

(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 11 月 21 日 (21.11.2019)



(10) 国际公布号
WO 2019/218751 A1

(51) 国际专利分类号:
G06Q 10/06 (2012.01)

(21) 国际申请号: PCT/CN2019/076524

(22) 国际申请日: 2019 年 2 月 28 日 (28.02.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810469782.1 2018 年 5 月 16 日 (16.05.2018) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 吴龙凤 (WU, Longfeng); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集

团法务部, Zhejiang 311121 (CN)。 陈葵 (CHEN, Yan); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 石秋慧 (SHI, Qiuhui); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 张泰玮 (ZHANG, Taiwei); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。 陈诗奕 (CHEN, Shiyi); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京博思佳知识产权代理有限公司 (BEIJING BESTIPR INTELLECTUAL PROPERTY LAW CORPORATION); 中国北京市海淀区上地三街 9 号嘉华大厦 B 座 409, Beijing 100085 (CN)。

(54) Title: PROCESSING METHOD, APPARATUS AND DEVICE FOR RISK PREDICTION OF INSURANCE SERVICE

(54) 发明名称: 一种保险业务风险预测的处理方法、装置及处理设备

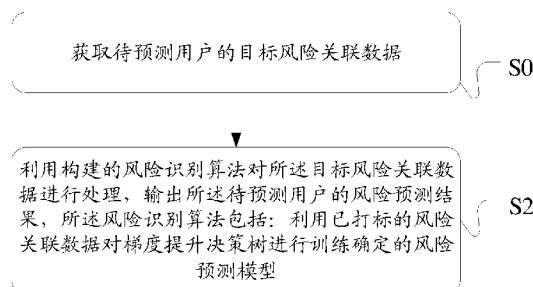


图 1

S0 ACQUIRE TARGET RISK-ASSOCIATED DATA OF A USER TO BE PREDICTED
S2 PROCESS THE TARGET RISK-ASSOCIATED DATA USING A CONSTRUCTED RISK IDENTIFICATION ALGORITHM, AND OUTPUT A RISK PREDICTION RESULT OF THE USER TO BE PREDICTED. THE RISK IDENTIFICATION ALGORITHM COMPRISING: A RISK PREDICTION MODEL DETERMINED BY TRAINING USING RISK-ASSOCIATED MARKED DATA ON THE BASIS OF A GRADIENT BOOSTING DECISION TREE

(57) Abstract: A processing method, apparatus and device for risk prediction of an insurance service. A gradient boosting decision tree can be used in advance to construct a risk prediction model, the risk prediction model being able to be trained using risk-associated marked data associated with the insurance service. When the training of the risk prediction model satisfies a prediction requirement, the risk prediction model can be used as online risk prediction, performs insurance service risk prediction on a user to be predicted, and outputs a prediction result. Multidimensional nonlinear variables in the insurance service can be reasonably and effectively used, and the risk prediction model based on the nonlinear relationship of the gradient boosting decision tree can be well compatible with linear

(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))。

and nonlinear variables. Compared to the traditional linear model, the accuracy of the prediction result is significantly improved, the deficiency of the traditional linear model is effectively compensated for, and the service experience of the insurance service is improved.

(57) 摘要: 一种保险业务风险预测的处理方法、装置及处理设备。可以预先使用梯度提升决策树来构建风险预测模型, 该风险预测模型可以使用已打标的与保险业务相关联的风险关联数据进行训练。当风险预测模型训练达到预测要求时可以作为线上风险预测使用, 对待预测用户进行保险业务风险预测, 输出预测结果。可以合理有效地应用保险业务中多维度的非线性变量, 基于梯度提升决策树的非线性关系的风险预测模型可以很好地兼容线性和非线性变量, 相对于传统的线性模型, 预测结果的准确性有着显著的提升, 有效弥补传统线性模型的不足, 提高保险业务服务体验。

一种保险业务风险预测的处理方法、装置及处理设备

技术领域

[01] 本说明书实施例方案属于保险业务风险预测的计算机数据处理技术领域，尤其涉及一种保险业务风险预测的处理方法、装置及处理设备。

5 背景技术

[02] 机动车辆保险即汽车保险（或简称车险），是指对机动车辆由于自然灾害或意外事故所造成的人身伤亡或财产损失负赔偿责任的一种商业保险。随着经济的发展，机动车辆的数量不断增加，当前，车险已成为中国财产保险业务中最大的险种之一。

[03] 用户在进行车辆投保时，保险公司通常会对用户进行风险评估，风险评估的结果会直接影响到用户投保金额、优惠待遇等。通过对用户的风险评估，保险公司可以更加准确、合理的进行保险业务的处理，有效规避或减少业务风险。目前，在车险风险预测领域，基于广义线性模型（generalized linear model, GLM）的风险预测已成为业内的主流风险预测技术体系。广义线性模型主要处理的为线性相关的数据对象，如上网时长降低 1 个百分点，年龄增大 1 岁，可以基于网龄数据与年龄数据的线性关系实现 GLM 的建模。

[04] 但随着车险业务的不断增加，互联网数据已呈现多种类、海量数据增长，传统的 GLM 模型体系已越来越受到限制，例如如果“年龄”不是单纯随上网时长变化而变化，而是同时与人群的购物以及习惯等方面相关，不同消费习惯随自身变化改变年龄分布呈非线性影响的形式。GLM 模型可以通过分箱将非线性变量进行分段汇总，但是会损失很多变量的精准性，难以适应当前大数据、多维度的风险预测要求。因此，业内亟需一种可以在多维度数据中更加有效和高效的进行车险业务风险预测的处理方式。

发明内容

[05] 本说明书实施例目的在于提供一种保险业务风险预测的处理方法、装置及处理设备，可以通过在保险业务风险预测中引入梯度提升决策树，实现兼容非线性关系的保险业务数据的风险预测，有效提高保险业务风险预测的准确性。

[06] 本说明书实施例提供的一种保险业务风险预测的处理方法、装置及处理设备是包

括以下方式实现的:

[07] 一种保险业务风险预测的处理方法, 所述方法包括:

[08] 获取待预测用户的目标风险关联数据;

[09] 利用构建的风险预测模型对所述目标风险关联数据进行处理, 输出所述待预测用户的风险预测结果, 所述风险预测模型法包括: 利用已打标的风险关联数据对梯度提升决策树进行训练确定的预测模型。

[10] 一种保险业务风险预测处理装置, 包括:

[11] 预测数据获取模块, 用于获取待预测用户的目标风险关联数据;

[12] 风险预测模块, 用于利用构建的风险预测模型对所述目标风险关联数据进行处理, 输出所述待预测用户的风险预测结果, 所述风险预测模型法包括: 利用已打标的风险关联数据对梯度提升决策树进行训练确定的预测模型。

[13] 一种保险业务风险预测处理设备, 包括处理器以及用于存储处理器可执行指令的存储器, 所述处理器执行所述指令时实现:

[14] 获取待预测用户的目标风险关联数据;

[15] 利用构建的风险预测模型对所述目标风险关联数据进行处理, 输出所述待预测用户的风险预测结果, 所述风险预测模型法包括: 利用已打标的风险关联数据对梯度提升决策树进行训练确定的预测模型。

