

# 数据挖掘在运动自行车量身定做中的应用

王建华<sup>1</sup>, 赵静楠<sup>1</sup>, 张聪敏<sup>1</sup>, 刘海云<sup>2</sup>

(1. 北京工业大学 机械工程与应用电子技术学院, 北京 100124;

2. 北京航轮运动器材有限公司, 北京 100076)

**摘要:** 基于自组织数据挖掘的基本原理, 根据运动自行车车架设计的背景知识, 选取影响车架尺寸的骑行者的身体特征, 并选择相关的数据作为学习样本, 在 Knowledge Miner 中建模, 得到车架尺寸和身体特征之间的定量关系, 实现了运动自行车车架量身定做的功能.

**关键词:** 运动自行车; 车架设计; 人体尺寸; 量身定做; 自组织数据挖掘; 集群数据处理方法 (GMDH)

**中图分类号:** TP 18

**文献标志码:** A

**文章编号:** 0254-0037(2010)06-0742-06

目前, 国内关于自行车车架量身定做的服务很少, 可以提供此类服务的几个厂家也是引进国外的技术, 多数厂家还是根据设计人员的经验反复试算进行设计的, 这样不仅浪费时间, 也不能保证设计的车架完全满足骑行者的需要. 另外, 经过调查, 引进的设计规则不适合直接用来设计亚洲人的车架, 主要是因为自行车的特殊性, 其使用者的骑行习惯和身体比例与欧美国家有所不同, 导致在实际的设计中会遇到很多问题<sup>[1]</sup>. 运用数据挖掘的原理, 从收集到的车架尺寸信息和身体特征信息中挖掘出它们之间的关系, 建立一套合适的自主设计规则, 将会提高车架的设计水平和效率, 为自行车运动在国内的发展做出贡献.

通俗地说, 自组织数据挖掘 (SODM) 依靠的是系统内部的相互作用达到“发现知识”的目的<sup>[2]</sup>. 自组织建模的过程实质上是确定系统最优复杂度模型的过程. 它处理的对象为若干输入变量, 1 个或多个输出变量构成的变量间关系待定的一个封闭系统. 通过各输入变量相互结合产生众多候选模型集, 利用外准则选出若干项最优模型, 再将其结合, 由此得到下一代, 如此不断重复, 直到新产生的模型不比上一代更加优秀为止, 则倒数第 2 代中的模型就是寻找的最优复杂度模型<sup>[3]</sup>.

## 1 GMDH (group method of data handling) 方法简介

自组织数据挖掘的核心是 GMDH, 即将数据分为训练集和测试集, 在训练集上使用内准则进行参数估计得到中间待选模型, 而在测试集上使用外准则进行中间候选模型的选择, 这个过程不断重复直到外准则值不能再改善才停止, 这样则可以保证在一定噪声水平下得到数据拟合精度和预测能力实现最优平衡的最优复杂度模型. 这也是自组织数据挖掘不同于遗传算法等其他“他组织数据挖掘”的特点之一<sup>[4-5]</sup>.

GMDH 算法的基本步骤<sup>[6-8]</sup>如下:

- 1) 数据预处理 确定主要的输入和输出以及待定的关系  $f$
- 2) 样本数据分组 样本数据分为训练集和测试集, 待选模型将在训练集上产生, 而测试集则用于检验和筛选产生的模型.
- 3) 建立参考函数 经典的  $K-G$  多项式如下

$$y = a_0 + \sum_{i=1}^m \sum_{j=1}^m a_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m a_{ijk} x_i x_j x_k + \dots$$

其中,  $a_0, a_{ij}, \dots$  是局部函数的系数;  $x_i, x_j, x_k, \dots$  是作为此次选择过程的输入.

收稿日期: 2008-10-29.

作者简介: 王建华 (1963-), 女, 山西浑源人, 副教授.

4) 选择目标准则 从具有外补集性质的标准选择准则表中选出目标准则.

5) 选择局部函数 根据所选择的目标准则, 在检测集上检验局部函数的性质, 选择具有最佳性质的函数作为最优模型. 如果函数不是最优的, 则选择出 F 个局部函数作为下一次的输入 (F 为自由度).

6) 最优模型产生 经过多次的选择, 产生最优的模型.

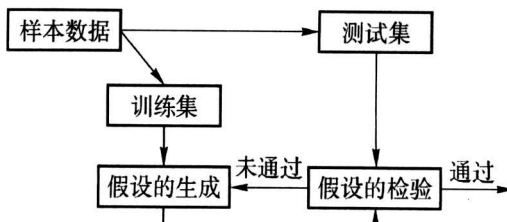


图 1 自组织数据挖掘的流程图<sup>[6]</sup>

Fig 1 The flow chat of SODM<sup>[6]</sup>

图 1所示为自组织数据挖掘的工作流程图.

