

(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
2019 年 1 月 31 日 (31.01.2019)



(10) 国际公布号
WO 2019/019375 A1

(51) 国际专利分类号:
G06Q 40/08 (2012.01) **G06F 17/30** (2006.01)

(21) 国际申请号: PCT/CN2017/104598

(22) 国际申请日: 2017 年 9 月 29 日 (29.09.2017)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201710618080.0 2017 年 7 月 26 日 (26.07.2017) 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区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(72) 发明人: 邵正铂(SHAO, Zhengbo); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 李斌(LL, Bin); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 陈杰(CHEN, Jie); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 高雪(GAO, Xue); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 马向东(MA, Xiangdong); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 丁杰(DING, Jie); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。 张捷(ZHANG, Jie); 中国广东省深圳市福田区八卦岭八卦三路平安大厦, Guangdong 518000 (CN)。

(54) Title: METHOD AND APPARATUS FOR CREATING UNDERWRITING DECISION TREE, AND COMPUTER DEVICE AND STORAGE MEDIUM

(54) 发明名称: 核保决策树的创建方法、装置、计算机设备及存储介质

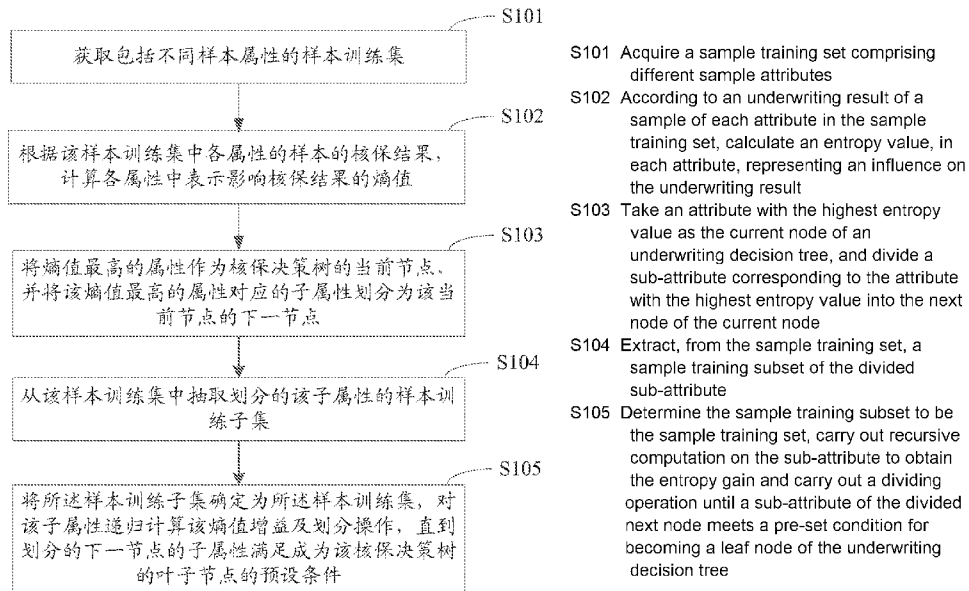


图 1

(57) Abstract: A method for creating an underwriting decision tree. The method comprises: acquiring a sample training set comprising different sample attributes (S101); according to an underwriting result of a sample of each attribute in the sample training set, calculating an entropy value, in each attribute, representing an influence on the underwriting result (102); taking an attribute with the highest entropy value as the current node of the underwriting decision tree, and dividing a sub-attribute corresponding to the attribute with the highest entropy value into the next node of the current node (S103); extracting, from the sample training set, a sample training subset of the divided sub-attribute (S104); and determining the sample training subset to be the sample training set, carrying out recursive computation on the sub-attribute to obtain the entropy gain and carrying out a dividing operation until a sub-attribute of the divided next node meets a pre-set condition for becoming a leaf node of the underwriting decision tree (105).



(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区花城大道 85 号 3901 房, Guangdong 510623 (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, 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))。

(57) 摘要: 一种核保决策树的创建方法, 该方法包括: 获取包括不同样本属性的样本训练集 (S101); 根据该样本训练集中各属性的样本的核保结果, 计算各属性中表示影响核保结果的熵值 (102); 将熵值最高的属性作为该核保决策树的当前节点, 并将该熵值最高的属性对应的子属性划分为该当前节点的下一节点 (S103); 从该样本训练集中抽取划分的该子属性的样本训练子集 (S104); 将该样本训练子集确定为该样本训练集, 对该子属性递归计算该熵值增益及划分操作, 直到该划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件 (105)。

核保决策树的创建方法、装置、计算机设备及存储介质

本申请要求于 2017 年 7 月 26 日提交中国专利局，申请号为 201710618080.0，发明名称为“核保决策树的创建方法、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
5 中。

技术领域

本申请涉及保险技术领域，特别是涉及一种核保决策树的创建方法、装置、计算机设备及存储介质。

10

背景技术

在保险这一领域中，经常需要对用户的投保单进行核保，依据核保人的年龄、所从事的行业、收入、性别的等相关信息审核对应的保单是否可以承保。

15

目前对于用户的投保单的核保工作主要是人工来审核，审核的依据一方面是基于被保人的相关信息，另一方面还基于审核人的工作经验，但是对于个人而言其工作经验毕竟有限，如果没有对应的直观历史数据作为参考难以对用户的保单进行准确的审核。

20

发明内容

根据本申请的各种实施例，提供一种核保决策树的创建方法、计算机设备及存储介质。

25

一种核保决策树的创建方法，该方法包括：获取包括不同样本属性的样本训练集；根据该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点；从该样本训练集中抽取划分的该子属性的样本训练子集；将所述样本训练子集确定为所述样本训练集，对该子属性递归计算上述熵值增益及划分操

作，直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

一种核保决策树的创建装置，该装置包括：样本获取模块，用于获取包括不同样本属性的样本训练集；熵值增益计算模块，用于根据该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；
5 节点划分模块，用于将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点；子集抽取模块，用于从该样本训练集中抽取划分的该子属性的样本训练子集；递归模块，用于将所述样本训练子集确定为所述样本训练集，对该子属性递归
10 计算上述熵值增益及划分操作，直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：获取包括不同样本属性的样本训练集；根据
15 该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点；从该样本训练集中抽取划分的该子属性的样本训练子集；将所述样本训练子集确定为所述样本训练集，对该子属性递归计算上述熵值增益及划分操作，直到划分的
20 下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

一个或多个存储有计算机可读指令的非易失性可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：获取包括不同样本属性的样本训练集；根据该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点；从该样本训练集中抽取划分的该子属性的样本训练子集；将所述样本训练子集确定为所述样本训练集，对该子属性递归计算上述熵值增益及划分操作，直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。
25

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

5 附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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它的附图。

