

**(19) 대한민국특허청(KR)**
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0095496

(43) 공개일자 2023년06월29일

(51) 국제특허분류(Int. Cl.)

G16H 50/80 (2018.01) G06N 3/08 (2023.01)

G16H 50/20 (2018.01) G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

(52) CPC특허분류

G16H 50/80 (2018.01)

G06N 3/08 (2023.01)

(21) 출원번호 10-2021-0185040

(22) 출원일자 2021년12월22일

심사청구일자 2021년12월22일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

황인준

서울특별시 성동구 매봉길 15(옥수동, 래미안 옥수 리버젠)

문재욱

서울특별시 광진구 아차산로78길 7(광장동, 유진스텔)

정승원

서울특별시 노원구 우이천로 102(월계동, 초안아파트)

(74) 대리인

특허법인주연케이알피

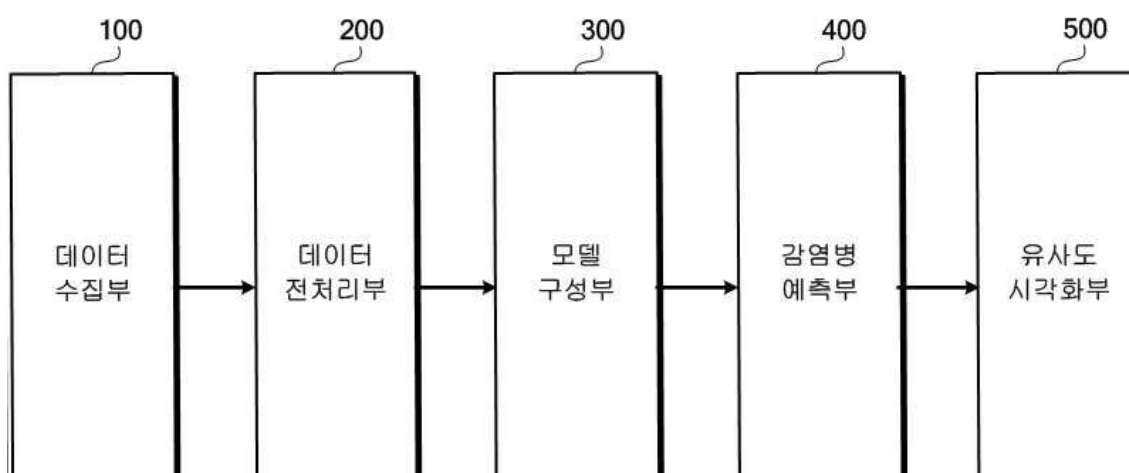
전체 청구항 수 : 총 13 항

(54) 발명의 명칭 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치 및 방법

(57) 요약

본 출원은 지역별 감염병을 예측할 수 있는 장치 및 방법에 관한 것이다. 본 출원의 실시 예에 따른 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치는 서로 다른 복수의 지역들의 감염병 데이터를 수집하는 데이터 수집부; 상기 데이터 수집부로부터 상기 서로 다른 복수의 지역들의 감염병 데이터를 수신하며, 상기 서로 다른 복수

(뒷면에 계속)

대표도 - 도110

의 지역들의 감염병 데이터에 대한 전처리 동작을 수행하여 입력 데이터를 생성하는 데이터 전처리부; 상기 데이터 전처리부로부터 입력 데이터를 수신하며, 상기 입력 데이터에 기초하여 각 지역에 대응하는 쿼리, 키, 밸류를 생성하는 모델 구성부; 및 상기 모델 구성부로부터 각 지역에 대응하는 쿼리, 키, 밸류를 수신하며, 상기 각 지역에 대응하는 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하는 감염병 예측부를 포함한다. 본 출원의 실시 예에 따른 지역별 감염병 예측 장치 및 방법은 자기 주의 집중 모델을 이용하여 지역들 사이의 데이터 상관관계를 학습함으로써, 미래의 지역별 감염병 발생량을 정확하게 예측할 수 있다.

(52) CPC특허분류

G16H 50/20 (2018.01)

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465033193
과제번호	HG19C0682020021
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	감염병위기대응기술개발(R&D)
연구과제명	감염병 예측 정확도 향상을 위한 딥러닝 모델 고도화 연구
기 여 율	1/1
과제수행기관명	고려대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

명세서

청구범위

청구항 1

서로 다른 복수의 지역들의 감염병 데이터를 수집하는 데이터 수집부;

상기 데이터 수집부로부터 상기 서로 다른 복수의 지역들의 감염병 데이터를 수신하며, 상기 서로 다른 복수의 지역들의 감염병 데이터에 대한 전처리 동작을 수행하여 입력 데이터를 생성하는 데이터 전처리부;

상기 데이터 전처리부로부터 입력 데이터를 수신하며, 상기 입력 데이터에 기초하여 각 지역에 대응하는 쿼리, 키, 밸류를 생성하는 모델 구성부; 및

상기 모델 구성부로부터 각 지역에 대응하는 쿼리, 키, 밸류를 수신하며, 상기 각 지역에 대응하는 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하는 감염병 예측부를 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 2

제1 항에 있어서,

상기 데이터 전처리부는 상기 서로 다른 복수의 지역들의 감염병 데이터 중 결측 데이터는 0으로 대체하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 3

제1 항에 있어서,

상기 모델 구성부는

상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 쿼리를 생성하는 쿼리 레이어;

상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 키를 생성하는 키 레이어; 및

상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 밸류를 생성하는 밸류 레이어를 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 4

제3 항에 있어서,

상기 쿼리 레이어 및 상기 키 레이어는 LSTM을 통하여 구현되는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 5

제4 항에 있어서,

상기 밸류 레이어는 FC 레이어를 통하여 구현되는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 6

제1 항에 있어서,

상기 감염병 예측부는

상기 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하는 어텐션 레이어; 및

상기 지역들 사이의 상관관계에 대한 학습 결과에 기초하여, 미래의 각 지역에서의 감염병 발생량을 예측하는 예측 레이어를 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 7

제6 항에 있어서,

상기 어텐션 레이어는

상기 쿼리 및 키에 대한 내적 연산을 통하여, 상기 지역들 사이의 유사도를 출력하는 내적 모듈;

상기 지역들 사이의 유사도를 확률 값으로 변환하여 확률 유사도를 출력하는 소프트맥스 모듈; 및

상기 확률 유사도 및 상기 밸류 사이의 곱셈 연산을 통하여, 상기 지역들 각각에 대한 주의 집중 값을 출력하는 곱셈 모듈을 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 8

제6 항에 있어서,

상기 예측 레이어는 FC 레이어를 통하여 구현되는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 9

제1 항에 있어서,

상기 감염병 예측부로부터 상기 지역들 사이의 유사도 행렬을 수신하고, 수신된 유사도 행렬을 시각화하여 디스플레이하는 유사도 시각화부를 더 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치.

