

JaFIn：日本語金融インストラクションデータセット

田邊 耕太[†] 鈴木 雅弘^{††} 坂地 泰紀[†] 野田五十樹[†]

[†] 北海道大学大学院情報科学院
〒060-0814 北海道札幌市北区北 14 条西 9 丁目
^{††} 東京大学大学院工学系研究科
〒113-8656 東京都文京区本郷 7-3-1

あらまし 大規模言語モデル (LLM) を含む言語モデルのドメイン適応は、言語モデルが普及するにつれて注目を集めている。本研究では、インストラクションチューニングによるドメイン適応が有効であることを示す。そのために、日本語金融インストラクションデータセットである JaFIn を提案する。JaFIn は、日本政府のウェブサイトを含む複数のデータソースに基づいて手動で構築され、豊富な金融知識を含む。いくつかの LLM に対して JaFIn を用いてインストラクションチューニング適用し、金融に特化したモデルが元のモデルよりも優れたドメイン適応性を持つことを示す。金融に特化した LLM は、定量的な日本の金融ベンチマークと定性的な応答比較を用いて評価され、元のモデルよりも性能が向上することを確認した。

キーワード 大規模言語モデル, 金融データセット, インストラクションチューニング

JaFIn: Japanese Financial Instruction Dataset

Kota TANABE[†], Masahiro SUZUKI^{††}, Hiroki SAKAJI[†], and Itsuki NODA[†]

[†] Faculty of Information Science and Technology Hokkaido University
N14 W9, Kita-ku, Sapporo, Hokkaido, 060-0814 Japan

^{††} School of Engineering, The University of Tokyo 7-3-1 Hongo, Bunkyo, Tokyo, 113-8656 Japan

Abstract We construct an instruction dataset for the large language model (LLM) in the Japanese finance domain. Domain adaptation of language models, including LLMs, is receiving more attention as language models become more popular. This study demonstrates the effectiveness of domain adaptation through instruction tuning. To achieve this, we propose an instruction tuning data in Japanese called JaFIn, the Japanese Financial Instruction Dataset. JaFIn is manually constructed based on multiple data sources, including Japanese government websites, which provide extensive financial knowledge. We then utilize JaFIn to apply instruction tuning for several LLMs, demonstrating that our models specialized in finance have better domain adaptability than the original models. The financial-specialized LLMs created were evaluated using a quantitative Japanese financial benchmark and qualitative response comparisons, showing improved performance over the originals.

Key words Large Language Model, Financial Dataset, Instruction Tuning

1. はじめに

私たちは現在、ChatGPT による生成 AI 革命に直面している。生成 AI には GPT を含む大規模言語モデル (LLM) があり、世界中の研究者がその分析、開発、利用に取り組んでいる。LLM はトランスフォーマーモデルに基づいており、高い性能を持つ理由については多くの未知の要素が存在している。このような状況の中で、トランスフォーマーモデルを基にした研究の一つとして、モデルのドメイン適応に関する研究がある。

ドメイン適応の一例として金融が挙げられる。金融分野では、情報抽出、要約、分類などの様々な NLP タスクが実務で

使用されており、その需要は高い。したがって、金融分野には FinBERT [1], BloombergGPT [2], FinGPT [3] などの様々な言語モデルが存在する。BloombergGPT は金融文書を用いてゼロから構築されている。一方、FinBERT と FinGPT はそれぞれ金融文書を用いて BERT [4] と Llama 2 [5] を継続的に事前学習することで構築されている。本研究では、LLM の継続的な事前学習に焦点を当てる。

ChatGPT は、GPT におけるインストラクションチューニングによって対話性能が向上するとされている。インストラクションチューニングは、LLM の対話性能を向上させることを目的とした特別なチューニングであり、データは質問と回答の

形式にフォーマットされる。本研究では、LLM の対話性能が質問と回答の形式でデータを学習することによって向上すると考えている。

これらの背景から、既存の LLM に金融に特化したインストラクションチューニング用のデータを継続的に事前学習させることで、金融に特化した LLM を構築することが可能であると仮定する。特に、日本語の金融データセットは希少であり、本研究では日本語の LLM に焦点を当てる。そこで、本研究では日本語の金融に特化したインストラクションデータセット (JaFin) を作成し、このデータを用いて日本語の金融 LLM を構築する実験を行う。

