

用模糊聚类法研究中药成分特征谱

周敬泉¹, 颜春兰², 袁 鹏¹, 曾启杰¹, 袁 敏³, 张铭光⁴

(1. 华南理工大学 自动化科学与工程学院, 广东 广州 510640; 2. 广东粤安信息工程有限公司, 广东 广州 510630;

3. 华南师范大学 分析测试中心, 广东 广州 510631; 4. 华南师范大学 生命科学学院, 广东 广州 510631)

摘要: 为了寻求和评价不同产地中药裂解气相色谱图之间的相似关系和差异程度, 建立指纹图谱, 选取了 10 个郁金样品的裂解气相色谱图, 提取图谱中反映不同样本在化学成分和含量上有差异的信息特征(保留时间、相对浓度、峰面积等), 采用模糊聚类 6 种不同分析法在 16 个峰参数上进行分类, 得到简洁明了的模糊聚类图. 实验结果反映了图谱之间的相似关系, 并能快速准确地选出具有代表性的图谱.

关键词: 模糊聚类; 郁金; 裂解气相色谱; 指纹图谱

中图分类号: TP399 **文献标识码:** A

Research and analysis of componential characteristic spectrum of Chinese traditional medicine with fuzzy clustering

ZHOU Jing-quan¹, YAN Chun-lan², YUAN Peng¹, ZENG Qi-jie¹, YUAN Min³, ZHANG Ming-guang⁴

(1. College of Automation Science & Engineering, South China University of Technology, Guangzhou Guangdong 510640, China;

2. Guangdong Yue'an Information Engineering Co., LTD, Guangzhou Guangdong 510630, China;

3. Center of Analysis & Test, South China Normal University, Guangzhou Guangdong 510631, China;

4. Life Science College, South China Normal University, Guangzhou Guangdong 510631, China)

Abstract: 10 Pyrolysis gas chromatography samples of Curcuma aromatica Salisb were selected to evaluate the resemblance and difference of pyrolysis gas chromatography of Chinese traditional medicine coming from different areas and found fingerprint spectrum. Useful information was collected to reflect the difference of the samples in chemical component and content, such as the remaining time, relative concentration and peak acreage. They were classified by 6 fuzzy clustering analysis methods in 16 peak parameters. Simple and clear fuzzy clustering charts were gotten. The results show the resemblance of spectrums, and a quick, correct choice can be achieved of typical chromatography.

Key words: fuzzy clustering; Curcuma aromatica Salisb; pyrolysis gas chromatography; fingerprint spectrum

1 引言(Introduction)

在生物医学上, 利用裂解气相色谱(pyrolysis gas chromatography, PGC)技术和指纹图谱(fingerprint spectrum, FS)技术进行中药的真伪鉴别和质量评价, 近年来受到了广泛的重视, 被认为是对图谱进行宏观分析的可行手段之一. 由于中药成分具有复杂性和多样性, 尽管同一种中药的裂解气相色谱指纹图谱(PGC-FS)具有相似性, 但很难获得两张完全一样的图谱. 借助于样本数据的相似和差异信息, 建立相应的模式识别法, 实现图谱的辨认和鉴别是可行的, 但就具体方法而言, 目前都是视具体情况灵活而定, 并无普遍适用的方法^[1].

由于中药裂解气相色谱所反映的信息极为复杂, 而且指纹图谱技术对于图谱中的各峰, 并不要求每个组分的化学结构都清楚, 也不需要呈现在图谱中的每个组分都清楚地定量, 仅需对同批号的同一种药品的指纹图谱基本保持一致, 也就是说, 图谱之间的关系具有很大的模糊性, 因此, 采用模糊聚类的方法来分析显得更自然、更符合客观规律.

近年来, 模糊聚类法在生物医学上的应用研究日益增多: 赵蔡斌等人利用模糊聚类法讨论了小柴胡汤的配伍规律, 得到与传统中医理论完全相符的结论^[2]; 董顺福等人利用模糊聚类分析了治疗心血管疾病的中药中无机元素与疗效的关系^[3]; 马玉

荣等人利用模糊聚类法进一步评价了药品“尼群地平片”的质量^[4].但是将模糊聚类分析应用于测定指纹图谱的研究,即研究图谱之间的相关系数,进而评价图谱间的相似和差异程度,报道还不多.将模糊聚类分析法应用于郁金裂解气相色谱测定指纹图谱,其数据处理结果表明,以模糊聚类分析法来解析图谱,进而建立指纹图谱的方法是有效可行的;模糊聚类的结果基本反映了图谱之间的相似关系,并能快速准确地选出具有代表性的图谱.

