审计署计算机审计中级培训系列教材

审计线索的特征发现

刘汝焯　等 著
王智玉　审定

清华大学出版社

审计署计算机审计中级培训系列教材

审计线索的特征发现

刘汝焯　等 著
王智玉　审定

清华大学出版社
北京

内容简介

本书总结了审计线索特征发现的一般过程，介绍了特征发现的常用技术和方法。从一个全新的角度把数据分为数值型数据和非数值型数据两大类型，对两类数据的关联、分析方法，如何在两类数据的平台上进行特征捕捉进行了详尽的介绍。书中列举了大量的案例，详细阐述了特征发现的思路。

本书具有较强的指导性和可操作性，对于从事计算机审计理论研究和实务工作的人员具有参考价值，同时可供高等院校相关专业的师生参考。

图书在版编目（CIP）数据

审计线索的特征发现/刘汝焯等著. —北京：清华大学出版社，2009.5
（审计署计算机审计中级培训系列教材）
ISBN 978-7-302-19689-1

Ⅰ. 审…　Ⅱ. 刘…　Ⅲ. 计算机应用—审计—技术培训—教材　Ⅳ. F239.1

中国版本图书馆 CIP 数据核字（2009）第 034280 号

责任编辑：王　青
责任校对：宋玉莲
责任印制：李红英

出版发行：清华大学出版社　　**地　　址**：北京清华大学学研大厦 A 座
http://www.tup.com.cn　　**邮　　编**：100084
社　总　机：010-62770175　　**邮　购**：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
印 装 者：北京市清华园胶印厂
经　　销：全国新华书店
开　　本：185×260　　**印　张**：8.25　　**字　数**：187 千字
版　　次：2009 年 5 月第 1 版　　**印　次**：2009 年 5 月第 1 次印刷
印　　数：1～4000
定　　价：18.00 元

前　言

任何一项具体的经济活动都具有行为特征，表现在管理和财政财务收支核算上，这种行为特征又会转换为数据特征。如何从海量的电子数据中发现特征和分析特征，是计算机审计在理论上必须研究和探索的一个问题，也是在审计实务中必须解决的一个问题。只有解决好这个问题，才能引导审计人员自觉地、有效地在把握总体的基础上，突出重点，捕捉审计线索，收集证据，对经济活动做出客观公正的评价。从2007年开始，审计署京津冀特派办设立了专门的课题，成立了专门的研究小组，研究审计线索的特征发现和特征分析，并把研究成果及时推广到审计实践中，在具体的审计项目中实践和验证，收到了良好的效果。

本书总结了审计线索特征发现的一般过程，介绍了特征发现的常用技术和方法。从一个全新的角度把数据分为数值型数据和非数值型数据两大类型，对两类数据的关联、分析方法，如何在两类数据的平台上进行特征捕捉进行了详尽的阐述。与以按图索骥的形式介绍技术、方法、步骤的书籍不同的是，本书的着力点是详细阐述特征发现的思路，书中列举了大量的案例，以期从实际出发，对审计实务工作者和理论研究者有所启发。

刘汝焯设计了全书的章节，并负责统稿，孙俭、陈峰、王秦辉、郭宁、李金龙、程建勤执笔撰写了有关章节。王智玉审定了全书。

审计线索的特征发现和特征分析还处在探索的过程中。书中一定有很多不成熟的地方，甚至可能有错讹之处。真诚地希望广大读者批评指正！

刘汝焯

2008年11月6日于津门

目　录

第1章 概 述

所谓特征，是指可以作为事物特点的征象、标志等。在信息化环境下，审计线索会通过电子数据表现出一定的特征，捕捉到这些特征进而进行分析取证，是计算机审计发展到目前阶段的一种有效做法。这是一个从海量数据中提取符合条件的数据并获取相关信息的过程，是一种基于审计中间表的知识发现(Knowledge Discovery in Database，KDD)的技术。审计线索的特征发现强调的是一种思路，不是就技术讲技术，就方法讲方法，而是计算机审计向前迈进的又一个坚实的步伐，是从必然王国迈向自由王国的重要一步。

