

3-3 大規模 Web 情報分析システム WISDOM X 深層学習版

3-3 Large-scale Web Information Analysis System WISDOM X: Deep-learning Version

呉 鍾勲 クロエツェー ジュリアン
OH Jong-Hoon and KLOETZER Julien

国立研究開発法人情報通信研究機構 (NICT) データ駆動知能システム研究センター (DIRECT) は、Web60 億ページの情報を基に「なに」「なぜ」「どうなる」「どうやって」といったタイプの様々な質問に回答することができる大規模 Web 情報分析システム WISDOM X「深層学習版」の Web 上での試験公開を 2021 年 3 月 31 日より開始した。この「深層学習版」のシステムは、深層学習が注目を浴びる前の世代の機械学習技術を用いて開発した WISDOM X (2015 年より試験公開) を、近年注目を集めている深層学習技術を用いて改良したもので、より広範な質問に対してより高い精度で回答可能である。本稿では、WISDOM X「深層学習版」で利用している深層学習ベースの質問応答技術を中心に紹介する。

At the Data-driven Intelligent System Research Center (DIRECT), NICT, we have released on March 31st, 2021, a neural version of our question-answering system WISDOM X that is able to answer to a variety of “what”, “why”, “how-to” and “what-happens-if” type questions using data extracted from 6 billion Web pages.

We had previously released a non-neural version of the system in 2015, before deep-learning took the machine learning world by storm; thanks to recent progress in said deep-learning research, we were able to expand the answering capabilities and improve the answer quality of WISDOM X. In this paper, we will present the deep-learning capabilities and technologies that were developed for the neural version of WISDOM X.

1 まえがき

国立研究開発法人情報通信研究機構 (NICT) データ駆動知能システム研究センター (DIRECT) では、ネット等にテキストとして流布している、社会における知、すなわち社会知を意味的に深く分析し有効活用できるようにするための自然言語処理技術を研究開発している。2015 年から試験公開し、さらに 2021 年 3 月 31 日に「深層学習版」(図 1 は試験公開のウェブページ) にアップデートして試験公開を開始した大規模 Web 情報分析システム WISDOM X はその研究成果の主要なものの一つであり、Web60 億ページから抽出した情報をもとに「なに」「なぜ」「どうなる」「どうやって」といったタイプの質問に回答することで、ユーザが求める情報やその関連情報などを迅速かつ容易に把握できるようにするための質問応答システムである。



図 1 WISDOM X 深層学習版 (<https://www.wisdom-nict.jp>) のトップページ

3 社会知コミュニケーション技術

我々の質問応答技術と関連する技術で一般に利用可能なものとしては検索エンジンによる情報検索があるが、既存の検索エンジンでは多くの場合、検索キーワードを含む文書を一度に十件程度ユーザに提示するだけであり、ユーザが探したい情報を網羅的に把握するためには検索で得られた文書を大量に読む必要があった。これに対し、WISDOM Xではユーザが入力した質問の端的な回答のリストを瞬時に提示することが可能であり、これによりユーザが求める情報の全体像を迅速かつ容易に把握することが可能である。

さらには、得られた回答に関して、更に深掘りする質問を入力することで価値ある想定外の情報を発見することも容易になる。例えば、「地球温暖化が進むとどうなる」という質問に対してWISDOM Xは「海水温が上昇する」や「ホッキョクグマは2030年に絶滅する」のような可能な出来事を回答し、さらに、ユーザが「海水温が上昇するとどうなる」といった深掘りをする質問を繰り返すことで、「地球温暖化が進む」→「海水温が上昇する」→「腸炎ビブリオ菌が増殖する」→「食中毒が発生する」のような情報を容易に得ることができる。この情報は、「海水温が上昇する」→「腸炎ビブリオ菌が増殖する」のような、因果関係をつないだ、出来事の連鎖を含むシナリオのようなものであり、「風が置けば桶屋が儲かる」式に解釈すれば「地球温暖化が進む」→「食中毒が発生する」という仮説を示すものと考えることができ、また多くの人にとっては想定外な情報であろう。この仮説は実際に2007年に収集したWebデータをもとにWISDOM Xを用いて発見されたものだが、その時点ではこの仮説を直接記述したページは収集したWebページには含まれていなかった。ところが、その後2013年のNature Climate Changeに掲載された論文[1]でバルト海における人為的な気候変動による海水温上昇とその周辺での食中毒増加の相関が科学的に確認された。つまり、我々の地球温暖化と食中毒の発生に関する仮説はたかだか一般のWebデータにある情報を組み合わせて作ったものであるが、その後、有力なジャーナルで科学的に確認された価値ある情報、更には、価値ある想定外であったということになる。今後、例えば、新たに海水魚に関する養殖ビジネスを始めるとしたら、上述の地球温暖化によるビブリオ菌の増加が食中毒の上昇につながるといったWeb由来の仮説的な将来シナリオを意識して、本物の海水を使った養殖ではなく、人工的な海水を使った陸上養殖を選ぶといった意思決定を行う時代も来るかもしれない。近年、イノベーションのための知識探索やリスク管理のための問題発見などの重要度が増しており、その際にはユーザが知りたい情報の全体像の把握や、上で例示したような価値ある想定外の発見が重要であ

