(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 10 月 24 日 (24.10.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/200742 A1

(51) 国际专利分类号:
G06K 9/62 (2006.01)

(21) 国际申请号: PCT/CN2018/095483

(22) 国际申请日: 2018 年 7 月 12 日 (12.07.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810345257.9 2018年4月17日 (17.04.2018) 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福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(72) 发明人: 王义文 (WANG, Yiwen); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 王健宗 (WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。 肖京 (XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路5033号平安金融中心23楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市明日今典知识产权代理事务所（普通合伙） (SHENZHEN MINGRIJINDIAN INTELLECTUAL PROPERTY AGENCY FIRM (GENERAL)); 中国广东省深圳市南山区南头街道智恒新兴产业园E区01B栋405室, Guangdong 518000 (CN)。

(81) 指定国(除另有指明，要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, 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,

(54) **Title:** SHORT-TERM PROFIT PREDICTION METHOD, APPARATUS, COMPUTER DEVICE, AND STORAGE MEDIUM

(54) **发明名称**: 短期盈利的预测方法、装置、计算机设备和存储介质

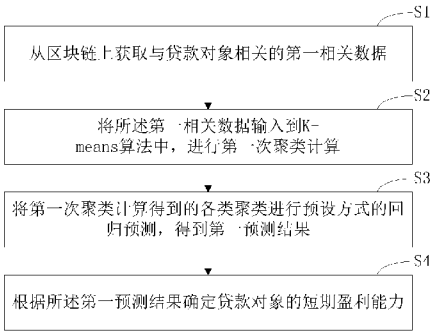

图 1

(57) **Abstract:** The present application discloses a short-term profit prediction method, apparatus, computer device, and storage medium. The prediction method comprises: obtaining first relevant data related to a loan borrower from a blockchain; inputting the first relevant data into a *k*-means algorithm to perform first clustering; performing a regression prediction task in a preset manner on various types of clusters obtained from the first clustering so as to obtain a first prediction result; and according to the first prediction result, determining the ability of the loan borrower to make short-term profit.

(57) **摘要**: 本申请揭示了一种短期盈利的预测方法、装置、计算机设备和存储介质，其中预测方法，包括：从区块链上获取与贷款对象相关的第一相关数据；将第一相关数据输入到K-means算法中，进行第一次聚类计算；将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果；根据第一预测结果确定贷款对象的短期盈利能力。

WO 2019/200742 A1

[见续页]

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))。

短期盈利的预测方法、装置、计算机设备和存储介质

本申请要求于2018年4月17日提交中国专利局、申请号为2018103452579，申请名称为“短期盈利的预测方法、装置、计算机设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及到互联网技术领域，特别是涉及到一种短期盈利的预测方法、装置、计算机设备和存储介质。

背景技术

区块链是一种去中心化、无需信任的新型数据架构，它由网络中所有的节点共同拥有、管理和监督，不接受单一方面的控制。由于区块链是一种新型的数据架构，所以在区块链布局的前期数据量较少，银行等金融机构很难通过目前的“小数据”完成短期盈利预测，从而存在无法发放合适的贷款额度等问题。

技术问题

本申请的主要目的为提供一种在区块链布局前期企业相关数据量少的情况下，对企业的短期盈利的预测方法、装置、计算机设备和存储介质。

技术解决方案

本申请提出一种短期盈利的预测方法，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测方法，包括：从区块链上获取与贷款对象相关的第一相关数据；

将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算；

将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果；

根据所述第一预测结果确定贷款对象的短期盈利能力。

本申请还提供一种短期盈利的预测装置，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测装置，包括：

获取单元，用于从区块链上获取与贷款对象相关的第一相关数据；

聚类单元，用于将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算；

回归单元，用于将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果；

确定单元，用于根据所述第一预测结果确定贷款对象的短期盈利能力。

本申请还提供一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现上述任一项所述方法的步骤。

本申请还提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现上述任一项所述的方法的步骤。

有益效果

本申请的短期盈利的预测方法、装置、计算机设备和存储介质，先对获取到的少量数据通过K-means

算法进行聚类，然后通过回归算法进行预测得到预测结果，最后根据预测结果确定贷款对象的短期盈利能力。解决了银行等金融机构在各企业数据链前期布局阶段相关数据较少的情况下，无法准确预测贷款企业的短期盈利能力的问题，便于相对准确地限定贷款对象的贷款额度，以减小银行机构的借贷风险。

附图说明

图1 为本发明一实施例的短期盈利的预测方法的流程示意图；

图2 为本发明一实施例的短期盈利的预测方法的流程示意图；

图3 为本发明一实施例的短期盈利的预测装置的结构示意框图；

图4 为本发明一实施例的回归单元的结构示意框图；

图5 为本发明一实施例的聚类单元的结构示意框图；

图6 为本发明一实施例的短期盈利的预测装置的结构示意框图；

图7 为本发明一实施例的计算机设备的结构示意框图。

本发明的最佳实施方式

参照图 1，本申请提供一种短期盈利的预测方法，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用。

本申请中，银行等金融机构流动资金贷款一般分为临时贷款、短期贷款和中期贷款，其中短期贷款期限一般为三个月至一年（不含三个月含一年）的流动资金贷款。因为市场变化反复无常，利用历史数据提炼出的规律在一定时间内可能是正确的，但是过一段时间后其正确的概率降低。按预测时间范围长短不同，可将其分为短期预测、中期预测和长期预测三种。一般地，预测时间范围越短，预测质量越高；反之，预测结果的准确性越低。本申请中，区块链上数据量小于预设量是一个限定条件，主要限定本方法针对各企业在数据链布局的前期，各种数据相对较少情况下使用，本申请中“小于预设量的数据量”相对目前“大数据”而言，可以称之为“小数据”。

上述预测方法，包括步骤:

S1、从区块链上获取与贷款对象相关的第一相关数据;

S2、将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

S3、将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

S4、根据所述第一预测结果确定贷款对象的短期盈利能力。

如上述步骤 S1 所述，上述贷款对象为需要到银行等金融机构贷款的企业或个人。上述第一相关数据可以是在区块链上与贷款对象相关的全部数据，也可以根据指定要求检索到的数据，比如根据不同的企业或项目，获取区块链上不同的数据，比如采购代理融资企业，其可以获取金融机构区块数据、核心企业区块数据、仓储物流区块数据、经销商区块数据等。

如上述步骤 S2 所述，上述 K-means 算法是一种输入聚类个数 k，以及包含 n 个数据对象的数据库，

输出满足方差最小标准 k 个聚类的一种算法。k-means 算法接受输入量 k ；然后将 n 个数据对象划分为 k 个聚类以便使得所获得的聚类满足：同一聚类中的对象相似度较高；而不同聚类中的对象相似度较小。其原理为：先初设几个中心的位置，计算所有点到这几个中心的距离，然后找出属于这几个中心的点，比如 A 点与 1 号中心距离最近就属于 1 号。将所有属于 1 号的点求平均就得到新的中心点。不断迭代直到属于每个中心的中心点不变，得到最后的中心位置，以完成数据的聚类。

本申请请中，上述步骤 S2 的具体过程如下：

S21、对于给定的一个包含 n 个 d 维数据点的相关数据的数据集（第一相关数据）$X=\{x_1,x_2,\cdots,x_n\}$，其中，$x_i \in R^d$，选择数据集中 K 个点作为初始聚类中心，每个对象代表一个类别的中心 $\mu_k(k=1,2,\cdots,K)$。

