Better Decision Boundary for Pattern Recognition with Supplementary Information

Masakazu IWAMURA†, Yoshio FURUYA†, Koichi KISE†,

Shinichiro OMACHI††, and Seiichi UCHIDA†††

† Graduate School of Engineering, Osaka Prefecture University
1–1 Gakuencho, Naka, Sakai, 599–8531 Japan
†† Graduate School of Engineering, Tohoku University
6–6–05 Aoba, Aramaki, Aoba, Sendai, 980–8579 Japan
††† Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi, Fukuoka, 819–0395 Japan
E-mail: †yoshio@m.cs.osakafu-u.ac.jp, †{masa,kise}@cs.osakafu-u.ac.jp,
††machi@ecei.tohoku.ac.jp, †††uchida@is.kyushu-u.ac.jp

Abstract Pattern recognition with supplementary information which differs from the conventional pattern recognition has been proposed. This framework decreases misrecognition rates by using not only a pattern itself but also supplementary information that assists recognition. In the framework, a classifier suitable for the conventional pattern recognition has simply been used. Thus, there is room for improvement in the classifier. In this research, in order to improve the recognition performance with supplementary information, we attempt to change the recognition results of the classifier by the shift of decision boundary from the Bayesian one. Though this reduces recognition rates in the conventional pattern recognition, this increases those in the pattern recognition with supplementary information. We have confirmed this by recognition experiments.

Key words classifier, supplementary information, confusion matrix, Mahalanobis distance, Bayesian decision boundary

1. Introduction

Pattern recognition is an assignment problem of an input sample to a class only with its features. Recently, a different pattern recognition framework, called *pattern recognition with supplementary information* (PRSI), has been focused especially in the field of character recognition and document analysis [1]〜[3]. In this framework, it is an assignment problem of an input sample with both its features and *supplementary information* related to the true class. As an example of the supplementary information, let us think two symbols 0 and 1 for confusing classes A and B. By assigning the symbols to the samples as supplementary information, such as 0 to A and 1 to B, recognition error is reduced.

In the framework of the PRSI, achievable recognition accuracy is approximately determined by the quantity of supplementary information; if a large amount of such information is available, the true class is determined only by the information (namely, recognition is not required!). Therefore, our main interest is the problem of how to obtain a true class correctly by combining a recognition result of a classifier and a limited amount of supplementary information. Thus, if the amount of the supplementary information is determined, we have proposed a method to determine the (approximately) best assignment of the supplementary information for a classifier [1].

In the usual pattern recognition, a classifier is designed to achieve the best performance without any additional information. However, in the PRSI, the classifier is not used alone but together with the supplementary information. This means that the best classifiers in the usual pattern recognition can be the no best one in the PRSI due to the difference of the criterions on the best classifiers.

In this paper, we demonstrate there is the case that a best classifier in the usual pattern recognition is worse in the PRSI. In the usual pattern recognition, the decision boundary which achieves the best performance is called Bayesian decision boundary. In the experiment, we shift the decision

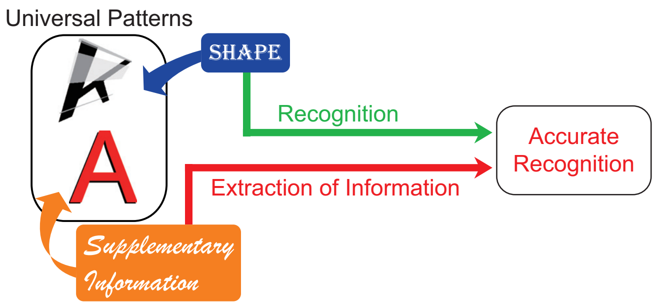


Figure 1 The model of the pattern recognition with supplementary information proposed in [1]. The characters which are called "universal patterns" [2], [3] have both a shape and supplementary information. Combination of such two kinds of information improves recognition accuracy.

boundary of a classifier from the Bayesian one. Though this causes the error rate increase in the usual pattern recognition, this decrease the error rate in the PRSI.

2. Pattern Recognition with Supplementary Information [1]

We briefly explain a part of the result of [1] because it is the important basis of the current work. The recognition model of the PRSI in [1] is illustrated in Fig. 1.

