

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0105453

(43) 공개일자 2022년07월27일

(51) 국제특허분류(Int. Cl.)

G06F 1/3234 (2019.01) G06F 3/06 (2006.01)

(52) CPC특허분류

G06F 1/3275 (2013.01)

G06F 3/0616 (2013.01)

(21) 출원번호 10-2021-0008172

(22) 출원일자 2021년01월20일

심사청구일자 2021년01월20일

(71) 출원인

서강대학교산학협력단

서울특별시 마포구 백범로 35 (신수동, 서강대학교)

(72) 발명자

김영재

서울특별시 마포구 백범로 35 서강대학교 AS909

박중언

서울특별시 마포구 백범로 35 서강대학교 AS911

(74) 대리인

이철희

전체 청구항 수 : 총 14 항

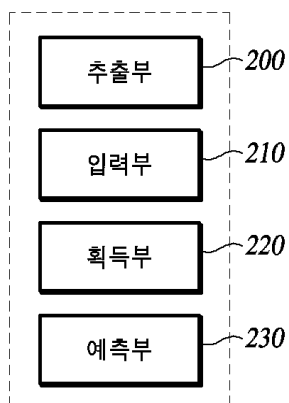
(54) 발명의 명칭 메모리 프로파일링 장치 및 그의 동작 방법

(57) 요약

메모리 프로파일링 장치 및 그의 동작 방법을 개시한다.

본 발명의 일 측면에 의하면, 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 예측하는 메모리 프로파일링 장치에 있어서, 애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 추출부; 상기 접근 정보를 학습된 예측 모델에 입력하는 입력부; 상기 예측 모델로부터 상기 메모리 오브젝트의 스케일링 비율을 획득하는 획득부; 및 상기 애플리케이션의 워크로드가 변화한 경우, 상기 스케일링 비율을 상기 접근 정보에 적용함으로써, 상기 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 예측부를 포함하는 메모리 프로파일링 장치를 제공한다.

대표도 - 도2

20

(52) CPC특허분류

G06F 3/0622 (2013.01)

G06F 3/0625 (2013.01)

G06F 3/0653 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711104176
과제번호	2017M3C4A7080243
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	차세대정보·컴퓨팅기술개발
연구과제명	단일 서버에서 에너지/발열 최적 시스템 SW 개발
기 여 율	1/1
과제수행기관명	고려대학교
연구기간	2017.11.01 ~ 2021.03.31

명세서

청구범위

청구항 1

워크로드(workload) 변화에 따른 메모리 오브젝트의 접근 정보(access pattern)를 예측하는 메모리 프로파일링(profiling) 장치의 동작 방법에 있어서,

애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 과정;

상기 접근 정보를 학습된 예측 모델에 입력하는 과정;

상기 예측 모델로부터 상기 메모리 오브젝트의 스케일링 비율(scaling rate)을 획득하는 과정; 및

상기 애플리케이션의 워크로드가 변화한 경우, 상기 스케일링 비율을 상기 접근 정보에 적용함으로써, 상기 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 과정

을 포함하는 메모리 프로파일링 장치의 동작 방법.

청구항 2

제1항에 있어서,

상기 접근 정보는,

상기 메모리 오브젝트의 크기(size), 메모리 접근 크기(accessed volume) 및 수명(lifetime) 중 적어도 하나를 포함하는 메모리 프로파일링 장치의 동작 방법.

청구항 3

제1항에 있어서,

상기 추출하는 과정은,

상기 메모리 오브젝트의 식별자를 오브젝트-레벨에서 추출하는 과정; 및

상기 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 시간 중 적어도 하나를 명령어-레벨에서 추출하는 과정

을 포함하는 메모리 프로파일링 장치의 동작 방법.

청구항 4

제1항에 있어서,

상기 학습된 예측 모델은,

학습용 메모리 오브젝트들의 접근 정보들을 입력 받고 출력 스케일링 비율을 출력하도록 학습된 것인 메모리 프로파일링 장치의 동작 방법.

청구항 5

제1항에 있어서,

상기 학습된 예측 모델은,

선형 회귀(Linear Regression) 모델, K-최근접 이웃 회귀(K-Nearest Neighbor Regression) 모델 및 랜덤 포레스트 회귀(Random Forest Regression) 모델 중 어느 하나인 메모리 프로파일링 장치의 동작 방법.

청구항 6

제1항에 있어서,

상기 학습된 예측 모델이 랜덤 포레스트 회귀 모델인 경우, 상기 학습된 예측 모델의 트리(tree)의 수는 학습용 메모리 오브젝트들의 스케일링 비율들에 기초하여 결정되는 것인 메모리 프로파일링 장치의 동작 방법.

청구항 7

제6항에 있어서,

상기 학습된 예측 모델의 트리의 수는,

상기 학습용 메모리 오브젝트들의 스케일링 비율들이 불규칙적일수록 높게 결정되는 것인 메모리 프로파일링 장치의 동작 방법.

청구항 8

워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 예측하는 메모리 프로파일링 장치에 있어서,

애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 추출부;

상기 접근 정보를 학습된 예측 모델에 입력하는 입력부;

상기 예측 모델로부터 상기 메모리 오브젝트의 스케일링 비율을 획득하는 획득부; 및

상기 애플리케이션의 워크로드가 변화한 경우, 상기 스케일링 비율을 상기 접근 정보에 적용함으로써, 상기 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 예측부

를 포함하는 메모리 프로파일링 장치.

청구항 9

제8항에 있어서,

상기 접근 정보는,

상기 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 중 적어도 하나를 포함하는 메모리 프로파일링 장치.

청구항 10

제8항에 있어서,

상기 추출부는,

상기 메모리 오브젝트의 식별자를 오브젝트-레벨에서 추출하고, 상기 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 시간 중 적어도 하나를 명령어-레벨에서 추출하는 것인 메모리 프로파일링 장치.

청구항 11

제8항에 있어서,

상기 학습된 예측 모델은,

학습용 메모리 오브젝트들의 접근 정보들을 입력 받고 출력 스케일링 비율을 출력하도록 학습된 것인 메모리 프로파일링 장치.

