

文章编号:1000-6893(2009)08-1380-05

递阶偏最小二乘回归在飞机研制费用预测中的应用

王礼沅, 郭基联, 张恒喜

(空军工程大学 工程学院, 陕西 西安 710038)

Application of Hierarchical Partial Least Squares Regression to Development Cost Prediction of Aircraft

Wang Liyuan, Guo Jilian, Zhang Hengxi

(Engineering College, Air Force Engineering University, Xi'an 710038, China)

摘要: 分析了飞机研制费用样本数据少、费用驱动因子众多的特点,考虑到递阶偏最小二乘回归(Hi-PLS)方法在变量规模巨大情形下进行回归建模的优势,应用递阶偏最小二乘回归方法对飞机研制费用进行预测。以战斗机机体研制费用预测为例进行研究,首先对战斗机机体研制费用驱动因子进行分组,然后应用递阶偏最小二乘回归方法对分组后的费用驱动因子进行回归建立机体研制费用预测模型。实例表明,在飞机研制费用预测方面,采用递阶偏最小二乘回归方法预测更能体现研制费用与飞机性能参数之间的关系。

关键词: 飞机; 研制费用; 递阶偏最小二乘回归; 费用预测模型; 费用驱动因子

中图分类号: V37;O241.5 **文献标识码:** A

Abstract: In the analysis of the development cost of an aircraft there are usually only small samples with a large number of cost drive factors. In view of this fact and considering the advantages of hierarchical partial least squares regression (Hi-PLS) in building a regressive model in the condition of a large number of variables, an application of Hi-PLS to aircraft development cost prediction is proposed using the development cost prediction of a fighter plane fuselage as an example. After the cost drive factors of the fighter plane fuselage development cost are grouped, Hi-PLS is applied to the regress of the grouped cost drive factors, and a fuselage development cost prediction model is established. The calculated results in the example show that Hi-PLS model can reflect satisfactorily the correlation between fuselage development cost and performance parameters of a military aircraft.

Key words: aircraft; development cost; hierarchical partial least squares regression; cost prediction model; cost drive factor

由于中国航空工业起步较晚,自主发展的飞机型号十分有限。对于飞机的研制费用预测来说,反映飞机质量特性的变量繁多,而样本数相对于变量数则明显偏少,如何尽可能多地利用各变量所蕴含的信息成为一个难题。文献[1]~文献[3]对此问题进行了深入研究和探讨,把偏最小二乘法应用到这样一个典型的小样本多元数据问题上,取得了较好的效果,然而,其选取的变量仍然比较有限。为更好地解决这一问题,本文将采用递阶偏最小二乘回归(Hierarchical Partial Least Squares Regression, Hi-PLS)方法进行研究。

在很多数据分析工作中,所谓海量可能是指大规模的样本点数量,但也可能是成百上千的变量。尽管偏最小二乘在处理多变量的建模问题中

是十分有效的,但当自变量太多时,得到的模型结果(比如自变量的因子载荷图、回归参数和变量重要性指标(VIP)值等指标)就会显得十分庞杂、难以分析和解释^[3]。

为了解决上述困难,人们试图找到一种更有效的方法将众多的自变量减少到一个较少的、易于处理的数量。然而,简单地删除一些变量,选取原变量集合中的一个子集用于建模显然是不可靠的,因为变量的减少有可能伴随着信息的丢失,从而导致错误的模型结果以及对系统虚假的分析解释。一种更好的解决办法是,将自变量集合划分成若干具有相同或相似内涵的组(Block),例如,可以按照时间、空间或某些属性等进行划分,这种对变量的分块自然引出了对建模过程的分层处理,即模型根据变量的层次结构分成:上层模型(Upper Level)——子块(Sub-block)间的模型;

下层模型 (Lower Level)——各子块内的模型。在每一层上,可应用偏最小二乘回归 (PLS)、主成分分析等标准的建模方法;并且,相应求得的主成分、因子载荷和残差分析等结果,可以更方便地用于模型的解释和评价。可以看出,在递阶建模过程:一方面可以得到形式简约、易于分析的整体模型 (上层模型关系);另一方面,还可以进一步解释所概括信息的内涵意义 (下层模型关系)^[4]。

