(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월28일
(11) 등록번호 10-2426812
(24) 등록일자 2022년07월26일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) *G06N 7/00* (2022.01)
G06Q 50/20 (2012.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06N 7/005 (2013.01)
(21) 출원번호 10-2020-0107453
(22) 출원일자 2020년08월25일
심사청구일자 2020년08월25일
(65) 공개번호 10-2022-0026449
(43) 공개일자 2022년03월04일
(56) 선행기술조사문헌
KR102096301 B1*
KR1020010097914 A*
KR1020190025873 A
KR1020200072046 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
세종대학교산학협력단
서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(72) 발명자
김형석
서울특별시 광진구
무하마드 주베르 이슬람
서울특별시 광진구 능동로 209 세종대학교
카시프 메흐무드
서울특별시 광진구 능동로 209 세종대학교
(74) 대리인
김현승

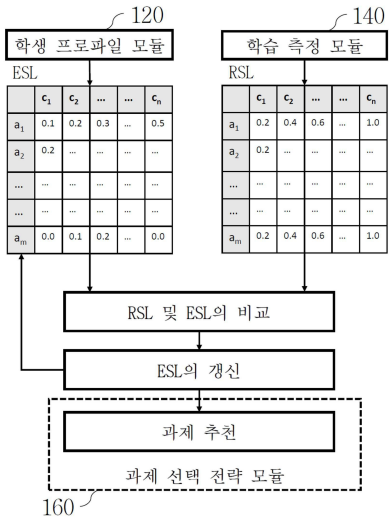
전체 청구항 수 : 총 7 항 심사관 : 박승철

(54) 발명의 명칭 **강화 학습 기반의 상호작용 향상 방식**

(57) 요 약

강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 방법이 제공된다. 개시된 방법은, 학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 위한 알고리즘의 매 라운드에서, 다수의 과제 각각의 선택 확률을 설정하는 것과, 설정된 선택 확률에 따라 다수의 과제 중 하나를 추천되는 과제로서 선택하는 것과, 학생의 주어진 역량 수준에서 선택된 과제의 요구되는 기량 수준 및 추정된 기량 수준을 비교하는 것과, 비교의 결과로부터 보상을 산출하고, 보상에 기반하여 추정된 기량 수준을 갱신하는 것과, 선택된 과제의 설정된 선택 확률 및 산출된 보상에 기반하여 상기 선택된 과제의 보상 추정치를 설정하는 것과, 보상 추정치를 사용하여, 선택된 과제가 다음에 선택될 확률의 감소를 방지하는 것을 포함하는 동작을 수행하는 단계를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
G06Q 50/20 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711110816
과제번호 2019R1A4A1023746
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 집단연구지원(R&D)
연구과제명 Tactile Internet을 위한 초저지연 고신뢰 햅틱 데이터 통신 기술 및 촉각 센서/액튜에이터 기술 연구
기 여 율 70/100
과제수행기관명 세종대학교
연구기간 2020.03.01 ~ 2021.02.28
이 발명을 지원한 국가연구개발사업
과제고유번호 1711108436
과제번호 2019R1F1A1060799
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)(R&D)
연구과제명 대규모 머신타입 통신/고신뢰 저지연 통신 성능 향상을 위한 평균 필드 게임 이론 및 비직교 다중 액세스 기반 기술 연구
기 여 율 30/100
과제수행기관명 세종대학교
연구기간 2020.03.01 ~ 2021.02.28

공지예외적용 : 있음

명 세 서

청구범위

청구항 1

강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 컴퓨팅 장치로서,

프로세서와,

프로세서 실행가능 명령어가 저장된 메모리를 포함하되, 상기 프로세서 실행가능 명령어는 상기 프로세서에 의해 실행되는 경우 상기 프로세서로 하여금, 학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 위한 알고리즘의 매 라운드에서 동작을 수행하게 하되, 상기 알고리즘은 적대적 밴딧(adversarial bandit) 설정에서의 상기 RL의 사용을 위한 Exp3 알고리즘(Exponential-weight algorithm for Exploration and Exploitation)이고, 상기 동작은,

상기 다수의 과제 각각의 선택 확률을 설정하는 것 - 상기 각각의 과제의 상기 선택 확률은 미선택된 과제의 탐색(exploration)을 가능하게 하는 균일 분포, 기선택된 과제의 이용(exploitation)을 가능하게 하는 가중 분포, 그리고 상기 탐색 및 상기 이용의 트레이드오프(tradeoff) 관계를 나타내는 파라미터에 기반하여 설정됨 - 과,

상기 설정된 선택 확률에 따라 상기 다수의 과제 중 하나를 추천되는 과제로서 선택하는 것과,

상기 선택된 과제의 요구되는 기량 수준 및 추정된 기량 수준을 비교하는 것과,

상기 비교의 결과로부터 보상(reward)을 산출하는 것과,

상기 산출된 보상에 기반하여 상기 추정된 기량 수준을 갱신하는 것과,

상기 선택된 과제의 상기 설정된 선택 확률 및 상기 산출된 보상에 기반하여 상기 선택된 과제의 보상 추정치를 설정하는 것과,

상기 보상 추정치를 사용하여 특정한 가중치를 상기 알고리즘의 다음 라운드에서의 상기 선택 확률의 설정을 위해 갱신하는 것 - 상기 특정한 가중치는 상기 선택된 과제가 상기 가중 분포에 따라 이용될 확률을 정의하는 데에 사용되고, 상기 특정한 가중치를 갱신하는 것은 상기 특정한 가중치에 지수 항을 곱하는 것을 포함하되, 상기 지수 항의 지수에 상기 보상 추정치가 사용됨 - 을 포함하는,

컴퓨팅 장치.

청구항 2

삭제

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인지 또는 오답인지에 따라 상이하게 상기 비교의 상기 결과로부터 산출되는,

컴퓨팅 장치.

청구항 5

제1항에 있어서,

상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 클수록 더 커지게 산출되거나, 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 오답인

경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 작을수록 더 작아지게 산출되거나, 양자 모두인,

컴퓨팅 장치.

청구항 6

강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 컴퓨팅 장치에 의해 수행되는 방법으로서,

학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 위한 알고리즘의 매 라운드에서 동작을 수행하는 단계를 포함하되, 상기 알고리즘은 적대적 밴딧(adversarial bandit) 설정에서의 상기 RL의 사용을 위한 Exp3 알고리즘(Exponential-weight algorithm for Exploration and Exploitation)이고, 상기 동작은,

상기 다수의 과제 각각의 선택 확률을 설정하는 것 - 상기 각각의 과제의 상기 선택 확률은 미선택된 과제의 탐색(exploration)을 가능하게 하는 균일 분포, 기선택된 과제의 이용(exploitation)을 가능하게 하는 가중 분포, 그리고 상기 탐색 및 상기 이용의 트레이드오프(tradeoff) 관계를 나타내는 파라미터에 기반하여 설정됨 - 과,

상기 설정된 선택 확률에 따라 상기 다수의 과제 중 하나를 추천되는 과제로서 선택하는 것과,

상기 선택된 과제의 요구되는 기량 수준 및 추정된 기량 수준을 비교하는 것과,

상기 비교의 결과로부터 보상을 산출하는 것과,

상기 산출된 보상에 기반하여 상기 추정된 기량 수준을 갱신하는 것과,

상기 선택된 과제의 상기 설정된 선택 확률 및 상기 산출된 보상에 기반하여 상기 선택된 과제의 보상 추정치를 설정하는 것과,

상기 보상 추정치를 사용하여 특정한 가중치를 상기 알고리즘의 다음 라운드에서의 상기 선택 확률의 설정을 위해 갱신하는 것 - 상기 특정한 가중치는 상기 선택된 과제가 상기 가중 분포에 따라 이용될 확률을 정의하는 데에 사용되고, 상기 특정한 가중치를 갱신하는 것은 상기 특정한 가중치에 지수 항을 곱하는 것을 포함하되, 상기 지수 항의 지수에 상기 보상 추정치가 사용됨 - 을 포함하는,

방법.

