属性维概念及其操作的研究*)

袁　霖　李战怀
(西北工业大学计算机系1108教研室　西安710072)

The Concept of Attribute Dimension and Corresponding Operations

YUAN Lin　LI Zhan-Huai
(1108 Staff Room, Department of Computer, Northwest Polytechnical University, Xi'an 710072)

Abstract Dimension member attribute is used to describe the property of dimension members. It is not fully understood and well defined in OLAP research area. This paper focuses on a special kind of dimension member attributes, which can be used as dimensions by themselves. We call them attribute dimensions. In order to facilitate this kind of necessity of multidimensional data modeling in many real-world applications, the traditional multidimensional structure is extended and a group of operations are given to formulate corresponding multidimensional queries.

Keywords Dimension member attribute, Virtual dimension, Multidimensional data model, OLAP

1. 引言

数据仓库技术研究如何将分散在企业内部多个异构数据源的操作型数据，经过抽取、清洗转换、集成，以一致的方式进行存储；OLAP(联机分析处理)技术则为决策人员提供这些数据的多维视图及相应的多维操作，而相应的OLAP工具使用多维数据模型对数据进行表示、分析、显示。

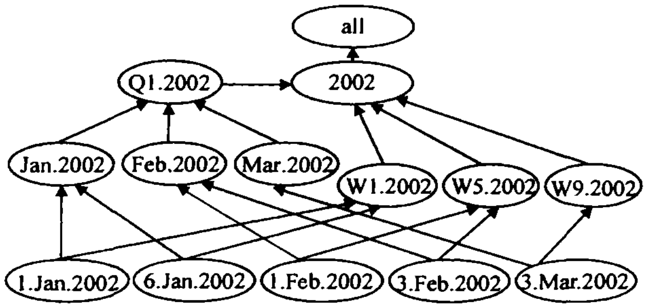

图1　时间维的实例

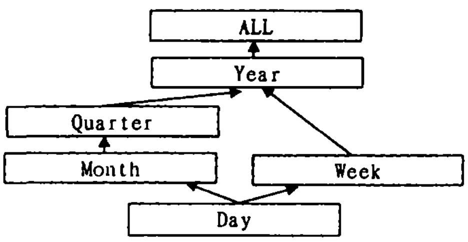

图2　时间维的多层次模式

在多维数据模型中，数据被分为两类——作为被分析对象的度量和用于描述度量的若干个维。度量被视为多维空间中的点。维自身具有结构，通过模式和实例来描述。维模式是由若干个粒度等级(Level)所构成的有向无环图，每一条从最低等级到最高等级的路径称为一个层次(hierarchy)，如图2所示的时间维模式中包含了两个层次。维实例是由每一个等级中所包含的维成员的集合以及属于不同等级的维成员之间的映射组成，这种映射是满射的被称为rollup函数，如图1所示。每一个低等级中的维成员属于唯一的一个高等级中的维成员，这种标准的维层次结构是严格的(strict)和划分的(partitioning)[1]。

2. 成员属性及属性维的概念

与层次、等级等概念相比，"成员属性"这一概念并没有被学术界和工业界很好地认识和定义，大多数厂商将这一概念定义为"对于维成员特性的描述，它的存在依赖于维的某一具体的等级"。例如，可用图3-5分别表示某连锁商店有关销售数据的三个维模式，图中用--○或—●表示成员属性；在图5中，Profession、Gender、IncomeLevel、Address、Married等成员属性描述了Customer等级中每一位顾客(维成员)的特征。

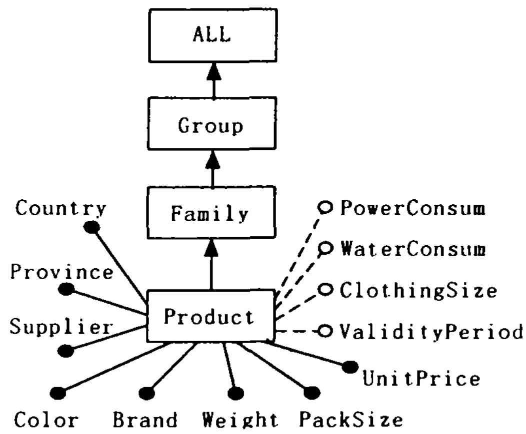

图3　产品维模式

现有的OLAP产品在成员属性的定义和使用方式上存在很大的差异，通过对现有OLAP产品和相关研究的调查，我们将成员属性分为以下四类：

*)本文研究得到国家自然科学基金、西北工业大学博士论文创新基金资助。袁霖　博士研究生，主要研究领域为对象关系数据库，数据仓库与OLAP技术。李战怀　博士，教授，主要研究领域为数据库理论与技术。

