様式 C-19

科学研究費補助金研究成果報告書

平成２３年　　4月　20日現在

機関番号：14603
研究種目：若手 B
研究期間：2009～2010
課題番号：21700304
研究課題名（和文）
情報統合のためのアンサンブル学習アルゴリズムの開発と解析
研究課題名（英文）
Analysis and development of ensemble learning algorithms for information fusion
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（和文）：本研究ではアンサンブル学習アルゴリズムに関する研究を行った．特にECOCに基づく多値判別，順序付きラベルデータ，レーティングデータを対象に新たなアルゴリズムの提案と解析を行った．ECOCに基づく多値判別では，ブラッドリー･テリーモデルに基づく従来の2値判別器統合法の問題点を解決する枠組みを提案することで，計算量を大幅に削減しつつ従来法を上回る性能を達成した． 順序付きラベルデータに関しては，従来直接最適化を行う事が困難であったAUCを最大化可能な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．レーティングデータに関しては，行列因子化法に混合モデルを援用することでユーザーの個性を表現可能なモデルを提案し，大規模な実データに対してその有効性を示した．

研究成果の概要（英文）：In this program, we developed and analyzed ensemble learning algorithms. Especially, we focused on multi-class classification based on the framework of ECOC, ordered label classification and rating data analysis. For the multi-class classification based on ECOC, we tackled a problem of conventional Bradley-Terry model based method and proposed a novel method, which outperforms the conventional method with drastically lower computational cost. For the ordered label classification, we proposed a Boosting algorithm which can maximize AUC and analyzed its statistical properties. For the rating data, we extended the matrix factorization method using a mixture model and showed effectiveness of the proposed method using a large scale real dataset.

交付決定額

（金額単位：円）

	直接経費	間接経費	合　計
2009 年度	1,400,000	420,000	1,820,000
2010 年度	1,400,000	420,000	1,820,000
年度			
年度			
年度			
総　計	2,800,000	840,000	3,640,000

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：パターン認識．アンサンブル学習, ECOC，判別

１．研究開始当初の背景
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を解き，現実のアプリケーションに適用することが可能になりつつある．このような大規模問題に対して，問題を比較的解きやすい複数の小規模問題に分割して解いておき，各々の結果を統合する方法、アンサンブル学習は有効なアプローチのひとつであり，大きな注目を集めていた．

２．研究の目的
複数の情報源からの情報を統合することにより，高精度の情報を復号するための新しい手法を提案しその理論的側面を明らかにする．具体的には，ECOCに基づく多値判別，順序付きラベルのための判別，レーティングデータの解析などを行うためのアンサンブル学習アルゴリズムの提案とその解析を行い，得られた知見を更にフィードバックすることで，より有用な手法の提案を目指す．

３．研究の方法
ECOCを用いた多値判別，順序付きラベルの判別の問題に対しては，代表的なアンサンブル学習アルゴリズムであるブースティングのアイディアを援用することで，大規模問題への適用を可能とした．レーティングデータに対しては，個人毎の個性を表現するために混合モデルを導入して対応した．

４．研究成果
(1)ECOCに基づく多値判別：アンサンブル学習の枠組みで，2 値判別器の統合によって多値判別を行うための新たな手法を提案した．ブラッドリー･テリーモデルに基づく従来法で問題となっていた点を解決する枠組みを提案したことで，提案法は従来法を上回る性能を発揮した．またコスト関数の変更により精度を保ったまま計算量を削減することができた．

(2)順序付きデータへの対応：2値判別問題においてラベルに順序が付与されているデータではROCカーブの下側面積（AUC）が判別器評価の指標として用いられる．従来，AUCの最適化は非凸な問題であるため直接最適化することは難しかったが，適切な近似コスト関数を考案することで，AUCを最大化する判別器を直接構成するためのアンサンブル学習アルゴリズムを提案した．また，提案アルゴリズムの統計的性質などを議論し，外れ値に影響を受けにくいロバストなコスト関数を考案した．

(3)レーティングデータ解析法：あるアイテムに対して複数のユーザーが評価を行ったデータを元に推薦を行うシステムにおいて基幹技術として用いられている行列（テンソル）因子化法に対して，2つの拡張を行った．①ユーザーの嗜好を反映するようなグループが存在するデータを対象として，混合モデルを用いて拡張を行い，その有効性を大規模な実データで示した．②データの各変量が異なる素性を持つような場合を対象として，指数型分布族を用いた拡張を行い，効率的な最適化を行うための近似法を提案した．実データを用いて提案法と従来手法と比較し，提案法が精度の高い予測性能を発揮することを確認した．

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　4 件）

1. T. Takenouchi, and S. Ishii.
Ternary Bradley-Terry model-based decoding for multi-class classification and its extensions.
Machine Learning in press, 2011. 査読有

2. S. Kozawa, T. Takenouchi, and K. Ikeda.
Subsurface imaging by Bayesian super-resolution for anti-personal mine detection using ground penetrating radar.
Journal of Signal Processing, 14(4), pp. 297-300, 2010. 査読有

3. T. Takenouchi, and S. ishii.
A multi-class classification method based on decoding of binary classifiers.
Neural Computation, 21(7), pp.2049-2081, 2009. 査読有.

