Offline Reinforcement Learning for Microgrid Voltage Regulation

Shan Yang , Yongli Zhu
Sun-Yat Sen University, Guangzhou, China
yangsh237@mail2.sysu.edu.cn, yzhu16@alum.utk.edu

ICLR International Conference On Learning Representations

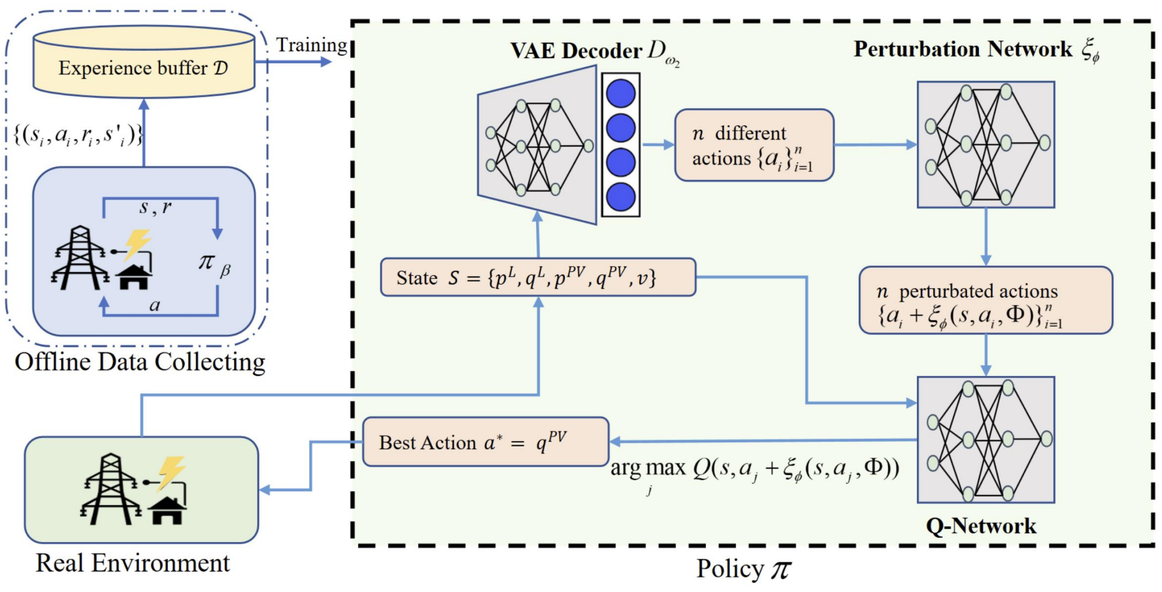
1. Introduction

Challenge: Voltage regulation is difficult under high solar penetration due to power fluctuations.

Limitations of Traditional Methods: Traditional and online RL methods rely on real-time interaction, which is **unsafe** in critical systems like microgrids.

Proposed Solution: Use Offline RL to learn voltage control policies from historical data, avoiding real-time risks.

2. Methods



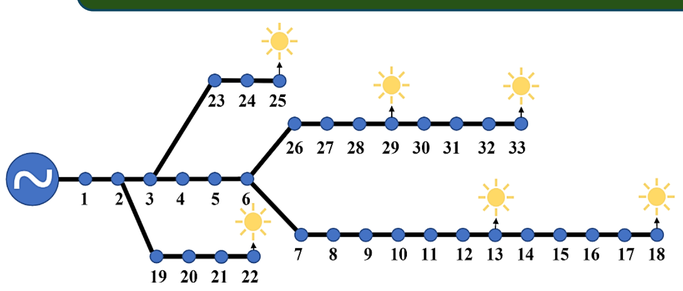
BCQ (Batch-Constrained Q-learning):

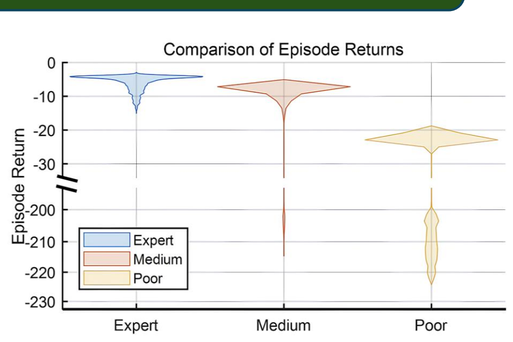
- **Restricts** actions to those in **collected dataset**
- Uses **VAE** to generate **safe actions**
- Fine-tunes with **perturbation** model
- **Reduces unsafe** out-of-distribution actions

CQL (Conservative Q-Learning):

- Adds a **conservative penalty** to the Q-value loss function
- **Penalizes overestimation** of out-of-distribution actions
- **Encourages** the policy to favor **in-distribution actions**
- Ensures **stable** performance with **low-quality experience**

3. Environment And Datasets





Environment

- Simulates a microgrid using the **distflow** model on the **IEEE 33-bus system**.
- **State:** Load, PV generation, and bus voltages.
- **Action:** Adjust reactive power of PV inverters.
- **Reward:** Penalizes voltage deviation and reactive power usage:

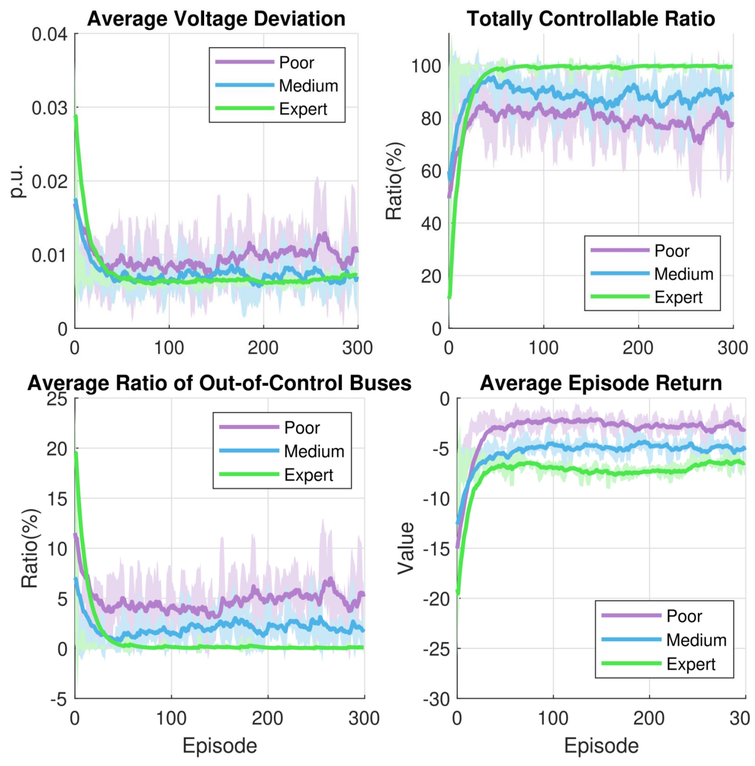
$$r = -\sum_i (\alpha_1 |v_i| + \alpha_2 |qPV_i|)$$

Performance Metric	Expert	Medium	Poor
Average Reward	-0.0242	-0.0678	-0.4325
Reward Variance	1.97E-04	6.5995	7.6837
Reward Std. Deviation	0.014	2.5691	2.227
Average Episode Return	-5.7464	-15.8355	-85.3963
Maximum Episode Return	-2.5771	-13.75	-20.349
Return Variance	4.9976	1.44E+3	7.91E+3

Dataset

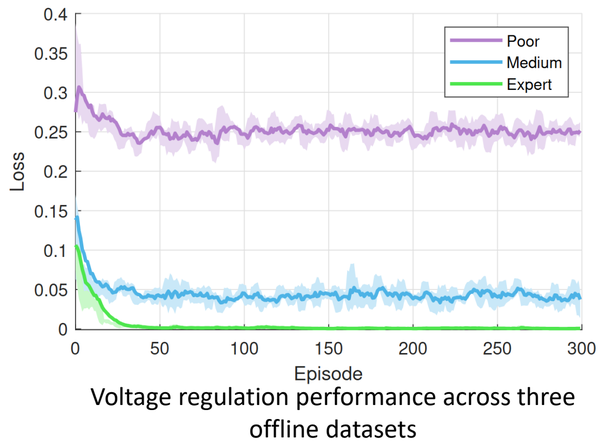
- **Expert:** High-quality data from a well-trained agent.
- **Medium:** Adds 5% noise to expert actions.
- **Poor:** Random actions with low-quality transitions.
- **Each dataset:** 200,000 transitions (s,a,r,s')

4. Experiment Results



Detailed Evaluation of BCQ's Performance

- Expert data gives best results across all metrics
- Voltage deviation: lowest in Expert, highest in Poor
- Controllable ratio drops with data quality
- Out-of-control buses more frequent in Poor
- Even Poor data allows BCQ to learn workable policies

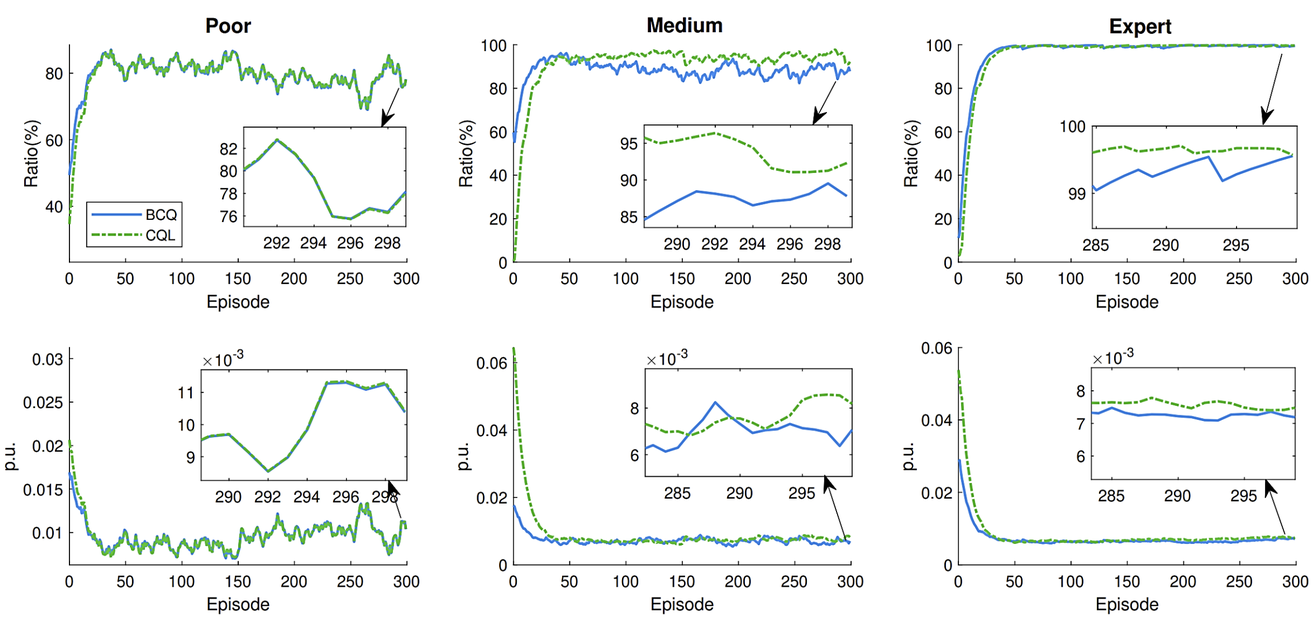


Voltage regulation performance across three offline datasets

Measures how closely generated actions match dataset

Lower = better (more realistic actions)

- Expert Lowest – clean, structured data
- Medium Moderate – some noise
- Poor Highest – noisy, inconsistent data



Performance comparison of BCQ and CQL across three datasets: Totally Controllable Ratio (top row) and Average Voltage Deviation (bottom row).

BCQ vs CQL:

- CQL performs better than BCQ on Poor & Medium datasets.
- Both work well on Expert data; CQL slightly more stable.
- CQL is more robust to low-quality data.

5. Conclusion

- **Offline RL achieves satisfiable voltage control effect with less live-interaction effort.**
- **CQL performs better on lower-quality data than BCQ.**
- **Potentially useful for other topics of power grid control.**