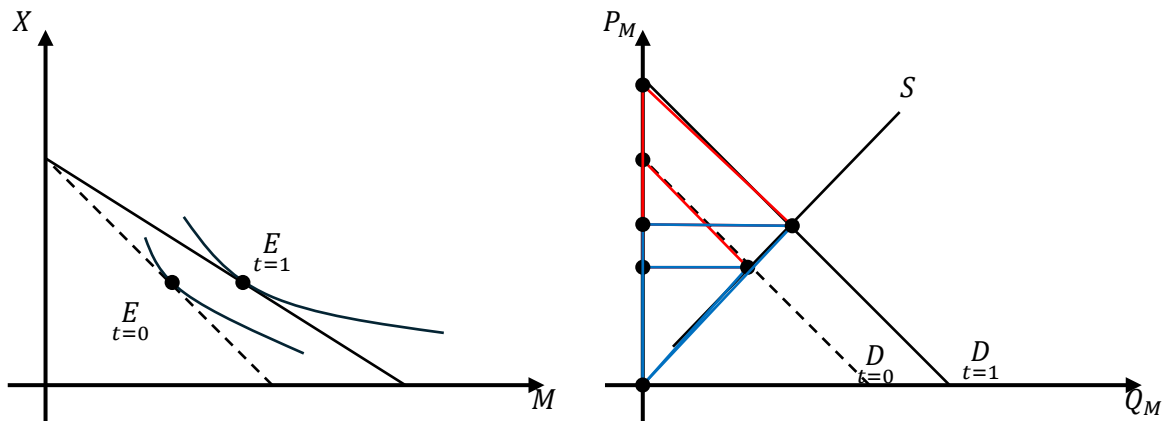


강화학습과 게임이론을 이용해 분석한 과잉의료의 가능성

2020-18537 경제학부 김유로

1. Utility Maximization and Market Equilibrium

의료서비스 이용자(이하 사용자)는 보험사와 보험계약을 체결하여 매월 일정 금액의 보험료를 보험사에 납부한다. 건강경제학 수업시간에 다룬 이후 건강상태에 이상이 발생한 경우 의료행위자를 통해 의료서비스를 소비한다. 이후 사용자는 건강보험료를 보험사에 청구하여 의료지출에 대한 보상을 받는다. 사용자가 보험사로부터 보상을 받는 경우 이용자의 입장에서 의료서비스의 단위 가격 P_M 이 하락하여 예산제약선이 밖으로 이동한다. 예산제약선을 우측으로 이동하고, 이에 따라 의료서비스에 대한 지출의사가 증가한다. 이 경우 사용자(소비자)와 의료행위자(공급자)의 잉여가 모두 증가한다. 이를 수업시간에 다룬 그래프로 나타내면 다음과 같다.



$t = 0$ 는 보험계약을 체결하기 전, $t = 1$ 는 보험계약을 체결한 이후를 나타낸다. 위의 그래프는 보험계약이 존재하는 경우 각 경제주체는 추가적인 의료서비스가 소비할 인센티브가 존재한다는 것을 보여준다.

2. Application of Game Theory and Payoff Structure

효용극대화와 시장균형 모형은 의료서비스가 추가적으로 소비되는 이유를 설명하는데 용이하지만, 개인의 선택이 다른 개인의 선택에 영향을 미치는 경우를 온전히 반영하지 못한다는 문제점이 존재한다. 현실세계에서는 사용자의 선택과 의료행위자의 선택 모두 서로의 효용에 영향을 미칠 수 있다.

가령 사용자는 보험금을 초과로 수취할 목적으로 과잉의료를 받기를 희망할 수 있다. 반대로 의료행위자의 경우 초과 수입을 얻기 위해 과잉의료를 제공하기를 희망할 수 있

다. 그런데 사용자와 의료행위자는 서로가 어떤 성향(type)인지 알 수 없다. 각 참여자(Player)는 행위(Action)를 행한 이후에만 상대방의 성향을 확인할 수 있다. 만약 사용자가 과잉의료를 요구하였는데 의료행위자가 정직한 행위자라면, 의료행위자가 사용자를 고발할 수 있다. 이 경우 정직한 의료인으로서 의료행위자의 사회적 위신 상승, 신뢰성 증가로 인한 홍보효과 등 효용을 얻지만, 사용자는 보험료 인상, 사회적 망신 등의 비효용을 얻는다. 반대로 사용자가 적절한 수준의 의료행위를 원함에도 불구하고 의료행위자가 과잉의료를 제안하는 경우 사용자가 의료행위자를 고발할 수 있다. 이 경우 의료행위자는 사회적 망신과 법적 제재 등의 비효용을 얻게 된다.

