

ニューラル文生成モデルを利用した英文読解問題の自動解法

Solving English reading comprehension problems using neural sentence generation models

加藤 秀大 三輪 誠 佐々木 裕

Hidehiro Kato Makoto Miwa Yutaka Sasaki

豊田工業大学

Toyota Technological Institute

{sd16407, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

1 はじめに

近年, 簡単な質問応答について, 半自動的に作成された大規模なデータを利用することで, 深層ニューラルネットワーク (NN) を用いた手法が高い性能を示している. しかし, より複雑な英文読解問題については, まだその性能は十分とは言えず, その学習データも少ないため, 学習データに依存しない読解問題の解答手法の確立が必要である. 本研究では, 文書に対する質問について複数の選択肢から正しいものを選択する TOEIC の読解問題を対象に, 大量のデータを利用したニューラル文生成モデルの利用可能性を探る. そのためにニューラル文生成モデルを利用した英文読解問題の自動解答手法を提案し, TOEIC 読解問題での正答率の向上を目指す.

2 関連研究

2.1 英文読解問題の自動解法

佐藤ら [1] は TOEIC の Part 7 を対象に, 本文中の解答の鍵となる重要文と各選択肢間の単語間類似度の計算を行うことで解答を選択する手法を提案した. 単語間類似度の計算には, WordNet::Similarity [2] 内の LESK を用いた. Webanhvan [3] から得た TOEIC の練習問題 45

問について評価し, 重要文の決定を人手で行った場合の正答数は 25 問, 単語間類似度で行った場合の正答数は 19 問であった.

2.2 ニューラル文生成モデル

Sutskever ら [4] は NN を利用して学習データペアである入力文と出力文から学習を行い, 学習内容に従った文を入力文から出力する sequence-to-sequence (S2S) モデルを提案した. 再帰型ニューラルネットワーク (Recurrent Neural Network; RNN) の一種である Long Short-Term Memory (LSTM) を利用しており, 過去の入力を保持したまま次の出力を行うことが可能となっている. S2S モデルに文を入力すると, 学習内容と入力内容からモデルが出力すべき単語の確率を LSTM が計算し, 最も確率の高い単語を出力していくことで出力文を形成する. 入力文と出力文に対する大量の学習データペアを用いることで文の内容のモデル化が可能となっている.

3 単語間類似度によるベースライン

佐藤らの行った英文読解問題の自動解法では解答選択が 1 つの類似度の計算結果に依存してしまうため, ベースラインとして単語間類似度の加重和を用いた手法を提案する. 単語間類似度の計算には WordNet [5] 内の 6 つの評価

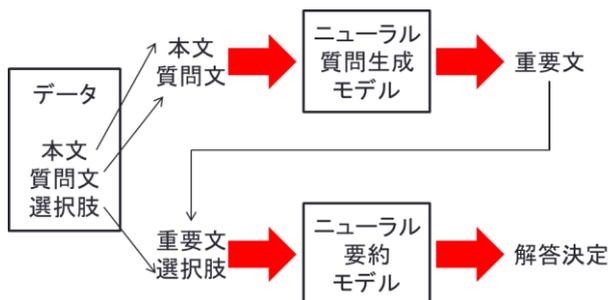


図1 解答選択の一連の流れ

指標を用いる．6つの評価指標で単語間類似度を計算した後，線形二値分類器である Support Vector Machine (SVM) を用いて各類似度の重み付けを行う．重みの学習には LIBLINEAR [6] を利用し，加重和が最大となった選択肢を解答とする．このベースラインの手法では単語同士の関係性しか考慮できないため，単語より大きな構造である文の内容理解が十分にできているとは言い難い．また，学習データが限られているため，学習を十分に行うことも難しい．

4 提案手法

本研究では文の内容のモデル化が可能なニューラル文生成モデルを用いた一連の解答手法を提案する．一連の解答を行うため，質問文から重要文を決定する手法と，文全体の内容を考慮した上で重要文から解答を選択する手法を提案する．一連の解答選択の流れは図1のようになっている．

4.1 ニューラル質問生成モデルによる重要文選択

ニューラル文生成モデルを利用した質問生成モデル [7] を参考にして，質問文から重要文を決定する．学習データとして質問内容が記されている文と質問文のペアを与えることで，入力文に対して質問文を出力するよう質問生成モデルを学習する．このモデルを用いて図2のように質問文を生成し，その生成確率から重要文を決定する．本文中の1文をモデルに入力し，出力文を問題で与えられた質問文で固定す