## 2 基于 GMDH 原理的车架尺寸与身体特征关系的模型研究

### 2.1 模型参数的定义

在运动自行车车架量身定做系统的研究中, 人体的特征和对应的车架尺寸构成了系统的输入和输出<sup>[9]</sup>. 仅仅依靠自行车设计的知识很难把它们之间的关系定量地确定出来, 这也是制约自行车尤其是运动自行车车架设计质量和效率的瓶颈问题. 本文采用数据挖掘中的自组织数据挖掘方法和 Knowledge Miner 软件进行建模, 研究身体特征和车架尺寸之间的模糊关系的定量表示.

根据行业经验和知识, 归纳出影响车架尺寸的人体尺寸, 如表 1 所示.

表 1 影响车架尺寸的人体尺寸

Table 1 The body sizes which affect the sizes of bicycle frame

人体尺寸	英文名称	说明	人体尺寸	英文名称	说明
X <sub>1</sub>	Height	身高	X <sub>5</sub>	Shoulder Width	肩宽
X <sub>2</sub>	Stemum Notch	躯干长	X <sub>6</sub>	Flexibility	身体的灵活度
X <sub>3</sub>	Inseam Length	腿长	X <sub>7</sub>	Weight	体重
X <sub>4</sub>	Arm Length	臂长	X <sub>8</sub>	Shoe Size	脚的尺寸

自行车车架的关键尺寸如表 2 和图 2 所示, 人体各部分尺寸测量方法如图 3 所示. 人体测量时的站姿要求为: 双脚间距 18 cm, 靠墙站立, 双肩放平, 光脚或只穿袜子. 身体灵活度的标准要求在上述站姿下身体前弯, 双臂伸直, 以手指可以触到的位置来划分等级.

表 2 自行车车架的关键尺寸

Table 2 The key sizes of bicycle frame

车架尺寸	英文名称	说明
X <sub>9</sub>	C-C	坐管中心到五通管中心距离
X <sub>10</sub>	C-T	坐管上表面到五通管中心距离
X <sub>11</sub>	Overall Reach	上管有效长度和把立长度之和
X <sub>12</sub>	Saddle Height	五通管中心到车座最高点的距离
X <sub>13</sub>	Handlebar Width	车把的宽度

身体灵活度分为 7 个等级<sup>[10]</sup>:

- 1) 0 级 手指在膝盖部位.
- 2) 3 级 手指触到胫骨部位.
- 3) 5 级 手指触到脚趾头.
- 4) 6 级 手指触到地板.
- 5) 7 级 指关节触到地板.
- 6) 8 级 手指平触地板.
- 7) 10 级 手掌平触地板.

选取若干已经被使用的自行车的车架尺寸数据和相对应的骑行者的人体尺寸数据作为学习样本, 前提是这些自行车的功能和舒适性已经被使用者认同. 选取 80 条基本车架数据和对应的人体特征的数据, 使用 Knowledge Miner 软件进行建模. 鉴于篇幅的原因, 本文仅对 Overall Reach 和 C-C 的车架尺寸进行尺

