(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0017616
(43) 공개일자 2025년02월04일

(51) 국제특허분류(Int. Cl.)
F22B 35/18 (2006.01) *F22B 13/04* (2006.01)
F23N 5/26 (2006.01) *G06N 20/00* (2019.01)
(52) CPC특허분류
F22B 35/18 (2025.01)
F22B 13/04 (2013.01)
(21) 출원번호 10-2023-0172208
(22) 출원일자 2023년12월01일
심사청구일자 2023년12월01일
(30) 우선권주장
1020230098373 2023년07월27일 대한민국(KR)

(71) 출원인
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
윤성환
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
임진혁
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
(뒷면에 계속)
(74) 대리인
특허법인 수

전체 청구항 수 : 총 56 항

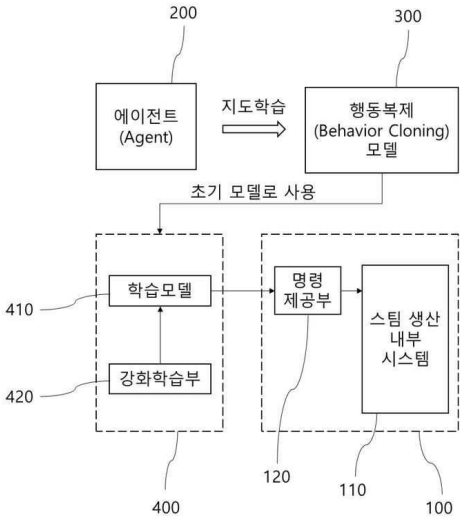
(54) 발명의 명칭 **강화학습 기반 연소 시스템 및 그 방법**

(57) 요 약

본 발명은 안정적인 학습을 위해 이전의 공장 시스템이 취했던 선택들을 이용해 공장 시스템을 모방하는 새로운 모델을 학습하고, 해당 모델의 정책에서 출발해 설정된 보상을 기반으로 효율을 올릴 수 있는 새로운 정책을 학습함으로써, 시스템 에이전트의 동작을 효과적으로 제어하고, 강화학습에 기반한 새로운 정책을 적용해 시스템 에이전트의 동작을 효과적으로 제어함으로써, LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시킬 수 있는 강화학습 기반 연소 시스템 및 그 방법에 관한 것이다.

본 발명의 강화학습 기반 연소 시스템은 순수한 물을 공급하여 스팀을 생산하는 스팀 생산 공장 시스템, 및 스팀 생산 공장 시스템의 상태 정보를 토대로 스팀 생산 공장 시스템으로 제어 정보를 전송하는 서버로 이루어진다.

대 표 도 - 도1



(52) CPC특허분류
F23N 5/265 (2013.01)
G06N 20/00 (2021.08)
F23N 2241/04 (2020.01)
(72) 발명자
권도현
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)
이재준
울산광역시 울주군 언양읍 유니스트길 50 (반연리, 울산과학기술원)

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	1/1
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

순수한 물을 공급하여 스팀을 생산하는 스팀 생산 공장 시스템; 및

상기 스팀 생산 공장 시스템의 상태 정보를 토대로 상기 스팀 생산 공장 시스템으로 제어 정보를 전송하는 서버;를 포함하는 강화학습 기반 연소 시스템.

청구항 2

제 1항에 있어서,

상기 스팀 생산 공장 시스템은, LNG(Liquefied Natural Gas)를 연료로 하여 스팀을 생산하는 스팀 생산 내부 시스템을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 3

제 2항에 있어서,

상기 스팀 생산 공장 시스템은, 상기 제어 정보를 분석하여 상기 스팀 생산 내부 시스템을 제어하는 명령 제공부를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 4

제 3항에 있어서,

상기 서버는, 인공지능 학습이 수행되어 상기 스팀 생산 공장 시스템으로 상기 제어 정보를 전송하는 학습 모델을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 5

제 4항에 있어서,

상기 서버는, 상기 상태 정보를 토대로 강화학습을 주관하여 상기 학습 모델의 파라미터를 업데이트하는 강화학습 제어부를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 6

제 4항에 있어서,

상기 학습 모델은, 행동복제(Behavior Cloning) 모델을 초기화 모델로 사용하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 7

제 6항에 있어서,

상기 행동복제(Behavior Cloning) 모델은, 에이전트(agent)에 의해 지도학습이 수행되는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 8

제 7항에 있어서,

상기 지도학습은, 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 상기 스팀 생산 공장 시스템을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행되는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 9

제 8항에 있어서,

상기 입력변수는, 목표 정보, 환경 정보, 위험 정보, 상기 액츄에이터 제어 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 10

제 9항에 있어서,

상기 목표 정보는, 생산 스팀량을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 11

제 9항에 있어서,

상기 환경 정보는, 상기 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 12

제 9항에 있어서,

상기 위험 정보는, 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 13

제 9항에 있어서,

상기 액츄에이터 제어 정보는, LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 14

제 9항에 있어서,

상기 지도학습은, 안정적인 학습 정보를 보유한 기존 상기 스팀 생산 공장 시스템이 취했던 선택들을 이용해 상기 스팀 생산 공장 시스템을 모방하는 새로운 모델을 학습 하는 행동 복제(Behavior Cloning)를 토대로 학습하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 15

제 14항에 있어서,

상기 지도학습의 입력은, 상기 입력변수에 대해 적어도 어느 하나 이상의 시간 간격으로 동시에 사용하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 16

제 15항에 있어서,

상기 지도학습은, 상기 입력변수에 대해 데이터 전처리 과정에서의 결측치를 제거한 후 상기 학습을 수행하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 17

제 6항에 있어서,

상기 강화학습 제어부는, 상기 행동복제(Behavior Cloning) 모델에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 상기 스팀 생산 공장 시스템의 제어 효율을 높이는 방향으로 새로운 정책(policy)을

학습하는 강화학습(Reinforcement Learning)을 수행하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 18

제 17항에 있어서,

상기 강화학습은, 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 상기 스팀 생산 공장 시스템을 제어할 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행되는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 19

제 18항에 있어서,

상기 입력변수는, 목표 정보, 환경 정보, 위험 정보, 상기 액츄에이터 제어 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 20

제 19항에 있어서,

상기 목표 정보는, 생산 스팀량을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 21

제 19항에 있어서,

상기 환경 정보는, 상기 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 22

제 19항에 있어서,

상기 위험 정보는, 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 23

제 19항에 있어서,

상기 액츄에이터 제어 정보는, LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 24

제 19항에 있어서,

상기 강화학습은, 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 25

제 24항에 있어서,

상기 강화학습은, 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 토대로 행동을 선택함으로써, 상기 정책(policy)를 업데이트하는 결정론적 행위 모듈을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 26

제 25항에 있어서,

상기 강화학습은, 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트하는 Q값 비평 모듈을 포함하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 27

제 26항에 있어서,

상기 보상(reward)은, LNG 가스 유량을 상기 외부 공기 투입량으로 나눈 값으로 사용하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 28

제 27항에 있어서,

상기 보상(reward) 중 상기 위험 정보가 일정값 이상일 경우 페널티(penalty) 요소를 추가하는 것을 특징으로 하는 강화학습 기반 연소 시스템.

