

# 基于聚类分析和可视化的增强遗传算法 —II. 算例分析及有效性验证

孙晓静, 王克峰, 姚平经

(大连理工大学化工过程系统工程研究所, 辽宁 大连 116023)

**摘要:**通过比较  $K$  平均算法与聚类约束映射(CCM)的聚类结果, 表明了 CCM 在降维过程中保持拓扑信息的有效性. 应用前文提出的增强遗传算法(IGA)对 3 个有约束优化算例进行了求解. 结果表明, 这种可视化、聚类分析与遗传算法相结合的方法可以帮助用户参与选择聚类参数, 比以往一些方法更有效.

**关键词:**可视化; 聚类分析; 带约束的优化; 遗传算法

中图分类号: TP391.75 文献标识码: A 文章编号: 1009-606X(2004)06-0536-08

## 1 前言

对于约束优化问题, 常规遗传算法的惩罚函数法需要设置适当的惩罚参数才能得到问题的可行解, 这给优化过程带来困难. 由前文<sup>[1]</sup>提出的多维数据可视化聚类分析和用户参与结合起来的增强遗传算法, 通过数据分析来获得可行域信息, 因此不需要惩罚参数.

可行域的相关信息由基于数据可视化的聚类分析得到. 这种基于可视化的聚类分析法, 首先提出了一种基于前向人工神经网络(ANN)的聚类约束映射(CCM)来进行从初始  $n$  维空间到 2 维的维数缩减映射, 同时在缩减的维空间中保存聚类信息. 然后, 用 Ward<sup>[2]</sup>的作用于 2 维空间的凝聚算法进行聚类分析, 参数通过可视化和后来的用户交互提供. 最后, 再将从 2 维空间得到的聚类信息映射回  $n$  维空间, 从而得到在初始维数下的可行域信息, 该信息最后用于改进的遗传算法 IGA.

对于给定的 1 个非线性规划(NLP)模型, 改进的遗传算法(IGA)通过数据可视化的聚类分析来获得可行域信息; 重新定义进化算子(交叉和变异), 利用已获得的可行域信息, 保证子代有较多的可行解; 如果个体不可行, 置该目标值为无穷大(如  $10^{20}$ ).

以下首先用 1 个实例详细分析了 CCM 的聚类分析算法的有效性, 然后用 3 个实例表明可视化、聚类分析和遗传算法集成方法的有效性.

## 2 聚类约束映射的有效性

对给定的一个  $n$  维带约束的优化问题, 基于 CCM 的聚类分析算法有以下 4 个步骤:

- (1) 随机产生  $10^n$  个节点, 并选出可行点来表征  $n$  维数集;
- (2) 应用 CCM 将  $n$  维数据映射到 2 维;
- (3) 应用聚类分析算法到可视化数据上来获得可行域信息;
- (4) 将 2 维空间的可行域信息映射回  $n$  维空间.

由于尚没有有效性的理论证明, 本工作通过比较分别由  $K$  平均算法<sup>[3]</sup>和 CCM 得到的聚类结果来表明 CCM 映射法的有效性.

收稿日期: 2003-09-26, 修回日期: 2003-11-24

基金项目: 国家 973 基金资助项目(编号: G2000263)

作者简介: 孙晓静(1980-), 女, 辽宁省辽阳市人, 硕士研究生, 化工过程系统工程专业, E-mail: minipaula@163.com.

应用 1 个已有的研究例子<sup>[4]</sup>进行证明. 该例子为 38 个化合物的物性参数(表 1), 其中每个化合物都有 5 个特性参数( $x_1, x_2, \dots, x_5$ ), 其中 4 组不同的随机数种子的 CCM 所应用的前向反馈人工神经网络(即前文所描述的 ANN)结构完全相同, 隐藏层节点都是  $n+1=6$  个( $n$  为特性参数数目). 4 个例子中, 目标训练迭代数均取 10 000(下例相同), 在映射中的每一个数据点都由它所表示的化合物的下标来表示.