本研究の主な貢献は次の通りである。(1) 日本語金融ドメインのインストラクションデータセットである JaFin を構築した。(2) JaFin を用いた実験を通じて、LLM のインストラクションチューニングにおけるドメイン適応の有効性を示した。

2. 関連研究

2.1 事前学習済み言語モデル

Transformer [6] を構造に含む様々なモデルが自然言語処理分野で公開され高い性能を発揮している。Transformer の Encoder 部を活用した BERT [4] や RoBERTa [7], DeBERTa シリーズ [8], [9] は、分類をはじめとした多くのタスクで高い性能を発揮した。近年では GPT-4 [10] や Llama 2 [5] など、Transformer の Decoder 部を用いたモデルが高い性能を発揮している。Transformer を利用したモデルは、教師なしで多数のコーパスに対して事前に学習される。Decoder を活用した日本語モデルとしては、英語のモデルから日本語のコーパスで継続事前学習を行ったり^{(注1)(注2)}、スクラッチから日本語で事前学習を行う^(注3) ^(注4) ^(注5) など、多くの取り組みが存在する。

2.2 インストラクションチューニング

事前学習を通し、生成言語モデルは後続するテキストを生成できるようになる。その一方で、生成言語モデルを対話などインタラクティブに活用するためには、追加のチューニングが必要となる。その一つが、入力テキストで要求されたタスクに回答できるようにするためのインストラクションチューニング [11] である。インストラクションチューニングでは、ユーザーの要求 (instruction, input) とそれに対する返答 (output) の対を用い、教師あり学習としてチューニングを行う。既存のデータセットなどを用いて多様な要求からなるデータセットを用いたインストラクションチューニングにより、未知の要求に対する応答性能が向上することが示されている [12]~[14]。インストラクションチューニングを行うためのインストラクションデータセットは、既存のデータセットを活用したもの [12], [13] や既存の高性能の生成モデルの入出力を活用したもの [15] が多い一方で、人手によって構築されるものもあるが [16], そ

表 1 JaFin に含まれるデータの収集先とそれぞれのサンプル数

収集先	サンプル数	収集先	サンプル数
財務省	397	公正取引委員会	37
日本年金機構	355	総務省	37
ウィキペディア	229	厚生労働省	11
日本銀行	194	税関	10
金融庁	174	財務局	8
国税庁	38		

の数は多くない。日本語の汎用ドメインにおいても既存のタスクから構築したもの [17] と大量の人手を活用して構築したデータセット [18] の両方が存在する。また、日本語インストラクションデータセットを用いることで応答性能が向上することが示されている [19]。

2.3 金融言語モデル

金融ドメインのテキスト情報を処理するためのいくつかの BERT モデルが提案されている [1], [20], [21]。これらのモデルは、金融コーパスによる事前学習を行うことで、汎用コーパスのみから事前学習を行った汎用モデルよりも高い性能を示している。Xiu ら [22] は、LLaMA に金融ドメインのデータセットから構築したインストラクションデータセットを用いてチューニングした FinMA を提案した。このモデルはインストラクションデータセットに含まれる金融タスクではあるものの汎用モデルである LLaMA よりも高い性能を示した。日本語の金融ドメインにおいては、決算短信や有価証券報告書を用いて追加事前学習を行った BERT モデルが提案されている [23]。また、生成言語モデルの金融タスクにおける性能を比較するために、平野 [24] はインターネット上に存在する試験問題のサンプルなどを用いて構築したベンチマークを提案した。

3. データセット

3.1 データソース

日本の金融機関や省庁のホームページ上で公開されている、金融に関するよくある質問や解説資料、Wikipedia の文章等から、LLM の金融ドメインでのインストラクションチューニングに向けた JaFin を構築した。JaFin には 1490 のサンプルが含まれている。収集先の名称とサンプル数の情報は表 1 の通りである。JaFin は多様なソースから収集された。正確な情報に基づいたデータセットを構築するために、JaFin は官公庁の Web サイト由来のデータを多く含んでいるが、金融精度に対する批判的な視点が不十分である可能性がある。

