Reinforcement Learning for Smart Grids and Stochastic Games

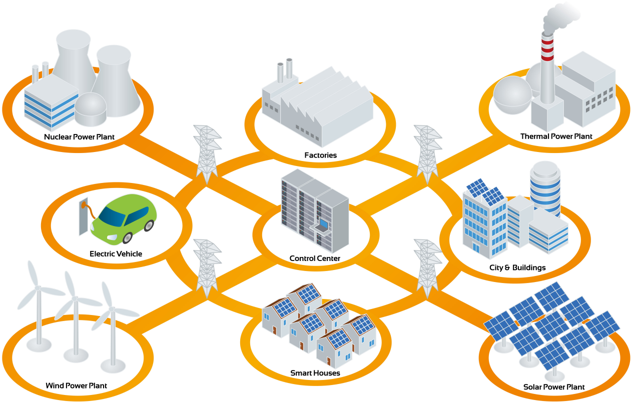
Sindhu P R
Advisor: Shalabh Bhatnagar

Dept. of Computer Science and Automation
Indian Institute of Science

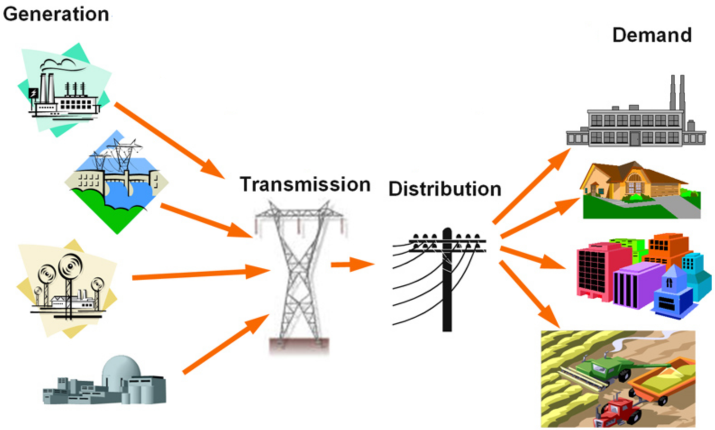
September 2, 2016

Overview

Smart Grid - Vision



Deregulated Markets



- Deregulated architecture: generation and distribution are open to competition

Power Pricing

Objective

Pricing of power in the wholesale and retail markets under generation and demand uncertainty

- Pricing is through day-ahead and spot markets
- Day-ahead market: Trading between sellers and buyers for the delivery of power on the following day
- Spot market: Power is traded for immediate delivery
- Day-ahead power price is determined by market clearing

Power Pricing

Objective

Pricing of power in the wholesale and retail markets under generation and demand uncertainty

- Pricing is through day-ahead and spot markets
- Day-ahead market: Trading between sellers and buyers for the delivery of power on the following day
- Spot market: Power is traded for immediate delivery
- Day-ahead power price is determined by market clearing

Market Clearing in Deregulated Market

- Power trading between multiple generators and retail utilities
- Market clearing price: Determined by auctions
- Auction Process: Matches electricity supply to demand at the lowest possible price point
- Each generator bids a generation capacity at a specific price
- Based on the demand, the lowest-priced combination of offers to meet demand are selected

Market Clearing - Example

Plant	Capacity Offered (MW)	Price (₹/MW)
Wind Farm	200	15
Nuclear Plant	1000	30
Coal Plant #1	500	40
Gas Plant #1	1000	50
Coal Plant #2	500	60
Gas Plant #2	200	200

Table: Price and Generation Bids

- Demand: $2,000$ MW
- Clearing price: ₹50 (Uniform Price Auction), ₹40 (Second Price Auction)

Issues in Restructuring

- Transmission Constraints
- Creation of market power:
  - Single or collection of generators profitably raise the price of power without losing market share
  - Significant market power occurs when prices exceed marginal cost
- Market power depends on:
  - Ease with which smaller generators can expand their output or new generators can enter the market
  - Market concentration

Problem Formulation

G: Number of generators, M: Number of consumers

On the t^{th} day,

- $q_t = (q_t^1, \ldots, q_t^G)$: Realized hourly quantity of power generated by generators on the $(t-1)^{\text{th}}$ day
- $p_t = (p_t^1, \ldots, p_t^G)$: Realized hourly prices on the $(t-1)^{\text{th}}$ day
- Assumption: $q_t^i \in Q$ and $p_t^i \in U$ take values in a discrete set
- Generator i bids prices for every hour of the following day, $a^i \in \mathbf{A}^i$

