



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0032307
(43) 공개일자 2022년03월15일

(51) 국제특허분류(Int. Cl.)

G16H 50/80 (2018.01) G16H 50/20 (2018.01)
G16H 50/50 (2018.01) G16H 50/70 (2018.01)
G16H 70/00 (2018.01)

(52) CPC특허분류

G16H 50/80 (2018.01)
G16H 50/20 (2018.01)

(21) 출원번호 10-2020-0113933

(22) 출원일자 2020년09월07일

심사청구일자 2020년09월07일

(71) 출원인

고려대학교 산학협력단

서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)

(72) 발명자

황인준

서울특별시 성동구 매봉길 15, 108동 1805호(옥수동, 래미안 옥수 리버젠)

정승원

서울특별시 노원구 우이천로 102, 204동 1504호(월계동, 초안아파트)

문재욱

서울특별시 광진구 아차산로78길 7, 유진스웰동 1204호(광장동, 유진스웰)

(74) 대리인

송인호, 윤형근, 최관락

전체 청구항 수 : 총 5 항

(54) 발명의 명칭 감염병 발생 패턴 분석 장치

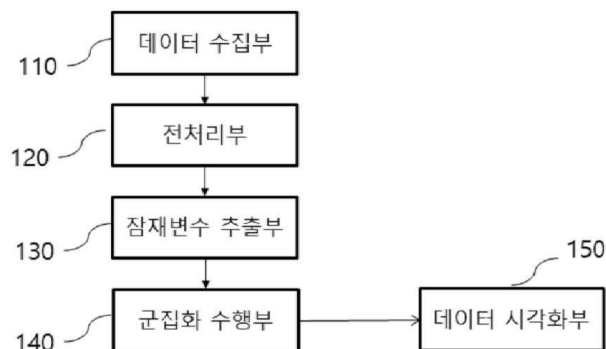
(57) 요약

본 발명은 군집화 기반 분석을 통해 감염병 발생 패턴을 분석할 수 있는 감염병 발생 패턴 분석 장치에 관한 것이다.

본 발명의 일실시예에 의한 감염병, 지역 및 특정 시간 별로 보고된 환자 수에 대한 데이터를 수집하는 데이터 수집부; 상기 수집된 데이터를 시간축에 따라 정규화를 수행하고, 상기 정규화된 데이터를 이용하여 상기 복수 차원의 텐서를 구성하는 전처리부; 상기 텐서를 상기 복수 차원 각각에 대응되게 분해하여 복수 개의 잠재변수 행렬을 추출하는 잠재변수 추출부; 상기 복수 개의 잠재변수 행렬 각각에 대해 K평균 군집화를 이용하여 군집화를 수행하는 군집화 수행부; 및 각 잠재변수 행렬별로 얻어진 군집화 결과를 이용하여 시각화 데이터를 디스플레이 하는 데이터 시각화부를 포함할 수 있다.

대표도 - 도1

100



(52) CPC특허분류

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

G16H 70/00 (2021.08)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465031070
과제번호	HG19C0682020020
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	2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

감염병, 지역 및 특정 시간 별로 보고된 환자 수에 대한 데이터를 수집하는 데이터 수집부;

상기 수집된 데이터를 시간축에 따라 정규화를 수행하고, 상기 정규화된 데이터를 이용하여 상기 복수 차원의 텐서를 구성하는 전처리부;

상기 텐서를 상기 복수 차원 각각에 대응되게 분해하여 복수 개의 잠재변수 행렬을 추출하는 잠재변수 추출부;

상기 복수 개의 잠재변수 행렬 각각에 대해 K평균 군집화를 이용하여 군집화를 수행하는 군집화 수행부; 및

각 잠재변수 행렬별로 얻어진 군집화 결과를 이용하여 시각화 데이터를 디스플레이 하는 데이터 시각화부를 포함하는 것을 특징으로 하는 감염병 발생 패턴 분석 장치.

청구항 2

제3항에 있어서,

상기 복수 차원의 텐서는 3차원 이상의 텐서이고,

상기 복수 차원은 감염병, 지역 및 시간 차원을 포함하는 것을 특징으로 하는 감염병 발생 패턴 분석 장치.

청구항 3

제2항에 있어서, 상기 잠재변수 추출부는

터커 분해를 이용하여 상기 텐서를 분해하여 상기 복수 차원에 대응되게 복수 개의 잠재변수 행렬을 추출하는 것을 특징으로 하는 감염병 발생 패턴 분석 장치.

청구항 4

제3항에 있어서, 상기 군집화 수행부는

엘로우 기법을 이용하여 각 차원마다 최적의 군집 개수를 결정하는 것을 특징으로 하는 감염병 발생 패턴 분석 장치.

