審査の結果の要旨

氏名　文　銘塡

DNA 塩基配列決定技術等の急速な進歩により、個人レベルのゲノムの違いや、各種疾病患者における体細胞変異等が続々と明らかにされつつあり、精密化医療と呼ばれる新しい医療が誕生しようとしている。そしてそこで産出されている大量データから有用情報を抽出するための情報処理技術が求められている。本論文は、このような背景において、癌のバイオマーカー発見、具体的には進行度で分けた癌細胞のサンプルで特徴的な発現を示す遺伝子集合、を高感度で検出する試みについて論じている。この問題の難しさは、探索される特徴（遺伝子）数と比べてサンプル数が十分でなく、患者の個人差や実験誤差等のために、真に有効なマーカー遺伝子の発見が難しい点にある。この点を克服するために、本論文では、機械学習法における各群の特徴のアンサンブル抽出という枠組みで、任意のステージの癌患者から得られたサンプル群において特徴的に発現する遺伝子セットを同定しようとしている。ここでアンサンブルとは、データに摂動をかけて擬似的に数を増やしたり（ブートストラップ、あるいはバギング）、別のデータと組み合わせたりすることによって得た多くの特徴抽出結果を総合しようというアプローチを指す。

本論文は、主に二つのセクションから成る。セクション１では、サポートベクターマシン（SVM）において L1 正則化法を使った変数選択法を腎細胞がん遺伝子発現データに適用した研究について論じており、セクション２では、遺伝子発現データに DNA メチル化データを組み合わせて、教師なしに変数選択を行う方法を同種のデータに適用した研究について述べている。

上述のように、セクション１で論文申請者は、L1 正規化を使った SVM の再帰的変数選択法を用いている。この方法では、寄与の小さい変数の重みはゼロにされるので、変数選択を効率的に行なうことができ、また雑音に強い答えが得られることが期待できる。データには、公共データベース TCGA から、淡明細胞型腎細胞癌のステージⅠ（268 個）とⅣ（84 個）を区別する遺伝子群をその RNA-seq 発現データに基づき約２万個の中から探索した。この癌を選んだ理由は、比較的ステージ分類が直感的にわかりやすいことによる。詳細は省くが、バギングによりサンプルをランダムに分割して、多くで利用された特徴を選択したところ、177 個の遺伝子が選択された。得られた予測法は、他の方法（*t*-検定に基づく FCBF 法、ランダムフォレスト法、L2 正則化による SVM 法）と比べて高い性能を示した（選択変数の数はアルゴリズムによって異なる）。また、変数をトップ 20 に絞ってもこの傾向は変わらず、バギングを繰り返したときに選択される変数の安定性でも、本方法が優れていた。一方、本方法の主な欠点は計算量を多く要することである。結果の医学生物学的考察や、方法論自体のオリジナリティに若干弱さがあるが、様々な客観的指標を使って、提案した方法の有効性を示したことは評価できる。

セクション２では、上述のように２種類のデータ（遺伝子発現量とメチル化情報）を組み合わせることで、より有効なマーカー遺伝子群を探索する試みについて述べている。生物学的には、高発現を示す遺伝子領域（転写開始点から 1.5KB 上流

から遺伝子本体の 3'末端まで）では DNA メチル化レベルが低いことが知られているので、各遺伝子のメチル化レベル情報を付加することで、より信頼度の高い結果が期待される。用いたデータは、基本的にはセクション１と同様の腎細胞癌のデータであるが、メチル化レベル情報が加わっている。発現量とメチル化レベルをそれぞれ[0, 1]区間に正規化したあと、３つの方法で融合させた。すなわち、カーネル付きと無しの主成分分析法（第１主成分のみ）とオートエンコーダ法（ニューラルネットを用いた次元圧縮法）である。３つの方法の結果に特に大きな差は見られなかったが、発現量やメチル化データを単独で用いたときと比べて、交差検証法で性能の向上が認められた。有効バイオマーカー数はおよそ 20 もあれば十分であった。その中でデータを組み合わせることで初めて有効性が示されたのは 11 個で、うち 6 個は既知のがん関連遺伝子であり、残りは新規のバイオマーカー候補となる可能性がある。本解析は、今後、より多元的なデータを組み合わせたマルチオミクス研究の先駆けとみることができ、比較的単純ながら意義が認められる。

なお、本論文セクション１の内容は、中井謙太との共著で専門誌に論文を発表済みであり（Moon & Nakai, 2016）、セクション２の内容も、同じく論文を投稿中であるが、どちらの研究も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。

したがって、博士（科学）の学位を授与できるものと認める。

以上 1992 字