

(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
2018 年 5 月 24 日 (24.05.2018)



(10) 国际公布号
WO 2018/090657 A1

- (51) 国际专利分类号:
G06Q 10/04 (2012.01)
- (21) 国际申请号: PCT/CN2017/093803
- (22) 国际申请日: 2017 年 7 月 21 日 (21.07.2017)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
2016110240477 2016 年 11 月 18 日 (18.11.2016) CN
- (71) 申请人: 同济大学 (TONGJI UNIVERSITY) [CN/CN]; 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。
- (72) 发明人: 蒋昌俊 (JIANG, Changjun); 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。

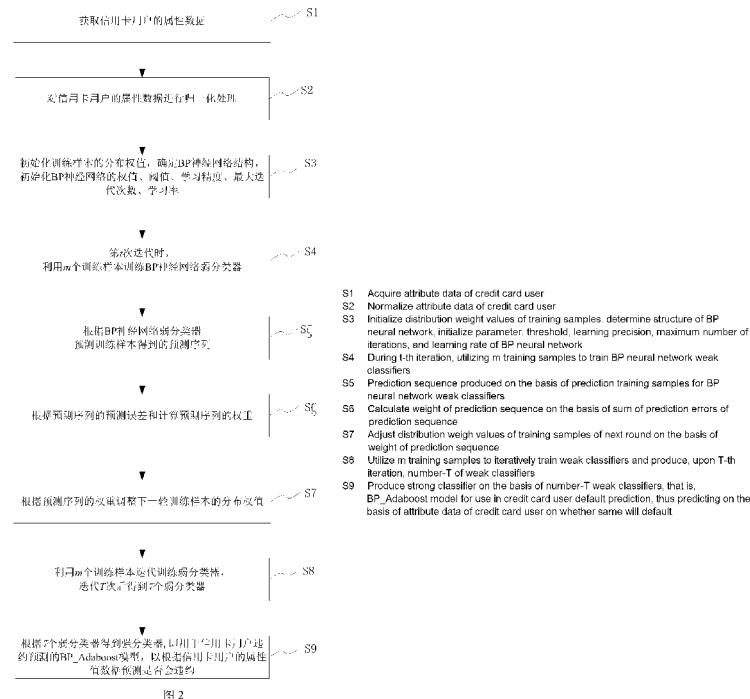
闫春钢 (YAN, Chungang); 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。 丁志军 (DING, Zhijun); 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。 张亚英 (ZHANG, Yaying); 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。 许朴 (XU, Pu); 中国上海市杨浦区四平路 1239 号, Shanghai 200090 (CN)。

(74) 代理人: 上海光华专利事务所 (J.Z.M.C. PATENT AND TRADEMARK LAW OFFICE); 中国上海市杨浦区国定路 335 号 5022 室余明伟, Shanghai 200433 (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,

(54) Title: BP_ADABOOST MODEL-BASED METHOD AND SYSTEM FOR PREDICTING CREDIT CARD USER DEFAULT

(54) 发明名称: 基于BP_Adaboost模型的信用卡用户违约的预测方法及系统



(57) Abstract: A BP_Adaboost model-based method and system for predicting credit card user default, comprising: acquiring attribute data of a credit card user and normalizing to produce a training sample set and a test sample set; initializing distribution weight values of training samples, determining the structure of a BP neural network, and initializing a parameter of the BP neural network; utilizing the training samples in training a number T of BP neural network weak classifiers; acquiring a strong classifier on the basis of the number T of weak classifiers, that is, a BP-Adaboost model for use in credit card user default prediction, thus predicting on the basis of the attribute

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))。

data of the credit card user on whether same will default. The BP_Adaboost model-based method and system for predicting credit card user default perform data analysis and training on the basis of credit history of bank credit card users and establish a BP_Adaboost model, thus increasing the accuracy of credit card user default prediction.

(57) 摘要: 一种基于BP_Adaboost模型的信用卡用户违约的预测方法及系统, 包括获取信用卡用户的属性数据, 并进行归一化处理, 得到训练样本集和测试样本集; 初始化训练样本的分布权值, 确定BP神经网络结构, 初始化BP神经网络的参数; 利用训练样本, 训练T个BP神经网络弱分类器; 根据T个弱分类器得到强分类器, 即用于信用卡用户违约预测的BP_Adaboost模型, 以根据信用卡用户的属性数据预测是否会违约。基于BP_Adaboost模型的信用卡用户违约的预测方法及系统根据银行信用卡用户的历史信用信息进行数据分析、训练, 建立BP_AdaBoost模型, 提升了信用卡用户违约预测的准确度。

基于 BP_Adaboost 模型的信用卡用户违约的预测方法及系统

技术领域

本发明涉及一种预测方法及系统，特别是涉及一种基于 BP_Adaboost 模型的信用卡用户违约的预测方法及系统。

背景技术

随着金融业、银行业的发展，信用卡的使用也越来越广泛，给消费者的生活带来了很多便利。然而，信用卡的不安全因素也暴露无遗，例如，很多信用卡发行银行为了增加市场份额，超额发行现金和信用卡给不合格的申请者。同时，大多数的信用卡持有者不考虑其偿还能力，在消费上过度使用信用卡，积累了大量的信用卡债务，给发卡银行造成了巨大的损失。

信用卡违约风险是由持卡人做出部分或全部支付的承诺时潜在的支付失败驱动的，它在持卡人在还款期限内出现资金紧张或不愿意偿还时发生。在一个健全的财务系统中，风险管理是在下游，而风险预测是在上游。风险预测的主要目的是利用企业绩效或者个人客户的信用风险，来降低风险和不确定性，因而具有很大的现实意义和价值。

目前，对于违约风险，银行往往是利用传统的“评分系统”或者所谓“专家评分”来确定申请者风险水平。具体地，对信用卡持有人的信息数据进行评分，对不同的属性如年龄、职业、学历、收入及其在外部征信系统中的信用历史记录信息，根据其在违约预测方面的重要程度不同，予以不同的分值，以评分结果来预测持卡人在未来一定时期内用卡发生违约拖欠的概率。这种传统的评分方法，因为人的参与程度很大，所以具有很大的随意性和不稳定性。

数据挖掘作为一种发现驱动型的知识发现技术，主要针对非常大型的数据进行研究和分析，采用自动或半自动的程序找出先前未知的、有趣的、可理解的隐含信息。利用数据挖掘方法，银行可以方便地部署一个预测模型，快速而有效地分析大量信用卡持有者的违约状况，已成为信用卡分析的重要工具。

为了降低信用评价中的主观因素影响和反映信用的动态变化，判别分析、逻辑回归、分类树和神经网络等方法逐渐被应用到信用评价领域，对信用卡用户的违约行为进行预测。

判别分析(Discriminant analysis, DA)又称分辨法，是在分类确定的条件下，根据某一研究对象的各种特征值判别其类型归属问题的一种多变量统计分析方法。判别分析的基本原理是按照一定的判别准则，建立一个或多个判别函数，用研究对象的大量资料确定判别函数中的待定系数，并计算判别指标，据此即可预测某一客户样本属于“好”客户还是违约用户。

逻辑回归（Logistic regression LR）可以被看作是线性回归模型的一种特例。该方法能产生一个简单的分类概率公式，判断样本属于各个类别的概率，但其缺点在于不能很好的处理非线性问题以及解释变量之间的交互性影响。

分类树方法（Classification Trees, CTs）是一种非参数统计方法，利用这种方法建立的信用卡违约预测模型并不是生成一个线性的评分表，其基本思想是按照某个特征变量的取值将客户总体分成两个子组，使得在不同组之间客户的违约概率差距尽量地大，而同一组内客户的违约概率尽量一致，然后再对这些子组进行再划分，达到一定的要求前，一直不断的重复这一过程，最后终止。整个过程可以用一个树状结构来表示，按照一定的规则，将每个根节点划分为“好”客户或“违约”客户。

