マルチエージェント強化学習によるEV群VPPの充放電協調制御

林 弘辰 * (筑波大学)

Coordinated Charging and Discharging Control of an EV-Based VPP Using Multi-Agent Reinforcement Learning

Koshin Hayashi (University of Tsukuba)

1 まえがき

VPP の需給調整市場参入では、リソースアグリゲーターが系統運用者から与えられる指令信号に対して、リアルタイムかつ低遅延でリソースを制御し追従する能力が求められる場合がある。一方、EV を多数束ねた VPP を集中最適化で制御する場合、個別 EV 状態の収集・通信およびオンライン計算がボトルネックとなり、応答レイテンシ制約下では運用が困難になり得る。

例えば情報をアグリゲータが集約し、Linear Programming (LP) により各 EV の行動を決定する VPP スケジューリングは日次計画等で広く検討されているが (1)、中央制御が必要なため通信遅延、計算遅延により市場の要求する応答時間内のデマンドレスポンスが間に合わない場合がある。一方、Multi-Agent Deep Deterministic Policy Gradient (MADDPG) に代表される MARL（Multi-Agent Reinforcement Learning）は CTDE（Centralized Training and Decentralized Execution）の性質を持ち、実行時に頻繁な全体通信や大規模最適化を要さずに協調方策を実装できる (2)。EV 群制御に MARL を適用した先行研究として、ユーザ側 QoS（充電費用や SoC 等）に重点を置くもの (3) や、系統側の指標に重点を置くもの (4) があるが、双方を同一の MARL 枠組みで明示的に最適化・評価する設計は限定的である。

そこで本研究では、充電ステーションをエージェントとする階層型 MADDPG 制御を設計し、指令追従（系統側）とユーザ制約（出発時 SoC）の同時最適化を、分散実行可能な形で実現することを目的とする。さらに、オープンソースデータを用い、合計 50 台の普通充電器からなる VPP が日本の需給調整市場（二次調整力相当）へ参入する状況を想定してケーススタディを行う。

2 提案手法

<2·1> 制御モデルの全体像 本研究では、VPP を構成する複数の充電ステーションをエージェントとし、各エージェントが当該ステーション配下の複数 EV の充放電出力を決定する階層型の構造を採る。実行時には、各ステーションは（i）自ステーションに接続中 EV の状態（SoC、残滞在時間、目標 SoC 等）と（ii）市場から与えられる指令信号のみを用いて行動し、ステーション間の高頻度な個別通信を前提としない。本構成の概念図を図 1 に示す。

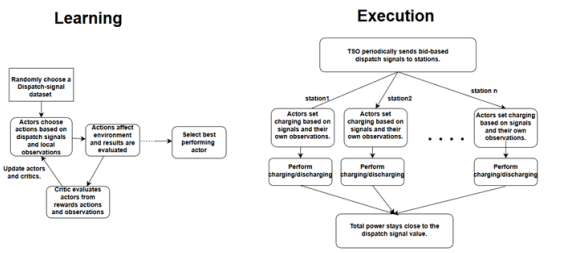

図 1 EV 充放電制御モデルの全体像

Fig. 1. Overview of the EV charging/discharging control model

<2·2> 階層型 MADDPG の実装 本モデルは MADDPG (2) を基礎とし、学習時は集中情報で価値を評価しつつ、実行時は各エージェントが局所観測のみで行動する。基本的には MADDPG で提唱されている決定論的方策勾配と 1 ステップ TD ターゲット（target critic を用いたブートストラップ）に基づき、actor および critic を更新する。ただし本研究では全体目標と個別目標を分離して扱うため、局所目的（出発時 SoC）を評価するローカル critic と、全体協調（指令追従）を評価する共有グローバル critic を併用し、actor を更新する。

ローカル critic はステーション i に関する局所情報（当該ステーションの観測および局所的に意味を持つ状態・行動）を入力として、出発時 SoC 等のユーザ制約に関わる価値を推定する。一方、グローバル critic は全ステーションの情報（全体状態および全エージェント行動）を入力として、VPP 集約としての指令追従に関わる価値を推定する。actor の更新では、MADDPG の方策勾配をローカル／グローバルの各 critic に対して評価し、両者を重み w で凸結合した方向にパラメータ更新を行う。この重み付き勾配合成の考え方は、複数の価値関数からの信号を統合する DE-MADDPG 型の拡張 (5) を参考にしており、指令追従とユーザ制約のトレードオフを学習に組み込むために用いる。

さらに、critic ネットワークには目的志向の attention 機構（goal-conditioned attention）を導入し、指令追従や出発時 SoC の達成に寄与する状態・行動成分に焦点を当てる。attention 自体は Transformer で用いられる scaled dot-product attention (6)

をベースとしつつ、本研究では目的（指令追従／ユーザ制約）に応じて重み付けの特徴が変化するよう設計している．

3　ケーススタディ

本研究は、日本の需給調整市場における二次調整力相当の参加を想定し (7,8)、指令信号が 5 分ごとに到来する離散時間環境（1 ステップ=5 分）で評価する。1 日を 288 ステップとして 1 エピソードを構成し、VPP は 5 ステーション、各 10 台の普通充電器（合計 50 台）からなるものとする。EV の到着分布および充電履歴等のデータには公開データセットを用いる。

