



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0042369
(43) 공개일자 2025년03월27일

(51) 국제특허분류(Int. Cl.)
G06N 3/0442 (2023.01) G01F 1/002 (2023.01)
G01W 1/10 (2006.01) G01W 1/14 (2006.01)
G06N 3/04 (2023.01) G06N 3/0985 (2023.01)
(52) CPC특허분류
G06N 3/0442 (2023.01)
G01F 1/002 (2022.01)
(21) 출원번호 10-2023-0125400
(22) 출원일자 2023년09월20일
심사청구일자 2023년09월20일

(71) 출원인
울산대학교 산학협력단
울산광역시 남구 대학로 93(무거동)
(72) 발명자
김종호
울산광역시 남구 봉월로8번길 20, 202호 (신정동)
전 득 충
울산광역시 남구 대학로52번길 5-3, 103호 (무거동)
(74) 대리인
특허법인태백

전체 청구항 수 : 총 11 항

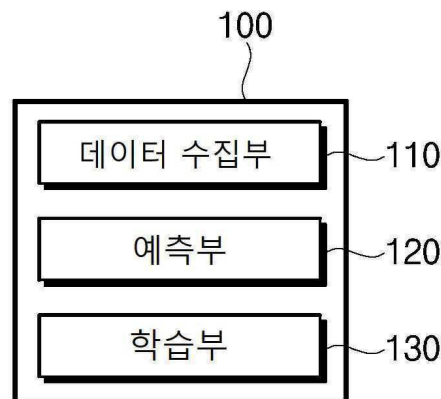
(54) 발명의 명칭 장기 예측을 위한 LSTM 기반 기계 학습 시스템

(57) 요약

본 발명은 장기 예측을 위한 LSTM 기반 기계 학습 시스템에 관한 것이다. 본 발명에 따르면, 현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 강수량 데이터와 시계열의 유량 데이터, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열 강수량 데이터를 포함한 입력 변수 데이터를 획득하는 데이터 수집부; 및 상기 입력 변수 데이터를 기 학습된 LSTM 모델에 적용하여 상기 미래 시간 구간에 대한 시계열의 유량 데이터를 예측하여 출력 변수 데이터로 제공하는 예측부를 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템을 제공한다.

본 발명에 따르면, LSTM에 기반한 체계화된 기계 학습 구조를 사용하여 과거 강수량과 유량을 바탕으로 미래 예상되는 하천의 유량을 효율적으로 장기 예측할 수 있다.

대표도 - 도1



(52) CPC특허분류

G01W 1/10 (2013.01)

G01W 1/14 (2013.01)

G06N 3/049 (2023.01)

G06N 3/0985 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711186324

과제번호 2022R1A2C2008584

부처명 과학기술정보통신부

과제관리(전문)기관명 한국연구재단

연구사업명 개인기초연구(과기정통부)

연구과제명 연구 패러다임의 전환:외삽능력을 탑재한 4세대 Physics-informed머신러닝-대체모형 개발

기 여 율 1/1

과제수행기관명 울산대학교

연구기간 2023.03.01 ~ 2024.02.29

명세서

청구범위

청구항 1

현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 강수량 데이터와 시계열의 유량 데이터, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열 강수량 데이터를 포함한 입력 변수 데이터를 획득하는 데이터 수집부; 및

상기 입력 변수 데이터를 기 학습된 LSTM 모델에 적용하여 상기 미래 시간 구간에 대한 시계열의 유량 데이터를 예측하여 출력 변수 데이터로 제공하는 예측부를 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 2

청구항 1에 있어서,

상기 설정 길이는 지체 시간(lag-time)을 나타내는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 3

청구항 1에 있어서,

상기 데이터 수집부는,

기후 예측 모델인 CMIP6 모델의 예측 데이터를 저장한 DB에 접속하여 상기 CMIP6 모델에 의해 예상된 상기 지역의 미래 시간 구간에 대한 시계열 강수량 데이터를 수집하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 4

청구항 1에 있어서,

해당 지역의 과거 빅데이터를 기반으로 상기 LSTM 모델을 하이퍼 파라미터 최적화를 통해 딥러닝을 기반으로 사전 학습시키는 학습부를 더 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 5

청구항 4에 있어서,

상기 학습부는,

베이지안 최적화 알고리즘을 이용하여 상기 LSTM 모델의 하이퍼 파라미터와 상기 입력 변수의 개수를 동시에 최적화하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 6

청구항 1에 있어서,

상기 LSTM 모델은 인접 셀 간에 서로 연결된 총 K개의 LSTM 셀에서 각 셀당 한 번의 실행으로 1개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해 총 K개의 예측 결과를 시간에 따라 점진적으로 도출하는 싱글-태스크 프로그레시브-런(single-task progressive-run) 구조의 spLSTM 모델로 구현되며,

상기 spLSTM 모델은,

k번째 LSTM 셀에서 도출된 t+k 시간의 하천의 유량 데이터를 이용하여 다음의 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하고, 기후 예측 DB로부터 획득된 t+k 시간의 하천의 강수량 데이터를 이용하여 상기 k+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정한 상태에서 상기 k+1 번째 LSTM 셀을 학습하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 7

청구항 6에 있어서,

상기 spLSTM 모델은,

다음의 k+1번째 LSTM 셀의 실행 시에는 이전에 도출된 t+k 시간의 하천의 유량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 유량 데이터 중에서 가장 최근 값으로 추가하는 동시에 가장 과거 값을 소거하는 방식으로 상기 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하면서,

기후 예측 DB로부터 획득된 t+k 시간의 하천의 강수량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 강수량 데이터 중에서 가장 최근 값으로 추가하는 동시에 가장 과거 값을 소거하는 방식으로 상기 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 8

청구항 7에 있어서,

상기 spLSTM 모델은,

아래 수학적식과 같이, t-L+k-1 시간 내지 t+K-1 시간 까지의 예측 강수량 데이터, 그리고 t-L+K-1 시간 내지 t+K-1 시간 까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, t+K 시간의 미래 시간의 하천의 유량 데이터를 출력 변수 데이터로 제공하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템:

$$y_{t+k}^{sim} = spLSTM_{t+k-1}(x_{t-L+k-1}^{sim}, \dots, x_{t+k-1}^{sim}, y_{t-L+k-1}^{sim}, \dots, y_{t+k-1}^{sim})$$

여기서, y_{t+k}^{sim} 은 t+k 시간의 예측된 유량, x^{sim} 는 예측 강수량, L은 지체 시간(lag-time), K는 출력 변수의 개수 이자 리드 타임(lead time)의 길이, k는 K의 인덱스(k=1,2,3,...,K)를 나타낸다.

