소수자 게임의 집단 효용 극대화: 강화학습의 보상 균형화 기반 신속 안정화 정책

백승호, 이상철, 조한얼, 김찬수*

과학기술연합대학원대학교, 세종대학교, *한국과학기술연구원

092236@kist.re.kr, eau@ust.ac.kr

Maximizing Collective Utility in the Minority Game: A Reward-Balancing Based Rapid Stabilization Policy for Reinforcement Learning

Seungho Baek, Haneol Cho, Chansoo Kim*

UST (University of Science and Technology), Sejong University,

*KIST (Korea Institute of Sci. and Technology)

요 약

본 연구는 경제학의 소수자 게임(minority game)의 대표적 사례인 El Farol Bar 문제를 강화학습(Q-learning) 기반으로 다룬다. 기존 접근은 집단적 참석자(attendance)가 수용량(capacity) 주변에서 불안정하게 진동하는 특성을 보였으며, 이는 사회적 낭비와 효율성 저하를 초래한다. 본 연구는 이러한 불안정성을 줄이고 집단적 효율성을 최대화하기 위해, 에이전트 특성을 반영한 보상 균형화(reward balancing) 방법을 제안한다. 제안 방법은 에이전트별 성향(초기 편향, 결정성, 관성)에 따라 유리/불리 유형을 구분하고, 이에 기반해 차등적인 보상 조정과 재분배를 수행한다. 시뮬레이션 결과, Q-learning에 보상 균형화 알고리즘을 더하여, attendance의 수렴 속도가 증가하고 자원 낭비(수용 인원 초과)가 감소함을 확인하였다. 이는 다중 에이전트 시스템에서 에이전트의 이질성을 고려한 정교한 보상 설계가 집단적 조정을 달성하는 데 매우 효과적임을 보인다.

Ⅰ. 서 론

엘 파롤 바(El Farol Bar) 문제는 소수자 게임(Minority Game)의 대표적 사례로, 제한된 자원을 두고 다수의 주체가 경쟁할 때 나타나는 집단적 의사결정 메커니즘을 설명한다. 경제학, 통계물리학, 복잡계 과학에서 집단 행동을 모사하는 전형적 모델로 다뤄져 왔으며, 개별 에이전트가 효용을 극대화하려는 과정이 오히려 집단적 비효율을 초래한다는 점을 보여준다. 이 문제를 분석하기 위해 Q-러닝과 같은 강화학습 알고리즘이 널리 활용되어 왔다. 에이전트들은 과거 경험으로부터 보상을 학습해 합리적 정책을 수립하지만, 전통적 접근에서는 총 참석자 수가 수용치(capacity)를 중심으로 크게 진동하며 예측 불가능한 변동성을 보인다. 이는 교통 체증, 전력망 과부하, 금융 시장 변동성처럼 현실 시스템의 비효율·사회적 비용과 직결된다.

그러나 기존 연구들은 대부분 모든 에이전트가 동일한 보상 구조와 학습 규칙을 따른다고 가정해, 현실의 다양한 성향(선호, 위험 감수, 학습 속도 등)을 충분히 반영하지 못한다. 이는 일부 에이전트가 국소 최적 전략에 고착되고, 전체 시스템의 불안정성이 완화되지 않는 한계로 작용한다.

본 연구는 표준 Q-러닝을 기준선으로 두고, 에이전트의 다양성을 반영한 보상 균형화 알고리즘을 제안한다. 개별 특성과 성과에 따라 보상을 동적으로 조정함으로써, 집단의 빠른 안정화와 사회적 비용 절감을 동시에 달성하는 가능성을 탐구한다.

Ⅱ. 본론

2.1 시뮬레이션 환경 및 행동(액션)

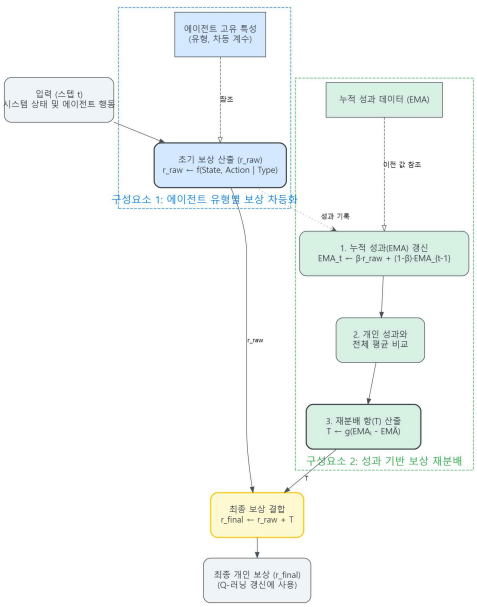

그림 1 보상 균형화 강화학습 알고리즘의 순서도

본 연구는 표준 보상 구조를 사용하는 기준 Q-러닝 모델과 보상 균형화가 적용된 Q-러닝 모델의 성능을 비교한다. 두 모델은 상태 공간, 행동 공간, Q-값 갱신 규칙 등 기본 프레임워크는 동일하지만 보상 설계에서 차이를 가진다. 기준 모델은 소수자 게임에 적용되는 전형적 Q-러닝 방식으

로, 에이전트는 최근 m 스텝의 혼잡도를 관찰해 과밀이 아닐 때 참석하거나 과밀일 때 불참하면 +1, 반대로 행동하면 -1의 보상을 받는다. 이 보상은 모든 에이전트에게 동일하게 주어진다.