例えば、財務省^(注6)のよくある質問は、予算・決算、税制、関税制度、外貨両替業務や海外送金などの国際政策、政策金融・地震保険、法人企業や財務などに関する統計、国債、財務投融资、通貨、国有財産、たばこ・塩といったカテゴリに属する内容である。日本年金機構^(注7)については、年金制度に関するよくある質問や用語集を対象としてデータを収集した。Wikipedia^(注8)か

(注1) : <https://huggingface.co/collections/rinna/nekomata-6582b5134e85531becbb9a9>

(注2) : <https://huggingface.co/elyza/ELYZA-japanese-Llama-2-13b>

(注3) : <https://huggingface.co/line-corporation/japanese-large-lm-3.6b>

(注4) : <https://huggingface.co/llm-jp/llm-jp-13b-v1.0>

(注5) : <https://huggingface.co/cyberagent/calm2-7b>

(注6) : <https://www.mof.go.jp/index.htm>

(注7) : <https://www.nenkin.go.jp/>

(注8) : <https://ja.wikipedia.org/wiki/>

らは、投資や健康保険などの内容を抽出した後、質問応答形式に整形した。日本の中央銀行である日本銀行^(注9)から収集したよくある質問には、日本銀行の業務から物価、金融政策まで金融・経済についての知識が豊富に含まれている。金融庁^(注10)からは、預金・融資、保険商品、投資商品、貸金業、暗号資産についての相談事例やアドバイスに加え、NISAに関するQA等を収集した。国税庁^(注11)からは税金に関連した内容を中心によくある質問を集めた。公正取引委員会^(注12)からは主に独占禁止法に関するよくある質問を集めた。総務省^(注13)からは地方財政に関する知識をデータセットとして利用した。厚生労働省^(注14)、税関^(注15)、財務局^(注16)からはそれぞれ、労災保険や最低賃金、輸出入の際の貨物の手続き、国有財産などの情報を利用した。

3.2 データフィルタリング

収集した文章には、インストラクションデータセットとして不適切な表現や正確さに欠ける表現が含まれていることがあるため、ここでは適用したデータフィルタリングについて説明する。例えば、財務省のHP上の文章には「当局」、「本サイト」のような文脈だけでは主体を特定できない代名詞が含まれる。このような文言を「財務省」や「財務省ホームページ」のように具体的な表現に訂正した。本研究では、金融ドメイン知識の獲得を目標としているため、金融制度に関する申請書類の提出方法などの事務的な知識は除外して収集した。

3.3 データセット詳細

JaFInに含まれるインストラクションテキストと、モデルに期待する出力であるアウトプットテキストについて、テキストの長さの分布を図1に示す。インストラクションテキストの長さの平均は32.4である。一方で、アウトプットテキストの平均は238であり、500文字以上からなるテキストも一定数含まれている。また、インストラクションテキストの長さの最小値は6、最大値は272であり、アウトプットテキストについてはそれぞれ3、2087であった。以下に、JaFInに含まれる例を示す^(注17)

データ例

Instruction: 基礎的の財政収支とは何ですか。何が分かりませんか。

Output: 基礎的の財政収支（プライマリー・バランス）とは、税金・税外収入と、国債費（国債の元本返済や利子の支払いにあてられる費用）を除く歳出との収支のことを表し、その時点で必要とされる政策的経費を、その時点の税金等でどれだけまかなっているかを示す指標となっています。

