(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116049888 A
(43)申请公布日 2023.05.02

(21)申请号 202310086159.9

(22)申请日 2023.01.19

(71)申请人 清华大学
地址 100084 北京市海淀区清华园
申请人 中国民用航空局空中交通管理局

(72)发明人 张毅 杨敬轩 晏松 何泓霖
陈宝刚 杨锐 马超

(74)专利代理机构 北京安信方达知识产权代理有限公司 11262
专利代理师 张建秀 栗若木

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 21/64(2013.01)
G06F 18/2431(2023.01)
G06N 5/01(2023.01)

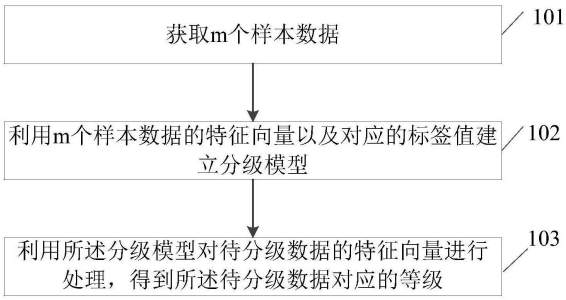
权利要求书2页 说明书7页 附图3页

(54)发明名称

一种数据安全性分级方法、装置、存储介质和电子装置

(57)摘要

本申请实施例公开了一种数据安全性分级方法、装置、存储介质和电子装置。所述方法包括:获取m个样本数据,其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i,其中,i=1,2,3,...,m,n和m均为大于或等于2的整数,y_i=1,2,3,...,R,其中R为等级的最大值;利用m个样本数据的特征向量以及对应的标签值建立分级模型;利用所述分级模型对待分级数据的特征向量进行处理,得到所述待分级数据对应的等级。



CN 116049888 A

1.一种数据安全性分级方法，包括：

获取m个样本数据，其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i，其中，i＝1,2,3,...,m，n和m均为大于或等于2的整数，y_i＝1,2,3,...,R，其中R为等级的最大值；

利用m个样本数据的特征向量以及对应的标签值建立分级模型；

利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级；

其中，所述分级模型的表达式如下：

$$f(x)=\arg\max_{\{y=1,2,..,R\}}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right);$$

在上述表达式中，f(x)为对待分级数据x进行分级操作确定的等级，$G_k(x)$为第k个决策树模型基于m个样本数据确定的分级参考结果，α_k为第k个决策树模型对应的权重系数，其中，k＝1,2,3,.....,K，K和R均为大于或等于2的整数；

其中，权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的。

2.根据权利要求1所述的方法，其特征在于，权重系数α_k是通过如下计算表达式得到的，包括：

$$\alpha_k=\frac{1}{2}\log\frac{1-e_k}{e_k}+\log(R-1)$$

。

3.根据权利要求2所述的方法，其特征在于，第k个决策树模型分级操作对应的分类误差e_k是通过如下计算表达式得到的，包括：

$$e_k=\sum_{i=1}^{m}w_{k,i}1_{\{G_k(x_i)\neq y_i\}};$$

在上述表达式中，$w_{k,i}$表示第k个决策树模型中第i个特征向量的权重，$G_k(x_i)$为第k个决策树模型基于第i个特征向量中的特征确定的标签参考结果。

4.根据权利要求3所述的方法，其特征在于：

在k等于1时，m组特征向量的权重$w_{1,i}$为预设值；

在k大于1时，第k个决策树模型分级操作m组特征向量的权重$w_{k,i}$是通过如下方式得到的，包括：

$$w_{k,i}=\frac{w_{k-1,i}\exp(-\alpha_{k-1}y_iG_{k-1}(x_i))}{\sum_{i=1}^{m}w_{k-1,i}\exp(-\alpha_{k-1}y_iG_{k-1}(x_i))};$$

在上述表达式中，$w_{k-1,i}$表示第k-1个决策树模型中第i个特征向量的权重，$G_k^{-1}(x_i)$为第k-1个决策树模型中基于第i个特征向量中的特征x_i确定的标签参考结果；α_{k-1}为第k-1个决策树模型分级操作对应的权重系数。