청구항 7

삭제

청구항 8

삭제

청구항 9

제6항에 있어서,

상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인지 또는 오답인지에 따라 상이하게 상기 비교의 상기 결과로부터 산출되는,

방법.

청구항 10

제6항에 있어서,

상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 클수록 더 커지게 산출되거나,

상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 오답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 작을수록 더 작아지게 산출되거나,

양자 모두인,

방법.

청구항 11

컴퓨터 프로세서에 의해 실행되는 경우 상기 컴퓨터 프로세서로 하여금 제6항, 제9항 및 제10항 중 어느 한 항에 기재된 방법을 수행하게 하는 컴퓨터 실행가능 명령어가 저장된 컴퓨터 판독가능 저장 매체.

발명의 설명

기 술 분 야

[0001] 본 개시는 강화 학습(Reinforcement Learning: RL) 기반의 교육용 상호작용 향상에 관한 것으로서, 더욱 구체적으로, RL을 사용하여 교육용 과제를 추천하기 위한 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 세계적으로 인구의 증가와 더불어 학교와 같은 교육 기관에서 많은 수(가령, 수백 명)의 학생을 한 명의 교사가 가르치는 일이 드물지 않다. 이에 양질의 교육이 큰 문제 중 하나가 되었다. 이러한 교육 시나리오는 개인맞춤형 교육(personalized teaching)을 훨씬 더 힘들게 하고, 결국 학생들은 그들의 역량 수준에 따른 개인맞춤형 과제(task)가 부족한 탓에 자신의 완전한 잠재력을 발휘하지 못한다. 그래서, 교사는, 제한된 시간과 학습 자원으로 인해 학생 개개인에게 개인별 수준에서 특별한 도움을 주기가 매우 어렵다. 개인맞춤형 교육이 학생의 학습 잠재력을 증진시키는 데에 유망한 접근법이라는 점은 상당한 양의 과학적 연구가 입증한다. 학생들의 수요를 수용하는 데에 사용될 수 있는 효과적인 자동 교육 시스템을 개발할 필요성이 높다. 스케줄링 문제(scheduling problem)를 위해 제안된 기존의 알고리즘이 확대되면서 그러한 시스템을 개발할 필요성이 현실화되었다.

[0003] 역량 기반의 지능형 시스템(intelligent system)의 개발과 함께, 학생의 지식으로부터 개인의 학업 속성을 연구하고 숙달 수준에 비례하여 추천을 하는 것이 가능하다. 역량 기반 추천기(competency-based recommender) 시스템의 최근 연구들은 학생의 학업 특성(성과, 경험, 성취도) 이력(history)을 사용하는 협업 필터링(collaborative filtering), 콘텐트 기반(content-based) 및 혼성(hybrid) 접근법들을 망라한다. 그러나, 새 학생의 이력 추적 없이 실시간 학습 환경에서 이들 기법을 활용하는 것은 콜드 스타트(cold-start) 문제 같은 몇몇 문제점을 제기한다. 적응적 학습(adaptive learning)을 위해 온톨로지 영역 표현(ontological domain representation)을 사용하여 이런 난점을 피하는 혼성 기법이 제시된 바 있다. 다만, 이 역시 학교 교육, 학습 및 숙달 수준과 같은 학생 이력 파라미터에 기반한다.

[0004] 따라서, 이러한 난제를 극복할 필요가 있다.

발명의 내용

해결하려는 과제

[0005] 본 문서는 강화 학습(Reinforcement Learning: RL) 기반의 교육용 과제의 추천을 위한 장치 및 방법을 개시한다.

과제의 해결 수단

[0006] 적어도 하나의 실시예에 따르면, 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 컴퓨팅 장치가 제공되는데, 상기 컴퓨팅 장치는 프로세서와, 프로세서 실행가능 명령어가 저장된 메모리를 포함하되, 상기 프로세서 실행가능 명령어는 상기 프로세서에 의해 실행되는 경우 상기 프로세서로 하여금, 학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 위한 알고리즘의 매 라운드에서, 다음을 포함하는 동작을 수행하게 한다: 상기 다수의 과제 각각의 선택 확률을 설정하는 것; 상기 설정된 선택 확률에 따라 상기 다수의 과제 중 하나를 추천되는 과제로서 선택하는 것; 상기 선택된 과제의 요구되는 기량 수준 및 추정된 기량 수준을 비교하는 것; 상기 비교의 결과로부터 보상(reward)을 산출하는 것; 상기 산출된 보상에 기반하여 상기 추정된 기량 수준을 갱신하는 것; 상기 선택된 과제의 상기 설정된 선택 확률 및 상기 산출된 보상에 기반하여 상기 선택된 과제의 보상 추정치를 설정하는 것; 및 상기 보상 추정치를 사용하여, 상기 선택된 과제가 다음에 선택될 확률의 감소를 방지하는 것.

[0007] 상기 알고리즘은 적대적 밴딧(adversarial bandit) 설정에서의 상기 RL의 사용을 위한 Exp3 알고리즘일 수

있다.

[0008] 상기 보상은 실수 값으로 산출될 수 있다.

[0009] 상기 비교는 상기 학생의 주어진 역량 수준에서 상기 요구되는 기량 수준 및 상기 추정된 기량 수준을 비교하는 것일 수 있다.

[0010] 상기 다수의 과제 중 상기 선택된 과제가 아닌 과제의 보상 추정치는 0으로 설정될 수 있다.

[0011] 상기 각각의 과제의 상기 선택 확률은 미선택된(not-yet-selected) 과제의 탐색(exploration)을 가능하게 하는 균일 분포, 기선택된(already-selected) 과제의 이용(exploitation)을 가능하게 하는 가중 분포, 그리고 상기 탐색 및 상기 이용의 트레이드오프(tradeoff) 관계를 나타내는 파라미터에 기반하여 설정될 수 있다.

[0012] 상기 방지는 상기 가중 분포에 따른 상기 선택된 과제의 이용 확률을 정의하는 데에 사용되는 특정한 가중치를 지수적으로(exponentially) 갱신하는 것을 포함할 수 있다.

[0013] 상기 지수적 갱신은 상기 특정한 가중치에 지수 항을 곱하는 것(상기 지수 항의 지수에 상기 보상 추정치가 사용됨)을 포함할 수 있다.

[0014] 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인지 또는 오답인지에 따라 상이하게 상기 비교의 상기 결과로부터 산출될 수 있다.

[0015] 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 클수록 더 커지게 산출되거나, 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 오답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 작을수록 더 작아지게 산출되거나, 양자 모두일 수 있다.

[0016] 적어도 하나의 실시예에 따르면, 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 방법이 제공되는데, 상기 방법은, 학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 위한 알고리즘의 매 라운드에서, 다음을 포함하는 동작을 수행하는 단계를 포함한다: 상기 다수의 과제 각각의 선택 확률을 설정하는 것; 상기 설정된 선택 확률에 따라 상기 다수의 과제 중 하나를 추천되는 과제로서 선택하는 것; 상기 학생의 주어진 역량 수준에서 상기 선택된 과제의 요구되는 기량 수준 및 추정된 기량 수준을 비교하는 것; 상기 비교의 결과로부터 보상을 산출하는 것; 상기 산출된 보상에 기반하여 상기 추정된 기량 수준을 갱신하는 것; 상기 선택된 과제의 상기 설정된 선택 확률 및 상기 산출된 보상에 기반하여 상기 선택된 과제의 보상 추정치를 설정하는 것; 및 상기 보상 추정치를 사용하여, 상기 선택된 과제가 다음에 선택될 확률의 감소를 방지하는 것.

[0017] 상기 알고리즘은 적대적 밴딧 설정에서의 상기 RL의 사용을 위한 Exp3 알고리즘일 수 있다.

[0018] 상기 보상은 실수 값으로 산출될 수 있다.

[0019] 상기 비교는 상기 학생의 주어진 역량 수준에서 상기 요구되는 기량 수준 및 상기 추정된 기량 수준을 비교하는 것일 수 있다.

