发票开具管理模块包括发票票源领购单元、发票开具单元和发票查验单元,所述发票查验单元用于发票的查询验证和举报异常发票。

[0011] 进一步地,所述形成发票使用异常数据历史库的步骤中,具体包括:

[0012] 从所述税收信息数据库中的税收数据中筛选出发票使用异常数据;

[0013] 对所述发票使用异常数据进行预处理;

[0014] 将所述预处理后的发票使用异常数据进行转换;

[0015] 形成所述发票使用异常数据历史库。

[0016] 进一步地,所述数据挖掘包括序列模式挖掘、时间序列分析、关联规则、聚类分析算法。

[0017] 更进一步地,所述形成异常模式知识库的步骤还包括:

[0018] 通过异常模式添加单元对所述数据挖掘产生的发票使用异常模式进行判断,如果确认为异常模式,则添加到所述异常模式知识库中;

[0019] 如果判断为非异常模式,则通过异常模式删除单元将其删除。

[0020] 更进一步地,还包括:所述数据检索单元对检索匹配满足多个异常模式的发票使用异常数据进行筛查分析。

[0021] 依据本发明的另一个方面,提供了一种发票使用异常预警系统,包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,其中:

[0022] 所述网络发票开具管理模块适用于在运行中生成包含税收数据的税收信息数据库,通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库;

[0023] 所述异常模式挖掘模块适用于对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库;

[0024] 所述预警评估模块适用于将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告,将所述预警评估报告发送给所述稽查处理模块进行处理;

[0025] 所述稽查处理模块适用于根据所述预警评估报告,将发票使用异常数据中的发票和纳税人数据信息进行跟踪处理和反馈。

[0026] 进一步的,所述网络发票开具管理模块包括发票票源领购单元、发票开具单元和发票查验单元,所述发票查验单元用于发票的查询验证和举报异常发票。

[0027] 进一步的,所述异常模式挖掘模块包括发票使用异常模式产生单元、发票使用异常模式添加单元和发票使用异常模式删除单元。

[0028] 进一步的,所述数据挖掘包括序列模式挖掘、时间序列分析、关联规则、聚类分析算法。

[0029] 进一步的,所述预警评估模块包括数据检索单元、预警评估报告生成单元和预警报告发送单元。

[0030] 本发明具有以下有益效果:本发明提出了一种新型的发票使用异常预警的方法和系统,依托网络发票开具管理模块的海量数据,动态挖掘发票使用异常模式,对发票使用异常数据进行识别和预警,有效的解决了当出现各种发票使用异常情况时,无法使用单一的模式识别发票使用异常数据,使得预警识别更具针对性,也更具灵活性。

[0031] 上述说明仅是本发明技术方案的概述,为了能够更清楚了解本发明的技术手段,而可依照说明书的内容予以实施,并且为了让本发明的上述和其它目的、特征和优点能够更明显易懂,以下特举本发明的具体实施方式。

附图说明

[0032] 通过阅读下文优选实施方式的详细描述,各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。说明书附图仅用于示出优选实施方式的目的,而并不认为是对本发明的限制。显而易见地,下面描述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。而且在整个附图中,用相同的参考符号表示相同的部件。在附图中:

[0033] 图1示出了根据本发明实施例一的发票使用异常预警方法流程图;

[0034] 图2示出了基于图1所示的发票使用异常预警方法中的异常模式挖掘模块功能实现流程图;

[0035] 图 3 示出了基于图 1 所示的发票使用异常预警方法中的预警评估模块功能实现流程图。

[0036] 图 4 示出了根据本发明实施例四的发票使用异常预警系统框图。

具体实施方式

[0037] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例,然而应当理解,可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反,提供这些实施例是为了能够更透彻地理解本公开,并且能够将本公开的范围完整的传达给本领域的技术人员。

[0038] 需要说明的是,在说明书及权利要求当中使用了某些词汇来指称特定组件。本领域技术人员应可以理解,硬件制造商可能会用不同名词来称呼同一个组件。本说明书及权利要求并不以名词的差异来作为区分组件的方式,而是以组件在功能上的差异来作为区分的准则。如在通篇说明书及权利要求当中所提及的“包含”或“包括”为一开放式用语,故应解释成“包含但不限于”。说明书后续描述为实施本发明的较佳实施方式,然所述描述乃以说明书的一般原则为目的,并非用以限定本发明的范围。本发明的保护范围当视所附权利要求所界定者为准。

[0039] 为了加强发票管理,提高税收征管效率,保证税收的公平公正,本发明公开了一种发票使用异常预警方法和系统。该方法或系统主要利用税收信息管理系统的数据,根据已知的发票使用异常数据筛选分类,挖掘出异常模式知识库,对不符合逻辑的异常数据实时检索分析,并进行匹配和预警,并对异常数据进行跟踪调查,直至税务人员完成巡查和确认,对异常模式进行反馈。

[0040] 为便于对本发明实施例的理解,下面将结合附图以几个具体实施例为例做进一步的解释说明,且各个附图并不构成对本发明实施例的限定。

[0041] 本发明是一种对发票异常开具或使用行为的挖掘,将近期税收信息数据库的数据进行数据挖掘和模式匹配,及时对发票异常使用数据进行预警和评估。本发明可分为四个子模块:网络发票开具管理模块,异常模式挖掘模块,预警评估模块,稽查处理模块。

