



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년11월30일
(11) 등록번호 10-2333271
(24) 등록일자 2021년11월26일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/04 (2006.01)

(52) CPC특허분류
G06N 3/082 (2013.01)
G01W 1/10 (2013.01)

(21) 출원번호 10-2020-0070002

(22) 출원일자 2020년06월10일

심사청구일자 2020년06월10일

(56) 선행기술조사문헌

성경민 등., 상호정보량 기법과 인공신경망을 이
용한 실시간 강우 자료 보정, 한국수자원학회:학
술대회논문집 (Proceedings of the Korea Water
Resources Association Conference),
1269-1273pages (2010)

(73) 특허권자

경상국립대학교산학협력단

경상남도 진주시 진주대로 501 (가좌동)

(72) 발명자

이태삼

경남 진주시 소호로 39

(74) 대리인

김종석

전체 청구항 수 : 총 9 항

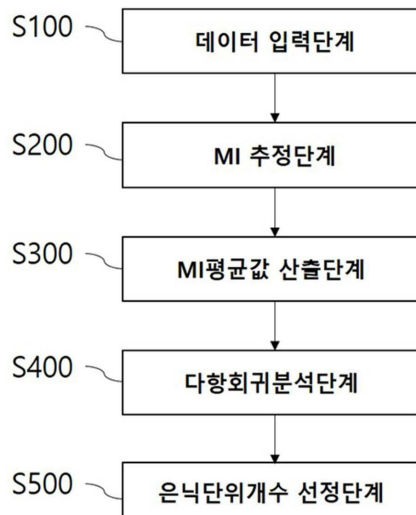
심사관 : 양대경

(54) 발명의 명칭 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법

(57) 요약

본 발명은 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법에 관한 것으로, 보다 구체적으로 인공 신경망 모델 내 은닉단위의 상관성(MI, Mutual Information)을 이용하여 상기 은닉단위의 개수가 선정되고, 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 기후인자 시계열 데이터가 분석되어 수문기후변수가 출력되도록 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법에 관한 것이다.

대표도 - 도1



(52) CPC특허분류

G06N 3/04 (2013.01)

Z01A 10/00 (2019.06)

이 발명을 지원한 국가연구개발사업

과제고유번호	2018R1A2B6001799
과제번호	2018R1A2B6001799
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	수문기상을 바탕으로 한 신재생 자연에너지의 기후변화 대비 지속공급가능성 평가
및 모니터링	
기 여 율	1/1
과제수행기관명	경상대학교
연구기간	2020.03.01 ~ 2021.02.28

명세서

청구범위

청구항 1

인공 신경망 모델에 시계열 데이터가 입력되는 데이터 입력단계;

상기 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정되는 MI 추정단계;

상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출되는 MI평균값 산출단계;

다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출되는 다항회귀분석단계; 및

상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되는 은닉단위개수 선정단계;를 포함하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법.

청구항 2

제 1항에 있어서,

상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은,

상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법.

청구항 3

제 1항에 있어서,

상기 다항회귀분석단계는,

하기 [수학식 1]로 2차 다항식이 산출되는 것을 특징으로 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법.

[수학식 1]

$$y = a_1 + a_2x + a_3x^2$$

여기서, y는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, x는 상기 은닉단위의 개수, a_1 , a_2 , a_3 는 상수이다.

청구항 4

제 1항에 있어서,

상기 은닉단위개수 선정단계는,

상기 은닉단위의 개수(x)가 소수점 첫째자리에서 반올림되어 양의 정수로 도출되는 것을 특징으로 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법.

청구항 5

인공 신경망 모델에 기후인자 시계열 데이터가 입력되는 데이터 입력단계;

상기 기후인자 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정되는 MI 추정단계;

상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출되는 MI평균값 산출단계;

다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출되는 다항회귀분석단계;

상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되는 은닉단위개수 선정단계; 및

상기 은닉단위 개수 선정단계로부터 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 상기 기후인자 시계열 데이터가 분석되어 수문기후변수가 출력되는 수문기후변수 출력단계;를 포함하는 은닉단위 개수 선정 방법을 이용한 수문기후변수 예측방법.

청구항 6

제 5항에 있어서,

상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은,

상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 하는 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법.

청구항 7

제 5항에 있어서,

상기 기후인자 시계열 데이터는 10년 주기 태평양 진동을 나타내는 PDO(Pacific Decadal Oscillation) 시계열 데이터인 것을 특징으로 하는 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법.

청구항 8

제 7항에 있어서,

상기 다항회귀분석단계는,

하기 [수학식 2]로 2차 다항식이 산출되는 것을 특징으로 하는 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법.

[수학식 2]

$$y = 0.006793x^2 - 0.1908x + 5.335$$

여기서, y는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, x는 상기 은닉단위의 개수이다.

청구항 9

제 5항에 있어서,

상기 은닉단위개수 선정단계는,

상기 은닉단위의 개수(x)가 소수점 첫째자리에서 반올림되어 양의 정수로 도출되는 것을 특징으로 하는 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법.

발명의 설명

기술분야

[0001] 본 발명은 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법에 관한 것으로, 보다 구체적으로 인공 신경망 모델 내 은닉단위의 상관성(MI, Mutual Information)을 이용하여 상기 은닉단위의 개수가 선정되고, 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 기후인자 시계열 데이터가 분석되어 수문기후변수가 출력되도록 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법에 관한 것이다.

배경기술

[0002] 과거에는 주로 전문가나 문헌에 의해 지식을 얻었지만, 문제의 범위가 복잡하고 넓어짐에 따라 과거의 전통적인 방법으로 지식을 획득하기 어려워졌다. 이에 대한 대안 및 4차 산업혁명의 진행으로 인해 데이터를 분석하여 패턴과 규칙을 찾아내어 지식을 추출하려는 방법들이 제시되고 있다.

