

研究終了報告書

「高速かつ高性能な広範にわたる逐次的意思決定問題の方策開発と解析」

研究期間： 2021 年 10 月～ 2024 年 3 月

研究者： 土屋 平

1. 研究のねらい

逐次的意思決定問題とは、逐次的に得られる情報をもとに意思決定者がある種の意味決定を行う数理モデルである。代表的な逐次的意思決定問題の 1 つにバンディット問題が挙げられる。この問題では、複数の行動の選択肢から行動を 1 つ選択し、選択した行動についての損失のみが観測できる状況で、選択した行動により得られる累積での損失を最小化する問題である。この問題の応用範囲は幅広く、推薦システムやオンライン経路制御などの多くの重要な応用が存在する。

近年の逐次的意思決定問題では、事前に与えられたある種の最適性を達成することに焦点を当て、多くの理論的に好ましい性質を持つアルゴリズムが考案されてきた。しかし、それらのアルゴリズムは複数の理由から、実問題で利用するための十分な性質を有していない場合が多い。例として、毎時刻非常に計算量の大きい最適化問題を解く必要があり中規模以上のデータに対して利用が困難である場合や、数値的なシミュレーションでは数値性能が非常に悪い場合、あるいは、仮定されている問題の構造が非常に単純であり実問題を十分にモデル化できていない場合がある。以上の課題により、理論的に優れたアルゴリズムが十分に普及するには至っていない。

本研究のねらいは実問題において重要と考えられるある種の最適性を持つアルゴリズムにおけるこれら 3 つの課題を解決し、より理論的に優れたアルゴリズムを社会に普及させることである。本研究では、最適性の指標として、両環境最適性を考える。バンディット問題には、観測がある確率分布に従って生成される確率的設定と、意思決定者にとって都合が悪いように生成される敵対的環境の大きく異なる 2 つの問題設定がある。しかし、事前に実問題がどちらの環境に属するのかを判断することは難しく、単一のアルゴリズムで確率的環境と敵対的環境において同時に最適性を達成するアルゴリズムを構築することが望ましい。このようなアルゴリズムは、両環境最適アルゴリズムと呼ばれ、確率的環境・敵対的環境だけでなく、これらの中間の環境である基本的には確率的であるが汚染の入った観測が得られる環境でも良好に動作することが知られている。本研究では、高速かつ高性能な広範の問題に適用可能な両環境最適アルゴリズムの構築することで、既存のアルゴリズムが抱える 3 つの課題を解決することを目指す。

2. 研究成果

(1) 概要

本研究では主に両環境最適アルゴリズムの高性能化と実問題をモデル化可能な複雑な問題への拡張に取り組んだ。複雑な問題への拡張については、行動の選択肢が組合せ的な構造を持つ組合せバンディット問題、及び、損失が直接観測される代わりに抽象的な観測が得られる部分観測問題において両環境最適アルゴリズムを構築した。高性能化については、主に両環境最適アルゴリズムが、それと同時にデータ依存上界と呼ばれる、観測系列の性質に依

存した理論保証を与えることで実現した。

(2) 詳細

本研究では、高性能かつ広範にわたる逐次的意思決定問題のアルゴリズム開発と解析を推し進めた。逐次的意思決定問題の目標の中でも、累積損失最小化(リグレット最小化)を目的とするアルゴリズムに焦点を当てる。ここで、リグレットとは、選択した行動によって被った累積損失と最適な行動の累積損失の差である。

研究開始当初の目標は、高速かつ広範の問題に利用可能な漸近最適性を有するアルゴリズムの構築であった。ここで、あるアルゴリズムが漸近最適であるとは、時刻が十分経過したときに、単位時間あたりのリグレットの増大が、与えられた行動集合とパラメータから定まるある下界を達成する、ということの意味する。特に行動の選択肢同士で相関がある問題においては、漸近最適性は、リグレットを押さえつつも、与えられた行動の集合からいかに未知パラメータの情報を効率的に得ているかの指標になり、漸近最適性を有するアルゴリズムを構築することは重要な研究課題だと考えられた。1年目から2年目の前半にかけた研究では、主にこの問題に取り組んだが、数値的性能と頑健性の問題が生じた。具体的には、漸近最適なアルゴリズムが漸近最適ではないすでに普及しているアルゴリズムに優位性を示すことができるのは、非現実的な問題設定であった。また、漸近最適性を考える時には、確率的環境を前提とするが、漸近最適なアルゴリズムは確率的環境に少しでもノイズが乗った場合に大きく性能が悪化することが明らかになった。以上の2点から、漸近最適性を有するアルゴリズムが真に実用的なのかに関する疑問が生じ、実問題において有用な他の最適性の指標について考えることにした。