①维成员的标识属性：为每个维成员提供唯一的标识。例如，对于商品维的 product 等级，productID(条形码)和 ProductName 可作为产品成员的标识属性；

②维成员的非标识属性：这类属性可以为维成员提供更多的信息，用于数据的显示。如产品维中 Product 等级的产品的单价 UnitPrice 为平凡成员属性，通过在产品轴上添加成员属性来扩展成员信息。

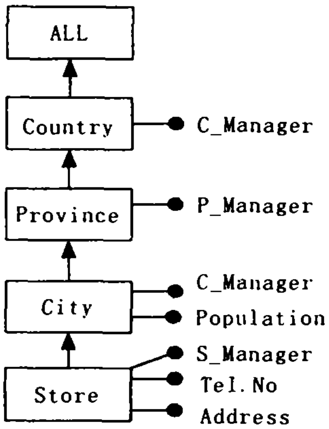

图4　商店维模式

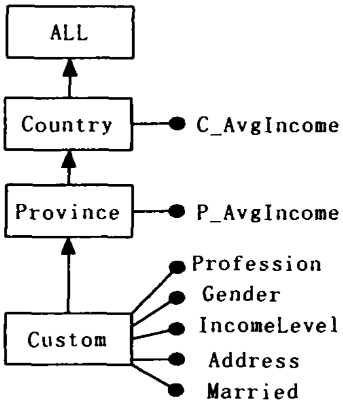

图5　客户维模式

表1　产品的年销售量

		1997年	1998年	1999年	2000年
Product	UnitPrice	销售量	销售量	销售量	销售量
调扇	15	4935	5498	6290	7724
75W 电灯	254	103443	128956	138405	148240
电视	1538	4544960	7329650	8432065	9534765
鞋	80	11034	13296	15969	17934

表2　不同产品大类(Group)不同颜色的产品销售量

	Blue	Brown	Green	Red	Yellow
Group	销售量	销售量	销售量	销售量	销售量
电器	4738	3726	5345	5249	4890
食品	964	846	1298	1345	1895
服装	3476	1011	4823	5692	3851
玩具	300	240	600	893	459

③强成员属性：是一类特殊的"非标识性成员属性"，相应等级中的每一个成员在属性上都有一个合法的取值，且维成员与属性值之间是多对一的关系。这一类成员属性在多维分析中可以作为与所在维不同的另一个独立的坐标轴，用来表示聚集数据的粒度。如图3所示，Group 是产品维的一个等级，而 color 是一成员属性(图中用一●表示强成员属性)，它可以作为另一个维而出现。

④弱成员属性：与强成员属性不同的是弱成员属性对相应等级中的某些成员不存在合法的取值，在图3所示的产品维中，弱成员属性(图中用--○表示)PowerConsum(耗电量)只能用来描述电器产品，而对于"服装"、"食品"是没有意义的。在适当的限定条件下，弱成员属性可以作为一个维来使用。

表3　电器类中不同产品对于不同用电量的销售量

		0-1000W	1001-2500W	2501-4000W	>4000W
Group＝电器	family	销售量	销售量	销售量	销售量
	空调	2456	6930	7290	5930
	冰箱	4927	3452	1284	980
	炊具	3859	4627	2748	1203

在下文中，我们用"成员属性"这一词专指"非标识性成员属性"，其中强成员属性可以被看作弱成员属性的一个特例，作为一个独立的维来使用是二者的共性，我们将这种特殊的基于成员属性的维统称为属性维，如何表示属性维及其相关操作是本文的研究重点。

3. 相关产品及研究中对于属性维概念的支持

表4　OLAP 产品的功能对比

功能 产品	成员属性				基于成员属性的维	
	所用术语	1	2	3	术语	4
Pilot 5	属性	√	√	×	属性维	√
微软 SQL Server2000	成员属性	√	√	×	虚拟维	√
Essbase 6.0	属性	√	√	×	属性维	√
Oracle9i OLAP	属性	√	√	×	×	×
Informix Metacube	属性	√	√	×	*	×