청구항 10

지역별 감염병 발생량에 대한 시계열 데이터를 수집하는 단계;

상기 시계열 데이터에 대한 전처리 동작을 수행하여 입력 데이터를 생성하는 단계;

상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 쿼리, 키, 밸류를 지역별로 정의하는 단계; 및

상기 쿼리, 키 및 밸류를 사용하여 지역들 사이의 상관관계를 학습하고, 미래의 감염병 발생량을 예측하는 단계를 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 방법.

청구항 11

제10 항에 있어서,

상기 자기 주의 집중 모델의 쿼리 및 키는 각각 LSTM을 통하여 정의되고, 상기 자기 주의 집중 모델의 밸류는 FC 레이어를 통하여 정의되는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 방법.

청구항 12

제10 항에 있어서,

상기 쿼리, 키 및 밸류를 사용하여 지역들 사이의 상관관계를 학습하는 단계는

상기 쿼리 및 키를 사용하여 상기 지역들 사이의 유사도를 계산하는 단계;

상기 지역들 사이의 유사도를 확률 유사도로 변환하는 단계; 및

상기 확률 유사도 및 상기 밸류에 기초하여, 주의 집중 값을 연산하는 단계를 포함하는, 자기 주의 집중 모델 기반의 지역별 감염병 예측 방법.

청구항 13

하드웨어와 결합되어 제10항 내지 제13항 중 어느 하나의 항의 방법을 실행시키기 위하여 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 출원은 지역별 감염병을 예측할 수 있는 장치 및 방법에 관한 것이다.

배경 기술

[0002] 감염병은 세균, 바이러스, 진균, 기생충, 스피로헤타, 리케차 등과 같은 여러 병원체에 의하여 감염되어 발병하는 질환이다. 예를 들어, 감염병 중 하나인 인플루엔자는 매년 전 세계적으로 큰 피해를 발생시키고 있으며, 최근에는 COVID-19가 세계 경제에 막대한 타격을 입히고 있다. 따라서, 이러한 감염병의 피해를 예방하고 효과적으로 대응하기 위해서는, 미래의 감염병의 발생 규모나 시기를 예측하는 것이 필요한 실정이다.

발명의 내용

해결하려는 과제

[0003] 본 출원의 목적은 미래의 지역별 감염병 발생량을 정확하게 예측할 수 있는 장치 및 방법을 제공하는데 있다.

과제의 해결 수단

[0004] 본 출원의 실시 예에 따른 자기 주의 집중 모델 기반의 지역별 감염병 예측 장치는 서로 다른 복수의 지역들의 감염병 데이터를 수집하는 데이터 수집부; 상기 데이터 수집부로부터 상기 서로 다른 복수의 지역들의 감염병 데이터를 수신하며, 상기 서로 다른 복수의 지역들의 감염병 데이터에 대한 전처리 동작을 수행하여 입력 데이터를 생성하는 데이터 전처리부; 상기 데이터 전처리부로부터 입력 데이터를 수신하며, 상기 입력 데이터에 기초하여 각 지역에 대응하는 쿼리, 키, 밸류를 생성하는 모델 구성부; 및 상기 모델 구성부로부터 각 지역에 대응하는 쿼리, 키, 밸류를 수신하며, 상기 각 지역에 대응하는 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하는 감염병 예측부를 포함한다.

[0005] 실시 예에 있어서, 상기 데이터 전처리부는 상기 서로 다른 복수의 지역들의 감염병 데이터 중 결측 데이터는 0으로 대체한다.

[0006] 실시 예에 있어서, 상기 모델 구성부는 상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 쿼리를 생성하는 쿼리 레이어; 상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 키를 생성하는 키 레이어; 및 상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 밸류를 생성하는 밸류 레이어를 포함한다.

[0007] 실시 예에 있어서, 상기 쿼리 레이어 및 상기 키 레이어는 LSTM을 통하여 구현된다.

[0008] 실시 예에 있어서, 상기 밸류 레이어는 FC 레이어를 통하여 구현된다.

[0009] 실시 예에 있어서, 상기 감염병 예측부는 상기 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하는 어텐션 레이어; 및 상기 지역들 사이의 상관관계에 대한 학습 결과에 기초하여, 미래의 각 지역에서의 감염병 발생량을 예측하는 예측 레이어를 포함한다.

[0010] 실시 예에 있어서, 상기 어텐션 레이어는 상기 쿼리 및 키에 대한 내적 연산을 통하여, 상기 지역들 사이의 유사도를 출력하는 내적 모듈; 상기 지역들 사이의 유사도를 확률 값으로 변환하여 확률 유사도를 출력하는 소프트맥스 모듈; 및 상기 확률 유사도 및 상기 밸류 사이의 곱셈 연산을 통하여, 상기 지역들 각각에 대한 주의 집중 값을 출력하는 곱셈 모듈을 포함한다.

[0011] 실시 예에 있어서, 상기 예측 레이어는 FC 레이어를 통하여 구현된다.

[0012] 실시 예에 있어서, 상기 감염병 예측부로부터 상기 지역들 사이의 유사도 행렬을 수신하고, 수신된 유사도 행렬을 시각화하여 디스플레이하는 유사도 시각화부를 더 포함한다.

[0013] 본 출원의 실시 예에 따른 자기 주의 집중 모델 기반의 지역별 감염병 예측 방법은 지역별 감염병 발생량에 대한 시계열 데이터를 수집하는 단계; 상기 시계열 데이터에 대한 전처리 동작을 수행하여 입력 데이터를 생성하는 단계; 상기 입력 데이터에 기초하여, 자기 주의 집중 모델의 쿼리, 키, 밸류를 지역별로 정의하는 단계; 및 상기 쿼리, 키 및 밸류를 사용하여 지역들 사이의 상관관계를 학습하고, 미래의 감염병 발생량을 예측하는 단계를 포함한다.