5.根据权利要求4所述的方法，其特征在于：

在k等于1时，每个特征向量对应权重均为1/m。

6.根据权利要求1至5任一项所述的方法，其特征在于：

所述特征向量的维度是根据分级操作的分级属性信息设置的。

7.根据权利要求6所述的方法,其特征在于:

所述特征向量中至少两个维度的信息用于描述同一分级属性信息。

8.一种数据安全性分级装置,包括:

获取模块,设置为获取m个样本数据,其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i,其中,i=1,2,3,...,m,n和m均为大于或等于2的整数,y_i=1,2,3,...,R,其中R为等级的最大值;

建立模块,设置为利用m个样本数据的特征向量以及对应的标签值建立分级模型;

分级模块,设置为利用所述分级模型对待分级数据的特征向量进行处理,得到所述待分级数据对应的等级;

其中,所述分级模型的表达式如下:

$$f(x)=\underset{\{y=1,2,\ldots\ldots,R\}}{\arg\max}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right);$$

在上述表达式中,f(x)为对待分级数据x进行分级操作确定的等级,$G_k(x)$为第k个决策树模型基于m个样本数据确定的分级参考结果,α_k为第k个决策树模型对应的权重系数,其中,k=1,2,3,.....,K,K和R均为大于或等于2的整数;

其中,权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的。

9.一种存储介质,其特征在于,所述存储介质中存储有计算机程序,其中,所述计算机程序被设置为运行时执行所述权利要求1至7任一项中所述的方法。

10.一种电子装置,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运行所述计算机程序以执行所述权利要求1至7任一项中所述的方法。

一种数据安全性分级方法、装置、存储介质和电子装置

技术领域

[0001] 本申请实施例涉及数据处理领域，尤指一种数据安全性分级方法、装置、存储介质和电子装置。

背景技术

[0002] 数据在经济社会发展具有非常高的重要性，一旦遭到非法利用，会造成较大危害程度，因此，对数据实行分类分级保护。

[0003] 由于人工逐字段进行数据安全分级，效率太低，且误差较大，目前可以借助数据安全分级算法实现自动化或半自动化的安全分级，极大地减少人工工作量，同时可避免人为操作的主观性和不稳定性。

[0004] 采用基于本体论和词频的数据分级方法的分级准确率有待进一步提高。

发明内容

[0005] 为了解决上述任一技术问题，本申请实施例提供了一种数据安全性分级方法、装置、存储介质和电子装置。

[0006] 为了达到本申请实施例目的，本申请实施例提供了一种数据安全性分级方法，包括：

[0007] 获取m个样本数据，其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i，其中，i＝1,2,3,...,m,n和m均为大于或等于2的整数，y_i＝1,2,3,...,R,其中R为等级的最大值；

[0008] 利用m个样本数据的特征向量以及对应的标签值建立分级模型；

[0009] 利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级；

[0010] 其中，所述分级模型的表达式如下：

[0011] $$f(x)=\arg\max_{\{y=1,2,\ldots,R\}}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right);$$

[0012] 在上述表达式中，f(x)为对待分级数据x进行分级操作确定的等级，$G_k(x)$为第k个决策树模型基于m个样本数据确定的分级参考结果，α_k为第k个决策树模型对应的权重系数，其中，k＝1,2,3,.....,K,K和R均为大于或等于2的整数；

[0013] 其中，权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的。

[0014] 一种数据安全性分级装置，包括：

[0015] 获取模块，设置为获取m个样本数据，其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i，其中，i＝1,2,3,...,m,n和m均为大于或等于2的整数，y_i＝1,2,3,...,R,其中R为等级的最大值；

[0016] 建立模块，设置为利用m个样本数据的特征向量以及对应的标签值建立分级模型；

[0017] 分级模块，设置为利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级；

[0018] 其中，所述分级模型的表达式如下：

[0019] $$f(x)=\underset{\{y=1,2,\ldots\ldots,R\}}{\arg\max}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right)$$

[0020] 在上述表达式中，f(x)为对待分级数据x进行分级操作确定的等级，$G_k(x)$为第k个决策树模型基于m个样本数据确定的分级参考结果，α_k为第k个决策树模型对应的权重系数，其中，k＝1,2,3,.....,K,K和R均为大于或等于2的整数；