1.1 什么是特征发现

审计之所以能够发现问题，首先要有审计线索。而几乎所有的问题，在信息化环境下，都会在电子数据或信息系统中存在一些蛛丝马迹，这些蛛丝马迹就是审计线索的特征。本节以一个案例来证明上面的陈述，把读者的思路引导到特征发现上来。

1.1.1 福尔摩斯与特征发现

侦探小说因其曲折的故事情节、严密的逻辑推理而深受很多读者的喜爱。在众多的侦探小说中，福尔摩斯是英国推理小说家柯南道尔塑造的一个著名侦探。我们在看完一系列关于福尔摩斯的侦探小说后会这样描述他的外貌特征：六英尺高，身材显得格外颀长；目光锐利，细长的鹰钩鼻子；下颌正方突出，招牌形象就是头戴一顶鸭舌帽，嘴叼烟斗。有了这些外貌特征，即使我们没有看见过福尔摩斯，也能很清晰地勾画出他的形象。

福尔摩斯的神奇之处在于他经常能准确地描述凶手的特征。例如他在一次观察作案现场后说："凶手是个男人，他身高六英尺多，正当中年。脚小了一些，穿着一双粗皮方头鞋子，抽的是印度雪茄烟，他是和被害者同乘一辆马车来的，马车用一匹马拉着，右前蹄的蹄铁是新的。凶手很可能是脸色赤红，右手指甲很长。这只是几点迹象，但也许对你们也有点帮助。"

这就是福尔摩斯的探案模式：一言不发地在现场附近仔细勘查，并不时把什么东西放进信封，然后站起身把凶手的特征说得清清楚楚。于是，所有的人都去找全世界符合特征的马车夫。我们暂且不探究福尔摩斯是如何发现这些特征的，指出如此具体、形象的特征对于破案来说无疑是巨大而直接的帮助。

1.1.2 审计线索的特征发现

审计过程中能否如同福尔摩斯一样有效地列举线索的特征，然后通过一定的技术和方法准确地找出符合这些特征的证据，将在很大程度上关乎审计的成败。下面以利用银

行承兑汇票骗取银行资金问题为例，分析该问题在电子数据上的特征以及特征类型，并对特征发现过程进行介绍，使读者对审计线索的特征发现有一个初步的印象。

1. 寻找特征

在福尔摩斯的探案模式中，归根结底是与人打交道，这是一个“人—人”系统，侦探是系统的一方，罪犯是系统的另一方。要与对方交往进而驾驭对方，侦探必须透彻地了解和把握对方的心理特征、心理过程和个性倾向。信息化环境下审计线索的特征发现则是一个“人—机”系统，审计人员是系统的一方，要从被审计单位信息系统的海量数据中寻找特征，就必须首先从业务出发，在深入了解业务规律的基础上进一步了解和把握信息系统和电子数据的特点和规律，进而列出审计线索的特征。

下面以如何寻找违规签发银行承兑汇票中的特征为例来说明。银行承兑汇票是银行在商业汇票上签章承诺付款的远期汇票，是由银行承担付款责任的短期债务凭证。目前，银行承兑汇票已成为企业的重要融资工具，也是商业银行的一项重要业务。但是随着这项业务的不断发展，商业银行办理银行承兑汇票的风险越来越大，利用银行承兑汇票骗取银行资金的案件也时有发生。审计中，把握这类案例在电子数据上的特征，采用计算机审计技术进行数据分析，可以迅速发现问题线索，实现突破。

按照正常业务规范，商业银行开具银行承兑汇票有两个很重要的条件：一是要有真实的商品或劳务交易；二是申请人必须提供足额的保证金存款或质押。如果符合这两个条件，一般来说就不可能发生骗取银行资金的问题。因此，犯罪分子必然会采取伪造贸易背景、提供虚假质押和存单等手段以达到骗取资金的目的。此外，由于银行承兑汇票的期限一般较短，犯罪分子为了达到长期占有资金的目的，往往会采取滚动开票的方法，即当一张汇票到期时再签发一张新汇票以兑付该汇票。