人工神经网络具备其他很多数据挖掘方法不具备的优势，因其非线性假设、泛化能力和映射能力等，在很多领域都得到了应用。其中 BP（Back Propagation，反向传播）神经网络的应用最为广泛。BP 网络是 1986 年由 Rinehart 和 McClelland 为首的科学家小组提出，是一种按误差逆传播算法训练的多层前馈网络，它具有强大并行处理机制，高度自学习、自适应能力。而且 BP 神经网络算法非常的灵活，因为其内部的很多参数都是可以调节的。BP 神经网络算法的主要设计思想是：首先输入信号通过隐藏层和输出层节点的处理计算，得到网络的实际输出，随后将其与期望得到的输出相比较，并计算实际输出与期望输出的误差；然后将误差作为修改权值的依据反向传播至输入层。这一过程中，计算输出层、隐藏层的权值、阈值增益，并更新权值、阈值，反复这一过程，直到实际输出与期望输出的误差达到预先设定的误差收敛标准或者达到最高迭代次数，从而获得最终的神经网络权值，构建信用卡违约预测模型，对样本进行预测。

上述方法在信用卡违约预测方面，都能实现一定的准确率，但都忽略了信用卡数据的不均衡性，也就是违约客户的样本与诚实客户的样本数目相差太大的问题。

Boosting 是一种提高任意给定学习算法准确度的方法，其思想起源于 Valiant 提出的 PAC（Probably Approximately Correct）学习模型。1995 年，Freund 和 Schapire 改进了 Boosting 算法，提出了 AdaBoost（Adaptive Boosting）算法，Adaboost 算法的核心思想是针对同一个训练集训练不同的分类器（弱分类器），然后把这些弱分类器集合起来，构成一个更强的最终分类器（强分类器）。因此，Adaboost 算法对不平衡数据集有着很好的鲁棒性，同时可以加强分类效果，进一步提高神经网络的泛化能力。该算法效率和 Freund 于 1991 年提出的 Boosting 算法几乎相同，但不需要任何关于弱学习器的先验知识，因而更容易应用到实际问题当中。使用 Adaboost 分类器可以排除一些不必要的训练数据特征，并将重点放在关键的训

练数据上面。在产生单个的基分类器时可用相同的分类算法，也可用不同的分类算法。这些算法一般是不稳定的弱分类算法，如 BP 算法。

因此，如何结合 BP 算法和 Adaboosting 算法来对信用卡用户进行违约预测成为一个极具前景的课题。

发明内容

鉴于以上所述现有技术的缺点，本发明的目的在于提供一种基于 BP_Adaboost 模型的信用卡用户违约的预测方法及系统，基于 BP 算法和 Adaboosting 算法，根据银行信用卡用户的历史信用信息进行数据分析、训练，建立 BP_AdaBoost 模型，从而实现对新的信用卡用户是否会违约的预测。

为实现上述目的及其他相关目的，本发明提供一种基于 BP_Adaboost 模型的信用卡用户违约的预测方法，包括以下步骤：获取信用卡用户的属性数据，所述信用卡用户的属性为信用卡用户的信用评价指标；对信用卡用户的属性数据进行归一化处理，并将归一化后的每个用户的属性数据作为一个样本，将所有样本分为训练样本集和测试样本集；初始化训练样本的分布权值 $D_1(i) = 1/m$ ， m 为训练样本总数， $i = 1, 2, \dots, m$ 表示训练样本序号，根据训练样本输入、输出维数确定 BP 神经网络结构，初始化 BP 神经网络的权值、阈值、学习精度、最大迭代次数、学习率；第 t 次迭代时，利用 m 个训练样本训练 BP 神经网络弱分类器 f_t ；根据 BP 神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t ，计算预测误差和 $e_t = \sum_i D_t(i)$ ，其中 i 为 $g_t(x_i) \neq y_i$ 时所对应的训练样本序号， x_i 为 m 个训练样本中第 i 个样本点； $g_t(x_i)$ 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的预测分类结果； y_i 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的期望分类结果；根据预测序列的预测误差和 e_t 计算预测序列的权重 $a_t = \frac{1}{2} \ln \left(\frac{1 - e_t}{e_t} \right)$ ；根据预测序列的权重 a_t 调整下一轮训练样本的分布权值为 $D_{t+1}(i) = \frac{D_t(i)}{B_t} \bullet \exp[-a_t y_i g_t(x_i)]$ ， $i = 1, 2, \dots, m$ ，其中 B_t 是归一化因子；利用 m 个训练样本迭代训练 BP 神经网络弱分类器，迭代 T 次后得到 T 个 BP 神经网络弱分类器； T 为自定义参数；根据 T 个弱分类器得到强分类器 $h(x) = \text{sign} \left[\sum_{t=1}^T a_t \bullet f_t \right]$ ，即用于信用卡用户违约预测的 BP_Adaboost 模型，以根据信用卡用户的属性数据预测是否会违约。

于本发明一实施例中，所述信用卡用户的属性包括个人信贷额度、性别、教育程度、婚

姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额。

于本发明一实施例中，对信用卡用户的属性数据进行归一化处理时，采用如下公式：

$$x_{ij}' = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1,2,3,\dots,m$; $j=1,2,3,\dots,n$; m 为样本总数, n 为属性个数; x_{ij}' 表示归一化后第 i 个样本的第 j 个属性值, x_{ij} 表示第 i 个样本的第 j 个属性值; $\max(j)$ 表示 m 个样本中第 j 个属性的最大值, $\max(i)$ 表示 m 个样本中第 i 个样本的最大值; $\min(j)$ 表示 m 个样本中第 j 个属性的最小值, $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

于本发明一实施例中，还包括利用测试样本，检测所述用于信用卡用户违约预测的 BP_Adaboost 模型的性能。

于本发明一实施例中，所述最大迭代次数为 50 次， T 取值为 10。

同时，本发明还提供一种基于 BP_Adaboost 模型的信用卡用户违约的预测系统，包括获取模块、归一化模块、初始化模块、训练模块、预测误差和计算模块、权重计算模块、分布权值计算模块、迭代模块和预测模块；

所述获取模块用于获取信用卡用户的属性数据，所述信用卡用户的属性为信用卡用户的信用评价指标；

所述归一化模块用于对信用卡用户的属性数据进行归一化处理，并将归一化后的每个用户的属性数据作为一个样本，将所有样本分为训练样本集和测试样本集；

所述初始化模块用于初始化训练样本的分布权值 $D_1(i)=1/m$ ， m 为训练样本总数， $i=1,2,\dots,m$ 表示训练样本序号，根据训练样本输入、输出维数确定 BP 神经网络结构，初始化 BP 神经网络的权值、阈值、学习精度、最大迭代次数、学习率；

所述训练模块用于在第 t 次迭代时，利用 m 个训练样本训练 BP 神经网络弱分类器 f_t ；

所述预测误差和计算模块用于根据 BP 神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t ，计算预测误差和 $e_t = \sum_i D_t(i)$ ，其中 i 为 $g_t(x_i) \neq y_i$ 时所对应的训练样本序号， x_i 为 m 个训练样本中第 i 个样本点； $g_t(x_i)$ 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的预测分类结果； y_i 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的期望分类结果；

