

令和 5 年 3 月 1 日現在

機関番号：23903

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K01735

研究課題名(和文) リスク鋭敏型のマルコフ決定過程に関する研究

研究課題名(英文) A Study of Risk-Sensitive Markov Decision Processes

研究代表者

影山 正幸 (MASAYUKI, Kageyama)

名古屋市立大学・大学院芸術工学研究科・准教授

研究者番号：50516903

交付決定額(研究期間全体)：(直接経費) 3,100,000円

研究成果の概要(和文)：マーケットで売買する人間の意志決定の不確実性も考慮したモデルを考えるために、B.Liu(清華大学)により提唱されているUncertainty Theoryをモデルの中に組み込むことを考え、実証分析では、種々の経済指標(ビッグデータ)を利用したMCMCによるリスク値(Value at Risk, Conditional Value at Risk)の近似値の推定を行い理論値との比較を行った。その後、ベイズ推定の枠組みの中で強化学習によるパラメータの推定を研究していく予定であったが体調不良に伴う長期入院のため研究継続が困難になり、研究成果をまとめるまでには至らなかった。

研究成果の学術的意義や社会的意義

確率最適化問題において、一般的には与えられた制約条件の下で最適な政策を見つけることが主眼になることが多い。しかし、現実問題を対象とした場合、事前に正確な確率分布を同定することは不可能に近い。よって、古典的な統計学的推論では特定の確率分布の族を仮定したパラメトリックな手法が主流であった。しかし、1990年代から、応用上適応範囲が広いベイズ推定が普及し、現在は人工知能分野の研究に広く活用されていることから、旧来の計量経済学の解析手法を踏まえつつ、確率分布のパラメータがある確率分布に従う(事前分布の想定)というベイズ流の手法で経済データ(株価、為替)の推定モデルの構築が期待できる。

研究成果の概要(英文)：In order to consider a model that also takes into account the uncertainty of human decision making in the market, we considered incorporating the Uncertainty Theory proposed by B. Liu (Tsinghua University) into the model, and in the empirical analysis, we estimated approximate MCMC risk values (Value at Risk and Conditional Value at Risk) using various economic indicators (big data) and compared them with theoretical values. In the empirical analysis, we estimated approximate values of risk (Value at Risk, Conditional Value at Risk) by MCMC using various economic indicators (big data) and compared them with theoretical values. After that, we planned to study parameter estimation by reinforcement learning within the framework of Bayesian estimation, but due to a long-term hospitalization for health problems, it became difficult to continue the research, and we were unable to summarize the research results.

研究分野：応用数学、統計数学

キーワード：マルコフ決定過程 リスク測度 Uncertainty Theory

1. 研究開始当初の背景

MDPs は実社会の重要な問題の解を与えるツールであり、多くのビジネスの現場で利用されてきた。MDPs は、関数方程式、確率論の観点からも多くの研究者の興味を引きつけてきた。しかし、MDPs 理論の研究は、ほとんどは、利得の期待値の最適方程式とその解を見つけるための Policy iteration と Value iteration に関するものである。また、従来、最適化理論に於いては、分散をなるべく小さくして、効用関数の期待を最大化することを目標としていたが、近年、分散の代わりに Value at Risk や Conditional Value at Risk を制約式とする研究が数多く行われている。本研究では、Conditional Value at Risk を用いて、MDPs のリスク評価を考える。具体的には、Conditional Value at Risk に関するベルマン方程式を導き、その最適解を見つけるためのアルゴリズムを提案することを目指す。

2. 研究の目的

近年非常に活発かつ啓発的に研究される計算機科学、人工知能論などと関連される学習理論は、その出発の原点には統計的多段決定過程、意思決定理論、マルコフ決定過程の流れの中にある。これら最適化理論にはこれに伴う関数方程式理論のいわゆる「次元の呪い」を克服するために、これまで多くの研究がなされてきた。その主要な解決方法は、最適方程式とその解を見出す手法の政策最適化(Policy iteration)、最適値改良法(Value iteration)に関するものである。ここでは確率制御理論のなかで、分散リスクを含む多目的最適化論での評価リスク汎関数としてリスク考慮型のモデルによるリスク測度を導入する。具体的には Conditional Value at Risk とリスク鋭敏的価値尺度による最適方程式を導き、その解析を目標とする。

