# 基于贝叶斯分类的可疑金融交易识别研究

张成虎，赵小虎
（西安交通大学 经济与金融学院，陕西 西安 710061）

**摘　要**：国内外反洗钱工作的大量实践表明，金融交易活动是洗钱犯罪行为的一个重要环节，通过分析金融机构的客户信息和交易数据，采用科学的方法识别可疑金融交易进而发现洗钱线索，已成为反洗钱研究的核心工作。文章将数据挖掘方法与金融领域知识相结合，首先通过对金融交易信息的多层次分析，总结出不同信息层次上的可疑金融交易特征；其次针对不同层次的交易信息，选择合适的数据挖掘方法，并结合客户背景资料，识别出可疑金融交易记录；最后依据贝叶斯判定原理，综合各层次的可疑信息，得到交易记录的整体可疑度，最终为反洗钱监测提供快速准确的参考。实验结果证明该方法是可行和有效的。

**关键词**：反洗钱；数据挖掘；可疑金融交易；贝叶斯分类
**中图分类号**：F830.4　**文献标识码**：A　**文章编号**：1001-9952(2009)10-0070-11

可疑金融交易识别是一个比较复杂的过程，能否有效识别业已发生的可疑金融交易，并对未来可能发生的洗钱模式做出预测，在很大程度上取决于所采用的识别方法、技术和手段的有效性。当把数据挖掘技术应用于可疑金融交易识别时，不单要选择适当的挖掘方法，还要结合相关领域知识对其进行优化和创新。由于金融交易方式的多样性，交易主体行为的不确定性和交易记录的时变性，使得在可疑金融交易识别领域，单一检测方法往往存在适用性、效率和条件约束等问题，难以对金融交易的整体可疑度进行判断。要达到理想的识别效果，必须在深入分析金融交易信息的基础上，结合领域知识，选择科学合理的信息汇总判定方式，综合各种检测方法发现的可疑线索，对金融交易记录的可疑度作出快速准确的整体判定。

作为一种基于概率的不确定性推理方法，贝叶斯法则在处理不确定信息

**收稿日期**：2009-05-13
**基金项目**：国家自然科学基金资助项目(70771087)
**作者简介**：张成虎(1958—)，男，陕西延安人，西安交通大学经济与金融学院教授，博士生导师；
赵小虎(1972—)，男，陕西宝鸡人，西安交通大学经济与金融学院博士生。

的智能化系统中已得到了广泛的应用。朴素贝叶斯分类是一种基于统计学的分类方法，用于预测类成员关系的可能性。应用于海量数据分析，贝叶斯分类表现出高准确率和高速度。同时，贝叶斯分类具备自适应功能，通过学习新的洗钱交易及正常交易样本，贝叶斯分类能反映最新的洗钱交易手法变换，为反洗钱监测提供一种快捷高效的方法。

## 一、相关研究

国外关于反洗钱信息技术的应用研究起步较早，20 世纪 70 年代就开始了反洗钱立法方面的研究，而将信息技术应用到反洗钱领域的研究也在 20 世纪 90 年代就已开始。Senator(1995)较为系统地介绍了 FinCEN(Financial Crimes Enforcement Network，金融犯罪执法网)的 FAIS(FinCEN Artificial Intelligence System，FinCEN 人工智能系统)的系统结构、监测识别关键技术及其应用。FAIS 综合采用各种人工智能技术，通过智能分析被提交的交易报告，发现各类可疑金融交易行为。FinCEN 系统中的交易可疑度评价模块，采用贝叶斯模型判断可疑度，然后再对高可疑度的交易数据进行进一步的分析调查。Stofella(1997)介绍了意大利中央银行监管部门(UIC)如何运用高性能数据库和数据可视化技术构建数据挖掘环境，对意大利整个金融系统的交易信息进行监测。Petrus C Van Duyne(1999)通过分析荷兰 1994－1996 年的可疑金融交易数据，指出可疑金融交易监测系统和反洗钱策略方面存在的问题，并提出改进建议。Kingdon J 和 Feldman K S(2002)设计了银行交易数据监测和分析系统，通过该系统可以自动检测到支付欺诈和金融领域的洗钱活动。Kingdon J(2004)设计出一套可自动识别客户行为模式的人工智能系统，应用该系统可以高效识别出客户的异常交易行为(unusual behavior)。针对传统的基于规则的洗钱监测系统不能适应洗钱手法的快速多变，误报率高，对海量交易数据健壮性差等局限性，国外学者展开了基于智能代理的反洗钱监测系统研究，指出该方法可以提高反洗钱监测效率，实现系统整合，增强适应性，同时还能降低监测成本。

