

Portfolio

MINSEON CHO

조민선

[ABOUT | MINSEON CHO \(C4FFEIN3.GITHUB.IO\)](#)

Profile

◆ Education

- 2016.03 ~ 2021.02 창원대학교 컴퓨터공학과 학사 (졸업)
- 2021.03 ~ 2023.02 창원대학교 컴퓨터공학과 석사 (졸업 예정)

◆ Experience

- 2020.08.03 ~ 2020.08.27 네이버 주식회사 인턴
- 2020.12 ~ Noslab 석사과정 연구원

◆ Language skills

- 영어 TOEIC 925/990 2019.11.24 ~ 2021.11.23
- 영어 OPIc AL 2022.08.27 ~ 2024.08.26

Tech stacks



Language

C/C++

python



Systems

Linux kernel

Ubuntu

Microkernel (seL4)

Bash

Strace

blktrace/blkparse

filefrag



File & Storage

File systems (F2FS)

Fio

Filebench

Sysbench

YCSB



Hadoop ecosystem

HDFS

Spark



AI/ML

Tensorflow

Keras

sklearn

Pandas

Projects

◆ AI/ML 기반 페이지 캐시 교체

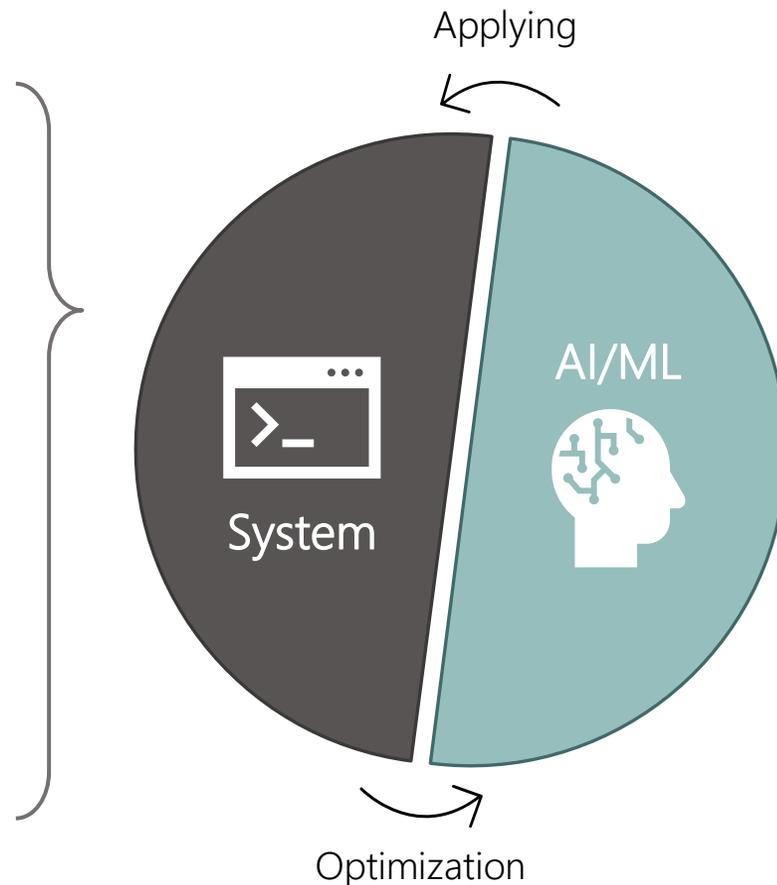
- L2LRU (2021.09) 🏆
- 기계 학습 기반 페이지 교체 정책 (2021.06) 🏆
- ML-CLOCK (2021.10)

◆ F2FS 관련 연구

- RocksDB 환경의 F2FS 파일 시스템 분석 (2021.12) 🏆

◆ 파일 시스템 단편화 관련 연구

- 실시간 파일 시스템 단편화 모니터링 도구 설계 (2022.06)



◆ AI RUSH 2020 🏆

- 네이버 AI 프로젝트 챌린지 AI RUSH (naver.com)
- 본선 진출자 중 임의로 구성된 2인 1조 팀으로 진행 (airushNcash 팀)
- “글의 분위기, 톤, 문법을 통한 종합 평가/추천” 분야 1위

◆ Music for your mood

- <https://music-for-your-mood.github.io>
- 사용자의 입력 텍스트에서 감정을 분석하여 어울리는 음악을 추천해주는 사이트
- Tensorflowjs 사용

◆ 기계학습 최적화 관련 연구

- Interleaved Data Processing Scheme for Optimizing Tensorflow Framework (2021.09)

System Projects

DETAILS

◆ AI/ML 기반 페이지 캐시 교체

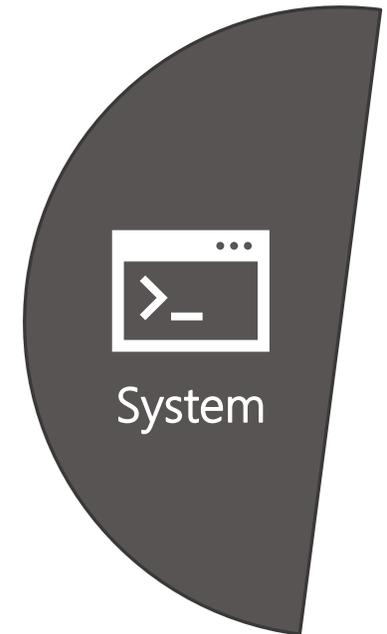
- L2LRU (2021.09) 
- 기계 학습 기반 페이지 교체 정책 (2021.06) 
- ML-CLOCK (2021.10)

◆ F2FS 관련 연구

- RocksDB 환경의 F2FS 파일 시스템 분석 (2021.12) 

