

氏 名 プリチャード 真理

学位（専攻分野） 博士（統計科学）

学位記番号 総研大甲第 1383 号

学位授与の日付 平成 22 年 9 月 30 日

学位授与の要件 複合科学研究科 統計科学専攻
学位規則第 6 条第 1 項該当

学位論文題目 Boosting method via the sparse learner approach
for high-dimensional gene expression data

論文審査委員 主 査 教授 松井 茂之
教授 江口 真透
准教授 藤澤 洋徳
部長 松浦 正明
(財団法人癌研究会癌研究所)

論文題目： Boosting method via the sparse learner approach for high-dimensional gene expression data

(高次元の遺伝子発現データのためのスパース学習アプローチによるブースティング法)

論文は5章と補遺の構成で全97ページからなり、英語で書かれている。高次元の遺伝子発現データの判別分析のためのスパース学習によるブースティング法が提案され、その数値的性能の比較・検討と、乳がんの遺伝子発現データへの適用がなされている。

第1章は序章である。近年、多くの医学研究では、マイクロアレー技術により、数千の遺伝子からなる高次元の遺伝子発現データが収集されており、これを用いた病型・病態の予測が統計科学上の重要な問題であることが述べられている。

第2章では、ブースティングの代表的な方法であるアダブーストとその修正版であるイータブーストが詳しく紹介されている。これらは、あらかじめ用意した学習器のセットから、最適な学習器を逐次的に選択し、その線型結合によって判別スコア関数を構成する方法である。新たな特徴ベクトルが与えられたときに、その判別スコアの値の符号によってクラスラベルの判別が行われる。しかし、データの次元が高くなることで、学習器のセットが過度に複雑になり、その結果、学習データへの過適合がおこり、判別精度が著しく低下することが指摘されている。

第3章では、この問題に対処するために、学習器セットから一部の学習器をデータに基づいて削除することが提案されている。すなわち、学習器のスパースセットを用いたブースティング法の提案である。具体的には、偽陽性確率と偽陰性確率がある値より大きい学習器が学習器のセットから削除される。この提案によって、判別モデルの複雑さは、VC (Vapnik-Chervonenkis) 次元という指標において、従来の学習器セットを用いた場合に比べて、学習器の削減率を乗じたものに減少することが示されている。つぎに、数値実験によって、従来の学習器セットを用いたアダブースト、イータブースト法と比較して、提案法の性能が優れていることが確認されている。公開されている乳がん細胞の遺伝子発現データを用いた遠隔再発の予測分析では、従来法と比較して、テストエラーの意味で、予測精度が改善することが示されている。

第4章では、高次元の遺伝子発現データの判別分析における困難な問題である多重解の存在（同程度の判別精度を達成しうる複数の遺伝子セットの存在）が上記の乳がんデータにおいても確認され、遺伝子セットの選択に関する考察がなされている。

第5章では全体の結論が述べられている。

博士論文の審査結果の要旨

論文題目： Boosting method via the sparse learner approach for high-dimensional gene expression data

(高次元の遺伝子発現データのためのスパース学習アプローチによるブースティング法)

高次元の遺伝子発現データに対して、学習器のスパースセットを用いてブースティングを行うというアイデアはこれまでなかったものである。このアイデアを具現化するために、偽陽性確率と偽陰性確率に基づく学習器の削減基準の提案とそれを組み込んだブースティング法を開発したことの新規性は高く評価できる。VC次元に基づく考察は、学習器のスパースセットを用いることの有効性を定量的に評価するものであり、提案法の理論的な解釈を与えるものである。さらに、数値実験、並びに、乳がん研究データへの適用を通して、提案法の優位性を示している。以上より、本論文は、高次元の遺伝子発現データを用いた有効な判別分析法の確立に十分資する研究であり、学位を授与するに値するものと判断する。なお、第3章のスパース学習器の提案法に関しては、申請者の単著による英文論文1編が雑誌「IPSJ Transactions on Bioinformatics」(査読付き)に受理されており、第4章の多重解の検討に関しては、申請者を筆頭著者とする1編の論文が日本統計学会誌38巻シリーズJ(査読付き)に掲載されている。

複合科学研究科における博士課程の学位授与に係る論文審査等の手続き等に関する規定第10条に基づいて口述による試験を実施した。この結果、出願者は博士(統計科学)の学位を授与するに十分な学識を有するものと判断し、合格と判定した。