청구항 12

제8항에 있어서,

상기 학습된 예측 모델은,

선형 회귀 모델, K-최근접 이웃 회귀 모델 및 랜덤 포레스트 회귀 모델 중 어느 하나인 메모리 프로파일링 장치.

청구항 13

제8항에 있어서,

상기 학습된 예측 모델이 랜덤 포레스트 회귀 모델인 경우, 상기 학습된 예측 모델의 트리의 수는 학습용 메모리 오브젝트들의 스케일링 비율들에 기초하여 결정되는 것인 메모리 프로파일링 장치.

청구항 14

제13항에 있어서,

상기 학습된 예측 모델의 트리의 수는,

상기 학습용 메모리 오브젝트들의 스케일링 비율들이 불규칙적일수록 높게 결정되는 것인 메모리 프로파일링 장치.

발명의 설명

기술 분야

[0001] 본 발명의 실시예들은 메모리 오브젝트의 에너지 소모량을 추정하기 위해, 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 예측 모델을 통해 예측하는 머신러닝 기반 메모리 프로파일링 장치 및 그의 동작 방법에 관한 것이다.

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 발명에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 최근 데이터 처리량이 다양한 분야에서 매년 기하급수적으로 증가하고 있다. 이러한 데이터 증가를 수용하기 위해, 기업 또는 기관들은 최적화된 데이터 센터를 채용하고 있다. 하지만, 데이터 센터에서 데이터를 저장하고 분석하는 것은 많은 에너지를 소모하고 방대한 양의 온실가스를 배출한다.

[0004] 데이터 센터에서 에너지 소모는 주로 두 가지 요인에 의해 결정된다. 한 가지는 중앙처리장치(Central Processing Unit; CPU)이고, 나머지 하나는 메인 메모리(main memory)이다. CPU들은 단일 서버에서 소요되는 총 에너지의 30 퍼센트 내지 60 퍼센트에 해당하는 에너지를 소모하고, 메인 메모리는 28 퍼센트 내지 40 퍼센트의 에너지를 소모한다.

[0005] 메모리-레벨에서의 에너지 소모를 줄이기 위해, 메모리 뱅크의 전력을 낮추는 방법, 기본 메모리 전압 및 주파수를 제어하는 방법, 휘발성 메모리와 비휘발성 메모리를 함께 사용하는 하이브리드 메모리 시스템 등 다양한 기술들이 연구되고 있다. 더불어, 메모리 시스템에서는 메모리-레벨에서의 에너지 효율을 개선하기 위해 세분화된 오브젝트-레벨(object-level)에서의 접근 정보(access information)를 추가적으로 고려한 소프트웨어 기반 솔루션이 개발되고 있다. 이러한 소프트웨어 기반 솔루션을 이용하기 위해서는, 메모리 오브젝트로부터 접근 정보를 추출하고, 접근 정보로부터 에너지 소모량을 추정할 필요가 있다. 여기서, 메모리 오브젝트의 접근 정보를 추출하는 과정을 메모리 프로파일링(memory profiling)이라 한다. 메모리 프로파일링을 통한 에너지 소모량 추정 방법은 한국 등록공고번호 KR 2178931 B1에 개시되어 있다.

[0006] 하지만, 애플리케이션의 오브젝트-레벨 메모리 프로파일링은 시간이 오래 걸리고, 에너지를 소모한다는 문제점이 있다. 게다가, 애플리케이션의 워크로드(workload) 크기가 변화할 때 오브젝트의 접근 정보도 함께 변화하므로, 워크로드 크기가 변화할 때마다 메모리 프로파일링을 수행해야 한다는 문제점이 있다.

[0007] 최근 메모리 프로파일링 시간을 줄이되 메모리 오브젝트의 접근 정보를 예측하기 위해, 선형 스케일링 비율(Linear Scaling Rate; LSR) 모델이 연구되었다. LSR 모델은 애플리케이션의 워크로드가 변화할 때 메모리 오브젝트의 접근 정보들이 선형적으로 변화하는 것으로 추정한다. 예를 들면, 애플리케이션이 크기가 N인 워크로드와 수명이 L인 오브젝트를 가지는 경우, 워크로드 크기가 두 배만큼 증가하면 오브젝트의 수명도 선형적으로 두 배만큼 증가하는 것으로 추정한다. 이를 통해, 워크로드의 크기가 변화할 때마다 접근 정보를 추출하는 것이 아니라, 스케일링 비율을 이전 접근 정보에 적용하여 변화된 워크로드에 대응되는 접근 정보를 예측하는 것이다.

[0008] 하지만, LSR 모델은 워크로드와 메모리 오브젝트가 선형관계임을 전제로 하므로, 다양한 애플리케이션에 적용되는 경우 예측 오류를 포함할 수 있다. 메모리 오브젝트의 접근 정보들과 스케일링 워크로드 사이의 상호관계가 적을수록, LSR 모델에 기초하여 예측된 메모리 오브젝트의 접근 정보들은 더 많은 오류율을 포함하게 된다. 즉, 애플리케이션의 워크로드가 증가할 때 메모리 오브젝트의 접근 정보들은 선형적으로 증가하지 않는 것이다.

[0009] 따라서, 워크로드의 크기가 변화할 때 스케일링 비율을 이용하여 접근 정보를 예측하여 시간 및 에너지 소모를

줄이되, 워크로드와 메모리 오브젝트가 다양한 상호관계를 가지더라도, 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 정확하게 예측할 수 있는 방법에 대한 연구가 필요하다.

발명의 내용

해결하려는 과제

[0010] 본 발명의 실시예들은, 워크로드 변화에 따른 메모리 프로파일링에 드는 시간과 에너지 소모를 줄이기 위한 머신러닝 기반 메모리 프로파일링 장치 및 그의 동작 방법을 제공하는 데 주된 목적이 있다.

[0011] 본 발명의 다른 실시예들은, 워크로드와 메모리 오브젝트 간 선형관계 여부에 상관없이 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 정확하게 예측하기 위한 메모리 프로파일링 장치 및 그의 동작 방법을 제공하는 데 일 목적이 있다.

