

统计推断在数模转换系统中的应用

组号：27 葛嘉昊 5130309646，周浩宇 5130309633

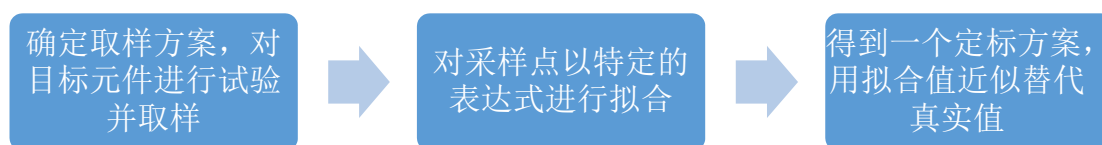
关键词：拟合，插值，遗传算法，启发式搜索，单点测定成本，单点定标误差成本

1. 引言

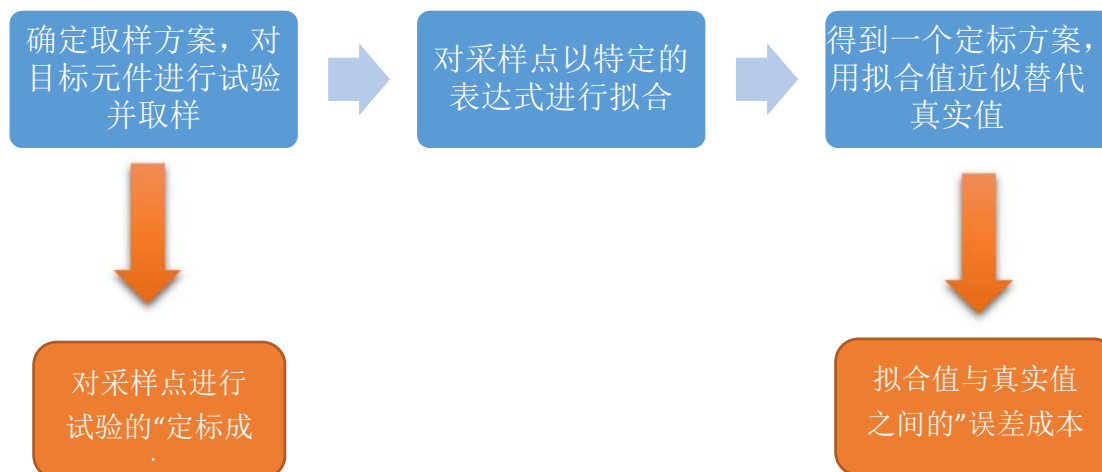
为某种产品内部的一个监测模块，寻求校准工序的优化方案。假定有某型投入批量试生产的电子产品，其内部有一个模块，功能是监测某项与外部环境有关的物理量（可能是温度、压力、光强等）。该监测模块中传感器部件的输入输出特性呈明显的非线性。本文为该模块的批量生产设计一种成本合理的传感特性校准（定标工序）方案。

2. 问题求解的途径

为了寻求定标方案，我们需要选择 2 个要素：对传感特性曲线拟合的方案，以及为特定的某个模块进行拟合提供基础数据的取样测量方案。故确定某个特定元件的定标方案大致可分为以下三步：



