农药学学报 2007, 9(4): 324-329
Chinese Journal of Pesticide Science

·研究论文·

基于支持向量机回归与 K-最近邻法的组合预测用于除草剂 QSAR建模

谭泗桥[a, b], 袁哲明[*a], 柏连阳[a], 熊洁仪[a]
(湖南农业大学, a 生物安全科学技术学院, b. 信息科学技术学院, 长沙 410128)

摘 要:为了提高定量构效关系(QSAR)研究的预测精度,发展了一种基于支持向量机回归(SVR)与 K最近邻法(KNN)的组合预测方法:以均方误差(MSE)最小为原则,基于 SVR实施核函数寻优;基于 MSE最小原则与最优核函数以 SVR进行描述符筛选并得到保留描述符;通过"多轮末尾强制淘汰法"揭示各保留描述符对预测精度的影响程度;从保留描述符出发,以不同 KNN预测值反映样本集异质性并构建子模型,然后基于 SVR以留一法实施组合预测。运用该组合预测方法研究苯乙酰胺类除草剂 QSAR建模,结果表明:基于 SVR与 KNN的组合预测方法在参比模型中预测精度最高,具结构风险最小、非线性、能有效克服过拟合、泛化推广能力优异等优点,在 QSAR研究中具有广泛的应用前景。

关键词:支持向量机回归; K最近邻法;组合预测;定量构效关系
中图分类号: O641 **文献标志码**: A **文章编号**: 1008-7303(2007)04-0324-06

The QSAR Modeling of Herbicide Using a Combinatorial Model Based on Support Vector Regression and K-nearest Neighbor

TAN Si-qiao[a, b], YUAN Zhe-ming[*a], BAI Lian-yang[a], XIONG Jie-yi[a]
(a. Bio-safety Science and Technology College, b. College of Information Science and Technology, Hunan Agricultural University, Changsha 410128, China)

Abstract: To improve the precision of QSAR modeling, a novel nonlinear combinatorial forecast method based on support vector regression (SVR) and K-nearest neighbor (KNN) was proposed. Kernels and descriptors optimization based on SVR were evaluated by the rule of minimum MSE value, and "multi-round enforcement optimization" was taken to illuminate the effect of retained descriptors on forecasting precision. The heterogeneity of sample set was characterized by different KNN and multiple sub-models were assembled, then combinational forecast was carried out based on leave-one-out method. The new method had been employed to study for QSAR on a series of herbicidal materials, N-phenylacetamides, and has the highest prediction precision in all reference models. It has the advantages of structural risk minimization, non-linear characteristics, avoiding the over-fit and strong generalization ability, etc. The novel combination model, so, can be widely used in QSAR.

Key words: support vector regression; K-nearest neighbor; combinatorial forecasting; QSAR

收稿日期: 2007-08-05;**修回日期**: 2007-10-26.
作者简介:谭泗桥(1974-),男,湖南茶陵人,在读博士生,讲师,研究方向为模式识别与预测; *通讯作者(Author for correspondence):袁哲明(1971-),男,湖南岳阳人,博士,教授,研究方向为模式识别与预测. **E-mail:** zhmyuan@sina.com
基金项目:国家自然科学基金(30570351);教育部新世纪优秀人才支持计划项目.

定量构效关系 (Quantitative structure-activity relationship, QSAR)是研究化合物分子结构与其生物活性之间规律性关系的重要手段,已在药物设计中被普遍采用[1]。QSAR研究中常用的、基于经验风险最小的数学模型如多元线性回归、逐步线性回归、主成分回归、偏最小二乘回归等对高维、非线性、小样本问题的解析能力有限[2],人工神经网络 (ANN)虽然对非线性问题有良好的解析能力,但易陷入局部极小,且收敛速度慢[3]。

