

文章编号:1001-9081(2009)09-2581-05

# 基于交易网络特征向量中心度量的可疑洗钱识别系统

喻 炜, 王建东

(南京航空航天大学 信息科学与技术学院, 南京 210016)

(yuwei\_36f@yahoo.com.cn)

**摘 要:**针对可疑洗钱识别问题,提出了基于交易网络特征向量中心度量的可疑洗钱行为检测系统。检测系统包含交易数据预处理、特征向量中心度量、时序检测分析三个关键子算法。同时还给出了三种用于检测异常交易行为的度量指标。通过对银行交易数据进行的实验仿真,验证了系统的有效性。

**关键词:**洗钱;反洗钱;可疑洗钱行为;特征向量中心度量;交易网络;时序分析;余量误差检测

**中图分类号:** TP274 **文献标志码:** A

## Suspicious money laundering detection system based on eigenvector centrality measure of transaction network

YU Wei, WANG Jian-dong

(College of Information Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing Jiangsu 210016, China)

**Abstract:** For anti-money laundering, a detection system based on eigenvector centrality measure was introduced, which included pre-processing of transaction data, eigenvector centrality measure and time-series analysis. Three key indexes of suspicious activities detection were also provided. Through the simulation on the transaction data of bank, the validity of the detection system was proved.

**Key words:** money laundering; anti-money laundering; suspicious money laundering activities; eigenvector centrality measure; transaction network; time series analysis; residual error detection

### 0 引言

由于洗钱犯罪对经济活动的巨大危害,世界各国都加大了对洗钱犯罪的打击力度,不但成立了相关的执法部门,而且还加强了信息技术在辅助反洗钱方面的学术研究和应用系统开发<sup>[1]</sup>。数据上报制度是反洗钱工作的核心措施<sup>[2]</sup>,但由于洗钱行为的隐蔽性、复杂性,反洗钱在实际开展中还是面临了很多问题<sup>[3-4]</sup>。

1)海量数据报送与高误报率。每天从全国各地报送的海量数据使得分析困难,效率与效果都无从保证。同时正常交易数据的增长对异常交易的识别会带来很大影响,正常数据对类似噪声的异常交易数据的“掩盖保护效应”也会更加严重;

2)上报制度很容易通过各种合法手段进行规避。比如,控制交易额使之小于上报额度、利用多个账户进行交易以控制单个账户的交易频度等。另外上报制度无法自动适应洗钱形势变化,无法体现交易周期等重要信息<sup>[5]</sup>。

3)洗钱交易数据的噪声特性。洗钱的交易行为都是隐藏在海量的正常交易数据之中,可疑交易占比极小,甚至可以视为一种噪声,很多洗钱行为都是在正常交易的掩护下完成的。如果直接使用数据挖掘技术,训练的数据太少从而特征不明显,而识别的数据又过大,从而造成识别的误报率很高。

4)现在的上报制度主要在还是集中在对单个账户进行监控,还没有拓展到整个交易网络的分析。现行报表也仅仅是大量交易记录的罗列,缺乏对这些数据的产生过程做出明确的解释说明能力,每条数据看不出所代表风险和嫌疑程度的高低水平,对调查起不到明确的指导作用。

### 1 数据挖掘技术在反洗钱中的应用分析

从国内外的研究情况来看<sup>[5-16]</sup>,数据挖掘技术虽然已经被大量引入到可疑洗钱行为识别的研究中,但是对于实际的效果还没有定论和评价标准。由于样本数据的严重缺乏也阻碍了研究工作的开展。综合国内外学者的研究成果,学者们提出了很多解决这些问题的算法,其中大多数都是分类和聚类算法,但对于交易网络的分析和研究还是比较稀少。从实际效果来看,现有的数据挖掘技术和反洗钱检测的实际需求还有很大的差距。

直接使用数据挖掘技术对上报的交易数据进行可疑洗钱行为的检测难度比较大。现提出以下假设:设数据挖掘算法的自身误差为  $\varepsilon_a$ ,可疑交易行为占全部交易比重为  $\varepsilon_t$ ,如果  $\varepsilon_a \gg \varepsilon_t$ ,则会导致识别无效。因此,应该有两个方面需要加强:

1)确立数据挖掘模型与实际数据的误差度量,即  $\varepsilon_a$  和  $\varepsilon_t$  的关系对识别效果的影响;