[0003] 즉, 데이터로부터 지식을 추출하는 기법으로 전통적으로는 통계기법을 넘어 최근에는 인공 신경망, 의사결정나무, 유전자알고리즘, 사례추론시스템, 퍼지시스템 등의 인공지능(Artificial Intelligent, AI) 기법 등이 사용되고 있다.

[0004] 특히, 인공 신경망은 분류 및 예측 문제의 해결하기 위한 다방면의 문제영역에서 사용되고 있다. 예컨대, 인공 신경망은 부실기업 예측모형, 채권등급 평가 등 재무관련 자료들을 분석하고 활용하여 결과를 예측할 수 있다. 그리고 과거부터 현재까지의 기후 변화와 이에 따른 영향을 평가 및 분석하고, 이를 활용하여 수자원 관리, 수문 설계 및 많은 기후관련 응용분야에 사용될 수 있다.

[0005] 다만, 인공 신경망은 데이터의 잡음에 민감하지 않고 그 구조가 견고하나, 자료를 학습하는 내부과정이 복잡한 수학적 모델에 의해서 생성되기 때문에 사용자들이 결과를 이해하기 어려우며, 인공 신경망의 복잡한 구조에 의해 입력된 데이터를 분석하는데 긴 시간이 걸리는 것이 가장 큰 문제점으로 지적되고 있다.

[0006] 관련문헌 1은 딥러닝 기반 기후 변화 예측 시스템의 동작 방법에 관한 것으로, 누락된 기상 데이터가 있는지를 감지하여 정확한 기후 변화를 예측할 수 있는 장점이 있으나, 은닉층의 수를 고려하지 않고 많은 데이터를 분해하고 분해된 데이터들을 분석하므로 신속한 분석이 어려운 단점이 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) KR 10-2020-0052806

발명의 내용

해결하려는 과제

[0008] 본 발명은 상기와 같은 문제점을 해결하기 위한 것으로 종래 은닉단위의 개수를 산정하기 위하여 자료를 분할하고 다른 자료로부터 은닉단위의 개수를 산정하는 과정을 생략할 수 있도록 은닉단위의 개수와 은닉단위의 상관성(MI, Mutual Information)을 이용하여 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되도록 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법을 제공하는 것을 목적으로 한다.

[0009] 또한, 본 발명의 목적은 인공 신경망의 복잡한 구조에 의해 속도가 느려지는 단점을 해소하고, 상기 인공 신경망 모델에 입력된 입력 데이터의 연산속도가 증가되어 분석결과를 보다 빠르게 도출할 수 있도록 인공 신경망 모델이 가장 적절한 은닉단위의 개수를 갖도록 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법을 제공하는 것이다.

[0010] 또한, 본 발명의 목적은 수문학적으로 유의미한 예측자료를 신속하게 출력할 수 있도록 은닉단위개수 선정단계로부터 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 기후인자 (예컨대, PDO) 시계열 데이터가 분석되어 수문기후변수가 출력되도록 하는 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선

정방법 및 이를 이용한 수문기후변수 예측방법을 제공하는 것을 목적으로 한다.

과제의 해결 수단

- [0011] 상기 목적을 달성하기 위하여, 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정 방법은 인공 신경망 모델에 시계열 데이터가 입력되는 데이터 입력단계; 상기 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정되는 MI 추정단계; 상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출되는 MI평균값 산출단계; 다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출되는 다항회귀분석단계; 및 상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되는 은닉단위개수 선정단계;를 제공한다.
- [0012] 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법에 있어서, 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은, 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 한다.
- [0013] 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법에 있어서, 상기 다항회귀분석단계는, 하기 [수학식 1]로 2차 다항식이 산출되는 것을 특징으로 한다.
- [0014] [수학식 1]
- $$y = a_1 + a_2x + a_3x^2$$
- [0015]
- [0016] 여기서, y는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, x는 상기 은닉단위의 개수, a_1 , a_2 , a_3 는 상수이다.
- [0017] 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법에 있어서, 상기 은닉단위 개수 선정단계는, 상기 은닉단위의 개수(x)가 소수점 첫째자리에서 반올림되어 양의 정수로 도출되는 것을 특징으로 한다.
- [0018] 다음으로, 상기 목적을 달성하기 위하여, 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법은 인공 신경망 모델에 기후인자 시계열 데이터가 입력되는 데이터 입력단계; 상기 기후인자 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정되는 MI 추정단계; 상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출되는 MI평균값 산출단계; 다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출되는 다항회귀분석단계; 상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되는 은닉단위개수 선정단계; 및 상기 은닉단위 개수 선정단계로부터 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 상기 기후인자 시계열 데이터가 분석되어 수문기후변수가 출력되는 수문기후변수 출력단계;를 제공한다.
- [0019] 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법에 있어서, 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은, 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 한다.
- [0020] 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법에 있어서, 상기 기후인자 시계열 데이터는 10년 주기 태평양 진동을 나타내는 PDO(Pacific Decadal Oscillation) 시계열 데이터인 것을 특징으로 한다.
- [0021] 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법에 있어서, 상기 다항회귀분석단계는, 하기 [수학식 2]로 2차 다항식이 산출되는 것을 특징으로 하는 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법.

[0022] [수학식 2]

[0023]
$$y = 0.006793x^2 - 0.1908x + 5.335$$

[0024] 여기서, y는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, x는 상기 은닉단위의 개수이다.

[0025] 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법에 있어서, 상기 은닉단위개수 선정단계는, 상기 은닉단위의 개수(x)가 소수점 첫째자리에서 반올림되어 양의 정수로 도출되는 것을 특징으로 한다.

발명의 효과

[0026] 이상과 같이 본 발명에 의하면 은닉단위의 개수와 은닉단위의 상관성(MI, Mutual Information)을 이용하여 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정되도록 함으로써, 종래 은닉단위의 개수를 산정하기 위하여 자료를 분할하고 다른 자료로부터 은닉단위의 개수를 산정하는 과정을 생략 가능한 효과가 있다.

