

混合物质谱的简单分类算法

殷龙彪

(华东化工学院化学系, 上海 200237)

摘要 有毒化合物质谱分类是以光谱数据的数值特性为基础的。这些数据的计算机分类提供了各种参数,用以确定属于有毒化合物中某一类的对应化合物。质谱数据的简单分类算法常被用于自相关变换质谱的主分量建模。主分量的有效值以相互确认来确定。对环境空气中 119 种被监测的有毒化合物,用色谱-质谱联用的试验和校正数据所得到的 4 类预报正确率为 76.5%。

关键词 简单分类算法, 质谱, 有毒化合物。

在化学计量学研究中,模式识别技术在计算机信息处理时的作用已引人关注^[1,2],其中统计模式识别更在化学中广泛应用^[3]。模式识别的计算机方法则可方便地作质谱数据的分类^[1]。简单分类算法(简称 SIMCA),是按样本类别分别处理样本数据的一种模式识别^[5]。它以其所提供的一组参数,给各类别以特征,是阐明数据结构中其它描述量的基础^[6,7]。结果残差可用于类别间距的计算,残差的标准偏差直接对应于目标和类别之间的距离^[8]。在复杂混合物的模式识别应用中,模式识别的简单分类算法极其有用^[9,10],N-亚硝化物的致癌性预报^[11,12],大气中多氯联苯残留物及污染物分析等均用此标法^[13,14]。本文用此法对大气中 119 种化合物的质谱分析数据进行分类。结果正确率为 76.5%。

1 方法

1.1 仪器

数据记录和处理用 IBM-PC/XT 计算机进行。存于该机上的数据,以 SIMCA-3B 软件包和取自文件 Finnigan GC-MS 的数据库(INCOS, 即 Integrated Control System)作分析处理。

1.2 过程

119 种化合物列于表 1。4 类化合物表示为目标和试验组。类别 1 为氯化物,类别 2 为无卤苯和多环芳香烃,类别 3 为溴(或氟)化物,类别

4 为无卤直链烃。在试验组中,最高质量为 350,间隔为 35m/g—350m/g。整个质谱分类计算机辅助研究的流程图如图 1 所示。1 类化合物列于表 2。SIMCA-3B 软件包的程序原理如图 2 所示。三维图象分类如表 3 所示。

2 结果与讨论

2.1 分类与建模

按照所有化合物结构类似状况,可分成 4 类。在每一类中选择一些化合物建模,然后将所有化合物来凑合该模型,根据简单分类算法的计算结果,可得到以该模型对应化合物的正确分类预报,同时取得正确率的结果。包括总分类在内的所有结果分别列于表 4 及表 5。结果表明,建模与分类时各类别的化合物数目有关,笔者曾用三维图象系统对所有化合物作分类,在对座标轴旋转观测中,取得位于空间的 8 簇特征点聚集,因而作出分为 8 类的判定。对此分类结果的正确程度及实用价值将另作进一步研究。

2.2 分类结果

模式识别法用于光谱分析结果不理想,可能有如下因素^[15]:

- (1)来自光谱模式的原始特征因素;
- (2)有关光谱鉴别的化学知识欠缺;