Modeling framework: General-Sum Stochastic game

Problem Formulation

G: Number of generators, M: Number of consumers

On the t^{th} day,

- $q_t = (q_t^1, \ldots, q_t^G)$: Realized hourly quantity of power generated by generators on the $(t-1)^{\text{th}}$ day
- $p_t = (p_t^1, \ldots, p_t^G)$: Realized hourly prices on the $(t-1)^{\text{th}}$ day
- Assumption: $q_t^i \in Q$ and $p_t^i \in U$ take values in a discrete set
- Generator i bids prices for every hour of the following day, $a^i \in \mathbf{A}^i$

Modeling framework: General-Sum Stochastic game

General-Sum Stochastic Games

Definition

Stochastic game with n agents: $\langle S, \mathbf{A}, r, P \rangle$

- State space: $S = (p_t, q_t)$
- Aggregate action space: $\mathbf{A} = \{\mathbf{A}^i\}$, $i = 1, \ldots, n$
- Set of actions available in state s: $\mathbf{A}(s) = \bigtimes_{i=1}^{n} \mathbf{A}^i(s)$
- Probability of transition from state s to s' under action $\mathbf{a}$: $P(s'|s, \mathbf{a})$
- Reward vector of all agents for the $(s, \mathbf{a})$ tuple: $r(s, \mathbf{a}) = [r^i(s, \mathbf{a}) : i = 1, \ldots, n]$, $\mathbf{a} \in \mathbf{A}(s)$
- Reward of generator i = quantity of power sold $\times$ clearing price

General-Sum Stochastic Games

Dynamics

- Markov property: Agents independently select actions $\mathbf{a} = (a^1, \ldots, a^n)$, based only on the current state
- Actions yield reward $r^i(s, \mathbf{a})$ to agent i

Strategy

- Stationary randomized strategy of agent i: $\pi^i = [\pi^i(s) : s \in S]$, $\pi^i \in \tilde{\Pi}^i$
- $\pi^i(s) \in H(\mathbf{A}^i(s))$, class of probability distributions over the set $\mathbf{A}^i(s)$
- Stationary strategies of all agents: $\pi = (\pi^1, \ldots, \pi^n) \in \tilde{\Pi}$
- Transition probability under policy π: $P^\pi(s, s') = \mathbb{E}\left[P(s'|s, \mathbf{a})|\pi\right]$

Objective

Expected Sum of Discounted Rewards

Agent i needs to maximize :

$$v^i(s,\pi) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}\left[r_t^i | \pi, s_0 = s\right]$$

where $0 < \beta < 1$ is the discount factor

Infinite Horizon Average Reward

Agent i needs to maximize :

$$v^i(s,\pi) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}\left[r_t^i | \pi, s_0 = s\right]$$

Objective

Expected Sum of Discounted Rewards

Agent i needs to maximize :

$$v^i(s,\pi) = \sum_{t=0}^{\infty} \beta^t \mathbb{E}\left[r_t^i | \pi, s_0 = s\right]$$

where $0 < \beta < 1$ is the discount factor

Infinite Horizon Average Reward

Agent i needs to maximize :

$$v^i(s,\pi) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}\left[r_t^i | \pi, s_0 = s\right]$$

Nash Equilibrium

Best Response Strategies

Tuple of n strategies $\pi_* = (\pi_*^1, \ldots, \pi_*^n)$ such that $\forall s \in S$, $\forall i \in (1, \ldots, n)$ and $\forall \pi^i \in \Pi^i$,

$$v^i(s, \pi_*^i, \pi_*^{-i}) \geq v^i(s, \pi^i, \pi_*^{-i})$$

Existence of Nash Equilibrium

- Require finite S and $\mathbf{A}^i$, $\forall i \in \{1, \ldots, n\}$
- At least one Nash equilibrium is guaranteed to exist in the space of stationary strategies ($\tilde{\Pi}$) for discounted and average reward criteria

Nash Equilibrium

Best Response Strategies

Tuple of n strategies $\pi_* = (\pi_*^1, \ldots, \pi_*^n)$ such that $\forall s \in S$, $\forall i \in (1, \ldots, n)$ and $\forall \pi^i \in \Pi^i$,

$$v^i(s, \pi_*^i, \pi_*^{-i}) \geq v^i(s, \pi^i, \pi_*^{-i})$$

Existence of Nash Equilibrium

- Require finite S and $\mathbf{A}^i$, $\forall i \in \{1, \ldots, n\}$
- At least one Nash equilibrium is guaranteed to exist in the space of stationary strategies $(\tilde{\Pi})$ for discounted and average reward criteria

Future Directions

- Find stationary deterministic pricing strategies for smart grid using iterative tabular model-free algorithms [1,2]
- Compute Herfindahl-Hirschman Index (HHI) to determine the market power of the generators
- Cardinality of state-space: $|Q|^{24*M} \times |U|^{24*G}$
- Methods are not scalable
- Scalable algorithms need to be designed - function approximation + actor-critic structure

