



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0045241
(43) 공개일자 2025년04월01일

- | | |
|--|--|
| <p>(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) H04L 41/14 (2022.01)
H04L 41/16 (2022.01)</p> <p>(52) CPC특허분류
G06N 20/00 (2021.08)
H04L 41/14 (2022.05)</p> <p>(21) 출원번호 10-2023-0128277</p> <p>(22) 출원일자 2023년09월25일
심사청구일자 2023년09월25일</p> | <p>(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(72) 발명자
김지수
울산광역시 울주군 언양읍 유니스트길 50
강민현
울산광역시 울주군 언양읍 유니스트길 50</p> <p>(74) 대리인
특허법인 무한</p> |
|--|--|

전체 청구항 수 : 총 18 항

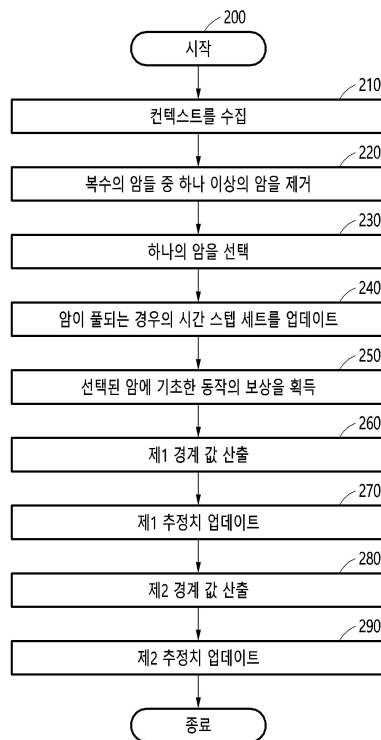
(54) 발명의 명칭 **밴딧 모델의 동작 및 트레이닝 방법 및 장치**

(57) 요약

일 실시예에 따른 네트워크 성능 분석 방법 및 장치는 목표 주기까지 컨텍스트(context)를 수집하는 단계, 상기 목표 주기가 강제 샘플링(forced sampling) 시간 스텝 세트에 포함되지 않는 경우, 상기 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암(arm)들 중 하나 이상의 암을 제거(discard)하는 단계, 상기 복수의

(뒷면에 계속)

대표도 - 도2



암들 중 상기 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 상기 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택하는 단계, 상기 목표 주기에 기초하여 암이 풀(pull)되는 경우의 시간 스텝 세트를 업데이트하는 단계, 상기 선택된 암에 기초한 보상(reward)을 획득하는 단계, 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1 경계(bound) 값을 산출하는 단계, 상기 산출된 제1 경계 값과 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트하는 단계, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하는 단계 및 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 단계를 포함할 수 있다.

(52) CPC특허분류

H04L 41/16 (2022.05)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193264
과제번호	2022-0-00469-002
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	사람중심인공지능핵심원천기술개발
연구과제명	자율 드론 실용화를 위한 목적지향 강화학습 핵심기술 개발
기 여 율	50/100
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193983
과제번호	2020-0-01336-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능대학원지원(울산과학기술원)
기 여 율	50/100
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

프로세서에 의해 수행되는 다중 암 밴딧 모델의 동작 및 트레이닝시키는 방법에 있어서,

목표 주기까지 컨텍스트(context)를 수집하는 단계;

상기 목표 주기가 강제 샘플링(forced sampling) 시간 스텝 세트에 포함되지 않는 경우, 상기 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암(arm)들 중 하나 이상의 암을 제거(discard)하는 단계;

상기 복수의 암들 중 상기 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 상기 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택하는 단계;

상기 목표 주기에 기초하여 암이 풀(pull)되는 경우의 시간 스텝 세트를 업데이트하는 단계;

상기 선택된 암에 기초한 보상(reward)을 획득하는 단계;

상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1 경계(bound) 값을 산출하는 단계;

상기 산출된 제1 경계 값과 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트하는 단계;

상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하는 단계; 및

상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 단계

를 포함하는 밴딧 모델의 동작 및 트레이닝 방법.

청구항 2

제1항에 있어서,

상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암을 풀하는 단계

를 더 포함하는 밴딧 모델의 동작 및 트레이닝 방법.

청구항 3

제1항에 있어서,

상기 목표 주기의 제1 추정치는,

상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트되는,

밴딧 모델의 동작 및 트레이닝 방법.

청구항 4

제1항에 있어서,

상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 목표 주기의 암이 풀되는 경우

의 시간 스텝 세트에서의 압에 기초하여 제2 경계 값을 산출하는 단계; 및
 상기 산출된 제2 경계 값과 상기 압이 풀되는 경우의 시간 스텝 세트에서의 압에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 단계를 더 포함하는 밴딤 모델의 동작 및 트레이닝 방법.

청구항 5

제4항에 있어서,
 상기 목표 주기의 제2 추정치는,
 상기 압이 풀되는 경우의 시간 스텝 세트에서의 압에 기초하여 가장 작은 손실함수를 가지도록 업데이트되는,
 밴딤 모델의 동작 및 트레이닝 방법.

청구항 6

제3항 또는 제5항에 있어서,
 상기 손실함수는,
 상기 강제 샘플링 시간 스텝 세트 또는 상기 압이 풀되는 경우의 시간 스텝 세트에서의 압에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값 이하이면 제곱 손실(square loss) 함수이고,
 상기 강제 샘플링 시간 스텝 세트 또는 상기 압이 풀되는 경우의 시간 스텝 세트에서의 압에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값보다 크면 절대 손실(absolute loss) 함수인,
 밴딤 모델의 동작 및 트레이닝 방법.

청구항 7

제1항에 있어서,
 상기 이전 목표 주기의 제1 추정치는 상기 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트에서의 압에 기초하여 추정되고,
 상기 이전 목표 주기의 제2 추정치는 상기 이전 목표 주기까지의 압이 풀되는 경우의 시간 스텝 세트에서의 압에 기초하여 추정되는,
 밴딤 모델의 동작 및 트레이닝 방법.

청구항 8

제1항에 있어서,
 상기 하나 이상의 압을 제거하는 단계는,
 이전 목표 주기에서 가장 큰 보상 값 추정치와 상기 이전 목표 주기의 제1 추정치에 기초한 보상 값 간의 차이가 미리 정해진 범위 값을 초과하는 압을 제거하는 단계를 포함하는 밴딤 모델의 동작 및 트레이닝 방법.

청구항 9

제1항에 있어서,
상기 선택하는 단계는,
상기 이전 목표 주기의 제2 추정치에 기초하여 산출된 보상 값 중 가장 큰 보상 값을 가진 암을 선택하는 단계를 포함하는 밴딤 모델의 동작 및 트레이닝 방법.

청구항 10

목표 주기까지 컨텍스트를 수집하고, 상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암들 중 하나 이상의 암을 제거하고, 상기 복수의 암들 중 상기 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 상기 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택하고, 상기 목표 주기에 기초하여 암이 풀되는 경우의 시간 스텝 세트를 업데이트하고, 상기 선택된 암에 기초한 보상을 획득하고, 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1 경계(bound) 값을 산출하고, 상기 산출된 제1 경계 값과 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트하고, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하며, 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 프로세서를 포함하는 밴딤 모델의 동작 및 트레이닝 장치.

청구항 11

제10항에 있어서,
상기 프로세서는,
상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암을 풀하는,
밴딤 모델의 동작 및 트레이닝 장치.

청구항 12

제10항에 있어서,
상기 목표 주기의 제1 추정치는,
상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트되는,
밴딤 모델의 동작 및 트레이닝 장치.

청구항 13

제10항에 있어서,
상기 프로세서는,
상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하고, 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는,

밴딧 모델의 동작 및 트레이닝 장치.

청구항 14

제13항에 있어서,

상기 목표 주기의 제2 추정치는,

상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트되는,

밴딧 모델의 동작 및 트레이닝 장치.