【両環境最適性: 確率的環境と敵対的環境における同時最適性】

そこでアルゴリズムの性能の指標として両環境最適性を考える。バンディット問題を含む多くの逐次的意思決定問題においては確率的環境と敵対的環境と呼ばれる2つの理論的枠組みがある。確率的環境は観測がある確率分布に従って得られると仮定する楽観的な設定であり、一方、敵対的環境は任意の有界な観測系列を仮定する

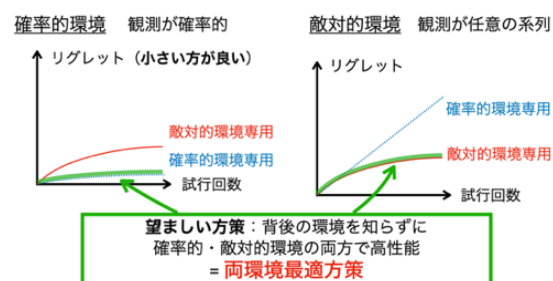


図 1 両環境最適アルゴリズム

非常に悲観的な設定である。確率的環境と敵対的環境で大きく異なるアルゴリズムが考案されてきた。しかし、実問題においては、背後の環境が確率的であるか敵対的であるかを事前に知ることは難しい。さらに、確率的環境専用のアルゴリズムは敵対的環境で全く動作せず、一方で敵対的環境専用のアルゴリズムは確率的環境において非常に性能が悪いことが知られている。そこで、単一のアルゴリズムで確率的環境・敵対的環境で同時に最適性を達成するアルゴリズムが望まれ、このようなアルゴリズムは両環境最適アルゴリズムと呼ばれている(図 1)。両環境最適アルゴリズムは実問題で最も想定される確率的環境と敵対的環境の中間の環境(ノイズのある確率的環境)においても、ほぼ最適な性能を達成することが知られており、両環境最適性は実問題における高性能なアルゴリズムの条件として重要な性質であると考えられる。

両環境最適アルゴリズムは Follow-the-Regularized-Leader (FTRL) と呼ばれるオンライン最適化の枠組みを用いて実現可能であることが知られている。FTRL はこれまでの累積損失と

適切に定めた強凸な正則化関数を最小化するように行動選択確率を決定する枠組みである。両環境最適性は、FTRL の凸正則化関数やその係数である学習率を観測データに対して適応的に定めることで実現可能であることが知られている。本研究では、問題の構造や活用したい観測データの構造に応じて適切に正則化関数と学習率を設計することで所望の性質を達成することを目指す。

【研究 1. 広範な問題で利用可能なアルゴリズム: 組合せ半バンディット問題と部分観測問題】

両環境最適性は、エキスパート問題や多腕バンディット問題などの比較的単純な逐次的意思決定問題においては数多くの研究がなされてきた。一方で実問題に現れるような複雑な問題においては十分研究がなされておらず、より実問題に近い複雑な構造を持つ逐次的意思決定問題において両環境最適性を持つアルゴリズムを実現可能であるかは十分に知られていなかった。そこで 1 つ目の研究では、行動の選択肢に組合せ的な構造が入った問題や抽象的なフィードバックしか得られない設定において両環境最適アルゴリズムを構築することを目指した。

本研究では、行動の選択肢に組合せ構造が入った組合せ半バンディット問題や抽象的なフィードバックから意思決定を行う非常に一般的な逐次的意思決定問題の枠組みである部分観測問題において、両環境最適方策を実現可能であることを明らかにした(成果論文 1, 2, 3)。以下では、特に部分観測問題における両環境最適アルゴリズムの実現法についての詳述する。

