

(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
2020 年 8 月 6 日 (06.08.2020)



(10) 国际公布号
WO 2020/155752 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2019/117291

(22) 国际申请日: 2019 年 11 月 12 日 (12.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201910078953.2 2019 年 1 月 28 日 (28.01.2019) 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區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: 杨志鸿(YANG, Zhihong); 中国广东省深圳市福田區福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 徐亮(XU, Liang); 中国广东省深圳市福田區福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。 阮晓雯(RUAN, Xiaowen); 中国广东省深圳市福田區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: 深圳市精英专利事务所(SHENZHEN TALENT PATENT SERVICE); 中国广东省深圳市福田區深南中路 6009 号绿景广场 B 栋 20 层 B, Guangdong 518000 (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,

(54) Title: OUTLIER DETECTION MODEL VERIFICATION METHOD AND APPARATUS, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 异常点检测模型验证方法、装置、计算机设备及存储介质

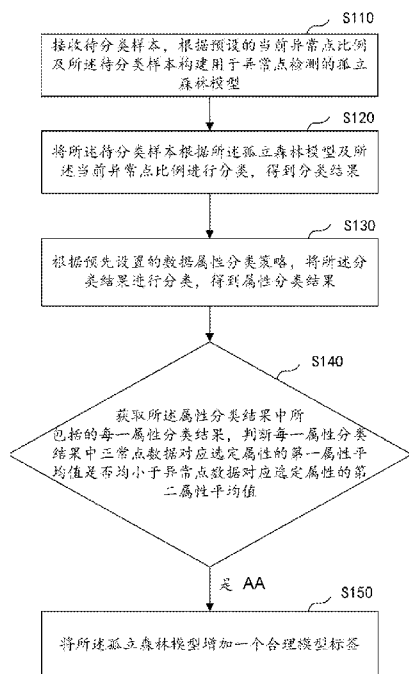


图 1

(57) Abstract: Disclosed are an outlier detection model verification method and apparatus, and a computer device and a storage medium. The method comprises: constructing an isolation forest model according to a preset current outlier proportion and a sample to be classified; classifying, according to the isolation forest model and the current outlier proportion, the sample to be classified to obtain a classification result; classifying the classification result according to a preset data attribute classification policy to obtain an attribute classification result; acquiring each attribute classification result in the attribute classification result; and if a first attribute average value

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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

in each attribute classification result is less than a second attribute average value, adding a rational model label to the isolation forest model. In the method, rapidly and accurately determining the classification rationality of an isolation forest model in an automatic manner is realized by means of an intelligent decision, thereby preventing low efficiency and high costs caused by manual determination.

(57) 摘要: 本申请公开了异常点检测模型验证方法、装置、计算机设备及存储介质。该方法包括: 根据预设的当前异常点比例及待分类样本构建孤立森林模型; 将待分类样本根据孤立森林模型及当前异常点比例进行分类得到分类结果; 根据预先设置的数据属性分类策略, 将分类结果进行分类, 得到属性分类结果; 获取属性分类结果中每一属性分类结果; 若每一属性分类结果中第一属性平均值均小于第二属性平均值, 将孤立森林模型增加一个合理模型标签。该方法采用智能决策实现了对孤立森林模型的分类合理性自动进行快速而准确的判断, 避免了因人工判断而导致的效率低下和高成本。

异常点检测模型验证方法、装置、计算机设备及存储介质

本申请要求于2019年1月28日提交中国专利局、申请号为201910078953.2、申请名称为“异常点检测模型验证方法、装置、计算机设备及存储介质”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

本申请涉及智能决策技术领域，尤其涉及一种异常点检测模型验证方法、装置、计算机设备及存储介质。

背景技术

异常值分析是检验数据是否有录入错误以及含有不合常理的数据的过程，忽视异常值的存在是十分不利的，不加剔除地把异常值包括进数据的计算分析过程中，对结果会产生不良影响。

异常值分析常用的是无监督异常检测模型，而对于无监督异常检测模型的输出结果，往往需要从现实中判断该模型分类的合理性。现有最常用的方法是专家法，即请教相关的专家以其业务知识，判断模型分类是否合理，这就导致判断成本高，而且效率低下。

发明内容

本申请实施例提供了一种异常点检测模型验证方法、装置、计算机设备及存储介质，旨在解决现有技术中对于无监督异常检测模型的输出结果通过相关的专家以其业务知识，判断模型分类是否合理，这就导致判断成本高，而且效率低下的问题。

第一方面，本申请实施例提供了一种异常点检测模型验证方法，其包括：

接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性

分类结果；

获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

第二方面，本申请实施例提供了一种异常点检测模型验证装置，其包括：

初始构建单元，用于接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

数据点分类单元，用于将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

属性分类单元，用于根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果；

平均值比较单元，用于获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

第一标签设置单元，用于若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

第三方面，本申请实施例又提供了一种计算机设备，其包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现上述第一方面所述的异常点检测模型验证方法。

第四方面，本申请实施例还提供了一种计算机可读存储介质，其中所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行上述第一方面所述的异常点检测模型验证方法。

附图说明

为了更清楚地说明本申请实施例技术方案，下面将对实施例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。

图 1 为本申请实施例提供的异常点检测模型验证方法的流程示意图；

图 2 为本申请实施例提供的异常点检测模型验证方法的子流程示意图；

图 3 为本申请实施例提供的异常点检测模型验证方法的另一子流程示意图；

图 4 为本申请实施例提供的异常点检测模型验证方法的另一流程示意图；

图 5 为本申请实施例提供的异常点检测模型验证装置的示意性框图；

图 6 为本申请实施例提供的异常点检测模型验证装置的子单元示意性框图；

图 7 为本申请实施例提供的异常点检测模型验证装置的另一子单元示意性框图；

图 8 为本申请实施例提供的异常点检测模型验证装置的另一示意性框图；

图 9 为本申请实施例提供的计算机设备的示意性框图。

具体实施方式

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，所描述的实施例是本申请一部分实施例，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，都属于本申请保护的范围。

