

修士論文

コンピュータエージェントを用いた認知症検出のための 質問分析と自動化手法

宇城 毅犠

2019 年 3 月 4 日

奈良先端科学技術大学院大学
情報科学研究科 情報科学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
修士（工学）授与の要件として提出した修士論文である。

宇城 毅犠

審査委員：

中村 哲 教授 （主指導教員）

安本 慶一 教授 （副指導教員）

田中 宏季 助教 （副指導教員）

コンピュータエージェントを用いた認知症検出のための 質問分析と自動化手法*

宇城 毅犠

内容梗概

認知症の早期発見には，定期的に高齢者の認知機能の状態を把握する必要がある．これまで，音声言語が高齢者の認知機能を反映させていることに着目して，音声言語を用いた認知症の検出手法が多く提案されてきた．その中でも，質問応答から認知症を検出する手法は，高齢者にとって課題の難度が低く，負担が少ない．しかし，同じ質問を提示する場合，高齢者が回答に慣れる可能性がある．先行研究では，音声言語を用いた認知症の検出手法について検討されているが，音声言語を用いた認知症検出のための有効な質問については検討の余地がある．そこで，本研究では，音声言語を用いた認知症検出において有効な質問について分析した．自己認識や自身の関心，遠隔記憶などに係る質問を用意し，機械学習アルゴリズム (L2 正則化付きロジスティック回帰，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) によって検出能力を算出した．その結果，自身の関心に係る質問のカテゴリが，検出能力が高く (AUC 値: 0.95)，音声言語を用いた認知症検出において有効であることがわかった．また，本研究では，コンピュータエージェントの質問に対する応答から認知症を自動検出する手法も提案した．コンピュータエージェントは，13 問の質問からランダムで 5 問を 24 人の参加者 (認知症患者 12 名，非認知症者 12 名) に問いかけた．質問への応答を記録し，音声区間を自動的に検出し，音声特徴と言語特徴を抽出した．抽出した特徴量を用いて，機械学習アルゴリズム (L2 正則化付きロジスティック回帰，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) によって 2 つのグループ (認知症，非認知症) に分類した．結果は，ROC 曲線の AUC で 0.88 の検出能力を示した．これは，認知症のスクリーニングテストで一般的に用いられる MMSE と同等以上の検出能力である

*奈良先端科学技術大学院大学 情報科学研究科 修士論文, 2019 年 3 月 4 日.

(AUC: 0.85). これより，コンピュータエージェントを用いて認知症の自動検出する手法の有用性を示した．

キーワード

認知症検出，コンピュータエージェント，質問分析，自動化，音声，言語

Question Analysis and Automated Method for Dementia Detection through Computer Agents*

Tsuyoki Ujiro

Abstract

Regular dementia inspection will help to detect early dementia. Previous studies focused on assumption that spoken utterances may reflect cognitive decline in elderly people and proposed a dementia detection method using speech and language information. The methods which detect dementia from question answering are easy and less stress for the elderly. However, when the same tasks were given, the elderly may get used to the task. Previous studies examined the methods for dementia detection, however, there is room for discussion about effective questions for dementia detection using speech and language. Therefore, I analyzed effective questions for dementia detection. I prepared questions which related to self-recognition, interest, remote memory, and so on and calculated detection performance by machine learning algorithms (support vector machines, L2 regularized logistic regression, and L1 regularized logistic regression) for each category. As a result, I found that the questions which are related to interest are effective (AUC: 0.95). Next, I proposed an automated detection method of dementia from responses to a computer agent's questions. The agent asked participants about randomly chosen five questions from the 13 questions and obtained answers from 24 participants (12 dementia and 12 non-dementia). I applied voice activity detection (VAD) to the recorded responses and extracted speech and language features. We classified the participants into two groups (dementia / non-dementia) by machine learning algorithms (support

*Master's Thesis, Graduate School of Information Science, Nara Institute of Science and Technology, March 4, 2019.

vector machines, L2 regularized logistic regression, and L1 regularized logistic regression). The results showed a 0.88 detection performance in the area under the curve of the receiver operating characteristic (AUC). We confirmed that this score was greater than a score of the Mini-Mental State Examination which is one of the frequently used tests for screening dementia.

Keywords:

Dementia detection, Computer agents, Automated, Question analysis, Speech, Language

目次

図目次	vii
表目次	1
第 1 章 はじめに	2
1.1 背景	2
1.2 研究目的	3
1.3 本論文の構成	3
第 2 章 関連研究	4
2.1 認知症の症状	4
2.2 神経心理学的検査	5
2.3 音声言語による認知症検出	8
2.4 関連研究の課題	9
2.5 本章のまとめ	10
第 3 章 提案手法	11
3.1 研究協力者	11
3.2 質問の分析手法	12
3.3 質問セット	12
3.4 認知症の自動検出手法	12
3.5 コンピュータエージェント	14
3.6 音声区間検出と音声認識	15
3.7 特徴量抽出	15
3.8 分類モデル	17
3.8.1 L2 正則化付きロジスティック回帰	17
3.8.2 L1 正則化付きロジスティック回帰	19
3.8.3 線形サポートベクトルマシン	19
3.9 評価方法	20

3.9.1	交差検証	21
3.9.2	ハイパーパラメータのチューニング	21
3.9.3	ROC 曲線と AUC	22
3.9.4	マン・ホイットニーの U 検定	23
第 4 章	結果	25
4.1	認知症検出に有効な質問の分析	25
4.2	認知症検出の自動化	26
第 5 章	考察	30
5.1	質問分析	30
5.2	認知症検出の自動化	31
第 6 章	総括	32
	謝辞	33
	参考文献	34
	発表リスト	39

図目次

2.1	時間による記憶の分類	5
3.1	自動検出手法の概要	14
3.2	エージェントシステム	15
3.3	ロジスティック回帰の概念図	17
3.4	決定境界とサポートベクトルの関係	20
3.5	入れ子式の交差検証の概念図	22
3.6	混同行列	23
4.1	手動 VAD における各質問の Gap	26
4.2	分類モデルの ROC 曲線	27
4.3	非認知症グループと認知症グループの Gap	28

表目次

2.1	MMSE の設問	7
3.1	研究協力者のデータ (平均 (標準偏差) で表記)	11
3.2	質問セット	13
3.3	特徴量一覧	16
4.1	カテゴリ別の検出能力と正解率	25
4.2	分類モデルの評価	28
4.3	質問数の変化による検出能力の違い	29

第 1 章 はじめに

1.1 背景

認知症は、認知機能が病前よりも低下し、日常生活へ支障をきたす状態を指す [1]. 認知機能の低下に伴い、記憶や理解、判断など、さまざまな領域が障害される [2, 3]. 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 [4]. 現在、認知症の治療方法は確立されていないため、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. 認知症があまり進行していない初期段階の場合、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る. そのため、認知症の早期診断による早期治療が重要である. しかし、病識の欠如や老化による症状の混同などにより、早期段階で認知症を発見することは難しい [5, 6]. 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る.

医療機関では、認知症のスクリーニングのために、Mini-Mental State Examination (MMSE)[7] や改訂長谷川式簡易知能評価スケール (HDS-R)[8] などの神経心理学的検査が用いられる. これらの検査は、口頭試問で行われるため、訓練された検査者が必要である. また、約 10 分の検査だが、問診よりも時間がかかる. このため、簡便な認知症のスクリーニングツールが必要である.

認知症により認知機能が低下して、さまざまな領域が障害されるが、コミュニケーションに係る音声や言語についても障害される [9, 10, 11, 12]. Aramaki ら [13] は、認知機能の低下と言語能力との関係を調査し、話し言葉の分析から認知機能の低下の検出の可能性を示した. これより、音声情報や言語情報に着目した認知症検出手法が提案されている [14, 15, 16]. さらに、音声言語を用いた認知症の自動検出手法も提案されている [17, 18, 19, 20, 21, 22].

