



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0127669
(43) 공개일자 2024년08월23일

(51) 국제특허분류(Int. Cl.)

G06N 20/10 (2019.01) G06N 3/12 (2023.01)

G06N 5/01 (2023.01) G06T 17/05 (2011.01)

(52) CPC특허분류

G06N 20/10 (2021.08)

G06N 3/126 (2023.01)

(21) 출원번호 10-2023-0020693

(22) 출원일자 2023년02월16일

심사청구일자 2023년02월16일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

아불가셈

서울특별시 광진구 능동로 209, 대양 AI 센터 433호(군자동, 세종대학교)

최수미

서울특별시 광진구 능동로 209, 대양 AI 센터 720호(군자동, 세종대학교)

바히드 라자비 테르메

서울특별시 광진구 능동로 209, 대양 AI 센터 1001호(군자동, 세종대학교)

(74) 대리인

양정보

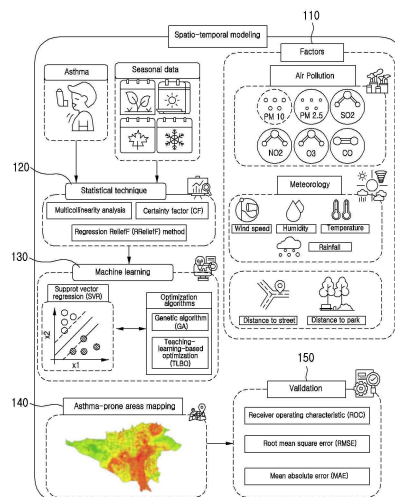
전체 청구항 수 : 총 12 항

(54) 발명의 명칭 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법 및 시스템

(57) 요약

메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법 및 시스템이 제시된다. 본 발명에서 제안하는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법은 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계, 입력 데이터 모델링에 대해 통계 분석을 수행하는 단계, GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계, GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하는 단계 및 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계를 포함한다.

대표도 - 도1



(52) CPC특허분류

G06N 5/01 (2023.01)

G06T 17/05 (2013.01)

G06T 2207/20081 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711174142
과제번호	00156354
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	대학ICT연구센터지원사업
연구과제명	실-가상 연계 메타버스를 위한 초실감 XR 기술 연구
기 여 율	1/2
과제수행기관명	세종대학교산학협력단
연구기간	2022.07.01 ~ 2022.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1415183506
과제번호	P0016038
부처명	산업통상자원부
과제관리(전문)기관명	한국산업기술진흥원
연구사업명	산업기술국제협력(R&D)
연구과제명	가상·증강현실을 위한 지능형 콘텐츠 제작도구 개발
기 여 율	1/2
과제수행기관명	세종대학교 산학협력단
연구기간	2022.12.01 ~ 2022.12.31

명세서

청구범위

청구항 1

시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계;

입력 데이터 모델링에 대해 통계 분석을 수행하는 단계;

GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계;

GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하는 단계; 및

상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계를 포함하는 시공간 모델링 방법.

청구항 2

제1항에 있어서,

상기 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계는,

상기 시공간 모델링을 수행하기 위해 위치를 포함하는 종속 데이터와 환경적 요인을 포함하는 독립 데이터를 수집하고, 상기 수집된 종속 데이터 및 독립 데이터는 훈련 데이터 및 테스트 데이터를 위한 두 가지 하위 세트를 생성하고, 테스트 데이터는 모델을 훈련하는 데 사용되지 않고, 평가 및 예측 결과를 실제 값과 비교하는 데 사용되는

시공간 모델링 방법.

청구항 3

제1항에 있어서,

상기 입력 데이터 모델링에 대해 통계 분석을 수행하는 단계는,

독립 변수 간의 상관관계를 결정하는 다중 공선성 분석(multicollinearity analysis), 모델링에서 이기종 데이터를 결합하기 위해 입력 변수의 모호성을 처리하는 CF(Certainty Factor) 및 데이터 세트의 표본 중 랜덤하게 표본을 선택하고, 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 분석하는 RReliefF(Regression Relief)를 분석하는

시공간 모델링 방법.

청구항 4

제1항에 있어서,

상기 GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계는,

상기 GA 알고리즘을 수행하고, 상기 상기 GA 알고리즘은 적합도에 따라 적절한 수의 염색체를 선택하고, 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성하며, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행하고, 상기 염색체들의 유전자를 수정함으로써 새로운 정보를 입력한 후 최신 모집단을 선택하는

시공간 모델링 방법.

청구항 5

제1항에 있어서,

상기 GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계는,

상기 TLBO 알고리즘을 수행하고, 변수의 상한과 하한, 생산 대수, 적합성 함수를 입력 데이터로 하여 변수에 따라 클래스 멤버가 생성되고, 적합도 함수에 따른 최적 솔루션이 교사로 선택되며, 새로운 솔루션 업데이트 여부에 따라 적합도 함수를 개선하는

시공간 모델링 방법.

청구항 6

제1항에 있어서,

상기 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계는,

RMSE(root mean square error), MAE(Mean absolute error), ROC(Receiver operating characteristic) 지수를 종합하여 시공간 모델링 결과를 평가하는

시공간 모델링 방법.

청구항 7

시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 데이터 수집부;

입력 데이터 모델링에 대해 통계 분석을 수행하는 분석부;

GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 학습부;

GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하고, 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 시공간 모델링부

를 포함하는 시공간 모델링 시스템.

청구항 8

제7항에 있어서,

상기 데이터 수집부는,

상기 시공간 모델링을 수행하기 위해 위치를 포함하는 종속 데이터와 환경적 요인을 포함하는 독립 데이터를 수집하고, 상기 수집된 종속 데이터 및 독립 데이터는 훈련 데이터 및 테스트 데이터를 위한 두 가지 하위 세트를 생성하고, 테스트 데이터는 모델을 훈련하는 데 사용하지 않고, 평가 및 예측 결과를 실제 값과 비교하는 데 사용하는

시공간 모델링 시스템.

