Identification of Electromagnetic Radiation Source with Support Vector Machines

Dan Shi[1] Junjian Bi[2], Chao Li[3], Zhiliang Tan[2], Hongbo Wang[4] and Yougang Gao[5]
(1)beijing university of posts and telecommunications, Beijing, China
(2)Shijiazhuang Mechanical Engineering, Shijiazhuang, China
(3)ministry of industry and information technology, Beijing, China
(4)ministry of industry and information technology, beijing, China
(5)beijing university of posts and telecommunications, beijing, China

***Abstract*—A method for electromagnetic radiation source identification is proposed. The spatial characteristic of a radiation source is taken as the unique parameter for support vector machines (SVMs) to identify. First, the location of radiation source is determined by the triangulation method, and then its spatial characteristic is collected by a band receiver array with simulation, which removes the limit of absolute similarity between test data and training data. The 3D data are converted into a 1D vector with subscripts as inputs for SVMs, which are trained by the inputs to identify radiation source types intelligently. The identification time needs a few seconds, much faster than artificial neural networks (ANNs). The influence of parameters (e.g., noise from ambient environment, data collection method, scaling method for inputs, and parameters of SVMs) is discussed. The proposed method has good performance in noisy environment and the identification accuracy is 76.57 %, even though the signal to noise ratio decreases to 10 dB.**

Keywords—support vector machines; electromagnetic radiation source identification; spatial characteristics; band receiver array

I. INTRODUCTION

Electromagnetic radiation source identification (ERSI) has been a hot topic due to its applications in electromagnetic interference diagnosis (EMID) and radio monitoring, as well as radar sensors [1-5]. Nowadays EMID is mainly confronting the challenge of identification and localization of different interference sources. [1] proposed an ERSI method based on independent component analysis. The method requires that the sources are independent signals and is inapplicable when frequencies overlap. [2-3] investigated the possibility of detecting and identifying electronic devices based on their electromagnetic emissions, in which different frequencies and a high signal to noise ratio (SNR) are required. In radar detection, published literature focuses on emitter identification. Conventional identification approaches based on basic pulse parameters are sensitive to noise and not applicable when EM source frequencies overlap [4]. In radio monitoring, the distinction between radiation sources mainly depends on frequency separation using a spectrum analyzer [5].

Our group has been working on ERSI when the frequencies of sources overlap in a noisy environment. First, we proposed to identify the EM sources by artificial neural networks (ANNs) with 27 receivers in a 3D cube, which has high accuracy without additive noise. However, the accuracy is significantly affected by noise [6]. Afterwards an improved method was brought forward with support vector machines (SVMs), which can withstand strong noise [7]. The above two methods require that test data are collected in the same way as training data are collected, since they both use 3D cube receivers. In this paper, the receiver array is improved by using band shape instead of the 3D cube, which is more consistent with pattern recognition theory. It relaxes the requirement of absolute similarity between test data and training data.

In Section II, the SVMs for classification are introduced. Section III describes data collection and the model. Section IV discusses the influence of various parameters on identification accuracy. Finally, conclusions are made in Section V.

II. SVMS FOR CLASSIFICATION

The data used for SVMs classification is often a pair comprising an input object (typically a vector, called the attribute or feature) and a desired output value (called the label). A supervised learning algorithm analyzes the training data and produces inferred functions, which should predict the correct output value for any valid input object. In classification, however, it often happens that the data sets to discriminate are not linearly separable in a finite dimensional space. Thus, it is proposed that the original finite dimensional space is mapped into a much higher dimensional space. To keep the computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function [8-9]. There are four types of kernel function and radial basis function (RBF) is chosen in our model:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\gamma \| \boldsymbol{x}_i - \boldsymbol{x}_j \|^2), \quad \gamma > 0 . \tag{1}$$

The hyper-planes are constructed in a high dimensional space separating the samples with maximal margin. A good separation is achieved by the hyper-plane that has the largest distance to the nearest training data point of any class, since in general the larger the margin the lower the generalization error. If there exists no hyper-plane that can split the "yes" and

