2019年電子情報通信学会ソサイエティ大会
2019年9月10日（火）～13日（金），大阪大学　豊中キャンパス，大阪

2 値ECOCと3 値ECOCの システム評価

早稲田大学　平澤　茂一
早稲田大学　雲居　玄道
電気通信大学　八木　秀樹
早稲田大学　小林　学
早稲田大学　後藤　正幸
青山学院大学　稲積　宏誠

目次

1　まえがき
　　システム評価モデル
2　Exhaustive 符号と短縮Exhaustive 符号
3　実験
4　考察とむすび・・・正規化した平均的性能

1 まえがき

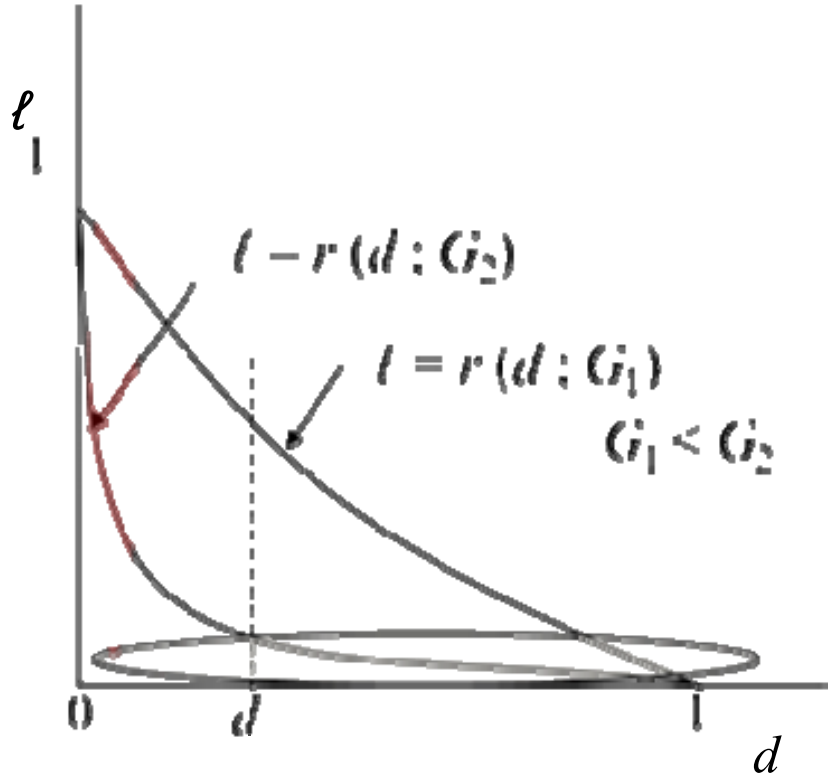
- 質問回答システム(QAシステム) J. Pearl and A.Crolotte [3]
- 情報縮約論
  - –「フレキシブル(Flexible)」,
  - –「エラスティック(Elastic)」
- **システム（トレードオフ）評価モデル**[4]

- 適用事例[3]

[本研究]

- 対象：「2値/3値判別器による多値判別問題」
  - **ECOC** (Error Correcting Output Codes) [1]
- **分類誤り確率P_eを最小化することが目的ではない.**

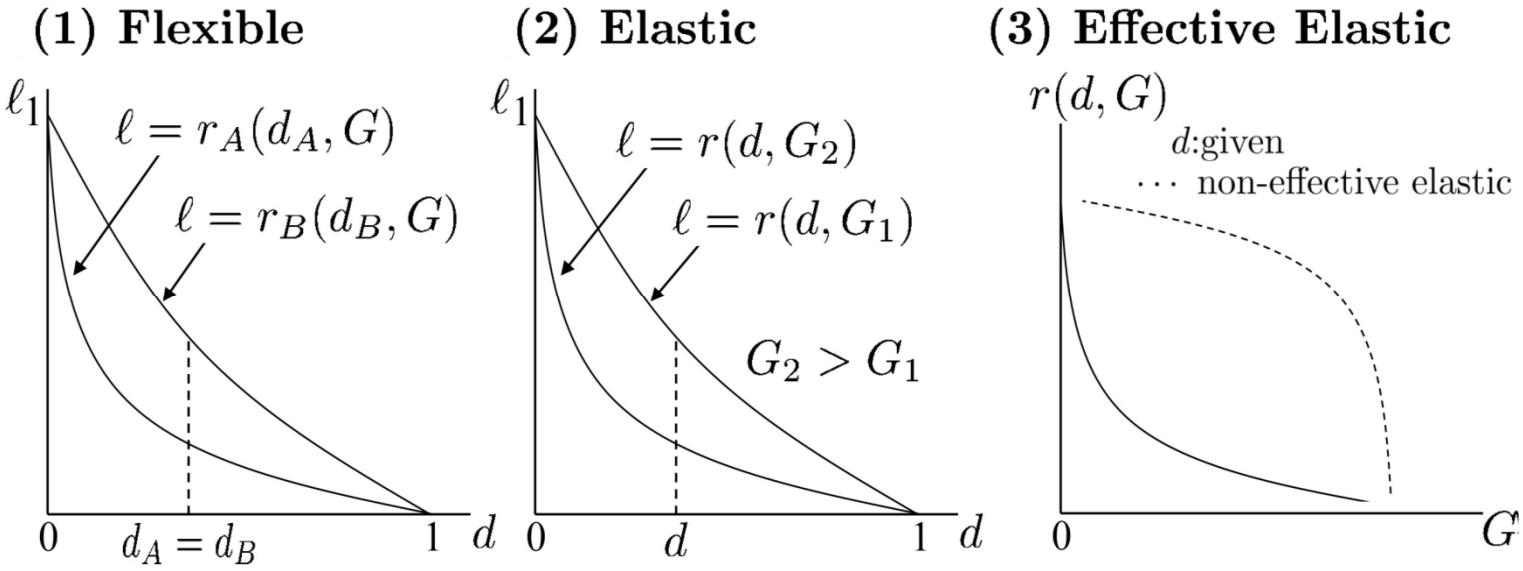
Flexible性・・・ほんの僅かな歪（性能劣化）を許容すれば,
レート（投資コスト）を大幅に減少できる.
Elastic性・・・システムサイズが大きくなればFlexible性が増す.



