(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0021905
(43) 공개일자 2023년02월14일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) *G06N 5/02* (2023.01)
G06N 7/00 (2023.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06N 5/02 (2023.01)
(21) 출원번호 10-2021-0103852
(22) 출원일자 2021년08월06일
심사청구일자 2021년08월06일

(71) 출원인
서울대학교산학협력단
서울특별시 관악구 관악로 1 (신림동)
숙명여자대학교산학협력단
서울특별시 용산구 청파로47길 100 (청파동2가, 숙명여자대학교)
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
조명희
서울특별시 강남구 언주로30길 57, G동 1002호 (도곡동, 타워팰리스)
백승훈
경기도 고양시 일산서구 강선로 70, 802동 204호 (주엽동, 강선마을8단지아파트)
(뒷면에 계속)
(74) 대리인
특허법인엠에이피에스

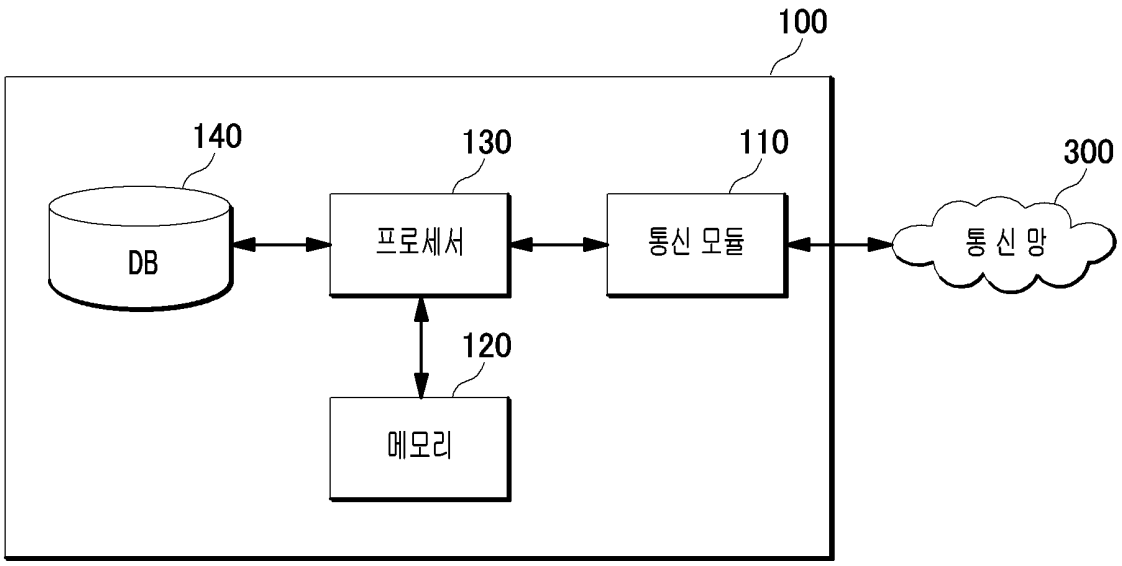
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신 문제의 해를 산출하는 장치 및 방법**

(57) 요 약

본 발명의 일 측면에 따른 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신(contextual multi-armed bandit) 문제의 해를 산출하는 방법은 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 수학식을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출하는 단계를 포함하되, 톰슨 샘플링(Thompson sampling) 알고리즘에 따라 하기의 사용자 선호도가 샘플링 되도록 한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 7/01 (2023.01)
(72) 발명자
최영근
서울특별시 용산구 청파로47길 100, 사회교육관 503호 (청파동2가, 숙명여자대학교)
김지수
울산광역시 울주군 언양읍 유니스트길 50, 112동 301-13호 (언양읍, 울산과학기술원)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711116545
과제번호 2020R1A2C1A01011950
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 불완전 및 순차적 자료에 적합한 심화학습 방법론 개발과 생의학 자료에의 적용
기 여 율 65/100
과제수행기관명 서울대학교
연구기간 2020.03.01 ~ 2021.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1711116471
과제번호 2020R1G1A1A01006229
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 규제된 다수준 모형 기반의 공간적 위험 맵핑과 이상지역 탐지
기 여 율 25/100
과제수행기관명 숙명여자대학교
연구기간 2021.03.01 ~ 2022.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1711131922
과제번호 2021R1G1A1009801
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 순차적 의사 결정 문제를 위한 다중 슬롯 머신 방법론 개발 및 응용
기 여 율 10/100
과제수행기관명 울산과학기술원
연구기간 2021.01.01 ~ 2022.02.28

명 세 서

청구범위

청구항 1

다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신(contextual multi-armed bandit) 문제의 해를 산출하는 방법에 있어서,

시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 하기의 수학식 1을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출하는 단계를 포함하되, 톰슨 샘플링(Thompson sampling) 알고리즘에 따라 하기의 사용자 선호도가 샘플링되는 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

[수학식 1]

$$\sum_{j=1}^{n}\sum_{\tau\in T_{j,t-1}}\left(r_{a(\tau),j}(\tau)-b_{a(\tau)}(\tau)^{\top}\mu_j\right)^2+\lambda\mu^{\top}(L\otimes I_d)\mu$$

j 는 n명의 사용자 각각을 대표함.

$T_{j,t-1}=\{\tau : j_\tau=j, 1\le\tau\le t\}$ 임.

$b_i(\tau)\in\mathbb{R}^d$ 는 사용자의 행동(i)의 시간(τ)에서의 상황 벡터를 나타냄.

a(τ)는 시간 τ 에서 사용자가 선택한 행동을 나타냄.

$r_{a(\tau),j}$ 사용자 j가 시간 τ 에서 선택한 행동에 따른 보상을 나타냄.

$\mu_j\in\mathbb{R}^d$ 는 사용자의 선호도를 나타내는 사용자별 계수임.

L 은 사용자의 관계를 나타내는 그래프의 라플라시안을 나타냄.

$\otimes$ 는 각 행렬의 크로네커 곱(Kronecker product)을 나타냄.

I_d는 d 차원의 단위 행렬을 나타냄.

