第15卷 第5期 自 动 化 学 报 Vol. 15, No. 5
1989 年 9 月 ACTA AUTOMATICA SINICA Sept., 1989

微机实现非线性映照

罗学才 程兆年 汤锋潮 缪 强 陈念贻
(中国科学院上海冶金研究所)

摘 要

本文提出了在微机上实现非线性映照的下山法和逐步松弛法相结合的计算方法。编制了在 IBM-PC 机上运行的程序。双倍精度计算，最大允许样本集250个。程序结构采用积木块式，结合菜单选择，使用方便灵活。

关键词：非线性映照，模式识别，微机应用。

随着计算机的广泛应用，模式识别已不仅是一种设计建立自动识别机器的基础方法，还广泛应用于图谱分析[1]、物料配方[2]、地质勘探[3]、文物鉴定[4]、医疗诊断[5]等各种自动测试分析领域。由模式识别应用于工业诊断引伸的优化决策，也开始在国民经济中直接发挥作用[6]。

交互模式识别可直接在计算机屏显图中引入人的判断，在一定程度上可弥补人们对某些概率分布知识的缺乏。虽然面向显示的各种线性映照方法计算简单，但较多损失了样本分布的信息。非线性映照则可以较真实地反映模式空间中样本点的聚集状况，有较大的实际意义。

一、面向显示的非线性映照

设有一组 n 个样本的数据集，每一样本用一 m 维矢量表示，其矢量集构成矩阵 $[X]_{n\times m}$，$X_i=(x_{i1}, x_{i2}, \cdots, x_{im})$，$i=1, 2, \cdots, n$。作非线性映照后，$[X]_{n\times m}\rightarrow[Y]_{n\times 2}$，$Y_i=(y_{i1}, y_{i2})$ 为二维矢量。

映照前，样本点 i 与 j 间欧氏距离为

$$d_{ij}^{*}=\|X_i-X_j\|_2=\left[\sum_{k=1}^{m}(x_{ik}-x_{jk})^2\right]^{\frac{1}{2}}. \tag{1}$$

映照到二维平面后，i 与 j 间欧氏距离为

$$d_{ij}=\|Y_i-Y_j\|_2=\left[\sum_{k=1}^{2}(y_{ik}-y_{jk})^2\right]^{\frac{1}{2}}. \tag{2}$$

Sammon 定义误差函数为[7]

本文于1987年6月9日收到。

$$E=\frac{1}{\sum_{i<j} d_{ij}^{*}}\sum_{i<j}\frac{(d_{ij}^{*}-d_{ij})^{2}}{d_{ij}}. \tag{3}$$

问题归结为寻找$[Y]_{n\times 2}$使E达到极小或尽可能小.常采用迭代法,先选取初始$[Y]_{n\times 2}$,计算E的初始值 E_0，再调整 $[Y]_{n\times 2}$ 使新的E值小于初值 E_0。 如此循环使E值不断变小,直至达到下述三个终止准则之一[8]：(1) E已达到预先给定的值；(2) 迭代已达到预定次数；(3) 当前的分布已使观察者满意. 准则(3)很有实际意义.

二、计 算 方 法

Sammon[7] 提出下山法寻找E的最小值。设 Y_i，Y_i' 分别代表某一步迭代前后的样本点i，则按下山法可得:

$$y_{ik}'=y_{ik}-\alpha\frac{\partial E}{\partial y_{ik}}\Big/\left|\frac{\partial^{2}E}{\partial y_{ik}^{2}}\right|. \tag{4}$$

收敛因子α为一经验常数. 式中 $i=1,2,\cdots,n$; $k=1,2$. 并且,

$$\frac{\partial E}{\partial y_{ik}}=-\frac{2}{c}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{d_{ij}^{*}-d_{ij}}{d_{ij}^{*}d_{ij}}(y_{ik}-y_{ik}), \tag{5}$$

$$\frac{\partial^{2}E}{\partial y_{ik}^{2}}=-\frac{2}{c}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{d_{ij}^{*}d_{ij}}\left[(d_{ij}^{*}-d_{ij})-\frac{d_{ij}^{*}}{d_{ij}^{2}}(y_{ik}-y_{ik})^{2}\right], \tag{6}$$