청구항 9

제7항에 있어서,

상기 분석부는,

독립 변수 간의 상관관계를 결정하는 다중 공선성 분석(multicollinearity analysis), 모델링에서 이기종 데이터를 결합하기 위해 입력 변수의 모호성을 처리하는 CF(Certainty Factor) 및 데이터 세트의 표본 중 랜덤하게 표본을 선택하고, 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 분석하는 RReliefF(Regression Relief)를 분석하는

시공간 모델링 시스템.

청구항 10

제7항에 있어서,

상기 학습부는,

상기 GA 알고리즘을 수행하고, 상기 상기 GA 알고리즘은 적합도에 따라 적절한 수의 염색체를 선택하고, 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성하며, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행하고, 상기 염색체들의 유전자를 수정함으로써 새로운 정보를 입력한 후 최신 모집단을 선택하는

시공간 모델링 시스템.

청구항 11

제7항에 있어서,

상기 학습부는,

상기 TLBO 알고리즘을 수행하고, 변수의 상한과 하한, 생산 대수, 적합성 함수를 입력 데이터로 하여 변수에 따라 클래스 멤버가 생성되고, 적합도 함수에 따른 최적 솔루션이 교사로 선택되며, 새로운 솔루션 업데이트 여부에 따라 적합도 함수를 개선하는

시공간 모델링 시스템.

청구항 12

제7항에 있어서,

상기 시공간 모델링부는,

RMSE(root mean square error), MAE(Mean absolute error), ROC(Receiver operating characteristic) 지수를 종합하여 시공간 모델링 결과를 평가하는

시공간 모델링 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법 및 시스템에 관한 것이다.

배경 기술

[0002] 천식은 2016년에 전 세계적으로 3억 3900만 명 이상의 사람들에게 영향을 미치는 가장 흔한 비전염성 질병 중 하나이다. 세계보건기구(World Health Organization; WHO) 통계에 따르면 천식으로 인해 2016년 전 세계적으로 417,918명이 사망했으며, 천식으로 인한 장애 조정 수명(Disability-Adjusted Life-Years; DALY)은 2,480만 명이다(WHO 2018). 이란 어린이와 성인의 천식 발병률은 각각 13%와 10%이다. 테헤란 시에서는 어린이의 34%가 기침, 쌉쌉거림, 숨가쁨과 같은 천식 증상을 앓고 있다. 천식 및 호흡기 질환은 개발도상국과 선진국에서 중요한 건강 문제이다. 천식은 폐의 개폐를 방해하는 알레르기 및 호흡기 감염으로 인해 기도가 좁아지고 기관지염이 발생할 때 발생하는 합병증이다. 천식은 기도의 염증을 특징으로 하는 급성 및 만성 폐 질환의 한 형태이다. 천식이 있는 사람의 기도가 막히면 쌉쌉거림, 기침, 숨가쁨, 가슴 답답함 등의 증상이 나타나 자연스러운 호흡이 불가능해진다. 천식 발작은 천식 증상이 평소보다 더 심해지는 경우를 말한다.

[0003] 천식 증상은 심각도가 다양하며 일부는 사망에 이를 정도로 심각하여 천식을 의학적 응급 상황으로 만들 수 있다.

[0004] 천식을 악화시키는 대부분의 요인은 본질적으로 공간적 및 환경적이며, 이는 서로 다른 장소와 시간에 따라 다르다는 것을 의미한다. 또한 이러한 요인의 필수적인 부분은 인간의 생활환경과 관련이 있기 때문에 알레르기성 천식 발작을 예측하는 것은 그 영향을 줄이는 데 중요한 역할을 할 수 있다. 그 결과 천식이 있는 사람들의 생활환경에서 적절한 정보를 수집함으로써 알레르기성 천식발작에서 다양한 환경요인이 시간과 장소에 따라 어떤 역할을 하는지 예측할 수 있다. 통계 및 컴퓨터 기반 공간 모델링의 사용은 정보의 새로운 측면을 발견하는 데 도움이 된다.

[0005] 세계 여러 지역에서 지리 정보 시스템(Geographic Information System; GIS)은 도시 의료 서비스의 필수적인

부분이 되었다. 질병 데이터를 맵핑하고 시각화하고 해석하는 데 사용된다. 정보 통신 기술은 사회적 상호작용, 라이프스타일, 직장을 변화시킨다. 의료는 가장 유망한 정보 기술 애플리케이션 중 하나이다. 의료는 급성 상태에 대한 대응에서 예방적 접근으로 이동하고 있다. 천식 환자에 대한 빠른 결정이 필요하기 때문에 공간 분석 도구는 건강 관리 및 계획을 포함한 의사 결정을 도울 수 있다. 대부분의 의료 및 계획 선택은 위치를 기반으로 한다. 결과적으로 헬스케어 네트워크는 많은 공간 데이터를 포함한 빅데이터를 다루기 때문에 더 나은 관리를 위한 GIS가 필수적이다.

[0006] 대부분의 연구는 천식의 공간적 분석과 환경 요인과의 관계 분석을 수행했다. 그러나 이전 연구에서는 기계 학습 모델을 사용하여 천식이 발생하기 쉬운 영역의 시공간적 모델링에 대한 관심이 적었다. 천식이 발생하기 쉬운 영역 모델링에는 천식에 영향을 미치는 많은 환경 기준이 있다.