청구항 2

제1항에 있어서,

상기 보상을 나타내는 모델은 선형 보상 모델로서, 아래의 수학식 2에 의해 정의되는 것이며, 행동(i)의 시간(τ)에서의 상황 벡터와 추정량의 곱과 노이즈의 합으로 이루어진 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

[수학식 2]

$$r_{i,j}(t)=b_i(t)^{\top}\mu_j+\eta_{i,j}(t)$$

$\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

청구항 3

제2항에 있어서,

상기 각 사용자별로 최적의 행동을 산출하는 단계는

각 라운드 별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하는 단계를 포함하되,

사용자의 상태에 변경이 없는 사용자에 대해서는 이전 상태를 유지하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

청구항 4

제3항에 있어서,

상기 각 사용자가 선택하는 행동에 따른 보상을 반영하는 갱신을 수행하는 단계는,

수학식 3에 의해 정의되는 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 를 산출하고,

[수학식 3]

$$\hat{\mu}_{j_t} := \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{(-1)} \sum_{k \neq j_t} l_{j_t,k} \bar{\mu}_k(t)$$

$$B_k(t) := \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau) b_{a(\tau)}(\tau)^\top + \lambda l_{k,k} I_d$$

$$\bar{\mu}_k(t) := B_k(t)^{(-1)} \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau)\, r_{a(\tau),k}(\tau)$$

각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 샘플링하고, 샘플링된 d차원 확률 변수에 따라 보상이 최대가 될 것으로 추정되는 행동($a(t) = argmax_i \left\{ b_i(t)^\top \tilde{\mu}_{j_t}(t) \right\}$)을 선택하도록 하고, 선택된 행동에 따른 실제 보상을 획득하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

청구항 5

제1항에 있어서,

상기 보상을 나타내는 모델은 준모수 보상 모델로서, 수학식 4에 의해 정의되는 것이며, 행동(i)의 시간(τ)에서의 상황 벡터와 추정량의 곱과 노이즈 및 절편(v(t))의 합으로 이루어진 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

[수학식 4]

$$r_{i,j}(t) = v(t) + b_i(t)^\top \mu_j + \eta_{i,j}(t)$$

$\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

청구항 6

제5항에 있어서,

상기 각 사용자별로 최적의 행동을 산출하는 단계는

각 라운드 별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하는 단계를 포함하되,

사용자의 상태에 변경이 없는 사용자에 대해서는 이전 상태를 유지하는 것인, 다중 슬롯 머신 문제의 해를 산출

하는 방법.

청구항 7

제6항에 있어서,

상기 각 사용자가 선택하는 행동에 따른 보상을 반영하는 갱신을 수행하는 단계는,

수학식 5에 의해 정의되는 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 를 산출하고,

[수학식 5]

$$\hat{\mu}_{j_t} \coloneqq \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{-1} \sum_{k \neq j_t} l_{j_t k} \bar{\mu}_k(t)$$

$$B_k(t) = \hat{\Sigma}_{k,t} + \Sigma_{k,t} + \lambda l_{kk} I_d$$

$$\bar{\mu}_k(t) \coloneqq B_k(t)^{-1} \sum_{\tau \in T_{k,t-1}} 2X_\tau \, r_{a(\tau),k}(\tau)$$

$X_\tau = b_{a(\tau)}(\tau) - E(b_{a(\tau)}(\tau) \mid F_{\tau-1})$ 이고, E는 기대값을 나타냄.

$$F_{\tau-1} = \left\{ j_t, \{b_i(t)\}_{(i=1)}^{N} \right\} \cup \left(\bigcup_{\tau=1}^{t-1} \left\{ j_\tau, \{b_i(t)\}_{i=1}^{N_\tau}, a(\tau), r_{a(\tau),j_\tau}(\tau) \right\} \right)$$

$$\hat{\Sigma}_{k,t} = \sum_{\tau \in T_{k,t-1}} X_\tau X_\tau^\top$$

$$\Sigma_{k,t} = \sum_{\tau \in T_{k,t-1}} E(X_\tau X_\tau^\top | F_{\tau-1})$$

N은 사용자 수, $F_{\tau-1}$ 은 시간(τ)에서 학습자가 가진 정보를 나타냄.

각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 샘플링하고, 샘플링된 d 차원 확률 변수에 따라 보상이 최대가 될 것으로 추정되는 행동($a(t) = argmax_i \left\{ b_i(t)^\top \tilde{\mu}_{j_t}(t) \right\}$)을 선택하도록 하고, 선택된 행동에 따른 실제 보상을 획득하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 방법.

청구항 8

다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신(contextual multi-armed bandit) 문제의 해를 산출하는 장치에 있어서,

다중 슬롯 머신 문제 해 산출 프로그램이 저장된 메모리; 및

상기 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하며,

상기 다중 슬롯 머신 문제 해 산출 프로그램은 각 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 하기의 수학식 1을 최소화하는 알고리즘을 사용하여 각 사용자

별로 최적의 행동을 산출하되, 톰슨 샘플링(Thompson sampling) 알고리즘에 따라 하기의 사용자 선호도가 샘플링 되는 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

[수학식 1]

$$\sum_{j=1}^{n}\sum_{\tau\in T_{j,t-1}}\left(r_{a(\tau),j}(\tau)-b_{a(\tau)}(\tau)^{\top}\mu_j\right)^2+\lambda\mu^{\top}(L\otimes I_d)\mu$$

j 는 n명의 사용자 각각을 대표함.

$T_{j,t-1}=\{\tau:j_\tau=j,1\le\tau\le t\}$ 임.

$b_i(\tau)\in\mathbb{R}^d$ 는 사용자의 행동(i)의 시간(τ)에서의 상황 벡터를 나타냄.

a(τ)는 시간 τ에서 사용자가 선택한 행동을 나타냄.

$r_{a(\tau),j}$ 사용자 j가 시간 τ에서 선택한 행동에 따른 보상을 나타냄.

$\mu_j\in\mathbb{R}^d$ 는 사용자의 선호도를 나타내는 사용자별 계수임.

L 은 사용자의 관계를 나타내는 그래프의 라플라시안을 나타냄.

$\otimes$ 는 각 행렬의 크로네커 곱(Kronecker product)을 나타냄.

I_d는 d 차원의 단위 행렬을 나타냄.

청구항 9

제8항에 있어서,

상기 보상을 나타내는 모델은 선형 보상 모델로서, 아래의 수학식 2에 의해 정의되는 것이며, 행동(i)의 시간(τ)에서의 상황 벡터와 추정량의 곱과 노이즈의 합으로 이루어진 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

[수학식 2]

$$r_{i,j}(t)=b_i(t)^{\top}\mu_j+\eta_{i,j}(t)$$

$\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

청구항 10

제9항에 있어서,

상기 다중 슬롯 머신 문제 해 산출 프로그램은 상기 각 사용자별로 최적의 행동을 산출하기 위해, 각 라운드 별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하되,

사용자의 상태에 변경이 없는 사용자에 대해서는 이전 상태를 유지하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

청구항 11

제10항에 있어서,

상기 다중 슬롯 머신 문제 해 산출 프로그램은 상기 각 사용자가 선택하는 행동에 따른 보상을 반영하는 갱신을 수행하기 위해,

수학식 3에 의해 정의되는 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 를 산출하고,

[수학식 3]

$$\hat{\mu}_{j_t} := \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{(-1)} \sum_{k \neq j_t} l_{j_t,k} \bar{\mu}_k(t)$$

$$B_k(t) := \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau) b_{a(\tau)}(\tau)^{\top} + \lambda l_{k,k} I_d$$

$$\bar{\mu}_k(t) := B_k(t)^{(-1)} \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau)\, r_{a(\tau),k}(\tau)$$

