

文章编号: 1001-0920(2003)04-0494-03

## 一种基于案例推理的欺诈分析方法

柳炳祥<sup>1</sup>, 盛昭瀚<sup>2</sup>

(1. 东南大学 经济管理学院, 江苏 南京 210018; 2. 南京大学 管理科学与工程研究院, 江苏 南京 210093)

**摘要:** 为了对欺诈行为进行分析和识别, 提出一种基于案例推理的欺诈分析方法。研究了应用案例推理方法对各种欺诈信息进行识别的一些关键技术, 并通过一个信用卡欺诈行为的分析实例对该方法进行检验。研究表明, 该方法应用于欺诈行为的分析和识别是有效的。

**关键词:** 案例推理; 欺诈; 案例库; 检索; 知识表示

中图分类号: TP183

文献标识码: A

## Case-based reasoning for cheat risks diagnosis

LIU Bing-xiang<sup>1</sup>, SHENG Zhao-han<sup>2</sup>

(1. Institute of Economic and Management, Southeast University, Nanjing 210018, China; 2. Graduate School of Management Science and Engineering, Nanjing University, Nanjing 210093, China)

**Abstract:** A new method based on case-based reasoning theory is proposed to analysis the cheat risks. Some critical techniques are investigated such as knowledge representation, retrieval of similar cases in case library and case learning. A credit card cheat example is given to validate the method based on case-based reasoning. The results show that it is reliable for identifying various cheat risks.

**Key words:** Case-based reasoning; Cheat; Case library; Searches; Knowledge denotation

### 1 引言

随着信息技术和网络化经济的迅速发展, 传统的商业模式发生了根本性的变化。企业面临的外部环境风险和内部经营风险越来越多, 企业危机管理及其预警已成为管理领域新出现的一个研究热点问题。如何对企业面临的各种危机进行识别、分析、评价和管理, 是企业取得竞争优势, 实现可持续发展战略的关键。

在企业面临的所有危机中, 有一类危机即欺诈变得越来越大, 危害也越来越大。一旦发生欺诈行为, 企业将面临管理活动的失效、市场份额的丧失和商务活动的失败, 导致企业失去业务、顾客、竞争力和信誉<sup>[1]</sup>。据统计, 英国有 62% 的企业认为欺诈行

为比以前更加普遍, 有 2/3 以上的企业曾发生过欺诈事件, 这还不包括那些发生了欺诈行为但由于各种原因没有向外界披露的企业<sup>[2]</sup>。因此, 建立一种科学的欺诈分析方法, 准确、及时、有效地预测企业可能发生的欺诈行为是非常有意义的。

传统的欺诈分析方法主要有头脑风暴法、统计分析法、主观概率法、指标分析法、现场调查法和指数法等<sup>[2]</sup>。这些方法在实际的欺诈分析和预测中, 存在着许多难题, 结果并不理想。这是因为欺诈行为是一类带有大量不确定因素的半结构化或非结构化问题, 产生欺诈的因素复杂, 种类繁多, 很难科学地计算和评估, 只能采用定性和定量相结合的方法来确定。

收稿日期: 2002-05-08; 修回日期: 2002-07-22。

基金项目: 国家自然科学基金资助项目(70171028)。

作者简介: 柳炳祥(1966—), 男, 江西九江人, 副教授, 博士生, 从事数据挖掘、企业危机管理的研究; 盛昭瀚(1944—), 男, 江苏镇江人, 教授, 博士生导师, 从事复杂系统理论等研究。

表1 欺诈案例框架

案例编号:
框架名: 欺诈案例名称
层次1: 欺诈信息描述
层次1.1: 欺诈种类
层次1.2: 欺诈描述
层次1.3: 欺诈发现的原因
层次2: 欺诈行为特征描述
层次2.1: 欺诈特征1(指标1, 值1, 权重1; 指标2, 值2, 权重2; ...)
层次2.2: 欺诈特征2(指标1, 值1, 权重1; 指标2, 值2, 权重2; ...)
⋮
层次2.n: 欺诈特征n(指标1, 值1, 权重1; 指标2, 值2, 权重2; ...)
层次3: 结果集
层次3.1: 欺诈行为多长时间发现
层次3.2: 欺诈带来的损失
层次4: 相关知识
层次4.1: 欺诈产生的原因
层次4.2: 欺诈发生的过程
层次4.3: 欺诈发生后的相关对策

