

⑩
54-58决策支持系统新结构体系^①陈文伟^② 黄金才[✓] 陈元
(国防科技大学系统工程与数学系)TP399
C934

【摘要】90年代中期发展起来的决策支持新技术:数据仓库、联机分析处理和数据开采,它们是从数据库中获取辅助决策信息,为决策支持系统开辟了新途径。本文在讨论了决策支持系统的发展过程、关键技术和研究进展基础上分析决策支持新技术,提出了综合数据仓库、联机分析处理、数据开采、模型库、专家系统的决策支持系统新结构体系。

关键词:决策支持系统,数据仓库,数据开采,模型库,专家系统,联机分析处理

1 基于模型库的决策支持系统

1.1 系统结构形式

自从70年代提出决策支持系统以来,已经得到很大发展。决策支持系统是在管理信息系统(MIS)基础上发展起来的。MIS是利用数据库技术实现各级管理者的管理业务,在计算机上进行各种事务处理工作。DSS要达到为各级管理者辅助决策的能力。

1980年Sprague提出了决策支持系统三部件结构,即对话部件、数据部件(数据库DB和数据库管理系统DBMS)、模型部件(模型库MB和模型库管理系统MBMS)。该结构形式明确了DSS的组成,也间接地反映了DSS关键技术:模型库管理系统,部件的接口,系统的综合集成。该结构形式为DSS的发展起了很大推动作用。

1981年Bonczak等提出了DSS三系统结构,即语言系统(LS)、问题处理系统(PPS)、知识处理系统(KS)。该结构明确了问题处理和语言系统的重要性,统一了知识看法。该结构在一定范围内有其影响,但它忽略了DSS中数据部件与模型部件的关系,不适合与其他系统(如专家系统ES)的区别,对DSS开发是不利的。

90年代,智能决策支持系统(IDSS)逐渐形

结构体系

成,即在三部件的基础上,增加知识部件(知识库KB和知识库管理系统KBMS以及推理机),它是把决策支持系统和专家系统结合起来。

根据实际决策问题, IDSS为实现数据部件、模型部件、知识部件有机综合,形成系统,原DSS的对话部件应该改为“综合部件”或“问题综合与交互系统”,它将模型、数据的定量辅助决策的信息和知识推理的定性辅助决策信息结合起来,更有效地辅助决策。

1.2 基于模型库的决策支持系统的关键技术以及开发的困难

基于模型库的决策支持系统的各部件中只有数据部件有成熟的商品软件可以直接使用,其他部件及部件的接口和集成问题需要开发者自己完成。开发一个决策支持系统的关键技术有:(1)模型库系统的设计和实现,它包括模型库的组织结构,模型库管理系统的功能,模型库语言等方面的设计和实现;(2)知识部件(专家系统)的设计和实现,它包括知识库结构、知识库管理系统功能以及语言、推理机、动态数据库等方面的设计和实现;(3)部件接口,各部件之间的联系通过接口来完成,部件接口包括:①对数据部件的数据的存取,②对模型部件的模型的调用和运行,③对知识部件的知识完成推理;系统综合集成,由综合部件根