我国对反洗钱的研究于 20 世纪 90 年代后逐渐增多，目前对于人工智能技术、数据仓库技术、数据挖掘技术等技术在反洗钱中应用的研究仅仅处于起步阶段。徐志春、肖伟平、何宏(2003) 提出了基于数据开采技术的反洗钱系统的实现框架，介绍了反洗钱系统中用到的几个关键数据开采技术，包括数据集成、数据分类、关联分析、聚类分析和可视化技术。汤俊(2005)分析了我国现行交易报告制度存在的问题，指出国内对于金融交易客户行为模式识别的技术研究处于空白状态，并提出了相应的框架体系。杨胜刚、王鹏(2005)在探讨数据挖掘技术在大额和可疑交易报告制度中应用的必要性与可行性，在全面把握数据挖掘技术的各种主要算法及其在大额和可疑金融交易数据分析中

的应用前景的基础上，针对我国反洗钱工作的实际，设计了一套人民币大额和可疑支付交易数据挖掘系统。陈云开(2006)提出分布式异构计算环境下基于数据挖掘技术的洗钱侦测系统体系结构，并从逻辑层次结构、系统基本框架和系统基本流程三方面对洗钱侦测系统的体系结构进行了阐述。孙景等(2008)根据逻辑回归原理与数据挖掘技术，建立了企业大额可疑外汇资金交易识别模型，用于分析银行的企业客户洗钱概率及洗钱事件发生的可能性，并通过对具有洗钱嫌疑的银行企业客户进行识别和预测，为银行反洗钱提供参考。

从国外相关研究来看，由于洗钱与反洗钱间存在的博弈关系，各国洗钱监测手段具有一定的保密性，各类研究主要侧重于数据挖掘技术应用等方面，与反洗钱领域知识相结合的研究细节很少公开。同时国外的反洗钱实际上与我国的洗钱活动特征和反洗钱管理方式存在较大的差异，国外的研究成果无法为我国可疑金融交易识别提供一个有效的策略及方法。从国内相关研究来看，由于受我国反洗钱实践经验的限制，国内在可疑金融交易识别方面的研究成果相对较少。很少有基于反洗钱实践的、应用真实交易报告数据进行验证的数据挖掘算法研究，同时，针对单一检测方法可能具有片面性，综合各类检测结果的可疑金融交易识别研究还很少见，可疑金融交易的自动识别研究则更为少见。

本研究以数据挖掘技术为手段，应用金融领域知识和反洗钱领域知识，通过对真实交易报告数据的层次分析，选择合适的数据挖掘方法对可疑金融交易进行多层次识别，同时针对单一识别方法在可疑金融交易识别中的片面性，建立一套可疑金融交易线索整体判定方法，得出交易记录的整体可疑度，不仅在更大程度上发挥了各种识别方法的优势，而且可以为可疑金融交易的识别提供一种新的思路和方法，目的在于提高反洗钱监测的效率和水平。

## 二、可疑金融交易特征与识别方法

可疑金融交易藏匿于金融机构成千上万的海量交易数据中，对其甄别的难度很大。数据挖掘技术能够根据分析人员的需要，从海量数据中提取有价值的模式和规律，它的发展和广泛应用为其在识别可疑金融交易和挖掘洗钱线索方面的应用奠定了基础。同时，数据挖掘技术在可疑金融交易识别中的应用离不开交易信息的深入分析和算法的合理选择。

