

Vehicle-to-Grid (V2G)를 위한 알고리즘 및 인공지능 모델 개발



202055577 이선진

202055582 이지은

황원주

목 차

1. 서론	1
1.1. 연구 배경	1
1.2. 기존 문제점	2
1.3. 연구 목표	3
2. 연구 배경	4
2.1. EV2Gym	4
2.2. RL 알고리즘	4
3. 연구 내용	6
3.1. 선행 연구	6
3.1.1. PST(Power Setpoint Tracking) Problem 평가	6
3.1.2. V2G Profit Maximization 평가	9
3.1.3. Battery Degradation 관점 평가	10
3.2. 이해관계자에게 유리한 V2G 충방전 보상/상태함수 알고리즘 설계	12
3.2.1. 전력망 운영자	13
3.2.2. 전기차 사용자	14
3.2.3. 충전 서비스 사업자	16
3.3. 이해관계 최적화를 위한 통합된 V2G 충방전 보상/상태함수 알고리즘 설계	18
3.3.1. 상태함수	19
3.3.2. 보상함수	20
3.3.3. 결과	25
3.4. 전기차 충전 요금을 고려한 충방전 알고리즘 설계	29
3.4.1. step 알고리즘	29
3.4.2. 결과 분석	31
4. 연구 결과 분석 및 평가	33
5. 결론 및 향후 연구 방향	34
6. 참고 문헌	35

1. 서론

1.1. 연구 배경

V2G(Vehicle to Grid)는 자동차에서 전력망으로 전기를 이동하는 것을 의미하는데, 전기차에 저장한 배터리를 에너지저장장치(Energy Storage System, ESS)로 활용하여 전력계통에 연계하는 기술을 뜻한다. 전기 자동차는 엔진 대신 배터리에 있는 전력으로 모터를 구동한다. 전기차 운전자는 전기요금이 저렴한 심야 시간대에 배터리를 충전하고, 출근 이후에는 배터리에 남아있는 전력을 되판다. 이러한 거래 행위를 통해 피크 시간대에는 전력회사에 전력을 팔아 고객은 돈을 벌고, 전력회사는 발전소 가동률을 줄여 수요관리를 하는 양방향 기술이다. 전기차의 전력망으로부터 전력을 공급받아 움직일 수 있는 관점에서 전기차는 에어컨이나 냉장고와 똑같이 전력을 소비하는 장치이나, 배터리에 저장 되어있는 에너지를 다시 전력망으로 되돌린다는 점에서 전기차는 하나의 발전기 역할을 하는 것으로 볼 수 있다.

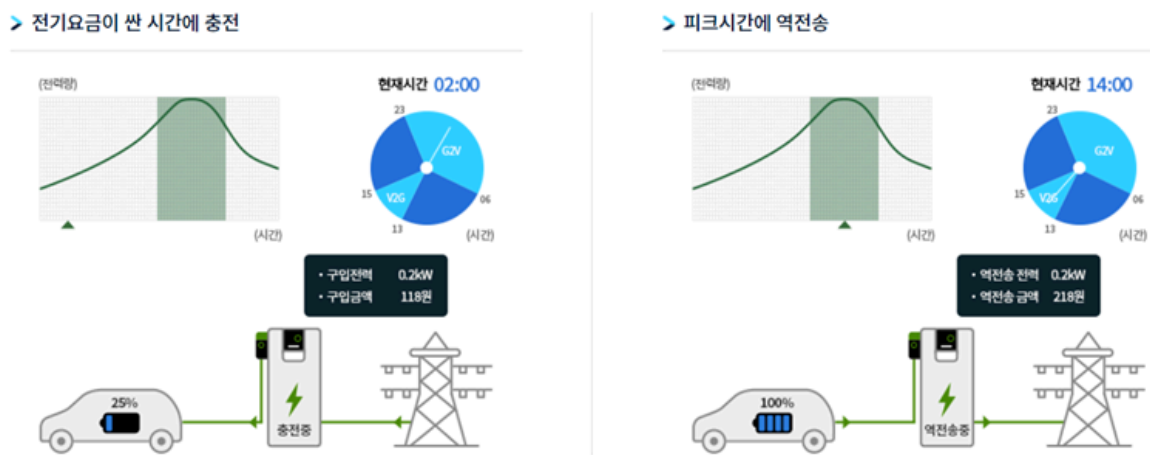


그림 1. V2G 시스템의 기본 개념

국가적 관점에서 [그림 1]과 같이 전력 소모가 낮은 밤 시간에는 전기차를 충전하여 에너지를 저장해 두었다가 전력 수요가 급증하는 낮 시간에 에너지를 전력망으로 되돌림으로써 V2G 기술을 활용할 수 있다.

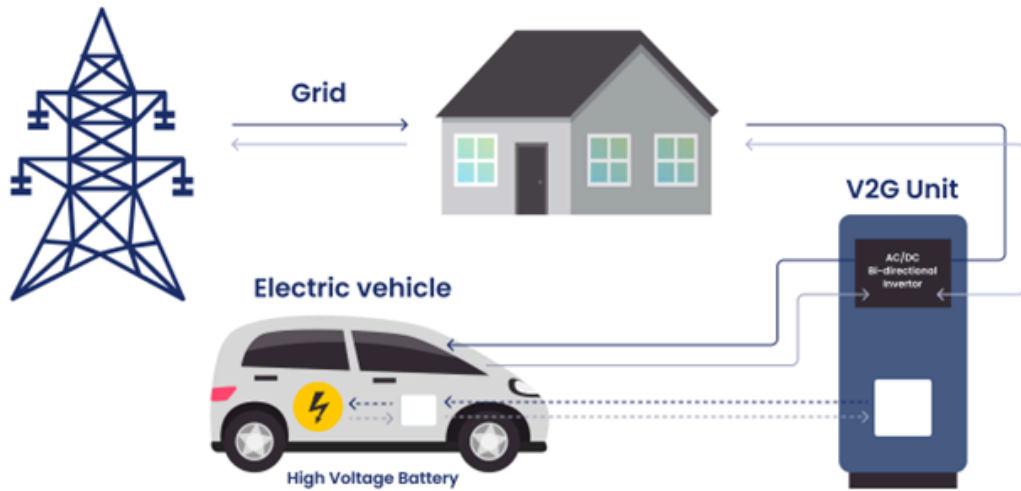


그림 2. V2G 시스템 구성도

현재 V2G 기술은 이산화탄소의 배출을 줄여 친환경적이며, 경제성을 갖추어 지속 가능한 성장을 이어갈 수 있는 산업 모델로 주목받고 있다. [1] V2G를 지원하는 전기차의 보급이 증가하면서, 전력망의 효율성에 기여할 수 있는 잠재력 또한 주목받고 있다. [2] 전기차가 충전 및 방전하는 시간대를 조절함으로써 피크 수요를 완화하고, 전력망의 안정성을 높여 다양한 전력 보조 서비스와 부가가치 창출이 가능할 것으로 기대된다. 다만, 전기차 배터리는 충방전 주기에 따라 성능이 저하되기 때문에, 충전 및 방전 과정에서 배터리의 수명을 최대한 유지하면서도 효율적인 에너지 관리를 수행하는 알고리즘이 필요하다. 또한 전기차 소유자들의 V2G 시스템 참여를 유도를 위해 배터리 사용의 효율성을 극대화하면서 사용자에게 경제적인 혜택을 제공할 수 있는 알고리즘이 필요하다.

1.2. 기존 문제점

전기차(Electric Vehicle, EV) 시장이 급속도 성장하고 배터리 용량이 증가함에 따라, 사용자의 다양한 요구를 충족하며 높은 신뢰도로 안전하게 충전 가능한 인프라가 필수적이다. 전기차의 보급이 증가하며, 특히 피크 시간대의 더 많은 전기 에너지가 필요하며, 충전소의 가용성 및 사용자의 충전 요구 사항(시작 시간 및 희망 SoC(State of

Charge))이 중요해지고 있다.

EV 소유자가 차량을 충전 포트에 연결하자마자 배터리 충전을 시작하는 단순 충전 방식은 하루 전기 비용이 동일하다고 가정한다. 이와 달리 이중 가격 충전 방식은 사용자가 전기 요금에 따라 차량을 충전할 수 있도록 하여 피크 시간대 수요를 제어하는 것을 목표로 한다. 그러나 비피크 시간대에 모든 차량이 거의 동시에 충전되기 시작하여 전력 수요가 급증된다는 단점이 있다.

이러한 전기차의 충방전을 최적화하기 위해 효율적이고 경제적인 방식으로 충전 전력을 관리하는 스마트 충전소 출현이 요구된다. 스마트 충전소로 그리드에 연결된 EV가 비이용 시간대에는 전력망에 에너지를 제공하고 운행 전에 재충전이 가능하다. 그러나 사용자의 충전 요구 사항을 준수하며 보조 서비스 수요를 충족하는데 필요한 에너지 제공이 가능해야 하며, 해당 방식은 EV 배터리의 충방전 사이클로 인한 배터리 열화를 가속화할 수 있다는 우려가 있다. 고출력 오프보드 충전기 장비는 배터리를 80%의 SoC로 단시간 내에 충전한다는 장점이 있지만, 이러한 장비의 대량 사용은 전력망에 부정적인 영향을 미칠 수 있다. 따라서 전력망의 부담을 완화하기 위해 분산 에너지 저장 시스템의 이점을 활용하는 V2G 기술 구현이 필요하다. [3]

1.3. 연구 목표

논문 및 시뮬레이션 연구를 통한 V2G 시스템에 대한 이해를 바탕으로 RL(Reinforcement Learning)을 이용한 두 가지 알고리즘 개발을 목표로 한다. 'V2G 이해관계자별 이익을 고려한 상태 및 보상 함수 설계'를 통해 전력망 운영자, 충전 서비스 사업자, 사용자 3가지의 V2G 이해관계자 입장에서 각자의 이익을 최대화하는 알고리즘을 설계한다. 이를 통해 각 이해관계자가 V2G 시스템을 보다 효과적으로 활용하여, 시스템의 효율성을 높인다. 'V2G 이해관계 최적화를 위한 충방전 알고리즘 통합 상태 및 보상 함수 설계'를 통해 각 이해관계자의 입장을 통합함으로써 상호 이익을 극대화하는 최적의 운영 모델을 구축하고자 한다. 해당 알고리즘을 통해 전기차의 활용을 극대화하며, 전기차 소유자와 전력망 운영자, 충전 서비스 사업자 모두에게 경제적, 환경적 이점을 제공할 수 있다. 추가적으로 '전기차 충전 요금을 고려한 충방전 알고리즘 설계'를 통해 전기차 사용자에게 가장 높은 이익을 제공하는 알고리즘을 찾고 이를 시각화할 수 있는 자료를 제공할 것이다.

2. 연구 배경

2.1. EV2Gym

OVERVIEW OF EXISTING EV SIMULATORS FOCUSING ON SMART CHARGING STRATEGIES.

Simulator Name	V2G	Power Network Impact	EV Models	EV Behavior	Charging Stations	Available Baseline Algorithms	RL Ready	Programming Language	Comments
V2G-Sim [12]	Yes	Partial	Diverse	Real Charging Transaction Probability Distributions	Uniform	- Heuristics, Mathematical Programming	No	Not Open-source	- Customizable V2G simulations.
EVLlibSim [13]	Yes	No	Diverse	Randomized	Uniform	- Heuristics, Mathematical Programming	No	Java	- Easily customizable simulations with a visual interface.
EV-EcoSim [14]	Yes	Complete	Uniform	Randomized	Uniform	- Mathematical Programming	No	Python	- Grid-Impact analysis
evsim [15]	No	Partial	Diverse	Real Charging Transaction Probability Distributions	Uniform	- Heuristics, Mathematical Programming	No	R	- Simulate and analyze the charging behavior of EV users.
OPEN [16]	Yes	Complete	Uniform	Randomized	Uniform	- Heuristics, Mathematical Programming	No	Python	- Modelling, control & simulations for smart local energy systems.
ACN-Sim [17]	No	Complete	Uniform	Real Charging Transactions	Uniform	- Heuristics, MPC, RL, Mathematical Programming	Yes	Python	- Designing a complete simulator framework.
SustainGym [18]	No	Complete	Uniform	Real Charging Transactions	Uniform	- RL	Yes	Python	- Providing a benchmark for sustainable RL applications.
Chargym [19]	Yes	Very Limited	Uniform	Randomized	Uniform	- Heuristics, RL	Yes	Python	- Comparing RL algorithms for smart charging.
EV2Gym (Ours)	Yes	Partial	Diverse	Real Charging Transaction Probability Distributions	Diverse	- Heuristics, MPC, RL, Mathematical Programming	Yes	Python	- Comprehensive simulator for any control algorithm.

