語りに基づく認知症傾向判別

晩 鴻翔 [1]　岡崎 直観 [1]
[1] 東京工業大学
{hongxiang.wan,okazaki} at nlp.c.titech.ac.jp

1 はじめに

認知症とは「生後いったん正常に発達した種々の精神機能が慢性的に減退・消失することで，日常生活・社会生活を営めない状態」である*1．アルツハイマー病を代表とした認知症は不可逆的な脳疾患であり，その治療は困難である．認知症になると，治療費だけではなく，介護にも費用がかかってしまう．世界最速で高齢化が進んでいる日本では，認知症高齢者の増加が医療・社会問題となっている．今後，認知症患者の者の数は急速に増加し，2060 年に 850 万人に達すると予測され，国が負担する医療費を含む社会コストは 24 兆 2,630 億円となると推計されている [15]．しかし，初期段階で発見できれば，認知症の進行を抑制し，社会的負担を減らすことができる．したがって，認知症傾向の早期診断が重要である．

現在，医療機関での主な診断方法として用いられているのが脳画像検査である．例えば，Magnetic Resonance Imaging (MRI) 検査や Computed Tomography (CT) 検査である．それらの方法を利用することで認知症の診断ができるが，一定の身体拘束時間が必要であるため，日常的な受診は難しい．そのため，更に簡易的な検査法を開発する必要がある．

近年，自然言語処理技術の発展とともに，言語能力に基づく認知症の迅速かつ簡易的な診断が行えるスクリーニング技術が注目されている．通常，認知症初期段階，いわゆる軽度認知障害（MCI）の段階では，記憶，特に言語能力を司る短期記憶に障害があると考えられるため [13]，認知症に関連する言語的な特徴を特定する研究がしばしば行われている．英語圏では，カーネギーメロン大学が収集した画像説明課題における健常者と認知症患者の大量の音声と書き起こしデータ DementiaBank*2が多くの認知症に関する先行研究に用いられている．日本でも近年対照群付き高齢者コーパス*3が公開され，日本語における認知症スクリーニング研究も可能となった．柴田らはこのデータセットの書き起こしテキストから，語彙難易度，意味密度といったテキスト特徴を抽出し，機械学習による分類を行ったところ，最大 85% の精度を得たことを確認した [11]．しかし，このデータセットに沿って認知症のスクリーニングを行う場合，被験者は指定されたタスクを解く必要がある．そこで，日常会話からの抽出が可能で，タスクに依存しない特徴として，本研究では音声特徴の効果も検証する．

*1 https://www.mhlw.go.jp/kokoro/speciality/detail_recog.html
*2 https://dementia.talkbank.org/

2 関連研究

前述のように，英語圏では DementiaBank が存在するため，多くの認知症判別実験が行われてきた．DementiaBank で被験者の発話の収集に用いられたのが cookie-theft [5] という画像である．被験者は絵について，思ったこと全てを説明するように指示される．そして，その発話は録音され，人手により書き起こされる．先行研究では，認知症自動診断モデルを構築するために，音声特徴，テキスト特徴及びその組み合わせを特徴量として用いる手法が多く提案された．

韻律の特徴 (prosodic features) によって，認知症患者を判別できることも報告された [4]．Hoffman らの実験により，ポーズ頻度と話す速度（ポーズを含めた 1 秒あたりの音素数）は認知症の判断の指標として用いることができ，認知症症状の初期，中期，後期の判別にも有効であることが示された．他にも，話す速度はミニメンタルステート検査（MMSE: Mini Mental State Examination）の結果と正の相関をもつことが報告された [6]．また，認知症患者は非文法的なポーズがより多く出現する傾向があることもわかっている [3]．König らは認知症患者，軽度認知症患者，健常者から音声上の特徴である Mel-Frequency Cepstrum Coefficients (MFCC) を抽

*3 https://www.gsk.or.jp/catalog/gsk2018-a

出し，Support Vector Machines (SVM) による分類を行ったところ，健常者と軽度認知症患者，健常者と認知症患者，軽度認知症患者と認知症患者それぞれの識別精度が，それぞれ 79%，87%，80% を達成した [1]．

また，多くの先行研究では，音声の特徴とテキストの特徴を組み合わせた実験を行っている．Fraser らは認知症患者，健常者から構成されるコーパス (DementiaBank) からテキストの特徴（例えば品詞の頻度や構文的複雑さ，語彙の量）と音声の特徴（MFCC など）を抽出し，ロジスティック回帰による分類を行ったところ，81% の精度を達成した [14]．また近年では深層学習が注目され，Karlekar らは，DementiaBank を用いて，CNN-RNN モデルによる分類を行い，91.1% の精度を実現した [7]．