청구항 9

청구항 1에 있어서,

상기 LSTM 모델은 인접 셀 간에 서로 연결된 총 J+1개의 LSTM 셀(J는 점프-런의 횟수)에서 각 셀당 한 번의 실행으로 M개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해, 총 (J+1)×M 개의 예측 결과를 시간에 따라 도출하는 멀티-태스크 점프-런(multi-task jump-run) 구조의 mjLSTM 모델로 구현되며,

상기 mjLSTM 모델은,

J+1개의 LSTM 셀 중에서 j번째 LSTM 셀에서 도출된 [t + (j-1)×M + 1] 내지 [t + (j-1)×M + M] 시간의 M개의 하천의 유량 데이터를 다음의 j+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터로 활용하고, 기후 예측 DB로부터 획득된 미래의 M개의 하천의 강수량 데이터를 상기 j+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터로 활용하여, 상기 j+1 번째 LSTM 셀을 학습하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

청구항 10

청구항 9에 있어서,

상기 mjLSTM 모델은,

아래 수학적식과 같이, [t + j×M - M + 1] 내지 [t + j×M] 시간 까지의 예측 강수량 데이터, 그리고 [t + j×M - M + 1] 내지 [t + j×M] 시간 까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, [t + j×M + 1] 내지 [t + j×M + M] 시간의 미래 시간의 하천의 유량 데이터를 출력 변수 데이터로 제공하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템:

$$(y_{t+j\times M+1}^{sim}, \dots, y_{t+j\times M+M}^{sim}) = mjLSTM_{t+j\times M}(x_{t+j\times M-M+1}^{sim}, \dots, x_{t+j\times M}^{sim}, y_{t+j\times M-M+1}^{sim}, \dots, y_{t+j\times M}^{sim})$$

여기서, y^{sim} 은 예측 유량, x^{sim} 는 예측 강수량, L은 지체 시간(lag-time), j는 점프-런의 인덱스(j=1,...,J)로, j 번째 점프-런을 나타내며, J는 점프-런의 횟수를 나타낸다.

청구항 11

현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 제1 인자 측정값 및 시계열의 제2 인자 측정값, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열의 제1 인자 예측값을 포함한 입력 변수 데이터를 획득하는 데이터 수집부; 및

상기 입력 변수 데이터를 기 학습된 LSTM 모델에 적용하여 상기 미래 시간 구간에 대한 시계열의 제2 인자 예측값을 예측하여 출력 변수 데이터로 제공하는 예측부를 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 장기 예측을 위한 LSTM 기반 기계 학습 시스템에 관한 것으로서, 보다 상세하게는 체계화된 기계 학습 구조를 사용하여 하천의 유량을 장기 예측할 수 있는 장기 예측을 위한 LSTM 기반 기계 학습 시스템에 관한 것이다.

배경 기술

[0002] 하천 흐름 예측은 수문 관리의 기본이며 저수지의 안전한 제어 및 운영을 위한 필수적인 도구이다. 효과적인 저수지 관리는 홍수, 가뭄 등의 극한 상황에서 지역사회의 회복력, 지속 가능한 물 공급, 도시의 홍수 방지, 깨끗하고 재생 가능한 전력 생산을 가능하게 한다. 하천 흐름을 정확하게 예측하지 못하면 최근 오로빌 저수지 유출로 붕괴에서 목격된 것과 같은 저수지 오작동이 발생할 수 있다. 하천 흐름의 단기 예측과 함께 유역 내의 기후 변화를 포착하는 하천 흐름의 장기 예측은 물 관리 전략을 알리는 데 도움이 된다.

[0003] 하천의 흐름 예측을 위하여 다양한 모델이 개발되었으며, 프로세스 기반과 데이터 기반의 두 가지 범주로 나눌 수 있다. 1차원, 2차원 및 3차원 시뮬레이션 기능을 갖춘 프로세스 기반 모델은 이론적으로 관찰 가능하고 확장 가능한 상태 변수와 플럭스로 구성된 물리적 원리를 기반으로 한다. 이들은 과정의 공간적 가변성, 경계 조건, 유역의 물리적 특성을 명시적으로 다루고 있다. 그런데, 이러한 모델들은 물리적 프로세스에 대한 통찰력을 제공할 수 있지만 대규모 데이터 세트가 필요하고 종종 엄청난 계산 비용이 발생하고, 상당한 불확실성을 수반할 수 있다.

[0004] 최근 몇 년 동안 컴퓨팅 과학의 획기적인 발전으로 딥러닝 알고리즘으로 무장한 데이터 기반 모델이 시계열 예측에서 인기를 얻고 있다. 이러한 모델은 최소한의 과거 데이터를 기반으로 하며 기본 물리적 프로세스에 대한 지식이 필요하지 않는다. 기계학습(Machine learning, ML) 기술은 물리적 모델에 비해 상대적으로 덜 복잡하여 구현이 쉽고 낮은 계산 비용으로 만족스러운 성능을 제공한다.

[0005] 시계열 예측에는 컨볼루션 신경망, 순환 신경망, LSTM(long short-term memory), GRU(Gated Recurrent Unit) 등의 딥러닝 모델이 사용되며 여러 기술이 통합된 형태로 개발되고 있다. 가장 널리 사용되는 하이브리드 기술은 웨이블릿 변환(입력 예측기를 위한 분해 기술)과 LSTM 또는 GRU 알고리즘의 조합으로, 복잡성을 크게 증가시키지 않으면서 시계열의 극단값에 대한 정확한 예측을 제공한다. 최근 주목받고 있는 하이브리드 기법으로, 중요한 입력 예측기를 선택하는데 강력한 성능을 보이는 컨볼루션 신경망은 높은 시계열 예측력을 가진 LSTM 또는 GRU 기법과 결합될 수 있다. 훈련 데이터 세트와 유사하지 않은 기상 이변과 같은 극한 상황에 대한 예측 추론 능력을 향상시킬 수 있는 모델도 개발되고 있다.

[0006] streamflow forecast, hindcasting, forecasting, projection 등을 포함한 많은 시계열 예측에는 이들을 명확히 구분하지 않고 사용한다. hindcasting은 과거 데이터를 입력으로 사용하여 관찰된 값(참값)이 존재하는 현상을 재현하고 모델의 정확성을 테스트하고, 과거 이벤트에 대한 입력이 모델에 입력되어 모델이 출력을 얼마나 잘 생성하는지를 검사한다. hindcasting을 사용하여 미래 데이터를 예측하는 것을 예측된 미래 기간에 따라 forecasting 또는 projection이라고 한다.

[0007] 여기서 리드 타임(Lead time)이 클수록 예측의 오차가 커지기 때문에 리드 타임은 짧은 경향이 있다. 예를 들어, 수문 분야 일일 예측에 대한 최대 리드타임은 약 14일이다. 따라서 예측은 일반적으로 단기 예측을 포함하며, 프로세스 기반 및 데이터 기반 모델을 사용하여 날씨, 홍수 및 하천 유량 예측에 광범위하게 적용된다.

[0008] 프로젝션(projection)은 수년에서 심지어 한 세기에 이르는 장기 예측을 의미한다. 일반적으로 프로젝션은 기후 변수에 대응하기 위해 장기 기후 프로젝션 데이터를 사용하여 장기적인 유량과 댐 유입을 예측한다. 프로세스

기반 모델을 사용하는 연구가 일반적이지만 데이터 기반 모델을 사용한 연구는 드물다.

