第 33 卷 第 8 期 计 算 机 工 程 2007 年 4 月
Vol.33 No.8 Computer Engineering April 2007
·人工智能及识别技术· 文章编号：1000—3428(2007)08—0190—03 文献标识码：A 中图分类号：TP18

基于多 Agent 客户识别的反洗钱系统研究

陈 起，崔颖安，崔杜武
(西安理工大学计算机科学与工程学院，西安 710048)

摘 要：针对现行的反洗钱系统的缺陷，该文将模式识别 Agent 引入反洗钱系统设计，充分发挥 Agent 系统自主性、反应性、协作性的特点，构造了基于多 Agent 客户识别的反洗钱系统模型。并对多 Agent 通信、协商、实现与支持向量机模式识别等关键技术进行了研究。从而为反洗钱系统的实现提供了支持。

关键词：反洗钱；多 Agent；模式识别

Research on Anti-money Laundering System Based on Multi-agent and Customer Pattern Recognition

CHEN Qi, CUI Yingan, CUI Duwu
(School of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048)

【Abstract】 Aiming at the faults of current anti-money laundering system, this paper takes full advantage of agent system which is autonomy, reactivity and initiative to constructs the model of anti-money laundering system based on multi-agent with pattern recognition. And researches on Rey problems such as communication, negotiation and realization of multi-agent and pattern recognition using SVM are made to offer support for realization of anti-money laundering system.

【Key words】 Anti-money laundering; Multi-agent; Pattern recognition

洗钱通常是指为了掩盖非法收入的真实来源和存在，通过各种手段使其合法化的过程。美国的FASI、澳大利亚的Screen IT反洗钱系统，均采用专家系统、人工智能的方式高效识别未知的、潜在的可疑金融交易行为，发现可疑的客户。然而，我国的反洗钱工作仍然以传统的手工方式和用信息技术对有关交易报告信息进行简单的汇总为主，利用信息技术辅助反洗钱工作的水平仍然较低。我国现行的反洗钱数据报告制度，规定在人民币和外汇交易的报告制度中超出一定数额的交易属于大额交易，金融机构必须加以记录并向上级主管部门报告。报告制度同时还规定了可疑支付报告要求。该系统具有以下弊端[5]：(1)巨量数据报表与高误报率。(2)预设标准易于为洗钱分子规避。(3)无法自动适应洗钱形势变化。(4)缺乏对上报数据的解释能力。(5)不能跨部门全局监控。

本文提出了一种基于多 Agent 客户识别的反洗钱系统，有效地解决了以上问题，将为我国反洗钱系统的建设提供一种全新的思路。

1 系统模型

1.1 系统结构

该系统为数据采集层、业务处理层、应用层 3 层结构，其中包含识别 Agent、识别服务 Agent 两种 Agent。采集层各 Agent 与业务层识别服务 Agent 分工协作，通过对业务数据的处理与知识交流，实现联合多个部门，甄别洗钱行为与用户的功能。系统结构如图 1 所示。识别 Agent 分别具备各自部门的领域知识，用客户模式识别的方式，识别各自数据系统中的洗钱行为，同时接受识别服务 Agent 管理，与其他部门识别 Agent 合作协商。

Agent 服务提供黑板作为通信模型，同时是识别服务 Agent 的容器，包含如下模块：

(1)识别服务 Agent 本身具有数据库与知识仓库，分别存放从识别 Agent 与黑板获取的洗钱交易数据与知识。通过读取 Agent 配置信息，决定系统中启用的识别 Agent，管理启用 Agent 间的通信，综合各识别 Agent 的识别结果，管理 Agent 知识库的扩充，协调 Agent 协商得出最终结论。

