

## 3-4 DIRECT における深層学習を用いた大規模自然言語処理

### 3-4 Large-scale Natural Language Processing using Deep Learning Techniques at DIRECT

飯田 龍  
IIDA Ryu

大規模言語モデルは現在の自然言語処理研究の根幹をなす要素技術となってきた。本稿では、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) で研究開発を行っている深層学習に基づくテキストの自動分類と自動生成に関する技術を紹介する。まず最初に、深層学習版の大規模 Web 情報分析システム WISDOM X 等の DIRECT の多くのシステム開発で活用されている、DIRECT で作成したテキスト分類タスク用の大規模言語モデルについて紹介する。次に、その分類用言語モデルを生成に利用するために追加学習して作成したテキスト生成タスク用言語モデルとそのモデルの次世代音声対話システム WEKDA での利用についても紹介する。最後に、深層学習版の WISDOM X の軽量化／高速化に関する取組についても紹介する。

Huge language models are becoming an elemental technology that forms the basis of current natural language processing research. In this paper, we present deep learning-based classification and generation techniques that we recently studied and developed at the Data-driven Intelligent System Research Center (DIRECT), NICT. First, we introduce the huge language models that we created for text classification tasks. We exploit these models for developing many systems at DIRECT, such as the deep learning version of our open-domain question answering system WISDOM X. Then, we present our language models for text generation tasks, which we created by additionally pretraining the aforementioned models for classification and modifying them to generate texts, and explain how we use these models in our next-generation spoken dialog system WEKDA. Finally, we present our efforts to make the deep learning version of WISDOM X lighter and faster.

#### 1 はじめに

ここ数年の深層学習の圧倒的な進歩により、他の人工知能の分野と同様に、自然言語処理の分野においても研究の様相が大きく変わってきている。従来の自然言語処理のアプローチでは、機械翻訳、情報抽出、質問応答といった最終的に開発したいアプリケーションが解くべき問題を部分的な問題に分割し、部分問題それぞれを段階的に処理することで最終的なアプリケーションの出力を得るということがなされてきた。例えば、文中の語の品詞 (名詞や動詞等) を特定する処理や、文の構造を解析する統語解析、入力中の個々の単語を他言語の同義の単語に置き換えるなどの部分問題が設計され、その部分問題の処理方法や、もしくはある部分問題の出力が与えられた状態で次の部分問題を適切に処理するための研究が進められてきた。しかし、近

年ではそのような部分問題への分割を行わず、未処理のテキストから、いきなり最終的な結果を出力するという、いわゆる end-to-end の処理が研究者の注目を集め、盛んに研究が行われている。この研究の盛り上がるの理由としては、自然言語処理分野における深層学習技術の発展により、そうした end-to-end の処理が容易になったことがあげられるが、特にニューラルネットの1つである Transformer [1] を利用した言語モデル BERT (Bidirectional Encoder Representations from Transformers) [2] が提案され、その有効性が広く知れ渡ったことの影響が大きい。

本特集号 3-1 [3] でも概要が紹介されているように、BERT では事前学習とファインチューニングという2段階の処理を行う。ここで事前学習とは言語モデルを巨大なテキスト集合を対象に教師なしで実施可能な問題に関して学習をすることをいい、一方、ファイン

チューニングとは事前学習で学習済みの言語モデルのパラメタを初期パラメタとして読み込み、個別の問題に関する学習データで追加で学習することをいう。この2段階処理の重要な点としては、より高品質な事前学習済み言語モデルを用いてファインチューニングするだけで、他の設定は変更せずとも、個別の様々な問題でより高い性能を得られるというものである。一般的に、質の高い事前学習済み言語モデルを作成するには、言語モデルの学習に利用するニューラルネットをより巨大にして、かつ、大規模で多様なテキストを用いて多くのGPGPUで大規模に事前学習を行う必要がある。このため、BERTが提案された後に、主に大規模GPGPUクラスタを保持する研究機関・企業によって事前学習手法の改良が進められている。

近年の事前学習に関する取組としては、主にパラメタ数をより巨大にすることでより汎用的な言語モデルを構築する研究が進められている。例えば、英語についてはGPT-3 [4] 等、日本語についてもHyperCLOVA [5] 等の開発がなされているが、それらのビジネス化については、必要な計算パワーが膨大であり、運用コストが非常に高くなることやヘイトのような問題のあるテキストを生成するリスク等の問題等もあり未知数である。そこで、我々としては実運用上問題とはならないと考えられる規模の大規模言語モデル事前学習を実施し、そのモデルを我々が作成済み、もしくは問題によっては新規に作成する高品質な学習データでファインチューニングすることで、より実用化が容易な環境、設定で動く先進的なアプリケーションの研究開発を進めている。より具体的には、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)で研究開発した技術について、従来の機械学習アルゴリズムであるSVM [6] からBERTへ移行し、大規模Web情報分析システムWISDOM X<sup>\*1</sup>の深層学習版を2021年3月末に公開した。また、今までは処理が難しかった自動生成系の問題において大規模言語モデルがうまく機能し、対話システム等の高度なアプリケーションでの積極的な利用が進んでいる。DIRECTでも、大規模言語モデルを用いて、次世代音声対話システムWEKDA [7]、マルチモーダル音声対話システムMICSUS (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の共同開発)等の研究開発を行っている。

上記のシステムは様々な深層学習の技術の複合的な組み合わせにより実現されているが、多くの部分問題は、例えば、「何で地球温暖化を防ぐ?」といった質問に対して与えられたテキストが適切な回答を含んでいるかといった分類問題と、例えば、ある質問と回答候補を含むテキストが与えられた時に、そのテキストを

