(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0128191
(43) 공개일자 2023년09월04일

(51) 국제특허분류(Int. Cl.)
G06N 7/00 (2023.01) *C02F 1/00* (2023.01)
G05B 11/36 (2006.01)
(52) CPC특허분류
G06N 7/01 (2023.01)
C02F 1/00 (2013.01)
(21) 출원번호 10-2022-0025289
(22) 출원일자 2022년02월25일
심사청구일자 2022년02월25일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
유창규
경기도 수원시 영통구 웰빙타운로 20, 8303동 202호 (이의동, 호반가든하임)
허성구
서울특별시 강남구 선릉로72길 37(대치동)
남기전
경기도 수원시 영통구 청명북로7번길 16-10, 201호(영통동)
(74) 대리인
특허법인 아이퍼스

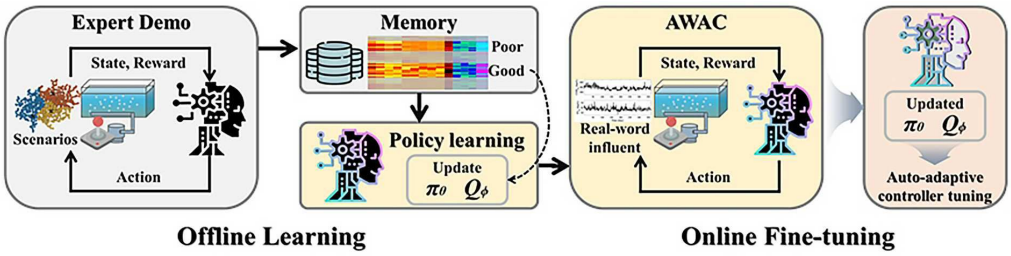
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 **강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템**

(57) 요 약

본 발명은 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템에 관한 것으로, 보다 상세하게는 하수처리장 공정 운영 제어기 파라미터 튜닝시스템으로서, 상기 하수처리장의 공정운영데이터를 수집, 축적하는 데이터수집부; 및 수집, 축적된 데이터에 따라 오프라인 강화학습을 통해 상기 제어기의 파라미터 값을 결정하는 파라미터 산출부;를 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템에 관한 것이다.

대 표 도 - 도1



(52) CPC특허분류
G05B 11/36 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711127448
과제번호 2021R1A2C2007838
부처명 과학기술정보통신부
과제관리(전문)기관명 (통합)한국연구재단
연구사업명 (혁신법)이공분야기초연구사업
연구과제명 [중견연구 유형1-2]DeepAI 기반 하폐수처리장 스마트워터 자율운전 & 자율설계 원천기술 개발
기 여 율 1/1
과제수행기관명 경희대학교산학협력단
연구기간 2021.03.01 ~ 2026.02.28

명 세 서

청구범위

청구항 1

하수처리장 공정 운영 제어기 파라미터 튜닝시스템으로서,

상기 하수처리장의 공정운영데이터를 수집, 축적하는 데이터수집부; 및

수집, 축적된 데이터에 따라 오프라인 강화학습을 통해 상기 제어기의 파라미터 값을 결정하는 파라미터 산출부;를 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 2

제 1항에 있어서,

상기 공정운영데이터는 기존 공정 제어기 파라미터 값 변화에 따른 에러값의 수집데이터인 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 3

제 2항에 있어서,

상기 오프라인 강화학습 알고리즘은 수집데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 4

제 3항에 있어서,

상기 제어기는 PID 제어기이고, 상기 파라미터 산출부는 상기 PID 제어기 3개의 파라미터 값을 자동적으로 튜닝하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 5

제 4항에 있어서,

상기 파라미터 산출부를 통해 도출된 파라미터 값으로 상기 PID 제어기가 운영되고,

운영 시, 공정유입-악조건 발생시 온라인 학습을 통해 상기 PED 제어기의 파라미터를 튜닝하는 인공지능 학습모듈;을 더 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 6

하수처리장 공정 운영 제어기 파라미터 튜닝방법으로서,

데이터수집부가 상기 하수처리장의 공정운영데이터를 수집, 축적하는 단계; 및

파라미터 산출부가 수집, 축적된 데이터에 따라 오프라인 강화학습을 통해 상기 제어기의 파라미터 값을 결정하는 단계;를 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝방법.

청구항 7

제 6항에 있어서,

상기 축적단계에서,

상기 공정운영데이터는 기존 공정 제어기 파라미터 값 변화에 따른 에러값의 수집데이터인 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템.

청구항 8

제 7항에 있어서,

상기 결정하는 단계에서,

상기 오프라인 강화학습 알고리즘을 통해 수집데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝방법.

청구항 9

제 8항에 있어서,

상기 제어기는 PID 제어기이고,

상기 결정하는 단계에서, 상기 파라미터 산출부는 상기 PID 제어기 3개의 파라미터 값을 자동적으로 튜닝하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝방법.

청구항 10

제 9항에 있어서,

상기 결정하는 단계 후에,

상기 파라미터 산출부를 통해 도출된 파라미터 값으로 상기 PID 제어기가 운영되는 단계; 및

운영 시, 공정유입-악조건 발생시 온라인 학습을 통해 상기 PED 제어기의 파라미터를 튜닝하는 단계;을 더 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 오프라인 강화 학습의 데이터 기반 스마트 결정을 사용하여 폐수 처리 프로세스를 위한 자동 적응 컨트롤러 튜닝 시스템에 관한 것이다.

배 경 기 술

[0002] 하수 처리 공정은 현대 사회에서 필수적인 역할로 여겨져 왔다. 환경을 보호하기 위해 물리적 생물학적 메커니즘을 사용하여 처리함으로써 산업 및 도시 폐수를 줄일 수 있다. 변동하는 유입 유량과 매우 다양한 유입 오염물질 농도에 영향을 받으므로 비선형 유입은 처리 과정의 복잡성과 처리 과정의 에너지 사용량을 증가시킨다. 처리공정 중 용존산소(DO)를 기체에서 액체로 이동시키는 것은 에너지 집약도가 높은 공정일 뿐만 아니라 생물학적 공정의 만족스러운 운전을 위한 중요한 요소이다. 폭기 시스템은 가장 큰 에너지 소비 공정으로 비선형 유입 조건에서도 견고하고 안정적인 폭기 제어가 필요하다.

[0003] 폭기 시스템의 제어는 하수처리장이 엄격한 폐수 품질 제한을 충족하면서 에너지 효율적인 방식으로 운영되어야

하는 경우 매우 중요하다. 비선형 유입수는 미생물의 복잡한 생물학적 및 물리적 메커니즘으로 인해 하수처리장(WWTP) 반응기의 DO 농도에 영향을 미칠 수 있습니다. 또한, 큰 시간 지연과 강한 비선형성을 포함하는 산소 전달 과정이 유입수의 유량 및 수질 변화와 결합될 때 DO 농도 제어는 더 어려울 것이다. 따라서 DO 농도는 동적 유입수 조건에서 원활하고 견고하게 제어되기 어렵다.

[0004] PID(Proportional-Integral-Derivative)는 DO 수준을 충족하는 폭기 시스템의 주요 제어 기술이었다. 그러나 선형 제어기의 고정 매개변수는 시간에 따라 변하는 경향, 높은 비선형성 및 불확실성으로 인해 만족스러운 제어 DO 성능을 유지하는 데 약점이 있다. 모델 예측 제어(MPC)는 WWTP의 제어 시스템에서 널리 구현되는 또 다른 제어 기술이다. MPC는 가까운 미래만을 예측하여 모든 시간 단계에서 제어 시스템을 재귀적으로 계산하여 폐쇄 루프를 수행한다. 실시간 교란 및 모델 플랜트 불일치에 대한 프로세스의 불확실성 및 비선형성을 효과적으로 극복할 수 있다. 그러나 MPC 알고리즘을 실제 WWTP에 적용하는 것은 경쟁 예측 모델을 구축하기 위해 많은 시변 매개변수를 식별하는 것이 어렵기 때문에 실제 WWTP에 적용하는 데 한계가 있다.

[0005] 최근 몇 년 동안 강화 학습(RL)은 복잡한 비선형 제어 문제를 해결하는 강력한 솔루션이었다. RL은 알려진 모든 방법을 능가하는 기능을 제공하여 기존 제어 시스템의 한계를 극복한다. 추적 및 오류 검색 솔루션과 후속 조치로 인한 보상을 통해 RL은 기존 제어를 능가하고 지식의 기존 경계를 확장할 수 있다. 로봇 공학, 난방-환기-공조 시스템, 화학 공학 분야의 펌핑 제어에 RL의 성공적인 구현을 통해, RL 기반 제어 시스템의 적용은 WWTP 연구 분야로 확장되었다. Hernandez-del-Olmo et al은 경제적인 WWTP 운영을 위해 산소 농도를 제어하기 위해 Q-learning 및 SARSA 알고리즘을 제안했다. Yang et al은 유출 질산염 농도를 최소화하면서 DO 제어를 위한 배우 비평가 기반 추적 제어 시스템을 제안했다. Chen et al은 지속 가능한 최적화를 위해 DDPG(Deep Deterministic Policy Gradient)를 기반으로 WWTP의 운영 체제를 제어했다. 이러한 연구는 뛰어난 결과를 나타내었지만 온라인 RL을 기반으로 했다. 온라인 학습 패러다임은 광범위한 적용에 장애물이 있다. 정책 개선을 위해 환경과 상호작용하여 경험을 수집해야 하므로 데이터 수집이 비용이 많이 들고 지속 가능한 환경 운영을 방해하는 WWTP 분야에서도 위험하기 때문에 온라인 상호 작용은 비실용적이다.

[0006] 즉, 하수처리장을 포함한 다양한 환경-화공 공정 시스템에서 공정 운영 설정치에 따라 공정을 운영하기 위한 제어기가 많이 사용되고 있으나, 동적이며 다양한 유입은 실제 공정의 비선형성을 증가시켜 기존의 제어 시스템인 proportional-integral-derivative (PID) 제어기 내의 변수를 튜닝하는데 어려움이 있다. 또한, 공정 운영에 있어 운영자의 실험적 제어기 튜닝은 공정의 안정적 운영을 제한할 수 있다.

[0007] 데이터 기반 학습 방법의 발전을 고려할 때 오프라인 RL(데이터 기반 RL)은 WWTP의 제어 시스템에 적합한 도구가 될 것이다. WWTP 운영 체제의 경우 실제 작업 조건은 일반적으로 외부 환경에 따라 변경된다. 따라서 실시간 온라인 RL 교육은 제어 시스템과 에너지 및 폐수 품질과 같은 프로세스 출력을 직접적으로 방해하는 지연되거나 바람직하지 않은 제어 성능을 제안한다. 오프라인 RL은 불안정한 온라인 RL과 비교하여 추가 온라인 상호 작용 없이 이전에 수집된 오프라인 데이터만 활용하므로 탐색 시도와 실제 환경 간의 직접적인 상호 작용 위험이 줄어든다. 특히 AWAC(Advantage Weighted Actor Critic)는 수집된 오프라인 데이터에 대한 고성능 정책을 빠르게 학습하고 학습된 성공적인 정책을 기반으로 실제 프로세스와 상호 작용하여 온라인 미세 조정을 수행하는 능력을 가지고 있다. 따라서 수집된 과거 운전 데이터 세트를 기반으로 하는 WWTP용 AWAC 기반 DO 제어 시스템은 극심한 유입수를 포함한 비선형 조건에서 제어 성능을 안정적으로 향상시킬 것이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 등록특허 10-1209944

(특허문헌 0002) 대한민국 등록특허 10-1629240

(특허문헌 0003) 대한민국 등록특허 10-1927503

발명의 내용

해결하려는 과제

[0009] 따라서 본 발명은 상기와 같은 종래의 문제점을 해결하기 위하여 안출된 것으로서, 본 발명의 실시예에 따르면,

인공지능기술 중 오프라인 강화학습 (advantage weighted actor critic, AWAC)을 이용하여 축적된 공정 운영 데이터를 기반으로 스스로 학습하고 제어기 파라미터 값을 결정하는 자동 적응 제어 튜닝 시스템을 제공하는데 그 목적이 있다.

[0010] 본 발명의 실시예에 따르면, 오프라인 강화학습 인공지능 기술을 이용하여 공정 시스템의 PID 제어기의 제어기 파라미터 값을 스스로 결정하고, 기존 공정 제어기 파라미터 값 변화에 따른 에러 값의 수집 데이터에 따라, 오프라인 강화학습 알고리즘은 데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습할 수 있으며, 또한, 본 발명의 자동 적응 제어 튜닝 시스템을 운영 중에 공정 유입 오염물이 크게 증가하는 악조건이 감지되면 온라인 학습을 통해 제어기의 성능을 개선할 수 있는, 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템을 제공하는데 그 목적이 있다.

[0011] 그리고 본 발명의 실시예에 따르면, 국내 하수처리장 데이터를 이용해 검증하여, 제어기의 에러를 19~39% 감소시킬 수 있고, 정확한 제어기 성능을 바탕으로 국내 및 국외 하수처리장 및 환경-화공 제어 시스템 분야에 적용을 기대할 수 있으며, 사업화로는 국내외 하수처리장, 환경-화공 시스템 기업 등에 적용하여 공정 운영 개선 등에 이용될 수 있는, 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템을 제공하는데 그 목적이 있다.

[0012] 한편, 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0013] 본 발명의 제1목적은 하수처리장 공정 운영 제어기 파라미터 튜닝시스템으로서, 상기 하수처리장의 공정운영데이터를 수집, 축적하는 데이터수집부; 및 수집, 축적된 데이터에 따라 오프라인 강화학습을 통해 상기 제어기의 파라미터 값을 결정하는 파라미터 산출부;를 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝시스템으로서 달성될 수 있다.

[0014] 그리고 상기 공정운영데이터는 기존 공정 제어기 파라미터 값 변화에 따른 에러값의 수집데이터인 것을 특징으로 할 수 있다.

[0015] 또한 상기 오프라인 강화학습 알고리즘은 수집데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습하는 것을 특징으로 할 수 있다.

[0016] 그리고 상기 제어기는 PID 제어기이고, 상기 파라미터 산출부는 상기 PID 제어기 3개의 파라미터 값을 자동적으로 튜닝하는 것을 특징으로 할 수 있다.

[0017] 또한 상기 파라미터 산출부를 통해 도출된 파라미터 값으로 상기 PID 제어기가 운영되고, 운영 시, 공정유입-악조건 발생시 온라인 학습을 통해 상기 PED 제어기의 파라미터를 튜닝하는 인공지능 학습모듈;을 더 포함하는 것을 특징으로 할 수 있다.

[0018] 본 발명의 제2목적은 하수처리장 공정 운영 제어기 파라미터 튜닝방법으로서, 데이터수집부가 상기 하수처리장의 공정운영데이터를 수집, 축적하는 단계; 및파라미터 산출부가 수집, 축적된 데이터에 따라 오프라인 강화학습을 통해 상기 제어기의 파라미터 값을 결정하는 단계;를 포함하는 것을 특징으로 하는 강화학습을 이용한 제어기의 자동 적응 제어튜닝방법으로서 달성될 수 있다.