[0014] 실시 예에 있어서, 상기 자기 주의 집중 모델의 쿼리 및 키는 각각 LSTM을 통하여 정의되고, 상기 자기 주의 집

주 모델의 밸류는 FC 레이어를 통하여 정의된다.

[0015] 실시 예에 있어서, 상기 쿼리, 키 및 밸류를 사용하여 지역들 사이의 상관관계를 학습하는 단계는 상기 쿼리 및 키를 사용하여 상기 지역들 사이의 유사도를 계산하는 단계; 상기 지역들 사이의 유사도를 확률 유사도로 변환하는 단계; 및 상기 확률 유사도 및 상기 밸류에 기초하여, 주의 집중 값을 연산하는 단계를 포함한다.

발명의 효과

[0016] 본 출원의 실시 예에 따른 지역별 감염병 예측 장치 및 방법은 자기 주의 집중 모델을 이용하여 지역들 사이의 데이터 상관관계를 학습한다. 따라서, 지역별 감염병 예측 장치 및 방법은 미래의 지역별 감염병 발생량을 정확하게 예측할 수 있다.

도면의 간단한 설명

[0017] 도 1은 본 출원의 실시 예에 따른 지역별 감염병 예측 장치(10)를 보여주는 블록도이다.
 도 2는 도 1의 데이터 수집부(100)에서 수집하는 데이터의 일 예를 보여주는 도면이다.
 도 3은 도 1의 모델 구성부(300)의 구성을 좀 더 자세히 보여주는 블록도이다.
 도 4 및 도 5는 각각 모델 구성부(300) 중 쿼리 레이어(310) 및 키 레이어(320)의 일 예를 보여주는 블록도이다.
 도 6은 도 1의 감염병 예측부(400)를 좀 더 자세히 보여주는 블록도이다.
 도 7은 모델 구성부(300) 및 감염병 예측부(400)에 의하여 소정 지역(R_i)에서의 주의 집중 값(a_i)이 연산되는 과정을 보여주는 도면이다.
 도 8은 도 1의 유사도 시각화부(500)에 의하여 디스플레이되는 유사도의 일 예를 보여주는 도면이다.
 도 9는 도 1의 지역별 감염병 예측 장치(10)의 성능의 일 예를 보여주는 실험 결과이다.
 도 10은 본 출원의 실시 예에 따른 감염병 예측 방법을 보여주는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0018] 이하에서는, 본 출원의 기술분야에서 통상의 지식을 가진 자가 본 출원의 기술적 사상을 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여, 첨부된 도면을 참조하여 본 출원의 실시 예들이 설명될 것이다.

[0019] 도 1은 본 출원의 실시 예에 따른 지역별 감염병 예측 장치(10)를 보여주는 블록도이다. 도 1을 참조하면, 지역별 감염병 예측 장치(10)는 데이터 수집부(100), 데이터 전처리부(200), 모델 구성부(300) 및 감염병 예측부(400)를 포함한다. 또한, 추가적으로, 지역별 감염병 예측 장치(10)는 유사도 시각화부(500)를 더 포함할 수 있다.

[0020] 데이터 수집부(100)는 서로 다른 복수의 지역들에 대한 감염병 데이터를 수집한다. 예를 들어, 데이터 수집부(100)는 T주 동안의 N개 지역의 인플루엔자 발생량 시계열 데이터를 수집할 수 있으며, 이는 예측 모델의 입력 변수로서 예측 모델을 학습하는데 활용될 수 있다. 또한, 데이터 수집부(100)는 L주 후의 N개 지역의 인플루엔자 발생량 데이터를 수집할 수 있으며, 이는 예측 모델의 출력 변수로서 예측 모델을 학습하는데 활용될 수 있다.

[0021] 데이터 전처리부(200)는 데이터 수집부(100)로부터 서로 다른 복수의 지역들에 대한 감염병 데이터를 수신한다. 데이터 전처리부(200)는 예측 모델의 입력 변수로 활용될 수 있도록, 수집된 데이터에 대한 전처리 동작을 수행한다.

[0022] 예를 들어, 서로 다른 복수의 지역들에 대한 감염병 데이터 중에서 누락된 데이터인 데이터 결측치가 존재하는 경우, 데이터 전처리부(200)는 데이터 결측치를 '0'으로 대체할 수 있다.

[0023] 또한, 예를 들어, 지역별로 데이터 범위가 크게 차이가 나는 경우에는 예측 모델의 학습에 방해가 될 수 있다. 따라서, 효과적인 예측 모델의 학습을 위하여, 데이터 전처리부(200)는 지역별로 최소값은 '0', 최대값은 '1'로 정규화하는 최소-최대 정규화(min max normalization)를 수행할 수 있다.

[0024] 모델 구성부(300)는 데이터 전처리부(200)로부터 전처리된 지역별 감염병 데이터를 수신한다. 전처리된 지역별

감염병 데이터는 입력 데이터로 모델 구성부(300)를 학습하는데 사용될 수 있다.

