# 自律分散型スマートグリッド上の電力取引に対する自然方策勾配法によるマルチエージェント強化学習の有効性検証

○谷口忠大, 榊原一紀, 西川郁子 (立命館大学情報理工学部)

# Effectiveness of multiagent reinforcement learning based on natural actor-critic method for electric power trading on an autonomous-decentralized smart grid

∗Tadahiro TANIGUCHI, Kazutoshi SAKAKIBARA, Ikuko NISHIKAWA (Ritsumeikan University)

**Abstract:** In this paper, we describe an adaptive trading agent which can sell and buy electric power effectively in a locally produced and consumed electric energy network. The trading agents manage the amount of electric power generated by solar panels or other renewable energies and stored in a storage battery in several bases. The agents learnsa trading strategies by maximizing future cumulative reward based on reinforcement learning method. Especially, we build autonomous trading agents based on the natural actorcritic method, which is a type of natural policy gradient methods. Several experiments in which we use electric power consumption and generation data measured in actual houses show that the adaptive trading agents can reduce useless energy consumption and a deficiency in most cases.

**Key Words:** スマートグリッド，電力取引，強化学習，人工市場

## 1　はじめに

近年，石油の枯渇や二酸化炭素排出による温暖化の問題を受けて，新たなエネルギー経済への移行が求められている [1]．電力ネットワークに注目すると，旧来の電力供給システムは都市部から離れた場所にある大規模な発電所で集中的に発電し，それを巨大な送電網を用いて配給するという系統電力網によって支えられて来た．この構造は旧来の電力が火力，原子力といったエネルギー密度が高く，また発電システムが大規模であるという一次エネルギーの特性と密接に関係している．このようなシステムでは長距離送電時に発生する電力ロスが問題視されている．その一方で，太陽光，風力を始めとした再生可能エネルギーの利用の促進が図られている．そのために，電気会社による買取り制度が整備されているが，非定常な発電消費パターンに基づく余剰電力が交流系統電力網に逆潮流した際には，既存の系統電力網を不安定化させる懸念がある [2]．

そこで，再生可能エネルギーを普及させるために，マイクログリッドやローカルグリッドと呼ばれる地産地消型の電力ネットワークの研究開発が為されている．地域で発電した電力を地域で上手く消費することにより，出来る限り逆潮流を起こさずに再生可能エネルギー普及を促進させようとするものである．本発表では，その一種である自律分散型直流スマートグリッド i-Rene (Inter Renewable Energy NEtwork) において，Natural Actor-Critic 法に基づく強化学習により自律分散的な電力融通を行う方法について説明し，その後に家庭消費電力及び太陽光発電の実測値に基づいたシミュレーションにより，その有効性を検証する．

## 2　自律分散型直流スマートグリッド i-Rene

筆者らは再生可能エネルギーに基づく電力を地産地消するための枠組みとして自律分散型直流スマートグリッド i-Rene を提案している [3]．図 1 に i-Rene の概念図を示す．i-Rene は松本らにより提唱された ECO ネットに余剰電力融通のために人工知能を用いた人工市場に基づく自由取引による電力売買を導入したものである．i-Rene には複数の発電・消費拠点（住宅，事業所など）が存在しており，それぞれが基本的には発電装置，負荷装置，蓄電池，電力ルータを持つ事を前提としている．電力ルータは蓄電池の残量や電力の発電・消費量を監視しつつ他の拠点との間で電力の送受信を行う情報処理装置である．電力ルータによって各拠点間の電力融通を行う事でロスのない再生可能エネルギーの利用を実現する事を目指す．

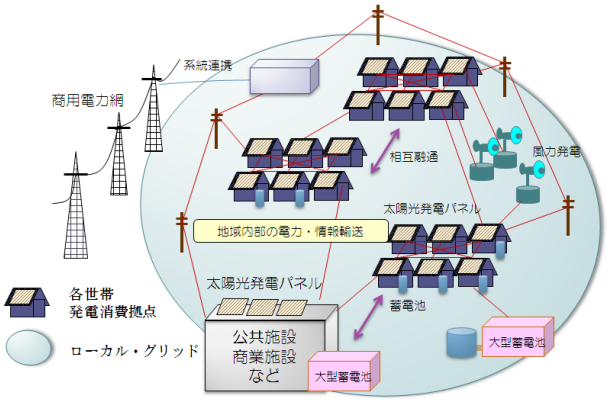


Fig. 1: 自律分散型スマートグリッド i-Rene の概観図

i-Rene では，各発電消費拠点が，自らの利潤を最大化しようとするという行動基準に基づき電力取引を行う地産地消型の電力取引市場を仮定する．市場を介した自由な売買に基づき効率的な電力融通を実現することを目指している．しかしながら，電力事業者ではなく，一般家庭が電力取引に参加することを考えると，常時取り引きを担う事の出来る人員の存在を前提とは出来ない．よって，その状況下で実時間的な電力取引を実現する為に，人手を介さず人工知能を用いる事を目指している．

一方で，自動的な電力売買を行うにしても，各拠点が為すべき取引は一定ではない．各世帯ごとに電力発電・消費・蓄電の諸条件は異なり，また，時間帯によっても変化する．よって，電力取引を自動的に行う知能は，み

ずからの環境に適応する必要がある．筆者らは電力取引を強化学習の枠組みに基づいてモデル化し，Natural Actor-Critic 法を用いてシミュレーション上での実装を行った [3].