从银行承兑汇票业务的电子数据来分析，由于犯罪分子伪造了必要的资料，所以往往很难从数据上来断定其商品劳务交易、提供的保证金和质押是否真实，但滚动开票则具有“前一笔承兑汇票的到期日与后一笔的出票日相同或相近”这一明显的数据特征。而从正常业务逻辑规律来分析，如果某项银行承兑汇票业务具有真实贸易背景，则其出票日期会依据交易的需要而确定，出现滚动开票数据特征的概率较小。因此，把握滚动开票在电子数据中所表现出的不符合正常业务逻辑规律的数据特征，就容易确定审计重点。

2. 寻找证据

在柯南道尔的小说中，福尔摩斯在列举出凶手的特征后，警察当局是发动所有的人满世界去寻找符合这些特征的马车夫。而在信息化环境下的审计过程中，由于电子数据反映了被审计单位的经济活动，因此特征列举后我们可以用计算机语言描述这些特征，然后在海量的数据中高效、准确地找出符合这些特征的数据。

在利用银行承兑汇票骗取银行资金的案例中，与银行承兑汇票业务相关的电子数据包括“承兑合同表”和“出票人基本信息表”。对相关数据整理后，生成构建该分析模型所需的审计中间表——“承兑汇票信息表”。“承兑汇票信息表”包括“机构编码”、“承兑合同编号”、“出票人代码”、“客户名称”、“经营范围”、“出票人开户银行”、“出票人存款账号”、“收款人全称”、“收款人开户银行”、“收款人存款账号”、“币种”、“汇票金额”、“出票日期”、“到期日期”、“实存保证金总和”、“担保方式”、“合同状态”和“已承兑金额”共18个字段，

其中与滚动开票数据特征相关的有“客户名称”、“汇票金额”、“到期日期”和“出票日期”4个字段。

在滚动签发银行承兑汇票中，应重点关注同日滚动（即前一笔承兑汇票的到期日与后一笔承兑汇票的出票日相同）的记录，因为这种滚动签发的方式下承兑申请人实际上没有对上一笔银行承兑业务进行解付。因此，我们可以用如下的分析模型来描述这个思路：

```
select a.客户名称,a.经营范围,a.汇票金额,a.出票日期,a.到期日期,
    b.汇票金额 as 下一笔汇票金额,b.出票日期 as 下一笔出票日期,
    b.到期日期 as 下一笔到期日期
into 分析表_同日滚动开票表
from 承兑汇票信息表 a
join 承兑汇票信息表 b
on a.客户名称 = b.客户名称 and a.汇票金额 = b.汇票金额 and a.到期日期 = b.出票日期
```

上述运算模型可以筛选出同日滚动开票的记录，再结合审计实际情况，综合考虑企业经营范围、单笔汇票金额等因素，即可确定审计重点，进行精确延伸。

3. 典型案例

在2005年对某商业银行的审计中，审计人员把握住了利用银行承兑汇票骗取银行资金问题的数据特征，根据审计分析模型运算的结果，再结合审计中了解的实际情况，将滚动开票金额大、笔数多的A集团公司、B投资有限公司以及这两家公司的关联企业J置业有限公司确定为重点进行延伸审计。

经延伸审计，A集团公司、B投资公司、J置业公司滚动开票存在明显问题和突出风险。首先，票据无真实贸易背景；其次，资金流向异常，开票后立即由关联企业贴现，甚至是在承兑行直接贴现，贴现资金划往证券公司，用于虚假注资和转入个人账户；最后，审计期间A集团公司停止滚动开票，转为国债质押贷款。上述三家公司办理的滚动开票业务分别是以国债质押、50%保证金加企业担保和全额保证金这三种方式进行担保，表面上看风险较低，但根据其资金流向异常的现象，审计组决定对承兑银行进行风险提示，提醒该商业银行关注三家公司的银行资金风险。此后银行方面虽采取措施防范风险，但为时已晚。2005年5月底，该商业银行持A集团公司质押的到期国债去财政部门兑付，被认定为虚假国债，拒绝兑付。该行对A集团公司发放的贷款余额8 000多万元面临损失风险。2005年6月初，J置业公司的5 000万元银行承兑汇票和5 000万元贷款，由于担保方以担保合同系伪造为由，拒绝承担担保责任，该行1亿元资金面临损失风险。延伸审计还发现，当地另外一家银行向A集团公司及其关联企业发放的1亿元贷款也存在重大风险。