과제의 해결 수단

[0012] 본 발명의 일 측면에 의하면, 워크로드(workload) 변화에 따른 메모리 오브젝트의 접근 정보(access pattern)를 예측하는 메모리 프로파일링(profiling) 장치의 동작 방법에 있어서, 애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 과정; 상기 접근 정보를 학습된 예측 모델에 입력하는 과정; 상기 예측 모델로부터 상기 메모리 오브젝트의 스케일링 비율(scaling rate)을 획득하는 과정; 및 상기 애플리케이션의 워크로드가 변화한 경우, 상기 스케일링 비율을 상기 접근 정보에 적용함으로써, 상기 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 과정을 포함하는 메모리 프로파일링 장치의 동작 방법을 제공한다.

[0013] 본 실시예의 다른 측면에 의하면, 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 예측하는 메모리 프로파일링 장치에 있어서, 애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 추출부; 상기 접근 정보를 학습된 예측 모델에 입력하는 입력부; 상기 예측 모델로부터 상기 메모리 오브젝트의 스케일링 비율을 획득하는 획득부; 및 상기 애플리케이션의 워크로드가 변화한 경우, 상기 스케일링 비율을 상기 접근 정보에 적용함으로써, 상기 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 예측부를 포함하는 메모리 프로파일링 장치를 제공한다.

발명의 효과

[0014] 이상에서 설명한 바와 같이 본 발명의 일 실시예에 의하면, 머신러닝 기반 예측 모델을 이용하여 메모리 오브젝트의 스케일링 비율을 예측하고, 워크로드가 변화하더라도 스케일링 비율을 메모리 오브젝트 접근 정보에 적용함으로써, 워크로드 변화에 따른 접근 정보 예측에 드는 시간과 에너지 소모를 감소시킬 수 있다.

[0015] 본 발명의 다른 실시예에 의하면, 머신러닝 기반 예측 모델을 이용하여 워크로드 변화에 따라 변화된 메모리 오브젝트의 접근 정보를 정확하게 예측할 수 있다.

도면의 간단한 설명

[0016] 도 1은 본 발명의 일 실시예에 따른 애플리케이션과 메모리 프로파일링 장치를 설명하기 위해 예시한 도면이다.

도 2는 본 발명의 일 실시예에 따른 메모리 프로파일링 장치의 구성도이다.

도 3a는 본 발명의 일 실시예에 따른 메모리 프로파일링 과정을 설명하기 위해 예시한 도면이다.

도 3b는 본 발명의 일 실시예에 따른 예측 모델의 학습 및 예측 모델을 이용한 에너지 소모량 추정 과정을 설명하기 위해 예시한 도면이다.

도 4는 본 발명의 일 실시예에 따른 메모리 오브젝트의 접근 정보 추출 과정을 설명하기 위한 도면이다.

도 5a 및 도 5b는 본 발명과 종래기술의 메모리 프로파일링 결과를 비교한 도면이다.

도 6은 본 발명의 일 실시예에 따른 메모리 프로파일링 과정을 설명하기 위해 예시한 순서도다.

발명을 실시하기 위한 구체적인 내용

[0017] 이하, 본 발명의 일부 실시예들을 예시적인 도면을 통해 상세하게 설명한다. 각 도면의 구성요소들에 참조부호를 부가함에 있어서, 동일한 구성요소들에 대해서는 비록 다른 도면 상에 표시되더라도 가능한 한 동일한 부호

를 가지도록 하고 있음에 유의해야 한다. 또한, 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다.

[0018] 또한, 본 발명의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 '포함', '구비' 한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 또한, 명세서에 기재된 '~부', '모듈' 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어나 소프트웨어 또는 하드웨어 및 소프트웨어의 결합으로 구현될 수 있다.

[0019] 도 1은 본 발명의 일 실시예에 따른 애플리케이션과 메모리 프로파일링 장치를 설명하기 위해 예시한 도면이다.

[0020] 도 1을 참조하면, 애플리케이션(100), 메모리 시스템(110), 메모리 프로파일러(memory profiler, 112) 및 메모리 디바이스(114)가 도시되어 있다. 메모리 시스템(110)은 메모리 프로파일러(112) 및 메모리 디바이스(114)를 포함할 수 있다.

[0021] 본 발명의 일 실시예에 따른 메모리 프로파일링 장치는 메모리 시스템(110) 또는 메모리 프로파일러(112) 중 어느 하나일 수 있다.

[0022] 메모리 디바이스(114)는 휘발성 메모리(Volatile Memory) 또는 비휘발성 메모리(Non-Volatile Memory) 중 어느 하나이거나, 또는 두 종류의 메모리가 결합된 이기종 메모리 디바이스일 수 있다. 여기서, 메모리 디바이스(114)의 종류에 따라 메모리 오브젝트(memory object)의 에너지 소모량이 달라진다. 만일 메모리 디바이스(114)가 휘발성 메모리인 경우, 에너지 정보는 전처리 명령어에 의해 소비되는 에너지 소비량, 읽기 및 쓰기 명령어에 의해 소비되는 에너지 소비량 또는 리프레시 명령어에 의해 소비되는 에너지 소비량 중 적어도 하나를 포함한다.

[0023] 메모리 프로파일러(112)는 애플리케이션(100)으로부터 프로파일링 요청을 수신하면, 애플리케이션(100)에 의해 메모리 디바이스(114)에 할당되는 메모리 오브젝트(memory object)가 소비하는 에너지 소비량을 예측한다.

[0024] 이는 메모리 디바이스(114)의 종류 자체도 에너지 소비량에 영향을 주지만, 애플리케이션(100)이 사용하는 메모리 오브젝트가 메모리 디바이스(114)에 접근하는 패턴(access pattern) 또한 에너지 소비량에 영향을 끼치기 때문이다. 예를 들면, 메모리가 데이터를 기록하는 데 있어 많은 에너지를 소모하는 제1 메모리 디바이스와 적은 에너지를 소모하는 제2 메모리 디바이스가 존재하는 것으로 가정한다. 이때, 쓰기를 많이 수행하는 메모리 오브젝트가 제1 메모리 디바이스에 할당되면 제2 메모리 디바이스에 할당될 때보다 많은 에너지를 소비한다. 이처럼 어떤 메모리 오브젝트가 어떤 메모리 디바이스에서 이용되는지에 따라 총 에너지 소모량이 달라질 수 있다.

