

文章编号:1674-2869(2008)04-0042-04

用分子连接性研究有机化合物溶解度的 QSPR

吴桂玲, 吴启勋

(青海民族学院化学系, 青海 西宁 810007)

摘要:利用分子连接性指数对五种类型共139种有机化合物的水溶解度进行了预测, 得出有机化合物的溶解度($\log S$)与其零阶指数(0X)、零阶价指数($^0X^v$)和极化率(Φ)有良好的相关性, 采用 SPSS11.5 软件处理数据, 建立了一个具有四个变量的方程, 预测值准确, 该法具有计算简单, 相关性好等显著特点, 有一定的推广应用价值。

关键词:分子连接性指数; 溶解度; 极化率; QSPR

中图分类号: O641

文献标识码: A

分子连接性指数是一种拓扑指数, 是根据分子结构计算出的非经验性参数, 拓扑指数是分子结构数值化的一种方式, 它通过对表征分子图的矩阵实施某种数学运算而获得, 它是一种图的不变量, 直接产生于分子结构, 反映化合物分子中各骨架原子排列、连接方式等, 即化合物的结构特征, 拓扑指数有三个基本的要求: 第一, 拓扑指数应为一个或一组数据, 即分子图拓扑不变量的数值化; 第二, 能充分反映分子图的连接信息和化学环境; 第三, 能有效地表达化合物结构与性质(活性)的关系。许多研究工作已经证明, 分子连接性指数与物质的多种理化参数用来描述化学结构, 并且与化合物的物理性质、热力学参数、化学性能、生物活性以及致癌性相关^[1]。

分子结构与性质关系的拓扑指数法研究引起了化学家们的广泛兴趣^[2]。迄今已提出100多种拓扑指数, 其中得到广泛应用的是由 Randic^[3]提出、后经 Kier 等人^[4]进一步发展的价连接性指数($^mX^v$)。余训民等^[5]在36种卤代苯的溶解度- $\lg S_w$ 研究中, 采用零阶价连接性指数($^0X^v$)建立 QSPR 方程, 在他们的工作基础上, 本文应用了 Nagamann 的研究方法^[6], 利用分子连接性指数与极化率(Φ)作为参数, 对139种有机化合物的溶解度进行 QSPR 分析, 预测结果好。

1 计算方法

1.1 极化率(Φ)的计算

极化率的计算方法是由 Kestelar 提出的^[7], 可用下面方程表示:

$$\Phi = A(\text{H 原子数}) + B(\text{C 原子数}) +$$

$$C(\text{Cl 原子数}) + D(\text{Br 原子数}) +$$

$$E(\text{双键数})$$

其中, 上式中的 A~E 均为常数, 分别为: 0.42、0.93、2.82、3.34、0.58。

在此基础上, 王连生先生对极化率公式进行了改进, 得到的新方程^[8]为:

$$\Phi = -0.963n_{\text{Cl}} - 0.361n_{\text{H}} - 0.767n_{\text{双键}} \quad (n_{\text{Cl}} \text{ 为 Cl 原子数, } n_{\text{H}} \text{ 为 H 原子数, } n_{\text{双键}} \text{ 为双键数})$$

例: 用上述新方程解苯的极化率 Φ 值。

H 原子数为 6, Cl 原子数为 0, 双键数为 3。

$$\Phi = -0.963 \times 0 - 0.361 \times 6 - 0.767 \times 3 = -4.47$$

同理, 计算了139种有机化合物的极化率 Φ 值, 结果如表1所示。

由于极化率仅仅是有机化合物溶解度相关信息的一部分, 因此有必要加入一个参数, 即分子连接性指数。

1.2 分子连接性指数的计算

分子连接性指数是用点价乘积平方根的倒数来表示, 它是根据分子构型计算出来的常数, 反映了分子的连接情况和分枝的多少。

1.2.1 杂原子的点价计算 Kier 等在改进了 Randic 分枝指数的基础上提出了连接性指数 $^mX^v$, 其实质是采用了原子的核外电子总数 Z 、价电子数 Z_i 和相连的氢原子数 h_i 来定义新的原子点价 δ_i^v ^[10],

$$\text{即} \quad \delta_i^v = \frac{Z_i - h_i}{Z - Z_i - 1}$$

例: 对氯代烃中的氯来说, 核外电子总数 Z 为 17, 价电子数 Z_i 为 7, 相连的氢原子数 h_i 为 0,

$$\text{则, 原子点价 } \delta_i^v = \frac{7-0}{17-7-1} = 0.78$$

收稿日期: 2008-04-24

作者简介: 吴桂玲(1983-), 女, 青海乐都人, 物理化学专业硕士, 研究方向: 计算机化学。

表1 化合物名称及水溶解度(logS)的实验值和预测值

Table 1 List of compounds and their experimental and predicted water solubility (logS)