그림 3. 기존 EV Simulator 비교

본 연구는 EV 충전 시나리오를 시뮬레이션하기 위한 프레임워크인 EV2Gym 환경[4][5] 을 기반으로 EV 충전 전략을 실험하였다. EV2Gym은 여러 제어 알고리즘을 적용할 수 있는 시뮬레이터로, 기존의 EV 시뮬레이터들에 비해 유연성과 확장성이 우수하다. 특히 RL 알고리즘을 적용한 실험이 가능해, 다양한 요소를 고려한 EV 충전 전략 최적화에 적합한 실험 환경을 제공한다.

또한 EV2Gym은 전력 설정값 추적, 이익 극대화, 태양광 활용, 수요 반응 이벤트 등을 포함하여 실제 환경과 유사한 충전 시나리오를 효과적으로 재현할 수 있다. 사용자 정의가 가능한 시뮬레이션을 통해 가격, 부하 및 전력 변압기 수준 등 다양한 요소를 직접 설정 가능해 특정 조건에 맞춰 시뮬레이션을 조정하고, EV 충전 전력 최적화에 도움을 준다.

2.2. RL 알고리즘

다양한 RL 알고리즘의 적용이 용이한 EV2Gym 환경에서 다음과 같은 9가지 알고리즘을 중심으로 연구에 사용하였다.

(1) A2C (Advantage Actor-Critic)

정책 기반과 가치 기반 방법을 결합하여, 안정적이고 빠른 학습이 가능하다. 연속적인 행동 공간이 필요한 V2G 충전 및 방전 전략 최적화에 적합한 알고리즘이다.

(2) ARS (Augmented Random Search)

간단한 구조로 신속한 탐색이 가능해 초기 정책을 빠르게 계산할 수 있다는 특징이 있다. 랜덤하게 정책 업데이트가 이루어져, local minimum에 빠질 위험이 적다.

(3) DDPG (Deep Deterministic Policy Gradient)

연속적인 행동 공간을 처리하기 위한 학습 방법으로, Actor-Critic 구조를 기반으로 한다.

(4) PPO (Proximal Policy Optimization)

안정성과 효율성을 동시에 가진 정책 최적화 알고리즘으로, 업데이트 시 정책의 변화 폭을 제한하여 다양한 이해관계자의 이익을 고려하는 복잡한 환경에 적합하다.

(5) RPPO (Robust Proximal Policy Optimization)

PPO의 변형된 알고리즘으로, 불확실성과 변동성이 큰 환경에서 안정적인 성능을 유지한다. 전력망의 변동성이 있는 V2G 환경에서 신뢰성 있는 학습이 가능하다.

(6) SAC (Soft Actor-Critic)

엔트로피 강화 학습을 적용하여 탐색과 활용의 균형을 고려하는 알고리즘으로, 다양한 충전 및 방전 전략을 동시에 고려하는 연구에 적합하다.

(7) TD3 (Twin Delayed Deep Deterministic Policy Gradient)

강화 학습에서 특정 state에서 특정 action을 취했을때 기대되는 보상에 대한 값을 나타내는 Q-value가 DDPG에서는 과대 추정되는 경향이 있는데, 이러한 DDPG의 단점을 개선하기 위해 정책 업데이트를 지연시켜 안정적인 상태에서 정책 변경이 가능하도록 한다.

(8) TQC (Truncated Quantile Critics)

Q-value를 여러 구간으로 나누어 가치 추정을 제공하는 알고리즘으로, 정밀한 학습이 필요한 환경에 적합하다.

(9) TRPO (Trust Region Policy Optimization)

정책 업데이트 시 신뢰 구간을 설정하여 안정성을 높여, 복잡한 정책 최적화 과정에서 신뢰성 높은 성능을 보장한다.

해당 알고리즘의 특징들을 활용하여 V2G 이해관계자 각각에게 적합한 알고리즘을 판단하고, 이를 기반으로 EV2Gym 환경에서 이해관계 최적화에 대한 연구를 진행할 예정이다.

3. 연구 내용

3.1. 선행 연구

본 과제를 수행하기에 앞서 V2G 시뮬레이션 환경인 EV2Gym을 활용하여 PST Problem과 V2G Profit Maximization Problem, 그리고 Battery Degradation 관점에서 성능을 평가해보았다.

3.1.1. PST(Power Setpoint Tracking) Problem 평가

실제 전력 소비와 목표 전력 간의 차이를 최소화하는 것을 목표로 시스템이 목표 전력에 도달하도록 유도한다.

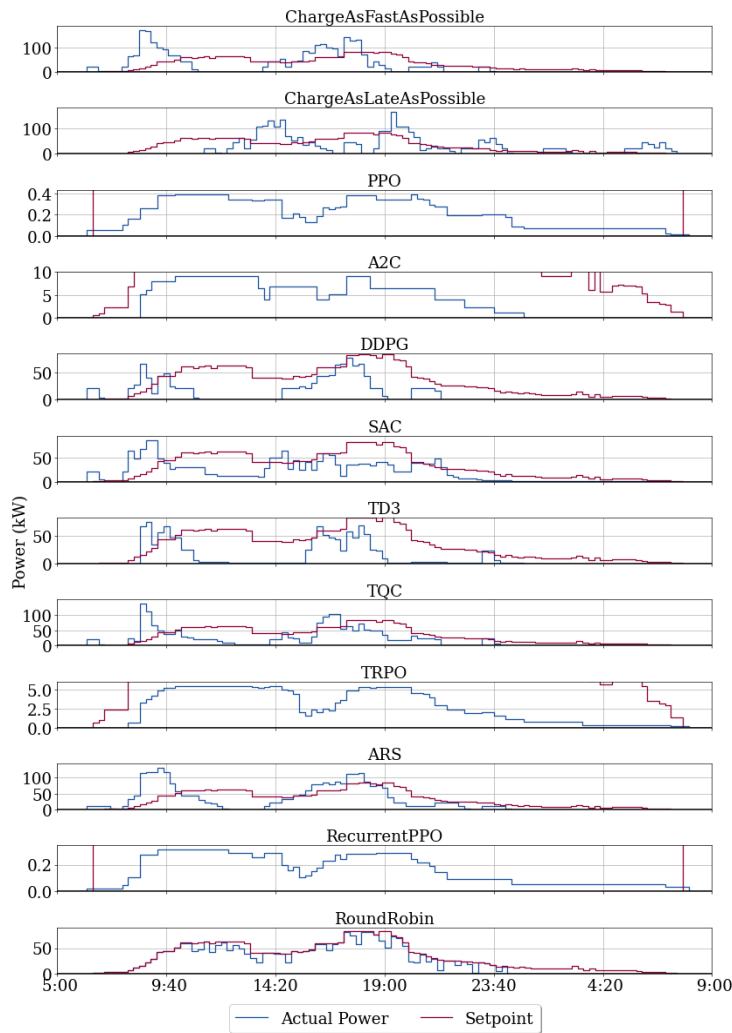


그림 4. PST Problem 평가 그래프

Heuristic 알고리즘 3가지, RL 알고리즘 9가지에 총 12개의 알고리즘에 대해 PST Problem을 평가해 보았다. *Charge As Fast As Possible* (AFAP) 알고리즘은 EV가 연결되자마자 가능한 최대 전력으로 즉시 충전을 시작한다. 변압기의 제약 조건을 고려하지 않고 충전 속도를 최우선으로 한다. *Charge As Late As Possible* (ALAP) 알고리즘은 EV가 목표 SoC에 도달할 수 있도록 가능한 한 늦게 최대 속도로 충전을 시작한다. *Proximal Policy Optimization* (PPO)는 강화 학습 기반 알고리즘으로, EV 충전 시 그리드의 상태를 고려하여 적절한 충전 패턴을 학습하고 주어진 목표에 따라 충전 속도를 유연하게 조절한다. *Advantage Actor-Critic* (A2C)는 강화 학습 기반 알고리즘으로, 충전 패턴을 최적화하기 위해 그리드 상태와 EV의 목표를 동시에 고려하여 점진적으로 충전 속도를 조절한다. *Deep*

Deterministic Policy Gradient (DDPG)는 연속적인 action 공간을 다루는 강화 학습 알고리즘으로, EV 충전 시 전력망 상태에 따라 실시간으로 충전 속도를 최적화한다. *Soft Actor-Critic* (SAC)는 그리드 제약 조건을 고려하여 EV 충전 시 전력 소비를 효율적으로 관리한다. *Twin Delayed Deep Deterministic Policy Gradient* (TD3)는 DDPG의 개선 알고리즘으로 그리드 상태와 EV 목표를 기반으로 충전 속도를 조절한다. *Truncated Quantile Critics* (TQC)는 분포 기반 강화 학습 알고리즘으로, 다양한 시나리오에서 안정적인 충전 성능을 보장하기 위해 충전 속도를 조절하고 전력망 제약의 반영하여 최적의 충전 패턴을 학습한다. *Trust Region Policy Optimization* (TRPO)는 강화 학습 알고리즘으로, 정책 변화의 안정성을 유지하면서 최적의 충전 전략을 학습하고, 그리드 상태를 반영하여 EV 충전 속도를 조절한다. *Augmented Random Search* (ARS)는 random 탐색을 통해 최적의 충전 패턴을 찾는 강화학습 알고리즘으로, 그리드 제약 조건을 고려하여 EV 충전 속도를 조절한다. *Recurrent PPO* (RPPO)는 과거의 충전 상태를 기억하여 현재 충전 결정을 최적화하는 알고리즘으로 시간에 따른 전력망 상태 변화를 고려하여 충전한다. *Round Robin* (RR)은 각 EV가 공평하게 에너지 나누도록 순환 방식으로 순차적 EV 최대 전력 setpoint 까지만 충전한다.

Algorithm	Energy Charged (kWh)	$\epsilon^{usr} (\%)$	Tracking Error (10^3)	Energy Error (kWh)	Reward (10^3)
A2C	114±38	72±4	140±42	809±158	-140±42
ARS	783±140	100±0	105±46	667±163	-59±23
AFAP	784±140	100±0	171±74	814±176	-95±44
ALAP	784±140	100±0	137±52	715±141	-137±52
DDPG	368±95	83±4	126±42	725±124	-126±42
PPO	6±2	68±4	182±52	930±158	-182±52
RPPO	5±1	68±4	182±52	931±157	-182±52
RR	745±135	98±1	14±5	181±34	-8±4
SAC	551±102	92±3	75±27	572±129	-75±27
TD3	306±93	81±4	121±40	709±134	-121±40
TQC	548±116	91±3	94±36	629±140	-94±36
TRPO	85±20	71±4	156±48	858±154	-156±48

그림 5. Baseline Algorithms for the PST Problem with 20 Chargers

20개의 Charging Station으로 15분 단위 총 85단계 PST Problem에 대해 학습 및 평가하였다. RL 알고리즘 중에서는 DDPG, SAC, TD3, TQC, ARS가 상대적으로 목표치와 근사하게 결과를 보임을 알 수 있다. 결과 그래프 상 Heuristic의 RR 알고리즘이 가장 우세한 성능을 보이나, RR은 충전소 수의 증가, 제약 조건의 증가에 따라 한계를 보여, RL을 중점으로 알고리즘을 설계 및 평가할 것이다.

3.1.2. V2G Profit Maximization 평가

V2G 이익 최대화 문제와 관련하여 RL 알고리즘 별 전력 소비 패턴과 사용자 만족도를 비교해보았다.

