

## 発言、表情、行動に基づく 認知症患者の不穏度合い推定

重酒成海<sup>1</sup>, 山内雅明<sup>1</sup>, 辻寛<sup>1†</sup>, 下西英之<sup>1††</sup>, 村田正幸<sup>1</sup>, 杉田美和<sup>1†††</sup>, 木多道宏<sup>1†††</sup>

<sup>1</sup>大阪大学 大学院情報科学研究科  
<sup>†</sup>一般社団法人日本モンテッソーリア協会  
<sup>††</sup>大阪大学 サイバーメディアセンター  
<sup>†††</sup>大阪大学 大学院工学研究科



1

## 認知症患者と不穏

- **認知症患者**
  - 様々な脳の病気により、脳の神経細胞の働きが徐々に低下し、認知機能（記憶力、判断力等）が低下して、社会生活に支障をきたした状態
  - 世界で毎年1,000万人のペースで増加
  - 認知機能の低下により記憶や将来に対して不安を抱えている
- **不穏、BPSD**
  - 認知症患者が抱える不安やストレス、体の不調等が表面に現れた状態
  - 不穏：落ち着きがなくなり興奮した状態
  - BPSD：行動心理症状 (Behavioral and Psychological Symptoms of Dementia)
    - 例：怒りっぽくなる、妄想がある、意欲と元気が無くなる、興奮状態となる、暴言や暴力
  - 暴言や暴力といった攻撃的な行動は、認知症患者自身に限らず、一緒に生活する家族や、介護を行う介護職員等に対して心理的・身体的な負担
- **認知症患者の増加に伴い、介護職員の不穏対応負担が増加**
  - 早期対応により、患者本人や介護者の負担が軽減
  - 不穏の手前の段階である「不安」の状態を捉えることが重要

2

## CADATY index の提案 [1]

- **Caregivers Assess Dementia's Anxiety designed by Tsuji and Yamauchi index**
  - 認知症患者が感じている不安感情の強さを推定する指標
  - 不穏や BPSD の前段階である不安の状態を検出可能
  - 介護職員が経験的に会得している不穏の予兆を点数化したもの
- **点数化方法**
  - 認知症患者の「発言」「表情」「行動」等の日常の様子から、対象者の不安および不穏度合いを点数化
    - 対象者の認知症患者の普段の様子をよく知る介護職員等が評価
    - -1~-3 は不安状態 (不穏の予兆)、-3~-5 は不穏状態の目安

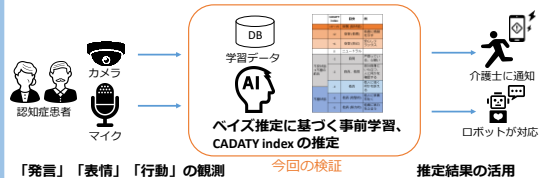
	CADATY index	目安	例
	+3~+5	興奮 (段状態)	他者に感謝を示す
	+2	安寧 (他者)	安心しリラックス
	+1	安寧 (自己)	戸惑っている、心細い
	0	ニュートラル	自分自身にいら立つ人に何かを確認する
不安状態 → 不穏の予兆	-1	自問	他人に強く何かを訴える
	-2	自責、他問	他人に暴言を吐く
	-3	他責	他者に暴力をふるう
不穏状態	-4	他責 (攻撃的)	
	-5	他責 (暴力的)	

[1] H. Tsuji, M. Yamauchi, M. Tsuji, and M. Kizu, "Appearance Transcription (CADATY Index and Memory Bank Edit History Service) Considered for a Psychological State Estimation Index for People with Dementia," in Proc. 20th Int. Workshop on Intelligent Information Systems, Aug. 2012.

