

2. 알파고 | 인간을 능가하는 기계의 등장

1. 딥 블루는 어떻게 체스 챔피언이 되었을까?

- 체스 특화 인공지능 컴퓨터 딥 블루

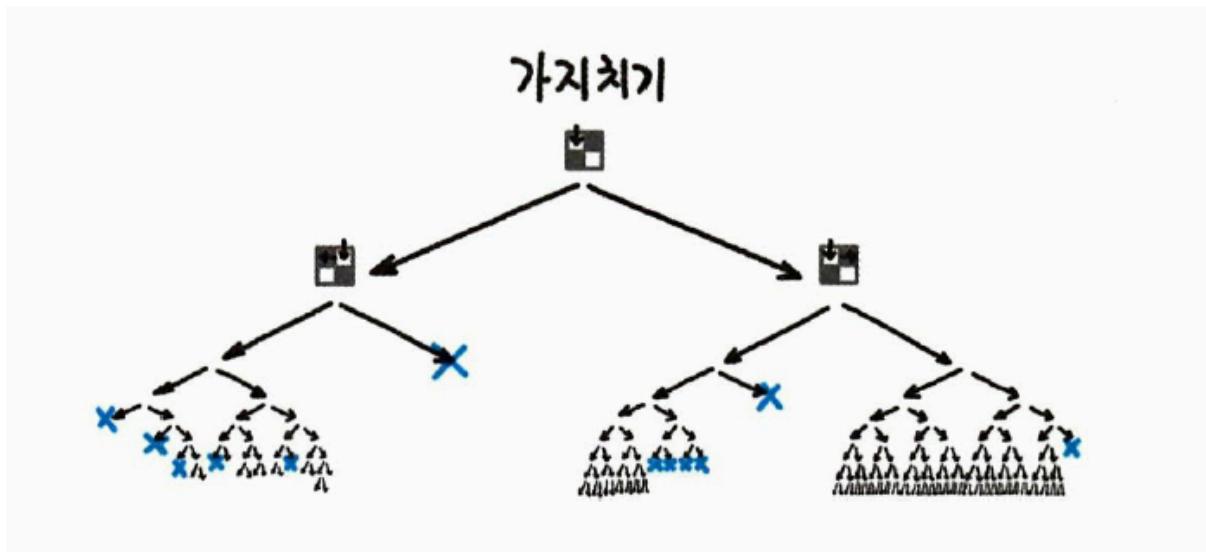


IBM이 만든 체스 특화 인공지능 컴퓨터이다. 딥 블루는 칩테스트 프로젝트의 일환으로 당시 세계 체스 챔피언이던 가리 카스파로프를 이겼으며 컴퓨터는 인간을 이길 수 없다는 오랜 패러다임을 깨는 계기가 되었다.

딥 블루는 체스의 수를 계산할 때 아래와 같은 방법을 사용

완전 탐색을 하지 않고 한 번 탐색해 복고 성과가 없다면 그쪽은 더 이상 탐색하지 않도록 표시.

→ 가지치기



2. 바둑, 필요한 모든 수를 계산할 수 있을까?

- 경우의 수가 너무 방대해서 예전처럼 “가능한 수를 다 계산하는 방식”으로는 접근이 안 됨
- 체스처럼 평가 함수 + 완전탐색(에 가까운) 방식이 아니라 **직관/형세 판단**이 중요한 영역으로 여겨짐
- 바둑은 “규칙이 명확한데도” **계산으로 해결하기 어려운 문제**라서 상징성이 컸다

3. 인간을 대표하는 천재 기사 이세돌

- 세계 최고 수준의 바둑 실력자
- 역시 아직은 인간의 직관력이나 이런 감각들을 컴퓨터인 인공지능이 따라오기는 약간 무리가 있지 않나 생각해서 자신감이 있다.