[1] Li, Jun, K Ramachandran, and T K. Das. "A Reinforcement Learning (Nash-R) Algorithm for Average Reward Irreducible Stochastic Games." Journal of Machine Learning Research (2007)

[2] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." Journal of machine learning research (2003)

Future Directions

- Find stationary deterministic pricing strategies for smart grid using iterative tabular model-free algorithms [1,2]
- Compute Herfindahl-Hirschman Index (HHI) to determine the market power of the generators
- Cardinality of state-space: $|Q|^{24*M} \times |U|^{24*G}$
- Methods are not scalable
- Scalable algorithms need to be designed - function approximation + actor-critic structure

[1] Li, Jun, K Ramachandran, and T K. Das. "A Reinforcement Learning (Nash-R) Algorithm for Average Reward Irreducible Stochastic Games." Journal of Machine Learning Research (2007)

[2] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." Journal of machine learning research (2003)

Future Directions

- Find stationary deterministic pricing strategies for smart grid using iterative tabular model-free algorithms [1,2]
- Compute Herfindahl-Hirschman Index (HHI) to determine the market power of the generators
- Cardinality of state-space: $|Q|^{24*M} \times |U|^{24*G}$
- Methods are not scalable
- Scalable algorithms need to be designed - function approximation + actor-critic structure

[1] Li, Jun, K Ramachandran, and T K. Das. "A Reinforcement Learning (Nash-R) Algorithm for Average Reward Irreducible Stochastic Games." Journal of Machine Learning Research (2007)

[2] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." Journal of machine learning research (2003)

Future Directions

- Find stationary deterministic pricing strategies for smart grid using iterative tabular model-free algorithms [1,2]
- Compute Herfindahl-Hirschman Index (HHI) to determine the market power of the generators
- Cardinality of state-space: $|Q|^{24*M} \times |U|^{24*G}$
- Methods are not scalable
- Scalable algorithms need to be designed - function approximation + actor-critic structure

---

[1] Li, Jun, K Ramachandran, and T K. Das. "A Reinforcement Learning (Nash-R) Algorithm for Average Reward Irreducible Stochastic Games." Journal of Machine Learning Research (2007)

[2] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." Journal of machine learning research (2003)

Future Directions

- Find stationary deterministic pricing strategies for smart grid using iterative tabular model-free algorithms [1,2]
- Compute Herfindahl-Hirschman Index (HHI) to determine the market power of the generators
- Cardinality of state-space: $|Q|^{24*M} \times |U|^{24*G}$
- Methods are not scalable
- Scalable algorithms need to be designed - function approximation + actor-critic structure

[1] Li, Jun, K Ramachandran, and T K. Das. "A Reinforcement Learning (Nash-R) Algorithm for Average Reward Irreducible Stochastic Games." Journal of Machine Learning Research (2007)

[2] Hu, Junling, and Michael P. Wellman. "Nash Q-learning for general-sum stochastic games." Journal of machine learning research (2003)

Overview

Constrained Setting

- Agent incurs cost while picking an action
- Need to keep the costs incurred within bounds
- B_i: Total number of cost functions associated with agent i
- $c_j^i(s, \mathbf{a})$: j^{th} cost function of agent i when state is $s \in S$ and the vector of actions $\mathbf{a}$ is chosen, $(1 \leq j \leq B_i)$
- D_j^i: Bound on the discounted cost for agent i's j^{th} discounted cost function

Constrained Setting (contd.)

Notation

- $c_j^i(\pi) = (c_j^i(1,\pi), \ldots, c_j^i(|S|,\pi))^\top$: Expected j^{th} cost vector of agent i for policy π
- $M_j^i(\pi) = [M_j^i(s,\pi) : s \in S] = (I - \beta P^\pi)^{-1} c_j^i(\pi)$: Discounted type j cost under policy π
- $r^i(\pi) = (r^i(1,\pi), \ldots, r^i(|S|,\pi))^\top$: Expected Reward vector of agent i for policy π
- $v^i(\pi) = (I - \beta P^\pi)^{-1} r^i(\pi)$

Feasible Strategies

- $h(s)$: Initial distribution on states
- $\Pi_f^i = \left\{ \pi \in \tilde{\Pi} : (1-\beta) \sum_{s \in S} h(s) M_j^i(s,\pi) \leq D_j^i, 1 \leq j \leq B_i, \right\}$
- $\Pi_f = \bigcap_{i=1}^{n} \Pi_f^i$: Set of all feasible strategies

Nash Equilibrium

$\pi_* \in \Pi_f$ and $\forall i \in \{1, \ldots, n\}$, $\forall [\pi^i, \pi_*^{-i}] \in \Pi_f$,

$$h^\top v^i(\pi_*) \geq h^\top v^i(\pi^i, \pi_*^{-i})$$

Does Nash equilibrium exist in the set of stationary strategies ?