2)提供新的数据挖掘算法,或对交易数据进行转化或提取特征,对于转化和特征提取后的数据再进行传统方式的数据挖掘。

### 2 基于交易网络特征向量中心度量的检测系统

#### 2.1 识别检测算法概述

##### 2.1.1 交易实体和交易网络

首先引入实体  $E$  的概念,实体可以是一个银行账户或一个人,也可以是一个公司。在一个时间片段内,多个实体之间的资产或资金流动关系(流动方向和数量)可以视为一个交易网络,一个交易网络可以通过有向图或者矩阵方式进行表达。每

收稿日期:2009-04-07;修回日期:2009-05-25。

作者简介:喻炜(1981-),男,湖南临湘人,硕士研究生,主要研究方向:数据挖掘;王建东(1945-),男,江苏沭阳人,教授,博士生导师,主要研究方向:数据挖掘、机器学习、知识工程、人工智能。

个实体在交易网络都有一定的重要性并且扮演一定的角色。

2.1.2 系统的算法流程

由于反洗钱检测的时效性,算法必须能够较快地处理海量的交易,因此检测算法对交易数据进行时间分帧切片处理,帧长或时间切片一般为季度、月、半月。基于一个时间切片,再将交易数据转化为交易网络进行相关处理。图 1 为系统的算法流程。

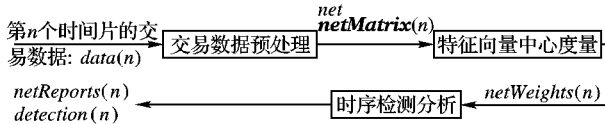


图 1 基于交易网络的识别检测系统

系统包含以下 3 个子算法。

1) 交易数据预处理:将无关联的第  $n$  个时间片的交易数据  $data(n)$  转化成为带有链接关系的全部交易网络的集合  $net$  和对应的交易矩阵  $netMatrix(n)$ ;

2) 交易网络的特征向量中心度量:将交易网络的资金流入、流出的要素特征进行了特征向量中心度量,即计算  $netMatrix(n)$  和  $netMatrix(n)^T$ , 将特征向量生成  $netWeights(n)$ 。

对于复杂网络进行特征向量中心度量,有助于发现复杂网络中的重要节点,因而具有重要的应用价值。特征向量中心度量已经在社会网络、病毒传播网络、交通网络和恐怖分子网络等方面的应用上有重大的实际意义<sup>[17]</sup>。

3) 可疑交易行为的时序检测分析:使用交易网络相似性周期搜索,寻找到交易网络的最小误差下的交易周期,形成交易报告  $netReports(n)$ ,通过误差余量进行可疑交易分析形成检测结果  $detection(n)$ ,时序检测分析的核心是余量的误差检测。

2.2 交易数据预处理

预处理包含交易网络初始化和交易网络重建两步操作。

2.2.1 交易网络初始化

通过递归算法将离散的交易数据  $MT\_data$  转化为包含有网络标志、实体标志的交易网络  $net$ 。

1) 数据结构。

① 交易数据,结构为:

$MT\_data = \{交易月度(M), 交易时间(rtt), 源实体账户(sa), 目标实体账户(da) 和 交易金额(amount)\}$ ;

② 交易网络集合,用于保存网络和实体的对应关系,一个实体只能对应一个网络,结构为:

$net = \{网络 ID(NID), 实体 ID(EID), 实体账户(EA)\}$ ;

③ 临时网络: $netTemp = \{实体账户(EA), 处理标志(PF)\}$ 。

2) 算法流程。

算法的基本流程如下:

① 读取  $MT\_data$  的第  $i$  条交易数据  $MT\_data(i)$ ,初始为第 1 条;

② 如果第  $i$  条交易数据没有处理,则读取  $MT\_data(i)$  中的  $sa$  和  $da$ ,

③ 判断  $sa$  和  $da$  是否已经在  $netTemp$  的  $EA$  存在,如果不存在则插入  $netTemp$  的  $EA$  集合中;

④ 获取  $netTemp$  中第一个实体  $EA$  的全部链接实体,即  $EA(1)$  全部的流入链接实体节点  $sa'$  和流出链接实体节点  $da'$ ,并且将  $sa'$  和  $da'$  插入到  $netTemp$  的  $EA$  集合中;