"no" examples, the soft margin method will choose a hyper-plane that splits the examples as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples [10]. The method introduces non-negative slack variables ξ_i to mark a degree of classification error. Given a training set of attribute-label pairs $(\boldsymbol{x}_i, \boldsymbol{y}_i)$, $i = 1,\mathrm{L}\ ,m$, where $\boldsymbol{x}_i \in \mathrm{R}n$ and $\boldsymbol{y}_i \in \{-1,1\}^m$, the SVMs require the solution of the following optimization problem:

$$\min_{\boldsymbol{w},b,\xi} \left\{ \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{m}\xi_i \right\}$$
$$\text{subject to } \begin{matrix} \boldsymbol{y}_i(\boldsymbol{w}^T\phi(\boldsymbol{x}_i)+b) \geqslant 1-\xi_i \\ \xi_i \geqslant 0 \end{matrix}. \qquad (2)$$

Here, training vectors $\boldsymbol{x}_i$ are mapped into a higher dimensional space by the function ϕ. $\boldsymbol{w}$ is the normal vector to the hyper-planes. We should minimize $\|\boldsymbol{w}\|$ to find the maximum distance between these hyper-planes. SVMs find a separating hyper-plane with the maximal margin in this higher dimensional space. C>0 is the penalty parameter of the error term, and the compromise between maximal margin and classification error can be achieved by adjusting C. (2) tries to increase the margin and decrease the error introduced by ξ_i .

Furthermore, the RBF has the relationship with function ϕ, that is, $K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \phi(\boldsymbol{x}_i)^T\phi(\boldsymbol{x}_j)$.

The steps using SVMs for classification are listed as follows:

1) Collect data sets comprising attributes and labels, and randomly divide them into training sets and test sets.

2) Scale the attributes.

3) Choose the RBF kernel function.

4) Perform cross-validation to find optimal parameters C and γ .

5) Utilize the optimal parameters C and γ to train the data sets and then find appropriate inferred function.

6) Test and predict the accuracy with test data sets, and check the validity of the method.

III. MODEL AND CALCULATION

Intentional electromagnetic radiation is often generated from the radiated antenna. For instance, the radiation from a mobile phone usually comes from the phone antenna, such as planar inverse-F antenna (PIFA), the radiation from a base station is often generated by a planar antenna, and radars transmit electromagnetic waves through phased array antennas, etc. On the other hand, unintentional radiation is often caused by equivalent antennas. For example, the bare wire in an electronic device causing electromagnetic interference is usually modeled with a dipole antenna. The radiation caused by the printed circuit board trace is often modeled with a loop antenna. In a word, the devices causing radiation interference can be modeled with their primary radiated antennas. Obviously, the spatial characteristics of different antennas are different, which are unique features used for identification of the antennas, thereby achieving ERSI.

In [6] and [7], the spatial characteristics of radiation sources are collected by a 3D cube. The dimensions of this cube are 50 mm x 50 mm x 50 mm. The 27 receivers distributed in the cube are located at the center of edges, faces and volume. Two neighboring receivers are set 25 mm away from each other. The cube moves with a 20 mm steps in the x, y and z directions in a cone around the main lobe to collect the data sets. In those models, the way collecting the data sets for recognition should be similar with the way collecting data sets for training. That is, the 27 receivers collect data sets in enlarged or shrunken cubes in the narrow cone around the main lobe of radiation sources. In other words, the high identification accuracy is achieved under these conditions: ① The test data sets are collected at the same plane as the training data sets, shown in Fig. 1.

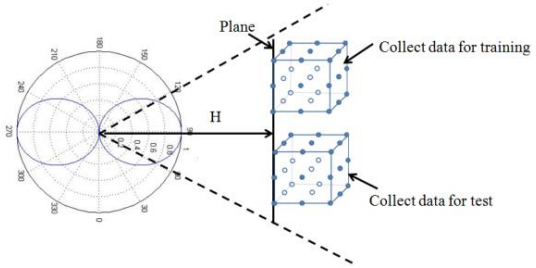

Fig.1 Data collection for dipole antenna

② For an alternative method, if the test data sets are collected at different plane from the training data sets, the distance between neighboring receivers should vary with the distance from the radiation source to receiver cube, and the formula is presented in [6].

$$L = H/50\sqrt{6} \ . \qquad (3)$$

Where L is the distance between neighboring receivers, H is the distance from the radiation source to receiver cube.