청구항 5

제3항에 있어서, 상기 시각화부는

상기 전처리부에서 구성된 텐서를 더 이용하여 상기 시각화 데이터를 디스플레이 하는 것을 특징으로 하는 감염병 발생 패턴 분석 장치.

발명의 설명

기술 분야

[0001] 본 발명은 감염병 발생 패턴 분석 장치에 관한 것으로, 보다 상세하게는 군집화 기반 분석을 통해 감염병 발생 패턴을 분석할 수 있는 감염병 발생 패턴 분석 장치에 관한 것이다.

배경 기술

[0002] 일반적으로 전염병 또는 감염병은 질병 중 전염 또는 감염의 가능성이 있는 질병들을 말하는데, 이러한 감염병은 특정 병원체나 병원체의 독성물질로 인하여 발생하는 질병으로 감염된 사람으로부터 감수성이 있는 숙주(사람)에게 감염되는 질환을 의미한다.

[0003] 그리고 그러한 전염병의 전파 방법은 사람간 접촉, 식품이나 식수, 곤충 매개, 동물에서 사람으로 전파 등에 의

한다.

[0004] 예컨대, 21세기 최초의 신종인플루엔자와 같은 전염병은 병원체인 인플루엔자 바이러스가 공기 중으로 퍼져나가 호흡과 함께 인체에 침투한다. 이러한 인플루엔자는 치사율이 높으면서 전염력도 높은 악성변종의 출현가능성이 있으며, 변이가 심하기 때문에 정확한 예측이 어렵다.

[0005] 또한, 최근 전세계적으로 신종코로나바이러스(COVID-19)가 대유행중인 상황에서 이례없는 확진자수를 기록하고 있는 추세이다. 특히, 한국과 중국에서는 2020년 1~2월 기점으로 확진자가 폭발적으로 증가하고 있으며, 국가에서는 이러한 추가 확산 방지를 위한 확산 방지 대책을 적극 도입하고 있는 실정이다.

[0006] 감염병 발생에 효과적으로 대응하기 위해 감염병의 발생 패턴을 파악하고 미래에 발생할 감염병을 예측하여 미리 예방해야 한다. 하지만, 현재 감염병 발생에 대한 연구 결과는 부족하다. 우선 여러 상황에 따라 환경, 문화 또는 기후와 같은 감염병 발병과 관련이 있는 것으로 알려진 다양한 요인들의 영향력이 변화하기 때문에 각 요인들의 영향력을 정량적으로 평가하기 어렵다. 더욱이, 각 요인 사이에서도 서로 영향을 주고받기 때문에 감염병 발생에 중요한 요인을 찾아내는 것도 매우 어렵다.

[0007] 따라서 감염병 발생에 중요한 요인을 효과적으로 찾아낼 수 있도록 하는 기술에 대한 연구가 필요한 실정이다.

선행기술문헌

특허문헌

[0008] (특허문헌 0001) 대한민국 등록특허 제10-1960504호

(특허문헌 0002) 전염병 확산 예측 모델링 장치 및 방법

발명의 내용

해결하려는 과제

[0009] 본 발명의 목적은 감염병 발생에 중요한 요인을 찾아낼 수 있도록 군집화 방법을 기반으로 하여 유사한 발생 패턴을 보이는 시간대, 지역, 감염병 등 요소들을 그룹화하고 분석할 수 있는 감염병 발생 패턴 분석 장치를 제공하는 데 있다.

과제의 해결 수단

[0010] 상기 목적을 달성하기 위해 본 발명의 일실시예에 의하면, 감염병, 지역 및 특정 시간 별로 보고된 환자 수에 대한 데이터를 수집하는 데이터 수집부; 상기 수집된 데이터를 시간축에 따라 정규화를 수행하고, 상기 정규화된 데이터를 이용하여 상기 복수 차원의 텐서를 구성하는 전처리부; 상기 텐서를 상기 복수 차원 각각에 대응되게 분해하여 복수 개의 잠재변수 행렬을 추출하는 잠재변수 추출부; 상기 복수 개의 잠재변수 행렬 각각에 대해 K평균 군집화를 이용하여 군집화를 수행하는 군집화 수행부; 및 각 잠재변수 행렬별로 얻어진 군집화 결과를 이용하여 시각화 데이터를 디스플레이 하는 데이터 시각화부를 포함하는 감염병 발생 패턴 분석 장치가 개시된다.

