Full-length article

Using support vector classification for SAR of fentanyl derivatives[1]

Ning DONG, Wen-cong LU[2], Nian-yi CHEN, You-cheng ZHU[3], Kai-xian CHEN[3]

[2]Laboratory of Chemical Data Mining, Department of Chemistry, School of Science, Shanghai University, Shanghai 200436, China; [3]Shanghai Institute of Materia Medica, Chinese Academy of Sciences, Shanghai 201203, China

Key words

structure-activity relationship; support vector machine; fentanyl derivatives; support vector classification

[1] Project supported by the National Natural Science Foundation of China (No 20373040).
[2] Correspondence to Prof. Wen-cong LU.
Phn 86-21-6613-3513. Fax 86-21-6613-4275.
E-mail wclu@mail.shu.edu.cn

Received 2004-06-15
Accepted 2004-09-10

doi: 10.1111/j.1745-7254.2005.00014.x

Abstract

Aim: To discriminate between fentanyl derivatives with high and low activities. **Methods:** The support vector classification (SVC) method, a novel approach, was employed to investigate structure-activity relationship (SAR) of fentanyl derivatives based on the molecular descriptors, which were quantum parameters including ΔE [energy difference between highest occupied molecular orbital energy (HOMO) and lowest empty molecular orbital energy (LUMO)], MR (molecular refractivity) and M_r (molecular weight). **Results:** By using leave-one-out cross-validation test, the accuracies of prediction for activities of fentanyl derivatives in SVC, principal component analysis (PCA), artificial neural network (ANN) and K-nearest neighbor (KNN) models were 93%, 86%, 57%, and 71%, respectively. The results indicated that the performance of the SVC model was better than those of PCA, ANN, and KNN models for this data. **Conclusion:** SVC can be used to investigate SAR of fentanyl derivatives and could be a promising tool in the field of SAR research.

Introduction

Fentanyl, a synthetic opioid commonly used during anesthesia, is also used to relieve pain in terminally ill patients[1]. Fentanyl is lipophilic and has high potency as an analgesic or anesthetic, which can rapidly penetrate the central nervous system once taken by patients[2].

Support vector classification (SVC) is a machine learning method based on the support vector machine (SVM) proposed by Vladimir N Vapnik[3]. It has been recently proposed as a very effective method for pattern recognition. It has also been successfully used in such research fields as vowel recognition[4], drug design[5], combinatorial chemistry[6], prediction of beta-turn and alpha-turn types of proteins etc[7,8]. In the present work, the qualitative model was built based on SVC, with structural descriptors calculated by using the software Hyperchem, to explore the structure-activity relationship of fentanyl derivatives. The outstanding performance of the SVC model proved the significance of this method.

Methodology

Computational theory The SVC method was used in this work. The geometrical interpretation of SVC is that it chooses the optimal separating surface, ie the hyperplane equidistant from two classes. This optimal separating hyperplane has many nice statistical properties, which are detailed by Vapnik[3,9].

Consider the problem of separating the set of training vectors belonging to two separate classes, $(y_1, \mathbf{x}_1)$, ..., $(y_n, \mathbf{x}_n)$, $\mathbf{x}\in R^m$, $y\in$-1, +1, with a hyperplane

$$\mathbf{w}^T\mathbf{x}+b=0$$

If the training data are linearly separable, then there exists a pair $(\mathbf{w}, b)$ such that:

$$y_i(\mathbf{w}^T\mathbf{x}_i+b)-1\geq 0,\ i=1, 2, ..., l$$
$$\mathbf{w}^T\mathbf{x}+b\geq+1, \text{ for all } \mathbf{x}\in\boldsymbol{T};$$
$$\mathbf{w}^T\mathbf{x}+b\geq-1, \text{ for all } \mathbf{x}\in\boldsymbol{F};$$

The decision rule is:

$$f_{\mathbf{w},b}(\mathbf{x})=\text{sgn}(\mathbf{w}^T\mathbf{x}+b)$$

where $\mathbf{w}$ is termed the weight vector and b the bias. Without loss of generality the pair $(\mathbf{w}, b)$ can be rescaled such that:

$$\min_{i=1,2,\ldots,l} | \mathbf{w}^T\mathbf{x}_i+b | =1$$

The learning problem is hence reformulated as: minimize $\|\mathbf{w}\|^2$ subject to the constraints of linear separability. This is equivalent to maximizing the distance, normal to the hyperplane, between the convex hulls of two classes. The optimization is now a quadratic programming (QP) problem:

$$\underset{\mathbf{w},b}{Minimize}\ \phi(\mathbf{w})=\frac{1}{2}\|\mathbf{w}\|^2$$

subject to $y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1$, $i=1, 2, \ldots, l$.