もとに、質問の簡潔な回答を要約されたテキストとして生成するといった生成問題に大別できる。このうち、分類問題に関しては、前述のBERTもしくはその拡張(RoBERTa [8]、ALBERT [9]、ELECTRA [10]、DeBERTa [11] 等)を個別にファインチューニングすることにより、様々な自然言語処理の分類問題で高い性能を得ることができていることが知られている。ただし、この技術が公開された当初は事前学習をどのようなテキストを用いて、どのような規模のニューラルネット、ハイパーパラメタで事前学習すべきかについての知見が少なく、NICT DIRECTでも試行錯誤を繰り返しながら日本語を対象とした高品質の巨大言語モデルの事前学習の構築に取り組んだ。また現在もより巨大で高性能な事前学習済みモデルを得るために継続して研究を実施中である [12]。この過程で得られた日本語Wikipediaに関するBERT<sub>BASE</sub>モデルはNICT BERT 日本語 Pre-trained モデルという名称で一般に公開済みで3,500件以上もダウンロードされて広く利用されている。またBERT<sub>BASE</sub>モデルより巨大なニューラルネットを利用し、かつ、Web文書から得た質の高い大規模テキスト集合を用いて事前学習したNICT BERT<sub>LARGE</sub>モデルはDIRECTの各種システム中で各種分類問題を高性能に処理するために活用されている。公開版NICT BERT<sub>BASE</sub>とNICT BERT<sub>LARGE</sub>はほぼ同時期である2020年1~2月に事前学習が完了しており、その後、NICT DIRECTの各種分類器等の研究開発で利用されている。また、MICSUSで音声認識エラーに頑健な処理を実現するため、通常の仮名漢字混じり文の入力に加え、単語の読み情報も入力可能にし、学習データにノイズを一定量加えて事前学習を実施したHBERTというBERTの改良版も作成している(HBERTについては本特集号3-5 [13]を参照)。本稿では、まずこのNICT BERT<sub>LARGE</sub>の構築とその評価結果について紹介する。

また、DIRECTで対象とする生成問題には、WISDOM Xの回答(特に「なぜ」型や「どうやって」型の質問に対する長い回答)をWEKDAやMICSUSのような音声対話システムでコンパクトに提示するために、長い回答の要約を行ったり、また、Webテキストから音声対話で使える自然な発話を生成するといった課題がある。これらの生成課題に対しては、研究開発の当初はrecurrent neural network (RNN)を使用したEncoder Decoder ネットワーク [14]を利用していましたが、そこからUniLM [15]やTransformer ネットワーク [1][16]等のより質の高い生成結果が得られる技術へ移

\*1 <https://www.wisdom-nict.jp/>

行していった経緯がある。本稿ではそのような生成課題のうち長い回答の要約の技術について紹介する。最後に、近年研究開発を行った深層学習版の WISDOM X の高速化／軽量化に関する取組についても紹介する。

## 2 NICT BERT の事前学習

BERT [2] は Transformer ネットワーク [1] の encoder 部分を利用したニューラルネットとして構成されており、前述したように、大規模な教師無しテキストを用いて事前学習を行う。より詳しくは、(1)入力テキスト中の単語をある確率で [MASK] という特殊なトークンに置き換えてニューラルネットに入力し、もとの単語が何であったかを予測する masked language modeling (masked LM) と、(2)もとの単語列とその後続する単語列の対である場合を正例、もとの単語列とランダムに選択した単語列の対である場合を負例とした学習を行う next sentence prediction の学習を大規模に行う。学習に利用する Transformer ネットワークは入力の各単語を単語 embedding というベクトルに変換、また、その出現位置を表す位置 embedding、何個目の単語列のまとまりであるかを表すセグメント embedding を用いて、それらを合算して入力のベクトルを作成し、以降の多段から成る Transformer 層へ入力する。Transformer 層の各層では multi-head attention と呼ばれる複数の観点からの学習を行うが、上述の (1) (2) の学習を通じて、例えば、ある [MASK] となっている語を予測するために他のどの入力単語のどのような言語的な特性を捉えて予測をすべきかといった観点そのものを自動で学習し続けることにより、入力された言語の文の多様な言語的な特性に関する学習が BERT の内部で行われることになる。その後、ファインチューニングで BERT のパラメータを特定の問題に関して更に更新することでその問題に関して非常に高い性能が得られることが知られている (BERT のアーキテクチャの概要は図 1 を参照)。

BERT の事前学習の結果得られるモデルの質、つまり、その後のファインチューニングを行って得られるモデルが個別の問題で高い性能を得られるかは、モデルのパラメータの規模と後述するバッチサイズの事例を 1 ステップで学習するとして、事前学習全体で何ステップ分の処理を行うのか、別の言い方をすると、入力のテキストデータを何回読み込んで学習を行うのが影響することが知られている。さらには事前学習で利用するテキストの質も影響することが知られている。BERT の学習の事前学習の設定として代表的なものが BERT<sub>BASE</sub> と BERT<sub>LARGE</sub> である。前者は Transformer 層が 12 層、隠れ層が 768 次元のベクトルを利用したも

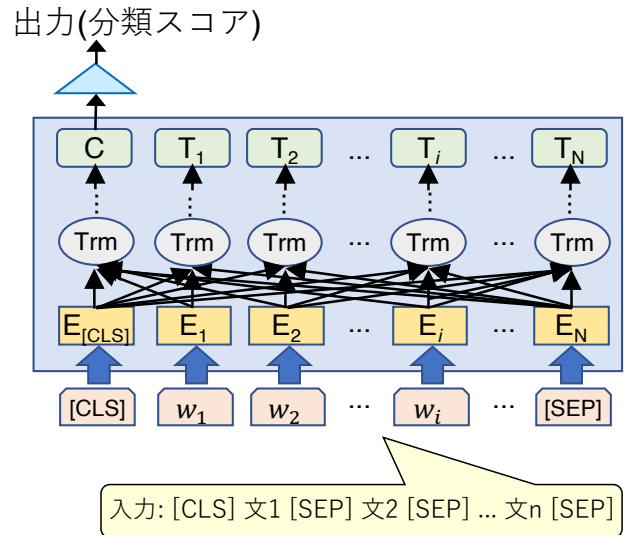


図 1 BERT のアーキテクチャ

のであり、後者の BERT<sub>LARGE</sub> は Transformer 層が 24 層、隠れ層が 1,024 次元のベクトルを用いたものとなっている。Devlin らが提案したオリジナルの BERT [2] の事前学習時のハイパーパラメータはバッチサイズが 256、最大系列長が 512、ステップ数が 100 万、学習率は  $1e-4$  である。また一般的には、BERT<sub>LARGE</sub> を事前学習したモデルをファインチューニングすることで BERT<sub>BASE</sub> よりも性能が高くなることが知られている。また、さらには学習時のバッチサイズ、つまり、1 ステップで何事例を同時に学習するも重要となる。例えば、BERT の拡張である RoBERTa というモデル [8] ではバッチサイズを 256 から最大 8,192 まで増やすが、ステップ数は 50 万ステップ<sup>\*2</sup> に設定して事前学習することでファインチューニング後の各種タスクの性能が向上したことが報告されている。