S22、计算各点到中心 μ_k 的欧氏距离，按距离最近的准则分别将它们分配给与其最相似的聚类中心代表的类，形成 K 个簇 $C=\{c_k,k=1,2,\cdots,k\}$。每个簇 c_k 代表一个类。计算该类各点到聚类中心 μ_k 的距离平方和 $J(c_k)$：

$$J(c_k)=\sum_{x_i \in c_k}||x_i-\mu_k||^2$$

S23、计算各类样本到其所在类别聚类中心 μk 总的距离平方和 ，直至最小：

$$J(C)=\sum_{k=1}^{K}J(c_k)=\sum_{k=1}^{K}\sum_{x_i \in c_k}||x_i-\mu_k||^2=\sum_{k=1}^{K}\sum_{i=1}^{n}d_{ki}||x_i-\mu_k||^2$$

式中：若 $x_i \in c_k$，$d_{ki}=1$；$x_i \notin c_k$，$d_{ki}=0$，则计算类内所有对象的均值作为该类的新聚类中心。

S24、判断聚类中心和值是否发生改变，若发生改变则转回步骤 S22，若不再改变则聚类结束。

本申请使用 K-means 算法进行数据聚类，简单、快速，算法保持可伸缩性和高效性，当簇接近高斯分布时，效果更佳。

如上述步骤 S3 所述，上述回归预测就是把预测的相关性原则作为基础，把影响预测目标的各因素找出来，然后找出这些因素和预测目标之间的函数关系的近似表达，并且用数学的方法找出来。上述第一预测结果即为将第一次聚类计算得到的各类聚类通过预设方式的回归预测计算得出的结果，又因为上述第一相关数据是贷款对象的相关数据，所以第一预测结果在一定程度上可以反映贷款对象在短期内的盈利能力。回归预测的基本步骤如下：1、根据预测目标，确定自变量和因变量。具体地，明确预测的具体目标，也就确定了因变量。如预测具体目标是下一年度的销售量，那么销售量 Y 就是因变量。通过市场调查和查阅资料，寻找与预测目标的相关影响因素，即自变量，并从中选出主要的影响因素。2、建立回归预测模型。具体地，依据自变量和因变量的历史统计资料进行计算，在此基础上建立回归分析方程，即回归预测模型。3、进行相关分析。具体地，回归分析是对具有因果关系的影响因素(自变量)和预测对象(因变量)所进行的数理统计分析处理。只有当变量与因变量确实存在某种关系时，建立的回归方程才有意义。因此，作为自变量的因素与作为因变量的预测对象是否有关，相关程度如何，以及判断这种相关程度的把握性多大，就成为进行回归分析必须要解决的问题。进行相关分析，一般要求出相

关关系，以相关系数的大小来判断自变量和因变量的相关的程度。4、检验回归预测模型，计算预测误差。具体地，回归预测模型是否可用于实际预测，取决于对回归预测模型的检验和对预测误差的计算。回归方程只有通过各种检验，且预测误差较小，才能将回归方程作为预测模型进行预测。5、计算并确定预测值。具体地，利用回归预测模型计算预测值，并对预测值进行综合分析，确定最后的预测值。本申请中，先对数据进行聚类，然后在对聚类后数据进行回归预测，预测速度更快。

如上述步骤 S4 所述，即为根据第一预测结果确定贷款对象的短期盈利能力。然后银行等金融机构既可以根据其盈利能力确定上述贷款对象的贷款额度，即可以给上述贷款对象的贷款金额上限。上述第一预测结果可以是代表等级的数字，比如，分为 1-10 级，随着等级的提高，代表贷款对象的短期盈利能力越强，其贷款的额度也就相应的越高，本实施例中，贷款额度还与贷款对象的注册资金、市场价值等数据相关。

本实施例中，上述将第一次聚类计算得到的各类聚类进行预设方式的回归预测的步骤 S3，包括:

S31、将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

如上述步骤 S31 所述，上述 SVR（Support Vector Regression，支持向量回归），是支持向量机(SVM)的重要的应用分支。本实施例中，通过极小化目标函数来确定回归函数，回归函数为 f(x)=wx+b。其具体过程为:

$$\min_{w,b,\zeta,\zeta^*,\varepsilon} \frac{1}{2}w^T w + C(v\varepsilon + \frac{1}{l}\sum_{i=1}(\zeta_i + \zeta_i^*))$$

限制条件为: $(w^T\Phi(x_i) + b)\text{-}\, c \le \varepsilon + \zeta_i$

$$z_i - (w^T\Phi(x_i) + b) \le \varepsilon + \zeta_i^*$$

$$\zeta_i,\ \zeta_i^* \ge 0,\ i = 1, \ldots, i$$

对偶问题为:

$$\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + z^T(\alpha - \alpha^*)$$

限制条件为: $e^T(\alpha - \alpha^*) = 0, e^T(\alpha + \alpha^*) \le Cv$

$$0 \le \alpha_i,\ \alpha_i^* \le C/l, i = 1, \ldots, i$$

近似函数为:

$$\sum_{i=1}^{l}(-\alpha_i + \alpha_i^*)K(x_i, x) + b$$

类似于 2002 年提出的 v-SVC,$e^T(\alpha + \alpha^*) \le Cv$不等式可以由等式进行替换。而且,由于用户经常选择 C =1 类似的小常量，导致C/l太小。因此，在 LIBSVM（是台湾大学林智仁(Lin Chih-Jen)教授等开发设计的一个简单、易于使用和快速有效的 SVM 模式识别与回归的软件包） 中,将用户指定的参数作为C/l. 即，$\bar{c}$= C/l是用户指定的, LIBSVM 解决了以下问题:

$$\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha - \alpha^*)^T Q(\alpha - \alpha^*) + z^T(\alpha - \alpha^*)$$

限制条件为： $e^T(\alpha-\alpha^*)=0, e^T(\alpha+\alpha^*)\le \overline{c}/v$

$$0\le \alpha_i,\ \alpha_i^*\le \overline{c}, i=1,\dots,l$$

$\varepsilon-SVR$在参数 $(\overline{c},\varepsilon)$ 下，其与$v-SVR$ 在参数$(l\overline{c},v)$ 下具有相同的解。

上式中，l 为训练样本个数，这里 $l=k$；C 为平衡模型复杂性 $(1/2)w^Tw$ 和训练误差项的权重参数；ε为不敏感损失函数；ζ为松弛因子。$K(x_i,x)$为核函数。

上述 SVR（支持向量回归算法）主要是通过将聚类结果升维后，在高维空间中构造线性决策函数来实现线性回归，用 e 不敏感损失函数时，其基础主要是 e 不敏感损失函数和核函数算法。若将拟合的数学模型表达多维空间的某一曲线，则根据 e 不敏感损失函数所得的结果，就是包括该曲线和训练点的“e 管道”。在所有样本点中，只有分布在“管壁”上的那一部分样本点决定管道的位置。这一部分训练样本称为“支持向量”。为适应训练样本集的非线性，传统的拟合方法通常是在线性方程后面加高阶项。此法诚然有效，但由此增加的可调参数未免增加了过拟合的风险。SVR 采用核函数解决这一矛盾。用核函数代替线性方程中的线性项可以使原来的线性算法“非线性化”，即能做非线性回归。与此同时，引进核函数达到了“升维”的目的，而增加的可调参数是过拟合依然能控制。本申请，利用技术成熟的 SVR 算法，计算结果可靠，而且可以达到准确预测的效果。

在一个实施例中，上述将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算的步骤 S2，包括：

S21、将所述第一相关数据进行特征提取；

S22、将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据；

S23、将所述第一相关数据中与所述不相关特征数据对应的第一相关数据清除后输入到 K-means 算法中，进行第一次聚类计算。

如上述步骤 S201 至 S203 所述，将上述贷款对象相关的第一相关数据进行特征提取，进行相关性分析找出特征数据中与其它特征数据不相关的不相关特征数据，然后将这些不相关特征数据对应的第一相关数据从第一相关数据中剔除，使用留下的第一相关数据进行聚类计算，得到的聚类更加准确，因为将不相关特征数据对应的第一相关数据提出，所以提高聚类计算的效率。