This problem has a global optimum. The Lagrangian for this problem is:

$$L(\mathbf{w}, b, \Lambda)=\frac{1}{2}\|\mathbf{w}\|^2-\sum_{i=1}^{l}\lambda_i[y_i(\mathbf{w}^T\mathbf{x}_i+b)-1]$$

where $\Lambda=\{\lambda_1,\mathrm{K},\ldots,\lambda_l\}$ are the Lagrange multipliers, one for each data point.

Hence we can write:

$$F(\Lambda)=\sum_{i=1}^{l}\lambda_i-\frac{1}{2}\|\mathbf{w}\|^2=\sum_{i}^{l}\lambda_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\lambda_i\lambda_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j$$

note that the Lagrange multipliers are only non-zero when $y_i(\mathbf{w}^T\mathbf{x}_i+\mathrm{b})=1$, vectors for these cases are called support vectors since they lie closest to the separating hyperplane. Then the optimal separating hyperplane is given by:

$$\mathbf{w}^*=\sum_{i=1}^{l}\lambda_i^*\mathbf{x}_i y_i$$

and the bias is given by:

$$b^*=-\frac{1}{2}(\mathbf{w}^*)^T(\mathbf{x}_s+\mathbf{x}_r)$$

where $\mathbf{x}_r$ and $\mathbf{x}_s$ are any support vector from each class satisfying

$$y_r=1, y_s=-1$$

The hard classifier is then,

$$f(\mathbf{x})=\mathrm{sgn}[(\mathbf{w}^*)^T\mathbf{x}+b^*]$$

In the case where a linear boundary is inappropriate the SVC can map the input vector, $\mathbf{x}$, into a high dimensional feature space, F. By choosing a non-linear mapping Φ, the SVC constructs an optimal separating hyperplane in this higher dimensional space. Among the acceptable mappings are polynomials, radial basis functions and certain sigmoid functions. Then the optimization problem becomes,

$$W(\alpha)=\sum_{i=1}^{l}\alpha_i-\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}y_i y_j\alpha_i\alpha_j<\Phi(\mathbf{x}_i)\cdot\Phi(\mathbf{x}_j)>$$

In this case, the decision function in SVC is as follows:

$$g(\mathbf{x})=\mathrm{sgn}(f(\mathbf{x}))=\mathrm{sgn}\{\sum_{i\in SV}\alpha_i y_i<\Phi(\mathbf{x})\cdot\Phi(\mathbf{x}_i)>+b\}$$

$$=\mathrm{sgn}\{\sum_{i\in SV}\alpha_i y_i K(\mathbf{x},\mathbf{x}_i)]+b\}$$

where $\mathbf{x}_i$ is the support vectors and $K(\mathbf{x}, \mathbf{x}_i)$ is called kernel function.

Implementation of SVC The SVM software package including SVC was programmed according to the literature[3]. The software was tested in some applications in chemistry and chemical engineering[9,10]. All computations were carried out on a Pentium IV computer with a 1.3G Hz processor.

Results

Data set The data set consists of 14 fentanyl derivatives available[11]. The molecular formula investigated in this work is shown in Figure 1. The substituents of the compounds include R_1, R_2, and R_3. The data set can be divided into two classes according to the analgesic bioactivities ED_{50} (hot plate method in mice)[11] of samples. Here Class 1 contains the compounds with high activities, ie the molecules with ED_{50} $<1.0\times10^{-6}$ (mol/kg). Class 2 contains the compounds with low activities, ie the molecules with $ED_{50}>1.0\times10^{-6}$ (mol/kg).

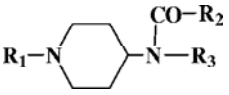


Figure 1. Structure of fentanyl derivatives.