[0007] 따라서 시공간 모델링에서는 빅데이터를 도입한다. 기계 학습은 데이터를 더 잘 해석하고 빅 데이터에서 종속 변수와 독립 변수 간의 연관성을 탐색하여 현상을 예측하는 데 유용한 도구이다. 지금까지 다양한 연구에서 천식의 공간 분석을 위해 기계 학습 모델을 사용했다.

선행기술문헌

비특허문헌

[0008] (비특허문헌 0001) Kazem A, Sharifi E, Hussain FK, Saberi M, Hussain OK. 2013. Support vector regression with chaos-based firefly algorithm for stock market price forecasting. Appl Soft Comput. 13(2):947-958.

(비특허문헌 0002) Longley PA, Goodchild MF, Maguire DJ, Rhind DW. 2005. Geographic information systems and science. Chichester (UK): John Wiley & Sons.

발명의 내용

해결하려는 과제

[0009] 본 발명이 이루고자 하는 기술적 과제는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법 및 시스템을 제공하는데 있다. 본 발명은 SVR(support vector regression) 모델을 이용한 기계 학습을 제안하고, SVR 모델의 최적 매개변수의 값을 결정하기 위해 유전 알고리즘(Genetic Algorithm; GA) 및 TLBO(Teaching-Learning-Based Optimization)를 이용하여 머신 러닝 모델을 이용한 천식 다발 부위의 시공간적 모델링하고, 천식 취약 지역 맵핑하기 위한 방법 및 시스템을 제공한다.

과제의 해결 수단

[0010] 일 측면에 있어서, 본 발명에서 제안하는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법은 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계, 입력 데이터 모델링에 대해 통계 분석을 수행하는 단계, GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계, GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하는 단계 및 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계를 포함한다.

[0011] 상기 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계는 상기 시공간 모델링을 수행하기 위해 위치를 포함하는 종속 데이터와 환경적 요인을 포함하는 독립 데이터를 수집하고, 상기 수집된 종속 데이터 및 독립 데이터는 훈련 데이터 및 테스트 데이터를 위한 두 가지 하위 세트를 생성하고, 테스트 데이터는 모델을 훈련하는 데 사용되지 않고, 평가 및 예측 결과를 실제 값과 비교하는 데 사용된다.

[0012] 상기 입력 데이터 모델링에 대해 통계 분석을 수행하는 단계는 독립 변수 간의 상관관계를 결정하는 다중 공선성 분석(multicollinearity analysis), 모델링에서 이기종 데이터를 결합하기 위해 입력 변수의 모호성을 처리하는 CF(Certainty Factor) 및 데이터 세트의 표본 중 랜덤하게 표본을 선택하고, 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 분석하는 RReliefF(Regression Relief)를 분석한다.

[0013] 상기 GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계는 상기 GA 알고리즘

을 수행하고, 상기 상기 GA 알고리즘은 적합도에 따라 적절한 수의 염색체를 선택하고, 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성하며, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행하고, 상기 염색체들의 유전자를 수정함으로써 새로운 정보를 입력한 후 최신 모집단을 선택한다.

[0014] 상기 GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계는 상기 TLBO 알고리즘을 수행하고, 변수의 상한과 하한, 생산 대수, 적합성 함수를 입력 데이터로 하여 변수에 따라 클래스 멤버가 생성되고, 적합도 함수에 따른 최적 솔루션이 교사로 선택되며, 새로운 솔루션 업데이트 여부에 따라 적합도 함수를 개선한다.

[0015] 상기 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계는 RMSE(root mean square error), MAE(Mean absolute error), ROC(Receiver operating characteristic) 지수를 종합하여 시공간 모델링 결과를 평가한다.

[0016] 또 다른 일 측면에 있어서, 본 발명에서 제안하는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 시스템은 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 데이터 수집부, 입력 데이터 모델링에 대해 통계 분석을 수행하는 분석부, GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 학습부, GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하고, 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 시공간 모델링부를 포함한다.

발명의 효과

[0017] 본 발명의 실시예들에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법 및 시스템을 통해 SVR 모델을 이용한 기계 학습을 제안하고, SVR 모델의 최적 매개변수의 값을 결정하기 위해 유전 알고리즘(Genetic Algorithm; GA) 및 TLBO(Teaching-Learning-Based Optimization)를 이용하여 머신 러닝 모델을 이용한 천식 다발 부위의 시공간적 모델링하고, 천식 취약 지역 맵핑할 수 있다.

도면의 간단한 설명

[0018] 도 1은 본 발명의 일 실시예에 따른 시공간 모델링의 전체 개략도를 나타내는 도면이다.
 도 2는 본 발명의 일 실시예에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법을 설명하기 위한 흐름도이다.
 도 3은 본 발명의 일 실시예에 따른 천식 환자가 있는 지역을 나타내는 지도이다.
 도 4 내지 도 7은 본 발명의 일 실시예에 따른 봄, 여름, 가을, 겨울의 유효 요인 지도의 변화를 나타내는 도면이다.
 도 8은 본 발명의 일 실시예에 따른 유전 알고리즘의 수행 과정을 설명하기 위한 도면이다.
 도 9는 본 발명의 일 실시예에 따른 하이브리드 모델의 수행 과정을 설명하기 위한 도면이다.
 도 10은 본 발명의 일 실시예에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 시스템의 구성을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

[0019] 이하, 본 발명의 실시 예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0021] 도 1은 본 발명의 일 실시예에 따른 시공간 모델링의 전체 개략도를 나타내는 도면이다.