청구항 15

제12항 또는 제14항에 있어서,

상기 손실함수는,

상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값 이하이면 제공 손실 함수이고,

상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값보다 크면 절대 손실 함수인,

밴딧 모델의 동작 및 트레이닝 장치.

청구항 16

제10항에 있어서,

상기 이전 목표 주기의 제1 추정치는 상기 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 추정되고,

상기 이전 목표 주기의 제2 추정치는 상기 이전 목표 주기까지의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 추정되는,

밴딧 모델의 동작 및 트레이닝 장치.

청구항 17

제10항에 있어서,

상기 프로세서는,

이전 목표 주기에서 가장 큰 보상 값과 상기 이전 목표 주기의 제1 추정치에 기초한 보상 값 간의 차가 미리 정해진 범위 값을 초과하는 암을 제거하는,

밴딧 모델의 동작 및 트레이닝 장치.

청구항 18

제10항에 있어서,

상기 프로세서는,

상기 이전 목표 주기의 제2 추정치에 기초하여 산출된 보상 값 중 가장 큰 보상 값을 가진 암을 선택하는,

밴딧 모델의 동작 및 트레이닝 장치.

발명의 설명

기술 분야

[0001] 이하, 두꺼운 꼬리 분포를 가지는 데이터에 적용가능한 밴딧 알고리즘에 관한 기술이 제공된다.

배경 기술

[0003] 머신 러닝(ML, Machine Learning)은 컴퓨터가 스스로 방대한 데이터를 분석하여 앞으로의 행동을 예측하는 기술이다. 머신 러닝(ML)은 크게 지도학습(Supervised Learning), 비지도학습(Unsupervised Learning) 및 강화 학습(RL, Reinforcement Learning)을 포함할 수 있다. 강화 학습(RL)은 현재의 상태(state)에서 어떠한 행동(action)(예: 암(arm))을 선택하는 것이 최적인지를 학습하는 방법이다. 학습 방법에 있어서, 강화 학습(RL)은 에이전트(agent)가 행동할 때마다 보상(reward)이 주어지고 이러한 보상을 최대화하는 방향으로 학습을 진행할 수 있다. 강화 학습(RL)은 자율 주행 차량, 산업 자동화, 경제 분석, 질병 치료 및 선호도에 기반한 추천 콘텐츠 제공과 같이 다양한 분야에서 적용되고 있다.

[0004] 위에서 설명한 배경기술은 발명자가 본원의 개시 내용을 도출하는 과정에서 보유하거나 습득한 것으로서, 반드시 본 출원 전에 일반 공중에 공개된 공지기술이라고 할 수는 없다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국 등록 특허공보 10-1935654(2018년 12월 28일 공고)에는 사용자 그룹에 대한 아이템 추천 장치 및 방법이 제시된다.

(특허문헌 0002) 미국 등록 특허공보 11443346(2022년09월13일 공고)에는 Group item recommendations for ephemeral groups based on mutual information maximization이 제시된다.

발명의 내용

해결하려는 과제

[0007] 일 실시예에 따른 밴딧 모델의 동작 및 트레이닝 방법 및 장치는 손실함수에 기초하여 업데이트되는 제1 추정치를 이용하여 최적의 행동이 아닌 행동들을 필터링하고 남은 행동들 중 최적의 행동을 손실함수에 기초하여 업데이트되는 제2 추정치를 이용하여 선택할 수 있다.

[0008] 다만, 기술적 과제는 상술한 기술적 과제들로 한정되는 것은 아니며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0010] 일 실시예에 따른 밴딧 모델의 동작 및 트레이닝 방법은 목표 주기까지 컨텍스트(context)를 수집하는 단계, 상기 목표 주기가 강제 샘플링(forced sampling) 시간 스텝 세트에 포함되지 않는 경우, 상기 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암(arm)들 중 하나 이상의 암을 제거(discard)하는 단계, 상기 복수의 암들 중 상기 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 상기 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택하는 단계, 상기 목표 주기에 기초하여 암이 풀(pull)되는 경우의 시간 스텝 세트를 업데이트하는 단계, 상기 선택된 암에 기초한 보상(reward)을 획득하는 단계, 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1

경계(bound) 값을 산출하는 단계, 상기 산출된 제1 경계 값과 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트하는 단계, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하는 단계 및 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 단계를 포함할 수 있다.

[0011] 상기 방법은 상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암을 풀하는 단계를 더 포함할 수 있다.

[0012] 상기 목표 주기의 제1 추정치는, 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트될 수 있다.

[0013] 상기 방법은 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하는 단계 및 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 단계를 더 포함할 수 있다.

[0014] 상기 목표 주기의 제2 추정치는, 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트될 수 있다.

[0015] 상기 손실함수는, 상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값 이하이면 제곱 손실(square loss) 함수이고, 상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값보다 크면 절대 손실(absolute loss) 함수일 수 있다.

[0016] 상기 이전 목표 주기의 제1 추정치는 상기 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 추정되고, 상기 이전 목표 주기의 제2 추정치는 상기 이전 목표 주기까지의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 추정될 수 있다.

[0017] 상기 하나 이상의 암을 제거하는 단계는, 이전 목표 주기에서 가장 큰 보상 값 추정치와 상기 이전 목표 주기의 제1 추정치에 기초한 보상 값 간의 차가 미리 정해진 범위 값을 초과하는 암을 제거하는 단계를 포함할 수 있다.

[0018] 상기 선택하는 단계는, 상기 이전 목표 주기의 제2 추정치에 기초하여 산출된 보상 값 중 가장 큰 보상 값을 가진 암을 선택하는 단계를 포함할 수 있다.

[0019] 일 실시예에 따른 밴드 모델의 동작 및 트레이닝 장치는 목표 주기까지 컨텍스트를 수집하고, 상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암들 중 하나 이상의 암을 제거하고, 상기 복수의 암들 중 상기 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 상기 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택하고, 상기 목표 주기에 기초하여 암이 풀되는 경우의 시간 스텝 세트를 업데이트하고, 상기 선택된 암에 기초한 보상을 획득하고, 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1 경계 값을 산출하고, 상기 산출된 제1 경계 값과 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트하고, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하며, 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트하는 프로세서를 포함할 수 있다.

[0020] 상기 프로세서는, 상기 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되는 경우, 상기 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암을 풀할 수 있다.

[0021] 상기 목표 주기의 제1 추정치는, 상기 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실함수를 가지도록 업데이트될 수 있다.

[0022] 상기 프로세서는, 상기 목표 주기가 상기 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 상기 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출하고, 상기 산출된 제2 경계 값과 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 상기 이전 목표 주기의 제2 추정치를 목표 주기

의 제2 추정치로 업데이트할 수 있다.

- [0023] 상기 목표 주기의 제2 추정치는, 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 가장 작은 손실 함수를 가지도록 업데이트될 수 있다.
- [0024] 상기 손실함수는, 상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값 이하이면 제곱 손실 함수이고, 상기 강제 샘플링 시간 스텝 세트 또는 상기 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값보다 크면 절대 손실 함수일 수 있다.
- [0025] 상기 이전 목표 주기의 제1 추정치는 상기 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 추정되고, 상기 이전 목표 주기의 제2 추정치는 상기 이전 목표 주기까지의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 추정될 수 있다.
- [0026] 상기 프로세서는, 이전 목표 주기에서 가장 큰 보상 값 추정치와 상기 이전 목표 주기의 제1 추정치에 기초한 보상 값 간의 차가 미리 정해진 범위 값을 초과하는 암을 제거할 수 있다.
- [0027] 상기 프로세서는, 상기 이전 목표 주기의 제2 추정치에 기초하여 산출된 보상 값 중 가장 큰 보상 값을 가진 암을 선택할 수 있다.