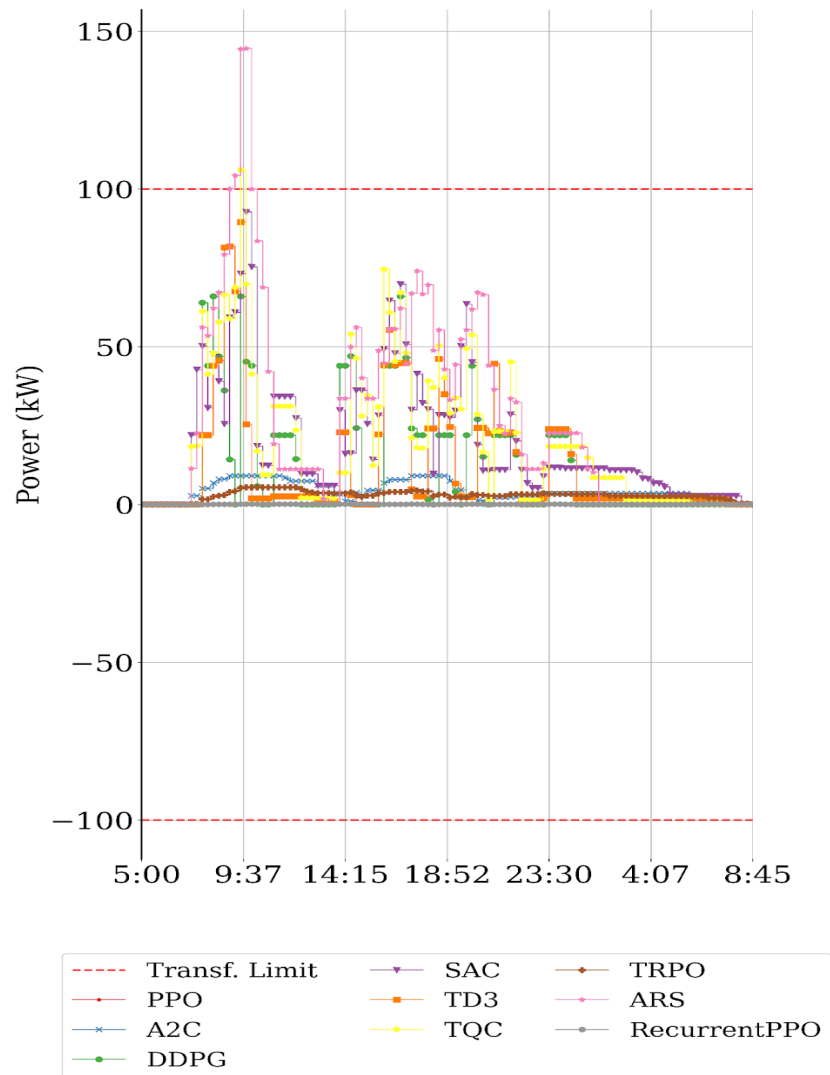


그림 6. 전력 소비 패턴

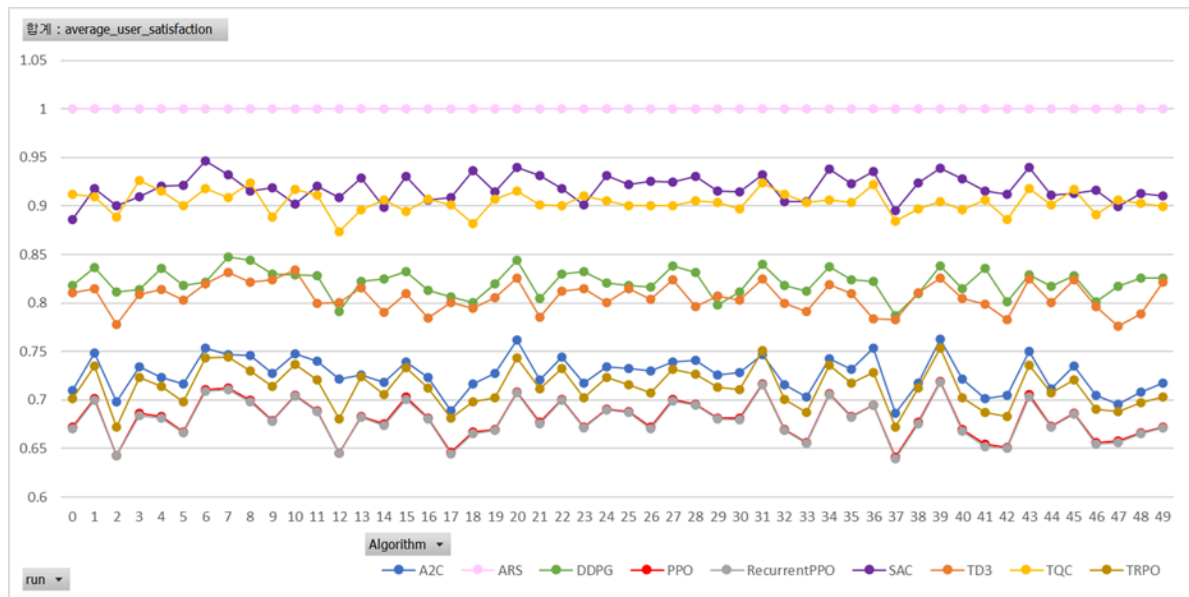


그림 7. RL 알고리즘 별 사용자 만족도

그래프에서 SAC, TQC RL 알고리즘이 전력 제한 범위 내에서 유연한 형태를 보이며, 사용자 만족도에서 또한 높은 만족도를 보임을 알 수 있다. 반면 전력 제한 범위에서 변동성을 거의 보이지 않은 A2C, TRPO, PPO, RPPO 알고리즘의 경우 사용자 만족도에서 낮은 결과를 보인다.

3.1.3. Battery Degradation 관점 평가

20개의 충전기로 하루 15분 간격 총 85단계 설정하고, 9가지 RL 알고리즘에 대해 각 5,000회 학습 후 배터리 열화 관점에서 각 알고리즘 별 성능을 평가해보았다.

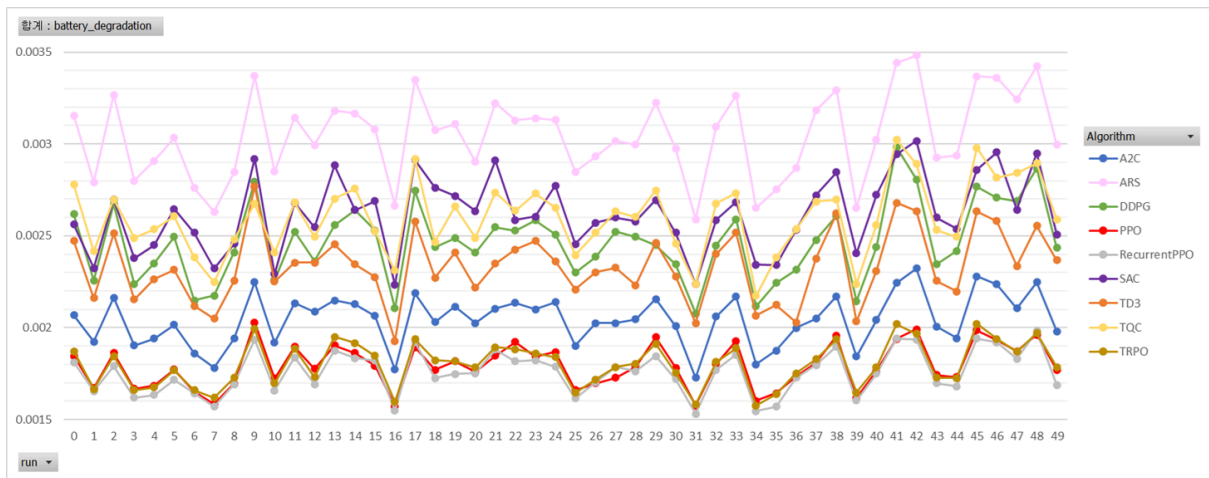


그림 8. RL 알고리즘 별 배터리 열화 평가 그래프

전기차 배터리 열화 범위가 연간 평균 0.5% ~ 1% [6], 하루 단위 약 0.0013% ~ 0.0027%임을 고려하였을 때, 가장 높은 열화를 보이는 ARS 알고리즘이 비효율적이며 하단의 RPPO, PPO, TRPO 알고리즘이 비교적 효율적인 결과를 보임을 알 수 있다.

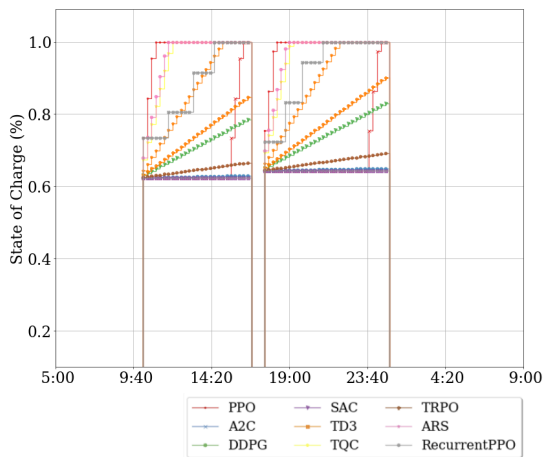


그림 9. SoC 추이 그래프

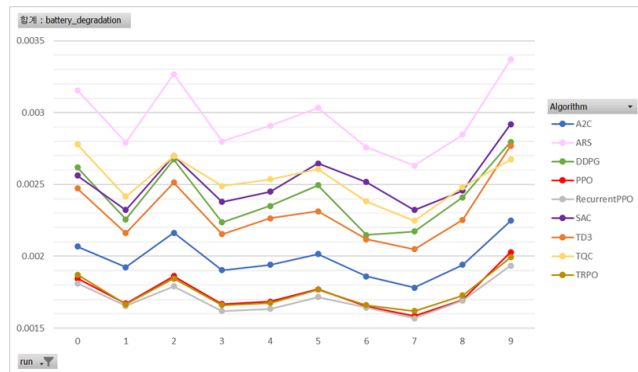


그림 10. 배터리 열화 그래프 (그림 8. 일부)

전기차 충전 상태 (SoC)와 배터리 열화 상태를 비교해보았을 때, SoC의 값이 1이 되는 즉, 완전 충전을 진행하는 ARS 알고리즘이 배터리 열화 측면에서도 비효율적인 결과를 보임을 알 수 있다.

위 실험 결과를 토대로 적정 수준의 SoC 범위 내에서 충방전이 이루어지고, 배터리 열화 지표 또한 고려하여 높은 사용자 만족도를 보이는 효율적인 배터리 충방전 알고리즘을 설계할 예정이다.

3.2. 이해관계자에게 유리한 V2G 충방전 보상/상태함수 알고리즘 설계

전력망 운영자, 전기차 사용자, 충전 서비스 사업자 총 3가지의 이해관계자별로 각자의 이익에 중요한 지표를 생각해보고, 이를 개선하기 위한 보상 함수 및 상태 함수를 설계해 보았다. 우선, 9가지 알고리즘을 사분면 그래프를 이용해 특징을 기준으로 구분해보았다.

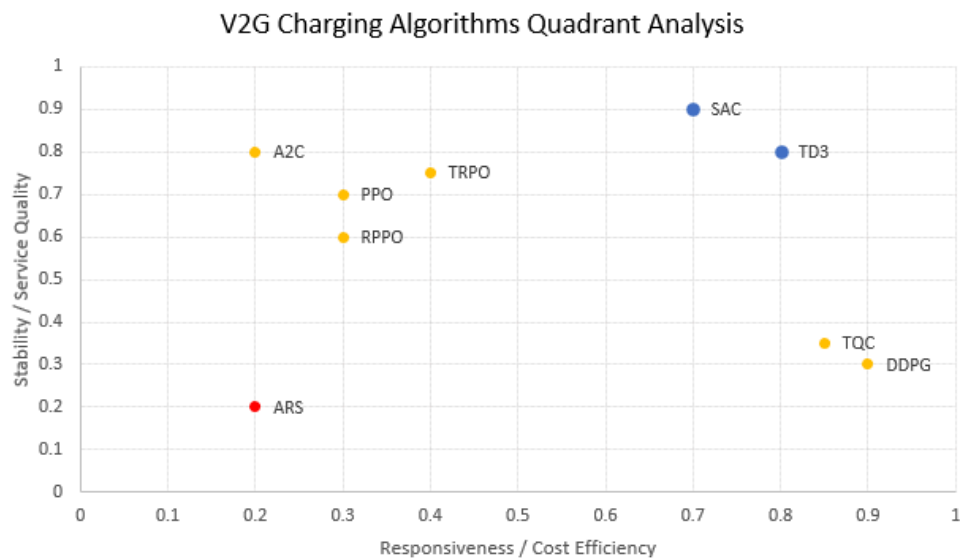


그림 11. 알고리즘 특징별 비교

제1사분면의 경우, 빠르게 전력 수요에 대응하면서도 높은 안정성을 제공하므로 전력망 운영자에게 적합한 알고리즘이라 판단했고, 제 2사분면의 경우 충전 시간과 안정성 측면에서 EV 사용자가 선호할 수 있는 특징을 제공하므로 전기차 사용자에게 적합한 알고리즘이라 생각했다. 또, 제 4사분면의 경우 비용 효율성을 극대화하면서도 빠르게 응답할 수 있는 능력을 제공하므로 충전 서비스 사업자에게 유리한 알고리즘으로 설정하고 실험을 진행했다.

