数据挖掘在金融领域中的应用研究

张　焱　欧阳一鸣　王　浩　汪曦东
（合肥工业大学计算机与信息学院，合肥 230009）
E-mail：zy_79@sina.com.cn

摘　要　论文介绍了数据挖掘的几种主流技术，并将其应用于金融领域。针对金融领域中的反洗钱活动，分析了数据挖掘技术的应用特点，提出了一个实际的应用系统原型，论证了其中的一些关键技术，并给出相关的解决方案。该系统的实现对于防范和打击金融犯罪活动，具有重要的现实意义。

关键词　数据挖掘　金融犯罪　反洗钱　实时监控

文章编号 1002-8331-(2004)18-0208-04　**文献标识码** A　**中图分类号** TP311；TP18

Application of Data Mining in the Financial Field

Zhang Yan　Ouyang Yiming　Wang Hao　Wang Xidong
(School of Computer and Information，Hefei University of Technology，Hefei 230009)

Abstract：This paper introduces data mining and applies it to the financial field.We analyze the characteristic of data mining and propose a practical applying prototype in anti-money laundering.In details，we demonstrate some key techniques and provide relative solutions.Its realization has the significance for keeping away and striking financial crimes.

Keywords：Data Mining，financial crime，anti-money laundering，real time monitoring and controlling

1　引言

随着数据库技术的成熟和普及，人类积累的数据正快速增长，但对于这些数据还未能充分利用其价值，数据挖掘(DM——Data Mining)便应运而生。目前，数据挖掘技术已在市场营销、客户关系管理等领域得到应用，作为新兴技术，在金融领域中的应用研究也于近几年展开，内容涉及：银行客户关系管理、信贷风险预警、金融市场变化分析等方面[4]。

金融犯罪是当今业内面临的棘手问题之一，其中洗钱活动日益猖獗，严重威胁全球经济发展和国家安全。目前银行业对于洗钱的防范和打击主要是通过建立法律法规，银行的具体防范措施有：为客户办理业务应当核对其真实身份信息；对可疑外汇现金交易进行核查；对于大额现金的流动应警惕跟踪，把有洗钱嫌疑的账户列入“黑名单”等。可见，目前对金融犯罪的打击还处于人为干预阶段，仍未见有成熟的数据挖掘技术应用于其中。如果能建立一个相关的应用系统来自动地识别洗钱活动、分类洗钱账户、预警洗钱行为，将能大大提高反金融犯罪的效率和力度。

下面介绍论文提出的反洗钱数据挖掘系统，并论述其中的一些关键技术。

2　任务分析

2.1　金融领域中洗钱犯罪活动的特点分析

从金融机构的角度来看，洗钱是指犯罪集团或分子以银行或金融系统为媒介，利用其转移、储存、支付犯罪活动资金，以隐瞒或掩盖犯罪资金的来源、流向和违法性质的行为[6]。典型的洗钱交易过程是：(1)入账，即通过存款、电汇或其它途径把不法钱财放入金融机构；(2)分账，也就是通过多层次复杂的转账交易，使犯罪活动得来的钱财脱离其来源；(3)融合，以一项显示合法的转账交易为掩护，隐瞒不法钱财。通过这些过程，罪犯就可把非法所得转移并融合到有合法来源的资金中。

但实际上，犯罪分子洗钱的形式是多样的。例如，可以与境外公司签订虚假合同，如购货、合资在境外办企业，出口不收汇、进口不到货，将境内资金或权益转移到境外等等。

所以，我们需要在银行的大量数据上建立一个应用系统，在结构上，它能够完成数据挖掘的一般任务，在内容上，能够综合金融领域知识，有监督地发掘洗钱活动的各种模式，找出洗钱活动的规律和特点，得出相关知识，帮助人们自动识别出手段不同的洗钱行为。