[16] 本说明书实施例提供的一种保险业务风险预测的处理方法、装置及处理设备, 可以预先使用梯度提升决策树来构建风险预测模型, 该风险预测模型可以使用已打标的与保险业务相关联的风险关联数据进行训练。当风险预测模型训练达到预测要求时可以作为线上风险预测使用, 对待预测用户进行保险业务风险预测, 输出预测结果。利用本说明书实施例提供的方法可以合理有效地应用保险业务中多维度的非线性变量, 基于梯度提升决策树的非线性关系的风险预测模型可以很好地兼容线性和非线性变量, 相对于传统的线性模型, 预测结果的准确性有着显著的提升, 有效弥补传统线性模型的不足, 提高保险业务服务体验。

附图说明

[17] 为了更清楚地说明本说明书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说明书中记载的一些实施例，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。

[18] 图 1 是本说明书提供的一种保险业务风险预测处理方法实施例的流程示意图；

5 [19] 图 2 是本说明书提供的所述方法中一种构建风险预测模型的处理过程示意图；

[20] 图 3a 及图 3b 是本说明书提供的所述方法中决策树学习训练的处理过程示意图；

[21] 图 4 是本说明书提供的一种应用保险业务风险预测处理方法的服务器的硬件结构框图。

[22] 图 5 是本说明书提供的一种保险业务风险预测处理装置的模块结构示意图。

10 具体实施方式

[23] 为了使本技术领域的人员更好地理解本说明书中的技术方案，下面将结合本说明书实施例中的附图，对本说明书实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本说明书中的一部分实施例，而不是全部的实施例。基于本说明书中的一个或多个实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都应当属于本说明书实施例保护的范围。

15 [24] 随着计算机互联网技术的发展，数据量飞速增长。保险业务风险预测时数据特征的分类也越来越多维度、细致化。很多变量对筛选分类的影响是以非线性存在的，例如上网时长和年龄呈现相关性，但是该相关性可以是多种多样的。例如可以是简单的线性关系，例如上网时长降低 1 个百分点，年龄增大 1 岁；也可以是比较复杂的关系，例如
20 指数关系，上网时长降低 4 个百分点，年龄增大 2 岁，此时可以通过一定数学变化转化为线性的均可以用广义线性模型解决。现实生活中，除了一些基本线性关系的变量外，还存在大量非线性变量。例如在预测年龄时，如果“年龄”不是单纯随上网时长变化而变化，而是同时与人群的购物以及习惯等方面相关，不同消费习惯随自身变化改变年龄分布呈非线性影响的形式。因为预测“用户年龄”是目的之一，若一些 GLM 模型无法
25 识别非线性关系将会大幅降低模型的预测性能。现有解决的方式中，可以通过分箱将变量进行分段汇总，但是会损失很多变量的精准性，降低预测结果。本说明书实施例提供的区别于现有常规实施方案的另一种保险业务中风险预测的实现方法，引入了 GBDT(Gradient Boosting Decision Tree, 梯度提升决策树)，可以在风险预测中合理有效

地应用非线性变量构建风险预测模型，该模型能很好地兼容线性和非线性变量，相对于传统的线性模型，预测结果的准确性有着显著的提升。

[25] GBDT(Gradient Boosting Decision Tree)是一种迭代的决策树算法，该算法由多棵决策树组成，所有树的结论累加起来做最终答案。GBDT 中的树都是回归树，可以用来做回归预测。本说明书提供的保险业务风险预测的处理方法中，可以预先使用已打标的风险关联数据构建决策树模型，通过回归的机器学习（分布迭代）对决策树中的参数逐步调整优化。当模型预测结果符合保险业务风险预测的精度要求时，可以线上使用来预测待预测用户的风险数值或赔付率等。

[26] 下面以一个具体的车险业务风险预测处理的应用场景为例对本说明书实施方案进行说明。具体的，图 1 是本说明书提供的所述保险业务风险预测的处理方法实施例的流程示意图。虽然本说明书提供了如下述实施例或附图所示的方法操作步骤或装置结构，但基于常规或者无需创造性的劳动在所述方法或装置中可以包括更多或者部分合并后更少的操作步骤或模块单元。在逻辑性上不存在必要因果关系的步骤或结构中，这些步骤的执行顺序或装置的模块结构不限于本说明书实施例或附图所示的执行顺序或模块结构。所述的方法或模块结构的在实际中的装置、服务器或终端产品应用时，可以按照实施例或者附图所示的方法或模块结构进行顺序执行或者并行执行（例如并行处理器或者多线程处理的环境、甚至包括分布式处理、服务器集群的实施环境）。

[27] 当然，下述车险业务风险预测的实施例的描述并不对基于本说明书的其他可扩展到的技术方案构成限制。例如其他的实施场景中，本说明书提供的实施方案同样可以应用到基金风险预测、医疗保险风险预测等的实施场景中，其他实施场景中的应用参照本说明书车险业务的实施例描述，不再进行替代性的重复描述。具体的一种实施例如图 1 所示，本说明书提供的一种保险业务风险预测处理方法可以包括：

[28] S0：获取待预测用户的目标风险关联数据；

[29] S2：利用构建的风险识别算法对所述目标风险关联数据进行处理，输出所述待预测用户的风险预测结果，所述风险识别算法包括：利用已打标的风险关联数据对梯度提升决策树进行训练确定的风险预测模型。

[30] 本说明书的一个或多个实施例中，可以预先构建基于 GBDT 的风险预测模型。具体的 GBDT 模型的训练和构建可以根据实际业务场景需求和数据进行相应的模型结构和参数设置，如可以以单棵树进行单独训练，训练的残差作为另一个树的输入继续进行

训练；或者多棵树多级连接进行训练，训练残差再作为另一个多级连接的数的输入。当然，其他的实施例中还可以应用基于 GBDT 算法进行一些变形、变换或改进的处理算法实现的非线性关系的保险业务数据的风险预测处理，本说明书不再对 GBDT 模型构建的实现过程逐一赘述。

- 5 [31] 本实施例中可以预先根据历史车险业务保单数据采集确定风险预测模型的训练数据，根据风险划分或者设置要求对训练数据进行打标。在本实施例保险业务风险预测的实施场景中，所述的训练数据可以称为风险关联数据，这些风险关联数据通常与保险业务相关联，用于对风险预测模型的样本训练。例如风险关联可以为包括多个维度的用户特征数据，一个用户相关联的用户特征数据为一组训练数据，每组风险关联数据可以打
- 10 标设置相应的风险分值。具体的，本说明书所述方法的一个实施例中，所述风险关联数据可以包括与至少一个类别的用户特征数据，所述用户特征数据包括与保险业务相关联的非线性关系的数据信息。例如一个示例中，用户 A 的风险关联数据可以包括（A1，A2，A3...，A9）9 个维度的用户特征数据。可以根据车险预测的需求相应的选取不同维度的用户特征数据，例如上述示例的 9 个维度可以包括年龄，性别，职业，年收入，
- 15 历史出险次数，月均消费，征信等级，婚姻状况，负债资产。或者可以预先采集获取 10 个或 10 个维度以上的用户特征数据，在确定风险关联数据时从多个维度的用户特征数据中选取需要进行模型训练的用户特征数据。例如，具体的风险关联数据可以包括如下表 1 所示：

表 1 模型训练的风险关联数据示意表

用户	年龄	性别	职业	年收入 (万)	历史出 险次数	月均消 费(千)	征信等 级	婚姻 状况	负债资 产(万)
A	36	男	律师	30	0	15	优	已婚	10
B	34	女	科员	25	0	5	优	未婚	0
C	26	女	学生	8	0	3.7	优	未婚	0
D	24	男	学生	6	1	2.5	良	未婚	2
...	...	...	...	...	...	...	...	...	...