2 基于模糊聚类法的郁金裂解气相色谱测定指纹图谱的研究 (Research method of determining Curcuma aromatica Salisb's fingerprint spectrum by pyrolysis gas chromatography based on fuzzy clustering)

一般情况下,模糊聚类分析分三个步骤进行:第一步,预处理和标准化.预处理的目的是选择最具代表性的特征参与模糊聚类.标准化的目的是消除量纲的影响,并使数据分布在 $[0,1]$ 的闭区间里;第二步,求模糊相似关系. $R(x_i, x_j) = (r_{ij})_{n \times n}$. r_{ij} 表示被分类对象之间相似程度,采用何种方法来求 r_{ij} 对模糊聚类的结果有决定性的影响;第三步,聚类.把模糊关系矩阵改造成模糊等价关系矩阵后,在适当的阈值 λ 上进行截取,得到所需的分类,并可根据这个过程画出模糊聚类图.

具体做法是在裂解气相色谱法测定郁金指纹图谱的研究中,设 $X = \{x_1, x_2, \dots, x_n\}$ 为待分类的 n 张郁金裂解气相色谱图(例如,4号郁金裂解气相色谱图如图1所示),由于裂解峰众多,而且每张图谱中峰的各项参数并不完全一致,必须根据保留时间、浓度和峰面积等参数进行预处理,从而筛选出能代表图谱特征的一组参数 $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$ 作为参与模糊聚类的数据矩阵.

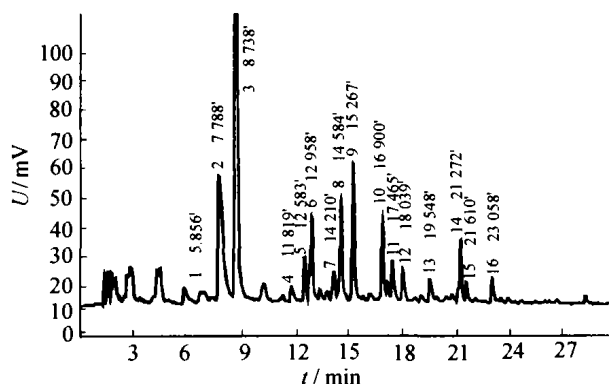


图1 4号郁金样品的裂解色谱图

Fig. 1 PGC of No.4 curcuma aromatica salisb

标准化可分为两步完成:先做平移-标准差变换,再做平移-极差变换.但是从实验的结果来看,由于预处理后所有的参数都具有相同的量纲,因而可不经平移-标准差变换,而是直接做平移-极差变换:

$$x'_{ik} = (x_{ik} - \min_{1 \leq i \leq n} \{x_{ij}\}) / (\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}).$$

其中, x'_{ik} 是标准化后的变量,且 $x'_{ik} \in [0,1]$.

计算 r_{ij} 的方法有很多,采用不同方法进行模糊聚类,其结果是不同的.在实际应用时,应先采用多种方法进行实验,再选取分类最符合实际的算法.(例如,分别采用数量积法、相关系数法、最大最小法、算术平均最小法、几何平均最小法和绝对值减数法的计算公式,对10种郁金裂解气相色谱图进行模糊聚类分析,结果如图2所示.)

以上得到的模糊相似关系矩阵一般说来只满足自反性和对称性,而不满足传递性.即 R 只是相似矩阵,还需将它改造成模糊等价矩阵,也就是求传递闭包 $t(R)$. 采用逐次平方法来求比较方便快捷,计算公式为

$$R \Rightarrow R^2 \Rightarrow (R^2)^2 = R^4 \Rightarrow \dots \Rightarrow R^{2^k} = t(R),$$

其中, k 为自然数.

$$R^2 = \{r_{ij}^{(2)}\} = \bigvee_{k=1}^n [r_{ik} \wedge r_{kj}],$$

$$R^4 = \{r_{ij}^{(4)}\} = \bigvee_{k=1}^n [r_{ik} \wedge r_{kj}], \dots,$$

直到 $R^{2^k} = R^k$, 计算完成.