2.2　反洗钱数据挖掘系统的主要功能

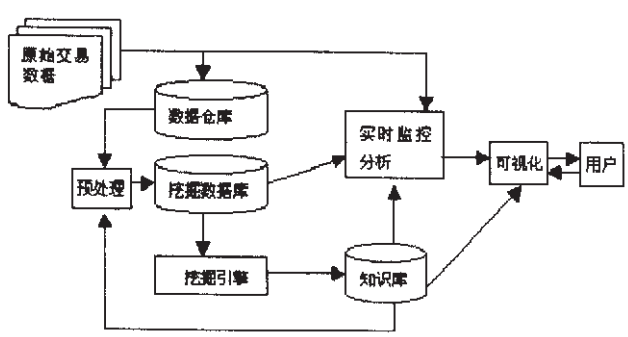

图1　反洗钱数据挖掘系统原型

反洗钱数据挖掘系统原型如图1所示，系统流程为：原始交易数据存储进数据仓库后，预处理模块在数据仓库中选取和

基金项目：安徽省教育厅自然科学重点科研项目资助(编号：2000jl168zd)

洗钱活动相关的数据，并用数理统计的方法产生有利于发现反洗钱的新属性。由于金融机构数据的海量性，需要额外为预处理后的数据增建一个数据库，有利于提高挖掘效率。挖掘引擎对预处理过的数据进行挖掘，发现知识。挖掘好的规则存入知识库备用，知识库中的规则和模式也可以经由可视化模块分析提交给管理者。作者增加了一个实时监控和分析模块，每当有新的交易数据产生时，该模块会参照挖掘数据库中的统计信息和已挖掘出的知识来判断这笔交易是否有洗钱的嫌疑。另外值得一提的是，在知识库和预处理两个模块间也有信息交互，因为账户的很多属性是不断变化的，例如，它的资金流量可能会随着企业的发展而增加，也可能由于经营不善而日渐萎缩，所以，根据挖掘出的规则有监督地调整预处理过程是十分必要的。

3 反洗钱数据挖掘系统

3.1 数据预处理

每次洗钱活动都牵涉到一段时间内多个账户之间的多笔交易，有效的预处理能够提取特征，利于下一步挖掘算法的高效运行。数据预处理可以从以下几个方面考虑。

(1)属性过滤。交易数据中并不是每个属性都和洗钱活动有关，例如开户人的姓名。也有一些无关的属性需要相关性分析技术来判断是否过滤该属性。问题的形式化描述如下：

设数据集 $D=\{d_1,\cdots d_n\}$，数据的属性集 S。现在检测 $x\in S$ 是否和目标属性 $y\in S$ 相关。

一个较直观的方法是，对于所有 x 的值排序后可以得到相应 y 属性值的分布。设 x_i 表示第 i 条记录的 x 属性值，$0<i<n$，定义距离函数 $F=\sum_{2}^{n}\sqrt{(x_i-x_{i-1})^2+(y_i-y_{i-1})^2}$，可以描述此分布的离散性，若超过设定的阈值，则认为这两个属性没有相关性，可以过滤属性 x。

(2)基于领域知识的特征提取。有些特征，如某时段内的资金流动量、出入账频率等，和洗钱行为密切相关，但在原始交易数据中没有记录，可以用统计的方法计算，作为下一步挖掘的参考属性。可以用线性回归模型描述这种关系：

记注册资金属性为 f，固定时段内的出入账总额为 s，f_i 和 s_i 表示账户 i 的相应属性，假设 f 和 s 存在近似线性关系 $f=\alpha+\beta s+\varepsilon$，其中 α,β 是常数，ε 服从正态分布 $N(0,\sigma^2)$。用最小二乘法估计 α,β，得：

$$\hat{\beta}=\frac{n\sum_{i=1}^{n}f_is_i-(\sum_{i=1}^{n}s_i)(\sum_{i=1}^{n}f_i)}{n\sum_{i=1}^{n}s_i^2-(\sum_{i=1}^{n}s_i)^2},\quad \hat{\alpha}=\bar{f}-\hat{\beta}\bar{s}$$