발명의 효과

- [0029] 일 실시예에 따른 밴딤 모델의 동작 및 트레이닝 방법 및 장치는 손실함수에 기초하여 업데이트되는 제1 추정치를 이용하여 최적의 행동이 아닌 행동들을 필터링하고 남은 행동들 중 최적의 행동을 손실함수에 기초하여 업데이트되는 제2 추정치를 이용하여 선택함에 따라, 두꺼운 꼬리(heavy-tailed) 분포를 가진 이상치나 극단값이 많은 보상분포 혹은 변동성이 큰 보상분포에 적용될 수 있다.

도면의 간단한 설명

- [0031] 도 1은 일 실시예에 따른 밴딤 모델의 동작 및 트레이닝 장치의 블록도를 도시한다.
- 도 2는 일 실시예에 따른 밴딤 모델의 동작 및 트레이닝 방법의 흐름도를 도시한다.
- 도 3 내지 도 5는 시간에 따라 일 실시예에 따른 밴딤 모델의 동작 및 트레이닝 방법과 다른 기술들을 이용하여 산출된 리그렛(regret)의 누적합(R(T))을 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0032] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0033] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0034] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.
- [0035] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0036] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로

해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0037] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.

[0038] 강화 학습(RL, Reinforcement Learning)의 일종인 밴딤(bandit) 알고리즘은 순차적 의사 결정(sequential decision) 과정에서 발생하는 보상이나 손실을 최적화하는 방법이다. 밴딤 알고리즘 중 하나인 멀티-암드 밴딤(MAB, Multi-armed Bandit) 알고리즘은 각 시점마다 학습자(예: 알고리즘)에게 다수의 후보 암(arm) (예: 행동(action))이 주어지며, 학습자는 이 가운데 하나의 암을 선택(예: 풀(pull))하고 그에 따른 보상(reward)(또는 손실)을 받을 수 있다. 그런 후, 멀티-암드 밴딤 알고리즘은 보상값이 새로 축적될 때마다 보상 모형을 학습하며 이를 다음 시점의 행동 선택에 반영할 수 있다. 이는 멀티-암드 밴딤 알고리즘이 순차적으로 축적되는 데이터로부터 정보를 학습하는 온라인 학습 방법과 보상이 큰 행동을 선택하는 의사 결정 전략을 접목한 방식임을 의미할 수 있다. 이러한 멀티-암드 밴딤 알고리즘은 진술한 특성들에 의해 임상 시험, 웹 페이지 뉴스 기사 추천 시스템, 광고 배치 시스템, 모바일 헬스 어플리케이션 등의 분야에서 활용되고 있다.

[0039] 컨텍슈얼 밴딤 알고리즘(contextual bandit algorithm)은 멀티-암드 밴딤 알고리즘에서 컨텍스트(context)가 추가된 알고리즘으로서, 선택 가능한 각 행동에 대한 컨텍스트가 유한 차원의 공변량 형태로 주어질 때 이를 선택에 이용할 수 있다. 예를 들어, 사용자의 운동량(예: 보상)을 증가시킬 목적으로 컨텍슈얼 밴딤 알고리즘이 적용된 모바일 헬스케어 시스템이 사용자에게 운동을 권장하는 다양한 메시지(예: 행동) 가운데 하나를 선택하여 푸시(push) 하고자 하는 경우, 모바일 헬스케어 시스템은 공변량으로 주어진 각 메시지에 대한 정보 및 사용자의 나이, 성별, 건강상태 등의 정보(예: 컨텍스트)에 기초하여 메시지를 선택함에 따라 사용자에게 푸시할 수 있다.

[0040] 이러한 컨텍슈얼 밴딤 알고리즘은 수학적식1을 이용하여 보상($r_i(t)$)을 획득할 수 있다. $r_i(t)$ 은 임의의 시간(t)에서 임의의 암(i)(예: 행동)을 선택했을 때 받을 수 있는 보상을 의미할 수 있다. 이는 학습자가 임의의 시간(t)에서 d차원의 공변량(X_t)을 관찰하고, K개의 선택 가능한 행동 중에 하나를 선택하는 것을 의미할 수 있다. 여기서, 모수(β_i)(예: 추정치)는 d차원으로 임의의 암(i)과 관련되었으나 학습자가 알 수 없는 값으로 학습을 통해 추정해야 하는 값이다. 또한, 학습자는 임의의 시간(t)에서 선택한 임의의 암(i)에 대한 보상($r_i(t)$)만을 관측하고 나머지 K-1개 행동에 대한 보상 값을 전혀 알지 못할 수 있다.

수학적식 1

$$r_i(t) = X_t^T \beta_i + \epsilon_{i,t}$$

[0041]

[0042] 진술한 수학적식 1에서 $\epsilon_{i,t}$ 는 서브-가우시안(sub-gaussian) 분포를 따르는 랜덤 에러(error)를 나타낼 수 있다.

[0043] 또한, 보상($r_i(t)$)은 수학적식 1에서와 같이 한 시점(t)에서의 암(i)의 모수(β_i)와 하나의 공변량(X_t)으로 나타내는 대신 수학적식 2와 같이 하나의 고정된 모수(β)와 암(i)에 따른 공변량($X_{i,t}^T$)으로도 나타낼 수 있다.

수학적식 2

$$r_i(t) = X_{i,t}^T \beta + \epsilon_{i,t}$$

[0044]

[0045] 또한, 학습자는 수학적식3을 통해 획득되는 리그렛(regret)의 평균적인 누적합(R(t))을 높은 확률로 최소화하는

것을 목적으로 할 수 있다.

수학식 3

$$R(T) = \sum_{t=1}^T \text{regret}(t)$$

수학식 4를 통해 획득할 수 있는 리그렛(regret)은 t 번째 단계에서 보상의 평균($X_t^T \beta_i$)을 최대화하는 행동이 $a^*(t)$ 인 경우, t 번째 단계에서의 최적의 보상과 학습자가 받는 보상의 차이를 의미할 수 있다.

수학식 4

$$\text{regret}(t) = X_t^T \beta_{a^*(t)} - X_t^T \beta_{a(t)}$$

또한, 밴딧 알고리즘의 보상은 행동의 성공 여부에 따라서 정해지며, 보상이 높을수록 해당 행동이 좋다는 것을 의미할 수 있다. 그러나 밴딧 알고리즘은 지도학습과 달리, 학습자가 선택하지 않은 행동에 대한 보상을 알지 못하기 때문에 매 시점마다 ‘활용(exploitation)’과 ‘탐색(exploration)’ 사이의 균형(trade-off)을 잘 맞추어야 한다. 여기서, ‘활용’은 현재까지 축적한 정보가 가리키는 최상의 행동을 선택하는 것을 의미할 수 있다. ‘탐색’은 현재에서 최상의 선택이 아닐 수 있지만 미래의 선택에 도움이 될 만한 정보를 줄 행동을 선택하는 것을 의미할 수 있다. ‘활용’과 ‘탐색’ 사이의 균형을 조절하기 위해 밴딧 알고리즘에서는 신뢰 상한(UCB, Upper Confidence Bound) 알고리즘을 사용할 수 있다. 이에 밴딧 알고리즘은 각 행동의 보상값에 대한 UCB를 계산한 후, UCB를 최대화하는 행동을 선택할 수 있다. UCB는 보상값의 추정치에 불확실성이 더해진 값으로 표현됨에 따라, 추정치가 크거나 불확실성이 크면 선택확률이 높아지므로 ‘활용’과 ‘탐색’이 자연스럽게 조절될 수 있다.

시점 t에서 i번째 행동에 대한 보상의 UCB는 수학식 5로 표현되는 전술한 수학식 2에 기반한 설정에서 모수(β)의 추정을 위한 최소 제곱 추정치와 수학식 6으로 표현되는 공변량의 그람(Gram) 행렬(B(t))에 기초하는 수학식 7에 의해 산출될 수 있다.