表 1 物性参数

Table 1 The property data of materials

$N$	Compound	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
1	ZnS	3.90	3.823	6.261	97.434	53
2	ALN	6.20	3.110	4.980	40.990	25
3	ZnO	3.30	3.251	5.209	81.369	22
4	AgGaS <sub>2</sub>	2.64	5.751	10.238	241.718	53
5	CuGaS <sub>2</sub>	2.43	5.351	10.470	197.388	53
6	LiIO <sub>3</sub>	4.00	5.481	5.171	181.836	22
7	Se	1.70	4.361	4.954	78.960	66
8	GaS	2.50	3.586	15.496	101.784	53
9	SiC	6.00	4.359	4.359	40.090	29
10	SiO <sub>2</sub>	8.40	4.913	5.405	60.078	22
11	Te	0.33	4.457	5.939	236.550	82
12	AgI	2.80	6.473	6.473	234.770	126
13	CuCl	3.17	5.405	5.405	98.993	77
14	CuI	2.95	6.042	6.042	190.440	96
15	InSb	0.23	6.479	6.479	236.550	89
16	AgGaSe <sub>2</sub>	1.80	5.981	10.865	335.510	66
17	AgInSe <sub>2</sub>	1.20	6.099	11.691	286.798	66
18	InAs	0.36	6.268	6.479	189.790	71
19	CdGeAs <sub>2</sub>	0.57	5.943	22.217	334.970	71
20	GaSb	0.72	6.095	6.095	191.470	89
21	InSe	1.25	4.002	24.946	193.760	66
22	InP	1.35	5.868	5.868	145.770	59
23	Ag <sub>3</sub> AsS <sub>3</sub>	2.00	10.80	8.690	494.792	53
24	GaAs	1.40	5.653	5.653	144.710	71
25	CuGaSe <sub>2</sub>	1.70	5.606	11.006	242.468	66
26	GaSe	2.02	3.747	23.910	148.680	66
27	CuInS <sub>2</sub>	1.53	5.489	11.101	242.468	53
28	HgS	2.10	4.145	9.496	232.654	53
29	$\beta$ -SiC	2.26	4.359	4.359	40.090	29
30	GaP	2.30	5.450	5.450	100.690	59
31	ZnTe	2.30	6.101	6.101	192.970	82
32	ZnSe	2.70	5.667	5.668	144.330	66
33	CuBr	2.91	5.690	5.690	143.449	82
34	CdGeP <sub>2</sub>	2.91	5.740	10.776	246.930	59
35	ZnSiAs <sub>2</sub>	1.74	5.606	10.880	243.430	71
36	ZnGeP <sub>2</sub>	2.05	5.463	10.731	199.900	59
37	CdGa <sub>2</sub> S <sub>4</sub>	3.05	5.568	10.040	380.096	53
38	LiNbO <sub>3</sub>	4.00	5.148	13.860	147.842	22

首先应用 CCM 将  $n$  维数据映射到 2 维, 再基于该映射结果, 分别对给定的聚类数 3, 4, 6 和 8 进行平方根求和聚类分析, 结果见图 1. 图中 CCM 自动地映射到几个数据点群的区域, 通过比较可知各组映射彼此间有相似性, 即域分布基本一致. 同时应用  $K$  平均聚类算法进行分析(精确解取决于  $K$  聚类中心的初始位置), 类数分别设定为 3, 4, 6, 8, 所得分级聚类数见图 2. 比较两种方法的聚类结果, 邻近点如(6, 38), (2, 9, 10)在图 1 中也相近, 证明了在初始 6 维空间中相邻较近的数据点在降维的 2 维空间中仍很相近, 即 CCM 映射在降维映射过程中保持了节点原有的拓扑信息.

虽然  $K$  平均法对  $n$  维数据的聚类比较直观, 但用户在算法运行过程中不能进行干预, 也不能得到任何中间结果, 且其输出结果仅是聚类数, 使其很难在优化过程中得到进一步应用. 而基于

