

# 2장 알파고

인간을 능가하는 기계의 등장

## 딥 블루는 어떻게 체스 챔피언이 되었을까?

### 1. 역사적 배경

- 1997년, 뉴욕
- 인간 대표: 가리 카스파로프
  - 1984년 최연소 체스 세계 챔피언
  - 10년 이상 세계 랭킹 1위 유지
- 컴퓨터 대표: IBM 슈퍼컴퓨터 딥 블루
- 결과: 딥 블루 승리
  - 컴퓨터가 정규 체스 대회에서 인간 세계 챔피언을 이긴 최초의 사례

### 2. 딥 블루의 의미

- 인간 지능을 상징하던 체스에서 **컴퓨터가 인간을 능가**
- 2016년 알파고 vs 이세돌 이전의 상징적 사건
- 인공지능 시대의 본격적인 시작을 알린 계기

### 3. 딥 블루의 핵심 강점: 계산 능력

- 체스 한 수당 선택지: 약 **40가지**
- 몇 수만 앞을 봐도 경우의 수: **수십억 개**
- 딥 블루의 성능:
  - 초당 **2억 번 계산**
  - 인간이 수백 년 걸릴 계산을 **약 20초 만에 처리**

### 4. 게임 트리(Game Tree)

- 체스의 모든 가능한 수를 나무 구조로 표현한 것
- 한 수 → 여러 갈래로 분기되는 구조
- 체스 게임 트리의 크기:  **$10^{120}$** 
  - 우주에 존재하는 원자 수보다 많음
- 전체를 탐색하는 것은 불가능

### 5. 효율적인 탐색 전략

## 5-1. 가지치기(Pruning)

- 승산이 없는 경로는 더 이상 탐색하지 않음
- 미로에서 막다른 길을 표시해 두는 것과 유사
- 탐색 속도와 효율을 크게 향상

## 5-2. 데이터 활용

- 오프닝 데이터: 게임 초반의 최적 수 활용
- 엔드게임 테이블베이스: 종반의 완벽한 해답 미리 저장
- 불필요한 계산 최소화

## 6. 인간 vs 컴퓨터의 사고 방식

### 인간 (카스파로프)

- 약 10만 가지 패턴을 직관적으로 대조
- 몇 가지 후보 수를 깊이 분석
- 심리 상태에 따라 판단이 흔들릴 수 있음

### 컴퓨터 (딥 블루)

- 확률과 점수 기반으로 수를 평가하고 정렬
- 실수 없음, 감정 없음
- 항상 동일한 성능 유지

## 7. 심리 전략

- 반응 속도를 일부러 빠르거나 느리게 조절
- 카스파로프가 딥 블루의 의도를 예측하기 어렵게 만듦
- 인간에게 심리적 압박 유발

## 8. 결론

- 딥 블루의 승리는 단순한 계산 승부가 아님
- 계산 능력 + 알고리즘 최적화 + 데이터 + 심리 전략의 결합
- 인간 중심이던 지적 게임의 패러다임 전환점
- 이후 알파고로 이어지는 인공지능 발전의 출발점

## 바둑, 필요한 모든 수를 계산할 수 있을까?

- 바둑판은 19×19, 총 361개 지점
- 가능한 경우의 수가 체스보다 압도적으로 많음
- 단순 계산 시 경우의 수는  $361! \approx 10^{768}$

- 실제 규칙을 반영해도 게임 트리 크기는 약  $10^{360}$

## 게임 트리 크기 비교

- 우주의 원자 수:  $10^{80}$
- 체스:  $10^{120}$
- 바둑:  $10^{360}$

→ 바둑은 체스보다 단순히 몇 배 어려운 수준이 아니라,  
현재의 계산 능력으로는 완전 탐색이 사실상 불가능

- 1997년 전문가 예측:
  - "컴퓨터가 바둑에서 인간을 이기려면 100년은 걸릴 것"

결론:

바둑은 계산 중심 접근으로는 한계가 있으며,  
알파고는 이를 극복하기 위해 전혀 다른 방식의 AI를 사용했다.

