

平成21年度科学研究費補助金実績報告書（研究実績報告書）

1. 機関番号 1 4 6 0 3      2. 研究機関名      奈良先端科学技術大学院大学
3. 研究種目名 若手研究 (B)      4. 研究期間      平成 21年度 ～ 平成 22年度
5. 課題番号 2 1 7 0 0 3 0 4
6. 研究課題名 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

7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
5 0 4 0 3 3 4 0	フリガナ: タケノウチ タカシ 竹之内 高志	情報科学研究科	助教

8. 研究分担者(所属研究機関名については、研究代表者の所属研究機関と異なる場合のみ記入すること。)

研究者番号	研究分担者名	所属研究機関名・部局名	職名
	フリガナ		

9. 研究実績の概要

下欄には、当該年度に実施した研究の成果について、その具体的内容、意義、重要性等を、交付申請書に記載した「研究の目的」、「研究実施計画」に照らし、600字～800字で、できるだけ分かりやすく記述すること。また、国立情報学研究所でデータベース化するため、図、グラフ等は記載しないこと。

今年度はアンサンブル学習の枠組みで、ROCカーブ最適化のためのブースティング法、バイオインフォマティクスデータに対して有効な判別法、行列因子化法の考案を行った。

- ・ 2値ラベルにおいて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手法を提案した。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。
- ・ 病理診断システムなどに用いられるバイオインフォマティクスデータは例題数が入力の次元（遺伝子数）に比べて著しく少ない。このようなデータに対して、判別に有効な遺伝子を選択することは重要な問題であるが、従来の選択法では、例題数の少なさとデータに含まれるノイズの影響により、信頼性の高い選択を行うことができなかった。この問題に対し、ブートストラップ法を用いて性能の最悪評価を行うことで信頼性の高い遺伝子数選択、判別器を構成するための手法を提案し、人工データ、実データでその有効性を確かめた。
- ・ 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。

10. キーワード

- (1) アンサンブル学習      (2) 判別分析      (3) ブートストラップ  
 (4) ブースティング      (5)      (6)  
 (7)      (8)      (裏面に続く)

11.研究発表（平成21年度の研究成果）

〔雑誌論文〕 計（2）件    うち査読付論文 計（2）件

著者名	論文標題			
T. Takenouchi.	A multi-class classification method based on decoding of binary classifiers.			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Neural Computation	有	21(7)	2009	2049-2081

著者名	論文標題			
I. Suzuki.	Robust model selection for classification of microarrays.			
雑誌名	査読の有無	巻	発行年	最初と最後の頁
Cancer Informatics	有	7	2009	141-157

著者名	論文標題			
雑誌名	査読の有無	巻	発行年	最初と最後の頁

〔学会発表〕 計（9）件    うち招待講演 計（1）件

発表者名	発表標題	
林 浩平	Exponential Family Tensor Factorization for Missing Values Prediction and Anomaly Detection	
学会等名	発表年月日	発表場所
第12回DMSM研究会	2010.3.30	統計数理研究所

発表者名	発表標題	
S. Kozawa	Subsurface imaging by Bayesian super-resolution for anti-personal mine detection using ground penetrating radar.	
学会等名	発表年月日	発表場所
International Workshop on Nonlinear Circuits, Communications and Signal Processing	2010.3.4	Hawaii, USA

発表者名	発表標題	
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.	
学会等名	発表年月日	発表場所
電子情報通信学会技術研究報告NC,	2010.1.19	北海道大学

発表者名	発表標題	
武田 学	経験尤度を用いた統計量推定法とその性質	
学会等名	発表年月日	発表場所
電子情報通信学会技術研究報告NC,	2010.1.19	北海道大学

発表者名	発表標題	
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.	
学会等名	発表年月日	発表場所
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

発表者名	発表標題		
林 浩平	Sparse Exponential Family PCA with Heterogeneous Attributes.		
学会等名	発表年月日	発表場所	
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	

発表者名	発表標題		
Takashi Takenouchi	A multi-class classification by ECOC ensemble and its extension		
学会等名	発表年月日	発表場所	
2009年度統計関連学会連合大会	2009.9.7	同志社大学	

発表者名	発表標題		
T. Takenouchi	Extension of ROC curve.		
学会等名	発表年月日	発表場所	
IEEE International Workshop on Machine Learning For Signal Processing	2009.9.4	Grenoble, France	

発表者名	発表標題		
T. Takenouchi	Robust classification with mislabeling model.		
学会等名	発表年月日	発表場所	
Mathematical Aspects of Generalized Entropies and their Applications	2009.7.8	Kyoto, Japan	

〔図書〕 計 ( 1 ) 件

著者名	出版社		
金森敬文	共立出版株式会社		
書名	発行年	総ページ数	
<a href="#">パターン認識 (Rで学ぶデータサイエンス 5)</a>	2009	288	

12. 研究成果による産業財産権の出願・取得状況

〔出願〕 計 ( 0 ) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

〔取得〕 計 ( 0 ) 件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別

13. 備考

※ 研究者又は所属研究機関が作成した研究内容又は研究成果に関するwebページがある場合は、URLを記載すること。

--