- [0009] 데이터 기반 모델은 정확성, 효율성, 속도 및 간단한 셋업 구조로 인해 유량, 댐 유입 및 홍수 예측에 점점 더 많이 사용되고 있다.
- [0010] 기존의 일반적인 ML 방식을 이용한 예측에서는 단순히 과거의 관측값을 입력 데이터로 하여 미래의 예측 값 도출하도록 구현된다. 그런데, 이러한 기존의 ML 알고리즘을 위한 함수 구조는 예측을 위한 최적의 형태로 변경되어야 한다.
- [0011] 본 발명의 배경이 되는 기술은 한국등록특허 제10-2409155호(2022.06.16 공고)에 개시되어 있다.

발명의 내용

해결하려는 과제

- [0012] 본 발명은 LSTM를 기반으로 하는 체계화된 기계 학습 구조를 사용하여 과거 강수량과 유량을 바탕으로 하천의 유량을 효율적으로 장기 예측할 수 있는 장기 예측을 위한 LSTM 기반 기계 학습 시스템을 제공하는데 목적이 있다.

과제의 해결 수단

- [0013] 본 발명은, 현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 강수량 데이터와 시계열의 유량 데이터, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열 강수량 데이터를 포함한 입력 변수 데이터를 획득하는 데이터 수집부; 및 상기 입력 변수 데이터를 학습된 LSTM 모델에 적용하여 상기 미래 시간 구간에 대한 시계열의 유량 데이터를 예측하여 출력 변수 데이터로 제공하는 예측부를 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템을 제공한다.
- [0014] 또한, 상기 설정 길이는 지체 시간(lag-time)을 나타낼 수 있다.
- [0015] 또한, 상기 데이터 수집부는, 기후 예측 모델인 CMIP6 모델의 예측 데이터를 저장한 DB에 접속하여 상기 CMIP6 모델에 의해 예상된 상기 지역의 미래 시간 구간에 대한 시계열 강수량 데이터를 수집할 수 있다.
- [0016] 또한, 상기 LSTM 기반 기계 학습 시스템은, 해당 지역의 과거 빅데이터를 기반으로 상기 LSTM 모델을 하이퍼 파라미터 최적화를 통해 딥러닝을 기반으로 사전 학습시키는 학습부를 더 포함할 수 있다.
- [0017] 또한, 상기 학습부는, 베이지안 최적화 알고리즘을 이용하여 상기 LSTM 모델의 하이퍼 파라미터와 상기 입력 변수의 개수를 동시에 최적화할 수 있다.
- [0018] 또한, 상기 LSTM 모델은 상기 LSTM 모델은 인접 셀 간에 서로 연결된 총 K개의 LSTM 셀에서 각 셀당 한 번의 실행으로 1개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해 총 K개의 예측 결과를 시간에 따라 점진적으로 도출하는 싱글-태스크 프로그레시브-런(single-task progressive-run) 구조의 spLSTM 모델로 구현되며, 상기 spLSTM 모델은, k번째 LSTM 셀에서 도출된 t+k 시간의 하천의 유량 데이터를 이용하여 다음의 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하고, 기후 예측 DB로부터 획득된 t+k 시간의 하천의 강수량 데이터를 이용하여 상기 k+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정한 상태에서 상기 k+1 번째 LSTM 셀을 학습할 수 있다.
- [0019] 또한, 상기 spLSTM 모델은, 다음의 k+1번째 LSTM 셀의 실행 시에는 이전에 도출된 t+k 시간의 하천의 유량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 유량 데이터 중에서 가장 최근 값으로 추가하는 동시에 가장 과거 값을 소거하는 방식으로 상기 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하면서, 기후 예측 DB로부터 획득된 t+k 시간의 하천의 강수량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 강수량 데이터 중에서 가장 최근 값으로 추가하는 동시에 가장 과거 값을 소거하는 방식으로 상기 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정할 수 있다.
- [0020] 또한, 상기 spLSTM 모델은, 아래 수식식과 같이, t-L+k-1 시간 내지 t+K-1 시간 까지의 예측 강수량 데이터, 그리고 t-L+K-1 시간 내지 t+K-1 시간 까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, t+K 시간의 미래 시간의 하천의 유량 데이터를 출력 변수 데이터로 제공할 수 있다.

[0021]
$$y_{t+k}^{sim} = spLSTM_{t+k-1}(x_{t-L+k-1}^{sim}, \dots, x_{t+k-1}^{sim}, y_{t-L+k-1}^{sim}, \dots, y_{t+k-1}^{sim})$$

[0022] 여기서, y_{t+k}^{sim} 은 $t+k$ 시간의 예측된 유량, x^{sim} 는 예측 강수량, L 은 지체 시간(lag-time), K 는 출력 변수의 개수이자 리드 타임(lead time)의 길이, k 는 K 의 인덱스($k=1,2,3,\dots,K$)를 나타낸다.

[0023] 또한, 상기 LSTM 모델은 인접 셀 간에 서로 연결된 총 $J+1$ 개의 LSTM 셀(J 는 점프-런의 횟수)에서 각 셀당 한 번의 실행으로 M 개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해, 총 $(J+1) \times M$ 개의 예측 결과를 시간에 따라 도출하는 멀티-태스크 점프-런(multi-task jump-run) 구조의 mjLSTM 모델로 구현되며, 상기 mjLSTM 모델은, $J+1$ 개의 LSTM 셀 중에서 j 번째 LSTM 셀에서 도출된 $[t + (j-1) \times M + 1]$ 내지 $[t + (j-1) \times M + M]$ 시간의 M 개의 하천의 유량 데이터를 다음의 $j+1$ 번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터로 활용하고, 기후 예측 DB로부터 획득된 미래의 M 개의 하천의 강수량 데이터를 상기 $j+1$ 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터로 활용하여, 상기 $j+1$ 번째 LSTM 셀을 학습할 수 있다.

[0024] 또한, 상기 mjLSTM 모델은, 아래 수식과 같이, $[t + j \times M - M + 1]$ 내지 $[t + j \times M]$ 시간까지의 예측 강수량 데이터, 그리고 $[t + j \times M - M + 1]$ 내지 $[t + j \times M]$ 시간까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, $[t + j \times M + 1]$ 내지 $[t + j \times M + M]$ 시간의 미래 시간의 하천의 유량 데이터를 출력 변수 데이터로 제공할 수 있다.

$$[0025] (y_{t+j \times M+1}^{sim}, \dots, y_{t+j \times M+M}^{sim}) = mjLSTM_{t+j \times M}(x_{t+j \times M-M+1}^{sim}, \dots, x_{t+j \times M}^{sim}, y_{t+j \times M-M+1}^{sim}, \dots, y_{t+j \times M}^{sim})$$

[0026] 여기서, y^{sim} 은 예측 유량, x^{sim} 는 예측 강수량, L 은 지체 시간(lag-time), j 는 점프-런의 인덱스($j=1,\dots,J$)로, j 번째 점프-런을 나타내며, J 는 점프-런의 횟수를 나타낸다.

