

クイズビジネスにおける作問作業支援

折原良平¹ 鶴崎修功² 森岡靖太¹ 島田克行¹ 狭間智恵¹ 市川尚志¹

¹ キオクシア株式会社 ² 東京大学大学院

{ryohei.oriyara,yasuhiro.morioka,katsuyuki1.shimada,

chie.hazama,takashi2.ichikawa}@kioxia.com

htsurusaki1929@gmail.com

概要

クイズには、それが使われる場面に応じて、問題文に盛り込むべき情報の質や量が異なるという特徴があり、近年様々な応用で成功を収めている「大規模コーパスで事前学習した言語モデルのファインチューニング」というアプローチでは、訓練データの調達が困難であることが予想される。本稿では、QuizKnock のサービスである「朝 Knock」風のクイズを作問することを題材に、クイズビジネスにおける作問作業に機械学習を適用する上での課題と、その解決法について考察する。

1 はじめに

テレビ番組表サイト「テレビ王国」[1]で「クイズ」と検索すると134件がヒットする¹⁾ことからわかるように、クイズは、様々な変遷を遂げつつ[2]も、現代日本のマスメディアにおける主要コンテンツとしての地位を確立している[3]。その活躍の場がテレビにとどまらず、オンライン・オフラインの各種イベントにも広がってきていることと呼応して、クイズ関連のイベント催行やコンテンツ提供を基幹ビジネスとする企業も複数出現している[4]。

クイズを活用できる場面は数多い。クイズは1940年代にはすでに日本でもラジオコンテンツとしてのジャンルを確立していたが、その普及の背景には、クイズの持つ教育的効果の影響が大きかった[2]。結婚式や組織の懇親会など、肩ひじ張らない集会における余興として多用されるのはよく経験するところである。賞品付きのクイズがイベントへの集客促進を狙った企画として使われることも多い。社会課題としての認知症への興味が増す中で、高齢者の認知機能維持向上を目的とする心理療法も注目されるようになってきている。その中には、高齢者に過去

の思い出を想起するように働きかける回想法や、参加者の用意した素材について会話を促す共想法があり[5]、適切に選ばれたクイズが役に立つであろうことは想像に難くない。実際、精神科病棟でのクイズ大会が統合失調症患者の状態を改善したとの報告を行った医師が、クイズを、認知症まで至らない軽度の認知障害という段階での予防に活用することに言及している[6]。

一方、様々な場面にクイズが進出していくことは、それぞれの場面に合った様々な種類の作問能力が求められることを意味する。結婚式や、高齢者向け心理療法において、クイズの主題が適切でなければならぬのは自明であるが、ある程度主題の選択に自由があるケースでも、出題されるクイズが満たすべき条件が明示的・暗示的に存在し、それに沿った作問がなされる必要がある。例えば、イベントに用いられるクイズにおいては、イベント内でクイズの果たすべき役割に応じて難易度を調節²⁾する必要がある。また、クイズ出題側の意図として、問題への解答とそれによる解答者の選別というクイズ本来の機能を越えた効果を狙っている場合もある。QuizKnock のサービスである「朝 Knock」[7]は、クイズに接することによりユーザが享受するエンタテインメントとして、単に正答する喜びにとどまらず、時事問題に対する興味を喚起し、題材となっているニュースに関連する学びを与えることを狙っている。図1は、朝 Knock に掲載されたクイズの例である。これを見たユーザは、正答を選ぼうとするほかに、平井氏の来歴を調べたり、「人にやさしいデジタル化」の詳細を知ろうとデジタル庁のホームページにアクセスするかもしれない。すなわち、インターネットビジネスのエコシステムにおいて波及力を持つ可能性がある。一方、これが図2であった

2) 例えば、余興においては多くの人に参加できるよう簡単にする、競争率の高い賞品を争う場合には難しくする、など。

1) 2021年12月8日に検索実施

らどうだろうか。ユーザが潜在的に得られる喜びの広がりや、図 1 の場合に比べて限定的と言わざるを得ない。

多彩な作問能力が求められる需要側の事情に対し、作問作業自体はクイズ作家頼みの高コスト構造となっているのが現状である。クイズにおいては、ビジネス化以前にアマチュアクイズ大会の開催においても問題の収集に悩んできたという歴史 [2] があり、需要の増大が供給のひっ迫に拍車をかけていると考えられる。

