Project RL

Energy storage optimization

Final Presentation

Emile Dhifallah, Wenhua Hu, Felix Nastar

Content for today

Data

Problem & environment setup

Method

Experimental results

Visualization on test

Perspectives

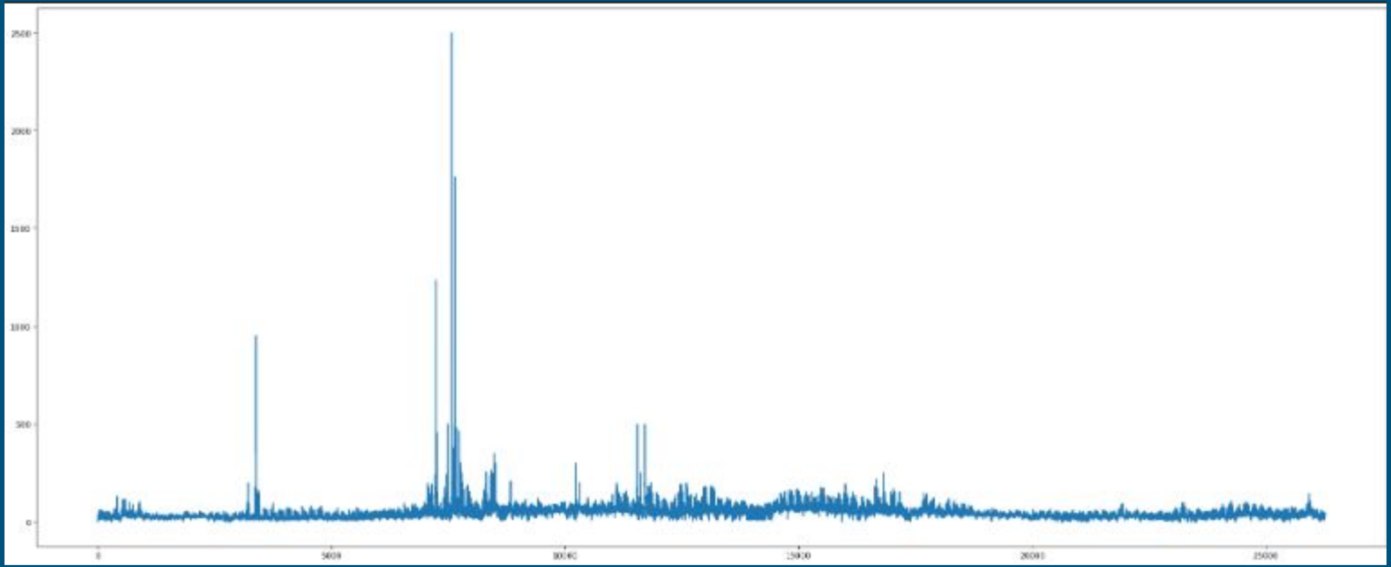
Data

26k hourly price samples

Seasonality

Slight fluctuations

Big Outliers
(1%) > 150



Problem & environment setup

Actions: Discrete action space, ranging from -1 (sell) to +1 (3-5)

State: 1) electricity price (3-5 bins)
2) hour of the day (3-24)
3) battery level (6-11)

One episode of the whole trajectory, no termination for different days

Reward: positive reward for selling electricity, Negative reward for buying

Methods

Tabular methods: Random & Q Learning

Discrete action space: [-1,0,1] or [-1, -0.5, 0, 0.5, 1]

State discretized for Battery Levels, Electricity Price and Hours

- Battery Levels (0-50 kwh): 6 or 11 bins

  E.g. [0,10) [10,20) [20,30) [30,40) [40,50) [50, +inf)

- Electricity Price (0-2500 €): 3 or 5 bins

  Use quartiles

  E.g. [0.01, 29.9 , 43, 65, 150] => [0.01, 29.9), [29.9 , 43), [43, 65), [65, 150), [150, **+inf**)