对一个定标方案，产生成本之处有 2 个：



为使总成本最小，需要尽可能的降低这两处成本，根据成本公式：

$$S_i = \sum_{j=1}^{s_i} s_{i,j} + q \cdot n_i \quad (2)$$

（其中 $s_{i,j}$ 为单点误差， n_i 为采样点个数， q 取 20）

若想降低成本，一需要减小误差，二需要减少采样点个数

2.1 减少误差的办法

2.1.1 选择合适的拟合表达式

随机选取 4 个样品进行 4 次全点实验并画图，得到图 2-1 如下：

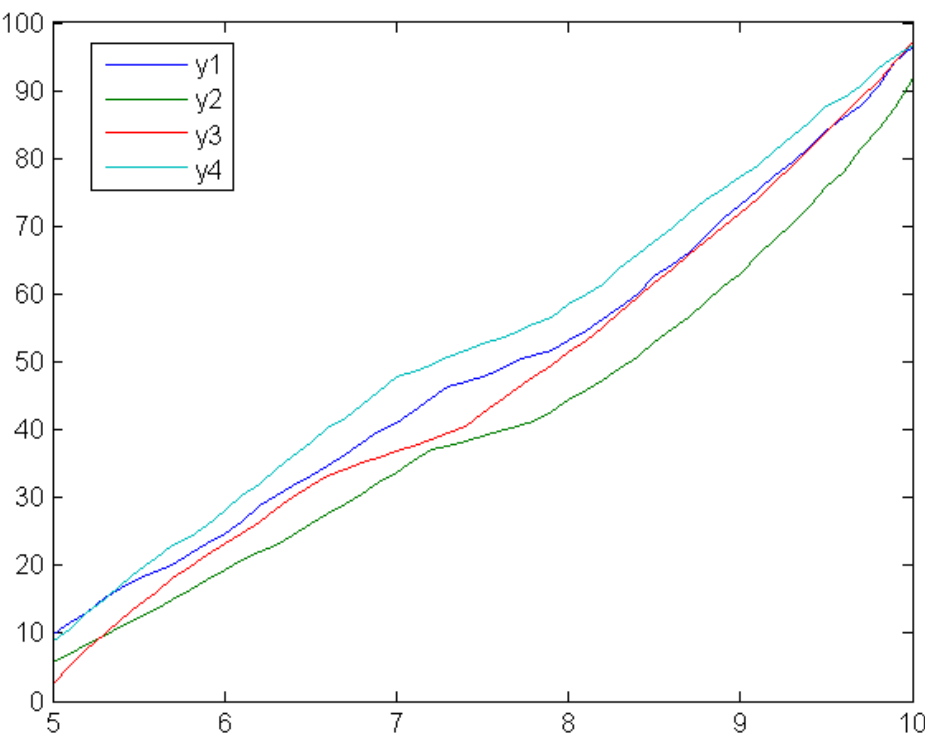


图 2-1

由上图可以看出，该模块的传感特性曲线基本成正比例，曲率变化并不大，进行进一步数据分析后得表 2-1 如下，

随机选取曲线	Y1	Y2	Y3	Y4
相关系数	0.9952	0.9882	0.9934	0.9976

表 2-1

其相关性较高，在随机选取一个采样方案（采样率为 1/2）后分别使用 MATLAB 提供的几种拟合形式测试，根据运算速度和误差大小出如下表 2-2

拟合方式	误差大小（四次总和）	运算时间/s
Poly1	913.5	0.0358
Poly2	517	0.0388
Poly3	229	0.0407
Poly4	141	0.0411

Poly5	124	0.0433
Poly6	92.5	0.0712
Poly7	78.5	0.0876
Poly8	57.5	0.0895
fourier1	571	0.3207
Fourier2	107.5	0.1013
Fourier3	69.5	0.2080
Fourier4	40.5	0.1720
Fourier5	40.5	0.1787
Fourier6	68	0.1188
Fourier7	29.5	0.1968
Fourier8	123.5	0.1803
Gauss1	1268.5	0.1661
Gauss2	124.5	0.5847
Gauss3	36	0.7601
Gauss4	67	1.5257
Gauss5	58	1.8232
Gauss6	45.5	1.6629
Gauss7	61.5	1.7301
Gauss8	19.5	1.7158
Power1	1334.5	0.0998
Power2	572.5	0.1306
Sin1	914.5	2.7235
Sin2	124	3.4170
Sin3	132.5	3.4244
Sin4	28.5	3.4669
Sin5	70	3.5172

Sin6	135.5	3.5977
Sin7	196	3.5856
Sin8	246.5	3.5755
Interpl(spline)	8.5	0.00098360
Interpl(nearest)	488.5	0.00081083
Interpl(linear)	23.5	0.00069579
Interpl(cubic)	12	0.00075267

表 2-2

由于拟合误差使用的是公式（3），所以并不用特意追求均方差 SSE 最小。由上表数据可以直观的看出，内插值 interpl（spline）与（cubic）无论是误差还是计算速度都有明显的优势，而且不需要高阶拟合方式所要求的最少数据点，无论对怎样的采样方案都是可以计算的。综合考虑下，选择 interpl(spline)为最终的拟合方式。

