

一种评估数据仓库设计质量的方法

刘益江,毛宁,陈庆新

(广东工业大学 广东省计算机集成制造重点实验室,广东 广州 510006)

摘要:数据仓库(Data Warehouse, DW)是支持决策管理过程的、面向主题的、集成的、随时间而变的、持久的数据集合,它集成了源数据库(Database, DB)和其他平面文件系统的相关数据,以支持决策管理活动。数据仓库结构是一个多维模型,主要分星形、雪花星和星座形三种。有别于传统关系型数据库的设计,数据库仓库的设计往往以数据作为驱动,其设计的好坏,直接影响了数据仓库系统的建设与应用。文中提出了一种评估数据仓库设计质量的定量分析方法,主要是量化了和数据仓库有血缘关系的数据源的相关质量指标,就是通过分析所选取的表和属性两方面的数据质量(Data Quality, DQ)指标,最终联合这些指标以计算数据仓库设计质量的评估值。其中,分析数据源相关质量指标的过程,也可以支持数据库仓库的设计。

关键词:数据仓库;数据质量;数据仓库设计质量

中图分类号:TP392;TP302.7

文献标识码:A

文章编号:1673-629X(2012)09-0161-05

A Methodology for Data Warehouse Design Quality Assessment

LIU Yi-jiang, MAO Ning, CHEN Qing-xin

(Guangdong Provincial Key Laboratory of Computer Integrated Manufacturing System,

Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Data warehouse which embraces operation database and other relative data from flat files is subject-oriented, integrative, updated and long-lasting and can support decision making. Data warehouse is a multi-dimension model. There are star, snowflake and constellation, all together three kinds of data warehouse. It is different from traditional relation database, the design of data warehouse is data-driven. And the design can directly affect the establishment and application of the data warehouse. It proposes a methodology of quantitative analysis to evaluate the design quality of data warehouse. A way is found to quantify the quality index of the data source which is relative to the data warehouse. The value of the data warehouse is assessed by analyzing the data quality from selection tables and attribute. Through analyzing data quality of data source, the quality design of data warehouse can be evaluated.

Key words: data warehouse; data quality; quality of data warehouse design

0 引言

随着各行各业现有信息系统数据量的增长,越来越多的企业、组织热衷于将现有的历史数据集成为数据仓库,以作分析决策之用,或者挖掘其中有用的信息。良好的数据质量是建设数据仓库系统成功与否的关键,没有过硬质量的数据支持的数据仓库系统,只能对决策分析者起到误导的作用,这会产生不可估量的后果。在选取数据这个问题上,Chengalur-Smith认为从操作型系统中选择质量好的数据能很好地达到决策的需求^[1]。在此之前, Daniel Aebi, Louis Perrochon给

数据质量作了定义,数据质量主要指一个信息系统在多大程度上实现了模式和数据实例的一致性,及模式和数据实例在多大程度上实现了正确性(correctness)、一致性(consistency)、完整性(completeness)和最小性(minimality)^[2]。对于数据质量的评估,大多从语义上分析,很少以定量的方式去分析。当然,也有一些学者通过统计的方法去评估数据质量,文献[3]提出通过最终的用户去评估数据质量,主要是通过问卷调查或者访问的方式去评估。而文献[4]采用了众所周知的常用统计方法来获取数据质量的信息,其中属性的统计分布就是重点的一个部分。

回到之前所谈,数据仓库需要高质量数据的支持,因此有人提出在设计数据仓库的时候需要考虑其质量因素。文献[5]提出了一种支持数据仓库物理设计的框架,这种结构考虑了支持设计的一些质量因素。文献[6]提出了一种方法用以检查所选择的数据视图是

收稿日期:2012-01-31;修回日期:2012-05-06

基金项目:国家自然科学基金(51175094);省部产学研结合项目(2010A090200054)