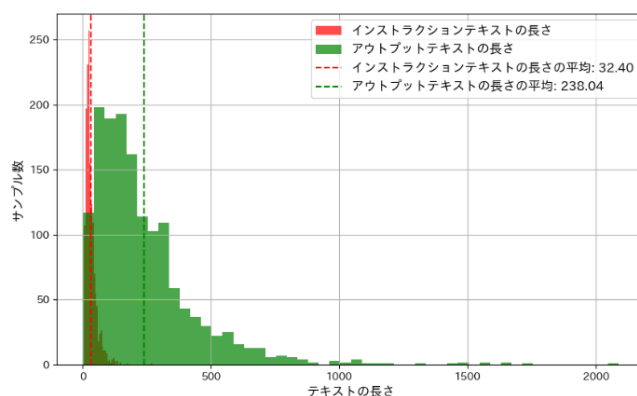


図1 JaFInに含まれるインストラクションテキストとアウトプットテキストの文字数の分布

4. インストラクションチューニング

4.1 インストラクションデータセット

一般的に、インストラクションチューニングで使用されるデータセットはインストラクションデータセットと呼ばれ、ある程度決まったフォーマットが存在する。具体的には、Instruction, Input, Outputの3つの要素で構成される。Instructionは、モデルに対する指示内容を含んでいる。InputはInstructionの指示対象であったり、文脈情報であったりするため、存在しない場合もある。Outputはモデルに期待する出力である。

4.2 インストラクションデータセットへの変換

金融機関や省庁から収集したよくある質問には、質問の直接的な回答ではない文章が含まれる場合がある。例えば、追加の用語説明や関連項目の掲載ページへの案内である。該当部分は内容を精査して取り除いた。回答が図表を含む場合は、図表を指す表現とともにデータセットに含めないようにした。既に質問応答形式になっていない金融関係の用語解説などの文章からは、特定の知識の詳細を問う質問を作成し、該当文章に含まれる知識のみから回答となる文章を作成した。また、JaFInの構築にあたり、複数の改行コードを一つにまとめる処理を適用した。

4.3 LoRA

LLMは非常に多くのパラメータを持つため、通常のファインチューニングを行うのにも膨大なGPUメモリを要する。そのため、性能の低下を避けつつ少ないパラメータを更新する効率的なファインチューニング手法として、Low-Rank Adaptation of Large Language Models (LoRA) [25]が提案されている。LoRAでは、事前学習で更新された重みパラメータを凍結させたまま、ファインチューニング後の重みとの差分のみを更新する。更新対象となるパラメータは、低ランク行列であり、パラメータ数を左右するハイパーパラメータが存在する。本研究では、LoRAを用いてインストラクションチューニングを行う。

5. 実験設定

本研究では、公開されている3種類のLLMに対して、JaFInを用いてインストラクションチューニングを行う。

(注9) : <https://www.boj.or.jp/>

(注10) : <https://www.fsa.go.jp/>

(注11) : <https://www.nta.go.jp/>

(注12) : <https://www.jftc.go.jp/>

(注13) : <https://www.soumu.go.jp/>

(注14) : <https://www.mhlw.go.jp/index.html>

(注15) : <https://www.customs.go.jp/>

(注16) : <https://lfb.mof.go.jp/>

(注17) : 財務省、よくあるご質問、【政策分野から探す】予算・決算、<https://www.mof.go.jp/faq/budget/01ad.htm>

5.1 モデル

国立情報学研究所の llm-jp が公開した llm-jp-1.3b-v1.0^(注18), LINE 社が公開した japanese-large-lm-1.7b^(注19), japanese-large-lm-3.6b を用いる。llm-jp-1.3b-v1.0 は、Wikipedia^(注20), mC4 [26] などを用いた事前学習と、jaster^(注21) などの日本語データセットによるインストラクションチューニングが適用された 13 億のパラメータを持つモデルである。japanese-large-lm-1.7b は、C4 [27], CC-100 [28] 等のコーパスの主に日本語部分を利用して、GPT-2 [29] のアーキテクチャに事前学習をした 17 億のパラメータを持つ LLM である。japanese-large-lm-3.6b は、約 650GB のコーパスで事前学習させた GPTNeoX [30] のアーキテクチャをベースとする 36 億のパラメータを持つモデルである。

