

[身近になった対話システム]

② 機械読解による自然言語理解

基
専

西田京介 日本電信電話（株）NTT 人間情報研究所

自然言語理解への挑戦

機械読解とは

人工知能（AI）が自然言語（我々が日常的に用いる言葉）で書かれたテキストを読み、その意味を理解できるようにすることは AI および自然言語処理分野における主要な研究目標である。近年では、機械読解（Machine Reading Comprehension ; MRC）が自然言語理解のベンチマークタスクとして大きな注目を集めている。人間が読解力の試験を受けるように、このタスクでは機械読解モデルにテキストが与えられ、そのテキストに関する質問に正しく答えられるかが評価される。ルールベースや知識ベースに基づく質問応答と比べて、機械読解は質問に回答するために必要な知識源としてテキストを用いることが特徴である。

機械読解の研究は、現状のモデルが解くことのできない新たな研究課題となるデータセットを整備し、そのデータセットを用いて技術を磨くことを繰り返すことで発展し続けている。このため、さまざまな問題設定が存在しており、回答の形式には大きく3種類存在する。

- **多肢選択** 回答候補の集合の中から正しい回答を選択する。
- **スパン抽出** テキスト中のスパンを回答として抽出する。すなわち、回答として適切なテキスト中の開始単語と終了単語を選択する。

- **自由記述** 回答は文章中のスパンに限定されず、自由な形式で回答文を生成する。

ここで、機械読解モデルを検索エンジンや対話システムなど実際のサービスに適用するには、スパン抽出式や自由記述式で回答可能なモデルが適している。たとえば、コンタクトセンタにおける対話システムが機械読解の能力を備えると、検索用にマニュアルから FAQ を事前に準備しなくても、システムがマニュアルに書かれた内容を正確に読み解いてピンポイントに回答を発見し、ユーザに返答できるようになる（図-1）。その一方で、AI の自然言語能力を確かめる意味では多肢選択式が正解率を測定しやすいため適している。実際に、人間用に作られた多肢選択式のテスト問題を AI に解かせて評価することも行われている。

人間を上回る回答スコアの達成

機械読解による自然言語理解の発展に関するターニングポイントは、2016 年の 6 月に深層学習を行う

あんしん保険の弁護士費用特約はどのような場合に対象外になりますか？

「事故の相手が不明である場合など、相手の方に法律上の損害賠償請求を行うことができない時」です

回答範囲の抽出 **知識源となるテキスト**

あんしん保険の弁護士費用特約は、自動車事故などにより保険契約者が怪我などをされたり、自らが所有する自動車・家屋などの財物を壊されたりすることによって、相手の方に法律上の損害賠償請求をするために支出された弁護士費用や、弁護士などへの法律相談・書類作成費用などを保険金としてお支払いする特約です。ただし、保険金のお支払い対象となる費用に関しては、当社の同意を得た上で支出された費用に限ります。また**事故の相手が不明である場合など、相手の方に法律上の損害賠償請求を行うことができない時**は、本特約は対象外となりますのでご注意ください。

図-1 機械読解能力を備えた対話システムの例

小特集 Special Feature

ために必要な大規模なデータセットである SQuAD¹⁾ がスタンフォード大から発表されたことであった。

SQuAD は知識源となるテキストから、質問に対して回答となる文字列のスパンを抽出するタスクである (図-2)。平均 140 単語程度の Wikipedia の段落テキストに対して、クラウドソーシングにより大規模に作成された 10 万件を超える質問・回答ペアのデータから構成されるこのデータセットは、広く深層学習・自然言語理解の研究者にインスピレーションを与えた。図-3 に示すようにデータセットの発表当初は人間と AI の回答スコアに大きな開きがあったが、次々と新たなニューラル機械読解モデルが多数発表され、最高スコアが更新された。2018 年 10 月には Google により発表された BERT²⁾ が人間を大きく上回る回答スコアを達成し、「AI の読解力が人間を超えた」と報じられるなど社会にインパクトを与えた。