## 3 Natural Actor-Critic による適応的取引エージェントの構築

本章では i-Rene 上における適応的取引エージェントの構築を行う．まず，用いる強化学習手法である Natural Actor-Critic の導入を行い，次に電力取引過程を強化学習で扱いうる形に定式化する．最後に Natural Actor-Critic を取引エージェントに適用する．

### 3.1 Natural Actor-Critic

MDP に基づく強化学習では，エージェントが時間 $t$ において環境の状態 $x_t \in \boldsymbol{X}$ を観測した後に方策 $\pi(x,u)$ に従い行動意思決定を行い行動 $u_t \in \boldsymbol{U}$ を環境に出力した後に遷移後の環境の状態 $x_{t+1}$ を観測し，同時に報酬 $r_t$ がエージェントに与えられる [4] という一連の流れを前提にしてエージェントは学習を行う．エージェントは将来に亘る獲得報酬の期待値を最大化するように学習を行う．

Natural Actor-Critic は広い意味では方策勾配法と呼ばれる手法に含まれる．方策勾配法では確率的な方策を陽に表現し最適方策を直接近似する事を目指す．逐次的に方策自体を直接的に改善していけるため，Q-learning などの価値に基づく方法より，POMDP 下でも学習を進める事が出来ると言われている．本研究で扱う系はマルチエージェント環境であり，部分観測性を持つため，方策勾配法の利用が適切であると考えられる．

$$J(\boldsymbol{\theta}) = E\{(1-\gamma)\sum_{t=0}^{\infty}\gamma^t r_t \mid \boldsymbol{\theta}\} \quad (1)$$

$$= \int_{\boldsymbol{X}} d^{\pi_\theta}(x)\int_{\boldsymbol{U}}\pi_{\boldsymbol{\theta}}(x,u)r(x,u)dxdu \quad (2)$$

$$d^{\pi_\theta} = (1-\gamma)\sum_{t=0}^{\infty}\gamma^t p(x_t = x) \quad (3)$$

収益の期待値 $J(\boldsymbol{\theta})$ は将来に亘って得られる累積報酬の割引和になっており，この評価値を最大化するように方策を最適化する事が強化学習の課題となる．ここで，方策 $\pi_{\boldsymbol{\theta}}(x,u) = p(u \mid x;\boldsymbol{\theta})$ はパラメータベクトル $\boldsymbol{\theta}$ によって表される方策であり，$d^\pi$ は割り引かれた状態分布である．

収益の期待値をパラメータ $\theta$ に関して偏微分した $\nabla_{\boldsymbol{\theta}} J$ は収益の期待値における $\theta$ の最急上昇方向を表すとされるが，パラメータ空間が歪んでいるときにはプラトー現象などにより学習性能が悪化する事が知られている．このプラトー問題を回避する有効な手段として提案された手法が自然勾配法である．自然勾配法では通常の勾配にフィッシャー情報行列 $F$ の逆行列を乗ずる事で自然勾配を求めるが，強化学習においてはこの自然勾配が比較的容易に計算しうることが示されている [5]．Natural Actor-Critic とは自然勾配法を方策勾配法に適用した自然方策勾配法の一つである．自然勾配はアドバンテージ関数を $\nabla_{\boldsymbol{\theta}}\log\pi$ を基底関数として推定したときの係数ベクトルに一致することが知られている．アドバンテージ関数とは行動価値と状態価値の差分によって表される関数であり，状態価値の寄与を除いた行動の価値を表現する関数である [6]．具体的な実装としては価値関数の推定方法として LSTD-Q($\lambda$) [7, 8] を用いた無限区間タスクでの Peters らによって提案された Natural Actor-Critic [9] を用いる．Natural Actor Critic では行動価値関数の一部であるアドバンテージ関数を正確に求める必要があるため，統計量から最適な価値関数を直接求める事の出来る LSTD-Q($\lambda$) を用いている[1]．アルゴリズムは以下である．

まず，パラメータ表現された方策 $\pi(x,u) = p(u \mid x,\boldsymbol{\theta})$(初期パラメータ $\boldsymbol{\theta} = \boldsymbol{\theta}_0$)，その導関数 $\nabla_{\boldsymbol{\theta}}\log\pi(x,u)$，状態価値関数 $V^\pi(x)$ のパラメータ化のための基底関数 $\boldsymbol{\phi}(x)$ とおく．

