

## 3 社会知コミュニケーション技術

### 3 *Data-driven Intelligent Communication Technology*

#### 3-1 社会知コミュニケーション技術の概要

##### 3-1 *Data-driven Intelligent Communication Technologies at NICT*

大竹 清敬

OHTAKE Kiyonori

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、2021 年度から 2025 年度の第 5 期中長期目標期間に、インターネット等社会に自然言語で書かれたテキストとして広く存在する知識 (社会知) の有効活用を可能にする社会知コミュニケーション技術の研究開発に取り組んでいる。そうした技術の例としてはユーザに寄り添った音声対話を行いつつ、有用な知識や情報をユーザとやり取りできる音声対話システムがあり、また、大量の Web ページ等の情報の深い意味的分析に基づき有用な情報を容易に取得することを可能とする質問応答システム、情報分析システム等がある。本稿ではまず、近年、関連する AI やその一種である自然言語処理の研究開発で、中心的役割を担っている深層学習技術について簡単に述べ、本特集で紹介する社会知コミュニケーション技術の各技術の概要、その狙いについて述べる。

During the medium- to-long-term plan period from FY 2021 to FY 2025, the Data-driven Intelligent System Research Center (DIRECT), NICT is engaged in research and development of data-driven intelligent communication technologies that effectively use Internet's "social wisdom", i.e., the huge amount of knowledge that is produced by society and is ubiquitous in the form of natural language texts on the Internet. Examples of such technologies include a dialogue system that can exchange useful knowledge and information with users through user-friendly dialogues, as well as question answering systems that can easily and instantly discover valuable information from a vast amount of Web pages. In recent years, deep learning has played a central role in research and development in the fields of AI and natural language processing. This paper overviews the situation surrounding deep learning technologies and introduces the novel systems and technologies we developed as part of data-driven intelligent communication technologies.

#### 1 はじめに

NICT では、これまで 10 年以上にわたってインターネット等に存在する膨大な知識を活用するための技術の研究開発に取り組んでいる。特に 2016 年度から 2020 年度までの第 4 期中長期目標期間においては、DIRECT において我々が「社会知」と呼ぶ、インターネット等にテキストとして広く存在する知識を意味的に深く解析し活用する技術、すなわち「社会知解析技術」の研究開発及びその社会実装に取り組み、後ほど紹介するように、いくつかのシステムをインターネット上で試験公開しているほか、一部の技術については民間企業へのライセンスも実施している。2021 年度か

らの第 5 期中長期目標期間においては、社会知解析技術を発展させ、社会知を活用しつつユーザに寄り添った対話を行い、ユーザに有用な社会知を提供したり、ユーザから必要な情報を取得したりする音声対話技術、社会知を表現している膨大なテキストを深く分析し、有用な知識、情報を容易に取得可能とする質問応答技術、やはり有用な知識を膨大なテキストより抽出し、そうした知識を用いて推論を行い、テキストに書かれていない仮説・知識を得る技術等の研究開発に取り組んでおり、そうした一連の技術を社会知コミュニケーション技術と称している。

これらの技術は、最終的には、日本社会において深刻化している具体的な社会課題の解決に資することを

目的として研究開発を進めている。また、より長期的には、我々はそうした技術を発展、融合させることで、2030年ごろには、膨大な社会知を活用し、多様な話題に関して人間とブレインストーミングを行い、適切な意思決定やイノベーションを促進することができる対話システムの開発を目指している。本稿では、この社会知コミュニケーション技術の概要やその狙いについて説明する。

