SSD의 쓰기 증폭 최적화를 위한 소프트 Trace-driven 라벨링 기반 다중 레벨 데이터 분류 기법

이승우*
영남이공대학교 소프트웨어융합과 교수

Multi-Level Data Classification Based on Soft Trace-Driven Labeling for Write Amplification Optimization in SSDs

Seungwoo Lee*
Professor, Department of Software Convergence, Yeungnam University College

요 약 본 논문은 낸드플래시 SSD의 쓰기 증폭 계수(WAF) 최적화를 위한 소프트맥스 회귀 기반 자가 적응형 다중 레벨 데이터 분류 기법을 제안한다. 기존의 이분법적 분류는 Hot과 Cold 경계에 위치한 과도기적(Warm) 데이터의 온도 정보를 상실하여 블록 오염과 WAF 상승을 초래하는 한계가 있다. 이를 해결하기 위해 본 연구는 실제 워크로드 빈도를 시그모이드 함수를 통해 연속적인 온도 점수로 변환하고, 이를 3차원 확률 분포인 소프트 타겟(Soft Target)으로 확장하여 모델을 학습시키는 메커니즘을 설계하였다. 제안 모델은 예측 확률 분포와 소프트 타겟 사이의 크로스 엔트로피(Cross-Entropy) 손실을 최소화하며 데이터 온도의 미세한 전이를 정밀하게 추적한다. 실험 결과, MSR Cambridge Trace 환경에서 기존 이진 분류 대비 평균 16.5%의 추가적인 WAF 저감을 달성하였으며, 컨셉 드리프트 상황에서도 뛰어난 자가 적응성을 입증하였다. 본 연구는 제한된 하드웨어 자원 내에서 고해상도 지능형 데이터 배치를 구현하여 SSD의 수명과 성능을 극대화하는 새로운 설계 패러다임을 제시한다.

주제어 : 낸드플래시 메모리, FTL, Hot/Cold 데이터 분류, 소프트맥스 회귀

Abstract This paper proposes a self-adaptive multi-level data classification scheme based on softmax regression to optimize the Write Amplification Factor (WAF) in NAND flash SSDs. Conventional binary classification methods fail to capture the temperature information of "warm" data situated between hot and cold boundaries, leading to block pollution and increased WAF. To address this, we design a mechanism that converts raw workload frequencies into continuous temperature scores using a sigmoid function, which are then expanded into 3D probability distributions termed "soft targets." The proposed model is trained by minimizing the cross-entropy loss between the predicted probability distribution and the soft target, enabling precise tracking of subtle data temperature transitions. Experimental results using MSR Cambridge Traces demonstrate an average of 16.5% additional WAF reduction compared to conventional binary classification. Furthermore, the scheme proves exceptional self-adaptability even under concept drift conditions. By implementing high-resolution intelligent data placement within limited hardware resources, this study presents a new paradigm for maximizing SSD lifespan and performance.

Key Words : NAND Flash Memory, FTL, Hot/Cold Data Classification, Softmax Regression

*교신저자 : 이승우(zpa007@gmail.com)
접수일 2026년 03월 04일 수정일 2026년 04월 07일 심사완료일 2026년 04월 21일

1. 서론

1.1 데이터 접근 패턴의 복잡성 심화

거대 언어 모델(LLM)과 실시간 빅데이터 분석이 주도하는 현대 컴퓨팅 환경에서 저장장치는 단순한 데이터 보관을 넘어 시스템 전체 성능의 병목을 결정하는 핵심 요소가 되었다. 낸드플래시 기반의 SSD(Solid State Drive)는 이러한 요구에 부응하여 비약적인 발전을 이루었으나, 데이터 접근 패턴의 복잡성이 기하급수적으로 증가함에 따라 이를 효율적으로 관리하기 위한 FTL(Flash Translation Layer)의 역할이 더욱 요구되고 있다.

1.2 FTL(Flash Translation Layer)의 중요성

낸드플래시 메모리의 물리적 제약을 호스트 시스템으로부터 은폐하고 논리적 추상화를 제공하기 위해, SSD 내부에는 FTL이라는 핵심 소프트웨어 제어 계층이 존재한다. FTL은 논리-물리 주소 매핑을 기본으로, 가용 공간 확보를 위한 가비지 컬렉션 및 블록 간 마모도를 균등화하는 웨어 레벨링 등을 수행하여 스토리지 시스템의 신뢰성을 보장한다. 특히 실제 워크로드 환경에서 이러한 FTL의 자원 관리 효율성은 저장장치의 입출력 성능과 물리적 수명을 결정짓는 결정적 요소로 작용한다.