最后把传递闭包矩阵 $t(R)$ 中不同元素的值按从大到小的顺序依次排列为 $1 = \lambda_1 > \lambda_2 > \dots > \lambda_t$. 可以得到一系列 λ 的截矩阵, 即 $t(R)_\lambda \triangleq (\bar{r}_{ij}(\lambda))_{n \times n}$, 其中,

$$\bar{r}_{ij}(\lambda) = \begin{cases} 1, & \bar{r}_{ij} \geq \lambda, \\ 0, & \bar{r}_{ij} < \lambda. \end{cases}$$

对于对象 x_i, x_j , 若 $\bar{r}_{ij}(\lambda) = 1$, 则在水平 λ 上将它们归为一类, 从而达到将对象集进行模糊聚类的目的. 在适当的 λ 值上进行截取, 就可得到所需的分类. 随着水平 λ 在 $[0,1]$ 中发生改变, 相应的分类也发生变化. 当 λ 由1逐步降到0时, 类的个数由多变少, 而类中的元素由少变多, 逐步归并, 最后所有对象都成一类. 根据以上分析, 可以画出模糊聚类图, 此图能简洁明了地表示模糊聚类的结果(见图2).

3 实验结果与讨论 (Experiment result and discussion)

在样品质量有保障的基础上, 测定10种经厂家筛选过质量较好的郁金药材, 其编号与产地来源见

表 1,4 号郁金的裂解气相色谱图如图 1 所示.

表 1 郁金样品来源

Table 1 Samples' source

编号	样品来源	编号	样品来源
1	浙江陶山镇	6	广州天河
2	浙江陶山镇	7	广州天河
3	四川双流	8	河北安国
4	广东电白县	9	河北安国
5	广西桂林	10	广西玉林

从实验结果可见,10 个样品的图谱既有相似性

表 2 郁金裂解气相色谱指纹图谱参数值

Table 2 Parameter vaule of curcuma aromatica salib's PGC-FS Sr/(%)

峰号	α	样 品									
		1	2	3	4	5	6	7	8	9	10
1	0.22	11.32	10.86	12.45	17.60	12.34	11.13	11.57	12.11	10.70	9.53
2	0.31	39.12	30.81	46.42	36.42	50.76	31.70	27.40	37.29	41.83	41.54
3	0.64	1.63	2.53	1.88	1.83	1.34	2.35	2.43	2.85	1.78	1.65
4	0.72	2.70	2.65	1.90	1.94	0.69	2.40	2.94	4.10	2.65	1.78
5	0.75	10.89	5.47	5.88	6.92	2.01	4.47	4.28	4.15	4.12	2.75
6	0.89	2.41	3.70	2.04	2.43	3.33	5.17	4.83	3.15	5.56	4.56
7	0.93	5.30	7.72	6.68	6.75	9.78	11.08	10.26	8.98	9.89	10.24
8	1	8.83	10.30	6.73	7.93	5.10	8.47	11.50	8.54	7.14	6.44
9	1.17	4.96	8.13	4.06	4.38	5.39	9.75	8.23	4.03	4.74	7.34
10	1.23	2.93	3.83	2.29	3.01	1.85	3.79	3.25	3.48	3.05	2.63
11	1.29	1.74	2.64	1.85	2.45	1.13	2.20	3.05	2.78	1.60	2.14
12	1.45	1.53	3.82	1.97	1.94	1.82	2.09	4.22	2.32	1.66	3.13
13	1.63	4.07	3.51	3.65	3.68	2.61	2.72	3.69	3.73	3.26	3.26
14	1.67	1.07	1.91	1.11	1.27	1.05	1.64	2.32	1.71	1.13	1.74
15	1.82	1.51	2.51	1.48	1.43	0.79	1.23	2.02	1.36	0.91	1.59

利用 Matlab 软件编程可直接得出采用不同方法的模糊聚类结果,如图 2 所示.图中竖轴表示相似系数,横轴表示图谱编号.在一定的相似系数下,划分为同一类的图谱被认为是相似的.相似系数越接近 1,同一类中的裂解气相色谱图之间的相似程度就越大,把它们分为一类的置信度越大.也就是说,按照指纹图谱技术要求,以指定参考的指纹图谱为基准,对其它样品的指纹图谱进行比较,相似系数越大,则建立指纹图谱的可靠性越高.

