电信欺诈综合分析与系统架构研究

罗 丹[1] 刘万军[1] 罗 超[1] 操龙兵[2] 戴汝为[3]

(辽宁工程技术大学电子与工程系 辽宁阜新 123000)[1]

(Faculty of Information Technology, University of Technology, Australia)[2]

(中国科学院自动化所复杂系统与智能科学实验室 北京 100080)[3]

摘 要 多年来,无论在西方发达国家还是在发展中国家,电信欺诈都正变得越来越严重与普遍。越来越多的电信运营商与欺诈研究专家认识到这个问题的严峻性与长期性,并分别在理论与工程等方面研究欺诈侦测、分析与预防的策略与措施。然而,从总体上看,上述工作实际效果不尽人意。本文从系统科学角度,按照综合集成的方法论,探讨进行电信欺诈智能分析与控制的综合策略。文章首先从多维视角分析电信欺诈的复杂性与演化性,并进而提出一套智能分析与监控的综合策略:包括混合智能分析与监控的一整套流程,从五个模型与四层分析进行这类系统的分析与设计等,并进而基于上述策略提出欺诈智能分析与监控的闭环商业智能系统的框架。最后以移动欺诈客户为例,讨论了欠费客户的分类分析,介绍了一个基于上述策略的智能分析与监控系统原型。

关键词 电信欺诈,智能分析与监控,综合策略,综合集成

Hybrid Analyses and System Architecture for Telecom Frauds

LUO Dan[1] LIU Wan-Jun[1] LUO Chao[1] CAO Long-Bing[2] DAI Ru-Wei[3]

(Department of Electronics and Information, Liaoning Technical University, Liaoning Fuxing 123000)[1]

(Faculty Of Information Technology, University of Technology, Sydney, Australia)[2]

(Lab. of Complex System and Intelligent Science, Institute of Automation, CAS, Beijing 100080)[3]

Abstract In recent years, telecom frauds have been getting more and more widespread in either western developed countries or other developing countries. Many telecom operators and experts in this field have recognized the seriousness and persistence of telecom frauds, they have worked together to develop strategies, algorithms and tools for detecting, analyzing and preventing fraudulent activities. However, the above work including industrial practices has shown to be unsatisfactory. In this paper, we focus on finding a hybrid strategy for intelligent analysis and control of frauds in terms of systematism and the methodology of metasynthesis. We first discuss the complexity of telecom fraudulent activities from multiple-dimension point of view. Furthermore, we present an approach for system analysis and design of fraud system: a package solution of intelligent analysis and control from life cycle of fraudulent activity, and a five-model and four-analysis method. A framework of intelligent analysis and control system for dealing with frauds is also shown. Our case study of owe fee classification and a system prototype shows that our strategy is valuable and feasible for systematic and scientific decision making of dealing with telecom frauds.

Keywords Telecom frauds, Intelligent analysis and control, Hybrid strategy, Metasynthesis

1 前言

据估计,近年来世界电信业由于各种形式的欺诈所造成的损失每年达到几十亿美元。在中国,2001 年欺诈给电信业造成的损失高达 200 亿元[1]。在西方国家,虽然在不同的电信服务与时间内欺诈带来的损失有所不同,但是,每年的损失仍高达收入的 3%～6%[2]。

越来越多的电信运营商与电信业主管部门认识到这个问题的严峻性、复杂性与长期性,并从多方面探讨进行欺诈分析、预防与堵截的途径与措施。目前,在欺诈侦测与管理中,行政手段与人干预应用得比较多。为了应对欺诈行为,有关部门提出并探讨多种可能的途径,比如法律规则、客户信息收集、客户信用认证与评估、在现有的电信系统中附加安全机制等。这些措施使得电信网络与服务抵御欺诈的能力得以提升。然而,有一些商业性的欺诈活动是难以采用一般的技术手段加以排除的。此外,引入的新兴服务也会产生相应的欺诈活动。

智能而系统化的欺诈分析与监控技术[3],可以进行早期的欺诈侦测与预防,并有效应对技术欺诈活动。另一方面,这些技术有助于进行欺诈前、欺诈过程中与欺诈后的活动分析、监测与控制。因此,深入研究欺诈活动的各种可能情形与复杂性,针对欺诈可能的源头、途径、趋势等进行综合分析,提出比较全面的欺诈侦测、预防、分析、监测与控制,无疑具有极大的经济与社会效益。本文重点探讨设计电信欺诈智能分析与监控系统的策略。该系统提供电信欺诈的侦测、分析、预测、预警、监测、控制等功能,试图为电信企业提供一整套的、闭环的分析与监控体系。

本文第 2 节总结与分析可能的电信欺诈情形及其复杂性表现,第 3 节概要介绍电信欺诈智能分析与监控的综合策略,第 4 节进一步提出一个电信欺诈混合智能系统框架,第 5 节以欠费分类为例介绍基于商业智能技术的欺诈专题分析,第 6 节给出了一个面向中国移动用户的欺诈分析与监控系统原型。最后总结全文并介绍将要进行的工作。

2 电信欺诈的复杂性分析

为了更好地研究应对电信欺诈的措施,有必要首先弄清楚存在于电信网络与服务中的电信欺诈的可能情形、特征,从

罗 丹 工程师,研究生. 刘万军 教授,研究生. 罗 超 研究生. 操龙兵 博士. 戴汝为 教授,院士.

