반복 죄수의 딜레마 게임에서 역공학기법을 이용한 상대 플레이어 의사결정 모델링

박현수 ○ 김경중

세종대학교 컴퓨터공학과

hspark@sju.ac.kr, kimkj@sejong.ac.kr

Opponent's Player's Decision Modeling in Iterated Prisoner's Dilemma using Reverse Engineering Technique

Hyunsoo Park○ Kyung-Joong Kim

Dept. of Computer Engineering, Sejong Univ.

요 약

사용자의 요구에 적합한 서비스를 제공하는 시스템을 설계하기 위해서는 사용자의 의사결정, 의도, 미래행동 등을 예측할 수 있어야 한다. 이를 위해 대량의 데이터를 이용한 기계학습 기법이 주목 받고 있다. 그러나, 수집할 수 있는 데이터의 양이 제한적이거나, 비용이 발생하는 경우 이런 방법이 적합하지 않다. 본 논문에서는 이러한 문제를 해결하기 위해 Estimation Exploration Algorithm(EEA)을 이용한 방법을 제안한다. EEA를 이용하면 사용자와의 제한된 상호작용(데이터 수집)을 통해서 모델을 생성할 수 있다. 본 논문에서는 반복 죄수의 딜레마 게임에 참여하는 플레이어를 EEA를 통해 모델링하는 실험을 진행했다. 그 결과 제안한 방법의 가능성과 한계를 확인할 수 있었다.

1. 서론

지능형 에이전트가 사용자에게 적합한 서비스를 제공하기 위해서 사용자의 의도 및 생각을 이해하고 그에 적합한 서비스를 제공할 필요성이 있다. 하지만, 상대방의 의도 및 생각은 숨겨져 있으며, 에이전트가 감지할 수 있는 것은 사용자의 외부로 표현되는 정보의 일부분이다. 다행히도, 최근에는 센서 기술의 발달과 대중화로 인하여 다양한 정보를 대량으로 수집할 수 있으며, 이러한 데이터를 이용하면 상대방을 모델링 할 수 있다. 그리고 상대방의 행동을 관측했을 때 이 모델을 이용하여 상대방의 다음 행동을 예측할 수 있다.

하지만, 이러한 작업에는 다량의 관측 데이터가 필요하다. 데이터를 수집하기 위해서는 상대방과 수많은 상호작용을 통해서 데이터를 수집하거나, 상대방의 행동을 무수히 관측함으로써 데이터를 수집해야 한다. 이런 작업은 큰 비용이 필요한 경우가 많다.

본 논문에서는 이러한 결점을 극복하기 위해 역공학 기법 중 하나인 Estimation Exploration Algorithm (EEA)[1]를 이용하여 능동적으로 중요한 데이터를 수집하고, 이를 이용하여 상대방의 모델을 생성하는 방법을 제안한다. 이 알고리즘을 이용하면 중요한 데이터를 우선적으로 수집할 수 있으며, 최소의 상호작용으로 데이터를 수집할 수 있다. 또한 알고리즘의 특성상 다량의 모델을 생성하므로 이를 이용하면 안정적인 결과를 도출할 수 있다.

본 연구는 간단한 게임 중의 하나인 반복 죄수 딜레마(Iterated Prisoner's Dilemma; IPD) 게임을 이용한다. 이 연구에서는 게임의 참가자에 상대방과 나라는 역할을 부여한다. 나는 상대방과 게임을 진행하며, EEA를 이용하여 상대방의 의사결정 과정을 추측한다. EEA를 이용하여 상대방의 내부에 있는 의사결정 모델을 추측할 수 있음을 보이기 위해서 상대방을 일반적인 사용자(인간) 대신에 쉽게 의사결정 모델을 설계 하고 비교할 수 있는 프로그래밍된 플레이어를 이용하여 실험을 진행했다.

게임 초기에는 나는 행위자 상대방에 대한 데이터가 없지만 게임을 진행해 나감에 따라 상대방의 정보를 수집하고 그 행동을 예측해 나갈 수 있다. 그리고 상대방이 아직 보여주지 않은 부분 중 현재 모델을 개선하기에 필요한 데이터를 찾고, 그 데이터를 수집하기 위해 능동적으로 상대방과 상호작용을 한다.