る。我々はそうした高度に知的な作業をWISDOM Xによって支援可能であると考えている。

2015年から試験公開したWISDOM X(以降、従来版と呼ぶ)はSVM[2]等、深層学習以前の機械学習技術を使ったシステムであったが、新たに深層学習技術を導入し改良した深層学習版についても、2021年3月31日に試験公開を開始した。我々は近年注目を集めているBERT[3]という巨大ニューラルネットを、約350GBという大量のWebテキストやDIRECTで構築した高品質かつ大量の学習データで学習し、さらに必要に応じてBERTを敵対的学習[4]と組み合わせた独自の改良版[5][6]も導入することで「深層学習版」WISDOM Xを開発した。この深層学習版は従来版のWISDOM Xと比較してより広範な質問に回答可能であり、また質問応答の精度も大幅に向上している。具体的には、従来版でも回答可能であった「なに」型質問(「AIって、どんな社会問題の解決に使えるのかな?」、「高齢者のケアができるAIを使った技術には何がある?」のように「何/どこ/いつ/誰/どんな」等に対する回答を求める質問)や、「なぜ」型質問(「高齢者介護でコミュニケーションロボットが必要なのはなぜ?」のようにある事象を起こす原因や理由の説明を求める質問)、「どうなる」型質問(「量子コンピュータが実用化されるとどうなる」のようにある事象が招きうる出来事の説明を求めるタイプの質問)のそれぞれに関して、従来版と比較して大幅に高い精度で回答が可能である。例えば、図2で示している質問「AIが解決できそうな高齢化の問題は何がある?」は従来版では1つも回答を出力することができなかったが、深層学習版では「介護問題」、「孤独社会」を含む合計132件の回答を出力することが可能である。

さらに深層学習版では、従来版では対象外であった

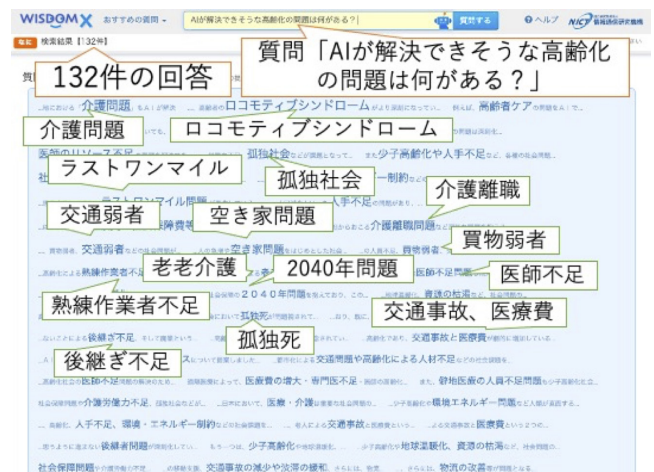


図2 質問「AIが解決できそうな高齢化の問題は何がある?」に対するWISDOM X 深層学習版の回答

「AIはどうしたらCO2回収の技術の開発に貢献できる?」のような「どうやって/どうしたら(How-to)」型の質問(ある目的/目標を達成するための方法・手続き・ノウハウ等を問う質問)についても回答できる新規の質問応答技術を導入している。これらの改良により、Web60億ページから抽出した大量の情報を用いて「チャーハンをパラパラにするにはどうしたらいい?」、「伊豆のB級グルメをおしえて」、「奈良観光はどうしたらいい?」、「有名なピアニストのホロヴィッツのおすすめの演奏は?」、「なぜギリシャで哲学が発展した?」、「花がきれいで、食べられる実をつける木で庭に植えるのに良いのは何?」等の多様な話題に関する質問に回答可能となっている。

本稿では、WISDOM X 深層学習版に使われている深層学習技術、特に一般に難しいと知られているノンファクトイド型の質問(「なぜ/どうやって/どうなる」型の質問)に対する質問応答技術について説明する。2では「なぜ/どうやって」型質問に対する質問応答技術について紹介し、3で「どうなる」型質問応答とその関連技術について紹介する。

2 「なぜ/どうやって」型質問応答

「なぜ」型質問や「どうやって」型質問の質問応答は、名詞等の単語レベルの回答を出力する「なに」型質問応答、いわゆるファクトイド型質問応答とは異なり、1文もしくはより長いテキストで回答を出力する必要があるノンファクトイド型の質問応答である。本節では、WISDOM X 深層学習版で利用されている「なぜ/どうやって」型質問の質問応答技術について説明する。

2.1 質問応答機構の構成

WISDOM Xにおける「なぜ/どうやって」型の質問応答では、図3に示すように質問解析、回答候補パッセージ検索、回答パッセージ特定、回答パッセージ中

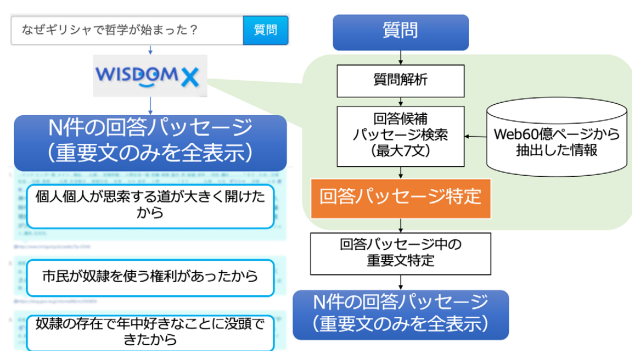


図3 WISDOM X 深層学習版における「なぜ/どうやって」型質問応答の処理手順

の重要文特定という4つの処理を行う。まず、入力された質問に対して質問解析と呼ばれる、質問からの検索キーワードの抽出処理を行う。ついで、回答候補パッセージ検索として、事前にWeb60億ページから抽出しておいたテキストパッセージ(連続する最大7文から成る文の集まり、以降回答候補パッセージと呼ぶ)を対象に、質問解析の結果得られた検索キーワードを用いて検索し、回答を含む可能性があるパッセージを取得する。次に、回答パッセージ特定用のニューラルネットワークを用いて回答候補パッセージが質問の回答を含むか否かを判定し、その判定スコアの上位N件を回答パッセージとして選択する。最後に、回答パッセージ中で特に回答とみなせる内容を含む重要文(1文)を同様にニューラルネットワークで特定し、特定した重要文全体と回答パッセージの残りの部分(ただし、重要文以外のテキストについては名詞、動詞、形容詞などの内容語のみを表示)を、回答パッセージのダウンロード元のURLと共に、最終回答として出力する。