## 인간을대표하는천재기사이세돌

### 인간 천재의 직관과 고정관념

- 이세돌 9단
  - 이창호에 이어 세계 최강 계보를 잇는 기사
  - 과감하고 창의적인 묘수로 유명
  - 스승 권갑용: "100년에 한 번 나올 천재"
  - 17세에 32연승 → '불패소년'

## 도박의 확률을 이용하는 몬테카를로 방법

### 1. 몬테카를로 방법이란?

- 확률적 시뮬레이션으로 결과를 추정하는 방법
- 도박처럼 무작위 시행을 반복해 전체 경향을 파악
- 예: 잭팟 확률이 0.03%라면
  - 1만 번 → 약 3번
  - 10만 번 → 약 30번 발생

### 2. 이름의 유래

- 몬테카를로: 카지노로 유명한 도시
- 1940년대 핵무기 개발 중이던 과학자 스타니스와프 울람

- 복잡한 계산 대신 무작위 실험을 반복하는 방식을 고안
- 비밀 프로젝트의 암호명으로 '몬테카를로' 사용

### 3. 바둑에서 왜 필요한가?

- 바둑은:
  - 모든 돌의 가치가 같음
  - 현재 국면만 보고 승패 판단이 매우 어려움
- 체스처럼 점수 합산 방식 사용 불가
- 모든 경우를 끝까지 두어보는 것도 불가능

### 4. 바둑에서의 몬테카를로 아이디어

- 유리한 쪽은 무작위로 뒤도 이길 확률이 높다
- 현재 국면에서:
  - 무작위로 끝까지 여러 번 뒤봄
  - 승률이 높은 수를 좋은 수로 판단
- 여론조사처럼:
  - 전수조사 대신 무작위 표본으로 전체 경향 추정

### 5. 몬테카를로 트리 탐색 (MCTS)

- 2006년, 게임 트리 탐색에 몬테카를로 방법 결합
- 알고리즘 이름: Monte Carlo Tree Search
- 효과:
  - 바둑 인공지능 실력 급상승

### 6. 핵심 인물

- 레미 쿨롱: MCTS 최초 제안
- 아자 황
  - 바둑 AI 연구자
  - 딥마인드 합류 → 알파고 핵심 개발자
  - 이세돌 대국에서 알파고의 대리 착점자

### 7. 한계

- MCTS 도입 후:
  - 바둑 AI 실력: 아마 6단 수준
- 그러나:
  - 프로 기사에게는 여전히 4점 이상 접히는 수준
- → 이 한계를 넘기 위해 알파고는 딥러닝을 결합

# 정책망, 어디에 돌을 내려놓을까?

## 알파고의 정책망(Policy Network)

### 1. 기보학습 정책망 (프로 기사처럼 천천히 두는 사람)

- 사람의 기보로 학습
  - KGS 바둑 사이트
  - **6단 이상** 고수 기보만 사용
- 학습 데이터:
  - 약 **16만 판**, **3,000만 수**
- 입력: 현재 바둑판 상태
- 출력: **361개 위치 중 착점 확률**
- 성능:
  - 정확도 **57%**
  - 인간 기준 약 **5단 수준**
- 한계:
  - 기사들의 기풍이 섞여 평균화됨
  - 학습되지 않은 국면에 약함

### 2. 롤아웃 정책망 (아마 고수 수준이지만 엄청 빨리 두는 사람)

- 기보학습 정책망을 단순화한 경량 모델
- 특징:
  - 정확도 **24%**
  - 대신 **1,500배 빠른 속도**
- 역할:
  - 빠른 시뮬레이션용
  - 몬테카를로 트리 탐색에서 핵심적으로 사용

### 3. 강화학습 정책망

- 알파고끼리 대국하며 스스로 학습
- 시작은 기보학습 정책망을 기반으로 함
- 학습 방식:
  - 이기면 그 대국에서 둔 모든 수를 강화
  - 점수 차이는 중요하지 않음 (반 집 승리도 동일한 승리)
- 결과:
  - 기보학습 정책망과 대국 시 약 **80% 승률**
  - 즉, **5단을 80%로 이기는 수준**

#### 4. 정책망 정리 표

정책망 종류	특징
기보학습 정책망	인간 기보 학습, 정확도 57%, 약 5단
롤아웃 정책망	정확도 24%, 대신 1,500배 빠름
강화학습 정책망	자가 대국으로 학습, 5단 상대 80% 승

#### 5. 중요한 포인트

- 강화학습 정책망이 가장 강하지만:
  - 단독으로는 이세돌을 이길 수 없음
- 따라서:
  - 정책망 자체로 대국하지 않음
  - 더 강한 다른 신경망을 학습시키는 재료로 사용

## 가치망, 형세를 판단하다

### 가치망(Value Network)

- 현재 국면의 승리 확률을 예측하는 신경망
- 출력:
  - 승률이 50% 이상 → 유리
  - 50% 미만 → 불리
- 차이:
  - 정책망: 어디에 둘지 결정
  - 가치망: 이길 수 있는지 판단

### 왜 중요한가?

- 바둑은:
  - 돌의 점수 개념이 없음
  - 판의 형태만 보고 형세 판단이 매우 어려움
- 알파고는:
  - 현재 국면만 보고 승패를 예측하는 방법을 처음으로 구현