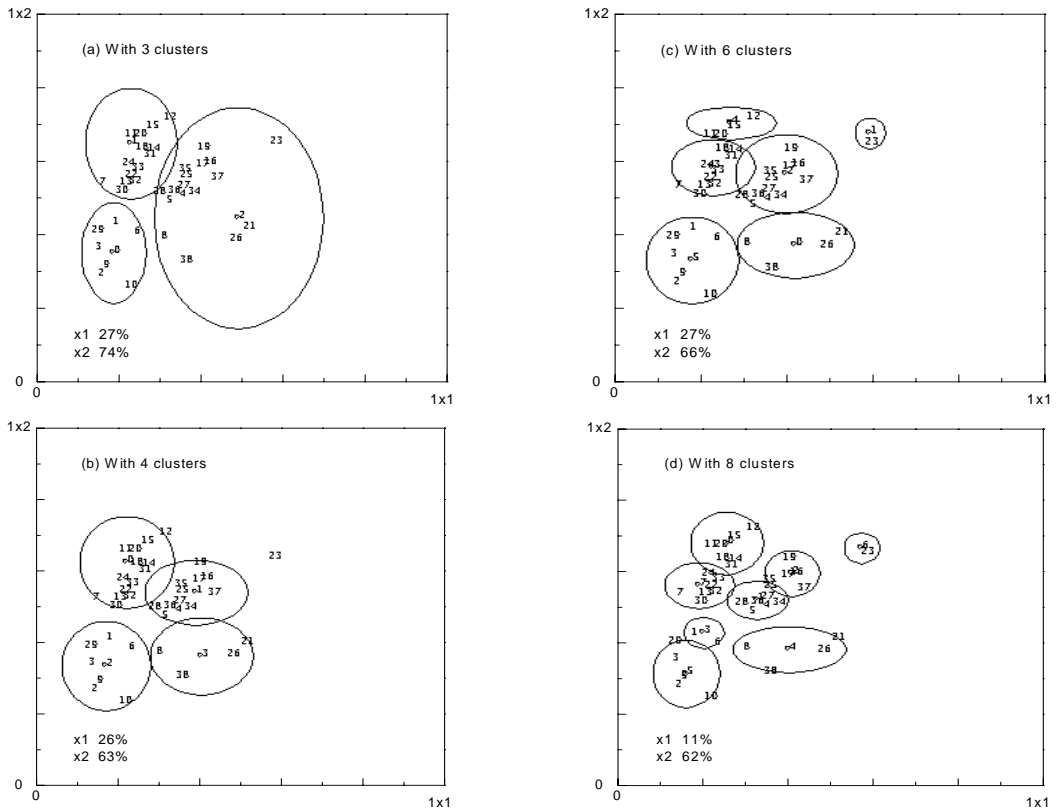


图 1 CCM 映射聚类分析

Fig.1 The cluster analysis based on CCM mapping

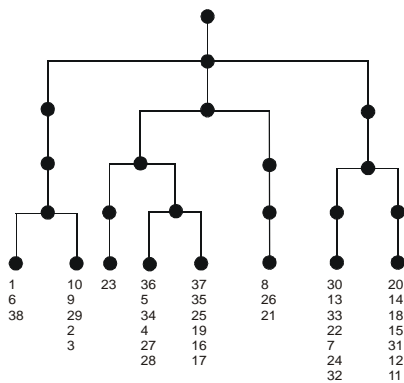


图 2 物性参数的分级聚类结构

Fig.2 Hierarchical cluster structure of materials property data (The K-means algorithm was used to obtain the cluster results)

CCM 映射的聚类分析是一种面向用户的方法，用户可参与到确定聚类数的过程中。如本例，点 23 离其它数据点较远，如果单由算法选用 3 个聚类，就要用一个很大径的聚类才能覆盖该点[图 1(a)]，这使后续优化过程中的计算工作量增大；或者如图 1(b)所示，选用 4 个聚类却都未能覆盖点 23。而可视化过程允许用户直接调节每一聚类的径—由数据可视化自行定义两个算法参数  $G$ (群体数)和  $\epsilon$  (用于在计算工作量与可行点覆盖之间取得平衡)，来确定适合的聚类数，此例确定为 6。因此，将用户与计算机结合起来比任何一种单独的方法更有效，这已由 Wegner<sup>[5]</sup>证实。

### 3 生成可行点通用问题的通用方法

上述方法的假定条件是只通过随机生成就可得到足够数量的可行点。但在一些例子中，要只靠随机生成来找到足够的可行点是很困难的，对此可用一种标准遗传算法(对应于本研究的改进遗

传算法 IGA)来生成有充分多样性的可行点.