2. 관련 연구

2.1 Estimation Exploration Algorithm

이 알고리즘은 내부를 알 수 없는 시스템의 입력과 출력 만으로 내부 상태를 추측한다. 최근에는 데이터의 부족으로 인해 모델링하기 힘든 시스템을 추론하기 위해 사용되었다[2].

이 알고리즘은 최소의 실험으로 내부를 추측하기 위해 현재 수집한 데이터를 이용해 여러 개의 후보 모델을 생성하고, 후보 모델간의 가장 큰 불일치 점이 가장 현재 모델을 개선하는 데 가장 중요하다고 보고 반복학습을 수행한다. 실험을 수행하여 추가로 얻은 데이터를 학습을 위한 현재 데이터베이스에 추가하여 모델간의 차이가 없을 때까지 위 과정을 반복한다.

2.2 반복 죄수의 딜레마

반복 죄수의 딜레마는 죄수의 딜레마 게임의

변형으로, 플레이어는 상대 플레이어의 이전 결정을 바탕으로 다음 행동에서 배신 또는 협력을 결정할 수 있다. R. M. Axelrod는 이 게임을 이용하여 협력행위가 어떻게 진화 하였는지를 설명하였다[3].

이 게임에서는 상대방이 계속 협력할 것이라는 사실을 알기만 한다면 계속 협력을 선택하는 것이 참가자 모두에게 가장 좋은 선택이지만, 상대방이 언젠가 배신할 것이라면 먼저 배신하는 것이 유리하다. 이 게임의 특징 때문에 상대방의 숨겨진 의도(배신, 협력) 및 미래행동을 사전에 예측하는 것이 중요하다.

3. 제안하는 알고리즘

3.1 알고리즘 개요

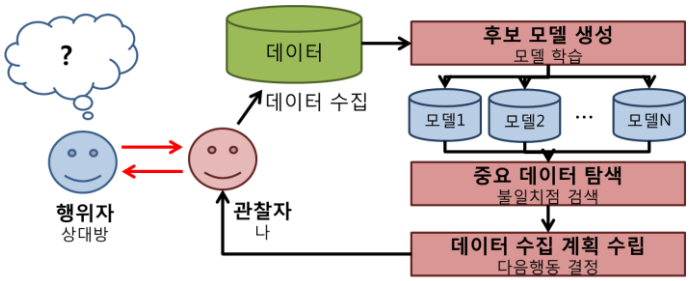

그림 1. 전체 시스템 개요

그림 1은 제안하는 알고리즘의 개요이다. 그림에서 "상대방" 플레이어는 행위자의 역할을 하며, "나" 플레이어는 관찰자의 역할을 한다. 나의 목적은 상대방과 상호작용을 하면서 상대방의 내부에 있는 의사결정 모델을 추정하는 것을 목적으로 한다. 이를 위해 게임을 진행하면서 데이터를 수집하여 상대방의 모델들을 예측하고, 예측된 모델들의 부족한 부분을 보완하기 위해, 모델들을 분석하여 모델을 개선하기에 유용한 데이터를 추론한 다음, 이 데이터를 수집하기 위해 다음 상호작용 때 나의 행동을 결정한다. 이 과정을 통해 나 플레이어는 상대방의 행동을 모델링할 수 있는 중요 데이터를 우선적으로 수집한다. 이 과정을 반복하면 최소의 상호작용으로 데이터를 수집하여 모델을 생성할 수 있다.

3.2 후보 모델 생성

본 연구에서 플레이어의 의사결정 모델은 테이블 형태로 되어있다고 가정하였다. 상대방은 나의 이전행동(M_{-2n}, M_{-n})들과 상대방의 이전 행동(O_{-2n}, O_{-n})들을 고려하여 상대방의 다음 행동(O_n)을 결정한다고 가정하였다. 후보 모델을 생성하는 방법은 진화연산(Genetic Algorithm)이용하였다. 진화연산은 자연계의 진화 기작을 모방한 최적화 알고리즘으로써 다양한 분야에 널리 적용되었다. 본 논문에서는 이를 이용하여 상대방의 의사결정 모델을 추측하였다. 이미 관측된 데이터와 더 잘 맞는 모델일수록 높은 적합도를 가지게 된다.

표 1은 플레이어들이 이전행동만 고려할 경우에 사용할 수 있는 의사결정 모델의 예이다. 여기서 C는 협력(Cooperate)를 뜻하며 D는 배신(Defect)을 의미한다. U는 관측 안됨(Unseen)을 뜻하는데, 이는 게임의 첫 번째 선택이라서 상대 플레이어의 이전 행동/자신의 이전행동이 존재하지 않음을 뜻한다.

