Proceedings of the 2008 International Conference on Wavelet Analysis and Pattern Recognition, Hong Kong, 30-31 Aug. 2008

A RBF NEURAL NETWORK MODEL FOR ANTI-MONEY LAUNDERING

LIN-TAO LV, NA JI, JIU-LONG ZHANG

Institute of Computer Science and Engineering, Xi' an University of Technology, Xi 'an 710048, China
E-MAIL: lvlintao@xaut.edu.cn

Abstract:

Money laundering (ML) is a serious crime which makes it necessary to develop detection methods in transactions. Some researches have been carried on, but the problem is not thoroughly solved. Aiming at the low detection rate of suspicious transaction at home and abroad in financial field, and with an analysis of radial basis function (RBF) neural network, we propose a radial basis function neural network model based on APC-III clustering algorithm and recursive least square algorithm for anti-money laundering (AML). APC-III clustering algorithm is used for determining the parameters of radial basis function in hidden layer, and recursive least square (RLS) algorithm is adopted to update weights of connections between hidden layer and output layer. The proposed method is compared against support vector machine (SVM) and outlier detection methods, which show that the proposed method has the highest detection rate and the lowest false positive rate. Thus our method is proved to have both theoretical and practical value for anti-money laundering.

Keywords:

Anti-money laundering; Neural network; Radial basis function; APC-III clustering; Recursive least square; Support vector machine; Outlier detection

1. Introduction

Money laundering (ML) usually refers to such activity or process that deals with criminal proceeds to disguise their illicit origin and make them appear legitimate. In recent years, money laundering activity is becoming more and more rampant all over the world. The International Monetary Fund (IMF) estimates that the aggregate size of ML in the world could be 5% of the global gross domestic product (GDP), equivalent to approximately 3,000 billion. This shows that money laundering has seriously threatened the economic development and social safety of the global. From reference [1], we can see that anti-money laundering issue is much more obvious and needs urgent settlement.

Several detecting technologies nowadays have been applied in AML field are proposed, in reference [2, 3], decision tree classification and outlier analysis tools were introduced in extracting suspicious financial transaction. And in reference [4, 5, 6], such as classification analysis tools (Bayesian classification, Genetic algorithms etc), cluster analysis tools, support vector machine(SVM), case-based reasoning, rough set, link analysis, intelligent agents and neural network were mentioned. Outlier detection is used to discover data object which is not consistent with the features of general data, but it can't recognize the suspicious behavior of peer comparison. SVM can get good result, but the time cost is high. A high accuracy detection method in financial field is badly needed due to the massive data in financial institution and high error detection rate. Since RBF neural network can perform calculations from time to time with its units, it can determine whether the capital flow is involved in money laundering activity. It does not involve the complex process, thus saving much time in tracking ML activity. Consequently, the RBF neural network model proposed in this paper for AML is a new attempt to automate the process of AML as much as possible.

2. Architecture of RBF neural network

We adopt the radial basis function (RBF) method introduced in [7～14]. It can overcome the longer training time and the difficulty in determining hidden layer units of back propagation (BP) network to a large extent. An RBF network is a three-layer feed-forward neural network which consists of an input layer, a hidden layer, and an output layer. Let the number of input layer, hidden layer and output layer be p, m, n respectively. For any sample $x=[x_1,x_2,..,x_P]$ in the training set $X=[X_1,X_2,...,X_N]^T$, the output is $y=[y_1,y_2,\ldots,y_n]$ .The model of a RBF network can be described with equations:

$$\{x\in R_p\}\xrightarrow{\varphi(.)}\{h\in R_m\}\xrightarrow{w_j}\{y\in R_n\} \tag{1}$$