基于统计学习理论 (Statistical learning theory, SLT)的支持向量机 (Support vector machines, SVM)较好地解决了小样本、非线性、高维数、局部极小值等实际问题。SVM结构风险最小,包括分类 (Support vector classification, SVC)和回归 (Support vector regression, SVR),其中 SVR更适用于 QSAR研究[4~8]。由于 SVR采用留一法时训练样本的选取是基于全局的,其预测结果往往并非最优且计算量较大;而 K最近邻法 (K-nearest neighbor, KNN)只选取 K个训练样本,理论上更能反映样本集的异质性并有较优的预测精度与较小的计算开销,但先验地给出每个待测样本的最优 K值相当困难。Bates等证明:将两种无偏的单项预测进行组合,能够得到优于每个单项预测的预测结果[9],即把各子模型的预测结果组合在一个模型中进行预测的方法能提高预测精度。基于以上因素,笔者构建了一种基于 SVR与 KNN的组合预测新方法,并在苯乙酰胺类除草剂对莎草科植物萤蔺 Scirpus juncoides的抑制活性的 QSAR研究中进行了应用,同时与逐步线性回归 (SLR)、基于正交最小二乘法的径向基函数网 (OLS-RBFN)、基于二次多项式的循环子空间回归 (Q-CSR)、采用遗传算法优化的径向基循环子空间回归 (EGA-RBF-CSR)等模型[10]进行了比较,结果令人满意。

1　基于 SVR与 KNN的组合预测方法

1.1　支持向量机回归 (SVR)

支持向量机回归的基本思想是将输入样本空间非线性变换到另一个特征空间,在这个特征空间中构造回归估计函数,而这种非线性变换是通过定义适当的核函数 $K(x_i, x_j)$来实现的。设给定的输入样本 x为 p维向量,n个样本及其输出值可表示为:

$$(x_1, y_1) \cdots (x_n, y_n) \in R^p \times R \qquad (1)$$

则 SVR的学习问题就是一个二次规划问题。通常采用 Vapnik的 ε不敏感损失函数来表示,即指定容许误差 ε,若样本 x误差为 ξ,则当 $|\xi| \leq \varepsilon$ 时不计损失,否则损失计为 $|\xi| - \varepsilon$。回归函数表示为:

$$f(x) = \sum_{i=1}^{n} (a_i - a_1^*) K(x, x_i) + b \qquad (2)$$

其中 a和 a^* 为求解的 p维向量,$\Phi(x)$为从样本空间到高维特征空间的映射函数,$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$为核函数,表示为两个 $\Phi(x)$的点积。可以看出,核函数的选择对于 SVR的预测性能有着重要影响。核函数的采用使得映射函数 $\Phi(x)$不必明确求出,使求解非线性回归成为可能。式 (2)中对应于权值 $(a - a^*)$不为 0的样本 x_i称为支持向量。显然,支持向量的数目决定了计算的复杂度,并且与预测精度存在较强关联[4~8, 11]。

1.2　K最近邻法 (KNN)

KNN是一种基于类比的算法,其基本思想是在多维空间中找到与待测样本最近邻的 K个点 ($1 \leq K \leq n-1$,其中 $K=1$对应最近邻法,采用留一法时 $K=n-1$对应全局预测),并根据这 K个点的类别来判断未知样本的类别。类似的,KNN也可以用于回归估计,即以这 K个点作为训练样本来计算待测点的值。其近邻性用欧氏距离来定义,设两个样本分别为 $x_i = \{x_{i,1}, x_{i,2}, \cdots x_{i,m}\}$与 $x_j = \{x_{j,1}, x_{j,2}, \cdots x_{j,m}\}$,则欧氏距离如式 (3):

$$Dist(x_i, x_j) = \sqrt{\sum_{a=1}^{m} (x_{i,a} - x_{j,a})^2} \qquad (3)$$

SVR以留一法实施预测时选取训练样本是基于全局的,即待测样本之外所有的样本参与训练。由于样本集的异质性,基于全局预测的精度并非总是最优:例如某样本集明显地分为两个亚类,则预测第一亚类的某个样本时,若训练集中包含有第二亚类样本,将对预测结果产生干扰;此时,合适 K值的 KNN预测精度可能高于全局预测。KNN的缺点在于尽管可以利用系统聚类、非线性映射等方法来估算 K值大小,但要先验地给出每一个待测样本的最优 K值仍然是非常困难的[12~14]。