[0019] 그리고 상기 축적단계에서, 상기 공정운영데이터는 기존 공정 제어기 파라미터 값 변화에 따른 에러값의 수집데이터인 것을 특징으로 할 수 있다.

[0020] 또한 상기 결정하는 단계에서, 상기 오프라인 강화학습 알고리즘을 통해 수집데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습하는 것을 특징으로 할 수 있다.

[0021] 그리고 상기 제어기는 PID 제어기이고, 상기 결정하는 단계에서, 상기 파라미터 산출부는 상기 PID 제어기 3개의 파라미터 값을 자동적으로 튜닝하는 것을 특징으로 할 수 있다.

[0022] 또한 상기 결정하는 단계 후에, 상기 파라미터 산출부를 통해 도출된 파라미터 값으로 상기 PID 제어기가 운영되는 단계; 및 운영 시, 공정유입-악조건 발생시 온라인 학습을 통해 상기 PED 제어기의 파라미터를 튜닝하는 단계;을 더 포함하는 것을 특징으로 할 수 있다.

발명의 효과

[0023] 본 발명의 실시예에 따르면, 인공지능기술 중 오프라인 강화학습 (advantage weighted actor critic, AWAC)을 이용하여 축적된 공정 운영 데이터를 기반으로 스스로 학습하고 제어기 파라미터 값을 결정하는 자동 적응 제어 튜닝 시스템을 제공할 수 있다.

[0024] 본 발명의 실시예에 따른 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템에 따르면, 오프라인 강화학습 인공지능 기술을 이용하여 공정 시스템의 PID 제어기의 제어기 파라미터 값을 스스로 결정하고, 기존 공정 제어기 파라미터 값 변화에 따른 에러 값의 수집 데이터에 따라, 오프라인 강화학습 알고리즘은 데이터를 기반으로 제어기의 에러를 줄일 수 있는 파라미터 값을 제시하는 정책을 스스로 학습할 수 있으며, 또한, 본 발명의 자동 적응 제어 튜닝 시스템을 운영 중에 공정 유입 오염물이 크게 증가하는 악조건이 감지되면 온라인 학습을 통해 제어기의 성능을 개선할 수 있는 효과를 갖는다.

[0025] 그리고 본 발명의 실시예에 따른 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝방법 및 시스템에 따르면, 국내 하수처리장 데이터를 이용해 검증하여, 제어기의 에러를 19~39% 감소 시킬 수 있고, 정확한 제어기 성능을 바탕으로 국내 및 국외 하수처리장 및 환경-화공 제어 시스템 분야에 적용을 기대할 수 있으며, 사업화로는 국내외 하수처리장, 환경-화공 시스템 기업 등에 적용하여 공정 운영 개선 등에 이용될 수 있는 효과를 갖는다.

[0026] 한편, 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0027] 본 명세서에 첨부되는 다음의 도면들은 본 발명의 바람직한 실시예를 예시하는 것이며, 발명의 상세한 설명과 함께 본 발명의 기술적 사상을 더욱 이해시키는 역할을 하는 것이므로, 본 발명은 그러한 도면에 기재된 사항에만 한정되어 해석 되어서는 아니 된다.

도 1은 본 발명의 실시예에 따른 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝시스템의 블록도,

도 2는 본 발명의 실시예에 따른 WWTP 제어 시스템을 위한 AWAC 기반 자동 적응 제어기 튜닝 시스템의 도식적 프레임워크,

도 3은 FOPTD 모델의 값을 결정하기 위한 공정 반응 곡선 방법.

도 4는 폭기 공정의 컨트롤러 구성: (a) 기존 PID 컨트롤러 및 (b) 시간 지연 보상기의 Smith 예측기

도 5는 온라인 미세 조정을 통한 오프라인 강화 학습 알고리즘 기반 자동 적응 컨트롤러 조정 시스템의 그래픽 표현,

도 6은 Deep Q 네트워크와 경험 재생 버퍼로 구성된 DQN 알고리즘의 개략도,

도 7은 오프라인 학습 및 온라인 미세 조정을 포함한 AWAC 알고리즘 구조의 개략도,

도 8은 비선형 유입수 조건 및 제어 입력 및 출력을 나타내는 대상 WWTP의 폭기 시스템,

도 9는 Smith 예측자와 상호 작용하는 오프라인 RL 기반 자동 적응 컨트롤러 튜닝 시스템의 그래픽 구조,

도 10은 전문가 데모를 통한 오프라인 데이터 수집 결과: (a) 수렴된 보상을 보여주는 훈련된 전문가 데모, (b) 전환 수집을 위한 유입 시나리오, (c) 상태, 동작 및 보상을 포함한 전환의 정적 데이터 수집,

도 11은 (a) 보상 값의 커널 밀도 및 (b) 보상 추세에 따른 전문가 데모, 오프라인 학습 AWAC 및 온라인 미세 조정 AWAC의 비교,

도 12는 본 발명의 실시예에 따른 제어 시스템의 제어 성능을 평가하기 위한 유입수 조건: (a) 낮음, (b) 정상 및 (c) 높은 유입수 조건

도 13은 (a) 낮음, (b) 정상 및 (c) 높은 유입 조건에서 AWAC 기반 자동 적응 컨트롤러 튜닝 시스템에 의한 폭기 시스템의 성능 제어,

도 14는 (a) 낮음, (b), 보통 및 (c) 높은 유입수 조건에서 AWAC 기반 자동 적응 컨트롤러 튜닝 시스템과 기존

제어 시스템의 비교그래프를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0028] 이상의 본 발명의 목적들, 다른 목적들, 특징들 및 이점들은 첨부된 도면과 관련된 이하의 바람직한 실시예들을 통해서 쉽게 이해될 것이다. 그러나 본 발명은 여기서 설명되는 실시예들에 한정되지 않고 다른 형태로 구체화될 수도 있다. 오히려, 여기서 소개되는 실시예들은 개시된 내용이 철저하고 완전해질 수 있도록 그리고 통상의 기술자에게 본 발명의 사상이 충분히 전달될 수 있도록 하기 위해 제공되는 것이다.

[0029] 본 명세서에서, 어떤 구성요소가 다른 구성요소 상에 있다고 언급되는 경우에 그것은 다른 구성요소 상에 직접 형성될 수 있거나 또는 그들 사이에 제 3의 구성요소가 개재될 수도 있다는 것을 의미한다. 또한 도면들에 있어서, 구성요소들의 두께는 기술적 내용의 효과적인 설명을 위해 과장된 것이다.

[0030] 본 명세서에서 기술하는 실시예들은 본 발명의 이상적인 예시도인 단면도 및/또는 평면도들을 참고하여 설명될 것이다. 도면들에 있어서, 막 및 영역들의 두께는 기술적 내용의 효과적인 설명을 위해 과장된 것이다. 따라서 제조 기술 및/또는 허용 오차 등에 의해 예시도의 형태가 변형될 수 있다. 따라서 본 발명의 실시예들은 도시된 특정 형태로 제한되는 것이 아니라 제조 공정에 따라 생성되는 형태의 변화도 포함하는 것이다. 예를 들면, 직각으로 도시된 영역은 라운드지거나 소정 곡률을 가지는 형태일 수 있다. 따라서 도면에서 예시된 영역들은 속성을 가지며, 도면에서 예시된 영역들의 모양은 소자의 영역의 특정 형태를 예시하기 위한 것이며 발명의 범주를 제한하기 위한 것이 아니다. 본 명세서의 다양한 실시예들에서 제1, 제2 등의 용어가 다양한 구성요소들을 기술하기 위해서 사용되었지만, 이들 구성요소들이 이 같은 용어들에 의해서 한정되어서는 안 된다. 이들 용어들은 단지 어느 구성요소를 다른 구성요소와 구별시키기 위해서 사용되었을 뿐이다. 여기에 설명되고 예시되는 실시예들은 그것의 상보적인 실시예들도 포함한다.

[0031] 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 '포함한다(comprises)' 및/또는 '포함하는(comprising)'은 언급된 구성요소는 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다.

[0032] 아래의 특정 실시예들을 기술하는데 있어서, 여러 가지의 특정적인 내용들은 발명을 더 구체적으로 설명하고 이해를 돕기 위해 작성되었다. 하지만 본 발명을 이해할 수 있을 정도로 이 분야의 지식을 갖고 있는 독자는 이러한 여러 가지의 특정적인 내용들이 없어도 사용될 수 있다는 것을 인지할 수 있다. 어떤 경우에는, 발명을 기술하는 데 있어서 흔히 알려졌으면서 발명과 크게 관련 없는 부분들은 본 발명을 설명하는데 있어 별 이유 없이 혼돈이 오는 것을 막기 위해 기술하지 않음을 미리 언급해 둔다.

[0034] 이하에서는 본 발명의 실시예에 따른 오프라인 강화 학습의 데이터 기반 스마트 결정을 사용하여 폐수 처리 프로세스를 위한 자동 적응 컨트롤러 튜닝 시스템 및 튜닝방법에 대해 설명하도록 한다. 도 1은 본 발명의 실시예에 따른 강화학습을 이용한 PID 제어기의 자동적응 제어튜닝시스템의 블록도를 도시한 것이다.

[0035] 본 발명의 실시예에서는 여러 비선형 유입수 조건에서 새로운 오프라인 RL 기반 DO 제어 시스템을 제안한다. 먼저, 시스템 식별 기술을 사용하여 호기성 반응기에서 폭기 유량과 DO 농도 사이의 관계를 나타내는 공정 모델을 개발하였다. 다음으로 폭기 시스템에 대한 시간 지연 효과를 고려한 PID 제어기와 Smith 예측기를 사용하여 폭기 제어 전략을 설계한다. 셋째, PID 제어기의 튜닝 매개변수를 적응적으로 제안하기 위해 전문 데모 알고리즘과 비선형 유입수 조건 시나리오를 통해 오프라인 데이터를 수집한다. 넷째, AWAC 알고리즘을 구현하여 실제 유입수 조건에서 세 가지 조정 매개변수를 제공하여 호기성 반응기의 DO 농도를 자동으로 제어한다. 마지막으로 제안된 완전 데이터 구동 제어 시스템의 제어 성능을 기존 PID 제어기 및 스미스 예측기와 오차 측정 제어 성능 기준 측면에서 비교한다.

[0037] 먼저, 컨트롤러의 적응 튜닝을 위해 제안된 프레임워크는 도 2에 도시되어 있다. 첫 번째 단계는 공정 역학을 설명하고 공정 컨트롤러를 구현하기 위한 폭기 시스템을 식별하는 것이다. 낮은 차수 플러스 시간 지연 프로세스는 프로세스의 동적 특성을 설명하는 데 유용한 도구이기 때문에 사용되었다. 다음은 PID(Proportional-Integral-derivative) 컨트롤러와 같은 컨트롤러 구조의 선택이다. 또한 본 발명의 실시예에서는 프로세스의 시간 지연 문제를 고려하여 제어 성능을 향상시키기 위해 Smith 예측기를 사용하였다. 세 번째 단계는 채택된

Smith 예측기에 대한 자동 적응 컨트롤러 튜닝 시스템의 구현이다. 제안된 튜닝 시스템은 Smith 예측 변수의 제어기 상수 값을 자동으로 제공한다. 이 단계에서는 자동 적응 컨트롤러 튜닝 시스템에 AWAC(Advanced Weighted Actor Critic) 알고리즘이 사용ㄷ된. 이 알고리즘은 데이터 기반 방법으로 오프라인 데이터에서 사전 학습하고 온라인 데이터 수집으로 개선할 수 있으므로 실제 문제에 효과적으로 적용할 수 있다. 마지막 단계는 제안된 자동 적응 컨트롤러 튜닝 시스템에 대한 평가이다. 호기성 반응기에서 용존 산소(DO) 농도의 제어 성능은 (1) 시간 가중 절대값 오차(ITAE) 조정 규칙의 적분을 사용하는 기존 PID 제어기, (2) ITAE 조정이 있는 스미스 예측기, (3) AWAC 기반 자동 튜닝 시스템이 있는 Smith 예측기를 위해 평가되었다.

[0039] FOPTD(first-order plus time-delay) 모델은 프로세스 엔지니어링 분야에서 프로세스 컨트롤러를 설계하고 사용하는 데 널리 사용되었다. 프로세스의 동적 추세를 설명하는 데 이점이 있다. FOPTD 모델은 하수처리장(WWTP)의 정확한 동적 모델을 설명하고 얻는 데에도 효과적이다. 시스템 식별을 통해 물리적 및 생물학적 프로세스에 대한 인식 없이 광범위한 유압 시스템 역학을 처리할 수 있다. 따라서 폭기 시스템을 설명하고 제어하기 위해 폭기 시스템에 대한 프로세스 식별을 수행했다.

[0040] 호기성 반응기의 DO 농도는 폭기 시스템에 의한 공기 유량과 기체에서 물로의 산소 전달 효율에 따라 달라진다. 따라서 압축기에서 폭기 장치로의 공기 유량은 수학식 1로 표현된 반응기의 산소 전달 속도를 설명하는 데 사용할 수 있다.

[0041] [수학식 1]

[0042] $$K_L a = k_1 \cdot \left(1 - e^{k_2 \cdot q\_air(t)}\right)$$

[0043] 여기서 $k_1(d^{-1})$ 및 $k_2(d/m^3)$는 대상 식물과 관련된 경험적 상수이고, q_air(t)는 공기 유량(m3/d), $K_L a$는 산소 전달 속도(d^{-1})입니다. 수학식 1은 공기 유량이 증가할 때 전달 속도가 어떻게 계산되는지 설명한다. 호기성 반응기의 DO 농도 변화는 폭기 시스템(수학식 1)을 통한 공기 추가에 의해 고려될 수 있으며, 그 관계는 활성 슬러지 모델 가용성 미생물 생성물(ASM-SMP) 모델에서 다음 식으로 설명된다.

[0044] [수학식 2]

[0045] $$K_L a\left(S_{O.at} - S_O\right)$$

[0046] 여기서 $S_{0.sat}$는 수중 산소 포화 농도이고 S_0는 호기성 반응기의 DO 농도이다.

[0047] 호기성 반응기의 기류율과 DO 농도의 관계는 수학식 3과 같이 표현되는 FOTPD를 이용하여 확인하였다.