1. 可疑金融交易特征分析。可疑金融交易是指金融交易的金额、频率、来源、流向和用途等有异常特征的交易行为。从反洗钱监测实践来看，可疑金融交易行为经常表现为以下几类特征：(1)交易金额、交易频率异常。例如短期内频繁发生资金收付，但与客户身份、财务状况、经营业务明显不符；长期闲置的账户原因不明地突然启用或者平常资金流量小的账户突然有异常资金流入，且短期内出现大量资金收付等。(2)交易流向或交易来源异常。例如与来

自贩毒、走私、恐怖活动、赌博严重地区或者避税型离岸金融中心客户之间的资金往来活动在短期内明显增多，或者频繁发生大量资金收付；多个境内居民接受一个离岸账户汇款，其资金的划转和结汇均由一人或者少数人操作等。(3)交易用途或交易性质异常。例如没有正常原因的多头开户、销户，且销户前发生大量资金收付；证券经营机构指令银行划出与证券交易、清算无关的资金，与其实际经营情况不符；保险机构通过银行频繁大量对同一家投保人发生赔付或者办理退保等。

以上是从交易行为的角度对可疑金融交易特征进行剖析，借助反洗钱领域知识，从金融交易记录的角度分析，交易金额的异常通常体现在单笔或相关交易记录中，交易频率的异常通常体现在基于时间序列的交易记录中，而交易用途或交易性质的异常则通常体现在交易主体间的交易往来中。

2. 可疑金融交易识别方法。基于交易记录层面的交易金额异常，通常与交易数据异常点对应，可选择聚类算法加以识别。聚类作为一种重要的数据挖掘技术，通过无指导学习将数据划分成多个簇，聚类结果表现为簇内成员的相似和不同簇中成员的差异。对金融交易数据进行聚类分析，聚类结果中的孤立点多为交易金额异常记录。在可疑金融交易识别中孤立点的检测成为发现和识别可疑金融交易的重点。基于聚类分析的这一特点，它在可疑金融交易识别中被广泛应用于交易金额异常的监测。

针对交易账户层面的交易金额、频率异常情形，可选择时间序列分析方法加以识别。交易账户信息中所反映出的交易金额、频率异常，通常表现为交易记录时间序列的信号突变。小波分析是发现时间序列信号异常的有效方法。由 Morlet 提出的小波分析(wavelet analysis)是一种具有时频多分辨功能的调和分析方法，将小波分析引入可疑金融交易识别研究中，与可疑金融交易特征相结合，选择合适的小波函数，不依赖于经验模型，对金融序列进行小波变换，可多尺度揭示交易序列的变化规律，挖掘出隐藏于交易时间序列中的单笔异常交易和密集频繁交易，为反洗钱监测提供一种快捷高效的方法。

针对关联账户层面的交易流向、来源以及用途或性质异常情形，可选择链接分析方法加以识别。交易变量之间的相关性是知识发现的重要方面，链接分析可用于识别不同交易主体间交易活动的联系，而交易流向及用途等的异常通常表现为交易变量之间的异常关联。通过约束性链接分析，可以更好地发现可疑金融交易信息的内在联系。由于链接分析不用构造频繁项目集，不用设置最小支持度和置信度阈值，同时具备可视化特点，因此在挖掘交易信息内在相关性方面独具优势。在反洗钱中，通过对交易主体与交易流向、交易编码之间的链接分析，可发现资金流向或交易性质异常的可疑金融交易。需要指出的是，链接分析的结果不代表交易信息中内在的因果关系，但从洗钱侦测角度来看，这种相关性能为可疑金融交易识别提供有潜在价值的线索。

运用离群点聚类、小波分析和链接挖掘技术识别可疑金融交易，这三种方法识别出的可疑结果不尽相同且相互独立，依据每种识别方法所得的结果对交易信息予以可疑度标识，设定三种可疑参数，分别为“Suspicious _Cluster”、“Suspicious _Wavelet”和“Suspicious _Link”，将每种检测方法认定的可疑金融交易参数值标识为“1”，其余标识为“0”，在此基础上建立以各可疑参数作为属性的新的数据集，作为下一步运用贝叶斯准则进行可疑度整体判定的数据源。