[0025] 따라서, 메모리 오브젝트의 에너지 소비량을 정확하게 예측하기 위해서는, 메모리 디바이스(114) 자체의 특성뿐만 아니라 메모리 오브젝트가 메모리 디바이스(114)에 접근하는 패턴 또한 고려해야 한다. 특히, 에너지 소모량 추정과 관련도가 높은 접근 정보들을 추출할 필요가 있다.

[0026] 본 발명의 일 실시예에 따른 메모리 프로파일러(112)는 애플리케이션(100)으로부터 프로파일링 요청을 수신하면, 애플리케이션(100)이 메모리 디바이스(114)에 할당하는 메모리 오브젝트가 메모리 디바이스(114)에 접근하는 접근 정보(access pattern)를 추출한다. 여기서, 에너지 추정에 이용하기 위해 추출되는 접근 정보는 메모리 오브젝트의 크기(size), 메모리 오브젝트의 메모리 접근 크기(accessed volume) 및 메모리 오브젝트의 수명(lifetime) 중 적어도 하나를 포함한다. 메모리 프로파일러(112)는 한 번 이상의 프로파일링을 수행함으로써, 메모리 오브젝트의 접근 정보를 추출할 수 있다.

[0027] 하지만, 메모리 프로파일러(112)가 애플리케이션(100)으로부터 프로파일링 요청을 수신할 때, 메모리 오브젝트의 접근 정보를 직접 추출하는 것은 많은 시간과 에너지를 소요한다. 또한, 애플리케이션(100)의 워크로드의 크기 등이 변화하는 경우 메모리 오브젝트의 접근 정보도 변화하므로, 워크로드의 크기 등이 변화할 때마다 메모리 오브젝트의 접근 정보를 추출하는 것은 시간과 에너지 측면에서 비효율적이다.

[0028] 이와 달리, 본 발명의 일 실시예에 따른 메모리 프로파일링 장치는 예측 모델을 이용하여 워크로드 변화에 따른 접근 정보의 변화 정도를 의미하는 스케일링 비율(scaling rate)을 예측하고, 워크로드가 변화하는 경우 메모리 오브젝트의 접근 정보에 스케일링 비율을 적용함으로써 변화된 워크로드에 대한 메모리 오브젝트의 접근 정보를 예측할 수 있다. 그리고 LSR 모델과의 차별화를 위해 새로운 예측 모델을 이용한다. 이를 통해, 메모리 오브젝

트의 접근 정보를 추출하는 데 드는 시간과 에너지 소모를 줄이되, 특히 예측 모델을 통해 접근 정보를 더 정확하게 예측할 수 있다.

[0029] 도 2는 본 발명의 일 실시예에 따른 메모리 프로파일링 장치의 구성도이다.

[0030] 도 2를 참조하면, 본 발명의 일 실시예에 따른 메모리 프로파일링 장치(20)는 추출부(200), 입력부(210), 획득부(220) 및 예측부(230)를 포함할 수 있다. 메모리 프로파일링 장치(20)는 애플리케이션의 워크로드 변화에 따른 메모리 오브젝트의 접근 정보를 예측하기 위한 장치이다.

[0031] 추출부(200)는 애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출하는 구성요소다.

[0032] 본 발명의 일 실시예에 의하면, 추출부(200)는 메모리 오브젝트가 메모리 디바이스에 할당되었을 때 소요되는 에너지 소모량과 관련된 접근 정보들을 추출한다. 여기서, 에너지 소모량 추정에 이용하기 위한 접근 정보는 메모리 오브젝트의 크기, 메모리 오브젝트의 메모리 접근 크기 및 메모리 오브젝트의 수명 중 적어도 하나를 포함한다.

[0033] 구체적인 추출 과정으로써, 추출부(200)는 메모리 오브젝트의 접근 정보를 두 단계에 걸쳐 추출할 수 있다. 첫 번째 단계는 추출부(200)가 메모리 오브젝트의 식별자를 오브젝트-레벨에서 추출하는 단계다. 두 번째 단계는 추출부(200)가 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 시간 중 적어도 하나를 명령어-레벨에서 추출하는 단계다. 더 자세한 내용은 도 4에서 설명한다.

[0034] 입력부(210)는 접근 정보를 학습된 예측 모델에 입력하는 구성요소다.

[0035] 획득부(220)는 학습된 예측 모델로부터 메모리 오브젝트의 스케일링 비율을 획득하는 구성요소다.

[0036] 예측 모델은 메모리 오브젝트의 접근 정보를 입력 받고, 메모리 오브젝트의 스케일링 비율(scaling rate)을 출력하도록 학습된 모델이다.

[0037] 스케일링 비율이란 워크로드 변화에 따른 접근 정보의 변화 정도를 의미한다. 구체적으로, 다양한 워크로드에 따라 접근 정보가 변화되는 비율을 메모리 오브젝트의 접근 정보의 스케일링 비율이라고 한다. 스케일링 비율은 메모리 오브젝트의 크기에 대한 스케일링 비율, 메모리 접근 크기에 대한 스케일링 비율 및 수명에 대한 스케일링 비율 중 적어도 하나를 포함할 수 있다. 한편, 메모리 오브젝트들은 접근 정보들에 대해 다른 스케일링 비율을 가질 수 있다.

[0038] 본 발명의 일 실시예에 의하면, 학습된 예측 모델은 학습 방법에 따라 선형 회귀(Linear Regression) 모델, K-최근접 이웃 회귀(K-Nearest Neighbor Regression) 모델 및 랜덤 포레스트 회귀(Random Forest Regression) 모델 중 어느 하나일 수 있다.

[0039] 예측부(230)는 애플리케이션의 워크로드가 변화한 경우, 스케일링 비율을 접근 정보에 적용함으로써, 메모리 오브젝트의 스케일링된 접근 정보를 예측하는 구성요소다.