각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 샘플링하고, 샘플링된 d차원 확률 변수에 따라 보상이 최대가 될 것으로 추정되는 행동($a(t) = argmax_i \left\{ b_i(t)^{\top} \tilde{\mu}_{j_t}(t) \right\}$)을 선택하도록 하고, 선택된 행동에 따른 실제 보상을 획득하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

청구항 12

제8항에 있어서,

상기 보상을 나타내는 모델은 준모수 보상 모델로서, 수학식 4에 의해 정의되는 것이며, 행동(i)의 시간(τ)에서의 상황 벡터와 추정량의 곱과 노이즈 및 절편(v(t))의 합으로 이루어진 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

[수학식 4]

$$r_{i,j}(t) = v(t) + b_i(t)^{\top} \mu_j + \eta_{i,j}(t)$$

$\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

청구항 13

제12항에 있어서,

상기 다중 슬롯 머신 문제 해 산출 프로그램은 상기 각 사용자별로 최적의 행동을 산출하기 위해, 각 라운드 별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하되,

사용자의 상태에 변경이 없는 사용자에 대해서는 이전 상태를 유지하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

청구항 14

제13항에 있어서,

상기 다중 슬롯 머신 문제 해 산출 프로그램은 상기 각 사용자가 선택하는 행동에 따른 보상을 반영하는 갱신을 수행하기 위해,

수학식 5에 의해 정의되는 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 를 산출하고,

[수학식 5]

$$\hat{\mu}_{j_t} \coloneqq \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{-1} \sum_{k \neq j_t} l_{j_t k} \bar{\mu}_k(t)$$

$$B_k(t) = \hat{\Sigma}_{k,t} + \Sigma_{k,t} + \lambda l_{kk} I_d$$

$$\bar{\mu}_k(t) \coloneqq B_k(t)^{-1} \sum_{\tau \in T_{k,t-1}} 2X_\tau \, r_{a(\tau),k}(\tau)$$

$X_\tau = b_{a(\tau)}(\tau) - E(b_{a(\tau)}(\tau) \mid F_{\tau-1})$ 이고, E는 기대값을 나타냄.

$$F_{\tau-1} = \left\{ j_t, \{b_i(t)\}_{(i=1)}^{N} \right\} \cup \left(\bigcup_{\tau=1}^{t-1} \left\{ j_\tau, \{b_i(t)\}_{i=1}^{N}, a(\tau), r_{a(\tau),j_\tau}(\tau) \right\} \right)$$

$$\hat{\Sigma}_{k,t} = \sum_{\tau \in T_{k,t-1}} X_\tau X_\tau^\top$$

$$\Sigma_{k,t} = \sum_{\tau \in T_{k,t-1}} E(X_\tau X_\tau^\top \mid F_{\tau-1})$$

N은 사용자 수, $F_{\tau-1}$ 은 시간(τ)에서 학습자가 가진 정보를 나타냄.

각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 샘플링하고, 샘플링된 d 차원 확률 변수에 따라 보상이 최대가 될 것으로 추정되는 행동($a(t) = argmax_i \left\{ b_i(t)^\top \tilde{\mu}_{j_t}(t) \right\}$)을 선택하도록 하고, 선택된 행동에 따른 실제 보상을 획득하는 것인, 다중 슬롯 머신 문제의 해를 산출하는 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신 문제의 해를 산출하는 장치 및 방법에 관한 것이고, 이를 응용하여 다수의 대상체 중 어느 하나를 추천하는 장치 및 방법을 제공하고자 한다.

배 경 기 술

[0002] 본 발명은 전통적인 강화학습 문제 중 하나인 상황별 다중 슬롯 머신(contextual multi-armed bandit, MAB) 문제에 대한 것이다. 강화학습은 학습자가 이전 정보를 활용하여 선택을 계속해 나가면서 누적 보상의 합이 최대가 되도록 하는 것이며, 상황별 MAB는 전통적인 강화학습 문제 중 하나이다. 상황별 MAB 문제를 해결하는 알고리즘은 뉴스 기사, 광고, 행동 개입 등의 여러 종류의 개인화된 추천 문제에서 그 효용성이 입증고 있으며, 광고뿐만 아니라 많은 분야에서 개인 데이터 양이 급증함에 따라, 상황별 MAB 문제를 이용한 사용자 맞춤형 추천 서비스는 더욱 증가할 것으로 예측되고 있다.

[0003] 상황별 다중 슬롯 머신 모형은 학습자가 ‘이용(exploitation)’과 ‘탐색(exploration)’의 균형을 맞추면서 순차적으로 행동(action)을 선택하는 문제이다. ‘이용’은 지금까지의 정보를 이용해 보상을 최대화하는 행동을 선택하는 것을 의미하고, ‘탐색’은 더 좋은 예측을 위해 상황(context)의 함수로서의 보상(reward) 메커니즘을 추론하는 것이다. 상황별 MAB 문제를 해결하는 알고리즘은 뉴스 기사, 광고, 행동 개입 등의 여러 종류의

개인화된 추천 문제에서 그 효용성이 입증되었다. 특히 여러 MAB를 해결하는 여러 알고리즘 중에서도 매 라운드마다 사후확률을 업데이트하는 톰슨 샘플링(TS, Thompson sampling) 기반 알고리즘들은 실증적인 이점을 보여주는 것으로 알려져 있다.

[0004] 여러 사용자가 관계되어 있는 현실 문제에서는, 소셜 네트워크 상의 사용자들의 관계를 부가정보로 이용할 수 있다. 많은 연구에서 소셜 네트워크에 대한 정보는 추천에 활용되었으며, 사용자별 계수(user-specific coefficient)를 사용해 개인의 선호를 학습하는데 있어 이러한 부가정보를 활용하는 그래프 기반의 상황별 MAB 알고리즘들도 제안되었다. 특히 Vaswani et al.(Vaswani, S., Schmidt, M., and Lakshmanan, L. V. Horde of Bandits Using Gaussian Markov Random Fields. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, pp. 690?699, 2017)은 사용자의 관계를 나타내는 그래프의 라플라시안(Laplacian)을 이용하여 사용자별 계수의 사후평균과 분산을 정규화하는 TS기반 알고리즘을 제시했다.

[0005] 확장성(Scalability)은 여러 사용자와 이들의 그래프를 사용하는 선형 보상 모델의 상황별 MAB에서 어려운 점 중 하나이다. 모델이 커짐에 따라 계산 비용 역시 증가하는데, 너무 빠른 계산 비용의 증가는 실사용에 있어 불리하기 때문이다.