应当理解，当在本说明书和所附权利要求书中使用时，术语“包括”和“包含”指示所描述特征、整体、步骤、操作、元素和/或组件的存在，但并不排除一个或多个其它特征、整体、步骤、操作、元素、组件和/或其集合的存在或添加。

还应当理解，在此本申请说明书中所使用的术语仅仅是出于描述特定实施例的目的而并不意在限制本申请。如在本申请说明书和所附权利要求书中所使用的那样，除非上下文清楚地指明其它情况，否则单数形式的“一”、“一个”及“该”意在包括复数形式。

还应当进一步理解，在本申请说明书和所附权利要求书中使用的术语“和/或”是指相关联列出的项中的一个或多个的任何组合以及所有可能组合，并且包括这些组合。

请参阅图 1，图 1 为本申请实施例提供的异常点检测模型验证方法的流程示意图，该异常点检测模型验证方法应用于服务器中，该方法通过安装于服务器中的应用程序进行执行。

如图 1 所示，该方法包括步骤 S110~S150。

S110、接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型。

在本实施例中，例如，服务器接收了上传端所上传的待分类样本后，也同时获取所设置初始的当前异常点比例为 0.5（如将初始的当前异常点比例记为 m_0 ），表示所期望的孤立森林模型的分类结果中正常点样本和异常点样本比例为 1:1。由于假设正常点数量比异常点多，因此此时异常点类别中含有大量的错分正常点。当异常点比例减少的时候，异常点类别中的正常点会被剔除。

孤立森林算法，即 iForest 算法，用于挖掘异常数据，或者是离群点挖掘，能在海量数据中，检索出与其它数据的规律不太符合的数据。例如，通常用于网络安全中的攻击检测和流量异常等分析，金融机构则用于挖掘出欺诈行为。

在一实施例中，如图 2 所示，步骤 S110 包括：

S111、从所述待分类样本中随机获取数据属性，及由数据属性和当前异常点比例所确定的分裂值；

S112、根据所述数据属性及所述分裂值将所述待分类样本进行划分，得到多个孤立树，由多个孤立树组合得到用于异常点检测的孤立森林模型。

在本实施例中，例如从训练数据集 $D=\{d_1, d_2, \dots, d_n\}$ 中随机选择一个数据属性 A ，并由数据属性 A 和当前异常点比例确定一个分裂值 p_1 ；然后对训练数据集中每个数据对象 d_i ，按照数据属性 A 的分裂值 p_1 进行划分。若 $d_i(A)$ 小于 p_1 ，则放在左子树，反之则在右子树。此时再随机选择一个数据属性 B ，并由数据属性 B 和当前异常点比例确定一个分裂值 p_2 ；然后对左子树和右子树均根据按照数据属性 B 的分裂值 p_2 进行划分，得到与左子树对应的次级左子树和次级右子树，以及与右子树对应的次级左子树和次级右子树。以此迭代，直至满足一下条件之一：（1） D 中剩下一条数据或者多条相同的数据；（2）孤立树达到最大高度。由于每一个孤立树在形成的过程中，所随机得到数据属性及与数据属性对应的分裂值不同，这就导致了孤立森林中能包括多个孤立树。孤立树中若设置异常点比例得当，即可提升异常点的检测效果。

S120、将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点。

在本实施例中，先根据所述孤立森林模型及当前异常点比例将所述待分类

样本进行分类后，得到了包括正常数据点和异常数据点的分类结果。为了对分类结果的合理性进行验证，可以继续观测正常类与异常类在某些特征上的均值上是否存在同向的变化，如在各个类别中正常点数据的某些特征值均大于或小于异常点数据。

在一实施例中，步骤 S120 之后、步骤 S130 之前还包括：

获取所述分类结果中正常类别的正常点中心；

获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离，以作为当前状态平均欧式距离；

通过所述当前异常点比例减去预设的步长，以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离；

通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长，得到平均欧式距离变动幅度；

若所述平均欧式距离变动幅度超出预设的变动幅度阈值，将当前异常点比例加上步长作为最优异常点比例，通过所述最优异常点比例以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果。

在本实施例中，步骤 S120 中将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类是初次分类，此时的当前异常点比例不一定是最佳的异常点比例。为了获取最佳的异常点比例，可通过上述方式获取。

当根据步骤 S120 中的当前异常点比例将待分类样本由所述孤立森林模型进行分类后，可以确定分类结果中正常类别的数据点对应的正常点中心，这一正常点中心在后续过程中是恒定不变的。

为了判断异常类别的每一数据点与正常点的距离关系，需计算异常类别的每一数据点与所述正常点中心的欧式距离后求平均，得到所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离，以作为当前状态平均欧式距离，从该当前状态平均欧式距离可以看出异常类别的每一数据点是否均远离正常点中心。

通过将当前异常点比例减去所述步长以更新当前异常点比例，此时无需再次确定正常点中心，只需得到分类结果中的异常类别的数据点，再计算异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离。

初始状态下时当前状态平均欧式距离视为 d_0 ，则初次得到的下一状态平均欧式距离视为 d_1 ，则第二次得到的下一状态平均欧式距离视为 d_2 （此时对应的当前状态平均欧式距离为 d_1 ），……，第 N 次得到的下一状态平均欧式距离视为 d_N （此时对应的当前状态平均欧式距离为 d_{N-1} ）。若将预设的步长记为 l ，则是通过 $(d_N - d_{N-1}) / l$ 来计算平均欧式距离变动幅度，其中 N 为大于 0 的正整数。

当平均欧式距离变动幅度陡然变大，表示此刻最新的当前异常点比例不是最优异常点比例，可考虑将此刻最新的当前异常点比例之前一个状态的当前异常点比例作为最优异常点比例。

若平均欧式距离变动幅度超出预设的变动幅度阈值，表示有部分真实的异常点被划分为正常点，导致异常点到正常中心点的平均欧式距离突增，此时当前异常点比例的上一状态（即当前异常点比例加上步长）即可作为最优异常点比例。