1.3　组合预测

用两个或两个以上不同的预测方法各自对同一预测对象进行预测,再将各个单独的预测结果进行适当组合并得到最终预测结果的预测方法即为组合预测方法。对某一预测对象,利用 m种预测方法 (子模型)得到 m个模型的预测值 Φ_i ($i = 1, 2, \cdots, m$),其组合预测函数为 $\omega = \phi(\Phi_1, \Phi_2, \cdots \Phi_m)$,其中 $\phi(x)$为组合预测所采用的函数。由于

组合预测综合了所有单个预测方法包含的有用信息,所以 ω常常比 φ_i (i = 1, 2, …, m)的结果精度更高[15]。

1.4 基于 SVR与 KNN的组合预测模型构建

KNN 预测要解决最优 K值选择问题,逐一试算搜索最优 K值计算量太大,且其获得的是公用最优 K值。事实上由于待测样本间存在差异,各样本的最优 K值应该是不同的。考虑到组合预测往往较单一预测精度更高,同时为避免最优 K值选择难问题,现提出基于不同 K值的 KNN 实施组合预测。

步骤 1:核函数寻优。不同核函数对 SVR 预测结果有重要影响,本文基于 5种常用核函数以留一法预测,取 MSE最小者为最优核函数。5种核函数分别为:线性核函数 (t = 0),多项式核函数 I(t = 1, d = 2),多项式核函数 II(t = 1, d = 3),径向基核函数 (t = 2)以及 sigmoid核函数 (t = 3)。

步骤 2:描述符筛选。基于最优核函数,以"多轮末尾淘汰法"从包含全部描述符的 SVR模型中以留一法依 MSE最小标准逐次剔除对提高预测精度有不利影响的变量:对第一轮筛选,记 MSE(x_1, x_2, …x_i, …x_p)为全部 p个描述符的均方误差,MSE(x_1, x_2, …x_{i-1}, x_{i+1}, …x_p)为剔除第 i个描述符后的均方误差。如 min[MSE(x_1, x_2, …x_{i-1}, x_{i+1}, …x_p)]≤MSE(x_1, x_2, …x_i, …x_p),表明没有描述符可剔除,汰选结束;反之,剔除第 i个描述符后进入下一轮筛选 (注意此时 p变为 p - 1),直至没有描述符可剔除为止。汰选结束后保留的描述符称为保留描述符,用于后续建模预测。为揭示各保留描述符对预测精度的影响程度,进一步发展了"多轮末尾强制淘汰法":基于最优核函数与保留描述符,每轮强制剔除一个对预测值 MSE影响最小的保留描述符,直至只剩下一个保留描述符为止。

步骤 3:子模型构建。从最优核函数与保留描述符出发,分别计算待测样本与其余 n - 1个样本保留描述符向量的欧氏距离,根据 K值取不同数目近邻样本作为训练样本,以留一法进行预测,预测值即构成一个子模型。根据样本集的规模可在 K∈[1, n - 1]中均匀取 3～10个 K值构建 3～10个子模型;对于大样本,取 K∈[1, L],其中 L ≪ n。活性实测值与预测值组成组合预测样本集,用于后续组合预测 (在组合预测样本集中,子模型活性预测值相当于原样本集的描述符)。