[0006] 또한, 확장성 문제 외에도, 시간에 무관한 절편(intercept)을 가정하는 선형 보상 모델은 제한적이다. 예를 들어, 뉴스기사 추천시 평소 스포츠의 관심이 별로 없는 사용자들도 올림픽 기간에는 스포츠 기사에 관심을 가질 수 있고, 영화 추천에서도 아카데미상 발표 이후 사람들 사이에 다양한 추세가 있을 수 있다.

[0007] 본 발명은 상황별 다중 슬롯 머신 문제의 해결을 위해, 다중 사용자 상황에서 실현 가능한 컴퓨터 비용을 가지면서도 시간에 따른 추세를 반영할 수 있는 새로운 그래프 기반 톰슨 샘플링 기반 알고리즘을 제시한다. 또한, 새로이 제안한 두 알고리즘의 중간 과정으로, 알고리즘의 중추적 역할을 하는 새로운 국소적 추정량(local estimator)을 제안한다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신 문제의 해를 산출하는 장치 및 방법으로서, 선형 보상 모델을 사용하는 방법과 준보수 모상 모델을 사용하는 방법을 제공하는데 목적이 있다.

[0009] 다만, 본 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0010] 상술한 기술적 과제를 해결하기 위한 기술적 수단으로서, 본 발명의 일 측면에 따른 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신(contextual multi-armed bandit) 문제의 해를 산출하는 방법은 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 수학식을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출하는 단계를 포함하되, 톰슨 샘플링(Thompson sampling) 알고리즘에 따라 하기의 사용자 선호도가 샘플링 되도록 한다

[0011] 또한, 본 발명의 다른 측면에 따른 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신(contextual multi-armed bandit) 문제의 해를 산출하는 장치는 다중 슬롯 머신 문제 해 산출 프로그램이 저장된 메모리; 및 상기 메모리에 저장된 프로그램을 실행하는 프로세서를 포함하며, 상기 다중 슬롯 머신 문제 해 산출 프로그램은 각 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 수학식을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출하되, 톰슨 샘플링(Thompson sampling) 알고리즘에 따라 하기의 사용자 선호도가 샘플링 되도록 한다.

발명의 효과

[0012] 전술한 본원의 과제 해결 수단에 의하면, 다중 사용자에 대한 그래프 기반 상황별 다중 슬롯 머신 문제의 해를 효율적으로 산출할 수 있게 된다.

[0013] 특히, 본 발명에서 제안하는 선형 보상 모델의 경우 기존 방법들 보다 계산 속도 측면과 이론적인 후회 한계(regret bound) 측면 모두에서 더 향상된 효과를 가진다. 앞서 소개한 최신의 기술인 Vaswani et al. 에 비하여 고확률 후회 한계가 $\sqrt{n}$ 배만큼 작으며, 계산 복잡도는 n배(n은 전체 사용자의 수) 만큼 작은 것을 실험을 통해 확인하였다.

[0014] 또한, 본 발명에서 제안하는 준모수 보상 모델의 경우 시간에 따라 기본 보상이 변화하는 많은 현실적인 문제들에서, 시간에 따른 변화를 반영할 수 있는 효과를 제공한다. 또한, 다중 사용자 상황의 준모수 보상 모델에서 그래프를 이용한 최초의 알고리즘이며, 선형 보상 가정에서 앞서 설명한 선형 보상 모델과 마찬가지로, 고확률 후회 한계를 제공한다.

도면의 간단한 설명

[0015] 도 1은 본 발명의 일 실시예에 따른 다중 슬롯 머신 문제 해 산출 장치의 구성을 도시한 블록도이다.

도 2는 본 발명의 일 실시예에 따른 선형 보상 모델 알고리즘을 설명하기 위한 도면이다.

도 3은 본 발명의 일 실시예에 따른 준모수 보상 모델 알고리즘을 설명하기 위한 도면이다.

도 4는 본 발명의 일 실시예에 따른 다중 슬롯 머신 문제 해 산출 방법을 도시한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0016] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0017] 본원 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다.

[0018] 본원 명세서 전체에서, 어떤 부재가 다른 부재 “상에” 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.

[0019] 본 발명은 크게 2가지 알고리즘을 제안한다.

[0020] 첫 번째 알고리즘은 선형 보상 모델(linear reward model)을 사용하는 그래프 기반의 상황별 MAB 문제에 대해 TS(Thompson sampling)를 사용하는 알고리즘이다. 종래 기술과 대비하여 개선된 부분은 두 가지로, 알고리즘의 후회(regret)가 더 작은 고확률 상한(upper bound)을 가진다는 점, 사용자의 수나 계수(coefficient)의 차원이 증가하더라도 실제 사용이 용이한 더 좋은 확장성(scalability)을 가진다는 점이다.

[0021] 두 번째 알고리즘은 준모수 보상 모델(semi-parametric reward model)을 사용하는 그래프 기반의 상황별 MAB 문제에 대해 TS를 사용하는 알고리즘이다. 다중 사용자 상황에서의 준모수 보상 모형에서 그래프를 이용하는 최초의 알고리즘일 뿐만 아니라, 선형 보상 가정에서 우리가 새로이 개발한 알고리즘이 달성한 것과 같은 고확률 후회 상한을 가진다. 특히, 준모수 보상 모델은 시기적절한 트렌드를 잘 수용하는 것으로 나타났으며, 이를 통해 뉴스기사 추천시 평소 스포츠에 관심이 별로 없는 사용자들에게도 올림픽 기간에는 스포츠 기사에 관심을 가질 수 있다는 트렌드를 반영하여 스포츠 기사를 추천하거나, 영화에 관심이 없는 사용자들에게도 아카데미상 발표 이후에는 영화 기사에 관심을 가질 수 있다는 트렌드를 반영하여 영화 관련 기사를 추천하는 등의 서비스 제공이 가능하다.

[0022] 이하 첨부된 도면을 참고하여 본 발명의 일 실시예를 상세히 설명하기로 한다.

[0023] 도 1은 본 발명의 일 실시예에 따른 다중 슬롯 머신 문제 해 산출 장치의 구성을 도시한 블록도이다.

[0024] 도시된 바와 같이 다중 슬롯 머신 문제 해 산출 장치(100)는 통신 모듈(110), 메모리(120), 프로세서(130) 및 데이터베이스(140)를 포함할 수 있다.