수학식 5

$$\hat{\beta}(t) = \left\{ \sum_{r=1}^{t-1} X_{a(r)}(r) X_{a(r)}(r)^T \right\}^{-1} \sum_{r=1}^{t-1} X_{a(r)}(r) r_{a(r)}(r)$$

수학식 6

$$B(t) = \sum_{r=1}^{t-1} X_{a(r)}(r) X_{a(r)}(r)^T$$

수학식 7

$$UCB_i(t) = X_i(t)^T \hat{\beta}(t) + \gamma \sqrt{X_i(t)^T B(t)^{-1} X_i(t)}$$

[0053]

[0054]

수학식 7에 있어서, $\hat{\beta}(t)$ 는 추정치를 나타내고, $\sqrt{X_i(t)^T B(t)^{-1} X_i(t)}$ 는 공변량($X_i(t)$)에 대한 현 시점의 불확실성을 나타낼 수 있다. 또한, γ 는 양의 상수로서, 보상값 에러의 분포에 따라 달라지는 값을 나타낼 수 있다.

그람 행렬($B(t)$)은 직전 시점까지 선택된 행동들의 공변량 $\{X_{a(r)}(r)\}_{r=1}^{t-1}$ 으로 표현되는 행렬이다. 이에 시점을 구간으로 나누어 해당 구간 내에서 같은 공변량을 선택하는 방식에 기초하여 추정치를 업데이트하는 방법을 사용할 수 있으나, 실제 데이터를 이용하는 경우에는 이를 적용시키기 어려운 단점이 있다.

[0055]

또한, 최소제곱법(least square) 또는 리지드 추정치(ridge estimator)를 사용한 UCB 알고리즘 성능은 전술한 수학식 1 및 수학식 2에서의 에러의 분포가 서브-가우시안인 경우에만 리그렛(regret)(t)의 누적합($R(T)$)이 높

은 확률로 $R(T) \leq O(d\sqrt{T} \log T)$ 를 만족하는 성능을 보장하는 문제가 있다. 에러의 분포가 서브-가우시안인 것은 에러의 극단값 쪽인 꼬리쪽 확률의 두께가 가우시안 분포(예: 정규분포)의 꼬리쪽 확률보다 얇거나 같은 경우를 의미할 수 있다. 따라서, 해당 UCB 알고리즘은 에러의 분포가 가우시안 분포의 꼬리쪽 확률보다 두꺼운 경우인 이상치나 극단값이 많은 보상분포 혹은 변동성이 큰 보상분포에 적합하지 않는 것을 의미할 수 있다.

[0056]

이에 일 실시예에 따른 밴딧 모델의 동작 및 트레이닝 방법 및 장치는 연속적인 공변량 선택이 필요하지 않으며 에러의 분포가 두꺼운 꼬리 분포를 가질 때에도 적용할 수 있다.

[0057]

도 1은 일 실시예에 따른 밴딧 모델의 동작 및 트레이닝 장치의 블록도를 도시한다.

[0058]

밴딧 모델의 동작 및 트레이닝 장치(100)는 프로세서(110) 및 메모리(120)를 포함할 수 있다.

[0059]

프로세서(110)는 목표 주기(t)까지 컨텍스트(context)(X_t)를 수집할 수 있다. 컨텍스트는 밴딧 모델이 다양한 암(arm)(예: 행동(action)) 중에 적절한 것을 선택하기 위한 정보를 의미할 수 있다. 예를 들어, 밴딧 모델이 온라인 광고를 추천하는 경우, 컨텍스트는 이용자의 나이, 성별 및 관심사 등일 수 있다.

[0060]

프로세서(110)는 목표 주기(t)가 강제 샘플링(forced sampling) 시간 스텝 세트($T_{i,t}$)에 포함되지 않는 경우, 컨텍스트(X_t)와 이전 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t-1})$)에 기초하여 복수의 암(i)들 중 하나 이상의 암을 제거(discard)할 수 있다. 프로세서(110)는 수학식 8을 이용하여 이전 목표 주기에서 가장 큰 보상(reward) 값($\max_{j \in [K]} X_t^T \hat{\beta}(T_{j,t-1})$)과 이전 목표 주기의 제1 추정치에 기초한 보상 값($X_t^T \hat{\beta}(T_{i,t-1})$) 간의 차가 미리 정해진 범위 값($\frac{h}{2}$)을 초과하는 암을 제거할 수 있다. 이는 임의의 암(i)의 이전 목표 주기의 제1 추정치에 기초한 보상 값($X_t^T \hat{\beta}(T_{i,t-1})$)이 이전 목표 주기에서 가장 큰 보상 값($\max_{j \in [K]} X_t^T \hat{\beta}(T_{j,t-1})$)과 유사할수록 해당 보상 값의 암이 최적의 암(i)에 해당되므로 세트($\mathcal{S}$)에 포함될 수 있다. 이전 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t-1})$)는 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트($T_{i,t-1}$)에 기초하여 추정될 수 있다. 또한, 이전 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t-1})$)가 사용하는 데이터 수는 $O(\log T)$ 로 리그렛(regret)의 누적합에 유의미한 영향을 끼치지 않으면서 최적의 행동이 아닌 행동들을 1차적으로 걸러줄 수 있는 적정의 수에 해당될 수 있다.

수학식 8

$$\mathcal{D} = \{i \in [K] \mid \max_{j \in [K]} X_i^T \hat{\beta}(T_{j,t-1}) - X_i^T \hat{\beta}(T_{i,t-1}) \leq \frac{h}{2}\}$$

수학식 8에 있어서, $[K]$ 는 전체 행동의 수로서, 전체 암(i) 세트를 나타낼 수 있다. $[K]$ 에는 차선의(suboptimal) 암(i)들과 최적의(optimal) 암(i)들로 구성되며, 이는 상호 중첩되지 않는다. h 는 하이퍼파라미터(hyperparameter)로서, 밴딧 모델의 동작 및 트레이닝 시작 시 입력되는 값일 수 있다.

또한, 강제 샘플링 시간 스텝 세트($T_{i,t}$)는 ‘활용(exploitation)’과 ‘탐색(exploration)’ 사이의 균형(trade-off)을 맞추기 위해 강제로 샘플링하는 주기를 의미할 수 있다. 이에 프로세서(110)는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되는 경우, 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암(i)을 풀(pull)할 수 있다. 이는 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대한 행동($a(t)$)이 미리 결정되어 해당 시점에 이르면 미리 정한 행동($a(t)$)을 선택하는 것을 의미할 수 있다.

프로세서(110)는 복수의 암들 중 제거된 하나 이상의 암 이외의 나머지 암들(예: 세트($\mathcal{D}$)) 중 하나의 암(i)을 컨텍스트와 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)에 기초하여 선택할 수 있다. 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)는 이전 목표 주기까지의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t-1}$)에 기초하여 추정될 수 있다. 또한, 프로세서(110)는 수학식 9를 통해 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)에 기초하여 산출된 보상 값($X_t^T \hat{\beta}(S_{i,t-1})$) 중 가장 큰 보상 값을 가진 암을 최적의 암으로 선택할 수 있다.

수학식 9

$$a(t) = \arg \max_{i \in \mathcal{D}} X_i^T \hat{\beta}(S_{i,t-1})$$

프로세서(110)는 목표 주기(t)에 기초하여 암이 풀되는 경우의 시간 스텝 세트($S_{a(t),t}$)를 업데이트하며, 선택된 암에 기초한 보상(y_t)을 획득할 수 있다. 여기서 $S_{a(t),t}$ 는 K 개의 암 중에 임의의 암(i)에 대한 목표 주기의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서 목표 주기(t)에 선택한 암($a(t)$)에 대한 목표 주기의 암이 풀되는 경우의 시간 스텝 세트를 의미할 수 있다. 목표 주기의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)는 강제 샘플링 시간 스텝 세트($T_{i,t}$)가 포함되는($T_{i,t} \subset S_{i,t}$) 전체 시간 스텝 세트를 의미할 수 있다.

