System Evaluation of Ternary Error-Correcting Output Codes for Multiclass Classification Problems*

Shigeichi Hirasawa[1], *Life Fellow*, *IEEE*, Gendo Kumoi[2], Hideki Yagi[3], *Member*, *IEEE*,
Manabu Kobayashi[4], *Member*, *IEEE*, Masayuki Goto[2], *Member*, *IEEE*,
Tetsuya Sakai[5], and Hiroshige Inazumi[6]

***Abstract*— To solve multiple classification problems with $M(\geq 3)$ categories, many studies have been devoted using $N(\geq \lceil \log_2 M \rceil)$ binary ($\{0,1\}$) classifiers, where these systems are known as binary Error-Correcting Output Codes (binary ECOC). As an extended version of the binary ECOC, the ternary ($\{0,*,1\}$) ECOC have also been discussed, where ternary classifiers classify data into positive examples when the element is 1, into negative examples when the element is 0, and no classification when the element is $*$. In this paper, we discuss the ternary ECOC system from the view point of the system evaluation model based on rate-distortion function. First, we discuss a table of M code words with length N which is given by a ternary matrix W of M rows and N columns. Next, by leveraging the benchmark data for multi-class document classification which is widely used in Japan, the relationships between the probability of classification error P_e and the number of the ternary classifiers N for a given M are experimentally investigated. In addition, by assuming the M-dimensional Normal distribution for a classification data model, the relationship between P_e and N for a given M is also examined. Finally, we show by the system evaluation model that the ternary ECOC systems have desirable properties such as *"Flexible"*, *"Elastic"*, and *"Effective Elastic"*, when M becomes large.**

I. INTRODUCTION

In recent research and development on data science, especially in the field of machine learning, the importance of techniques for extraction of necessary information from large-scale data and for automatic data classification is rapidly increasing [5]. Various kinds of sets such as a set of documents, a set of patterns, a set of images, etc. to be classified have M classes (categories), where $M \geq 3$. Although the Support Vector Machines (SVMs) have been improved in performance to directly solve multiclass classification problems using as a single multi-valued classifier, the attainable performance is still insufficient and the configuration method for such SVMs become complicated, as M becomes large. Therefore, a method to solve them using multiple simple q-ary classifiers in parallel is introduced and discussed, where $q < M$. Since this method uses the concept of error-correcting codes, it is called error-correcting output code (ECOC) [1], [2], [3], [4], [5], [13], [14], [15]. Here, the coding of ternary (q=3) ECOC is given by an $M \times N$ matrix W, where M is the number of categories, and N is the number of q-valued classifiers, and W=$[w_{ij}]$ is called a *code word table*. While a decoding method of ternary ECOC is to classify data as a positive example when w_{ij} = 1, as a negative example when w_{ij} = 0, and to not classify when w_{ij} = * (don't care). Compared to the binary ECOC, it has been shown that the ternary ECOC has significantly improved its expressive power of the code word table, and the latter has smaller probability of classification error than the former [4],[15]. However, the number of required classifiers is significantly increased for the ternary ECOC compared to the binary ECOC for a given M.

On the other hand, in the late 1970s, J. Pearl et al. constructed a system evaluation model using a rate-distortion function and performed a detailed evaluation for Question Answering (QA) systems [16]. The trade-off between the storage space and the probability of error which corresponds to rate and distortion, respectively, inherent in the QA systems was theoretically clarified. This idea pays attention to the trade-off curve saying, *"If we tolerate only a small error probability, we can drastically reduce the storage space,"* for which the system is called *elastic* as the system size becomes large. The desirable properties such as *flexible* and *elastic*, are defined and made it possible to evaluate systems of interest. In our previous work, we applied this model to information systems such as network structures [12] and files with consecutive retrievable [11]. Pearl et al.'s model has, however, some strong restrictions to apply to the target information systems. We successfully reduced these restrictions, and introduced new properties to make them useful by a generalized trade-off model [6],[7]. By using this model, it is possible to ask whether the target information systems have desirable properties or not. Hence, prior to researching, developing or designing the target systems, it is useful to

*Research leading to this paper was partially supported by the Mistry of Education, Science, Sports and Culture Grant-in-Aids for Scientific Research (B) 26282090, and (C) 25330045.