Hereafter, we deal with an N-class problem with K symbols ($K \leqq N$). In the problem, one of K symbols is assigned to each class as the supplementary information. We assume no error occur in the process of "Extraction of Information" in Fig. 1.

2.1 Confusion Matrix

First of all, we have to prepare a confusion matrix (CM) of the classifier in order to discuss the tendency of recognition errors of the classifier. Let $\boldsymbol{W}$ be an $N \times N$ CM whose (i, j) element represents the probability where a pattern of a class ω_i is recognized as one of a class ω_j, that is $P(\omega_j|\omega_i)$. Fig. 2(a) shows a example of such a CM.

Then, we define $\{\mathcal{H}_k\}$ which represent the assignment of the supplementary information. For the assignment, the N classes are divided into K groups each of which has a diverse symbol as the supplementary information. Namely, symbols are assigned to all the rows in the matrix $\boldsymbol{W}$ as in Fig. 2(b). A set of the rows to which the k-th symbol is assigned is defined as $\mathcal{H}_k = \{l | l = l_1, \ldots, l_{|\mathcal{H}_k|}\}$. For example, $\mathcal{H}_1 = \{1, 2\}$, $\mathcal{H}_2 = \{3, 4\}$ and $\mathcal{H}_3 = \{5\}$ in Fig. 2(b). Note that rows in $\mathcal{H}_k$ are not necessarily continuous.

2.2 Symbols Which Achieve the 100% Recognition Rate

With Fig. 2(a), we consider the condition that the symbols which achieve the 100% recognition rate. This will be the theoretical basis of calculation of error rates in Sect. 2.3.

Fig. 2(a) shows that if a recognition result is the class A, the true class can be either the class A, C or E. In this case, the true class cannot be determined by the classifier because it can cause an error. Therefore, at least three kinds of symbols are required as the supplementary information in order to distinguish such confusing three classes.

Similarly, if a recognition result is the class B, the true class can be either the class B, D or E. Thus, different three symbols have to be assigned to the three classes. However, in this case, three symbols are enough and no additional symbols are required because classes A and B are not confusing, and classes C and D are also. Consequently, the 100% recognition rate is achieved when different three symbols are assigned to either "A and B," "C and D," and "E" as in Fig. 2(b), or "A and D," "B and C," and "E."

Finally, the condition of the supplementary information which achieves the 100% recognition rate is that "for a given pair of a recognition result and a symbol, the true class is determined uniquely." By associating this with the CM representation, the condition is that "if for every recognition result and every symbol, the number of non-zero elements of the CM in the intersections of the column corresponding to the recognition result and the rows corresponding to the symbol is less than two." For example, in the CM of Fig. 2(b), there is only one non-zero element in $(1, 1)$ and $(2, 1)$ elements which are corresponding to the combination of the recognition result A and the symbol 1.

2.3 Calculation of Recognition Error Rates

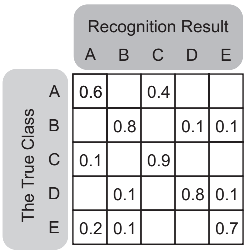
In this section, we calculate an error rate. In order to do so, we have to consider the case other than the 100% recognition rate. Namely, it is the case that the true class is not determined uniquely from a pair of a recognition result and a symbol. In this case, the best way to minimize the recognition error is to choose *the most feasible class*. For example, in Fig. 2(c), if the recognition result is the class A and the symbol is 2, the classes C and E are the candidates. However, the class E should be chosen because it is more probable.

Now we are ready to calculate an error rate. For the given assignment of symbols, the error rate R_{error} is calculated as

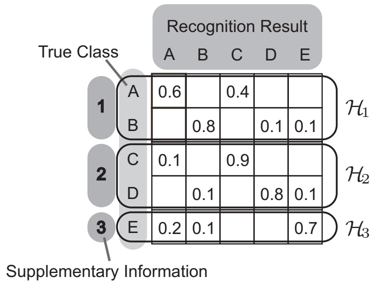
$$R_{\text{error}} = \frac{1}{N} \sum_j \sum_k \left\{ \sum_{l \in \mathcal{H}_k} w_{lj} - \max_{l \in \mathcal{H}_k} w_{lj} \right\}. \qquad (1)$$

In Eq. (1), the first term in the parentheses is the sum of the elements of the matrix $\boldsymbol{W}$ in the intersections of the column corresponding to the j-th recognition result and the rows corresponding to the k-th symbol. The second term is the corresponding element to the class employed as the output.