[0022] 본 발명의 실시예에 따른 천식이 발생하기 쉬운 영역의 시공간 모델링은 5단계로 수행될 수 있다. 처음에는 종속 데이터(예를 들어, 천식 환자 위치)와 독립 데이터(예를 들어, 대기오염 매개변수, 기상 매개변수, 도로 및 공원까지의 거리)를 계절별로 수집한다(110). 두 번째 단계에서는 입력 데이터 모델링에 대해 통계 분석(예를 들어, 다중 공선성 분석(multicollinearity analysis), RReliefF(Regression Relief), CF(Certainty Factor) 방법)을 수행한다(120). 세 번째 단계에서 GA 및 TLBO 알고리즘을 사용하여 SVR(support vector regression) 모델의 최적 매개변수를 획득한다(130). 네 번째 단계에서는 GA 알고리즘, TLBO 알고리즘 및 SVR 모델 모델을

통해 획득한 천식이 발생하기 쉬운 지역의 지도를 생성한다(140). 시공간 모델링은 다섯 번째 단계에서 메트릭 지표를 사용하여 평가된다(150).

- [0024] 도 2는 본 발명의 일 실시예에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법을 설명하기 위한 흐름도이다.
- [0025] 제안하는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법은 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집하는 단계(210), 입력 데이터 모델링에 대해 통계 분석을 수행하는 단계(220), GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득하는 단계(230), GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑하는 단계(240) 및 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가하는 단계(250)를 포함한다.
- [0026] 단계(210)에서, 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집한다.
- [0027] 본 발명의 실시예에 따르면, 상기 시공간 모델링을 수행하기 위해 위치를 포함하는 종속 데이터와 환경적 요인을 포함하는 독립 데이터를 수집하고, 상기 수집된 종속 데이터 및 독립 데이터는 훈련 데이터 및 테스트 데이터를 위한 두 가지 하위 세트를 생성하고, 테스트 데이터는 모델을 훈련하는 데 사용되지 않고, 평가 및 예측 결과를 실제 값과 비교하는 데 사용된다.
- [0028] 단계(220)에서, 입력 데이터 모델링에 대해 통계 분석을 수행한다.
- [0029] 본 발명의 실시예에 따르면, 독립 변수 간의 상관관계를 결정하는 다중 공선성 분석(multicollinearity analysis), 모델링에서 이기종 데이터를 결합하기 위해 입력 변수의 모호성을 처리하는 CF(Certainty Factor) 및 데이터 세트의 표본 중 랜덤하게 표본을 선택하고, 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 분석하는 RRelief(Regression Relief)를 분석한다.
- [0030] 단계(230)에서, GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득한다.
- [0031] 본 발명의 실시예에 따르면, 상기 GA 알고리즘을 수행하고, 상기 상기 GA 알고리즘은 적합도에 따라 적절한 수의 염색체를 선택하고, 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성하며, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행하고, 상기 염색체들의 유전자를 수정함으로써 새로운 정보를 입력한 후 최신 모집단을 선택한다.
- [0032] 본 발명의 실시예에 따르면, 상기 TLBO 알고리즘을 수행하고, 변수의 상한과 하한, 생산 대수, 적합성 함수를 입력 데이터로 하여 변수에 따라 클래스 멤버가 생성되고, 적합도 함수에 따른 최적 솔루션이 교사로 선택되며, 새로운 솔루션 업데이트 여부에 따라 적합도 함수를 개선한다.
- [0033] SVR 모델의 최적 매개변수는 GA 알고리즘 및 TLBO 알고리즘을 통해 획득된다.
- [0034] 단계(240)에서, GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑한다.
- [0035] 단계(250)에서, 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가한다.
- [0036] 본 발명의 실시예에 따르면, RMSE(root mean square error), MAE(Mean absolute error), ROC(Receiver operating characteristic) 지수를 종합하여 시공간 모델링 결과를 평가한다.
- [0037] 도 3 내지 도 9를 참조하여, 본 발명의 실시예에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 방법에 관하여 더욱 상세히 설명한다.
- [0039] 도 3은 본 발명의 일 실시예에 따른 천식 환자가 있는 지역을 나타내는 지도이다.
- [0040] 본 발명의 실시예에서는 테헤란에 대한 시공간 모델링을 예시로서 설명한다. 테헤란은 이란에서 가장 인구가 많은 도시로, 약 870만 명(2015년 통계), 면적은 615km^2 이다. 테헤란은 북위 $35^\circ 36'$ 에서 북위 $35^\circ 44'$ 사이에 위치해 있으며 경도 $51^\circ 17'$ 에서 동경 $51^\circ 33'$ 사이에 위치해 있다. 많은 차량들이 테헤란에서 운행되고 있으며, 서풍은 공장과 다른 오염 요소들을 일년 내내 테헤란의 도시로 가져온다. 테헤란은 최근 몇 년간 불리한 환경

조건과 대기 오염이 있었다. 테헤란의 호흡기 질환의 원인으로는 대기오염, 산업센터, 교통 등이 있다.

