

認知症病因物質がもたらす会話内容への影響分析と 発症前アルツハイマー病の予測

岡田智哉¹ 入部百合絵¹ 横井克典² 中村昭範³ 北岡教英⁴ 勝野雅央⁵

¹愛知県立大学 大学院情報科学研究科 ²国立研究開発法人 国立長寿医療研究センター 脳神経内科部

³国立研究開発法人 国立長寿医療研究センター 認知症先進医療開発センター

バイオマーカー開発研究部 ⁴豊橋技術科学大学 情報・知能工学系 ⁵名古屋大学 大学院医学系研究科

im222003@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

概要

アルツハイマー型認知症(AD)の早期発見を目的とした研究の多くは、神経細胞死が起こり始めた軽度認知症を早期段階としている。しかし、神経細胞が死後に回復しないことを考えると、原因物質の蓄積が認められるプレクリニカルAD期での予測が必要である。そのプレクリニカルADの診断には侵襲性を有する高価な検査が必要である。本研究では、雑談対話音声を用いてプレクリニカルADを識別した。BERTやRoBERTaを特徴量抽出器として用い、ADの識別に有用な言語特徴量と組み合わせた識別器を構築することで、F値0.8、Recall1.0を得ることができ、雑談対話音声からのプレクリニカルAD早期予測への応用が期待できる。

1 はじめに

近年、認知症者は増加し、介護や病院などの逼迫による社会問題を引き起こしている。認知症全体の60-70%を占めるAD[1]は早期治療により、症状の緩和や遅延が可能である。昨年、日本でも承認されたレカネマブ[2]は投与対象をADの早期段階としており、依然として早期発見が求められている。

ADの早期発見を目的とした研究の多くは軽度認知症(MCI)を対象としている[3, 4]。しかし、MCIは神経細胞死が起こり始めた段階であり、既に死滅した神経細胞は回復しないと言われている。そのため、MCI期以前の検出が必要である。そこで注目したのが、健常者(HC)でありつつも、ADの原因物質であるアミロイドβ(Aβ)の蓄積が認められるプレクリニカルAD[5]である。

現在プレクリニカルADの検出方法には、アミロイド・イメージング(ABPET)やタウ・イメージング(タウPET)がある。これらのPET検査はプレクリニカルADの検出に有効な一方で、侵襲性を有し、検

査費が高額などのデメリットを有している。そのため、プレクリニカルADを検出できる安価かつ非侵襲性を有する検査が求められている。そこで、ADの症状が表出される情報の中で、日常生活から容易に収集が可能であり、収集する際に受検者への負担の少ない自由会話音声に着目した。

ADの発見やADの早期発見に自由会話音声を利用した先行研究は多い。しかしそれらの多くは、ある絵を説明しているなど、一定のタスク実行時の音声を利用している[3, 4]。そのため、認知障害の表れていないプレクリニカルADの識別には適していない。そのため、用いる音声は、日常の会話に近い雑談対話音声望ましい。

上記背景の下、本研究は雑談対話音声から得たテキストを用いたプレクリニカルADの識別を目的とする。まず、プレクリニカルAD期の認知機能検査の結果や言語特徴量への影響を調査した上で、言語特徴量を用いてプレクリニカルADを識別する。さらに、BERT(Bidirectional Encoder Representations from Transformers)とその派生モデルRoBERTa(A Robustly Optimized BERT Pretraining Approach)を用いた手法、BERT、RoBERTaとアンサンブル学習を組み合わせた手法の2種類を実施した。次に、精度向上を目標に、BERTやRoBERTaを特徴量抽出器として用い、言語特徴量と組み合わせて識別器の入力とした識別モデルを新たに提案する。

なお、本研究は名古屋大学大学院医学系研究科生命倫理委員会の承認を得て実施した。

2 収集データ

2.1 収集した認知機能検査と書き起こし

本研究では、「認知症関連疾患の会話への影響—自然言語処理を用いた検討[6]」により収集したデータを用いている。複数の認知機能検査の結果と検

査時の音声、および専門医が診断した A β PET とタウ PET の結果を収集した。実施した認知機能検査は Alzheimer's Disease Assessment Scale(ADAS), GD S-15-J, Mini-Mental State Examination(MMSE), CD R, Logical memory(LM1, LM2), Japanese Adult Reading Test(JART), Frontal Assessment Battery(FAB), 数唱(D span F, D span B)の 10 種類である。

