

研究简报

# 人工神经网络法预测炸药组分的色谱保留值参数

章婷曦<sup>1</sup>, 黄俊<sup>2</sup>, 周申范<sup>2</sup>

(1. 南京师范大学化学与环境科学学院, 江苏 南京 210097; 2. 南京理工大学化工学院, 江苏 南京 210094)

**摘要** :以分子拓扑指数作为炸药组分的结构描述符,利用反向传播算法(BP)人工神经网络,以 Sigmoid 函数为传递函数,分子连接性指数 $\chi^0, \chi^1, \chi^2, \chi^3$ 与边邻接指数( $\epsilon$ )为输入向量,反相高效液相色谱保留值参数  $\log k_w$  和  $S$  为输出向量,将输入向量归一化至  $-3 \sim 3$  区间,输出向量归一化至  $0 \sim 1$  区间,网络精度取 0.5,学习步长  $\eta$  的初始值取 0.2,动量因子  $\alpha$  取 0.5,通过对 20 种炸药的神经网络模型进行训练,建立了炸药分子结构与  $\log k_w$  和  $S$  之间的定量模型。结果表明,该模型较好地反映了炸药分子结构与保留值之间的关系,预测值与文献所报道实验值的相对误差大部分在  $\pm 5\%$  内,具有较高的预测精度。

**关键词** :保留值参数;人工神经网络;定量结构-色谱保留相关;分子拓扑指数

中图分类号:O658 文献标识码:A 文章编号:1000-871X(2001)04-0319-04

## 1 引言

近年来,定量结构-色谱保留相关(quantitative structure retention relationship, QSRR)的研究得到了广泛关注。自从 Snyder 等<sup>[1]</sup>提出了高效液相色谱中的二参数保留方程以来, Schoenmaker 等<sup>[2]</sup>、卢佩章等<sup>[3]</sup>和耿信笃等<sup>[4]</sup>对该方程进行了改进,并由经验到理论对参数的物理意义作了说明。但在实际应用中,广泛采用的还是 Snyder 提出的公式  $\log k' = \log k_w - S\phi$ 。其中  $k'$  为容量因子,  $\phi$  为流动相中强溶剂的体积分数,  $S$  为一经验常数,  $k_w$  是  $\phi$  为零时的  $k'$  值。对特定溶质来说,采用一定的反相液相色谱柱和一定类型的二元溶剂,其  $\log k_w$  和  $S$  是一常数,根据该式可预测色谱峰的出峰位置。因此  $\log k_w$  和  $S$  值的预测也成为了 QSRR 研究的一项重要内容。

一些方法已被用于  $\log k_w$  和  $S$  值的预测,如利用 Taft 常数及疏水性常数表征分子结构差异并采用回归方法来进行预测<sup>[5]</sup>、利用多个物理常数表征分子差异并采用人工神经网络方法来进行预测<sup>[6-9]</sup>等,但其前提是必须知道化合物的物理化学常数或其他特征常数。我们采用基于化学图论的分子拓扑指数来表征分子结构,并采用反向传播算法(BP)人工神经网络对 20 种炸药组分建立了 QSRR 定量关系式,预测所得到的  $\log k_w$  和  $S$  具有较好的

精度。

## 2 理论部分

我们采用的是目前比较成熟且应用最广泛的 3 层 BP 神经网络模型<sup>[10]</sup>,它由输入层、输出层和隐层组成,同层各节点互不连接,相邻层的节点通过权连接。输入层各点的输入信号经权重耦合到隐层的各点,由作用函数  $f(x)$  转换后再耦合到输出层的各点。将输出信号与学习样本的目标数值进行比较,两者之间的误差利用“反传算法”(back-propagation learning algorithm)沿原通道返回,通过修改各层节点的连接权重,使误差达到最小。

图 1 所示是一个带偏置节点的  $m-n-l$  3 层 BP 神经网络,随机 BP 算法可以描述如下<sup>[11]</sup>:

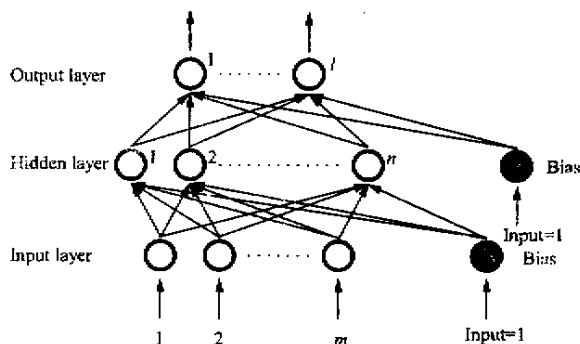


图 1 一个典型的 BP 网络(带偏置的  $m-n-l$  体系)示意图  
Fig. 1 Schematic diagram of a typical back-propagation algorithm network ( $m-n-l$  with bias)