[0027] 또한, 본 발명은 인공 신경망 모델이 가장 적절한 은닉단위의 개수를 갖도록 함으로써, 인공 신경망의 복잡한 구조에 의해 속도가 느려지는 단점을 해소하고, 상기 인공 신경망 모델에 입력된 입력 데이터의 연산속도가 증가되어 분석결과를 더욱 빠르게 도출할 수 있는 효과가 있다.

[0028] 또한, 본 발명은 은닉단위개수 선정단계로부터 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 기 후인자 (예컨대, PDO) 시계열 데이터가 분석되어 수문기후변수가 출력되도록 함으로써, 수문학적으로 유의미한 예측자료를 신속하게 출력할 수 있는 효과가 있다.

도면의 간단한 설명

[0029] 도 1은 본 발명에 따른 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 흐름도이다.

도 2는 본 발명의 실시예에 따른 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법 흐름도이다.

도 3은 본 발명의 실시예에 따른 PDO 시계열 데이터를 표시한 도면이다.

도 4는 본 발명의 실시예에 따른 LSTM4의 각 은닉단위의 상태를 표시한 도면이다.

도 5는 본 발명의 실시예에 따른 LSTM10의 각 은닉단위의 상태를 표시한 도면이다.

도 6은 본 발명의 실시예에 따른 LSTM6과 LSTM14의 각 은닉단위의 상관성을 표시한 도면이다.

도 7은 본 발명의 실시예에 따른 다항회귀분석단계로부터 산출된 [수학식 2]를 표시한 그래프이다.

도 8은 본 발명의 실시예에 따른 은닉단위의 개수에 따른 RMSE와 은닉단위의 상관성을 표시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 명세서에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 발명의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 발명에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 발명의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0031] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

[0032] 이하, 본 발명에 따른 실시예를 첨부한 도면을 참조하여 상세히 설명하기로 한다. 도 1은 본 발명에 따른 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법 흐름도이다.

[0033] 도 1을 보면, 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법은 데이터 입력단계(S100), MI 추정단계(S200), MI평균값 산출단계(S300), 다항회귀분석단계(S400), 및 은닉단위개수 선정단계(S500)를 포함한다.

- [0034] 보다 구체적으로 설명해보면, 상기 데이터 입력단계(S100)는 인공 신경망 모델에 시계열 데이터가 입력된다.
- [0035] 일반적으로 인공 신경망은 입력층, 은닉층, 출력층을 포함하고, 상기 입력층에 입력 데이터가 입력되면 상기 은닉층이 상기 입력 데이터를 분석하고 출력층에 출력데이터를 제공한다. 상기 은닉층은 다수 개의 은닉단위 또는 은닉노드가 포함될 수 있다. 본 명세서에는 은닉층 내 다수 개의 은닉단위가 포함되는 것을 전제로 작성된 것이다.
- [0036] 또한, 인공 신경망은 다양한 종류가 있으며, 그 중 시간이력을 고려할 수 있는 모델로 순환 신경망(RNN)이 있고, 상기 순환 신경망(RNN)은 일반적인 순환신경망과 LSTM(Long and Short Term Memory)기반의 순환 신경망이 있다.
- [0037] 가장 바람직하게, 본 발명의 인공 신경망 모델 내 은닉단위의 상관성을 이용한 은닉단위 개수 선정방법은 장기간 시계열 데이터를 분석하기 위하여 일반 인공신경망, RNN 및 LSTM(Long and Short Term Memory) 기반의 순환 신경망이 사용될 수 있고, 이에 따라 상기 입력 데이터, 은닉단위 및 출력 데이터 모두 시계열 데이터 일 수 있다.
- [0038] 그리고 상기 인공 신경망 모델은 상기 은닉단위의 개수에 따라 LSTM1, LSTM4 또는 LSTM10으로 명명할 수 있다. 예컨대, LSTM1은 은닉층에 은닉단위 1개를 갖는 인공 신경망이고, LSTM4는 은닉층에 은닉단위 4개를 갖는 인공 신경망인 것이다.
- [0039] 다음으로, 상기 MI 추정단계(S200)는 상기 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정된다.
- [0040] 한편, 상기 MI 추정단계(S200)는 이변량 경험누적 분포함수를 이용하여 은닉단위의 상관성(MI, Mutual Information)이 추정된다. 여기서, 이변량 경험누적 분포함수는 경험적 척도와 관련된 누적분포함수으로써, 확률 표본에 대하여 통계량을 구하고 작은 값부터 순서대로 배치한 계단형 그래프로 도출될 수 있다.
- [0041] LSTM4를 이용하여 예를 들어보면, LSTM4의 은닉층에는 4개의 은닉단위가 존재한다. 각각의 은닉단위는 시계열 형태로 나타낼 수 있고, 은닉단위들 하나하나 서로 간의 은닉단위의 상관성(MI, Mutual Information)을 구할 수 있다.
- [0042] 이때, 상기 MI 추정단계(S200)는 4개의 은닉단위에 대해 서로 간의 은닉단위의 상관성(MI, Mutual Information)이 이변량 경험누적 분포함수를 이용하여 추정될 수 있고, 제 1 MI부터 제 n MI로 나타낼 수 있다.
- [0043] 다음으로, MI평균값 산출단계(S300)는 상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된다.
- [0044] LSTM4를 이용하여 예를 들어보면, 상기 MI 추정단계(S200)로부터 추정된 제 1 MI부터 제 n MI를 평균하면 LSTM4 전체에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 최종적으로 산출될 수 있다.
- [0045] 즉, 상기 MI평균값 산출단계(S300)는 앞서 LSTM4에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된 방식으로, LSTM1부터 LSTMn까지의 인공 신경망 모델에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된다.
- [0046] 여기서, 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은, 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 한다.
- [0047] 다음으로, 다항회귀분석단계(S400)는 다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출된다.
- [0048] 일반적으로, 다항회귀분석(Polynomial Regression)은 비선형적인 관계를 갖는 독립변수와 종속변수를 분석하기 위한 방법으로, 하기 [수학식 1]로 2차 다항식이 도출될 수 있다.