作者简介:刘益江(1986-),男,广西人,硕士研究生,主要研究方向为信息化制造;毛宁,教授,研究方向为工业工程。

否达到数据完整性和一致性的要求和是否达到最终数据流通质量的目标。文献[7]描述了一种质量元模型,这个模型不仅可以用于数据仓库的设计,也可以用于衡量最终数据仓库的质量。文献[8]提出了一种如何应用质量因素和达到用户质量要求的方法。文献[9]通过统计数据仓库架构的一些指标,如星形数据集市维度表的数目等,然后通过 H. Zuse 提出的软件评估框架^[10]去确定哪些指标是可行的,从而最终评定数据仓库的设计质量。上述这些分析,主要采用了定性分析的方法,缺乏客观量化的分析。由于定性的分析缺乏客观性和可重现性,定量评估技术成为一个值得关注的方向。因为数据仓库为数据驱动,所以更关注于数据的统计解决方法。在此基础上,文中参考了数据事务系统和数据仓库系统数据质量的一些定义和统计方法(如文献[4]和文献[9]的统计分析方法),提出了一种定量统计分析方法,通过衡量数据仓库从数据源中所选取的那部分数据的数据质量,并最终评估数据仓库的设计质量,其分析过程也可以作为数据质量驱动设计的参考。

1 数据仓库设计质量评估指标体系

目标是通过常用的统计分析方法去评估业务系统中所要集成到数据仓库的那部分数据的数据质量,也就是定量分析源数据库表、属性两方面的数据质量指标,最后联合这两方面的指标计算最终设计的数据仓库的设计质量。要计算的目标指标是 $W_q(t_j, a_i)$, 其中 $q = m$ 或者 $q = d$, m 代表量值, d 代表维度。此指标衡量了属于表 t_j 的属性 a_i 适合作为最终数据仓库的量值或者维度的程度。而最终要评估的数据仓库设计质量 $W(DW)$ 则通过最终数据仓库量值属性和维度属性的 $W_q(t_j, a_i)$ 计算所得。

1.1 数据表指标

此指标衡量了表 t_j 适合抽取量值和维度的程度,支持选择那些表作为数据仓库的结构。定义表的整体指标如下:

$$WT_q(t_j) = \frac{\sum_{o=1}^n \theta_{q,o} * W_o(t_j)}{n} \quad (1)$$

其中, j 标识了源数据库中具体的某张表, $q = m$ 或者 $q = d$, m 表示量值, d 表示维度, $W_o(t_j)$ 为表特征的度量, o 标识了特征度量 $W_o(t_j)$, $\theta_{q,o}$ 为对应于 $W_o(t_j)$ 的系数。下面将对表特征度量 $W_o(t_j)$ 展开论述。

1.1.1 记录比率

此指标主要衡量了所分析数据表的记录占数据仓库所涉及的所有表的记录总数的比率。对于操作型数据库系统中的表,可以分为业务表和非业务表两种。

业务表多用于存储业务数据,表更新的频率比较高,而非业务表(如用户表)多存储一些静态数据,表更新的频率就比较低。鉴于此情况,在 ETL 过程中,不同表的数据抽取周期也不一样。因此,有必要对表的记录数作一个时间上的规范化。

现定义 $t_{j=1..p}$ 为数据仓库结构中所涉及的一组数据源表, $isActData_{j=1..p}$ 用于鉴定表 t_j 是否包含业务数据,如果 $isActData_j = true$,则表示 t_j 包含业务数据,如果 $isActData_j = false$,则表示 t_j 不包含业务数据。对于每张表 t_j , s_j 表示表 t_j 的时间间隔。对于业务表而言, s_j 等同于表的 ETL 周期,而对于非业务表而言,由于表的数据几乎静止不动,则令 $s_j = 0$ 。对于每张表 t_j ,时间规范化函数 $f(t_j)$ 作如下定义:

$$f(t_j) = \begin{cases} nRec(t_j) * \frac{\max(s_{j=1..p})}{s_j}, & isActData_j = true \\ nRec(t_j), & isActData_j = false \end{cases}$$

其中, $nRec(t_j)$ 表示表 t_j 的记录总数。最后记录比率可由 $f(t_j)$ 计算所得:

$$W_1(t_j) = \frac{f(t_j)}{\sum_{j=1}^p f(t_j)}$$

如果表 t_j 适合抽取数据作为量值,则 $\Theta_{m,1} > 0$,因为存储业务数据的表的记录数一般都比较大。反之,存储非业务数据的表的记录数都比较小,适合抽取数据作为维度,则 $\Theta_{d,1} < 0$ 。