所述权重计算模块用于根据预测序列的预测误差和 e_t 计算预测序列的权重

$$a_t = \frac{1}{2} \ln \left(\frac{1 - e_t}{e_t} \right);$$

所述分布权值计算模块用于根据预测序列的权重 a_i 调整下一轮训练样本的分布权值为

$$D_{t+1}(i) = \frac{D_t(i)}{B_t} \cdot \exp[-a_t y_i g_t(x_i)], \quad i = 1, 2, \dots, m, \quad \text{其中 } B_t \text{ 是归一化因子};$$

所述迭代模块用于利用 m 个训练样本迭代训练 BP 神经网络弱分类器, 迭代 T 次后得到 T 个 BP 神经网络弱分类器; T 为自定义参数;

所述预测模块用于根据 T 个弱分类器得到强分类器 $h(x) = \text{sign}\left[\sum_{t=1}^T a_t \cdot f_t\right]$, 即用于信用卡用户违约预测的 BP_Adaboost 模型, 以根据信用卡用户的属性数据预测是否会违约。

于本发明一实施例中, 所述信用卡用户的属性包括个人信贷额度、性别、教育程度、婚姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额。

于本发明一实施例中, 所述归一化模块对信用卡用户的属性数据进行归一化处理时, 采用如下公式:

$$x_{ij}' = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1, 2, 3, \dots, m$; $j=1, 2, 3, \dots, n$; m 为样本总数, n 为属性个数; x_{ij}' 表示归一化后第 i 个样本的第 j 个属性值, x_{ij} 表示第 i 个样本的第 j 个属性值; $\max(j)$ 表示 m 个样本中第 j 个属性的最大值, $\max(i)$ 表示 m 个样本中第 i 个样本的最大值; $\min(j)$ 表示 m 个样本中第 j 个属性的最小值, $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

于本发明一实施例中, 还包括检测模块, 用于利用测试样本, 检测所述用于信用卡用户违约预测的 BP_Adaboost 模型的性能。

于本发明一实施例中, 所述最大迭代次数为 50 次, T 取值为 10。

如上所述, 本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法及系统, 具有以下有益效果:

(1) 基于 BP 算法和 Adaboosting 算法, 根据银行信用卡用户的历史信用信息进行数据分析、训练, 建立 BP_AdaBoost 模型, 从而根据新的信用卡用户的相关信用指标, 对其是否违约进行预测与判定;

(2) 节省了银行人员的大量决策时间、并可辅助其做出最佳决策, 从而有效地规避相关风险;

(3) BP_Adaboost 模型对不平衡数据集有很好的处理能力, 进一步提高神经网络的泛化能力, 提升了信用卡用户违约预测的准确度。

附图说明

图 1 显示为单隐藏层 BP 神经网络模型的结构示意图；

图 2 显示为本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法的流程图；

图 3 显示为本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法的一个实施例中实验结果的 ROC 曲线示意图；

图 4 显示为本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测系统的结构示意图。

元件标号说明

- | | |
|---|-----------|
| 1 | 获取模块 |
| 2 | 归一化模块 |
| 3 | 初始化模块 |
| 4 | 训练模块 |
| 5 | 预测误差和计算模块 |
| 6 | 权重计算模块 |
| 7 | 分布权值计算模块 |
| 8 | 迭代模块 |
| 9 | 预测模块 |

具体实施方式

以下通过特定的具体实例说明本发明的实施方式，本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用，本说明书中的各项细节也可以基于不同观点与应用，在没有背离本发明的精神下进行各种修饰或改变。需说明的是，在不冲突的情况下，以下实施例及实施例中的特征可以相互组合。

需要说明的是，以下实施例中所提供的图示仅以示意方式说明本发明的基本构想，遂图式中仅显示与本发明中有关的组件而非按照实际实施时的组件数目、形状及尺寸绘制，其实际实施时各组件的型态、数量及比例可为一种随意的改变，且其组件布局型态也可能更为复杂。

如图 1 所示，BP 神经网络由输入层（I）、隐层（H）和输出层（O）组成。其中，X、Z 分别表示网络的输入、输出，每一神经元用一个节点表示。

BP 算法由信息的正向传递与误差的反向传播组成。信息的正向传递过程中，输入信息从输入层经隐含层逐层计算传向输出层，每一层神经元的状态只影响下一层神经元的状态。如果在输出层没有得到期望输出，则计算输出层的误差变化值，然后转向反向传播，通过网络将误差信号沿原来的连接通路反传回来修改各神经元的权值直至达到期望输出。神经网络理论已经证明 BP 神经网络具有强大的非线性映射能力和泛化功能，任一连续函数或映射均可采用三层网络加以实现。

BP 学习算法的具体过程如下：

a) 确定 BP 神经网络的架构，即确定输入层节点的个数、输出层节点的个数，以及隐藏层的层数和每层的节点个数。

其中，输入层节点的个数由样本属性的维度决定。输出层节点的个数由样本分类个数决定。隐藏层的层数和每层的节点个数由用户自定义。如图 2 中隐藏层数目为 1，为单隐层。

1989 年 Robert Hecht-Nielsen 证明了对于任何闭区间内的一个连续函数都可以用一个隐含层的 BP 网络来逼近，这就是万能逼近定理。所以一个三层的 BP 神经网络就可以完成任意的 m 维到 n 维的映射。其中，隐藏层节点数是影响神经网络性能的重要参数之一。隐藏层节点数少，那么网络则不能充分反映输入节点与输出节点之间的复杂函数关系；但隐藏层节点数过多时，又会出现过拟合现象。通过经验公式 $h = \sqrt{m+n} + a$ 可以确定隐藏层节点数目。其中 h 为隐藏层节点数目， m 为输入层节点数目， n 为输出层节点数目， a 为 1-10 之间的调节常数。

经过多次实验结果表明：当隐藏层节点数为 9 时，BP 神经网络的性能较好。

b) 对 BP 神经网络中的参数进行初始化。

权值 w_{ij} 表示前一层节点和后一层节点之间的权值，初始为 $[-1, 1]$ 的随机值。阈值 θ_j 用来改变节点的活性，初始为 $[0, 1]$ 随机值，前层节点至后层节点之间通过权值系数相连接。

c) 正向传递

设节点 i 和节点 j 间的权值为 w_{ij} ，节点 j 的阈值为 b_j ，每个节点的输出值为 x_j ，而每个节点的输出值是根据上层所有节点的输出值、当前节点与上一层所有节点的权值和当前节点的阈值还有激活函数来实现的。

节点的输出值的计算方法如下：

隐藏层和输出层的输入表示为 $I_j = \sum_i w_{ij} O_i + \theta_j$ 。其中， O_i 是上一层的单元 i 的输出； θ_j 表示单元 j 的阈值。

神经元的输出是经由激活函数计算得到的，激活函数一般使用 simoid 函数。故输出表示

为 $O_j = \frac{1}{1+e^{-I_j}}$ 。

故正向传递的过程比较简单，按照上述公式计算即可。在 BP 神经网络中，输入层节点没有阈值。

d) 逆向反馈

BP 神经网络的输出层的误差公式为 $Err_j = O_j(1-O_j)(T_j - O_j)$ 。其中， O_j 是节点 j 的实际输出，而 T_j 是节点 j 基于给定训练样本的已知类标号的真正输出。

隐藏层的每个节点的误差公式为 $Err_j = O_j(1-O_j)\sum_k Err_k \bullet w_{kj}$ 。其中， w_{kj} 是由下一较高层中节点 k 到节点 j 的连接权值，而 Err_k 是节点 k 的误差。

