



Press RELEASE ご取材案内

配布先：文部科学記者会、科学記者会

報道各社：科学部・社会部 ご担当者各位

2023年2月2日

学校法人 早稲田大学

国立大学法人 電気通信大学

2値分類器の結合による多値分類法の理論的解析に成功 深層学習モデルにおける性能向上に期待

発表のポイント

- 機械学習分野において、2値分類器を組み合わせて多値分類器を構築する Error-correcting output coding (ECOC)は実験的に良い性能を示すことが知られていましたが、その理由は理論的に明らかになっていませんでした。
- 本研究では2値分類器の推定性能に着目し、その推定誤差に対して頑健な ECOC の枠組みを数理的にモデル化することで、分類誤り率を最小とする最大事後確率分類に限りなく近づける方法を理論的に明らかにしました。
- 近年脚光を浴びている深層学習モデルに本研究成果を適用することにより、その性能向上に大きく寄与することが期待されています。

早稲田大学データ科学センター講師 雲居 玄道（くもい げんどう）、電気通信大学大学院情報理工学研究科准教授 八木 秀樹（やぎ ひでき）、および早稲田大学データ科学センター教授 小林 学（こばやし まなぶ）、早稲田大学理工学術院創造理工学部教授 後藤 正幸（ごとう まさゆき）、早稲田大学名誉教授 平澤 茂一（ひらさわ しげいち）による共同研究グループは、2値分類器を組み合わせて多値分類器を構築する Error-correcting output coding (以下、「ECOC」)の枠組みを数理的にモデル化し、ECOC が良い分類性能を示すための、2値分類器の組み合わせの条件を理論的に明らかにしました。

本研究成果は、シンガポール・World Scientific 社発行の『International Journal of Neural Systems, Vol. 33, No. 2 (2023)』に、論文名 “Performance Evaluation of Error-Correcting Output Coding Based on Noisy and Noiseless Binary Classifiers” として掲載されました。

(1) これまでの研究で分かっていたこと（科学的・歴史的な背景など）

機械学習分野において、データに紐づけられた内容を表すカテゴリを学習し、新しい入力データのカテゴリを推定する分類問題は大きなテーマの1つです。対象とするカテゴリ数が2つのものを2値分類問題、3つ以上のものを多値分類問題と呼びます。このカテゴリ数が多くなるほど問題が複雑になるため、分類問題を解決するには多くの困難があります。

この多値分類問題を解決する方法の1つに、2値分類問題の組み合わせと捉える ECOC が1994年に T. G. Dietterich らによって示されました（図1）。ECOC は符号理論における誤り訂正の考え方を援用した方法で、複数の2値分類器を用いることで、いくつかの2値分類器の出力が間違っても正しいカテゴリを推定することができるという仕組みです。

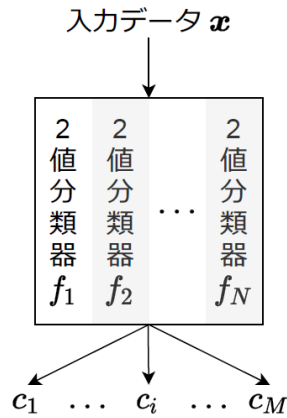


図 1 : Error-correcting output coding

そのため、多値分類を 2 値分類として様々なパターンに 2 値変換して 2 値分類器を構築します。例えば、カテゴリ数が 4 の場合、最も単純な方法は、2 値変換に 4 つのカテゴリを 1 対 3 に分割する 4 種類の 2 値分類器を用いる方法で One-vs-Rest と呼ばれます。これ以外に、2 対 2 に分ける 2-vs-Rest が考えられ、最大で 7 種類の 2 値分類器を用いることが可能です (図 2)。この最大の 2 値分類器数をもつ Exhaustive Code から良い性能を示す 2 値分類器の組み合わせを選択することが ECOC の大きな課題です。そのため、多くの研究が成され、様々なデータに対して ECOC が良い性能を示すことは実験的に知られています。しかし、実験条件を変えれば良い性能を示す 2 値分類器の組み合わせが変わることから、2 値分類器の組み合わせ方法の明確な答えは示されていませんでした。