- 10 图 1 为一个实施例中核保决策树的创建方法的流程图；
- 图 2 为另一实施例中核保决策树的创建方法的流程图；
- 图 3 为又一实施例中核保决策树的创建方法的流程图；
- 图 4 为一个实施例中使用场景示意图；
- 图 5 为一个实施例中核保决策树的创建装置的示范性结构框图；
- 15 图 6 为一个实施例中计算机设备的内部结构示意图。

具体实施方式

- 为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处所描述的具体实施
- 20 例仅仅用以解释本申请，并不用于限定本申请。

图 1 为根据本申请的一个实施例的核保决策树的创建方法的流程图，下面结合图 1 来详细描述根据本申请的一个实施例的核保决策树的创建方法，如图 1 所示，该方法包括以下步骤 S101、S102、S103、S104 及 S105。

S101、获取包括不同样本属性的样本训练集。

- 25 根据本实施例的一个示例，该样本训练集的来源为从历史核保记录中选取的样本数据，将从历史核保记录中选取的样本数据作为创建核保决策树的依据对于审核人的工作具有更强的指导性。

在该步骤中，上述的属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率，其中年龄这一属性的子属性包括低龄、高龄和中龄，行

业风险这一属性的子属性包括高风险、低风险和中风险，既往病史这一属性的子属性包括是和否，赔付率这一属性的子属性包括赔付率高和赔付率低。

根据本实施例的一个示例获取的样本训练集如下表（1）所示：

年龄	行业风险	既往病史	赔付率	是否核保通过	计数
低龄	高	否	高	通过	640
高龄	高	否	高	不通过	1280
中龄	中	否	高	不通过	600
低龄	高	否	低	通过	640
中龄	低	是	高	不通过	640
中龄	低	是	低	通过	640
高龄	低	是	低	不通过	640
低龄	中	否	低	通过	1280
中龄	中	是	高	不通过	1320
低龄	中	是	低	不通过	640
高龄	中	否	低	不通过	320
低龄	低	是	高	不通过	640
高龄	高	是	高	不通过	320
中龄	中	否	低	通过	630
中龄	中	否	低	不通过	10

表（1）

- 5 其中，各年龄段的年龄区间可以通过实际业务的需求的来设定，根据本实施例的一个示例例如：可以将 0~25 岁设为低龄，可以将 26~45 岁设为中龄，可以将 46 岁及以上岁数以上设为高龄。

S102、根据该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益。

- 10 根据本实施例的一个示例，该步骤 S102 具体还包括：从该样本训练集中抽取同一属性的样本的核保结果，再根据同一属性的核保结果计算该属性的熵值增益。

在一个实施例中，该步骤 S102 中的核保结果包括是否核保通过及对应属

性的核保通过率和核保不通过率，通过以下公式计算熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2 (1 - M)) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i}) \quad (1);$$

其中，M表示样本训练集中总核保通过率， A_i 表示对应属性A的子属性i的数量占样本训练集中的总数量的比例， B_i 表示子属性i以属性A的数量为基数的核保通过率，n表示对应属性A的子属性的个数， G_A 表示计算的属性A的熵值增益。

在一个实施例中，可以依据该样本训练集先计算出该样本训练集总的决策熵值，再计算出样本训练集中的其中一个属性的熵值，然后将决策熵值与样本训练集中的其中一个属性的熵值的差值作为该属性的熵值增益，熵值增益的意义在于可以表示该属性对于核保结果的影响力，熵值增益越大表示对于核保结果的影响力度越大。

根据本实施例的一个使用场景，参照例如在上述样本训练集中抽取的年龄这一属性的样本的核保结果如下表（2）所示：

年龄	核保决定	数量
中龄	通过	1270
中龄	不通过	2570
低龄	通过	2560
低龄	不通过	1280
高龄	不通过	2560
高龄	通过	0

表（2）

根据上述表（1）可以得出：

$$\text{总的通过率 } M = \frac{3830}{6410 + 3830}, \text{ 总的不通过率 } (1 - M) = \frac{6410}{6410 + 3830};$$

当上述属性A表示年龄这一属性时，该属性A的子属性i包括中龄、低龄和高龄，根据上述表（1）和表（2）可以得出：

年龄属性的子属性为低龄的数量占样本训练集中的总数量的比例

$A_i = \frac{3840}{10240}$ ，该子属性低龄以属性 A 的数量为基数的核保通过率

$B_i = \frac{2560}{1280 + 2560}$ ，该子属性低龄以属性 A 的数量为基数的核保不通过率

$(1-B_i) = \frac{1280}{1280 + 2560}$ ，可以计算出决策熵值 S_0 ：

$$S_0 = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)})$$

$$5 \quad = -\left(\frac{3830}{6410 + 3830} \times \log_2 \frac{3830}{6410 + 3830} + \frac{6410}{6410 + 3830} \times \log_2 \frac{6410}{6410 + 3830}\right) = 0.9573$$

；

还可以计算出低龄这一子属性的熵值 S_{A1} ：

$$S_{A1} = -B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i}$$

$$10 \quad = -\left(\frac{2560}{1280 + 2560} \times \log_2 \frac{2560}{1280 + 2560} + \frac{1280}{1280 + 2560} \times \log_2 \frac{1280}{1280 + 2560}\right) = 0.9138 ;$$

同理，可以计算出中龄这一子属性的熵值为 0.9157，老龄这一属性的熵值为 0，然后利用以下公式（2）计算出年龄属性 A 的熵值：

$$S_A = \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i}) \quad (2);$$

可以计算出年龄属性 A 的熵值大小为：

$$15 \quad S = \frac{3840}{10240} \times 0.9183 + \frac{3840}{10240} \times 0.9157 + \frac{2560}{10240} \times 0 = 0.6877 ;$$

再通过上述公式（1）可以得出年龄这一属性的熵值增益 G_A 为：

$$G_A = 0.9537 - 0.6877 = 0.2660。$$

同理，参照上述表（1）可计算出行业风险、既往病史、赔付率的熵值增益分别为 0.0176、0.1726 和 0.0453。

S103、将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点。

由于熵值增益的意义在于可以表示该属性对于核保结果的影响力，熵值增益越大表示对于核保结果的影响力度越大，将熵值增益最大的属性作为该核保决策树的当前节点有利于核保人员对于处于该核保决策树上层节点的属性进行重点审核，有利于提高核保的准确性。

根据本实施例的使用场景例如，当通过上述步骤 S102 的步骤计算的年龄这一属性的熵值增益最高的属性时，则将年龄这一属性作为该核保决策树的当前节点。

S104、从该样本训练集中抽取划分的该子属性的样本训练子集。

根据本实施例的一个示例，当上述熵值增益最高的属性为年龄这一属性，年龄属性中对应的子属性包括低龄、高龄和中龄，根据本实施例的一个使用场景抽取出来的低龄这一子属性的样本训练子集如下表（3）所示：