而把握电信欺诈的内在特点与规律性。

电信欺诈者指的是那些故意甚至有组织地通过各种可能的途径欺骗或者盗用电信服务的电信用户。他们可能采用的进行欺诈的途径有:服务订购欺诈、PBX 欺诈、呼叫卡欺诈、PRS 欺诈、帐务欺诈、盗用服务内容、预付费欺诈、盗用呼叫、内部盗用、漫游欺诈、投币电话欺诈、病毒攻击、偷听等等。这些欺诈主要是通过技术手段进行或者达到目的。

随着 IP 网络与 IP 服务的普及,它所采用的技术和推行的商业模型与传统的电话服务有很大的差异。Internet 与 IP 在电信业的广泛应用,为电信业带来新的运营与商业模型。在 IP 业务中所涉及的用户角色与用户交互模式比传统的电话网更多、更丰富。更为重要的,以 Internet 为网络基础设施的新的电信商业模型使得电信服务类型、角色、用户及作用等都处于动态演化状态,而不像传统电话网中那样单一、稳定。在 IP 业务中可能涉及的角色包括内容提供商、服务提供商、网络运营商、客户以及欺诈者。电信运营商可能会同时担负多个角色,比如既是网络运营商又是服务提供商。不幸的是,IP 业务中的这种多角色合一的现象也为产生新型欺诈创造了条件,比如非法重分配、过渡下载、订购欺诈等。

关于欺诈的另三个观察角度分别是:(i)是否是以赢利为目的技术欺诈,或者(ii) 不是以牟取私利而是为方便个人使用为目的,亦或是(iii) 从电信企业内部通过滥用职权和操纵服务达到使用服务的目的。上述对于网络滥用情况的观察进而可以归结为下面三类欺诈之一的结果:行政欺诈、采购欺诈或者应用欺诈[4]。

进一步分析,电信无线与有线网络中的电信欺诈可以归结为下面三种类型:(i) 技术欺诈,比如投币电话与预付费欺诈、PBX 性能滥用、盗用信用卡或者信用卡号、盗用或者伪装话机、电话线私自嫁接话机、克隆本地或漫游服务信息、PRS 欺诈等;(ii) 订购欺诈,比如帐务欺诈、盗用服务内容、预付费欺诈、偷听、盗用身份、克隆 SIM 卡、IP 欺诈、呆坏帐、呼叫转移或漫游欺诈等;(iii) 内部欺诈,例如盗用电信数据、开辟系统安全缺口、从欺骗性的销售中牟取提成、无授权提供服务等。

更为复杂的是,随着电信客户在技术与服务方面的知识水平的提高,电信欺诈者越来越熟悉电信业务在技术上与行政上可能存在的漏洞。换言之,电信欺诈者的欺诈手段与行为也随着电信业的发展而发展,欺诈活动具有了演化性,从而使得应对欺诈变得越来越复杂、困难。

3 欺诈智能分析与监控综合策略

随着新兴服务类型的不断涌现,传统的欺诈监测方法论与技术不足以有效应对新服务的复杂性,实施工具管理。电信运营商期待着能够提出一整套的解决方案进行事前与事后欺诈处理。

正如在第 2 节所讨论到的,欺诈活动的不断演化使得对于处理电信欺诈的指导思想与技术途径的要求也变得越来越苛刻。为此,有必要从交叉学科里吸取新的指导思想与技术手段,比如按照综合集成的思想[5,6]从系统科学角度全面地分析欺诈及其所处的环境,采用数据挖掘与知识发现技术、自学习与自适应技术、突现模式侦测与识别、数据流挖掘等技术。问题是,如何从系统工程的角度构建综合多学科技术的欺诈管理、分析与监控系统,这是我们应对新兴欺诈与欺诈演化所重点研究的问题。

本节我们将从混合智能系统的角度提出一套比较全面的处理电信欺诈的综合解决方案。我们结合在实践中的体会首先讨论综合解决方案的体系架构,并进而讨论具体实施该体系架构中的两大关键问题:面向系统分析与设计的五个模型、面向商业智能的四类分析[7]。

3.1 欺诈处理的综合策略

综观电信欺诈研究,包括已经进行过的和目前正在进行的,大多数侧重在理论上从不同的学科、采用不同的方法研究进行欺诈侦测与分析的算法,而在将理论与工业实际相结合研究面向工业真实环境的欺诈综合处理系统方面则相对较弱。