이와 같은 인센티브는 게임이론의 보상체계(Payoff Structure)로 표현할 수 있다. 각 Player는 Action을 통해 자신의 Type을 드러낸다. 각 행위의 집합을 통해 얻는 효용을 (*사용자 편익, 의료행위자 편익*)으로 나타내면 다음과 같다. 의료행위자와 달리 사용자는 상대방의 과잉의료를 지적하더라도 얻을 수 있는 편익이 존재하지 않는다. 반대로 의료행위자는 사용자의 과잉의료 요구를 거절하면 큰 편익을 얻는다.

		의료행위자	
		정직한 의료행위	과잉의료 제안
사용자	정직한 의료 요구	(1,1)	(1,0)
	과잉의료 요구	(0,3)	(2,2)

3. Pure Strategy Nash Equilibrium

우선 1회성 게임을 살펴보면 강하게 지배된(Strictly Dominated) 행위(Action)를 제거하여 Pure Nash Equilibrium을 도출할 수 있다.

사용자가 정직한 의료를 요구하는 경우 의료행위자가 정직하나 의료행위를 한다면 1을 얻지만 과잉의료를 제안하는 경우 0을 얻는다. 사용자가 과잉의료를 요구하는 상황에서 정직한 의료를 행한다면 3을 얻지만 과잉의료를 하면 2를 얻는다. 따라서 사용자가 어떤 행동을 선택하든 의료행위자의 입장에서는 정직한 의료 행위를 하는 것이 우위전략이다.

의료행위자가 정직한 의료행위를 한다는 것이 자명하기 때문에 사용자는 의료행위자가 과잉의료를 제안할 경우는 고려하지 않아도 된다. 의료행위자가 정직한 의료행위를 한다면 사용자는 정직한 의료를 요구하는 것이 우위전략이다. 즉, 사용자 역시 정직한 의료를 요구하게 된다.

결국 사용자가 정직한 의료를 요구하고 의료행위자가 정직한 의료행위를 제공하는 것이 순수전략 내쉬균형(Pure Strategy Nash Equilibrium)이고, 해당 균형에서는 다른 행동(action)으로 이탈(deviate)하려는 유인(incentive)가 존재하지 않는다.

4. Repeated Game

총 n 번 게임을 반복하는 경우에도 담합은 발생하지 않는다. 담합이 성립하기 위해선 상대방이 협조하지 않으면 이에 대해 보복을 한다는 전제가 성립해야 한다. 가령 i 번째 게임에서 사용자와 의료행위자가 담합하기 위해선, 만약 의료행위자가 i 번째 게임에서 과잉의료를 제안하지 않으면 $i + 1$ 번째 게임에서 사용자 역시 과잉의료를 선택하지 않을 것이라는 위협이 전제가 되어야 한다. 이 경우 의료행위자는 $i + 1$ 번째 게임에서의 보복을 피하기 위해 i 번째 게임에서 담합을 하게 된다. 그런데 n 번째가 마지막 게임이라는 것이 알려져 있다면 두 Player 모두 보복의 두려움이 없기 때문에 정직한 의료를 요구하게 된다. n 번째 게임에서 정직한 의료가 균형으로 확정된다면, 해당 효과는 $n - 1$ 번째 게임으로도 파급된다. n 번째 게임에서 보복의 두려움이 없기 때문에 $n - 1$ 번째 게임에서도 정직한 의료가 균형으로 확정된다. 이와 같은 연쇄파급은 첫번째 게임으로까지 이어져 담합은 불가능해지고, 첫번째 게임부터 n 번째 게임까지 담합은 발생하지 않는다.

5. Reinforcement Learning

그런데 현실 속에서 의료행위는 1회성이 아닐 뿐더러, 시행횟수가 정해져 있지도 않다. 따라서 담합의 위협이 여전히 존재한다. 이를 위해 강화학습(Reinforcement Learning, 이하 RL)을 이용하고자 한다. 사용된 RL 모형은 Proximal Policy Optimization Algorithm¹에 기반한 Github Repository를 사용하였다.²

