

大会企画 | 2023年11月23日

📅 2023年11月23日(木) 9:20 ~ 11:20 🏢 A会場 (KFMホール"イオ")

大会企画1

生成AIの医療への応用

オーガナイザー：土井 俊祐（千葉大学）

座長：土井 俊祐（千葉大学）、河添 悦昌（東京大学大学院医学系研究科）

[2-A-2-02] 大規模言語モデルの構築とドメイン適応

*相澤 彰子¹ (1. 国立情報学研究所)

キーワード：large language models、domain adaptation、biomedical natural language processing

大規模言語モデルの登場によって、人間の言語を扱うAI技術は飛躍的な進歩を遂げた。特に、与えられた入力文に対して最適な応答を予測する生成系モデルは、検索、推薦、要約、質疑応答などの多様なタスクを対話的に実行することが可能で、その汎用性は従来のAIシステムとは一線を画すものである。大規模言語モデルには膨大な文書から獲得した情報が埋め込まれており、プログラムコードや構造化文書を含む多様なスタイルにも対応できる。さらに画像、音声、データなどの異なるモダリティへの拡張、行動シーケンスの自動生成によるエージェントやロボットの制御など想定される応用範囲は広い。一方で、大規模言語モデルは構築に膨大なコストがかかることが知られている。また、社会バイアスや情報の正確さなど様々な問題点に対応するためには、適応言語や分野にあわせた調整が必要である。本発表では、日本語を中心とした大規模言語モデルの構築の試みについて、国内の自然言語処理研究者を中心とした活動状況を踏まえながら紹介する。まず大規模言語モデルの仕組みや構築のフローを概観し、次にドメイン適応の方法を簡単に述べる。また、生物医学などの学術分野を想定しながら、知識の抽出・検索および活用に関する研究の最新動向を紹介する。

大規模言語モデルの構築とドメイン適応

- 研究動向の紹介 -

相澤 彰子^{*1}

^{*1} 国立情報学研究所

Constructing Large Language Models with Domain Adaptation

- A View into the Current Research Trends -

Akiko Aizawa^{*1}

^{*1} National Institute of Informatics

Abstract

This presentation will introduce an attempt to construct a large-scale language model. First, the mechanism and construction flow of large-scale language models will be briefly presented, followed by a short introduction of domain adaptation methods. Also, the latest trends in knowledge and question answering from academic papers using language models will be introduced.

Keywords: Large Language Models, Domain Adaptation, Biomedical Natural Language Processing.

1. 大規模言語モデル

大規模言語モデル (large language models) の登場によって、人間の言語を扱う AI 技術は飛躍的な進歩を遂げた。特に、与えられた入力文に対して最適な応答を予測する生成系モデルは、検索、推薦、要約、質疑応答などの多様なタスクを対話的に実行することが可能で、その汎用性は従来の AI システムとは一線を画すものである。大規模言語モデルには膨大な文書から獲得した情報が埋め込まれており、プログラムコードや構造化文書を含む多様なスタイルにも対応できる。さらに画像、音声、データなどの異なるモダリティへの拡張、行動シーケンスの自動生成によるエージェントやロボットの制御など想定される応用範囲は広い。

一方で、大規模言語モデルは構築に膨大なコストがかかることが知られている。また、社会バイアスや情報の正確さなど様々な問題点に対応するためには、適応言語や分野にあわせた調整が必要である。

2. 大規模言語モデルの構築

基本的な言語タスクを想定する場合、大規模言語モデルの構築から利用までは、一般に以下のステップを踏んで行われる。

- ① コーパスの収集とクリーニング
- ② モデルの学習
- ③ 指示チューニング (instruction tuning)
- ④ 入力 (prompt) による調整

ここで①では、モデルのサイズに見合った十分な量のコーパスを準備することがポイントとなる。②は、もともと計算資源を使うステップである。モデルは大きければよいというわけではなく、モデルのパラメタ数、計算資源の量、コーパスの大きさのバランスが重要であるとされている。¹⁾ ③の指示チューニングは、分類、知識抽出、質問応答、要約、翻訳などのさまざまなタスクに言語モデルが応えられるよう、いわば「問題を解くためのメタ知識」をモデルに教え込むために有効とされる。また、社会適用のための価値観の調整 (value / human alignment) も指示チューニングの役割である。人間によるフィードバックを組み込んだ強化学習 (reinforcement learning with human feedback) を用いる方法が広く知られている。²⁾ 最

後に④は、実際に言語モデルを使用する場面でユーザが工夫をすることで、モデル出力の品質を高めようとするもので、プロンプトエンジニアリングやプロンプトプログラミングとして広く実践されている。

3. ドメイン適応

ChatGPT に代表される大規模言語モデル³⁾は汎用的な基盤となるものであるが、一方で、透明性、信頼性、説明可能性などの観点から、今後は個々のドメインへの適応も重要になる。

ドメイン適応は前節で述べた各ステップにおいて可能で、利用可能な計算資源やコーパスを踏まえて、適切な方法を選択することになる。ここでは構築にかかわる①～③について簡単に述べる。まず①について、コーパスとして対象分野の論文、特許、医療記録などを収集して整形する必要がある。次に②について、想定しているモデルの大きさに対して十分な量のコーパスが集まれば、何もない状態から言語モデルを訓練することが可能である。このときに Web などから収集した他の汎用的なコーパスを併用することも考えられる。一方で、近年の大規模なモデルの大きさを踏まえると、汎用コーパスで学習したオープンな事前学習済みのモデルに対して、ドメインに特化したコーパスを用いて追加事前学習することも現実的な方法である。さらに③では、ドメインに特化したインストラクションの準備が不可欠である。⁴⁾

4. 活動紹介

現在、国内の自然言語処理研究者を中心として、日本語に強いオープンな大規模言語モデル構築の試みがはじまっている。⁵⁾ 生物医学分野はドメイン適応の主要なターゲットの1つであり、日本語の論文や診療記録などへの展開が期待される。

参考文献

- 1) Hoffmann, Jordan, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, et al. 2022. "Training Compute-Optimal Large

- Language Models.” arXiv [cs.CL]. arXiv.
<http://arxiv.org/abs/2203.15556>.
- 2) Ouyang, Long, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, et al. 2022. “Training Language Models to Follow Instructions with Human Feedback.” arXiv [cs.CL]. arXiv.
<http://arxiv.org/abs/2203.02155>.
 - 3) Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, et al. 2023. “A Survey of Large Language Models.” arXiv [cs.CL]. arXiv.
<http://arxiv.org/abs/2303.18223v10>.
 - 4) Wu, Chaoyi, Xiaoman Zhang, Ya Zhang, Yanfeng Wang, and Weidi Xie. 2023. “PMC-LLaMA: Further Finetuning LLaMA on Medical Papers.” arXiv [cs.CL]. arXiv.
<http://arxiv.org/abs/2304.14454>.
 - 5) <https://llm-jp.nii.ac.jp/>