- [0025] 모델 구성부(300)는 전처리된 지역별 감염병 데이터에 기초하여, 자기 주의 집중 모델(self-attention model)의 세 가지 구성 요소인 쿼리(query), 키(key), 밸류(value)를 지역별로 정의할 수 있다.
- [0026] 좀 더 자세히 설명하면, 일반적인 주의 집중 모델(attention model)은 주어진 입력 데이터에서 예측에 도움이 되는 데이터에 가중치를 두어 학습하는 심층 학습 모델이다. 주의 집중 모델은 두 데이터 사이의 관계를 학습하는데 많이 활용되며, 예를 들어, 자연어 처리, 이미지 처리, 시계열 예측 분야에서 활용되고 있다. 본 출원의 실시 예에 따른 지역별 감염병 예측 장치(10)는 이러한 일반적인 주의 집중 모델이 아닌, 하나의 데이터 안에서 두 개체 사이의 관계를 학습하는 자기 주의 집중 모델(self-attention model)을 사용한다. 이를 통해, 지역별 감염병 예측 장치(10)는 서로 다른 지역의 감염병 발생량 데이터로 구성된 입력 데이터 안에서 지역들 사이의 상관관계를 학습할 수 있다. 특히, 지역별 감염병 예측 장치(10)는 자기 집중 모델의 세 가지 구성요소 중 소정 지역의 쿼리(Query)를 해당 지역의 쿼리*Query와 유사한 키(Key)를 가진 다른 지역을 탐색하기 위한 기준의 역할로 사용한다. 지역별 감염병 예측 장치(10)는 쿼리(Query)와 키(Key)의 유사도를 구하여 이 유사도를 Value에 곱하여 최종 예측에 사용한다.
- [0027] 모델 구성부(300)는, 예를 들어, LSTM(Long Short-Term Memory)을 통하여 쿼리 및 키를 정의하고, FC 레이어(Fully-Connected layer)를 통하여 밸류를 정의할 수 있다. 모델 구성부(300)의 구성 및 동작은 이하의 도 3 내지 도 5 그리고 도 7에서 좀 더 자세히 설명될 것이다.
- [0028] 감염병 예측부(300)는 모델 구성부(300)에서 정의된 지역별 쿼리, 키, 밸류를 수신한다. 감염병 예측부(300)는 지역별 쿼리, 키, 밸류에 기초하여 지역들 사이의 상관관계를 학습하고, 이를 통해 미래의 감염병의 발생량을 예측할 수 있다.
- [0029] 예를 들어, N개 지역으로 구성된 입력 데이터에서 소정 지역(Ri)과 다른 지역 사이의 상관관계를 학습하고, 이를 해당 지역(Ri)의 미래 감염병 예측에 활용할 경우, 감염병 예측부(300)는 해당 지역(Ri)의 쿼리와 다른 지역들의 키 사이의 유사도를 계산할 수 있다. 감염병 예측부(300)는 계산된 유사도를 이용하여 해당 지역(Ri)의 감염병 발생량을 예측할 수 있다. 감염병 예측부(300)의 구성 및 동작은 이하의 도 6 및 도 7에서 좀 더 자세히 설명될 것이다.
- [0030] 유사도 시각화부(500)는 감염병 예측부(400)로부터 지역 사이의 유사도에 대한 정보를 수신하고, 이를 시각화할 수 있다. 예를 들어, 유사도 시각화부(500)는 시간에 따라 변화하는 각 지역 사이의 유사도를 시각화하여 디스플레이할 수 있다. 유사도 시각화부(500)는 이하의 도 8에서 좀 더 자세히 설명될 것이다.
- [0031] 상술한 바와 같이, 본 출원의 실시 예에 따른 지역별 감염병 예측 장치(10)는 자기 주의 집중 모델을 이용하여 지역들 사이의 데이터 상관관계를 학습한다. 따라서, 지역별 감염병 예측 장치 및 방법은 미래의 지역별 감염병 발생량을 정확하게 예측할 수 있다.
- [0032] 도 2는 도 1의 데이터 수집부(100)에서 수집하는 데이터의 일 예를 보여주는 도면이다. 도 2에서는, 예시적으로, 미국의 CDC로부터 수집한 미국의 여러 지역에서의 인플루엔자 발생량 데이터를 보여준다.
- [0033] 예를 들어, 데이터 수집부(100)는 미국의 여러 지역에서 발생하는 인플루엔자 발생량 데이터를 수집할 수 있다. 이 경우, 데이터 수집부(100)는 미국의 CDC(Centers for Disease Control and Prevention)로부터 지역별 인플루엔자 데이터를 수집할 수 있다.
- [0034] 미국의 CDC로부터 수집 가능한 데이터는 미국을 10개의 지역으로 나눈 데이터와 51개의 주 단위로 나눈 두 가지의 데이터가 있다. 두 가지의 데이터 모두 지역별 인플루엔자 발생량을 기록한 데이터이며, 각 지역의 시계열 데이터로 구성되어 있다.
- [0035] 데이터 수집부(100)에서 수집된 데이터는 데이터 전처리부(200)로 송신되며, 데이터 전처리부(200)는 예측 모델에 대한 입력 데이터로 사용될 수 있도록 수집된 데이터에 대한 전처리 동작을 수행한다.
- [0036] 도 3 내지 도 5는 도 1의 모델 구성부(300)를 좀 더 자세히 설명하기 위한 도면이다. 구체적으로, 도 3은 도 1의 모델 구성부(300)의 구성을 좀 더 자세히 보여주는 블록도이며, 도 4 및 도 5는 각각 모델 구성부(300) 중 쿼리 레이어(310) 및 키 레이어(320)의 일 예를 보여주는 블록도이다.
- [0037] 먼저 도 3을 참조하면, 모델 구성부(300)는 쿼리 레이어(query layer, 310), 키 레이어(key layer, 320) 및 밸류 레이어(value layer, 330)를 포함한다.