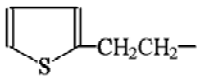
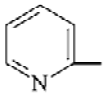
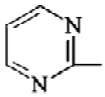
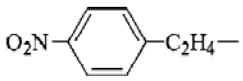
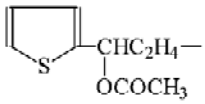
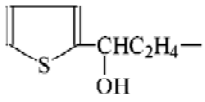
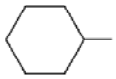
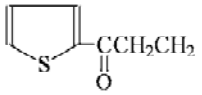
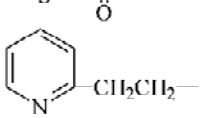
Computation of descriptors The three-dimensional structures of the molecules were drawn, and optimized with the software Hyperchem3 (Release 7.0 for Windows Molecular Modeling System, Hypercube Inc. 2002), which was utilized for the computation of MM+ and PM3 later. Prior to the semi-empirical computation of quantum chemistry, all structures of the compounds were submitted to MM+ computation of molecular mechanics for energy optimization. The computations were carried out at a restricted Hartree-Fock level with no configuration interaction. The molecular structures were optimized using the Polak-Ribiere algorithm until the root-mean-square gradient was 0.001. Only the most stable conformation of molecule has been used to obtain the structural descriptors via the computational results of semi-empirical method PM3. Using the software Hyperchem3, the descriptors obtained were as follows: HOMO (highest occupied molecular orbital energy), LUMO (lowest empty molecular orbital energy), ΔE (energy difference between HOMO and LUMO), TE (total energy), HF (heat of formation), EE (electronic energy), SA (surface area), MV (molecular volume), lgP (partition coefficient), MR (molecular refractivity), MP (molecular polarizability), M_r (molecular weight), N1 (charge density of the atom N connecting with R_1), C2 (charge density of the atom C connecting with R_2), N3 (charge density of the atom N connecting

with R_3).

Selection of descriptors The selection of descriptors is a relatively tough job due to the redundancy of some parameters. The result used to depend on the experience of the researcher. Recently, some of promising results have been reported on the problem of feature selection[12,13]. In this work, the entropy method was applied to the selection of descriptors[14]. Through the computation of entropy for the data set available, the three descriptors (ΔE, MR, M_r) are determined to be more important than the others. Table 1 lists the samples with bioactivity ED_{50} and selected descriptors, including ΔE, MR, and M_r. It should be mentioned that there possibly exist other combinations of descriptors useful for the classification of data set used here, but the three descriptors above are enough to be used as the determining factors for the prediction of activities of the compounds (refer to the good results described in the following sections).

Selection of the kernel function and the capacity of parameter *C* used in the SVC model Similar to other multivariate statistical models, the performance of SVC is related to dependent and independent variables as well as the

Table 1. The descriptors of structures and bioactivities of the samples.

No	R_1	R_2	R_3	ED_{50}/mol·kg^{-1}	ΔE/eV	MR	M_r
1	$PhCH_2CH_2$–	Et–	Ph–	1.70×10^{-10}	8.767	103.5	336.5
2	PhCH(OH)-CH(CH_3)-	Et–	Ph–	1.40×10^{7}	8.947	109.1	366.5
3	$PhCH_2CH_2$–	CH_2=CH–	Ph–	2.20×10^{-7}	9.136	103.5	334.5
4	$PhCH_2CH_2$–	FCH_2CH_2–	Ph–	3.30×10^{7}	9.186	103.6	354.5
5		Et–	Ph–	5.70×10^{-7}	8.752	105.0	342.5
6	PhCH(OH)CH_2—	Et–	Ph–	8.40×10^{-7}	8.797	104.7	352.5
7	$PhCH_2CH_2$–	Et–		8.80×10^{-7}	8.523	104.2	337.5
8	$PhCH_2CH_2$–	Et–		8.90×10^{-7}	8.856	101.3	338.5
9		Et–	Ph–	1.10×10^{-5}	8.057	110.8	381.5
10		Et–	Ph–	1.60×10^{-5}	8.811	120.1	414.6
11		Et–		2.40×10^{-5}	8.502	112.2	378.6
12		Et–	Ph–	3.70×10^{-5}	8.235	110.1	370.5
13		Et–	Ph–	3.80×10^{-5}	8.911	105.4	337.5
14	CH_2=$CHCH_2$–	Et–	Ph–	4.60×10^{-5}	9.238	83.28	272.4

combination of parameters used in a model. In the computation of SVC, we have to deal with the capacity parameter *C* (also called the regularization parameter) and the kernel type used in modeling.