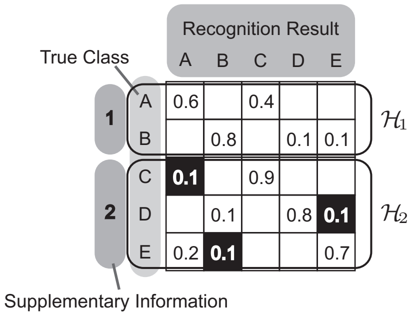
Because an error rate is calculated with Eq. (1), the best assignment of symbols, which gives the smallest error rate, should be found. Let $R_{\text{error}}^{\min}(K)$ be the smallest error rate with K symbols. However, the problem of finding the best



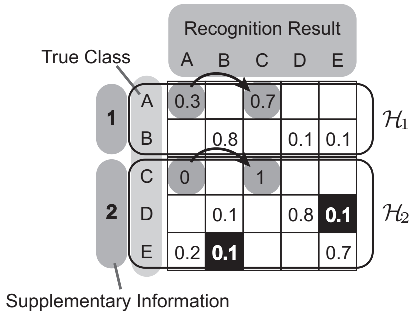
(a) An example of a probabilistic confusion matrix.



(b) Symbols that achieve the 100% recognition rate. # of symbols:3, error rate:0%.



(c) Symbols which cause recognition errors. The reverse colored elements will be recognition errors. # of symbols:2, error rate:0.3/5=6%.



(d) Change of the CM in (c) reduces the error rate. # of symbols:2, error rate:0.2/5=4%.

Figure 2 An example of CMs, assignments of symbols and their error rates. The CMs in (a), (b) and (c) are the same. The assignments of symbols in (c) and (d) are the same. In (d), four elements changes by the shift of the decision boundary between the classes A and C. Empty elements of CMs represent 0.

assignment is too complex to calculate because it is an NP-hard problem. Therefore, to minimize the error rate approximately, we utilize a greedy algorithm to calculate $R^{\min}_{\text{error}}(K)$ shown in Algorithm 1[1]. When K is specified, the algorithm calculates the assignment of K symbols and an approximately smallest error rate $R^{\min}_{\text{error}}(K)$. In the algorithm, an approximately smallest error rate is calculated by decreasing the number of symbols s from N to K by 1. Note that the case of $K = 1$ is the same as the usual pattern recognition method without supplementary information.

3. Proposed Method

In this paper, we propose a method to decrease the error rate in PRSI by shifting the decision boundary from the Bayesian one. In this section, we explain the detail of the method and the reason of the decrease of the error rate.

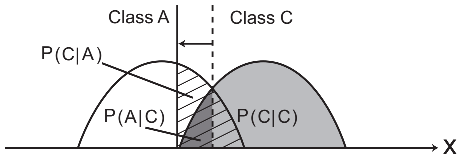
Algorithm 1: A greedy algorithm.

```
Assign different symbols to all classes. Namely, H_k = {k}, for k = 1, ..., N.
/* The number of symbols s is decreased from N − 1 to
   K by one, and finally calculate R^min_error(K).  Note
   that obviously R^min_error(N) = 0.                        */
for s = N − 1 to K do
    Choose a pair of sets H_s and H_t that achieve the minimal
    R_error if the same symbol is assigned to them.
    The same symbol is assigned to the rows of H_s and H_t.
end
```

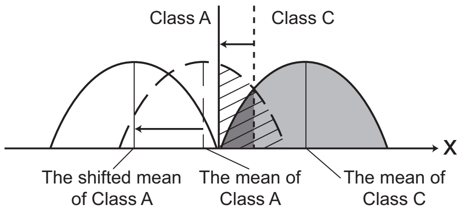
3.1 Relationship Between the Position of the Decision Boundary and the Error Rate

First of all, we review the well-known Bayesian decision theory in the usual pattern recognition. Fig. 3(a) illustrates the probabilities of a sample in one-dimensional feature space. In the region of $P(\mathrm{C}|\mathrm{C})$, a sample of the class C is correctly recognized. In the region of $P(\mathrm{A}|\mathrm{C})$, a sample of the class

[1] Algorithm 1 is slightly modified from the original one in [1] in order to compare with Algorithm 2.