$$c=\sum_{i<j} d_{i}^{*}. \tag{7}$$

Chang 和 Lee[9] 提出松弛法寻找E的极小值。 为适合于微机，此处采用逐步松弛法[10]. 下山法计算一步，所有 Y_i 一起变动；而逐步松弛法计算一步只变动 $\left\|\frac{\partial E}{\partial Y_i}\right\|_2^2=\left(\frac{\partial E}{\partial y_{i1}}\right)^2+\left(\frac{\partial E}{\partial y_{i2}}\right)^2$ 为最大的一个样本点 Y_p. 首先,确定 $\left\|\frac{\partial E}{\partial Y_i}\right\|_2^2$ 最大的点为被松弛点p，然后使被松弛点沿牛顿方向移动 dY_p:

$$\begin{cases} y_{p1}'=y_{p1}+\beta dy_{p1},\\ y_{p2}'=y_{p2}+\beta dy_{p2}, \end{cases} \tag{8}$$

$$\begin{pmatrix} dy_{p1}\\ dy_{p2}\end{pmatrix}=\begin{pmatrix} \frac{\partial^{2}E}{\partial y_{p1}^{2}} & \frac{\partial^{2}E}{\partial y_{p1}\partial y_{p2}}\\ \frac{\partial^{2}E}{\partial y_{p1}\partial y_{p2}} & \frac{\partial^{2}E}{\partial y_{p2}}\end{pmatrix}^{-1}\begin{pmatrix} \frac{\partial E}{\partial y_{p1}}\\ \frac{\partial E}{\partial y_{p2}}\end{pmatrix}. \tag{9}$$

式中收敛因子β亦为一经验常数. $\frac{\partial^{2}E}{\partial y_{pk}^{2}}$ 由(6)式给出,

$$\frac{\partial^{2}E}{\partial y_{p1}\partial y_{p2}}=-\frac{2}{c}\sum_{\substack{j=1\\ j\neq p}}^{n}\left[-\frac{1}{d_{pj}^{3}}(y_{p1}-y_{j^1})(y_{p2}-y_{j^2})\right]. \tag{10}$$

松弛一步后,误差函数 $E'=E-\Delta E$,

$$\Delta E = \frac{1}{c}\sum_{\substack{j=1\\ j\neq p}}^{n}\frac{1}{d^*_{pj}}(d'_{pj}-d_{pj})(2d^*_{pj}-d_{pj}-d'_{pj}). \tag{11}$$

显然，松弛一步要比下山一步节省机时．但由于仅改变一点 p，误差函数 E 的减少也小得多．分析两种算法的计算量可知，逐步松弛法计算一步的时间约为下山法的$\frac{1}{10}$—$\frac{1}{20}$（与 n 有关）．实际计算表明，在计算早期下山法收敛较快，在计算后期逐步松弛法收敛较快．因此，先使用下山法，当 E 下降到一定程度时再使用逐步松弛法是合理的．

综上所述，我们提出如下计算方法：（1）用主成分分析第一、第二主成分的线性映照结果为初始 $[Y]_{n\times 2}$；（2）用下山法计算若干步，一般可计算到二步间的误差函数差 ΔE 小于 E_0 的百分之一；（3）最后用逐步松弛法计算至结束．

笔者编制了基于上述方法的程序，程序结构采用积木块式．由 LOTUS 1-2-3 数据库软件产生数据文件．程序中设置若干选择菜单，用以选择计算方法、收敛因子和终止准则，使用方便、灵活．缺省项的算法是：先用下山法，在 ΔE 小于 E_0 的百分之一后进入松弛法，ΔE 小于 E_0 的万分之一时终止．程序运行时可通过键盘随时中断计算并屏显中间结果，并可通过菜单选择是否继续计算，直至结果满意．屏显可显示样本的类别，也可显示样本编号．采用专门打印软件打印图形．

三、计 算 实 例

图 1 和图 2 给出了 38 个癌组织与溃疡组织七种微量元素含量的 P_1-P_2 主成分的线性映照结果与非线性映照结果．非线性映照计算约需 10 分钟．比较图 1 和图 2 看出非线性映照比线性映照在聚类效果上有显著改善．图 2 中虚线给出了两类的目测边界，仅一个样本点不在此边界内，说明七种微量元素含量确能构成癌与溃疡的特征模式．