[0048] [수학식 3]

[0049] $$G(s) = \frac{y(s)}{u(s)} = \frac{k \cdot \exp(-\theta s)}{\tau s + 1}$$

[0050] 여기서 G(s)는 공정 입력 u(s)와 y(s) 간의 관계를 나타내는 라플라스 변환 전달 함수이고, u(s)는 폭기 과정에 의한 공기 유량을 나타내고, y(s)는 호기성 산소 농도 리액터, k, τ, Θ는 각각 정적 이득, 시간 상수 및 시간 지연이다. FOPTD 모델은 단계 입력 테스트에서 공정 반응 곡선(PRC) 방법으로 얻었다. 단계 입력 u는 프로세스 출력 y가 정상 상태 값이 될 때까지 대상 프로세스에 입력된다. 그런 다음 도 3와 같이 시간 상수 τ와 시간 지연 Θ를 결정합니다. 정적 이득 k는 설정된 입력에 대한 프로세스 출력의 비율로 추정된다. PRC는 폭기 과정에 대한 정확한 FOPTD 모델을 식별하는 데 도움이 될 수 있다.

[0052] 프로세스 제어는 프로세스 입력을 자동 방식으로 조작하여 프로세스 출력이 원하는 방식으로 작동하도록 한다. WWTP에 대한 안정적인 운영 체제의 전제는 합리적인 범위 내에서 설정값에 따라 공정 입력을 조작하는 효율성에

달려 있다. WWTP 운영 시스템에서 대부분의 시스템은 일반적으로 선형 제어 기술에 의해 제어된다. 선형 컨트롤러는 레귤레이터가 교정기 법칙에 따라 원하는 설정값으로 액츄에이터를 조작하는 단순화된 구조와 저비용 제어 작동을 보장하기 때문이다. 따라서 본 발명의 실시예에서는 도 4와 같이 선형 제어 시스템인 PID 컨트롤러)과 Smith 예측기를 사용하여 WWTP의 복잡한 비선형 조건에서 원하는 DO 설정값을 달성하기 위한 폭기 시스템을 조절했다.

[0053] WWTP 연구 분야에서 가장 널리 사용되는 제어 알고리즘은 PID이다. 이 방법은 수식식 4 내지 7에 표시된 대로 측정된 출력과 원하는 설정값 간의 차이인 오차의 함수로 프로세스 입력을 조작한다.

[0054] [수학식 4]

[0055] $$\text{Proportional (P) part}: \; u_p(t) = k_c(y_s(t) - y(t))$$

[0056] [수학식 5]

[0057] $$\text{Integral (I) part}: \; u_I(t) = \frac{k_c}{\tau}\int_0^t (y_s(\tau) - y(\tau))d\tau$$

[0058] [수학식 6]

[0059] $$\text{Derivative (D) part}: \; u_D(t) = k_c\tau_d \frac{d(y_s(t) - y(t))}{dt}$$

[0060] [수학식 7]

[0061] $$u(t) = u_P(t) + u_I(t) + u_D(t)$$
$$= k_c(y_s(t) - y(t)) + \frac{k_c}{\tau_i}\int_0^t (y_s(\tau) - y(\tau))d\tau + k_c\tau_d \frac{d(y_s(t) - y(t))}{dt}$$

[0062] 여기서 $y_s(t)$는 설정값, y(t)는 프로세스 출력, u(t)는 프로세스 입력(PID 컨트롤러의 제어 출력), k_c는 비례 이득, τ_i는 적분 시간, τ_d는 미분 시간이다. PID 제어기는 도 4(a)와 같이 입력이 $y_s(t)$-y(t)이고 출력이 u(t)인 단순한 기능이다. 세 가지 조정 매개변수 k_c, τ_i 및 τ_d는 프로세스의 역학을 고려하여 적절하게 설정되어야 한다.

[0063] [수학식 8]

[0064] $$k \cdot k_c = 0.965\left(\theta / \tau\right)^{-0.850}$$

[0065] [수학식 9]

[0066] $$\tau / \tau_i = 0.796 - 0.1465\left(\theta / \tau\right)$$

[0067] [수학식 10]

[0068] $$\tau_d / \tau = 0.308\left(\theta / \tau\right)^{0.929}$$

[0069] 폭기 제어 시스템은 폭기 역학 및 DO 포화 단계에 따라 시간 지연이 발생하므로 폭기 제어의 응답 시간은 약

5~30분이다. 따라서 피드백 제어 루프(PID)는 이러한 가변적인 시간 지연으로 인해 강력한 성능을 제공하는 데 한계가 있다. 일정한 시간 지연으로 제한된 제어 성능을 극복하기 위해 Smith 예측기는 제어 문제에서 시간 지연 효과를 처리하는 솔루션이다. Smith 예측기는 시간 지연 보상기 중 하나로, 시간 지연을 기다리지 않고 미래 프로세스 출력에 대한 현재 제어 조치의 영향을 예측한다. Smith 예측기의 목적은 시간 지연을 제거하고 지연 없는 전달 함수를 얻는 것이다.

[0070] 스미스 예측기는 도 4(b)와 같이 시간지연 없는 모델($G_m^*(s)$), 시간지연이 있는 과정(폭기과정), 시간지연($\exp(-\Theta s)$), PID 제어기($G_C(s)$)의 네 부분으로 구성된다. 스미스 예측기(Smith predictor) PID 컨트롤러가 피드백 루프에서 시간 지연 없이 조정될 수 있는 구조로, 결과적으로 빠른 폐쇄 루프 응답이 나타난다. 실제로 $y(s)$와 $y_m(s)$ 간의 모델링 오류를 고려하기 위해 모델링 오류를 보상하기 위한 피드백 루프가 포함된다. 모델링 오차로 설정값을 감소시킨다.

[0072] 선형 제어 시스템은 복잡하지 않은 구조와 저비용 피드백 제어를 통해 우수한 제어 성능을 제공할 수 있는 능력이 있다. 그러나 프로세스 역학이 크게 변경되면 컨트롤러가 부정확하고 성능이 저하된다. 폭기 시스템 제어에서 직면하는 주요 어려움은 동적 및 비선형 유입수 조건이다. 따라서 유입 교란에 해당하는 가능한 제어 오류를 수용하기 위해 컨트롤러를 보수적으로 조정할 필요가 있다. 적응형 컨트롤러는 시간 지연 또는 프로세스 역학이 크게 다를 때 만족스러운 성능을 달성하기 위한 핵심 역량이 될 수 있다.

[0073] 적응 제어는 다양한 프로세스 조건을 보상하기 위해 컨트롤러 매개변수를 자동으로 조정한다. k_c, τ_i 및 τ_d와 같은 컨트롤러 매개변수는 새 데이터가 측정됨에 따라 업데이트되고, 업데이트된 조건을 기반으로 제어 계산이 수행된다. 실시간 매개변수 튜닝은 모델 매개변수를 정확하게 추정하기 위해 때때로 외부 알고리즘을 도입해야 한다. 이러한 맥락에서 강화 학습(RL)은 비선형 시스템에 대한 최적의 컨트롤러를 찾는 강력한 알고리즘이다. RL 알고리즘의 에이전트는 제어 시스템과 상호 작용하는 휴리스틱 및 모델 없는 학습을 통해 보상을 최대화하며, 이는 제어 비용의 최소화와 동일하다. 그러나 온라인 RL 접근 방식은 실시간 측정을 수집하기 위해 제어 시스템과의 지속적인 상호 작용이 필요하다. 이는 안전 및 잠재적으로 위험한 애플리케이션에 대한 몇 가지 문제를 야기한다.

[0074] 따라서 본 발명의 실시예에서는 도 5와 같이 폭기 시스템의 적응형 제어기에 대한 튜닝 매개변수를 제안하기 위해 오프라인 RL 알고리즘을 사용하였다. 오프라인 RL은 수집된 데이터에서 최대한 활용 가능한 주요 정책을 추출할 수 있으므로 의사 결정 문제를 자동화할 수 있다. 데이터 기반 오프라인 RL 알고리즘을 튜닝 시스템에 적용하기 위해 상태, 동작 및 관련 보상의 궤적으로 구성된 오프라인 데이터 세트를 수집했다. 오프라인 데이터 세트는 원하는 작업, 덜 최적의 정책 및 환경과의 부분적 탐색-이용 상호 작용에 대한 데모가 될 수 있다. 효과적인 데이터 기반 RL의 사용은 오프라인 데이터의 품질에 달려 있다. 오프라인 RL은 데이터를 사용하여 오프라인으로 사전 훈련하는 동시에 온라인 미세 조정으로 성능을 향상시키기 때문이다. 따라서 이전 공개된 연구의 권장 사항에 따라 본 발명의 실시예에서는 도 5a에 도시된 바와 같이 훈련된 DQN이 전문가 데모라는 가정 하에 전환을 위해 3개의 에이전트가 있는 심층 Q 네트워크(DQN)를 사용했다. 또한 폭기 시스템의 비선형 거동을 반영하기 위해 다양한 유입수 조건이 생성되었다. 본 발명의 실시예에서는 k-means 클러스터링 알고리즘을 이용한 데이터 생성을 활용하였다.

[0075] 도 5(b)는 AWAC(Advantage Weighted Actor Critic) 알고리즘인 오프라인 RL의 구조를 나타낸다. AWAC는 기존 오프라인 RL과 비교하여 다양한 작업 정책 기능 분포를 가진 다중 모드 환경에서 성능을 능가할 수 있다. AWAC의 세 에이전트는 DQN의 동작을 기반으로 적응 제어 구조의 오류를 줄이기 위해 컨트롤러 매개 변수를 제안하는 정책을 배웠다. 사전 훈련 절차 후 실제 측정된 유입수 데이터를 이용하여 AWAC 알고리즘의 온라인 미세 조정을 수행했다.

[0077] DQN 알고리즘은 훈련 과정에서 경험 재생 버퍼에 지속적으로 경험을 저장하고 환경과 상호 작용할 때 버퍼에서 저장된 경험을 무작위로 샘플링한다. 경험 버퍼와 훈련된 Q-함수의 지속적인 변경은 개선된 목표를 선험적으로 계산할 수 없음을 나타내므로 복사된 목표 Q-함수의 매개변수는 메모리에 유지된다. Q-값의 훈련은 목표 Q 함수의 이전 매개변수를 사용하는 Q-함수 네트워크의 복사된 버전인 목표 네트워크에 의해 최적화된다. 목표 Q 함수의 매개변수는 모든 훈련 에포크에서 Q 함수의 현재 매개변수와 동일하도록 업데이트된다. DQN 알고리즘은 로봇

제어, 기계 제어, 교통 제어 및 차량 제어에서 우수한 성능을 보였다.

[0078] DQN 알고리즘은 도 5(a)와 도 6과 같이 상태 s, 액션 a, 다음 상태 s` 및 보상 r을 포함하는 전이 D = (s,a,s`,r)의 정적 데이터 세트를 수집하는 데 사용되었다. 본 발명의 실시예에서는 DQN 실험이 프로세스 운영자의 경험을 반영할 수 있는 전문가 데모를 대표한다고 가정했다. 컨트롤러 시스템에 대한 측정 데이터가 부족하고 컨트롤러 매개변수가 실제 시스템에서 변경될 때의 위험 때문에 본 발명의 실시예에서는 DQN을 전문 지식으로 가정했다. 또한 DQN은 WWTP의 비선형 및 동적 작동 조건을 반영하기 위해 많은 유입수 시나리오를 포함했다.

[0079] AWAC 알고리즘은 정책 외 비평가와 묵시적 정책 제약이 있는 행위자를 학습하고 훈련한다. AWAC는 정책 평가 단계에서 정책 $\pi(Q^{\pi})$를 사용하여 행동 가치 Q 함수를 기울이고 정책 개선 단계 π를 업데이트하는 행위자 비평가 알고리즘의 표준 패러다임을 사용한다. AWAC는 정책 외 시간차 학습 방법을 사용하여 Q를 계산합니다. 정책 평가 단계에서 정책을 개선한다. 이는 온라인 미세 조정 절차로 성능을 향상시키면서 이전 데이터 세트에서 오프라인 교육이 진행될 때 오프라인 RL 알고리즘의 이점을 쉽게 얻을 수 있다.

[0080] 정책평가 단계에서 학습한 비평가의 가치를 극대화하는 정책을 학습하여 정책개선을 수행한다. AWAC는 모든 반복 $k(Q^{\pi k}(s,a))$에서 추정된 Q-함수(비계)를 증가시키도록 정책을 최적화한다. 한편, 전문가 시연을 나타내는 Bellman 연산자(B)에 해당하는 이전 데이터 세트(D)에서 관찰된 동작에 근접하는 정책을 제한한다. 여기서 제약 조건은 모델의 동작을 명시적으로 학습하지 않고 암시적으로 적용된다. 정책 개선 중 액터 π는 critic Q^{π}를 통해 업데이트된다. 액터 업데이트는 지도 학습의 가중 최대 가능성 방법 중 하나이다. 여기서 목표는 학습된 critic Q로부터 예측된 이점에 의해 데이터 세트에서 관찰된 상태-동작 쌍에 의해 계산되며, 단순히 버퍼에서 상태와 동작을 샘플링한다.

[0081] 요약하자면, AWAC는 동적 프로그래밍 방법을 사용하여 critic Q_{ϕ}의 매개변수를 훈련하고 감독 학습을 사용하여 제약 있는 actor π를 훈련한다. AWAC 알고리즘에는 최고 감독 학습과 배우 critic 알고리즘의 결합 구조가 채택되었다. 동적 프로그래밍은 정책 외 데이터를 활용하며 효율적인 샘플 학습이 가능하다. 간단한 감독 액터 업데이트는 오프라인 데이터 세트에서 학습할 때 이상값 작업의 영향을 완화하는 제약 조건을 암시적으로 촉진한다. AWAC의 그래픽 구조는 도 7과 같다.

[0083] 본 발명의 실시예에서는 Y-city에 위치한 WWTP의 폭기시스템에 대한 Smith 예측변수의 튜닝 파라미터를 제안하는 것을 목적으로 하였다. 도 8은 양분 제거의 질산화 조건을 제공하기 위한 호기 반응기에서 호기 조건을 유지하기 위한 목표 폭기 시스템을 보여준다. 질화의 성공적인 성능은 DO 농도가 합리적인 오류 경계에서 얼마나 효과적으로 제어되는지에 달려 있다. 비선형 유입수 조건은 실제 WWTP에서 DO 농도를 제어하는 데 장애가 된다.

[0084] 따라서 호기성 반응기에서 적절한 DO 농도를 유지하려면 강력한 컨트롤러가 필요하다. 기존 PID 컨트롤러는 목표 WWTP의 설정값에 따라 DO 농도를 유지하기 위해 사용하는 것으로 가정했다. 폭기유량, 산소전달율, DO 농도의 관계는 측정된 데이터와 기계적 특성을 이용하여 실증적으로 계산하였다. 본 발명의 실시예에서는 수학식 1에서 k_1과 k_2로 각각 400/d와 -0.01 min/m^3를 사용하였다. 계산된 산소 전달 속도 값은 보정된 활성 슬러지 모델 및 가용성 미생물 제품(ASM-SMP) 모델의 입력으로 사용되었다. 그런 다음 포화 DO 농도, 오염물질 성분 및 미생물을 고려하여 호기성 반응기에서의 DO 농도를 추정하였다. 따라서 폭기 유량은 공정 입력(즉, 컨트롤러의 출력)이었고 호기성 반응기의 DO 농도는 공정 출력이었다.