청구항 29

스팀 생산 공장 시스템에서 순수한 물을 공급하여 스팀을 생산하는 스팀 생산단계; 및

서버에서 상기 스팀 생산 공장 시스템의 상태 정보를 토대로 상기 스팀 생산 공장 시스템으로 제어 정보를 전송하는 상태수집 및 제어단계;를 포함하는 강화학습 기반 연소 방법.

청구항 30

제 29항에 있어서,

상기 스팀 생산단계는, 상기 스팀 생산 공장 시스템 내의 스팀 생산 내부 시스템에서 LNG를 연료로 하여 스팀을 생산하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 31

제 30항에 있어서,

상기 스팀 생산단계는, 상기 스팀 생산 공장 시스템 내 명령 제공부에서 상기 제어 정보를 분석하여 상기 스팀 생산 내부 시스템을 제어하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 32

제 31항에 있어서,

상기 상태수집 및 제어단계는, 상기 서버 내 학습 모델에서 인공지능 학습이 수행되어 상기 스팀 생산 공장 시스템으로 상기 제어 정보를 전송하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 33

제 32항에 있어서,

상기 상태수집 및 제어단계는, 상기 서버 내 강화학습 제어부에서 상기 상태 정보를 토대로 강화학습을 주관하여 상기 학습 모델의 파라미터를 업데이트하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 34

제 32항에 있어서,

상기 상태수집 및 제어단계는, 상기 학습 모델의 초기화 모델로 행동복제(Behavior Cloning) 모델을 사용하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 35

제 34항에 있어서,

상기 상태수집 및 제어단계는, 상기 행동복제(Behavior Cloning) 모델이 에이전트(agent)에 의해 지도학습이 수

행되는 지도학습 단계를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 36

제 35항에 있어서,

상기 지도학습 단계는, 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 상기 스팀 생산 공장 시스템을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행되는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 37

제 36항에 있어서,

상기 입력변수는, 목표 정보, 환경 정보, 위험 정보, 상기 액츄에이터 제어 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 38

제 37항에 있어서,

상기 목표 정보는, 생산 스팀량을 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 39

제 37항에 있어서,

상기 환경 정보는, 상기 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 40

제 37항에 있어서,

상기 위험 정보는, 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 41

제 37항에 있어서,

상기 액츄에이터 제어 정보는, LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 42

제 37항에 있어서,

상기 지도학습 단계는, 안정적인 학습 정보를 보유한 기존 상기 스팀 생산 공장 시스템이 취했던 선택들을 이용해 상기 스팀 생산 공장 시스템을 모방하는 새로운 모델을 학습 하는 행동 복제(Behavior Cloning)를 토대로 학습하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 43

제 42항에 있어서,

상기 지도학습 단계에서의 입력은, 상기 입력변수에 대해 적어도 어느 하나 이상의 시간 간격으로 동시에 사용하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 44

제 43항에 있어서,

상기 지도학습 단계는, 상기 입력변수에 대해 데이터 전처리 과정에서의 결측치를 제거한 후 상기 학습을 수행하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 45

제 34항에 있어서,

상기 상태수집 및 제어단계는, 강화학습 제어부에서, 상기 행동복제(Behavior Cloning) 모델에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 상기 스팀 생산 공장 시스템의 제어 효율을 높이는 방향으로 새로운 정책(policy)을 학습하는 강화학습(Reinforcement Learning)을 수행하는 강화학습 단계를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 46

제 45항에 있어서,

상기 강화학습 단계는, 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 상기 스팀 생산 공장 시스템을 제어할 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행되는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 47

제 46항에 있어서,

상기 입력변수는, 목표 정보, 환경 정보, 위험 정보, 상기 액츄에이터 제어 정보 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 48

제 47항에 있어서,

상기 목표 정보는, 생산 스팀량을 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 49

제 47항에 있어서,

상기 환경 정보는, 상기 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 50

제 47항에 있어서,

상기 위험 정보는, 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 51

제 47항에 있어서,

상기 액츄에이터 제어 정보는, LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 52

제 47항에 있어서,

상기 강화학습 단계는, 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 53

제 52항에 있어서,

상기 강화학습 단계는, 상기 강화학습 제어부 내 결정론적 행위 모듈에서 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 토대로 행동을 선택함으로써, 상기 정책(policy)를 업데이트하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 54

제 53항에 있어서,

상기 강화학습 단계는, 상기 강화학습 제어부 내 Q값 비평 모듈에서 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 55

제 54항에 있어서,

상기 보상(reward)은, 상기 LNG 가스 유량을 상기 외부 공기 투입량으로 나눈 값으로 사용하는 것을 특징으로 하는 강화학습 기반 연소 방법.

청구항 56

제 55항에 있어서,

상기 보상(reward) 중 상기 위험 정보가 일정값 이상일 경우 페널티(penalty) 요소를 추가하는 것을 특징으로 하는 강화학습 기반 연소 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 강화학습 기반 연소 시스템 및 그 방법에 관한 것으로, 상세하게는, LNG(Liquefied Natural Gas) 가스와 외부 공기를 이용하여 스팀을 생산하는 것이다. 또한, 본 발명은 인공지능(Artificial Intelligence) 기술 중 강화학습(Reinforcement learning)을 기반으로 효율적으로 스팀을 생산할 수 있는 강화학습 기반 연소 시스템 및 그 방법에 관한 것이다.

배 경 기 술

[0002] 기존에 스팀을 생산하여 이용하는 공장 환경에서는 LNG(Liquefied Natural Gas) 가스와 외부 공기를 이용해 스팀을 생산하는 것이 일반적이며, 이때 외부 부하에 맞춰 적절하게 스팀을 생산해야 한다.

[0003] 한편, 스팀 생산 시 LNG 가스를 적게 사용할수록 효율이 상승하나, 공장의 안전 측면에서 고려해야 할 변수들도 존재한다. 스팀 생산 시 고려해야 할 변수로는 크게 다음의 11가지 정도를 들 수 있다.

[0004] 즉, (1) LNG 가스 유량, (2) 외부 공기 투입량, (3) 생산 스팀량, (4) 진동수치, (5) 연소 공기 농도, (6) 주증기압력, (7) 배출 오염 물질의 양, (8) 외부 공기 온도, (9) 순수 온도, (10) 외부에서 공급되는 스팀 (11) 외부에서의 스팀 수요량 등이다.

[0005] 상술한 변수 중에서, (1) LNG 가스 유량 및 (2) 외부 공기 투입량 의 2가지 변수가 시스템 에이전트(agent)가 조절하는 변수이며, 시스템 에이전트는 LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시키도록 동작하게 된다.

[0006] 따라서, 이러한 시스템의 동작을 여러 가지 변수를 고려하여 효과적으로 제어할 필요가 있으나, 지금까지는 운전 경험이나 동작 데이터를 단순 적용하는 정도여서 효율성을 극대화 하기 어려운 문제점이 있어 이를 해결하려

는 연구가 지속되어 왔다.