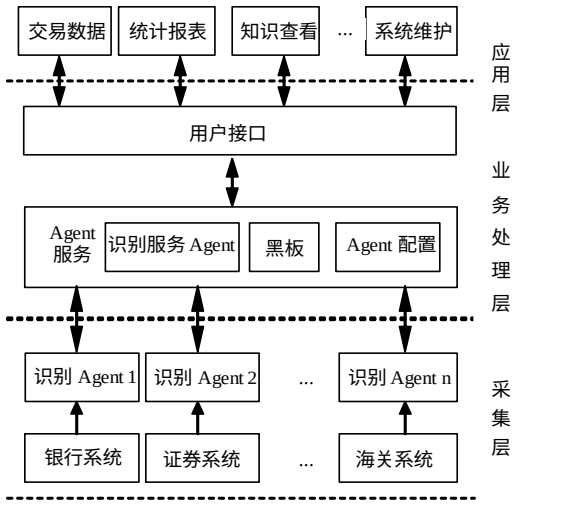

图 1 系统整体结构

基金项目：陕西省教育厅科研基金资助项目(03JC16)
作者简介：陈 起(1979－)，男，硕士生，主研方向：数据挖掘，J2EE；崔颖安，硕士、讲师；崔杜武，博导
收稿日期：2006-05-15 **E-mail：**xachenqi@163.com

(2)黑板是分布式人工智能中经常采用的通信模型。黑板有一个公共区与若干专用区组成，每个Agent都拥有各自独立的专用区存储各自的私有数据与中间信息，公共区的数据用于共享使用，并负责对Agent通知有用的中间识别结果、最终识别结果、通知发送者与接收者[4]。

(3)Agent 配置通过与用户接口的交互，产生启用识别 Agent 的列表，并负责增加与删除识别 Agent 服务。

通过对用户接口的调用，终端用户能够获得异常交易数据、统计报表、规则知识的信息，并维护系统整体运行。

1.2 识别 Agent 的结构

识别 Agent 具有用户接口、数据采集、识别知识库、推理机、本地数据库与通信接口 6 个模块。用户接口负责接收用户请求，数据采集模块负责采集所在子系统业务数据，推理机负责识别运算，识别知识库准备支持向量机算法与静态洗钱规则，本地数据库保存识别出的业务数据，通信接口负责与业务处理层、各识别 Agent 通信。其结构如图 2 所示。

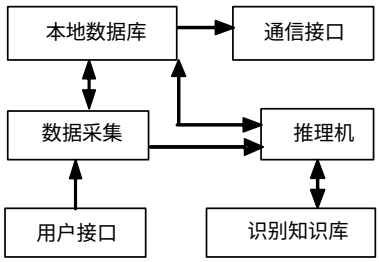

图 2 识别 Agent 结构

1.3 系统工作流程

系统工作流程如下：

(1)系统启动，根据 Agent 配置信息，由识别服务 Agent 决定启用的识别 Agent，并启动黑板模型负责 Agent 通信，启动识别服务 Agent 等待各系统采集信息。

(2)各部门定时扫描日常交易数据，由具备各自领域知识的识别 Agent 甄别出洗钱交易数据与知识，并将各自获取的信息发送至黑板公共区与专用区，同时附上交易人实名信息。

(3)识别服务 Agent 扫描黑板公共区，检测交易人实名信息，将与此人相关的洗钱交易数据保存至自身数据库并发各识别 Agent 协商，将送回的协商结果保存到自身数据库，发现的知识保存到自身知识库。同时更新各识别 Agent 的知识库。

(4)通过用户接口，相关人员可以执行管理 Agent 配置、查看洗钱交易数据、查看洗钱行为统计报表、查看洗钱知识规则等操作。

2 实现中的关键技术

2.1 Agent 相关技术

Agents 是一种在分布式系统或协作系统中能持续自主的发挥作用的计算实体，通常称为智能体。本系统中各业务部门通常独立运作。利用 Agent 自主性、交互性、反应性和主动性的特征，整合多个识别 Agent 运算能力，从而构造出基于多 Agent 的客户识别反洗钱系统。

2.1.1 多 Agent 通信机制

系统采用 KQML(Knowledge Query and Manipulation Language)语言处理多个 Agent 间信息格式与消息处理协议，该语言包含一系列可扩充的行为原语，定义了 Agent 对知识的目标和各种操作，在其上可以建立 Agent 互操作的高层模型。主要用于支持 Agent 之间知识与信息的实时交流，确保多个 Agent 共同协商取得最终论断。KQML 分为内容层、消息层和通信层、其一般结构如下：

{MSG
：Type<消息的类型：查询或断言>
：Qualifiers<消息的限定词：sender，receiver 等>
：Languages<所用的语言>
：Topic<知识的主题>
：Content<消息的实际内容>
}