本实施例中，对第一相关数据进行特征能提取的方法具体为：使用 Relief 算法（Relief 算法是一种特征权重算法(Feature weighting algorithms)，根据各个特征和类别的相关性赋予特征不同的权重，权重小于某个阈值的特征将被移除）进行特征提取。Relief 算法从训练集 D 中随机选择一个样本 R，然后从和 R 同类的样本中寻找最近邻样本 H，称为 Near Hit，从和 R 不同类的样本中寻找最近邻样本 M，称为 NearMiss，然后根据以下规则更新每个特征的权重：如果 R 和 Near Hit 在某个特征上的距离小于 R 和 Near Miss 上的距离，则说明该特征对区分同类和不同类的最近邻是有益的，则增加该特征的权重；反之，如果 R 和 Near Hit 在某个特征的距离大于 R 和 Near Miss 上的距离，说明该特征对区分同类和不

同类的最近邻起负面作用，则降低该特征的权重。以上过程重复 m 次，最后得到各特征的平均权重。特征的权重越大，表示该特征的分类能力越强，反之，表示该特征分类能力越弱。Relief 算法的运行时间随着样本的抽样次数 m 和原始特征个数 N 的增加线性增加，因而运行效率非常高。具体算法如下所示:

设训练数据集为 D，样本抽样次数 m，特征权重的阈值 δ，最近邻样本个数输出为各个特性的特征权重 T:

1、置所有特征权重为 0，*T*为空集。

2、*for i=1 to m do*

1）、随机选择一个样本 *R*;

2）、从同类样本集中找到 *R* 的最近邻 *H*，从不同类样本集中找最近邻样本 *M*。

3）、*for A=1 to N do*

$$W(A) = W(A) - diff(A,R,H)/m + diff(A,R,M)/m$$

3、*for A=1 to N do*

if $W(A) \geq \delta$

把第 *A* 个特征添加到 *T* 中。

在一个实施例中，上述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤 S202，包括:

S2021、将所述特征数据制作成散点图，将所述散点图中的离散点对应的特征数据记为所述不相关特征数据。

如上述步骤 S2021 所述，上述散点图（scatter diagram）在回归分析中是指数据点在直角坐标系平面上的分布图；通常用于比较跨类别的聚合数据。散点图中包含的数据越多，比较的效果就越好。本实施例中上述特征数据一般为矩阵，此时可利用散点图矩阵来同时绘制各自变量间的散点图，这样可以快速发现多个变量间的主要相关性。将上述特征数据制作成散点图的过程即为可视化的过程，特征数据可视化处理，所以人可以个通过肉眼在图形或图像上直观的分辨出离散点的存在，然后选择出离散点，计算机设备会将选择的离散点对应的特征数据记为不相关特征数据。

在另一实施例中，上述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤 S202，包括:

S2022、将所述特征数据进行相关矩阵分析，提取出与其它特征数据不相关的所述不相关特征数据。

如上述步骤 S2022 所述，上述相关矩阵也叫相关系数矩阵，其是由矩阵各列间的相关系数构成的。也就是说，相关矩阵第 i 行第 j 列的元素是原矩阵第 i 列和第 j 列的相关系数。本实施例中一般用到协方差矩阵进行分析，协方差用来衡量两个变量的总体误差，如果两个变量的变化趋势一致，协方差就是正

值，说明两个变量正相关。如果两个变量的变化趋势相反，协方差就是负值，说明两个变量负相关。如果两个变量相互独立，那么协方差就是 0，说明两个变量不相关，当变量大于或等于三组的时候，即会使用相应的协方差矩阵。

参照图 2，在本实施例中，上述根据所述第一预测结果确定贷款对象的短期盈利能力的步骤 S4 之后，包括：

S5、获取非区块链上的与所述贷款对象相关的第二相关数据；

S6、将所述第二相关数据输入到 K-means 算法中，进行第二次聚类计算；

S7、将第二次聚类计算得到的各类聚类进行预设方式的回归预测，得到第二预测结果；

S8、判断所述第一预测结果与所述第二预测结果的差值是否小于预设的阈值；

S9、若所述差值小于所述阈值，则判定根据所述第一预测结果确定贷款对象的短期盈利能力的结果为可用结果。

如上述步骤 S5 至 S9 所述，上述非区块链上的第二相关数据，是指没有记录在区块链上的数据，一般为大数据网络中数据。对第二相关数据的聚类算法和回归预测方法与上述的第一相关数据完全相同，再此不在赘述。本实施例中，将根据第一相关数据得到的第一预测结果与根据第二相关数据得到的第二预测结果进行比较，即为设置一道验证的步骤，以判断第一预测结果是否可用。本申请中，因为主要是针对区块链布局的前期，所以各企业的历史数据会有大量的存在与大数据的互联网上，如企业自己的服务器中，或则与企业相关的其它企业的服务器中，只要在互联网环境中，就有可能被获取到。本步骤中，主要将利用互联网上的“大数据”得到的第二预测结果验证利用区块链上的“小数据”得到的第一预测结果，只有第二预测结果和第一预测结果的差值小于预设的阈值才判定第一预测结果基本正确，可以使用。

在一个实施例中，上述将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算的步骤 S2 之前，包括：

S201、判断所述第一相关数据的数据量是否大于预设的数据阈值；

S202、若是，则将所述第一相关数据输入到预设的基于大数据的预测算法中进行预测。

如上述步骤 S201 和 S202 所述，就是设定了一个数据阈值，当获取到的第一相关数据的数据量大于数据阈值时，其已经脱离了上述短期盈利的预测方法适用的“小数据”范围，所以会停止后续的聚类、回归预测等步骤，而是切换预测方法。具体切换的方法可以是，将获取到的第一相关数据输入到预设的现有的相对成熟的预测模型中，比如基于 TD-ABC 模型的企业盈利模型等。

在一个实施例中，还可以分析上述的第一相关数据中是否含有欺诈数据，具体的方法可以为：将获取的第一相关数据进行特征提取，以得到特征数据；在所述特征数据中提取出与其它特征数据不相关的不相关特征数据；然后通过 Voronoi 算法对所述不相关特征数据进行异常值识别，得出欺诈数据。可以

通过欺诈数据的多少等情况，分析出贷款对象的借贷信誉值。然后结合信誉值和短期盈利能力确定贷款对象的贷款额度。

在一具体实施例中，a 企业需要找 P 银行进行贷款，P 银行则需要对 a 企业进行评估，其评估的过程为：1、通过在区块链上收集与该 a 企业相关的全部数据，如 a 企业的销售数据、生产数据、财务数据等。然后对获取到的数据进行特征提取，将无用的数据提前删除，已提高后续聚类计算的速度与效率。具体的删除方法为，先将提取出的数据进行可视化地形成散点图，然后将散点图中的离散点删除。2、将从区块链上获取到的 a 企业的数据通过 K-means 算法进行聚类计算。3、将聚类计算的结果进行 SVR 回归预测，进而得到该 a 企业盈利能力等结果；4、还通过上述欺诈数据的识别方法判断 a 企业的信誉等；5、P 银行根据 a 企业的信誉、盈利能力等确定是否可以贷款给 a 企业，以及最大贷款限额等。具体的，如果 a 企业的信誉小于预设值，则拒绝贷款给 a 企业；如果 a 企业的信誉为预设值则可以到款给 a 企业，此时在结合该 a 企业的盈利能力，计算最大的贷款限额等，从而有效地提高 P 银行规避风险的能力。具体获取 a 企业在数据链上的数据包括：采购货物种类，以及该采购经费数据；海关出口货物、关税，进口货物、关税；国内销售数据；销售产品数据；贷款数据；还贷信誉数据；货物库存数据；物流相关数据（仓库数量、仓库地理分布、每个仓库的存储数据、销售地域分布）等。