[1] Shigeichi Hirasawa is with Research Institute for Science and Engineering, Waseda University, Tokyo, 169-8555 Japan hira@waseda.jp

[2] Gendo Kumoi and Masayuki Goto are with School of Creative Science and Engineering, Waseda University, Tokyo, 169-8555 Japan m.kumoi@kurenai.waseda.jp and masagoto@waseda.jp

[3] Hideki Yagi is with Graduate School of Informatics and Engineering, the University of Electro-Communications, Tokyo 182-8585 Japan h.yagi@uec.ac.jp

[4] Manabu Kobayashi is with Center for Data Science, Waseda University, Tokyo, 169-8050 Japan mkoba@waseda.jp

[5] Tetsuya Sakai is with School of Fundamental Science and Engineering, Waseda University, Tokyo, 169-8555 Japan tetsuyasakai@acm.org

[6] Hiroshige Inazumi is with Faculty of Informatics, Aoyama Gakuin University, Kanagawa, 229-8558 Japan inazumi@si.aoyama.ac.jp

check the properties of the target information systems in advance. We apply this model to the disk allocation problem for questionnaire data files, and have found a method which enables them to have the desired properties efficiently by using unequal error-correcting codes [8].

In this paper, we apply the system evaluation model to the ternary ECOC. The trade-off relationship between the average probability of classification error (performance degradation) P_e and the number of ternary classifiers (investment cost) N as two variables and the number of categories (the scale of the system) M as a parameter, is experimentally examined by using the benchmark data of multiclass classification problem (Japanese newspaper articles of the 2015 Yomiuri Shimbun [17]) and M-dimensional Normal distribution for classification data model [10]. It should be noted that it is not the purpose of this research to obtain a code word table which minimizes the probability of classification error [13] using ternary ECOC. That is, this paper studies another aspect on ECOC.

Throughout this paper, we evaluate the average performance of the ternary ECOC. Section II briefly describes the system evaluation model. How to construct the ECOC is discussed in Section III. Section IV shows the construction methods for code word tables, and Section V reports on our experimental results. Section VI provides the concluding remarks of this paper.

II. Preliminaries

A. Outline of Rate-Distortion Theory

Rate-distortion theory discusses data compression by the trade-off property between rate and distortion [16]. The rate-distortion function can be written as:

$$L = R(D) \tag{1}$$

where L is the rate defined by $L = (1/n^*) \log |C|$, where $|C|$ is the number of code words, n^*, the code length, and D, the distortion. The $L = R(D)$ is usually a convex downward and non-increasing function of D.

B. System Evaluation Model

Generally, the rate L discussed in the previous subsection corresponds to the investment cost of a system, and distortion D, the performance degradation of the system [16]. By extending the rate distortion model, we have proposed the trade-off model for system evaluation [6], [7], where we have also introduced a parameter G as the scale of the system.

Let the rate L be normalized by the maximum of L, $L_{\max}$, and the distortion D, by the maximum of D, $D_{\max}$, then we have the following normalized function by $\ell = L/L_{\max}$, and $d = D/D_{\max}$, and introducing G:

$$\ell = r(d; G). \tag{2}$$

For evaluation of the systems, we define the following properties to the normalized trade-off system evaluation function (2):

Definition 1

1) *Flexible* [16]: The system is said to be *flexible*, if $\ell = r(d)$ is a decreasing and convex downward function of d. And the system A with $\ell = r_A(d; G)$ is said to be *more flexible* than the system B with $\ell = r_B(d; G)$, if $r_A(d; G) < r_B(d; G)$ for arbitrary $d(0 < d < 1)$, and $G(G > 1)$. (See Fig. 1 (1)).
2) *Elastic* [16]: The system with $\ell = r(d; G)$ is said to be *elastic*, if $\ell = r(d; G)$ is a decreasing function of G for arbitrary $d(0 < d < 1)$. (See Fig. 1 (2)).
3) *Effective elastic* [6]: The system is said to be *effective elastic*, if the system is *elastic* and $\ell = r(d; G)$ is a convex downward function of G. (See Fig. 1 (3)).
4) *Trivial elastic* [16]: The system with $\ell = r(d; G)$ is said to be *trivial elastic*, if $d = r^{-1}(0; G)$ is a decreasing function of G, where $d = r^{-1}(\ell; G)$ is the inverse function of $\ell = r(d; G)$.
5) *Marginal elastic* [6]: The system with $d = r^{-1}(\ell, G)$ is said to be *marginal elastic*, if $d = r^{-1}(0, G)$ is a convex downward function of G.

Although the system discussed here is not applicable to 4) *Trivial elastic* and 5) *Marginal elastic*, they are sometimes observed depending on the structure of systems (See [11]).

