2024 5th International Conference on Computer Engineering and Application (ICCEA) | 979-8-3503-8677-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICCEA62105.2024.10603658

Optimizing EV Charging Station Placement in New South Wales: A Soft Actor-Critic Reinforcement Learning Approach

Jinyi Huang*
Monash University
Melbourne, Australia
geohuang2000@163.com

Xiaozhou Zhou
Jiangsu University
Zhenjiang, China
zxz03ppsw@163.com

***Abstract*—In this work, the Soft Actor-Critic (SAC) reinforcement learning algorithm is employed to investigate the optimization of new electric vehicle (EV) charging station layouts in New South Wales. This work integrates traffic density data and existing charging station locations, aiming to minimize the distance between electric vehicles and charging stations. It will ease the growing pressure on demand for EV infrastructure. Tailored reward mechanisms are designed within the SAC framework to accommodate adaptive dimensionality, demonstrating strong compatibility with urban scenarios. The model's performance offers convincing evidence of its efficacy in real-world settings. Integrating more granular data and more detailed reward mechanism design could enhance local decision-making processes, leading to more coherent overall infrastructure planning. This work contributes to advancing sustainable urban mobility by showcasing how advanced machine-learning techniques can be utilized in infrastructure planning.**

Keywords-Electric Vehicle (EV) Infrastructure, Charging Station Optimization, Reinforcement Learning (RL), Soft Actor-Critic (SAC) Algorithm, Deep Q-Networks (DQN)

I. INTRODUCTION

A. Background

The rise of electric vehicles (EVs) has necessitated a sophisticated EV charging infrastructure, catalyzing new directions in computational research. Among these, Reinforcement Learning (RL) techniques have been pivotal, particularly suited for dynamic and complex decision-making[1]. Deep Q-Networks (DQN), which integrate deep learning with Q-learning, have been notable for managing high-dimensional spaces[2]. Building on such foundations, the Soft Actor-Critic (SAC) algorithm has excelled due to its stable and efficient learning, adeptly balancing exploration and exploitation[3] and finding applications across various fields[4]. This work employs the SAC algorithm to optimize EV charging station distribution in New South Wales, tapping into RL's capabilities for managing urban planning complexities.

B. The related work

Optimizing EV charging station locations has embraced geographical data and RL. Chen et al. [5] developed a deep RL-based strategy for EV charging navigation, leveraging real-time EV operation data to optimize locations. Tuchnitz et al. [6] introduced an intelligent charging strategy for EV fleets, incorporating power grid loads and using neural networks for function approximation, highlighting reward function design for optimization. Qian et al. [7] combined smart grid and intelligent transportation systems to optimize EV charging, focusing on travel time and charging costs reductions. Shukla et al. [8] presented a data-driven method for optimizing EV charging station locations, emphasizing the efficient use of geographic data for location accuracy.

C. Our work

Most of the current layout models, focusing on traditional site selection methods, are found to only partially utilize the complexity of traffic and geographic data. It reveals two main issues to be solved: First, these RL models need help handling geographic coordinate data types. Second, existing models need help to effectively and accurately locate new charging poles based on specific circumstances.

We monitor the real-time generation of vehicles and the installation of new charging stations in our simulation. Our model optimizes charging station locations, targeting high-demand areas. For instance, in high-traffic zones, new charging poles are positioned strategically, matching the real distribution of charging needs. Our reward function, based on the proximity of vehicles to charging stations, ensures new installations address demand effectively without saturating areas already dense with facilities.

As shown in Fig1, iterative testing of our model confirms optimized station placements. Experimental results suggest the model efficiently strategizes geographical locations for new charging poles, enhancing the layout of EV charging stations across New South Wales. The base map for our analysis was sourced from Esri through ArcGIS online.