3

## 研究目的とアプローチ

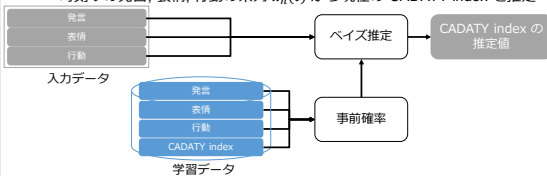
- **目的**
  - 認知症患者の発言、表情、行動といったマルチモーダルな観測情報から CADATY index を推定することで、不穏度合いを推定
- **アプローチ**
  - ベイズ推定を用いて、発言、表情、行動から CADATY index を推定
    - 入力：発言、表情、行動
    - 出力：CADATY index (不穏度合いの指標)



4

## 提案手法の概要

- ① **データセットの収集**
  - 映像カメラや音声マイク等の記録映像や音声から、時刻  $t$  における対象者の認知症患者の発言  $a(t)$ 、表情  $\beta(t)$ 、行動  $\gamma(t)$  の情報を抽出
  - 対象者の普段の様子についてよく知る介護者等が CADATY index の目安に基づき、時刻  $t$  における CADATY index の値  $y(t)$  を記録
- ② **事前確率の学習**
  - CADATY index が  $y_c$  のとき、発言  $a_t$ 、表情  $\beta_t$ 、行動  $\gamma_k$  が発現する確率を学習
- ③ **CADATY index の推定**
  - 時刻  $t$  の発言、表情、行動の系列  $x_n(t)$  から現在の CADATY index を推定



5

5

## ① データセットの収集

- **映像カメラや音声マイクから「発言」「表情」「行動」を取得**
  - 発言：音声マイクを使って、対象者の会話や独り言を記録した音声データの文字起こしを行い、得られたテキストデータから、対象者が高い頻度で発するフレーズを抽出
    - 例：「ありがとう」「ごめん」等の一般的に皆が使うフレーズ
    - 例：認知症患者が頻繁に発する「わからない」「もう帰らない」と
    - 例：対象者特有の口癖
  - 表情：映像カメラを使って、対象者の顔画像の抽出を行い事前に定義した表情ラベルのどれに該当するか分類し、表情ラベルを入力として蓄積
    - 対象者の顔のパーツの様子 (目や口角の角度、特定のしわの有無) に着目したものを事前に定義
      - 例：口元が緩み目を細める、眉間に縦じわ
  - 行動：映像カメラを使って、対象者の様子を抽出し、事前に定義した行動ラベルのどれに該当するか分類し、行動ラベルを入力として蓄積
    - 動作および移動の情報を事前に定義
      - 例：立つ、座る、右後方を見る、左後方を見る、特定の部位を触る
- **学習データ用の CADATY index をラベル付け**
  - 対象者をよく知る介護者等が観察を行い、一定時間間隔で-5~+5の値を記録

6

## ② 事前確率の学習

### ベイズ推定に基づく事前確率の算出

- ベイズ推定
  - 原因 A が発生した際に事象 B が発生する確率  $P(B|A)$  から、事象 B が発生した際に原因 A が生じている確率  $P(A|B)$  を算出
    - $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- CADATY index が  $y_c$  であるときに、発言  $\alpha_i$  が発現する確率  $P(\alpha_i|y_c)$  を算出
  - $P(\alpha_i|y_c) = \frac{\text{各時刻 } t \text{ のうち } y(t)=y_c \text{ かつ } \alpha_i(t)=\alpha_i \text{ となる } t \text{ の総数}}{\text{各時刻 } t \text{ のうち } y(t)=y_c \text{ となる } t \text{ の総数}}$
  - 観測時刻  $t = 0, 1, \dots, T$
  - 表情  $\beta_j$ 、行動  $\gamma_k$  が発現する確率  $P(\beta_j|y_c)$ 、 $P(\gamma_k|y_c)$  も同様
- CADATY index が  $y_c$  となる確率  $P(y_c)$  を算出
  - $P(y_c) = \frac{\text{各時刻 } t \text{ のうち } y(t)=y_c \text{ となる時刻}}{T}$

7

## ③ CADATY index の推定

### 現時時刻 $t$ の発言、表情、行動の系列 $x_n(t)$ から現在の CADATY index が $y_c$ である確率 $P(y_c|x_n(t))$ を推定

$$P(y_c|x_n(t)) = \frac{P(y_c)P(x_1(t), \dots, x_N(t)|y_c)}{P(x_1(t), \dots, x_N(t))}$$

$$= \frac{P(y_c) \prod_{n=1}^N P(x_n(t)|y_c)}{P(x_1(t), \dots, x_N(t))}$$

$$\propto P(y_c) \prod_{n=1}^N P(x_n(t)|y_c)$$

各事象を独立とみなすナイーブベイズを適用  
 $P(X|A, B, \dots, Z)$   
 $= \frac{P(X)P(A|X)P(B|X) \dots P(Z|X)}{P(A, B, \dots, Z)}$

