

〈협력의 진화 - 로버트 액셀로드 지음〉(1984, 2006) 정리

- 조항현 h2jo23@gmail.com

제2부 협력의 창발

- 컴퓨터 대회에서 ‘눈에는 눈, 이에는 이’ 전략(TFT)이 거둔 성공

		경기자 2	
		협력	배반
경기자 1	협력	3,3	0,5
	배반	5,0	1,1

1차 대회

- 14개 프로그램(즉 규칙)이 참가, 각 규칙은 다른 모든 규칙뿐 아니라 자기 자신, 그리고 랜덤 규칙과 200번씩 게임한다.
- 점수는 원래 0점부터 1000점까지 가능하지만 한쪽은 협력만 하고 다른 쪽은 배반만 하는 경우는 없으므로, 가장 좋은 경우는 서로 협력만 하여 얻는 600점이다. 가장 나쁜 경우는 둘 다 배반만 하여 얻는 200점이다.
- 높은 점수를 얻는 규칙 집단(472~504점)과 낮은 점수를 얻는 규칙 집단(최대 401점)을 구분하는 건 먼저 배신하지 않는 신사적(nice) 특성이다.
- 높은 점수를 얻는 규칙 집단 내의 순위는 ‘성과 극대화’ 원칙을 따른 다운닝(Downing)과 같은 킹메이커에 의해 결정되었다. 다운닝은 상대의 행동 패턴으로부터 확률적 예측을 한 후 자신의 전략을 선택한다. 하지만 첫 두 게임을 배반으로 시작하므로 TFT 정도를 제외한 나머지 신사적 규칙들과는 서로 배반하여 낮은 점수에 머무르게 한다.
- 비신사적 규칙들이 부진한 이유는 그들이 대부분 용서(상대가 배반해도 이후에 협력해주는 행동)할 줄 모른다는 것이다. TFT은 단 한 차례의 응징 후 용서함으로써 다시 협력의 조건을 마련한다.
- 결과적으로 TFT이 가장 단순하면서도 가장 높은 점수를 얻어 1등이 되었다.

2차 대회

- 62개 규칙이 참가, 매 게임 후 0.99654의 확률로 게임이 반복되도록 했다.(평균 151번 게임)
- 테스터라는 규칙의 경우, 일단 배반하는데, 상대가 배반으로 맞서면 사과의 뜻으로 협력하고 이후는 TFT을 하며, 상대의 반응이 느긋하면 두 번 더 협력하다가 이후에는 두 번에 한 번 꼴로 배반한다. ‘느긋한 반응’을 하는 규칙들이 테스터에 의해 착취당했다. TFT은 즉각 보복함으로써 테스터로부터 보상을 받을 수 있었다.
- 1차 대회의 교훈을 “신사적이고 관대하라” (교훈 1)로 받아들일 수도 있고, “신사적이고 관대한 상대를 이용하라” (교훈 2)고 받아들일 수도 있다. 교훈 1을 배운 규칙은 교훈 2를 배운 규칙에게 당했고, 교훈 2를 배운 규칙은 종종 결국 응징을 당해서로 낮은 점수에 머물렀다.
- 결국 교훈 2를 배운 규칙으로부터 보상을 받은 TFT가 2차 대회에서도 1등을 차지했다.

TFT이 강건한지 보기 위해 생태학적 모의실험을 했다.(75쪽 그림 2)

- 해링턴(그림 2에서 8번 규칙)은 다른 성공적인 규칙들과 달리 비신사적이었으나 자신이 착취하는 규칙들이 살아 있는 동안 성공적일 수 있었다. 하지만 자신이 착취한 규칙들이 사라지면서 결국 함께 사라진다.

- TFT은 최종적으로 1등으로 남았다. TFT의 강건한 성공은 다음 특성들이 조합된 결과다.
 - 신사적(쓸데없는 문제에 휘말리지 않는다)
 - 보복적(상대의 배반 시도를 더 이상 못하게 한다)
 - 관대함(상호협력을 회복하는데 도움이 된다)
 - 명료함(상대가 이해하기 쉽게 하여 장기적 협력을 이끌어낸다)

어떤 전략의 침범으로부터도 이겨내는 전략은 총체적으로 안정하다고 한다.

- TFT을 침범할 수 있는 조건은?
 - 일단 한 번은 TFT을 배반해야 한다. 하지만 다음 게임에서 보복당하므로 결과적으로 서로 협력할 때보다 낮은 점수를 얻는다.
 - 사실 게임이 반복될 확률 w 가 충분히 크다면 TFT을 침범할 수 있는 전략은 없다. (명제 2) 반대로 w 가 충분히 작다면 협력 관계가 쉽게 깨진다. “역경에 처하면 친구가 적이 된다.” (85쪽)

명제 3. 먼저 협력할 수 있는 임의의 전략은 오직 w 가 충분히 클 때에만 총체적으로 안정할 수 있다.