[0040] 이로써, 메모리 프로파일링 장치(20)는 애플리케이션의 워크로드가 변화할 때, 변화된 워크로드에 따른 접근 정보를 다시 추출하는 것이 아니라 예측 모델을 통해 미리 구한 스케일링 비율을 이용함으로써 적은 시간 및 적은 에너지로도 메모리 오브젝트의 접근 정보를 예측할 수 있다.

[0041] 또한, 워크로드와 메모리 오브젝트의 접근 정보가 비선형관계를 포함한 다양한 관계에 있더라도 예측 모델을 이용함으로써, 워크로드 변화에 따른 메모리 오브젝트의 스케일링 비율 및 접근 정보를 정확하게 예측할 수 있다. 예측 모델의 종류에 대해서는 도 3a 및 도 3b에서 자세히 설명한다.

[0042] 도 3a는 본 발명의 일 실시예에 따른 메모리 프로파일링 과정을 설명하기 위해 예시한 도면이다.

[0043] 도 3a를 참조하면, 메모리 프로파일링 장치는 애플리케이션으로부터 메모리 오브젝트에 대한 프로파일링 요청을 수신하면, 2 단계 메모리 프로파일링 과정을 수행함으로써, 메모리 오브젝트의 접근 정보를 추출한다. 여기서, 접근 정보는 메모리 오브젝트의 크기, 메모리 오브젝트의 메모리 접근 크기 및 메모리 오브젝트의 수명 중 적어도 하나를 포함한다.

[0044] 도 3b는 본 발명의 일 실시예에 따른 예측 모델의 학습 및 예측 모델을 이용한 에너지 소모량 추정 과정을 설명하기 위해 예시한 도면이다.

[0045] 도 3b를 참조하면, 메모리 프로파일링 장치가 메모리 오브젝트의 접근 정보에 기초하여 예측 모델을 학습시키고, 학습된 예측 모델을 이용하여 접근 정보를 예측하며, 최종적으로 메모리 오브젝트의 에너지 소모량

을 추정하는 과정이 순서대로 도시되어 있다.

- [0046] 이하에서, 예측 모델의 구체적인 학습 과정을 설명한다.
- [0047] 메모리 프로파일링 장치는 학습용 메모리 오브젝트들에 대해 접근 정보들을 추출한다. 여기서, 접근 정보들은 에너지 추정과 관련된 정보들을 의미한다.
- [0048] 이후, 메모리 프로파일링 장치는 워크로드를 변화시키고, 변화된 워크로드에 따라 스케일링된 학습용 메모리 오브젝트들의 접근 정보들을 추출한다.
- [0049] 메모리 프로파일링 장치는 위의 정보들을 이용하여 메모리 오브젝트들의 스케일링 비율들을 구할 수 있다. 구체적으로, 메모리 프로파일링 장치는 워크로드 크기 변화에 따른 메모리 오브젝트의 접근 정보 변화를 의미하는 스케일링 비율을 메모리 오브젝트별로 연산하여 구한다.
- [0050] 이후, 메모리 프로파일링 장치는 학습용 메모리 오브젝트의 접근 정보 및 스케일링 비율을 예측 모델에 학습시킨다. 예측 모델이 메모리 오브젝트의 접근 정보와 스케일링 비율 간 관계를 학습하는 것이다. 예측 모델의 학습 데이터는 메모리 오브젝트의 접근 정보 및 스케일링 비율이 된다.
- [0051] 본 발명의 일 실시예에 의하면, 예측 모델은 메모리 오브젝트의 접근 정보를 입력 받아 스케일링 비율을 출력하도록 학습될 때, 메모리 오브젝트의 접근 정보들 사이의 상관 관계도 함께 학습한다. 여기서, 접근 정보 간 상관 관계는 하나의 접근 정보가 변화할 때 다른 접근 정보에 영향을 미치는지 여부를 의미한다. 예를 들어, 메모리 오브젝트의 접근 정보 중 메모리 접근 크기와 수명 간에는 양의 상관 관계가 있음을 의미한다. 다시 말하면, 메모리 오브젝트의 메모리 접근 크기가 클수록 수명도 길다는 것을 의미한다. 반면, 메모리 오브젝트의 크기는 다른 접근 정보와 상관 관계에 있지 않다. 예측 모델은 메모리 오브젝트의 접근 정보와 스케일링 비율 간 관계뿐만 아니라 메모리 오브젝트의 접근 정보들 간 상관 관계를 함께 학습할 수 있다.
- [0052] 이하에서는, 예측 모델의 종류를 설명한다. 본 발명의 일 실시예에 따른 예측 모델은 선형 회귀 모델, K-최근접 이웃 회귀 모델 및 랜덤 포레스트 회귀 모델 중 어느 하나일 수 있다.
- [0053] 선형 회귀 모델은 애플리케이션의 워크로드와 메모리 오브젝트의 접근 정보가 선형 관계에 있는 것을 전제로, 워크로드 변화에 따라 접근 정보도 선형적으로 변화하는 것으로 추정하는 모델이다. 워크로드와 학습용 메모리 오브젝트들의 접근 정보들 간 스케일링 비율들을 선형 회귀 방식으로 학습한다.
- [0054] 본 발명의 일 실시예에 의하면, 선형 회귀 모델은 메모리 오브젝트별로 선형 회귀를 적용하거나 메모리 오브젝트의 접근 정보별로 선형 회귀를 적용할 수 있다.
- [0055] 선형 회귀 모델은 데이터 세트들 사이에 존재하는 선형 접근 정보들을 이용하여 스케일링 비율을 예측하므로, 접근 정보들이 선형 관계에 있을 때, 스케일링 비율의 예측 정확도가 높다. 또한, 선형 회귀 모델에서는 스케일링 비율이 오직 선형 접근 정보들에 기초하여 예측되므로, 접근 정보들 사이의 관계가 복잡하지 않고 가벼운 예측 모델이라는 특성을 가진다. 다만, 워크로드의 메모리 오브젝트의 접근 정보들 사이에 상호관계가 존재한다면, 선형 회귀 모델은 스케일링 비율을 정확하게 예측하지 못한다. 따라서, 선형 회귀 모델은 다양한 메모리 오브젝트의 접근 정보들을 예측하는 데 적합하지는 않다.
- [0056] K-최근접 이웃 회귀 모델은 존재하는 데이터들 중 가장 최근의 K 번째 데이터들의 평균으로부터 새로운 데이터를 예측하는 모델이다. k 번째까지의 학습용 메모리 오브젝트들의 스케일링 비율의 평균을 학습하는 것이다. K-최근접 이웃 회귀 모델은 워크로드 변화에 따라 각 메모리 오브젝트들의 스케일링 비율이 제각각인 경우, k 번째까지의 메모리 오브젝트들의 스케일링 비율을 평균한 값을 출력하도록 학습된 모델이다.
- [0057] K-NNR 모델은 최근접 이웃 데이터의 개수에 따라 예측 정확도가 달라진다. k 값이 작을 경우, 데이터의 지역적 특성에 오버피팅될 수 있다. 반대로, k 값이 클 경우, K-NNR 모델은 언더피팅에 의해 과도하게 일반화될 수 있다. 예측 정확도 관점에서, 이웃한 메모리 오브젝트들 사이에 상호관계가 존재할 때, K-NNR 모델에 따른 예측 정확도는 높아진다. K-NNR 모델은 애플리케이션 내 메모리 오브젝트들 사이의 상호관계를 증명하고, 스케일링 비율을 예측할 수 있다. 또한, K-NNR 모델은 선형 회귀 모델과 유사하게, 복잡한 관계에 있는 메모리 오브젝트들로부터 스케일링 비율을 예측하는 것이 아니라 k 번째 이웃들을 이용하여 스케일링 비율을 예측하기 때문에, K-NNR 모델은 가벼운 모델일 수 있다.
- [0058] 랜덤 포레스트 회귀 모델은 각 데이터들이 가진 속성들로부터 패턴을 찾아내서 결과를 출력할 수 있도록 하는 지도학습 머신러닝 모델이다. 학습 데이터 중 일부를 임의로 추출하여 트리를 만든다. 일부를 추출하는 것을 배