[0021] 其中，权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的。

[0022] 一种存储介质，所述存储介质中存储有计算机程序，其中，所述计算机程序被设置为运行时执行上文所述的方法。

[0023] 一种电子装置，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运行所述计算机程序以执行上文所述的方法。

[0024] 上述技术方案中的一个技术方案具有如下优点或有益效果：

[0025] 由于系数α_k是决策树在第k次分级操作对应的分类误差e_k确定的，因此，分级模型能够体现决策树的分类准确度，使得确定的输入数据的等级具有较高的准确度，达到提高分级操作的准确度的目的。

[0026] 本申请实施例的其它特征和优点将在随后的说明书中阐述，并且，部分地从说明书中变得显而易见，或者通过实施本申请实施例而了解。本申请实施例的目的和其他优点可通过在说明书、权利要求书以及附图中所特别指出的结构来实现和获得。

附图说明

[0027] 附图用来提供对本申请实施例技术方案的进一步理解，并且构成说明书的一部分，与本申请实施例的实施例一起用于解释本申请实施例的技术方案，并不构成对本申请实施例技术方案的限制。

[0028] 图1为本申请实施例提供的数据安全性分级方法的流程图；

[0029] 图2为本申请实施例提供的特征向量与分级属性信息的对应关系示意图；

[0030] 图3为本申请实施例提供的数据集中标签值的分布示意图；

[0031] 图4为本申请实施例提供的测试集的标签图；

[0032] 图5为本申请实施例提供的测试集分级结果的准确度的对比图；

[0033] 图6为本申请实施例提供的数据安全性分级装置的结构示意图。

具体实施方式

[0034] 为使本申请实施例的目的、技术方案和优点更加清楚明白，下文中将结合附图对本申请实施例的实施例进行详细说明。需要说明的是，在不冲突的情况下，本申请实施例中的实施例及实施例中的特征可以相互任意组合。

[0035] 图1为本申请实施例提供的数据安全性分级方法的流程图。如图1所示，所述方法包括：

[0036] 步骤101、获取m个样本数据，其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i，其中，i＝1,2,3,...,m，n和m均为大于或等于2的整数，y_i＝1,2,3,...,R，其中R为等级的最大值；

[0037] 步骤102、利用m个样本数据的特征向量以及对应的标签值建立分级模型；

[0038] 步骤103、利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级；

[0039] 其中，所述分级模型的表达式如下：

[0040] $$f(x)=\arg\max_{\{y=1,2,\ldots,R\}}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right);$$

[0041] 在上述表达式中，f(x)为对待分级数据x进行分级操作确定的等级，$G_k(x)$为第k个决策树模型基于m个样本数据确定的分级参考结果，α_k为第k个决策树模型对应的权重系数，其中，k＝1,2,3,.....,K，K和R均为大于或等于2的整数；

[0042] 其中，权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的；

[0043] 在上述示例性实施例中，$$1_{\{G_k(x)=y\}}=\begin{cases}1, & G_k(x)=y\\ 0, & G_k(x)\neq y\end{cases};$$

[0044] 其中，可以使用ID3(Iterative Dichotomiser 3，二叉树三代)、C4.5、C5.0、CART(Classification and Regression Trees，分类回归树)等具体算法其中；

[0045] 其中，argmax()是求自变量最大的函数，因此，上述分级模型利用argmax()函数来获取系数α_k在K次总和的最大值所对应的y的取值。

[0046] 由于系数α_k是决策树在第k次分级操作对应的分类误差e_k确定的，因此，根据系数α_k在K次总和的最大值可以记录有分类误差，利用系数α_k来体现每次分级操作中决策树的分类准确度，并将K次所确定的系数α_k的总和作为K次分级操作的准确度。

[0047] 由于系数α_k在K次总和的最大值对应的y的取值能够体现分类准确度，因此，利用该分级模型确定的输入数据的等级，具有较高的准确度。

[0048] 本申请实施例提供的方法，由于系数α_k是决策树在第k次分级操作对应的分类误差e_k确定的，因此，分级模型能够体现决策树的分类准确度，使得确定的输入数据的等级具有较高的准确度，达到提高分级操作的准确度的目的。