표 1. 의사결정 모델의 예

M_{-2n}	M_{-n}	O_{-2n}	O_{-n}	O_n
C	C	C	C	**C**
C	D	C	C	**D**
U	C	U	D	**C**

3.3 능동적 데이터 수집

능동적 데이터 수집이란 현재 모델들을 개선하는데 어떤 데이터가 가장 중요한지 추론하고, 그 데이터를 수집하기 위해 관찰자의 다음 행동을 결정하는 것이다. 이미 수집된 데이터를 이용해서 생성한 모델들은 모든 조건에서 동일한 결론을 내리지 않을 수 있다. EEA에서 말하는 중요한 데이터는 생성된 후보 모델간에 의견차가 가장 큰 데이터를 뜻한다. 예를 들어 표 1에서 조건 CCCC에 대해 모든 모델이 C라는 예측을 한다면 CCCC에 대한 데이터는 충분히 수집되었다고 가정한다. 하지만, CDCC에 대해 모든 후보 모델의 예측이 다르다면 조건 CDCC에 대한 데이터 수집이 부족하다고 가정한다.

능동적 관측이란 모델간의 결과가 가장 차이 나는 지점을 찾는 탐색 문제로 볼 수 있다. N 개의 모델이 있을 때, 조건 i에 대한 일치도(A_i)를 구하고자 한다면, 각 모델(M_k)의 조건 i에 대한 예측 $c_i = M_k(i)$ 가 얼마나 일치하는지를 구해야 한다. 가장 중요한 데이터는 일치도(A_i)가 가장 낮은 조건을 가진 데이터이다.

$$A_i = \left| \sum_{k=1}^{N} v_i \right|, \; c_i = M_k(i), \; v_i = \begin{cases} 1, where\, c_i = C \\ -1, where\, c_i = D \end{cases}$$

가장 중요한 데이터를 추론했다면, 그 데이터를 수집하기 위한 다음 행동을 결정해야 한다. 만약 상대방이 계획적인 행동을 수행하는 개체라면 과거의 상대방의 행동도 미래의 행동에 영향을 준다고 보는 것이 타당하다. 하지만, 나는 상대방의 행동을 직접 조작할 수 없기 때문에, 내가 할 수 있는 행동만을 이용하여 상대방에 대한 데이터를 수집해야 한다. 이를 위해서 나 플레이어는 비록 중요한 데이터지만, 자신이 관측할 수 있는 데이터와 자신의 의사만으로는 관측할 수 없는 데이터를 구분할 수 있어야 한다. 만약 나의 행동에 대한 상대방의 행동으로 데이터를 분류했을 때 데이터가 균일하게 분포하지 않는다면, 이런 경우로 취급할 수 있다. 예를 들어 내 행동 CC에 대한 상대방의 행동이 CC인 경우가 12개, CD가 0개, DC가 0개, DD가 0개라면 나의 행동에 대해 상대방의 행동이

차이가 없는 경우(내 행동 CC에 상대방은 언제나 CC로 대응함)이므로, 더 이상 데이터 수집을 시도할 필요가 없다. 내 특정 행동에 대해서 상대방 대응의 모든 경우의 수의 리스트를 $counts$라고 한다면 위의 경우엔 다음과 같이 수정한다.

$$A'_i = A_i + \{\max(counts) - \min(counts)\}$$

위 과정을 거친 결과 가장 낮은 일치도 (A_i')를 가진 조건이 가장 중요한 데이터라고 볼 수 있고 이에 해당하는 나의 행동이 내가 다음에 할 행동들(Action sequence, eg. CD)이 된다.

3.4 모델 병합

간단한 다수결 방법을 사용하여 생성된 모델들을 하나로 병합하였다. 동일한 조건(나의 행동들, 상대방의 행동들)에 대해 더 많은 모델이 지지하는 결론을 최종 결론으로 인정하였다.

4. 실험 및 결과

4.1 실험조건

표 2. 알고리즘의 중요 파라미터

조건	값
EEA: 후보모델 개수	5
EEA: 상호작용의 개수	10
IPD: 기억의 길이	3
GA: 집단크기	20
GA: 세대	50

표 2는 이 실험에 사용하는 알고리즘의 중요 파라미터이다. 중요한 것은 IPD게임에서 상대방가 얼마나 과거의 정보를 고려하는지 결정하는 것이다. 이에 따라 탐색 범위, 문제의 난이도 및 생성된 모델의 표현력이 결정된다.