- 20 [32] 当然，其他的实施例中，所述的风险关联数据还可以包括按照预定规则生成的人工数据，例如作业人员可以根据预期的风险可能包括的情况自定义设置进行模型训练的风险关联数据。或者，在设置的数据生成规则后由计算机自动生成所需的风险关联数据。这里的生成的人工数据更加符合预期的风险预测情况，而历史车险案件数据则更接近真

实的风险情况，一些实施应用场景中，可以使用其中的一种或者，同时结合人工数据和历史车险案件数据进行风险预测模型的训练，以提高预测结果的准确性。

[33] 获取的风险关联数据可以作为训练数据在 GBDT 模型中进行训练，经过学习训练后风险预测模型中决策树分枝时的决策特征的阈值（可以是全部的阈值，或者部分的阈值）能满足模型最终输出结果的精度要求（通常还可以要求连续稳定的输出符合精度要求）。本说明书实施例中使用的 GBDT 是一种迭代的决策树算法，主要可以分为决策树（Regression Decision Tree，DT）和梯度提升（Gradient boosting，GB）。决策树主要分为两类：分类树和回归树，分类树常用来解决分类问题，比如用户性别、网页是否是垃圾页面、用户是不是作弊等。而回归树一般用来预测真实数值，比如用户的年龄、用户点击的概率、网页相关程度等等。前者用于分类标签纸，后者用于预测实数值。这里要强调的是，回归树的结果加减是有意义的，如 10 岁+5 岁-3 岁=12 岁，后者则是没有办法累加或累加结果无意义，如男+男+女=到底是男是女。本说明书实施例可以利用回归树来预测车险的车险分值，如累加所有树的结果作为最终风险预测的结果，

[34] 回归树大致流程与分类树类似，区别在于，回归树的每一个节点都会得到一个预测值，以年龄为例，该预测值等于属于这个节点的所有人年龄的平均值。分枝时穷举每一个特征寻找最优切分变量和最优切分点，本实施例中衡量的准则不再是分类树中的基尼系数，而是平方误差最小化。也就是被预测错误的人数越多，平方误差就越大，通过最小化平方误差找到最可靠的分枝依据。分枝直到每个叶子节点上人的对游戏感兴趣都是唯一的或者达到预设的终止条件(如叶子个数上限)，若最终叶子节点上年龄不是唯一的，则以该节点上所有人的平均年龄作为该叶子节点的预测结果。

[35] 梯度提升（Gradient boosting）是一种用于回归、分类和排序任务的机器学习技术，属于 Boosting 算法族的一部分。Boosting 是一族可将弱学习器提升为强学习器的算法，属于集成学习（ensemble learning）的范畴。Boosting 方法基于这样一种思想：对于一个复杂任务来说，将多个专家的判断进行适当的综合所得出的判断，要比其中任何一个专家单独的判断要好。通俗地说，就是“三个臭皮匠顶个诸葛亮”的道理。梯度提升同其他 boosting 方法一样，通过集成（ensemble）多个弱学习器，通常是决策树，来构建最终的预测模型。boosting 方法通过分步迭代（stage-wise）的方式来构建模型，在迭代的每一步构建的弱学习器都是为了弥补已有模型的不足。

[36] 例如具体的一个处理过程中，训练的时候可以设定树的棵数，树的棵数达到指定数值后(如八十棵)可以停止训练了；或者残差很小（满足停止训练的条件）的时候，这

两个条件满足一个训练就可以停止训练。

[37] 若在第 N 棵残差不全为 0 或不满足停止条件时, 使用第 N 棵树的节点的残差结果替代相应的原值代入到第 N+1 棵树中进行学习;

[38] 直至第 N+K 颗数叶子节点的残差和预测值相等或小于阈值, 输出当前叶子节点对应的风险预测结果 (风险值或赔付率)。具体的可以将所有残差累加作为最终风险预测值。

[39] 图 2 是本说明书提供的一种构建风险预测模型的处理过程示意图。如图 2 所示, 本说明书提供的所述方法的另一个实施例中, 采用下述方式训练得到所述风险预测模型:

[40] S20: 确定决策树的总数量, 以及决策树在每次分枝时所使用的决策特征的阈值, 所述决策特征为所述用户特征数据的类别中的一种;

[41] S22: 当对一组风险关联数据进行训练时, 若训练的决策树的棵树达到预设数值或者决策树的残差满足停止训练条件, 则停止该组风险关联数据的训练, 所述预设数值小于等于所述总数量;

[42] S24: 通过风险关联数据的训练结果调整相应决策树的决策特征的阈值, 直至调整后的所述阈值满足风险预测模型的预测结果输出要求时, 确定所述风险预测模型。

[43] 本实施例中, 可以预先确定训练使用的决策树的数量, 通过梯度迭代逐渐优化确定一颗决策树进行分枝时的决策特征的阈值。如可以使用 80 棵决策树, 每棵树每一棵树学的是之前所有树结论和的残差。初始的数的阈值可以根据经验值进行设置。假如 A 的真实分值 (打标分值为 80 分), 但第一棵树的根据年龄的决策特征预测分值是 60 分, 差了 20 分, 残差为 20。那么在第二棵树 (决策特征为用户的职业) 里把 A 的分值设为 20 分去学习, 如果第二棵树真的能把 A 分到 20 分的叶子节点, 那累加两棵树的结论就是 A 的真实分值 (预测分值 60 分+残差 20 分); 如果第二棵树的结论是 18 分, 则 A 仍然存在 2 分残差, 第三棵树 (决策特征为年收入) 里 A 的年龄就变成 2 分, 继续学习。每一步的残差计算相当于变相地增大了分错事件的权重, 而已经分对的时间则都趋向于 0, 如, 根据年龄过大或过小, 则风险越大, 以及, 收入越高风险越小; 如果一个用户年龄过大为 60 岁, 但被分入了风险较小的分支 L1, 但风险较小的分组 L1 上的平均年龄在 20-40 岁之间, 则得到的残差值就会相应的增大, 该用户可以通过后续的收入、婚姻状况、驾龄等逐渐将其分向靠近实际风险的叶子节点。

[44] 若训练的决策树的数量达到预定数值后, 如从根节点一直到叶子节点的 10 棵树均

训练一遍，或者当前数的参数满足停止训练条件，如残差为 0 或者其他残差停止阈值，此时可以停止该组数据的训练。当每个阈值找最好的分割点，或者符合训练要求的分割点，则可以确定决策树的决策特征的阈值，直至调整后的所述阈值满足风险预测模型的预测结果输出要求时，确定所述风险预测模型。例如初始设置风险分值分为 60 和 80 的
5 阈值为年龄是否大于 20 岁。经过大量数据训练优化后，最终可以将从年龄维度进行风险评估这个决策特征调整年龄是否大于 24 岁，以符合多数情况下的真实预测结果。

[45] 下面以一个简单的年龄预测示例来说明本说明书实施例中如何利用 GBDT 实现决策树的训练。下述的示例在本说明书保险业务风险预测中，将年龄替换为车险风险分值或赔付率，将月购物消费和是否经常提问相应的替换为用户特征数据的类别即可，其中
10 的阈值相应的进行设置。具体的实施过程可以包括：