### 가치망 학습 방법

- 가장 강한 강화학습 정책망끼리 자가 대국
- 약 3,000만 판 대국
- 각 대국에서 한 장면씩 추출
- 해당 국면 이후의 실제 승패 결과로 학습

- 예:
  - 승률 74% → 같은 상황에서 이길 확률 74%

## 의미

- 끝까지 두지 않아도 형세 판단 가능
- 알파고의 가장 혁신적인 기술 중 하나

## 알파고 전체 학습 흐름

1. 인간 기보로 정책망 지도 학습
2. 자가 대국으로 정책망 강화학습
3. 자가 대국 결과로 가치망 학습

## 알파고가수를두는방법

### 1. 기본 아이디어

- 바둑은 경우의 수가 너무 많아 완전 탐색 불가능
- 대신 몬테카를로 트리 탐색(MCTS) 사용
- 일부만 똑똑하게 탐색해도 충분히 좋은 결과 가능

### 2. 몬테카를로 트리 탐색(MCTS)

- 무작위 탐색의 단점:
  - 낮은 확률의 묘수를 놓칠 수 있음
- 해결:
  - 유망한 수부터 집중적으로 탐색
  - 무작위 대신 학습된 신경망 사용

### 3. 탐색 과정 요약

1. 유망한 수 선택
  - 정책망으로 승리 가능성 높은 수부터 고름
2. 확장
  - 기보학습 정책망으로 다음 수 후보 생성
3. 시뮬레이션
  - 게임이 끝날 때까지 가상 대국 수행
  - 빠른 계산을 위해 롤아웃 정책망 사용
4. 평가
  - 시뮬레이션 결과 (50%)

- 가치망의 승률 예측 (50%)
- 두 결과를 반반씩 결합

#### 5. 업데이트

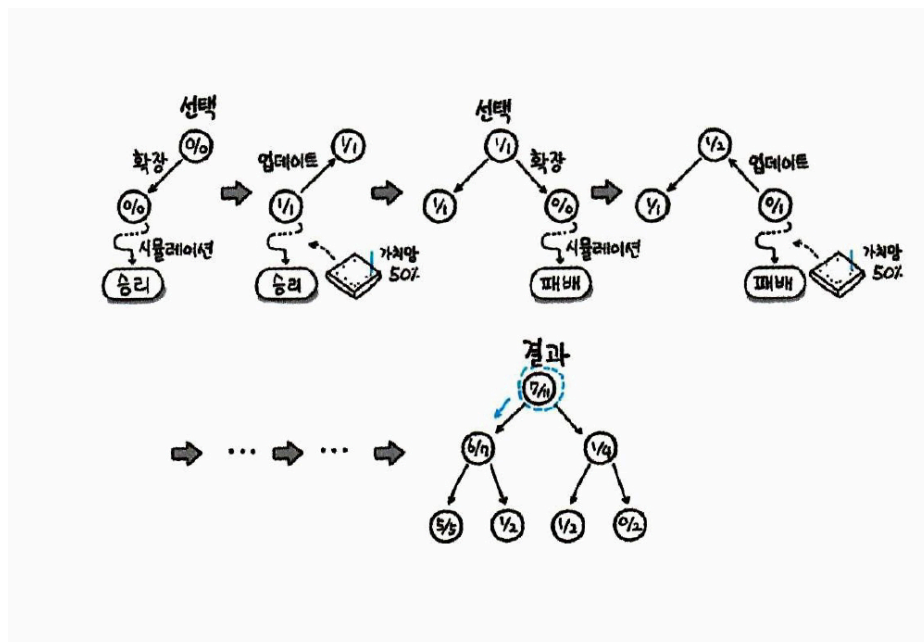
- 승리한 경우: 승리 점수 + 진행 점수
- 패배한 경우: 진행 점수만 반영

#### 4. 다음 수 선택 기준

- 점수가 가장 높은 수 ❌
- 가장 많이 시뮬레이션된 수 ○
- 이유:
  - 많이 시도된 수일수록 신뢰도 높음
  - 리뷰 많은 음식점이 더 믿을 만한 것과 같은 원리

#### 5. 핵심 특징

- 딥 블루: 계산력으로 모든 수를 최대한 탐색
- 알파고:
  - 계산 대신 확률 + 학습 + 선택적 탐색
  - 정책망 + 가치망 + MCTS의 결합



1. 어떤 수에서 시 작할지 승리 할 가능성 이 높아 보이는 수를 선택 합니다.
2. 기보학습 정책망을 이용해 다음 수를 어디에 둘지 확장합니다.
3. 롤아웃 정책망을 이용해 끝까지 시뮬레 이션합니다.
4. 가치망의 점수를50% 반영한승패 여부를모든수에 업데이트합니다.
5. 1~4의 과정을 계속해서 반복한 후 최종적으로 가장 많이 진행한 수를 다음 수로 결정 합니다.



