

平成 21 年 4 月 21 日現在

研究種目：若手研究 (B)

研究期間：2007 ~2008

課題番号：19700264

研究課題名 (和文) アンサンブル学習による新たなアルゴリズムと情報論的解析による洗練、及びその応用

研究課題名 (英文) Development, sophistication and applications of new ensemble learning algorithms based on information theoretic view point.

研究代表者 竹之内 高志 (Takenouchi Takashi)

奈良先端科学技術大学院大学 情報科学研究科 助教

研究者番号：50403340

研究成果の概要：本研究では近年注目を集めている2種類のアンサンブル学習、ブースティングとECOCに基づく多値判別に関する研究を行った。

(1) ブースティングに関しては統計的考察に基づく最もロバストなロス関数の考案とミスラベルを考慮した多値判別アルゴリズムを提案し、アルゴリズムの収束性などに関する考察を行った。パターン識別問題においては例題に外れ値などが含まれることが多く、このような状況に対してロバストなアルゴリズムは有効に働く。

(2) ECOCに基づく多値判別問題に関しては、確率モデルを用いて複数の2値判別器を組み合わせて多値判別器(デコーダー)を構成するための枠組みを提案した。提案した枠組みにおいては各2値判別器同士の相関関係を考慮してデコーダーを構成することができる。また提案手法では確率モデルをベースとして推定を行っており、対象とするデータに対して適応的にデコーダーを構成することができる。確率モデルをベースとしたことで、従来の手法に用いられている様々な拡張や正則化を容易に適用出来ることが提案手法の利点であり、提案手法は従来の最新の多値判別手法を優越する性能を示す事がわかった。

交付額

(金額単位：円)

	直接経費	間接経費	合計
19年度	1,200,000	0	1,200,000
20年度	1,300,000	390,000	1,690,000
年度			
年度			
年度			
総計	2,500,000	390,000	2,890,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：分類, パターン認識, ブースティング

1. 研究開始当初の背景

近年の計算機のめざましい進歩により、従来では解くことの出来なかった超大規模な最適化問題に注目が集まり、現実の問題への適用

も徐々になされつつあった。その様な超大規模問題へのアプローチとして、問題を比較的解きやすい小規模な問題に分割し、結果を統合(分割統治)する方法、アンサンブル学習

に大きな関心が寄せられていた。その中の中心的な2つのアプローチ、ブースティングと誤り訂正符号(ECOC)に基づく多値判別法に着目した。

2. 研究の目的

(1)ブースティングを用いた判別において、典型的なアルゴリズムとしてアダブーストが有名であるが、このアルゴリズムは訓練に用いた例題にオーバーフィットしてしまい、未知の例題に対する汎化性能が悪化する場合がありますが知られており、何らかの正則化が必要であった。ただし交差検証法などによる途中停止では計算量的な問題もあり、より簡易にこの問題に対応できる手法を構築する事を目指した。

(2)ECOCに基づく多値判別法は、多値判別問題を複数の2値判別問題に分割し、各々学習した2値判別器を組み合わせる事で多値判別(デコーダー)を行う方法であるが、従来の方法においては各2値判別器をどのように組み合わせが事前に固定されている場合が多く、うまく働く場合とそうでない場合の差が大きかった。この問題に対処するため、確率モデルに基づくデータ適応的な組み合わせ法の提案を目指した。

3. 研究の方法

(1)ブースティングに関しては、判別問題の例題において本質的に分離できない部分や、ラベルのつけ間違い(ミスラベル)に対してアルゴリズムが過学習してしまう事が問題であった。そこでラベル反転を考慮した確率モデルに基づく新しいブースティングアルゴリズムの構築を目指した。また提案したアルゴリズムの統計的性質についての解析や情報幾何を用いた解析を行った。

(2)ECOC判別に関しては、2値判別器の判別誤りを、情報送信を行う際のビット反転ノイズと同様のものとみなして確率モデルを用いて定式化した。従来の通信路モデルは2元対称通信路、つまり判別問題に対応させて考えると2種の誤り(False Positive Error, False Negative Error)が等しいという仮定の下に成り立っているが、そもそも多値判別問題においてはこの2種の誤りの対称性は大きく崩れる。このような非対称的なノイズを考慮した確率モデルの構築によって従来の手法と比較

して大きな性能の向上を図ることを試みた。また各2値判別器は訓練の過程でデータを共有しているため、各2値判別器を組み合わせ最終的な結果を得る過程において、各々の関係性を記述できる確率モデルの構築を試みた。

4. 研究成果

(1)ブースティングに関して：前述の確率モデルに基づく多値判別のためのブースティングアルゴリズムを提案し、その統計的性質を明らかにした。ベンチマークデータを用いた実験においても既存の手法を上回る性能を発揮した。この結果は機械学習の分野におけるトップジャーナル Neural Computationに掲載された。

ミスラベルを考慮したブースティングアルゴリズムにより、過学習に強いロバストな判別関数の構成が可能になったが、情報量基準による停止基準の提案には至らなかったため、今後は適切な停止基準の提案を目指していきたい。

ECOC判別に関して：提案した確率モデルを用いてデータ適応的にデコーダーを構成するための手法を確立した。提案法は確率モデルに基づくため、従来の手法に用いられている様々な拡張や正則化を容易に適用出来る。また各2値判別器間の関係性を考慮する事で性能が向上し、機械学習、バイオインフォマティクスなどの各種ベンチマークデータに対して、従来の最新の多値判別法を大きく上回る性能を発揮している。この結果をまとめた論文はブースティングの結果と同じく機械学習のトップジャーナル Neural computation に採択され掲載予定である。

現段階で基礎的な手法を確立し精度の良い判別器を構成する事ができている。今後はさらなる性能向上のために提案モデルへ局所化、正則化などを適用し、現実の問題への適用も視野に入れて実用化を目指していきたい。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 8 件)

1. T. Takenouchi, and S. Ishii.

A multi-class classification method based on decoding of binary classifiers.