In this work, the cross validation test, using the leaving-one-out (LOO) method was undertaken to find a suitable capacity parameter *C* and the appropriate kernel function for the SVC model. Suppose that P_w is the number of samples misclassified using the LOO method; it can then be employed as a criterion to obtain the appropriate kernel function and the optimal capacity parameter *C*. Figure 2 illustrated P_w (concerned with different kernel functions including linear, radial, polynomial and sigmoid functions) versus the capacity parameter *C* from 0.1 to 250. It was found that the SVC model with the best performance could be ascertained by using the radial kernel function with capacity parameter *C* from 50 to 100.

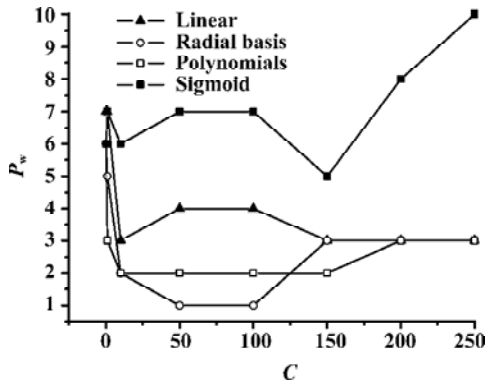


Figure 2. P_w (the number of samples misclassified using LOO method) versus the *C* (capacity parameter) using LOO method with different kernel functions.

Modeling of SVC According to the results we abtained, the optimal model of SVC for discriminating between high and low activities of compounds could be built as follows, using the radial kernel function with capacity parameter *C*=100:

$$g(\mathbf{x})=\mathrm{sgn}\left(\sum_{i\in SV}\alpha_i y_i \exp\{-\frac{\|\mathbf{x}-\mathbf{x}_i\|^2}{\sigma^2}\}+b\right)$$

where σ=1, *b*=0.68, α_i=36.1 (*i*=2), 21.8 (*i*=3), 55.1 (*i*=5), 19.3 (*i*=7), 0 (*i*=10), 30.5 (*i*=11), 100 (*i*=13), 1.84 (*i*=14), correspond to the Lagrange multipliers of support vectors, while all the others. α_i=*0*. y_i=1 for the samples of class 1; y_i=-1 for the samples of class 2. $\mathbf{x}_i$ is a vector (pattern of sample) with unknown activity to be discriminated, $\mathbf{x}_i$ is one of the support vectors. Based on this SVC model, the samples were discriminated as those of high bioactivities (ED_{50}< 1.0×10^{-6} mol/kg), if $g(\mathbf{x})\geq0$. Using SVC model for the classification of activities of fentanyl derivatives, the accuracy of classification was 93% by using radial basis kernel functions with the capacity parameter *C*=100. Table 2 lists the trained results from SVC model obtained above. Figure 3 illustrated the effect of classification with trained SVC model. It was found that only one sample (compound No 13) was misclassified.

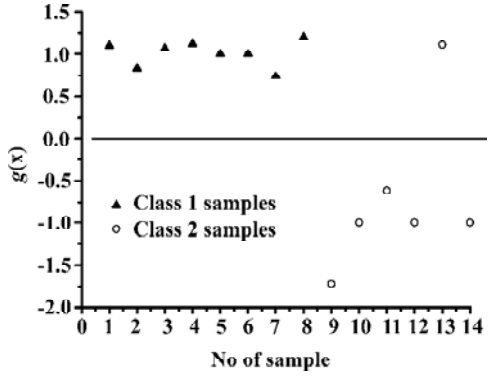


Figure 3. Effect of classification with trained SVC model.

Results of cross validation tests Figure 4 illustrated the effect of the cross validation test using LOO method of SVC. It is obvious that the quality of prediction results (Figure 4) is as good as that of trained results (Figure 3), also with only one sample (compound No 13) discriminated wrong. Table 2 lists the predicted results obtained by using LOO method of SVC model.

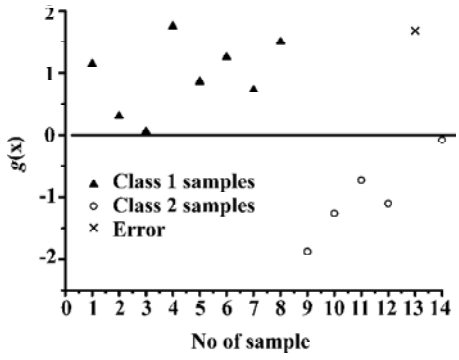


Figure 4. Predicted effect of classification with LOO (leaving-one-out) test of SVC.