2.1.2 多 Agent 跨部门协商机制

为实现跨部门的协同甄别，需要在由识别服务 Agent 组织各部门识别 Agent 协同分析，识别服务 Agent 将本次扫描中全部的实名信息发送至黑板公共区，与黑板中各识别 Agent 的专用区对比，并将补集信息送至识别 Agent 推理机，抽取与该实名信息相关的客户交易记录，各识别 Agent 诊断结果汇总至识别服务 Agent，仲裁识别结果，更新数据库与知识库。简单的识别服务 Agent 仲裁机制可以使用如下的方式：

每个识别 Agent 的识别结果表示为 CONCLUSION=<异常描述><可信度>

识别服务 Agent 的仲裁结果表示为

ARBITRATION=a*CF+b*C+c*T

线性组合，其中 a,b,c 为待定系数，CF 为可信度的均值，C 为一致度的均值，T 行业相关度的均值。设 ω_i 为各部门权值，则

$$\sum_{i=1}^{n}\omega_i = 1$$

$$CF = \sum_{i=1}^{n} CF_i * \omega_i$$

CF_i 为各识别 Agent 的可信度，一致度用 CF 的均差

$$C = \sum_{i=1}^{n} |CF - C_i| * \omega_i$$

表示。

$$T = \sum_{i=1}^{n} T_i \omega_i$$

T_i 为各识别 Agent 行业相关系数。根据实际应用的需要，需要适当地调整关于一致度的计算方式，设置行业相关度的值。

2.1.3 Agent 实现

基于 KQML 的 Agent 通信体系结构，其精髓在于构建一个实时的、分布式的网络化计算环境。在实现时可以将 Facilitator 和 Message Router 转换为 CORBA 对象，并为这些对象创建应用程序访问接口，编程语言可采用 Java 或 C++。CORBA 规范中的对象请求代理(ObjectRequest Broker, ORB) 可以用于 Agent 之间的定位，接口定义语言 (Interface DefinitionLanguage, IDL)可用于 Agent 内部对象之间的通信，同时对象服务可以用于诊断 Agent 的命名、复制、删除等 Agent 的管理。考虑到系统分布式的异构数据，CORBA 的平台无关性和语言无关性对分布式诊断系统的研制和功能实现至关重要。

2.2 模式识别

模式识别是随计算机技术的发展形成的一种模拟人的各种识别能力和方法技术。基本上属于自动判别与分类的理论。本系统引入模式识别技术中的支持向量机(Support Vector Machine, SVM)，将识别出的业务属性作为 SVM 的运算维，并利用其完成对识别 Agent 推理机的构造。

2.2.1 识别属性的选取

洗钱活动总是有其特定的规律，以银行系统为例，异常的行为通常涉及到一个较短时间段内多个账户的频繁交易，如何从海量的交易信息中发现可疑的数据，交易数据属性的识别选取，成为第 1 步，也是最为关键的一步。据此提出属性选取的参考[7]：

(1)基于业务常规的选取。频繁的开户、销户、存款、取款；多个账户款项短时间内向同一账户聚集或者同一账户款项短时间内向多账户分散。

(2)基于行业相关性的选取。参照行业相关系数，判断双方交易是否正常。此外，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={$d_1,d_2,...,d_n$}，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

$$F = \sum_{2}^{n} \sqrt{(x_i - x_{i1})^2 + (y_i - y_{i1})^2}$$

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。

(3)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'的相应属性，假设f和s存在近似线性关系：$f = \alpha + \beta x + \varepsilon$，其中$\alpha$，$\beta$是常数，$\varepsilon$服从正态分布。用最小二乘法估计$\alpha$、$\beta$，得

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} f_i s_i - (\sum_{i=1}^{n} fi)(\sum_{i=1}^{n} s_i)}{n\sum_{i=1}^{n} s_i^2 - (\sum_{i=1}^{n} s_i)^2}$$

$$\hat{\alpha} = \bar{f} - \hat{\beta}\bar{s}$$

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。

2.2.2 支持向量机识别客户模式

洗钱交易客户行为模式识别的主要困难是数据量大、需要分析的特征变量过多、缺乏训练样本集。洗钱手法、客户经营活动的多变性也决定了没有一成不变的判别模式。因此，此类客户行为模式是一种非监督(unsupervised) 判别模式。而由统计学习理论(Statistical Learning Theory, STL)发展起来的SVM 成为解决此问题的出路。

