

強化学習を用いた RPS 制度の評価に関する基礎検討

非会員	北海道大学	芦村 陽子	ASHIMURA Yoko
01109481	北海道大学	*北 裕幸	KITA Hiroyuki
非会員	北海道工業大学	西谷 健一	NISHIYA Ken-ichi
01307651	北海道大学	長谷川 淳	HASEGAWA Jun

1. はじめに

電力産業における規制緩和の流れが世界的に広まり、日本においても電力市場の創設が検討されている。また、いち早く自由化がなされた欧米では、環境負荷低減の観点から、様々な CO2 削減政策が提案・運用され、同様に日本においても運用が決定・検討されている。この電力市場の動向と CO2 削減政策は、相互に依存した関係であると考えられ、電力系統および市場参加者の計画・運用や意思決定などにも影響を与えるであろう。著者らは、このような複雑な意思決定主体の動的挙動・相互関係を分析するために、市場参加者・電力市場などをマルチエージェントの概念に基づいてモデル化した「電力システム解析モデル^[4]」を提案している。本稿では、この解析モデルに近年欧米豪で導入された RPS 制度(Renewable Portfolio Standard)を組み込み、市場動向・環境政策が、再生可能エネルギー発電事業者の経済性、環境政策に関わるコストなどにどのように影響を与えるのかを評価するための基礎検討を行う。

2. RPS 制度とそのモデル化

<2.1> RPS 制度 RPS 制度とは、再生可能エネルギーシステムの導入量を目標値として定め、その環境的付加価値を証書化し、需要側には、販売電力シェアに応じた保持義務を課するという制度である。証書(Renewable Energy Credit, 以下 REC)は市場において取引されることで、より経済的な調達が可能となり、また、再生可能エネルギーシステム間での競争も促進されると期待できる。RPS 制度は、再生可能エネルギーシステムの導入量を最初に定めることから、コスト構造が分かりにくい環境税や補助金と比較して、CO2 の削減量が見積もりやすいという利点がある。

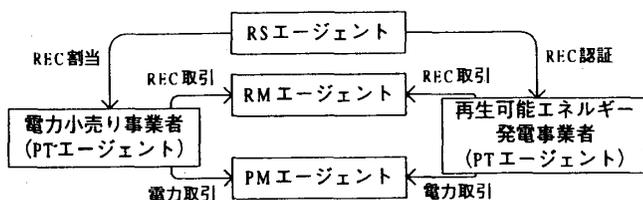


図1 RPS 制度を考慮した電力システム解析モデル

<2.2> RPS 制度のモデル化 本稿では、RPS 制度のモデルとして米国テキサス州の例を取り上げている。モデル化にあたっては、新たに RPS 制度認証機関(RPS System, RS)エージェントおよび RPS 市場(RPS Market, RM) エージェントを追加した。図1にその概要を示す。ここでは、電力系統(PS)エージェントおよび、再生可能エネルギー発電事業者以外の「発電」の要素を持つ市場参加者(PT)エージェントを省略している。

1) RPS 認証機関(RS)エージェント RS エージェントはこの制度の管理・運用を行うエージェントであり、テキサスの例では ERCOT(テキサスエネルギー安定供給協議会: Electric Reliability Council of Texas)の役割を担う。REC の需要側である電力小売り事業者に対して、販売電力量のシェアに応じた暫定的な REC 割当量を 15 分毎に情報伝達する。認証期間終了時には REC の納付受付を行い、不足分に対してはペナルティーを課す。また、REC の供給側である再生可能エネルギー発電事業者に対しては、実績値に応じて REC の認証を行う。

2) RPS 市場(RM)エージェント REC の市場取引を行うエージェントであり、市場は 15 分毎の取引とする。電力市場と同様のプール型の市場を想定し、REC の価格・取引量を決定する。

3. 市場参加者のモデル化

市場参加者の意思決定のアルゴリズムとして、強化学習を適用する。図2にその概要を示す。強化学習はニューラルネットワークのような「教師あり学習」に対して「教師なし学習」と呼ばれ、学習主体が置かれた環境に対する情報が乏しい状態で、自らの試行錯誤を通じて学習する体系である。意思決定に強化学習を適用する理由は以下のようなものである。