⑤ 令  $[m,n] = size(EA)$ ,即读取集合  $EA$  中实体的个数  $m$ ;

⑥ 令  $IdCount = 1$ ,即初始化  $EA$  的引用计数器;

⑦ while  $IdCount \leq m$ ;

⑧ 如果  $EA(IdCount)$  没有被处理,获取  $EA(IdCount)$  全部的流入链接实体节点  $sa'$  和流出链接实体节点  $da'$ ,并且将  $sa'$  和  $da'$  插入到  $netTemp$  的  $EA$  集合中;重置  $[m,n] = size(netTemp)$ ;

⑨ 累计计数器  $IdCount = IdCount + 1$ ,跳转到 ⑦;

⑩ 完成遍历后,将  $netTemp$  复制到  $net$  中,并且清空  $netTemp$ ,跳转到 ①。

2.2.2 交易网络重建

由于交易网络的实体提取是从交易记录中提取,那么在实际的操作中,经过对交易记录的多次处理后,需要重建交易网络操作。网络重建操作分为网络融合操作和插入操作,插入操作比较简单。

1) 数据结构。

① 原交易网络集合, $net\_old = \{网络 ID(NID), 实体 ID(EID), 实体账户(EA)\}$ ;

② 新交易网络集合, $net\_new = \{网络 ID(NID), 实体 ID(EID), 实体账户(EA)\}$ ;

③ 重建后交易网络集合, $net = \{网络 ID(NID), 实体 ID(EID), 实体账户(EA)\}$ ;

④ 临时网络: $netTemp = \{实体账户(EA), 处理标志(PF)\}$ 。

2) 算法流程。

① 算法的基本流程如下。

a) 读取  $net\_new$  的第  $i$  个交易网络  $net\_new(i)$ ,其  $NID$  为  $i$ ,初始为第 1 个;

b) 提取完整的交易网络  $NID$  全部的交易实体  $E$ ;

c) 遍历交易网络实体每一个  $E(j)$ ,判断  $E(j)$  是否已经存在  $net\_old$  中;

d) 如果有任一个  $E(j)$  存在  $net\_old$  中,融合到原网络  $net\_old$  中;

e) 如果没有任一个  $E(j)$  存在  $net\_old$  中,插入到原网络  $net\_old$  中;

f)  $i++$  后返回到第 a) 步。

② 网络融合。

a) 设  $NID$  为  $i$  和  $j$  两个交易网络  $net(i), net(j)$ ;

b) 令  $net(i)$  中的全部实体为集合  $EA, net(j)$  中的全部实体为集合  $EB$ ;

c) 如果有  $EA \cap EB \neq \emptyset$ ,则执行融合。

图 2 为网络融合的一个示例。

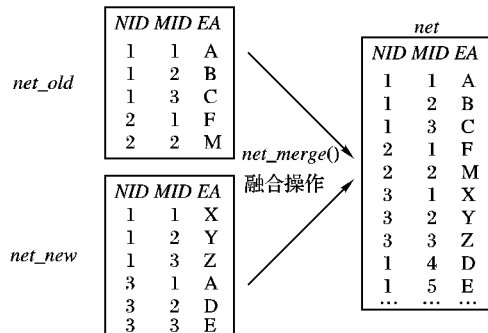


图 2 融合示例

2.3 交易网络的特征向量中心度量算法

2.3.1 交易实体重要性

在真实的洗钱行为中,很多交易都是通过化整为零、多重中间环节转账等方法进行的。因此交易网络中的异常交易行为对于可疑洗钱交易识别是非常重要的。图 3 是一个时间段内  $a, b, c, d, e, f$  六个实体账户之间资金流向构成的有向图。

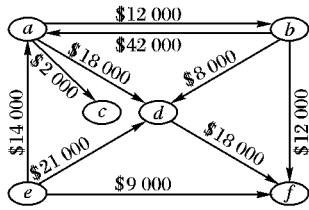


图 3 交易网络

令  $x_i$  代表第  $i$  个实体的重要性,  $X$  代表全部交易实体的重要性向量,  $A$  为交易矩阵,  $A(i, j)$  代表从第  $j$  个实体流向第  $i$  个实体的资产或资金。图 4 为相应的交易矩阵。

$$A = \begin{matrix} & \begin{matrix} a & b & c & d & e & f \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \\ f \end{matrix} & \begin{bmatrix} 0.00 & 42\,000 & 0.00 & 0.00 & 14\,000 & 0.00 \\ 12\,000 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 2\,000 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 18\,000 & 8\,000 & 0.00 & 0.00 & 21\,000 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 12\,000 & 0.00 & 0.00 & 9\,000 & 0.00 \end{bmatrix} \end{matrix}$$