另一方面,当前使用的或者正在筹建的电信欺诈侦测或者管理系统,主要提供基于预定义报表、即席报表和 OLAP 报表分析等手段的报表汇总分析。这些系统在一定程度上满足了电信运营商对欺诈分析的要求。数据挖掘技术[8]从企业信息系统的海量数据里挖掘隐含的知识,从而提升生产率、改善客户关系、降低收益损失。但是,在开发面向欺诈真实环境的挖掘专题、模型的设计与演化、结果的可视化与解释性、如何将数据挖掘融入到处理问题的一整套环节等方面,都存在许多值得探索的问题。

更进一步,由于电信欺诈是在电信生产系统面向社会提供服务的过程中衍生出来的,涉及技术与服务运营环境即社会环境,更关系到经营效益与效率问题。因此,电信欺诈处理系统本质上是一个电信欺诈的分析与决策支持系统[3]。该系统应该提供哪些功能以结合欺诈生命周期开展全程性欺诈处理工作?系统能否将技术分析的结果自动反馈到生产系统或者经营与决策者,从而形成一个闭环系统?建设这样的系统所涉及的系统分析与设计方法、主要的技术手段有哪些?

此外,电信运营商,无论是提供传统业务还是基于 IP 新服务,在期待处理欺诈的系统中,都希望得到一个能够对欺诈发生与演化过程、活动、行为、结果、未来等进行全面管理的解决方案。这就是我们研究处理欺诈综合解决方案的起因。我们的目标,是探讨贯穿欺诈整个生命周期进行监测、分析、管理与控制的综合解决方案的主要研究内容、各环节的主要任务与目标、建立该系统的主要问题等,从而形成一套面向欺诈全流程的处理体系。

为此,结合我们在电信业中运用商业智能技术进行的一些探索,我们提出一个电信欺诈综合解决方案,该方案试图从事前、事中、事后等三大环节进行欺诈的全生命周期分析与监控。图 1 描述了该系统所包含的主要功能模块、系统闭环机制,标明了该系统与生产系统的关系。

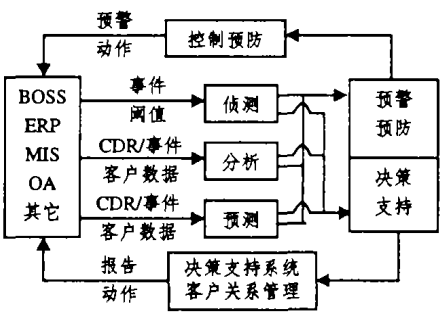

图 1 欺诈分析与监控系统结构示意图

该欺诈综合解决方案包含具有如下特点:(i) 是一个集成的全流程监控与管理系统,能实现从侦测、分析、预测、预警、预防到控制、管理;(ii) 是一个混合智能平台,根据需要采用多种人工智能技术,比如用于侦测的智能方法、针对分析的挖掘技术等;(iii)是一个闭环商业智能系统,从电信运营支撑

系统及其它外部系统中抽取、转换数据、形成统一的欺诈数据仓库，支持全面的智能分析，基于分析结果提出决策支持报告或建议，反馈到生产系统或者经营环境，以最终实现欺诈处理目标。

我们将该系统设计成一个闭环商业智能系统[7,9]，还进一步表现为将进行欺诈分析、预测与预警、预防与控制等所发现的建议、报告、事件或者希望采取的动作等直接反馈给相关的生产系统，或者反馈给经营分析师或决策者，并通过他们执行最后的预防、控制或者其它处理工作。无论是以通过事件激活、规则匹配还是以阈值判断侦测、分析或者预测发现的欺诈，都会向生产系统或者决策者发出预警与动作，警示、预防或控制异常的服务运营或使用和/或欺诈活动。同时上述信息连同新发现的欺诈行为，将进行统一编码与归档，作为案例知识记录到欺诈处置库与欺诈行为库。欺诈处置库与欺诈行为库伴随着欺诈活动的演化而不断积累，从而使得欺诈分析与监控系统具有演化适应性。

3.2 基于五个模型的系统分析与设计

总结我们在探讨如何实施欺诈分析与监控系统的过程中的经验与感受，我们认为该系统的分析与设计可以基于下列五个模型[7]进行：面向电信运营环境的概念模型、面向客户关系管理的客户模型、面向数据仓库与集市的数据模型、面向侦测与分析与挖掘的算法模型、以及面向欺诈分析与监控的系统模型。这五个模型贯穿了欺诈分析与监控系统的分析与设计的全过程、与欺诈有关的运营支撑系统及其环境。五个模型中，针对电信欺诈的概念模型与客户模型是实施电信欺诈分析与监控系统的基础。

