



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0002682
(43) 공개일자 2024년01월05일

(51) 국제특허분류(Int. Cl.)
G16H 50/80 (2018.01) G06N 20/00 (2019.01)
G06N 3/04 (2023.01) G06N 3/08 (2023.01)
G16H 10/20 (2018.01) G16H 10/60 (2018.01)
G16H 50/50 (2018.01) G16H 70/00 (2018.01)
(52) CPC특허분류
G16H 50/80 (2018.01)
G06N 20/00 (2021.08)
(21) 출원번호 10-2023-0061833
(22) 출원일자 2023년05월12일
심사청구일자 2023년05월12일
(30) 우선권주장
1020220079373 2022년06월29일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
황인준
서울특별시 성동구 매봉길 15(옥수동, 래미안 옥수리버젠)
정원용
서울특별시 서초구 잠원로 88(잠원동, 신반포아파트)
(74) 대리인
(뒷면에 계속)
특허법인주연케이알피

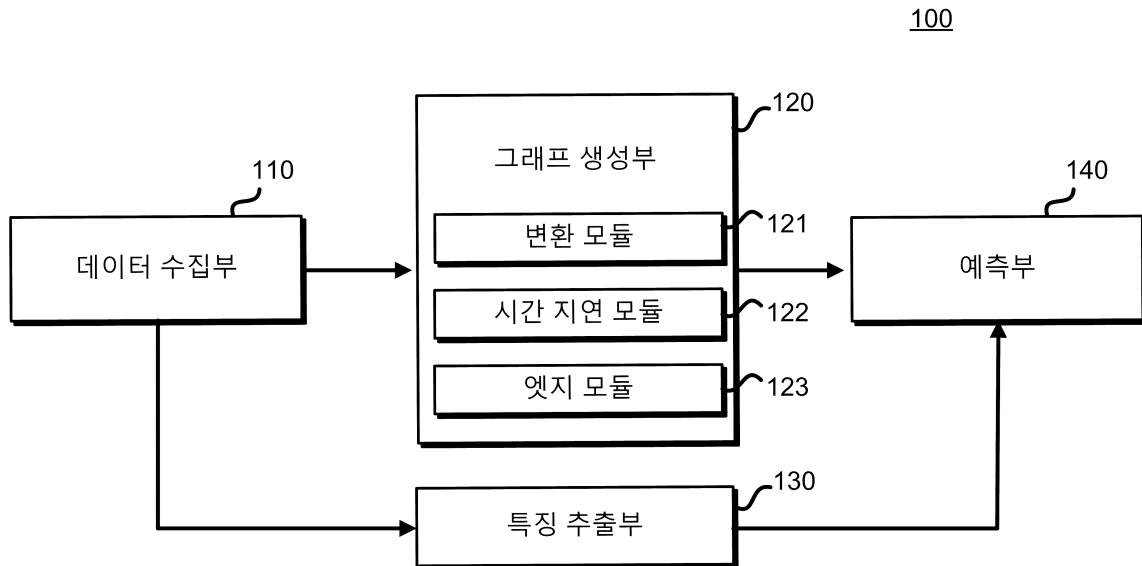
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 지역 유사도를 통한 인플루엔자 예측 장치 및 방법

(57) 요약

본 발명은 인플루엔자 확산 예측 장치에 관한 것이다. 본 발명의 일 실시예에 따른 인플루엔자 확산 예측 장치는 복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집하는 데이터 수집부, 상기 다변량 시계열 데이터에 기초하여 그래프 구조를 생성하는 그래프 생성부; 상기 다변량 시계열 데이터의 적어도 하나 이상의 신경망

(뒷면에 계속)
대표도 - 도1



을 활용하여 시간적 특징을 추출하는 특징 추출부; 및 상기 그래프 구조와 상기 시간적 특징을 이용하여 연산 그래프를 생성하고, 상기 연산 그래프로부터 인플루엔자 확산을 예측하는 예측부를 포함하고, 상기 그래프 생성부는: 상기 다변량 시계열 데이터의 노드 간의 공간 유사도를 계산하고, 상기 노드 간의 상관 관계를 설정하는 변환 모듈; 상기 그래프 구조에서 시계열적 특징이 상이한 상기 노드 간의 유사도를 확인하여 가중치를 설정하는 시간 지연 모듈; 및 상기 그래프 구조에서 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 엣지 모듈을 포함할 수 있다.