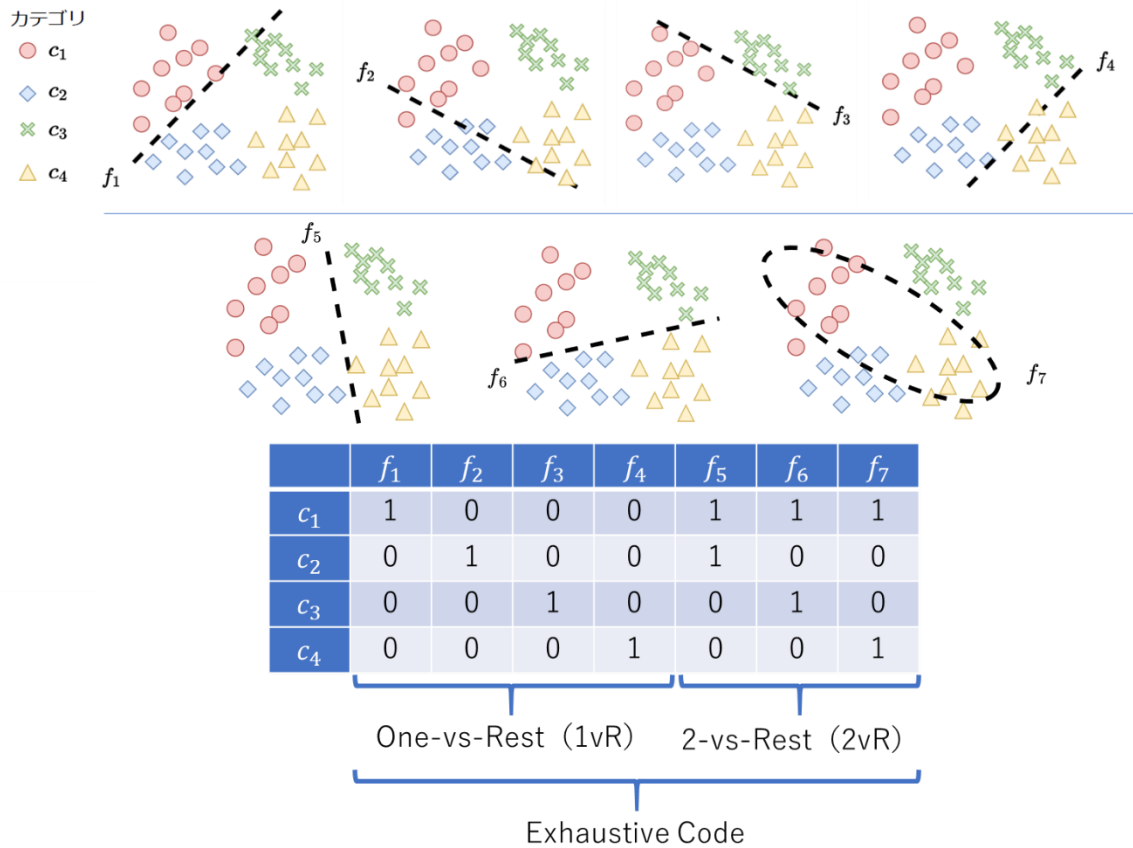


図 2 : カテゴリ数が 4 つの場合の 2 値分類器の組み合わせ例

(2) 今回の研究で新たに実現しようとしたこと、明らかになったこと

本研究では、よい性能を示す2値分類器の組み合わせを明らかにするべく、2値分類器の推定性能に着目しました。一般的にデータから分類規則を学習した2値分類器は、データの不足などから推定誤差が生じると考えられます。この推定誤差に対して頑健な2値分類器を組み合わせるための条件を明らかにしました。この結果、2値分類器の構成を示す符号語表(※1)に対し、符号語間のハミング距離(※2)を大きくすれば、2値分類器の誤差の分散が大きくなったとしても分類誤り率を最小とする最大事後確率分類(※3)に限りなく近づくことを明らかにしました(図3)。

