

自由会話のトピックモデルに基づいた軽度認知障害の検出

長江勇樹¹ 岡田智哉² 入部百合絵¹ 横井克典³ 中村昭範⁴ 北岡教英⁵ 勝野雅央⁶

¹ 愛知県立大学 情報科学科 ² 愛知県立大学 大学院情報科学研究科

³ 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

⁴ 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

⁵ 豊橋技術科学大学 情報・知能工学系 ⁶ 名古屋大学 大学院医学系研究科

is201063@cis.aichi-pu.ac.jp iribe@ist.aichi-pu.ac.jp

概要

先行研究では、話題が限定されたタスク型音声を用いて認知症傾向を検出していることが多い。本研究では、複数の認知機能検査および画像検査から専門医が診断した診断結果と雑談対話音声を用いて研究を行う。また、認知症検出の新たな特徴量として話題の類似度を提案する。話題の類似度と複数の言語特徴量に Kruskal Wallis 検定を施し、特徴量選択とデータ拡張を行った上で、アルツハイマー型認知症者 (AD)、軽度認知症者 (MCI)、健常者 (HC) の3群間を分類した。その結果、形容詞総発話数割合と対話間のトピックの類似度の分散に有意差が確認された。また、認知症の分類は Accuracy 0.658, Recall 0.7200 の精度を得た。

1 はじめに

2012年時点で、日本に住む高齢者の約4人に1人が認知症またはその予備軍であるとされており、以後も増加していくことが予想されている [1]。

認知症の中で最も患者数の多いADの主な診断方法には、CTやMRI検査など、脳の「画像検査」やSPECT(脳血流シンチグラフィ)やPET(ポジトロン断層撮影)など、脳の血流や代謝を調べる「機能画像検査」がある。これらの方法は、被験者の精神的・肉体的・経済面での負担が大きい。そのため、簡易的で被験者への負担の少ない検出方法が求められる。そのような背景の下、自由会話音声からAD、または認知症の初期状態に当たるMCIを検出する研究が活発化している。その多くは、ボストン失語症のクッキー・セフト絵 [2][3] など、比較的話題の絞られたタスク型の会話である。しかし、早期の検出を目指すならば、特定場面の自由会話音声ではな

く、日常生活で交わされる雑談対話音声から認知症を検出できることが望ましい。石原 [4] らは、高齢者の雑談対話音声から文章の複雑さに関係する特徴量を抽出しているが、ADの言語症状は多岐に亘るため、文章の複雑さを評価するだけでは十分でない。一方、Zhang [5] らは、トピックモデルを用いて特定の話題とうつ病の進行度の関連性を報告している。ADの症状には記憶障害があり、近似記憶が特に損なわれやすいため、会話内容の忘却により、対話相手との話題にズレが生じたり、急な話題転換などが発生する可能性がある。そのため、話題の一貫性をHCと比較することで、AD、MCIの症状を捉えることができる可能性が高い。

本研究では、雑談会話中の対話相手との話題の類似性に着目し、AD、MCIを識別する。

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 使用データについて

2.1 音声収録

本研究は、「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]」により得た音声一式を用いている。それらには、50代以上の男女82名を対象に、ADAS(Alzheimer's Disease Assessment Scale)を含む複数の認知機能検査による検査結果と雑談対話音声が含まれている。被験者の年齢と性別、認知症傾向の有無の分布を表1に示す。本研究では、ADASの音声の一部を手書き起こし、自由会話音声の書き起こし文として利用した。ADASには、自由会話を通して口語言語能力や言語の聴覚的理解などを評価する項目がある。その自由会話部分での会話内容は「趣味」、「仕事」、「年末の予定」

表 1: 被験者の年齢, 性別と認知症傾向

		AD	MCI	HC	合計
50代	男性	1	0	1	2
	女性	3	0	1	4
60代	男性	2	1	5	8
	女性	0	1	1	2
70代	男性	4	4	24	32
	女性	5	0	14	19
80代	男性	0	0	6	6
	女性	2	2	5	9
合計		17	8	57	82

などが中心であるため, 日常の雑談会話に近い。

2.2 認知症の判断基準

実施された認知機能検査は, ADAS, GDS15-J, MMSE(Mini-Mental State Examination), CDR, Logical memory(LM1, LM2), Japanese Adult Reading Test(JART), Frontal Assessment Battery(FAB), 数唱(DspanF, DspanB)の10種類である。本研究におけるAD, MCI, HCのラベルは, 複数の認知機能検査および画像検査から専門医が診断した結果である。

