（ 続紙 1 ）

京都大学	博士（情報学）	氏名	土屋 平
論文題目	Environment Adaptive Regret Analysis in Bandit Problems （バンディット問題における環境適応的リグレット解析）		

（論文内容の要旨）

機械学習においては不確実な知識のもとで逐次的に試行錯誤を伴う意思決定を行い利益を最大化する問題が数多くあり、その代表的なタスクの一つとしてバンディット問題がある。これは学習者がある環境下で複数の行動の選択肢から一つを選択し、選択した行動についてのみ損失を観測できるという設定のもとで累積損失の最小化を目指す問題である。この問題に対しては、従来では最悪時における性能に理論保証を与える方策が多く研究されてきたが、現実のデータは実際には過度に悪い性質をもっておらず、その場合では最悪時のみを考慮する既存の方策が必ずしも良い性能とはならないという課題があった。本学位論文は、現実的な性質をもつデータに対して適応的にその性質を利用することで性能を改善する方策をバンディット問題およびその拡張に対して提案するとともに、その性能解析を行ったものである。

第１章は序論であり、バンディット問題の歴史ならびに本学位論文の主要な貢献の概要が述べられている。第２章では、主要な貢献となる以降の章の準備としてバンディット問題の基礎的な理論とその方策の応用例、および本学位論文の背景となる主要な技術についての説明が述べられている。

第３章では、部分観測問題における実用的な方策について述べられている。部分観測問題はバンディット問題をはじめ非常に多くの設定を含む逐次意思決定問題である。これに対して本章では環境の内部状態が何らかの確率分布によって生成される場合を考え、その場合にトンプソン抽出とよばれるバンディット問題において高性能な方策を部分観測問題に自然に拡張した方策を提案している。さらに、その方策において必要となる計算を高速に行う手法を新たに提案し、既存方策に比べて大幅に良い性能を達成することを実験的に示している。さらに、部分観測問題における不確実性を連続拡張した線形部分観測問題とよばれる設定において、トンプソン抽出の拡張の累積損失が対数オーダーとなることが示されている。また、この設定が線形バンディット問題とよばれる広く用いられている設定を含み、それらにおいてトンプソン抽出が対数期待損失を達成可能であることを示す初の結果であることが併せて説明されている。

第４章では、部分観測問題における両環境最適な方策の構築について述べられている。前章で構築した方策は内部状態が確率的に生成されている場合にのみ性能保証をもつものであった。これに対して本章では内部状態が確率的・敵対的いずれの方法で生成されている場合にも対応可能な方策の構築を行っている。このような性質は両環境最適性とよばれ、この性質をもつ方策の構築は従来ではバンディット問題といった比較的単純な設定に限られていた。これに対して本章では、方策の安定性を最適化計算を通じて改善する既存の枠組みにおいて、最適化の実行可能領域を適切に制限することで部分観測問題において望ましい性質を実現できることを明らかにしている。これを用いることで提案法は確率的設定では対数多項式オーダーの、敵対的設定では最適オーダーである多項式オーダーの累積損失をそれぞれ達成可能であることが示されており、これが部分観測問題における初の両環境最適方策であることが説明されている。

第５章では、汎用的なオンライン学習方策における新たな学習率の決定法とその解析について述べられている。両環境最適な方策のほとんどはFollow-The-Regularized-Leader (FTRL)とよばれる汎用方策の枠組みに属しており、FTRLの学習率とよばれるパラメータを達成したい適応性に対応した統計量に依存して定めることで理論保証を得ていた。一方で、従来の解析手法では学習率を複数の統計量に同時に依存させると理論が破綻する問題があり、そのために達成可能な適応性が限られていた。これに対して、本章では複数の統計量に同時に依存して学習率を定める新たな規準およびそれに対する汎用的な理論保証を導出し、それを適用することで損失系列が疎性をもつバンディット問題への方策、および部分観測問題において両環境最適性をより精密に達成するような方策を新たに構築している。