在使误差不断减小的原则下，将输出层和隐藏层的误差公式依次展开至隐藏层和输入层，应使权值的调整量与误差的负梯度成正比。网络中每个权值、阈值的增量分别为 $\Delta w_{ij} = (l)Err_j O_j$ 和 $\Delta \theta_j = (l)Err_j$ ，则分别更新权值、阈值为 $w_{ij} = w_{ij} + \Delta w_{ij}$ 和 $\theta_j = \theta_j + \Delta \theta_j$ 。其中， l 代表学习率，增量公式表示学习率乘以后面两个参数。

学习率也是 BP 网络训练中的一个重要参数，学习率过小，则收敛过慢；学习率过大，则可能修正过头，导致振荡甚至发散。经过多次实验，当 BP 神经网络的学习率设置为 0.00004 时，样本预测的正确率最高，少数类的 F 值最大。

至此完成一次了对权值、阈值的修正，在不满足算法终止条件，即算法误差达到预设精度或学习次数达到设定的最大次数之前，不断的进行正向与反向传播。

因此，BP 神经网络的学习训练过程如下：

- (1) 初始化 BP 神经网络，对网络参数及各权值系数、阈值进行赋值，设定学习精度、最大迭代次数、学习率等。
- (2) 向前传播，输入训练样本，计算输出层的实际值，并与期望值相比较，计算出网络的输出误差。
- (3) 依据误差反向传播规则，调整隐藏层之间以及隐藏层与输入层之间的权值系数、节点的阈值。
- (4) 重复步骤 (2) 和 (3)，直至预测误差满足预设学习精度或学习次数达到设定的最大迭代次数。

本发明将 BP 算法与 Adboost 算法结合，即通过 BP_Adaboost 模型预测信用卡用户是否违约。

参照图 2，本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法包括以下步骤：

步骤 S1、获取信用卡用户的属性数据，其中信用卡用户的属性为信用卡用户的信用评价指标。

具体地，信用卡用户的信用评价指标构成信用卡用户的评价指标体系。为了尽量准确地对信用卡违约状况进行预测，该评价指标体系要全面、真实的反映信用卡持有者的信用状况。

其中，每个信用评价指标作为信用卡用户的一个属性。在本发明中，信用卡用户的属性包括个人信贷额度、性别、教育程度、婚姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额等 23 项。其中，优选为近 6 个月的偿还记录、近 6 个月的账单金额和近 6 个月的实际还款金额。

具体地，可以从网址<http://archive.ics.uci.edu/ml/>下载信用卡用户的属性数据。

步骤 S2、对信用卡用户的属性数据进行归一化处理，并将归一化后的每个用户的属性数据作为一个样本，将所有样本分为训练样本集和测试样本集。

由于某些属性值如个人信贷额度、账单金额等的数值差别太大，影响分类效果，故需要对属性数据进行[-1,1]的归一化处理。

具体地，对信用卡用户的属性数据进行归一化处理时，采用如下公式：

$$x'_{ij} = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1,2,3,\dots,m$; $j=1,2,3,\dots,n$; m 为样本总数， n 为属性个数； x'_{ij} 表示归一化后第 i 个样本的第 j 个属性值， x_{ij} 表示第 i 个样本的第 j 个属性值； $\max(j)$ 表示 m 个样本中第 j 个属性的最大值， $\max(i)$ 表示 m 个样本中第 i 个样本的最大值； $\min(j)$ 表示 m 个样本中第 j 个属性的最小值， $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

经过上述归一化处理，每个属性值都限制在[-1,1]之间。

步骤 S3、初始化训练样本的分布权值 $D_1(i)=1/m$ ， m 为训练样本总数， $i=1,2,\dots,m$ 表示训练样本序号，根据训练样本输入、输出维数确定 BP 神经网络结构，初始化 BP 神经网络的权值、阈值、学习精度、最大迭代次数、学习率。

优选地，最大迭代次数为 50。

在本发明中，每个样本的输入为 23 维，分别表示每个信用评价指标；输出为 1 维，表示个人违约状况预测结果。

步骤 S4、第 t 次迭代时，利用 m 个训练样本训练 BP 神经网络弱分类器 f_t 。

步骤 S5、根据 BP 神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t ，计算预测误差

和 $e_i = \sum_i D_i(i)$ ，其中 i 为 $g_i(x_i) \neq y_i$ 时所对应的训练样本序号， x_i 为 m 个训练样本中第 i 个样本点； $g_i(x_i)$ 为 BP 神经网络弱分类器 f_i 对样本点 x_i 的预测分类结果； y_i 为 BP 神经网络弱分类器 f_i 对样本点 x_i 的期望分类结果。

步骤 S6、根据预测序列的预测误差和 e_i 计算预测序列的权重 $a_i = \frac{1}{2} \ln \left(\frac{1-e_i}{e_i} \right)$ 。

步骤 S7、根据预测序列的权重 a_i 调整下一轮训练样本的分布权值为 $D_{i+1}(i) = \frac{D_i(i)}{B_i} \cdot \exp[-a_i y_i g_i(x_i)]$ ， $i = 1, 2 \dots m$ 。其中 B_i 是归一化因子，目的是在权重比例不变的情况下使分布权值和为 1。

步骤 S8、利用 m 个训练样本迭代训练 BP 神经网络弱分类器，迭代 T 次后得到 T 个 BP 神经网络弱分类器； T 为自定义参数。

优选地，本发明中 T 设置为 10。

步骤 S9、根据 T 个弱分类器得到强分类器 $h(x) = \text{sign} \left[\sum_{i=1}^T a_i \cdot f_i \right]$ ，即用于信用卡用户违约预测的 BP_Adaboost 模型，以根据信用卡用户的属性数据预测是否会违约。

优选地，在训练基于神经网络为弱分类器的 Adaboost 数学模型后，还包括输入测试样本集中的样本，检测该用于信用卡用户违约预测的 BP_Adaboost 模型的性能。具体地，通过与其他模型 F 值、ROC 的对比实验来检测预测模型。

当有新的银行信用卡用户数据时，将数据进行归一化处理后，输入用于信用卡用户违约预测的 BP_Adaboost 模型，即可得到这部分信用卡用户的违约预测结果，预测客户未来的违约情况。

下面结合具体实施例来进一步阐述本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法。

利用台湾某银行的信用卡交易数据为研究对象，共 30000 条记录，其中 6636 条违约记录，23364 条诚实记录，每条记录包括 23 个解释变量，1 个反映变量（违约与否）。信用卡用户的属性数据如表 1 所示。

表 1、信用卡用户的属性数据表

属性名称	属性描述	属性取值
LIMIT_BAL	个人信贷额度	正整数
SEX	性别	1-男性 2-女性

EDUCATION	教育程度	1-研究生 2-大学 3-高中 4-其他
MARRIAGE	婚姻状况	1-已婚 2-单身 3-其他
AGE	年龄	正整数
PAY_0	2005 年 9 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
PAY_2	2005 年 8 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
PAY_3	2005 年 7 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
PAY_4	2005 年 6 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
PAY_5	2005 年 5 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
PAY_6	2005 年 4 月偿还记录	-1-按时还款 1-延期一个月 9-延期 9 个月及以上
BILL_AMT1	2005 年 9 月账单金额	正整数
BILL_AMT2	2005 年 8 月账单金额	正整数
BILL_AMT3	2005 年 7 月账单金额	正整数
BILL_AMT4	2005 年 6 月账单金额	正整数
BILL_AMT5	2005 年 5 月账单金额	正整数
BILL_AMT6	2005 年 4 月账单金额	正整数
PAY_AMT1	2005.9 实际还款金额	正整数
PAY_AMT2	2005.8 实际还款金额	正整数
PAY_AMT3	2005.7 实际还款金额	正整数
PAY_AMT4	2005.6 实际还款金额	正整数
PAY_AMT5	2005.5 实际还款金额	正整数
PAY_AMT6	2005.4 实际还款金额	正整数
Default payment next month	预测下个月是否会违 约	1-违约 0-不违约