발명의 효과

[0011] 본 발명의 일실시예에 의한 감염병 발생 패턴 분석 장치는 텐서 분해와 K평균 군집화를 이용하여 시간, 지역, 감염병 측면에서 유사한 감염병 발생 패턴을 가진 요소를 정확하게 분석할 수 있다.

[0012] 또한, 본 발명의 일실시예에 의하면, 도출된 군집들을 분석하여 군집 간의 공통점과 차이점을 탐색함으로써, 감염병 발생 패턴의 영향을 주는 요인을 탐색할 수 있다.

[0013] 또한, 본 발명의 일실시예에 의하면, 시간, 지역, 감염병 측면뿐만 아니라 성별, 연령대 등 다양한 정보가 추가되어도 동일한 프로세스로 분석 가능하다.

[0014] 또한, 본 발명의 일실시예에 의하면, 분석에 용이하도록 도출된 군집들을 시각화하여 사용자가 쉽게 결과를 파악 가능하다.

도면의 간단한 설명

- [0015] 도 1은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치의 블록도이다.
- 도 2는 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 데이터를 수집하는 과정을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 수집된 데이터를 전처리하는 과정을 설명하기 위한 도면이다.
- 도 4는 도 3의 정규화(Normalization) 과정을 설명하기 위한 도면이다.
- 도 5는 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 텐서(Tensor)를 분해(Decomposition)하는 과정을 설명하기 위한 도면이다.
- 도 6은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 군집화를 수행하는 방법을 설명하기 위한 도면이다.
- 도 7 내지 도 9는 군집화 수행의 결과를 이용하여 생성된 시각화 데이터의 일례를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0016] 이하, 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에 대해 도면을 참조하여 설명하도록 하겠다.
- [0017] 본 명세서에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "구성된다" 또는 "포함한다" 등의 용어는 명세서상에 기재된 여러 구성 요소들, 또는 여러 단계들을 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0018] 도 1은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치의 블록도이다.
- [0019] 도시된 바와 같이, 감염병 발생 패턴 분석 장치(100)는 데이터 수집부(110), 전처리부(120), 잠재변수 추출부(130), 군집화 수행부(140) 및 데이터 시각화부(150)를 포함할 수 있다.
- [0020] 상기 데이터 수집부(110)는 감염병, 지역 및 시간(예: 날짜) 별로 보고된 환자 수에 대한 데이터를 수집할 수 있다. 필수 수집 정보는 감염병, 지역, 시간, 환자 수를 포함할 수 있다. 상기 데이터 수집부(110)는 성별, 연령대 등 추가 정보를 수집 가능할 경우 이 역시 수집할 수 있다.
- [0021] 본 명세서에서 환자 수를 제외하고 감염병, 지역, 시간, 추가 정보는 차원이라 부른다.
- [0022] 상기 전처리부(120)는 상기 데이터 수집부(110)에 의해 수집된 데이터를 시간축에 따라 정규화(Normalization)를 수행하고, 정규화된 데이터를 다차원(또는 '복수 차원'이라 부를 수 있음) 텐서(tensor)로 구성할 수 있다.
- [0023] 상기 잠재변수 추출부(130)는 상기 텐서를 상기 복수 차원 각각에 대응되게 분해하여 복수 개의 잠재변수 행렬을 추출할 수 있다.
- [0024] 상기 군집화 수행부(140)는 잠재변수 추출부(130)에서 도출한 잠재변수 행렬들을 기반으로 하여 각 차원마다 군집화를 수행할 수 있다.
- [0025] 상기 데이터 시각화부(150)는 군집화 수행 결과를 이용하여 각 차원마다 시각화 데이터를 생성하여 출력할 수 있다.
- [0026] 도 2는 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 데이터를 수집하는 과정을 설명하기 위한 도면이다.
- [0027] 도시된 바와 같이, 상기 데이터 수집부(110)는 감염병, 지역 및 시간(예: 날짜) 별로 보고된 환자 수에 대한 데이터를 수집할 수 있다. 상기 데이터는 한국질병관리본부의 감염병 포털을 통해 제공될 수 있다.
- [0028] 상기 데이터 수집부(110)는 웹 크롤러(web crawler)를 통해 감염병 데이터를 한국질병관리본부의 감염병 포털에서 자동으로 수집할 수 있다.
- [0029] 도 3은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 수집된 데이터를 전처리하는 과정을 설명하기 위한 도면이다.
- [0030] 먼저, 전처리부(120)는 상기 데이터 수집부(110)에서 수집된 데이터를 시간 축에 따라 정규화(Normalization)를

수행할 수 있다(S210).