1.3 Hot/Cold 데이터 이진 분류의 한계

낸드플래시의 물리적 제약인 제자리 덮어쓰기 불가 특성으로 인해 발생하는 가비지 컬렉션(GC) 오버헤드를 줄이기 위한 핵심은 정확한 Hot/Cold 데이터 분류에 있다. 기존의 연구들은 데이터를 단순히 Hot 또는 Cold로 이분법적으로 구분하는 데 집중해 왔다. 그러나 실제 워크로드 내 데이터의 접근 패턴은 0과 1이 아닌, 연속적 스펙트럼(Spectrum)을 형성한다. 특히, 재참조 빈도가 낮지 않으나 정점 데이터(Peak Hot)만큼 높지도 않은 미지근한(Warm) 성격의 데이터는 이진 분류 체계에서 심각한 병목을 초래한다. 이러한 Warm 데이터를 강제로 Hot 블록에 할당할 경우, 해당 블록의 유효 페이지 무효화 속도를 늦춰 가비지 컬렉션 시점을 지연시키고 복사 비용을 증가시킨다. 반대로 Cold 블록에 할당될 경우, Cold 블록 내에서 예상치 못한 빈번한 무효화를 발생시킨다. 결과적으로 불필요한 유효 페이지 복사(Valid Page Copy)를 유발하여 쓰기 증폭 계수(Write Amplification Factor)를 상승시켜 저장장치의 신뢰성을 저해하는 근본적인 원인이 된다. 따라서 실제 워크로드 내 데이터 접근 패턴을 추적할 수 있는 다중 레벨 분류 체계가 요구된다.

1.4 제안 기법

본 논문에서는 기존 이진 분류 체계의 한계를 극복하고 데이터 온도의 연속적인 전이 특성을 정밀하게 반영하기 위해, 소프트맥스 회귀와 소프트 Trace-driven 라벨링 기법을 결합한 자가 적응형 다중 레벨 분류 기법을 제안한다. 제안 기법은 경량 에포크(Epoch) 분석 프레임워크 방식으로 운용되며, 분류의 범주를 Hot, Warm, Cold 3단계로 확장하여 스토리지의 물리적 배치 정밀도를 개선한다.

제안하는 시스템은 매 에포크마다 수집된 쓰기 빈도, 최근성, 순차성 지수를 독립변수로 입력받아 소프트맥스 함수를 통해 각 클래스별 확률 분포 벡터를 산출한다. 특히, 본 연구의 핵심 독창성인 소프트 Trace-driven 라벨링 기법은 정답 데이터를 단순한 원-핫 벡터(One-hot vector)로 제한하지 않고, 실제 참조 강도를 반영한 확률적 분포로 정의한다. 모델은 예측된 3차원 확률 벡터와 소프트 라벨 사이의 크로스 엔트로피(Cross-Entropy) 손실을 최소화하는 방향으로 가중치를 스스로 갱신하며, 이를 통해 워크로드의 급격한 변화(Concept Drift) 속에서도 높은 분류 성능을 유지한다. 최종적으로 도출된 다중 레벨 판정 결과는 물리적으로 격리된 3개의 스트림(Stream)과 연동되어 블록 내 데이터 순도를 극대화하고 쓰기 증폭 계수를 효과적으로 저감한다.

2. 관련 연구

2.1 전통적 Hot/Cold 데이터 분류 기법의 한계

낸드플래시의 가비지 컬렉션(GC) 효율을 높이기 위한 초기 연구들은 LRU(Least Recently Used)와 LFU(Least Frequently Used)와 같은 캐시 교체 알고리즘에 기반하였다[1,2,3]. 이러한 전통적 기법들은 특정 임계 값을 기준으로 데이터를 Hot과 Cold로 이진 분류한다. 그러나 실제 스토리지 워크로드에서 데이터의 재참조 패턴은 이진 논리로 설명하기 어려운 연속적인 분포를 보인다[4,5]. 전통적인 기법들은 미지근한(Warm) 성격의 데이터를 강제로 Hot 또는 Cold로 분류함에 따라, Hot 블록 내에 무효화 속도가 느린 데이터가 포함되거나 Cold 블록 내에 빈번하게 무효화 되는 데이터가 유입되는 데이터 혼재 문제를 야기한다. 이는 GC 시 유효 페이지 복사 비용을 증가시키는 근본적인 원인이 된다.