WISDOM Xの従来版と比較して、深層学習版は回答の情報源をWeb40億ページからWeb60億ページに増強しており、さらに、回答パッセージ特定部に我々が独自に研究開発した深層学習のニューラルネットワークを導入したことでより高い精度で多様な回答を提示可能である。

以降では、まず2.2で我々が「なぜ/どうやって」型質問応答の性能向上のために独自に研究開発した回答パッセージ特定技術[5][6]を紹介し、次に2.3でその性能評価の結果について示す。

2.2 回答パッセージ特定技術

質問応答の回答パッセージ特定において、入力である質問や回答の候補(ここではテキストパッセージ)だけでなく、その質問と回答のつながりを表す「背景知識」が高品質な回答パッセージ特定のための重要な手がかりとなることが広く知られている。一例をあげると、質問「なぜ地球温暖化が起きる?」に対して、地球温暖化の原因を表す「温室効果ガスが増える」(原因)→「地球温暖化を加速させる」(帰結)という因果関係知識は、帰結部分が質問に対応し、原因部分がその質問の回答とみなせるため、回答パッセージ特定のための重要な手がかりとして使える背景知識といえる。我々はこのような背景知識を有効に活用する点に焦点を当て、深層学習を用いた新規の回答パッセージ特定技術を研究開発した。

関連する技術としてBERT[3]のような大規模言語モデルの事前学習時に特定の背景知識を学習させ、学習した言語モデルを回答抽出用に利用する従来研究[7]-[11]は存在するが、自然言語処理の個々の問題に応

じて異なる重要な背景知識を大規模言語モデルの事前学習に利用するのは、事前学習自体の実行コスト、つまり必要な計算機と計算時間、が非常に大きいことから、現実的なアプローチとは言い難い。そこでDIRECTでは、個別の問題で重要となる背景知識をより低コストで、かつ、柔軟に大規模言語モデルに組み込むことができる新たな枠組みBERTAC(BERT-style TLM with an Adversarially pretrained Convolutional neural network)を提案した [6]。さらにこのBERTACを利用して「なぜ／どうやって」型質問応答技術を開発し、「深層学習版」のWISDOM Xに搭載した。

BERTACを用いた学習では、事前に作成した背景知識に関するベクトルを出力する生成器と、通常的事前学習済み言語モデル、つまり、特定のタスクや特定のタイプの背景知識意識しない事前学習を実施したBERTを組み合わせて1つのネットワークを作成し、それを用いて最終的に処理を行いたい問題（例：「なぜ」型質問応答等）の学習を行う（図4：このような事前学習済みの言語モデルを用いた個別の問題に対する追加の学習をファインチューニングと呼ぶ）。より詳しくは、まず、大量のテキストから抽出した意味的関係の表現（例えば、上記例での因果関係「温室効果ガスが増える」(原因)→「地球温暖化を加速させる」(帰結)を表す言語表現)を用い、敵対的学習により背景知識のベクトル表現を自動生成する生成器 [5][6]をそれ単体で事前に学習する。次に、図4に示すように、BERTの上段に設けたTIER(Transformer for Integrating External Representation)という特別なTransformer層を用いて事前学習した生成器から得た背景知識のベクトル表現をBERTの処理過程に組み込み、背景知識に関する情報を効果的に回答パッセージ特定等の最終的なタスクで活用する。

BERTACについてさらに詳しく紹介するために、まず2.2.1で背景知識のベクトル表現を生成するための事前の学習について説明し、2.2.2でBERTACによる回答パッセージ特定について述べる。

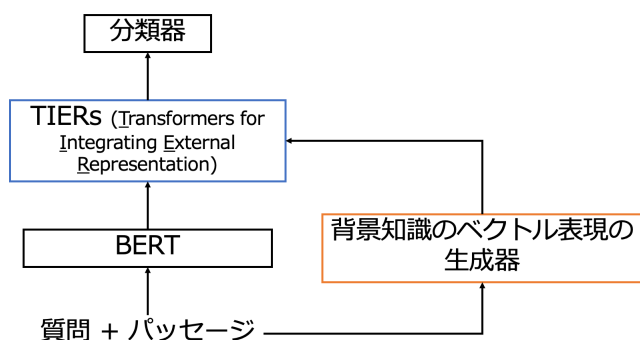


図4 BERTACによる回答パッセージ特定モデル

2.2.1 敵対的学習による背景知識のベクトル表現生成器の事前学習

敵対的学習は、学習対象の機械学習のモデルとそのライバルモデルを競合させ、モデルを強化する方法の一種である。なかでも近年注目されているのはIan Goodfellowが提案した「敵対的生成ネットワーク」(GAN: Generative Adversarial Networks) [4]で、主に画像生成・合成に活用されている。GANは学習対象の「生成モデル」とそのライバルの「識別モデル」(画像生成の場合は、入力された画像が生成モデルによって生成されたものなのか、あるいは、実在の画像なのかを識別するモデル)という2つのニューラルネットが互いに競争し成長していく学習方法で、この学習方法を用いることで、例えば、実存しないが実物のような画像を生成する「生成モデル」を学習することができる。AIが描いた作品「Edmond De Belamy (エドモンド・ベラミーの肖像)」*¹は、GANの代表的な適用例である。DIRECTでは、GANを画像生成ではなく質問応答に適用し、敵対的学習を行うことで、入力の問題と回答候補のパッセージから背景知識のベクトル表現を生成する「背景知識のベクトル表現生成器」を開発した。