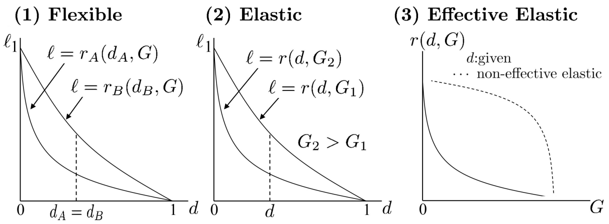

Fig. 1. Trade-off model for system evaluation

Fig.1 illustrates typical examples given in Definition 1. As shown in Fig. 1 (1), ℓ is a decreasing and convex downward function of d, hence we can decrease ℓ drastically tolerating a slight increase in d.

III. Multi-class Classification System using Ternary Classifier

A. Configuration of Multi-class Classification System

In the automatic classification problem of real data such as newspaper articles and hand-written characters, the number of categories to be classified is $M(M \geq 3)$. For example, the former [17] has 10 categories in terms of their publication space (See Table IV) and the latter has 26 categories.

Classification problems of $M(\geq 3)$ categories can be solved directly using multi-valued classifiers such as the SVMs. However, as M increases, the configuration of multi-valued classifiers becomes complicated and it becomes difficult to achieve high performance. Therefore, a method has been proposed for performing multiclass classification using multiple simple binary classifiers [3]. The key idea behind it is to reduce the M-class classification problem to a series of N binary problems. Since it uses the concept of Error-Correcting Codes (ECC), the method is called (binary)

Error-Correcting Output Codes (ECOC). Furthermore, the (binary) ECOC method is extended to the ternary ECOC method [1].

Coding

Usually, in coding theory, the i $(i = 1, 2, \cdots, M)$ of the i-th category C_i is an information symbol, and check (redundant) symbols are added to this to generate a code word of length N. If the symbol is binary, the i is converted to a binary sequence of length K, and a $K \times N$ generator matrix is multiplied to obtain a code word $\boldsymbol{c}_i$. If an error[1] (noise) occurs in the communication channel, the added redundancy is used to correct and detect the error. Therefore, the distance (measured with Hamming distance, Lee distance, Euclidean distance etc.) between arbitrary code words obtained is designed to be as large as possible. The Hamming codes, the BCH codes, the RM (Reed-Muller) codes, the RS (Reed-Solomon) codes etc. are well-known as powerful codes.

In ECOC, using these known codes, M code word vectors of length N are extracted, and a *code word table*[2] W which meets the target data is obtained. Here, we represent matrix W as follows:

$$\begin{aligned} W &= [w_{ij}](w_{ij} \in 0, *, 1, i = 1, 2, \cdots, M, j = 1, 2, \cdots, N) \\ &= [\boldsymbol{d}_1^{\mathrm{T}}, \boldsymbol{d}_2^{\mathrm{T}}, \cdots, \boldsymbol{d}_N^{\mathrm{T}}] \\ &= [\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_M]^{\mathrm{T}} \qquad (3) \end{aligned}$$

where T represents the transpose of a matrix (or a vector). Note that matrix W implicitly describes a decomposition scheme of the original multiclass classification problem.

In the training phase, letting the element of matrix W, $w_{ij} \in \{0, *, 1\}$ in each classifier $\boldsymbol{d}_j$ $(j = 1, 2, \cdots, N)$, if $w_{ij} = 1$, then training data are used as positive examples, if $w_{ij} = 0$, then they are used as negative examples, and if $w_{ij} = *$, then they are ignored.

Example 1. Simple examples of the code word table are shown in TABLE I. □

TABLE I
EXAMPLES OF CODE WORD TABLES.

(a) Binary "one vs. the rest" method ($M = 4$ and $N = 4$)

	$\boldsymbol{d}_1$	$\boldsymbol{d}_2$	$\boldsymbol{d}_3$	$\boldsymbol{d}_4$
$\boldsymbol{c}_1$	1	0	0	0
$\boldsymbol{c}_2$	0	1	0	0
$\boldsymbol{c}_3$	0	0	1	0
$\boldsymbol{c}_4$	0	0	0	1

(b) Ternary "one vs. one" method ($M = 4$ and $N = 4$)

	$\boldsymbol{d}_1$	$\boldsymbol{d}_2$	$\boldsymbol{d}_3$	$\boldsymbol{d}_4$
$\boldsymbol{c}_1$	1	*	*	0
$\boldsymbol{c}_2$	0	1	*	*
$\boldsymbol{c}_3$	*	0	1	*
$\boldsymbol{c}_4$	*	*	0	1

This extension increases the expressive power of ECOC, so that ternary ECOC can be applied to nearly all binary ECOC. For example, pairwise classification, where one classifier is trained for each pair of classes, could not be modeled in the original framework, but can be modeled with ternary ECOC (See TABLE I (b)).