Tanaka ら [23] や Kato ら [24], Mirheidari ら [25] は、実験協力者の質問応答から認知症を検出する手法を提案している. 質問応答による認知症検出は、神経心理学的検査よりも所要時間が短く、課題も簡単であるため、高齢者に与える負担は軽減される. これらの研究では、認知症を検出するために有効な特徴量やモデルを検討しており、その有用性が示されている. 認知症の早期発見には、定期的に状態を

把握する必要がある，認知症検出ツールとして繰り返し使用できることが望ましい．しかし先行研究では，高齢者に与える課題（質問）が固定されており，繰り返し使用することによって高齢者が課題に慣れてしまう可能性があり，検出能力に影響を与える [26]．そのため，同じ課題を与える認知症のスクリーニングツールを定期的に用いることは望ましくない．

1.2 研究目的

質問応答における音声言語から認知症を検出することは有用であることが示されたが，認知症を検出するための有効な質問については検討の余地がある．そこで本研究では，質問応答から音声言語を用いて認知症を検出するにあたり，以下の課題について検討する．

1. 認知症検出に有効な質問の分析
2. エージェントシステムを用いた認知症検出の自動化

1.3 本論文の構成

以下，本稿では，質問応答から早期認知症の自動検出手法について述べる．まず，2 章で認知症について述べ，関連研究について説明する．続いて，3 章で提案手法について説明する．そして，4 章で結果について述べ，5 章で考察を述べる．最後に，6 章で本論文の総括を行う．

第 2 章 関連研究

本章では、本稿の目的である認知症検出に関する話題をまとめる。まず、2.1 節で、認知症の症状について述べる。さらに 2.2 節で、医療機関において一般的に用いられる神経心理学的検査について述べる。次に、2.3 節で、質問応答から音声言語情報を用いて認知症を検出する手法についてまとめる。そして、2.4 節で、関連研究の課題について述べ、2.5 節でこの章をまとめる。

2.1 認知症の症状

認知症は、認知機能が病前よりも低下し、日常生活へ支障をきたす状態を指す [1]。臨床でよく用いられる診断基準に、米国精神医学会による DSM-IV-TR[27] や DSM-V[28] がある。本研究では、認知症患者の精神科医師による診断に DSM-IV-TR を用いた。認知症自身は疾患単位ではなく、さまざまな病因による疾患の集合体であり、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、様々な種類が存在する [29]。その中でも、AD が最も割合を占めている [4]。

認知症は、日常生活に支障をきたしていることが必須であったが、日常生活に支障をきたさない程度の認知機能の障害の場合は、軽度認知障害 (MCI) とみなされる。1 年あたり、約 10~15% の MCI 患者が AD に移行することがいわれており、MCI は認知症の前段階ないしリスク群であるといえる。

認知症の症状には、認知症の症状の中核をなす中核症状と認知症に伴う行動異常および心理症状の周辺症状 (BPSD) の 2 つのグループに分けられる。中核症状は、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。記憶障害はその中でも主体をなすものであり、進行性に憎悪していく。また、記憶障害は、比較的早期の段階で障害が認められることが多く、典型的な AD においては、顕著に記憶障害が認められる [3]。

記憶は、時間によって図 2.1 のように分類することができる。即時 (短期) 記憶は、情報の記銘の後、すぐに想起させるものである。近時記憶は、情報の記銘の後、数分から数日ほど情報を保持する。遠隔記憶は、近時記憶よりもさらに長期間保持し、個人の生活史や歴史的イベントなどを記憶する。即時記憶は記銘から想起までに干

渉がなく，近時記憶は記銘から想起までに干渉が介在する．近時記憶と遠隔記憶との時間的な境界は明確に定められていない [2]．

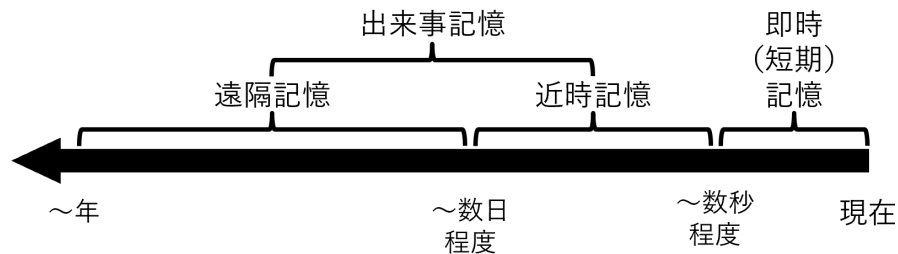


図 2.1 時間による記憶の分類

認知症における記憶障害は，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ことや，新しい情報を学習できないということが特徴である．つまり，近時記憶が日常生活で最も急速に忘却が起こる部分である．また，症状が進行するにつれて，遠隔記憶にまで障害が及ぶといわれている [30]．

また，見当識障害についても，比較的初期段階において，障害が確認される [31]．見当識とは，時間や場所，人物などの基本的な状況を把握する能力である．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，周りの人物へと広がっていく．

BPSD に含まれる問題としては，幻覚・妄想，易怒性，無関心・意欲低下，徘徊などが挙げられる．BPSD の発症は，原因疾患に加えて身体・心理・環境的な要因も関与している．AD における BPSD の典型的な症候は，無関心・意欲低下である [3]．また，うつ状態も頻度の高い症候である．

2.2 神経心理学的検査

神経心理学的検査は，中核症状と BPSD を定量化する目的で使用され，スクリーニングとしての検査，進行度合いや治療効果の評価としての検査，鑑別診断の補助としての検査に分類することができる．神経心理学的検査の多くは，高齢者と検査者のやり取りで行われる質問式の検査のため，そのときの高齢者の気分や検査者の質問の仕方や態度などが結果に影響を及ぼすことがあり，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ことが望ましいとされている [32].

認知症のスクリーニングテストとして、Mini-Mental State Examination (MMSE) [7] が臨床および研究において国際的に広く用いられている。MMSE は、見当識（時間、場所）、記銘（物品名の復唱）、注意（計算）、想起（物品名の想起）、言語（物品名の呼称、文章の反復、3 段階の口頭命令、読解、書字）、図形模写の 11 項目で構成され、認知機能を評価する検査である。表 2.1 に MMSE の設問の概要を示す。検査方式は、高齢者と検査者のやり取りで行われる質問式で、所要時間は 10 分程度である。合計スコアは 30 点で、23 点以下で認知症の疑いがあるとする [3]。そして、MMSE の感度（認知症患者が正しく認知症と検出される割合）は 0.76-0.87、特異度（非認知症者が正しく非認知症と検出される割合）は 0.82-0.97 と報告されている [33]。

日本では、改訂長谷川式簡易知能評価スケール（HDS-R）[8] もスクリーニングテストとして一般的に使われている。HDS-R は、年齢（1 点）、日時の見当識（4 点）、場所の見当識（2 点）、3 つの言葉の記銘（3 点）、計算（2 点）、数字の逆唱（2 点）、3 つの言葉の遅延再生（6 点）、5 つの物品記銘（5 点）、言語流暢性（5 点）の 9 つの項目があり、すべて言語を用いる検査である。MMSE と同様、高齢者と検査者のやり取りで行われる質問式で、所要時間は 5 分程度である。合計スコアは 30 点で、20 点以下を認知症の疑いがあるとする [32]。そして、HDS-R の感度は 0.90、特異度は 0.82 と報告されている [33]。

表 2.1 MMSE の設問

項目	点数	内容
時間の見当識	1	「今年は何年ですか.」と質問する.
	1	「今の季節は何ですか.」と質問する.
	1	「今日は何曜日ですか.」と質問する.
	1	「今日は何月ですか.」と質問する.
	1	「今日は何日ですか.」と質問する.
場所の見当識	1	「この病院の名前は何ですか.」と質問する.
	1	「ここは何県ですか.」と質問する.
	1	「ここは何市ですか.」と質問する.
	1	「ここは何階ですか.」と質問する.
	1	「ここは何地方ですか.」と質問する.
物品名の復唱	3 (各 1)	3 つの関連のない物の名前を伝え、復唱させる. 例: 桜, 猫, 電車
注意	5	100 から 7 を順に引き算させる. 5 回できれば 5 点. 間違えた時点で終了.
物品名の想起	3 (各 1)	『物品名の復唱』で提示した物品を再度復唱させる.
物品名の呼称	2 (各 1)	腕時計と鉛筆をみせて、名称を答えさせる.
文章の反復	1	次の文章を繰り返す. 「みんなで、力を合わせて綱を引きます」
3 段階の動作命令	3 (各 1)	次の (口頭による) 命令を実行させる. 「右手にこの紙を持ってください.」 「それを半分に折ってください.」 「机の上に置いてください.」
読解	1	文章に書かれた指示に従わせる. 「目を閉じなさい」
書字	1	「何か文章を書いてください.」
図形模写	1	「次の図形を書き写してください.」 図形は角が 10 個あり、2 つの五角形が交わっている. 

2.3 音声言語による認知症検出

Roark ら [14] 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 [34] 中の音声を収録し、非認知症と軽度認知障害 (MCI) の分類を行った。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。

Aramaki ら [13] は、健常者と MCI の疑いがある被験者の言語能力を分析した。実験では、生活の中で最も幸せな出来事というテーマの作文と同じテーマの発話の課題を与えた。そして、作文を解析した結果、健常者と MCI の疑いがある被験者との間に大きな違いは見られなかった。しかし、発話を解析した結果、健常者よりも MCI の疑いがある被験者の方が、語彙数が多くなるという結果となった。これより、話し言葉の分析が認知症の検出の可能性を示した。

Tanaka ら [23] は、コンピュータ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 問を実験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者よりも Gap が大きいことが報告された。

Kato ら [24] は、出生地、卒業した小学校の名前、日時見当識、3 桁の数字の逆唱の 4 問を出題し、回答中の音声から認知症の検出を試みた。特徴量には、ピッチやフォルマント、音声の大きさ、メル周波数ケプストラム、回答するまでの反応時間に関する合計 128 個を抽出した。そして、主成分分析による変数合成を行った後、赤池情報量規準を用いたフォワードステップワイズ法で特徴量選択を行った。そして、選択された特徴量を説明変数とし、HDS-R のスコアを目標属性として重回帰分析を行うと、HDS-R のスコアと選択された特徴量との間に中程度に有意な相関を示した。また、軽度の AD と非認知症者の分類結果は、出生地を尋ねる質問については、67.2% の正解率、卒業した小学校の名前を尋ねる質問については、74.7%、日時見当識について問う質問については、89.5% の正解率、3 桁の数字の逆唱の課題については、78.1% の正解率であった。さらに、MCI と非認知症者の分類結果