프로세서(110)는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되는 경우, 수학식 10을 통해 목표 주기(t)의 강제 샘플링 시간 스텝 세트($T_{i,t}$)에서의 암에 기초하여 제1 경계 값($\tau(T_{i,t})$)을 산출할 수 있다.

수학식 10

$$\tau(T_{i,t}) = v_\delta (|T_{i,t}| / \log(t^2(2d+1)/\alpha))^{1/(1+\delta)}$$

수학식 10에 있어서, v_δ 는 오차의 분포와 관련되어 학습자가 알고 있다고 가정하는 값을 나타내고, α 는 하이퍼파라미터를 나타내고, d 는 행동(예: 암) 컨텍스트 벡터의 차원(dimension)을 나타낼 수 있다.

[0070] 프로세서(110)는 산출된 제1 경계 값($\tau(T_{i,t})$)과 강제 샘플링 시간 스텝 세트($T_{i,t}$)에서의 암에 기초하여 이전 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t-1})$)를 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t})$)로 업데이트할 수 있다. 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t})$)는 수학적 식 11을 이용하여 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 기초하여 가장 작은 손실함수(l_t)를 가지도록 업데이트될 수 있다. 수학적 식 11은 후술되는 수학적 식 13에 기초하여 도출될 수 있으며, 수학적 식 13의 특징에 따라 특정 시점의 오차가 과도한 비중을 차지하지 않는 특징을 또한 가질 수 있다.

수학적 식 11

[0071]
$$\hat{\beta}(T_{i,t}) = \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{|T_{i,t}|} \sum_{r \in T_{i,t}} l_{\tau}(y_r - X_r^T \beta)$$

[0072] 또한, 프로세서(110)는 목표 주기의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서의 암에 기초하여 제2 경계 값($\tau(S_{i,t})$)을 산출할 수 있다. 프로세서(110)는 산출된 제2 경계 값($\tau(S_{i,t})$)과 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서의 암에 기초하여 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)를 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)로 업데이트할 수 있다.

[0073] 또한, 프로세서(110)는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되지 않는 경우, 전술한 수학적 식 10에서 $T_{i,t}$ 부분이 $S_{i,t}$ 로 변경된 수식을 통해 목표 주기의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서의 암에 기초하여 제2 경계 값($\tau(S_{i,t})$)을 산출할 수 있다.

[0074] 프로세서(110)는 산출된 제2 경계 값($\tau(S_{i,t})$)과 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에 기초하여 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)를 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)로 업데이트할 수 있다. 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)는 전술한 수학적 식 11에서 $T_{i,t}$ 부분이 $S_{i,t}$ 로 변경된 수식을 통해 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에 기초하여 가장 작은 손실함수(l_t)를 가지도록 업데이트될 수 있다. 이는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되는 경우, 제1 경계 값($\tau(T_{i,t})$), 제1 추정치($\hat{\beta}(T_{i,t})$), 제2 경계 값($\tau(S_{i,t})$) 및 제2 추정치($\hat{\beta}(S_{i,t})$)를 산출하거나 업데이트하는 것을 의미할 수 있다. 그러나, 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되지 않는 경우, 제2 경계 값($\tau(S_{i,t})$)과 제2 추정치($\hat{\beta}(S_{i,t})$)만을 산출하거나 업데이트하는 것을 의미할 수 있다.

[0075] 또한, 손실함수(l_t)는 강제 샘플링 시간 스텝 세트($T_{i,t}$) 또는 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에 기초한 실제 보상과 예측 보상 간의 차이($y_r - X_r^T \beta$)(예: x)가 각 경계 값($\tau(T_{i,t})$ 또는 $\tau(S_{i,t})$) 이하이면 제곱 손실(square loss) 함수이고, 강제 샘플링 시간 스텝 세트($T_{i,t}$) 또는 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에 기초한 실제 보상과 예측 보상 간의 차이($y_r - X_r^T \beta$)(예: x)가 각 경계 값($\tau(T_{i,t})$ 또는 $\tau(S_{i,t})$)보다 크면 절대 손실(absolute loss) 함수일 수 있다. 이는 수학적 식 12로 나타낼 수 있다.

수학식 12

$$l_{\tau}(x) = \begin{cases} x^2/2, & \text{if } |x| \leq \tau, \\ \tau|x| - \tau^2/2, & \text{if } |x| > \tau, \end{cases}$$

수학식 13

$$\hat{\beta}(t) = \operatorname{argmin}_{\beta \in \mathbb{R}^d} 1/t \sum_{r=1}^{r=t} l_{\tau}(r_{a(r)}(r) - X_{a(r)}(r)^T \beta)$$

수학식 13을 이용하여 산출된 추정치($\hat{\beta}(t)$)는 논문 'Sun, Qiang, Wen-Xin Zhou, and Jianqing Fan. "Adaptive huber regression." Journal of the American Statistical Association 115, no. 529 (2020): 254-265.'에서 오차 상계가 $O(t^{\frac{\delta}{1+\delta}})$ 임이 증명되었지만 바로 밴딧 알고리즘에 적용할 수 없다. 수학식 13을 이용하여 산출된 추정치($\hat{\beta}(t)$)는 고정된 시점 t 에서 독립적이고 동일한 t 개의 데이터로 추정하였기 때문에 현재시점의 선택이 이전시점의 선택에 영향을 받는 밴딧 알고리즘에 적절하지 않다. 이에 논문 'Alexander Goldenshluger and Assaf Zeevi. "A linear response bandit problem" Stoch. Syst. 3(1): 230-261 (2013).' 및 'Hamsa Bastani and Mohsen Bayati. "Online Decision Making with High-Dimensional Covariates" Operations Research, INFORMS, vol. 68(1), pages 276-294, January'에서의 강제 샘플링(forced sampling) 기법과 슈퍼마팅게일 불평등(supermartingale inequality)을 이용하여 밴딧 알고리즘에 수학식 13을 적용할 수 있다.

밴딧 알고리즘에 수학식 13을 적용하기 위하여 독립적이고 동일한 데이터 조건은 두 부분에서 사용될 수 있다.

수학식 14

$$\|\hat{\beta}(S_{i,t}) - \beta_i\|_2 \leq 2|S_{i,t}|^{-1}c_l^{-1} \left\| \sum_{r \in S_{i,t}} \psi_{\tau}(\epsilon_r) X_r \right\|_2$$

하나는 비선형 함수인 손실함수(l_{τ})의 미분함수(ψ_{τ})의 합을 구하는 과정에서 사용될 수 있다. 여기서, 함수(ψ_{τ})는 손실함수(l_{τ})의 미분함수일 수 있다. 밴딧 알고리즘에 적용되기 위하여 미분함수(ψ_{τ})의 합을 구하는 과정은 독립적이고 동일한 데이터에 적용되는 부등식 대신에 상관관계가 있는 변수들에 적용가능한 슈퍼마팅게일 불평등(supermartingale inequality)을 사용할 수 있다.

$\exp\left(\sum_{r=1}^t \left[(X_{rj}/\tau)\psi_{\tau}(\epsilon_r) - \frac{\nu_{\delta}}{\tau^{1+\delta}}\right]\right)$ 로 정의된 슈퍼마팅게일(M_t)은 수학식 15를 통해 $\mathcal{F}_t = \sigma(X_1, \dots, X_{t+1}, a(1), a(2), \dots, a(t), y_1, \dots, y_t)$ 에 대하여 증명될 수 있다.