① 国家自然科学基金资助项目(79670019)。

② 陈文伟,教授,博士生导师,研究方向:智能决策、机器学习、数据开采、数据仓库。通讯地址:湖南长沙国防科技大学703教研室,邮编:410073。

据实际决策问题的要求,通过综合集成语言完成对各部件的有机综合形成一个完整的系统。

决策支持系统的开发困难与计算机语言的发展有紧密的联系。计算机数值处理语言如 FORTRAN、C、PASCAL 等不具备对数据库的操作功能,而数据库语言如 FoxPro、ORACLE 等适合数据处理但是不适合数值计算,用它们开发决策支持系统都不合适。开发决策支持系统所需的语言(DSS 语言)要有人机交互能力、数值计算能力、数据处理能力及模型调用能力等。解决这一问题的途径有两个:一是针对以上功能自行研制开发一套 DSS 语言,这种途径工作量大,开发周期长,但它是彻底解决 DSS 的各个问题的有效途径,比较典型的如 GFKD-DSS 语言体系;二是扩充现有的语言,选用一种功能强大的语言作为宿主语言,增加其不足的能力,使之达到 DSS 语言的功能要求,比较典型的如 DM IDSSP 选用 Borland C++ 3.1 作为宿主语言,以 CodeBase 作为宿主语言与数据库的接口。近年来,ODBC (Open Database Connection)、BDE (Borland Database Engineer) 等逐渐成为大家接受的数据访问标准。以支持这种标准的语言如 Virtual C++ 等作为宿主语言,以 ODBC 等作为数据库接口标准,已成为近年来开发决策支持系统的趋势。

1.3 决策支持系统的研究进展

作者从 1987 年开始研究决策支持系统,一直紧跟国际先进技术,对 DSS 进行了广泛的研究。1989 年研制完成了“决策支持系统开发工具 GFKD-DSS”,自行研制了一套 DSS 语言体系(DSS 核心语言、模型库管理语言、接口语言、工具操作语言)。用 GFKD-DSS 工具和自己研制的 PROLOG 产生器 P3 开发了南京林业大学的“松毛虫智能预测系统”。

1995 年完成了“分布式多媒体智能决策支持系统平台 DM-IDSSP”。该平台集成了多媒体技术、分布式处理技术、专家系统开发工具、机器学习、神经网络、模型库系统、地形处理等多项新技术。该平台的集成环境是以 Borland C++ 语言为主语言,嵌入数据库接口语言 Codebase 和自行研制的多媒体人机交互语言、推理机语言、分布式支持语言形成的 IDSS 宿主语言体系。用该平台开发了一个大型的军事指挥智能决策系统。

2 决策支持新技术

90 年代中期,由数据库的发展,兴起了 3 个新的辅助决策技术:数据仓库、联机分析处理和数据开采。数据库是为事务处理服务的,是管理信息系统的基础。而数据仓库是根据决策主题的需要汇集大量的数据库,并对它进行重新组织和提取,通过综合和分析得到辅助决策的信息。联机分析处理是以客户机/服务器的方式完成多维数据分析。数据开采是利用一系列方法,从大量的数据库中获取有用的知识。这些知识是对数据库中数据的浓缩,为决策服务。数据开采的概念是从知识发现概念引伸出来的。

2.1 数据仓库

数据仓库可以认为是具有如下特性:(1)面向主题的(subject-oriented) (2)集成的(integrated) (3)时变的(time-variant) (4)稳定的(nonvolatile) 数据集合(collection),用以支持经营管理的决策制定过程。

数据仓库是在原有关系型数据库基础上发展形成的,但不同于数据库系统的组织结构形式,它从原有的业务数据库中获得的基本数据和综合数据被分成一些不同的级(levels)。一般数据仓库的结构组成包括:当前基本数据(current detail data);历史基本数据(older detail data);轻度综合数据(lightly summarized data);高度综合数据(highly summarized data);元数据(meta data)。

整个数据仓库的组织结构是由元数据来组织的,它不包含任何业务数据库中的实际数据信息。元数据在数据仓库中扮演了重要的角色,它被用在以下几种用途:起到辅助 DSS 分析过程中定位数据仓库的目录作用;数据从业务环境向数据仓库环境传送时数据仓库的目录内容;指导从近期基本数据到轻度综合数据,轻度综合数据到高度综合数据的综合算法选择。元数据至少包括以下一些信息:数据结构、用于综合的算法、从业务环境到数据仓库的映射。

2.2 联机分析处理(OLAP)

1993 年,Codd 提出 OLAP 概念,主要是实现客户/服务器方式的多维数据分析。维是同类数据的集合,如各个商店的集合是一维,时间集合是

一维,商品集合是一维.多维数据结构是从数据中提取有用信息的有效组织形式.多维数据结构的物理存储形式有两种:基于关系数据库的ROLAP和多维数据库的MOLAP.

