

異種混合学習を用いたアルツハイマー病発症予測解析

菊地 正隆^{*1}, 小林 香織^{*1,2}, 湯本 英二^{*3}, 伏見 泰人^{*2}, 武田 理宏^{*4}, 真鍋 史朗^{*4},
松村 泰志^{*4}, 上條 憲一^{*2}, 中谷 明弘^{*1}

^{*1} 大阪大学大学院医学系研究科 ゲノム情報学共同研究講座, ^{*2} NEC 医療ソリューション事業部, ^{*3} NEC データサイエンス研究所, ^{*4} 大阪大学大学院医学系研究科 医療情報学

Prediction of Conversion to Alzheimer's Disease using Heterogeneous Mixture Learning

Masataka Kikuchi^{*1}, Kaori Kobayashi^{*1,2}, Eiji Yumoto^{*3}, Yasuto Fushimi^{*2},
Toshihiro Takeda^{*4}, Shiro Manabe^{*4}, Yasushi Matsumura^{*4}, Kenichi Kamijo^{*2},
Akihiro Nakaya^{*1}

^{*1} Dept. of Genome Informatics, Osaka University Graduate School of Medicine,

^{*2} Medical Solutions Division, NEC Corporation,

^{*3} Data Science Research Laboratories, NEC Corporation,

^{*4} Dept. of Medical Informatics, Osaka University Graduate School of Medicine

アルツハイマー病 (AD: Alzheimer's disease) を含む認知症の前段階は軽度認知障害 (MCI: Mild cognitive impairment) として知られているが、必ずしも全員が AD へ移行するわけではない。本研究では脳 fMRI データや髄液バイオマーカーデータといった様々なデータを異種混合学習技術により統合し、健常者と AD 患者を判別する複数の学習器からなる決定木を生成した。生成した決定木に基づき MCI 患者を予測した結果、AD 患者と予測された MCI 患者は 60.1% が 3 年以内に AD へ移行していた。一方、健常者と予測された MCI 患者の AD 移行率は 13.2% だった。以上の結果から複数のデータを統合することにより MCI 患者の AD 移行率を予測した。

キーワード アルツハイマー病, 軽度認知障害, 発症予測, 異種混合学習, 決定木

1. はじめに

認知症は現在世界中で約 5,000 万人が罹患していると言われ、その数は 2030 年には 1 億人を超えると試算されている[1]。認知症の中でもアルツハイマー病 (AD: Alzheimer's Disease) は過半数を占めており、その克服は喫緊の課題である。現在 AD の根本治療薬開発は重大な副作用によりことごとく失敗に終わっている。そこで AD 発症の前段階である軽度認知障害 (MCI: Mild cognitive impairment) を伴う患者を効率よく見つけ出し、早期治療介入する取り組みが進んでいる。しかし、すべての MCI 患者が必ずしも AD へ進行するとは限らず、その病態は不均一である。そこでは我々は異種混合学習技術を用い、fMRI による脳容積データや髄液バイオマーカーデータといった様々な客観的なデータを統合することで MCI 患者の AD 移行率を予測した。

2. 方法

1) 異種混合学習技術

我々は様々なデータを統合した学習器を生成するために異種混合学習技術を用いた[2]。本手

法は複数の学習器を決定木に基づいて統合する階層混合エキスパートモデルの一種である。異種混合学習では異なる規則性をもつデータ (例えば、線形な性質と非線形な性質が混在するデータ) を似た性質を示すグループに分け、それぞれのグループの性質に応じた適切な予測モデルを生成する。

2) 決定木の構成

決定木はエキスパートと呼ばれる末端ノードと、ゲート関数と呼ばれるそれ以外のノードによる二分木に基づいて構成される。ゲート関数はある説明変数の閾値に従い入力サンプルを左または右の枝に振り分ける。エキスパートは学習器であり、ここでは二項ロジスティック回帰式を用いた。構築した決定木に基づき入力サンプルを適切なエキスパートに分類し判別を行う。

3) 条件設定

決定木の深さは 3 から 6 に設定した。パラメータ推定に用いる EM アルゴリズムは局所最適解に陥る可能性がある。局所最適解を回避するために、各深さで初期値が異なる 30 回の試行を行った。こ

