A Reinforcement-learning based Energy Plan Selection Approach for Energy Markets with Retail Choice

Yuze He
Department of Information Enginerring
The Chinese University of Hong Kong
Shatin, Hong Kong
hy019@ie.cuhk.edu.hk

Xiang Pan
Department of Information Enginerring
The Chinese University of Hong Kong
Shatin, Hong Kong
px018@ie.cuhk.edu.hk

Abstract

Recently, customers are provided with the opportunity to switch energy plans among competitive retail energy suppliers. However, few customers are likely to switch energy plans even if they may get better savings due to the complex time-dependent decision-making process for selecting the best energy plan under the uncertainty of the demand. In this paper, we propose a reinforcement-learning (RL) based approach for the energy plan selection problem. We formulate the energy plan selection problem as a Markov Decision Process (MDP) problem with the switching costs and train an off-policy reinforcement learning (RL) algorithm to solve it. In the execution phase, the trained RL agent determines the best energy plan at any time step.

1 Introduction

Retail choice in energy markets aims to provide diverse options to residential, industrial and commercial customers, where the customers are able to purchase electricity and natural gas from multiple competitive retail suppliers [1]. In competitive energy retail markets (including electricity and natural gas), energy suppliers can compete in an open marketplace by providing a range of services and tariff schemes. With retail choice, customers can compare and choose different energy plans from multiple suppliers, and determine the best energy plans that are the suitable. Usually, the switching from one supplier to another can be attained smoothly via an online platform, or through third-party assistance services.

Despite the promising benefits of retail choice in energy markets, there appears subsided participation particularly from residential customers. Even worse, declining residential participation rates and diminishing market shares of competitive retailers have been reported in the recent years [2, 3]. While there are some reasons behind the customers' reactions, the major reason is the complication of the available energy plans in energy retail markets [4]. Usually, the energy retailers' tariff structures are complex, which is not friendly to the consumers. Meanwhile, savings and incentives among energy retailers are not easy to compare. Without discerning the expected benefits of switching their energy plans, most customers are reluctant to participate in energy retail markets. Besides, the increasing market complexity with a growing number of retailers and agents obscures the benefits of retail choice. To sum up, the challenges arisen in the decision-making processes for energy plan selection are follows [5]:

- **Complex Tariff Structures.** There are diverse tariff structures e.g., different contract periods (e.g., 6 or 12 months), which may lead to different peak tariffs as well as dynamic pricing depending on renewable energy sources.

- **Uncertain Future Information.** Future information including the usage and the fluctuation of energy tariffs is important in deciding the best energy plan. However, it is difficult to predict the uncertain information accurately.

The government are always anticipating a proper decision-making processes for energy plan selection in energy markets with retail choice [6, 7, 8]. For this, this proposal aims to leverage reinforcement learning for the energy plan selection problem. Suppose we can train an RL agent based on the historical data, then customer can directly use the agent to select the most suitable energy plan without facing the complication of the available energy plans themselves.

1.1 Related Work

The energy plan selection problem is closely related to the category of the online convex optimization problem with the switching cost [9]. For the general online optimization problem, when the state space set is discrete, the problem is known as Metrical Task System (MTS) problem [10] in a general setting, and several works e.g., the energy generation scheduling [11] also leverage the studies in this field. Meanwhile, several works focus on the online convex optimization problem with switching costs and continuous state space like LCP [12]. Recently, [5] construct a competitive online algorithm based on LCP for the energy plan selection problem with temporally dependent switching cost.

2 Framework

We present a formulation of the simplified energy plan selection problem in this section. [1]

2.1 Model

Like [5], we consider a typical setting where a consumer can select the energy plan from one energy retailers, where the fluctuating demands, the time-varying prices, and the cancellation fees are considered.

2.1.1 Electricity Demands

Let the electricity demand at time t be $e(t)$. We show an illustration of the demands of four clients as examples in Fig. 1.

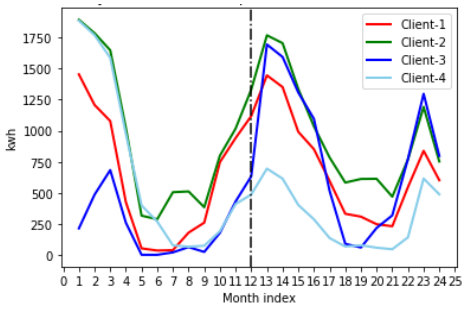


Figure 1: Illustrations of the electricity demand of families from 2015 to 2016.