年龄-低龄	行业分险	既往病史	赔付率	是否核保通过	计数
低龄	高	否	高	通过	640
低龄	高	否	低	通过	640
低龄	中	否	低	通过	1280
低龄	中	是	低	不通过	640
低龄	低	是	高	不通过	640

表（3）

S105、将样本训练子集确定为样本训练集，对该子属性递归计算上述熵值增益及划分操作，直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

根据本实施例的一个示例，该步骤中的递归操作是指将对应的子属性确定为上述公式（1）中的属性 A，依据抽取的样本训练子集，计算属性 A 的子属性的熵值增益并进行核保决策树中分支的划分，直到划分出的划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

根据本实施例的一个使用场景参照表（3）所示，即将上述表（3）确定为上述样本训练集，将低龄这一属性确定为上述公式（1）中的属性 A，逐一

计算行业分险、既往病史、赔付率这些属性的熵值增益，依次递归直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

在一个实施例中，对年龄属性下的每个子属性都进行上述步骤 S104 的抽取操作和步骤 S105 中的递归操作，直到其中的划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

本实施例通过计算样本训练集中各个属性的熵值增益，并将熵值增益最大的属性作为核保决策树的根节点，再通过递归的方式划分出该核保决策树的中间节点属性以及叶子节点属性，以创建基于各个属性的核保决策树，使得核保人可以依据该核保决策树对处于该核保决策树例如根节点等上层的属性进行着重审核，以给用户提供核保依据，提高核保的准确性。

图 2 为根据本申请的另一实施例的核保决策树的创建方法的流程图，如图 2 所示，该核保决策树的创建方法在包括上述步骤 S101 至 S104 的基础上，上述步骤 S105 进一步包括以下步骤 S201。

S201、将样本训练子集确定为样本训练集，对该子属性递归计算上述熵值增益及划分操作，直到划分的子属性只有一个或划分的子属性的核保结果均为通过或者不通过或当该子属性的熵值增益小于预设的阈值时，将该子属性确定为该核保决策树的叶子节点。

根据本实施例的一个示例，当该子属性的熵值增益小于预设的阈值时，还可以对小于预设的阈值的熵值增益对应的子属性做剪枝操作，将该子属性的上一节点属性作为该核保决策树的叶子节点。

图 4 为根据本申请的一个实施例的使用场景示意图，根据本实施例的叶子节点的确定的使用场景如图 4 所示，当子属性老年这一叶子节点的核保结果均为不通过时，将老年这一子属性确定为该核保决策树的叶子节点。根据本实施例的另一使用场景参照上述表 (3)，例如核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——既往病史中，是既往病史的核保通过率均为不通过，非既往病史均为不通过，因此可以将既往病史作为该核保决策树中年龄——低龄——既往病史这一分支的叶子节点。

根据本实施例的另一使用场景例如通过上述递归算法划分的核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——行业风险

中——既往病史否下的子属性仅包括赔付率时，该子属性赔付率可作为核保决策树的叶子节点。

根据本实施例的又一使用场景通过上述递归算法划分的核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——行业风险中——既往病史否，其中既往病史的熵值增益如果小于预设的阈值，则可将既往病史作为该核保决策树的叶子节点，也可将既往病史这一叶子节点做剪枝操作，将该叶子节点的上一节点行业风险作为该核保决策树的叶子节点。

本实施例将熵值增益很小的子属性做剪枝操作，可以将对核保决策影响很小的属性剔除在核保决策树之外，进一步提高了核保决策树所呈现的核保决策的准确性。

图 3 为根据本申请的又一实施例的核保决策树的创建方法的流程图，如图 3 所示，该核保决策树的创建方法在包括上述步骤 S101 至 S105 的基础上还包括以下步骤 S301。

S301、显示该核保决策树并在该核保决策树的叶子节点中显示对应属性的核保结果。

根据本实施例的一个示例，该步骤中的核保结果可以是如图 4 中所示子属性核保通过的数量和核保不通过的数量，还可以是该叶子节点对应子属性的核保的通过率和/或核保的不通过率。

根据本申请的一个实施例还提供了一种利用核保决策树自动进行核保的方法，该方法包括：获取待核保的保单中的各属性，将获取的各属性与核保决策树的各个节点的属性进行匹配，将与该核保决策树的叶子节点匹配成功的属性中该叶子节点所对应的核保结果作为该保单的核保结果。

其中，将获取的各属性与核保决策树的各个节点的属性进行匹配的步骤进一步包括：获取该核保决策树的当前节点的属性，将该待核保的保单中与当前节点属性相同的属性确定为与当前节点的属性匹配成功，进一步获取该保单中与当前节点匹配成功的属性的子属性，然后在核保决策树中查询与获取的子属性相同的分支属性，进一步再在该子属性的其它属性中依次与核保决策树的中间节点进行匹配，直到匹配到该核保决策树的叶子节点，将该叶子节点的核保结果作为该待核保保单的核保结果。

根据本实施例的一个使用场景例如，核保决策树的其中一个分支中从当前节点到叶子节点依次为：年龄——低龄——行业风险高，其中叶子节点属性——行业风险高的核保结果为均不通过，如果待核保的保单中的年龄属性属于低龄的范围，既往病史为是，行业风险为高，则依次将年龄匹配到核保决策树的子属性低龄中，然后获取核保决策树中低龄的下一节点为行业风险高，将待核保的保单中的行业风险为高这一属性匹配到核保决策树中低龄的行业风险高中，然后确定行业风险高为该核保决策树的叶子节点，且该叶子节点的核保结果为均不通过，则不需要再匹配待核保的保单中的既往病史这一属性，可以直接作出该待核保的保单核保不通过的决定。

根据本实施例的一个示例，上述步骤 S101~S301 的标号并不用于限定本实施例中各个步骤的先后顺序，各个步骤的编号只是为了使得描述各个步骤时可以通用引用该步骤的标号进行便捷的指代，只要各个步骤执行的顺序不影响本实施例的逻辑关系均视为在本申请的保护范围之内。

图 5 为根据本申请的一个实施例的核保决策树的创建装置的示范性结构框图，下面结合图 5 来详细描述根据本申请的一个实施例的核保决策树的创建装置，如图 5 所示，该核保决策树的创建装置 10 包括：样本获取模块 11，用于获取包括不同样本属性的样本训练集；熵值增益计算模块 12，用于根据该样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；节点划分模块 13，用于将熵值增益最高的属性作为核保决策树的当前节点，并将该熵值增益最高的属性对应的子属性划分为该当前节点的下一节点；子集抽取模块 14，用于从该样本训练集中抽取划分的该子属性的样本训练子集；递归模块 15，用于将样本训练子集确定为样本训练集，对该子属性递归计算上述熵值增益及划分操作，直到划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

根据本实施例的一个示例，上述样本获取模块具体用于从历史核保记录中选取的样本数据，将从历史核保记录中选取的样本数据作为创建核保决策树的依据对于审核人的工作具有更强的指导性。

