1>

안녕하세요 저는 1조 포커페이스의 제안서 발표를 맡은 서민정입니다. 저희 조는 블러핑이 가능한 강화학습 기반의 인공지능 포커게임의 구현을 주제로 프로젝트를 진행하고자 합니다.

2>

목차는 다음과 같습니다

3>

-

4>

저희 조는 포커게임 중에서 텍사스 홀덤 포커를 구현하기로 했습니다. 텍사스 홀덤은 개인의 손패와 순차적으로 공개되는 공유 카드의 조합을 고려해 승패의 확률을 계산하는 게임입니다.

저희는 이 텍사스 홀덤 포커에서 이성과 논리를 기반으로 하는 AI플레이어가 감정적 요인에도 영향을 받는 인간 플레이어보다 높은 승률을 가질 것을 예상할 수 있었습니다.

그래서 이번 과제에서 턴 별로 핸드와 커뮤니티 카드를 기반으로 변화하는 상황에서의 승률을 고려해 플레이하는 것뿐만 아니라 블러핑까지도 가능한 AI플레이어의 구현을 목표로 설정했습니다.

5>

이 게임에서는 딜러시스템, AI플레이어, 사용자 플레이어 3가지 결과물이 개발됩니다.

먼저, 딜러시스템은 플레이어 수를 확인하고, 카드를 분배하며, 턴마다 베팅 금액을 확인해, 최종적으로 게임의 승패를 판정하는 등 게임을 진행하는 역할을 합니다. 이 모든 과정은 자동으로 이루어집니다.

그리고 저희는 두 종류의 AI플레이어를 구현할 것인데요, 하나는 강화학습 기반의 AI플레이어이고, 다른 하나는 이 AI를 학습시키기 위해 개발한 교육용 AI플레이어입니다. 강화학습을 할 AI는 교육용 AI를 상대하며 최적의 블러핑 수준을 학습하게 됩니다. 그리고 강화학습된 AI는 인간이 AI의 플레이를 예측할 수 없는 수준의 블러핑을 구사하도록 만드는 것이 목표입니다.

마지막으로 사용자 플레이어는 사람이 직접 플레이할 수 있는 플레이어입니다. 사용자 플레이어는 하나, 혹은 복수의 AI플레이어를 상대하며 게임을 즐기게 됩니다.

6>

-

7>

왼쪽의 시스템 구성도를 보면, 구성요소에는 딜러, AI플레이어와 사람플레이어가 있습니다. 이들은 텍사스 홀덤 룰에 따라 게임을 진행하는데, 딜러와 각 플레이어는 카드 정보, 게임참여여부, 베팅금액, 승패에 대한 정보를 주고받습니다.

예상되는 GUI는 오른쪽 그림과 같습니다. 기본적으로 사용자가 플레이할 때는 사용자 외의 플레이어들의 카드는 공개하지 않도록 합니다. 하지만 강화학습 진행 중에는 학습 확인을 위해 강화학습 AI 플레이어와 교육용 AI 플레이어 카드가 공개됩니다. 게임이 시작되면 테이블의 가운데에 위치한 딜러가 각 플레이어에게 두 장씩 카드를 분배하고 다섯장의 공유패를 뒷면이 보이도록 테이블에 위치시킵니다. 게임은 프리플랍, 플랍, 턴, 리버의 단계로 구성되며, 프리플랍에서는 핸드만 공개하고, 플랍에서는 공유패 세 장이 공개됩니다. 턴과 리버 단계에서 공유패가 각 한 장씩 공개됩니다. 각 플레이어의 옆에는 자신의 잔고, 게임 참여여부가 표시되며, 사용자의 경우 베팅 금액 조절을 위한 버튼이 존재합니다.

8>

기본적으로 사용자와 AI플레이어의 일대일 포커 게임을 진행하는데, 다수의 AI플레이어를 추가하는 기능으로 플레이어의 수를 선택할 수 있습니다. 사용자는 상대 플레이어를 강화학습 이후의 AI와 학습 이전의 AI 중에서 선택해 게임의 난이도를 설정합니다. 그 다음으로 게임의 판수를 지정합니다. 게임이 시작되면 사용자는 카드를 받고, 매 턴마다 상황을 판단해 게임 참여 여부를 결정한 후 베팅 금액을 정해 게임을 진행합니다. 게임은 앞서 설정한 판수를 채우거나 중간에 사용자나 AI플레이어의 잔고가 0이 되면 판수를 채우기 전에 게임이 종료됩니다. 사용자가 게임을 여러 번 실행하더라도 AI플레이어가 정석적인 플레이 패턴을 벗어난 행동을 하기 때문에, 게임마다 다른 사람과 포커를 하는 듯한 재미를 느낄 수 있습니다.

9>

저희 조는 AI플레이어의 학습방법으로 강화학습을 선택했습니다. 강화학습에서 에이전트와 환경이 주어지는데, 에이전트는 정해진 환경 속에서 현재의 상태를 인식하고, 행동을 통해 보상을 최대화하는 방향으로 학습합니다.

보상이 최대화되는 최적의 정책을 찾기 위해서 에이전트에게 주어진 상태에 대한 가치와 에이전트가 할 수 있는 행동에 대한 가치판단이 이루어져야 합니다.