(52) CPC특허분류

G06N 3/04 (2023.01)
G06N 3/08 (2023.01)
G16H 10/20 (2021.08)
G16H 10/60 (2021.08)
G16H 50/50 (2018.01)
G16H 70/00 (2021.08)

노윤아

서울특별시 도봉구 도봉로150길 42(방학동, 삼성래미안아파트2차)

(72) 발명자

문재욱

서울특별시 광진구 아차산로78길 7(광장동, 유진스텔)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465036120
과제번호	HG19C0682020022
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	감염병위기대응기술개발(방역연계 범부처 감염병 연구개발사업단)
연구과제명	감염병 예측 정확도 향상을 위한 딥러닝 모델 고도화 연구
기 여 율	1/1
과제수행기관명	고려대학교산학협력단
연구기간	2022.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집하는 데이터 수집부,
 상기 다변량 시계열 데이터에 기초하여 그래프 구조를 생성하는 그래프 생성부;
 상기 다변량 시계열 데이터의 적어도 하나 이상의 신경망을 활용하여 시간적 특징을 추출하는 특징 추출부; 및
 상기 그래프 구조와 상기 시간적 특징을 이용하여 연산 그래프를 생성하고, 상기 연산 그래프로부터 인플루엔자 확산을 예측하는 예측부를 포함하고,
 상기 그래프 생성부는:
 상기 다변량 시계열 데이터의 노드 간의 공간 유사도를 계산하고, 상기 노드 간의 상관 관계를 설정하는 변환 모듈;
 상기 그래프 구조에서 시계열적 특징이 상이한 상기 노드 간의 유사도를 확인하여 가중치를 설정하는 시간 지연 모듈; 및
 상기 그래프 구조에서 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 엣지 모듈을 포함하는 인플루엔자 확산 예측 장치.

청구항 2

제1항에 있어서,
 상기 변환 모듈은,
 상기 다변량 시계열 데이터로부터 고속 푸리에 변환을 통해 상기 노드 간의 유사도를 계산하는 인플루엔자 확산 예측 장치.

청구항 3

제2항에 있어서,
 상기 엣지 모듈은,
 상기 노드 간의 연결 관계를 복수로 가지는 허브 노드의 가중치를 설정하는 인플루엔자 확산 예측 장치.

청구항 4

제1항에 있어서,
 상기 특징 추출부는,
 상기 다변량 시계열 데이터로부터 k (여기서, k 는 자연수)개의 필터를 이용하여 k 개의 시간적 특징을 추출하는 인플루엔자 확산 예측 장치.

청구항 5

제1항에 있어서,
 상기 다변량 시계열 데이터는 인플루엔자 또는 유사 질병의 지역별 확진자 수, 날짜별 확진자 수, 사망자 수 및 감염 재생산 지수를 포함하는 인플루엔자 확산 예측 장치.

청구항 6

인플루엔자 확산 예측 장치의 동작 방법으로서,

복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집하는 단계;

상기 다변량 시계열 데이터에 기초하여 그래프 구조를 생성하는 단계;

상기 다변량 시계열 데이터의 적어도 하나 이상의 신경망을 활용하여 시간적 특징을 추출하는 단계; 및

상기 그래프 구조와 상기 시간적 특징을 이용하여 연산 그래프를 생성하고, 상기 연산 그래프로부터 인플루엔자 확산을 예측하는 단계를 포함하고,

상기 그래프 구조를 생성하는 단계는:

상기 다변량 시계열 데이터의 노드 간의 공간 유사도를 계산하고, 상기 노드 간의 상관 관계를 설정하는 단계;

상기 그래프 구조에서 시계열적 특징이 상이한 상기 노드 간의 유사도를 확인하여 가중치를 설정하는 단계; 및

상기 그래프 구조에서 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 단계를 포함하는 인플루엔자 확산 예측 방법.

청구항 7

제6항에 있어서,

상기 노드 간의 상관 관계를 설정하는 단계는,

상기 다변량 시계열 데이터로부터 고속 푸리에 변환을 통해 상기 노드 간의 유사도를 계산하는 인플루엔자 확산 예측 방법.

청구항 8

제7항에 있어서,

상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 단계는,

상기 노드 간의 연결 관계를 복수로 가지는 허브 노드의 가중치를 설정하는 인플루엔자 확산 예측 방법.

청구항 9

제6항에 있어서,

시간적 특징을 추출하는 단계는,

상기 다변량 시계열 데이터로부터 k (여기서, k 는 자연수)개의 필터를 이용하여 k 개의 시간적 특징을 추출하는 인플루엔자 확산 예측 방법.

청구항 10

제6항에 있어서,

상기 다변량 시계열 데이터는 인플루엔자 또는 유사 질병의 지역별 확진자 수, 날짜별 확진자 수, 사망자 수 및 감염 재생산 지수를 포함하는 인플루엔자 확산 예측 방법.

발명의 설명

기술 분야

[0001] 본 발명은 지역 간 유사도를 통한 인플루엔자 예측 장치 및 방법에 관한 것이다.

배경 기술

[0002] 최근 세계화에 따른 활발한 인적, 물적 교류로 인해, 인플루엔자와 같은 감염병에 의한 피해가 증가하고 있다. 이러한 피해를 줄이기 위해 감염병의 미래 발생량을 예측하기 위한 다양한 연구들이 진행되었으며, 심층학습 기반의 예측 모델을 활용한 연구가 활발하게 이루어지고 있다.