该实施例中训练样本集和测试样本集的分布情况如表 2 所示。

表 2、信用卡用户的样本集组成

样本集	样本总数	训练样本	测试样本
下个月违约	6636	5370	1266
下个月不违约	23364	18630	4734

在该实施例中，于 BP_Adaboost 模型的信用卡用户违约的预测方法包括以下步骤：

1) 通过观察样本属性值可知，样本中某些属性，如信贷额度、账单金额等，数值差别太大，影响分类效果，故先对其进行 $[-1,1]$ 的归一化处理。

2) 建立基于 BP 神经网络为弱分类器的 Adaboost 数学模型，得到用于信用卡用户违约预测的 BP_Adaboost 模型。

如上所述，共有 30000 组客户信用状况数据，每组数据的输入为 23 维，分别代表表 1 中列出的前 23 个指标；输出为 1 维，代表个人违约状况预测结果。从中随机选取 24000 组数据作为训练数据，6000 组数据作为测试数据。

3) 输入 6000 个测试样本得到测试样本的分类结果，验证所得用于信用卡用户违约预测的 BP_Adaboost 模型的性能。

实验结果的受试者工作特征曲线（Receiver Operating Characteristic curve, ROC）如图 3 所示。

参照图 4，本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测系统包括获取模块 1、归一化模块 2、初始化模块 3、训练模块 4、预测误差和计算模块 5、权重计算模块 6、分布权值计算模块 7、迭代模块 8 和预测模块 9。

获取模块 1 用于获取信用卡用户的属性数据，其中信用卡用户的属性为信用卡用户的信用评价指标。

具体地，信用卡用户的信用评价指标构成信用卡用户的评价指标体系。为了尽量准确地对信用卡违约状况进行预测，该评价指标体系要全面、真实的反映信用卡持有者的信用状况。

其中，每个信用评价指标作为信用卡用户的一个属性。在本发明中，信用卡用户的属性包括个人信贷额度、性别、教育程度、婚姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额等 23 项。其中，优选为近 6 个月的偿还记录、近 6 个月的账单金额和近 6 个月的实际还款金额。

具体地，可以从网址<http://archive.ics.uci.edu/ml/>下载信用卡用户的属性数据。

归一化模块 2 与获取模块 1 相连，用于对信用卡用户的属性数据进行归一化处理，并将归一化后的每个用户的属性数据作为一个样本，将所有样本分为训练样本集和测试样本集。

由于某些属性值如个人信贷额度、账单金额等的数值差别太大，影响分类效果，故需要对属性数据进行 $[-1,1]$ 的归一化处理。

具体地，对信用卡用户的属性数据进行归一化处理时，采用如下公式：

$$x_{ij}' = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1,2,3,\dots,m$; $j=1,2,3,\dots,n$; m 为样本总数, n 为属性个数; x_{ij}' 表示归一化后第 i 个样本的第 j 个属性值, x_{ij} 表示第 i 个样本的第 j 个属性值; $\max(j)$ 表示 m 个样本中第 j 个属性的最大值, $\max(i)$ 表示 m 个样本中第 i 个样本的最大值; $\min(j)$ 表示 m 个样本中第 j 个属性的最小值, $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

经过上述归一化处理, 每个属性值都限制在 $[-1,1]$ 之间。

初始化模块 3 与归一化模块 2 相连, 用于初始化训练样本的分布权值 $D_1(i) = 1/m$, m 为训练样本总数, $i=1,2,\dots,m$ 表示训练样本序号, 根据训练样本输入、输出维数确定 BP 神经网络结构, 初始化 BP 神经网络的权值、阈值、学习精度、最大迭代次数、学习率。

优选地, 最大迭代次数为 50。

在本发明中, 每个样本的输入为 23 维, 分别表示每个信用评价指标; 输出为 1 维, 表示个人违约状况预测结果。

训练模块 4 与归一化模块 2 和初始化模块 3 相连, 用于在第 t 次迭代时, 利用 m 个训练样本训练 BP 神经网络弱分类器 f_t 。

预测误差和计算模块 5 与训练模块 4 相连, 用于根据 BP 神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t , 计算预测误差和 $e_t = \sum_i D_t(i)$, 其中 i 为 $g_t(x_i) \neq y_i$ 时所对应的训练样本序号, x_i 为 m 个训练样本中第 i 个样本点; $g_t(x_i)$ 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的预测分类结果; y_i 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的期望分类结果。

权重计算模块 6 与预测误差和计算模块 5 相连, 用于根据预测序列的预测误差和 e_t 计算预测序列的权重 $a_t = \frac{1}{2} \ln \left(\frac{1 - e_t}{e_t} \right)$ 。

分布权值计算模块 7 与权重计算模块 6 相连, 用于根据预测序列的权重 a_t 调整下一轮训练样本的分布权值为 $D_{t+1}(i) = \frac{D_t(i)}{B_t} \bullet \exp[-a_t y_i g_t(x_i)]$, $i=1,2,\dots,m$ 。其中 B_t 是归一化因子, 目的是在权重比例不变的情况下使分布权值和为 1。

迭代模块 8 与训练模块 4、预测误差和计算模块 5、权重计算模块 6 和分布权值计算模块 7 相连, 用于利用 m 个训练样本迭代训练 BP 神经网络弱分类器, 迭代 T 次后得到 T 个 BP 神经网络弱分类器; T 为自定义参数。

优选地，本发明中 T 设置为 10。

预测模块 9 与迭代模块 8 相连，用于根据 T 个弱分类器得到强分类器 $h(x) = \text{sign} \left[\sum_{t=1}^T a_t \cdot f_t \right]$ ，即用于信用卡用户违约预测的 BP_Adaboost 模型，以根据信用卡用户的属性数据预测是否会违约。

优选地，还包括检测模块，用于在训练基于神经网络为弱分类器的 Adaboost 数学模型后，输入测试样本集中的样本，检测该用于信用卡用户违约预测的 BP_Adaboost 模型的性能。具体地，通过与其他模型 F 值、ROC 的对比实验来检测预测模型。

当有新的银行信用卡用户数据时，将数据进行归一化处理后，输入用于信用卡用户违约预测的 BP_Adaboost 模型，即可得到这部分信用卡用户的违约预测结果，预测客户未来的违约情况。

综上所述，本发明的基于 BP_Adaboost 模型的信用卡用户违约的预测方法及系统基于 BP 算法和 Adaboosting 算法，根据银行信用卡用户的历史信用信息进行数据分析、训练，建立 BP_AdaBoost 模型，从而根据新的信用卡用户的相关信用指标，对其是否违约进行预测与判定；节省了银行人员的大量决策时间、并可辅助其做出最佳决策，从而有效地规避相关风险；BP_Adaboost 模型对不平衡数据集有很好的处理能力，进一步提高神经网络的泛化能力，提升了信用卡用户违约预测的准确度。所以，本发明有效克服了现有技术中的种种缺点而具高度产业利用价值。

上述实施例仅例示性说明本发明的原理及其功效，而非用于限制本发明。任何熟悉此技术的人士皆可在不违背本发明的精神及范畴下，对上述实施例进行修饰或改变。因此，举凡所属技术领域中具有通常知识者在未脱离本发明所揭示的精神与技术思想下所完成的一切等效修饰或改变，仍应由本发明的权利要求所涵盖。

