Journal of Advanced Technology Research Vol.8, No.2, pp.9-13, ISSN: 0974-6269

심층 강화학습 기반 2단계 조세 정책 최적화 시뮬레이션 환경 분석 및 실험

Deep Reinforcement Learning-based 2-level Tax Policy Optimization Simulation Environment Analysis and Experiment

허 주 성(Joo-Seong Heo), 최 요 한(Yo-Han Choi), 한 연 희(Youn-Hee Han)

Advanced Technology Research Center
Korea University of Technology and Education
{chil1207,yowief,yhhan}@koreatech.ac.kr

요 약

AI는 실제로 많은 영역에서 꾸준히 발전되고 있지만, 특히, 경제 분야는 다양한 환경, 변수로 인해 AI를 통해 정책을 실험하고 평가 할 수 있는 방법이 제한적이다. 본 논문에서는 Salesforce 팀의 AI 기반 경제 시뮬레이션 환경인 AI Economist를 활용하여 경제 활동 에이전트와 정부 에이전트 간의 2단계 학습을 통하여 조세 정책 최적화를 위한 심층 강화학습 기반 조세 정책 최적화 시뮬레이션 환경 분석 및 실험을 진행하였다.

키워드: 조세 정책 최적화, 경제 시뮬레이션 분석, 강화학습

Abstract

AI is actually developing steadily in many areas, but in particular, in the economic field, there are limited ways to experiment and evaluate policies through AI due to various environments and variables. In this paper, we used AI Economist, the Salesforce team's AI-based economic simulation environment, to analyze and experiment on a tax policy optimization simulation environment based on deep reinforcement learning for tax policy optimization through two-step learning between economic activity agents and government agents..

Key words: Tax Policy Optimization, Economic Simulation Analaysis, Reinforcement

1. 서 론

전통적인 경제학에서 경제 정책은 정부와 다양한 경제 주체들 간의 상호작용을 고려하여 많은 실험과 공식, 계산 등의 복잡한 절차를 통해 연구, 수립된다. 수많은 경제 활동, 경제 활동 주체들 간의 상호작용 등을 고려해야하기 때문에 강화학습을 통해 경제 정책을 수립한다는 것은 현실적으로 많은 제한사항이 있다.

4차 산업 혁명으로 AI가 사회 전반에 걸쳐 상용화되고 꾸준히 발전하고 있지만, 경제 분야는 여전히 데이터 부족, 다양한 환경, 변수 등으로 AI 적용이 어렵다. 실제 사회 경제적 문제를 해결하기 위해선 경제 주체 간 다양한 환경과 상호작용 요인들을 확인하며 경제 활동 및 정책 수립 과정을 설계하고 테스트해야 하지만, 경제 관련 데이터가 부족하고, 실제 정책을 실험하기 위한 환경 구성이 어렵기 때문이다.

예를 들어, 실제 경제 활동 인구는 개인마다 직업, 숙련도 등이 다르고 그에 따라 받는 소득 또한 다르다. 따라서, 자연스레 소득에 따른 부의 격차가 발생하게 된다. 이때, 정부의 소득 구간별 조세 정책은 일반적으로 소득이 높은 경제 활동 인구에게 세금을 걷어 소득이 낮은 구간의 경제 활동 인구에게 나눠줌으로써 부의 재분배를 통해 불평등 해소에 중요한 역할을 한다. 하지만, 불평등 해소를 위해 무리하게 세율을 올리게 되면 오히려 능력이 높은 즉, 소득 구간이 높은 곳에 있는 경제 활동 인구들이 의욕을 잃고, 생산성이 감소하게 되는 문제가 발생할 수 있으며, 반대로 생산성을 걱정하여 세율을 낮게 책정하면 생산성은 올라가지만 여전히 불평등 격차는 다시 심해지게 된다.

본 논문에서는 이러한 다양한 경제 활동, 정책 수립에 따른 상호 요인 분석 등을 실험하기 위하여 Salesforce 팀에서 개발한 AI Economist 환경을 사용하여 경제 활동 에이전트들을 학습하고 조세 정책을 최적화하기 위한 실험을 진행하였다[1][2].

2. 관련 연구

일반적으로 정부의 조세 정책은 부의 재분배를 통해 불평등 해소에 중요한 역할을 한다. 최적의 조세 정책을 수립하기 위한 핵심 과제는 과도한 세금 징수로 인한 생산성 감소 즉, 노동자들의 일할 동기에 영향을 미칠 수 있어 부의 재분배를 통한 평등과 생산성 사이의 상충관계를 적절히 조율하는 것으로 불평등 해소를 위해 세율을 올리게 되면 생산성이 감소하고 반대로 세율이 낮아지면 생산성은 올라가지만 불평등 격차는 심해지게 된다.