[0007] 그 일례로, 대한민국 공개특허공보 제10-2011-0065514호에서는 엔진에 공급할 LNG 를 기화시키기 위한 열원을 별도의 가열 장치를 통해 공급하는 대신에 보일러에서 생산된 스팀으로부터 얻도록 한 보일러 스팀을 이용한 직접 열교환 방식의 연료가스 공급장치에 관해 개시하고 있다.

[0008] 그러나, 이 경우에도 학습에 의한 제어를 수행하지 않아 시스템의 동작을 효과적으로 제어하지 못하는 단점이 있다.

선행기술문헌

특허문헌

[0009] (특허문헌 0001) 대한민국 공개특허공보 제10-2011-0065514호(2011.07.01)

발명의 내용

해결하려는 과제

[0010] 본 발명의 목적은, 안정적인 학습을 위해 이전의 공장 시스템이 취했던 선택들을 이용해 공장 시스템을 모방하는 새로운 모델(Behavior Cloning Model)을 학습하고, 해당 모델의 정책(policy)에서 출발해 설정된 보상(reward)을 기반으로 효율을 올릴 수 있는 새로운 정책(policy)을 학습(Reinforcement Learning; 강화학습)함으로써, 시스템 에이전트의 동작을 효과적으로 제어하는 강화학습 기반 연소 시스템 및 그 방법을 제공하는 것이다.

[0011] 본 발명은 강화학습에 기반한 새로운 정책(policy)을 적용해 시스템 에이전트의 동작을 효과적으로 제어함으로써, LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시킬 수 있는 강화학습 기반 연소 시스템 및 그 방법을 제공하는데 또 다른 목적이 있다.

과제의 해결 수단

[0012] 본 발명에 따른 강화학습 기반 연소 시스템은 순수한 물을 공급하여 스팀을 생산하는 스팀 생산 공장 시스템 및 스팀 생산 공장 시스템의 상태 정보를 토대로 스팀 생산 공장 시스템으로 제어 정보를 전송하는 서버를 포함할 수 있다.

[0013] 여기서, 스팀 생산 공장 시스템은 LNG(Liquefied Natural Gas)를 연료로 하여 스팀을 생산하는 스팀 생산 내부 시스템을 포함할 수 있다.

[0014] 또한, 스팀 생산 공장 시스템은 제어 정보를 분석하여 스팀 생산 내부 시스템을 제어하는 명령 제공부를 포함할 수 있다.

[0015] 여기서, 서버는 인공지능 학습이 수행되어 스팀 생산 공장 시스템으로 제어 정보를 전송하는 학습 모델을 포함할 수 있다.

[0016] 또한, 서버는 상태 정보를 토대로 강화학습을 주관하여 학습 모델의 파라미터를 업데이트하는 강화학습 제어부를 포함할 수 있다.

[0017] 여기서, 학습 모델은 행동복제(Behavior Cloning) 모델을 초기화 모델로 사용할 수 있다.

[0018] 또한, 행동복제(Behavior Cloning) 모델은 에이전트(agent)에 의해 지도학습이 수행될 수 있다.

[0019] 여기서, 지도학습은 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0020] 또한, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보 중 적어도 어느 하나를 포함할 수 있다.

[0021] 여기서, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0022] 또한, 환경 정보는 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부

에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함할 수 있다.

[0023] 여기서, 위험 정보는 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함할 수 있다.

[0024] 또한, 액츄에이터 제어 정보는 LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함할 수 있다.

[0025] 여기서, 지도학습은 안정적인 학습 정보를 보유한 기존 스팀 생산 공장 시스템이 취했던 선택들을 이용해 스팀 생산 공장 시스템을 모방하는 새로운 모델을 학습 하는 행동 복제(Behavior Cloning)를 토대로 학습할 수 있다.

[0026] 또한, 지도학습의 입력은 입력변수에 대해 적어도 어느 하나 이상의 시간 간격으로 동시에 사용할 수 있다.

[0027] 여기서, 지도학습은 입력변수에 대해 데이터 전처리 과정에서의 결측치를 제거한 후 학습을 수행할 수 있다.

[0028] 또한, 강화학습 제어부는 행동복제(Behavior Cloning) 모델에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 스팀 생산 공장 시스템의 제어 효율을 높이는 방향으로 새로운 정책(policy)을 학습하는 강화학습(Reinforcement Learning)을 수행할 수 있다.

[0029] 여기서, 강화학습은 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템을 제어할 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0030] 또한, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보 중 적어도 어느 하나를 포함할 수 있다.

[0031] 여기서, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0032] 또한, 환경 정보는 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함할 수 있다.

[0033] 여기서, 위험 정보는 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함할 수 있다.

[0034] 또한, 액츄에이터 제어 정보는 LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함할 수 있다.

[0035] 여기서, 강화학습은 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용할 수 있다.

[0036] 또한, 강화학습은 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 토대로 행동을 선택함으로써, 정책(policy)를 업데이트하는 결정론적 행위 모듈을 포함할 수 있다.

[0037] 여기서, 강화학습은 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트하는 Q값 비평 모듈을 포함할 수 있다.

[0038] 또한, 보상(reward)은 LNG 가스 유량을 외부 공기 투입량으로 나눈 값으로 사용할 수 있다.

[0039] 한편, 보상(reward) 중 위험 정보가 일정값 이상일 경우 페널티(penalty) 요소를 추가할 수 있다.

[0040] 본 발명의 다른 실시예에 따른 강화학습 기반 연소 방법은 스팀 생산 공장 시스템에서 순수한 물을 공급하여 스팀을 생산하는 스팀 생산단계 및 서버에서 스팀 생산 공장 시스템의 상태 정보를 토대로 스팀 생산 공장 시스템으로 제어 정보를 전송하는 상태수집 및 제어단계를 포함할 수 있다.

[0041] 여기서, 스팀 생산단계는 스팀 생산 공장 시스템 내의 스팀 생산 내부 시스템에서 LNG를 연료로하여 스팀을 생산할 수 있다.

[0042] 또한, 스팀 생산단계는 스팀 생산 공장 시스템 내 명령 제공부에서 제어 정보를 분석하여 스팀 생산 내부 시스템을 제어할 수 있다.

[0043] 여기서, 상태수집 및 제어단계는 서버 내 학습 모델에서 인공지능 학습이 수행되어 스팀 생산 공장 시스템으로 제어 정보를 전송할 수 있다.

[0044] 또한, 상태수집 및 제어단계는 서버 내 강화학습 제어부에서 상태 정보를 토대로 강화학습을 주관하여 학습 모델의 파라미터를 업데이트할 수 있다.

[0045] 여기서, 상태수집 및 제어단계는 학습 모델의 초기화 모델로 행동복제(Behavior Cloning) 모델을 사용할 수 있다.

[0046] 또한, 상태수집 및 제어단계는 행동복제(Behavior Cloning) 모델이 에이전트(agent)에 의해 지도학습이 수행되는 지도학습 단계를 포함할 수 있다.

[0047] 여기서, 지도학습 단계는 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0048] 또한, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보 중 적어도 어느 하나를 포함할 수 있다.

[0049] 여기서, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0050] 또한, 환경 정보는 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함할 수 있다.

[0051] 여기서, 위험 정보는 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함할 수 있다.