We have tested the ability of the method proposed in [7] to withstand deviation between test data sets and training data sets. It is found that the identification accuracy decreases significantly with the deviation, as shown in Table I.

TABLE I
IDENTIFICATION ACCURACY VS. DEVIATION

Training data location / Test data location	5 m	10 m
1.5 m	13.59%	19.03%
3 m	50.46%	29.37%
5 m	89.86%	46.2%
10 m	63.7%	91.49%

To solve the problem, a new receiver array is presented. In practice, before collecting the data, the position of radiation source should be located. The triangulation method can be applied. As Fig.2 shows, a directional antenna with a narrow range beamwidth or antenna array is located at TP1. The direction of the arrival of radiation source is determined,

which is in the line of r1. Similarly, another directional antenna or antenna array is located at TP2, which determines that the radiation source is in the line of r2. The intersection point between r1 and r2 is the location of radiation source, which is the red point in Fig.2. In the sphere around the radiation source, there are several measurement points, and some of them are marked from MP1 to MP9.

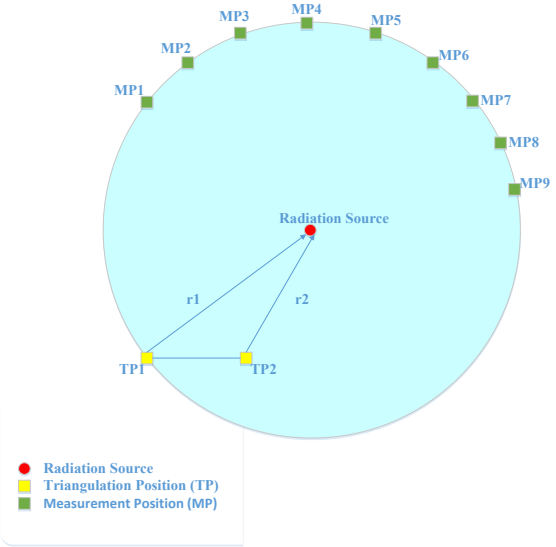

Fig. 2 Triangulation for determining position of radiation source

After location of the radiation source is determined, the receiver array is moved around the equator of the radiation source to collect the electric field distribution of the source, and then the data are sent to SVMs for training and recognition. The steps of ERSI are shown in Fig. 3.

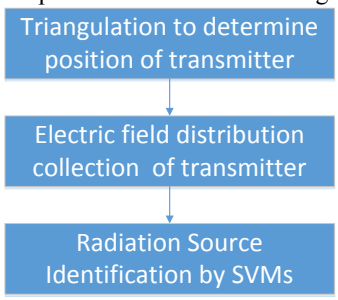

Fig.3 Steps of ERSI

The receiver array is composed by 27 elements and they are located along a band, which is 30^0 in latitude and 120^0 in longitude. All the receivers are arranged in 3 rows and 9 columns. The receivers at 9 columns are shown as MP1 to MP9 in Fig.1. The neighboring receivers are 15^0 intervals in latitude and 15^0 intervals in longitude. Then the receiver band moves along the equator of radiation sources with 0.1^0 steps. After the band moves 360^0 along the equator, 3600 data sets are collected, and every data set includes 27 values, which can be expressed as a 1D vector with 27 elements:

$$\boldsymbol{E}_{\mathrm{kp}} = [\boldsymbol{E}_{11}^{\mathrm{kp}}, \boldsymbol{E}_{12}^{\mathrm{kp}}, \boldsymbol{E}_{13}^{\mathrm{kp}}, \boldsymbol{E}_{14}^{\mathrm{kp}}, \boldsymbol{E}_{\mathrm{mn}}^{\mathrm{kp}} \mathrm{L}\ , \boldsymbol{E}_{39}^{\mathrm{kp}}]. \quad (4)$$

Here k is the label of radiation source type, p is the position label of the band receiver array, m is the row label of receiver in the band, and n is the column label of receiver in the band. Three radiation sources, bare wires, mobile phones, and RFID systems, are modelled with their primary radiators, that is, a dipole antenna, a PIFA, and a microstrip antenna. All antennas are working at 3GHz to ensure that the SVMs identify them not based on the frequency. The electric field distributions around these sources, called "spatial characteristics", are collected by simulation with HFSS. The band receiver arrays vs. dipole antenna pattern is displayed in Fig. 4.