[0086] 폭기 시스템의 역학은 공정을 구성하는 비선형 요소에 대한 호기성 반응기의 DO 농도를 설명하기 위해 전달 함수의 형태로 표현되었다. 수학적 모델은 호기성 반응기의 DO 농도를 폭기 유량과 같은 폭기 시스템에 영향을 미치는 요인과 관련시켰다. 식별된 정보는 상대적으로 정확도가 낮더라도 프로세스 모델에 대해 사용할 수 있다. 프로세스 모델의 적용은 모델 프리 RL 알고리즘에 도메인 지식을 제공할 수 있기 때문에 RL이 대상 프로세스를 더 빠르고 쉽게 학습하는 데 도움이 된다.

[0087] 보정된 활성슬러지 모델과 용해성 미생물 생성물(ASM-SMP) 모델을 사용하여 동정을 실험하였다. ASM-SMP 모델의 입력은 유입수 COD 362 mg/L, 유입수 TN 46 mg/L, 유입유량 14,000 m^3/day인 정상상태 평균 유입수 조건으로 설

정하였다. 폭기 유량은 단계 응답을 제공하기 위해 50에서 55m^3/day로 증가되었다. 입력 및 출력의 단계 응답을 포함하여 획득한 데이터를 사용하여 PRC 방법을 사용하여 FOPTD 모델을 식별했다. 폭기 유량과 DO 농도를 특징으로 하는 폭기 시스템은 수학식 11에 추정 및 주어졌다.

[0088] [수학식 11]

$$G_m(s) = \frac{0.14378 \cdot \exp(-0.002523s)}{0.008259s + 1}$$

[0089]

[0090] 공정 게인(0.14378)의 양수 값은 더 높은 폭기 유량이 호기성 반응기의 DO 설정값을 증가시켰음을 나타낸다. 시간 지연은 폭기 시스템이 제어 신호가 발행된 후 DO 농도의 첫 번째 변화를 일으키는 데 필요한 시간을 보여준다. 확인된 시간지연은 3.63분으로 폭기제어시스템의 일반적인 시간지연 값에 포함된 값이다. 시간 상수는 제어 출력 값이 변경될 때 프로세스 변수가 설정값으로 변경되는 속도를 나타낸다. 여기서 시상수는 11.89분으로 추정되었다. 결과적으로, 식별된 FOPTD 모델은 폭기 프로세스가 비선형 유입수 및 복잡한 작동 조건에서 느리고 불안정한 폭기 시스템 작동에 함정이 될 시간 상수와 시간 지연을 가지고 있음을 보여주었다. 이러한 맥락에서 시간 지연 문제를 해결하기 위한 시간 지연 보상기와 비선형 조건을 반영하는 적응 제어는 폭기 시스템에 효과적이고 안정적인 솔루션이 될 것이다.

[0091] 표 1은 기존의 PID 제어기와 확인된 폭기 시스템에 따른 Smith 예측기를 포함한 제어 시스템의 사양을 요약한 것이다. PID 컨트롤러는 폭기 유량을 조작하여 호기성 반응기에서 설정값과 DO 농도 사이의 차이를 최소화하기 위해 사용되었다. ITAE-1 튜닝 매개변수는 PID 제어기와 Smith 예측기의 성능을 향상시키는 데 사용되었다. 튜닝 매개변수의 양수 값은 DO 농도가 원하는 설정값보다 낮을 때 컨트롤러에 의한 증가된 긍정적인 동작이 설정값을 향해 증가하는 경향이 있음을 나타낸다. 비례 게인 K_P의 절대량은 시간이 지남에 따라 오차를 줄이려고 한다. 적분 이득은 시간에 따른 오차의 합을 정의한다. 따라서 식별된 시스템이 DO 농도를 변경하는 데 걸리는 시간 상수가 11.89분이기 때문에 K_I 값이 높으면 설정값 값에 빠르게 도달하려는 상당한 제어 조치가 생성된다. 그러나 과도한 양의 K_I는 진동하고 불안정한 응답을 초래하여 폭기 유량의 변동이 큰 폭기 시스템이 제대로 제어되지 않을 수 있다. 미분 이득(K_D)은 수학식 6과 같이 오차율을 기준으로 하며 오차가 커짐에 따라 제어기가 더 빠르게 반응하도록 한다. 그러나 K_D 값이 높으면 고주파 진동이 발생한다. 따라서, 튜닝 매개변수의 결정은 안정적이고 효과적인 제어 성능을 보장하기 위한 제어 시스템의 중요한 문제이며, 비선형 유입수 조건에 따라 WWTP에 대한 적응 제어 시스템이 필요하다.

[0092] [표 1] ITAE-1 튜닝 방식의 튜닝 파라미터를 포함한 PID 제어기 및 스미스 예측 사양

	Parameters and equations	Descriptions
PID controller	$K_p(k_c)$: 12.3902	Proportional gain
	$K_I(k_c/\tau_i)$: $1.1270 \cdot 10^3$	Integral gain
	$K_D(k_c \cdot \tau_d)$: 0.0105	Derivative gain
Smith predictor	PID controller (K_p : 12.3902, K_I : $1.1270 \cdot 10^3$, and K_D : 0.0105)	ITAE-1-tuned PID controller
	$G_m^*(s) = \frac{0.14378}{0.008259s + 1}$	Time-delay-free model
	$\exp(-0.002523s)$	Time-delay in the model

[0093]

[0094] RL 알고리즘의 구조는 시간 단위로 Smith 예측자에 대한 최적의 튜닝 매개변수를 적응적으로 제안하도록 결정되었다. 튜닝 파라미터의 빈번한 변경은 불안정하고 주의를 요하는 제어를 요하므로, 본 발명의 실시예에서는 시간당 튜닝 파라미터를 제안하였다. 또한, 본 발명의 실시예에서는 센서 샘플링-기록 시간, 기계적 제어기 작동 등 WWTP의 동작 특성을 고려하여 시뮬레이션 샘플링 시간은 5분, 제어 샘플링 시간은 15분으로 가정하였다.

[0095] Smith 예측기에서 PID 컨트롤러의 k_c, τ_i 및 τ_d와 같은 매개변수를 시간당 조정하기 위해 3개의 에이전트가 사

용되었다. 단일 에이전트 RL은 극단적이고 높은 차원의 관절 행동 공간으로 인해 효율적으로 행동을 제공하는 데 한계가 있다. 따라서 다중 에이전트 RL 알고리즘 구조를 사용하여 자동 적응 컨트롤러 튜닝 시스템의 계산 효율성을 높이고 확장성을 보장했다. 보상 인터프리터의 구조는 에이전트에게 공동 보상 값을 제공하고 중앙 집중식 교육을 위해 사용되었다. 다음은 RL 알고리즘의 세부 구조와 컨트롤러 성능 향상을 위한 추가 기술에 대해 설명한다.

[0096] - States:

[0097] 세 에이전트에게 동시에 주어진 상태(들)에는 다음 사이의 오류가 포함되었습니다. 3개의 에이전트에 동시에 주어진 상태(s)는 1시간 전의 DO 설정값과 $e(t)$, $\Delta e(t)$, $\Delta^2 e(t)$, and $\int e(t)$와 같은 제어 출력 사이의 오차를 포함했다. 1시간 동안의 제어출력 오차, error의 1차 차분($\Delta e(t) = e(t) - e(t-1)$), error의 2차 차분($\Delta^2 e(t) = e(t) - 2e(t-1) + e(t-2)$), error의 적분($\int e(t) = e(t-3) + e(t-2) + e(t-1) + e(t)$)을 나타낸다.시간당 컨트롤러의 성능을 나타내는 총 10개의 오류 값을 RL 알고리즘의 3개 에이전트의 상태로 사용했다.

[0098] - Actions:

[0099] 개별 증가 또는 감소 조정 매개변수로 구성된 조치 세트가 RL 에이전트에 사용되었다. RL 기반의 튜닝 파라미터 변경은 ITAE-1 튜닝 규칙에 의해 튜닝된 파라미터로부터 시작되었으며, 여기서 k_c는 12.3902, τ_i는 0.0110, τ_d는 0.0008이다. 또한 1시간 전 이전 오류가 0.1 이상이었다는 제약 조건에서 에이전트를 활성화하여 튜닝 매개변수를 제안하고, 그렇지 않으면 매개변수를 동일한 값으로 유지했다. 튜닝 비례 게인(kc) 에이전트, 적분 시간(τ_i), 미분 시간τ_d)의 동작 세트는 각각 [-2, -1, 0, +1, +2], [-0.002, -0.001, 0, + 0.001 및 +0.002] 및 [-0.0002, -0.0001, 0, +0.0001 및 +0.0002]이다.

[0100] - Rewards:

[0101] 보상 해석기의 기능은 제안된 자동 적응 컨트롤러 튜닝 시스템의 목표에 따라 결정되었다. 3개의 에이전트는 보상 해석기를 통해 보상 값을 최대화하면서 제어 오류를 줄이려고 했다. 따라서 기존의 제어기를 개선하기 위해 공동 보상 해석기의 기능에서 제어 오류와 제어 성능의 변화를 동시에 고려하였다. 공동 보상 인터프리터는 수학식 12로 표현된다.

[0102] [수학식 12]

$$r(t) = \int (\omega_1 \cdot r_1(t) + \omega_2 \cdot r_2(t))$$

[0103]
$$r_t(t) = \begin{cases} +1 & |e(t)| \le 0.25 \\ -1 & 0.25 < |e(t)| \le 0.5 \\ -2 & 0.5 \le |e(t)| \end{cases}, \quad r_2(t) = \begin{cases} +1 & |e(t)| < |e(t-1)| \\ -1 & otherwise \end{cases}$$

[0104] 여기서 r(t)는 보상 해석기의 시간당 보상 값, $r_1(t)$는 제어 오류의 보상, $r_2(t)$는 제어 오류의 변화, $|e(t)|$는 제어 오류의 절대 값, ω_1 및 ω_2는 각각 $r_1(t)$ 및 $r_2(t)$(ω_1=0.4 및 ω_2=0.6)에 대한 가중치 인자아다. 제어오차의 변화에 대한 더 높은 가중치는 설정값의 변화에 빠르게 반응하면서 과감쇠된 응답을 방지하기 위해 결정되었다.

[0105] - Anti-windup:

[0106] PID 제어기의 적분 부분이 급격히 증가하고 큰 오버슈트의 발생이 불가피한 적분 와인드업을 방지하기 위해 안티 와인드업 기술이 활용되었다. 제어 출력이 프로세스의 포화 값(또는 하한 상한 값)에 위치할 때 적분 부분이

크게 증가하면 제어 성능이 느려진다. 따라서 공정(폭기 액츄에이터)이 한계값에 있을 때 적분 제어 입력을 제한하여 적분 권취를 방지할 필요가 있습니다. 사용된 Anti-windup은 수학식 13과 같다.

[0107] [수학식 13]

[0109] 여기서 u_{max}는 목표 폭기 시스템을 고려한 최대 폭기 유량 150m^3/d, u_{min}은 최소 폭기 유량 30m^3/d이다.

[0110] 폭기 시스템에 대해 시간 단위로 조정 매개변수를 제공하기 위한 RL의 구조는 표 2에 요약되어 있고 도 9에는 그래픽으로 표시되어 있다.

[0111] [표 2] 튜닝 매개변수 제안을 위한 전문가 데모 RL 및 오프라인 RL 알고리즘의 구조

Structures	Descriptions
Agent 1 (P part)	- Action: [-2, -1, 0, +1, and +2] for proportional gain (k_c)
Agent 2 (I part)	- Action: [-0.002, -0.001, 0, +0.001, and +0.002] for integral time (τ_I)
Agent 3 (D part)	- Action: [-0.0002, -0.0001, 0, +0.0001, and +0.0002] for derivative time (τ_d)
State	$e(t)$, $\Delta e(t)$, $\Delta^2 e(t)$, and $\int e(t)$ for the previous one hour
Reward	$r(t) = \int (\omega_1 \cdot r_1(t) + \omega_2 \cdot r_2(t))$ $r_1(t) = \begin{cases} +1 & \lvert e(t)\rvert \le 0.25 \\ -1 & 0.25 < \lvert e(t)\rvert \le 0.5 \\ -2 & 0.5 \le \lvert e(t)\rvert \end{cases}$, $r_2(t) = \begin{cases} +1 & \lvert e(t)\rvert < \lvert e(t-1)\rvert \\ -1 & otherwise \end{cases}$
Anti-windup	If $u(t) > u_{max}$, $u_I(t) = u_I(t-1)$ (no update) *and* $u(t) = u_{max}$ If $u(t) < u_{min}$, $u_I(t) = u_I(t-1)$ (no update) *and* $u(t) = u_{min}$
DQN (Expert demo)	- Q networks: 10-64-32-5
AWAC (Offline RL)	- Actors: 10-64-32-5 - Critics: 10-64-64-5

[0112]

[0113] 3개의 에이전트로 구성된 DQN 알고리즘은 메모리 수집을 위한 전문가 데모로 사용되도록 훈련되었다. 오프라인 RL 알고리즘이 환경과 상호 작용하고 해당 정책을 사용하여 전환을 수집하는 것은 불가능하기 때문이다. 오프라인 RL의 학습 알고리즘을 활성화하려면 전환의 정적 데이터 세트가 필요하다. 따라서 Nair et al의 권장 사항에 따라 폭기 제어 시스템을 사용한 이전 RL 실험과 상호 작용하는 전이의 정적 데이터 D=(s,a,s`,r)가 훈련된 DQN 알고리즘을 통해 수집되었다.

[0114] 도 10은 유입 시나리오에 따른 전문가 데모(DQN 알고리즘)를 이용하여 수집된 오프라인 데이터셋의 결과를 보여준다. DQN 알고리즘은 도 10(a)와 같이 오류를 최소화하고 컨트롤러 성능을 향상시키기 위해 10,000 epoch 동안 훈련되었다. 도 10(b)는 COD 및 TN 조성비에 따른 높음(고)-정상-낮음(저) 유입 시나리오를 보여준다. 목표 WWTP의 비선형 및 동적 유입수 조건을 반영하기 위해 모든 훈련 시기에 다양하고 다른 유입수 시나리오를 사용했다. 음수 값은 전문가 데모 교육 절차의 시작 부분에서 튜닝 매개변수를 제안할 때 낮은 성능을 나타낸다. 최적의 정책 탐색을 위한 탐색이 탐색보다 높기 때문에 해당 절차에서 임의의 조치를 취했다. Epoch 횟수가 증가함에 따라 전문가 데모 알고리즘의 평균 보상은 음수 값에서 양수 보상 값으로 증가했다. 평균 보상 값은 더 높은 보상 범위로 수렴되었다. 전문가 데모 알고리즘의 훈련 절차가 끝날 때 훈련된 매개변수(도 10(a)에서 강조 표시된 상자)를 적용하여 제어 시스템의 정보를 포함하는 전환의 정적 데이터 세트를 수집했다. 또한, 본 발명의 실시예에서는 전문가 데모가 0.3의 엡실론 값을 부여하여 차선책 정책을 가지고 있다고 가정하므로 전문가 데모는 30% exploration 및 70% greedy action(exploitation)으로 조정 매개변수를 조정했다.