[0003] 다만, 종래의 예측 모델은 여러 지역의 인플루엔자 발생량을 동시에 예측하는데 어려움이 있으며, 여러 지역 간의 상관관계를 학습하지 않아 여러 지역에서 동시에 발생하는 감염병을 정확하게 예측하기에 어려움이 있다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 대한민국 등록특허 제 10-1728506호

발명의 내용

해결하려는 과제

[0005] 본 발명은 상술한 과제를 해결하기 위한 것으로서, 본 발명의 목적은 지역 간 유사도를 통한 인플루엔자 예측 장치 및 방법을 제공하는 데 있다.

과제의 해결 수단

[0006] 본 발명은 복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집하는 데이터 수집부, 상기 다변량 시계열 데이터에 기초하여 그래프 구조를 생성하는 그래프 생성부; 상기 다변량 시계열 데이터의 적어도 하나 이상의 신경망을 활용하여 시간적 특징을 추출하는 특징 추출부; 및 상기 그래프 구조와 상기 시간적 특징을 이용하여 연산 그래프를 생성하고, 상기 연산 그래프로부터 인플루엔자 확산을 예측하는 예측부를 포함하고, 상기 그래프 생성부는: 상기 다변량 시계열 데이터의 노드 간의 공간 유사도를 계산하고, 상기 노드 간의 상관 관계를 설정하는 변환 모듈; 상기 그래프 구조에서 시계열적 특징이 상이한 상기 노드 간의 유사도를 확인하여 가중치를 설정하는 시간 지연 모듈; 및 상기 그래프 구조에서 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 엣지 모듈을 포함할 수 있다.

[0007] 본 발명의 일 실시예에 있어서, 상기 변환 모듈은, 상기 다변량 시계열 데이터로부터 고속 푸리에 변환을 통해 상기 노드 간의 유사도를 계산할 수 있다.

[0008] 본 발명의 일 실시예에 있어서, 상기 엣지 모듈은, 상기 노드 간의 연결 관계를 복수로 가지는 허브 노드의 가중치를 설정할 수 있다.

[0009] 본 발명의 일 실시예에 있어서, 상기 특징 추출부는, 상기 다변량 시계열 데이터로부터 k (여기서, k 는 자연수)개의 필터를 이용하여 k 개의 시간적 특징을 추출할 수 있다.

[0010] 본 발명의 일 실시예에 있어서, 상기 다변량 시계열 데이터는 인플루엔자 또는 유사 질병의 지역별 확진자 수, 날짜별 확진자 수, 사망자 수 및 감염 재생산 지수를 포함할 수 있다.

[0011] 본 발명은 인플루엔자 확산 예측 장치의 동작 방법으로서, 복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집하는 단계; 상기 다변량 시계열 데이터에 기초하여 그래프 구조를 생성하는 단계; 상기 다변량 시계열 데이터의 적어도 하나 이상의 신경망을 활용하여 시간적 특징을 추출하는 단계; 및 상기 그래프 구조와 상기 시간적 특징을 이용하여 연산 그래프를 생성하고, 상기 연산 그래프로부터 인플루엔자 확산을 예측하는 단계를 포함하고, 상기 그래프 구조를 생성하는 단계는: 상기 다변량 시계열 데이터의 노드 간의 공간 유사도를 계산하고, 상기 노드 간의 상관 관계를 설정하는 단계; 상기 그래프 구조에서 시계열적 특징이 상이한 상기 노드 간의 유사도를 확인하여 가중치를 설정하는 단계; 및 상기 그래프 구조에서 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 단계를 포함할 수 있다.

[0012] 본 발명의 일 실시예에 있어서, 상기 노드 간의 상관 관계를 설정하는 단계는, 상기 다변량 시계열 데이터로부터 고속 푸리에 변환을 통해 상기 노드 간의 유사도를 계산할 수 있다.

[0013] 본 발명의 일 실시예에 있어서, 상기 노드 간의 연결 관계를 확인하여 가중치를 설정하는 단계는, 상기 노드 간의 연결 관계를 복수로 가지는 허브 노드의 가중치를 설정할 수 있다.

[0014] 본 발명의 일 실시예에 있어서, 시간적 특징을 추출하는 단계는, 상기 다변량 시계열 데이터로부터 k (여기서, k 는 자연수)개의 필터를 이용하여 k 개의 시간적 특징을 추출할 수 있다.

[0015] 본 발명의 일 실시예에 있어서, 상기 다변량 시계열 데이터는 인플루엔자 또는 유사 질병의 지역별 확진자 수, 날짜별 확진자 수, 사망자 수 및 감염 재생산 지수를 포함할 수 있다.

발명의 효과

[0016] 본 발명에 따른 인플루엔자 확산 예측 장치는 지역 간의 유사도 및 상관관계를 학습할 수 있다. 이에 따라, 인플루엔자 확산 예측 장치는 보다 정확한 예측 성능을 제공할 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치의 일 예를 보여주는 블록도이다.