### 推定された CADATY index の期待値から CADATY index の推定値 $y_{est}(t)$ を決定

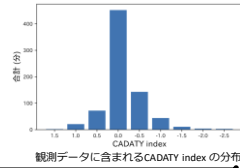
$$y_{est}(t) = \frac{\sum_c y_c P(y_c) \prod_{n=1}^N P(x_n(t)|y_c)}{C}$$

$C = 1, 2, \dots, C$

8

## 評価用データセットの取得

- 取得場所
  - サービス付き高齢者向け住宅 (サ高住) の共有スペース (1階リビング、廊下)
- 対象者
  - 1名: 自立移動可能な90歳代の認知症女性
- データ収集期間 (6日間)
  - 15:00~19:00 から最大4時間
- 観測データ
  - 発言、表情、行動
    - 手動で記録し、映像、音声記録を用いて補充
  - CADATY Index (正解ラベル)
    - $y_c = \{+5.0, +4.5, \dots, -4.5, -5.0\}$
    - 映像、音声記録等を参考に当該施設で長期間被験者の様子を観察した者 (社) が1分毎に記録
    - 実際に観測データに含まれた値は +1.5~-2.5 の範囲



本実験は、大阪大学大学院理工学部の研究開発推進委員会による実験許可 (許可番号: 202305)、および大阪大学大学院理工学部の研究開発推進委員会による実験承認 (承認番号: 4-7-1) を得て実施した。

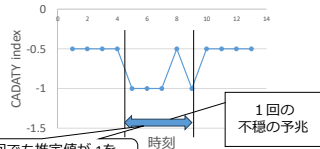
9

9

## 評価指標

### 推定値 $y_{est}(t)$ と正解値 $y(t)$ を比較し、どれだけ不確合いを推定できたかを確認する評価指標を設定

- 正解値  $y(t)$  が -1.0 以下となったタイミングを何回推定できるかで評価
  - 正解値  $y(t)$  が -1.0 以下の場合は不穩の予兆と考えられる
  - 正解値  $y(t)$  が -1.0 以下の値を連続して記録する期間 (ただし  $y(t)$  が -1.0 を上回った場合でも1分以内であれば連続とみなす) を1回と計算し、期間内に推定値  $y_{est}(t)$  が -1.0 以下の値をとることができたかどうか、つまり不穩の兆候を検出できたかどうかを確認
- 各時刻における推定値  $y_{est}(t)$  と正解値  $y(t)$  との差分を計算し、誤差の平均、誤差の中央値、誤差の最大値、誤差の標準偏差を算出
- 各モーダルの組み合わせを変更した場合についても同様に評価し、各モーダルの組み合わせによる推定の有効性を評価



この間に1回でも推定値が-1を下回りアラートを上げる必要

10

10

## 3モーダルの情報を用いた推定結果 (検出できた場合)

- 発言、表情、行動の全てのモーダルが欠けることなく観測できている場合、ほとんどの不穩の予兆が検出可能
  - 複数のモーダルを掛け合わせることで、CADATY index の推定範囲が限定されたため
  - 図7: 4回の予兆のうち3回を検出
    - モーダルを掛け合わせて精度が向上した例: 15:44 に「アホ」という発言、耳を覆う、顔を前後に揺らす、右後方を見るという行動が観測

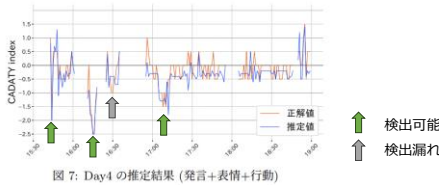


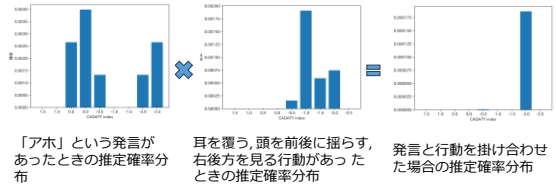
図7: Day4の推定結果 (発言+表情+行動)

