

多肢選択問題の音声認識を用いた解答方法の一検討 A Study of Method of Answering Multiple Choice Questions Using Speech Recognition

藤村 誠[†] 岩崎 純弥[†] 渡邊 毅[†]
Makoto Fujimura Junya Iwasakii Tsuyoshi Watanabe

1. はじめに

医療面接では、患者の診断において病歴聴取などの情報収集および医師患者間の良好な関係構築が目的となる[1]。医療面接の演習では、医師患者間の関係構築を考慮した対話内容の習得のため、学習者は適切な文を選択して対象との会話を進めることを学ぶことになる。学習の初期段階では、学習者に多肢選択問題として、数個の文が提示され選択する方法がとられる場合がある。対話に用いるべき適切な文および用いるべきではない不適切な文を同時に提示することで、学習者の理解を深める効果が期待できる。

我々は、この多肢選択問題で提示された文候補の選択方法として、文を読み上げて選択する方式を検討している。音声認識を用いて選択文を判定する方法であり、選択肢の文に付与された番号を単純に選ぶ方法より有効であると考えられる。

しかし、現時点の音声認識精度では、学習者の発声内容の認識結果をそのまま用いることは難しい。そこで、学習者が 5 択問題文を読み上げ、音声認識によりテキストデータに変換し、選択した選択肢文を判定するが、テキストデータと選択肢の文とを比較して、選択された文を単純に推定することは難しい。一方、文字列のスペル誤り訂正の一つに、文字間の相違の程度を表すものとしてレーベンシュタイン距離を利用する方法がある[2]。そこで、変換されたテキストデータと選択肢の文との比較方法として、文の類似度を計算する方法を検討する。この方法により、音声認識精度がある程度低くても、十分な成果率が得られると考えられる。

以降、2 で医療面接演習システムと提案する読み上げによる解答方法の提案を行い、3 の評価実験では、正答率を測定し、提案した方法を検証した。

2. 医療面接演習システム

2.1 システムの概要

図 1 に医療面接演習システムの概要図を示す。学習者はヘッドマウントディスプレイ（以後、HMD と称する）を装着し、仮想空間中の仮想患者を相手にして、図 2 に示す手順で対話を行う。まず、学習者が 5 択選択肢の対話文を選んで読み上げる。次に、読み上げた音声テキストデータに変換する。次いで、変換したテキストデータと 5 択選択肢のそれぞれの文との間のレーベンシュタイン距離を求める。最後に、最小のレーベンシュタイン距離の文章を選択されたものと判定する。

なお、対話内容は、JSON(JavaScript Object Notation)形式のファイルに記述したものを用いる、

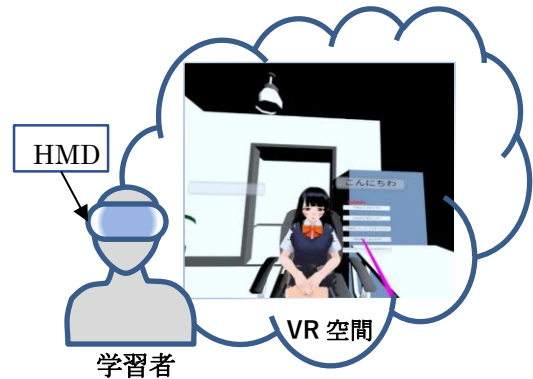


図 1 医療面接演習システムの概要図

1. 学習者が質問文を読み上げる
2. 音声認識を行い、テキストデータに変換する
3. 変換したテキストデータと、5 択の質問文それぞれとのレーベンシュタイン距離を計算する
4. レーベンシュタイン距離が最小の質問文を求め、学習者が選択した質問文とする

図 2 対話手順

2.2 多肢選択問題の音声認識を用いた解答方法

レーベンシュタイン距離は、ある文字列からもう一方の文字列に変わるために必要な、1 要素の挿入、削除および置換のそれぞれの操作の回数合計である。文字列 $X = x_1, x_2, \dots, x_m$ と $Y = y_1, y_2, \dots, y_n$ の間のレーベンシュタイン距離 $d(i, j)$ は、 X から Y に文字列を変換するときの編集操作の最小コストとして定義され、次の式によって求められる。

$$d(i, j) = \begin{cases} 0, & i = 0 \text{ and } j = 0 \\ i, & i > 0 \text{ and } j = 0 \\ j, & i = 0 \text{ and } j > 0 \\ \min \begin{cases} T[i-1][j] + 1 \\ T[i][j-1] + 1 \\ T[i-1][j-1] + \delta_{x[i-1] \neq y[j-1]} \end{cases}, & \text{otherwise} \end{cases}$$