도 2는 본 발명의 실시 예에 따른 인플루엔자 확산 예측 모델의 구조도를 보여주는 도면이다.

도 3은 본 발명의 실시 예에 따른 그래프 생성부의 일 예를 보여주는 도면이다.

도 4는 본 발명의 실시 예에 따른 지연 시간의 일 예를 보여주는 도면이다.

도 5는 본 발명의 실시 예에 따른 인플루엔자 확산 예측 모델의 일 예를 보여주는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 이하에서, 본 발명의 기술 분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있을 정도로, 본 발명의 실시 예들이 명확하고 상세하게 기재될 것이다.

[0019] 도 1은 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치의 일 예를 보여주는 블록도이다.

[0020] 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치(100)는 하나 이상의 결과 변수를 포함하는 다변량(Multivariate) 시계열 데이터에 기초하여 유사도를 계산하고, 인공지능 모델을 통해 인플루엔자 확산을 예측할 수 있다.

[0021] 도 1을 참조하여 좀더 자세히 설명하면, 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치(100)는 데이터 수집부(110), 그래프 생성부(120), 특징 추출부(130) 및 예측부(140)를 포함할 수 있다.

[0022] 데이터 수집부(110)는 복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집할 수 있다. 예를 들어, 다변량 시계열 데이터는 인플루엔자 또는 유사 질병의 지역별 확진자 수, 날짜별 확진자 수, 사망자 수, 감염 재생산 지수 등일 수 있다.

[0023] 그래프 생성부(120)는 수집한 다변량 시계열 데이터의 노드(예를 들어, 지역) 별 유사도를 계산하고, 각 노드 및 엣지에 따라 가중치를 부여하고, 그래프 구조를 생성할 수 있다. 이를 위해 그래프 생성부(120)는 변환 모듈(121), 시간 지연 모듈(122) 및 엣지 모듈(123)을 포함할 수 있다.

[0024] 변환 모듈(121)은 다변량 시계열 데이터를 공간적 영역으로 변환할 수 있다. 예를 들어, 변환 모듈(121)은 고속 푸리에 변환(Fast Fourier Transform, FFT)을 통해 다변량 시계열 데이터 노드 간의 공간 유사도를 계산하여 그래프 구조를 생성할 수 있다. 이를 통해, 변환 모듈(121)은 서로 연관이 있는 지역을 연결하고, 연관된 정도에 따라 지역 간 상관 관계를 개별로 설정할 수 있다.

[0025] 시간 지연 모듈(122)은 그래프 구조에서 시계열적 특징이 상이한 노드 간의 유사도를 확인하여 가중치를 설정할 수 있다. 예를 들어, 시간 지연 모듈(122)은 데이터의 시계열적 특징으로 인해 시간 지연(Time Delay)가 발생하더라도 노드 간의 유사도를 판단하여 각 노드에 가중치를 설정할 수 있다.

[0026] 엣지 모듈(123)은 그래프 구조에서 노드 간의 연결 관계를 확인하여 가중치를 설정할 수 있다. 예를 들어, 엣지 모듈(123)은 허브 노드의 가중치를 높이기 위해 각 노드의 연결 수에 따라 가중치를 설정할 수 있다.

[0027] 특징 추출부(130)는 수집한 다변량 시계열 데이터의 적어도 하나 이상의 신경망을 활용하여 시간적 특징을 추출할 수 있다. 신경망은 합성곱 신경망(Convolutional Neural Network, CNN), 심층 신경망(Deep Neural Network, DNN), 순환 신경망(Recurrent Neural Network, RNN), 제한된 볼츠만 머신(Restricted Boltzmann Machine, RBM), 심층 신경망(Deep Belief Network, DBN), 양방향 순환 심층 신경망(Bidirectional Recurrent Deep Neural Network, BRDNN) 또는 심층 Q-네트워크(Deep Q-Networks) 등이 있지만, 이에 한정되는 것은 아니다.

[0028] 본 발명의 일 실시예에 따른 특징 추출부(130)는 순환 신경망(RNN)을 이용하여 다변량 시계열 데이터에서 시간

적 특징을 포함하는 은닉 상태(Hidden State)를 어텐션 매트릭스(Attention Matrix)로 임베딩(Embedding)할 수 있다. 특징 추출부(130)는 합성곱 신경망(CNN)을 이용하여 다변량 시계열 데이터에서 시간적 특징을 추출할 수 있다.

[0029] 예측부(140)는 그래프 구조와 시간적 특징을 이용하여 연산 그래프(Computation Graph)를 생성하고, 그래프 신경망(Graph Neural Network, GNN)을 이용하여 연산 그래프와 시간적 특징을 결합하여 인플루엔자 확산을 예측할 수 있다.