图 4 交易矩阵

设第  $i$  个实体从第  $j$  个实体分到的重要性为  $W(i, j)x_j$ , 并且将所有流向第  $i$  个实体的重要性相加, 得到第  $i$  个实体最终的重要性, 即:

$$x_i = \sum_{j \in M(i)} W(i, j)x_j \quad (1)$$

其中  $M(i)$  是有资产或资金流向第  $i$  个实体的其他实体集合,  $W(i, j)$  是第  $i$  个实体从第  $j$  个实体得到的流入重要性的权重。令

$$D = \text{Diag} \left( \frac{1}{\sum_{i=1}^N A(i, 1)}, \frac{1}{\sum_{i=1}^N A(i, 2)}, \dots, \frac{1}{\sum_{i=1}^N A(i, j)} \right) \quad (2)$$

其中  $N$  是所有交易实体的数量。

可以证明:  $W = A \cdot D$ , 其中:

$$W(i, j) = \frac{A(i, j)}{\sum_{i=1}^N A(i, j)} \quad (3)$$

并且可以得到:  $X = W \cdot X = A \cdot D \cdot X$ , 可知  $X$  是  $W$  特征值为 1 的特征向量, 即一个交易网络中全部实体的重要性就是权重矩阵  $W$  特征值为 1 的特征向量。

为了保证  $W$  一定有特征值, 在计算之前可以给每个实体赋一个基本重要性  $\delta$ , 即:

$$x_i = pW(i, j) \sum_{j \in M(i)} x_j + \delta; \delta = (1 - p)/N \quad (4)$$

则

$$X = W \cdot X \Rightarrow p \cdot A \cdot D \cdot X + \delta \cdot e \quad (5)$$

设  $e = (1, 1, \dots, 1)^T$ ,  $\sum_{i=1}^N x_i = 1$ , 则:

$$1 = \sum_{i=1}^N x_i = e^T \cdot X \quad (6)$$

$$X = p \cdot A \cdot D \cdot X + \delta \cdot e \cdot 1 = p \cdot A \cdot D \cdot X + \delta \cdot e \cdot e^T \cdot X = (p \cdot A \cdot D + \delta \cdot e \cdot e^T) \cdot X \quad (7)$$

令

$$B = (p \cdot A \cdot D + \delta \cdot e \cdot e^T) \quad (8)$$

根据 Perron-Frobenius 定理<sup>[18]</sup>, 上述方程组的解存在并且唯一。

### 2.3.2 双向特征向量中心度量

对于交易矩阵  $A$ ,  $A(i, j)$  代表从第  $j$  个实体流向第  $i$  个实体的资产或资金。对  $A$  进行转置得到  $T$ , 即:

$$A(i, j)^T = T \quad (9)$$

则  $T(i, j)$  代表从第  $i$  个实体流向第  $j$  个实体的资产或资

金, 再进行特征向量中心度量计算, 可得到  $A$  的流出重要性。

根据式 (6) ~ (7), 令:

$$B_i = (p \cdot A \cdot D + \delta \cdot e \cdot e^T) \quad (10)$$

$$X_i = B_i \cdot X_i \quad (11)$$

令:

$$B_o = (p \cdot T \cdot F + \delta \cdot e \cdot e^T) \quad (12)$$

其中:

$$F = \text{Diag} \left( \frac{1}{\sum_{i=1}^N T(i, 1)}, \frac{1}{\sum_{i=1}^N T(i, 2)}, \dots, \frac{1}{\sum_{i=1}^N T(i, j)} \right) \quad (13)$$