수학식 1

$$y = a_1 + a_2x + a_3x^2$$

[0049]

[0050] 여기서, y 는 종속변수이고, x 는 독립변수이다. 그리고 해당하는 데이터를 대입하여, a_1 , a_2 , a_3 상수를 구하면, 해당하는 데이터의 비선형적인 관계를 2차 다항식 그래프로 나타낼 수 있는 것이다.

[0051] 즉, 본 발명의 y 는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, x 는 상기 은닉단위의 개수, a_1 , a_2 , a_3 는 상수이다. 그리고 상기 다항회귀분석단계(S400)는 다수 개의 데이터 중 임의의 (은닉단위의 개수 (x_1), 은닉단위의 상관성의 평균값(y_1))과 (은닉단위의 개수 (x_2), 은닉단위의 상관성의 평균값(y_2))를 대입 하여 상기 a_1 , a_2 , a_3 을 구할 수 있다.

[0052] 다음으로, 상기 은닉단위개수 선정단계(S500)는 상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정된다.

[0053] 즉, 상기 은닉단위개수 선정단계(S500)는 상기 다항회귀분석단계(S400)로부터 상기 [수학식 1]이 도출되면 상기 종속변수인 은닉단위의 개수(x)에 대해 상기 [수학식 1]이 하기 [수학식 3]과 같이 1차 미분될 수 있다.

[0054] 1차 미분을 하는 이유는 상기 언급한 것과 같이 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 특징이 있으므로, 상기 최소점을 구하기 위해서이다.

[0055] 그리고 하기 [수학식 3]과 같이 상기 독립변수인 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0을 대입하여, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정될 수 있는 것이다.

수학식 3

$$y' = a_2 + 2a_3x$$

$$y' = 0 \text{ 이면, } x = -\frac{a_2}{2a_3}$$

[0056]

[0057] 이때, 상기 은닉단위개수 선정단계(S500)는 상기 은닉단위의 개수(x)가 소수점 첫째자리에서 반올림되어 양의 정수로 도출되는 것을 특징으로 한다. 이는 인공 신경망 모델 내 상기 은닉단위의 개수(x)는 반드시 양의 정수여야 하기 때문이다.

[0059] **은닉단위 개수 선정방법 및 이를 이용한 수문기후변수 예측방법**

[0060] 도 2는 본 발명의 실시예에 따른 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법 흐름도이다.

[0061] 도 2를 보면, 본 발명의 은닉단위 개수 선정방법을 이용한 수문기후변수 예측방법은 데이터 입력단계(S110), MI 추정단계(S210), MI평균값 산출단계(S310), 다항회귀분석단계(S410), 은닉단위개수 선정단계(S510) 및 수문기후변수 출력단계(S610)를 포함한다.

[0062] 보다 구체적으로 설명해보면, 상기 데이터 입력단계(S110)는 인공 신경망 모델에 기후인자 시계열 데이터가 입력된다.

[0063] 가장 바람직하게, 상기 기후인자 시계열 데이터는 10년 주기 태평양 진동을 나타내는 PDO(Pacific Decadal Oscillation) 시계열 데이터인 것을 특징으로 한다. 상기 PDO 시계열 데이터는 가장 대표적인 기후지수인 태평

양 기후 변동 패턴을 시계열로 나타낸 데이터이다.

