

機関番号	研究種目番号	審査区分番号	整理番号
13901	14	1002	0005

令和8(2026)年度 研究活動スタート支援 研究計画調書

令和 8 年 4月28日
2版

新規

研究種目	研究活動スタート支援						
審査区分	人間情報学、応用情報学およびその関連分野						
研究代表者 氏名	(フリガナ)	キムラ ユウスケ					
	(漢字等)	木村 優介					
所属研究機関	名古屋大学						
部 局	情報学研究科						
職	研究員						
学 位	博士（文化情報学）						
エフォート	30%						
応募要件	(A)令和7(2025)年9月18日以降に科学研究費助成事業の応募資格を得、かつ文部科学省及び日本学術振興会が公募を行う研究種目に応募していない者						
研究課題名	未十分学習ドメインへの言語モデル軽量適応に向けた学習時更新資源配分制御の研究						
研究経費 〔千円未満の端数は切り捨てる〕	年度	研究経費 (千円)	使用内訳(千円)				
			設備備品費	消耗品費	旅費	人件費・謝金	その他
	令和8年度	1,913	1,141	7	473	0	292
	令和9年度	1,005	0	0	473	0	532
	総計	2,918	1,141	7	946	0	824
開示希望の有無	審査結果の開示を希望する						

1 研究目的、研究方法など

本研究計画調書は「令和 8（2026）年度研究活動スタート支援 審査区分表」の審査区分で審査される。記述に当たっては、「科学研究費助成事業における審査及び評価に関する規程」（公募要領参照）を参考にすること。

本研究の目的と方法などについて、3 頁以内で記述すること。

冒頭にその概要を簡潔にまとめて記述し、本文には、(1) 本研究の学術的背景や本研究の着想に至った経緯、研究課題の核心をなす学術的「問い」、(2) 本研究の目的及び学術的独自性と創造性、(3) 関連する国内外の研究動向と本研究の位置づけ、(4) 本研究で何をどのように、どこまで明らかにしようとするのか、(5) 本研究の目的を達成するための準備状況、について具体的かつ明確に記述すること。

（概要）

本研究は、事前学習および既存のポストトレーニングで十分に経験されていない専門的分布を「未十分学習ドメイン」と捉え、そのようなドメインに対する大規模言語モデル（LLM）の軽量適応を、個別の PEFT 手法の選択問題ではなく、**限られた追加更新をどのトークン由来の学習信号として、どの層・どの追加パラメタへ反映するかという更新資源配分の問題**として定式化する。医療・法務・金融などでは、機密性、計算資源、既存推論基盤との互換性の制約から、ベースモデル本体を更新せずに専門化したい場面が多い。しかし既存の PEFT 手法では、精度向上のために adapter 容量や rank の増加、推論時 routing、動的重み生成などを導入する方向に進みやすく、追加パラメタ量または推論時コストとのトレードオフが生じる。そこで本研究では、推論時のモデル構造・パラメタ数・実行手順を既存の軽量適応手法の枠内に保ったまま、学習時のみトークンおよび層ごとの更新配分を制御する。代表的な PEFT 手法を基準条件とし、追加パラメタ量を増やす条件、推論時に選択機構を導入する条件、およびベースモデル本体を更新する条件との比較を通じて、**既存 PEFT で十分な条件、学習時の更新重点化で改善可能な条件、ベースモデル改良や構造拡張が必要となる条件**を切り分ける。これにより、未十分学習ドメインへの軽量適応において、何をどの層・どの追加パラメタへ書き込むべきかという設計原理と、推論基盤を大きく変更せずに専門化精度を高めるための実証的判断基準を提示する。

（本文）

【(1) 学術的背景と着想に至った経緯、および研究課題の核心をなす学術的問い】 近年の LLM は、多様な自然言語処理課題に高い性能を示しているが、医療・法務・金融・低資源言語のような、事前学習や既存のポストトレーニングで十分に経験していない分布では、そのままでは十分な性能が得られないことが多い。このような未十分学習ドメインでは、継続事前学習やフルファインチューニングによりベースモデル本体を改良することが本筋となる場合がある。一方、実際の導入現場では、機密性、計算資源、既存推論基盤との互換性の制約から、ベースモデル本体を更新せずに適応したい場面も多い [1, 2]。