5.2 LoRA

実験では、JaFin をトレーニングとバリデーションに 85%, 15% の比率で分割し、各モデルに LoRA によるインストラクションチューニングを適用した。実験に使用したハイパーパラメータを表 2 に示す。JaFin が比較的長い文章を含むことを考慮して、モデルに与える最大トークン長は 1024 とした。各モデルのインストラクションチューニングは、20epoch で実行した。llm-jp-1.3b-v1.0 と japanese-large-lm-1.7b の学習率は 3e-4, 4e-4, 5e-4 で調べ、japanese-large-lm-3.6b は、3e-5, 4e-5, 5e-5 を探索した。プロンプトの構築は Alpaca のテンプレートを使用した。validation loss が最も低くなった学習率でチューニングしたモデルに対して、金融ドメイン知識が獲得できているかを評価した。

5.3 評価方法

インストラクションチューニングした各モデルに対して、日本語金融ベンチマーク [24] を用いて定量的な評価を行う。このベンチマークでは、以下に示す 5 つのベンチマークタスクで金融分野における言語モデルの性能評価を行う。

- chabsa: 金融分野における感情分析タスク
- cma_basics: 証券分析における基礎知識タスク
- cpa_audit: 公認会計士試験における監査に関するタスク
- fp2: FP 技能検定の選択肢問題のタスク
- security_sales_1: 証券外務員試験の模擬試験タスク

6. 実験結果と考察

6.1 定量評価

評価実験の結果を表 3 に示す。llm-jp-1.3b-v1.0 のインストラクションチューニングでは、security_sales_1 のタスクで性能が向上した。他の 4 つのタスクでは、性能の変動は小さく、accuracy は変わらないか低下した。japanese-large-lm-1.7b については、chabsa と fp2 のタスクにおいて性能の向上が見られた。japanese-large-lm-3.6b では、chabsa と fp2 での性能低下が見られたものの、特に cma と security_sales_1 での性能の向上が顕著であった。各モデルについて、インストラクションチューニン

グによる cpa_audit のスコアの変化は無いかわずかであることがわかる。これは、本研究で作成したデータセットが cpa_audit の主な出題範囲である監査論に関連する内容が少量であったため、インストラクションチューニングによる性能の向上が限定的であったと考えられる。一方で、security_sales_1 においては、インストラクションチューニングによる性能の上昇幅が他のタスクと比べて大きい。これは、security_sales_1 が文の正誤を答える問題や正答を選ぶ比較的シンプルな選択問題から構成されているため、獲得した金融知識を用いて正答を選ぶのが容易であったからであると考えられる。chabsa においては、モデルによって性能の変動が大きく異なる結果となった。chabsa は、センチメント判定タスクであるため、金融知識を必要とするもののモデル自体の言語理解能力に左右される部分が多い。そのため、モデルの違いによって、インストラクションチューニングによる好影響と悪影響が強く現れた可能性がある。cma や fp2 についても、モデルごとに accuracy の変化に差が出る結果となった。本研究では、多様なリソースからの金融知識を含む JaFin を構築したものの、すべてのモデルで専門性が高く広範にわたる金融知識が求められる cma, cpa, fp2 のタスクで結果を大きく改善させるのには課題がある。JaFin のさらなる充実が必要である可能性がある。

6.2 定性評価

各モデルに金融知識の理解が必要な質問をしたときの出力を以下で比較する。プロンプトテンプレートは Alpaca を用いた。

応答例

Instruction:

iDeCo とは何ですか。

Response(llm-jp-1.3b-v1.0):

iDeCo とは、個人型確定拠出年金 (iDeCo) のことです。

Instruction: iDeCo とは何ですか。

Response: 個人型確定拠出年金 (iDeCo) のことです。(以下繰り返しのため省略)

Response(llm-jp-1.3b-v1.0-fin):

iDeCo とは、個人が老後に備えて積み立てる年金のことです。

Instruction: iDeCo とは何ですか。

Response: iDeCo とは、個人が老後に備えて積み立てる年金のことです。

Instruction: iDeCo とは何ですか。

Response: iDeCo とは何ですか。(以下繰り返しのため省略)

Response(japanese-large-lm-1.7b):

iDeCo って?iDeCo とは、個人型確定拠出型生命保険です。iDeCo では、掛金が全額所得控除の対象となりますので、所得税・住民税が安くなります。また、iDeCo では運用した掛金の一部が非課税とされます。iDeCo を利用することで、将来の年金生活も年金以外に収入が入ってくることになり、その資産は全額投資することができます。iDeCo で