- [0064] 또한, 가장 바람직하게, 상기 인공 신경망 모델은 장기간 시계열 데이터를 분석하기 위하여 LSTM(Long and Short Term Memory) 기반의 순환신경망이 사용될 수 있다. 이에 따라, 입력 데이터인 상기 PDO 시계열 데이터, 은닉단위 및 출력데이터인 수문기후변수가 모두 시계열 데이터일 수 있다.
- [0065] 그리고 상기 인공 신경망 모델은 상기 은닉단위의 개수에 따라 LSTM1, LSTM4 또는 LSTM10으로 명명할 수 있다. 예컨대, LSTM1은 은닉층에 은닉단위 1개를 갖는 인공 신경망 모델이고, LSTM4는 은닉층에 은닉단위 4개를 갖는 인공 신경망 모델인 것이다.
- [0066] 다음으로, 상기 MI 추정단계(S210)는 상기 기후인자 시계열 데이터에 있어서, 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)이 추정된다.
- [0067] 한편, 상기 MI 추정단계(S210)는 이변량 경험누적 분포함수를 이용하여 은닉단위의 상관성(MI, Mutual Information)이 추정된다. 여기서, 이변량 경험누적 분포함수는 경험적 척도와 관련된 누적분포함수으로써, 확률 표본에 대하여 통계량을 구하고 작은 값부터 순서대로 배치한 계단형 그래프로 도출될 수 있다.
- [0068] LSTM4를 이용하여 예를 들어보면, LSTM4의 은닉층에는 4개의 은닉단위가 존재한다. 각각의 은닉단위는 시계열 형태로 나타낼 수 있고, 은닉단위들 하나하나 서로 간의 은닉단위의 상관성(MI, Mutual Information)을 구할 수 있다.
- [0069] 이때, 상기 MI 추정단계(S210)는 4개의 은닉단위에 대해 서로 간의 은닉단위의 상관성(MI, Mutual Information)이 이변량 경험누적 분포함수를 이용하여 추정될 수 있고, 제 1 MI부터 제 n MI로 나타낼 수 있다.
- [0070] 다음으로, MI평균값 산출단계(S310)는 상기 은닉단위의 개수별 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된다.
- [0071] LSTM4를 이용하여 예를 들어보면, 상기 MI 추정단계(S210)로부터 추정된 제 1 MI부터 제 n MI를 평균하면 상기 은닉단위의 개수 4개를 갖는 LSTM4에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 최종적으로 산출될 수 있다.
- [0072] 즉, 상기 MI평균값 산출단계(S310)는 앞서 LSTM4에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된 방식으로, LSTM1부터 LSTMn까지의 인공 신경망 모델에 대한 은닉단위의 상관성(MI, Mutual Information)의 평균값이 산출된다.
- [0073] 이때, 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은, 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가되는 것을 특징으로 한다.
- [0075] 이를 증명하고자 간단한 실험이 진행된다.
- [0076] 도 3은 본 발명의 실시예에 따른 PDO 시계열 데이터를 표시한 도면이다. 도 4는 본 발명의 실시예에 따른 LSTM4의 각 은닉단위의 상태를 표시한 도면이다. 도 5는 본 발명의 실시예에 따른 LSTM10의 각 은닉단위의 상태를 표시한 도면이다.
- [0077] 우선, 도 3을 보면, 본 실험에서는 PDO 시계열 데이터를 입력 데이터로 한다. 상기 PDO 시계열 데이터는 가장 대표적인 기후지수인 태평양 기후 변동 패턴을 시계열로 나타낸 데이터이다.
- [0078] 도 4를 보면, 은닉단위의 개수로 4개를 갖는 LSTM4 인공 신경망 모델을 이용하여 상기 PDO 시계열 데이터가 분석된 것으로, 상기 각 은닉단위의 상태(검은색 그래프)가 상기 PDO 시계열 데이터 그래프에 겹쳐져 표시되어 있다.
- [0079] 보다 구체적으로, 도 3의 ht-1 그래프를 보면 제 1 은닉단위는 고주파수 변동성을 갖고 있다. 상기 제 1 은닉단위는 고주파를 나타내면서 정상(normal)에 가깝다. 대조적으로 제 2 내지 4 은닉단위는 양 또는 음의 단 하나의 위상에만 민감하기 때문에 왜곡이 심한 것을 볼 수 있다.
- [0080] 1920-1940년 및 1980-2000년과 같은 주파수가 높은 기간 동안 도 3의 ht-2 그래프를 보면 제 2 은닉단위의 상태는 높은 양의 값을 얻는 반면, 도 3의 ht-4 그래프를 보면 제 4 은닉단위의 상태는 높은 음의 값을 얻는다.
- [0081] 또한, 제 2 은닉단위와 제 3 은닉단위는 거의 동일한 모습을 보이며 서로 매우 의존적이고, 제 1 은닉단위와 제

4 은닉단위는 거의 의존하지 않는다.

- [0082] 다음으로, 비교적 많은 수의 은닉단위를 갖는 구조의 특성을 파악하기 위해 LSTM10 모델이 추가로 분석된다.
- [0083] 도 5를 보면, 은닉단위의 개수로 10개를 갖는 LSTM10 인공 신경망 모델을 이용하여 상기 PDO 시계열 데이터가 분석된 것으로, 상기 각 은닉단위의 상태(검은색 그래프)가 상기 PDO 시계열 데이터에 겹쳐져 표시되어 있다.
- [0084] 1900-1930년인 제 1 기간의 변동성은 제 9 내지 10 은닉단위로 캡처되는 반면, 이외 은닉단위는 제 1 기간 동안 -1 또는 1의 값을 일관되게 유지한다. 즉, 상기 10개의 은닉단위 중 중요한 스펙트럼은 제 1 기간에서 제 9 내지 10 은닉단위이다.
- [0085] 그리고 제 3 은닉단위는 제 4 은닉단위와 음의 관계가 있고, 제 7 은닉단위와 양의 관계가 있는 것으로 보인다. 제 1 은닉단위는 도 3의 ht-1에 표시된 것처럼 상기 PDO 시계열 데이터의 피크값에 매우 민감하지만, 제 8 단위는 음의 피크에 역으로 민감하다. 제 6 은닉단위는 제 1 기간을 제외하고는 장기적으로 상기 PDO시계열 데이터와 음의 관계를 나타낸다.
- [0086] 다만, 이외의 단위들은 서로 의존하지 않으므로, 상기 LSTM10 모델 내 은닉단위는 상기 LSTM4 모델 내 은닉단위보다 낮은 상관성을 갖는 것을 볼 수 있다.
- [0087] 추가로, 도 6은 본 발명의 실시예에 따른 LSTM6과 LSTM14의 각 은닉단위의 상관성을 표시한 도면이다.
- [0088] 도 6은 플롯 매트릭스이고, LSTM의 은닉단위를 대각선 패널에 위치시키고 상관성을 표시한 그래프이다. 즉, 하나의 은닉단위와 다른 하나의 은닉단위에서 겹쳐지는 패널에 흠어짐이 적은 선명한 그래프가 도출될 경우 두 은닉단위의 양/음의 상관성이 높다고 할 수 있다.
- [0089] 도 6의 (a)는 6개의 은닉단위를 갖는 LSTM6의 플롯 매트릭스이고, 모든 패널에서 흠어짐이 적고 양 또는 음의 상관성을 명확히 보여주는 선명한 그래프가 도출된 것을 볼 수 있다. 도 5의 (b)는 14개의 은닉단위를 갖는 LSTM14의 플롯 매트릭스이고, 모든 패널에서 흠어짐이 많은 불명확한 그래프가 도출된 것을 볼 수 있다.
- [0090] 즉, 적은 수의 은닉단위에서 각 은닉단위는 입력 데이터로 사용된 상기 PDO 시계열 데이터가 충분히 분리되지 않았기 때문에 은닉단위 간 강한 상관성을 갖는다. 반면에, 많은 수의 은닉단위에서 각 은닉단위는 입력 데이터로 사용된 상기 PDO 시계열 데이터가 충분히 분리되었기 때문에 은닉단위 간 약한 상관성을 갖는다.
- [0091] 다만, 은닉단위의 상관성(MI, Mutual Information)의 평균값이 최소인 은닉단위의 개수를 초과하게 되면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 유지되거나 다시 증가될 수 있다. 이는 하기에서 보다 상세히 설명한다.
- [0093] 다음으로, 상기 다항회귀분석단계(S410)는 다항회귀분석(Polynomial Regression)을 이용하여 독립변수가 상기 은닉단위의 개수이고, 종속변수가 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값인 2차 다항식이 도출된다.
- [0094] 가장 바람직하게, 상기 다항회귀분석단계(S410)는, 상기 기후인자 시계열 데이터가 10년 주기 태평양 진동을 나타내는 PDO(Pacific Decadal Oscillation) 시계열 데이터일 경우 하기 [수학식 2]로 2차 다항식이 산출되는 것을 특징으로 한다.