[0052] 또한, 액츄에이터 제어 정보는 LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함할 수 있다.

[0053] 여기서, 지도학습 단계는 안정적인 학습 정보를 보유한 기존 스팀 생산 공장 시스템이 취했던 선택들을 이용해 스팀 생산 공장 시스템을 모방하는 새로운 모델을 학습 하는 행동 복제(Behavior Cloning)를 토대로 학습할 수 있다.

[0054] 또한, 지도학습 단계에서의 입력은 입력변수에 대해 적어도 어느 하나 이상의 시간 간격으로 동시에 사용할 수 있다.

[0055] 여기서, 지도학습 단계는 입력변수에 대해 데이터 전처리 과정에서의 결측치를 제거한 후 학습을 수행할 수 있다.

[0056] 또한, 상태수집 및 제어단계는 강화학습 제어부에서, 행동복제(Behavior Cloning) 모델에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 스팀 생산 공장 시스템의 제어 효율을 높이는 방향으로 새로운 정책(policy)을 학습하는 강화학습(Reinforcement Learning)을 수행하는 강화학습 단계를 포함할 수 있다.

[0057] 여기서, 강화학습 단계는 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템을 제어할 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0058] 또한, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보 중 적어도 어느 하나를 포함할 수 있다.

[0059] 여기서, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0060] 또한, 환경 정보는 스팀 생산 내부 시스템의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량 중 적어도 어느 하나를 포함할 수 있다.

[0061] 여기서, 위험 정보는 연소 공기 농도 및 주증기 압력 중 적어도 어느 하나를 포함할 수 있다.

[0062] 또한, 액츄에이터 제어 정보는 LNG 가스 유량 및 외부 공기 투입량 중 적어도 어느 하나를 포함할 수 있다.

[0063] 여기서, 강화학습 단계는 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책 그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용할 수 있다.

[0064] 또한, 강화학습 단계는 강화학습 제어부 내 결정론적 행위 모듈에서 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 토대로 행동을 선택함으로써, 정책(policy)를 업데이트할 수 있다.

[0065] 여기서, 강화학습 단계는 강화학습 제어부 내 Q값 비평 모듈에서 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트할 수 있다.

[0066] 또한, 보상(reward)은 LNG 가스 유량을 외부 공기 투입량으로 나눈 값으로 사용 할 수 있다.

[0067] 여기서, 보상(reward) 중 위험 정보가 일정값 이상일 경우 페널티(penalty) 요소를 추가할 수 있다.

발명의 효과

[0068] 본 발명에 의한 강화학습 기반 연소 시스템 및 그 방법은 안정적인 학습을 위해 이전의 공장 시스템이 취했던 선택들을 이용해 공장 시스템을 모방하는 새로운 모델(Behavior Cloning Model)을 학습하고, 해당 모델의 정책(policy)에서 출발해 설정된 보상(reward)을 기반으로 효율을 올릴 수 있는 새로운 정책(policy)을 학습(Reinforcement Learning; 강화학습)함으로써, 시스템 에이전트의 동작을 효과적으로 제어하는 장점이 있다.

[0069] 또한, 본 발명에 의한 강화학습 기반 연소 시스템 및 그 방법은 강화학습에 기반한 새로운 정책(policy)을 적용해 시스템 에이전트의 동작을 효과적으로 제어함으로써, LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시킬 수 있는 장점이 있다.

도면의 간단한 설명

[0070] 도 1은 본 발명의 일 실시예에 따른 강화학습 기반 연소 시스템을 나타낸 개략구성도이다.

도 2는 도 1의 스팀 생산 내부 시스템을 상세히 나타낸 도면이다.

도 3은 도 1의 강화학습 제어부를 상세히 나타낸 도면이다.

도 4는 본 발명의 일 실시예에 따른 강화학습 기반 연소 방법을 나타낸 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0071] 이하, 본 발명의 실시를 위한 구체적인 실시예를 첨부된 도면들을 참조하여 설명한다.

[0072] 본 발명을 설명함에 있어서 제 1, 제 2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 구성요소들은 용어들에 의해 한정되지 않을 수 있다. 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제 1 구성요소는 제 2 구성요소로 명명될 수 있고, 유사하게 제 2 구성요소도 제 1 구성요소로 명명될 수 있다.

[0073] 어떤 구성요소가 다른 구성요소에 연결되어 있다거나 접속되어 있다고 언급되는 경우는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해될 수 있다.

[0074] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함할 수 있다.

[0075] 본 명세서에서, 포함하다 또는 구비하다 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것으로서, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해될 수 있다.

[0076] 또한, 도면에서의 요소들의 형상 및 크기 등은 보다 명확한 설명을 위해 과장될 수 있다.

[0077] 이하, 첨부된 도면을 참조하여 본 발명에 따른 강화학습 기반 연소 시스템 및 그 방법에 대해 상세히 설명한다.

[0079] 도 1은 본 발명의 일 실시예에 따른 강화학습 기반 연소 시스템을 나타낸 개략구성도이며, 도 2 및 도 3은 도 1을 상세히 설명하기 위한 세부 도면이다.

[0080] 이하, 도 1 내지 도 3을 참조하여 본 발명의 일 실시예에 따른 강화학습 기반 연소 시스템을 설명한다.

[0081] 먼저, 도 1을 참조하면, 본 발명의 일 실시예에 따른 강화학습 기반 연소 시스템은 순수한 물을 공급하여 스팀을 생산하는 스팀 생산 공장 시스템(100), 및 스팀 생산 공장 시스템(100)의 상태 정보를 토대로 스팀 생산 공장 시스템(100)으로 제어 정보를 전송하는 서버(400)로 이루어진다.

[0082] 여기서, 스팀 생산 공장 시스템(100)은 LNG를 연료로 하여 스팀을 생산하는 스팀 생산 내부 시스템(110), 및 제어 정보를 분석하여 스팀 생산 내부 시스템(110)을 제어하는 명령 제공부(120)로 이루어진다.

[0083] 이때, 서버(400)는 인공지능 학습이 수행되어 스팀 생산 공장 시스템(100)으로 제어 정보를 전송하는 학습 모델(410) 및 스팀 생산 공장 시스템(100)의 상태 정보를 토대로 강화학습을 주관하여 학습 모델(410)의 파라미터를 업데이트하는 강화학습 제어부(420)로 이루어진다.

[0084] 한편, 학습 모델(410)은 강화학습 제어부(420)에서 학습 시 위험한 선택을 하는 것을 방지하기 위해, 행동복제(Behavior Cloning) 모델(300)을 초기화 모델로 사용할 수 있다. 여기서, 행동복제(Behavior Cloning) 모델(300)은 에이전트(agent)(200)에 의해 지도학습이 수행될 수 있으며, 이에 대해 도 2에서 상세히 설명한다.

[0085] 또한, 강화학습 제어부(420)는 행동복제(Behavior Cloning) 모델(300)에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 스팀 생산 공장 시스템(100)의 제어 효율을 높이는 방향으로 새로운 정책(policy)을 학습하는 강화학습(Reinforcement Learning)을 수행하며, 이에 대해서는 도 3에서 상세히 설명한다.