在一个实施例中，上述熵值增益计算模块 12 具体还用于从该样本训练集中抽取同一属性的样本的核保结果，再根据同一属性的核保结果计算该属性

的熵值增益。

根据本实施例的一个示例，上述熵值增益计算模块 12 具体还用于：依据该样本训练集先计算出该样本训练集总的决策熵值 S_0 ，再计算出样本训练集中的其中一个属性的熵值 S_A ，然后将决策熵值与样本训练集中的其中一个属性的熵值之差作为该属性的熵值增益，熵值增益的意义在于可以表示该属性对于核保结果的影响力，熵值增益越大表示对于核保结果的影响力度越大。

其中， $S_0 = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)})$ ；

$S_A = \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i})$ ，其中，属性 A 的每个子属性的熵值为 S_{Ai} ：

$$S_{Ai} = -B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i} \quad (1)$$

在一个实施例中，该熵值增益计算模块通过以下公式计算熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i})$$

其中，M 表示样本训练集中总核保通过率， A_i 表示对应属性 A 的子属性 i 的数量占样本训练集中的总数量的比例， B_i 表示子属性 i 以属性 A 的数量为基数的核保通过率，n 表示对应属性 A 的子属性的个数， G_A 表示计算的属性 A 的熵值增益。

其中，该属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率，其中年龄这一属性的子属性包括低龄、高龄和中龄，行业风险这一属性的子属性包括高风险、低风险和中风险，既往病史这一属性的子属性包括是与否，赔付率这一属性的子属性包括赔付率高和赔付率低。

其中，上述递归模块 15 具体用于将对应的子属性确定为上述公式 (1) 中的属性 A，依据抽取的样本训练子集，计算属性 A 的子属性的熵值增益并进行核保决策树中分支的划分，直到划分出的划分的下一节点的子属性满足成为该核保决策树的叶子节点的预设条件。

由于熵值增益的意义在于可以表示该属性对于核保结果的影响力，熵值增益越大表示对于核保结果的影响力度越大，上述节点划分模块 13 将熵值增益最大的属性作为该核保决策树的当前节点有利于核保人员对于处于该核保决策树上层的属性进行重点审核，有利于提高核保的准确性。

根据本实施例的一个示例，该递归模块 15 还包括：第一叶子节点确定单元，用于当划分的子属性只有一个时，将该子属性确定为该核保决策树的叶子节点；或 第二叶子节点确定单元，用于当划分的子属性的核保结果均为通过或者不通过时，将该子属性确定为该核保决策树的叶子节点；或 第三叶子节点确定单元，用于当该子属性的熵值增益小于预设的阈值时，将该子属性确定为该核保决策树的叶子节点。

根据本实施例的另一示例，该第三叶子节点确定单元还用于当该子属性的熵值增益小于预设的阈值时，对小于预设的阈值的熵值增益对应的子属性做剪枝操作，将该子属性的上一节点属性作为该核保决策树的叶子节点。

根据本实施例的一个使用场景例如当子属性老年这一叶子节点的核保结果均为不通过时，将老年这一子属性确定为该核保决策树的叶子节点。根据本实施例的另一使用场景参照上述表（3），例如核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——既往病史中，是既往病史的核保通过率均为不通过，非既往病史均为不通过，因此可以将既往病史作为该核保决策树中年龄——低龄——既往病史这一分支的叶子节点。

根据本实施例的另一使用场景例如通过上述递归算法划分的核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——行业风险中——既往病史否下的子属性仅包括赔付率时，该子属性赔付率可作为核保决策树的叶子节点。

根据本实施例的又一使用场景通过上述递归算法划分的核保决策树的根节点至叶子节点的某一个分支的属性依次为年龄——低龄——行业风险中——既往病史否，其中既往病史的熵值增益如果小于预设的阈值，则可将既往病史作为该核保决策树的叶子节点，也可将既往病史这一叶子节点做剪枝操作，将该叶子节点的上一节点行业风险作为该核保决策树的叶子节点。

根据本实施例的一个示例，该核保决策树的创建装置 10 还包括：

显示模块，用于显示该核保决策树并在该核保决策树的叶子节点中显示对应属性的核保结果。

根据本实施例的一个示例，该显示模块具体用于显示子属性核保通过的数量和核保不通过的数量，还可以是该叶子节点对应子属性的核保的通过率

和/或核保的不通过率。

其中上述第一叶子节点确定单元、第二叶子节点确定单元及第三叶子节点确定单元中的“第一”、“第二”和“第三”的意义仅在于将不同的叶子节点确定单元加以区分，并不用于限定哪个叶子节点确定单元的优先级更高或者其它的限定意义。

上述核保决策树的创建装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于终端的存储器中，也可以以软件形式存储于终端的存储器中，以便于处理器调用执行以上各个模块对应的操作。该处理器可以为中央处理单元（CPU）、微处理器、单片机等。

上述核保决策树的创建装置可以实现为一种计算机可读指令的形式，在如图 6 所示的计算机设备上运行。

在一个实施例中，提供了一种计算机设备，计算机设备的内部结构可对应于如图 6 的结构，即该计算机设备既可以是服务器也可以是终端，包括存储器和一个或多个处理器，存储器中存储有计算机可读指令，计算机可读指令被处理器执行时，使得一个或多个处理器执行以下步骤：获取包括不同样本属性的样本训练集；根据样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；将熵值增益最高的属性作为核保决策树的当前节点，并将熵值增益最高的属性对应的子属性划分为当前节点的下一节点；从样本训练集中抽取划分的子属性的样本训练子集；及将样本训练子集确定为样本训练集，对子属性递归计算熵值增益及划分操作，直到划分的下一节点的子属性满足成为核保决策树的叶子节点的预设条件。

在一个实施例中，处理器所执行的根据样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益的步骤，包括：

通过以下公式计算熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i});$$

其中，M 表示样本训练集中总核保通过率， A_i 表示对应属性 A 的子属性 i 的数量占样本训练集中的总数量的比例， B_i 表示子属性 i 以属性 A 的数量为

基数的核保通过率， n 表示对应属性 A 的子属性的个数， G_A 表示计算的属性 A 的熵值增益。

在一个实施例中，处理器所执行的判断子属性是否满足成为核保决策树的叶子节点的预设条件的步骤，包括：当划分的子属性只有一个时，将子属性确定为核保决策树的叶子节点；或当划分的子属性的核保结果均为通过或者不通过时，将子属性确定为核保决策树的叶子节点；或当子属性的熵值增益小于预设的阈值时，将子属性确定为核保决策树的叶子节点。

在一个实施例中，在处理器所执行的将样本训练子集确定为样本训练集，对子属性递归计算熵值增益及划分操作，直到划分的下一节点的子属性满足成为核保决策树的叶子节点的预设条件的步骤之后，处理器执行计算机可读指令还用于实现以下步骤：显示核保决策树并在核保决策树的叶子节点中显示对应属性的核保结果。