[0027] 그리고, 본 발명은, 현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 제1 인자 측정값 및 시계열의 제2 인자 측정값, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열의 제1 인자 예측값을 포함한 입력 변수 데이터를 획득하는 데이터 수집부; 및 상기 입력 변수 데이터를 기 학습된 LSTM 모델에 적용하여 상기 미래 시간 구간에 대한 시계열의 제2 인자 예측값을 예측하여 출력 변수 데이터로 제공하는 예측부를 포함하는 장기 예측을 위한 LSTM 기반 기계 학습 시스템을 제공한다.

발명의 효과

[0028] 본 발명에 따르면, LSTM를 기반으로 하는 체계화된 기계 학습 구조를 사용하여 과거 강수량과 유량을 바탕으로 하천의 유량을 효율적으로 장기 예측할 수 있다.

도면의 간단한 설명

[0029] 도 1은 본 발명의 실시예에 따른 장기 예측을 위한 LSTM 기반 기계 학습 시스템의 구성을 나타낸 도면이다.

도 2는 본 발명의 실시예에 따른 장기 예측에 적용되는 spLSTM 모델을 구체적으로 설명한 도면이다.

도 3는 본 발명의 실시예에 따른 장기 예측에 적용되는 mjLSTM 모델을 구체적으로 설명한 도면이다.

도 4는 한강유역에 속한 소양유역의 연구영역과 강우량계의 위치를 나타낸 도면이다.

도 5는 본 발명의 실시예에 따른 두 가지 구조의 LSTM 모델을 이용한 단기 예측 결과 성능을 보여주는 도면이다.

도 6은 본 발명의 실시예에 따른 두 가지 구조의 LSTM 모델을 이용한 장기 예측 결과 성능을 보여주는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0030] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시 예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0031] 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결"되어 있는 경우도 포함한다. 또한 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.

- [0032] 본 발명은 LSTM 기반의 기계 학습 시스템에 관한 것으로, 특히 장기 예측에 특화된 두 가지 구조의 LSTM 구조를 제안한다. 제안한 시스템에 따르면, LSTM을 기반으로 하는 체계화된 기계 학습 구조를 사용하여 과거 강수량과 유량 데이터를 바탕으로 하천의 유량을 장기 예측할 수 있다.
- [0033] 도 1은 본 발명의 실시예에 따른 장기 예측을 위한 LSTM 기반 기계 학습 시스템의 구성을 나타낸 도면이다.
- [0034] 도 1에 나타난 것과 같이, 본 발명의 실시예에 따른 LSTM 기반의 기계 학습 시스템(100)은 데이터 수집부(110) 및 예측부(120)를 포함하며, 학습부(130)를 더 포함할 수 있다. 여기서 각 부(110,120,130)의 동작과 각 부 간의 데이터 흐름은 제어부(미도시)에 의해 제어될 수 있다.
- [0035] 데이터 수집부(110)는 현재 시간으로부터 설정 길이의 과거 시간에 관측된 지역의 시계열의 강수량 데이터와 시계열의 유량 데이터, 그리고 현재 시간으로부터 설정된 시간 길이의 미래 시간 구간에 대해 예상된 상기 지역의 미래 시계열 강수량 데이터를 포함한 총 3가지의 입력 변수 데이터를 획득한다.
- [0036] 여기서, 과거 시간에 대한 설정 길이(L)는 유역의 지체 시간(lag-time)을 의미할 수 있다. 아울러, 미래 시간 구간에 대한 설정된 시간 길이란, 리드 타임(lead time)의 길이를 나타낼 수 있다.
- [0037] 데이터 수집부(110)는 입력 변수 데이터를 획득하고 이를 예측부(120)로 전달할 수 있다. 여기서, 데이터 수집부(110)는 네트워크 연결된 사용자 단말, 저장 장치, 외부 서버, 기상 서버, 외부 데이터베이스 등으로부터 관련 데이터를 수집 또는 입력받을 수 있다.
- [0038] 특히, 데이터 수집부(110)는 미래 시간 구간의 예상된 미래 시계열 강수량의 경우, 기 알려진 기후 예측 모델인 CMIP6 모델을 통해서 획득할 수 있다. 구체적으로, 데이터 수집부(110)는 CMIP6 모델의 예측 데이터를 저장한 DB(기후 예측 DB)에 접속하여 CMIP6 모델에 의해 예상된 지역의 미래 시간 구간에 대한 시계열 강수량 데이터를 수집할 수 있다.
- [0039] 그리고, 예측부(120)는 수집된 입력 변수 데이터를 기 학습된 LSTM 모델에 적용하여, 미래 시간 구간에 대한 시계열의 유량 데이터를 예측하여 출력 변수 데이터로 제공한다.
- [0040] 도 2는 본 발명의 실시예에 따른 장기 예측에 적용되는 spLSTM 모델을 구체적으로 설명한 도면이고, 도 3는 본 발명의 실시예에 따른 장기 예측에 적용되는 mjLSTM 모델을 구체적으로 설명한 도면이다.
- [0041] 본 발명의 실시예에서 장기 예측을 위한 LSTM 모델은 도 2와 같이 싱글-태스크 프로그레시브-런(single-task progressive-run) 구조의 spLSTM 모델로 구현될 수 있고, 도 3과 같이 멀티-태스크 점프-런(multi-task jump-run) 구조의 mjLSTM 모델로 구현될 수도 있다.
- [0042] 먼저, 본 발명의 실시예에 따른 장기 예측을 위한 LSTM 모델이 싱글-태스크 프로그레시브-런 구조의 spLSTM 모델로 구현된 경우를 설명한다.
- [0043] 도 2에 나타난 것과 같이, spLSTM 모델의 경우, 인접 셀 간에 서로 연결된 총 K개의 LSTM 셀에서 각 셀당 한 번의 실행으로 1개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해 총 K개의 예측 결과를 시간에 따라 점진적으로 도출하는 싱글-태스크 프로그레시브-런(single-task progressive-run) 구조를 갖는다.
- [0044] 이 구조의 경우 하나의 셀에서 하나의 예측 결과를 도출함으로써 총 K개의 예측 결과(outcome)가 도출되는 구조로, 셀당 한번의 실행으로 한 개의 결과가 얻어지므로 싱글 태스크에 해당한다. 또한, 이전 셀에서 도출된 1개의 예측 값을 현재 실행 단계의 입력 데이터로 삽입하여 활용하게 된다.
- [0045] 도 2에 나타난 것과 같이, spLSTM 모델은 k번째 LSTM 셀에서 도출된 t+k 시간의 하천의 유량 데이터를 이용하여 다음의 k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하고, 이와 함께, 기후 예측 DB로부터 획득된 t+k 시간의 예측된 하천의 강수량 데이터를 이용하여 k+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정하여 사용함으로써, k+1 번째 LSTM 셀을 학습할 수 있다.
- [0046] 즉, spLSTM 모델은, 다음의 k+1번째 LSTM 셀의 실행 시에는 이전 셀에서 도출된 유량 예측값과 기후 예측 DB를 통해 얻은 강수량 예측값을 기반으로, k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정하여 학습에 사용할 수 있다.