수학식 2

$$y = 0.006793x^2 - 0.1908x + 5.335$$

- [0095]
- [0096] 여기서, 종속변수인 y는 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이고, 독립변수인 x는 상기 은닉단위의 개수이다.
- [0097] 도 7은 본 발명의 실시예에 따른 다항회귀분석단계로부터 산출된 [수학식 2]를 표시한 그래프이다.
- [0098] 도 7을 보면, 상기 MI평균값 산출단계(S310)로부터 산출된 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 각 은닉단위의 개수에 따라 좌표가 표시되어 있다. 그리고 다항회귀분석단계(S410)로부터 산출된 [수학식 2]를 표시한 그래프가 보라색으로 표시되어 있다.
- [0099] 도 7을 보면, 은닉단위의 상관성(MI, Mutual Information)의 평균값이 최소인 은닉단위의 개수를 초과하게 되면

상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 다시 증가되는 것을 볼 수 있다.

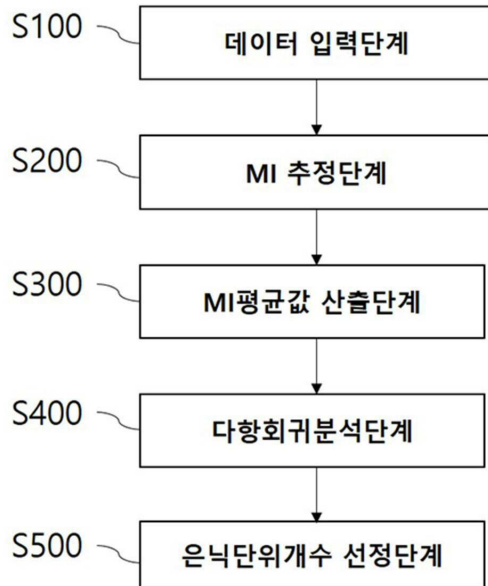
- [0100] 즉, 일정 이상으로 은닉단위의 개수가 증가하면 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값이 다시 증가하게 되고, 이로 인해 분석부하가 증가하여 속도가 느려질 수 있는 단점이 있다. 그리고 은닉단위의 수에 따라 매개변수가 기하급수적으로 증가하기 때문에 적절한 은닉단위의 수를 선정해주는 것이 매우 중요하다.
- [0101] 다음으로, 상기 은닉단위개수 선정단계(S510)는 상기 2차 다항식이 은닉단위의 개수에 대해 1차 미분되고, 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되어, 상기 인공 신경망 모델의 가장 적절한 은닉단위의 개수가 선정된다.
- [0102] 또한, 상기 은닉단위개수 선정단계(S510)는 상기 다항회귀분석단계(S410)로부터 하기 [수학식 2]로 2차 다항식이 산출된다면 상기 [수학식 2]가 은닉단위의 개수(x)에 대해 1차 미분된다.
- [0103] 1차 미분을 하는 이유는 상기 언급한 것과 같이 상기 은닉단위의 개수가 증가할수록 각각의 은닉단위의 고유한 특성이 강화되어 상기 은닉단위의 상관성(MI, Mutual Information)이 감소되고, 최소점을 지나면 상기 은닉단위의 상관성(MI, Mutual Information)이 유지되거나 다시 증가되는 특징이 있으므로, 상기 최소점을 구하기 위해서이다.
- [0104] 그리고 상기 은닉단위개수 선정단계(S510)는 상기 은닉단위 간 상관성이 제거되도록 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값에 0이 대입되면 상기 인공 신경망 모델의 은닉단위의 개수(x)로 14.04가 산출된다.
- [0105] 이때, 상기 은닉단위의 개수(x)는 양의 정수로 도출되어야함으로, 소수점 첫째자리에서 반올림되어 14개인 것이다.
- [0106] 즉, 상기 은닉단위개수 선정단계(S510)는 상기 다항회귀분석단계(S410)로부터 도출된 상기 [수학식 2]를 미분함으로써, 상기 은닉단위의 상관성(MI, Mutual Information)이 최소가 되는 은닉단위의 개수를 용이하게 찾을 수 있다.
- [0107] 다음으로, 상기 수문기후변수 출력단계(S610)는 상기 은닉단위 개수 선정단계(S500)로부터 선정된 은닉단위의 개수를 갖는 인공 신경망 모델을 이용하여 상기 PDO 시계열 데이터가 분석되어 수문기후변수가 출력된다.
- [0108] 즉, 상기 PDO 시계열 데이터는 과거로부터 추출된 값으로, 상기 인공 신경망 모델에 입력되면 상기 수문기후변수 출력단계(S610)는 미래에 발생할 수 있는 유의미한 수문학적 데이터가 출력되는 것이다.
- [0110] **종래 비교**
- [0111] 도 8은 본 발명의 실시예에 따른 은닉단위의 개수에 따른 RMSE와 은닉단위의 상관성을 표시한 도면이다.
- [0112] 본 발명의 정확성 및 신뢰성을 증명하기 위해서 일반적으로 사용되는 은닉단위의 개수에 따른 RMSE 값과 본 발명의 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)의 평균값을 비교해본다.
- [0113] 도 8을 보면, 파란색 그래프가 일반적으로 사용되는 은닉단위의 개수에 따른 RMSE 값이고, 빨간색 그래프가 본 발명의 은닉단위의 개수에 따른 은닉단위의 상관성(MI, Mutual Information)의 평균값이다.
- [0114] 상기 평균 제곱근 오차(RMSE, Root Mean Square Error)는 k-폴드 교차검증을 통해 도출될 수 있다. 여기서, k-폴드 교차검증은 k는 양의 정수이고, 전체 데이터를 k 수 만큼 비슷한 크기의 집합으로 나눈다.
- [0115] 즉, k=5일 때, 전체 데이터를 5개의 부분집합으로 분할한 후, 각 분할마다 하나의 폴드를 테스트용으로 사용하고 나머지 4개의 폴드는 훈련용으로 사용하고 이 과정을 반복하여 각 분할마다 정확도를 측정하는 것이다.
- [0116] 가장 바람직하게, 도 8은 K=10일 때 은닉단위의 개수에 따른 평균 제곱근 오차(RMSE, Root Mean Square Error)를 도출한 것이다.
- [0117] 즉, 도 8을 보면, LSTM5의 RMSE 값과 유사할 때까지 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은 상당한 감소를 보여준다. 이와 같이 두 그래프가 완벽하게 일치하지 않더라도 은닉단위의 개수가 증가함에 따라 RMSE와 상기 은닉단위의 상관성(MI, Mutual Information)의 평균값은 유사하게 동작되는 것을 알 수 있다. 이에 따라, 상기 은닉단위의 상관성(MI, Mutual Information)은 종래 사용되는 RMSE 값을 대체할 수 있다.
- [0119] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는바, 당업계의 통상의 지식을 가진 자에게 있어서 이러한