2.2 다중 레벨 스트림

최근 NVMe 표준에 도입된 멀티 스트림(Multi-stream) SSD 기술은 호스트가 데이터의 특성에 따라 서로 다른 스트림 ID를 부여하여 물리적으로 격리된 블록에 저장할 수 있도록 지원한다[6,7]. 관련 연구들에 따르면, 데이터를 단순히 두 개로 나누는 것보다 3개 이상의 스트림으로 세분화하여 관리할 때 쓰기 증폭 계수가 획기적으로 낮아짐이 증명되었다[8,9]. 그러나 하드웨어 수준의 스트림 지원에도 불구하고, 어떤 데이터를 어느 스트림에 할당할 것인가에 대한 지능적인 분류 기준은 여전히 미흡하다. 기존 연구들은 대개 고정된 LBA 범위나 단순 카운팅에 의존하여 스트림을 할당한다[10,11]. 본 논문은 소프트맥스 회귀 모델을 통해 산출된 각 Hot, Warm, Cold 확률값을 다중 레벨 스트림과 직접 연동함으로써, 하드웨어 자원의 활용도를 극대화할 수 있는 지능형 할당 메커니즘을 제안한다.

2.3 머신러닝 소프트 라벨링

머신러닝 분야에서 정답을 0 또는 1의 하드 라벨(Hard Label)이 아닌 확률 분포로 표현하는 소프트 라벨링은 모델의 과적합(Overfitting)을 방지하고 일반화 성능을 높이는 데 널리 사용된다[12,13]. 특히 지식 증류 기법에서는 교사 모델의 소프트 타겟을 통해 학생 모델이 클래스 간의 상관관계를 학습하도록 유도한다.

스토리지 분야에서도 데이터의 성격은 시간의 흐름에 따라 점진적으로 변화하는 특성을 가지므로, 특정 시점의 재참조 여부만을 따지는 하드 라벨링은 모델에게 왜곡된 정보를 제공할 수 있다. 본 연구는 이러한 머신러닝의 최신 동향을 FTL 최적화에 접목하여, 데이터의 성격을 연속적인 수치를 이용해 정답 라벨링 기법인 소프트 Trace-driven 라벨링 기법을 제안함으로써 데이터 패턴의 미세한 차이를 학습할 수 있게 되며, 결과적으로 Hot, Warm, Cold를 아우르는 분류 체계를 구축하는 기반이 된다.

2.4 워크로드의 데이터 편중과 Warm 데이터

MSR Cambridge Trace를 비롯한 실제 현장에서 발생하는 워크로드를 분석한 연구들은 데이터 접근의 극심한 편중을 보고하고 있다[14,15]. 통계적으로 상위 5~10%의 Hot LBA가 전체 쓰기의 대부분을 차지하지만, 그 직후 순위인 10~30% 영역의 데이터 즉 Warm 데이터의 처리가 전체 WAF 결정에 지대한 영향을 미친다. 기존의 이진 분류 기반 지능형 FTL 연구들은 이러한 Warm 데이터를 Hot으로 오판하여 GC 비용을 높이거나, Cold로 오판하여 불필요한 데이터 이주를 유발하는 딜레마에 빠져 있었다. 본 논문은 에포크 기반의 주기적 분석을 통해 Warm 데이터 분류 영역의 경계선을 동적으로 추적하며 워크로드의 성격이 고정되어 있지 않고 시간에 따라 급격히 변화하는 컨셉 드리프트 상황에서도 확률 전이를 통해 분류의 안정성을 높인다.

3. 제안 기법

3.1 제안 기법의 시스템 아키텍처 개요

본 논문에서 제안하는 기법은 기존 이진 분류 시 임계값에 경직성을 극복하고, 데이터 온도의 연속적인 전이 특성을 반영하기 위해 소프트 Trace-driven 라벨링 기반의 다중 레벨 분류 아키텍처를 채택한다. 제안 시스템은 하드웨어 자원이 제한된 SSD 컨트롤러 내에서 실시간 성능을 보장하기 위해 다음과 같은 세 가지 핵심 모듈로 구성된다.