[46] 假设训练集（风险关联数据）只有 4 个人，A，B，C，D，他们的年龄分别是 14，16，24，26。其中 A、B 分别是高一和高三学生；C，D 分别是应届毕业生和工作两年的员工。如果是用一棵传统的回归决策树来训练。选取使用 GBDT 来做年龄预测，由于数据太少，我们限定叶子节点做多有两个，即每棵树都只有一个分枝，并且限定只学两
15 棵树，可以得到图 3a 及图 3b 所示的结果。在第一棵树分枝中，由于 A，B 年龄较为相近，C，D 年龄较为相近，他们被分为两拨，每拨用平均年龄作为预测值。此时计算残差（残差的意思就是： $A \text{ 的预测值} + A \text{ 的残差} = A \text{ 的实际值}$ ），所以 A 的残差就是 $16 - 15 = 1$ （注意，A 的预测值是指前面所有树累加的和，这里前面只有一棵树所以直接是 15，如果还有树则需要都累加起来作为 A 的预测值）。进而可以分别得到 A，B，C，
20 D 的残差分别为 -1，1，-1，1。然后拿残差替代 A，B，C，D 的原值，到第二棵树去学习，如果我们的预测值和它们的残差相等，则只需把第二棵树的结论累加到第一棵树上就能得到真实年龄了。第二棵树只有两个值 1 和 -1，直接分成两个节点。此时所有人的残差都是 0，即每个人都得到了真实的预测值。

[47] 经过两棵树的处理现在 A，B，C，D 的预测值都和真实年龄一致了：

25 [48] A：14 岁高一学生，购物较少，经常问学长问题；预测年龄 $A = 15 - 1 = 14$ ；

[49] B：16 岁高三学生；购物较少，经常被学弟问问题；预测年龄 $B = 15 + 1 = 16$ ；

[50] C：24 岁应届毕业生；购物较多，经常问师兄问题；预测年龄 $C = 25 - 1 = 24$ ；

[51] D：26 岁工作两年员工；购物较多，经常被师弟问问题；预测年龄 $D = 25 + 1 = 26$ 。

[52] 另一种实施例中，确定风险预测模型使用的决策树的数量时，可以基于所述用户

特征数据对应的类别的数量确定。例如选取了 80 个维度的用户特征数据, 每一个维度可以代表一棵树的决策特征, 这样可以使用 80 个决策树来构建非线性的风险预测模型。

一般情况, 可以设置一个维度对应多棵树, 具体的可以根据预测模型处理的数据量和应用场景的处理要求进行相应的设置。当然, 本说明书其他的实施例中, 具体决策树的总

5 数量可以根据采集是数据、树的分支数、树的上下级连接关系等进行确定。

[53] 如前所述, 本说明书提供的实施例不仅可以用于车险业务风险预测的实施场景中, 还可以应用到基金风险预测、医疗保险风险预测等的实施场景中。具体的在车险业务风险预测的应用场景中, 所述风险预测模型包括基于与车险业务相关联的风险关联数据进行训练得到的车险风险预测模型;

10 [54] S26: 所述风险预测结果包括待预测用户的赔付率、车险风险分值中的任意一种。

[55] 当然, 上述所述的赔付率、车险风险分值仅仅是一种或多种实施例对非线性关系风险预测模型的一种输出表征方式。本说明书不限制其他的实施例中还可以有其他的表征方式或者所述赔付率、车险风险分值经过变形、变换的表征方式, 如赔付率经过线性变换后可以得到车险分, 车险分越大, 风险越小 (车险风险分值相反, 风险分值越大, 15 风险越高)。

[56] 需要说明的是, 通常所述的线性关系是指两个变量之间存在一次方函数, 本说明书实施例中所述的保险或车险中变量的线性关系可以包括 $y=ax+b$ 形式, x 为自变量, y 为因变量。本说明书实施例在具体的保险或车险业务应用场景中, 所述的线性关系广义的理解可以是指两个变量之间的关系是明确的、固定的, 一些情况下可以用直线表述或者通过一定的数学变化后转化为线性关系 (转化的信息损失在一定范围内)。所述的非 20 线性关系主要是指变量之间的关系是不断变化的, 无法用公式描述, 一些情况下只能用曲线、曲面或不规则的线来表示, 如风险分值与职业、风险分值与性别。

[57] 本说明书一个或多个实施例中, 所述的构建风险预测模型的处理, 可以采用离线预先构建的方式生成, 可以预先选取包含非线性关系的训练数据进行 GBDT 决策树的学习训练, 训练完成后再在线上使用。本说明书不排除所述风险预测模型可以采用在线构建或更新/维护的方式, 例如在计算机能力足够的情况下, 可以在线构建出风险预测模型, 构建出风险预测模型可以同步在线使用, 对待预测用的目标风险关联数据进行处理。

[58] 本说明书实施例提供的一种保险业务风险预测的处理方法, 可以预先使用梯度提升决策树来构建风险预测模型, 该风险预测模型可以使用已打标的与保险业务相关联的

风险关联数据进行训练。当风险预测模型训练达到预测要求时可以作为线上风险预测使用，对待预测用户进行保险业务风险预测，输出预测结果。利用本说明书实施例提供的方法可以合理有效地应用保险业务中多维度的非线性变量，基于梯度提升决策树的非线性关系的风险预测模型可以很好地兼容线性和非线性变量，相对于传统的线性模型，预测结果的准确性有着显著的提升，有效弥补传统线性模型的不足，提高保险业务服务体验。

[59] 上述所述的方法可以用于客户端一侧的风险识别，如移动终端的支付应用中提供的保险业务的风险评估。所述的客户端可以为 PC (personal computer) 机、服务器、工控机 (工业控制计算机)、移动智能电话、平板电子设备、便携式计算机 (例如笔记本电脑等)、个人数字助理 (PDA)、或桌面型计算机或智能穿戴设备等。移动通信终端、手持设备、车载设备、可穿戴设备、电视设备、计算设备。也可以应用在保险公司或第三方保险服务机构的系统服务器中，所述的系统服务器可以包括单独的服务器、服务器集群、分布式系统服务器或者处理设备请求数据的服务器与其他相关联数据处理的系统服务器组合。例如，一种实现中可以包括建立在阿里云开放数据处理服务 (Open Data Processing Service, 简称 ODPS) 平台上。可以为来自不同用户需求的各种数据处理任务提供统一的编程接口和界面。基于 ODPS 进行系统性能的保障，实施本说明书实施例方法的系统可以并行处理海量数据并达到最佳的运算性能。

[60] 如前所述，本说明书实施例所提供的方法实施例可以在移动终端、计算机终端、服务器或者类似的运算装置中执行。以运行在服务器上为例，图 4 是本说明书提供的一种应用保险业务风险预测处理方法的服务器的硬件结构框图。如图 4 所示，服务器 10 可以包括一个或多个 (图中仅示出一个) 处理器 102 (处理器 102 可以包括但不限于微处理器 MCU 或可编程逻辑器件 FPGA 等的处理装置)、用于存储数据的存储器 104、以及用于通信功能的传输模块 106。本领域普通技术人员可以理解，图 4 所示的结构仅为示意，其并不对上述电子装置的结构造成限定。例如，服务器 10 还可包括比图 4 中所示更多或者更少的组件，例如还可以包括其他的处理硬件，如数据库或多级缓存，或者具有与图 4 所示不同的配置。

[61] 存储器 104 可用于存储应用程序的软件程序以及模块，如本发明实施例中的搜索方法对应的程序指令/模块，处理器 102 通过运行存储在存储器 104 内的软件程序以及模块，从而执行各种功能应用以及数据处理，即实现上述导航交互界面内容展示的处理方法。存储器 104 可包括高速随机存储器，还可包括非易失性存储器，如一个或者多个磁

