

大規模データ分析による精神神経系疾患診断系の検討 Development of Psychiatric Diagnosis System using Big Data Analysis

石井 一夫[†] 沼田周助[‡] 木下 誠[‡] 大森 哲郎[‡]
Kazuo Ishii[†] Shusuke Numata[‡] Makoto Kinoshita[‡] Tetsuro Ohmori[‡]

1. はじめに

近年の国内経済状況の長期悪化、および社会状況の変化により、精神神経系疾患に罹患する患者数が急激に増えている。平成23年の第19回社会保障審議会医療部会資料の「患者調査」に基づく資料によると、平成11年から患者数は急激に増加しており、平成20年にはその患者数は、がん患者の2倍以上で、糖尿病よりも多くなっている。なかでも、平成8年には43.3万人であったうつ病の患者数がうつ病等の気分障害の総患者数は、平成20年には104.1万人と12年間で2.4倍に増した。一方、国内の自殺者数も、平成10年以降、継続して毎年3万人を超えており同資料においても3万1千人で、悪性新生物、心疾患、脳血管疾患、肺炎、老衰、不慮の事故に続き、死因の第9位となっており、社会的な問題となっている。また、自殺の大きな要因としてあげられるのは、うつ病などの精神疾患との因果関係である。トーマス・E・エリスらによれば、自殺既遂者の95%は何らかの精神疾患を患っていて、その大半が治療可能だったという研究結果もある。しかし、うつ病および総合失調症などの精神神経系疾患の原因はいまだ不明であり、その診断は、もっぱら医師による患者からの愁訴に基づく問診により実施されてきた。科学的エビデンスによる診断法の存在しない、いわゆる「アンメットメディカルニーズ」(未充足の医療ニーズ)の疾患である。

本研究は、複数の診断マーカによる説明変数を用いた精神神経系疾患の診断系を確立することを目標に、東京農工大学農学府ゲノム科学人材育成プログラムと徳島大学大学院ヘルスバイオサイエンス研究部精神医学科との共同研究により精神神経系疾患のゲノムレベルのデータを用いたデータ分析を開始した。その結果、うつ病、総合失調症、および双極性感情障害の3つの精神神経系疾患において複数の変動マーカ(ゲノムメチル化および発現変動遺伝子マーカ)を元にした判別分析により、非常に高い感度および特異度で疾患患者と健常者を識別できる数理モデルを構築することが可能となったので報告する

[†] 東京農工大学農学府 Graduate School of Agricultural and Life Sciences, Tokyo University of Agriculture and Technology
[‡] 徳島大学大学院ヘルスバイオサイエンス研究部 Institute of Health Biosciences, The University of Tokushima Graduate School

データ解析のワークフロー

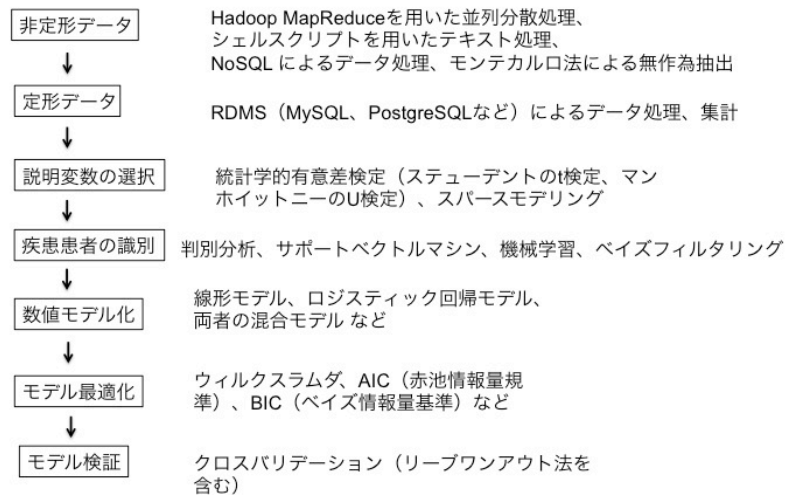


図1 本研究で用いられるデータ解析の全体のワークフロー

2. データ分析行程

図1にデータ分析の大まかな全体のワークフローを示す¹⁻⁴⁾。本研究では、まずメチル化マイクロアレイを用いて数理モデル化を試みた。その後、次世代シーケンサーの少数データでその有効性を確認した。