(3) 研究分类问题时, 太多地依据传统的化学亚结构, 而不着眼于模式空间的聚类结构的研究。

表 1 119 种试验化合物

序号	名称	序号	名称
1	氯代甲烷	31	氯苯
2	溴代甲烷	32	乙苯
3	氯乙烯	33	4-溴氯代苯
4	氯乙烷	34	苯乙烯
5	二氯甲烷	35	二甲苯
6	丙酮	36	2-氯苯酚
7	二硫化碳	37	五氯苯酚
8	1,1-二氯乙烯	38	苯酚
9	1,1-二氯乙烷	39	二(2-氯乙基)醚
10	反-1,2-二氯乙烯	40	2-氯苯酚
11	氯仿	41	1,3-二氯苯
12	1,2-二氯乙烷	42	1,4-二氯苯
13	2-丁酮	43	苯甲醇
14	1,1,1-三氯乙烷	44	1,2-二氯苯
15	四氯化碳	45	邻甲酚
16	乙烯基乙酸盐	46	二(2-氯异丙基)醚
17	溴二氯甲烷	47	对甲酚
18	1,2-丙烷	48	N-亚硝基-二-N-丙胺
19	顺-1,3-二氯丙烯	49	六氯乙烷
20	三氯乙烯	50	硝基苯
21	二溴氯代甲烷	51	异佛尔酮
22	1,1,2-三氯乙烷	52	2-硝基苯酚
23	反-1,3-二氯丙烯	53	2,4-甲基苯酚
24	苯	54	2-氯乙氧基(双)甲烷
25	溴仿	55	苯甲酸
26	4-甲基-2-戊酮	56	2,4-二氯苯酚
27	2-己酮	57	1,2,4-三氯苯
28	1,1,2,2-四氯乙烷	58	萘
29	四氯乙烯	59	4-氯苯胺
30	甲苯	60	六氯丁二烯
61	4-氯-3-甲基苯酚	91	菲
62	2-甲基萘	92	蒽
63	六氯环戊二烯	93	δ-六氯化苯
64	2,4,6-三氯苯酚	94	七氯
65	2,4,5-三氯苯酚	95	二-N-丁基邻苯二甲酸盐
66	2-氟联苯	96	氯甲桥萘
67	2-氯萘	97	蒽蒽
68	2-硝基苯胺	98	芘
69	二甲基邻苯二甲酸盐	99	α-内液体三氧化硫
70	萘	100	P,P'-双对氯苯基三氯乙烷
71	2,6-二硝基甲苯	101	氧桥氯甲桥萘
72	3-硝基苯胺	102	恩台林
73	二氢萘	103	β-内液体三氧化硫
74	2,4-二硝基苯酚	104	P,P'-双对氯苯基二氯乙烷
75	4-硝基苯酚	105	丁基苯基邻苯二甲酸盐
76	氧芴	106	P,P'-双对氯苯基三氯乙烷
77	2,4-二硝基苯酚	107	内液体三氧化硫硫酸盐
78	邻苯二甲酸二乙酯	108	甲氧氯
79	芴	109	苯并(A)蒽
80	4-氧苯-苯基醚	110	蒈
81	4-硝基苯酚	111	二(2-乙基己基)邻苯二甲酸盐
82	2,4,6-三溴苯酚	112	二-N-辛基邻苯二甲酸盐
83	4,6-二硝基邻甲酚	113	苯并(B)蒽
84	N-亚硝二苯胺	114	苯并(K)蒽
85	4-溴苯基-苯基醚	115	苯并(A)芘
86	α-六氯化苯	116	苝并(1,2,3-COL)芘
87	六氯苯	117	二苯并(A-H)蒽
88	β-六氯化苯	118	苯并(B-H-I)芘
89	五氯苯酚	119	七氯环氧化物
90	γ-六氯化苯		

表 2 4 类化合物分类

类别	化合物(以序号表示, 名称见表 1)
1	1,3,4,5,8,9,10,11,
	12,14,15,18,19,20,22,
	23,28,29,31,38,40,41,
	42,44,49,54,56,57,59,
2	60,61,63,64,65,67,80,86,
	87,88,89,90,93,94,96,
	100,101,102,104,106,108,119.
	24,30,32,34,35,38,43,45,
3	47,50,51,52,53,55,58,
	62,68,69,70,71,72,73,74,
	75,76,77,78,79,81,83,
	84,91,92,95,97,98,99,103,
4	105,107,109,110,111,112,113,
	114,115,116,117,118,
	2,17,21,25,33,36,37,
	66,82,85,
5	6,7,13,16,26,27,48,

表 3 三维图象分类的 8 类化合物分类

类别	化合物(以序号表示, 名称见表 1)
1	24,30,34,36,43,45,47
2	6,7,13,16,32,36,37,39,48,50,
3	1,3,4,12,26,27,33,35,40,
4	2,18,29,39,54,58,59,61,66,
5	67,73,75,79,85,96,
6	98,101,109,110,113,114,115,116,118,
7	86,88,91,93,99,103,
8	20,60,63,89,107,
	9,15,51,52,53,55,62,68,
	69,71,72,74,76,77,78,
	80,82,83,84,91,92,94,95,
	100,105,108,111,112,117,
	5,8,10,11,14,17,19,21,
	22,23,25,28,31,41,42,
	44,49,53,57,64,65,70,75,
	87,97,102,104,106,119