Feasible Strategies

- $h(s)$: Initial distribution on states
- $\Pi_f^i = \left\{ \pi \in \tilde{\Pi} : (1-\beta) \sum_{s \in S} h(s) M_j^i(s,\pi) \leq D_j^i, 1 \leq j \leq B_i, \right\}$
- $\Pi_f = \bigcap_{i=1}^{n} \Pi_f^i$: Set of all feasible strategies

Nash Equilibrium

$\pi_* \in \Pi_f$ and $\forall i \in \{1, \ldots, n\}$, $\forall [\pi^i, \pi_*^{-i}] \in \Pi_f$,

$$h^\top v^i(\pi_*) \geq h^\top v^i(\pi^i, \pi_*^{-i})$$

Does Nash equilibrium exist in the set of stationary strategies ?

Future Directions

- Existence[3] of Nash equilibrium: Guaranteed under strict inequality constraints on costs
- Equilibrium can be computed if model parameters are known
- P ✗, R ✗, C ✗
- RL algorithms to find Nash equilibria in constrained general-sum stochastic games
- Explore learning of correlated equilibria

[3] Vinayaka G. Yaji, Shalabh Bhatnagar, " Necessary and sufficient conditions for optimality in constrained general sum stochastic games", Systems & Control Letters, Volume 85, November 2015

Future Directions

- Existence[3] of Nash equilibrium: Guaranteed under strict inequality constraints on costs
- Equilibrium can be computed if model parameters are known
- P ✗, R ✗, C ✗
- RL algorithms to find Nash equilibria in constrained general-sum stochastic games
- Explore learning of correlated equilibria

[3] Vinayaka G. Yaji, Shalabh Bhatnagar, " Necessary and sufficient conditions for optimality in constrained general sum stochastic games", Systems & Control Letters, Volume 85, November 2015

Future Directions

- Existence[3] of Nash equilibrium: Guaranteed under strict inequality constraints on costs
- Equilibrium can be computed if model parameters are known
- P ✗, R ✗, C ✗
- RL algorithms to find Nash equilibria in constrained general-sum stochastic games
- Explore learning of correlated equilibria

[3] Vinayaka G. Yaji, Shalabh Bhatnagar, " Necessary and sufficient conditions for optimality in constrained general sum stochastic games", Systems & Control Letters, Volume 85, November 2015

Future Directions

- Existence[3] of Nash equilibrium: Guaranteed under strict inequality constraints on costs
- Equilibrium can be computed if model parameters are known
- P ✗, R ✗, C ✗
- RL algorithms to find Nash equilibria in constrained general-sum stochastic games
- Explore learning of correlated equilibria

[3] Vinayaka G. Yaji, Shalabh Bhatnagar, " Necessary and sufficient conditions for optimality in constrained general sum stochastic games", Systems & Control Letters, Volume 85, November 2015

Future Directions

- Existence[3] of Nash equilibrium: Guaranteed under strict inequality constraints on costs
- Equilibrium can be computed if model parameters are known
- P ✗, R ✗, C ✗
- RL algorithms to find Nash equilibria in constrained general-sum stochastic games
- Explore learning of correlated equilibria

[3] Vinayaka G. Yaji, Shalabh Bhatnagar, " Necessary and sufficient conditions for optimality in constrained general sum stochastic games", Systems & Control Letters, Volume 85, November 2015

Thank You

Questions

References I

J. Filar and K. Vrieze, *Competitive Markov decision processes*. Springer-Verlag, New York, 1997.

C. Joe-Wong, S. Sen, S. Ha, and M. Chiang, "Optimized day-ahead pricing for smart grids with device-specific scheduling flexibility," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 6, pp. 1075–1085, 2012.

M. Rahimiyan and H. R. Mashhadi, "An adaptive Q-learning algorithm developed for agent-based computational modeling of electricity market," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 40, no. 5, pp. 547–556, 2010.

References II

- P. P. Reddy and M. M. Veloso, "Strategy learning for autonomous agents in smart grid markets," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence - Volume Volume Two*, ser. IJCAI'11. AAAI Press, 2011, pp. 1446–1451. [Online]. Available: http://dx.doi.org/10.5591/978-1-57735-516-8/IJCAI11-244
- M. Peters, W. Ketter, M. Saar-Tsechansky, and J. Collins, "A reinforcement learning approach to autonomous decision-making in smart electricity markets," *Machine learning*, vol. 92, no. 1, pp. 5–39, 2013.
- V. Nanduri and T. K. Das, "A reinforcement learning model to assess market power under auction-based energy pricing," *IEEE Transactions on Power Systems*, vol. 22, no. 1, pp. 85–95, Feb 2007.