[0020] 상기 다수의 과제 중 상기 선택된 과제가 아닌 과제의 보상 추정치는 0으로 설정될 수 있다.

[0021] 상기 각각의 과제의 상기 선택 확률은 미선택된 과제의 탐색을 가능하게 하는 균일 분포, 기선택된 과제의 이용을 가능하게 하는 가중 분포, 그리고 상기 탐색 및 상기 이용의 트레이드오프 관계를 나타내는 파라미터에 기반하여 설정될 수 있다.

[0022] 상기 방지는 상기 가중 분포에 따른 상기 선택된 과제의 이용 확률을 정의하는 데에 사용되는 특정한 가중치를 지수적으로 갱신하는 것을 포함할 수 있다.

[0023] 상기 지수적 갱신은 상기 특정한 가중치에 지수 항을 곱하는 것(상기 지수 항의 지수에 상기 보상 추정치가 사용됨)을 포함할 수 있다.

[0024] 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인지 또는 오답인지에 따라 상이하게 상기 비교의 상기 결과로부터 산출될 수 있다.

[0025] 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 정답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 클수록 더 커지게 산출되거나, 상기 보상은 상기 선택된 과제에 대한 상기 학생의 답이 오답인 경우, 상기 요구되는 기량 수준이 상기 추정된 기량 수준보다 작을수록 더 작아지게 산출되거나, 양자 모두일

수 있다.

[0026] 적어도 하나의 실시예에 따르면, 컴퓨터 실행가능 명령어가 저장된 컴퓨터 판독가능 저장 매체가 제공되는데, 상기 컴퓨터 실행가능 명령어는 컴퓨터 프로세서에 의해 수행되는 경우 상기 컴퓨터 프로세서로 하여금 상기 방법을 수행하게 할 수 있다.

[0027] 전술된 개요는 상세한 설명에서 추가로 후술되는 몇몇 양상을 단순화된 형태로 소개하기 위해 제공된다. 이 개요는 청구된 주제(subject matter)의 중요 특징 또는 필수적 특징을 식별하도록 의도되지 않고, 청구된 주제의 범위를 정하는 데 사용되도록 의도되지도 않는다. 나아가, 청구된 주제는 본 명세서에서 논의되는 임의의 또는 모든 이점을 제공하는 구현에 한정되지 않는다.

발명의 효과

[0028] 본 발명의 실시예에 따르면, 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 데에 적합한 개인맞춤형 기량 기반 추천기가 제공된다.

[0029] 본 발명의 실시예에 따른 추천기는 추천 전략(policy)이 비추계적(non-stochastic) 문제로서 규정되는 적대적 밴딧 설정(adversarial bandit setting)에서 RL을 사용하며, 그러한 설정 및 사용을 위한 알고리즘으로서 탐색 및 이용을 위한 지수 가중 알고리즘(Exponential-weight algorithm for Exploration and Exploitation)(약칭으로, Exp3 알고리즘 또는 Exp3)을 활용하는바, 온라인 방식으로 실시간으로 동작할 수 있다.

도면의 간단한 설명

[0030] 도 1은 본 발명의 실시예에 따른 개인맞춤형 기량 기반 추천기(personalized skill-based recommender)의 개략적인 블록도이다.

도 2는 본 발명의 실시예에 따른 개인맞춤형 기량 기반 추천기를 위한 프레임워크를 개략적으로 도시한다.

도 3은 본 발명의 실시예에 따라 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 프로세스를 의사코드(pseudocode)의 형태로 제시한다.

도 4는 본 발명의 실시예에 따라 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하기 위한 컴퓨팅 장치의 개략적인 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0031] 이하에서는, 첨부된 도면을 참조하여 본 발명의 실시예를 상세히 설명한다. 본 발명은 여러 가지 실시예를 가질 수 있고, 몇몇 실시예가 본 명세서에 개시된다. 그러나, 이는 본 발명에 대한 한정이 아니라 예시로서 제공되며, 본 발명의 사상 및 범주에 속하는 모든 변환, 균등물 내지 대체물을 망라하는 것으로 이해되어야 한다. 개시된 실시예에 따른 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 다음의 상세한 설명에서 특정한 세부사항이 제공되는데, 몇몇 실시예는 이들 세부사항 중 일부 또는 전부가 없더라도 실시될 수 있다. 또한, 본 발명의 다양한 양상을 불필요하게 모호하게 하지 않도록 공지 기술의 구체적인 설명은 생략될 수 있다.

[0032] 후술되는 용어는 단지 특정 실시예를 설명하기 위해 사용된 것으로, 한정적 의미로 고려되고자 의도된 것이 아니다. 단수 형태의 표현은 명확하게 달리 사용되지 않는 한, 복수 형태의 의미를 포함한다. 또한, 이 문서에서, "포함하다" 또는 "가지다"와 같은 용어는 어떤 특징, 숫자, 단계, 동작, 구성요소, 정보 또는 이들의 조합이 존재함을 나타내려는 것이며, 하나 또는 그 이상의 다른 특징, 숫자, 단계, 동작, 구성요소, 정보 또는 이들의 조합의 존재 또는 가능성을 배제하지 않는 것으로 이해되어야 한다.

[0033] 개관

[0034] 본 발명의 실시예에 따르면, 개인맞춤형 기량 기반 추천기를 위한 프레임워크가 제안된다. 이는 학생의 속성에 따라 과제를 추천하는 데에 온라인 추정 접근법(online estimation approach)을 사용하여 실시간으로 학생의 역량 수준을 측정한다. 이 프레임워크는 학생 기량 수준을 제안된 과제에 대한 학생의 응답에 기반하여 추정하는데, 어떤 다른 정보나 학생의 학업 이력 데이터도 사용하지 않는다. 학생의 지식 속성을 추정하기 위해, 그러한 추천기는 각각의 개별 학생에 대해 개념 학습 기량 프로파일(profile)을 만든다. 따라서, 이 시스템의 사용은 학습자가 자신의 페이스(pace)로 학습할 수 있게 한다. 또한, 본 발명의 실시예에서, 최적의 과제를 학생에게

추천하는 데에 낮은 복잡도의 알고리즘인, 탐색 및 이용을 위한 지수 가중 알고리즘(Exponential-weight algorithm for Exploration and Exploitation)(약칭으로, Exp3 알고리즘 또는 Exp3)이 활용된다. 이는 몇몇 기존의 시스템에서 높은 계산량의 전문가 알고리즘(expert algorithm)인, 전문가 조언을 사용한 탐색 및 이용을 위한 지수 가중 알고리즘(Exponential-weighted algorithm for Exploration and Exploitation using Expert advice)(약칭으로, Exp4 알고리즘 또는 Exp4) 및 퍼지 언어 접근법(Fuzzy linguistic approach)이 사용되는 것과 대조적이다.

[0035] 이하에서 논의되는 바와 같이, 실시예에 따른 추천기와 관련하여, (i) 실시간으로 학생의 기존 학업 기량을 파악하기 위해 학생의 행동이 학생 프로파일로서 모델링되고, (ii) 과제 및 학생 기량 관계가 M-매트릭스(M-matrix)에 의해 특징지어지는데, 학생 프로파일 정보를 업데이트하고 추정하는 데에 이것이 활용되며, (iii) 지수 가중치 알고리즘 Exp3이 Exp4에 비해 많은 수의 액션을 다룰 수 있고 실행 시간이 빨라 최적의 과제를 추천하는 데에 활용된다.

[0036] 시스템 모델

[0037] 이제, 도 1 및 도 2를 참조하여, 본 발명의 실시예에 따른 개인맞춤형 기량 기반 추천기를 위한 프레임워크를 살펴본다.

[0038] 도 1에 도시된 바와 같이, 예시적인 추천기(100)는 추천 전략을 위한 컴포넌트로서 학생 프로파일 모듈(120), 학습 측정 모듈(140) 및 과제 선택 전략 모듈(160)을 포함한다.