実施した検査の 1 つである ADAS は自由会話を通して口語言語能力を評価する項目がある。この評価項目では、仕事や趣味などに関する会話が交わされており、日常会話と内容が近い。そのため本研究では、この自由会話音声を人手で書き起こし、被験者の発話部分のみを雑談対話テキストとして用いた。

A β PET 陽性から AD 発症までは 20 年以上の歳月がかかるため、50 歳以上を対象とした。さらに、AD 以外の神経変性疾患の可能性のあるデータは除外している。

2.2 ラベルの付与

本研究では、AD ラベルとプレクリニカル AD ラベルを付与している。AD ラベルは、複数の認知機能検査の結果を基に AD, MCI, HC のいずれかを付与した。プレクリニカル AD ラベルは、HC の被験者に対して、PET 検査の結果を基にプレクリニカル AD 後期、プレクリニカル AD 前期、正常のいずれかを付与した。以降、本稿内の図表は、プレクリニカル AD 後期、プレクリニカル AD 前期、正常を、A β +tau+, A β +tau-, A β -tau-と表記する。被験者の年代、性別、ラベルの分布を表 1 に示す。

3 言語特徴量の抽出と検定

AD の症状の多くは、記憶障害から症状が現れ始

表 1 被験者の内訳

AD ラベル	男女	50代	60代	70代	80代	合計
AD	男	1	4	4	1	10
	女	3	0	5	2	10
MCI	男	0	1	4	0	5
	女	0	1	1	3	5
HC (A β +tau+)	男	0	0	12	2	14
	女	1	0	6	3	10
HC (A β +tau-)	男	0	2	3	0	5
	女	0	0	0	1	1
HC (A β -tau-)	男	1	3	7	4	15
	女	0	1	7	1	9
合計		6	11	49	18	84

※AD ラベルの()内はプレクリニカル AD ラベル

め、徐々に時間や場所などが認識できなくなる見当識障害が現れる[7]。先行研究[8]は、認知機能の変化が主観的に経験される前のプレクリニカル AD の状態でも言語機能に関連があることが明らかになっている。そのため、本研究でも GiNZA[9]を用いて言語特徴量の抽出を行った。抽出した特徴量を表 2、表 3 に示す。

3.2 認知機能検査と言語特徴量の分析

言語特徴量あるいは現行の認知機能検査がプレクリニカル AD の検出に有効であるか否かを明らかにするため、プレクリニカル AD ラベルに基づく 3 群間に対して、認知機能検査あるいは言語的特徴量を用いて検定を実施した。言語特徴量については AD ラベルに基づく 3 群間への検定も行った。有意差の確認には、事前検定に Kruskal-Wallis 検定 (H 検定)、事後検定に Steel-Dwass 検定を採用した。事前検定および事後検定により、有意水準 5% で有意差の認められた個数を表 4 と表 5 に示す。認知機能検査の中で、プレクリニカル AD ラベルに基づく 3 群に有意差が認められたのは ADAS のみであった。

表 4 より、プレクリニカル AD に基づく 3 群について、複数の言語特徴量と認知機能検査の中では認知機能検査のみに有意差が認められた。言語特徴量は抽出に ADAS の音声の一部を用いたため、ADAS よりも容易に収集できる。ADAS よりも言語特徴量の方が収集も容易であり、有意差が認められる数も多いため、プレクリニカル AD の識別には認

表 2 品詞割合に関わる特徴量

特徴量名	品詞
総語数割合	名詞, 形状詞, 動詞, 形容詞, 助動詞, 助詞, 連体詞, 副詞, 接続詞, 感動詞, フィラー, 接頭辞, 接尾辞, 格助詞, 普通名詞, 代名詞, 固有名詞, 数詞, 助動詞語幹, 普通名詞+固有名詞, 代名詞-人系, 代名詞-こそあど系
総発話数割合	上記のもの, 総語数
総名詞数割合	普通名詞, 代名詞, 固有名詞, 数詞, 助動詞語幹, 普通名詞+固有名詞, 代名詞-人系, 代名詞-こそあど系, 異なり名詞