既存の PEFT は、少数の追加パラメタのみを学習することで、この要求に応えてきた。しかし一般に、精度を上げようとすると rank を増やす、adapter 数を増やす、推論時に routing や expert 選択を導入するなど、追加パラメタ量あるいは推論時コストを増やす方向に進みやすい。すなわち、軽量適応では、**追加パラメタサイズと精度向上の間に強いトレードオフ**が存在する [2, 3]。

応募者はこれまで、専門分野の文書分類を対象に、追加の人手アノテーションを必要としない補助タスク設計と、トークン単位の重点化によるドメイン適応を研究してきた。博士論文では、どのようなトークンが重点的に学習されるべきかという観点から、補助タスクが有効となる条件を整理している [7, 8]。この知見は、**未十分学習ドメインへの軽量適応では、限られた更新をどこへ書き込むべきかが本質ではないか**という着想につながった。

本研究の核心をなす学術的問いは、「未十分学習ドメインへの言語モデル適応において、ベ-

【1 研究目的、研究方法など（つづき）】

スモデル本体を更新しない条件の下で、限られた追加更新をどのトークンに由来する更新として、どの層へ重点配分すれば、高精度な適応を実現できるのか」である。

【(2) 研究目的及び学術的独自性・創造性】 本研究の独自性は、軽量適応の問題を「どの PEFT 手法を選ぶか」という手法比較としてではなく、**未十分学習ドメインにおいて、限られた追加更新がどこへ配分されるべきかという更新資源配分の問題として定式化する点にある**。既存研究では、rank の増加、adapter 数の増加、推論時 routing の導入、shared adapter の変調など、多様な改良法が個別に提案されている。しかし、それらは主として「どの手法が高性能か」を示すものであり、**既存 PEFT がどの条件でなぜ不足するのか、またどの条件で学習時制御だけで十分かは十分に整理されていない** [2, 3]。

これに対し本研究は、まず PEFT の失敗を、単なる追加パラメタ量の不足ではなく、**重要なトークンや task 適応に敏感な層に更新が十分に届いていないこととして捉える**。その上で、どのトークン/層への重点配分が性能改善に結びつくかを診断し、**標準 PEFT で十分な条件、学習時の更新重点化が必要な条件、さらに最小限の構造拡張が必要な条件を切り分ける**。すなわち、本研究の創造性は、新しい手法を一つ追加することではなく、**未十分学習ドメインへの軽量適応において、何をどこへ書き込めばよいかという設計原理と判断基準を与える点にある**。

【(3) 関連する国内外の研究動向と本研究の位置づけ】 関連研究は、大きく三つの流れに整理できる。第一は、未十分学習ドメインへの適応において、継続事前学習や full fine-tuning が本質的に重要であることを示す研究である。これらは、ベースモデル側の知識不足が支配的な条件を見極める上で重要である [1]。第二は、LoRA を中心とする PEFT 改良の研究である。LoRA-Pro は LoRA を低ランク勾配最適化として捉えた手法 [3] であり、MTL-LoRA は複数タスクが混在した条件でのタスク間の悪影響を緩和するために、一つの A 行列に対して複数の B 行列などを導入している [4]。さらに Sine-LoRA は、低ランク行列の表現力自体を高める方向を示している [5]。第三は、TopLoRA に代表される、トークン単位の選択性に着目した手法である。TopLoRA は、全トークンが同一射影を共有する LoRA の限界に着目し、入力トークンごとに異なる入出力の射影を学習することで、同一 rank でも LoRA を上回る性能を報告している [6]。

なかでも TopLoRA は本研究に最も近い先行研究である。ただし TopLoRA のトークンの選択性は、入力トークンに応じた動的な重み生成によって実現されており、ファインチューニング後に事前学習重みに直接統合できず、推論時追加計算と推論遅延を伴う [6]。したがって、トークンごとの選択性が有効であること自体は既に示されている一方で、その利益を推論時の動的変調なしにどこまで回収できるかは未解明である。