我々が実施した日本語を対象とした BERT の事前学習でも、上述の既存研究で性能が向上した各種設定を反映もしくは一部拡張することで学習を実施した。まず、日本語 Wikipedia の記事を対象に BERT<sub>BASE</sub> のモデルサイズを事前学習した際には、最初にバッチサイズを 4,096、ステップ数を 100 万、最大系列長を 128、学習率を  $1e-4$  に設定して学習し、次に、より長い入力にも対応できるように、バッチサイズを 4,096、ステップ数を 10 万、最大系列長を 512、学習率を  $2e-5$  に変更して追加で学習するといった 2 段階の事前学習を実施した。この結果得られた BERT モデルは NICT BERT 日本語 Pre-trained モデルという名称で一般に

\*2 バッチサイズを増やしたことで事前学習のステップ数はオリジナルの 100 万ステップよりも少ない最大 50 万ステップが採用されている。



表1 BERT 評価に利用した分類問題の概要 (具体例はいずれも著者が作成した作例である)

因果関係認識 [18][20][21]	入力された<名詞、助詞、述語>の対とその対が係り受け関係で出現する文を入力し、与えられた句の対が因果関係となるか否かを判定する課題	これまでの調査結果をまとめると、 <b>地球温暖化が進み、海水温が上昇し</b> 、さらにはその影響で海水中の有毒な微生物が増えることがわかっている。(太字箇所の「地球温暖化が進み」と「海水温が上昇し」の対が因果関係となるか否かを判定する)
ファクトイド質問応答 [19]	入力された「なに」型の質問と回答となる名詞、その名詞を含む文の3つ組に対して、質問の回答がその文中に含まれるか否かを判定する課題	質問: AI をどんな社会課題に利用する / 回答名詞: リモートワーク化 / 文: コロナ禍の状況において重要となるリモートワーク化に AI を効果的に利用することが考えられる。
バイナリパターン含意関係認識 [22][23]	変数 X と Y を含む句の対 (例: 「X が Y で眠る」と「X が Y にいる」の対) の間に含意関係が成り立つか否かを判定する課題	彼が横で眠る → 彼が横にいる / 本人が目の前で眠る → 本人が目の前にいる / 親がそばで眠る → 親がそばにいる (X と Y に入り得る語をそれぞれ X と Y に代入したものを複数個列挙して入力する)
解答可能性付き読解 [24]	質問、回答、文書の3つ組に対して、文書の読解によって質問に答えることができるかどうかを判定する課題	質問: 木下順二が執筆した『平家物語』を題材とした戯曲は何でしょう? / 回答: 子午線の祀り / 文書: 木下順二は 1978 年に『子午線の祀り』を発表した。この作品は『平家物語』に基づいて書かれたものであり、「群読」と呼ばれる独自の朗読形式を取り入れている。

公開している\*3。

また、より高品質な BERT モデルを構築すべく、DIRECT では BERT<sub>LARGE</sub> のモデル (以降、NICT BERT<sub>LARGE</sub> と呼ぶ) の事前学習も実施している (ただし、NICT BERT<sub>LARGE</sub> は一般には公開していない)。事前学習時のバッチサイズやステップ数、最大系列長の設定、2段階による事前学習は上述と同じだが、事前学習に利用するテキストを日本語 Wikipedia 記事から大規模 Web テキストに変更し、さらに、Web テキストの中でも、Oh ら [17] の因果関係検出器を適用して特定された因果関係を含むテキストパッセージのみを対象に事前学習用のテキスト集合を作成、その結果、日本語 Wikipedia の約 100 倍のサイズとなる 350 GB もの規模の学習データを作成して事前学習を行っている。我々が実施した予備調査の結果、この因果関係を含むテキストを事前学習へ利用することで、Wikipedia 単体やランダムにサンプリングした同じ規模の Web テキスト集合を利用した場合よりも質の高いモデルが得られることがわかっており、後述する因果関係認識やファクトイド質問応答で性能が向上することを確認している [18][19]。特に因果関係認識は DIRECT で研究対象としている「なぜ」型、「どうやって」型の質問応答をはじめとする多くの問題に関連しており、そのため、その因果関係認識で性能が向上することで他の関連する問題においても同様に性能が向上することが期待できる。また、本特集号 3-5 [13] で紹介する対話システム WEKDA で行う雑談発話は因果関係知識を用いた「どうなる」型の質問応答等を用いてユーザ発話に対するチャンスやリスクに関する応答を返すが、そこで利用される因果関係知識は因果関係認識の結果を利用し

て獲得されるため、そういった意味でも因果関係認識の性能向上は我々のシステム開発に直接的に貢献することになる。このように事前学習の設定を試行錯誤することで作成された NICT BERT<sub>LARGE</sub> は深層学習版 WISDOM X、WEKDA や MICSUS といった DIRECT が研究開発しているシステム内の様々な分類問題で利用されている。

### 2.1 NICT BERT の性能評価

NICT BERT 日本語 Pre-trained モデル公開時 (2020 年 3 月) に、その時期までに公開されていた日本語の BERT モデルと NICT BERT<sub>BASE</sub> を比較するために、DIRECT で過去に作成した分類問題のデータセットに適用して分類性能の評価を行った。比較に利用した日本語の BERT モデルの一覧を表 2 にまとめる。今回の研究報告では追加で NICT BERT<sub>LARGE</sub> についても評価を実施した。評価に利用した分類問題は、因果関係認識 [18][20][21]、ファクトイド質問応答 [19]、バイナリパターン含意関係認識 [22][23]、解答可能性付き読解 [24] の 4 種類であり、いずれも自然言語理解を対象とする重要な課題である (各問題の概要については表 1 を参照)。

評価では、比較用の BERT との実験条件を揃えるために、同じ学習率等のハイパーパラメタ集合を探索し、それぞれ開発用データでベストな性能を得たパラメタをベストパラメタとして、そのパラメタに関して得られたモデルで評価用データの性能を評価した。ただし、

\*3 <https://alaginrc.nict.go.jp/nict-bert/index.html>

表 2 評価に利用した BERT モデル一覧

モデル	モデルサイズ	語彙数	語彙	学習データ
NICT BERT <sub>BASE</sub> (日本語 Wikipedia, BPE 無)	BASE	100k	juman	日本語 Wikipedia
NICT BERT <sub>BASE</sub> (日本語 Wikipedia, BPE 有)	BASE	32k	juman+BPE	日本語 Wikipedia
NICT BERT <sub>LARGE</sub> (因果関係コーパス, BPE 無)	LARGE	100k	juman	因果関係コーパス
BERT <sub>BASE</sub> Multilingual Cased <sup>*4</sup>	BASE	120k (多言語)	heuristic+wordpiece	104 言語の Wikipedia
BERT 日本語 Pretrained モデル (Base WWM 版) <sup>*5</sup>	BASE	32k	juman+WWM	日本語 Wikipedia
BERT 日本語 Pretrained モデル (Large WWM 版) <sup>*5</sup>	LARGE	32k	juman+WWM	日本語 Wikipedia
Pretrained Japanese BERT models <sup>*6</sup>	BASE	32k	ipadic+WWM	日本語 Wikipedia
BERT with SentencePiece for Japanese text <sup>*7</sup>	BASE	32k	setencepiece	日本語 Wikipedia
hottoSNS-BERT <sup>*8</sup>	BASE	32k	setencepiece	大規模日本語 SNS コーパス

