

コンピュータエージェントによる質問への応答からの 認知症の自動検出

宇城 毅^{*} 田中 宏季^{*} 足立 浩祥[†] 数井 裕光[‡] 池田 学[§] 工藤 喬[†] 中村 哲^{*}

{ujiro.tsuyoki.uq5, hiroki-tan}@is.naist.jp

1 はじめに

1.1 背景

2015年現在、認知症患者数は世界で約4700万人いるとされ、2050年にはその数が1億3000万人に増加すると推測されている[1]。認知症患者への医療対策は世界的な規模で重要な問題である。認知症は、記憶や思考、行動などの認知機能が低下し、日常生活へ支障をきたす状態を指す[2]。多くの医療機関では、DSM-IV-TR[3]やDSM-V[4]に基づいて認知症を診断する。認知症には、アルツハイマー型(AD)や正常圧水頭症型(NPH)、レビー小体型など、さまざまな種類が存在する。その中でも、ADが最も割合を占めている。現在、ADの発症原因は分かっておらず、症状を回復させる方法も確立されていない。しかし、障害があまり進行していない場合、進行を遅らせることができる効果的な治療を受けることができ、支援サービスを受けることもできる。そのため、認知症の早期診断は重要である。

認知症の症状の中核をなす症状は、記憶や見当識、理解、計算、言語、判断などの認知機能障害である。その中でも、記憶障害は主体をなすものであり、進行性に憎悪していく。典型的なADにおいては、顕著に記憶障害が認められる。認知症における記憶障害では、比較的最近(数分から数日程度)に自分が体験した出来事を想起できないという近時記憶の障害が目立つ。また、記憶障害に伴い、周囲の状況を把握できなくなる見当識障害が目立つようになる。見当識障害は、日にちや季節などの時間から始まり、自分のいる場所、周りの人物へと障害の範囲が広がっていく。

認知症の症状は、自分で認識することがほとんどなく、家族や友人などの周囲の人々が認識することが大

多数を占める。これは、自分が認知症ではないという思い込みが一つの要因として挙げられる。ただ、周囲の人々も、認知症の症状を通常の老化の症状と混同してしまうことが多く、症状を早期に気付くのは容易ではない。そのため、ある程度障害が進行した状態で診断を受けることが多い[5]。この問題を踏まえ、自宅においても定期的に状態を把握することが必要であり、これは早期認知症の診断につながる。そのためには、高齢者が自宅においても容易に認知症かどうかを把握するためのツールが必要である。

医療機関では、スクリーニングとしての検査のために神経心理学的検査が用いられる。代表的な検査に、Mini-Mental State Examination(MMSE)や改訂長谷川式簡易知能評価スケール(HDS-R)がある。検査に必要な時間は5~10分と短い、被検者と検査者のやり取りで行われる質問式の検査でかつ、臨床心理や神経心理などに係る専門教育を受け、検査を熟知した者が行う必要があるため、自宅において容易に用いることはできない。

1.2 関連研究

音声情報や言語情報を用いて認知症を検出する研究がいくつか行われている。Roarkら[6]は、神経心理学的検査の一つであるウェクスラーの記憶検査[7]中の音声を収録し、非認知症と軽度認知障害(MCI)の方との分類を行った。MCIは、厳密な定義では認知症ではないが、認知症の前段階ないしはリスク群といわれている。分類には、構文木の複雑さや音声特徴を用いており、結果は、ウェクスラーの記憶検査と同等の検出能力であった。これより、認知症の検出に言語情報や音声情報が有用である可能性を示した。しかし、神経心理学的検査中の研究協力者の発話に着目しており、熟練のスタッフが必要である。さらには、発話内容の書き起こしが必要であるため、自宅において容易に使用することが難しい。Tanakaら[8]は、コンピュータ