[0049] 下面对本申请实施例提供的方法进行说明：

[0050] 在一个示例性实施例中，权重系数α_k是通过如下计算表达式得到的，包括：

[0051] $$\alpha_k=\frac{1}{2}\log\frac{1-e_k}{e_k}+\log(R-1)$$。

[0052] 在上述表达式中，不但使用了分类误差e_k用以体现分类准确度信息，还使用了等级的最大值以适应当前分级操作对应的等级，进一步提高提高分级模型对当前等级的适配性。

[0053] 在一个示例性实施例中，第k个决策树模型分级操作对应的分类误差e_k是通过如下计算表达式得到的，包括：

[0054] $e_k = \sum_{i=1}^{m} w_{k,i} 1_{\{G_k(x_i) \neq y_i\}}$；

[0055] 在上述表达式中，$w_{k,i}$表示第k个决策树模型中第i个特征向量的权重，$G_k(x_i)$为第k个决策树模型基于第i个特征向量中的特征确定的标签参考结果；

[0056] 其中，$1_{\{G_k(x_i) \neq y_i\}} = \begin{cases} 1, & G_k(x_i) \neq y_i \\ 0, & G_k(x_i) = y_i \end{cases}$。

[0057] 上述分类误差的计算方式仅为示意，并不作限定，可以采用其他方式进行分类误差的计算。

[0058] 在一个示例性实施例中，在k等于1时，m组特征向量的权重$w_{1,i}$为预设值；

[0059] 优选的，在k等于1时，每个特征向量对应权重均为1/m。

[0060] 在一个示例性实施例中，在k大于1时，第k次分级操作m组特征向量的权重$w_{k,i}$是通过如下方式得到的，包括：

[0061] $w_{k,i} = \frac{w_{k-1,i} \exp(-\alpha_{k-1} y_i G_{k-1}(x_i))}{\sum_{i=1}^{m} w_{k-1,i} \exp(-\alpha_{k-1} y_i G_{k-1}(x_i))}$；

[0062] 在上述表达式中，$w_{k-1,i}$表示第k-1个决策树模型中第i个特征向量的权重，$G_{k-1}(x_i)$为第k-1个决策树模型中基于第i个特征向量中的特征x_i确定的标签参考结果；α_{k-1}为第k-1个决策树模型分级操作对应的权重系数。

[0063] 采用上述方式更新下一次分级操作对应的权重，能够提高权重的取值的准确度。

[0064] 在一个示例性实施例中，所述特征向量的维度是根据分级操作的分级属性信息设置的。

[0065] 以应用场景为民航通信系统所产生的数据为例进行说明：

[0066] 随着民航空管信息化建设的全面推进，民航通信系统产生的数据非常庞大。对上述数据进行安全属性的分级操作时，可以将数据的可用性、完整性和机密性作为分级属性信息执行分级操作。

[0067] 进一步的，根据分级属性信息确定特征向量，且同一分级属性信息还可以通过至少两个特征向量来描述。

[0068] 图2为本申请实施例提供的特征向量与分级属性信息的对应关系示意图。如图2所示，分级属性信息为可用性、完整性和机密性，其中：

[0069] 可用性对应有5个特征向量，其中：

[0070] 使用频率A_1，用于表征数据项使用频率高低；

[0071] 恢复时间要求A_2，用于表征数据项出现问题后恢复时间要求；

[0072] 是否有替代方式A_3，用于表征数据项是否有替代数据；

[0073] 不可用的影响范围A_4，用于表征数据项不可用的影响范围大小；

[0074] 不可用的损失大小A_5，用于表征数据项不可用的损失大小；

[0075] 完整性对应有3个特征向量，其中：

[0076] 是否需要完整性检查I_1，用于表征数据项是否需要完整性检查；

[0077] 完整性被破坏影响范围I_2，用于表征数据项完整性被破坏影响范围为内部还是外部；

[0078] 完整性被破坏损失大小I_3,用于表征数据项完整性被破坏损失大小;

[0079] 机密性对应有5个特征向量,其中:

[0080] 数据类型C_1,用于表征数据项的数据类型:公开,内部,敏感,高敏感数据泄密影响范围C_2,用于表征数据项泄密影响范围为内部还是外部数据泄密经济损失大小C_3,用于表征数据项泄密造成的经济损失大小;

[0081] 数据泄密经济社会影响C_4,用于表征数据项泄密造成的经济社会影响大小;