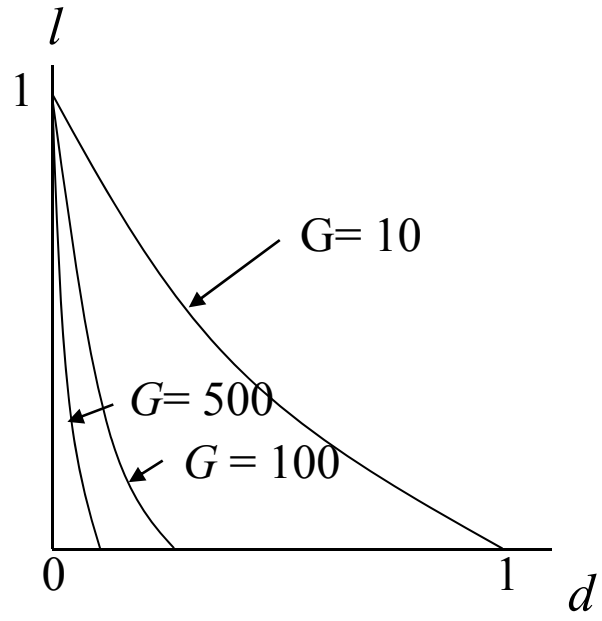
レート歪関数・・・トレードオフ評価モデル

トレードオフ評価モデル[3][4]

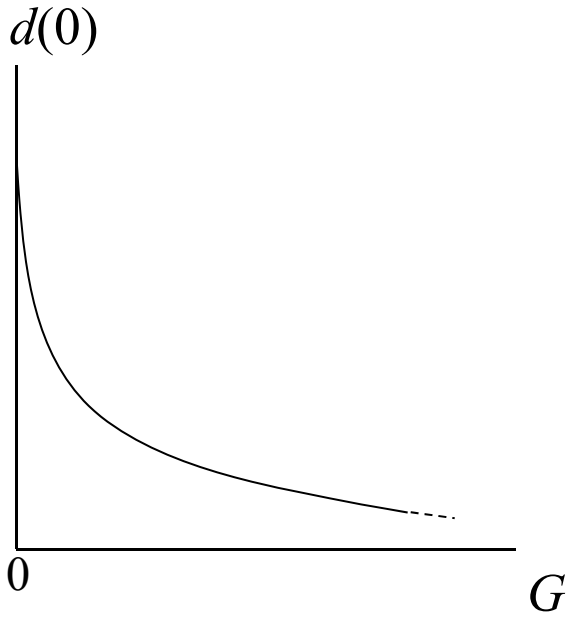
- 情報縮約論：レート歪関数　$L= R\,(D)$ (A.1)
- システム評価モデル：　$l = r\,(d\,;\,G)$ (A.2)
- $L\,,(l\,)$：レート，（投資コスト）
- $D\,,(d\,)$：歪，（性能劣化）
- G：システムの規模