[0025] 다중 슬롯 머신 문제 해 산출 장치(100)는 여러 추천 대상에 대한 추천 결과를 제시하는 서비스를 응용 서비스로 제공할 수 있다. 예를 들어, 사용자의 취향을 반영한 뉴스, 영화 또는 음악 컨텐츠 등 다양한 디지털 컨텐츠를 추천하는데 사용될 수 있다. 이러한, 응용 서비스로의 활용을 위해 다중 슬롯 머신 문제 해 산출 장치

(100)는 각 사용자의 프로필 정보, 각 사용자간의 연결성 정보, 각 사용자가 행동을 통해 선택하고자 하는 대상체에 대한 각종 정보, 행동의 수행 시점에서의 주변의 여러 상황 정보를 수집한다. 예를 들어, 각 사용자 간의 연결성 정보는 후술할 수학식 1에서 L 로 표현되는 것으로서, 각종 SNS 플랫폼에 가입된 사용자들 간의 친구관계 또는 팔로잉-팔로워 관계에 의해서 특정된다. 또한, 대상체에 대한 각종 상황 정보는 수학식 1에서 $b_i(\tau) \in \mathbb{R}^d$ 로 표현되는 것으로서, 예를 들어, 대상체가 음악 컨텐츠라고 할때, 해당 음악의 가수, 음원의 재생길이, 음원의 발매 시점, 음원의 구성 주파수 등에 대한 정보를 포함한다. 또 다른 예로서, 대상체가 뉴스라면, 기사의 작성자, 기사를 공개한 언론사, 기사의 텍스트 원문, 텍스트의 길이등의 각종 정보를 포함한다.

[0026] 또한, 다중 슬롯 머신 문제 해 산출 장치(100)는 각 사용자 단말로부터 각 개인 상태에 대한 여러 데이터를 수신하고, 그로부터 추천 서비스를 제공하는 서버로서 동작할 수 있다. 이러한 경우, 다중 슬롯 머신 문제 해 산출 장치(100)는 SaaS (Software as a Service), PaaS (Platform as a Service) 또는 IaaS (Infrastructure as a Service)와 같은 클라우드 컴퓨팅 서비스 모델에서 동작할 수 있다. 또한, 다중 슬롯 머신 문제 해 산출 장치(100)는 사설(private) 클라우드, 공용(public) 클라우드 또는 하이브리드(hybrid) 클라우드와 같은 형태로 구축될 수 있다.

[0027] 통신 모듈(110)은 사용자의 각종 정보 또는 상황 정보 등의 다양한 데이터를 외부 기기로부터 통신망(300)을 통해 수신할 수 있다. 통신 모듈(110)은 다른 네트워크 장치와 유무선 연결을 통해 제어 신호 또는 데이터 신호와 같은 신호를 송수신하기 위해 필요한 하드웨어 및 소프트웨어를 포함하는 장치일 수 있다.

[0028] 메모리(120)에는 각 사용자에 대한 정보를 기반으로 최적의 보상을 제공할 수 있는 최적의 행동을 추천하는 다중 슬롯 머신 문제 해 산출 프로그램이 저장된다. 이러한 메모리(120)에는 다중 슬롯 머신 문제 해 산출 장치(100)의 구동을 위한 운영 체제나 다중 슬롯 머신 문제 해 산출 프로그램의 실행 과정에서 발생되는 여러 종류가 데이터가 저장된다.

[0029] 이때, 다중 슬롯 머신 문제 해 산출 프로그램은, 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 수학식을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출한다. 예를 들어, 본 발명의 알고리즘을 영화 추천 시스템에 적용한다면, 행동은 추천 시스템이 화면에 표시하는 영화와 대응될 수 있고, 이에 대한 보상은 각 영화에 대한 사용자들의 피드백 (별점, 클릭여부 등)이 될 수 있다. 또한, 상황 벡터는 각 영화의 정보에 대한 상세 정보들, 예를 들면 영화의 장르, 감독, 박스오피스 성적, 비디오 영상의 디지털화된 벡터 (예를들어 가로축,세로축,시간축의 3D픽셀들), 비디오 음성의 디지털화된 벡터들 등의 다양한 정보를 포함할 수 있다.

[0030] 프로세서(130)는 메모리(120)에 저장된 프로그램을 실행하되, 다중 슬롯 머신 문제 해 산출 프로그램의 실행에 따라, 본 발명이 제안하는 선형 보상 모델에 기반한 알고리즘 또는 준모수 보상 모델에 기반한 알고리즘을 수행할 수 있다.

[0031] 이러한 프로세서(130)는 데이터를 처리할 수 있는 모든 종류의 장치를 포함할 수 있다. 예를 들어 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 이와 같이 하드웨어에 내장된 데이터 처리 장치의 일 예로써, 마이크로프로세서(microprocessor), 중앙처리장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티프로세서(multiprocessor), ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 처리 장치를 망라할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

[0032] 데이터베이스(140)는 프로세서(130)의 제어에 따라, 다중 슬롯 머신 문제 해 산출 프로그램의 실행에 필요한 데이터를 저장 또는 제공한다. 이러한 데이터베이스(140)는 메모리(120)와는 별도의 구성 요소로서 포함되거나, 또는 메모리(120)의 일부 영역에 구축될 수도 있다.

[0033] 도 2는 본 발명의 일 실시예에 따른 선형 보상 모델 알고리즘을 설명하기 위한 도면이다.

[0034] 먼저, 초기 설정 단계(1번 라인)로서 사용자의 관계를 나타내는 그래프의 튜닝 변수(tuning parameter) λ는 그래프 구조의 영향을 조절하는데 사용되며, 0 보다 크도록 설정된다.

[0035] 또한, TS 샘플링을 위한 정규분포를 나타내는 v는 $v = 2R\sqrt{dlog\{\left(\frac{T^4}{\delta}\right)(1+\lambda^{-1})\}} + 3\sqrt{\lambda}$ 로 설정된다.

[0036] 다음으로, 크게 2개의 반복 루프가 수행되는데, 각 시간 별로 수행되는 첫번째 루프(2~14 번 라인)와 그 내부에 포함되는 각 사용자별로 수행되는 두번째 루프(4~12번 라인)를 포함하며, 이를 통해 각 라운드(시간 t 별)별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하고(7~12번 라인), 변경이 없는 사용자에 대해서는 이전 상태를 유지한다(5~6번 라인). 이와 같이 종래에 알려진 알고리즘에 대비하여, 모든 사용자에 대한 변수를 매번 갱신하지 않고, 각 단계별로 특정 시간에 대한 사용자(j_t) 에 대한 변수만 업데이트 한다는 점에서, 계산 복잡도 차원을 감소시킬 수 있는 효과를 제공한다.

[0037] 이때, 각 사용자가 선택하는 행동은 각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 TS샘플링하고(9번 라인), $a(t) = argmax_i\left\{b_i(t)^\top \tilde{\mu}_{j_t}(t)\right\}$ 와 같이, 보상이 최대가 될 것으로 추정되는 행동 a(t)를 선택한다(10번 라인). 그리고, 그 행동에 따른 결과로서 실제 보상($r_{a(\tau),j}$)을 획득한다(10번 라인).