Decoding Method

The ternary classifier used in this study is the SVM. When test data (whose category is unknown) $\boldsymbol{y}$ is input to the j-th classifier, $h_j(\boldsymbol{y}) \in (-\infty, +\infty)$ is output. Based on this, the test data $\boldsymbol{y}$ is classified into $C_{i'}$ estimated by $f(y) = C_{i'} \in C$, where

$$f(\boldsymbol{y}) = \arg\max_{C_{i'} \in \boldsymbol{C}} g_i(\boldsymbol{y}). \qquad (4)$$

This decoding method is called the maximum margin soft decision SVM [1], [5], which is calculated as

$$g_i(\boldsymbol{y}) = \sum_j I(w_{ij}) h_j(\boldsymbol{y}), \qquad (5)$$

where

$$I(w_{ij}) = \begin{cases} 1, & w_{ij} = 1 \\ 0, & w_{ij} = * \\ -1, & w_{ij} = 0. \end{cases}$$

It should be noted that the ternary ECOC has different meanings of $*$ depending on the decoding methods. There are Hamming distance decoding method, Euclidean distance decoding method, loss based decoding method, etc. and classification performances are different. The decoding method of equations (4) and (6) is the one of the attenuated soft decision decoding methods.

B. Correspondence Between System Evaluation Model and ECOC

In Section II, we have shown the rate-distortion function (1), and the trade-off system evaluation function (2). The corresponding trade-off function (9) for the ternary ECOC will be shown later in Section IV, where the number of ternary classifiers n corresponds to the investment cost ℓ, and the probability of classification error p_e, the performance degradation d, respectively. And also, the scale of system G corresponds to the number of category M. Total correspondence between system evaluation model and ECOC is given by TABLE II.

TABLE II
EVALUATION OF ECOC. (CORRESPONDENCE TABLE) [3]

Rate-Distortion Theory	System Evaluation Model	Ternary ECOC
Rate (L)	Investment Cost (ℓ)	Number of Ternary Classifiers (n)
Distortion (D)	Performance Degradation (d)	Probability of Classification Error (p_e)
	Scale of System (G)	Number of Categories (M)

[1] In the case of ECOC, the error is caused by external influences such as a too small sample size [15].

[2] This matrix is called the ECOC Matrix [11], the coding matrix [1], the classifier structure, the code word structure, etc.

[3] In the following sections, unlike Fig. 1, the horizontal axis is used for n, and the vertical axis, for p_e.

TABLE III

TERNARY EXHAUSTIVE CODE. ($M = 4$, AND $N_{\max} = 25$)

	$\boldsymbol{d}_1$	$\boldsymbol{d}_2$	$\boldsymbol{d}_3$	$\boldsymbol{d}_4$	$\boldsymbol{d}_5$	$\boldsymbol{d}_6$	$\boldsymbol{d}_7$	$\boldsymbol{d}_8$	$\boldsymbol{d}_9$	$\boldsymbol{d}_{10}$	$\boldsymbol{d}_{11}$	$\boldsymbol{d}_{12}$	$\boldsymbol{d}_{13}$	$\boldsymbol{d}_{14}$	$\boldsymbol{d}_{15}$	$\boldsymbol{d}_{16}$	$\boldsymbol{d}_{17}$	$\boldsymbol{d}_{18}$	$\boldsymbol{d}_{19}$	$\boldsymbol{d}_{20}$	$\boldsymbol{d}_{21}$	$\boldsymbol{d}_{22}$	$\boldsymbol{d}_{23}$	$\boldsymbol{d}_{24}$	$\boldsymbol{d}_{25}$
$\boldsymbol{c}_1$	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	$*$	$*$	$*$	1	1	1	$*$	$*$	$*$
$\boldsymbol{c}_2$	0	0	0	0	1	1	1	0	1	1	0	1	1	$*$	$*$	$*$	0	0	0	0	$*$	$*$	1	1	$*$
$\boldsymbol{c}_3$	0	0	1	1	0	0	1	1	0	1	$*$	$*$	$*$	0	1	1	0	1	1	$*$	0	$*$	0	$*$	1
$\boldsymbol{c}_4$	0	1	0	1	0	1	0	$*$	$*$	$*$	1	0	1	1	0	1	1	0	1	$*$	$*$	0	$*$	0	0

