(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0076831
(43) 공개일자 2025년05월30일

(51) 국제특허분류(Int. Cl.)
G06N 3/0985 (2023.01) *G06N 3/04* (2023.01)
G06N 3/0442 (2023.01) *G06Q 50/06* (2024.01)
G06Q 50/26 (2024.01)
(52) CPC특허분류
G06N 3/0985 (2023.01)
G06N 3/0442 (2023.01)
(21) 출원번호 10-2023-0163945
(22) 출원일자 2023년11월23일
심사청구일자 2023년11월23일

(71) 출원인
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
김종호
울산광역시 남구 봉월로8번길 20, 202호 (신정동)
전 옥 빈
울산광역시 남구 대학로 93 (무거동)
(74) 대리인
특허법인태백

전체 청구항 수 : 총 12 항

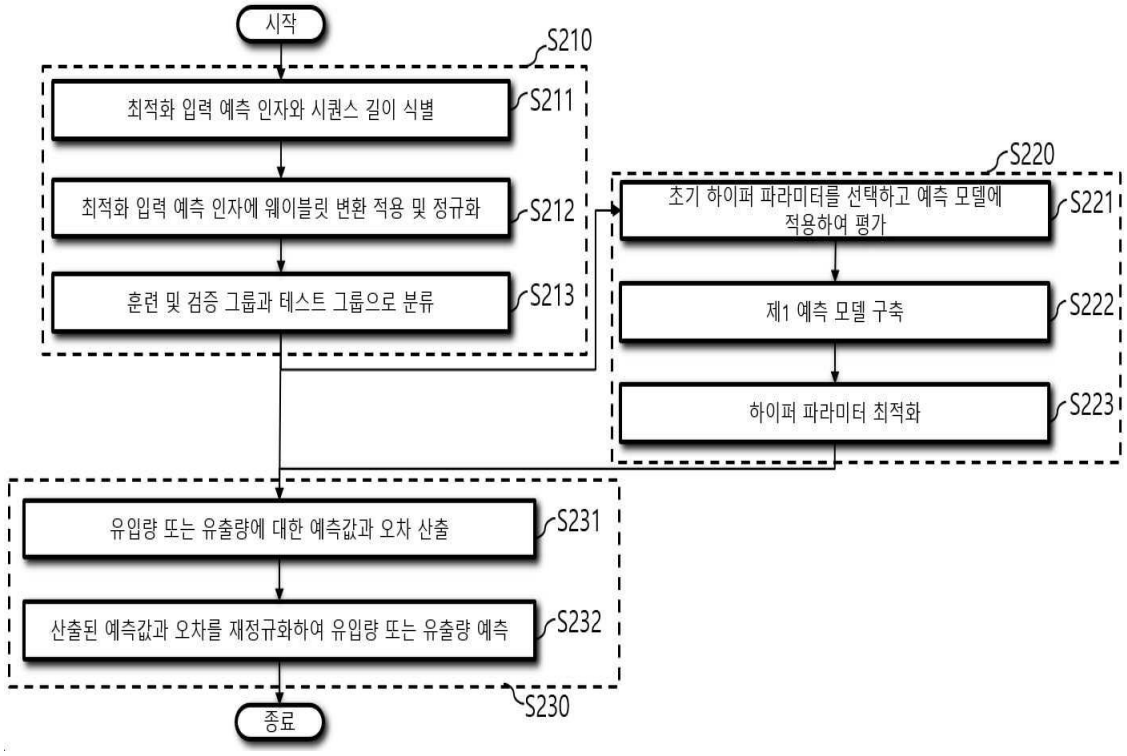
(54) 발명의 명칭 **캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법 및 그 시스템**

(57) 요 약

본 발명은 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법 및 그 시스템에 관한 것이다. 본 발명의 실시예에 따르면 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법에 있어서, 전처리부가 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고 전처리를 수행하는 단계; 학습부는 훈련 및 검증 그룹에 기 지정된

(뒷면에 계속)

대 표 도 - 도2



방법을 적용하여 예측 모델을 학습시키는 단계; 및 예측부는 전처리가 수행된 데이터에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측하는 단계를 포함할 수 있다.

이와 같이 본 발명에 따르면, 불확실성의 원인을 우연적 불확실성 및 인식적 불확실성으로 구분하고, 우연적 불확실성 및 인식적 불확실성을 동시에 정량화하여 저수지의 유입량 또는 유출량을 예측할 수 있다. 또한, 종래에 비해 캐스케이드 저수지 유입량 또는 유출량에 내포된 불확실성을 더 정확하게 예측할 수 있다.

(52) CPC특허분류
G06N 3/049 (2023.01)
G06Q 50/06 (2024.01)
G06Q 50/26 (2024.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1315002008
과제번호	ND641012
부처명	행정안전부
과제관리(전문)기관명	국립재난안전연구원
연구사업명	재난안전공동연구기술개발(R&D)
연구과제명	도시침수 모니터링 기술 개발 및 관리체계 구축
기 여 율	1/1
과제수행기관명	강원대학교 삼척산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

전처리부가 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고 전처리를 수행하는 단계;

학습부는 훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시키는 단계; 및

예측부는 전처리가 수행된 데이터에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측하는 단계를 포함하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법.

청구항 2

제 1 항에 있어서,

상기 전처리를 수행하는 단계는,

상기 전처리를 수행하여 예측 모델을 학습시키기 위하여 전처리된 데이터를 훈련 및 검증 그룹과 테스트 그룹으로 분류하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법.