性存储装置、闪存、或者其他非易失性固态存储器。在一些实例中，存储器 104 可进一步包括相对于处理器 102 远程设置的存储器，这些远程存储器可以通过网络连接至计算机终端 10。上述网络的实例包括但不限于互联网、企业内部网、局域网、移动通信网及其组合。

- 5 [62] 传输模块 106 用于经由一个网络接收或者发送数据。上述的网络具体实例可包括计算机终端 10 的通信供应商提供的无线网络。在一个实例中，传输模块 106 包括一个网络适配器（Network Interface Controller，NIC），其可通过基站与其他网络设备相连从而可与互联网进行通讯。在一个实例中，传输模块 106 可以为射频（Radio Frequency，RF）模块，其用于通过无线方式与互联网进行通讯。
- 10 [63] 基于上述所述的设备型号识别方法，本说明书还提供一种保险业务风险预测处理装置。所述的装置可以包括使用了本说明书实施例所述方法的系统（包括分布式系统）、软件（应用）、模块、组件、服务器、客户端等并结合必要的实施硬件的设备装置。基于同一创新构思，本说明书提供的一种实施例中的处理装置如下面的实施例所述。由于装置解决问题的实现方案与方法相似，因此本说明书实施例具体的处理装置的实施可以
- 15 参见前述方法的实施，重复之处不再赘述。尽管以下实施例所描述的装置较佳地以软件来实现，但是硬件，或者软件和硬件的组合的实现也是可能并被构想的。具体的，如图 5 所示，图 5 是本说明书提供的一种保险业务风险预测处理装置实施例的模块结构示意图，可以包括：

[64] 预测数据获取模块 201，可以用于获取待预测用户的目标风险关联数据；

- 20 [65] 风险预测模块 202，可以用于利用构建的风险预测模型对所述目标风险关联数据进行处理，输出所述待预测用户的风险预测结果，所述风险预测模型法包括：利用已打标的风险关联数据对梯度提升决策树进行训练确定的预测模型。

[66] 需要说明的是，本说明书实施例上述所述的装置和，根据相关方法实施例的描述还可以包括其他的实施方式。具体的实现方式可以参照方法实施例的描述，在此不作一

25 一赘述。

[67] 本说明书实施例提供的服务器或客户端可以在计算机中由处理器执行相应的程序指令来实现，如使用 windows 操作系统的 c++ 语言在 PC 端或服务器端实现，或其他例如 Linux、系统相对应的应用设计语言集合必要的硬件实现，或者基于量子计算机的处理逻辑实现等。上述的处理设备可以具体的为保险服务器或第三方服务机构提供风险预

测的服务器，所述的服务器可以为单独的服务器、服务器集群、分布式系统服务器或者处理设备请求数据的服务器与其他相关联数据处理的系统服务器组合。本说明书还提供一种保险业务风险预测处理设备，具体的可以包括处理器以及用于存储处理器可执行指令的存储器，所述处理器执行所述指令时实现：

5 [68] 获取待预测用户的目标风险关联数据；

[69] 利用构建的风险预测模型对所述目标风险关联数据进行处理，输出所述待预测用户的风险预测结果，所述风险预测模型法包括：利用已打标的风险关联数据对梯度提升决策树进行训练确定的预测模型。

10 [70] 基于前述方式实施例所述，本说明书提供的所述处理设备的另一个实施例中，所述风险关联数据包括至少一个类别的用户特征数据，所述用户特征数据包括与保险业务相关联的非线性关系的数据信息。

[71] 基于前述方式实施例所述，本说明书提供的所述处理设备的另一个实施例中，所述处理器采用下述方式训练得到所述风险预测模型：

15 [72] 确定决策树的总数量，以及决策树在每次分枝时所使用的决策特征的阈值，所述决策特征为所述用户特征数据的类别中的一种；

[73] 当对一组风险关联数据进行训练时，若训练的决策树的棵树达到预设数值或者决策树的残差满足停止训练条件，则停止该组风险关联数据的训练，所述预设数值小于等于所述总数量；

20 [74] 通过风险关联数据的训练结果调整相应决策树的决策特征的阈值，直至调整后的所述阈值满足风险预测模型的预测结果输出要求时，确定所述风险预测模型。

[75] 基于前述方式实施例所述，本说明书提供的所述处理设备的另一个实施例中，所述决策树的总数量基于所述用户特征数据对应的类别的数量确定。

25 [76] 基于前述方式实施例所述，本说明书提供的所述处理设备的另一个实施例中，所述风险预测模型包括基于与车险业务相关联的风险关联数据进行训练得到的车险风险预测模型；

[77] 所述风险预测结果包括待预测用户的赔付率、车险风险分值中的任意一种。

[78] 上述的指令可以存储在多种计算机可读存储介质中。所述计算机可读存储介质可以包括用于存储信息的物理装置，可以将信息数字化后再以利用电、磁或者光学等方式

的媒体加以存储。本实施例所述的计算机可读存储介质有可以包括：利用电能方式存储信息的装置如，各式存储器，如 RAM、ROM 等；利用磁能方式存储信息的装置如，硬盘、软盘、磁带、磁芯存储器、磁泡存储器、U 盘；利用光学方式存储信息的装置如，CD 或 DVD。当然，还有其他方式的可读存储介质，例如量子存储器、石墨烯存储器等等。上述所述的装置或服务器或客户端或处理设备中的所涉及的指令同上描述。

[79] 需要说明的是，本说明书实施例上述所述的装置和处理设备，根据相关方法实施例的描述还可以包括其他的实施方式。具体的实现方式可以参照方法实施例的描述，在此不作一一赘述。

[80] 本说明书中的各个实施例均采用递进的方式描述，各个实施例之间相同相似的部分互相参见即可，每个实施例重点说明的都是与其他实施例的不同之处。尤其，对于硬件+程序类实施例而言，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。

[81] 上述对本说明书特定实施例进行了描述。其它实施例在所附权利要求书的范围内。在一些情况下，在权利要求书中记载的动作或步骤可以按照不同于实施例中的顺序来执行并且仍然可以实现期望的结果。另外，在附图中描绘的过程不一定要求示出的特定顺序或者连续顺序才能实现期望的结果。在某些实施方式中，多任务处理和并行处理也是可以的或者可能是有利的。

[82] 本说明书实施例提供一种保险业务风险预测的处理方法、装置及处理设备，可以预先使用梯度提升决策树来构建风险预测模型，该风险预测模型可以使用已打标的与保险业务相关联的风险关联数据进行训练。当风险预测模型训练达到预测要求时可以作为线上风险预测使用，对待预测用户进行保险业务风险预测，输出预测结果。利用本说明书实施例提供的方法可以合理有效地应用保险业务中多维度的非线性变量，基于梯度提升决策树的非线性关系的风险预测模型可以很好地兼容线性和非线性变量，相对于传统的线性模型，预测结果的准确性有着显著的提升，有效弥补传统线性模型的不足，提高保险业务服务体验。

[83] 虽然本申请提供了如实施例或流程图所述的方法操作步骤，但基于常规或者无创造性的劳动可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式，不代表唯一的执行顺序。在实际中的装置或系统服务器产品执行时，可以按照实施例或者附图所示的方法顺序执行或者并行执行（例如并行处理器或者多线程处理的环境）。