## 2 近年のAI、自然言語処理における深層学習の動向

社会知コミュニケーション技術の根幹は、人工知能(AI)に関する技術のうち人間が用いる言葉やその意味を知的に処理する自然言語処理技術と呼ばれる技術である。近年広く知られるようになった深層学習(deep learning: ディープラーニング)と呼ばれる技術がAIの様々な課題に対してこれまでの最高精度を更新する結果を出すようになってから、深層学習はその技術を支えるハードウェアの発展とあわせてAI研究の必要不可欠な要素となった。自然言語処理技術においても、こうした深層学習の発展の恩恵を受けている。代表的な事例が、Googleが発表した巨大ニューラルネットワークBERT [1]であるが、これは自然言語処理における様々なタスクで過去最高の精度を達成し、その有用性を世に知らしめた。その後、様々な亜種も提案され、現在は、自然言語処理の分野において非常に良く用いられるニューラルネットワークの一つとなっている。

BERTの特徴は、事前学習とファインチューニングという二段階の学習を行うことである。従来のニューラルネットワークを用いた自然言語処理では、質問応答等、特定のタスクのための学習データ(教師データ)を使って学習(教師あり学習)を行い、そのタスクだけを実行できるニューラルネットワークを構築していた。この際、学習データは通常、人間の作業者が作成していた。例えば、質問応答を行うニューラルネットワークを開発したければ、作業者が大量の質問と、あえて誤った回答も含め、その回答の候補を作文し、質問とそれら回答候補をペアにした上で、回答が質問の適切な回答になっているかどうかをラベルとしてペアに付与し、それを質問応答のための学習データとして利用していた。一方、BERTの一段階目の学習、すなわち、事前学習では、インターネット等にある膨大なテキストデータを用いて、上記のような人間による学習データ作成の作業をせずに学習(教師なし学習と呼ぶ)を行う。

BERTの事前学習では、典型的には、膨大なテキストデータの単語を一部削除して、いわゆる「穴埋め問題」を大量に作成し、その穴埋め問題で正解となる削

除された単語を当てるというタスクで学習を行う。こうした穴埋め問題は完全に自動で作成できるため、膨大なテラバイト単位のテキストデータであっても学習で活用できるわけである。一見、こうした穴埋め問題で学習すると一体何が学習できるのか訝しく思われるかもしれないが、BERTのようなニューラルネットワークはこうした穴埋め問題を介して、言語の多岐にわたる性質、構造等を学習できることがわかっており、また、そうした性質、構造等をあらかじめニューラルネットワークが持っていることによって、その後行われる二段階目の学習、つまり、ファインチューニングの効果が高められ、最終的に高い精度が達成できることになる。

こうした事前学習済みのニューラルネットワークは、非常に大雑把な言い方ではあるが、言語の性質を学習済みであることから言語モデルもしくは、大規模なニューラルネットワークであることから大規模言語モデルと呼ばれることもある。第二段階の学習であるファインチューニングでは、質問応答等の個別のタスクごとに(典型的には人間の作業者がテキストデータへのラベル付等の作業をして)用意した学習データを用いて、事前学習済みのニューラルネットワークに対し追加で学習を行う。

また、一般により大量のデータでより大量のパラメータをもつ巨大ニューラルネットワークで事前学習をすれば、ファインチューニング用の学習データが少なくとも高い精度を得られると言われている。これは、同等の精度を達成するのに、人間による学習データ作成作業を減らすことができるということであり、この考えをある意味極限まで推し進めたのが、BERTの514倍<sup>\*1</sup>のパラメータをもつ、GPT-3 [5]のような超巨大なニューラルネットワーク、超大規模言語モデルである。これらの超大規模言語モデルは、人間が作成した大量の学習データなしで、タスクの説明のみ(Zero-shot)やタスクの説明と少量のデモンストレーション<sup>\*2</sup>(Few-shot)を与えてタスクを実行することが可能であり、タスクによっては極めて高い精度を達成している。また、これらの超大規模言語モデルの出力はテキストであり、なかには人が書いたとの区別がつかないような高品質なテキストが出力されるといった事例も多々報告されている。