3 特徴抽出

本研究では，対照群付き高齢者コーパスを用いて，既存研究で有効であると報告された音声とテキストの特徴量を抽出し，機械学習による分類実験を行う．

3.1 音声特徴

音声の特徴量の抽出には，日本語連続音声認識システム Julius[*4]と Python のライブラリ python_speech_features を用いた．

3.1.1 韻律の特徴 (Prosodic features)

Forced alignment とは，書き起こしテキストと音声ファイルを与えることで，書き起こしテキストの内容に従って音声の位置付けを行うプロセスである．Julius を用いて，音声に音素単位でラベル付けを行い，発話のポーズ頻度，スピーチテンポといった特徴量を抽出した．結果を図1に示す．発話の流暢さを表現するには，「あのー」「えーと」といった言い淀みの情報も重要であるが，テキストに書き起こされていないため，julius によるラベル付けだけではそのデータを取得することができない．そのため，さらに Praat[*5]を用いて，手作業で音声に言い淀みのラベル付けを行い，その有声区間をフィラーポーズとした．区別のため，前述の無声区間をサイレンスポーズとする．サイレンスポーズとフィラーポーズを合わせてポーズとする．このようにして得られたアライメント情報に基づき，下記の特徴量を算出した．

[*4] http://julius.osdn.jp/
[*5] http://www.fon.hum.uva.nl/praat/

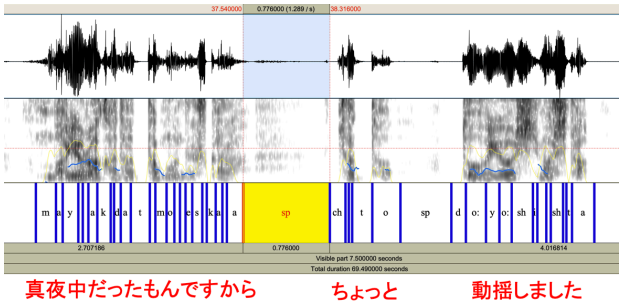

図1: 音声への音素単位でのラベル付け

- 各ポーズ数：発話時間 1 秒あたりのサイレンスポーズ数，フィラーポーズ数とポーズ数．
- 各ポーズの平均持続時間：サイレンスポーズ，フィラーポーズ，ポーズの平均持続時間．
- 各ポーズの割合：合計発話時間に対するサイレンスポーズ，フィラーポーズ，ポーズのそれぞれの持続時間の割合．
- スピーチテンポ：1 秒あたりの音素数 (ポーズの持続時間を含む)．
- アーティキュレーション率：1 秒あたりの音素数 (ポーズの持続時間を除外する)．
- ロングサイレンスポーズ数：発話時間あたりの，1 秒以上持続したサイレンスポーズの数．

3.1.2 MFCC

MFCC とは，音声のスペクトル包絡（声道成分に由来した周波数特性）を表現できる対数ケプストラムの低次成分に対して，ヒトの周波数知覚特性を考慮した重み付けをした特徴量である．この特徴量は広く話者認識や音声認識などに応用されている．本研究では，各フレームの 13 次元 MFCC の平均，分散，前後 2 フレームの MFCC の傾き，尖度，歪度，および平均の尖度と歪度を算出した．

3.1.3 基本周波数とフィルターバンクエネルギー

基本周波数とは，信号を正弦波の合成で表したときの最も低い周波数成分の周波数を意味する．基本周波数の変化は単一の音素，単語，または発話の範囲にわたって発生するため，イントネーションの変動の指標として扱われている．この変動が少ない場合，発話はフラットであると考えられている．フィルターバンクエネルギーとは信号のスペクトルエネルギーを表す特徴量である．本研究では各フレームの基本周波数の平均，分散，および

13 次元フィルターバンクエネルギーの平均，分散を算出した．

3.2 テキスト特徴

テキスト特徴量は，柴田らの実験を再現したものであるため，各特徴の抽出方法は原論文を参照されたい．日本語形態素解析器 MeCab[*6] を用いて，名詞，動詞，形容詞と副詞を抽出し，基礎特徴，意味密度，Bag of Words と分散表現を算出した．