若所述平均欧式距离变动幅度未超出所述变动幅度阈值，将当前异常点比例减去步长以更新当前异常点比例，通过下一状态平均欧式距离以更新当前状态平均欧式距离，返回执行将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离的步骤。

当平均欧式距离变动幅度仍保持平稳过渡，表示所降低的异常点比例不足以明显影响异常类别的每一数据点与所述正常点中心的平均欧式距离，此时需将当前异常点比例减去步长以更新当前异常点比例，并通过下一状态平均欧式距离以更新新当前状态平均欧式距离。例如当 $(d_N - d_{N-1}) / l$ 未超出预设的变动幅度阈值，此时将 d_1 作为当前状态平均欧式距离，将 $(m_0 - l)$ 作为当前异常点比例重新返回计算以得到 d_2 ；之后即是以 $(d_2 - d_1) / l$ 作为平均欧式距离变动幅度，以此类推，直至执行到平均欧式距离变动幅度超出预设的变动幅度阈值即可。

在一实施例中，所述将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果之后，还包括：

将所述分类结果及所述最优异常点比例发送至所述待分类样本对应的上传端，并将所述分类结果及所述最优异常点比例同步发送至云服务器；

将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除。

在本实施例中，若在服务器中完成了获取了与所述待分类样本对应的分类结果及所述最优异常点比例后，可以及时的将该分类结果及所述最优异常点比例发送至所述待分类样本对应的上传端，实现对上传端进行分类结果的有效通知。

而且为了降低服务器中的数据存储压力，此时可及时的将所述分类结果及所述最优异常点比例同步发送至云服务器，通过云服务器实现对与所述待分类样本对应的分类结果及所述最优异常点比例的有效存储。此过程中，还可以将与所述分类结果及所述最优异常点比例对应的所述待分类样本同步至云服务器。上述的待分类样本、分类结果及最优异常点比例在由服务器同步至云服务器中时，需以上传端的唯一机器识别码（如IMEI 串号）为数据标识位来进行唯一数据标识。

此时将所述分类结果及所述最优异常点比例同步发送至云服务器之后，则可对服务器中将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除，从而有效释放出存储空间。

在一实施例中，所述将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除之前，还包括：

根据预设的当前异常点比例与所述最优异常点比例之差除以所述步长，得到迭代次数；

将所述迭代次数发送至所述待分类样本对应的上传端，并将所述迭代次数同步发送至云服务器。

在本实施例中，为了清楚的获知预设的当前异常点比例与最优异常点比例之间经过了多少次迭代，此时可以根据预设的当前异常点比例与最优异常点比例之差除以所述步长，得到迭代次数。当获知了所述迭代次数后，可以将所述迭代次数发送至所述待分类样本对应的上传端，上传端对应则可积累设置最优异常点比例的经验。

S130、根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果。

在一实施例中，如图3所示，步骤S130包括：

S131、根据所述数据属性分类策略判断每一正常数据点所属的分组，以得到正常类别属性分类结果；

S132、根据所述数据属性分类策略判断每一异常数据点所属的分组，以得到异常类别属性分类结果；

S133、由所述正常类别属性分类结果及所述异常类别属性分类结果组成所述属性分类结果。

在本实施例中，将所述待分类样本根据所述孤立森林模型进行分类（例如根据最优异常点比例进行分类，得到分类结果）后，所述分类结果包括正常点数据和异常点数据，也即待分类样本中的每一数据点在分类完成后都对应有属于正常类别或是异常类别的标签。此时为了进一步分析分类结果中的数据特点，可以选取分类结果中的每一正常数据点和每一异常数据点中的某一属性，将每一正常数据点再次进行分类得到正常类别属性分类结果，同时将每一异常数据点再次进行分类得到异常类别属性分类结果。

正常数据点组成的正常类别数据集合、和异常数据点组成的异常类别数据集合中包括的每一数据点，除了拥有一个主属性（如身份证号），还有与主属性对应的多个从属属性（如年龄，投保保单数目等属性）。此时可以选择以投保保单数目作为分类属性，将所述待分类样本进行分类，得到属性分类结果。

例如，可以在正常类别数据集合中选择将投保保单数目不超过1的数据点分第一类属性分类结果，在正常类别数据集合中选择将投保保单数目超过1的数据点分为第二类属性分类结果。在异常类别数据集合中选择将投保保单数目不超过1的数据点分第三类属性分类结果，在异常类别数据集合中选择将投保保单数目超过1的数据点分为第四类属性分类结果。由第一类属性分类结果和第三类属性分类结果组成第一属性分类结果，由第二类属性分类结果和第四类属性分类结果组成第二属性分类结果。

S140、获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值。

在本实施例中，例如判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值如下：

判断第一属性分类结果中第一类属性分类结果中各正常数据点在投保保单数目这一属性下的第一属性平均值是否小于第一类属性分类结果中各异常数据点在投保保单数目这一属性下的第二属性平均值。同时也判断第二属性分类结果中第三类属性分类结果中各正常数据点在投保保单数目这一属性下的第一属性平均值是否小于第四类属性分类结果中各异常数据点在投保保单数目这一属性下的第二属性平均值。通过观察每个类别中，正常类别数据集合与异常类别数据集合在某些特征上的均值上是否存在同向的变化，可以辅助判断所述孤立森林模型分类的合理性。

S150、若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

在本实施例中，例如第一属性分类结果中第一类属性分类结果中各正常数据点在投保保单数目这一属性下的第一属性平均值小于第一类属性分类结果中各异常数据点在投保保单数目这一属性下的第二属性平均值。而且第二属性分类结果中第三类属性分类结果中各正常数据点在投保保单数目这一属性下的第一属性平均值小于第四类属性分类结果中各异常数据点在投保保单数目这一属性下的第二属性平均值。这就表示正常类别数据集合与异常类别数据集合在投保保单数目上的均值上存在同向的变化，说明所述孤立森林模型是合理的，具有较好的分类效果，可增加一个合理模型标签进行标识。

