Randomized and Dimension Reduced Radial Basis Features for Support Vector Machine*

Akinori Hidaka† and Takio Kurita‡

Generally, the dimension of the Kernel matrices of the kernel Support Vector Machines (SVM) increases as the number of training samples increases. However high dimensional features often bring redundant computation and decline of the generalization ability. Also kernel functions have several hyper-parameters which are fixed to the same values for all training samples. By considering the kernel matrix of Radial Basis Function (RBF) as a new high dimensional nonlinear feature for SVM, there is no limitation of which those hyper-parameters have to be a single fixed number. In this paper, we develop a nonlinear feature extraction method based on the selection of kernel seeds and the fine tuning of the kernel parameters, called randomized and dimension Reduced Radial Basis Features (RRBF).

1. Introduction

Support Vector Machine (SVM)[1,7–9] is one of most popular and powerful classifiers, and it has been applied to many pattern recognition and machine learning problems. Although deep neural networks[10] has recently appeared as a strong alternative, DNN has a bottleneck that it needs high computational cost to obtain a fine feature representation specialized for a target problem. Therefore SVM still has been an important method for many machine learning problems of which their feature representation is identified or fixed.

Since SVM is a linear classifier, the Kernel approach is used to handle nonlinear classification problems. Especially, Kernel SVM using polynomial Kernel function or Radial Basis Function (RBF) has been one of the most powerful nonlinear classifiers in the field of machine learning. However, generally training of Kernel SVM requires high computational cost. Therefore, many previous works tried to relieve these difficulties. For example, Lin[4] proposed to select a random subset of the training set, though it could not reduce the number of basis functions to attain the accuracy close to full SVM solution. Nishida[6] proposed RANSAC-SVM method, which trains a number of small SVMs for randomly selected subsets of training set. Their hyper-parameters are tuned by applying the trained SVMs to whole training set. In this wise, it is a rational plan to reduce computational cost to train Kernel SVM by using only a part (subset) of training samples.

In real applications, the high computational cost of the Kernel functions often becomes a limiting factor. Therefore, it is important that we reduce unnecessary dimensions in Kernel matrices while keeping the classification performance. Gram matrices of Kernel functions constructed from N training samples x_i have N×N components. They can be considered as N dimensional nonlinear feature representation for the original feature x which is d-dimensions. Since N is usually larger than d in normal problem setting of machine learning, Kernel matrices will often include redundant information for classification more than x. Kurita proposed feature extraction from empirical Kernel vector[3], which is closely related with the Kernel principal component analysis, to reduce the redundant dimensions of kernel representation. The method can extract a dimension reduced feature vector in which the dominant information of the given empirical Kernel vector is included, and be equivalent to the original Kernel SVM if the full dimensions are used.

In General, Kernel functions have several hyper-parameters p = (p_1,p_2,...) which have to be determined based on appropriate estimation methods, such as Grid Search (GS) with Cross Validation (CV). For example, the polynomial Kernel function k(x,x') = (x·x' + h)^p has the threshold h and the order p, and
the radial basis function (RBF) \( \exp(-||x-x'||^2/\sigma) \) has the coefficient \( \sigma \). Usually, these parameters are fixed to same value for every training sample.

In this paper, we investigate what kind of influence occurs on RBF Kernel when we assigned different \( \sigma \) for each training sample independently. We call this variant of RBF Kernel “Randomized Radial Basis Features” (RRBF). The preliminary experiments of the proposed algorithm were reported in our previous conference paper[5]. We show that RRBF can fine-tune the decision boundary of usual RBF SVM, by experiments using artificial and real data.

Also, in this paper, we proposed the randomized selection of the training samples to eliminate the redundant dimensions from our RRBF. Our dimension reduced RRBF could bring us more efficient feature representation while maintaining the classification performance.

The rest of this paper is organized as follows: Section 2. reviews linear and Kernel SVM, and also looks back on a nature of RBF kernel. Then, RRBF are introduced and their property is investigated in Section 3.2. The classification experiments using benchmark data sets are described in Section 4. Finally, Section 5. concludes the paper.

2. Support Vector Machine

Support Vector Machine (SVM) is one of most powerful linear classifiers which determines the separation, which outputs +1 when \( y > 0 \) and outputs -1 when \( y < 0 \).