表中注释：

1. 是否为查询结果提供附加信息？
2. 是否可用于选择操作？
3. 是否区分强/弱成员属性？
4. 是否可具有多层次结构？

* 允许聚集操作，但不允许上钻、下钻操作

对成员属性这一重要概念，OLAP 委员会[2]并未给出与之相似的概念。不同的厂商在这一概念的名称、可适用的操作，以及基于这一概念的维的名称、结构等诸多方面都存在很大差异，如表4所示，其中没有一个产品在概念上区分"强/弱成员属性"。

与工业界对这一概念的关注相比，学术界的相关研究很少。我们对现有的多维数据模型[3~10]进行了调查，其中只有文[3,4]试图表达这一概念，其不足之处将分别在3.1、4.3节予以论述。在下文中，我们将不支持成员属性，且不支持基于成员属性的维的所有多维数据模型统称为"经典模型"。

4. 经典模型的局限性

在经典模型中，成员属性只有两种可能的表示方式：①将成员属性作为它所在维的一个等级；②将它直接定义成一个独立的维。但从模型和实现的角度上看，这两种表示方式均存在缺陷。

4.1　将成员属性表示为等级

如图6所示，我们将产品维中的 brand 成员属性分别以两

种不同的方式表示为一个 level，但是对一个多维数据立方来说，一个维中最多只能允许其中的一个 level 参与数据立方粒度的表示，因此不可能得到 Group 和 Color 分别作为两个不同的坐标轴的查询结果。

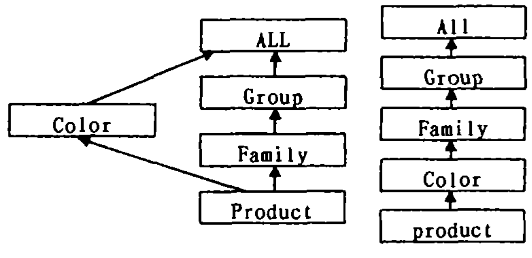

图6　产品维(将 Color 表示为 level)

假如我们允许一个维中可以有多个 level 同时参与粒度的表示，相当于一个维中包含了多个“子维”，文[4]提出的多维数据模型就是这种结构。但是，这种维的“多子维结构”与维的“多层次结构”又将在概念上产生混乱，例如图2的时间维包含了两个层次，如果将 Month 和 Week 分别作为两个不同的坐标轴所表示的数据立方是没有意义的。

因此，如果将成员属性表示为一个 level，则丧失了它可以作为另一个独立维的能力。

4.2　将成员属性表示为一个独立的维

如果将成员属性直接表示为经典模型中的一个维，将会导致：(1)数据立方的维数将急剧增加，仅对于图3所示的产品维就有8个强成员属性，他们都要成为独立的维(而弱成员属性根本不能成为维)。当人们谈到一个数据立方具有几十甚至上百个维时，往往是这样造成的。(2)由于这些维之间存在函数依赖关系，例如，图3所示产品维中的 product 等级与新增加的 Brand，color，weight，packsize 维之间存在多对一的函数依赖关系，无论对于 ROLAP(关系型的 OLAP)还是 MOLAP(多维型的 OLAP)，都会导致实现上的问题。

对于 MOLAP 实现，产品维中的每一种产品，在 Brand 中只有一个值，对 Brand 的其它可能的取值只能为空。维数越多，数据立方的稀疏性越大，将会给数据的装载和查询计算带来困难。

对于 ROLAP 实现，每一个独立的维在事实表中都有一个相应的外码，而这些外码之间存在函数依赖关系，由此导致事实表中存在大量的冗余，这不仅影响数据立方装载和查询，而且会给动态模式修改带来更大的困难。如果需要向数据立方增加一个原本可作为成员属性的维，则需要对事实表中的全部元组进行更改，这简直就是一场灾难。

综上所述，从实现的角度来看，不能将成员属性“硬”建模为一个独立的维。

5. 经典模型的扩展思路

由于经典模型不能合理表达成员属性和属性维，故须将其扩展。

5.1　属性维与成员属性之间的映射

成员属性与维是两个不同的概念，如果要将成员属性作为维来使用，必须要有一个中间层，将成员属性翻译成属性维，相当于视图。经典模型不能合理表达属性维的根本原因就在于缺少这样一个中间层。