本稿では、「多様な用途に対して安価に問題を提供する」というクイズビジネスニーズに対しての現実的なアプローチを提案する。

1 日付で、新たな中央省庁が発足しました。平井卓也氏が大臣を務め、「誰一人取り残さない、人に優しい〇〇化」を掲げる、この省庁は何でしょう。

1. スマート庁
2. デジタル庁
3. IT 庁

図 1 朝 Knock 風クイズの例 [8]

1 日付で発足した、新たな中央省庁は何でしょう。

1. …

図 2 朝 Knock 風でないクイズの例

2 関連研究

クイズを解答するシステムの研究開発 [9] や、そのための技術向上を図る取り組み [3] はすでに多数行われており、ここでは取り上げない。

前述の通り、クイズの教育的効果が古くから認識されていたことから、教育の文脈で学生の理解度を可視化するツールとしてのクイズを生成する研究が行われてきた [10]。

奥原ら [11] は、近年の教育で求められている教科・科目間の関連を図った横断的な学習において有効と考えられる、幅広い関連情報を含み全体像をとらえさせるような問題を人手で生成することが高コストであることに着目し、Linked Data として与えられた知識源から問題文と等価な Question Graph を生成する方法を提案した。問題の視点の幅広さ(俯瞰度)を、Question Graph に含まれるべき語句の正答からの意味的非類似度を使って表現し、俯瞰度が高くなるような工夫がなされている。

Willis ら [12] は、パラグラフを入力すると正答

の候補を抽出しそれを問う問題文を生成する end-to-end システムを提案した。システムはパラグラフから正答候補を抽出するモデルと、正答とパラグラフから問題文を生成するモデルとからなる。[12] では、SQuAD データセット [13] を使って前者を学習し、後者としては学習済みの QG-Net[14] を用いた。教育の専門家から収集したデータによる検証を通して、システムがベースラインより教育者の選択に近いことを示した。

一方、クイズには典型的な文としての形式が存在することから、これを分析することは自動作問に対し重要な示唆を与える。伊沢 [2] は、早押しクイズの文型を 25 通りに分類し、その分析を通して、問題が備えるべき条件と、クイズプレイヤーの取りうる戦略について考察した。橋元 [15] は、早押しクイズの中でも「パラレル問題」と呼ばれる形式に着目し、エキスパートプレイヤーが問題文の途中でも正答できる理由を考察し、その実装方法を提案した。この中で、未知の問題を解く能力と、問題を作成する能力は、本質的に同じであると述べられている。

深層学習に基づく言語生成が広く行われるようになったことから、その一分野としての質問生成も勃興しつつある。上で述べた QG-Net もその一つである。「朝 Knock」が扱うことの多い時事問題については、Lelks らの研究 [16] がある。ニュース記事がどれほど人々に理解されているかを計測するためのツールとしてクイズを用いる、という文脈におけるクイズの自動生成を扱っており、記事を入力すると問題文と正答、および外れ選択肢を生成する。データセットとしては 5,000 記事から人手で作成した 20,000 問のクイズを用いている。アプローチとしては、要約タスク向けのモデルをクイズデータセットでファインチューニングするというもので、システムの出力に対する軽微な編集が求められるものの、ニュースに対する理解度の計測という目的を果たすことができることを示した。この中で、SQuAD データセットは読解力テストのためのデータセットであって、クイズのためのデータセットとしては必ずしも適切でないと述べられている。

3 学習済み日本語言語モデルのファインチューニング

上述の [16] のように、大規模コーパスから学習された言語モデルをクイズデータセットでファインチューニングしてクイズ生成に用いる、というアプローチは、日本語については園部による開発例 [17]

がある。これは、Wikipedia の日本語ダンプデータなど約 100GB のデータから学習された日本語 T5 モデル [18] をベースにしている。ファインチューニングの詳細は明らかにされていない。

我々は、園部が mC4 の日本語コーパスなど約 890GB のデータから学習させた日本語 T5 モデル [19] をベースに、「朝 Knock」を手本にニュース記事から独自に作成した(記事, 正答, 問題文)という 3 つ組のデータセット約 220 件によりファインチューニングする³⁾ことを試みた。データの記事と正答を与えて問題文を出力するタスクにおいて、学習に用いなかったテストデータ 17 件に対する自動評価指標の値、および、ベースラインとして、ベースの言語モデルを [18] にした場合⁴⁾と、[17] をファインチューニングなしにそのまま用いた場合の値を表 1 に示す。表中の「相対長」とは、生成された問題文の文字列長を ground truth のそれで割った値である。