これに対し本研究は、トークンごとの選択性が有効であるという問題意識を受け継ぎつつ、その実現を推論時の動的重み生成ではなく、学習時の更新資源配分制御に求める点に独自性がある。すなわち本研究は、どのトークンに由来する更新を、どの層の adapter に重点的に書き込むかを制御し、順伝播・推論インタフェースは LoRA と同一に保つことを目指す。したがって本研究は、LoRA の一改良を目指すものではなく、未十分学習ドメインへの軽量適応を、単なる rank 不足ではなく更新配分の不適合として再定式化し、推論時不変な条件の下で精度-追加パラメタ trade-off をどこまで緩和できるかを問う研究として位置づけられる。

【(4) 本研究で何をどのように、どこまで明らかにしようとするのか】 本研究では、次の三つの問いを中心に据える。第一に、**軽量適応における精度低下の主因は何か**、という問いである。精度が上がらないのは、単純に追加パラメタ量が不足しているからなのか、それとも限られた更新が本来重要なトークンや層に十分配分されていないからなのかを切り分ける。第二に、**どのトークンとどの層に更新を重点配分すべきか**、という問いである。予測確率、surprisal、初期勾配統計などを手がかりとして、追加適応の余地が大きいトークンと、task 適応に敏感な層を同定し、それらへの重点配分が性能改善に結びつくかを検証する [8]。第三に、**推論時を一切変えずに、どこまで精度と追加パラメタのトレードオフを緩和できるか**、という問いである。この問いに対し、

【1 研究目的、研究方法など（つづき）】

本研究では、まず通常の LoRA を基準とし、学習時のみトークン/層ごとの勾配配分を制御する方法を検証する。それでも不十分な条件に限って、共有 adapter の変調や最小限の構造拡張を発展的に比較する。

これらの問いに答えるため、本研究は段階的に研究を進める。第1段階では、代表的な設定として、医療や法務等の専門分野の QA を対象に、LoRA とベースモデル本体も更新する比較条件を比べ、ベースモデル本体を固定した条件で本当に追究すべき余地がどこにあるかを明らかにする [1, 3]。第2段階では、7B 級オープンモデルを主対象として、通常の LoRA を基準に、トークン/層単位の更新重点化を導入する。ここでは、順伝播は通常の LoRA と同一に保ち、逆伝播時のみ勾配スケールリングや stop-gradient 等を用いて更新の流れを制御する。この段階では、追加パラメタサイズを固定したときの精度改善量、最も弱い task の改善、task 間ばらつきの変化を評価する。第3段階では、この更新重点化を複数領域・複数 seed で再現確認し、一般性と失敗境界を整理する。さらに、学習時の更新制御だけでは不足する条件に限って、共有 adapter の変調や最小限の構造拡張を発展的に検討し、**学習時のみの選択性で足りる条件と、追加構造が必要になる条件**を明らかにする。

以上により、本研究では、(i) ベースモデル改良が本質的に必要な条件、(ii) ベースモデル本体を固定したまま更新重点化で改善可能な条件、(iii) そのときトークン/層ごとの重点化がどのように効くかを整理し、軽量適応における精度-追加パラメタ trade-off を緩和するための実証的な判断基準を提示する。

【(5) 研究目的を達成するための準備状況】 応募者は、文書分類におけるマルチタスク学習の条件分析、L3Masking によるトークン単位の重点化、LoRA マルチタスク学習における更新空間の分離など、本研究に直結する知見を既に有している。博士論文では、補助タスクが有効となるトークン条件を整理しており、これは「どこを重点的に学習させるべきか」という本研究の視点の予備的整理に相当する [7, 8]。

また、応募者は現在、学習時のみ重要トークン/層に更新を集中させ、推論時の構造・手順を一切変えない軽量適応法を試作している。予備実験では、数学 QA および医療 QA の設定において、既存の LoRA 系ベースラインを上回る初期結果を確認しており、ベースモデル本体を固定した条件でも学習時の更新制御に改善余地があることを把握している。

さらに、応募者は Miyabi や TSUBAME 等のスパコンを用いた共同研究課題を通じて、大規模言語モデルを限られた計算資源下で扱うための実装・運用経験を有している。そのため、本研究における基礎比較、トークン/層ごとの更新重点化、複数 seed による再現確認を段階的に進める現実的な見通しを持つ。