3.2.1 基礎特徴

- Token：コーパス中の単語数．
- Type：コーパス中の単語の異なり数．
- Type Token Ratio (TTR)：コーパスの語彙量．
- 語彙難易度 (JEL: Japanese Educational Level)：日本語学習辞書[*7] に収載されている語彙レベルを用いて，コーパスの語彙難易度を評価する．
- 潜在語彙量 (PVS: Potential Vocabulary Size)：潜在語彙量 [2] は，ジップ則を用いて，ある人物が無限時間発話した際に使用すると予測される語彙量である．
- 品詞割合：コーパス中の名詞，動詞，形容詞および副詞の割合を算出する．

3.2.2 意味密度

意味密度は，話者が発話した文の意味の豊かさを評価する指標である．意味密度を算出するのに，一つの文章に含まれる命題数を求める必要がある．命題数は専門家の作業により求めることができるが，コストが大きい．そのため，英語圏では，意味密度の自動計算方法 [10] が提案されたが，日本語における意味密度の算出方法は未だに確立されていない．しかし，日本語コーパスを機械翻訳を用いて英語に翻訳し，英語における算出方法を適用することで，近似的な値を算出できることが報告された [9]．そのため，機械翻訳を用いて，先行研究と同様に D-ID，DR-ID，DRA-ID と意味密度の増減割合を示す二つの指標を算出し，発話の意味密度とした．

3.2.3 Bag of words

Bag of words(BoW)BoW とは文書に単語が含まれているかのみを考えて，単語の順序は考慮しないモデルである．本研究では，コーパスの全体集合に対して，名詞のみで構成される BoW モデルを作成する．コーパスの各文書の BoW に対して TF-IDF による重み付けを行った後，潜在的意味解析による次元圧縮を行う．このプロセスによって得られた特徴ベクトルを素性とした．

[*6] http://taku910.github.io/mecab/
[*7] http://jhlee.sakura.ne.jp/JEV.html

	全体		健常者		高齢健常者		軽度認知症	
	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差	平均値	標準偏差
参加者数	80	-	20	-	45	-	15	-
年齢	65.2	16.3	39.6	11.4	73.8	4.4	73.5	3.5
教育年数	14.2	2.1	15.1	2.2	14.0	4.4	14.0	1.7
MMSE	21.3	12.5	-	-	29.3	0.7	25.9	1.0

表1: コーパスに収録されている実験参加者の詳細

3.2.4 分散表現

本研究では Word2Vec [8] で学習済みの分散表現ベクトルである日本語 Wikipedia エンティティベクトル [12] を素性として用いた．

4 実験

対照群付き高齢者コーパスは，実験参加者に対して三つの課題を行い，その発話の録音，書き起こしテキストと MMSE の結果から構成されている．MMSE は 65 歳以上の実験参加者のみに対して行い，23 点以上かつ 27 点以下の実験参加者を軽度認知症とし，それ以外の実験参加者を健常者とする．

各節で説明した韻律の特徴 (PFs)，MFCC，基本周波数とフィルターバンクエナジー (FBE)，テキスト基礎特徴と意味密度 (BAD)，BoW，分散表現 (W2V) を素性とし，ロジスティック回帰（LR），SVM，ランダムフォレスト（RF）による識別実験を行なった．

今回の実験データには，1 人あたりに複数の音声データが存在するため，これらを別々の訓練・評価事例と見なすことができる．訓練データと評価データに同じ話者が含まれているのは，厳密な評価とは呼べないが，先行研究がどのような設定で実験を行っているのか不明である．また，今回のデータの被験者は 80 人と少ないため，それぞれの被験者を訓練・評価事例と見なすと，学習データや評価データのサンプル数が不足してしまう．そこで，本研究では被験者を分けずに音声データの単位で分けて 10 分割交差検定を行ったケース（表2）と，被験者の単位で分けて 10 分割交差検定を行ったケース（表3）の両方の設定における評価結果を報告する．

表2では，音声特徴を用いた場合に識別精度が最も高く，特に MFCC は単独でも高い識別制度を示している．これは，訓練データと評価データの両方に同一人物の音声データが含まれるため，タスクが認知症傾向識別ではなく話者識別になってしまった可能性がある．

	Accuracy		
	LR	SVM	RF
PFs	82.25	82.75	82.25
MFCC	87.50	87.25	85.75
FBE	85.75	84.75	86.25
Voice	**90.75**	89.25	86.75
BAD	82.29	81.96	82.52
BoW	81.71	79.48	82.26
W2V	81.70	79.82	81.44
Text	82.25	76.07	83.09
Voice+Text	87.71	84.73	86.25