[0115] 도 10(c)는 전문가 데모 알고리즘에서 수집된 전환 데이터 세트를 나타낸다. 상태, 행동 및 보상을 포함하여 수

집된 데이터 세트는 변형을 분석하기 위해 0에서 1로 정규화되었다. 오프라인 데이터셋에서 다양한 상태값을 수집했는데, 밝은 색(높은 오차)과 어두운 색(낮은 오차)으로 다양한 오차가 관찰되었다. 훈련된 전문가 데모 RL의 결정에 따라 튜닝 매개변수(밝은 색)의 증가와 감소(어두운 색)에 의해 동작이 메모리에 저장되었다. 수집된 보상 값은 튜닝 매개변수에 의한 개선된 성능에 대해 전문가 데모 RL에게 주어졌다.

[0116] 더 높은 보상과 더 낮은 보상을 가진 오프라인 데이터 세트는 도 10(c)와 같다. 행동의 현저한 경향은 낮은 성과와 높은 성과에 따라 감지되었다. 낮은 보상 값은 튜닝 파라미터 값을 줄이는 행동과 튜닝 파라미터 중 하나의 파라미터가 편향적으로 증가했기 때문이다. 낮은 성능에서 모든 튜닝 매개변수의 증가는 제어 출력이 크고 고주파수 진동을 만들고 오랫동안 설정값 아래에 머물게 한다. 따라서 제어 시스템의 교란이 발생할 수 있는 비선형 유입수 조건에서 자동 적응 컨트롤러 튜닝 시스템에 의해 최적의 안정적인 튜닝 매개변수를 제공하는 것이 필요하다. 전문가 데모에 의한 강력하고 높은 성능은 도 10(c)에 나와 있다. 조정된 매개변수를 사용하기 전에 제어 시스템이 높고 낮은 오류와 같은 다양한 상태를 가지고 있었지만 전문가 데모는 동적 유입 조건에서 강력한 폭기 제어 시스템을 공식화하기 위해 조정 매개변수를 제안했다. 따라서 전문가 데모에 의한 튜닝 매개 변수의 변경은 높은 보상 값을 획득했다. 수집된 전환의 정적 데이터는 저장된 고성능 결과로부터 최적의 정책 $\pi^*(\mathbf{a}|\mathbf{s})$ 을 학습하기 위한 오프라인 AWAC 알고리즘을 사전 훈련하는 데 활용되었다.

[0118] 도 11은 전문가 데모, 오프라인 학습 AWAC, 온라인 미세 조정 AWAC의 보상 값을 커널 밀도와 보상 추세를 이용하여 비교한 그림이다. 커널 밀도는 획득한 샘플에 대한 기본 경험적 분포를 설명하기 위한 비모수적 방법이다. 전문가 데모에 대한 커널 밀도의 피크 포인트는 0 부근에 있으며, 이는 도 11(a)와 같이 전문가 데모가 제어 성능을 크게 향상시키는 데 한계가 있음을 나타낸다. 0.3의 엡실론 값은 훈련된 DQN의 임의적인 동작을 강제로 조정 매개변수를 조정하도록 했기 때문에 전문가 데모는 차선책 정책을 갖는다. 반면, 오프라인 데이터셋으로 훈련된 AWAC는 전문가 데모에 비해 성능이 향상되었고 더 높은 보상 값을 받았다. AWAC는 높은 행동 차원과 넓은 보상으로 어려운 작업에 대한 구별 정책을 학습하는 능력이 있으며, AWAC는 차선책 및 무작위 탐색 데이터에서 성공적으로 활용될 수 있음이 입증되었다. 따라서 오프라인으로 훈련된 AWAC에 대한 커널 밀도의 피크 포인트는 4.2의 보상 값에 배치된다. Smith 예측자에 사용된 AWAC는 수집된 전환의 정적 데이터에서 상태, 행동 및 보상 간의 관계를 해석할 수 있다고 결론지을 수 있다. AWAC 알고리즘은 오프라인 교육 과정에서 환경(폭기 공정의 제어 시스템)과 직접 상호 작용하지 않았지만 완전히 데이터 기반 정책으로 제어 성능이 향상되었다.

[0119] 온라인 미세 조정을 위해 2017년에 실제 측정된 데이터에 완전히 데이터 기반으로 훈련된 AWAC를 사용했다. 오프라인 RL을 실제 제어 시스템에 적용하기 위해 컨트롤러의 작동 조건이 미세 조정 조건을 만족할 때 AWAC 알고리즘의 미세 조정이 활성화되었다. 본 발명의 실시예에서는 다음과 같이 세 가지 미세 조정 활성화 조건을 제안했다.

[0120] (1) 매우 낮고 높은 유입 COD 농도

[0121] (2) 매우 낮고 높은 유입수 TN 농도

[0122] (3) 낮은 제어 성능 감지

[0123] AWAC의 온라인 미세 조정 성능은 도 11(a)에 나와 있다. 전문가 데모 및 오프라인 훈련된 AWAC와 비교할 때 온라인으로 미세 조정된 AWAC는 더 높은 보상 값을 얻기 위해 더 나은 성능을 보였다. 따라서 커널 밀도는 더 높은 보상 값의 위치에 배치되었다. 또한, 커널 밀도는 더 높은 보상 방향으로 이동하고 훈련 단계에 따라 좁은 모양으로 변경되었다. 온라인으로 미세 조정된 AWAC는 비선형 유입수 조건에서 조정 매개변수를 안정적으로 제안하는 기능이 있었다. 앞서 언급한 극한 조건을 포함한 훈련 과정에서 가장 낮은 음의 보상 값과 높은 보상을 받았기 때문에 강력한 제어 튜닝 시스템이 될 것이다.

[0124] 도 11(b)는 전문가 데모, 오프라인으로 훈련된 AWAC 및 온라인으로 미세 조정된 AWAC의 95% 신뢰 구간을 포함하는 보상 값의 추세를 보여준다. 전문가 데모의 보상은 매우 불안정하게 변동했다. DQN 알고리즘은 30% 탐색(exploration)과 70% greedy actio을 통해 매개변수 조정을 위한 차선책을 수집하므로 보상 값은 에포크에 따라 다르다. 전문가 데모 DQN 알고리즘에 의한 전환의 정적 데이터를 기반으로 AWAC는 actors와 critics에 대한 가중치 값을 업데이트했다. AWAC는 수집된 데이터에서 더 높은 성능의 정책을 추출하여 점차적으로 제어 성능을 높였다. 따라서 오프라인으로 훈련된 AWAC의 보상 값은 Epoch의 숫자에 따라 증가했다. 마지막으로 2017년에 테스트 및 업데이트된 온라인 미세 조정 AWAC는 별표로 표시된 극한 조건에서도 견고하고 안정적인 제어 성능을

보여주었다. 극단적인 상황은 40번 이상 발생했으며, 이는 1년에 12% 이상에 해당한다. 이는 대상 플랜트의 실제 유입 조건이 매우 비선형적이었고 유입 시나리오로 분류하기 어려웠음을 의미한다. 극한의 유입수 조건이 빈번하게 발생함에도 불구하고 온라인으로 미세 조정된 AWAC는 긍정적인 보상 값을 얻기 위한 제어 작업을 수행했다. 결과적으로 온라인으로 미세 조정된 AWAC는 수집된 오프라인 데이터와 제어 시스템 간의 상호 작용을 사용하여 가장 높은 보상을 받는 최고의 제어 성능을 보여주었다.

[0126] AWAC 기반 자동 적응 컨트롤러 튜닝 시스템은 저유입, 정상 및 고유입 조건에서 평가되었다. 도 12은 2018년 5일 동안 실제 측정된 유입수 COD 및 TN 농도를 보여준다. 특히 유입수 COD 및 TN 조성비는 저유입, 정상, 고유입 조건에서 각각 (0.72, 0.81), (0.93, 1.06), (1.29, 1.09)였다. 5일의 유입수 데이터셋은 2월에 유입 오염물질 농도가 증가하고 7월에 감소하는 계절적 유입 특성을 기반으로 선택되었다. 유입수 오염물질 농도는 시간에 따라 크게 변하는 비선형 및 동적 경향을 보였다. 또한 오전 10시부터 오전 1시까지 유입부하가 주로 증가하는 가사활동에 따라 COD와 TN 농도가 변화하는 명백한 일변동이 관찰되었다. 유입수 조건의 이러한 비선형 및 동적 특성은 기존 제어 시스템을 사용하는 경우 유입수 조건의 불확실성으로 인해 제어 성능이 악화되는 문제를 야기한다. 따라서 기존 제어 시스템의 한계를 극복하기 위해 AWAC 기반 자동 적응 제어기 튜닝 시스템을 제안하고 도 12와 같이 저, 중, 고 조건의 실제 측정 유입수 데이터에서 평가하였다.

[0127] 도 13은 3가지 유입수 조건에서 제안된 AWAC 기반 튜닝 시스템의 제어 성능을 보여주고 있다. 비례 이득, 적분 시간 및 미분 시간을 포함한 조정 매개변수의 값은 AWAC 알고리즘에 의해 자동으로 조정되었다. AWAC는 도 13(a)와 같이 DO 설정값의 평균값이 1.91 mg/L인 저유입 조건에서 안정적인 제어 성능을 보였다. AWAC 알고리즘은 ITAE-1 튜닝 규칙에 의해 계산된 튜닝 매개변수에 대한 비례 이득의 유사한 값을 제안했다. 또한 유입 부하가 증가했지만 DO 설정값이 변경되지 않은 경우 비례 이득 값이 증가했다(0.5에서 1일로 시간 참조). AWAC는 유입 부하가 증가할 때 비례 이득을 증가시켜 공정 출력을 증가시키려는 의도였다. 다른 시간에 비례 이득은 일반적으로 호기성 반응기에 대한 폭기 유량을 유지했다. 제안된 적분 시간 값은 ITAE-1 기반 조정 매개변수와 유사했지만 증가된 설정값 아래에 머무를 수 있는 DO 농도를 방지하기 위해 5일째에 감소했다. 따라서 감소된 적분 시간 값은 약간의 진동에도 불구하고 프로세스 출력을 빠르게 증가시켰다. AWAC 기반의 미분시간 튜닝값은 일주 유입수에 따른 경향을 보였다. 미분시간의 감소는 제어시스템에 의한 진동의 감소를 의미하므로, 유입부하 증가에 따라 점진적으로 감소하는 값으로 DO 농도의 진동을 회피하였다.

[0128] 도 13(b)는 평균 DO 설정값이 2.29 mg/L인 정상적인 유입 조건에서 자동 적응 튜닝 시스템의 성능을 보여준다. AWAC 알고리즘은 DO 설정값의 증가와 관련하여 비례 이득 값을 자주 증가시켰다. 이는 AWAC 알고리즘이 증가하는 DO 설정값과 호기성 반응기의 낮은 DO 농도 사이의 오차가 감지되었을 때 비례 이득을 증가시켜 공정 출력(즉, 폭기 유량)을 빠르게 증가시키려고 시도했음을 의미한다. 한편, 적분 시간 값은 DO 설정점에 비해 더 낮은 DO 농도를 증가시키기 위해 ITAE-1 조정 매개변수와 유사하거나 감소되었다. 증가하는 DO 설정값에 의해 발생하는 제어 오류에 해당하는 폭기 유량을 조작하기 위해 제안된 비례 이득과 적분 시간의 경향이 함께 변경되었다. 또한 비례 게인의 증가로 인해 발생할 수 있는 큰 진동을 줄이기 위해 비례 게인의 증가에 따라 미분 시간이 감소했다.

[0129] 도 13(c)는 COD 및 TN 조성비가 (1.29, 1.09)인 극도로 높은 유입수 조건에서의 제어 성능을 나타낸다. DO 설정값의 평균값은 2.69mg/L로 많은 양의 통기가 필요했다. 저유입 및 정상 유입 조건에서의 결과와 비교하여 비례 이득 값이 크게 증가하였다. 시스템 식별은 평균 유입수 조건에서 수행되었기 때문에 기존 PID 제어기는 유입수 조건이 크게 증가하는 경우에는 제한적이었다. AWAC 알고리즘은 컨트롤러가 작동 설정값을 추적하도록 성공적으로 안내하면서 비례 이득 값을 증가시켜 과잉 감쇠 응답을 방지했다. 한편, 폭기유량을 빠르게 증가시키기 위해 적분시간값을 감소시켰다. 한편, P 부분과 I 부분 제어기에 의해 폭기유량을 증가시켰을 때, 감소된 미분시간은 빈번한 진동을 방지하였다. 따라서 AWAC 기반 자동 적응 제어기 튜닝 시스템을 사용한 폭기 시스템은 작동 설정값을 추적하고 DO 농도를 효율적으로 증가시키면서 기존 제어 시스템에서 일반적으로 발생할 수 있는 진동 현상을 안정적으로 방지했다.

[0131] 도 14는 AWAC 기반 자동 적응 제어기 튜닝 시스템, PID 제어 및 Smith 예측기 간의 제어 성능 비교를 보여준다. PID와 Smith 예측기를 포함한 기존 제어 시스템은 ITAE-1 조정 규칙에 따라 조정되었다. 도 14(a)는 낮은 유입수 조건에서 제어 성능을 비교한 것이다. 기존 PID 제어기 시스템은 낮은 DO 설정값을 추적하는 데 한계가 있었고 높은 진동 경향을 보였다. 따라서 PID 제어의 오류가 많이 발생하였다. 제어 오류로 인해 공기 유량도 시간

에 따라 크게 변했다. 이러한 진동은 주로 시간 지연에 의해 발생했다. 프로세스의 상당한 시간 지연은 피드백 신호 또는 입력 지연으로 인해 제어 시스템의 성능을 심각하게 제한한다. 따라서 시간지연 보상기는 도 14(a)와 같이 향상된 제어 성능을 보였다. Smith 예측기와 AWAC 기반 제어 시스템 모두 폭기 시스템에 대한 시간 지연 효과를 고려하여 제어 시스템을 개선했다. Smith 예측자와 비교하여 제안된 데이터 기반 제어 시스템은 설정값이 증가할 때 조정 매개변수를 적응적으로 업데이트하여 설정값을 잘 추적했다(0.5일 및 4.5일의 최고점 참조).