수학식 15

$$\begin{aligned}
 & \mathbb{E}[M_t | \mathcal{F}_{t-1}] \\
 &= M_{t-1} \mathbb{E}[\exp((X_{tj}/\tau)\psi_\tau(\epsilon_t)) | \mathcal{F}_{t-1}] / \exp\left(\frac{\nu_\delta}{\tau^{1+\delta}}\right) \\
 &\leq M_{t-1} \mathbb{E}[1 + X_{tj}(\epsilon_t/\tau) + |\epsilon_t/\tau|^{1+\delta} | \mathcal{F}_{t-1}] / \exp\left(\frac{\nu_\delta}{\tau^{1+\delta}}\right) \\
 &\leq M_{t-1} \left[1 + \frac{\nu_\delta}{\tau^{1+\delta}}\right] / \exp\left(\frac{\nu_\delta}{\tau^{1+\delta}}\right) \\
 &\leq M_{t-1} \exp\left(\frac{\nu_\delta}{\tau^{1+\delta}}\right) / \exp\left(\frac{\nu_\delta}{\tau^{1+\delta}}\right) \\
 &= M_{t-1}
 \end{aligned}$$

[0083]

[0084] 다른 하나는 그람 행렬의 최소 고유값인 $c_l \leq \lambda_{\min}(\frac{1}{t} \sum_{r \in [t]} X_r X_r^T)$ 을 구하는 과정에서 사용될 수 있다. 논문 'Tropp, Joel A. "User-friendly tail bounds for sums of random matrices." Foundations of computational mathematics 12 (2012): 389-434.'에서 증명된 바와 같이 독립적이고 동일한 데이터에 대해서는 고확률 상계가 증명되었다. 이에 밴딧 알고리즘에서 최소 고유치 값을 구하기 위해서는 강제 샘플링 기법이 사용될 수 있다. 강제 샘플링 기법은 특정 시점(t)(예: 목표 주기)에 대한 행동($a(t)$)이 미리 결정되어 해당 시점에 이르면 미리 정한 행동($a(t)$)을 선택하는 것을 의미할 수 있다. 이를 통해 강제 샘플링 시간 스텝 세트($T_{i,t}$) 내에서의 데이터는 독립적이며 동일한 분포를 가질 수 있다. 이러한 특성을 통해 밴딧 알고리즘에서도 전체 데이터의 최소 고유값 경계(minimum eigenvalue bound)(C_l)를 획득할 수 있다.

[0085]

또한, 일 실시예에 따른 밴딧 모델은 전술한 방식을 적용함에 따라 수학식 16에서의 리그렛 오더 이하인 리그렛(regret)의 누적함을 가질 수 있다. 이에 따라 일 실시예에 따른 밴딧 모델은 랜덤 에러가 서브-가우시안 분포를 따르는 것이 아닌 두꺼운 꼬리 분포를 가지는 경우에도 적용될 수 있다.

수학식 16

$$O(\sqrt{dT}^{\frac{1}{1+\delta}} (\log dT)^{\frac{\delta}{1+\delta}})$$

[0086]

[0087]

메모리(120)는 밴딧 모델의 초기 트레이닝을 수행하기 위한 데이터를 저장할 수 있으나, 이에 한정되지 않는다.

[0088]

다양한 실시예에 따르면, 메모리(120)는 휘발성 메모리 장치 또는 비휘발성 메모리 장치로 구현될 수 있다. 휘발성 메모리 장치는 DRAM(dynamic random access memory), SRAM(static random access memory), T-RAM(thyristor RAM), Z-RAM(zero capacitor RAM), 또는 TTRAM(Twin Transistor RAM)으로 구현될 수 있다. 비휘발성 메모리 장치는 EEPROM(Electrically Erasable Programmable Read-Only Memory), 플래시(flash) 메모리, MRAM(Magnetic RAM), 스핀전달토크 MRAM(Spin-Transfer Torque(STT)-MRAM), Conductive Bridging RAM(CBRAM), FeRAM(Ferroelectric RAM), PRAM(Phase change RAM), 저항 메모리(Resistive RAM(RRAM)), 나노 튜브 RRAM(Nanotube RRAM), 폴리머 RAM(Polymer RAM(PoRAM)), 나노 부유 게이트 메모리(Nano Floating Gate Memory(NFGM)), 홀로그래픽 메모리(holographic memory), 분자 전자 메모리 소자(Molecular Electronic Memory Device), 및/또는 절연 저항 변화 메모리(Insulator Resistance Change Memory)로 구현될 수 있다.

[0089]

다양한 실시예에 따르면, 프로세서(110)는 메모리(120)에 저장된 컴퓨터로 읽을 수 있는 코드(예를 들어, 소프

트웨어) 및 프로세서(110)에 의해 유발된 인스트럭션(instruction)들을 실행할 수 있다. 프로세서(110)는 목적하는 동작들(desired operations)을 실행시키기 위한 물리적인 구조를 갖는 회로를 가지는 하드웨어로 구현된 데이터 처리 장치일 수 있다. 목적하는 동작들은 예를 들어, 프로그램에 포함된 코드(code) 또는 인스트럭션들(instructions)을 포함할 수 있다. 하드웨어로 구현된 데이터 처리 장치는 예를 들어, 마이크로프로세서(microprocessor), 중앙 처리 장치(central processing unit), 프로세서 코어(processor core), 멀티-코어 프로세서(multi-core processor), 멀티프로세서(multiprocessor), ASIC(Application-Specific Integrated Circuit), FPGA(Field Programmable Gate Array)를 포함할 수 있다.

- [0090] 이외에도 밴딩 모델의 동작 및 트레이닝 장치(100)는 통신부를 추가로 구성하여 외부 구성(예: 사용자 플랫폼)으로부터 밴딩 모델에 적용하기 위한 데이터를 수신받을 수 있다.
- [0091] 도 2는 일 실시예에 따른 밴딩 모델의 동작 및 트레이닝 방법의 흐름도를 도시한다.
- [0092] 단계(210)에서, 프로세서는 목표 주기까지 컨텍스트를 수집할 수 있다.
- [0093] 단계(220)에서, 프로세서는 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되지 않는 경우, 컨텍스트와 이전 목표 주기의 제1 추정치에 기초하여 복수의 암들 중 하나 이상의 암을 제거할 수 있다. 이전 목표 주기의 제1 추정치는 이전 목표 주기까지의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 추정될 수 있다.
- [0094] 단계(230)에서, 프로세서는 복수의 암들 중 제거된 하나 이상의 암 이외의 나머지 암들 중 하나의 암을 컨텍스트와 이전 목표 주기의 제2 추정치에 기초하여 선택할 수 있다. 이전 목표 주기의 제2 추정치는 이전 목표 주기까지의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 추정될 수 있다.
- [0095] 단계(240)에서, 프로세서는 목표 주기에 기초하여 암이 풀되는 경우의 시간 스텝 세트를 업데이트할 수 있다.
- [0096] 단계(250)에서, 프로세서는 선택된 암에 기초한 보상을 획득할 수 있다.
- [0097] 단계(260)에서, 프로세서는 목표 주기가 강제 샘플링 시간 스텝 세트에 포함되는 경우, 목표 주기의 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 제1 경계(bound) 값을 산출할 수 있다.
- [0098] 단계(270)에서, 프로세서는 산출된 제1 경계 값과 강제 샘플링 시간 스텝 세트에서의 암에 기초하여 이전 목표 주기의 제1 추정치를 목표 주기의 제1 추정치로 업데이트할 수 있다.
- [0099] 단계(280)에서, 프로세서는 목표 주기의 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 제2 경계 값을 산출할 수 있다.
- [0100] 단계(290)에서, 프로세서는 산출된 제2 경계 값과 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초하여 이전 목표 주기의 제2 추정치를 목표 주기의 제2 추정치로 업데이트할 수 있다.
- [0101] 일 실시예에 따른 밴딩 모델의 동작 및 트레이닝 방법은 표1의 알고리즘을 이용하여 구체적으로 설명하도록 한다.