- [0038] 쿼리 레이어(310)는 데이터 전처리부(200)로부터 N 개의 지역으로 구성된 입력 데이터(R)를 수신한다. 쿼리 레이어(310)는 N 개 지역으로 구성된 입력 데이터에 대한 학습을 통하여, 각 지역별로 쿼리(Q)를 정의할 수 있다.
- [0039] 본 출원의 일 실시 예에 있어서, 쿼리 레이어(310)는 LSTM을 통하여 구현될 수 있다. 예를 들어, 도 4에 도시된 바와 같이, 쿼리 레이어(310)는 복수의 LSTM 셀을 포함하도록 구현될 수 있다.
- [0040] 예를 들어, 소정 지역(Ri)의 T주 동안의 인플루엔자 발생량 시계열 데이터가 입력 데이터로 제공되는 경우, 쿼리 레이어(310)는 해당 지역(Ri)의 인플루엔자 발생량에 대한 시간적 패턴을 학습할 수 있다. 특히, 쿼리 레이어(310)는 LSTM을 사용하여 구현되었기 때문에, 소정 지역(Ri)의 시간적 패턴을 효과적으로 학습할 수 있다. 이후, 쿼리 레이어(310)는 소정 지역(Ri)에 대응하는 쿼리(qi)를 정의할 수 있다.
- [0041] 키 레이어(320)는 데이터 전처리부(200)로부터 N 개의 지역으로 구성된 입력 데이터(R)를 수신한다. 키 레이어(320)는 N 개 지역으로 구성된 입력 데이터에 대한 학습을 통하여, 각 지역별로 키(K)를 정의할 수 있다.
- [0042] 본 출원의 일 실시 예에 있어서, 키 레이어(320)는 LSTM을 통하여 구현될 수 있다. 예를 들어, 도 5에 도시된 바와 같이, 키 레이어(320)는 복수의 LSTM 셀을 포함하도록 구현될 수 있다. 이 경우, 키 레이어(320)는 소정 지역(Ri)의 시간적 패턴을 효과적으로 학습할 수 있으며, 이에 따라 키 레이어(320)는 소정 지역(Ri)에 대응하는 키(ki)를 정의할 수 있다.
- [0043] 밸류 레이어(330)는 데이터 전처리부(200)로부터 N 개의 지역으로 구성된 입력 데이터(R)를 수신한다. 밸류 레이어(330)는 N 개 지역으로 구성된 입력 데이터에 대한 학습을 통하여, 각 지역별로 밸류(V)를 정의할 수 있다. 밸류 레이어(330)는, 예를 들어, FC 레이어(Fully-Connected Layer)를 통하여 구현될 수 있다.
- [0044] 도 6은 도 1의 감염병 예측부(400)를 좀 더 자세히 보여주는 블록도이다.
- [0045] 감염병 예측부(400)는 모델 구성부(300)로부터 쿼리(Q), 키(K) 및 밸류(V)를 수신하며, 수신된 쿼리, 키 및 밸류를 사용하여 지역들 사이의 상관관계를 학습한다. 감염병 예측부(400)는 학습된 상관관계를 통하여 각 지역에 대한 미래의 감염병 발생량을 예측한다.
- [0046] 도 6을 참조하면, 감염병 예측부(400)는 어텐션 레이어(410) 및 예측 레이어(420)를 포함할 수 있다.
- [0047] 어텐션 레이어(410)는 모델 구성부(300)로부터 쿼리(Q), 키(K) 및 밸류(V)를 수신한다. 어텐션 레이어(410)는 수신된 쿼리(Q), 키(K) 및 밸류(V)를 이용하여 소정 지역(Ri)과 다른 지역들 사이의 유사도를 계산하고, 계산된 유사도를 사용하여 해당 지역(Ri)에 대한 주의 집중값(ai)을 출력한다. 어텐션 레이어(410)는 내적 모듈(411), 소프트맥스 모듈(412) 및 곱셈 모듈(413)을 포함한다.
- [0048] 내적 모듈(411)은 모델 구성부(300)로부터 쿼리(Q) 및 키(K)를 수신하며, 내적 연산(dot product)을 통하여 지역들 사이의 유사도를 계산한다. 예를 들어, 소정 지역(Ri)과 다른 지역들 사이의 유사도를 계산하는 경우, 내적 모듈은 다음과 같은 수학적 식 1에 의하여 내적 연산을 수행할 수 있다.
- [0049] [수학적 식 1]
- $$in_{i,j} = q_i \cdot k_j$$
- [0050]
- [0051] 여기서, "in_{i,j}"는 "q_i"와 "k_j"의 내적 연산을 의미하고, "q_i"는 소정 지역(Ri)의 쿼리를 의미하며, "k_j"는 다른 지역(Rj)의 키를 의미한다.
- [0052] 소프트맥스 모듈(412)은 내적 모듈(411)로부터 유사도 정보를 포함하는 내적 연산 결과를 수신한다. 소프트맥스 모듈(412)은 소프트맥스 함수(softmax function)를 사용하여 수신된 유사도를 확률로 변환하며, 이는 다음과 같은 수학적 식 2에 의하여 수행될 수 있다.
- [0053] [수학적 식 2]
- $$r_{i,j} = \frac{e^{in_{i,j}}}{\sum_{n=1}^N e^{in_{i,n}}}$$
- [0054]
- [0055] 여기서, "r_{i,j}"는 지역들 사이의 모든 쌍들(즉, 지역(Ri)과 지역(Rj) 사이의 모든 쌍들)에 대한 자기 집중 가중치 값(attent weight)으로, 확률 유사도라고 칭해질 수도 있다. 소프트맥스 모듈(412)은 소정 지역(Ri)과 다른

지역들 각각에 대한 확률 유사도를 계산하여, N 개의 확률 유사도를 출력할 수 있다.

[0056] 곱셈 모듈(413)은 소프트맥스 모듈(412)로부터 확률 유사도를 수신하고, 모델 구성부(300)로부터 벨류(V)를 수신한다. 곱셈 모듈(413)은 N개의 확률 유사도와 벨류(V)를 곱하여 주의 집중 값(ai)을 연산하며, 이는 다음과 같은 수학적 식 3에 의하여 수행될 수 있다.

[0057] [수학적 식 3]

$$a_i = \sum_{j=1}^N r_{i,j} \times v_j$$

[0058]

[0059] 예측 레이어(420)는 어텐션 레이어(410)로부터 소정 지역(Ri)에 대한 주의 집중 값(ai)을 수신한다. 예측 레이어(420)는 주의 집중 값(ai)에 기초하여, 소정 시간 후(예를 들어, L 주 후)의 소정 지역(Ri)에서의 감염병 발생량을 예측할 수 있다. 예측 레이어(420)는 FC 레이어를 통하여 구현될 수 있으며, 소정 지역(Ri)에 대한 감염병 발생량 예측은 이하의 수학적 식 4에 의하여 수행될 수 있다.

[0060] [수학적 식 4]

$$y_{i,t+L} = \text{act}_f(W_f \times a_i + b_f)$$

[0061]

[0062] 여기서, "actf"는 활성화 함수를 의미하며, Wf는 가중치 행렬을 의미하고, bf는 예측 레이어(420)의 편향(bias)을 의미한다.

[0063] 한편, N개의 지역 모두에 대하여, 어텐션 레이어(410)에서 유사도, 확률 유사도 및 주의 집중 값을 계산하는 동작 및 예측 레이어(420)에서 감염병 발생량을 예측하는 동작이 반복될 수 있으며, 결과적으로 모든 N개의 지역에 대한 미래의 감염병 발생량을 예측할 수 있다.

[0064] 한편, L 주 후의 실제 감염병 발생량과 예측된 감염병 발생량 사이의 손실 함수는 이하의 수학적 식 5에 의하여 정의될 수 있다.

[0065] [수학적 식 5]

$$L = \sum_{i=1}^N (x_{i,t+L} - y_{i,t+L})^2 / N$$

[0066]

[0067] 도 7은 모델 구성부(300) 및 감염병 예측부(400)에 의하여 소정 지역(Ri)에서의 주의 집중 값(ai)이 연산되는 과정을 보여주는 도면이다.