表 3 その他の特徴量

特徴量名
総語数, 総発話数, TTR(Type-Token Ratio), 文字単語親密度の平均と分散, 最大係り受け距離の平均と分散, 最大係り受け深度の平均と分散

表 4 プレクリニカル AD ラベルに基づく 3 群の検定結果

	H 検定	Steel-Dwass 検定		
		A β -tau- と A β +tau+	A β -tau- と A β +tau-	A β +tau- と A β +tau+
認知機能検査	1	1	0	0
言語特徴量	13	5	0	6

表 5 AD ラベルに基づく 3 群の検定結果

	H 検定	Steel-Dwass 検定		
		HC と AD	HC と MCI	MCI と AD
言語特徴量	1	1	0	0

認知機能検査よりも言語特徴量が有効な可能性がある。

また、正常群とプレクリニカル AD 前期群間に有意差の認められた特徴量はないが、他の 2 群間には有意差の認められる特徴量が多い。これらの結果より、プレクリニカル AD 後期から言語特徴量への影響が表出する可能性が高い。そのため、言語特徴量は正常+プレクリニカル AD 前期群とプレクリニカル AD 後期群の識別に有効である可能性が示された。

表 4 と表 5 より、有意差の認められた言語特徴量の数は、AD ラベルに基づく 3 群よりもプレクリニカル AD ラベルに基づく 3 群の方が多い。そのため、言語特徴量への影響は MCI/AD 期よりも、プレクリニカル AD 期の方が大きい可能性がある。

4 様々な手法での識別

4.1 言語特徴量を用いた識別

前章で認知機能検査よりも識別に有効な可能性が示された言語特徴量を用いて、正常+プレクリニカル AD 前期群とプレクリニカル AD 後期群、正常群とプレクリニカル AD 群の識別を行った。識別器は、Support Vector Classification(SVC), LogisticRegression, XGBClassifier を利用し、leave-one-person-out cross-validation (LOPOCV) 法により評価した。特徴量は言語特徴量に加えて、性別と年齢を使用した。ハイパーパラメータ最適化は optuna[10]を使用した。各識別群の F 値の最も高い識別結果を表 6 に示す。

表 6 4.1 節の手法での識別結果

識別群	識別器	F 値	正解率	再現率
A β +tau- と A β +tau+	Logistic Regression	0.68	0.70	0.71
A β -tau- と A β +tau+	XGBClassifier	0.73	0.69	0.77

表 6 より、正常+プレクリニカル AD 前期群とプレクリニカル AD 後期群よりも正常群とプレクリニカル AD 群の識別の方が効果的に予測しているといえる。3.2 節での考察とは異なる結果となった理由として、特徴量単独と各群間の検定と、複数の特徴量を組み合わせた各群の識別の違いが考えられる。一方で、正解率は 0.7 程度であるため、識別精度向上の余地がある。

4.2 BERT, RoBERTa を用いた識別

自然言語の事前学習モデル BERT とその派生モデル RoBERTa を用いて識別を行った。BERT のモデルには東北大学が公開した cl-tohoku/bert-large-japanese-v2[11](cl-tohoku モデル)を、RoBERTa のモデルには早稲田大学が公開した, nlp-waseda/roberta-large-japanese-seq512[12](nlp-waseda モデル), nlp-waseda/roberta-large-japanese-seq512-with-auto-jumanpp[13](nlp-waseda-jumanpp モデル)を用いた。学習データと検証データは LOPOCV 法を用いて分割し、Epoch 数は 20 とした。識別結果を図 1 に示す。

図 1 より、識別性能の一番高い Epoch では、言語特徴量を用いた識別と同等か、それ以上の性能で識別できていることがわかる。一方で、Epoch 毎に識別性能の振り幅が大きいことが問題である。