[0030] 이상에서 살펴본 바와 같이, 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치(100)는 유사도, 시간 지연 및 연결관계의 가중치를 가지는 그래프 구조를 생성할 수 있다. 이에 따라, 인플루엔자 확산 예측 장치(100)는 예측 데이터의 정확도를 높일 수 있다.

[0032] 도 2는 본 발명의 실시 예에 따른 인플루엔자 확산 예측 모델의 구조도를 보여주는 도면이다.

[0033] 도 2를 참조하면, 인플루엔자 확산 예측 장치(100)는 다변량 시계열 데이터로서 여러 지역의 인플루엔자 감염자 수 데이터를 수집하고, 각 지역의 감염자 수를 예측할 수 있다.

[0034] 그래프 생성부(120)는 여러 지역의 인플루엔자 감염자 수 데이터를 고속 푸리에 변환을 통해 스펙트럼 단위로 감염자 수 데이터를 분해할 수 있다. 예를 들어, 그래프 생성부(120)는 서로 지역이 상이한 i 지역과 j 지역의 감염자 수 데이터인 X_i 와 X_j 에 고속 푸리에 변환을 적용하여 주파수 t 까지의 진폭 I 와 J 의 합으로 표현할 수 있다. 또한, 그래프 생성부(120)는 두 지역의 스펙트럼 데이터를 내적하여 i 지역과 j 지역의 유사도 g_{ij} 를 계산할 수 있다. 이에 따라 그래프 생성부(120)는 모든 지역 간 유사도 g_{ij} 들의 집합이 지역 간 관계 행렬 G 로 정의할 수 있으며, 아래 수학적 식 1으로 정의될 수 있다.

수학적 식 1

$$FFT(X_i) \approx \sum_{k=1}^t I_k, \quad FFT(X_j) \approx \sum_{k=1}^t J_k$$

$$g_{ij} = FFT(X_i) \cdot FFT(X_j)$$

[0035]

[0036] 특징 추출부(130)는 각 지역의 인플루엔자 발생량 데이터의 시간적 특징(Temporal Feature)을 추출할 수 있다. 예를 들어, 특징 추출부(130)는 여러 지역의 인플루엔자 감염자 수 데이터에 확대 비율 k 만큼의 필터 크기를 증가시킨 여러 필터를 이용하여 $1 \times k$ dilated 컨볼루션을 사용할 수 있다. 이에 따라, 특징 추출부(130)는 k 개 주기의 시간적 특징을 추출할 수 있으며, 추출된 시간적 특징을 하나로 연결하여 지역 행렬의 값으로 설정할 수 있다.

[0037] 예측부(140)는 그래프 컨볼루션을 사용하여 각 지역의 정보를 지역 행렬로 설정하고, 지역을 연결한 지역 간 관계 행렬을 통해 이웃 지역의 정보를 받아 지역 정보를 업데이트할 수 있다. 즉, 예측부(140)는 시간적 특징을 포함한 지역 행렬과 공간적 특징을 포함한 지역 간 관계 행렬 G 를 사용하여 시공간 패턴을 학습할 수 있다.

[0038] 또한, 예측부(140)는 학습한 시공간 패턴을 기반으로 인플루엔자 감염자 수 데이터를 가지는 다른 지역의 데이터를 업데이트할 수 있다. 이에 따라, 예측부(140)는 타겟 지역 벡터와 이웃 지역 정보를 합친 벡터를 연결하여 하나의 벡터로 생성할 수 있으며, 생성한 벡터를 다층 퍼셉트론(Multi-Layer Perceptron, MLP)에 입력하여 최종 예측을 진행할 수 있다.

[0039] 이상에서 살펴본 바와 같이, 본 발명의 실시 예에 따른 인플루엔자 확산 예측 장치(100)는 시공간 패턴을 학습하고, 다른 지역의 데이터를 업데이트할 수 있다. 이에 따라, 인플루엔자 확산 예측 장치(100)는 보다 정확한 예측 데이터를 생성할 수 있다.

[0041] 도 3은 본 발명의 실시 예에 따른 그래프 생성부의 일 예를 보여주는 도면이다.