[0068] 도 7을 참조하면, 모델 구성부(300)는 N개의 지역 각각에 대한 쿼리, 키 및 벨류를 정의할 수 있다. 소정 지역(Ri)의 쿼리(qi)는 해당 지역의 쿼리와 유사한 키를 갖는 다른 지역을 탐색하기 위한 기준의 역할을 한다. 감염병 예측부(400)는 쿼리와 키의 유사도를 구하고, 이를 확률 유사도로 변환하며, 확률 유사도를 벨류에 곱하여 최종 예측에 사용되는 주의 집중 값(ai)을 도출할 수 있다.

[0069] 도 8은 도 1의 유사도 시각화부(500)에 의하여 디스플레이되는 유사도의 일 예를 보여주는 도면이다.

[0070] 유사도 시각화부(500)는 감염병 예측부(400)로부터 NxN 형태를 가지는 유사도 행렬을 수신한다. 예를 들어, 감염병 예측부(400)는 소정 지역(Ri)과 다른 지역들 사이의 유사도를 1xN의 행렬로 출력할 수 있다. 감염병 예측부(400)는 이를 모든 지역에 대하여 반복함으로써, 결과적으로 NxN 형태를 갖는 유사도 행렬을 유사도 시각화부(500)에 송신할 수 있다.

[0071] 유사도 시각화부(500)는 NxN 형태를 갖는 행렬을 시각화할 수 있다. 이때, 입력 데이터로 사용되는 T주 동안의 기간에서, 모든 시간의 행렬이 추출될 수 있으며, 이에 따라 유사도 시각화부(500)는 시간에 따라 변하는 지역 사이의 유사도를 디스플레이할 수 있다.

[0072] 예를 들어, 유사도 시각화부(500)는, 도 8에 도시된 바와 같이, Heat map을 사용하여 디스플레이할 수 있다. 도 8의 (a), (b), (c)는 각각 T 주 동안의 기간에서 첫 번째 주, 열 번째 주, 마지막 주를 의미한다. 또한 X축은 쿼리 지역, Y축은 키 지역을 나타내고, 빨간색으로 갈수록 높은 유사도를 의미한다.

- [0073] 도 9는 도 1의 지역별 감염병 예측 장치(10)의 성능의 일 예를 보여주는 실험 결과이다.
- [0074] 도 9에서는, 예시적으로, RMSE(root mean square error)와 PCC(Pearson correlation coefficient)를 이용하여, 도 1의 지역별 감염병 예측 장치(10)의 성능을 평가하였다. 실험 결과에서, 가장 우측의 SAIFlu_Net이 도 1의 지역별 감염병 예측 장치(10)의 성능 평가 결과이며, 미국 CDC로부터 수집된 인플루엔자 발생량 데이터를 이용하여 성능을 평가하였다.
- [0075] 도 9에 도시된 바와 같이, 도 1의 지역별 감염병 예측 장치(10)가 가장 높은 정확도를 갖는 것을 확인할 수 있다.
- [0076] 도 10은 본 출원의 실시 예에 따른 감염병 예측 방법을 보여주는 순서도이다.
- [0077] S110 단계에서, 지역별 감염병 발생량 데이터가 수집된다. 예를 들어, 데이터 수집부(100, 도 1 참조)는 N개 지역에 대한 인플루엔자 발생량에 대한 시계열 데이터를 수집할 수 있다.
- [0078] S120 단계에서, 지역별 감염병 발생량 데이터에 대한 전처리 동작이 수행된다. 예를 들어, 데이터 전처리부(200, 도 1 참조)는 누락된 데이터를 '0'으로 대체할 수 있다. 또한, 데이터 전처리부(200)는 지역별로 최소값은 '0', 최대값은 '1'로 정규화하는 최소-최대 정규화(min max normalizaion)을 수행할 수 있다.
- [0079] S130 단계에서, 각 지역에 대한 쿼리, 키, 밸류가 정의될 수 있다. 예를 들어, 모델 구성부(300)는 LSTM을 통하여 각 지역에 대한 쿼리 및 키를 정의하고, FC 레이어를 통하여 각 지역에 대한 밸류를 정의할 수 있다.
- [0080] S140 단계에서, 각 지역에 대한 쿼리, 키, 밸류를 사용하여, 지역들 사이의 상관관계를 학습하고 미래의 감염병 발생량을 예측할 수 있다. 예를 들어, 감염병 예측부(400)는 각 지역에 대한 쿼리, 키 및 밸류를 사용하여 유사도 및 주의 집중 값을 연산하고, 이를 통하여 각 지역의 감염병 발생량을 예측할 수 있다.
- [0081] 한편, 상술한 설명은 예시적인 것이며, 본 출원의 기술적 사상은 이에 한정되지 않음이 이해될 것이다. 예를 들어, 도 1 내지 도 10에서는, 예시적으로 미국에서의 인플루엔자 발생량 데이터를 이용하여 성능을 평가하였다. 다만, 이는 예시적인 것이며, 본 출원의 감염병 발생량 예측 장치는 다양한 지역의 다양한 감염병에 대하여도 적용이 가능함이 이해될 것이다. 예를 들어, 일본, 한국 등의 국가에 대하여 COVID-19의 발생량 예측에도 본 출원의 감염병 발생량 예측 장치가 적용될 수도 있다.
- [0082] 이상에서는 본 발명에 따른 바람직한 실시 예들에 대하여 도시하고 또한 설명하였다. 그러나 본 발명은 상술한 실시 예에 한정되지 아니하며, 특허 청구의 범위에서 첨부하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술 분야에서 통상의 지식을 가진 자라면 누구든지 다양한 변형 실시가 가능할 것이다.