3.2.1 *面向电信运营环境的概念模型* 建立电信欺诈系统的基础是建立面向电信运营商经营环境的概念模型。概念模型的基本任务是基于电信领域与欺诈情景分析，定义与电信欺诈有关的主要电信服务、电信运营系统、电信产品、企业运营工作流、信息采集、信息发布与使用策略、经营决策流程、企业内部欺诈处理流程与制度、主管部门欺诈与信用管理的有关规定与政策、继往发现的欺诈类型、实施欺诈的途径、欺诈行为特点、欺诈活动处理案例、积累的经验与教训等。

3.2.2 *面向客户关系管理的客户模型* 面向客户关系管理的客户模型是按照客户关系管理[10]的思想，定义面向电信运营环境的欺诈问题概念模型中的系统角色，各角色类型、责任、管理或者使用的电信产品与服务，企业客户关系管理有关的规定、流程，客户背景、联系、状况、帐户、行为、服务以及扩展等属性，客户总量、新增、流失、转网情况，使用各类电信产品与服务的客户类型、组成、行为等，客户信用度、风险黑白名单、高额客户、重要客户、恶意欺诈客户等，继往发现的欺诈客户类型、行为、途径等。

3.2.3 *面向数据仓库与集市的数据模型* 在概念模型与客户模型的基础上，基于商业智能技术，将欺诈作为一个特定的分析专题，进行数据准备与预处理、属性选择，确定进行欺诈专题分析的子专题、分析角度（维度）、分析指标（度量）等，并基于星型或者雪花型模型，设计数据模型，包括事实表、概要表、维表、度量等。由于只针对欺诈进行商业智能分析，而欺诈问题却牵涉到计费、帐务、结算、客户服务等多个生产系统，因此欺诈专题分析涉及客户、业务、收益、营销、市场竞争、重要客户、合作服务方等多个主题层面，有必要从上述多方面出发设计欺诈专题的子专题、分析角度与指标等。比如，客户新增/流失分析、客户转网分析、客户欠费分析、客户行为分析、客户信用与风险分析、高额客户分析、大客户分析等等。

3.2.4 *面向侦测、分析与挖掘的算法模型* 针对欺诈情形与欺诈处理的目标需求，设计、选择进行欺诈侦测、预测、分类、聚类、转网、流失等分析的算法。算法模型需要根据目标问题进行算法选择，从训练、测试到应用与打分，以及模型优化等。

3.2.5 *面向欺诈分析与监控的系统模型* 在上述各模型与系统分析设计的基础上，建立面向欺诈分析与监控的系统模型。系统模型定义该系统主要功能、组成结构、开发模式、层次，如何实现从多个数据源进行数据抽取与更新逻辑、数据转换与集成机制，从数据 ETL、数据仓库、智能分析到结果展现与决策支持的具体技术与途径，如何实现将分析与发现的事件与动作反馈到生产系统或者决策经营者等等。一些重要问题还包括如何设计系统体系结构与设计模式、元数据管理、统一认证、统一稽核机制，并最终形成一个统一的企业知识门户。

通过建立上述五个模型，对欺诈分析与监控系统的从需求分析到问题求解全过程的主要问题应该都已经解决。此时实施该系统就会变得相对清楚、容易。图 2 显示了上述五个模型之间的相互关系。

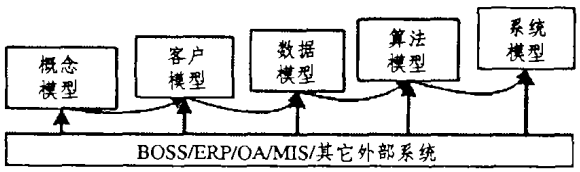

图 2 五个模型

3.3 四类分析

在电信欺诈处理中，数据分析扮演着重要甚至关键的角色。由于欺诈涉及的环节与生产系统比较复杂，因此电信欺诈分析需要从多个专题、多个分析角度与多个观察指标进行综合考察，每个专题都需要采用以下四类分析：预定义分析、即席分析、OLAP 分析及数据挖掘分析，才能全面、准确地把握欺诈内外部情形与规律，从而准确地预测、预警、预防与控制欺诈。

四类分析的基础，是从电信运营支持系统（包括计费系统、运营与帐务系统、结算与财务系统，以及企业资源规划、管理信息系统等外部系统）中按照概念模型、客户模型与数据模型的设计，抽取并清洗转换数据，建立专用的欺诈数据仓库或者数据集市。在欺诈数据仓库或者数据集市上进行包括上述四类分析在内的欺诈综合分析。如图 3 所示。

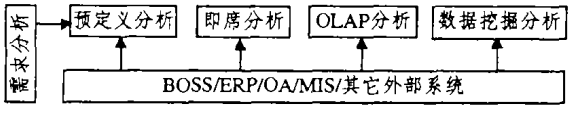

图 3 四类分析

预定义分析是按照客户日常具体观测需要，采用统计技术，对细节数据、轻型或者中度聚集的数据按预先定义的分析角度、观测指标等进行汇总。预定义分析是日常进行欺诈行为与异常动作分析的最主要形式，占据日常报表的多数。