本文根据实验结果,从实际出发,经过比较,发现采用最大最小法、算术平均最小法及几何平均最小法等方法来建立模糊相似矩阵,能够正确地反映郁金裂解气相色谱图之间的相似关系,其中最大最小法和算术平均最小法的聚类结果很相似,见

又有特殊性,郁金的裂解峰有近 30 个色谱峰,10 种样品的共有峰占 90% 以上,选择 Sr(各组分峰面积与共有峰面积之和的比值) > 1% 的共有峰进行比较,发现不同产地的郁金其指纹图谱主要变化是组分间峰值强度的改变.

选 10 种样品都共有、峰值较稳定的 8 号峰为参照峰,设其 α 值为 1,分别求各组分的 α 值(各组分的调整保留值与峰 8 的调整保留值的比值).由 α 值和 Sr 值构成的郁金裂解气相色谱指纹图谱参数作为参与模糊聚类分析的数据矩阵,见表 2.

图 2(c) ~ (f).

可见无论采用哪种方法,所有样品的相似系数都很大,说明样品质量相对很稳定,没有差异很大的图谱.结合表 1,地域因素对于图谱有一定的影响,6 和 7 号样品、8 和 9 号样品来自同一产地,5 和 10 号样品产地相近,它们图谱之间的相似系数相对大些,但 1 和 2 号样品可能受其它因素的影响较大,没有相对较大的相似系数.

以图 2(c)为例,在相似系数为 $0.85 \geq \lambda > 0.84$ 时,10 种郁金可分为 6 类,5,10,8,9 号样品为一类,6,7 号样品为一类,3,1,4,2 号样品独自各成一类.在相似系数为 $0.82 \geq \lambda > 0.81$ 时,可分为两类,5,10,8,9,3,1,4 号样品为一类,其余的为另一类.当相似系数为 $0.81 \geq \lambda > 0$ 时,全部样品都归为一类.

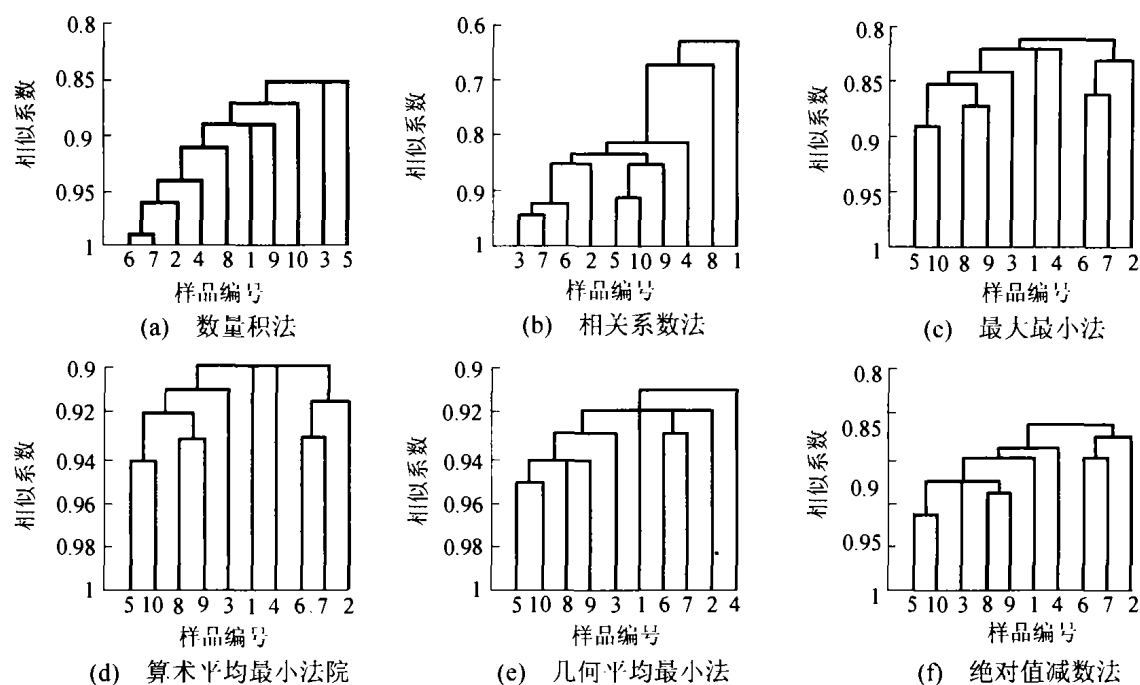


图2 采用各种不同方法得出的模糊聚类图
Fig. 2 Fuzzy clustering chart used various methods

以上的郁金裂解气相色谱图经过证明,完全达到指纹图谱的技术要求,并且其结果与水蒸汽提取挥发油的指纹图谱分析结果基本一致.总的说来,用模糊聚类的方法能比较全面、综合地反映不同产地郁金裂解气相色谱之间的相似关系,可为指纹图谱的建立提供有力的依据.