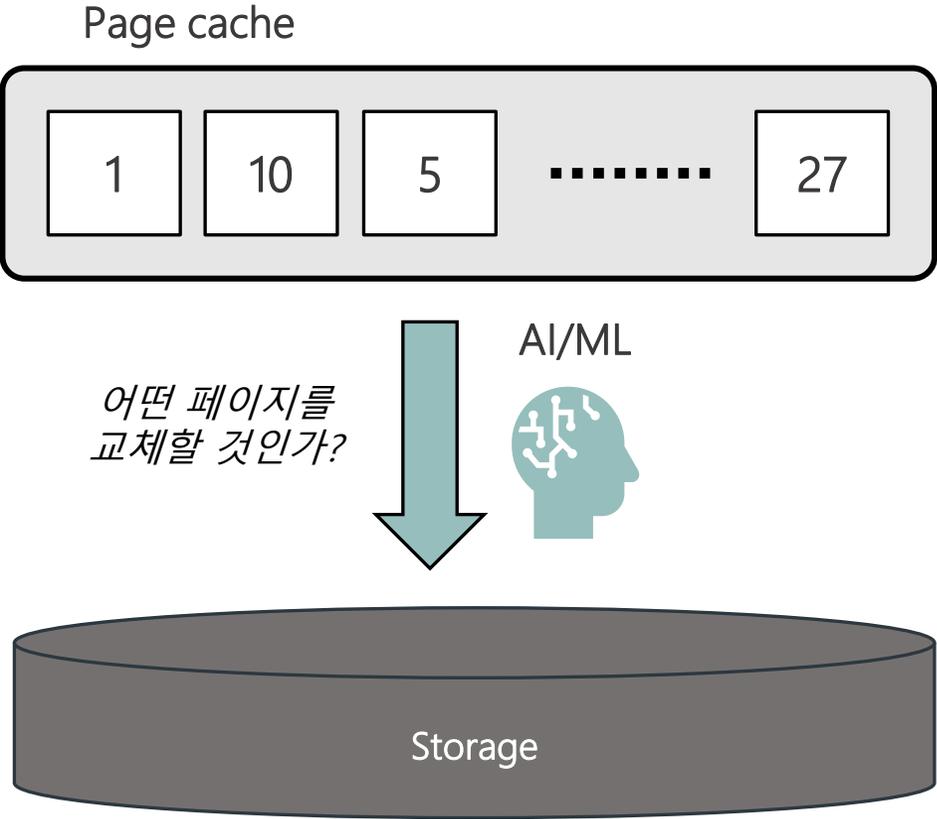
◆ 파일 시스템 단편화 관련 연구

- 실시간 파일 시스템 단편화 모니터링 도구 설계 (2022.06)





AI/ML 기반 페이지 캐시 교체



◆ L2LRU

- 한국정보과학회논문지 (KCI 등재)
- 2022 정보과학회 논문지 시스템 및 이론 분야 우수논문상 🏆
- 페이지의 재사용거리를 ML을 통해 예측하여 페이지 교체에 사용

◆ 기계 학습 기반 페이지 교체 정책

- 한국컴퓨터종합학술대회
- 특허 출원 📄
- 페이지의 추후 재접근 여부를 ML을 통해 예측하여 페이지 교체에 사용

◆ ML-CLOCK

- MDPI Electronics (SCI 등재)
- 캐시 적중 뿐만 아니라 순차 쓰기를 일으켜 플래시 친화적일 수 있도록 고려

L2LRU

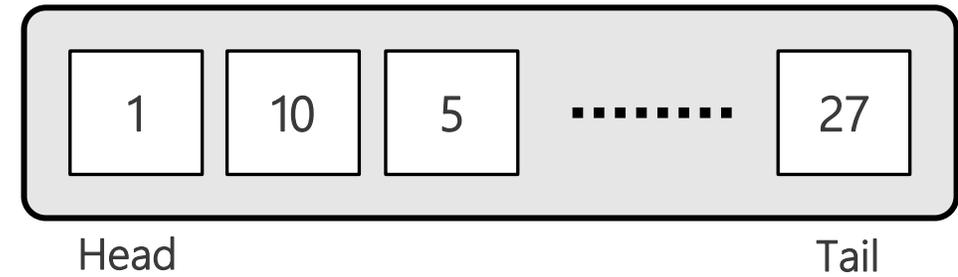
◆ L2LRU

- 조민선, 강동현, "L2LRU: LRU 페이지 교체 기법을 위한 러닝 기반 페이지 이동 정책," *한국정보과학회논문지 (JOK)*, 제 48권, 9호, pp. 981-987, 2021.
- 2022 정보과학회 논문지 시스템 및 이론 분야 우수논문상

◆ Technical skills

- 트레이스 기반 시뮬레이터 형식으로 구현 (C언어)
- 딥러닝 모델 설계 및 학습 (Python, Tensorflow, Keras)
- 학습 데이터 가공 (Python, Pandas)

LRU Page cache queue



재접근된 페이지를 어디로 이동시킬 것인지 결정

L2LRU

◆ 재사용 거리 (Reuse distance)

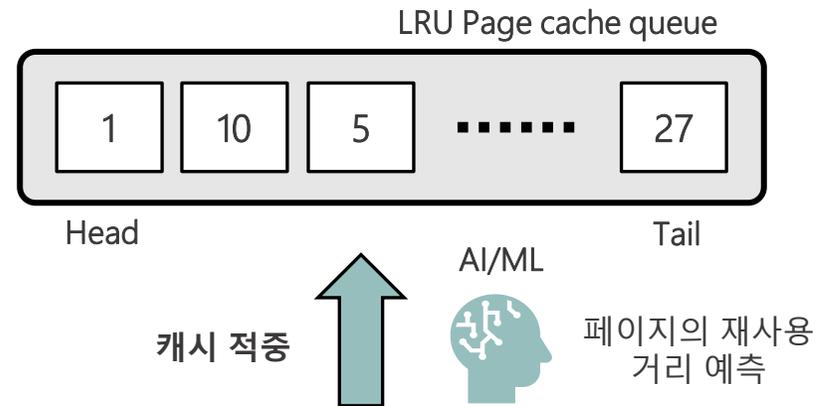
- 페이지가 재접근되기까지의 논리적 거리

Time	1	2	3	4	5	6	7	8	9	10
Page #	1	2	3	4	5	1	6	2	1	1
Reuse distance		←————→			4					
		←————→				5				
						←————→		2		
										0

그림 출처: 조민선, 강동현, "L2LRU: LRU 페이지 교체 기법을 위한 러닝 기반 페이지 이동 정책," 한국정보과학회논문지 (JOK), 제 48 권, 9호, pp. 981-987, 2021.09.

