

Research on Credit Card Risk Detection and Control Based on SVDD

Peicai Cui¹, You Zhou², Fanrong Meng¹

¹College of Mathematics and System Science, Shandong University of Science and Technology, Qingdao Shandong

²College of Civil Engineering and Architecture, Shandong University of Science and Technology, Qingdao Shandong

Email: 1983947440@qq.com

Received: Jun. 22nd, 2018; accepted: Jul. 13th, 2018; published: Jul. 20th, 2018

Abstract

With the increase of credit card fraud transactions, the loss of banks and users caused by inadequate detection and early warning untimely is also increased. Firstly, by the TP and FP parameters of the fraud detection accuracy, our paper determined the algorithm training set and test set through probability estimation. Then, using the hypersphere as a criterion, the Lagrange function was used to train the radius and center of the optimal hypersphere, and the SVDD risk identification and evaluation model for credit card fraud consumption was established. Finally, Credit card fraud risk control probabilistic models and anti-fraud systems were proposed.

Keywords

SVDD Model, Hypersphere, Fraud Detection, Risk Management

基于SVDD的信用卡欺诈风险检测与管控研究

崔佩彩¹, 周游², 孟凡荣¹

¹山东科技大学, 数学与系统科学学院, 山东 青岛

²山东科技大学, 土木工程与建筑学院, 山东 青岛

Email: 1983947440@qq.com

收稿日期: 2018年6月22日; 录用日期: 2018年7月13日; 发布日期: 2018年7月20日

摘要

随着全球信用卡欺诈交易数量的与日俱增, 由信用卡欺诈交易检测和预警不力所带来的银行及用户损失也日益增加。首先, 本文通过判断欺诈检测准确性的TP和FP参数, 运用概率估计确定算法训练集和测试

集的容量。其次，以超球体为判断准则，运用拉格朗日函数训练得到最优超球面的半径和中心，建立信用卡欺诈消费的SVDD风险识别与评估模型。最后，提出信用卡欺诈的风险管控概率模型与反欺诈系统等处理措施。

关键词

SVDD模型，超球体，欺诈检测，风险管控

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着全球信用卡欺诈交易数量的与日俱增，数据分析服务成为信用卡欺诈检测和贷款投资风控工作的重要手段之一。现存的部分成功案例是银行等金融服务单位购买数据分析公司已经训练好的成熟模型，但这存在资金花销较大和技术垄断问题。部分数据分析公司试图找到一种折中道路：数据分析公司仍然作为第三方出现，但由于无法得到金融单位的原始数据，所以需要通过一些不完整的、经过数据预处理的或者质量不高的数据来训练模型。

本文以第四届“好贷杯”中国风险管理控制与控制能力挑战赛的赛题为例，参照赛题所提供的数据集，从以下几个方面对信用卡欺诈交易检测问题进行探讨：

问题一：建立信用卡欺诈检测模型。根据提供的预处理数据集，进行相关的数据分析和模型准备，建立信用卡欺诈的检测模型。但在识别模型建立过程中，需考虑“无标签”数据的处理。

问题二：信用卡欺诈风险的识别与评估。运用基于主成分分析后的支持向量机模型，根据核函数建立最优超平面，使在新信用卡交易数据产生时，对其进行判断与分类。

问题三：提出信用卡欺诈的对策与管控方案。通过分析检测模型的分类结果，对信用卡欺诈交易提出了相应的风险控制对策和建议，同时提出了风险管控的模型。

2. 风险识别

信用卡风险的识别是控制和管理业务处理过程中潜在风险的重要前提，只有准确的对信用卡风险的种类进行确定，分析风险来源，才能针对性的提出一套较为完善的信用卡管理与防范措施，对信用卡风险进行有效的控制与防范。