4 结论(Conclusion)

指纹图谱技术是使用多学科交叉、综合技术手段对物质组成体系质量稳定性进行评价的检测方法,是目前国内外生物医药界研究的热点和发展趋势.但由于其分析与运算的复杂性,人们正寻求更先进的技术和更强大的数据处理能力来实现.目前建立指纹图谱的方法不少,但是能将色谱技术、人工智能技术和计算机技术融合为一体,实现快速、简便、精确判断的方法还不多.

本文采用模糊聚类分析方法,提取图谱中反映不同样本在化学成分和含量上有差异的信息特征(保留时间、相对浓度和峰面积等),并对图谱进行综合评价,得到了一系列与相似系数对应的分类.整个过程采用 Matlab 6.5 编程实现,将图谱信息转化为计算机能够接受的数量化矩阵,借助于计算机强大的数据处理能力,快速、准确地求出相似关系并画出简洁明了的模糊聚类图.结果表明,此法基本上可为郁金裂解气相色谱测定指纹图谱的分析和鉴别提供一种直观、简便、快捷的方法.

此外,作者还做了如广藿香、郁金等其它中药的

裂解色谱指纹图谱、气相色谱指纹图谱等数据的分析,实验结果表明模糊聚类分析法对于相似性、重复性、质量控制等方面都可获得较好的效果.

参考文献(References):

- [1] 苏薇薇.计算机辅助的色谱、光谱——聚类分析法在中药鉴定中的应用[J].广东药学院学报,1996,12(2):99-101.
(SU Weiwei. Chromatogram and spectrogram assisted by computer: Application of clustering analysis in appraising Chinese traditional medicine [J]. *J of Guangdong College of Pharmaceutical*, 1996, 12 (2): 99-101.)
- [2] 赵蔡斌,周鲁,付超,等.中药复方的模糊分析[J].中国实验方剂学杂志,2003,9(2):62-63.
(ZHAO Caibin, ZHOU Lu, FU Chao, et al. Fuzzy analysis of Chinese traditional medicine compound [J]. *Chinese J of Experimental Traditional Medicine Formulae*, 2003, 9(2): 62-63.)
- [3] 董顺福,韩丽琴,朱志国,等.治疗心血管疾病中成药中无机元素模糊聚类分析[J].中成药,2003,25(4):316-319.
(DONG Shunfu, HAN Liqin, ZHU Zhiguo, et al. Fuzzy clustering analysis of inorganic elements in traditional Chinese medicines cured for vascular disease [J]. *Chinese Traditional Patent Medicine*, 2003, 25(4): 316-319.)
- [4] 马玉荣,杨梁,郭瑞峰,等.模式识别法评价尼群地平片的质量[J].中国药事,2002,16(4):242-244.
(MA Yurong, YANG Liang, GUO Ruifeng, et al. Evaluation for the quality of Nitrendipine tablets with pattern recognition [J]. *Chinese Pharmaceutical Affairs*, 2002, 16(4): 242-244.)
- [5] 杨伦标,高英仪.模糊数学原理及应用[M].第二版.广州:华南理工大学出版社,1998.