在一实施例中，如图4所示，步骤S150之后还包括：

S160、若每一属性分类结果中所述第一属性平均值有大于或等于所述第二属性平均值，将所述孤立森林模型增加一个非合理模型标签。

在本实施例中，若每一属性分类结果中所述第一属性平均值有大于或等于所述第二属性平均值，表示正常类别数据集合与异常类别数据集合在对应选定属性的均值上不存在同向的变化，说明所述孤立森林模型不是合理的，需要重新训练模型，可增加一个非合理模型标签进行标识。

该方法实现了对孤立森林模型的分类合理性自动进行快速而准确的判断，避免了因人工判断而导致的效率低下和高成本。

本申请实施例还提供一种异常点检测模型验证装置，该异常点检测模型验证装置用于执行前述异常点检测模型验证方法的任一实施例。具体地，请参阅图5，图5是本申请实施例提供的异常点检测模型验证装置的示意性框图。该异

常点检测模型验证装置 100 可以配置于服务器中。

如图 5 所示, 异常点检测模型验证装置 100 包括初始构建单元 110、数据点分类单元 120、属性分类单元 130、平均值比较单元 140、第一标签设置单元 150。

初始构建单元 110, 用于接收待分类样本, 根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型。

在一实施例中, 如图 6 所示, 初始构建单元 110 包括:

分类参数获取单元 111, 用于从所述待分类样本中随机获取数据属性, 及由数据属性和当前异常点比例所确定的分裂值;

模型获取单元 112, 用于根据所述数据属性及所述分裂值将所述待分类样本进行划分, 得到多个孤立树, 由多个孤立树组合得到用于异常点检测的孤立森林模型。

数据点分类单元 120, 用于将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类, 得到分类结果; 其中, 所述分类结果包括正常数据点和异常数据点。

在一实施例中, 异常点检测模型验证装置 100 还包括:

正常点中心获取单元, 用于获取所述分类结果中正常类别的正常点中心;

第一平均值获取单元, 用于获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离, 以作为当前状态平均欧式距离;

第一比例更新单元, 用于通过所述当前异常点比例减去预设的步长, 以更新当前异常点比例;

第二平均值获取单元, 用于将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类, 得到当前异常类别的数据点, 获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离;

幅度计算单元, 用于通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长, 得到平均欧式距离变动幅度;

最优比例获取单元, 用于若所述平均欧式距离变动幅度超出预设的变动幅度阈值, 将当前异常点比例加上步长作为最优异常点比例, 通过所述最优异常点比例以更新当前异常点比例;

最优分类单元, 用于将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类, 得到分类结果。

属性分类单元 130, 用于根据预先设置的数据属性分类策略, 将所述分类结果进行分类, 得到属性分类结果。

在一实施例中, 如图 7 所示, 属性分类单元 130 包括:

第一类别分类单元 131, 用于根据所述数据属性分类策略判断每一正常数据点所属的分组, 以得到正常类别属性分类结果;

第二类别分类单元 132, 用于根据所述数据属性分类策略判断每一异常数据点所属的分组, 以得到异常类别属性分类结果;

属性分类结果获取单元 133, 用于由所述正常类别属性分类结果及所述异常类别属性分类结果组成所述属性分类结果。

平均值比较单元 140, 用于获取所述属性分类结果中所包括的每一属性分类结果, 判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值。

第一标签设置单元 150, 用于若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值, 将所述孤立森林模型增加一个合理模型标签。

在一实施例中, 如图 8 所示, 异常点检测模型验证装置 100 还包括:

第二标签设置单元 160, 用于若每一属性分类结果中所述第一属性平均值有大于或等于所述第二属性平均值, 将所述孤立森林模型增加一个非合理模型标签。

该装置实现了对孤立森林模型的分类合理性自动进行快速而准确的判断, 避免了因人工判断而导致的效率低下和高成本。

上述异常点检测模型验证装置可以实现为计算机程序的形式, 该计算机程序可以在如图 9 所示的计算机设备上运行。

请参阅图 9, 图 9 是本申请实施例提供的计算机设备的示意性框图。该计算机设备 500 是服务器, 服务器可以是独立的服务器, 也可以是多个服务器组成的服务器集群。

参阅图 9, 该计算机设备 500 包括通过系统总线 501 连接的处理器 502、存储器和网络接口 505, 其中, 存储器可以包括非易失性存储介质 503 和内存储器 504。

该非易失性存储介质 503 可存储操作系统 5031 和计算机程序 5032。该计算机程序 5032 被执行时, 可使得处理器 502 执行异常点检测模型验证方法。

该处理器 502 用于提供计算和控制能力，支撑整个计算机设备 500 的运行。

该内存储器 504 为非易失性存储介质 503 中的计算机程序 5032 的运行提供环境，该计算机程序 5032 被处理器 502 执行时，可使得处理器 502 执行异常点检测模型验证方法。

该网络接口 505 用于进行网络通信，如提供数据信息的传输等。本领域技术人员可以理解，图 9 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备 500 的限定，具体的计算机设备 500 可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

其中，所述处理器 502 用于运行存储在存储器中的计算机程序 5032，以实现本申请实施例公开的异常点检测模型验证方法。

本领域技术人员可以理解，图 9 中示出的计算机设备的实施例并不构成对计算机设备具体构成的限定，在其他实施例中，计算机设备可以包括比图示更多或更少的部件，或者组合某些部件，或者不同的部件布置。例如，在一些实施例中，计算机设备可以仅包括存储器及处理器，在这样的实施例中，存储器及处理器的结构及功能与图 9 所示实施例一致，在此不再赘述。

应当理解，在本申请实施例中，处理器 502 可以是中央处理单元 (Central Processing Unit, CPU)，该处理器 502 还可以是其他通用处理器、数字信号处理器 (Digital Signal Processor, DSP)、专用集成电路 (Application Specific Integrated Circuit, ASIC)、现成可编程门阵列 (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、分立硬件组件等。其中，通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