1.2 审计线索特征发现的概念

特征是指可以作为事物特点的征象、标志等。审计线索的特征发现则是指从大量的数值型和非数值型数据中提取有用的信息和知识的过程。

审计线索的特征发现可以分为已知事件的特征发现、未知事件的特征发现以及征兆发现等。已知事件的特征发现是指审计人员主要依据历史案例、业务处理逻辑等建立模

型进行特征发现。在分析过程中，通常已知某些违规违纪行为的特征表现，列举出特征，然后运用一定的技术方法寻找符合特征的数据，并进一步分析取证。未知事件的特征发现是指运用数据挖掘等技术方法发现事件的特征，这些特征在得出挖掘结果之前审计人员是无法预测的。而征兆发现与一般特征的发现有很大的差异：特征是指事件（问题、案件）已经发生，而征兆则是指事件尚未发生或正在进行当中。因此，对已知事件、未知事件的特征发现以及征兆发现的一般过程和技术方法都是不同的。

需要说明的是，本书所指的审计线索的特征发现是指以计算机处理为基础的信息化环境下的特征发现，因此审计人员面对的是电子数据。在这里，我们把电子数据分成两大类，即数值型数据和非数值型数据。数值型数据是指数据库系统中的数值类型、货币类型结构的数据。这是我们在实践中遇到最多的情况，并且已经在这方面作了较为深入的研究，例如关系数据库数据等。非数值型数据是指文本文件、图像、声音乃至网页等其他结构的数据，如何利用非数值型数据进行特征发现是值得我们进行探索和研究的全新领域。

1.3 审计线索特征发现的技术

纵观计算机审计的发展历程，计算机技术和方法在其中扮演着十分重要的角色。可以说，计算机审计的发展在一定程度上是以相关技术和方法的发展为导向的。广大审计人员在计算机技术和方法领域已经并正在进行深入而富有成效的探索和研究，有力地推动着计算机审计实践的发展。

SQL 查询和多维分析技术是目前计算机审计较为成熟的技术。SQL 查询技术是审计人员需要掌握的一门基本功，掌握之后审计人员在审计过程中就可以根据不同的分析需要，通过编写 SQL 语句设置各种条件对数据进行查询分析。

多维分析是充分利用电子数据的特点和规律，对海量数据进行切片、切块、旋转、钻取、挖掘等多角度立体分析处理，发现趋势和异常的一种分析技术。它支持审计人员从不同的角度快速灵活地对数据库中的海量数据进行多角度查询和分析，并以直观易懂的形式将查询和分析结果展现给审计人员。

查询分析和多维分析已经在审计实践中得到了广泛应用，同样也是审计线索特征发现过程中十分重要的技术。关于这些技术的操作方法和具体运用可以参考《审计数据的多维分析技术》（刘汝焯，北京：清华大学出版社）等书，本书不再赘述。然而随着审计实践的发展，上述传统分析技术在很多项目中已经不能完全满足我们的全部需求。如何在缺乏审计经验的情况下对海量数据进行特征挖掘，如何处理非数值型数据等等是值得我们探索的新领域。因此，诸如数据挖掘、非数值型数据的文本挖掘、征兆发现、探索性数据分析等技术日益成为计算机审计重要的研究方向和技术。

1.3.1 数据挖掘技术

数据挖掘（Data Mining），是指从大量的、不完全的、有噪声的、模糊的、随机的实际应用数据中，提取隐含在其中的、人们事先不明确但又潜在有用的信息和知识的过程。其他类型的数据分析，如查询分析、多维分析等，一般是根据已知的知识去提取符合条件的数