978-1-4244-2239-5/08/$25.00 ©2008 IEEE

$$y_j = f_j(x) = w_0 + \sum_{i=1}^{m} w_j \phi(\|x - c_i\|)$$
$$(i = 1, 2, ..., m, j = 1, 2, ..., n) \quad (2)$$

where $x \rightarrow \phi_i(x)$ is a non-linear mapping from the input layer to the hidden layer, while the mapping from hidden layer to output layer, i.e. $\phi_i(x)$ to y_j , is linear. $w_j = [w_{1j}, w_{2j}, ..., w_{mj}]^T \quad (j = 1, 2, ..., n)$ is the connection weight value between the hidden layer and the output layer, $\|.\|$ represents the Euclidean norm, $\phi(.) = \{\varphi_0 = 1, \varphi_1(.), ..., \varphi_h(.)\}$ is a radial basis function from R^+ to R ,it has several forms, we generally choose the Gaussian function:

$$\phi(\|x - c_i\|) = \begin{cases} \exp(-\frac{\|x-c_i\|^2}{\sigma_i^2}) & i = 1, 2, \ldots, m \\ 1 & i = 0 \end{cases} \quad (3)$$

where $c_i = [c_i^1, c_i^2, ..., c_i^m] \in R_m (i = 1, 2, ..., m)$ represents the ith center in the hidden layer, σ_i controls the attenuation speed of Gaussian function, m is the number of units in the hidden layer.

As shown in (1), (2) and (3) , there are three elements in designing a RBF neural network i.e. the centre vectors c_i , the width of the radial basis function σ_i , and the linear weights w_j . Fig.1 illustrates the architecture of RBF neural network,

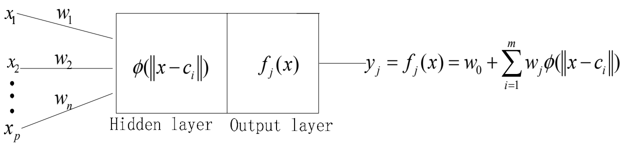

Figure 1 Structure of RBF neural network

3. RBF neural network model for anti-money laundering

By investigating the accounts of the bank, we discovered that if we start with the transaction among the accounts, the relatively real information of the clients can be acquired. The training set is composed of a set of records in database. Each record includes some attributes, forming a feature vector. Besides, there is a unique class label corresponding to each sample in training set, the specific form of a training sample data sets is expressed as:

$$E = (v_1, v_2, v_{3,...} v_p; s) \quad (4)$$

where $v_1, v_2, v_3, \ldots, v_p$ denote the input attributes for anti-money laundering, s is the value of the class label. The available attributes include client number, client name, capital account, certificate number of client, transaction date, business types affiliated with the transaction account, code of the transaction area, transaction sum, transaction time, transaction currency, transaction types, and frequencies of transaction.

Through the analysis of the paper above, the RBF neural network model proposed for anti-money laundering is depicted in Fig.2.

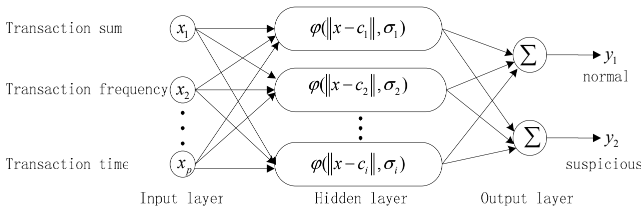

Figure 2 RBF neural network model for AML

4. Model realization of RBF neural network for anti-money laundering

The large and suspicious transaction data used in AML is provided by the financial institution, we can consider these transaction data provided and the case data summarized by money laundering case as the training samples. We then use trained RBF neural network to judge whether the transaction data preprocessed is illegal.

When the suspicious transaction data is confirmed to be real money laundering transaction data, we add it to the case data and start to learn the parameters again. The flow chart of model realization is below.