청구항 3

제 2 항에 있어서,

상기 전처리를 수행하는 단계는,

상기 입력된 데이터에 대해 상관 분석을 수행하여 적어도 하나의 최적화 입력 예측 인자와 시퀀스 길이를 식별하는 단계;

상기 식별된 적어도 하나의 최적화 입력 예측 인자에 웨이블릿 변환을 적용하고 정규화하는 단계; 및

상기 정규화된 적어도 하나의 입력 예측 인자를 기초로 훈련 및 검증 그룹과 테스트 그룹으로 분류하는 단계를 포함하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법.

청구항 4

제 1 항에 있어서,

상기 예측 모델을 학습시키는 단계는,

상기 훈련 및 검증 그룹의 기 지정된 하이퍼 파라미터 범위에서 초기 하이퍼 파라미터를 선택하고 예측 모델에 적용하여 평가하는 단계;

상기 초기 하이퍼 파라미터를 기초로 가우시안 프로세스를 적용하여 제1 예측 모델을 구축하는 단계; 및

상기 하이퍼 파라미터를 최적화하는 단계를 포함하고,

현재 선택된 하이퍼 파라미터보다 최적의 하이퍼 파라미터가 없을 때까지 상기 과정을 반복하여 수행하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법.

청구항 5

제 4 항에 있어서,

상기 하이퍼 파라미터를 최적화 단계는,

아래의 수학식을 이용하여 하이퍼 파라미터를 최적화하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법:

$$EI(\boldsymbol{\delta}_i) = \begin{cases} \left(\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})\right)\Phi\left(\frac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right) + \sigma(\boldsymbol{\delta})\phi\left(\frac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right), & if\ \sigma(\boldsymbol{\delta}_i) > 0 \\ 0, & if\ \sigma(\boldsymbol{\delta}_i) = 0 \end{cases}$$

여기서, EI는 새로운 하이퍼 파라미터를 선택하기 위해 사용된 획득함수(Acquisition function)이고, $\hat{\delta}$는 현재 선택된 하이퍼 파라미터이고, δ_i는 i번째 선택된 하이퍼 파라미터이며, Φ는 $\frac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 누적 분포 함수(Cumulative distribution function, CDF)이고, ϕ는 $\frac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 확률 밀도 함수(Probability density function, PDF)이며, μ(δ)는 예상 유입량 또는 유출량이고, σ(δ)는 분산 값이다.

청구항 6

제 1 항에 있어서,

상기 저수지의 유입량 또는 유출량을 예측하는 단계는,

상기 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 산출하는 단계; 및

상기 산출된 예측값과 오차를 재정규화하여 저수지의 유입량 또는 유출량을 예측하는 단계를 포함하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법.

청구항 7

캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고 전처리를 수행하는 전처리부;

훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시키는 학습부; 및

전처리가 수행된 데이터에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측하는 예측부를 포함하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템.

청구항 8

제 7 항에 있어서,

상기 전처리부는,

상기 전처리를 수행하여 예측 모델을 학습시키기 위하여 전처리된 데이터를 훈련 및 검증 그룹과 테스트 그룹으로 분류하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템.

청구항 9

제 8 항에 있어서,

상기 전처리부는,

상기 입력된 데이터에 대해 상관 분석을 수행하여 적어도 하나의 최적화 입력 예측 인자와 시퀀스 길이를 식별하고, 상기 식별된 적어도 하나의 최적화 입력 예측 인자에 웨이블릿 변환을 적용하고 정규화하며, 상기 정규화된 적어도 하나의 입력 예측 인자를 기초로 훈련 및 검증 그룹과 테스트 그룹으로 분류하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템.

청구항 10

제 7 항에 있어서,

상기 학습부는,

상기 훈련 및 검증 그룹의 기 지정된 하이퍼 파라미터 범위에서 초기 하이퍼 파라미터를 선택하고 예측 모델에 적용하여 평가하고, 상기 초기 하이퍼 파라미터를 기초로 가우시안 프로세스를 적용하여 제1 예측 모델을 구축하며, 상기 하이퍼 파라미터를 최적화하고, 현재 선택된 하이퍼 파라미터보다 최적의 하이퍼 파라미터가 없을 때까지 상기의 과정을 반복하여 수행하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템.

청구항 11

제 10 항에 있어서,

상기 학습부는,

아래의 수학식을 이용하여 하이퍼 파라미터를 최적화하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템:

$$EI(\boldsymbol{\delta}_i) = \begin{cases} \left(\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})\right)\Phi\left(\dfrac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right) + \sigma(\boldsymbol{\delta})\phi\left(\dfrac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right), & if\ \sigma(\boldsymbol{\delta}_i) > 0 \\ 0, & if\ \sigma(\boldsymbol{\delta}_i) = 0 \end{cases}$$

여기서, EI는 새로운 하이퍼 파라미터를 선택하기 위해 사용된 획득함수(Acquisition function)이고, $\hat{\boldsymbol{\delta}}$는 현재 선택된 하이퍼 파라미터이고, $\boldsymbol{\delta}_i$는 i번째 선택된 하이퍼 파라미터이며, Φ는 $\dfrac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 누적 분포 함수(Cumulative distribution function, CDF)이고, ϕ는 $\dfrac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 확률 밀도 함수(Probability density function, PDF)이며, μ(δ)는 예상 유입량 또는 유출량이고, σ(δ)는 분산 값이다.

청구항 12

제 7 항에 있어서,

상기 예측부는,

상기 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 산출하고, 상기 산출된 예측값과 오차를 재정규화하여 저수지의 유입량 또는 유출량을 예측하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템.