- [0042] 도 3을 참조하면, 변환 모듈(121)은 적어도 하나 이상의 노드를 포함하는 다변량 시계열 데이터에 고속 푸리에 변환(FFT)을 적용하여 유사도 매트릭스(Similarity Matrix)를 생성할 수 있다.
- [0043] 시간 지연 모듈(122)은 유사도 매트릭스에서 시간 지연을 고려하여 특정 노드와 유사도가 높은 노드에 가중치를 설정하여 업데이트된 매트릭스(Updated Matrix)를 생성할 수 있다.
- [0044] 엣지 모듈(123)은 특정된 각 노드들의 연결 관계를 확인하여 가중치를 설정할 수 있다.
- [0045] 상술한 바와 같이, 본 발명의 실시 예에 따른 그래프 생성부(120)는 변환된 매트릭스 각 노드에 유사도, 시간 지연 및 연결관계의 가중치를 부여하여 그래프 구조를 생성하고, 예측부(140)에 출력할 수 있다.
- [0047] 도 4는 본 발명의 실시 예에 따른 지연 시간의 일 예를 보여주는 도면이다. 설명의 편의상, 도 4에서는 다변량 시계열 데이터로서 캘리포니아 지역과 조지아 지역의 인플루엔자 감염자 수 데이터를 이용해 예시될 것이다.
- [0048] 인플루엔자 확산 예측 장치(100)는 다변량 시계열 데이터로서 여러 지역들의 인플루엔자 감염자 수 데이터가 수집되면, 예컨대, 도 4에 도시된 바와 같이 지역 간의 유사도를 계산하여 그래프 구조를 생성할 수 있다. 예를 들어, 인플루엔자 확산 예측 장치(100)는 캘리포니아 지역과 조지아 지역의 인플루엔자 감염자 수 데이터의 감염자 비율, 감염자의 증가량 및 감소량 등 유사도를 계산할 수 있다.
- [0049] 한편, 시간 지연 모듈(122)은 시간 지연이 발생하더라도 감염자 수 데이터의 유사도를 판단하여 노드(즉, 지역)의 가중치를 설정할 수 있다.
- [0050] 상술한 바와 같이, 인플루엔자 확산 예측 장치(100)는 지역 간의 상관 관계를 계산하고, 시간의 변수를 포함함으로써 감염병 확산을 보다 정확하게 예측할 수 있다.
- [0052] 도 5는 본 발명의 실시 예에 따른 인플루엔자 확산 예측 모델의 일 예를 보여주는 순서도이다.
- [0053] 도 5를 참조하면, S110 단계에서, 데이터 수집부(110)는 복수개 이상의 결과 변수를 포함하는 다변량 시계열 데이터를 수집할 수 있다.
- [0054] S120 단계에서, 그래프 생성부(120)는 다변량 시계열 데이터의 노드 별 유사도를 계산하고, 각 노드 및 엣지에 따라 가중치를 부여하고, 그래프 구조를 생성할 수 있다.
- [0055] S130 단계에서, 특징 추출부(130)는 다변량 시계열 데이터에서 시간적 특징을 추출할 수 있다.
- [0056] S140 단계에서, 예측부(140)는 그래프 구조와 시간적 특징을 이용하여 연산 그래프를 생성할 수 있다. 다시 말해, 예측부(140)는 그래프 구조와 시간적 특징에 그래프 컨볼루션하여 시공간 패턴을 학습한 연산 그래프를 생성할 수 있다.
- [0057] S150 단계에서, 예측부(140)는 연산 그래프에 기반하여 각 노드를 업데이트하고, 노드 데이터의 벡터를 추출하여 예측값을 출력할 수 있다.
- [0059] 이하, 실험예를 통해서 본 발명을 보다 상세히 설명하기로 한다. 하지만, 이들은 본 발명을 보다 상세하게 설명하기 위한 것일 뿐 본 발명의 권리범위가 이에 한정되는 것은 아니다.
- [0060] **실험예 1 : 예측을 위한 통계적 기법, 심층학습 기법 및 본 발명에 대한 평가 비교**
- [0061] 본 발명의 인플루엔자 확산 예측 방법과 종래의 예측 모델과의 성능을 비교하기 위해 실험을 실시하였다.
- [0062] 본 실험은 미국과 일본의 인플루엔자 감염자 수 데이터셋을 기반으로 실험하였으며, 비교 모델은 Auto Regressive(AR), ARMA, MLP RNN, Long Short-Term Memory(LSTM), LSTNet-Skip, ST-GCN 및 Cola-GNN을 사용하였다. 예측 실험은 입력 시퀀스 50이 입력될 때, 2, 3, 5, 10, 15주의 감염자 수를 예측하는 방식이며, 훈련-검증-테스트 데이터셋은 60%-20%-20%의 비율로 설정하였으며, 배치 크기는 32, 학습 반복 횟수는 1000회, 최적화 함수는 Adam을 사용하였다.
- [0063] 또한, 본 실험의 모델의 예측 성능을 평가하기 위한 지표로 Root Mean Square Error(RMSE)와 Pearson Correlation Coefficient(PCC)를 사용하였다.

표 1

Horizon	2		3		5		10		15	
Metric	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC
AR	161	0.94	204	0.909	251	0.863	306	0.773	327	0.723
ARMA	161	0.939	200	0.909	250	0.862	306	0.773	326	0.725
MLP	159	0.938	200	0.917	236	0.891	294	0.791	320	0.739
RNN	149	0.948	181	0.922	217	0.886	274	0.821	315	0.758
LSTM	150	0.948	180	0.922	213	0.889	276	0.820	307	0.771
LSTNet-skip	199	0.913	249	0.850	299	0.759	292	0.760	292	0.802
ST-GCN	189	0.907	209	0.778	256	0.823	289	0.769	292	0.774
Cola-GNN	136	0.955	167	0.933	205	0.887	241	0.822	237	0.856
Our model	141	0.947	151	0.949	202	0.897	229	0.841	221	0.880

표 1은 미국 인플루엔자 감염자 수 데이터셋의 예측 결과로서 비교 모델보다 예측 시점과 평가 지표에 대해서 우수한 것을 확인할 수 있다. 또한, 사전에 지역 거리를 기반으로 그래프 구조를 설정하여 입력하는 종래 기술보다 우수한 성능을 확인할 수 있다.

