

## 平成24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 大規模医療データに基づく高度臨床診断
3. 研究期間 平成24年4月23日 ~ 平成25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

本共同研究では、医学の分野で得られる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) 治療前の患者の胸部大動脈の座標情報を CT 画像から取得し、その情報から血管形状の曲率・振率の定量化を行った。さらに、それらをパラメータとして、胸部大動脈ステント治療後に生じる副作用を予測するモデル作成を行い、各種の統計手法によってそのモデルの妥当性を評価した。その結果、ステントが挿入されている箇所曲率・振率が大きい患者ほどステント治療後の副作用が生じやすくなるという医師の経験的知見を客観的に強めることができた。

(2) 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をしたり、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。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。

## 2. 空間構造を利用した SNP データの解析

(1) 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。推定された連鎖不平衡係数  $D'$  の値及び位置情報を用いて、こうした領域の検出及び空

(研究成果のつづき)

間的な構造を検討するとともに他の空間集積性の方法との比較検討を行った。また、ブロック同定後の関連分析についても検討した。

(2) 大量の SNP データに対する Echelon 解析を行うために、データ値の高低を考えずに構造のみで階層化を表現させ、パターン認識やカットオフ手法に基づいた尺度を用いて複雑な階層構造を指標化し、それに基づいたクラスター化を行った。

### 3, 医学ビッグデータの解析とその応用

医薬品の有害事象（副作用）と因果関係の可能性に関する情報を見つけるシグナル検出において、数百万件のデータを解析して、有害事象を検出した。また、従来 Windows サーバー用のプログラムをスパコン用に改良し、約 20 分の 1 の実行時間で解析することができた。