3.2.1. 전력망 운영자

전력망 운영자는 전력 수요와 공급의 안정성을 중시하므로, 실제 전력 수요와 목표 수요 간의 차이를 줄이는 것이 중요하다. 따라서 전력망 운영자에게 중요한 지표로는 Total Energy Discharged, Power Tracker Violation, Tracking Error Total Transformer Overload, Total Profits를 살펴보았다. 다만, Total Energy Discharged의 경우 값이 0으로 설정되어 있기 때문에 이를 제외하고, 다른 지표에서 개선된 값을 가질 수 있는 보상/상태함수를 설계하였다.

(1) 보상함수

기존 SquaredTrackingErrorReward 보상함수 식은 다음과 같다.

$$reward = - (\min(P_{set}(t), P_{pot}(t)) - P_{cur}(t))^2$$

$P_{set}(t)$: 시간 t 에서의 목표 전력 설정 값 (power setpoint)

$P_{pot}(t)$: 시간 t 에서의 전력 잠재력 (power potential)

$P_{cur}(t)$: 시간 t 에서의 실제 전력 사용량 (current power usage)

전력망 운영자의 핵심 관심사는 총 수익 및 전력 안정성이므로 전력 수익 극대화와 더불어 전력 공급의 안정성을 높이는 데 중점을 두었다. 따라서 이를 위해 수익과 변압기 과부하 및 추적 오차 페널티를 추가해 다음과 같이 보상함수를 구현하였다.

$$reward = 1.0 * Profit_{tot} - 0.5 * O_{tr} - 0.5 * (\min(P_{set}(t), P_{pot}(t)) - P_{cur}(t))^2$$

$Profit_{tot}$: 총 이익

O_{tr} : 변압기 과부하량

$P_{set}(t)$: 시간 t 에서의 목표 전력 설정 값

$P_{pot}(t)$: 시간 t 에서의 전력 잠재력

$P_{cur}(t)$: 시간 t 에서의 실제 전력 사용량

(2) 상태함수

기존 상태 함수 PublicPST는 현재 시간, 전력 설정값, 현재 전력 사용량 및 변압기와 EV 상태만 반영하므로, 변압기 과부하와 총 방전 에너지를 포함하여 전력 안정성 및 에너지 방출 주기를 고려하도록 확장해 구현하였다.

$$state = [s_t, P_{set}(t), P_{cur}(t), E_{dis}, O_{tr}]$$

s_t : 현재 시간 스텝

$P_{set}(t)$: 시간 t 에서의 전력 설정 값

$P_{cur}(t)$: 시간 t 에서의 현재 전력 사용량

E_{dis} : 방전된 총 에너지

O_{tr} : 변압기 과부하량

(3) 결과

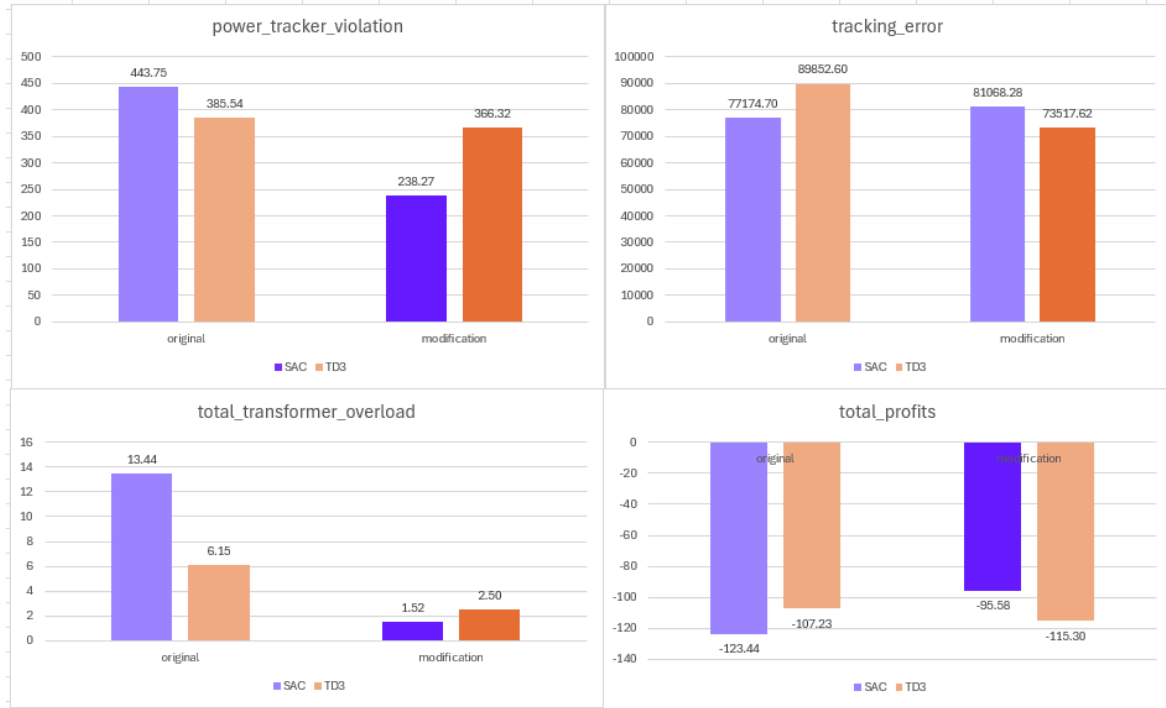


그림 12. 전력망 운영자 입장에서의 알고리즘 결과 그래프

위 그림에서 보면 대부분의 경우 수정된 보상, 상태 함수하에서 성능이 개선됨을 확인할 수 있다. 특히 total_transformer_overload의 값을 매우 줄이는 등 좋은 결과를 보였지만, SAC의 경우 tracking error가 소폭 증가, TD3의 경우 total profits이 소폭 하락하는 결과를 보였다. 하지만 전력망 운영자의 입장에서 볼 때, 보상함수 및 상태함수의 변경으로 인해 전체적으로 운영 효율성이 증가했음을 알 수 있다.

3.2.2. 전기차 사용자

전기차 사용자는 충전 품질 및 서비스 안정성을 중요하게 생각하므로, 사용자 만족도를 높이는 동시에 과부하를 방지해야 한다. 따라서 Average User Satisfaction, Energy User

Satisfaction, Total Energy Charged, Battery Degradation (Calendar and Cycling), Total Reward를 주요 지표로 설정하고 개선된 값을 가질 수 있는 보상/상태함수를 설계하였다. 이때, Energy User Satisfaction은 고정된 값이기 때문에 결과 그래프에서는 제외하고 나타내었다.

(1) 보상함수

기존 ProfitMax_TrPenalty_UserIncentives 보상함수 식은 다음과 같다.

$$reward = C_{tot} - 100 * \sum(O_{tr}) - 100 * \sum(e^{-10*S_i})$$

C : 총 비용

O_{tr} : 변압기 과부하 비율

S_i : 사용자 만족도 목록

전기차 사용자의 주요 관심사는 충전 비용을 최소화하고, 배터리 열화를 줄이며, 사용자 만족도를 극대화하는 것이다. 따라서, 전기차 사용자의 관점에서 평균 사용자 만족도, 에너지 사용자 만족도, 배터리 열화, 총 충전 에너지 및 보상 등을 고려해 보상 함수를 설계해보려고 했다.

$$reward = -1.0 * C_{tot} + 100.0 * S_{avg} + 100.0 * E_{sat} - 50.0 * D_{cal} - 50.0 * D_{cyc}$$

C_{tot} : 총 비용

S_{avg} : 평균 사용자 만족도

E_{sat} : 에너지 만족도

D_{cal} : calendar 기반 배터리 열화

E_{cyc} : cycling 기반 배터리 열화

(2) 상태함수

기존 상태 함수 V2G_profit_max_loads는 현재 전력 사용량 및 충전 가격을 반영하고 있다. 그러나, 배터리 상태와 총 충전 에너지, 사용자 만족도를 추가하면 사용자 경험을 더욱 개선할 수 있다고 판단해서 다음과 같이 설계하였다.

$$state = [s_t, P_{cur}(t), P_{charge}, E_{charged}, B_{stat}, S_{avg}]$$

s_t : 현재 시간 단계

$P_{cur}(t)$: 시간 t 에서의 현재 전력 사용량

P_{charge} : 충전 가격

$E_{charged}$: 충전된 총 에너지

B_{stat} : 배터리 상태

S_{avg} : 평균 사용자 만족도

(3) 결과

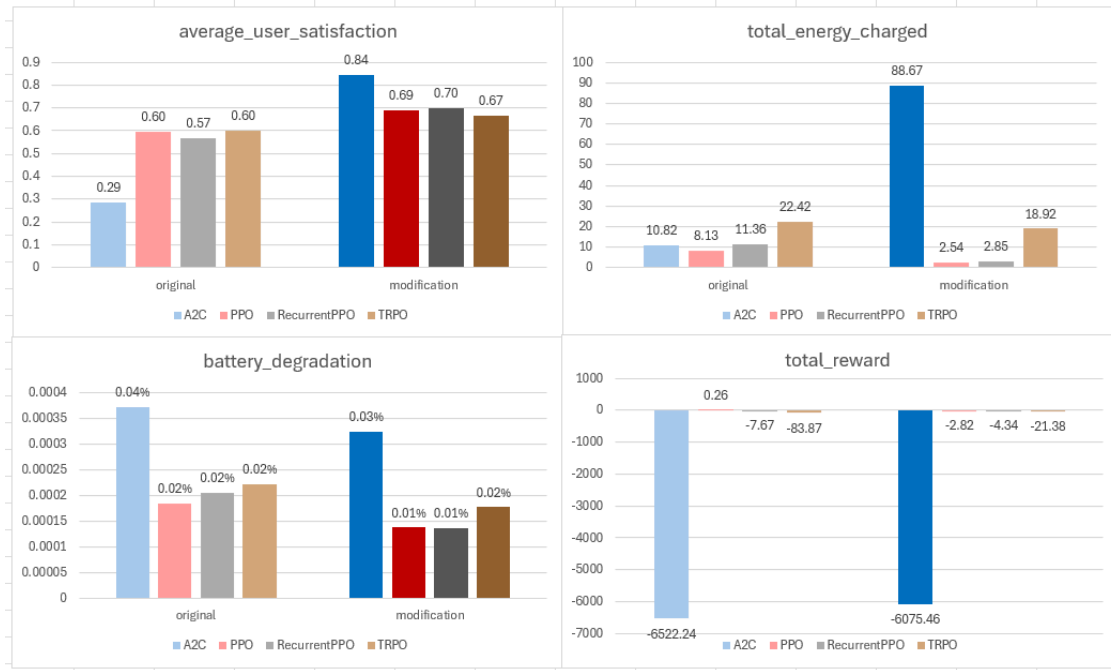


그림 13. 전기차 사용자 입장에서의 알고리즘 결과 그래프

그림 13에서 보면, total_energy_charged는 A2C에서만 성능 개선이 있고 다른 알고리즘의 경우 수치가 소폭 하락했으나, 나머지 지표에서 모두 성능 개선이 이루어졌다. 전기차 사용자의 경우 배터리 열화 지표와 사용자 만족도 지표를 더 중요하다 판단했기에 해당 지표의 보상 및 페널티 가중치를 크게 주었고, 해당 결과로 전기차 사용자의 목표를 더 잘 반영하고 성능을 향상시킬 수 있음을 확인했다.