本ケーススタディでは、性能指標として（i）Target SoC satisfaction rate、（ii）charging dispatch tracking rate、（iii）discharging dispatch tracking rate を用いる。ここで、Target SoC satisfaction rate は、各 EV の出発時刻において SoC がユーザの設定した目標値以上であった出発事象の割合として定義する。また tracking rate は、全 288 ステップの内、VPP 集約出力が指令値の許容帯（本研究では入札幅 ×0.1 の許容幅）内に収まったステップの割合として算出する．

4　結果

学習の進行に伴う性能指標の推移を図 2 に示す。学習の初期段階から指令追従性能が改善し、収束近傍では安定した追従が観察される。収束付近の方策で評価した結果、Target SoC satisfaction rate は 63.5 %、charging dispatch tracking rate は 99.5 %、discharging dispatch tracking rate は 100.0 % であった．

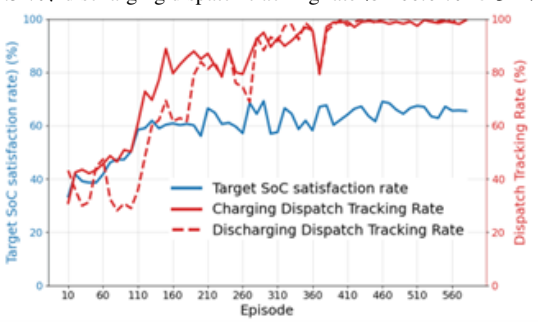

図 2 学習に伴う性能指標の推移
Fig. 2. Performance metrics over training episodes

代表的なテストエピソードにおけるステーション出力を図 3 に示す。ステーション単位の分散実行でありながら、紫の破線により表現される集約出力が、黒の実線で示される系統指令値に追従する協調挙動が得られていることが確認できる．

あわせて、EV の充放電器滞在時の SoC 曲線を確認したところ、出発ステップ前に目標 SoC を達成している EV が積極的に放電指令に協力したり、逆に目標 SoC に達していない EV が優先的に充電するなど、ユーザ制約を考慮した柔軟な挙動が観察された。しかし、目標達成が困難な EV に対しては充電を放棄する、余裕がある EV に対しても過剰に放電してしまうなどの課題も確認された。

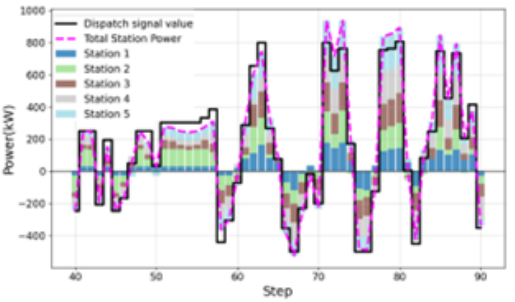

図 3 代表テストにおけるステーション電力
Fig. 3. Station power in a representative test episode

5　まとめ

本稿では、需給調整市場における EV–VPP の低遅延制御を目的とし、階層型 MADDPG 制御を示した。ローカル critic（ユーザ制約）と共有グローバル critic（指令追従）を併用し、両者の評価に基づく方策更新を重み付きに統合することで、分散実行のまま二目的を同時に扱う枠組みを与えた。日本市場（二次調整力相当）を想定した 50 台規模のケーススタディにより、高い追従性能（99.5%/100.0%）と一定のユーザ満足（63.5%）を確認し、集中制御がレイテンシ制約で困難となる状況における実装可能な代替案としての有効性を示した。しかし実用にはまだ距離があるため、ハイパーパラメータの最適化やルールベース制御との併用によるユーザ満足度指数の向上を今後の課題とする．

文献

(1) A. Shayegan-Rad, A. Badri, and A. Zangeneh. Day-ahead scheduling of virtual power plant in joint energy and regulation reserve markets under uncertainties. *Energy*, 121:114–125, February 2017.

(2) Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *CoRR*, abs/1706.02275, 2017.

(3) K. Park and I. Moon. Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid. *Applied Energy*, 328:120111, December 2022.

(4) P. Fan, J. Hu, S. Ke, Y. Wen, S. Yang, and J. Yang. A frequency–pressure cooperative control strategy of multi-microgrid with an electric–gas system based on MADDPG. *Sustainability*, 14(14):8886, July 2022.

(5) Hassam Ullah Sheikh and Ladislau Bölöni. Multi-agent reinforcement learning for problems with combined individual and team reward. *IEEE Transactions on Games*, 2020. DE-MADDPG algorithm.

(6) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems*, 30:5998–6008, 2017.

(7) Japan Electric Power Exchange (JEPX). Supply and Demand Adjustment Market Product Guide Ver.5, May 2025. Tokyo, Japan.

(8) Ministry of Economy, Trade and Industry (METI). Second interim report of the 30th meeting of the Working Group on System Review, Electricity and Gas Basic Policy Subcommittee [11.1], July 2019. Tokyo, Japan.