2.1.2 Energy Plan

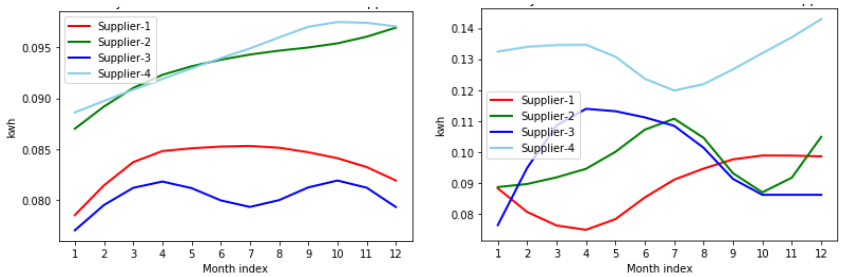
We consider two energy plans [5]: the fixed-rate plan (represented by 0) and the variable-rate plan (represented by 1). Here we use the case with single supplier to explain the model, which can be easily extend to multiple supplier case by enlarging the dimension of the action space. We show an illustration of the demands of four clients as examples in Fig. 3. According to [5], the energy plan of the current is determined at the end of the previous month, which means the clients need to select the

[1] The multiplier supplier case can be extended by enlarging the set of action space.

energy plan without knowing the actual demand. It means we cannot directly applies the classical dynamic programming to solve this problem.

- **Variable-Rate Plan.** The consumers can always switch between different variable-rate plans without any cost. Suppose the price per electricity unit for the variable-rate plan is p_t^1, the consumption charge at time t is $p_t^1 \cdot e_t$.
- **Fixed-Rate Plan.** Fixed-rate plans is characterized by a stable ranking for a long period of time. Meanwhile, there is a relatively high cancellation fee for the fixed-rate plan. With this understanding, we assume that a consumer will not switch between fixed-rate plans. A fixed-rate plan will not offer the same electricity price for arbitrary demand, but a tiered-pricing scheme up for a certain level of demand [11]. Let B_t be a base load for a consumer for each time t on the previous year, and p_t^0 be the price per electricity unit for a fixed-rate plan. A consumer will pay $B_t \cdot p_t^0$ if the demand is between $0.9B_t$ and $1.1B_t$. Otherwise, an under-usage fee will be charged at rate H when e_t is less than $0.9B_t$, or an over-usage fee will be charged at the same rate of a variable-rate plan when e_t is more than $1.1B_t$. The cost function $g_t(s_t)$ based on the input $\sigma_t \triangleq \left(e_t, p_t^0, p_t^1, B_t\right)$ and the selected plan s_t at time t is defined as:

$$g_t\left(s_t\right) = \begin{cases} p_t^1 \cdot e_t, \text{if } s_t = 1 \\ e_t \cdot p_t^0 + \left(p_t^1 - p_t^0\right) \cdot \left(e_t - 1.1B_t\right)^+ - H \cdot \left(0.9B_t - e_t\right)^+, \text{if } s_t = 0 \end{cases} \tag{1}$$



(a) (b)

Figure 2: Illustrations of electricity prices of both fixed-rate plan and variable-rate plan from different suppliers.

2.1.3 Cancellation Fee

When a customer cancels a fixed-rate plan, there will be a cancellation fee of the following types:

- **No Fee.** The customer does not need to pay any cancellation fee.
- **Constant Fee.** If the residual number of months in the contract is within a certain level, then a fixed amount of cancel fee is required.

Let β_t be the cancellation fee at time t.