表 1 自動評価結果

	ROUG E-1	ROUG E-2	ROUG E-L	ROUG E-BE	BLEU	相対長
[19]+FT	0.620	0.410	0.548	0.302	27.5	0.946
[18]+FT	0.615	0.396	0.548	0.235	25.9	0.999
[17]	0.354	0.129	0.288	0.120	3.6	0.387

生成された問題文に対し、問題文として文法的に成立しているか、さらに正答を問う意味になっているかを著者が主観的に評価し、テストデータ全体に対する割合を計算した結果を表 2 に示す。

表 2 主観評価結果

	文法 OK	意味 OK
[19]+FT	57.6%	36.4%
[18]+FT	72.7%	24.2%
[17]	93.4%	69.7%

[17] に比べて独自データセットでファインチューニングしたモデルの自動評価スコアが良いのは、テストデータに近い訓練データで学習しているのが当然の結果である。[17] は相対長が短い傾向があり、それが他の自動評価指標に影響している。一方、このタスクを、正答を問う問題文を生成する、というタスクであると考え、表 2 からわかるように、[17] の性能は極めて良好である。この結果から、[17] のファインチューニングには、正答に対応する短く簡明な質問文が使われており、それが相対長に表れていると考えられる。

3) 本モデルを「[19]+FT」と呼ぶ。

4) 本モデルを「[18]+FT」と呼ぶ。

また、[19]+FT と [18]+FT の性能の比較から、ベースとなる言語モデルの訓練データの規模は、問題文生成の主観評価・自動評価性能に正の相関がある。

[19]+FT の出力結果のうち、正答を問う意味になっているものの例を表 3、4 に示す。

表 3 ファインチューニング結果の例: 成功

記事	任天堂は、Nintendo Switch 用シミュレーション「あつまれどうぶつ森」において、11月5日配信予定の無料アップデート Ver.2.0 にて「ラジオ体操」を実装する。(約 100 文字中略) ボタン操作はもちろろん、Joy-Con を握って実際に体を動かしてプレイヤーもどうぶつたちと一緒に「ラジオ体操」を体験できる。[20]	2020 年 3 月に新型コロナウイルス感染症に伴う肺炎のため、この世を去った志村けんさんの半生を描いたドラマ「志村けんとドリフの大爆笑物語」(フジテレビ系)が、2021 年 12 月に放送されることが決定。フジテレビ系ドラマ初主演となる山田裕貴が志村さんを演じる。(約 1700 文字中略) ドリフタワズの物語をぜひ家族そろって笑顔でご覧ください。[21]
正答	あつまれどうぶつ森	山田裕貴
生成された問題文	任天堂が 2021 年 11 月 5 日配信予定の無料アップデート ver.2.0 にて「ラジオ体操」を実装すると発表した、nintendo switch 用シミュレーションソフトは何でしょう?	志村けんさんの半生を描いたドラマ「志村けんとドリフの大爆笑物語」が、2020 年 12 月に放送されるのでしょうか? 志村さんを演じるのは誰でしょう?

これらのうち、表 3 に挙げたものは、クイズ作家の観点から「朝 Knock」風であると評価されたもの、4 に挙げたものはそうでないとされたものである。両者とも指定された正答を問う問題文にはなっているのだが、前者では、ニュース記事に含まれる情報が問題文にもバランスよく配され、問題文を読んだだけで読者にとっての学びがある内容となっているのに対し、後者は「正答を問う」という最低限の機能を果たしているに過ぎず、第 1 章で述べた「朝 Knock」に込められた陰の狙いを実現できていない。

こうした課題は、機械学習の観点からは、目的に合ったファインチューニング用のデータセットを拡充することで解決されることが期待される。しかし、クイズが多様な場面で用いられる可能性を考えると、すべての応用において十分な規模のデータセットが用意できることは考えにくい。また、新規な応用においては常にデータが不足する状況が予想される。一方、各応用の「陰の狙い」は仕様化できる場合があり、その実現は必ずしも機械学習に基づく必要はないと考えられる。