据并获取相关的信息，而数据挖掘则是在知识未知或不明确的前提下去发现知识、挖掘信息。数据挖掘包括对数值型数据的挖掘和对非数值型数据的挖掘。本书介绍的数据挖掘指的是对数值型数据的挖掘。反映被审计单位经济活动的财务数据和业务数据绝大部分都是数值型数据。从审计业务角度来看，对数值型数据的挖掘就是根据事先明确的审计目标，对被审计单位的大量业务数据进行分析，揭示其中潜在的逻辑关系和规律，进而形成明确且有效的审计思路的过程。

作为一个知识发现的过程，数据挖掘技术在审计中应用的重要价值主要体现在可以依据其发现的知识来构建审计分析模型。构建审计分析模型的一个先决条件是有明确的审计思路，而审计思路的正确与否也关系到构建的审计分析模型是否准确有效。数据挖掘发现的知识经过解释和评估就可以形成明确的审计思路，并且知识发现的过程不受审计人员主观因素的影响，也使审计思路的客观性得以保障。例如，通过对商业银行信用卡恶意透支的历史数据进行数据挖掘，可以发现“信用卡恶意透支与信用卡客户的收入状况、平均消费额、职业、年龄等客户属性之间的联系”这类的知识，以此为基础构建相应的审计分析模型并运用到被审计商业银行的当前信用卡业务数据中，审计人员就可以快速确定审计重点。

1.3.2 征兆发现技术

征兆发现是一种新兴的数据分析技术。所谓征兆(Sign)，是指系统中已经发生的事件或者已经存在的状态，该事件/状态能够对相关的决策产生重要的影响。作为能够影响决策的征兆可以被理解为机遇或者风险。征兆发现(Sign Discovery)是指从大量的事件和系统状态中甄别征兆的过程。

1.3.3 探索性数据分析

探索性数据分析(Exploratory Data Analysis，EDA)由图基(Tukey)于1972年提出，是数据分析的一种有效方式，茎叶图、字母值、箱线图和批比较、数据变换、耐抗线、双向表、残差分析、稳健估计量等是其分析数据的有效手段。探索性数据分析一方面强调通过分析结果的可视化提高数据分析结果的直观性，同时也可根据分析人员的反馈进行继续的分析，逐步接近问题的实质。与提供汇总结论式的统计分析方式相比，探索性数据分析是一种更直接、更有效的数据分析方式。

数据挖掘、征兆发现是从海量数据中获取高价值事件的有效手段，而探索性数据分析技术一方面可以对海量数据集中的特征数据集进行直观的分析，获得其中的趣味事件；另一方面可以为使用数据挖掘、征兆发现技术获得的事件提供直观的分析结果。

1.3.4 对非数值型数据的处理

数据的形式是多种多样的，可以是数值型的，也可以是文字、图形、图像和声音等非数值型数据。审计中间表涵盖的范围不仅包括数值型的业务数据和财务数据，还有大量的被审计单位以及审计机关在自身的机关管理过程中产生的文本文件等非数值型数据。对这些非数值型数据的分析则涉及文本挖掘、文本自动分类等技术工具。

1. 数字信息检索技术

传统文献信息进行数字化之后，得到的数字化数据大多是非数值型数据，不能直接转化成字段信息，很难用传统的关系型数据软件来管理这些数据。

以文本数据为主要处理对象的数字信息检索系统提供了强大的检索功能，可以直接根据文本的内容进行检索。同时，对于信息资源的综合利用，给予多角度、多层面的支持。全文检索技术是数字化检索系统的主要技术基础。现在的全文查询检索系统按查询方式可分为以下两类。

(1) 主题目录查询。这种检索方式与传统的人工检索系统类似。首先将信息按不同的标准进行分类，一般在大类目下再细分成若干小类目，类目之间按照等级大小排列。用户通过主题目录的指引，逐级、逐层浏览，找到自己所需要的有用信息。按照主题目录进行查询的主要优点是信息通过筛选和系统组织，检索质量较高，层次性、条理性强，检索结果接近用户需求。其缺点是对于原始信息进行加工处理的速度远远跟不上信息的增长速度，检索到的有用信息的数量有限。