㉠ 1

Algorithm 1

```

1: Input:  $h, \nu_\delta, \alpha$ 
2:  $\hat{\beta}(T_{i,0}) = \hat{\beta}(S_{i,0}) = 0^d$ 
3: for  $t \in [T]$  do
4:   Observe  $X_t \sim \mathcal{P}_{\mathcal{X}}$ 
5:   if  $t \in T_i$  then
6:      $a(t) = i$ 
7:   else
8:      $\mathcal{D} = \{i \in [K] \mid \max_{j \in [K]} X_t^T \hat{\beta}(T_{j,t-1}) - X_t^T \hat{\beta}(T_{i,t-1}) \leq \frac{h}{2}\}.$ 
9:      $a(t) = \arg \max_{i \in \mathcal{D}} X_t^T \hat{\beta}(S_{i,t-1})$ 
10:  end if
11:  Update  $S_{a(t),t} = S_{a(t),t-1} \cup \{t\}$ 
12:  Observe reward  $y_t = X_t^T \beta_{a(t)} + \varepsilon_t$ 
13:  if  $t \in T_i$  then
14:     $\tau(T_{i,t}) = \nu_\delta(|T_{i,t}| / \log(t^2(2d+1)/\alpha))^{1/(1+\delta)}$ 
15:     $\hat{\beta}(T_{i,t}) = \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{|T_{i,t}|} \sum_{r \in T_{i,t}} l_\tau(y_r - X_r^T \beta)$ 
16:  else
17:     $\tau(S_{i,t}) = \nu_\delta(|S_{i,t}| / \log(t^2(2d+1)/\alpha))^{1/(1+\delta)}$ 
18:     $\hat{\beta}(S_{i,t}) = \arg \min_{\beta \in \mathbb{R}^d} \frac{1}{|S_{i,t}|} \sum_{r \in S_{i,t}} l_\tau(y_r - X_r^T \beta)$ 
19:  end if
20: end for

```

[0102]

[0103] 알고리즘의 1번째 줄에서, 프로세서는 하이퍼파라미터(h, ν_δ, α)를 입력할 수 있다.

[0104] 알고리즘의 2번째 줄에서, 프로세서는 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t})$) 및 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)의 초기값을 임의로 d차원의 0벡터로 설정할 수 있다.

[0105] 알고리즘의 3,4번째 줄에서, 프로세서는 목표 주기(t)까지 컨텍스트(X_t)를 수집할 수 있다. 이는 프로세서가 목표 주기(t)의 컨텍스트(X_t)만을 수집하는 것을 의미할 수 있다.

[0106] 알고리즘의 5,6번째 줄에서, 프로세서는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되는 경우, 강제 샘플링 시간 스텝 세트의 인덱스(index)에 대응되는 암(i)을 풀할 수 있다.

[0107] 알고리즘의 7, 8번째 줄에서, 프로세서는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되지 않는 경우, 이전 목표 주기에서 가장 큰 보상 값($\max_{j \in [K]} X_t^T \hat{\beta}(T_{j,t-1})$)과 이전 목표 주기의 제1 추정치에 기초한 보상 값($X_t^T \hat{\beta}(T_{i,t-1})$) 간의 차가 미리 정해진 범위 값($\frac{h}{2}$)을 초과하는 암을 제거할 수 있다.

[0108] 알고리즘의 9번째 줄에서, 프로세서는 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)에 기초하여 산출된 보상 값($X_t^T \hat{\beta}(S_{i,t-1})$) 중 가장 큰 보상 값을 가진 암(i)을 선택할 수 있다.

[0109] 알고리즘의 11번째 줄에서, 프로세서는 목표 주기(t)에 기초하여 암(i)이 풀되는 경우의 시간 스텝 세트

$(S_{a(t),t})$ 를 업데이트할 수 있다.

- [0110] 알고리즘의 12번째 줄에서, 프로세서는 시간 경계(time horizon)(T)(예: 전체 시간) 동안 선택된 암(i)에 기초한 보상(y_t)을 획득할 수 있다.
- [0111] 알고리즘의 13, 14번째 줄에서, 프로세서는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되는 경우, 목표 주기(t)의 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 기초하여 제1 경계 값($\tau(T_{i,t})$)을 산출할 수 있다.
- [0112] 알고리즘의 15번째 줄에서, 프로세서는 산출된 제1 경계 값($\tau(T_{i,t})$)과 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 기초하여 이전 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t-1})$)를 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t})$)로 업데이트할 수 있다. 목표 주기의 제1 추정치($\hat{\beta}(T_{i,t})$)는 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 기초하여 가장 작은 손실함수(l_τ)를 가지도록 업데이트될 수 있다.
- [0113] 알고리즘의 16, 17번째 줄에서, 프로세서는 목표 주기의 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서의 암에 기초하여 제2 경계 값($\tau(S_{i,t})$)을 산출할 수 있다.
- [0114] 알고리즘의 15번째 줄에서, 프로세서는 산출된 제2 경계 값($\tau(S_{i,t})$)과 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에서의 암에 기초하여 이전 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t-1})$)를 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)로 업데이트할 수 있다. 목표 주기의 제2 추정치($\hat{\beta}(S_{i,t})$)는 암이 풀되는 경우의 시간 스텝 세트($S_{i,t}$)에 기초하여 가장 작은 손실함수(l_τ)를 가지도록 업데이트될 수 있다. 알고리즘의 15번째 줄과 18번째 줄에서의 수학식은 전술한 수학식 13에서 사용하는 시간들의 집합을 변경하여 도출된 수학식으로서, 이를 통해 이전 시점의 선택이 반영되어 최적의 추정치가 추정될 수 있다. 또한, 프로세서는 목표 주기(t)가 강제 샘플링 시간 스텝 세트($T_{i,t}$)에 포함되지 않는 경우, 제1 경계 값($\tau(T_{i,t})$) 및 제1 추정치($\hat{\beta}(T_{i,t})$)의 산출 및/또는 업데이트없이 제2 경계 값($\tau(S_{i,t})$) 및 제2 추정치($\hat{\beta}(S_{i,t})$)만을 산출하거나 업데이트할 수 있다.
- [0115] 손실함수는, 강제 샘플링 시간 스텝 세트 또는 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값 이하이면 제곱 손실 함수이고, 강제 샘플링 시간 스텝 세트 또는 암이 풀되는 경우의 시간 스텝 세트에서의 암에 기초한 실제 보상과 예측 보상 간의 차이가 각 경계 값보다 크면 절대 손실 함수일 수 있다.
- [0116] 도 3 내지 도 5는 시간에 따라 일 실시예에 따른 밴딧 모델의 동작 및 트레이닝 방법과 다른 기술들을 이용하여 산출된 리그렛(regret)의 누적합(R(T))을 나타낸다.
- [0117] 실험은 총 100번 진행하였고, [K]는 10, d는 5, 전체시간([T])은 1000으로 설정하여 실험하였다. 또한, 모수(β_i)와 공변량(X_i)은 -1 부터 1 까지의 균등분포에서 생성하고, L_2 -norm이 1이 되게 하는 $\|X_t\|_2 \leq 1, \|\beta_i\|_2 \leq 1$ 를 수행하였다. 또한, 에러(ϵ_t)는 {1.1, 1.5, 2}의 자유도(df, degree of freedom)를 가진 t-분포에서 생성하였다. 도 3 내지 도 5에서의 라인(line)은 평균 값을 나타내고, 그림자는 분산값을 나타낸다. 전체적으로 일 실시예에 따른 밴딧 모델의 리그렛(regret)의 누적합(R(T))(310, 410, 510)은 다른 기술들의 리그렛의 누적합에 비하여 제일 낮은 값으로 형성됨에 따라 가장 뛰어난 성능을 가지고 있음을 확인할 수 있다.
- [0118] 이상에서 설명된 실시예들은 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치, 방법 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 범용 컴퓨터 또는 특수 목적 컴퓨터를

이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0119] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 저장될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0120] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있으며 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