- [0031] 도 4는 도 3의 정규화(Normalization) 과정을 설명하기 위한 도면이다.
- [0032] 도시된 바와 같이, 수집된 데이터가 감염병, 지역, 날짜, 환자 수로 구성되었다면 특정 감염병의 지역별 감염자 수의 최대값을 찾고 특정 감염병의 지역별 감염자 수에 대해 정규화한다. 정규화는 몇몇 특정 감염병에 지나치게 의존하는 분석 결과를 막을 수 있다. 도 4에 도시된 발생 환자 수 96, 64, 48은 정규화를 통해 1.0, 0.67, 및 0.5로 변환될 수 있다.
- [0033] 한편, 전처리부(120)는 데이터를 정규화한 후, 정규화된 데이터를 다차원 텐서로 구성할 수 있다(S220). 도 3에 서는 전처리부(120)는 지역, 감염병, 시간을 각 차원으로 가지는 3차원 텐서를 구성한다.
- [0034] 도 5는 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 텐서(Tensor)를 분해(Decomposition)하는 과정을 설명하기 위한 도면이다.
- [0035] 도시된 바와 같이, 상기 잠재변수 추출부(130)는 터커 분해(Tucker Decomposition)를 이용하여 3차원의 데이터 텐서를 3개의 잠재 변수 행렬(Matrix)로 나눌 수 있다. 상기 잠재변수 추출부(130)는 각 차원의 요소마다 잠재 변수를 추출할 수 있다. 잠재변수 추출 결과의 예시로, "Seoul", "Cholera" 등과 같은 요소마다 잠재 변수로 표현된 것을 확인할 수 있다. d_R 는 지역 차원에서의 잠재변수의 수이고, d_T 는 시간 차원에서의 잠재변수의 수이고, d_I 는 감염병 차원에서의 잠재변수의 수이다. d_R , d_T , d_I 의 값은 군집화에 용이함을 위해 실험적으로 결정될 수 있다.
- [0036] 도 6은 본 발명의 일실시예와 관련된 감염병 발생 패턴 분석 장치에서 군집화를 수행하는 방법을 설명하기 위한 도면이다.
- [0037] 도시된 바와 같이, 군집화 수행부(140)는 도출한 잠재변수 행렬들을 기반으로 하여 각 차원마다 K평균 군집화(K-means Clustering)를 사용하여 군집화를 수행할 수 있다. 이 경우, 군집화를 수행하기에 앞서 최적의 군집 개수를 정해주어야 한다. 상기 최적의 군집 개수는 엘보우(Elbow) 기법을 사용하여, 결정될 수 있다. 상기 엘보우 기법은 군집 개수를 군집 개수를 1부터 증가시켜가며, 오차 제곱합(Sum of Squared Error)을 구하여 오차 제곱합 값의 변화량이 급격히 감소하는 군집 개수를 최적의 군집 개수라고 가정하는 기법을 의미한다.
- [0038] 데이터 시각화부(150)는 군집화 수행부(140)에서 얻은 군집화 결과 및 데이터 텐서를 이용하여 각 차원마다 시각화 데이터를 생성하여 디스플레이 할 수 있다.
- [0039] 도 7 내지 도 9는 군집화 수행의 결과를 이용하여 생성된 시각화 데이터의 일례를 나타낸다.
- [0040] 도 7은 감염병 차원의 군집화 결과 시각화 예시이다. 도시된 군집은 다음과 같이 설명될 수 있다.
- [0041] 군집 1: 계절에 관계없이 대부분의 지역에서 환자가 계속 보고되는 감염병
- [0042] 군집 2: 주기가 약 6개월인 감염병으로 대부분의 지역에서 비슷한 경향을 보임.
- [0043] 군집 3, 4: 한국에서 발병하지 않거나, 거의 발견되지 않는 감염병.
- [0044] 군집 5: 가을 동안 열병으로 분류되는 세가지 감염병.
- [0045] 군집 6: 봄부터 가을까지 자주 보고되는 감염병.
- [0046] 도 8은 시간 차원의 군집화 결과 시각화 예시이다.
- [0047] 시간적 경향을 관찰하기 위해 발생 패턴에 따라 월별 군집화가 수행될 수 있다. 도시된 예에서는 46개월동안 4개의 군집이 여러 색상으로 표현된다. 군집 1, 2, 3을 보면 감염병들의 계절성을 보여준다. 질병 군집 1은 시간 군집 3과 관련이, 질병 군집 4, 6은 시간 군집 1, 2의 결정 요인이 될 수 있다. 또한, 시간 군집 4는 2019년의 감염병 발생 패턴이 크게 변화했음을 나타내는데, 주요 질병과 그 발생을 조사하여 설명이 가능하다.
- [0048] 도 9는 지역 차원의 군집화 결과 시각화 예시이다.
- [0049] 도시된 예시는 17개의 영역에서 6개의 군집이 생성된 결과를 보여준다. 일반적으로 지리적으로 인접한 지역은 동일한 군집으로 분류될 가능성 높다. 도시된 예시를 통해 잠재 변수와 군집화에 관한 연관성도 분석 가능하다.
- [0050] 전술한 바와 같이, 본 발명의 일실시예에 의한 감염병 발생 패턴 분석 장치는 텐서 분해와 K평균 군집화를 이용하여 시간, 지역, 감염병 측면에서 유사한 감염병 발생 패턴을 가진 요소를 정확하게 분석할 수 있다.