在本申请的另一实施例中提供计算机可读存储介质。该计算机可读存储介质可以为非易失性的计算机可读存储介质。该计算机可读存储介质存储有计算机程序，其中计算机程序被处理器执行时实现本申请实施例公开的异常点检测模型验证方法。

所述存储介质为实体的、非瞬时性的存储介质，例如可以是 U 盘、移动硬盘、只读存储器 (Read-Only Memory, ROM)、磁碟或者光盘等各种可以存储程序代码的实体存储介质。

所属领域的技术人员可以清楚地了解到，为了描述的方便和简洁，上述描

述的设备、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。

以上所述，仅为本申请的具体实施方式，但本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到各种等效的修改或替换，这些修改或替换都应涵盖在本申请的保护范围之内。因此，本申请的保护范围应以权利要求的保护范围为准。

权利要求书

1. 一种异常点检测模型验证方法，包括：

接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果；

获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

2. 根据权利要求1所述的异常点检测模型验证方法，其中，所述根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果之前，还包括：

获取所述分类结果中正常类别的正常点中心；

获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离，以作为当前状态平均欧式距离；

通过所述当前异常点比例减去预设的步长，以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离；

通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长，得到平均欧式距离变动幅度；

若所述平均欧式距离变动幅度超出预设的变动幅度阈值，将当前异常点比例加上步长作为最优异常点比例，通过所述最优异常点比例以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，

得到分类结果。

3. 根据权利要求1所述的异常点检测模型验证方法，其中，所述根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型，包括：

从所述待分类样本中随机获取数据属性，及由数据属性和当前异常点比例所确定的分裂值；

根据所述数据属性及所述分裂值将所述待分类样本进行划分，得到多个孤立树，由多个孤立树组合得到用于异常点检测的孤立森林模型。

4. 根据权利要求1所述的异常点检测模型验证方法，其中，所述根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果，包括：

根据所述数据属性分类策略判断每一正常数据点所属的分组，以得到正常类别属性分类结果；

根据所述数据属性分类策略判断每一异常数据点所属的分组，以得到异常类别属性分类结果；

由所述正常类别属性分类结果及所述异常类别属性分类结果组成所述属性分类结果。

5. 根据权利要求2所述的异常点检测模型验证方法，其中，所述获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值之后，还包括：

若每一属性分类结果中所述第一属性平均值有大于或等于所述第二属性平均值，将所述孤立森林模型增加一个非合理模型标签。

6. 根据权利要求2所述的异常点检测模型验证方法，其中，所述将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果之后，还包括：

将所述分类结果及所述最优异常点比例发送至所述待分类样本对应的上传端，并将所述分类结果及所述最优异常点比例同步发送至云服务器；

将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除。

7. 根据权利要求6所述的异常点检测模型验证方法，其中，所述将所述分

类结果及所述最优异常点比例对应的存储区域进行格式化删除之前，还包括：

根据预设的当前异常点比例与所述最优异常点比例之差除以所述步长，得到迭代次数；

将所述迭代次数发送至所述待分类样本对应的上传端，并将所述迭代次数同步发送至云服务器。

8.一种异常点检测模型验证装置，包括：

初始构建单元，用于接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

数据点分类单元，用于将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

属性分类单元，用于根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果；

平均值比较单元，用于获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

第一标签设置单元，用于若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

9.根据权利要求8所述的异常点检测模型验证装置，其中，还包括：

正常点中心获取单元，用于获取所述分类结果中正常类别的正常点中心；

第一平均值获取单元，用于获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离，以作为当前状态平均欧式距离；

第一比例更新单元，用于通过所述当前异常点比例减去预设的步长，以更新当前异常点比例；

第二平均值获取单元，用于将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离；

幅度计算单元，用于通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长，得到平均欧式距离变动幅度；

最优比例获取单元，用于若所述平均欧式距离变动幅度超出预设的变动幅

度阈值，将当前异常点比例加上步长作为最优异常点比例，通过所述最优异常点比例以更新当前异常点比例；

最优分类单元，用于将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果。

10.根据权利要求8所述的异常点检测模型验证装置，其中，所述属性分类单元，包括：

第一类别分类单元，用于根据所述数据属性分类策略判断每一正常数据点所属的分组，以得到正常类别属性分类结果；

第二类别分类单元，用于根据所述数据属性分类策略判断每一异常数据点所属的分组，以得到异常类别属性分类结果；

属性分类结果获取单元，用于由所述正常类别属性分类结果及所述异常类别属性分类结果组成所述属性分类结果。

11.一种计算机设备，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时实现以下步骤：

接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果；

获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

12.根据权利要求11所述的计算机设备，其中，所述根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果之前，还包括：

获取所述分类结果中正常类别的正常点中心；

获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式

距离，以作为当前状态平均欧式距离；

通过所述当前异常点比例减去预设的步长，以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离；

通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长，得到平均欧式距离变动幅度；

若所述平均欧式距离变动幅度超出预设的变动幅度阈值，将当前异常点比例加上步长作为最优异常点比例，通过所述最优异常点比例以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果。

13. 根据权利要求 11 所述的计算机设备，其中，所述根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型，包括：

从所述待分类样本中随机获取数据属性，及由数据属性和当前异常点比例所确定的分裂值；

根据所述数据属性及所述分裂值将所述待分类样本进行划分，得到多个孤立树，由多个孤立树组合得到用于异常点检测的孤立森林模型。

14. 根据权利要求 11 所述的计算机设备，其中，所述根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果，包括：

根据所述数据属性分类策略判断每一正常数据点所属的分组，以得到正常类别属性分类结果；

根据所述数据属性分类策略判断每一异常数据点所属的分组，以得到异常类别属性分类结果；

由所述正常类别属性分类结果及所述异常类别属性分类结果组成所述属性分类结果。

15. 根据权利要求 12 所述的计算机设备，其中，所述获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值之后，还包括：