在一个实施例中，属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率。

图6为根据本申请的一个实施例的计算机设备的内部结构示意图，该计算机设备可以为服务器。参照图6，该计算机设备包括通过系统总线连接的处理器、非易失性存储介质、内存储器、输入装置和显示屏。其中，该计算机设备的非易失性存储介质可存储操作系统和计算机可读指令，该计算机可读指令被执行时，可使得处理器执行本申请各实施例的一种核保决策树的创建方法，该方法的具体实现过程可参考图1至4各实施例的具体内容，在此不再赘述。该计算机设备的处理器用于提供计算和控制能力，支撑整个计算机设备的运行。该内存储器中可储存有计算机可读指令，该计算机可读指令被处理器执行时，可使得处理器执行一种核保决策树的创建方法。计算机设备的输入装置用于各个参数的输入，计算机设备的显示屏用于进行显示。本领域技术人员可以理解，图6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的限定，具体的计算机设备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。

在一个实施例中，提供了一个或多个存储有计算机可读指令的非易失性

可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：获取包括不同样本属性的样本训练集；根据样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；将熵值增益最高的属性作为核保决策树的当前节点，并将熵值增益最高的属性对应的子属性划分为当前节点的下一节点；从样本训练集中抽取划分的子属性的样本训练子集；及将样本训练子集确定为样本训练集，对子属性递归计算熵值增益及划分操作，直到划分的下一节点的子属性满足成为核保决策树的叶子节点的预设条件。

在一个实施例中，处理器所执行的根据样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益的步骤，包括：通过以下公式计算熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2 (1 - M)) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2 B_i - (1 - B_i) \times \log_2 (1 - B_i));$$

其中，M表示样本训练集中总核保通过率， A_i 表示对应属性A的子属性i的数量占样本训练集中的总数量的比例， B_i 表示子属性i以属性A的数量为基数的核保通过率，n表示对应属性A的子属性的个数， G_A 表示计算的属性A的熵值增益。

在一个实施例中，处理器所执行的判断子属性是否满足成为核保决策树的叶子节点的预设条件的步骤，包括：当划分的子属性只有一个时，将子属性确定为核保决策树的叶子节点；或当划分的子属性的核保结果均为通过或者不通过时，将子属性确定为核保决策树的叶子节点；或当子属性的熵值增益小于预设的阈值时，将子属性确定为核保决策树的叶子节点。

在一个实施例中，在处理器所执行的将样本训练子集确定为样本训练集，对子属性递归计算熵值增益及划分操作，直到划分的下一节点的子属性满足成为核保决策树的叶子节点的预设条件的步骤之后，处理器执行计算机可读指令还用于实现以下步骤：显示核保决策树并在核保决策树的叶子节点中显示对应属性的核保结果。

在一个实施例中，属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率。

根据本实施例的一个示例，上述实施例方法中的全部或部分流程，可以通过计算机可读指令来指令相关的硬件来完成，所述程序可存储于一计算机可读取存储介质中，如本申请实施例中，该程序可存储于计算机系统的存储介质中，并被该计算机系统中的至少一个处理器执行，以实现包括如上述各方法的实施例的流程。该存储介质包括但不限于磁碟、优盘、光盘、只读存储记忆体（Read-Only Memory，ROM）等。

本实施例通过计算样本训练集中各个属性的熵值增益，并将熵值增益最大的属性作为核保决策树的当前节点，再通过递归的方式划分出该核保决策树的中间节点属性以及叶子节点属性，以创建基于各个属性的核保决策树，使得核保人可以依据该核保决策树对处于该核保决策树例如根节点等上层的属性进行着重审核，并使得核保人可以依据核保决策树中叶子节点中显示的核保结果直接作出核保决定，以给用户提供核保的数据依据，提高核保的准确性和核保的效率。

以上所述实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种核保决策树的创建方法，包括：

获取包括不同样本属性的样本训练集；

根据所述样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；

5 将熵值增益最高的属性作为核保决策树的当前节点，并将所述熵值增益最高的属性对应的子属性划分为所述当前节点的下一节点；

从所述样本训练集中抽取划分的所述子属性的样本训练子集；及

10 将所述样本训练子集确定为所述样本训练集，对所述子属性递归计算所述熵值增益及划分操作，直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件。

2、根据权利要求1所述的方法，其特征在于，所述根据所述样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益，包括：

通过以下公式计算所述熵值增益：

15
$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i});$$

其中，M表示样本训练集中总核保通过率， A_i 表示对应属性A的子属性i的数量占样本训练集中的总数量的比例， B_i 表示所述子属性i以所述属性A的数量为基数的核保通过率，n表示对应属性A的子属性的个数， G_A 表示计算的所述属性A的熵值增益。

20 3、根据权利要求1所述的方法，其特征在于，判断所述子属性是否满足成为所述核保决策树的叶子节点的预设条件的步骤，包括：

当划分的子属性只有一个时，将所述子属性确定为所述核保决策树的叶子节点；或

25 当划分的子属性的核保结果均为通过或者不通过时，将所述子属性确定为所述核保决策树的叶子节点；或

当所述子属性的熵值增益小于预设的阈值时，将所述子属性确定为所述核保决策树的叶子节点。

4、根据权利要求1所述的方法，其特征在于，在所述将所述样本训练子集确定为所述样本训练集，对所述子属性递归计算所述熵值增益及划分操作，直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件之后，还包括：

5 显示所述核保决策树并在所述核保决策树的叶子节点中显示对应属性的核保结果。

5、根据权利要求1至4任一项所述的方法，其特征在于，所述属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率。

6、一种核保决策树的创建装置，其特征在于，所述装置包括：

10 样本获取模块，用于获取包括不同样本属性的样本训练集；

熵值增益计算模块，用于根据所述样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；

节点划分模块，用于将熵值增益最高的属性作为核保决策树的当前节点，并将所述熵值增益最高的属性对应的子属性划分为所述当前节点的下一节点；

15 子集抽取模块，用于从所述样本训练集中抽取划分的所述子属性的样本训练子集；及

递归模块，用于将所述样本训练子集确定为所述样本训练集，对所述子属性递归计算所述熵值增益及划分操作，直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件。

7、根据权利要求6所述的装置，其特征在于，所述熵值增益计算模块通过以下公式计算所述熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i});$$

其中，M表示样本训练集中总核保通过率， A_i 表示对应属性A的子属性i的数量占样本训练集中的总数量的比例， B_i 表示所述子属性i以所述属性A的数量为基数的核保通过率，n表示对应属性A的子属性的个数， G_A 表示计算的所述属性A的熵值增益。