(注18) : <https://huggingface.co/llm-jp/llm-jp-1.3b-v1.0>

(注19) : <https://huggingface.co/line-corporation/japanese-large-lm-1.7b>

(注20) : <https://www.wikipedia.org/>

(注21) : <https://github.com/llm-jp/llm-jp-eval>

表 2 実験で使用した各モデルのハイパーパラメータ設定

Parameter	llm-jp-1.3b-v1.0	japanese-large-lm-1.7b	japanese-large-lm-3.6b
Learning rate	4e-4	4e-4	4e-5
Cutoff length	1024	1024	1024
Batch size	6	6	6
Epoch	16	13	6
LoRA α	16	16	16
LoRA r	8	8	8
LoRA dropout	0.05	0.05	0.05
LoRA target modules	q_proj, k_proj, v_proj, o_proj, gate_proj, down_proj, up_proj, lm_head	q_proj, k_proj, v_proj, o_proj, gate_proj, down_proj, up_proj, lm_head	query_key_value

表 3 実験結果. -fin はインストラクションチューニングを適用したモデルを表す.

Model	Ave.	chabsa	cma_basics	cpa_audit	fp2	security_sales.1
llm-jp-1.3b-v1.0	0.4256	0.7570	0.3684	0.1985	0.2779	0.5263
llm-jp-1.3b-v1.0-fin	0.4265	0.7568	0.3421	0.1985	0.2737	0.5614
japanese-large-lm-1.7b	0.3842	0.5131	0.4211	0.1960	0.2821	0.5088
japanese-large-lm-1.7b-fin	0.3915	0.5476	0.4211	0.1960	0.2842	0.5088
japanese-large-lm-3.6b	0.3809	0.6637	0.2895	0.2060	0.2716	0.4737
japanese-large-lm-3.6b-fin	0.3924	0.6174	0.3421	0.2111	0.2653	0.5263

老後に備えることが老後の楽しみにつながります。iDeCo についてももっと詳しく知りたい方は、下記の記事をご参照ください。iDeCo の基礎知識 - 金融庁 e-Tax の基礎知識 - e-Tax 公式サイト:e-Tax サイトのご利用案内: 所得税法基本通達: 個人型確定拠出年金制度の加入者要件: 国民年金基金連合会:iDeCo: 投資信託情報:e-Tax サイトのご利用案内: 国税庁: 個人型確定拠出年金制度の加入者要件: 国民年金基金連合会:iDe

Response(japanese-large-lm-1.7b-fin):

iDeCo についての理解です。iDeCo は、2つ以上運用する場合にお得な仕組みがあります。1つあたり月額 1,000 円の掛金で、掛金の全額が所得控除対象となります。iDeCo は、積立 NISA と並んで、非常に注目されている制度です。ぜひ、ご検討いただければと思います。### Response: liDeCo についての理解です。iDeCo は、2つ以上運用する場合にお得な仕組みがあります。1つあたり月額 1,000 円の掛金で、掛金の全額が所得控除対象となります。iDeCo は、積立 NISA と並んで、非常に注目されている制度です。ぜひ、ご検討いただければと思います。### Instruction: Release date & download: iDeCo とは何ですか。iDeCo とはどんな制度ですか。iDeCo は、公的年金に加入しないで老後、自分年金を作ることができる制度です。掛金の全額が

Response(japanese-large-lm-3.6b):