The soft- margin SVM is defined as an optimization problem for the following evaluation function

\[
L_D(\alpha) = \frac{1}{2}||\alpha||^2 + c \sum_{i=1}^{N} \xi_i
\]

under the constraints

\[
\xi_i \geq 0,
\]

\[
t_i(\alpha^T \phi(x) - h_i) \geq 1 - \xi_i,
\]

where \( \xi_i \) is the measure of the error for the training sample \( x_i \).

The dual problem is obtained as the optimization problem that maximizes the objective function

\[
y(x) = \text{sign} (\sum_{i=1}^{N} \alpha_i t_i x_i^T x - h^*)
\]

where \( S \) is a set of support vectors and \( \alpha_i^* \) and \( h^* \) are the optimal solutions of Lagrange multiplier. Since the function \( \sum_{i \in S} \alpha_i^* t_i x_i^T x - h^* \) is linear for \( x \), this is called linear SVM.

2.2 Kernel Support Vector Machine

By using the Kernel-trick, linear SVM can be extended to nonlinear. In the Kernel SVM, input vectors are mapped to higher dimensional feature space by non-linear function \( \phi(x) \) and the linear SVM is applied to the mapped features. Since the linear SVM depends only on the inner products \( x_i^T x \) about the input \( x \), if we replace \( x_i^T x \) with \( \phi(x_i)^T \phi(x) \), we can define the object function as

\[
L_D(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j t_i t_j K(x_i, x_j)
\]

where \( K(x_i, x_j) = \phi(x_i)^T \phi(x_j) \) is the Kernel function.

Usually the Kernel function \( K(x, y) \) is defined a priori. The polynomial function,

\[
K(x, y) = (x^T y + h)^p
\]

and the Radial Basis Function (RBF),

\[
K(x, y) = \exp \left( -\frac{||x-y||^2}{\sigma} \right)
\]
are often used as the Kernel function.

Then the optimal classification function of the Kernel SVM can be derived as

$$ y(x) = \text{sign} \left( \sum_{i \in S} \alpha_i^* t_i \phi(x_i)^T \phi(x) - h^* \right) $$

$$ = \text{sign} \left( \sum_{i \in S} \alpha_i^* t_i K(x_i, x) - h^* \right) $$

In this paper, RBF is focused on as the Kernel function of SVM. The hyper-parameter of RBF Kernel, $\sigma > 0$ in Eq. (11), have to be determined based on appropriate estimation methods, such as Cross Validation (CV).

3. Tuning of RBF Kernel

In this section, at first we look back on the property of RBF Kernel and RBF SVM. Then, we describe how to tune the parameter $\sigma$ of RBF Kernel to improve the classification accuracy.

3.1 Nature of RBF Kernel

RBF Kernel $K(x, y)$ (11) is calculated based on the Euclid distance between $x$ and $y$. This function takes the value between 0 to 1, and will become small if the distance of $x$ and $y$ becomes large. Thus, we can interpret this function as one of a similarity between $x$ and $y$.

The hyper-parameter $\sigma > 0$ controls the speed of which this similarity decreases to 0; Fig. 1 shows the responses of 1-dimensional RBF depending on several different $\sigma$. As can be seen from the figure, a large $\sigma$ brings a longer-tail response. Meanwhile, the Kernel value will decrease more quickly if $\sigma$ is small. This means that RBF Kernel with small $\sigma$ will show “peaky” responses in which each sample only reacts to sufficiently nearby samples.

Fig. 2 shows the difference of influence for RBF SVM between large $\sigma$ and small $\sigma$. For this simulation experiment, we generate 2-dimensional artificial data of 2 classes problem of which each class has 100 training samples and 50 test samples. The top row shows that, since $\sigma = 2^{10}$ ($\sigma = 2^{-10}$). Circle and square markers represent samples of class 1 and 2, respectively. Filled markers mean those samples are misclassified by each SVM.
Let us consider one of the most extreme cases. If \( \sigma \) is infinitesimal, the \( i \)-th sample \( x_i \) only responds to itself. In this case, the \( i \)-th dimension of the Kernel matrix will become active only for \( i \)-th sample itself. It implies that this sample can be perfectly classified even if the sample is located in the center of the opposite class in the \( N-1 \) dimensional subspace excepting the \( i \)-th dimension.

