

# 基于数据仓库的公安决策支持系统<sup>\*</sup>

王素青 李玉忱

(山东大学 计算机科学与技术学院 济南250061)

**摘要** 公安系统近年来在信息领域取得了长足发展,并积累了大量数据,如果在这些数据中发现趋势、找到发案规律,就能给办案带来很大帮助。因此针对公安机关在决策支持方面的不足,提出了基于数据仓库的公安决策支持系统(PDSS)。分析了决策支持系统的设计和结构,研究了系统实现过程中用到的数据仓库、OLAP 和数据挖掘的关键技术,最后结合实例给出了该系统的功能。

**关键词** 数据仓库,决策支持系统,数据挖掘,OLAP 在线分析处理,DSO 决策支持对象

## Decision Supporting System of Police Based on Data Warehouse

WANG Su-Qing LI Yu-Chen

(College of Computer Science and Technology, Shandong University, Jinan 250061, China)

**Abstract** Recently, great progress is made in the information fields of the police. A great deal of data is accumulated. If the rules of the data is found, it would be greatly beneficial to the cases. So a police decision support system(PDSS) is proposed based on data warehouse in order to overcome the deficiencies of the police. The architecture of the system is analyzed. The key technology in data warehouse, olap and datamining are researched. The functions of the system with real instances are given.

**Keywords** Data warehouse, Decision support system, Data mining, On line analysis process (OLAP), DSO (decision support object)

## 1 引言

公安机关在“向科技要警力”、“科技兴警”的目标的指引下,近二十年来在信息领域的建设中取得了长足的发展。各地公安机关建立了大量的业务应用系统,目前许多业务信息系统其性能愈加完善,功能愈加强大,用户越来越多,但同时也暴露出缺点:大量的数据存放在不同的事务处理系统中,这些系统独立运行,难以进行信息的关联。各系统都只提供了某些数据对象的部分特征,缺乏一个跨越多系统对某个数据对象的完整表达。

公安系统在办案和处理日常事务的过程中积累了大量的数据,这些是操作型数据。它是面向事务驱动的。如果能在这些数据中发现趋势、找到发案规律,就能够有助于今后的办案及案情分析。因此建立一个公安决策支持系统(Police Decision Support System,以下简称 PDSS)是十分必要的。如何建立一个公安决策支持系统呢?这里传统的数据库技术不再适用,因为它不能为数据的集成、分析和综合提供强大的功能。为了分析嫌疑人的情况,必须组织起以嫌疑人为中心的数据,这些是传统数据库解决不了的。数据仓库(Data Warehouse,以下简称 DW)的

出现无疑给决策支持系统注入了新的活力。它把决策所需的信息从原始数据中抽取出来,把分散的、难以访问的操作型数据转化为集中统一、随时可用的信息,即 DW 对数据进行合理、全面而高效的管理。

本系统是在充分了解公安业务和公安部金盾工程要求的基础之上,结合数据仓库和数据挖掘的知识,开发出的一套适用于公安系统办案和案情趋势分析的决策软件。系统实现了各种信息之间的共享、互访和关联,为领导和侦察员提供了强有力的信息查询、统计和分析的能力,为破案提供了辅助决策手段。

文章的创新之处在于应用了可视化数据挖掘工具,增加了用户和系统的交互能力。

## 2 数据仓库的结构与设计

为了利用公安系统中的操作型数据,需要经过数据抽取转换、建立数据仓库,才能进行数据分析,为公安系统决策支持提供依据。

### 2.1 DW 体系结构:

由以下四部分组成(如图1所示):

(1)数据源来源于操作型数据库,包括嫌疑人和案件表,SQL Server 2000作为数据库管理工具;本

系统的数据源有两个,一个是本局内的警务动态管理系统中的案件、嫌疑人数据(存放在 SQL Server 2000上),另一个是刑警大队的刑事案件管理系统数据(存放在 Oracle 上)。

(2)数据管理模块:数据抽取转换用到 SQL Server 自带的 DTS 工具,目的数据库是 SQL Server 数据库。建立后的数据仓库是 GongAn\_DW,有事实表和多个维表。数据仓库管理工具用到 SQL Server Analysis Services。