상태가치판단에서는 현재 상태의 좋고 나쁨을 나타내는데, 저희가 구현하려는 텍사스 홀덤에서는 AI 본인이 가지는 핸드의 순위와 팟 오즈라고 판돈과 AI가 베팅할 금액을 비교했을 때 기대할 수 있는 수익률 그리고 공유패, 상대 레이즈 등이 에이전트의 상태가 될 것입니다. 이런 상태에서 에이전트는 정석 플레이와 블러핑 플레이 중에서 선택하게 되고, 선택한 플레이를 기반으로 콜, 레이즈, 폴드의 행동을 하게 됩니다. 이 때 상태가치와 행동가치가 높은 것을 선택하다 보면 누적 보상을 최대화하면서 손익률을 높이려는 저희 조의 목표에 도달하게 됩니다.

10>

저희 조는 강화학습 중에서도 몬테카를로 방법을 활용해 AI를 학습시키기로 했습니다. 몬테카를로 방법은 환경에 대한 완전한 정보가 있지 않은 경우에 대해 알 수 없는 값을 확률적으로 근사하여 구합니다. 따라서 상대의 패와 공개되지 않은 커뮤니티 카드에 대한 정보가 제한된 텍사스 홀덤에 적합하다 판단해 해당 방법을 적용하게 되었습니다.

몬테카를로 방법에서 정책에 따라 행동하는 에이전트가 에피소드 생성, 행동가치 갱신, 최적 행동 추출, 정책 갱신의 단계를 거치면서 최적의 정책을 찾아갑니다. 에피소드 생성단계에서 에이전트는 카드 정보랑 상대 베팅 금액을 입력받아 자신의 베팅금액을 출력합니다. 이런 일련의 과정을 통해 에피소드 결과를 샘플링하게 됩니다. 이런 샘플링을 평균하여 행동가치를 갱신하고 이를 기반으로 기대수익률을 높이는 최적 행동들을 추출해 현재 상태, 에피소드에 대해 가장 큰 가치를 지니는 행동을 선택해 정책을 갱신합니다. 최적인 행동만을 하다보면 정석적인 플레이로 굳어지게 될 수 있습니다. 이를 막기 위해 ε-탐욕정책을 도입합니다. 탐욕정책은 최적 행동을 최대화하려는 정책이며 ε 값을 조절해 탐욕정책을 선택할 확률을 조정합니다.

이를 통해 핸드에 따른 승률과 블러핑 학습을 통한 수익률 개선에 대해 학습할 수 있습니다.

11>

최우선으로 자동화된 딜러와 AI플레이어를 통해 게임을 진행할 환경을 구현해야 합니다. 이것을 기반으로 프로젝트의 연구와 개발을 확대해나갈 수 있기 때문입니다. 그 다음으로 몬테카를로 방법을 통해 AI플레이어를 학습시키는 것이 중요합니다. 이때 AI는 핸드에 따라 승률을 예측 방법, AI의 플레이 패턴을 숨기는 블러핑을 학습합니다. 이를 통해 앞서 개발한 AI플레이어보다 나은 손익률을 가지는 AI를 개발할 수 있습니다. 그리고 앞의 과정으로부터 얻은 실험 데이터를 통해 학습된 AI가 최적의 결과를 도출함을 증명할 수 있는 방법에 대해 연구해야 합니다. 마지막으로 실험 데이터와 플레이 데이터를 통해 AI플레이에 일정한 패턴을 찾을 수 없음을 밝히는 연구를 진행합니다.

12>

업무 분담에 대해서는, 권기원 학우와 제가 딜러시스템과 GUI를 구현하기로 했고, 김태호 학우와 양혜원 학우가 AI플레이어의 기본기능 구현을 담당하였습니다. AI플레이어의 강화학습, 테스트와 검증은 조원 모두가 참여하기로 정했습니다.

전체 일정 계획은 아래와 같습니다. 중간발표 전까지는 딜러시스템과 게임 플레이가 가능한 게임 플레이가 가능한 기본 AI를 구현해 게임이 플레이되는 것을 중간 결과물로 제시할 예정입니다. 최종 발표 때는 이에 더해 강화학습된 AI와 이것의 테스트 결과 그리고 GUI를 구현해 결과물로 제시하려고 합니다.

14>

저희 조의 달성 목표는 총 5가지로, 우선 상황을 부여하고 승패를 판단하는 딜러를 구현한 후에 상황에 따라 적절하게 판단할 수 있는 AI플레이어를 구현합니다. 이 두가지 목표를 달성해 먼저 일대일 게임 플레이 환경을 만들 것입니다. 그 다음으로 강화학습을 하는 AI를 구현해 AI플레이어의 손익률을 높일 것입니다. 그 후에는 게임 플레이 환경을 GUI로 구현하고, 마지막으로 일대일 대결뿐만 아니라 최대 9명의 AI플레이어가 플레이하는 게임을 구현하는 것을 목표로 정했습니다.

15>

이번 프로젝트의 의의는 크게 3가지가 있습니다. 자동화된 딜러시스템과 AI플레이어를 두어 인간의 개입없이 진행되는 포커 게임을 구현하는 것입니다.

이런 시스템 안에서 상태/행동가치 판단 방법과 보상기준을 비교 연구하는 과정을 통해, 저희 조만의 독자적인 AI플레이어 강화학습 방법을 제시할 수 있습니다.

이런 과정에서 AI플레이어가 블러핑 기술에 대해 학습하면서 기계가 할 수 있는 블러핑 기술에 대한 강화학습 방법을 제시합니다.