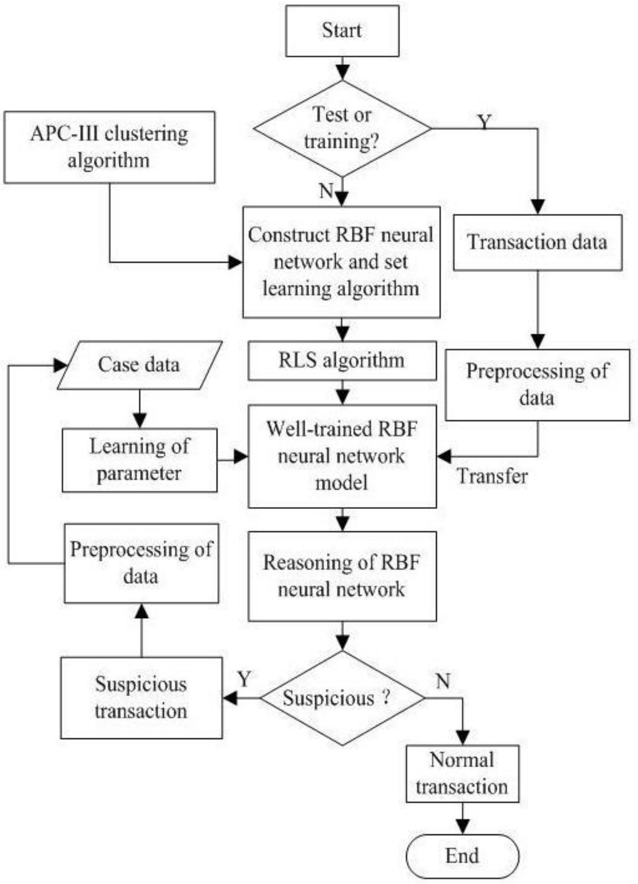

Figure 3 Flow chart of model realization

4.1. Preprocessing of attribute parameter

Since bank transaction records are typical high dimensional and heterogeneous data, and RBF neural network can only deal with numerical data, we should preprocess the attributes before they are presented to the work. For the qualitative attribute variable, we can classify them with the grade. For a given dataset U, suppose Q is a subset of U, then for each data in Q there is a mathematical function C which maps the data to a value between zero and one, i.e. $\forall Q \subseteq U, C: Q \rightarrow [0,1]$. In this paper we utilize the normalized difference

$$\xi_i = \max_i(x_{ij}) - \min_i(x_{ij}) \tag{5}$$

where ξ_i represents the normalized difference (difference between the maximum and minimum value of each attribute in a training sample), x_{ij} denotes the value of jth attribute of the ith sample.

The normalization is to convert each data x_{ij} to $x_{ij}' = (x_{ij} - \min(x_{ij}))/\xi_i$. After this process, the value of the new data falls in the interval $[0,1]$. In this way, the variables have a certain comparability and computability. With these new normalized data, the calculation cost is reduced.

4.2. Learning process of RBF network

4.2.1. APC-III clustering algorithm

Traditionally K-means clustering is used to determine the centers of RBF neural network, where the center is typically obtained by randomly selecting points from among all of the data points, thus the learning speed is relatively slow, and the result is unsatisfactory. In reference [15,16], a kind of one-pass clustering algorithm called APC-III clustering algorithm is proposed. It can make full use of information contained in training data, and by clustering the patterns class by class instead of the entire patterns at the same time, it obtains a higher learning speed. The clustering radius (the minimum distance between training patterns) of APC-III clustering algorithm can be obtained through formula (6), so

$$R_0 = \alpha \frac{1}{p} \sum_{i=1}^{p} \min_{i \neq j}(\|X_i - X_j\|) \tag{6}$$

where R_0 is the clustering radius, p is number of the training samples, α is a predetermined constant, R_0 expresses the radius of each cluster and consequently it controls the number of the clusters (the number of hidden units).

The idea of APC-III clustering algorithm is to cluster the data firstly, then to modify the number in hidden layer and clustering center of RBF neural network according to prescribed principle. Then the distance between each given training sample and the existing clusters is calculated. If the minimum value of all the distances is less than R_0, then we include the pattern in the cluster and recalculate the new center of the cluster. If the pattern is not included in any cluster then we create a new cluster with this sample.

Based on the analysis and study above, the outline of APC-III algorithm can be stated as follows.

Input: training samples $X = \{x_i \mid x_i \in R^p, i = 1, 2, ..., p\}$

Output: centers of clusters $c_i (i = 1, 2)$

Variables:

L: number of clusters

c_i : Center of the ith cluster

n_i :Number of samples in the ith cluster

d_{ji} :Distance between x_j and the ith cluster

Step1: Initialize the parameters /* for each sample */

set L=1， $c_1 = x_1$， n_1=1；

Step2: for (j=2; j≤p; j++) /* for each cluster */

Step3: for (i=1; i ≤L; i++)

{

compute d_{ji} ;

}

Step4: if $(d_{ji} \le R_0)$ /* include x_j into the ith cluster */

{

$$R_0 = \alpha \frac{1}{p}\sum_{i=1}^{p} \min_{i \ne j}(\|X_i - X_j\|);$$

$$c_i = (c_i n_i + x_j)/(n_j + 1);$$

$$n_j = n_j + 1;$$

}

Step5: if ($x_j \notin c_i$) /* create a new cluster */

{

L=L+1;

$c_L = x_j$;

n_L =1;

}

4.2.2. Determining the width parameter of Gaussian function

So far we have determined the center c_i of RBF neural network, the width parameter of its kernel function can be calculated by

$$\sigma_i = d_i \Big/ \sqrt{2m} \tag{7}$$

where d_i denotes the maximum distance between the ith center and other training samples, m denotes the number of centers (the number of hidden layer nodes).