用于适值估计(对种群多样性的度量)的目标函数:

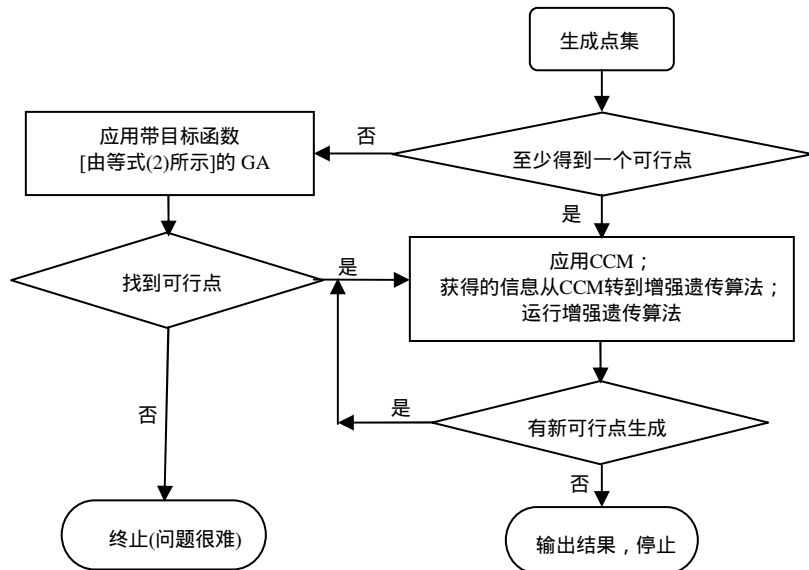
$$\max \sum_{p=1}^N \sum_{p'=1}^N d_{pp'}, \quad (1)$$

式中  $p, p'$  为可行点,  $N$  为可行点总数, 其中:

$$d_{pp'} = \begin{cases} 0 & \text{若两个个体中的任何一个为不可行} \\ \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{x_i^{(p)} - x_i^{(p')}}{x_i^u - x_i^l} \right)^2} & \text{若两个个体都可行} \end{cases}, \quad (2)$$

$n$  为空间维数,  $x_i$  ( $i=1, 2, \dots, n$ ) 为初始  $n$  维空间第  $i$  维坐标下变量, 其中  $(x_i^l, x_i^u)$  为第  $i$  个变量的变化范围. 若对任一第  $i$  个变量都有  $x_i^l = x_i^u$ , 则贡献值为 0.

用下述过程来处理通用问题:



上述程序运行需结合运用下列工具: Jacaranda<sup>[6]</sup>用于系统建模; 应用 GALib<sup>[7]</sup>实现遗传算法; 部分聚类分析和可视化程序自行编写并用 JAVA 运行.

## 4 简例

在利用前述方法获得信息的帮助下 本节中将改进遗传算法 IGA 应用于 3 个不同的约束问题: 例 1 取于文献, 例 2 为石油稳定过程工业实例, 例 3 说明了该算法对可行域很小的问题优化有局限性. 所有的例子都对种群进行了 10 次不同的初始化, 并将最优、最差优化结果连同第 1 次得到的收敛种群世代在表格中列出进行对比.

为使遗传算法进行有效的优化, 确定正确的种群大小是很重要的. Goldberg 等<sup>[8]</sup>建议种群大小应随着问题规模的增大而增大, 本研究采用由维数确定的简单方法来计算种群大小, 设置为  $10n(n$  为维数).

例 1: 有 7 个变量、4 个非线性约束<sup>[9]</sup>问题:

$$\begin{aligned}
 &\text{Minimize} && (x_0 - 10)^2 + 5(x_1 - 12)^2 + x_2^4 + 3(x_3 - 11)^2 + 10x_4^6 + 7x_5^2 + x_6^4 - 4x_5x_6 - 10x_5 - 8x_6 \\
 &\text{Subject to} && g_1: 127 - 2x_0^2 - 3x_1^4 - x_2 - 4x_3^2 - 5x_4 \geq 0 \\
 &&& g_2: 282 - 7x_0 - 3x_1 - 10x_2^2 - x_3 + x_4 \geq 0 \\
 &&& g_3: 196 - 23x_0 - x_1^2 - 6x_5^2 + 8x_6 \geq 0 \\
 &&& g_4: -4x_0^2 - x_1^2 + 3x_0x_1 - 2x_2^2 - 5x_5 + 11x_6 \geq 0 \\
 &&& (-10 \leq x_i \leq 10, \quad i = 0, \dots, 6).
 \end{aligned} \tag{3}$$

最优解是 : $X^*=(2.3305, 1.9514, -0.4775, 4.3657, -0.6245, 1.0381, 1.5942)$ ,  $f^*=680.63$ . 研究表明<sup>[9]</sup>该问题的可行域仅占搜索空间的 0.5%.

首先进行  $100/0.005=20\ 000$  次随机模拟, 得到约 100 个可行点, 然后用 CCM 映射法将这些可行点映射到 2 维空间. 图 3 示出得到的聚类信息(聚类数分别给定为 3 和 5). GA 对不同聚类信息的性能见表 2, 比较了 3 个例子:(A) IGA+3, (B) IGA+5 和(C) DebGA.