4.3 BERT, RoBERTa とアンサンブル学習を組み合わせた識別

BERT, RoBERTa を用いた識別では、Epoch により性能が不安定であったため、安定性を図る必要が

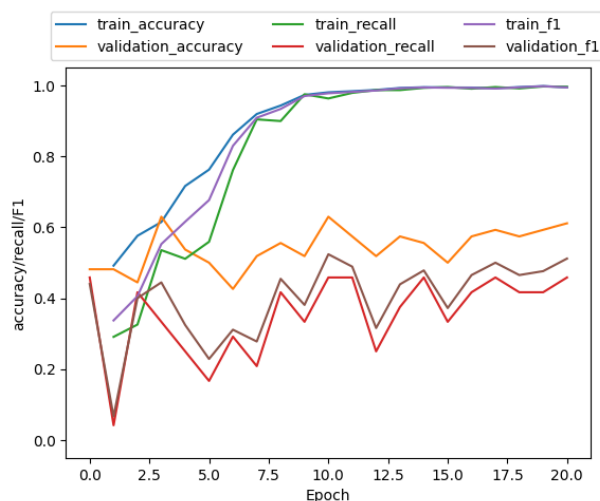


図 1 nlp-waseda-jumanpp モデルを用いた A β +tau- と A β +tau+ の識別結果

ある。先行研究から、BERT や ERNIE の識別結果に対してアンサンブル学習することが精度の安定化に貢献することが明らかになっている[14]。そこで、本研究でも BERT や RoBERTa を用いた識別結果に対してアンサンブル学習を採用した。本研究では各モデルの Epoch 毎の識別結果に対して多数決を行い、クラスラベルを予測する。各識別群の F 値の最も高い識別結果を表 7 に示す。

表 7 より、いずれの結果も言語特徴量を用いた識別性能を上回ることができなかった。理由としては、学習精度の低い Epoch がノイズになっている可能性が考えられる。

4.4 言語特徴量と BERT, RoBERTa を組み

合わせた識別

言語特徴量では抽出できない特徴を BERT や RoBERTa では捉えられることを期待し、BERT と RoBERTa を特徴量抽出器として活用し、言語特徴量を組み合わせる手法を新たに提案する。ただし、それらの特徴量を組み合わせるだけでは、識別器に入力する情報量として少ないと考えられる。そのため、BERT と RoBERTa の各モデルの出力値ではなく、出力前の活性化関数による変換前の値 (logit) を特徴量 (言語モデル特徴量) として用いた。識別器には、SVC, LogisticRegression, XGBoostClassifier を利用し、評価は LOPOCV 法を用いた。入力特徴量には、言語モデル特徴量, 言語特徴量, 性別, 年齢を使用し、ハイパーパラメータ最適化は optuna を利用した。さらに、4.3 より学習精度の低い Epoch がノイズとなる可能性が示唆されたため、ラッパー法の変数減少法を用いて特徴量を削減した。各識別群の F 値の最も高い識別結果を表 8, 正常群とプレクリニカル AD 群の F 値の最も高い識別の混同行列を表 9 に示す。

表 8 より、正常+プレクリニカル AD 前期群とプレクリニカル AD 後期群よりも正常群とプレクリニカル AD 群の識別の方が 2 群を効果的に予測しているといえる。さらに、表 6 の言語特徴量を用いた識別よりも性能が高いことが分かる。

表 7 4.3 節の手法での識別結果

識別群	言語モデル	F 値	正解率	再現率
Aβ+-tau- と Aβ+tau+	nlp-waseda-jumanpp	0.32	0.54	0.25
Aβ-tau- と Aβ+tau+-	nlp-waseda-jumanpp	0.69	0.59	0.8

表 8 4.4 節の手法での識別結果

識別群	言語モデル	F 値	正解率	再現率
Aβ+-tau- と Aβ+tau+	nlp-waseda	0.72	0.74	0.75
Aβ-tau- と Aβ+tau+-	nlp-waseda-jumanpp	0.8	0.72	1.0

表 9 正常群とプレクリニカル AD 群の F 値の最も高い識別時の混同行列

		予測	
		Aβ+tau+-	Aβ-tau-
正解	Aβ+tau+-	30	0
	Aβ-tau-	15	9

表 9 より、特に正常群とプレクリニカル AD 群の識別では、正解がプレクリニカル AD かつ正常に予測された人数が 0 人、つまりプレクリニカル AD の見落とし件数が 0 件である。よって、この提案手法が、受検者の負担の大きな PET 検査の必要性の有無の判定への活用と日常で交わす雑談からのプレクリニカル AD の候補の早期発見への貢献が期待できる。

5 おわりに

本研究では、雑談対話テキストを用いて、HC でありつつも AD の原因物質の蓄積が認められるプレクリニカル AD の識別を行った。

雑談対話音声の収集と抽出した言語特徴量を分析することで、下記について明らかとなった。

- 言語特徴量への影響は MCI/AD 段階よりも、プレクリニカル AD 段階の方が大きいプレクリニカル AD の識別には認知機能検査より言語特徴量の方が効果的である可能性が高い