[84] 尽管本说明书实施例内容中提到线性关系/非线性关系的定义、GBDT 中决策树的结构、GBDT 模型算法的处理过程等之类的数据获取、存储、交互、计算、判断等操作和数据描述，但是，本说明书实施例并不局限于必须是符合行业通信标准、标准 GBDT 模型算法处理、通信协议和标准数据模型/模板或本说明书实施例所描述的情况。某些行业标准或者使用自定义方式或实施例描述的实施例基础上略加修改后的实施方案也可以实现上述实施例相同、等同或相近、或变形后可预料的实施效果。应用这些修改或变形后的数据获取、存储、判断、处理方式等获取的实施例，仍然可以属于本说明书的可选实施方案范围之内。

[85] 在 20 世纪 90 年代，对于一个技术的改进可以很明显地区分是硬件上的改进（例如，对二极管、晶体管、开关等电路结构的改进）还是软件上的改进（对于方法流程的改进）。然而，随着技术的发展，当今的很多方法流程的改进已经可以视为硬件电路结构的直接改进。设计人员几乎都通过将改进的方法流程编程到硬件电路中来得到相应的硬件电路结构。因此，不能说一个方法流程的改进就不能用硬件实体模块来实现。例如，可编程逻辑器件（Programmable Logic Device, PLD）（例如现场可编程门阵列（Field Programmable Gate Array, FPGA））就是这样一种集成电路，其逻辑功能由用户对器件编程来确定。由设计人员自行编程来把一个数字系统“集成”在一片 PLD 上，而不需要请芯片制造厂商来设计和制作专用的集成电路芯片。而且，如今，取代手工地制作集成电路芯片，这种编程也多半改用“逻辑编译器（logic compiler）”软件来实现，它与程序开发撰写时所用的软件编译器相类似，而要编译之前的原始代码也得用特定的编程语言来撰写，此称之为硬件描述语言（Hardware Description Language, HDL），而 HDL 也并非仅有一种，而是有许多种，如 ABEL（Advanced Boolean Expression Language）、AHDL（Altera Hardware Description Language）、Confluence、CUPL（Cornell University Programming Language）、HDCal、JHDL（Java Hardware Description Language）、Lava、Lola、MyHDL、PALASM、RHDL（Ruby Hardware Description Language）等，目前最普遍使用的是 VHDL（Very-High-Speed Integrated Circuit Hardware Description Language）与 Verilog。本领域技术人员也应该清楚，只需要将方法流程用上述几种硬件描述语言稍作逻辑编程并编程到集成电路中，就可以很容易得到实现该逻辑方法流程的硬件电路。

[86] 控制器可以按任何适当的方式实现，例如，控制器可以采取例如微处理器或处理器以及存储可由该（微）处理器执行的计算机可读程序代码（例如软件或固件）的计算机可读介质、逻辑门、开关、专用集成电路（Application Specific Integrated Circuit, ASIC）、可编程逻辑控制器和嵌入微控制器的形式，控制器的例子包括但不限于以下微控制器：

ARC 625D、Atmel AT91SAM、Microchip PIC18F26K20 以及 Silicone Labs C8051F320，存储器控制器还可以被实现为存储器的控制逻辑的一部分。本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又可以是硬件部件内的结构。

[87] 上述实施例阐明的处理设备、装置、模块或单元，具体可以由计算机芯片或实体实现，或者由具有某种功能的产品来实现。一种典型的实现设备为计算机。具体的，计算机例如可以为个人计算机、膝上型计算机、车载人机交互设备、蜂窝电话、相机电话、智能电话、个人数字助理、媒体播放器、导航设备、电子邮件设备、游戏控制台、平板计算机、可穿戴设备或者这些设备中的任何设备的组合。

[88] 虽然本说明书实施例提供了如实施例或流程图所述的方法操作步骤，但基于常规或者无创造性的手段可以包括更多或者更少的操作步骤。实施例中列举的步骤顺序仅仅为众多步骤执行顺序中的一种方式，不代表唯一的执行顺序。在实际中的装置或终端产品执行时，可以按照实施例或者附图所示的方法顺序执行或者并行执行（例如并行处理器或者多线程处理的环境，甚至为分布式数据处理环境）。术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、产品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、产品或者设备所固有的要素。在没有更多限制的情况下，并不排除在包括所述要素的过程、方法、产品或者设备中还存在另外的相同或等同要素。

[89] 为了描述的方便，描述以上装置时以功能分为各种模块分别描述。当然，在实施本说明书实施例时可以把各模块的功能在同一个或多个软件和/或硬件中实现，也可以将实现同一功能的模块由多个子模块或子单元的组合实现等。以上所描述的装置实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

[90] 本领域技术人员也知道，除了以纯计算机可读程序代码方式实现控制器以外，完

全可以通过将方法步骤进行逻辑编程来使得控制器以逻辑门、开关、专用集成电路、可编程逻辑控制器和嵌入微控制器等的形式来实现相同功能。因此这种控制器可以被认为是一种硬件部件，而对其内部包括的用于实现各种功能的装置也可以视为硬件部件内的结构。或者甚至，可以将用于实现各种功能的装置视为既可以是实现方法的软件模块又

5 可以是硬件部件内的结构。

[91] 本发明是参照根据本发明实施例的方法、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。

10

[92] 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。

15

[93] 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。

20

[94] 在一个典型的配置中，计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[95] 内存可能包括计算机可读介质中的非永久性存储器，随机存取存储器(RAM)和/或非易失性内存等形式，如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

25

[96] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括，但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储

器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带, 磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质, 可用于存储可以被计算设备访问的信息。按照本文中的界定, 计算机可读介质不包括暂存电脑可读媒体(transitory media), 如调制的数据信号和载波。

[97] 本领域技术人员应明白, 本说明书的实施例可提供为方法、系统或计算机程序产品。因此, 本说明书实施例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, 本说明书实施例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。

[98] 本说明书实施例可以在由计算机执行的计算机可执行指令的一般上下文中描述, 例如程序模块。一般地, 程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。也可以在分布式计算环境中实践本说明书实施例, 在这些分布式计算环境中, 由通过通信网络而被连接的远程处理设备来执行任务。在分布式计算环境中, 程序模块可以位于包括存储设备在内的本地和远程计算机存储介质中。

[99] 本说明书中的各个实施例均采用递进的方式描述, 各个实施例之间相同相似的部分互相参见即可, 每个实施例重点说明的都是与其他实施例的不同之处。尤其, 对于系统实施例而言, 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。在本说明书的描述中, 参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本说明书实施例的至少一个实施例或示例中。在本说明书中, 对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且, 描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外, 在不相互矛盾的情况下, 本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

[100] 以上所述仅为本说明书实施例的实施例而已, 并不用于限制本说明书实施例。对于本领域技术人员来说, 本说明书实施例可以有各种更改和变化。凡在本说明书实施例的精神和原理之内所作的任何修改、等同替换、改进等, 均应包含在本说明书实施例的权利要求范围之内。

权利要求书

1、一种保险业务风险预测的处理方法，所述方法包括：

获取待预测用户的目标风险关联数据；

5 利用构建的风险预测模型对所述目标风险关联数据进行处理，输出所述待预测用户
的风险预测结果，所述风险预测模型法包括：利用已打标的风险关联数据对梯度提升决
策树进行训练确定的预测模型。