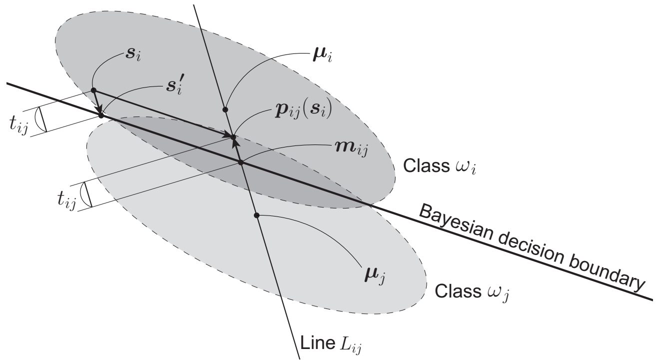


(a) Change of the error rate by the shift of the decision boundary.

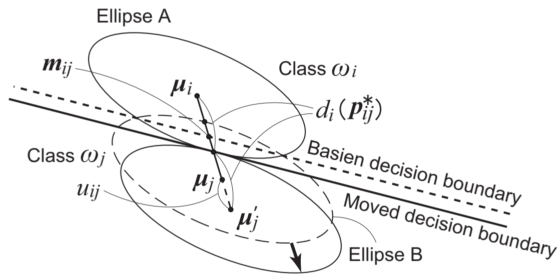


(b) Two distributions does not overlap by the shift of the mean vector of the class A.

Figure 3 Probability distributions and decision boundary.



(a) Calculation of the radius of the class ω_i against the class ω_j.



(b) The amount of the shift of the mean vector of the class ω_j. The ellipse of the class ω_j in the dashed line is the original position of the sample distribution. The one in the solid line is the shifted position.

Figure 4 Shift of decision boundary.

C is misrecognized as the one of the class A. According to the Bayesian decision theory, the Bayesian decision boundary (the dashed line in Fig. 3(a)) minimizes the error rate. We consider the shift of the decision boundary in the usual pattern recognition. If the decision boundary is shifted as shown in Fig. 3(a), the error rate in total increases; though the misrecognition region of $P(\mathrm{A|C})$ decreases, that of $P(\mathrm{C|A})$ increases much more. This is the well-known result.

Then, we consider the shift of the decision boundary in the PRSI. If the decision boundary is shifted, the error rates increases as mentioned in the usual pattern recognition case. However, in the PRSI, we use the supplementary information! If the classes A and C have diverse symbols, no error occurs between the classes A and C wherever the decision boundary is. This means that such decision boundaries are no use for the relevant classes, e.g., the classes A and C. Therefore, we attempt to decrease the error rate by shifting such a decision boundary.

3. 2 Method to Shift the Decision Boundary

In this paper, the Mahalanobis distance is used. The shift of decision boundary is done by the shift of "mean vector." This mean vector is a parameter used in the calculation of the Mahalanobis distance. Usually, such a mean vector is estimated from the samples, and used without any modification for the calculation of the distance. However, in the proposed method, we dare to shift such a mean vector. Note that even if we shift such a mean vector, no sample is shifted because we only change an estimate of the distribution.

We can shift such a mean vector in any direction and in any amount. Thus we restrict them. For the direction, we only take into account the direction between every pair of class means. For the amount of shift, by assuming short-tail distributions of samples as shown in Fig. 3, the mean vector is shifted until two distributions does not overlap (see Fig. 3(b)). For the direction, because it is difficult to estimate the best set of decision boundaries, all the pairs of class means are examined and the best one is chosen. For the direction to shift, In the following sections, we explain how to determine the amount of shift for each pair of class means.

3. 3 Calculation of Class Radius

In order to determine how much the decision boundary should be shifted, a *radius* is determined for a pair of classes. Hereafter, we explain the process to determine the radius of a class ω_i against a class ω_j (see Fig. 4(a)).