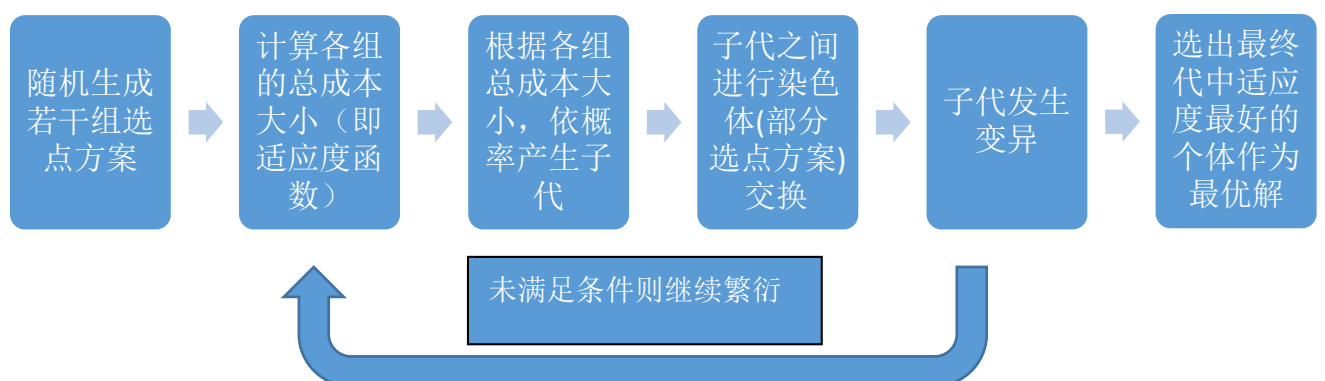
（以上代码详见附录）

2.2 减少采样点个数的办法

由于选点个数最少与误差最小两者存在矛盾，故需要找到一个总成本最小的优化方案，这里本组使用遗传算法进行求解：

2.2.1 遗传算法

通过对自然界仿生的遗传算法，以选点方案作为数据模拟生物体的染色体，以总成本大小作为适应度函数模拟自然选择，以种群为单位进行多代的繁衍，最终可得到相对最优的定标方案，其大致流程如下图 2-3：



根据本课题的要求，我们定义每个个体由“拟合曲线的类型”与“选点方案”组成。其中拟合类型为表 2-2 中提供的，选点方案为一个 1x51 的向量，只存有 0 和 1，有 1 的点表示对应位置被取样。

其中计算的适应度函数，为计算方便，选择为总成本大小：

$$C = \sum_{i=1}^M S_i = \sum_{i=1}^M \left(\sum_{j=1}^{51} s_{i,j} + q \cdot n_i \right) \quad (3)$$

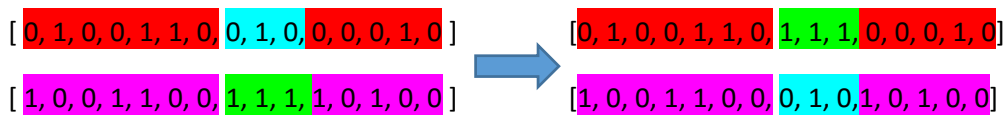
而个体将自己的染色体（选点方案）传给子代的几率为：

$$P_i = \frac{(C_{\max} - C_i)^2}{\sum_{j=1}^G (C_{\max} - C_j)^2} \quad (C_{\max} > C_i)$$

$$\text{或} \quad P_i = 0 \quad (C_{\max} \leq C_i) \quad (4)$$

其中 G 为种群大小。将成本超过 C_{\max} 的个体直接消去。 C_{\max} 为一个与种群无关的常量，根据初始种群的总成本、遗传算法的收敛速度，期望得到的最终结果大小等确定，本实验中采用过多个 C_{\max} 值。

染色体交换过程为随机抽取 2 个子代个体，随机在选点方案中抽取一点作为起点，随机选择交换长度（长度上限为总长度除 8），将其片段与另一个个体的片段交换：



突变过程中，根据给定的突变概率，对选点方案中的每个点进行判断，落在概率中的点的数据取反：



在实验中我们发现，经过多轮进化后，在总成本下降到 200 以下时，经常出现优秀个体（总成本 150 以下）无法稳定的把优良性状遗传下去的情况。分析后认为是在总成本在 200 以下时，选点数较少，突变与交叉遗传对总成本影响变得十分大。而变异大多都是有害变异，故导致较差个体占满种群。为了保证种群不退化，我们使当代总成本最小的最优个体无条件复制到下一代以保持优良基因。

（以上代码详见附录）

3. 结论

经过实验，种群大小 15，繁衍 200 代，突变率 1%， $C_{\max}=170$ ，得出最终结果如下：
选点方案为：

1	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	1	0	0	0
0	0	0	0	0	0	0	1	0	0	0

0 0 0 1 0 0 0

即[1, 11, 22, 30, 41, 48]

拟合方式为 `interp1(spline)`

该方案总成本=96.9072，其中测量成本 72，误差成本 24.9072

4. 参考文献

- [1] 袁焱等.《统计推断课程讲义》.上海交通大学电子工程系. 2009.9~2009.11
- [2] 武爱文、冯卫国等.《概率论与数理统计》.上海交通大学出版社 2011

5. 附录

6. 基本流程（平台：matlab，具体代码见附件）