## 三、贝叶斯分类与交易可疑度整体判定

面对金融交易的复杂性和不确定性，每一类识别方法都有其应用的约束条件，有其优点和缺陷。为了反映真实交易变化趋势，发挥各种数据挖掘方法在可疑金融交易识别方面具有的优势，将各种可疑线索应用于反洗钱实践，需对交易记录的整体可疑度做出科学的判断。作为一种基于概率的不确定性推理方法，贝叶斯判定具备整体判定优势，可从总体和细节两方面把握可疑金融交易特征，将各类识别方法所采集的可疑金融交易线索进行综合分析，得出交易可疑度的整体判定结果，同时该方法可操作性强，能够为反洗钱决策较好的参考。

1. 贝叶斯分类与交易可疑度参数设定。贝叶斯分类有朴素贝叶斯分类和贝叶斯信念网络两种。前者是一种简单而高效的分类方法。由此，本文将利用朴素贝叶斯分类方法实现金融交易可疑度整体判定。

在可疑金融交易识别中，每个数据样本用一个 n 维特征向量 $X=\{x_1, x_2, x_3, \cdots, x_n\}$ 表示，分别描述对 n 个属性 $A_1, A_2, A_3, \cdots, A_n$：可疑属性 1，可疑属性 2，可疑属性 3，可疑属性 n 样本的 n 个度量。数据分为正常($C_1$)和洗钱($C_2$)两类。给定一个未知的数据样本 X(即没有类标号)，可疑金融交易识别中交易记录可疑度的判定只需计算出可疑概率的大小即可，即 $P(C_2|X)$。

根据贝叶斯定理：$P(C_2|X)=\frac{[P(X|C_2)P(C_2)]}{P(X)}$，类的先验概率 $P(C_2)=s_2/s$，其中 $s_2$ 是训练集中洗钱的样本数，s 是训练样本总数。由于属性间不存在依赖关系，$P(X)=\prod_{k=1}^{n} p(x_k)$ ①，$P(X|C_2)\prod=\prod_{k=1}^{n} p(x_k|C_2)$ ②。

基于贝叶斯原理的可疑概率公式：P=P(洗钱|X)，设 P=P(洗钱|X)，将 P 定为交易记录可疑度。通过对样本数据的判定得到合适的可疑度判定阈值 K，该阈值为多次试验比较所得的较为理想的数值，以该值为可疑度判定的下限值促使对训练数据可疑与否的判断正确率尽可能高、遗漏率尽可能低。同时，会随着洗钱活动的不断变化进行调整。

运用整体判定准则进行交易可疑度判定，判定准则的选择至为重要，直接影响可疑金融交易识别的有效性。本文选用贝叶斯判定准则作为整体判定准

则，一方面在理论上是科学的，因为基于贝叶斯分类判定准则的分类方法有严谨的理论基础，已被广泛应用于各类科学研究；另一方面，采用可疑度参数作为贝叶斯分类判定的源数据，完全满足朴素贝叶斯定理的“类条件独立假设”和“概率分布可知”的要求，在应用实践中也是可行的。如果在以后研究中发现有更佳的可疑金融交易整体判定准则，可进一步优化可疑金融交易整体判定模式。

2. 交易可疑度整体判定流程。针对金融交易数据，基于三类可疑金融交易特征，分别利用基于 CURE 聚类的交易数据离群点分析、基于小波分析的交易序列突变点检测和基于链接挖掘的交易路径异常识别方法进行处理，得到金融交易记录的三项可疑属性数据集，然后利用贝叶斯准则进行判断。结合训练集学习所得结果，计算出交易活动洗钱的概率为：

$$P(\text{洗钱}|X)=\frac{P(X|\text{洗钱})P(\text{洗钱})}{P(X)}=\frac{\prod_{k=1}^{n}p(x_k|\text{洗钱})P(\text{洗钱})}{\prod_{k=1}^{n}p(x_k)}$$