本申请的短期盈利的预测方法，先对获取到的“小数据”据通过 K-means 算法进行聚类，然后通过回归算法进行预测得到预测结果，最后根据预测结果确定贷款对象的短期盈利能力。解决了银行等金融机构在各企业数据链前期布局阶段相关数据较少的情况下，无法准确预测贷款企业的短期盈利能力的问题，便于相对准确地限定贷款对象的贷款额度，以减小银行机构的借贷风险。

参照图 3，本申请实施例还提供一种短期盈利的预测装置，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用。

本申请中，银行等金融机构流动资金贷款一般分为临时贷款、短期贷款和中期贷款，其中短期贷款期限一般为三个月至一年（不含三个月含一年）的流动资金贷款。因为市场变化反复无常，利用历史数据提炼出的规律在一定时间内可能是正确的，但是过一段时间后其正确的概率降低。按预测时间范围长短不同，可将其分为短期预测、中期预测和长期预测三种。一般地，预测时间范围越短，预测质量越高；反之，预测结果的准确性越低。本申请中，区块链上数据量小于预设量是一个限定条件，主要限定本方法针对各企业在数据链布局的前期，各种数据相对较少情况下使用，本申请中“小于预设量的数据量”相对目前“大数据”而言，可以称之为“小数据”。

上述预测装置，包括:

获取单元 10，用于从区块链上获取与贷款对象相关的第一相关数据;

聚类单元 20，用于将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

回归单元 30，用于将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

确定单元40，用于根据所述第一预测结果确定贷款对象的短期盈利能力.

在上述获取单元10中，上述贷款对象为需要到银行等金融机构贷款的企业或个人。上述第一相关数据可以是在区块链上与贷款对象相关的全部数据，也可以根据指定要求检索到的数据，比如根据不同的企业或项目，获取区块链上不同的数据，比如采购代理融资企业，其可以获取金融机构区块数据、核心企业区块数据、仓储物流区块数据、经销商区块数据等。

在上述聚类单元 20 中，上述 K-means 算法是一种输入聚类个数 k，以及包含 n 个数据对象的数据库，输出满足方差最小标准 k 个聚类的一种算法。k-means 算法接受输入量 k ；然后将 n 个数据对象划分为 k 个聚类以便使得所获得的聚类满足：同一聚类中的对象相似度较高；而不同聚类中的对象相似度较小。其原理为：先初设几个中心的位置，计算所有点到这几个中心的距离，然后找出属于这几个中心的点，比如 A 点与 1 号中心距离最近就属于 1 号。将所有属于 1 号的点求平均就得到新的中心点。不断迭代直到属于每个中心的中心点不变，得到最后的中心位置，以完成数据的聚类。

本申请请中，上述的聚类单元 20 的具体聚类过程如下：

（1）、对于给定的一个包含 n 个 d 维数据点的相关数据的数据集（第一相关数据）$X=\{x_1,x_2,\cdots,x_n\}$，其中，$x_i \in R^d$，选择数据集中 K 个点作为初始聚类中心，每个对象代表一个类别的中心 $\mu_k(k=1,2,\cdots,K)$。

（2）、计算各点到中心 μ_k 的欧氏距离，按距离最近的准则分别将它们分配给与其最相似的聚类中心代表的类，形成 K 个簇 $C=\{c_k,k=1,2,\cdots,k\}$。每个簇 c_k 代表一个类。计算该类各点到聚类中心 μ_k 的距离平方和 $J(c_k)$:

$$J(c_k)=\sum_{x_i \in c_k}||x_i-\mu_k||^2$$

（3）、计算各类样本到其所在类别聚类中心 μk 总的距离平方和 ，直至最小：

$$J(C)=\sum_{k=1}^{K}J(c_k)=\sum_{k=1}^{K}\sum_{x_i \in c_k}||x_i-\mu_k||^2=\sum_{k=1}^{K}\sum_{i=1}^{n}d_{ki}||x_i-\mu_k||^2$$

式中：若 $x_i \in c_k$，$d_{ki}=1$；$x_i \notin c_k$，$d_{ki}=0$，则计算类内所有对象的均值作为该类的新聚类中心。

（4）、判断聚类中心和值是否发生改变，若发生改变则转回步骤 S22，若不再改变则聚类结束。

本申请使用K-means算法进行数据聚类，简单、快速，算法保持可伸缩性和高效性，当簇接近高斯分布时，效果更佳

在上述回归单元30中，上述回归预测就是把预测的相关性原则作为基础，把影响预测目标的各因素找出来，然后找出这些因素和预测目标之间的函数关系的近似表达，并且用数学的方法找出来。上述第一预测结果即为将第一次聚类计算得到的各类聚类通过预设方式的回归预测计算得出的结果，又因为上述第一相关数据是贷款对象的相关数据，所以第一预测结果在一定程度上可以反映贷款对象在短期内的盈利能力。回归预测的基本步骤如下：（1）根据预测目标，确定自变量和因变量。具体地，明确预测的具体目标，也就确定了因变量。如预测具体目标是下一年度的销售量，那么销售量Y就是因变量。通

过市场调查和查阅资料，寻找与预测目标的相关影响因素，即自变量，并从中选出主要的影响因素。（2）建立回归预测模型。具体地，依据自变量和因变量的历史统计资料进行计算，在此基础上建立回归分析方程，即回归预测模型。（3）进行相关分析。具体地，回归分析是对具有因果关系的影响因素(自变量)和预测对象(因变量)所进行的数理统计分析处理。只有当变量与因变量确实存在某种关系时，建立的回归方程才有意义。因此，作为自变量的因素与作为因变量的预测对象是否有关，相关程度如何，以及判断这种相关程度的把握性多大，就成为进行回归分析必须要解决的问题。进行相关分析，一般要求出相关关系，以相关系数的大小来判断自变量和因变量的相关的程度。（4）检验回归预测模型，计算预测误差。具体地，回归预测模型是否可用于实际预测，取决于对回归预测模型的检验和对预测误差的计算。回归方程只有通过各种检验，且预测误差较小，才能将回归方程作为预测模型进行预测。（5）计算并确定预测值。具体地，利用回归预测模型计算预测值，并对预测值进行综合分析，确定最后的预测值。本申请中，先对数据进行聚类，然后在对聚类后数据进行回归预测，预测速度更快。

在上述确定单元40中，即为用于根据第一预测结果确定贷款对象的短期盈利能力。然后银行等金融机构既可以根据其盈利能力确定上述贷款对象的贷款额度，即可以给上述贷款对象的贷款金额上限。上述第一预测结果可以是代表等级的数字，比如，分为1-10级，随着等级的提高，代表贷款对象的短期盈利能力越强，其贷款的额度也就相应的越高，本实施例中，贷款额度还与贷款对象的注册资金、市场价值等数据相关。

参照图 4，本实施例中，上述回归单元 30，包括:

SVR 预测模块 31，用于将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

在上述 SVR 预测模块 31 中，上述 SVR（Support Vector Regression，支持向量回归），是支持向量机(SVM)的重要的应用分支。本实施例中，通过极小化目标函数来确定回归函数，回归函数为 f(x)=wx+b。其具体过程为:

$$\min_{w,b,\zeta,\zeta^*,\varepsilon} \frac{1}{2}w^T w + C(v\varepsilon + \frac{1}{l}\sum_{i=1}(\zeta_i + \zeta_i^*))$$

限制条件为: $(w^T\Phi(x_i)+b)\text{-}\, c \le \varepsilon + \zeta_i$

$$z_i - (w^T\Phi(x_i)+b) \le \varepsilon + \zeta_i^*$$

$$\zeta_i,\ \zeta_i^* \ge 0,\ i = 1,\dots,i$$

对偶问题为:

$$\min_{\alpha,\alpha^*} \frac{1}{2}(\alpha-\alpha^*)^T Q(\alpha-\alpha^*) + z^T(\alpha-\alpha^*)$$

