



## 《第58回》モデルベース制御と強化学習を融合した次世代制御技術の開発

大川 佳寛

### 1. はじめに

「社会に貢献する研究開発がしたい」、ぼんやりとそう考えていた私は、大学入学後、「制御工学・制御理論」という学術分野に出会いました。そして、自動車やロボットから大規模な社会システムまで、ありとあらゆる対象を思いのままに動かせる「制御」の魅力に取りつかれ、「制御のことをもっと深く知り、使いこなしたい」、その思いが私の研究者としての人生をスタートさせました。本稿では、そんな私が在学時代から企業に入社してこれまで行ってきた、モデルベース制御および強化学習とその両者を融合した新たな制御技術の研究開発、そして開発した技術の社会システムへの応用例について紹介いたします。

### 2. 博士課程における研究

私は、慶應義塾大学大学院後期博士課程において、滑川徹先生のご指導の下、スマートグリッドにおける分散的な動的電力価格決定に関する研究に取り組みました。この研究は、電力価格を操作することで、電力の需給バランスを取りながら、社会効用の最大化を図る問題であり、物理と経済を融合させて社会課題を解決する新たな取り組みでした。特に、次世代の効率的なエネルギーマネジメントシステム構築に向け、電力消費量の削減を行った需要家に金銭的なインセンティブを与える「ネガワット取引」に関して、地域間の電力融通や、風力発電や太陽光発電による発電量の不確かさを考慮して蓄電システムの充放電を含む最適インセンティブ価格設計方法<sup>1)</sup>を開発しました。

これらの価格設計方法は、電力需給バランスに関しては電力網の物理モデルを用いたモデルベースな制約条件を用いる一方で、電力需要家の効用関数や電力供給者

の費用関数などのプライベートな情報は秘匿したまま、市場取引量の情報のみで社会効用の最大化を行うデータドリブンな手法です。したがって、現在取り組んでいるモデルベースな手法とデータドリブンな手法の融合の先駆けになっていたと今回気づき、本研究テーマの立ち上げから手厚くご指導いただいた滑川先生には、この場を借りて改めて感謝申し上げます。

### 3. 入社後の研究

私が富士通研究所（現富士通）に入社した2017年は、ちょうどその前年にGoogle DeepMind社が開発したAlphaGoが囲碁の五番勝負でプロ棋士に勝つなど、強化学習が注目を集め始めていました。そのような中、私は人工知能研究所にて、強化学習を用いた制御手法の開発に取り組み始めましたが、すぐにAI・強化学習、すなわちデータドリブンな手法だけでは、実際の制御問題への適用は難しいとの考えに至りました。とりわけ強化学習は、未知の環境においても適切な方策（制御則）を獲得するために、探索要素を含む行動とその結果に従って試行錯誤をする必要があり、学習時の性能や安全性の保証は大きな課題でした。

この課題解決に向け、私は多くの制御問題においては、制御対象に関して、線形近似した簡易モデルなどの部分的に既知な情報を事前に得ている場合があるという事実に着目し、モデルベース制御とデータドリブンな強化学習を融合させた新たな制御技術の開発に取り組みました。ここで図1に、開発した技術の一例として、外乱存在下における強化学習のための安全探索技術の概要を示しております。本技術では、事前知識として制御対象に関する部分的に既知な情報を用いることで、学習時においても所望の確率以上で制約条件を充足することを可能にしました。本技術に関する研究成果は、制御および機械

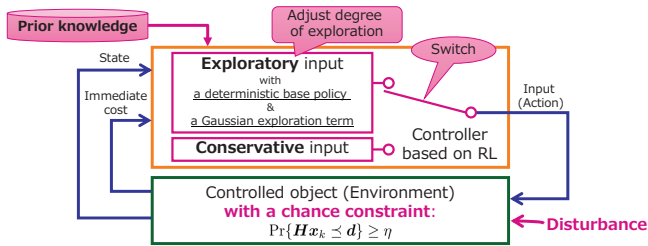


図1 既知の事前知識を用いた強化学習のための安全探索<sup>3)</sup>  
 強化学習制御器の探索要素の有無の切り替えや探索度合いの調整を自動で行うことで、学習中も所望の確率以上で制約条件を充足。

