

大規模 Web 情報分析システム WISDOM X「深層学習版」の試験公開を開始 ～Web60 億ページの情報に基づき質問に回答する質問応答システムを BERT で強化～

【ポイント】

- 約 350GB の Web テキストで事前学習した BERT 及びその独自改良版を用い、様々な質問に回答
- 既存の非深層学習版に比べ大幅に精度向上。新たに「どうやって」(How-to) 型の質問にも対応
- 関連技術は、高齢者介護や防災等、社会課題解決のための AI システム等に展開予定

国立研究開発法人情報通信研究機構(NICT、理事長: 徳田 英幸)は、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の開発した、大規模 Web 情報分析システム WISDOM X の「深層学習版」の試験公開を <https://www.wisdom-nict.jp/>にて 2021 年 3 月 31 日から開始しました。

WISDOM X は 2015 年から試験公開していましたが、深層学習を使用しない非深層学習版でした。一方、このたび公開開始となった「深層学習版」は、近年注目を集めている BERT^{*1} という巨大ニューラルネットワークを、約 350GB という大量の Web テキストや NICT で構築した高品質かつ大量の学習データで学習させ、さらに独自技術と組み合わせた改良版も使って、より広範な質問へのより高い精度での回答を実現したものです。質問応答の情報源としては Web60 億ページから抽出した情報を使っています。

今後、関連技術を高齢者介護、防災等、重要な社会課題の解決に貢献する対話システム等、様々な AI システムに展開していく予定です。上記 URL から、商用目的を除き、どなたでもお試しください。

【背景】

NICT では過去 10 年以上にわたり、ネット上のテキストを対象とする質問応答技術の研究開発を進めてきました。既存の検索エンジンは多くの場合、検索キーワードを含む文書を一度に十件程度提示するだけであり、質問の回答を網羅的に集めるためにはユーザが提示された文書を大量に読む必要があります。一方で、WISDOM X のような質問応答システムは、質問の端的な回答のリストを提示することができ、関連する情報の全体像を迅速かつ容易に把握可能にし、さらには価値ある想定外の発見を容易にします。近年、重要性を増しているイノベーションやリスク管理といった不確実性に対処する作業では、こうした情報の全体像や価値ある想定外の把握が必須であり、WISDOM X はこのような把握を容易にし、価値ある考えるヒントを提供できると考えています。

【今回の成果】

今回、NICT は、大規模 Web 情報分析システム WISDOM X を「深層学習版」にバージョンアップし、2021 年 3 月 31 日から試験公開を開始しました。
(<https://www.wisdom-nict.jp/>)

2015 年に試験公開した WISDOM X の初期バージョンは、深層学習が注目を浴びる前の世代の機械



図 1: 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

学習技術を使っていました。今回、試験公開を開始する「深層学習版」は、近年注目を集めているニューラルネットワークである BERT や、BERT と敵対的学習²と呼ばれる深層学習の技法を組み合わせた独自改良版を使い、Web60 億ページから抽出した情報を基にして、様々なタイプの質問に柔軟に回答します。

より具体的には、初期バージョンでも回答可能であった「何／どこ／いつ／誰／どんな」等のタイプの質問(例:「AI って、どんな社会問題の解決に使えるのかな?」、「AI が解決できそうな高齢化の問題は何がある?」、「高齢者のケアができる AI を使った技術には何がある?」)や、「なぜ」型質問(例:「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?」、「どうしてフレイルを防ぐのに会話が重要なのか?」、「どうなる」型質問(例:「量子コンピュータが実用化されるとどうなる」)に関して、より多様な質問により高い精度で多様な回答を提示します。

例えば、図 1 で示した質問「AI が解決できそうな高齢化の問題は何がある?」は初期バージョンでは回答ができませんでしたが、今回のバージョンアップで回答可能になったものです。加えて、初期バージョンでは回答ができなかった「AI はどうやって高齢化の問題を解決できる?」、「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)」型の質問にも回答します。また、「チーズとネギがあるけど、つまみになにをつくったらいいかな?」のような日常的な言い回しをそのまま入力しても回答できます(図 2 参照)。さらに、大量の Web 情報を使うことで、「チャーハン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、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、多様な話題の質問に回答します。