限制条件为: $e^T(\alpha-\alpha^*) = 0, e^T(\alpha+\alpha^*) \le Cv$

$$0 \le \alpha_i,\ \alpha_i^* \le C/l, i = 1,\dots,i$$

近似函数为:

$$\sum_{i=1}^{l}(-\alpha_i+\alpha_i^*)K(x_i,x)+b$$

类似于 2002 年提出的 v-SVC,$e^T(\alpha+\alpha^*)\le Cv$不等式可以由等式进行替换。而且,由于用户经常选择 C =1 类似的小常量，导致C/l太小。因此，在 LIBSVM 中,将用户指定的参数作为C/l. 即，$\overline{c}=C/l$是用户指定的,LIBSVM 解决了以下问题:

$$\min_{\alpha,\alpha^*}\frac{1}{2}(\alpha-\alpha^*)^TQ(\alpha-\alpha^*)+z^T(\alpha-\alpha^*)$$

限制条件为: $e^T(\alpha-\alpha^*)=0, e^T(\alpha+\alpha^*)\le\overline{c}/v$

$$0\le\alpha_i,\ \alpha_i^*\le\overline{c}, i=1,\dots,l$$

$\varepsilon-SVR$在参数 $(\overline{c},\varepsilon)$ 下，其与$v-SVR$ 在参数$(l\overline{c},v)$ 下具有相同的解。

上式中，l为训练样本个数，这里 l=k; C 为平衡模型复杂性 (1/2)w^Tw 和训练误差项的权重参数; ε为不敏感损失函数; ζ为松弛因子。$K(x_i,x)$为核函数。

上述 SVR（支持向量回归算法）主要是通过将聚类结果升维后，在高维空间中构造线性决策函数来实现线性回归，用 e 不敏感损失函数时，其基础主要是 e 不敏感损失函数和核函数算法。若将拟合的数学模型表达多维空间的某一曲线，则根据 e 不敏感损失函数所得的结果，就是包括该曲线和训练点的“e 管道”。在所有样本点中，只有分布在“管壁”上的那一部分样本点决定管道的位置。这一部分训练样本称为“支持向量”。为适应训练样本集的非线性，传统的拟合方法通常是在线性方程后面加高阶项。此法诚然有效，但由此增加的可调参数未免增加了过拟合的风险。SVR 采用核函数解决这一矛盾。用核函数代替线性方程中的线性项可以使原来的线性算法“非线性化”，即能做非线性回归。与此同时，引进核函数达到了“升维”的目的，而增加的可调参数是过拟合依然能控制。本申请，利用技术成熟的 SVR 算法，计算结果可靠，而且可以达到准确预测的效果。

参照图 5，在一个实施例中，上述聚类单元 20，包括:

提取模块 21，用于将所述第一相关数据进行特征提取;

分析模块 22，用于将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据;

聚类模块 23,用于将所述第一相关数据中与所述不相关特征数据对应的第一相关数据清除后输入到 K-means 算法中，进行第一次聚类计算。

在上述提取模块 21、分析模块 22 和聚类模块 23 中，将上述贷款对象相关的第一相关数据进行特征提取，进行相关性分析找出特征数据中与其它特征数据不相关的不相关特征数据，然后将这些不相关特征数据对应的第一相关数据从第一相关数据中剔除，使用留下的第一相关数据进行聚类计算，得到的聚类更加准确,因为将不相关特征数据对应的第一相关数据提出,所以提高聚类计算的效率。本实施例中，对第一相关数据进行特征能提取的方法具体为：使用 Relief 算法（Relief 算法是一种特征权重算法

(Feature weighting algorithms)，根据各个特征和类别的相关性赋予特征不同的权重，权重小于某个阈值的特征将被移除）进行特征提取。Relief 算法从训练集 D 中随机选择一个样本 R，然后从和 R 同类的样本中寻找最近邻样本 H，称为 Near Hit，从和 R 不同类的样本中寻找最近邻样本 M，称为 NearMiss，然后根据以下规则更新每个特征的权重：如果 R 和 Near Hit 在某个特征上的距离小于 R 和 Near Miss 上的距离，则说明该特征对区分同类和不同类的最近邻是有益的，则增加该特征的权重；反之，如果 R 和 Near Hit 在某个特征的距离大于 R 和 Near Miss 上的距离，说明该特征对区分同类和不同类的最近邻起负面作用，则降低该特征的权重。以上过程重复 m 次，最后得到各特征的平均权重。特征的权重越大，表示该特征的分类能力越强，反之，表示该特征分类能力越弱。Relief 算法的运行时间随着样本的抽样次数 m 和原始特征个数 N 的增加线性增加，因而运行效率非常高。具体算法已在方法实施例中描述，所以不再赘述。

在一个实施例中，上述分析模块 22，包括：可视分析子模块，用于将所述特征数据制作成散点图，将所述散点图中的离散点对应的特征数据记为所述不相关特征数据。

在上述可视分析子模块中，上述散点图（scatter diagram）在回归分析中是指数据点在直角坐标系平面上的分布图；通常用于比较跨类别的聚合数据。散点图中包含的数据越多，比较的效果就越好。本实施例中上述特征数据一般为矩阵，此时可利用散点图矩阵来同时绘制各自变量间的散点图，这样可以快速发现多个变量间的主要相关性。将上述特征数据制作成散点图的过程即为可视化的过程，特征数据可视化处理，所以人可以个通过肉眼在图形或图像上直观的分辨出离散点的存在，然后选择出离散点，计算机设备会将选择的离散点对应的特征数据记为不相关特征数据。

在另一实施例中，上述分析模块 22，包括：矩阵分析子模块，用于将所述特征数据进行相关矩阵分析，提取出与其它特征数据不相关的所述不相关特征数据。

在上述矩阵分析子模块中，上述相关矩阵也叫相关系数矩阵，其是由矩阵各列间的相关系数构成的。也就是说，相关矩阵第 i 行第 j 列的元素是原矩阵第 i 列和第 j 列的相关系数。本实施例中一般用到协方差矩阵进行分析，协方差用来衡量两个变量的总体误差，如果两个变量的变化趋势一致，协方差就是正值，说明两个变量正相关。如果两个变量的变化趋势相反，协方差就是负值，说明两个变量负相关。如果两个变量相互独立，那么协方差就是 0，说明两个变量不相关，当变量大于或等于三组的时候，即会使用相应的协方差矩阵。

参照图 6，在本实施例中，上述短期盈利的预测装置，还包括：

数据获取单元 50，用于获取非区块链上的与所述贷款对象相关的第二相关数据；

数据聚类单元 60，用于将所述第二相关数据输入到 K-means 算法中，进行第二次聚类计算；

聚类回归单元 70，用于将第二次聚类计算得到的各类聚类进行预设方式的回归预测，得到第二预测结果；

比较单元 80，用于判断所述第一预测结果与所述第二预测结果的差值是否小于预设的阈值；

判定单元 90，用于若所述差值小于所述阈值，则判定根据所述第一预测结果确定贷款对象的短期盈利能力的结果为可用结果。

上述非区块链上的第二相关数据，是指没有记录在区块链上的数据，一般为大数据网络中数据。对第二相关数据的聚类算法和回归预测方法与上述的第一相关数据完全相同，再此不在赘述。本实施例中，将根据第一相关数据得到的第一预测结果与根据第二相关数据得到的第二预测结果进行比较，即为设置一道验证的步骤，以判断第一预测结果是否可用。本申请中，因为主要是针对区块链布局的前期，所以各企业的历史数据会有大量的存在与大数据的互联网上，如企业自己的服务器中，或则与企业相关的其它企业的服务器中，只要在互联网环境中，就有可能被获取到。本步骤中，主要将利用互联网上的“大数据”得到的第二预测结果验证利用区块链上的“小数据”得到的第一预测结果，只有第二预测结果和第一预测结果的差值小于预设的阈值才判定第一预测结果基本正确，可以使用。