NO	Compound name	Φ	Exp.		OLS models		NO	Compound name	Φ	Exp.		OLS models	
			logS		Pred.	Cal.				logS		Pred.	Cal.
1	benzene	-4.47	-0.75	-1.24	0.49		71	perchloroethene	-4.62	-1.82	-1.34	-0.48	
2	toluene	5.19	1.29	1.53	0.24		72	3-chloroprop-1-ene	3.54	0.54	0.69	0.15	
3	1-ethylbenzene	-5.91	-1.82	-1.96	0.14		73	(Z)-1,3-dichloroprop-1-ene	-4.14	-1.00	-1.01	0.01	
4	p-xylene	5.91	1.73	1.88	0.15		74	perchloroprop-1-ene	6.55	2.77	2.97	0.20	
5	m-xylene	5.91	1.79	1.88	0.09		75	perchlorobuta-1,3-diene	7.31	3.39	3.66	0.27	
6	o-xylene	-5.91	-1.76	-1.88	0.12		76	bromochloromethane	-1.69	0.17	-0.04	0.21	
7	1,2,3-trimethylbenzene	6.63	2.13	2.28	0.15		77	bromomethane	1.08	0.18	0.33	0.15	
8	1,2,4-trimethylbenzene	-6.63	-2.24	-2.28	0.04		78	dibromomethane	-0.72	0.06	0.02	0.04	
9	mesitylene	6.63	2.01	2.28	0.27		79	bromoform	0.36	0.51	0.59	0.08	
10	1-propylbenzene	6.63	2.26	2.42	0.16		80	perbromomethane	0.00	1.62	1.19	0.43	
11	cumene	-6.63	-2.30	-2.37	0.07		81	bromochloroethane	-1.81	-0.05	-0.05	0.00	
12	1-ethyl-2-methylbenzene	6.63	2.03	2.35	0.32		82	(Z)-1,2-dibromoethene	1.49	0.05	0.64	0.58	
13	1-ethyl-4-methylbenzene	-6.63	-2.03	-2.35	0.32		83	1,2-dibromomethane	-1.44	-0.38	-0.41	0.03	
14	1-butylbenzene	7.36	2.92	2.93	0.01		84	1,1,1,2-tetrabromoethane	0.72	1.17	1.89	0.72	
15	1-isobutylbenzene	-7.36	-3.00	-2.88	-0.12		85	1-bromopropane	-2.53	-0.62	-0.43	-0.18	
16	1-sec-butylbenzene	-7.36	-2.77	-2.88	0.11		86	2-bromopropane	-2.53	-0.51	-0.29	-0.22	
17	1-tert-butylbenzene	7.36	2.47	2.82	0.35		87	1,2-dibromo-3-chloropropane	2.77	1.00	1.39	0.39	
18	1-bromo-2-chlorobenzene	-4.71	-1.91	-2.06	0.15		88	1,2-dibromopropane	-4.09	-0.85	-0.72	-0.13	
19	1-bromo-3-chlorobenzene	4.71	1.93	2.06	0.13		89	1-bromobutane	3.25	1.27	0.85	0.42	
20	1-bromo-4-chlorobenzene	-4.71	-2.35	-2.06	-0.29		90	1-bromo-2-methylpropane	-3.25	-1.29	-0.71	-0.58	
21	1-bromobenzene	-4.11	-1.39	-1.61	0.22		91	1-bromo-3-methylbutane	-3.97	-1.70	-1.30	-0.40	
22	1,2-dibromobenzene	3.75	2.13	2.20	0.07		92	butan-1-ol	3.25	0.86	0.94	0.08	
23	1,3-dibromobenzene	-3.75	-2.17	-2.20	0.02		93	2-methylpropan-1-ol	-3.25	0.93	1.01	-0.08	