BERT によるパラダイム・シフト

事前学習とファインチューニング

BERT とは、巨大なニューラルネットワークを、人間により正解情報が与えられていない大量のテキストで学習した汎用言語モデルである。この事前に学習されたモデルを、応用タスクの学習データを用いて適応 (ファインチューニング) させることで、機械読解を含めたさまざまな自然言語処理タスクで高い性能を実現した。

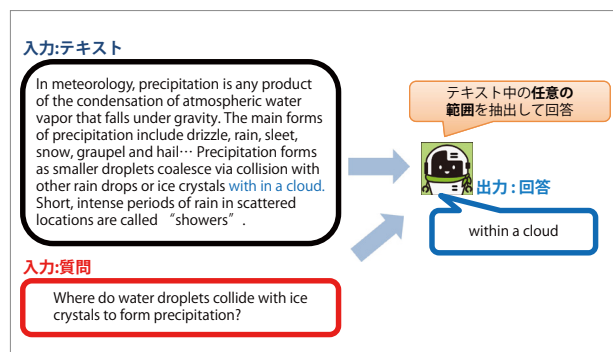


図-2 SQuAD の問題例

ニューラルネットワークを用いた応用タスクの学習には通常大量の目的タスクの学習データが必要であるが、事前学習したモデルは従来に比べて少量のデータによるファインチューニングでも高い性能が得られるようになった。また、従来は応用タスクごとに特化したニューラルネットワークのアーキテクチャを人間が設計していたが、図-4 に示すように BERT は同一の汎用的なアーキテクチャにタスクごとの出力層を 1 つ追加するだけで、特化型のニューラルネットワークを超える性能を実現した。これらの優れた特徴から、汎用の事前学習モデルをファインチューニングにより目的タスクに適応させるアプローチが BERT の登場以降は主流になった。

穴埋め問題から言語をモデリング

前述したように BERT は汎用の「言語モデル」で

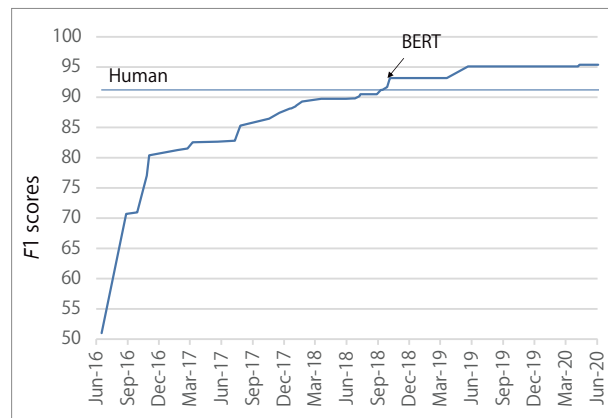


図-3 SQuAD の F1 スコア (回答 (単語列) の部分一致に関する精度) の最高値の変動

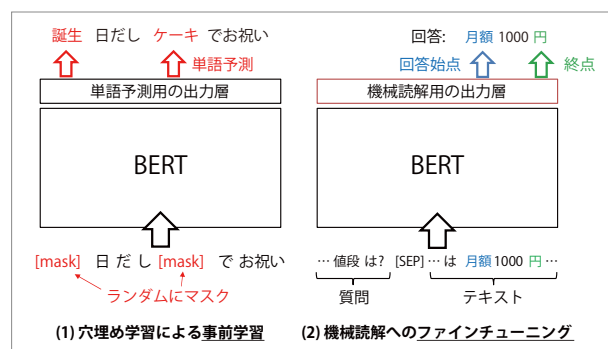


図-4 BERT における穴埋め問題による事前学習とファインチューニングのイメージ

小特集 Special Feature

ある。この言語モデルとは、より自然な、より頻出する文や単語列に高い確率を与えるモデルである。BERTでは、図-4に示すように、テキストの一部をランダムにマスクして、マスクした部分にどの単語が入るのが自然であるかを予測する穴埋め問題（Masked Language Modeling）により言語モデルを学習する。こうした穴埋め問題による事前学習は、人間が正解（教師）情報のアノテーションを行わずにテキストデータを集めるだけで実現できる利点があり、自己教師あり学習と呼ばれる。BERTでは、英語のWikipediaページすべてと小説のコーパスという大量のテキストデータ（約16ギガバイト）を用いて事前学習を行っている。