5.2　数据立方中不同坐标轴之间的独立性

对于一个 n 维数据立方，每一个坐标轴的刻度是由维中的每个 level 决定的，这些来自不同维中的 level 之间不能具有一对多的函数依赖关系，只能是多对多的关系。而属性维与真实维之间并不总能满足这一约束条件，如图5中基于 gender 的属性维与客户维中的 customer 等级之间是一对多的关系，因此不能同时参与数据立方粒度的表示。这一约束条件必将影响多维数据立方的表示和操作。

5.3　基于成员属性的维的动态定义

属性维除了可以静态声明外(称为虚拟维)，还应该允许用户在查询过程中动态定义(称为动态维)，因为：(1)静态定义只能针对强成员属性，而弱成员属性只能在满足相应约束条件下动态定义；(2)动态定义可以使分析人员把注意力集中在他所需要的信息上。

5.4　与相关研究及 OLAP 产品的比较

Lehner 曾提出了一个扩展的多维数据模型[3]，引入了“维属性”的概念，用来专指一个维中最低等级中的维成员属性，缺乏普遍性。通过两个操作 split 和 merge，动态地将维属性作为维使用，但一个不合理的限制是：在每个查询中，每个维至多只能有一个维属性出现。

我们在表4中对 Microsoft SQL Serve2000的 Analysis 进行了调查，Analysis 对于强成员属性和弱成员属性不作区分，当某些成员在弱成员属性上没有合法取值时，用 NULL 表示，但 NULL 也同时用来表示一个值未知，因此造成歧义；基于“成员属性”的维，只能静态声明，不能在查询中动态定义；不满足5.2节中的约束条件，即数据立方的坐标轴之间可以具有函数依赖关系，从而导致稀疏性。

我们的扩展模型克服了上述局限，对于每一个成员属性都有一个约束与之对应；不仅可以对基于强成员属性的虚拟维进行静态声明，而且可以在查询中定义动态维(可基于弱成员属性)。下面给出这一扩展模型的形式化定义。

6. 一个扩展的多维数据模型

6.1　维模式和维实例

设 $LEVELS$ 是 level(等级)名称的有限集合，对于每一个 $L \in LEVELS$，都有一个维成员的集合与之对应，称为 L 的域，用 $dom(L)$表示，且 ALL$\in LEVELS$，dom(ALL)={all}。

定义6.1.1(常规维模式)　是一个六元组 $DS=(Dname, Lset, \leqslant_L, Aset, AL, P_{Aset}$，其中 $Dname$ 是维名，$Lset \subseteq LEVELS$ 是 level 的有限集合，$\leqslant_L$ 是 Lset 上的偏序关系，具有唯一的最底层 L_{inf}和唯一的最高层 ALL。对每一个 $L \in Lset$，都有 $L_{inf} \leqslant_L L \leqslant_L ALL$。Aset 是成员属性的有限集合。$AL$：Aset→Lset 是一函数。$P_{Aset}$是作用在 $L_p \in Lest$ 上的，满足 $AL(A_i) \leqslant_L L_p$，形为 $L_p=c$(c 为常数值)的谓词的集合，P_{Aset}与 $Aset$ 之间是一一映射关系。

定义6.1.2(常规维实例)　常规维 D 的实例是一个三元组 D=(DS,Rup,la)，其中 $DS=Schema(D)$是维 D 的模式，我们用 $D.L_i$ 表示 $DS=Schema(D) \wedge L_i \in D$；Rup 是一个满射函数的有限集合，设$<_L$ 为 Lset 上的拟序关系，对于 $L_1, L_2 \in DS \wedge L_1 <_L L_2$，存在一个满射函数 $Rup_{L_1}^{L_2}: dom(L_1) \rightarrow dom(L_2)$。如果存在 $L_1 <_L L_2 <_L L_3 <_L L <_L L_{i-1} <_L L_i$，则 $Rup_{L_1}^{L_i} = Rup_{L_{i-1}}^{L_i}$ oL o$Rup_{L_2}^{L_3}$ o $Rup_{L_1}^{L_2}$。la 是一个部分函数的集合，对于每一个 $AL(A_j)=L_i$，简写为 $L_i.A_j$，设 $e \in dom(L_i)$，$P_{A_j} \in P_{Aset}$是一以 e 为变元的谓词，if $P_{A_j}(e)=true$ THEN $la_{L_i}^{A_j}(e)=$ v，$v \in dom(A_j)$ ELSE $la_{L_i}^{A_j}(e)$无定义。如果 $P_{A_j} \Leftrightarrow true$，则 A_j