구체적인 기술은 단지 바람직한 구현예일 뿐이며, 이에 본 발명의 범위가 제한되는 것이 아닌 점은 명백하다.

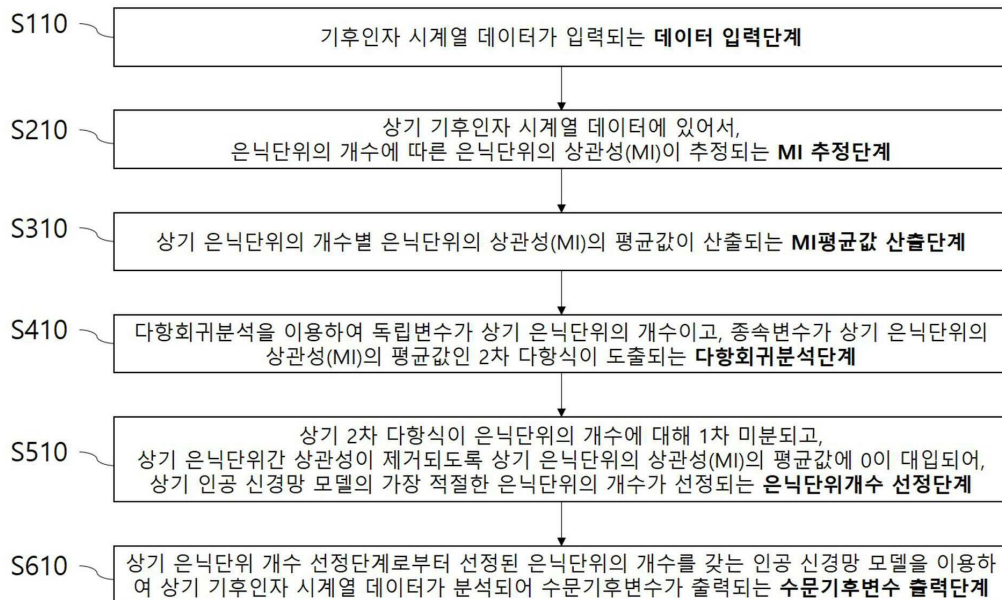
따라서 본 발명의 실질적인 범위는 첨부된 청구항과 그의 등가물에 의하여 정의된다고 할 것이다.