[0039] 실시예에서, 학생 프로파일 모듈(120)은 학생 프로파일을 제공하도록 구성된다. 예를 들어, 학생 프로파일 모듈(120)은 학생의 실시간 학습 결과(예컨대, 역량 및 학습 수준)를 관리하고, 따라서 학생의 특성을, 가령, Q-매트릭스 표현(Q-Matrix representation)을 사용하여, M-매트릭스 내에 저장하는 데에 사용될 수 있다. Q-매트릭스 표현에서, 매트릭스 $Q_{(M\times N)}$는 M×N 차원 행렬인데, 각각의 행은 질문을 나타내고 각각의 열은 각 질문에 대해 요구되는 역량(또는 숙달)에 해당한다. Q-매트릭스는 질문과 그것의 요구되는 개념 간의 관계를 표현함으로써 학생의 성과를 이해하는 데에 사용된다. 이 맥락에서, 추천기(100)를 위한 프레임워크에서, Q-매트릭스와 유사하지만, 학생의 기량(또는 진도) 수준을 수치로 표현하는 매트릭스 $M_{(A\times C)}$가 생성된다. 이 매트릭스는, 도 2에 도시된 바와 같이, 과제 세트 A 및 역량 세트 C 간의 관계를 나타낸다. 세트 $A = (a_1, a_2, a_3, \dots, a_m)$ 로부터의 각각의 과제는 역량 세트 $C = (c_1^a, c_2^a, c_3^a, \dots, c_n^a)$ 와의 관계를 갖는다. 과제 및 역량의 쌍 각각에 대응하는 수치는 학생의 추정된 기량 수준(Estimated Skill Level: ESL)을 나타낸다. 따라서, 학생에 의해 획득된 역량 수준 c_j에 속하는 각각의 과제 a_i의 기량 수준에 의해 학생 프로파일이 표현된다. 예를 들어, 도 2에 도시된 학생 프로파일에서, 셀 (a_1, c_2)는 학생이 역량 수준 c_2에 비견되는 난이도를 가진 과제 a_1에 대해 20 퍼센트를 성취하였음(20%의 진도)을 보여준다. 실시예에서, 프로파일 모듈(120)은 과제에 대응하는 학생의 답(맞음 또는 틀림)을 활용하여 학생의 학습 진도를 저장할 수 있다.

[0040] 실시예에서, 학습 측정 모듈(140)은 숙달 수준과 연관된 특정 과제를 획득하는 데에 요구되는 값의 세트를 정의한다. 구체적으로, 학습 측정 모듈(140)은 각각의 과제를 풀기 위한 요구되는 기량 레벨(Required Skill Level: RSL)(이는 도 2에 보여진 바와 같이 (a_i, c_j)로 인덱싱됨(indexed))을 제공하도록 구성된다. 예를 들어, 이 측정 모듈(140)은 학생의 학습 기량을 이해하고 추정하는 데에, 또 학습 자료(learning material)의 특성을 밝히는 데에도, 사용된다. 예로서, 수학식 1에 주어진 바와 같이, 자동으로 학습 자료 내의 각각의 과제에 대해 독립적인 RSL이 정의될 수 있는데, 이 파라미터 값 $RSL_{(a_i, c_j)}$ 는 학생의 현재의 ESL을 추정하는 데에 사용되는 제안된 과제에 대한 요구되는 기량 수준을 보여준다.

수학식 1

[0041] $$RSL_{(a_i, c_j)} = \left(\frac{T_j \times (T_i + 1)}{len(A)}\right) - \left(\frac{T_j \times T_i}{len(A)}\right)$$

[0042] 여기서 T_j는 (매트릭스의 차원에서의 열 번호로서) 선택된 과제의 요구되는 개념 이해를 나타내고, T_i는 (매트릭스의 차원에서의 행 번호로서) 제안된 과제를 나타내며, len(A)는 학습 자료 내의 과제의 총 개수이다. 추천 전략으로 학생에게 과제의 추천이 행해지는 경우에, 학습 측정 모듈(140)은 추천된 과제에 대한 응답으로서 학생으로부터 맞는 또는 틀린 답을 수신한다. 학습 측정 모듈(140)은 제안된 과제의 대응하는 파라미터 값 $RSL_{(a_i, c_j)}$ 을 통해서 학생의 성과를 추정하도록 돕는다. 학습 측정 모듈(140)은 과제마다 학생을 평가하고 개별 학생 각각에게 학습 자료 내에서의 그의 진도에 따라 보상(reward)을 할당한다. 예로서, 학습 측정 모듈(140)의 보상 함수는 수학식 2 및 수학식 3과 같이 주어진다.

수학식 2

[0043]
$$R_{(a_i, c_j)} = max\left\{\left(RSL_{(a_i, c_j)} - ESL_{(a_i, c_j)}\right), 0\right\}$$

수학식 3

[0044]
$$R_{(a_i, c_j)} = min\left\{\left(RSL_{(a_i, c_j)} - ESL_{(a_i, c_j)}\right), 0\right\}$$

[0045] 여기서 $R_{(a_i, c_j)}$ 는 역량 c_j에 있어서 과제 a_i가 추천되는 경우에 학생의 진도에 알맞은 보상이다. 그래서, 수학식 2 및 수학식 3은 제안된 과제에 대한 학생의 답이 맞든 틀리든 두 경우 다 포섭한다. 즉, 수학식 2는 정답에 대한 보상을 측정하는 데에 사용되고, 수학식 3은 오답의 경우에 사용된다. 그러면, 수학식 1, 수학식 2 및 수학식 3에서의 서술 파라미터 값으로써, 각 학생에 대해 학습 속성을 갱신하고 추정할 수가 있다. 따라서, 학습 시간 동안 임의의 추천된 과제에 대한 학생의 ESL은 다음과 같이 갱신된다.

수학식 4

[0046]
$$ESL_{(a_i, c_j)} \leftarrow ESL_{(a_i, c_j)} + \eta R_{(a_i, c_j)}$$

[0047] 여기서 η은 학생의 학습 속도를 나타내는 가중치인데, 학습 자료로부터의 학생의 기량 향상에 달려 있다. 이 값은 보상 선호(preference)를 판정하는 데에 사용된다. 적합한 값에 대한 최선의 설정인즉 학습 프로세스에서 더 나은 결과를 위해 과제가 추천된다는 뜻이다. 수학식 4는 학생의 성과가, 학습 숙달을 증가시키는 데에서는 물론 그 동기를 부여하는 데에도 사용되는 추천된 과제의 보상을 받는 것에 달려 있음을 보여준다. 따라서, 수학식 4에 따르면, 학생이 그의 이전의 추정된 기량 수준 대신에 어려운 과제에 정확한 답을 하는 경우에만 그는 진도를 나아갈 수 있다.

[0048] 실시예에서, 과제 선택 전략 모듈(160)은 과제 추천 전략을 수립하고 실행하도록 구성된다. 예를 들어, 과제 선택 모듈(160)은 학생 프로파일 속성에 따른 ESL에 기반하여 과제를 추천하기 위해 최적의 추천 전략을 선택하는 데에 사용될 수 있다. 그러나, 과제 풀 A로부터의 최적의 과제의 선택은 차원 문제(dimensionality problem)로 인해 매우 어려울 수 있다. 몇몇 예시적인 구현에서, 추천기(100)는 적대적 밴딧(adversarial bandit) 문제로 규정된 프레임워크를 갖고서 최적화 및 차원 축소(dimensionality reduction)를 위해 Exp3 알고리즘을 활용함으로써, 그러한 난제를 극복한다.

