

発話の韻律情報と頭部移動情報に基づく 会話態度評価値予測モデルの有用性評価

Evaluating the Usefulness of a Prediction Model for the Conversation Attitude Assessment
based on the Prosodic and Head Movement Information

山中 俊貴[†]

高瀬 裕[‡]

中野 有紀子[‡]

Toshiki Yamanaka

Yutaka Takase

Yukiko Nakano

[†] 成蹊大学理工学研究科理工学専攻

[‡] 成蹊大学理工学部情報科学科

[†] Graduate School of Science and Technology, Seikei University

[‡] Dept. of Computer and Information Science, Seikei University

In the recent aging society, the number of older adult with Dementia is increasing. In such a situation, recognizing and assessing cognitive and health status of the elderly with Dementia are important. In this study, we proposed a prediction model, which uses speech and visual features, for assessing conversational activity of the elderly during interacting with a listener agent. The coefficient of determination of this model was 0.469. In order to evaluate the model, we applied the model to new data including conversations with a listener robot. We employed a binary classification ("good" and "not-good"), and found that our model can predict whether the user's conversation attitude is good or not-good with 69.2% accuracy.

1. はじめに

高齢化社会の中で、認知症高齢者の数が年々増加傾向にある。厚生労働省の調査[1]によると、65歳以上の認知症高齢者の数は462万人(平成24年)、高齢者の約15%、全国のMCI(正常でもない、認知症でもない状態の者)の数は約400万人(平成24年)と推計されている。また日常生活自立度II以上の高齢者数は2010年には280万人であった。この数が2025年には470万人と推定され、介護や見守りが必要な高齢者がますます増える傾向にある。

しかしながら、介護士の数は不足している状況にあり、このような状況を改善するために、認知的障害を持つ高齢者の支援を目的とした情報技術の研究・開発が活発に行われている。その1つとして、患者の状態を把握、記録し、必要に応じて介護者や家族に通知を行う見守りサービスは、今後さらに増加が予測される一人暮らしの認知症高齢者やその家族に向けたサービスとして必要性が高まると考えられる。

2. 研究の背景と目的

認知症高齢者の日々の健康状態・心理状態・認知状態の把握を目指し、我々は、在宅で患者の体調等について尋ねることにより話の聞き役にもなる傾聴エージェントの開発を進めている。

図1に我々が開発した初期の傾聴エージェントシステム[5]と高齢者とのやり取りの様子を示す。本システムでは、エージェントが高齢者に対して日ごろの生活の様子や趣味などについて質問し、高齢者が答えるという一問一答式の形で会話を行う。エージェントは、高齢者の発話の韻律情報の特徴に応じて頷き・相槌などを行う。傾聴エージェントとの対話終了後に生成されるログファイル(xml形式)には、エージェントの発話終了時間やユーザの発話開始/終了時間、音声認識結果などが保存されており、加えて対話時の各発話の音声ファイル(wav形式)、動画ファイル(flv形式)が保存される。

我々はこれまでに、この傾聴エージェントと高齢者との会話を



図1. 傾聴エージェントと高齢者の対話の様子

分析し、高齢者がエージェントからの問いかけに対して、テンポよく、リズムよく、また積極的に応答しているか否かを会話活動評価値と定義し、この会話活動評価値を韻律情報から自動的に判定する技術を開発した[2].

一方、会話活動の良好性を左右するのは韻律的な情報だけではない。Bavelasら[3]は、会話中に身振りで対話のトピックに関連する情報を伝えるのと同様に、表情でもトピックに関連する情報を伝えると述べている。表情は、発話した単語と2秒未満で同期し、対面での対話では、聞き手は、頻繁に表情を変化させている。

そこで我々は韻律情報に加えて、視覚的な情報である頭部移動情報を使って会話活動評価値を推定する重回帰モデルを以下の3種類作成した。[4]