上述多变量递阶建模的思想是十分简单可行的。1996 年,法国数学家伍德 (Wood) 提出的递阶偏最小二乘回归就是基于这种思路。该方法首先将自变量集合划分为若干子块,使得在每一子块中的变量均具有相同或相似的含义。然后,分别建立每一子块与因变量的偏最小二乘回归模型,提取相应的偏最小二乘成分;再使用在各子块中提出的偏最小二乘成分作为自变量,与因变量建立整体的偏最小二乘回归模型。在这个过程中,由于子块的数量远远小于原始变量的总数,并且各个子块有其特定的意义,因此,采用这种方法得到的结果,对信息的综合概括能力更强,具有更强的解释作用和应用价值^[4-6]。

1 递阶偏最小二乘回归方法的建模步骤

下面介绍递阶偏最小二乘回归方法的建模步骤。

首先将数据作标准化处理。为简单起见,仍记经标准化处理后的数据矩阵为 X 和 Y 。设自变量 X 被分为 p 个子块,即 $X = [X_1 \ X_2 \ \dots \ X_p]$, 每个子块 $X_i (i=1, 2, \dots, p)$ 含有 $p_i (i=1, 2, \dots, p)$ 个自变量,即 $X_i = [x_{i1} \ x_{i2} \ \dots \ x_{ip_i}]$ 。

(1) 建立底层模型 (Base Model)。分别建立因变量 Y 与每一个自变量子块 X_i 的偏最小二乘回归模型。根据交叉有效性,提取相应的偏最小二乘成分 $t_{i1}, t_{i2}, \dots, t_{im_i} (i=1, 2, \dots, p)$, 记 $m_i (i=1, 2, \dots, p)$ 为所提取的 X_i 中的偏最小二乘成分个数。

(2) 建立顶层模型 (Top Model)。对所提取的全部自变量子块 X_i 的偏最小二乘成分 $t_{i1}, t_{i2}, \dots, t_{im_i}$ 与因变量 Y 进行偏最小二乘回归,根据交叉有效性,提取偏最小二乘成分 $t_{Top,1}, t_{Top,2}, \dots, t_{Top,m}$ 。

(3) 由于底层模型 $t_{i1}, t_{i2}, \dots, t_{im_i}$ 是 X_i 的线性组合,代入顶层模型,可以得到 Y 关于 X_i 的线性回归模型。

概括起来,递阶偏最小二乘回归建模过程示意图如图 1 所示。

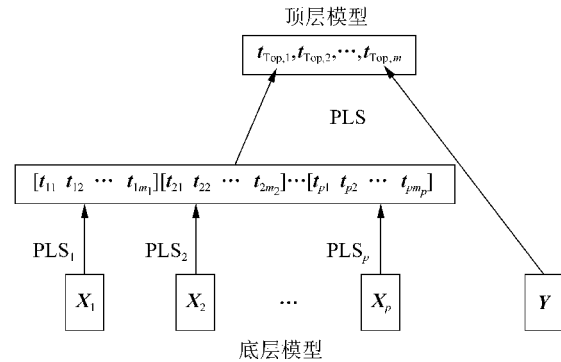


图 1 递阶偏最小二乘回归建模示意图

Fig. 1 Modeling schematic of Hi-PLS

2 飞机研制费用估算建模

一般采用参数法对飞机研制费用进行预测。在费用预测模型中,影响费用的特征参数为费用驱动因子,以往的预测模型^[7-11]大多选取几个相对重要的参数作为费用驱动因子:如重量、最大飞行速度、首飞时间及可靠性等。然而,现代飞机使用要求日益提高使得影响飞机研制费用的因素繁多,特别是先进机载电子设备和复合材料的大量采用,重量等因素已经不能全面反映费用的情况了,采用原有的模型对研制费用进行预测的误差在不断增大,此外,费用驱动因子选择的偏少也不利于费用敏感性分析的实施,因此,必须采用更多的费用驱动因子来建立新的研制费用预测模型。而递阶偏最小二乘回归在处理多变量数据方面具有独特的优势,因此可以采用递阶偏最小二乘回归方法对飞机研制费用进行预测。本文以战斗机机体研制费用预测为例,以国产战斗机为估算样本,采用递阶偏最小二乘回归方法进行建模和验证。

