

文章编号:1004-4736(2008)01-0091-04

# 基于指标体系的决策支持模型研究

刘黎志<sup>1</sup>, 陈传波<sup>2</sup>

(1. 武汉工程大学计算机科学与工程学院, 湖北 武汉 430074;

2. 华中科技大学计算机科学与技术学院, 湖北 武汉 430074)

**摘要:**讨论了构成指标体系的关系结构,在说明指标、分组、阶码的基础上,重点描述了实用指标的形成过程,说明了指标体系多维数据集中的共享维度和事实数据表,最后对基于指标体系的决策支持模型作了分析.利用实用指标、时间、地区维度构成的多维数据集,可以方便的存储各种指标数据,并可以方便的进行时间序列数据抽取,从而为决策提供直观、有效的数据.由于指标是一种抽象的统计数据,所以基于指标体系的决策支持模型可以广泛的用于政府各职能部门及银行、电力等行业.

**关键词:**指标体系;实用指标;数据仓库;决策支持模型

**中图分类号:**TP 311

**文献标识码:**A

## 0 引言

决策的过程就是对多个可行方案进行选择的过程,决策的过程一般包括:对问题进行评价;收集和分析与问题相关的信息;根据信息确定可行方案;对各种可行方案导致的决策结果进行预测;对预测结果进行全面的评估并最终做出选择;对决策结果进行评估<sup>[1]</sup>.决策支持帮助决策者进行决策,但不代替决策过程.决策支持的主要任务是以历史数据为依据,对未来的数据进行预测,为决策者提供可行性方案.决策支持所需的数据一般包括地域信息、时间信息、数据自身的层次结构关系,人们通常将这类数据定义为各种指标.将构成指标的各种成分有机的结合在一起形成指标体系,并在指标体系的基础上形成实用指标,实用指标是数据的实际载体.数据仓库技术为基于海量实用指标的决策支持的数据分析、预测提供了基础.

## 1 指标体系

指标体系是构成实用指标维度的底层关系结构,指标体系包含构建它所需的基本关系实体,并在基本关系实体的基础上形成实用指标.

### 1.1 基本关系实体

**指标:**度量国家经济发展、生产效益、人口情况、自然资源等各方面的统计值.

**指标分类:**根据指标的统计领域确定其分类,一般分为工业、农业、国民经济、人口等.

**指标小类:**在某个指标类别中,为更好归纳不同统计目的指标而自定义的分类.

**指标度量单位:**指衡量指标值的单位,如米、美元、升等.

**指标度量单位分类:**按度量单位的用途而进行的分类,如长度、质量、面积等.

**阶码:**指同一度量单位的不同度量范围.对度量单位元,若阶码为3,则表示千元,阶码为9则表示亿元.

**分组:**指标反映的是某个统计值的综合信息,如“亏损企业数”指标,可能反映的是某个地区{特大型企业、大型企业、中型企业、小型企业}的亏损数的合计值,也可能反映的是某年{国有企业、私有企业、三资企业}的亏损数的合计值.具体确定指标统计对象的集合称为分组,{特大型企业、大型企业、中型企业、小型企业...}为企业规模分组,{国有企业、私有企业、三资企业...}为工商登记类型分组,可能的分组还有性别、学历、营业状况、本期值和末期值等.

**实用指标:**确定指标的具体统计对象及度量范围的统计值,实用指标是指标、分组、阶码的有效笛卡儿集.

### 1.2 实用指标的形成

**定义:**指标元组  $i = \langle \text{指标码}, \text{指标名称}, \text{指标父结点码}, \text{指标兄弟结点码}, \text{指标度量单位码}, \text{指标所属小类} \rangle$

**指标**  $I = \{i_1, i_2, i_3, \dots, i_j\}$

收稿日期:2006-10-24

基金项目:科学技术部科技型中小企业创新基金项目(03C26214201044).

作者简介:刘黎志(1973-),男,湖北武汉人,讲师,硕士,研究方向:基于网络的计算机应用、数据仓库及数据挖掘等.