- I. 韻律情報のみを用いた推定モデル
- II. 韻律情報および、ユーザ発話中の頭部移動情報を用いた推定モデル
- III. 韻律情報およびエージェントの問いかけ開始からユーザの応答発話終了までの区間の頭部移動情報を用いた推定モデル

その結果、韻律情報およびエージェントの問いかけ開始からユーザの応答発話終了までの区間の頭部移動情報を用いた推

定モデルは他のモデルに比べて決定係数が最も高く、データに対するあてはまりが最も良いことが分かった

そこで本研究では、作成した予測モデルに未知のデータを適用することで、その予測モデルの有用性を評価することを目的とする。評価方法としては 2 クラス分類による分類精度によって評価する。

3. データ分析

本研究で会話活動評価推定モデルの構築のために用いた対話データは、先行研究[2]で得られた、傾聴エージェントと認知症高齢者との対話データである。被験者である高齢者は計 11 名(男性 8 名, 女性 3 名, 平均年齢:74.2 歳)であり, MMSE (認知機能検査) の平均スコアは 22.2 点である。各ユーザにエージェントから 30 問の質問を行ったが、データ欠損等の理由により計 236 の質問・応答ペアを分析対象データとした。以下に韻律情報と視覚的情報である頭部動作情報のパラメータについて記す。ただし、分析の詳細については参考文献[4]に詳しく述べる。

3.1 会話活動評価

心理学研究において提案された参考文献[6]の対人コミュニケーションに関する評価指標を参考に、会話活動評価用の質問紙を作成した。質問紙は、話す姿勢・表情・姿勢に関する 20 の質問からなる。作成した質問紙と各ユーザのエージェントとの対話の様子を収めたビデオを用いて、外部評価者による会話活動の評価を行った。評価者は大学生 15 名(男性 11 名, 女性 4 名), 回答者の平均年齢は 20.8 歳(19~22)であった。各質問に対して 5 段階のリッカート尺度を用いて、全評価者 15 名の結果の平均値をそのユーザの会話活動評価値とした。評価値の平均は 64.5 点(最大値 87.4, 最小値 52.6)であった。

3.2 韻律情報の分析対象パラメータ

発話の韻律情報としてモデル構築に用いた予測パラメータは以下の 4 つである。

- I. ピッチ: ユーザの発話区間のピッチ(声の高さ)の平均
- II. 発話長: ユーザの発話開始時間から発話終了時間までの時間
- III. 抑揚: ユーザの発話区間におけるピッチの最高値と最低値の差
- IV. 反応時間: エージェントの発話終了からユーザが発話を開始するまでの時間

3.3 視覚情報の分析対象パラメータ

視覚情報としてモデルの構築に用いた予測パラメータは以下 2 つである。

- I. フレームごとの頭部移動量
- II. 頭部移動量の離散フーリエ解析の結果

4. 会話活動評価値予測モデル

3.2 節で用いた 4 種類の韻律情報(ピッチ, 発話長, 抑揚, 反応時間)の説明変数に加え, 3.3 節で用いた頭部移動情報(頭部移動量, 頭部移動量の離散フーリエ解析の結果)を説明変数に用い, 3.1 節で評価者によるビデオ観察によって得られた

会話活動評価値を予測するモデルを重回帰分析により作成した。

ユーザの韻律情報はユーザがエージェントの問いかけに対して応答している区間を分析対象としている。一方, ユーザの頭部動作はエージェント発話中にユーザが聞き手として振舞っている時間の動きと, エージェントの問いかけに対して応答する, つまり発話者として振舞っている時間に分けることができる。今回はエージェントの質問開始からユーザの応答終了時間までを分析区間とした。

エージェント発話開始から, それに対するユーザ応答発話の終了までのフレームを 768 のデータ点として正規化し, 離散フーリエ解析を行った。離散フーリエ解析の結果である 385 点のうち 0Hz を除く 384 点を 4 点ずつ統合することで, 約 0.125 Hz 刻みに 96 の周波数帯に等分割し, 各周波数帯 F1~F96 における頻度を求めた結果を説明変数として利用した。したがって説明変数 F1 は 0.032 Hz, 0.064 Hz, 0.096 Hz, 0.128 Hz の頻度の和となっている。