[0042] 所述方法采用数据挖掘的方法,循环不断分析归纳发票使用的异常模式,采用模式匹配的方法,对近期的税收信息检索,识别发票使用的异常数据。网络发票开具管理模块是生产系统,产生大量的税收发票数据,形成税收信息数据库。发票异常使用数据历史库是税收信息数据库中已经被系统认定为发票异常使用的数据,数据经过长时间的积累,具有各式各样表现模式。异常模式挖掘模块通过对异常使用数据历史库的数据挖掘分析,通过各种方法,比如聚类分析、时间序列分析、关联分析、序列模式挖掘,找出各种发票使用异常的模式,形成异常模式知识库。所述数据挖掘是指:数据挖掘是指从数据库的大量数据中揭示出隐含的、先前未知的并有潜在价值的信息的非平凡过程。预警评估模块可以根据异常模式知识库中每个异常模式,对近期的税收信息数据库进行分析匹配,形成预警信息和评估报告,并将其发送给稽查处理模块。稽查处理模块是决策处理模块,由稽查人员对异常数据进行追踪确认,决定对每个异常模式采取什么措施,并对发现的异常模式反馈,验证和维护异常模式并对其进行新增、删除、调整,保证预警系统的高适应性。

[0043] 本发明基于知识发现的发票使用异常预警系统,它使得识别具有高度适应的特

点。在面对发票使用异常模式方法变化时,很多发票预警方法,只考虑当前的发票异常使用情况,而忽略了随着技术或者政策的发展虽然能够阻止和监控一些发票异常的产生,但是同时又有一些新的发票异常使用的手段和行为衍生。因此,本方案引入数据挖掘概念,发票异常使用监控不局限于某一种或某些发票使用的异常模式上,而是通过对历史数据的分析、知识识别、人工配置即可定制出具有针对性的预警方案,从而提高方案可适应性和扩展性。

[0044] 实施例一、发票使用异常预警方法。

[0045] 下面结合图 1 对本实施例一的方法进行详细说明。

[0046] 图 1 为本发明实施例一的发票使用异常预警方法流程图,如图 1 所示,本发明实施例包括网络发票开具管理模块、异常模式挖掘模块、预警评估模块和稽查处理模块,包括以下步骤:

[0047] 步骤 S101、所述网络发票开具管理模块在运行中生成包含税收数据的税收信息数据库。

[0048] 具体的,所述网络发票开具管理模块包括发票票源领购单元、发票开具单元和发票查验单元,所述发票查验单元用于发票的查询验证和举报异常发票。

[0049] 所述网络发票开具管理模块是本实施例的基础,只有大力推行机打发票,推行网络开具,将所有的发票信息纳入税收管理信息系统,为发票使用人快速网络实时查询、辨识开具信息的真伪,为发票使用异常预警奠定基础。该模块是税收信息数据库数据的生产模块,包括发票票源领购单元、发票开具单元以及发票查验单元。所述发票票源领购单元产生发票票源的分配分发数据,包括每个纳税人的能开具的发票种类以及领购过的发票票源结存。发票开具产生大量的发票数据,包括正票、负票、废票,其中每张发票中有详细的发票元素数据,如合计金额、开票单位等。所述发票查验单元产生发票的查询验证和举报异常发票的数据,包括每张发票的查询次数、验证发票是否正确、举报不正确的发票。如举报发票开具金额纸质票和电子票不符、举报纸质发票票样和电子票不符等。

[0050] 所述网络发票开具管理模块在正常运行中,会产生海量的税收数据,将这些数据定义为税收信息数据库,其中大部分的数据都是合法的符合正常逻辑的税收信息。

[0051] 步骤 S102、通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库。

[0052] 具体的,所述形成发票使用异常数据历史库的步骤中,具体包括:

[0053] 从所述税收信息数据库中的税收数据中筛选出发票使用异常数据;

[0054] 对所述发票使用异常数据进行预处理;

[0055] 将所述预处理后的发票使用异常数据进行转换;

[0056] 形成所述发票使用异常数据历史库。

[0057] 本发明实施例中优选的,异常数据的原始库是要保存的,而要产生异常数据模式,需要对数据做处理,发票相关数据的数据信息非常多,很多数据项是发票异常预警不关心的,只选择构成发票异常的关键数据进行知识学习和模式挖掘。因此,所述发票使用异常数据的预处理和转换需要把异常数据的原始数据转化为处理算法识别的数据集合。具体的实现步骤包括:

[0058] 提取构成异常模式的关键数据项;

[0059] 和具体处理算法适配,将原始数据转化为处理算法识别的数据集合。

[0060] 所述发票使用异常模式产生单元中,所述异常模式的产生首先要准备数据:所述税收信息数据库包含了海量的税收数据,发票使用异常的数据也存在其中。本实施例将已知的发票使用异常数据筛选出来,筛选过程不是一次完成的,是经过不断异常模式的挖掘,实时检索税收信息数据库,发现发票使用异常数据,形成并更新发票使用异常数据历史库。

[0061] 步骤 S103、所述异常模式挖掘模块对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库。

[0062] 本发明实施例中,异常模式挖掘模块包括 3 个功能单元:发票使用异常模式产生单元、发票使用异常模式添加单元、发票使用异常模式删除单元。

[0063] 本实施例中优选的,所述异常模式的产生通过两种方式。

[0064] 第一种是通过税务人员的经验,政策,或者数据统计得到的异常模式。另一种是通过对所述发票使用异常数据历史库的数据分析来挖掘产生发票使用异常模式。所有的异常模式组成所述异常模式知识库。