第６章では、組合せバンディット問題における両環境最適方策について述べられている。組合せバンディット問題は各時刻において複数の選択肢を選ぶようなバンディット問題の拡張である。この問題に対しては両環境最適方策が知られていたが、この「最適」というのはオーダーについての緩い意味であり、その係数部分についてはいずれかの設定に特化した方策より大きく劣るものであった。これに対して本章では観測の分散への適応性を考慮することでより優れた理論保証をもつ両環境最適方策を構築している。さらに、これが片方の環境のみへ特化した方策に比べていずれの環境においても同等に近い性能を達成することを実験的にも確認している。

第７章は学位論文全体の結論となっている。これまでに得られた研究成果のまとめが述べられるとともに、今後の課題ならびに実用化に関する展望について議論されている。

（続紙 ２ ）

（論文審査の結果の要旨）

機械学習における主要なトピックの一つである逐次意思決定問題においては、観測や報酬の系列に関する最悪時において達成可能な性能について古くから解析が行われてきた。一方、現実の問題においては観測や報酬が極端に悪い性質をもつことは稀であり、何らかの良い性質をもつデータに対してはその性質に対して適応的に動作する方策が実応用の観点から求められている。一方、このような特性をもつ既存の方策のほとんどはバンディット問題をはじめとした比較的単純な設定のものに限られている。本学位論文は、部分観測問題をはじめとした複雑な観測モデルにおいてデータの性質に適応的に動作する方策を構築し、その性能を実験と理論の両面から検証している。具体的には、以下に示す研究成果を得ている。

1. 部分観測問題において状態が確率的に生成される設定で対数オーダーのリグレットを達成する方策は、従来では複雑な最適化計算を要する非実用的なもののみが知られていた。これに対して本論文では、バンディット問題において優れた性質が知られているトンプソン抽出とよばれる手法を部分観測問題に適切に拡張した方策を構築し、この方策が最適化計算を行わず対数期待損失を達成可能であることを示すとともに、実験的にも優れた性能となることを明らかにした。さらに、状態が敵対的に生成されている場合にも対応可能な方策を併せて構築し、これが部分観測問題における初の両環境最適な方策であることを示した。

2. 両環境最適となる方策のほとんどはFollow-The-Regularized-Leader (FTRL)を用いた一定の枠組みに沿って構築が行われていたが、この枠組みでは扱うことのできる適応性について限界があった。これに対して本論文ではFTRLの学習率パラメータを複数の統計量に依存して定める枠組みを新たに提案している。このような着想により学習率を素朴に定めた方策は従来の枠組みのもとでは解析を行うことができないが、提案手法では学習率を陰形式で定めることで理論保証が可能となることを新たに示している。さらにこれを応用することで、損失系列が疎性をもつバンディット問題への方策や両環境最適性を従来法より精密に達成するような部分観測問題への方策を新たに構築した。

3. 組合せバンディット問題に関する既存研究では、各選択肢の報酬期待値にのみ注目した方策の構築および解析がなされていたが、現実の問題では報酬の分散が小さい場合が多く、その場合には既存の方策は過度に保守的になる問題があった。これに対して本論文では観測の分散への適応性を有しつつ両環境最適となる方策を構築し、これが実験的にも現実的なさまざまな設定において非常に優れた性能を達成することを示した。

このように、本論文はバンディット問題やその一般化に対して両環境最適性をはじめとするさまざまな適応性を達成する方策を提案し、その有効性を理論的・実験的に示したものであり、関連分野における高度な学術を含むとともに当該研究分野の今後の発展に大きく寄与しうる内容を含んでおり、博士（情報学）の学位論文として価値あるものとして認める。また、令和5年8月23日に論文内容とそれに関連した事項について口頭試問を行った結果、合格と認めた。本論文のインターネットでの全文公表についても支障がないことを確認した。