称为强成员属性；否则为弱成员属性。

例1 对于图3所示的产品维，定义在 Product 等级上的强成员属性有：Color、Brand、Weight、PackSize、Supplier、Province、Country，相应谓词是一永真式，例如：$P_{color} \Leftrightarrow 1$，而对于弱成员属性 PowerConsum，与之相应的谓词 $P_{owerConsum} \Leftrightarrow Group =$“电器”，它的解释为：对于任意一个属于 Product 等级的维成员，如果满足 $e \in dom(Product) \wedge Rup_{product}^{Group}(e) =$“电器”，则必存在一个“用电量”的值，否则，没有定义。

定义6.1.3（虚拟维模式） 是一个五元组 $VDS = (Dname, Lset, \leqslant_L, D_R.L_R, LR)$，LR 是一单射函数，它将 Lset 中的每一个等级 L_i 映射到常规维的 $D_R.L_R.A_R$ 上，$LR: Lset \rightarrow Aset(D_R.L_R)$，$Aset(D_R.L_R)$ 为定义在 $D_R.L_R$ 上的所有成员属性的集合。

定义6.1.4（虚拟维实例） 一个三元组 VD=(VDS, Rup, Proj)，其中 VDS 是维 VD 的模式；对于每一个 $LR(L_i) = D_R.L_R.A_R$，维 D_R 的 L_R 等级中的每一个维成员在属性 A_R 上的值都成为 VD 在 L_i 中的一个维成员，Rup 的定义同定义5.1.2，是虚拟维必须满足的约束条件。Proj 是一函数映射 $Proj_{D_R.L_R}^{VD.L_{inf}}: dom(D_R.L_R) \rightarrow dom(VD.L_{inf})$。

例2 对于图3所示产品维，我们可以基于它的 Product 等级上的强成员属性 Supplier、Province、Country 定义一个名为“*SupplierDim*”的虚拟维，有四个等级：$Vsupplier < Vprovince < Vcountry < ALL$。

定义6.1.5（维成员操作符） 两个维操作符 descendants、elements 定义如下：

如果 $L_i, L_j \in Lset(DS) \wedge L_i <_L L_j, e_i \in dom(L_i)$，则 $descendants(e_j, L_i) = \{e_i \mid Rup_{L_i}^{L_j}(e_i) = e_j \wedge e_j \in dom(L_j)\}$。对于某常规维的 $L.A$，设 $a_j \in dom(A)$，$elements(a_j) = \{e_i \mid a_j = al_L^A(e_i)\}$。

6.2 度量及多维数据集

定义6.2.1（度量模式） 是一个3元组 $MS = (Mname, MT, O)$，其中 $Fname$ 是度量的名称，MT=(idT, MVT)是度量单元(cell)的类型，idT 是标识类型，MVT 是度量值的类型。O 是可用于 MVT 上的聚集函数的集合。

定义6.2.2（度量域） MS 是度量模式，每一个度量单元是一个以 MT 为型的对象，存在双射函数 $rep: dom(MT) \leftrightarrow dom(idT)$，度量域 $dom(M) \subseteq dom(MT)$。

定义6.2.3（多维数据集模式） 是一个三元组 MDS=(DSset, G, MS)，MS 是度量模式，DSset 是维模式的集合 $DSset = \{DS_i \mid 1 \leqslant i \leqslant n\}$，$G = \{L_i \mid 1 \leqslant i \leqslant n\}$ 用来表示多维数据集的粒度，其中 $L_i \in level(DS_i)$。

定义6.2.4（多维数据集） 是一个五元组 $MD = (MDname, MDS, Dset, M, ML)$，MDS 是多维数据集的模式，M 是以 MS 为模式的度量，Dset 是以 DSset 为模式的维的集合；对于 n 维多维数据集，$ML = \{ml_i: dom(M) \rightarrow dom(L_i) \mid L_i \in G, 1 \leqslant i \leqslant n\}$。