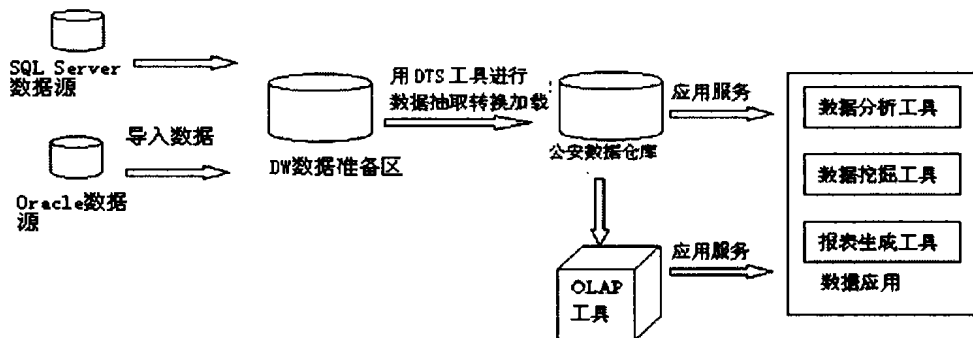


图1 数据仓库体系结构

## 2.2 DW 的设计

考虑到星型模型简单易用,所以采用星型模型作为数据模型。

第一步:通过对业务需求的理解与对可用数据的理解组合起来而确定建模的业务处理内容。公安系统日常的事务处理就是刑事案件、行政案件的处理,以及案件牵涉到的嫌疑人,这样可以确定两大主题:案件和嫌疑人。

第二步:定义粒度,所谓粒度就是数据仓库中数据单元的详细程度和级别<sup>[3]</sup>。原子型数据可以根据维度的级别进行汇总。汇总后的数据与原子数据并存在数据仓库中。本系统中的原子粒度数据就是每天的案件和嫌疑人数;低度细节级数据是每月的汇总数据;高度细节级是每年的汇总数据。

第三步:选定维度,案件的维度包括办案单位、立案时间、破案时间、案别等;嫌疑人的维度包括时间、性别、职业、文化程度、作案手段、身份、年龄等。

## 3 公安决策支持系统的实现

在以上分析设计的基础上建立数据仓库和 OLAP 多维立方体。

### 3.1 加载数据

数据清洗转换,存入数据仓库,数据源有两个,一个是 SQL Server 2000,另一个是 Oracle 数据库。在 SQL Server 2000 的企业管理器中,导入 Oracle 数据库中的数据,把抽取转换出来的数据放在一起,建立数据仓库准备区;对准备区里的数据经过清洗以后,存储到数据仓库 GongAn\_DW 中。装载数据仓库的历年数据时,要根据需求分析中的要求重新

(3)数据分析模块:在 Analysis Services 中建造 OLAP<sup>[1,2]</sup>多维模型,对数据仓库或数据集市中的数据进行多维分析,提供高度交互的在线分析处理功能,可以即时进行反复分析、切片、切块、旋转、钻取等操作,迅速获得所需的结果。同时还提供预定义的数据挖掘及数据查询功能。

(4)数据应用模块:将 OLAP 分析的结果以适当的形式展示给决策者,并对数据仓库、数据集市中的数据进行查询、挖掘、分析等应用。

组织表的结构和格式。利用数据转换服务设置 DTS 包进行简单清洗转换工作,复杂的清洗和转换,需要使用 ADO 的 Recordset 编写程序来实现。具体实施步骤是:①连接数据库,即数据准备区;②清空数据库表;③调用 DTS 包;④清洗;⑤转换;⑥转换后数据存入数据仓库。

### 3.2 建立 OLAP 数据库

在 OLAP Manager Console 中创建 OLAP Services 数据库 GongAn\_OLAP。对数据源进行编辑,在连接项选择服务器和源数据库 GongAn\_DW,这样就创建了与数据仓库中数据的连接。

3.2.1 创建维度 OLAP Manager Console 中创建基于源数据表的各维度。用新建维度向导或维度编辑器建立维度,考虑到数据更新维护和下面创建多维数据集的需要,创建了20个星型架构的公共维度,新建维度要进行处理,处理后的维度能进行维度层次和维度数据的浏览。

3.2.2 创建多维数据集 用新建多维数据集向导或多维数据集编辑器创建多维立方体,将数据仓库事实表和相关各维连接起来。一个多维数据集(Cellset)就是一个立方体,是包含维度和度量值的多维结构<sup>[4]</sup>。维度定义立方体的结构,而度量值提供最终用户感兴趣的数值。事实表包括案件和嫌疑人表;系统根据业务需求,创建了22个多维立方体。对多维立方体存储,我们用 MOLAP 方式。创建完毕,在多维数据集编辑器中给度量值定义聚合函数 COUNT。最后进行多维立方体处理,对立方体进行处理以后就可以浏览数据,图2所示就是嫌疑人文化程度立方体的数据,可以对分析维度文化程度进行