24	1,4-dibromobenzene	3.75	2.70	2.20	0.50		94	butan-2-ol	3.25	1.26	1.01	0.25	
25	1,2,3-tribromobenzene	-3.38	-3.54	-3.01	-0.52		95	pentan-1-ol	-3.97	0.33	0.54	-0.20	
26	1,2,4-tribromobenzene	-3.38	-3.00	-3.01	0.01		96	2-methylbutan-1-ol	-3.97	0.46	0.62	-0.16	
27	1,3,5-tribromobenzene	3.38	2.70	3.01	0.32		97	3-methylbutan-1-ol	3.97	0.43	0.62	0.19	
28	1,2,4,5-tetrabromobenzene	-3.02	-4.30	-4.09	-0.22		98	2,2-dimethylpropan-1-ol	-3.97	0.52	0.70	-0.18	
29	1-chlorobenzene	5.07	1.30	1.57	0.27		99	pentan-2-ol	3.97	0.63	0.62	0.01	
30	1,2-dichlorobenzene	-5.67	-2.04	-1.97	-0.07		100	pentan-3-ol	-3.97	0.70	0.62	0.08	
31	1,3-dichlorobenzene	5.67	1.91	1.97	0.06		101	3-methylbutan-2-ol	3.97	0.73	0.68	0.05	
32	1,4-dichlorobenzene	5.67	2.05	1.97	0.08		102	2-methylbutan-2-ol	3.97	1.03	0.70	0.33	
33	1,2,3-trichlorobenzene	-6.27	-2.51	-2.44	-0.07		103	hexan-1-ol	-4.69	-0.23	0.11	-0.35	
34	1,2,4-trichlorobenzene	6.27	2.52	2.44	0.09		104	2-methylpentan-1-ol	4.69	0.10	0.18	0.28	
35	1,3,5-trichlorobenzene	-6.27	-3.18	-2.44	-0.74		105	4-methylpentan-1-ol	-4.69	-0.13	0.18	-0.31	
36	1,2,3,4-tetrachlorobenzene	6.88	3.37	2.99	0.37		106	2,2-dimethylbutan-1-ol	4.69	0.03	0.26	0.29	
37	1,2,3,5-tetrachlorobenzene	6.88	3.46	2.99	0.46		107	3,3-dimethylbutan-1-ol	4.69	0.51	0.26	0.25	
38	1,2,4,5-tetrachlorobenzene	-6.88	-3.25	-2.99	-0.26		108	2-ethylbutan-1-ol	-4.69	-0.16	0.18	-0.34	
39	dichloromethane	2.65	0.12	0.10	0.21		109	hexan-3-ol	4.69	0.19	0.18	0.01	
40	chloroform	-3.25	-0.10	-0.41	0.30		110	3-methylpentan-2-ol	-4.69	0.27	0.26	0.01	
41	perchloromethane	3.85	1.10	0.76	0.34		111	4-methylpentan-2-ol	4.69	0.20	0.26	0.06	
42	chloroethane	-2.77	-0.25	-0.08	-0.17		112	2-methylpentan-3-ol	-4.69	0.30	0.26	0.04	
43	1,1-dichloroethane	-3.37	-0.32	-0.38	0.06		113	3,3-dimethylpentan-2-ol	-4.69	0.37	0.32	0.05	
44	1,2-dichloroethane	3.37	0.07	0.46	0.39		114	2-methylpentan-2-ol	4.69	0.50	0.26	0.24	
45	1,1,1-trichloroethane	-3.97	-0.82	-0.88	0.06		115	3-methylpentan-3-ol	-4.69	0.62	0.26	0.36	
46	1,1,2-trichloroethane	3.97	0.36	0.81	0.46		116	2,3-dimethylbutan-2-ol	4.69	0.60	0.32	0.28	
47	1,1,1,2-tetrachloroethane	-4.57	-0.96	-1.23	0.28		117	2,4-dimethylpentan-2-ol	-5.42	0.10	-0.17	0.28	