表 2 例 1 的 GA 结果比较

Table 2 Comparison of GA results for Example 1

Case	Optimal Value			Terminal generation, $n$
	Best	Medium	Worst	
(A) IGA+3	690.17	695.00	704.21	181
(B) IGA+5	680.80	683.07	687.188	252
(C) DebGA	680.63	680.64	680.635	577

由图3和(A)与(B)的GA性能比较可以看出,有5个聚类信息要优于3个聚类. 比较(B)和(C),对IGA和DebGA也可得到与前文相同的结论. 又对5个变量、6个非线性约束问题进行了求解(例子未列出),也得到了与前文和本例相同的结论.

本例中,由于维数大于3,DebGA仅给出了基于目标函数的最优结果,而没有给出任何其它信息,如可行域特性等,来使用户理解算法并接受结果.而IGA不仅给出了相近的较优结果,还提供了关于可行域的可视化信息(图3),且这些由平行坐标系(Parallel Coordinate System, PCS)实现可视化的聚类分析结果(图4)可以帮助用户对可行域信息进一步理解. PCS是指应用  $n$  个平行轴对应  $n$  维坐标,则 1 个  $n$  维点的各维坐标就可由相应平行轴上的对应值标出,这样 1 个  $n$  维点就可表示为有  $n-1$  个节点的连续折线. 在图 4 中,用  $n=7$  个距离均等的平行轴表示 7 维变量,每 1 维对应值  $(\mu - x_i^l)/(x_i^u - x_i^l)$  的平均值( $\mu$ )在相应轴上标出,并用直线连接不同维相邻轴上的平均值,这样,

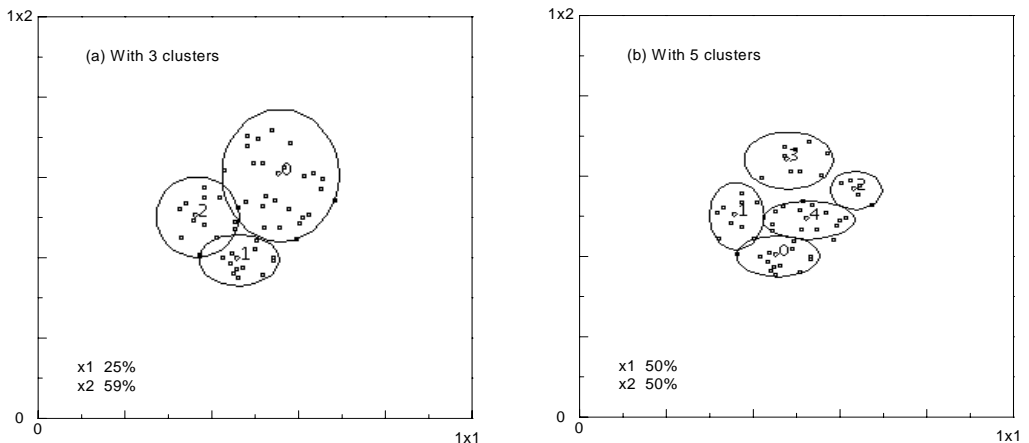


图 3 CCM 映射的聚类分析

Fig.3 The cluster analysis based on CCM mapping

$n$  维空间的中心点就可由折线表示出来,该折线由  $n$  个 1 维线段横跨  $n$  个平行轴. 每个平行轴上用两个短横线表示变量的方差( $\sigma$ )对应值,两横线的中心值是变量的对应平均值( $\mu$ ),每一聚类的下标都在中心点处标出. PCS 表示法可以使用户更深入理解可行空间的概念,例如,由图 4 用户可以直观地看出,在聚类  $O$  中,  $x_1$  的可行域较小;而  $x_2$  和  $x_3$  之间有明显的反比关系等.