步骤 4:组合预测样本集的核函数寻优及核函数筛选。以 MSE最小为原则,参照步骤 1、2分别进行核函数寻优和子模型筛选,得到最优核函数与保留子模型。

步骤 5:组合预测。基于最优核函数和保留子模型以留一法实施预测。

1.5 预测评价指标及算法实现

预测结果的优劣采用均方误差 (Mean Squared Error, MSE)和平均绝对误差百分率 (Mean Absolute Percentage Error, MAPE)作为评价指标[16]:

$$MSE = \frac{1}{n}(y - \hat{y})^2$$

$$MAPE(\%) = \frac{1}{n}\sum\frac{|y - \hat{y}|}{y} \times 100$$

式中 y为实测毒性值,$\hat{y}$为拟合或预测毒性值,n为样本数。

本研究以自编程序通过调用 LIBSVM 完成上述过程并经逐步验证通过。LIBSVM 是一个简单有效、易于操作的通用 SVM 软件包,可以解决分类问题、回归问题以及分布估计等。LIBSVM 是开源的软件包,可以免费从网址 http://www.csie.ntu.edu.tw/～cjlin/libsvm 处获得。有关 LIBSVM 的详细内容参见文献[6～8]。

2 结果与讨论

2.1 样本数据说明

样本集数据采自文献[17, 18](见表 1),共 50种 N-(1-甲基-1-苯乙基)苯乙酰胺类化合物,7个理化参数描述符分别为:疏水性参数 lgp及其平方项 lg^2p,电性效应参数 δ, Taft立体参数 E_S,摩尔折射度 M_R,一阶分子连接性指数 ^1X,二阶分子连接性指数 ^2X。除草活性为化合物对莎草科植物萤蔺 Scirpus juncoides生长的抑制活性,记为 PI_{50}, $PI_{50} = lg(1/IC_{50})$。

2.2 基于 SVR的核函数寻优与描述符筛选

基于 SVR,采用 5种核函数分别对样本集 (50个样本,7个描述符)以留一法预测算得 MSE值,MSE最小的 t = 0 (MSE = 0.230)为最优核函数。以 t = 0进行描述符筛选,7个描述符全部得到保留 (表 2),表明这 7个描述符与化合物活性均有关联,应全部保留用于后续 KNN 预测。"多轮末位强制淘汰"结果显示,各保留描述符对毒性预测影响的相对重要性排序如下:$M_R < E_S < lg^2p < {}^2X < lgp < {}^1X < \delta$,即 δ对预测精度影响最大,M_R对预测精度影响最小 (表 2)。

表 1 苯乙酰胺类化合物结构参数与其对萤蔺生长的抑制活性

Table1 Structural parameters and anti-Scirpus juncoides activity of N-phenylacetamides