[0132] AWAC 기반 제어 시스템은 도 14(b)와 같이 정상적인 유입수 조건에서 안정적인 제어 성능을 보였다. 기존 PID 제어 시스템에 비해 진동을 효과적으로 감소시켰다. 또한 AWAC 알고리즘은 비례 이득을 증가시켜 과감쇠 응답을 방지하기 위해 튜닝 매개변수를 조정했다. 따라서 AWAC 기반 제어 시스템을 사용할 때 제어 오류가 감소했다. 극도로 높은 유입수 조건에서 AWAC 알고리즘은 도 14(c)와 같이 강력한 제어 성능을 보였다. AWAC 기반 제어 시스템은 도 13(c)와 같이 대부분의 작동 시간 동안 가장 높은 비례 이득을 선택하여 더 높은 설정값으로 빠르게 수렴한다. 더 높은 오류가 감지됨에 따라 언더슈트를 줄이면서 빠르게 증가하는 프로세스 입력을 갖는 것이 높은 유입수 조건에서 AWAC 알고리즘의 주요 목표였다.

[0133] 결과적으로, 본 발명의 실시예에 따른 AWAC 기반 제어 시스템은 DO 농도를 안정적으로 제어하고 낮은 진동 범위에서 공기 유량을 조작하는 능력을 가졌다. 호기성 반응기의 DO 농도가 질산화의 핵심 요소이기 때문에 설정값과 비교하여 과도하거나 부족한 폭기는 불안정한 생물학적 처리 공정을 유발할 수 있다. 또한 DO 제어의 넓은 범위는 생물학적 행동을 악화시킬 것이다. 따라서 DO 농도가 엄격한 배출 요구 사항을 충족하려면 만족스러운 제어가 필요하다. 이러한 맥락에서 본 발명의 실시예에 따른 제어 시스템은 기존의 폭기 제어 시스템과 비교하여 만족스러운 성능으로 DO 설정값을 추적함으로써 성능을 능가했다. 한편, 장비의 관점에서 볼 때 불안정한 컨트롤러에 의한 폭기유량의 변화는 기계적 마모 및 폭기 액츄에이터의 오작동을 유발하여 공정 운전에 심각한 단점이 된다. 또한 빈번한 제어는 더 높은 에너지 수요를 필요로 하다. 기존 PID 시스템에 의해 크게 변동하는 공기 유량은 지속 가능한 WWTP 운영의 함정이 될 수 있다. 결과적으로 AWAC 기반 자동 적응 컨트롤러 튜닝 시스템은 기계적 과부하 및 과도한 에너지 요구를 줄이기 위해 폭기 시스템의 빠르고 안정적인 제어 성능을 보장한다.

[0134] 또한, 세 가지 제어 시스템의 제어 성능을 비교하여 표 3에 요약하였다. 성능은 3개의 유입수 조건에서 5일 동안의 오차 집계를 측정하기 위해 2개의 성능 평가 지표로 평가되었다. 이때 제어기 평가기준은 오차절대값(IAE)의 적분과 시간가중절대오차(ITAE)의 적분을 사용하였다. 결과는 Smith 예측기가 PID 제어 시스템에 비해 더 적은 오류를 생성하므로 향상된 성능을 보여주었음을 확인했다. AWAC 기반 자동 튜닝 시스템을 구현한 후 IAE로 표시되는 절대 오류와 ITAE로 표시되는 영구 오류가 최대 31.86% 및 39.38% 감소했다.

[0136] AWAC 기반 자동 적응 컨트롤러 튜닝 시스템이 12.89%와 39.38% 사이에서 오류가 개선됨에 따라 기존 PID 제어 및 Smith 예측 시스템에 비해 상당한 개선을 제공했음을 보여준다. 더욱이, 평가는 제어 시스템에서 시간 지연을 고려하고 컨트롤러 매개변수를 조정하는 것의 중요성을 보여준다. 그들은 설정값 추적과 비선형 및 다양한 유입 조건에서 필수적인 역할을 한다. 따라서 완전 데이터 기반 오프라인 RL 알고리즘에 의해 호기성 반응기의 DO 농도를 유지하는 데 더 적합한 제어 응답이 달성되었다. AWAC는 전문가 데모의 수집된 데이터 세트에서 가능한 유입 시나리오와 해당 제어 성능을 배웠다. 따라서 기존 제어 시스템과 온라인 RL이 안정적인 제어 성능을 제한할 수 있는 극한의 유입수 조건에도 불구하고 최상의 튜닝 매개변수를 제안할 수 있는 능력이 있었다.

[0137] [표 3] 컨트롤러 평가 기준에 따른 제어 시스템의 성능 수치 비교

Influent conditions	Control systems	IAE	Improvement	ITAE	Improvement
Low	PID	176.31	-	81.68	
	Smith predictor	126.81	28.08 %	54.45	33.34 %
	AWAC-based auto system	**120.13**	**31.86 %**	**49.51**	**39.38 %**
Normal	PID	144.90		63.25	
	Smith predictor	138.88	4.15 %	62.23	1.61 %
	AWAC-based auto system	**126.22**	**12.89 %**	**55.03**	**13.00 %**
High	PID	132.21		58.49	
	Smith predictor	132.49	-0.21 %	60.24	-3.00 %
	AWAC-based auto system	**114.72**	**13.23 %**	**47.16**	**19.36 %**

[0138]

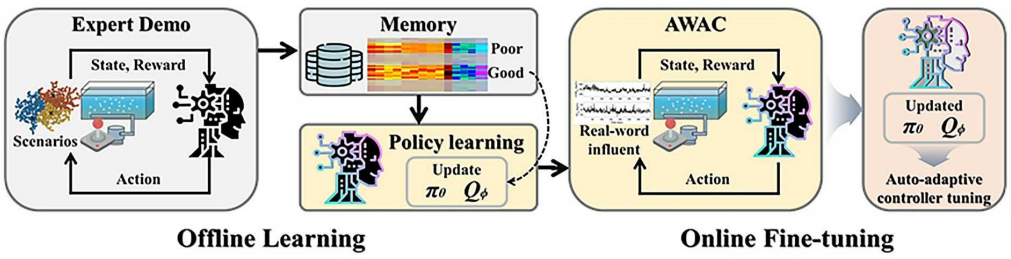
[0140] 본 발명의 실시예에서는 시간 지연 및 비선형 유입수 조건을 고려하여 WWTP의 폭기 시스템에 대한 새로운 자동 적응 컨트롤러 튜닝 시스템을 제공한다. 다양한 유입수 조건 시나리오에서 전문가 데모 알고리즘으로 오프라인 데이터 세트를 수집했다. 그런 다음 AWAC 알고리즘은 수집된 데이터로부터 상태(오류), 동작(조정 매개변수), 보상 정보를 포함하는 조정 매개변수 결정 정책만을 학습했다. 완전 데이터 구동 AWAC 알고리즘을 사용하여 비례 이득, 적분 시간 및 미분 시간을 포함한 튜닝 매개변수의 값을 제안했다.

[0141] AWAC 기반 자동 적응 제어기 튜닝 시스템의 결과는 폭기 시스템의 제어 성능을 증가시키기 위해 튜닝 매개변수가 적절하게 제안되었음을 보여주었다. 또한 새로 측정된 유입수 데이터를 추가하여 AWAC 기반 제어 시스템이 극한의 유입수 조건에 적응적으로 대처할 수 있음을 나타낸다. 제안된 시스템은 기존 PID 제어 시스템과 비교하여 작동 설정값을 빠르게 추적하고 진동을 줄이며 오버슈트 및 오버댐핑 응답을 방지하여 제어 성능을 향상시켰다. 따라서 AWAC 알고리즘은 매우 낮은 유입 조건에서 제어 오류를 39.38%, 매우 높은 유입 조건에서 19.36%까지 극적으로 줄였다. 결과적으로 AWAC 기반 자동 튜닝 시스템은 매우 다양하고 비선형적인 작동 조건에 대한 적응성을 통해 WWTP의 운영 프로세스에 대한 안정적이고 강력한 제어 구조에 혁신적인 기여를 보여주었다. 제어 시스템의 구조를 변경하지 않고 효과적인 튜닝 매개변수를 제안함으로써 실현 가능한 제어 성능이 필요한 실제 WWTP 시스템에 활용될 수 있다.

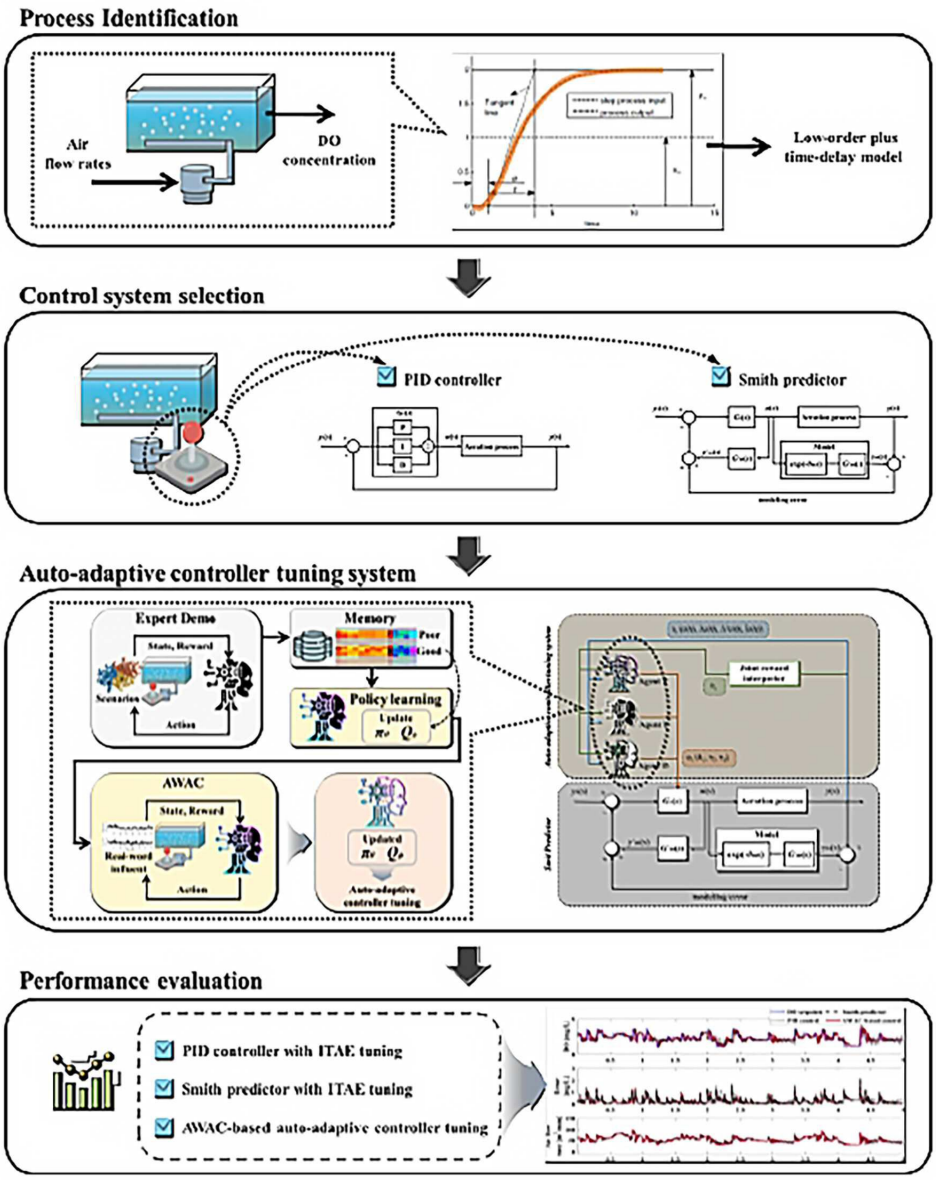
[0143] 또한, 상기와 같이 설명된 장치 및 방법은 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

도면

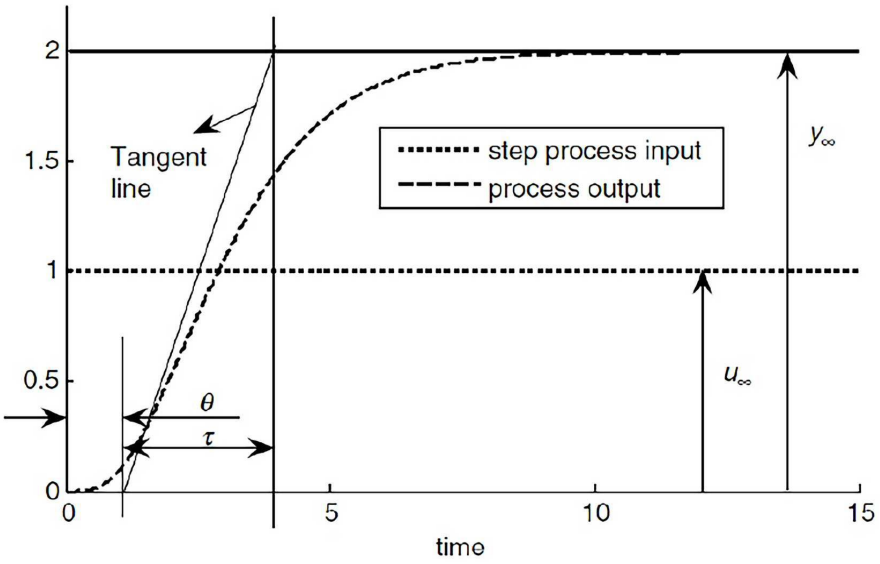
도면1



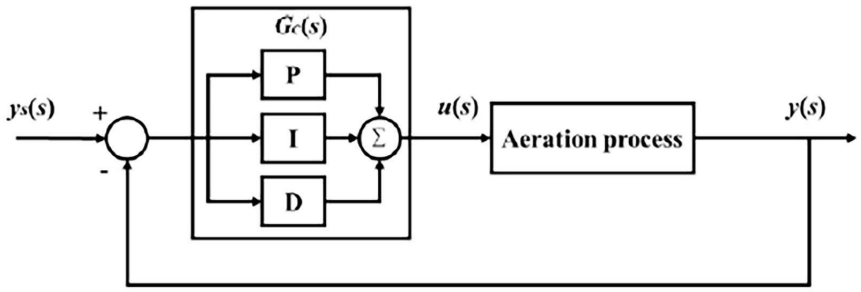
도면2


Process Identification
Air flow rates
DO concentration
Low-order plus time-delay model
Control system selection
PID controller
Smith predictor
Auto-adaptive controller tuning system
Expert Demo
State, Reward
Scenarios
Action
Memory
Poor
Good
Policy learning
Update
AWAC
State, Reward
Real-word influent
Action
Updated
Auto-adaptive controller tuning
Performance evaluation
PID controller with ITAE tuning
Smith predictor with ITAE tuning
AWAC-based auto-adaptive controller tuning