4. I. Suzuki, T. Takenouchi, M. Ohira, S. Oba, and S. Ishii.
Robust model selection for classification of microarrays.
Cancer Informatics, 7, pp. 141-157, 2009. 査読有.

〔学会発表〕(計 15 件)

1. Satoshi Kozawa.
Subsurface imaging for anti-personal mine detection by Bayesian super-resolution with Smooth-gap prior.
International Symposium on Artificial Life and Robotics (AROB 16th '11), 2011.
2011.1.27, 大分B-conプラザ

2. 林浩平.
大規模データのための指数族テンソル因子化法.
第 13 回情報論的学習理論ワークショップ (IBIS'10), 2010.
2010.11.4, 東京大学.

3. Kohei Hayashi.
Exponential family tensor factorization for missing-values prediction and anomaly detection.
The IEEE International Conference on Data Mining (ICDM'10), December 2010.
2010.12.17, オーストラリア, シドニー.

4. T. Takenouchi.
Theoretical analysis of Cross-Validation(CV)-EM algorithm.
20th International Conference on Artificial Neural Networks (ICANN 2010).
2010.9.18,
ギリシャ, テッサロニキ.

5. T. Takenouchi.
Bayesian decoder for multi-class classification by mixture of divergence.
Information Geometry and its Applications III, 2010.
2010.8.5, ドイツ, ライプチヒ.

6. 竹之内 高志.
混合モデルによる行列因子化法の拡張とその応用
第 54 回システム制御情報学会研究発表講演会(SCI'10).
2010.5.20, 京都.

7. 林 浩平.
指数族テンソル因子化法による欠損値予測と異常検知
人工知能学会 データマイニングと統計数理研究会 (SIG-DMSM).
2010.3.30, 東京, 統計数理研究所.

8. S. Kozawa.
Subsurface imaging by Bayesian super-resolution for anti-personal mine detection using ground penetrating radar.
International Workshop on Nonlinear Circuits, Communications and Signal Processing, 2010.
2010.3.4, アメリカ合衆国, ハワイ.

9. 中村 政義.
行列因子化の混合モデルへの拡張と映画レーティング予測への応用
電子情報通信学会技術研究報告, vol. 109, no. 363, NC2009-86, pp. 89-93, 2010.
2010.1.19, 北海道大学.

10. 武田 学.
経験尤度を用いた統計量推定法とその性質
電子情報通信学会技術研究報告, vol. 109, no. 363, NC2009-84, pp. 77-81, 2010.
2010.1.19, 北海道大学.

11. 中村 政義.
各要素が混合ガウス分布に従う行列に対する行列因子化による欠損値予測.
第 12 回情報論的学習理論ワークショップ (IBIS 2009).
2009.10.20, 九州大学.

12. 林 浩平.
Sparse Exponential Family PCA with Heterogeneous Attributes.
第 12 回情報論的学習理論ワークショップ (IBIS 2009).
2009.10.20, 九州大学.

13. Takashi Takenouchi.
A multi-class classification by ECOC

ensemble and its extension
2009 年度統計関連学会連合大会, p.127, 2009.
2009.9.7,同志社大学.

14. T. Takenouchi.

Extension of ROC curve.

IEEE International Workshop on Machine

Learning For Signal Processing, 2009.

2009.8.4, フランス, グルノーブル.

15. T. Takenouchi.
Robust classification with mislabeling model.
Mathematical Aspects of Generalized Entropies and their Applications, RIMS workshop, 2009.
2009.7.8, 京都.

〔図書〕(計1件)
金森 敬文(著), 竹之内 高志 (著), 村田 昇 (著), 金 明哲 (編集). 共立出版.
パターン認識 (Rで学ぶデータサイエンス 5) (単行本). 2009. 273ページ.

〔産業財産権〕
○出願状況 (計 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
( )
研究者番号:

(3)連携研究者
( )

研究者番号: