

## 基于特征差异的仿真模型验证及选择方法

李伟<sup>1</sup> 焦松<sup>1</sup> 陆凌云<sup>1</sup> 杨明<sup>1</sup>

**摘要** 为了在系统存在多个数据类型各异的输出时, 多个备选仿真模型的验证和择优, 提出了基于特征差异的仿真模型验证及选择方法. 首先, 将系统输出分为静态、缓变和速变三类数据, 并分别给出了每类数据的特征差异度量模型; 然后, 采用主成分分析法从多个具有相关性的特征差异中提取出少数几个独立的主成分变量; 再者依据主成分数据, 采用  $K$ -均值聚类分析方法将多个备选仿真模型的输出划分为  $K$  类; 最后, 基于 Fisher 判别分析法判定参考输出是否属于其中的某类, 进而实现对多个备选仿真模型的验证和选择. 通过实例应用, 表明了该方法的有效性.

**关键词** 模型验证, 模型选择, 特征差异, 主成分分析, 聚类分析, 判别分析

**引用格式** 李伟, 焦松, 陆凌云, 杨明. 基于特征差异的仿真模型验证及选择方法. 自动化学报, 2014, 40(10): 2134–2144

**DOI** 10.3724/SP.J.1004.2014.02134

### Validation and Selection of Simulation Model Based on the Feature Differences

LI Wei<sup>1</sup> JIAO Song<sup>1</sup> LU Ling-Yun<sup>1</sup> YANG Ming<sup>1</sup>

**Abstract** To validate alternative models and select the most credible one when the models have many outputs of different kinds, a validation and selection method of simulation model based on feature differences is proposed. The outputs are divided into three kinds: static data, gradual data and fast data, and the measure models of feature differences for each kind data are given. The correlation among the feature differences is eliminated via principal component analysis, and several independent principal components are gained. Furthermore, the outputs of simulation models are divided into  $K$  kinds of clusters based on  $K$ -means clustering according to the principal components. Which cluster the output of actual system belongs to is judged by Fisher discriminant analysis. So the validation and selection of alternative models are realized. Finally, the method is validated in an application.

**Key words** Validation of simulation model, selection of simulation model, feature difference, principal component analysis, clustering analysis, discriminant analysis

**Citation** Li Wei, Jiao Song, Lu Ling-Yun, Yang Ming. Validation and selection of simulation model based on the feature differences. *Acta Automatica Sinica*, 2014, 40(10): 2134–2144

随着仿真技术的发展, 其在国防、经济、材料以及化工等众多领域得到了广泛的应用. 由于仿真是一种基于模型的活动, 因此模型的可信性成为模型用户关注的焦点之一<sup>[1–3]</sup>. 在仿真活动中, 通常存在多个仿真模型可供选择. 此时, 就需要验证它们的可信性, 并从中选出最可信者.

通过比较分析仿真输出与参考输出之间的特征差异是常用的仿真模型验证方法<sup>[4]</sup>. 1967 年, Naylor 等指出可根据直觉对仿真模型进行表面验证, 还可以从二者的典型事件、假设和其他内在特

征等不同角度分析仿真模型和真实对象之间的一致性, 由此提出了很多基于特征的验证方法, 如方差分析和一些非参数检验, 包括  $\chi^2$  检验, Kolmogorov-Smirnov 检验等<sup>[5]</sup>. 1978 年, Kheir 等利用 Theil 不等式系数 (Theil inequality coefficient, TIC) 对导弹仿真系统有效性进行了验证<sup>[6]</sup>. 文献 [7] 分析了灰色关联度模型的特点, 并将其应用于导弹仿真系统可信度评估工作中. 文献 [8] 采用误差分析方法评估了计量经济模型的可信度. 文献 [9] 论述了可信度与相似度之间的联系, 并基于相似度给出了仿真模型可信度的量化方法. 文献 [10] 运用经典的傅里叶方法估计频谱, 通过频谱分析实现对导弹系统仿真模型可信度的评估, 其他频谱分析方法还有窗谱<sup>[11]</sup>、最大熵谱<sup>[12]</sup> 和交叉谱<sup>[13]</sup> 等. 此外, Sargent、Balci 以及 Kleijnen 通过对前人工作的研究和总结, 逐步形成了较为统一的理论和方法体系<sup>[1, 14–15]</sup>.

按与时间是否相关进行分类, 可以将系统输出分为静态数据和动态数据两类<sup>[11]</sup>. 对于动态数据, 按随时间变化频率的大小进行分类, 又可以分为缓

收稿日期 2013-11-13 录用日期 2014-04-18  
Manuscript received November 13, 2013; accepted April 18, 2014

国家自然科学基金 (61273226), 中央高校基本科研业务费专项资金 (HIT.NSRIF.2015035) 资助

Supported by National Natural Science Foundation of China (61273226) and Fundamental Research Funds for the Central Universities (HIT.NSRIF.2015035)

本文责任编辑 陈杰

Recommended by Associate Editor CHEN Jie

1. 哈尔滨工业大学控制与仿真中心 哈尔滨 150080

1. Control and Simulation Center, Harbin Institute of Technology, Harbin 150080

变数据和速变数据两类. 例如: 导弹的脱靶量为静态数据, 位姿状态为缓变数据, 振动信号为速变数据. 在上述方法中, 方差分析、 $\chi^2$  检验以及 Kolmogorov-Smirnov 检验等能较好地解决系统输出为静态数据的仿真模型验证问题; TIC 法、灰色关联度分析法、误差分析法以及相似度法等能较好地解决系统输出为缓变数据的仿真模型验证问题; 窗谱分析法、最大熵谱分析法以及交叉谱分析法能较好地解决系统输出为速变数据的仿真模型验证问题. 但是这些经典方法未能充分考虑仿真输出与参考输出之间的差异, 且仅能适用于某种特定的数据类型. 此外, 当仿真输出存在多个分量时, 一般处理方法是多个仿真输出与参考输出之间的差异结合仿真应用领域知识转化为统一的评估指标后, 综合得到验证及选择结果. 由此可知, 结果中存在较强的主观性. 为了解决参考输出已知时的多个仿真模型验证和选择问题, 在经典方法的基础上还需要从两方面进行改进: 1) 尽可能充分地刻画仿真输出和参考输出之间的差异; 2) 尽可能减少验证及选择结果的主观性.

针对上述问题, 本文提出了基于特征差异的仿真模型验证方法, 首先, 针对不同类型的数据, 给出各自的特征差异度量模型; 然后, 基于主成分分析从多个相关的数据特征差异中提取出少数几个独立的有代表性的主成分; 再者基于  $K$ -均值聚类将多个待选仿真模型输出分为  $k$  类; 最后, 基于 Fisher 判别分析判定参考输出是否属于  $k$  类中的某一类, 若是, 则该类中所有元素对应的仿真模型均认为可信, 且离参考输出“最近者”对应的仿真模型视为最可信者.

## 1 问题描述及分析

用  $S$  表示系统,  $I = \{i_1, i_2, \dots, i_n\}$  和  $O = \{o_1, o_2, \dots, o_m\}$  分别表示系统的输入和输出. 用  $S_s$  和  $S_r$  分别表示仿真系统和参考系统,  $I_s = \{i_{s1}, i_{s2}, \dots, i_{sn}\}$  和  $I_r = \{i_{r1}, i_{r2}, \dots, i_{rn}\}$  分别表示仿真系统和参考系统的输入,  $O_s = \{O_{s1}, O_{s2}, \dots, O_{sm}\}$  和  $O_r = \{O_{r1}, O_{r2}, \dots, O_{rm}\}$  分别表示仿真系统和参考系统的输出.

用  $C(O_s, O_r)$  表示在  $I_r = I_s$  时,  $O_s$  相对于  $O_r$  的一致性程度, 且  $C(O_s, O_r) \in (0, 1]$ . 当  $O_s$  与  $O_r$  完全一致, 即仿真模型完全可信, 则有  $C(O_s, O_r) = 1$ ; 当  $O_s$  相对于  $O_r$  一致性程度越差, 即仿真模型越不可信时, 则有  $C(O_s, O_r) \rightarrow 0$ .

针对参考系统  $S_r$ , 假定存在  $p$  个备选的仿真模型  $S_i, i = 1, 2, \dots, p$ , 当它们的输入与参考输入一致性时, 输出集合为  $O_{si}, i = 1, 2, \dots, p$ . 现需要验证  $p$  个仿真模型的可信性, 进而确定最可信者. 解决

该问题的一般方法如下:

- 1) 获取仿真输出与参考输出之间的差异

$$e_{ij} = F_{ij}(O_{si}, O_{ri}) \quad (1)$$

其中,  $i = 1, 2, \dots, m, j = 1, 2, \dots, a_i; e_{ij}$  为仿真输出和参考输出第  $i$  分量的第  $j$  个数据特征差异;  $F_{ij}(\cdot)$  为特征差异度量模型.

- 2) 确定仿真输出与参考输出的一致性程度

$$C(O_s, O_r) = G(e_{11}, e_{12}, \dots, e_{1a_1}, \dots, e_{m1}, e_{m2}, \dots, e_{ma_m}) \quad (2)$$

其中,  $G(\cdot)$  为一致性程度评估模型.

- 3) 模型的验证和选择

$$T_i = \text{sgn}(C_i(O_s, O_r) - \xi) \quad (3)$$

其中,  $i = 1, 2, \dots, p; \xi > 0$ , 为可信性判断阈值;  $\text{sgn}(\cdot)$  为符号函数;  $T_i$  为判断结论, 当  $T_i \geq 0$  代表  $S_i$  可信, 反之不可信.

$$C_r(O_s, O_r) = \max_{i=1,2,\dots,p} \{C_i(O_s, O_r)\} \quad (4)$$

其中,  $C_r(O_s, O_r)$  表示一致性程度最大者, 其对应的仿真模型  $S_r$  为最可信者.

如果采用经典方法解决多个备选仿真模型的验证及选择问题, 则存在如下几个问题需要进一步解决:

- 1) 仿真输出与参考输出之间差异的刻画是仿真模型验证的基础, 需要尽可能充分地刻画二者之间的差异.

- 2) 特征差异  $e_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, a_i$  的含义各异. 式 (2) 利用它们得到  $C(O_s, O_r)$  时, 需要结合仿真应用领域知识将它们转换为统一的评估指标后再进行综合. 由此,  $C(O_s, O_r)$  的主观性较强.

- 3) 式 (3) 中仿真模型可信性阈值的设定也往往具有较强的主观性.

针对上述第一个问题, 本文将仿真/参考输出分为静态、缓变以及速变数据三类, 然后从多个视角分别给出二者之间的差异. 为了降低验证和选择结果的主观性, 本文给出了一种新的解决多个备选仿真模型验证及选择问题的方法. 在此, 利用  $e_{ij}, i = 1, 2, \dots, m, j = 1, 2, \dots, a_i$  定义数据特征差异向量如下:

$$\mathbf{E} = [e_{ij}]_{1 \times (a_1 + a_2 + \dots + a_m)} \quad (5)$$

令  $a_1 + a_2 + \dots + a_m = r$ , 表示特征差异向量的个数. 用  $\mathbf{E}_j = [e_i^j]_{1 \times r}, j = 1, 2, \dots, p$  表示  $p$  个待选仿真模型输出与参考输出之间的数据特征差异.

定义“最理想特征差异向量” $\mathbf{E}^* = [e_i^*]_{1 \times r} = 0$ , 其表示仿真输出与参考输出的所有数据特征均相等. 本文将通过分析 $\mathbf{E}_i, i = 1, 2, \dots, p$ 与 $\mathbf{E}^*$ 之间的关系, 实现对 $p$ 个待选仿真模型的验证和选择.

## 2 特征差异度量模型

对仿真输出和参考输出之间差异的刻画是仿真模型验证和选择工作的基础. 经典的仿真模型验证方法仅从单个或者少数几个方面刻画仿真输出与参考输出之间的差异. 例如: 假设检验法仅关注静态数据均值或者方差之间的差异; TIC系数法、误差分析法仅关注缓变数据之间的距离差异; 灰色关联系数法以及相似度法仅关注缓变数据外形之间的差异; 谱分析法仅关注速变数据平稳项谱密度之间的差异. 本节依据数据特征先将仿真输出和参考输出分为静态、缓变和速变三类数据, 然后分别给出每类数据的特征差异度量模型, 进而充分地刻画仿真输出和参考输出之间的差异.

考虑到在一些特殊的仿真应用中, 需要格外关注某些物理量, 例如: 控制系统响应中的上升时间、超调量以及稳态误差, 位置数据中的变化趋势, 测量数据中的噪声等. 此时, 度量仿真输出和参考输出之间差异时, 就需要先依据具体的领域知识将这些物理量从数据中提取出来; 然后, 再根据它们的数据类型, 采用下文中的方法刻画各自的差异; 最后, 与原始仿真数据与参考数据之间的差异一起构造得到数据特征差异向量, 用以支持仿真模型验证和选择工作.

### 2.1 仿真/参考输出分类

系统输出的数据类型不同, 所蕴含的特征也不同. 因此, 在刻画仿真输出与参考输出之间的差异之前, 需要对二者进行分类.

假定 $X$ 为一个随机变量,  $SX$ 为其对应的样本空间,  $X(i), i = 1, 2, \dots, N$ 表示 $X$ 的第 $i$ 次观察值. 将 $X(i), i = 1, 2, \dots, N$ 构成的集合 $\{X(1), X(2), \dots, X(N)\}$ 称为静态数据; 若 $X(i), i = 1, 2, \dots, N$ 为按时间先后依次观察得到的数据, 将 $X(i), i = 1, 2, \dots, N$ 构成的时间序列 $\langle X(1), X(2), \dots, X(N) \rangle$ 称为动态数据. 进一步, 假定动态数据 $\langle X(1), X(2), \dots, X(N) \rangle$ 对应的时间变化序列为 $\langle T(1), T(2), \dots, T(N) \rangle$ , 由此定义 $\langle X(1), X(2), \dots, X(N) \rangle$ 随时间的变化频率如下:

$$F = \frac{\sum_{i=1}^{N-1} \left| \frac{\Delta X(i)}{\Delta T(i)} \right|}{|\bar{X}|} \quad (6)$$

其中,  $F \geq 0$ 为 $\langle X(1), X(2), \dots, X(N) \rangle$ 的数据

变化频率;  $\Delta X(i) = X(i+1) - X(i)$ ;  $\Delta T(i) = T(i+1) - T(i)$ ;  $|\bar{X}| = \sum_{i=1}^N |X(i)|/N \neq 0$ .

由此, 假定 $f_0 \geq 0$ 为评定数据变化快慢的临界值. 如果 $F \geq f_0$ , 则认为 $\langle X(1), X(2), \dots, X(N) \rangle$ 为速变数据; 否则, 认为 $\langle X(1), X(2), \dots, X(N) \rangle$ 为缓变数据.  $f_0$ 可以由专家依据具体的领域设定.

### 2.2 静态数据特征差异度量模型

假定参考输出 $X_r$ 和仿真输出 $X_s$ 均为静态数据, 分别由相应的系统运行 $N$ 次和 $M$ 次后得到的输出集合为

$$X_r = \{x_r(1), x_r(2), \dots, x_r(N)\} \quad (7)$$

$$X_s = \{x_s(1), x_s(2), \dots, x_s(M)\} \quad (8)$$

其中,  $x_r(i), i = 1, 2, \dots, N$ 表示参考系统第 $i$ 次运行的输出;  $x_s(j), j = 1, 2, \dots, M$ 表示仿真系统第 $j$ 次运行的输出.