は、出生地を尋ねる質問については、73.3%の正解率、卒業した小学校の名前を尋ねる質問については、70.9%、日時見当識について問う質問については、75.9%の正解率、3桁の数字の逆唱の課題については、76.4%の正解率であった。これより、提案手法がADやMCIのスクリーニングに有効なツールである可能性を示した。

Bahman ら [25] は、コンピュータエージェントを用いて、アルツハイマー病などの神経変性疾患（ND）を発症している患者と記憶障害（FMD）を有し、認知症を有していない患者の自動分類システムを提案した。コンピュータエージェントは、次のような質問を投げかけた。

- 長期記憶に関する質問（例：学校を卒業した時は何歳でしたか？）
- 一度の発話に複数の質問が含まれているもの（例：なぜ今日はここに来たのですか？また、あなたの期待は何ですか？）
- オープンエンド型の質問（例：学校を卒業した後、何の仕事をしましたか？）
- 自身の記憶に関する質問（例：誰があなたの記憶について心配していますか？）

ND は、一度の発話に複数の質問が含まれている質問を答えるのが難しい傾向にある。また、FMD は、自身の記憶に関する質問を答えやすいと感じる傾向がある。音声特徴と言語特徴、そして、ND や FMD が現しやすい会話の特徴（例：特定のワードやフィラーなど）を用いて、ND と FMD の分類を行った結果、約 90% の精度であった。これより、コンピュータエージェントを用いた認知症の自動検出の可能性を示した。

2.4 関連研究の課題

関連研究では、音声言語から認知症を検出するために有効な特徴量や分類手法が提案されているが、有効な質問は明らかにされていない。また、認知症の早期発見には定期的に状態を把握する必要があり、認知症検出ツールとして繰り返し使用できることが望ましいが、関連研究においては、同じ質問を提示するため、繰り返し使用した場合、高齢者が慣れてしまう可能性がある。

2.5 本章のまとめ

2.1 節では，認知症の診断基準や MCI との違いを述べ，比較的早期の段階から症状の主体をなす記憶障害と見当識障害について説明した．記憶は，時間によっていくつかの種類に分けることができ，認知症では近時記憶の障害が早期段階から目立つ．2.2 節では，医療機関で認知症スクリーニングとして使われる神経心理学的検査を説明した．代表的なものとして，MMSE や HDS-R が挙げたが，これらのテストには訓練された検査者が必要である．2.3 節では，音声言語による認知症の検出に関する研究を紹介した．神経心理学的検査の課題中や口頭による質問への応答中の音声言語に着目し，認知症検出に有効な特徴量や分類手法が提案されている．そして，質問応答の音声言語に着目することで，認知症を検出できる可能性を示している．2.4 節では，音声言語から認知症を検出するための有効な質問は明らかにされておらず，定期的に使用することが想定されていないという関連研究の課題を述べた．

第 3 章 提案手法

本章では、(1) 認知症検出に有効な質問を分析し、(2) エージェントシステムを用いた認知症検出の自動化手法を提案する。まず、3.1 節で本研究の研究協力者について述べる。次に、3.2 節で質問の分析手法を述べ、3.4 節で繰り返し使用できる認知症検出の自動化手法について述べる。そして、3.5 節で使用したエージェントシステムを説明し、3.3 節で用意した質問セットについて説明する。また、3.6 節では、自動検出における音声区間検出と音声認識について述べる。そして、3.7 節で抽出した特徴量について説明し、3.8 節で分類モデルについて説明する。最後に、3.9 節で評価方法について述べる。

3.1 研究協力者

合計 24 名 (認知症患者 12 名, 非認知症者 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てもらい、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 3.1 に示す。認知症患者は、DSM-IV-TR [27] 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。認知症の原因疾患の内訳は、アルツハイマー型 9 名, 正常圧水頭症 1 名, MCI 1 名, アルツハイマー + 正常圧水頭症 1 名である。また、参加者の全員が、与えたタスクを完遂することができた。

表 3.1 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 質問の分析手法

まず、用意した質問セット（詳細は 3.3 節で述べる）をカテゴリ別に分ける．そして、それぞれのカテゴリにおける検出能力を算出する．次に、非認知症患者よりも認知症患者の方が反応するまでの時間に遅れが生じることが示されており [35]，Tanaka ら [23] が質問応答から認知症を検出するために、Gap として特徴量に加えている．そして、Gap が認知症検出において最も有効な特徴量であることを示した．これより、各質問における非認知症患者と認知症患者の Gap について比較し、その質問が有効かどうかを検討する．

3.3 質問セット

本研究で用意した質問セットを表 3.2 に示す．この質問セットは、精神科医師との相談により作成された．まず、Q1 については、回答の際に家族の状況について把握している必要があり、人の見当識に係る質問であると考えられる．次に、Q2,Q11,Q12,Q13 は、自己認識に係る質問であるといえる．なお、この 4 問は、問診の技法などの臨床技法を評価する OSCE [36] に基づいて作成され、問診における会話の導入などを想定している．ここで、Q2 はオープンエンド型の質問（回答内容が限定されていない質問）であり、Q11-Q13 はクローズドエンド型の質問（回答の選択肢が決められている質問）である．これらの質問は、問診における会話の導入によく使用され、日常会話においても次に、Q3 や Q4 は、本人の関心に係る質問と考えることができる．他にも、Q5,Q6,Q7 は、過去の有名人に関する質問で、日常会話における昔話の語りを想定しており、遠隔記憶に係る質問と考えることができる．また、Q8 は常識に関する質問であり、Q9,Q10 については、時の見当識に係る質問であると考えられる．この 3 問に関しては神経心理学的検査に基づく質問であるが、その他の質問は神経心理学的検査には基づいていない．

3.4 認知症の自動検出手法

本節では、コンピュータエージェントの質問に対する応答から認知症を自動検出手法について述べる．図 3.1 にその概要を示す．

表 3.2 質問セット

	カテゴリ	質問文
Q1	人の見当識	ご家族のことを教えてください.
Q2	自己認識	日頃ストレスを感じることを話してください.
Q3	関心	趣味は何ですか.
Q4	関心	好きな歌は何ですか.
Q5	遠隔記憶	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	遠隔記憶	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	遠隔記憶	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	神経心理学的検査	今の日本の総理大臣は誰ですか.
Q9	神経心理学的検査	今の季節は何ですか.
Q10	神経心理学的検査	今年は何年ですか.
Q11	自己認識	あなたは左利きですか, 右利きですか.
Q12	自己認識	夜は眠れていますか.
Q13	自己認識	食欲はありますか.

提案手法は, コンピュータエージェントが高齢者に質問を投げかけ, その応答から認知症を検出する. まず, 3.5 節でコンピュータエージェントについて述べ, 3.3 節で用意した質問セットについて述べる. 次に, 高齢者の応答から発話区間を自動的に抽出する手法と音声認識の手法を 3.6 節で述べる. さらに, 特徴量の抽出方法を 3.7 節で述べる. 最後に, 分類モデルの詳細を 3.8 節で述べ, モデルの評価方法について 3.9 節で述べる.

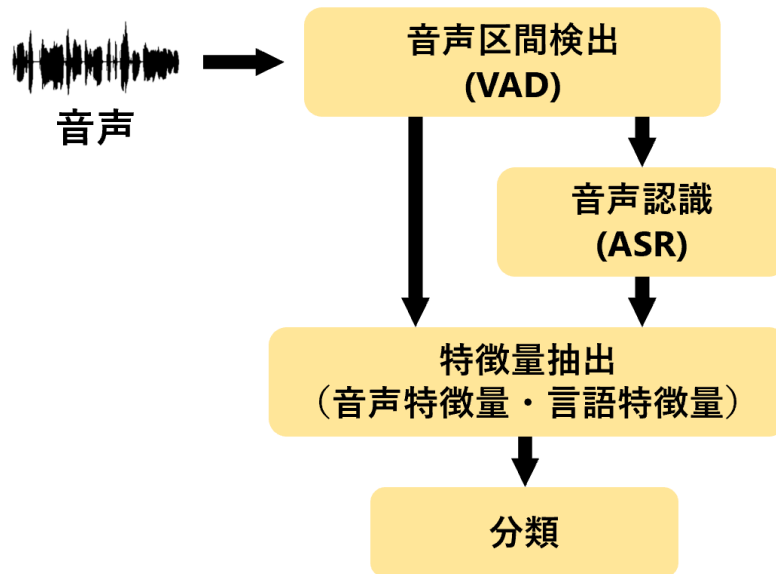


図 3.1 自動検出手法の概要

3.5 コンピュータエージェント

本研究では, Tanaka ら [23] が作成したエージェントシステムを使用してデータを収集した. 図 3.2 にエージェントシステムを示す. エージェントは, MMDAgent* を使用した. エージェントの発話速度が下げられており, 発話内容が字幕として表示されるため, 高齢者が容易に質問が理解できるよう工夫されている. エージェントは, 合計で 13 問の質問セット (詳細は 3.3 節で述べる) からランダムで 5 問選択し, 高齢者に問う. 13 問の中から 5 問をランダム選択することで, 課題 (質問) の重複を防ぐことができ, 高齢者の慣れに対応することができる. さらにエージェントは, 質問応答中の高齢者の音声と映像を記録する. 高齢者が 15 秒以上無言が続いた場合, 自動的に次の質問へ移動する. 本研究では, 高齢者の応答時における発話について着目し, 音声と言語情報について分析を行った.