(2) 关键词、主题词查询。使用这种检索方式时，首先由用户提出检索要求，再由查询软件负责代替用户在信息库中进行检索，并将检索到的结果及时反馈给用户。在检索过程中，系统软件还可以利用特殊的算法，计算、评估检索到的信息与用户所需信息的相关性，并根据相关性的大小将信息排序，将相关性最大、与用户需求最贴切的信息排在输出信息的前列。

2. 文本挖掘技术

文本挖掘(Text Mining)，顾名思义，就是从非数值型的文本中发现潜在的概念以及概念间的相互关系。传统的信息检索软件所查询的信息可能仅仅从字面上符合查询要求，并不是人们真正需要的信息。文本挖掘技术能够根据用户的真正需要，把与之相关联、有价值以及用户以前未曾注意的有用信息都检索出来。文本挖掘技术是建立在对原始信息的分类和聚类基础上的技术。

第 2 章　审计线索特征的表现

审计是围绕审计线索收集证据的过程，发现线索是审计的重要工作。审计线索是有特征的，这些特征首先表现为行为特征。经济活动，不管是正常的或是异常的，都有一定的行为特征；行为特征又一定会反映在数据中，形成数据特征。审计线索的数据特征可以表现在很多方面，例如可以表现在数据内容上，也可以表现在数据结构上，或者可以表现在功能模块上等。捕捉到数据特征后，可以对数据进行直接分析和比较，也可以通过一些可视化工具（如 Excel 的数据透视图或 Crystal Analysis 等）形象地、直观地观察和分析。

2.1　审计线索特征的表现方式

2.1.1　数据内容

在审计实践中，对大部分违纪违规问题的查处都是从数据内容来发现特征的，如某条记录的值大于既定的阈值、某储蓄所年底储蓄量的异常增长等。

2.1.1.1　审计线索特征在数值型数据中的表现

审计人员在审计实务中通常要面对大量的数值型数据，审计线索的特征也大量地表现在数值型数据中。

下面以某海关审计中审计人员通过分析出口货物申报重量大于码头装船过磅重量的异常信息，发现企业虚假出口、涉嫌走私问题为例，说明如何从数值型数据入手，寻找审计线索的特征表现。

1. 审计思路

货物出口通关流程中，一方面，海关要求出口企业办理通关业务时，必须在舱单、报关单中如实向海关申报装载于出口集装箱内的货物重量；另一方面，出口码头在进场卡口处对装载有出口货物的集装箱进行过磅称重。正常情况下，码头的过磅称重结果扣除出口集装箱自身箱重和合理误差因素后，应和企业舱单、报关单中的申报重量是一致的。但是，考虑到出口货物申报重量是企业行为，有可能存在虚假申报的情况，而码头对出口货物进行过磅称重是为保证装货作业安全和运输工具航行安全的必要措施，相对独立、客观、可信。因此，如果企业存在“多报少出”骗核进口保税料件或骗取出口退税等问题，那么在数据内容上就会表现为企业申报重量与码头过磅重量存在较大差异，特别是申报重量远大于过磅重量这一数据特征。

2. 典型案例

在对某海关的审计中，审计人员统计出了“一票一箱”（即一份报关单申报出口的货物

只装载于一个集装箱，不包括“一票多箱”和“多票一箱”）情况下，“主表_出口舱单集装箱表”和“主表_码头过磅数据表”两张审计中间表，中间表的构建过程本章不作详细介绍，分析模型如图 2-1 所示。