表 4 简单分类算法分类建模的结果

类别	建模的 化合物数	119 种化合物 与各类模型 的配合数	与各类模型正确 预报的化合物 数和总数的比
1,CH	52	86	35/52
1,CH1	46	107	48/52
1,CH2	35	93	41/52
1,CH3	27	70	35/52
1,CH4	34	87	42/52
1,CH5	42	106	50/52
1,CH6	23	45	29/52
1,CH7	29	59	34/52
1,CH8	14	17	16/52
1,CH9	26	40	24/52
1,CH10	31	57	34/52
2,NH1	50	72	40/50
2,NH2	40	83	42/50
2,NH3	42	77	45/50
2,NH4	38	83	43/50
2,NH5	36	57	39/50
3,BF1	10	108	10/10
3,BF2	5	48	8/10
3,BF3	3	30	6/10
3,BF	7	100	10/10
4,NH1	7	44	7/7
4,NH2	6	33	6/7
4,NH3	3	12	3/7

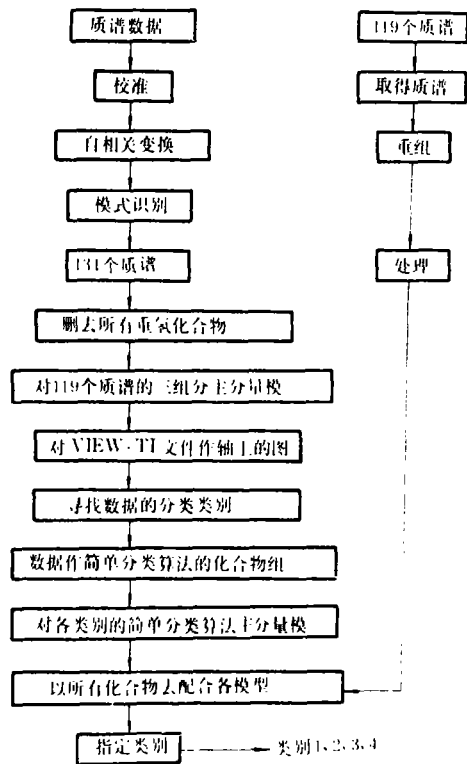


图 1 质谱分类框图

表 5 简单分类算法总的分类结果

类别	正确分类占总数的比例	正确率(%)
1	35/52	67.3
2	42/50	84.0
3	9/10	90.0
4	5/7	71.4
总计	91/119	76.5

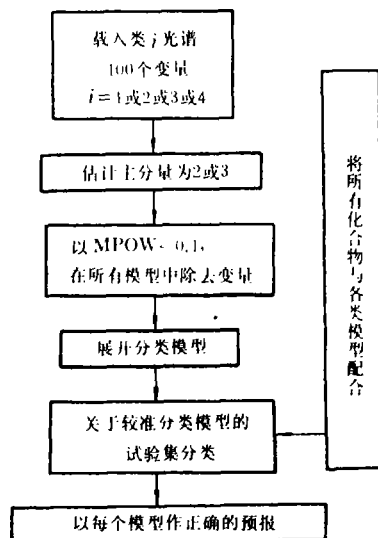


图 2 SIMCA-3B 软件包程序图

本研究中,119 种化合物有 52 种氯化物被归于类别 1,50 种无卤苯和多环芳香烃被放在类别 2。52 种氯化物中有 1/3 不符合正确分类,这些化合物都不是多环烃。用简单分类算法的数据来校正监测数据已有报道^[16]。实际分类研究表明分类模型选择,开发质谱分类方法新体系至关重要。用简单分类算法作色谱-质谱联用的有毒化合物鉴定,不但可探讨它的变量选择和分类规律^[17],在相应的研究中还发现了更准确的结果^[18]。许多年来,模式识别方法一直作为化学计量学的同义词得到高度评价^[19],不但在化合物的分离和鉴定中取得成果^[20],而且已发展到作为模式识别的人工神经网络方法^[21]。国内尚未见这方面报道,因而企盼今后活跃该领域的研究,以促进化学计量学在环境科学中的应用。

(下转第 42 页)

有机质;设计了适合于光合细菌生长的处理装置,该装置适合于光照(可利用日光),内部的纤维性填料为光合细菌提供了附着的场所;返回柱中的菌液大大增加了液相中的菌数,使整个处理系统中光合细菌菌体浓度较大,保持了光合细菌的优势生长;所得到的菌体污泥可作饲料添加剂,即减少二次污染,又带来经济效益。