若每一属性分类结果中所述第一属性平均值有大于或等于所述第二属性平均值，将所述孤立森林模型增加一个非合理模型标签。

16. 根据权利要求 12 所述的计算机设备，其中，所述将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果之后，还包括：

将所述分类结果及所述最优异常点比例发送至所述待分类样本对应的上传端，并将所述分类结果及所述最优异常点比例同步发送至云服务器；

将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除。

17. 根据权利要求 16 所述的计算机设备，其中，所述将所述分类结果及所述最优异常点比例对应的存储区域进行格式化删除之前，还包括：

根据预设的当前异常点比例与所述最优异常点比例之差除以所述步长，得到迭代次数；

将所述迭代次数发送至所述待分类样本对应的上传端，并将所述迭代次数同步发送至云服务器。

18. 一种计算机可读存储介质，所述计算机可读存储介质存储有计算机程序，所述计算机程序当被处理器执行时使所述处理器执行以下操作：

接收待分类样本，根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型；

将所述待分类样本根据所述孤立森林模型及所述当前异常点比例进行分类，得到分类结果；其中，所述分类结果包括正常数据点和异常数据点；

根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果；

获取所述属性分类结果中所包括的每一属性分类结果，判断每一属性分类结果中正常点数据对应选定属性的第一属性平均值是否均小于异常点数据对应选定属性的第二属性平均值；以及

若每一属性分类结果中所述第一属性平均值均小于所述第二属性平均值，将所述孤立森林模型增加一个合理模型标签。

19. 根据权利要求 18 所述的计算机可读存储介质，其中，所述根据预先设置的数据属性分类策略，将所述分类结果进行分类，得到属性分类结果之前，还包括：

获取所述分类结果中正常类别的正常点中心；

获取所述分类结果中异常类别的每一数据点与所述正常点中心的平均欧式距离，以作为当前状态平均欧式距离；

通过所述当前异常点比例减去预设的步长，以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及当前异常点比例进行分类，得到当前异常类别的数据点，获取当前异常类别的每一数据点与所述正常点中心的平均欧式距离以作为下一状态平均欧式距离；