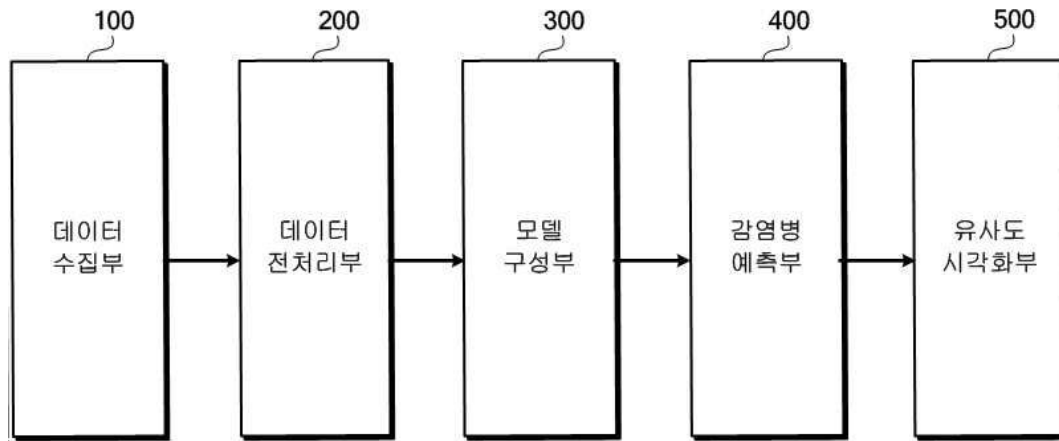
부호의 설명

- [0083] 100: 데이터 수집부
200: 데이터 전처리부
300: 모델 구성부
400: 감염병 예측부
500: 유사도 시각화부

도면

도면1

10



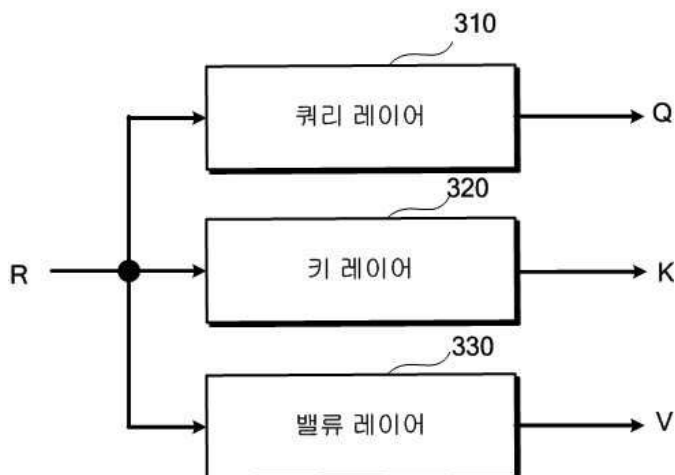
도면2

Time ↓

	A	B	C	D	E	F	G	H	I	J
1	지역 0	지역 1	지역 2	지역 3	지역 4	지역 5	지역 6	지역 7	지역 8	지역 9
2	179	192	184	319	753	15	71	183	631	41
3	219	249	142	500	655	39	120	263	658	30
4	220	347	267	795	752	146	122	405	638	51
5	297	747	324	881	962	130	269	416	722	49
6	519	692	380	916	1059	249	325	537	846	34
7	509	400	353	956	1327	183	400	517	906	73
8	552	311	317	946	1620	229	469	334	1031	115
9	405	290	287	1029	1698	213	356	282	1144	56
10	304	237	351	980	1544	206	296	246	899	41
11	157	222	251	489	1073	117	262	171	839	50
12	126	298	252	359	801	114	157	164	660	46