采用上述框架结构的描述方法, 将所搜集到的欺诈案例存放在一系列相互关联的数据库中, 利用关系数据库技术可以方便地建立案例索引。

### 3.3 欺诈案例的检索方法

案例推理的基本原理是在案例库中检索出最佳案例, 并根据以往的编程经验, 对当前的问题提供解决方案。检索算法效率的高低, 对于案例推理方法而言是非常重要的。案例推理的检索方法有串行检索和并行检索两种策略: 串行检索采用的是由上至下逐层求精的方式, 越往下相似程度越高; 并行检索策略就是同时检索多个案例, 并返回一个相似程度最高的案例。案例推理常用的检索方法主要有最近相邻法、归纳法和知识导引法3种。本文采用最近相邻法<sup>[7]</sup>, 其算法为

$$\text{Sim}_i = \sum_{j=1}^n \omega_j \text{Sim}_{ij} \quad (1)$$

其中:  $\text{Sim}_i$  是案例库中第  $i$  个旧案例与问题案例的综合相似度;  $\omega_j$  是第  $j$  个属性或特征在参与匹配检索的属性或特征指标中所占的权重, 称为第  $j$  个指标的权重;  $\text{Sim}_{ij}$  是第  $i$  个旧案例第  $j$  个属性指标与问题案例第  $j$  个属性指标的相似度。

本文将案例推理方法引入企业欺诈行为的分析, 提出一种基于案例推理的欺诈分析方法, 可在一定程度上弥补上述分析方法的不足, 为欺诈危机的预警提供了一种新的研究方法。

## 2 案例推理方法

案例推理是人工智能领域新出现的一种重要的推理方法<sup>[3]</sup>, 它在通用问题求解、法律案例、医疗诊断、故障诊断、计算机辅助设计等许多领域得到广泛的应用<sup>[4-7]</sup>。案例推理方法是基于人类认知过程的一种方法, 其核心思想是在求解问题时, 在以前类似的求解问题的成功范例基础之上进行推理, 而不必一切从头开始。

典型的事例推理过程, 首先按一定的形式向系统描述当前案例; 其次从案例库中检索出与当前问题相应的案例, 若该案例与当前案例完全匹配, 则输出该案例的求解方案, 否则修正该案例, 形成当前问题的求解并进行评价; 最后将新的案例加入到案例库中, 以便在以后问题求解时使用。由此可见, 案例推理方法实际上是一种类比推理方法, 其主要步骤包括待求问题的案例表达、案例库检索、相似案例解的修正及问题案例的学习等<sup>[5]</sup>。

## 3 基于案例推理的欺诈分析方法

将案例推理方法用于企业欺诈行为的分析和预警问题, 首先是欺诈案例的搜集、整理和分类; 其次是欺诈案例的表达、案例库的组织、建立和维护; 再次是案例检索和相应的匹配算法; 最后是案例的调整策略和案例的学习策略等。

### 3.1 欺诈案例的搜集、整理和分类

首先要重点搜集企业生产、经营、市场、开发等各个方面发生的欺诈案例, 并对这些欺诈案例进行分析、整理和分类。根据分析结果按一定的分类体系和模式描述这些案例, 并把这些案例存储起来组成案例库。案例推理方法的有效性和工作质量, 在很大程度上取决于案例库的规模、所存储的案例质量、案例知识的表示方法和索引机制的确定等。因此, 案例的搜集、整理和分类是非常重要的。

### 3.2 欺诈案例知识的表示方法

案例推理的第1步是将问题表达成案例的形式, 称为案例的知识表示。案例是知识表示的一种模式, 它将专家的知识 and 经验用描述案例的数据结构和众多的案例表示出来, 这实际上是专家求解问题的一个具体例子。本文采用框架的知识表示方法来表示欺诈案例<sup>[7]</sup>, 如表1所示。

最近相邻法的关键是确定属性指标权重和计算单个属性指标的相似度。属性指标权重一般采用专家评分、统计分析和主观概率等方法设定,但它不可避免地带有主观臆断性。本文采用层次分析法来确定属性指标权重,使属性指标权重更具客观性。层次分析法是将复杂的属性指标重要性问题层次化,明确内部的层次结构关系及其各组成因素之间的相互关系,通过专家对两两因素之间相对重要程度的比较和判断而建立判断矩阵,并运用相应的数学方法进行分析和处理,以得出不同指标间相对重要性权重。它是一种定性与定量相结合的数学方法<sup>[8]</sup>。