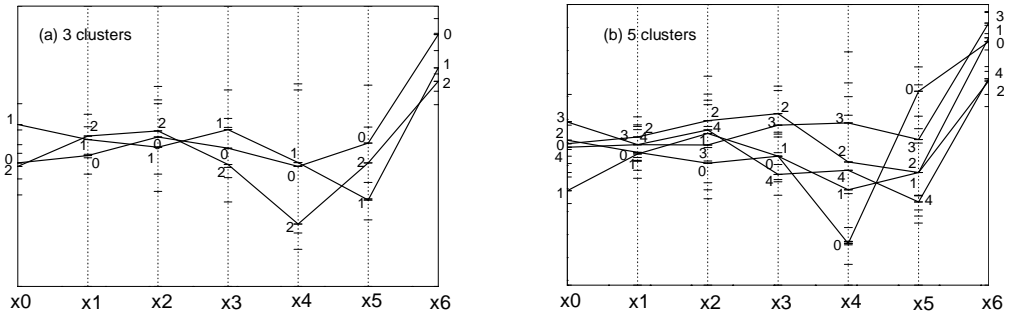


图 4 对可行信息的 PCS 描述

Fig.4 The PCS representation for feasible knowledge

例 2: 用石油稳定过程(Oil Stabilization Process, OSP)<sup>[10]</sup>来说明 IGA 对工业过程优化问题的优化效果. 图 5 给出了流股结构,其中 1, 2, 3, 4 代表闪蒸器, M 是混合器, V 为压力阀,输入流股由 12 个碳氢化合物组成. 有  $n=7$  个连续优化变量: 对每一输入流股的闪蒸温度  $x_0, x_1, x_3, x_4$ , 设定初始搜索范围  $[0, 1]$ ; 阀门的目标压力  $x_5 \in [1, 30] \times 10^5$  Pa, 混合器  $x_2, x_6$  的压力本例定为 3 MPa.

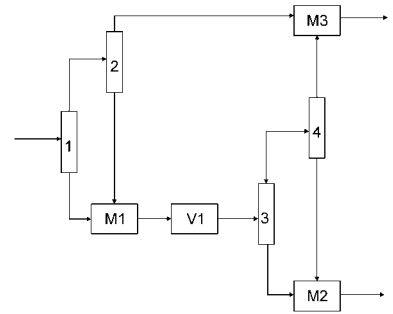


图 5 流股结构图

Fig.5 The flowsheet structure

问题优化目标为费用最小: 操作费用-成品价值. 主要约束是达到油产品所要求的蒸汽压和气体产品中重组分含量的规定要求. 出现不可行点主要原因有二: (1) 过程无效. 闪蒸器的温度可能超出了物性规定范围, 或所用设备不能达到工业要求. 这种情况下费用为无穷大. (2) 在单个过程单元中, 过程有效, 但产品流股组分不满足含量要求, 成品价值设为 0. 所以, 过程费用值将为正值. 对一个可行过程(即满足所有的约束条件), 假定得到了足够量的油产品, 费用将为负值.

用 Jacaranda<sup>[6]</sup>在初始搜索空间(其中 1%是典型可行的)内生成  $10^n$  个( $n=7$ )初始数据, 其 CCM 映射结果如图 6 所示. 基于可视化, 用户可以通过可视化交互作用来调节每一个聚类的半径. 3 个可行域由 PCS 系统表示, 如图 7 得到的 3 个可行域为

- Cluster 0:  $x_0 \in (0.866 \pm 0.089), x_1 \in (0.400 \pm 0.077), x_2 \in (30.000 \pm 0.000), x_3 \in (21.302 \pm 5.422), x_4 \in (0.633 \pm 0.247), x_5 \in (0.433 \pm 0.196), x_6 \in (30.000 \pm 0.000);$
- Cluster 1:  $x_0 \in (0.574 \pm 0.169), x_1 \in (0.582 \pm 0.079), x_2 \in (30.000 \pm 0.000), x_3 \in (16.383 \pm 8.268), x_4 \in (0.695 \pm 0.188), x_5 \in (0.387 \pm 0.193), x_6 \in (30.000 \pm 0.000);$
- Cluster 2:  $x_0 \in (0.687 \pm 0.130), x_1 \in (0.499 \pm 0.109), x_2 \in (30.000 \pm 0.000), x_3 \in (15.347 \pm 7.611), x_4 \in (0.479 \pm 0.151), x_5 \in (0.614 \pm 0.106), x_6 \in (30.000 \pm 0.000).$