2.1. 信用卡欺诈风险的来源

1)商业欺诈：主要来源于与欺诈者勾结的不法商家或者合法商家的不法员工。

2)持卡人欺诈：主要是由于不道德的持卡人所导致的，而非不法分子冒用持卡人身份进行的信用卡欺诈。

3)第三方欺诈：信用卡欺诈主要来源，不法分子通过非法方式获取信息，并利用这些信息伪造或者骗取信用卡进行交易[1]。

2.2. 信用卡欺诈风险的分类

信用卡欺诈可以分为不同的类型，但其中多数属于第三方欺诈。欺诈风险的分类如表 1 所示。

3. 数据分析

为了充分理解提供的信用卡欺诈消费数据，我们首先对数据集中的有关内容进行了初步分析，对信用卡欺诈数据来源、企业与银行的互信关系以及数据内容上进行了初步的探索。

3.1. “无标签”数据获取过程

由于数据的隐私保护问题，数据的提供要考虑到企业与银行的互信关系问题。数据集中的 V1 到 V28 列是由原数据经过主成分分析提取的 28 个主成分，由于隐私保护的原因，这 28 个字段均无详细的说明，即提供的数据集为“无标签”数据。

考虑数据具体信息抹去的过程不可能由数据分析公司完成，否则就会造成隐私信息的泄露，所以从原始数据的主成分分析、标签的抹去均由银行等数据提供方完成，甚至其中的数据挖掘也有银行完成。“无标签”数据的获得流程如图 1 所示。

3.2. 数据集的元素分析

3.2.1. 时间元素

时间字段 *Time* 记录了从第一条记录开始以后的计算时间，单位为秒，共记录 172,792 秒。但我们发现时间的记录并非连续序列，且存在相同时间点多次被记录的现象，由此分析为时间元素的记录应该为多台记录机器，连续记录间断信用卡用户交易的过程。

3.2.2. 金额元素

金额元素 *Amount* 为刷卡消费的金额。数据集中并未给定金额元素的单位大小，观察金额的数据可以猜测金额元素的计量单位应为百元。

3.2.3. 金额元素

判断元素 *Class* 指本次信用卡交易是否存在欺诈行为，在欺诈的情况下其值为 1，否则为 0。

Table 1. Classification of credit card fraud risk

表 1. 信用卡欺诈风险的分类

风险类型	导致原因	风险后果
盗窃	信用卡被盗取持卡人挂失不及时	信用卡被大量盗刷信用卡被转卖到通信网络不发达的国家
复制	信用卡被违法复制空白卡上复制磁条	支付交易签名一致造成“白卡”欺诈交易
ATM 欺诈	隐蔽处被安装微型读卡或记录设备	密码被窃取或伪造甚至会遭遇暴力抢劫
伪造	使用压卡设备伪造剪切信用卡改造签账单	可能附有真实的持卡人信息仅用于无卡环境的欺诈交易
身份冒用	欺诈组织贩卖用户信息欺诈者违法使用	个人信息被顶替冒用隐私信息暴露

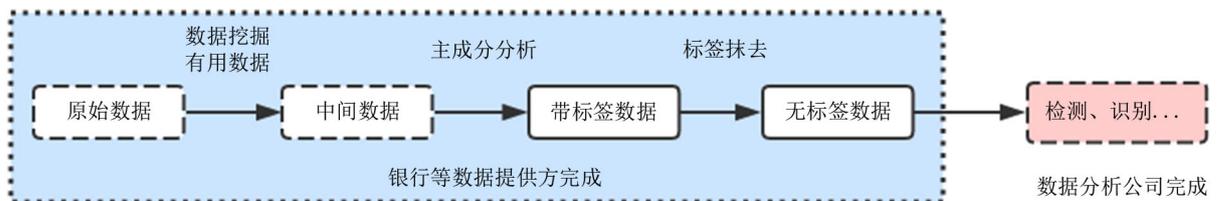


Figure 1. The sharing and obtaining process of unlabeled data

图 1. 无标签数据的分享与获取过程

3.3. 数据内容分析

在信用卡欺诈案例数据集中, 欺诈案例所占的比例甚小, 仅为 0.17%, 说明在常规的信用卡消费中, 发生欺诈事件仅为小概率事件(表 2)。在建模和测试计算过程中应该考虑到这种小概率事件的问题。

3.4. 参数指标的选取

根据中心极限定理, 若一个量是由许多极小的独立随机因素影响的结果, 则可以认为这个量具有正态分布的特征。信用卡欺诈性质的数据由于在整体数据中所占的比例极小, 且持卡人的每次刷卡行为都是独立事件, 所以欺诈数据符合正态分布。