第一步 权值(包括与偏置节点相连的各个权值)初始化,一般采用随机函数来进行赋值;

第二步 随机给定训练样本,即给定特定的输入  $I$  和输出  $T$ ;

第三步 计算实际输出,这里标记偏置节点序号为 0;

设  $O_{pk}$  为第  $p$  个样本的实际输出,则

$$O_{pk} = f\left(\sum_{j=0}^n W_{kj}f\left(\sum_{i=0}^m W_{ji}I_i\right)\right), k = 1, 2, \dots, l$$

第四步 修正权值,即从输出层开始,将误差信号沿连接通路反向传播,通过修正各层节点的连接权值,使误差最小。

设  $T_{pk}$  为第  $p$  个样本的期望输出,则:

(1) 输出层到隐含层的传递误差项为

$$\delta_{pk} = (T_{pk} - O_{pk})O_{pk}(1 - O_{pk}), k = 1, 2, \dots, l$$

(2) 隐含层到输入层的传递误差项为

$$\delta_{pj} = O_{pj}(1 - O_{pj})\sum_k \delta_{pk}W_{kj}, k = 1, 2, \dots, l$$

节点  $j$  与上一层节点  $i$  间的权重调整公式为

$$\Delta W_{ji(N+1)} = \eta\delta_{pj}O_{pi} + \alpha\Delta W_{ji(N)}$$

式中:  $W$  为权重,  $\Delta W_{ji(N+1)}$  表示当前的权值修正值,  $\Delta W_{ji(N)}$  表示上一学习周期的权值修正值;  $N$  为训练次数;  $\eta$  和  $\alpha$  分别为学习步长和动量因子,  $0 < \alpha < 1$ 。

当所有样本输出值与目标期望值之间的均方差 RMS 值满足要求时,停止迭代,网络训练完毕,各节点的连接权值就固定下来,同时也就得到了所研究对象的输入输出之间的拓扑关系,即建立了输入向量与输出向量之间的数学定量关系。利用训练好的网络即可方便地进行新样本的预测。

由于 BP 的计算较为复杂,我们利用 FORTRAN 语言编制了源程序,并经 MS-FORTRAN PWB 5.1 编译成可执行文件以辅助计算。

### 3 建模方法

#### 3.1 分子结构特征参数的选择

本 QSRR 研究中,我们采用的是零阶至二阶的分子连接性指数 ( ${}^0\chi \sim {}^2\chi$ ) 和边邻接指数 ( $\epsilon$ ),其值由自编的 FORTRAN 程序计算得到。关于分子连接性指数的计算方法可参见文献 [12] 和 [13]。

#### 3.2 网络结构选择

选用一个带偏置节点的 4-5-2 体系,即 4 个输入节点、5 个隐层节点及两个输出节点的体系进行模型训练及参数预测,其中,隐层及输出层均选择 Sigmoid 函数<sup>[10]</sup>作为传递函数,分子连接性指数 ${}^0\chi$ ,

${}^1\chi$ ,  ${}^2\chi$  与边邻接指数  $\epsilon$  为输入向量,  $\log k_w$  和  $S$  为输出向量。向量采用线性标准化方法,其中输入向量归一化至  $-3 \sim 3$  区间,而输出向量归一化至  $0 \sim 1$  区间。网络精度取 0.5,学习步长  $\eta$  的初始值取 0.2,动量因子  $\alpha$  取 0.5。采用随机数矩阵进行权值矩阵的初始化,设置随机数生成器的最大和最小输出为 +3 和 -3。

#### 3.3 预测模型建模方法

将文献 [14] 中的 20 种炸药组分组成样本集(见表 1),采用文献 [15] 中的神经网络建模方法,每次从 20 个样本中随机选出一个样本作为预测集,其余的 19 个样本组成训练集进行模型训练,直至得到预测精度符合要求的模型。

表 1 炸药的人工神经网络模型训练与预测结果

Table 1 Artificial neural network training and predicted results for the explosives