[0082] 数据泄密其他损失大小C_5,用于表征数据项泄密造成的非经济损失大小。

[0083] 在上述每个特征向量的可选值分别为:A_1={低/0.25,一般/0.50,频繁/0.75,非常频繁/1.00},A_2={长/0.25,一般/0.50,短/0.75,实时切换/1.00},A_3={是/0.00,否/1.00},A_4={内部/0.50,外部/1.00},A_5={间接/0.25,小/0.50,中/0.75,大/1.00},I_1={否/0.00,是/1.00},I_2={内部/0.50,外部/1.00},I_3={间接/0.25,小/0.50,中/0.75,大/1.00},C_1={公开/0.25,内部/0.50,敏感/0.75,高敏感/1.00},C_2={内部/0.50,外部/1.00},C_3={间接/0.25,小/0.50,中/0.75,大/1.00},C_4={间接/0.25,小/0.50,中/0.75,大/1.00},C_5={间接/0.25,小/0.50,中/0.75,大/1.00}。

[0084] 在上述应用场景中,分级操作从5个等级中确定数据所属的一个等级,即,R=5。其中:

[0085] 1级表示极低共享风险:数据一旦遭到篡改、破坏、泄露或者非法获取、非法利用,可能对个人合法权益、组织合法权益造成轻微危害,但不会危害国家安全、公共利益;

[0086] 2级表示低共享风险:数据一旦遭到篡改、破坏、泄露或者非法获取、非法利用,可能对个人合法权益、组织合法权益造成一般危害,或者对公共利益造成轻微危害,但不会危害国家安全;

[0087] 3级表示中共享风险:数据一旦遭到篡改、破坏、泄露或者非法获取、非法利用,可能对个人合法权益、组织合法权益造成严重危害,或者对公共利益造成一般危害,但不会危害国家安全;

[0088] 4级表示高共享风险:数据一旦遭到篡改、破坏、泄露或者非法获取、非法利用,可能对个人合法权益、组织合法权益造成特别严重危害,可能对公共利益造成严重危害,或者对国家安全造成轻微或一般危害;

[0089] 5级表示极高共享风险:数据一旦遭到篡改、破坏、泄露或者非法获取、非法利用,可能对国家安全造成严重或特别严重危害,或对公共利益造成特别严重危害。

[0090] 以某一民航通信系统中标注数据集为例进行说明:

[0091] 该标注数据集包含数据项986个,其中前20项的数据的特征向量的标签如表1所示:

[0092] 表1

[0093]

A1	A2	A3	A4	A5	I1	I2	I3	C1	C2	C3	C4	C5	L
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	3
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	3
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	2
0.75	0.75	1	0.5	0.25	1	0.5	0.25	0.25	1	0.5	0.5	0.5	1
1	0.75	1	1	0.75	1	1	0.75	0.75	1	0.75	0.75	0.5	3

0.75	0.75	1	1	0.5	1	1	0.5	0.75	1	0.75	0.75	0.5	2
1	0.75	1	1	0.75	1	1	0.75	0.75	1	0.75	0.75	0.5	3
0.75	0.75	1	0.5	0.5	1	0.5	0.5	0.75	1	0.75	0.75	0.5	2
0.5	0.75	1	0.5	0.5	1	0.5	0.5	0.75	1	0.75	0.75	0.5	3
1	0.75	1	1	0.5	1	1	0.5	0.25	1	0.5	0.5	0.5	2
1	0.75	1	1	0.5	1	1	0.5	0.25	1	0.5	0.5	0.5	2
1	0.75	1	1	0.5	1	1	0.5	0.25	1	0.5	0.5	0.5	2
0.25	0.75	1	0.5	0.25	1	0.5	0.25	0.5	1	0.5	0.5	0.5	2
0.75	0.75	1	1	0.5	1	1	0.5	0.75	1	0.75	0.75	0.5	2
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	3
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	2
1	0.75	1	1	0.75	1	1	0.75	0.25	1	0.5	0.5	0.5	3
0.75	0.75	1	0.5	0.25	1	0.5	0.25	0.5	1	0.5	0.5	0.5	1
1	0.75	1	1	0.75	1	1	0.75	0.75	1	0.75	0.75	0.5	3
1	0.75	1	1	0.5	1	1	0.5	0.25	1	0.5	0.5	0.5	2