この「背景知識のベクトル表現生成器」の学習では、前述のように、大量のテキストから抽出した意味的関係の表現を学習データとして用いる。より具体的には、「なぜ」型質問応答の重要な背景知識として知られている因果関係知識（例：「温室効果ガスが増える」(原因)→「地球温暖化を加速させる」(帰結)と、「どうやって」型質問応答で回答特定のための手がかりとなる道具・目的関係知識（例：「マスクを着用する」(道具)→「インフルエンザを予防する」(目的)）をWeb40億ページから抽出して敵対的学習の学習データとして利用した。因果関係知識の抽出にはOhらの手法 [12]（「ため」、「ので」などの因果関係抽出用の手がかり単語で入力のテキストから因果関係を含む可能性のある文を抽出し、抽出した文における因果関係の原因部と帰結部をCRF (Conditional Random Field)で特定する手法)を用い、約1億件の因果関係知識を獲得した。次に約1億件の知識からランダムにサンプリングした100万件を学習データとし、図5の左側に示すように因果関係の帰結部に対してその原因部のベクトル表現を生成するように背景知識のベクトル表現生成器を学習した（学習の詳細は後述）。

作成した背景知識のベクトル表現生成器を回答パッセージ特定に利用する際には、図5の右側に示すよう

*1 <https://obvious-art.com/portfolio/edmond-de-belamy>

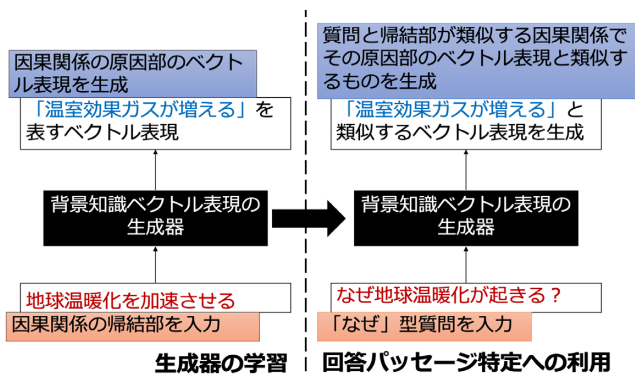


図5 背景知識のベクトル表現生成器の学習と回答パッセージ特定への利用例

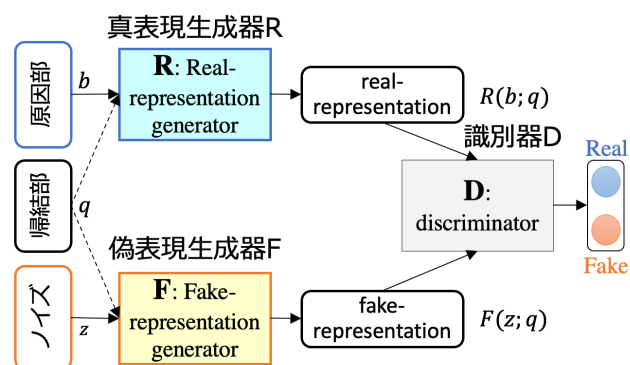


図6 敵対的学習による背景知識のベクトル表現生成器の学習

にベクトル表現生成器に質問を入力し、質問と類似している帰結部を持つ因果関係の原因部のベクトル表現と類似したベクトル表現を生成させる。例えば、質問「なぜ地球温暖化が起きる？」を入力した場合、学習に利用した因果関係知識「温室効果ガスが増える」(原因)→「地球温暖化を加速させる」(帰結)に基づいて、「温室効果ガスが増える」という原因部に相当するベクトル表現に類似したものが生成されることが期待できる。このベクトルは入力された質問の回答(の一つ)を端的に表している可能性が高いため、これを追加の入力とすることで最終的な回答パッセージ特定の精度向上につながると思われる。

また、道具・目的関係の場合は、「マスクを着用する」、「インフルエンザを予防する」のように〈名詞、助詞、述語〉から成る句の対をテキストから抽出し、BERTでその対が道具・目的関係に成り立っているかを判定するKadowakiらの手法[13]を用いて約1億件の道具・目的関係の言語表現を獲得した。また、因果関係の場合と同様にランダムにサンプリングした100万件を学習データとし、道具・目的関係の目的部に対してその道具部のベクトル表現を生成するように背景知識のベクトル表現生成器を学習した。回答パッセージ特定時には、学習した生成器を用いて質問に書かれている目的を達成するための道具・行為を表すベクトル表現を

生成し回答パッセージ特定の手がかりとして用いる。

次に、敵対的学習を利用して背景知識のベクトル表現生成器をどのように学習したかについて述べる。この学習は図6に示すニューラルネットによって行われるが、この中に含まれる偽表現生成器F(Fake-representation generator)と書かれている部分的なニューラルネットが学習終了後に背景知識のベクトル表現生成器となる。学習には、前述したように、因果関係の(原因部、帰結部)の対から成る学習データが使われ、入力の帰結部に対してその原因部のベクトル表現を生成するように偽表現生成器Fを学習させる。より具体的には(原因部 b 、帰結部 q 、ランダムノイズ z)の3つ組に対して、1) 帰結部 q と原因部 b から直接的に原因部のベクトル表現を生成する真表現生成器R(Real-representation generator)、2) 帰結部 q とランダムノイズ z から原因部のベクトル表現を自動生成することを目的とする偽表現生成器F(Fake-representation generator)、そして、3) 与えられたベクトルが真表現生成器Rの出力なのか、それとも偽表現生成器Fの出力なのかを識別する識別器D(Discriminator)を含む図6のニューラルネットを構成する。また、偽表現生成器Fと真表現生成器RはCNN(Convolutional Neural Network)、識別器DはFFN(Feed-Forward Network)といった軽量で学習が早いニューラルネットとなっているため、前述のように低コストで背景知識のベクトル表現生成器を事前学習することが可能である。