参考文献

- [1] Gururangan, S., Marasovic, A., Swayamdipta, S., et al., “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks,” In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8342–8360, 2020.
- [2] Hu, E. J., Shen, Y., Wallis, P., et al., “LoRA: Low-Rank Adaptation of Large Language Models,” In *The Tenth International Conference on Learning Representations*, 2022.
- [3] Wang, Z., Liang, J., He, R., Wang, Z., and Tan, T., “LoRA-Pro: Are Low-Rank Adapters Properly Optimized?” In *The Thirteenth International Conference on Learning Representations*, 2025.
- [4] Yang, Y., Muhtar, D., Shen, Y., et al., “MTL-LoRA: Low-Rank Adaptation for Multi-Task Learning,” *arXiv preprint arXiv:2410.09437*, 2025.
- [5] Ji, Y., Saratchandran, H., Gordon, C., Zhang, Z., and Lucey, S., “Efficient Learning with Sine-Activated Low-Rank Matrices,” In *The Thirteenth International Conference on Learning Representations*, 2025.
- [6] Li, S., Luo, X., Wang, H., et al., “Beyond Higher Rank: token-wise Input-Output Projections for Efficient Low-Rank Adaptation,” In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- [7] Kimura, Y., Komamizu, T., and Hatano, K., “An Automatic Labeling Method for Subword-Phrase Recognition in Effective Text Classification,” *Informatica*, 47(3), 315–326, 2023.
- [8] Kimura, Y., Komamizu, T., and Hatano, K., “L3Masking: Multi-task Fine-tuning for Language Models by Leveraging Lessons Learned from Vanilla Models,” In *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)*, pp. 53–62, 2024.

2 応募者の研究遂行能力及び研究環境

応募者の研究計画の実行可能性を示すため、(1)これまでの研究活動（主要な研究業績を含む）、(2)研究環境（研究遂行に必要な研究施設・設備・研究資料等を含む）について2頁以内で記述すること。

「(1)これまでの研究活動」の記述には、研究計画に関連した国際的な取組（国際共同研究の実施歴や海外機関での研究歴等）がある場合には必要に応じてその内容を含めること。また、研究活動を中断していた期間がある場合にはその説明などを含めてもよい。

【(1) これまでの研究活動】 応募者はこれまで、自然言語処理におけるマルチタスク学習、文書分類、ドメイン適応、および大規模言語モデルの軽量適応に継続して取り組んできた。とくに一貫して扱ってきた問いは、限られた学習資源の下で、どの情報を重点的に学習させるべきか、という点にある。博士論文では、専門文書分類を対象に、追加の人手アノテーションを必要としない補助タスク設計とトークン単位の重点化を検討し、その成果を査読付き論文および国際ワークショップ論文として公表してきた [7, 8]。

具体的には、文書分類における汎用的サブタスクの有効条件の分析、サブワード句認識の自動ラベリング法、およびドメイン適応のためのトークン単位マスク戦略を通じて、どのトークンが主タスク適応に寄与するかを一貫して検討してきた [7]。さらに L3Masking では、事前学習済み言語モデルの予測特性を踏まえて重点的に学習すべきトークンを選別する考え方を提示しており、本研究のトークン側の問い、すなわち「どのトークンに由来する学習信号を優先すべきか」を考える直接の基盤となっている [8]。

また最近では、LoRA マルチタスク学習における更新空間の分離による学習偏りの抑制にも取り組んでおり、これは本研究の層側の問い、すなわち「その学習信号をどこへ書き込むべきか」に直結する。したがって本研究は、補助タスク設計、トークン 重点化、LoRA 更新空間という三つの蓄積を、LLM の軽量適応へ接続した自然な発展に位置づける。

研究の進め方の面でも、応募者は単に新手法を提案するだけでなく、データ条件、モデル条件、評価条件を統制しながら複数条件を比較し、性能差を有効条件・失敗条件として解釈する研究スタイルを培ってきた。本研究でも同様に、どの条件で標準 PEFT で十分か、どの条件で学習時更新制御が有効か、どの条件でベースモデル改良や構造拡張が必要かを切り分けることを狙うため、この比較実験設計能力はそのまま研究遂行能力の中核となる。