然而，作为灵活而动态的监测欺诈的途径，即席查询提供了运行时进行属性选择、条件过滤、规则生成的机制，可以基于一个或者多个主题或专题（或者子主题与子专题），自由选择一个或者多个概要表、多个维表、多个分析度量，在线交互

式动态生成报表。OLAP 分析提供了多维、关系或者混合型的在线分析处理机制，可以同时从多个角度观察客户行为、欺诈活动，发现异常情况。

作为一种深层的数据分析与决策支持技术，数据挖掘[8]在从大规模、历史或者动态的电信服务数据集中发现不寻常的行为和模式方面发挥着不可替代的作用，可以发现那些采用前三种分析方法无法观测到的隐藏的信息或者行为模式。

由于欺诈情形本身的复杂性与演化性，进行欺诈分析的挖掘算法需要结合各分析专题的具体任务进行算法选择，必要时需要进行多类算法的集成应用。另外，由于大量、连续的电信服务数据流的快速输入与对电信生产系统的实时更新，进行欺诈挖掘分析需要研究流数据挖掘[11]新技术。流数据挖掘新技术的研究与应用必将增强在线、动态地进行欺诈侦测、预防、预测、分析、预警与控制的能力。

4 电信欺诈混合智能监控系统框架

本节描述了建设一个更为具体而实用的电信欺诈混合智能分析与监控系统所应解决的主要问题、包含的主要模块及其功能，最后给出了一个系统框架结构图。

系统目标：基于上面关于闭环智能分析系统的设计理念，电信欺诈混合智能分析与监控系统提供技术支持欺诈侦测、预测、预警、预防与控制，并将发现的事件或者动作反馈给生产环节或者经营决策者，以支持更为系统化与全面的决策支持。

对于异常或者负面行为的警示信号，无论是基于事件还是阈值侦测、分析或者预测得到的，报警信息或者动作请求将被反馈给相应的生产子系统或者触发生产系统的相应控制环节，或者发送给系统管理员、经营与技术分析师或者经营决策者，供他们进行人工控制或者预防。这些日常或者周期性的报告或者信号还将被用于决策支持与客户关系管理。

数据集成：如前所述，由于欺诈信息隐含在多个生产子系统中，在建立面向欺诈分析的数据仓库或者数据集市以前，需要从各个数据源进行数据抽取、清洗、约减、聚合等处理，生产高质量的数据。但是，由于实际生产系统的复杂性，还存在结构与语义的不一致现象，需要进行整合处理才能形成可以灌入数据仓库与集市的数据。

另一方面，数据集成还涉及从面向业务的用户层到面向实体的数据源层直接的映射。为此，可以建立一个基于本体的三层结构，实现从业务层概念到仓库层属性到企业信息系统中字段的三层映射。在此过程中需要建立三层的本体元数据库、元数据之间的映射与查询解析机制[12]。三层本体库的建设与维护是基于具体问题领域的需求、系统分析（概念模型与客户模型）、企业信息系统，通过本体元数据的管理与维护进行的。数据仓库与数据集市除了要设计与灌入技术元数据信息，还要管理业务元数据与 SQL 操作类元数据项。

数据集成的结果不仅要建立一个集成而透明的从业务用户接口至 EIS 数据源的通道，隐藏异构数据源的复杂性，在用户端还应提供一个统一的、面向业务的易用的企业门户与工作环境空间，支持用户进行业务分析、系统管理等。

欺诈处理：在数据仓库与数据集市之上进行四类欺诈处理：侦测、分析、预测与预警。对于每个主题与专题，根据问题性质与需求进行四层的分析：预定义分析、即席分析、OLAP 分析及数据挖掘分析。上述处理结果，比如异常消费行为的发现、离网趋势的预测等，反馈到企业门户或者相关的生产系统或环节，进行预警、预防甚至控制。