在一个实施例中，上述短期盈利的预测装置，还包括：

判断单元，用于判断所述第一相关数据的数据量是否大于预设的数据阈值；

切换单元、用于则将所述第一相关数据输入到预设的基于大数据的预测算法中进行预测。

如上述判断单元和切换单元中，就是设定了一个数据阈值，当获取到的第一相关数据的数据量大于数据阈值时，其已经脱离了上述短期盈利的预测装置的适用 的“小数据”范围，所以会停止后续的聚类、回归预测等预测过程，而是切换预测方法。具体切换的方法可以是，将获取到的第一相关数据输入到预设的现有的相对成熟的预测模型中，比如基于 TD-ABC 模型的企业盈利模型等。

在一个实施例中，上述短期盈利的预测装置还包括：

欺诈分析单元，用于分析上述的第一相关数据中是否含有欺诈数据，具体的方法可以为：将获取的第一相关数据进行特征提取，以得到特征数据；在所述特征数据中提取出与其它特征数据不相关的不相关特征数据；然后通过 Voronoi 算法对所述不相关特征数据进行异常值识别，得出欺诈数据。可以通过欺诈数据的多少等情况，分析出贷款对象的借贷信誉值。然后结合信誉值和短期盈利能力确定贷款对象的贷款额度。

在一具体实施例中，a 企业需要找 P 银行进行贷款，P 银行则需要对 a 企业进行评估，其评估的过程为：1、通过在区块链上收集与该 a 企业相关的全部数据，如 a 企业的销售数据、生产数据、财务数据等。然后对获取到的数据进行特征提取，将无用的数据提前删除，已提高后续聚类计算的速度与效率。具体的删除方法为，先将提取出的数据进行可视化地形成散点图，然后将散点图中的离散点删除。2、将从区块链上获取到的 a 企业的数据通过 K-means 算法进行聚类计算。3、将聚类计算的结果进行 SVR 回归预测，进而得到该 a 企业盈利能力等结果；4、还通过上述欺诈数据的识别方法判断 a 企业的信誉等；5、P 银行根据 a 企业的信誉、盈利能力等确定是否可以贷款给 a 企业，以及最大贷款限额等。具体

的，如果 a 企业的信誉小于预设值，则拒绝贷款给 a 企业；如果 a 企业的信誉为预设值则可以到款给 a 企业，此时在结合该 a 企业的盈利能力，计算最大的贷款限额等，从而有效地提高 P 银行规避风险的能力。具体获取 a 企业在数据链上的数据包括：采购货物种类，以及该采购经费数据；海关出口货物、关税，进口货物、关税；国内销售数据；销售产品数据；贷款数据；还贷信誉数据；货物库存数据；物流相关数据（仓库数量、仓库地理分布、每个仓库的存储数据、销售地域分布）等。

本申请的短期盈利的预测装置，先对获取到的“小数据”据通过K-means算法进行聚类，然后通过回归算法进行预测得到预测结果，最后根据预测结果确定贷款对象的短期盈利能力。解决了银行等金融机构在各企业数据链前期布局阶段相关数据较少的情况下，无法准确预测贷款企业的短期盈利能力的问题，便于相对准确地限定贷款对象的贷款额度，以减小银行机构的借贷风险。

参照图7，本发明实施例中还提供一种计算机设备，该计算机设备可以是服务器，其内部结构可以如图7所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。其中，该计算机设计的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储获取的第一相关数据和第二相关数据、K-means算法模型等数据。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现上述各方法实施例的流程。

本发明一实施例还提供一种计算机非易失性可读存储介质，其上存储有计算机可读指令，计算机可读指令被处理器执行时实现上述各方法实施例的流程。

以上所述仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

1、一种短期盈利的预测方法，其特征在于，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测方法，包括:

从区块链上获取与贷款对象相关的第一相关数据;

将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

根据所述第一预测结果确定贷款对象的短期盈利能力。

2、根据权利要求 1 所述的短期盈利的预测方法，其特征在于，所述将第一次聚类计算得到的各类聚类进行预设方式的回归预测的步骤，包括:

将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

3、根据权利要求 1 所述的短期盈利的预测方法，其特征在于，所述将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算的步骤，包括:

将所述第一相关数据进行特征提取;

将提取的特征数据进行相关性分析，得到与所述特征数据中的其它特征数据不相关的不相关特征数据;

将所述第一相关数据中与所述不相关特征数据对应的目标数据清除后输入到 K-means 算法中，进行第一次聚类计算。

4、根据权利要求 3 所述的短期盈利的预测方法，其特征在于，所述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤，包括:

将所述特征数据制作成散点图，将所述散点图中的离散点对应的特征数据记为所述不相关特征数据。

5、根据权利要求 3 所述的短期盈利的预测方法，其特征在于，所述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤，包括:

将所述特征数据进行相关矩阵分析，提取出与其它特征数据不相关的所述不相关特征数据。

6、根据权利要求 1 所述的短期盈利的预测方法，其特征在于，所述根据所述第一预测结果确定贷款对象的短期盈利能力的步骤之后，包括:

获取非区块链上的与所述贷款对象相关的第二相关数据;

将所述第二相关数据输入到 K-means 算法中，进行第二次聚类计算;

将第二次聚类计算得到的各类聚类进行预设方式的回归预测，得到第二预测结果;

判断所述第一预测结果与所述第二预测结果的差值是否小于预设的阈值;

若所述差值小于所述阈值，则判定根据所述第一预测结果确定贷款对象的短期盈利能力的结果为可用结果。

7、根据权利要求 1 所述的短期盈利的预测方法，其特征在于，所述将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算的步骤之前，包括:

判断所述第一相关数据的数据量是否大于预设的数据阈值;

若是，则将所述第一相关数据输入到预设的基于大数据的预测算法中进行预测。

8、一种短期盈利的预测装置，其特征在于，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测装置，包括:

获取单元，用于从区块链上获取与贷款对象相关的第一相关数据;

聚类单元，用于将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

回归单元，用于将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

确定单元，用于根据所述第一预测结果确定贷款对象的短期盈利能力。

9、根据权利要求 8 所述的短期盈利的预测装置，其特征在于，所述回归单元，包括:

SVR 预测模块，用于将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

10、根据权利要求 8 所述的短期盈利的预测装置，其特征在于，所述聚类单元，包括:

提取模块，用于将所述第一相关数据进行特征提取;

分析模块，用于将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据;

聚类模块，用于将所述第一相关数据中与所述不相关特征数据对应的第一相关数据清除后输入到 K-means 算法中，进行第一次聚类计算。

11、根据权利要求 10 所述的短期盈利的预测装置，其特征在于，所述分析模块，包括:

矩阵分析子模块，用于将所述特征数据进行相关矩阵分析，提取出与其它特征数据不相关的所述不相关特征数据。

12、根据权利要求 10 所述的短期盈利的预测装置，其特征在于，所述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤，包括:

将所述特征数据进行相关矩阵分析，提取出与其它特征数据不相关的所述不相关特征数据。

13、根据权利要求 8 所述的短期盈利的预测装置，其特征在于，所述短期盈利的预测装置，还包括:

数据获取单元，用于获取非区块链上的与所述贷款对象相关的第二相关数据;

数据聚类单元，用于将所述第二相关数据输入到 K-means 算法中，进行第二次聚类计算;

聚类回归单元，用于将第二次聚类计算得到的各类聚类进行预设方式的回归预测，得到第二预测结果;

比较单元，用于判断所述第一预测结果与所述第二预测结果的差值是否小于预设的阈值;

判定单元，用于若所述差值小于所述阈值，则判定根据所述第一预测结果确定贷款对象的短期盈利能力的结果为可用结果。

14、根据权利要求 8 所述的短期盈利的预测装置，其特征在于，所述短期盈利的预测装置，还包括:

判断单元，用于判断所述第一相关数据的数据量是否大于预设的数据阈值;