[0065] 定义(发票使用异常模式):<异常类型、算法、异常表示公式>,每一个三元组代表一种异常模式。其中,所述异常类型:表示发票使用异常的类型或者名称。所述算法:是监测发票异常需要匹配使用的算法。所述异常表示公式:异常模式的表现形式,可以通过算法监测数据异常的异常模式公式。

[0066] 第一种异常模式包括农产品收购发票抵扣进项占比过高、进项税额变动与销项税额变动弹性异常、企业所得税申报与增值税申报比对异常等等。

[0067] 具体的,在本发明实施例中进行举例说明:农产品收购发票抵扣进项占比过高。农产品收购凭证抵扣进项分析主要通过当期有农产品收购凭证抵扣进项的纳税人当期农产品收购凭证抵扣进项税额占当期全部抵扣进项税额的比例分析,分析纳税人农产品收购凭证抵扣进项的合理性,对比例高于同行业预警值 w 的,可能存在虚开农产品收购凭证、虚抵进项税额的纳税人进行预警。

[0068] 其中,农产品收购凭证抵扣进项占全部进项比例 $f(x) = \text{当期农产品收购凭证进项税额} / (\text{当期全部进项抵扣税额} - \text{当期全部固定资产抵扣进项税额}) * 100\%$

[0069] 具体的,农产品收购发票抵扣进项占比过高的异常模式可以表示为:

[0070] $\langle \text{Model}, \text{Algorithm1}, f(x) \rangle w$

[0071] 其中,所述 Model: 进项税额变动与销项税额变动弹性异常模式;

[0072] 所述 Algorithm1: 实现农产品收购凭证抵扣进项占全部进项比例公式计算相匹配的算法。

[0073] 第二种异常模式通过序列模式挖掘、关联规则、聚类分析、时间序列分析等算法挖掘产生的异常模式。每种算法会产生若干个发票使用异常模式。所述关联规则分析是指:关联分析就是从大量数据中发现项集之间有趣的关联和相关联系。所述聚类分析是指将数据分类到不同的类或者簇这样的一个过程,所以同一个簇中的对象有很大的相似性,而不同簇间的对象有很大的相异性。其中,所述时间序列分析是指:时间序列是按时间顺序的一组数字序列,时间序列分析就是利用这组数列,应用数理统计方法加以处理,以预测未来事物的发展。

[0074] 本实施例中优选的,采用序列模式挖掘对发票操作序列进行挖掘。例如,发现不同

的纳税人多次执行发票查询操作,发票可以认定为一种备选发票使用异常模式,可能一张电子票对应多张纸质票。其次,如果发现一张发票查询操作,紧接着是发票举报操作,可以认定为是另一种备选发票使用异常模式。

[0075] 具体的,首先通过原始数据查询统计、预处理转换将近期的发票数据转化为每张发票操作的序列数据;其次,设定序列模式类的发票异常模式有:

[0076] 1. $\{*, a \text{ 纳税人对发票 } I \text{ 做发票查验操作、} \dots、 b \text{ 纳税人对发票 } I \text{ 发票查验操作}, *\}$ 即多个纳税人对同一张发票进行发票查验操作,此模式可能是一票多开发票异常。

[0077] 2. $\{*, \text{ 发票举报}, *\}$,只要包含发票举报的操作序列就认定该张发票为异常发票,此模式可能是发票关键要素不符的发票异常。如一段时间里对发票代码 000000000001 发票号码为 00000001 的发票的操作序列为 { 发票票源下载、发票开具、发票查验 },那么这种发票就是异常发票。

[0078] 然后通过序列模式挖掘算法,将最近一段发票操作序列数据进行发票数据模式匹配检索,发现近期的异常发票。

[0079] 其中,所述的发票相关操作包括发票票源下载、发票开具、发票查验、发票举报等。所述发票操作序列,就是在一定时间内对每张发票的操作集合。为了表示方便,我们用字母表示发票操作。

[0080] 定义(发票操作序列): $I = \{i_1, i_2 \dots i_m\}$ 为发票操作项集,其中 $i_k, 1 < k < m$ 是一个发票操作。发票操作序列 $S = \langle s_1, s_2 \dots s_n \rangle$,其中 $s_j (1 < j < n)$ 为项集,即 $s_j \in I$ 。序列中每个元素由不同的项组成。序列也可表示为 $S = \langle i_1, i_2, \dots i_k \rangle$ 。序列包含所有项的个数成为序列的长度。长度为 L 的序列记为 L -序列。

[0081] 定义(发票操作子序列):序列 $T = \langle t_1, t_2 \dots t_m \rangle$ 是一个序列 $S = \langle s_1, s_2, \dots s_n \rangle$ 的子序列,需要满足对于任意 $i (1 < i < m-1)$, 存在 $k_1 (1 < k_1 < n)$ 满足 $t_i = s_{k_1}$, 对于任意 $i+1 (1 < i < m-1)$, 存在 $k_2 (1 < k_2 < n)$ 满足 $t_{i+1} = s_{k_2}$, 且 $k_1 < k_2$ 。即序列 S 包含序列 T , 表示为 $T \subseteq S$ 称序列 T 为序列 S 的子序列。若一个序列 S 不包含在其他的序列当中,则称序列 S 是最大的。