(4) Trivial Elastic



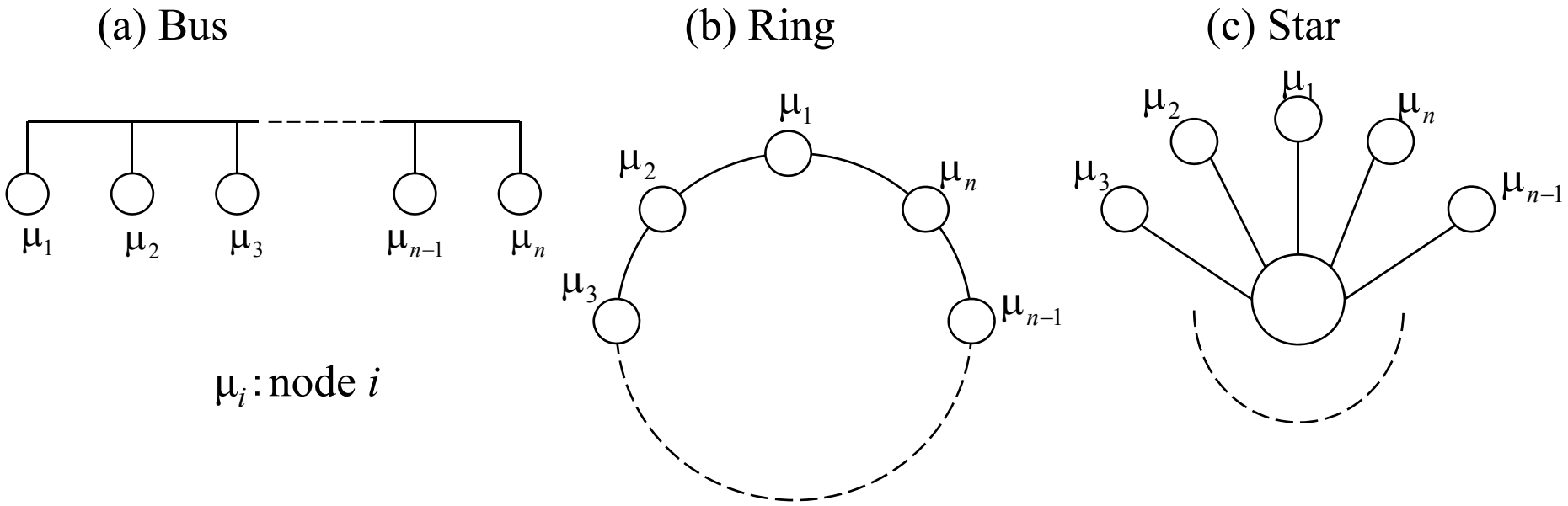
(5) Marginal Elastic



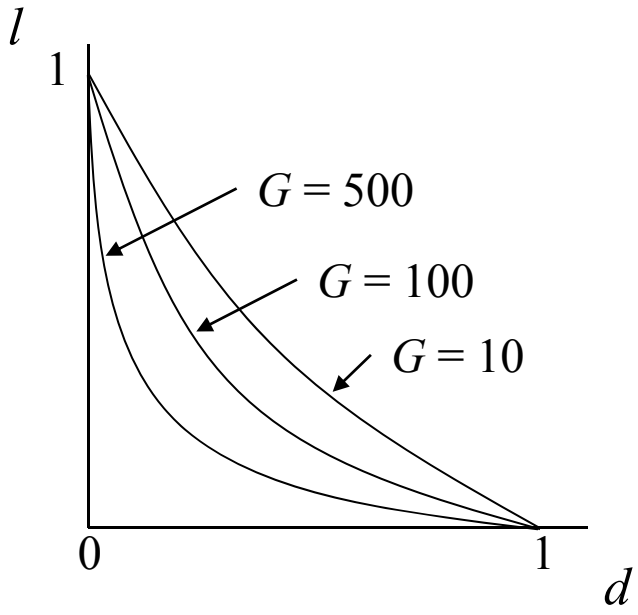
[例1]　分散データベース

- l: ファイルの冗長性
- d: ファイルのアクセスコスト
- G: ノード数(n)

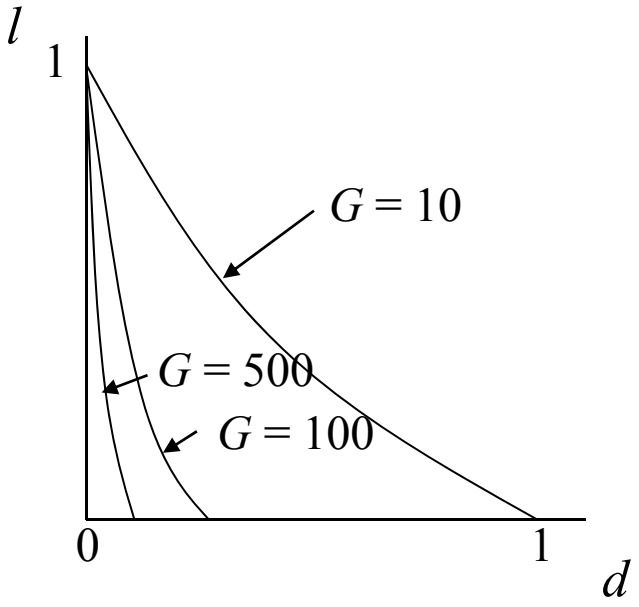
ネットワークトポロジー



(a) Bus, (b) Ring : **Elastic**



(c) Star : **Trivial Elastic**



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.