1.1.2 属性比率

此指标表示了所分析数据表的属性总数占数据仓库结构中所涉及所有数据源表的属性总数的比率,作如下定义:

$$W_2(t_j) = \frac{nAttr(t_j)}{\sum_{j=1}^p nAttr(t_j)}$$

上式中, $nAttr(t_j)$ 表示表 t_j 的属性总数。对于量值而言, $\Theta_{m,2} < 0$,因为业务表的属性数量一般都比较小。反之,对于维度而言, $\Theta_{d,2} > 0$ 。

1.2 表属性指标

文中主要考虑了两种属性类型:数值类型(如 short, integer, float, double 等)和字符数值类型(如 character 和 string)。通过定义一组度量值 $W_k(a_i)$ 来衡量属性的数据特征。

定义了 $WA_d(a_i)$ 和 $WA_m(a_i)$ 两个指标评估了属性 a_i 是否适合分别作为度量和维度,具体定义如下:

$$WA_q(a_i) = \frac{\sum_k \theta_{q,k} * W_k(a_i)}{m} \quad (2)$$

上式中 $q = m$ 或者 $q = d$, m 表示度量值, d 表示维度。 i 标识了具体的某个属性, $W_k(a_i)$ 为属性的特征

度量值,其中 k 标识了 $W_k(a_i)$, $\Theta_{q,k}$ 为对应于 $W_k(a_i)$ 的系数。

在讨论 $W_k(a_i)$ 之前,先自定义一些关于属性的自定义函数,具体描述如下:

(1) $nValue(arr, v)$,该函数表示值 v 在序列 arr 中出现的次数。

(2) $nValues(a_i)$,该函数适用于字符数值型属性,表示属性 a_i 当中有多少种不同的值。

(3) $nDiff(a_i)$,该函数适用于字符数值型属性,将属性 a_i 的值归为 $nValues(a_i)$ 种类型,而该函数表示了每种类型的实例个数。

(4) $nDiff(a_i, n)$,该函数适用于数值型属性,将属性 a_i 的值域分成 n 等份,该函数返回的是一组属性值落在每个跨度内的个数。

(5) $nRec(t_j)$,该函数返回表 t_j 的记录总数。

1.2.1 属性的数据质量

文献[4,11~13]提出了一些衡量数据质量的指标,在进行数据质量评估时候,需要根据实际情况做一些取舍。在对数据仓库进行ETL操作的时候可能更关注的是数据的一致性(Consistent)、完整性(Complete)、有效性(Valid)和精确性(Accurate)等指标,具体的定义如下:

一致性:描述同一实体的属性值在不同的系统中是否一致。

完整性:数据集中是否有缺失的记录或者属性,在文中,如果属性为空即可定义为不完整。

有效性:数据是否满足用户自定义的规则,或者在一定的值域范围内。

精确性:数据是否跟其描述的客观实体特征相一致。

为了评估属性集的数据质量,定义了一个五元组的评估模型,该模型为: $M = \langle a, i, r, w, s \rangle$,各个元素分别表示:

a :需要评估的属性集合。

i :属性集 a 上需要评估的数据质量指标,如上面所列举的那些指标。

r :与评估指标的相对应的规则,如属性要求的值域范围和非空等。

w :赋予规则 r 的权值,表示了规则 r 在数据质量评估中所占的比重。

s :表示对应于规则 r 的最终评估结果,是0到1的实数。

属性集有不同的质量指标,而同一个指标又可以制定不同的评估规则,根据上面提出的模型,可以制定属性集的不同评估规则。假如属性 $a_i(a_i \in t_j)$ 制定了评估规则 $r = \langle r_1 = \text{非空}(\text{指标:完整性,权值:80}), r_2 =$

范围在0到100(指标:有效性,权值:20)>,根据对应的规则可以编写SQL语句来统计出符合的数目 n_1 和 n_2 ,则符合规则的评估结果为 $s = \langle s_1 = n_1/nRec(t_j), s_2 = n_2/nRec(t_j) \rangle$ 。