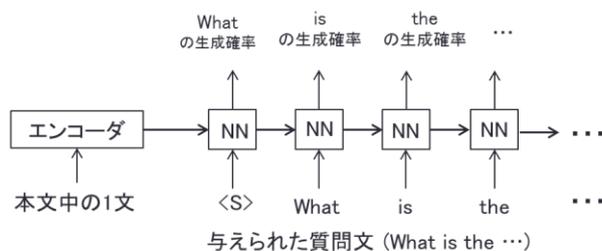


図2 重要文決定モデルの重要文決定の過程

る．モデルの出力文を固定することで，その出力文の各単語が本来出力される生成確率を求めることができる． $\langle S \rangle$ は出力を開始するスタートシンボルである．質問文を Q ，本文中の1文を T_m (m ：本文中の文の数)，質問文中の各単語が出力される確率を $\{q_1, q_2, q_3, \dots, q_n\}$ (n ：質問文中の単語数) とした際のスコアの計算式を式 (1) に示す．

$$\text{score}(Q, T_m) = \prod_{k=1}^n q_k \quad (1)$$

式 (1) より，与えられた質問文中の各単語の生成確率をかけ合わせたものが，入力として与えられた文が重要文となるスコアである．このスコアが最大となる本文中の文を重要文と決定する．また，スコア最大の文とのスコア差が一定以下の文も重要文として用いることで，解答選択に必要な情報が複数文に分かれている問題への対応を目指す．

4.2 ニューラル要約モデルによる解答選択

読解問題の解答を本文の要約であるとみなすことで，ニューラル文生成モデルを利用した要約モデル [8] を参考にして，重要文から解答となる選択肢を決定する．学習データとして要約元の文と要約文のペアを与えることで，入力文に対する要約を出力するよう要約モデルを学習する．このモデルを用いて図3のように解答となる選択肢の文を生成し，その生成確率から選択肢を決定する．4.1 節と同様に，モデルの出力文を与えられた選択肢で固定することで，重要文から選択肢が生成される確率を求めることができる．重要文を T ，選択肢を S_l (l ；選択肢の数)，選択肢中の各単語が生成される

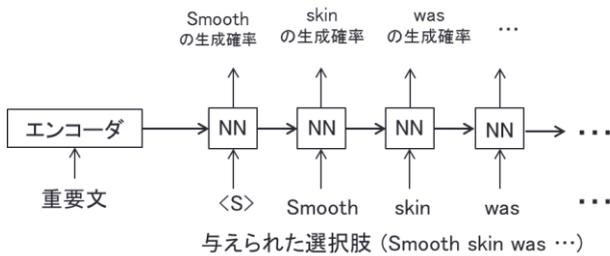


図 3 解答選択モデルの解答決定の過程

確率を $\{s_1, s_2, s_3, \dots, s_j\}$ (j : 選択肢中の単語数) とした際のスコアの計算式を式 (2) に示す.

$$\text{score}(T, S_i) = \prod_{i=1}^j s_i \quad (2)$$

式 (2) より, 与えられた選択肢中の各単語の生成確率をかけ合わせたものが, 重要文に対してその選択肢が解答となるスコアである. 重要文が複数選択された場合, 各重要文で求めたスコアの平均値が解答選択に用いられるスコアとなる. このスコアが最大となる選択肢を解答とする.

5 学習データセットの作成

重要文選択と解答選択の学習を行うためデータセットを作成した.

重要文選択に用いる質問生成モデルは, 入力文に対応する質問文を出力するモデルであるため, 学習データペアを質問応答データセットである NewsQA [9] から取得した. NewsQA はニュース記事と記事の内容を問う質問文から構成されており, 解答となる単語または単語列の本文中の位置も記載されている. ここから, 質問文と解答が含まれる文を取得し, 34,520 個の学習データペアを作成した.

解答選択に用いる要約モデルは入力に対する要約を出力するモデルであるため, 学習データペアを Cable News Network (CNN) データセット [10] から取得した. CNN データセットはニュース記事とニュース記事に対する人手で書かれた要約が各 3 個から 4 個含まれるデータセットである. ニュース記事の各文と要約間で

重要文選択モデルを用いてスコアを計算し, スコア 0.4 を閾値と決定してそれ以上のスコアを出した最大スコアの一文を要約前の文と決定した. この方法で得られた要約前の文と要約文 66,265 個を学習データペアとした.