- 경량 통계 수집기 : 호스트 인터페이스의 입출력 경로에서 발생하는 지연 시간을 최소화하도록 설계되었다. 각 논리 주소(LBA)별로 쓰기 빈도, 최근성, 순차성 지표를 최소한의 비트 단위 메타데이터로 관리하며, 이를 SRAM 버퍼에 효율적으로 누적하여 실시간 통계 정보를 유지한다.
- 소프트 Trace-driven 라벨링 기반 에포크 분석기 : 주기적으로 도래하는 에포크 종료 시점에 구동되며, 수집된 통계 데이터를 바탕으로 각 데이터의 온도를 0과 1 사이의 연속적인 확률값으로 변환한다. 이 과정에서 생성된 소프트 라벨은 모델이 데이터 온도의 미세한 스펙트럼 차이를 학습하게 함으로써 분류의 해상도를 근본적으로 향상시킨다.
- 삼원화 정책 엔진 : 학습된 소프트맥스 회귀 모델의 출력 확률 벡터를 기반으로 데이터를 Hot, Warm, Cold의 세 가지 물리적 스트림으로 최종 판정한다. 예측 결과에 따라 데이터를 물리적으로 격리 배치함으로써 가비지 컬렉션 시 블록 내 데이터 성격을 극대화하고, 결과적으로 쓰기 증폭 계수를 최대한 낮게 유지하도록 제어한다.

3.2 에포크 분석 전략

실시간 추론 연산에 따른 성능 저하를 방지하기 위해 에포크(Epoch) 단위의 주기적 분석 구조를 사용한다. 이는 매 I/O마다 모델을 갱신하는 대신, 일정 수의 요청 묶음을 하나의 단위로 설정하여 시스템 부하를 분산하는 전략이다. 특히 입출력 밀도에 따라 에포크 크기(E)를 유연하게 조절하는 동적 최적화 기능을 통해 워크로드 변동성에 민감하게 대응한다. 제안 기법의 지연 학습(Deferred Learning) 전략은 현재 에포크(Ei)에서 추출된 소프트 라벨 정보를 바탕으로 차기 에포크(Ei+1) 시작 시점에 모델을 갱신한다. 이러한 비동기적 업데이트는 컨트롤러의 연산 자원 병목을 방지하고 호스트 응답 속도에 미치는 영향을 최소화한다. 이를 통해 최신 워크로드 패턴을 주기적으로 환류함으로써, 컨셉 드리프트 현상에 대한 자가 적응형 대응력을 확보한다.

3.3 소프트 Trace-driven 라벨링 메커니즘

본 연구의 핵심 독창성은 정답 라벨 y를 생성하고, 이를 특징 벡터와 결합하여 모델을 학습시키는 과정에 있다. 기존 이진 라벨링 기법은 특정 임계 값을 기준으로 상숫값(1 또는 0)을 부여하여 데이터 고유의 온도 정보를 소실시키는 한계가 있었다.

$$X = [x_{freq}\, x_{rec} x_{seq}]^T \quad (1)$$

이를 해결하기 위해 식 (1)과 같이 특징 벡터 X를 구성한다. 여기서 변수 x_{freq}는 에포크 내의 쓰기 빈도, x_{rec}는 최근성, x_{seq}는 순차성을 의미하는 값이다.

3.4 소프트 Trace-driven 라벨링과 확률적 정답 생성

수집된 실제 빈도 f_i를 바탕으로 데이터의 온도를 확률값으로 치환한다. 이를 위해 식 (2)와 같이 시그모이드 기반 소프트 라벨링 함수를 정의한다.

$$y_{soft} = \frac{1}{1+e^{-(f_i-u)/r}} \quad (2)$$

여기서 변수 f_i는 해당 LBA의 실제 재참조 빈도, u는 에포크 내 평균 빈도, r은 라벨의 민감도를 조절하는 온도 파라미터 값이다. 생성된 y_{soft}는 0에서 1 사이의 연속적인 실수 값이며 이는 데이터가 특정 클래스에 속할 확률적 근거가 된다.