权利要求书

1、一种基于 BP_Adaboost 模型的信用卡用户违约的预测方法，其特征在于：包括以下步骤：

获取信用卡用户的属性数据，所述信用卡用户的属性为信用卡用户的信用评价指标；

对信用卡用户的属性数据进行归一化处理，并将归一化后的每个用户的属性数据作为一个样本，将所有样本分为训练样本集和测试样本集；

初始化训练样本的分布权值 $D_1(i) = 1/m$ ， m 为训练样本总数， $i = 1, 2, \dots, m$ 表示训练样本序号，根据训练样本输入、输出维数确定 BP 神经网络结构，初始化 BP 神经网络的权值、阈值、学习精度、最大迭代次数、学习率；

第 t 次迭代时，利用 m 个训练样本训练 BP 神经网络弱分类器 f_t ；

根据 BP 神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t ，计算预测误差和 $e_t = \sum_i D_t(i)$ ，其中 i 为 $g_t(x_i) \neq y_i$ 时所对应的训练样本序号， x_i 为 m 个训练样本中第 i 个样本点； $g_t(x_i)$ 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的预测分类结果； y_i 为 BP 神经网络弱分类器 f_t 对样本点 x_i 的期望分类结果；

根据预测序列的预测误差和 e_t 计算预测序列的权重 $a_t = \frac{1}{2} \ln \left(\frac{1 - e_t}{e_t} \right)$ ；

根据预测序列的权重 a_t 调整下一轮训练样本的分布权值为

$D_{t+1}(i) = \frac{D_t(i)}{B_t} \bullet \exp[-a_t y_i g_t(x_i)]$ ， $i = 1, 2, \dots, m$ ，其中 B_t 是归一化因子；

利用 m 个训练样本迭代训练 BP 神经网络弱分类器，迭代 T 次后得到 T 个 BP 神经网络弱分类器； T 为自定义参数；

根据 T 个弱分类器得到强分类器 $h(x) = \text{sign} \left[\sum_{t=1}^T a_t \bullet f_t \right]$ ，即用于信用卡用户违约预测的 BP_Adaboost 模型，以根据信用卡用户的属性数据预测是否会违约。

2、根据权利要求 1 所述的基于 BP_Adaboost 模型的信用卡用户违约的预测方法，其特征在于：所述信用卡用户的属性包括个人信贷额度、性别、教育程度、婚姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额。

3、根据权利要求 1 所述的基于 BP_Adaboost 模型的信用卡用户违约的预测方法，其特征在于：对信用卡用户的属性数据进行归一化处理时，采用如下公式：

$$x_{ij}' = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1,2,3,\dots,m$; $j=1,2,3,\dots,n$; m 为样本总数, n 为属性个数; x_{ij}' 表示归一化后第 i 个样本的第 j 个属性值, x_{ij} 表示第 i 个样本的第 j 个属性值; $\max(j)$ 表示 m 个样本中第 j 个属性的最大值, $\max(i)$ 表示 m 个样本中第 i 个样本的最大值; $\min(j)$ 表示 m 个样本中第 j 个属性的最小值, $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

- 4、根据权利要求1所述的基于BP_Adaboost模型的信用卡用户违约的预测方法,其特征在于:还包括利用测试样本,检测所述用于信用卡用户违约预测的BP_Adaboost模型的性能。
- 5、根据权利要求1所述的基于BP_Adaboost模型的信用卡用户违约的预测方法,其特征在于:所述最大迭代次数为50次, T 取值为10。
- 6、一种基于BP_Adaboost模型的信用卡用户违约的预测系统,其特征在于:包括获取模块、归一化模块、初始化模块、训练模块、预测误差和计算模块、权重计算模块、分布权值计算模块、迭代模块和预测模块;

所述获取模块用于获取信用卡用户的属性数据,所述信用卡用户的属性为信用卡用户的信用评价指标;

所述归一化模块用于对信用卡用户的属性数据进行归一化处理,并将归一化后的每个用户的属性数据作为一个样本,将所有样本分为训练样本集和测试样本集;

所述初始化模块用于初始化训练样本的分布权值 $D_1(i) = 1/m$, m 为训练样本总数, $i=1,2,\dots,m$ 表示训练样本序号,根据训练样本输入、输出维数确定BP神经网络结构,初始化BP神经网络的权值、阈值、学习精度、最大迭代次数、学习率;

所述训练模块用于在第 t 次迭代时,利用 m 个训练样本训练BP神经网络弱分类器 f_t ;

所述预测误差和计算模块用于根据BP神经网络弱分类器 f_t 预测训练样本得到的预测序列 g_t , 计算预测误差和 $e_t = \sum_i D_t(i)$, 其中 i 为 $g_t(x_i) \neq y_i$ 时所对应的训练样本序号, x_i 为 m 个训练样本中第 i 个样本点; $g_t(x_i)$ 为BP神经网络弱分类器 f_t 对样本点 x_i 的预测分类结果; y_i 为BP神经网络弱分类器 f_t 对样本点 x_i 的期望分类结果;

所述权重计算模块用于根据预测序列的预测误差和 e_t 计算预测序列的权重

$$a_i = \frac{1}{2} \ln \left(\frac{1 - e_i}{e_i} \right);$$

所述分布权值计算模块用于根据预测序列的权重 a_i 调整下一轮训练样本的分布权值为 $D_{i+1}(i) = \frac{D_i(i)}{B_i} \bullet \exp[-a_i y_i g_i(x_i)]$, $i = 1, 2 \dots m$, 其中 B_i 是归一化因子;

所述迭代模块用于利用 m 个训练样本迭代训练 BP 神经网络弱分类器, 迭代 T 次后得到 T 个 BP 神经网络弱分类器; T 为自定义参数;

所述预测模块用于根据 T 个弱分类器得到强分类器 $h(x) = \text{sign} \left[\sum_{i=1}^T a_i \bullet f_i \right]$, 即用于信用卡用户违约预测的 BP_Adaboost 模型, 以根据信用卡用户的属性数据预测是否会违约。

7、根据权利要求6所述的基于BP_Adaboost模型的信用卡用户违约的预测系统,其特征在于: 所述信用卡用户的属性包括个人信贷额度、性别、教育程度、婚姻状况、年龄、近几个月的偿还记录、近几个月的账单金额和近几个月的实际还款金额。

8、根据权利要求6所述的基于BP_Adaboost模型的信用卡用户违约的预测系统,其特征在于: 所述归一化模块对信用卡用户的属性数据进行归一化处理时, 采用如下公式:

$$x_{ij}' = \frac{\{\max(j) - \min(j)\} * \{x_{ij} - \min(i)\}}{\max(i) - \min(i)} + \min(j)$$

其中 $i=1,2,3,\dots,m$; $j=1,2,3,\dots,n$; m 为样本总数, n 为属性个数; x_{ij}' 表示归一化后第 i 个样本的第 j 个属性值, x_{ij} 表示第 i 个样本的第 j 个属性值; $\max(j)$ 表示 m 个样本中第 j 个属性的最大值, $\max(i)$ 表示 m 个样本中第 i 个样本的最大值; $\min(j)$ 表示 m 个样本中第 j 个属性的最小值, $\min(i)$ 表示 m 个样本中第 i 个样本的最小值。

9、根据权利要求6所述的基于BP_Adaboost模型的信用卡用户违约的预测系统,其特征在于: 还包括检测模块, 用于利用测试样本, 检测所述用于信用卡用户违约预测的 BP_Adaboost 模型的性能。