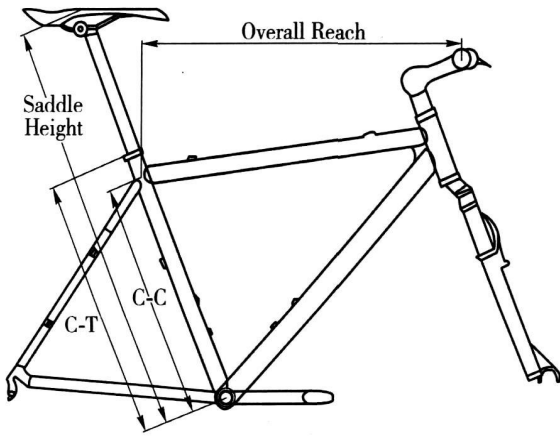


图 2 车架简图

Fig 2 Sketch of bike frame

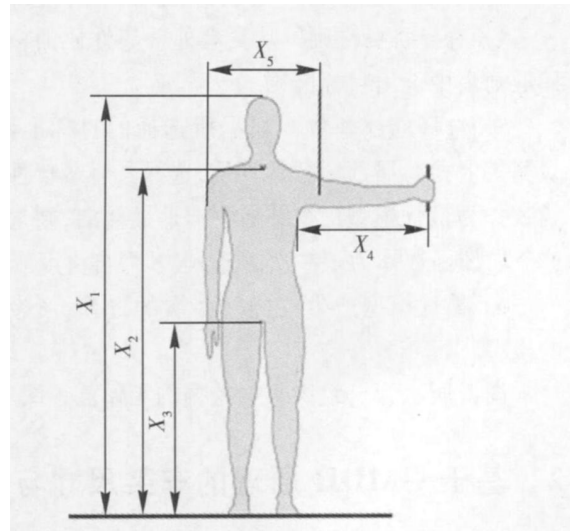


图 3 人体测量方法

Fig 3 Method of body measurements

寸与身体特征的关系建模,对于其他车架管件尺寸可以用同样的方法进行分析.

## 2.2 模型与结果分析

### 1) Overall Reach模型与结果分析

在 Knowledge Miner中使用“Create-Input/Output Model”,建立以  $X_{11}$ 即 Overall Reach为输出,以  $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ 为输入的模型,经过 5层运算得出

$$X_{11} = 3.62041 \times 10^{-1} a + 2.77354 b + 6.67139 \times 10$$

$$a = a_1$$

$$a_1 = a_{11}$$

$$a_{11} = 1.54829 \times 10^{-1} X_2 - 2.25521 \times 10$$

$$b = -3.82756 \times 10^{-1} b_1 + 1.35987 \times 10 b_2$$

$$b_1 = b_{11}$$

$$b_{11} = 2.61115 \times 10^{-1} X_4 - 1.65398 \times 10$$

$$b_2 = -5.87916 \times 10^{-1} b_{21} + 1.52631 \times 10^{-1} b_{22}$$

$$b_{21} = 2.24071 \times 10^{-1} X_3 - 1.87238 \times 10$$

$$b_{22} = 7.66580 \times 10^{-2} X_2 + 1.35124 \times 10^{-1} X_4 - 1.97250 \times 10$$

模型的参数值分别为:误差率为 0.02%,模型拟合值与实际值方差为 0.0003,模型样本描述能力为 99.97%,模型的拟合度为 98%.  $X_2, X_4, X_3$ 对模型的贡献比率分别为 46%、32%、22%.

归纳后的模型为

$$X_{11} = 4.0760 + 4.9735 \times 10^{-1} X_2 + 5.0067 \times 10^{-1} X_4 - 4.9686 \times 10^{-1} X_3 \quad (1)$$

图 4中,原始数据曲线与模型拟合数据曲线拟合度非常好.由图 4和各项指标可以看出,本模型反映出一个有效输入输出关系,正确描述 98%的数据,且对于样本来说,描述能力为 99.97%.对于选择的模型类型,本模型具有足够的信息.

### 2) C-C模型与结果分析

在 Knowledge Miner中使用“Create-Input/Output Model”,建立以  $X_9$ 即 C-C为输出,以  $X_1, X_2, X_3, X_4, X_5, X_7$ 为输入的 non-linear/static模型,经过 3层运算得出

$$X_9 = 6.34605 \times 10^{-1} X_3 + 7.98990 \times 10^{-3} X_7 + 7.35347 \times 10^{-1} \quad (2)$$