为判定欺诈检测的正确性, 我们设置了 TP (True Positive Fraction)和 FP (False Positive Fraction)两个指标[2]。其具体含义为:

- 1) **TP 指标:** 将欺诈交易正确检测为异常交易的比例;
- 2) **FP 指标:** 将合法交易检测为异常交易的比例。

3.5. 模型样本的选取

为了准确选择训练数据和测试数据, 样本大小的选择也是非常重要的。其既要求模型精度达到一定的要求, 有要使花费的时间不致太高。

在欺诈数据正态分布的前提下, 设样本 (X_1, X_2, \dots, X_n) 来自正态分布总体 $N(\mu, \sigma^2)$, 总体均值 μ 的点估计为样本均值 \bar{X} 。若以 \bar{X} 估计 μ 时的绝对误差为 Δ , 可靠度为 $1-\alpha$, 即要求

$$P\left\{\left|\bar{X}-\mu\right|\leq\Delta\right\}=1-\alpha \quad (1)$$

由正态分布的原理

$$P\left\{\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right|\leq Z_{\frac{\alpha}{2}}\right\}=1-\alpha$$

可得

$$P\left\{\left|\bar{X}-\mu\right|\leq Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right\}=1-\alpha \quad (2)$$

结合(1)、(2)两式, 令绝对误差

$$\Delta=Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

从而解得样本选取的数量 n 为

$$n=\frac{Z_{\frac{\alpha}{2}}^2\sigma^2}{\Delta^2} \quad (3)$$

设 TP 指标的标准差为 $\sigma_1=0.15$, FP 指标的标准差为 $\sigma_2=0.05$, 在置信区间为 0.95, TP 和 FP 的误

Table 2. Probability of fraudulent events

表 2. 欺诈事件发生概率

欺诈事件发生数	统计信用卡交易数	欺诈发生概率
483	284,785	0.17%

差分别设定为 0.2%和 1.5%，其中 $Z_{0.025} = 1.96$ 。根据式(3)计算得到估计 TP 样本数量需要 21,609 个， FP 样本需要 43 个，共需样本 21,652 个。

基于所建模型理论，在此取训练集样本(无欺诈消费数据)数量为 20,000 个，同时选取测试集样本(欺诈消费数据)数量 400 个。

4. 模型准备

提供的数据集是经过数据预处理和主成分分析之后的数据，当一个新的用户数据产生时，需要重新完成相应的准备工作，才能将测试数据代入所建模型中进行检测。所以在此对数据预处理、数据挖掘和主成分分析，以及分类识别中的聚类分析过程进行简述。

4.1. 数据预处理

信用卡采集数据的预处理主要包括：

- 1) 异常值剔除：对收集到的信用卡消费的原始数据，按照拉伊达法则进行坏值和异常值的剔除；
- 2) 缺失值补足：对缺失数据可以采取忽略、插补或者加权调整的方式进行补足；
- 3) 归一化处理：对主成分间具有不同量纲的，采用归一化处理，消除数据单位与数量级造成的偏差；
- 4) 噪声数据处理：对收集到的信用卡消费数据中的随机误差或偏差进行去噪或降噪处理。

4.2. 数据挖掘

对原始数据进行数据挖掘，可以初步得到有利于检测模型建立和算法训练的中间数据。按照一般情况下的有标签数据，可以数据挖掘到持卡人用户信息和交易记录历史数据的两大类数据。数据挖掘的有利信息如表 3 所示。

由于隐私保护，中间数据经过后期银行的标签抹去后变成“无标签”数据，即我们得到的信用卡欺诈消费数据集。

4.3. 数据挖掘

信用卡数据在应用到模型之前首先要进行主成分分析提取相应的主成分，从而提取对识别模型有利的数据成分。

4.4. 聚类分析

信用卡消费数据的聚类主要针对交易属性的聚类，即区分合法交易和欺诈交易，分析客户的交易特征。信用卡欺诈检测一般采用 k -means 聚类算法[3]。

5. 理论概述

5.1. 信用卡欺诈检测模型

5.1.1. SVDD 模型适用分析

改进的支持向量机 SVDD 是根据传统的支持向量机 SVM 改进而来，其基本思想是为输入的样本计算出一个球状的决策边界，称为超球体[4]，它将整个空间划分为两部分：