도면

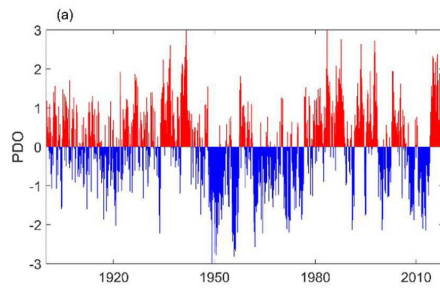
도면1



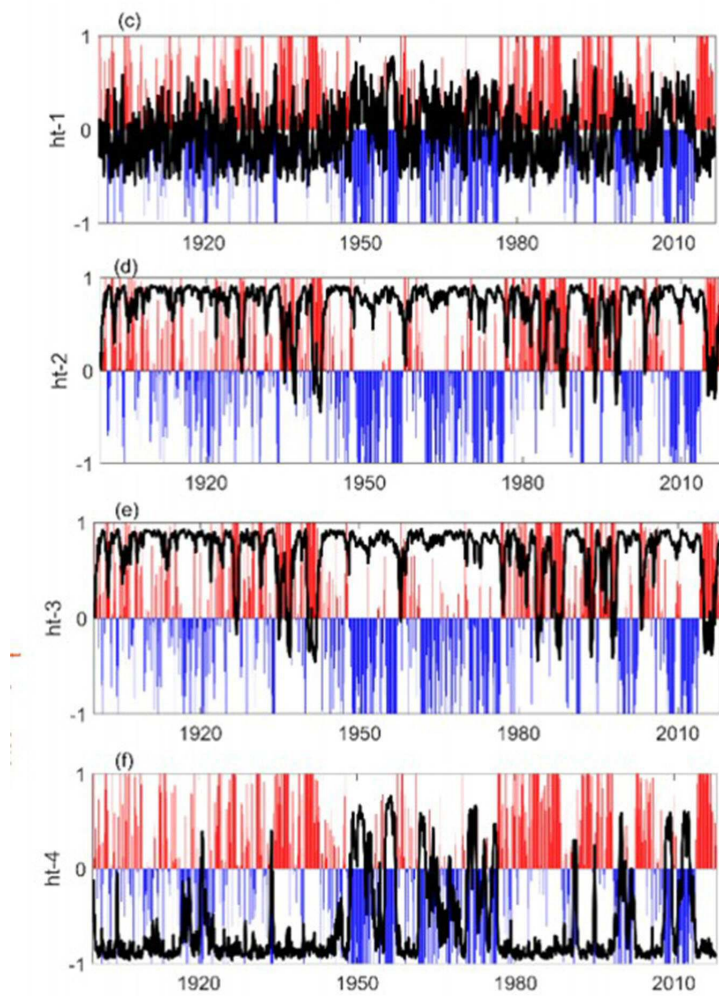
도면2



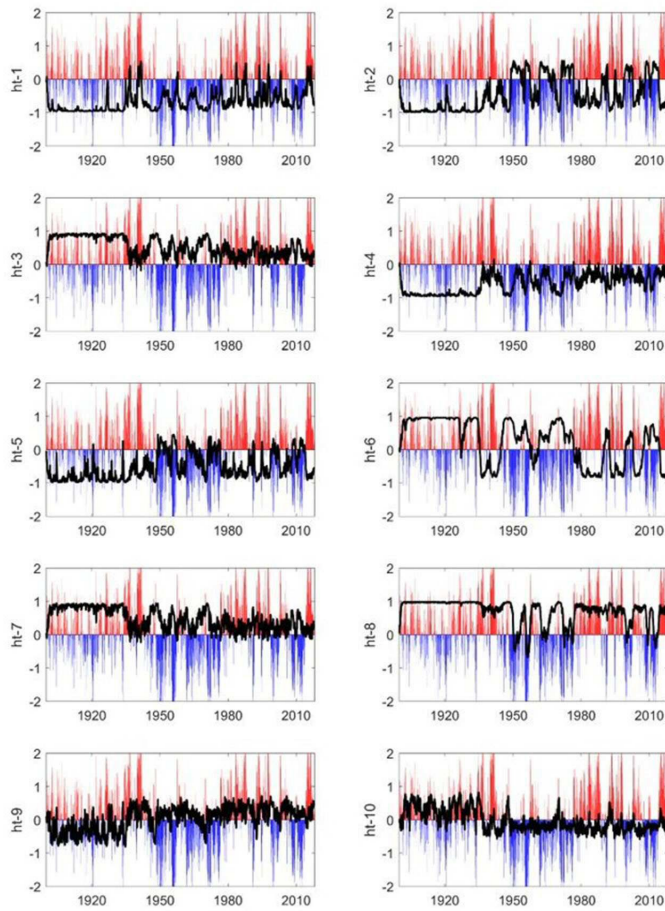
도면3



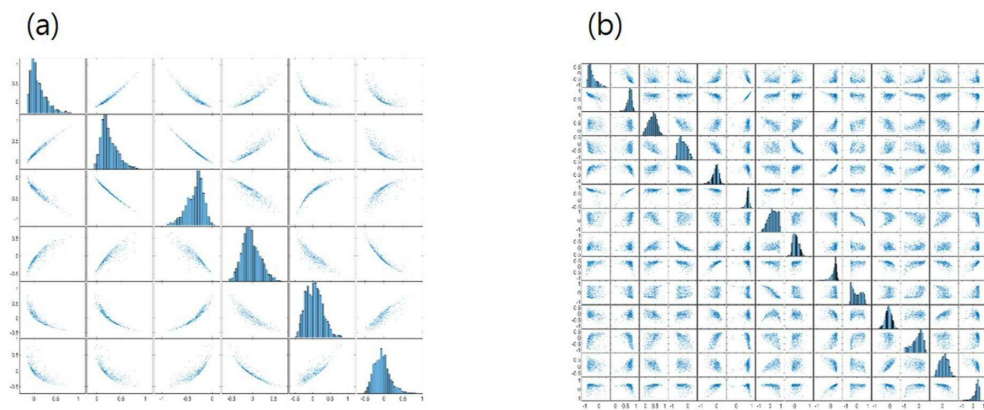
도면4



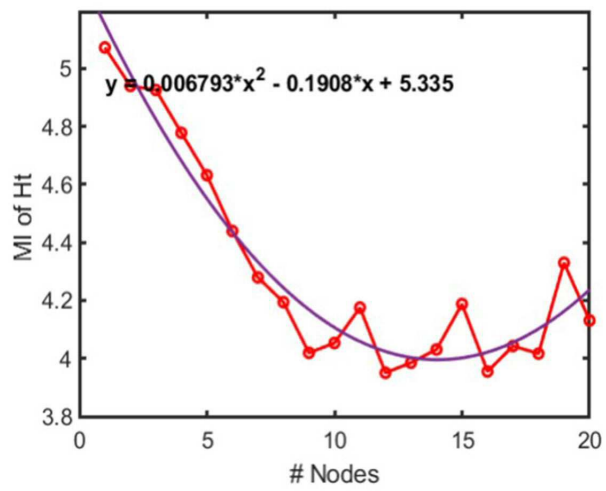
도면5



도면6



도면7



도면8