- [0051] 또한, 본 발명의 일실시예에 의하면, 도출된 군집들을 분석하여 군집 간의 공통점과 차이점을 탐색함으로써, 감염병 발생 패턴의 영향을 주는 요인을 탐색할 수 있다.
- [0052] 또한, 본 발명의 일실시예에 의하면, 시간, 지역, 감염병 측면뿐만 아니라 성별, 연령대 등 다양한 정보가 추가되어도 동일한 프로세스로 분석 가능하다.
- [0053] 또한, 본 발명의 일실시예에 의하면, 분석에 용이하도록 도출된 군집들을 시각화하여 사용자가 쉽게 결과를 파악 가능하다.
- [0054] 상기와 같이 설명된 감염병 발생 패턴 분석 장치는 상기 설명된 실시예들의 구성과 방법이 한정되게 적용될 수 있는 것이 아니라, 상기 실시예들은 다양한 변형이 이루어질 수 있도록 각 실시예들의 전부 또는 일부가 선택적으로 조합되어 구성될 수도 있다.

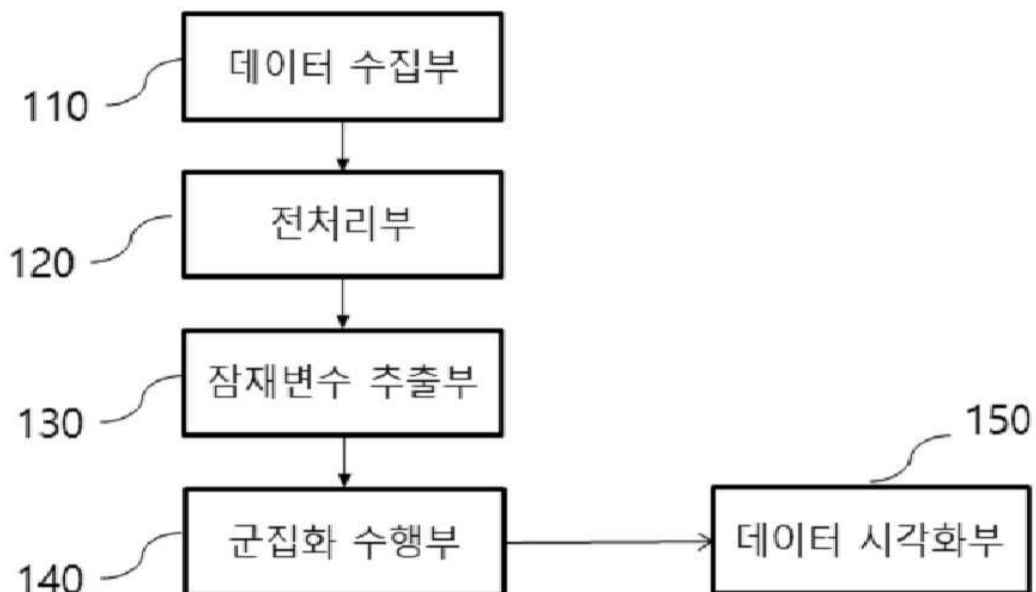
부호의 설명

- [0055] 100: 감염병 발생 패턴 분석 장치
110: 데이터 수집부
120: 전처리부
130: 잠재변수 추출부
140: 군집화 수행부
150: 데이터 시각화부