조세 정책을 위해선 경제학에서 고려해야 할 다양한 환경과 이론이 있지만, 기본적인 정책은 소득의 양에 따른 소득 구간별 세금 징수로 이루어진다. 2018년 미국의 소득 구간과 세율은 소득이 높아질수록 점점 증가하는 형태로 고소득자가 세금을 더욱 많이 납부하는 계급별 세율 구조를 나타내는 반면 Saez의 조세 정책은 오히려 높은 소득에 따라 세율이 감소하는 구조를 나타낸다. 이에 따라 개인의 행동 또한 달라지게 된다. 즉, 소득 구간별 납부 해야하는 세금의 양에 따라 생산성이 차이나게 된다.

Salesforce는 고객 관계 관리를 위한 플랫폼 서비스를 제공하는 IT 기업으로, 최근 Salesforce Research 팀은 주요 글로벌 이슈 중 하나인 경제적 불평등 개선을 목적으로 AI 기반 경제 정책 설계에 도움을 줄 수 있는 프레임워크를 개발하였다.

AI Economist는 경제 활동 에이전트와 정부 에이전트를 통해 강화학습 알고리즘을 적용하여 학습한 조세 정책이 실제 경제적 불평등을 개선할 수 있는지 평가하기 위한 시뮬레이션 환경을 제공하며, 실제 AI Economist가 학습한 조세 정책은 기존의 잘 알려진 조세 정책과 비교했을 때 개선된 성능을 보여준다[3]. 또한, 강화학습을 사용하여 동적인 경제에서 경제 모델의 가정이 아닌 관찰된 데이터를 활용하여 주어진 환경에서 불평등과 생산성의 균형을 위한 최적의 조세 정책을 시뮬레이션 할 수 있다[4].

3. 경제 시뮬레이션 환경

AI Ecomist에선 경제 시뮬레이션을 위한 25X25 크기의 2차원의 그리드 공간의 Gather and Build 환경을 제공한다. 환경에서 Resident(경제 활동 에이전트)들은 이동하며 자원인 돌과 나무를 수집하고, 자원을 활용하여 집을 지어 코인을 얻거나, 다른 에이전트 간의 거래를 통해 코인으로 교환을 하는 다양한 경제 활동을 할 수 있다. 또한, Planner(정부 에이전트)는 생산성 향상과 불평등 해소의 균형을 맞추기 위해 소득 구간별 세율을 조정할 수 있다. 초기 Resident들은 서로 다른 스킬 레벨을 가진채 4구역으로 분리된 공간에 위치한다. 맵 임의의 공간에 생성된 돌과 나무를 수집 혹은 거래를 통해 모아 집을 지을 수 있으며, 돌과 나무 하나씩 소모하여 집 한 채를 짓고 보상으로 코인을 받는다. 이 때, 스킬 레벨에 따라 자원의 수집량과 집을 짓고 받는 코인은 다르게 지급된다. 본 논문에서는 그림 1과 같이 AI Economist 환경을 분석하고, 시각화하여 평가하였다.

그림 1. step별 학습 결과(왼쪽부터 step-0, step-300, step-999)

3.1 행동 및 관찰정보

Resident의 행동은 상하좌우 이동을 위한 행동 4개와 집 짓기 1개, 거래를 위한 행동 44개 마지막으로 아무 행동도 하지 않는 것까지 총 50개로 스텝별 하나의 행동을 선택할 수 있다. 자원의 수집은 돌과 나무 위로 Resident가 위치하면 자동으로 수집되며, [표 1]은 Resident의 행동 분류를 나타낸다.

표 1. Resident의 행동 분류

행 동		개 수
이동		4
집 짓기		1
거래	Buy / Sell	2
	Wood / Stone	2
	Coin(0 ~ 10)	11
	거래 행동합계	44 (=2*2*11)
건너뛰기		1
Resident 행동합계		50

Resident의 관찰 정보는 총 4가지를 포함한다. 첫 번째, 맵은 전체 맵에서 각 Resident를 중심으로 11X11 크기만큼의 주변 정보를 관리하며, 수집 가능한 돌과 나무, 집의 개수, 맵 정보 등이 포함된다.

두 번째, 거래소는 자원별 거래 시세와 내역, 거래된 매수, 매도 호가를 관리한다. 세 번째, Resident가 보유하고 있는 코인과 자원의 양을 관리하는 자산정보와 네 번째, 수집, 집 짓기에 사용될 스킬 레벨 정보를 관리하고 있다.