統計科学、数理工学、経営科学の研究分野における確率的制御システム等のテーマでの数理モデルが、マルコフ決定過程(Markov Decision Processes 以後、MDPs) により定式化されることがある。MDPs は、離散時間の確率過程の最適化問題を考える数理モデルであり、1950 年代に研究が始まり、様々な理論や応用研究が生み出されてきた。特に、オペレーションズ・リサーチの分野に於いて、多くの問題を解決するツールになっている。1980 年代までの MDPs の研究は主として、ベルマン方程式とその解を見つけるための、Policy iteration、Value iteration に関する Dynamic Programming(以後、DP) とよばれるアルゴリズムの研究であった(E. A. Feinberg, A. Shwartz, Handbook of Markov decision processes: Methods and Applications, Kluwer Academic Publishers, (2002))。これらの研究の目的は状態空間と行動空間が有限な場合に、総合期待利得の期待値を最大にする Policy をみつけることである。しかし、複数の目的関数を最大化するような問題や、ある制約の下で目的関数を最適化するような問題では、直接 DP を適応させることが難しいことが、1980 年代頃の研究で徐々に明らかになっていった。この問題を克服するために、1990 年代から現在に至るまで、強化学習の理論を用いて MDPs を解析する研究が、理論、応用、両方の側面から行われている。特に、Bertsekas(マサチューセッツ工科大学) と Tsitsklis(マサチューセッツ工科大学) により提案されたニューロ動的計画法(Dimitri P. Bertsekas, Jhon N. Tsitsklis, Neuro-Dynamic Programming, Athena Scientific, (1996)) は、推移確率が未知の MDPs を解析する際の強力な道具となった。一般には、推移確率を完全に把握することは不可能なので、この手法は非常に有益であり、数理モデルとして MDPs を用いる分野で広く利用されている。昨今、脚光を浴びている人工知能理論の中核を成す Deep Learning と MDPs は密接に関わっていることも特筆すべき点である。しかし、囲碁や将棋などの応用面だけに注目が注がれているきらいがあり、その根底に根ざしている数学的性質には関心があまり寄せられていないことは、長期的なスパンで考えた時には、汎用性という観点で欠けていると言わざるえない。科学技術の進歩は、理論と応用が並列して初めて成立するものであり、本研究では、アプリケーションを常に意識しながらも普遍的な数理モデルの構築を目指していく。

3. 研究の方法

2000 年代に入り、MDPs の研究は、それまでの利得の期待値を最大化する研究から、リスクを最小化することに重点が置かれるようになった。特に、リスク考慮型強化学習の研究は、多くの研究者の関心を引きつけている。MDPs におけるリスク考慮型強化学習は大きく分けて、次の 3 つのタイプに分類される(牧野貴樹、他、これからの強化学習、森北出版、2016)。

1. 利得が最も少なくなる Policy を回避するための学習方法
2. ある効用関数を用いてリスクを評価し、そのリスクを最小にすることを目指す手法
3. 期待値以外の汎関数を導入してリスクを定式化する方法

本研究では、3 について考える。Rockafellar(ワシントン大学名誉教授)ら(R. T. Rockafellar, "Conditional Value-at-Risk", Journal of Risk, (2000), Vol.2, No.3, 21-41) により提案された、Conditional Value at Risk(以後、CV @R(X)) を MDPs におけるリスク指標として用い

る。

4. 研究成果

マーケットで売買する人間の意志決定の不確実性も考慮したモデルを考えるために、B.Liu(清華大学)により提唱されている Uncertainty Theory をモデルの中に組み込むことを考え、実証分析では、種々の経済指標(ビッグデータ)を利用した MCMC によるリスク値(Value at Risk、Conditional Value at Risk)の近似値の推定を行い理論値との比較を行った。

その後、ベイズ推定の枠組みの中で強化学習によるパラメータの推定を研究していく予定であったが体調不良に伴う長期入院のため研究継続が困難になり、研究成果をまとめるまでには至らなかった。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 0件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 王琦, 影山正幸, 張景耀	4. 巻 2126
2. 論文標題 マルコフ決定過程による船の避難パターンを推定するモデル	5. 発行年 2019年
3. 雑誌名 京都大学数理解析研究所講究録	6. 最初と最後の頁 63,66
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 影山正幸	4. 巻 2114
2. 論文標題 マルコフ決定過程におけるリスク解析	5. 発行年 2019年
3. 雑誌名 京都大学数理解析研究所講究録	6. 最初と最後の頁 171,173
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計5件（うち招待講演 1件/うち国際学会 1件）

1. 発表者名 影山正幸
2. 発表標題 ベイジアンマルコフ決定過程
3. 学会等名 日本数学会秋季総合分科会
4. 発表年 2019年

1. 発表者名 影山正幸
2. 発表標題 ノイズを含んだマルコフ決定過程について
3. 学会等名 ファジイシステムシンポジウム
4. 発表年 2019年

1. 発表者名 影山正幸
2. 発表標題 Bayesian decision processes with a fuzzy prior distribution
3. 学会等名 Workshop (招待講演)
4. 発表年 2019年

1. 発表者名 Qi Wang, 影山正幸, Jingo Zhang
2. 発表標題 A new criteria in Markov decision processes
3. 学会等名 Stochastic Analysis, Random Fields and Integrable Probability (国際学会)
4. 発表年 2019年

1. 発表者名 影山正幸, 布和額尔敦
2. 発表標題 A construction of Bayesian Markov decision processes
3. 学会等名 動的決定モデルとその応用
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------