- 두 경기자 사이에 게임이 충분히 반복된다면, 즉 장기적인 관계를 맺어야 한다면 신사적 전략이 살아남을 수 있다.

명제 4. 신사적 전략이 총체적으로 안정하려면 상대의 최초 배반을 응징해야 한다.

- 상대의 배반을 응징하지 않으면 침범당해서 사라질 수 있다.

명제 5. 무조건 배반전략은 항상 총체적으로 안정하다.

명제 6. 최소의 p 값을 가지고 무리지어 무조건 배반전략을 침범할 수 있는 전략은 TFT처럼 최대 판별력을 가진 전략이다.

- 무조건 배반전략만으로 이루어진 집단에 TFT이 소수라도 침범할 수 있고, 이로부터 협력하는 전략이 살아남을 수 있다.

명제 7. 신사적 전략이 한 개체에 의해 침범당할 수 없다면 그 개체들이 모인 어떤 무리에 의해서도 침범당하지 않는다.

결론: 무조건적인 배반의 세상에서도 협력은 일어날 수 있다. 식별력 있는 개인들이 모인 작은 무리로부터, 작은 규모나마 상호 작용하기만 하면, 협력은 창발된다.(명제 6) 뿐만 아니라 신사적 전략이 마침내 모두에 의해 채택되면 이들은 서로 관대하게 대해서도 괜찮은 여유가 생긴다. 그렇게 함으로써 다른 전략을 쓰는 개체(들)로부터 자신들을 보호할 수 있다.(명제 7) 그러나 신사적 전략이 총체적으로 안정하려면 상대의 배반을 눈감지 말고 반드시 응징해야 한다.(명제 4) 상호협력은 호혜주의를 기반으로 하는 개인들이 모인 무리에서 시작되어 중앙 통제 없이도 이기주의자들의 세상에서 창발될 수 있다.

제3부 우정이나 지능 없이도 가능한 협력

사례 1. 우정 없이 가능한 협력: 1차 대전 참호전에 나타난 공존공영 시스템

- 1차 세계대전 서부전선에서 프랑스와 벨기에 영토의 800킬로미터에 걸친 여러 전선에서 적군끼리 서로 상당히 자제하는 일이 허다했다.
- 극한의 적대관계에 있을 때에도 그 관계가 오래 지속되는 경우 상호협력이 나타날 수 있다. 하지만 전쟁이 끝날 무렵 이러한 협력은 무너졌다.
- 전쟁 초기와 달리 전선이 고착된 후, 완충지의 양쪽에서 같은 시각에 식사가 제공되었다. 날씨가 매우 좋지 않을 때의 휴전으로 상대에게 총을 쏘지 않는 일이 자주 일어났다. 기본 욕구와 활동이 서로 비슷했으므로 병사들은 상대가 무조건 배반하지 않을 것이라고 생각했다.
- 서로 자제하더라도 한쪽의 도발에 대해 다른쪽은 분명히 보복할 것임을 보여주기 위해 서로 애썼다. 보복 능력의 과시 등.

용납되는 수준 이상의 도발에 대해서는 두세 배로 갚아주었다. 전방의 부대가 후방과 교대하면서 협력을 위한 방법들을 그대로 전수했다.

- 양쪽 지휘관들은 이런 묵시적 협력을 중지시키고자 기습 공격 같은 명령을 내리고 이를 확인했다. 이로 인해 공존공영 시스템이 무너졌다.

사례 2. 지능 없이 가능한 협력: 생명계

- 자연에 존재하는 협력을 설명하기 위한 진화 이론으로서 유전자공유 친족이론(혈연선택 이론)과 호혜주의이론 등이 제시되었다.

- 호혜주의이론의 예: 개미야까시나무는 개미에게 집과 식량을 제공하고 개미는 그 보답으로 나무를 보호해준다.

- 생물학적 진화에서도 반복되지 않는 게임이라면 배반이 최선의 전략이다.

- 생물체는 전략을 구사하는데 뇌가 없어도 된다. 박테리아는 게임을 할 수 있다. (1) 박테리아는 화학적 환경에 민감하며 (2) 주변 다른 개체들의 행동에 따라 다르게 반응할 수 있고 (3) 그런 조건부 전략 행동은 유전되며 (4) 박테리아들은 서로의 적합도에 영향을 줄 수 있기 때문이다.

협력의 연대기

- 무조건 배반전략은 태고 상태에서 진화적으로 안정하다.

- 돌연변이 전략들 사이에 혈연관계가 있을 경우, 혈연관계에 있는 상대에게 협력하는 것이 유전자 관점에서 더 높은 적합도(즉 포괄적 적합도)를 줄 수 있다.