깅(bagging)이라 하고, 데이터들의 특징(feature)들을 제한하여 구성된 트리를 이용하여 결과를 예측한다. 즉, 학습용 메모리 오브젝트들을 이용하여 트리를 구성하고, 트리를 이용하여 할당된 메모리 오브젝트의 스케일링 비율을 예측한다.

- [0059] 본 발명의 일 실시예에 의하면, 예측 모델이 랜덤 포레스트 회귀(RFR) 모델인 경우, 예측 모델의 정확성을 위해 RFR을 구성하는 트리(tree) 수도 고려되어야 한다. 예측 모델의 정확도는 트리의 수에 따라 다르기 때문에 각 애플리케이션은 트리의 수에 따라 정확도가 포화되는 지점을 찾아 학습 과정에 반영한다. 이를 위해, 학습된 예측 모델의 트리의 수는 학습용 메모리 오브젝트들의 스케일링 비율들에 기초하여 결정될 수 있다. 구체적으로, 학습된 예측 모델의 트리의 수는 학습용 메모리 오브젝트들의 스케일링 비율들이 불규칙적일수록 높게 결정되는 것일 수 있다. 이는, 워크로드와 메모리 오브젝트의 접근 정보 간 관계가 복잡할수록 트리의 수를 증가시킴으로써, 예측 모델의 학습량을 늘려 스케일링 비율 및 접근 정보의 예측 정확도를 높이기 위한 것이다.
- [0060] 학습이 완료된 예측 모델은 메모리 오브젝트의 접근 정보를 입력 받으면 메모리 오브젝트의 스케일링 비율을 출력한다.
- [0061] 도 4는 본 발명의 일 실시예에 따른 메모리 오브젝트의 접근 정보 추출 과정을 설명하기 위한 도면이다.
- [0062] 도 4를 참조하면, 메모리 프로파일링 장치는 빠른 단계와 느린 단계를 포함하는 두 단계를 거쳐 학습용 메모리 오브젝트의 접근 정보를 추출한다. 이를 2 단계 메모리 프로파일링(Two Pass Memory Profiling; TPMP)이라 한다.
- [0063] 첫 번째 단계로써, 메모리 프로파일링 장치는 학습용 메모리 오브젝트에 대해 1차 프로파일링을 수행한다. 1차 프로파일링은 첫 번째 단계로서, 빠른 단계에 해당한다. 빠른 단계에서, 메모리 프로파일링 장치는 애플리케이션의 서로 다른 변수들을 구분하고, 각 변수의 오브젝트들을 해당 변수와 맵핑(mapping)하기 위해 함수 호출 스택의 해시 값(hash value)을 추출한다. 여기서, 해시 값은 메모리 오브젝트의 식별자(identifier)에 해당한다.
- [0064] 구체적으로, 메모리 프로파일링 장치는 래퍼 라이브러리(wrapper library)를 이용하여 malloc, calloc, realloc 함수와 같은 동적 할당 함수를 가로챌 다음, 동적 할당 함수의 호출 스택을 문자열로 저장하여 이를 해싱(hashing)한다. 메모리 프로파일링 장치는 모든 메모리 오브젝트들의 해시 값을 추출하여 결과 파일에 쓰면, 이후 오프라인 작업(offline processing)이 수행되면서 2차 프로파일링을 수행할 주요 변수들을 선정하여 해시 값을 파일에 나열한다.
- [0065] 이후, 메모리 프로파일링 장치는 학습용 메모리 오브젝트에 대해 2차 프로파일링을 수행한다. 2차 프로파일링은 두 번째 단계로서, 느린 단계에 해당한다. 느린 단계에서, 메모리 프로파일링 장치는 주요 변수들을 핀 툴(PIN tool)을 이용하여 인스트럭션 단위에서 접근 정보에 대해 프로파일링을 수행한다. 이때, 핀 툴 라이브러리를 통해 메모리 프로파일링 장치는 에너지 추정에 필요한 접근 정보들만을 추출하고, 에너지 추정에 관련이 없는 접근 정보들을 생략할 수 있다.
- [0066] 예를 들면, 메모리 프로파일링 장치는 Intel Pin tool을 이용하여 런타임에 모든 학습용 메모리 오브젝트들의 접근에 대해 프로파일링을 수행할 수 있으며, 이 경우 학습용 메모리 오브젝트들에 대한 프로파일링 결과를 첫 번째 단계에서 생성한 해시 테이블을 참조하여 해당 오브젝트가 속한 학습용 메모리 오브젝트의 결과로 저장할 수 있다.
- [0067] 두 번째 단계로써, 메모리 프로파일링 장치는 학습용 메모리 오브젝트에 대해 2차 프로파일링을 수행하여 학습용 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 중 적어도 하나를 추출한다.
- [0068] 메모리 프로파일링 장치는 추출된 접근 정보를 이용하여 학습용 메모리 오브젝트가 메모리 디바이스에 할당되었을 때 예측되는 에너지 소비량을 결정한다. 구체적으로, 메모리 프로파일링 장치는 학습용 메모리 오브젝트로 인해 메모리 디바이스 상에서 수행되는 명령어에 따라 소비되는 에너지 정보를 결정하고, 접근 정보 및 에너지 정보를 이용하여 에너지 소비량을 결정한다.
- [0069] 한편, 본 명세서에서는 메모리 프로파일링 장치가 학습용 메모리 오브젝트의 접근 정보를 추출하기 위해 첫 번째 단계 및 두 번째 단계를 통해 프로파일링하는 Two-pass 메모리 프로파일링을 사용하는 것으로 설명하였으나, 학습용 메모리 오브젝트의 접근 정보를 추출하기 위해 다른 메모리 프로파일링 기법을 사용할 수도 있다.
- [0070] 도 5a 및 도 5b는 본 발명과 종래기술의 메모리 프로파일링 결과를 비교한 도면이다.
- [0071] 도 5a를 참조하면, 종래 기술인 선형 스케일링 비율 모델과 본 발명인 메모리 프로파일링 장치 간 에너지 추정