◆ 딥러닝을 이용한 재사용 거리 예측

- 캐시 적중(cache hit)이 발생하였을 때, 해당 페이지의 미래 재사용 거리를 딥러닝으로 예측하여 그에 따라 추가적인 기회를 부여



If 예측된 재사용 거리 > 현재 Tail까지의 거리:
추가적인 기회를 부여할 수 있는 위치로 이동

Else:
페이지를 이동하지 않음

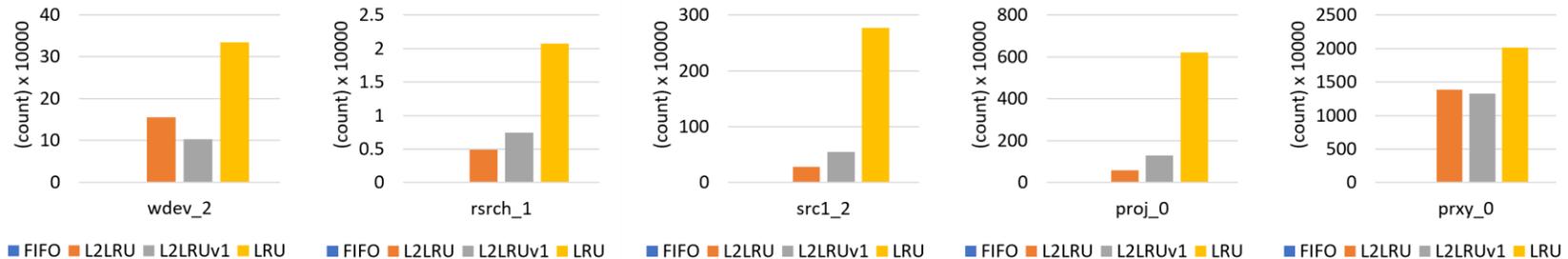
L2LRU

◆ 장점 ✓

- 캐시 적중이 일어났을 때 모든 페이지를 이동시키지 않으므로, **락-언락 명령어 호출로 인한 오버헤드 감소**
- **전기 에너지 소모량 감소**

◆ 단점 ✗

- 기존 LRU 대비 **높지 않은 캐시 적중률**
- 딥러닝 모델의 **연산 오버헤드**
- 딥러닝 모델 **사전학습에 장시간 소모**
- 사전학습을 위한 **다량의 학습 데이터 필요**



락-언락 명령어 호출 횟수

그림 출처: 조민선, 강동현, "L2LRU: LRU 페이지 교체 기법을 위한 러닝 기반 페이지 이동 정책," 한국정보과학회논문지 (JOK), 제 48권, 9호, pp. 981-987, 2021.09.

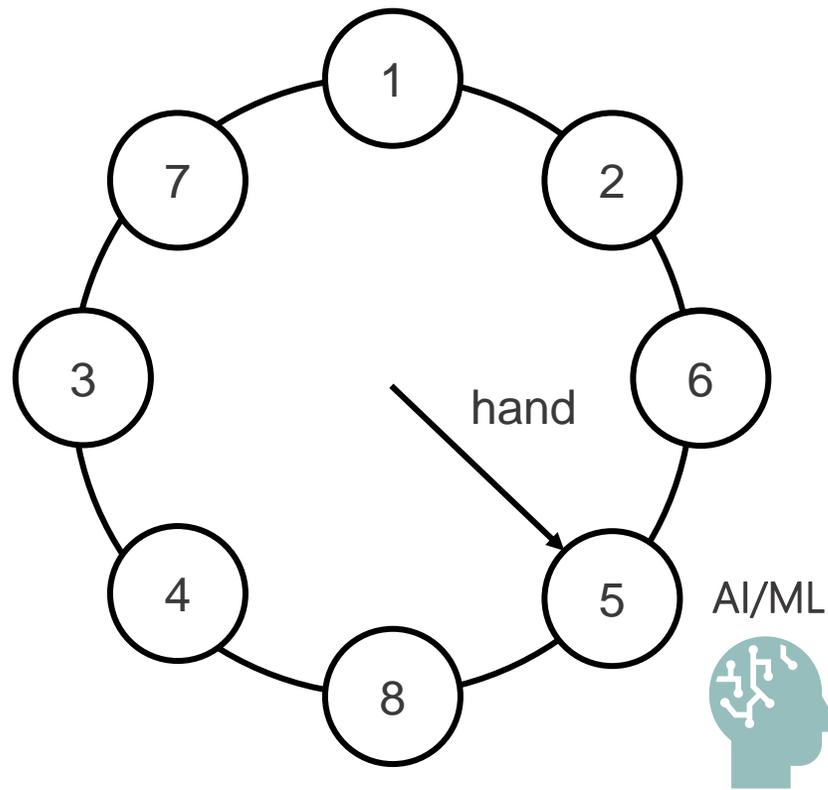
기계 학습 기반 페이지 교체 정책

◆ 기계 학습 기반 페이지 교체 정책

- 조민선, 강동현, "기계 학습 기반 페이지 교체 정책," 2021년 한국컴퓨터종합학술대회 논문집 (KCC), Jun. 2021.
- 특허 출원 (출원번호: 10-2021-0089771)

◆ Technical skills

- 트레이스 기반 시뮬레이터 형식으로 구현 (C언어)
- 퍼셉트론 모델 구현, 학습 및 예측 (C언어)
- 성능 평가 벤치마크 사용 (filebench)
- 실험용 트레이스 생성 (blktrace, blkparse)