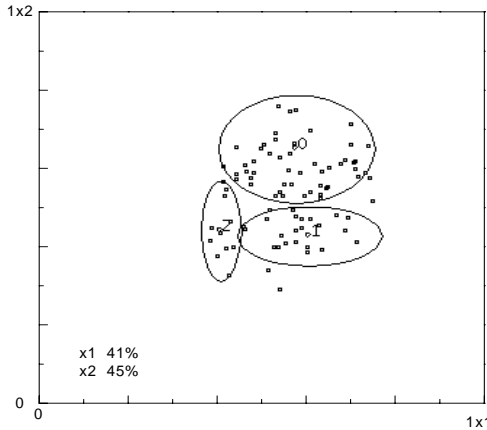


图6 基于3个聚类 CCM 映射的聚类分析  
Fig.6 The cluster analysis based on CCM mapping with 3 clusters

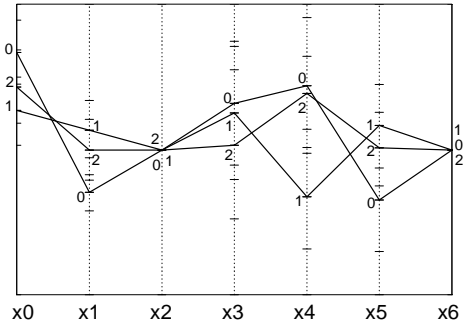


图7 对3个聚类可行信息的 PCS 描述  
Fig.7 The PCS representation for 3 clusters feasible knowledge

通过模拟，用户可以知道：

(1) 在聚类分析的辅助下，可行搜索域明显缩小了。如  $x_0$  的搜索域从(0,1)缩减到(0.574±0.169), (0.687±0.130)和(0.866±0.089)。

(2) 可行搜索域是基于变量间关系的。比较图7中不同的聚类信息可知，若  $x_0$  被缩小了，那么  $x_1$  就应增加来保持其可行性。对  $x_4$  和  $x_5$  有同样的规则。

(3)  $x_1$  的最小可行范围约为整个给定搜索域的 30%。可以看出， $x_1$  是敏感变量。在该例中，由 PCS 描述可直观得到变量的敏感顺序为  $x_1 > x_0 > x_5 > x_4 > x_3 > (x_2, x_6)$ 。

为了评价利用搜索空间信息的效率，考虑应用 IGA。考虑两个例子 (A) DebGA; (B) IGA+3。表3给出了 GA 性能，其中，(A)为了收敛需要更多的世代数，但若考虑到寻求信息所需的额外计算工作量，那么(B)的计算量更大。在(A)中，DebGA 没能找到最优解，原因是本例所给出的约束并不像前面的例子一样明确，无法计算违反约束惩罚量，这限制了 DebGA 在实际工业问题上的应用。本例表明改进的 GA 对实际工业化问题的优化效果较优。

表3 GA 结果比较  
Table 3 Comparison of GA results

Case	Optimal value			Terminal generation
	Best ( $\times 10^6$ )	Medium ( $\times 10^6$ )	Worst ( $\times 10^6$ )	
(A) DebGA	6.50	6.53	6.60	30
(B) IGA+3	-363	-361	-357	26

例3：有13个变量和9个非线性约束的最小值问题，最优解(1, 1, 1, 1, 1, 1, 1, 1, 3, 3, 3, 1)，最优值为-15。

$$\begin{aligned}
 &\text{Minimize} && 5\sum_{i=1}^4 x_i - 5\sum_{i=1}^4 x_i^2 - \sum_{i=5}^{13} x_i \\
 &\text{Subject to} && g_1 : 2x_1 + 2x_2 + x_{10} + x_{11} \leq 10, & g_2 : 2x_1 + 2x_2 + x_{10} + x_{12} \leq 10, \\
 & && g_3 : 2x_1 + 2x_2 + x_{11} + x_{11} \leq 10, & g_4 : -8x_1 + x_{10} \leq 0, & g_5 : -8x_2 + x_{11} \leq 0 \\
 & && g_6 : -8x_3 + x_{12} \leq 0, & g_7 : -2x_4 - x_5 + x_{10} \leq 0 \\
 & && g_8 : -2x_6 - x_7 + x_{11} \leq 0, & g_9 : -2x_8 - x_9 + x_{12} \leq 0 \\
 & && 0 \leq x_i \leq 1, \quad i = 1, \dots, 9; & 0 \leq x_i \leq 100, \quad i = 10, 11, 12; & 0 \leq x_{13} \leq 1.
 \end{aligned} \tag{4}$$

本例可行域很小，仅由随机模拟很难生成可行点。但由约束  $g_1, g_2, g_3$ ，容易对可行域的  $x_{10}, x_{11}$ ,

$x_{12}$  从(0, 100)重新定义为(0, 10), 由此生成初始点  $10^{13}$  个. 再用式(2)判断生成可行点, 迭代 48 代后得到与 DebGA<sup>[9]</sup>结果相同的最优值(-15). 本例说明了 IGA 对可行域很小的问题优化有局限性.