[0049] 적대적 밴딧은 멀티 암드 밴딧(Multi-Armed Bandit: MAB)의 변형이다. MAB의 개념은 시간에 걸쳐 불확실성 하에서 결정을 하는 알고리즘을 위해 사용된다. MAB는 주로 추계적 스케줄링 및 의사결정 문제에서 연구된다. 밴딧 문제에서, 플레이어(player)는 최대의 보상을 얻기 위해 순차적으로 어느 액션을 취할지(가령, 어느 암을 플레

이할지)를 결정한다. MAB 모델은 에이전트(agent)가 K개의 암 중 임의의 것을 매 시간 스텝(time step)(이는 라운드(round) 또는 시도(trial)로도 지칭됨) t에서 선택하여 플레이하고 그 암은 각자의 확률 P∈[0,1]로 어떤 보상(가령, 실수 x)을 준다는 것이다. MAB는 효율적인 온라인 추정 방식으로서, 이에 따르면, 최선의 결정을 하기 위해 최적의 액션을 찾고자 새로운 데이터 점이 주어지면 전략은 온라인 방식으로 알고리즘을 훈련시킬 수 있다. 적대적 또는 비통계적 밴딧은 그러한 MAB 프레임워크에, 사전정의된 고정된 확률 분포의 보상 대신, 보상 생성 프로세스에 대한 비통계적(non-statistical) 가정이 들어간 것이다. 이 적대형 모델에서, 보상은 제한된 범위로부터 무제한적인 컴퓨팅 파워로 플레이에 따라 적대자(adversary)에 의해 선택된다고 볼 수 있다.

[0050] 예시적인 추천기(100)는 온라인 방식으로 학생을 훈련시키기 위해 비추계적 밴딧 문제로 규정된 프레임워크에서 동작하도록 구성될 수 있다. 적대적 밴딧 프레임워크의 비추계적 환경에서, 학생을 훈련시키기 위해 여러 과제 중 택일하는 것은 랜덤 분포를 갖는 것으로 모델링될 수 있다. 적대적 밴딧 문제에 있어서, Exp3는 매우 단순하면서도 강력한 알고리즘이다. 이 알고리즘은 MAB 문제에서 탐색 및 이용 트레이드오프(exploration and exploitation tradeoff)를 푸는 데에 사용되는 헤지(Hedge) 알고리즘에 기반한다. Exp3 알고리즘은 수학식 5를 참조하여 다음과 같이 설명될 수 있다.

수학식 5

[0051]
$$E[\hat{x}_i(t) \mid i_1, \dots i_{t-1}] = p_i(t) \cdot \frac{x_i(t)}{p_i(t)} + \big(1 - p_i(t)\big) \cdot 0 = x_i(t)$$

[0052] Exp3는 매 라운드 t에서 K개의 가능한 액션(각각의 액션은 i=1, 2, ..., K에 의해 표기됨) 중에서 액션 i_t를 확률 분포 $p_1(t)$, $p_2(t)$, ..., $p_K(t)$에 따라 뽑는다. 위 식에서, $E[\hat{x}_i(t) \mid i_1, \dots i_{t-1}]$ 는 라운드 t에서 취해진 액션에 대응하는 이른바 추정된 보상(estimated reward)(이는 보상 추정치(reward estimate)로도 지칭될 수 있음) $\hat{x}_i(t)$ 의 조건부 기대치(expectation)(이전의 t-1 번의 라운드에서의 선택 i_1, i_2, ..., i_{t-1}이 주어진 경우에 t번째 라운드에서의 선택에 대한 보상의 기대치)이고, $p_i(t) \cdot \frac{x_i(t)}{p_i(t)} + \big(1 - p_i(t)\big) \cdot 0$ 는 보상 추정 프로세스를 시사하는데, 이에 따르면, 액션이 플레이된 관측되는 암에 대해서는 실제의 보상 $x_i(t)$을 확률 $p_i(t)$로 나눈 값이 보상으로 추정되고, 관측되지 않는 암에 대해서는 0의 보상이 할당되는 경우, 전술된 기대치는 관측된 보상과 같게 됨을 알 수 있다. 확률 $p_i(t)$는 수학식 6에서와 같이 각각의 암에 대한 균일 분포 및 가중 분포의 혼합을 나타내도록 정의될 수 있다.

수학식 6

[0053]
$$p_i(t) = (1 - \gamma) \frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K}$$

[0054] 여기서 γ ∈[0,1]는 액션 i를 위한 탐색 파라미터(이는 평등주의 인자(egalitarianism factor)로도 지칭됨)로서 초기에 선택된 값이고, w_i는 액션 i의 지수 가중치(exponential weight)(이는, 가령, 각각의 액션에 대해 균일하게 초기화됨)이다. 가중치 w_i는 다음과 같이 보상 추정치를 사용하여 갱신된다.

수학식 7

$$w_i(t+1) = w_i(t) \cdot e^{\left(\frac{\gamma}{K}\hat{x}_i(t)\right)},$$

$$\hat{x}_i(t) = \begin{cases} x_i(t)/p_i(t), & i = i_t \\ 0, & i \neq i_t \end{cases}$$

[0055]

[0056] 이와 같이, Exp3 알고리즘은 좋은 암을 당길 균일한 확률을 계산하는 것과, 각 라운드에서 당겨진 암에 대한 가중치 리스트를 유지하고 갱신하는 것을 수반한다. 실시예에서, 학생의 주어진 역량 수준에 있어서, Exp3는 여러 가능한 액션 중에서 각 액션의 확률 분포에 기반하여 뽑힌 액션 i_t 에 따라 세트 A로부터 과제를 선택하고, 이에 대한 보상 $x_{i_t}(t)$ (가령, $x_{i_t}(t)$ ∈[0,1])을 관측한다. 또한, 실시예에서, Exp3 알고리즘은 선택된 액션에 대한 추정된 보상을, 수학식 5에 의해 시사되는 바와 같이, 실제 관측된 보상을 그 액션의 확률로 나눈 값으로 설정하며, 다른 액션에 대한 추정된 보상은 0으로 설정한다. 예를 들어, 예시적인 추천기(100)에서, Exp3를 위한 추정된 보상은 다음과 같이 $\hat{R}_i(t)$ 로서 정의될 수 있다.

수학식 8

$$\hat{R}_i(t) = \begin{cases} R_i(t)/p_i(t), & i = i_t \\ 0, & i \neq i_t \end{cases}$$

[0057]

[0058] 여기서 $\hat{R}_i(t)$ 는 학생의 역량 수준 c에 연관된 과제 a에 대한 학생의 답에 따라 수학식 2 또는 수학식 3과 같이 주어지는 보상 $R_{(a,c)}$ 를 나타낸다.

[0059] 교육용 과제의 추천을 위한 예시적인 프로세스

[0060] 도 3은 본 발명의 실시예에 따라 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하는 프로세스(300)를 의사코드(pseudocode)의 형태로 제시한다. 도 3의 예시적인 프로세스(300)는 단지 예로서 제공될 뿐이며, 몇몇 다른 실시예에서는 도 3의 프로세스(300)의 동작 중 일부가 생략되고/되거나 다른 동작이 추가될 수 있음에 유의하여야 한다. 또한, 몇몇 예시적인 구현에서, RL은 적대적 밴딧 설정에서 사용될 수 있고, 그러한 사용을 위한 알고리즘은 Exp3 알고리즘일 수 있다.

[0061] 예시적인 프로세스(300)는 일련의 초기화 동작(310)을 수행함으로써 시작된다. 초기화 동작(310)은, 예컨대, 탐색 파라미터 γ 를 설정하는 것, 가중치 w_i를 모든 액션 i에 대해 균일한 초기 값으로 설정하는 것, 학생의 각각의 가능한 역량 수준에서 각각의 과제의 ESL을 초기화하는 것 등을 포함할 수 있다.

[0062] 이어서, 예시적인 프로세스(300)는 학생의 교육을 위한 다수의 과제로부터의 반복되는 택일을 수행한다. 그러한 과제 선택을 위한 알고리즘(가령, Exp3 알고리즘)의 매 라운드에서, 가령, 금번 라운드에서의 과제 선택, 이로부터의 보상의 관측, 그리고 다음 라운드에서의 과제 선택을 위한 소정의 갱신을 포함하여, 다음과 같은 동작이 수행될 수 있다.