钻取,或是改变分析角度,从时间角度分析嫌疑人数据。

Class A	Class B	犯罪嫌疑人人数
文化程度	文化程度: 会计	71
	职业: 教师	
	大学专科(普通大学)	
	大学	
	硕士研究生	
	硕士研究生(大学硕士)	
教育水平	教育水平: 会计	66
	职业: 教师	
	大学	
	职业: 教师	
	职业: 教师	

图2 嫌疑人多维立方体数据浏览

### 3.3 数据更新操作

数据仓库中数据加载更新以后,OLAP 数据库中的维度数据和立方体数据并不会随之更新,要手动控制这部分操作。维度更新处理包括增量更新和重建维度结构;立方体的处理在维度处理之后,不然就会处理出错。立方体的处理包括增量更新、刷新数据和完整处理。增量更新是将数据追加到多维数据集,而不会重新计算聚合。要想周期性地更新数据仓库,也必须定义更新周期,调用增量更新包,用 VB 中的 DSO(decision support object)模块编程实现。

DSO 更新算法: DsoRefresh 算法

输入: OLAP 数据库名

输出: 更新数据成功或失败

方法: ①定义 DSO 对象模型; ②DSO 对象赋值; ③连接 OLAP 数据库; ④更新 OLAP 数据库; ⑤for num=0 to 获取的立方体数对 OLAP 立方体或者维度进行更新操作; ⑥Next num。

### 3.4 OLAP 数据的前台展示

要访问创建的多维立方体必须用到两种技术: MDX 语言和 ADOMD 组件。我们用 VB 的 ADOMD(ADO to include objects specific for multidimensional data)编程展示 OLAP 立方体的数据分析结果。具体过程如下: 前台应用程序根据用户的查询条件生成多维表达式(MDX),然后将此表达式提交给 ADO MD 的 CellSet 对象执行,CellSet 对象根据此表达式从多维数据集中返回相应的数据。下面以图表展示为例说明这两种技术应用:

图表展示算法: DataShowing 算法

输入: chart 图形展示种类; chart 图形展示标题以及 MDX 查询语句。

输出: MSChart 展示效果图

方法: ①连接 OLAP 数据库; ②Set Cellset. Source = MDX 查询语句; ③打开记录集 Cellset; ④调用图表展示过程 fMSChartZhanshi; ⑤图表展示记录集 Cellset; ⑥关闭记录集 Cellset。

## 3.5 数据挖掘应用

数据挖掘工具根据系统需求需要用到聚类和关联规则。利用 Kohonen 的 SOM (Self-organizing Maps)<sup>[5]</sup>作为交互式可视化数据挖掘的起点,由用户动态选择感兴趣的区域进行深入分析。数据挖掘过程如下: ①SOM 模型定义; ②SOM 模型管理; ③SOM 图形生成及显示; ④聚类和关联规则算法库。

## 4 系统的功能特点及优势

### 4.1 系统的设计思想

在开发和设计中,立足于数据处理和分析;立足于信息关联共享;立足于数据处理的广度;立足于为侦案破案提供有力依据和手段;立足于为公安管理提供辅助决策。具备数据分析与辅助决策功能,是公安决策支持系统的设计目标之一。它与一般统计功能区别在以下几方面:

1. 公安综合系统的分析与决策功能,是为高层领导和指挥决策人员服务的。它能在大量数据中,发现规律,找出问题,从而为打击犯罪、预防犯罪以及提高公安系统内部管理水平提供有力的辅助手段。

2. 一般统计功能往往只能输出预先定义好的报表,而决策与分析系统则是针对特定主题,从宏观的,历史的角度对信息进行分析挖掘,用户可以和系统进行交互,从而得出结论的一个过程。

3. 由于存在以上特点,所以一般的报表统计打印软件是不能胜任的。决策与分析系统需要在具有大量参考数据和高性能数据仓库系统的基础之上,利用专用的数据分析与挖掘工具软件实现辅助决策的功能。

### 4.2 系统的功能

分析主题有刑事嫌疑人、刑事案件、行政嫌疑人、行政案件。分析的角度: 嫌疑人的作案时间、文化程度、职业、身份、性别、年龄、作案手段等等。图的类型有棒图(如图3),和折线图、散点图等。本系统具有以下功能:

1. 利用综合信息系统提供的综合数据及关联信息进行统计分析,辅助领导决策,根据统计分析的结果,可以发现公安业务中的薄弱环节,对改进公安工作有很大帮助。如:对发案情况的分析。分析发案的趋势和规律,各种案件之间的联系,有效地预防、减少犯罪。