将交易记录可疑度指标 P 与选定的最佳可疑度判定值 K 进行比较，若 $P\geqslant K$，则将该交易记录标定为可疑；若 $P<K$，则记录标定为正常。

## 四、实验验证

“可疑金融交易整体判定方法”对金融交易信息的处理分为两个步骤：第一步针对不同的可疑金融交易特征，利用适合的数据挖掘方法，识别出各类可疑金融交易，标识可疑金融交易记录；第二步将交易可疑标识量化成交易可疑参数，利用整体判定准则对交易可疑度进行整体判定，得到最终的判定结果。通过这两个步骤，达到更好综合不同的检测方法对可疑金融交易进行识别的目的。本文运用真实的金融交易数据对整体判定模式进行实验分析。实验通过 SAS 8.0 工具软件编程和调用 SAS EM（企业数据挖掘）模块来完成，最后对实验结果进行了评估。

1. 数据准备。本实验所采用的源数据是选取某省企业 2003－2007 年外汇账户交易数据，共计 11939 条记录，1274 个账户，其中包括业已确认的犯罪线索记录 210 条。[③] 从企业外汇账户交易数据中抽取交易序号、交易币种、交易发生日、企业代码、资金收付标志、交易编码、交易对象、交易对象所属国家或地区等信息形成客户原始交易数据，对客户原始交易数据进行数据预处理操作，将交易金额折合为美元，对当日没有交易发生的情形，设定交易金额为零。对每笔外汇资金交易数据的客户代码、交易金额、资金收付标志、交易发生日、交易编码、交易对象等重要字段做逐一检查，对一些错误和缺失值（missing value）使用经验值或背景资料进行补充，经数据预处理后的企业外

汇交易数据集属性如表 1 所示。随机抽取 70%数据进行综合识别方法训练学习,30%留作验证使用。

**表 1　企业外汇交易数据集属性及说明**

| 属性名称 | 说　明 |
|---|---|
| 交易序号 | 银行交易记录标识号,具有唯一性。 |
| 企业代码 | 企业代码采用国家技监局规定的全国统一的 9 位标识码。 |
| 交易发生日 | 交易日期以记账日为准。 |
| 交易编码 | 记录交易性质和交易用途。 |
| 资金收付标志 | 表示资金流向,"1"为流入,"0"为流出。 |
| 交易对象名称(姓名) | 交易对象企业名称(个人姓名)。 |
| 交易对象所属国家或地区 | 采用三位英文缩写国别标识码。 |
| 交易金额 | 统一为原币折美元金额。 |
| 洗钱标识 | 业已认定的洗钱交易记录标识为"1",否则为"0"。 |
| 可疑标识 | 初始为"0",经分析认定的可疑金融交易标识为"1",正常为"0"。 |

2. 实验过程。整体判定方法包括基于各种检测方法的可疑参数获取和基于贝叶斯准则的判定两个过程。

(1)基于各种检测方法的可疑参数获取。运用基于 CURE 聚类的金融交易数据离群点分析方法对交易数据中资金转移异常进行检测。8357 条观测数据中的 8304 条数据被聚为 6 类,另外有 183 条观测数据被归到可疑金融交易集合当中,对该集合中客户的身份特征进行分析,发现客户背景资料与交易特征差异较大,进行与其身份不相符合的大额频繁外汇交易,该集合被列为可疑外汇交易的重点监控对象,在相应交易记录中标记为可疑,即"Suspicious _Cluster"属性值为"1"。④

运用基于小波技术的交易序列突变点检测方法对交易账户层面的交易金额、频率异常情形进行识别。针对 915 户账户信息,⑤根据每一个账户每天的交易信息构建金融交易时间序列,实现对金融交易时间序列的小波分析。从 915 个账户中提取出交易信号异常账户 23 个。由于交易主体的不同,针对筛选出的 23 个账户,结合企业的背景信息对小波分析结果进一步筛选,对带有普遍行业特征的属于正常交易的小波分解细节信号异常应予以排除,同时排除企业经营状况好转出现的资金往来突然活跃情况,认定交易可疑度较大,需进一步调查审核的账户 11 个,将其包含的 216 条异常交易记录标记为可疑,即"Suspicious _Wavelet"属性值为"1"。⑥