For comparisons with other data mining methods, three commonly used chemometric methods including principal

Table 2. Data set available and results of computation using the LOO test.

Sample No	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Actual Class	1	1	1	1	1	1	1	1	2	2	2	2	2	2
TSVC[a] Class	1	1	1	1	1	1	1	1	2	2	2	2	1	2
PSVC[b] Class	1	1	1	1	1	1	1	1	2	2	2	2	1	2
TPCA[c] Class	1	1	1	1	1	1	1	1	2	2	2	2	1	2
PPCA[d] Class	1	1	2	1	1	1	1	1	2	2	2	2	1	2
TANN[e] Class	2	1	2	1	2	1	1	1	2	2	2	2	2	2
PANN[f] Class	1	2	2	1	1	1	2	1	2	2	1	2	1	1
PKNN[g] Class	1	1	1	1	1	1	1	1	1	2	1	1	1	2

a: TSVC Class means trained class by using SVC method; b: PSVC Class means predicted class by using leaving-one-out (LOO) test of SVC method; c: TPCA Class means trained class by using PCA method; d: PPCA Class means predicted class by using leaving-one-out (LOO) test of PCA method; e: TANN Class means trained Class by using ANN method; f: PANN Class means predicted class by using leaving-one-out (LOO) test of ANN method; g: PKNN Class means predicted class by using K-nearest neighbor (KNN) method.

Table 3. The number of samples predicted wrongly using the LOO test of different models.

Algorithm	SVC	PCA	ANN	KNN
N_w (No in Table 2)	1 (13)	2 (3,13)	6 (2,3,7,11,13,14)	4 (9,11,12,13)

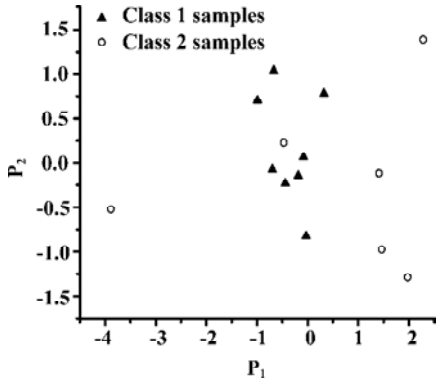


Figure 5. Classification diagram using PCA method.

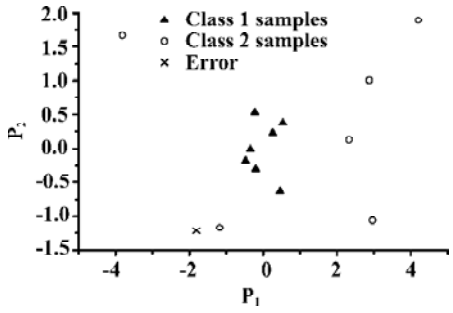


Figure 6. Location of No 3 sample predicted using PCA

component analysis (PCA), K-nearest neighbor (KNN) and artificial neural network (ANN) were utilized to investigate SAR of fentanyl derivatives, with special consideration of their predictive ability (generalization ability) from cross validation tests using LOO method.

Figure 5 illustrated the trained results of classification for the same data set using PCA method. It could be seen from Figure 5 that the quality of classification results was as good as thoses from the SVC model, with only one sample misclassified. However, there were two samples predicted wrong from the results of cross validation test using LOO method of PCA. Figure 6 and Figure 7 illustrated the locations of samples marked No 3 and No 13, whose predicted classes did not agree with the actual ones determined using the LOO test of PCA method.

KNN method, as a helpful pattern recognition tool, was utilized to discriminate between high and low activities of compounds for the same data set, with the accuracy of prediction being 71% (K=5). Obviously, the predictive ability of KNN model was poor compared to that of SVC model in this situation.