3.2.3. 충전 서비스 사업자

충전 서비스 사업자는 수익을 극대화하는 것을 목표로 하며, 비용 최소화과 수익 극대화에 집중한다. 따라서 Total Profits, Total EV Served, Discharge Price Factor, Control Horizon, Battery Degradation 을 주요 지표로 설정하고 개선된 값을 가질 수 있는 보상/상태함수를

설계하였다. Discharge Price Factor, Control Horizon은 고정된 값으로 설정되어있으므로 결과 그래프에 제외하고 나타내었다.

(1) 보상함수

기존 profit_maximization 보상함수 식은 다음과 같다.

$$reward = C_{tot} - 100 * \sum(1 - S_i)$$

C_{tot} : 총 비용

S_i : 사용자 만족도 목록 내 각 요소

충전 서비스 사업자는 총 수익을 극대화하고, 고객 만족도를 높이며, 방전 가격을 반영해 수익성을 개선, 배터리 열화를 최소화하는 등의 요소를 고려해야 한다고 생각했다. 따라서, 수정된 보상 함수에서는 수익을 극대화하면서 사용자 만족도와 배터리 열화 페널티를 추가한 모습을 볼 수 있다.

$$reward = 1.0 * Profit_{tot} - 0.3 * D_{bat} - 80.0 * (1 - S_{avg}) - 50.0 * F_{dis}$$

$Profit_{tot}$: 총 수익

D_{bat} : 배터리 열화

S_{avg} : 평균 사용자 만족도

F_{dis} : 방전 가격 요소

(2) 상태함수

기존 상태 함수 V2G_profit_max는 현재 전력 사용량 및 충전 가격을 반영한다. 하지만 고객 요구와 충전 서비스의 효과성을 높이기 위해 배터리 상태 및 충전 고객 수를 추가해야 한다고 생각해 총 방전 에너지, 충전소에서 서비스하는 총 전기차 수를 요소로 추가해 상태 함수를 수정하였다.

$$state = [s_t, P_{cur}(t), Price_{charge}, E_{charged}, EV_{served}]$$

s_t : 현재 시간 단계

$P_{cur}(t)$: 시간 t 에서의 현재 전력 사용량

$Price_{charge}$: 충전 가격

$E_{charged}$: 충전된 총 에너지

EV_{served} : 서비스된 EV의 총 수

(3) 결과

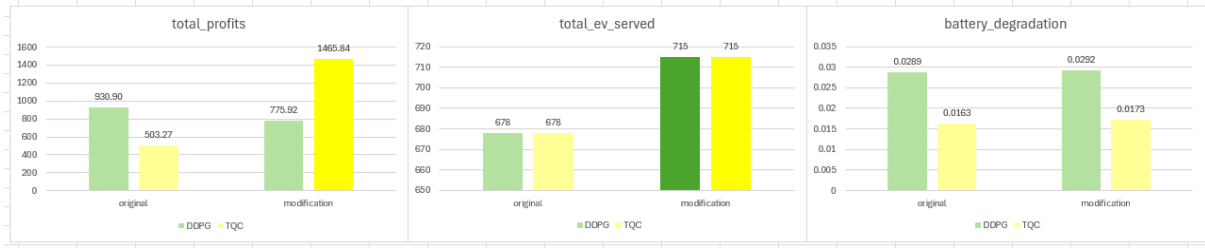


그림 14. 충전소 서비스 사업자 입장에서의 알고리즘 결과 그래프

S_{avg} 의 가중치를 100.0으로 두고 실험을 진행해보았을 때, DDPG 알고리즘의 경우 total_profits이 930.9에서 1495.8로 개선되었지만, TQC의 경우 하락했고, 배터리 열화 수치가 0.03 이상이 되는 등 부족한 점을 보였다. 따라서 S_{avg} 의 가중치를 80으로 조정 후 실험을 진행했고, TQC의 total_profits 및 배터리 열화, 서비스한 총 전기차 수 지표에서 이전보다 향상된 결과를 얻을 수 있었다.

이를 통해 보상함수에서는 총 수익, 사용자 만족도, 방전 가격, 배터리 열화를 포함하여 수익성을 높였고, 상태함수는 총 방전 에너지와 총 EV 수를 포함하여 서비스의 효과성을 강화할 수 있었다고 생각한다.

3.3. 이해관계 최적화를 위한 통합된 V2G 충방전 보상/상태함수 알고리즘 설계

충전 서비스 사업자, 전기차 사용자, 전력망 운영자를 모두 고려하여 통합 상태 함수와 통합 보상 함수를 설계하고, 이를 통해 각 이해관계자의 목표를 균형 있게 반영하고자 했다. 이해관계자 모두에게 중요한 공통 지표는 average_user_satisfaction, total_energy_charged, total_energy_discharged, battery_degradation이라고 생각했는데, 이유는 다음과 같다.

- (1) average_user_satisfaction : 전기차 사용자의 만족도는 충전 서비스와 전력망의 성공을 평가하는 중요한 요소이다.

(2) *total_energy_charged*, *total_energy_discharged* : 전력망 운영자는 전력 수급과 배출을 조정해야 하며, 충전 서비스 제공자는 수익성과 고객 요구를 충족시켜야 하고, 전기차 사용자들은 필요에 따라 충전과 방전을 원활히 진행하길 원한다.

(3) *battery degradation* : 배터리 수명과 유지 비용은 충전 서비스 제공자와 사용자 모두에게 직접적인 영향을 주고, 전력망 운영자에게도 간접적인 영향을 미칠 수 있다.

따라서 이 지표들을 중점적으로 생각하며 3개의 통합된 상태, 보상 함수를 설계해 보았다.

3.3.1. 상태함수

이 통합 상태 함수는 전력망 운영자, 충전 서비스 제공자, 전기차 소유자의 각 상태 정보를 조화롭게 결합하여, 세 이해관계자 간의 상충되는 목표를 균형 있게 고려하는 것을 목표로 하였다. 3가지의 통합 함수에 모두 동일하게 적용하였으며, 전력망 운영자에게는 에너지 수요와 전력 제한 정보, 충전 서비스 제공자에게는 평균 출력과 과부하 정보, 전기차 소유자에게는 배터리 SOC와 충전량, 배터리 손상 정도를 반영함으로써 각자의 특화된 요구를 구체적으로 반영하고자 했다. 이를 통해 개별 이해관계자가 중요한 결정을 내릴 수 있는 충분한 상태 정보를 제공해 전체적인 시스템의 효율성을 높일 수 있을 것이라 생각한다. 설계한 통합 상태 함수식은 다음과 같다.

$$state = [s_t, P_{cur}(t), charge\ price, s_{load}, s_{power\ limit}, s_{avg\ power}, s_{avg\ overload}, s_{setpoint}]$$

s_t : 현재 시간 스텝

$P_{cur}(t)$: 시간 t 에서의 현재 전력 사용량

charge price : 충전 가격

s_{load} : 변압기 에너지 부하 정도

$s_{power\ limit}$: 변압기 최대 허용 전력 한계치

$s_{avg\ power}$: 평균 충전 전력

$s_{avg\ overload}$: 평균 전압 과부하

$s_{setpoint}$: 목표 전력 수준

3.3.2. 보상함수

(1) 통합 함수 1

각 이해관계자의 성과 지표를 포함해 총 수익, 사용자 만족도, 배터리 열화, 전력 사용 효율성 등을 반영하려고 했다. 특히, 전력망 운영자의 안정성을 위해 전력 사용량과 관련된 패널티를 강조하며, 충전 서비스 사업자의 수익과 전기차 사용자의 만족도를 적절히 조화시키는 데 중점을 두어 설계했다.

$$reward = 0.5 * reward_{grid} + 0.5 * reward_{provider} + 0.5 * reward_{ev}$$

$$reward_{grid} = -100 * \max(0, m - P_{actual}) - 50 * o - 100 * I - 10 * T - 5 * E$$

P_{actual} : 변압기의 실제 전력 사용량

m : 변압기의 최소 전력

o : 변압기 과부하 수준

I : 과부하 발생 여부

T : 추적 오차

E : 전력 추적 오차

$$reward_{provider} = -100 * C + 10 * E_{ch} - 50 * \max(0, Th - r) + 100 * I - 30 * \max(0, 0.7 - S)$$

C : 총 비용

E_{ch} : 총 충전된 에너지

Th : efficiency threshold

r : 에너지 효율 비율

I : 과부하 발생 여부

S : 평균 사용자 만족도

$$reward_{ev} = 0.1 * H + 0.1 * E_c + 0.1 * E_d - 0.0005 * E_{exchanged} - 0.02 * (P_c - P_d) + 0.01 * A$$

H : 평균 배터리 건강 상태

E_c : 비피크 시간대 충전 에너지

E_d : 피크 시간대 방전 에너지

$E_{exchanged}$: 교환된 에너지

P_c : 충전 가격

P_d : 방전 가격

A : 평균 전력 출력

위와 같이 전력망 운영자, 충전 서비스 사업자, 전기차 사용자의 보상함수를 결합하는 방식으로 통합된 보상함수를 만들었다.

Algorithm	Energy Charged (kWh)	ϵ^{usr} (%)	Tracking Error (10^3)	Energy Error (kWh)	Reward (10^3)
A2C	282±99	79±3	61±27	407±116	-37660±3430
ARS	152±88	69±9	88±53	503±161	-42241±4086
DDPG	247±80	63±1	168±89	636±143	-47835±6194
PPO	8±2	67±11	98±53	536±146	-43512±207
RPPO	8±2	67±11	95±51	528±143	-43177±134
SAC	407±102	92±1	68±41	426±150	-33683±2959
TD3	405±95	92±1	66±33	418±122	-33611±3908
TQC	360±110	83±3	79±39	455±125	-36911±3051
TRPO	180±73	71±9	81±51	486±148	-41518±3010

그림 15. 20개의 충전소로 평가한 통합 함수 1 알고리즘 결과

20개의 Charging Station으로 15분 단위 총 85단계, 9개의 RL 알고리즘으로 통합 함수 1에 대해 학습 및 평가한 후, 그림 15와 같이 결과를 정리해보았다. A2C는 낮은 Tracking Error 값을 유지하며 높은 에너지 충전량 수치로 우수한 성능을 보인다. ARS는 상대적으로 높은 에너지 충전량을 보이나, 그에 반해 에너지 오류 값이 크다는 단점이 있다. SAC 알고리즘은 사용자 만족도, 보상에 대해 가장 높은 값을 보이며 안정적인 성능을 유지함을 알 수 있다. TD3 알고리즘 또한 높은 사용자 만족도와 낮은 추적 오류 값을 보이며 안정적인 성능을 보여준다.