[0063] 동작(320)에서, 다수의 과제 각각의 선택 확률(probability of selection)이 설정된다. 실시예에서, 그러한 선택 확률은 다수의 과제 중의 미선택된(not-yet-selected) 과제의 탐색을 가능하게 하는 균일 분포, 다수의 과제 중의 기선택된(already-selected) 과제의 이용을 가능하게 하는 가중 분포, 그리고 탐색 및 이용의 트레이드오

프 관계를 나타내는 파라미터(γ)에 기반하여 설정될 수 있다. 예를 들어, 수학식 6에 주어진 바와 같이, 각 과제의 선택 확률 p_i은 각각의 과제가 균일 분포에 따라 선택될 확률(이는 해당 과제가 탐색될 확률(또는 해당 과제의 "탐색 확률"(probability of exploration))로 지칭될 수 있고, 수학식 6에서는 다수의 과제의 총 개수의 역수, 즉 $\frac{1}{K}$ 임) 및 해당 과제가 가중 분포에 따라 선택될 확률(이는 해당 과제가 이용될 확률(또는 해당 과제의 "이용 확률"(probability of exploitation))로 지칭될 수 있고, 수학식 6에서는 모든 과제의 지수 가중치의 합에 대한 해당 과제의 지수 가중치의 비율, 즉 $\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)}$ 임)의 가중 조합(단, 이 가중 합에서의 이 두 확률의 가중치는 각각 γ 및 $1-\gamma$ 임)일 수 있다.

[0064] 동작(330)에서, 설정된 선택 확률에 따라 다수의 과제 중 하나가 선택된다. 실시예에서, 그러한 선택은 랜덤할 수 있다. 이와 같이 선택된 과제는 추천되는 과제로서 학생에게 제시될 수 있다.

[0065] 이제, 보상의 수신과 관측이 이어진다. 예를 들어, 보상은 (가령, [0, 1]의 범위 내의) 실수 값일 수 있다. 동작(340)에서, 예컨대, 학생의 주어진 역량 수준에 대해, 학생에게 추천된 과제의 RSL 및 ESL이 비교되며, 그러한 비교의 결과로부터 보상이 산출된다. 실시예에서, 그러한 보상은 학생에게 추천된 과제에 대한 학생의 답이 정답인지 또는 오답인지에 따라 상이하게 산출될 수 있다. 추가적으로 또는 대안적으로, 그러한 보상은 학생에게 추천된 과제에 대한 학생의 답이 정답인 경우 RSL이 ESL보다 클수록 더 커지게 산출(가령, 수학식 2 참조)될 수 있고/거나, 학생에게 추천된 과제에 대한 학생의 답이 오답인 경우 RSL이 ESL보다 작을수록 더 작아지게 산출(가령, 수학식 3 참조)될 수 있다.

[0066] 동작(350)에서, 산출된 보상에 기반하여 ESL이 갱신된다.

[0067] 이어서, 동작(360)에서, 추천된 과제의 보상 추정치 $\hat{R}_j(t)$ 가 설정되고, 그 과제의 선택 확률의 갱신(즉, 다음 라운드에서의 설정)을 위해 지수 가중치 $w_i(t)$ 가 갱신된다. 실시예에서, 이 갱신은 추천된 과제가 다음에 이용될 확률(그리고, 결국, 그 과제가 다음에 선택될 확률)을 증가시키거나 적어도 유지하기 위해, 즉, 그러한 확률의 감소를 방지하기 위해, (가령, 이용 확률의 정의에 사용되는 가중치 $w_i(t)$ 를 지수적으로 갱신하는 데에) 보상 추정치를 사용하는 것을 수반할 수 있다(가령, 수학식 7에 주어진 바와 같이, 갱신된 가중치 $w_i(t+1)$ 를 획득하기 위해 현재의 가중치 $w_i(t)$ 에 곱해지는 지수 항(exponential term)의 지수에 보상 추정치가 사용될 수 있음).

[0068] 한편, 동작(360)에서, 다수의 과제 중 나머지 과제의 보상 추정치는 0으로 설정된다.

[0069] 예시적인 컴퓨팅 장치

[0070] 도 4는 본 발명의 실시예에 따라 강화 학습(Reinforcement Learning: RL)을 사용하여 교육용 과제를 추천하기 위한 컴퓨팅 장치(400)의 개략적인 블록도이다. 도 4를 참조하면, 예시적인 컴퓨팅 장치(400)는 프로세서(processor)(410) 및 메모리(memory)(420)를 포함하는 것으로 예시된다.

[0071] 실시예에서, 프로세서(410)는 RL을 사용하여 교육용 과제를 추천하기 위한 동작(가령, 프로세스(300)와 관련하여 기술된 동작 중 전부 또는 일부)을 수행하도록 구성된다. 예를 들어, 프로세서(410)는 중앙 처리 유닛(Central Processing Unit: CPU), 디지털 신호 프로세서(Digital Signal Processor: DSP), 그래픽 처리 유닛(Graphics Processing Unit: GPU), 프로세서 코어(processor core), 마이크로프로세서(microprocessor), 마이크로제어기(microcontroller), 필드 프로그램가능 게이트 어레이(Field-Programmable Gate Array: FPGA), 애플리케이션 특정 집적 회로(Application Specific Integrated Circuit: ASIC), 다른 하드웨어 및 로직 회로, 또는 이의 임의의 적합한 조합을 포함할 수 있다.

[0072] 실시예에서, 메모리(420)는 프로세서(410)에 의해 실행되는 경우 프로세서(410)로 하여금 본 발명의 실시예에 따라 몇몇 동작(가령, 전술된 예시적인 프로세스(300)와 관련하여 기술된 동작 중 적어도 일부)을 수행하게 하

는 프로세서 실행가능(processor executable) 명령어 및/또는 그러한 동작에서 사용되는 정보, 데이터, 변수, 상수, 데이터 구조, 기타 등등(가령, 전술된 바와 같이 M-매트릭스의 형태로 된 RSL, ESL, 또는 다른 정보)을 저장하도록 구성된다. 예를 들어, 컴퓨터 판독가능 저장 매체는 판독 전용 메모리(Read-Only Memory: ROM), 랜덤 액세스 메모리(Random-Access Memory: RAM), 휘발성(volatile) 메모리, 비휘발성(non-volatile) 메모리, 착탈가능(removable) 메모리, 비착탈가능(non-removable) 메모리, 플래시(flash) 메모리, 다른 저장 디바이스 및 저장 매체, 또는 이의 임의의 적합한 조합을 포함할 수 있다.

[0073] 예시적인 실시예는 본 문서에 기술된 동작, 기법, 프로세스, 또는 이의 어떤 양상이나 부분이 체현된 컴퓨터 프로그램을 포함하는 컴퓨터 판독가능 저장 매체로서 구현될 수 있다. 이러한 컴퓨터 판독 가능 저장 매체에는 프로그램 명령어, 로컬 데이터 파일, 로컬 데이터 구조 등등이 단독으로 또는 조합되어 포함될 수 있다. 개시된 동작, 기법, 프로세스, 또는 이의 어떤 양상이나 부분을 구현하거나 이용할 수 있는 프로그램은 컴퓨터에 의해 실행될 수 있는 어떤 유형의 (가령, 컴파일형(compiled) 또는 해석형(interpreted)) 프로그래밍 언어, 예컨대, 어셈블리(assembly), 기계어(machine language), 프로시저형(procedural) 언어, 객체지향(object-oriented) 언어 등등으로 구현될 수 있고, 하드웨어 구현과 조합될 수 있다. 용어 "컴퓨터 판독가능 저장 매체"는, 컴퓨팅 장치에 의한 실행을 위한 명령어(실행 시에 컴퓨팅 장치로 하여금 개시된 기법을 수행하게 함)를 저장할 수 있고, 그러한 명령어에 의해 사용되거나 이와 연관된 데이터 구조를 저장할 수 있는 임의의 매체를 포함할 수 있다. 컴퓨터 판독가능 저장 매체의 예는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM, DVD와 같은 광 기록 매체, 플롭티컬 디스크와 같은 자기-광 매체, 그리고 ROM, RAM, 플래시 메모리, 솔리드 스테이트(solid-state) 메모리와 같은 메모리 디바이스를 포함하되, 이에 한정되지 않는다.