IV. CONSTRUCTION METHODS OF CODE WORD TABLE

A. Exhaustive Codes

In our previous work which discussed binary ECOC [9], [10], we started from the binary exhaustive codes, since exhaustive codes are exhaustively extracting all column vectors which contribute to classification. Such a method is, however, not realistic because the code length increases exponentially as M increases. In fact, the code length $N_{\max}$ of the binary exhaustive codes is given by:

$$N_{\max} = 2^{M-1} - 1. \tag{6}$$

We can extract only the useful column vectors to construct efficient exhaustive codes with short code lengths (later we call them shortened exhaustive codes). Here, an $M \times N_{\max}$ ternary exhaustive code for M categories and code length $N_{\max}$ are derived as follows:

Let the number of $*'$s, $|*| = \alpha$, a binary exhaustive code for $M - \alpha$ categories be represented by $ex(M - \alpha)$, and a part of the ternary exhaustive code for which each column vector includes $\alpha*'$s be represented by $EX(\alpha)$. First, let $\alpha = 0$, and $EX(0) = ex(M)$. Next, set $\alpha = 1$ and add $\alpha*'$s to each column of $ex(M - \alpha)$ to make a column vector of length M (in this case, there are $\binom{M}{M-\alpha}$ combinations), and $EX(1) = ex(M - 1)$. Letting $\alpha = \alpha + 1$ until $\alpha = M - 2$ (since each column includes at least 2 symbols, i.e., 0 and 1). Then, we have a ternary exhaustive code which is given by the concatenation of $EX(0), EX(1), \cdots, EX(M-2)$, and the code length $N_{\max}$ of the ternary exhaustive codes is given by the following equation:

$$N_{\max} = \sum_{\alpha=0}^{M-2} \binom{M}{M-\alpha} (2^{M-1-\alpha}). \tag{7}$$

Example 2: An example of $M = 4$, a ternary exhaustive code is shown in TABLE III. □

B. Shortened Exhaustive Codes

If we consider a shortened[4] version of exhaustive code, then we can decrease the investment cost (decreasing the number of binary classifiers) N by tolerating the performance degradation (increasing the probability of classification error) P_e. For given M categories, we shall evaluate the relationships between N and P_e, where P_e corresponding to N column vectors selected from the $N_{\max}$ column vectors is obtained and is averaged over all $\binom{N_{\max}}{N}$ combinations. The result is experimentally obtained as the following relation:

$$P_e = S(N, M). \tag{8}$$

V. EXPERIMENTS AND DISCUSSIONS

A. Conditions of Experiments

In this paper, we primarily focus on a document classification task, since it is a typical multiclass classification problem. As the benchmark data, we use the Yomiuri 2015 text classification problem with Yomiuri Shimbun articles [17]. TABLE IV shows the specifications of benchmark data and experiment data.

TABLE IV

SPECIFICATION OF EXPERIMENTAL DATA.

Benchmark data (2015 Yomiuri Shimbun article) Specification	
Number of Categories	10
Word feature extraction by feature vector (dimension)	Morphological analysis of documents (7,432 words)
Category(number of data)	Politics (23,719); Economy (19,490); Local (16,414); Sports (30,495); Culture (15,127); Life (10,747); Crime case (24,545); Science (2,369); International (1,667); Imperial family (257)
Experiment Data (data extracted and used) Specification	
Categories (number of categories)	Politics, Economy, Sports, Local, Culture, Life, Crime Case, Science (8)
Total number of experimental data	12,000
Number of training data / category	1,350 (10,800 in total)
Number of test data / category	150 (1,200 in total)

B. Relationship between P_e and N

Experiment 1: First we show the relationship between the average probability of classification error P_e and the code length N which corresponds to (8) for the case of $M = 4$ in Fig. 2. Note that for given N, M categories are selected from $M_{\max}$ categories, and corresponding P_e is also averaged over all $\binom{M_{\max}}{M}$ combinations, where $M_{\max} = 8$.