3 雑談対話音声に含まれるトピック分析と言語特徴量

3.1 抽出する言語特徴量の概要

収録した雑談対話音声から認知症の症状に関係があると考えられる特徴量を抽出し, 分析する。分析する特徴量は以下の通りである。日本語自然処理ライブラリGinzaを用いた言語特徴量は, 先行研究[4][7]を参考に抽出した。

BERTopicを用いて抽出する言語特徴量

- 応答の正確さ
 - 文章間のトピック類似度の平均, 分散

GiNZAを用いて抽出する言語特徴量

- 語彙
 - 各品詞の総語数割合, 総発話数割合
 - TTR (Type Token Ratio)
 - 日本語学習語彙レベル
- 言い淀み・迂言
 - フィラーの総語数割合, 総発話数割合
 - 異なり名詞割合
- 文章の複雑さ
 - 最大係り受け距離, 深度の平均, 分散

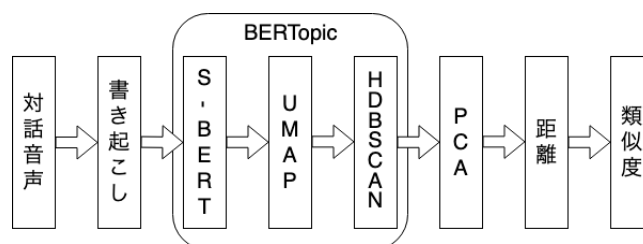


図 1: トピック類似度の計算手順

3.2 文章間のトピック類似度の平均, 分散

ADの主な症状は記憶障害や, 時間や場所などが認識できなくなる見当識障害などがあり, 多くは記憶障害から症状が現れるが, 病気の自覚はない。ADでは, 個人的な体験や社会的なできごとを記録するエピソード記憶が障害され, 特に, 数分前から数日前に記憶された近似記憶が損なわれる。即時記憶能力の低下に伴う, 会話の内容の不自然な変化を検出する指標として, 文章から成分が各トピックに属する確率によるベクトルをトピックモデルBERTopicを利用して表現し, インタビュアーの言葉に対する高齢者の応答の間でトピックの類似度を算出する。本研究では図1の示す手順を経て, 文内に属する各トピックの確率のベクトル表現を抽出する。

3.2.1 BERTopic

BERTopicは, sentence-BERTやspacyなどによって生成される文章の分散表現を利用したトピックモデルである。BERTopicはモジュール性を有しており, 本研究では以下のモジュールを使用した。

sentence-BERT BERTをベースに, 文章の入力からベクトル表現を取得する。本研究ではデータ数が少ないため, 日本語用に学習が行われたモデルbert-base-japanese-v3に対し全ての文章データを用いて訓練を行った。

UMAP(Uniform manifold approximation and projection) データの情報をできるだけ損なわないように, 次元削減を行う手法。次元が減ることにより, クラスタリングが容易となる。

HDBSCAN(Hierarchical Density-Based Spatial Clustering of Applications with Noise) データ群をクラスタリングする手法の一つ。これにより話題のクラスに文章をクラスタリングする。

表 2: 特徴量の分析結果

特徴	分散			p 値
	AD	MCI	HC	
形容詞総発話数割合	0.139	0.244	0.172	0.031*
チェビシェフ距離の分散	0.296	0.305	0.315	0.027*

$p < 0.05$:*

3.2.2 PCA(Principal Component Analysis)

PCA は複数の次元を統合して新たな指標を生み出すことで主成分を探る手法である。本研究では PCA により同次元の多次元データを生成し、累積寄与率 50%以下の主成分を抽出した。

3.2.3 類似度の算出方法

本研究では、文章間の類似度を計算する際に、以下の 4 種類の手法を用いた。x, y はベクトル、n は x, y の次元数である。

ユークリッド距離 2つのベクトルの直線での最短距離を計算する。

$$d_e(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$$

マンハッタン距離 2つのベクトルの各次元から距離を求め、合計した値とする。

$$d_m(x, y) = \sum_{i=0}^n |x_i - y_i|$$

チェビシェフ距離 2つのベクトルの各次元から距離を求め、その中で最大の値をとる。

$$d_c(x, y) = \max(|x_0 - y_0|, |x_1 - y_1|, \dots, |x_n - y_n|)$$

$\max(x_1, \dots, x_n)$ は、 x_1 から x_n の内最も大きい値を解とする。

コサイン類似度 2つのベクトルの角度から類似度を計測する。

$$Sim(x, y) = \frac{xy}{\|x\| \cdot \|y\|}$$

3.3 雑談対話音声の特徴量の分析

対話会話の各言語特徴量に対して AD, MCI, HC の 3 群間で有意差の有無を検定し、AD, MCI の識別に有効な特徴量を明らかにした。有意差の確認には、事前検定として Kruskal Wallis 検定を、事後検定として steel Dwass 検定を使用した。検定の結果、事前検定に 5%の有意差が認められた特徴量を表 2, 図 2, 3 に示す。