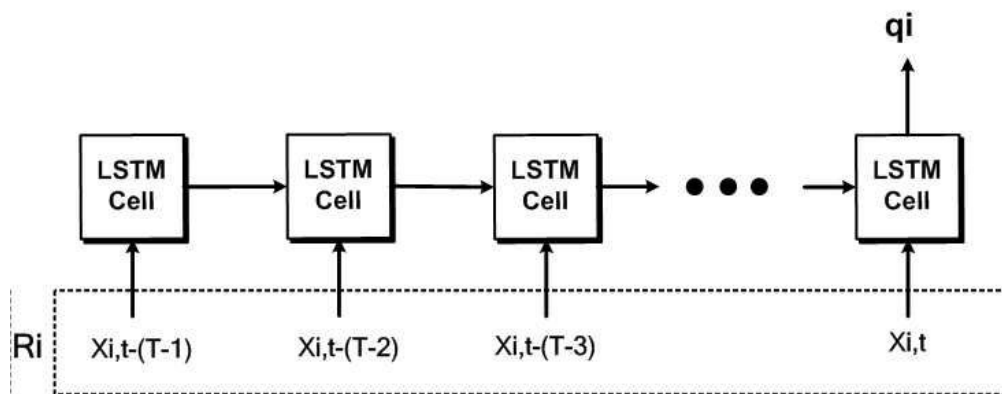
도면3

300



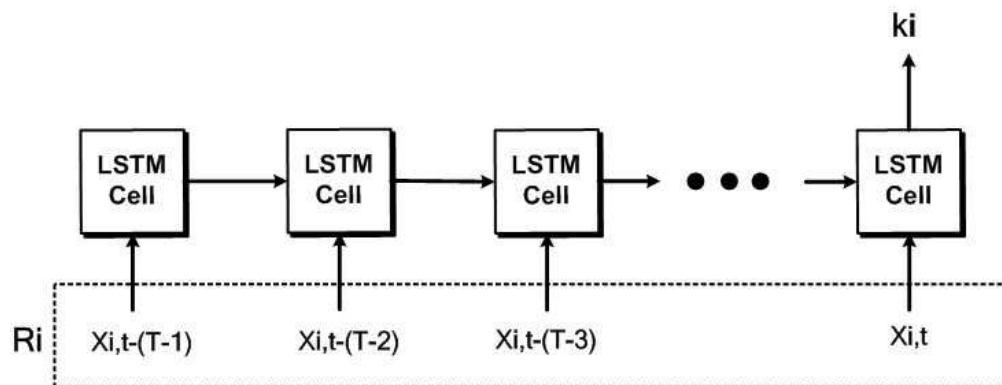
도면4

310

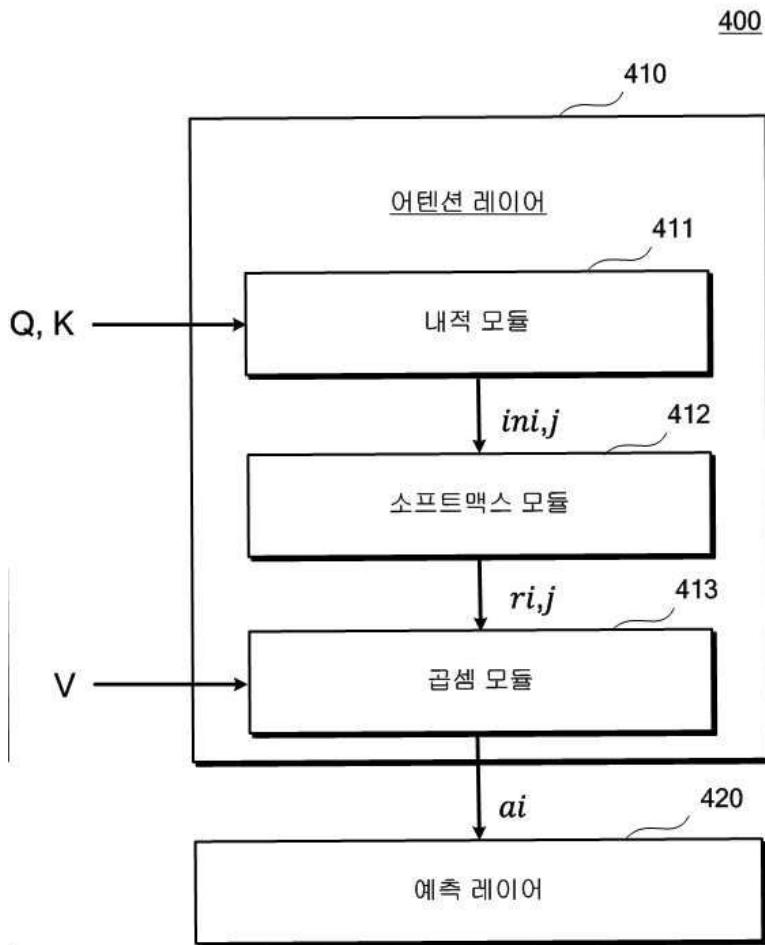


도면5

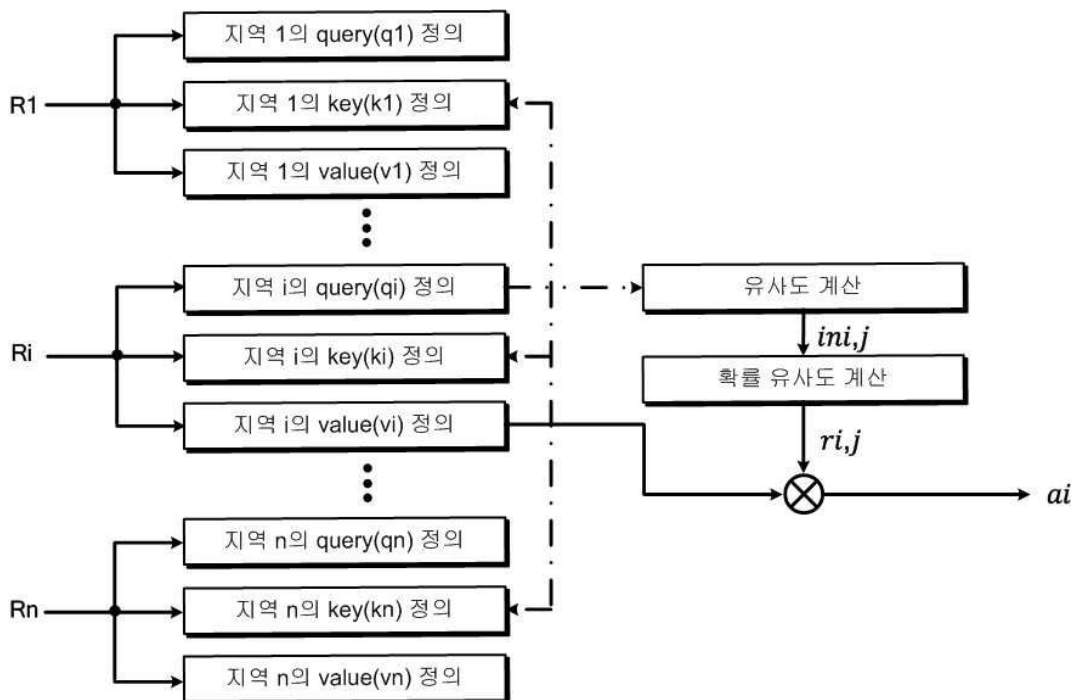
320



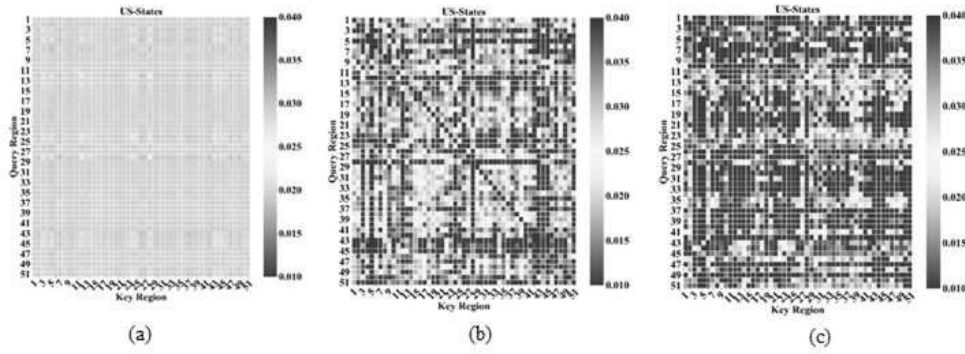
도면6



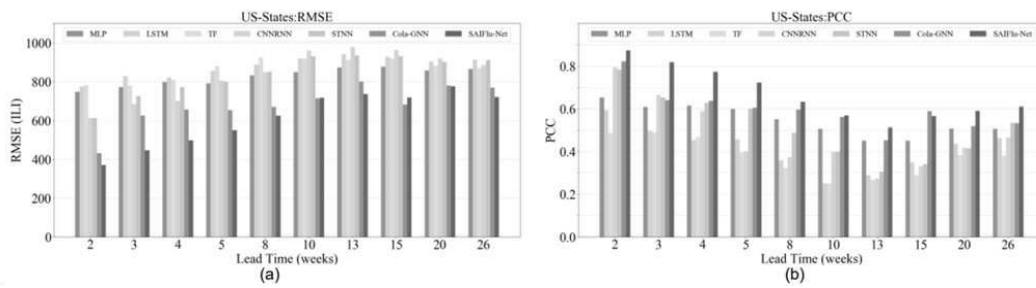
도면7



도면8



도면9



도면10