。(推导方法见文献[7])。

这样，就可以估计出某个行业内账户注册资金和资金流动量的线性关系。

(3)统计出和某个账户有资金往来的所有账户，参考领域专家的意见分析账户行业之间的相关性，以一个相关系数表示，作为下一步挖掘的参考属性。各行业的相关性可以先存储为一个矩阵形式，例如表 1。

(4)由于洗钱行为经常是和大笔现金或外汇的存入取出相关，所以，对于所有超过一定限额或频率的现金或外汇交易进行统计，结合相关的账户，作为挖掘的参考属性。

表 1

入帐方 \ 出帐方	钢铁	机械	食品	化工
钢铁	0.5	0.9	0.1	0.1
机械	0.7	0.5	0.6	0.7
食品	0.2	0.2	0.7	0.1
化工	0.4	0.3	0.3	0.5

3.2 挖掘算法

挖掘算法是整个系统最核心的部分，高效准确的算法能够发掘出深刻全面的知识。根据金融行业的数据组织和洗钱的特殊模式，可以把数据从微观到宏观分为交易层、帐户层、商业网络层这 3 个层次，分别用合适的方法挖掘出各个层次的规则，低层挖掘出的知识用于高层的挖掘过程，逐层向上，联合挖掘多层次的知识。

(1)交易层是最低层的数据，洗钱必定牵涉到一系列交易，比如现金存入、取出、电子交易、支票。这一层的挖掘任务主要在于提取各个帐户的交易特点和各行业的交易特点，主要方法是基于统计的方法。

对于特定账户的时间序列数据，可以利用数理统计方法发掘账户的行为规律。问题的形式化描述如下。设账户进货出帐额近似服从正态分布，$X\sim N(\mu,\sigma^2)$ 用子样中位数和极差估计参数 μ 和 σ，其中有定理：

若 Me 是样本 $X_1,X_2,\cdots Xn$ 的中位数，则对任意 x，有 $\lim_{n\to\infty}P\{\sqrt{\frac{2n}{\pi}}(Me-\mu)\le x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$，那么可以认为，当 n 很大时可取 $\hat{\mu}=Me$。另外，子样极差 R 的数学期望和方差分别为：$E(R)=dn\sigma$，$D(R)=vn^2\sigma^2$，那么可以推出 $\hat{\sigma}=\frac{R}{dn}$。($dn$ 和 vn 值参考文献[7])。

确定参数后，给定单侧置信度 α，那么置信区间即为 $(0,\mu+\mu_\alpha\frac{\sigma}{\sqrt{n}})$，若某笔交易金额太大，落在置信区间外，我们就有理由怀疑这笔资金的合法性。

假设检验的数理统计方法还能应用于账户出入账频率的概率统计方面，假若发现久置不用的账户突然频繁出入资金，或者有人频繁销户开户，且账户在注销之前有大量资金流动，也有理由怀疑该账户。

(2)帐户层数据是关于帐户交易特点的数据，这些数据一部分来源于预处理过程中提取的属性，一部分是对帐户在交易层上挖掘出的统计学特点。可行的挖掘方法是，按行业在大量不同账户之间进行分类或聚类分析。

典型的决策树分类问题描述如下。根据银行对历史数据的分析，为每个账户加上分类标签“可疑”和“正常”，就成为了训练数据。例如提取的表如表 2。

表 2 分类训练集

账户	注册资金	注册资金/月资金流动额	来往账户的相关性	是否可疑
1	<50 万	>1	0.3	N
2	50–500 万	0.8–1.0	0.6	Y
3	500–1000 万	0.6–0.8	0.8	N
…	…	…	…	…