加えて、応募者は量子化スケジューラによる効率的な LLM アダプタ学習や、学習進度に応じた層別動的量子化に基づく高効率アダプタ学習法に関する共同研究課題にも取り組んでおり、計算資源制約下で LLM を実験的に扱う経験を有している。また、SPRING 支援対象学生、若手研究者育成フェロシップ制度の支援対象学生、NTT コミュニケーション科学基礎研究所でのインターンシップを通じて、独立して研究計画を立案・遂行し、外部研究者との議論を通じて課題設定を洗練させる経験を積んできた。さらに、査読付き国際ワークショップ CustomNLP4U 2024 での発表や EMNLP Workshop Reviewer の経験は、国際的な研究コミュニティとの接点として位置づけられる [8]。

主要な研究業績は以下のとおりである。とくに 1, 2 は査読付き国際会議および国際ジャーナル、3, 4 は本研究計画に最も近い国内業績であり、いずれも本研究と直接連続している。

1. Kimura, Y., Komamizu, T., & Hatano, K. (2024). L3Masking: Multi-task fine-tuning for language models by leveraging lessons learned from vanilla models. In *Proceedings of the 1st Workshop on Customizable NLP: Progress and Challenges in Customizing NLP for a Domain, Application, Group, or Individual (CustomNLP4U)* (pp. 53–62). Association for Computational Linguistics.
2. Kimura, Y., Komamizu, T., & Hatano, K. (2023). An automatic labeling method for subword-phrase recognition in effective text classification. *Informatica*, 47(3), 315–326.

【2 応募者の研究遂行能力及び研究環境（つづき）】

3. 木村 優介, 駒水 孝裕, 波多野 賢治 (2024). ドメイン適応のためのトークン単位の擬似尤度に基づくマスク戦略. 情報処理学会研究報告, 154(1), 1-6.
4. 平子 翔太, 木村 優介, 波多野 賢治 (2026). LoRA マルチタスク学習における更新空間の分離による学習偏りの抑制. 情報処理学会研究報告, 2026(4), 1-6.

以上より、応募者は、未十分学習ドメインへの適応に関する継続的な研究蓄積、条件統制を重視した比較実験の設計能力、国際発信の経験、および言語モデルの軽量適応を実装・評価するための技術的基盤を有しており、本研究を主体的に遂行する能力を備えている。

【(2) 研究環境】 応募者は、2026 年 4 月より名古屋大学大学院情報学研究科に所属し、知能情報学、機械学習、自然言語処理に関する議論が可能な研究環境にある。所属先では、モデル設計、学習アルゴリズム、応用自然言語処理にまたがる研究交流が可能であり、本研究課題のように、理論的な設計判断と比較実験の両方を要するテーマを進めるうえで適した環境が整っている。また、JST 戦略的創造研究推進事業研究員として研究に従事しており、研究時間を確保しやすい環境にある。

応募者はすでに、公開データセットを対象とした教師ありファインチューニングおよび PEFT 用のデータ整形、評価スクリプト、誤り分析環境、実験ログ管理基盤を整備しており、既存コードの再利用により初年度から比較実験を開始できる。特に、本研究では LoRA を基準としてトークン/層ごとの更新制御を段階的に追加するため、ベースライン実装、学習条件の統一、複数 seed の再実行を同一パイプラインで扱えることが重要であるが、そのための実装基盤は既に確立している。

計算資源については、学内資源として名古屋大学情報基盤センターの次期スーパーコンピュータ「不老・式」を運用開始後に活用する予定である。加えて、学外資源としては JCAHPC の Miyabi、東京科学大学の TSUBAME4.0、および東京大学情報基盤センターの共同利用資源の活用を予定している。応募者は、既に東京科学大学 TSUBAME 若手・女性利用者支援制度の対象、学際大規模情報基盤共同利用・共同研究拠点の萌芽型共同研究課題、東京大学情報基盤センター若手・女性利用者推薦による研究課題を通じて、外部計算資源を確保・活用する経験を有している。したがって本研究でも、必要な計算資源の申請・運用を具体的に進められる見通しがある。