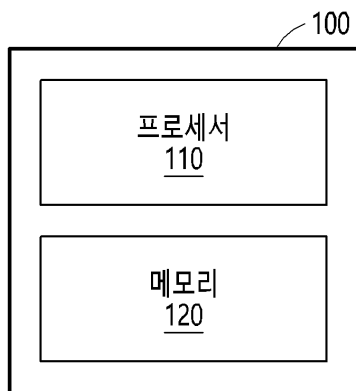
[0121] 위에서 설명한 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 또는 복수의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0122] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

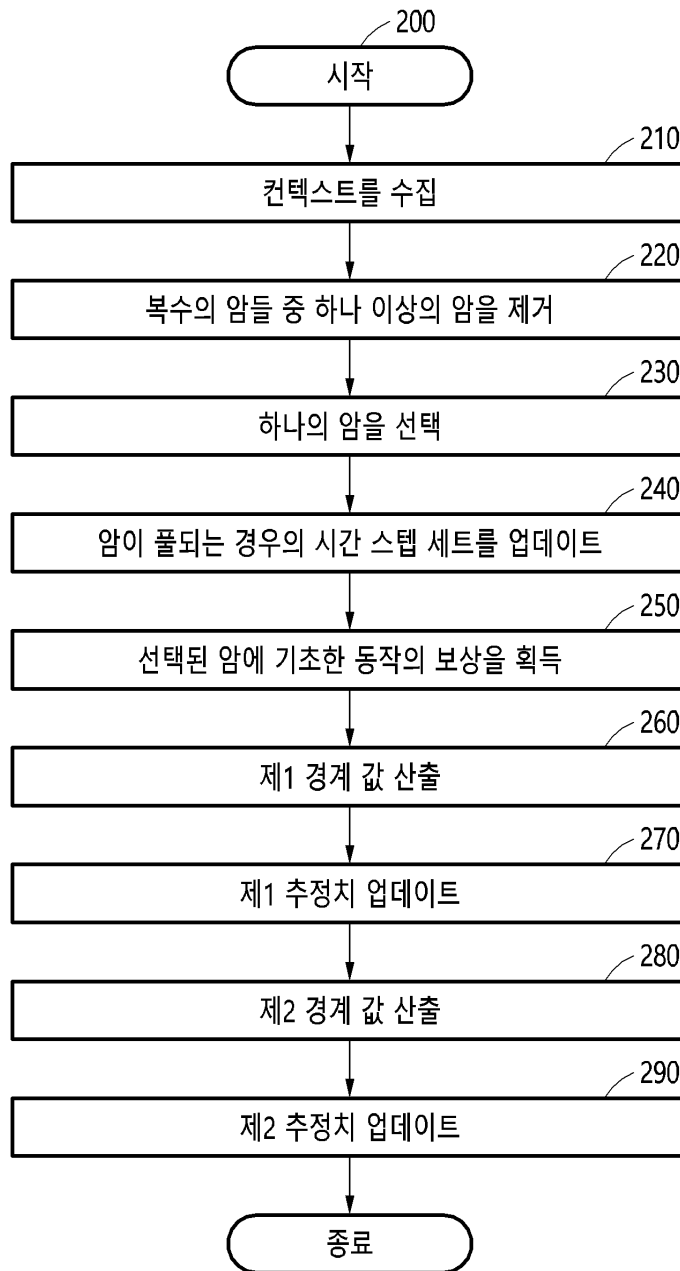
[0123] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

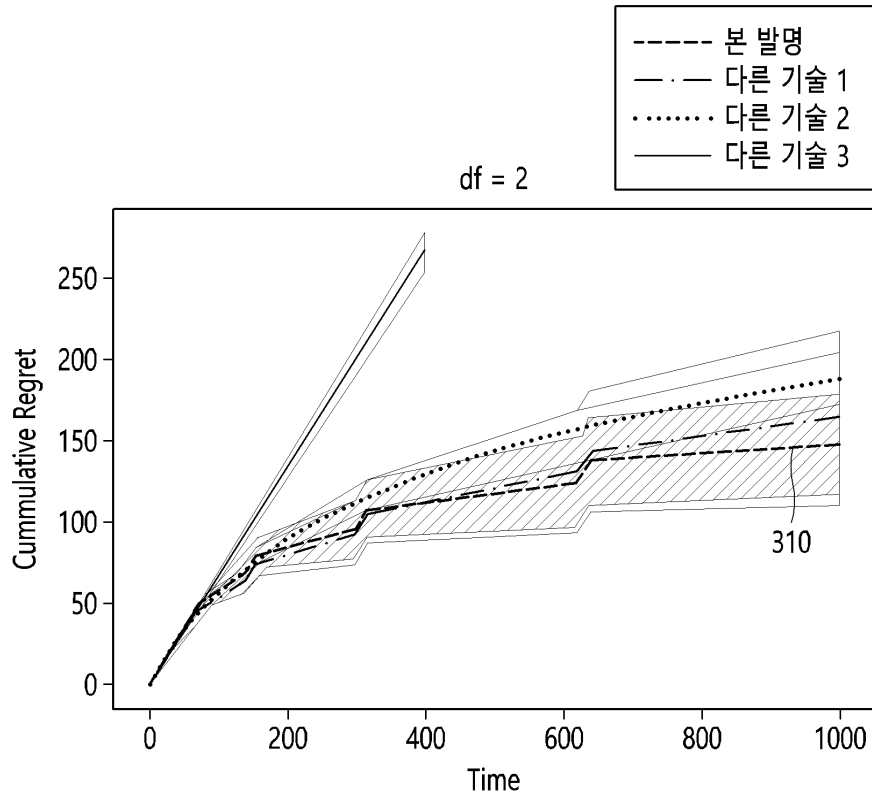
도면1



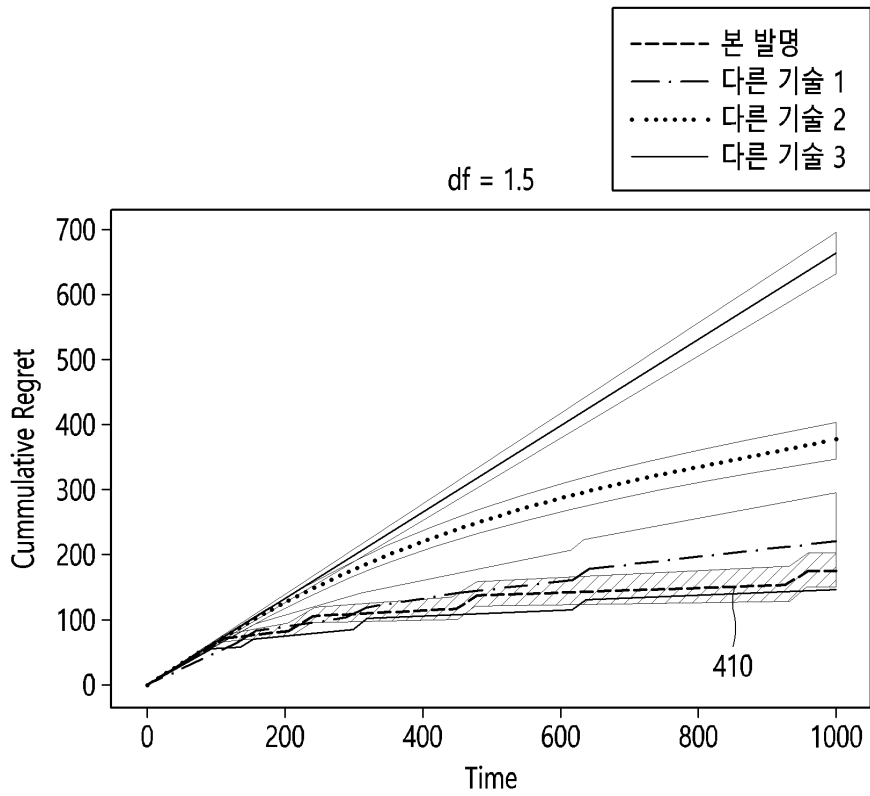
도면2



도면3



도면4



도면5