4.2.3. Recursive least square (RLS) algorithm

Due to the fact that the mapping from hidden layer to output layer is linear, the weights determination become a linear optimization problem. Since the convergence speed of Gradient Descent algorithm is too low, the RLS algorithm is introduced in this paper. We regard the weights as status vector

$$w_j(k) = [w_{1j}(k), w_{2j}(k), \ldots, w_{nj}(k)]^T \; 1 \le j \le n \tag{8}$$

When it goes to the kth step, the output vector of the middle layer is

$$\phi(k) = [\phi_1(k), \ldots, \phi_{nl}(k)]^T = [\phi_1(l_1(k)), \sigma, \ldots, \phi_{nl}(l_{nl}(k)), \sigma]^T \tag{9}$$

The estimated output of the jth unit in the kth step is

$$\widehat{y_j}(k) = \sum w_{ij}\phi_l[l_i(k), \sigma] \tag{10}$$

Supposing the actual output obtained by RBF neural network is $y_i(k)$， the output error is calculated by

$$\varepsilon_j(k) = y_j(k) - \widehat{y_j}(k) \tag{11}$$

The weight of RBF neural network is updated as follows

$$w_j(k+1) = w_j(k) + \mu(k)\phi(k)\varepsilon_j(k)\frac{1}{\lambda(k)}\left[\mu(k) - \frac{\mu(k-1)\phi(k)\phi^T(k)\mu(k-1)}{\lambda(k)+\phi^T(k)\mu(k-1)\phi(k)}\right] \tag{12}$$

where μ is the error covariance matrix, λ $(0<\lambda<1)$ is a forgetting factor, it can weaken the influence of historical data on current value gradually, therefore the current value can reflect the characteristics of present sample data as much as possible. In this paper, we propose a method to adjust $\lambda(k)$ dynamically

$$\lambda(k) = 1 - \exp\left(-\frac{k}{\tau_0}\right) \tag{13}$$

where τ_0 is an initial smoothing factor, which is set empirically.

5. Experiment

In this paper we compare RBF neural network algorithm with the other two typical methods: support vector machine and outlier analysis. Our attention is mainly focused on the detection rate(DR) and the false positive rate(FPR), the detection rate is defined as the number of unusual instances detected by the system divided by the total number of unusual instances presented in the test set. The false positive rate is defined as the total number of normal instances that are incorrectly classified as unusual divided by the total number of normal instances.

5.1. Experiment data set

A real financial transaction record database set acquired from one commercial bank is adopted in the experiment; it comprises 6000 accounts, one million records transactions over 8 months. Table 1 is the main structure of the raw data sets. According to the characteristics of case data, and in order to acquire the primary features to build a customer behavior profile, we pick up three attributes processed according to the statistical theory, namely, frequency of withdrawals (d_1), frequency of deposits (d_2) and transaction sum (d_3). We use the preprocessing methods described in 4.1 to deal with the three attributes, then we obtained training data as listed in table 2.

Table 1.Raw data set

Client number	Business type	Transaction date	Transaction type	Times of deposits	Times of withdrawals	Transaction sum
187366	Manufacturing	20051206	Deposits	7	6	109060
157801	Service	20060305	Withdrawals	5	2	8610
199643	Private	20060418	Withdrawals	4	3	76431
…	…	…	…	…	…	…

Table 2. Attribute parameters (d_1, d_2, and d_3) for AML

Sample	Frequency of withdrawals(d_1)	Frequency of deposits(d_2)	Transaction sum(d_3)
S_1	0.04	0.21	0.65
S_2	0.18	0.15	0.71
…	…	…	…

Because the real money laundering data is hard to acquire, we can add some suspicious event to the normal transaction data to detect the effectiveness of RBF neural network. Let $\psi_1(\kappa)$ denote normal transaction data, then suspicious transaction data is

$$\psi_2(k) = \psi_1(k) + \gamma . e(k), k=1, 2, ..., 200 \quad (14)$$

where $e(k) = \begin{cases} \sin(\frac{\pi}{2}k) & k \in [80,90] \\ 0 & other \end{cases}$ is an unusual event, γ is a constant measuring the intensity of an unusual transaction.