SVM是从线性可分情况下的最优分类面发展而来的，基本思想是使用最优分类线H将两类正确分开(训练错误率为0)，而且使分类间隔最大，假定H_1、H_2分别为过各类中离分类线最近的样本且平行于分类线的直线，则它们之间的距离叫做分类间隔。分类线方程为

$$\boldsymbol{x} \cdot \boldsymbol{w} + b = 0$$

对它进行归一化，使得对线性可分的样本集$(\boldsymbol{x}_i, y_i)$，$i = 1, ..., n$，$\boldsymbol{x} \in R^d$，$y \in \{+1, -1\}$，满足

$$y_i[(\mathbf{w} \cdot \mathbf{x}_i) + b] - 1 \geq 0, \quad i = 1, \cdots, n$$

此时分类间隔等于 $2/\|\boldsymbol{w}\|$，使间隔最大等价于使$\|\boldsymbol{w}\|^2$最小。满足上述条件且使$\frac{1}{2}\|\boldsymbol{w}\|^2$最小的分类面就叫做最优分类面，$H_1$、$H_2$上的训练样本点就称作支持向量。利用Lagrange优化方法可以把上述最优分类面问题转化为其对偶问题，即得到[6]

$$f(\mathbf{x}) = \mathrm{sgn}\{(\mathbf{w} \cdot \mathbf{x}) + b\} = \mathrm{sgn}\left\{\sum_{i=1}^{n} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}) + b\right\}$$

其中，α 为约束条件(1)对应的 Lagrange 乘子，$\alpha_i \geq 0$ $i = 1, ..., n$。b 是分类阈值。对非线性问题，相应的分类函数为

$$f(\mathbf{x}) = \mathrm{sgn}(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, +\mathbf{x}) + b)$$

SVM算法通过一个变换函数，将需要计算的变量维数大大减少，使得计算的复杂度不再取决于特征变量维数,而是取决于样本数，尤其是样本中的支持向量数。这些特点使维数爆炸问题得到有效地解决。据此，微软研究院的Scholkopf 等人的研究表明支持向量机(SVM) 可以用于密度估计和孤立点发现，适合于金融领域客户行为模式的高维异构不均衡数据集内的异常发现[5]。因此，利用SVM技术对客户异常行为模式进行识别在技术上是可行的。

本系统中，各部门识别 Agent 加载 SVM 算法，处理机根据部门业务特性选取的属性值，调用算法处理业务数据，甄别洗钱交易，获得相应的知识并发送黑板公共区由各 Agent 协商，产生最终的识别结论。

3 总结

本文构建了基于多 Agent 客户识别的反洗钱系统。本系统具有多 Agent 发挥集体智能的优势，同时又具有从海量数据中迅速有效地识别洗钱行为、兼容动静态洗钱规则、跨部门识别的领域特色，将为智能化的反洗钱的实现提供理论基础与技术支持。但是，应当看到该系统在具体的实现中还存在如何保障公网上 Agent 信息交换的安全性问题，这将是作者下一步要开展的工作。

参考文献

1 Kingdon J. AI Fights Money Laundering[EB/OL]. 2003. http://www.searchspace.com/ ibrary/showPress.php?id = 20.

2 Wang H Q. Intelligent Agent-assisted Decision Support System: Integration of Knowledge Discovery, Knowledge Analysis and Group Decision Support[J]. Expert System with Applications, 1997, 12(3): 323-335.

3 杨文恺, 李海刚. 基于多 Agent 技术的企业协同框架研究[J]. 计算机工程, 2005, 31(11): 208-210.

4 左万里. 基于多 Agent 的智能故障诊断系统研究[J]. 计算机与现代化, 2003, (8): 4-6.

5 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.

6 杜树新, 吴铁军. 模式识别中的支持向量机方法[J]. 浙江大学学报(工学版), 2003, 37(5): 521-527.

7 张　焱, 欧阳一鸣, 王　浩. 数据挖掘在金融领域中的应用研究[J]. 计算机工程与应用, 2004, 40(18): 208-211.