[0087] 도 2는 도 1의 스팀 생산 내부 시스템(110)을 상세히 나타낸 도면이다.

[0088] 도 2에서 볼 수 있는 바와 같이, 스팀 생산 내부 시스템(110)은 LNG를 사용하여 순수(pure water)를 스팀으로 만들고 이를 외부 부하(500)로 제공한다.

[0089] 일반적으로 스팀 생산 내부 시스템(110)은 스팀 생산 내부 시스템(110)의 운영자에 의해 운영되며, 이때 운영한 각종 상태 및 제어 정보를 토대로 지도학습을 수행할 수 있다.

[0090] 여기서, 지도학습은 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템(100)을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0091] 이때, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보로 분류할 수 있다.

[0092] 여기서, 스팀 생산 공장 시스템(100)의 최종 목표가 외부 부하(500)로 스팀을 공급하는 것이므로, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0093] 또한, 환경 정보는, 스팀 생산 내부 시스템(110)의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량을 포함할 수 있다.

[0094] 한편, 위험 정보는 연소 공기 농도 및 주증기 압력을 포함할 수 있는데, 위험 정보가 일정값 이상일 경우에는 스팀 생산 내부 시스템(110)의 운영이 위험함을 나타낼 수 있다.

[0095] 또한, 액츄에이터 제어 정보로 LNG 가스 유량 및 외부 공기 투입량을 포함하여 스팀 생산 내부 시스템(110)을 제어하고 스팀을 외부 부하(500)로 제공할 수 있다.

[0096] 여기서, 지도학습은, 안정적인 학습 정보를 보유한 기존의 스팀 생산 공장 시스템이 취했던 선택들을 이용해 스팀 생산 공장 시스템(100)을 모방하는 새로운 모델을 학습하는 행동 복제(Behavior Cloning)를 토대로 이루어질 수 있다.

[0097] 이때, 지도학습의 입력은 입력변수에 대해 하나 이상의 시간 간격으로 동시에 사용하여 시간적인 변화를 학습의 결과에 포함할 수 있도록 한다.

[0098] 또한, 지도학습은 입력변수에 대해 데이터 전처리 과정에서의 결측치(기록 불량 등)를 제거한 후 학습을 수행하여 학습 결과가 안정적으로 될 수 있도록 한다.

[0100] 도 3은 도 1의 강화학습 제어부(420)를 상세히 나타낸 도면이다.

[0101] 도 3에서 알 수 있는 바와 같이, 본 발명의 강화학습 제어부(420)는 도 2에서 언급한 지도학습과 동일한 입력변수와 출력변수를 사용할 수 있다.

[0102] 즉, 강화학습 제어부(420)는 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템(100)을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습을 수행할 수 있다.

[0103] 이때, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보로 분류할 수 있으며, 상기의 각 입력변수에 대해서는 도 2에서 이미 설명한 바가 있으므로 이에 대한 자세한 설명은 생략한다.

[0104] 한편, 본 발명에서 강화학습은 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용할 수 있다.

[0105] 즉, 도 3에서와 같이, 본 발명의 강화학습 제어부(420)는 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되, 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고 이를 토대

로 행동을 선택함으로써 정책(policy)를 업데이트하는 결정론적 행위 모듈(421), 및 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트하는 Q값 비평 모듈(422)로 이루어진다.

[0106] 여기서, 결정론적 행위 모듈(421)에서의 행위자(Actor)와 Q값 비평 모듈(422)에서의 비평자(Critic)는 강화학습에서 사용되는 알고리즘 중 하나로, 주어진 환경에서 에이전트가 학습을 통해 최적의 행동을 선택하는 방법을 학습하는 방법론이다. 이 알고리즘은 정책(policy)과 가치함수(value function)를 동시에 학습하는 형태로 구성되어 있다.

[0107] 즉, 행위자는 에이전트의 행동을 결정하는 정책(policy) 함수를 학습한다. 정책 함수는 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하는 역할을 한다. 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 기반으로 행동을 선택한다.

[0108] 비평자는 주어진 상태(state)에서의 가치함수(value function)를 학습한다. 가치함수는 주어진 상태에서의 기대값 또는 평균적인 보상(reward)을 추정하는 함수이다. 비평자는 에이전트가 선택한 행동의 가치를 평가하고, 이를 통해 행동의 가치를 업데이트한다.

[0109] 여기서, 보상(reward)은 LNG 가스 유량을 외부 공기 투입량으로 나눈 값으로 사용하며, 이때 보상(reward)으로 인해 위험 정보가 일정값 이상으로 갈 경우 페널티(penalty) 요소를 추가하여 스팀 생산 내부 시스템(110)을 위험 요소로부터 안정적으로 운영할 수 있도록 한다.

[0111] 이상과 같은 본 발명에 의한 강화학습 기반 연소 시스템은 안정적인 학습을 위해 이전의 공장 시스템이 취했던 선택들을 이용해 공장 시스템을 모방하는 새로운 모델(Behavior Cloning Model)을 학습하고, 해당 모델의 정책(policy)에서 출발해 설정된 보상(reward)을 기반으로 효율을 올릴 수 있는 새로운 정책(policy)을 학습(Reinforcement Learning; 강화학습)함으로써, 시스템 에이전트의 동작을 효과적으로 제어할 수 있다. 또한, 강화학습에 기반한 새로운 정책(policy)을 적용해 시스템 에이전트의 동작을 효과적으로 제어함으로써, LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시킬 수 있다.

[0113] 다음, 도 4는 본 발명의 일 실시예에 따른 강화학습 기반 연소 방법을 나타낸 순서도이다.

[0114] 도 4에서 볼 수 있는 바와 같이, 본 발명의 강화학습 기반 연소 방법은 스팀 생산 공장 시스템(100)에서 순수한 물을 공급하여 스팀을 생산하는 스팀 생산단계(S100), 및 서버(400)에서 스팀 생산 공장 시스템(100)의 상태 정보를 토대로 스팀 생산 공장 시스템(100)으로 제어 정보를 전송하는 상태수집 및 제어단계(S200)로 이루어진다.

[0115] 여기서, 스팀 생산단계(S100)는 스팀 생산 공장 시스템(100) 내의 스팀 생산 내부 시스템(110)에서 LNG를 연료로 하여 스팀을 생산하고 스팀 생산 공장 시스템(100) 내 명령 제공부(120)에서 제어 정보를 분석하여 스팀 생산 내부 시스템(110)을 제어한다.

[0116] 또한, 상태수집 및 제어단계(S200)는 서버(400) 내 학습 모델(410)에서 인공지능 학습이 수행되어 스팀 생산 공장 시스템(100)으로 제어 정보를 전송하고, 서버(400) 내 강화학습 제어부(420)에서 상태 정보를 토대로 강화학습을 주관하여 학습 모델(410)의 파라미터를 업데이트한다.

[0117] 여기서, 상태수집 및 제어단계(S200)는 학습 모델(410)은 강화학습 제어부(420)에서 학습 시 위험한 선택을 하는 것을 막기 위해, 행동복제(Behavior Cloning) 모델(300)을 초기화 모델로 사용할 수 있다.