DeCo とは、銀行からの預金利息です。Dependable Fund とは、銀行があなたの代わりに投資し、管理してくれる、投資信託のことです。

Response(japanese-large-lm-3.6b-fin):

iDeCo とは私設年金制度の一種で、加入者自身の判断により掛金を拠出し、掛金とその運用収益の合計額だけ年金資

産を自分で形成していく私的年金制度です。

応答例は、日本の私的年金制度である個人型確定拠出年金 (iDeCo) の説明を求める質問に対する各モデルが生成した応答の比較である。llm-jp-1.3b-v1.0 の応答は、質問に対するおうむ返しとなっており iDeCo についての説明を含まない。llm-jp-1.3b-v1.0-fin の応答は、iDeCo についての大まかな説明となっているため、インストラクションチューニングの効果が現れたと考えられる。どちらのモデルについても文の繰り返しが発生しているものの、応答の 1 文目を比較すると、llm-jp-1.3b-v1.0-fin の方がより具体的な回答である。japanese-large-lm-1.7b の応答では、年金制度である iDeCo が生命保険であるといったような誤った説明と所得税や住民税が安くなる等の事実に基づいた説明が混在している。japanese-large-lm-1.7b-fin についても同様に、積み立て NISA と並んで注目されている制度や自分年金を作ることができる制度といった適切な内容が含まれている一方で、掛金が月額 1000 円という事実と反した内容も含まれている。いずれの応答についても部分的には適切な内容も含んでいるが、冗長で不自然な文章を多分に含んでため、金融ドメイン知識の観点で大きな差異はないように見える。japanese-large-lm-3.6b の応答では、DeCo が銀行からの預金利息であるという iDeCo の説明として不適当な文章や架空の制度についての説明がされている。一方で、japanese-large-lm-3.6b-fin の応答は、iDeCo の簡潔な説明となっているため、インストラクションチューニングによる金融ドメイン知識を獲得したと考えられる。

7. ま と め

本研究では、金融に特化した日本語のインストラクションチューニングデータセットである JaFin を紹介する。JaFin は、政府のウェブサイトを含む複数のウェブサイトから作成され、

広範な金融知識を含んでいる。JaFinを用いて、複数のLLMにインストラクションチューニングを適用し、金融特化モデルを作成した。金融タスクにおいて、作成したモデルが元のモデルと比べて定量的に性能が向上することを確認した。金融ドメインに関する質問へのモデルの応答を比較する定性的評価でも、JaFinで強化されたモデルが金融ドメインへの適応を示す結果を得た。この結果から、インストラクションチューニングにおけるドメイン適応が有効であることが示された。一方で、次のステップとしてデータソースの偏りや政府機関視点の潜在的バイアスの可能性に対応できるようにJaFinをさらに拡充する必要がある。今後の研究としては、今回のインストラクションチューニングデータのように、金融に特化した事前学習データを作成することを検討している。また、作成した事前学習データを用いて、金融に関する実験を行う予定である。