设上述的训练集为 T，属性分别为 $A1,A2,A3,A4$，分类决策树的基本算法是：

步 1 :开始,T 中所有数据都在根结点。属性都被离散化为种类字段。

步 2 :选择一个基于启发式规则或统计度量,如信息增益(information gain)或基尼系数(Gini index),按此属性的分类将当前训练集分叉。

步 3 :以深度遍历的方法对每个分叉继续重复步骤 2,直至分叉后的训练集类别一致。

这样经过训练后建立分类模型,就可以对其他数据进行分类来判断账户是否可疑。

(3)商业网络层的数据挖掘。这一层数据描述的对象是有商务往来的若干企业和个人,他们的交易形成一个网络。因为一个完整的洗钱过程必定牵涉到多个帐户的多笔交易,那么对于这些帐户的关系及他们之间交易的特点进行分析是很有价值的。

拟先构造一个有向图数据结构描述账户间的关系,以及他们之间交易的特点,这些特点属性可以从交易层数据和账户层数据中统计得到,结构及形式描述如下:

节点结构:

```
node{
  Count :long ;//账号
  CountAttr :String ;//属性
  N :Int ;//出帐单位数量
  Out[1..N] :Line ;//出帐边
    }
```

边结构:

```
Line{
  Node :node ;//指向节点
  Acount :Int ;//平均月出帐额
  Freq :Int ;//平均月出帐次数
  S :float ;//月出帐次数的方差
    }
```

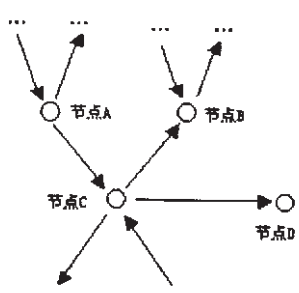

图 2　商业网络层的有向图表示

其中,每个节点代表一个帐户,帐户之间如有交易则用有向箭头连接。节点包含的信息有帐户的交易层特点、帐户层特点,箭头上蕴含的信息有这两个帐户之间的交易特点,例如交易频率、交易量等。

对于这样的数据,要在其中独立地挖掘出和洗钱有关的知识是不可行的,因为这种数据模型并不能体现洗钱的具体模式。但是,如能结合工商、海关等调查数据挖掘,会高效快速挖掘出可疑节点。例如,工商机构查出一个皮包公司,根据这个公司的账户构造上述有向图,我们可以查到所有和其有金钱来往的账户及其和这间公司的业务关系,结合这些账户的交易层和账户层数据,就能够大大缩小搜索范围、轻易地找出可疑的账户。笔者称这种方法为连接分析。

3.3　实时监控模块

实时监控模块的主要功能是,每当有新的交易发生时,就实时将这笔交易与该账户过去的交易对比,用已挖掘出的知识判断这笔交易或账户是否有洗钱的可疑。该模块存储挖掘出的规则,类似于一个专家系统。具体可以有两种方法。

(1)结合该账户已挖掘出的概率模型,利用数理统计的假设检验来验证该笔交易的可疑度,如果可疑度超过了用户设定的阈值,则提交可疑报告。方法类似于 3.2 节第一种挖掘方法。

(2)利用已分类或聚类挖掘的规则来判断该笔交易是否可疑。

4　实验结果

首先对数据预处理部分做了初步实验。原始数据表是某个机械制造业的企业账户在一段时间内的所有交易记录,共有 174 条,这些数据并非银行原始数据结构,为了研究方便经过一定处理,结构如表 3。

表 3

交易时间	转帐方向	对方帐号所属行业	金额
1997-06-30 15:46	入帐	生产资料制造业	108300
1997-07-14 8:00	出帐	私人帐号	800
1997-07-22 10:02	出帐	钢铁	24700
…	…	…	…

根据金融领域知识,以月平均为时间单位,拟提取的该账户的特征罗列如下:出账总额、入账总额、平均单笔出账额、平均单笔入账额,出账单位相关性、入账单位相关性。

各统计量计算方法如下:

```
Procedure PickChara(D :array of table,R :array of table)
Var OutSum,InSum,OutAver,InAver,OutRelat,InRelat :float ;
    I :interger ;
Begin
    Set all variables to 0 ;
    For i :=0 to 142 do
    Begin
      If direct='out'
      then Sum the variables OutSum and InSum ;
      else Sum the variables InSum and InRelat ;
    End ;
    Set OutAver=OutSum/(the number of payment) ;
    Set InAver= InSum/(the number of gathering) ;
End ;
```