[0047] 구체적으로, 도 2에서 처럼, spLSTM 모델은, 다음의 k+1번째 LSTM 셀의 실행 시에는 이전에 도출된 t+k 시간의 하천의 유량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 유량 데이터 중에서 가장 최근 값으로 추가하는 동시에, 가장 과거 값을 소거하는 방식으로, k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터를 수정하여 사용하면서, 아울러 기후 예측 DB로부터 획득된 t+k 시간의 하천의 강수량 데이터를 기존 k번째 LSTM 셀에 사용된 과거 하천의 시계열 강수량 데이터 중에서 가장 최근 값으로 추가하는 동시에, 가장 과거 값을 소거하는 방식으로, k+1번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터를 수정할 수 있다.

[0048] 결과적으로, spLSTM 모델은, 아래 수학적 식 1과 같이, t-L+k-1 시간 내지 t+K-1 시간 까지의 예측 강수량 데이터, 그리고 t-L+K-1 시간 내지 t+K-1 시간 까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, t+k 시간의 미래의 하천의 유량 데이터를 출력 변수 데이터로 제공할 수 있다.

수학적 식 1

$$[0049] \quad y_{t+k}^{sim} = spLSTM_{t+k-1}(x_{t-L+k-1}^{sim}, \dots, x_{t+k-1}^{sim}, y_{t-L+k-1}^{sim}, \dots, y_{t+k-1}^{sim})$$

[0050] 여기서, y_{t+k}^{sim} 은 t+k 시간의 예측된 유량, x^{sim} 는 예측 강수량, L은 지체 시간(lag-time), K는 출력 변수의 개수이자 리드 타임(lead time)의 길이, k는 K의 인덱스로, k={1,2,3,...,K}로 표현될 수 있다.

[0051] 이러한 spLSTM 모델의 경우, 기상학적 강제력에 대한 예측값을 입력으로 활용하여 지속적으로 스스로 업데이트되면서 시간 변동 모델을 구현하게 된다. 본 발명의 실시예에서는 CMIP6 데이터베이스로부터 다운스케일된 미래 강수량 예측 데이터를 기상학적 강제력에 대한 예측값으로 사용하였다.

[0052] 다음으로, 본 발명의 실시예에 따른 장기 예측을 위한 LSTM 모델이 멀티-태스크 점프-런 구조의 mjLSTM 모델로 구현된 경우를 설명한다.

[0053] spLSTM 모델은 예측 변수의 수가 KSTM 셀의 수와 동일한 K개를 갖는다. 합리적인 수의 K개 예측치를 도출하는 단기 예측에서는 LSTM 모델을 실행하는 계산 과정이 큰 부담이 되지 않지만, 단기 예측에 비해 K값이 큰 장기 예측의 경우 수학적 식 1의 방식은 계산 부담이 크게 증가할 수 있다.

[0054] 따라서 본 발명의 실시예에서는 spLSTM 모델에 비해 계산 시간과 비용을 줄일 수 있는 mjLSTM 모델로 구현된 LSTM을 추가적으로 제안한다. mjLSTM 구조에서 LSTM 셀은 spLSTM 구조에서 처럼 각 시간 단계가 아닌, 특정 시간 간격(jump) 단위로 업데이트되도록 설계되었다.

[0055] 즉, 앞서 상술한 spLSTM과 다른 mjLSTM 구조의 특징은 LSTM 셀당 M개의 유량 예측치를 도출하며 J+1개의 LSTM 셀을 통해 셀간 점프-런하면서 LSTM 모델을 학습함으로써 결과적으로 낮은 시간과 비용으로 총 (J+1)×M 개의 시계열 유량 예측치를 도출할 수 있다.

[0056] 도 3에 나타낸 것과 같이, mjLSTM 모델의 경우, 인접 셀 간에 서로 연결된 총 J+1개의 LSTM 셀(J는 점프-런의 횟수)에서 각 셀당 한 번의 실행으로 M개의 예측 결과를 도출 후 다음 셀의 입력으로 제공하여 학습하는 것을 통해, 총 (J+1)×M 개의 예측 결과를 시간에 따라 도출하는 멀티-태스크 점프-런(multi-task jump-run) 구조를 갖는다. 이때, (J+1)×M은 mjLSTM 모델의 출력 변수의 개수이자 리드 타임(lead time)의 길이가 된다. 또한, 이 구조의 경우 하나의 셀당 한 번의 실행으로 M개의 결과(outcome)가 도출되는 구조로, 실행 하나당 여러 개의 결과가 얻어지므로 멀티 태스크에 해당한다.

[0057] 이러한 mjLSTM 모델은, 실질적으로 J+1개의 LSTM 셀 중에서 j번째 LSTM 셀에서 도출된 [t + (j-1)×M + 1] 내지 [t + (j-1)×M + M] 시간의 M개의 하천의 유량 데이터를 다음의 j+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 유량 데이터로 활용하고, 기후 예측 DB로부터 획득된 미래의 M개의 하천의 강수량 데이터를 j+1 번째 LSTM 셀에 입력되는 과거 하천의 시계열 강수량 데이터로 활용하여, j+1 번째 LSTM 셀을 학습한다.

[0058] 쉽게 말해서, j+1 번째 LSTM 셀의 경우, 바로 앞의 j번째 LSTM 셀에서 학습을 통해 도출된 다음의 M개 시간의 유량 예측치와 기후 예측 DB에서 얻은 동일한 M개 시간의 강수량 예측치를 입력 데이터로 활용하여 학습을 수행하게 된다. 이에 따라, 첫 번째 LSTM 셀을 제외한 나머지 LSTM 셀에서는 관측치 대신에 시뮬레이션 예측치들이 입력 변수 데이터로 활용된다.

[0059] 결과적으로, 이러한 mjLSTM 모델은, 아래 수학적 식 2와 같이, [t + j×M - M + 1] 내지 [t + j×M] 시간 까지의

예측 강수량 데이터, 그리고 $[t + j \times M - M + 1]$ 내지 $[t + j \times M]$ 시간 까지의 예측 유량 데이터를 포함한 입력 변수 데이터를 딥러닝 분석하여, $[t + j \times M + 1]$ 내지 $[t + j \times M + M]$ 시간의 미래 시간의 하천의 유량 데이터를 출력 변수 데이터로 제공할 수 있다.

수학식 2

$$(y_{t+j \times M+1}^{sim}, \dots, y_{t+j \times M+M}^{sim}) = mjLSTM_{t+j \times M}(x_{t+j \times M-M+1}^{sim}, \dots, x_{t+j \times M}^{sim}, y_{t+j \times M-M+1}^{sim}, \dots, y_{t+j \times M}^{sim})$$

여기서, y^{sim} 은 예측 유량, x^{sim} 은 예측 강수량, L 은 지체 시간(lag-time), j 는 점프-런의 인덱스($j=1, \dots, J$)로, j 번째 점프-런을 나타내며, J 는 점프-런의 횟수를 나타낸다.

점프-런이 총 J 회라면, LSTM 셀은 총 $J+1$ 개 존재하며, 각 LSTM 셀당 한 번의 실행으로 M 의 변수가 도출되므로, mjLSTM 모델을 통해 총 $(J+1) \times M$ 개의 예측 값이 도출될 수 있다.