In this wise, choosing small \( \sigma \) for RBF means to build nonlinear classifiers which pay its attention only to local relationships between the samples, and may cause over-fitting if too small \( \sigma \) is chosen. In contrast, larger \( \sigma \) will bring classifiers which focus on more global information, however it may bring slightly simple decision boundary similar to linear classification. The appropriate values of \( \sigma \) which gives good generalization performance for nonlinear problems will usually be intermediate values, and we have to find such moderate values by appropriate estimation methods, such as GS with CV.

### 3.2 Randomized Radial Basis Feature

Given an input \( x \) and \( N \) training samples \( x_i \), we can compute the \( N \) dimensional vector

\[
k(x) = (K(x_1,x), \ldots, K(x_N,x))^T,
\]

(14)
called the Empirical Kernel Vector (EKV). Kurita et al. showed that this vector includes enough information to construct the classification function of the Kernel SVM[3]. They derived the dimension reducing mapping for EKV defined by using the eigen values/vectors of RBF Kernel, and they used the mapped kernel as a new nonlinear feature for linear SVM. In this wise, this paper also investigate the feature extraction method from RBF Kernel matrix.

In RBF Kernel, usually the same \( \sigma \) is used for every training sample \( \{x_i\} \). However, using the same \( \sigma \) for all the samples \( \{x_i\} \) will mean to disregard the difference on importance of each sample. It is expected that the performance of RBF Kernel can be improved if the optimal parameter \( \sigma_i \) can be individually determined for every sample \( x_i \). Fig. 3 illustrates this idea visually.

According to this motivation, let us consider the variance radius RBF[5],

\[
K_\sigma(x_i,x) = \exp\left(-\frac{|x_i-x|^2}{\sigma_i}\right),
\]

(15)
of which the fixed number \( \sigma \) becomes the variable \( \sigma_i \) for each training sample \( x_i \), and the corresponding empirical Kernel vector,

\[
k_\sigma(x) = (K_\sigma(x_1,x), \ldots, K_\sigma(x_N,x))^T.
\]

(16)

Then, there is the problem that how \( \sigma_i \) should be decided. As shown by the simulation experiment in Section 4.1, arbitrary assignment of \( \sigma_i \) for important samples will cause over-fitting directly. However, the optimization of \( N \) (i.e. hundreds or thousands of) parameters by using CV is clearly computationally difficult. In this research, we also adopt the random approach[5]; We randomly assign the Kernel parameters \( \sigma_i > 0 \) for each training sample \( x_i \).

However, it is difficult to determine a good set of \( \{\sigma_i\} \) by completely at random. Meanwhile, the value of \( \sigma^* \), which implies the \( \sigma \) determined by GS and CV for standard RBF SVM, will become a good clue for the random selection. Let \( \mathcal{N}(\mu,\sigma) \) be normal random number with mean \( \mu \) and standard deviation \( \sigma \). In this paper, we generate randomized \( \sigma_i \) by

\[
\sigma_i = 2^p, \quad p = \mathcal{N}(\log_2 \sigma^*, \hat{\sigma})
\]

(17)

where \( \hat{\sigma} \) is a standard deviation which should be determined by manually.

The empirical Kernel vector (16) where \( \sigma_i \) are given by Eq. (17) is called randomized radial basis feature (RRBF). We use RRBF as a new nonlinear feature for linear SVM. We call it RRBF SVM.

### 3.3 Kernel Seeds Selection

In normal cases of pattern recognition problems, RBF Kernel matrix (and also RRBF vector) is usually high dimension than original feature \( x \), and then it will include many redundant or unnecessary information for classification. However, when one consider the Kernel matrix is one of nonlinear feature expressions, there is no limitation that all training samples have to be used to construct Kernel matrices.