ステップワイズ法を用いた重回帰分析の結果を表 1 に示す。本モデルの自由度調整済決定係数 R^2 は 0.469 (重相関係数 R は 0.699) となった。標準化係数ベータについてはピッチが最も高く, モデルへの影響度が最も高く, 次に縦方向の頭部移動量を示す特徴量 y , その次に F5 (0.513~0.64 Hz 付近)をはじめとする, 6 種類の周波数帯が有意なパラメータとして抽出された。

表 1 韻律情報および頭部移動情報を用いた予測モデル

説明変数	標準化されていない係数		標準化係数 ベータ
	β	標準誤差	
(定数)	50.227	4.103	
ピッチ	4.553	0.543	0.467
y	5920.790	1641.185	0.221
F5	-165.017	54.922	-0.162
F3	142.190	40.336	0.190
F67	713.443	210.804	0.175
F46	369.920	145.528	0.129
F78	-433.630	207.925	-0.106
F1	52.036	25.431	0.115

5. 会話活動評価値予測モデルの評価

4 章で作成した会話活動評価値予測モデルの有用性の評価をモデル作成に使われていない未知のデータを使って行った。今回評価に使ったデータは 2 種類の方法で取得している。その方法について以下に説明する。

5.1 データ収集

5.1.1. 傾聴エージェントによるデータ収集

図 2 に示す新しい傾聴エージェントシステムを開発し, これに伴いアニメーションキャラクターのデザイン改良やモーションの追加を行った。

エージェントの発話バリエーションをモデル作成時の 30 個から 300 個に拡張し, 24 種類の話題(音楽, 動物, 学生時代など)を設定, 各話題に 5 回分の発話を 1 セットとした発話セットを複数作成した。これらの発話セットをランダムに選択することによりエージェントから問いかけの発話を決定した。



図 2. 評価データ収集に用いたエージェント

なお、評価に用いたデータは、介護施設に約 1 ヶ月本システムを常設し、介護施設の方にシステムを起動してもらうことにより収集した。

5.1.2. ロボットによるデータ収集

傾聴エージェントの技術をロボットに適用し、図 3 に示す傾聴ロボットを開発し、介護施設で認知症高齢者の方に使用していただいた。なお、ロボット使用時は介護施設の方ではなく、実験者がロボットの起動などを行った。



図 3. 評価データ収集に用いたロボット

ロボットもエージェントと同じく、高齢者の発話の韻律情報の特徴に応じて頷き・相槌などを行う。傾聴エージェントとの対話終了後に生成されるログファイル(xml 形式)には、ロボットの発話終了時間やユーザの発話開始/終了時間、音声認識結果などが保存されており、加えて対話時の各発話の音声ファイル(wav 形式)、動画ファイル(flv 形式)が保存される。また、ロボットの問いかけ発話についてもエージェントと同じく、300 個の質問を用意し、24 種類的话题を設定している。ただし、ロボットでは、ロボットの質問に対するユーザの応答が肯定か否

定を推定するモデル[7]を実装し、図 4 示すように、肯定の場合は同じ話題に関する質問、否定の場合は別の話題を選び質問できる機能を追加した。

5.2 評価方法

モデル作成時と同じ方法で韻律情報および頭部移動情報のパラメータを抽出し、4 章で作成したモデルに適用した。次に、これにより得られた会話活動評価値の推定値を中央値または平均値により“よい”、“悪い”の 2 クラスに分類した。3.1 節で収集した人による評価結果についても同じく中央値と平均値により“よい”、“悪い”に分類した。この結果と推定値を用いた分類結果とを比較し、正解率を算出した。