*<http://www.mmdagent.jp/>



図 3.2 エージェントシステム

3.6 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，検出した区間の音声認識を行う．音声区間検出と音声認識には，Julius [37] 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．

音声区間検出の評価は，False Acceptance Rate (FAR) と False Rejection Rate (FRR) を用いた．FAR と FRR は次のように定義される．

$$FAR = \frac{N_{FA}}{N_{ns}} \quad (3.6.1)$$

(N_{FA} : 誤受理したフレーム数, N_{ns} : 雑音区間のフレーム数)

$$FRR = \frac{N_{FR}}{N_s} \quad (3.6.2)$$

(N_{FR} : 誤棄却したフレーム数, N_s : 音声区間のフレーム数)

3.7 特徴量抽出

本研究では，質問への応答から音声区間検出を行い，検出された区間から音声特徴量を，検出された区間の音声認識結果から言語特徴量を抽出し，合計 5 問の平均値を計算した．特徴量については，先行研究 [22, 23, 24] に基づいて抽出した．抽出した特徴量について，表 3.3 にまとめた．ここで，F0，および，パワーの抽出に

は, Snack Sound Toolkit[†]を使用した. また, トークンやフィラー, 品詞 (動詞, 名詞, 形容詞, 副詞) の抽出には, MeCab Toolkit[‡]を使用した.

表 3.3 特徴量一覧

特徴量	説明
Gap	質問の応答にかかった時間.
Pause_count	発話と発話の間の中で 1 秒以上のものの個数.
Pause_time_mean	発話と発話の間 (1 秒以上) の平均値.
Pause_time_max	発話と発話の間 (1 秒以上) の最大値.
F0_mean	有声区間における基本周波数の平均値.
F0_min	有声区間における基本周波数の最小値.
F0_max	有声区間における基本周波数の最大値.
F0_median	有声区間における基本周波数の中央値.
F0_range	有声区間における基本周波数の最大値と最小値の差.
F0_sd	有声区間における基本周波数の標準偏差.
F0_cov	有声区間における基本周波数の変動係数.
Power_mean	有声区間における声量の平均値.
Power_sd	有声区間における声量の標準偏差.
Utterance length	発話の長さ.
トークン数	発話内の形態素数
フィラー	発話内のフィラーの総数
動詞	発話内の動詞の総数.
名詞	発話内の名詞の総数.
形容詞	発話内の形容詞の総数.
副詞	発話内の副詞の総数.
Speech rate	発話時間あたりのトークン数.

[†]<http://www.speech.kth.se/snack/>

[‡]<http://taku910.github.io/mecab/>

3.8 分類モデル

前処理として、各特徴量を平均 0，標準偏差 1 に正規化を行い，3 つの機械学習アルゴリズムを用いて，認知症患者と非認知症者の分類を行った．分類モデルには，L2 正則化付きロジスティック回帰（LRL2），L1 正則化付きロジスティック回帰（LRL1），線形サポートベクトルマシン（SVM）を使用した．分類モデルについては，以下の 3.8.1 項，3.8.2 項，3.8.3 項で説明する．

3.8.1 L2 正則化付きロジスティック回帰

本項では，L2 正則化付きロジスティック回帰について述べる．まず，ロジスティック回帰の概念図を図 3.3 に示す．

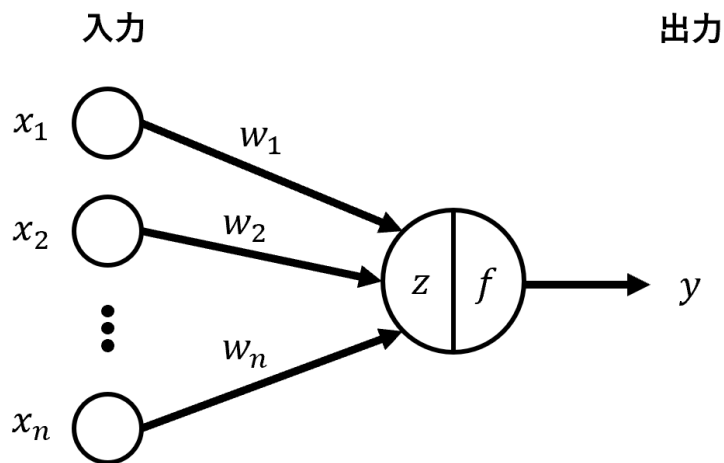


図 3.3 ロジスティック回帰の概念図

図 3.3 のように，ロジスティック回帰では，入力 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ に対して，重み $\mathbf{w} = (w_1, w_2, \dots, w_n)$ とバイアス b を用いて線形結合を行っている．線形結合した値を z とすると，

$$z = \mathbf{w}^T \mathbf{x} + b \quad (3.8.1)$$

と表すことができる．この z は，シグモイド関数 (f) によって，無限区間 $(-\infty,$

∞) が区間 $(0, 1)$ の確率値に写像される．なお、シグモイド関数は次のように表される．

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (3.8.2)$$

最終的な出力 y は、以下に示すように、予測された確率値を用いて二値の結果に変換される．

$$y = \begin{cases} 1 & f(z) \geq 0.5 \\ 0 & f(z) < 0.5 \end{cases} \quad (3.8.3)$$

学習では、重み $\mathbf{w}$ とバイアス b を変化させ、コスト関数 J の値を確認する．コスト関数 J は、以下のように定義され、この値が最小となるような重み $\mathbf{w}$ とバイアス b を決める．

$$J(\mathbf{w}) = \sum_{i=1}^n [-y^{(i)} \log(f(z^{(i)})) - (1 - y^{(i)}) \log(1 - f(z^{(i)}))] \quad (3.8.4)$$

次に、L2 正則化について述べる．まず、正則化は、トレーニングデータセットではうまく機能するモデルが、テストデータではうまく汎化されないという問題（過学習）への対処する手法の一つである．過学習は、モデルがトレーニングセットの観測結果にパラメータを適合させすぎている状態であり、データセットに対してモデルが複雑すぎるのが原因である．正則化を行うことで、重みにペナルティを科し、モデルの複雑さを低減させる．正則化にはいくつか種類があるが、L2 正則化は一般的な正則化手法であり、以下のように定義される．

$$\|\mathbf{w}\|^2 = \sum_{j=1}^m w_j^2 \quad (3.8.5)$$

L2 正則化が適用されることにより、式 3.8.4 が以下の式で表される．

$$J(\mathbf{w}) = C \sum_{i=1}^n [-y^{(i)} \log(f(z^{(i)})) - (1 - y^{(i)}) \log(1 - f(z^{(i)}))] + \frac{1}{2} \|\mathbf{w}\|^2 \quad (3.8.6)$$

式 3.8.6 の C は、逆正則化パラメータと呼ばれ、トレーニングデータの適合度合いを制御するパラメータである．この C の値を減らすことは、正則化の強さを強めることを意味する．

3.8.2 L1 正則化付きロジスティック回帰

L2 正則化がモデルの複雑さを低減させる手法であることを 3.8.1 項で述べた．モデルの複雑さを低減させる別の手法として，L1 正則化がある．L1 正則化は，以下のように定義される．

$$\|\mathbf{w}\| = \sum_{j=1}^m |w_j| \quad (3.8.7)$$

L2 正則化と比較すると，二乗の和の重みが重みの絶対値の和に置き換わっていることがわかる．これより，L1 正則化を用いたロジスティック回帰のコスト関数は，以下のように表すことができる．

$$J(\mathbf{w}) = C \sum_{i=1}^n [-y^{(i)} \log(f(z^{(i)})) - (1 - y^{(i)}) \log(1 - f(z^{(i)}))] + \frac{1}{2} \|\mathbf{w}\| \quad (3.8.8)$$

L1 正則化によって，一部の特徴量の重みが 0 となり，次元が削減される．

3.8.3 線形サポートベクトルマシン

線形サポートベクトルマシンも線形分類器であるため，ロジスティック回帰と同様に，入力 $\mathbf{x}$ に対して式 3.8.1 と表すことができる．線形サポートベクトルマシンでは，重み $\mathbf{w}$ とバイアス b を決定するために，マージン最大化という手法を用いる．マージンは，正のクラスと負のクラスを分類する境界である決定境界（超平面）とそこから最も近いトレーニングサンプル（サポートベクトル）との距離で定義される．マージンと決定境界，サポートベクトルベクトルの関係を図 3.4 に示す [38]．

