学位論文審査の要旨

博士の専攻分野の名称　博士 (情報科学)　氏名　渡邊　僚

審査担当者　主　査　准教授　中村 篤祥
副　査　教　授　工藤 峰一
副　査　教　授　有村 博紀

学位論文題名

Analysis of Policies for Budgeted Multi-Armed Bandit Problem and Matching-Selection Bandit Problem
（予算制約付き多腕バンディット問題およびマッチング選択バンディット問題に対する方策の解析）

多腕バンディット問題とは、オンライン学習に属する問題であり、選択とそれに対する報酬 (フィードバック) を得ることを繰り返す場合に、選んだもの以外の情報は得られないという設定において、累積報酬を最大化する選択法 (方策) を求める問題である。選択しないと情報が得られないため、情報を得るために選ぶ「探索」と過去の平均報酬が大きいものを選択する「知識利用」をバランスよく行う必要がある問題として知られている。

多腕バンディット問題は、1930 年頃から患者にどの治療を施すかといった治験の問題として統計学において研究され、その後、強化学習や機械学習などの分野で盛んに研究されてきた。最近では、インターネットの普及により、広告配信システムや推薦システムが広く利用されるようになったが、その配信の最適化に多腕バンディット問題の方策が使われている。

本論文では、報酬が確率的に生成されるという仮定をおく、確率的多腕バンディット問題の 2 つの拡張問題、「予算制約付き多腕バンディット問題」及び「マッチング選択バンディット問題」について、主にリグレット解析と呼ばれる方法で方策を評価し、理論的に分析を行っている。

ある方策のリグレットとは、神のみぞ知る期待累積報酬最大化方策との累積期待報酬の差として定義され、方策の評価に一般的に用いられる指標である。漸近的な評価において、どのような方策を用いてもこれ以上リグレットを下げることができないという限界が存在し、その限界を達成する方策は漸近最適方策と呼ばれる。

本論文では, 第 1 章で概要と貢献について述べ, 第 2 章で古典的多腕バンディット問題の定義、評価法、基本方策について説明している。第 3、4 章で扱う問題は古典的多腕バンディット問題の拡張であり、そこで用いられている方策も第 2 章で説明する方策の拡張になっている。第 3 章では予算制約付き多腕バンディット問題を扱っている。この問題は、選んだものに依存してコストが (確率的に) 変わり、予算を使い切るまで続けるという設定であり、最近のオークション広告などはこの問題として扱うことができる。この章では、古典的多腕バンディット問題の方策である KL-UCB と UCB の拡張である KL-UCB-SC と UCB-SC(UCB for Stochastic Costs) を提案し、それらの方策のリグレット解析を行っている。2 つの方策のリグレット上界を求めるにあたり難しいところは、終了までの選択回数が確率的に変化するということである。本論文では、予算を使い切るまでの可変選択回数によるリグレットの期待値を、固定選択回数のリグレットから変換する式を求め、その式を用いて 2 ステップで証明している。KL-UCB-SC のリグレット上界と (予算 B の関数として)

漸近的に一致するリグレット下界を証明することにより、KL-UCB-SC の漸近最適性を証明している。また、KL-UCB-SC と違い、選択指標が効率的に計算可能な閉形式でかける UCB-SC に対しては、従来法のリグレットより小さいリグレットを証明している。シミュレーションではこれらの理論結果をサポートする結果が示されている。第 4 章ではマッチング選択バンディット問題を扱っている。マッチング選択バンディット問題は、各ラウンドにおいて完全 2 部グラフにおける最大マッチング (共通の端点を持たない最大辺集合) の 1 つを選択し、選択した辺集合に対応する報酬を得る設定の問題である。報酬は辺毎に発生し、選択した辺集合の辺各々から報酬が得られるという半バンディット (semi-bandit) 設定を扱っている。本論文では、$M \leq N$ を満たす自然数 M, N に対し、完全 2 部グラフ $K_{M,N}$ 上のマッチング選択バンディット問題において、新たな証明技法を用いることにより既存方策である LLR 方策のリグレットを $\Theta(M^{2/3})$ 倍改善している。第 5 章では本論文のまとめと今後の課題について述べている。

これを要するに、著者は、情報科学、特に機械学習分野で盛んに研究されている多腕バンディット問題の 2 つの重要な拡張において、提案方策または既存方策に対する理論的評価を行い、それらの性能が拡張問題に対して知られていた最良性能を超えること示しており、多腕バンディット問題における理論の発展に大きく貢献している。よって著者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める。