そこで、本稿では、特定の応用に関して機械学習モデルの出力が不十分となる場合のケーススタディとして、「朝 Knock」風でない [19]+FT モデルの出力

表4 ファインチューニング結果の例: 失敗

記事	卓球金メダルで「水谷隼カレー」人気。通販サイトで品切れ... メーカー「リオ以上の反響になるのでは」。東京五輪・卓球混合ダブルス（決勝、2021年7月26日）で、金メダルを獲得した水谷隼・伊藤美誠ペア。日本卓球史上初の快挙に、選手ゆかりの企業は「嬉しい悲鳴」をあげている。水谷選手をパッケージに起用したレトルトカレーは、金メダル獲得を受け通販サイトで「品切れ」状態に。（約700文字中略）広報担当者は「オリンピックは『これで最後』と本人も考えているようですので、悔いのないように頑張ってください」と語った。[22]	16日、米フロリダ州ケープカナベラル空軍基地から無人探査機「ルーシー」を載せて打ち上げられる「アトラスV」ロケット(NASA提供)(EPA時事)【ワシントン時事】米航空宇宙局(NASA)は16日早朝(日本時間同日夕)、木星軌道上の「トロヤ群」と呼ばれる小惑星群を目指す無人探査機「ルーシー」を打ち上げた。(約150文字中略)27~33年に木星トロヤ群の小惑星に接近する。12年間で探査する小惑星は計八つで、太陽を周回する天体を目指すミッションとしては過去最多になる。[23]
正答	レトルトカレー	ルーシー
生成された問題文	7月26日、食肉加工のフリーデンが品切れとなった商品は何でしょう?	木星軌道上の小惑星群を目指す無人探査機を何というでしょう?

を後処理で改善する方法を提案する。

4 後処理

表3、4の比較からわかるのは、後者ではニュース記事に含まれる情報が問題文に十分に活かせていないことである。これを緩和する方法として、2通りの後処理を考案した。なお、開発にあたっては、日本語NLPライブラリGiNZA[24]を用いた。

4.1 正答の連体修飾節の追加

問題文の情報を増やす方法として、正答に関する情報を付加することが考えられる。これは、「朝Knock」の狙いからすれば読者の学びを増やすことであるし、他の目的のクイズにおいてはヒントを与えることとも考えられる。具体的な方法の一つとして、ニュース記事から正答の連体修飾節を抽出し、問題文中で正答に対応する語句の前に置くことが考えられる。たとえば、表4左側の例では、正答「レトルトカレー」の連体修飾節として「水谷をパッケージに起用した」が抽出される。その結果、図3に示す問題文が生成される。

4.2 導入文の追加

正答以外の語句に関する情報の付加も、読者の学びを増やしたりヒントを与える効果を持ちうる。具

7月26日、食肉加工のフリーデンが品切れとなった、水谷をパッケージに起用した商品は何でしょう?

図3 連体修飾節の追加例

体的な方法の一つは、正答に関連する固有表現に関する情報を表す文を、問題文への導入文として前置することである。導入文の生成は本質的には要約であり、既存の要約生成手法を用いることができる。

本研究では、関連固有表現が与えられるという前提で、石原らによるニュース記事要約[25]で用いられた抽出型要約の手法を参考に、関連固有表現を含むニュース記事から導入文を生成する方法を実装した。[25]は、MMRによる文選択とTF-IDFを用いた文圧縮からなる。本研究では、より多くの情報を付加するという観点から、文選択は関連固有表現を含む文のうち最も長いものを選択する。文圧縮については、選択された文の構文木から評価関数に従って部分木を選択するのは[25]と同様であるが、評価関数としてTF-IDFではなく、関連固有表現との関連度を利用する。

たとえば、表4左側の例では、関連固有表現「水谷」に対して導入文「東京五輪・卓球混合ダブルスで、金メダルを獲得した水谷隼」が抽出される。体言止め文についてのヒューリスティックを適用した結果、図4に示す問題文が生成される。

東京五輪・卓球混合ダブルスで、水谷隼が金メダルを獲得しました。さて、7月26日、食肉加工のフリーデンが品切れとなった、水谷をパッケージに起用した商品は何でしょう?