- 1) **接受部分**：指超球体边界内空间，即合法的信用卡交易行为；
- 2) **拒绝部分**：指超球体边界外空间，即信用卡欺诈交易行为。

这就使得 SVDD 具有一类样本的分类特征，在训练过程中通过控制超球体的范围和大小，使球面内的样本保持该类别的特殊属性。

对于信用卡欺诈交易的案例数据,本问题的超球体维数为29维(即28维的主成分,加1维度的金额)。SVDD的超球体如图2所示。

对于无标签的信用卡欺诈检测问题,SVDD模型的适用性有如下的几点特征:

- 1) 无标签数据的处理:对无标签数据的处理问题,SVDD模型可以做到描述无标签数据的多类分类;
- 2) 数据类别不均衡问题:由于信用卡欺诈数据的比例占校园内消费总体数据的极小,SVDD可以采用分类精度对数据不均衡的机器学习问题;
- 3) 算法运行速度快:SVDD算法结合k-means聚类算法,进行“分治”并行计算加快了算法的运行速度。

5.1.2. SVDD 多类描述原理

1) SVDD 的训练方法

假设空间中存在样本集 $X = [x_i, i=1, 2, \dots, n]$, SVDD 目标需要找到一个超球体,使得 x_i 尽可能的包含在超球体内。设超球体的中心为 a 和半径为 R , 则寻找的超球面应满足

$$(x_i - a)(x_i - a)^T \leq R^2 \tag{4}$$

由于要使超球的半径最小,我们记(4)式为

$$\min f(R, a) = R^2 \tag{5}$$

为增强分类的鲁棒性,我们引入松弛因子 $\xi_i \geq 0, i=1, 2, \dots, n$, 同时引入控制错分样本惩罚程度的惩罚系数 C , 使得(5)式变为

$$\min f(R, a, \xi) = R^2 + C \sum_{i=1}^n \xi_i \tag{6}$$

Table 3. Favorable information for data mining (before processing)

表 3. 数据挖掘的有利信息(处理前)

数据挖掘信息	举例
持卡人用户信息	持卡人卡号、身份证号、学历、职业、年收入等
交易记录历史数据	交易日期、交易金额、交易类型等

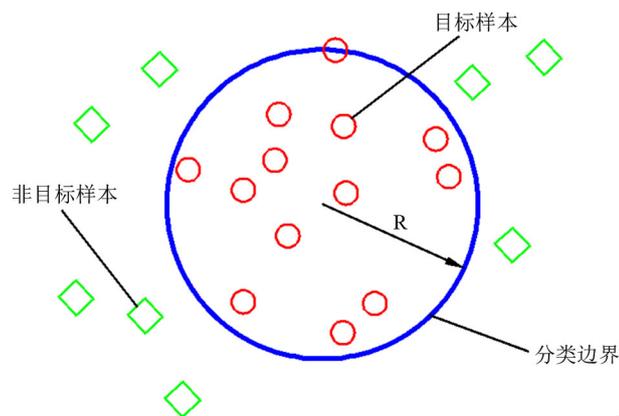


Figure 2. The schematic diagram of a super sphere (as an example of a two-dimensional space)

图 2. 超球体示意图(以二维空间为例)

其约束条件为

$$s.t. (x_i - a)(x_i - a)^T \leq R^2 + \xi_i, \text{ 且 } \xi_i \geq 0$$

式(6)是一个考虑中心 a 和半径 R 的二次规划问题, 我们引入拉格朗日算子, 建立拉格朗日函数, 对于 n 维问题有

$$L(R, a, \xi, \alpha, \gamma) = R^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i [R^2 + \xi_i - (x_i^2 - 2ax_i + a^2)] - \sum_{i=1}^n \gamma_i \xi_i \quad (7)$$