6.3 多维操作

假设，被操作的 n 维数据集为：$MD = (MDname, MDS, Dset, M, ML)$，$MDS = (DSset, G, MS)$，$Dset = \{D_1, L, D_i, L, D_n\}$，$G = \langle L_1, L_2, L, L_i, L, L_n \rangle$，$dom(G) = dom(L_1) \times dom(L_2) \times L \times dom(L_i) \times L \times dom(L_n)$。结果多维数据集为

$MD' = (MDname', MDS', Dset', M', ML')$，$MDS' = (DSset', G', MS')$

定义6.3.1（限定操作） $\sigma[P](MD) = MD'$

P 是作用在 MD 上的限定谓词，用文法产生式表示为：$P \rightarrow P_M \mid P_{D_i.L_j} \mid P_{D_i.L_j.A_K}, P \rightarrow P \wedge P$。$P_M$、$P_{D_i.L_j}$、$P_{D_i.L_j.A_k}$ 是分别作用在度量、维等级、成员属性上的谓词。针对度量、维等级的限定操作的语义与经典模型没有差别。为了将限定操作扩展到成员属性，我们有必要先给出针对维等级的限定操作的定义，而针对度量的限定操作的定义从略。

(1) $\sigma[P_{D_i.L_j}](MD) = MD'$

$D_i.L_i \in G$ 是 MD 的粒度所在的等级，查询应满足 $L_i \leqslant_L L_j$。

设 $dom(L_i \mid P_{D_i.L_j}) = U_{e_j \in dom(L_j) \wedge P_{D_i.L_j}(e_j)} descendant(e_j, L_i)$，

$dom(M') = \{m \in dom(M) \mid \exists\ e_i \in dom(l_i)(ml_i(m) = e_i \wedge dom(L_i \mid P_{D_i.L_j}))\}$

$ML' = \{(m', (e_1, e_2, L, e_n)) \in ML \mid m' \in dom(M')\}$，$MDS' = MDS$

(2) $\sigma[P_{D_i.L_j.A_K}](MD) = MD'$

$dom(L_j \mid P_{D_i.L_j.A_K}) = U_{a_j \in dom(A_j) \wedge P_{D_i.L_j.A_K}(a_j)} elements(a_j)$，令 $P_{D_i.L_j} \Leftrightarrow e \in dom(L_j \mid P_{D_i.L_j.A_K}) \wedge L_j = e$，$\sigma[P_{D_i.L_j.A_K}](MD) = \sigma[P_{D_i.L_j}](MD) = MD'$。用谓词变换的方式将作用在成员属性上的谓词变换为作用在等级上的谓词。

例3 “限定操作”是广义的“切片”、“切块”操作，通过扩展，可以在成员属性上施加限定谓词，例如：$\sigma[ProductDim.Brand =$“Haier”$\wedge ProductDim.Family =$“电器”$]MD$。

定义6.3.2（上翻操作） $Rollup[G', Agg](MD) = MD'$

$Agg \in o$，是作用在 M 上的聚集函数。查询应满足条件：满足 $L_i \leqslant_L L'_i, i = 1, 2, L, n$，且至少存在一个 $j \leqslant n$，使 $L_j < L'_j$ 成立。假设：$e \in \{\langle e_1, L, e_i, L, e_n \rangle \mid ML(m) \wedge m \in dom(M)\}$，为了便于表达，我们定义辅助函数 $RUP_G^{G'}$：$RUP_G^{G'}(e) = e' = \langle e'_1, L, e'_i, L, e'_n \rangle \in dom(G')$ if $L_i = L'_i$, then $e'_1 = e_1$ else $e'_i = Rup_{L_i}^{L'_i}(e_i)$，则查询结果为：

$ML' = \{(m', e') \mid \exists\ e' \in G'(m' = Agg(\{m \mid m \in M \wedge RUP_G^{G'} o ML(m) = e'\}))\}$

$dom(M') = \{m' \mid (m', e') \in ML'\}$。

例4 “上翻操作”的语义与“经典模型”类似，例如：输入多维数据集 MD 的模式为：

$MDS = \{DSset = \{ProductDim, StoreDim, CustomDim, TimeDim\}$，

$G = \langle Product, Store, Custome, Day \rangle$，$MS = (Sales, (idT, number))\}$.