[0038] 이러한 알고리즘을 통해 시간별로 각 사용자가 선택한 행동에 따른 보상과 시간별로 각 사용자가 선택한 행동과 연관된 상황 벡터의 차이에 기반한 항과 그래프에서 연결된 사용자들간의 평탄함을 사용하기 위한 라플라시안 정규화 항을 포함하는 하기의 수학식 1을 최소화하는 알고리즘을 사용하여 각 사용자별로 최적의 행동을 산출한다.

[0039] [수학식 1]

[0040] $$\sum_{j=1}^{n}\sum_{\tau\in T_{j,t-1}}\left(r_{a(\tau),j}(\tau) - b_{a(\tau)}(\tau)^\top \mu_j\right)^2 + \lambda\mu^\top (L\otimes I_d)\mu$$

[0041] j 는 n명의 사용자 각각을 대표함.

[0042] $T_{j,t-1} = \{\tau : j_\tau = j, 1 \le \tau \le t\}$ 임.

[0043] $b_i(\tau)\in\mathbb{R}^d$ 는 사용자의 행동(i)의 시간(τ)에서의 상황 벡터를 나타냄.

[0044] a(τ)는 시간 τ 에서 사용자가 선택한 행동을 나타냄.

[0045] $r_{a(\tau),j}$ 사용자 j가 시간 τ 에서 선택한 행동에 따른 보상을 나타냄.

[0046] $\mu_j\in\mathbb{R}^d$ 는 사용자의 선호도를 나타내는 사용자별 계수임.

[0047] L 은 사용자의 관계를 나타내는 그래프의 라플라시안을 나타냄.

[0048] $\otimes$ 는 각 행렬의 크로네커 곱(Kronecker product)을 나타냄.

[0049] I_d는 d 차원의 단위 행렬을 나타냄.

[0050] 한편, 도 2에서 보상을 나타내는 모델은 선형 보상 모델이며, 선형 보상 모델은 아래의 수학식 2에 의해 정의된다. 선형 보상 모델은 행동(i)의 시간(τ)에서의 상황 벡터($b_i(t)$)와 추정량(μ_j)의 곱과 노이즈의 합으로 이루어진다.

[0051] [수학식 2]

[0052] $$r_{i,j}(t) = b_i(t)^{\top}\mu_j + \eta_{i,j}(t)$$

[0053] $\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

[0054] 그리고, 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 는 본 발명에서 새롭게 제안하는 것으로서 아래와 같은 수학식 3에 의해 정의되며, 도 2에서는 11번 라인의 형태로 실행된다.

[0055] 앞서 최소화하고자 하는 수학식 1에서 다른 모든 변수들은 고정된 상태이고, μ_{j_t}만 변화 가능한 경우, 이 식을 최소화하는 아이디어에서 수학식 3이 도출되었다.

[0056] [수학식 3]

[0057] $$\hat{\mu}_{j_t} := \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{(-1)} \sum_{k \neq j_t} l_{j_t,k}\bar{\mu}_k(t)$$

[0058] $$B_k(t) := \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau) b_{a(\tau)}(\tau)^{\top} + \lambda l_{k,k} I_d$$

[0059] $$\bar{\mu}_k(t) := B_k(t)^{(-1)} \sum_{\tau \in T_{kt-1}} b_{a(\tau)}(\tau)\, r_{a(\tau),k}(\tau)$$

[0060] 도 3은 본 발명의 일 실시예에 따른 준모수 보상 모델 알고리즘을 설명하기 위한 도면이다.

[0061] 먼저, 초기 설정 단계(1번 라인)로서 사용자의 관계를 나타내는 그래프의 튜닝 변수(tuning parameter) λ는 그래프 구조의 영향을 조절하는데 사용되며, 0 보다 크도록 설정된다.

[0062] 또한, TS 샘플링을 위한 정규분포를 나타내는 v는 $v = (4R + 12)\sqrt{dlog\{\left(\frac{T^4}{\delta}\right)(1+\lambda^{-1})\}} + 3\sqrt{\lambda}$ 로 설정된다.

[0063] 이후 2개의 반복 루프가 수행되는 과정이나, TS 샘플링을 통해 각 라운드(시간 t 별)별로, 각 사용자가 선택한 행동에 따른 보상을 반영하는 갱신을 수행하고(7~12번 라인), 변경이 없는 사용자에 대해서는 이전 상태를 유지하는(5~6번 라인) 과정은 도 2의 것과 동일하여 상세한 설명은 생략한다.

[0064] 도 3에서 보상을 나타내는 모델은 준모수 보상 모델로서, 아래의 수학식 4에 의해 정의된다. 준모수 보상 모델은 행동(i)의 시간(τ)에서의 상황 벡터($b_i(t)$)와 추정량(μ_j)의 곱, 노이즈 및 절편(ν(t))의 합으로 이루어진다.

[0065] [수학식 4]

[0066] $$r_{i,j}(t) = \nu(t) + b_i(t)^{\top}\mu_j + \eta_{i,j}(t)$$

[0067] $\eta_{i,j}$ 는 가우시안 분포를 갖는 노이즈를 나타냄.

[0068] 선형 보상 모델과 대비하여 사용자의 행동과는 독립적으로 시간에 따라 변화하는 절편(intercept, ν(t))를 추가한 것을 특징으로 한다.

[0069] 절편(ν(t))에 대한 가정은 $|\nu(t)| \leq 1$ 이며, 만약 $\nu(t) = 0$ 이면, 준모수 보상 모델은 선형 보상 모델로 환원된다.

[0070] 예를 들어, 영화를 추천하는 서비스에 본 발명을 적용할 경우, 각 영화의 특징과는 무관하게 시간에 따라 달라지는 트렌드로서, 사용자의 영화 시청 경향을 나타내기 위해 절편(v(t))이 추가된다. 이러한 절편에 의해 추가되는 사용자의 영화 시청 경향은 상황 벡터로는 표현하기 어려운 사용자의 감정, 상황, 주변 환경등에 의해 달라질 수 있음에 근거한 것이며, 이를 통해 사용자의 영화 시청 경향을 좀더 정밀하게 표현할 수 있다.

[0071] 한편, 준모수 항인 v(t)는 최적 행동(optimal arm)이나 후회(regret)의 정의에 변화를 미치지 않는데, 그 이유는 모든 행동에 대한 보상에 공통적으로 v(t)가 적용되기 때문이다. 하지만 그렇다고 v(t)를 무시하고 생각하는 것은 사용자의 선호도(μ_j)의 추정에 있어 혼란을 일으킬 수 있다. 왜냐하면, 매 시점 관찰되는 보상은 하나이기에 선형 부분과 준모수 부분의 영향을 구분하기 어렵기 때문이다. 이러한 어려움을 극복하기 위해 준모두 보상 모델을 새롭게 제안한 것이다.