分组元组  $g = \langle \text{分组码}, \text{分组名称}, \text{分组依据}, \text{分组父结点码} \rangle$

分组  $G = \{g_1, g_2, g_3 \dots g_k\}$

阶码元组  $n = \langle \text{度量单位码}, \text{阶码值}, \text{阶码名称} \rangle$

阶码  $N = \{n_1, n_2, n_3 \dots n_m\}$

实用指标元组  $pi = \langle \text{实用指标码}, \text{指标码}, \text{分组码}, \text{阶码} \rangle$

实用指标  $PI = \{pi \mid pi \in \langle \langle i_1, g_1, n_1 \rangle, \langle i_2, g_1, n_1 \rangle, \langle i_2, g_2, n_3 \rangle \dots \langle i_j, g_k, n_m \rangle \}$  其中  $\langle i_j, g_k, n_m \rangle$  为有效组合}

$j, k, m$  为自然数。

将实用指标定义为指标、分组、阶码的有效笛卡儿集提高了构造指标体系的灵活性,特别是分组的引入可以极大地减少系统中指标的个数。实用指标必须是指标,分组,阶码的有意义的组合,实用指标一般是在系统初始化时根据需要收集的具体指标统计值由系统管理员生成,报表设计工具可根据实用指标设计收集指标数据及汇总指标数据的表单格式。由于指标体系结构的灵活性,使得实用指标在使用中很方便系统管理员维护。

## 2 建立指标体系数据仓库

### 2.1 指标体系维度

时间维:时间维的设计根据系统的不同应用领域而定,对于应用指标体系的系统存在一个时间粒度控制问题。粒度是维划分的单位,体现着数据单元的详细程度和级别。数据越详细,粒度越小,级别越低。数据综合程度越高,粒度越大,级别越高<sup>[2]</sup>。时间维粒度的确定一般和报表的报送制度一致,若报表每月报送一次则时间维可为年、季、月,若报表每天报送一次则时间维可为年、季、月、天。时间维是典型的星型层次架构。

地区维:反映指标体系的地区信息,地区是具有父子层次架构的维度模型。指标体系的地区维度一般表现系统所应用区域的树型结构。将指标体系应用到信息数据统计平台,地区维表现为企业、地市、省、信息产业部;应用到湖北扶贫信息统计监测系统,地区维表现为村、乡镇、地市、省。数据仓库按地区的层次结构逐级汇总得出指标的聚合值,但其默认方式只是从叶结点开始逐级向上汇总。若应用系统需要在非叶结点输入实用指标值,例如扶贫统计中乡镇同样也要上报指标数据报表,则需设置地区维,允许其在非叶结点上接收数据。

实用指标维:指标、分组、阶码构成实用指标维的雪花型层次模型,以指标名称为第一级别、分

组名称为第二级别。实用指标维的雪花型层次模型如图 1 所示。

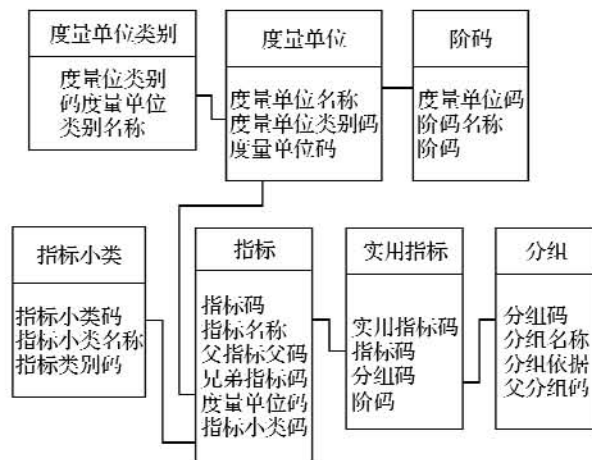


图 1 实用指标维雪花模型

Fig. 1 Snowflake model of practical indicator

指标类别维:确定指标所属的类别,可以设定为共享维度或虚拟维度。若设定为共享维度,则需指定其底层指标类别数据表。若设定为虚拟维度,可为实用指标维度指定指标类别成员属性。