敵対的学習はこのニューラルネット全体で行われるが、その学習の過程では、偽表現生成器Fは原因部が入力されていないにもかかわらず、それが入力されたかのようなベクトル、つまり、原因部を入力とする真表現生成器Rのベクトルと区別できないようなベクトルを出力できるように学習する。より大雑把な言い方をすると、偽表現生成器Fは真表現生成器Rの出力を「真似」しようとし、識別器DとRはこの真似ができないように学習を進めるということであり、より正確に述べれば、偽表現生成器Fは、識別器DがFの出力と原因部を入力とする真表現生成器Rの出力が区別できないように学習を進める一方で、識別器Dと真表現生成器RはFとRそれぞれの出力をDが区別できるように学習を進めるということである。当然のことながら、FとD、Rの学習の方向性は違いに矛盾するので、FとD、Rは違いに競合しているとみなせる。このような学習を繰り返すことで、偽表現生成器Fはノイズと帰結部(つまり、「なぜ」型質問回答の入力である質問)しか入力されていないにもかかわらず、あたかも帰結部を入力したかのようなベクトル表現を出力できるようになることが期待できる。このベクトル表現には「なぜ」型質問回答で重要となる因果関係に関す

る背景知識の情報、すなわち質問に類似した帰結部に対応する原因部の情報が表現されていると考えることができる。このため、生成器の出力を利用することで、最終的な「なぜ」型質問応答の性能が向上することが期待できる。

道具・目的関係の背景知識のベクトル表現生成器の学習の場合は、(道具部 b 、目的部 q 、ランダムノイズ z) の3つ組みに対して上記と同様な敵対的学習を行い、最終的に、目的部 q とランダムのノイズ z から道具部のベクトル表現を自動生成する偽表現生成器 F を道具部のベクトル表現生成器として用いる。

2.2.2 BERTAC による回答パッセージ特定

上述した因果関係と道具・目的関係のための2つの背景知識のベクトル表現生成器を柔軟に大規模言語モデルに組み込むことができる枠組み BERTAC [6] で、背景知識のベクトル表現生成器を大規模言語モデルの一種である BERT と組み合わせて回答パッセージ特定の実験を行った。

このファインチューニング時には質問と回答候補のパッセージの対を BERT への入力とするが、さらにその質問・回答候補パッセージ対を背景知識のベクトル表現生成器への入力とし、背景知識のベクトル表現を生成してこれも最終的な回答パッセージ判定に利用する。より正確には、図6の帰結部 q (或いは、目的部) の代わりに質問を、ノイズ z の代わりに回答候補のパッセージを生成器の入力とすることで、質問の回答と関連する因果関係の原因部と道具・目的関係の道具部のベクトル表現を生成する。

生成器によって自動生成された背景知識のベクトル表現は、前述したように、図4に示す TIER (Transformers for Integrating External Representation) と我々が呼ぶ、特別な Transformer 層を経由して BERT の出力と統合する。TIERs は背景知識のベクトル表現から質問・回答パッセージ対に関する BERT の出力ベクトル表現への Attention 機構を備えており、これにより回答パッセージの中で背景知識のベクトル表現と関連ある部分により強い Attention を当てた回答パッセージ特定が可能となる。

2.3 性能評価

本稿では、既存手法との比較を容易にするため、まず過去に我々の論文 [5] で使われた「なぜ／どうやって」型質問応答データセットでの評価実験の結果を示し、次に学習データを増強しニューラルネットを学習させた現在公開中の WISDOM X「深層学習版」の「なぜ／どうやって」型質問応答の精度を示す。

我々の論文 [5] では質問応答の性能評価のため、DIRECT で作成した表1の「なぜ」型質問応答、「どう

やって」型質問応答の評価データセットを利用して評価実験を行った。このデータは人手で作成した「なぜ」型質問、「どうやって」型質問に関して Murata らの手法 [14] で各質問最大で20件の回答候補パッセージを Web6 億ページから収集し、さらに3人のアノテータが対象の質問に対してパッセージ中に適切な回答を含むか否かを判定し、最後に多数決で正解ラベルを決定したものである。

実験では本特集号 3-4 [15] で紹介する NICT BERT_{LARGE} を用い、以下の3つの手法での比較実験を行った。

- 1) BERT : BERT のみで回答パッセージ特定を行う手法
- 2) BERT+ 背景知識のベクトル表現 : 背景知識のベクトル表現と BERT の出力を組み合わせ単純に最終段の分類層の入力とする回答パッセージ特定の手法。
- 3) BERTAC : BERTAC で TIERs を経由して背景知識のベクトル表現と BERT を組み合わせて回答パッセージ特定を行う手法

上記2)と3)の手法では 2.2.1 で述べたように Web40 億ページから抽出した因果関係と道具・目的関係を表す言語表現を学習データとして学習し、その結果で得られた2つの生成器(因果関係用の生成器と道具・目的関係用の生成器)を用いて背景知識のベクトル表現を生成し、回答パッセージ特定に用いた。

評価時には、入力の質問と回答候補のパッセージに対して各手法が出力する回答パッセージ判定のスコアを利用して入力の質問に関する回答候補パッセージをランキングし、最上位の精度「P@1 (Precision at top answer)」と質問ごとにランク付けした結果における平均精度の平均「MAP (Mean Average Precision)」の2つの評価尺度で性能評価を行った。

表2に3つの手法の性能評価の結果を示す。この結果から「BERT」と比較して「BERT+ 背景知識のベクトル表現」と「BERTAC」がいずれも高い性能を示しており (BERT に対して P@1 で 0.4 ~ 4.2% 向上)、また「BERT+ 背景知識のベクトル表現」よりも「BERTAC」の性能が高いことから (P@1 で 2.8 ~ 3.8% 向上)、背景知識のベクトルそのものとそのベクトルを活用する際に TIER を用いた背景知識のベクトル表現と BERT の出力の統合の両方が有効であることがわかる。また、英語の質問応答に対する評価結果であるため表2の結果との直接比較はできないが、我々の論文 [6] では敵対的学習を使わずに学習した背景知識のベクトル表現生成器を用いた設定での BERTAC と敵対的学習で学習した生成器を用いる上記3)の手法の間の比較実験