对于静态数据而言, 样本数据的集中和分散趋势是最主要的两个特征, 可分别采用均值和方差进行刻画. 由此, 得到 $X_s$ 相对 $X_r$ 的集中和分散特征差异如下:

$$e_c = |\bar{x}_s - \bar{x}_r| = \left| \frac{\sum_{i=1}^M x_s(i)}{M} - \frac{\sum_{i=1}^N x_r(i)}{N} \right| \quad (9)$$

$$e_d = |s_s - s_r| = \left| \sqrt{\frac{\sum_{i=1}^M [x_s(i) - \bar{x}_s]^2}{M-1}} - \sqrt{\frac{\sum_{i=1}^N [x_r(i) - \bar{x}_r]^2}{N-1}} \right| \quad (10)$$

其中,  $e_c$ 表示集中趋势特征差异;  $e_d$ 表示离散趋势特征差异.

### 2.3 缓变数据特征差异度量模型

假定参考输出 $X_r$ 和仿真输出 $X_s$ 为缓变数据, 分别由相应的系统运行 $N$ 次和 $M$ 次后得到的时间序列为

$$X_{ri} = \langle x_{ri}(1), x_{ri}(2), \dots, x_{ri}(T) \rangle \quad (11)$$

$$X_{sj} = \langle x_{sj}(1), x_{sj}(2), \dots, x_{sj}(T) \rangle \quad (12)$$

其中,  $i = 1, 2, \dots, N, j = 1, 2, \dots, M$ .

对多次运行输出在每个采样点进行平均处理后, 得到的参考输出和仿真输出如下:

$$X_r = \langle x_r(1), x_r(2), \dots, x_r(T) \rangle \quad (13)$$

$$X_S = \langle x_s(1), x_s(2), \dots, x_s(T) \rangle \quad (14)$$

其中,  $x_r(j) = \sum_{i=1}^N |X_{ri}(j)|/N$  和  $x_s(j) = \sum_{i=1}^M |X_{si}(j)|/M$ ,  $j = 1, 2, \dots, T$  分别表示参考系统和仿真系统在第  $j$  时刻的输出.

在此需要进一步说明,  $X_r$  和  $X_s$  的采样时刻可能会不一致, 采样点个数也可能不同. 此时, 需要以二者之一为基准, 采用插值的方法对另一个进行时间一致性处理, 具体方法见文献 [16]. 本文所处理的数据假定都是经过时间一致性处理后的数据.

在此, 令  $Z = \langle z(1), z(2), \dots, z(T) \rangle = \langle x_s(1) - x_r(1), x_s(2) - x_r(2), \dots, x_s(T) - x_r(T) \rangle$ . 如果  $z(1), z(2), \dots, z(T)$  彼此间相差不大, 则表明  $X_r$  与  $X_s$  的外形较相似, 特别当  $X_r$  与  $X_s$  重合或平行时, 有  $z(1) = z(2) = \dots = z(T)$ ; 如果  $z(1), z(2), \dots, z(T)$  彼此间相差较大, 则表明  $X_r$  与  $X_s$  外形的相似性较差. 由此, 从位置差异和外形差异两方面刻画  $X_r$  与  $X_s$  之间的差异<sup>[6, 10]</sup>, 定义如下:

$$e_s = \frac{1}{T} \sqrt{\sum_{i=1}^T z(i)^2} \quad (15)$$

$$e_t = \frac{1}{T} \sqrt{\sum_{i=1}^T [z(i) - \bar{z}]^2} \quad (16)$$

其中,  $z(i) = x_s(i) - x_r(i)$ ,  $i = 1, 2, \dots, T$ ;  $\bar{z} = \sum_{i=1}^T z(i) / T$ .

#### 2.4 速变数据特征差异度量模型

如果参考输出  $X_r$  和仿真输出  $X_s$  为速变数据, 与缓变数据处理方式相同, 得到:

$$X_r = \langle x_r(1), x_r(2), \dots, x_r(T) \rangle \quad (17)$$

$$X_s = \langle x_s(1), x_s(2), \dots, x_s(T) \rangle \quad (18)$$

其中,  $x_r(i)$  和  $x_s(i)$ ,  $i = 1, 2, \dots, T$  分别表示参考系统和仿真系统在第  $i$  时刻的输出.

由于速变数据一方面包含反映波动特征的平稳项, 另一方面包含反映数据整体变化趋势特征的趋势项. 基于经验模态分解算法可将二者分离, 具体算法见文献 [17-18]. 由此给出如下定义:

$$X_r = X_{rc} + X_{rf}, \quad X_s = X_{sc} + X_{sf} \quad (19)$$

其中,  $X_{rc}$  和  $X_{sc}$  分别为  $X_r$  和  $X_s$  的趋势项;  $X_{rf}$  和  $X_{sf}$  分别为  $X_r$  和  $X_s$  的平稳项.

$X_{rc}$  和  $X_{sc}$  为缓变数据, 可以直接依据式 (15) 和 (16) 计算  $e_s$  和  $e_t$  来刻画二者的差异.  $X_{rf}$  和  $X_{sf}$  可以通过窗谱分析转换至频域中得到各自的谱

密度, 进而通过谱密度差异刻画  $X_{rf}$  和  $X_{sf}$  之间的差异, 具体原理见文献 [10]. 由此定义如下:

$$e_f = 1 - \frac{m}{M} \quad (20)$$

式中,  $e_f$  表示  $X_{rf}$  和  $X_{sf}$  的谱密度差异;  $M$  表示  $X_{rf}$  和  $X_{sf}$  转换至频域中的点数;  $m$  表示通过相容性检验的点数.

### 3 仿真模型验证方法

#### 3.1 基于主成分分析的特征差异向量构造方法

依据  $O_s$  和  $O_r$  各分量的不同数据类型, 分别采用式 (9)、(10)、(15)、(16) 以及式 (20) 得到数据特征差异向量  $\mathbf{E} = [e_{ij}]_{1 \times (a_1 + a_2 + \dots + a_m)}$ .

由于  $\mathbf{E}$  中的  $r$  个数据特征差异之间存在相关性, 也就是说它们所蕴含的信息存在一定的“重叠”. 在此, 采用主成分分析法从  $r$  个数据特征差异中提取少数几个相互独立且包含了大部分信息的主成分<sup>[19]</sup>. 首先对  $\mathbf{E}_j$ ,  $j = 1, 2, \dots, p$  及  $\mathbf{E}^*$  进行标准化处理如下:

$$\tilde{e}_i^j = \frac{e_i^j - \bar{e}_i}{s_i} \quad (21)$$

$$\tilde{e}_i^* = \frac{0 - \bar{e}_i}{s_i} \quad (22)$$

其中,  $i = 1, 2, \dots, r$ ,  $j = 1, 2, \dots, p$ ;  $\tilde{e}_i^j$  为  $e_i^j$  经过标准化处理的结果;  $\tilde{e}_i^*$  为  $e_i^*$  经过标准化处理的结果;  $\bar{e}_i = \sum_{j=1}^p e_i^j / p$ ;  $s_i = \sqrt{\sum_{j=1}^p (e_i^j - \bar{e}_i)^2 / p}$ .

由此, 得到  $\mathbf{E}_j$ ,  $j = 1, 2, \dots, p$  及  $\mathbf{E}^*$  进行标准化处理后的结果为  $\tilde{\mathbf{E}}_j = [\tilde{e}_i^j]_{1 \times r}$ ,  $j = 1, 2, \dots, p$  和  $\tilde{\mathbf{E}}^* = [\tilde{e}_i^*]_{1 \times r}$ . 进一步, 构造协方差阵为

$$A = [a_{ij}]_{r \times r} \quad (23)$$

其中,  $A$  为协方差阵,  $a_{ij} = [\sum_{t=1}^p (\tilde{e}_i^t - \bar{\tilde{e}}_i)(\tilde{e}_j^t - \bar{\tilde{e}}_j) + (\tilde{e}_i^* - \bar{\tilde{e}}_i)(\tilde{e}_j^* - \bar{\tilde{e}}_j)] / (p + 1)$ ;  $\bar{\tilde{e}}_i = (\sum_{j=1}^p \tilde{e}_i^j + \tilde{e}_i^*) / (p + 1)$ .