部分観測問題では、直接損失が観測される代わりにそれらの損失に対して部分的な情報を与えるフィードバック記号のみから意思決定を行う問題の枠組みである。この問題は、図 2 に示すように、多腕バンディット問題や動的価格設定、効率的ラベル予測など、多くの代表的な逐次的意思決定問題特別なケースとして含む非常に一般的な問題である。

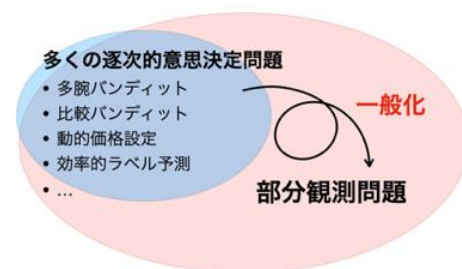


図 2 部分観測問題

部分観測問題の重要な性質は、その行動を選択することにより得られる損失は大きい一方で、他の行動の損失について多くの情報を持つような行動を考慮する必要がある点にある。このような問題に対処する最近考案された方法は、これまでの累積損失のみに基づいた FTRL の出力を直接行動選択確率として採用する代わりに、累積損失に加えて各行動の選択によって得られる情報量を考慮した凸最適化問題を解くことで FTRL の出力を別の確率分布に変換することである。しかし、この変換は両環境最適性を示すための標準的なテクニックであるリグレットの下界を用いた方法と同時に利用することができない。そこで本研究では、FTRL の出力から行動選択確率へ変換を定める最適化問題において、その実行可能領域を制限することを考えた。これは最小化問題における実行可能領域の制限のため、一見性能が悪化するように見えるが、最適解の性質を解析することで最悪ケースの最適値が悪化しないように実行可能領域の制限が可能であることを明らかにした。これによって、部分観測問題における初めての両環境最適アルゴリズムを実現することに成功した(研究成果 1)。さらに、本研究は正則化関数の更新を精緻化することで、最適化問題を解かずに高速に実行可能な枠組みも考案している。

【研究 2. 高性能な両環境最適アルゴリズム: 両環境最適性とデータ依存上界の同時実現】

本研究では、データ依存上界を同時に達成可能な両環境最適アルゴリズムの構築に取り組んだ。データ依存上界を有するアルゴリズムとは、敵対的環境において背後の損失系列が良

性である場合、その良性度合いに依存した性能を達成するアルゴリズムのことである。良性なデータの例として、ある行動を取り続けた場合の累積損失が非常に小さい場合(図 3 左)や観測値の変動が小さい場合(図 3 右)が挙げられる。これらの設定はそうでない設定と比較して、比較的容易で、リグレットを小さくしやすい。そこで、このような損失の性質の良さに依存したリグレット上界はデータ依存上界と呼ばれ、データ依存上界を持つアルゴリズムを構築できれば、対応する良性度合いを有する問題において小さなリグレットを達成することが期待される。そこで、本研究では両環境最適性の性能をさらに向上させるために、同時にデータ依存上界を達成可能なアルゴリズムの構築を目指す。

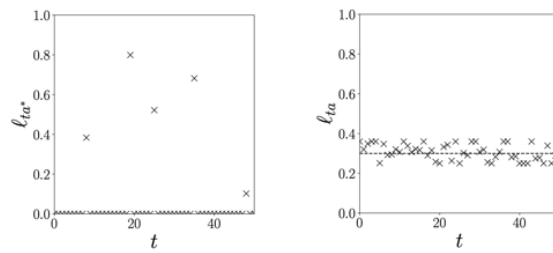


図 3 データ依存上界

本研究では、データ依存上界の中でも、損失ベクトルのスパース性の活用に注目した。損失のスパース性は多くの実問題に登場する。例として、手元にある広告からどの広告をウェブサイトに掲載するかを逐次的に決定し利益の最大化を目指すオンライン広告配置問題(図 4 左)ではほとんどの広告

