

問題と模範解答を同時に生成する難易度調整機能付き読解問題自動生成手法

Joint Generation of Questions and Reference Answers for Reading Comprehension with Difficulty-Controllability

後藤照佳*¹, 富川雄斗*¹, 宇都雅輝*¹
 Teruyoshi Goto*¹, Yuto Tomikawa*¹, Masaki Uto*¹
 *¹ 電気通信大学

*¹The University of Electro-Communications
 Email: {goto, tomikawa, uto}@ai.lab.uec.ac.jp

あらまし： 近年，学習者の読解力に合った難易度の読解問題を自動生成する手法が提案されている．そのような難易度調整機能付き読解問題自動生成の最先端手法は，まず読解対象文と所望の難易度値から答えを生成し，次に生成した答えと読解対象文，所望の難易度値から問題を生成する枠組みを2つの独立な深層学習モデルで実現した．しかし，この手法では，先に生成された答えに問題の生成が強く制約され，所望の難易度に対して多様な問題を生成できない問題がある．本研究では，この問題を解決するため，2つの深層学習モデルで構築された既存手法を，T5と呼ばれる深層学習モデルを拡張し一つのモデルに統合することで，問題と答えを同時生成する手法を提案する．

キーワード： 読解問題自動生成，項目反応理論，深層学習，自然言語処理

1 はじめに

読解力養成のためには，学習者にまとまった量の文章を読ませ，様々な読解問題に取り組ませる方法が有効である．また，その際には学習者の読解力にあった読解問題の出題が望ましい．このような背景から，近年，学習者の読解力に合った難易度の読解問題を任意の文章から自動生成できる手法が提案されている．そのような難易度調整機能付き読解問題自動生成の最先端手法 [1] は，答えの生成と問題の生成を次の2段階の手続きで行う．(1) 読解対象文と所望の難易度値をBERT (Bidirectional Encoder Representations from Transformers) と呼ばれる大規模言語モデルに入力して答えを生成する．(2) 読解対象文と難易度値，および答えを組み合わせることでT5 (Text-to-Text Transfer Transformer) と呼ばれる大規模言語モデルに入力し，答えに対応する問題文を生成する．しかし，この手法では答えと問題の生成が異なるモデルで行われるため，答えと問題の組に対して難易度値を対応させることができず，生成される問題と答えの多様性が制限される問題点がある．

この問題を解決するために，本研究では答えと問題の2段階生成を単一の深層学習モデルで実現できる読解問題生成手法を提案する．提案手法は，従来研究で問題生成に用いられていたT5に，BERTによる答え生成手法と同等の機構を追加した手法として設計する．具体的には，T5のEncoderがBERTと同等の機構であることを踏まえ，T5のEncoderで答えを生成しつつ，その答え情報をT5のDecoderに直接引き渡せるようにT5を拡張する．これにより，問題と答えの組み合わせに対して難易度値を対応づけることができ，難易度値制御による問題生成のパターンを既存手法より向上できると期待される．

2 提案手法

提案モデルの概略図を図1に示す．提案手法では，

Encoderへの入力として読解対象文 r と難易度 b を $b, r_{SEP}, r_1, \dots, r_O$ という形に変換した系列データを用いる．ここで， r_{SEP} とは難易度と読解対象文の区切りを表す特殊単語であり， r_o は読解対象文 r の o 番目の単語を表し， O は読解対象文 r の単語数をそれぞれ表している．Encoderはこの入力に対して読解文中の答えの開始位置と終了位置を出力することで答え生成を行う．具体的には， i 番目の入力単語に対応するEncoderの出力を T_i とし，Encoderの出力系列 T を $T = (T_1, T_2, \dots, T_{(O+2)})$ とするとき，学習可能なパラメータ S を用いて，読解対象文の o_s 番目の単語が答えの開始位置である確率を

$$P_{o_s}^{(s)} = \frac{\exp(S \cdot T_{o_s})}{\sum_i \exp(S \cdot T_i)} \quad (1)$$