발명의 설명

기 술 분 야

[0001] 본 발명은 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법 및 그 시스템에 관한 것으로서, 더욱 상세하게 설명하면 캐스케이드 저수지 또는 댐 예측에 내포된 불확실성을 우연적(Aleatoric) 또는 인식적(Epistemic) 불확실성으로 구분하고, 우연적 불확실성 및 인식적 불확실성의 정량화 과정을 통해 캐스케이드 저수지 또는 댐 유입량 및 유출량을 예측하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법 및 그 시스템에 관한 것이다.

배 경 기 술

[0002] 저수지는 농업용수 또는 하천의 수량을 조절할 목적으로 물을 가두어 놓은 것을 의미한다. 저수지의 유입 및 유출을 정확하게 예측하고 적절한 저수지 운영 계획을 세우는 것은 경제적, 환경적 및 사회적 비용을 경감시키기 위해 매우 중요하다.

[0003] 일 예로, 강우를 앞두고 예비 방류를 시행하기 위해서는 저수지 유입량을 예측하여 홍수 등의 피해를 방지해야 한다.

[0004] 저수지의 유입량 및 유출량을 예측하기 위해 물리적 기반의 방법과 데이터 기반의 방법으로 분류할 수 있다. 물리적 기반의 방법은 유입량 및 유출량 예측을 위해 저수지의 유입 및 유출을 개념화하는 방법으로 실제 저수지의 유입량 및 유출량과 차이가 있다는 한계가 존재한다. 데이터 기반의 방법은 복잡한 비선형성과 고차원성을 포함하고 있어, 저수지의 유입량 및 유출량을 예측하기 제한적이라는 한계가 존재한다.

[0005] 특히, 딥러닝과 같은 인공지능 기술을 이용하여 저수지의 유입량 및 유출량을 예측하는 경우 데이터로부터 발생하는 우연적(Aleatoric) 불확실성과 인공지능 모델로 인하여 발생하는 인식적(Epistemic) 불확실성으로 인한 한계가 존재한다.

[0006] 따라서, 우연적(Aleatoric) 불확실성과 인식적(Epistemic) 불확실성을 정량화하여 저수지의 유입량 및 유출량을 예측하는 모델을 구축하는 기술이 필요한 실정이다.

[0007] 본 발명의 배경이 되는 기술은 한국공개특허 제10-2009-0093520호(2009.09.02. 공개)에 기재되어 있다.

발명의 내용

해결하려는 과제

[0008] 이와 같이 본 발명에 따르면, 저수지 또는 댐 예측에 내포된 불확실성을 우연적 불확실성 또는 인식적 불확실성으로 구분하고, 우연적 불확실성 및 인식적 불확실성의 정량화 과정을 통해 캐스케이드 저수지 또는 댐 유입량 및 유출량을 예측하는 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법 및 그 시스템을 제공하기 위한 것이다.

과제의 해결 수단

[0009] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따르면 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법에 있어서, 전처리부가 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고 전처리를 수행하는 단계; 학습부는 훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시키는 단계; 및 예측부는 전처리가 수행된 데이터에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측하는 단계를 포함할 수 있다.

[0010] 상기 전처리를 수행하는 단계는, 상기 전처리를 수행하여 예측 모델을 학습시키기 위하여 전처리된 데이터를 훈련 및 검증 그룹과 테스트 그룹으로 분류할 수 있다.

[0011] 상기 전처리를 수행하는 단계는, 상기 입력된 데이터에 대해 상관 분석을 수행하여 적어도 하나의 최적화 입력 예측 인자와 시퀀스 길이를 식별하는 단계; 상기 식별된 적어도 하나의 최적화 입력 예측 인자에 웨이블릿 변환을 적용하고 정규화하는 단계; 및 상기 정규화된 적어도 하나의 입력 예측 인자를 기초로 훈련 및 검증 그룹과 테스트 그룹으로 분류하는 단계를 포함할 수 있다.

[0012] 상기 예측 모델을 학습시키는 단계는, 상기 훈련 및 검증 그룹의 기 지정된 하이퍼 파라미터 범위에서 초기 하이퍼 파라미터를 선택하고 예측 모델에 적용하여 평가하는 단계; 상기 초기 하이퍼 파라미터를 기초로 가우시안 프로세스를 적용하여 제1 예측 모델을 구축하는 단계; 및 상기 하이퍼 파라미터를 최적화하는 단계를 포함하고, 현재 선택된 하이퍼 파라미터보다 최적의 하이퍼 파라미터가 없을 때까지 상기 과정을 반복하여 수행할 수 있다.

[0013] 상기 하이퍼 파라미터를 최적화 단계는, 아래의 수학식을 이용하여 하이퍼 파라미터를 최적화할 수 있다:

[0014] $$EI(\boldsymbol{\delta}_i) = \begin{cases} \left(\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})\right)\Phi\left(\frac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right) + \sigma(\boldsymbol{\delta})\phi\left(\frac{\mu(\boldsymbol{\delta}_i) - f(\hat{\boldsymbol{\delta}})}{\sigma(\boldsymbol{\delta}_i)}\right), & if\ \sigma(\boldsymbol{\delta}_i) > 0 \\ 0, & if\ \sigma(\boldsymbol{\delta}_i) = 0 \end{cases}$$

[0015] 여기서, EI는 새로운 하이퍼 파라미터를 선택하기 위해 사용된 획득함수(Acquisition function)이고, $\hat{\delta}$는 현재 선택된 하이퍼 파라미터이고, δ_i는 i번째 선택된 하이퍼 파라미터이며, Φ는 $\frac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 누적분포 함수(Cumulative distribution function, CDF)이고, ϕ는 $\frac{\mu(\boldsymbol{\delta}_i)-f(\boldsymbol{\delta}^*)}{\sigma(\boldsymbol{\delta}_i)}$에 대한 확률 밀도 함수(Probability density function, PDF)이며, μ(δ)는 예상 유입량 또는 유출량이고, σ(δ)는 분산 값이다.