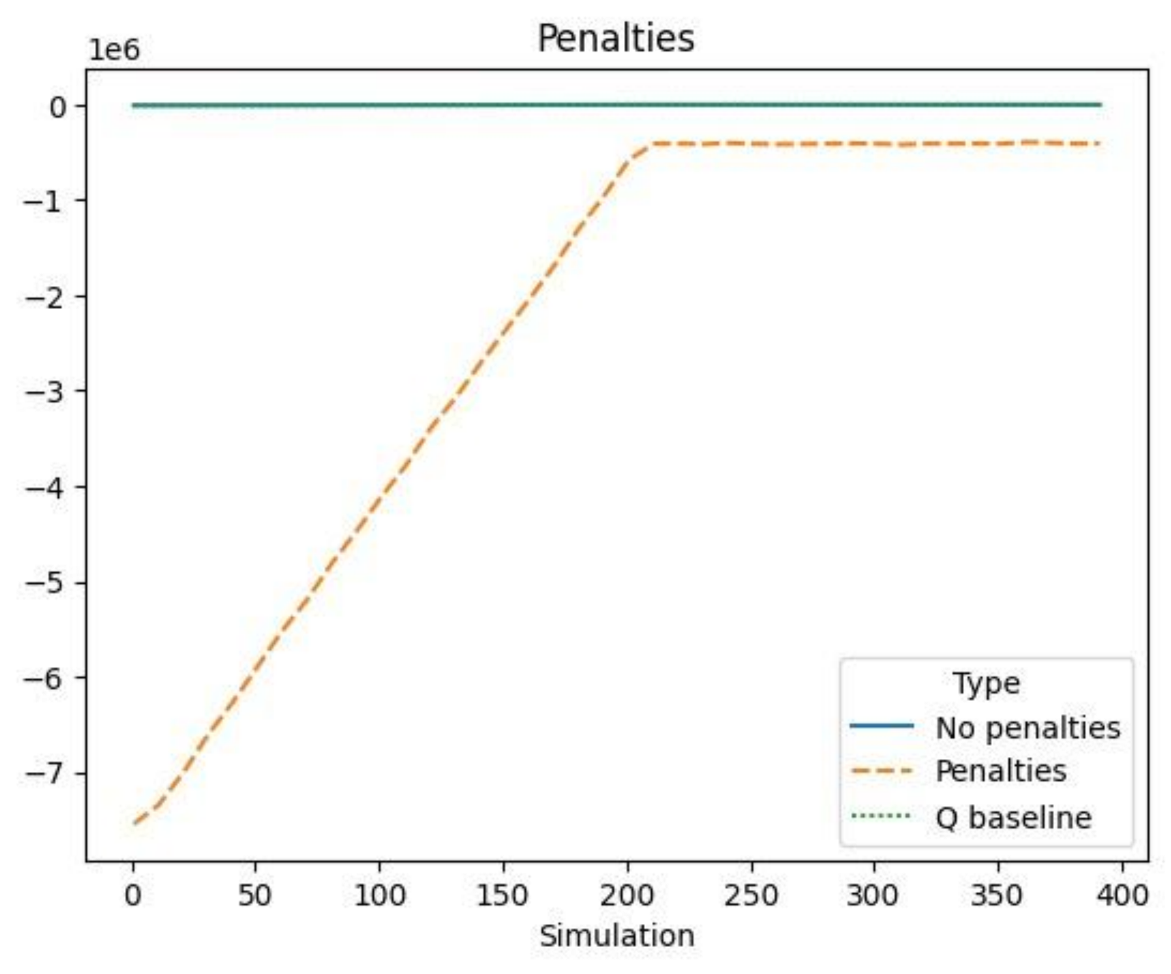
- Hours (24 hs): 3 or 24 bins

Methods

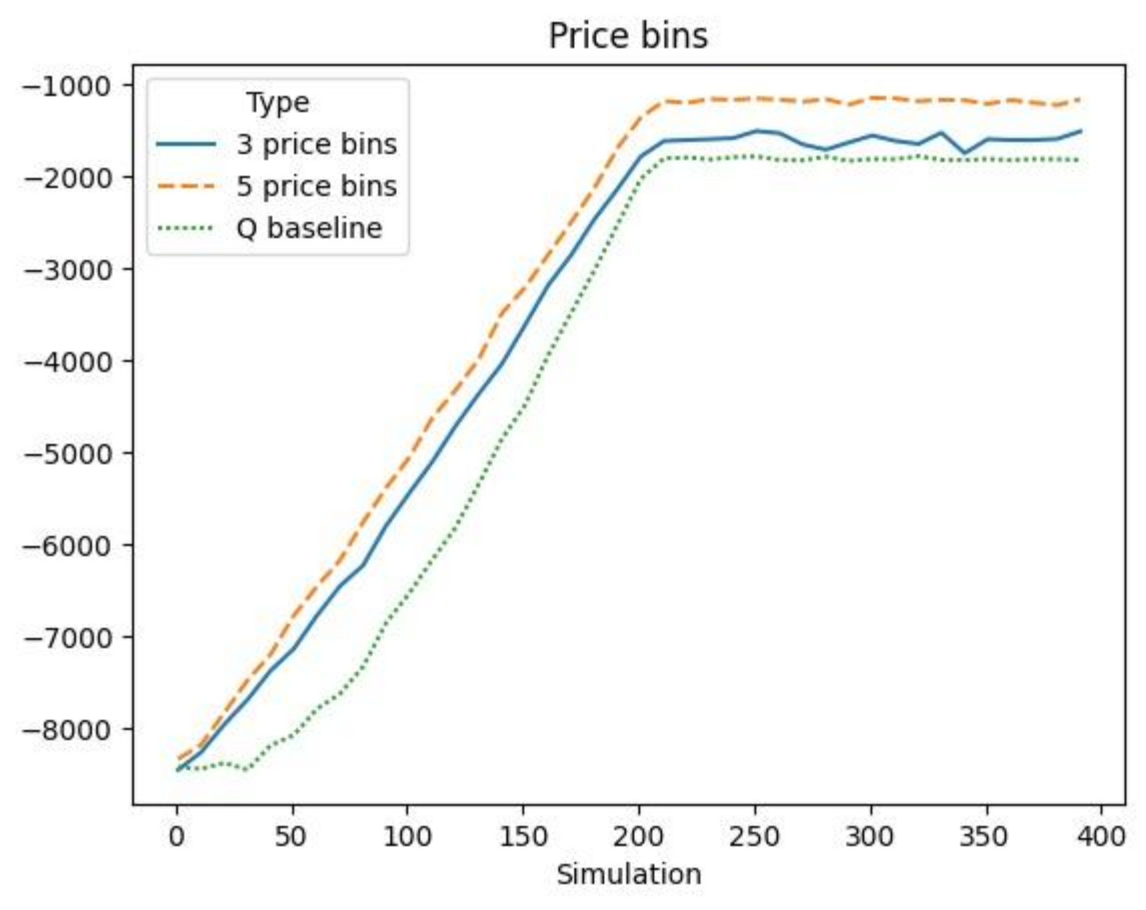
Reward shaping

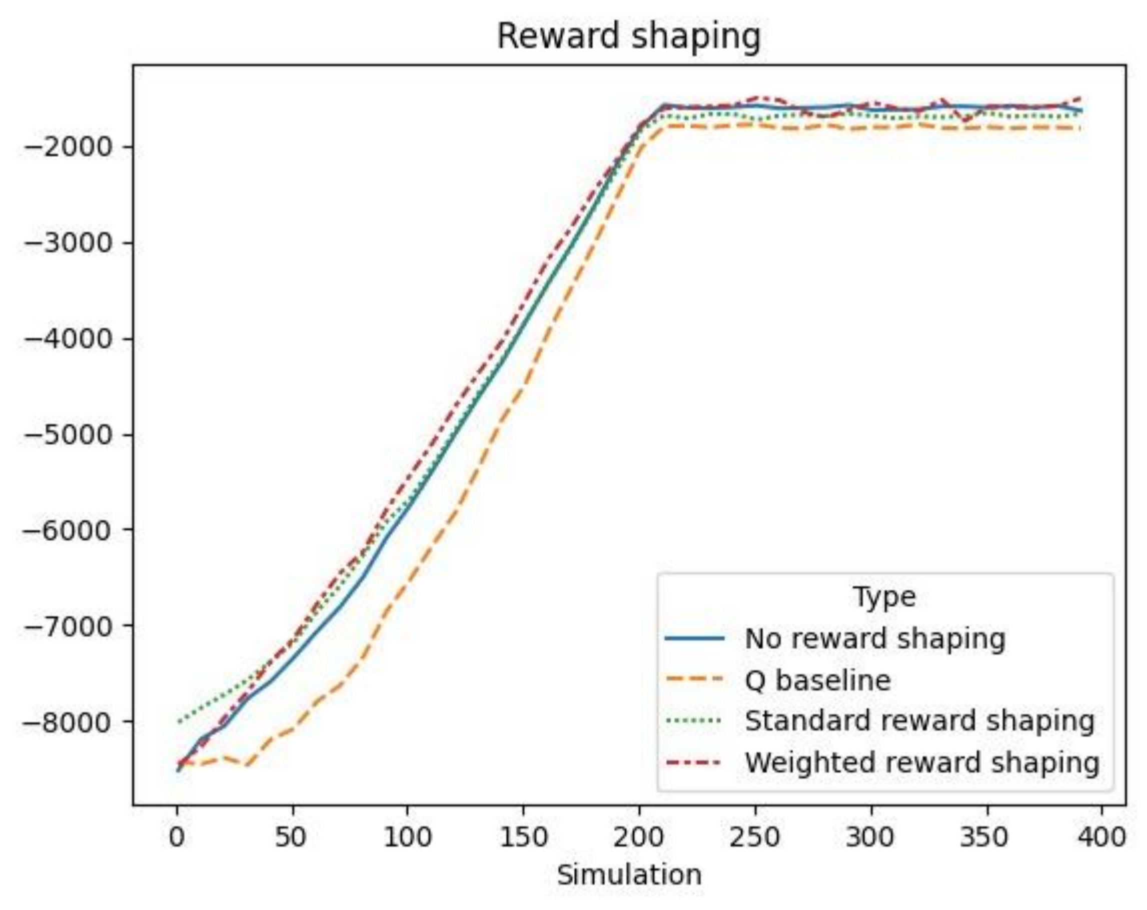
```
       weight = 100.0

        if next_battery_level < battery_level: # sell
            extra_reward = (electricity_price - 2 * next_electricity_price) * (battery_level -
next_battery_level)
        if next_battery_level > battery_level: # buy
            extra_reward = (2 * electricity_price - next_electricity_price) * (battery_level -
next_battery_level)

        return weight * extra_reward
```