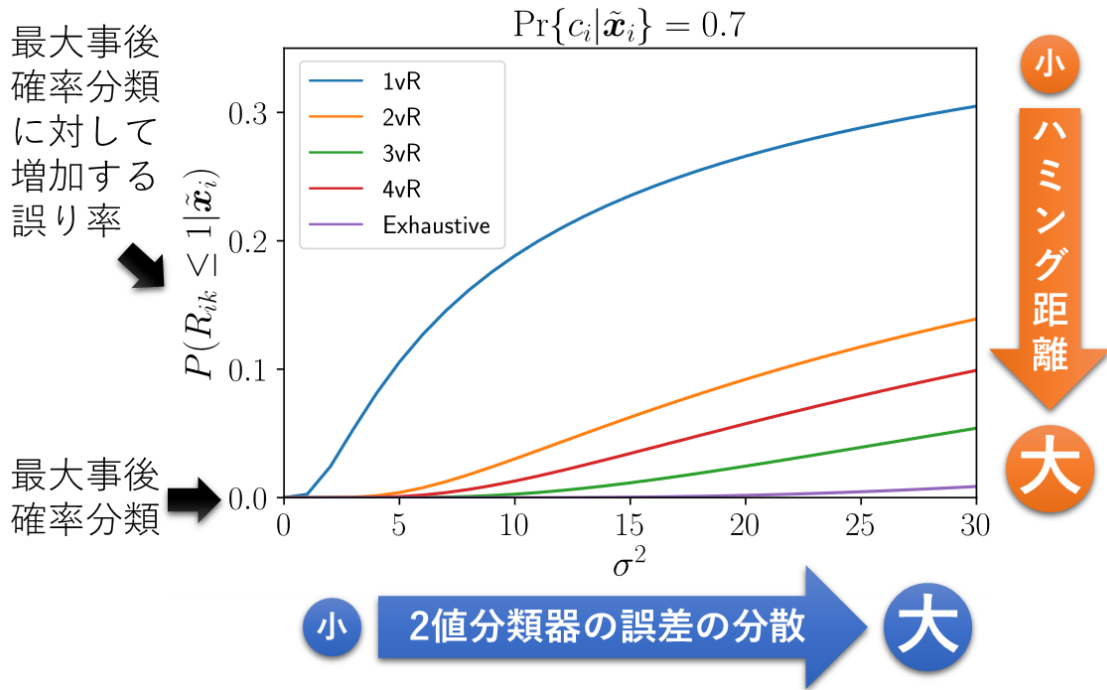


図3: 定理に基づく解析結果(カテゴリ数8の場合)(※4)

(3) そのために新しく開発した手法

従来、ECOCの理論的解析が困難であった理由は、分類問題がもつ複雑性が要因であると考えられます。この原因は、データに含まれるノイズと2値分類器の推定誤差であるノイズが混在しているため、実験では実験条件に依存した結果しか得られていなかったことと考え、この2種のノイズを仮定した数理モデルを新たに考案しました(図4)。

この数理モデルに基づき、本研究では2値分類器のノイズに統計的な性質を仮定し、符号語間のハミング距離を用いてカテゴリ間の分類誤り率を解析する手法を開発しました。

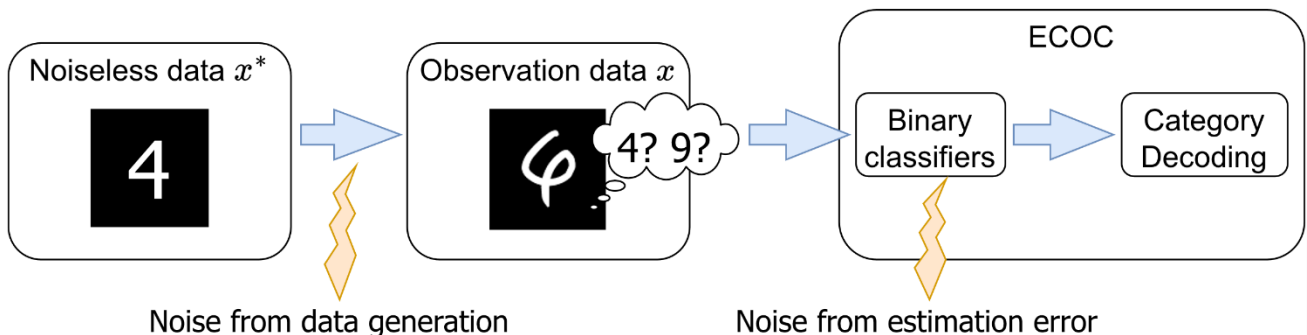


図4: ECOC ノイズモデル

(4) 研究の波及効果や社会的影響

本研究は、実データに対する実験結果から性能が良いとされてきた ECOC に対し、理論的な保証を与えるものです。そのため、ECOC を実応用に適用する際に、想定される 2 値分類器の誤差や計算機環境に応じて、本研究の成果を用いて組み合わせる 2 値分類器を決定することが可能になります。