Transformer とセルフアテンション

BERTはGoogleが2017年6月に発表したTransformer³⁾という構造を採用している。Transformerは、これまで利用されてきた再帰的ニューラルネットワーク（Recurrent Neural Network; RNN）や畳み込みニューラルネットワーク（Convolutional Neural Networks; CNN）をセルフアテンションに置き換えていることが特徴である。

一般的に、アテンションとは $\{h_1, h_2, \dots, h_n\}$ をベクトル集合、 $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$ を各ベクトルに対するアテンションの分布（重み； $\sum_i \alpha_i = 1$ ）としたとき、ベクトル集合の重み付き総和 $\sum_i \alpha_i h_i$ を出力

する単純な計算処理である。ここで、アテンションの重み α_i は、ベクトル h_i とクエリ q を引数とするスコア関数 $S(h_i, q)$ の値に基づいて計算される。従来、アテンションは機械翻訳や画像キャプションニングを始めとした系列から別の系列への変換（sequence-to-sequence）タスクで有効であることが確認されてきたが、Transformerで用いるセルフアテンションでは、クエリおよびベクトル集合のどちらも同じ入力系列を対象とする（図-5）。このセルフアテンションによって、従来問題とされていた文章中の遠く離れた単語間の関係も高精度に理解できるようになり、自然言語理解のレベルは大きく上がった。さらに、Transformerでは1つの層内で複数のアテンション機構（アテンションヘッドと呼ばれる）を用いる（図-5はヘッド数2の場合）ことで、異なる観点で単語間の関係をモデリングすることを可能にしている。

より巨大なモデル、より大量のデータ

SQuADの最高精度を達成したBERTは3.4億個のモデルパラメータ（float値）を持つ。実験的な検証によりTransformerのモデルパラメータ数を多くすると言語モデルの性能が上がっていくことが確認され、2020年5月には1,750億個、2021年1月には1.5兆個のパラメータ数を持つモデルが発表されるなど、言語モデルのパラメータ数は指数関数的に増加している。これに伴って、事前学習に用いるテキストコーパスのサイズも数百ギガバイト～テラバイトのレベルまで増大している。現在のところ、モデルパラメータ数の増加による性能向上の限界はまだ確認されておらず、今後もより巨大な言語モデルの研究が続いていくと予想される。

機械読解研究の展望

SQuADではAIが人間を上回るスコアを達成した。しかし、SQuADは比較的単純な問題設定であるた

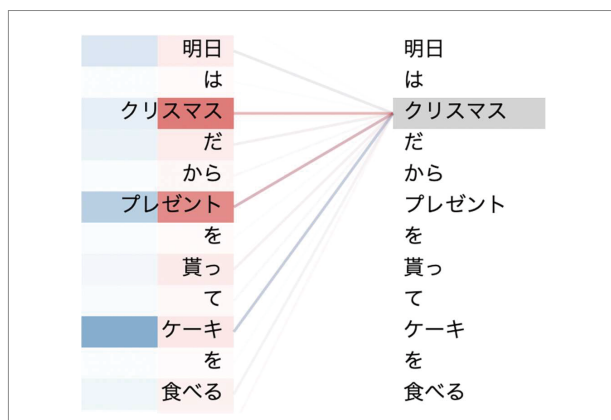


図-5 セルフアテンションを可視化した例

小特集 Special Feature

め、AI が人間を上回る読解力を獲得したとはいまだ言えず、下記に述べるさまざまな課題が残っている。

オープンドメイン質問応答

SQuAD では知識源となる回答に必要な情報を含むテキストが与えられていた。こうした問題設定の場合は、たとえば人名が答えになるような質問においてテキスト中に人名が1つしかなければ、自然言語の意味を理解せずとも回答を見つけることができる。しかし、検索エンジンのように（任意のドメインの）質問のみが与えられる場合は、最初にどのテキストに回答に必要な情報があるかを発見する必要があるため、難易度が大きく上がる。