得られた回答は通常の検索エンジンと同様に情報源の Web ページにリンクされており、そうしたリンクをさらなる周辺情報の収集や回答の適切さの判断に使うことも可能です。なお、WISDOM X で利用している BERT やその独自改良版は、大量の Web テキスト約 350GB に加えて NICT が構築した大量の高品質な学習データで学習しています。

【今後の展望】

2015 年に試験公開を開始した WISDOM X の初期バージョンで使われた技術は、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 DISAANA³、災害状況要約システム D-SUMM⁴ 等に展開され、民間企業による商用化も開始されています。同様に、今回公開開始の深層学習版の技術も災害関連技術に導入され、商用ライセンスが行われているほか、現在開発中の次世代音声対話システム WEKDA⁵、高齢者介護用マルチモーダル音声対話システム MICSUS⁶ において、Web の情報を用いた多様な雑談的対話を実現するために既に組み込まれています。

また、2021 年 3 月 31 日にフリーソフトウェアとして NICT が公開を開始した自動並列化深層学習ミドルウェア RaNNC⁷ を使って、さらに巨大かつ強力なニューラルネットワークの学習も行っており、そうした言語モデルも今後 WISDOM X 等のシステムに導入していくほか、民間企業等へのライセンスも行っていく予定です。今後もこうした技術を防災や高齢者介護のように日本の重要課題の解決に貢献できるシステム等に展開していきます。

＜関連するプレスリリース＞

- 2015 年 3 月 31 日 大規模 Web 情報分析システム「WISDOM X(ウィズダム エックス)」を Web 上に試験公開
<https://www.nict.go.jp/press/2015/03/31-1.html>
- 2021 年 3 月 31 日 自動並列化深層学習ミドルウェア RaNNC(ランク)をオープンソースで公開
<https://www.nict.go.jp/press/2021/03/31-2.html>

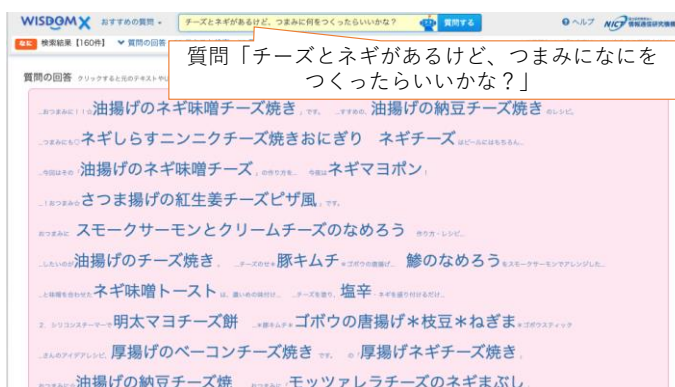


図 2: 日常的な言いまわし「チーズとネギがあるけど、つまみになにをつくったらいいかな?」に対する WISDOM X 深層学習版の回答(一部のみ)

＜ 本件に関する問合せ先 ＞

国立研究開発法人情報通信研究機構
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
鳥澤 健太郎
E-mail: wisdom-contact@ml.nict.go.jp

＜ 広報（取材受付） ＞

広報部 報道室
Tel: 042-327-6923
E-mail: publicity@nict.go.jp

<用語解説>

*1 BERT

2018年にGoogleから発表されたニューラルネットワーク。質問応答など、言語処理分野における様々なタスクで、従来の最高性能を更新した。その後、BERTを拡張あるいは参考にして多くのニューラルネットワークが提案されるなど、言語処理分野の深層学習研究に極めて大きな影響を与えた。

*2 敵対的学習

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) (詳細は参考文献[1]を参照)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、実物の画像なのかを識別するモデル)という二つのネットワークが互いに競合し成長していく学習方法で、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)」は、GANの代表的な適用例である。