^{*}奈良先端科学技術大学院大学 情報科学研究科

[†]大阪大学 キャンパスライフ健康支援センター

[‡]高知大学 医学部 神経精神科学教室

[§]大阪大学大学院 医学系研究科

エージェントを用いて認知症を検出する手法を提案した。この研究では、神経心理学的検査で使用されている質問の中から3問を研究協力者にコンピュータエージェントと音声合成を用いて提示し、質問応答中の音声や言語、画像情報から認知症の検出を試みた。結果は、ROC 曲線の AUC で 0.94 と高い検出能力だった。また、Gap (質問の投げかけから応答するまでにかかった時間) が最も有効な特徴量で、認知症患者の方が非認知症患者よりも Gap が大きいことが報告された。この手法では、3問の質問で固定されているため、定期的を使用することには向いていない。Ujira ら [9] は、神経心理学的検査に用いられていないトピックの質問を用いて、コンピュータエージェントによる質問への応答から認知症検出を試みた。音声特徴と言語特徴を用いて分類した結果、ROC 曲線の AUC で 0.95 と高い検出能力だった。しかし、発話内容の書き起こしを手で行っているため、自宅において使用することは難しい。

本研究では、自宅で定期的に状態を把握するために、コンピュータエージェントを用いて、複数のトピックの質問に対する応答から認知症を自動検出する手法を提案する。

2 認知症の自動検出手法

2.1 エージェントシステム

データ取得のために Tanaka ら [8] のエージェントシステムを利用した。このシステムは、エージェントとして、MMDAgent を使用しており、高齢者(ユーザ)が理解しやすいように、エージェントの発話速度を下げ、発話内容に字幕を付与している。

2.2 質問セット

質問セットには、エージェントの発話内容として、神経心理検査の質問よりも非定型な質問を 13 問用意した。質問は、精神科医師との相談の上で決定した。その質問を表 1 に示す。エージェントは、13 問の質問からランダムで 5 問を選び、高齢者に質問する。エージェントの会話中において、システムはユーザの音声と動画をそれぞれマイクと内蔵カメラで記録した。システムは、15 秒の無回答時間を検出すると、自動的に次の質問へ移行する。

表 1: 用意した質問セット

質問	内容
Q1	ご家族のことを教えてください
Q2	日頃ストレスを感じていることを話してください
Q3	趣味は何ですか
Q4	好きな歌は何ですか
Q5	石原裕次郎さんについて知っていることを教えてください
Q6	長嶋茂雄さんについて知っていることを教えてください
Q7	美空ひばりさんについて知っていることを教えてください
Q8	今の日本の総理大臣は誰ですか
Q9	今の季節は何ですか
Q10	今年は何年ですか
Q11	あなたは左利きですか、右利きですか
Q12	夜は眠れていますか
Q13	食欲はありますか

2.3 音声区間検出と音声認識

質問応答の音声データから、音声区間検出を行い、検出した発話区間の書き起こしを行った。音声区間検出と書き起こしには、Julius¹を使用した。Julius のパラメータに関しては、すべてデフォルトのままで行った。

3 分類実験

3.1 研究協力者

合計 24 名 (認知症患者 12 名, 非認知症患者 12 名) の参加者で、分類実験を行った。実験は、12 名の認知症患者が大阪大学医学部附属病院にて、12 名の非認知症患者が奈良先端科学技術大学院大学にて行われた。研究協力者全員に対して事前に説明をし、この実験の参加に同意し、インフォームドコンセントを得た。研究協力者の詳細な情報を表 2 に示す。認知症患者の詳細は、AD9 名, NPH1 名, MCI1 名, AD + NPH1 名である。これらは、DSM-IV-TR[3] に基づいて、大阪大学医学部附属病院の精神科医師によって診断された。参加者の全員が、与えたタスクを完遂することができた。

¹<http://julius.osdn.jp/>

表 2: 研究協力者のデータ (平均 (標準偏差) で表記)