6 実験設定

英語の読解問題として TOEIC の Part 7 を対象とする. 評価データとして, Webanhvan [3] から図表の読み取りや数字の計算を必要としない TOEIC 用練習問題 91 問を利用した. ベースラインでは 91 問を 45 問と 46 問に分割し, 交差検証を用いて評価を行った. 重要文選択モデルの学習データは NewsQA データセット [9] から質問文と質問内容が記された文のペアを半自動で取得し, 30,000 件を訓練データ, 3,000 件をテストデータとして用いた. 解答選択モデルの学習データは CNN データセット [10] から要約前の文と要約文のペアを半自動で取得し, 60,000 件を訓練データ, 1,000 件をテストデータとして用いた. ニューラル文生成モデルとして TensorFlow ライブラリ [11] の S2S モデルを用いて各モデルの学習を行い, LSTM は 3 層, 隠れ状態の次元数は 256 次元, 単語ベクトルは 256 次元に設定した. 学習アルゴリズムには Adam [12] を用いており, Adam の各パラメータは論文の推奨値である $\alpha = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$ を用いた. バッチサイズは 32 であり, buckets はニューラル質問生成モデルが [(50, 20)], ニューラル要約モデルが [(50, 30)] とした. 質問生成モデルが重要文を複数選ぶ際のスコア差はスコア最大の文のスコアの 1 割以下とした.

7 実験結果と考察

ベースラインの手法との正答率の比較を表 1 に示す. 表 1 より, ベースライン手法は佐藤ら

[1] の手法と同等の正答率を得られる事がわかる。また、提案手法は人手で重要文を決定した際も、質問生成モデルで重要文を複数決定した際も、若干ではあるが従来手法より多くの問題に正答しており、ニューラル文生成モデルの利用が有効である可能性がある。ニューラル文生成モデルは文の内容をモデル化することができるため、既存手法でできなかった単語より大きな構造を考慮することができると考えられる。重要文を1つしか選択しなかった際の正答率より複数選択した際の正答率が表1より2問多く、重要文を複数用いることでより質問の答えとなる内容を選択できた可能性がある。しかし、TOEICの問題では不正解の選択肢も本文中の内容から作成されているため、間違った選択肢を選ばないために重要文選択の精度の向上が必要であると思われる。

8 おわりに

英文読解問題の正答率向上を目的として、ニューラル文生成モデルを用いた重要文選択のための質問生成モデルと解答選択のための要約モデルを提案した。学習データを半自動で作成した結果、ベースラインよりも人手で重要文を決定した際に3問、モデルで重要文を決定した際に5問多く正答でき、ニューラル文生成モデルを用いて文の内容理解を行うことが有効である可能性がわかった。また、重要文を複数選択することで正答が2問増えたことから、文の内容理解において複数文を考慮すること

表1 提案手法の正答率 (% (正解数/問題数))

	モデルで 重要文を決定	人手で 重要文を決定
ベースライン	42.9 (39/91題)	57.1 (52/91題)
提案手法 (重要文1つ)	46.1 (42/91題)	60.4 (55/91題)
提案手法 (重要文複数)	48.3 (44/91題)	60.4 (55/91題)

が必要である可能性を示した。

今後の課題として、誤った選択肢を選ばないための重要文選択の精度の向上が挙げられる。

参考文献

- [1] 佐藤敦美. 英語読解問題の自動解法. 言語処理学会, pp105-108, 2014.
- [2] Ted Pedersen et al., WordNet::Similarity - Measuring the Relatedness of Concepts. In AAAI, pages 1024-1025, 2004.
- [3] Webanhvan, <http://webanhvan.com>.
- [4] Ilya Sutskever et al., Sequence to Sequence Learning with Neural Network. In NIPS, pages 3104-3112, 2014.
- [5] George A. Miller, WordNet: A Lexical Database for English, Communications of the ACM Vol. 38, No. 11, pages 39-41, 1995.
- [6] Rong-En Fan et al., LIBLINEAR: A library for large linear classification. In Machine Learning Research 9, pages 1871-1874, 2008.
- [7] Alexander M. Rush et al., A Neural Attention Model for Abstractive Sentence Summarization. In EMNLP, pages 379-389, 2015.
- [8] Nan Duan et al., Question Generation for Question Answering. In ACL, pages 866-874, 2017.
- [9] Adam Trischler et al., NewsQA: A Machine Comprehension Dataset. In ACL, pages 191-200, 2017.
- [10] DeepMind Q&A Dataset, <https://cs.nyu.edu/~kcho/DMQA/>.
- [11] Martín Abadi et al., TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2015.
- [12] Diederik P et al., Adam: A method for stochastic optimization. In ICLR, 2015.