假定  $A$  的特征值为  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r \geq 0$ . 对应的单位特征向量为  $\mathbf{t}_i$ ,  $i = 1, 2, \dots, r$ . 主成分分析的具体原理见文献 [20]. 由此知, 如果:

$$\frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^r \lambda_i} \geq \zeta \quad (24)$$

其中,  $\zeta \in [0, 1]$  为主成分判定阈值.

$\mathbf{E}_j, j = 1, 2, \dots, p$  及  $\mathbf{E}^*$  经过主成分分析后得到的新的特征差异向量如下:

$$\hat{\mathbf{E}}_j = [\tilde{\mathbf{E}}_j \mathbf{t}_1, \tilde{\mathbf{E}}_j \mathbf{t}_2, \dots, \tilde{\mathbf{E}}_j \mathbf{t}_q] = [\hat{e}_j^i]_{1 \times q} \quad (25)$$

$$\hat{\mathbf{E}}^* = [\tilde{\mathbf{E}}^* \mathbf{t}_1, \tilde{\mathbf{E}}^* \mathbf{t}_2, \dots, \tilde{\mathbf{E}}^* \mathbf{t}_q] = [\hat{e}_i^*]_{1 \times q} \quad (26)$$

其中,  $j = 1, 2, \dots, p$ .

### 3.2 基于 $K$ 均值的仿真输出聚类算法

$\hat{\mathbf{E}}_i$  和  $\hat{\mathbf{E}}_j$  之间的距离定义如下:

$$d(\hat{\mathbf{E}}_i, \hat{\mathbf{E}}_j) = \sqrt{\sum_{t=1}^q (\hat{e}_t^i - \hat{e}_t^j)^2} \quad (27)$$

假设  $\hat{\mathbf{E}}_i, i = 1, 2, \dots, p$  经过聚类分析后, 得到  $k$  类输出集合分别为  $C_i, i = 1, 2, \dots, k$ , 且满足如下条件:

$$C_1 \cup C_2 \cup \dots \cup C_k = \{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \dots, \hat{\mathbf{E}}_p\} \quad (28)$$

$$C_i \cap C_j = \emptyset, \quad i \neq j \quad (29)$$

令  $C_i = \{\psi_1^i, \psi_2^i, \dots, \psi_{n_i}^i\}, i = 1, 2, \dots, k$ , 定义其中心为

$$\psi_0^i = \frac{1}{n_i} \sum_{j=1}^{n_i} \psi_j^i \quad (30)$$

由于  $K$  均值聚类分析算法原理简单、易于实现. 本文将采用该方法对  $\hat{\mathbf{E}}_i, i = 1, 2, \dots, p$  进行分类, 具体算法见文献 [21]. 若聚类数目为  $k$ , 则基于  $K$  均值的聚类目标是使所有待分类的元素到其所在类中心的距离之和最小, 如下所示:

$$J = \min \sum_{i=1}^k \sum_{j=1}^{n_i} d(\psi_j^i, \psi_0^i) \quad (31)$$

算法实现过程中存在如何确定初始点以及聚类数目  $k$  的问题, 本文对这两个问题的处理如下:

#### 1) 初始点的确定

依据  $\hat{\mathbf{E}}_i, i = 1, 2, \dots, p$  得<sup>[22]</sup>:

$$\tilde{e}_j^0 = \frac{1}{p} \sum_{i=1}^p \tilde{e}_j^i, \tilde{\delta}_j = \sqrt{\frac{\sum_{i=1}^p (\tilde{e}_j^i - \tilde{e}_j^0)^2}{p-1}} \quad (32)$$

$$u_j^i = (\tilde{e}_j^i - \tilde{e}_j^0) + \frac{2\tilde{\delta}_j \times (i-1)}{k-1} \quad (33)$$

其中,  $i = 1, 2, \dots, k, j = 1, 2, \dots, q$ .

进一步, 构造得到的  $k$  个向量为

$$\mathbf{U}_i = [u_1^i, u_2^i, \dots, u_q^i] \quad (34)$$

其中,  $i = 1, 2, \dots, k$ .

假定已从  $\hat{\mathbf{E}}_i, i = 1, 2, \dots, p$  中确定与  $\mathbf{U}_j, j = 1, 2, \dots, i-1$  距离最近的元素为  $\psi_0^j, j = 1, 2, \dots, i-1$ . 现从  $\hat{\mathbf{E}}_i, i = 1, 2, \dots, p$  中确定与  $\mathbf{U}_j$  距离最近且与  $\psi_0^j, j = 1, 2, \dots, i-1$  相异的元素.

$$\begin{cases} d(\psi_0^i, \mathbf{U}_i) = \min_{j=1,2,\dots,p} \{d(\hat{\mathbf{E}}_j, \mathbf{U}_i)\} \\ \psi_0^i \in \{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_2, \dots, \hat{\mathbf{E}}_p\} \\ \psi_0^i \notin \{\psi_0^1, \psi_0^2, \dots, \psi_0^{i-1}\} \end{cases} \quad (35)$$

其中,  $i = 1, 2, \dots, k$ .

#### 2) 聚类数目 $k$ 的确定

为了确定聚类数目  $k$ , 本文引入 Calinski Harabasz (CH) 准则如下<sup>[23]</sup>:

$$\psi_0^i = \frac{1}{n_i} \sum_{j=1}^{n_i} \psi_j^i, \quad i = 1, 2, \dots, k \quad (36)$$

$$\psi_0 = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} \psi_j^i}{n_1 + n_2 + \dots + n_k} \quad (37)$$

$$TB = \sum_{i=1}^k n_i d(\psi_0^i, \psi_0), \quad TW = \sum_{i=1}^k \sum_{j=1}^{n_i} d(\psi_j^i, \psi_0^i) \quad (38)$$

$$CH = \frac{\frac{TB}{(k-1)}}{\frac{TW}{(n-k)}} \quad (39)$$

$TB$  和  $TW$  分别刻画了类间分离性和类内紧致性, 并认为类内部样本间越接近, 即  $TW$  越小, 且类与类之间相隔越远, 即  $TB$  越大, 则这样的聚类结果则越佳. 由此可知, 对于聚类数目  $1 < k < n$  而言, 当所得的  $CH$  最大时, 则此时的聚类结果最佳.

### 3.3 基于 Fisher 的参考输出判别算法

针对由聚类分析得到的  $k$  个类  $C_i, i = 1, 2, \dots, k$ , 现需要判定“最理想特征差异向量” $\hat{\mathbf{E}}^*$  是否属于其中的某类. 若存在, 则认为该类中元素所对应的仿真模型均有效. 本文采用 Fisher 判别分析方法判定  $\hat{\mathbf{E}}^*$  所属的类, 其基本思想是通过投影变换, 使得各类在投影之后, 相互之间

尽可能分开, 具体算法见文献 [24–25]. 投影方向  $\mathbf{a} = [a_1, a_2, \dots, a_q]^T$  的确定方法如下:

$$H = \sum_{i=1}^k n_i (\boldsymbol{\psi}_0^i - \boldsymbol{\psi}_0) (\boldsymbol{\psi}_0^i - \boldsymbol{\psi}_0)^T \quad (40)$$

$$M = \sum_{i=1}^k \sum_{j=1}^{n_i} n_i (\boldsymbol{\psi}_j^i - \boldsymbol{\psi}_0^i) (\boldsymbol{\psi}_j^i - \boldsymbol{\psi}_0^i)^T \quad (41)$$

假定  $M^{-1}H$  的非零特征值为  $r_1 \geq r_2 \geq \dots \geq r_s \geq 0$ , 且  $s \leq \min(k-1, \nu)$ . 对应的特征向量为  $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots, \boldsymbol{\eta}_s$ . 如果:

$$\frac{\sum_{l=1}^l r_l}{\sum_{j=1}^s r_j} \geq \delta \quad (42)$$

其中,  $\delta$  为设定的累积贡献率阈值.