[0118] 한편, 상태수집 및 제어단계(S200)는 행동복제(Behavior Cloning) 모델(300)이 에이전트(agent)(200)에 의해 지도학습이 수행되는 지도학습 단계(S210), 및 강화학습 제어부(420)에서, 행동복제(Behavior Cloning) 모델(300)에서 사용된 정책(policy)에서 시작하여 설정된 보상(reward)를 기반으로 스팀 생산 공장 시스템(100)의 제어 효율을 높이는 방향으로 새로운 정책(policy)을 학습하는 강화학습(Reinforcement Learning)을 수행하는 강화 학습 단계(S220)로 이루어진다.

[0119] 이때, 지도학습 단계(S210)는 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템(100)을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0120] 이때, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보로 분류할 수 있다.

[0121] 여기서, 스팀 생산 공장 시스템(100)의 최종 목표가 외부 부하(500)로 스팀을 공급하는 것이므로, 목표 정보는 생산 스팀량을 포함할 수 있다.

[0122] 또한, 환경 정보는, 스팀 생산 내부 시스템(110)의 진동수치, 매출 오염 물질의 양, 외부 공기 온도, 순수 온도, 외부에서 공급되는 스팀, 및 외부에서의 스팀 수요량을 포함할 수 있다.

[0123] 한편, 위험 정보는 연소 공기 농도 및 주증기 압력을 포함할 수 있는데, 위험 정보가 일정값 이상일 경우에는 스팀 생산 내부 시스템(110)의 운영이 위험함을 나타낼 수 있다.

[0124] 또한, 액츄에이터 제어 정보로 LNG 가스 유량 및 외부 공기 투입량을 포함하여 스팀 생산 내부 시스템(110)을 제어하고 스팀을 외부 부하(500)로 제공할 수 있다.

[0125] 여기서, 지도학습 단계(S210)는, 안정적인 학습 정보를 보유한 기존의 스팀 생산 공장 시스템이 취했던 선택들을 이용해 스팀 생산 공장 시스템(100)을 모방하는 새로운 모델을 학습하는 행동 복제(Behavior Cloning)를 토대로 이루어질 수 있다.

[0126] 이때, 지도학습 단계(S210)에서의 입력은 입력변수에 대해 하나 이상의 시간 간격으로 동시에 사용하여 시간적인 변화를 학습의 결과에 포함할 수 있도록 한다.

[0127] 또한, 지도학습 단계(S210)에서는 입력변수에 대해 데이터 전처리 과정에서의 결측치(기록 불량 등)를 제거한 후 학습을 수행하여 학습 결과가 안정적으로 될 수 있도록 한다.

[0129] 한편, 본 발명의 강화학습 단계(S220)는 스팀 생산 시 고려해야 할 상태를 입력변수로 하고 스팀 생산 공장 시스템(100)을 제어할 액츄에이터(actuator)의 액츄에이터 제어 정보를 출력변수로 하여 학습이 수행될 수 있다.

[0130] 이때, 입력변수는 목표 정보, 환경 정보, 위험 정보, 액츄에이터 제어 정보로 분류할 수 있으며, 상기의 각 입력변수에 대해서는 앞서 이미 설명한 바가 있으므로 이에 대한 자세한 설명은 생략한다.

[0131] 한편, 강화학습 단계(S220)는 정책(policy)와 가치함수(value function)을 동시에 업데이트하는 심층결정론적 정책그라디언트(DDPG; Deep Deterministic Policy Gradient) 모델을 사용할 수 있다.

[0132] 여기서, 강화학습 단계(S220)는, 강화학습 제어부(420) 내 결정론적 행위 모듈(421)에서 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하되, 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고 이를 토대로 행동을 선택함으로써 정책(policy)를 업데이트하고, 강화학습 제어부(420) 내 Q값 비평 모듈(422)에서 주어진 상태(state)에서의 기대값 또는 평균적인 보상(reward)을 추정하여 가치함수(value function)을 업데이트 한다.

[0133] 이때, 결정론적 행위 모듈(421)에서의 행위자(Actor)와 Q값 비평 모듈(422)에서의 비평자(Critic)는 강화학습에서 사용되는 알고리즘 중 하나로, 주어진 환경에서 에이전트가 학습을 통해 최적의 행동을 선택하는 방법을 학습하는 방법론이다. 이 알고리즘은 정책(policy)과 가치함수(value function)를 동시에 학습하는 형태로 구성되어 있다.

[0134] 즉, 행위자는 에이전트의 행동을 결정하는 정책(policy) 함수를 학습한다. 정책 함수는 주어진 상태(state)에서 에이전트가 어떤 행동(action)을 선택해야 할지를 결정하는 역할을 한다. 주어진 상태에서 가능한 행동들에 대한 확률 분포를 출력하고, 이를 기반으로 행동을 선택한다.

[0135] 비평자는 주어진 상태(state)에서의 가치함수(value function)를 학습한다. 가치함수는 주어진 상태에서의 기대값 또는 평균적인 보상(reward)을 추정하는 함수이다. 비평자는 에이전트가 선택한 행동의 가치를 평가하고, 이를 통해 행동의 가치를 업데이트한다.

[0136] 여기서, 보상(reward)은 LNG 가스 유량을 외부 공기 투입량으로 나눈 값으로 사용하며, 이때 보상(reward)으로 위험 정보가 일정값 이상으로 갈 경우 페널티(penalty) 요소를 추가하여 스팀 생산 내부 시스템(110)을 위험 요소로부터 안정적으로 운영할 수 있도록 한다.

[0138] 이상과 같은 본 발명에 의한 강화학습 기반 연소 방법은 안정적인 학습을 위해 이전의 공장 시스템이 취했던 선택들을 이용해 공장 시스템을 모방하는 새로운 모델(Behavior Cloning Model)을 학습하고, 해당 모델의 정책

(policy)에서 출발해 설정된 보상(reward)을 기반으로 효율을 올릴 수 있는 새로운 정책(policy)을 학습(Reinforcement Learning; 강화학습)함으로써, 시스템 에이전트의 동작을 효과적으로 제어할 수 있다. 또한, 강화학습에 기반한 새로운 정책(policy)을 적용해 시스템 에이전트의 동작을 효과적으로 제어함으로써, LNG 가스 유량을 가능한 적게 사용하면서 생산되는 스팀량을 최대한 증가시킬 수 있다.

[0140] 본 발명의 기술 분야에서 통상의 지식을 가진 자는 여기에 개시된 실시예들과 관련하여 설명된 다양한 예시적인 논리 블록들, 모듈들, 프로세서들, 수단들, 회로들 및 알고리즘 단계들이 전자 하드웨어, (편의를 위해, 여기에서 소프트웨어로 지칭되는) 다양한 형태들의 프로그램 또는 설계 코드 또는 이들 모두의 결합에 의해 구현될 수 있다는 것을 이해할 것이다. 하드웨어 및 소프트웨어의 이러한 상호 호환성을 명확하게 설명하기 위해, 다양한 예시적인 컴포넌트들, 블록들, 모듈들, 회로들 및 단계들이 이들의 기능과 관련하여 위에서 일반적으로 설명되었다. 이러한 기능이 하드웨어 또는 소프트웨어로서 구현되는지 여부는 특정한 애플리케이션 및 전체 시스템에 대하여 부과되는 설계 제약들에 따라 좌우된다. 본 발명의 기술 분야에서 통상의 지식을 가진 자는 각각의 특정한 애플리케이션에 대하여 다양한 방식들로 설명된 기능을 구현할 수 있으나, 이러한 구현 결정들은 본 발명의 범위를 벗어나는 것으로 해석되어서는 안 될 것이다.

