

43 딥러닝을 이용한 SSD 성능 개선 연구

소속 정보컴퓨터공학부

분과 D

팀명 SSDeep Learning

참여학생 이준형, 강인석, 김지수

지도교수 안성용

과제 배경 및 목표

과제 배경

SSD는 빠른 속도로 인해 주요 저장 장치가 되었지만, 데이터를 쓰고 지우는 과정에서 **쓰기 증폭(Write Amplification)** 문제가 발생해 수명과 성능을 떨어뜨린다.

기존의 SSD 데이터 관리 방식(FTL)은 정해진 규칙을 따르므로, 다양한 입출력 패턴에 유연하게 대응하지 못해 쓰기 증폭 문제를 효과적으로 해결하지 못하고 있다.

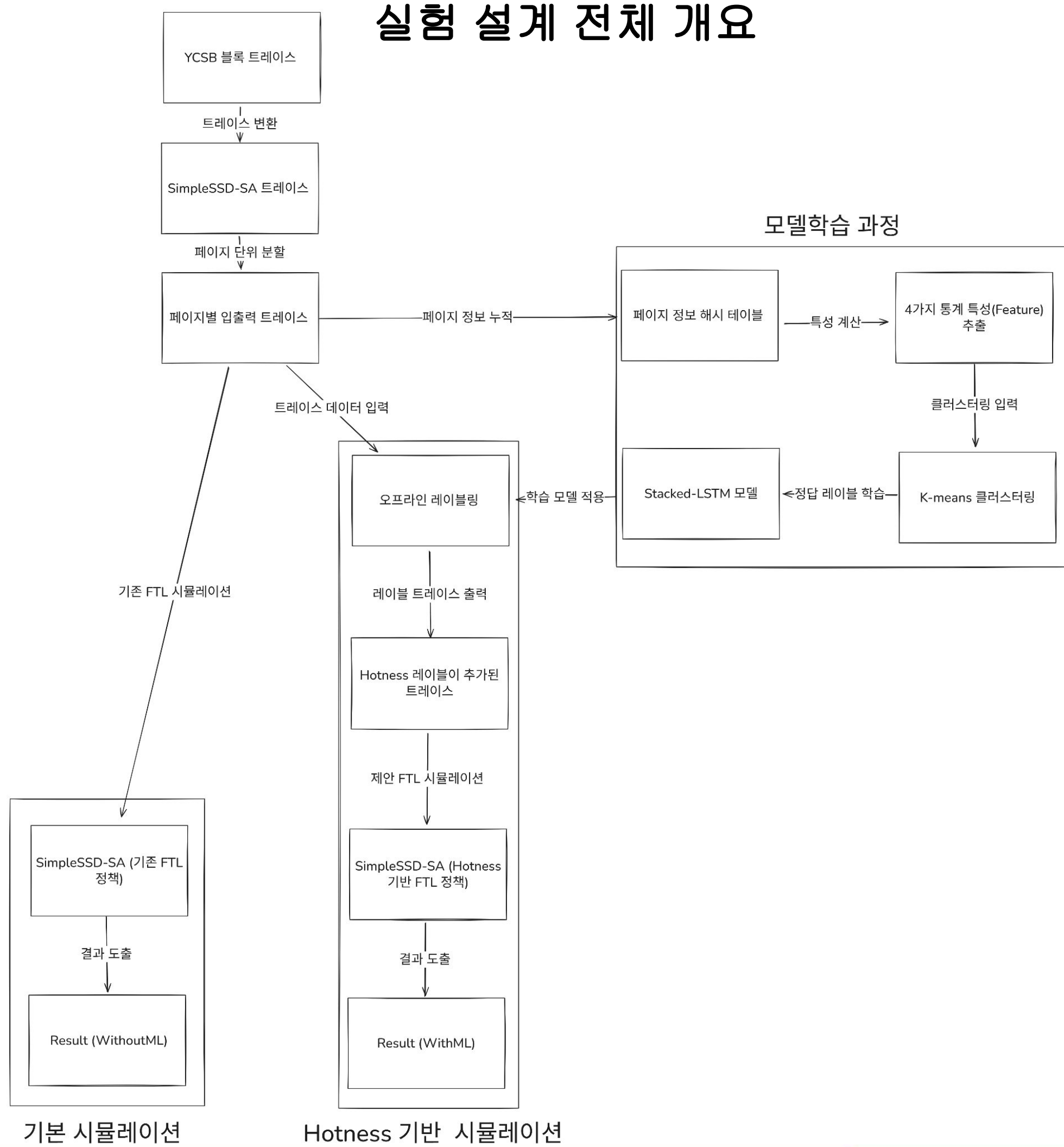
과제 목표

딥러닝 기술을 도입하여 데이터의 사용 빈도(Hotness)를 예측하고, 이를 바탕으로 SSD의 데이터 관리 방식을 개선한다.

쓰기 증폭을 줄여 SSD의 성능과 수명을 동시에 향상시키는 것을 목표로 한다.

연구 내용 및 방법

실험 설계 전체 개요



딥러닝을 활용한 데이터 Hotness 예측

1. 예측을 위한 학습 데이터 생성

방대한 I/O 트레이스를 분석하여 페이지의 접근 패턴을 나타내는 **시간 간격 평균, 시간 간격 표준편차, 요청 크기 평균, 마지막 수명의 네 가지 핵심 통계 특징**을 추출한다. 이렇게 추출된 특징들을 기반으로 K-means 군집화(K=4)를 수행하여 Hotness를 **Cold(0), Medium(1), Warm(2), Hot(3)** 네 가지 라벨로 정의하고, 이를 LSTM 모델의 학습 정답으로 사용한다.

2. Stacked-LSTM 모델 설계

페이지 접근 패턴의 시계열적 특성을 효과적으로 학습하기 위해 **두 개의 LSTM 레이어**를 쌓은 모델을 설계했다. 이 모델은 과거의 패턴을 바탕으로 미래의 Hotness를 예측하며, 99.81%라는 정확도를 기록했다.

3. FTL 정책에 Hotness 적용

학습된 모델의 예측 결과(Hotness 라벨)를 기반으로 FTL(Flash Translation Layer)이 **Hot 데이터와 Cold 데이터를 물리적으로 다른 블록에 분리하여 저장**하는 동적 정책을 구현했다. 이로써 가비지 컬렉션(GC) 시 발생하는 불필요한 유효 페이지 복사량을 최소화하여 쓰기 증폭을 근본적으로 완화한다.

결과 및 결론

본 연구는 기존 Page Mapping과 제안한 Hotness 기반 정책의 GC 동작을 비교 분석하였다. 기존 방식은 GC가 빈번히 발생하며 유효 페이지를 대량 재복사하여 내부 쓰기량을 증가시켰다. 반면 Hotness 기반 정책은 GC 빈도를 줄이고, 발생 시에도 무효 페이지 중심으로 처리하여 유효 데이터의 강제 복사를 효과적으로 억제하였다.

중요한 점은 W/A 절대값이 아니라 "무엇을 옮겼는가"이다. Hot/Cold 분리를 통해 불필요한 유효 데이터 이동을 방지함으로써 지연 및 내구성 측면에서 실질적 성능 향상을 확인하였다. Hotness 기반 접근법은 GC 최소화 → 무효 데이터 중심 처리 → 유효 데이터 복사 비용 제거의 연쇄 효과로 SSD 성능을 구조적으로 개선하는 것으로 나타났다.

IOPS(초당 처리량 비교)