序号 No.	lgP	σ	M_R	E_S	^1X	^2X	lg^2P	除草活性 Activity, PI_{50}
1	1.12	-0.34	9.24	0	7.31	5.59	1.254	5.85
2	0.70	-0.11	4.51	0.78	6.93	5.17	0.490	5.68
3	1.27	0.06	9.62	0.27	7.42	5.70	1.613	5.72
4	0.54	-0.44	11.46	0.69	7.43	5.48	0.292	6.07
5	1.12	-0.24	9.24	0	7.31	5.65	1.254	5.89
6	1.27	0.20	9.62	0.27	7.41	5.77	1.613	6.00
7	1.27	0.06	9.62	0.27	7.42	5.70	1.662	5.58
8	0.85	0.29	4.89	1.05	7.03	5.28	0.722	5.44
9	1.42	0.46	10.00	0.54	7.52	5.81	2.016	5.65
10	0.69	-0.04	11.84	0.96	7.53	5.59	0.476	5.46
11	1.27	0.16	9.62	0.27	7.41	5.76	1.613	5.57
12	1.42	0.60	10.00	0.54	7.51	5.88	2.016	5.65
13	0.69	0.10	11.84	0.96	7.52	5.63	0.476	5.42
14	1.27	0.06	9.62	0.27	7.41	5.76	1.613	5.18
15	1.42	0.46	10.00	0.54	7.51	5.88	2.016	5.41
16	1.12	-0.24	9.24	0	7.31	5.65	1.254	5.79
17	0.70	-0.01	4.51	0.78	6.92	5.23	0.490	5.92
18	1.27	0.16	9.62	0.27	7.41	5.75	1.613	6.10
19	0.54	-0.34	11.46	0.69	7.42	5.53	0.292	5.79
20	1.12	-0.14	9.24	0	7.30	5.71	1.254	5.82
21	1.27	0.30	9.62	0.27	7.40	5.83	1.613	6.22
22	1.44	0.36	8.61	-1.16	7.39	5.66	2.074	5.73
23	1.42	0.16	12.47	0.08	7.80	6.27	2.016	4.64
24	0.94	-0.31	16.06	0.03	8.00	5.80	0.884	4.32
25	0.70	0.17	4.51	0.78	6.92	5.20	0.490	6.47
26	0.85	0.57	4.89	1.05	7.02	5.31	0.722	6.57
27	1.27	0.20	9.62	0.27	7.41	5.76	1.613	6.49
28	0.85	0.43	4.89	1.05	7.02	5.34	0.722	7.04
29	1.42	0.60	10.00	0.54	7.51	5.87	2.016	6.71
30	0.69	0.10	11.84	0.96	7.52	5.65	0.476	6.53
31	1.27	0.30	9.62	0.27	7.40	5.83	1.613	6.38
32	1.42	0.74	10.00	0.54	7.51	5.94	2.016	6.95
33	1.57	0.62	12.85	0.35	7.90	6.32	2.465	6.47
34	0.56	-0.34	9.24	0	7.31	5.64	0.314	5.82
35	0.70	-0.11	4.51	0.78	6.92	5.22	0.490	6.01
36	1.27	0.06	9.62	0.27	7.41	5.75	1.613	6.22
37	0.54	-0.44	11.46	0.69	7.42	5.53	0.292	5.82
38	0.56	-0.24	9.24	0	7.30	5.70	0.314	5.89
39	1.27	0.20	9.62	0.27	7.40	5.82	1.613	5.92
40	0.70	-0.11	4.51	0.78	6.92	5.19	0.490	6.25
41	0.85	0.29	4.89	1.05	7.02	5.31	0.722	7.05
42	0.12	-0.21	6.73	1.47	7.03	5.08	0.014	6.44
43	0.70	-0.01	4.51	0.78	6.91	5.26	0.490	6.50
44	0.85	0.43	4.89	1.05	7.02	5.37	0.722	6.39
45	1.27	0.06	9.62	0.27	7.41	5.76	1.613	6.68
46	0.85	0.29	4.89	1.05	7.02	5.34	0.722	6.63
47	1.42	0.46	10.00	0.54	7.51	5.87	2.016	6.91
48	1.27	0.16	9.62	0.27	7.40	5.82	1.613	6.84
49	1.00	0.29	7.74	0.86	7.41	5.79	1.00	6.60
50	1.57	0.60	12.85	0.35	7.90	6.39	2.465	6.49

表 2 基于 MSE值的描述符法选结果

Table 2 The results of descriptors optimization based on MSE

描述符 Descriptors	lgP	σ	M_R	E_S	^1X	^2X	lg^2P	淘汰顺序 Sequence
第一轮 Round 1	0.248 5	0.251 1	0.239 5	0.243 0	0.248 7	0.242 9	0.246 3	M_R
第二轮 Round 2	0.240 4	0.249 9	—	0.238 1	0.265 3	0.242 3	0.238 3	E_S
第三轮 Round 3	0.240 8	0.260 8	—	—	0.266 6	0.240 7	0.238 0	lg^2P
第四轮 Round 4	0.246 5	0.282 1	—	—	0.261 5	0.235 8	—	^2X
第五轮 Round 5	0.236 1	0.308 0	—	—	0.260 9	—	—	lgP
第六轮 Round 6	—	0.318 0	—	—	0.287 2	—	—	^1X