### 2.2 指标体系多维数据集

数据仓库多维数据集有星型模型和雪花模型两种。星型模型的设计比较简单,是基于关系型数据库的面向 OLAP 的一种多维数据模型的数据组织形式。星型模型由事实表和多个维度表组成,通过使用一个包括主题的事实表和多个包含事实的非正规化描述的维度表来执行指标体系查询,由于数据仓库在存储事实数据表时会自动计算指标数据的聚合值,因此可获得比一般 SQL 语句分组查询更高的查询性能。星型模型的中心是指标数据,对应实用指标事实数据表,四周是访问的角度,对应维度表,每一个维度表通过一个关键字直接与事实表关联。事实表中每条记录都包含指向各个维度表的外键和实用指标度量值。雪花模型与星型模型的区别仅在于对维度表的描述是正规的,对实用指标维的描述就是正规的。指标体系雪花模型多维数据集如图 2 所示。

对事实数据表和维度表所作的修改在对其数据进行刷新后可以从底层关系表中反映到多维数据集中,对数据集的刷新往往有一定的逻辑规则及时间规则。DTS(数据转换服务)以工作流的方式实现数据集刷新的逻辑规则,数据仓库管理员可以规定多维数据集的刷新次序,并在数据集刷新完成后收到 DTS 发送的完成消息。数据集刷新的逻辑规则以 DTS 包的形式组织,数据仓库管理员设置 DTS 包的时间规则完成对 DTS 包的调度。

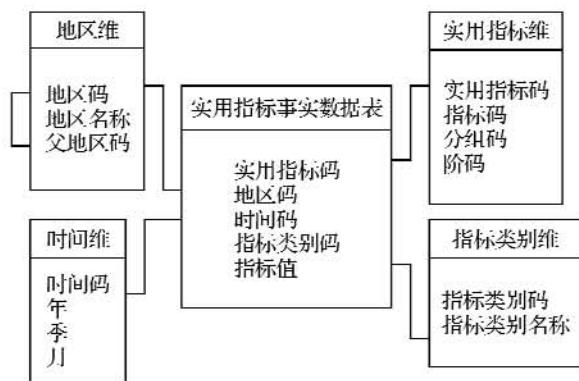


图 2 指标体系多维数据集

Fig. 2 Multi-dimension dataset of indicator system

### 2.3 阶码转换

由于实用指标的阶码可能与汇总报表中的阶码不一致,所以向多维数据集添加数据记录的指标值时,程序逻辑要根据实用指标的阶码对用户输入的指标值进行转换,公式为:

令:  $M$  = 实用指标阶码

存储到事实数据表的指标值 = 用户输入的指标值  $\times 10^M$

从多维数据集中查询的指标值在向用户展示时同样也要进行阶码转换,公式为:

展示给用户的指标值 = 存储到事实数据表的指标值  $\times 10^{-N}$   $N$  = 用户指定的阶码。

## 3 基于指标体系的决策支持模型

基于指标体系的决策支持过程为:客户端使用报表工具设计收集有针对性的指标报表;将报表通过 Web 服务器发布;地区维的叶结点、非叶结点在规定报告期内填报数据,从而完成数据收集;对系统历史数据进行查询、分析、比较、预测,为决策提供依据。基于指标体系的决策支持模型如图 3 所示。