表A.1. 多値分類システムの評価（対応表）

情報縮約論	システム評価モデル	多値分類システム
レート (L)	投資コスト(l)	(2値/3値)判別器数 (n)
歪 (D)	性能劣化(d)	分類誤り確率 (p_e)
	規模(G)	カテゴリ数 (M)

2. Exhaustive 符号と短縮Exhaustive 符号

符号語表 （code word table）: $W=[w_{ij}],\ i=1,2,\ldots,M,\ j=1,2,\ldots,N$

M：カテゴリ数（符号語数）

N：弁別器数（符号長）

[例2] 符号語表

表A.2. 符号語表の例（$M=4, N=4$）

(a) 2値 one vs. the rest 法

	d_1	d_2	d_3	d_4
c_1	1	0	0	0
c_2	0	1	0	0
c_3	0	0	1	0
c_4	0	0	0	1

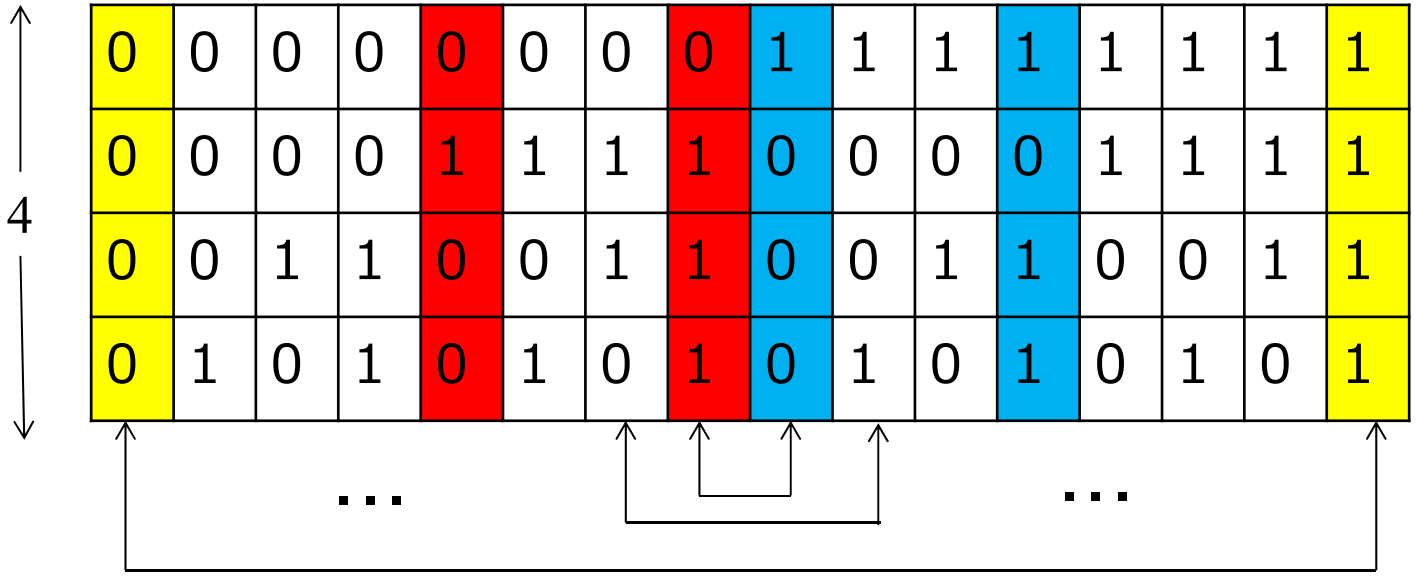
(b) 3値 one vs. one 法

	d_1	d_2	d_3	d_4
c_1	1	*	*	0
c_2	0	1	*	*
c_3	*	0	1	*
c_4	*	*	0	1

$$w_{ij}=\begin{cases} 0, & \text{負例} \\ *, & \text{don't care} \\ 1, & \text{正例} \end{cases}$$

2値Exhaustive 符号の生成法

[例3] $M = 4$



全てのM[ビット]の列ベクトル $\boldsymbol{d}_j$ (j=1,2,…, 2^M) を生成し，その中から
① 補元を取り除く.
② その後，全1（または全0）の列ベクトルを取り除く.

$$N = 2^{M-1} - 1 \qquad \text{(A.3)}$$

[例4] 2値と3値のExhaustive 符号（$M=4$）

表A.3．2値 Exhaustive 符号（$M=4.\ N_{\max}=7$）

	d_1	d_2	d_3	d_4	d_5	d_6	d_7
c_1	1	1	1	1	1	1	1
c_2	0	0	0	0	1	1	1
c_3	0	0	1	1	0	0	1
c_4	0	1	0	1	0	1	0