表 1 「なぜ」型、「どうやって」型質問応答の性能評価に使われたデータセット

	「なぜ」型質問応答		「どうやって」型質問応答	
	質問数	パッセージ数	質問数	パッセージ数
学習データ	850	17,000	1,000	20,000
開発データ	100	2,000	200	4,000
評価データ	500	10,000	1,200	24,000
合計	1,450	29,000	2,400	48,000

表 2 「なぜ／どうやって」型質問応答の比較実験の結果

	「なぜ」型質問応答		「どうやって」型質問応答	
	P@1	MAP	P@1	MAP
BERT	63.8	61.4	55.1	55.2
BERT+ 背景知識のベクトル表現	64.2	61.9	56.0	55.4
BERTAC	68.0	62.5	58.8	56.6

表 3 因果関係と道具・目的関係の生成器に対する Ablation test の結果

	「なぜ」型質問応答		「どうやって」型質問応答	
	P@1	MAP	P@1	MAP
BERTAC	68.0	62.5	58.8	56.6
因果関係の生成器を除いた場合	67.0	62.4	57.7	55.9
道具・目的関係の生成器を除いた場合	66.8	62.0	57.3	56.5

も行い、背景知識のベクトル表現生成器の学習における敵対的学習の有効性を確認している。

また、BERTACに使われた因果関係と道具・目的関係の2つの背景知識のベクトル表現の生成器がそれぞれの程度性能向上に貢献するかを調査するため、Ablation testを行いそれぞれの生成器をBERTACから除いた場合の性能を評価した。表3にその結果を示すが、2種類の生成器のいずれかを除くと性能低下が見られるため、2種類の生成器がともにBERTACの性能向上に貢献しているといえる。

現在公開中のWISDOM X「深層学習版」は、「なぜ」型質問応答と「どうやって」型質問応答それぞれの更なる性能向上のために、各々約20万件の質問・パッセージ対から成る学習データ(表1の約10倍以上で合計40万件以上)を構築し、それらのデータで学習したBERTACのモデルを用いて回答パッセージ特定を行っている。この学習データを利用した性能評価のため、まず、DIRECTで作成した200件の「なぜ」型質問

表 4 WISDOM X「深層学習」の「なぜ／どうやって」型質問応答の精度

「なぜ」型質問応答		「どうやって」型質問応答	
P@1	AR@3	P@1	AR@3
83.0	92.5	73.5	89.0

と200件の「どうやって」型質問を現在公開中のWISDOM X「深層学習版」の質問として入力して各々の質問に対する上位3件の回答パッセージを収集した。次に3人のアノテータが入力の質問で得られた回答パッセージに適切な回答が含まれているか否かを判定し、最後に多数決で正解ラベルを決定した。性能評価には、表2、3と同様に質問ごとの最上位の精度「P@1 (Precision at top answer)」と、また、上位3位以内に適切な回答を含む比率を表すAR@3を評価尺度として用いた。

表4に「なぜ／どうやって」型質問応答の性能評価の

結果を示す。評価方法、データが異なるため、表2のBERTACとの直接比較は難しいが、学習データを増強したWISDOM X「深層学習版」は73.5～83.0%という表2よりも高いP@1を示している。さらに、AR@3では89%以上の高い性能となっており、この結果はWISDOM X「深層学習版」に「なぜ／どうやって」型質問のいずれかを10回入力すると9回程度は上位3件以内に回答を見つけられることを意味する。現在、多様な形式の質問を考慮した更なる学習データの増強を実施しており、増強した学習データの利用やその他の技術の改良を行うことで更なる性能向上を目指す予定である。

3 「どうなる」型質問応答

我々は、言語によるコミュニケーションの最も重要な目的の一つが、ある出来事が生じた後、どのような出来事が生じるかの予測や、予測するための知識を社会において共有することであると考えている。こうした予測やそのための知識があるコミュニティで正しく共有されれば、そのコミュニティ自体やその構成員の生存はもとより、コミュニティ全体の繁栄につながるであろう。こうした予測やその共有を可能にする知識が「出来事Aが生じるならば出来事Bが生じやすくなる」ことを示す因果関係と呼ばれるものである。WISDOM Xの「どうなる」型質問応答は、まさにそのような因果関係に関する知識を、質問応答を通じてユーザーに提示するものであり、Webページから抽出した因果関係を表現するテキストの巨大な知識データベース（以降、因果関係データベースと呼ぶ）を活用し、例えば「地球温暖化が進むとどうなる」という質問に対して「海水温が上昇する」のような地球温暖化という出来事が起こった結果を回答として、情報源となったWebページのURLと共にユーザーに提示する。さらにユーザーがその回答をさらに深堀りする質問を繰り返すことで、1に例示した様に、「地球温暖化が進む」→「海水温が上昇する」→「腸炎ビブリオ菌が増殖する」→「食中毒が発生する」のような潜在的に起こり得る因果関

係の連鎖、つまり、一種の将来シナリオを発見することもできる（図7を参照）。この地球温暖化から食中毒につながる将来シナリオが科学的に確認された[1]ということは、どこにでもあるWebテキストを組み合わせることで作成したシナリオが実際に未来の予測につながる価値ある内容を含み得ることを示している。