[0074] 학생의 학습 잠재력이 시간에 따라 변하는 많은 교육 시나리오에서, 기존의 역량 기반 추천 시스템에 비해, 본 발명의 실시예에 따른 개인맞춤형 기량 기반 추천기는, 예컨대, 다음을 포함하여, 여러 주목할 만한 특징을 갖는다.

[0075] 일반적으로, 역량 기반 추천기 시스템의 기본적인 컴포넌트는 학생 역량을 모델링하는 방법과 더불어 기술된다. 본 발명의 실시예에 따른 추천기를 위한 모델에서는, 학생 프로파일로서 학생의 획득된 기량 수준을 나타내는 데에 [0, 1] 사이의 연속적인 숫자가 활용된다(가령, 0.3은 30 퍼센트 획득됨을 의미하고 0은 전혀 획득된 것이 없음을 의미한다). 이에 반해, 퍼지 언어적 접근법과 같은 몇몇 다른 접근법은 학생이 성취한 바를 각각의 역량으로 표현하는 데에 사용되는 학생 모델을 생성하지만, 그러한 표현은 학생 역량을 수치로 표현하는 것만으로는 불충분하다.

[0076] 또한, 본 발명의 실시예에 따른 추천기는 추천 전략을 비추계적 문제로서 규정하며, 고정된 학습 절차를 사용하지 않는다. 이는 학습 프로세스에서의 고정된 절차를 가정함으로써 추천 전략을 추계적(stochastic) 문제로 규정하는 접근법, 예컨대, 적응적 학습 시스템을 위해 마르코프 결정 프로세스(Markov decision process)를 사용하는 것과 같은, 학습자의 심리측정적 평가에 기반한 수학적 프레임워크에서, c-μ 규칙 및 기틴스(Gittins) 지수와 같은 갖가지 최적 전략을 사용하여 각각의 기량 수준과 연관된 학습자에게 학습 자료를 추천하는 접근법과는 대조적이다.

[0077] 나아가, 본 발명의 실시예에 따른 추천기는 개인맞춤화를 위한 각각의 액션(action)에 대응하는 과제의 최적화를 위해 전문가 조언에 의존하지 않는 Exp3 알고리즘을 활용한다는 점에서, 개인맞춤형 지도를 위한 교육 전략을 채택하는 데에 전문가 알고리즘 내지 전문가 조언을 사용하는 접근법, 예컨대, 학생 개개인의 최적의 교육 시퀀스를 위한 교습 모델에서 멀티 암드 밴딧(Multi-Armed Bandit: MAB)을 사용하는 접근법과는 구별된다. 본 문서에서 논의되는 예시적인 추천기는 유리하게도 실시간으로 동작가능한 반면, 개인맞춤형 과제 선택을 위한 액션의 가능한 세트를 찾기 위해 탐색-이용 트레이드오프에 있어서 전문가 지식을 사용하는 것, 예컨대, 각각의 활동(activity)에 대해 낙관적 액션을 선택하는 데에 전문가 알고리즘 Exp4를 활용하는 것은, 특히 각각의 활동에 다수의 액션과 전문가가 대응하는 경우에, 실시간 환경에 적합하지 않다.

[0078] 이상에서 본 발명의 몇몇 실시예가 상세하게 기술되었으나, 이는 제한적이 아니고 예시적인 것으로 간주되어야 한다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 개시된 실시예의 세부사항에 대해 본 발명의 범주에서 벗어나지 않고서 다양한 변경이 행해질 수 있음을 이해할 것이다. 그러므로 본 발명의 범주는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위 및 그 균등물에 의해 정해져야 한다.