정확도를 애플리케이션 별로 비교한 그래프가 도시되어 있다. 그래프는 여러 애플리케이션에서 메모리 프로파일링 장치에 따른 에너지 소모량 추정의 예측 정확도가 종래 기술보다 높다는 것을 보여준다.

- [0072] 도 5b를 참조하면, 종래 기술인 선형 스케일링 비율 모델과 본 발명이 메모리 프로파일링 장치 간 메모리 오브젝트의 접근 정보 예측 정확도를 애플리케이션 별로 비교한 그래프가 도시되어 있다. 메모리 프로파일링 장치는 예측 모델로부터 구한 스케일링 비율을 통해 접근 정보를 예측한 것이고, 종래기술은 선형 스케일링 비율을 통해 접근 정보를 예측한 것이다.
- [0073] 예측 모델이 RFR 모델인 경우, 메모리 프로파일링 장치의 접근 정보 예측 정확도는 평균적으로 92.85 퍼센트이며, 에너지 소모량 추정 정확도는 91.3 퍼센트이다.
- [0074] 도 6은 본 발명의 일 실시예에 따른 메모리 프로파일링 과정을 설명하기 위해 예시한 순서도다.
- [0075] 도 6을 참조하면, 메모리 프로파일링 장치는 애플리케이션에 의해 할당된 메모리 오브젝트의 접근 정보를 추출한다(S600). 여기서, 접근 정보는 메모리 오브젝트의 크기(size), 메모리 접근 크기(accessed volume) 및 수명(lifetime) 중 적어도 하나를 포함한다.
- [0076] 구체적인 추출 과정은, 메모리 오브젝트의 식별자를 오브젝트-레벨에서 추출하는 과정 및 메모리 오브젝트의 크기, 메모리 접근 크기 및 수명 시간 중 적어도 하나를 명령어-레벨에서 추출하는 과정을 포함하여 두 단계로 수행된다.
- [0077] 메모리 프로파일링 장치는 접근 정보를 학습된 예측 모델에 입력한다(S602). 여기서, 예측 모델은 학습용 메모리 오브젝트들의 접근 정보들을 입력 받고 출력 스케일링 비율을 출력하도록 학습된 것이다. 예측 모델은 학습 종류에 따라 선형 회귀(Linear Regression) 모델, K-최근접 이웃 회귀(K-Nearest Neighbor Regression) 모델 및 랜덤 포레스트 회귀(Random Forest Regression) 모델 중 어느 하나일 수 있다. 만약 학습된 예측 모델이 랜덤 포레스트 회귀 모델인 경우, 학습된 예측 모델의 트리(tree)의 수는 학습용 메모리 오브젝트들의 스케일링 비율들에 기초하여 결정될 수 있다. 구체적으로는, 학습용 메모리 오브젝트들의 스케일링 비율들이 불규칙적일수록 학습된 예측 모델의 트리의 수가 높게 결정될 수 있다.
- [0078] 메모리 프로파일링 장치는 예측 모델로부터 메모리 오브젝트의 스케일링 비율을 획득한다(S604).
- [0079] 메모리 프로파일링 장치는 애플리케이션의 워크로드가 변화한 경우, 스케일링 비율을 접근 정보에 적용함으로써, 메모리 오브젝트의 스케일링된 접근 정보를 예측한다(S606).
- [0080] 추후, 메모리 프로파일링 장치는 스케일링 비율을 이용하여 획득한 스케일링된 접근 정보로부터 메모리 오브젝트의 에너지 소모량을 추정할 수 있다.
- [0081] 도 6에서는 과정 S600 내지 과정 S606을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 본 발명의 일 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것이다. 다시 말해, 본 발명의 일 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 일 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 6에 기재된 순서를 변경하여 실행하거나 과정 S600 내지 과정 S606 중 하나 이상의 과정을 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 6은 시계열적인 순서로 한정되는 것은 아니다.
- [0082] 한편, 도 6에 도시된 과정들은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 즉, 이러한 컴퓨터가 읽을 수 있는 기록매체는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등의 비일시적인(non-transitory) 매체를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [0083] 또한, 본 발명의 구성 요소들은 메모리, 프로세서, 논리 회로, 룩-업 테이블(look-up table) 등과 같은 직접 회로 구조를 사용할 수 있다. 이러한 직접 회로 구조는 하나 이상의 마이크로 프로세서 또는 다른 제어 장치의 제어어를 통해 본 명세서에 기술된 각각의 기능을 실행한다. 또한, 본 발명의 구성 요소들은 특정 논리 기능을 수행하기 위한 하나 이상의 실행 가능한 명령을 포함하고 하나 이상의 마이크로 프로세서 또는 다른 제어 장치에 의해 실행되는 프로그램 또는 코드의 일부에 의해 구체적으로 구현될 수 있다. 또한, 본 발명의 구성 요소들은 각각의 기능을 수행하는 중앙 처리 장치(CPU), 마이크로 프로세서 등을 포함하거나 이에 의해 구현될 수 있다. 또한, 본 발명의 구성 요소들은 하나 이상의 프로세서에 의해 실행되는 명령어들을 하나 이상의 메모리에 저장

할 수 있다.