10、根据权利要求6所述的基于BP_Adaboost模型的信用卡用户违约的预测系统,其特征在于: 所述最大迭代次数为50次, T 取值为10。

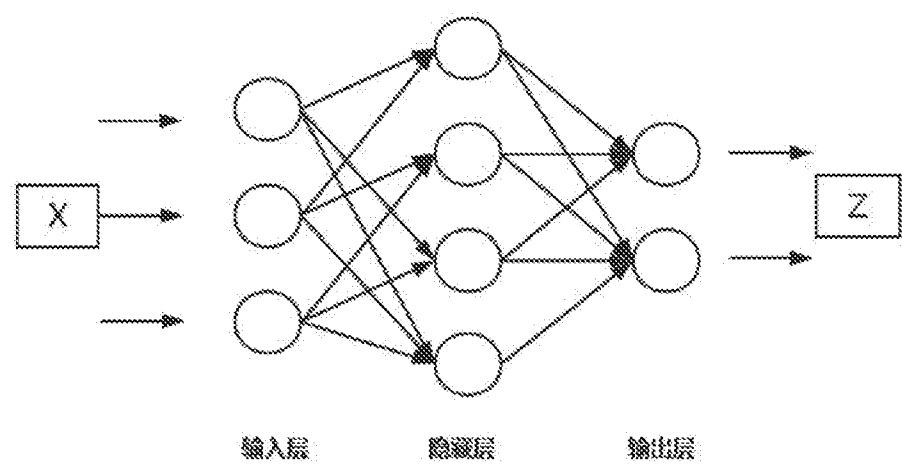


图 1

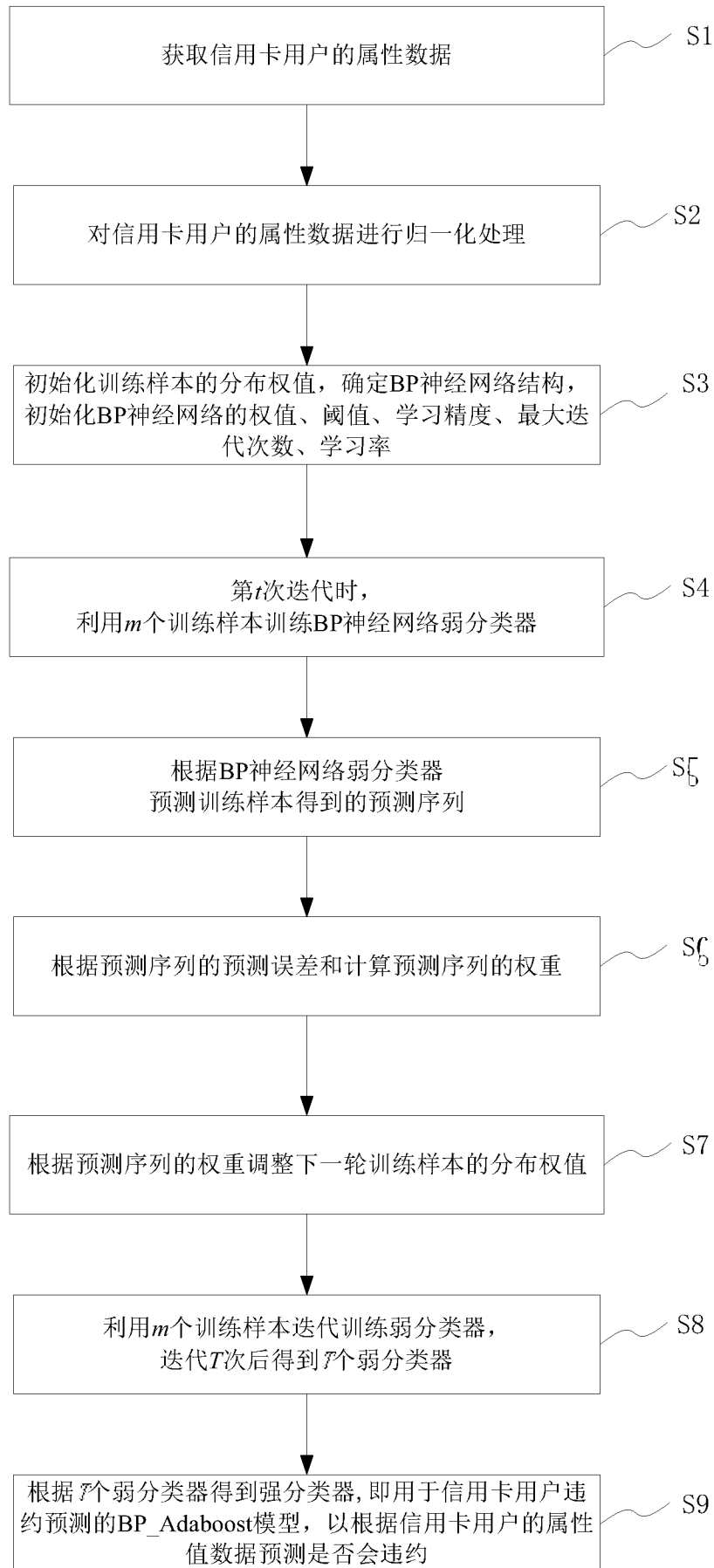


图 2

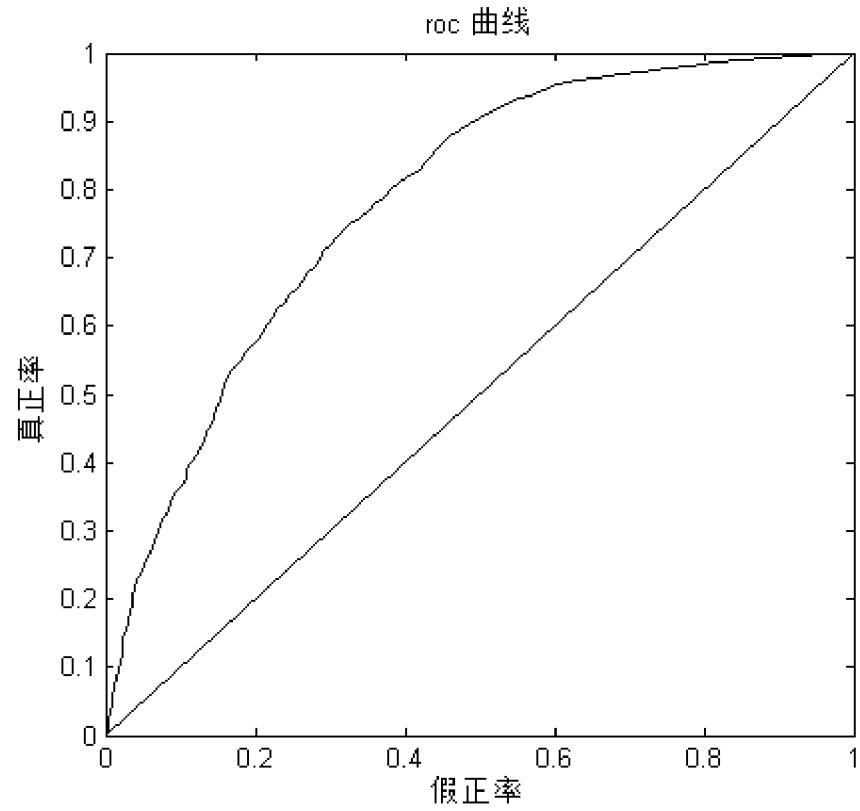


图 3

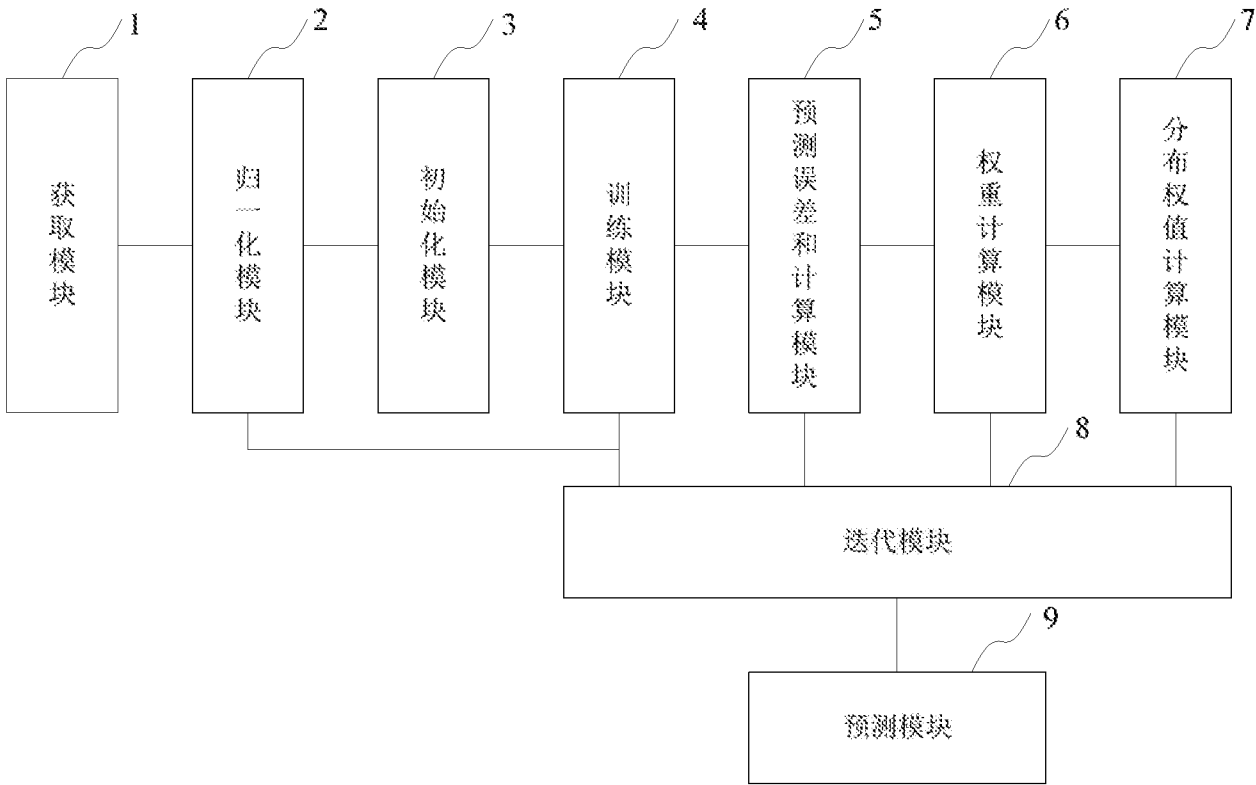


图 4

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2017/093803

A. CLASSIFICATION OF SUBJECT MATTER

G06Q 10/04 (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

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: 信用卡, 违约, 预测, 样本, 归一化, 神经网络, 弱分类器, 强分类器, 预测序列, 权重, 迭代, 调整, BP, adaboost, sample, prediction, sequence, weight, adjust, classifier

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 106529729 A (同济大学), 22 March 2017 (22.03.2017), claims 1-10	1-10
X	王娜. “基于 BP - Adaboost 的上市公司财务危机预警实证研究”, 中国集体经济, no. 1, 31 January 2015 (31.01.2015), ISSN: 1008-1283, pages 144-146, (China Collective Economy), non-official translation (WANG, Na. “Empirical Study on the Pre-warning of Listed Company Finance Crisis Based on BP-Adaboost”)	1-10
X	陆正华等, “基于 BP_Adaboost 算法的上市公司财务预警研究”, 财会通讯, no. 18, 31 December 2013 (31.12.2013), ISSN: 1002-8072, pages 117-119, (Communication of Finance and Accounting), non-official translation (LU, Zhenghua et al., “Study on the Pre-warning of Listed Company Finance Based on BP_Adaboost”)	1-10
A	CN 104240720 A (PEKING UNIVERSITY SHENZHEN GRADUATE SCHOOL), 24 December 2014 (24.12.2014), entire document	1-10
A	CN 106022229 A (BEIHANG UNIVERSITY), 12 October 2016 (12.10.2016), entire document	1-10

☐ 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 11 October 2017	Date of mailing of the international search report 19 October 2017
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 ZOU, Yuting Telephone No. (86-10) 62414037

International application No.
PCT/CN2017/093803

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

国际检索报告

国际申请号

PCT/CN2017/093803

A. 主题的分类 G06Q 10/04 (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 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: 信用卡, 违约, 预测, 样本, 归一化, 神经网络, 弱分类器, 强分类器, 预测序列, 权重, 迭代, 调整, BP, adaboost, sample, prediction, sequence, weight, adjust, classifier																				