[0016] 상기 저수지의 유입량 또는 유출량을 예측하는 단계는, 상기 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 산출하는 단계; 및 상기 산출된 예측값과 오차를 재정규화하여 저수지의 유입량 또는 유출량을 예측하는 단계를 포함할 수 있다.

[0017] 본 발명의 다른 실시 예에 따르면, 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템에 있어서, 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고 전처리를 수행하는 전처리부; 훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시키는 학습부; 및 전처리가 수행된 데이터에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측하는 예측부를 포함할 수 있다.

발명의 효과

[0018] 이와 같이 본 발명에 따르면, 불확실성의 원인을 우연적 불확실성 및 인식적 불확실성으로 구분하고, 우연적 불확실성 및 인식적 불확실성을 동시에 정량화하여 저수지의 유입량 또는 유출량을 예측할 수 있다. 또한, 종래에 비해 캐스케이드 저수지 유입량 또는 유출량에 내포된 불확실성을 더 정확하게 예측할 수 있다.

도면의 간단한 설명

[0019] 도 1은 본 발명의 일 실시예에 따른 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템의 구성도이다.

도 2 및 도 3은 본 발명의 다른 실시예에 따른 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법의 흐름도이다.

도 4는 본 발명의 다른 실시예에 따른 하이퍼 파라미터의 결과를 도시하는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0020] 이하 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시예를 상세히 설명하기로 한다. 이 과정에서 도면에 도시된 선들의 두께나 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다.

[0021] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0022] 또한 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서, 이는 사용자, 운용자의 의도 또는 관례에 따라 달라질 수 있다. 그러므로 이러한 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

[0023] 도 1은 본 발명의 일 실시예에 따른 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템의 구성도이다.

[0024] 도 1에 도시된 바와 같이, 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템(100)은 전처리부(110), 학습부(120) 및 예측부(130)를 포함할 수 있다.

[0025] 먼저, 전처리부(110)는 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고, 상관분석(Correlation analysis), 웨이블릿 변환(Wavelet transform) 및 정규화(Nomalization)를 통해서 전처리를 수행

하며, 전처리가 수행된 데이터를 훈련 및 검증 그룹과 테스트 그룹으로 분류할 수 있다. 여기서, 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터는 기 설정된 기간(예를 들면, 10년, 20년 등)동안의 (시간별, 일자별, 월별) 강우량, 하천 또는 강의 수위, 유역 평균 유량, 저수지 유입량 및 유출량, 실시간 강우 예측 데이터, 온도, 증발량 등을 포함할 수 있다.

[0026] 구체적으로, 전처리부(110)는 입력된 데이터에 대해 상관 분석을 수행하여 적어도 하나의 최적화 입력 예측 인자(Optimal input predictor)와 시퀀스 길이(Sequence length)를 식별할 수 있다.

[0027] 또한, 전처리부(110)는 식별된 적어도 하나의 최적화 입력 예측 인자에 웨이블릿 변환을 적용하고 정규화할 수 있다.

[0028] 또한, 전처리부(110)는 정규화된 적어도 하나의 최적화 입력 예측 인자를 기초로 훈련 및 검증 그룹(Training and validation set)과 테스트 그룹(Test set)으로 분류할 수 있다.

[0029] 다음으로, 학습부(120)는 훈련 및 검증 그룹을 입력받고, 훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시킬 수 있다.

[0030] 구체적으로, 학습부(120)는 기 지정된 하이퍼 파라미터 범위(Hyperparameter range)와 입력된 훈련 및 검증 그룹에 기 지정된 방법(예를 들어, 베이지안 최적화 알고리즘(Bayesian optimization algorithm, BOA), K-폴드 교차검증(K-fold cross validation))을 적용하여 예측 모델을 학습시킬 수 있다. 이때, 예측 모델은 데이터(예를 들어, 기 설정된 기간(예를 들면, 10년, 20년 등)동안의 (시간별, 일자별, 월별) 강우량, 하천 또는 강의 수위, 유역 평균 유량, 저수지 유입량 및 유출량, 실시간 강우 예측 데이터, 온도, 증발량 등)을 입력받고 기 지정된 하이퍼 파라미터를 최적화하여 유입량 및 유출량을 예측하도록 학습될 수 있다.

[0031] 다음으로, 예측부(130)는 테스트 그룹을 입력받고, 입력된 테스트 그룹에 기 학습된 예측 모델을 적용하여 우연적 불확실성 및 인식적 불확실성 정량화 과정을 포함하여 저수지의 유입량 또는 유출량을 예측할 수 있다.

[0032] 구체적으로, 예측부(130)는 입력된 테스트 그룹에 기 학습된 예측 모델을 적용하여 이분산 회귀분석(Heteroscedastic regression, HR)을 수행하고, 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 생성하며, 생성된 예측값과 오차를 재정규화(Renormalization)하여 저수지의 유입량 또는 유출량을 예측할 수 있다.

[0033] 더 나아가, 예측부(130)는 저수지의 유입량 및 유출량을 예측하기 위한 데이터를 입력받고, 입력된 데이터에 기 학습된 예측 모델을 적용하여 우연적 불확실성 및 인식적 불확실성을 고려한 저수지의 유입량 또는 유출량을 예측할 수 있다.

