公開

研 究 終 了 報 告 書

「限られた情報に基づく統計的機械学習と数理最適化アルゴリズムの開発」

研究期間: 2020 年 11 月～2023 年 1 月
研 究 者: 黒木 祐子

1. 研究のねらい

膨大な量のデータが世の中に放たれている「ビッグデータ時代」の到来により，ビッグデータを利用した新たな技術革新が現代の我々の生活を取り巻いている．近年の著しい情報工学分野の研究発展により，大量のデータからの機械学習法は実応用レベルで高精度な学習を達成できることが示唆されている．しかし，多くの応用では大量の教師付きデータの入手は最初から得ることは難しい．特に実世界では不完全なデータやノイズの乗ったデータが散開しており，**「何を情報として選択し，学習データとするのか」**を議論することの重要性がより一層高まっている．このような背景から，比較的データの取れるドメインから逐次的にデータの観測を行うオンライン学習法の研究開発が望まれている．

統計的機械学習とは，得られた観測データからデータの背後に潜む確率的な規則を推定する数理技術である．大量のデータが満足に得られない，あるいは部分的な観測しか得られない場合でも有効である頑健な統計的機械学習法の開発は未だ挑戦的な研究課題の一つである．本研究は限られた情報に基づく統計的機械学習法の理論構築に向けて，**数理最適化理論および統計学的観点の両側面からの解決法の提案と理論解明**を目標とし，特にオンライン意思決定問題を主となる研究テーマとして扱っている．

ありうる選択肢候補の中から実際に試行したものについてのみ情報が得られる環境下での最適な意思決定を目指す問題は「バンディット問題」とよばれ，オンライン推薦システム，広告表示の最適化，薬の治験を始めとして実応用上重要な研究分野である．実世界における意思決定は，しばしば**組合せ的な構造**(例：推薦システムにおける類似商品の集合，通信ネットワークにおける接続形態，道路ネットワークにおける経路など)により特徴付けられている． ある組合せ構造をなす意思決定候補が与えられる場合のバンディット問題を，「組合せバンディット問題」と呼び，そのアルゴリズム的な性質の良さから実世界への応用は多岐に渡っている．

本研究のねらいは，実応用により即した**確率的オンライン意思決定モデルとそのアルゴリズム基盤の確立**である．**確率・統計学的な妥当性と数理最適化理論に基づく最適性の保証および計算効率性を同時に達成する学習アルゴリズム**を開発し，「限られた観測情報に基づく確率的組合せバンディット問題」という挑戦的な課題に対する理論限界の解明を目指す．

2. 研究成果

(1)概要
最適腕識別問題 (Best arm identification of multi-armed bandits)とは統計的機械学習分野

における重要なオンライン意思決定問題の一つである．(確率的)最適腕識別問題は，期待報酬が未知の確率分布に従う複数の選択肢(arm/action)が与えられている状況で,期待報酬が最大となる選択肢を学習するための逐次的意思決定問題として定式化できる．この最適腕識別問題は，オンライン推薦システム，広告表示の最適化，薬の治験を始めとして実応用上重要な研究分野であるとともに，統計的強化学習における最も単純な枠組みでもあるが故，情報理論・アルゴリズム理論的観点からも古くから現在に至るまで活発に研究されている分野である．

組合せ最適腕識別問題(Combinatorial pure exploration of multi-armed bandits)は，従来の最適腕識別問題を一般化した問題であり，ある組合せ構造をなす組合せ的選択肢の集合(Combinatorial action space)が与えられる．例えば，電子機器の周波数の割当や労働者のタスク割当などの「マッチング」，道路ネットワークや通信ネットワークなどのグラフ上の「木」や「パス」といった単一の腕だけでは捉えられない組合せ構造が組合せ的選択肢の集合として挙げられる．このように，現実的な意思決定問題をモデル化した組合せ最適腕識別問題は，その数学的，アルゴリズム的な性質の良さから実世界への応用は多岐に渡っている．さらに組合せ最適腕識別問題は，入力(グラフ上の枝重みなど)が不完全な場合の組合せ最適化問題とも密接に関わっており，その学習可能性の解明は数理最適化および確率的バンディットの双方の観点から重要な研究課題である．