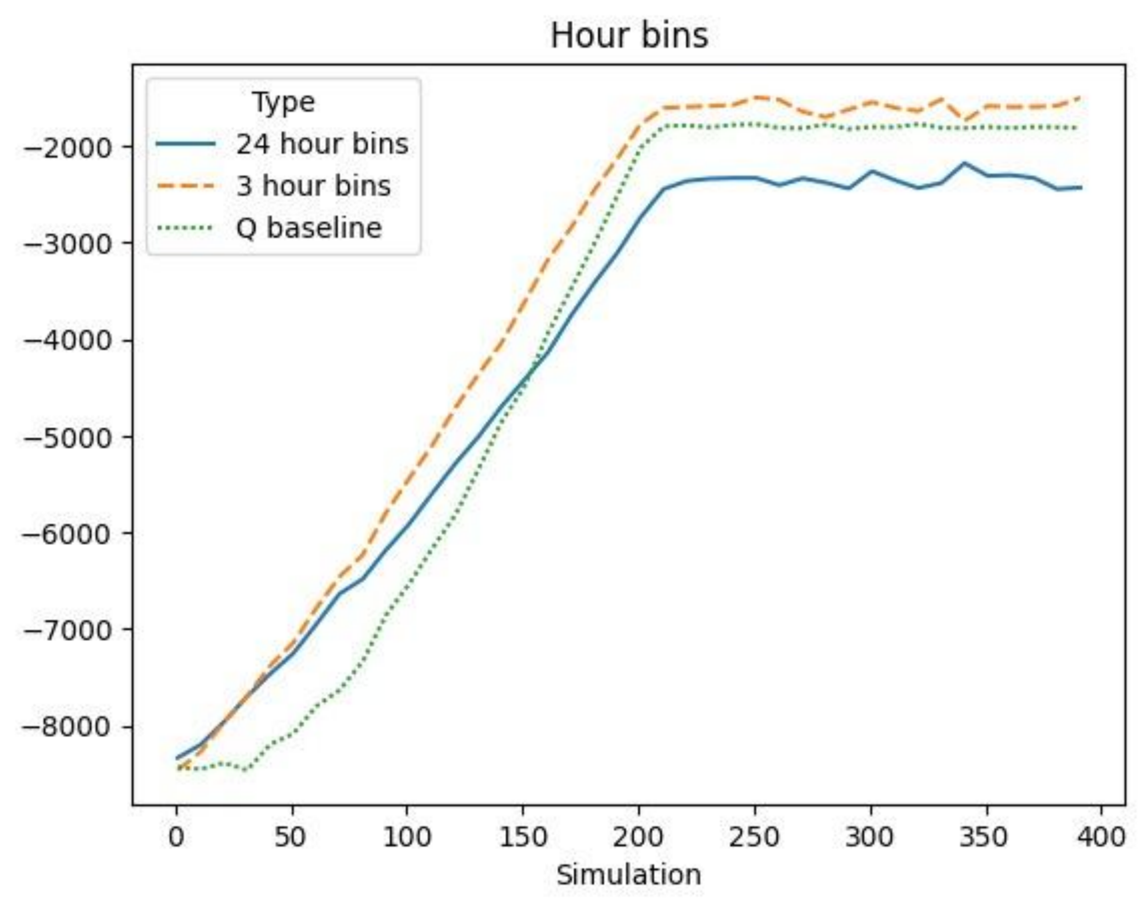
Penalty to reward for illegal/unfavourable actions

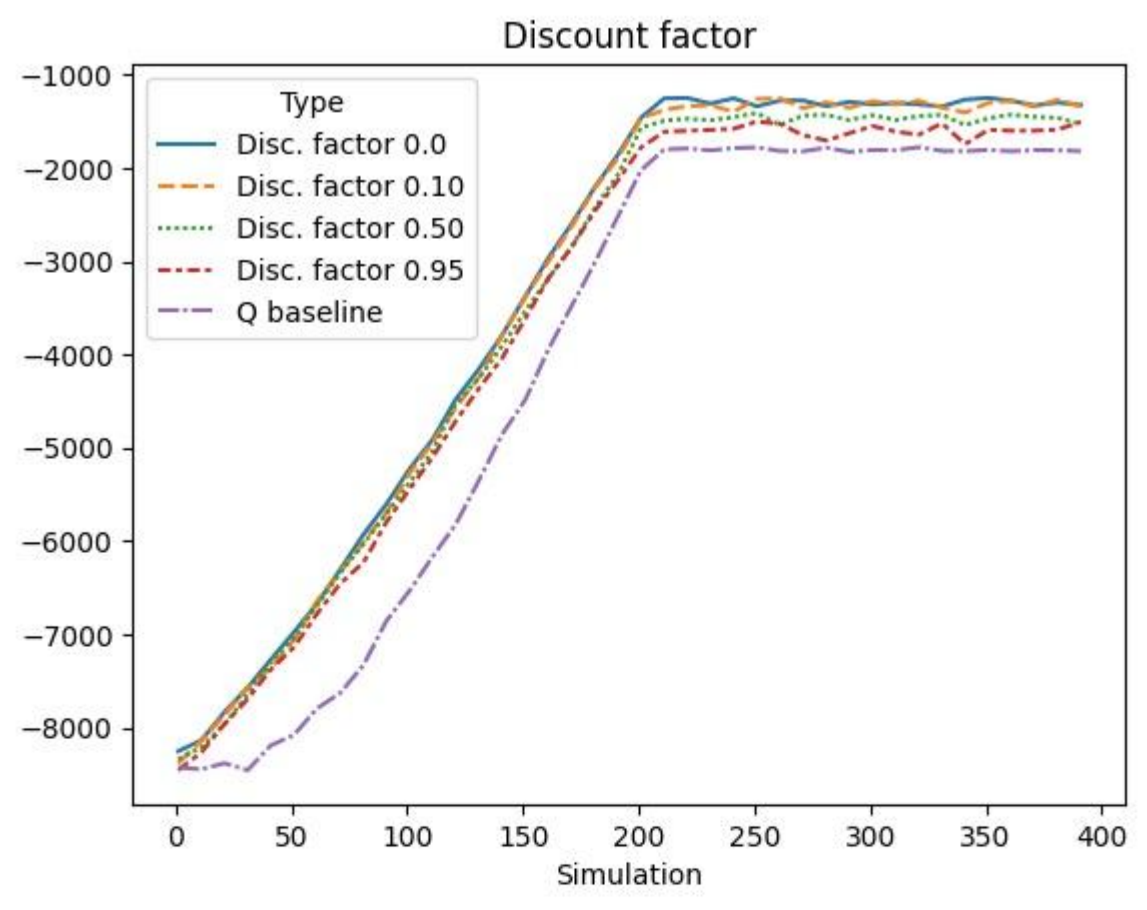


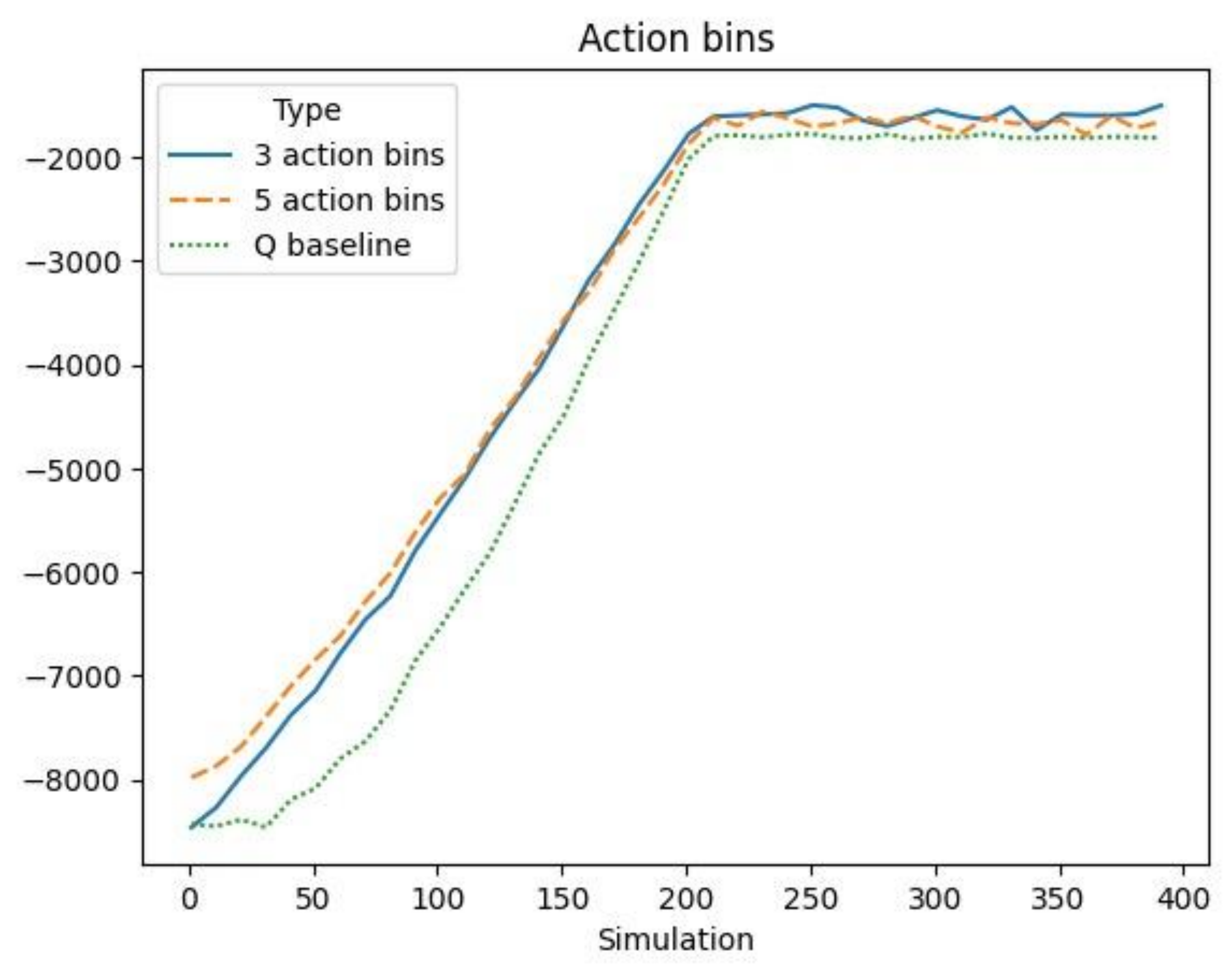
Penalties
1e6
0
−1
−2
−3
−4
−5
−6
−7
0
50
100
150
200
250
300
350
400
Simulation
Type
No penalties
Penalties
Q baseline

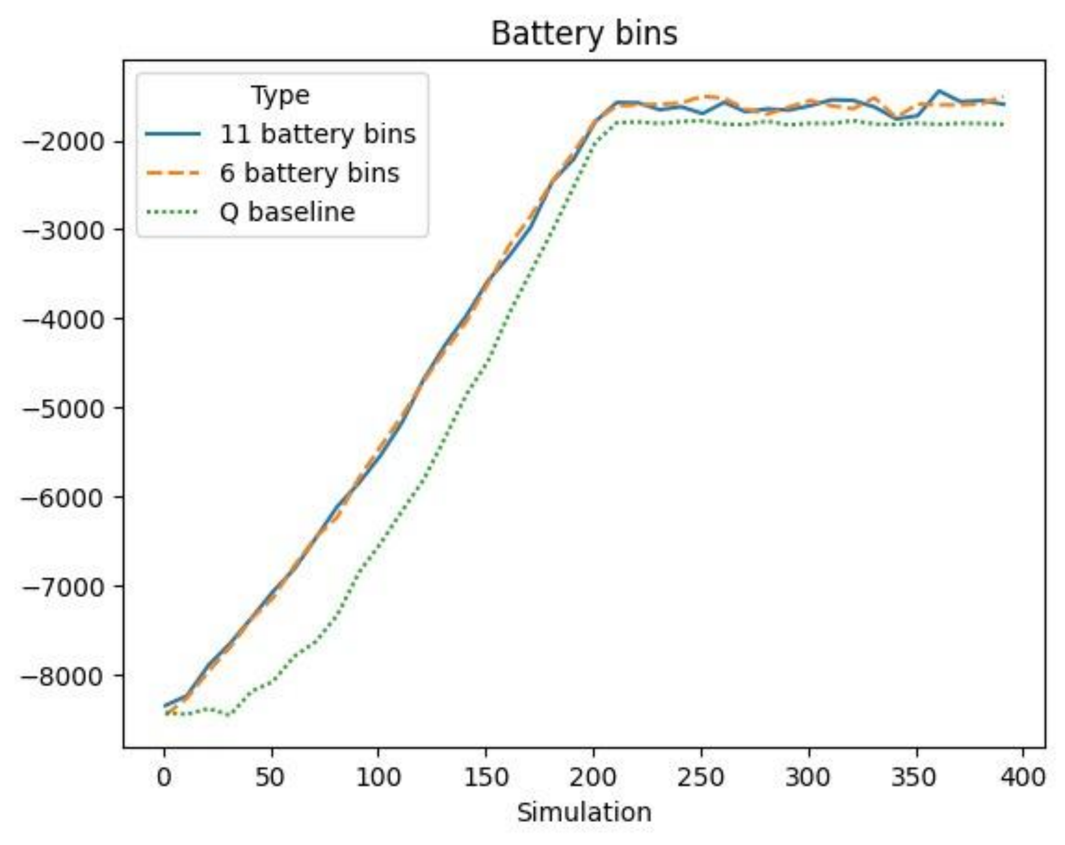

Price bins
Type
3 price bins
5 price bins
Q baseline
−1000
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation


Reward shaping
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation
Type
No reward shaping
Q baseline
Standard reward shaping
Weighted reward shaping

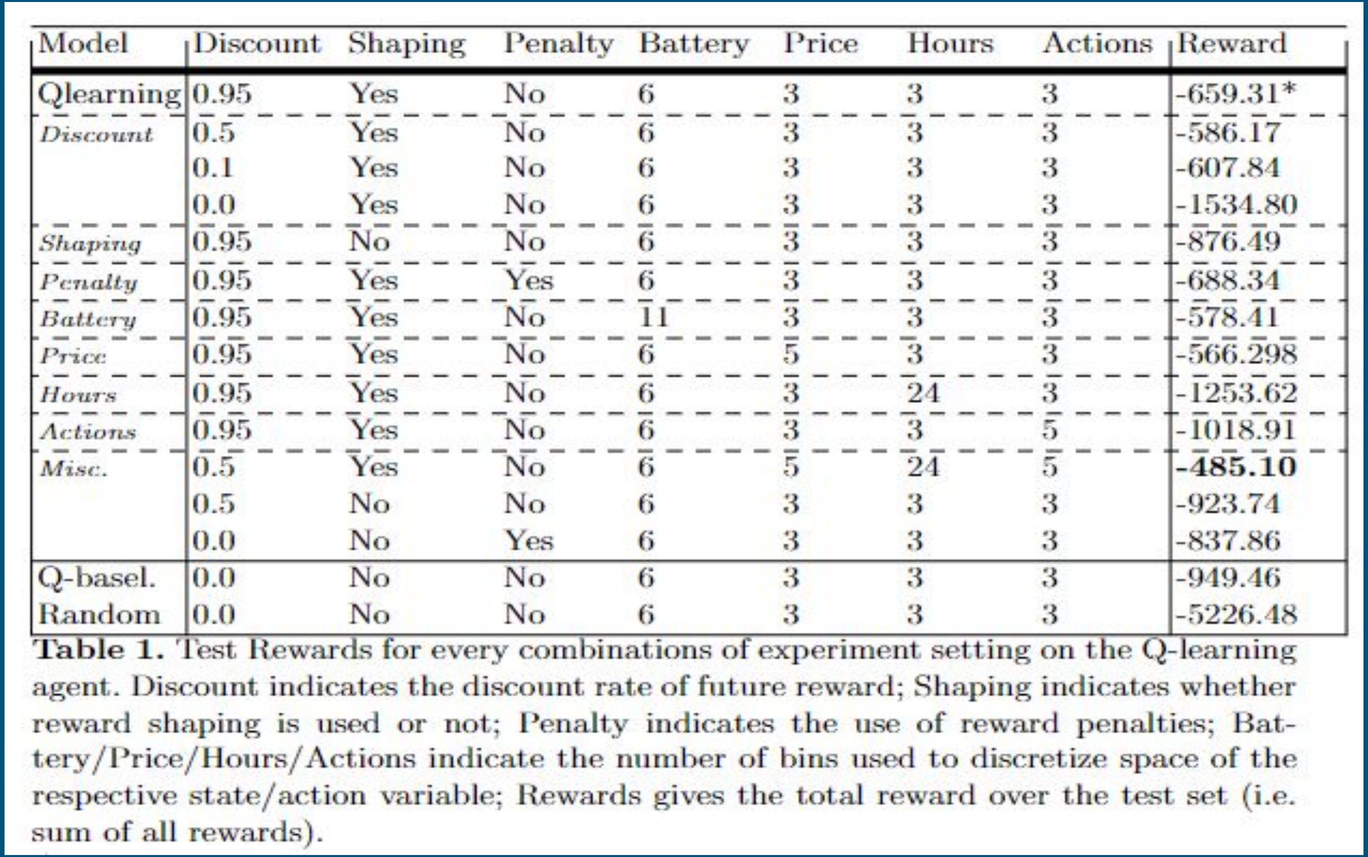

Hour bins
Type
24 hour bins
3 hour bins
Q baseline
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation


Discount factor
Type
Disc. factor 0.0
Disc. factor 0.10
Disc. factor 0.50
Disc. factor 0.95
Q baseline
−1000
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation


Action bins
Type
3 action bins
5 action bins
Q baseline
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation


Battery bins
Type
11 battery bins
6 battery bins
Q baseline
−2000
−3000
−4000
−5000
−6000
−7000
−8000
0
50
100
150
200
250
300
350
400
Simulation

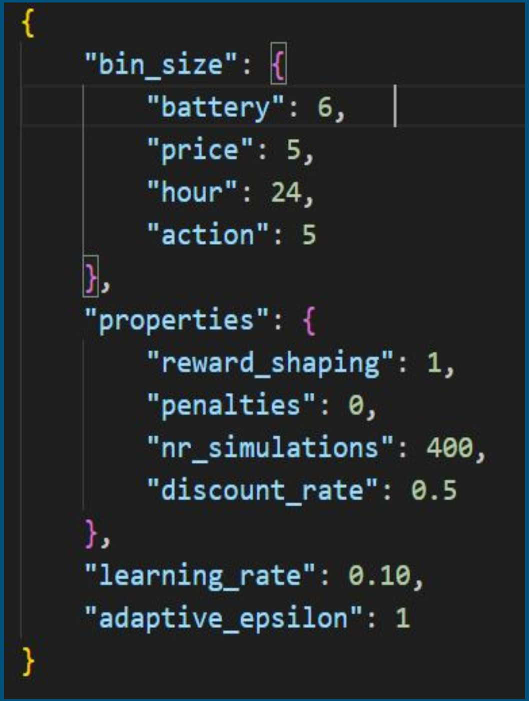
Experimental Result



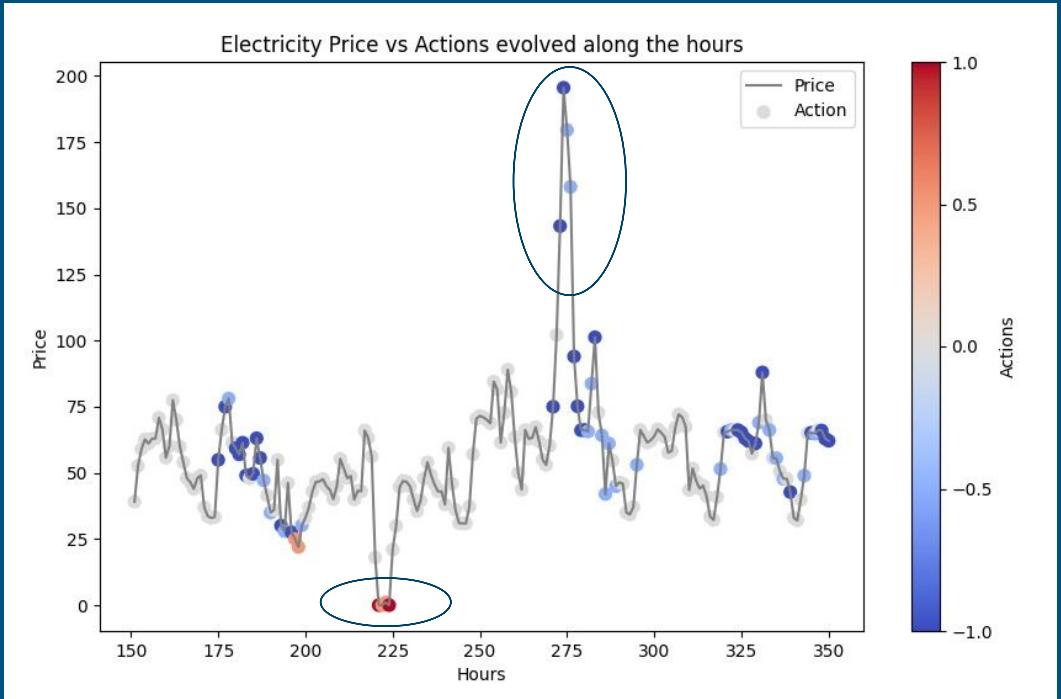
Model	Discount	Shaping	Penalty	Battery	Price	Hours	Actions	Reward
Qlearning	0.95	Yes	No	6	3	3	3	-659.31*
Discount	0.5	Yes	No	6	3	3	3	-586.17
	0.1	Yes	No	6	3	3	3	-607.84
	0.0	Yes	No	6	3	3	3	-1534.80
Shaping	0.95	No	No	6	3	3	3	-876.49
Penalty	0.95	Yes	Yes	6	3	3	3	-688.34
Battery	0.95	Yes	No	11	3	3	3	-578.41
Price	0.95	Yes	No	6	5	3	3	-566.298
Hours	0.95	Yes	No	6	3	24	3	-1253.62
Actions	0.95	Yes	No	6	3	3	5	-1018.91
Misc.	0.5	Yes	No	6	5	24	5	**-485.10**
	0.5	No	No	6	3	3	3	-923.74
	0.0	No	Yes	6	3	3	3	-837.86
Q-basel.	0.0	No	No	6	3	3	3	-949.46
Random	0.0	No	No	6	3	3	3	-5226.48