점프 후 예측되는 변수는 $t+1$ 에서 $t+(J+1) \times M$ 이고, 첨자 $t+(J+1) \times M$ 은 수학식 2에서 첨자 $t+K$ 에 해당한다. 즉, 이 경우 리드 타임 $K=(J+1) \times M$ 이다. 예를 들어, 99번의 점프-런을 수행하고($J = 99$), 각 점프-런에서 10개의 테스트(변수)를 예측하였다면($M = 10$), 이는 총 1,000개($=(J+1) \times M$)의 미래 변수 값을 예측하는 것과 같다.

상술한 구조에 따라, 제안한 바와 같은 splSTM 모델과 mjLSTM 모델은 "time-variant"한 특성(originality)을 가지며, 새로운 입력 자료로부터 splSTM 모델 또는 mjLSTM 모델이 새롭게 지속적으로 업데이트 및 학습될 수 있다.

본 발명에 의해 제안한 LSTM 모델은 학습부(130)에 의해 사전 학습될 수 있다. 학습부(130)는 해당 지역의 과거 빅데이터를 기반으로 LSTM 모델을 딥러닝을 기반으로 사전 학습시킬 수 있다. 또한, 학습부(130)는 베이지안 최적화(Bayesian Optimization, BO)를 기반으로 하이퍼 파라미터 최적화를 수행하여 딥러닝 학습되는 LSTM 모델 전체를 최적화시킬 수 있다.

본 발명의 실시예에서, 학습부(130)는 베이지안 최적화 알고리즘을 사용하여 LSTM 모델의 하이퍼 파라미터와 입력 변수의 개수를 동시에 최적화하도록 한다.

본 발명의 실시예에서 하이퍼 파라미터(초매개변수)는 LSTM 모델의 은닉층(hidden layer)의 수, 은닉층을 구성한 은닉 유닛(hidden unit)의 수, 에포크(epoch)의 수, 드랍아웃 비율(dropout rate), 그리고 배치 사이즈(batch size) 중 적어도 하나를 포함할 수 있다.

위와 같이 본 발명의 실시예에서 학습 모델의 하이퍼 파라미터와 입력 변수의 개수를 동시 최적화하는 이유는 다음과 같다. 기계 학습 모델의 신뢰성 있는 예측력을 보장하기 위해서는 훈련 전 모델의 구성과 관련된 입력 변수 및 하이퍼 파라미터가 적절히 사용되어야 한다.

하이퍼 파라미터의 최적화를 위한 주요 과제는 최적화 시간과 계산 비용을 절약해야 한다는 것이다. 즉, 최적화에 소요되는 시간과 계산 비용을 최소화하면서 상술한 5가지 하이퍼 파라미터의 무수히 많은 조합 중에서 가장 정확한 세트를 선택할 수 있어야 한다. 기존에는 과학적 근거 없이 임시 혹은 무작위로 하이퍼 파라미터를 선택하였으며, 상대적으로 간단한 그리드 검색(Grid search, GS) 또는 무작위 검색(Random search, RS) 기법을 사용한 연구는 많지 않다.

GS 기법과 RS 기법은 데이터 기반 모델의 초매개변수 최적화를 위한 가장 널리 사용되는 방법이다. GS 기법은 사용자가 지정한 범위의 모든 조합을 살펴보기 때문에 모델의 확장이 잘 이루어지지 않는데, 예를 들어 5개의 매개 변수에 대해 9개의 값을 검사해야 한다면 가능한 조합수는 $9 \times 9 \times 9 \times 9 \times 9 = 59,049$ 개가 된다. 그런데 각 하이퍼 파라미터에는 수백 또는 수천 개의 값이 존재할 수 있으므로 GS를 사용하기는 어렵다.

RS 기법은 GS 기법과 동일한 목표를 따르지만 가능한 모든 조합 대신에 확률 분포에서 추출된 무작위로 선택된 조합만 탐색을 수행하며, 일반적으로 RS 기법은 GS 기법에 비해 빠르고 정확한 결과를 제공한다.

하지만, 최적화 프로세스는 하이퍼 파라미터 최적화 외에도 추가적인 불안정한 변수(예: 입력 변수의 개수)도 포함된다. 하이퍼 파라미터와 입력 변수의 개수는 서로 독립적이지 않기 때문에, 어느 하나만 수정하면서 나머지를 최적화하는 것은 바람직하지 않다. 즉, 두 변수 모두 동시에 최적화되어야 한다.