NICTでは、GANを質問応答用に拡張し、与えられた質問の回答特定用の手がかりを生成する「生成モデル」を学習させ、この生成モデルとBERTを組み合わせて質問の回答を特定する質問応答モデルを開発した(この生成モデルの学習法については参考文献[2]を参照)。今回公開したWISDOM Xでは、「なぜ」型の質問と「どうやって/どうしたら」型の質問に回答するために、こうした質問応答モデルが使われている。

*3 対災害情報分析システム DISAANA(ディサーナ)

Twitterで発信された災害関連情報をリアルタイムに分析し、「熊本県で土砂災害が起きているのはどこか?」「熊本県で何が不足しているか?」といった質問への回答を地図上に可視化し、災害の被災状況の把握を容易にするシステム。NICTが開発し、2015年から<https://disaana.jp/>にて試験公開中。さらに、日本電気株式会社が、この技術の商用ライセンスを受け、2020年から商用サービスの販売を行っている。

*4 災害状況要約システム D-SUMM(ディーサム)

DISAANAと同様に、Twitterで発信された災害関連情報をリアルタイムに分析するが、質問に回答するのではなく、自治体名を指定すると関連する被災報告等を簡潔に要約し、被災状況の全体像の把握を容易にするシステム。内閣府の戦略的イノベーション創造プログラム(SIP)(第1期)による支援のもと、NICTが開発し、2016年から<https://disaana.jp/d-summ/>にて試験公開中。自治体等においても実際に活用され、豪雨による鉄橋流失を鉄道会社に先駆けて発見することに成功する等の事例もある。DISAANAと同様に、日本電気株式会社が、この技術の商用ライセンスを受け、2020年から商用サービスの販売を行っている。

*5 次世代音声対話システム WEKDA(ウェクダ)

多様な話題に関して、ユーザとブレインストーミング的な雑談を行うことを最終目標として狙った、NICTが開発している次世代音声対話システム。WISDOM Xとほぼ同じ仕組みで様々な質問に回答する他、「対話システムを作っています。」のような平叙文の音声入力に対しても、Webの情報をういて「対話システムを用いて回想法を行い、認知症の予防、改善をしよう」といった応答を行う。さらなる詳細は、NICTのプレスリリース「会話するAI、次世代音声対話システム『WEKDA(ウェクダ)』」(<https://www.nict.go.jp/press/2017/10/24-1.html>)を参照。

*6 高齢者介護用マルチモーダル音声対話システム MICSUS(ミクス)

要支援等の認定を受けている在宅高齢者に対して、介護モニタリングと呼ばれる、健康状態や生活習慣のチェックを音声での対話で行い、ケアプランの作成等に役立てるために開発中の対話システム。本来、介護モニタリングを実施するケアマネジャーと呼ばれる職種の介護職の負担を軽減するとともに、現在月一回とされている介護モニタリングの頻度を増やし、より高品質なケアにつなげることを目的とする。また、次世代音声対話システム WEKDA や WISDOM X の技術を使って Web 情報を用いた雑談的対話も行い、高齢者に飽きられることなく普段使いをしてもらうことも狙っており、さらには高齢者の健康状態を阻害する

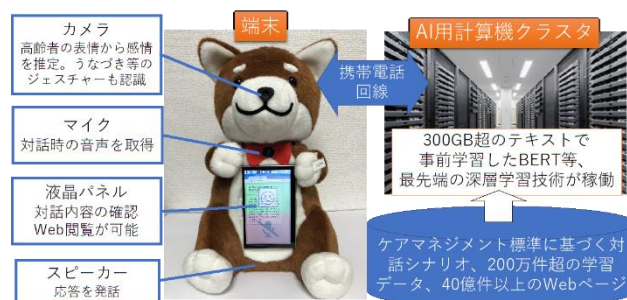


図3: MICSUSの構成

要因となるコミュニケーション不足の解消も狙う。内閣府の戦略的イノベーション創造プログラム(SIP)(第2期)の支援のもと、KDDI株式会社、NECソリューションイノベータ株式会社、株式会社日本総合研究所とNICTが共同開発を