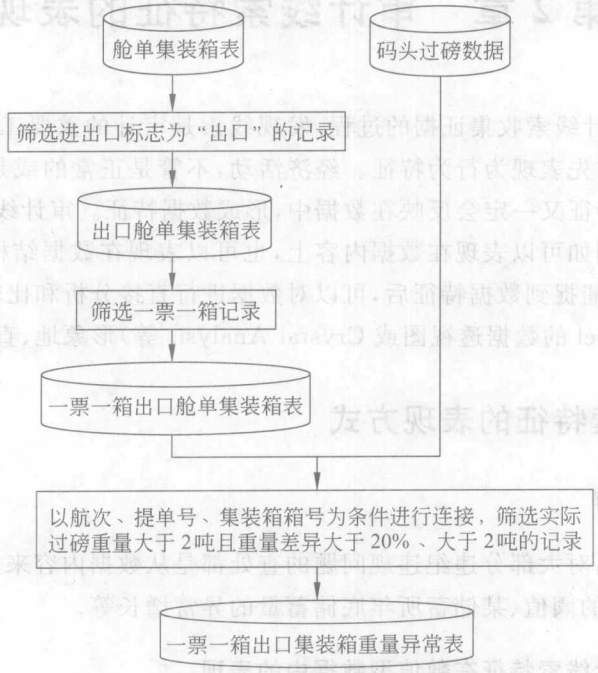

图 2-1　多报少出虚假出口分析模型图

利用舱单数据与码头过磅数据之间的关联关系，筛选出出口集装箱重量异常即企业申报重量与实际过磅重量差异较大的记录。

SQL 语句如下：

```
select b.船名,a.SHIP_ID as 船号,b.航次,b.提单号,b.箱号 as 集装箱号,
  a.CONTA_WT as 集装箱表申报重量,b.过磅重量 * 1000 as 过磅重量,
  a.CONTA_WT-b.过磅重量 * 1000 as 多报重量
into 分析表_一票一箱出口集装箱重量异常表
from 主表_一票一箱出口舱单集装箱表 a
join 主表_码头过磅表 b
  on a.VOYAGE_NO = b.航次
    and a.BILL_NO = b.提单号
    and a.CONTA_ID = b.箱号
where b.过磅重量 * 1000>2000
    and a.CONTA_WT/b.过磅重量/1000>1.2
    and a.CONTA_WT-b.过磅重量 * 1000>2000
order by a.CONTA_WT-b.过磅重量 * 1000 desc
```

经过对比分析出口货物的企业申报重量和相应的码头过磅重量，其中有 1 415 家企

业报关出口的3 972个集装箱货物，企业向海关申报的重量远远大于码头过磅重量。如表2-1所示，某公司4个出口集装箱货物，申报重量达到了过磅重量的1.7～3.8倍。

表2-1　出口集装箱重量比对情况表

集装箱号	出口报关单号	报关单申报重量(kg)	集装箱过磅重量(kg)
××××534720	××××××××××924563	10 385	2 700
××××507690	××××××××××877298	12 840	5 500
××××401180	××××××××××748000	27 592	16 600
××××629275	××××××××××888050	11 394	10 100
	××××××××××888017	10 615	

审计人员根据有关企业出口货物申报重量大于码头过磅重量的数据特征，对包括该企业在内的3家出口企业进行了延伸审计。核查过程中，以企业明细出口装箱单为突破口，发现有关企业均存在以“多报出口数量”和“高报单耗”方式，虚假核销保税料件的问题。

在对另一海关的审计中，审计人员根据同样的数据特征发现，2 985家企业货值××亿元的××××个集装箱，涉嫌虚报、高报出口重量。如某电子股份有限公司有86份报关单申报的86个出口集装箱货物重量远远高于码头过磅重量，涉及出口货值××××万元；其中35个集装箱码头过磅重量显示为空值，涉及出口货值××××万元。

上述问题的发生，与有关部门较为单一地依靠企业申报数据开展监管，缺乏更多的第三方数据进行验证，难以及时有效发现申报数据异常信息有直接联系。在审计过程中，审计人员恰恰可以利用两者数据的独立性及不一致性，通过数据内容的对比分析，发现异常特征，从而为问题的查证提供直接的、有价值的线索。在信息化环境下的审计过程中，大部分违规违纪问题的数据特征表现在数据的内容上。当然，从数据内容入手来进行特征发现的方法是多种多样的。本案例是从内外部数据的对比关系入手来发现特征，此外，还可以从对业务规律的把握入手发现特征、从钩稽关系入手进行特征发现等等。这些都属于特征发现的一般方法，我们将从第4章开始分别进行论述。