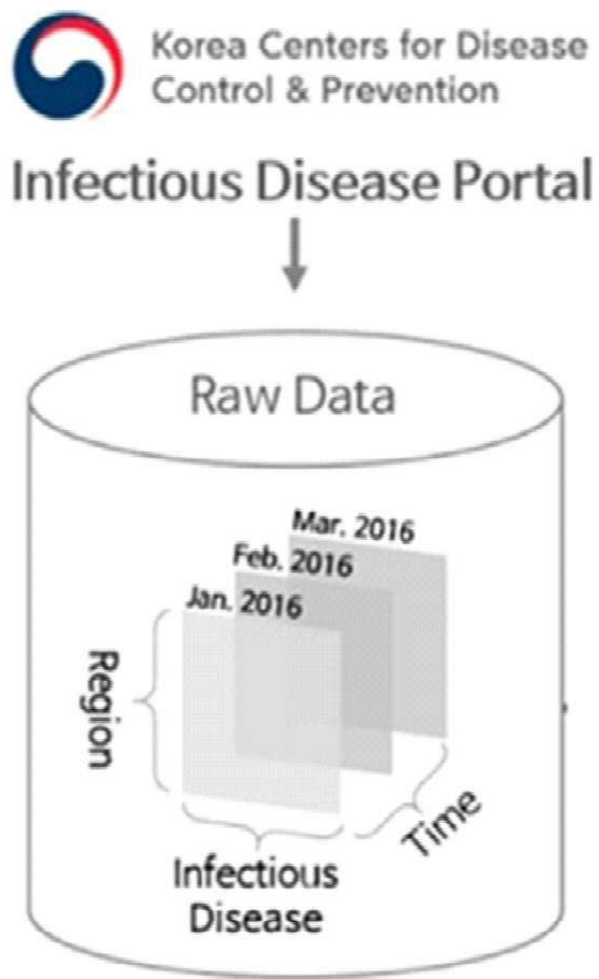
도면

도면1

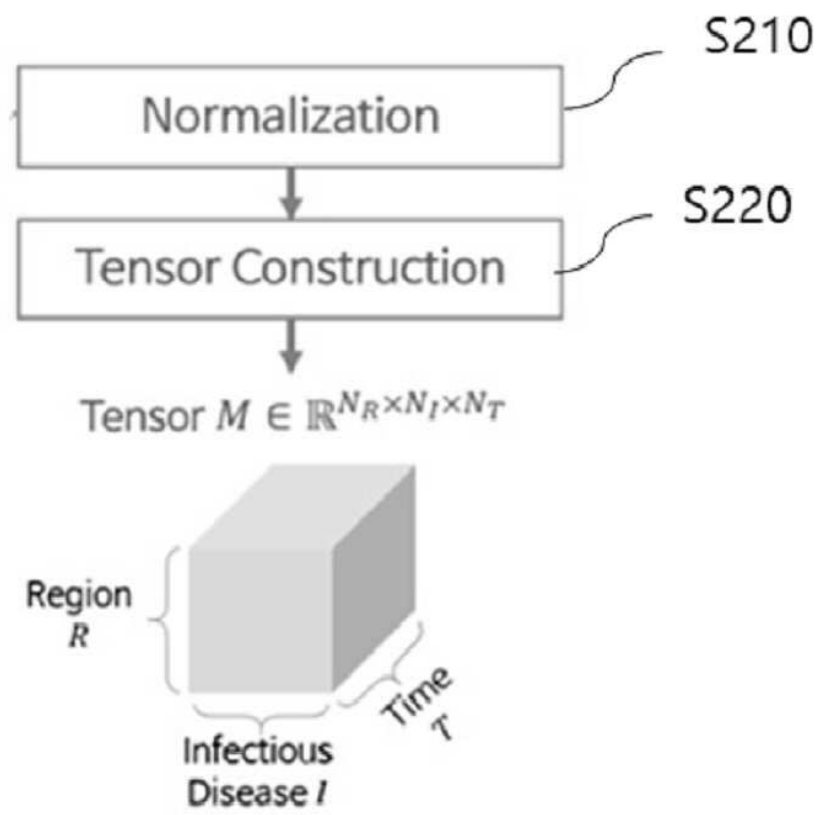
100



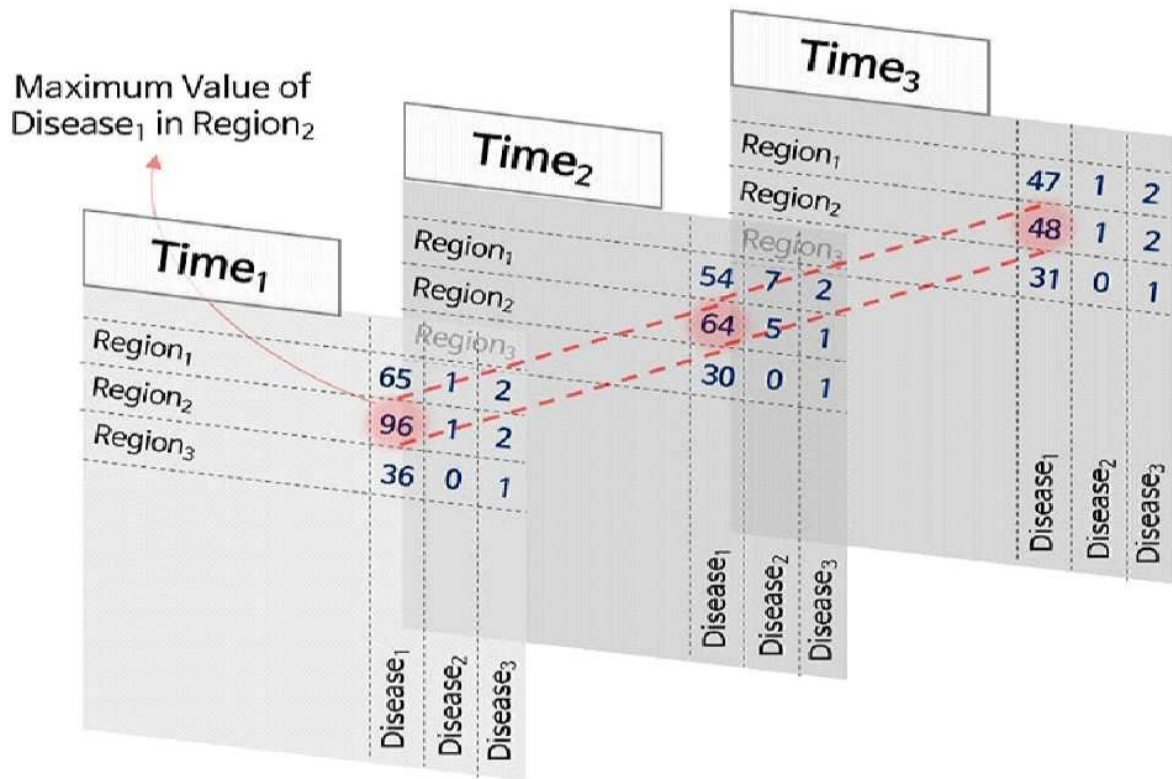