For the preparation, let $\boldsymbol{\mu_i}$ and $\boldsymbol{\mu_j}$ be the mean vectors of classes ω_i and ω_j. Let L_{ij} be the line through $\boldsymbol{\mu_i}$ and $\boldsymbol{\mu_j}$. The Mahalanobis distance between a certain point $\boldsymbol{x}$ and $\boldsymbol{\mu_i}$ is given as

$$d_i^2(\boldsymbol{x}) = \left\{ (\boldsymbol{x} - \boldsymbol{\mu}_i)^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_i) \right\}^{1/2}, \qquad (2)$$

where $\boldsymbol{\Sigma}$ is the common covariance matrix shared by all the classes. Thus, the decision boundary is a hyperplane. An arbitrary point $\boldsymbol{x}$ which is on the Bayesian decision boundary between the classes ω_i and ω_j satisfies $d_i^2(\boldsymbol{x}) = d_j^2(\boldsymbol{x})$.

Let $\boldsymbol{s}_i$ be a sample vector of the class ω_i. Now we consider the parallelogram related to $\boldsymbol{s}_i$, which is determined by $\boldsymbol{s}_i$, $\boldsymbol{s}'_i$, $\boldsymbol{m}_{ij}$ and $\boldsymbol{p}_{ij}(\boldsymbol{s}_i)$ in Fig. 4(a). Firstly, $\boldsymbol{s}'_i$ is defined as

$$\boldsymbol{s}'_i = \boldsymbol{s}_i + t_{ij}(\boldsymbol{s}_i)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) \tag{3}$$

with a certain scalar value $t_{ij}(\boldsymbol{s}_i)$[2]. Since $\boldsymbol{s}'_i$ is on the boundary, $d_i^2(\boldsymbol{s}'_i) = d_j^2(\boldsymbol{s}'_i)$ satisfies. By solving for $t_{ij}(\boldsymbol{s}_i)$, we obtain

$$t_{ij}(\boldsymbol{s}_i) = -\frac{\boldsymbol{\mu}_i^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_i - \boldsymbol{\mu}_j^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}_j - 2\boldsymbol{s}_i^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}{2(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)}. \tag{4}$$

Secondly, $\boldsymbol{m}_{ij}$ is the midpoint of the two mean vectors $\boldsymbol{\mu_i}$ and $\boldsymbol{\mu_j}$. Thirdly, $\boldsymbol{p}_{ij}(\boldsymbol{s}_i)$ is the intersection between the line L_{ij} and the hyperplane on $\boldsymbol{s}_i$ which is parallel to the Bayesian decision boundary. Using the characteristics of a parallelogram, we have

$$\boldsymbol{p}_{ij}(\boldsymbol{s}_i) = -t_{ij}(\boldsymbol{s}_i)(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) + \boldsymbol{m}_{ij}. \tag{5}$$

Then, from all the $p_{ij}(\boldsymbol{s}_i)$, the furthest point from $\boldsymbol{\mu_i}$ in the direction of $\boldsymbol{\mu}_j - \boldsymbol{\mu}_i$ is calculated. The points is denoted by p^*_{ij}, and is given as

$$\boldsymbol{p}^*_{ij} = -t^*_{ij}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) + \boldsymbol{m}_{ij}, \tag{6}$$

where $t^*_{ij} = \min_{\boldsymbol{s}_i} t_{ij}(\boldsymbol{s}_i)$.

Finally, the radius of the class ω_i against the class ω_j is calculated by $d_i(\boldsymbol{p}^*_{ij})$, that is the Mahalanobis distance between $\boldsymbol{p}^*_{ij}$ and $\boldsymbol{\mu}_i$.

3.4 Shift of Mean Vector

In this section, we explain the method to shift the mean vector. As shown in Fig. 4(b), the mean vector of the class ω_j is shifted. The amount of the shift, which is denoted by u_{ij}, is determined in order that two ellipses "Ellipses A" and "Ellipses B" in Fig. 4(a) do not overlap. Hereafter, u_{ij} is calculated. Let $\boldsymbol{\mu}'_j$ be the shifted mean vector. $\boldsymbol{\mu}'_j$ is given as

$$\boldsymbol{\mu}'_j = \boldsymbol{\mu}_j + u_{ij}(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) \tag{7}$$