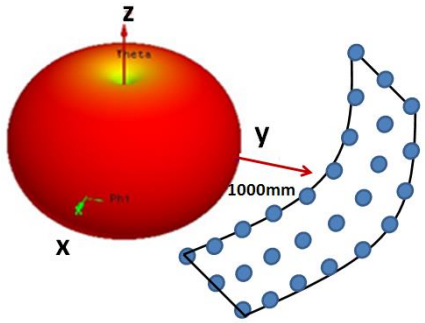

Fig.4 Receivers vs. pattern of the dipole antenna

Obviously, the spatial characteristics of these three antennas are different. They are collected by band receiver array, which is a circular and part of a big ring around the radiation source, and are used as unique features for SVMs to identify. The spatial information is converted from 3D array to 1D vector with subscripts as inputs to simplify the model. Since there are 3600 data sets, that is, 3600 1D vectors for each antenna, 10800 data sets for these three antennas are obtained. These data sets are expressed as a combined vector set $\boldsymbol{E}_{total} = [\boldsymbol{E}_{11}, \boldsymbol{E}_{12}, \mathrm{L}\ , \boldsymbol{E}_{33600}]'$. The first number of subscript in the vector set represents the type of radiation source, and the other number of subscript in the vector set represents the position of band receiver array. The vector set $\boldsymbol{E}_{total}$ is scaled to the range [0, 1] with linear scaling and marked with the source type. For instance, the data sets belonging to the dipole antenna are marked with "1" at the end of the data sets, the data sets belonging to PIFA antenna are marked with "2", and the third one is marked with "3". This type mark is for parameter modification in training process and result validation in test process, respectively. The 10800 vectors are randomly divided into two categories for training and test. In the training process, RBF extract the relationship between spatial characteristics and source types, and constructs the hyper-plane classifying the sources based on the vectors. Since the distinction among spatial characteristics of radiation sources is apparent, the hyper-plane is easy to construct. The RBF kernel function has two important parameters, called the "kernel parameter γ and penalty parameter C", that need to be determined. A genetic algorithm is applied to find the optimal parameters C and γ by iteration searching, which are 2 and 1 in this model respectively. After being trained with 9000 data sets within a few minutes, the SVMs can identify the source

types by their spatial characteristics rapidly and accurately. The 1800 data sets left are applied for the test. After the vectors are input, the SVMs can identify the source types which these vectors belong to, and the identification accuracy is up to 100%. The $F1$ measure is used for checking the validity of our model. The $F1$ combining recall (r) and precision (p) is:

$$F1 = 2rp/(r+p). \quad (5)$$

where p is the number of correct results divided by the number of all returned results, and r is the number of correct results divided by the number of results that should have been returned. The $F1$ in this model is 1, which reflects the validity of our method.

It should be noted that the identification accuracy is closely related with the diversity between radiation sources. Since the spatial characteristics of radiation sources in this paper differ greatly, the identification accuracy is up to 100%. However, the accuracy would fall down when the pattern of radiation sources are similar.

IV. Discussion

A. The Influence of Noise

Gaussian noise is added in our model. The parameter of SNR is applied, which is expressed as follows:

$$\text{SNR (dB)} = 10 \times \log_{10}\left(P_{\text{avg}}/P_{\text{noise}}\right). \quad (6)$$

where P_{avg} is the average power of input data and P_{noise} is the power of Gaussian noise. The identification accuracy versus SNR is listed in Table II. It also gives the comparison between SVMs and ANNs. In ANNs, the back-propagation algorithm is applied and the iterations are stopped when the mean square error reaches 0.01. Two hidden layers and 30 neurons are constructed. Table II shows that if no noise is added, the identification accuracy with both SVMs and ANNs is very high. When the SNR decreases from 20 dB to 15 dB, the accuracy varies from 99.94% to 97.12%. Even in a strong noise environment, where the SNR is only 10 dB, the accuracy is still 76.57%. Thus, the method using SVMs can be applied in noisy environments. However, the accuracy with ANNs significantly decreases as the noise increases. When the SNR is 15 dB, the accuracy rate with ANNs degrades to 64.1%, and it doesn't work when the SNR is 10 dB.