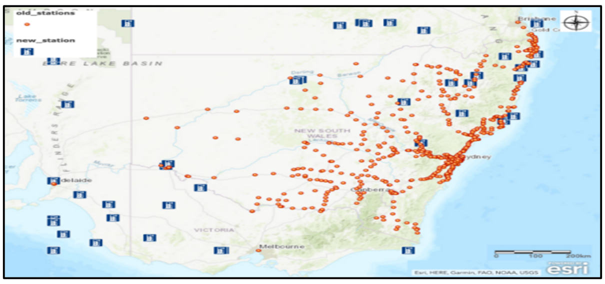

Fig. 1. Visualization of Existing and Newly generated EV Charging Stations

II. FRAMEWORK ANALYSIS AND MODEL ADAPTATION

A. Theories and Methodology

1) SAC theory

The Soft Actor-Critic (SAC) algorithm, a leading approach in reinforcement learning, is designed to optimize both the expected return and the policy's entropy, thereby maximizing cumulative rewards while promoting exploration through uncertainty[9].The calculation of policy entropy is given by:

$$\mathrm{H}(\pi(\cdot\,|s)) = -\int \pi\,(a|s)\log \pi(a|s)da \qquad (1)$$

Here, a denotes action, and s denotes states.

2) Objective Function of SAC:

SAC aims to optimize the sum of the expected return and the entropy of the policy. Its objective function can be expressed as:

$$\mathrm{J}(\pi) = \mathrm{E}_{s_t \sim D, a_t \sim \pi}[\textstyle\sum_t \gamma^t(\mathrm{R}(s_t, a_t) + \alpha \mathrm{H}(\pi(\cdot\,|s_t)))] \quad (2)$$

Here, π denotes the policy, $R(s_t,a_t)$ is the reward function, H is the entropy of the policy, γ is the discount factor, and α controls the importance of entropy.

3) Q-Value Function Update in SAC:

The Q-value function update in SAC uses the Bellman equation[10]:

$$\mathrm{Q}(s_t, a_t) = \mathrm{R}(s_t, a_t) + \gamma \mathrm{E} \qquad (3)$$

E is Expectation over the probability distribution:

$$\mathrm{E}_{s_{t+1} \sim P, a_{t+1} \sim \pi}[\mathrm{Q}(s_{t+1}, a_{t+1}) - \alpha \log \pi(a_{t+1}, s_{t+1})] \quad (4)$$

$Q(s_t,a_t)$: Q-value function, representing the expected return for taking action a_t in state s_t.

$R(s_t,a_t)$: Reward function, providing the immediate reward for taking action a_t in state s_t.

γ: Discount factor, indicating the importance of future rewards.

$s_{t+1} \sim P$: Next state s_{t+1}, following the state transition probability P.

$a_{t+1} \sim \pi$: Next action a_{t+1}, according to the policy π.

α: Coefficient balancing the importance of the entropy term in the policy. Here, P represents the state transition probability.

4) Policy Gradient in SAC:

The policy gradient is used to optimize the policy network and is defined as:

$$\nabla_\theta \mathrm{J}(\pi_\theta) = \mathrm{E}_{s \sim D, a \sim \pi_\theta}[\nabla_\theta \log \pi_\theta(a|s)(\mathrm{Q}_{\pi_\theta}(s,a) - \mathrm{V}_{\pi_\theta}(s))] \quad (5)$$

Where θ are the parameters of the policy network, and $Q^{\pi\theta}$ and $V^{\pi\theta}$ are the Q-value and value functions calculated according to policy π_θ.

5) Reward Function:

Our environment settings focus on minimizing the distance between vehicles and charging stations. We set the reward function as the negative distance:

$$\mathrm{R}(s,a) = -\mathrm{dist}(s,a) \qquad (6)$$

Where $dist(s,a)$ denotes the distance generated between a vehicle and the nearest charging station under state s and action a.