本研究課題では，限られた観測情報に基づく組合せ最適腕識別問題に対して，統計量的かつ計算量的に効率の良いアルゴリズム開発および理論解明を目指す．単一腕を直接サンプルする既存研究に対して，本研究では選択した腕集合からのランダム報酬の和（全バンディット），さらに腕集合の部分集合のみからの和(部分観測)などの限られた観測情報からの学習アルゴリズムの開発を行う．理論解析として，高確率で最適な腕集合を出力するために必要なサンプルの数,つまり標本複雑度(Sample complexity)の上界を与える．標本複雑度の解析においては，最適値とある腕集合の報酬関数の差(Reward gap)や最適値と次に良い最適値の差(Minimal reward gap)への依存性を明らかにする．また，任意の多項式時間のバンディットアルゴリズムが必要とする標本複雑度の下界の解明を試みる．

本提案で扱う限られた観測情報に基づく組合せ最適腕識別問題は，高速なアルゴリズムの設計が実用上必要である．しかしながら従来の確率的多腕バンディット問題の研究コミュニティはこのような計算量的観点を度外視していた．従来の研究では，計算量を無視したアルゴリズムに対しての標本複雑度の下界のみが解析されていた．一方で多項式時間アルゴリズムの下界は未解明であり，この多項式時間アルゴリズムの解明は当該分野の理論研究の更なる発展に求められている．統計的観点から良い性質を持ったアルゴリズム設計と，計算量的にも効率の良いアルゴリズムの設計には，統計的学習理論だけでなく数理最適化理論に関する深い知見を要する．本提案では，その双方の観点から探求し，限られた情報に基づく高精度な学習アルゴリズム開発と理論解明に挑戦する．

（2）詳細

研究テーマ A「全バンディット組合せ最適腕識別」

背後にある報酬関数が線形であり，選択した行動からのランダムな報酬の線形和がフィード

バックとして観測できる場合のアルゴリズム開発をおこなった．真の値と推定値の差を表す信頼区間(Confidence bound)とその確率集中不等式は,不確実性を評価する上で重要な数理概念である．また，全バンディット観測を扱うために有用な推定方法は最小二乗推定量(Least square estimator)による．最小二乗推定量に対する信頼区間は，全バンディット観測の場合,一般に楕円型信頼区間(Confidence ellipsoids)と呼ばれ，期待報酬と推定報酬の差は楕円型ノルムで定義される．統計的に効率の良い，つまり標本複雑度を最小にするための腕の引き方の決定には実験計画(Experimental design)の考え方を用いる．組合せ的選択肢集合に対する実験計画デザイン手法はこれまでに提案されてこなかったが，本研究では組合せ実験計画(Combinatorial experimental design)に対する有効な手法の開発やその計算量的困難性についての考察を試みた．

具体的には，動的サンプル戦略を計算効率よく用いるために，指数個ある組合せ的アクション集合を多項式サイズのアクション集合に落とす準備段階を実行する．この準備のための実行では計算効率の良い静的サンプル戦略を用いる．静的サンプル戦略としては，楕円型信頼バウンドの体積を小さくする方向にアームを引いていく G-optimal design という実験計画の文脈で知られている戦略を用いる．しかし既存の G-optimal design の計算では指数時間を要してしまうため，サポートを多項式サイズに制限する場合の G-optimal design を用いる．サポートを制限してしまった場合，楕円型信頼バウンドの最大値が行動の次元で抑えられる保証がなくなってしまうが，本研究で新たな上界を作った．次に，多項式サイズの行動集合に対して動的サンプル戦略を用いて最適なアクションを出力するアルゴリズムを設計する．理論解析としては，1. 第一段階の出力の中に最適なアクションが高確率で含まれていること 2. 第二段階の出力が高確率で最適なアクションであること 3. アルゴリズム全体で要したサンプル数の上界　の保証を与えた．計算機実験を用いて，最適値と次に良い最適値の差に依存してアルゴリズムのサンプル数がどのように変化するのか，既存の指数時間アルゴリズムと比較してどれほど計算時間が改善されたのかを検証した．完全マッチングなどのインスタンスを用いて，これらの振る舞いを示し有効性を確認した．

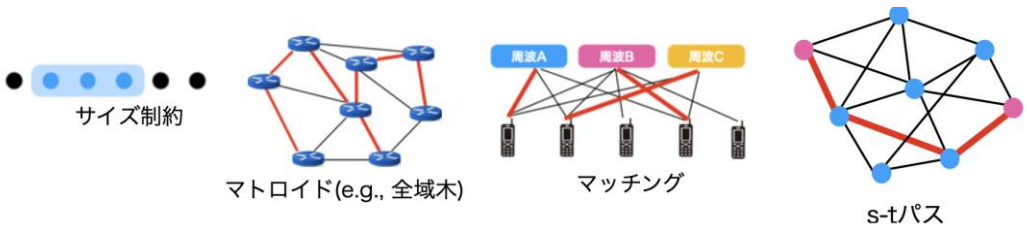

図1：組合せ的行動集合の例

研究テーマ B「部分観測組合せ最適腕識別」

前項の全バンディット及び選んだ要素全ての観測が可能な半バンディットを拡張，またさらに限定された観測を扱うことのできる部分観測設定に対して初となる手法を提案した．また非線形報酬関数を扱うための妥当な仮定として，関数の一様連続性を表すリプシッツ連続性(Lipschitz continuity)を用いる．推定量には最小二乗推定量の代わりにムーア-ペンロー