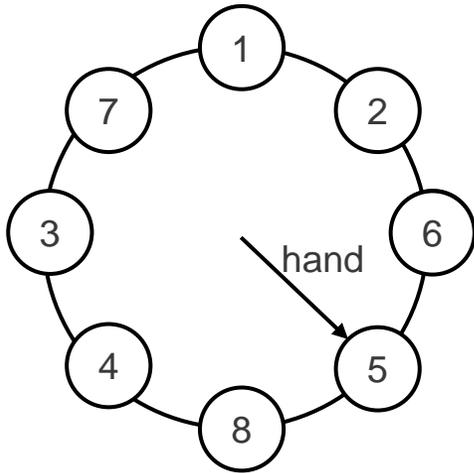
페이지의 재접근 여부 예측



기계 학습 기반 페이지 교체 정책

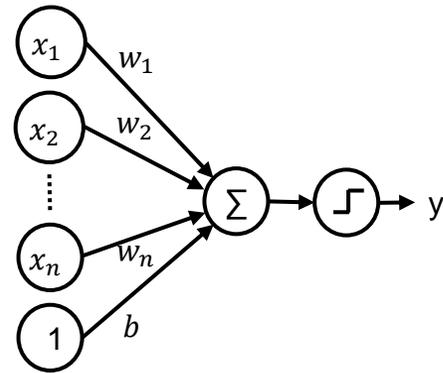
◆ Clock 알고리즘 기반 구조

- Circular queue와 hand를 사용하여 hand와 조우하는 페이지의 교체 여부를 결정
- 교체 여부를 결정하기 위해 ML을 사용해 해당 페이지의 재접근 여부 예측



◆ 온라인 퍼셉트론 모델을 이용한 페이지 재접근 가능성 예측

- 페이지의 최신성과 접근 빈도를 토대로 퍼셉트론이 해당 페이지가 추후 재접근될 것인지를 예측
- 온라인 모델이므로 워크로드 실행 중 워크로드의 패턴을 실시간으로 학습 가능



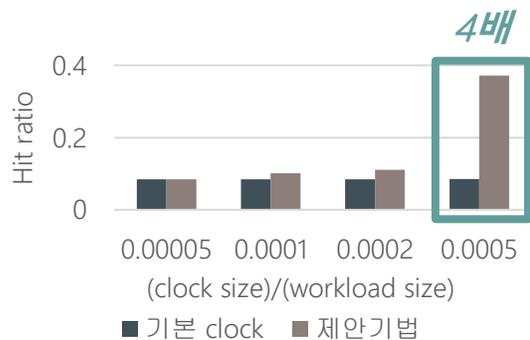
기계 학습 기반 페이지 교체 정책

◆ 장점 ✓

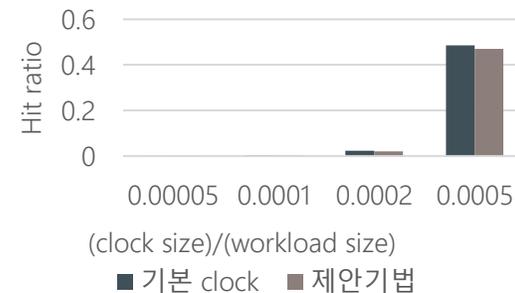
- 기존 clock 알고리즘 대비 **최대 4배의 캐시 적중률**
- 퍼셉트론 모델은 복잡한 구조를 가지고 있지 않으므로 **연산 오버헤드가 적음**
- 온라인 모델은 실시간 학습을 하므로 **사전학습의 오버헤드가 존재하지 않음**

◆ 단점 ✗

- 모든 종류의 워크로드에서 캐시 적중률이 clock 알고리즘을 능가하는 것은 아님
- 충분한 학습이 일어나지 않으면 성능 감소



Varmail



Fileserver

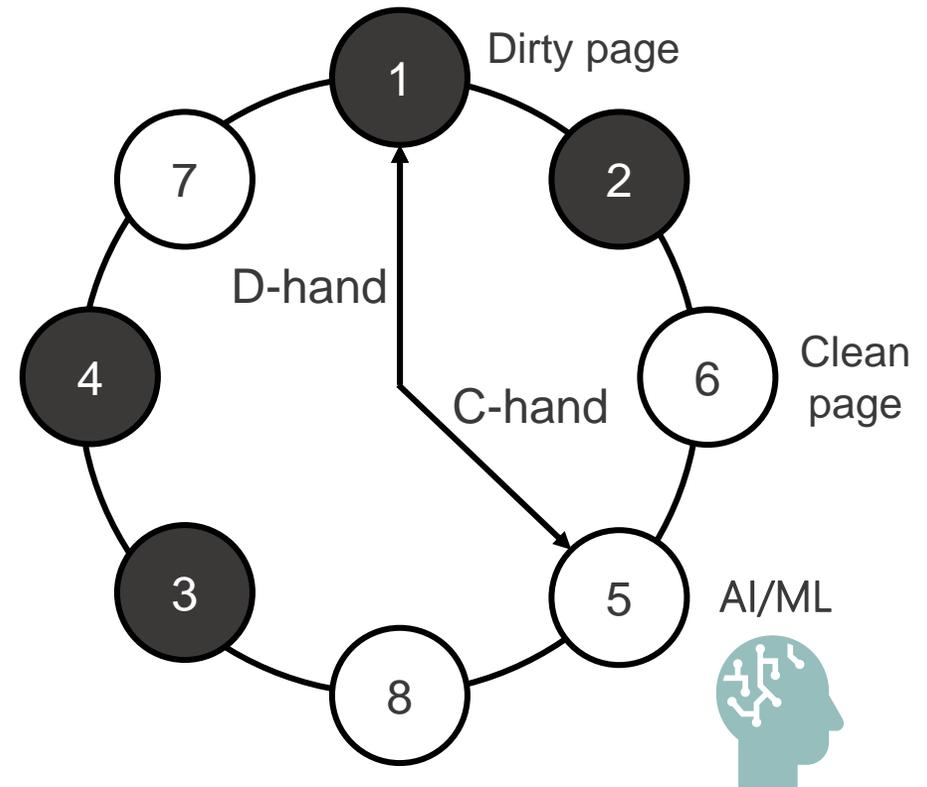
ML-CLOCK

◆ ML-CLOCK

- M. Cho and D. Kang, "ML-CLOCK: Efficient Page Cache Algorithm based on Perceptron-based Neural Network," *MDPI Electronics*, Vol. 10, No. 20, pp. 1-16, Oct. 2021.