由于要求得 L 关于 R, a, ξ 的最小值, 对式(7)求其偏导并令其为 0, 得

$$\begin{aligned} \frac{\partial L}{\partial R} &= 0, \quad \sum_{i=1}^n \alpha_i = 1 \\ \frac{\partial L}{\partial a} &= 0, \quad a = \sum_{i=1}^n \alpha_i x_i \\ \frac{\partial L}{\partial \xi} &= 0, \quad C - \alpha_i - \gamma_i = 0 \end{aligned}$$

将上述得到的条件重新带入式(7)中, 得到最终的优化问题为

$$L(R, a, \xi, \alpha, \gamma) = \sum_{i=1}^n \alpha_i (x_i \cdot x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \quad (8)$$

其约束条件为

$$s.t. \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i = 1 \quad (9)$$

2) SVDD 的识别方法

假设 z 为测试样本, 当下式满足时即判定 z 为正常类别, 即落在超球体内部, 否则为异常类别[5]

$$\|z - a\|^2 = (a \cdot a)^2 - 2 \sum_{i=1}^n \alpha_i (a \cdot x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \quad (10)$$

利用指示函数对(10)式进行判断

$$\begin{aligned} f_{SVDD}(z, a, R) &= I(\|z - a\|^2 \leq R^2) \\ &= I\left\{ (a \cdot a)^2 - 2 \sum_{i=1}^n \alpha_i (a \cdot x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \right\} \end{aligned} \quad (11)$$

这里指示函数 I 的定义为

$$I(A) = \begin{cases} 1, & \text{if } A \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

这即是我们最终检测信用卡是否存在欺诈交易行为的准则, 也是数据集 *class* 字段中 0 与 1 的来源。

3) 核函数的引入

当输入空间的样本点不满足球状分布时, 可以通过引入核函数使其映射到高维空间, 在此映射法则下于高维空间中进行求解。也就是将式(18)中的内积变换为

$$x_i \cdot x_j \rightarrow \varphi(x_i) \varphi(x_j) = K(x_i, x_j) \quad (13)$$

引入核函数后

$$f_{SVDD}(z, a, R) = I(\|\varphi(z) - \varphi(a)\|^2 \leq R^2) = I\left\{K(a, a)^2 - 2\sum_{i=1}^n \alpha_i K(a, x_i) + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i, x_j) \leq R^2\right\} \quad (14)$$

核函数的几何意义如图 3 所示。常用的核函数形式有

1) 线性核函数:

$$K(x, y) = x \cdot y$$

2) 多项式核函数:

$$K(x, y) = (x \cdot y + 1)^d, d = 1, 2, \dots$$

3) 高斯核函数:

$$K(x, y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad (15)$$

为了提高算法的分类准确率，我们在求解过程中采用的核函数为高斯核函数。

6. 风险评估

根据 SVDD 理论建立适用于本数据的信用卡欺诈风险的识别模型，信用卡欺诈风险控制的总体流程图如图 4 所示。

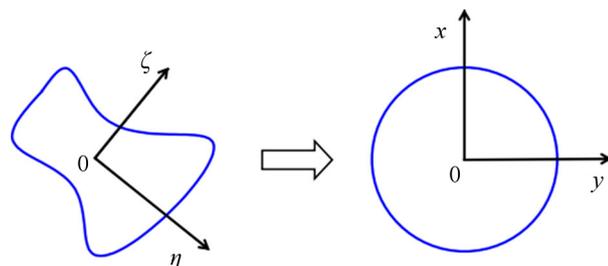


Figure 3. The schematic diagram of the geometric meaning of a kernel function

图 3. 核函数的几何意义示意图

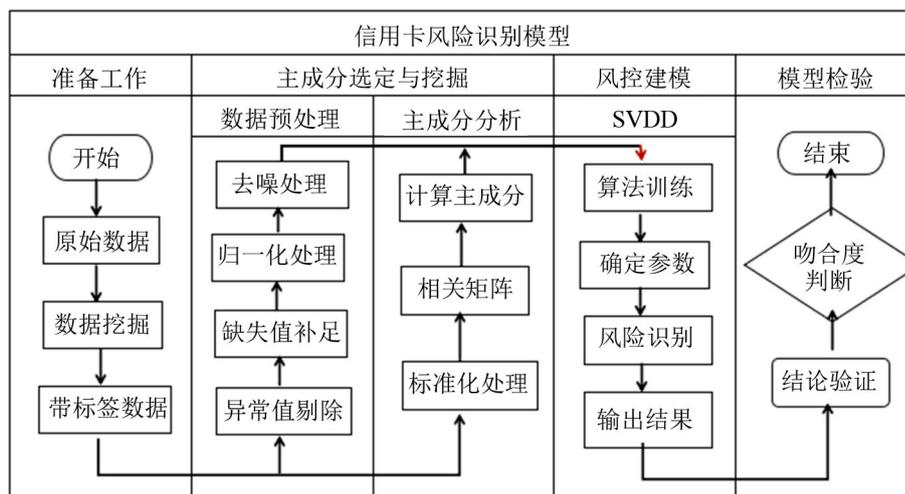


Figure 4. The overall flow chart of credit card fraud risk control

图 4. 信用卡欺诈风险控制的总体流程图

6.1. 参数确定

6.1.1. 核函数 K 的确定

选取 SVDD 模型的核函数为高斯径向基核函数, 选取核函数的调节参数的值为 0.5, 则式(14)中的高斯核函数可表示为

$$K(x_i, x_j) = e^{-\frac{\|x_i - x_j\|^2}{0.5^2}} \quad (16)$$

6.1.2. 惩罚系数 C 的确定

惩罚系数 C 用来平衡超球体积大小和正常样本上的错误率, 实际应用中一般通过程序调试不断改进该系数值, 但也可按照下式计算

$$C = \frac{1}{\nu l} \quad (17)$$

式中, l 是样本个数, ν 是用户设置的参数, 其表示目标类上错分率的上界, 也是支持向量占训练样本的百分率的下界。

针对选取的 20,000 训练样本, 我们取用户设置参数 $\nu = 0.002$, 因此由公式(17)得惩罚系数 C 值为 0.25。

6.1.3. 消费金额的范围约束

消费金额 $Amount$ 是体现交易是否存在异常的重要的即时信息, 由于所给数据中消费金额相比较于 28 个主成分而言, 其数值变化范围明显大于 28 个主成分间的数值范围, 在模型求解中严重影响超球体的半径, 增加了模型的误差。

为此, 通过改变消费金额的数量级, 将消费金额约束在合理的变化范围内, 即新的消费金额数据为

$$Amount' = Amount \times 0.01 \quad (18)$$

式中, $Amount'$ 为经过范围约束后的消费金额, $Amount$ 为消费金额的原数据。

6.2. 模型求解

将主成分分析提取得到的 28 个主成分, 和消费金额 $Amount$ 数据组合后, 得到一个 29 维的特征向量, 即样本数据应该满足 $x_i \in R^{29}$ 。则式(14)具体变为

$$\min f(R, a, \xi) = R^2 + 0.25 \sum_{i=1}^{29} \xi_i, (i = 1, 2, \dots, 29) \quad (19)$$

$$s.t. (x_i - a)(x_i - a)^T \leq R^2 + \xi_i, \text{ 且 } \xi_i \geq 0$$

利用 *MATLAB* 求解最优超球面, 得到 29 维的最优超球面中心 a 和半径 R 为

$$\begin{cases} a = (-22.37, -4.29, -19.34, 7.13, -15.19, -0.49, -5.35, 9.25, -1.89, -5.76, 2.55, -1.81, 1.26, -1.49, \\ \quad 0.10, -2.25, -5.22, -1.87, -1.20, 8.77, 3.25, -3.17, -4.16, 0.14, -0.21, -0.72, -0.07, 0.87, 39.89) \\ R = 47.1386 \end{cases}$$

其中, 中心 a 为 29 维向量, 即 $a \in R^{29}$ 。则得到最终识别函数为

$$f_{SVDD}(z, a, R) = I \left\{ K(a, a)^2 - 2 \sum_{i=1}^{29} \alpha_i K(a, x_i) + \sum_{i=1}^{29} \sum_{j=1}^{29} \alpha_i \alpha_j K(x_i, x_j) \leq 47.1386^2 \right\} \quad (20)$$