表 3 各種 BERT の複数タスクでの性能評価

モデル	因果関係認識		ファクトイド質問応答		バイナリパターン間含意関係認識		解答可能性付き読解	
	F1	平均精度	F1	平均精度	F1	平均精度	EM	F1
NICT BERT <sub>BASE</sub> (日本語 Wikipedia, BPE 無)	58.56	59.01	73.47	82.52	57.92	63.01	76.42	77.75
NICT BERT <sub>BASE</sub> (日本語 Wikipedia, BPE 有)	56.34	57.96	69.72	77.20	59.18	61.16	77.92	79.49
NICT BERT <sub>LARGE</sub> (因果関係コーパス, BPE 無)	<b>64.84</b>	<b>70.44</b>	<b>79.38</b>	<b>87.40</b>	<b>66.46</b>	<b>72.10</b>	<b>78.41</b>	<b>79.83</b>
BERT <sub>BASE</sub> Multilingual Cased <sup>*4</sup>	48.48	47.19	48.27	61.75	51.51	54.01	70.10	70.16
BERT 日本語 Pretrained モデル (Base WWM 版) <sup>*5</sup>	53.06	51.76	67.48	75.29	53.90	56.06	73.89	75.65
BERT 日本語 Pretrained モデル (Large WWM 版) <sup>*5</sup>	54.74	54.76	70.34	77.89	56.62	61.18	75.79	77.49
Pretrained Japanese BERT models <sup>*6</sup>	54.99	54.63	67.00	75.08	57.86	62.45	77.68	78.87
BERT with SentencePiece for Japanese text <sup>*7</sup>	54.76	55.67	72.87	80.28	55.53	58.74	73.66	76.83
hottoSNS-BERT <sup>*8</sup>	48.94	47.09	67.26	73.43	49.04	48.22	61.14	64.93

NICT BERT<sub>LARGE</sub> に関しては比較的低い学習率で性能が高くなることがわかったため、低めの学習率に関してもパラメタ探索を実施している。

NICT BERT<sub>BASE</sub> 公開当時の比較結果としては、NICT BERT<sub>BASE</sub> が因果関係認識、ファクトイド質問応答、バイナリパターン間含意関係認識の3つの問題については Byte Pair Encoding (BPE) 無しの設定で、また、解答可能性付き読解の問題については BPE 有りの設定で、他の研究機関等で公開されている BERT を上回る最高性能を得た。この結果から高バッチサイズで、十分な量のステップ数で事前学習することが重要であることがわかる。また、表 2 の NICT BERT<sub>BASE</sub> と日本語 Pretrained モデル (Large WWM 版) の結果を比較すると、NICT BERT<sub>BASE</sub> (日本語 Wikipedia, BPE 無) はモデルサイズが小さくパラメタ数が約半分であるにもかかわらず、日本語 Pretrained モデル (Large WWM 版) よりも高い性能を得ていることがわかる。この結果から、単純にモデルサイズを大きくして巨大なニューラルネットにするだけで最終的な性能向上につながるわけではなく、そのモデルサイズに適切な事前学習時の設定(バッチサイズやステップ数等)も合わせて適切に調整する必要があり、高品質な大規模言語モデルを作成することは容易でないことがわかる。

また、語彙数に関しては表 2 を見てわかるように単

一言語に関して BPE などを利用せずに 100k (=10 万語) という巨大な語彙数で事前学習を実施しているのは我々だけであるが、表 3 を見る限り、解答可能性付き読解以外の問題では全て語彙数 320k+BPE よりも語彙数 100k のモデルのほうが高い性能を得ている。BPE を用いた場合の性能低下の要因としては、語を過剰に BPE 等で分割すると、その短く分割された語の組み合わせでもとの語の意味を学習させる必要がでてきて、その分事前学習の問題が難しくなったのだと考えられる。一方で、解答可能性付き読解の場合は固有名を回答とする問題が多いため、回答となる語が事前学習時に 100k の語彙に含まれずに学習できない、もしくは語彙に含まれたとしても出現回数が少なく十分に学習できていないという問題がある。このため、そのような低頻度語の扱いについては今後も検討を続けていく必要がある。

最後に、追加で実施した NICT BERT<sub>LARGE</sub> の結果については他の BERT モデルと比較してより高い性能を得ており、学習の規模(高バッチサイズ、十分な量

\*4 <https://github.com/google-research/bert>

\*5 [https://nlp.ist.i.kyoto-u.ac.jp/index.php?ku\\_bert\\_japanese](https://nlp.ist.i.kyoto-u.ac.jp/index.php?ku_bert_japanese)

\*6 <https://github.com/cl-tohoku/bert-japanese>

\*7 <https://github.com/yoheikikuta/bert-japanese>

\*8 [https://www.hottolink.co.jp/blog/20190311\\_101674/](https://www.hottolink.co.jp/blog/20190311_101674/)

のステップ数)を保ちつつ、モデルサイズをより大きくし、かつ、学習に利用するテキストを選別しつつもより大規模にすることが重要であることがわかる。

また、因果関係認識については、BERTの事前学習の性質を調査するために、追加で入力を変更した場合についても実験を行った。表1に示すように、オリジナルの問題設定では因果関係の候補となる句の対が出現している文も含めて入力とし、分類対象となる句の対とそれ以外の単語の位置を追加のembeddingで区別するというを行っているが、この入力に加えて、因果関係の背景知識に相当する入力も加えた実験も実施した。ここでいう背景知識とは、例えば「地球温暖化が進む→海水温が上昇する」のような因果関係を考えた場合の「地球温暖化」と「海水温」をつなぐ「地球温暖化で海水温が上がる」のような言語パターンや「地球温暖化」と「海水温」と「ため」のような手がかり表現を含む1文等の入力された因果関係(の候補)を捉えるための付加的な情報をいう。我々の過去の研究[20][21]では、このような背景知識を収集し、SVM[6]と呼ばれる深層学習以前の機械学習アルゴリズム、もしくは、convolutional neural network (CNN)と呼ばれる事前学習を利用しないニューラルネットの追加の入力とすることで性能が向上することを確認している。今回の実験でもその実験と同様に、入力の1文全体に背景知識(言語パターンと1文の検索結果)を単純につなぐことで1つの長い入力単語列を作成し、それをBERTへ入力して実験を行った。この実験でNICT BERT<sub>LARGE</sub>を用いた場合の結果を表4に示すが、この結果からBERTへ背景知識を入力した場合には性能が低下することがわかる。