- (YANG Lunbiao, GAO Yingyi. *Principle and Application of Fuzzy Mathematics* [M]. 2nd ed. Guangzhou: South China University of Technology Press, 1998.)
- [6] 袁敏, 曾志, 宋力飞, 等. 气相色谱指纹图谱用于连翘的质量控制[J]. 分析化学, 2003, 31(4): 455 - 458.
(YUAN Min, ZENG Zhi, SONG Lifei, et al. The quality assessment of Forsythia suspensa (Thunb.) Vahl by gas chromatographic fingerprint [J]. *Chinese J of Analytical Chemistry*, 2003, 31(4): 455 - 458.)
- [7] 梁晓峰, 吕春祥, 李香兰, 等. 模糊聚类分析评价炭纤维裂解色谱图重现性[J]. 色谱, 1998, 16(5): 439 - 441.
(LIAN Xiaoyi, LU Chuanxiang, LI Xianglan, et al. Fuzzy clustering analysis evaluating the reproducibility of pyrolysis gas chromatography of carbon fibers [J]. *Chromatography*, 1998, 16(5): 439 - 441.)
- [8] 颜春兰. 模糊聚类分析法在中药成分特征谱分析系统中的应用研究[D]. 广州: 华南理工大学, 2003.
(YAN Chunlan. *Research and application using fuzzy clustering in the componential analysis system of chinese traditional medicine characteristic spectrum* [D]. Guangzhou: South China University of Technology, 2003.)

作者简介:

周敬泉 (1943 —), 男, 副教授, 主要研究方向为智能控制、仪表与装置、计算机应用、电子产品研发, E-mail: aujqzhou@scut.edu.cn;

颜春兰 (1974 —), 女, 硕士研究生, 主要研究方向为智能仪表与自动装置、计算机应用;

袁 鹏 (1972 —), 男, 讲师, 博士研究生, 主要研究方向为人工智能、智能控制、计算机应用、电子产品研发;

曾启杰 (1977 —), 男, 硕士研究生, 主要研究方向为智能仪表与自动装置、计算机应用;

袁 敏 (1962 —), 女, 高级实验师, 主要研究方向为气相色谱分析;

张铭光 (1956 —), 男, 副教授, 主要研究方向为植物研究.

(上接第 568 页)

选择控制参数 k_1 的值可以将混沌 Chua 电路稳定到所需的目标轨道. 为了验证该控制方法及讨论其在工程应用中的潜力, 文中设计了控制电路, 电路仿真的结果与理论分析是一致的. 在现有的混沌控制方法中, 有许多方法具有复杂的描述形式且停留在理论实现和数值计算阶段, 而本文中所采用的控制方法不仅控制量易于测得, 而且控制原理很容易在电路中实现, 因此具有较好的工程应用价值.

参考文献(References):

- [1] OTT E, GREBOGI C, YORKE J A. Controlling chaos [J]. *Physical Review Letters*, 1990, 64(11): 1196 - 1199.
- [2] GUEMEZ J, MATIAS M A. Control of chaos in unidimensional maps [J]. *Physical Letter A*, 1993, 181(1): 29 - 32.
- [3] PYRAGAS K. Continuous control of chaos by self-controlling feedback [J]. *Physical Letter A*, 1992, 170(6): 421 - 428.
- [4] MASOLO S, GRASSI G. Controlling chaotic dynamics using back-stepping design with application to the Lorenz system and Chua's circuit [J]. *Int J of Bifurcation and Chaos*, 1999 9(7): 1425 - 1434.
- [5] BARONE K, SINGH S N. Adaptive feedback linearizing control of Chua's circuit [J]. *Int J of Bifurcation and Chaos*, 2002, 12(7): 1599 - 1604.
- [6] LAVAL L, M' Sirdi N K. Stabilization global invariant sets for chaotic systems: An energy based control approach [J]. *Int J of Bifurcation and Chaos*, 2002, 12(6): 1403 - 1409.
- [7] GUO S M, SHIEH L S, CHEN G. Ordering chaos in Chua's circuit: A sampled-data feedback and digital redesign approach [J]. *Int J of Bifurcation and Chaos*, 2000, 10(9): 2221 - 2231.
- [8] CHUA LO, ITOH M, KOCAREV L, et al. Chaos synchronization in Chua's circuits [J]. *J of Circuits, Systems and Computers*, 1993, 3(1): 93 - 108.
- [9] ALLIGOOD K T, SAUER T D, YORKE J A. *Chaos: An Introduction to Dynamical Systems* [M]. New York: Springer-Verlag, 1996.

作者简介:

邹艳丽 (1972 —), 女, 讲师, 研究方向为混沌控制与同步, E-mail: zouyanli@sjtu.edu.cn;

罗晓曙 (1961 —), 男, 博士, 教授, 主要研究方向包括混沌理论与应用、混沌控制与混沌同步;

陈关荣 (1948 —), 男, 博士, 首席教授, 主要研究方向包括混沌控制与同步、复杂性网络研究.