则  $C_i = \{\boldsymbol{\psi}_1^i, \boldsymbol{\psi}_2^i, \dots, \boldsymbol{\psi}_{n_i}^i\}$ ,  $i = 1, 2, \dots, k$  以及  $\hat{\mathbf{E}}^*$  得到投影转换后的结果如下:

$$\boldsymbol{\varphi}_j^i = [\boldsymbol{\psi}_j^i \boldsymbol{\eta}_1, \boldsymbol{\psi}_j^i \boldsymbol{\eta}_2, \dots, \boldsymbol{\psi}_j^i \boldsymbol{\eta}_l] \quad (43)$$

$$\boldsymbol{\Phi} = [\hat{\mathbf{E}}^* \boldsymbol{\eta}_1, \hat{\mathbf{E}}^* \boldsymbol{\eta}_2, \dots, \hat{\mathbf{E}}^* \boldsymbol{\eta}_l] \quad (44)$$

其中,  $i = 1, 2, \dots, k$ ,  $j = 1, 2, \dots, n_i$ .

由此判断如果  $\boldsymbol{\Phi}$  满足式 (43) 的条件, 则  $\tilde{\mathbf{E}}^*$  属于第  $r$  类, 即认为该类中所有元素对应的仿真模型均有效; 否则,  $\tilde{\mathbf{E}}^*$  不属于任何类, 即认为所有的仿真模型均不可信.

$$\begin{cases} d(\boldsymbol{\Phi}, \boldsymbol{\varphi}_0^r) = \min_{1 \leq i \leq k} \{d(\boldsymbol{\Phi}, \boldsymbol{\varphi}_0^i)\} \\ \min_{1 \leq j \leq n_i} \{d(\boldsymbol{\Phi}, \boldsymbol{\varphi}_j^r)\} \leq \max_{1 \leq i \leq k} \max_{1 \leq j, p \leq n_i, j \neq p} \{d(\boldsymbol{\varphi}_j^i, \boldsymbol{\varphi}_p^i)\} \end{cases} \quad (45)$$

其中,  $\boldsymbol{\varphi}_0^r = \sum_{j=1}^{n_r} \boldsymbol{\varphi}_j^r / n_r$ .

进一步, 如果  $\tilde{\mathbf{E}}^*$  属于第  $r$  类, 则得到  $\boldsymbol{\varphi}_c^r$  所对应的仿真模型最为可信:

$$d(\boldsymbol{\Phi}, \boldsymbol{\varphi}_c^r) = \min_{1 \leq i \leq n_r} \{d(\boldsymbol{\Phi}, \boldsymbol{\varphi}_i^r)\} \quad (46)$$

#### 4 应用实例与分析

假定研究对象为某导弹的比例导引系统, 其攻击平面为铅垂面, 与目标的相对位置关系如图 1 所示. 其中,  $M$  表示导弹;  $T$  表示目标;  $V$  为导弹速度;  $V_T$  为目标速度;  $q$  为目标线角;  $\eta$  为导弹速度矢量前置角;  $\theta$  为导弹弹道角, 由于攻击平面为铅垂面, 所以其也是弹道倾角;  $\eta_T$  为目标速度矢量前置角;  $\sigma_T$  为目标航向角.

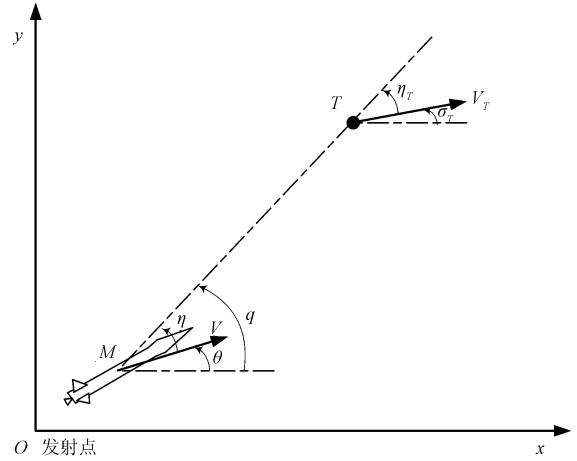


图 1 目标与导弹的相对位置

Fig. 1 Relative position between the missile and the target

依据各物理量之间的关系以及导引关系, 建立的系统模型如下所示:

$$\begin{cases} \dot{r} = V_T \cos \eta_T - V \cos \eta \\ \dot{q} = \frac{V \sin \eta - V_T \sin \eta_T}{r} \\ \dot{\theta} = k(\dot{q} + \varepsilon) \\ q = \eta + \theta = \eta_T + \sigma_T \\ r = \sqrt{(x - x_T)^2 + (y - y_T)^2} \end{cases} \quad (47)$$

式中,  $r$  为导弹和目标之间的距离;  $\varepsilon$  为测量目标线角的噪声信号;  $x$  和  $y$  分别为导弹位置的横坐标和纵坐标;  $x_T$  和  $y_T$  分别为目标位置的横坐标和纵坐标.

为了验证本文方法的有效性, 以某参数下的仿真模型为参考系统, 其他 10 组不同参数下的仿真模型为验证和选择对象. 由于噪声信号  $\varepsilon$  为随机信号, 因此参考系统和备选仿真模型均运行 30 次, 且采样周期相同. 关注的输出包含导弹的位姿信息、弹道倾角、弹道倾角速度以及脱靶量. 在这多个系统输出中, 导弹的位姿信息和弹道倾角为缓变数据, 弹道倾角速度中由于包含了噪声, 因此将其视为速变数据, 而脱靶量为静态数据.

对于动态数据, 在求取特征差异时, 需要选取仿真输出和参考输出均有采样数据的时间段. 本实例中的目标设定为典型的匀速直线运动, 脱靶量为导弹攻击目标时, 二者的最小距离. 其样本为参考系统和备选仿真模型分别运行 30 次的结果. 在此, 将采用本文方法对供选择的仿真模型进行有效性分析, 并提供给模型用户最为可信的模型. 如图 2~图 10 所示, 为第一个模型输出与参考输出的对比图.

针对不同数据类型的输出, 分别采用相应的数据特征差异度量模型, 得到的数据特征差异如表 1

所示. 由于 11 个特征差异之间存在着相关性, 通过主成分分析可知, 前三个主成分的累积方差贡献率就达到了 98.19%. 由此, 得到的最理想特征差异向量为  $\hat{\mathbf{E}}^* = [-4.800, 1.045, -0.050]$ , 得到新的特征差异向量为  $\hat{\mathbf{E}}_i = [\hat{e}_1^i, \hat{e}_2^i, \hat{e}_3^i]$ ,  $i = 1, 2, \dots, 10$ , 如式 (48) 所示, 特征差异的主成分如表 2 所示.