表1．3値Exhaustive 符号（$M=4.\ N_{\max}=25$）

	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}	d_{16}	d_{17}	d_{18}	d_{19}	d_{20}	d_{21}	d_{22}	d_{23}	d_{24}	d_{25}
c_1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	*	*	*	1	1	1	*	*	*
c_2	0	0	0	0	1	1	1	0	1	1	0	1	1	*	*	*	0	0	0	0	*	*	1	1	*
c_3	0	0	1	1	0	0	1	1	0	1	*	*	*	0	1	1	0	1	1	*	0	*	0	*	1
c_4	0	1	0	1	0	1	0	*	*	*	1	0	1	1	0	1	1	0	1	*	*	0	*	0	0

$$2\text{ 値}：N_{\max}=2^{M-1}-1, \tag{1}$$

$$3\text{ 値}：N_{\max}=\sum_{\alpha=0}^{M-2}\binom{M}{M-\alpha}(2^{M-1-\alpha}-1), \tag{2}$$

短縮Exhaustive 符号の構成と性能評価

(1) 短縮Exhaustive符号
Mが与えられたとき

$$N_{\min} \leqq N \leqq N_{\max} \qquad \text{(A.4)}$$

となるような短縮Exhaustive符号を構成する．ここで，$N_{\min} = [\log_2 M]$とし，$P_{e,\max} = S(N_{\min}, M)$である．

(2) Exhaustive符号の2^{M-1}-1個の列ベクトルからランダムに　(2値判別器数)N個を選択し，その時の分類誤り確率P_eを求める．ここで，次式を実験的に求める．

$$P_e = S(N, M) \qquad \text{(A.5)}$$

(3) システム（トレードオフ）評価関数は，$n = N/N_{\max}$，$p_e = P_e/P_{e\max}$と正規化して次式で与えられる．

$$p_e = s(n, M) \qquad (3)$$

，

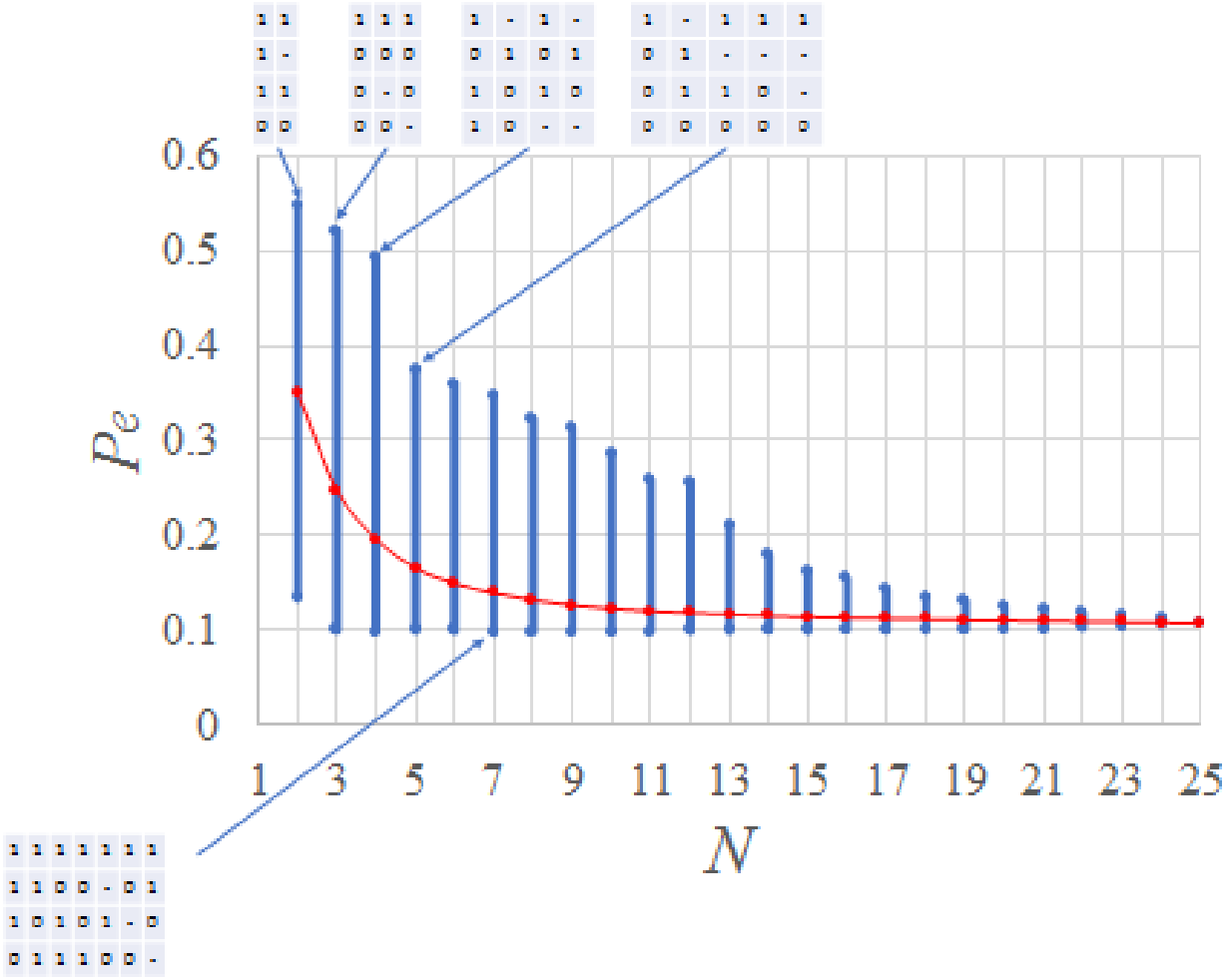
3. 実験

3.1 実験データ

表A.2. ベンチマークデータ（2015年読売新聞記事[5]）

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)