切换单元、用于则将所述第一相关数据输入到预设的基于大数据的预测算法中进行预测。

15、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现短期盈利的预测方法，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测方法，包括:

从区块链上获取与贷款对象相关的第一相关数据;

将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

根据所述第一预测结果确定贷款对象的短期盈利能力。

16、根据权利要求 15 所述的计算机设备，其特征在于，所述将第一次聚类计算得到的各类聚类进行预设方式的回归预测的步骤，包括:

将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

17、根据权利要求 15 所述的计算机设备，其特征在于，所述将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算的步骤，包括:

将所述第一相关数据进行特征提取;

将提取的特征数据进行相关性分析，得到与所述特征数据中的其它特征数据不相关的不相关特征数据;

将所述第一相关数据中与所述不相关特征数据对应的目标数据清除后输入到 K-means 算法中，进行第一次聚类计算。

18、根据权利要求 17 所述的计算机设备，其特征在于，所述将提取的特征数据进行相关性分析，得到与其它特征数据不相关的不相关特征数据的步骤，包括:

将所述特征数据制作成散点图，将所述散点图中的离散点对应的特征数据记为所述不相关特征数据。

19、一种计算机非易失性可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现短期盈利的预测方法，用于在区块链上获取到与贷款对象相关的数据量小于预设量时使用，所述预测方法，包括:

从区块链上获取与贷款对象相关的第一相关数据;

将所述第一相关数据输入到 K-means 算法中，进行第一次聚类计算;

将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果;

根据所述第一预测结果确定贷款对象的短期盈利能力。

20、根据权利要求 19 所述的计算机非易失性可读存储介质，其特征在于，所述将第一次聚类计算

得到的各类聚类进行预设方式的回归预测的步骤，包括：

将计算得到的各类聚类输入到预设的 SVR 预测模型中进行回归预测。

5

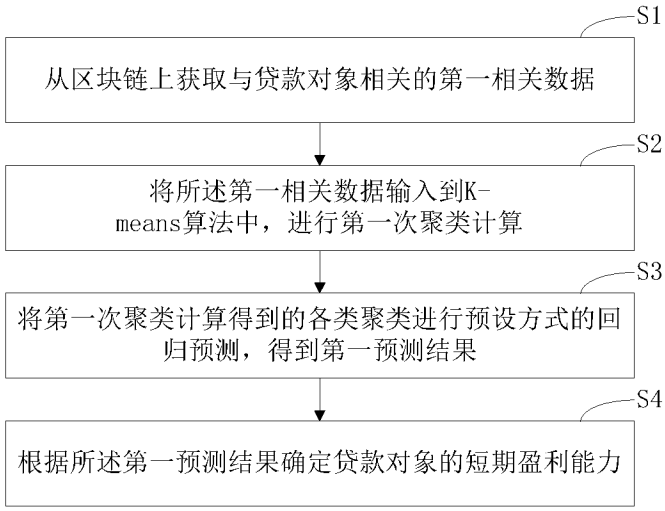
S1
从区块链上获取与贷款对象相关的第一相关数据
S2
将所述第一相关数据输入到K-means算法中，进行第一次聚类计算
S3
将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果
S4
根据所述第一预测结果确定贷款对象的短期盈利能力

图 1

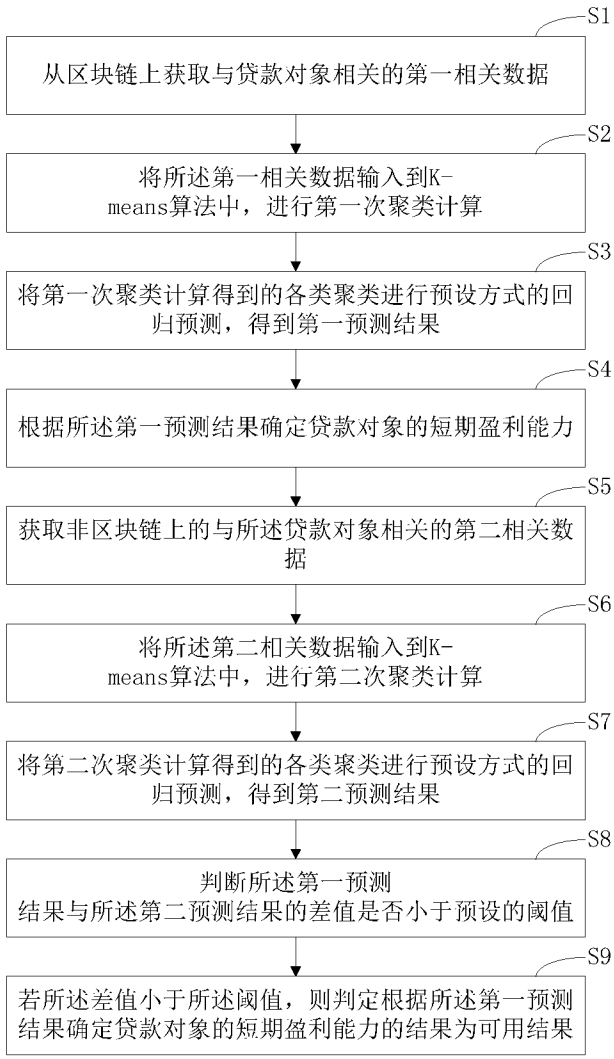
S1
从区块链上获取与贷款对象相关的第一相关数据
S2
将所述第一相关数据输入到K-means算法中，进行第一次聚类计算
S3
将第一次聚类计算得到的各类聚类进行预设方式的回归预测，得到第一预测结果
S4
根据所述第一预测结果确定贷款对象的短期盈利能力
S5
获取非区块链上的与所述贷款对象相关的第二相关数据
S6
将所述第二相关数据输入到K-means算法中，进行第二次聚类计算
S7
将第二次聚类计算得到的各类聚类进行预设方式的回归预测，得到第二预测结果
S8
判断所述第一预测结果与所述第二预测结果的差值是否小于预设的阈值
S9
若所述差值小于所述阈值，则判定根据所述第一预测结果确定贷款对象的短期盈利能力的结果为可用结果

图 2

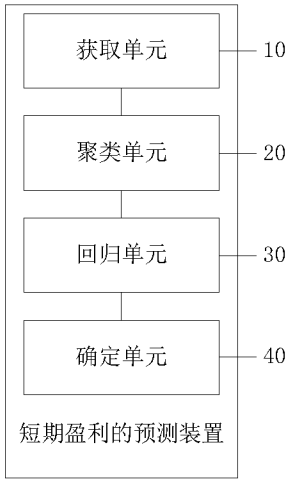
获取单元
10
聚类单元
20
回归单元
30
确定单元
40
短期盈利的预测装置