8、根据权利要求6或7所述的装置，其特征在于，所述递归模块还包括：

第一叶子节点确定单元，用于当划分的子属性只有一个时，将所述子属性确定为所述核保决策树的叶子节点；或

第二叶子节点确定单元，用于当划分的子属性的核保结果均为通过或者不通过时，将所述子属性确定为所述核保决策树的叶子节点；或

5 第三叶子节点确定单元，用于当所述子属性的熵值增益小于预设的阈值时，将所述子属性确定为所述核保决策树的叶子节点。

9、根据权利要求 6 所述的装置，其特征在于，所述装置还包括：

显示模块，用于显示所述核保决策树并在所述核保决策树的叶子节点中显示对应属性的核保结果。

10 10、根据权利要求 6 至 9 任一项所述的装置，其特征在于，所述属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率。

11、一种计算机设备，包括存储器和一个或多个处理器，所述存储器中存储有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：

15 获取包括不同样本属性的样本训练集；

根据所述样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益；

将熵值增益最高的属性作为核保决策树的当前节点，并将所述熵值增益最高的属性对应的子属性划分为所述当前节点的下一节点；

20 从所述样本训练集中抽取划分的所述子属性的样本训练子集；及

将所述样本训练子集确定为所述样本训练集，对所述子属性递归计算所述熵值增益及划分操作，直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件。

12、根据权利要求 11 所述的计算机设备，其特征在于，所述处理器所执行的根据所述样本训练集中各属性的样本的核保结果，计算各属性中表示影响核保结果的熵值增益的步骤，包括：

通过以下公式计算所述熵值增益：

$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i});$$

其中, M 表示样本训练集中总核保通过率, A_i 表示对应属性 A 的子属性 i 的数量占样本训练集中的总数量的比例, B_i 表示所述子属性 i 以所述属性 A 的数量为基数的核保通过率, n 表示对应属性 A 的子属性的个数, G_A 表示计算的所述属性 A 的熵值增益。

5 13、根据权利要求 11 所述的计算机设备, 其特征在于, 所述处理器所执行的判断所述子属性是否满足成为所述核保决策树的叶子节点的预设条件的步骤, 包括:

当划分的子属性只有一个时, 将所述子属性确定为所述核保决策树的叶子节点; 或

10 当划分的子属性的核保结果均为通过或者不通过时, 将所述子属性确定为所述核保决策树的叶子节点; 或

当所述子属性的熵值增益小于预设的阈值时, 将所述子属性确定为所述核保决策树的叶子节点。

14、根据权利要求 11 所述的计算机设备, 其特征在于, 在所述处理器所
15 执行的将所述样本训练子集确定为所述样本训练集, 对所述子属性递归计算所述熵值增益及划分操作, 直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件的步骤之后, 所述处理器执行所述计算机可读指令还用于实现以下步骤:

20 显示所述核保决策树并在所述核保决策树的叶子节点中显示对应属性的核保结果。

15、根据权利要求 11 至 14 任一项所述的计算机设备, 其特征在于, 所述属性包括以下情况中的至少两种: 年龄、行业风险、既往病史及赔付率。

16、一个或多个存储有计算机可读指令的非易失性可读存储介质, 所述计算机可读指令被一个或多个处理器执行时, 使得所述一个或多个处理器执
25 行以下步骤:

获取包括不同样本属性的样本训练集;

根据所述样本训练集中各属性的样本的核保结果, 计算各属性中表示影响核保结果的熵值增益;

将熵值增益最高的属性作为核保决策树的当前节点, 并将所述熵值增益

最高的属性对应的子属性划分为所述当前节点的下一节点;

从所述样本训练集中抽取划分的所述子属性的样本训练子集; 及

将所述样本训练子集确定为所述样本训练集, 对所述子属性递归计算所述熵值增益及划分操作, 直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件。

17、根据权利要求 16 所述的存储介质, 其特征在于, 所述处理器所执行的根据所述样本训练集中各属性的样本的核保结果, 计算各属性中表示影响核保结果的熵值增益的步骤, 包括:

通过以下公式计算所述熵值增益:

$$G_A = -(M \times \log_2 M + (1 - M) \log_2^{(1-M)}) - \sum_{i=1}^{i=n} A_i \times (-B_i \times \log_2^{B_i} - (1 - B_i) \times \log_2^{1-B_i});$$

其中, M 表示样本训练集中总核保通过率, A_i 表示对应属性 A 的子属性 i 的数量占样本训练集中的总数量的比例, B_i 表示所述子属性 i 以所述属性 A 的数量为基数的核保通过率, n 表示对应属性 A 的子属性的个数, G_A 表示计算的所述属性 A 的熵值增益。

18、根据权利要求 16 所述的存储介质, 其特征在于, 所述处理器所执行的判断所述子属性是否满足成为所述核保决策树的叶子节点的预设条件的步骤, 包括:

当划分的子属性只有一个时, 将所述子属性确定为所述核保决策树的叶子节点; 或

当划分的子属性的核保结果均为通过或者不通过时, 将所述子属性确定为所述核保决策树的叶子节点; 或

当所述子属性的熵值增益小于预设的阈值时, 将所述子属性确定为所述核保决策树的叶子节点。

19、根据权利要求 16 所述的存储介质, 其特征在于, 在所述处理器所执行的将所述样本训练子集确定为所述样本训练集, 对所述子属性递归计算所述熵值增益及划分操作, 直到划分的下一节点的子属性满足成为所述核保决策树的叶子节点的预设条件的步骤之后, 所述处理器执行所述计算机可读指令还用于实现以下步骤:

显示所述核保决策树并在所述核保决策树的叶子节点中显示对应属性的核保结果。

20、根据权利要求 16 至 19 任一项所述的存储介质，其特征在于，所述属性包括以下情况中的至少两种：年龄、行业风险、既往病史及赔付率。

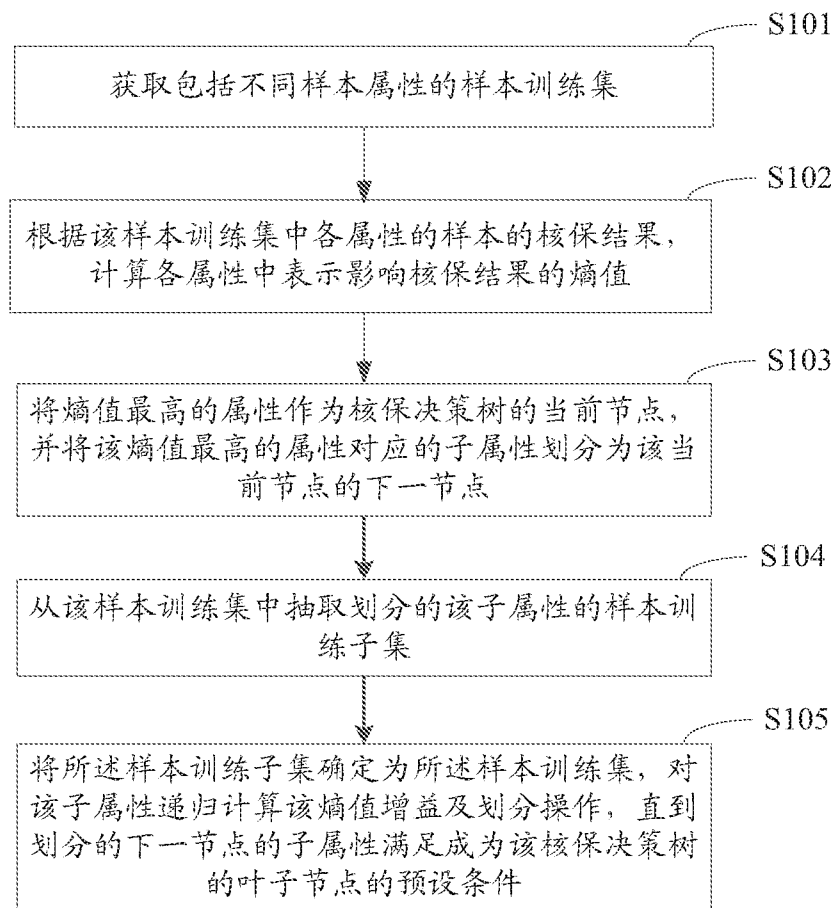


图 1

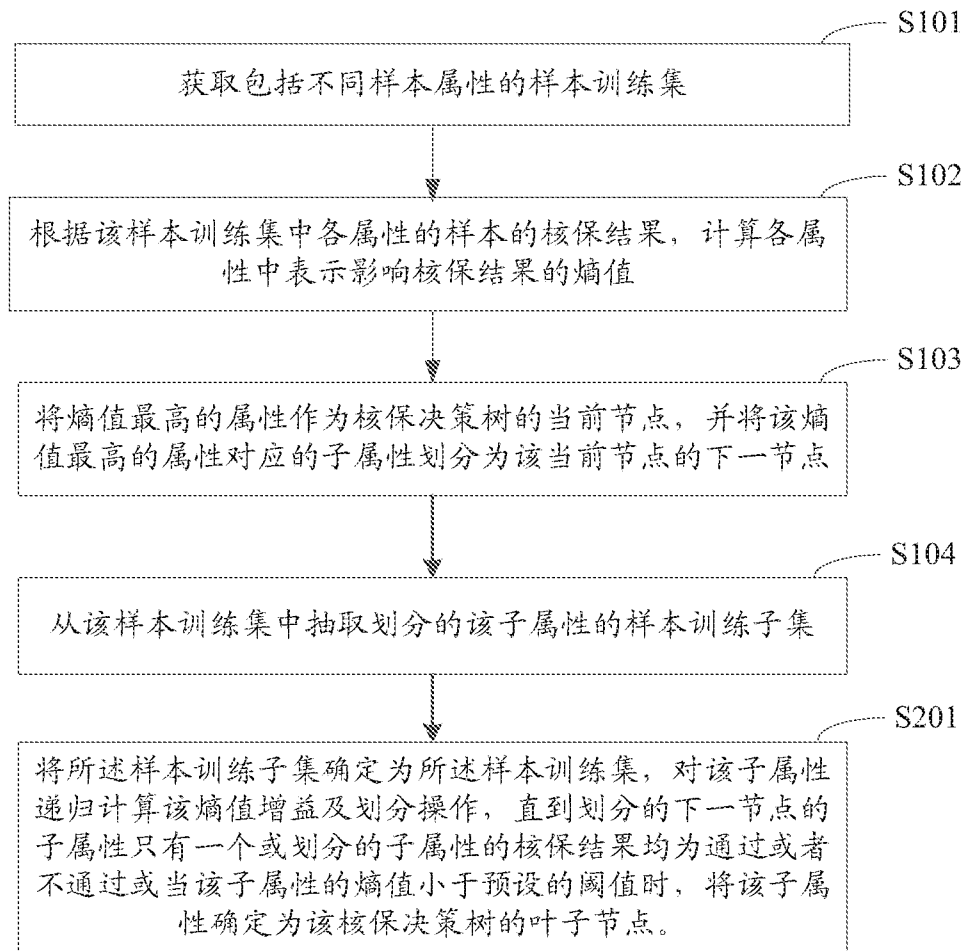


图 2

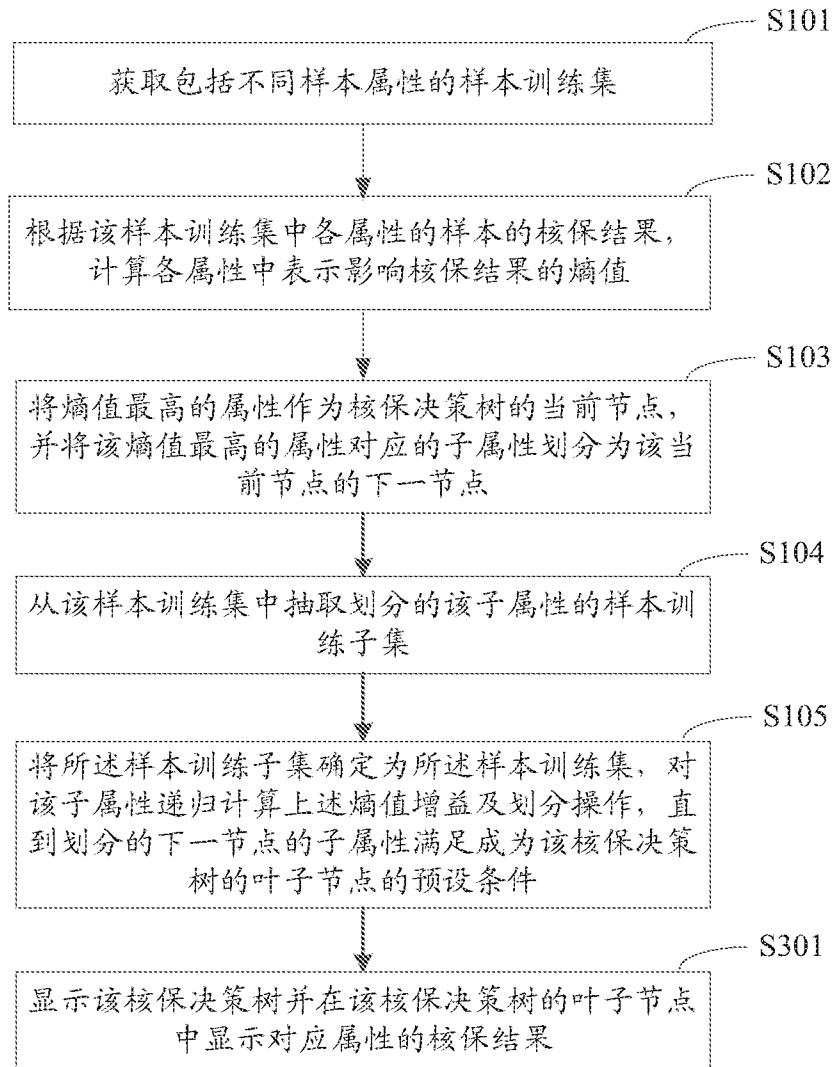


图 3

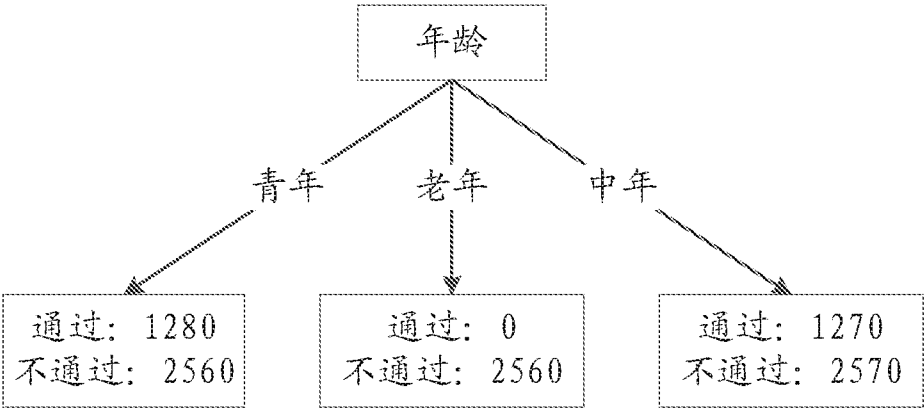


图 4

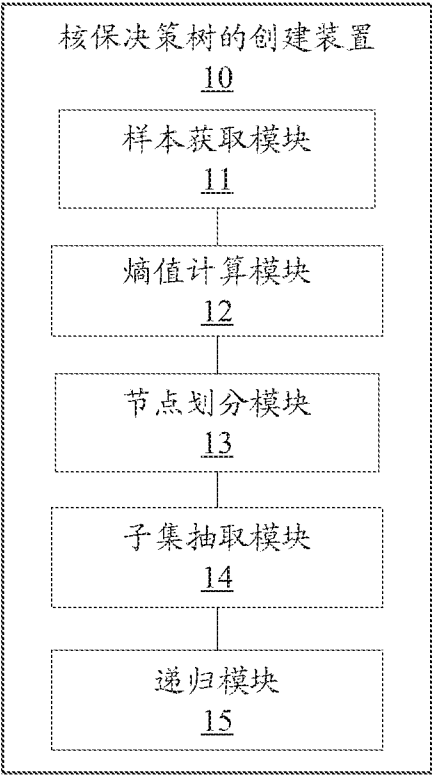


图 5

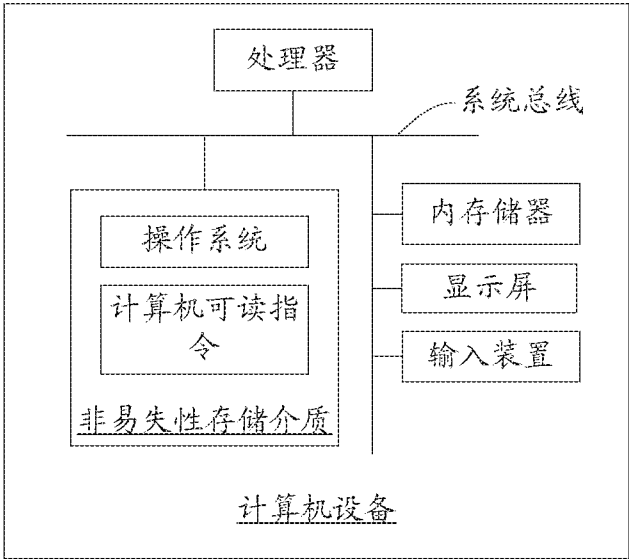


图 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/104598

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 40/08 (2012.01) i; G06F 17/30 (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)

G06Q; G06F

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)

CNKI, CNPAT, WPI, EPODOC, IEEE: 保险, 决策树, 属性, 样本, 训练样本集, 增益, 节点, 结点, insurance, decision tree, attribute, sample, training sample set, gain, node

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 104765839 A (XIANGTAN UNIVERSITY), 08 July 2015 (08.07.2015), description, paragraphs [0009]-[0015]	1-20
A	CN 104778250 A (NANJING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS), 15 July 2015 (15.07.2015), entire document	1-20
A	CN 103996287 A (JIANGSU UNIVERSITY), 20 August 2014 (20.08.2014), entire document	1-20
A	CN 106600423 A (SHENZHEN CLOUDYBI INFORMATION TECHNOLOGY CO., LTD.), 26 April 2017 (26.04.2017), entire document	1-20
A	US 2003229630 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA), 11 December 2003 (11.12.2003), entire document	1-20

☐ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"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
"A" document defining the general state of the art which is not considered to be of particular relevance	"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
"E" earlier application or patent but published on or after the international filing date	"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