运用基于链接挖掘对关联账户层面的交易流向、来源以及用途或性质异常情形进行检测。由于可疑金融交易相对较少,运用链接挖掘处理大量金融交易信息时,必须根据掌握的异常特征,给出约束性规则,对属性值进行筛选,从中挑选出感兴趣的交易主体加以分析。通过交易关系挖掘共发现可疑金融交易记录 209 条,其"Suspicious _Link"属性值为"1"。⑦

(2)基于贝叶斯准则的整体判定。利用数据挖掘方法对训练样本数据在不同交易层面的可疑情况予以识别,以各可疑标识作为集合属性得到新的数据集,将新数据集作为交易记录可疑度整体判定的数据源,运用贝叶斯准则予以推理判定。贝叶斯整体判定源数据集属性如表 2 所示。

**表 2　贝叶斯整体判定源数据集属性**

| 属　性 | 说　明 |
|---|---|
| 企业代码 | 技监局 9 位标识码,具备唯一性。 |
| 交易发生日 | 以记账日为准。 |
| 交易金额 | 折美元后金额。 |
| 交易编码 | 记录交易性质、用途。 |
| 资金流向 | 交易对象的国别或地区。 |
| …… | …… |
| 洗钱标识 | 业已认定的洗钱记录为“1”,否则为“0”。 |
| 聚类分析异常标识 | 检测资金转移异常,“1”为可疑,“0”为正常。 |
| 小波分析异常标识 | 检测交易波动异常,“1”为可疑,“0”为正常。 |
| 链接挖掘异常标识 | 检测交易关系异常,“1”为可疑,“0”为正常。 |

确定贝叶斯判定源数据集后,经过已知可疑与否的训练数据的学习(学习过程一和学习过程二所得如表 3 和表 4 所示),选定最佳可疑度判定指标,得到贝叶斯判定方法的可疑度判定阈值。

**表 3　学习过程一**

| 属　性 | 各属性值概率(即 $p(x_k)$) |
|---|---|
| 聚类分析可疑标识 | $p_1(0)=0.978102, p_1(1)=0.021898$ |
| 小波分析可疑标识 | $p_2(0)=0.974153, p_2(1)=0.025847$ |
| 链接分析可疑标识 | $p_3(0)=0.974991, p_3(1)=0.025009$ |

**表 4　学习过程二**

| 属性 | 类 $C_2$ 即洗钱样本中各属性的概率(即 $p(x_k\mid C_2)$) |
|---|---|
| 聚类分析可疑标识 | $p_1(0\mid$洗钱$)=0.119718, p_1(1\mid$洗钱$)=0.880282$ |
| 小波分析可疑标识 | $p_2(0\mid$洗钱$)=0.176056, p_2(1\mid$洗钱$)=0.823944$ |
| 链接分析可疑标识 | $p_3(0\mid$洗钱$)=0.260563, p_3(1\mid$洗钱$)=0.739437$ |
| 类 $C_2$ 即洗钱概率 | $p($洗钱$)=0.016992$ |

经过训练学习,结合反洗钱领域知识,选定可疑度判定阈值 K 为 0.49。K 值为可疑度判定的下限值,K 值的最终确定是挖掘方法和交易信息两方面综合的经验值。K 值的选取犹如选择不同尺寸的滤网,K 值越大,网眼越大,识别结果的可疑度越高,但同时可疑交易线索相对较少,遗漏率较高。K 值越小,网眼越小,可疑交易线索较多,遗漏率较低,但同时识别结果的可疑度相对较低。因此,K 值的选择需要在训练数据可疑与否的判断正确率与挖掘效率间做出权衡,在保证较高的挖掘效率的前提下,促使正确率尽可能高、遗漏率尽可能低。在实际运用过程中,可以选择不同的阈值对交易记录进行处理,一