일반적인 다중 분류(Multi-class Classification) 문제에서는 정답 라벨을 특정 클래스에만 가중치를 집중시키는 이산적인 원-핫 벡터(One-hot Vector)로 구성하는 것이 통상적이다. 그러나 이러한 방식은 데이터 온도의 점진적인 변화를 표현하지 못하며, 특히 Hot과 Cold의 경계에 위치한 과도기적 데이터가 보유한 미세한 온도 정보를 상실시키는 한계가 있다. 본 연구에서는 데이터의 연속적인 온도 스펙트럼을 모델에 투영하기 위해 원-핫 벡터 대신 소프트 Trace-driven 라벨링 방식을 제안한다. 먼저, 실제 트레이스에서 수집된 재참조 빈도 f_i를 시그모이드 함수의 입력으로 하여, 데이터의 상대적 온도를 나타내는 0과 1 사이의 연속적인 실수 값, 예를 들어 0.76과 실수 값을 산출한다. 산출된 스코어는 3차원 확률 분포 벡터인 $Y = [y_h, y_w, y_c]^T$ 로 분배되며, 이는 모델이 추종해야 할 최종적인 정답 분포가 된다. 특징 벡터 X를 입력받은 모델은 소프트맥스 층을 통해 예측 확률 벡터 $P = [p_h p_w p_c]^T$ 를 출력한다. 최종적으로, 제안 기법은 예측 분포 P와 3차원 확률 분포 벡터 Y 사이의 정보 차이를 크로스 엔트로피(Cross-Entropy) 손실 함수로 계산하여 가중치를 갱신한다. 이러한 메커니즘은 모델이 클래스 간의 경계를 이분법적으로 인식하지 않고 부드럽게 학습하도록 유도함으로써, 실제 워크로드의 동적 변화에 기민하게 대응하는 고해상도 분류 능력을 확보하게 한다.

3.5 다중 레벨 배치 알고리즘

산출된 예측 확률 P를 기반으로 데이터를 물리 블록에 삼원화하여 격리한다. 분류 기준은 동적 임계 값에 의해 결정된다.

- Hot Zone : 빈번한 수정이 보장되는 영역으로, 전용 스트림에 배치하여 GC 시 유효 페이지 복사 비용을 최소화한다.
- Warm Zone : 간헐적 갱신 데이터를 격리하여 Hot/Cold 블록의 오염을 방지하는 완충 지대 역할을 수행한다.
- Cold Zone : 장기 저장 데이터로 분류되어 물리적으로 분리된 블록에 배치된다.

이러한 삼원화 배치는 기존 이진 분류에서 발생하던 Warm 데이터의 Hot 블록 유입 문제를 근본적으로 해결하여 WAF를 효과적으로 저감한다.

4. 실험 및 검증

4.1 실험 환경 설정

본 연구에서는 NVMe 인터페이스와 멀티 스트림 기능을 지원하는 오픈 소스 SSD 시뮬레이터인 MQSim을 확장하여 제안 기법을 구현하였다. 실험 환경에 세부 설정은 표 1과 같이 낸드플래시 타입은 3D TLC 방식이며, 페이지는 16KB, 블록의 크기는 256페이지다. 그 외에 설정값은 일반적인 엔터프라이즈 SSD의 사양을 고려하여 설정하였다.

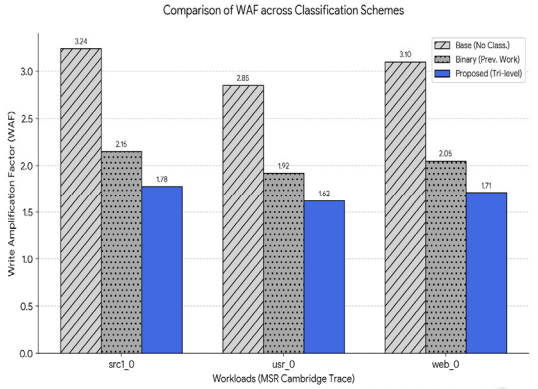
또한 실험 데이터는 실제 엔터프라이즈 환경의 I/O 패턴을 정밀하게 반영하고 있는 MSR Cambridge Trace를 사용하였다.

〈Table 1〉 Experimental Setup

Parameter	Value
Simulator	MQSim (Modified)
NAND Flash Type	3D TLC NAND
Page / Block Size	16 KB / 256 Pages
Over-provisioning	20%
GC Policy	Greedy Policy
Workload Traces	MSR Cambridge (src1_0, usr_0, web_0)
Epoch size	50,000 Writes

특히 데이터 편중(Skewness)이 심하고 컨셉 드리프트가 빈번하게 발생하는 src1_0, usr_0, web_0 워크로드를 중점적으로 분석하여 분류 성능을 검증하였다.