假如定义的规则为 $r = \langle r_1, r_2, \dots, r_v \rangle$,则对应地有评估结果 $s = \langle s_1, s_2, \dots, s_v \rangle$,那么对于属性 a_i 的数据质量评估值则取评估结果 s 的一个加权平均值:

$$W_1(a_i) = \frac{\sum_{s=1}^v w_s * s_s}{\sum_{s=1}^v w_s}, a_i \in t_j$$

当属性 a_i 越是符合所要评估的规则越是适合作为数据仓库的维度或则度量值,所以取 $\Theta_{m,1} > 0, \Theta_{d,1} > 0$ 。

1.2.2 属性的聚簇度

此指标主要衡量的是某属性的聚簇度,主要考察的是属性主要集中分布在哪些值。对于不同的属性类型,有不同的统计方法,对于数值型属性的聚簇度,作如下考虑:

$$W_2(a_i) = \frac{nValue(nDiff(a_i, n), 0)}{n},$$

n 的大小可以自定义,如果数值型属性值几乎不相等的话,那么, $W_2(a_i)$ 就接近于0,说明该属性的聚簇度很低。

对于字符数值型的聚簇度,作如下考虑:

$$W_2(a_i) = 1 - \frac{nValues(a_i)}{nRec(t_j)}$$

如果字符数值型属性 a_i 的值的分布种类很少,那么 $W_2(a_i)$ 接近于1,如果为主键或者复合主键的一个属性,那么 $W_2(a_i)$ 接近于0。

对于聚簇度高的属性,适合作为维度,而不太适合作为量值,那么取 $\Theta_{d,2} > 0, \Theta_{m,2} < 0$ 。

1.2.3 属性值的分布

此指标主要衡量的是属性值在其值域内的分布,对于不同的属性类型,有不同的统计方法,对于数值型属性的分布,作如下考虑:

$$W_3(a_i) =$$

$$\frac{\text{stdDev}(a_i)}{\max\{[\text{avg}(a_i) - \min(a_i)], [\max(a_i) - \text{avg}(a_i)]\}}$$

其中, $\text{stdDev}(a_i)$ 为属性 a_i 的标准差, $\text{avg}(a_i)$ 为 a_i 的平均值, $\min(a_i)$ 为 a_i 的最小值,而 $\max(a_i)$ 为 a_i 的最大值,而函数 $\max(a, b)$ 则取 a 和 b 两者的最大值。

对于字符数值型属性而言,先定义一个向量 V , V 的长度为 $nValues(a_i)$,定义 V 的第一个元素为:

$$v[1] = nRec(t_j) - (nValues(a_i) - 1), a_i \in t_j$$

而其他的 $nValues(a_i) - 1$ 个元素等于1,最后将

字符数值型属性的分布定义为:

$$W_3(a_i) = 1 - \frac{\text{stdDev}(n\text{Diff}(a_i))}{\text{stdDev}(V)}$$

对于分布趋于统一的属性适合作为维度,那么取 $\Theta_{d,3} > 0, \Theta_{m,3} < 0$ 。

1.2.4 属性的类型

不同类型的属性对于作为数据仓库的度量值或者维度的适合度不同,下面根据不同的类型返回一个值,定义如下:

$$W_4(a_i) = \begin{cases} 0, a_i \text{ 为 String 类型} \\ 0.5, a_i \text{ 为 Short 或 Integer 类型} \\ 1, a_i \text{ 为 Float 或 Double 类型} \end{cases}$$

相对于字符数值型属性来说,数值型属性更适合作为数据仓库的度量值,所以取 $\Theta_{m,4} > 0$,反之取 $\Theta_{d,4} < 0$ 。

1.3 数据仓库质量

对于表 t_j 的属性 a_i 通过 $W_m(t_j, a_i)$ 和 $W_d(t_j, a_i)$ 两个指标衡量其是否适合分别作为数据仓库的量值和维度,这个指标通过联合 $WT_q(t_j)$ 和 $WA_q(a_i)$ 计算机所得,具体计算过程如下:

$$W_q(t_j, a_i) = WT_q(t_j) * WA_q(a_i), a_i \in t_j \tag{3}$$

其中, i 和 j 分别标识了某个属性和某张表, $WT_q(t_j)$ 和 $WA_q(a_i)$ 分别是表和属性两方面的指标。

最后,用 M_{dw} 表示数据仓库的 n_m 个量值,而 D_{dw} 表示 n_d 个维度,这样数据仓库的设计质量 $W(DW)$ 可作如下定义:

$$W(DW) = \frac{\sum_{a_i \in M_{dw}} W(t_j, a_i) + \sum_{a_i \in D_{dw}} W(t_j, a_i)}{n_m + n_d} \tag{4}$$

2 实例分析

文中的实验例子来源于某模具企业的“管理驾驶舱”项目,该项目为基于包含库存、生产、销售、薪酬、设备、人事等几大主题的数据仓库的一个决策支持管理系统(DSS)。数据仓库的存储环境为 SQL Server2008 数据库管理系统。