Sample name*	$\log k_w$			$S$		
	ref.	predicted	relative error (%)	ref.	predicted	relative error (%)
$\alpha$ -TNT	6.102	6.225	2.0	7.050	7.107	0.8
2,4-DNT	7.022	6.966	-0.8	7.674	7.698	0.3
2,6-DNT	6.805	6.511	-4.3	7.571	7.309	-3.5
2,5-DNT	5.511	5.735	4.1	6.190	6.563	6.0
$o$ -MNT	7.828	8.504	8.6	8.513	9.062	6.4
$m$ -MNT	7.704	7.350	-4.6	8.150	7.945	-2.5
$p$ -MNT	7.723	7.416	-4.0	8.208	8.009	-2.4
RDX	3.592	3.560	-0.9	4.534	4.486	-1.1
Tetryl	5.627	5.511	-2.1	6.750	6.439	-4.6
NQ	3.692	3.725	0.9	4.644	4.664	0.4
Resorcinol	4.058	4.241	4.5	4.915	5.003	1.8
C1	11.526	10.795	-6.3	12.038	11.391	-5.4
C2	9.617	9.989	3.9	10.340	10.640	2.9
DMP	5.075	5.140	1.3	5.867	5.925	1.0
DEP	7.949	7.985	0.5	8.710	8.704	-0.1
PET	4.633	4.648	0.3	5.338	5.644	5.7
DPA	9.487	9.622	1.4	10.167	10.109	-0.6
Xylene	9.980	9.676	-3.0	10.430	10.192	-2.3
Phenol	5.417	5.394	-0.4	5.938	5.978	0.7
DINA	4.425	4.493	1.5	5.598	5.402	-3.5

\*. TNT: 2,4,6-trinitrotoluene; DNT: dinitrotoluene; MNT: nitrotoluene; RDX: cyclotrimethylene trinitramine; NQ: guanidine nitrate; C1: 1,3-dimethyl-1,3-diphenylurea; C2: 1,3-diethyl-1,3-diphenylurea; DMP: dimethyl( $o$ -)phthalate; DEP: diethyl( $o$ -)phthalate; PETN: tetranitropentaerythrite; DPA: diphenylamine; DINA: dinitro-diethyl-nitramine.

### 4 模型的训练与预测结果

根据第 2 节所述理论,用 FORTRAN 语言编制计算机程序,在 Pentium 166 PC 机上,采用 3.2 节中选择的 BP 网络,按 3.3 节所提供的方法进行建

模。经过约 260 s 的迭代, 网络训练完毕, 程序输出各种参数的结果。并得到输入层与隐层之间、隐层与输出层之间各节点的连接权值(见表 2 和表 3)。

表 2 输入层与隐层之间的连接权值

Table 2 Connective weights between input layer and hidden layer

Hidden node	Input node				Bias node
	1	2	3	4	
1	0.196918	-2.607561	0.910841	3.339612	-2.005541
2	-1.872654	1.423686	-0.428141	0.885255	2.777074
3	-1.757940	-0.899393	-0.998899	0.715878	-7.407116
4	1.562153	-6.021570	4.859765	0.401442	-0.501271
5	-4.230594	4.352118	3.764239	0.486999	5.460709

表 3 隐层与输出层之间的连接权值

Table 3 Connective weights between hidden layer and output layer

Output node	Hidden node					Bias node
	1	2	3	4	5	
1	4.909540	2.990715	-4.311211	-3.961316	-4.595607	0.147922
2	-5.107095	-1.921578	4.390949	4.097011	4.634393	-1.133527

确定了 BP 网络各节点间的连接权值, 我们便建立了炸药组分的拓扑参数与保留值参数  $\log k_w$  和  $S$  之间的人工神经网络模型, 将该模型的预测值与文献 [14] 所报道的实验值相比较, 其结果见表 1。

从表 1 中的数据可以看出, 以 20 种炸药组分为样本所建立的人工神经网络模型较好地反映了炸药分子结构与保留值参数之间的复杂关系, 模型的预测值与文献 [14] 所报道的实验值非常接近。该模型的预测精度较高, 误差绝大部分落在  $\pm 5\%$  内, 说明所建立的定量关系具有一定的可类推性, 可以用于炸药组分的反相 HPLC 色谱保留值估算。

### 参考文献:

- [1] Valko R, Snyder L R, Glajch J L. J Chromatogr, 1993, 656:501-521
- [2] Schoenmaker P J, Billiet H A H, Tijssen R, et al. J Chromatogr, 1978, 149:519-537
- [3] LU Pei-zhang, DAI Chao-zheng. Fundamental theory of chromatography. 2nd Ed. Beijing: Science Press, 1997. 263-291  
卢佩章, 戴朝政. 色谱理论基础. 第 2 版. 北京: 科学出版社, 1997. 263-291
- [4] Geng X D, Regnies F E. J Chromatogr, 1985, 332:147-168
- [5] LU Pei-zhang, ZHANG Yu-kui, LIANG Xin-miao.