1. 初期状態 $x_0 \sim p(x_0)$, パラメータを選択 $\mathbf{A}_{t+1} = \mathbf{0}, \mathbf{b}_{t+1} = \mathbf{z}_{t+1} = \mathbf{0}$.
2. t=0,1,2,... で以下 (3.~8.) 繰り返し
3. 行動 $u_t \sim \pi(x_t,u_t)$ を決定，次の状態 $x_{t+1} \sim p(x_{t+1} \mid x_t,u_t)$ とそれにともなう報酬 $r_t = r(x_t,u_t)$ を観測.
4. LSTD-Q による状態価値評価 (Critic 器):
   価値関数を
   $$V^\pi(x_t) = \boldsymbol{\phi}(x_t)^\top \boldsymbol{\nu}_t \quad (4)$$
   として，Actor に適合するアドバンテージ関数の近似式を
   $$f^\pi_\omega(x_t,u_t) = \nabla_{\boldsymbol{\theta}}\log\pi(u_t \mid x_t)^\top \boldsymbol{\omega}_t \quad (5)$$
   とする.
5. 基底関数の更新:
   $$\tilde{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}(x_{t+1})^\top, \mathbf{0}^\top]^\top \quad (6)$$
   $$\hat{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}(x_t)^\top, \nabla_{\boldsymbol{\theta}}\log\pi(x_t,u_t)^\top]^\top \quad (7)$$
6. 十分統計量の更新:
   $$\mathbf{z}_{t+1} = \lambda\mathbf{z}_t + \hat{\boldsymbol{\phi}}_t \quad (8)$$
   $$\mathbf{A}_{t+1} = \mathbf{A}_t + \mathbf{z}_{t+1}(\hat{\boldsymbol{\phi}}_t - \gamma\tilde{\boldsymbol{\phi}}_t)^\top \quad (9)$$
   $$\mathbf{b}_{t+1} = \mathbf{b}_t + \mathbf{z}_{t+1}r_t \quad (10)$$
7. Critic のパラメータ更新
   $$[\boldsymbol{\nu}_{t+1}^\top, \boldsymbol{\omega}_{t+1}^\top]^\top = \mathbf{A}_{t+1}^{-1}\mathbf{b}_{t+1} \quad (11)$$
8. Actor のパラメータ更新
   アドバンテージ関数のパラメータベクトルで表される自然勾配 $\omega_t$ が時間枠 $W_h$ で収束したとき (例えば $\forall\tau \in [0,...,W_h]$ に対し, $\cos(\angle(\boldsymbol{\omega}_{t+1},\boldsymbol{\omega}_{t-\tau})) \geq \epsilon$ ) 方策パラメータを更

[1] 自然方策勾配法の定式化において価値関数を逐次的に求める形のものも試されているがそのような実装は学習が不安定化する事が多いと考えられる [10].

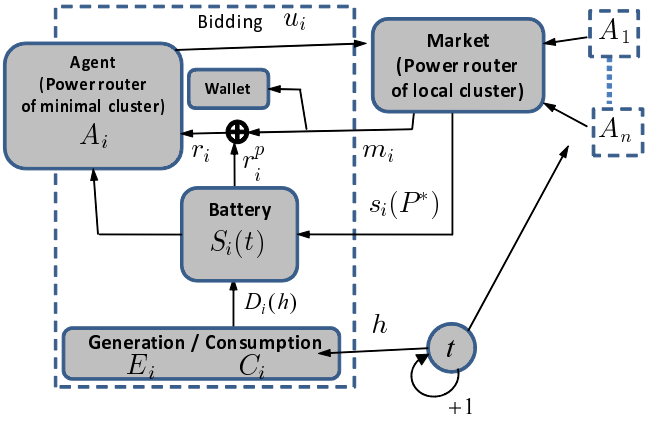


Fig. 2: 本稿のモデルにおける電力取引全体プロセス

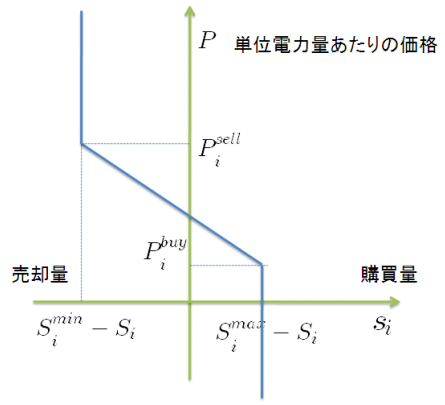


Fig. 3: 入札曲線の例

新した上で，十分統計量を減衰させる．

$$\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha\boldsymbol{\omega}_{t+1} \quad (12)$$
$$\mathbf{z}_{t+1} \leftarrow \beta\mathbf{z}_{t+1} \quad (13)$$
$$\mathbf{A}_{t+1} \leftarrow \beta\mathbf{A}_{t+1} \quad (14)$$
$$\mathbf{b}_{t+1} \leftarrow \beta\mathbf{b}_{t+1} \quad (15)$$

9. 終了

本研究ではこの Natural Actor-Critic に基づく方策勾配法を用いて自動取引エージェントを構築す [3].