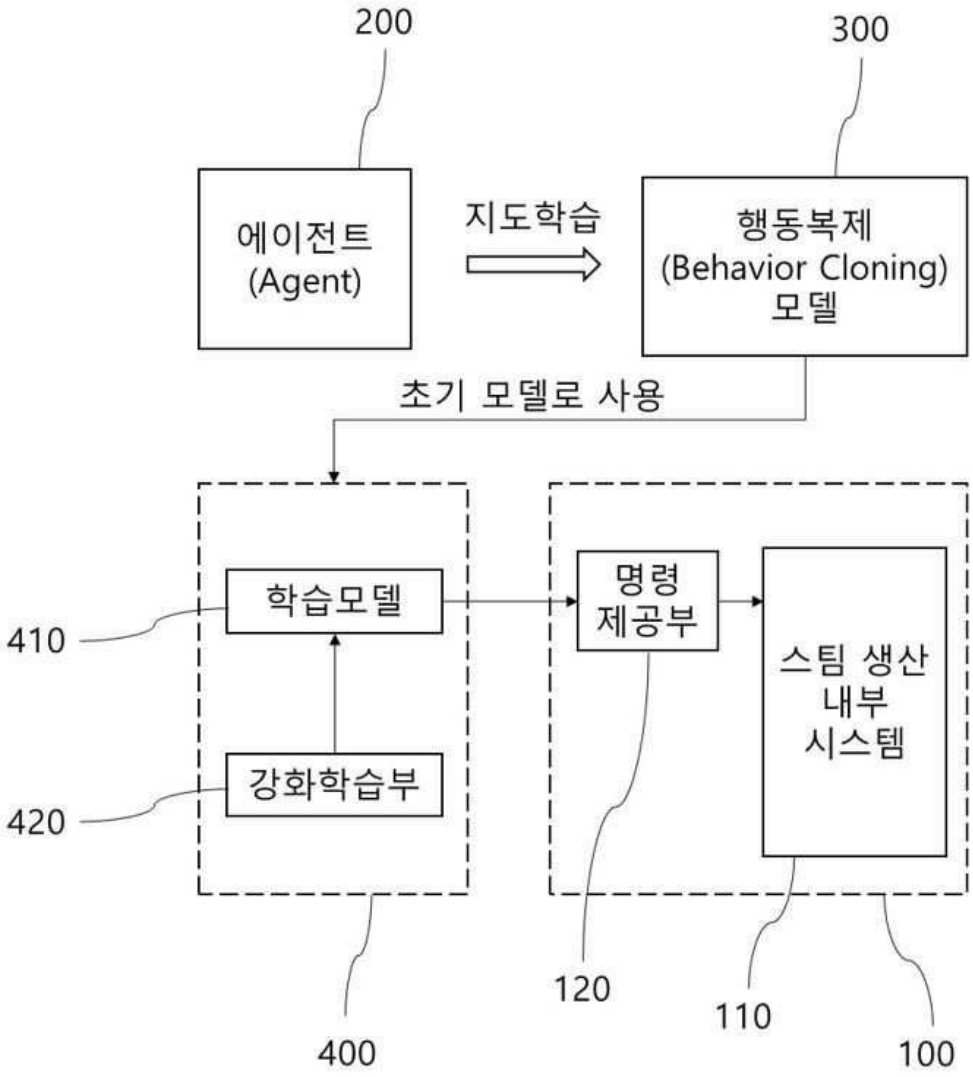
[0141] 여기서 제시된 다양한 실시예들은 방법, 장치, 또는 표준 프로그래밍 및/또는 엔지니어링 기술을 사용한 제조 물품(article)으로 구현될 수 있다. 용어 제조 물품은 임의의 컴퓨터-판독가능 저장장치로부터 액세스 가능한 컴퓨터 프로그램, 캐리어, 또는 매체(media)를 포함한다. 예를 들어, 컴퓨터-판독가능 저장매체는 자기 저장 장치(예를 들면, 하드 디스크, 플로피 디스크, 자기 스트립, 등), 광학 디스크(예를 들면, CD, DVD, 등), 스마트 카드, 및 플래쉬 메모리 장치(예를 들면, EEPROM, 카드, 스틱, 키 드라이브, 등)를 포함하지만, 이들로 제한되는 것은 아니다. 또한, 여기서 제시되는 다양한 저장 매체는 정보를 저장하기 위한 하나 이상의 장치 및/또는 다른 기계-판독가능한 매체를 포함한다.

[0142] 제시된 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조는 예시적인 접근들의 일례임을 이해하도록 한다. 설계 우선순위들에 기반하여, 본 발명의 범위 내에서 프로세스들에 있는 단계들의 특정한 순서 또는 계층 구조가 재배열될 수 있다는 것을 이해하도록 한다. 첨부된 방법 청구항들은 샘플 순서로 다양한 단계들의 엘리먼트들을 제공하지만 제시된 특정한 순서 또는 계층 구조에 한정되는 것을 의미하지는 않는다.

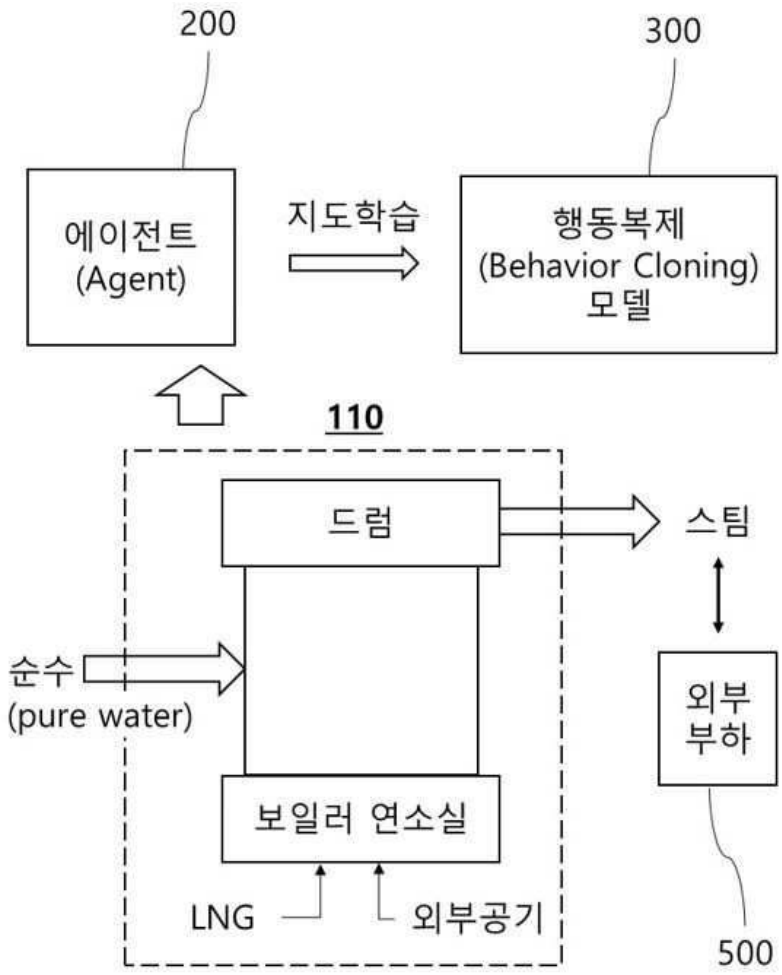
[0143] 제시된 실시예들에 대한 설명은 임의의 본 발명의 기술 분야에서 통상의 지식을 가진 자가 본 발명을 이용하거나 또는 실시할 수 있도록 제공된다. 이러한 실시예들에 대한 다양한 변형들은 본 발명의 기술 분야에서 통상의 지식을 가진 자에게 명백할 것이며, 여기에 정의된 일반적인 원리들은 본 발명의 범위를 벗어남이 없이 다른 실시예들에 적용될 수 있다. 그리하여, 본 발명은 여기에 제시된 실시예들로 한정되는 것이 아니라, 여기에 제시된 원리들 및 신규한 특징들과 일관되는 최광의의 범위에서 해석되어야 할 것이다.