2.1 自变量子块的构建

现在,无论国外还是国内,利用综合参数来构建费用估算参数模型是大势所趋。目前,国外使用的设备费用估算的最新软件 PRICE 尽管输入参数极其繁杂,但追溯到最根本的核心模型,使用的主要说明性变量就是描述设备先进性的综合参数。在估算飞机费用时,对于每一个分系统,上百个输入参数几乎都是为了得到该分系统的综合参数的值,文献^[11]提出了采用效能作为估算费用的综合参数。本文基于文献^[11]的建模思想,把反映效能的战斗机各项能力作为自变量子块引入模型,采用文献^{[12]~[14]}对战斗机作战效能分项能力的描述方法,对费用估算的自变量子

块进行构建,即自变量分组,将反映战斗机研制费用的变量分组:反映机动性的为一组,反映火力的为一组,反映态势感知能力的为一组,反映生存力的为一组,其他的如机体空重、航程等对研制费用有明显驱动的参数为一组,其对应的变量如表1所示。

表1 变量分组
Table 1 Grouping of variables

组	组中变量
机动性(ξ_{mane})	SEP($x_{\text{mane},1}$)、推重比($x_{\text{mane},2}$)和最大瞬时转弯角速度($x_{\text{mane},3}$)
火力(ξ_{fire})	导弹挂架数量($x_{\text{fire},1}$)、最大载弹量($x_{\text{fire},2}$)和挂载武器效能($x_{\text{fire},3}$)
态势感知能力(ξ_{dete})	雷达探测能力系数($x_{\text{dete},1}$)、座舱视场角($x_{\text{dete},2}$)和信息系数($x_{\text{dete},3}$)
生存力(ξ_{sur})	机翼面积($x_{\text{sur},1}$)、雷达反射截面($x_{\text{sur},2}$)和电子对抗能力($x_{\text{sur},3}$)
其他(ξ_{else})	航程($x_{\text{else},1}$)、机体空重($x_{\text{else},2}$)和机体首翻期($x_{\text{else},3}$)

注:SEP—最大单位重量剩余功率

表2 样本数据

Table 2 Data of aircraft samples

机型	机动性			火力			态势感知能力			生存力			其他			研制费用/亿元
	$x_1/(\text{m} \cdot \text{s}^{-1})$	x_2	$x_3/((^\circ) \cdot \text{s}^{-1})$	x_1	x_2/kg	x_3	x_1	$x_2/(\text{^\circ})$	x_3	x_1/m^2	x_2/m^2	x_3	x_1/km	x_2/kg	x_3/h	
1	150	0.900	10	5	1 500	0.60	7	120	0.60	23.00	3.8	1.05	1 692	4 470	1 000	3.05
2	200	0.890	10	6	2 500	0.70	8	120	0.60	42.19	9.0	1.05	2 300	7 430	1 200	6.65
3	200	0.890	10	7	5 000	0.75	10	180	0.70	56.00	10.1	1.10	3 650	12 160	1 500	15.60
4	106	0.710	8	6	1 500	0.30	3	120	0.50	25.00	3.5	1.00	1 100	5 530	800	0.80
5	300	1.050	25	11	6 000	0.90	10	270	0.90	46.50	12.0	1.20	4 000	13 200	2 000	133.00
6	200	0.890	10	5	7 000	0.80	9	180	0.90	56.00	10.1	1.15	3 300	12 160	2 000	41.00
7	135	0.687	10	10	1 500	0.50	4	120	0.60	23.00	3.8	1.00	1 450	3 830	800	0.84
8	235	1.140	25	7	6 000	0.90	10	270	0.90	40.00	8.4	1.20	3 000	12 780	2 000	110.00
9	180	0.800	10	6	2 000	0.65	5	120	0.60	42.19	10.0	1.05	2 200	6 850	1 000	4.00
10	310	1.200	25	9	6 500	0.90	10	360	0.95	51.75	0.2	1.20	3 500	10 000	2 500	80.00
11	300	1.200	25	11	8 000	0.90	10	360	0.95	79.56	0.2	1.20	4 000	12 000	2 500	120.00

归模型中,提取了成分 t_{mane} ,它与原变量的回归关系为

$$t_{\text{mane}} = 0.6048x_{\text{mane},1} + 0.5978x_{\text{mane},2} + 0.5261x_{\text{mane},3} \quad (1)$$

式中: t_{mane} 对费用的解释程度为89.12%。

在以火力为自变量的偏最小二乘回归模型中,提取了成分 t_{fire} ,它与原变量的回归关系为

$$t_{\text{fire}} = 0.1412x_{\text{fire},1} + 0.7232x_{\text{fire},2} + 0.6760x_{\text{fire},3} \quad (2)$$