通过下一状态平均欧式距离与当前状态平均欧式距离之差除以所述步长，得到平均欧式距离变动幅度；

若所述平均欧式距离变动幅度超出预设的变动幅度阈值，将当前异常点比例加上步长作为最优异常点比例，通过所述最优异常点比例以更新当前异常点比例；

将所述待分类样本根据所述孤立森林模型及所述最优异常点比例进行分类，得到分类结果。

20. 根据权利要求 18 所述的计算机可读存储介质，其中，所述根据预设的当前异常点比例及所述待分类样本构建用于异常点检测的孤立森林模型，包括：

从所述待分类样本中随机获取数据属性，及由数据属性和当前异常点比例所确定的分裂值；

根据所述数据属性及所述分裂值将所述待分类样本进行划分，得到多个孤立树，由多个孤立树组合得到用于异常点检测的孤立森林模型。

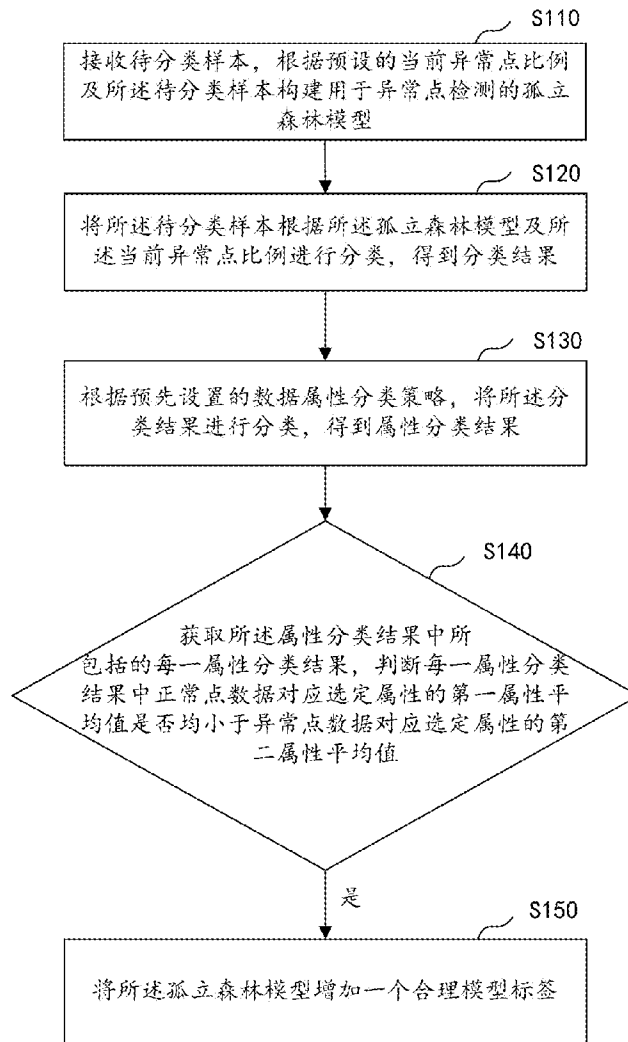


图 1

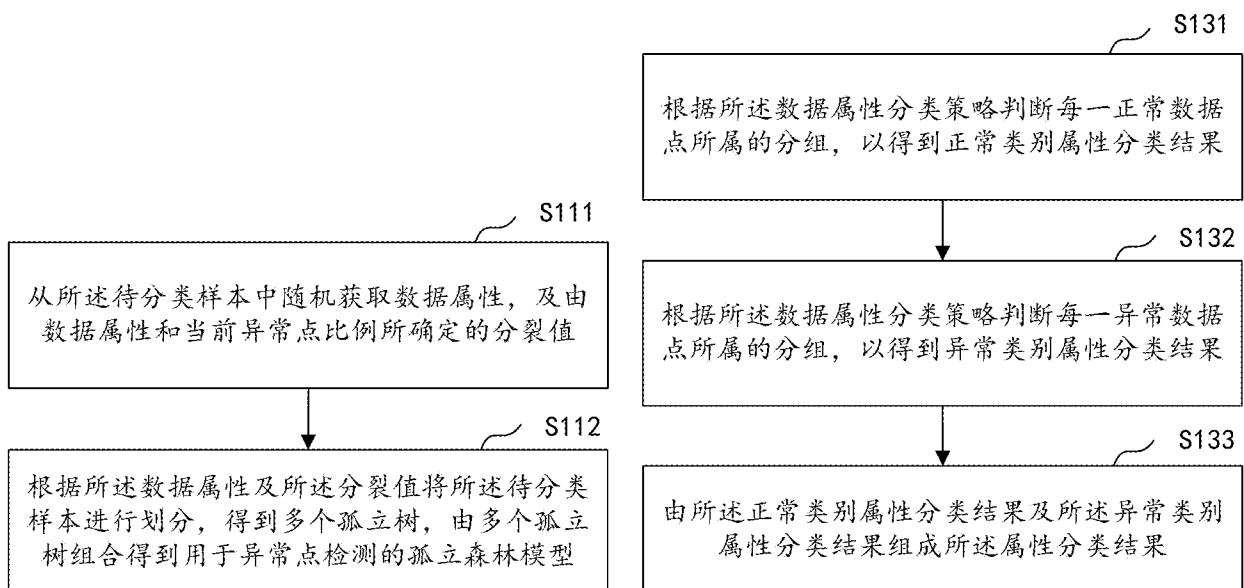


图 2

图 3

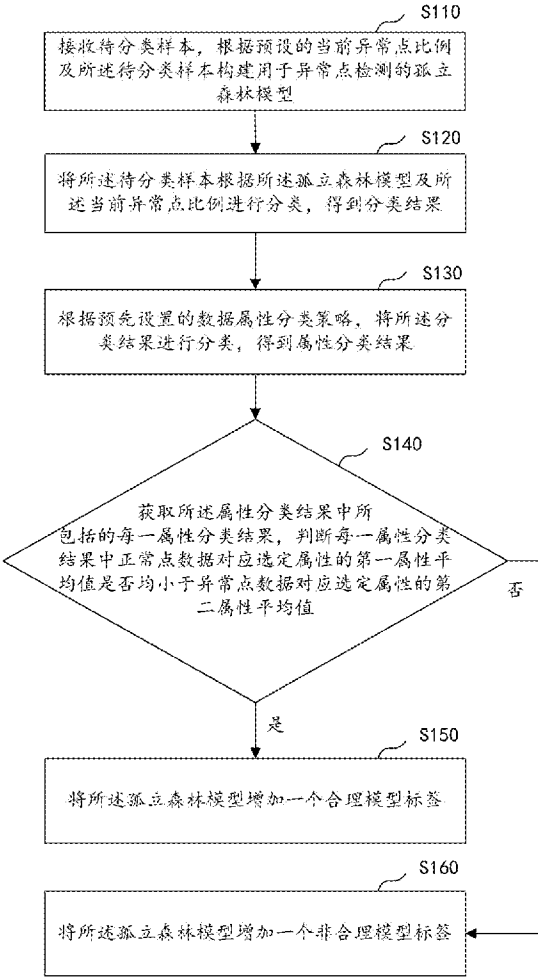


图 4

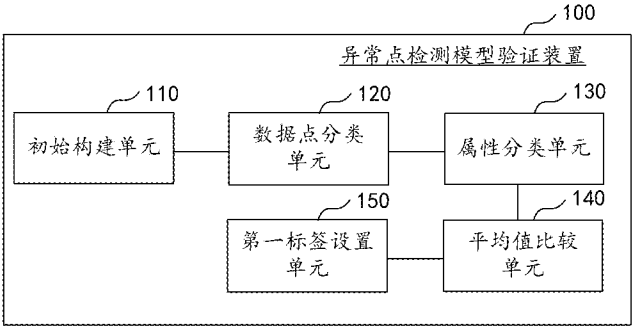


图 5

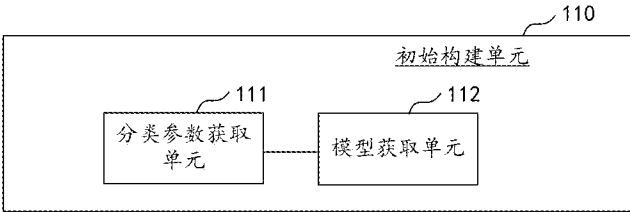


图 6

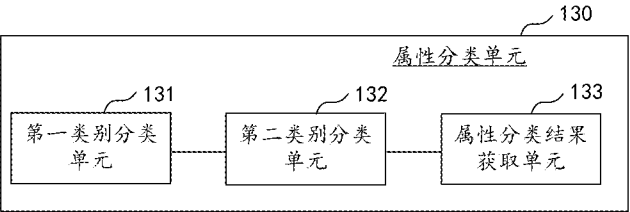


图 7

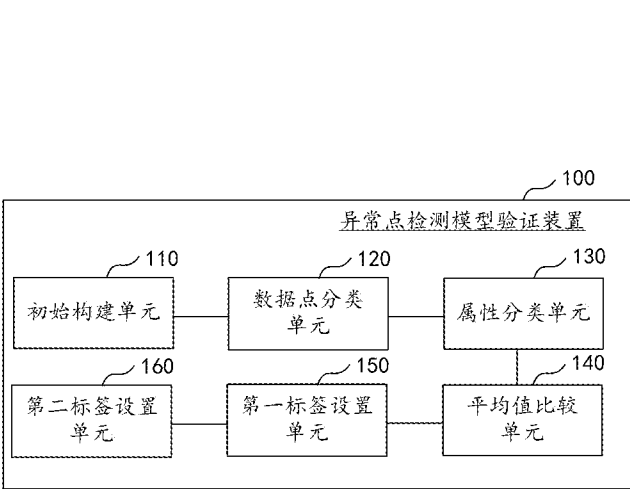


图 8

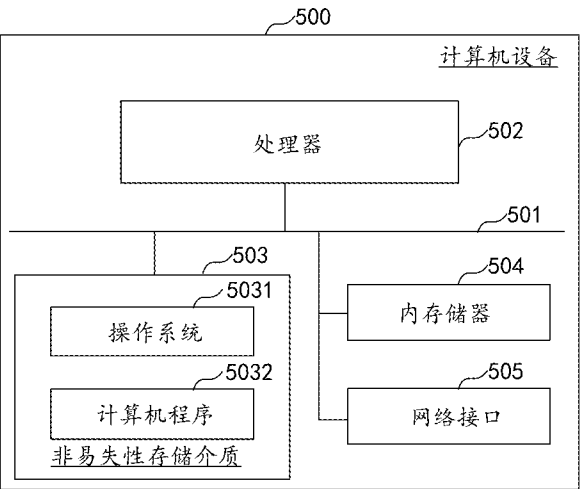


图 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/117291

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)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)

G06K9/-

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)

USTXT; EPTXT; CNTXT; CNABS; WOTXT; VEN; CNKI: 孤立森林, 模型, 检测, 样本, 异常点, 比例, 分类, 属性, 平均值, 合理, 标签, iforest, isolation forest, model, detection, sample, abnormal point, ratio, classification, characteristic, average value, label, reasonable

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109902721 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 18 June 2019 (2019-06-18) description, paragraphs 35-79	1-5, 8-15, 18-20
A	CN 108777873 A (JIANGNAN UNIVERSITY) 09 November 2018 (2018-11-09) description, paragraphs 44-94	1-20
A	CN 107357790 A (ALIBABA GROUP HOLDING LIMITED) 17 November 2017 (2017-11-17) description, paragraphs 73-77 and 93	1-20
A	US 10045218 B1 (ARGYLE DATA INC.) 07 August 2018 (2018-08-07) entire document	1-20
A	CN 108921440 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 30 November 2018 (2018-11-30) 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

18 December 2019

Date of mailing of the international search report

18 February 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

International application No.

PCT/CN2019/117291

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LIU, F. T. et al. "Isolation Forest" <i>IEEE International Conference on Data Mining 2008 (ICDM 08)</i> , 19 December 2008 (2008-12-19), pp. 413-422	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/117291

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109902721	A	18 June 2019	None	
CN	108777873	A	09 November 2018	None	
CN	107357790	A	17 November 2017	None	
US	10045218	B1	07 August 2018	None	
CN	108921440	A	30 November 2018	None	

国际检索报告

国际申请号

PCT/CN2019/117291