流入、流出双向特征向量  $X_i$  和  $X_o$  可用于评价  $A$  的流入和流出特性。

### 2.4 时序检测分析

#### 2.4.1 相似性度量

同类事物之所以属于同一类, 在于它们的某些属性相似, 因此可选择适当的度量方法检测出它们之间的相似性。统计模式识别各种方法实际上都是直接或间接以距离度量为基础的。常用的距离度量是  $D$  维特征空间中的欧氏距离:

$$\delta(X_k, X_j) = \sqrt{\sum_{i=1}^D (x_{ki} - x_{ji})^2} \quad (14)$$

其中:  $X_k, X_j$  表示两个样本的特征向量;  $x_{ki}, x_{ji}$  则是相应的第  $i$  个分量。

通过将交易网络的特征向量中心度量, 可将复杂的交易结构进行了二维量化。一个交易网络发生交易行为的变化, 则对应的流入、流出双向特征向量也会发生变化:

设第  $n$  个时间片交易网络的流入、流出双向特征向量  $X(n)_i$  和  $X(n)_o$ , 第  $n-1$  时间片交易网络的流入、流出双向特征向量  $X(n-1)_i$  和  $X(n-1)_o$ 。根据式 (14) 可以定义出第  $n$  个时间片和第  $n-1$  时间片交易网络流入、流出相似度  $d(n, n-1)_i, d(n, n-1)_o$  和平均相似度  $e(n, n-1)$ :

$$d(n, n-1)_i = \delta(X(n)_i, X(n-1)_i) \quad (15)$$

$$d(n, n-1)_o = \delta(X(n)_o, X(n-1)_o) \quad (16)$$

$$e(n, n-1) = \frac{d(n, n-1)_i + d(n, n-1)_o}{2} \quad (17)$$

如果平均相似度  $e(n, n-1)$  大于特定的阈值, 则可以视为交易网络的交易状态发生改变。同样可以定义  $e(n, m)$  为第  $n$  个时间片和第  $m$  时间片交易网络的平均相似度。

#### 2.4.2 最小周期搜索

由于企事业单位的财务管理制度和银行信息系统的处理特征, 银行柜台中不论对私业务还是对公业务, 一般而言大量的正常交易都具有比较强的周期性。比如对私业务中的代扣水电费、代发工资、信用卡还款以及个人 ATM 终端存取款, 在现实的生活中都是有很强的周期性。而商务结算也更是如此, 如各种贸易结算、采购支出等。同样交易网络的交易特征也具备强周期性。一个交易网络在时间片为  $1, 2, \dots, n$  中有  $n$  个不同的双向特征向量中心度量, 因此一共可以得到  $n-1$  个平均相似度:

$$e_{(n,m)} = \frac{d(n, m)_i + d(n, m)_o}{2}; 1 \leq m \leq n-1 \quad (18)$$

$e_{(n,m)}$  为第  $n$  个时间片与第  $m$  个时间片的双向特征向量的平均相似度。为减少检测误报的情况, 在判断第  $n$  个时间片的交易是否发生异常时, 应该首先找到其最小周期  $p, 1 \leq p \leq n-1$ , 再进行异常交易行为的判断。

对于交易网络的第  $n$  个时间片的双向特征向量, 其最小周期  $p$  可以定义如下:

$$e_{(n)} = e_{(n,t)} = \min(e_{(n,1)}, e_{(n,2)}, \dots, e_{(n,n-1)}) \quad (19)$$

则:

$$p = n - \max(t) \quad (20)$$

2.4.3 余量误差检测

设第  $n$  个时间片的交易网络可以寻找到最小交易周期  $p$ , 对于周期性的交易网络, 则第  $n-1$  个时间片的交易网络可以寻找到最小交易周期  $p-1$ , 因此可以得到流入和流出的余量误差。

第  $n$  个时间片的流入、流出余量:

$$\mathbf{R}(n)_i = \mathbf{X}(n)_i - \mathbf{X}(p)_i \quad (21)$$

$$\mathbf{R}(n)_o = \mathbf{X}(n)_o - \mathbf{X}(p)_o \quad (22)$$

第  $n-1$  个时间片的流入、流出余量:

$$\mathbf{R}(n-1)_i = \mathbf{X}(n-1)_i - \mathbf{X}(p-1)_i \quad (23)$$

$$\mathbf{R}(n-1)_o = \mathbf{X}(n-1)_o - \mathbf{X}(p-1)_o \quad (24)$$

对于交易网络, 可以利用  $\mathbf{R}(n)_i$  和  $\mathbf{R}(n)_o$  来的范围定义异常交易检测的度量. 具体可以这样实现, 首先对向量  $\mathbf{R}(n)_i$  的每个分量求绝对值, 即:

$$\text{ABS}(\mathbf{R}(n)_i) = |\mathbf{R}(n)_i| \quad (25)$$

同理可以定义  $\text{ABS}(\mathbf{R}(n)_o)$ , 再对向量  $\text{ABS}(\mathbf{R}(n)_i)$  的分量求最大值, 即:

$$r(n)_i = \max(\text{ABS}(\mathbf{R}(n)_i)) \quad (26)$$

则  $r(n)_i$  代表第  $n$  个时间片的交易网络与其最相似交易网络的流入特征的误差变化范围, 同理可以定义  $r(n-1)_i$ ,  $r(n)_o$  和  $r(n-1)_o$ .

对一个交易网络对最近的交易进行余量误差检查, 可以定义出三种异常交易行为的度量条件: 1)  $e(n) > e(n-1)$ ; 2)  $r(n)_i > r(n-1)_i$ ; 3)  $r(n)_o > r(n-1)_o$ .

在实际使用中, 该过程还可以简化。对包含  $1, 2, \dots, n$  个切片的交易网络, 设  $e(n), r(n)_i, r(n)_o$  的最大值为  $\max(e), \max(r_i), \max(r_o)$ , 则: 1)  $e(n) > \max(e)$ ; 2)  $r(n)_i > \max(r_i)$ ; 3)  $r(n)_o > \max(r_o)$ 。

当满足一个条件时, 可以视为异常交易。余量误差检测主要利用了正常交易网络普遍存在的周期性和稳定性, 通过相似性误差、流入特征误差和流出特征误差进行异常交易行为的检测。

3 仿真实验

仿真工具采用 Matlab, 版本是 7.3.0 (R2006b)。仿真数据为从 2008 年 3 月 1 日至 2008 年 8 月 31 日的银行交易数据 12871 条, 总共涉及 998 个账户。检测度量条件为: 1)  $e(n) > \max(e)$ ; 2)  $r(n)_i > \max(r_i)$ ; 3)  $r(n)_o > \max(r_o)$ , 如果达到可疑标准则为  $S$ , 否则为  $N$ 。仿真最后生成每个交易网络的交易报告, 交易报告生成 3 个字符串, 分别代表 3 个检查条件的检测结果。实验总共进行 2 次: 实验 1, 数据被按月度进行切割, 一共生成 6 个时间片的数据, 以 6 个时间片为时间序列进行仿真模拟, 如表 1 所示; 实验 2, 数据被按半月进行切割, 一共生成 12 个时间片的数据, 再以 12 个时间片为时间序列进行仿真模拟, 如表 2。

对于  $n$  个时间切片数据, 可疑检测一共检测  $n-1$  次, 第一次都是生成交易网络的初始化操作。

表 1 中, 从左到右为网络 ID, 从上到下为 6 次或 12 次时间片的操作, 字段为空表明本次切片中该交易网络没有交易或者是第一次被初始化。

表 1 实验 1 的检测结果

网络	网络 ID												
	1	2	3	4	5	6	7	8	9	10	11	12	13
切片 1	—	—	—	—	—	—	—	—	—	—	—	—	—
切片 2	SNN	SNN	SNN	SNN	SNN	SNS	NNN	SNN	NNN	—	—	—	—
切片 3	SNN	NNN	NNN	NSN	NNN	NNN	NNN	NNN	NNN	—	—	—	—
切片 4	NNN	NNN	NNN	NNN	NNN	SNS	NNN	NNN	NNN	—	—	—	—
切片 5	NNN	NNN	NNN	NNN	NNS	NNN	NNN	NNN	NNN	—	—	—	—
切片 6	NNN	SNN	NNN	NNN	NNN	NNN	NNN	NNN	NNN	SSS	NNN	SNN	NNN