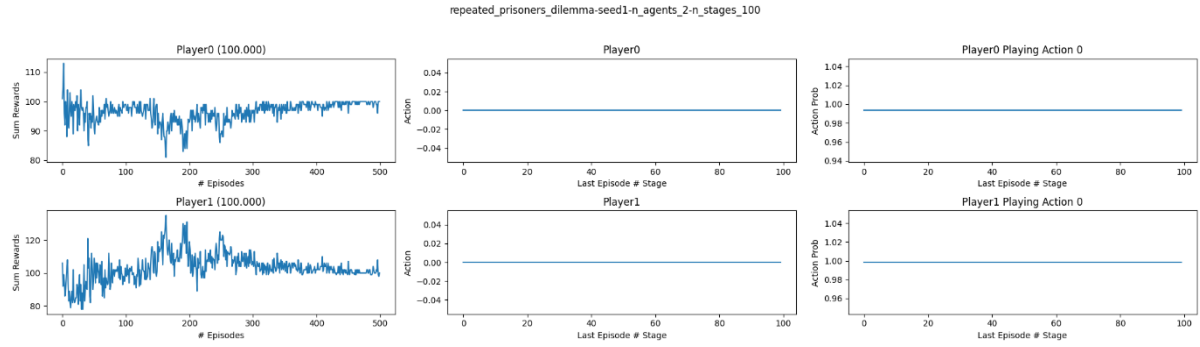
Episode란 각 게임이 종료되기까지의 시나리오를 말한다. 가령 체스게임의 경우, 첫 움직임으로부터 체크메이트(혹은 기권, 무승부)에 이르기까지의 연속된 동작들을 Episode라고 한다. Batch는 학습에 사용되는 Action의 묶음이다. Epoch는 총 학습 사이클 수를 나타낸다. 본 보고서에서는 총 Epoch의 크기를 500, 모형 업데이트 당 사용할 Episode의 크기를 5, 그리고 Batch Size의 크기를 5로 설정하여 훈련을 시행하였다. Episode의 크기가 5라는 것은 게임을 총 5회 반복한다는 것을 의미하고, Batch Size가 5라는 것은 훈련 1회 당 Episode 5개가 사용된다는 것을 의미한다. Model(각 Player의 전략)을 업데이트 하는 데까지 사용되는 게임의 횟수이다.

사용자를 Player0, 의료행위자를 Player1으로 지정한 뒤 총 3번의 서로 다른 Random Seed를 이용해 결과를 도출하였다.

Seed0

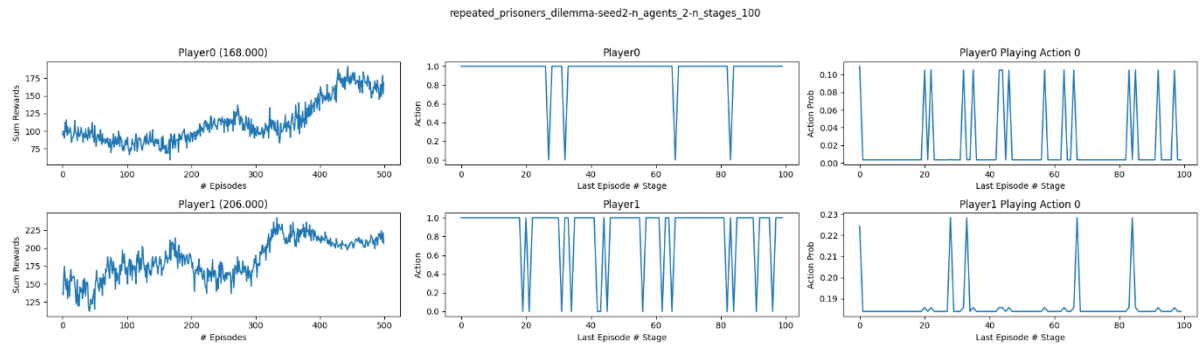
¹ <https://arxiv.org/abs/1707.06347>

² https://github.com/yoonniverse/solving_game_theory_with_reinforcement_learning



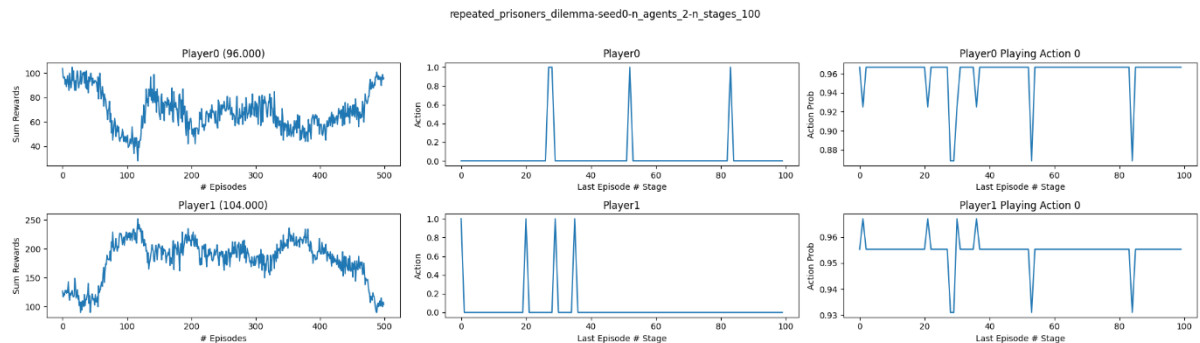
Seed0의 경우 초창기에는 사용자가 정직한 의료행위를 요구하되 의료행위자가 과잉의료를 제안하는 모습이 관측되었다. Episode가 150회 정도에 이르렀을 때에는 반대로 사용자가 과잉의료를 요구하지만 의료행위자가 정직한 태도를 보였으며, Episode 300 이후부터는 대체로 두 Player 모두 정직한 태도를 보였다.