- [0073] 여기서, 베이지안 최적화(BO)는 가능한 가장 적은 반복을 사용하여 하이퍼 파라미터의 최적값을 결정한다. 본 발명의 실시예에서는 낮은 계산 비용으로 높은 효율성을 제공하는 보다 발전된 방법인 베이지안 최적화(BO)에 기초한 이중 최적화 시스템을 사용하여 하이퍼 파라미터 H와 입력 변수의 개수 N를 동시에 최적화하도록 한다.
- [0074] 학습부(130)는 베이지안 최적화가 적용된 이중 루프 구조를 사용하여 하이퍼 파라미터와 입력 변수의 개수의 최적 값을 각각 결정할 수 있다. 이중 루프 구조는, 하이퍼 파라미터를 최적화하기 위한 내부 루프와, 상기 내부 루프의 완료 후 실행되어 입력 변수의 개수를 최적화하고 내부 루프의 동작을 제어하면서 하이퍼 파라미터와 입력 변수의 개수를 동시에 최적화하기 위한 외부 루프를 포함할 수 있다.
- [0075] 이와 같이, H와 N을 최적화하는 이중 최적화 시스템은 첫 번째 루프에 해당하는 내부 루프와 두 번째 루프에 해당하는 외부 루프를 포함하는 두 개의 연속적인 반복 루프로 구성될 수 있다. 하이퍼 파라미터 H의 하위 루프는 입력 변수의 개수 N의 반복(Iteration)에 의해 영향을 받는다.
- [0076] 내부 루프는 하이퍼 파라미터 세트의 최적화와 관련된다. 하이퍼 파라미터 H에 대한 수렴(정지) 기준이 충족되지 않으면, 새로운 H에 대해 반복적으로 내부 루프가 실행된다. 내부 루프가 완료되면, 더 큰 외부 루프가 시작된다. 외부 루프는 내부 루프를 포함하며 H와 N의 최적 값을 동시에 결정하며, 이전에 내부 루프에서 결정된 H의 값에 대해, N에 대한 다른 기준이 만족되는지 여부에 대해서도 유사한 결정이 이루어진다. 기준이 충족되지 않으면, 반복 알고리즘은 새로운 N을 계산하고, 외부 루프의 시작 단계로 돌아가 N에 대한 기준이 만족될 때까지 외부 루프를 반복한다. H는 새로운 N이 선택되어 외부 루프가 실행될 때마다 내부 루프 내에서 새롭게 결정된다. 이중 최적화 프레임워크를 완성하기 위해서는 N과 H에 대한 기준이 모두 충족되어야 한다. 일단 두 개의 루프로부터 최적의 입력 예측 개수 N 및 하이퍼 파라미터 H가 결정되면, LSTM 모델은 훈련 데이터 세트에 의해 훈련되고, 테스트 세트에 의해 평가된다.
- [0077] 즉, 제안한 이중 루프 구조는, 외부 루프의 실행을 통해 결정된 입력 변수의 개수 N이 설정 기준을 충족하지 못하면, 새로운 N을 결정하여 새로운 N이 설정 기준을 만족할 때까지 외부 루프를 반복 실행하고, 아울러 이와 같이 새로운 N이 선택되어 외부 루프가 재 실행될 때마다 내부 루프에서 또한 새로운 H를 결정하여 새로운 H가 설정 기준을 만족할 때까지 내부 루프를 반복 실행하도록 한다. 이러한, 새로운 H와 새로운 N이 모두 해당 기준을 충족할 때까지 루프 실행을 반복하는 것을 통해, 하이퍼 파라미터 H와 상기 입력 변수의 개수 N의 최적 값을 최종 결정하게 된다.
- [0078] 이와 같이, 본 발명은 BO 알고리즘을 도입한 이중 루프 최적화 시스템을 사용하여 LSTM 모델 전체를 최적화할 수 있는 프레임워크를 제시하고 이를 기반으로 장기 예측을 위한 STTM 모델을 훈련할 수 있다.
- [0079] 이하에서는 본 발명에서 제안한 LSTM 모델 구조와 이중 최적화 프레임워크를 기반으로 임의 유역에 대한 미래 유량의 변화를 장기 예측하고 평가한 결과를 설명한다.
- [0080] 도 4는 한강유역에 속한 소양유역의 연구영역과 강우량계의 위치를 나타낸 도면이다. 남한의 소양댐 유역은 독립적으로 위치하여 인간의 활동에 영향을 적게 받기 때문에 사례 연구로 선정되었다. 소양댐은 발전용량 353 GWh의 수력발전댐으로 설계되었다. 댐에 의해 생성된 유역의 면적은 2,703 km²에 달하며, 연평균 유역 유입량과 강우량은 각각 약 55.5 m³/s와 1,100 mm이다.
- [0081] 과거 빅데이터로 1980년 1월 1일부터 2020년 12월 31일까지 41년 동안 27개 강우 관측소의 일별 유량(y)과 일별 강수량(x)을 수집했다. 수집된 데이터(x, y)와 그에 대응된 시차 시간(lag time)을 포함하여 구성된 데이터 세트에서 후보 입력 예측 변수의 선택은 ACF(Autocorrelation function)와 PACF(Partial autocorrelation function)를 사용하여 수행되었다. 유량에 영향을 미치는 강우 관측소 및 시차 시간을 적절히 선택하기 위해 ACF 및 PACF에 대한 상관 계수를 계산하고 사전 정의된 상관 임계값 이상의 변수만 선택하였다.
- [0082] 본 연구에서는 임계값 0.2를 사용하여 유량과 밀접한 상관 관계가 있는 변수를 모델에 대해 선택하였다. 이 임계값에 의해 9개의 강우 관측점 $x = x_1, x_2 \dots x_9$ 및 5개 유량 $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}$ 의 지체 시간(lag time)으로 이루어진 총 14개의 관측 변수를 선택할 수 있다.
- [0083] 훈련 시작 시 Xavier 초기화를 사용하여 하이퍼 파라미터 매개변수를 초기화한 후 손실 함수(NSE, Nash-Sutcliffe 효율)를 Adam 알고리즘으로 최적화하였다. 데이터 세트의 처음 7,671일(1980년부터 2000년까지)은 모델의 훈련에 사용되었고 다음 7,305일(2001부터 2020년까지)은 모델의 테스트에 사용되었다. 효율성 성능을 조사하기 위해 모든 시뮬레이션은 2.20GHz에서 실행되는 Intel Xeon CPU E5-4660 v4의 동일한 컴퓨터 구성을 사용

하여 수행되었다.

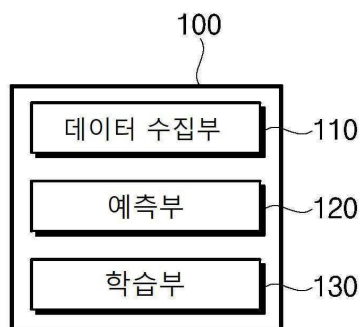
- [0084] 도 5는 본 발명의 실시예에 따른 두 가지 구조의 LSTM 모델을 이용한 장기 예측 결과 성능을 보여주는 도면이다. 도 5는 K = 5 즉, 리드 타임이 매우 짧은 단기 예측의 경우로, spLSTM 모델과 mjLSTM 모델을 단기 예측에 활용하였을 때의 성능을 보여준다.
- [0085] 두 가지 모델 모두 K=5에 대해 만족스런 성능을 보여주었으며, 제안된 두 가지 LSTM 모델을 사용하여 예측된 유량의 전체 크기와 시간적 변화가 실제 관측치와 일치하는 것을 확인할 수 있다. NSE(Nash-Sutcliffe Efficiency)는 0.78~0.81, PE(peak error)는 15.27~23.41, RMSE(Root Mean Square Error)는 77.16~86.67, R=0.89~0.91로 나타냈으며, 예측치와 관찰치 간 1:1 상관(R) 결과에서 mjLSTM 모델 보다 spLSTM 모델이 약간 더 우수한 성능을 나타내었다. NSE와 RMSE는 모델 정확도 평가에 사용되는 일반적인 인자에 해당하며 PE는 예측치 최대값과 관측치 최대값 간 차이를 최대 관측값으로 나눈 비율을 나타낼 수 있다.
- [0086] 이와 같이, spLSTM 모델과 mjLSTM 구조는 단기 예측에서도 만족스러운 성능을 보여주나, 단기 예측 활용에는 다소 복잡도가 있고 시간과 비용 면에서 비효율적인 방안으로 볼 수 있다.
- [0087] 도 6은 본 발명의 실시예에 따른 두 가지 구조의 LSTM 모델을 이용한 장기 예측 결과 성능을 보여주는 도면이다. 도 6은 K=7,305를 적용한 경우 즉, 리드 타임이 긴 장기 예측의 경우로, 리드 타임이 7,305로 증가하더라도, spLSTM 모델과 mjLSTM 모델은 높은 정확도를 보여준다. NSE가 0.70~0.72, PE가 36.95~41.28, RMSE가 112.76~117.54, R이 0.80~0.81로 나타났으며, R이 0.8 이상으로 단기 예측보다 크게 떨어지지 않았다. 결과적으로 리드 타임이 매우 긴 장기 예측에서도 두 모델 모두 허용 가능한 높은 성능을 나타내었다.
- [0088] 이와 같이, 본 발명에서는 장기 예측 문제를 해결하기 위해 이중 루프 프레임워크를 통한 효과적인 구조와 최적화 방법을 사용하였으며, 제안된 LSTM 모델 구조와 프레임워크는 LSTM 모델 구현과 최적화에 적용되어 특정 유역에 대한 하천 흐름을 시뮬레이션하는데 활용될 수 있다. 물론, 제안한 LSTM 구조는 시계열 예측이 가능한 다른 데이터 기반 모델에도 광범위하게 적용될 수 있으며, 다양한 지역의 시계열 변수를 추정할 때 효과적으로 사용될 수 있다.
- [0089] 아울러, 본 발명의 제안 시스템은 강수량을 이용하여 유량을 단기 예측하는 것을 대표 예시로 설명하고 있으나 발명의 적용 대상이 반드시 이에 국한되지 않음은 물론, 다양한 원인과 인과 관계를 가지는 다양한 예측 대상 값의 장기 예측에 활용 가능하다.
- [0090] 본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