4. 도박의 확률을 이용하는 몬테카를로 방법

- **무작위(random) 표본 추출**을 반복해서 사용해 **복잡한 문제의 근사 해답**을 구하는 **확률적 계산 방법**이다. 해석적으로 풀기 어렵거나 계산량이 매우 큰 문제에서 특히 강력하다.
- ▼ 몬테 카를로 방법

핵심 아이디어

1. 무작위로 값들을 뽑는다

2. 각 표본에 대해 결과를 계산한다

3. 결과를 많이 모아 평균·분포로 근사한다

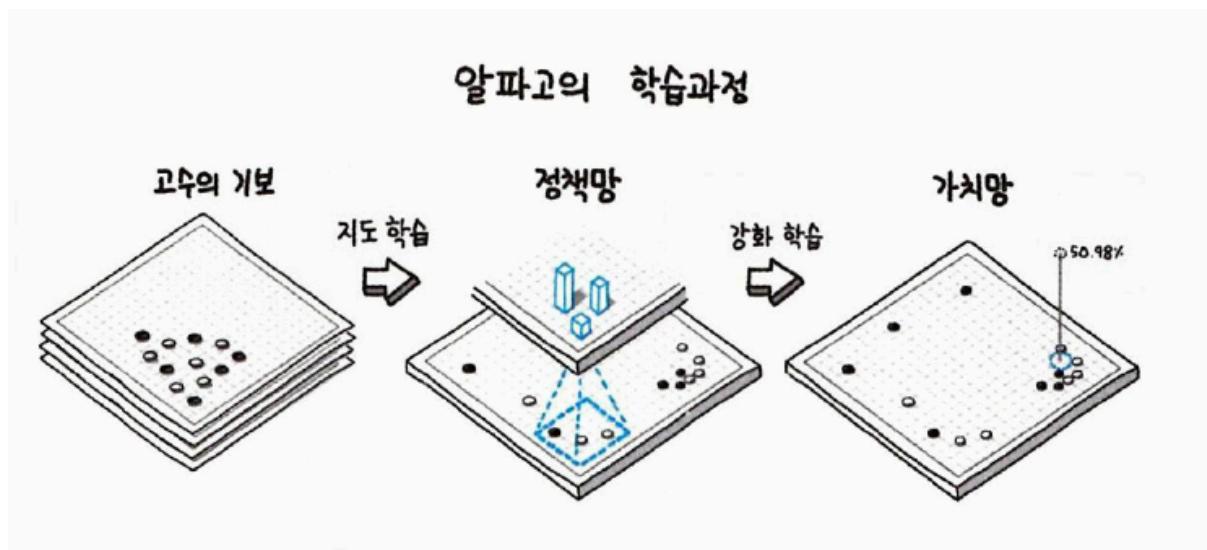
표본 수가 늘어날수록 결과는 확률적으로 실제 값에 수렴한다(큰 수의 법칙).

무작위 표본을 반복 생성해 평균으로 해답을 근사하는 확률적 계산 기법이다.

5. 정책망, 어디에 돌을 내려놓을까?