決定境界に沿った正のクラスと負のクラスの超平面はそれぞれ以下のように表すことができる．

$$\mathbf{w}^T \mathbf{x}_{pos} + b = 1 \quad (3.8.9)$$

$$\mathbf{w}^T \mathbf{x}_{neg} + b = -1 \quad (3.8.10)$$

ここで，ベクトルの長さを以下のように定義する．

$$\|\mathbf{w}\| = \sqrt{\sum_{j=1}^n w_j^2} \quad (3.8.11)$$

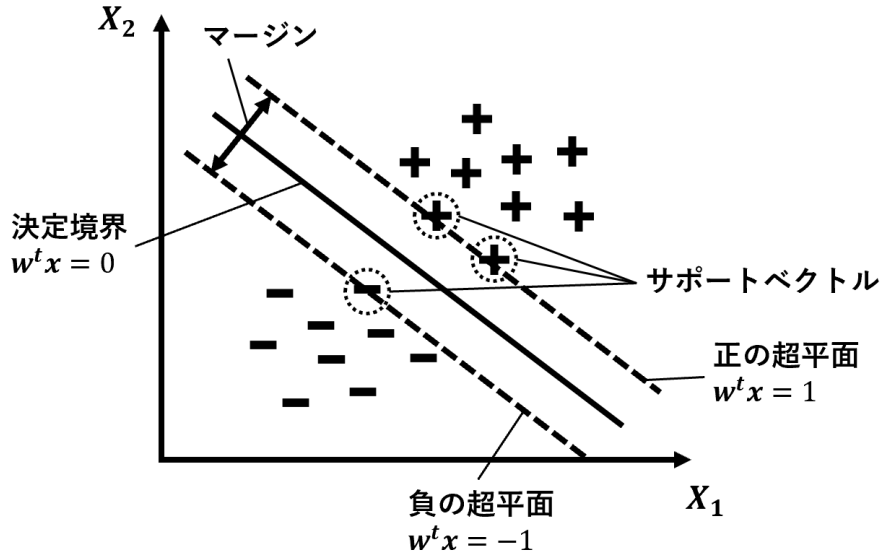


図 3.4 決定境界とサポートベクトルの関係

これより,

$$\frac{w^T(x_{pos} - x_{neg})}{\|w\|} = \frac{2}{\|w\|} \quad (3.8.12)$$

と変形することができる．式 3.8.12 の左辺は，正の超平面と負の超平面との距離であり，これはマージンである．マージンの最大化と $\|w\|$ の最小化は等価であり， $\|w\|$ の最小化の方が簡単であるため， $\|w\|$ の最小化が考えられる．

3.9 評価方法

まず，構築した分類モデルを評価するために，Leave-one-out nested cross validation（入れ子式の 1 つ抜き交差検証）[39]を行った．そして，ROC 曲線に基づく AUC による検出能力の評価を行った．また，特徴量の分析において，認知症患者のグループと非認知症者のグループとの間に差が見られるかを，ノンパラメトリック検定の 1 つで，対応のない 2 群の比較に用いられるマン・ホイットニーの U 検定を行い，効果量を算出した．以下の項に，詳細を述べる．

3.9.1 交差検証

モデル性能を評価するには、モデルにとって未知のデータを用いる。一般的なアプローチは、元のデータセットからトレーニングデータセットとテストデータセットに分割し、トレーニングデータセットを用いてモデルを学習し、テストデータセットを用いて評価を行う方法（ホールドアウト法）である。しかし、ホールドアウト法は、データの分割方法によっては、データの偏りが含まれる可能性があり、性能の評価に影響がおよぶという問題点がある。そこで、 k 分割交差検証法という手法を用いる。この手法では、元のデータセットをランダムに k 個のデータセットに分け、そのうちの $k - 1$ 個のデータセットを学習に用いて、残りの 1 個のデータセットで評価を行う。すべてのデータセットで評価するように、この操作を k 回繰り返す。

Leave-one-out cross validation（1 つ抜き交差検証）は、 k 分割交差検証の k が、データのサンプル数と同じに設定することで実現できる。主に、元データのサンプル数が少ないときに用いられるモデル評価の手法である。

3.9.2 ハイパーパラメータのチューニング

3.8.1 項で登場した逆正則化パラメータは、ハイパーパラメータとも呼ばれ、適切にチューニングすることでモデルの性能の改善が期待できる。ハイパーパラメータの最適化手法に、グリッドサーチがある。グリッドサーチは、しらみつぶしの網羅的に適切なハイパーパラメータを探索する手法である。

本研究では、交差検証を 2 回組み合わせた入れ子式の交差検証を用いた。入れ子式の交差検証の概念図を図 3.5 に示す。図 3.5 より、外側のループで、元のデータセットからトレーニングデータセットとテストデータセットに分割している。そして、内側のループで、分割されたトレーニングデータセットをトレーニングサブセットとバリデーション（調整）サブセットに分割している。内側のループでグリッドサーチを用いてハイパーパラメータをチューニングし、外側のループで最適なパラメータで学習する。

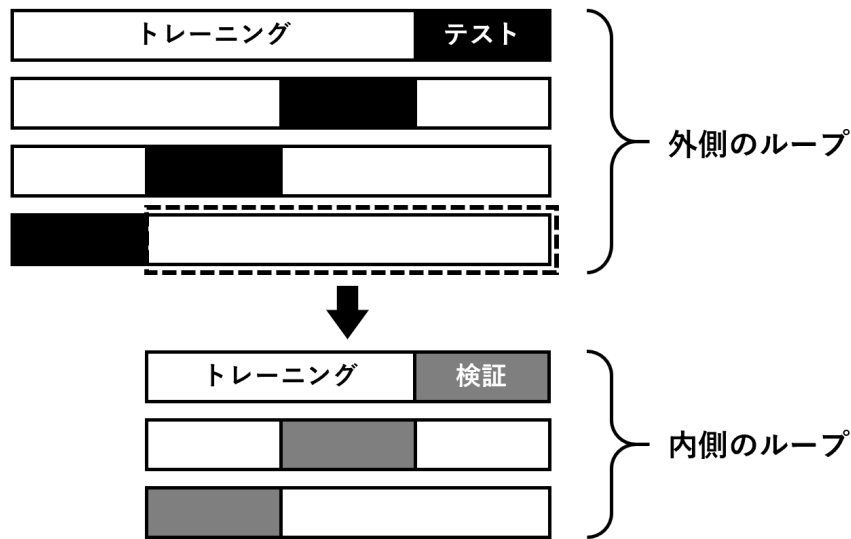


図 3.5 入れ子式の交差検証の概念図

3.9.3 ROC 曲線と AUC

分類結果をまとめる方法の一つに混同行列と呼ばれるものがある。これは、真陽性 (True Positive: TP)、偽陰性 (False Negative: FN)、偽陽性 (False Positive: FP)、陽陰性 (True Negative: TN) の 4 つの指標をまとめた行列である。まず、混同行列を図 3.6 に示す。

図 3.6 における真陽性は、予測が陽性 (Positive) であり、結果も陽性であるものである。偽陰性は、予測が陰性 (Negative) であり、結果は陽性のものである。偽陽性は、予測が陽性であり、結果は陰性のものである。陽陰性は、予測が陰性であり、結果も陰性のものである。

真陽性率 (True Positive Rate: TPR) と偽陽性率 (False Positive Rate: FPR) は、不均衡なクラスの問題 (例: がん患者と健常者の分類) に有効な指標である。TPR と FPR は以下のように表される。

$$TPR = \frac{TP}{TP + FN} \quad (3.9.1)$$

$$FPR = \frac{FP}{FP + TN} \quad (3.9.2)$$

		予測クラス	
		P	N
正解 クラス	P	真陽性 (TP)	偽陰性 (FN)
	N	偽陽性 (FP)	真陰性 (TN)

図 3.6 混同行列

受信者操作特性 (Receiver Operator Characteristic: ROC) 曲線は、分類器のしきい値を変えることによって計算される FPR と TPR から描くことができる。この際、横軸が FPR、縦軸が TPR となる。ROC 曲線下の面積 (Area Under the Curve: AUC) は、分類器の性能を表す指標となる。完全に分類が可能な場合、AUC は 1 となり、ランダムな分類の場合、AUC は 0.5 となる。

3.9.4 マン・ホイットニーの U 検定

本研究では、認知症患者のグループと非認知症者のグループとの間で Gap の値に差が見られるかどうかを検定するために、マン・ホイットニーの U 検定を用いた。マン・ホイットニーの U 検定は、対応のない 2 つのグループの差を検定する手法であり、正規性の仮定を必要としないノンパラメトリック検定の 1 つである。あるグループの標本を $X = (X_1, X_2, \dots, X_{n_1})$ 、もう一方のグループの標本を $Y = (Y_1, Y_2, \dots, Y_{n_2})$ とし、2 つのグループの標本をまとめたものを $Z = (X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2})$ とした場合の検定統計量 U は、次の手順で求めることができる。まず、 Z を大きいものから順に順位 R_i をつける (同順位の場合

は平均順位をつける). すると, X の順位和 R_X と Y の順位和 R_Y はそれぞれ,

$$R_X = \sum_{i=1}^{n_1} R_i, R_Y = \sum_{i=n_1+1}^{n_2} R_i \quad (3.9.3)$$

と表せる. そして,

$$U_X = R_X - \frac{n_2(n_2 + 1)}{2}, U_Y = R_Y - \frac{n_1(n_1 + 1)}{2} \quad (3.9.4)$$

とし, $U = \min(U_X, U_Y)$ で求めることができる.