[0082] 定义(序列支持度):序列数据库 D 是元组 $\langle \text{sid}, S \rangle$ 的集合, sid 是序列标识号,如果序列 T 是 S 的子序列(即 $T \subseteq S$),称元组 $\langle \text{sid}, S \rangle$ 包含序列 T 。序列 T 在数据库 D 中支持度,是数据库 D 中包含序列 T 的元组数,即 $\text{Support}(T) = |\{ \langle \text{sid}, S \rangle \mid \langle \text{sid}, S \rangle \in D \wedge T \subseteq S \}|$ 。

[0083] 定义(频繁序列模式):给定正整数 ϵ 为支持度阈值,如果数据库 D 中最少有 ϵ 个元组包含序列 S ,则称序列 S 为数据库 D 中的一个(频繁)序列模式。长度为 L 的序列模式成为 L -模式。

[0084] 所述发票操作序列模式挖掘的任务是找出数据库中所有的序列模式即那些发票操作序列集合中出现频率超过最小支持度的子序列(用户指定最小支持度阈值)。

[0085] 下面具体举例如下:

[0086] 假设发票操作序列数据库 D 为:

[0087] 1. $\langle a, e, b, c, d \rangle$;

[0088] 2. $\langle a, c, d, c, e \rangle$;

[0089] 3. $\langle a, b, c, d \rangle$;

[0090] 4. $\langle a, c, e \rangle$;

[0091] 5. $\langle d, e \rangle$ 。

[0092] 设定支持度阈值 $\epsilon = 2$ 来检索最大的频繁序列,挖掘过程如下:

[0093] L1

[0094]

1- 序列	支持度
$\langle a \rangle$	4
$\langle b \rangle$	2
$\langle c \rangle$	4
$\langle d \rangle$	4
$\langle e \rangle$	4

[0095]

[0096] L2

[0097]

2- 序列	支持度
$\langle a, b \rangle$	2
$\langle a, c \rangle$	4
$\langle a, d \rangle$	3
$\langle a, e \rangle$	3
$\langle b, c \rangle$	2
$\langle b, d \rangle$	2
$\langle c, d \rangle$	3
$\langle c, e \rangle$	2
$\langle e, f \rangle$	2

[0098] L3

[0099]

3- 序列	支持度
$\langle a, b, c \rangle$	2

<a, b, d>	2
<a, c, d>	3
<a, c, e>	2
<b, c, d>	2

[0100] L4

[0101]

4- 序列	支持度
<a, b, c, d>	4

[0102] 最大的频繁序列为

[0103]

序列	支持度
<a, b, c, d>	2
<a, c, e>	2
<d, e>	2

[0104] 本发明实施例采用 SPADE 序列模式挖掘算法 (参见 SPADE:an efficient algorithm for mining frequent sequences ZAKI M J,2001,42(01),Machine Learning) 对发票使用异常数据历史库中的发票操作序列数据,进行频繁序列模式挖掘,获得发票使用异常数据的频繁序列模式,并将得到的频繁序列模式表示为异常模式,如果有多个最大频繁序列,则有多个发票使用异常模式。

[0105] 在上述具体举例中,可以表示为下面的形式:

[0106] <Mode2, Algorithm2, T<a, b, c, d>>

[0107] <Mode2, Algorithm2, T<a, c, e>>

[0108] <Mode2, Algorithm2, T<d, e>>

[0109] 其中, Mode2 表示为 :发票使用异常序列模式 M1 ;

[0110] Algorithm2 :SubSequence 最大频繁序列匹配算法,检测最大频繁序列模式是否是待检索发票操作序列的子序列;

[0111] T<t1, t2, ... tm> 表示最大频繁序列,而 S<s1, s2, ... sn> 表示为待检测发票操作序列,具体程序语言如下:

[0112]

```

    BOOL SubSequence(T, S)
    {
        Int M=0;//记录上一次比对项在 S 中位置
        FOR (i=1;i<=m;i=i+1)
        {
            If (  $t_i \in T$ , 存在  $s_k \in S$  满足  $t_i=s_k$  并且  $k>M$ )
            [0113]
            {
                M=k;
                Continue;
            }
            Return False;
        }
        Return TRUE;
    }

```

[0114] 通过税务人员的经验,政策,或者数据统计结果,或者通过各类数据分析算法,对异常历史数据库进行知识发现,产生所有的备选的发票使用异常模式。

[0115] 本实施例优选的,所述形成异常模式知识库的步骤还包括:

[0116] 通过所述发票使用异常模式添加单元对所述数据挖掘产生的发票使用异常模式进行判断,如果确认为异常模式,则添加到所述异常模式知识库中;

[0117] 如果判断为非异常模式,则通过所述发票使用异常模式删除单元将其删除。

[0118] 其中所述发票使用异常模式添加单元,对上述数据挖掘产生的备选发票使用异常模式由决策人员进行决策,将其确认为异常模式,并添加到所述异常模式知识库中。

[0119] 所述发票使用异常模式删除单元,是将所述数据挖掘产生的不合理的备选异常模式由决策人员,将其删除,或者由于政策变化,一些原本比较合理的异常模式,现在变的不合理了,也可以由所述发票使用异常模式删除单元将其删除。

[0120] 步骤 S104、所述预警评估模块将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估报告。