行っている。さらなる詳細は、<https://www.youtube.com/watch?v=gCUrC3f9-Go> を参照。(YouTube で”MICSUS”と検索しても閲覧可能)

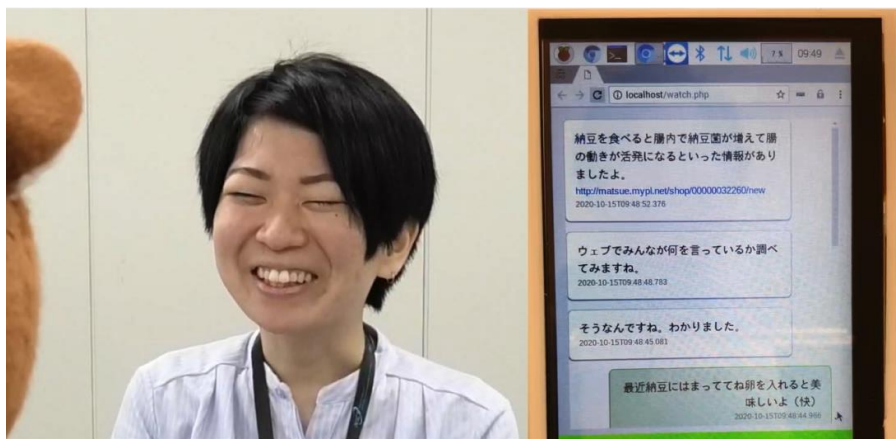


図 4: MICSUS での Web を使った雑談の様子
(コロナの影響で高齢者対象の実証実験ができないため、NICT の技術者が対話)

*7 自動並列化深層学習ミドルウェア RaNNC(ランク)

GPU1 枚に収まりきらない巨大ニューラルネットワークを高速に学習するために、NICT が東京大学と共同で開発したミドルウェア。代表的な深層学習ソフトウェアである PyTorch のために記述された既存のニューラルネットワークの定義を与えられると、実行速度を最適化しつつ、各々の GPU のメモリに収まるように自動的にニューラルネットワークを分割した上で、複数の GPU を使って並列に学習を行う。これまでも、同様に巨大なニューラルネットワークを分割し、並列で学習を行うソフトウェアは存在したが、ニューラルネットワークの定義を並列化のために開発者が書き換える必要があり、GPU 等に精通した開発者でなければ使えないものであるか、もしくは並列化は自動であるものの十分に大きなサイズのニューラルネットワークには適用できないものであるか、のいずれかであり、また、BERT のような特定のタイプのニューラルネットワークにしか適用できないものも多かった。RaNNC はこれらの制約を取り払うものであり、巨大ニューラルネットワーク構築の裾野を広げるものと期待される。RaNNC はオープンソースで GitHub に公開されている(URL: <https://github.com/nict-wisdom/rannnc>)。また、RaNNC の仕組みや性能評価は並列分散処理分野におけるトップレベルの国際会議である IPDPS (IEEE International Parallel and Distributed Processing Symposium) で発表予定である[3]。さらに RaNNC の概要は、GPU テクノロジカンファレンス(GTC) (2021 年 4 月 12-16 日開催)でも講演予定である。

参考文献

- [1] Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). Generative Adversarial Networks (PDF). Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014). pp. 2672–2680
- [2] Jong-Hoon Oh, Kazuma Kadowaki, Julien Kloetzer, Ryu Iida and Kentaro Torisawa, Open Domain Why-Question Answering with Adversarial Learning to Encode Answer Texts, In the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019), pp.4227--4237, Florence, Italy, July, 2019
- [3] Masahiro Tanaka, Kenjiro Taura, Toshihiro Hanawa and Kentaro Torisawa, Automatic Graph Partitioning for Very Large-scale Deep Learning, In the Proceedings of 35th IEEE International Parallel and Distributed Processing Symposium (IPDPS 2021), May, 2021. (発表予定)