# 審 査 の 結 果 の 要 旨

氏　　名　　太田　力文

本論文は6章からなる。第1章では、多因子性疾患に罹患するリスクを推定する方法に関する過去の研究事例を詳しく紹介している。1塩基多型は通常は2つのアレル A,a (高頻度を A, 低頻度を a と表記)から構成される。アレルの組合せは AA, Aa, aa の3通りあり遺伝子型と呼ばれる。疾患に関わるアレルがあるとすれば、通常は低頻度アレル a であり、リスクアレルとここでは呼ぶ。既存のアルゴリズムは1塩基多型のリスクアレルの加算的効果のみを使い、例えば a をリスクアレルとすれば遺伝子型 aa は遺伝子型 Aa の2倍のリスクを持つこととなる。しかし伝統的なメンデル性遺伝病では相加的な遺伝効果だけではなく、リスクアレル a が1つでもある遺伝子型（Aa もしくは aa）の時に罹患する顕性的効果(dominant)、もしくはリスクアレル a が2つ揃う遺伝子型 aa の時に罹患する潜性的効果(recessive)が考慮される。

第2章では、過去の研究では活用されてこなかった顕性・潜性等の非相加的な遺伝効果を罹患リスク推定に取り込むことの意義、その手段として機械学習手法の1つであるブースティング理論を拡張する着想に至った動機を述べている。第3章では、本論文が提案する新手法 GenoBoost の予測精度を評価する際に使う英国の大規模ゲノムコホート UK Biobank の遺伝子型と表現型（疾患情報）の統計情報と、正解情報が利用できるシミュレーション・データの生成方法を述べている。

第4章では、本論文の主題である新手法 GenoBoost を記述している。ブースティング理論では様々な定式化が提案されている。その中で logistic loss を最適化する LogitBoost アルゴリズムが、罹患リスク推定に親和性があることを述べ、最適解をニュートン法で計算できることを示している。次に、非相加的な遺伝効果を柔軟に表現できる関数を用意し、最適解を構成する関数を解析的に計算できることを示した定理（具体的には定理 4.1.4）を証明している。この定理は実用的に大変重宝であり、UK Biobank が提供する数百万個の遺伝子型と数十万個のサンプルデータに適用しても、罹患リスク式を現実的な時間で計算することができる。

第5章では、UK Biobank が提供する互いに相関の無い 12 多因子性疾患（関節リウマチ、乾癬、痛風、炎症性腸疾患、喘息、全原因型認知症、アルツハイマー病、心房細動、乳がん、結腸直腸がん、冠動脈疾患、2 型糖尿病）に対し、新手法 GenoBoost と既存の7手法（snpboost, snpnet, lassosum, LDpred, PRS-CS, SBayesR, C+T）を適用し、罹患リスク予測精度の違いを分析している。12 疾患の中で、GenoBoost の予測精度は4疾患で1位、3疾患で2位という良好な結果が得られている。さらに、非相加的な遺伝効果を考えることの利点が顕著だった関節リウマチと乾癬について分析し、自己免疫疾患に関連する MHC 遺伝子座の非相加的効果を捉えられたことを示している。非相加的な遺伝効果を予測したバリアントを含む遺伝子の中には，文献で疾患との関連性が報告されていたものもあった.

第6章では、提案手法 GenoBoost を今後どのように改良し拡張してゆくべきかについて考察

している。特に、UK Biobank が提供を予定している 50 万人の全ゲノム再解読データが利用可能になったとき、頻度の低いバリアントを適切に扱うことの重要性を述べて締めくくっている。

なお、本論文の中心である第４章と第５章は、東京大学の森下真一と鈴木裕太、MIT の谷川洋介と Manolis Kellis との共同研究だが、アルゴリズムは全て太田が設計し開発し、実験結果も全て太田が取得し検証している。論文提出者が主体となって分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

よって本論文は博士（医科学）の学位請求論文として合格と認められる。

以上１６８０字