と計算する．同様に，式(1)の右辺の学習可能なパラメータ S を E に置き換えることで， o_e 番目の単語が答えの終了位置である確率 $P_{o_e}^{(e)}$ を計算する．

次に，Encoderで得られた読解対象文と難易度情報に，答え情報を加味した情報をDecoderに渡すことで問題生成を行う．具体的には，答え位置の情報を反映する行列 A をEncoderの出力系列 T に足し合わせた $T + U \cdot A$ で表されるデータをDecoderに渡す．ここで， $U \in \mathbb{R}^{D \times 2}$ (D はベクトル T_i の次元数) は学習可能なパラメータであり，行列 A は答えの開始位置 o_s と終了位置 o_e に基づいて計算される以下の要素を持つ2行 $(O+2)$ 列の行列である．

$$A_{1j} = \begin{cases} 1 & (o_s \leq j \leq o_e) \\ 0 & (\text{otherwise}) \end{cases} \quad A_{2j} = \begin{cases} 0 & (o_s \leq j \leq o_e) \\ 1 & (\text{otherwise}) \end{cases}$$

訓練時には， o_s, o_e に真の答えの位置を入れて A を計

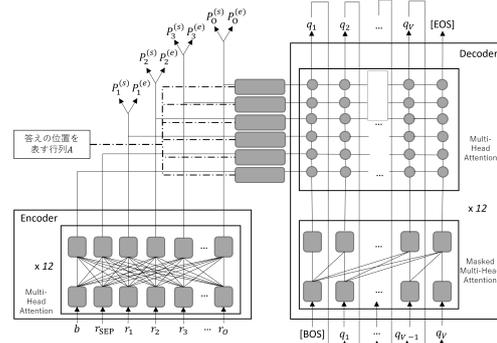


図1 提案手法のモデルの概念図

算し、推論時には、 $P_{o_s}^{(s)} + P_{o_e}^{(e)}$ ($o_s \leq o_e$) が最大となる o_s, o_e を使って \mathbf{A} を計算する。これによって、 \mathbf{T} に 答えの位置の情報を埋め込むことができる。

提案モデルでは、答え生成と問題生成の損失関数を組み合わせた以下の損失関数 L を最小化することで、答え生成機構と問題生成機構を一つの深層学習モデルで同時に訓練することができる。

$$L = \sum_{i=1}^n L_{\text{qg}}^{(i)} + L_{\text{ae}}^{(i)} \quad (2)$$

ここで、 n は訓練データの総数を表す。 $L_{\text{qg}}^{(i)}$ は i 番目の訓練データに対する問題生成の損失関数であり、真の問題文に対する負の対数尤度である。また、 $L_{\text{ae}}^{(i)}$ は i 番目のデータに対する答え生成の損失関数であり、真の答えの開始・終了位置とのクロスエントロピー誤差である。なお、提案モデルの訓練データとなる難易度付きデータセットは、提案手法の元となっている富川ら [1] と同様の手順で作成する。具体的には以下の通りである。(1) 読解対象文とそれに関する問題文と答えからなる SQuAD データセット (問題生成研究のベンチマークデータセット) を用い、そのデータセットの中の各問題を異なる性能を持つ複数の QA (Question Answering) システムに解かせ、正誤反応データを得る。(2) 得られた正誤反応データから問題の難易度 b を推定するために、IRT (項目反応理論: Item Response theory) の数理モデルの一つである Rasch モデルを適用する。ここで、Rasch モデルは、能力値 θ の受験者が難易度 b の問題に正答する確率 p を式 (3) で表すモデルである。

$$p = \frac{1}{1 + \exp(-(\theta - b))} \quad (3)$$

(3) SQuAD データセット内の読解対象文と問題文、答えに、推定された難易度値を組み込むことで、難易度調整可能な問題生成モデルのための訓練データを構築する。

3 提案手法の有効性評価実験

SQuAD データセットを利用して提案手法の性能を評価する。ここでは、まず、SQuAD データセットの訓練データを利用して提案モデルを訓練した。その後、SQuAD データセットのテストデータ中の様々な読解対象文に様々な難易度を指定して提案モデルで問題文と答えの組を生成し、その質を評価する実験を行った。