이와 대조적으로, 본 논문이 제안하는 핵심 모델은 기준 Q-러닝 프레임워크 위에 '보상 균형화' 메커니즘을 통합한다. 그림 1에 나타난 바와 같이, 먼저 입력 단계에서 각 스텝의 시스템 상태와 에이전트 행동이 관찰된다. 이어서 구성요소 1: 유형 기반 보상 차등화에서 에이전트 고유 특성(행동 편향, 관성 등)을 참조하여 초기 보상 r_{raw} 이 산출된다. 이때 혼잡을 유발할 가능성이 큰 그룹에는 과밀 상황에서 더 큰 페널티를, 불리한 그룹에는 상대적으로 완화된 보상이 부여된다. 다음으로 구성요소 2: 성과 기반 보상 재분배에서는 각 에이전트의 누적 성과가 지수이동평균(EMA)으로 갱신되고, 이를 집단 평균과 비교해 성과 편차를 계산한다. 이 편차에 기반하여 재분배 항 T가 산출되며, 이는 성과가 과도하게 높은 에이전트의 보상 일부를 평균 이하 성과자에게 이전하는 역할을 한다. 마지막으로 두 신호가 결합되어 최종 보상 $r_{final} = r_{raw} + T$ 가 결정되고, 이는 Q-값 갱신에 사용된다. 이러한 방법을 통해 에이전트 간 이질성을 수용하면서도 집단 전체의 안정적 균형을 위한 보상 체계를 구현한다.

2.2 시뮬레이션 결과

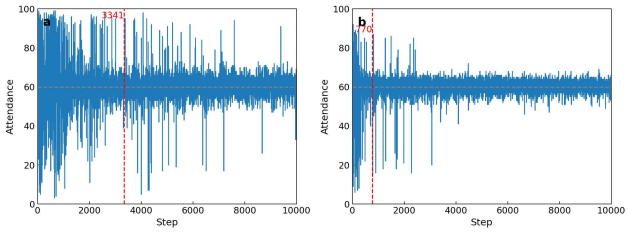

그림 2 Q-Learning 베이스와 보상 균형화 알고리즘이 추가된 모델의 비교

그림 2의 시계열 데이터는 두 모델의 동태적 성능 차이를 명확히 보여준다. 기준 Q-러닝 모델은 획일적인 보상 구조로 인해 에이전트들의 진동을 효과적으로 억제하지 못하였다. 그 결과, 시스템은 수용 인원을(60 명) 중심으로 큰 진폭의 진동을 지속하며 안정화되기까지 3341 스텝이라는 시간이 소요되었다.

반면, 개인화된 보상 균형화 모델은 시스템의 변동성을 현저히 감소시켜 수렴 시간을 770 스텝으로 단축했다. 이는 수렴에 필요한 시간을 77% 개선한 것으로, 제안된 메커니즘의 효율성을 입증한다. 이러한 결과는 유형별 보상 차등화와 성과 기반 재분배를 통해 에이전트들의 행동을 조율하고, 비효율적인 탐색 과정을 최소화하여 시스템을 안정화할 수 있었다. 결과적으로, 본 연구에서 제안한 보상 구조가 시스템의 불안정성을 효과적으로 제어하고 빠르게 수렴함을 실험적으로 확인하였다.

Ⅲ. 결론

본 연구는 엘 파롤 문제에서 표준 Q-러닝의 성능을 개선하기 위해 '보상 균형화 강화학습'을 제안했으며, 시스템의 수렴 시간을 77% 단축하는 효과를 확인했다. 이는 집단적 효율성을 높이는 데 보상 구조의 설계가 학습 알고리즘의 선택만큼 중요하다는 것을 보여준다. 따라서 본 연구가 제시한 '보상 균형화' 전략은, 교통, 에너지, 보건과 같은 소수자 게임을 포함하는 실제 복잡계 시스템에서 더욱 효율적인 전략을 수립할 가능성을 제공한다.

ACKNOWLEDGMENT

This research was funded by the grant Nos. 2023-00262155; 2024-00339583; 2024-00460980; and 2025-02304717 (IITP) funded by the Korea government (the Ministry of Science and ICT).

참 고 문 헌

[1] W. Brian Arthur, "Inductive Reasoning and Bounded Rationality: The El Farol Problem," American Economic Review: Papers and Proceedings, vol. 84, no. 2, pp. 406 - 411, 1994.

[2] Esteban Moro, "The Minority Game: An Introductory Guide," arXiv preprint, 2004.

[3] David Sherrington and Tobias Galla, "The minority game: effects of strategy correlations and timing of adaptation," arXiv preprint, 2003.

[4] Willemien Kets, "The minority game: An economics perspective," arXiv preprint, 2007.

[5] Radu Manuca, Yi Li, Rick Riolo, Robert Savit, "The Structure of Adaptive Competition in Minority Games,"

[6] Si-Ping Zhang, Jia-Qi Dong, Li Liu, Zi-Gang Huang, Liang Huang, Ying Cheng Lai, "Reinforcement learning meets minority game: Toward optimal resource allocation," Physical Review E, vol. 99, no. 3, Article 032302, 2019. DOI: 10.1103/PhysRevE.99.032302

[7] Guozhong Zheng, Weiran Cai, Guanxiao Qi, Jiqiang Zhang, Li Chen, "Optimal Coordination in Minority Game: A Solution from Reinforcement Learning," arXiv preprint, Dec. 2023.

[8] Duncan Whitehead, "The El Farol Bar Problem Revisited: Reinforcement Learning in a Potential Game," University of Edinburgh Technical Paper, 2008.