- 또는 돌연변이 전략들이 무리지어 나타나 이들 사이의 상호작용이 무시하지 못할 정도의 비율이 되는 것이다. TFT이라면 무조건 배반전략을 침범한 후 총체적으로 안정할 수 있다.

협력의 조건

- 상대의 배반을 응징할 수 있어야 한다. 고등 생물은 다른 개체를 식별하는 인지능력으로 이 문제를 극복하고, 하등 생물은 상호작용하는 개체/집단의 수를 작게 제한함으로써 해결한다.

- 두 개체가 다시 만날 확률, 즉 w 가 충분히 커야 한다. 감염을 통해 아직 다른 숙주로 퍼져갈 기회가 있는 공생체는, 기존 숙주와의 상호작용의 지속가능성이 감소하면 상리공생에서 기생으로 돌변한다.

제4부 죄수의 딜레마 참가자와 개혁가를 위한 조언

죄수의 딜레마 상황에 놓인 사람들을 위한 조언

1. 질투하지 마라.

- 상대방의 성공을 질투하고 그걸 깎아내리기 위해 배반할 수밖에 없는데 그러면 악순환에 빠져 서로 실패할 수밖에 없다. 내가 주의해서 잘하는 한, 상대들이 나와 같거나 조금 높은 점수를 얻도록 내버려두어도 좋다. 오래 반복되는 죄수의 딜레마 게임에서 상대의 성공이 곧 나의 성공을 위한 전제조건이다.

2. 먼저 배반하지 마라.

- 신사적인 전략들끼리 워낙 좋은 성과를 올리기 때문에 신사적인 전략이 모여 있는 집단은 비신사적 전략이 침범하기 가장 어렵다.

3. 협력이든 배반이든 그대로 되갚아라.

- TFT은 상대가 한 차례 배반하면 자기도 정확하게 한 차례만 배반한다. 만일 그보다 더 배반한다면 끝없는 보복의 악순환으로 이어질 수도 있다. 한 차례 미만으로 배반한다면 상대로부터 이용당할 수도 있다.

4. 너무 영악하게 굴지 마라.

- 컴퓨터 대회의 결과 성과 극대화 규칙들은 상호배반의 고리에 갇혀 좋은 성적을 내지 못했다. 이들은 상호 적응을 통한 변화를 고려하는 대신 상대를 이용하려고만 했으며, 이들의 예측은 모두 틀렸다. 그에 비해 TFT는 단순명료하여 상대가 대번에 파악할 수 있고 TFT와 협력하는 게 제일 좋다는 것을 깨닫는다.

개혁가를 위한 조언

1. 현재에 드리우는 미래의 그림자를 확대하라.

- 현재에 비해 미래가 충분히 중요하다면, 즉 보복의 효과가 나타날 만큼 상호작용하는 기간이 충분히 길다면 상호협력은 안정적이다. 상호작용이 보다 많이 일어나게 하고, 미래의 상호작용의 중요성이 크게 하라. 예를 들어 새들의 텃세권, 지역 기반이 튼튼한 기업, 소수 집단으로 제한된 상호작용 등. 계층체계와 조직 역시 상호작용을 집중시키는데 효과적이다. 군비축소나 비무장 조약을 잘게 쪼개어 여러 번 선택하게 함으로써 호혜주의가 보다 효과적으로 나타날 수 있다.

2. 보수 자체를 바꾸어라.

- 보수 자체를 바꾸어 협력이 최적의 전략이 되도록 할 수도 있다. 하지만 그렇게 보수를 큰 폭으로 바꾸지 않아도 상호협력의 장기적인 동기를 배반의 단기적인 동기보다 높게만 하면 된다.

3. 서로에 대한 배려를 가르쳐라.

- 죄수의 딜레마에 빠지더라도 상대를 배려함으로써 협력을 훨씬 쉽게 이끌어낼 수 있다.(혈연선택 가설의 응용?)

4. 호혜주의(TFT)를 가르쳐라.

- 무조건 협력전략은 가장 도덕적이지만 다른 사람에게 이용당할 수 있으며, 상대방을 망치는 경향이 있다. 호혜주의는 남을 착취하는 전략이 살아남기 어렵게 함으로써 자신뿐 아니라 다른 사람들에게도 좋다. 호혜주의가 이기주의보다 낫지만 최선은 아닐 수 있다. 하지만 이보다 나은 대안이 있을까? 호혜주의의 문제 중 하나는 일단 불화가 시작되면 영원히 계속된다는 점이다. 이런 보복의 메아리를 줄이는 방법은 상대의 배반에 9/10만 되갚는 것일지도 모른다.

5. 인식 능력을 높여라.

- 과거에 상호작용했던 상대를 알아보고, 그 상호작용의 특성을 기억하는 능력은 협력을 유지하는데 필수적이다.