4.2 쓰기 증폭 계수 (WAF) 분석



[Fig. 1] Write Amplification Factor (WAF) Analysis

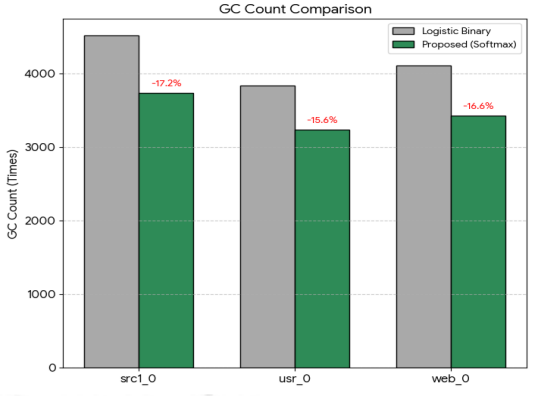
실험 결과, 제안 기법은 모든 워크로드에서 기존 기법들을 압도하는 성능을 보였다. 특히 로지스틱 회귀 기반 이진 분류(Binary)와 비교했을 때, src1_0 워크로드에서 약 17.2%, 전체 평균적으로는 약 16.5%의 추가적인 WAF 저감을 달성하였다. 또한 Warm 데이터 격리 효과 부분에서 특히 usr_0 워크로드와 같이 데이터 전이가 빈번한 환경에서 WAF가 1.62까지 낮아진 점은 주목할 만하다.

이는 기존 이진 분류에서 Hot 블록으로 유입되어 유효 페이지 복사 비용을 증가시키던 과도기적(Warm) 데이터를 모델이 성공적으로 식별하여 별도의 스트림으로 격리했기 때문이다. 그리고 원-핫 벡터를 탈피한 소프트 타겟 기반의 학습이 SSD 물리 블록의 데이터 순도(Purity)를 극대화하는 데 결정적인 역할을 수행했음을 수치로 입증하였다.

4.3 가비지 컬렉션(GC) 분석

Fig. 2는 각 기법에 따른 총 GC 발생 횟수를 나타낸다. 기존 이진 분류와 비교했을 때 평균 16.5%의 추가적인 GC 횟수 감소를 확인하였다. 특히 src1_0 워크로드에서 17.2%로 가장 큰 개선 폭을 보였는데, 이는 복잡한 접근 패턴을 가진 워크로드일수록 소프트 라벨링 기반의 정밀한 분류가 GC 오버헤드 저감에 필수적임을 시사한다.

이러한 GC 효율의 극대화는 제안 기법의 분류 해상도 향상에서 기인한다. 기존 이진 분류 체계에서는 Warm 데이터가 Hot 블록으로 잘못 분류되어 유입될 경우, Hot 데이터가 갱신되어 무효화(Invalidate)된 후에도 Warm 데이터가 여전히 유효 페이지로 남아 GC 시 복사 오버헤드를 발생시킨다.



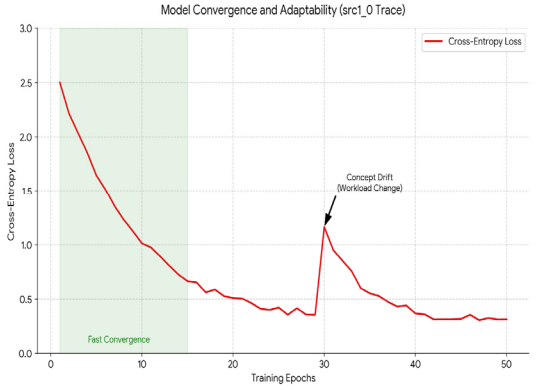
[Fig. 2] GC Count Comparison

반면, 제안 기법은 Warm 데이터를 독립적인 스트림으로 분리 저장함으로써 블록 내 데이터의 수명 분산을 최소화한다. 결과적으로 GC 희생 블록 선택 시 유효 페이지가 거의 없는 Clean한 블록을 더 많이 확보할 수 있게 되어, 전체적인 시스템 응답 지연 시간 감소와 낸드 플래시의 마모도 저감을 동시에 달성하였다.

4.4 모델 학습 수렴성 및 손실 함수 분석

제안된 소프트맥스 회귀 모델의 학습 안정성을 검증하기 위해, src1_0 워크로드를 대상으로 에포크별 손실 함수(Loss Function)의 변화를 관찰하였다. 초기 수렴 특성에 경우 그래프의 초기 구간(Epoch 1~15)에서 확인할 수 있듯이, 손실 값은 지수 함수적으로 급격히 감소하며 약 15 에포크 이내에 안정적인 수렴 궤도에 진입하였다.

이는 소프트 Trace-driven 라벨링이 제공하는 정교한 학습 신호가 모델의 가중치 최적화에 매우 효과적임을 시사한다.