单个属性指标的相似度采用下式计算

$$\text{Sim}_{ij} = |Y_j^* - Y_{ij}| / Y_{ij} \quad (2)$$

其中:  $Y_j^*$  是问题案例第  $j$  个属性指标的值,  $Y_{ij}$  是案例库中第  $i$  个旧案例第  $j$  个属性指标的值。

### 3.4 欺诈案例的学习方法

人类学习是人类不断获得知识和技巧的过程,而案例推理的学习是从案例库中不断获得新知识和改进旧知识的过程。案例推理的学习分为成功学习和失败学习两种。

成功学习包括推理成功和案例库学习两层含义。推理成功是指相似案例的解经过调整和修正,能作为问题案例的解决方案。案例库学习是指如果案例库中存在相似度大于预先设定的阈值的旧案例,则问题案例不加入到案例库中;否则,问题案例作为新案例加入到案例库中。

同样,失败学习包括推理不成功和案例库学习两层含义。推理不成功是指问题案例不能在相似案例中找到适用解,旧案例与问题案例的最大相似度低于预先设定的阈值,旧案例的解不适合作为问题案例的解决方案。案例库学习是指如果领域专家能给出问题案例的解,则问题案例作为新案例入库;否则,问题案例不入库。

## 4 基于案例推理的信用卡欺诈实例分析

企业危机管理及其预警是近年来管理科学与工程领域中一个令人关注的热点研究问题。由于危机管理所涉及的问题大多是一些半结构化或非结构化问题,很难科学地计算和评估,需要应用其他技术和方法来分析、度量和预测危机。将基于案例推理方法运用到危机管理中,能克服传统分析预测方法解决非线性问题的局限性,为企业的危机管理和预测提供了一种新的思路。下面以信用卡欺诈行为为分析实例,具体说明案例推理方法如何应用于欺诈行

为的分析和预测。

本文利用 <http://www.amadon.ibm.com> 所提供的 1986 ~ 1995 年 10 年间客户信用卡数据库,通过统计抽样抽取出 500 条记录。这些记录数据包括客户的基本信息、客户的有关行为特征、客户使用信用卡的信息及欺诈的相关信息等<sup>[9]</sup>。然后对这些数据进行分析、整理和分类,总结出客户信用卡欺诈案例共 395 条,并按框架的知识表示方法,将所搜集到的欺诈案例存放到一系列相互关联的数据库中,建立了客户信用卡欺诈行为的案例库。

为了检验案例推理方法的有效性,采用 Visual C++ 6.0 为软件开发工具,用 SQL6.0 管理案例库和知识库,编程实现了此算法。以 1996 年发生的 63 起客户信用卡欺诈数据作为实验数据,采用本文介绍的案例推理方法进行检验。通过 63 起欺诈案例的学习和推理,实际预测的欺诈数为 57 起,预测准确率为 90.4%,取得了令人满意的结果。

## 5 结 语

本文提出一种基于案例推理的欺诈分析方法,用于分析和识别欺诈行为,并通过一个客户信用卡的欺诈实例进行分析。研究表明,该算法应用于欺诈行为的分析和预测是有效的,预测的准确率为 90.4%。利用此分析方法,能分析出欺诈为什么会发生? 哪些因素会导致欺诈? 欺诈风险主要来自于何处? 如何预测可能发生的欺诈? 采取何种措施以减少欺诈的发生? 以便准确、及时地对各种欺诈行为进行监视、预警和管理,并采取有效的规避措施,防范欺诈危机的发生。

基于案例推理的欺诈分析方法经进一步拓展,还可应用到各种虚假商业信息的识别以及信用等级的划分等方面。特别是利用此方法对各种虚假的财务数据、会计报表以及各种审计和检察机关的造假等欺诈行为进行分析、识别和预警管理,将是今后需要进一步研究的内容。

### 参考文献(References):

- [1] 基特 塞德格洛夫. 罗平岩译. 商务风险管理完全指南 [M]. 沈阳: 沈阳出版社, 2001. 153-172.
- [2] 詹姆斯. 丁一兵译. 企业的泛风险管理——动态处理风险与机遇的系统化策略 [M]. 长春: 吉林人民出版社, 2001. 116-167.
- [3] Barletta B. An introduction to case-based reasoning [J]. *AI Expert*, 1991, 8(1): 58-66.