B. Integration with Custom Model for EV Charging Stations

In optimizing the Electric Vehicle (EV) charging station layout in New South Wales, the Soft Actor-Critic (SAC) model has been adapted to urban planning needs. As illustrated in Fig2, this includes integrating geographic data such as existing charging stations and high-traffic locations into state representations, customizing the policy network for geographic coordinates output, and refining the reward function to focus on proximity to high traffic areas while avoiding oversaturation.

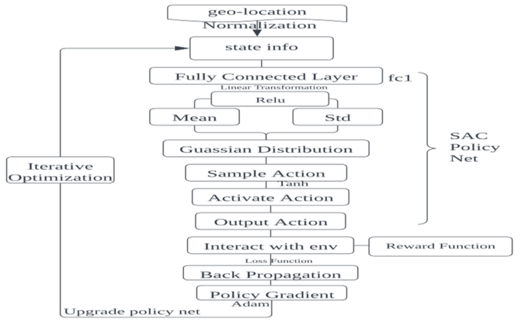

Fig. 2. Model Structure

C. The Enhanced SAC Approach: Merging Theory with Practical Application

Our enhanced SAC approach integrates theoretical frameworks into practical applications. It customizes action spaces to suggest specific coordinates for new charging stations and adjusts the reward mechanism to consider distances between proposed sites and high-traffic areas, applying penalties for clustering too many stations. The policy network uses advanced neural networks[11], including fc1 layers for linear transformations and ReLU for non-linear activations, essential for handling complex geographic data. The network is continuously updated through backpropagation, guided by a policy gradient tailored from the refined reward function.

D. Conclusion: Bridging SAC with Urban EV Infrastructure

The model is uniquely equipped to address the challenges of EV charging station placement. During simulations, the agent interacts with the environment through actions representing proposed station coordinates, with the SAC algorithm processing these actions to update the state and refine the network's policy for optimal placement. This iterative process and the precise reward function enable the model to render strategic, data-informed decisions that align with real-world scenarios, as evidenced by the empirical results.

III. DATA SETS

We adopted comprehensive datasets to analyze and optimize the Electric Vehicle (EV) charging station layout in New South Wales. We utilized a dataset of 831 high-traffic sites from Geoscience Australia's GeoNetwork, crucial for understanding traffic flow and identifying potential high-demand areas for EV charging. Another dataset from the New

South Wales Open Data Portal included 58 existing EV charging stations, providing insights into the current infrastructure and guiding the strategic placement of new stations to prevent oversaturation. Additionally, we incorporated public transport data from NSW's opendata website, crucial for assessing the broader transportation landscape and future EV charging requirements.

Our simulation environment integrates these datasets, presenting a comprehensive view of the current EV charging and traffic scenarios. It includes the positions of existing charging stations and high-traffic sites, with proposed station locations aimed at optimizing distance and accessibility.

The scatter plot (Fig3) visualizes the normalized positions of high-traffic sites and existing charging stations, establishing a geographic baseline and depicting the state of EV infrastructure and traffic concentration in New South Wales. This visualization is instrumental in identifying where additional infrastructure is necessary.

Preliminary steps included pretraining and normalization processes to ensure consistent model inputs. We applied min-max normalization to the geographical data, crucial for effective spatial data processing. The input data can be calculated using the minimum and maximum values within the dataset, with lat_min = -37.82, lat_max = -27.47, $long\_min$ = 138.59, and $long\_max$ = 153.61. To verify the model's utility, we output the coordinates of result points and then denormalize these coordinates.