표 2

Horizon	2		3		5		10		15	
Metric	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC	RMSE	PCC
AR	1377	0.752	1705	0.579	2013	0.31	2107	0.238	2042	0.483
ARMA	1371	0.754	1703	0.579	2013	0.31	2105	0.238	2041	0.486
MLP	1190	0.817	1437	0.716	1712	0.46	1905	0.497	1827	0.604
RNN	1001	0.892	1259	0.833	1376	0.821	1696	0.616	1629	0.709
LSTM	1052	0.896	1246	0.873	1335	0.853	1622	0.681	1649	0.695
LSTNet-skip	1133	0.846	1459	0.728	1883	0.432	1811	0.518	1884	0.515
ST-GCN	996	0.902	1115	0.88	1129	0.872	1541	0.735	1527	0.773
Cola-GNN	929	0.915	1051	0.901	1117	0.89	1372	0.810	1475	0.753
Our model	918	0.922	1030	0.914	1088	0.898	1398	0.813	1446	0.758

표 2는 일본 인플루엔자 감염자 수 데이터셋의 예측 결과로서 비교 모델보다 예측 시점과 평가 지표에 대해서 우수한 것을 확인할 수 있다. 또한, 사전에 지역 거리를 기반으로 그래프 구조를 설정하여 입력하는 종래 기술보다 우수한 성능을 확인할 수 있다.

상술한 바와 같이, 본 발명의 일 실시 예에 따른 인플루엔자 확산 예측 장치(100)는 보다 정확한 인플루엔자 확산 예측 모델을 구성할 수 있다.

상술한 내용은 본 발명을 실시하기 위한 구체적인 실시 예들이다. 본 발명은 상술한 실시 예들 이외에도, 단순히 설계 변경되거나 용이하게 변경할 수 있는 실시 예들도 포함될 것이다. 또한, 본 발명은 실시 예들을 이용하여 용이하게 변형하여 실시할 수 있는 기술들도 포함될 것이다. 따라서, 본 발명의 범위는 상술한 실시 예들에 국한되어 정해져서는 안되며, 후술하는 특허청구범위뿐만 아니라 이 발명의 특허청구범위와 균등한 것들에 의해 정해져야 할 것이다.