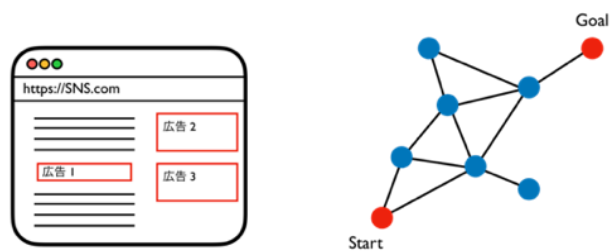


図 4 オンライン広告配置(左)とオンライン経路制御(右)

はクリックされることがなく、従ってほとんどの行動で得られる報酬値が 0 であるとみなすことができる。また、有向グラフが与えられ、そのグラフ上のあるスタートからゴールへ逐次的にデータを輸送するオンライン経路制御問題(図 4 右)では、ほとんどの経路でデータの損失が起きないと仮定できるとすると、損失ベクトルのほとんどの要素が 0 であるとみなせる。このように観測のスパース性を活用したデータ依存上界はスパース性依存上界と呼ばれる。

前述のように両環境最適アルゴリズムやこれらのデータ依存上界は FTRL と呼ばれる枠組みで達成可能であることが知られている。スパース性依存上界やゲーム依存上界は、FTRL の安定性と密接に関連することが知られており、FTRL 中の安定性を制御するように学習率を設計することでこれらの上界を導出できる。一方で、本研究の主要なテーマである両環境最適性を実現する代表的な方法の 1 つは、FTRL の正則化関数の強さに対して適応的な学習率を設計することである。そこで、両環境最適性とデータ依存上界を同時に実現するために、FTRL の安定性と正則化関数の強さに同時に適応的な学習率の枠組みの構築を行った。これによって、多腕バンディット問題では、スパース性依存上界を有する両環境最適アルゴリズムを実現した(成果論文 3)。構築した学習率の枠組みは非常に一般的であり、その変種は、研究 1 で議論した部分観測問題において、両環境最適性と意思決定者が実際に扱っている問題クラスの難易度への自動適応性(ゲーム依存上界)を達成することができる。また、研究 1 で議論した組合せ半バンディット問題において、最適アームの損失の大きさや損失の変動に対して適応的なデータ依存上界も導出した(成果論文 2)。

3. 今後の展開

今後の展開としては、それらは主に本期間中に完了することができなかった、より高性能であり広範な問題に利用可能な両環境最適の実現と、サーベイ論文執筆と数値実験による両環境最適性の利用可能性向上が挙げられる。

両環境最適方策の課題の残っている課題の 1 つはそれらが真に最適ではないということである。非最適性の例として試行回数 T に対する依存性が挙げられる。多くの逐次的意思決定問題においては、試行回数 T に対して、確率的環境では $O(\ln T)$ 、敵対的環境では $O(\sqrt{T})$ のリグレット上界を達成可能であり、これらの上界は T について最適であることが知られている。しかし、部分観測問題などの逐次的意思決定問題における両環境最適方策は、確率的環境における T に関する依存が $O((\ln T)^2)$ 以上であるものが多い。この非最適性は数値的な性能に大きな影響を及ぼし、 $O(\ln T)$ の上界を達成可能な方策と比較すると大きく性能が劣化することが知られている。このような非最適性を解消し、両環境最適アルゴリズムのさらなる性能向上を目指すことは重要な研究課題である。またよりアルゴリズムの適用可能性を広げるとために、毎時刻得られる文脈と呼ばれる情報を活用した枠組み文脈付バンディット問題における両環境最適アルゴリズムの実現は重要な研究課題である。

他の大きな今後の展開の 1 つとしては、両環境最適方策のサーベイ論文の執筆と包括的数値実験による性能評価によるユーザの利用可能性向上が挙げられる。機械学習のアルゴリズム、特にバンディット問題の方策は、必ずしも理論的に優れた方策が社会において普及するとは限らない。実問題においては、UCB 方策やトンプソンサンプリングと呼ばれるすでに普及した方策が用いられ、実応用の論文や他分野の研究者においてそれらの方策と比較して優れた理論保証を有する方策が利用されることは稀である。そこで今後の研究では、両環境最適アルゴリズムに関するサーベイ論文の執筆と包括的数値実験による性能評価を目指す。