表2: 識別性能（訓練と評価で話者は混じっている）

	Accuracy		
	LR	SVM	RF
PFs	81.50	81.00	79.50
MFCC	79.00	81.00	79.75
FBE	79.75	74.50	79.00
Voice	68.75	74.00	81.00
BAD	81.41	81.14	80.57
BoW	82.23	79.22	82.23
W2V	81.16	81.42	81.42
Text	81.70	75.38	82.97
Voice+Text	71.83	78.14	80.29

表3: 識別性能（訓練と評価で話者が混じらない）

これに対し，表3の設定では，音声特徴の識別制度が大幅に低下し，テキストの特徴量がわずかに貢献を示した程度になった．この設定ではサンプル数が少なすぎる（80 件）ため，実験結果の安定性・信頼性が低下してしまう．先行研究において MFCC が認知症傾向の識別に有効である，という報告が見られたが，表3の設定にすると MFCC の貢献は消滅し，この実験では音声特徴の有効性を確認できなかった．

5 まとめ

本研究では，対象群付き高齢者コーパスを用いて，柴田らの先行研究に加えて，音声特徴量も抽出し，機械学習による分類実験を行なった．コーパス中のサンプル数の不足などもあり，現状では音声特徴の効果を確認することはできなかった．今後は，コーパスのサンプル数を増やすことを検討するとともに，話者の識別には使えないものの，認知症傾向の識別には有用となるような素性を探求し，限られた言語資源を有効活用する方向性を考えていく必要がある．

参考文献

[1] Alexandra König. et al. “Automatic speech analysis for the assessment of patients with predementia and Alzheimer’ s disease”. In: *Alzheimer’ s and Dementia: Diagnosis, Assessment and Disease Monitoring*. 2015, 1:112–124. Elsevier.

[2] Aramaki E. et al. “Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment”. In: *PLOS ONE*. 2016, Vol.11, No.5, pp.1–13.

[3] Frederique Gayraud. et al. “Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects”. In: *Clin Ling and Phon*. 2010, 25(3):198–209.

[4] Frederique Gayraud. et al. “Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects”. In: *Clin Ling and Phon*. 2011, 25(3):198–209.

[5] Goodglass H. et al. “The assessment of aphasia related disorders”. In: *Philadelphia, Pa,: Lea and Febiger*. 1983, 1:102.

[6] Ildikó Hoffmann. et al. “Temporal parameters of spontaneous speech in Alzheimer’ s disease”. In: *J of Speech-Language Pathology*. 2010, 12(1), 29–34.

[7] Karlekar S. et al. “Detecting Linguistic Characteristics of Alzheimer’ s Dementia by Interpreting Neural Models”. In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2018, Volume 2(Short Papers), pp.701–707.

[8] Mikolov T. et al. “Distributed representations of words and phrases and their compositionality”. In: *Advances in neural information processing systems*. 2013, pp.3111–3119.

[9] Shibata D. et al. “Idea density in Japanese for the early detection of dementia based on narrative speech”. In: *PLoS ONE*. 2018, 13(12):e0208418.

[10] Sirts K. et al. “Idea density for pre-dicting Alzheimer’ s disease from transcribed speech”. In: *Proceedings of the 21st Conference on Computational Natural Language Learning*. 2017, pp.322–332.

[11] 柴田 大作. et al. “対照群付き高齢者コーパスの構築とそれを用いた認知症予備軍スクリーニング技術の開発”. In:

[12] 鈴木正敏. et al. “Wikipedia 記事に対する拡張固有表現ラベルの多重付与”. In: 言語処理学会第 *22* 回年次大会. 2016, pp.797–800.

[13] DittmannJ. et al. BlankenG. “Spontaneous speech in senile dementia and aphasia: implications for a neurolinguistic model of language production”. In: *Cognition*. 1987, 27:247–74.

[14] Graeme Hirst. Kathleen C Fraser. “Detecting semantic changes in Alzheimer’ s disease with vector space models”. In: *LREC Workshop: Resources and Process-Ing of linguistic and extra-linguistic Data from people with various forms of cognitive/psychiatric impairments (RaPID)*. 2016, Pp.1–8.Portorož Slovenia.

[15] 佐渡充洋. “日本における認知症の社会的コスト”. In: ダイヤニュース. 2016, No.84, 10–11.