また, マン・ホイットニーの U 検定における効果量は, 次の手順で求めることができる. まず, 検定統計量 U の平均 $E[U]$, 分散 $V[U]$ を求め, 次のように変換する.

$$Z = \frac{U - E[U]}{\sqrt{V[U]}} \quad (3.9.5)$$

そして,

$$r = \frac{Z}{\sqrt{(n_1 + n_2)}} \quad (3.9.6)$$

で求めることができる.

第 4 章 結果

本章では，4.1 節で音声言語を用いた認知症検出に有効な質問の分析結果を述べ，4.2 節で認知症検出の自動化の結果について述べる．

4.1 認知症検出に有効な質問の分析

表 3.2 より，質問をいくつかのカテゴリに分け，それぞれのカテゴリにおける認知症の検出能力を測定した．カテゴリ分けは以下の通りである．

- 自己認識に係る質問（Q2, Q11, Q12, Q13）
- 自身の関心に係る質問（Q3, Q4）
- 遠隔記憶に係る質問（Q5, Q6, Q7）
- 神経心理学的検査で使用されている質問（Q8, Q9, Q10）

なお，Q1 に関しては，上記のカテゴリとは別カテゴリに属すると考えたが，データ数が少ないため，今回の分析では除外した．

手動による音声区間検出と書き起こしで行い，L2 正則化付きロジスティック回帰，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を用いて分類した．その中で，最も検出能力（AUC）が高かったものとその正解率をカテゴリ別で表 4.1 に示す．

表 4.1 カテゴリ別の検出能力と正解率

	AUC	正解率
自己認識	0.67	0.62
関心	0.95	0.73
遠隔記憶	0.60	0.32
神経心理学的検査	0.63	0.73

これより，遠隔記憶に係る質問が最も検出能力が低く，自身の関心に係る質問が高い検出能力を示していることがわかる．

次に、質問ごとの Gap の値を図 4.1 に示す．先行研究 [35, 23] より、認知症患者の方が非認知症者よりも Gap の値が大きい傾向にあることが示されており、Q1, Q2, Q8, および Q9 においては同様の傾向であることを確認した．しかし、Q3, Q5, Q6, Q7, Q11, Q12, および、Q13 においては、認知症患者の方が非認知症者よりも Gap の値が大きいという傾向が見られなかった．

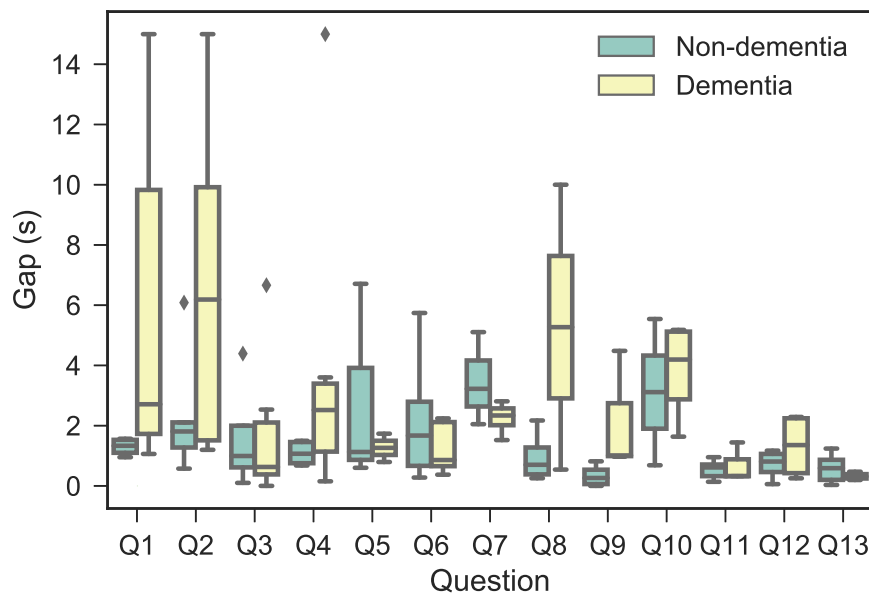


図 4.1 手動 VAD における各質問の Gap

4.2 認知症検出の自動化

まず、音声区間検出の評価として、FAR と FRR を計算した．その結果、FAR は 3.7%、FRR は 7.3% となった．

次に、手動 VAD + 書き起こしで最も検出能力が高かったモデルの ROC 曲線と自動 VAD + 音声認識で最も検出能力が高かったモデルの ROC 曲線を図 4.2 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．また、ROC 曲線の AUC と正解率をまとめたものを表 4.2 に示す．

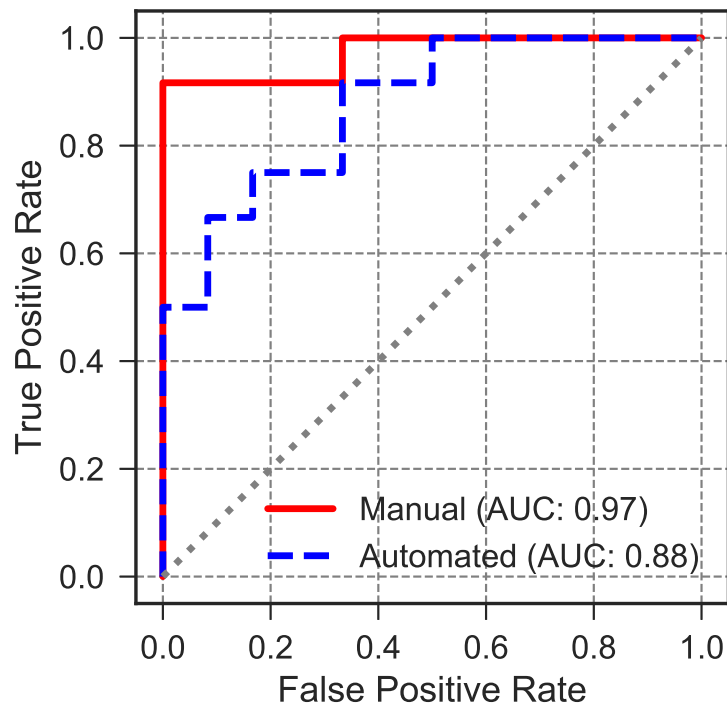


図 4.2 分類モデルの ROC 曲線. Manual は、手動 VAD を意味し, Automated は、自動 VAD を意味する. Manual と Automated のどちらも、入力が音声特徴量のみで、分類モデルが L1 正則化付きロジスティック回帰である.

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 VAD + 書き起こしでは、Gap, F0_range, Power_mean, Power_sd, 形容詞, Speech rate であった. また、自動 VAD + 音声認識では、Gap, F0_median, Pause_mean, 形容詞であった.

次に、L2 正則化付き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 VAD + 書き起こしにおける上位 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 VAD + 音声認識における上位 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、Gap が分類に有効な特徴量であることがわかる. ここで、非認知症者と認知症患者の Gap

表 4.2 分類モデルの評価. Sp. は音声特徴量を意味し, La. は言語特徴量を意味する.

	手動 VAD + 書き起こし		自動 VAD + 音声認識	
	AUC	正解率	AUC	正解率
MMSE	0.85	0.83	-	-
Sp. (SVM)	0.85	0.83	0.74	0.75
Sp. (LRL2)	0.90	0.83	0.85	0.79
Sp. (LRL1)	0.97	0.96	0.88	0.71
Sp.+La. (LRL1)	0.91	0.88	0.85	0.67
Sp.+La. (LRL2)	0.92	0.92	0.71	0.58
Sp.+La. (SVM)	0.95	0.92	0.71	0.58

を図 4.3 に示す. マン・ホイットニーの U 検定で p 値を計算すると, 手動 VAD で $p = 0.04$, 自動 VAD で $p = 0.02$ と有意差を示した. また, 効果量は, 手動 VAD で $r = 0.42$, 自動 VAD で $r = 0.47$ であった. これより, 先行研究 [35, 23] と同様, 非認知症患者よりも認知症患者の方が Gap が大きいことを確認した.