2、如权利要求 1 所述的方法，所述风险关联数据包括至少一个类别的用户特征数
据，所述用户特征数据包括与保险业务相关联的非线性关系的数据信息。

3、如权利要求 2 所述的方法，采用下述方式训练得到所述风险预测模型：

10 确定决策树的总数量，以及决策树在每次分枝时所使用的决策特征的阈值，所述决
策特征为所述用户特征数据的类别中的一种；

当对一组风险关联数据进行训练时，若训练的决策树的棵树达到预设数值或者决策
树的残差满足停止训练条件，则停止该组风险关联数据的训练，输出风险关联数据对应
的风险预测数值，所述预设数值小于等于所述总数量；

15 通过风险关联数据的训练结果调整相应决策树的决策特征的阈值，直至调整后的所
述阈值满足风险预测模型的预测结果输出要求时，确定所述风险预测模型。

4、如权利要求 3 所述的方法，所述决策树的总数量基于所述用户特征数据对应的
类别的数量确定。

20 5、如权利要求 1 至 4 中任意一项所述的方法，所述风险预测模型为基于与车险业
务相关联的风险关联数据进行训练得到的车险风险预测模型；

所述风险预测结果包括待预测用户的赔付率、车险风险分值中的任意一种。

6、一种保险业务风险预测处理装置，包括：

预测数据获取模块，用于获取待预测用户的目标风险关联数据；

25 风险预测模块，用于利用构建的风险预测模型对所述目标风险关联数据进行处理，
输出所述待预测用户的风险预测结果，所述风险预测模型法包括：利用已打标的风险关
联数据对梯度提升决策树进行训练确定的预测模型。

7、一种保险业务风险预测处理设备，包括处理器以及用于存储处理器可执行指令
的存储器，所述处理器执行所述指令时实现：

获取待预测用户的目标风险关联数据；

30 利用构建的风险预测模型对所述目标风险关联数据进行处理，输出所述待预测用户
的风险预测结果，所述风险预测模型法包括：利用已打标的风险关联数据对梯度提升决

策树进行训练确定的预测模型。

8、如权利要求 7 所述的处理设备，所述风险关联数据包括至少一个类别的用户特征数据，所述用户特征数据包括与保险业务相关联的非线性关系的数据信息。

5 9、如权利要求 8 所述的处理设备，所述处理器采用下述方式训练得到所述风险预测模型：

确定决策树的总数量，以及决策树在每次分枝时所使用的决策特征的阈值，所述决策特征为所述用户特征数据的类别中的一种；

10 当对一组风险关联数据进行训练时，若训练的决策树的棵树达到预设数值或者决策树的残差满足停止训练条件，则停止该组风险关联数据的训练，所述预设数值小于等于所述总数量；