Seed 1



Seed 1에서는 담합의 가능성이 드러난다. 초창기에는 두 Player 모두 정직한 태도를 보이지만, Episode가 진행될수록 과잉의료를 요구하고 과잉의료를 제공하는 모습이 관찰되어 두 Player 모두 Payoff가 200에 가까워졌다. 결과적으로 모든 Seed 중에서 Seed 1일 때 두 Player 모두 가장 많은 Payoff를 갖게 되었다.

Seed 2



Seed 2의 경우 마찬가지로 초창기 두 Player 모두 정직한 태도를 보이고 Episode가 진행될

수록 양측 모두 교대로 과잉의료를 추구하는 모습이 나타나다가, 다시 정직한 태도로 선회하였다.

6. 결론 및 시사점

의료서비스의 보상은 보험회사에 의해 이루어지는데, 보험회사는 의료서비스가 어느정도 수준에서 이루어져야 하는지 정확하게 알지 못한다. 대신 보험회사는 사회적 위신, 보험료 인상, 보험사기 고발 등의 조치를 취할 수 있고, 이러한 행위의 가능성은 사용자와 의료행위자의 Payoff Matrix에 반영되어 도덕적 헤이를 억제하도록 하는 기재로 작용한다. RL의 시사점을 정리하면 다음과 같다.

첫째, Payoff가 제대로 구성되지 않은 경우 반복게임의 상황에서 의료서비스는 과대 공급될 가능성이 존재한다. 만약 과잉의료를 요구하는 것이 사용자와 의료행위자 모두에게 우위전략이라면 1회성 게임이라고 할지라도 과잉의료가 내쉬균형이 될 수 있다. 둘째, 순수게임에서 Payoff가 제대로 구성되었다고 하더라도, Seed1에서 드러났듯 게임이 반복되면 담합(collusion)의 가능성이 존재한다. 셋째, 게임의 반복을 통해 발생할 수 있는 도덕적 헤이를 방지하기 위해서는 추가적인 규제가 필요하다. 가령 동일한 의료기관을 여러 번 이용할 수 없도록 하는 장치는 게임의 참여자를 계속해서 바꿈으로써 담합의 가능성을 배제할 수 있다. 또 한쪽이 과잉의료를 요구한 상황에서 다른 한쪽이 정직한 태도를 요구하였을 때 발생하는 Payoff의 구조를 개편하는 것도 해결방안이 될 수 있다. 위의 모형에서 사용자는 의료행위자의 과잉의료를 고발하더라도 직접적으로 얻을 수 있는 효용이 존재하지 않는다. 그런데 과잉의료 신고 포상제를 도입하여 사용자에게도 인센티브를 제공한다면 반복게임에서의 담합행위는 더욱 발생하기 어려워질 것이다.

7. 부록

사용한 Github Repository에서 보고서 작성을 위해 수정한 부분은 다음과 같다.

```
#envs.py
class RepeatedPrisonersDilemma:

    def __init__(self, **kwargs):
        self.n_agents = kwargs['n_agents']
        self.n_stages = kwargs['n_stages']
        assert self.n_agents == 2
        self.action_space = [('disc', 2), ('disc', 2)]
        self.observation_space = [('disc', 2), ('disc', 2)]
        self.stage = 0

    def step(self, action, turn):
        payoff_matrix_1 = np.array([[1, 1], [0, 2]])
        payoff_matrix_2 = np.array([[1, 0], [3, 2]])
        action_1, action_2 = action
        payoff_1 = payoff_matrix_1[action_1, action_2]
        payoff_2 = payoff_matrix_2[action_1, action_2]
        self.stage += 1
        done = 1 if self.stage >= self.n_stages else 0
```

```
        return np.array(action, dtype=np.float32), [payoff_1, payoff_2], done, -1

def reset(self):
    self.stage = 0
    return np.array([0, 0], dtype=np.float32), -1
```