So far as BP ANN model is concerned, ANN with three

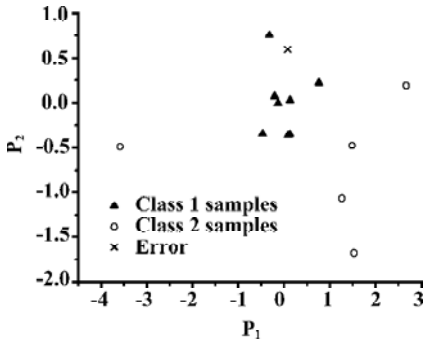


Figure 7. Location of No 13 sample predicted using PCA method.

layers was used to build the relationship between the features and activities of compounds. The number of hidden nodes was three; the transformation function used was Sigmoid; the number of training steps was 250 000. Table 2 lists both trained results and predicted results using LOO method, based on the ANN model built. It was found from Table 2 that the trained results of classification were not as good as the predicted ones. There were only three (No 1, No 3, and No 5) samples misclassified for the trained set, with the accuracy of classification being 79%. However, there were six samples that were wrongly classified from the results of the cross validation test using LOO method of ANN model. Table 3 lists *N*w (the number of samples to be predicted wrong) using LOO test of different models. It could be concluded that the predictive ability of SVC model was superior to that of PCA, KNN, and ANN models for the data set available.

Discussion

The SVC has been introduced as a robust and highly accurate intelligent classification technique, likely to be a useful chemometrics tool. On a simple but real chemometric problem the predictive ability of SVC for the data set available here outperforms that of PCA, KNN and ANN methods, which are the most frequently used chemometric techniques. The SVC exhibits better overall performance because it embodies the structural risk minimization principle. It has an advantage over the other techniques because it converges to the global optimum, and not to a local optimum that depends on the initialization and parameters affecting the rate of convergence. It can be concluded that (1) the selected descriptors can account for the features of the fentanyl derivatives; (2) the SVC is a very promising tool for the approximation of qualitative classification and (3) the SVC is especially suitable for finding the regularities of the small data set, ie, data set with fewer samples, giving modeling results with good generalization ability.

Training and optimization using SVC are easier and faster compared with other machine learning techniques, because there are fewer free parameters and only support vectors (only a fraction of all data) are used in the generalization process. The results show that the SVC is a good approach for predicting the classes of fentanyl derivatives. At the same time, the models proposed could identify and provide some insight into what features are related to the classification of these compounds and afford some instruction for further recognizing new fentanyl derivatives. It should be noted that no single method or paradigm is uniformly superior, although the preliminary evidence presented in this work suggests that the SVC is a data-mining tool with great potential in chemometric application.

References

1 Maddocks I. Subcutaneous administration of fluid and drugs in palliative care. Aust J Hosp Pharm 1992; 22: 181–4.
2 Gerak LR, Moerschbaecher JM, Bagley JR, Brockunier LL, France CP. Effects of a novel fentanyl derivative on drug discrimination and learning in rhesus monkeys. Pharmacol Biochem Behav 1999; 64: 367–71.
3 Vapnik VN. Statistical learning theory. New York: John Wiley and Sons; USA; 1998.
4 Wan V, Campbell WM. Support vector machines for speaker verification and identification. Neural networks for signal processing X, Proceedings of the 2000 IEEE Signal Processing Workshop. IEEE 2000; 2: 775–84.
5 Burbidge R, Trotter M, Buxton B, Holden S. Drug design by machine learning support vector machines for pharmaceutical data analysis. Comput Chem 2001; 26: 5–14.
6 Trotter M, Buxton B, Holden S. Support vector machines in combinatorial chemistry. Meas Control 2001; 34: 235–9.
7 Cai YD, Liu XJ, Li YX, Xu XB, Chou KC. Prediction of beta-turns with learning machines. Peptides 2003; 24: 665–9.
8 Cai YD, Feng KY, Li YX, Chou KC. Support vector machine for predicting alpha-turn types. Peptides 2003; 24: 629–30.
9 Lu WC, Chen NY, Ye CZ, Li GZ. Introduction to the algorithm of support vector machine and the software ChemSVM. Comput Appl Chem 2002; 19: 697–702.
10 Chen NY, Lu WC. Support vector machine applied to chemistry and chemical technology. Comput Appl Chem 2002; 19: 673–6.
11 Zhu YC, Fang SN, Ge BL, Li QZ, Dai QY, Huang ZM, *et al*. Studies on potent analgesics 2. Synthesis and analgesic activity of the derivatives. Acta Pharm Sin 1981; 16: 97–104. [in Chinese]
12 Li GZ, Wang ZX, Yang J, Yao LX, Chen NY. A SVM-based feature selection method and its applications. Comput Appl Chem 2002; 19: 703–5.
13 Dash M, Liu H. Feature selection for classification. Intell Data Anal 1997; 1: 131–56.
14 Chen NY, Qin P, Chen RL, Lu WC. Pattern recognition applied to chemistry and chemistral industry. Beijing: Science Press; 2000.