また、上記の実験に加え、1文全体の入力のうち冗長だと思われる箇所を除外してBERTへ入力する実験も実施した。具体的には、1文全体を入力するのではなく、「地球温暖化が進む [SEP] 海水温が上昇する」のような、因果関係認識で着目している原因候補の句と帰結候補の句の対のみを入力とし、NICT BERT<sub>LARGE</sub>を用いて実験を行った。この実験結果も表4に示すが、1文全体を入力した場合と比較して句の対だけを入力した場合のほうがより性能が良くなっていることがわかる。深層学習、特に事前学習が流行する以前のアプローチでは、背景知識のような入力テキストに書かれ

ていない情報や1文全体から得られる情報等をいかに効果的に処理に取り込むかが重要な研究の観点であり、因果関係認識の既存研究[20][21]でも実際にそのような処理を導入することで性能が向上していたが、表4の結果そのようなことを考慮せずとも問題によっては高い性能が得られることを示しており、これはつまり人間の内省に基づいて考案していた付加的な背景知識等の情報でさえ事前学習によって学習されてしまっているかのように見える。

一方で、本特集号3-3[25]では、「なぜ」型質問応答の回答パッセージ特定においてより洗練された背景知識の利用に関する手法[26][27]が紹介されており、この手法を用いた場合にはBERTを利用した場合であっても背景知識を利用することで性能が向上することが報告されている。この手法では、単純に背景知識に相当する単語をBERTの入力に加えるのではなく、あるCNNをBERTとは独立に事前学習し、入力に対して背景知識に相当するベクトルを出力できるようにしておき、その出力をBERTの出力と統合して利用するということが行われており、そのような処理を行った場合には少なくとも回答パッセージ特定においては背景知識が性能向上に貢献している。このようにBERTでの背景知識の利用一つとっても技術的に深掘りする余地があり、我々は今後もそのような観点からも研究を続けていく予定である。

また、上記のBERTの品質向上に関する調査の更なる発展版として、言語モデルをより巨大にすることによる性能限界の調査を現在実施している。具体的にはRaNNC[12]を用いて200億パラメタというより巨大なBERT(NICT BERT<sub>LARGE</sub>のパラメタ数約4億の50倍)の事前学習を進めている。現時点の学習段階のモデルをファインチューニングした場合であっても上述の因果関係認識でNICT BERT<sub>LARGE</sub>と比較して性能改善が見られ、言語モデルの更なる巨大化によって更に性能が向上する見込みがあることが確認できている。ただし、ありとあらゆる言語的な知識が巨大モデルに蓄積できる可能性は低いと考えており、大規模言語モデルと他の知識の効果的な併用等も今後の研究課題となる。

### 3 生成問題における NICT BERT の利用

NICT DIRECTでは、様々な話題に対してユーザが関心を持ちそうな情報を適用し、ユーザの知識を深め、知的視野を広げることを目的とした「雑談」を行う音声対話システムWEKDA[7]を研究開発している。また、その「雑談」の機能はマルチモーダル音声対話システムMICSUSでも利用されている。(WEKDAとMICSUS

表4 因果関係認識：入力フォーマットの違いによる性能評価

入力フォーマット	F1	平均精度
1文全体	64.84	70.44
1文全体 + 背景知識 [21]	63.55	69.55
原因、帰結候補の句の対	67.25	74.44



の詳細については本特集号 3-5 [13] を参照。) WEKDA や MICSUS で雑談を実現する基本的な原理は、入力されたユーザ発話を質問の形式に自動で変換し、その質問を WISDOM X に入力して得られた回答を整形してシステムの応答とするというものである。例えば、「介護施設でロボットは使われるのかな」というユーザ発話を「介護施設でロボットを使うとどうなる」や「介護施設でロボットを何に使う」のような質問の形式に変換し、その質問を WISDOM X に入力して得られた回答を利用して応答を作成することで「介護ロボットが介護施設に普及が進めば離職率が下がりますよ」といった応答を出力することができる。このように質問応答を経由して応答を作成する一連のプロセスは「ユーザがUというユーザ発話をしたので、Qという質問の回答を教えてあげれば今後の行動等で有用もしくは参考になる可能性があるため、その回答を含む応答をする」という「応答の理由」に基づいた応答生成とみなすことができ、WEKDA や MICSUS による雑談はまさにその「応答の理由」を考慮した応答生成の処理だといえる。また、「介護施設でロボットを使うとどうなる」のような質問の回答から応答を作成することで、「介護施設でロボットを使う」ことに関する将来のチャンスやリスク等に関する情報を提供できる雑談につながる。このように質問応答を経由した雑談の作成はユーザにとって興味深い応答を返すことにつながる。期待できるが、質問応答の出力形式と雑談の応答として出力したい形式に齟齬があるため、そのまま質問応答の出力を無加工で雑談の応答に利用することはできない。例えば、WISDOM X を用いた質問応答では知識源である Web テキストに書かれた内容をそのまま回答として出力するため、それをそのまま雑談の応答に利用すると冗長で不自然な発話になってしまう。そこで、回答を含む文から発話として利用できる部分を抜き出し、自然な文として整形するといった生成の処理が必要となる。

また、WEKDA や MICSUS では雑談だけでなく「なに」「どうやって」「なぜ」といった質問を行い、WISDOM X 経由で回答を得ることが可能である。これは、雑談でシステム応答として提示された内容を質問応答で深掘りして知識を得る場合に有益であり、例えば、上の発話例の場合は「離職率が下がる」という情報が提示された後で、ユーザが「なぜ介護施設でロボットを使うと離職率が下がるのか？」のような質問をすることで、更に深掘りして知識を得ることができる。ただし、この場合も WISDOM X が提示する回答は、「なぜ」や「どうやって」といった質問に対しては表 5 の回答パッセージに示すような回答箇所を含む比較的長いテキストを、一定の表現は省略されるものの出力するよ

表 5 「なぜ」型質問応答の回答要約の具体例 (作例)