TABLE II
IDENTIFICATION ACCURACY VS. SNR

SNR/dB	No noise	25	20	15	10
Accuracy of SVMs/%	100	100	99.94	97.12	76.57
Accuracy of ANN/%	99.98	98.43	82.3	64.1	38.87

B. The Influence from Data Collection

The method of data collection is important to identification accuracy. In [7], the 3D cube moves along the cone with 60^0 angle to collect data sets, which mainly represent the spatial characteristics of the main lobe of the antennas. In contrast, the data sets here are collected with a band receiver array, and the band moves around the equator of antennas in this paper. According to the pattern recognition theory, it is preferable to identify the radiation sources with the global information other than local information, which has been proven by the result comparison. Some results also support this point. First, we use a band receiver array with 9 receivers to collect data sets. The neighboring receivers are 1^0 intervals in longitude. The accuracy versus SNR is shown with the blue line in Fig.5, and the best accuracy is only 71.2%. Then the receiver number is increased to 30, and the neighboring receivers are also 1^0 intervals in longitude. The accuracy versus SNR is shown with the red line in Fig.5. Unfortunately, the identification accuracy is still not satisfactory. Finally, we increase the intervals between neighboring receivers to 15^0 in longitude and 27 receivers are applied, and the accuracy versus SNR is shown with the green line in Fig.5. It can be found that good results are obtained.

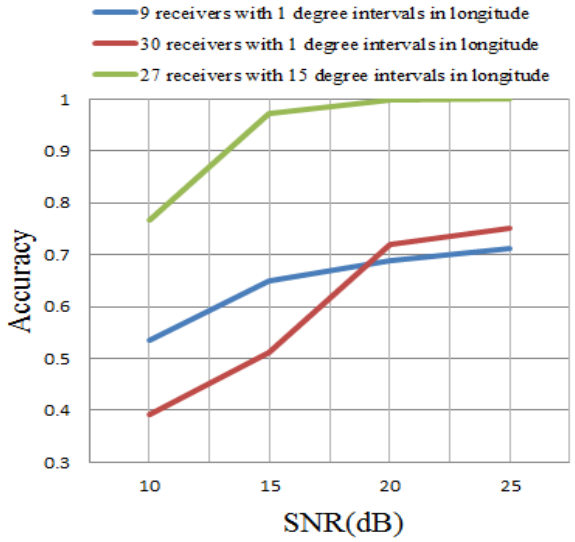

Fig.5 Accuracy of three data collection methods

C. The Influence of the Scaling Method

Scaling methods such as tangent scaling, logarithm scaling and linear scaling are investigated. As mentioned in Section III, a total of 10800 input vectors are expressed as a combined vector set $\boldsymbol{E}_{total} = [\boldsymbol{E}_{11}, \boldsymbol{E}_{12}, \mathrm{L}\ , \boldsymbol{E}_{33600}]'$. The elements in $\boldsymbol{E}_{total}$ are scaled by these three methods.

The following formula is applied for arc tangent scaling:

$$\boldsymbol{E}_s = a\tan(\boldsymbol{E}_{total}) \times 2/\pi. \quad (7)$$

where $\boldsymbol{E}_s$ is the input vector sets after scaling.

(8) is used for logarithm scaling:

$$\boldsymbol{E}_s = \log_{10}(\boldsymbol{E}_{total}). \quad (8)$$

(9) is applied for linear scaling:

$$\boldsymbol{y}_i = \left(\boldsymbol{x}_i - x_{\min}\right) / \left(x_{\max} - x_{\min}\right). \quad (9)$$

where $\boldsymbol{x}_i$, $x_{\max}$, and $x_{\min}$ are the element, maximum, and minimum elements of $\boldsymbol{E}_{total}$; $\boldsymbol{y}_i$ is the element of $\boldsymbol{E}_s$ after scaling.

The accuracy versus scaling methods is shown in Table III.

TABLE III
ACCURACY VS. SCALING METHOD

SNR/dB	25	20	15	10
Accuracy with arc tangent scaling/%	95.09	93.66	84.16	67.21
Accuracy with logarithm scaling/%	100	99.47	92.4	68.67
Accuracy with linear scaling/%	100	99.94	97.12	76.57

Obviously, the linear scaling method has the best performance among these three scaling methods.