数据仓库的数据集成于该模具企业支撑 ERP 系统的各个数据库,数据库的存储环境为 SQL Server2005 数据库管理系统。文中正是通过量化的手段来统计分析数据仓库涉及的一些数据源质量指标,并计算最终的数据仓库设计质量。对于需要计算的一些自定义函数(上文所提),则通过编写 SQL 语句来计算统计。如需要计算表 T 的记录总数,则通过 SQL 语句 SELECT COUNT(*) FROM T 来统计。而对于上文所提到的对应于特征度量值的参数 Θ ,为了便于计算,如果 $\Theta > 0$ 则取 $\Theta = 1$,反之取 $\Theta = -1$ 。另外,时间维度是

一个特殊的维度,且文中并没有考虑时间类型的属性,所以时间维度并没有纳入计算范围。

表 1 为文中分析的一个数据集市例子,第二列为事实表 Fact_HoursAmt_Release 的属性,其中包括了维度和量值。第三列 SourceAttribute 为事实表属性在数据源中对应的属性,第四列 SourceTable 为第二列属性所在的数据源表。

根据上面提出的模型(1)的分析过程,和数据仓库 ETL 的周期可得各张表的时间间隔。另外,表的记录数和属性数,可以通过 SQL 语句统计出来。最终,可以计算出数据源表的质量指标,具体结果归列于表 2。从结果中可以看出,表 Department 的 WT_d 值比较大,可见其非常适合抽取数据作为数据仓库的维度,其他结果的分析就不再赘述。

表 1 事实表 Fact_HoursAmt_Release

	Fact_Hours Amt_Release	SourceAttribute	SourceTable
Dimension	ReleaseDate_Key		
	POrderKey	POrderID	PP_Order
	Dept_Key	DepartmentID	Department
Measure	ReleaseHours	OSchedulePHours	PP_OperationSchedule
	ReleaseAmt	OScheduleCost	PP_OperationSchedule

表 2 数据源表的质量指标

Dimension	
Table	WT_d
PP_Order	0.874
Department	0.962
Measure	
Table	WT_m
PP_OperationSchedule	0.781

根据模型(2)的分析过程,对于属性数据质量的计算,定义了“非空”、“数据范围”、“数据长度”等评估规则。另外,对于属性的分布和聚簇度的计算,可以通过 SQL 语句计算可得。最终,属性质量指标的计算结果归列于表 3。从计算结果可以看出,属性 POrderID、DepartmentID、OSchedulePHours、OScheduleCost 都很适合作为数据仓库的量值或者维度。

表 3 属性质量指标

Dimension	
Attribute	WA_d
POrderID	0.700
DepartmentID	0.700
Measure	
Attribute	WA_m
OSchedulePHours	0.844
OScheduleCost	0.759

根据模型(3),用表 2 和表 3 的计算结果可以计算出属性适合作为维度和量值的程度,其计算结果如表 4 所示。

根据模型(4)可以计算出所设计的数据集市设计

质量值 $W(DW)$ 为 0.634,按照“[0,0.5]:一般、[0.5,0.8]:良好、[0.8,1]:优秀”的划分,实验验证的数据集市的设计质量为良好。对于其他数据集市的设计质量,可通过同样的计算方法得出。

表 4 $W_q(t_j,a_i)$ 计算结果

Dimension	
Attribute	$W_d(t_j,a_i)$
POrderID	0.612
DepartmentID	0.673
Measure	
Attribute	$W_m(t_j,a_i)$
OSchedulePHours	0.659
OScheduleCost	0.593

3 结束语

文中提出了一种自动验证数据仓库设计质量的定量分析方法,用数据去衡量数据仓库的设计质量,进一步研究的话,可以用文中提出的方法开发一个自动验证数据仓库设计质量的辅助判断工具。当然,一个数据仓库设计的好坏,往往是一个综合评定的结果。文中得出的结果只不过是定量分析的结果,可以作为判断数据仓库设计质量好坏的一个判断因素,另外,可以结合数据仓库的逻辑结构、设计经验等非定量因素去综合判定数据仓库设计质量的好坏。

参考文献:

[1] Chengalur-Smith I N, Ballou D P, Pazer H L. The impact of data quality information on decision making: an exploratory analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 1999, 11(6): 853-864.