## 5 结语

表明了前文提出的基于数据模拟与分析的 CCM 聚类分析法的有效性. 通过实例可以看出, 该法提供了及时直观的可行域信息, 且用户可以直接参与到决定聚类分析的参数设置之中. 通过这种交互模拟提高了发现可行域的效率. 另外, 由平行坐标系统(PCS)模拟得到的信息, 可使用户对可行域有更深入的理解. 很明显, 本研究所提出的基于模拟、聚类分析相结合的 IGA, 在大部分例子中, 不仅高效(所得解更接近最优解), 鲁棒性好(GA 对不同的随机数种子通过迭代相近的世代数找到的问题解都接近已知最优解), 而且还为用户提供了更多关于可行域的信息, 使用户理解模型并接受优化结果.

例 3 中 IGA 在处理可行域很小的问题上有局限性, 在结合使用 IGA 和 DebGA 中将进行改进. 参考文献:

- [1] 王克峰, 孙晓静, 姚平经. 基于聚类分析和可视化的增强遗传算法—I. 算法的引出、原理与分析 [J]. 过程工程学报, 2004, 4(5): 438-444.
- [2] Ward J H. Hierarchical grouping to optimize an objective function [J]. J. Am. Statist. Assoc., 1963, 58: 236-244.
- [3] Hartigan J A. Clustering Algorithms [M]. New York: John Wiley & Sons, 1975.
- [4] Pao Y H. Dimension Reduction, Feature Extraction and Interpretation of Data with Network Computing [J]. Int. J. Pat. Recognition Artif. Intell., 1996, 1(5): 521-535.
- [5] Wegner P. Why Interaction is Not More Powerful than Algorithms [J]. Comm. ACM, 1997, 40(5): 80-91.
- [6] Fraga E S, Steffens M A, Bogle I D L, et al. Foundation of Computer-aided Process Design [J]. AIChE Symposium Series, 2000, 96: 446-449.
- [7] Wall M. GALib: A C++ Library of Genetic Algorithm Components [CP/OL]. <http://lancet.mit.edu/ga>, 1999-12-12.
- [8] Goldberg D E, Deb K, Clark J H. Genetic Algorithms, Noise, and the Size of Population [J]. Complex Systems, 1992, 6: 333-362.
- [9] Deb K. An Efficient Constraint Handling Method for Genetic Algorithms [J]. Comput. Methods Appl. Mech. Eng., 2000, 186: 311-338.
- [10] McCarthy E, Fraga E S, Ponton J W. An Automated Procedure for Multicomponent Product Separation Synthesis [J]. Comput. Chem. Eng., 1998, 22(Suppl.): s77-s84.

## Cluster Analysis and Visualization Enhanced Genetic Algorithm —II. Analysis of Cases and Validation

SUN Xiao-jing, WANG Ke-feng, YAO Ping-jing

(Inst. Chem. Process Systems Eng., Dalian University of Technology, Dalian, Liaoning 116012, China)

**Abstract:** This paper validated that the Cluster Constrained Mapping (CCM) can keep the “topological” information of the points in the reduced dimension map by comparing the cluster results obtained using the  $K$ -means algorithm. The enhanced GA proposed in Part I was applied to three constrained optimization cases. The results show that the combination of visualization, cluster analysis and genetic algorithms can help users to participate in selecting appropriate parameters of clusters, and the combination of a computer and the user is more powerful than either alone, which is an effective process optimal design tool with high solution quality and consistency. In the new cluster analysis method, the data are visualized by CCM that provides immediate direct information about the feasible domain, and the user is directly involved in determining the parameters for the cluster analysis and increasing the effectiveness of feasible regions discovery by visual interaction; the obtained knowledge is visualized by Parallel Coordinate Systems (PCS), thus the user has a deeper understanding of the feasible regions. It is clear that in most cases the proposed IGA based on the combination of visualization and cluster analysis has performed not only with the high efficiency (in terms of getting closer to the best-known solution) and with more robustness (in terms of the number of GA runs finding solutions close to the best known solution), but also with providing more information about the feasible regions for the user to understand the model and accept the optimal results.

**Key words:** visualization; cluster analysis; constrained optimization; genetic algorithm