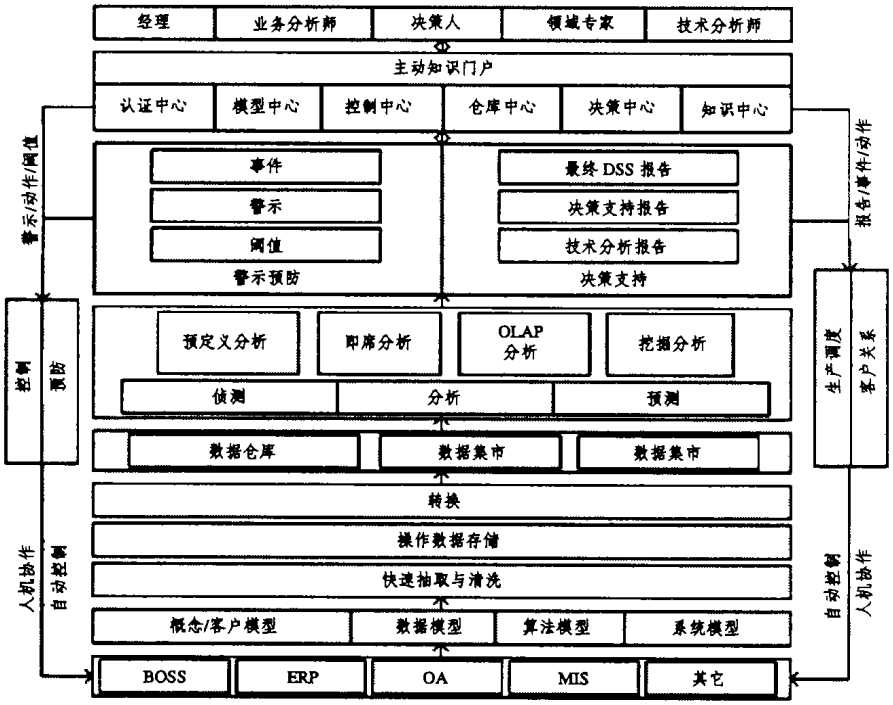

图 4 电信欺诈分析与监控系统框架

决策支持：欺诈分析的目的是进行决策支持。对于欺诈处理的发现与结果，将进一步反馈到相关的市场经营、综合业务运营支撑、客户服务等部门，进行生产调度、运营控制与管理、经营决策分析、客户关系管理等。反馈给上述部门相关人员供决策支持的材料包括详细报告、技术分析报告、决策建议报告（由专门的技术分析与决策建议报告模板通过规则定义灌入数据与文字描述）等，与系统角色之间的通信渠道包括消息传递、Web 消息、短消息、电子邮件等。

企业知识门户：企业知识门户[7]为客户提供一个统一的信息与知识表达、服务与管理，分析主题与专题的组织与浏览，统一的用户认证，数据与安全稽核，ETL 逻辑与稽核，元数据与本体管理，灵活的工作流程与功能重构，统一的信息、应用、视图与服务等的集成入口。无论底层的企业信息系统如何复杂，完成欺诈分析与处理所涉及的环节与技术有多复杂，

企业门户提供了一个对用户透明的、一站式的工作与服务空间。

系统框架:根据上述讨论,图 4 描述了一个基于商业智能思想的欺诈分析与监控系统的框架结构。如图所示,在运营支撑系统 BOSS、外部系统 OA 与 MIS、客户需求的基础上建立面向客户问题领域的运营概念模型、客户模型、数据模型、算法模型,以及欺诈分析与监控系统模型,作为进行欺诈分析与系统设计的基础与依据。从 EIS 中抽取与清洗后的数据可以灌入操作数据存储系统 ODS,并根据数据模型设计与建立数据仓库或者集市。利用算法模型中的算法和四类分析手段进行欺诈侦测、分析、预测等,发现的提示、动作、事件、异常等结果直接或者以决策支持报告的形式反馈到相应的生产系统或者相关的经营分析师与决策者,进行预防、控制、生产调度、客户关系管理等。系统各类角色,包括经营分析师、项目经理、领域专家、技术分析师等,通过包含在企业知识门户中的多个中心与系统进行人机对话。

5 案例分析——欠费客户分类分析

电信欺诈分析与监控是一项复杂的系统工程。为了全面地把握欺诈行为规律,需要以欺诈作为一个特定的专题设计分析数据模型,在这个专题中包括多个子专题、分析角度与分析指标。表 1 描述了进行欺诈分析的部分子专题、分析角度与指标。下面以客户欠费分类为例简要介绍数据挖掘分析在客户欠费子专题中的应用。

表 1 欠费欺诈专题分析

	内容名称
子专题	欠费状态统计分析,欠费原因因素分析,欠费客户特征分析,欠费业务特征分析,欠费趋势预测分析,欠费客户结构分析等
分析维度	时间(日/月/年)、帐期、业务属性、客户背景属性、客户行为属性、客户帐户属性、客户服务属性、营业点属性等
分析指标	欠费客户数、欠费金额、欠费周期、欠费率、欠费回收率、高额欠费率、恶意欠费率等

5.1 特征选择

在进行数据模型的设计时,为了增强数据模型的可靠性,常需要采用数理统计、OLAP 分析、粗糙集、信息熵等方法数据分布与数据特征进行分析,从而发现并保留与所关注的任务敏感的属性,排除无关属性,即属性特征选择。比如,在对某移动运营商进行客户欠费分类分析的数据模型设计与优化过程中时,我们发现与欠费类别敏感的属性与贡献率不大的属性如表 2 所示。