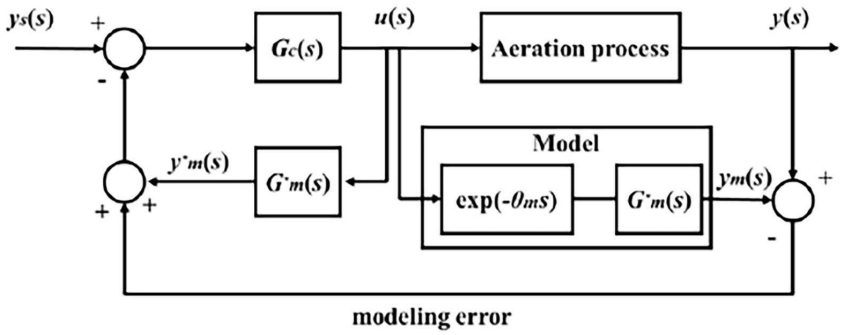
도면3


2
1.5
1
0.5
0
0
5
10
15
time
Tangent line
step process input
process output
y_∞
u_∞
θ
τ

도면4

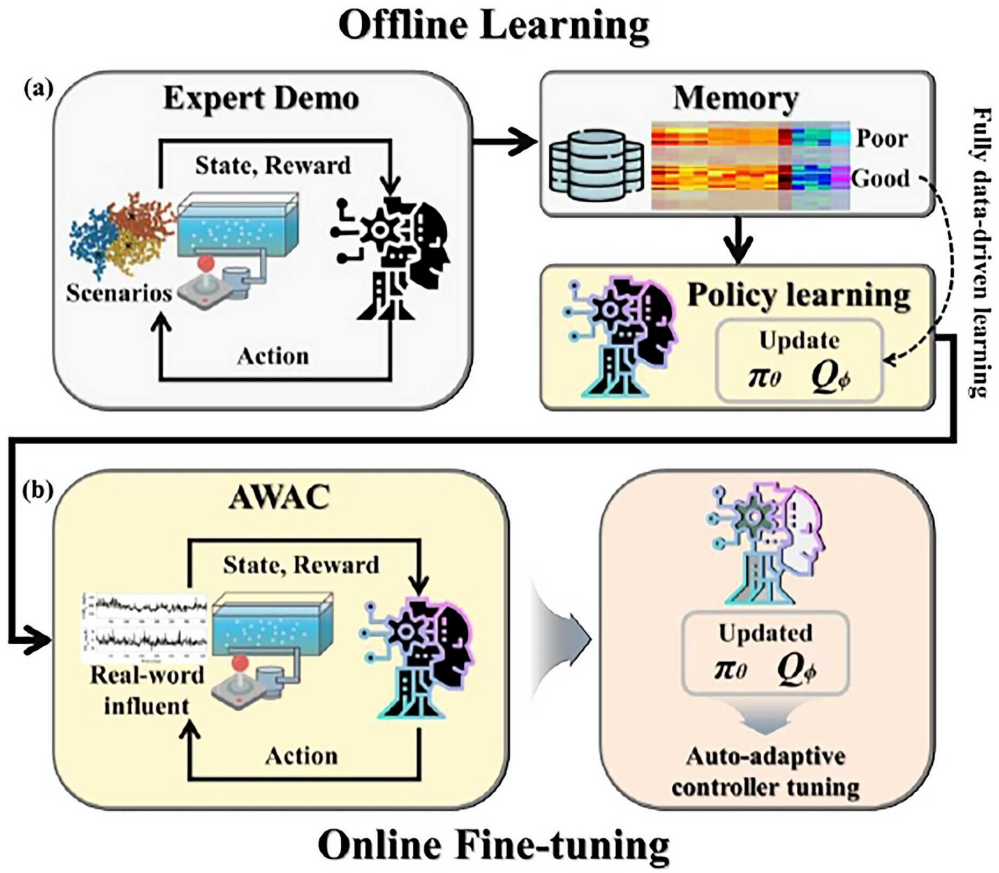

$G_c(s)$
P
I
D
$y_s(s)$
+
-
$u(s)$
Aeration process
$y(s)$

(a)

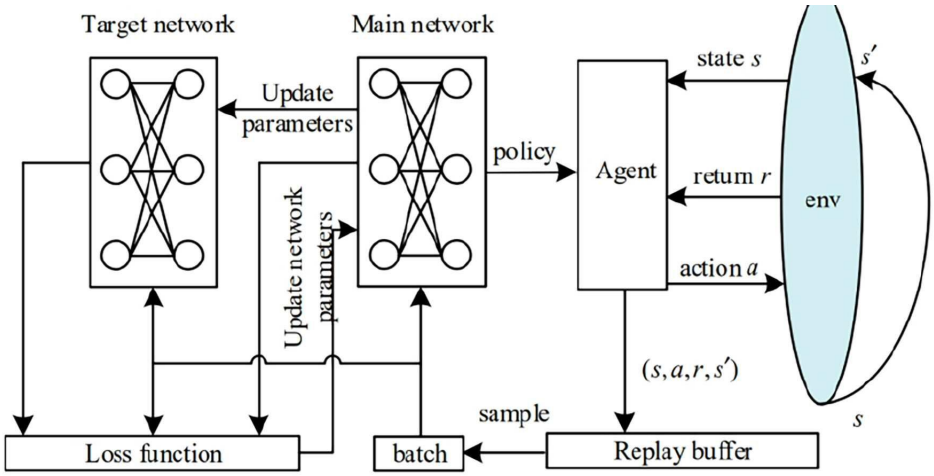

$y_s(s)$
+
-
$G_c(s)$
$u(s)$
Aeration process
$y(s)$
Model
$y'_m(s)$
$G'_m(s)$
exp(-$\theta_m s$)
$G'_m(s)$
$y_m(s)$
+
+
+
-
modeling error

(b)

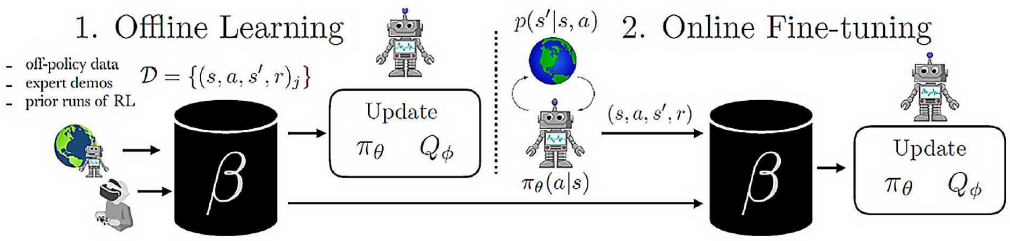
도면5


Offline Learning
(a)
Expert Demo
State, Reward
Scenarios
Action
Memory
Poor
Good
Policy learning
Update
π_θ Q_ϕ
Fully data-driven learning
(b)
AWAC
State, Reward
Real-word influent
Action
Updated
π_θ Q_ϕ
Auto-adaptive controller tuning
Online Fine-tuning

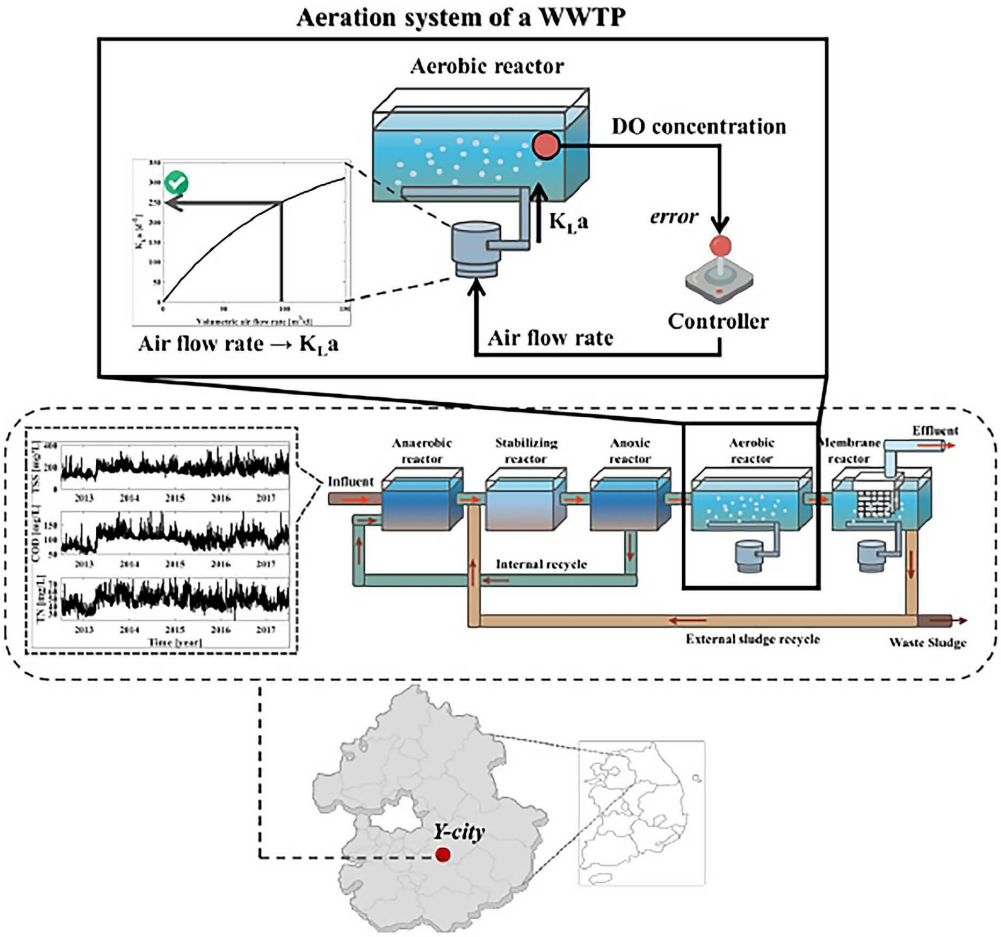
도면6


Target network
Main network
Update parameters
policy
Agent
state s
return r
action a
env
s'
s
Update network parameters
(s,a,r,s')
sample
Loss function
batch
Replay buffer

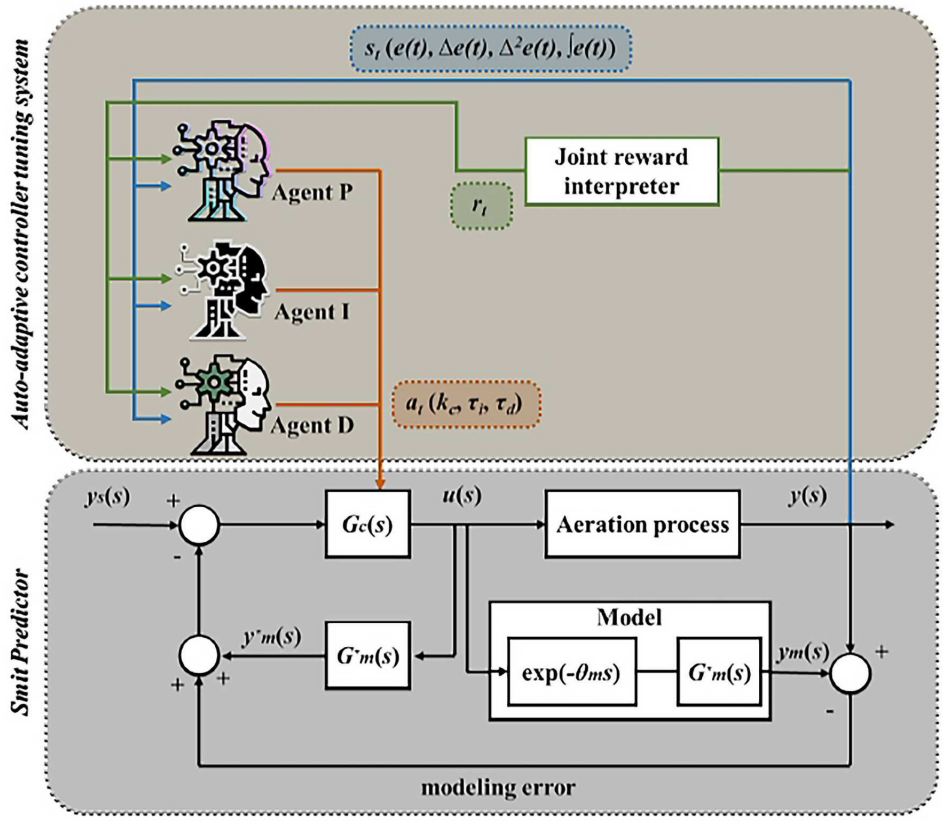
도면7


1. Offline Learning
- off-policy data
- expert demos
- prior runs of RL
$\mathcal{D} = \{(s, a, s', r)_j\}$
Update
π_θ Q_ϕ
$p(s'|s,a)$
2. Online Fine-tuning
(s, a, s', r)
$\pi_\theta(a|s)$
Update
π_θ Q_ϕ

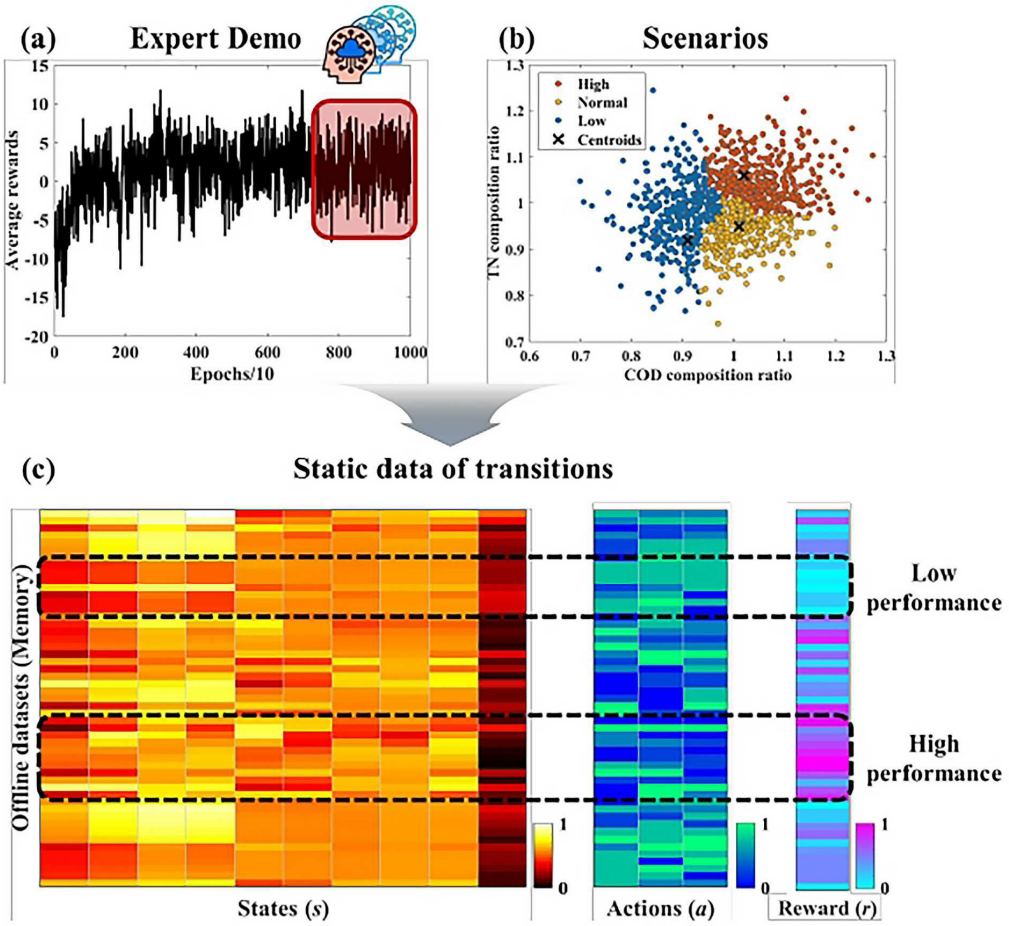
도면8


Aeration system of a WWTP
Aerobic reactor
DO concentration
K_La
error
Controller
Air flow rate
Air flow rate → K_La
TSS [mg/L]
COD [mg/L]
TN [mg/L]
2013
2014
2015
2016
2017
Time [year]
Influent
Anaerobic reactor
Stabilizing reactor
Anoxic reactor
Aerobic reactor
Membrane reactor
Effluent
Internal recycle
External sludge recycle
Waste Sludge
Y-city