[0034] 이하에서는 도 2 내지 도 4를 이용하여 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법에 대해 보다 구체적으로 설명한다.

[0035] 도 2 및 도 3은 본 발명의 다른 실시예에 따른 캐스케이드 저수지 유입 및 유출 예측 모델 구축 방법의 흐름도이다.

[0036] 도 2 및 도 3에 도시된 바와 같이, 전처리부(110)는 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터를 입력받고, 전처리를 수행하며 전처리가 수행된 데이터를 훈련 및 검증 그룹과 테스트 그룹으로 분류할 수 있다(S210). 여기서, 캐스케이드 저수지 유입량 및 유출량 예측을 위한 데이터는 기 설정된 기간(예를 들면, 10년, 20년 등)동안의 (시간별, 일자별, 월별) 강우량, 하천 또는 강의 수위, 유역 평균 유량, 저수지 유입량 및 유출량, 실시간 강우 예측 데이터, 온도, 증발량 등을 포함할 수 있다.

[0037] 구체적으로, 전처리부(110)는 입력된 데이터에 대해 상관 분석을 수행하여 적어도 하나의 최적화 입력 예측 인자와 시퀀스 길이를 식별할 수 있다(S211). 이때, 상관 분석은 교차상관함수(Cross correlation function, CCF)와 편자기상관함수(Partial autocorrelation function, PACF)를 이용할 수 있다.

[0038] 또한, 전처리부(110)는 식별된 적어도 하나의 최적화 입력 예측 인자에 웨이블릿 변환을 적용하고 정규화할 수 있다(S212).

[0039] 또한, 전처리부(110)는 정규화된 적어도 하나의 최적화 입력 예측 인자를 기초로 훈련 및 검증 그룹(Training and validation set)과 테스트 그룹(Test set)으로 분류할 수 있다(S213)

[0040] 다음으로, 학습부(120)는 전처리가 수행된 데이터에 기 지정된 방법을 적용하여 예측 모델을 학습시킬 수 있다

(S220).

[0041] 즉, 학습부(120)는 입력된 훈련 및 검증 그룹에 기 지정된 방법을 적용하여 예측 모델을 학습시킬 수 있다(S220). 이때, 예측 모델은 데이터(예를 들어, 기 설정된 기간(예를 들면, 10년, 20년 등)동안의 (시간별, 일자별, 월별) 강우량, 하천 또는 강의 수위, 유역 평균 유량, 저수지 유입량 및 유출량, 실시간 강우 예측 데이터, 온도, 증발량 등)을 입력받고 기 지정된 하이퍼 파라미터를 최적화하여 유입량 및 유출량을 예측하도록 학습될 수 있다.

[0042] 구체적으로, 예측 모델은 장단기 메모리(Long short-term memory, LSTM)를 기초로 생성된 것으로 망각 게이트(f_t), 입력 게이트(i_t) 및 출력 게이트(o_t)를 포함하고, 시간(t)에서 입력된 데이터($\boldsymbol{x}_t = [x_{1,t}, x_{2,t}, \ldots, x_{N_{in},t}]$)에 대하여 N_{in}은 입력된 데이터의 수이고, 장단기 메모리의 셀 상태(c_t)와 히든 상태(h_t)는 아래의 [수학식 1]을 이용하여 산출될 수 있다.

수학식 1

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \tilde{c}_t$$

[0043]
$$h_t = o_t \times \tanh(c_t)$$

[0044] 여기서, 망각 게이트(f_t)는 현재 장단기 메모리의 셀 상태(c_t)에서 이전 셀 상태(c_{t-1})가 얼마나 유지되는지를 나타내고, 입력 게이트(i_t)는 현재 장단기 메모 메모리의 셀 상태(c_t)를 산출하기 위한 현재 입력된 데이터(x_t)의 정보량이며, 출력 게이트(o_t)는 현재 장단기 메모 메모리의 셀 상태(c_t)에서 새로운 히든 상태(h_t)로 전달되는 정보의 양이고, 망각 게이트, 입력 게이트 및 출력 게이트는 아래의 [수학식 2]를 이용하여 산출될 수 있으며, $\tilde{c}_t$는 잠재 상태로 [수학식 3]을 이용하여 산출될 수 있다.

수학식 2

$$i_t = \sigma(\boldsymbol{W}_i \boldsymbol{x}_t + \boldsymbol{U}_i h_{t-1} + \boldsymbol{b}_i)$$

$$f_t = \sigma(\boldsymbol{W}_f \boldsymbol{x}_t + \boldsymbol{U}_f h_{t-1} + \boldsymbol{b}_f)$$

[0045]
$$o_t = \sigma(\boldsymbol{W}_o x_t + \boldsymbol{U}_o h_{t-1} + \boldsymbol{b}_o)$$

[0046] 여기서, W_i, U_i, b_i, W_f, U_f, b_f, W_o, U_o, b_o는 기 결정된 가중치이고, σ는 시그모이드 함수이다.

수학식 3

[0047]
$$\tilde{c}_t = \tanh(\boldsymbol{W}_{\tilde{c}} x_t + \boldsymbol{U}_{\tilde{c}} h_{t-1} + \boldsymbol{b}_{\tilde{c}})$$

[0048] 여기서, $W_{\tilde{c}}$, $U_{\tilde{c}}$, $b_{\tilde{c}}$는 기 결정된 가중치이다.

[0049] 또한, 예측 모델은 산출된 히든 상태를 [수학식 4]에 적용하여 최종적으로 출력되는 유입량 또는 유출량(y_t)을

산출할 수 있다.