- [0041] 도 3을 참조하면, 천식이 발생하기 쉬운 영역의 시공간 모델링을 위해 천식 환자의 위치를 종속 데이터로 고려했다. 테헤란 병원 정보 시스템은 2019년 테헤란에서 천식 환자의 위치를 수집하는 데 사용되었다. 훈련과 테스트는 데이터 세트의 두 가지 하위 집합이다. 훈련 데이터 세트는 모델을 맞추는 데 사용되는 첫 번째 하위 집합이다. 두 번째 부분 집합(테스트)은 모델을 훈련하는 데 사용되지 않는다. 적합한 머신 러닝 모델을 평가하고 예측 결과를 실제 값과 비교하는 데 사용된다. 훈련 및 테스트 데이터는 홀드아웃 접근법을 사용하여 분할되었다. 총 872개의 천식 환자 위치가 수집되었으며, 데이터의 70%(611개 위치)는 모델링(훈련)에, 30%(261개 위치)는 평가(테스트)에 사용되었다.
- [0043] 도 4 내지 도 7은 본 발명의 일 실시예에 따른 봄, 여름, 가을, 겨울의 유효 요인 지도의 변화를 나타내는 도면이다.
- [0044] 본 발명에서는 도 4 내지 도 7과 같이 대기오염, 기상, 주차거리, 도로거리 등 천식 발생에 영향을 미치는 환경적 요인을 고려하였다. 도 4 내지 도 7을 참조하여 이러한 각각의 독립적인 기준에 대한 설명이다.
- [0045] 도 4 내지 도 7은 본 발명의 일 실시예에 따른 봄(도 4), 여름(도 5), 가을(도 6), 겨울(도 7)의 다음과 같은 유효 요인들의 지도를 도시한다: a) PM 2.5, b) PM 10, c) SO₂, d) O₃, e) NO₂, f) CO, g) 강수량, h) 풍속, i) 온도, j) 습도, k) 거리까지의 거리 1) 공원까지의 거리.
- [0046] 본 발명에서 고려한 대기오염 매개변수로는 입자물질(Particle Matter) PM_{2.5}, PM₁₀, 이산화질소(NO₂), 이산화황(SO₂), 오존(O₃), 일산화탄소(CO) 등이 있다. 실외 오염물질은 1차 오염물질과 2차 오염물질로 나뉜다. 차량에서 배출되는 CO와 산업 공정에서 배출되는 SO₂는 1차 오염물질의 예이다. 사람의 건강에 해로운 영향을 미치기 때문에, NO₂는 대기 오염에 중요한 역할을 한다. NO₂ 배출원은 차량과 산업용 연료이다. 2차 오염물질(O₃, PM_{2.5}, PM₁₀)은 1차 오염물질과의 화학반응으로 인해 환경에서 생성된다. 태양 광선과 공기의 습도는 이러한 오염 물질을 생성한다. 황산염과 같은 대기 중 휘발성 탄화수소 또는 2차 미세 입자에 대한 햇빛의 영향은 O₃를 생성한다. PM은 화학적 또는 물리적 고체 입자와 공기역학적 직경을 기준으로 허용 또는 초미세로 분류된 액체 입자로 구성된다. 이러한 오염물질은 산소 공급을 감소시키고 천식을 악화시켜 기침과 호흡곤란을 유발한다. 2009년부터 2019년까지 테헤란의 23개 대기오염 측정소에서 계절별 평균 대기오염 데이터가 수집되었다.
- [0047] 본 발명의 실시예에 따른 Kriging 보간 기술은 ArcGIS 10.3 소프트웨어의 대기오염 매개변수를 맵핑한다. 일반적인 Kriging 방법과 구면 변광도를 사용하여 보간하였다. 다른 보간 방법에 비해 Kriging 방법의 장점은 추정 표면에 대한 오차 또는 불확실성의 측정을 제공한다는 것이다.
- [0048] 본 발명에서 고려된 기상학적 변수는 강우, 온도, 습도, 풍속을 포함한다. 기상학적 매개변수는 대기 중 오염물질의 농도를 증가시키거나 감소시키는 또 다른 필수 요소이다. 중요한 기상학적 역할은 대기 중의 대기 오염을 분산, 변환 및 제거하는 것이다. 2009년부터 2019년까지 테헤란 지역의 12개 기상 관측소에 대한 이러한 매개변수의 연평균은 기상 매개변수 지도를 만드는 데 사용되었다. ArcGIS 10.3 소프트웨어의 kriging 보간법은 기상학적 매개변수의 래스터 지도를 준비하는 데 사용되었다.
- [0049] 본 발명의 실시예에 따르면, 녹지는 대기오염으로 인한 납, 먼지, 그을음 등을 흡수하여 대기오염을 감소시킨다. 그러므로, 녹지와 공원과의 근접성은 천식의 심각성을 줄이는데 도움을 줄 수 있다. 도로는 교통 체증으로 인해 대기 오염을 일으킬 수 있다. 따라서 도로와 더 가까운 거리에서는 천식 발생 가능성이 증가한다. 공원의 점층은 ArcGIS 10.3 소프트웨어의 유클리드 거리 도구를 사용하여 래스터 형식으로 변환되었으며 값은 0 ~ 5179.08m이다. 거리의 선형층은 ArcGIS 10.3 소프트웨어에서 유클리드 거리 도구를 사용하여 래스터 지도로 변환된 메인 도로와 사이드 도로를 포함하며, 이 층은 0에서 2342.85m까지 다양하다.
- [0050] 본 발명의 실시예에 따르면, CF(Certainty Factor)는 입력 변수의 모호성을 처리하기 위한 최적의 모델 중 하나이다. CF 모델은 모델링에서 효과적인 이기종 데이터를 결합하는 문제를 해결한다. 일반적으로 지도가 병합되는 방식은 이 모델과 다른 이변량 모델의 주요 차이점이다. 모델링에서 유효 기준의 각 클래스의 가중치는 식 (1)을 사용하여 계산한다. 여기서, pp_a는 클래스 내 천식 발생 횟수 대 해당 클래스의 총 픽셀의 비율이고, pp_s는 연구 영역의 전체 픽셀에 대한 전체 천식 발생 비율이다. CF 모델에서 가중치 출력 값은 -1과 1 사이이다.

$$CF = \begin{cases} \frac{PP_a - PP_s}{PP_s(1 - PP_s)} & \text{if } PP_a > PP_s \\ \frac{PP_s - PP_a}{PP_s(1 - PP_a)} & \text{if } PP_a < PP_s \end{cases} \quad (1)$$

[0051]

[0052]

본 발명의 실시예에 따른 다중 공선성 분석은 모델링에서 독립 변수 간의 상관 관계를 결정한다. 이를 위해 분산 인플레이션 계수(Variance Inflation Factor; VIF) 지수가 다중 공선성 분석에 사용된다. 종래 기술에 따르면 VIF 값이 5 미만이면 다중 공선성 분석이 가능하다. 즉, 독립적 요인 간에는 상관관계가 없다.