算法中的变量 OutSum,InSum,OutAver,InAver,OutRelat,InRelat 即是要得到的统计量。该算法用 VC++实现,运行的环境是 P4-2G/256M,在 174 条记录数据上运行时间是 10 毫秒,实验结果如表 4。

表 4

平均出帐笔数	平均入帐笔数	平均每笔入帐金额	平均每笔出帐金额	平均入帐相关度	平均出帐相关度	入帐总额	出帐总额
14	15	9394	7166	0.73	0.51	272446	207836

结果表明从原始交易数据中能够提取基于领域知识的统计量,把大量账户的这些统计量计算出来并离散化后,就可以进行下一步账户层的分类或聚类挖掘。作者后续的实验就是从更多账户交易记录中提取特征统计量,然后再一起进行聚类和分类挖掘。

5 进一步的工作

论文从理论上提出了反洗钱数据挖掘系统的框架，下一步，将在模拟数据上分析和实现各个功能模块，结合已有的算法理论，探索出适合反洗钱的有效算法。

另外，反洗钱领域仍有许多挑战，包括：

(1)与工商、海关、税务部门的合作和数据共享。洗钱行为通常都涉及到多家金融机构之间的资金流动，以及国内和海外企业的经营，如果能集合上述相关部门的相关数据，无疑会提高反洗钱数据挖掘系统的精度和效率。

(2)由于银行拥有海量数据，并且每天都有大量交易数据增加和更新，在此情况下开发高效实时的挖掘算法和监控模块将是很大的挑战。

(3)该方案采用了不同的挖掘算法，它们丰富了挖掘出的规则，起到互补的作用，但规则之间是否存在矛盾仍然未知，有待于在实践中验证。

6 结论

论文针对金融领域的反金融犯罪方面，尤其是反洗钱活动，分析了数据挖掘技术的应用，提出了一个系统原型，具有现实意义。(收稿日期：2003 年 11 月)

参考文献

1.Jiawei Han，Micheline Kamber.Data Mining Concepts and Techniques [M].Morgan Kaufmann Publishers，2000

2.刘红岩，陈剑，陈国青.数据挖掘中的分类算法综述[J].清华大学学报，自然科学版，2002，42(6)

3.Zijian Zheng，Ron Kohavi，Llew Mason.Real World Performance of Association Rule Algorithms[C].In：ACM SIGKDD international conference on Knowledge discovery and data mining，2001

4.金融时报网.http：//www.financialmews.com.cn/

5.金网在线.http：//www.fcc.com.cn/

6.中国人民银行.金融机构反洗钱规定；人民币大额和可疑支付交易报告管理办法.http：//www.pbc.gov.cn/jinrongfagui/

7.汪荣鑫.数理统计[M].西安交通大学出版社，1986

(上接 165 页)

的生成，文本数字水印算法。

该方案及其所用算法有如下特点。

可靠性高：由于诸如修改语句、添加语句、删除语句、移动语句等语句修改对原文破坏操作量小，这些小量的修改对整个待嵌入水印文本的语句分布影响甚微，因此这种根据语句分布来定位水印信息的位置有较高的可靠性。

抗攻击性强：在修改、添加、删除语句的攻击下，最好的情况是可以完整提取水印；在平均情况下，修改两个语句破坏 b 位水印信息的概率为 $9(a/b*a/b)/(n*n)$，如果语句数比水印信息位数大很多，这个数会很小。

对移动语句的攻击，最好的情况是可以完整提取水印；平均情况为修改两个语句破坏 b 位水印信息的概率达 $2*9(a/b*a/b)/(n*n)$(a 为水印信息的位数，即字数乘以 16)。当水印信息为 12 个汉字，文本为 10000 句时，用移动语句攻击两个语句破坏一位水印信息的概率为 0.0066。