A. 主题的分类 G06K 9/62 (2006.01) 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) G06K9/- 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)) USTXT; EPTXT; CNTXT; CNABS; WOTXT; VEN; CNKI: 孤立森林, 模型, 检测, 样本, 异常点, 比例, 分类, 属性, 平均值, 合理, 标签, iforest, isolation forest, model, detection, sample, abnormal point, ratio, classification, characteristic, average value, label, reasonable																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109902721 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 说明书第35-79段</td> <td>1-5、8-15、18-20</td> </tr> <tr> <td>A</td> <td>CN 108777873 A (江南大学) 2018年 11月 9日 (2018 - 11 - 09) 说明书第44-94段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107357790 A (阿里巴巴集团控股有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第73-77、93段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>US 10045218 B1 (ARGYLE DATA INC) 2018年 8月 7日 (2018 - 08 - 07) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108921440 A (平安科技深圳有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>LIU, Fei Tony等. "Isolation Forest" IEEE International Conference on Data Mining 2008 (ICDM 08), 2008年 12月 19日 (2008 - 12 - 19), 第413-422页</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109902721 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 说明书第35-79段	1-5、8-15、18-20	A	CN 108777873 A (江南大学) 2018年 11月 9日 (2018 - 11 - 09) 说明书第44-94段	1-20	A	CN 107357790 A (阿里巴巴集团控股有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第73-77、93段	1-20	A	US 10045218 B1 (ARGYLE DATA INC) 2018年 8月 7日 (2018 - 08 - 07) 全文	1-20	A	CN 108921440 A (平安科技深圳有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文	1-20	A	LIU, Fei Tony等. "Isolation Forest" IEEE International Conference on Data Mining 2008 (ICDM 08), 2008年 12月 19日 (2008 - 12 - 19), 第413-422页	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 109902721 A (平安科技深圳有限公司) 2019年 6月 18日 (2019 - 06 - 18) 说明书第35-79段	1-5、8-15、18-20																					
A	CN 108777873 A (江南大学) 2018年 11月 9日 (2018 - 11 - 09) 说明书第44-94段	1-20																					
A	CN 107357790 A (阿里巴巴集团控股有限公司) 2017年 11月 17日 (2017 - 11 - 17) 说明书第73-77、93段	1-20																					
A	US 10045218 B1 (ARGYLE DATA INC) 2018年 8月 7日 (2018 - 08 - 07) 全文	1-20																					
A	CN 108921440 A (平安科技深圳有限公司) 2018年 11月 30日 (2018 - 11 - 30) 全文	1-20																					
A	LIU, Fei Tony等. "Isolation Forest" IEEE International Conference on Data Mining 2008 (ICDM 08), 2008年 12月 19日 (2008 - 12 - 19), 第413-422页	1-20																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 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 </td> <td> "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 </td> </tr> </table>			* 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																			
* 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 2019年 12月 18日		国际检索报告邮寄日期 2020年 2月 18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 夏雪 电话号码 86-(20)-28950718																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/117291

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	109902721	A	2019年 6月 18日	无	
CN	108777873	A	2018年 11月 9日	无	
CN	107357790	A	2017年 11月 17日	无	
US	10045218	B1	2018年 8月 7日	无	
CN	108921440	A	2018年 11月 30日	无	