11

11

## 3モーダルの情報を用いた推定結果 (検出できた場合)

- 発言、表情、行動の全てのモーダルが欠けることなく観測できている場合、ほとんどの不穩の予兆が検出可能
  - 複数のモーダルを掛け合わせることで、CADATY index の推定範囲が限定されたため
  - 図7: 4回の予兆のうち3回を検出
    - モーダルを掛け合わせて精度が向上した例: 15:44 に「アホ」という発言、耳を覆う、顔を前後に揺らす、右後方を見るという行動が観測



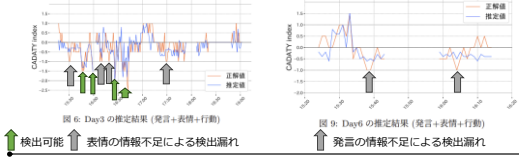
「アホ」という発言があったときの推定確率分布 × 耳を覆う、顔を前後に揺らす、右後方を見る行動があったときの推定確率分布 = 発言と行動を掛け合わせた場合の推定確率分布

12

12

### 3モーダルの情報を用いた推定結果 (検出できなかった場合)

- データの一部分が取得できない場合
  - 表情の情報不足 (図6)
    - 映像カメラの画角に顔が映っていない
    - カメラの増設や他モーダルの活用によりある程度対応可能
  - 発言の情報不足 (図9)
    - 声がいさくフレーズが聞き取れなかった
    - かなり精度が低下
    - カメラから何かを話している状態は判別可能
    - 音声マイクの性能向上や設置位置の工夫による改善
- 生活の場においてどこで発言を行うか等は予想できない部分も多く、このような場合でも他のモーダルから補足可能とする方法を検討中



13

### 各モーダルを組み合わせた評価結果

- 複数のモーダルを活用することで検出精度が向上
  - 発言+行動 により精度が向上
    - Day3で発言 → 発言+行動 で検出数が 2 → 5
- 今回のケースでは表情による不穏の予兆の検出回数は少なかつた
  - 発言+行動 → 発言+行動+表情 で検出回数(は同じ)
  - 発言 → 発言+表情 で誤差が 1.04 → 0.78 に改善
    - 表情により正の値の範囲が多く検出できたため
  - 表情のデータ数が少なかつたことも影響
- 不穏の予兆の検出に重要なモーダルは、行動 ≥ 発言 > 表情 とみられる
  - 性格、性別、認知症の型等で個人差があると考えられるため確認予定

各モーダルの組み合わせにおいて推定できた不穏の予兆の回数

モーダル	Day3	Day4	Day5	Day6	モーダル	誤差の平均	中央値	最大値	標準偏差
発言	2	1	0	0	発言	1.04	1.00	2.50	0.41
表情	0	1	0	0	表情	0.98	1.00	2.50	0.52
行動	2	1	1	0	行動	0.59	0.37	2.34	0.52
発言+表情	2	2	0	0	発言+表情	0.78	0.79	2.50	0.46
表情+行動	2	1	1	0	表情+行動	0.51	0.37	2.34	0.49
発言+行動	5	3	1	0	発言+行動	0.44	0.28	2.34	0.42
発言+表情+行動	5	3	1	0	発言+表情+行動	0.37	0.28	2.34	0.39
正解値	11	4	4	2					

14

### まとめと今後の課題

- まとめ
  - 認知症患者の発言、表情、行動といったマルチモーダルな情報から、認知症患者の不安の強さの指標である CADATY index を推定する手法を提案
  - ベース推定により、発言、表情、行動が全て観測できる場合、不穏の予兆を推定可能
  - 発言および表情の情報欠損した場合、推定できなかった
    - 映像カメラや音声マイクの増設等で対応可能な部分も
    - 他モーダルの利用、確率分布のばらつき具合の利用で対応検討
- 今後の課題
  - 表情、発言、行動を AI 抽出した場合の評価
    - 表情分析技術や行動分析技術を利用して推定可能か確認
    - 本稿では映像や音声データから手動で情報を抽出
  - CADATY index の事前推定によるさらに早期の不穏対応
  - 被験者を増やすこと

本研究は、大阪大学 NEC Beyond 5G 協働研究所における成果です。実験にご協力頂いた、一般社団法人日本モンテッソーリアクア協会・サービスピック高齢者向け住宅「柴原モカメゾン」の皆様へ感謝いたします。

15