- [0091] 100: 기계 학습 시스템 110: 데이터 수집부
120: 예측부 130: 학습부

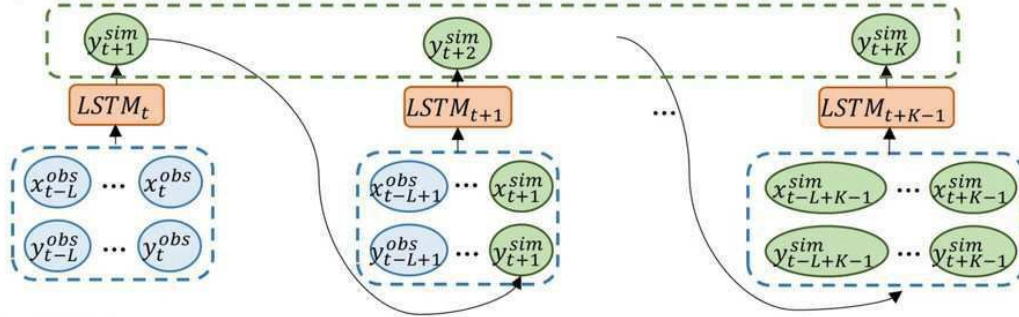
도면

도면1



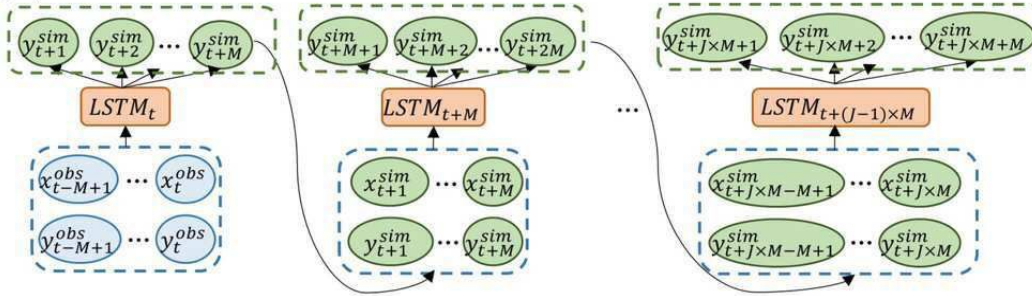
도면2

(a) *spLSTM*

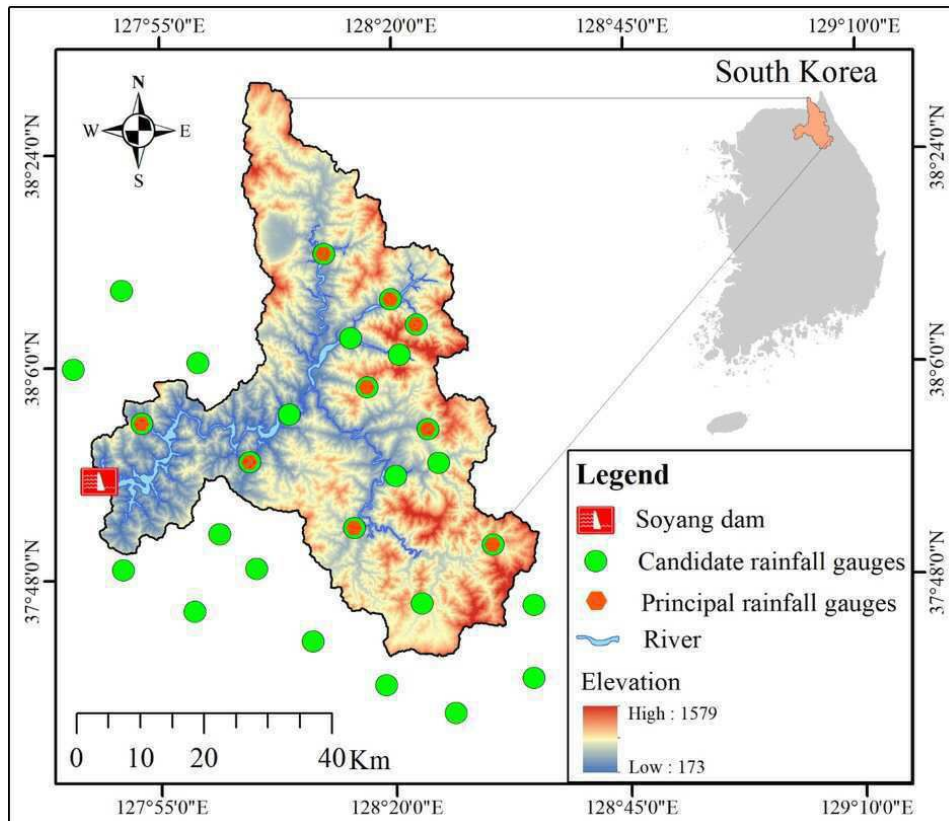


도면3

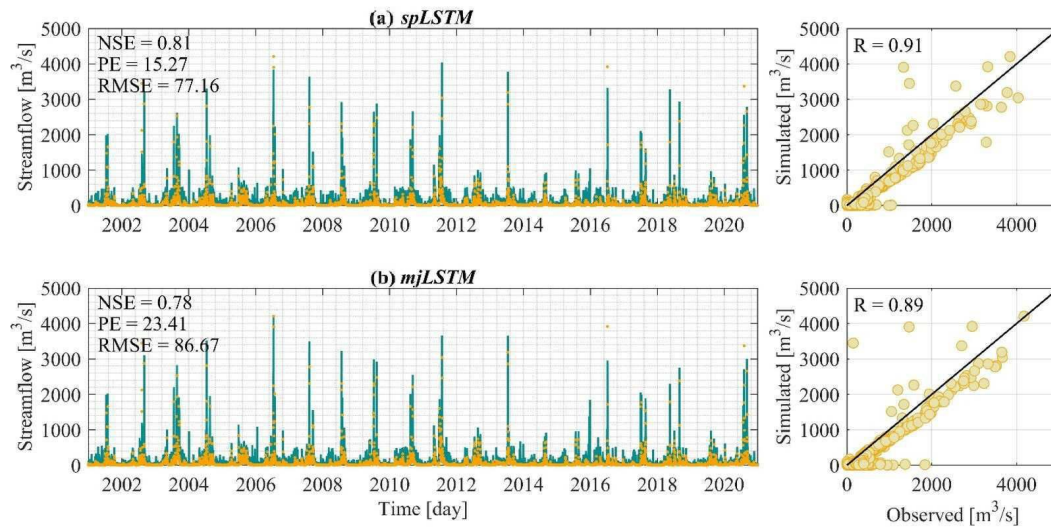
(b) *mjLSTM*



도면4



도면5



도면6