[0053]

본 발명의 실시예에 따르면, 특징 선택을 위해 인스턴스 기반 알고리즘에서 영감을 얻은 예측 솔루션을 사용한다. 알고리즘의 개선된 버전인 RReliefF는 회귀 문제에 대한 효율적인 알고리즘이다. 이 방법에서는 데이터 세트의 표본 중에서 랜덤하게 표본을 선택한다. 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 업데이트된다. 선택한 표본 속성 중 하나의 점수는 동일한 클래스의 인접 표본의 관련 속성과 다를 경우 감소한다. 반면, 선택된 표본의 동일한 속성이 다른 부류의 인접 표본의 관련 속성과 다를 경우 속성 점수가 상승한다.

[0054]

본 발명의 실시예에 따른 SVM(Support Vector Machine)은 분류 및 회귀 문제에 대한 머신 러닝모델이다. SVM은 데이터의 n 차원 공간에서 범주의 경계를 정의하는 점들의 집합이다. 회귀 및 예측 문제를 해결하는 데 사용되는 SVR은 SVM의 응용 프로그램 중 하나이다. m 개의 서로 다른 특징을 가진 공간에서 x 의 비선형 입력은 SVR 모델에서 비선형적이며, 이러한 특징을 사용하여 선형 모델을 생성한다. 선형 모델은 식 (2)를 사용하여 계산된다.

[0055]

$$f(x) = \omega^T \phi(x) + b \quad (2)$$

[0056]

여기서 ω 는 가중치 계수 벡터를 나타내고 b 는 바이어스 상수를 나타낸다. 특징 벡터 $\phi(x)$ 는 공간 고차원에 대한 x 의 맵핑을 나타낸다. 추정치의 품질은 $L_\epsilon(\xi)$ 함수에 의해 측정된다. 새로운 유형의 L 함수를 사용하여 $\in$ 함수에 둔감하게 한다. 이 함수의 식은 다음과 같다:

[0057]

$$L_\epsilon(\xi) = \begin{cases} 0, & \text{if } |\xi| < \epsilon \\ |\xi| - \epsilon, & \text{otherwise} \end{cases} \quad (3)$$

[0058]

ω 및 b 는 다음과 같은 최적화 문제를 최소화하여 결정할 수 있다:

[0059]

$$\begin{aligned} & \text{minimize } \frac{1}{2} |\omega|^2 \\ & \text{subject to } \begin{cases} |y_i - \langle \omega, \phi(x_i) \rangle - b| \leq \epsilon \\ i = 1, 2, \dots, m \end{cases} \end{aligned} \quad (4)$$

[0060]

SVR은 ϵ -insensitive를 사용하여 특징 공간의 큰 차원에서 선형 회귀를 수행하고 최소 $|\omega|^2$ 로 모델의 복잡성을 줄이려고 한다: 이는 변수 ξ_i 와 ξ'_i 를 도입하여 ϵ -insensitive 훈련 샘플의 편차를 측정함으로써 설명할 수 있다. 따라서 SVR은 식 (5)와 (6)과 같은 함수 최소화로 정의된다.

[0061]

$$\text{minimize } \frac{1}{2} |\omega|^2 + C \sum_{i=1}^m (\xi_i + \xi'_i) \quad (5)$$

[0062]

$$\text{subject to } \begin{cases} y_i - \langle \omega, \phi(x_i) \rangle - b \leq \epsilon + \xi_i \\ \langle \omega, \phi(x_i) \rangle + b - y_i \leq \epsilon + \xi'_i \\ \xi_i \geq 0, \xi'_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad (6)$$

[0063]

여기서 C 는 모델 복잡성과 적합 오차 사이의 균형을 결정하는 정규화 상수이다.

[0065]

도 8은 본 발명의 일 실시예에 따른 유전 알고리즘의 수행 과정을 설명하기 위한 도면이다.

[0066]

본 발명의 실시예에 따른 유전 알고리즘(Genetic Algorithm; GA)은 다윈의 진화론에 기반을 두고 있다. GA 알고

리즘은 염색체로 대표되는 일련의 답변으로 시작한다. 이 솔루션 집합을 초기 모집단이라고 한다.

[0067] 도 4를 참조하면, 본 발명의 실시예에 따른 GA 알고리즘은 먼저 적합도에 따라 적절한 수의 염색체를 선택한다(Selection). 적합도가 높은 염색체는 생산 과정에서 여러 번 선택할 수 있지만 적합도가 낮은 염색체는 절대 선택되지 않을 수 있다. 그런 다음 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성할 수 있다(Crossover). 그런 다음, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행한다(Mutation). 그것은 이 염색체들의 유전자를 수정함으로써 새로운 정보를 입력하는 방법을 만든다. 그런 다음 최신 모집단을 선택하여 알고리즘의 다음 단계로 이동한다. 이것은 염색체의 적합도를 더 적절한 염색체가 선택된 것과 비교함으로써 이루어진다. 마지막으로, 새로운 모집단의 모든 구성원이 평가된다. 알고리즘은 종료 조건이 충족되면 종료된다. 그렇지 않으면 다음 단계의 시작 모집단은 기존 모집단에서 도출된다.

[0069] 도 9는 본 발명의 일 실시예에 따른 하이브리드 모델의 수행 과정을 설명하기 위한 도면이다.

[0070] 본 발명의 일 실시예에 따른 하이브리드 모델의 수행 과정은 먼저, 입력 데이터를 입력 받는다. 이후, GA 알고리즘(920) 및 TLBO 알고리즘(930)을 각각 수행한다.