2.3 子模型构建与组合预测

实际应用时我们无法先验地预知最优 K值。从训练集出发搜索最优 K值一方面计算量太大,另一方面每个待测样本的最优 K值显然不同,不存在一个公用的最优 K值。由于组合预测往往有更高的预测精度和稳定性,现考虑采用组合预测方法。根据样本集数目 (n=50),均匀选取 K=1、7、14、21、28、35、42、49(其中 K=1对应最近邻法,K=49对应全局法,其余 K值对应近邻群预测)参与组合预测。基于最优核函数 t=0根据不同 K值分别构建 8个子模型,子模型预测值与实测值构成"组合预测样本集"(50×9的矩阵)。以留一法用 SVR对"组合预测样本集"进行核函数寻优和子模型筛选,结果最优核函数为 t=3,保留子模型为 K=1、14、21、28、42、49。

基于 SVR与 KNN的组合预测及各参比模型留一法预测结果见表 3。从 MSE看,在 8个 KNN子模型中,基于全局的 K=49预测精度不是最优的,表明样本集存在异质性;虽然 K=1的预测精度在 8个子模型中最高,但是有可能还有精度更高的 K值,且无法先验获知最优 K值;而基于 SVR与 KNN的组合预测(模型 E)的预测精度优于近邻群预测、全局预测和最近邻预测,表明组合预测比单个模型预测的精度更高,稳定性更好。

逐步线性回归 SLR(模型 D)预测精度较低,其 MSE=0.307, MAPE=7.703%,远大于其他非线性模型,反映出苯乙酰胺类化合物分子结构与其活性之间更多地呈现为非线性关系;基于 SVR与 KNN的组合预测同样明显优于基于径向基函数网及其改进算法的参比模型 A、B、C等非线性模型[10],在所有参比模型中具最优的预测精度。

表 3 基于不同预测方法的苯乙酰胺类化合物 QSAR建模性能比较

Table 3 Comparison of QSAR modeling for N-phenylacetamides based on different methods

评价指标 Evaluation index	K最近邻法(KNN) K-near neighbor model								模型 A①	模型 B②	模型 C③	模型 D④	模型 E⑤
	K=1	K=7	K=14	K=21	K=28	K=35	K=42	K=49					
MSE	0.063	0.211	0.278	0.267	0.261	0.278	0.340	0.294	0.296	0.275	0.176	0.307	0.061
MAPE(%)	2.887	5.505	6.859	6.854	6.146	7.345	7.815	7.456	7.214	6.814	5.270	7.703	2.689

注:①基于最小正交二乘法的循环子空间回归模型;②基于二次多项式的循环子空间回归模型;③遗传算法优化的径向基一循环子空间回归模型;④逐步线性回归模型;⑤基于 SVR与 KNN的组合预测模型。

Note: ①: Model A, OLS-RBFN model; ②: Model B, Q-CSR model; ③: Model C, EGA-RBF-CSR model; ④: Model D, SLR model; ⑤: Model E, Combinatorial forecasting model based on SVR & KNN.

3 结论

基于 SVR与 KNN的组合预测具有很多优点。首先,其结构风险最小并较好地解决了小样本、非线性、过拟合、维数灾难和局部极小等问题,泛化推广能力优异;其次,其描述符和子模型筛选是非

线性的;第三,它从两个方向对训练集进行了优化,包括列方向的描述符筛选和行方向的近邻群选取;第四,其最终组合预测也是非线性的并具有组合预测精度高与稳定性强的特性。SVR所固有的两个缺点在基于近邻群的非线性 SVR组合预测中也得到了明显或部分的克服:对大样本而言,SVR计算复杂度高,但通过选取合适的近邻群 K值 (K∈[1, L],其中 L << n)可明显降低计算复杂度;由于不存在一个显性的表达式,SVR对描述符欠缺解释能力,但通过强制汰选可给出各描述符对毒性预测影响的相对重要性排序,因而具有部分的解释能力。

致谢:本文数据集由施彦、陈德钊提供,谨表谢忱。

参考文献:

[1] ZHONG Guo-hua(钟国华), HU Mei-ying(胡美英). QSAR及其在农药设计中的应用和进展[J]. Chin J Pestic Sci(农药学学报), 2001, 3(2): 1-11.