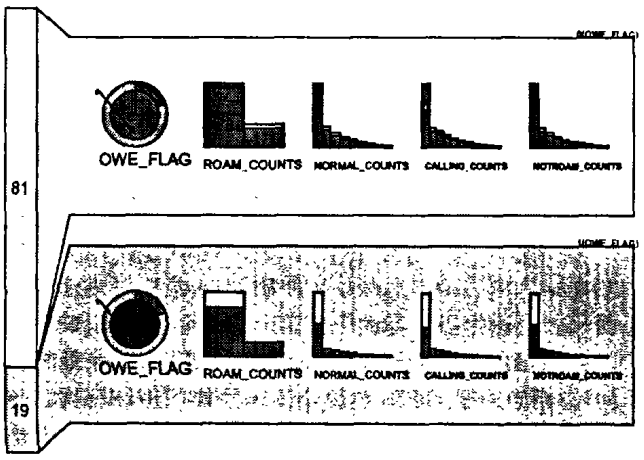

图 5 欠费分类特征选择

从图 5 中(图中符号标识参见表 2)可以进一步发现,对于表中"选择的属性",欠费客户的分布(红色部分)与全部用户(深色部分)的分布有明显的差异;而"排除的属性"等属性对于识别欠费贡献不大,因此从一开始设计的数据模型中排除这些属性。

表 2 数据模型特征选择

选择的属性	排除的属性
漫游次数 (roam_counts)	呼叫联通次数(Unicom_dura)
正常时段呼叫次数(normal_counts)	停机次数(stop_counts)
主叫次数 (calling_counts)	呼叫亲情次数(love_counts)
非漫游次数 (notroam_count)	呼叫小灵通次数(phs_counts)

5.2 决策树分类

图 6(图中符号含义见表 3)是采用决策树算法得到的欠费客户分类的结果。以 lost_flag 节点为例,从图中我们可以发现如下欠费客户类别规则:

```
if (prepay_fee<0.005000)
  and (total_counts<3.5000000)
  and (prepay_fee>=-0.005000)
  and mesg_fee<10.700000)
  and (should_fee<0.500000)
then Class =1
lost_flag
```

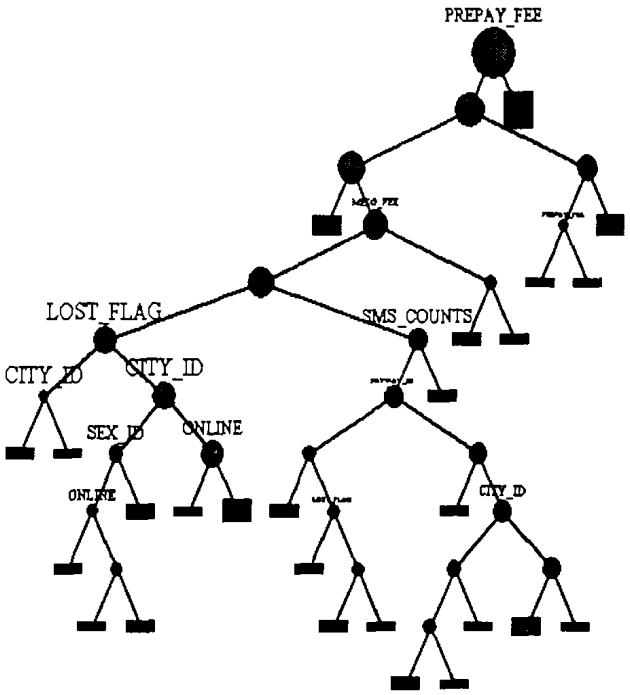

图 6 欠费决策树分类

表 3 决策树数据模型特征说明

	特征名称	符号表示
1	预存款	prepay_fee
2	信息费	mesg_fee
3	短信次数	sms_counts
4	地区标识	city_id
5	性别标识	sex_id

在原始数据中抽样 28084 个作测试数据,测试模型后的分析结果是给每个新用户加上新的类标签。与原来的结果相比较,测试结果如图 7 所示。

从图中可以明显地看出来模型将测试样本分为两类,类标签 OWE_FLAG=1 的为欠费客户;OWE_FLAG=0 的为不欠费客户。在抽取的 28084 个样本中共有 576 个样本被错

分，错误率为2.051%。在欠费的5362个样本中有5169个样本被正确分类，正例样本的准确率达到96.40%；在未欠费的22722个样本中被错误贴上“欠费”类标签的有383个，占总样本数的1.686%。最终此训练模型为5552个样本贴上“欠费”的类标签，其余的22532个样本被分为“未欠费”类。表4汇总了应用决策树算法进行欠费客户分类预测的结果。