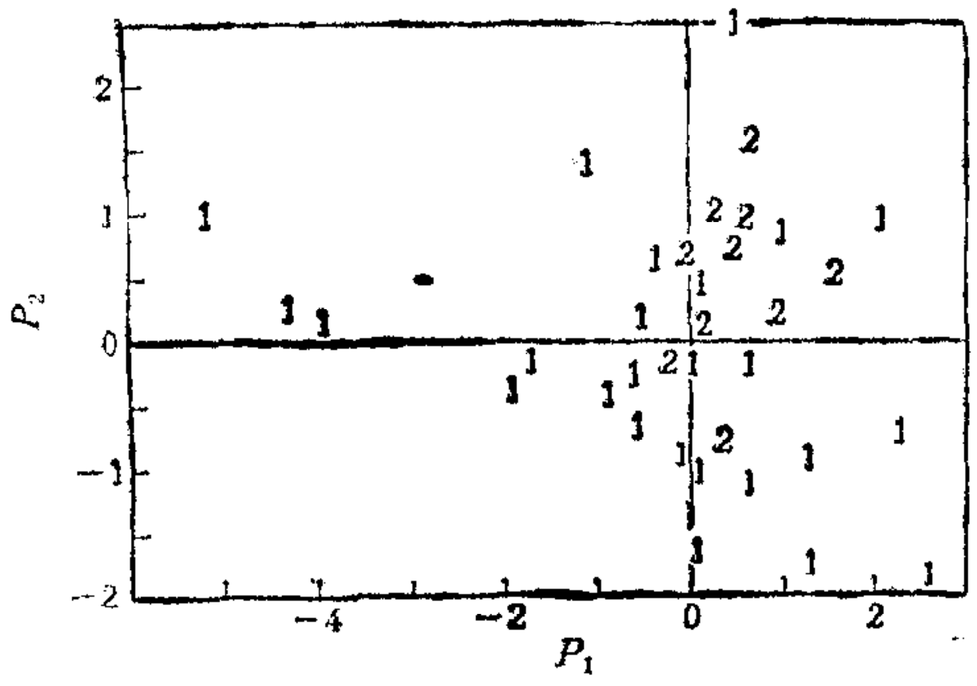

图 1 38 个癌组织与溃疡组织 P_1-P_2 平面映照结果

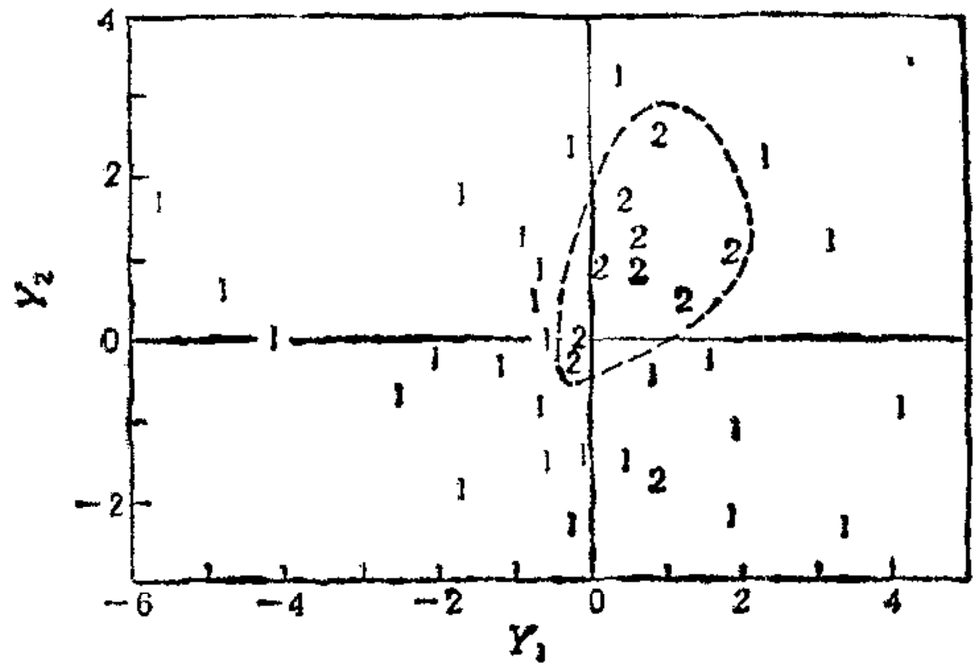

图 2 38 个癌组织与溃疡组织非线性映照图

图 3 是某厂 41 种橡胶配方（十种物料含量）按橡胶性能分类的非线性映照图．由图可见，优质配方（用 1 代表）与非优质配方（用 2 代表）确处于所选十个变量构成的模式空间的不同区域．说明所选的十个成分含量确能表征橡胶的性能．结合其它方法，有可能设计出更优质的配方．