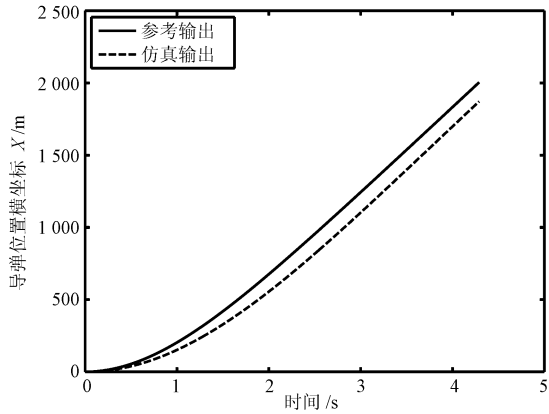


图 2 导弹位置横坐标

Fig. 2 Horizontal position of the missile

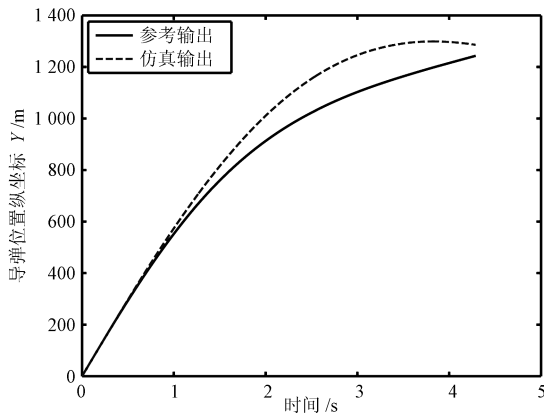


图 3 导弹位置纵坐标

Fig. 3 Longitudinal position of the missile

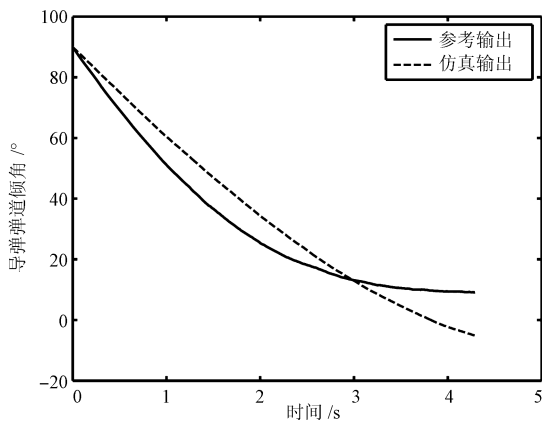


图 4 导弹弹道倾角

Fig. 4 Flight-path angle of the missile

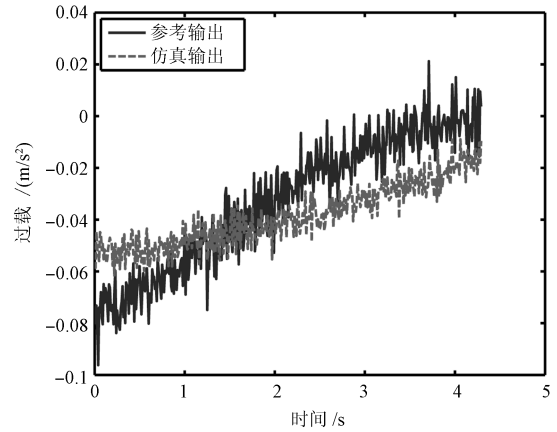


图 5 导弹弹道倾角速度

Fig. 5 Rate of flight-path angle

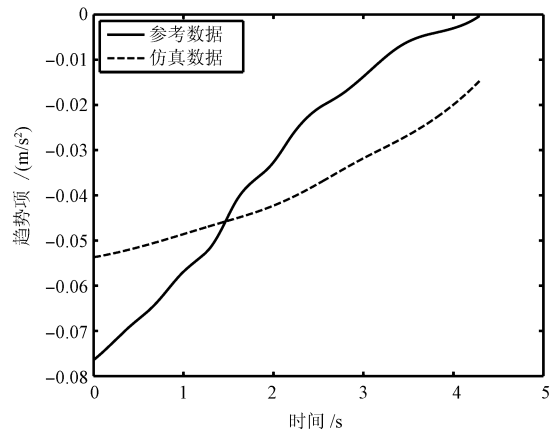


图 6 导弹弹道倾角速度趋势项

Fig. 6 Trend item of rate of flight-path angle

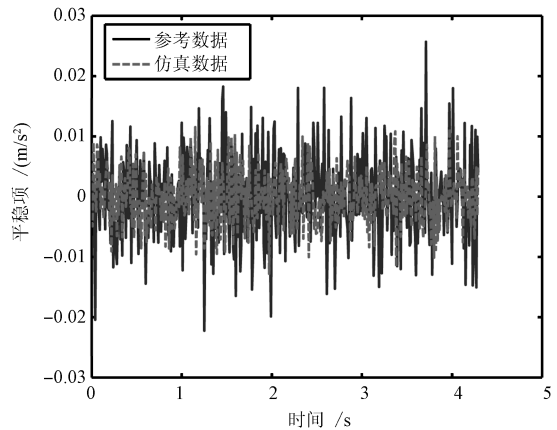


图 7 导弹弹道倾角速度平稳项

Fig. 7 Stationary item of rate of flight-path angle

表 1 仿真输出与参考输出的特征差异

Table 1 Feature differences between simulation outputs and reference outputs

变量	差异	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组	第 7 组	第 8 组	第 9 组	第 10 组
$d_{\text{脱靶量}}$	$e_c$	1.041	2.083	1.002	0.110	0.567	1.800	6.873	0.143	5.864	5.500
	$e_d$	0.734	2.729	5.718	1.676	3.331	2.159	2.753	4.859	5.348	6.889
$x$	$e_s$	5.278	0.972	3.898	2.755	2.149	0.435	1.577	4.334	4.128	5.013
	$e_t$	49.235	7.681	30.687	25.078	20.128	3.260	13.058	34.052	38.384	38.488
$y$	$e_s$	4.463	0.608	2.795	2.367	1.991	0.220	1.195	3.168	3.587	3.557
	$e_t$	51.309	7.551	34.867	27.427	22.764	3.743	13.411	39.128	41.024	44.675
$\theta$	$e_s$	0.377	0.070	0.313	0.210	0.159	0.035	0.116	0.348	0.298	0.403
	$e_t$	7.450	1.437	6.482	4.239	3.121	0.733	2.368	7.206	5.884	8.341
$\dot{\theta}$	$e_s$	0.042	0.011	0.032	0.022	0.017	0.006	0.015	0.043	0.034	0.044
	$e_t$	0.814	0.221	0.556	0.419	0.346	0.130	0.291	0.875	0.642	0.837
	$e_f$	0.760	0.333	0.496	0.388	0.411	0.310	0.388	0.713	0.550	0.589

表 2 特征差异的主成分

Table 2 Principal components of feature differences

主成分变量	第 1 组	第 2 组	第 3 组	第 4 组	第 5 组	第 6 组	第 7 组	第 8 组	第 9 组	第 10 组
$\hat{e}_1$	3.700	-2.761	1.644	-0.213	-0.882	-3.549	-1.837	2.921	2.214	3.564
$\hat{e}_2$	1.468	-0.045	-0.081	1.085	0.499	0.128	-1.521	0.567	-1.451	-1.694
$\hat{e}_3$	-1.297	0.187	1.027	-0.134	0.383	0.123	-0.962	0.792	-0.420	0.350