1) 電力システムに関わる市場は、その市場の制度設計によって大きく状況が異なり、一般的といえる知見には乏しく、統計的な処理ができない。

2) 市場参加者が入手できる情報は、実際においてもごく限られると予想され、その限られた情報をもとに意思決定し、行動の結果から学習する必要がある。

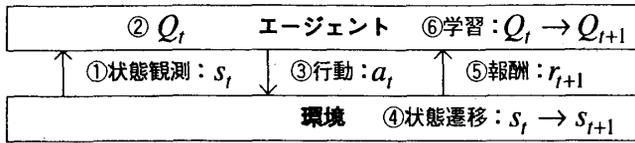


図2 強化学習の概要

<3.1> 強化学習(Q-learning^[2]) 強化学習は、学習意思決定主体であるエージェントが環境の状態を観測し、ある方策に基づいて意思決定をする。このときエージェントは、自らの行動によって環境が状態遷移した結果として得られる報酬を最大化するように行動選択し、これを繰り返すことによって学習する。観測によって得られた状態と、行動の結果得られた報酬をどのように取り扱うことで意思決定するのか、その繰り返しの結果、どのように学習していくのか。そのアルゴリズムが Q-learning である。Q-learning は学習結果の利用と探索との兼ね合いを決めるものである。ここで、報酬 r の時系列として「価値 V 」を以下の式で定義する。

$$V_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (1)$$

価値 V とは、あるタイムステップ t で行動した結果、将来にわたって得られると推定される報酬に不確定の度合いを考慮する重み係数である割引率 ($0 \leq \gamma < 1$) を乗じて合算したものであり、有限の値である。エージェントの目的は評価期間にわたる「価値」の最大化であるが、将来得られる報酬は不確定であるので、価値を期待値で表し、Action-Value 関数 $Q(s, a)$ として定義する。エージェントがタイムステップ t で状態 s_t を観測した時に推定される価値の期待値の関数 $Q(s_t, a)$ を最大化する行動が最適な行動 a^* である。原則として、エージェントは a^* を a_t として選択する。しかしながら、これではローカルミニマムに陥る可能性があるため、 ϵ の確率で行動をランダム選択する (ϵ -greedy 選択)。Q-learning はこの $Q(s, a)$ を以下の式により繰り返し更新することで、学習するアルゴリズムである。

$$Q_{t+1}(s_t, a) \leftarrow \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) \right] + (1 - \alpha) Q_t(s_t, a) \quad (2)$$

右辺第1項はタイムステップ t における行動の結果を反映させる部分であり、どの程度反映させるかは学習率 α ($0 < \alpha \leq 1$) による。行動 a_t に対する報酬 r_{t+1} と、行動 a_t の結果、遷移した状態 s_{t+1} における $Q(s_{t+1}, a)$ の最大値に割引率をかけたものを足し合わせることで、未来の報酬を推定する。第2項は t の時点で持っていた学習結果を残す部分である。このように、第1項の探索結果に重みをつけた部分

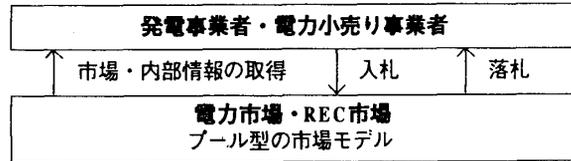


図3 電力システムへの適用

と、第2項の学習結果利用の部分とを足し合わせて $Q(s, a)$ を更新する。

<3.2> 電力システムへの適用 前節の強化学習の枠組を電力システムに適用し、図2に対応する形で示したものが図3である。エージェントは市場参加者であり、発電事業者および電力小売りにあたると。また、環境は電力およびRPS市場にあたる。状態観測にあたる「市場・内部情報の取得」とは、前回のタイムステップでの市場における、電力・RECの均衡価格・流通量および市場参加者のREC保持量である。行動にあたる「入札」は、それぞれの市場に対する希望価格・量である。報酬にあたる「落札」は、エージェントの目的である自己利益最大化を達成するように設定された「信号」であり、以下の式で示す。

発電事業者(Generator)

$$r_G = \text{電力市場からの収入} + \text{REC市場からの収入} \quad (3)$$

電力小売り事業者(Retail)

$$r_R = -(\text{電力市場への支出} + \text{REC市場への支出}) + \text{REC逸脱量} \times \text{REC単価} \quad (4)$$

REC逸脱量とは15分毎に伝達される暫定的なREC割当量と保持量との差分であり、REC単価はそのタイムステップでのRECの均衡価格である。したがって、電力小売り事業者は割当量に追従する形でRECを取得するように行動する。

4. まとめ

本稿では、「電力システム解析モデル」に新たに組み込むRPS制度のモデルを、米国テキサス州の例をもとに提案した。また、電力市場およびRPS市場への参加者の意思決定・学習のモデルとして、Q-learningを用いた強化学習の適用を提案した。

参考文献

- (1) 山口, 他: 「競争環境を考慮した電力システム解析モデルの設計に関する検討」, 電気学会電力技術・電力系統技術合同研究会資料, PE-01-162/PSE-01-156(2001)
- (2) Sutton, R.S. & Barto, A.: 「Reinforcement Learning, An Introduction」, A Bradford Book, The MIT Press (1998)