### 3.2 電力取引エージェントへの適用

#### 3.2.1 モデルの全体像

図 2 に i-Rene の各拠点での発電・消費からエージェントの取引条件の意思決定，市場取引，そしてその結果からの収益と電力量の流入，に至る情報の流れを簡単に示す．まず $i$ 番目のエージェント $A_i$ の時刻 $t$ での発電量，消費量を $E_i(t), C_i(t)$ とする．また，時刻 $t$ における売買による流入は $A_i$ の購買量 $s_i(t)$ と等しくなる[2]．各ミニマル・クラスターは蓄電池を有しており時刻 $t$ において蓄電残量を $S_i(t)$ とする．また，その最大，最小容量を $S_i^{max}, S_i^{min}$ とする．これらの流入により蓄電残量 $S_i(t)$ は次式に基づいて毎時刻変化する．

$$\bar{S}_i(t) = S_i(t-1) + s_i(t-1) + D_i(t) \quad (16)$$
$$S_i(t+1) = \max((\min(\bar{S}_i(t), S_i^{max}), S_i^{min}) \quad (17)$$

取引は 2 時間に一度行われるものとし 2 時間毎に $t$ は 1 加算される．$h$ は $t$ の周期成分を取り出した 24 時間周期の値であり，具体的には $h$ は $t$ を 12 で割った余りである．本研究では太陽光発電の下で一定の生活パターンで居住者が生活する系を想定し，系はノイズ程度の変化を除き 24 時間周期の周期性を持つものとする．これらに基づき，エージェントは毎時刻，状態変数 $x_i = (S_i, h)$ を観測出来るものとする．この状態変数に基づきエージェントは取引条件の意思決定を行い，行動出力とし $u_i = (P_i^{buy}, P_i^{sell})$ を出力する．

[2]簡単の為，今回は送電ロスや蓄電ロスは考慮しない．流入に係数などをかけることにより拡張は可能である．

$P_i^{buy}, P_i^{sell}$ は市場に提出する需要・供給曲線の内それぞれ最大限買う，最大限売る場合の価格であり，このときの価格 $p$ に対する購入量 $s$ を決定する入札曲線は図 3 の様になる．購入量 $s$ の負の部分は売却量を表す．

$$s_i(p) = \begin{cases} S_i^{min} - S_i(t) & (p > P_i^{sell}) \\ \frac{S_i^{min} - S_i^{max}}{P_i^{sell} - P_i^{buy}}(p - P_i^{sell}) + (S_i^{min} - S_i(t)) & \\ \qquad (P_i^{sell} \geq p \geq P_i^{buy}) & \\ S_i^{max} - S_i(t) & (p < P_i^{buy}) \end{cases} \quad (18)$$

エージェントは毎時刻 $P_i^{buy}, P_i^{sell}$ を市場に出力する事で，図 3 に示すような入札曲線を生成する．入札曲線は価格から数量への関数であるが，関数を上下限を持つ一次関数に制約することで，端点二点の出力で入札曲線の出力に替える事ができる．図 4 に示すように，各エージェントは $P_i^{buy}, P_i^{sell}$ を出力する事で取引条件を市場に出力するが，取り引き数量 $s_i$ について正の部分が買い手として，負の部分が売り手としての取引条件となる，これらを数量の方向に足し合わせる事で市場における需要曲線・供給曲線が生成される．これらの曲線の交点を求める事で決済される際の価格 $P^*$ が求まる．具体的には数式 (19) を満たす均衡価格 $P^*$ を決済価格とする．

$$\sum_{s.t. A_i \in M} s_i(P^*) = 0 \quad (19)$$

時刻 $t$ での市場での決済価格を $P^*(t)$ とすると，エージェント $A_i$ の売買収支は $m_i(t) = -s_i(t) * P^*(t)$ となる．決済後，各エージェントに購買量 $s_i(t)$ と売買収支 $m_i(t)$ が流入し，一サイクルが終了する．報酬 $r_t$ は売買収支 $m_i$ と電力超過・不足ペナルティ $r_i^p$ などに基づいて計算される．

#### 3.2.2 学習器の設定

入札曲線の学習について述べる．1 エージェントについて述べるので，簡単の為，エージェントの添え字 $i$ は省略する．方策関数 $\pi(x, u; \boldsymbol{\theta})$ を方策パラメータ $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma_1, \sigma_2\}$ とガウス関数を用いて下の様に定義する．

$$\pi(x, u; \boldsymbol{\theta}) = \prod_{k=1,2} \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(\frac{-1}{2}\frac{(u_k - f_k(x, \boldsymbol{\theta}_k))^2}{\sigma_k^2}\right) \quad (20)$$