[0084]

이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0085]

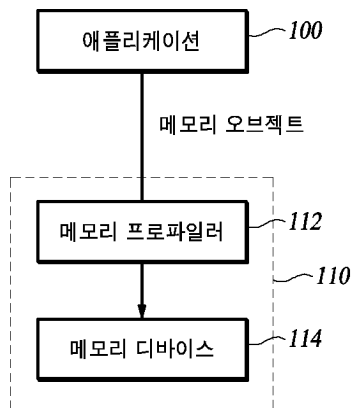
20: 메모리 프로파일링 장치 200: 추출부

210: 입력부 220: 획득부

230: 예측부

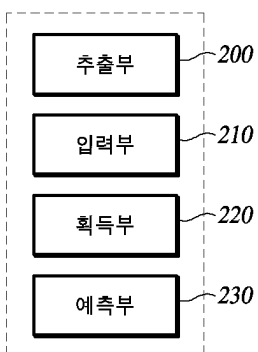
도면

도면1

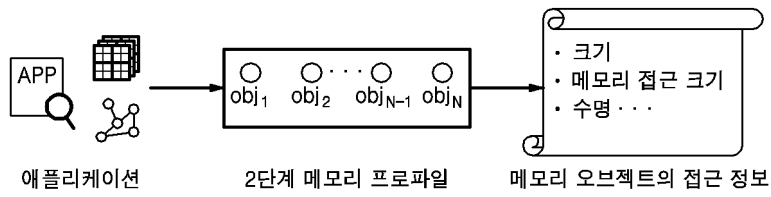


도면2

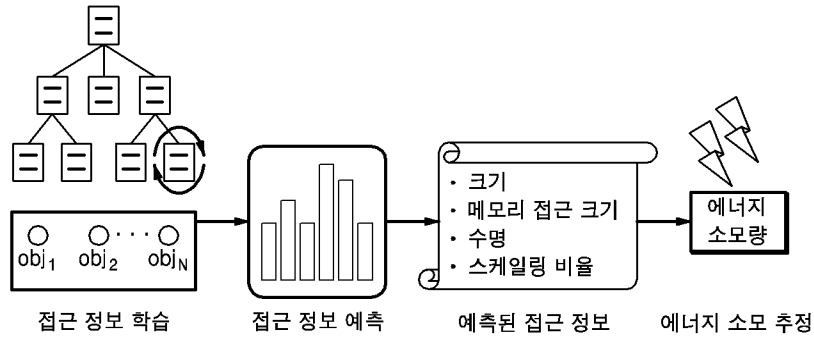
20



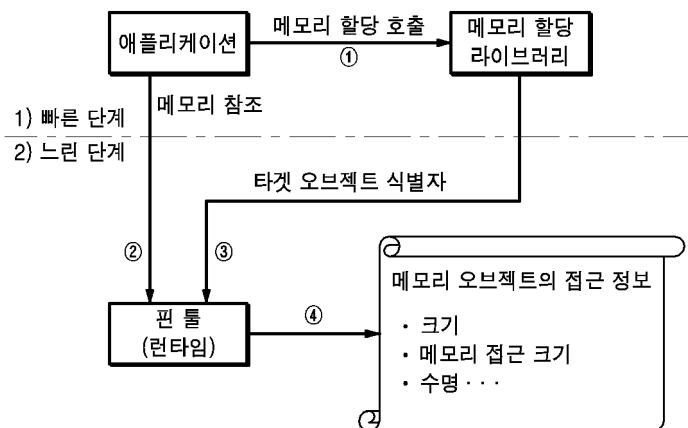
도면3a



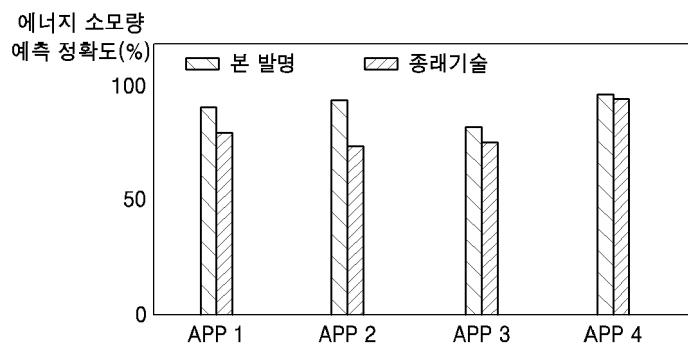
도면3b



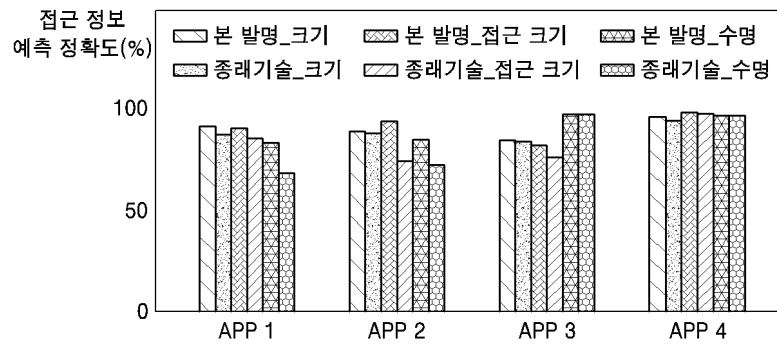
도면4



도면5a



도면5b



도면6