Neural Computation, in press.

査読有.

2. J. Hirayama, M. Nakatomi, T. Takenouchi and S. Ishii.

Collaborative prediction by multiple Bayesian networks and its application to printer usage modeling.

Behaviormetrika, 35(2), pp. 99-114, 2008.
査読有.

3. J. Hirayama, M. Nakatomi, T. Takenouchi and S. Ishii.

α -Bayesian collaboration of multiple predictors and its applications to hybrid recommendation and user modeling.

Neural Information Processing Letters and Reviews, 12(1-3), pp. 11-20, 2008.

査読有.

4. T. Takenouchi, S. Eguchi, N. Murata, and T. Kanamori.

Robust boosting algorithm against mislabeling in multi-class problems.

Neural Computation, 20(16), pp. 1596-1630, 2008.

査読有.

5. S. Osaga, J. Hirayama, T. Takenouchi and S. Ishii.

A probabilistic modeling of MOSAIC learning

Artificial Life and Robotics, 12(1-2), pp. 167-171, 2008.

査読有.

6. N. Yukinawa, T. Takenouchi, S. Oba and S. Ishii.

Combining multiple decisions: applications to bioinformatics.

Journal of Physics: Conference Series 95 012018 (13pp) doi:

10.1088/1742-6596/95/1/012018, 2008.

査読有.

7. Kanamori, T., Takenouchi, T., Eguchi, S., and Murata, N.

Robust Loss Functions for Boosting.

Neural Computation, 19, pp. 2183-2244, 2007.

査読有.

8. Tanabe, A., Fukumizu, K., Oba, S., Takenouchi, T. and Ishii, S.

Parameter Estimation for von Mises-Fisher distributions.

Computational Statistics, 22(1), pp. 145-157, 2007.

査読有.

[学会発表] (計 15 件)

1. 竹之内高志.

ECOCに基づく多値判別法とその拡張
大阪市立大学数学研究所 情報幾何研究集会
2009 (OCAMI meeting on "Information
Geometry 2009").

2009.1.25, 大阪市立大.

2. T. Takenouchi.

Ternary Bradley-Terry model-based decoding for multi-class classification.

IEEE International Workshop on Machine Learning For Signal Processing.

2008.10.18, Cancun, Mexico.

3. 竹之内高志.

ブースティングによるパターン識別.

日本鉄鋼協会秋季講演大会.

2008.9.24, 熊本大学.

4. 大佐賀智.

ダイナミクスのベイズ推定に基づくマルチ
モジュール型適応制御モデル.

日本神経回路学会第 18 回全国大会, pp.
68-69, P1-22.

2008.9.24, 産業技術総合研究所.

5. 大山賀己.

ECOC 復号法に基づく階層的多値判別法.

ニューロコンピューティング研究会 (NC) .

2008.3.13, 玉川大学.

6. A. Ishihara.

An unified approach to collaborative and feature-based recommendation based on probabilistic latent semantic models.

International Symposium on Artificial Life and Robotics (AROB 13th '08), GS17-3.

2008. 1. 30, B-Con Plaza, Beppu, Oita, Japan

.

7. 石原愛美.

アスペクトモデルに基づいた特徴ベース協
調フィルタリング.

人工知能学会研究会資料, SIG-FPAI-A703,

pp. 59-63.

2008. 1. 17, 北海道大学.

8. 山下哲矢.

1 分子蛍光イメージングに対する尤度比検
定を用いたアプローチ.

ニューロコンピューティング研究会 (NC) .

2008.1.15, 玉川大学.

9. T. Takenouchi.

A multi-class classification with a probabilistic localized decoder.

IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), pp. 861-865. 2007.12.17, Cairo, Egypt.

10. Hirayama, J.
Bayesian collaborative predictors for general user modeling tasks.
International Conference on Neural Information Processing (ICONIP).
2007. 11.16, Kita-Kyuushu, Japan.

11. S. Ishii.
Combining multiple decisions: applications to bioinformatics.
The International Workshop on Statistical-Mechanical Informatics, pp. 151-161.
2007.9.18, Paruru Plaza, Kyoto, Japan.

12. 竹之内高志.
ECOCに基づく多値判別法の局所化.
2007 年度統計関連学会連合大会, pp.158.
2007.9.8, 神戸大学.

13. T. Takenouchi.
A probabilistic decoding approach to multi-class classification.
Interantional Joint Conference on Neural Networks (IJCNN), pp. 2671-2676.
2007.8.14, Orlando, Florida, USA.

14. T. Takenouchi.
Multiclass classification as a decoding problem.
The First IEEE Symposium on Foundations of Computational Intelligence(FOCI'07), pp.470-475.
2007. 4.4, Hawaii, USA.

15. S. Osaga.
A probabilistic model of MOSAIC.
The First IEEE Symposium on Foundations of Computational Intelligence (FOCI'07), pp.41-46.
2007. 4.2, Hawaii, USA.

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

名称 :
発明者 :
権利者 :
種類 :
番号 :

出願年月日 :
国内外の別 :

○取得状況 (計◇件)

名称 :
発明者 :
権利者 :
種類 :
番号 :
取得年月日 :
国内外の別 :

[その他]
ホームページ等

6. 研究組織
(1) 研究代表者
竹之内 高志 (Takenouchi Takashi)
奈良先端科学技術大学院大学
情報科学研究科 助教
研究者番号 : 50403340

(2) 研究分担者
なし ()

研究者番号 :

(3) 連携研究者
なし ()

研究者番号 :