(1) ROLAP型数据组织形式一般采用“星型模型”,它有一个“事实表”和多个“维表”组成.“事实表”中每条元组都包括指向各个“维表”的外键和相应的测量数据,数据量很大,记录达到上万条.“维表”中记录有关这一维的属性,数据量一般较小.

(2) MOLAP数据组织是基于多维数据库的,一般采用(1)超立方体结构(hypercube),用三维以上的维数来描述一个对象,数据测量值发生在维的交叉点上;(2)多立方体(multicube)结构,将大的数据结构分成多个多维结构.这些多维结构是整个数据的子集.

OLAP的主要工作是多维数据分析,具体有:

(1)切片和切块(slice and dice)

在多维数据结构中按二维进行切片,按三维进行切块,得到所需的数据.

(2)钻取(drill)

钻取有向下钻取(drill down)和向上钻取(drill up)操作.

(3)旋转(pivoting)

通过旋转可以得到不同视角的数据.

2.3 数据开采

数据开采(DM)被认为是知识发现(KDD)过程中的一个特定步骤,它用专门算法从数据中抽取模式(patterns).知识发现(KDD)被认为是从数

据中发现有用知识的整个过程,它从在数据中鉴别出有效的模式的非平凡过程,该模式是新的,可能有用的和最终可理解的(Fayyad, Piatetsky-Shapiror, Smyth 1996).

KDD过程是多个步骤相互连接起来,反复进行人机交互的过程.在其中DM是一个非常重要的阶段,这也无怪乎有的论文作者把DM和KDD看成一个概念(Cure Hall 1995).目前,在KDD领导进行的研究主要集中在DM上,它是从大型数据库或数据仓库中发现并提取隐藏在其中的模式的一种新技术,目的是帮助决策者寻找数据间的潜在的关联,发现被忽略的因素.而这些因素对决策行为是十分有用的.DM技术涉及数据库、AI、机器学习和统计分析等技术,主要处理相关分析、聚类、概念描述、偏差检测及预测等任务.其具体方法和技术有:(1)决策树方法(2)神经网络方法(3)覆盖正例排斥反例方法(4)粗集(Rough Set)方法(5)概念树方法(6)遗传算法(7)公式发现(8)统计分析方法(9)模糊论方法(10)可视化技术.

2.4 基于数据仓库的决策支持系统

新的决策支持技术的发展和基于模型库的DSS的开发的困难,促使人们重新考虑DSS的结构体系.于是兴起了DW+OLAP+DM=DSS的解决方案.DW、OLAP和DM是作为3种独立的信息处理技术出现的.DW用于数据的存储和组织;OLAP集中于数据的分析;DM致力于知识的自动发现.这3种技术具有内在的联系性和互补性,把它们结合起来,形成了新的决策支持系统结构.如图1所示:

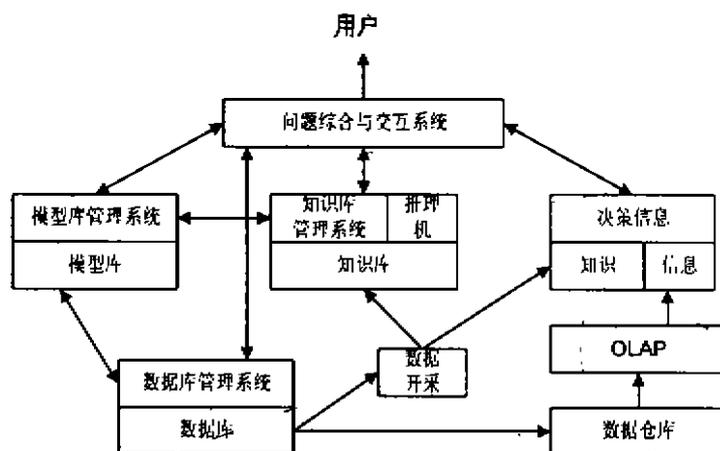


图1 DW+OLAP+DM的DSS结构

该结构的特点是以大量的数据为基础,系统由数据驱动, DW 对底层的业务数据进行集成、转换和综合,重新组织成面向决策主题的全局数据视图, OLAP 是在全局数据视图上进行多维数据分析, 数据开采是从数据库或数据仓库中进行挖掘, 获得辅助决策信息, 由于 DW、OLAP、DM 都是数据驱动的, 故它们的共同点是寻找大量数据中的规律性, 从三个不同的角度为决策者提供辅助决策信息, 这是决策支持系统新的发展方向。

不可否认, 这种 DW+OLAP+DM 的结构体系在一定程度上满足了决策问题的需要, 但是它完全屏弃模型库的做法并不妥当, 主要问题有:

- 许多决策问题必须要借助于模型来完成, 如一个地区的经济规划问题, 它要借助于许多运筹学模型(线性规划模型、投入产出模型等)来完成。

- DM 抽取得到的是知识, 而知识要有效发挥作用必须借助于各种推理机, 从本质上讲, 推理机也是一种模型, 它既不属于 OLAP 工具, 也不属于 DM 工具。

因此, 在实践的基础上提出决策支持系统新结构体系。

3 DW+OLAP+DM+MB+ES 的综合决策支持系统的新结构体系

3.1 智能决策支持系统和数据仓库、OLAP 和数据开采的结合

智能决策支持系统主要是以知识和模型为主体, 结合大量数据, 形成定性和定量相结合的辅助决策信息, 智能决策支持系统由于开发的困难, 发展较慢, 但已经形成了研究热潮, 取得了不少成果, 为各级决策者进行科学决策提供了强有力的手段。

数据仓库、OLAP 和数据开采这种新的决策支持技术是从大量数据库中获取有用信息和知识, 它来源于数据库内部, 它完全不同于智能决策支持系统的知识和模型是来源于数据库的外部, 可见, 这两种辅助决策的方式是互不相同的。

智能决策支持系统的新结构体系是将智能决策支持系统与数据仓库、OLAP 和数据开采结合

起来, 即将两种不同的辅助决策方式结合起来, 起到相辅相成的作用, 进一步提高辅助决策的效果。

3.2 综合 DSS 结构体系

决策支持系统的新结构是在智能决策支持系统的结构上增加数据仓库、OLAP 和数据开采的结构, 统一由“问题综合与交互系统”进行综合集成, 结构形式如下图 2 所示。

综合决策支持系统体系结构包括 3 个主体, 第 1 个主体是模型库系统和数据库系统的结合, 它是决策支持的基础, 它为决策问题提供定量分析(模型计算)的辅助决策信息, 第 2 个主体是数据仓库、OLAP, 它从数据仓库中提取综合数据和信息, 这些数据和信息反映了大量数据的内在本质, 第 3 个主体是专家系统和数据开采的结合, 数据开采从数据库和数据仓库挖掘知识, 放入专家系统的知识库中, 进行知识推理达到定型分析辅助决策。