通过风险关联数据的训练结果调整相应决策树的决策特征的阈值，直至调整后的所述阈值满足风险预测模型的预测结果输出要求时，确定所述风险预测模型。

10、如权利要求 9 所述的处理设备，所述决策树的总数量基于所述用户特征数据对应的类别的数量确定。

15 11、如权利要求 7-10 中任意一项所述的处理设备，所述风险预测模型包括基于与车险业务相关联的风险关联数据进行训练得到的车险风险预测模型；

所述风险预测结果包括待预测用户的赔付率、车险风险分值中的任意一种。

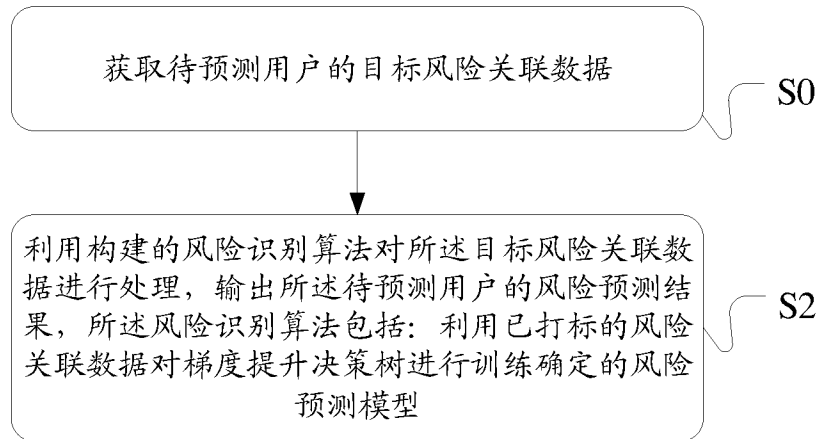


图 1

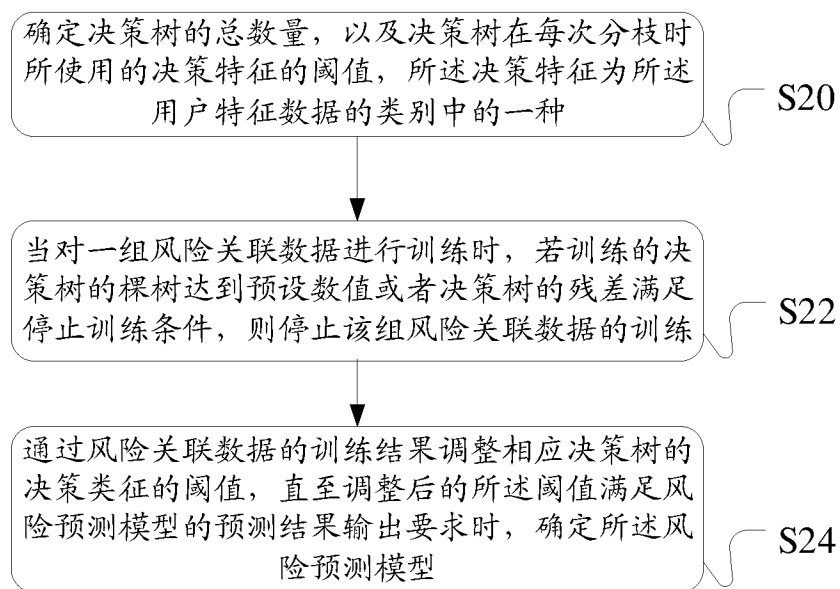


图 2

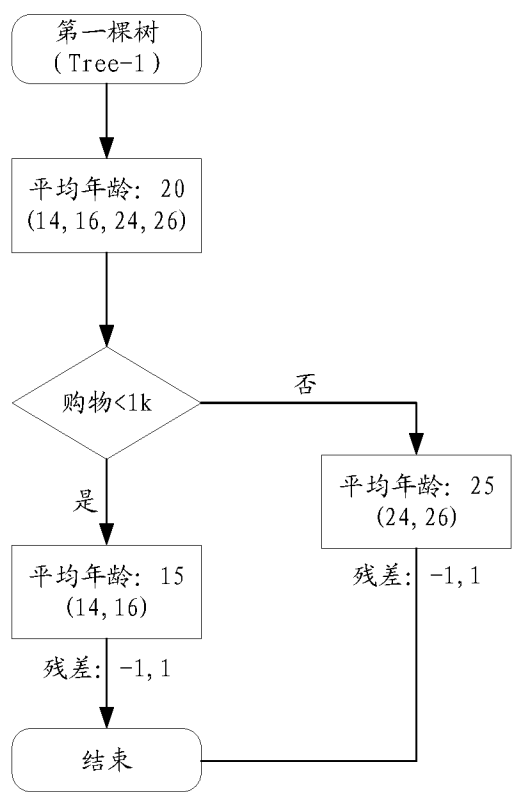


图 3a

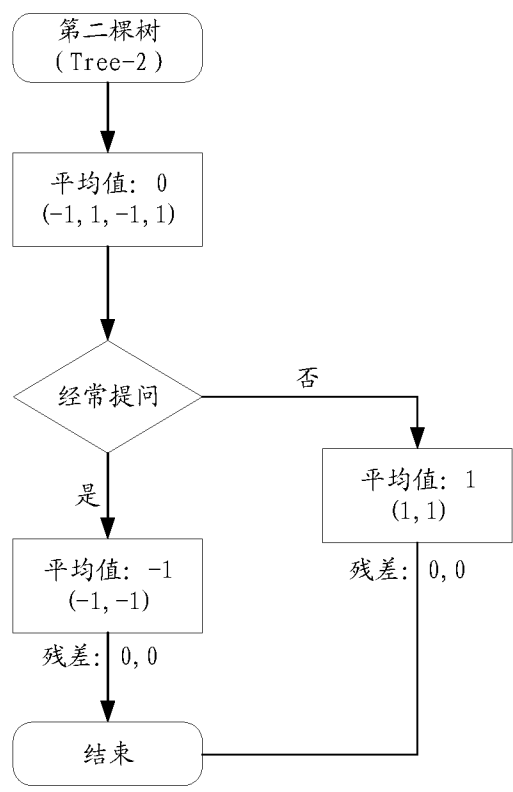


图 3b

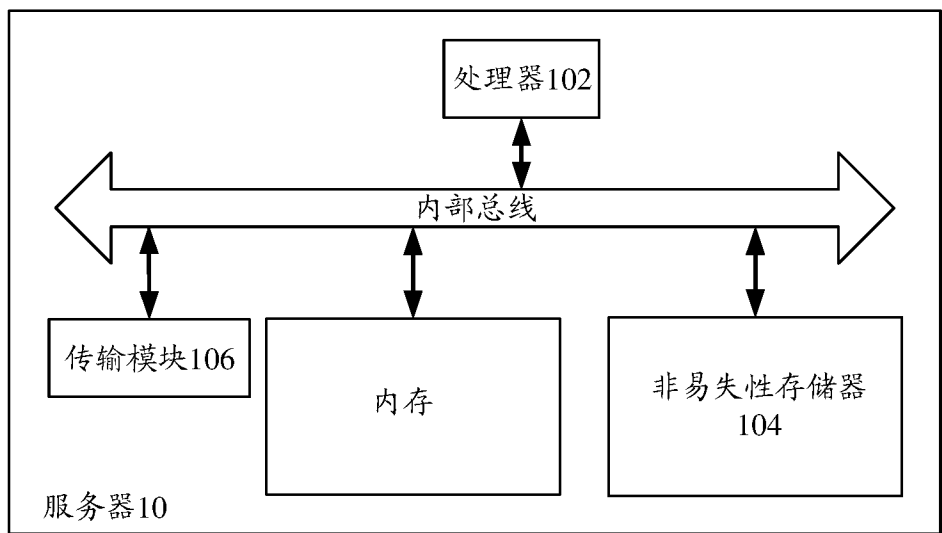


图 4

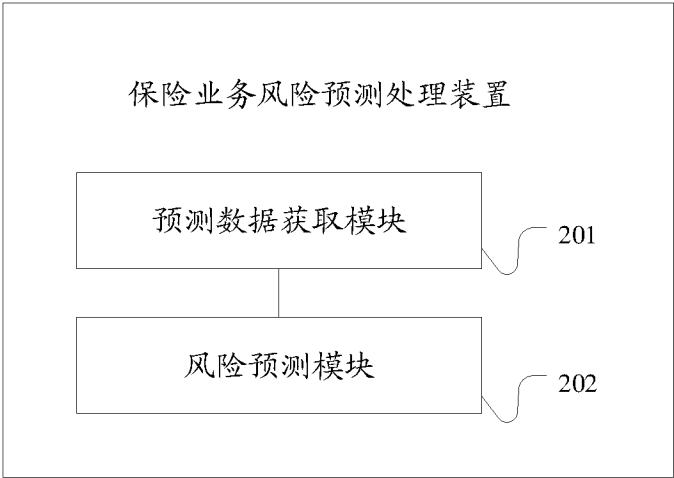


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/076524

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/06(2012.01)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; G06K

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, 车险, 风险, 预测, 预估, 评估, 决策树, 梯度, GBDT, insurance, assurance, risk, forecast, gradient, decision w tree

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108665175 A (ALIBABA GROUP HOLDING LIMITED) 16 October 2018 (2018-10-16) claims 1-11	1-11
X	CN 107292528 A (ALIBABA GROUP HOLDING LIMITED) 24 October 2017 (2017-10-24) claims 1-20, description, paragraphs [0031], [0071]-[0078] and [0137], and figures 1-10	1-11
X	CN 108009914 A (MASHANG XIAOFEI FINANCE CO., LTD.) 08 May 2018 (2018-05-08) description, paragraphs [0045]-[0106], and figures 1-4	1-11
A	CN 106503863 A (BEIJING HONGMA MEDIA CULTURE DEVELOPMENT CO., LTD.) 15 March 2017 (2017-03-15) entire document	1-11
A	CN 106447383 A (HANGZHOU QIGUAN NETWORK TECHNOLOGY CO., LTD.) 22 February 2017 (2017-02-22) entire document	1-11



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

07 May 2019

Date of mailing of the international search report

30 May 2019

Name and mailing address of the ISA/CN

**National Intellectual Property Administration, PRC (ISA/
CN)**
**No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088**
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/076524

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108665175	A	16 October 2018	None			
CN	107292528	A	24 October 2017	US	2019005586	A1	03 January 2019
				WO	2019006373	A1	03 January 2019
				HK	1245967	A0	31 August 2018
CN	108009914	A	08 May 2018	None			
CN	106503863	A	15 March 2017	None			
CN	106447383	A	22 February 2017	None			

国际检索报告

国际申请号

PCT/CN2019/076524

A. 主题的分类

G06Q 10/06 (2012.01) 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; G06K

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, EPDOC, IEEE: 保险, 车险, 风险, 预测, 预估, 评估, 决策树, 梯度, GBDT, insurance, assurance, risk, forecast, gradient, decision w tree

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
PX	CN 108665175 A (阿里巴巴集团控股有限公司) 2018年 10月 16日 (2018 - 10 - 16) 权利要求1-11	1-11
X	CN 107292528 A (阿里巴巴集团控股有限公司) 2017年 10月 24日 (2017 - 10 - 24) 权利要求1-20、说明书第0031, 0071-0078, 0137段、附图1-10	1-11
X	CN 108009914 A (马上消费金融股份有限公司) 2018年 5月 8日 (2018 - 05 - 08) 说明书第0045-0106段、附图1-4	1-11
A	CN 106503863 A (北京红马传媒文化发展有限公司) 2017年 3月 15日 (2017 - 03 - 15) 全文	1-11
A	CN 106447383 A (杭州启冠网络技术有限公司) 2017年 2月 22日 (2017 - 02 - 22) 全文	1-11

☐ 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

2019年 5月 7日

国际检索报告邮寄日期

2019年 5月 30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

受权官员

张雯

传真号 (86-10) 62019451

电话号码 86-(10)-53961314

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/076524

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	108665175	A	2018年 10月 16日	无			
CN	107292528	A	2017年 10月 24日	US	2019005586	A1	2019年 1月 3日
				WO	2019006373	A1	2019年 1月 3日
				HK	1245967	A0	2018年 8月 31日
CN	108009914	A	2018年 5月 8日	无			
CN	106503863	A	2017年 3月 15日	无			
CN	106447383	A	2017年 2月 22日	无			