By subtracting $\boldsymbol{\mu}_i$ from the both sides of Eq. (7), and then by using the Mahalanobis norm, we obtain

$$\|\boldsymbol{\mu}'_j - \boldsymbol{\mu}_i\|_{\boldsymbol{\Sigma}_{ij}} = \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_i\|_{\boldsymbol{\Sigma}_{ij}} + u_{ij}\|\boldsymbol{\mu}_j - \boldsymbol{\mu}_i\|_{\boldsymbol{\Sigma}_{ij}}. \tag{8}$$

By applying $\|\boldsymbol{x} - \boldsymbol{\mu}_i\|_{\boldsymbol{\Sigma}_{ij}} = d_i(\boldsymbol{x})$ for Eq. (8), we obtain

$$d_i(\boldsymbol{\mu}'_j) = d_i(\boldsymbol{\mu}_j) + u_{ij}d_i(\boldsymbol{\mu}_j). \tag{9}$$

Because the midpoint of $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}'_j$ is on the decision boundary, we have $d_i(\boldsymbol{\mu}'_j) = 2d_i(\boldsymbol{p}^*_{ij})$. By substituting this for Eq. (9), finally we obtain

$$u_{ij} = \frac{2d_i(\boldsymbol{p}^*_{ij}) - d_i(\boldsymbol{\mu}_j)}{d_i(\boldsymbol{\mu}_j)}. \tag{10}$$

[2] $t_{ij}(\boldsymbol{s}_i)$ can be a minus value.

3.5 Proposed Procedure

In summary of the above discussion, we show the proposed procedure in Algorithm 2. When K is specified, the algorithm calculates the assignment of K symbols, an approximately smallest error rate $R^{\min}_{\mathrm{error}}(K)$, and a set of mean vectors (including zero or one shifted mean vector). In the first step of the algorithm, Algorithm 1 is used to calculate the assignment of $K+1$ symbols. After that, all the pairs of class means are examined as candidate decision boundaries[3]. For each candidate, an approximately best assignment of K symbols is searched by the greedy algorithm, and an error rate is calculated. Then, the best decision boundary is chosen from the candidates. Finally, the assignment of K symbols, an approximately smallest error rate $R^{\min}_{\mathrm{error}}(K)$, and a set of mean vectors are calculated.

4. Experiments

In the framework of PRSI, we compared Algorithm 2 to Algorithm 1. Feature vectors were created as follows; the character images of digit samples in NIST Special Database 19 [4] were normalized nonlinearly [5] to fit in a 64×64 square, and 196-dimensional Directional Element Features [6] were extracted.

Experiments were carried out in 10-times cross validation. For each experiment, 3,600 samples per class were used in total: 3,400 samples were for training, 100 samples were for validation, and 100 samples were for testing. Because the Mahalanobis distance is used for discrimination, mean vectors for the classes and a pooled covariance matrix (the average of covariance matrices here) were calculated with the training samples. With the validation samples, a CM was estimated. Then, two assignments of symbols were carried out with the CM: one assignment labelled "Conventional method" is calculated by Algorithm 1, and the other one labelled "Proposed method" is calculated by Algorithm 2. Note that in Algorithm 2, the mean vector of a class was shifted. Finally, with the test samples and the symbols, an error rate was calculated for both recognition methods. The symbols were used without any error.

The average error ratios of both methods are shown in Fig. 5. As mentioned before, if the number of symbols is 1, it is the usual pattern recognition. The figure shows that unless the number of symbols was 1, the average error rates of the proposed method was decreased to less than 3/4 of the conventional method's ones. In the case that the number of symbols was 2, the average error rate was decreased the most, by 1.39%. On the other hand, the average error did not change in the case that the number of symbols was

[3] The candidates include the original decision boundary.

Algorithm 2: The proposed algorithm.