[0094] 在上述数据集中标签值的分布示意图如图3所示。

[0095] 在上述特征向量下，将数据集按照4:1划分为训练集和测试集，测试集的标签图如图4所示。

[0096] 在本申请中，以决策树为基分类器的Adaboost算法构建数分级模型，将申请提供的方法与Linear Regression,Ridge,Lasso,Elastic Net,Decision Tree,Random Forest,Extra Tree,Bagging,Gradient Boosting,XGB,XGBRF等方法在训练集上进行训练，并在测试集上进行对比，对比结果参见图5所示，本申请所提基于决策树为基分类器的Adaboost算法的分级准确率最高，可达95.5%。

[0097] 综上所述，本申请基于数据的安全特性设计了5级分级框架，基于以决策树为基分类器的Adaboost方法构造数据自动分级算法，可以高效准确实现民航空管数据安全性自动分级。

[0098] 图6为本申请实施例提供的数据安全性分级装置的结构示意图。如图6所示，所述装置包括：

[0099] 获取模块10，设置为获取m个样本数据，其中第i个样本数据包括n维的特征向量x_i以及与n维特征向量x_i一一对应的标签值y_i，其中，i＝1,2,3,...,m,n和m均为大于或等于2的整数，y_i＝1,2,3,...,R,其中R为等级的最大值；

[0100] 建立模块20，设置为利用m个样本数据的特征向量以及对应的标签值建立分级模型；

[0101] 分级模块30，设置为利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级；

[0102] 其中，所述分级模型的表达式如下：

[0103] $$f(x)=\underset{\{y=1,2,\ldots,R\}}{\arg\max}\left(\sum_{k=1}^{K}\alpha_k 1_{\{G_k(x)=y\}}\right);$$

[0104] 在上述表达式中，f(x)为对待分级数据x进行分级操作确定的等级，$G_k(x)$为第k个

决策树模型基于m个样本数据确定的分级参考结果，α_k为第k个决策树模型对应的权重系数，其中，k＝1,2,3,.....,K,K和R均为大于或等于2的整数；

[0105] 其中，权重系数α_k是基于第k个决策树模型分级操作对应的分类误差e_k以及等级最大值R确定的。

[0106] 本申请实施例提供的装置，由于系数α_k是决策树在第k次分级操作对应的分类误差e_k确定的，因此，分级模型能够体现决策树的分类准确度，使得确定的输入数据的等级具有较高的准确度，达到提高分级操作的准确度的目的。

[0107] 本申请实施例提供一种存储介质，所述存储介质中存储有计算机程序，其中，所述计算机程序被设置为运行时执行上文任一项中所述的方法。

[0108] 本申请实施例提供一种电子装置，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运行所述计算机程序以执行上文任一项中所述的方法。

[0109] 本领域普通技术人员可以理解，上文中所公开方法中的全部或某些步骤、系统、装置中的功能模块/单元可以被实施为软件、固件、硬件及其适当的组合。在硬件实施方式中，在以上描述中提及的功能模块/单元之间的划分不一定对应于物理组件的划分；例如，一个物理组件可以具有多个功能，或者一个功能或步骤可以由若干物理组件合作执行。某些组件或所有组件可以被实施为由处理器，如数字信号处理器或微处理器执行的软件，或者被实施为硬件，或者被实施为集成电路，如专用集成电路。这样的软件可以分布在计算机可读介质上，计算机可读介质可以包括计算机存储介质(或非暂时性介质)和通信介质(或暂时性介质)。如本领域普通技术人员公知的，术语计算机存储介质包括在用于存储信息(诸如计算机可读指令、数据结构、程序模块或其他数据)的任何方法或技术中实施的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于RAM、ROM、EEPROM、闪存或其他存储器技术、CD-ROM、数字多功能盘(DVD)或其他光盘存储、磁盒、磁带、磁盘存储或其他磁存储装置、或者可以用于存储期望的信息并且可以被计算机访问的任何其他的介质。此外，本领域普通技术人员公知的是，通信介质通常包含计算机可读指令、数据结构、程序模块或者诸如载波或其他传输机制之类的调制数据信号中的其他数据，并且可包括任何信息递送介质。