もっとも、本研究は資源依存的に過度に膨張する計画ではない。主対象を 7B 級オープンモデル 1 種に絞り、比較対象、データ規模、seed 数を段階的に制御する設計としているため、大規模 sweep や広範なモデル比較に依存せずに、本質的な仮説検証を進めることができる。具体的には、既存環境で実装と予備比較を行い、不老・式や Miyabi、TSUBAME4.0 では複数 seed による再検証と追加条件の評価を行うという分担が可能である。このように、本研究は計算資源が豊富なほど望ましい一方で、資源が一時的に制約されても本質的な検証が失われない縮退可能な計画として設計されている。

さらに、本研究で重視するのは、大規模学習を回すこと自体ではなく、条件統制の下で結果を解釈可能にすることである。そのためには、ドメインギャップの整理、評価指標の選択、結果のばらつきの解釈、失敗事例の分析が不可欠である。応募者はこれまでの研究を通じてこのような統制的・分析的アプローチを継続してきたため、現在の研究環境は、その研究スタイルを維持しながら、言語モデルの軽量適応という新しい対象へ展開するのに適している。

以上より、応募者は、本研究課題に必要な学術的蓄積、実験設計能力、研究環境、計算資源運用経験を備えており、本研究を着実に遂行できる。

3 人権の保護及び法令等の遵守への対応（公募要領参照）

本研究を遂行するに当たって、相手方の同意・協力を必要とする研究、個人情報の取扱いの配慮を必要とする研究、生命倫理・安全対策に対する取組を必要とする研究など指針・法令等（国際共同研究を行う国・地域の指針・法令等を含む）等に基づく手続が必要な研究が含まれている場合、講じる対策と措置を、1 頁以内で記述すること。

個人情報を伴うアンケート調査・インタビュー調査・行動調査（個人履歴・映像を含む）、提供を受けた試料の使用、ヒト遺伝子解析研究、遺伝子組換え実験、動物実験など、研究機関内外の倫理委員会等における承認手続が必要となる調査・研究・実験などが対象となります。

該当しない場合には、その旨記述すること。

本研究では、公開されている言語モデル、公開データセット、およびそれらに基づく派生データを用いる。人を対象とする実験は行わない。

学習および評価に用いるデータについては、利用規約、ライセンス、配布元の定める条件を遵守する。公開コーパスに個人情報やセンシティブな情報が含まれる場合には、配布元の条件に従い、必要に応じて匿名化済みデータまたは利用が許諾されたデータのみを用いる。研究成果の公表にあたっては、特定の個人や集団に不利益を与える表現を避け、モデルの偏りや限界についても適切に記述する。研究データ、実験ログ、学習済み追加パラメタ、評価コード等は所属機関の規程に従って適切に保存・管理する。

[illegible]

設備備品費、消耗品費の必要性

設備備品費として、外部計算資源（スパコン）および研究室ワークステーションを効率的に利用するための研究用ノートPC（コーディングや分析のためにRAM は32GB必要・Lenovo ThinkPad X9 15 Gen 1）と、予備実験・実装検証・前処理・小規模アプリケーションを行うためのGB10 搭載のGPUワークステーション（コニファイドメモリが非常に大きいため、LLM に対する提案手法の基礎分析や本実験の検証が可能・Lenovo ThinkStation PGX）を購入する。ノートPCは、大規模学習そのものを担う機器ではなく、スパコンやワークステーションに接続してジョブ管理、結果確認、解析、図表作成、論文執筆を行うためのクライアントPCとして用いる。また、国際会議や国内会議での出張先でも研究を行うためにクライアントPCはデスクトップPCではなく、ノートPCを選択する。GPUワークステーションは、スパコン利用前の条件確認や不具合検出を迅速に行い、外部計算資源の有効活用に資する。消耗品費としてモニターやプロジェクター接続用周辺機器（Lenovo USB-C 7-in-1 ハブ）を購入し、発表用に用いる。

研究活動スタート支援 8 - (1)

(金額単位：千円)

[illegible]

旅費、人件費・謝金、その他の必要性

旅費として、2026・2027年度に国内会議各1回、国際会議各1回に参加し、研究成果を発表するとともに、関連分野の最新動向を収集する。人件費・謝金は計上せず、データ整備、実験実施、評価、誤り分析、論文執筆までを研究代表者自身が一貫して行う。その他の経費として、Miyabi（初年度は TSUBAME より安い）・TSUBAMEの大規模実験用の外部計算資源利用料にあてる。