$f_1, f_2$ は行動出力分布の中心を表わし，それに標準偏差 $\sigma_1, \sigma_2$ のガウスノイズを付加し入札価格を決定する．こ

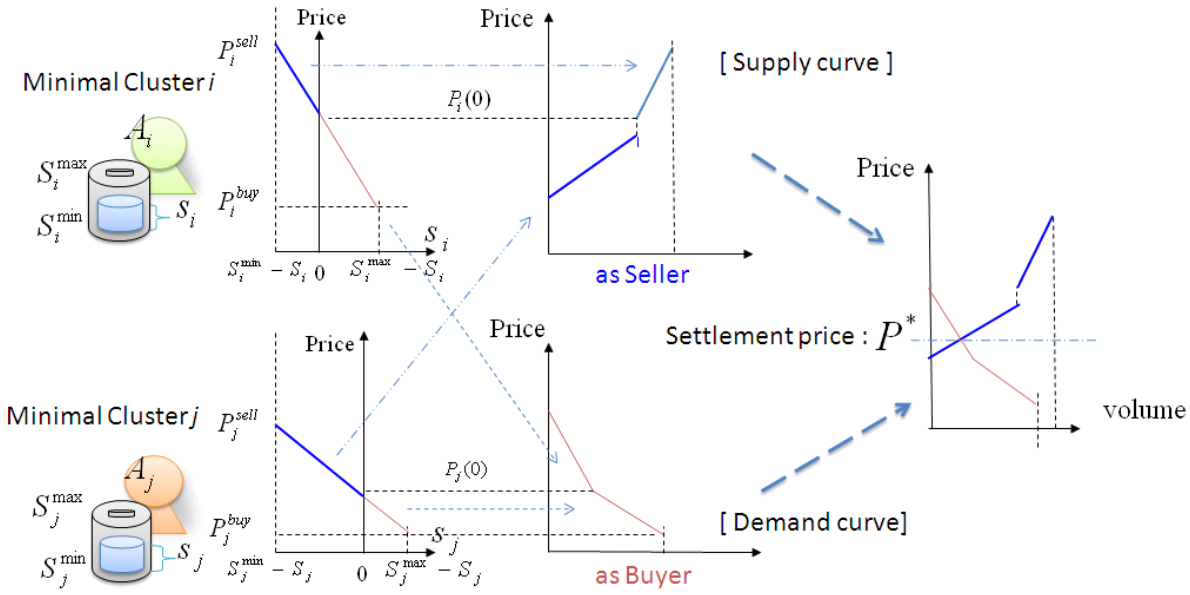


Fig. 4: 市場決済の概念図

こで $u_1 = P^{buy}$, $u_2 = P^{sell}$ とする．ここで，価値関数 $V^{\pi}(x)$ および $f_k$ を共通の基底関数 $\boldsymbol{\phi}$ を用いて下記のように表わす．

$$V^{\pi}(x) = \boldsymbol{\phi}(x)^{\top}\boldsymbol{\nu} \quad (21)$$

$$f_k(x;\boldsymbol{\theta}_k) = \boldsymbol{\phi}(x)^{\top}\boldsymbol{\theta}_k \quad (22)$$

また，アドバンテージ関数の係数ベクトル $\boldsymbol{\omega}$ については，上記方策関数の定義に従い，式 (5) に基づき定義する．ここで基底関数 $\boldsymbol{\phi}(x)$ の第 $j$ 成分 $\phi_k(\boldsymbol{x})$ はクロネッカーのデルタ $\delta_{ij}$ を用いて，

$$\phi_k(\boldsymbol{x}) = \begin{cases} \delta_{kh} & (0 \le k < 12) \\ \rho^{k-12} & (12 \le k \le 15) \end{cases} \quad (23)$$

$$\rho = \frac{S_{max} - S}{S_{max} - S_{min}} \quad (24)$$

と表わされるとする．上段のクロネッカーのデルタで表される部分は周期的に変化する時間に依存する項，下段の項は電池残量への依存性を三次の多項式で表現している．この前提の下で自然勾配計算の為に $\nabla_{\boldsymbol{\theta}} \log \pi(x,u)$ を計算する．$\boldsymbol{\theta}_k$ の第 $j$ 成分を $\theta_{kj}$ として，

$$\frac{\partial}{\partial \theta_{kj}} \log \pi = \frac{\phi_j}{\sigma_k^2}\varepsilon_k \quad (25)$$

$$\frac{\partial}{\partial \sigma_k} \log \pi = \frac{(\varepsilon_k^2 - \sigma_k^2)}{\sigma_k^3} \quad (26)$$

となる．ただし $\varepsilon_1 = P^{buy} - f_1, \varepsilon_2 = P^{sell} - f_2$ である．

よって，Natural Actor-Critic における各の基底関数の更新については

$$\tilde{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}(x_{t+1})^{\top}, \mathbf{0}^{\top}]^{\top} \quad (27)$$

$$\hat{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}(x_t)^{\top}, \nabla_{\boldsymbol{\theta}} \log \pi(x_t,u_t)^{\top}]^{\top} \quad (28)$$

