

추론 모델은 어떻게 만들어지는가?

이동재

2025년 1월 30일

Abstract

지난 설 연휴 중국 Deepseek의 R1모델이 저렴한 훈련 비용과 높은 성능으로 큰 화제를 모았다. R1 모델은 흔히 ‘추론 모델’이라 불리는 종류의 모델로, 매우 긴 길이의 생각의 사슬 (Chain of Thought)을 통해 복잡한 문제를 해결하도록 훈련된 모델이다. 본 글에서는 오픈 소스 진영에서 공개한 추론 모델의 구축 방법을 정리하고, 그 중 핵심 요소들을 하나씩 짚어볼 것이다.

OpenAI사에서 발표한 o1모델의 등장 이후, 매우 긴 생각의 사슬을 활용하면 모델의 성능을 크게 향상할 수 있음이 많은 사람들에게 공개되었다. 하지만, 올바른 추론 과정을 모델이 학습하려면 고품질의 추론 데이터가 있어야 하지만, 복잡한 문제를 해결한 풀이 과정 데이터를 구하는 작업은 매우 어렵고, 검증도 어렵다.

Deepseek R1에서 제시한 해결 방법은 모든 것을 모델에게 맡기는 것이다. 즉, 인간이 만들어낸 올바른 추론 과정을 모델이 학습하도록 하는 것이 아니라, 모델이 스스로 자신에게 가장 적합한 추론 방식을 찾아나서도록 훈련하는 것이다. 인간은 모델이 내놓은 답을 보고 채점만 하면 된다. 이러한 관찰은 과거 인간의 기보를 학습한 알파고 초기 버전과, 바둑의 규칙만 주어진 상태로 스스로 바둑을 학습한 알파고-제로의 사례와 일맥상통한다. 결국, 인간의 휴리스틱을 모델에게 학습하는 것보다 규칙만 제공하고, 모델이 그 속에서 최적의 답을 스스로 찾아 나서도록 만드는 것이 가장 효과적이라는 것이다.

이 시스템의 가장 큰 장점은 데이터 고갈 문제로부터 어느 정도 자유롭다는 점이다. 기존의 학습 방식은 인간이 만들어낸 풀이 과정을 요구했지만, 현재 학습 방식에서는 인간이 복잡한 문제를 만들고, 이를 채점할 시스템만 갖추면 모델이 스스로 학습하는 것이 가능하다. 문제와 그 풀이 과정까지 생성하는 것은 어렵지만, 문제를 생성하고 자동으로 채점해주는 시스템을 구축하는 것은 훨씬 수월한 작업이다.

프로그래밍 관점에서는 요구사항과 올바른 프로그램인지 검증하는 시스템을 구축하면 모델이 스스로 프로그래밍을 학습하게 할 수 있을 것이다. 이러한 방식을 통해 현존하는 소프트웨어 공학 문제들을 어디까지 해결할 수 있을지 가능해보고, 앞으로 우리의 역할이 무엇일지 고민해보아야 할 시점이다.