

クイズ AI システムを用いた ヒューマンロボットインタラクションのための検討

坪倉和哉¹ 久保谷空史¹ 小林邦和¹

¹ 愛知県立大学大学院情報科学研究科

{im212008,im211004}@cis.aichi-pu.ac.jp, kobayashi@ist.aichi-pu.ac.jp

概要

本研究では、「AI 王～クイズ AI 日本一決定戦～」のシステムをヒューマンロボットインタラクションに適用するための検討を行った。具体的には、クイズ問題に対して、クイズ難易度と早押し可能タイミングのアノテーションを行い、分析を行った。クイズ難易度の推定では 65% の正解率を得た。また、平均的にクイズの問題文の 3/4 が読み上げられた時点で早押し可能であることが明らかになった。また、ロボットにクイズシステムの実装を行うための検討を行った。

1 はじめに

近年、スマートスピーカーやスマートフォンのバーチャルアシスタントなどの質問応答システムが身近なものとなっている。そのため、質問応答に関する研究が行われている。日本語を対象とした質問応答の研究として「AI 王～クイズ AI 日本一決定戦～」¹⁾ (以下、「AI 王」と呼ぶ) が挙げられる。AI 王は、日本語を対象とした質問応答研究を促進させることを目的とし、クイズ問題を題材としたコンペティションである。AI 王では、クイズの問題文からクイズの解答を返すシステムの性能を競う。コンペティションでは、クイズ問題の全文を入力としているため、クイズ番組で見られるような、いわゆる「早押し」に関する問題には対応していない。また、システムがクイズを解答するだけでなく、クイズを通じた人とのインタラクションも考えられる。例えば、エンターテインメント分野や教育分野、介護分野などにおいて、ロボットやバーチャルエージェントとクイズを行うことで、ユーザーを楽しませたり、教育的な効果をもたらすことができると考える。

このようなインタラクションを行うためには、

1) <https://sites.google.com/view/project-aiocompetition2>

クイズを出題したり、解いたりするだけではなく、ユーザの知識レベルに合わせて出題するクイズの難易度を変更したり、難易度に応じて「この問題は難しいね」などと発話することでインタラクションの質を高めることが可能であると考えられる。

そこで本研究では、AI 王のデータセットとベースラインモデルを用いて、コンペティションでは対象とされていない、人と一緒にクイズを行うロボットシステムを構築するための検討を行う。具体的には、以下の 3 点について検討を行った。

- クイズの難易度推定
- 早押し可能なタイミングの分析
- AI 王のベースラインモデルをロボットに適用し、人とロボットが一緒にクイズを行うシステムの検討

2 章では AI 王のデータセットの一部に対して行った難易度と早押し可能タイミングのアノテーションについて述べる。3 章では、アノテーションデータを分析した結果をもとに考察する。4 章では AI 王のベースラインモデルをコミュニケーションロボットに適用するための検討を行う。最後に 5 章で本稿をまとめる。

2 クイズデータのアノテーション

本研究では、クイズ問題の分析を行うため、AI 王第 2 回コンペティションで公開されている「第 2 回コンペティション 開発データ v1.0」を用いる。このデータセットに収録されている 1,000 問のクイズ問題からランダムで抽出した 100 問に対してアノテーションを行った。アノテータは以下の 2 つのラベルのアノテーションを行った。なお、アノテータには問題の ID (qid), 問題文 (question), 正解のリスト (answers) が与えられる。

- 難易度 (1～5 の整数値)
- 早押し可能タイミング

難易度は、クイズ問題の難しさを表す指標である。クイズの問題文と正解から、問題の難易度を1（易しい）～5（難しい）の整数値でラベリングを行う。

早押し可能タイミングは、仮にクイズの問題文を読み上げられた場合、読み上げ途中のどの時点で答えを推測可能かを表す。AI王ではクイズ問題の全文が入力として与えられるが、クイズ番組のようにクイズ問題を読み上げるようなインタラク션을考えた場合には、問題の読み上げ途中で解答することも可能となる。アノテータがクイズ問題の答えを知らない場合、早押し可能タイミングのアノテーションは困難であるから、アノテータにはクイズの問題とクイズの答えを提示し、クイズの答えを知識として知っているとした場合、早押し解答可能だと思う時点でラベルを付与してもらった。