また、レーベンシュタイン距離は、文の間の差異が大きいくほど数値が大きくなるため、文字列 X と Y との類似度 S を次式で定義する。

[†]長崎大学 Nagasaki University

$$S = \frac{d(i, j) - L}{L}$$

ここで、 $L = \begin{cases} i, & i \geq j \\ j, & i < i \end{cases}$

3. 評価実験

3.1 実験目的および条件

提案した 5 択選択肢の読み解答方法の検証のため、評価実験を行った。質問は 1 セット(10 問)として合計 5 セット(50 問)を被験者に提示し読み上げを行わせる。このとき、読み上げた文と提示された 5 択選択肢の質問文の正解率を測定した。

被験者は 20 代前半の男性 9 名および女性 2 名である。実験環境は、二重壁構造、二重扉の外部音を遮断する部屋である。被験者には、HMD として Meta Quest2 を装着させる。Meta Quest2 のマイクとスピーカーを用いて、音声入力と音声の提示を行った。また、本実験では音声認識に IBM 社の Watson の Watson Speech To Text を用いている。

5 つの質問の選択肢の構成は、基本的に良いとも悪いとも言えない質問、1 番良い質問、禁忌の質問となります。禁忌の質問とは、医療面接において、医師が患者に絶対発言してはいけない質問を意味しており、禁忌の質問を選択した場合には仮想患者が返答することなく終了します。

1 つの選択肢の内容に関しては、質問を表す ID、質問、返答、次の質問セットの番号、スコアの情報を記述します。

提案した多肢選択肢の音声認識による判定方法を用いる。音声認識による 5 択質問の選択として、被験者に 5 つの質問の選択肢から 1 つの文を読み上げてもらい、発声した音声と 5 つの質問の選択肢の類似度をそれぞれ計算し、最も類似度が高いものを選択文と判定する。

3.2 実験結果

図 4 は、提案した解答方法を用いた場合の各質問における平均正答率である。横軸は、1 問目の 5 択を 1 から 5 まで、2 問目の 5 択を 6 から 10 というような順番に並べている。縦軸は平均正答率である。

図 4 から、ほとんどの質問の正答率が 90 から 100 % の高い正答率である。最も正答率が低かった場合は、17 問目の 64 % であった。正答率が低くなった原因の多くは、音声認識の精度の問題ではなく、システム側のトラブルが生じたためと考えられる。

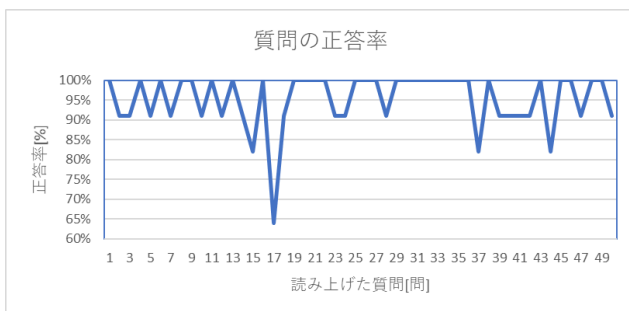


図 3 正答率

図 4 は、質問の正答率と平均類似度の相関を示している。図 4 より、質問の正答率は平均類似度との相関関係は低いことがわかる。このとき、平均類似度は、読み上げ音声を音声認識によるテキスト化されたデータが真の文との差異を表すものである。この結果より、音声認識精度がある程度低くても、正答率が高くなることからわかる。

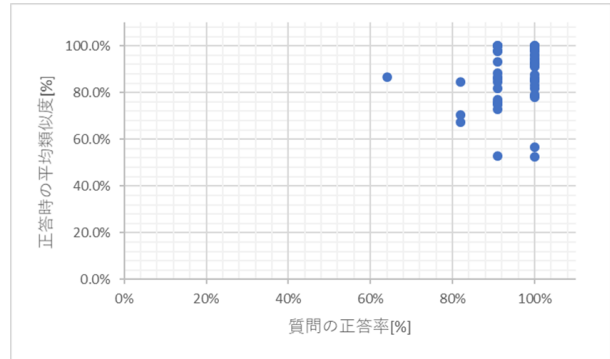


図 4 質問の正答率と類似度の相関

また、多くの被験者が同じ質問で類似度が低い結果であった文として、「素行が悪く、刑期が延びた」、「橋本くんが喋った」などがあつた。このため、音声認識精度が低くなるような文があるものと考えられる。

4. おわりに

本稿では、多肢選択問題の音声認識を用いた解答方法について検討した。提示された 5 択選択肢の文を読み上げ、レーベンシュタイン距離を用いた判定方法で選択した文を推定する方法である。評価実験では、良好な結果を得ている。今後の課題は、医療面接演習システムに提案した音声読み上げによる 5 択選択肢の回答法を適用することなどである。

参考文献

- [1] 南郷栄秀. 医療面接の基本と客観的臨床能力試験(OSCE). <http://cont. o. oo7. jp/40 2/p395-400. pdf> (参照 2024/02/01)
- [2] 畑田稔 and 遠藤裕英, 1997. 日本語 OCR における英字・カタカナのスペル誤り訂正法. 情報処理学会論文誌, 38(7), pp. 1317-1327.