Remark 1: In Fig. 2, we also show the range of the probability of classification error P_e for given N by the vertical bold lines, since the probability of classification error gives different values depending on the combination of selected column vectors, where each column $\boldsymbol{d}_j \in \{0, *, 1\}^M$. □

C. Trade-Off Curves for System Evaluation— Relationship between p_e and n

We let $N_{\min}$ be the code length of the exhaustive code with distinct M row vectors which has the smallest number

[4] In coding theory, when a code of length N and the number of information symbols K is represented by an (N, K) code, we call an $(N - s, K - s)$ code as a shortened code. Here, although exhaustive code is not a systematic code, where systematic code can separate the information symbols from the code word, we call a code which is obtained by removing the s symbols from the code word of length N as a shortened code.

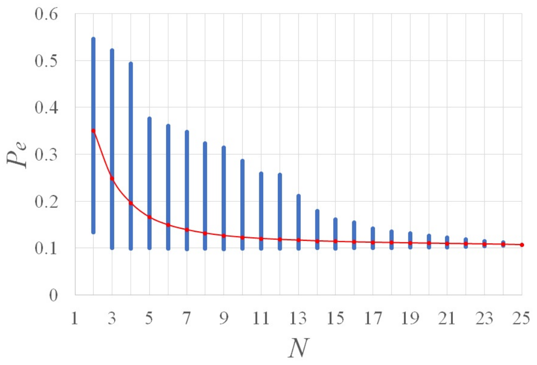

Fig. 2. Relationship between average probability of classification error P_e and code length N by ternary ECOC using Benchmark Data [17] for $M = 4$.

of column vectors, then we can choose[5] $N_{\min} = \lceil \log_2 M \rceil$. While, $N_{\max}$ is obviously given by the code length of the (full) exhaustive code.

By the above preparation, (normalized) trade-off function for system evaluation is given by letting $P_e/P_{e,\max} = p_e$, and $N/N_{\max} = n$, we have the following equation, where $P_{e,\max}$ corresponds to the value of P_e for $N = N_{\min}$, where we assume the function is monotonically decreasing:

$$p_e = s(n, M). \tag{9}$$

Experiment 2: We examine the same experiments as stated in Experiment 1 for $M = 4, 5, 6, \text{and } 7$, by ternary ECOC, the results obtained are shown in Fig. 3, together with those by binary ECOC.

Remark 2: From Fig. 3, we have

- The ternary ECOC is *more flexible* than the binary ECOC, since p_e of ternary ECOC is smaller than that of binary ECOC for any n and for a given M. □

Remark 3: For the ternary ECOC, we have from Fig. 3

- The trade-off curves are convex downward, and are shown to be *flexible*.
- Most of the trade-off curves go toward the origin as M becomes large except in the neighborhood of $n = 1$, and are almost *elastic*. □

D. Relationship between n and M for given p_e

Experiment 3: From the results of Fig. 3, we can obtain Fig. 4, which shows n as a function of M for a constant value of p_e, where $p_e = 0.7, 0.6, 0.5, \text{and } 0.4$.

Remark 4: For both (binary and ternary) cases, the curves are almost convex downward, hence are almost *effective elastic*. □

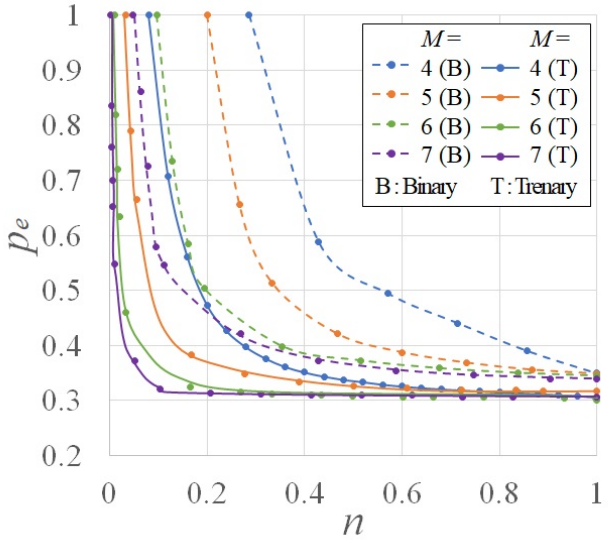

Fig. 3. Trade-off relationship between investment cost n and performance degradation p_e with scale of system M by ternary ECOC using Benchmark Data [17].

