002 강화학습

Reinforcement Learning

보상과 시행착오를 통해 스스로 행동을 학습하는 AI 기법

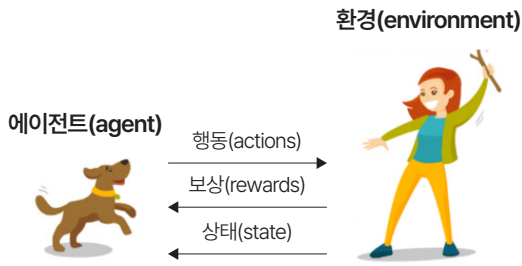
- 환경과 상호작용하며 행동 결과에 따른 보상을 받고, 이를 바탕으로 목표 달성을 위한 최적의 전략을 찾아가는 자기학습형 인공지능 학습 방식
- 정답 데이터를 주지 않고 경험을 통해 성능을 개선하는 자율 학습 기술

강화학습의 개념

강화학습은 인공지능이 환경 속에서 행동을 선택하고, 그 결과를 바탕으로 더 나은 선택을 학습하는 과정입니다. 시스템은 시행착오를 거치며 목표 달성에 유리한 행동을 찾아가는데, 이는 인간이나 동물이 경험을 통해 학습하는 방식과 유사합니다. 인공지능은 단순히 입력과 출력의 관계를 외우는 것이 아니라 행동과 보상의 관계를 스스로 파악해 전략을 세우며, 정답이 주어지지 않은 상황에서도 경험을 축적해 점차 더 나은 판단을 내릴 수 있습니다. 또한 강화학습은 시간의 흐름을 고려해, 현재의 행동이 미래에 어떤 영향을 미치는지를 평가하고 단기적 보상보다 장기적 이익을 극대화하는 방향으로 학습합니다. 이 과정에서 인공지능은 '지금의 보상'과 '앞으로의 이득' 사이의 균형을 조절하며 점차 효율적인 의사결정 구조를 만들어 갑니다.

강화학습의 구조

강화학습은 에이전트, 환경, 행동, 보상의 네 요소로 이루어진 순환 구조를 기반으로 합니다. 에이전트는 환경의 상태를 관찰하고 행동을 선택합니다. 환경은 그 결과를 보상으로 반환하고, 다시 에이전트는 이를 바탕으로 전략을 조정합니다. 이 과정이 반복되면서 에이전트는 어떤 행동이 유리한지 스스로 학습하고, 점점 더 높은 보상을 얻는 방향으로 전략을 발전시킵니다. 이러한 구조는 정답이 주어지지 않은 상황에서도 학습이 가능하다는 점에서 다른 학습 방식과 차별화됩니다. 단순한 예측이 아니라 환경과의 상호작용을 통해 점차 나은 판단을 내리도록 학습하기 때문에, 로봇 제어나 자율주행처럼 상황이 계속 변하는 환경에서 특히 효과적입니다.



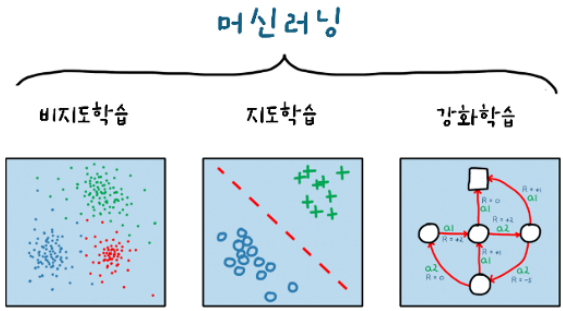
출처 : Mathworks

강화학습의 중요성

강화학습은 인공지능이 스스로 판단하고 환경 변화에 적응하도록 만드는 핵심 기술입니다. 단순히 주어진 데이터를 분석하는 수준을 넘어, 경험을 통해 전략을 조정하고 최적의 행동을 선택하는 능력을 제공합니다. 예를 들어 로봇은 강화학습으로 장애물을 회피하며 이동 경로를 최적화할 수 있습니다. 이러한 자율적 학습 구조는 불완전한 데이터나 예측이 어려운 상황에서도 성능을 지속적으로 개선할 수 있는 장점이 있습니다. 또한 최근 생성형 인공지능에서는 인간의 평가를 보상으로 활용하는 인간 피드백 기반 강화학습(RLHF)이 적용되어 모델의 응답 품질을 높이는 방식으로 활용되고 있습니다.

다른 학습 방식과의 비교

지도학습이 주어진 정답을 기반으로 예측 모델을 만드는 것이라면, 강화학습은 명시적인 정답 없이 행동의 결과를 평가해 최적의 전략을 스스로 찾아갑니다. 비지도학습이 데이터 속 패턴을 탐색하는 데 초점을 둔다면, 강화학습은 행동과 보상 간의 관계를 학습해 보상을 극대화합니다. 즉, 강화학습은 정적인 데이터 분석이 아닌, 변화하는 환경 속에서 전략을 개선하며 자율적 의사결정을 수행하는 지능형 학습 방식입니다. 이 때문에 강화학습은 정적인 데이터 분석이 아닌, 환경 변화에 따라 전략을 지속적으로 개선해야 하는 문제에 적합하며, 지능형 시스템의 핵심 기초로 평가됩니다.



출처 : Mathworks

관련 용어

인간 피드백 기반 강화학습(RLHF) vs AI 피드백 기반 강화학습(RLAIF)

인간 피드백 기반 강화학습(Reinforcement Learning from Human Feedback)과 AI 피드백 기반 강화학습(Reinforcement Learning from AI Feedback)은 강화학습 원리를 이용해 AI 모델의 출력을 개선하는 기술입니다. RLHF는 사람이 모델의 응답을 평가해 보상 신호로 활용함으로써, AI가 인간의 의도와 가치에 맞는 답변을 학습하도록 합니다. 반면 RLAIF는 검증된 AI가 다른 모델의 응답을 평가하는 방식으로, 인적 평가의 비용과 시간을 줄일 수 있습니다. 다만 AI의 오판이 누적될 위험이 있어, RLAIF는 RLHF를 대체하기보다 보완적으로 활용됩니다. 두 방식 모두 생성형 AI의 품질과 신뢰성을 높이는 핵심 학습 기술로 평가됩니다.