(2) 통합 함수 2

통합 함수 2에서는 각 이해관계자의 목표에 대해 더 강한 비중을 두어 최적화를 추구하였다. 특히, 전력망의 과부하와 전력 추적 위반에 대해 높은 패널티를 부여해 전력망 안정성을 강조했다. 반면 이전 함수는 각 요소에 0.5 비중을 부여하여 더 균형 있는 접근을 취하고 있어, 안정성보다는 전체적인 운영 효율성을 고려한 최적화를 목표로 한다는 차이점이 있다. 통합 함수 2는 과부하 및 사용자 만족도에 더 강한 패널티를 부여해 각각의 이해관계자 목표를 보다 강하게 반영하고 있다. 설계한 함수는 아래에 정리되어 있다.

$$reward = 1.0 * reward_{grid} + 1.0 * reward_{provider} + 1.0 * reward_{ev}$$

$$reward_{grid} = - \sum_{i=1}^N (1000 * \max(0, m_i - P_i) + 500 * o_i + 1000 * I_i) - 100 * Vio - 50 * Err$$

N : 변압기의 총 개수

m_i : 변압기 i 의 최소 전력

P_i : 변압기 i 의 실제 전력 사용량

o_i : 변압기 i 의 과부하 수준

I_i : 변압기 i 의 과부하 발생 여부

Vio : 전력 추적 위반량

Err : 추적 오차

$$reward_{provider} = -1000 * C + 10 * E_{ch} - 500 * \max(0, Th - r) + 1000 * I - 300 * \max(0, 0.7 - S)$$

C : 총 비용

E_{ch} : 총 충전된 에너지

Th : efficiency threshold

r : 에너지 효율 비율

I : 만족도 기준 초과 여부

S : 평균 사용자 만족도

$$reward_{ev} = 0.1 * H + \sum_{i=1}^M (0.1 * E_{c,i} + 0.1 * E_{d,i}) - 0.0005 * E_{exch} - 0.02 * (P_c - P_d) + 0.01 * A$$

H : 평균 배터리 건강 상태

E_c : 비피크 시간대 충전 에너지

E_d : 피크 시간대 방전 에너지

E_{exch} : 교환된 에너지

P_c : 충전 가격

P_d : 방전 가격

A : 평균 전력 출력

Algorithm	Energy Charged (kWh)	ϵ^{usr} (%)	Tracking Error (10^3)	Energy Error (kWh)	Reward (10^3)
A2C	317±72	75±3	93±32	497±121	-16164±5006
ARS	168±40	73±6	68±41	444±122	-15290±5640
DDPG	252±98	63±3	153±84	618±201	-20754±11375
PPO	4±3	66±7	100±50	546±113	-17771±6698
RPPO	5±2	66±6	96±48	536±113	-17496±6754
SAC	389±114	88±2	69±34	429±103	-13690±5812
TD3	313±111	75±2	102±52	513±153	-16683±7965
TQC	437±103	95±0	67±42	423±137	-12725±6845
TRPO	293±79	87±6	25±17	259±87	-11561±5503

그림 16. 20개의 충전소로 평가한 통합 함수 2 알고리즘 결과

20개의 Charging Station으로 15분 단위 총 85단계, 9개의 RL 알고리즘으로 통합 함수 2에 대해 학습 및 평가한 후, 그림 16과 같이 결과를 정리해보았다. A2C는 에너지 충전량이 높고, Tracking Error 값이 낮은 전반적으로 안정적이고 우수한 성능을 보이며, 사용자 만족도 또한 높은 평가를 받았다. RPPO 알고리즘도 다른 알고리즘에 비해 Tracking Error 수치가 낮아, 상대적으로 안정적인 성능을 보여준다.

(3) 통합 함수 3

통합 함수 3에서는 전력망 안정성에 대해 강화된 페널티를 제공할 수 있도록 SQE(Squared Error) 기반으로 보상함수를 설계해보았다. 과부하 시 고강도의 페널티를 부과함으로써 실시간 전력 사용량을 엄격히 조정할 수 있도록 했다. 또한, 사용자 경험 개선과 효율성 향상을 동시에 도모하기 위해 에너지 효율성 및 평균 사용자 만족도를 중요한 요소로 다루어 함수를 설계했다. 더불어 전기차 소유자의 장기적인 배터리 관리를 위해 평균 배터리 상태가 기준보다 낮거나 높을 때 페널티를 적용하는 방식으로 배터리의 건강 상태를 철저히 관리할 수 있도록 했다. 이를 기반으로 설계한 보상함수 식은 다음과 같다.

$$reward = 1.0 * reward_{grid} + 1.0 * reward_{provider} + 1.0 * reward_{ev}$$

$$reward_{grid} = - (\sum_{tr} (1000 * \max(0, m - P_{cur}(tr)) + 500 * O(tr) + 1000 * 1_{\{I(tr)\}})) + 100 * V^2 + 50 * E^2$$

m : 변압기의 최소 전력 합계

$P_{cur}(tr)$: 변압기의 현재 전력 사용량

$O(tr)$: 변압기의 과부하 정도

$1_{\{I(tr)\}}$: 변압기가 과부하일 경우 1, 아닐 경우 0

V : calculate power tracker violation (env) (전력 합계 위반)

E : calculate tracking error(env) (추적 오차)

$$reward_{provider} = - 1000 * C + 10 * E_c + I$$

$$I: - 500 * (T - R)^2 \text{ if } R < T, + 1000 * (A - T)^2 \text{ if } A > T, - 500 * (T - A)^2$$

C : 총 비용

E_c : 총 충전 에너지

$R = E_e / (E_d + \epsilon)$ (E_d : total energy discharged, ϵ : 작은 양수)

T : 효율성 기준치 (0.8)

A : 평균 사용자 만족도

$$reward_{ev} = -((H - 0.8)^2 + \sum_{cs} \sum_{port} (1_{\{c < 0\}} * (|I|^2 * 0.1)) + \sum_{cs} (-(|P_c - P_d|^2 * 0.02)) + (A_p - 0)^2 * 0.01)$$

H : 평균 배터리 상태

I : 현재 신호

P_c : 현재 충전 가격

P_d : 현재 방전 가격

A_p : 평균 전력 출력

Algorithm	Energy Charged (kWh)	ϵ^{usr} (%)	Tracking Error (10^3)	Energy Error (kWh)	Reward (10^6)
A2C	389±95	91±2	58±26	394±100	-53911±45669
ARS	176±39	73±6	69±30	452±123	-67407±77947
DDPG	264±69	67±3	132±49	576±123	-138412±175105
PPO	6±2	68±6	92±38	522±121	-82295±95699
RecurrentPPO	6±2	68±5	91±38	518±119	-81343±95030
SAC	243±90	63±3	164±73	636±168	-184278±188439
TD3	266±95	67±3	127±43	564±106	-135877±179302
TQC	200±62	55±2	209±66	715±123	-253553±295734
TRPO	257±68	85±6	30±17	287±94	-44297±34758

그림 17. 20개의 충전소로 평가한 통합 함수 3 알고리즘 결과

20개의 Charging Station으로 15분 단위 총 85단계, 9개의 RL 알고리즘으로 통합 함수 3에 대해 학습 및 평가한 후, 그림 17과 같이 결과를 정리해보았다. A2C는 에너지 충전량이 높고, Tracking Error 값이 낮은 전반적으로 안정적이고 우수한 성능을 보이며, 사용자 만족도 또한 높은 평가를 받았다. TRPO 알고리즘도 다른 알고리즘에 비해 Tracking Error 수치가 낮고 높은 사용자 만족도 결과를 보이며 상대적으로 안정적인 성능을 보여준다.

3.3.3. 결과

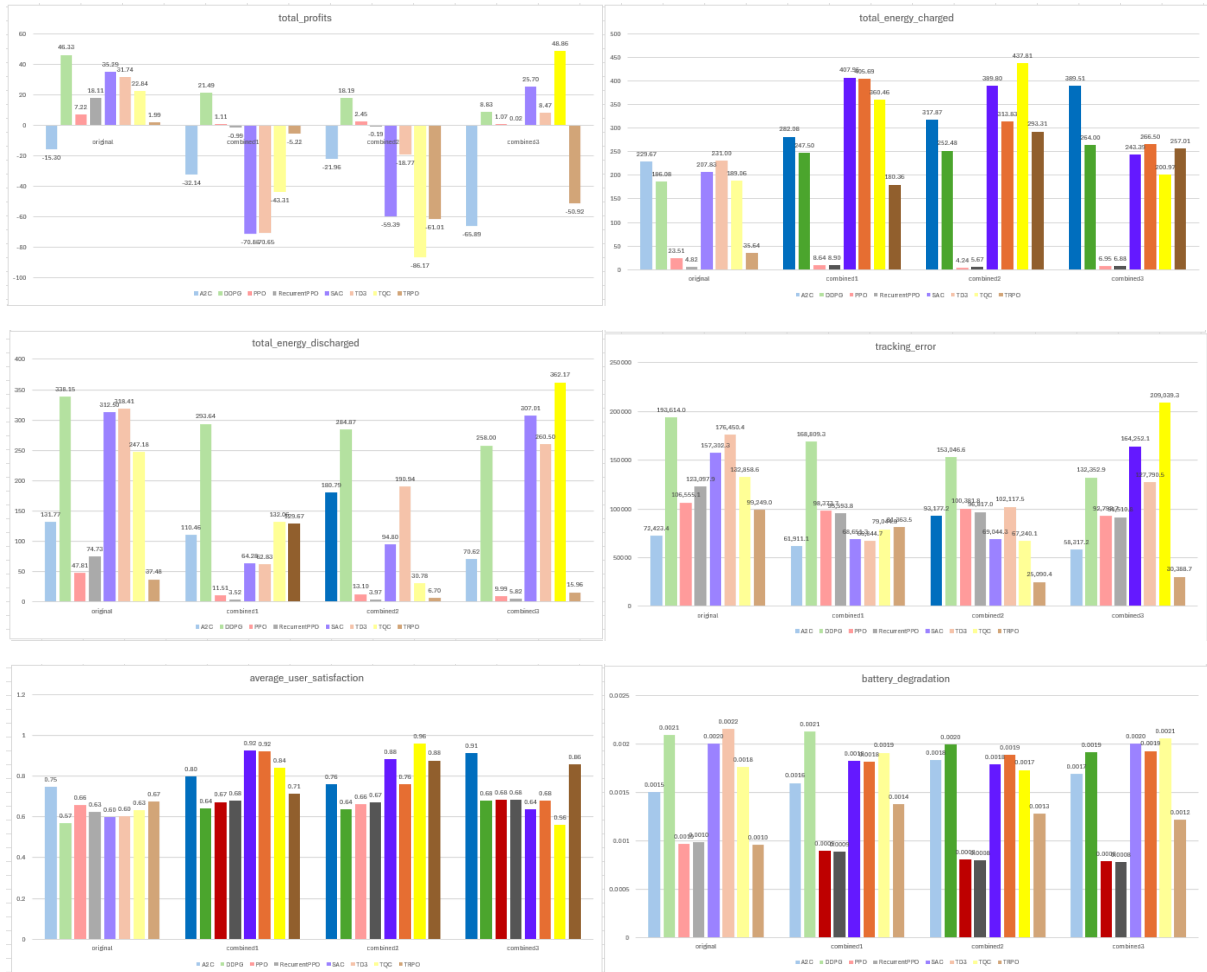


그림 15. 전체 결과 그래프

위의 그래프에서 진한 색으로 표시된 부분이 기존과 비교해 성능의 개선이 있는 지표이다.

먼저 ,total_energy_charged와 total_energy_discharged 지표를 자세히 살펴보면 다음과 같다.

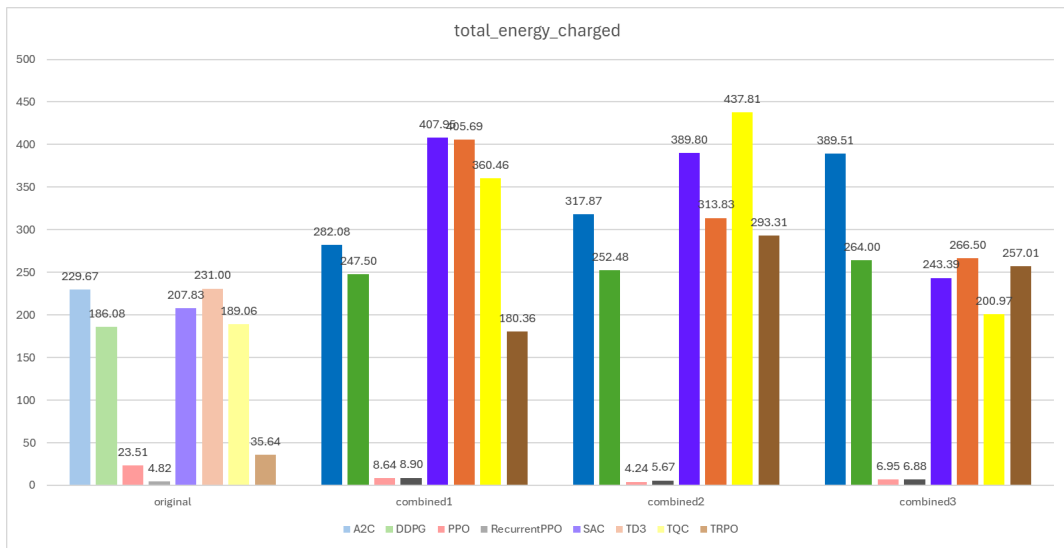


그림 16. total_energy_charged

통합 함수 1, 2, 3 모두 PPO를 제외하고 전체적으로 수치가 증가하는 모습을 보였다. PPO는 보상 함수의 변화에 대해 다소 민감하게 반응하지 않기 때문에, 충전 에너지 관련 보상이 높아지더라도 안정적인 수익을 유지하려는 성향을 유지해서 이러한 결과가 나온 것으로 생각된다.