◆ Technical skills

- 트레이스 기반 시뮬레이터 형식으로 구현 (C언어)
- 퍼셉트론 모델 구현, 학습 및 예측 (C언어)

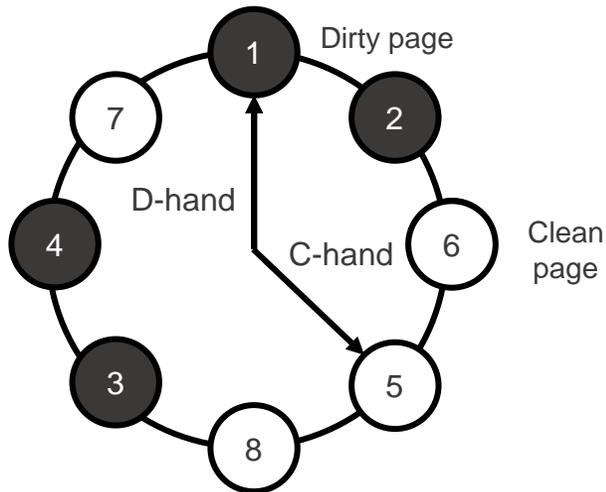


페이지의 재접근 여부 예측

ML-CLOCK

◆ Clock 알고리즘 기반 구조

- 클린 페이지를 위한 C-hand와 더티 페이지를 위한 D-hand가 별도로 존재
- 더티 페이지의 경우 LBA 순으로 정렬
- 순차 쓰기를 일으킬 가능성을 높여 플래시 친화적



◆ 온라인 퍼셉트론 모델을 이용한 페이지 재접근 가능성 예측

- D-hand와 C-hand 모두 기본 clock 알고리즘의 방식으로 각각의 교체 대상 후보 페이지를 선정
- 두 후보 중 어느 페이지를 교체할 것인지를 판단하기 위해 퍼셉트론 사용
- 페이지의 최신성과 접근 빈도를 토대로 퍼셉트론이 해당 페이지가 추후 재접근될 것인지를 예측
- 온라인 모델이므로 워크로드 실행 중 워크로드의 패턴을 실시간으로 학습 가능

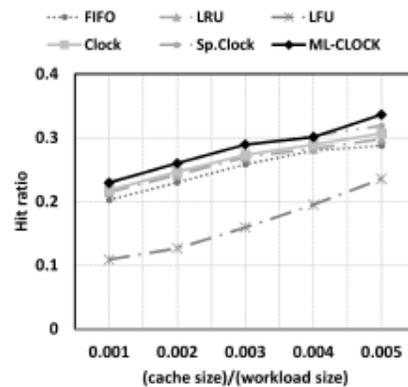
ML-CLOCK

◆ 장점 ✓

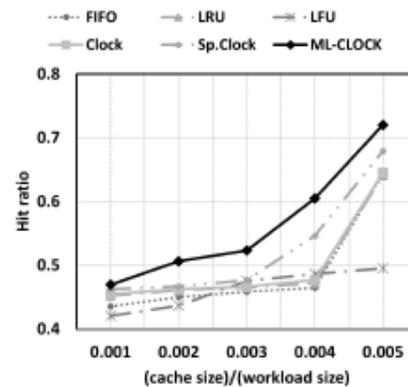
- 순차 쓰기를 고려하여 **플래시 친화적**
- 퍼셉트론 모델은 복잡한 구조를 가지고 있지 않으므로 **연산 오버헤드가 적음**
- 온라인 모델은 실시간 학습을 하므로 **사전학습의 오버헤드가 존재하지 않음**

◆ 단점 ✗

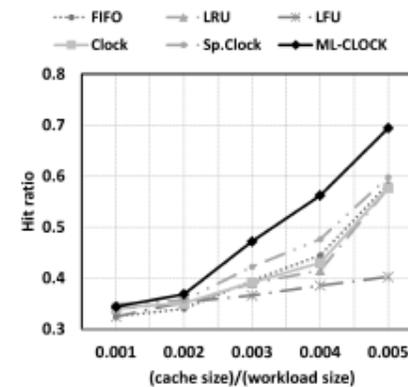
- 충분한 학습이 일어나지 않으면 성능 감소
- 캐시 적중률이 다른 알고리즘과 비교해 충분히 높지 않으면 **퍼셉트론의 컴퓨팅 오버헤드**가 드러날 수 있음



(a) ikki



(b) webmail



(c) web-vm



RocksDB 환경의 F2FS 파일 시스템 분석

◆ RocksDB 환경의 F2FS 파일 시스템 분석

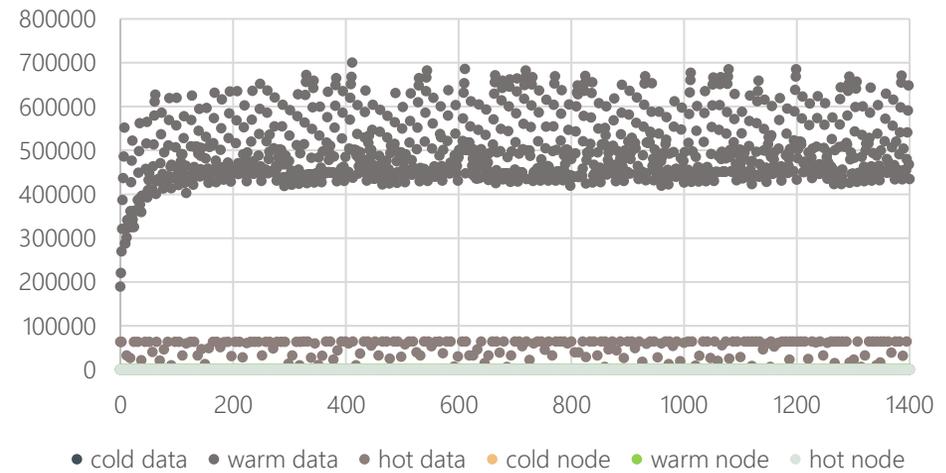
- 조민선, 강동현, "RocksDB 환경의 F2FS 파일 시스템 분석," 2021 전자통신학술대회(부산/울산/경남지부), Dec. 2021.
- 2021 전자통신학술대회(부산/울산/경남지부) 우수논문상