様々なタスクをそのための大量の学習データ(教師データ)なしにこなすことが可能な万能モデルという

\*1 BERT<sub>LARGE</sub>は3.4億パラメータ、GPT-3は1,750億パラメータを有する[4]。

\*2 GPT-3のZero-shot/Few-shotという用語は一般的な機械学習で用いられる用語とは異なり、GPT-3に与えるデモンストレーションの量で区別して用いられる用語である。なお、GPT-3では、Few-shotとしてデモンストレーションが与えられてもモデルのパラメータ更新は行われない。

のは、夢の技術であるといえるが、少なくとも現段階での社会実装は容易ではないと思われる。一つは、コストの問題である。超大規模言語モデルは、学習時のみならず、それを用いてユーザのリクエストに応じて様々なタスクを行う際にも大量の計算資源を必要とし、それを賄うだけの料金に見合ったサービスをユーザに提供することは現状難しいと考えている。もう一つは、倫理的な問題である。超大規模言語モデルは学習データなしで様々なタスクに応用することが可能であり、タスクの中には人間と同等、あるいは場合によっては人間をしのぐ精度を示すものもある [5] が、これは裏を返せば、開発者のコントロールが効きにくいということでもある。この結果、大規模言語モデルを対話システム等で用いた場合に、ヘイトスピーチやヘイトスピーチとまで行かなくても、ユーザを不快にさせる、いわば失言のようなシステム応答が出力されないことを保証することは難しく、また、仮にそうした応答をシステムがした場合に、誰が責任を負うのか合理的に決定することが難しい。もちろん、事前学習において学習データとして使われたテキストが原因となって不適切なシステム応答が出力されたことは間違いないのだが、特定のテキストの断片に原因を帰するのは、超大規模言語モデルの複雑さからほぼ不可能であろう。このような場合、責任を負うのは超大規模言語モデルをサービスで活用している運用者となる他なく、運用者はリスクを負うことになる。今後、何らかの新規技術でこうした課題が解決される可能性もあるが、現時点で超大規模言語モデルを社会実装するには、その使い方や応用領域等に関してよく考え抜く必要がある。

### 3 NICTにおける社会知コミュニケーション技術の研究開発

NICTでは、これまで10年以上にわたって日本語のWebページを研究開発用に収集してきており、現在、約400億ページの日本語Webページを収集済みである。さらに、現実的な社会課題の解決に資する自然言語処理を実現する上では、非常に多岐にわたる様々なタスクを処理する必要があるが、そうしたタスクをAIで処理するための学習データを柔軟に構築できる人間の作業者のチーム及びそのチームを指導する言語学者から成る体制を整備し、質問応答や対話に関する高品質かつ大量な学習データを整備、蓄積してきている。また、収集したWebページでBERT等の巨大ニューラルネットの事前学習を行い、それらを上述の高品質かつ大量な学習データでファインチューニングを行い、高品質な対話システム、質問応答システム等を開発しており、一部の技術に関しては、インター

ネット上での試験公開やビジネス化を意識した実証実験等も行っている。こうした大量の学習データの構築は、学習データを必要としないGPT-3のような超大規模言語モデルの研究開発のような最近の動向に逆行しているのではないかと、と思われるかもしれないが、前述したように、運用で必要となる計算リソースや出力の適切性の二点において、そうした超大規模言語モデルがすぐに社会実装に結びつくとは考えづらく、GPT-3ほどのサイズではないがそれなりの規模の大規模言語モデルを高品質かつ大量の学習データでファインチューニングし、リーズナブルなコストで運用できる技術やそれらを活用したシステムの方が普及や社会課題の解決に結びつきやすいと考えている。一方で、なんらかのゲームチェンジャーとなる技術によって、超大規模言語モデルが一気に実用化されるような事態にも備えて、そうした超大規模言語モデルを容易に開発可能とするミドルウェアの開発も行っている。