Decrease the number of symbols to $K+1$ by Algorithm 1.
`/* From the set of rows` $\{\mathcal{H}_k\}$`, a pair are chosen. Let the pair` $\mathcal{H}_s$ `and` $\mathcal{H}_t$ $(s < t)$`. */`
forall $\mathcal{H}_s, \mathcal{H}_t \in \{\mathcal{H}_k\}$ **do**
  Calculate an error rate e_0 by temporarily assigning a same symbol to the classes related to the rows in $\mathcal{H}_s$ and $\mathcal{H}_t$.
  `/* The lowest error rate is searched by shifting the decision boundary between the classes` ω_i `and` ω_j `for all` i `and` j `below. */`
  Let $e_{\min}$ be the lowest error rates. Initialize it by some large value.
  for $i = 1$ **to** N **do**
    for $j = 1$ **to** N **do**
      if $i \neq j$ *and* $w_{ij} \neq 0$ **then**
        Calculate the error rate e with the shifted mean vector of the class ω_j by the method in Sect. 3. 4.
        if $e < e_{min}$ **then**
          `/* Update the lowest error rate and the parameters. */`
          Substitute e for $e_{\min}$.
          Substituted ω_i, ω_j, s and t for ω'_i, ω'_j, s' and t', respectively.
        end
      end
    end
  end
end
Decrease the number of symbols by 1 to K by assigning a same symbol to $\mathcal{H}_{s'}$ and $\mathcal{H}_{t'}$.
`/* The decision boundary is shifted if the error rate decreases below. */`
if $e_{min} < e_0$ **then**
  Shift the mean vector of the class ω'_j against the class ω'_i.
end

1. This is because Algorithm 2 ignores increased error rates. In Algorithm 2, many decision boundaries are examined and only the smallest error rate is employed. If all the error rates are larger than the original one, all of them are ignored and the original one is employed. Therefore, in this case, all the examined decision boundaries were worse than the Bayesian decision boundary. Finally, we confirmed there is a case that (i) the shift of decision boundary increases the error rate in the usual pattern recognition, and (ii) it decreases the error rate in the PRSI.

5. Conclusion

Recently, a different pattern recognition framework from the usual one, called *pattern recognition with supplementary information* (PRSI), has been focused. In this paper, we demonstrated there is the case that a best classifier in the usual pattern recognition is not always the best in the PRSI. In other words, the shift of decision boundary from the Bayesian decision boundary can improve recognition rates in the PRSI framework though it decreases in the usual pattern recognition. This is due to the difference of the criterion of "the best classifier." Future work includes the design of the best classifier in the PRSI.

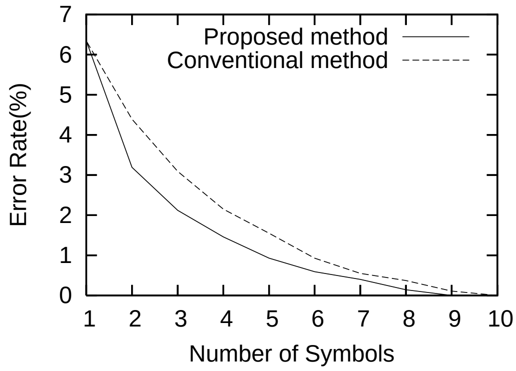


Figure 5 The average error rates of "Conventional method" and "Proposed method." In the "Conventional method," the assignment of the symbols was determined by Algorithm 1. In the "Proposed method," the assignment was determined and the mean vector was shifted by Algorithm 2.

References

[1] M. Iwamura, S. Uchida, S. Omachi and K. Kise, "Recognition with supplementary information —how many bits are lacking for 100% recognition?—, ," Proc. First International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2005), pp.68–75, Aug., 2005.

[2] S. Uchida, M. Iwamura, S. Omachi and K. Kise, "OCR fonts revisited for camera-based character recognition," Proc. 18th International Conference of Pattern Recognition (ICPR2006), Aug., 2006.

[3] S. Omachi, M. Iwamura, S. Uchida and K. Kise, "Affine invariant information embedment for accurate camera-based character recognition," Proc. 18th International Conference of Pattern Recognition (ICPR2006), Aug., 2006.

[4] P. J. Grother, "NIST special database 19 —handprinted forms and characters database—, Technical report, ," National Institute of Standards and Technology, Mar., 1995.

[5] H. Yamada, K. Yamamoto and T. Saito, "A nonlinear normalization method for handprinted kanji character recognition —line density equalization—," Pattern Recognition, vol.23, pp.1023–1029, 1990.

[6] N. Sun, Y. Uchiyama, H. Ichimura, H. Aso and M. Kimura, "Intelligent recognition of characters using associative matching technique, ," Proc. Pacific Rim Int'l Conf. Artificial Intelligence (PRICAI'90), pp.546–551, Nov., 1990.