关联性有保证：各版权实体嵌入水印时使用的算法不同，只有拥有密钥信封的密钥才能提取水印，且提取的水印信息里有版权实体独一无二的标志信息，这就保证了水印嵌入和提取过程的独立性和非干扰性。水印信息里含有的版权实体间相互协商的信息、时戳签名及信息摘要都可用于说明各版权实体的水印信息在 eBook 网络传播过程中的连贯性及不可抵赖性。

获得良好的继承性：一方面，各版权实体的水印信息内容里有对时戳的签名，这联系着他们之间嵌入的水印信息在时间上的顺序性和事件过程上的继承性(即发行商秘密水印信息的时戳必然晚于出版商的)。同时，水印信息内容里的文件摘要也为保证继承性提供依据；另一方面，各阶段所应用的算法也能体现继承性，在出版阶段和发行阶段，针对各阶段面临的不同问题，使用的是不同的水印技术，比如数字水印和数字指纹，而且每次嵌入的水印都是依据不同的特征。这样，水印信息的内容结合水印技术本身便可以唯一地确定具体环节和版权。可以通过提取不同阶段的秘密水印来确认所有权或泛滥渠道。

兼容性好：该方案所用水印算法，支持向后兼容，一旦新算法出现，仅需要做一下判断，就可将老版本的水印算法包括进去。

实用性强：该方案嵌入和提取简单易行，输入密钥信封或在便携式阅读器插入密钥卡即可提取水印。

由于以上特点，该方案和水印技术有着广阔的前景。不可否认，由于当前数字出版的相关法律还不完善，遇到具体问题时还会有这样那样的问题。因此还需要有关法律来支持数字出版中的技术问题。

5 结语

目前的水印技术大多是针对所有权的证明，还没人提及网络出版中版权保护还需要解决关联性与继承性的问题，针对此现状，提出了参考模型，建议了水印信息的格式与内容，并给出了一个解决方案来说明各种性能的获得。具体实现了水印系统里的水印密钥信封的生成，多重文本水印嵌入与检测算法，版权实体间契约的协商。该方案能够将文本水印完整地贯穿于网络出版的各个环节中，一旦发现盗版，作者、出版商、发行商、用户就可以根据水印信息追查盗版者，并提取水印信息作为诉诸法律的依据。

另外，文本水印算法本身还有许多需要完善的地方，需要许多自然语言处理工具，需要改进算法增加嵌入信息量等。虽然文本水印目前还没有达到有效而广泛应用。但随着数字版权管理的不断完善，文本水印必将成为网络出版解决方案中有效的版权保护技术手段之一，从而推动网络出版的发展。

(收稿日期：2003 年 7 月)

参考文献

1.W Bender，D Gruhl，N Morimoto et al.Techniques for data hiding[J].IBM Systems Journal，1996，35：332~335

2.van Schyndel R G，Tirkel A Z，Osborne CF.A digital watermark[C].In：Proceedings of the 1994 IEEE International Conference on Image Processing，1994-02：86~89

3.S H Low，N F Maxemchuk.Performance Comparison of two text marking methods[J].IEEE Journal on Selected Areas in Communications，1998，16：561~572

4.卢开澄.计算机密码学[M].第 2 版，清华大学出版社，1998-07

5.Atallah M J，V Raskin，M Crogan et al.Natural Language Watermarking：Design，Analysis，and a Proof-of-Concept Implementation[C].In：I S Moskowitz ed.Information Hiding：4th InternationalWorkshop，IH 2001，Pittsburgh，PA，USA，Proceedings，Berlin：Springer，2001-04：185~199

数据挖掘在金融领域中的应用研究

万方数据 WANFANG DATA 文献链接

作者： 张焱， 欧阳一鸣， 王浩， 汪曦东
作者单位： 合肥工业大学计算机与信息学院,合肥,230009
刊名： 计算机工程与应用 ISTIC PKU
英文刊名： COMPUTER ENGINEERING AND APPLICATIONS
年，卷(期)： 2004,40(18)
被引用次数： 19次