$Rollup[\langle Family, City, Custom, Day \rangle, Sum]MD = MD'$。

$MDS' = \{DSset' = DSset, G = \langle Family, City, Custome, Day \rangle, MS' = MS\}$。

定义6.3.3（增加虚拟维） $AddVirtualDim[L_i \rightarrow L'_i, \{VD_1, L, VD_K\}](MD) = MD'$ 向多维数据集中加入 k 个在物理上依赖于某一真实维 D_i 的虚拟维 VD_1, L, VD_K，同时，维 D_i 的粒度等级从 L_i 上升到 L'_i。假设 VD_1, L, VD_K 所基于的成员属性在 D_i 中所属的等级分别为 $D_i.L_1, D_i.L_2, L, D_i.L_k$，则操作应满足：$\forall\ L_j \in \{L_1, L_2, L, L_k\}(L_i \leqslant_L L_j <_L L'_i)$。$Dset' = Dset \cup \{VD_1, VD_2, L, VD_K\}$，$G' = \langle L_1, L_2, L, L'_i, L, L_n, L_{n+1}, L, L_{n+k} \rangle$，其中 $L_{n+1}, L_{n+2}, L, L_{n+k}$ 分别是 VD_1, L, VD_K 的最低一级的 level。

为了易于表示，我们先定义三个辅助函数 EXT、$RUP^{G_{ext}}_{G'_{ext}}$、$RUP^{G}_{G_{ext}}$：

$e\in\{\langle e_1,L,e_i,L,e_n\rangle\mid ML(m)\wedge m\in M\}$，对 e 进行扩展增加 k 个 e_i：

$e_{ext}=\langle e_1,L,e_i,L,e_n,e_i,L,e_i\rangle$，$|e_{ext}|=n+k$，定义双射函数 $EXT(e)=e_{ext}$。

以同样的方式对 G 进行扩展：$G_{ext}=\langle L_1,L,L_i,L,L_n,L_i,L,L_i\rangle$ 是 $n+k$ 元向量。

令 $G'_{ext}=\langle L_1,L,L'_i,L,L_n,L_{n+1},L,L_{n+k}\rangle=\langle L_1,L,L'_i,L,L_n,L_{RJ},L,L_{RK}\rangle$

定义函数 $RUP^{G'_{ext}}_{G_{ext}}(e_{ext})=e'_{ext}=\langle e'_1,L,e'_i,L,e'_n,e'_{n+1},L,e'_{n+k}\rangle\in dom(G'_{ext})$

$if\ L_i=L'_i,e'_1=e_1$ else $e'_i=Rup^{L'_i}_{L_i}(e_i)$

定义函数 $RUP^{G'}_{G'_{ext}}(e'_{ext})=e'=\langle e'_1,L,e'_i,L,e'_n,proj(e'_{n+1}),L,proj(e'_{n+k})\rangle\in dom(G')$

$ML'=\{(m',e')\mid\exists e'\in G'(m'=Agg(\{m\mid m\in M\wedge RUP^{G'}_{G'_{ext}}\circ RUP^{G'_{ext}}_{G_{ext}}\circ EXT\circ ML(m)=e'\}))\}$。

例5 我们将例2中定义的虚拟维 SupplierDim 加入多维数据集 MD，其模式 MDS 同例4，$AddVirtualDin[Product\rightarrow Family,\{SupplierDim\}](MD)=MD'$，$MDS'=\{Dset'=\{ProductDim,StoreDim,CustomDim,TimeDim,SupplierDim\},G'=\langle Family,Store,Custome,Day,Vsupplier\rangle,MS'=(Sales',(idT,number))\}$。

定义6.3.4(增加动态维) $AddDynaDim[L_i\rightarrow L'_i,\{A_1,L,A_k\}](MD)=MD'$

这一操作与增加虚拟维类似，不同的是：1.这些维不是静态声明的，而是在查询中动态定义的，每个动态维只有两个等级 $L_j<ALL$，$dom(L_j)=dom(A_j)$；2.对于每一个 $A_j\in\{A_1,L,A_k\}$，P_{A_j} 是使 A_j 有定义的谓词，在这些维被加入多维数据集 MD 之前，先要对 MD 施加限定操作，限定谓词为：$P=P_{A_1}\wedge P_{A_2}\wedge L\wedge P_{A_j}\wedge L\wedge P_{A_k}$。