質問: なぜ Google の機械翻訳の質は劇的に向上したのか。

回答パッセージ: 深層学習の技術は近年注目を集めており、高度な技術を持った企業が最新の深層学習の研究成果を各種サービスに統合することに対して意欲的だった。例えば、Google Brain チームは、そのような技術を Google の機械翻訳エンジンに組み込む技術を開発した。これにより、フランス語や中国語などの言語に関して機械翻訳の質は劇的に向上し、他の多くの言語についても今後同じように質が向上することが他所される。どの程度になるのかは不明だが、深層学習の技術によって今後我々の未来が急激に変わるというのは間違いのないようだ。

回答要約: 深層学習技術が導入されたため

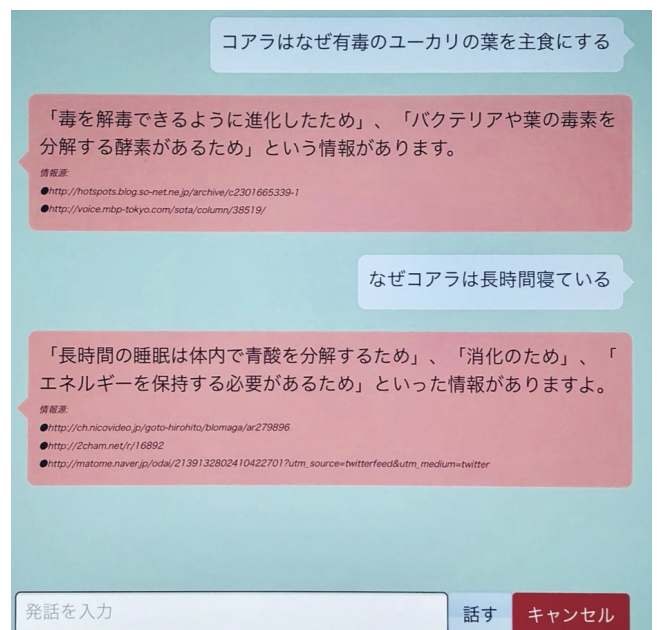


図 2 WEKDA における回答要約の実行例

うに設計されているが、そのままではシステムの応答として利用することができない。そこで、この場合も回答箇所を含み、かつ、コンパクトで自然な回答を生成するといった回答要約の処理が必要となる。実際の WEKDA の出力例を図 2 に示すが、このようにコンパクトに回答を要約することで、音声対話システムで音声のみで回答が提示された場合でも回答が容易に理解可能となる。

本稿では上述の雑談発話生成と回答要約の 2 種類の生成課題のうち、回答要約について紹介する。

### 3.1 「なぜ」型質問応答の回答要約

DIRECT における生成処理の一例として、表 5 の具体例を用いて「なぜ」型質問応答の回答要約の処理を説明する。回答要約の入力は「なぜ Google の機械翻訳の質は劇的に向上したのか」のようなユーザが入力した

質問と、その質問に対してDIRECTで開発された最新のノンファクトイド型の質問応答の回答パッセージ特定技術 [26][27] 等を利用して特定された、回答が含まれる可能性が高い、連続した文から成る回答パッセージである(この回答パッセージ特定技術については本特集号 3-3 [25] を参照)。回答要約の課題はそのような質問と回答パッセージが入力として与えられた際に、「深層学習技術が導入されたため」のような、入力質問の回答に相当する簡潔で自然な日本語の文を出力するというものである。この回答要約では音声対話システム等で読み上げられることを想定して、最大でも25文字というコンパクトな要約を出力するという課題設計となっている。また、別の特徴としては、パッセージ内から内容語をコピーし、それを自然な回答となるように整形するという課題になっている。そういった意味では、パッセージに無い内容語を生成するという難しい問題は含まないが、一方で、入力の質問に関連する部分を特定しなければならないという難しさを含む。

この回答要約のためにDIRECTでは独自に学習・評価用データを作成し [28][29]、回答要約技術の研究開発を行った。研究の初期段階(2018年5月頃)では、Pointer-Generator [30] と呼ばれる、RNN Encoder Decoder ネットワーク [14] で、入力から必要に応じて単語列のコピーも行えるニューラルネットを利用して回答要約を行った [28][29]。

その後、生成処理のためにも事前学習を実施した。具体的には、事前学習済みのBERTに対して更に生成用の事前学習を実施する UniLM [15] の事前学習手法を採用し、NICT BERT<sub>LARGE</sub> に対して生成用の追加学習を実施して、日本語版のUniLM モデルを作成した(2020年の3月に事前学習が完了)。このモデルに対して更に前述の回答要約に関する学習データでファインチューニングを行い、学習の結果得られた生成モデルを利用して「なぜ」型質問応答の回答要約の評価を行った。

結果を表6にまとめる。回答要約の性能評価には ROUGE スコア [31] と呼ばれる正解単語列との単語の一致率に基づく評価尺度で評価を行っている。表中の ROUGE-1、ROUGE-2、ROUGE-L はそれぞれ単語 unigram、bi-gram、最長一致に基づいた評価結果を表す。表6の結果から、以前の Pointer-Generator ベースの手

法 [29] と比較して UniLM を用いることでより質の高い回答要約を生成できることがわかる。

UniLM を用いた生成処理は、BERT、つまり、巨大な Transformer のエンコーダだけで構成されたニューラルネットで入力全体と出力された単語全体を生成の各ステップで全てエンコードするという負荷の高い処理を行うが、それは処理速度の面で問題となるため、効率的な生成のためにアーキテクチャの改善を行った。具体的には、通常の encoder と decoder に分かれた Transformer アーキテクチャ [1] を利用し、Rothe ら [16] と同様に、事前学習済みの BERT のパラメータを encoder と decoder の初期パラメータとし<sup>\*9</sup>、生成用の学習を行った。ただし、Rothe らの場合は BERT のパラメータで初期化後すぐに本番の生成課題のファインチューニングを行っているが、我々の手法(以降、これを BERT EncDec と呼ぶ)では、NICT BERT<sub>LARGE</sub> のパラメータで encoder と decoder を初期化した後に、さらに生成用の追加学習として Google T5 [32] と同様の encoder と decoder をまたぐ masked LM に相当する学習(学習データの具体例は図3を参照)を Web から取得したテキストを使って約1億事例分学習するというを行っている。これにより、モデルのパラメータをより生成に適したパラメータに更新した上で、本番生成課題のファインチューニングが可能となる。また、decoder の層数も encoder と同じ24層にする代わりに、1層に変更することで更に高速化を図っている。(生成用追加学習データとして1億事例を利用し、decoder の層を1層にした場合の学習は2021