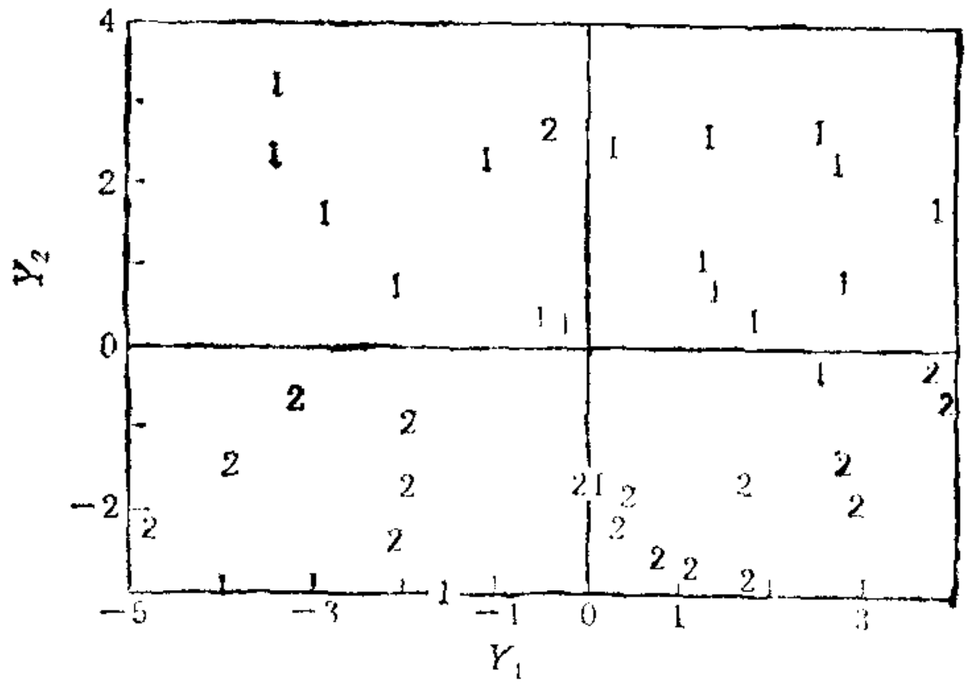

图 3　橡胶配方的非线性映照结果

参　考　文　献

[1] Kowalski, B.R., Anal. Chem., **47**(1975), 1152A.
[2] Kowalski, B. R., Chem. Technol., **4**(1974), 300.
[3] 徐驰等,科学通报, **22**(1986),1758.
[4] 叶礼萍等,第二届全国计算化学会议论文集, 77,1987.
[5] 陈念贻等,科学通报, **22**,(1986),1518.
[6] 张未名等,自动化学报, **15**(1989), No.1, 1—7.
[7] Sammon, J. W., *IEEE Trans. Computers*, **C-18**(1969), 401.
[8] 李介谷等编,计算机模式识别技术,上海交通大学出版社 (1986),208.
[9] Chang, C. L. and Lee, R. C., *IEEE Trans. Syst. Man. Cybern.*, **SMC-3** (1973), 197.
[10] 上海计算所编,电子计算机计算手册,上海教育出版社 (1982),879.

THE REALIZATION OF NONLINEAR MAPPING ON MICROCOMPUTER

Luo Xuecai　Cheng Zhaonian　Tang Fengchao　Miao Qiang　Chen Nianyi

(Shanghai Institute of Metallurgy, Academia Sinica)

Abstract

In this paper, we suggest a new algorithm in which the steepestdescent procedure is combined with the relaxation procedure for nonlinear mapping on microcomputer. Based on this algorithm we have written a program run on microcomputer. The allowed number of samples is extended to 250 with double precision. Using modular construction, the program contains a set of optional menus and can be used flexibly and conveniently.

Key words ——Nonlinear mapping; pattern recognition; application of microcomputer.