- 강화학습 (스스로 두고, 지고, 다시 배움)
 - 잘 두면 보상(강화)
 - 못 두면 벌점(약화)
 - 반복을 통해 전략이 다듬어짐

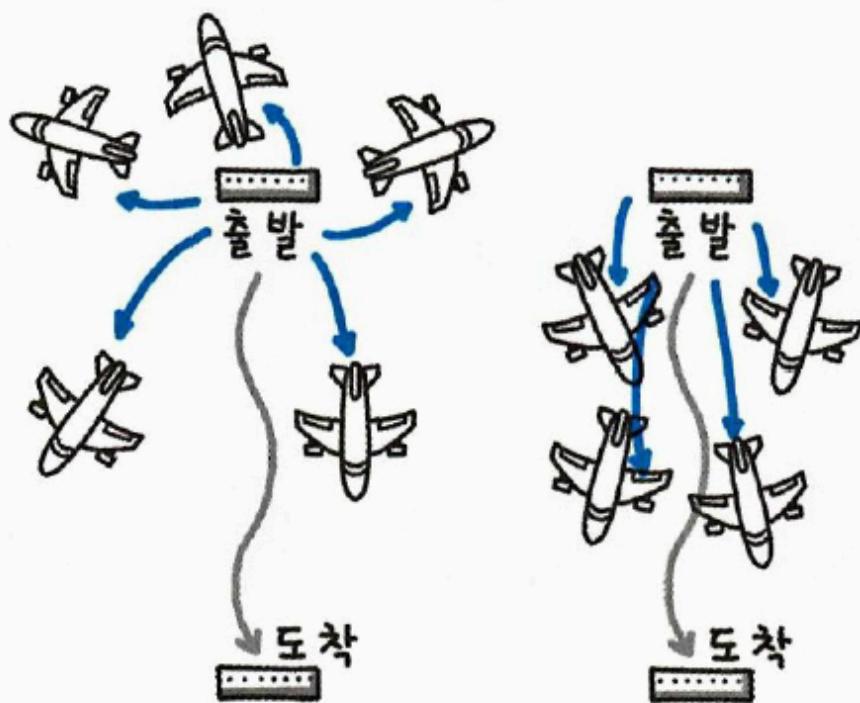
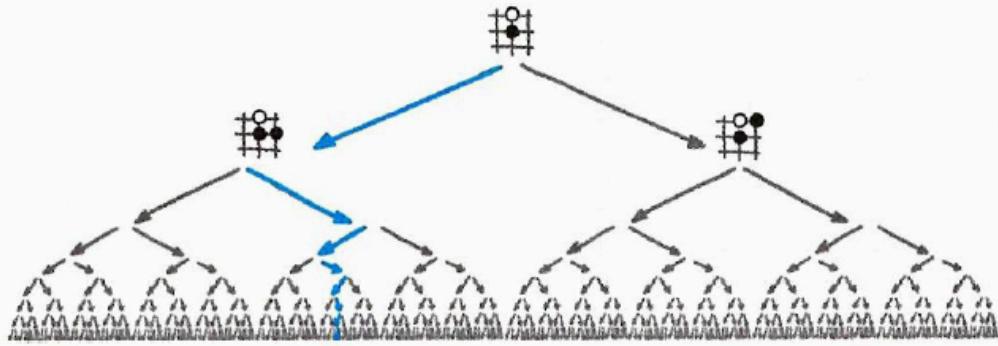
6. 가치망, 형세를 판단하다



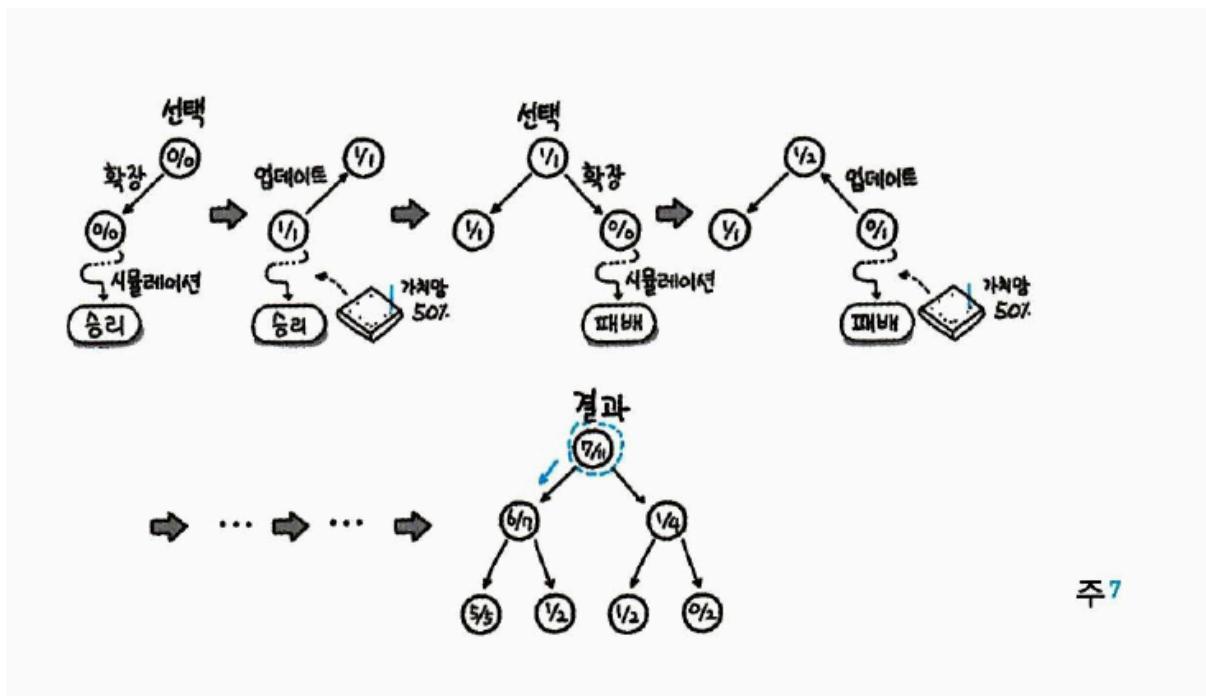
- 고수의 기보로 지도 학습을 진행
- 어디에 돌을 내려놓을지 판단하는 정책망을 만듬
- 정책망을 이용해 스스로 대국을 두어 강화학습을 진행
- 승리할 확률이 얼마나 되는지 알아내는 가치망을 만듬