반면, Planner는 7X22개의 행동이 있다. 사회복지를 실현하기 위한 조세 정책을 위한 행동으로 총 7개의 소득 구간과 각 구간별 세율을 위한 22개의 행동으로 나뉜다. (0.05 단위로 0부터 100을 나누기 위한 21개와 건너뛰기 1개) Planner의 관찰 정보는 Resident와 달리 부분이 아닌 전체 맵에 대한 정보를 관리한다. 전체 맵에서 Resident별 스킬 레벨 정보를 제외한 Resident의 위치, 집의 개수, 거래 정보 등을 포함한다.

3.2 보상

Resident의 보상은 한계효용을 적용한 유틸리티 함수를 사용한다. Resident의 1차 목표는 자원 수집, 거래, 집 짓기 등을 통한 자산의 최대화지만, 자산이 늘어갈수록 한계 효용 체감 법칙에 따라 만족감은 감소한다는 경제학 개념을 적용하였다. 따라서, Resident마다 각각 스킬 레벨이 다르고, 레벨에 따라 얻는 자원, 코인의 양 또한 다르기 때문에 노동 대비 얻는 자산의 양을 아래 수식 1과 같이 정의하

여 단순히 자산을 최대화 하는 것이 아닌 한계 효용이 적용된 노동대비 자산의 최대화를 목표로 한다.

$$u_i(x_{i,t}, l_{i,t}) = crra(x^c_{i,t}) - l_{i,t}, \quad crra(z) = \frac{z^{1-\eta}-1}{1-\eta}, \quad \eta > 0 \qquad (1)$$

위 식에서 i는 Resident, t는 타임스텝, x는 자산, l은 노동을 의미하며, crra는 Constant Relative Risk Aversion로 경제학에서의 효용함수를 나타낸다.

Planner의 보상은 사회복지 함수를 사용한다. 사회복지 지수는 다양한 것이 있지만 본 논문에서는 서로 상충관계에 있는 생산성과 불평등 지수를 활용하며, 사회 복지 함수를 아래 수식 2와 같이 적용하였다.

$$swf_t = Eq(coin_t) \cdot \mathrm{Prod}(coin_t) \qquad (2)$$
$$Eq(coin_t) = 1 - gini(x^c), \quad \mathrm{Prod}(coin_{i,t}) = \sum_{i=1}^{N} coin_{i,t}$$

위 식에서 Eq는 평등지수로 자산에 대한 지니 계수를 사용하며, Prod는 생산성으로 모든 Resident의 코인의 양을 사용한다.

4. 실험 및 평가

본 논문에서는 다양한 에이전트들과 경제 활동에 대한 상호작용을 테스트하기 위해 2단계 학습을 진행하였다. 1단계 학습은 Resident만 학습하는 것으로 정부의 별도 조세 정책이 없는 시나리오에서 Resident들이 이동, 자원 수집, 거래, 집 짓기 등의 행동으로 개인 보상을 최대화하는 것을 목표로 학습한다.

학습에 사용한 강화학습 알고리즘은 데이터를 효과적으로 사용하고, 여러 번의 업데이트를 위해 PPO(Proximal Policy Optimization Algorithm)를 사용하였다[8]. PPO 알고리즘은 2017년 OpenAI에서 개발된 강화학습 기법으로 대표적인 정책 경사 기반의 강화학습 기법이다. 기존 정책 경사 기반 기법들의 학습 불안정성, 속도 및 성능 저하 등을 개선하기 위하여 정책 갱신 전후의 비율을 클리핑하여 정책 갱신의 양을 간접적으로 제한하는 방식을 활용한다.

2단계 학습은 1단계에 이어 진행되며, Planner의 조세 정책을 적용하여 세율에 따른 Resident들의 행동 양식 변화와 각 Resident들의 행동에 따른 소득, 불평등 지수를 토대로 사회복지 함수를 최대화하기 위한 Planner의 학습이 함께 진행된다. 본 논문에서는 AI Economist, Saez, US Federal, Free Market 총 4가지 조세 정책에 대해 2단계 학습을 진행하고 비교 분석하였다. 첫 번째는 Free Market으로 조세 정책이 없어 1단계 학습과 차이가 없다. 두 번째는 US Federal 정책으로 2020년 미국의 소득 구간별 세율을 그대로 적용하여 실험하였다. 세 번째는 Saez 조세 정책으로 경제 개념 중 Saez 공식을 통해 계산된 소득 구간별 세율을 적용하였다. 마지막 네 번째는 AI Economist로 세율 학습을 통해 소득 구간별 사회복지를 최대화하기 위해 학습된 모델로 실험을 진행하였다.