2.2 Problem Formulation

We first consider a simplified version of this problem. we focus on the setting with a constant cancellation fee. Meanwhile, it is common for retailers to offer fixed-rate plans without a minimum contract period according to [5, 13]. Thus, the Energy Plan Selection Problem can be formulated as follows. Given the total time period T is divided into integer slots $\mathcal{T} \triangleq \{1,...,T\}$ each is assumed to last for one month. The goal is to find a solution $\boldsymbol{s} \triangleq (s_1, s_2, ..., s_T)$ to the following optimization problem:

$$\min \sum_{t=1}^{T} \left(g_t\left(s_t\right) + \beta \cdot \left(s_t - s_{t-1}\right)^+\right), \text{variables } s_t \in \{0,1\}. \tag{2}$$

where $(x)^+ = \max(x, 0)$. Function $g_t(s_t)$ is defined in (1). Without loss of generality, the initial state s_0 is set to be 0. The cost function is consisted of operational cost and switching cost when cancelling a fixed-rate plan. This formulation can be extended by adding the minimum contract length for the fixed-rate plan and the non-constant cancelling fee.

2.3 Problem Formulation

We first consider a simplified version of this problem. we focus on the setting with a constant cancellation fee. Meanwhile, it is common for retailers to offer fixed-rate plans without a minimum contract period according to [5, 13]. Thus, the Energy Plan Selection Problem can be formulated as follows. Given the total time period T is divided into integer slots $\mathcal{T} \triangleq \{1,...,T\}$ each is assumed to last for one month. The goal is to find a solution $\boldsymbol{s} \triangleq (s_1, s_2, ..., s_T)$ to the following optimization problem:

$$\min \sum_{t=1}^{T} \left(g_t\left(s_t\right) + \beta \cdot \left(s_t - s_{t-1}\right)^+ \right), \text{variables } s_t \in \{0,1\}\,. \tag{3}$$

where $(x)^+ = \max(x, 0)$. Function $g_t(s_t)$ is defined in (1). Without loss of generality, the initial state s_0 is set to be 0. This formulation can be extended by adding the minimum contract length for the fixed-rate plan and the non-constant cancelling fee.

3 Proposed Approach

To solve this energy plan selection problem, we need to formulate this problem into a MDP and use reinforcement learning to solve this problem.

3.1 Q-learning

From Fig. 1, we found that the demands for the clients for the same months on different year are similar. For example, we can see there is a peak at the beginning of the year and a valley in the middle of the year. With this observation, we first consider one simple and straight idea, which only uses the time index as the state (as we mentioned before, the time index somehow corresponds to the demand). The action set consists of all available plans at the current time, and the reward is regarded as the saving as compared to the baseline method e.g., apply fixed-rate plan for all months. Noted that here both the states and the action sets are discrete, thus we use the Q-learning method to train the agent.

Algorithm 1 Q-learning based algorithm to train agent for energy plan select problem (one supplier case).

Require: $S = 0, 1, 2, ..., 11$, $A = 0, 1$, $Q(s, a) \in R^{12\times 2}$.

1: Initialize each element in Q table to 0
2: **repeat**
3: Choose a from A using ϵ-greedy and take action a
4: Update the state to $s' = s + 1$
5: Calculate reward as the cost of baseline plan minus cost of chosen plan for current month
6: Update Q-value:$Q\left(s, a\right) \leftarrow Q\left(s, a\right) + \alpha\left[R + \gamma \max_a Q\left(s', a\right) - Q\left(s, a\right)\right]$
7: **until** s equals to 11

3.2 Deep-Q learning

We found that only using the time index may not be enough. With the demand in the previous months, the agent may can be able to make better choice. Thus, we further include the demand of the previous months into the state. Then, we the state set is not discrete; thus we apply the deep Q learning in this approach.

Algorithm 2 Deep-Q learning based algorithm to train agent for energy plan select problem (one supplier case).

Require: $S = 0, 1, 2, ..., 11$, $A = 0, 1$, Q-table represented by DNN.

1: Initialize replay buffer, and the actual network (set of parameters is θ) and the target network (set of parameters is θ')
2: pre-process and the environment and feed state to DQN (NN used in the experiment is with 1hidden layer, 128 neurons)
3: **repeat**
4: Choose a from A using ϵ-greedy and take action a
5: Update the state to $s' = s + 1$ and reward R
6: store transition in replay buffer as $< S, a, R, S' >$
7: sample some random batches of transitions and calculate the loss:

$$\left(R + \gamma \max_{a'} Q\left(S', a'; \theta\right) - Q\left(S', a'; \theta'\right)\right)^2$$

8: Tune network parameters to minimize this loss.
9: After every k steps, copy our actual network weights to the target network weights
10: **until** the number of training episodes equals to M

4 Experiments

We use real-world data of household demands and retailers' energy plan prices in our simulations.