アノテータは、理系大学院生（情報系専攻）3名（男性）である。アノテータの属性を理系大学院生（情報系専攻、男性）としたのは、属性の違いによる難易度の感じ方に対する影響を抑えるためである。クイズを題材として人とロボットがインタラク션을行う際には、ユーザの属性を考慮した出題問題の選択などにより、インタラクシオンの質を高めることが可能になると考えるが、ユーザの属性とクイズ問題の難易度の感じ方の関係性については本研究では対象とせず、今後の検討事項としたい。

3 分析と考察

本章では前章のアノテーション結果の分析と考察を行う。

3.1 クイズ難易度

人とロボットがクイズを題材としてインタラクシオンを行うことを考える上で、クイズの難易度を推定することはインタラクシオンの幅を広げることができる。例えば、インタラクシオン相手が子どもであれば、易しい問題を提示し、クイズが得意なユーザであれば難しい問題を提示するといったことが可能となる。そのため、クイズ難易度の推定を試みる。

3.1.1 アノテータによる難易度の違い

3名のアノテータによるクイズ難易度ラベルのアノテーションの結果を表1に示す。3名のアノテータ間で全アノテータが同一ラベルを付与した割合は

難易度	1(簡単)	2	3	4	5(難しい)
アノテータ1	25	13	13	18	31
アノテータ2	22	22	18	17	21
アノテータ3	18	26	18	22	16

22%、少なくとも2名が同一ラベルを付与した割合は75%であった。クイズ問題毎に3名のアノテータの難易度ラベルの平均値を求め、その値をクイズの難易度とした。全クイズの難易度の平均値を求めた結果、平均クイズ難易度は3.0となった。

3.1.2 クイズ問題から特徴量の抽出

クイズの問題文からクイズ難易度との相関の分析(3.1.3)や難易度推定(3.1.4)に用いるための特徴量の抽出を行う。

クイズ問題の難易度推定のために、クイズの問題文から以下の特徴量を抽出した。

- クイズの問題文の長さ
- 問題文中の各品詞の出現数

問題文中の各品詞の出現数は、オープンソースの日本語自然言語処理ライブラリである「GiNZA」²⁾を用いて、クイズの問題文を形態素解析し、品詞ごとに出現頻度をカウントした。なお、品詞のタグセットとして、日本語のUD品詞タグで定義されている17種の品詞タグを用いた[1]。

3.1.3 特徴量と難易度の相関の分析

クイズの難易度と3.1.2節で抽出した特徴量との相関の分析を行う。クイズの難易度は3名によるアノテーションの平均値を採用する。

クイズの難易度と前節で抽出した特徴量との間に弱い相関 ($|r| > 0.20$) のある特徴量と相関係数を表2に示す。表より、3つの特徴量においてクイズ難易度との弱い相関がみられた。問題文に登場する固有名詞を知っていないと、クイズに解答が困難であることから、固有名詞とクイズの難易度との間に弱い相関があったと考えられる。設置詞は、格助詞・副助詞・係助詞である。語と語の意味関係を示したり、意味を添えたりするため、クイズ問題文中に設置詞の数が多くなると、問題文の複雑度が上がり、難易度も高まると考えられる。同様に、問題文が長くなるほど問題文の複雑度が上がるため、難易度も上がると考えられる。