수학식 4

[0050] $$y_t = W_d h_t + b_d$$

[0051] 여기서, W_d는 가중치 행렬이고, b_d는 기 결정된 편향 항이다.

[0052] 또한, 학습부(120)는 입력된 훈련 및 검증 그룹에 기 지정된 방법(예를 들어, 베이지안 최적화 알고리즘)을 적용하여 기 지정된 하이퍼 파라미터 범위의 하이퍼 파라미터를 최적화할 수 있다.

[0053] 부연하자면, 학습부(120)는 입력된 훈련 및 검증 그룹의 기 지정된 하이퍼 파라미터 범위에서 초기 하이퍼 파라미터를 선택하고, 예측 모델에 적용하여 평가할 수 있다(S221).

[0054] 또한, 학습부(120)는 초기 하이퍼 파라미터를 기초로 가우시안 프로세스(Gaussian process, GP)를 적용하여 제1 예측 모델($f(\delta)$)을 구축할 수 있다(S222). 이때, 제1 예측 모델은 예측 모델과 선택된 하이퍼 파라미터를 기초로 구축된 모델로, 선택된 하이퍼 파라미터를 이용하여 유입량 및 유출량을 예측하도록 구축될 수 있다.

[0055] 또한, 학습부(120)는 아래의 [수학식 5]를 이용하여 하이퍼 파라미터를 최적화할 수 있다(S223).

수학식 5

[0056] $$EI(\delta_i) = \begin{cases} \left(\mu(\delta_i) - f(\hat{\delta})\right)\Phi\left(\dfrac{\mu(\delta_i) - f(\hat{\delta})}{\sigma(\delta_i)}\right) + \sigma(\delta)\phi\left(\dfrac{\mu(\delta_i) - f(\hat{\delta})}{\sigma(\delta_i)}\right), & if\ \sigma(\delta_i) > 0 \\ 0, & if\ \sigma(\delta_i) = 0 \end{cases}$$

[0057] 여기서, EI는 새로운 하이퍼 파라미터를 선택하기 위해 사용된 획득함수(Acquisition function)이고, $\hat{\delta}$는 현재 선택된 하이퍼 파라미터이고, δ_i는 i번째 선택된 하이퍼 파라미터이며, Φ는 $\dfrac{\mu(\delta_i)-f(\delta^*)}{\sigma(\delta_i)}$에 대한 누적 분포 함수(Cumulative distribution function, CDF)이고, ϕ는 $\dfrac{\mu(\delta_i)-f(\delta^*)}{\sigma(\delta_i)}$에 대한 확률 밀도 함수(Probability density function, PDF)이며, μ(δ)는 평균 유입량 또는 유출량이고, σ(δ)는 분산 값이다.

[0058] 이때, 학습부(120)는 현재 선택된 하이퍼 파라미터보다 더 정확한 예측을 할 것으로 예상되는 하이퍼 파라미터가 없을 때까지($\sigma(\delta_i) = 0$) 상기 S221 내지 S223의 과정을 반복할 수 있다.

[0059] 이로 인하여, 학습부(120)는 최적의 하이퍼 파라미터를 이용하여 예측 모델을 저수지의 유입량 및 유출량을 예측하도록 학습시킬 수 있다.

[0060] 도 4는 본 발명의 다른 실시예에 따른 하이퍼 파라미터의 결과를 도시하는 도면이다.

[0061] 도 4에 도시된 바와 같이, 도출된 최적의 하이퍼 파라미터(빨간 점으로 표시)는 95%의 정확도를 나타냄을 알 수 있다.

[0062] 다음으로, 예측부(130)는 테스트 그룹을 입력받고, 입력된 테스트 그룹에 기 학습된 예측 모델을 적용하고 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량을 예측할 수 있다(S230).

[0063] 구체적으로, 예측부(130)는 이분산 회귀분석과 몬테-카를로 드롭아웃(Monte-carlo dropout)을 통해 우연적 불확실성 및 인식적 불확실성을 정량화하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 산출할 수 있다

(S231).

[0064] 이때, 예측부(130)는 아래의 [수학식 6]을 이용하여 저수지의 유입량 또는 유출량에 대한 예측값과 오차를 산출할 수 있다.

수학식 6

[0065] $$L(\boldsymbol{W},\boldsymbol{U},\boldsymbol{b}) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}exp(-s_t)\left\|y_t^{Obs} - y_t\right\|^2 + \frac{1}{2}s_t$$

[0066] 여기서, W, U 및 b는 예측 모델의 매개변수이고, t는 시간지수이며, T는 테스트 그룹의 총 시간 단계 수이고, y_t^{Obs}는 시간 t에서의 실제 관찰값이고, y_t는 시간 t에서의 예측값이며, s_t는 시간 t에서의 로그 분산값이다.

[0067] 마지막으로, 예측부(130)는 산출된 예측값과 오차를 재정규화하여 저수지의 유입량 또는 유출량을 예측할 수 있다(S232).

[0068] 본 발명의 실시예에 따르면, 불확실성의 원인을 우연적 불확실성 및 인식적 불확실성으로 구분하고, 우연적 불확실성 및 인식적 불확실성을 동시에 정량화하여 저수지의 유입량 또는 유출량을 예측할 수 있다. 또한, 종래에 비해 캐스케이드 저수지 유입량 또는 유출량에 내포된 불확실성을 더 정확하게 예측할 수 있다.