より，$\boldsymbol{\phi}' = \boldsymbol{\phi}(x_{t+1}), \boldsymbol{\phi} = \boldsymbol{\phi}(x_t)$ として，

$$\tilde{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}'^{\top}, \mathbf{0}^{\top}]^{\top} \quad (29)$$

$$\hat{\boldsymbol{\phi}}_t = [\boldsymbol{\phi}^{\top}, \frac{\varepsilon_1}{\sigma_1^2}\boldsymbol{\phi}^{\top}, \frac{\varepsilon_2}{\sigma_2^2}\boldsymbol{\phi}^{\top}, \frac{\varepsilon_1^2 - \sigma_1^2}{\sigma_1^3}, \frac{\varepsilon_2^2 - \sigma_2^2}{\sigma_2^3}]^{\top} \quad (30)$$

となる．その後，Natural Actor-Critic アルゴリズムにそって状態価値関数，アドバンテージ関数の近似式のパラメータベクトルを更新し，ある方向に収束したと判定されれば方策を更新する．収束判定は時間 $t+1$ と時間 $t-\tau$ の間のそれぞれのパラメータベクトルのなす角の cos 値が十分 1.0 に近い閾値 $\epsilon$ に対し $\cos\theta = \frac{\langle \omega_{t+1}, \omega_{t-\tau} \rangle}{\|\omega_{t+1}\| \|\omega_{t-\tau}\|} > \epsilon$ を満たせば収束したとみなし，方策パラメータを更新する．

## 4 実験

先行研究 [3] における実験では，仮想的な世帯を計算機環境に構築し検証を行っていたが，そこでの実験条件は余剰電力の時間変化を正弦波で近似するという現実の消費電力パターンからはかけ離れた条件であった．本研究では，太陽光発電と家庭消費電力の実測値を用いて検証を行う．

### 4.1 実験条件

本実験では太陽光パネルを家に備えた一戸建て世帯を想定する．家は六軒とし，それらの中で電力がまかなえなかった時のみ，通常の電力価格で電力会社から電力を購入するものとする．本実験では発電量 $E_i$ と消費量 $C_i$ はそれぞれ実測値に基づくデータを用いて実験を行う．太陽光発電データについては滋賀県草津市の立命館大学テクノコンプレクス　ハイテクリサーチセンター屋上に備え付けられた定格発電容量 1,980[$W$] のアモルファスシリコン製のソーラーパネルによる 2007 年 8 月の実測データを用いた．図 6 に晴天時，雨天時の典型的な発電パターンを示す．晴天時の発電量は最大値が入射角を反映したコサイン波の正の部分で表わされ，一時的な影や曇りによる減少割合を含んだものとなる．図 6 の上図が典型的な発電パターンである[3]．家庭電力消費データについては前田らによって計測された福岡市の 19 戸ある集合住宅の中の 6 世帯の 2003 年 8 月の消費電力パターンを用いた [11, 12]．図 5 に典型

[3]以降，本研究では電力量の基本単位を MJ とする．

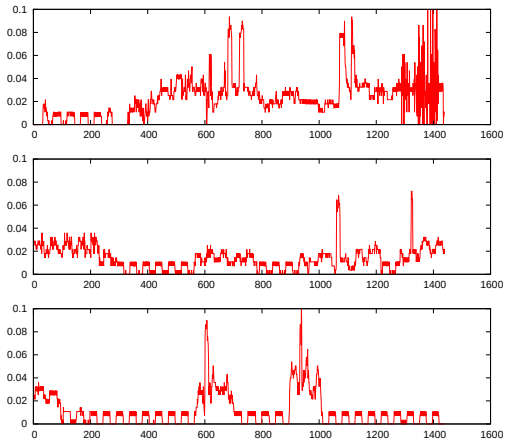

Fig. 5: 電力消費パターンの例：A4 の拠点での消費，上からそれぞれ 4 日目，10 日目，13 日目のパターン

的な消費電力パターンを示す．ソーラーパネルによる発電パターンに比べて，同一世帯であってもその消費パターンには多様性があり，予測は容易ではない．また，太陽光による発電量は説明変数として天候でほぼ足りるが，消費電力については説明変数の存在も不明瞭である．よって，これらの差分としての余剰電力は予測しがたいものとなる．実験はこれらのデータ一ヶ月分を繰り返し外生変数として各エージェントに与えた．太陽光パネルは同一のものを持っているとし，生活パターンは世帯毎に異なるものとしている[4]．