[2] Aebi D, Perrochon L. Towards improving data quality[C]//Proc of the international conference on information systems and management of data. [s. l.]: [s. n.], 1993: 273-281.

[3] Wang R Y, Strong D M. Beyond accuracy: what data quality

means to data consumers[J]. Journal of Management Information Systems, 1996, 12(4): 5-30.

[4] Karr A F, Sanil A P, Banks D L. Data quality: a statistical perspective[J]. Statistical Methodology, 2006, 3(2): 137-173.

[5] Bouzeghoub M, Kedad Z. A Quality-based Framework for Physical Data Warehouse Design[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2000). Stockholm, Sweden: [s. n.], 2000: 5-6.

[6] Theodoratos D, Bouzeghoub M. Data Currency Quality Factors in Data Warehouse Design[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'99). Heidelberg, Germany: [s. n.], 1999.

[7] Manfred A, Jeusfeld C Q. Design and Analysis of Quality Information for Data Warehouse[C]//European Commission in ESPRIT Long Term Research Project 22469 DWQ. [s. l.]: [s. n.], 1998.

[8] Vassiliadis P, Bouzeghoub M, Quix C. Towards Quality-oriented Data Warehouse Usage and Evolution[C]//The 11th International Conference on Advanced Information System Engineering. [s. l.]: [s. n.], 2000: 89-115.

[9] Calero C, Piattini M. Towards Data Warehouse Quality Metrics[C]//Proceedings of the International Workshop on Design and Management of Data Warehouses (DMDW'2001). Interlaken, Switzerland: [s. n.], 2001.

[10] Zuse H. A Framework of Software Measurement[M]. Hawthorne, NJ, USA: Walter de Gruyter & Co., 1997.

[11] Wang Y, Wang R Y. Anchoring Data Quality Dimension in Ontological Foundation[J]. Communication of the ACM, 1996, 39(11): 86-95.

[12] Wang R Y, Storey V C, Firth C P. A Framework for Analysis of Data Quality Research[J]. IEEE Transaction on knowledge and data engineering, 1995, 7(4): 623-640.

[13] Pipino L L, Lee Y W, Wang R Y. Data Assessment[J]. Communication of the ACM, 2002, 45(4): 211-218.

(上接第 160 页)

[7] Vapnik V. Statistical Learning Theory[M]. New York: John Wiley & Sons, 1998.

[8] Platt J, Cristianini N, Shawe-Taylor J. Large Margin DAGs for Multiclass Classification[C]//Proceedings of Neural Information Processing Systems. [s. l.]: MIT Press, 1999: 547-553.

[9] 张苗, 张德贤. 多类支持向量机文本分类方法[J]. 计算机技术与发展, 2008, 18(3): 139-141.

[10] Takahashi F, Abe S. Decision Tree-based Multiclass Support Vector Machines[C]//Proceedings of the 9th International Conference on Neural Information Processing. [s. l.]: [s. n.], 2002: 1418-1422.

[11] 胡国胜. 支持向量机及在电力系统中的应用[J]. 高电压技术, 2007, 33(4): 101-105.

[12] Lei H S, Govindaraju V. Half-Against-Half Multi-class Support Vector Machines[C]//Proceedings of Multiple Classifier Systems 2005. USA, California: [s. n.], 2005: 156-164.

[13] 刘健, 刘忠, 熊鹰. 改进的二叉树支持向量机多类分类算法研究[J]. 计算机工程与应用, 2010, 46(33): 117-120.

[14] 刘培胜, 贾银山, 韩云萍. 一种改进的简化支持向量机[J]. 辽宁石油化工大学学报, 2009, 29(1): 76-78.

[15] Srivani J, Swarup K S. Power system static security assessment and evaluation using external system equivalents[J]. Electrical Power and Energy Systems, 2008, 30(2): 83-92.

[16] Chang Chin-Chung, Lin Chih-Jen. LIBSVM: A Library for Support Vector Machines[EB/OL]. 2005. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.

一种评估数据仓库设计质量的方法

作者: [刘益江](#), [毛宁](#), [陈庆新](#)
作者单位: [广东工业大学 广东省计算机集成制造重点实验室, 广东 广州 510006](#)
刊名: [计算机技术与发展](#)
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2012(9)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjtz201209043.aspx