◆ Technical skills

- Linux kernel, F2FS에 대한 지식
- RocksDB에 대한 지식
- 파일 I/O 벤치마크 도구 사용 (YCSB)
- Python

◆ Overview

- RocksDB 실행 시 F2FS의 hot-cold separation (multi-head logging)을 분석
- log가 비효율적으로 사용되는 것을 관측





RocksDB 환경의 F2FS 파일 시스템 분석

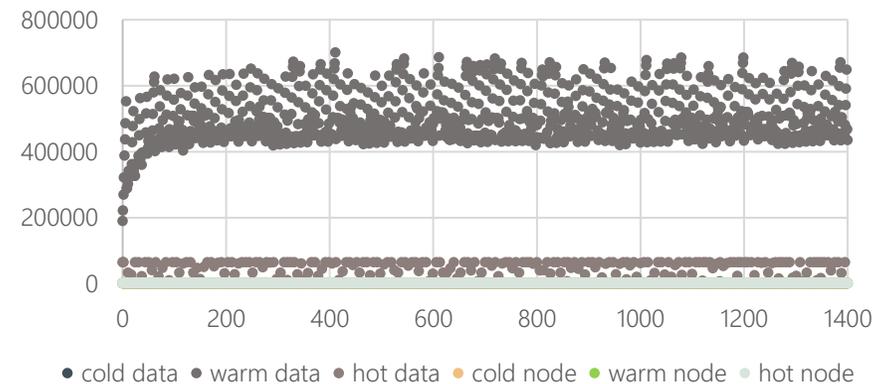
◆ Multi-head logging

- 블록의 종류에 따라 다른 로그에 할당하여 GC의 비용을 감소

Type	Temp.	Objects
Node	Hot	디렉토리의 디렉트 노드
	Warm	일반파일의 디렉트 노드
	Cold	인디렉트 노드
Data	Hot	디렉토리 엔트리 데이터
	Warm	사용자에 의해 생성된 데이터
	Cold	멀티 미디어 파일 데이터 등

◆ 분석 결과

- RocksDB 특성상 빈번한 데이터 수정이 발생시 SST 파일 데이터를 지속적으로 생성
- Warm data가 대부분을 차지하며 동일 로그에 할당된 warm data도 SST 파일의 레벨에 따라 다른 수명을 지니므로 비효율적
- RocksDB는 멀티 미디어 파일은 생성하지 않으므로 cold data 로그가 낭비되고 있음





실시간 파일 시스템 단편화 모니터링 도구 설계

◆ 실시간 파일 시스템 단편화 모니터링 도구 설계

- 조민선, 강동현, "실시간 파일 시스템 단편화 모니터링 도구 설계," 2022년 한국컴퓨터종합학술대회 논문집 (KCC), Jun. 2022.

◆ Technical skills

- Linux kernel, File systems에 대한 지식
- 모니터링 툴 구현 (Bash, AWK)
- 파일 시스템 단편화 모니터링 (filefrag)
- 단편화 데이터 가공 및 시각화 기능 구현 (Python)
- 파일 I/O 벤치마크 도구 사용 (YCSB)
- RocksDB 코드 분석 (RocksDB, C++)
- 어플리케이션 시스템 콜 분석 (strace)

◆ Overview

- 실시간으로 지정된 경로에 위치하는 파일 시스템 단편화 정보를 수집하여 시각화하는 과정을 자동화한 모니터링 툴

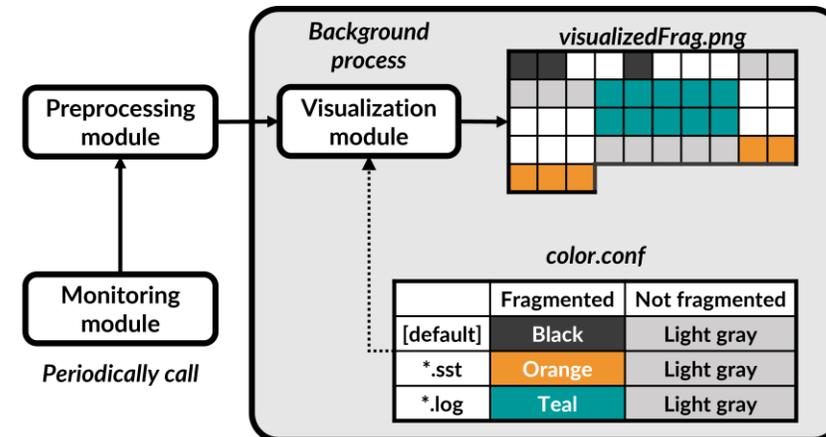


그림 출처: 조민선, 강동현, "실시간 파일 시스템 단편화 모니터링 도구 설계," 2022년 한국컴퓨터종합학술대회 논문집 (KCC), 2022.06.