[0069] 다시 말해, 종래의 캐스케이드 저수지 또는 댐 유입량 및 유출량 예측에 있어서 불확실성의 원인을 구분하지 않은 상태로 예측이 수행되었으나, 본 발명의 실시예에 따르면 불확실성을 크게 2개, 즉 우연적 불확실성과 인식적 불확실성으로 구분하고, 각각의 불확실성의 정량화함으로써 보다 정확한 예측이 가능하도록 한다.

[0070] 본 발명은 도면에 도시된 실시예를 참고로 하여 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술이 속하는 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 아래의 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

[0071] 100: 캐스케이드 저수지 유입 및 유출 예측 모델 구축 시스템

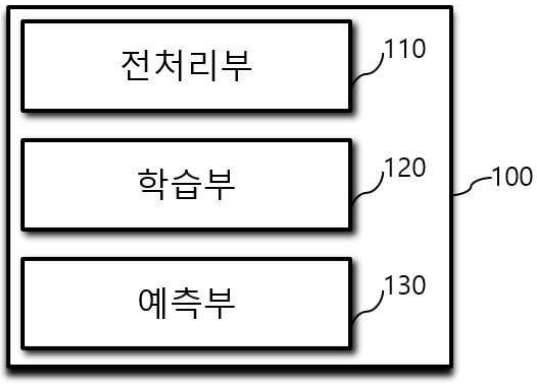
110: 전처리부

120: 학습부

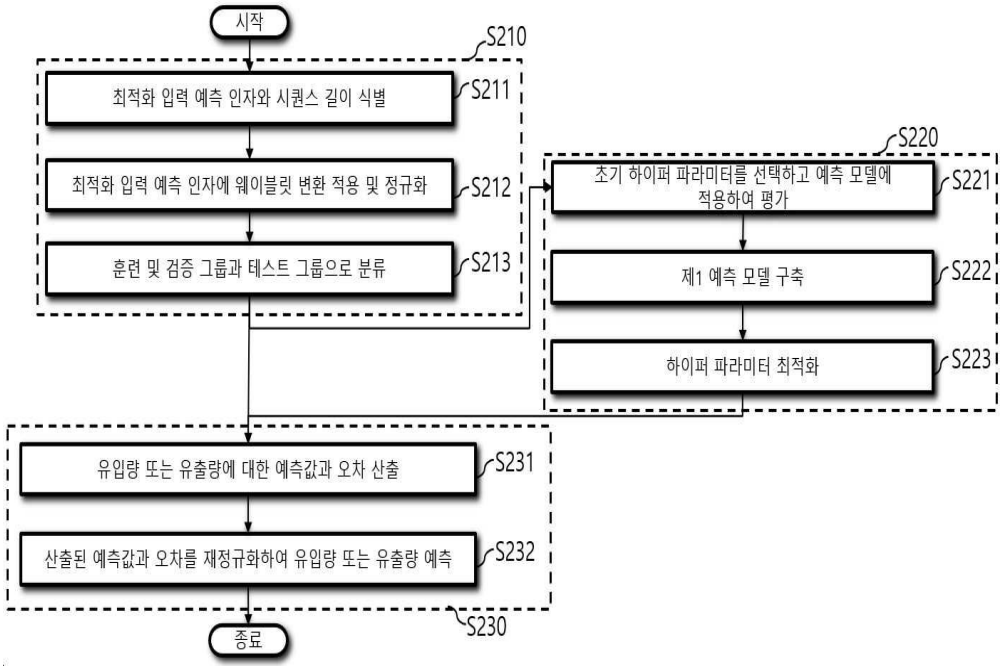
130: 예측부

도면

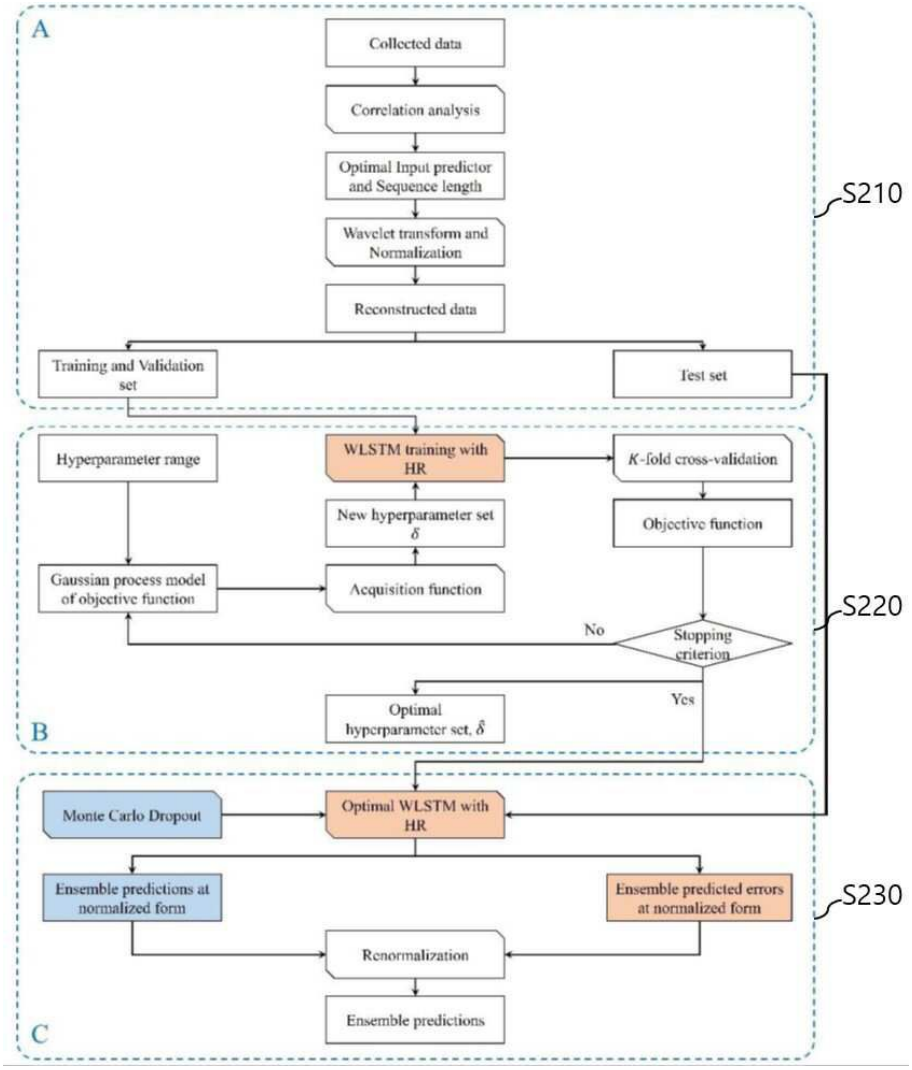
도면1



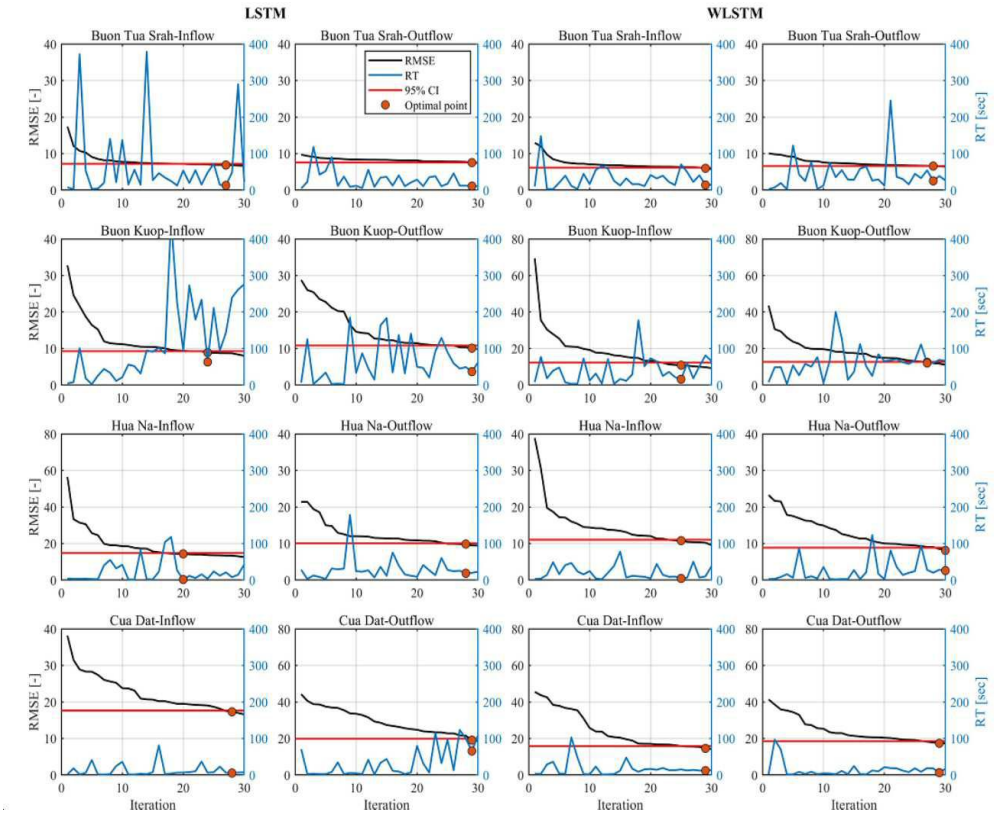
도면2


시작
S210
최적화 입력 예측 인자와 시퀀스 길이 식별
S211
최적화 입력 예측 인자에 웨이블릿 변환 적용 및 정규화
S212
훈련 및 검증 그룹과 테스트 그룹으로 분류
S213
S220
초기 하이퍼 파라미터를 선택하고 예측 모델에 적용하여 평가
S221
제1 예측 모델 구축
S222
하이퍼 파라미터 최적화
S223
유입량 또는 유출량에 대한 예측값과 오차 산출
S231
산출된 예측값과 오차를 재정규화하여 유입량 또는 유출량 예측
S232
S230
종료

도면3


A
Collected data
Correlation analysis
Optimal Input predictor and Sequence length
Wavelet transform and Normalization
Reconstructed data
Training and Validation set
Test set
S210
Hyperparameter range
WLSTM training with HR
K-fold cross-validation
New hyperparameter set δ
Objective function
Gaussian process model of objective function
Acquisition function
No
Stopping criterion
S220
Yes
Optimal hyperparameter set, δ̂
B
Monte Carlo Dropout
Optimal WLSTM with HR
Ensemble predictions at normalized form
Ensemble predicted errors at normalized form
S230
Renormalization
Ensemble predictions
C

도면4


LSTM
WLSTM
Buon Tua Srah-Inflow
Buon Tua Srah-Outflow
Buon Kuop-Inflow
Buon Kuop-Outflow
Hua Na-Inflow
Hua Na-Outflow
Cua Dat-Inflow
Cua Dat-Outflow
RMSE
RT
95% CI
Optimal point
RMSE [-]
RT [sec]
Iteration