グループ	N	年齢	MMSE
非認知症	12	74.5 (4.3)	27.5 (1.8)
認知症	12	75.9 (7.3)	21.2 (5.1)

3.2 特徴量抽出

特徴量として、音声特徴量と言語特徴量を抽出した。音声特徴量の抽出は、Snack sound toolkit²を使用した。本研究では、Gap や Pause、基本周波数、パワー、発話の長さに関する特徴を抽出した。Gap は、エージェントの質問終了時から研究協力者の応答開始までの時間差のことである。Pause は、発話と発話の間の中で時間差が 1 秒以上のものをカウントしている。また、時間差の平均と最長のものも算出した。基本周波数に関しては、平均値、標準偏差、変動係数、最大値、中央値、最小値、レンジを算出した。また、パワーに関しては、平均値、標準偏差を算出した。我々は、MeCab³を用いて日本語の形態素解析を行った。MeCab の出力より、トークン (形態素) 数、フィルター、名詞や動詞、形容詞、副詞の数、発話速度 (発話の長さに対するトークン数) を算出した。

3.3 分類モデル

我々は、3 種類の機械学習アルゴリズムを用いて、認知症患者グループと非認知症患者グループの分類を行った。使用した分類モデルは、ロジスティック回帰と L1 正則化付きロジスティック回帰、線形サポートベクトルマシンである。分類モデルの入力には、平均 0、分散 1 になるように正規化された特徴量を使用した。分類モデルの評価には、入れ子式の Leave-one-participant-out 交差検証 [10] を行い、ROC 曲線を描き、ROC 曲線の AUC を算出した。

3.4 実験結果

分類結果として、ROC 曲線の AUC と正解率をまとめたものを表 3 に示す。手動と自動のどちらも最も検出能力が高かった分類モデルは、L1 正則化付きロジスティック回帰で、音声特徴量のみを入力したものである。

表 3: 手動アノテーションおよび自動アノテーションによる検出能力。Sp. は音声特徴量、La は言語特徴量を指す。

	手動		自動	
	AUC	Acc.	AUC	Acc.
MMSE	0.85	0.83	*	*
Sp. (SVM)	0.85	0.83	0.74	0.75
Sp. (LR)	0.90	0.83	0.85	0.79
Sp. (L1)	0.97	0.96	0.88	0.71
Sp.+La. (L1)	0.91	0.88	0.85	0.67
Sp.+La. (LR)	0.92	0.92	0.71	0.58
Sp.+La. (SVM)	0.95	0.92	0.71	0.58

次に、ロジスティック回帰における特徴量の重みについて分析を行った。手動アノテーションにおける上位 5 つの特徴量は、1) Gap、2) F0 の最大値と最小値の差、3) F0 の最大値、4) Pause の平均値、5) 動詞、である。自動アノテーションにおける上位 5 つの特徴量は、1) Power の平均値、2) Gap、3) F0 の中央値、4) Pause の平均値、5) F0 の標準偏差、である。これより、分類に有効な特徴量であることがわかる。また、Gap は、認知症患者の方が非認知症患者よりも大きい傾向があった。ここで、マン・ホイットニーの U 検定で p 値を計算すると、手動アノテーションで $p = 0.04$ 、自動アノテーションで $p = 0.02$ と有意差を示した。また、効果量は、手動アノテーションで $r = 0.42$ 、自動アノテーションで $r = 0.47$ であった。

次に、L1 正則化によって選択された特徴量について分析を行った。Leave-one-participant-out の試行のすべてにおいて選択された特徴量は、手動アノテーションでは、Gap、F0 の最大値と最小値の差、パワーの平均値と標準偏差、形容詞、発話速度であった。また、自動アノテーションでは、Gap、F0 の中央値、Pause の平均値、形容詞であった。この結果からも、Gap は分類に有効な特徴量であるといえる。