4.1 Dataset

4.1.1 Household Electricity Consumption

Household Electricity Consumption. We use the electricity consumption data from Smart* project [14], which consists of 114 single-family apartments recorded in 2015-2016.

4.1.2 Energy Plan Prices

We consider the historic data of energy plan prices from the real-world official dataset in New York State, including 12-month fixed-rate energy plans and one-month variable-rate energy plans. Note that fixed-rate energy plans can be associated with different kinds of fees when customers switch their selections. Details can be found at http://documents.dps.ny.gov/PTC/.

4.1.3 Performance Evaluation

We test our proposed algorithms to evaluate the potential saving of selecting an appropriate energy plan based on the two datasets we mentioned before. We consider the case of a household switching between a fixed-rate energy plan and a variable-rate energy plan in a specific year. Specifically, we suppose that the test household choose the energy plan by taking the previous year consumption into account and the retailer can provide either a variable-rate energy plan or a fixed-rate energy plan. The length of a fixed-rate plan is set as 12 months, with constant cancellation fee. We compare our algorithm for the same application scenario with the fixed-rate plan for 12 months choice.[2]

4.2 Experiment Results

4.2.1 Experiment result of Q-learning

We use 80 families' electricity consumption data for training and use the remaining 34 families' electricity consumption data for testing. Fig. 3 is the illustration of final Q table.

[2]Here we only show the comparison results between the fixed-rate plan as an example. For the comparisons with other potential algorithm, we leave it to future investigation.

	0	1	2	3	4	5	6	7
0	10.105	1.778	1.335	-2.140	10.124	14.740	-8.180	-68.451
1	7.436	7.186	-1.245	1.620	11.689	-0.459	-0.341	-84.193
2	6.697	-55.545	-3.584	-97.613	11.220	-102.868	-0.789	-141.615
3	5.899	-85.241	-2.627	-91.649	9.696	-117.993	-0.984	-123.839
4	4.859	-93.376	0.660	-98.268	8.774	-94.369	-0.875	-115.072
5	3.757	-85.396	3.305	-99.495	7.204	-60.214	1.428	-67.074
6	3.515	-96.769	2.003	-101.666	6.292	-100.678	2.082	-89.446
7	2.798	-96.204	1.780	-62.604	5.879	-84.084	1.578	-55.286
8	2.865	-82.456	-0.418	-96.797	5.495	-96.778	0.336	-108.361
9	2.456	-48.487	-3.669	-94.779	5.121	-83.177	-5.803	-112.897
10	1.274	-109.133	-16.495	-111.130	4.087	-101.481	-11.331	-139.392
11	0.000	-99.250	-12.387	-106.899	1.807	-81.106	-10.316	-126.692

Figure 3: Q table

To evaluate Q-learning, the comparison between the total cost of 34 families using Q-learning algorithm and the total cost of 34 families staying in fixed-rate plan for 12 months is shown in the Fig. 4. It can be observed that the cost of Q learning is lower than the cost of baseline.

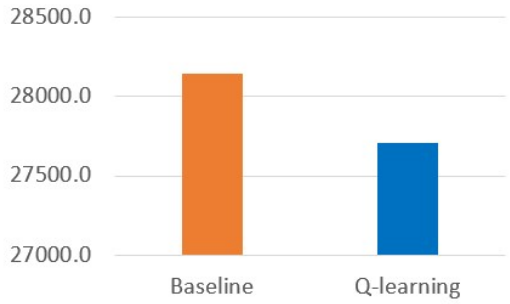


Figure 4: cost comparisons between applying Q-learning algorithm and baseline

4.2.2 Experiment result of Deep Q Network

We use 89 families' electricity consumption data for training, and the remaining 25 families' electricity consumption data for testing. NN used in the experiment is with 1 hidden layer, 128 neurons. The comparison of the 25 families' total cost between baseline plan and our Deep Q learning algorithm is shown in Fig. 5. It is apparent that the cost of Deep Q learning is much lower than the cost of baseline plan. More specifically, Fig. 6 shows the cost in a whole year of one of the 25 families, which illustrate that we can get savings in most months by applying Deep Q learning algorithm.