(1) 説明変数の選択

次世代シーケンサーやマイクロアレイによる網羅的発現解析や網羅的ゲノムメチル化解析を用いた場合、発現変動遺伝子や、ゲノムDNAメチル化部位など、大量の説明変数が得られる。これらの説明変数は、負の二項分布またはポアソン分布をすることが知られている。従って正規分布を仮定しているステューデントの t 検定を用いず、マンホイットニーの U 検定などのノンパラメトリック検定を用いて選択した。

(2) 識別方法と説明変数の最適化

複数の説明変数による疾患、患者の識別には、多変量解析(重回帰分析、判別分析、クラスタ分析など)、サポートベクトルマシン、機械学習(SOMなど)、ベイズフィルタリング、ランダムフォレストなどの利用が考えられる。今回は、このうち良好な結果の得られた判別分析を中心に紹介する。

説明変数の選択は最初に、各マーカについて健常者群と患者群の間でマンホイットニーの U 検定を実施し、両群で

有意差の見られたものを選択した。さらに、ウィルコクスラムダなどの別の指標をもちいて、説明変数を絞り込み、少数の説明変数で良好な識別の出来る数理モデルを選択した。識別の指標は、感度、特異度のほか、健常者群と患者群の間の判別スコアの平均値の差などを指標とした。

(3) 得られた数理モデルの検証

最終的に得られた数理モデルは、クロスバリデーション(交差検定法)により検証した。

(4) 分析結果の視覚化

分析結果は、グラフやヒートマップなどで視覚化した。

3. 各疾患での分析結果

うつ病および統合失調症における分析結果は、すでに他の場所で紹介しているので、そちらを参照していただきたい⁵⁾。今回はうつ病および統合失調症に加え、双極性感情障害 I 型においても、有効な識別能力を有する数理モデルの構築が可能であることが確認できたのでその結果を報告する。

双極性感情障害 I 型 20 名、健常者 20 名において、最適化を試みた 10 個のメチル化マーカを用いて判別分析を実施したときの判別分析の結果を図 2 に示す。

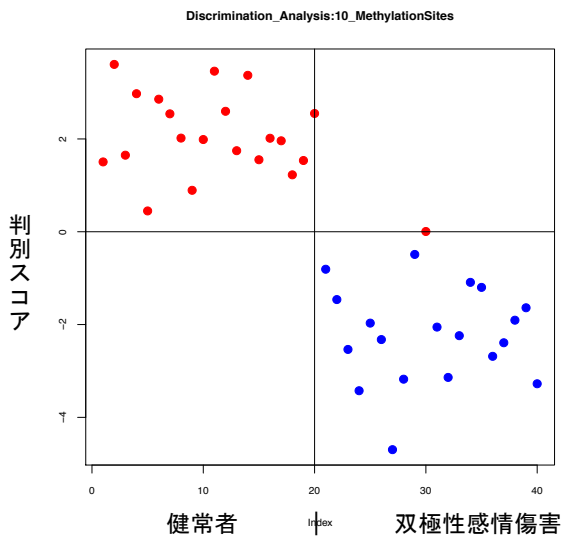


図 2 10 個のメチル化マーカを用いて、双極性感情障害 I 型および健常者を識別した判別分析の結果

10 個の説明変数(メチル化マーカ)を用いた判別分析により、双極性感情障害 I 型患者と健常者を良好に識別出来ていることが確認できる。

この 10 個の説明変数(メチル化マーカ)を用いた判別分析を、さらに別の双極性感情障害 I 型患者 12 名と健常者 12 名について実施した結果を図 3 に示す。図 2 で識別できた 10 個の説明変数を用いて、別の双極性感情障害 I 型患者群と健常者群を識別でき、この数理モデルが双極性感情障害 I 型の診断に有効であることが示唆された。