\*9 decoder の cross-attention 層は BERT には含まれないため、その層のパラメータはランダムに初期化している。

**入力:** 深層学習 [MASK1] は近年注目 [MASK2] 集めており、高度な技術を持った企業が最新の [MASK3] 研究成果を各種サービスに統合することに [MASK4] で意欲的である。  
**出力:** [MASK1] の技術 [MASK2] を [MASK3] 深層学習の [MASK4] 対し

図3 BERT EncDec の生成用追加学習の追加事例の具体例(作例)

表6 「なぜ」型質問応答の回答要約の評価結果

手法	ROUGE-1	ROUGE-2	ROUGE-L
Pointer-Generator ベースの手法 [29]	55.15	40.94	52.42
UniLM <sub>LARGE</sub>	63.29	50.81	62.53
BERT <sub>LARGE</sub> EncDec (enc=24層 / dec=24層)	<b>64.45</b>	<b>51.99</b>	<b>63.74</b>
BERT <sub>LARGE</sub> EncDec (enc=24層 / dec=1層)	63.81	51.23	63.18



年11月に完了し、以降はその生成モデルを利用した各種生成器の研究開発をDIRECTで行っている。

このようにして作成したBERT EncDec(decoderが24層の場合と1層の場合)を「なぜ」型質問応答の回答要約で評価した結果も表6にまとめるが、この結果より、decoderが24層の場合のBERT EncDecはUniLMよりも高い性能を得ている。また、decoderが1層の場合は24層の場合よりも若干性能が低下するものの、性能をほとんど損なわずに回答要約を生成できていることがわかる。実行速度に関する評価としては、UniLMで回答要約を生成した場合と比較して、decoderが24層のBERT EncDecの場合は約10倍<sup>\*10</sup>、1層の場合は約41倍の処理速度に改善できていることを確認しており、この速度の改善はWEKDA等の実アプリケーションで生成処理を行う上で非常に有益な結果であるといえる。

## 4 WISDOM X の軽量化／高速化

前述したように、NICT DIRECTでは大規模情報分析システムWISDOM Xの研究開発、ネット上の一般公開を行っており、このWISDOM Xを用いることで大規模なwebテキストを知識源として「なに」「どうなる」「どうやって」「なぜ」といった質問に回答することが可能である。さらに、WISDOM Xの質問応答技術は対話システムWEKDA [7]やマルチモーダル音声対話システムMICSUSで雑談を行う際の応答候補の生成等にも利用されており、その応用範囲も広い。一方で、そのような応用処理の実用を見越した処理の軽量化、高速化が重要な課題となっていた。特に処理のボトルネックとなっていたのが、入力の問題に対する回答候補パッセージ(web文書から抽出された連続7文から成るテキスト)のキーワード検索及び検索結果に深層学習を適用して、適切な回答を含むパッセージを特定する処理で、検索対象となるパッセージの規模

が大きくなり、また、より大量のパッセージに深層学習を適用する必要があるにつれて、その実行速度が問題となっていた。このような背景から、検索の効率化等の処理の改善に取り組んだ。

具体的には、SentenceBERT [33] (図4) と呼ばれる、2つの入力単語列(ここでは、質問とパッセージ)の類似度計算を行うニューラルネットを利用し、検索対象となるパッセージ集合のクラスタリング処理、さらに、そのクラスタリング結果を受けて、入力の問題に対して回答が含まれている可能性の高いクラスタを特定する処理を実現することで、全パッセージ集合を検索するのではなく、事前に絞り込まれたより小規模のパッセージ集合を検索するというを行い、処理の高速化を行った。さらに質問応答の回答パッセージ特定等に利用するGPUの割り当て等も見直すことにより、従来の処理と比較して、GPUの枚数を1/40に削減しつつ、処理速度を低下させることなく従来の1.3倍の速度で処理が可能となっている。これにより今後のWISDOM Xの技術展開で障害となるGPUリソースに関するハードルを劇的に下げることができたといえる。

## 5 まとめ

本稿では、NICT DIRECTで進めている深層学習を用いた自然言語処理について、テキストの分類処理と生成処理の観点からのそれぞれの取組を紹介した。現在の自然言語処理分野では深層学習の利用が広く浸透しているが、この中でも質の高い大規模言語モデル自体の開発並びにその言語モデルを利用した質の高いシステム開発がより重要になってきている。このような観点からの取組の一例として、公開版NICT BERT<sub>BASE</sub>並びにNICT BERT<sub>LARGE</sub>(非公開)の事前学習の概要と、因果関係認識等の4種類の問題に関する評価結果を示した。BERTを用いた分類処理の結果はそれが導入される以前では考えられないような高い性能を示しており、我々は今後もこのような大規模言語モデルを積極的に活用していくべきであることがわかる。一方で、モデルが巨大化することにとまらぬ、処理速度の面で問題が生じるが、本稿ではその問題を解決する例として、生成用ニューラルネットのアーキテクチャの改良による処理の高速化、さらに、WISDOM Xにおける軽量化／高速化について紹介をした。今後は、このような処理の効率化、高速化を実現しつつ、より質の高い技術や実アプリケーションの実現に向け、本稿

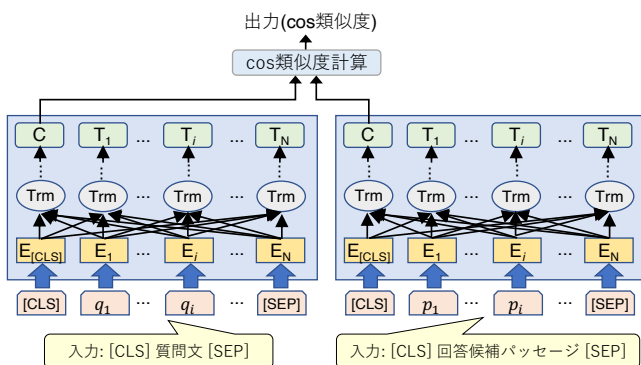


図4 SentenceBERTのアーキテクチャ

\*10 回答要約の場合は出力する単語列に比べて入力が質問と回答パッセージという長い入力を与えられているため、それらを生成の各ステップで全てエンコードすると非常に処理速度が遅くなってしまふ。

で紹介した深層学習技術やそれらを RaNNC [12] で更に大規模にしたより高品質な言語モデルを構築し、また、これまでに構築した言語リソース等も活用して、より興味深く、かつ、社会で役立つシステムの研究開発に取り組んでいきたいと考えている。