在废水处理中,以下几个问题值得注意:

(1)从实验结果可以看出利用光合细菌柱式生物膜法处理淀粉废水具有较高的去除效率,有机负荷也较高,但出水的 COD 仍在 1000mg/L 左右。在实际应用中发现光合细菌处理较高浓度有机废水时,COD 去除效率很高,能在短时间内使 COD 降至 1000mg/L 左右,如果继续处理 COD 降至 300mg/L 以下,则需增加几倍长的时间。因此从实际应用的角度出发,高浓度有机废水的处理可先用光合细菌法大幅度降低 COD,再用其它方法进一步处理。

(2)本实验所选废水可生化性较强,较适于光合细菌生长。用该方法处理其它废水时要先对废水水质进行认真考察,若污染物主要为大分子的脂肪、蛋白质类物质可考虑先进行可溶化处理或用物化法处理,再用光合细菌法处理。另外,无论处理哪种废水驯化均很重要,可使光合细菌耐受和代谢有机质的能力都有所提高。

参 考 文 献

- 1 Kobayashi M and Y T Tchan. *Water Resource*. 1973, **7**, 1219
- 2 小林达治. 发酵と工業. 1978, **36**(7):574
- 3 小林正泰. 发酵と工業. 1978, **36**(9):753
- 4 史家梁等. 环境科学研究. 1979, **2**:17
- 5 污染源统一监测分析方法编写组. 污染源统一监测分析. 北京:技术标准出版社, 1983:122, 127, 25
- 6 北京大学生物系生物化学教研室编. 生物化学实验指导. 北京:高等教育出版社, 1988:22-87
- 7 Weaver P F et al. *Arch. Microbiol.* 1975, **105**:207

(上接第 26 页)

参 考 文 献

- 1 Nillson N J. *Learning Machines*. New York: Mc Gran Hill, 1965:17
- 2 Batchelor B G. *Pattern Recognition - Idea in Practice*. New York: Plenum Press, 1978:23
- 3 Jurs P C et al. *Chemical Applications of Pattern Recognition*. New York: John Wiley and Sons, 1975:137
- 4 Voorhees K J et al. *Anal Chem*, 1985, **57**:1630
- 5 Wold S et al. In *Chemometrics: Theory and Application - ACS Symposium Series 52*, Washington D C, 1977:278
- 6 Wold S, J. *Pattern Recognition*, 1976, **8**:127
- 7 Albano C et al. *Anal Chim Acta*, 1978, **103**:429
- 8 Wold S et al. *Chemometrics - Mathematics and Statistics in Chemistry*. New York: Reidel D Publishing Company, 1984:17
- 9 Dunn W J III et al. *Analysis*, 1984, **12**:477
- 10 Lindberg W et al. *Anal Chem*, 1983, **55**:643
- 11 Dunn W J III et al. *Bioorganic Chemistry*, 1981, **10**:29
- 12 Dunn W J III et al. *J Chem Int Comput Sci*, 1981, **21**:8
- 13 Staling D L et al. *Environmental Applications of Chemometrics*, ACS, Washington D C, 1985:1
- 14 Scott D R. *Environmental Application of Chemometrics*, ACS, 1985:106
- 15 Varmuza K. *Computer Application in Chemistry*. Amsterdam: Elsevier Science Publishers B V, 1983:78
- 16 Arcos J C. *Environ Sci Technol*. 1987, **21**:743
- 17 Dunn W J III et al. *Environ Sci Technol*. 1989, **23**:1499
- 18 Scott D R. *Anal Chim Acta*, 1988, **211**:11
- 19 Brown S D. *Anal Chem*. 1990, **62**:84R-101R
- 20 殷龙彪. 分析化学. 1992, **20**(2):137
- 21 Bos A et al. *Anal Chim Acta*, 1992, **256**:133

conditions on the NiCu/REY indicated that the selectivity for alcohols is high from 443—473 K. The increase of temperature and pressure of the reaction lowered the selectivity for alcohols and increased the selectivity for methane and carbon monoxide. Experiments for the investigation of the effect of solvents on carbon dioxide hydrogenation indicated that polar solvents such as triethylamine promoted the reaction more effectively than less polar solvents. FT—IR spectra indicated that the intensity of the 1400 and 1437 cm^{-1} bands of the REY supported catalysts are higher than that of ZSM—5 supported catalysts. The peaks at 1400, 1437 and 1636 cm^{-1} might be assigned to the absorption of NH_3 coordinated to the metal ion center in the form of $[\text{Ni}(\text{NH}_3)_6]^{2+}$ and/or $[\text{Ni}(\text{NH}_3)_5\text{NO}_3]^+$.