2. 应用数据挖掘可视化技术,通过对数据进行各种算法分析,交互式发掘隐藏在数据中的信息,发现犯罪率升高或降低的原因。

3. 应用 OLAP 技术,将有关信息抽取出来进行展示,从文化程度角度、职业角度、作案手段等角度查询犯罪趋势。

4. 能够分析案件、嫌疑人的类型,有针对性的制

定应对措施。

5. 对图形进行旋转、缩放和筛选操作,增加信息展示的角度。比如在散点图中选择感兴趣的点,形成一个子集,进行钻取操作。

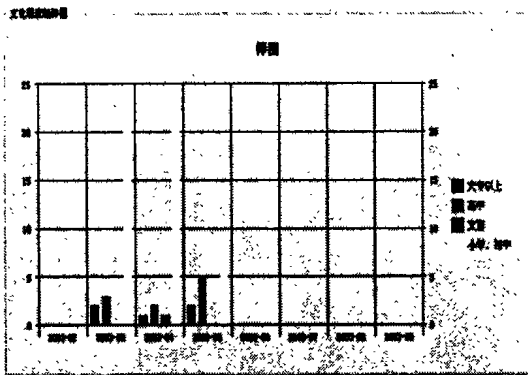


图3 棒图——嫌疑人趋势图

### 4.3 系统优势

- 1、任选时间段,灵活设定条件得出分析报告。
- 2、系统交互性强,用户可以以多种方式和挖掘模型交互。
- 3、可扩展性好。转换规则和挖掘算法是封装的、模块化的,系统提供了一个开放的、灵活通用的接口,使用户能够加入新的规则和算法。
- 4、对多个数据源进行自动高效率的数据收集。
- 5、分析结果深入、详尽,且易读性、图表化程度

(上接第259页)

请求,以 HTTP 协议与平台运行实例进行通信,获得平台运行的状态,并将结果返回给用户。使用户可以根据平台的状态对平台实例的运行进行实时干预,如启动、停止和重新启动等操作。

**结束语** 在电子政务系统结构中,中间件平台是应用支撑层的最核心的组成部分,它位于系统软件与应用软件之间,是政务系统建设中的重要软件基础设施。本文提出的平台以构件技术和中间件技术为基础,以软件框架技术为核心,从技术角度为电子政务系统的互联互通提供了安全可靠的解决方案。该平台已经在多个电子政务的工程项目中得到应用,很好地满足了客户的需求并取得较好的经济效益。

本平台的设计与实现是863课题的延伸,初衷是

高。

6、深入挖掘各类运行数据,适应高层次专业分析需求。

**结束语** 数据仓库、OLAP 和数据挖掘推动了决策支持系统的快速发展,目前市场上也推出不少的产品,但是用户要根据自身需要和所要解决的关键问题选择合适的 DSS 工具。本文从数据仓库的建立到 OLAP 立方体的建立,作了详细的介绍,并利用合适的数据挖掘算法,建立了一个完整的决策支持系统。用 VB 设计展示工具和查询处理工具,对用到的技术作了概述。

本设计现已应用在公安系统中,系统启用以后,充分利用了违法犯罪的资源,提高了办案效率和管理水平。系统在不断完善的同时,准备增加网上查询和网上分析数据的能力。

### 参考文献

- 1 Peterson T, Pinkelman J. Microsoft OLAP Unleashed. Sams, 2000
- 2 Thomsen E, Spofford G, Chase D. Microsoft OLAP Solutions. New York: Wiley, 1999
- 3 巴斯蒂安 M[德]. 数据仓库与数据挖掘. 冶金工业出版社, 2003. 50~56
- 4 George S[美]. Microsoft OLAP 解决方案. 人民邮电出版社, 2000. 12~17
- 5 Kohonen T. Self-organizing Maps springer-verlag. Berlin, 1997

为了解决银行业集成所面临的基本问题,我们将其中具有一定普适意义的部分抽象出来就形成一个通用的集成平台。针对银行业的特殊要求,集成平台还有许多深层次的分析 and 设计工作需要进一步展开。

### 参考文献

- 1 Len Bass, Paul Clements, Rick Kazman. Software Architecture in Practice. 北京:清华大学出版社, 2002
- 2 万麟瑞,李绪蓉. 系统集成方法学研究. 计算机学报. 2003, 22(10): 1025~1031
- 3 许焕梅,张思荣,王锋,等. 基于分层的软件系统体系结构. 计算机工程. 3002, 29(20): 57~59
- 4 段军玲,张曙光. 基于 MQ+XML 的安全数据交换模型. 计算机工程. 2003, 29(20): 194~197
- 5 刘瑞,陈微,李昭原. 一种基于构件的企业信息系统开发支撑框架. 计算机工程. 2003, 29(4): 35~37