To reduce the redundant information in RRBF vector, we choose \( M \) samples at random from all \( N \) training samples as kernel seeds \( s_i, (i = 1, \ldots, M \leq N) \), and assign the Kernel parameters \( \sigma_i > 0 \) for each seed \( s_i \) at random. Thus we obtain dimension reduced RRBF vector,

\[
K_s(s_i,x) = \exp\left(-\frac{|s_i-x|^2}{\sigma_i}\right),
\]

(18)

and corresponding empirical kernel vector,

\[
k_s(x) = (K_s(s_1,x), \ldots, K_s(s_M,x))^T
\]

(19)
of which the dimension was reduced to \( M \) from \( N \). Let the set of Kernel seeds and their parameters be

---

Fig. 3  Fixed hyper-parameters (left) and variable hyper-parameters (right)
To construct strong Kernel, we repeatedly generate such the set $S$ at random, and adopt the best one as the final Kernel. Table 1 shows the training procedure of RRBF SVM.

Now we can consider three types of RRBF: The first one has randomized $\sigma_i$ but uses all samples as Kernel seeds[5]. It is represented as “$\sigma$-RRBF”. The second one uses randomly selected Kernel seeds but they have fixed $\sigma$. It is represented as “D-RRBF” which implies dimensions are randomly reduced. The last one uses the random Kernel seeds with random $\sigma_i$[5]. It is represented as “Full RRBF”. Note that in the algorithm Table 1, these types of RRBF can be distinguished by how to make the random Kernel seeds $S_i$ in the item (1). If $M$ is set to $N$, it implies $\sigma$-RRBF. On the other hand, if $M < N$ but $\sigma_i$ are fixed to $\sigma^*$, it implies D-RRBF. If $M < N$ and $\sigma_i$ are randomly assigned, it implies Full-RRBF. Also if $M = N$ and $\sigma_i = \sigma^*$, it implies usual RBF Kernel.

4. Experiments

We evaluate our RRBF by using artificial data and standard benchmark dataset[2].

4.1 2-Dimensional Artificial Data

To clarify the property of our RRBF, we describe the simulation experiments using artificial data of 2 classes problem. We make 2-dimensional Gaussian mixture distribution where each class has three independent Gaussian components. Each class have 150 samples, then there are total 300 samples. 100 samples of each class are used for training, and remaining samples are used for testing. We train usual RBF SVM for this simulation by determining their hyper parameters $c, \sigma$ by grid search with 20-fold Cross Validation (CV). For grid search, we set the range of parameters as $c, \sigma \in \{2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{+14}, 2^{+15}\}$.

Let us consider a heuristic algorithm to change the kernel parameter $\sigma$. Only the training samples $x_i$ which are misclassified by usual RBF SVM are adjusted their parameter from $\sigma^*$ to $r\sigma^*$, where $r > 0$. We call it “heuristic RRBF”. Fig. 4 shows the decision boundary of several SVM which are trained by different ways. The top row shows the results of usual RBF SVM with the best $c$ and $\sigma$. This SVM has 94.5% CV accuracy, and misclassified 8 training samples (note that misclassified samples are illustrated as filled markers). The 2nd row shows the decision boundary of RRBF which has the heuristic parameter described above. Filled markers in the top left figure, i.e. samples which are misclas-

| Table 1 Training procedure of proposed method |
|-----------------------------------------------|
| • Input $N$ training samples and their teacher signals $\{x_i, t_i\}_{i=1}^N$. Let $X$ be $\{x_i\}_{i=1}^N$. |
| • Set $M$ which is the number of Kernel seeds ($M \leq N$). |
| • Set $F \geq 1$ which is the number of fold of Cross Validation (CV). |
| • Set range of soft margin $c = \{c_1, \ldots, c_m\}$ and Kernel parameter $s = \{\sigma_1, \ldots, \sigma_n\}$ for grid search of standard RBF SVM. |
| • By using $F$-fold CV, choose the best $c^*$ and $\sigma^*$ from $c$ and $s$ for standard RBF SVM. |
| • Set range of soft margin $c'$ in which the chosen $c^*$ is centered, for grid search of RRBF SVM. |
| • Set $T$ which is the number of iteration to generate random Kernel seeds $S_t$ ($t = 1, \ldots, T$). |
| • for $t = 1, \ldots, T$
| (1) Generate random Kernel seeds $S_t = \{s_i, \sigma_i\}_{i=1}^M$ of Eq. (20).
| (2) Calculate the empirical Kernel vector $K(x_i)$ for all $x_i \in X$ by using $S_t$.
| (3) By using $F$-fold CV, choose the best $\hat{c}^*$ from $c'$ for RRBF SVM.
| (4) Train linear SVM $L_t$ (with the best soft margin $\hat{c}^*$) by using $\{K(x_i)\}_{i=1}^N$ as the $M$ dimensional input vectors.
| (5) Do $F$-fold CV for this SVM and obtain CV accuracy $a_t$. |
| • Choose the best linear SVM $L^*$ which have the highest CV accuracy $a_t$. |