7. 模型检验

在进行模型训练时选取的训练样本为无欺诈消费数据，为了检验所得训练模型的准确性，我们将 $Class = 1$ 的欺诈消费数据(共 483 条)代入 SVDD 模型中进行识别，比较模型识别结果与数据真实结果的吻合度，以此检验模型的准确性。其中吻合度的定义为

$$\mu = \frac{m}{m_0} \quad (21)$$

式中， m_0 为原始欺诈事件数， m 为模型识别欺诈事件数。

由表 4 可知，将 483 条欺诈消费数据代入识别后，有 456 个样本点在超平面外，即模型识别出 456 条欺诈消费数据，吻合度高达 94.41%，说明建立的模型有较高准确性。

8. 风险对策

风险管理是识别面临风险，并采取适当措施使得损失或破坏的概率降低的系统方法。选择有效的风险管理方法来处理和避免风险的产生，对信用卡交易及欺诈事故处理有着重要意义。

8.1. 风险的处理方法与对策

从宏观角度看，一个完善有效的信用卡风险管理策略应该包括市场策略、信贷政策、利润评估、风险监控和风险评价等多个方面(表 5)。

8.2. 风险管控模型的构建

在完成信用卡欺诈的风险识别和检测后，建立风险管控的概率模型和反欺诈模型来对信用卡欺诈行为进行管理与控制。

Table 4. The result of matching
表 4. 吻合度结果

原始欺诈消费事件	识别出欺诈消费事件	吻合度
483	456	94.41%

Table 5. Risk treatment methods and Countermeasures
表 5. 风险处理方法和对策

编号	处理方法	具体对策
1	妥善的市场管理	1) 明确的市场定位，清晰的客户群体，针对性的需求分析； 2) 根据本行的优势和客户信用，分析信用卡办理风险； 3) 在信用卡发行、推广和客户选择中最大化的化解风险交易。
2	合理的信贷政策	1) 发卡银行采取正确的信用额度、还款周期、可透支金额等； 2) 做好信用监督和交易监控； 3) 在给予信用政策和风险管控之间寻找最佳平衡点。
3	准确的利润评估	1) 做好信用卡的利润与风险评估，预测收益风险； 2) 估算信用卡发行利润，确定措施可行性。
4	建立风险预警	在动态交易中进行风险监督和预警，利用已有数据库及时对信用卡新增交易行为进行风险预测与识别。
5	风险信息共享	建立各家银行间的风险信息共享体系，对欺诈案件的协助调查、立案通报，各成员银行之间应该给予充分的风险信息共享，追查和跟踪风险。
6	用户行为评分	1) 建立信用卡用户行为评分体系，对信用卡进行有效的跟踪和监控； 2) 利用历史消费记录等信息对用户潜在风险进行评估； 3) 对持卡人的实时交易行为进行跟踪与监控，预防风险。

8.2.1. 信用卡欺诈管控的概率模型

在考虑可能导致信用卡欺诈风险发生的因素和情况下，我们建立了风险欺诈管控的概率模型，使欺诈行为发生的概率低于允许概率，即

$$P = a \cdot f_i \cdot \lambda \cdot \phi \leq [P_0] \tag{22}$$

式中， P 为发生欺诈风险的概率， $[P_0]$ 为允许概率， a 为概率调节系数， ϕ 为本次消费金额， λ 为信贷程度。

根据风险约束机制的相关要求，设定信用卡欺诈交易行为的发生概率应该约束在 0.8% 以内，从而保证信用卡消费支付的正常交易。

8.2.2. 信用卡交易的反欺诈系统

1) 银行交易系统

信用卡风险的检测和识别一般在银行的中央计算机中进行。中央计算机完成数据提取、数据筛选、欺诈分析和反欺诈处理等过程，其中数据筛选、数据分析和发欺诈处理时银行反欺诈系统的核心功能(图 5)。

2) 反欺诈系统的设计

中央计算机在获取用户信息之后，首先对数据进行筛选，再利用数据挖掘技术进行欺诈分析，最终结果通过接口层返回交易前台[6]。如果系统判定为欺诈交易，前台的图形用户接口会显示警告。反欺诈系统的设计流程如图 6 所示。