7. 알파고가 수를 두는 방법

알파고의 몬테카를로 트리 탐색



- 기존 몬테카를로 방법이 무작위로 탐색해 속력을 높임
- 알파고의 몬테카를로 트리 탐색은 바둑 학습을 진행한 망을 사용해 가능성이 높은 수를 골라내 좀 더 현명하게 탐색



주 7

- 어떤 수에서 시작할지 승리할 가능성이 높아 보이는 수를 선택
- 기보학습 정책망을 이용해 다음 수를 어디에 둘지 확장
- 룰아웃 정책망을 이용해 끝까지 시뮬레이션
- 가치망의 점수를 50% 반영한 승패 여부를 모든 수에 업데이트
- 1~4의 과정을 계속 반복한 후 최종적으로 가장 많이 진행한 수를 다음 수로 결정

8. 신의 한수

- 알파고는 이세돌과의 네 번째 대국에서는 이세돌이 둔 78수를 막아내지 못함
- 이세돌이 둔 신의 한 수 지점에 착수할 확률을 알파고는 1만 분의 1로 매우 낮게 예측 따라서 충분히 탐색하지 않음
- 이후 더 이상 승률이 높은 지점을 찾아내지 못하고 엉뚱한 수를 남발하면서 패배

9. 알파고 제로, 인간은 필요 없다

- 기존 알파고는 학습을 수행할 인간 바둑기상의 기보가 필요하고 이를 습득해 식력을 키움
- 알파고 제로는 처음부터 자신과의 대국을 통해 실력을 쌓아감 (완전히 무에서 시작한다고 하여 알파고 '제로'라는 이름을 부여)

버전	엘로 평점	발표일	특징
이세돌	3,586	2010년 4월	이세돌의 최고 점수
알파고	3,739	2016년 3월	이세돌에 4승 1패
알파고 마스터	4,858	2017년 5월	프로 기사들에게 60전 전승, 커제에 5승 무패
알파고 제로	5,185	2017년 10월	이세돌을 이긴 알파고에 100전 전승, 알파고 마스터에 89승 11패

- 엘로 평점 (이기면 증가하고 지면 감소하는 방식, 어려운 미션에 더 큰 보상을 주는 매우 합리적인 점수 산정 방식)

10. 알파제로, 진정한 인공지능을 향하여

- 알파제로는 인간의 지식을 버리고, 스스로 학습하는 구조만으로도 최고 수준의 판단이 가능하다는 것을 증명