High performance liquid chromatography and its expert system. Shenyang: Liaoning Science and Technology Press, 1992. 429-461

卢佩章, 张玉奎, 梁鑫淼. 高效液相色谱法及其专家系统. 沈阳: 辽宁科学技术出版社, 1992. 429-461

- [6] ZHAO Rui-huan, YUE Bing-fang, SHAN Yi-chu, et al. Chinese Journal of Chromatography, 1999, 17(2):112-114  
赵瑞环, 岳丙方, 单亦初, 等. 色谱, 1999, 17(2):112-114
- [7] Gobburu J V S, Weilin L S, Shelver W H, et al. J Liq Chromatogr, 1995, 18(10):1957-1972
- [8] LI Rui, GAO Shou-guo, XIANG Bing-ren. Computers and Applied Chemistry, 2000, 17(1/2):113-114  
李睿, 高守国, 相秉仁. 计算机与应用化学, 2000, 17(1/2):113-114
- [9] HUANG Jun, ZHOU Shen-fan, GUO Zhen-shuang, et al. Chinese Journal of Chromatography, 2000, 18(1):1-4  
黄俊, 周申范, 国振双, 等. 色谱, 2000, 18(1):1-4
- [10] SHEN Qing, HU De-wen, SHI Chun. Applied technology of neural networks. Changsha: National Defence Technology University Press, 1993. 415-418  
沈清, 胡德文, 时春. 神经网络应用技术. 长沙: 国防科技大学出版社, 1993. 415-418
- [11] JIAO Li-cheng. Neural networks calculation. Xi'an: Xi'an University of Electronic Science & Technology Press, 1993. 53-65  
焦李成. 神经网络计算. 西安: 西安电子科技大学出版社, 1993. 53-65
- [12] WANG Lian-sheng, HAN Shuo-kui, ZHI Zheng-liang, et al. Organic quantitative structure-activity relationship. Beijing: China Environmental Science Press, 1993. 2-7  
王连生, 韩朔睽, 支正良, 等. 有机物定量结构-活性相关. 北京: 中国环境科学出版社, 1993. 2-7
- [13] DAI Jia-yin, JIN Li-jun, WANG Lian-sheng. Advances in Environmental Science, 1998, 4(4):56-63  
戴家银, 靳立军, 王连生. 环境科学进展, 1998, 4(4):56-63
- [14] ZHOU Shen-fan, WANG Zheng-ping, LIU Xiao-ming. Chinese Journal of Chromatography, 1989, 7(4):200-203  
周申范, 王正萍, 刘小明. 色谱, 1989, 7(4):200-203
- [15] GAO Shou-guo, XIANG Bing-ren, AN Deng-kui. Computers and Applied Chemistry, 1999, 16(5):381-382  
高守国, 相秉仁, 安登魁. 计算机与应用化学, 1999, 16(5):381-382

## Prediction of Retention Parameters of Explosives by Artificial Neural Network

ZHANG Ting-xi<sup>1</sup>, HUANG Jun<sup>2</sup>, ZHOU Shen-fan<sup>2</sup>

( 1. College of Chemistry & Environmental Science, Nanjing Normal University, Nanjing 210097, China; 2. College of Chemical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China )

**Abstract:** The quantitative relationship between the retention parameters and the structure of explosives is discussed. Molecular topological indices are used to represent the structure. Based on the back-propagation algorithm, a quantitative model was established after a training process of a train-set containing 20 explosives being completed. The Sigmoid function was chosen as the transmit function. The retention parameters ( $\log k_w$  and  $S$ ) acted as output vectors, while molecular connecting indices ( ${}^0\chi, {}^1\chi, {}^2\chi$ ) and edge adjacent indices ( $\epsilon$ ) acted as input vectors. The input vectors were normalized in the range of  $-3-3$  and the output vectors were normalized in the range of  $0-1$ . The accuracy of network was  $0.5$  and the beginning value of studying pace ( $\eta$ ) was  $0.2$ , the momentum factor ( $\alpha$ ) was  $0.5$ . The results showed that the yield model reflected the relationship between the structure and retention index of compounds, and had high accuracy. Most of the relative errors were below  $\pm 5\%$ .

**Key words:** retention parameter; artificial neural network; quantitative structure retention relationship; molecular topological index