[0121] 本实施例优选的,还包括:所述数据检索单元对检索匹配满足多个异常模式的发

票使用异常数据进行筛查分析。

[0122] 所述预警评估模块包括数据检索单元、预警评估报告生成单元和预警报告发送单元。

[0123] 所述数据检索单元是根据所述异常模式挖掘模块产生的所述异常模式知识库,对所述税务信息数据库近期的数据进行检索,根据每个异常模式对近期税务信息数据进行逐一匹配。对检索匹配的数据进行筛查分析,因为有些数据满足多个异常模式。所述筛查分析是指:通过异常模式知识库中的每个异常模式以及相应适配的数据挖掘算法,对目标数据进行分析,甄别符合异常模式的数据。本发明实施例优选的,还可以设定检索周期,定期的发现近期发票使用异常的数据,排重处理后,对发票使用异常数据做好标记,为产生预警评估报告做好数据准备。这里所述的排重处理是指:有可能多个异常模式都指向一张发票或者一个纳税人,那么将这些发票异常数据进行汇总排重处理。

[0124] 所述预警评估报告生成单元,是将检索发现的近期所有的发票使用异常数据进行汇总,统计使用异常的所有发票,识别发票使用异常的开票方和售票方,列出开具异常发票的纳税人的开票量。按照异常模式对近期发票使用异常进行分类统计,并且按照发票数量进行排序,识别出近期各类发票使用异常的现象的严重程度。

[0125] 步骤 S105、将所述预警评估报告发送给所述稽查处理模块进行处理。

[0126] 所述稽查处理模块是根据所述预警评估报告,将发票使用异常的发票和纳税人数据信息,分发给各个所属地区的税收管理员和稽查员。税务人员根据预警数据信息与相关的企业进行巡查和跟踪。确定每个发票数据是否是异常发票,如果确定是发票使用异常,根据相关法律法规,对相关人员进行处罚,经税务人员确认后此发票异常使用数据将存入所述发票使用异常数据历史库中。经过调查,如果所述发票使用异常数据是正常的发票使用逻辑,那么根据税务人员反馈,对该异常模式规则进行调整或删除。这样既保证了预警方法或系统的准确性,也提高了预警方法或系统的适应性。

[0127] 本实施例采用数据挖掘方法对发票异常的所有历史数据进行选择、预处理、转换,形成数据挖掘的元数据,根据各种数据分析算法,对发票异常数据进行聚类、异常操作序列模式识别、关联规则分析,同时结合税务人员的先验知识,形成发票异常使用的异常模式知识库。根据异常模式知识库,对近期的税收信息数据进行检索匹配,发现发票使用异常数据,并根据异常模式,对近期的数据汇总和计算,形成预警评估报告。预警信息按照不同的级别分发给不同的税务人员或稽查人员,他们对数据进行跟踪,巡查,确认异常数据是否属实,并对挖掘的异常模式规则进行反馈。

[0128] 本实施例可有效发现并解决纳税人领购发票、发票开具和发票流通过程中存在的异常用票问题,对各类发票异常数据提供预警,由以往被动依赖“以票控税”的管理模式,变为根据发票预警数据进行主动跟踪的“信息控税”的管理模式,杜绝发票管理漏洞,提升发票管理效率,提高税收征管质量。

[0129] 实施例二、所述发票使用异常预警方法中的异常模式挖掘模块功能实现方法。

[0130] 下面结合图 2 对本实施例的方法进行详细说明。

[0131] 图 2 为基于图 1 所示的发票使用异常预警方法中的异常模式挖掘模块功能实现流程图,如图 2 所示,本实施例中的异常模式挖掘模块功能实现方法包括以下步骤:

[0132] 步骤 S201、对所述税收信息数据库中的数据进行选择、预处理和转换。

[0133] 步骤 S202、形成所述发票使用异常数据历史库。

[0134] 步骤 S203、通过所述异常模式挖掘模块中的所述发票使用异常模式产生单元、发票使用异常模式添加单元和发票使用异常模式删除单元对所述发票使用异常数据历史库中的数据进行操作,形成所述异常模式知识库。

[0135] 具体的实现细节如上述实施例一中所述,在此不再赘述。

[0136] 本实施例中的方法使用所述发票使用异常模式挖掘模块,可以对异常发票数据的特征进行提取、量化,形成所述异常模式知识库,可以有效量化异常发票监测的标准和依据。

[0137] 实施例三、基于图 1 所示的发票使用异常预警方法中的预警评估模块功能实现流程图,如图 3 所示,包括以下步骤:

[0138] 步骤 S301、开始预警评估。

[0139] 步骤 S302、对所述税收信息数据库中待评估的数据进行检索。

[0140] 步骤 S303、从所述异常模式知识库中载入一个异常模式。

[0141] 步骤 S304、从所述待评估的数据中定位一个待评估数据。

[0142] 步骤 S305、判断所述数据是否匹配该异常模式;如果匹配,则进入下一步;反之,跳转到步骤 S307。

[0143] 步骤 S306、保存所述匹配的异常数据,并标记该异常数据的异常模式类型。

[0144] 步骤 S307、判断所述异常模式知识库中还有未载入的异常模式,如果是,则跳回到步骤 S303;反之,进入下一步。

[0145] 步骤 S308、判断是否还有未评估的数据,如果有,则跳回到步骤 S302;反之,进入下一步。

[0146] 步骤 S309、对匹配的异常数据进行排重处理,因为有的数据会同时满足多个异常模式。

[0147] 步骤 S310、生成预警评估报告。

[0148] 步骤 S311、将所述预警评估报告发送到所述稽查处理模块。

[0149] 具体的实现细节如上述实施例一中所述,在此不再赘述。

[0150] 本实施例实现的所述预警评估模块,可以实时动态的监测产生的异常发票数据,并及时做出异常发票预警评估报告,使异常发票数据被自动筛选出来,从而使以票管税向信息管税转变。