表 2 实验 2 的检测结果

网络	网络 ID												
	1	2	3	4	5	6	7	8	9	10	11	12	13
切片 1	—	—	—	—	—	—	—	—	—	—	—	—	—
切片 2	SNN	SNN	SNS	SNN	SNS	SNN	—	—	—	—	—	—	—
切片 3	NNN	NNN	NNN	NNN	SNN	—	NNN	NNN	SNN	—	—	—	—
切片 4	NNN	NNN	NNS	NNN	SNN	SNS	—	—	NNN	SNN	—	—	—
切片 5	NNN	NNN	NNN	—	—	—	NNN	NNN	SNN	—	—	—	—
切片 6	NNN	NNN	NNS	NNN	NNN	NNN	—	—	NNN	—	—	—	—
切片 7	NNN	NNN	NNN	—	—	—	NNN	NNN	NNS	—	—	—	—
切片 8	NNN	NNN	NNN	NNN	NNN	SNS	—	—	NNN	—	—	—	—
切片 9	NNN	NNN	NNN	NNN	—	NNN	NNN	NNN	NNN	—	—	—	—
切片 10	NNN	NNN	NNS	NNN	NNN	NNN	—	—	NNN	—	—	—	—
切片 11	NNN	NNN	NNN	—	—	—	NNN	NNN	NNN	—	—	—	—
切片 12	NNN	NNN	NNN	NNN	NNS	NNN	—	—	NNN	SSN	SNN	SNN	NNN

两次实验的趋势表明随着检测不断进行, 大部分网络的交易趋于稳定。同时两次实验的检测效果基本一致, 但是 12 次的交易切片更加能够体现一些交易的短周期性和相关细节。网络 10 的异常是由于静默了很久的账户突然发生了交易, 并且改变了实权重造成的。

通过对正常交易数据异常情况的识别可以说明本算法对异常交易敏感。

4 结语

本文的检测系统在反洗钱领域具有一定的创新性、开拓

性和实用性:

首先,通过具体的数据结构和算法流程介绍了交易网络的初始化和交易网络的重建;

其次,将交易网络的交易行为量化为特征向量,并且对交易网络进行流出和流入双向特征提取;

最后,使用了相似性度量对交易网络进行最小交易周期搜索,并将余量误差检测相结合的时序分析方法引入到异常交易行为的检测中。同时本文还给出了三种用于检测异常交易行为的度量指标。

算法除了具备良好的异常检测特性,还有其他两处优点:

1) 自适应性,算法不需要做训练就可以进行直接识别,这对完善上报制度中关于可疑交易有着极大意义;2) 压缩特性,算法从信息处理角度上来讲,就是一种有损压缩算法,由于反洗钱检测的处理数据量巨大,这为海量分析数据的存储带来便利。

#### 参考文献:

- [1] 欧阳卫民. 国际反洗钱的现状和趋势[J]. 中国金融, 2005(17): 29-32.
- [2] 张燕玲. 金融业反洗钱问题研究[J]. 国际金融研究, 2002(11): 4-12.
- [3] 刘兰英, 杨志. 中国金融业反洗钱问题研究[J]. 德宏师范高等专科学校学报, 2006(1): 28-32.
- [4] 钟会根. 基层人民银行反洗钱工作的难点与对策[J]. 金融与经济, 2007(3): 93-94.
- [5] 张成虎, 高薇. 朴素贝叶斯分类的可疑金融交易识别研究[J]. 情报杂志, 2006, 25(11): 46-51.
- [6] 汤俊. 基于可疑金融交易识别的离群模式挖掘研究[D]. 武汉: 武汉理工大学, 2007.
- [7] 汪素南. 智能技术在金融市场溢出效应和反洗钱中的应用研究[D]. 杭州: 浙江大学, 2007.
- [8] 仇芸斌. 基于CBR的反洗钱监控辅助决策支持系统研究[D]. 上

海: 复旦大学, 2007.