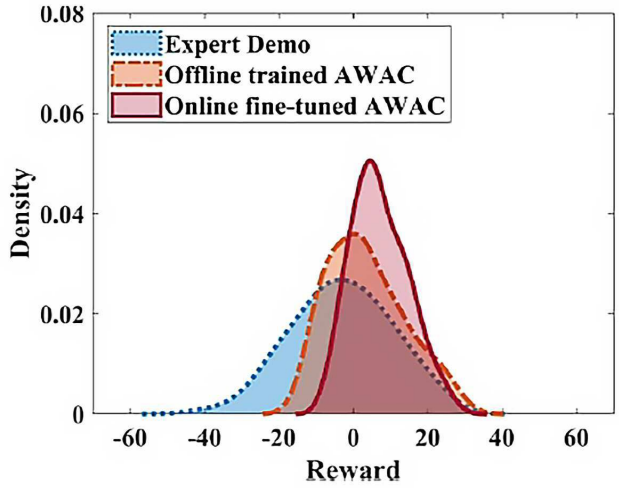
도면9


Auto-adaptive controller tuning system
s_t (e(t), Δe(t), Δ²e(t), ∫e(t))
Agent P
Agent I
Agent D
r_t
Joint reward interpreter
a_t (k_c, τ_i, τ_d)
Smit Predictor
ys(s)
+
-
Gc(s)
u(s)
Aeration process
y(s)
Model
exp(-θms)
G'm(s)
ym(s)
y'm(s)
G'm(s)
+
+
+
-
modeling error

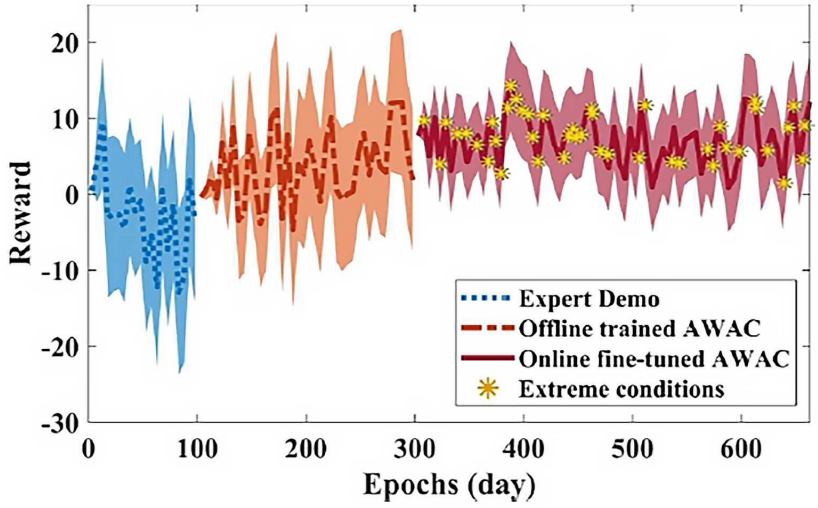
도면10


(a)
Expert Demo
Average rewards
Epochs/10
(b)
Scenarios
High
Normal
Low
Centroids
TN composition ratio
COD composition ratio
(c)
Static data of transitions
Offline datasets (Memory)
States (s)
Actions (a)
Reward (r)
Low performance
High performance

도면11

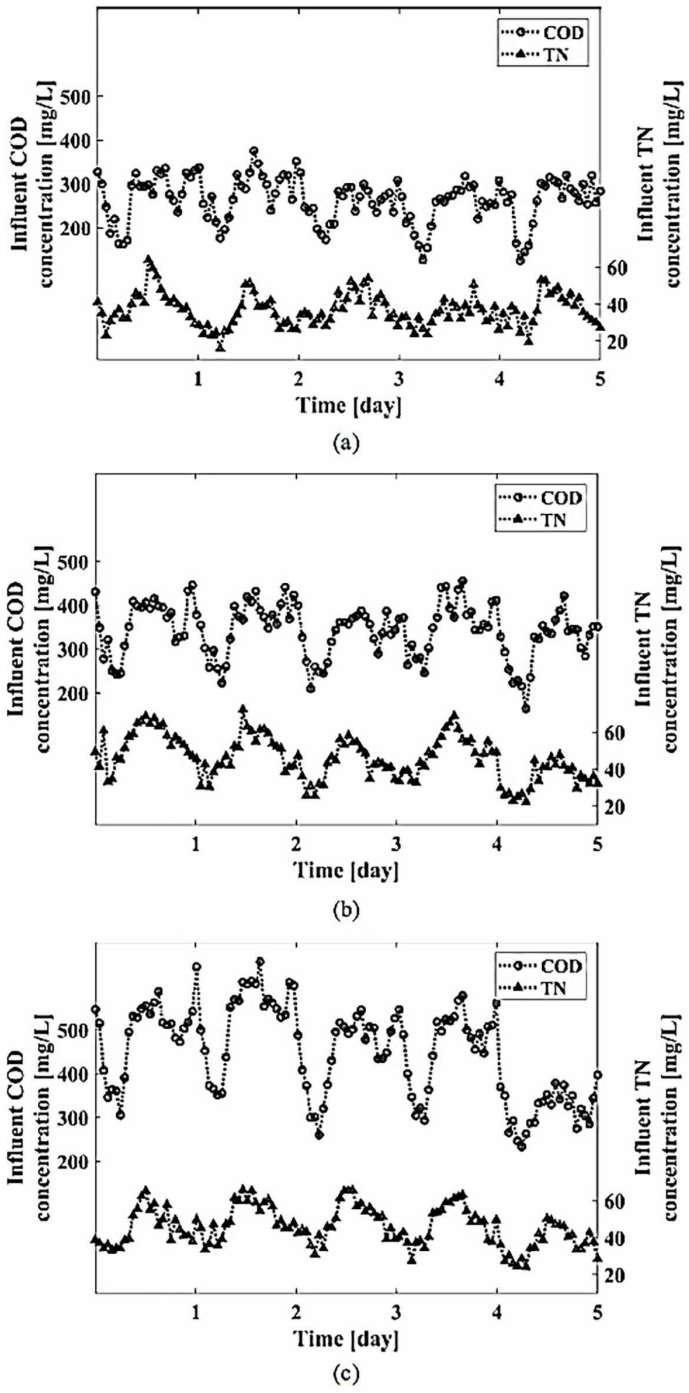

Expert Demo
Offline trained AWAC
Online fine-tuned AWAC
Density
Reward
0.08
0.06
0.04
0.02
0
-60
-40
-20
0
20
40
60

(a)

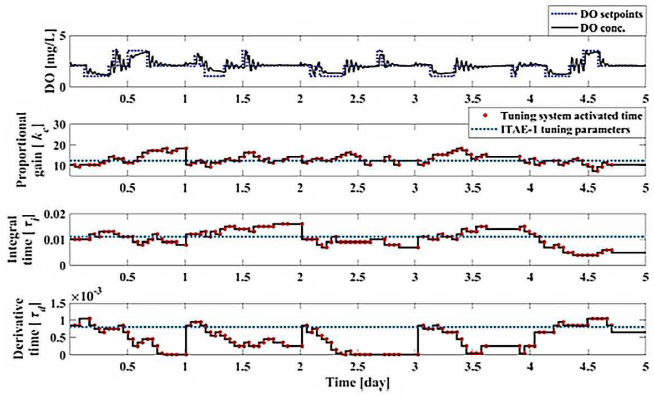

Expert Demo
Offline trained AWAC
Online fine-tuned AWAC
Extreme conditions
Reward
Epochs (day)
20
10
0
-10
-20
-30
0
100
200
300
400
500
600

(b)

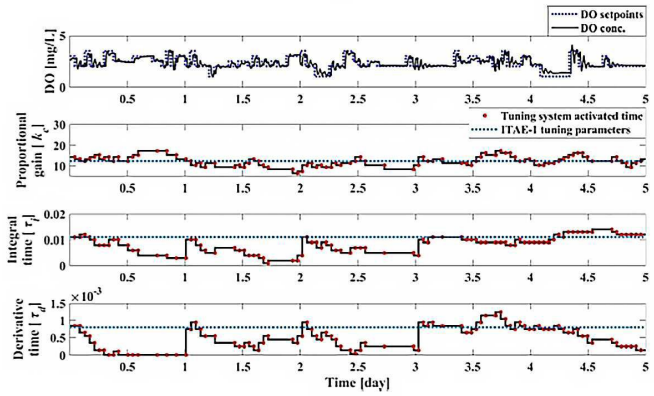
도면12


COD
TN
Influent COD concentration [mg/L]
Influent TN concentration [mg/L]
Time [day]
(a)
(b)
(c)

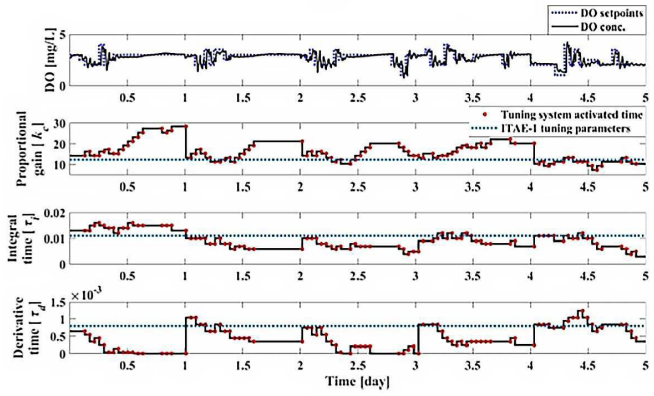
도면13


DO setpoints
DO conc.
DO [mg/L]
Tuning system activated time
ITAE-1 tuning parameters
Proportional gain [k_c]
Integral time [τ_I]
Derivative time [τ_d]
$\times 10^{-3}$
Time [day]

(a)

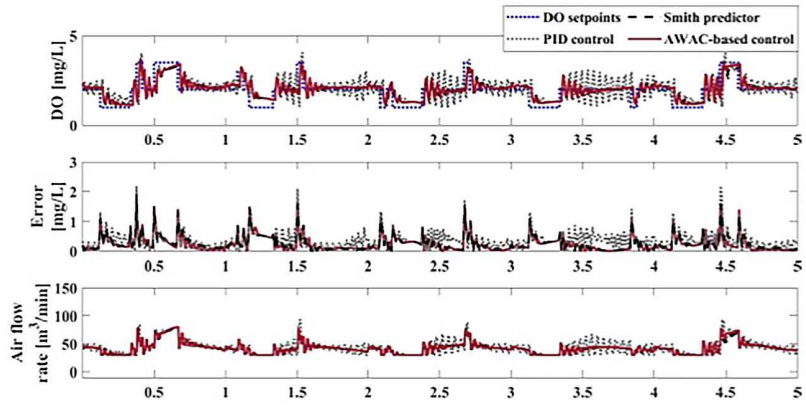

DO setpoints
DO conc.
DO [mg/L]
Tuning system activated time
ITAE-1 tuning parameters
Proportional gain [k_c]
Integral time [τ_I]
Derivative time [τ_d]
$\times 10^{-3}$
Time [day]

(b)

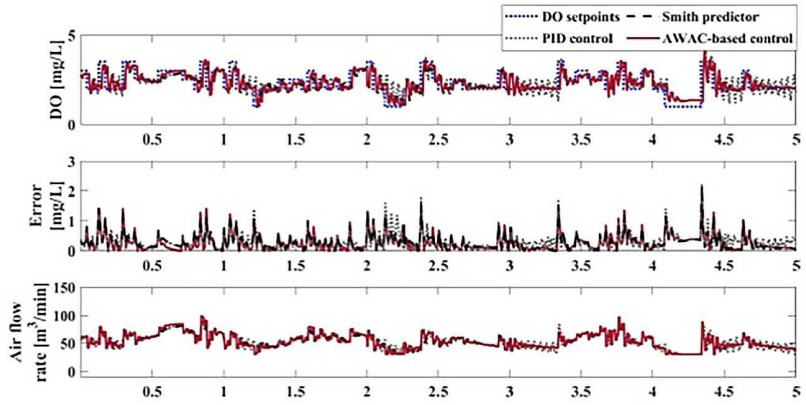

DO setpoints
DO conc.
DO [mg/L]
Tuning system activated time
ITAE-1 tuning parameters
Proportional gain [k_c]
Integral time [τ_I]
Derivative time [τ_d]
$\times 10^{-3}$
Time [day]

(c)

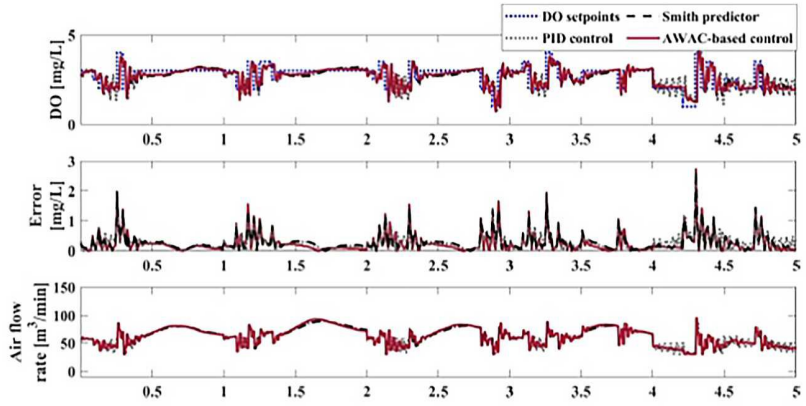
도면14


DO setpoints
Smith predictor
PID control
AWAC-based control
DO [mg/L]
Error [mg/L]
Air flow rate [m3/min]

(a)


DO setpoints
Smith predictor
PID control
AWAC-based control
DO [mg/L]
Error [mg/L]
Air flow rate [m3/min]

(b)


DO setpoints
Smith predictor
PID control
AWAC-based control
DO [mg/L]
Error [mg/L]
Air flow rate [m3/min]

(c)