2) <https://github.com/megagonlabs/ginza>

表 2 クイズ難易度ラベルと特徴量との相関係数

(r > 0.20)	
特徴量	相関係数
固有名詞	0.265
接置詞	0.235
問題文の長さ	0.235

表 3 Optuna で最適化した NN のハイパーパラメータ

ハイパーパラメータ	取り得る値
中間層のユニット数	10~500
ドロップアウト率	0~1
中間層の活性化関数	relu / sigmoid
最適化アルゴリズム	sgd / adam / rmsprop

3.1.4 クイズ難易度の推定

3.1.2 節で抽出した特徴量を用いてクイズの難易度の推定を試みる。17 個の品詞タグセットうち、「等位接続詞」と「その他」はデータセット中出现しなかったため、「等位接続詞」と「その他」を除く 15 個の品詞の出現頻度とクイズ問題文の長さを合わせた、合計 16 個の特徴量をクイズの難易度の推定に用いる。

クイズの難易度は 1~5 の範囲の値であるが、ここでは単純化のため、3 名のアノテーションの平均値が 3 以下のクイズを「易しい」、3 より大きいクイズを「難しい」とし、難易度（易しい/難しい）の 2 クラス分類問題として学習・評価を行う。

難易度推定には中間層 2 層からなる NN (Neural Network) を用いる。100 問のクイズ問題のうち、80 問を学習に、残りの 20 問を評価に用いる。各特徴量の標準化を行い、NN の入力とした。学習は 10epoch 行った。

なお、NN のハイパーパラメータ（中間層のユニット数、ドロップアウト率、中間層の活性化関数、最適化アルゴリズム）は、オープンソースのハイパーパラメータ自動最適化フレームワークである Optuna[2] を用いて最適化を行った。最適化したハイパーパラメータを表 3 に示す。

評価の結果、難易度の 2 クラス分類の正解率は 65% と低い精度となった。推定に用いた特徴量は品詞の出現頻度と問題文の長さを合わせた 16 個の特徴量であり、クイズの難易度推定には十分な情報が含まれていないことが確認された。

3.2 早押し可能タイミング

クイズ番組のようにクイズ問題が読み上げられ、クイズに解答するインタラクションを考えた場合、いわゆる「早押し」が可能となるため、早押し可能

なタイミングの分析を行った。

3 名のアノテータがクイズ問題文中に早押し可能タイミングの付与を行った。まず、3 名のアノテータによる問題文の長さに対する早押し可能位置の平均を求めた。クイズの問題文の長さを 100% とした場合、早押し可能位置の平均は 74.7% であり、問題文の 3/4 が読み上げられた時点でおおよそ半数の問題が回答可能であることが示唆された。

クイズの問題文の長さに対する早押し可能位置と難易度との相関係数を求めた結果、0.137 であった。そのため、早押し可能タイミングはクイズの難易度には依存しないことが示唆された。

4 ロボットへの実装

本章では、人とロボットが一緒にクイズを行うインタラクションの実装を検討する。人とロボットが一緒にクイズを行うことを考えた場合、クイズの司会者と解答者の 2 つの役割がある。人とロボットにこの役割をどう割り当てるかによって、いくつかのインタラクションの形式が考えられるが、ここでは、人が司会者としてクイズ問題を出題し、人とロボットがクイズの解答を行うことを考える。本研究では、対話ロボットとして、ソフトバンクロボティクス社の「NAO」を用いる。

4.1 クイズインタラクションの流れ

クイズインタラクションの流れを図 1 に示す。まず、司会者がクイズ問題を読み上げる。解答者（ロボット）は司会者が読み上げたクイズ問題を音声認識する。音声認識には Google Cloud Speech-to-Text³⁾ を用いる。音声認識結果は、クイズ解答推定システムに送られる。ここでは、クイズ解答推定システムとして、AI 王のベースラインシステムを用いる。このベースラインシステムは、文献 [3] をベースに AI 王コンペティションで扱う日本語クイズ問題で利用できるように改修されたシステムである。