4. 自己評価

当初の計画で改善を予定していたアルゴリズムの最適性の指標とは異なるものの、実問題で重要な両環境最適性に焦点を当て、アルゴリズムの性能の適用可能な問題の広さの観点から複数のアルゴリズムの開発とリグレット解析に成功した。これにより当初の計画は概ね達成できたと考えられる。本研究成果により、当初目的としていた従来の代表的なバンディットアルゴリズムに取って代わるようなアルゴリズムを開発できたと考えられる。研究費は主に研究代表者の研究成果の国際会議・国内会議での発表、および、フランスの研究機関である INRIA に滞在するために用いられた。ACT-X での領域会議を通じて隣接分野とのつながりが新たにでき、新たな共同研究のきっかけとなった。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 7件

1. T. Tsuchiya, S. Ito, and J. Honda, “Best-of-Both-Worlds Algorithms for Partial Monitoring,” In Proceedings of The 34th International Conference on Algorithmic Learning Theory (ALT 2023), pp.1484–1515, 2023.

部分観測問題は、損失が直接される代わりに、抽象的なフィードバック記号のみが観測されるもとでリグレット最小化を目指す、非常に一般的な逐次的意思決定問題の枠組みである。既存のほとんどの両環境最適方策は、完全情報下でのオンライン学習や多腕バンディット問題など比較的単純な設定のみを対象として構築されており、より複雑な問題において両環境最適方策が構築可能かは未知であった。そこで本論文では、それらの問題と比較して非常に複雑な逐次的意思決定問題である部分観測問題において両環境最適方策が構築可能であることを明らかにした。

2. T. Tsuchiya, S. Ito, and J. Honda, “Further Adaptive Best-of-Both-Worlds Algorithm for Combinatorial Semi-Bandits,” In Proceedings of 26th International Conference on Artificial Intelligence and Statistics (AISTATS 2023), pp.8117–8144, 2023.

組合せ半バンディット問題は、行動の選択肢の背後に組合せ構造を有する逐次的意思決定問題の一種である。本論文ではこの問題に対して、多くの適応性を有する両環境最適方策を構築した。中でも確率的環境において分散の小ささを活用可能にしたことは実用上重要である。実際、本問題の代表的な応用先であるオンライン広告配置やオンライン経路制御問題においては損失の分散が最悪ケースの場合と比較して非常に小さいことが想定される。本論文では実際に理論・実験の両面から分散の小ささを活用することで方策の性能が大きく改善することを明らかにした。

3. T. Tsuchiya, S. Ito, and J. Honda, “Stability-penalty-adaptive follow-the-regularized-leader: Sparsity, game-dependency, and best-of-both-worlds,” In Advances in Neural Information Processing Systems 36 (NeurIPS 2023), 2023. (2023 年 12 月に掲載予定)

研究成果の多くを占める両環境最適性を実現する最も基本的な学習の枠組みは follow-the-regularized-leader (FTRL) と呼ばれるオンライン凸最適化のための枠組みである。FTRL は両環境最適性の他にも多くの環境適応性を達成可能な枠組みであるが、既存の技術では複数の性質のうち1つの性質のみに依存した適応性のみが達成可能であり、複数の環境適応性を同時に実現することが困難であった。本論文では FTRL 中の最も重要なパラメータである学習率を複数の観測量に同時依存可能な枠組みを構築することで、観測量のスパース性や問題の難しさに対する適応性、両環境最適性などの複数適応性を同時に達成することを可能にした。

(2) 特許出願

研究期間全出願件数: 0 件(特許公開前のものは件数にのみ含む)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

受賞

- 学生最優秀プレゼンテーション賞, IBIS2022, 2022
- 学生優秀プレゼンテーション賞ファイナリスト, IBIS2023, 2023

招待公演

- “組合せ半バンディット問題における適応的 best-of-both-worlds 方策”, 第 22 回情報科学技術フォーラム(FIT2023), 2023