表4 欠费客户决策树分类预测结果

	欠费客户数	不欠费客户数	合计
欠费客户数	5169	193	5362
不欠费客户数	383	22339	22722
合计	5552	22532	28084

可见，采用决策树方法进行欠费客户分类预测中，错分概率是：

(193+383)/ 28084 =2.051%。

6 欺诈智能分析与监控原型系统

图7显示了一个基于上述策略的欺诈分析与监控系统的部分分析结果的界面快照。这个系统除了提供上述四类分析外，还提供面向业务部门或者板块的管理、用户及用户组管理、ETL规则稽核、统一客户认证、动态查询等。系统采用了IBM DB2数据仓库系统，建立了多个专题与主题分析，从侦测、分析、预测、预警等角度全面分析与欺诈有关的各个主题与专题，为电信运营商从事前预测、事中分析与事后控制等环节把握经营生产中欺诈的产生、发展与演化，及时采取措施进行预防、控制与管理提供了科学、及时的依据。

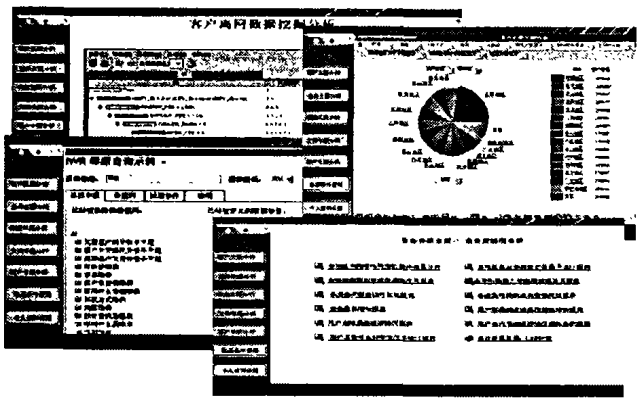

图7 欺诈分析与监控原型系统屏幕快照

结论与未来工作 近些年来，电信欺诈成为困扰世界电信运营商的顽症之一，占了年收入的3%～6%。随着电信新业务的不断推出，新的欺诈随之涌现。电信欺诈正表现出随电信业务与电信业运营政策而演化的特点。

在上述环境下处理电信欺诈问题，有必要探讨新的、更为系统化的应对策略。本文从系统科学的角度，采用综合集成的方法论，在多视角考察电信欺诈的可能复杂性的基础上，从电信欺诈产生、发展与演化生命周期出发，除了常规的分析欺诈行为外，还将滋生欺诈的环境——运营支撑系统、经营决策者、运营政策等纳入考察的范围，从而提出进行电信欺诈智能分析与监控的综合策略，即从欺诈侦测、分析、预测、预警、预防到管理与控制等全流程的闭环商业智能监控体系。在此基础上，总结提出了进行系统分析与设计的五个模型与四类分析，给出了电信欺诈分析与监控系统框架。在文章最后，结合某移动运营商欠费分类子专题，具体分析了欠费客户的敏感特征，并采用决策树算法进行了欠费客户分类分析；并基于上述策略展示了一个欺诈分析与监控系统原型。

本文今后的工作将结合电信运营商的欺诈现实，对本文提出的策略结合真实世界的环境进行具体、深入的工程实验与理论分析，从而建立更为翔实的电信欺诈智能分析与监控的理论与应用体系。

参 考 文 献

1 电信欠费上升的原因与对策. 亚太经济，2002(6)

2 de Jager P. Introduction to using intelligent techniques for telecommunications fraud detection. http://www.eurescom.de/~pub/seminars/past/2001/SecurityFraud/12-Jager/

3 Cao Longbing, Luo Chao, Luo Dan, Zhang Chengqi. Hybrid strategy of analysis and control of telecommunications frauds. In: Proc. of ICITA2004, 2004. 55～60

4 Burge P, Shawe-Taylor J, Cooke C, Moreau Y, Preneel B, Stoermann C. Fraud Detection and Management in Mobile Telecommunications Networks

5 钱学森，于景元，戴汝为. 一个科学新领域——开放的复杂巨系统及其方法论[J]. 自然杂志，1990，13(1)：3

6 戴汝为，王珏，田捷. 智能系统的综合集成[M]. 浙江科学技术出版社，1995

7 Cao Longbing, Luo Dan, Luo Chao, Zhang Chengqi. Systematic Engineering in Designing Architecture of Telecommunications Business Intelligence System. In: Proc. of Int. Conf. of Hybrid Intelligent System'03, 2003. 34～42

8 Han Jiawei, Kamber M. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, Aug. 2000

9 Whitehorn M, Whitehorn M. Business intelligence: the IBM solution, New Springer. London, 1999

10 Berson A, et al. Building data mining applications for CRM [M]. McGraw-hill Co., 2000

11 Babcock B, Babu S, Datar M, Motawani R, Widom J. Models and issues in data stream systems, PODS'02 (tutorial)

12 Cao Longbing, Zhang Chengqi, Luo Chao, Dan Luo. Ontology Services-Based Integration of Data Warehousing, OLAP and Data Mining in Mining Telecom Business Intelligence (submitted), 2004