### 【参考文献】

- 1 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," CoRR, vol. abs/1706.03762, 2017.
- 2 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 (NAACL-HLT 2019), pp.4171-4186, 2019.
- 3 大竹 清敬, "社会知コミュニケーション技術の概要," 情報通信研究機構研究報告, 本特集号, 3-1, 2022.
- 4 T.B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D.M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," CoRR, vol. abs/2005.14165, 2020.
- 5 B. Kim, H. Kim, S. Lee, G. Lee, D. Kwak, D.H. Jeon, S. Park, S. Kim, S. Kim, D. Seo, H. Lee, M. Jeong, S. Lee, M. Kim, S. Ko, S. Kim, T. Park, J. Kim, S. Kang, N. Ryu, K.M. Yoo, M. Chang, S. Suh, S. In, J. Park, K. Kim, H. Kim, J. Jeong, Y.G. Yeo, D. Ham, D. Park, M.Y. Lee, J. Kang, I. Kang, J. Ha, W. Park, and N. Sung, "What changes can large-scale language models bring? intensive study on HyperCLOVA: Billions-scale korean generative pretrained transformers," CoRR, vol. abs/2109.04650, 2021.
- 6 V.N. Vapnik, The Nature of Statistical Learning Theory, Springer, 1995.
- 7 水野 淳太, クロエツェー ジュリアン, 田仲 正弘, 飯田 龍, 呉 鍾勲, 石田 諒, 浅尾 仁彦, 福原 裕一, 藤原 一毅, 大西 可奈子, 阿部 憲幸, 大竹 清敬, 鳥澤 健太郎, "WEKDA:web40 億ページを知識源とする質問応答システムを用いた博学対話システム," 第 84 回言語・音声理解と対話処理研究会資料, pp.135-142, 2018.
- 8 Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," CoRR, vol. abs/1907.11692, 2019.
- 9 Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "ALBERT: A lite BERT for self-supervised learning of language representations," Proceedings of the 8th International Conference on Learning Representations (ICLR 2020), 2020.
- 10 K. Clark, M. Luong, Q.V. Le, and C.D. Manning, "ELECTRA: pre-training text encoders as discriminators rather than generators," Proceedings of the 8th International Conference on Learning Representations (ICLR 2020), 2020.
- 11 P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: decoding-enhanced BERT with disentangled attention," Proceedings of the 9th International Conference on Learning Representations (ICLR 2021), 2021.
- 12 M. Tanaka, K. Taura, T. Hanawa, and K. Torisawa, "Automatic graph partitioning for very large-scale deep learning," Proceedings of 35st IEEE International Parallel and Distributed Processing Symposium (IPDPS 2021), pp.1004-1013, 2021.
- 13 水野 淳太, 浅尾 仁彦, "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.
- 14 T. Luong, H. Pham, and C.D. Manning, "Effective approaches to attention-based neural machine translation," Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), pp.1412-1421, 2015.
- 15 L. Dong, N. Yang, W. Wang, F. Wei, X. Liu, Y. Wang, J. Gao, M. Zhou, and H.-W. Hon, "Unified language model pre-training for natural language understanding and generation," Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), 2019.
- 16 S. Rothe, S. Narayan, and A. Severyn, "Leveraging pre-trained checkpoints for sequence generation tasks," Transactions of the Association for Computational Linguistics, vol.8, pp.264-280, 2020.
- 17 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.
- 18 K. Kadowaki, R. Iida, K. Torisawa, J.-H. Oh, and J. Kloeetzer, "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.
- 19 関 直哉, 水野 淳太, 門脇 一真, 飯田 龍, 鳥澤 健太郎, "ファクトイド質問応答における BERT の pre-trained モデルの影響の分析," 言語処理学会第 26 回年次大会 発表予稿集 (NLP2020), pp.105-108, 2020.
- 20 C. Hashimoto, K. Torisawa, J. Kloeetzer, 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.
- 21 C. Kruengkrai, K. Torisawa, C. Hashimoto, J. Kloeetzer, J.-H. Oh, and M. Tanaka, "Improving event causality recognition with multiple background knowledge sources using multi-column convolutional neural networks," Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI 2017), pp.3466-3473, 2017.
- 22 J. Kloeetzer, K. Torisawa, C. Hashimoto, and J.-H. Oh, "Large-scale acquisition of entailment pattern pairs by exploiting transitivity," Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), pp.1649-1655, 2015.
- 23 二宮 大空, 門脇 一真, 飯田 龍, 鳥澤 健太郎, J. Kloeetzer, "BERT を用いたバイナリパターン間の含意関係認識," 言語処理学会第 26 回年次大会 発表予稿集 (NLP2020), pp.1423-1426, 2020.
- 24 鈴木 正敏, 松田 耕史, 岡崎 直観, 乾 健太郎, "読解による解答可能性を付与した質問応答データセットの構築," 言語処理学会第 24 回年次大会 発表予稿集 (NLP2018), pp.702-705, 2018.
- 25 呉 鍾勲, クロエツェー ジュリアン, "大規模 Web 情報分析システム WISDOM X 深層学習版," 情報通信研究機構研究報告, 本特集号, 3-3, 2022.
- 26 J.-H. Oh, K. Kadowaki, J. Kloeetzer, 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.
- 27 J.-H. Oh, R. Iida, J. Kloeetzer, 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.
- 28 R. Ishida, K. Torisawa, J.-H. Oh, R. Iida, C. Kruengkrai, and J. Kloeetzer, "Semi-distantly supervised neural model for generating compact answers to open-domain why questions," Proceedings of 32nd AAAI Conference on Artificial Intelligence (AAAI 2018), pp.5803-5811, 2018.
- 29 R. Iida, C. Kruengkrai, R. Ishida, K. Torisawa, J.-H. Oh, and J. Kloeetzer, "Exploiting background knowledge in compact answer generation for why-questions," Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019), pp.142-151, 2019.
- 30 A. See, P.J. Liu, and C.D. Manning, "Get to the point: Summarization with pointer-generator networks," Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), pp.1073-1083, 2017.
- 31 C.-Y. Lin, "ROUGE: A package for automatic evaluation of summaries," Proceedings of the ACL-04 Workshop on Text Summarization Branches Out, pp.74-81, 2004.
- 32 C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P.J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," Journal of Machine Learning Research, vol.21, no.140, pp.1-67, 2020.
- 33 N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP 2019), pp.3982-3992, 2019.

**飯田 龍** (いいだ りゅう)

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(工学)  
自然言語処理

【受賞歴】

- 2014年 言語処理学会 20周年記念論文賞
- 2007年 情報処理学会 2007年度(平成19年度)  
山下記念研究賞
- 2006年 COLING/ACL 2006 Asian  
Federation of Natural Language  
Processing, Best Asian NLP Paper  
Award