

Amazon Cloud ビッグデータ解析環境による 精神神経系疾患診断系の構築

石井一夫[†] 沼田 周介[‡] 木下 誠[‡] 伊賀 淳一[‡] 渡部 真也[‡] 飯田 満[§] 大森 哲
郎[‡]
東京農工大学[†] 徳島大学[‡] 大塚製薬[§]

要旨

東京農工大学及び徳島大学精神医学の共同により、大規模臨床データによる精神神経系疾患の診断系の開発を行なうため AWS (Amazon Web Services) のクラウド環境に Hadoop、MapReduce、NoSQL、Rなどを組み合わせたビッグデータ解析環境を構築し、大規模データ解析を開始した。今回、PCR アレイ、マイクロアレイなどの多次元データを対象にモンテカルロシミュレーションを用いた最適化によるデータ解析を実施し、多変量の遺伝子及び遺伝子修飾マーカーの組合せによる精神神経系疾患診断系の最適化を試み、そのパフォーマンスを検証したので報告する。

1. はじめに

国内の複数の大学医学部との共同研究により精神神経系疾患（主にうつ病、総合失調症など）の診断系の開発を開始した。現在、定量 PCR とメチル化アレイによる約百数十人分のデータを取得しており、次世代シーケンサーによる網羅的メチル化解析、網羅的発現解析も開始している。現在、うつ病、総合失調症の予備的な解析をおこなっており、100%の感度および特異度で診断が可能なマーカーのセットを見いだしている。今回これらを題材に臨床診断系におけるデータ解析と統計解析について概説する。

2. 材料

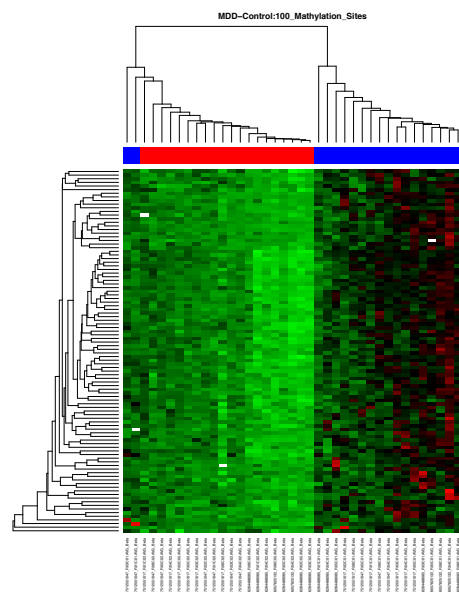
患者 20 名、健常者 19 名の血液から得た DNA を用いてメチル化アレイによる定量を行なったデータと、患者 25 名、健常者 25 名の血液から得た RNA を用いて PCR による解析を行なったデータを用いた。

2. 方法

AWS のクラウド環境にデータ分析環境を構築した。これにより統計学的有意差検定、多変量解析、機械学習などを行なえるようにした。統計学的検定により、有意差の見られる検査項目を上位から選択し、多変量解析、機械学習を行ないそのパフォーマンスを検証した。

3. 結果

メチル化部位につき、上位、100 個の検査項目を用いてクラスタ解析ヒートマップを行なった結果を以下に示す。青が健常者、赤が患者である。患者と健常者でうまく分別できる診断系が出来ていることが確認できた。

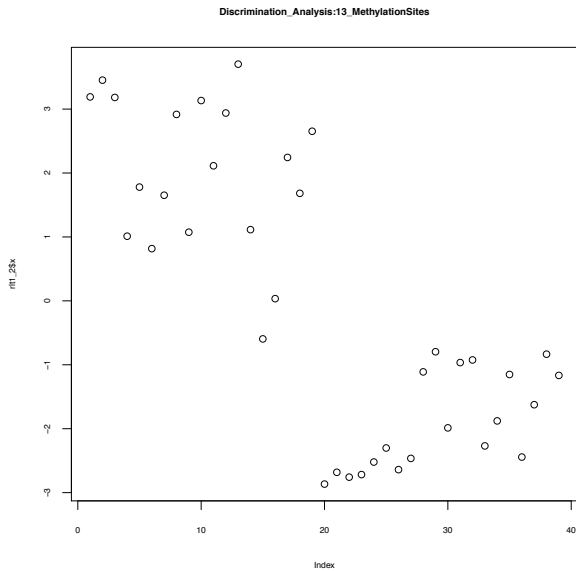


Construction of big data analysis system
for psychiatric diagnosis in Amazon cloud
services

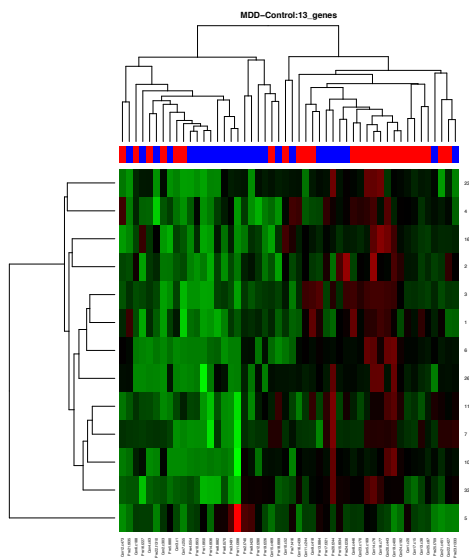
Kazuo Ishii[†] Syuusuke Numata[‡] Makoto
Kinoshita[‡] Jyunnichi Iga[‡] Shinya Watanabe[‡]
Mitsuru Iida[§] Tetsuro Ohmori[‡]
Tokyo University of Agriculture and
Technology[†] University of Tokushima[‡]
Otsuka Pharmaceutical Co., Ltd. [§]

メチル化部位につき、上位 13 個の検査項目で判別分析を行なった結果を示す。左側が健常者、右側が患者である。この場合もうまく分別できている。

参考文献
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 巻 1 号 90-111 頁 (2013)



PCR による上位 13 個の検査項目により、クラスター解析ヒートマップを行なった結果を以下に示す。青が健常者、赤が患者である。この場合は、メチル化の場合と比較してうまく分別出来ていなかった。



3. 結論

患者、健常者を分別できる診断系がアマゾン上で構築できた。発表ではそのシステムの詳細を紹介する。