48	1,1,2,2-tetrachloroethane	-4.57	-0.53	-1.24	0.71		118	heptan-1-ol	-5.42	-0.77	-0.34	-0.43	
49	1,1,1,2,2-pentachloroethane	5.18	1.30	1.74	0.44		119	2,4-dimethylpentan-1-ol	5.42	0.46	0.21	0.25	
50	perchloroethane	-5.78	-2.30	-2.29	-0.01		120	3,4-dimethylpentan-1-ol	-5.42	-0.54	-0.24	-0.29	
51	1-chloropropane	3.49	0.58	0.42	0.16		121	4,4-dimethylpentan-1-ol	5.42	0.49	0.21	0.28	
52	2-chloropropane	-3.49	-0.53	-0.35	-0.18		122	heptan-2-ol	-5.42	-0.49	-0.30	-0.19	
53	1,2-dichloropropane	-4.09	-0.53	-0.78	0.25		123	heptan-3-ol	-5.42	-0.38	-0.30	-0.08	
54	1,3-dichloropropane	4.09	0.57	0.84	0.27		124	heptan-4-ol	5.42	0.34	0.30	0.04	
55	1,2,3-trichloropropane	-4.69	-0.72	-1.26	0.54		125	5-methylhexan-2-ol	-5.42	-0.32	-0.24	-0.07	
56	1-chlorobutane	4.21	1.22	0.81	0.41		126	2-methylhexan-2-ol	5.42	0.26	0.21	0.05	
57	2-chlorobutane	-4.21	-1.00	-0.74	-0.26		127	2,2-dimethylpentan-3-ol	-5.42	-0.10	-0.17	0.08	
58	1,2-dichlorobutane	4.81	1.30	1.21	0.10		128	2,4-dimethylpentan-3-ol	5.42	0.17	0.18	0.01	
59	2,3-dichlorobutane	4.81	1.48	1.16	0.32		129	2-methylhexan-2-ol	5.42	0.03	0.21	0.19	
60	1-chloropentane	-4.93	-1.70	-1.25	-0.45		130	3-methylhexan-3-ol	-5.42	-0.06	-0.21	0.15	
61	2-chloropentane	4.93	1.60	1.17	0.43		131	2,3-dimethylpentan-2-ol	5.42	0.05	0.17	0.02	
62	3-chloropentane	-4.93	-1.60	-1.17	-0.43		132	2,3-dimethylpentan-3-ol	-5.42	0.20	-0.17	0.38	
63	1-chlorohexane	5.66	2.04	1.70	0.34		133	3-ethylpentan-3-ol	5.42	0.19	0.21	0.41	
64	1-chloro-2-methylpropane	4.21	1.04	0.74	0.30		134	2,3,3-trimethylbutan-2-ol	5.42	0.34	0.09	0.43	
65	2-chloro-2-methylbutane	-4.93	-1.48	-1.10	-0.38		135	octan-1-ol	-6.14	-1.26	-0.85	-0.41	
66	2,3-dichloro-2-methylbutane	5.54	1.54	1.58	0.04		136	2-ethylhexan-1-ol	6.14	1.00	0.80	0.20	
67	chloroethene	-2.81	-0.56	-0.35	-0.21		137	3-methylheptan-2-ol	-6.14	-0.61	-0.75	0.14	
68	1,1-dichloroethene	3.42	0.48	0.61	0.14		138	3-methylheptan-3-ol	6.14	0.49	0.74	0.25	
69	(Z)-1,2-dichloroethene	3.42	0.46	0.64	0.18		139	2,2,3-trimethylpentan-3-ol	6.14	0.16	0.63	0.48	
70	1,1,2-trichloroethene	-4.02	-0.96	-0.96	0.00								