このようなオープンドメイン質問応答と呼ばれる問題設定では、一般的には情報検索と機械読解の2段階のアプローチが採用される。図-6に示すように、まず検索モデルが知識源となるテキスト集合から質問に関連する複数のテキストに絞り込み、次に読解モデルがこれらのテキストを詳しく読み解き回答を発見する。回答に必要な情報が2つ以上の検索結果テキストに跨って存在する場合は、機械読解において多段の推論が必要になるため理解が難しくなる。

説明可能な AI

現在の機械読解モデルは回答が正解だったとしても、なぜその回答に至ったのかの説明を上手く行うことができない。また、機械読解モデルは一般的にニューラルネットワークで構築されており、たとえ

内部状態を可視化しても人間が回答の理由を解釈することは難しい。

機械読解においては、知識源となるテキストにおいてどの文が回答の根拠となったかを回答と併せて出力することで説明性を向上する取り組みが多くなされている。さらに、より実践的で人間に近いレベルを実現するには、知識源のテキストに書かれていない常識や共通感覚、演算などの知識も含めて説明することが求められる（図-7）。人間がAIの回答を信頼するには、精度の良い回答だけでなく納得のいく説明が必要になるため、説明可能なAIの実現には産業界からも大きな期待が寄せられている。

回答の評価指標

多肢選択式のタスクでは正解率を評価指標として利用可能であるが、その他の形式、特に自由記述式の回答では、機械読解モデルが回答として出力した文字列を、人間が与えた正解の文字列と比べて正しいのかを評価しなければならない。

人間による評価は品質の面では最良であるが、クラウドソーシングや専門家への依頼は高価で長い期間を要してしまうため、日常的なモデル改善のサイクルでは用いることが難しい。そのため、一般的には出力文字列と正解文字列の単語レベルでの重なり具合に基づく自動評価指標を多くの機械読解タスクでは用いている。しかし、機械読解の回答では1単語異なるだけでもまったく異なる意味になる場合

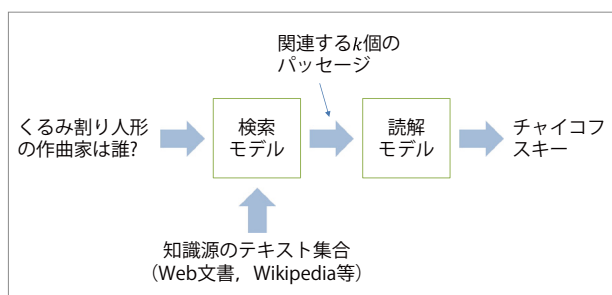


図-6 オープンドメイン質問応答のイメージ

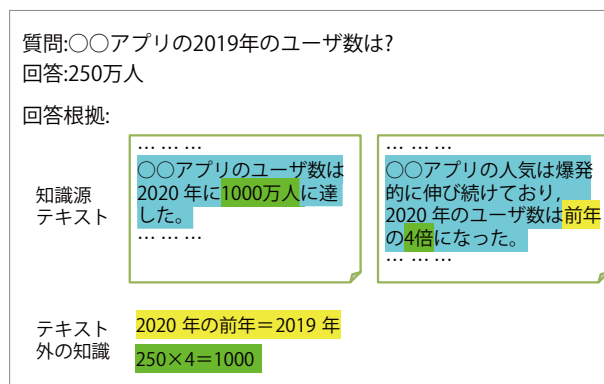


図-7 説明可能な機械読解モデルのイメージ

小特集 Special Feature

があるため、単語の重なりに基づく指標には限界がある。今後、さらなる自然言語理解の発展には、意味的な類似性に基づき、主観評価と高い相関を持つ自動評価指標の確立が必要不可欠である。