3.2　実験結果



図A.2.　3値ECOCにおける符号長Nと平均分類誤り確率P_eのトレードオフ関係 (M=4)

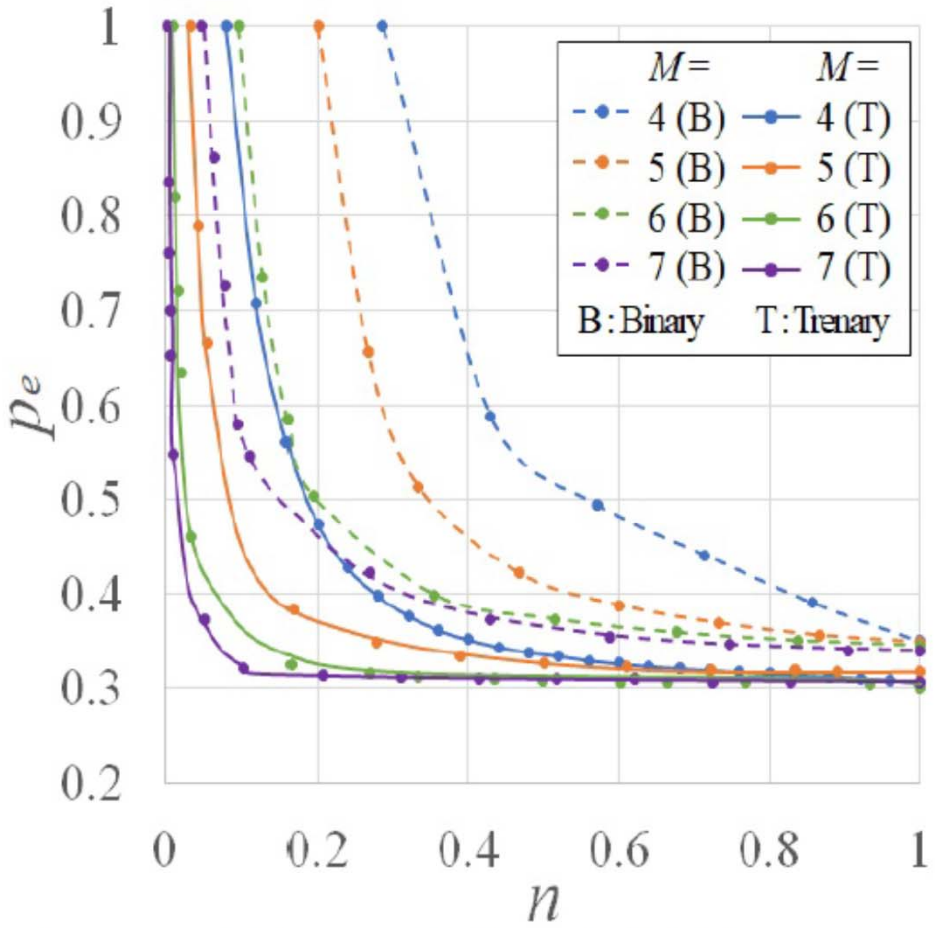


図 1　投資コスト割合 n と性能の劣化割合 p_e のトレードオフ関係（パラメータ：システムの規模 M）.