表1中的变量,有的是基本物理性能的参数,有的是综合评估参数,有的是专家评估参数,计算和评估方法可参见文献[12]~文献[14],对其他类型军用飞机的估算,可以根据飞机的设计特性和使命任务特点,采取不同的性能参数和分组方案,飞机性能数据见表2,取机型1~机型9为样本,机型10和机型11为验证机型。

2.2 模型的建立

(1) 底层模型的建立

考虑到研制费用与解释变量之间是一种指数关系的叠加^[15],为了对模型参数进行估计,采用对数变换,得到线性化的关系式,然后采用递阶偏最小二乘回归建模。相应地,需要对原始数据进行对数变换。

变换后,以研制费用为因变量,分别建立它对机动性、火力、态势感知能力、生存力及其他特性的5个底层偏最小二乘回归模型。根据交叉有效性原则^[4],在以机动性为自变量的偏最小二乘回

式中: t_{fire} 对费用的解释程度为91.51%。

在以态势感知能力为自变量的偏最小二乘回归模型中,提取了成分 t_{dete} ,它与原变量的回归关系为

$$t_{\text{dete}} = 0.5555x_{\text{dete},1} + 0.5816x_{\text{dete},2} + 0.5942x_{\text{dete},3} \quad (3)$$

式中: t_{dete} 对费用的解释程度为95.97%。

在以生存力为自变量的偏最小二乘回归模型中,提取了成分 $t_{\text{sur},1}$ 和 $t_{\text{sur},2}$,它们与原变量的回归关系为

$$\left. \begin{aligned} t_{sur,1} &= 0.5048x_{sur,1} + 0.5382x_{sur,2} + \\ &\quad 0.6749x_{sur,3} \\ t_{sur,2} &= -0.4633x_{sur,1} - 0.3284x_{sur,2} + \\ &\quad 0.8375x_{sur,3} \end{aligned} \right\} \quad (4)$$

式中: $t_{sur,1}$ 和 $t_{sur,2}$ 对费用的解释程度为 85.65%。

在以其他特性为自变量的偏最小二乘回归模型中,提取了成分 t_{else} ,它与原变量的回归关系为

$$t_{else} = 0.5607x_{else,1} + 0.5651x_{else,2} + 0.6051x_{else,3} \quad (5)$$

式中: t_{else} 对费用的解释程度为 93.09%。

(2) 顶层模型的建立

在底层模型的基础上,建立 $[t_{mane} \quad t_{fire} \quad t_{dete} \quad t_{sur,1} \quad t_{sur,2} \quad t_{else}]$ 对机体研制费的顶层模型,得到 2 个偏最小二乘成分为

$$\left. \begin{aligned} t_{Top,1} &= 0.4241t_{mane} + 0.4389t_{fire} + 0.4463t_{dete} + \\ &\quad 0.3747t_{sur,1} - 0.3082t_{sur,2} + 0.4390t_{else} \\ t_{Top,2} &= 0.5056t_{mane} + 0.0838t_{fire} + 0.3573t_{dete} - \\ &\quad 0.4004t_{sur,1} + 0.6724t_{sur,2} + 0.0606t_{else} \end{aligned} \right\} \quad (6)$$

对费用的解释程度为 98.24%,而在底层模型中,各偏最小二乘模型对费用的解释能力分别为 89.12%,91.51%,95.97%,85.65%和 93.09%,从而可以看出,顶层模型利用底层模型提取的成分来解释费用,所提取的偏最小二乘成分的解释能力得到了提高。各变量 $[t_{mane} \quad t_{fire} \quad t_{dete} \quad t_{sur,1} \quad t_{sur,2} \quad t_{else}]$ 的 VIP 为 $[1.0446 \quad 1.0509 \quad 1.0813 \quad 0.9246 \quad 0.8225 \quad 1.0508]$ 。机动性、火力和态势感知能力的 VIP 超过 1,说明它们对研制费用的影响作用较强,引入它们作为费用驱动因子是有意义的。

并且得到顶层偏最小二乘回归模型为

$$\hat{y} = -18.9436 + 1.2527t_{mane} + 0.5820t_{fire} + 0.9970t_{dete} + 0.2333t_{sur,1} + 0.2608t_{sur,2} + 0.5322t_{else} \quad (7)$$