[2] YIN Jia-jian(印家健), LI Meng-long(李梦龙), WEN Zhi-ning(文志宁). 支持向量回归用于氨基酸描述符在肽 QSAR建模中的性能评价[J]. J Sichuan Univ (Nat Sci Ed)(四川大学学报,自然科学版), 2006, 43(2): 396-402.

[3] ZHOU Peng(周鹏), ZENG Hui(曾晖), LI Bo(李波). 支持向量机分类和回归用于肽的 QSAR研究[J]. Chemistry(化学通报), 2006, 69(5): 342-346.

[4] DING Jun-jie(丁俊杰), DING Xiao-qin(丁晓琴), ZHAO Li-feng(赵立峰). 新型三维氨基酸结构描述符的研究及其在多肽 QSAR中的应用[J]. Acta Pharmaceutica Sinica(药学学报), 2005, 40(4): 340-346.

[5] AINa(艾娜), WU Zuo-wei(吴作伟), REN Jiang-hua(任江华). 支持向量机与人工神经网络[J]. J Shandong Univ Tech (Sci & Tech)(山东理工大学学报,自然科学版), 2005, 19(5): 45-49.

[6] VAPNIK V. The Nature of Statistical Learning Theory [M]. New York: Springer Verlag Press, 1995: 87-189.

[7] CRISTIANNIN, SHAWE-TAYLOR J. An Introduction to Support Vector Machines and other Kernel-based Learning Methods(支持向量机导论)[M]. LI Guo-zheng(李国正, 等译). Beijing(北京): Publishing House of Electronic Industry (电子工业出版社), 2004: 82-139.

[8] DENG Nai-yang(邓乃扬), TIAN Ying-jie(田英杰). The New Method of Data Mining-Support Vector Machine(数据挖掘中的新方法——支持向量机)[M]. Beijing(北京): Science Press(科学出版社), 2004.

[9] BATES J M, GRANGER C W J. Combination of Forecasts[J]. Operations Research Quarterly, 1969, 20(4): 451-468.

[10] LI Jian(李剑), CHEN De-zhao(陈德钊). 优化的径向基循环子空间网络为药物定量构效关系建模[J]. Chin J Anal Chem(分析化学), 2005, 33(6): 767-771.

[11] TIAN Sheng-feng(田盛丰), HUANG Hou-kuan(黄厚宽). 支持向量机回归的简化算法[J]. Journal of Software(软件学报), 2002, 13(6): 1169-1172.

[12] WANG Xiao-ye(王晓晔), WANG Zheng-ou(王正欧). K最近邻分类技术的改进算法[J]. Journal of Electronics & Information Technology(电子与信息学报), 2005, 27(3): 487-491.

[13] ZHANG Guo-ping(张国平). B-G组合预测剖析[J]. Forecasting(预测), 1988, 12(5): 24-27.

[14] CAI Gen-xiang(柴根象), WU Yue-qin(吴月琴). 相依样本下污染线性模型的最近邻估计[J]. Acta Mathematicae Applicatae Sinica(应用数学学报), 2006, 29(3): 542-554.

[15] LI Yuan-cheng(李元诚), LI Bo(李波), FANG Ting-jian(方廷健). 基于小波支持向量机的非线性组合预测方法研究[J]. Information and Control(信息与控制), 2004, 33(3): 303-306.

[16] TANG Qi-yi(唐启义), FENG Guang-ming(冯明光). DPS Data Processing System for Practical Statistics(实用统计分析及其 DPS数据处理系统)[M]. Beijing(北京): Science Press (科学出版社), 2002: 579-585.

[17] CHEN Y Q, CHEN D Z, HE C Y. Quantitative Structure-activity Relationships Study of Herbicides Using Neural Networks and Different Statistical Methods[J]. Chemometrics and Intelligent Laboratory Systems, 1999, 45: 267-276.

[18] KIRINO O, TAKAYAMA C, MINE A. Quantitative Structure-activity Relationships of Herbicidal N-(1-Methyl-1-phenylethyl) phenylacetamides[J]. J Pestic Sci, 1986, 11: 611-617.

(Ed. JIN S H)