参考文献(7条)

1. Jiawei Han;Micheline Kamber Data Mining Concepts and Techniques Morgan Kaufmann Publishers 2000

2. 刘红岩;陈剑;陈国青 数据挖掘中的分类算法综述[期刊论文]-清华大学学报（自然科学版） 2002(06)

3. Zijian Zheng;Ron Kohavi;Llew Mason Real World Performance of Association Rule Algorithms 2001

4. 查看详情

5. 查看详情

6. 中国人民银行 金融机构反洗钱规定;人民币大额和可疑支付交易报告管理办法

7. 汪荣鑫 数理统计 1986

本文读者也读过(4条)

1. 李志强.李灵 数据挖掘技术在金融反洗钱中的应用[期刊论文]-商场现代化2007(8)

2. 李金迎.詹原瑞 金融行业的数据挖掘技术研究[期刊论文]-现代管理科学2009(8)

3. 杨胜刚.王鹏.YANG Sheng-gang.WANG Peng 基于数据挖掘技术的人民币反洗钱系统设计[期刊论文]-财经理论与实践 2005,26(6)

4. 丁宁.刘富星 数据挖掘在金融中的应用[期刊论文]-广西轻工业2009,25(7)

引证文献(19条)

1. 贺兴时.于洁琼.李丽丽 基于互信息的特征子集选择[期刊论文]-西安工程大学学报 2008(3)

2. 李玉华.易鑫.孙小林 基于图熵的链接发现算法在反洗钱领域的应用[期刊论文]-计算机工程与科学 2007(11)

3. 欧阳一鸣.周强.胡学钢.江擒虎 数据挖掘中动态改变样本域提高预测精度[期刊论文]-合肥工业大学学报（自然科学版） 2006(4)

4. 周强.欧阳一鸣.胡学钢.王浩 数据挖掘中应用偏最小二乘法发现异常值[期刊论文]-微电子学与计算机 2005(1)

5. 徐宏宁.李代平.何利明.熊建斌 银行反洗钱系统的研究[期刊论文]-微型机与应用 2010(9)

6. 杨华明 数据挖掘在商业银行创利分析的应用[学位论文]硕士 2005

7. 张成虎.赵小虎 基于决策树算法的洗钱交易识别研究[期刊论文]-武汉理工大学学报 2008(2)

8. 汤俊 自适应反洗钱辅助信息系统开发框架设计[期刊论文]-管理学报 2005(z1)

9. 周强 改进C-均值算法实现时序数据模式自动生成[期刊论文]-黑龙江科技信息 2007(16)

10. 赵小虎.张成虎 可疑外汇交易的分类识别方法及其应用[期刊论文]-西安交通大学学报（社会科学版） 2008(5)

11. 张成虎.赵小虎 基于CURE聚类的可疑金融交易信息搜索研究[期刊论文]-情报杂志 2008(6)

12. 陈起.崔颖安.崔杜武 基于多Agent客户识别的反洗钱系统研究[期刊论文]-计算机工程 2007(8)

13. 支永安.欧阳一鸣 回归分析在安徽电信差异化服务中的应用[期刊论文]-合肥工业大学学报（自然科学版） 2011(3)

14. 刘晶晶 数据挖掘在我国商业银行中的应用[学位论文]硕士 2005

15. 吕昀卿 商业银行客户细分及应用研究[学位论文]硕士 2006

16. 贺超波.陈启买.石玉强.闫大顺 实验教学数据管理与分析系统的研究与实践[期刊论文]-实验技术与管理 2012(1)

17. 朱彦 数据挖掘在国税信息化中的应用[学位论文]硕士 2004

18. 黄晓东 数据仓库技术在证券行业的应用[学位论文]硕士 2005

19. 程照星 数据挖掘在电信企业客户细分中的应用[学位论文]硕士 2004

本文链接：http://d.g.wanfangdata.com.cn/Periodical_jsjgcyyy200418066.aspx