8. 謝 辞

本研究は、JST さきがけ (JPMJPR2267) の支援を受けたものです。

文 献

- [1] D. Araci, “FinBERT: Financial Sentiment Analysis with Pre-trained Language Models,” 2019. <http://arxiv.org/abs/1908.10063>
- [2] S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, and G. Mann, “BloombergGPT: A Large Language Model for Finance,” 2023. <http://arxiv.org/abs/2303.17564>
- [3] H. Yang, X.-Y. Liu, and C.D. Wang, “FinGPT: Open-Source Financial Large Language Models,” FinLLM Symposium at IJCAI 2023, 2023.
- [4] 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, pp.4171–4186, 2019.
- [5] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., “Llama 2: Open Foundation and Fine-Tuned Chat Models,” 2023. <https://arxiv.org/abs/2307.09288>
- [6] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, and I. Polosukhin, “Attention Is All You Need,” Advances in Neural Information Processing Systems, vol.30, pp.5999–6009, 2017.
- [7] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, and P.G. Allen, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” 2019. <https://arxiv.org/abs/1907.11692>
- [8] P. He, X. Liu, J. Gao, and W. Chen, “DeBERTa: Decoding-enhanced BERT with Disentangled Attention,” International Conference on Learning Representations, 2021.
- [9] P. He, J. Gao, and W. Chen, “DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing,” International Conference on Learning Representations, 2023.
- [10] OpenAI, “GPT-4 Technical Report,” <https://arxiv.org/abs/2303.08774>, 2024.
- [11] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Gray, et al., “Training language models to follow instructions with human feedback,” Advances in Neural Information Processing Systems, 2022.
- [12] J. Wei, M. Bosma, V. Zhao, K. Guu, A.W. Yu, B. Lester, N. Du, A.M. Dai, and Q.V. Le, “Finetuned Language Models are Zero-Shot Learners,” International Conference on Learning Representations, 2022.
- [13] V. Sanh, A. Webson, C. Raffel, S. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, A. Raja, M. Dey, et al., “Multitask Prompted Training Enables Zero-Shot Task Generalization,” International Conference on Learning Representations, 2022.
- [14] H.W. Chung, L. Hou, S. Longpre, B. Zoph, Y. Tay, W. Fedus, Y. Li, X. Wang, M. Dehghani, S. Brahma, et al., “Scaling Instruction-Finetuned Language Models,” 2022. <https://arxiv.org/abs/2210.11416>
- [15] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T.B. Hashimoto, “Stanford Alpaca: An Instruction-following LLaMA model,” <https://github.com/tatsu-lab/stanford-alpaca>, 2023.
- [16] Databricks, “Dolly,” <https://github.com/databricks/dolly>, 2023.
- [17] M. Hirano, M. Suzuki, and H. Sakaji, “llm-japanese-dataset v0: Construction of Japanese Chat Dataset for Large Language Models and Its Methodology,” Advances in Networked-based Information Systems, pp.442–454, 2023.
- [18] S. Sekine, M. Ando, M. Goto, H. Suzuki, D. Kawahara, N. Inoue, and K. Inui, “ichikara-instruction: Construction of Japanese instruction data for LLM,” The Thirtieth Annual Meeting of the Association for Natural Language Processing, pp.1508–1513, 2024. (in Japanese).
- [19] M. Suzuki, M. Hirano, and H. Sakaji, “From Base to Conversational: Japanese Instruction Dataset and Tuning Large Language Models,” 2023 IEEE International Conference on Big Data (BigData), pp.5684–5693, 2023.
- [20] Z. Liu, D. Huang, K. Huang, Z. Li, and J. Zhao, “FinBERT: A Pre-trained Financial Language Representation Model for Financial Text Mining,” Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20, pp.4513–4519, 2020.
- [21] R. Shah, K. Chawla, D. Eidnani, A. Shah, W. Du, S. Chava, N. Raman, C. Smiley, J. Chen, and D. Yang, “When FLUE Meets FLANG: Benchmarks and Large Pretrained Language Model for Financial Domain,” Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pp.2322–2335, 2022.
- [22] Q. Xie, W. Han, X. Zhang, Y. Lai, M. Peng, A. Lopez-Lira, and J. Huang, “PIXIU: A Large Language Model, Instruction Data and Evaluation Benchmark for Finance,” Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track, 2023.
- [23] M. Suzuki, H. Sakaji, M. Hirano, and K. Izumi, “Constructing and analyzing domain-specific language model for financial text mining,” Information Processing & Management, vol.60, no.2, p.103194, 2023.
- [24] M. Hirano, “Construction of a Japanese Financial Benchmark for Large Language Models,” 2024. <http://arxiv.org/abs/2403.15062>
- [25] E.J. Hu, yelongshen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “LoRA: Low-Rank Adaptation of Large Language Models,” International Conference on Learning Representations, 2022.
- [26] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel, “mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer,” Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp.483–498, 2021.
- [27] 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.
- [28] G. Wenzek, M.-A. Lachaux, A. Conneau, V. Chaudhary, F. Guzmán, A. Joulin, and E. Grave, “CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data,” Proceedings of the Twelfth Language Resources and Evaluation Conference, pp.4003–4012, 2020.
- [29] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, et al., “Language models are unsupervised multitask learners,” OpenAI blog, vol.1, no.8, p.9, 2019.
- [30] S. Black, S. Biderman, E. Hallahan, Q. Anthony, L. Gao, L. Golding, H. He, C. Leahy, K. McDonnell, J. Phang, M. Pieler, U.S. Prashanth, S. Purohit, L. Reynolds, J. Tow, B. Wang, and S. Weinbach, “GPT-NeoX-20B: An open-source autoregressive language model,” Proceedings of the ACL Workshop on Challenges & Perspectives in Creating Large Language Models, 2022.