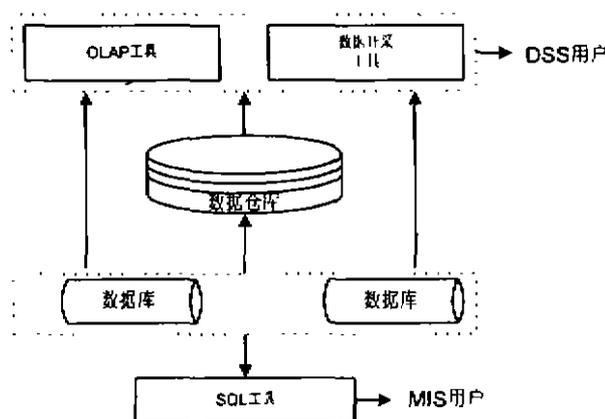


图 2 综合决策支持系统的结构

综合体系结构的 3 个主体既可以相互补充, 又可以相结合, 根据实际问题的规模和复杂程度决定是采用单个主体辅助决策还是采用两个或是三个主体的相结合辅助决策, 利用第 1 个主体(模型库和数据库结合)的辅助决策系统就是传统意义下的决策支持系统, 利用第 1 个主体和第 3 个主体(专家系统和数据开采)相结合的辅助决策系统就是智能决策支持系统, 利用第 2 个主体(数据仓库和 OLAP)的辅助决策系统就是新的决策支持系统, 在 OLAP 中可以利用模型库的有关模型将可以提高 OLAP 的数据分析能力, 将 3 个主体结合起来, 即利用“问题综合和交互系统”部

件集成3个主体,这样形成的综合决策支持系统,是一种更高形式的辅助决策系统,辅助决策能力将上一个大台阶.由于这种形式的决策支持系统包含了众多的关键技术,研制过程中将要克服很多困难.

3.3 综合 DSS 结构体系的关键技术

新结构的关键技术包括智能决策支持系统的关键技术以及数据仓库、OLAP 和数据开采的关键技术.前者已经在第1节说明.数据仓库和数据开采的关键技术有:

(1) 数据仓库的结构和数据仓库管理系统

数据仓库的数据量是大约100倍于数据库,它的存储结构形式的设计以及数据查询和提取的效率就显得很重要.一般采用分布式结构形式.数据仓库管理系统的一项重要工作是实现对传统数据库进行提取、清理和装载数据到数据仓库中.这项工作占整个管理工作的80%.

(2) 数据仓库中的分析工具

数据仓库虽然存有大量的数据,但提供辅助决策的信息需要利用各种分析工具,如 OLAP 工具、DM 工具、统计分析和查询优化工具等.这些工具直接影响辅助决策的效果.

(3) 数据开采的方法

目前,已经形成了多种数据开采方法,如第2节的说明.各种方法均有它一定的优点,也有它的

一些不足.还需研究更多的数据开采方法,以及多种方法的综合,使数据开采得到的知识具有更强的辅助决策效果.

3.4 辅助决策的新结构体系的研究

我们正在按新的结构体系进行研究解决各项关键技术,在已完成的“分布式多媒体智能决策支持系统平台 MD-IDSSP”中,已完成了数据开采的多个新方法.正在研制数据仓库原型系统,向综合决策支持系统的新结构体系迈进.

参 考 文 献

- 1 陈文伟编著.决策支持系统及其开发,清华大学出版社和广西科技出版社,1994.10
- 2 陈文伟等.GFKD-DSS 决策支持系统开发工具,计算机学报,1991,14(4)
- 3 陈文伟等.分布式多媒体智能决策支持系统平台(DM-IDSSP)技术报告,国防科技大学,1995.6
- 4 陈文伟等.“数据开采”技术专题(6篇论文).计算机世界报,1997.24
- 5 陈文伟等.“数据仓库与决策支持系统”技术专题(6篇论文).计算机世界报,1998.22
- 6 Fayyad U. Data mining and knowledge discovery in databases. Communications of the ACM, 1996, 39(11)
- 7 Inmon W H, Hackathorn R. Using the Data Warehouse, 1994

A New Architecture of Decision Support System

Chen Wenwei, Huang Jincal, Chen Yuan

System Engineering & Mathematics Department of National University of Defense Technology

Abstract The new decision support technologies, Data Warehouse, OLAP and Data Mining, developed in the middle of 1990s, open a new way to decision support system(DSS) by acquiring assistant decision information from databases. Based on the discuss of development progress, key technologies, research evolvments of DSS, the paper puts forward a new architecture of decision support system, which integrates Data Warehouse(DW), Data Mining(DM), Model Base(MB), Expert System(ES), On-Line Analytical Processing(OLAP).

Keywords: decision support system(DFSS), data warehouse(DW), data mining(DM), model base(MB), expert system(ES), on-line analytical processing(OLAP)