最後に、各質問における Gap を算出した結果を図 1 に示す。図 1 より、認知症患者の方が非認知症患者よりも大きい傾向が確認できた質問は、Q8 (今の日本の総理大臣は誰ですか)、Q9 (今の季節は何ですか) であった。これらの質問は、認知症の早期段階から障害される見当識に関する質問といえるため、認知症患者の方が非認知症患者よりも大きい傾向が確認できた。また、Q5、Q6、Q7 (石原裕次郎、長嶋茂雄、美空ひばりさんについて知っていることを教えてください)、

²<http://www.speech.kth.se/snack/>

³<http://taku910.github.io/mecab/>

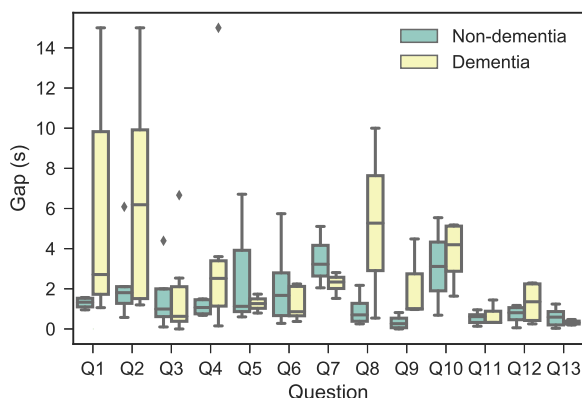


図 1: 手動アノテーションにおける各質問の Gap

Q11 (あなたは左利きですか, 右利きですか), Q12 (夜は眠れていますか), Q13 (食欲はありますか) に関しては, 認知症患者の方が非認知症患者よりも大きい傾向が確認できなかった. Q5, Q6, Q7 は過去の有名人に関する質問であり, Q11, Q12, Q13 はクローズドエンド型質問である.

4 まとめ

本研究では, コンピュータエージェントを用いて複数の話題の質問に対する応答から認知症の自動検出手法を提案した. 24名の参加者による分類実験でこの手法の有効性を示した. 認知症の自動検出の能力は, ROC 曲線の AUC で 0.88 だった. この結果は, 認知症のスクリーニングで一般的に用いられる検査の一つである MMSE と同等以上の検出能力であり, 提案した認知症の自動検出手法においても, 認知症を検出することができたといえる.

今後の課題として, まず, 本研究の結果の有用性を示すために研究協力者数を増やす必要がある. 次に, 認知症の原疾患の違いによる分類結果への影響の調査である. これは, 早期認知症の場合, 種類によって現れる症状が異なる場合があるからである. 最後に, 発話内容の詳細な分析である. Roark ら [6] は, 音声特徴に加え, 構文木の複雑さにも着目しており, 有効性を示した. 本研究では, 形態素解析を行ったのみであり, さらに詳細に分析する必要がある.

参考文献

- [1] Martin Prince et al. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. Technical report, 2015.
- [2] Guy M McKhann et al. The diagnosis of dementia due to alzheimer's disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic guidelines for alzheimer's disease. *Alzheimer's & dementia*, 7(3):263-269, 2011.
- [3] *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)*. American Psychiatric Association, 2000.
- [4] *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)*. American Psychiatric Association, 2013.
- [5] Ane Alberdi, Asier Aztiria, and Adrian Basarab. On the early diagnosis of alzheimer's disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, 71:1-29, 2016.
- [6] Brian Roark et al. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech and Language Processing*, 2011.
- [7] David Wechsler. *WAIS-III, Wechsler adult intelligence scale: Administration and scoring manual*. Psychological Corporation, 1997.
- [8] Hiroki Tanaka et al. Detecting Dementia Through Interactive Computer Avatars. *IEEE Journal of Translational Engineering in Health and Medicine*, 2017.
- [9] Tsuyoki Ujio et al. Detection of dementia from responses to atypical questions asked by embodied conversational agents. *Proc. Interspeech 2018*, pages 1691-1695, 2018.
- [10] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1):91, 2006.