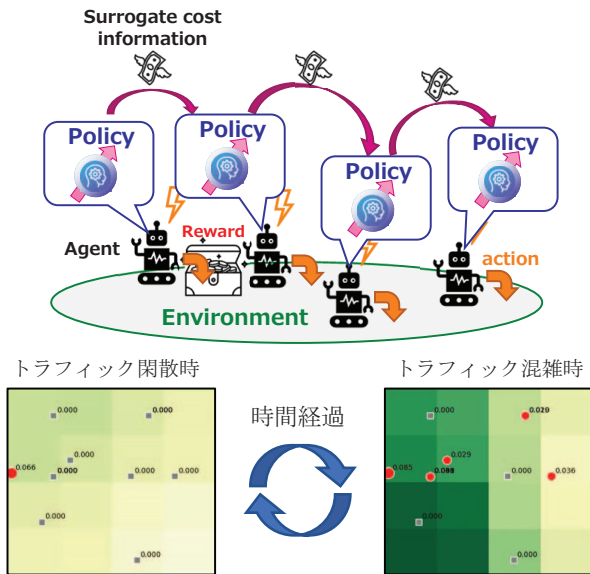


図2 (上) 制約付きマルチエージェント強化学習における分散的な方策更新, (下) 5G 基地局停波制御に対する制約付きマルチエージェント強化学習の適用結果

各地域のトラフィックの混雑状況(グリッドの色の濃淡)に応じて、基地局のRUのActive(●)/Sleep(■)を切り替えて、通信ネットワーク運用における省電力化を実現。

学習の双方の分野のトップカンファレンスで論文採択<sup>2), 3)</sup>されるなど、モデルベースな手法とデータドリブンな手法の融合の価値や有効性が両学術分野からも高く評価されています。

一方、強化学習技術の社会システムへの適用においては、制御対象の数量や範囲の拡大に伴う大規模複雑化に対処する必要もあります。この問題に対し、複数の強化学習エージェントを用いて分散的に学習を行いながらも、それらの行動が関与するようなシステム全体の制約条件を充足する制約付きマルチエージェント強化学習技術の開発<sup>4)</sup>を行いました。この開発技術は、図2(上)に示すように、各エージェントの方策の更新を、代理コスト情報を伝達しながら分散的に順次行います。これにより、システム全体の制約条件を充足しながら複数のエージェントによる自律的な制御を行う方策の学習が可能となり、

強化学習技術の今後の社会システムへの適用可能性を拡大させました。

最後に、現在取り組んでいる強化学習を用いた社会課題解決の例として、通信ネットワーク運用における省電力化について紹介いたします。通信ネットワークを取り巻く社会の状況が日々進化・発展する一方で、その運用における省電力化は、温室効果ガス排出量削減の観点からも大きな社会課題です。そこで上記の制約付きマルチエージェント強化学習技術を、イタリア、ミラノ市のオープンデータを用いて模擬した広域に設置された5G基地局の停波制御に適用し、その効果の検証を行いました。その結果、図2(下)に示すように各地域のトラフィックの混雑状況に応じて適切に基地局停波制御を行うことで、通信品質制約を充足しながら省電力効果があることを確認しました。

#### 4. おわりに

これまでの研究を通じ、社会に貢献する制御の実現には、モデルとデータは共に重要であると深く感じました。したがって、今後もモデルベース制御とデータドリブンな強化学習を融合した次世代制御技術を開発し、社会課題解決に取り組んでいきたいと存じます。

(2023年7月31日受付)

#### 参考文献

- 1) Y. Okawa and T. Namerikawa: Distributed Optimal Power Management via Negawatt Trading in Real-time Electricity Market, *IEEE Trans. on Smart Grid*, 8-6, 3009/3019 (2017)
- 2) Y. Okawa, T. Sasaki, and H. Iwane: Automatic Exploration Process Adjustment for Safe Reinforcement Learning with Joint Chance Constraint Satisfaction, In *Proc. of the 21st IFAC World Congress*, 1588/1595 (2020)
- 3) Y. Okawa, T. Sasaki, H. Yanami, and T. Namerikawa: Safe Exploration Method for Reinforcement Learning Under Existence of Disturbance, In *Machine Learning and Knowledge Discovery in Databases*, 132/147 (2023)
- 4) Y. Okawa, H. Dan, N. Morita, and M. Ogawa: Multi-Agent Reinforcement Learning with Information-sharing Constrained Policy Optimization for Global Cost Environment, In *Proc. of the 22nd IFAC World Congress* (2023)

#### [著者紹介]

おおかわ よしひろ  
 大川 佳寛君 (正会員)

2016年慶應義塾大学大学院理工学研究科総合デザイン工学専攻後期博士課程修了。同年日本学術振興会特別研究員(PD)。2017年株式会社富士通研究所入社。2021年会社統合に伴い富士通株式会社に転籍し、現在に至る。分散制御、分散最適化、強化学習とその社会システムへの応用に関する研究に従事。博士(工学)。

E-mail: okawa.y@fujitsu.com

所属: 富士通株式会社 神奈川県川崎市中原区上小田中 4-1-1