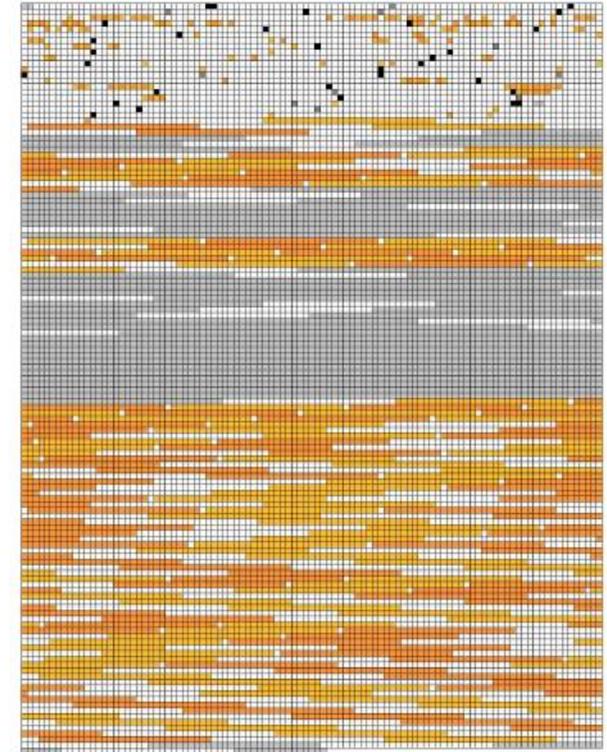
실시간 파일 시스템 단편화 모니터링 도구 설계

◆ 제안기법을 이용한 RocksDB의 쓰기 패턴 분석

- 순차쓰기 위주, 데이터의 수정이 일어나지 않는 플래시 친화적 DB이므로 **파일 시스템 단편화가 거의 일어나지 않을 것으로 예측**
- **예상과는 다른 결과 관측**
- 우측 이미지에서 회색은 **단편화가 일어나지 않은 SST 파일의 데이터 블록**을 의미
- 우측 이미지에서 **주황색**은 **단편화가 일어난 SST 파일의 데이터 블록**을 의미

◆ RocksDB의 쓰기 패턴 원인 추적

- RocksDB에서 **SST 파일 생성 시 일으키는 시스템 콜** 모니터링
- SST 파일 쓰기 중 **fsync**를 호출한 뒤 **ftruncate**를 호출하는 것에 의해 파일의 마지막 데이터 페이지가 더티 페이지가 되며 발생하는 현상

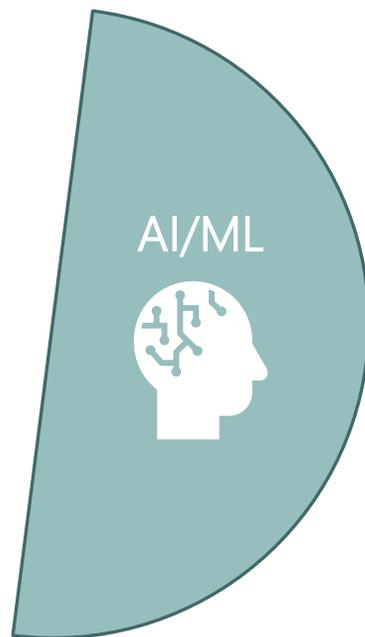


RocksDB 실행 중 파일 시스템 단편화 이미지

그림 출처: 조민선, 강동현, "실시간 파일 시스템 단편화 모니터링 도구 설계," 2022년 한국컴퓨터종합학술대회 논문집 (KCC), 2022.06.

AI/ML Projects

DETAILS



◆ AI RUSH 2020 🏆

- 네이버 AI 프로젝트 챌린지
[AI RUSH \(naver.com\)](https://airush.naver.com)
- 본선 진출자 중 임의로 구성된 2인 1조
팀으로 진행 (airushNcash 팀)
- “글의 분위기, 톤, 문법을 통한 종합
평가/추천” 분야 1위

◆ Music for your mood

- <https://music-for-your-mood.github.io>
- 사용자의 입력 텍스트에서 감정을
분석하여 어울리는 음악을 추천해주는
사이트
- Tensorflowjs 사용

◆ 기계학습 최적화 관련 연구

- Interleaved Data Processing Scheme for
Optimizing Tensorflow Framework
(2021.09)



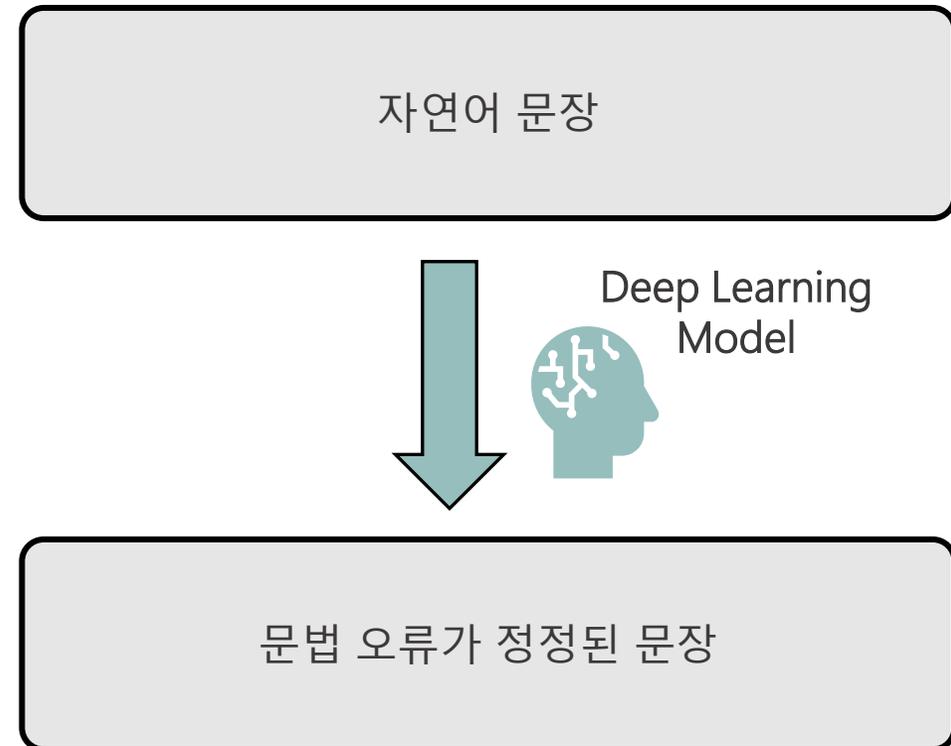
AI RUSH 2020

◆ AI RUSH 2020

- 네이버 AI 프로젝트 챌린지 [AI RUSH \(naver.com\)](http://AI RUSH (naver.com))
- 본선 진출자 중 임의로 구성된 2인 1조 팀으로 진행 (airushNcash 팀)
- “글의 분위기, 톤, 문법을 통한 종합 평가/추천” 분야 1위

◆ Technical skills

- 딥러닝 모델 구현 (Python, Tensorflow, Pytorch, scikit-learn)
- 자연어 처리에 대한 지식
- Language Model에 대한 지식