3.1 難易度ごとの平均正答率

まず、生成される問題の難易度調整が適切に行えているかを評価するために、異なる難易度を指定して生成させた問題群を 2 章のデータセット構築手順 (2) で用いた複数の QA システムに解かせ、その正答率を評価した。難易度ごとの平均正答率を図 2 に示す。横軸が難易

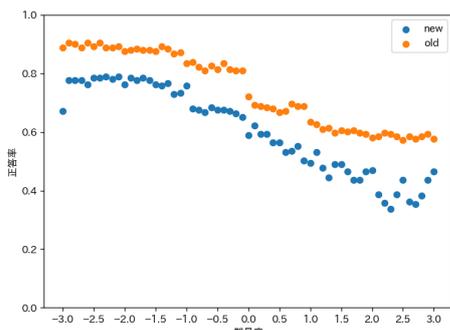


図 2 難易度ごとの正答率

表 1 提案手法と既存手法の問題、答えのパターン数の比較

	問題	答え
既存手法	885	757
提案手法	2464	941

表 2 先頭単語の疑問詞の出現率と 2 単語目のパターン数

	疑問詞出現割合		2 単語目のパターン数	
	提案手法	既存手法	提案手法	既存手法
what	62.89	56.93	110	93
who	11.84	17.28	58	52
when	9.09	7.02	7	3
how	7.65	9.5	14	10
where	3.82	1.54	8	6
why	1.37	1.37	5	5
which	0.54	2.18	23	13
尖度	-1.806	0.788	-	-

度値、縦軸が平均正答率を表しており、青点が提案手法、橙点が既存手法を表している。既存手法同様に難易度が上がるごとに正答率の低下が見られ、難易度調整が機能していることが読み取れる。

3.2 生成された問題・答えの多様性

次に、生成された問題と答えのパターン数を既存手法と提案手法のそれぞれで求めた結果を表 1 に示す。表から問題と答えのいずれも既存手法と比べてパターン数が増えていることがわかる。さらに、問題の多様性を分析するために、生成された問題のうち、先頭単語が what, who, when, how, where, why, which の疑問詞であった割合を表 2 の 2, 3 列目に示す。what 以外の疑問詞に着目すると、提案手法では出現割合が全体として均一に近づいている傾向が読み取れる。このことを確認するために、what を除いた 6 つの疑問詞の出現割合について、分布の一様度を表す尖度を計算した結果を表 2 の 2, 3 列目の最終行に示す。尖度は値が低いほど、分布が一様に近いことを表しており、提案手法が既存手法よりも著しく小さいことから、提案手法では問い方の多様性が増加していると考えられる。なお、what の割合が増加している理由は、what はそれ自体に多様な用法があるためと考えられる。加えて、先頭単語の疑問詞別に 2 単語目のパターン数を計算した結果を表 2 の 4, 5 列目に示す。すべての疑問詞に関して、提案手法では、2 単語目のパターン数が従来手法と同数以上になっていることから、より様々な問い方の問題が生成されるようになった、すなわち、問題の多様性が増加したことが示唆される。

4 まとめ

本研究では、最先端手法の難易度調整機能付き読解問題自動生成手法を拡張し、答えと問題の生成を 1 つの深層学習モデルに統合した読解問題自動生成手法を開発した。

参考文献

- [1] 富川雄斗, 鈴木彩香, 宇都雅輝. 項目反応理論に基づく難易度調整可能な読解問題自動生成手法. 電子情報通信学会論文誌 D, Vol. J107-D, No.02, 2024.