图 3

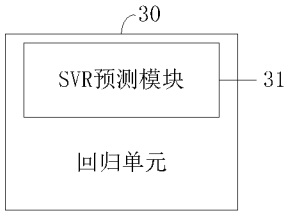
30
SVR预测模块
31
回归单元

图 4

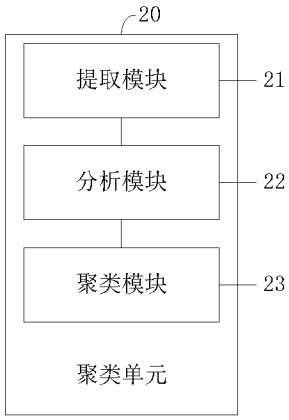
20
提取模块
21
分析模块
22
聚类模块
23
聚类单元

图 5

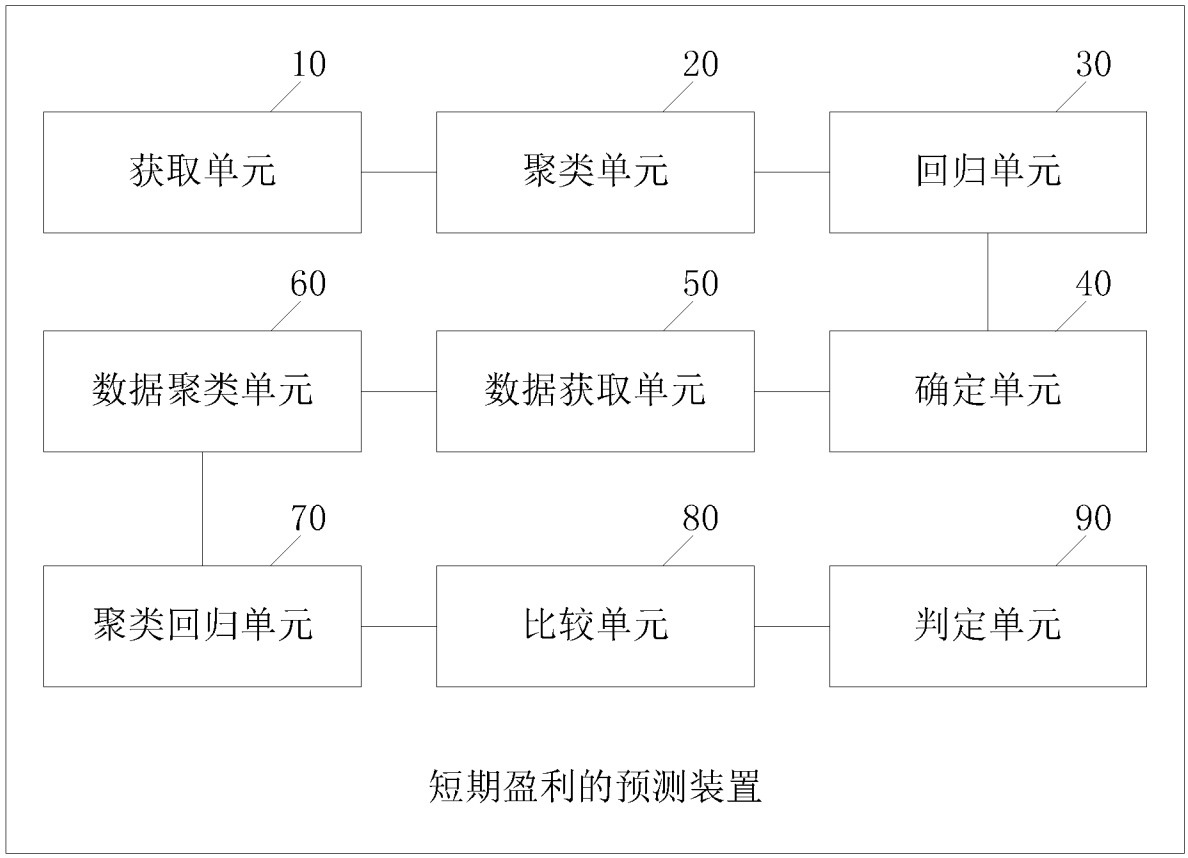
10
获取单元
20
聚类单元
30
回归单元
60
数据聚类单元
50
数据获取单元
40
确定单元
70
聚类回归单元
80
比较单元
90
判定单元
短期盈利的预测装置

图 6

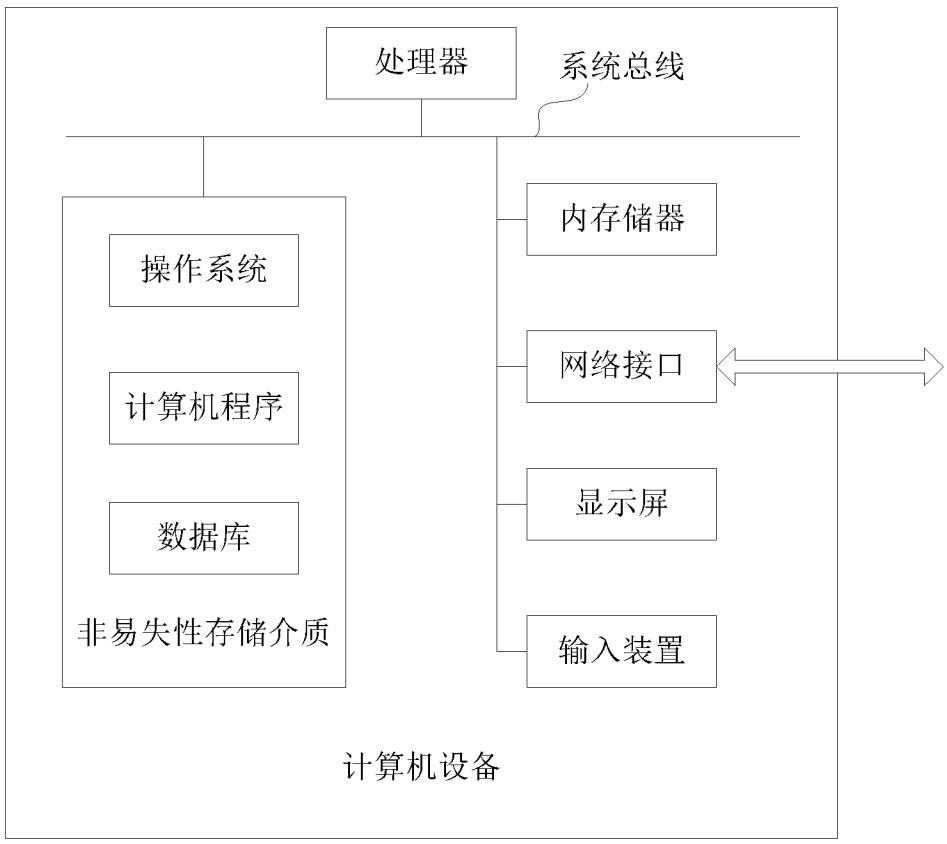
处理器
系统总线
操作系统
计算机程序
数据库
非易失性存储介质
内存储器
网络接口
显示屏
输入装置
计算机设备

图 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/095483

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/62(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K;G06F;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)

CNABS; SIPOABS; CNKI: 区块链, 盈利, 聚类, 回归, 贷款, 预测, block chain, profit, cluster, regression, loan, prediction

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 107844836 A (SUNYARD SYSTEM ENGINEERING CO., LTD.) 27 March 2018 (2018-03-27) entire document	1-20
A	CN 106127380 A (BEIJING TUOMING TECHNOLOGY CO., LTD.) 16 November 2016 (2016-11-16) entire document	1-20
A	CN 106980909 A (CHONGQING UNIVERSITY) 25 July 2017 (2017-07-25) entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 December 2018	**21 January 2019**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China **No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing** **100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/095483

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 107844836 A	27 March 2018	None	
CN 106127380 A	16 November 2016	None	
CN 106980909 A	25 July 2017	None	

国际检索报告

国际申请号
PCT/CN2018/095483

A. 主题的分类

G06K 9/62(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06K;G06F;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）)

CNABS;SIPOABS;CNKI:区块链，盈利，聚类，回归，贷款，预测，block chain, profit, cluster, regression, loan, prediction

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
A	CN 107844836 A (信雅达系统工程股份有限公司) 2018年 3月 27日 (2018-03-27) 全文	1-20
A	CN 106127380 A (北京拓明科技有限公司) 2016年 11月 16日 (2016-11-16) 全文	1-20
A	CN 106980909 A (重庆大学) 2017年 7月 25日 (2017-07-25) 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	国际检索报告邮寄日期
2018年 12月 18日	2019年 1月 21日

ISA/CN的名称和邮寄地址	受权官员
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	明媚
传真号 (86-10)62019451	电话号码 62411851

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/095483

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 107844836 A	2018年 3月 27日	无	
CN 106127380 A	2016年 11月 16日	无	
CN 106980909 A	2017年 7月 25日	无	