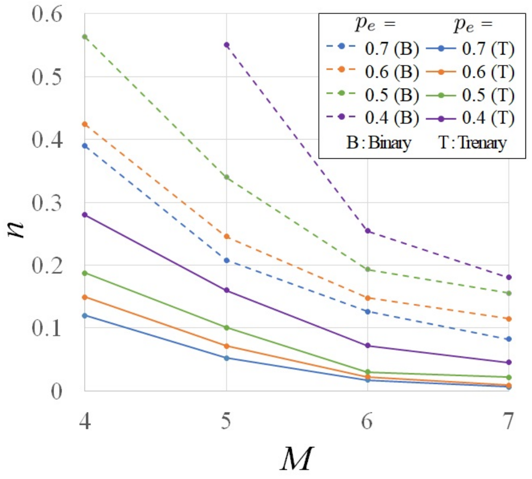

Fig. 4. Relationship between scale of system M and investment cost n by ternary ECOC using Benchmark Data [17].

E. Additional Experiment

Finally, we apply our method to the artificial data model which is given by the random variables generated by M-dimensional Normal distribution $N(\boldsymbol{\mu}, \sigma^2)$, where $\boldsymbol{\mu} = (\mu_1, \mu_2 \cdots, \mu_M)$ is the average, and $\sigma^2 = [\sigma_{ij}^2]$, the variance. The specification of training data and test data is given as shown in TABLE V.

Experiment 4: For the case of $\mu_i = 1 (i = 1, 2, \cdots, M)$, and $\sigma_{ii}^2 = 0.25, \sigma_{ij} = 0 \ (i \neq j)$, Fig. 5 shows only the results.

Remark 5: From Fig. 5, we have

- The trade-off curves are decreasing and convex downward, and are shown to be *flexible*.
- Most of the trade-off curves go toward the origin as M becomes large, and are *elastic*.
- The trade-off curves in Fig. 4 become narrower on the line where p_e is a constant, as M becomes large. Hence it is almost *effective elastic*. □

As we have described above, the remarkable results obtained by Experiments are highlighted as *Remarks* 1 to 5.

[5] $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Note that we can represent M by $\lceil \log_2 M \rceil$ column vectors by usual binary representation.

TABLE V
DESIGNATION OF ARTIFICIALLY GENERATED DATA.

Categories (number of categories)	4, 5, 6, 7
Number of training data / category	1000 (7,000 in total)
Number of test data / category	100 (700 in total)

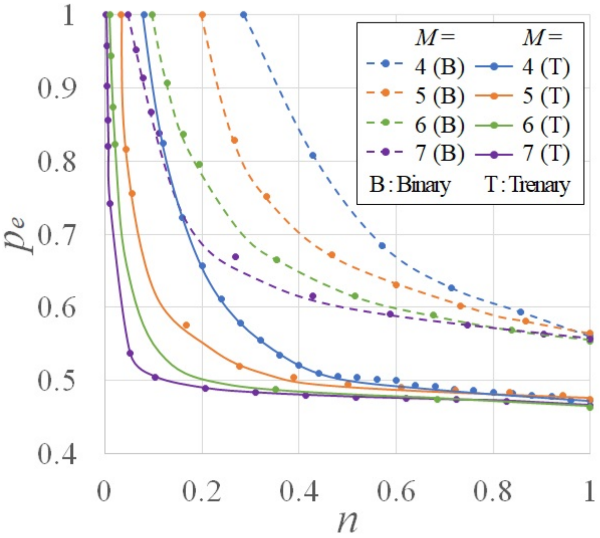

Fig. 5. Trade-off relationship between investment cost n and performance degradation p_e with scale of system M by ternary ECOC for $\mu_i = 1(i = 1, 2, \cdots, M)$, and $\sigma_{ii}^2 = 0.25, \sigma_{ij} = 0(i \neq j)$.

F. Additional Discussions

As additional discussions, we can state that:

In our previous work [9], [10], we chose the minimum code length of binary exhaustive codes, $N_{\min} = M - 1$, where we assume the trivial "modified one vs. the rest" method[6], on condition that the M row vectors must be distinct. Consequently, the code length of binary (shortened) exhaustive codes N is defined by $N_{\min} = M - 1 \leq N \leq N_{\max}$. Here, as described in Subsection V, C, $N_{\min} = \lceil \log_2 M \rceil$. However, when using binary classifiers and using ternary classifiers, both have *elastic* and *effective elastic*, and their trade off curves are almost the same. As a result, it can be said that the range of the code length of exhaustive codes has *robustness* from the view-point of system evaluation model.

VI. CONCLUSION

In this paper, it is shown that the ternary ECOC is better in the sense that it is *more flexible* than the binary ECOC from the view-point of system evaluation model by the trade-off curves. Also, both have *elastic* properties. This indicates that the probability of classification error can be drastically reduced, if the code length is relatively small. In a range of relatively small code length, it has a practical possibility. As a result, it is worthwhile to study and design the ECOC in more detail. If they are shown to be *unelastic* [16], then the scope of application of the ECOC is no longer limited.