本特集の一連の研究報告では、社会知コミュニケーション技術の研究開発の内容を紹介する。**3-2** 自動並列化深層学習ミドルウェア RaNNC (ランク) では、超大規模言語モデルの開発を容易にするために大規模なニューラルネットを自動で分割し、複数のGPUを活用して高速にその学習を行えるミドルウェア RaNNCの研究開発について述べる [6]。**3-3** 大規模 Web 情報分析システム WISDOM X 深層学習版では、インターネットに存在する膨大な社会知を活用すべく、誰もが容易にそういった知識にアクセスできるよう、第3期中長期目標期間(2011年度～2015年度)に研究開発を行った大規模 Web 情報分析システム WISDOM X (ウィズダムエックス) を最新の深層学習技術でより高精度にし、現在インターネット上で試験公開している<sup>\*3</sup> 質問応答システムについて説明する [7]。**3-4** DIRECT における深層学習を用いた大規模自然言語処理では、現在、WISDOM X 深層学習版やこの後に紹介する音声対話システムで実際に活用されている BERT 関連の一連の技術を紹介する [8]。**3-5** 高齢者支援用マルチモーダル音声対話システム MICSUS (ミクス) では、高齢化社会における課題解決を目的として内閣府の SIP 第2期にて研究開発を進めてきた音声対話システム MICSUS について説明する [9]。

## 4 むすび

本稿では、NICTで取り組んでいる社会知コミュニケーション技術の研究開発の概要について述べた。現

\*3 <https://www.wisdom-nict.jp/>

### 3 社会知コミュニケーション技術

在の自然言語処理において中心的役割を果たしている深層学習についてその状況を概観し、社会知コミュニケーション技術の研究開発として NICT で取り組んでいる内容の概要を説明した。今後の自然言語処理分野における研究開発の方向性としては、引き続き更なる大規模言語モデルの開発は続いていくと見られるが、既に述べたように社会実装上の課題も多く、そうした大規模言語モデルの構築は、当面はその可能性を示すに留まり、実際に社会で使われるのにはまだまだ時間を要すると考えている。一方で、超大規模言語モデルほどの規模ではなく、現実的な規模の言語モデルを利用する場合には、ファインチューニングして用いることが想定されるため、多様なタスクで高い精度を出すためには、それらのタスクの高品質な学習データを有しているか否かが成否を分けるとも考えられる。そのため、そういった学習データをはじめ Web テキストのような自然言語処理のための基礎的なデータの蓄積を継続的に実施していくことが非常に重要だと考えている。今後も NICT の特色を活かし、そうしたデータを蓄積しつつ、2030 年頃に社会知を駆使して多様な話題に関して人間とブレインストーミングしたり、適切な意思決定やイノベーションを促進することができる対話システムの実現を目指し、社会知コミュニケーション技術の研究開発に取り組んでいく。

### 謝辞

社会知コミュニケーション技術について日頃から議論している DIRECT の職員に深く感謝する。

#### 【参考文献】

- 1 J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proceedings of NAACL-HLT 2019, pp.4176-4186, 2019.
- 2 J. Kaplan et al., "Scaling Laws for Natural Language Models," arXiv preprint, arXiv:2001.08361, 2020.
- 3 A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," 2019.
- 4 田中 正弘, 田浦 健次郎, 塙 敏博, 鳥澤 健太郎, "自動並列化深層学習ミドルウェア RaNNC," 自然言語処理, vol.28, no.4, pp.1299-1306, 2021.
- 5 T. Brown et al., "Language Models are Few-Shot Learners," arXiv preprint, arXiv:2005.14165, 2020.
- 6 田中 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.
- 7 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.
- 8 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.
- 9 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.



大竹 清敬 (おおたけ きよのり)

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  
研究センター長

博士(工学)

自然言語処理、音声言語処理

【受賞歴】

2021年 第3回日本オープンイノベーション  
大賞、総務大臣賞

2019年 文部科学大臣表彰、科学技術賞

2014年 Twitter Data Grants 獲得