の中で accuracy が最も高い決定木とモデルパラメータを採用した。Accuracy の計算は後述する。

4) データセット

我々は Alzheimer's Disease Neuroimaging Initiative (ADNI) に登録されている健常者 272 人、MCI 患者 474 人、AD 患者 152 人に関する fMRI による脳容積データ(海馬, 嗅内皮質, 全脳, 脳室, 白質高信号域), 髄液バイオマーカー ($A\beta$ 量, タウ量, リン酸化タウ量, タウ/ $A\beta$ 比率, リン酸化タウ/ $A\beta$ 比率)を用いた。本研究は大阪大学倫理審査委員会(承認番号: 767-2)の承認を得ている。

5) 精度評価

我々は健常者と AD 患者のデータセットを学習用データとして用い決定木およびモデルパラメータを決定した。次に MCI 患者のデータセットを評価用データとして用い, 学習用データで決定した決定木およびモデルパラメータに基づいて健常者か AD 患者に分類した。AD と予測された MCI 患者のうち, 3 年以内に AD にコンバートした MCI を true positive, コンバートしなかった MCI を false positive とし, また, 健常者と予測された MCI 患者のうち, 3 年以内に AD にコンバートした MCI を false negative, コンバートしなかった MCI を true negative として分類の判別能を検証した。

3. 結果

1) 各深さにおける平均 accuracy の比較

それぞれの深さにおける平均 accuracy を比較した結果, 深さ 6 で 0.748 と最も高かった(Table.1)。

Table.1 各深さにおける平均判別能

	Sensitivity	Specificity	Precision	Accuracy
深さ3	0.724	0.750	0.546	0.742
深さ4	0.717	0.754	0.548	0.743
深さ5	0.717	0.752	0.547	0.742
深さ6	0.709	0.764	0.556	0.748

2) MCI のサブタイピング

最も平均 accuracy が高かった深さ 6 に注目し, この中で最も accuracy が高い決定木を用い, それぞれのエキスパートに分類された MCI 患者を判別した。用いた木は accuracy が 0.776 で 5 つのゲート関数と 6 つのエキスパートから構成された(Fig.1)。

3) AD への移行率

各エキスパートの AD への移行率を調べた結果, AD に分類された MCI 患者は 60.1% (=98 人/163 人) が 3 年以内に AD へ移行していた一方, 健常者に分類された MCI 患者の AD 移行率は 13.2% (41 人/311 人) だった(ログランク検定 $p=1.28 \times 10^{-14}$)。エキスパート別の移行率を Fig.2 に示す。

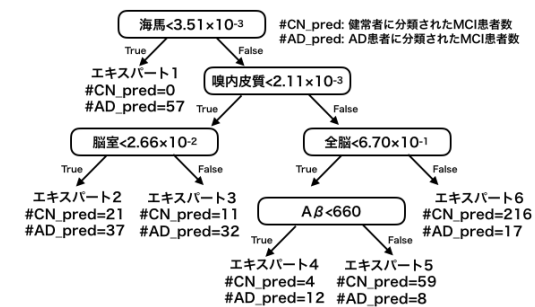


Fig.1 判別に用いた決定木

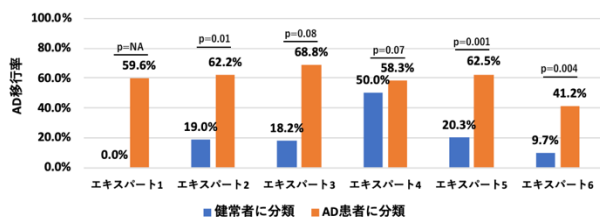


Fig.2 各エキスパートの AD 移行率

4. 考察

脳萎縮が進んでいるエキスパート 1 ではすべての MCI 患者が AD と予測された一方で, 脳萎縮が進んでいないエキスパート 6 や AD 原因タンパクである $A\beta$ タンパクが脳に蓄積せず髄液に露出しているエキスパート 5 では健常者と予測されている人数が多く妥当な判別結果であると考えられる。

5. 結語

様々な指標を異種混合学習技術により統合し解析を行った結果, AD に分類された MCI 患者は高い確率で AD へ移行した。

参考文献

- [1] World Alzheimer Report 2015: The Global Impact of Dementia. [https://www.alz.co.uk/research/world-report-2015]
- [2] 藤巻 遼平, 森永 聡: ビッグデータ時代の最先端データマイニング, NEC 技報, 65, 81-85, 2012.