

032 미세조정

Fine-Tuning

기준에 학습된 AI 모델을 새로운 데이터나 목적에 맞게 최적화하는 과정

- 사전 학습 모델을 기반으로, 특정 분야나 작업 목적에 맞게 추가 학습을 수행하는 단계적 조정 과정
- 새로운 데이터의 특성과 맥락을 반영해, 적은 자원으로 높은 효율을 달성하도록 돋는 학습 기술

● 미세조정의 개념

미세조정은 이미 학습된 AI 모델을 새로운 목적으로 환경에 맞게 다시 조정하는 학습 과정으로, 대규모 데이터로 사전 학습된 모델이 특정 분야에서 더 정확하고 안정적으로 작동하도록 추가 학습을 수행하는 기술을 의미합니다. 기존 학습 과정에서 모델이 얻은 일반 지식을 유지한 채, 새로운 데이터의 특성을 반영해 모델의 행동을 재정렬하는 절차라 할 수 있습니다. 최근 LLM의 발전 이후 미세조정은 단순한 성능 개선을 넘어, 응답 스타일·추론 방식·안전성 기준을 목적에 적합한 방향으로 정교하게 다듬는 핵심 단계로 활용됩니다.

● 미세조정의 과정

기존에 사용되었던 풀 미세조정은 모델의 모든 매개변수를 새로운 데이터에 맞춰 다시 학습시키는 방식입니다. 가장 강력한 조정 효과를 얻을 수 있지만 연산 비용이 매우 크고, 기존 능력이 손상되는 지식 파괴 위험도 존재한다는 한계 때문에 최신 모델 규모에서는 실제 산업 활용 사례가 제한적입니다. 현재 산업과 연구에서 주로 사용되는 방식은 필요한 부분에만 변화를 주어 모델의 행동을 업데이트하는 방식입니다. 대표적인 기법은 LoRA, Prefix-Tuning 등 PEFT(Parameter-Efficient Fine-Tuning) 계열로, 모델의 대부분을 고정한 채 극히 일부의 매개변수만 조정해 연산 비용을 줄이고, 기존 성능을 크게 훼손하지 않으면서 특정 능력을 강화할 수 있다는 장점이 있습니다. 특히 소량·고품질 데이터만으로도 특정 도메인 지식이나 말투, 응답 형식을 부여하는 데 효과적이나, 새로운 목표에 맞추는 과정에서 특정 작업에 최적화되면서 기존의 일반 능력이 감소하는 현상이 발생할 수 있으며, 그 효과는 데이터 품질·분포 적합성·튜닝 범위 등에 크게 의존합니다.

● 미세조정의 필요

사전 학습된 모델은 다양한 상황을 폭넓게 처리할 수 있는 범용 능력을 갖추었지만, 실제 서비스나 업무 환경에서는 특정 도메인의 표현 방식, 규제 요구사항, 기관별 문서 구조, 고객 응대 스타일처럼 맥락적으로 특수한 요소들이 중요하게 작용합니다. 미세조정은 이러한 간극을 해소하여 모델이 목표 작업에 더 적합한 예측과 판단을 수행하도록 만드는 핵심 수단입니다. 또한 기업·기관은 자체 데이터 기반으로 모델을 최적화함으로써 서비스 품질을 향상시키고, 모델 오작동을 줄이며, 경쟁력 있는 맞춤형 시스템을 구축할 수 있습니다. LLM 시대에는 특히 데이터 품질·안전성 기준·조직별 요구사항이 다양해지면서 사전 학습 모델만으로는 충분하지 않은 경우가 많아, 미세조정이 실질적 성능 확보와 안정성 강화의 필수 단계로 자리 잡고 있습니다.