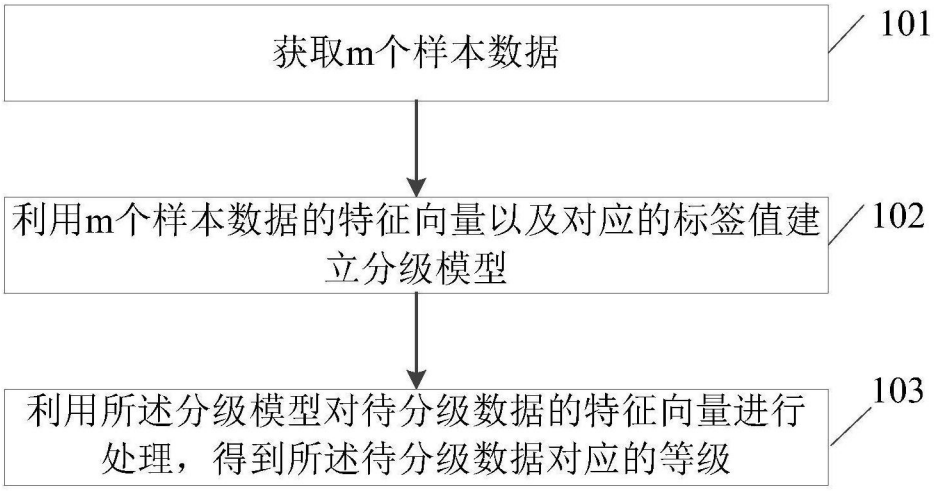

获取m个样本数据
101
利用m个样本数据的特征向量以及对应的标签值建立分级模型
102
利用所述分级模型对待分级数据的特征向量进行处理，得到所述待分级数据对应的等级
103

图1

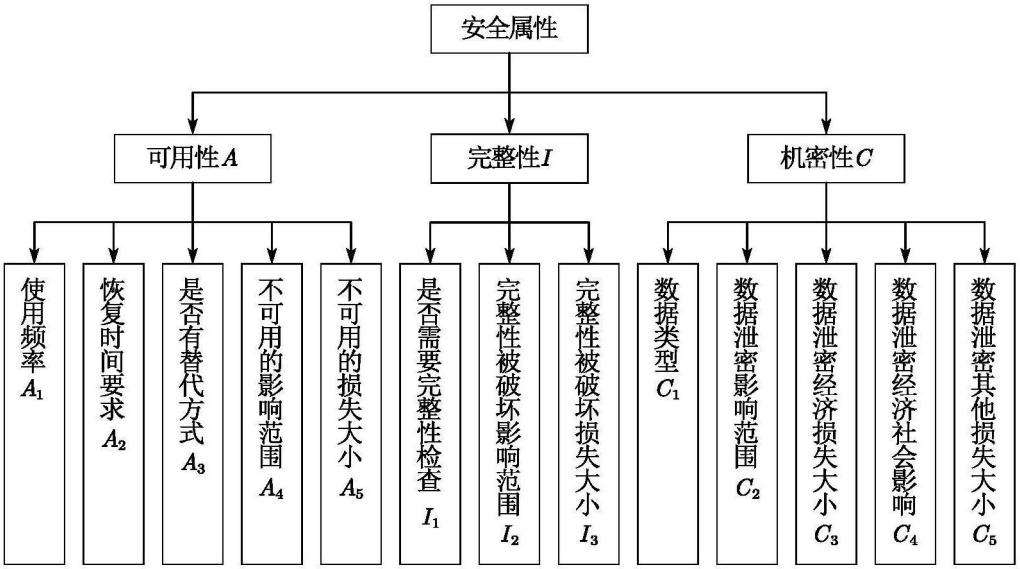

安全属性
可用性A
完整性I
机密性C
使用频率 A_1
恢复时间要求 A_2
是否有替代方式 A_3
不可用的影响范围 A_4
不可用的损失大小 A_5
是否需要完整性检查 I_1
完整性被破坏影响范围 I_2
完整性被破坏损失大小 I_3
数据类型 C_1
数据泄密影响范围 C_2
数据泄密经济损失大小 C_3
数据泄密经济社会影响 C_4
数据泄密其他损失大小 C_5