更进一步,将底层模型式(1)~式(5)代入顶层模型式(7),可以得到费用对自变量的线性回归模型为

$$\hat{y} = -18.9436 + 0.7576x_{mane,1} + 0.7488x_{mane,2} + 0.6590x_{mane,3} + 0.0821x_{fire,1} + 0.4209x_{fire,2} + 0.3934x_{fire,3} + 0.5538x_{dete,1} + 0.5798x_{dete,2} + 0.5924x_{dete,3} + 0.2386x_{sur,1} + 0.2112x_{sur,2} - 0.0609x_{sur,3} + 0.2984x_{else,1} + 0.3007x_{else,2} + 0.3220x_{else,3} \quad (8)$$

再对模型进行对数逆变换,得

$$\hat{y} = (5.9279 \times 10^{-9})x_{mane,1}^{0.7576} x_{mane,2}^{0.7488} x_{mane,3}^{0.6590} \cdot x_{fire,1}^{0.0821} x_{fire,2}^{0.4209} x_{fire,3}^{0.3934} x_{dete,1}^{0.5538} x_{dete,2}^{0.5798} x_{dete,3}^{0.5924} \cdot x_{sur,1}^{0.2386} x_{sur,2}^{0.2112} x_{sur,3}^{0.0109} x_{else,1}^{0.2984} x_{else,2}^{0.3007} x_{else,3}^{0.3220} \quad (9)$$

2.3 模型的验证

为检验回归方程,采用两架不同型号的国产战斗机的参数值对模型进行计算验证,军方实际投入的机体研制费用为 80 亿元和 120 亿元,采用本文模型进行预测的结果分别是 88.48 亿元和 116.66 亿元,预测残差率分别为:10.6%和 2.78%。一般来说,对于大型系统的研制费用估算,误差在 30%以内即可以说是有效的^[15],两个验证型号的误差都在 30%以内,说明所建立的模型是可行的。

3 结 论

(1) 从本文所建预测模型可以看出,生存力子块参数上的指数相对较小,雷达反射截面和电子对抗系数的指数接近于零,反映了隐身性设计参数和电子对抗能力对研制费用的驱动不大,甚至是逆向驱动,和实际情况并不相符。分析产生这一不合理现象的原因,主要是因为样本型号的研制中,隐身和电子对抗没有作为一项重要性指标纳入设计,样本数据不能反映隐身性要求和电子对抗对费用的驱动。

(2) 模型中机动性、火力和态势感知能力子块中参数的指数相对较大,表明这些参数是重要的费用驱动因子。同时,机体空重的指数较小,说明对于现代战机重量作为费用驱动因子的作用已经下降了。

(3) 从实例建模过程来看,递阶偏最小二乘回归方法具有较高的信息综合和筛选能力。在底层模型中,每个自变量子块与因变量的偏最小二乘回归可以有效提取该组自变量对因变量的最有用信息,并剔除原始数据中的部分噪声;在信息过滤的基础上,利用这些更加有效的综合变量再建立顶层模型,则所得结果解释能力更强。

参 考 文 献

[1] 尹力, 刘强, 王惠文. 偏最小二乘相关算法在系统建模中的两类型应用[J]. 系统仿真学报, 2003, 15(1): 135-137.
Yin Li, Liu Qiang, Wang Huiwen. Two typical applications of the related partial least square algorithms in the field of system modeling[J]. Journal of System Simula-