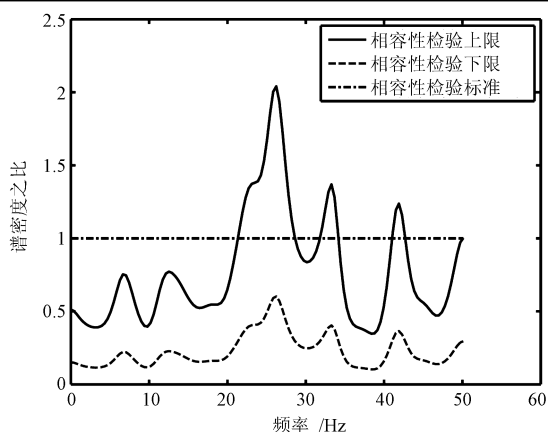


图 8 弹道倾角平稳项的相容性检验

Fig. 8 Compatibility check for stationary item of flight-path angle

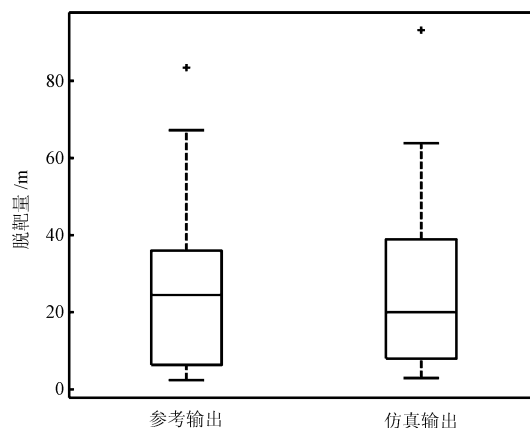


图 10 脱靶量箱线图

Fig. 10 Boxplot of the miss distance

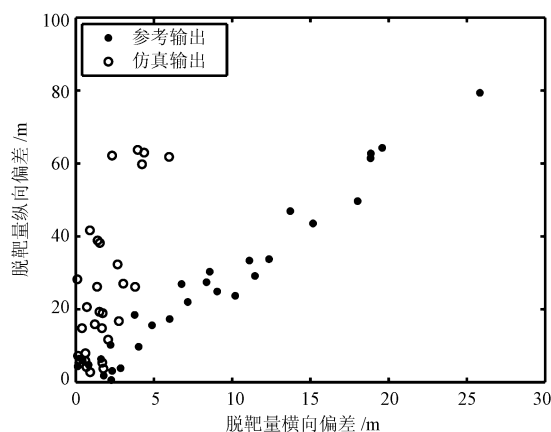


图 9 脱靶量分布图

Fig. 9 Dispersion of the miss distance

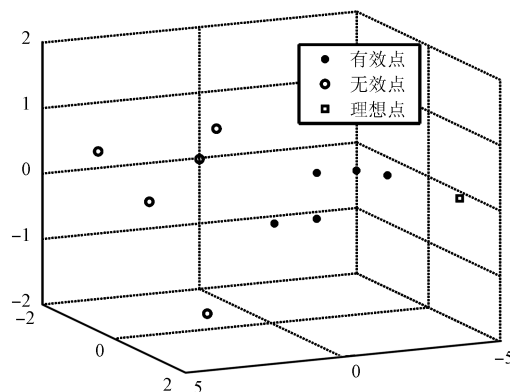


图 11 各仿真模型对应特征差异主成分的空间位置  
Fig. 11 Spatial position of principal components of feature differences for each simulation model

$$\hat{\mathbf{E}}_i = [\tilde{e}_1^i, \tilde{e}_2^i, \dots, \tilde{e}_{11}^i] \times T, \quad i = 1, 2, \dots, 10 \quad (48)$$

$$T = \begin{bmatrix} 0.007 & -0.832 & -0.544 \\ 0.155 & -0.531 & -0.759 \\ 0.342 & 0.013 & -0.041 \\ 0.335 & 0.072 & -0.202 \\ 0.336 & 0.068 & -0.188 \\ 0.340 & 0.052 & -0.103 \\ 0.342 & -0.008 & 0.074 \\ 0.341 & -0.020 & 0.115 \\ 0.334 & -0.007 & 0.060 \\ 0.331 & 0.008 & 0.042 \\ 0.256 & 0.111 & -0.126 \end{bmatrix} \quad (49)$$

其中,  $\tilde{\mathbf{E}}_i = [\tilde{e}_1^i, \tilde{e}_2^i, \dots, \tilde{e}_{11}^i], i = 1, 2, \dots, 10$  为  $\mathbf{E}_i = [e_1^i, e_2^i, \dots, e_{11}^i], i = 1, 2, \dots, 10$  经过标准化处理的结果.

进一步, 基于  $K$  均值聚类分析将  $C_1 = \{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_3, \hat{\mathbf{E}}_8, \hat{\mathbf{E}}_9, \hat{\mathbf{E}}_{10}\}$  和  $C_2 = \{\hat{\mathbf{E}}_2, \hat{\mathbf{E}}_4, \hat{\mathbf{E}}_5, \hat{\mathbf{E}}_6, \hat{\mathbf{E}}_7\}$  两类, 如图 11 所示. 接下来, 采用 Fisher 判别分析判定  $\hat{\mathbf{E}}^*$  属于其中的哪一类, 得到的投影向量如下:

$$\boldsymbol{\eta} = [0.859, -0.287, 0.424]^T \quad (50)$$

$C_1 = \{\hat{\mathbf{E}}_1, \hat{\mathbf{E}}_3, \hat{\mathbf{E}}_8, \hat{\mathbf{E}}_9, \hat{\mathbf{E}}_{10}\}, C_2 = \{\hat{\mathbf{E}}_2, \hat{\mathbf{E}}_4, \hat{\mathbf{E}}_5, \hat{\mathbf{E}}_6, \hat{\mathbf{E}}_7\}$  以及  $\hat{\mathbf{E}}^*$  投影变换结果如下式所示, 相对位置关系如图 12 所示.

$$\tilde{C}_1 = \{\hat{\mathbf{E}}_1\boldsymbol{\eta}, \hat{\mathbf{E}}_3\boldsymbol{\eta}, \hat{\mathbf{E}}_8\boldsymbol{\eta}, \hat{\mathbf{E}}_9\boldsymbol{\eta}, \hat{\mathbf{E}}_{10}\boldsymbol{\eta}\} = \{2.207, 1.871, 2.683, 2.140, 3.695\} \quad (51)$$

$$\tilde{C}_2 = \{\hat{\mathbf{E}}_2\boldsymbol{\eta}, \hat{\mathbf{E}}_4\boldsymbol{\eta}, \hat{\mathbf{E}}_5\boldsymbol{\eta}, \hat{\mathbf{E}}_6\boldsymbol{\eta}, \hat{\mathbf{E}}_7\boldsymbol{\eta}\} = \{-2.280, -0.551, -0.739, -3.033, -1.549\} \quad (52)$$

$$\boldsymbol{\Phi} = \hat{\mathbf{E}}^* \boldsymbol{\eta} = -4.444 \quad (53)$$

由投影结果易得:

$$\begin{cases} \boldsymbol{\Phi} \in \tilde{C}_2 \\ d(\boldsymbol{\Phi}, \tilde{\mathbf{E}}_6 T) = \min_{i=1,2,\dots,10} \{d(\boldsymbol{\Phi}, \tilde{\mathbf{E}}_i T)\} \end{cases} \quad (54)$$

由此可知,  $C_2$  中所有元素对应的仿真模型均有效, 且第 6 个模型为最有效模型.

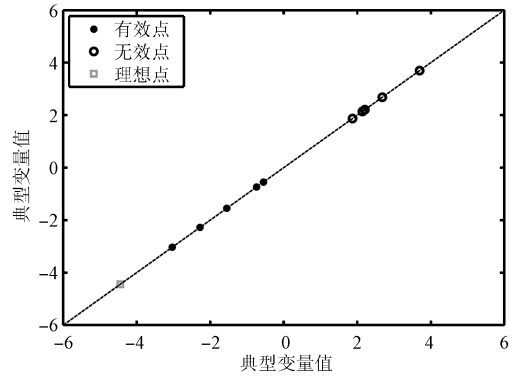


图 12 Fisher 判别中各典型变量的分布情况

Fig. 12 Dispersion of each canonical variable in Fisher discriminant analysis

### 5 结论

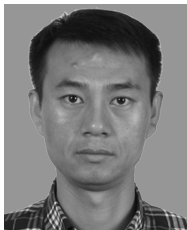
本文研究多个备选仿真模型的验证及选择问题. 通过比较仿真输出和参考输出之间的特征差异是常用手段之一. 当系统为单输出时, 已有的经典仿真模型验证方法能较好地解决此类问题. 但是, 当系统为多输出且数据类型各异时, 已有的经典方法则难以解决. 针对此问题, 将系统输出划分为静态、缓变和速变三类数据, 且分别刻画了各类数据的特征差异. 基于主成分分析从多个具有相关性的特征差异中提取出少数几个独立的主成分. 在此基础上, 采用  $K$ -均值聚类方法将多个待选仿真模型输出划分为  $k$  类, 并采用 Fisher 判别分析法判定参考输出是否属于其中某类, 进而实现对多个备选仿真模型的验证和选择. 在实例应用中, 本文方法有效地验证了备选仿真模型的可信性, 且选出了最可信模型. 本文方法仅适用于作为参考输出的实际系统输出已知的情形. 当实际系统输出未知时, 可以将仿真应用领域的先验知识作为参考数据. 此时就需要研究如何描述这些领域先验知识以及如何利用它们开展仿真模型的验证和选择工作.