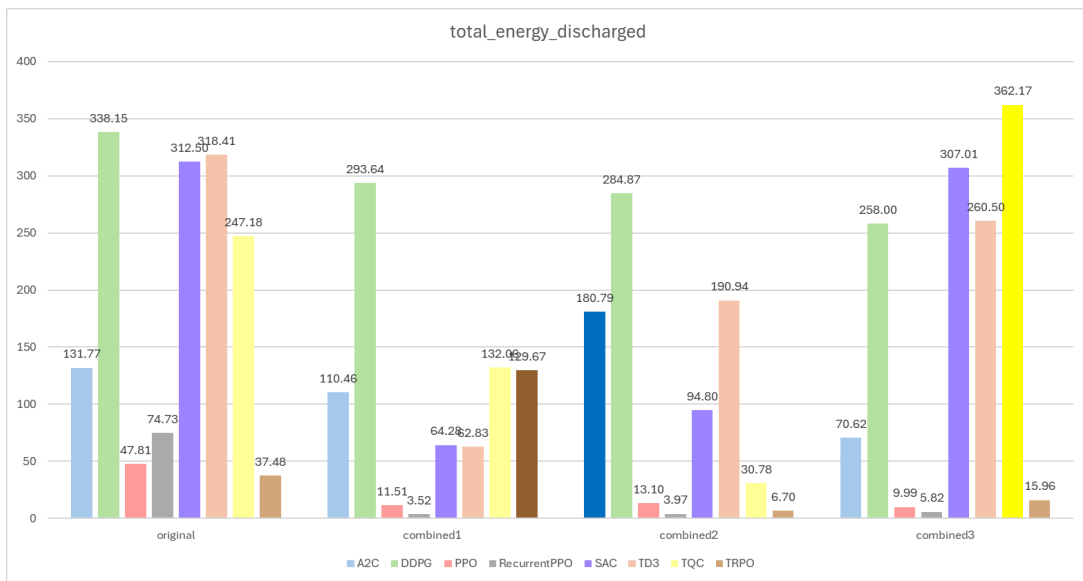


그림 17. total_energy_discharged

total_energy_discharged의 경우 total_energy_charged와 다르게 많은 알고리즘에서 개선이 일어나지는 않았다. 방전이 과도하게 진행되면 배터리 열화가 가속화되고 사용자

만족도가 낮아질 수 있기 때문에, 이러한 부분을 고려하여 보상 구조를 설계하였다. 따라서 대부분의 알고리즘이 방전보다는 충전에 초점을 맞추게 되어 방전 에너지 개선이 적게 나타난 것으로 추정된다.

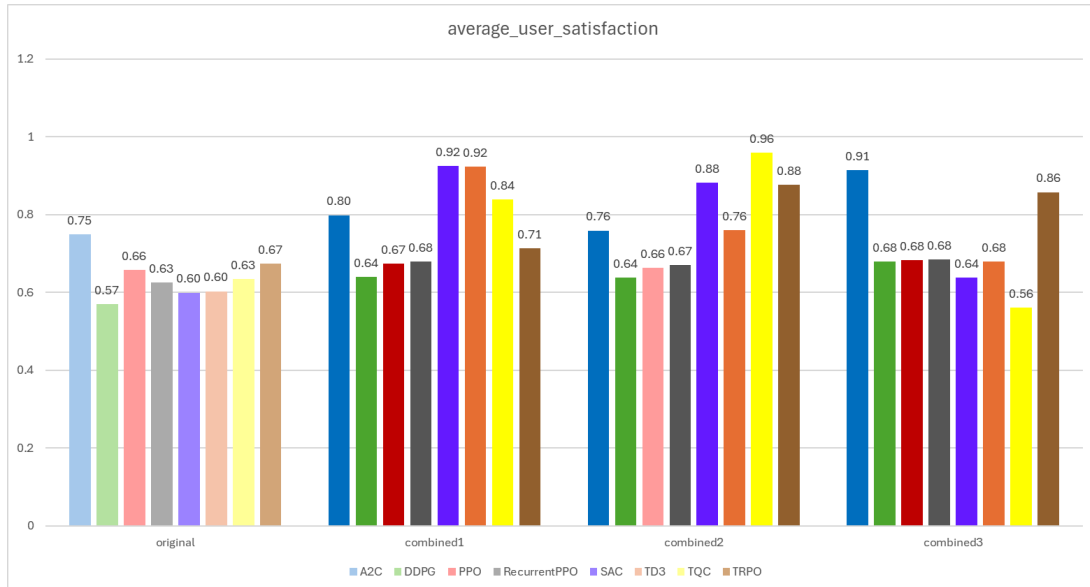


그림 18. average_user_satisfaction

사용자 만족도를 중요한 지표로 생각하고, 이를 고려하여 설계했기 때문에 사용자 만족도의 경우 거의 모든 알고리즘에서 개선된 모습을 보일 수 있었다.

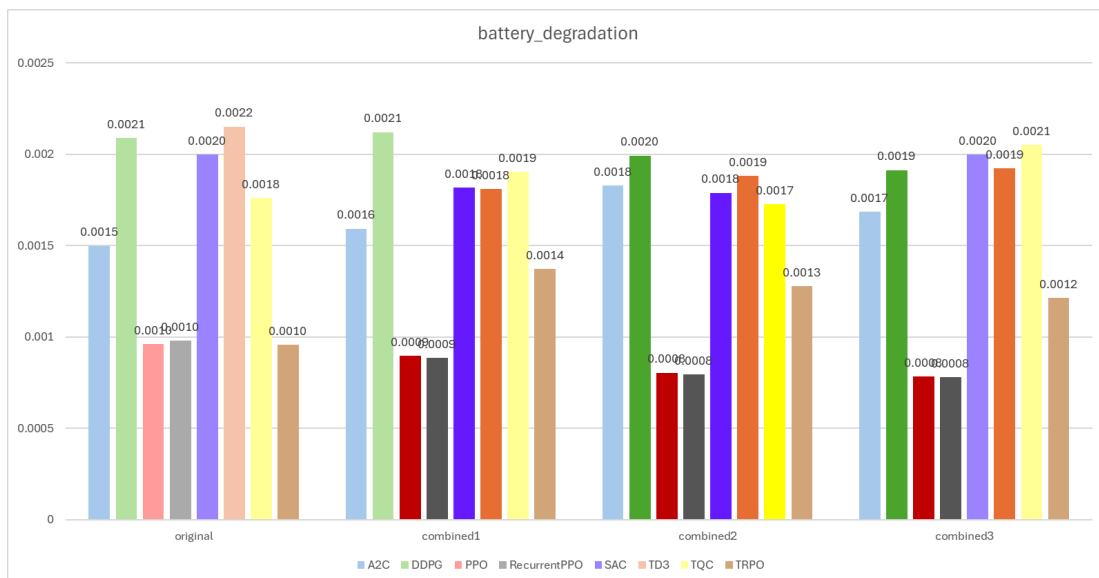


그림 19. battery_degradation

배터리 열화 지표의 경우, PPO와 RecurrentPPO,TD3의 경우 모든 통합 함수에서 수치가 향상되었으나, SAC는 통합 함수 1, 2에서만, DDPG는 통합 함수 2,3에서만, TQC는 통합함수 2에서만 성능이 개선되었다. PPO, RecurrentPPO, TD3는 상대적으로 적응력이 뛰어난 알고리즘으로, 다양한 보상 구조에서 빠르게 학습하고 최적화를 수행할 수 있기 때문에 상대적으로 일관된 성과를 보였다고 생각된다. SAC, DDPG, TQC의 경우는 특정 하이퍼파라미터에 민감하게 반응하거나, 특정 환경에서 최적화되지 않아 모든 경우의 성능 개선이 이루어지지 않았던 것으로 추정된다.

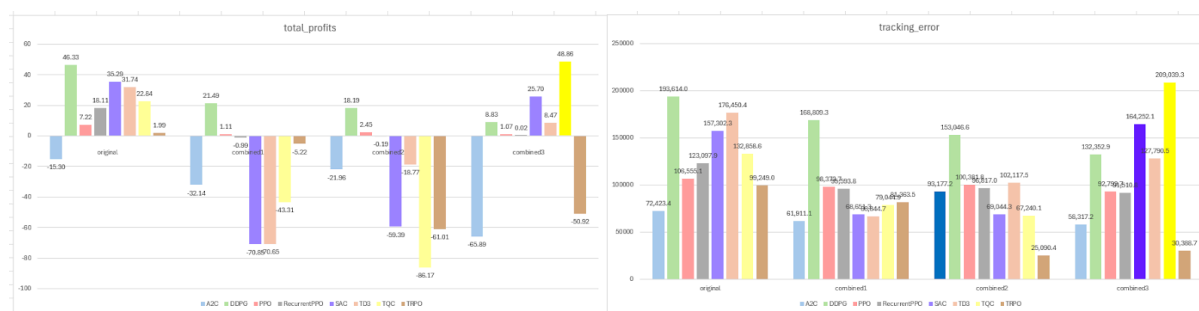


그림 20. total_profits & tracking_error

total_profits과 tracking_error의 경우 성능 개선이 거의 이루어지지 않았는데, 이는 보상함수 설계 시 다른 지표를 더 우선순위로 두고, 그에 더 가중치를 주었기 때문으로 추정된다. 추후 연구에서는 이 두 지표까지도 함께 수치를 향상시킬 수 있는 함수를 설계하도록 노력해야 할 것이다.

위의 결과를 통합해 보았을 때, 전체적으로 통합 함수 2에서 성능이 개선되었다고 판단했다. 통합 함수 2의 경우, 여러 지표에서 평균적으로 향상된 결과를 보였고, 수익은 다소 개선되었으며, 충전량도 증가한 모습을 볼 수 있었다. 이 때, 여러 알고리즘 중에서도 SAC, TD3 알고리즘이 가장 균형 잡힌 개선을 보였다고 생각하는데, SAC는 총 이익, 총 충전 에너지, 추적 오차, 배터리 열화 측면에서 다른 알고리즘과 비교해 높은 성과를 보였다. 이는 SAC가 정책 및 가치 네트워크를 동시에 학습해 정책의 최적화와 안정적인 학습이 가능하고, off-policy 학습 방법을 사용함으로써 다양한 상태에서 학습을 강화하고, 안정적인 수렴을 했기 때문이라고 생각된다. TD3 알고리즘 또한 사용자 만족도, 총 이익 측면에서 좋은 성과를 보였는데, 특히 과대평가 편향 완화를 통해 안정적인 정책 최적화가 가능해 사용자 만족도에서 효과적이었다고 추정된다.

종합적으로 볼 때, 이번 분석에서는 SAC와 TD3이 가장 균형 잡힌 성능 개선을 보여 두 알고리즘이 전력 관리 시스템의 성능을 향상하는 데 유망한 선택지임을 확인했고, 이후 연구에서는 유망했던 선택지들을 더욱 중점적으로 알고리즘을 발전시켜보고자 했다.

3.4. 전기차 충전 요금을 고려한 충방전 알고리즘 설계

3.4.1. step 알고리즘

전기차 충전 및 방전 시, 충전 요금을 고려하여 사용자에게 최대한의 이익을 제공할 수 있도록 RL 학습의 step 알고리즘을 설계하였다.

$$\begin{aligned} \text{normalized actions}[i] &= \frac{\text{action}[i]}{\Sigma \text{actions}} \text{ if } \Sigma \text{actions} > 1 \\ \text{normalized actions}[i] &= \frac{-\text{action}[i]}{\Sigma \text{actions}} \text{ if } \Sigma \text{actions} < -1 \end{aligned}$$

충전 action의 총합이 1보다 크거나 방전 action의 총합이 -1보다 작은 경우, 모든 action을 정규화하여 1 또는 -1이 되도록 조정한다.

action 값이 양수인 경우, EV의 충전을 시도한다.

$$\begin{aligned} \text{amps} &= \text{action} \times \text{max charge current} \\ \text{profit} &+= \text{abs(actual energy)} \times \text{charge price} \end{aligned}$$

충전 전류(amps)는 [-1, 1] 사이의 비율로 표현되는 action 값과 최대 충전 전류의 곱으로 두어, 상황에 맞게 충전 속도를 유연하게 조절할 수 있도록 하였다.