[0071] 본 발명의 실시예에 따른 GA 알고리즘(920)은 초기화 수행 후, 염색체로 대표되는 일련의 솔루션으로 시작한다(922). 이 솔루션 집합을 초기 모집단이라고 한다.

[0072] 먼저 적합도에 따라 적절한 수의 염색체를 선택한다(923). 적합도가 높은 염색체는 생산 과정에서 여러 번 선택할 수 있지만 적합도가 낮은 염색체는 절대 선택되지 않을 수 있다. 그런 다음 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성할 수 있다(924). 그런 다음, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행한다(925). 그것은 이 염색체들의 유전자를 수정함으로써 새로운 정보를 입력하는 방법을 만든다(926). 그런 다음 최신 모집단을 선택하여 알고리즘의 다음 단계로 이동한다(927). 이것은 염색체의 적합도를 더 적절한 염색체가 선택된 것과 비교함으로써 이루어진다. 마지막으로, 새로운 모집단의 모든 구성원이 평가된다(928). 알고리즘은 종료 조건이 충족되면 종료된다(929).

[0073] 본 발명의 실시예에 따른 TLBO 알고리즘(930)은 교사-학습 과정에서 도출된 것으로 교사가 교실에서 학생의 성과에 미치는 영향을 기반으로 한다. 변수의 상한과 하한, 생산 대수, 적합도 함수가 이 알고리즘의 입력값이다. 먼저, 초기화 수행 후(931), 변수의 창시자에 따라 클래스 멤버가 생성된다(932). 그러면 적합도 함수에 따라 가장 좋은 솔루션이 교사로 선정된다(933). 교사 단계라고 불리는 1단계에서는 교사가 클래스 구성원들의 이해 수준에 영향을 미쳐 클래스 구성원들의 평균적인 지식을 높이려 하고, 그 결과 지식이 증가하게 된다(934). 새로운 솔루션이 있는지 여부를 판단한 후(935), M_{new} 가 최선의 현재 솔루션으로 간주되고 M 이 현재 평균이라면, 각 현재 솔루션이 식 (7)을 통해 새로운 솔루션이 된다(936).

[0074]
$$X_{new,i} = X_{old,i} + r(M_{new} + T_F M) \quad (7)$$

[0075] 위의 관계에서, r 은 0과 1 사이의 랜덤 변수이고, T_F 는 학습 가속(-2와 2 사이의 난수)이다. 학습자 참여를 통해 얻은 지식은 학습 기간 동안 클래스의 평균 수준을 향상시킨다.

[0076] 다시 새로운 솔루션이 있는지 여부를 판단한 후(937), 두 번째 단계에서 알고리즘의 과정은 클래스의 두 멤버가 서로 랜덤으로 상호 작용하는 것이다(938). 두 멤버의 상호작용으로 인한 이해는 다른 구성원들의 지식 수준에 영향을 미친다. x_i 와 x_j 가 서로 상호작용하는 두 개의 클래스 멤버이고 $x_i = x_j$, $x_i > x_j$ 이면 변환 방법은 식 (10)을 기반으로 계산되며 그렇지 않으면 식 (9)에 따라 계산된다.

[0077]
$$X_{new,i} = X_{old,i} + r_i(x_i - x_j) \quad (10)$$

[0078]
$$X_{new,i} = X_{old,i} + r_i(x_j - x_i) \quad (9)$$

[0079] 적합도 함수가 개선되면 이 결과는 수용할 수 있다. 최적의 솔루션을 찾거나 종료 조건이 충족될 때까지 절차가 복제된다(939).

[0080] 본 발명의 실시예에 따른 하이브리드 모델은 아스트랩론 영역을 결정하기 위해 SVR, SVR-GA, SVR-TLBO 모델을 사용하였다(912). SVR 모델의 매개변수(C , γ , ϵ)는 GA 알고리즘과 TLBO 알고리즘을 사용하여 최적화되었다. 이

러한 하이브리드 모델의 목적은 예상 변수와 대상 변수 사이의 적절한 정확도와 최소 오차로 이러한 매개변수의 값을 찾는 것이다. RMSE(Root Mean Square Error) 지수는 목적 함수이다.

[0081] 천식이 발생하기 쉬운 영역의 모델링과 맵핑을 평가하기 위해 RMSE, MAE, ROC 지수가 사용되었다(913). 다음은 이러한 지표에 대한 설명이다:

[0082] RMSE는 하이브리드 모델에서 입력 및 출력 모델링 데이터 사이의 평균 차이를 나타낸다. 이 지수는 식 (10)을 사용하여 계산된다.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (p_i - o_i)^2}{n}} \quad (10)$$

[0084] MAE(Mean absolute error)는 관측된 변수와 하이브리드 모델에 의해 계산된 출력 간의 차이를 보여준다. MAE 지수는 식 (11)을 사용하여 계산된다.

$$\text{MAE} = \frac{\sum_{i=1}^n |p_i - o_i|}{n} \quad (11)$$

[0086] 식 (10) 및 (11)에서 p_i 는 모델의 계산된 값이고, o_i 는 관측 변수의 값이며, n 은 관측치의 수이다.

[0087] ROC(Receiver operating characteristic) 곡선은 수평축에 픽셀을 그려서 하이브리드 모델로 천식의 발생 여부를 정확하게 예측한 그래프이다. 하이브리드 모델에서 잘못 예상되는 픽셀의 비율도 수직 축에 표시된다. 이 곡선 아래 영역을 AUC라고 하며, AUC가 가장 높은 모델은 상대적 성능이 더 우수하다. 천식이 발생하기 쉬운 지역의 지도를 평가할 때 AUC가 남성 1명에 가까울수록 하이브리드 모델의 정확도가 높아진다.