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 106529729 A (同济大学) 2017年 3月 22日 (2017 - 03 - 22) 权利要求1-10</td> <td>1-10</td> </tr> <tr> <td>X</td> <td>王娜. “基于BP-Adaboost的上市公司财务危机预警实证研究” 中国集体经济, 第1期, 2015年 1月 31日 (2015 - 01 - 31), ISSN: 1008-1283, 第144~146页</td> <td>1-10</td> </tr> <tr> <td>X</td> <td>陆正华 等. “基于BP-Adaboost算法的上市公司财务预警研究” 财会通讯, 第8期, 2013年 12月 31日 (2013 - 12 - 31), ISSN: 1002-8072, 第117~119页</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 104240720 A (北京大学深圳研究生院) 2014年 12月 24日 (2014 - 12 - 24) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>CN 106022229 A (北京航空航天大学) 2016年 10月 12日 (2016 - 10 - 12) 全文</td> <td>1-10</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 106529729 A (同济大学) 2017年 3月 22日 (2017 - 03 - 22) 权利要求1-10	1-10	X	王娜. “基于BP-Adaboost的上市公司财务危机预警实证研究” 中国集体经济, 第1期, 2015年 1月 31日 (2015 - 01 - 31), ISSN: 1008-1283, 第144~146页	1-10	X	陆正华 等. “基于BP-Adaboost算法的上市公司财务预警研究” 财会通讯, 第8期, 2013年 12月 31日 (2013 - 12 - 31), ISSN: 1002-8072, 第117~119页	1-10	A	CN 104240720 A (北京大学深圳研究生院) 2014年 12月 24日 (2014 - 12 - 24) 全文	1-10	A	CN 106022229 A (北京航空航天大学) 2016年 10月 12日 (2016 - 10 - 12) 全文	1-10
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																		
PX	CN 106529729 A (同济大学) 2017年 3月 22日 (2017 - 03 - 22) 权利要求1-10	1-10																		
X	王娜. “基于BP-Adaboost的上市公司财务危机预警实证研究” 中国集体经济, 第1期, 2015年 1月 31日 (2015 - 01 - 31), ISSN: 1008-1283, 第144~146页	1-10																		
X	陆正华 等. “基于BP-Adaboost算法的上市公司财务预警研究” 财会通讯, 第8期, 2013年 12月 31日 (2013 - 12 - 31), ISSN: 1002-8072, 第117~119页	1-10																		
A	CN 104240720 A (北京大学深圳研究生院) 2014年 12月 24日 (2014 - 12 - 24) 全文	1-10																		
A	CN 106022229 A (北京航空航天大学) 2016年 10月 12日 (2016 - 10 - 12) 全文	1-10																		
☐ 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 2017年 10月 11日		国际检索报告邮寄日期 2017年 10月 19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 邹予婷 电话号码 (86-10)62414037																		

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2017/093803

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	106529729	A	2017年 3月 22日	无	
CN	104240720	A	2014年 12月 24日	无	
CN	106022229	A	2016年 10月 12日	无	