高品質な「どうなる」型質問応答を実現するためには、その回答の知識源となる大規模で高品質な因果関係データベースが重要となる。従来版のWISDOM Xで利用されていた因果関係データベースの構築には、深層学習以前の機械学習アルゴリズムであるSVM[2]を利用した因果関係認識技術[16]が利用されていたが、そこでは因果関係の候補である〈名詞、助詞、述語〉の対（例：「地球温暖化が進む」と「海水温が上昇する」の対）とその対が出現している文全体（例：「地球温暖化が進むと海水温が上昇するのは当たり前だね？」）を入力とし、更に多様な付加的な情報を背景知識として入力することで因果関係の抽出処理を実施していた。その後、我々が作成した高品質なBERTが利用可能になったことにより、そのBERTを利用した因果関係認識技術を開発した（詳細は下記で述べる）。さらに、因果関係に関する知識を獲得するための情報源であるWebテキストの規模をWeb60億ページに増やすことで、大規模で質の高い因果関係データベースを9千万件の規模で構築した。質問応答では、図8で表すように従来版WISDOM Xと同様に、ユーザーが入力した質問から〈名詞、助詞、述語〉の組を抽出し、それらの意味的特性[17]を考慮しつつ、因果関係データベースの原因部分のテキストと柔軟な照合を行うことで、検索された因果関係の帰結部分をその周辺のテキストと共に回答としてユーザーに提示する。

BERTを利用した因果関係認識技術の研究開発には、従来のSVM[2]を用いた因果関係認識技術[16]を開発した際に作成した学習・評価データを利用する（この学習・開発・評価用データの件数は表5を参照）。ただし、従来技術では因果関係候補となる〈名詞、助詞、述語〉の対に加え、その対が出現した文や付加的な情報も利用していたのに対し、BERTを使った因果



図7 WISDOM Xの「どうなる」質問応答機能で発見できる因果関係のシナリオの例

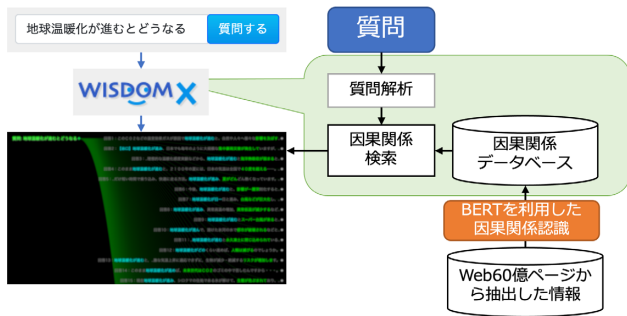


図 8 WISDOM X 深層学習版における「どうなる」質問応答機構

関係認識では、本特集号 **3-4** [15] で詳細を説明しているように、〈名詞、助詞、述語〉の対のみを入力した場合の性能がより高かったため、それらの対だけを入力として、因果関係の認識処理を行った。また、学習に利用する BERT は本特集号 **3-4** [15] で紹介する NICT BERT_{LARGE} を利用し、表 5 の学習データを利用してファインチューニングを行った。

性能評価の評価尺度には平均精度を用い、SVM を用いた従来版の因果関係認識技術 [16] との評価を行った。結果を表 6 にまとめる (BERT に関する評価は本特集号 **3-4** [15] で報告されるものと同一である)。この結果を見てわかるように BERT を利用した因果関係認識技術は、入力として因果関係候補の〈名詞、助詞、述語〉の対しか入力していないにもかかわらず、SVM に基づく技術と比較して圧倒的に高い平均精度を得ている。「深層学習版」WISDOM X の因果関係データベースにはこの BERT 版の因果関係認識技術を用いており、従来版の WISDOM X での「どうなる」型質問応答と比較して、質の高い「どうなる」型質問応答並びにそれを利用した興味深いシナリオの発見が期待できる。

最後に、上記の方法で獲得した因果関係と、本特集号でも取り上げている対話システム (特に雑談対話システム) との関係について補足しておきたい。本特集号 **3-5** [18]、**3-4** [15] で説明するように、我々は日常の対話、特に雑談等において、対話参加者の間で将来に関するリスクやチャンス共有するという意味で、現在対話中で注目されている出来事 (例えば、特定のタイプの野菜を食べること) が、次にどのような出来事 (例えば、一種の健康状態の改善) を引き起こすかを示す因果関係知識が大きな役割を果たすと考えている。また同時に、通常何気なく行われると想定される雑談的応答の動機・理由・目的を明示化し、より適切な雑談を実現する、あるいはそのための雑談内容の制御等を可能にするという点においても、因果関係は重要であると考えている。実際、この「どうなる」型質問応答の出力は、本特集号 **3-5** [18] で説明しているように、

表 5 因果関係認識の性能評価に使われたデータセット

	因果関係数
学習データ	107,068
開発データ	23,602
評価データ	23,650
合計	154,320

表 6 因果関係認識の比較実験の結果

手法	平均精度
SVM [16]	46.27 %
BERT	74.44 %

雑談対話システム WEKDA で中心的な役割を果たしている。高齢者介護の支援を目的とするマルチモーダル音声対話システム MICSUS では、この WEKDA を介することで、実証実験において高い品質の雑談的応答を生成することを実現している。

4 おわりに

本稿では、DIRECT で研究開発してきた大規模 Web 情報分析システム WISDOM X「深層学習版」とその WISDOM X に使われている深層学習技術について紹介した。従来版の WISDOM X で使われた技術は、その後、SNS を用いて災害時の情報の収集分析を行う対災害情報分析システム DISAANA、災害状況要約システム D-SUMM 等でも活用され、民間企業へ技術移転された。また、現在 DIRECT で開発中である「次世代音声対話システム WEKDA」や「高齢者介護用マルチモーダル音声対話システム MICSUS」(WEKDA と MICSUS の詳細については本特集号 **3-5** [18] を参照) でも、Web 上の情報を用いた多様な雑談的対話を実現するために深層学習版 WISDOM X の質問応答結果が利用されているなど、WISDOM X の技術は多岐にわたり利用されている。さらに今後も WISDOM X そのものもしくはその中で利用されている技術を、高齢者介護など、日本社会における重要課題の解決、解消に資する技術に応用していく予定である。

