

## 審査の結果の要旨

氏 名 李 鍾瑛

本論文の主要テーマは多腕バンディット問題(Multi-Armed Bandit problem)である。この問題は、機械学習分野における強化学習の一種であり、意思決定を行うエージェントが未知の環境において、どの行動を選択するかを学習する。具体的な設定としては、選択肢の集合から一つの要素を選択（アームを引く）し、その選択肢に対する報酬を得るが他の選択肢の報酬情報は得られない、というプロセスを繰り返す設定において、報酬和の最大値を目指す逐次決定問題となる。本論文では、各アームからの報酬が何らかの確率分布に従って生成される確率的バンディット(stochastic bandit)問題を扱っている。そして、報酬の確率分布のパラメータをベイズ統計的な枠組みでモデル化する「Thompson Sampling」に関するアプローチの理論解析が主結果である。

背景として、確率的バンディット問題に対しての主要アプローチであるUpper Confidence Boundアルゴリズムは、信頼区間が最大のアームを各試行で選ぶ。しかし、実際には信頼レベルをユーザが適切に決める必要があり、その決定は容易でない。一方でThompson Samplingは、各時点においてそれぞれのアームの報酬分布を推定し、期待値が最大のものを選択する。つまり、信頼レベルをユーザが決定する必要はなく自動的に決定されるという性質を持つ。

Thompson Samplingでは、データの生成元の分布に対して事前分布を設定し、設定した事前分布のもとで実際に観測したデータが生成される確率(尤度)を計算して、事後分布を得る。そして事後分布に基づいてアームを選択する。バンディット問題では、達成し得た最大の報酬から、現在の戦略で得られる報酬の差、すなわち「リグレット」で戦略の良さが評価されることが多い。

本論文は全6章からなる。第1章では、上記の背景を述べ、第2章では本論文に必要な数学定義が含まれている。また第2章では、事前分布として、もっとも単純な無情報事前(non-informative priors)分布を採択することを述べている。この事前分布は通常、一様分布に基づいており、すべての値が等しく信頼できることを表現する時に使用される。無情報事前分布では、以下のような分布が研究分野でよく利用されている。Uniform prior、Jeffreys prior、Reference prior。これらの事前分布は、ガウシアンモデルであれば、確率的多腕バンディット問題に対するリグレット解析に関して、Thompson samplingを利用したアプローチの事前分布として、ほぼOptimalな値を与えることが知られている。しかしながら、上記の解析は、ガウシアンモデル特有の性質を使っており、密度の台が未知母数に依存するような非正則な確率分布に対しては適用でき

ない。本論文の第3章では、このような背景のもと、非正則な確率分布に対しても、Uniform prior、Jeffreys prior、Reference priorは、確率的多腕バンディット問題に対するリグレット解析に関して、上記と同じようなことが成立することを示した。また数値実験を行い、その理論的結果をサポートしている。

本論文第4章は、Thomson samplingに対して、heavy-tailedなモデルを扱っている。背景として、今までの研究成果は、light-tailedなモデルでの解析が多かった。このような状況の下、第4章では、heavy-tailedなモデルとして知られているParetoモデルに関する解析を与えている。このモデルは、Webの解析で使われるなど、データマイニング分野では頻出するモデルである。第4章では、このモデルに対して、Uniform prior、Jeffreys prior、Reference priorに関するリグレット解析を行い、Reference priorの有効性を示している。

本論文の第4章までは、確率的多腕バンディット問題に対する報酬のリグレットについて、Thompson Samplingに基づくアプローチを解析した。本論文の第5章では、第3、4章で与えたThompson Samplingに基づくアプローチが、確率的バンディット問題における「ベストなアーム選択問題」に対しても、既存の結果と変わらないような性能を出すことを、理論的にも実験的にも示している。

第6章では、以上の成果をまとめ、将来課題について議論を行っている。以上のように、李鍾瑛氏の博士論文は、確率的多腕バンディット問題におけるThompson Samplingを利用したアプローチによる事前分布と報酬のリグレットの解析、および、ベストなアーム選択問題の解析に対して、非常に顕著な結果を与えている。よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。