例6 我们将例5中输出多维数据集 MD' 作为本例中的输入，$AddVirtualDim[Custom\rightarrow Province,\{Gender,Married\}](MD')=MD''$，$MDS''=\{Dset''=\{ProductDim,StoreDim,CustomDim,TimeDim,Supplier\ Dim,GenderDim,MarriedDim\},G''=(Product,Store,Province,Day,Supplier,Gender,Married),MS''=(Sales'',(idT,number))\}$。

在我们的操作符集合中没有包含下钻操作，因为下钻操作总是要访问更细粒度的数据，最终仍要转换为上钻操作。同样我们也没有定义删除维的操作，在一个多维数据集中，一个维的粒度等级位于 ALL 时，就相当于维的隐含删除。

结论 在 OLAP 的研究领域，“成员属性”是一个长久以来没有得到广泛重视，更没有得到很好认识和定义的概念。本文的研究重点是一类特殊的成员属性(强/弱成员属性)，它们在某些特定的条件下可以作为一个独立的维参与多维数据的表示，这种需求广泛存在于大量的 OLAP 的实际应用之中。为了给这类应用提供一个坚实的理论基础，本文提出了一个扩展的多维数据模型，用于合理表示“成员属性”及基于成员属性的虚拟维和动态维，并对相应的多维操作进行了扩展。与现有的 OLAP 产品相比，成员属性的完整性约束，以及对于动态维的支持是这一扩展模型的独特之处。

参 考 文 献

1 Hurtado C, Mendelzon A. Reasoning about summarizability in heterogeneous multidimensional schemas. In: Proc. of the 8th Intl. Conf. on Database Theory, London, UK, 2001

2 http://www.olapcoucil.org

3 Lehner W. Modeling Large Scale OLAP Scenarios. In: 6th Intl. Conf. on Extending Database Technology (EDBT'98, Valencia, Spain, 23-27 March)

4 Cabibbo L, Torlone R. A Logical Approach to Multidimensional Databases. In: Proc. of the EDBT 1998

5 Nguyen T B, TJOA A M, Wagner R R. An Object Oriented Multidimensional Data Model for OLAP. In: Proc. of the First Intl. Conf. on Web-Age Information Management (WAIM'00), Shanghai, China, June 2000. Lecture Notes in Computer Science (LNCS), Springer, 2000

6 Agrawal R, Gupta A, Sarawagi S. Modelling Mul-tidimensional Databases. In: Proc. of the ICDE 1997

7 Li C, Wang X S. A data model for supporting on-line analytical processing. In: Proc. Conf. on Infor-mation and Knowledge Management, Nov. 1996

8 Gyssens M, Lakshmanan L V S. A Foundation for Multi-Dimensional Databases. In: Proc. of the VLDB 1997, Athens, Greece

9 Vassiliadis P. Modeling multidimensional databases, cubes and cube operations. In: Proc. of 10 th SSDBM 1998, Capri

10 Pedersen T B, Jensen C S. Multidimensional Data Modeling for Complex Data. In: Proc. ICDE' 99

(上接第49页)

参 考 文 献

1 intel iSCSI protect. https://sourceforge.net/projects/intel-iscsi, 2002

2 Satran J, et al. iSCSI. http://www.ietf.org/internet-drafts/draft-ietf-ips-iscsi-11.pdf, 2002

3 Rajagopal M, et al. Fibre Channel Over TCP/IP (FCIP). http://www.ietf.org/internet-drafts/draft-ietf-ips-fcovertcpip-09.pdf, 2002

4 Monia C, et al. iFCP-A Protocol for Internet Fibre Channel Storage Networking. http://www.ietf.org/internet-drafts/draft-ietf-ips-ifcp-10.pdf, 2002

5 Fibre Channel Industry Association. Fibre Channel Standard. http://www.fibrechannel.com, 2002

6 Weber R, et al. FC Frame Encapsulation. http://www.ietf.org/internet-drafts/draft-ietf-ips-fcencapsulation-06.pdf, 2002

7 Tseng J, et al. Internet Storage Name Service (iSNS). http://www.ietf.org/internet-drafts/draft-ietf-ips-isns-09.txt, 2002