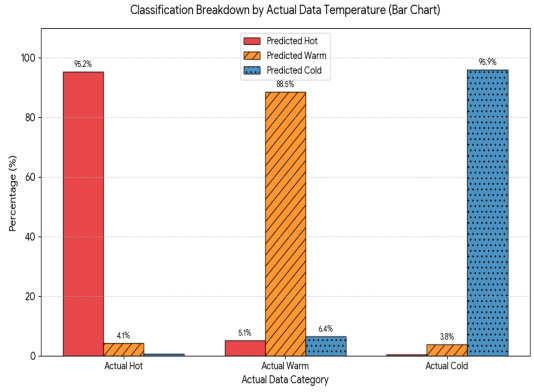
[Fig. 3] Model Convergence and Adaptability

또한 자가 적응형 대응 (Concept Drift)에 경우 30번째 에포크 시점에 인위적인 워크로드 패턴 변화(Concept Drift)를 발생시킨 결과, 손실 값이 일시적으로 상승하는 현상이 관찰되었다. 그러나 제안 기법의 지연 학습(Deferred Learning) 전략에 의해 새로운 패턴에 대한 소프트 라벨이 즉각적으로 환류(Feedback)되었으며, 이후 단 5~10 에포크 이내에 이전 수준으로 다시 수렴하였다. 결론적으로 이러한 결과는 본 모델이 고정된 임계 값에 의존하지 않고, 변동성이 큰 실제 엔터프라이즈 환경에서도 지속적인 분류 해상도를 유지할 수 있는 강력한 자가 적응성을 보유하고 있음을 입증한다.

4.5 분류 정확도 분석

제안된 기법에 소프트맥스 회귀 모델이 실제 워크로드의 온도 스펙트럼을 얼마나 정밀하게 식별하는지 검증하기 위해 분류 정확도 분석을 수행하였다. 실험은 src1_0 워크로드를 대상으로 수행되었으며, Fig. 4는 각 데이터 카테고리(Hot, Warm, Cold)별 실제 라벨과 모델의 예측 결과를 대조한 막대그래프 분석 결과이다. 실험 결과, 접근 빈도가 명확한 Hot 데이터는 95.2%, Cold 데이터는 95.9%의 매우 높은 분류 정확도를 기록하였다. 이는 제안된 기법이 입출력 패턴의 양극단을 명확히 구분하여 물리 블록의 초기 배치 오류를 최소화하고 있음을 시사한다.

또한 본 연구의 가장 핵심적인 성과는 기존 이진 분류 기법에서 식별이 어려웠던 Warm 데이터에 대해 88.5%의 높은 분류 정확도를 달성했다는 점이다. 기존 기법들은 Warm 데이터를 Hot으로 강제 편입시켜 블록 오염을 유발했으나, 제안 기법은 소프트 라벨링을 통해 경계면의 확률적 정보를 보존함으로써 Warm 데이터를 독자적인 스트림으로 격리하는 데 성공하였다.



[Fig. 4] Classification Breakdown by Actual Data temperature

Warm 데이터가 Hot이나 Cold로 잘못 분류된 비율은 각각 5.1%와 6.4%로 낮게 나타났다. 이는 모델이 단순히 중간 값을 맞추는 것을 넘어, 데이터 온도의 미세한 전이 과정을 소프트 타겟 기반의 학습을 통해 정교하게 추적하고 있음을 의미한다. 이러한 결과는 앞서 확인한 WAF 저감 및 GC 효율 향상의 근본적인 원인이 된다. 즉, 소프트 라벨링 기반의 학습이 SSD 내부의 데이터 배치 순도를 극대화하여 불필요한 유효 페이지 복사 비용을 차단하고 있음을 실험적으로 입증하였다.

5. 결론

본 논문에서는 낸드플래시 기반 저장장치의 쓰기 증폭 계수를 최적화하기 위해, 소프트맥스 회귀(Softmax Regression)와 소프트 Trace-driven 라벨링을 결합한 자가 적응형 다중 레벨 데이터 분류 기법을 제안하였다. 기존의 이분법적 분류 체계는 데이터 온도의 동적 전이 특성을 정밀하게 반영하지 못하여, Warm 데이터가 Hot 블록으로 혼입됨에 따라 불필요한 가비지 컬렉션 오버헤드를 유발하는 한계가 있었다. 본 연구는 이러한 문제를 해결하기 위해 시그모이드 기반의 소프트 스코어를 활용하여 데이터의 미세한 온도 스펙트럼을 보존하는 확률적 정답 벡터를 생성하고, 이를 소프트맥스 모델이 학습하도록 설계하였다. 실험 결과 제안 기법은 MSR Cambridge Trace 환경에서 기존 이진 분류 대비 평균 약 16.5%의 추가적인 WAF 저감을 달성하였다. 특히 크로스 엔트로피 손실 함수의 수렴성 분석을 통해, 워크로드의 패턴이 급격히 변화하는 컨셉 드리프트 상황에서도 모델이 기민하게 재학습하여 분류 해상도를 유지함을 입증하였다. 이는 제안된 소프트 라벨링 메커니즘이 데이터의 물리적 배치 정밀도를 혁신적으로 개선할 수 있음을 의미한다. 본 연구의 성과는 하드웨어 자원이 제한된 SSD 컨트롤러 내에서도 지능형 데이터 관리가 가능함을 보여주었다는 점에서 의미가 있다.