| Table 2 Statistics of experimental data set |
|-----------------------------------------------|
| classes | samples | dimensions |
| Heart | 2 | 270 | 13 |
| Ionosphere | 2 | 351 | 34 |
| Breastcancer | 2 | 683 | 10 |
| Australian | 2 | 690 | 14 |
| Pima | 2 | 768 | 8 |
| German | 2 | 1000 | 24 |
Fig. 4 The results of the simulation experiments. From top to bottom, each row shows the decision boundary of usual RBF SVM, heuristic RRBF and $\sigma$-RRBF, respectively. The left column shows training samples, and right column shows test samples. Circle and square markers represent samples of class 1 and 2, respectively. Filled markers mean those samples are misclassified by each SVM.
sified by usual RBF SVM, are adjusted their parameters $\sigma_i$ from $\sigma^*$ to $2^{-k}\sigma^*$. As can be seen in both the training and test cases, they clearly cause over-fitting. Every location of misclassified samples seem to have quite local and specific classification boundary only for each of them, as like singular points. As a result, although misclassified training samples are reduced 8 to 1, misclassification in test samples rather increases to 10 times from 9 times.

To overcome this problem, we have to use CV to evaluate the generalization performance of obtained classifiers. We train our $\sigma$-RRBF SVM with $T=100$, and then we select the best one having the highest CV accuracy. The bottom row of Fig. 4 shows the decision boundary of the selected $\sigma$-RRBF SVM. It seems that the boundary are naturally extended to improve misclassifications. This SVM has 95.5% CV accuracy and misclassified 5 training samples and 7 test samples. It can be considered that our $\sigma$-RRBF could improve the generalization performance of standard RBF SVM by fine tuning of decision boundary.

### 4.2 Benchmark Data Set

Next we evaluate the classification performance of our RRBF SVM by using six standard data sets (Heart, Ionosphere, Breastcancer, Australian, Pima and German in UCI Machine Learning Repository[2]). Table 2 shows the statistics of these data sets.

For classification experiments, each data set is divided into a training set (2/3 of all samples) and a test set (remaining samples) at random. A training and testing task is repeated 10 times with different divisions based on different random seeds. The averaged CV accuracy or test accuracy (i.e. classification rate for the test set) is shown in this section.

For all experiments, we used

\[ c = \{2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{+14}, 2^{+15}\}, \]
\[ s = \{2^{-15}, 2^{-14}, 2^{-13}, \ldots, 2^{+14}, 2^{+15}\}, \]

for grid search to train usual RBF SVM. When the cross validation is necessary, 20-fold CV was performed to all data sets and all SVM.

We have the hyper-parameter $\hat{\sigma}$ which have to be determined manually. We set $\hat{\sigma}=2.0$ for all data sets and all RRBF. Also we have to determin $T$ which is the number of generating random RRBF SVM. We set $T=200$ for Heart, Ionosphere and Breastcancer, and $T=100$ for Australian, Pima and German. We set $M=N/2$ for D-RRBF and Full RRBF where $N$ is the number of training samples of each data set. For grid search of RRBF SVM, we use the grid points $c' = \{2^{n-4}, 2^{n-2}, 2^n, 2^{n+2}, 2^{n+4}\}$ where $n = \log_2 c^*$. We compared the classification accuracy for the test set of the usual RBF SVM and our RRBF SVM. Table 3 shows the result of comparison. $\sigma$-RRBF shows +1.88% higher and −1.01% lower accuracy than usual RBF for Ionosphere and Pima, respectively. For other four data sets, the difference is smaller than ±1%. The bottom row in the table shows the averaged test accuracy for six data sets. $\sigma$-RRBF have a little bit better averaged accuracy than usual RBF. This seems to be a plausible result; Although our method can modify the decision boundary of RBF SVM, the degree of improvement will be limited because our algorithm is essentially a fine tuning for the precise classifier.