また、本特集号 **3-4** [15] で紹介しているように、今後の WISDOM X の技術展開を容易にするために深層学習版 WISDOM X の高速化・軽量化にも取り組んでおり、これによってより少ない規模の GPGPU でも WISDOM X を運用することが可能となっている。さらに、本特集号 **3-2** [19] で紹介している自動並列化深

層学習ミドルウェア RaNNC を用いてより巨大で高品質な言語モデルを構築中であり、こうした巨大言語モデルも WISDOM X のような各種質問応答技術で活用していきたいと考えている。

【参考文献】

- 1 C. Baker-Austin, J. Trinanes, N. Taylor, R. Hartnell, A. Siitonen, and J. Martinez-Urtaza, "Emerging Vibrio risk at high latitudes in response to ocean warming," Nature Climate Change, vol.3, no.1 pp.73-77, 2013.
- 2 V. N. Vapnik, "The nature of statistical learning theory," Springer-Verlag New York, Inc., New York, NY, US, 1995.
- 3 J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, vol.1 (Long and Short Papers), pp.4171-4186, 2019 .
- 4 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," Advances in Neural Information Processing Systems, vol.27, pp.2672-2680. Curran Associates, Inc, 2014.
- 5 J.-H. Oh, K. Kadowaki, J. Kloetzer, R. Iida, and K. Torisawa, "Open-domain why-question answering with adversarial learning to encode answer texts," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), pp.4227-4237, 2019.
- 6 J.-H. Oh, R. Iida, J. Kloetzer, and K. Torisawa, "BERTAC: Enhancing transformer-based language models with adversarially pretrained convolutional neural networks," Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021), pp.2103-2115, 2021.
- 7 B. He, D. Zhou, J. Xiao, X. Jiang, Q. Liu, N. J. Yuan, and T. Xu, "BERT-MK: Integrating graph contextualized knowledge into pre-trained language models," Findings of the Association for Computational Linguistics: EMNLP 2020, pp.2281-2290, 2020.
- 8 M. E. Peters, M. Neumann, R. Logan, R. Schwartz, V. Joshi, S. Singh, and N. A. Smith, "Knowledge enhanced contextual word representations," Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pp.43-54, 2019.
- 9 R. Wang, D. Tang, N. Duan, Z. Wei, X. Huang, J. Ji, G. Cao, D. Jiang, and M. Zhou, "K-adapter: Infusing knowledge into pre-trained models with adapters," CoRR, abs/2002.01808, 2020.
- 10 Y. Sun, S. Wang, Y. Li, S. Feng, H. Tian, H. Wu, and H. Wang, "ERNIE 2.0: A continual pre-training framework for language understanding," Proceedings of the AAAI Conference on Artificial Intelligence, pp.8968-8975, 2020.
- 11 W. Xiong, J. Du, W. Y. Wang, and V. Stoyanov, "Pretrained encyclopedia: Weakly supervised knowledge-pretrained language model," 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020.
- 12 J.-H. Oh, K. Torisawa, C. Hashimoto, M. Sano, S. De Saeger, and K. Ohtake, "Why-question answering using intra- and inter-sentential causal relations," Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013), pp.1733-1743, 2013.
- 13 K. Kadowaki, R. Iida, K. Torisawa, J.-H. Oh, and J. Kloetzer, "Event causality recognition exploiting multiple annotators' judgments and background knowledge," Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019), pp.5820-5826, 2019.
- 14 M. Murata, S. Tsukawaki, T. Kanamaru, Q. Ma, and H. Isahara, "A system for answering non-factoid Japanese questions by using passage retrieval weighted based on type of answer," Proceedings of NTCIR-6, 2007.
- 15 飯田 龍, "DIRECT における深層学習を用いた大規模自然言語処理," 情報通信研究機構研究報告, 本特集号, 3-4, 2022.

- 16 C. Hashimoto, K. Torisawa, J. Kloetzer, M. Sano, I. Varga, J.-H. Oh, and Y. Kidawara, "Toward future scenario generation: Extracting event causality exploiting semantic relation, context, and association features," Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014), pp.987-997, 2014.
- 17 C. Hashimoto, K. Torisawa, S. De Saeger, J.-H. Oh, and J. Kazama, "Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web," Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp.619-630. Association for Computational Linguistics, 2012.
- 18 水野 淳太, "高齢者介護支援用マルチモーダル音声対話システム MICSUS," 情報通信研究機構研究報告, 本特集号, 3-5, 2022.
- 19 田仲 正弘, "自動並列化深層学習ミドルウェア RaNNC," 情報通信研究機構研究報告, 本特集号, 3-2, 2022.



吳 鍾勳 (おー じょんぶん)

ユニバーサルコミュニケーション研究所
データ駆動知能システム研究センター
研究マネージャー
博士(工学)
自然言語処理、質問応答、言語モデル、情報検索

【受賞歴】

2015年 モバイル・コミュニケーション・ファン
ドコモ・モバイル・サイエンス賞
先端技術部門優秀賞
2014年 Twitter Data Grants 獲得



Kloetzer Julien (くろえつゑー じゅりあん)

ユニバーサルコミュニケーション研究所
データ駆動知能システム研究センター
主任研究員
博士(情報学)
自然言語処理、質問応答、音声対話システム

【受賞歴】

2015年 モバイル・コミュニケーション・ファン
ドコモ・モバイル・サイエンス賞
先端技術部門優秀賞
2014年 Twitter Data Grants 獲得