As future works, it is necessary to apply our approach to many other multi-valued classification problems such as hand-written character recognition [18] and image classification, to show that this kind of classification problems also has desirable properties. In addition, *elastic* property means that constructive methods of ECOC can reduce the probability of classification error with a relatively small code length, which also remains to be solved using fruits of code theory.

[6] The "modified one vs. the rest" method is given by excluding the column vector $(0, 0, \cdots, 0, 1)^{\mathrm{T}}$ from the "one vs. the rest" method.

REFERENCES

[1] E. L. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: A unifying approach for margin classifiers," *Journal of Machine Learning Research*, vol. 1, pp.113–141, 2000.

[2] G. Armano, C. Chira, and N. Hatami, "Error-correcting output codes for multi-label text categorization," *Proceedings of the Third Italian Information Retrieval Workshop*, IIR 2012, pp.26–37, Italy, Jan. 2012.

[3] T. G. Dietterich and G. Bakiri: "Solving multi-class learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol.2, pp.263–286, 1995.

[4] S. Escalera, O. Pujol, and P. Radeva: "On the decoding process in ternary error-correcting output codes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, No. 32, Vol. 1, pp. 120–134, 2010.

[5] M. Goto, and M. Kobayashi, *Introduction to Pattern Recognition and Machine Learning* (in Japanese), Corona-Sha, Tokyo, 2014.

[6] S. Hirasawa, and H. Inazumi, "A system evaluation model by using information theory," *The 30th Joint National Meeting*, ORSA/TIMS, MB35.3, Philadelphia, PE. USA, Oct. 1990.

[7] S. Hirasawa, and H. Inazumi, "A model for system evaluation based on information theory," *Proceedings of the The 2000 International Conference of Management Science and Decision Making*, Tamkang University, Taipei, ROC, Jun. 2000.

[8] S. Hirasawa, T. Saito, H. Inazumi, and T. Matsushima, "System evaluation of disk allocation methods for Cartesian product files by using error correcting codes," *Proceedings of 2011 IEEE International Conference on Systems*, Man, and Cybernetics, Anchorage, Alaska, USA, pp.2443–2448, Oct. 9–12, 2011.

[9] S. Hirasawa, G. Kumoi, M. Kobayashi, M. Goto, H. Inazumi, "System evaluation of construction methods for multi-class problems using binary classirers," *Proceedings of 6th World Conference on Information Systems and Technologies*, pp.909–919, 2018.

[10] S. Hirasawa, G. Kumoi, M. Kobayashi, M. Goto, H. Inazumi, "System evaluation of error correcting output codes for artificial data methods," *Proceedings of 2018 International Conference of Engineering, Technology, and Applied Science*, pp.112–122, 2018.

[11] H. Inazumi, *Studies on the Evaluations for Information Systems based on Rate Distortion Theory*, Dissertation of Dr. Eng., Waseda University, Nov. 1989.

[12] H. Inazumi, M. Kochiya, and S. Hirasawa, "On the trade-offs between the file redundancy and the communication costs in distributed database systems," *IEEE Trans. SMC*, vol. SMC-19, no. 1, pp.108–112, Jan. 1989.

[13] G. Kumoi, H. Yagi, M. Goto, and S. Hirasawa, "Binary Codeword Table for Multilevel Document Classification Using Information Theoretic Criterion of Binary Discriminations (in Japanese)," *Transactions on Mathematical Modeling and its Applications (TOM)*, 2019. (Under Review)

[14] Y. Luo, and K. Najarian, "Employing decoding of specific error correcting codes as a new classification criterion in multiclass learning problems," *Proceedings of 2010 International Conference on Pattern Recognition*, pp.4238–4241, 2010.

[15] S-H. Park, and J. Frnkranz, "Efficient decoding of ternary error-correcting output codes for multiclass classification," *Technical Report TUD-KE-2009-01*, *Technische Universitt Darmstadt*, pp. 1–20, 2009.

[16] J. Pearl, and A. Crolotte, "Storage space versus validity of answers in probabilistic question answering systems," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp.633–640, Nov. 1979.

[17] Yomiuri News Paper Articles 2015, *Nichigai Associates Inc.*

[18] UCI Machine Learning Repository, URL: http://www.trifields.jp/uci-machine-learning-repository-dataset s-956G.