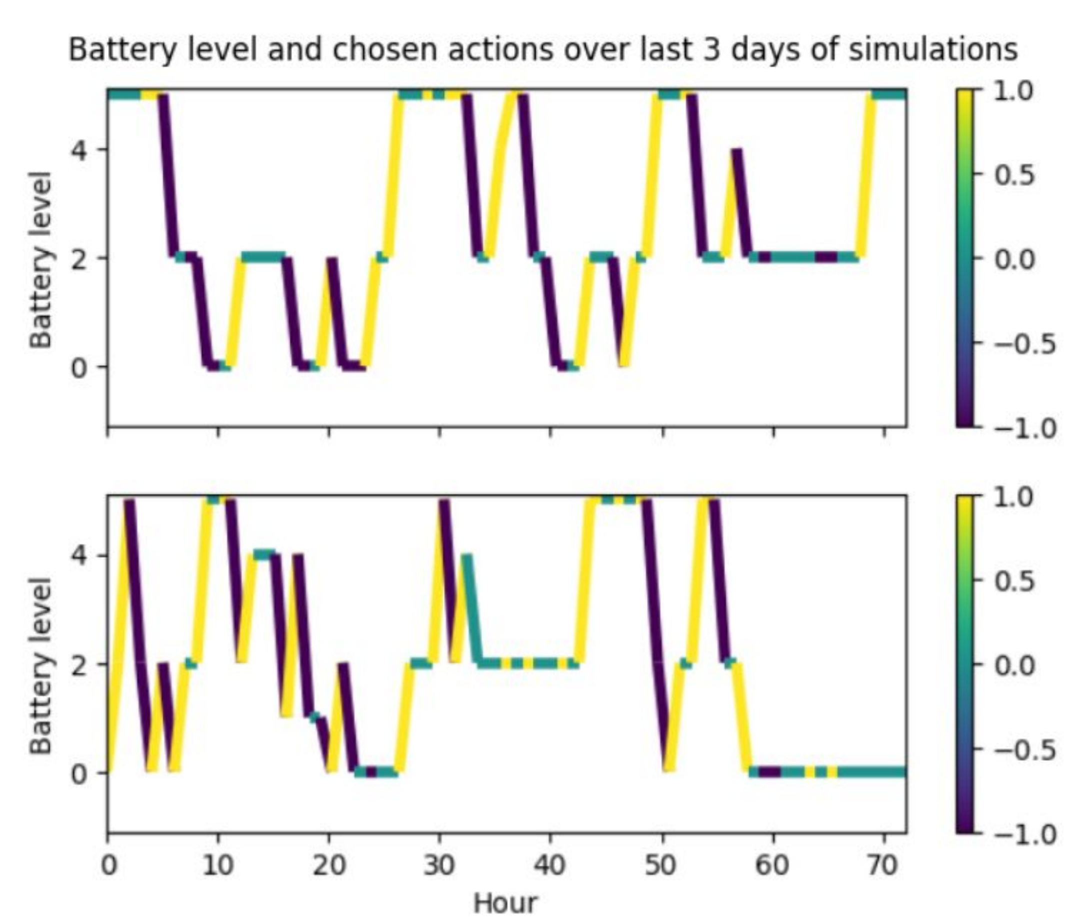
Table 1. Test Rewards for every combinations of experiment setting on the Q-learning agent. Discount indicates the discount rate of future reward; Shaping indicates whether reward shaping is used or not; Penalty indicates the use of reward penalties; Battery/Price/Hours/Actions indicate the number of bins used to discretize space of the respective state/action variable; Rewards gives the total reward over the test set (i.e. sum of all rewards).

Visualization on Test



```
{
    "bin_size": {
        "battery": 6,
        "price": 5,
        "hour": 24,
        "action": 5
    },
    "properties": {
        "reward_shaping": 1,
        "penalties": 0,
        "nr_simulations": 400,
        "discount_rate": 0.5
    },
    "learning_rate": 0.10,
    "adaptive_epsilon": 1
}
```




Battery level and chosen actions over last 3 days of simulations
Battery level
Battery level
Hour
0
10
20
30
40
50
60
70
0
2
4
1.0
0.5
0.0
−0.5
−1.0

Perspectives

Implement double DQN, policy gradient, potentially more methods

Extend research question, add extra factors

Test with different features