- [9] 陈起. 基于多 Agent 客户识别的反洗钱系统研究[J]. 计算机工程, 2007, 33(8): 190-193.
- [10] 张成虎, 赵小虎. 基于 CURE 聚类的可疑金融交易信息搜索研究[J]. 情报杂志, 2008, 27(6): 52-55.
- [11] 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): 58-62.
- [12] SENATOR T E, GOLDBERG H G, WOOTON J, *et al.* The FINCEN artificial intelligence system: Identifying potential money laundering from reports of large cash transactions [J]. AI Magazine, 1995, 16(4): 21-39.
- [13] TANG JUN, YIN JIAN. Developing an intelligent data discriminating system of anti-money laundering based on SVM [C]// Proceedings of the 4th International Conference on Machine Learning and Cybernetics. Washington, DC: IEEE Computer Society, 2005: 3453-3457.
- [14] TANG JUN. A peer dataset comparison outlier detection model applied to financial surveillance [C]// Proceedings of the 18th International Conference on Pattern Recognition. Washington, DC: IEEE Computer Society, 2006: 900-903.
- [15] GAO ZENGAN, YE MAO. A framework for data mining-based anti-money laundering research [J]. Journal of Money Laundering Control, 2007, 10(2): 170-179.
- [16] LIU XUAN, ZHANG PENG-ZHU, ZENG DA-JUN. Sequence matching for suspicious activity detection in anti-money laundering [C]// Proceedings of the IEEE ISI 2008 PAISI, PACCF, and SOCO International Workshops on Intelligence and Security Informatics, LNCS 5075. Berlin: Springer-Verlag, 2008: 50-61.
- [17] 王林, 张婧婧. 复杂网络的中心化[J]. 复杂系统与复杂性科学, 2006, 3(1): 15-20.
- [18] 房喜明, 黄荣里. 关于 Perron-Frobenius 定理的两个推论[J]. 湛江师范学院学报, 2004, 25(3): 9-11.

(上接第 2580 页)

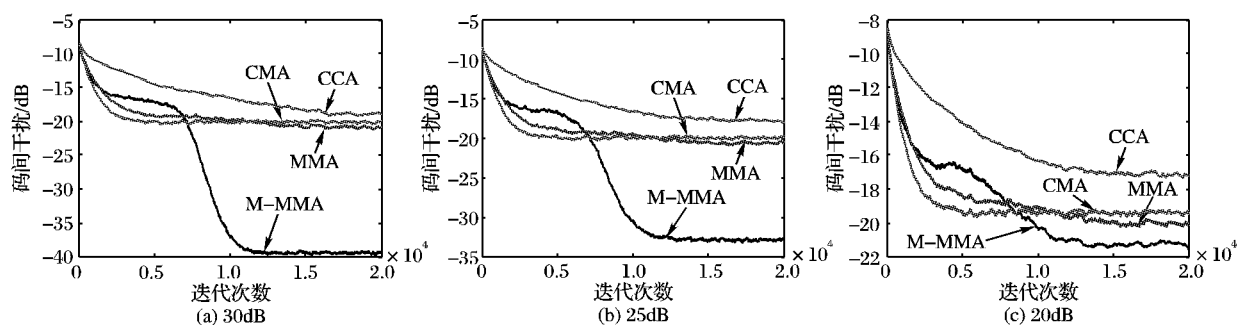


图4 不同信噪比下四种算法的 ISI 性能曲线(128QAM 信号)

#### 参考文献:

- [1] GODARD D N. Self-recovering equalization and carrier tracking in two-dimensional data communication systems [J]. IEEE Transactions on Communications, 1980, 28(11): 1867-1875.
- [2] OH K N, CHIN Y O. Modified constant modulus algorithm: Blind equalization and carrier phase recovery algorithm [C] // Proceedings of the 1995 IEEE International Conference on Communications. Seattle, WA, USA: IEEE Press, 1995: 498-502.
- [3] YANG J, WERNER J J, DUMONT G A. The multimodulus blind equalization algorithm [C] // Proceedings of the 13th International Conference on DSP. Santorini, Greece: IEEE Press, 1997: 127-130.
- [4] YANG J, WERNER J J, DUMONT G A. The multimodulus blind equalization and its generalized algorithms [J]. IEEE Journal on Se-

lected Areas in Communications, 2002, 20(5): 997-1015.

- [5] ABRAR S. Compact constellation algorithm for blind equalization of QAM signals [C] // Proceedings of the 2004 IEEE International Conference on Networking and Communication. Lahore, Pakistan: IEEE Press, 2004: 170-174.
- [6] 许小东, 戴旭初, 徐佩霞. 适合高阶 QAM 信号的加权多模盲均衡算法[J]. 电子与信息学报, 2007, 29(6): 1352-1355.
- [7] LABAT J, MACCHI O. Adaptive decision feedback equalization: Can you skip the training period? [J]. IEEE Transactions on Communications, 1998, 46(7): 912-930.
- [8] PICCHI G, PRATI G. Blind equalization and carrier recovery using a "stop-and-go" decision-directed algorithm [J]. IEEE Transactions on Communications, 1987, 35(9): 877-887.