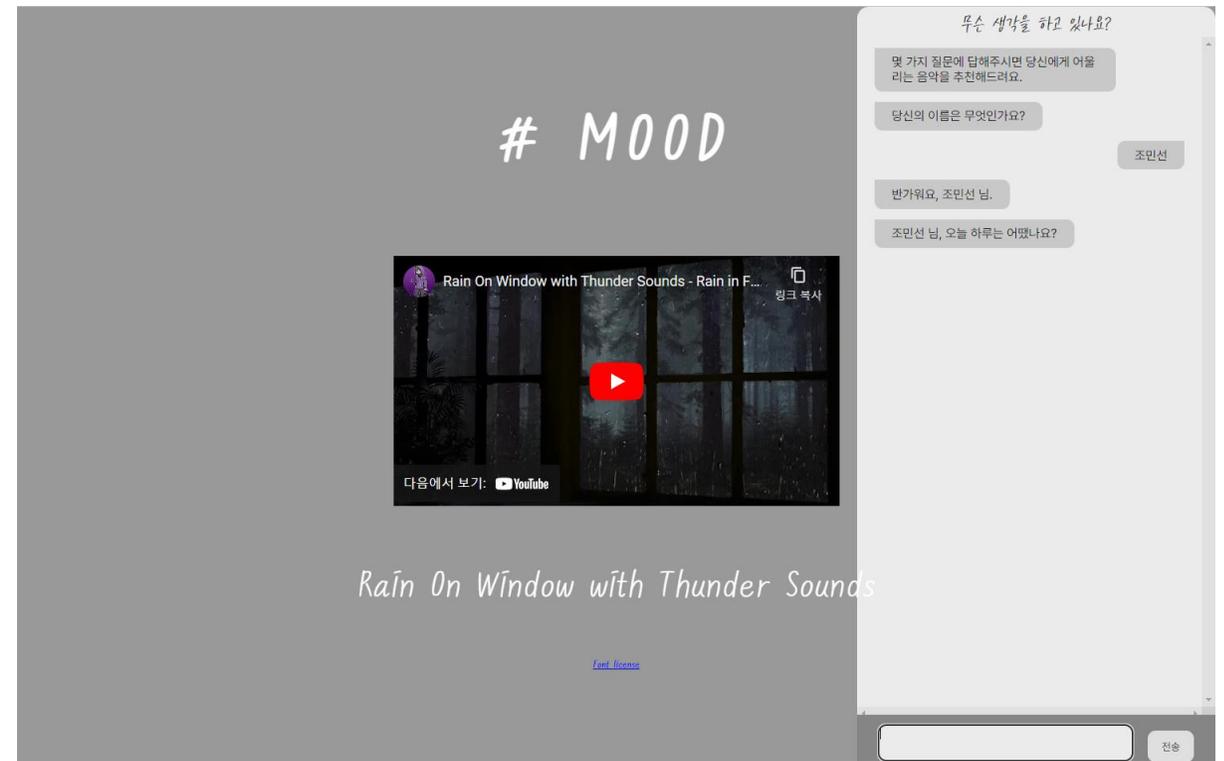
Music for your mood

◆ Music for your mood

- <https://music-for-your-mood.github.io>
- 사용자의 입력 텍스트에서 감정을 분석하여 어울리는 음악을 추천해주는 사이트

◆ Technical skills

- 사이트 구성 (HTML 5, CSS)
- 반응형 기능 구현 (Javascript, Tensorflowjs)
- 딥러닝 모델 디자인 및 학습 (Tensorflow)





기계학습 최적화 관련 연구

◆ 데이터 프로세싱 과정 최적화

- Jinseo Choi, Minseon Cho, Donghyun Kang, "Interleaved Data Processing Scheme for Optimizing Tensorflow Framework," *ICCE Berlin*, pp.1-3, Nov. 2021.
- 제2저자로 구현 관련 조연 및 학술대회 발표 역할을 맡음
- Training과 Inference 단계를 병렬로 진행하여 시간 단축

◆ Technical skills

- 기계학습 코드 구성 (Python, Tensorflow, Keras)
- DL, CNN에 대한 지식

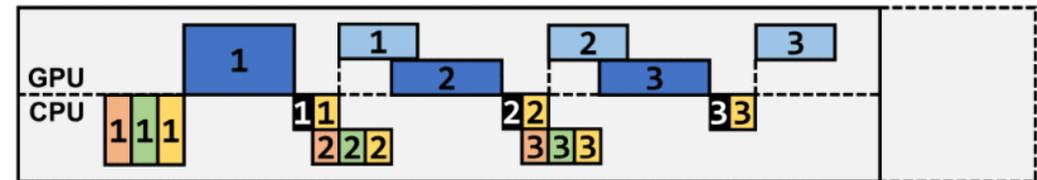
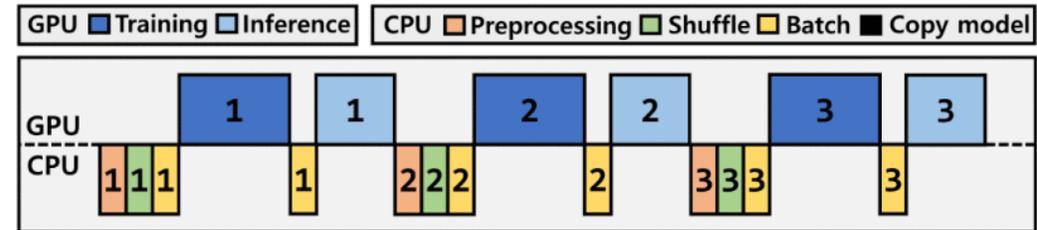


그림 출처: Jinseo Choi, Minseon Cho, Donghyun Kang, "Interleaved Data Processing Scheme for Optimizing Tensorflow Framework," *ICCE Berlin*, pp.1-3, Nov. 2021.

Thank you

읽어주셔서 감사합니다.

Email: bse59123@naver.com

Blog: <https://c4ffein3.github.io/>

Github: <https://github.com/C4FFEIN3>