Optimization of the Multi-armed Bandit Problem with Graphical Models: a Bayesian Perspective

氏　名	趙 辰
学位の種類	博　士（　工学　）
学位記番号	博　甲　第　９００９　号
学位授与年月日	平成　３１年　３月　２５日
学位授与の要件	学位規則第４条第１項該当
審査研究科	システム情報工学研究科
学位論文題目	Optimization of the Multi-armed Bandit Problem with Graphical Models: a Bayesian Perspective (確率的グラフィカルモデルによる多腕バンディット問題のベイズ最適化)

主　査	筑波大学 教授	博士(工学)	宇津呂 武仁
副　査	筑波大学 教授	博士(工学)	中内　靖
副　査	筑波大学 教授	博士(工学)	矢野 博明
副　査	筑波大学 教授	博士(工学)	山本 幹雄
副　査	筑波大学 准教授	博士(情報科学) 博士(デザイン学)	星野 准一

論　文　の　要　旨

本論文では、多腕バンディット(multi-armed bandit, MAB)問題を題材としている。多腕バンディット問題は、臨床試験、適応型ルーティング、オンライン広告などを含む多様な分野に応用できるリソース割り当ての最適化問題に一般化することが可能であるため、長年に渡り、多くの研究者の取り組みのもとで、学習アルゴリズムの改善が試みられてきている。

本論文では、まず、背景および理論の解説として、多腕バンディット問題の基礎となる古典的アプローチ、多腕バンディット問題、ガウス回帰手法、確率的グラフィカルモデルについて述べている。

その上で、本論文では、確率的環境設定の下での多腕バンディット問題に対する新しい解決法を提案している。具体的には、本論文では、本論文で提案する新しいグラフィカルモデルを、ベイジアン機械学習の観点のもとで設計している。そして、本論文では、Thompson sampling に基づく新しいサンプリング法を提案している。評価実験においては、グラフィカルモデル推論および適切なハイパーパラメータ調整のもとでは、提案手法により、従来の事後分布からのサンプリング方法と比較してより優れた予測能力を示すことを明らかにしている。特に、提案手法は、従来手法と比較して、訓練事例数に依存しない計算量を実現できており、評価実験においてこのことを定量的に実証した結果を報告している。

以上の議論をふまえて、本論文の主たる貢献は、多腕バンディット問題として解くことが望ましいと考えられる全てのシナリオにおいて適用可能な一般的学習技術を確立した点にある。

審　査　の　要　旨

【批評】

本論文で明らかにした事項は以下の通りである。

本論文で取り組んだ課題は、離散データサンプルに基づいて確率データパターンを正確にモデル化するという課題である。特に、本論文は、多腕バンディット(multi-armed bandit, MAB)問題を題材としている。この多腕バンディット問題では、アーム（腕）と呼ばれるランダム変数の集合によって問題が表現される。この課題における本質的な問題は、損失（regret）を最小限に抑える、すなわち報酬を最大化する選択肢を探索するための最適戦略を設計しながら、期待値を最大化する腕をできるだけすばやく見つけることである。多腕バンディット問題の損失とは、理想的な選択肢の下で取得できる報酬の期待値と、実際のサンプルのもとでの報酬との間の差分によって求められる。簡単そうに見える問題であるが、臨床試験、適応型ルーティング、オンライン広告などを含む多様な分野に応用できるリソース割り当ての最適化問題に一般化することが可能であるため、長年に渡り、多くの研究者の取り組みのもとで、学習アルゴリズムの改善が試みられてきている。多腕バンディット問題の設定は本質的に単一状態マルコフ決定過程であり、より複雑な強化学習問題も、その単一状態マルコフ決定過程である多腕バンディット問題に分割可能であると言える。

以上の背景のもとで、本論文は、多腕バンディット問題を解決するために一般的に使用されている古典的なアプローチの概要から述べ始め、その後、最適な損失の漸近限界を達成した既存の研究におけるガウス回帰手法の詳細を述べている。その上で、本論文では、確率的環境設定の下での多腕バンディット問題に対する新しい解決法を提案している。具体的には、与えられた腕から収集されたサンプルを訓練データとして使用し、選択可能な腕の集合は、モデルが予測する範囲に対応する行動の空間として扱っている。以上の設定をふまえた上で、本論文では、本論文で提案する新しいグラフィカルモデルを、ベイジアン機械学習の観点のもとで設計している。そして、本論文では、Thompson sampling に基づく新しいサンプリング法を提案している。評価実験においては、グラフィカルモデル推論および適切なハイパーパラメータ調整のもとでは、提案手法により、従来の事後分布からのサンプリング方法と比較してより優れた予測能力を示すことを明らかにしている。本論文の主たる貢献は、多腕バンディット問題として解くことが望ましいと考えられる全てのシナリオにおいて適用可能な一般的学習技術を確立した点にある。

以上の議論に基づき、本論文は博士論文に値するものと認められる。

【最終試験の結果】

平成　31 年 1 月 30 日、システム情報工学研究科において、学位論文審査委員の全員出席のもと、著者に論文について説明を求め、関連事項につき質疑応答を行った。その結果、学位論文審査委員全員によって、合格と判定された。

【結論】

上記の学位論文審査ならびに最終試験の結果に基づき、著者は博士（工学）の学位を受けるに十分な資格を有するものと認める。