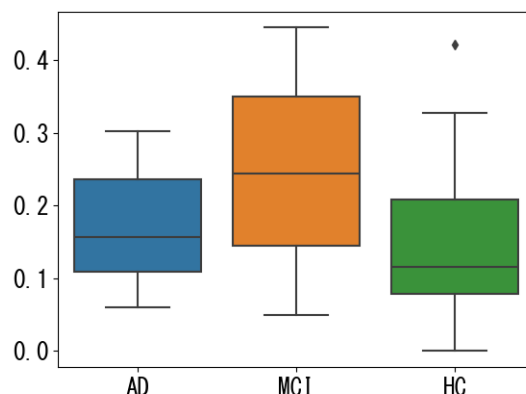


図 2: AD,MCI,HC の分布 (形容詞総発話数割合)

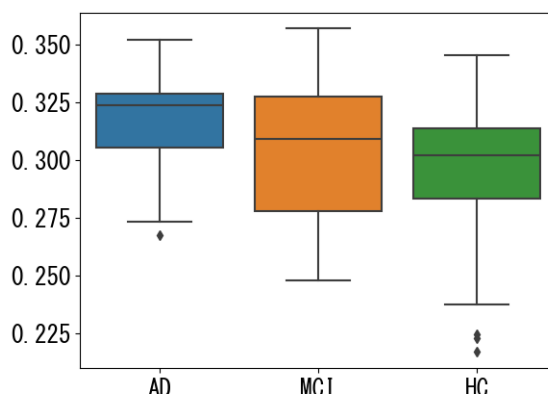


図 3: AD,MCI,HC の分布 (チェビシェフ距離の分散)

3.4 3 クラス間における言語特徴量の検定結果の考察

形容詞の総発話数割合は AD, MCI は HC と比べて有意に増加している。AD の症状には、言葉の意味や一般知識などの学習して得た知識を記憶する、意味記憶の能力が徐々に低下する記憶障害がある。この障害によって、表現に適した語句を思い出せず、平易な形容詞に言い換える会話が増加したのだと考えられる。チェビシェフ距離についても AD, MCI が有意に増加している。距離の値や分散値が大きいほど、対話相手との話題の相違が大きいことになる。ズレや AD の記憶障害は、過去数分から数日の記憶である近似記憶が、特に損なわれやすい。近似記憶障害により、被験者が直前の会話を忘却してしまうことで、インタビュアーとの話題にズレが生じ、トピック距離が増加したと考えられる。

表 3: 先行研究の特微量による分類結果

Accuracy		0.719		
	AD	MCI	HC	
Recall	0.294	0.125	0.929	
F-measure	0.370	0.181	0.841	
識別のクラス				
	AD	MCI	HC	
実際のクラス	AD	5	1	11
	MCI	2	1	5
	HC	3	1	53

表 4: トピック類似度を含む特微量による分類結果

Accuracy		0.634		
	AD	MCI	HC	
Recall	0.764	0.370	0.631	
F-measure	0.520	0.428	0.720	
識別のクラス				
	AD	MCI	HC	
実際のクラス	AD	13	0	4
	MCI	2	3	3
	HC	18	3	36

4 MCI の検出

先行研究の特微量 35 次元と、8 つのトピック類似度を加えた 43 次元の 2 つの条件で、AD, MCI, HC の 3 値分類を行い、分類結果の比較を行った。

4.1 分類方法

特微量選択手法 RFECV(Recursive Feature Elimination with Cross Validation) を用いて特微量を選択し、leave-one-out 法により評価した。RFECV は特微量選択手法 RFE(Recursive Feature Elimination) と交差検証 CV(Cross Validation) を組み合わせた手法である。AD, MCI は HC と比べて人数が少ないため、SMOTE を用いてデータ拡張をした上で、SVM(Support Vector Machine) の 4 つのカーネルによる三値分類を行った。精度の最も良いカーネルを用いた結果を表 3, 4 に、分類の際に選択された特微量を表 5 に示す。なお、先行研究の特微量では線形カーネル(表 3), 全特微量では多項式カーネル(表 4) が最良であった。

4.2 分類結果の考察

先行研究の特微量を用いた分類では、Accuracy が高い一方で AD, MCI の Recall と F-measure は低く、

表 5: 特微量選択の結果(トピック類似度を含む特微量)