ズの擬似逆行列による推定量を用いる. 未知の期待報酬ベクトルを推定するのに十分な選択肢の集合を大域的観測可能集合(Global observable set)と呼ぶ. まずは静的サンプル戦略に基づいたアルゴリズムを提案し, 初となる標本複雑度の上界を与えた.

標本複雑度を小さく抑えるような大域的観測可能集合の最適な選び方を求めるアルゴリズムを設計し, そのアルゴリズムを元に効率の良いサンプリングアルゴリズムを提案しその理論解析については今後の課題とする. また, G 最適計画や E 最適計画においてデータ点が組合せ的特徴をもつような場合を考え, 計算複雑性と連続緩和した問題に対する次元に関する多項式時間アルゴリズムの設計と理論解析も重要な今後の課題である.

研究テーマ C「非線形報酬の場合の組合せ最適腕識別」

報酬関数が非線形の場合に対しては, 1. 線形関数の場合と同じような解析が適用可能なリプシッツ連続性を満たす非線形報酬関数のクラス, 2. ボトルネック報酬関数を考える (例: パスを構成する枝重みの中で最も小さい枝重みを最大にするパスを選択する問題など). これらの問題の組合せ最適腕識別問題としての定式化を考え, 多項式時間適応的アルゴリズムの提案および標本複雑度およびその下界の証明を与えた.

さらにその発展として, 再生核ヒルベルト空間で定義されるカーネル関数を考え, その関数の性質と最適腕識別の関係を探究した. 固定信頼度および固定予算設定の両方を考慮しアルゴリズム設計と理論解析を試みた. この設定では組み合わせ的な行動集合を考えることは難しく, K 個の行動集合がある場合に限定して議論した. また n 人のエージェントが類似するタスクを同時に解き, 情報交換しながら標本複雑度を最小にする設定への拡張もおこなった.

3. 今後の展開

-本研究では主に背後にある組合せ最適化は厳密に解けるか P T A S が存在するクラスを考えていたが, 多くの組合せ最適化問題は NP 困難である. 今後の展開としてはそのようなクラスの組合せ最適化問題が含まれるような組合せ最適腕識別問題において, 出力がある近似保証を達成するまでに必要な標本複雑度を最小化するアルゴリズムの枠組みを提案していきたい.

-逐次的に情報を収集する枠組みは, 多くの学習問題で重要な課題であると. 本研究課題では古典的な逐次的意思決定問題の定式化に習ったバンディット問題の範疇で研究を進めてきたが, 他の学習タスクにおいても統計的最適性を担保するような逐次的な情報選択の枠組みの発展が不可欠である.

-グラフマイニングはネットワークから意味のある情報を抽出することを目指す技術であるが, 現状では必要なデータをどのように選択するかという点で議論したアルゴリズムなどはまだ未発展であり, 近い将来の発展として A C T-X で得られた研究成果をグラフマイニング分野へ展開することを目指したい.

-本研究課題では古典計算の枠組みでの標本複雑度を中心に学習可能性を探求していたが,

量子計算の枠組みにおいても同様な学習問題を定義し，標本複雑度に対応するサンプル効率性の観点で古典コンピュータに対して量子優位性があるのかを議論するような研究へ発展させることも一つの方向である．

4．　自己評価

-研究目的の達成状況: 当初の計画通り1．全バンディット組合せ最適腕識別及び部分観測に基づく組合せ最適腕識別問題について成果が得られ，さらにその非線形報酬関数を扱う枠組みへと発展させることに成功した．既に今後の他の統計的学習タスクへの拡張も視野に入れて研究が進行しており，研究提案時に構想していた研究課題以外にも研究を展開することができた．

-研究実施体制及び研究費執行状況: フランスにある Inria 研究所における強化学習の研究グループと Real-Life Bandits, Inria-Japan Associate Team を編成しオンラインで研究交流をすることができた．（ 実際の研究滞在に ACT-X の研究費を執行予定であったが，コロナ禍や本務の関係でやむを得ず実行することができなかった．） 本研究課題ではＭＳＲＡや清華大学の研究者と共同研究を行うことができた．実際に中国北京への研究滞在することは残念ながら叶わなかったが，代わりに日本での研究環境の整備に使用することができ，オンライン環境でも研究を無事に遂行することができた．