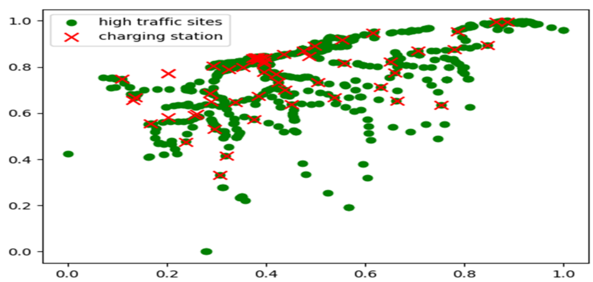

Fig. 3. Distribution of data points

To validate our data preprocessing and model compatibility, we conducted an in-depth analysis of the model dynamics by examining variations in ten state dimensions over time (Fig4). The state space dimensionality reflects the location of each proposed charging station and includes 100 dimensions—each of the 50 new charging stations represented by two spatial coordinates. This figure shows significant shifts at key time steps, reflecting likely moments for new station designations. The subplots track these dimensions, revealing decision-making patterns that balance clustered installations and individual decisions based on environmental feedback and algorithmic workings. Further analysis of decision point distributions across state dimensions (Fig5) reinforces the preprocessing strategy and model synergy. This pattern analysis offers a deeper insight into the strategic evolution of the model.

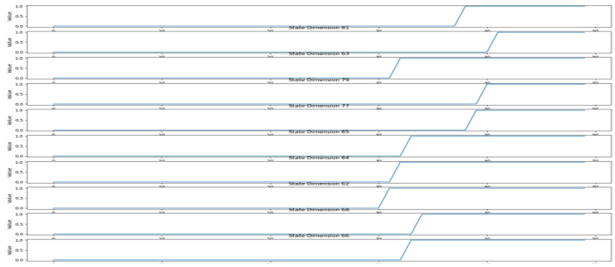

Fig. 4. Temporal Dynamics of Charging Station Positioning Across State Dimensions

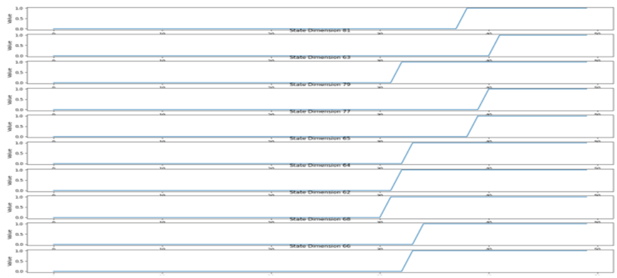

Fig. 5. Temporal Dynamics of Charging Station Positioning Across State Dimensions

IV. EXPERIMENTS AND RESULTS

A. Experimental Setting

Our model simulates the interaction between vehicle influx and charging station installations, aiming to optimize the placement of new stations in high-traffic areas to address urgent needs first. It assesses existing infrastructure and strategically plans installations where deficiencies are most severe, using a reward system that prioritizes areas with the greatest shortages. This approach enhances the utility and efficiency of the charging network across New South Wales, ensuring stations meet real-world demand effectively.

B. Experimental evaluation indicators

The Soft Actor-Critic (SAC) architecture incorporates dual critic networks to counteract overestimation bias—a prevalent issue in Q-value approximation. Each critic independently assesses Q-values for state-action pairs, providing a conservative estimate by adopting the lower of the two evaluations. This setup helps mitigate potential biases that could affect the agent's decision-making process.

To effectively implement the Soft Actor-Critic algorithm's dual-critic approach, key parameters are tuned as detailed in Table 1. These configurations above underpin the algorithm's ability to mitigate overestimation bias which is a common challenge in Q-value approximation. Key variables include dimensions for observations and actions, which represent the environmental state and possible station placements. Loss metrics for the actor and critics assess the policy network's effectiveness, while the return measures the cumulative reward received by the agent.

TABLE1. PARAMETER TUNING

Parameter Abbreviation	Meaning	Value Range	Description
actor_lr	Actor Learning Rate	1e-4	Optimization rate for the policy network.
critic_lr	Critic Learning Rate	1e-3	Optimization rate for the value network.
alpha_lr	Alpha Learning Rate	1e-4	Optimization rate for the temperature parameter.
gamma	Discount Factor	0.99	Discount rate for future rewards.
tau	Soft Update Coefficient	0.005	Update the rate for the target network parameters.
batch_size	Batch Size	64	Number of samples per batch during training.