9. 风险对策

本文首先基于数据挖掘和主成分分析方法，建立了信用卡欺诈交易识别的 SVDD 改进支持向量机模型，对欺诈交易进行识别和评估，且提出了对信用卡欺诈风险管控的模型和对策。

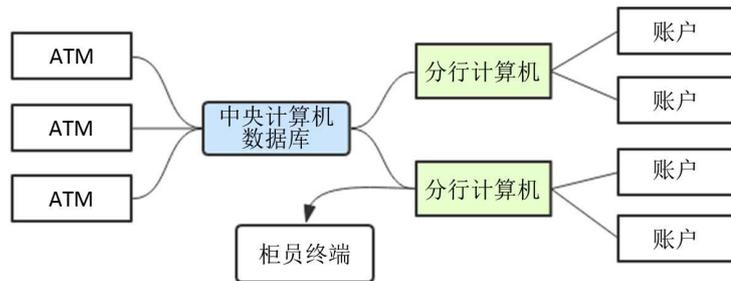


Figure 5. Bank trading system

图 5. 银行交易系统

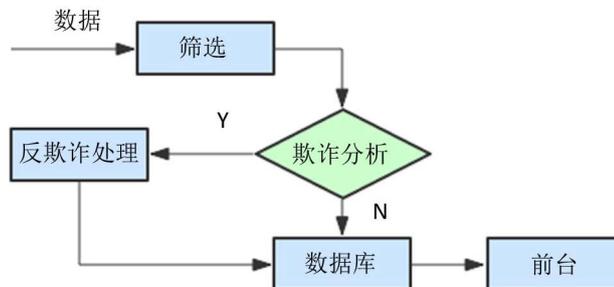


Figure 6. Anti fraud system design

图 6. 反欺诈系统设计

9.1. 检测模型的建立

在数据预处理、数据挖掘和主成分分析的基础上,通过考虑无标签数据的处理和数据类别不均衡问题,我们建立了 SVDD 模型对信用卡欺诈进行识别和检测。通过概率方法选取样本,引入核函数之后,选取惩罚系数和松弛因子建立起对风险识别的 SVDD 模型,以超球体为判别原则进行多分类。

9.2. 风险识别

在模型建立的基础上,通过求解得到超球体的半径和中心,以在超球体内的样本数据为接受数据,即交易合法数据;以超过超球体之外的数据为拒绝数据,即信用卡交易欺诈数据。通过,通过 SVDD 模型的指示函数对样本数据进行判断,最终识别本次交易是否存在欺诈行为。

9.3. 风险管控

通过分析风险的成因,我们建立了欺诈风险的概率模型,并提出了银行反欺诈系统的设计方法。分析所见模型的局限性和缺点,相应地提出了拓展建议。最后我们提出了针对信用卡交易风险的处理方法和对策。

参考文献

- [1] 杨玺. 基于支持向量机的信用卡欺诈检测研究[D]: [硕士学位论文]. 成都: 四川师范大学, 2008.
- [2] 李洁. 基于 HMM 模型的信用卡欺骗风险检测系统的仿真分析[D]: [硕士学位论文]. 马鞍山: 安徽工业大学, 2010.
- [3] 韩晓红, 胡彧. K-means 聚类算法的研究[J]. 太原理工大学学报, 2009, 40(3): 236-239.
- [4] 刘艳红. SVDD 算法研究及在信用卡欺诈检测中的应用[D]: [硕士学位论文]. 镇江: 江苏大学, 2010.
- [5] 朱帮助, 林健. 基于支持向量数据描述的无标签数据多类分类[J]. 智能系统学报, 2009, 4(2): 131-136.
- [6] 严华, 胡孟梁, 蔡瑞英. 防止信用卡欺诈的系统设计[J]. 微计算机信息, 2006(12): 63-66.

附 录

第四届“好贷杯”中国风险管理控制与控制能力挑战赛的赛题数据请见:

<http://fengkong.madio.net/index.php?m=content&c=index&a=lists&catid=62>

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2324-7991，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：aam@hanspub.org