# 신의 한 수

## 1. 사건 개요

- 2016년 알파고 vs 이세돌 4국
- 이세돌의 78수
  - 알파고와 바둑 전문가 모두 예상하지 못한 수
  - 단번에 판세를 뒤집으며 이세돌의 승리로 이어짐
- 결과:
  - 이세돌, 인류가 알파고에게 거둔 유일한 1승

## 2. 알파고가 막지 못한 이유

- 알파고는 확률이 높은 수 위주로 탐색
- 78수 위치의 착수 확률:
  - 약 1만 분의 1
- 따라서:
  - 해당 지점은 거의 탐색하지 않음
  - 다른 유망한 수는 수백만~수천만 번 탐색
  - 78수는 수십 번도 탐색하지 못함

## 3. 구조적 한계

- 몬테카를로 트리 탐색의 특성:
  - 가능성 낮은 수는 깊게 보지 않음
- 결과:
  - 극히 드문 묘수를 사전에 발견하지 못함
  - 수를 둔 이후에야 그 의미를 인식

## 4. 그 이후

- 78수 이후:
  - 알파고는 승률 높은 수를 찾지 못함
  - 비정상적인 수('떡수')를 두기 시작
- 이는:
  - 계산 오류가 아니라 탐색 전략의 한계

# 알파고 제로, 인간은 필요 없다

## 알파고 제로의 등장

- 알파고 vs 이세돌 대국(2016)은 전 세계에 큰 충격을 줌
- 하지만 구글 딥마인드는 여기서 멈추지 않고 알파고 제로(AlphaGo Zero) 개발

## 알파고 vs 알파고 제로 차이

- 기존 알파고
  - 인간 프로기사의 기보로 학습
- 알파고 제로
  - 인간 기보 없이 자기 자신과의 대국만으로 학습
  - 완전한 '무(無)'에서 시작 → 그래서 '제로'

## 압도적인 학습 속도

- 알파고 제로:
  - 하루에 100만 대국 가능
- 인간:
  - 평생 뒤도 약 13만 대국
- 결과:
  - 3일 만에 기존 알파고를 100대 0으로 압도
  - 40일 후 모든 이전 알파고 버전 제압

## 인간의 한계

- 알파고 마스터 → 중국 세계 최강 커제에게 5전 전승
- 알파고 제로는 그보다도 훨씬 강력
- 이제:
  - 인간은 바둑에서 컴퓨터를 이길 수 없음
  - 이세돌의 4국 승리는 인간의 마지막 승리로 기록

## 엘로(Elo) 평점

- 체스에서 시작된 실력 점수 체계 (바둑에도 적용)
- 특징:
  - 강한 상대를 이기면 점수 크게 상승
  - 약한 상대를 이기면 점수 소폭 상승
- 알파고와 이세돌의 실력 차이는 엘로 평점으로도 명확히 드러남

## 알파제로, 진정한 인공지능을 향하여

## 알파고 제로와 알파제로

- 알파고 제로
  - 인간 기보 없이 자기 대국만으로 학습
  - 바둑 최강의 인공지능
- 알파제로(2018)
  - 바둑을 넘어 체스·일본 장기까지 확장
  - 동일한 원리로 여러 보드게임 정복

## 알파제로 vs 스톡피시(체스)

- 스톡피시:
  - 기존 최강 체스 AI
  - 초당 7,000만 수 계산
- 알파제로:
  - 초당 약 8만 수만 계산 (875분의 1)
  - 사전 전략 없이 강화학습만으로 학습
- 결과:
  - 100판 중 28승 72무, 무패
  - 압도적 우위 입증

## 의미와 한계

- 알파제로는:
  - 모든 수를 계산하지 않고도 직관적 선택으로 최고 성능 달성
  - 인간 도움 없이 학습 가능성을 증명
- 하지만:
  - 정해진 규칙의 게임에서 계산을 잘한 것
  - 인공지능이 모든 면에서 인간을 능가한 것은 아님

## 인공지능에 대한 시각

- 호킹: 인공지능은 인류에 위협이 될 수 있음
- 앤드류 응: 과도한 공포는 불필요
- 페이페이 리: “자동차가 인간보다 빠른 게 놀라운가?”
- 네이트 실버:
  - 기술을 과대평가도, 과소평가도 말 것