[0072] 이때, 사용자의 선호도에 대한 추정량을 나타내는 $\hat{\mu}_{j_t}$ 는 수학식 5와 같이 정의된다. 수학식 5의 내용은 수학식 3에 의해 정의되는 추정량과는 달리, 상황 벡터의 중심화 (centering)를 통해 방해 변수(nuisance parameter)인 v(t)를 제거하는 방식을 사용했다.

[0073] [수학식 5]

$$\hat{\mu}_{j_t} \coloneqq \bar{\mu}_{j_t}(t) - \lambda B_{j_t}(t)^{-1} \sum_{k \neq j_t} l_{j_t k} \bar{\mu}_k(t)$$

[0074]

$$B_k(t) = \hat{\Sigma}_{k,t} + \Sigma_{k,t} + \lambda l_{kk} I_d$$

[0075]

$$\bar{\mu}_k(t) \coloneqq B_k(t)^{-1} \sum_{\tau \in T_{k,t-1}} 2X_\tau \, r_{a(\tau),k}(\tau)$$

[0076]

[0077] $X_\tau = b_{a(\tau)}(\tau) - E(b_{a(\tau)}(\tau) \mid F_{\tau-1})$ 이고, E는 기대값을 나타냄.

[0078] N은 사용자 수, $F_{\tau-1}$ 은 시간(τ)에서 학습자가 가진 정보를 나타낸 것으로, 아래와 같이 정의된다.

$$F_{\tau-1} = \left\{ j_t, \{b_i(t)\}_{(i=1)}^{N} \right\} \cup \left(\bigcup_{\tau=1}^{t-1} \left\{ j_\tau, \{b_i(t)\}_{i=1}^{N}, a(\tau), r_{a(\tau),j_\tau}(\tau) \right\} \right)$$

[0079]

$$\hat{\Sigma}_{k,t} = \sum_{\tau \in T_{k,t-1}} X_\tau X_\tau^\top$$

[0080]

$$\Sigma_{k,t} = \sum_{\tau \in T_{k,t-1}} E(X_\tau X_\tau^\top | F_{\tau-1})$$

[0081]

[0082] 이때, X_τ 와 $\Sigma_{k,t}$ 에서 기대값이 등장하는 이유는 주어진 $F_{\tau-1}$ 에서 a(t)를 선택할 때 무작위성이 들어가기 때문이다.

[0083] 이러한 기대값은 다음과 같이 계산할 수 있다.

[0084] 시간 τ 에서 i 번째 행동을 선택할 확률은 $\pi_i(\tau) = P(a(\tau) = i \mid F_{\tau-1})$ 이고, 이는 $\tilde{\mu}_{j_\tau}$의 사후분포

에 의해 결정된다.

$$E(b_{a(\tau)}(\tau) \mid F_{t-1}) = E\left(\sum_{i=1}^{N} I(a(\tau)=i)\, b_{a(\tau)}(\tau) \mid F_{\tau-1}\right) = \sum_{i=1}^{N} \pi_i(\tau) b_i(\tau)$$

[0085] 그러면, 가 도출된다.

[0086] 마찬가지 방식으로 $E(X_\tau X_\tau^\top | F_{\tau-1}) = \sum_{i=1}^{N} \pi_i(\tau)\,(b_i(\tau)-\bar{b}(\tau))\,(b_i(\tau)-\bar{b}(\tau))^\top$ 으로 계산할 수 있고, $\bar{b}(\tau) = E(b_{a(\tau)}(\tau)|F_{\tau-1})$ 이다.

[0087] 최종적으로는 각 라운드에서 $N_d\left(\hat{\mu}_{j_t}(t), v^2 B_{j_t}(t)^{-1}\right)$ 의 정규분포를 갖는 d차원 확률 변수 $\tilde{\mu}_{j_t}(t)$ 를 샘플링하고, 샘플링된 d 차원 확률 변수에 따라 보상이 최대가 될 것으로 추정되는 행동 ($a(t) = argmax_i \left\{ b_i(t)^\top \tilde{\mu}_{j_t}(t) \right\}$)을 선택하도록 하고, 선택된 행동에 따른 실제 보상을 획득한다.

[0088] 위와 같이 설명한 각 알고리즘의 성능을 살펴보면 다음과 같다.

[0089] 본 발명에 따른 선형 보상 모델의 경우 $O(d^2 + d \cdot \max_j \deg(j))$ 의 계산 복잡도를 가진다.

[0090] 도 2의 알고리즘에서와 같이, $j \neq j_t$ 인 경우, $B_j(t+1) = B_j(t),\ \bar{\mu}_j(t+1) = \bar{\mu}_j(t)$ 이므로, $j = j_t$에 대해서만, $B_j(t), \bar{\mu}_j$ 를 계산하면 되고, 이에 따라 $O(d^2 + d \cdot \max_j \deg(j))$ 의 계산 복잡도를 가지고, 사용자의 수 n과는 무관하므로, 계산 복잡도에서 n 의 차수를 하나 감소시킬 수 있다.

[0091] 도 4는 본 발명의 일 실시예에 따른 다중 슬롯 머신 문제의 해를 산출하는 방법을 도시한 순서도이다.

[0092] 먼저, 사용자에게 제공하고자 하는 추천 대상을 고려하여, 초기 설정 단계를 수행한다(S410).

[0093] 사용자들간의 관계를 설정하는 그래프를 생성할 수 있는데, 그래프는 노드와 링크로 구성되는 것으로서, 각각의 노드는 각 사용자를 나타낼 수 있다. 예를 들어, Erdos-Renyi 모델을 통해, 사용자의 수와 각 노드를 연결하는 확률을 설정하면, 정규 분포를 따르는 다양한 랜덤 그래프를 생성할 수 있다.

[0094] 또한, 추천하고자 하는 서비스의 특성에 맞게, 초기 설정의 λ 값을 조절한다. 예를 들어, 추천 과정에 친구 관계의 연결성이 큰 영향을 미친하면, λ 값을 크게 설정한다.

[0095] 다음으로, 앞서 설명한 선형 보상 모델 또는 준모수 보상 모델에 따른 알고리즘을 수행하여, 최적의 행동을 산출한다(S420).

[0096] 다음으로, 산출된 최적의 행동과 매칭되는 최적의 추천 결과를 제시한다(S430).

[0097] 예를 들어, 최적의 뉴스, 영화, 음악 등 다양한 컨텐츠를 추천할 수 있다.