続いて、解答者（ロボット）はクイズ解答推定システムにより予測されたクイズ解答を読み上げる。読み上げられた解答は司会者により正誤判定が行われ、「正解です」や「不正解です」などとフィードバックする。ロボットは司会者の正誤判定を音声認識し、正解の場合は喜びを、不正解の場合は悲しみを表出する発話や動きを行うことで、正誤判定に対してフィードバックする。その後は、クイズの出題

3) <https://cloud.google.com/speech-to-text>

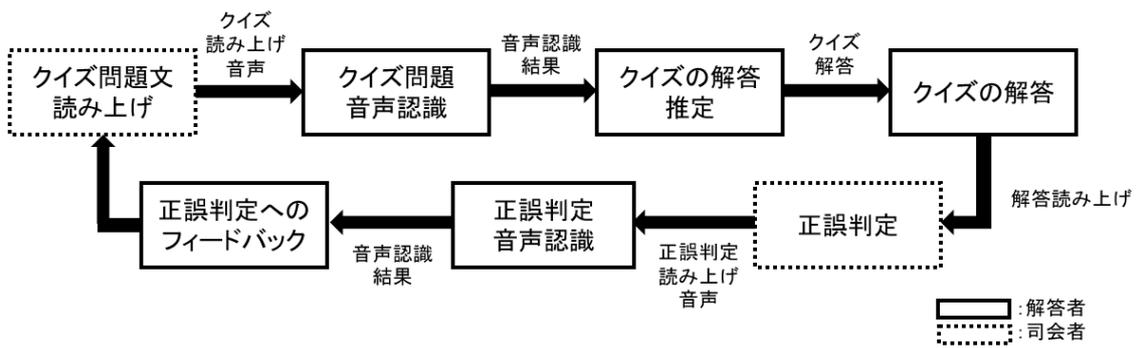


図1 クイズインタラクションの流れ

回数の上限まで上述の流れを繰り返す。

図1では考慮されていないが、通常クイズの解答者は複数である。そのため、クイズの解答段階では、クイズの解答権の付与を行う必要がある。本システムでは、クイズ解答前にロボットの名前(NAO)を呼ぶことで解答権の付与を行う。

5 おわりに

本研究では、「AI王～クイズAI日本一決定戦～」のシステムをヒューマンロボットインタラクションに適用するための検討を行った。具体的には、クイズ問題に対して、クイズ難易度と早押し可能タイミングのアノテーションを行い、分析を行った。クイズ難易度の推定では65%の正解率を得た。また、平均的にクイズの問題文の3/4が読み上げられた時点で早押し可能であることが明らかになった。また、ロボットにクイズシステムを実装するための検討を行った。

今後の課題として、本研究でクイズ難易度の推定に用いた特徴量16個では十分に難易度を推定できないことが明らかとなったため、単語や文章の意味表現ベクトルを用いるなど、特徴量を追加し、クイズ難易度推定の精度向上を目指す。また、多様な属性のアノータにクイズ難易度をアノテーションしてもらうことで、ユーザの属性とクイズ難易度の関係性を分析し、ユーザの属性を考慮した出題問題の選択を可能にすることを目指す。本研究では早押し可能タイミングの推定は行っていないため、今後データを増やして早押し可能タイミングの推定を行う。さらに、本研究ではロボットへの実装の検討を行ったが、実装と人を含めたインタラクション実験は行っていないため、今後人とロボットと一緒にクイズを解く実験を行い、インタラクションの評価をする予定である。

参考文献

- [1] 浅原正幸, 金山博, 宮尾祐介, 田中貴秋, 大村舞, 村脇有吾, 松本裕治. Universal dependencies 日本語コーパス. 自然言語処理, Vol. 26, No. 1, pp. 3–36, 2019.
- [2] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. *CoRR*, Vol. abs/1907.10902, , 2019.
- [3] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wentaoh Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6769–6781, Online, November 2020. Association for Computational Linguistics.