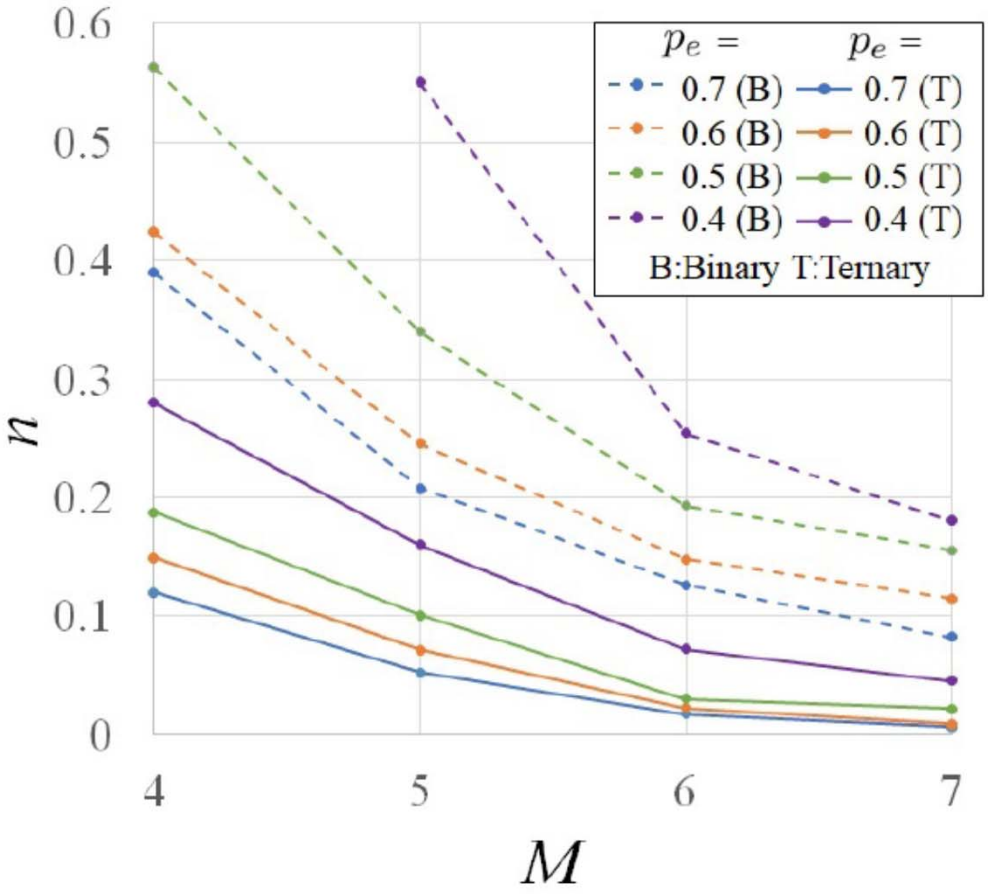
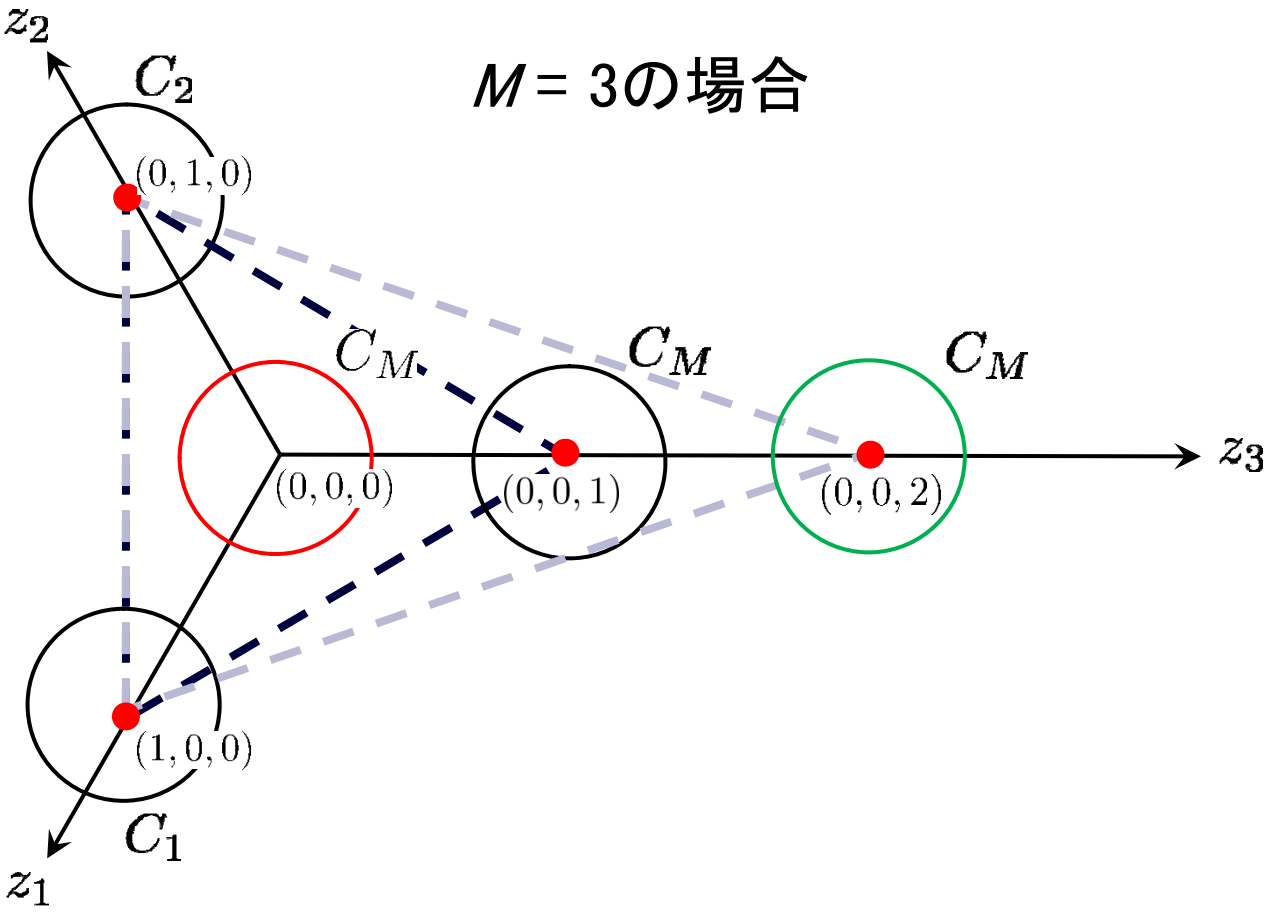


Fig. 4. Relationship between scale of system M and investment cost n by ternary ECOC using Benchmark Data [17].