5.2. Experiment analysis and comparison

Since SVM algorithm is suitable for the classifier designing and suspicious discovery among high dimensionality heterogeneous data sets. We use the SVM algorithm in the package provided by LIBSVM[17], the SVM algorithm is based on the HVDM distance, given the incorrectly classification punishment factor G is 100, the control parameter η is 1. And the outlier could be discriminated by the average and standard deviation. Given two datasets L ($l \in L$) and R ($r \in R$), then the average value is $\hat{n}(l,r,\alpha_1) = \frac{\sum_{q \in N_{L(l,r)}} n_R(q,\alpha_1 r)}{n_L(l,r)}$, $N_{L(l,r)} = \{q \in L | d(l,q) \le r\}$, $d(l,q)$ is the distance between point l and q, $n_L(l,r)$ is the number of r-neighbors of $l \in L$, the standard deviation is $\hat{\sigma}(l,r,\alpha_1) = \sqrt{\frac{\sum_{q \in N_{L(l,r)}} (n_R(q,\alpha_1 r) - \hat{n}(l,r,\alpha_1))^2}{n_L(l,r)}}$, α_1 is a parameter with its value $1/2^\delta$ (the factor δ is a positive number). If $\left|\hat{n}(l,r,\alpha_1) - n_R(l,\alpha_1 r)\right| > \varepsilon\hat{\sigma}(l,r,\alpha_1)$, then the point l is the outlier, ε is a constant.

According to the experiment analysis result of reference [16], we discovered that when parameter α which determines the clustering radius of APC-III algorithm is below 1.04, the accuracy of RBF neural network is better. So in this paper we choose α with its value 1.02, and we assume initial smoothing factor τ_0 is 0.1, The initial hidden layer number of RBF neural network is swing around the formula $\log_2^p$, p is the number of nodes in input layer.

We picked 200 instances from the raw data mixed with 70 suspicious transaction data to train the model. When the error is below 0.01, the learning is stopped. The error curve of RBF neural network is depicted as Fig. 4. As for the test set, we use another 50 of the raw data involved with another 40 deliberately simulated unusual records. Given these parameters $\alpha_1 = 1/8$, ε=3, γ=0.2. We examined our test data included 90 transaction data, out of which 27, 38 and 78 are detected in the three methods. The test result is shown in Fig. 5.

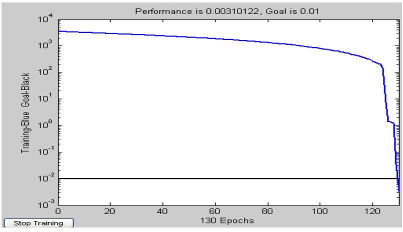

Figure 4 Error curve of RBF network

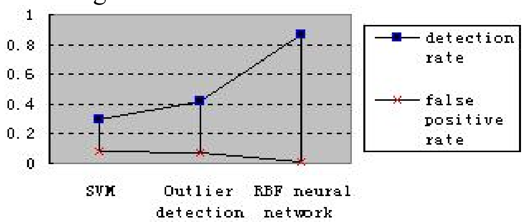

Figure 5 Detection rate and false positive rate comparison

6. Conclusions

The RBF neural network model for anti-money laundering presented in this paper can reach high correction rate. It shows promising results in reducing false positive rate and enhancing detection rate remarkably, providing a new method to detect the suspicious transaction. And it can adapt to the changing risk and means of money laundering. Experiment shows that the RBF neural network model is feasible and effective in money laundering detection. At the same time, there are many problems to be further studied, such as the actual detection system for anti-money laundering and the real money laundering transaction records for experiment analysis.