また、言語特徴量や BERT, RoBERTa を用いた複数の識別手法を提案した。その結果、いずれの学習においても、正常+プレクリニカル AD 前期群とプレクリニカル AD 後期群よりも、正常群とプレクリニカル AD 群の 2 群を効果的に予測できた。特に、言語特徴量と自然言語の事前学習モデル RoBERTa を組み合わせた提案手法により、プレクリニカル AD 群の再現率 1.0 を得ることができた。以上より、受検者が PET 検査の必要性を測る事前検査としての活用、日常で交わす雑談からプレクリニカル AD の候補の早期発見への貢献が期待できる。

今後は、さらなるデータ拡充を図るとともに、音響特徴量も組み合わせることで、精度向上を目指す。さらに、自動書き起こしを使用した人手を必要としない手法で同様の精度を得られることも目指す。

参考文献

1. World Health Organization. Dementia. <https://www.who.int/en/news-room/fact-sheets/detail/dementia>, 2023-12-19 閲覧.
2. エーザイ株式会社. 「レケンビ® 点滴静注」(一般名:レカネマブ)について、日本においてアルツハイマー病治療薬として製造販売承認を取得 - ニュースリリース:2023年. <https://www.eisai.co.jp/news/2023/news202359.html>, 2023-12-19 閲覧.
3. Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, Brian MacWhinney. Alzheimer's Dementia Recognition Through Spontaneous Speech: The ADReSS Challenge. Proc. Interspeech 2020, pp.2172-2176, doi: 10.21437/Interspeech.2020-2571, 2020.
4. Saturnino Luz, Fasih Haider, Sofia de la Fuente, Davida Fromm, Brian MacWhinney. Detecting Cognitive Decline Using Speech Only: The ADReSSo Challenge. Proc. Interspeech 2021, pp.3780-3784, doi:10.21437/Interspeech.2021-1220, 2021.
5. 嶋田裕之. アミロイドPETとpreclinical AD. http://www.rouninken.jp/member/pdf/20_pdf/vol.20_03-20-03.pdf, 2023-12-19 閲覧.
6. 国立長寿医療研究センター, 受付番号 No.22TB8. <https://www.ncgg.go.jp/ncgg-kenkyu/ekigaku/22TB8.html>, 2023-12-19 閲覧.
7. 河野和彦, ぜんぶわかる認知症の事典—4 大認知症をわかりやすくビジュアル解説, 2016, 成美堂出版.
8. Ihab Hajjar, et al. Development of digital voice biomarkers and associations with cognition, cerebrospinal biomarkers, and neural representation in early Alzheimer's disease, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 15(1), doi : 10.1002/dad2.12393, 2023.
9. GiNZA: 日本語自然言語処理オープンソースライブラリ - Megagon Labs - リクルート AI 研究機関. <https://www.megagon.ai/jp/projects/ginza-install-a-japanese-nlp-library-in-one-step/>, 2023-12-19 閲覧.
10. 株式会社 Preferred Networks. Optuna. <https://www.preferred.jp/ja/projects/optuna/>, 2023-12-19 閲覧.
11. Hugging Face. cl-tohoku/bert-large-japanese-v2. <https://huggingface.co/cl-tohoku/bert-large-japanese-v2>, 2023-12-19 閲覧.
12. Hugging Face. nlp-waseda/roberta-large-japanese-seq512. <https://huggingface.co/nlp-waseda/roberta-large-japanese-seq512>, 2023-12-19 閲覧.
13. Hugging Face. nlp-waseda/roberta-large-japanese-seq512-with-auto-jumanpp. <https://huggingface.co/nlp-waseda/roberta-large-japanese-seq512-with-auto-jumanpp>, 2023-12-19 閲覧.
14. Jiahong Yuan, Yuchen Bian, Xingyu Cai, Jiaji Huang, Zheng Ye, Kenneth Church. Disfluencies and FineTuning Pre-trained Language Models for Detection of Alzheimer's Disease. Proc. Interspeech 2020, pp.2162-2166, doi: 10.21437/Interspeech.2020-2516, 2020