D. The Influence of the Parameters of the SVMs

Table IV presents the identification accuracy versus C and γ . The optimized C and γ are 2 and 1. Obviously, the most accurate result is obtained with the optimized parameters. The accuracy decreases as C and γ deviate from the optimized values.

TABLE IV
ACCURACY VS. PARAMETERS

Parameters	C=0.1, γ=0.1	C=0.2, γ=0.2	C=2, γ=1	C=20, γ=20	C=100, γ=100
Accuracy/%	79.61	94.66	100	89.58	35.62

V. CONCLUSION

A new method for ERSI by using SVMs is proposed. The data collection method has been improved. After the position of radiation source is determined by the triangulation method, a band receiver array is applied for data collection. The band receiver array moves along the equator of the radiation sources rather than moving along the cone around the main lobe of radiation sources. It focuses on collecting data with global information other than local information. The results demonstrate that the method proposed in this paper has better identification accuracy compared with ANNs and a 3D cube receiver array. It can be applied in strong noise environment, and deviation between training data sets and test data sets is allowed.

As practical EMC issues are concerned, many electronic devices produce EMI problems intentionally or unintentionally. In EMI analysis, however, these interference sources have often been modeled with different radiation antennas which have unique spatial characteristics. Thus the method proposed in the paper can be applied to identify the radiation source type based on its unique spatial characteristics. The steps are as follows: ①Collect the spatial characteristics of typical EMI radiation sources. ② Put the spatial characteristics of these radiation sources into SVMs for training and establish the model. ③ In practice, the E-field at different positions caused by an unknown EMI source are measured. ④ Put these E-field data into the SVMs for test and the SVMs can identify the EMI source type intelligently. The identification accuracy depends on the similarity between training data and test data, which is mainly affected by the ambient noise.

REFERENCES

[1] Z. F. Song and D. L. Su, "A Novel Electromagnetic Radiated Emission Source Identification Methodology," Proceeding of 2010 Asia-Pacific International Symposium on Elec*tromagnetic Compatibility*, pp. 645-648, 2010.

[2] Haixiao Weng, Xiaopeng Dong, Xiao Hu, Daryl G. Beetner, Todd Hubing, and Donald Wunsch, "Neural Network Detection and Identification of Electronic Devices Based on Their Unintended Emissions, " *International Symposium on EMC*, Vol. 1, pp. 245-249, 2005.

[3] Xiaopeng Dong, Haixiao Weng, and Daryl G., " Detection and Identification of Vehicles Based on Their Unintended Electromagnetic Emissions," *IEEE Transaction on Electromagnetic Compatibility*, Vol. 48, NO. 4, pp. 752-759, 2006.

[4] C.-S. Shieh and C.-T. Lin, "A vector neural network for emitter identification," *IEEE Trans. Antennas Propag.*, vol. 50, no. 8, pp. 1120–1127, Aug. 2002.

[5] F. Mavromatis, A. Boursianis, T. Samaras, C. Koukourlis1, and J. N. Sahalos. "A Broadband Monitoring System for Electromagnetic-Radiation Assessment", *IEEE Antennas and Propagation Magazine*, vol.51, Issue 1, pp.71-79, 2009.

[6] D. Shi, and Y. G. Gao, "A New Method for Identifying Electromagnetic Radiation Sources Using Backpropagation Neural Network", *IEEE Transactions on Electromagnetic Compatibility*, vol. 55, Issue 5, pp. 842-848, 2013.

[7] D. Shi, and Y. G. Gao, "A method of identifying electromagnetic radiation sources by using support vector machines", *China Communications*, vol. 10, Issue 7, pp. 36-43, 2013.

[8] Cristianini Nello and Shawe-Taylor John, *An Introduction to Support Vector Machines and other kernel-based learning methods*, Cambridge University Press, 2000.

[9] Schölkopf Bernhard， Burges Christopher J. C.， and Smola Alexander J.， *Advances in Kernel Methods: Support Vector Learning*, MIT Press, Cambridge, MA, 1999.

[10] CORTES C, VAPNIK V. Support-Vector Networks. *Machine Learning*, 1995, vol. 20, No.3, pp. 273-297.