对溴代烃中的溴来讲,核外电子总数 Z 为 35, 价电子数 Z_i 为 7, 相连的氢原子数 h_i 为 0, 则,

$$\text{原子点价 } \delta_i^v = \frac{7-0}{35-7-1} = 0.26$$

1.2.2 分子连接性指数的定义 零阶指数 (0X) 是各个顶点之点价平方根的倒数之和, 表示为: $^0X = \sum (\delta_i)^{-\frac{1}{2}}$; 同理, 零阶价指数 ($^0X^v$) 表示为: $^0X^v = \sum (\delta_i^v)^{-\frac{1}{2}}$. 以苯分子为例, 计算其分子连接性指数为:

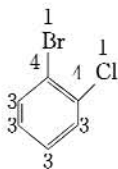


$$^0X = \sum (\delta)^{-\frac{1}{2}} = \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}}$$

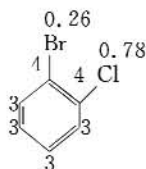
$$^0X^v = \sum (\delta^v)^{-\frac{1}{2}} = \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{3}}$$

$$\frac{1}{\sqrt{3}} = 3.46$$

再以 1 溴 2 氯苯为例, 计算其分子连接性指数:



$$^0X = \sum (\delta)^{-\frac{1}{2}} = \frac{1 \times 2}{\sqrt{1}} + \frac{1 \times 4}{\sqrt{3}} + \frac{1 \times 2}{\sqrt{4}} = 5.30$$



$$^0X^v = \sum (\delta^v)^{-\frac{1}{2}} = \frac{1}{\sqrt{0.26}} + \frac{1}{\sqrt{0.78}} + \frac{1 \times 4}{\sqrt{3}} + \frac{1 \times 2}{\sqrt{4}} = 6.40$$

计算出的 Φ 值, 0X , $^0X^v$, $(^0X^v)^2$ 如表所示 [其中 0X , $^0X^v$, $(^0X^v)^2$ 受篇幅的限制略去].

2 结果与讨论

有机化合物的溶解度与分子连接性指数和极化率具有相关性. 水溶性是一个很有实用意义的物理常数. 这里讨论的溶解性指溶剂是水, 溶质是有机液体. 溶解过程的影响因素很多, 分子间作用力、解离或缔合等都有影响, 这里排除溶质在溶剂中的解离或缔合作用以及含有多种杂原子有机物的情况. 整个样本集 139 个化合物按其分子特征

可分为五大类: 取代苯 (1~38, 共 38 个), 氯代烷烃 (39~66, 共 28 个), 氯代烯烃 (67~75, 共 9 个), 溴代烷烃 (76~91, 共 16 个), 脂肪醇 (92~139, 共 48 个). 有机化合物水溶解度实验数据来自文献[9], 水溶解度实验值为 $\log S$ (S 为溶质溶解在水中的摩尔浓度, 单位: $\text{mol} \cdot \text{L}^{-1}$), 化合物名称及其实验值列于表 1.

利用 SPSS11.5 软件, 对 139 种有机化合物的溶解度 ($\log S$) 与其零阶指数 (0X)、零阶价指数 ($^0X^v$) 和极化率 (Φ) 进行相关性分析, 其结果有良好的相关性且 t 检验 (方程相应系数底下列出 t 值) 都通过, 得如下回归方程:

$$\log S = -1.112 + 1.009\Phi - 0.870^0X^v + 1.611^0X - 0.034(^0X^v)^2$$

$$(3.808 \ 32.712 \ -7.270 \ 30.820 \ -3.890)$$

$$(n=139, R=0.991, R^2=0.989, S=0.23, F=740)$$

以上方程对 139 个化合物的溶解度进行了预测, 可看出具有四个变量的方程的预测值是比较好. 结果如表 1 所示. 从表中可以看出, 对于这 139 个化合物来说, 溶解度跨度为 5 个 \log 单位. 对于 96.4% 的化合物可以得到很好的解释, 共有 3.6% 的化合物有些偏离. 这表明建立的模型有好的可靠性和预测能力与文献中的类似工作相比^[8], 从相关系数来看 $R=0.991$ ($n=139$), 比其 $R=0.960$, $n=38$ 要好些.

由上述进行如下讨论.

a. 计算获得的分子连接性指数是从通过对化合物的拓扑结构的描述来研究物质的定量结构活性相关关系 (QSPR). 它代表了有机化合物溶解度相关信息的一部分, 有良好的预测能力. 溶解过程的推动力是熵的增加. 如果在溶解过程中没有热效应, 溶质和溶剂是混溶的; 如果有熵的减小或有热量放出, 则会增加溶解的推动力. 尽管溶解过程很复杂, 但它的主要特征为分子间作用力. 由于分子连接性能表示色散或极化作用, 因此, 它能解决溶解度的问题. 极化率与分子的平均偶极矩有关. 对于非极性分子, 若极化率越大, 则在外电场诱导出的偶极矩越大. 极性分子具有永久偶极矩, 它的极化率是原子极化、电子极化与定向极化的总和. 因此, 本文引入极化率这个参数对五种类型的有机化合物的水溶解度进行了相关分析, 建立了良好的数学模型.

b. 从我们对建立回归模型的统计检验可以看出, 所建立回归模型比较好. 并从残差散点图 (图 1 是以残差为纵坐标, 因变量 ($\log S$) 为横坐标的对划图) 和残差直方图 (图 2 是以频率为纵坐标, 残差为横坐标的对划图) 可以明确看出样本点随机