方面可以从中确定合理的 K 值，另一方面也便于从不同粒度对交易数据进行分析，深刻理解和把握源数据特征。

实验分别计算出训练数据和验证数据的正确率和遗漏率作为结果进行比较。实验结果显示，在总共 11939 条金融交易记录中（包括 1274 个账户，其中业已确认的犯罪线索记录 210 条），选择源数据的 70％记录作为训练数据（包括 8357 条金融交易记录，915 个账户，其中业已确认的犯罪线索记录 153 条），通过贝叶斯分析确定的可疑金融交易记录为 141 条，相对于业已确认的 153 条犯罪线索记录，准确率为 92.16％，遗漏率为 7.84％；将源数据的 30％作为验证数据（包括 3582 条金融交易记录，359 个账户，其中业已确认的犯罪线索记录 57 条），通过贝叶斯分析和可疑度判定值验证，确定可疑金融交易记录为 51 条，相对于业已确认的 57 条犯罪线索记录，准确率达 89.47％，遗漏率为 10.53％。与之相比而言，仅用离群点聚类分析、序列突变点检测或交易路径异常识别所得结果的准确率分别为 88.0282％、82.3944％、73.9437％（如表 4 所示）。这说明利用贝叶斯分类方法，能有针对性地综合单一数据挖掘方法的分析结果，提高可疑金融交易识别的准确率。

## 五、研究结论

研究结果证明整体判定方法在综合三种可疑金融交易检测方法的基础上运用贝叶斯准则进行整体判定是有效的，比单独应用一类检测方法的效果相对更好。这是因为贝叶斯分类通过计算完整的后验概率分布，充分汇总了可疑金融交易识别中涉及的各类可疑特征的全部信息，相对于单一可疑金融交易检测方法，在推理预测的准确性上有明显提高。

基于贝叶斯分类的可疑金融交易整体判定具有以下几方面的优势：一是有效利用各类识别结果，整合技术资源，形成优势互补，充分发挥基于数据挖掘的可疑金融交易识别能力。二是借助概率模型，很好地处理了不确定性，具备整体判定优势，能为反洗钱监测提供快速准确的参考。三是通过自适应性的监督学习，充分借鉴既往经验，与相关领域知识融合性好。四是对最新洗钱手法的变化反应灵敏，变被动识别为主动发现，在洗钱与反洗钱的博弈中争取先机。五是在大型数据库应用方面能达到高速度和高准确性较好的统一。基于以上五个特点，建立一种交易可疑度贝叶斯整体判定模式，能明显提高反洗钱监测的科学性和有效性。

由于洗钱交易与反洗钱监测存在博弈关系，识别方法必须能跟踪反映洗钱手法的最新动态。数据挖掘技术在可疑金融交易识别中的成功应用离不开交易数据的深入分析和算法的合理选择，更离不开反洗钱领域知识的熟习和理解。只有将丰富的反洗钱实践经验、权威的专家知识和数据挖掘技术优势相结合，才能建立一套基于数据挖掘的适合我国洗钱交易特征的可疑金融交

易综合识别方法，基于贝叶斯分类的交易可疑度整体判定模式正是这方面研究的有益尝试。

**注释：**

①概率 $P(x_1)$，$P(x_2)$，$P(x_3)$，…，$P(x_n)$可以由训练样本估值，其中 $P(x_k)=s_k/s$，$s_k$ 是在属性 $A_k$ 上具有值 $x_k$ 的训练样本数，s 是训练样本数。

②概率 $P(x_1|C_2)$，$P(x_2|C_2)$，$P(x_3|C_2)$，…，$P(x_n|C_2)$可以由训练样本估值，其中 $P(x_k|C_2)=s_{2k}/s_2$，$s_{2k}$是在属性 $A_k$ 上具有值 $x_k$ 的类 $C_2$ 的训练样本数，$s_2$ 是训练集中洗钱的样本数。

③外汇账户交易数据的取得经金融监管部门认可，对源数据涉及的客户名称、代码等敏感信息，在保证数据完整性的前提下，经一一对应的加密转换后仅供研究课题使用。

④改进的 CURE 算法针对可疑交易识别特征，充分利用 CURE 算法对孤立点处理的健壮性，将原算法聚类过程中删除的孤立点加以提取，建立可疑交易集合。参见笔者《基于 CURE 聚类的可疑金融交易识别研究》，《情报杂志》，2008 年第 6 期，第 52－54 页。