模型的参数值分别为:误差率为 0.44%,模型拟合值与实际值方差为 0.01,模型样本描述能力为

98.46%, 模型的拟合度为 90%。X<sub>3</sub>、X<sub>7</sub> 对模型的贡献比率分别为 99% 和 1%。

由图 5 和各项指标可以看出, 本模型反映出一个有效输入输出关系, 正确描述 90% 的数据, 且对于样本来说, 描述能力为 98.46%。

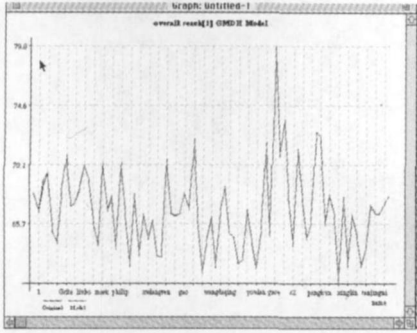


图 4 Overall Reach模型拟合效果图

Fig 4 Fitting effect of Overall Reach model

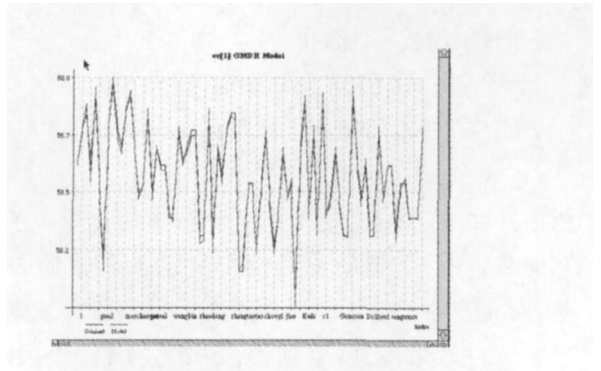


图 5 C-C模型拟合效果图

Fig 5 Fitting effect of C-C model

### 2.3 模型的预测能力及误差

为了检验模型的有效性, 使用样本以外的数据进行检验. 使用 What-if Prediction 进行预测, 分别得出表 3 和表 4 所示的预测数据.

表 3 Overall Reach 的实际值和预测值

Table 3 Overall Reach's actual value and predicted value

实际值 /cm	预测值 /cm	预测误差 /%
66.450 0	66.441 2	0.013
67.750 0	67.743 2	0.010
67.200 0	67.192 2	0.016

表 4 C-C 的实际值和预测值

Table 4 C-C's actual value and predicted value

实际值 /cm	预测值 /cm	预测误差 /%
59	59.241 13	0.401
54	53.846 98	0.283
49	49.177 83	0.363
57	56.385 41	1.070
60	59.431 51	0.947
57	56.529 63	0.825

检测结果可以进一步证明模型能准确地反映车架的尺寸 (C-C, Overall Reach) 和身体特征之间的关系. 其他的车架尺寸使用同样的建模方法可以确定与身体特征的关系.

## 3 模型解析

### 3.1 基于 GMDH 的 Overall Reach 的模型解析

为了方便得到人体尺寸与车架尺寸之间的关系, 将式 (1) 简化为

$$X_{11} = 4 + 0.5(X_2 - X_3 + X_4) \tag{3}$$

由式 (3) 可以看出, 设计上管有效长度和把立长度时需要考虑身高、臂长和腿长.

以山地车为例, 最佳的骑行姿势是肩关节呈 90°, 这个姿势不容易疲劳而且容易发挥力量; 肘关节弯曲度约 120°, 这种弯曲度最能发挥肩周围及手腕的肌肉的能量; 脚腕的角度呈 90°~120° 时, 最能发挥踩踏力, 膝盖和小腿肚也不宜疲劳, 可以长时间地保持体力. 如图 6 所示, AB 约为 Overall Reach

在简化公式里, (X<sub>2</sub> - X<sub>3</sub>) 是上身长, 可将 Overall Reach 的影响因素简化为上身长和臂长. 在图 6 所示的最佳骑行姿势中, 立管一定时, 上身长和臂长正好是影响 Overall Reach 的因素且是正比例的关系. 当然, 由于是量身定做, 对于骑行习惯不同的客户来说, 可以根据具体情况进行后期的调整.

### 3.2 基于 GMDH 的 C-C 模型解析

同样,将式(2)简化为

$$X_9 = 0.635X_3 + 0.008X_6 + 0.735 \quad (4)$$

式(4)与目前国际上流行的 Lemond公式

$$X_9 = 0.65X_3$$

相比, C-C模型同时还考虑了骑行者的体重. 在自行车的车架中,车架的菱形结构决定了车架的力学性能,而立管的长度是菱形中重要的因素. 对于腿长一定的骑行者来说,体重越大,施加在立管上的力就越大,虽然体重的影响因子只有 1%,但是忽略体重的车架设计将会为未来的骑行埋下安全隐患.

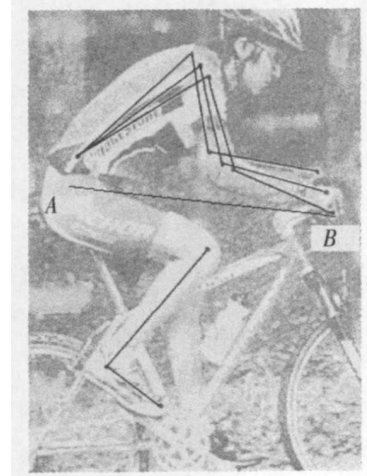


图 6 骑行姿势

Fig 6 Riding position of riders

## 4 结束语

运动自行车的车架量身定做是一项很难量化的工作,由于其使用者——人的适应性和多变性,以及其使用途径的多样性,量身定做的工作必须以客户为中心,以安全为指导. 基于自组织数据挖掘在复杂系统的数据挖掘中具有很好的作用,尤其是在先验知识不足的情况下,能很好地发现其中的规律. 本文利用自组织数据挖掘方法得到了部分车架尺寸和身体特征之间的定量关系,为实现运动自行车车架量身定做功能打下了基础.