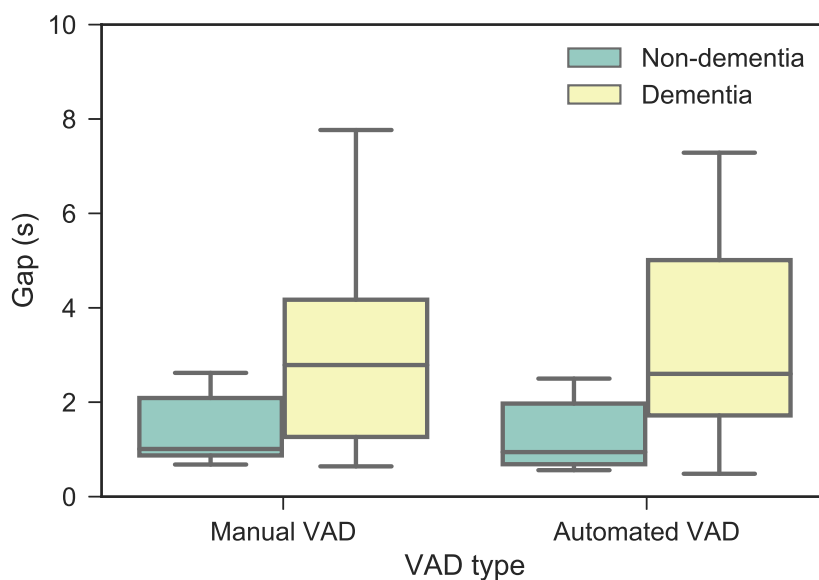


図 4.3 非認知症グループと認知症グループの Gap

次に，質問数の変化による検出能力の違いを分析した．質問数の変化による検出能力の違いを表 4.3 に示す．これより，質問数によって検出能力が変化していることがわかる．

表 4.3 質問数の変化による検出能力の違い

		1	2	3	4	5
手動 VAD + 書き起こし	Sp.	0.59	0.81	0.76	0.99	0.97
	Sp.+La.	0.96	0.58	0.75	0.94	0.91
自動 VAD + 音声認識	Sp.	0.52	0.86	0.79	0.84	0.88
	Sp.+La.	0.56	0.90	0.90	0.90	0.85

最後に，本研究の認知症患者には，AD のほかに NPH や MCI のタイプも含まれている．どのタイプの認知症でも，中核症状は同じであるが，障害の進み方は異なることがある．そこで，認知症患者を AD 型のみ ($n = 9$) にして分類を行った．最も検出能力 (AUC 値) が高かった結果は，手動 VAD + 書き起こしで 0.95，自動 VAD + 音声認識で 0.89 だった．どちらの結果も，入力が音声特徴量のみで，分類モデルは L1 正則化付きロジスティック回帰である．これより，AD 型認知症患者のみの場合でも，高い検出能力で認知症を検出することができた．

第 5 章 考察

5.1 質問分析

表 4.1 より，遠隔記憶に係る質問（Q5, Q6, Q7）が最も検出能力が低く，自身の関心・意欲に係る質問（Q3, Q4）が高い検出能力を示していることがわかる。

まず，Q5, Q6, Q7 は，他の質問よりも過去の話について尋ねており，高齢者が青年期・壮年期のころに活躍した有名人に関する質問である．一般的に記憶障害は，近時記憶から障害され，障害が進行するにつれて，遠隔記憶にまで障害が及ぶといわれており，早期認知症患者の場合は，遠隔記憶が障害されていない可能性が高い．つまり，今回の実験に参加した認知症患者は，遠隔記憶が障害されていなかったため，非認知症者と差が見られなかったと考える．これより，過去の有名人について尋ねるような，遠隔記憶に係る質問は，あまり効果的な質問ではないといえる。

次に，Q3, Q4 は，趣味や好きな歌を尋ねる質問で，自身の関心に係る質問である．認知症の早期段階から，無関心や意欲低下が認められることから，今回の実験に参加した認知症患者は，この影響を受けたものと考えられる．これらの質問は，音声言語を用いた認知症検出において有効であると考ええる。

Q8, Q9, Q10 は，神経心理学的検査に基づく質問であり，認知症検出において有効な質問である．しかし，今回の実験結果では，AUC 値で 0.63 と検出能力が高くなかった．この結果については，次のようなことが考えられる．まず，神経心理学的検査は，一連の課題の総合的なスコアから認知症の疑いがあるかどうか判定されるため，その検査の一部分だけを抜粋してきたとしても，効果的に認知症を検出できるとは限らないということ．次に，すべての課題において，回答が正しいかどうか重要であるため，回答中の音声言語に着目して作成されていないということ．これらが要因で，あまり検出能力が高くない結果となったと考える．

また，図 4.1 は，質問ごとに Gap の値を算出したものである．認知症患者の方が非認知症者よりも Gap の値が大きい傾向にあることが先行研究 [35, 23] よりで示されており，Q8, Q9 において同様の傾向であることを確認した．これらの質問は，時の見当識に関する質問であり，一般的に，認知症の初期段階から発症するといわれている．今回の実験に参加した認知症患者に関しても，時の見当識障害が認めら

れたと考えられるため、認知症患者の方が非認知症者に比べて Gap が大きい傾向が見られたと考える。Q11, Q12, Q13 については、自己認識に関する質問で、クローズド型の質問である。クローズド型の質問は、回答の内容が制限されるため、非認知症者との差が見られなかったと考える。

5.2 認知症検出の自動化

手動 VAD + 書き起こしの場合、音声特徴量のみを入力とした、L1 正則化付きロジスティック回帰における ROC 曲線の AUC が 0.98 となった。一方で、自動 VAD + 音声認識の場合、音声特徴量のみを入力とした、L1 正則化付きロジスティック回帰における ROC 曲線の AUC が 0.88 となった。自動化の検出能力は手動の場合と比べると、低い結果となった。しかし、MMSE の AUC 値は、0.85 であったことから、自動検出手法の結果は MMSE と同等以上の検出能力であるといえ、提案手法における有用性は示すことができたといえる。また、最も検出能力が高かった手法は、手動と自動のどちらも、入力音声特徴量のみで分類モデルが L1 正則化付きロジスティック回帰である。つまり、音声区間検出の結果のみが異なるため、音声区間検出の精度が検出能力に影響を与えていると考える (FAR: 3.7%, FRR: 7.3%)。

表 4.3 より、自動 VAD + 音声認識における検出能力は、2 問以上質問することで、比較的高く保っていることがわかる。また、質問の数に関しては、必ずしも多いほど検出能力が高くなるとは限らないことがわかるが、1 問のみで検出することは難しいことがわかる。質問数が増加することで、あまり効果がない質問の分類器への影響が小さくなり、検出能力が向上すると考える。

第 6 章 総括

本稿では、まず、質問応答から音声言語を用いて認知症を検出するために有効な質問を分析した。趣味に関する質問は、音声言語を用いた認知症検出で有効であることがわかった。無関心や意欲低下が音声言語に反映されることが理由であると考ええる。また、神経心理学的検査にも用いられる時の見当識に関する質問は、有効であることを確認した。しかし、過去の有名人に関する質問は、あまり有効でないことも確認した。これは、早期認知症患者が遠隔記憶の障害を発症していないことが理由であると考ええる。また、自己認識に関するクローズド型質問もあまり有効ではないことも確認した。これは、発話の内容が限定されることが理由であると考ええる。次に、コンピュータエージェントを用いて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。検出能力は、AUC で 0.88 を示した。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 の結果よりも高い。これより、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の数を増やす必要がある。次に、認知症の種類による影響についてあまり検討をしていないが、さらに調査する必要がある。そのためには、前述した通り、研究協力者の数を増やすことが必要になる。また、今回募集した研究協力者は、日本人のみであるため、外国人に対しても同じ手法が適用できるとは限らない。最後に、今回の実験では、コンピュータエージェントを使用した。人間は、コンピュータエージェントに対しても、自然で社会的な振る舞いをするといわれている [40, 41]。しかし、人間と全く同じ振る舞いをするわけではないため、人間対人間の質問応答と人間対コンピュータエージェントの質問応答による違いを検討する必要がある。

謝辞

本学情報科学研究科の中村哲教授には、主指導教員として常に熱心なご教示、ご鞭撻を受け賜り、本論文を執筆することができました。お忙しい中でも親身にご指導していただき、数多くの貴重なご助言を頂いたほか、研究活動を手厚くご支援いただき、大変有意義な研究生生活を送ることができました。心より感謝いたします。

本学情報科学研究科の安本慶一教授には、ご多忙の中にも関わらず、論文審査委員を引き受けていただき、心より感謝いたします。

本学情報科学研究科の田中宏季助教には、入学当初から様々な面でお世話になり、研究全般にわたり常に熱心なご指導を受け賜りました。心より感謝いたします。

本学情報科学研究科の須藤克仁准教授、吉野幸一郎助教には、数多くの貴重なご助言を頂いたほか、公私にわたり大変お世話になりました。心より感謝いたします。

本学情報科学研究科の Sakriani Sakti 特任准教授、安田圭志特任准教授、鈴木優特任准教授には、数多くの貴重なご助言を頂きました。心より感謝いたします。