各エージェントの初期方策は残存蓄電量を参照しながら，十分に残量がある際には安く，少なくなった際には高く値付けするよう下式に設定した．

$$P^{sell} = 10.0 - 1.0\rho \tag{31}$$

$$P^{buy} = 4.0 - 1.0\rho \tag{32}$$

単位は [円/MJ] とする．つまり $\theta_{1,12} = 4, \theta_{1,13} = -1, \theta_{2,12} = 10, \theta_{2,13} = -1$ とした．初期方策では時間に対する依存性は自明でないので全て 0 とした．さらに価格の出力の変動幅（ガウス分布に従う）は $\sigma_1 = \sigma_2 = 0.1$ としている．また出力する取引条件について $P_i^{sell} > P_i^{buy} > 0$ を制約条件として与える．実験条件として 1 日の取引回数は 12 回，学習率 $\alpha = 0.001$，適格度トレースの割引率 $\lambda = 0.98$，割引率 $\gamma = 0.99$，十分統計量の保持率 $\beta = 0.0$，方策勾配の収束判定のウィンドウ幅 $W_h = 12 \times 7 = 1$[week] 及び収束判定の閾値 $\varepsilon = 0.9$ とした．また，他のパラメータは $g_i = 20, b_i = 0, S_i^{max} = 100, S_i^{min} = 25$ とした．また，初期蓄電量は $S_i(0) = 50$ とした．

また，報酬であるが，蓄電残量が無くなった場合のみ 6.7[円/MJ] で系統電力を購入するものとし，時間あたりの売買収支を報酬とした．

### 4.2 実験結果

上記の実験条件の下でシミュレーション実験を行った結果を示す．全ての実験結果は同条件で五回行った結果の平均を示す．図 7 にエージェント A1 のみが学習を行った場合の月毎の獲得報酬値の変化を示すが，エージェント $A_1$ が徐々に上手く取引を行う事で獲得報酬値

[4]同一地域であれば，素材は違っても太陽光発電パターンは，定数倍程度でほぼ同様のものとなる．

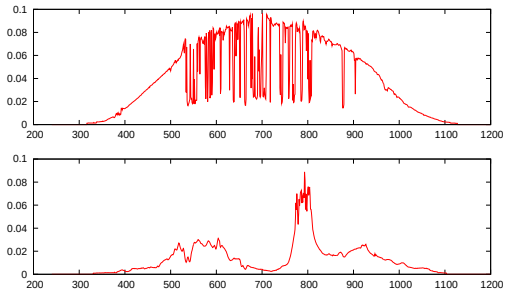

Fig. 6: 太陽光発電パターンの例：上から 10 日目（晴），30 日目（雨天）のパターン

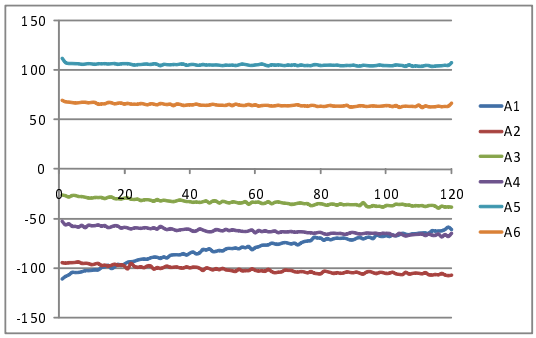


Fig. 7: A1 のみ学習した場合の月毎の各エージェントの報酬変化（単位は横軸は月，縦軸が円）

を増大させている事がわかる．次に，図 8 に全エージェントが学習を行った場合の結果を示す．地域内部における金銭の授受は基本的にゼロサムゲームであるために，報酬獲得を伸ばすエージェントも居れば，減らすエージェントも見られた．そこで，全エージェントの報酬の総和の変化を見た．図 9 に全エージェントの報酬総和の月変化を示す．徐々にではあるが，増加していっていることが分かる．今回の実験条件では，ゼロサムとなる地域内部での取引以外による報酬変化は，内部の電力融通で賄いきれなかった場合の系統電力からの買い入れを表わしている．故に，この買い入れ（負の報酬）が減少している事は，地域内での電力を無駄なく融通できるようになっていることを示している．

### 4.3 考察

先行研究 [3] では，余剰電力が毎日正弦波で近似され，日々の発電消費パターンに変化が無いという想定の下，Natural Actor-Critic 法による電力融通の可能性が示されたが，本実験により発電消費パターンの日変動