形容詞総語数割合	助動詞総語数割合
助詞総語数割合	代名詞総語数割合
格助詞総語数割合	形容詞総発話数割合
固有名詞総発話数割合	格助詞総発話数割合
日本語学習語彙レベル	最大係り受け距離の分散
最大係り受け深度の分散	マンハッタン距離の分散
チェビシェフ距離の平均	チェビシェフ距離の分散

HC へ偏った分類を行っていることがわかる。対して、トピックに関する特微量を加えた分類では、AD と HC の 6 割以上が分類できている。特に、AD の再現率は 47% も改善されている。一方、MCI の分類については低い結果となった。MCI は軽度の認知障害はあるが一般的な認知機能には問題のない状態 [8] を指し、AD と比べて識別の難易度が高い。加えて、本研究の MCI のデータ数は 8 名と少なく、高精度に AD, MCI 検出を行う上でデータ数の少なさが 1 つの課題である。一方、認知障害を検出するという観点から、表 4 の認知症状の有無を基準とした、AD+MCI と HC の 2 値分類も算出した、そのような 2 クラス間では Accuracy 0.658, Recall 0.7200, F-measure 0.5625 を得た。また、全特微量から RFECV により選択された特微量は、表 5 の 14 次元であった。話題に関する 3 つの特微量が選択され、分類精度が向上していることから、話題の距離を認知症検出に用いる有用性を示せた。

5 まとめ

本研究では、複数の認知機能検査より専門医が診断した認知症診断結果と雑談対話音声を用いて研究を行った。また、新たな特微量として話題の類似度を提案し、先行研究の特微量とともに KruskalWallis 検定を行い、有意差のある特微量を分析した。その結果、形容詞総発話数割合とトピックのチェビシェフ距離の分散に対して有意差が認められた。また、AD, MCI, HC の 3 群間で分類を行った。トピック類似度を用いることで、AD の再現率は 47% も向上した。以上の結果より、文間のトピック類似度を用いることが認知症識別に有効であることが示された。一方、認知症者の、特に MCI のデータ数の少なさによる問題が浮き彫りとなった。今後は、新たな認知症者の音声データの収集を続けつつ、認知機能の低下がもたらす、新たな言語的特徴の解明を目指す。

参考文献

- [1] 厚生労働省老健局: 知症施策の総合的な推進について (参考資料), 入手先 <<https://www.mhlw.go.jp/content/12300000/000519620.pdf>>, 2019.
- [2] Luz, S. et al. Alzheimer's Dementia Recognition Through Spontaneous Speech: The ADReSS Challenge. Proc. Interspeech 2020, pp.2172-2176, doi: 10.21437/Interspeech.2020-2571, 2020.
- [3] Luz, S. et al. Detecting Cognitive Decline Using Speech Only: The ADReSSo Challenge. Proc. Interspeech 2021, pp.3780-3784, doi: 10.21437/Interspeech.2021-1220, 2021.
- [4] 石原颯人, 入部百合絵, 北岡教英: 係り受け距離に着目した雑談対話からの認知症疑い検出, 電子情報通信学会論文誌 D Vol.J104-D No.4 pp.357-367, 2021.
- [5] Yuezhou Zhang, Amos A Folarin, Judith Dineley, Pauline Conde, Valeria de Angel, Shaoxiong Sun, Yatharth Rangan, Zulqarnain Rashid, Callum Stewart, Petroula Laiou, Heet Sankesara, Linglong Qian, Faith Matcham, Katie M White, Carolin Oetzmann, Femke Lamers, Sara Siddi, Sara Simblett, Björn W. Schuller, Srinivasan Vairavan, Til Wykes, Josep Maria Haro, Brenda WJH Penninx, Vaibhav A Narayan, Matthew Hotopf, Richard JB Dobson, Nicholas Cummins, RADAR-CNS consortium: Identifying depression-related topics in smartphone-collected free-response speech recordings using an automatic speech recognition system and a deep learning topic model, 2023.
- [6] 国立長寿医療研究センター, 受付番号 No.22TB8, 入手先 <<https://www.ncgg.go.jp/ncgg-kenkyu/ekigaku/22TB8.html>> (参照 2024-01-10).
- [7] Veronika Vincze, Martina Katalin Szabó, Ildikó Hoffmann, László Tóth, Magdolna Pákáski, János Kálmán, Gábor Gosztolya: Linguistic Parameters of Spontaneous Speech for Identifying Mild Cognitive Impairment and Alzheimer Disease, Computational Linguistics (2022) 48 (1): 119-153, 2022.
- [8] 河野和彦監修: ぜんぶわかる認知症の事典, 成美堂出版, 2016.