1번의 에피소드는 총1000 time-step으로 이루어져 있으며, Planner는 100 time-step마다 한 번씩 세금을 걷게 된다. 실험은 AI Economist의 소스코드를 내재화하여 진행했으며, python 버전 3.7, tensorflow 버전 1.14, ray[rllib] 버전 0.8.4 등의 환경에서 실험을 진행했다. 실험 결과를 모니터링 및 시각화하여 평가하였다.

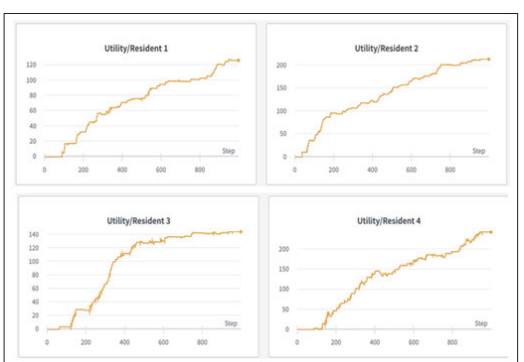

그림 2. Resident별 유틸리티 그래프

그림 2는 Resident의 스킬 레벨별 유틸리티를 나타낸 것으로 x축은 time-step, y축은 유틸리티 값을 나타낸다. 스킬 레벨이 높은 Resident일수록 유틸리티가 높았고, 각 Resident들이 자원 수집, 거래, 집 짓기를 통해 개인의 자산을 최대화하기 위해 학습되었으며 1번, 2번 Resident의 경우 상대적으로 3번 4번 Resident보다 스킬 레벨이 낮아 유틸리티가 낮은데, 집을 거의 짓지 않은 것에 비해 유틸리티가 있는 것은 집을 지을 때의 노동이 양이 상대적으로 거래보다 높기 때문에 집을 짓지 않고 거래를 통하여 자산을 얻은 것을 알 수 있다.

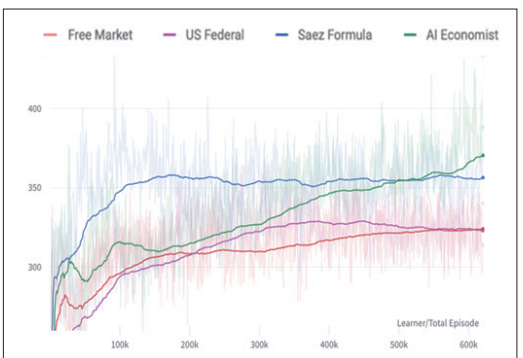

그림 3. 조세 정책별 사회복지 비교 그래프

그림 3은 정책별 사회복지를 비교한 그래프로 60만번 에피소드를 학습했을 때, AI Economist 정책의 사회복지 지수가 가장 높았고, Saez 정책이 두 번째, US Federal과 Free Market이 비슷하게 수렴하는 것을 알 수 있었다. 그림 7은 정책별 생산성과 평등 지수를 비교한 그래프로 Free Market이 세금이 없기 때문에 생산성이 가장 높고 평등 지수가 가장 낮으며, AI Economist 정책이 생산성, 평등 지수가 모두 적절히 높게 학습된 것을 확인하였다.

5. 결 론

본 논문에서는 전통적인 경제 개념과 공식을 적용하여 다양한 경제 주체들 간의 상호작용 속에서 조세 정책을 실험 및 평가하기 위해 AI Economist를 활용하여 심층 강화학습 기반 조세 정책 최적화 시뮬레이션 환경을 분석 및 실험하였다. 경제 활동 에이전트인 Resident와 조세 정책을 통해 사회복지를 실현하기 위한 Planner 에이전트를 2단계로 나누어 학습하였으며, 실험 결과 각 Resident의 스킬 레벨과 정부의 조세 정책에 따라 각 에이전트들이 서로 다른 행동 양식을 보이고, 실제 현실의 경제 환경을 일부 반영할 수 있다는 것을 확인하였다.

또한, 향후 소득 구간, 보상, 경제 활동 등 각종 변수들과 강화학습 알고리즘을 개선하여 다양한 경제 정책을 적용할 수 있도록 개선할 계획이다.

참고 문헌

[1] https://www.salesforceairesearch.com/projects/the-ai-economist
[2] https://github.com/salesforce/ai-economist
[3] S. Zheng et al., “The AI Economist: Improving Equality and Productivity with AI-Driven Tax Policies.” arXiv, 2020.
[4] S. Zheng, A. Trott, S. Srinivasa, D. C. Parkes, and R. Socher, “The AI Economist: Optimal Economic Policy Design via Two-level Deep Reinforcement Learning.” arXiv, 2021.