[0151] 实施例四、发票使用异常预警系统。

[0152] 如图 4 所示,本实施例发票使用异常预警系统,包括网络发票开具管理模块 401、异常模式挖掘模块 403、预警评估模块 404 和稽查处理模块 405,其中:

[0153] 所述网络发票开具管理模块 401 适用于在运行中生成包含税收数据的税收信息数据库,通过数据处理,从所述税收信息数据库中筛选出发票使用异常数据,形成发票使用异常数据历史库 402;

[0154] 所述异常模式挖掘模块 403 适用于对所述发票使用异常数据历史库进行数据挖掘,产生发票使用异常模式,形成异常模式知识库;

[0155] 所述预警评估模块 404 适用于将所述税收信息数据库中待评估的数据与所述异常模式知识库中的每个异常模式进行逐一匹配,将匹配的发票使用异常数据生成预警评估

报告,将所述预警评估报告发送给所述稽查处理模块进行处理;

[0156] 所述稽查处理模块 405 适用于根据所述预警评估报告,将发票使用异常数据中的发票和纳税人数据信息进行跟踪处理和反馈。

[0157] 本实施例中优选的,所述网络发票开具管理模块 401 包括发票票源领购单元、发票开具单元和发票查验单元,所述发票查验单元用于发票的查询验证和举报异常发票。

[0158] 本实施例中优选的,所述异常模式挖掘模块 403 包括发票使用异常模式产生单元、发票使用异常模式添加单元和发票使用异常模式删除单元。

[0159] 本实施例中优选的,所述数据挖掘包括序列模式挖掘、关联规则、聚类算法。

[0160] 本实施例中优选的,所述预警评估模块 404 包括数据检索单元、预警评估报告生成单元和预警报告发送单元。

[0161] 其中,具体的实施细节上述方法中所述,在此不再赘述。

[0162] 本发明提出了一种新型的发票使用异常预警方法和系统,依托税收信息管理系统的海量数据,动态挖掘发票使用异常模式,对发票使用异常数据进行识别和预警,有效的解决了当出现各种发票使用异常情况时,无法使用单一的模式识别发票使用异常数据,使得预警识别更具针对性,也更具灵活性。

[0163] 本发明根据事先挖掘的先验的异常模式对异常数据进行匹配和识别,不仅提高发现异常数据效率,更有利于发票使用异常的分类预警和管理。

[0164] 本发明是基于知识发现的预警方法,知识库在运行中,随着发票使用异常的行为的变化,异常模式也随着不断的进化,使得系统具有进化性、智能化的特点。

[0165] 本领域普通技术人员可以理解:附图只是一个实施例的示意图,附图中的模块或流程并不一定是实施本发明所必须的。

[0166] 在此处所提供的说明书中,说明了大量具体细节。然而,能够理解,本发明的实施例可以在没有这些具体细节的情况下实践。在一些实例中,并未详细示出公知的方法、结构和技术,以便不模糊对本说明书的理解。