図4 導入文の追加例

5 おわりに

クイズビジネスにおける多様な自動作問へのニーズと、学習済み言語モデルのファインチューニングに基づくアプローチの現状と課題、および後処理による現実的な改善方法について述べた。課題を解決する方法は本稿で挙げた方法以外にも多数あり、例えばクイズコミュニティの協力を得てデータセットを拡充することや、後処理で改善されたクイズをデータセットに含めることもあり得るであろう。本稿のような研究が、クイズビジネスの発展に自然言語処理技術が貢献する上での一つのヒントとなることを期待する。

謝辞

「朝 Knock」の狙いについて解説頂いた QuizKnock 様に感謝いたします。

参考文献

- [1] テレビ王国. <https://www.tvkingdom.jp/>.
- [2] 伊沢拓司. クイズ思考の解体. 朝日新聞出版, 2021.
- [3] 鈴木潤ほか. ライブコンペティション: 「AI 王へクイズ AI 日本一決定戦〜」. 自然言語処理, Vol. 28, No. 3, pp. 888–894, September 2021.
- [4] 伊沢拓司ほか. ユリイカ 2020 年 7 月号, 特集*クイズの世界. 青土社, 2020.
- [5] 大武美保子ほか. 回想法から見た共想法の考察と連携の可能性. 第 24 回人工知能学会大会論文集, 2010.
- [6] 精神科医・森隆徳が語る医療に生きるクイズ. http://www.qbik.co.jp/contents/mori_001/, 2019.
- [7] 朝 knock. <https://web.quizknock.com/tag/%E6%9C%9DKnock>.
- [8] 朝 knock. <https://web.quizknock.com/daily-sep2-21>, September 2021.
- [9] 金山博, 武田浩一. Watson: クイズ番組に挑戦する質問応答システム. 情報処理, Vol. 52, No. 7, pp. 840–849, 2011.
- [10] Azevedo P., et al. Exploring NLP and Information Extraction to Jointly Address Question Generation and Answering. In **AIAI2020: Artificial Intelligence Applications and Innovations**, pp. 396–407, 2020.
- [11] 奥原史佳ほか. Linked Data を用いた俯瞰的な多肢選択式問題自動生成手法の提案. 情報処理学会論文誌, Vol. 60, No. 10, pp. 1738–1756, 2019.
- [12] Willis A., et al. Key Phrase Extraction for Generating Educational Question-Answer Pairs. In **L@S '19: Proceedings of the Sixth (2019) ACM Conference on Learning**, pp. 1–10, 2019.
- [13] Rajpurkar P., et al. SQuAD: 100,000+ questions for machine comprehension of text. In **EMNLP 2016: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing**, pp. 2383–2392, 2016.
- [14] Wang Z., et al. QG-Net: A data-driven question generation model for educational content. In **L@S '18: Proceedings of the fifth annual ACM conference on learning at scale**, 2018.
- [15] 橋元佐知ほか. 競技クイズ・パラレル問題の基本構造と文型. 言語処理学会 第 27 回年次大会, pp. 1420–1424, 2021.
- [16] Adam D. Lelkes, Vinh Q. Tran, and Cong Yu. Quiz-Style Question Generation for News Stories. In **Proceedings of the Web Conference 2021 (WWW '21)**, 2021.
- [17] 園部 勲. deep-question-generation. <https://github.com/sonoisa/deep-question-generation>.
- [18] 園部 勲. t5-base-japanese. <https://huggingface.co/sonoisa/t5-base-japanese>.
- [19] 園部 勲. t5-base-japanese-mC4-Wikipedia. <https://huggingface.co/sonoisa/t5-base-japanese-mC4-Wikipedia>.
- [20] 「あつまれ どうぶつの森」に“ラジオ体操”が登場! joy-con を使って体を動かすことも可能. <https://game.watch.impress.co.jp/docs/news/1359018.html>, October 2021.
- [21] 志村けんさんの半生をドラマ化 主演・山田裕貴, 脚本・演出は福田雄一「志村けんとドリフの大爆笑物語」. <https://thetv.jp/news/detail/1055394/>, October 2021.
- [22] 卓球金メダルで「水谷隼カレー」人気 通販サイトで品切れ... メーカー「リオ以上の反響になるのでは」. <https://www.j-cast.com/2021/07/27416975.html?p=all>, July 2021.
- [23] 木星トロヤ群へ探査機 太陽系形成過程の解明目指す—N A S A. <https://www.jiji.com/jc/article?k=2021101600442&g=int>, October 2021.
- [24] 松田寛. GiNZA - Universal Dependencies による実用的日本語解析. 自然言語処理, Vol. 27, No. 9, September 2020.
- [25] 石原祥太郎, 澤紀彦. MMR による文選択と TF-IDF による文圧縮を用いたニュース記事要約. 第 35 回人工知能学会大会論文集, 2021.