EV가 충전하는 에너지량(actual energy), 현재 충전 가격 정보를 통해 이익을 계산한다.

action 값이 음수인 경우에는 EV의 방전을 시도한다.

$$amps = action \times max\ discharge\ current$$

$$profit += abs(actual\ energy) \times discharge\ price$$

방전 전류는 action 값과 최대 방전 전류의 곱의 수치로 이용하고, 방전된 에너지량과 현재 방전 가격 정보를 통해 사용자의 이익에 반영하도록 하였다.

$$current\ power\ output += actual\ energy \times \frac{60}{timescale}$$

$$currnet\ total\ amps += actual\ amps$$

EV의 충전 및 방전 시 소비 ,생성한 에너지의 양을 기록함으로써 에너지 사용량을 추적하고, 이를 기반으로 수익 및 손실을 계산할 수 있다. 또한 전력 출력의 변화를 업데이트하면서, 현재 전력 수요와 공급 상황을 파악하여 안정적인 전력 공급이 가능하도록 한다. 충전 및 방전을 마친 EV의 사용자 만족도까지 반영함으로써, 충전소 운영에 긍정적인 영향을 주도록 설계하였다.

3.4.2. 결과 분석



그림 21. 통합2 함수와 step 수정 알고리즘 결과 그래프

기존 통합2 알고리즘 결과와 유연한 충방전을 위해 수정한 step 알고리즘으로 학습된 통합2 알고리즘 결과를 그림 21과 같이 비교해보았다. 통합 2에서 가장 균형 잡힌 성능을 보인 것으로 판단했던 SAC, TD3와 더불어 total_profits 지표에서 다소 낮은 결과를 보인 A2C, PPO, TQC를 포함해 총 5가지 RL 알고리즘을 중심으로 학습 및 평가해보았다. 기존 통합2 알고리즘은 방전량에 비해 충전량이 상대적으로 많아 이익 측면에서 음수의 값을 보였으나, step 알고리즘 수정을 통해 유연한 충전 및 방전이 가능해짐에 따라 사용자가 얻는 이익이 증가했음을 알 수 있다. 다만, average_user_satisfaction 그래프에서도 확인 가능하듯 방전량이 상대적으로 증가하하면서 사용자의 만족도는 다소 낮은 결과를 보였다. power_tracker_violation 지표는 step 알고리즘 수정 이후 상당 수준 증가한 것을 볼 수

있는데, 이는 충전 및 방전량 사이의 균형이 불안정함을 나타낸다. tracking_error 지표 또한 부정적인 결과를 보이는데 step 함수의 action을 충전 및 방전 가격 조건에 따라 결정하게 함으로써, 실제 전압 및 전류 요구와 일치하지 않아 다소 낮은 결과를 보인 것으로 판단된다. 향후, 해당 알고리즘을 긍정적으로 활용하기 위해서는 충방전량 사이의 균형을 맞추는 과정을 통해 추적 오류를 줄이고 사용자 만족도를 높일 수 있는 방향으로 개선이 필요하다.

EV Charging Station Selection

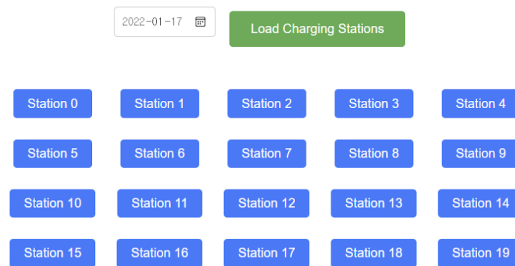


그림 22. 시각화 웹 구축

이후, 시뮬레이션 결과를 기반으로 EV 사용자에게 현재 충전 및 방전 상태, 이익에 대한 정보를 제공하기 위해 flask를 활용하여 시각화할 수 있는 웹을 구축하였다. 총 20개의 충전소에 대해 9가지 알고리즘에 대해 평가하여, 사용자가 이용 일자와 충전소를 선택하면 아래 그림과 같이 시뮬레이션 결과 중 가장 이익이 높은 충방전 상태를 확인할 수 있다.

EV Charging Station - 1



그림 23. 시뮬레이션 결과 확인

4. 연구 결과 분석 및 평가

본 연구에서는 EV2Gym 환경을 활용하여 V2G 이해관계자(전력망 운영자, 충전소 사업자, EV 사용자)의 입장에서 최대한의 이익을 요하는 알고리즘을 개발하여 학습 및 평가하고, 이해관계자의 입장을 통합함으로써 상호 이익을 극대화할 수 있는 모델을 구축하였다. 다만 EV2Gym 환경 내에서는 기존에 있는 자료들을 바탕으로 학습, 평가 및 시뮬레이션이 가능해 실시간으로 일어나는 변화에 대해서는 결과를 확인할 수 없다는 점이 큰 제약이었다.

전력망 운영자, 충전소 사업자, EV 사용자 3가지 V2G 이해관계자 입장을 고려하여 보상 함수 및 상태 함수를 설계하였고, 각각 설정한 주요 지표에서 대부분 성능이 개선된 모습을 확인하였으나, 모든 지표에서 동시에 성능 개선이 이루어지지 않았다는 한계점이 존재했다. 연구 조건 및 결과를 간단히 표로 정리해보면 아래와 같다.

이해관계자	알고리즘	관심 지표	보상함수 요약	상태함수 요약
전력망 운영자	SAC, TD3	1. Power Tracker Violation 2. Tracking Error 3. Total Transformer Overload 4. Total Profits	수익 극대화, 변압기 과부하 및 추적 오차 최소화	시간, 전력 설정 사용량, 과부하량 포함
전기차 사용자	A2C, PPO, RPPO, TPRO	1. Average User Satisfaction 2. Total Energy Charged 3. Battery Degradation 4. Total Reward	비용 최소화, 만족도 및 배터리 열화 반영	시간, 전력 사용량, 배터리 상태 포함
충전서비스 사업자	DDPG, TQC	1. Total Profits 2. Total EV Served 3. Battery Degradation	수익 극대화, 만족도 및 배터리 열화 고려	전력 사용량, 배터리 상태, 서비스 EV 수 포함

표 1. 이해관계자에게 유리한 V2G 충방전 보상/상태 함수 연구 요약

이해관계자별 보상, 상태 함수 설계 내용을 바탕으로, 이후 이를 종합하여 하나의 보상/상태함수를 설계하는 활동을 진행하였다.

함수	특징	결과 요약
통합 함수 1	1. 균형 중시 2. 사용자 만족도, 배터리 열화, 전력 효율성 반영	대부분의 지표에서 개선이 이루어졌으나, PPO는 충전량 개선에 민감하지 않음.
통합 함수 2	1. 각 이해관계자 목표 비중 강화 2. 과부하 및 전력 추적에 높은 페널티	모든 지표에서 개선, 특히 SAC와 TD3은 충전량과 사용자 만족도에서 뛰어난 성과를 보임.

통합 함수 3	1. 전력망 안정성 강화 2. 사용자 경험 및 배터리 관리 중점 반영	배터리 건강 상태에 강한 페널티를 부여해 장기적인 배터리 관리 효과 가능성을 보였으나 충전량 성과는 다소 제한적이었음.
---------	---	--

표 2. 통합된 V2G 충방전 보상/상태 함수 연구 요약

위 표에 정리된 바와 같이 통합 함수 2에서 여러 지표가 평균적으로 개선되었으며, 여러 알고리즘 중에서도 SAC와 TD3 가 사용자 만족도, 에너지 충전량 측면에서 가장 눈에 띄게 향상된 것을 볼 수 있었다. SAC, TD3에 비해 높은 수치는 아니지만 기존 연구에 비해 A2C 또한 개선된 결과를 보였다. A2C는 충전량에 비해 방전량이 상대적으로 낮는데, 사용자 만족도 지표가 높고 tracking error 또한 낮은 결과를 보여 안정적인 학습을 하고 있다 판단하였다.

SAC, TD3, A2C의 결과를 전반적으로 비교해 보았을 때 충전량 대비 방전량의 수치가 낮은 경우, 사용자 만족도에서 높은 결과를 보임을 알 수 있다. 이를 통해 충전량과 방전량의 균형을 맞추고 사용자 만족도를 높임과 동시에 전력망에 미치는 영향은 최소화 하는 연구가 필요할 것으로 보인다.

통합함수에서 결과가 우수했던 알고리즘을 바탕으로 step 함수를 수정해봄으로써 보완을 시도했다. 하지만 결과적으로 수정된 step 알고리즘은 충전 및 방전의 유연성을 통해 사용자 이익의 경우 증가되었지만, 방전량이 증가하면서 사용자 만족도와 전력 추적 지표에서는 다소 낮은 성능을 보였다. 이를 개선하기 위해서는 충전과 방전량 간의 균형을 맞추는 추가적인 조정을 통한 추후 연구가 더 필요할 것으로 생각된다.

이후 위에서 얻은 결과를 시각화할수 있는 방안에 대해 고민했으며, flask 프레임워크를 활용해 각 충전소별 충방전 상황을 확인할 수 있는 시뮬레이션 웹 페이지를 구축해보았다.

5. 결론 및 향후 연구 방향

전력망 운영자, 전기차 사용자, 충전 서비스 사업자, V2G의 세 이해관계자에 대해 각자의 입장에서 유리한 지표를 기준으로 알고리즘을 설계한 후, 이해관계를 최적화하기 위해 3가지 알고리즘으로 통합하여 성능을 비교 분석해 보았다. 이 과정에서 중요한 지표를 설정해두고 연구를 진행하였는데, 모든 지표가 동시에 개선되지는 못하였다는 한계점이

존재했다. 특정 지표의 최적화가 다른 지표에 부정적인 영향을 미친 경우라고 생각되며, 특히 통합 알고리즘에서는 total_profits 지표가 하락하는 한계가 있었다. 이는 각 이해관계자의 이익을 동시에 최적화하는 데 있어 균형을 맞추는 것이 어려움을 시사하며, 향후 추가적인 조정과 개선을 통해 각자의 이익을 극대화할 수 있는 또 다른 통합 방안을 모색할 필요가 있다.

또한 EV2Gym 환경 내에서는 기존 자료를 바탕으로 학습 및 시뮬레이션 결과를 도출하여, 실시간 자료에 대해서는 평가 및 학습이 불가하다는 한계가 있었다. 전력망 안정성이 최적화됨에 따라 평균적으로 전력망 투자 비용의 약 3~5% 절감 효과가 있는 바 [7], 실시간 데이터 처리 기능을 강화하여 변화에 즉각적인 대응이 가능하다면 향상된 전력망 안정성과 더불어 경제적 효과 또한 극대화될 것으로 기대된다.

6. 참고 문헌

- [1] KIBO. (2022, Jan. 31). 기술 동향 리포트 [Online]. Available: https://tb.kibo.or.kr/ktbs/board/tech-trend/tech_trend.do?mode=download&articleNo=569&attachNo=1447 (downloaded 2024, Aug. 23)
- [2] 한국자동차환경협회. 전기차 보급 현황 [Online]. Available: <https://ev.or.kr/nportal/evcarInfo/initEvcarSupplyPurposeAction.do>
- [3] Amry, Youssef, et al. "Electric Vehicle Traction Drives and Charging Station Power Electronics: Current Status and Challenges." *Energies*, vol. 15, no. 16, Aug. 2022
- [4] Stavros Orf. (2023). EV2Gym. [Online]. Available: <https://github.com/StavrosOrf/EV2Gym/tree/main>
- [5] Stavros Orfanoudakis, Cesar Diaz-Londono, Yunus E. Yilmaz, Peter Palensky, Pedro P. Vergara. (2024). EV2Gym: A Flexible V2G Simulator for EV Smart Charging Research and Benchmarking. [Online]. Available: <https://arxiv.org/abs/2404.01849> (downloaded 2024, June. 27)
- [6] Redondo-Iglesias, Eduardo, et al. "Efficiency Degradation Model of Lithium-Ion Batteries for Electric Vehicles" *IEEE Transactions on Industry Applications*, vol. 55, no. 2, Mar.-Apr. 2019
- [7] Y. Deng, Y. Zou, X. Luo, Z. Li, Y. Wang, X. Lei, and L. Wen, "Correlation Analysis between Power Grid Investment Cost and Load Curve Characteristics," 2022 4th International Conference on Smart Power & Internet Energy Systems (SPIES) Smart Power & Internet Energy Systems (SPIES), pp. 1696-1700, Dec. 2022