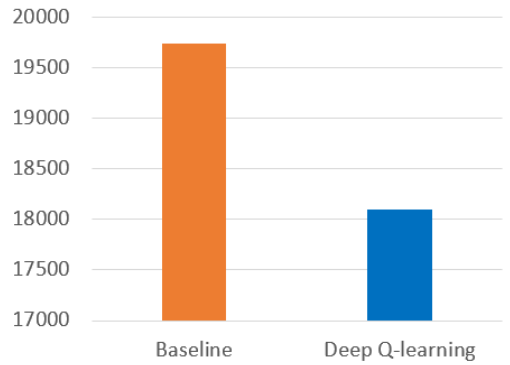


Figure 5: cost comparisons between applying Deep Q Network and baseline

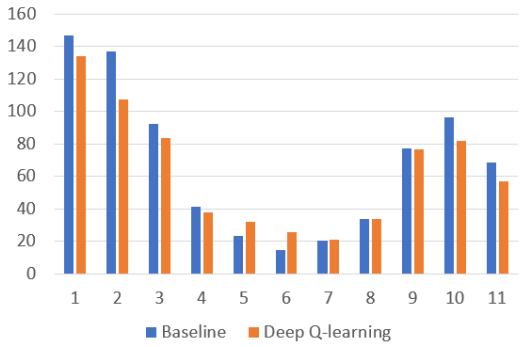


Figure 6: cost comparisons between applying Deep Q Network and baseline by month (one customer)

5 Conclusion

In this paper, a reinforcement-learning (RL) based approach is proposed for the energy plan selection problem, which can help customers to decide whether to switch energy plans even under the uncertainty of the demand for better savings. Specifically, we formulate the energy plan selection problem as a Markov Decision Process (MDP) problem with the switching costs and train an off-policy reinforcement learning (RL) algorithm to solve it. In the execution phase, the trained RL agent can be applied directly to determine the action at any time step.

References

[1] Shengru Zhou. An introduction to retail electricity choice in the united states. Technical report, National Renewable Energy Lab. (NREL), 2017.

[2] U.S. Energy Information Administration. Electricity residential retail choice participation has declined since 2014 peak, nov 2018.

[3] Frank Graves, Agustin Ros, Sanem Sergici, Rebecca Carroll, and Kathryn Haderlein. *Retail Choice: Ripe for Reform?* The Brattle Group, Boston, MA, USA, 2018.

[4] AEMC. 2018 retail energy competition review. Final report, Sydney, Australia, June 2018.

[5] Jianing Zhai, Sid Chi-Kin Chau, and Minghua Chen. Stay or switch: Competitive online algorithms for energy plan selection in energy markets with retail choice. In *the Tenth ACM International Conference*, 2019.

[6] Office of Gas and Electricity Markets (Ofgem). Compare gas and electricity tariffs: Ofgem-accredited price comparison sites, 2018.

[7] Public Utility Commission of Texas. Power to choose, 2018.

[8] Australian Energy Regulator. Energy made easy, 2018.

[9] Nikhil Bansal, Anupam Gupta, Ravishankar Krishnaswamy, Kirk Pruhs, Kevin Schewior, and Cliff Stein. A 2-Competitive Algorithm For Online Convex Optimization With Switching Costs. In Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 96–109, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[10] Allan Borodin and Ran El-Yaniv. *Online Computation and Competitive Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[11] Lian Lu, Jinlong Tu, Chi-Kin Chau, Minghua Chen, and Xiaojun Lin. Online energy generation scheduling for microgrids with intermittent energy sources and co-generation. *ACM SIGMETRICS Performance Evaluation Review*, 41(1):53, 2013.

[12] Minghong Lin, Adam Wierman, Lachlan L. H. Andrew, and Eno Thereska. Dynamic right-sizing for power-proportional data centers. In *2011 Proceedings IEEE INFOCOM*, pages 1098–1106. IEEE, 2011.

[13] Department of Public Service New York State. Nys power to choose, 2018.

[14] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, and Jeannie Albrecht. Smart*: An open data set and tools for enabling research in sustainable homes. In

Proceedings of the 2012 Workshop on Data Mining Applications in Sustainability, SustKDD 2012, 2012.