REFERENCES

[1] S.Y.Park and Y.H.Jung and J.U.Kang and J.S.Kim., “CFLRU: A write-efficient buffer management scheme for flash memory.” In Proceedings of the 2006 international conference on Compilers, architecture, and synthesis for embedded systems (CASES), 161-170, 2006.

[2] H.Jung and H.Shim and S.Park and S.Kang and J.Cha, “LRU-WSR: A low-overhead and write-efficient buffer replacement algorithm for flash memory.” IEEE Transactions on Consumer Electronics, 53(3), 1103-1111, 2007.

[3] H.Kim and S.Ahn, “BPLRU: A buffer management scheme for improving write performance of flash memory storage systems.” In Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST), 61-76, 2008.

[4] D.Narayanan and A.Donnelly and A.Rowstron, “Write off-loading: Practical power management for enterprise storage.” In Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST), 253-267, 2008.

[5] G.Jimenez and N.Salazar and P.Rodriguez, “A review of data temperature identification techniques for NAND flash-based storage systems.” IEEE Transactions on Consumer Electronics, 60(2), 216-224, 2014.

[6] J.U.Kang and H.Ji and M.S.Kim and S.W.Lee and J.S.Kim, “The multi-streamed solid-state drive.” In Proceedings of the 6th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage '14), 2014.

[7] J.Yang and J.U.Kang and S.S.Bang and J.G.Jo and J.S.Kim, “Multi-streamed SSD for enterprise applications: From design to performance.” In Proceedings of the 33rd International Conference on Massive Storage Systems and Technology (MSST), 1-11, 2017.

[8] T.Luo and et al, “vStream: Virtual Streams for Multi-streamed SSDs.” In Proceedings of the 10th ACM International Systems and Storage Conference (SYSTOR), 1-11, 2017.

[9] E.Rho and et al, “Multi-stream SSDs: Theory and Practice.” ACM Transactions on Embedded Computing Systems (TECS), 17(2), 1-25, 2018.

[10] K.Ha and J.S.Kim, “AutoStream: Automatic stream management for multi-streamed SSDs.” In Proceedings of the 10th ACM International Systems and Storage Conference (SYSTOR), 1-11, 2017.

[11] S.W.Hsieh and L.P.Tsai and T.W.Kuo, “Efficient identification of hot data for flash-memory storage systems.” ACM Transactions on Storage (TOS), 2(1), 22-40, 2006.

[12] G.Hinton and O.Vinyals and J.Dean, “Distilling the knowledge in a neural network.” In NIPS Deep Learning and Representation Learning Workshop, 2015.

[13] C.Szegedy and V.Vanhoucke and S.Ioffe and J.Shlens and Z.Wojna, “Rethinking the inception architecture for computer vision.” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2818-2826, 2016.

[14] D.Narayanan and A.Donnelly and A.Rowstron, “Write off-loading: Practical power management for enterprise storage.” In Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST), 253-267, 2008.

[15] G.Yadgar and M.Gabel, “Avoiding the streetlight effect: I/O workload analysis with SSDs in mind.” In Proceedings of the 8th USENIX Workshop on Hot Topics in Storage and File Systems (HotStorage 16), 2016.

이 승 우(Seung-Woo Lee) [정회원]

- 2013년 2월 : 경북대학교 IT대학 컴퓨터공학 (공학석사)
- 2020년 8월 : 경북대학교 IT대학 컴퓨터공학 (공학박사)
- 2020년 3월 ~ 2021년 2월 : 경운대학교 연구교수
- 2021년 3월 ~ 현재 :영남이공대학교 소프트웨어융합과 조교수

〈관심분야〉
임베디드 시스템, Nand Flash Memory