부호의 설명

100 : 인플루엔자 확산 예측 장치

110 : 데이터 수집부 121 : 변환 모듈

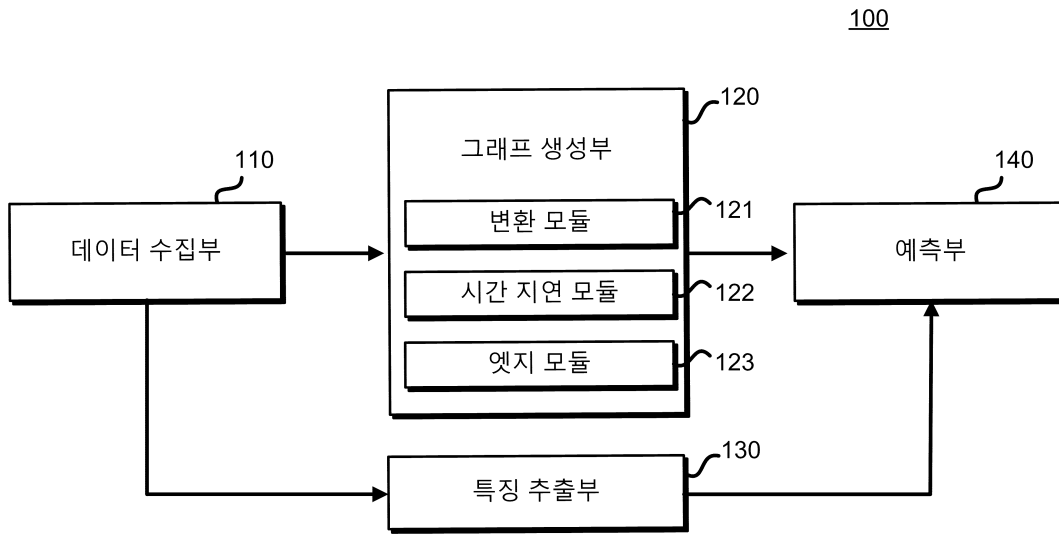
120 : 그래프 생성부 122 : 시간 지연 모듈

130 : 특징 추출부 123 : 엣지 모듈

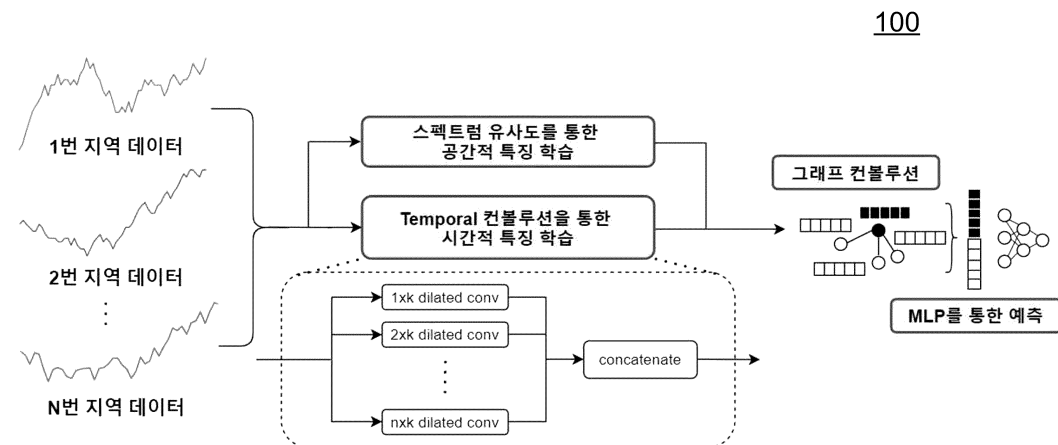
140 : 예측부

도면

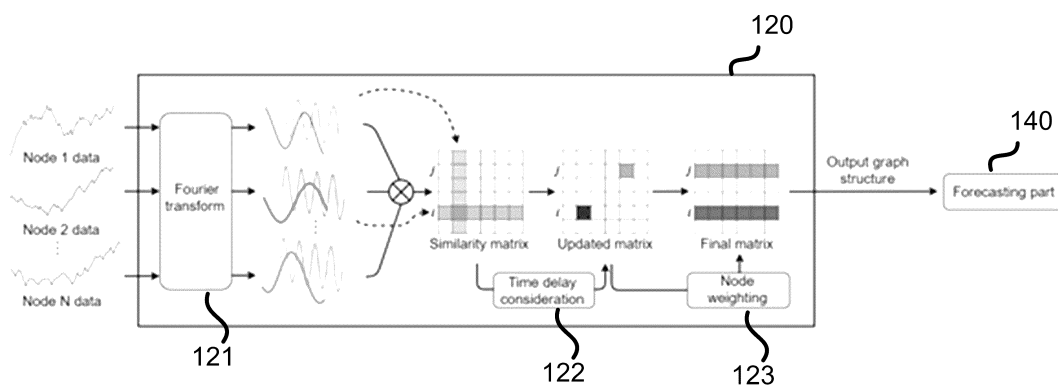
도면1



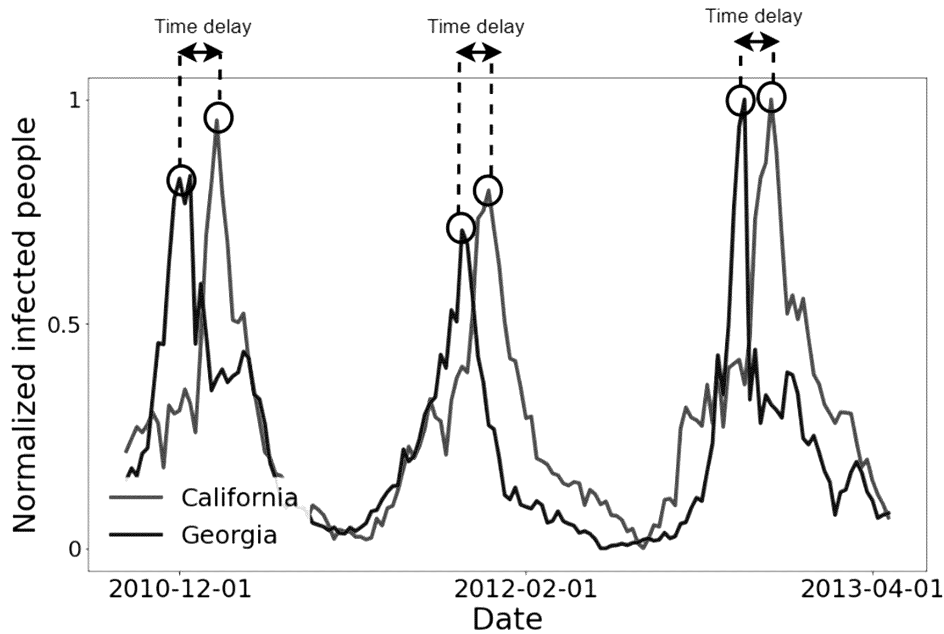
도면2



도면3



도면4



도면5