图2

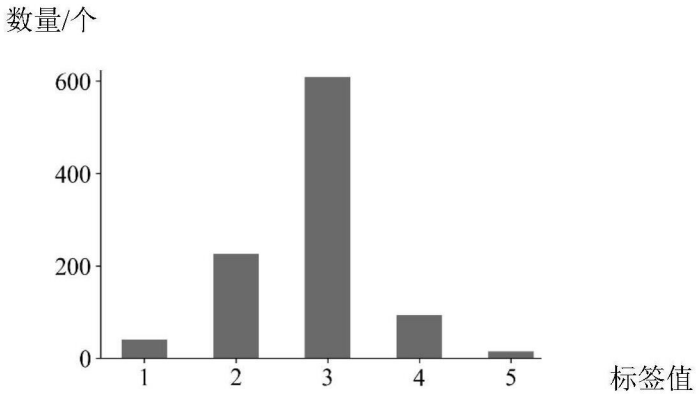


图3

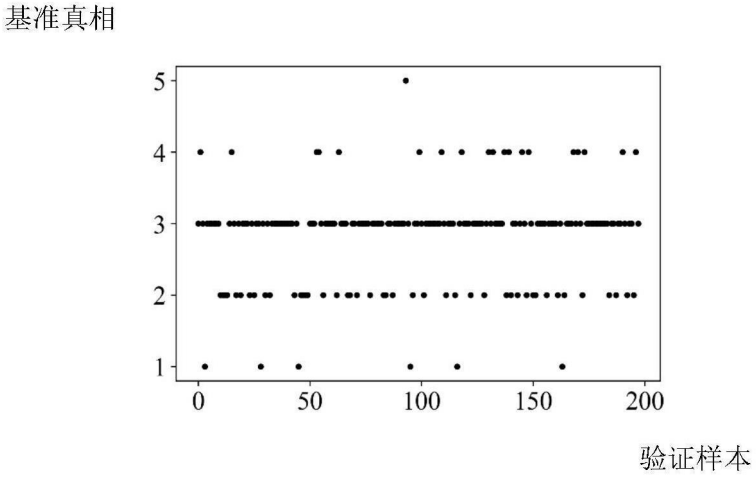


图4

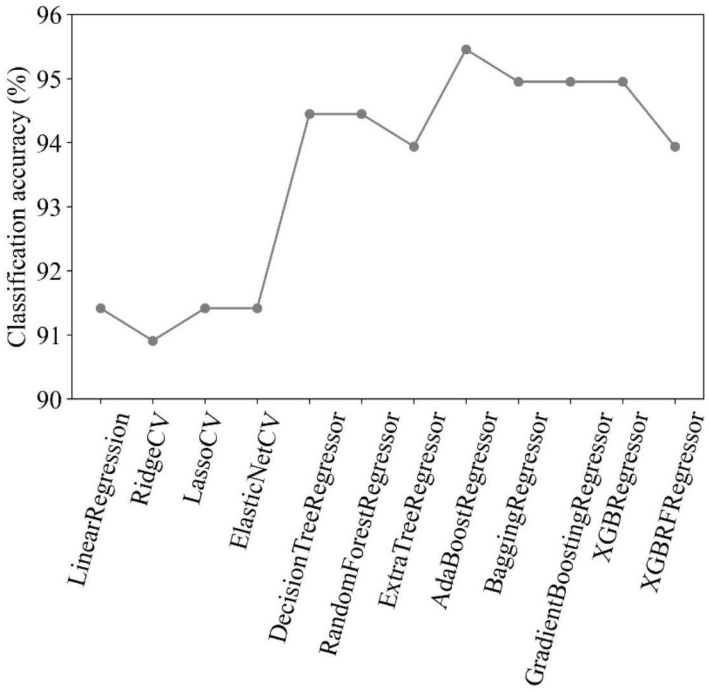

Classification accuracy (%)
96
95
94
93
92
91
90
LinearRegression
RidgeCV
LassoCV
ElasticNetCV
DecisionTreeRegressor
RandomForestRegressor
ExtraTreeRegressor
AdaBoostRegressor
BaggingRegressor
GradientBoostingRegressor
XGBRegressor
XGBRFRegressor

图5


10
20
30
获取模块
建立模块
分级模块

图6