상대방 플레이어는 세 종류를 사용했다. a) 반드시 협조만 하는 상대방, b) 배신에 대해서는 반드시 보복하고, 바로 용서하는 상대방, c) 내가 협력적으로 나오면 반드시 배신하는 상대방. b, c의 첫 번째 행동은 무작위로 하도록 했다. a는 나의 반응에 전혀 반응하지 않으며, b와 c는 각각 이전의 행동과 그 이전의 행동까지 고려하여 다음 행동을 결정한다. 이것을 무작위 탐색으로 가능한 모든 경우에 대해 탐색하고자 한다면 최대 92번 탐색이 필요하다.

4.2 실험 결과

그림 2는 모든 가능한 행동들에 대한 실제 행위자의 대응과 관찰자가 추측한 행위자의 의사결정 모델의 대응이 얼마나 동일한지를 보여준다. 길이 3과 10인 가능한 모든 행동(길이 3일 때 8가지, 10일 때 1024가지 경우, eg. CCC, CCD, …, DDD)으로 행위자와 예측모델이 게임을 진행했을 때 얼마나 유사한가를 보여준다. 가로축의 a, b, c는 상대방 플레이어를 의미하며, 3, 10은 테스트용 행동의 길이를 뜻한다. 4, 8, 12, 16은 상대 플레이어와의 상호작용 회수를 의미한다. 92번 탐색을 전부 하는 것 대신에, 최소 4개의 데이터부터 16개의 중요 데이터를 선별하여 수집한 것이다. 총 10번씩 실험을 진행하여 그 평균 정확도이다.

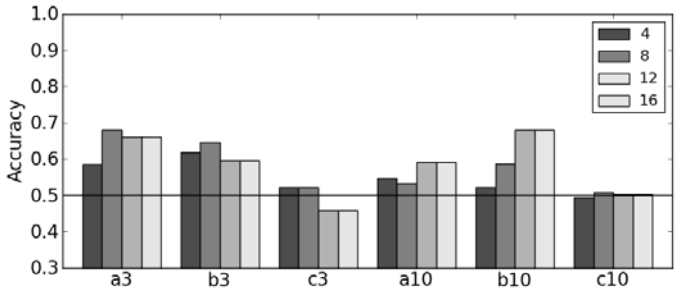

그림 2. 원본 모델과 추측 모델과의 차이

실험 결과 문제가 복잡할수록 더 많은 데이터가 유용하다고 할 수 있으며, 상대방의 의사결정 방식 마다 크게 다른 결과를 보였다. 0.5는 무작위로 예측했을 때와 같은 수준의 성능이므로 c3, c10의 경우에는 예측을 거의 못하고 있다. a3, b3, c3의 경우에는 무작위탐색에 비해 유용하다고 볼 수 없지만, a10, b10, c10의 경우처럼 테스트 범위가 커질수록 그 유용성이 높아질 것으로 추정된다.

5. 결론

본 연구는 역공학 기법 중 하나인 EEA를 이용하여 상대 플레이어의 내부 상태(의사결정과정, 생각, 의사 등) 및 미래 행동을 예측하는 것을 제안하고 반복 죄수의 딜레마 게임을 이용하여 플레이어의 내부 의사결정과정을 모델링하는 실험을 진행하였다. 행위자 플레이어와 관찰자 플레이어로 구분하여 관찰자 플레이어가 행위자 플레이어의 내부 모델을 EEA를 이용하여 추측하였다. 본 연구를 통해 제안한 방법의 가능성을 확인할 수 있었다.

6. 감사의 글

이 논문은 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 뇌과학 원천기술개발사업임 (2010-0018950)

7. 참고문헌

[1] J. C. Bongard and H. Lipson, “Nonlinear System Identification using Coevolution of Models and Tests,” *IEEE Trans. On Evol. Comp.,* vol. 9, no. 4, pp. 361-384, 2005.

[2] H. Lipson and J. C. bongard, “An Exploration-estimation Algorithm for Synthesis and Analysis of Engineering Systems using Minimal Physical Testing,” *In Proc. of the ASME Design Automation Conference (DAC04),* 2004.

[3] R. M. Axelrod, *The Evolution of Cooperation*, NY, Basic Books, 2006.