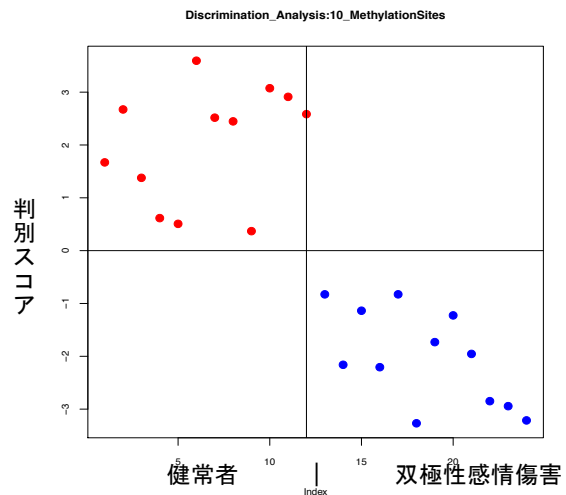


図 3 10 個のメチル化マーカを用いて、図 2 とは別の集団の双極性感情障害 I 型および健常者を識別した判別分析の結果

4. まとめ

今回、うつ病、統合失調症に加え、双極性感情障害の 3 つの精神神経系疾患において、良好な識別能力を有する診断系の構築に成功した。

その結果、従来では問診による経験的判断でしか、診断がつかなかった精神神経系疾患においてエビデンスに基づく診断を導入する糸口が見えてきた。

臨床医学的な知見の詳細は関連の臨床系の雑誌に報告するのでそちらを参照してほしい^{6,7)}。

謝辞

本研究を行なうにあたり、日本ヒューレット・パッカード社、Amazon Web Services, Inc. および、理化学研究所スーパーコンピュータ「京」と互換性を有する SCLS 計算機システムのご支援をいただいた。ここにお礼を申し上げる。

参考文献

- [1] 石井一夫: ゲノム科学におけるビッグデータ分析・大規模データマイニング, BIOINDUSTORY, 31, 6, pp67-73 (2014).
- [2] 石井一夫: 解説: 医療, 農学, 環境分野におけるビッグデータ解析, 生物工学会誌, 92, 2, pp92-93 (2014).
- [3] 石井一夫, 佐藤暁, 古崎利紀, 有江力, 寺岡徹: ゲノム科学におけるビッグデータ・データマイニング, 日本統計学会誌, 43, 1, pp90-111 (2013).
- [4] Rizzo M(著), 石井一夫, 村田真樹(共訳): R による計算機統計学, オーム社 (2011).
- [5] 石井一夫, 沼田周助, 木下 誠, 大森哲郎: ビッグデータ分析による精神神経系疾患診断系の検討, 日本計算機統計学会第 28 回大会中央大学後楽園キャンパス(2014)
- [6] Fuchikami M, Morinobu S, Segawa M, Okamoto Y, Yamawaki S, Ozaki N, Inoue T, Kusumi I, Koyama T, Tsuchiyama K, Terao T., DNA Methylation Profiles of the Brain-Derived Neurotrophic Factor (BDNF) Gene as a Potent Diagnostic Biomarker in Major Depression, PLoS One. 6, 8, e23881 (2011).
- [7] Kinoshita M, Numata S, Tajima A, Shimodera S, Ono S, Imamura A, Iga J, Watanabe S, Kikuchi K, Kubo H, Nakataki M, Sumitani S, Imoto I, Okazaki Y, Ohmori T. DNA methylation signatures of peripheral leukocytes in schizophrenia., Neuromolecular Med. 15, 1, 95-101 (2013).