分布,并得到了满意的正态分布曲线。

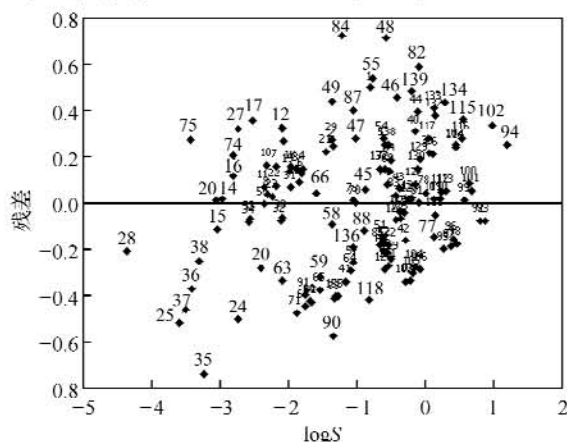


图1 残差散点图

Fig. 1 Residual scatter diagram

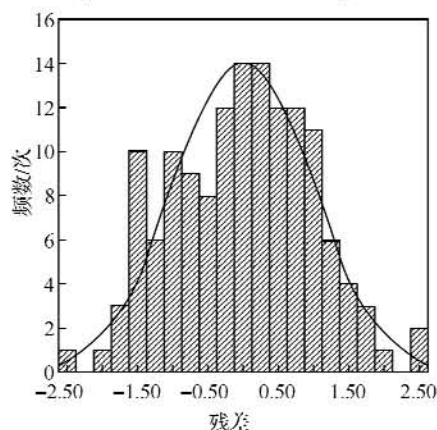


图2 回归标准化残差直方

Fig. 2 Regression standardized residual histogram

c. 有机化合物在水中的溶解度的数据对于化学工业,特别是石油化学工业的设计、科研、医药以及环保工作者极其重要。因此,进行有机物水溶解度的定量结构——性质关系研究非常重要。已有 QSPR 模型的研究对象通常是结构相似化合物或者是同种类型的化合物,所建模型往往不适用

于结构各异化合物的预测。本文则对五种类型的有机化合物的水溶解度进行了相关分析,建立了良好的 QSPR 模型,并对结构各异化合物具有良好的预测性,因而对实际生产有良好的指导作用。

综上所述,有机化合物的溶解度($\log S$)与其零阶指数(0X)、零阶价指数($^0X^v$)和极化率(Φ)之间有良好的相关性,利用本文得到 QSPR 模型可以预测取代苯,氯代烷烃,氯代烯烃,溴代烷烃,脂肪醇等五类有机化合物的水溶解度。

参考文献:

- [1] Kier L M, Hall L H. Molecular connectivity in chemistry and drug research [M]. New York: Academic Press, 1976; 82-90.
- [2] 王化云,江元生.应用拓扑指数预测化学性质[J].化学通报,1992,11:6-11.
- [3] Randic M. On the characterization of molecular branching[J]. Am Chem Soc, 1975, 97(23): 6609.
- [4] Kier L B, Hall L H. Molecular connectivity in chemistry and drug research [M]. New York: Academic Press, 1976; 95-101.
- [5] 余训民.一个新的拓扑指数用于芳香烃化合物的溶解度、分配系数及生物毒性的研究[J].环境化学, 2001, 2(1): 31.
- [6] 王连生.关于分子连接性指数与理化参数相关性的探讨[J].环境科学丛书, 2001, 10(3): 75.
- [7] Horvath A L. Halogenated Hydrocarbons Dekker [M]. New York: Academic Press, 1982; 110.
- [8] 王连生,支正良.分子连接性与分子结构-活性[M].北京:科学出版社, 1992: 95.
- [9] 王连生,支正良.分子连接性与分子结构-活性[M].北京:科学出版社, 1992: 101-105.
- [10] 王连生.有机污染化学[M].北京:高等教育出版社, 2004: 419.

Study on QSPR of organic compounds' solubility with molecular connectivity indexes

WU Gui-ling, WU Qi-xun

(Department of Chemistry, Qinghai University for Nationalities, Xi'ning 810007, China)

Abstract: A good relationship has been obtained in 0X , $^0X^v$, Φ and $\log S$, using molecular connectivity indexes to predict solubility on five different kinds of 139 organic compounds. This paper processes data with SPSS11. 5, and establishes an equation which has four variables, which also receives good predictive value. This method has the advantages of simple-calculation and good-relationship, and has good value for extension and application.

Key words: molecular connectivity indexes; solubility; polarizability; QSPR

本文编辑:张 瑞