

딥러닝을 이용한 SSD 성능 개선 연구

SSD가 더 빠르고 오래가는 비결

201924544 이준형 | 201914502 강인석 | 202155537 김지수

지도교수: 안성용

부산대학교 컴퓨터공학과





우리가 모르는 SSD의 비밀

01

페이지 단위 읽기/쓰기

낸드 플래시는 작은 페이지 단위로 데이터를 읽고 씁니다.

02

블록 단위 삭제

하지만 삭제는 훨씬 큰 블록 단위로만 가능합니다.

03

쓰기 증폭 발생

이 구조적 차이로 인해 불필요한 추가 쓰기가 발생합니다.

04

성능 저하 & 수명 단축

결과적으로 SSD 성능이 떨어지고 수명이 단축됩니다.

딥러닝, SSD를 더 똑똑하게 만들다

핵심 아이디어

- 자주 쓰는 Hot 데이터와 거의 쓰지 않는 Cold 데이터를 구분
- Stacked-LSTM 모델로 데이터의 'Hotness' 예측
- 예측 정보 기반으로 FTL이 데이터를 동적 관리
- 가비지 컬렉션 효율 극대화



문제 해결을 위한 여정



데이터 전처리

YCSB 트레이스를 페이지 단위로 분할하고 통계 특성 추출



모델 학습

K-means 클러스터링과 Stacked-LSTM으로 Hotness 예측 모델 구축



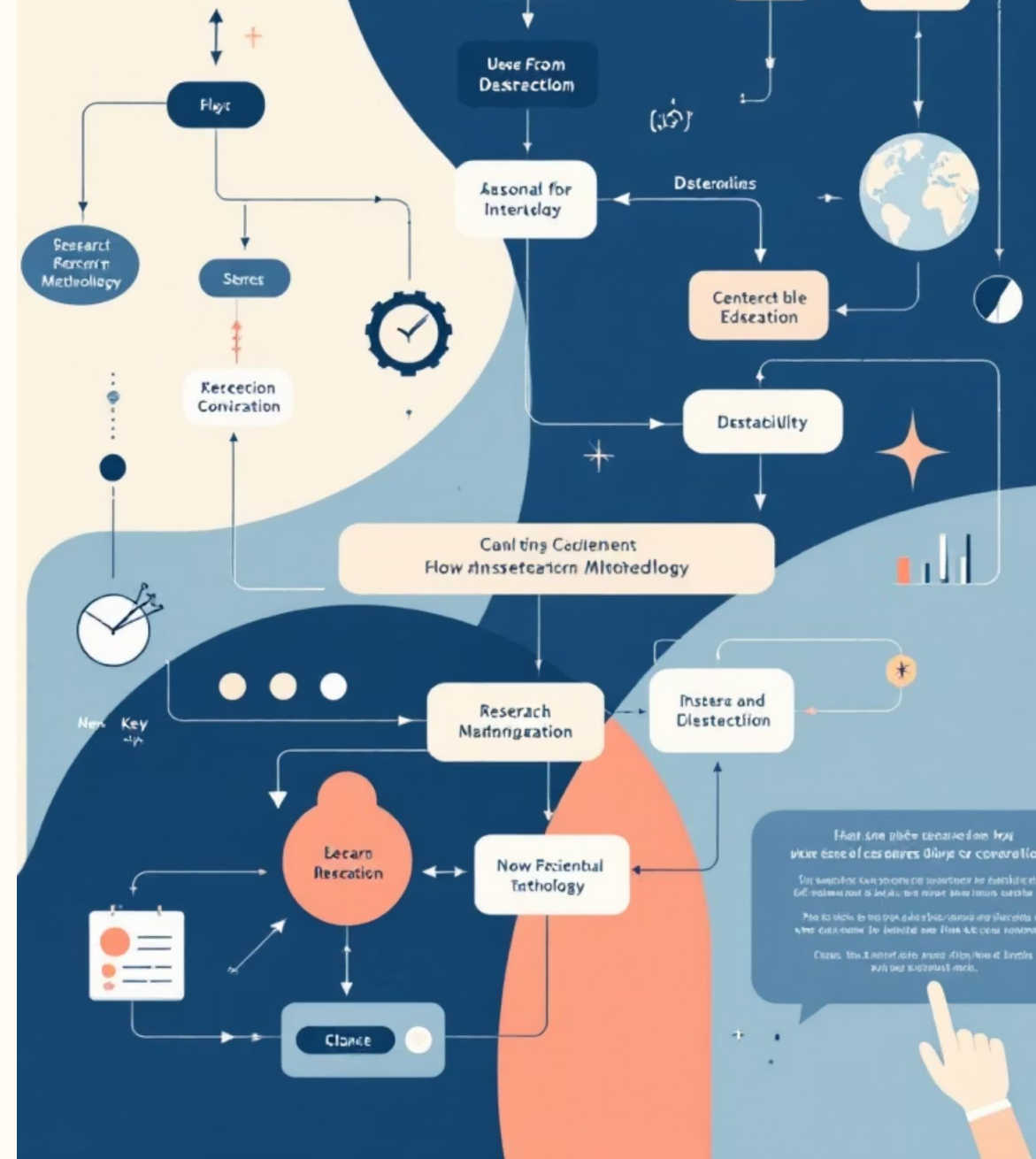
시뮬레이션 검증

SimpleSSD-SA를 활용한 성능 비교 및 검증

searchResearch Flowshollow

Is your sector business here and out, tested at long for wood dation a level.

Large over the past is the most over the past.