作者以 Overall Reach和 C-C的车架尺寸为例对自组织数据挖掘方法在运动自行车车架量身定做中的应用进行研究,说明了此种方法的应用准确性及可行性. 由于运动自行车车架各个尺寸之间的约束性,在所有的设计工作中还需要考虑其他因素,并不能仅依据此规则进行.

## 参考文献:

- [1] 陶浩, 段红杰. 人机工学在碳纤维赛车车架结构设计中的应用 [J]. 机械研究与应用, 1998(1): 10-12.  
TAO Hao DUAN Hong-jie The application of ergonomics in structure design of carbon fiber frame of competition bike [J]. Mechanical Research & Application, 1998(1): 10-12. (in Chinese)
- [2] JOHANN-ADOLF Muller Selforganizing modelling as a part of simulation process [J]. System Analysis Modelling Simulation, 2003, 43(10): 1283-1299.
- [3] 陈川, 何跃, 贺昌政. 基于自组织数据挖掘方法的经济预警研究 [J]. 成都信息工程学院学报, 2005, 20(1): 101-106.  
CHEN Chuan HE Yue HE Chang-zheng Economic pre-waming based on self-organization data mining [J]. Journal of Chengdu University of Information Technology, 2005, 20(1): 101-106. (in Chinese)
- [4] 高林, 贺昌政. GMDH方法在通货膨胀中的应用 [J]. 现代管理科学, 2006(3): 106-107.  
GAO Lin HE Chang-zheng Application of GMDH in inflation [J]. Modern Management Science, 2006(3): 106-107. (in Chinese)
- [5] 蒋志全. 基于 GMDH原理的自组织数据挖掘模型研究 [D]. 大连: 大连海事大学交通运输管理学院, 2004.  
JIANG Zhi-quan The research of self-organizing data mining based on GMDH principle [D]. Dalian: Transportation Management College Dalian Maritime University, 2004. (in Chinese)
- [6] 贺昌政, 俞海, 卢跃奇. 自组织组合预测方法及其应用 [J]. 数量经济技术经济研究, 2002, 19(2): 84-87.  
HE Chang-zheng YU Hai LU Yue-qi Self-organizing combination forecasting method and its application [J]. Quantitative and Technical Economics, 2002, 19(2): 84-87. (in Chinese)
- [7] JOHANN-ADOLF Muller FRANK Lenke Self-organizing data mining an intelligent approach to extract knowledge from data [M]. Dresden: Technische Universitaet Dresden Press, 1999.
- [8] HAND David MANNILA Heikki SMYTH Padhraic Principles of data mining [M]. Massachusetts: Massachusetts Institute of Technology Press, 2001.

- [9] 王建华, 杨文通, 胡凤来. 运动自行车车架结构的参数化设计 [J]. 北京工业大学学报, 2001, 27(4): 472-475.  
WANG Jian-hua, YANG Wen-tong, HU Feng-lai. Parameterized design of sport bike frame structure [J]. Journal of Beijing University of Technology, 2001, 27(4): 472-475. (in Chinese)
- [10] 梁艳霞. 人体尺度在产品中的应用: 以双层床的设计为例 [J]. 郑州轻工业学院学报: 社会科学版, 2009, 10(4): 68-71.  
LIANG Yan-xia. Application of anthropometric dimensions in product design: taking two-layer bed as example [J]. Journal of Zhengzhou University of Light Industry: Social Science Edition, 2009, 10(4): 68-71. (in Chinese)

## Application of Data Mining in the Customization Design of the Sport Bicycle

WANG Jian-hua<sup>1</sup>, ZHAO Jing-nan<sup>1</sup>, ZHANG Cong-min<sup>1</sup>, LIU Hai-yun<sup>2</sup>

(1. College of Mechanical Engineering and Applied Electronics Technology, Beijing University of Technology, Beijing 100124, China; 2. Beijing Hangleun Sports Apparatus Co., Ltd., Beijing 100076, China)

**Abstract:** Based on the basic theory of self-organizing data mining and the backgrounds of bicycle's frame design, several body sizes and body characters which may have impact on the bicycle's frame sizes and the corresponding frame sizes are selected as data sample, then create input/output models in Knowledge Miner are created. Finally, the quantitative relationship between body characters and frame sizes is found, and the customization design of the bicycle's frame is realized.

**Key words:** sport bicycle; design of bike frame; body sizes; customization design; self-organizing data mining; group method of data handling (GMDH)

(责任编辑 梁洁)