

平成25年度北海道大学情報基盤センター共同研究成果報告書

1. 研究領域番号 A3 大規模データ科学
2. 研究課題名 大規模医療データを有効活用した高度臨床診断と予測
3. 研究期間 平成25年 4月 1日～平成26年 3月31日
4. 研究代表者

氏名	所属機関・部局名	職名	備考
栗原 考次	岡山大学大学院・環境生命科学研究科	教授	

5. 研究分担者

氏名	所属機関・部局名	職名	備考
水田 正弘	北海道大学・情報基盤センター	教授	
南 弘征	北海道大学・情報基盤センター	准教授	
富田 誠	東京医科歯科大学・医学部附属病院	特任准教授	
石橋 雄一	株式会社スタットラボ	代表取締役	
田中 豊	岡山大学	名誉教授	
石岡 文生	岡山大学大学院・法務研究科	助教	

6. 共同研究の成果

下欄には、当該研究期間内に実施した共同研究の成果について、その具体的内容、意義、重要性等を、共同研究申請書に記載した「研究目的」と「研究計画・方法」に照らし、800字～1,000字で、できるだけ分かりやすく記載願います。文章の他に、研究成果を端的に表す図表を貼り付けても構いません。なお、研究成果の論文・学会発表等を行った実績（発表等の予定を含む。）があれば、あわせて記載して下さい。

本研究では、医学の分野で得られる各種の画像や SNP データなどの大規模医療データを有効活用した高度臨床診断及び予測を行うために、以下のような課題について統計的手法の開発とその応用について検討した。

1. 画像及びテキスト情報に基づく医療診断と予測

- (1) 乳腺疾患を対象として病理組織標本画像及び病理診断書における画像情報及びテキスト情報の効率的なデータベース化を行うとともに、画像情報、テキスト情報に基づき類似症例の検索可能なシステムの構築を行った。画像情報では、ガラススライド標本の画像ファイルを数値化し、データベースに蓄積された類似の画像標本の検索を可能にした。また、病理診断書をテキストマイニングにより解析を行って、診断書テキストから特定の疾患である確率の計算や、類似の診断書をデータベースから検索することも可能にした。さらに、診断書の内容と診断名との医学的な整合性のチェックを可能にした。これらの成果は、教育用の利用のみならず、病理診断の精度を高めることに貢献できる。

(研究成果のつづき)

- (2) 樹木モデルを用いた診断論理の統計解析として、肝細胞癌における悪性度を示す（低分化、中分化、高分化、良性）カテゴリカルデータに対して、ジニ係数またはエントロピーによる情報利得に基づいて分類木(classification tree)の作成を行った。その結果、肝細胞相の濃染程度、CTによる造影撮像結果(Ph1)及び造影撮像結果(Ph3)で分類できることなどがわかった。

2. 空間構造を利用した SNP データの解析

SNP 間の組換え頻度により、組換えがほとんど起こっていない連鎖不平衡の領域を同定することができる。この領域は多様性がおさえられており、study 群における共通の疾患(common disease)の原因となる共通の変異(common variant)を含む可能性が高いとされている(CDCV 仮説)。領域を絞り込んだ後には、その範囲で染色体上の配列を推定することができ、これを一般的にハプロタイプ推定という。絞り込む方法をハプロタイプ・ブロック同定という。さらに、ブロック内のハプロタイプの組合せ(diplotype configuration)が、ある多変量的変量との関連解析の手法は既に提案されている。これらの流れを他の先行研究と比較し、利点や問題点をまとめた(富田, 計算機統計学 25(1), 53-68, 2013)。