C. *Indicator performance*

The depicted graphs showcase the loss metrics during the training process of the Soft Actor-Critic (SAC) reinforcement learning model. SAC algorithm helps making decisions in continuous action spaces. It maximize the expected value of long-term rewards while encouraging exploration through entropy in behavior by learning a policy network and two value networks[12].

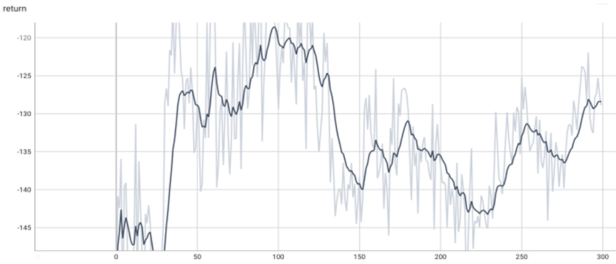

Fig6. Reward Optimization Over Training Episodes in SAC

Fig6 depicts the reward optimization process using the SAC algorithm. It shows the returns evolution, with cumulative rewards per episode on the y-axis and episode numbers on the x-axis. An initial upward trend indicates rapid learning improvement, though variability in returns highlights the model's exploratory nature and the environment's complexity. The graph's volatility signals ongoing efforts to balance exploration with exploitation, aiming for efficient EV charging station strategies. Despite fluctuations, there is evident convergence, demonstrating the SAC model's robustness in complex decision-making for urban transportation planning.

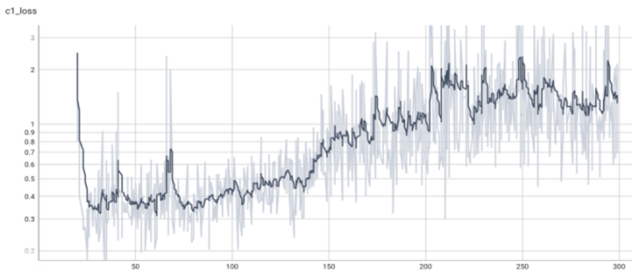

Fig7. Loss Variations of Critic 1 Network in SAC Model

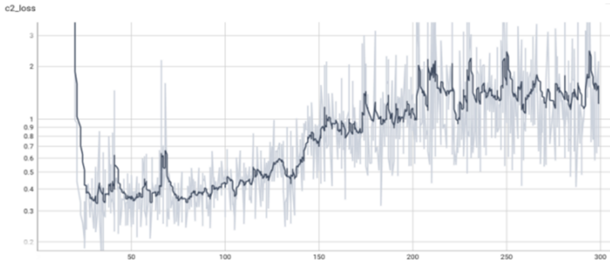

Fig8. Loss Variations of Critic 2 Network in SAC Model

Fig7 and Fig8 show the loss metrics for the two value networks, initially decreasing rapidly as the model learns to minimize vehicle distance to charging stations and later fluctuating, reflecting exploration-exploitation trade-offs. The congruence in loss trends across both networks indicates consistent value estimations.

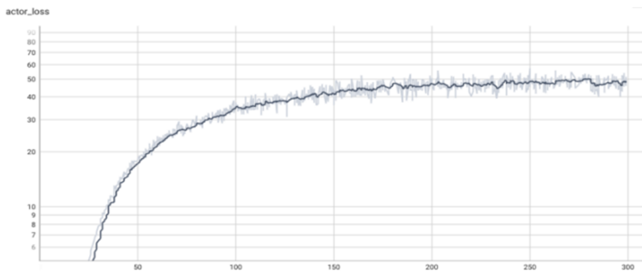

Fig9 Policy Network Loss Progression in SAC Model

Fig9 charts the policy network's loss, rising over time before stabilizing, suggesting increasing certainty in action selection while maintaining exploratory behavior to avoid local optima. This reflects the model's growing confidence in predicting beneficial actions while continuing to explore new strategies.