-研究成果の科学技術及び社会・経済への波及効果: 本研究成果は不確実性を伴う意思決定問題の発展及び組合せ最適化問題の標本複雑度から見た新しい理論発展の貢献に寄与すると考えている．

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:5件（うち投稿中3件）

1. Yihan Du*, Yuko Kuroki*(*共同第一著者), & Wei Chen, Combinatorial Pure Exploration with Partial or Full-Bandit Linear Feedback, In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI2021), 35(8), 7262-7270, 2021.
本研究では組合せ最適腕識別問題において，選んだ行動の複数要素からの(ノイズありの)報酬和しか観測できない設定を扱う．楕円型信頼区間最大化問題に対して直接的に近似アルゴリズムを用いた既存手法[Kuroki et a., 2020]では最適行動の報酬と次に良い行動の報酬の差に依存する標本複雑度を持っていたが，本研究ではこの報酬差への依存度を軽減する 2 段階の動的アルゴリズムを提案した．さらに部分観測及び一部の非線形報酬を扱う静的なアルゴリズムの設計を提案し，理論解析および計算機実験により提案手法の妥当性を示した．

2. Yihan Du, Yuko Kuroki, Wei Chen, Combinatorial Pure Exploration with Bottleneck Reward Function, In Proceedings of Advances in Neural Information Processing Systems (NeurIPS2021), 34: 23956--23967, 2021
本研究ではボトルネック報酬関数と呼ばれる非線形関数に着目した. ボトルネック関数とは, 組合せ的な行動について, その要素のうちでもっとも報酬が低いもので全体の報酬が定義される. この関数のクラスに対して, 報酬が未知な状態から, アルゴリズムが必要なサンプル回数の上界を示し, 高確率で最適な行動を出力することを保証した. さらに任意のアルゴリズムは必要とするサンプル数の下界を示し, 提案したアルゴリズムで達成する上界がこの下界 log タームを無視して一致することを示した.
3. Yihan Du, Wei Chen, Yuko Kuroki, Longbo Huang. Collaborative Pure Exploration in Kernel Bandit. In Proceedings of The Twelfth International Conference on Learning Representations (ICLR2023), 2023.
最適腕識別問題の多くの既存研究ではエージェントは一人しかいない場合を想定していたが, 実世界では複数のエージェントが同時に学習を行う場合がある. そのような実設定では複数のエージェントは類似する特徴をもつ学習タスクに取り組むことがある. このような場合において, エージェント間の通信コストを減らしながら最終的に学習に必要なサンプル数を最小化することを目指す枠組みを提案した. 具体的には報酬関数が RKHS で定義される広いクラスの報酬を扱える複数エージェントの逐次的学習モデルを提案し, 効率的なアルゴリズム設計および標本複雑度の解析としてエージェント間のタスク類似度を反映する新しい理論構築を行なった.

（2）特許出願　なし

（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] 第38回井上研究奨励賞, 井上科学振興財団, 2022 年 2 月.

[海外招待講演] Yuko Kuroki, *Combinatorial Pure Exploration with Limited Observation and Beyond,* 2022 Data-driven Optimization Workshop, host by Microsoft Research Asia, online, November 2022.

[国内招待講演]

1. 黒木祐子, 非線形報酬を持つ最適腕識別問題, 情報論的学習理論と機械学習研究会（IBISML）, 琉球大学 50 周年記念館, 2022. 6.27:
2. 黒木祐子, 組合せバンディット問題とその発展, 愛媛大学データサイエンスセミナー, オンライン. 2022. 5.26.

[和文解説記事]

1. 黒木祐子, 5 分で分かる!? 有名論文ナナメ読み Chen, S. et al. : Combinatorial Pure Exploration of Multi-armed Bandits., 情報処理 : 情報処理学会誌 : IPSJ magazine 63(5) 258-260 2022 年 5 月.
2. 黒木祐子,　離散数学に親しむ「人工知能と離散数学 組合せバンディット問題」, 数

理科学 59(12) 2021 年.
[プレプリント] Yuko Kuroki, Junya Honda, Masashi Sugiyama: Combinatorial Pure Exploration with Full-bandit Feedback and Beyond: Solving Combinatorial Optimization under Uncertainty with Limited Observation. In The Fields Institute Communications Series on Data Science and Optimization. [arXiv], 2023.