D-RRBF and Full RRBF show the test accuracy on the same level with $\sigma$-RRBF, though they use only
a half of the original dimensions. This result indicates that a half of RBF or RRBF dimensions is unnecessary for the classification, and we could reduce them while maintaining the generalization performance of the classifier.

In this paper, we performed only a simple random selection. However, more legitimate feature selection method might bring more better results. Also about $\sigma$, we only did the random optimization. It will be more suitable to use more orthodox optimization scheme.

5. Conclusions

In this paper, we developed the nonlinear feature extraction method based on RBF kernel SVM. We investigated what kind of influence occurs on the decision boundary of RBF SVM when we assigned different $\sigma$ for each training sample independently. We showed that it does not have much importance to fix $\sigma$ in the issue of classification, and variable $\sigma$ can improve the classification power of RBF SVM. We showed that we could fine-tune the decision boundary of RBF SVM by randomly perturbing the best $\sigma$ for RBF SVM, for every kernel seed.

Also we showed that a half of RRBF dimensions can be reduced by the simple random selection, without deterioration of the classification power. It implies kernel representations have many redundant or unnecessary information for classification, and we could reduce them to obtain more computationally efficient features.

As a future work, it is desired that a better random kernel seeds $S$ should be selected more efficiently. Several combinatorial optimization methods like genetic algorithm or particle swarm optimization may helpful for this purpose.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 23500211.

References

[1] N. Cristianini and J. Shawe-Taylor: An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press (2000)
[2] A. Frank and A. Asuncion: UCI Machine Learning Repository, University of California, School of Information and Computer Science.
[3] T. Kurita and Y. Harashima: Extraction of dimension reduced features from empirical Kernel Vector; The 21st International Conference on Neural Information Processing (ICONIP 2014) (2014)
[4] K.-M. Lin and C.-J. Lin: A study on reduced support vector machines; IEEE Transactions on Neural Networks, Vol. 14, pp. 1449–1459 (2003)
[5] A. Hidaka and T. Kurita: Randomized and dimension reduced Kernel generation for support vector machine; The 46th ISCIE International Symposium on Stochastic Systems Theory and Its Applications (SSS 2014), pp. 191–196 (2014)
[6] K. Nishida and T. Kurita: RANSAC-SVM for large-scale datasets; 19th International Conference on Pattern Recognition (ICPR 2008), pp. 1–4 (2008)
[7] B. Scholkopf, C. J. C. Burges and A. J. Smola: Advances in Kernel Methods - Support Vector Learning, The MIT Press (1999)
[8] V. N. Vapnik: Statistical Learning Theory, John Wiley and Sons (1998)
[9] T. Hastie, R. Tibshirani and J. Friedman: The Elements of Statistical Learning-Data Mining, Inference, and Prediction-, Second Edition, Springer (2006)
[10] G. E. Hinton, S. Osindero and Y. Teh: A fast learning algorithm for deep belief nets; Neural Computation, Vol. 18, pp. 1527–1554 (2006)

Authors

Akinori Hidaka (Member) received the B.Sci degree from Ibaraki University in 2004, and the M.Eng. and D.Eng. degree from the University of Tsukuba, in 2006 and 2009, respectively. He is currently an assistant professor of Tokyo Denki University, in Japan. His current research interests include computer vision based on statistical pattern recognition, especially object detection, tracking, generic object recognition and robot vision. He is a member of IEEE, IEICE and ISCIE.

Takio Kurita received the B.Eng. degree from Nagoya Institute of Technology and the Dr. Eng. degree from the University of Tsukuba, in 1981 and in 1993, respectively. He joined the Electrotechnical Laboratory, AIST, MITI in 1981. From 1990 to 1991 he was a visiting research scientist at Institute for Information Technology, National Research Council Canada. From 2001 to 2009, he was a deputy director of Nueroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST). Also he was a Professor at Graduate School of Systems and Information Engineering, University of Tsukuba from 2002 to 2009. He is currently a Professor at Hiroshima University. His current research interests include statistical pattern recognition and its applications to image recognition. He is a member of the IEEE, the IPSJ, the IEICE of Japan, Japanese Neural Network Society, The Japanese Society of Artificial Intelligence.