[実験1] 等方性のとき
$\mu_M = 1.0$ の場合

[実験2] 平均が異なるとき
① μ_M=0.0,とした場合

[実験3] 平均が異なるとき
② μ_M=2.0,とした場合

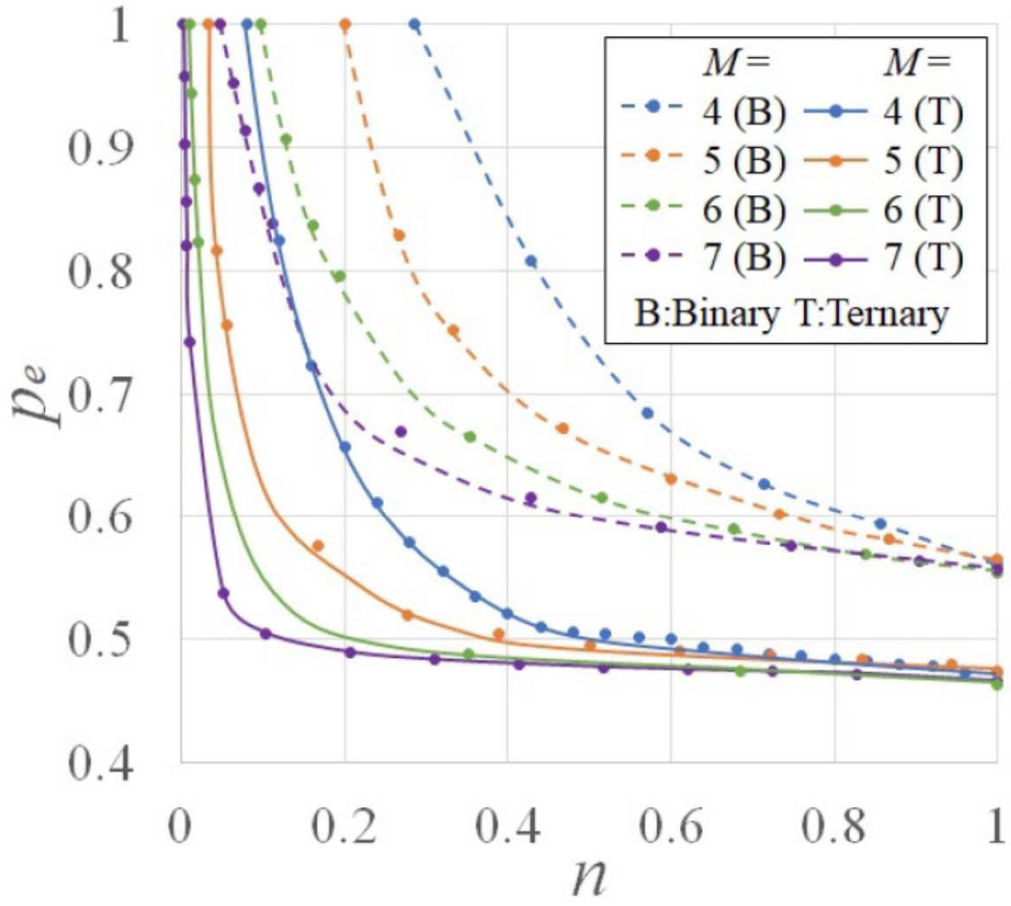


Fig. 5. Trade-off relationship between investment cost n and performance degradation p_e with scale of system M by ternary ECOC for $\mu_i = 1 (i = 1, 2, \cdots, M)$, and $\sigma_{ii} = 0.25, \sigma_{ij} = 0 (i \neq j)$.

4. 考察とむすび

$M = 4, 5, 6, 7$ に対し，図 1 よりトレードオフ曲線から

1. 全ての ECOC は Flexible である．
2. 与えられた M に対し，3 値 ECOC の方が 2 値 ECOC より，More Flexible である．
3. n=1 の近辺を除き，2 値 ECOC と 3 値 ECOC ともに Elastic ある．
4. 2 値 ECOC と 3 値 ECOC ともに Effective Elastic である（詳細は [6] 参照）．

以上より ECOC は情報システムの要素として，特に M が大きくなるとき望ましい性質を持っており，さらに深く詳細に検討する価値がある．