データの欠損等があったセッションを除き、65 の質問・応答ペアから得られたデータを用いて会話活動評価値推定モデルの評価を行った。

5.3 評価結果

5.2 で得られたデータをモデルに適用し、会話活動評価値を“よい”、“悪い”の 2 クラスに分類した結果を表 2 に示す。

表 2 会話活動評価値推定モデルの正解率

	正解率(%)
中央値による 2 値分類	69.23
平均値による 2 値分類	69.23

各データに対する中央値と平均値がほぼ同等であったため、今回の分類にはどちらを採用しても同じ結果となった。

6. 考察

評価用のデータは、主に 2 つの点でモデル作成時のデータと異なっていた。1 つ目はモデル作成時のデータ収集では、ユーザ全員に固定の 30 個の質問を行うという会話活動であったが、評価時のデータ収集では、エージェントシステムでは、10 個程度の質問をランダムに行い、ロボットではユーザの反応に応じて 10 個程度の質問を行った。2 つ目は、会話を行う相手として傾聴エージェントだけでなく、ロボットを使ったデータも含まれている点である。本研究の結果は、データ収集時のこのような差異を含みつつ 69%の正解率が得られたことを示している。このことから、作成したモデルに使用されているパラメータは、会話の内容やインターフェースが異なっても有用であると言える。

7. まとめ

本研究では、傾聴エージェントと会話をする認知症高齢者のビデオを観察してもらい、質問紙による評価を行うことで会話活動評価値を設定し、韻律情報および頭部移動情報を用いて重回帰分析により会話活動評価値を予測するモデルを提案した。さらに、新にロボットを加えた評価用データを収集し、未知のデータによる 2 クラス分類の正解率による評価を行った。その結果、会話活動評価値が“よい”か“悪い”かの判定において、正解率は 69.2%であった。今後はモデルの性能向上と評価の信頼性の向上のためにより多くのデータを収集する必要がある。また会話活動評価値予測モデルを傾聴エージェントに実装し、会話活動評価値に応じた対応が可能なエージェントの実現を目指す。

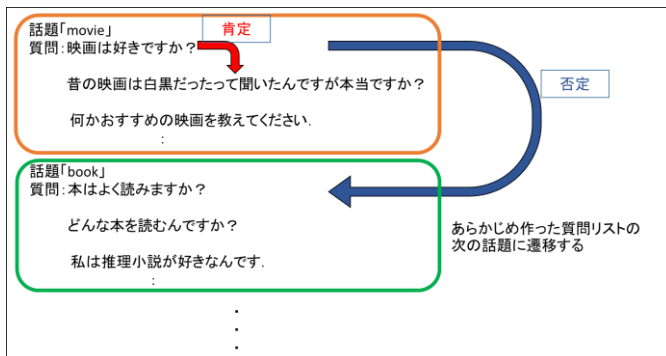


図 4. Robot 実験における質問の遷移

参考文献

- [1] 「認知症施策の現状について-厚生労働省-社保審-介護給付費分科会(115回)参考資料1」, 厚生労働省
<http://www.mhlw.go.jp/stf/shingi2/0000065658.html> : 参考資料 1
- [2] 野中 裕子, 高橋 明秀, 林 佑樹, 中野 有紀子: 発話中の非言語情報に基づく認知症高齢者の状態把握 , 第 28 回人工知能学会全国大会, 2H3-NFC-04a-3 (2014 年 5 月)
- [3] Bavelas, J. B. , & Gerwing, J. : Conversational hand gestures and facial displays in face-to-face dialogue(2007)
- [4] 山中 俊貴, 高瀬 裕, 中野 有紀子: 発話の韻律情報と頭部移動情報に基づく高齢者の会話活動評価値の予測, HCG シンポジウム 2015, (A-4-2)
- [5] 比企野純太, 中野有紀子, 安田清: 会話エージェントを利用した認知症患者のためのコミュニケーション支援, 第 73 回情報処理学会全国大会 4ZA-7 (2011)
- [6] 堀洋道: 『心理測定尺度集 I ~ VI』, サイエンス社
- [7] 功刀広基: 「認知症患者によるエージェント質問への応答における肯定／否定推定モデル」, 成蹊大学卒業論文 (2015)