KO: ASCender: Boid 로 유도된 Transformer Attention 의 구조적 편향

EN: ASCender:Boid-Guided Structural Biases in Transformer Attention

Abstract

본 논문에서는 군집 지능 (Swarm Intelligence) 에서 영감을 받은 구조적 편향을 도입한 새로운 Transformer 아키텍처인 ASCender 를 제안한다. 기존 Self-Attention 메커니즘은 모든 토큰 쌍을 동일하게 처리함으로써 계산 효율성이 떨어지고, 결과 해석 가능성이 제한되는 한계를 가진다. 이러한 균일한 처리 방식은 특히 계층적 구조나 공간적 추론이 필요한 과제에서 도메인 특유의 관계 구조를 충분히 반영하지 못한다.

ASCender 는 집단 행동 모델인 Boids 에서 유래한 세 가지 유도 편향 (Alignment, Separation, Cohesion) 을 Attention Score 계산 과정에 직접 통합한다. 이를 통해 의미적으로 유사한 토큰을 정렬 (Alignment) 하고, 불필요하거나 잡음이 많은 토큰과는 분리 (Separation) 하며, 의미 있는 군집 내로 응집 (Cohesion) 시키는 과정을 유도한다. 이러한 Swarm-aware Attention 설계는 해석 가능한 Attention Map 을 생성하고, 불필요한 토큰 상호작용을 억제하며, 장문 컨텍스트 처리 효율성을 향상시킨다.

다양한 자연어 처리 및 추론 벤치마크 실험에서 ASCender 는 기존 Transformer 대비 최대 x.x% 의 정확도 향상과 xx% 의 Attention FLOPs 절감을 달성하였다. 이 연구 결과는 생물학적 영감을 받은 유도 편향이 신경망 Attention 메커니즘의 효율성과 해석 가능성을 동시에 개선할 수 있는 잠재력을 보여주며, 향후 군집 지능 원리를 딥러닝 아키텍처 설계에 접목할 가능성을 제시한다.

Introduction

Transformer 는 자연어 처리, 컴퓨터 비전, 멀티모달 학습 전반에서 혁신을 일으키며, 다양한 과제에서 사실상의 표준 아키텍처로 자리잡았다. Transformer 의 핵심인 Self-Attention 메커니즘은 시퀀스 내 모든 토큰 쌍의 상호작용을 계산한다. 그러나 이러한 메커니즘은 모든 토큰 쌍을 균일하게 처리하며, 명시적인 구조적 가이던스 없이 학습된 Attention 가중치에만 의존한다. 이로 인해 불필요하거나 중복된 토큰 간 상호작용에 계산 자원이 소모되고, Attention 패턴이 해석하기 어렵게 나타나는 문제가 발생한다.

최근 연구에서는 Sparse Attention, 저랭크 근사 (Low-rank Approximation), Positional Encoding 개선 등으로 이러한 한계를 보완하려 하였다. 이러한 방법들은 계산 효율성이나 컨텍스트 모델링을 개선할 수 있지만, 특정 도메인에서 토큰이 서로 어떻게 관계 맺어야 하는지에 대한 명시적인 유도 편향 (Inductive Bias) 을 포함하지 않는 경우가 많다. 적절히 설계된 유도 편향은 학습을 보다 의미 있고 해석 가능한 관계 구조로 이끌 수 있다.

본 연구에서는 군집 지능 (Swarm Intelligence), 특히 집단 행동 모델인 Boids 에서 영감을 받아 Transformer 에 적용 가능한 새로운 형태의 구조적 편향을 설계하였다. Boids 모델은 단순하지만 강력한 세 가지 규칙을 통해 집단의 거동을 설명한다: Alignment(이웃의 평균 방향으로 정렬), Separation(과도한 밀집 회피), Cohesion(이웃의 평균 위치로 응집). 우리는 이 원리를 Self-Attention 맥락에서 재해석하여 Attention Score 계산 과정에 직접 반영하였다.

이를 위해 Alignment–Separation–Cohesion–enhanced Transformer 인 **ASCender** 를 제안한다. ASCender 는 의미적으로 유사한 토큰 군집을 형성하도록 유도하고, 불필요한 상호작용을 줄이며, 장문 컨텍스트 처리 효율을 향상시킨다. NLP 와 추론 벤치마크 실험을 통해 ASCender 가 기존 Transformer 보다 더 높은 정확도와 해석 가능성을 달성함과 동시에 계산 효율도 개선함을 보인다.

본 논문의 기여 사항은 다음과 같다.

1. Alignment, Separation, Cohesion 원리를 구조적 편향 형태로 Self-Attention 에 통합하는 새로운 접근법 제안
2. 편향을 Attention Score 행렬에 직접 삽입하여 해석 가능한 Attention 패턴과 효율적인 토큰 상호작용 구현
3. 다양한 벤치마크 실험을 통한 성능 및 계산 효율성 향상 검증