데이터 전처리 및 특성 추출

입출력 트레이스 변환

YCSB 블록 트레이스를
SimpleSSD-SA 형식으로 변환하
고 페이지 단위로 분할

4가지 통계 특성

시간 간격 평균, 표준편차, 요청 크
기 평균, 마지막 수명 추출

해시 테이블 구축

페이지별 접근 패턴 정보를 효율적으로 관리하는 구조 생성

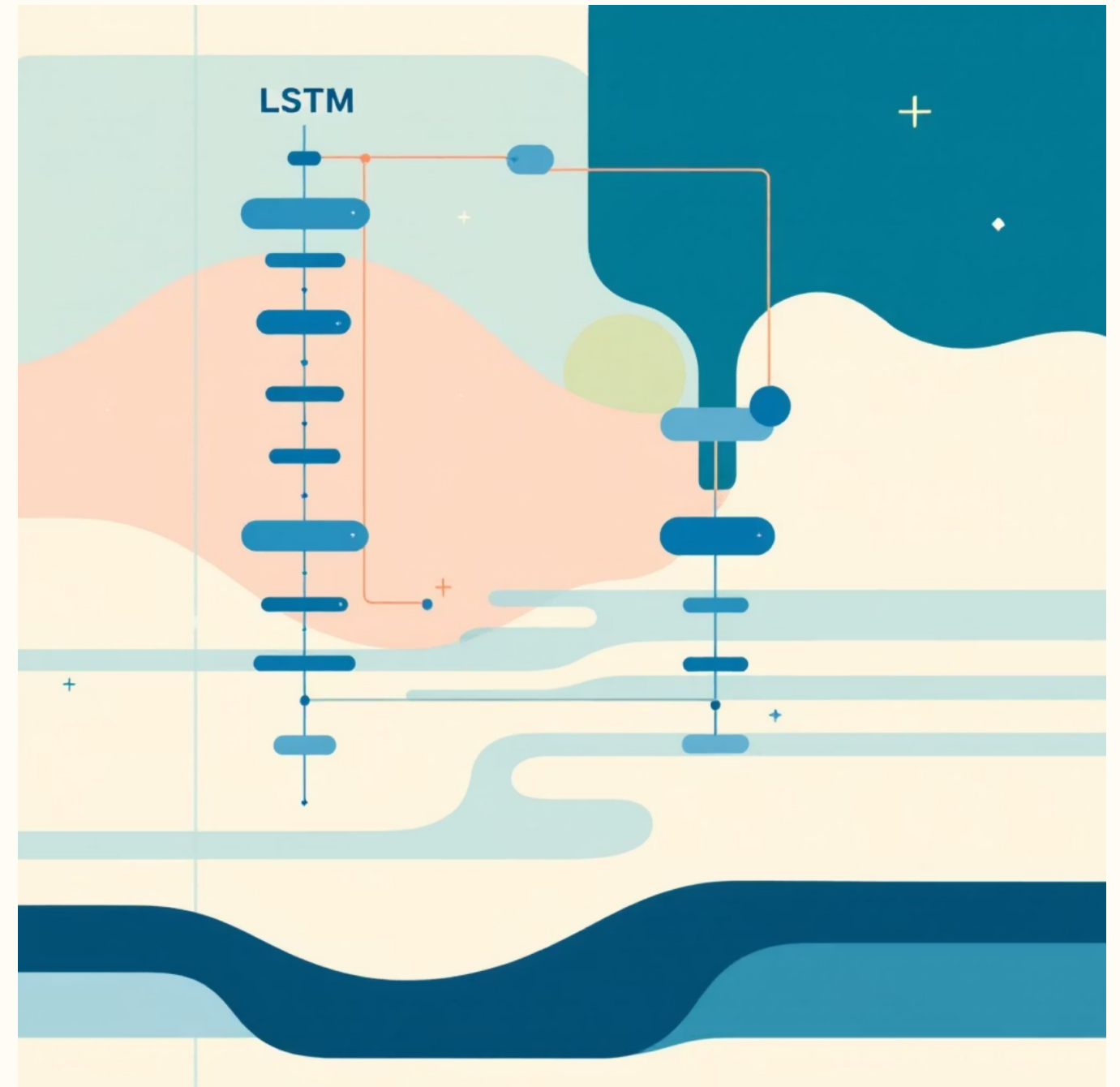
Stacked-LSTM 모델 설계

모델 구조

- 입력: 4차원 특징 벡터 (시퀀스 길이 2)
- LSTM 1: 64개 히든 차원
- LSTM 2: 16개 히든 차원
- 출력: 4개 Hotness 클래스 분류