⑤训练数据基本按照总交易记录数的 70％抽取，同时为了便于分析，保证同一账户数据不被分割到训练和验证两部分，因此，在训练数据中，账户个数和业已确认的犯罪线索记录不一定是总账户数与犯罪记录数的 70％。

⑥小波分析在数学上具有严格意义上的突变点诊断能力，不依赖于经验模型，适合检测金融交易序列中的可疑成分。基于小波技术的交易序列突变点检测方法，借助小波分析方法在信号奇异性检测方面具有的独特优势，从金融交易序列中识别出具有异常交易行为的账户。

⑦基于链接挖掘的交易路径异常识别方法，是基于图论理论，通过图形遍历方式和金融交易权重分析策略的组合，识别不同交易主体间交易活动的内在联系，发现交易流向、来源以及用途或性质异常等交易关系异常特征。

**参考文献：**

[1]徐志春，肖伟平，何宏.数据开采技术在反洗钱系统中的应用[J].湖南工程学院学报，2003，(3)：64－67.

[2]汤俊.基于客户行为模式识别的反洗钱数据监测与分析体系[J].中南财经政法大学学报，2005，(4)：62－67.

[3]杨胜刚，王鹏.基于数据挖掘技术的人民币反洗钱系统设计[J].财经理论与实践，2005，(11)：105－109.

[4]陈云开.分布式异构计算环境下的洗钱侦测系统体系结构——基于数据挖掘技术[J].计算机工程与应用，2006，(29)：202－204.

[5]孙景，李志伟，刘炜．基于逻辑回归的企业大额可疑外汇资金交易识别模型[J]．上海金融，2008，(6)：58－61.

[6]中国人民银行．金融机构大额交易和可疑交易报告管理办法[Z]．中国人民银行令 2006 第 2 号．

[7]Senator T E，Goldberg H G，Wooton J. The financial crimes enforcement network AI system(FAIS)—— Identifying potential money laundering from reports of large cash transactions[J]. AI Magazine，1995，16(4)：21－39.

[8]Stofella, Paolo. Proceedings of the IEEE international workshop on research issues in data engineering[J]. AI Magazine, 1997, (5):73—75.

[9]Petrus C van Duyne, Hervy de Miranda. The emperor's clothes of disclosure:Hot money and suspect disclosures [J]. Crime, Law & Social Change, 1999, 31(3):245—271.

[10]Kingdon J ,Feldman K S. Data monitoring and analysis system for bank transactions, constructs aggregate profiles of received data and investigates to identify its characteristic patterns of behavior[C]. SEARCHSPACE LTD (SEAR-Non-standard), 2002.

[11]Kingdon J. AI fights money laundering [J]. IEEE Intelligent Systems, 2004, 19(3):87—89.

[12]N R Jennings. On agent-based software engineering [J]. Artificial Intelligence, 2000, 117(2): 277—296.

[13]I Horobin. Applying technology to fight money laundering [J]. Money Laundering Bulletin, 2001, (4):5—8.

# Research on the Recognition of Suspicious Financial Transactions Based on Bayes Classification

ZHANG Cheng-hu, ZHAO Xiao-hu

(*School of Economic and Finance, Xi'an Jiaotong University, Xi'an* 710061, *China*)

**Abstract**: The practice of anti-money laundering at home and abroad indicates that financial transactions are the important step of money laundering crime. The core of anti-money laundering is to analyze the customer information and transaction data of financial organizations, recognize suspicious transactions based on scientific methods and find the clues of money laundering. Firstly, the paper summarizes the features of suspicious financial transactions through multi-level analysis of financial transaction information. Secondly,aiming at the transaction information with different levels, it recognizes the suspicious financial transaction records by using suitable data mining methods and customer information. At last, by Bayes' law and suspicious information with different levels, it obtains the whole suspicious degree of transaction records and provide references to the supervision over anti-money laundering. The empirical results provide support for the validity of the method above.

**Key words**: anti-money laundering; data mining; suspicious financial transactions; Bayes classification

(责任编辑 喜 雯)