"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)	"&" document member of the same patent family
"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	

Date of the actual completion of the international search 09 April 2018	Date of mailing of the international search report 18 April 2018
Name and mailing address of the ISA State Intellectual Property Office of the P. R. China No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088, China Facsimile No. (86-10) 62019451	Authorized officer DONG, Xianbin Telephone No 86-10-53961407

International application No.
PCT/CN2017/104598

Form PCT/ISA/210 (patent family annex) (July 2009)

国际检索报告

国际申请号

PCT/CN2017/104598

A. 主题的分类

G06Q 40/08(2012.01)i; G06F 17/30(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)

G06Q; G06F

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))

CNKI, CNPAT, WPI, EPODOC, IEEE: 保险, 决策树, 属性, 样本, 训练样本集, 增益, 节点, 结点, insurance, decision tree, attribute, sample, training sample set, gain, node

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 104765839 A (湘潭大学) 2015年 7月 8日 (2015 - 07 - 08) 说明书第[0009]-[0015]段	1-20
A	CN 104778250 A (南京邮电大学) 2015年 7月 15日 (2015 - 07 - 15) 全文	1-20
A	CN 103996287 A (江苏大学) 2014年 8月 20日 (2014 - 08 - 20) 全文	1-20
A	CN 106600423 A (云数信息科技深圳有限公司) 2017年 4月 26日 (2017 - 04 - 26) 全文	1-20
A	US 2003229630 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 2003年 12月 11日 (2003 - 12 - 11) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 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

2018年 4月 9日

国际检索报告邮寄日期

2018年 4月 18日

ISA/CN的名称和邮寄地址

中华人民共和国国家知识产权局(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

董显彬

电话号码 86-10-53961407

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/104598

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	104765839	A	2015年 7月 8日	无	
CN	104778250	A	2015年 7月 15日	无	
CN	103996287	A	2014年 8月 20日	无	
CN	106600423	A	2017年 4月 26日	无	
US	2003229630	A1	2003年 12月 11日	无	