부호의 설명

[0079] 100: 추천기

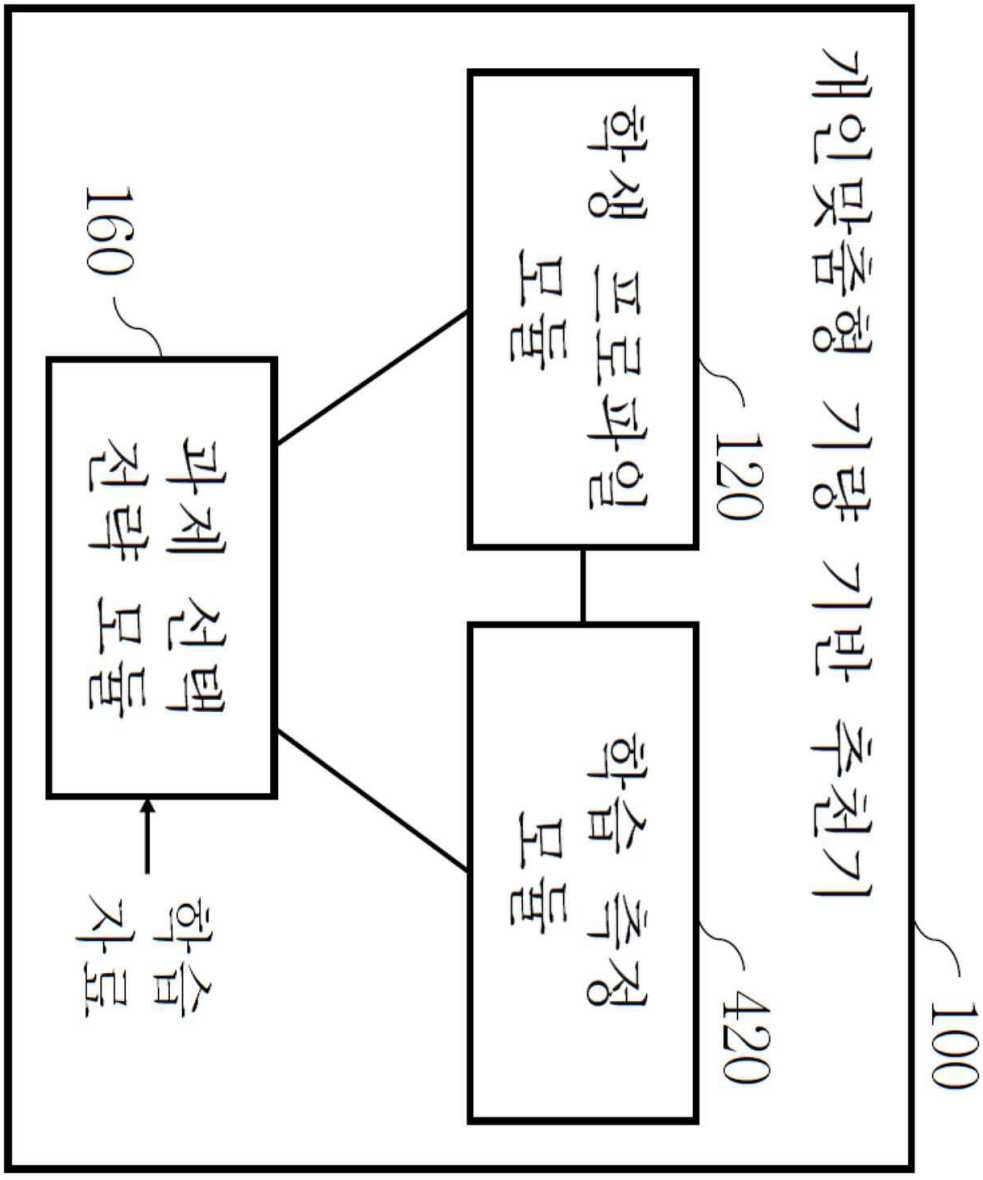
120: 학생 프로파일 모듈

140: 학습 측정 모듈

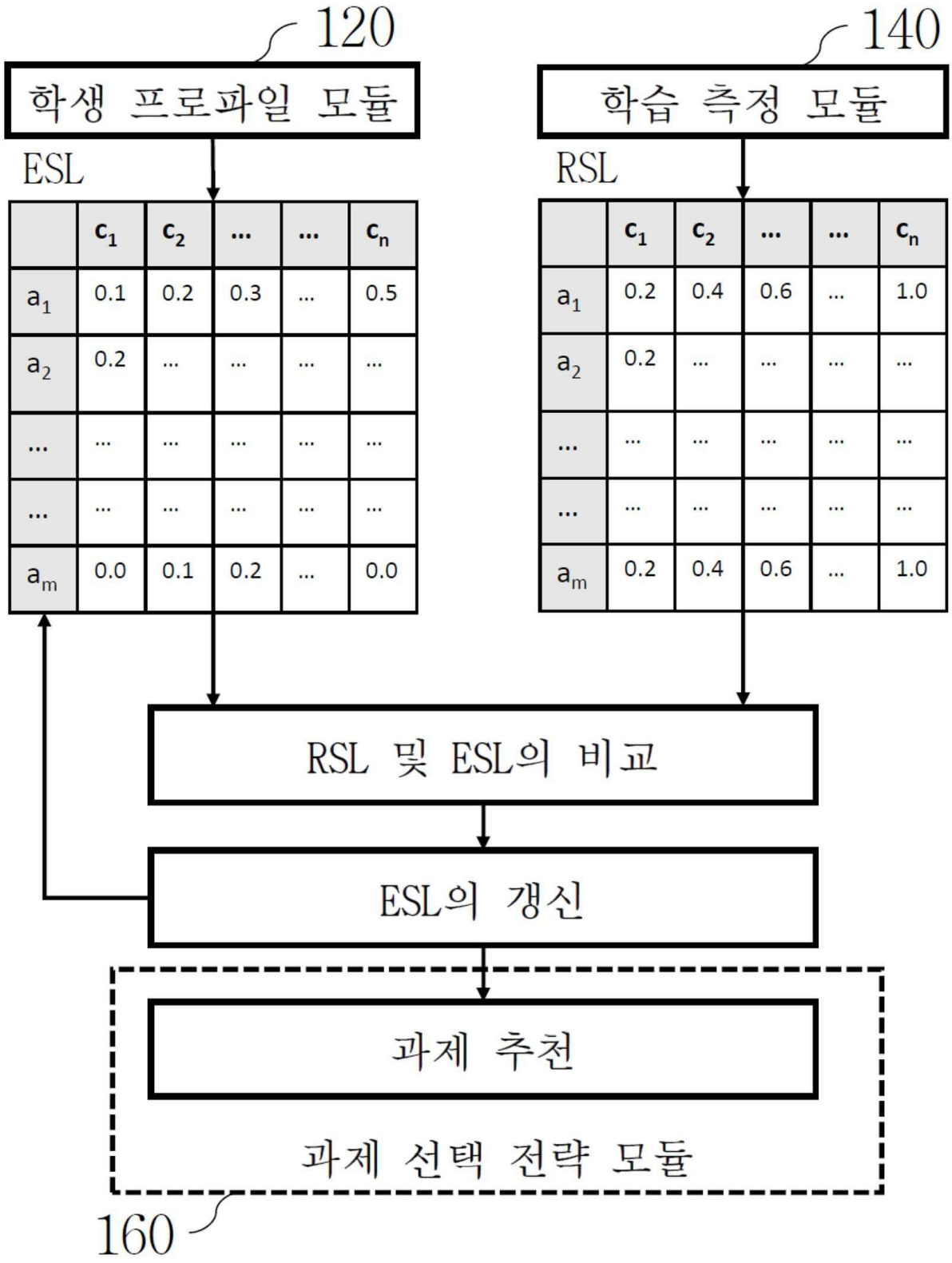
160: 과제 추천 전략 모듈

도면

도면1



도면2



	c_1	c_2	...	...	c_n
a_1	0.1	0.2	0.3	...	0.5
a_2	0.2	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
a_m	0.0	0.1	0.2	...	0.0

	c_1	c_2	...	...	c_n
a_1	0.2	0.4	0.6	...	1.0
a_2	0.2	...	...	...	...
...	...	...	...	...	...
...	...	...	...	...	...
a_m	0.2	0.4	0.6	...	1.0

도면3

300

초기화

Egalitarianism factor $\gamma \in [0,1]$

Exponential weight w_i

ESL

310

while *studying* **do**

선택 확률 설정

for all $i = 1, 2, \ldots, K$ **do**

$$p_i(t) = (1-\gamma)\frac{w_i(t)}{\sum_{j=1}^{K} w_j(t)} + \frac{\gamma}{K}$$

end

320

과제 추천

Recommend task $a=i_t \in (a_1, a_2, a_3, \ldots, a_K)$

330

보상 생성

if *answer* = 1

$$R_{(a,c)} = max\{(RSL_{(a,c)} - ESL_{(a,c)}), 0\}$$

else

$$R_{(a,c)} = min\{(RSL_{(a,c)} - ESL_{(a,c)}), 0\}$$

340

ESL 갱신

$$ESL_{(a,c)} \leftarrow ESL_{(a,c)} + \eta R_{(a,c)}$$

350

보상 추정 및 지수적 강화

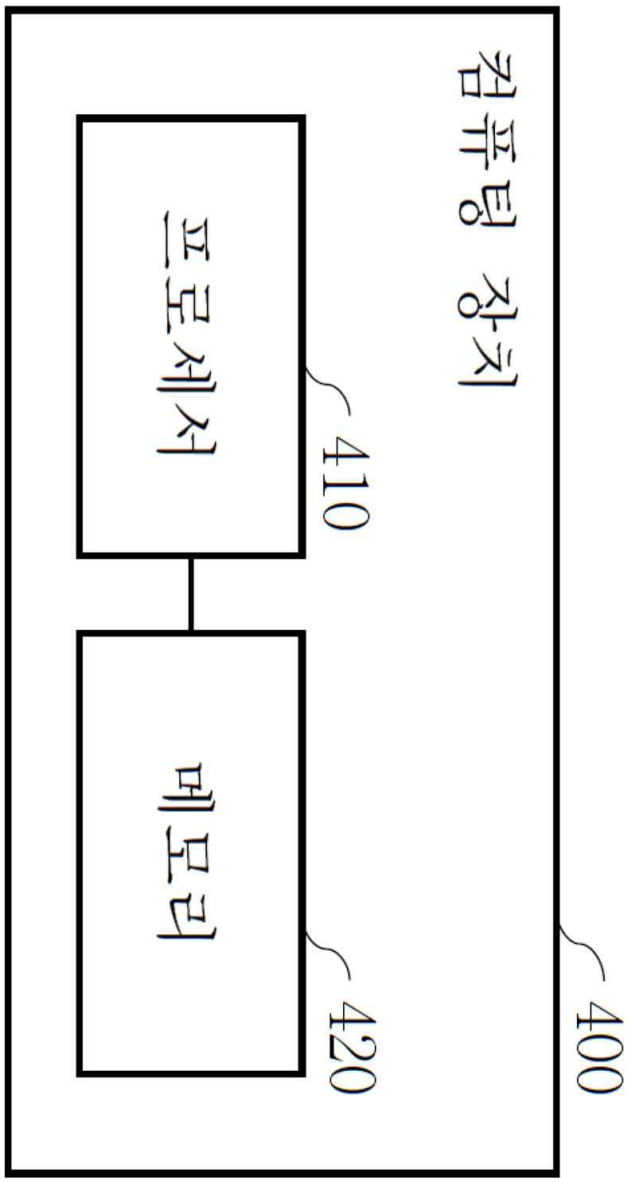
for $j = 1, 2, \ldots, K$ **do**

$$\widehat{R_j}(t) = \begin{cases} R_j(t)/p_j(t), & j = i_t \\ 0, & j \neq i_t \end{cases}$$

$$w_j(t+1) = w_j(t) \cdot e^{\left(\frac{\gamma}{K} \cdot \widehat{R_j}(t)\right)}$$

end

end

360

도면4