- tion, 2003, 15(1): 135-137. (in Chinese)
- [2] 郭基联, 张恒喜, 刘晓东. 飞机机体研制费用的偏最小二乘回归分析[J]. 数量经济技术经济研究, 2001(4): 121-124.
Guo Jilian, Zhang Hengxi, Liu Xiaodong. The partial least squares regression analyses of the aircrafts' fuselage development cost [J]. Quantitative & Technical Economics, 2001(4): 121-124. (in Chinese)
- [3] 李寿安, 张恒喜, 童中翔, 等. 偏最小二乘回归在军用飞机价格预测中的应用[J]. 航空学报, 2006, 27(3): 600-604.
Li Shouan, Zhang Hengxi, Tong Zhongxiang, et al. Application of partial least squares regression to the acquisition cost estimating of military aircrafts [J]. Acta Aeronautica et Astronautica Sinica, 2006, 27(3): 600-604. (in Chinese)
- [4] 王慧文, 吴载斌, 孟洁. 偏最小二乘回归的线性与非线性方法[M]. 北京: 国防工业出版社, 2006: 162-164.
Wang Huiwen, Wu Zaibin, Meng Jie. Partial least squares regression—linear and nonlinear methods [M]. Beijing: National Defence Industry Press, 2006: 162-164. (in Chinese)
- [5] Word S, Word K, Tjessem K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection[J]. Journal of Chemometrics, 1996, 10(5/6): 463-482.
- [6] Johan A W, Theodora K, John F M. Analysis of multiblock and hierarchical PCA and PLS models[J]. Journal of Chemometrics, 1998, 12(5): 301-321.
- [7] 张恒喜. 现代飞机效费分析[M]. 北京: 航空工业出版社, 2001.
Zhang Hengxi. Modern aircraft effectiveness-cost analysis [M]. Beijing: Aviation Industry Press, 2001. (in Chinese)
- [8] 徐哲, 刘荣. 偏最小二乘回归法在武器装备研制费用估算中的应用[J]. 数学的实践与认识, 2005, 35(3): 152-158.
Xu Zhe, Liu Rong. The application and research of PLS in estimating the cost of development in armament [J]. Mathematics in Practice and Theory, 2005, 35(3): 152-158. (in Chinese)
- [9] 金锋, 崔卉, 张玉平, 等. 武器装备研制阶段费用预测方法及仿真[J]. 北京理工大学学报, 2006, 26(11): 991-994.
Jin Feng, Cui Hui, Zhang Yuping, et al. Forecast and emulation research on the cost of development for military equipment [J]. Transactions of Beijing Institute of Technology, 2006, 26(11): 991-994. (in Chinese)
- [10] 陆凯. 无人战斗机系统及其全寿命周期费用研究[D]. 西安: 西北工业大学, 2002.
Lu Kai. Uninhabited combat air vehicle and research on it's life cycle cost [D]. Xi'an: Northwestern Polytechnical University, 2002. (in Chinese)
- [11] 周健, 黄栋, 徐宗昌. 无人机效能指数/费用模型初探[J]. 装甲兵工程学院学报, 2002, 16(3): 73-76.
Zhou Jian, Huang Dong, Xu Zongchang. A study on the efficiency-exponent/expenses model of unmanned aerial vehicle [J]. Journal of Academy of Armored Force Engineering, 2002, 16(3): 73-76. (in Chinese)
- [12] 朱宝莲, 朱荣昌, 熊笑非. 作战飞机效能评估[M]. 北京: 航空工业出版社, 1993.
Zhu Baoliu, Zhu Rongchang, Xiong Xiaofei. Fighter plane effectiveness assessment [M]. Beijing: Aviation Industry Press, 1993. (in Chinese)
- [13] 董彦非, 王礼沅, 张恒喜. 战斗机空战效能评估的综合指数模型[J]. 航空学报, 2006, 27(6): 1084-1087.
Dong Yanfei, Wang Liyuan, Zhang Hengxi. Synthesized index model for fighter plane air combat effectiveness assessment [J]. Acta Aeronautica et Astronautica Sinica, 2006, 27(6): 1084-1087. (in Chinese)
- [14] 王礼沅, 张恒喜, 徐浩军. 基于粗糙集的空战效能多指标综合评估模型[J]. 航空学报, 2008, 29(4): 880-885.
Wang Liyuan, Zhang Hengxi, Xu Haojun. A multi-indexes synthesize evaluation model based on rough set theory for aerial warfare efficiency [J]. Acta Aeronautica et Astronautica Sinica, 2008, 29(4): 880-885. (in Chinese)
- [15] 飞机设计手册编委会. 技术经济设计[M]. 北京: 航空工业出版社, 2001: 6-7.
The Compile Committee of Airplane Design Handbook. The technical economy design [M]. Beijing: Aviation Industry Press, 2001: 6-7. (in Chinese)

作者简介:

王礼沅(1980—) 男, 博士研究生。主要研究方向: 飞机总体设计与综合论证、作战效能评估。

Tel: 13661147247

E-mail: wangly1223@sohu.com

郭基联(1971—) 男, 博士, 副教授, 硕士生导师。主要研究方向: 飞机全寿命周期费用分析。

Tel: 029-84787482

E-mail: guo_jl71@yahoo.com.cn

张恒喜(1937—) 男, 教授, 博士生导师。主要研究方向: 军用飞机型号发展工程、效费分析和综合保障工程。

Tel: 029-84787296

(编辑: 徐晓)