학습 파라미터

- 배치 크기: 256
- 에폭: 100 (조기 종료 적용)
- 학습률: 0.001



K-means 클러스터링 결과

총 데이터 샘플: 142,221

11.8%

Cold 데이터

16,841개

29.7%

Medium 데이터

42,259개

36.8%

Warm 데이터

52,322개

21.7%

Hot 데이터

30,799개

모델 성능 평가



예측 정확도

테스트 데이터셋에서 달성한 높은 정확도



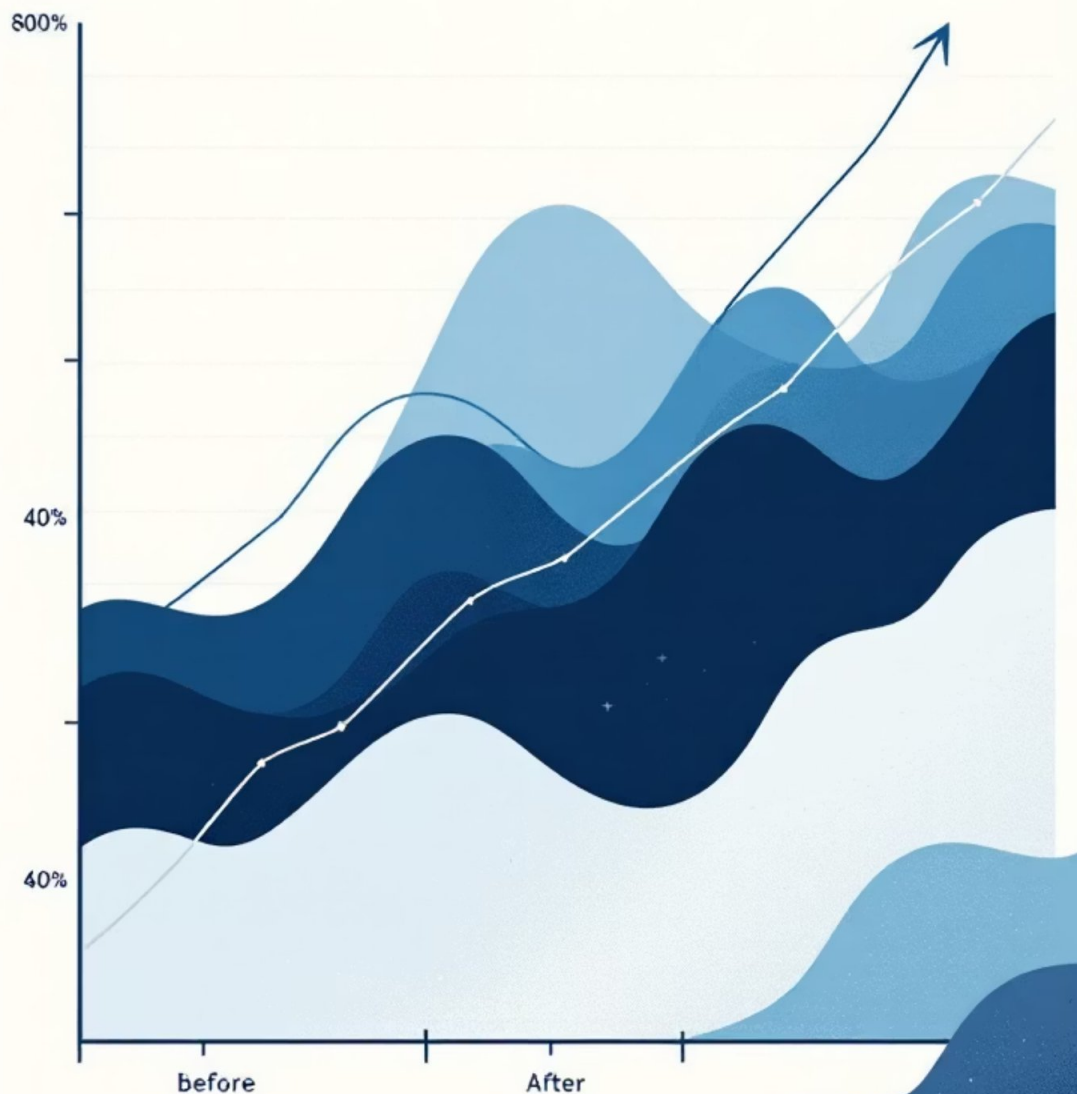
손실값

매우 낮은 손실로 안정적인 학습 완료

Stacked-LSTM 모델은 페이지의 Hotness를 거의 완벽하게 예측하는 뛰어난 성능을 보였습니다. 혼동 행렬 분석 결과 오분류된 샘플 수가 매우 적어 모델의 신뢰성을 확인했습니다.

● performance

improvement



시뮬레이션 결과

99.2%

WAF 감소

136.39 → 1.03으로 쓰기
증폭 감소

1103%

IOPS 증가

4,891 → 58,881로 처리
량 향상

98.2%

응답시간 단축

380,967ns →
6,980ns로 지연시간 감소

99.1%

GC 횟수 감소

24,394회 → 212회로
가비지 컬렉션 감소



결론 및 향후 연구 방향

연구 성과

딥러닝 기반 Hotness 예측으로 SSD 성능과 수명을 동시에 향상시키는 효과적인 방법 제시

핵심 기여

Hot/Cold 데이터 분리 관리를 통한 GC 효율 극대화 및 WAF 최소화 달성

향후 과제

온라인 학습 모델 도입, 현실적 환경에서의 추가 검증, Wear-Leveling과의 통합 연구

본 연구는 딥러닝을 스토리지 시스템에 적용하는 새로운 가능성을 제시하며, 차세대 지능형 SSD 개발의 기반을 마련했습니다.