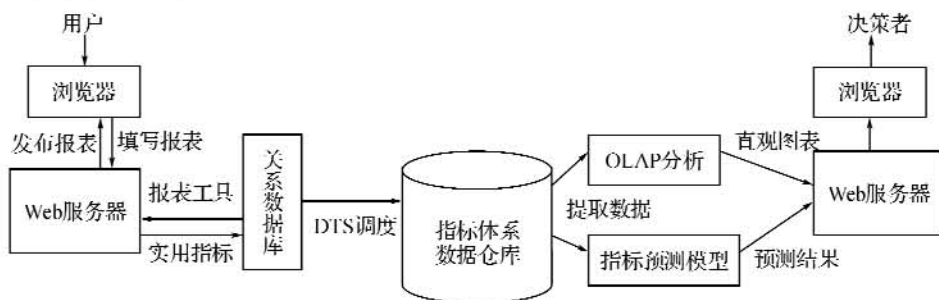


图 3 基于指标体系的决策支持模型

Fig. 3 Decision support model based on indicator system

### 3.1 OLAP 多维数据分析

基于数据仓库的指标多维数据集允许对数据进行切片和切块,可以很容易的得到某个地区、某个时间段、某个指标类别的某个指标的汇总和明细信息<sup>[3]</sup>。利用 MDX(多维表达式)来查询指标体系多维数据集时,地区、时间、指标类别维度一般作为切片器维度,而实用指标维度一般作为轴维度。由 MDX 查询返回的数据元组集合可以和基于 Web 的多维数据集浏览控件绑定,从而使得决策者可以根据维度进行查询条件筛选,可以深化以看到数据的细节,还可以浅化以看到汇总数据。通过 MDX 查询返回的二维数据集向决策者展示直观的指标数据关系,并为制作汇总指标报表及直方图提供数据源。由于多维数据集在存储时使用聚合预先计算汇总指标数据,所以可以极大地提高汇总指标查询的效率和响应时间,这对于决策过程来说,是非常重要的。利用钻取功能,又可以反向获得构成汇总指标的底层实用指标记录,从而得到汇总指标的详细成分。

### 3.2 实用指标预测

决策支持的主要任务就是以历史的实用指标

值作为依据,对未来的实用指标值进行预测。因为预测往往是针对某个特定地区的,因此可以将实用指标值看成是时间的函数。由一组在连续时间内的实用指标值组成的集合称为时间序列。

时间序列定义:

$X_t (t=1, 2, \dots, n)$   $t$  代表时间,  $X_t$  表示在第  $t$  个时间的实用指标值。

利用指标体系中的时间维作为 MDX(多维表达式)的切片器维度,可以方便地得到任何一个地区下任何一个指定的连续时间域的实用指标切片。

从经济学的角度出发,由实用指标组成的时间序列并不是一种单纯的数据序列,而是由若干种独立的基本成分按某种比重混合而成的复合体<sup>[4,5]</sup>。这些基本成分有性质不同的两类,一类是由事先无法预料的因素引起的随时间按随机规律波动的成分,称为随机成分;另一类是由某些系统因素引起的随时间呈系统变化的成分,称为系统成分。系统成分又可按成因和变化规律的不同分成若干独立的子成分,其中最重要的是长期趋势、

季节波动和周期波动<sup>[3]</sup>。基于实用指标值时间序列的预测方法可以分为定性预测法、定量预测法、半定量预测法。各类预测方法根据不同的应用情况各有优劣,决策支持系统可以根据成熟的各类预测方法建立实用指标预测模型,利用计算机自动完成预测,为决策者提供决策依据。

#### 4 结 语

指标体系的存储、管理、共享在政府决策支持、商业智能、经济发展分析与预测等各方面将发挥越来越重要的作用,基于以上领域的计算机应用系统都无一例外以大量的统计指标作为数据基础。在近几年的实践中,在指标体系的底层存储结构、实用指标的形成、指标体系多维数据集的构建、基于指标体系的决策支持等方面做了一些有益的探索。

#### 参考文献:

- [1] SolEuNet. Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise. EU co-founded project IST-1999-11495 [EB/OL]. <http://solcunct. ijs. si/website/html/euproject.html>, 2001.
- [2] 冉春玉,谷 川. 税务决策支持系统数据库的设计[J]. 计算机应用, 2005, (25):158-159.
- [3] Kantardzic M. Data Mining: Methods, Tools and Techniques[M]. USA:IEEE Press and John Wiley, 2002. 380.
- [4] 邓正宏,郑玉山,康慕宁. 基于辅助决策支持的虚拟经济性运行系统的设计[J]. 计算机工程, 2005, 31(17):180-181.
- [5] 李宝仁. 经济预测理论、方法及应用[M]. 北京:经济管理出版社,2005. 25-28.

## The research of decision support model based on indicator system

LIU Li-zhi<sup>1</sup>, CHEN Chuan-bo<sup>2</sup>

(1. School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430074, China;

2. College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

**Abstract:** In this paper, by explaining what are indicator, group and scale, the authors firstly discuss the relational structure of the indicator system and combining process of practical indicator, then we explain the shared dimension and fact data table within indicator system multi-dimension dataset. Finally the analysis of decision support model based on indicator system is given. Proved by practices, the multi dataset constructed by practical indicator, time and region can store kinds of indicator data and extract time serial data easily, so it can provide intuition and effective data for decision. Because indicator is an abstract statistical data, so the decision support model based on indicator system can be widely applied in government department and bank, power industry.

**Key words:** indicator system; practical indicator; data warehouse; decision support model

本文编辑:陈晓苹