[0167] 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。

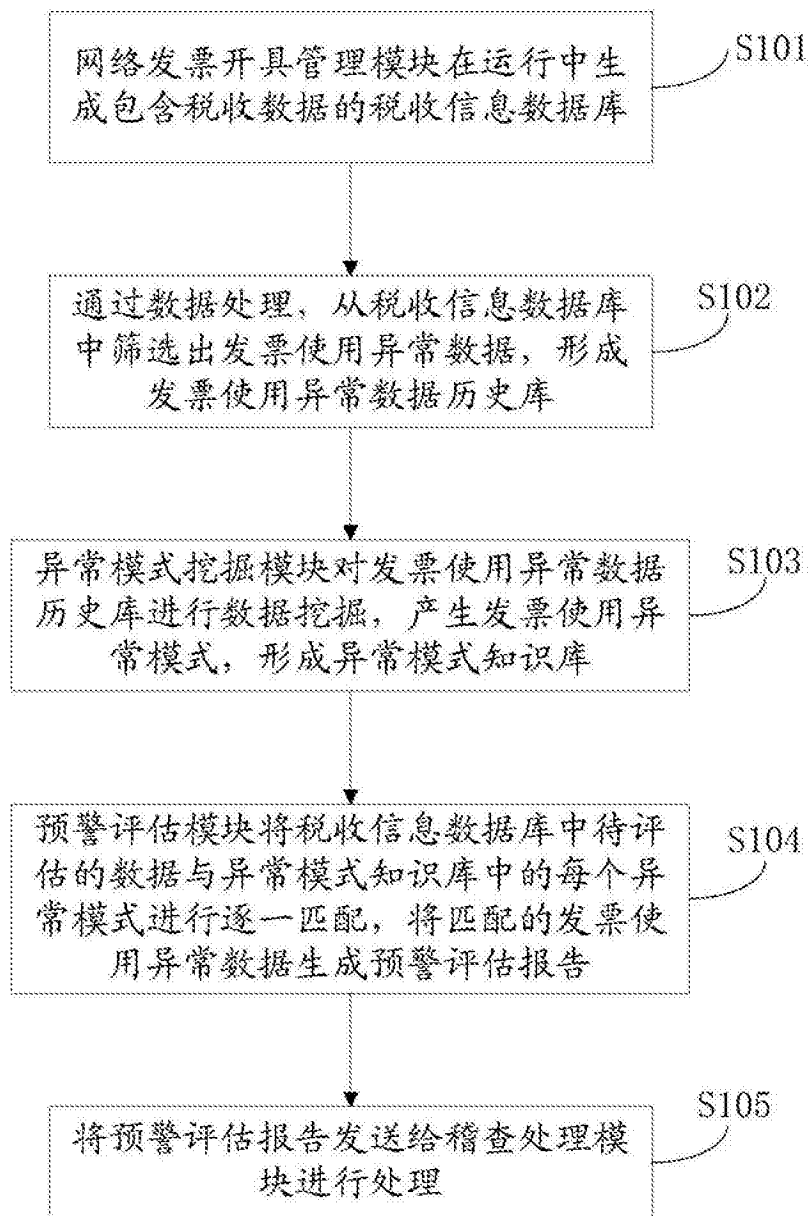


图 1