[0098] 본 발명의 일 실시예는 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터에 의해 실행가능한 명령어를 포함하는 기록 매체의 형태로도 구현될 수 있다. 컴퓨터 판독 가능 매체는 컴퓨터에 의해 액세스될 수 있는 임의의 가용 매체일 수 있고, 휘발성 및 비휘발성 매체, 분리형 및 비분리형 매체를 모두 포함한다. 또한, 컴퓨터 판독 가능 매체는 컴퓨터 저장 매체를 포함할 수 있다. 컴퓨터 저장 매체는 컴퓨터 판독가능 명령어, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현된 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함한다.

[0099] 본 발명의 방법 및 시스템은 특정 실시예와 관련하여 설명되었지만, 그것들의 구성 요소 또는 동작의 일부 또는 전부는 범용 하드웨어 아키텍처를 갖는 컴퓨터 시스템을 사용하여 구현될 수 있다.

[0100] 전술한 본원의 설명은 예시를 위한 것이며, 본원이 속하는 기술분야의 통상의 지식을 가진 자는 본원의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만

한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.

[0101] 본원의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본원의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

[0102] 100: 다중 슬롯 머신 문제 해 산출 장치

110: 통신 모듈

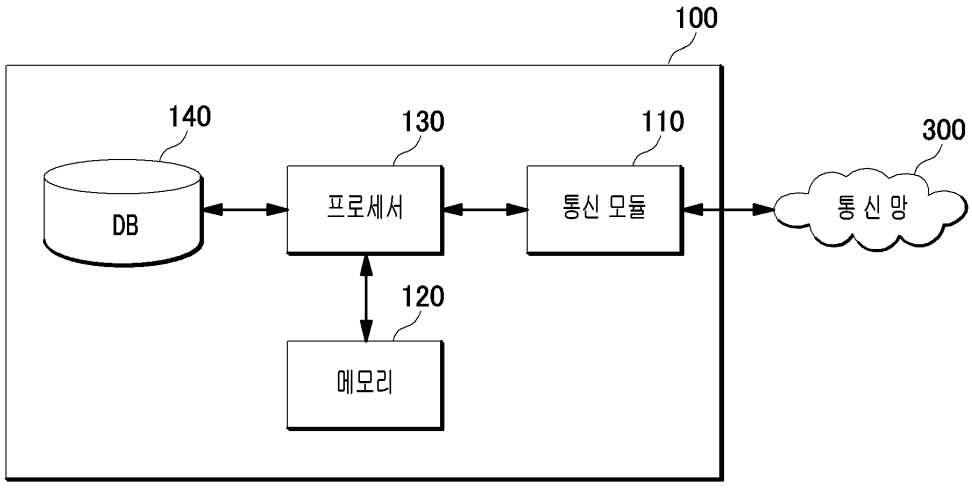
120: 메모리

130: 프로세서

140: 데이터베이스

도면

도면1



도면2

Algorithm 1 RGraphTS

Fix $\lambda > 0$. Set $B_j(1) = \lambda l_{jj} I_d$ and $y_j(1) = 0_d$ for $j = 1, \ldots, n$, and $v = 2R\sqrt{d \log\{(T^4/\delta)(1+\lambda^{-1})\}} + 3\sqrt{\lambda}$.
for $t = 1, 2, \ldots, T$ **do**
  Observe j_t.
  for $j = 1, 2, \ldots, n$ **do**
    if $j \neq j_t$ **then**
      Update $B_j(t+1) \leftarrow B_j(t)$, $\bar{\mu}_j(t+1) \leftarrow \bar{\mu}_j(t)$, and $y_j(t+1) \leftarrow y_j(t)$.
    else
      $\widehat{\mu}_j(t) \leftarrow \bar{\mu}_j(t) - B_j(t)^{-1} \sum_{k \neq j} \lambda l_{jk} \bar{\mu}_k(t)$.
      Sample $\widetilde{\mu}_j(t)$ from $\mathcal{N}_d(\widehat{\mu}_j(t), v^2 B_j(t)^{-1})$.
      Pull arm $a(t) = \operatorname{argmax}_{1 \leq i \leq N} \{b_i(t)^T \widetilde{\mu}_j(t)\}$ and get reward $r_{a(t),j}(t)$.
      Update: $B_j(t+1) \leftarrow B_j(t) + b_{a(t)}(t) b_{a(t)}(t)^T$, $y_j(t+1) \leftarrow y_j(t) + b_{a(t)}(t) r_{a(t),j}(t)$, and $\bar{\mu}_j(t+1) \leftarrow B_j(t+1)^{-1} y_j(t+1)$.
    end if
  end for
end for

도면3

Algorithm 2 Semi-RGraphTS

Fix $\lambda > 0$. Set $B_j(1) = \lambda l_{jj} I_d$ and $y_j(1) = 0_d$ for $j = 1, \ldots, n$, and $v = (4R + 12)\sqrt{d \log\{(T^4/\delta)(1 + \lambda^{-1})\}} + 3\sqrt{\lambda}$.
for $t = 1, 2, \ldots, T$ **do**
  Observe j_t.
  for $j = 1, 2, \ldots, n$ **do**
    if $j \neq j_t$ **then**
      Update $B_j(t+1) \leftarrow B_j(t)$, $\bar{\mu}_j(t+1) \leftarrow \bar{\mu}_j(t)$, and $y_j(t+1) \leftarrow y_j(t)$.
    else
      $\widehat{\mu}_j(t) \leftarrow \bar{\mu}_j(t) - B_j(t)^{-1} \sum_{k \neq j} \lambda l_{jk} \bar{\mu}_k(t)$.
      Sample $\widetilde{\mu}_j(t)$ from $\mathcal{N}_d(\widehat{\mu}_j(t), v^2 B_j(t)^{-1})$.
      Pull arm $a(t) = \text{argmax}_{1 \leq i \leq N}\{b_i(t)^T \widetilde{\mu}_j(t)\}$ and get reward $r_{a(t),j}(t)$.
      for $i = 1, \ldots, N$ **do**
        Compute $\pi_i(t) \leftarrow \mathbb{P}(a(t) = i | \mathcal{F}_{t-1})$.
      end for
      Compute $\bar{b}(t) \leftarrow \sum_{i=1}^{N} \pi_i(\tau) b_i(\tau)$.
      Compute $X_t \leftarrow b_{a(t)}(t) - \bar{b}(t)$.
      Update:
      $B_j(t+1) \leftarrow B_j(t) + X_t X_t^T + \sum_{i=1}^{N} \pi_i(t) X_t X_t^T$,
      $y_j(t+1) \leftarrow y_j(t) + 2 X_t r_{a(t),j}(t)$, and
      $\bar{\mu}_j(t+1) \leftarrow B_j(t+1)^{-1} y_j(t+1)$.
    end if
  end for
end for

도면4