도면

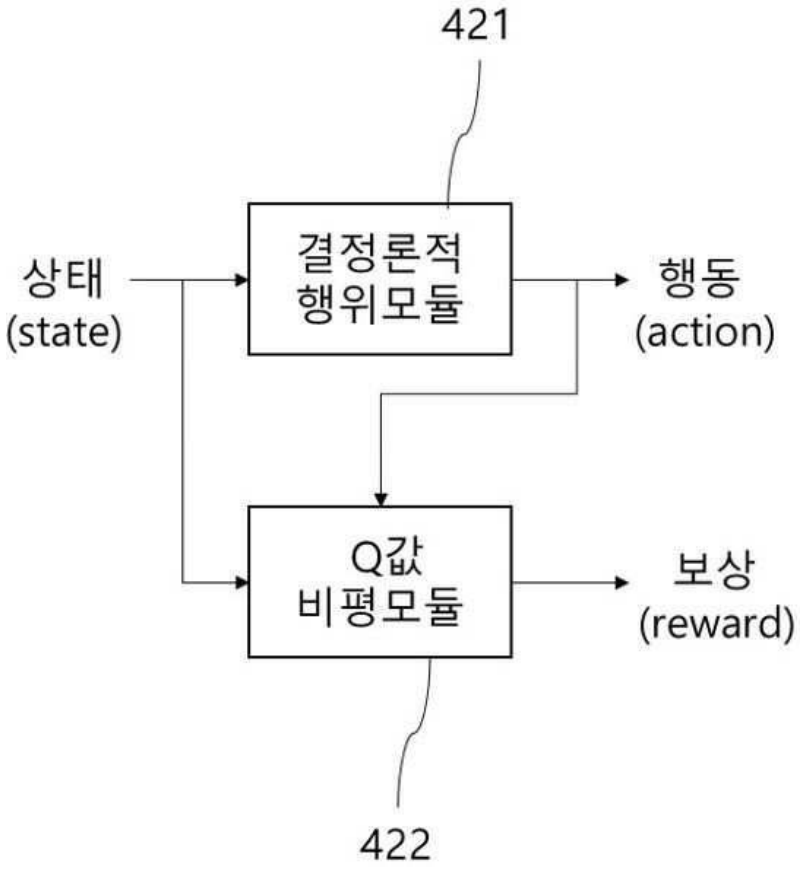
도면1


200
300
에이전트
(Agent)
지도학습
행동복제
(Behavior Cloning)
모델
초기 모델로 사용
학습모델
410
강화학습부
420
명령
제공부
스팀 생산
내부
시스템
400
120
110
100

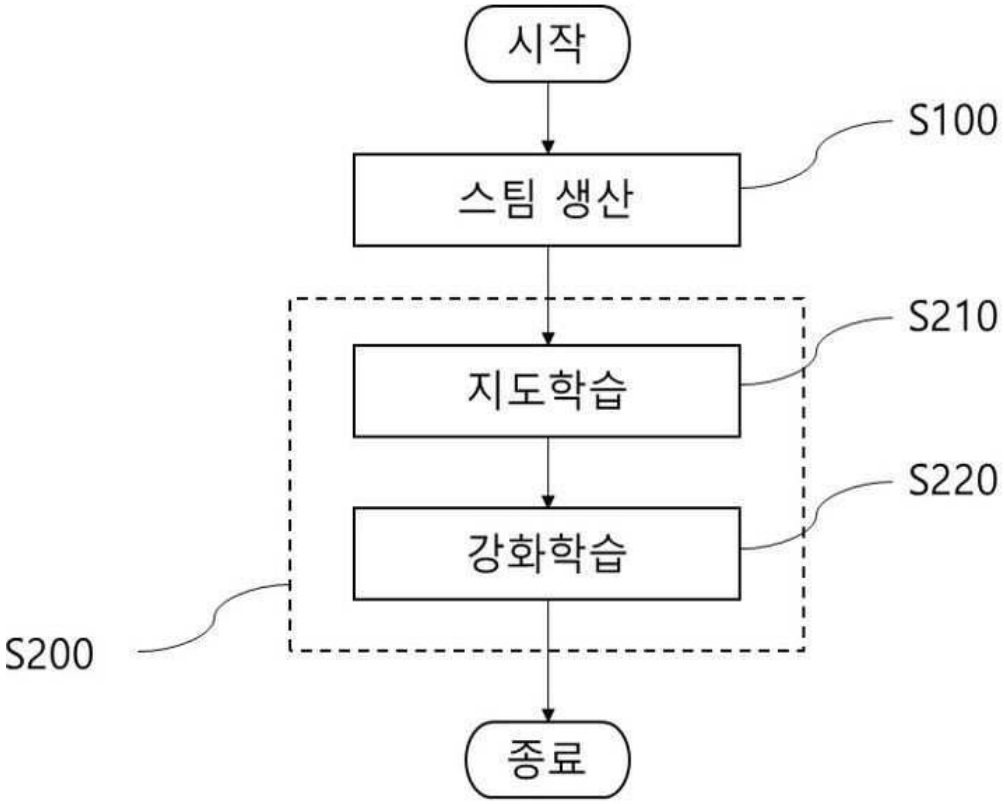
도면2


200
300
에이전트
(Agent)
지도학습
행동복제
(Behavior Cloning)
모델
110
드럼
스팀
순수
(pure water)
외부
부하
보일러 연소실
LNG
외부공기
500

도면3


421
결정론적
행위모듈
상태
(state)
행동
(action)
Q값
비평모듈
보상
(reward)
422

도면4


시작
S100
스팀 생산
S210
지도학습
S220
강화학습
S200
종료