### References

- 1 Sargent R G. Verification and validation of simulation model. In: Proceeding of the 2010 Winter Simulation Conference. Baltimore, MD: IEEE, 2010. 166–183
- 2 Liu Fei, Ma Ping, Yang Ming, Wang Zi-Cai. Research on credibility quantification of complex simulation systems. *Journal of Harbin Institute of Technology*, 2007, **39**(1): 1–3 (刘飞, 马萍, 杨明, 王子才. 复杂仿真系统可信度量化研究. 哈尔滨工业大学学报, 2007, **39**(1): 1–3)
- 3 Lu Shao-Wen. Two issues towards verification of simulation of superposed alternative renewal processes. *Acta Automatica Sinica*, 2009, **35**(5): 636–640

- (卢绍文. 重叠交替更新过程的 DTSS 仿真校验的两个问题. 自动化学报, 2009, **35**(5): 636–640)
- 4 Min Fei-Yan, Yang Ming, Wang Zi-Cai. Knowledge-based method for the validation of complex simulation models. *Simulation Modelling Practice and Theory*, 2010, **18**(5): 500–515
  - 5 Naylor T H, Finger J M. Verification of computer simulation models. *Management Science*, 1967, **14**(2): 92–101
  - 6 Kheir N A, Holmes W M. On validating simulation models of missile systems. *Simulation*, 1978, **30**(4): 117–128
  - 7 Wu Jing, Wu Xiao-Yan, Chen Yong-Xing, Teng Jiang-Chuan. Validation of simulation models based on improved grey relational analysis. *Systems Engineering and Electronics*, 2010, **32**(8): 1677–1679  
(吴静, 吴晓燕, 陈永兴, 滕江川. 基于改进灰色关联分析的仿真模型验证方法. 系统工程与电子技术, 2010, **32**(8): 1677–1679)
  - 8 Damborg M J. An example of error analysis in dynamic model validation. *Simulation*, 1985, **44**(6): 301–305
  - 9 Liu Shi-Kao, Liu Xing-Tang, Zhang Wen. Fixed quantity evaluation to reliability of simulation system with similar degree. *Journal of System Simulation*, 2002, **14**(2): 143–145  
(柳世考, 刘兴堂, 张文. 利用相似度对仿真系统可信度进行定量评估. 系统仿真学报, 2002, **14**(2): 143–145)
  - 10 Montgomery D C, Conard R G. Comparison of simulation and flight-test data for missile systems. *Simulation*, 1980, **34**(2): 63–72
  - 11 Liu Zao-Zhen. Simulation validation based on the data of the aero experimentation. *Journal of System Simulation*, 2002, **14**(3): 281–284  
(刘藻珍. 基于飞行试验数据的仿真模型验证方法的研究. 系统仿真学报, 2002, **14**(3): 281–284)
  - 12 Li Peng-Bo, Gao Xia. Application of MESA on validating missile simulation model. *Journal of National University of Defense Technology*, 1999, **21**(2): 9–12  
(李鹏波, 高霞. 应用最大熵谱估计进行导弹系统的仿真模型验证. 国防科技大学学报, 1999, **21**(2): 9–12)
  - 13 Wei Hua-Liang, Liu Zao-Zhen. Cross spectral estimation and its application in missile systems simulation validation. *Journal of System Simulation*, 1997, **9**(3): 116–121  
(魏华梁, 刘藻珍. 交叉谱估计及其在导弹系统仿真模型验证中的应用. 系统仿真学报, 1997, **9**(3): 116–121)
  - 14 Balci O. Verification, validation, and certification of modeling and simulation applications. In: Proceedings of the 2003 Winter Simulation Conference. New Orleans: ACM, 2003: 150–158
  - 15 Kleijnen J P C. An overview of the design and analysis of simulation experiments for sensitivity analysis. *European Journal of Operational Research*, 2005, **164**(2): 287–300
  - 16 Liu Hui-Ying, Sun Zhen, Liu Xin. Research and implementation of missile simulation model validation system. *Modern Electronics Technique*, 2012, **35**(3): 1–4  
(刘慧英, 孙真, 刘昕. 导弹仿真模型验证系统的研究与实现. 现代电子技术, 2012, **35**(3): 1–4)
  - 17 Huang N E, Shen Z, Long S R, Wu M C, Shih H H, Zheng Q, Yen N C, Tung C C, Liu H H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London A*, 1998, **454**(1971): 903–995
  - 18 Ivan M C, Baraniuk R G. Empirical mode decomposition based time-frequency attributes. In: Proceedings of the 69th SEG Meeting. Houston, USA, 1999.
  - 19 Huo Dong-Hai, Yang Dan, Zhang Xiao-Hong, Hong Ming-Jian. Principal component analysis based codebook background modeling algorithm. *Acta Automatica Sinica*, 2012, **38**(4): 591–600  
(霍东海, 杨丹, 张小洪, 洪明坚. 一种基于主成分分析的 Codebook 背景建模算法. 自动化学报, 2012, **38**(4): 591–600)
  - 20 Pan Chun-Guang, Chen Ying-Wu, Wang Hao. Principal component analysis' application to the software metrics-based for risk assessment. *Operations Research and Management Science*, 2005, **14**(5): 80–84  
(潘春光, 陈英武, 汪浩. 主成分分析法在基于度量的软件风险评估中的应用. 运筹与管理, 2005, **14**(5): 80–84)
  - 21 Jain A K. Data clustering: 50 years beyond  $K$ -means. *Pattern Recognition Letters*, 2010, **31**(8): 651–666
  - 22 Zhang Wen-Jun, Gu Xing-Fa, Chen Liang-Fu, Yu Tao, Xu Hua. An algorithm for initializing of  $K$ -Means clustering based on mean-standard deviation. *Journal of Remote Sensing*, 2006, **10**(5): 715–721  
(张文君, 顾行发, 陈良富, 余涛, 许华. 基于均值 - 标准差的  $K$  均值初始聚类中心选取算法. 遥感学报, 2006, **10**(5): 715–721)
  - 23 Maulik U, Bandyopadhyay S. Performance evaluation of some clustering algorithms and validity indices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, **24**(12): 1650–1654
  - 24 Cheng Zheng-Dong, Zhang Yu-Jin, Fan Xiang, Zhu Bin. Study on discriminant matrices of commonly-used Fisher discriminant functions. *Acta Automatica Sinica*, 2010, **36**(10): 1361–1370  
(程正东, 章毓晋, 樊祥, 朱斌. 常用 Fisher 判别函数的判别矩阵研究. 自动化学报, 2010, **36**(10): 1361–1370)

25 Chiang L H, Kotanchek M E, Kordon A K. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering*, 2004, **28**(8): 1389–1401

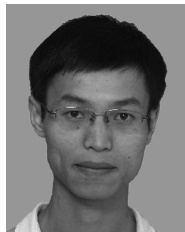


**李 伟** 哈尔滨工业大学讲师. 主要研究方向为仿真分析与评估, 分布式仿真.  
E-mail: fleehit@163.com  
(**LI Wei** Lecturer at Harbin Institute of Technology. His research interest covers simulation analysis and evaluation and distributed simulation.)



**焦 松** 哈尔滨工业大学博士研究生. 主要研究方向为仿真模型验证, 仿真试验设计, 武器效能评估.  
E-mail: jiaosong1985@163.com  
(**JIAO Song** Ph.D. candidate at Harbin Institute of Technology. His research interest covers simulation model validation, simulation experiment de-

sign, and operational effectiveness evaluation.)



**陆凌云** 哈尔滨工业大学博士研究生. 主要研究方向为仿真试验设计和仿真数据分析.

E-mail: lulingyun987@163.com

(**LU Ling-Yun** Ph.D. candidate at Harbin Institute of Technology. His research interest covers simulation experiment design and simulation data anal-

ysis.)



**杨 明** 哈尔滨工业大学教授. 主要研究方向为飞行器制导控制, 复杂系统仿真理论与方法. 本文通信作者.

E-mail: myang@hit.edu.cn

(**YANG Ming** Professor at Harbin Institute of Technology. His research interest covers vehicle guidance and control, complex system simulation theory and method. Corresponding author of this paper.)