視覚と言語の融合理解

これまで取り上げた内容は、すべてテキストに閉じた問題設定であった。しかし、我々が普段扱っている文書はプレーンテキストではない。機械読解の知識源となり得る Web ページ、PDF 文書、プレゼンテーションスライド等には、文字の大きさや色、図や表、グラフ、アイコン、レイアウトの情報などさまざまな視覚的要素が含まれており、BERT を始めとしたテキストを対象にした機械読解モデルではこれらの情報を考慮して回答することができない。

そこで、NTT では文書画像に対する機械読解のデータセット VisualMRC ⁴⁾ を構築して研究を進めている。このデータセットは英語の Web ページの

スクリーンショットの文書画像に対する自由記述型の質問応答データである。特徴として、文書内の領域 (Region Of Interest ; ROI) をタイトル・段落・リスト・画像・キャプションなど 9 クラスに分類してアノテーションしている点が挙げられる (図-8)。文書レイアウトを考慮可能な機械読解モデルの実現に向けて、文献 ⁴⁾ では、事前学習済の Text-to-Text モデルをベースに、物体認識技術を適用して抽出した文書中の ROI と、さらに文字認識技術 (OCR) を適用して抽出したトークンの位置・外観情報を追加入力とするモデルを提案した (図-9)。人間の質問応答精度にはまだ及ばないものの、ROI の意味クラスや、OCR で抽出された文字の位置座標や外観を加えることで性能が向上することが確認されている。

人間と AI の共生へ

AI が人間と同じように自然言語を理解し、人間と自然なコミュニケーションが取れるようになれば、我々の社会における AI の活躍シーンは大きく広がる。機械読解は人と AI の共生社会を実現するための鍵となる技術の 1 つであり、今後も大きな発展が期待される。

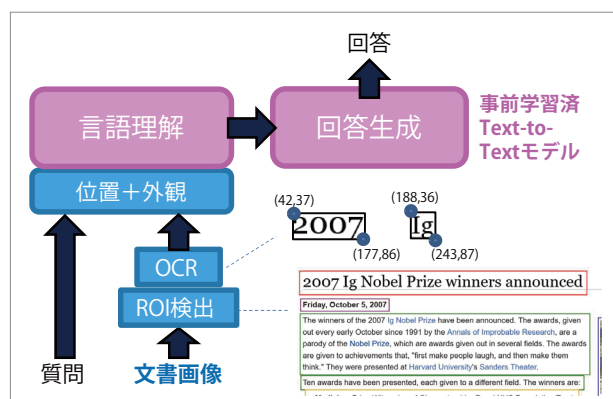
参考文献

- 1) Rajpurkar, P., Zhang, J., Lopyrev, K. and Liang, P. : SQuAD : 100,000+ Questions for Machine Comprehension of Text. In EMNLP, pp.2383-2392(2016).
- 2) Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In NAACL-HLT, pp.4171-4186 (2019).
- 3) Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. and Polosukhin, I. : Attention is All You Need. In NIPS, pp.5998-6008 (2017).
- 4) Tanaka, R., Nishida, K. and Yoshida, S. : VisualMRC : Machine Reading Comprehension on Document Images. In AAAI (2021).

(2021 年 5 月 31 日受付)



■ 図-8 VisualMRC データセットの例



■ 図-9 文書レイアウトを理解可能な機械読解モデル ⁴⁾

■ 西田京介 (正会員) kyosuke.nishida.rx@hco.ntt.co.jp

2008 年北海道大学大学院情報科学研究科博士課程了。博士 (情報科学)。2009 年日本電信電話 (株) 入社。現在、NTT 人間情報研究所特別研究員。