D. *Analysis and Explanation*

While our experiment demonstrates the applicability of reinforcement learning in the context of EV charging station placement, it concurrently underscores the necessity for enhancements. Enhancing the model's data input processes and decision-making algorithms will facilitate greater alignment with the intricate demands and realities of real-world scenarios.

Our experiment provides a mix of ideal and non-ideal outcomes that offer valuable insights for future enhancements. The model successfully discerns spatial relationships and distribution patterns, such as aligning charging points along roads with a coastal influence, which showcases its ability to grasp underlying dynamics despite data limitations. However, it falls short of accurately replicating real-world complexities, due in part to affine transformations causing unrealistic placements, and it often results in an excessive concentration of charging points in certain areas, reflecting a misalignment in decision-making that overlooks varied charging demands across regions.

For future improvements, incorporating road topology and vector data would align the model's recommendations more closely with actual roads and traffic patterns. Enhancing the reward function to factor in proximity to or distance from specific points, such as roads or high-traffic areas, would improve decision-making. Moreover, employing graph-based approaches to transform distribution scenarios into an undirected graph connected with road networks could provide a deeper understanding of spatial relationships and aid in more balanced charging station placement decisions.

V. Conclusion

This research explored optimizing EV charging station placement in New South Wales using the Soft Actor-Critic algorithm, highlighting the potential of machine learning to address complex urban planning challenges like EV infrastructure distribution. The model showed promise in identifying patterns and adapting to traffic density trends but also revealed gaps in fully reflecting real-world scenarios, primarily due to data representation and model processing limitations. Future work should focus on integrating more comprehensive and context-specific data, such as detailed road maps and traffic patterns, and refining the reward mechanism to incorporate geographical and infrastructural variables for more balanced and realistic outcomes.

References

[1] Wang, H., Liu, N., Zhang, Y., Feng, D., Huang, F., Li, D., & Zhang, Y. (2020). Deep reinforcement learning: a survey. Frontiers of Information Technology & Electronic Engineering, 21, 1726-1744.

[2] Mnih, V., Kavukcuoglu, K., Silver, D., et al. (2015). Human-level control through deep reinforcement learning. Nature, 518, 529-533.

[3] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. Proceedings of the 35th International Conference on Machine Learning, 80, 1861-1870.

[4] Lillicrap, T. P., Hunt, J. J., Pritzel, A., et al. (2015). Continuous control with deep reinforcement learning. CoRR, abs/1509.02971.

[5] Chen, L., Yang, F., Wu, S., & Xing, Q. (2021). Electric Vehicle Charging Navigation Strategy Based on Data Driven and Deep Reinforcement Learning. Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence.

[6] Tuchnitz, F., Ebell, N., Schlund, J., & Pruckner, M. (2021). Development and Evaluation of a Smart Charging Strategy for an Electric Vehicle Fleet Based on Reinforcement Learning. Applied Energy, 285, 116382.

[7] Qian, T., Shao, C., Wang, X., & Shahidehpour, M. (2019). Deep Reinforcement Learning for EV Charging Navigation by Coordinating Smart Grid and Intelligent Transportation System. IEEE Transactions on Smart Grid, 11(4), 1714-1723.

[8] Shukla, M., Singh, D., Yadav, A., Singh, A., Jumaa, N. A., & Mohammed, A. (2023). Optimizing Electric Vehicle Charging Station Placement in Urban Areas: A Data-Driven Approach. In 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2835-2838.

[9] Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction.

[10] Haarnoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic Algorithms and Applications. arXiv preprint arXiv:1812.05905.

[11] Subramanian, A., Chitlangia, S., & Baths, V. (2020). Reinforcement learning and its connections with neuroscience and psychology. Neural networks: the official journal of the International Neural Network Society, 145, 271-287.

[12] Vidyasagar, M. (2023). A tutorial introduction to reinforcement learning. SICE Journal of Control, Measurement, and System Integration,16,172-191