さらに、近年脚光を浴びている深層学習モデルも分類問題に適用する場合、多くの場合で出力層は図 2 に示した One-vs-Rest の形をとります (図 5)。そのため、深層学習モデルも ECOC の枠組みで捉えることが可能になることから、数多の深層学習モデルに対して、本研究の知見を適用することで性能向上をはかれる可能性があります。

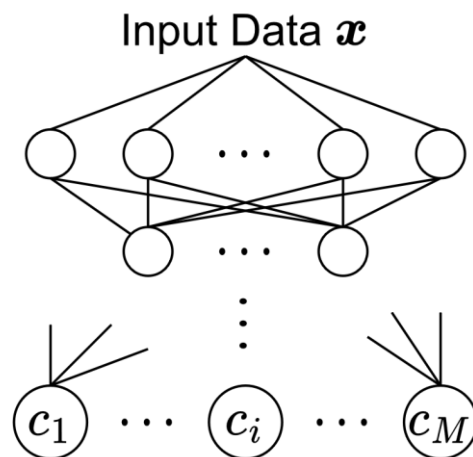


図 5: 分類問題における深層学習を用いた構成

(5) 今後の課題

本研究で示した理論的解析には、さまざま仮定を設定しています。そのため、ECOC における全てが明らかになったわけではありません。今後は、設定した仮定による制約を緩めるなど、より広い範囲で性能解析が可能な数理モデルを構築していきたいと考えております。

(6) 研究者のコメント

複雑な分類問題の研究に新たな解析手法を提案しました。この研究は、理論的解析が困難であったこの分野に一石を投じるものです。この研究成果は、石の上にも三年といった状況でようやく達成されました。この経験から、何事も軽々しく判断せず、物事の本質をしっかりと見極めることが大切であることを学びました。

(7) 用語解説

※1 符号語表

ECOC において 2 値分類器の構成を示す表のことです。例えば、図 1 に示す表のことであり列が 2 値分類器を意味します。行は各カテゴリに対応付けられており、各行が符号語と呼ばれます。

※2 ハミング距離

符号語間の 0,1 の異なっている数のこと。例えば、図 1 の符号語表のハミング距離は、すべての符号語間で 4 となります。

※3 最大事後確率分類

入力データ x に対して、各カテゴリの真の事後確率 $\Pr\{c_i|x\}$ の最大値をとるカテゴリに分類することです。

※4 Exhaustive

Exhaustive Code とは、構成可能な全ての 2 値分類器を用いたものです。図 3 に示したカテゴリ数 8 の場合は、1vR, 2vR, 3vR, 4vR に含まれる 2 値分類器を全て用いたものです。

(8) 論文情報

雑誌名 : International Journal of Neural Systems

論文名 : Performance Evaluation of Error-Correcting Output Coding Based on Noisy and Noiseless Binary Classifiers

著者名 : 雲居 玄道 (早稲田大学データ科学センター講師 (任期付))、
八木 秀樹 (電気通信大学大学院情報理工学研究科准教授)、
小林 学 (早稲田大学データ科学センター教授)、
後藤 正幸 (早稲田大学理工学術院創造理工学部教授)、
平澤 茂一 (早稲田大学名誉教授)

掲載日時 (現地時間) : 2023 年 1 月 9 日 (月) ※オンライン

掲載 URL : <https://www.worldscientific.com/doi/10.1142/S0129065723500041>

DOI : 10.1142/S0129065723500041

(9) 研究助成

研究費名 : 日本学術振興会 (JSPS) 科学研究費 (基盤研究 C)

研究課題名 : ビジネス価値創造のためのデータ解析プラットフォームと時変協調フィルタリングの研究

研究代表者名 (所属機関名) : 小林学 (早稲田大学)

研究費名 : 日本学術振興会 (JSPS) 科学研究費 (基盤研究 C)

研究課題名 : 高信頼・高安全な符号化システムの実現に向けた理論展開

研究代表者名 (所属機関名) : 八木秀樹 (電気通信大学)

【研究内容に関するお問い合わせ先】

早稲田大学データ科学センター 講師 (任期付) 雲居 玄道

Tel: 03-3204-9164 E-mail: kumoi@aoni.waseda.jp

電気通信大学 大学院情報理工学研究科 准教授 八木 秀樹

Tel: 042-443-5366 E-mail: h.yagi@uec.ac.jp

【発信元】

早稲田大学広報室広報課

Tel: 03-3202-5454 E-mail: koho@list.waseda.jp

電気通信大学総務企画課広報係

Tel: 042-443-5019 E-mail: kouhou-k@office.uec.ac.jp