Acknowledgements

This work is supported by the Education Department Natural Science Research Project of Shaanxi Province, China under Grant No. 07JK339

References

[1] Yang Qifeng[1], Feng Bin[2], Song Ping[3]. Study on Anti-Money Laundering Service System of Online Payment based on Union-Bank mode. International Conference on Wireless Communications, Networking and Mobile Computing, Sept.2007. pp: 4991-4994.

[2] ZHU TIANQING. An Outlier Detection Model Based on Cross Datasets Comparison for Financial Surveillance. Proceedings of the 2006 IEEE Asia-Pacific Conference on Services Computing (APSCC'06), Dec. 2006, pp: 601-604.

[3] SU-NAN WANG[1,2],JIAN-GANG YANG[1].A MONEY LAUNDERING RISK EVALUATION METHOD BASED ON DECISION TREE.Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 19-22 August 2007.

[4] LiuXuan, Zhang Pengzhu.An agent based Anti-Money Laundering system architecture for financial supervision. International Conference on Wireless Communications, Networking and Mobile Computing, Sept.2007, pp: 5472-5475.

[5] Shijia Gao[*],Dongming Xu.Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering.Expert System with Applications, doi:10.1016/j.eswa. 2007.11.059

[6] Shijia Gao, Dongming Xu, Huaiqing Wang, Yingfeng Wang. Intelligent Anti-money Laundering System. International Conference on Service Operation and Loqistics, and Informatics, 2006, SOLI'06, IEEE. pp: 851-856

[7] Haiping Du[a,*], Nong Zhang[a]. Time series prediction using evolving radial basis function networks with new encoding scheme. Neurocomputing, Vol:71, March 2008,pp:1388-1400.

[8] Zheng Rong Yang. A Novel Radial Basis Function Neural Network for Discriminant Analysis. IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL.17, NO.3, MAY 2006.

[9] Oliver Buchtala, Manuel Klimek, Bernhard Sick. Evolutionary Optimization of Radial Basis Function Classifiers for Data Mining Applications. IEEE TRANSACTION ON SYSTEMS, MAN, AND CYBERNETICS-PARTB: CYBERNETICS, VOL.35, NO.5, OCTOBER 2005.

[10] Jun Li, Feng Zhao. Identification of Dynamical Systems Using Radial Basis Function Neural Networks with Hybrid Learning Algorithm.1st International Symposium on Systems and Control in Aerospace and Astronautics, ISSCAA 2006. pp:4.

[11] Wei Qiu, Member, IEEE, Chunqi Chang, Wenqing Liu, Paul W.F.Poon, Yong Hu, F.K.Lam, Roger P.Hamernik, Gang Wei, and Francis H.Y.Chan. Real-Time Data-Reusing Adaptive Learning of a Radial Basis Function Network for Tracking Evoked Potentials. IEEE TRANSACTION ON BIOMEDICAL ENGINEERING, VOL.53, NO.2, FEBRUARY 2006.

[12] A.S.Parimal, Nick Ivanou, Wieslaw L. Nowinski. Transformation Of Images By Radial Basis Functions With Varying Center Point. 2006 IEEE.ICARCV2006

[13] Nor Irwan Ahmat Nor[1]; Sobri Harun[2]; and Amir Hashim Mohd. Kassim[3]. Radial Basis Function

Modeling of Hourly Streamflow Hydrograph. JOURNAL OF HYDROLOGIC ENGINEERING, ASCE/JANUARY/ FEBRUARY 2007/113

[14] HI-BINCHENG*,CHING-LUNGGHEN,CHUNG-JEN FU.Financial Distress Prediction by a Radial Basis Function Network with Logit Analysis Learning. Computers and Mathematics with Applications 51 (2006): 579-588.

[15] Mercedes Feández-Redondo, Joaquín Torres-Sospedra, and Carlos Hernández-Espinosa. Training RBFs Networks: A Comparison Among Supervised and Not Supervised Algorithms. I. King et al. (Eds.); ICONIP, Lecture Notes in Computer Science, volume: 4232, pp.477-486, 2006.

[16] Ali Idri, Alain Abran, Samir Mbarki. An Experiment on the Design of Radial Basis Function Neural Networks for Software Cost Estimation. Information and Communication Technologies: From Theory to Applications, 2nd, 24-28 April 2006, Volume:1, pp: 1612-1617.

[17] http://www.csie.ntu.edu.tw/~cjlin/libsvm.