また、知能コミュニケーション研究室秘書の松田真奈美様、林美保様には、諸手続きを始めとする様々な面でお世話になりました。心より感謝いたします。

知能コミュニケーション研究室の諸氏には、公私にわたり大変お世話になりました。心より感謝いたします。

最後に、学業生活を様々な面から支えてくれた家族や友人に感謝いたします。

参考文献

- [1] 日本神経学会. 認知症疾患診療ガイドライン 2010. 医学書院, 2010.
- [2] 尾崎 紀夫, 三村 將, 水野 雅文, and 村井 俊哉. 標準精神医学 第 7 版. 医学書院, 2018.
- [3] 日本神経学会. 認知症疾患診療ガイドライン 2017. 医学書院, 2017.
- [4] Martin Prince, Anders Wimo, Maëlen Guerchet, Ali Gemma-Claire, Yutzu Wu, and Matthew Prina. World Alzheimer Report 2015: The Global Impact of Dementia - An analysis of prevalence, incidence, cost and trends. Technical report, 2015.
- [5] Andrea Bradford, Mark E. Kunik, Paul Schulz, Susan P. Williams, and Hardeep Singh. Missed and delayed diagnosis of dementia in primary care: Prevalence and contributing factors, 2009.
- [6] Asier Basarab Adrian Alberdi, Ane Aztiria. On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey. *Artificial Intelligence in Medicine*, 71:1–29, Jul 2016.
- [7] Marshal F. Folstein, Susan E. Folstein, and Paul R. McHugh. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 1975.
- [8] Toru Hosokawa, Yoshiaki Yamada, Atsushi Isagoda, and Ryuichi Nakamura. Psychometric Equivalence of the Hasegawa Dementia Scale—Revised with the Mini-Mental State Examination in Stroke Patients. *Perceptual and Motor Skills*, 79(1):664–666, Aug 1994.
- [9] Vanessa Taler and Natalie A Phillips. Language performance in alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of clinical and experimental neuropsychology*, 30(5):501–556, 2008.
- [10] Vanessa Taler, Shari R Baum, Howard Chertkow, and Daniel Saumier. Comprehension of grammatical and emotional prosody is impaired in alzheimer's disease. *Neuropsychology*, 22(2):188, 2008.
- [11] Kathleen C Fraser, Jed A Meltzer, and Frank Rudzicz. Linguistic features

- identify alzheimer' s disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422, 2016.
- [12] 池田学. 認知症者のコミュニケーション. 高次脳機能研究 (旧 失語症研究), 35(3):292–296, 2015.
 - [13] Eiji Aramaki, Shuko Shikata, Mai Miyabe, and Ayae Kinoshita. Vocabulary Size in Speech May Be an Early Indicator of Cognitive Impairment. *PLOS ONE*, 11(5):e0155195, May 2016.
 - [14] Brian Roark, Margaret Mitchell, John Paul Hosom, Kristy Hollingshead, and Jeffrey Kaye. Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech and Language Processing*, 2011.
 - [15] Juan José G. Meilán, Francisco Martínez-Sánchez, Juan Carro, Dolores E. López, Lymarie Millian-Morell, and José M. Arana. Speech in alzheimer's disease: Can temporal and acoustic parameters discriminate dementia? *Dementia and Geriatric Cognitive Disorders*, 2014.
 - [16] Sylvester O. Orimaye, Jojo S.M. Wong, Karen J. Golden, Chee P. Wong, and Ireneous N. Soyiri. Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinformatics*, 2017.
 - [17] Aharon Satt, Ron Hoory, Alexandra König, Pauline Aalten, and Philippe H Robert. Speech-based automatic and robust detection of very early dementia. In *Fifteenth Annual Conference of the International Speech Communication Association*, pages 2538–2542, 2014.
 - [18] Eduardo Gonzalez-Moreira, Diana Torres-Boza, Héctor Arturo Kairuz, Carlos Ferrer, Marlene Garcia-Zamora, Fernando Espinoza-Cuadros, and Luis Alfonso Hernandez-Gómez. Automatic prosodic analysis to identify mild dementia. *BioMed research international*, 2015, 2015.
 - [19] Alexandra König, Aharon Satt, Alexander Sorin, Ron Hoory, Orith Toledo-Ronen, Alexandre Derreumaux, Valeria Manera, Frans Verhey, Pauline Aalten, Phillipe H. Robert, and Renaud David. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease.

- Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 2015.
- [20] Luke Zhou, Kathleen C Fraser, and Frank Rudzicz. Speech recognition in alzheimer's disease and in its assessment. In *INTERSPEECH*, pages 1948–1952, 2016.
 - [21] Laura Hernández-Domínguez, Sylvie Ratté, Gerardo Sierra-Martínez, and Andrés Roche-Bergua. Computer-based evaluation of Alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, 2018.
 - [22] Laszlo Toth, Ildiko Hoffmann, Gabor Gosztolya, Veronika Vincze, Greta Szatloczki, Zoltan Banreti, Magdolna Pakaski, and Janos Kalman. A Speech Recognition-based Solution for the Automatic Detection of Mild Cognitive Impairment from Spontaneous Speech. *Current Alzheimer Research*, 2018.
 - [23] Hiroki Tanaka, Hiroyoshi Adachi, Norimichi Ukita, Manabu Ikeda, Hiroaki Kazui, Takashi Kudo, and Satoshi Nakamura. Detecting Dementia Through Interactive Computer Avatars. *IEEE Journal of Translational Engineering in Health and Medicine*, 2017.
 - [24] Shohei Kato, Akira Homma, and Takuto Sakuma. Easy screening for mild alzheimer's disease and mild cognitive impairment from elderly speech. *Current Alzheimer Research*, 15(2):104–110, 2018.
 - [25] Daniel Walker Traci Reuber Markus Christensen Heidi Mirheidari, Bahman Blackburn. Dementia detection using automatic analysis of conversations. *Computer Speech & Language*, 53:65–79, Jan 2019.
 - [26] Leigh J Beglinger, Brenda Gaydos, Oranee Tangphao-Daniels, Kevin Duff, David A Kareken, Jane Crawford, Philip S Fastenau, and Eric R Siemers. Practice effects and the use of alternate forms in serial neuropsychological testing. *Archives of Clinical Neuropsychology*, 20(4):517–529, 2005.
 - [27] *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)*.

- American Psychiatric Association, 2000.
- [28] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)*. American Psychiatric Association, 2013.
 - [29] Louise Robinson, Eugene Tang, and John Paul Taylor. Dementia: Timely diagnosis and early intervention. *BMJ (Online)*, 2015.
 - [30] María Flavia Dorrego, Liliana Sabe, Agustina García Cuerva, Gabriela Kuzis, Cecilia Tiberti, François Boller, and Sergio E Starkstein. Remote memory in alzheimer’s disease. *The Journal of neuropsychiatry and clinical neurosciences*, 11(4):490–497, 1999.
 - [31] Anthony M Monacelli, Laura A Cushman, Voyko Kavcic, and Charles J Duffy. Spatial disorientation in alzheimer’ s disease the remembrance of things passed. *Neurology*, 61(11):1491–1497, 2003.
 - [32] 河月稔. 神経心理学的検査. *医学検査*, 66:11–21, 2017.
 - [33] 滝浦孝之 et al. 認知症スクリーニング検査. *広島修大論集. 人文編*, 48(1):347–379, 2007.
 - [34] David Wechsler. *WAIS-III, Wechsler adult intelligence scale: Administration and scoring manual*. Psychological Corporation, 1997.
 - [35] Francis J. Pirozzolo, Kathy J. Christensen, Kathleen M. Ogle, Edward C. Hansch, and W.Gary Thompson. Simple and choice reaction time in dementia: Clinical implications. *Neurobiology of Aging*, 2(2):113–117, jun 1981.
 - [36] Ronald M Harden and FA Gleeson. Assessment of clinical competence using an objective structured clinical examination (osce). *Medical education*, 13(1):39–54, 1979.
 - [37] Akinobu Lee and Tatsuya Kawahara. Recent development of open-source speech recognition engine julius. In *Proceedings: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference*, pages 131–137. Asia-Pacific Signal and Information Processing Association, 2009.
 - [38] 株式会社クイープ. *[第 2 版]Python 機械学習プログラミング達人データサイ*

エンティストによる理論と実践. インプレス, 2018.

- [39] Sudhir Varma and Richard Simon. Bias in error estimation when using cross-validation for model selection. *BMC bioinformatics*, 7(1):91, 2006.
- [40] 竹内勇剛 and 片桐恭弘. 人—コンピュータ間の社会的インタラクションとその文化依存性—互惠性に基づく対人的反応—. *認知科学*, 5(1):1_26–1_38, 1998.
- [41] Clifford Nass, Youngme Moon, Brian J Fogg, Byron Reeves, and Chris Dryer. Can computer personalities be human personalities? In *Conference companion on Human factors in computing systems*, pages 228–229. ACM, 1995.

発表リスト

学術論文誌（審査中）

[1] Tsuyoki Ujio, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo and Satoshi Nakamura. Fully Automated Dementia Detection from Responses to Questions Asked by Computer Agents. PLOS ONE. 2019.

国際会議 [査読有]

[1] Tsuyoki Ujio, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo and Satoshi Nakamura. Detection of dementia from responses to atypical questions asked by embodied conversational agents. Proc. Interspeech. 2018.

[2] Tsuyoki Ujio, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo and Satoshi Nakamura. Identifying Dementia Patients based on Behavioral Markers in Human-Avatar Interaction. WFSBP. 2018.

研究会 [査読無]

[1] 宇城毅犠, 田中宏季, 足立浩祥, 数井裕光, 池田学, 工藤喬, 中村哲. エージェントによる非定型質問への応答からの認知症検出. 研究報告高齢社会デザイン (ASD), 2018.

[2] (発表予定) 宇城毅犠, 田中宏季, 足立浩祥, 数井裕光, 池田学, 工藤喬, 中村哲. 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. 言語処理学会 (ANLP), 2019.