도면2



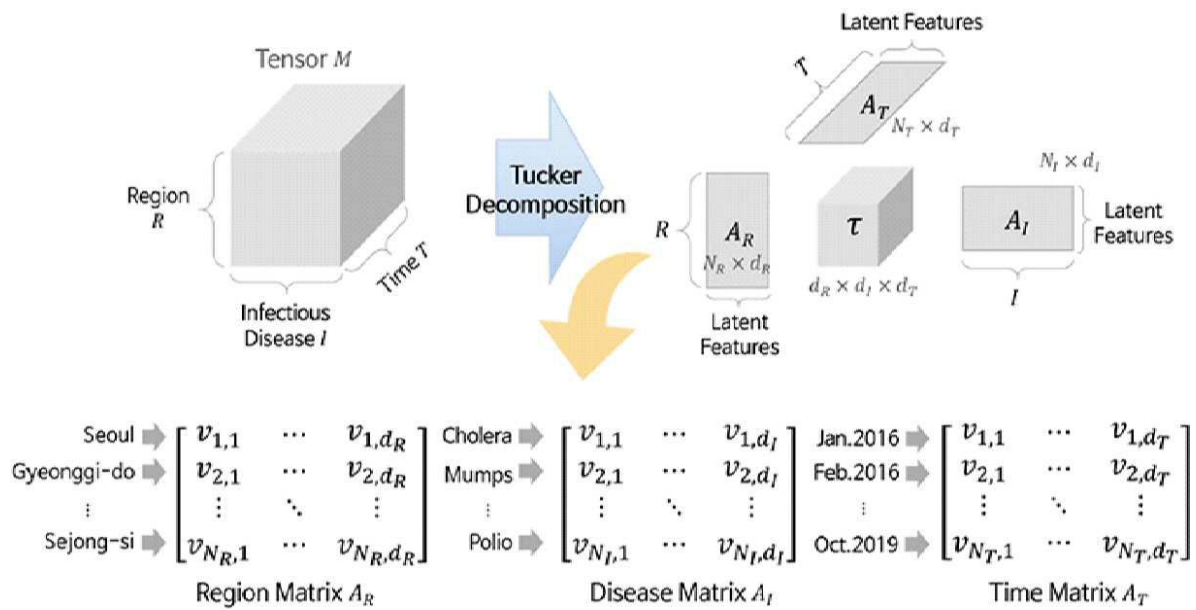
도면3



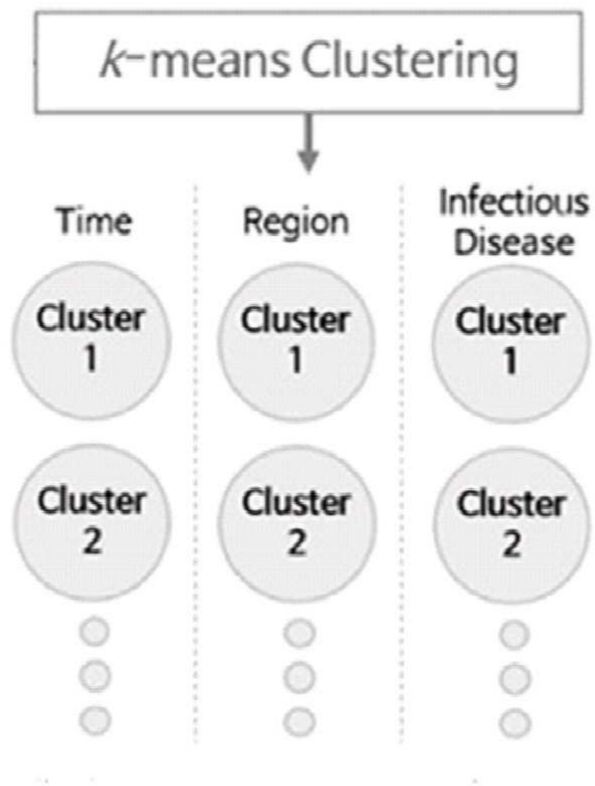