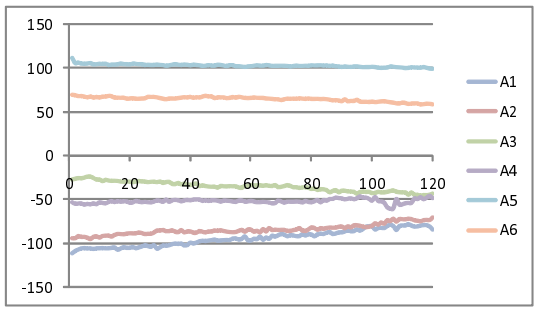


Fig. 8: 全てのエージェントが学習した場合の月毎の各エージェントの報酬変化（単位は横軸は月，縦軸が円）

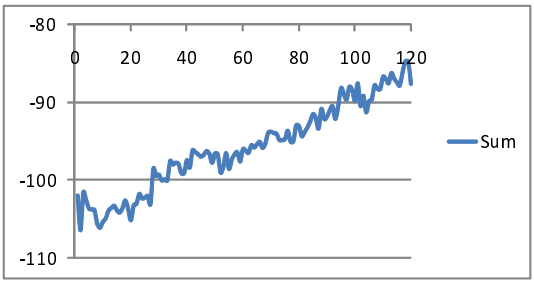


Fig. 9: 全てのエージェントが学習した場合の月毎の全エージェントの報酬総和の変化（単位は横軸は月，縦軸が円）

がある場合でも，可能である事が示された．

しかしながら，その学習スピードは決して速くなく，学習の成果が十分に出るためには年単位の時間が必要となる．学習率 $\alpha$ を増加させることで，学習を加速させる事は出来るが，方策勾配法の性質上，増加させすぎると発散してしまう事例が見られた．この結果は，基底関数の設計とも関連している．また，Natural Actor-Critic 法では設計すべきメタパラメータが多く存在し，この設計指針が決して問題やスケールと独立で無いためにしばしば困難が見られた．このような問題は，生活と密接に関わるエネルギー分野への応用を考える上では重要な問題となる．

また，本研究では天候情報などの付加的情報を方策の状態変数に加える事は行わなかった．先にも述べたように，発電データは天候予測により予測性が増すとは言え，生活行動パターンに従属する消費データを予測することは困難である．今後のアプローチとしては，そのような付加的な情報を増し状態変数の次元を増していくアプローチと，逆にモデルを簡素化し安定性やロバスト性を増していくアプローチの両方が存在すると考えられる．

## 5　まとめ

本研究では，Natural Actor-Critic 法に基づいて地域電力市場で電力売買を行うためのエージェントの構築手法について概説し，その有効性を実発電・消費データに基づいて検証した．気象予測などを状態変数に新たに追加することなくとも，取引条件を学習し無駄の少ない電力融通を実現する事ができるようになった．しかしながら，本研究では，実運用に耐えうる動作の安定性を保証するという事は出来ず，その点を考慮した適応的電力取引エージェントの構築が今後の課題である．

### 謝辞

本研究は 科学研究費補助金 基盤研究 (B)「不便の効用を活用したシステム論の展開」21360191 科研費 学術創成研究「記号過程を内包した動的適応システムの設計論」19GS0208　及び，平成 21 年度　経済産業省　低炭素社会に向けた技術発掘・社会システム実証モデル事業「自律分散型直流スマートグリッドの基本機能実証と地産地消電力取引の社会実験」の一部支援を受けた．また，太陽光発電データは立命館大学　高倉研究室に提供頂いた．また，家庭消費電力データについては産業技術総合研究所の前田哲彦先生より共同研究に基づき使用させて頂いた．ここに感謝の意を示す．

## 参考文献

[1] ヘルマンシェーア．ソーラー地球経済．岩波書店, 2001.

[2] 新エネルギー・産業技術総合開発機構．「自律分散型電力システムネットワークの可能性調査」委託業務成果報告書　．平成 15 年度調査報告書 NEDO-P-0304, 2004.

[3] 谷口忠大, 高木圭太, 榊原一紀, 西川郁子. 地産地消型電力ネットワークの為の natural actor-critic を用いた自動取引エージェントの構築. 知能と情報（日本知能情報ファジィ学会論文誌）, Vol. 6, , 2009.

[4] R.S. Sutton, A.G. Barto, 三上貞芳, 皆川雅章共. 強化学習. 森北出版, 東京, 2000.

[5] J. Peters and S. Schaal. Natural actor-critic. *Neurocomputing*, Vol. 71, No. 7-9, pp. 1180–1190, 2008.

[6] L.C. Baird. Advantage updating. 1993.

[7] S.J. Bradtke and A.G. Barto. Linear least-squares algorithms for temporal difference learning. *Machine Learning*, Vol. 22, No. 1, pp. 33–57, 1996.

[8] J.A. Boyan. Technical update: Least-squares temporal difference learning. *Machine Learning*, Vol. 49, No. 2, pp. 233–246, 2002.

[9] J. Peters, S. Vijayakumar, and S. Schaal. Reinforcement learning for humanoid robotics. In *Proceedings of the Third IEEE-RAS International Conference on Humanoid Robots*, 2003.

[10] 木村元. 適性度の履歴を用いた自然勾配 actor-critic 法．第 19 回自律分散システムシンポジウム, pp. 67–72, 2007.

[11] 前田哲彦, 長谷川裕夫, 伊藤宏充, 嶋川成浩, 松本和博. 九州地区の集合住宅におけるエネルギー需要の計測と解析（その１）. 第２０回エネルギーシステム・経済・環境コンファレンス 21-3, 2004.

[12] 前田哲彦, 丸山康司, 安幸, 伊藤宏充, 嶋川成浩. 九州地区の集合住宅におけるエネルギー需要の計測と解析（その２）. 第２０回エネルギーシステム・経済・環境コンファレンス 26-6, 2005.