(下转第 503 页)

高了重建图像的质量。

## 5 结 语

本文提出一种正则化广义逆 ERT 图像重建算法,并与反投影算法和灵敏度系数算法进行对比。重建出的图像经过统一的门限滤波后,反投影算法、灵敏度系数算法和正则化广义逆 ERT 图像重建算法重建图像的 CSIE 平均值分别为 12%, 9% 和 6%。由此说明,正则化广义逆 ERT 图像重建算法能显著地提高重建图像的质量,而其计算量与一步灵敏度系数算法相当,重建速度较快,适合于工业过程应用,是一种较好的 ERT 图像重建算法。

## 参考文献(References):

[1] Dickin F J, Wang M. Electrical resistance tomography

for process application[J]. *Meas Sci Tech*, 1996, 7(3): 247-260.

[2] 何旭初. 广义逆矩阵的基本理论和计算方法[M]. 上海: 上海科学技术出版社, 1985.

[3] 魏颖. 电阻层析成像技术(ERT)及其在两相流测量中的应用研究[D]. 沈阳: 东北大学, 2001.

[4] 李海青, 黄志尧. 特种检测技术及应用[M]. 杭州: 浙江大学出版社, 2000.

[5] Tadakuni Murai, Yukio Kagawa. Electrical impedance computed tomography based on a finite element model [J]. *IEEE Trans Biom Eng*, 1985, 32(3): 177-184.

[6] Thomas J Y, John G W, Willis J T. Comparing reconstruction algorithms for electrical impedance tomography[J]. *IEEE Trans Biom Eng*, 1987, 34(11): 843-852.

(上接第 496 页)

[4] 袁小红, 王钰. 基于事例的推理: 综述与分析[J]. 模式识别与人工智能, 1995, 8(12): 19-31.

(Yuan Xiaohong, Wang Yu. Case-based reasoning: Outline and analysis[J]. *Pattern Recognize and Artificial Intelligence*, 1995, 8(12): 19-31.)

[5] 周凯波, 魏莹, 冯珊. 基于案例推理的金融危机预警支持系统[J]. 计算机工程与应用, 2001, 14(1): 18-21.

(Zhou Kaibo, Wei Ying, Feng Shan. Case-based reasoning for financial crisis predicting support system[J]. *Computer Engineering and Application*, 2001, 14(1): 18-21.)

[6] Watson. Case based reasoning is a methodology not a

technology [J]. *Knowledge-based System*, 1999, 10(12): 303-308.

[7] Kyung-ahik Shin, Ingoo Han. Case-based reasoning supported by genetic algorithms for corporate bond rating [J]. *Expert System with Applications*, 1999, 13(16): 85-95.

[8] 王莲芬, 许树柏. 层次分析法引论[M]. 北京: 中国人民大学出版社, 1990. 80-126.

[9] Huges. *The Complete Database Marketer: Second Generation Strategies and Techniques for Tapping the Power of Your Customer Database*[M]. Chicago: Irwin Professional, 1996. 232-245.

(上接第 499 页)

## 5 结 语

本文以 agent 技术作为整个系统建模的基础,在信息搜索中引入了并行搜索资源的主动协调和负载自平衡机制。在信息的评价筛选机制中,利用社会选择理论的多准则决策模型,建立了一个基于偏好的评价模型。

## 参考文献(References):

[1] Gerhard Weiss. *A Modern Approach to Distributed Artificial Intelligence. Multi-agent Systems* [M]. Boston:

MIT Press, 1999.

[2] Nils J Nilsson. *A New Synthesis, Artificial Intelligence* [M]. Vermont: Morgan Kaufmann Publishers, 1998.

[3] Danny Blange. *Social Choice and Individual Values* [M]. London: Cowles Foundation of Yale University, 1970.

[4] Shoubong Wang. *Social Choice and Multicriterion Decision Making*[M]. Boston: MIT Press, 1986.

[5] Dennis C Mueller. *Public Choice-* [M]. Cambridge: Cambridge University Press, 1989.