도면4



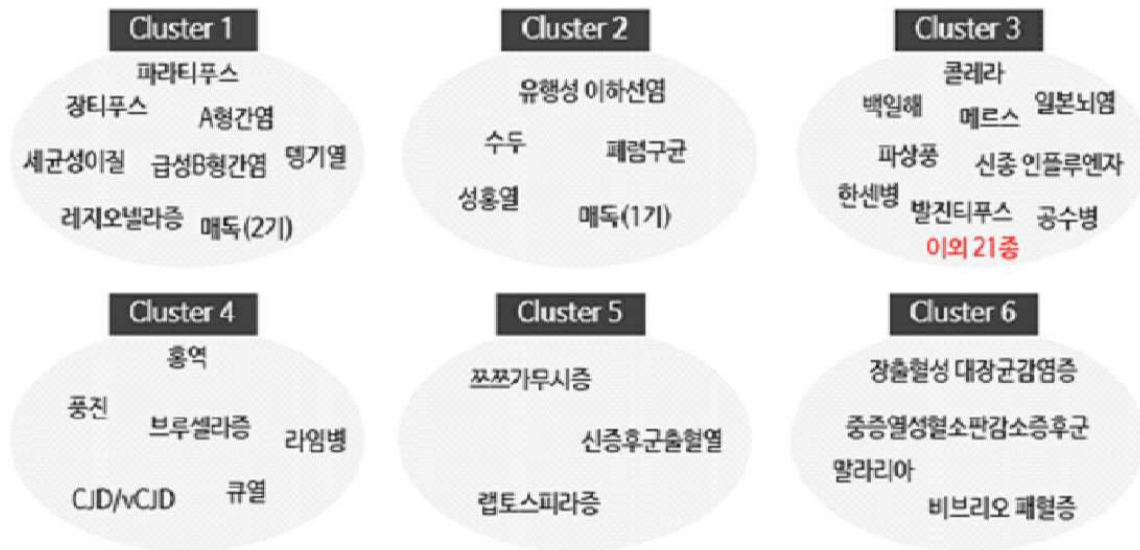
도면5



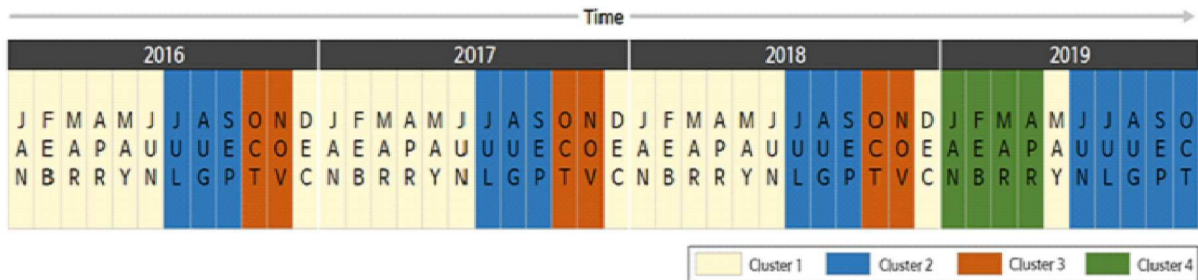
도면6



도면7



도면8



도면9