2.1.1.2　审计线索特征在非数值型数据中的表现

在审计时，我们越来越多地关注文本文件等非数值型数据，如翻阅被审计单位的收发文簿，查阅总结、报告、合同等文件，到审计现场进行现场观察等。审计实践告诉我们，相当一部分重大审计事项的突破不是从数值型数据开始的，而是从非数值型数据切入的。非数值型数据给予线索性提示，数值型数据起到描述和验证的作用，这已被实践证明是行之有效的方法。

例如在对某省民政厅进行救灾资金审计时，审计组收集了该厅救灾资金下拨及其二级单位使用救灾资金的相关文件、报告等电子和纸质数据。从该省民政厅备灾中心《关于2005—2006年度救灾物资、资金的来源、使用与管理情况的报告》(以下简称《报告》)中了

解到，省备灾中心是省民政厅下属的自收自支事业单位，主要职能是承担中央和省级救灾物资储备与管理、省级救灾捐赠款物的接收与管理以及全省救灾人员的培训等工作。从《报告》中还发现，该中心的经费来源主要是中央和省财政下拨的救灾物资储备管理费、省财政下拨的捐赠工作经费以及中心临街房屋租赁收入。在查阅报告时，审计人员对“中心临街房屋租赁收入”产生了如下疑问：省备灾中心是不是把本该用作救灾物资储备仓库的房屋去搞出租了？经延伸调查备灾中心大楼的规划设计和资金来源后发现，省民政厅确实存在挪用救灾资金建设综合楼的行为，并把大部分面积用于对外出租，而不是用于救灾物资储备。

上述过程就是一个典型的从非数值型数据入手发现审计线索特征的案例。这种思路将会在以后的审计项目中得到越来越多的重视。至于如何运用文本挖掘等技术方法从非数值型数据中发现特征，我们将在本书以后的相关章节中进行讨论。

2.1.2 数据结构

1. 审计思路

在审计项目中，有些线索特征表现在数据结构上。虽然这种情况不是很多，但确实是发现问题的一种重要方法。我们可以通过观察被审计单位业务或财务软件的流程和操作界面，从中发现一些线索。

2. 典型案例

在对 A 航空公司的审计项目中，审计人员根据调查了解的情况，首先通过观察收入结算系统的数据结构发现疑问，进而在数据分析的基础上，通过跟踪被审计单位的业务过程和数据处理流程，发现了被审计单位收入结算系统中存在的非法销售暗扣处理模块。

（1）观察数据字段，发现疑问

在航空运输企业中，收入结算不仅是一个重要的业务环节，而且与财务核算密切相关。因此，审计人员重点关注 A 航空公司的收入结算系统，对系统的情况进行了认真地调查了解。审计人员在前台观察过程中，发现客票的录入、修改等界面包含以下字段：销售类型、代理人、出票日期、承运人、航班号、日期、面额、毛额、净额、批号等。在上述字段中，与金额相关的字段就有三个：面额、毛额和净额。通过浏览数据发现，面额与毛额均相等，而毛额则是大于或等于净额。对此，审计人员就产生了疑问：为什么会有小于毛额的净额？系统是否会以净额生成运输报告，存在净价结算问题呢？因为从调查了解的情况看，航空公司为了提高市场占有份额，可能存在暗扣销售和净价结算行为。那么，数据结构中“净额”字段的存在能否作为航空公司暗扣销售的一个特征表现呢？如果果真如此，那么根据收入结算系统生成的运输报告，进而确认的财务收入就是不完整的（收入结算系统与财务系统之间的业务关系如图 2-2 所示），会影响收入的真实性并存在侵蚀营业税税基的问题。

（2）进行数据对比，验证问题线索

为了验证上述疑问，审计人员根据业务关系，将财务数据与收入结算数据进行了对比分析，以此审查该系统是否以净额生成运输报告。在数据采集、整理完成后，审计人员首先从财务数据中提取了 2002—2004 年的运输收入数据，然后从收入结算数据中提取了相