Key words: CO_2 , catalytic hydrogenation, bimetal catalyst zeolite, infrared spectra.

Simple Classification Algorithm (SIMCA) Studies on Mass Spectra of the Mixture. Yin Longbiao (Department of Chemistry, East China University of Chemical Technology, Shanghai 200237); *Chin. J. Environ. Sci.*, **14**(5), 1993, pp. 24—26

Classification of mass spectra of toxic compounds is based on the numerical features derived from spectral data. Computer calculations on these data provide various parameters, which are used to determine the corresponding compounds belonging to a certain class of toxic compounds. SIMCA studies on spectral data from mass spectra were used with SIMCA principal components modeling of the autocorrelation transformed mass spectra. The significant number of principal components was determined by cross validation. With this technique, correct rates for the four classes were obtained with GC/MS from training, and calibration data for the 119 toxic compounds targeted in ambient air is 76.5%.

Key words: simple classification algorithm, mass spectra, toxic compound.

The Studies of sampling and determining malodors of organic sulfur compounds. Wang Jizong et al. (Beijing Municipal Institute of Labour protection, Beijing 100054); *Chin. J. Environ. Sci.*, **14**(5), 1993, pp. 27—31
Complex techniques of sampling and determining main malodors have been developed in this paper. Sampling on solid adsorbents followed by thermodesorption and GC/FPD

and/or GC/MS was evaluated as a method for the quantitative analysis of CH_3SH and $(\text{CH}_3)_2\text{S}_2$ in air. Some apparatuses—Adsorbent Sampling Tube, Constant-flow sampling Pump, Thermodesorption Device and Gas Mixer for the preparation of calibration gas mixtures were developed and produced. The experimental results showed that the chinese-made GDX-105 was found suitable for adsorption and collection of CH_3SH and $(\text{CH}_3)_2\text{S}_2$ at the room temperature. The method affords recoveries between 60% and 100%. The limits of detection were 0.3 and 0.2 ng for CH_3SH and $(\text{CH}_3)_2\text{S}_2$. An accuracy of $\pm 3\%$ was obtained for the standard gases of CH_3SH and $(\text{CH}_3)_2\text{S}_2$ in the air or N_2 .

Key words: malodor, organic sulfur, sampling, GC, GC/MS.

Experimental Study on the Transport Patterns of Contaminants in Aeration Zone. Yan Xianliang (Environmental Geology Monitoring Station, Prospecting Institute of Geological Engineering, Taiyuan 030008); *Chin. J. Environ. Sci.*, **14**(5), 1993, pp. 32—38

Based on the laboratory experiment of leaching, there were elaborated the accumulation and transport patterns of phenols, cyanides, arsenic, mercury and chromium in sandy clay and neutral dene, including: ①the absorbing purification by soil can be classified into powerful absorption, saturating absorption and absorbing saturation stages; ②the soil absorption and purification intensity of contaminants in aeration zone soil is usually in a negative correlation to the size of soil particles; ③the reduction of phenols and cyanides in aeration zone is mainly due to their respective own volatility and biological decomposition. It was demonstrated that the main transport form of contaminants in aeration zone is H_3AsO_3 , H_2AsO_3^- , CrO_4^{2-} , $\text{Hg}(\text{OH})^+$, HgOHCl . Meanwhile, the pH value in the aeration zone affected the capability of contaminants transporting. Thus, it is suggested that in the studies on environmental geology and groundwater pollution control the relative factors should be fully taken into account, including the kinds of contaminants, the kinds and thickness of aeration zone, and the pH value.

Key words: aeration zone, contaminants, leaching, adsorbing purification, transport patterns.

Starch Waste Water Treatment by Photosynthetic Bacterium Using Columnar Biofilm Method. Wang yuxin et