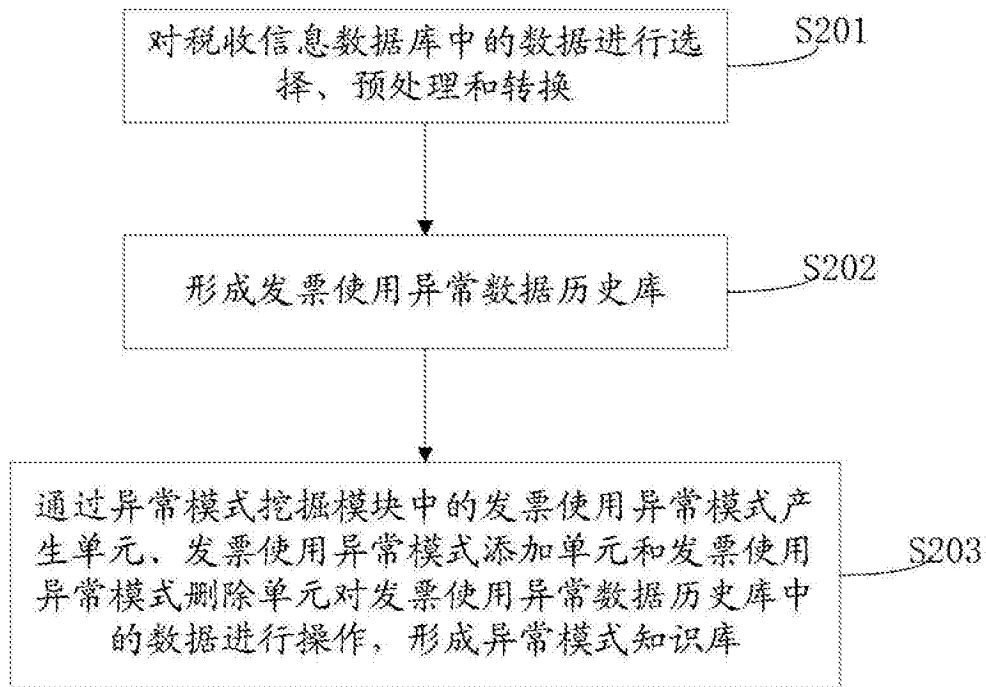


图 2

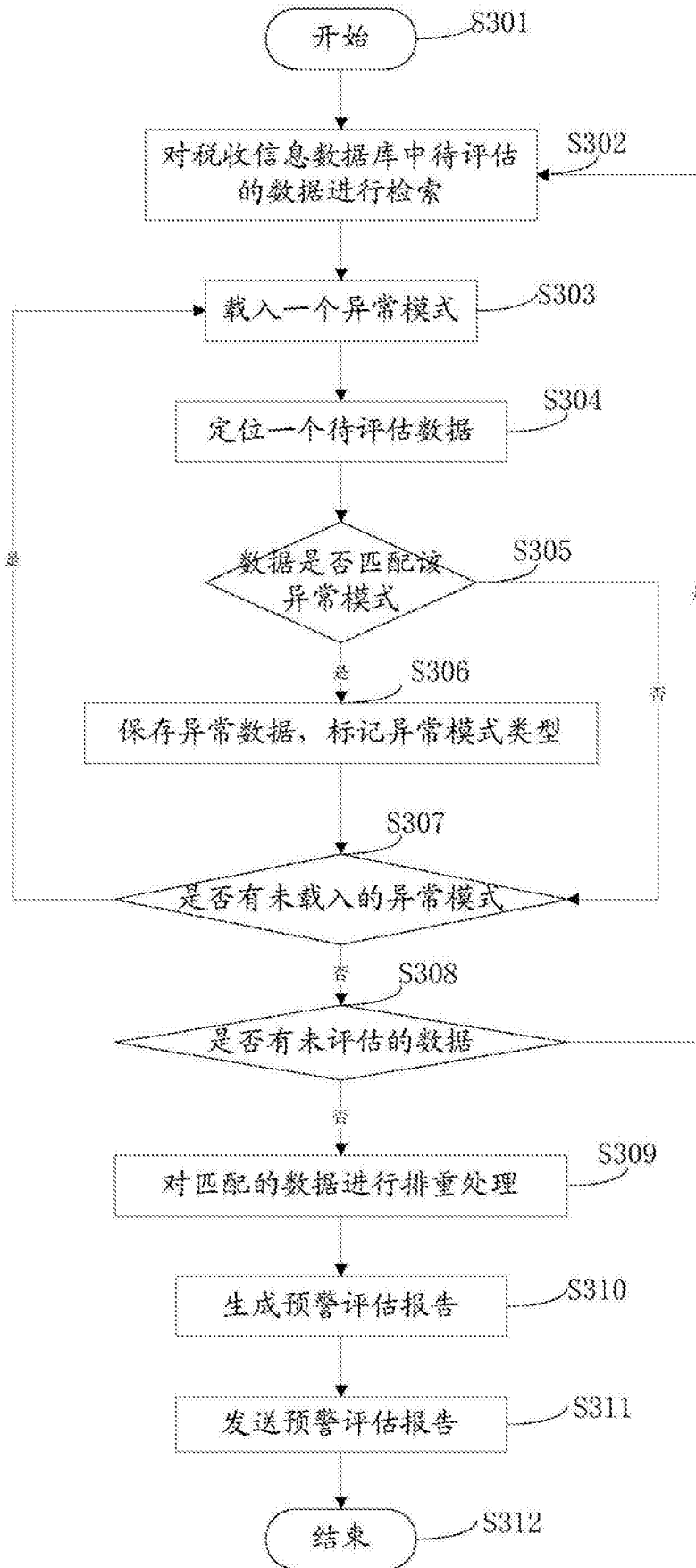


图 3

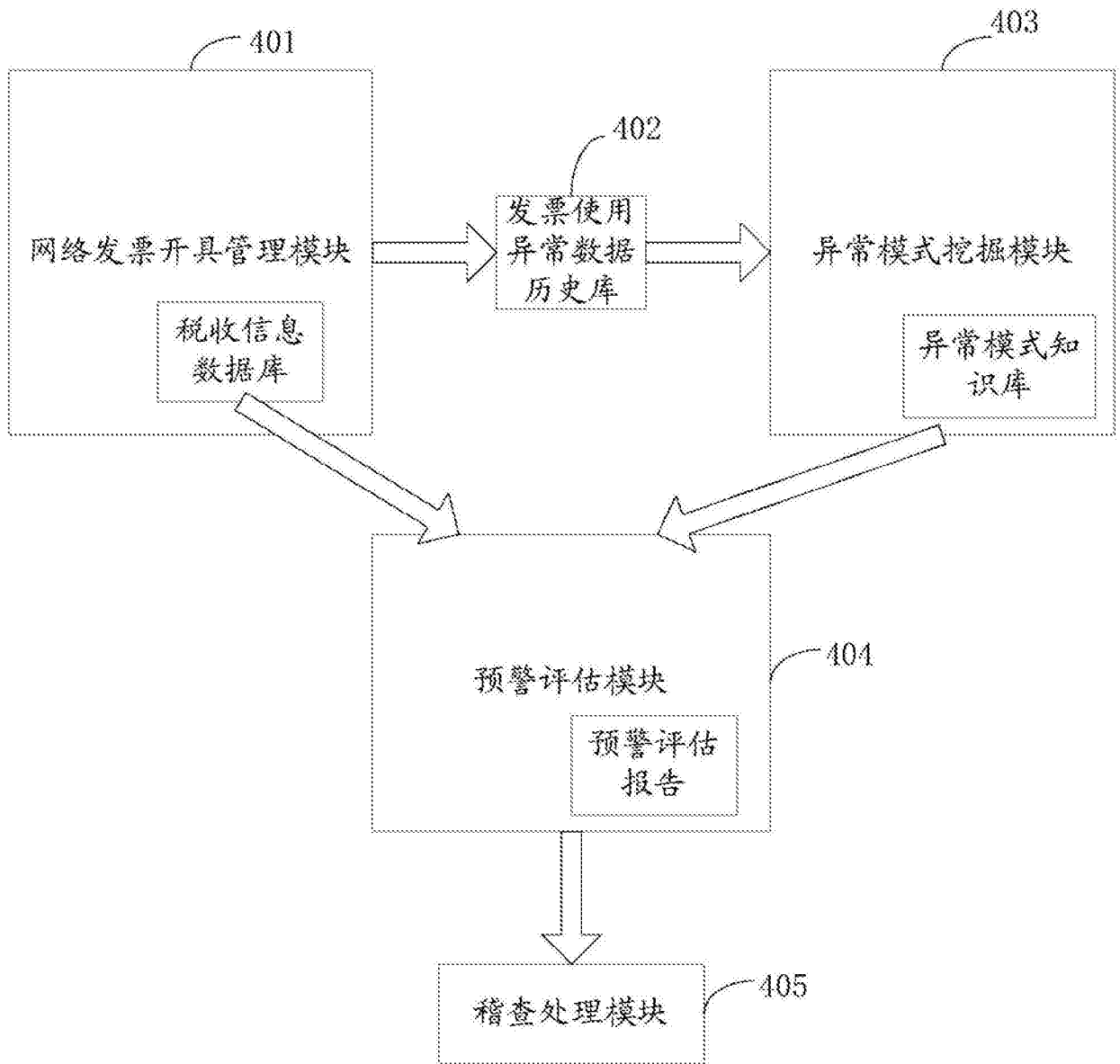


图 4