[0088] 본 발명의 실시예에 따른 하이브리드 모델은 RMSE, MAE, ROC 지수를 종합하여 천식이 발생하기 쉬운 영역의 모델링과 맵핑을 평가한다(914).

[0090] 도 10은 본 발명의 일 실시예에 따른 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 시스템의 구성을 나타내는 도면이다.

[0091] 본 실시예에 따른 시공간 모델링 시스템(1000)은 프로세서(1010), 버스(1020), 네트워크 인터페이스(1030), 메모리(1040) 및 데이터베이스(1050)를 포함할 수 있다. 메모리(1040)는 운영체제(1041) 및 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 루틴(1042)을 포함할 수 있다. 프로세서(1010)는 데이터 수집부(1011), 분석부(1012), 학습부(1013) 및 시공간 모델링부(1014)를 포함할 수 있다. 다른 실시예들에서 시공간 모델링 시스템(1000)은 도 10의 구성요소들보다 더 많은 구성요소들을 포함할 수도 있다. 그러나, 대부분의 종래기술적 구성요소들을 명확하게 도시할 필요성은 없다. 예를 들어, 시공간 모델링 시스템(1000)은 디스플레이나 트랜시버(transceiver)와 같은 다른 구성요소들을 포함할 수도 있다.

[0092] 메모리(1040)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(1040)에는 운영체제(1041)와 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링 루틴(1042)을 위한 프로그램 코드가 저장될 수 있다. 이러한 소프트웨어 구성요소들은 드라이브 메커니즘(drive mechanism, 미도시)을 이용하여 메모리(1040)와는 별도의 컴퓨터에서 판독 가능한 기록 매체로부터 로딩될 수 있다. 이러한 별도의 컴퓨터에서 판독 가능한 기록 매체는 플로피 드라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리 카드 등의 컴퓨터에서 판독 가능한 기록 매체(미도시)를 포함할 수 있다. 다른 실시예에서 소프트웨어 구성요소들은 컴퓨터에서 판독 가능한 기록 매체가 아닌 네트워크 인터페이스(1030)를 통해 메모리(1040)에 로딩될 수도 있다.

[0093] 버스(1020)는 시공간 모델링 시스템(1000)의 구성요소들간의 통신 및 데이터 전송을 가능하게 할 수 있다. 버스(1020)는 고속 시리얼 버스(high-speed serial bus), 병렬 버스(parallel bus), SAN(Storage Area Network) 및/또는 다른 적절한 통신 기술을 이용하여 구성될 수 있다.

[0094] 네트워크 인터페이스(1030)는 시공간 모델링 시스템(1000)을 컴퓨터 네트워크에 연결하기 위한 컴퓨터 하드웨어

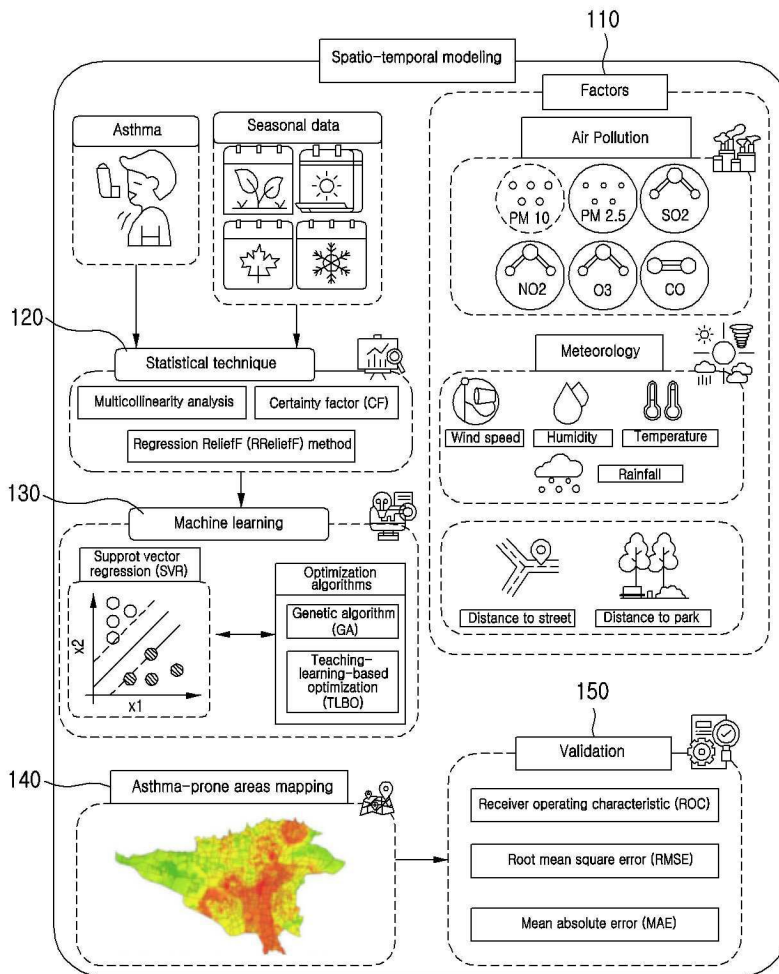
구성요소일 수 있다. 네트워크 인터페이스(1030)는 시공간 모델링 시스템(1000)을 무선 또는 유선 커넥션을 통해 컴퓨터 네트워크에 연결시킬 수 있다.

- [0095] 데이터베이스(1050)는 메타휴리스틱 알고리즘으로 최적화된 기계 학습을 이용한 천식 취약 지역의 시공간 모델링을 위해 필요한 모든 정보를 저장 및 유지하는 역할을 할 수 있다. 도 10에서는 시공간 모델링 시스템(1000)의 내부에 데이터베이스(1050)를 구축하여 포함하는 것으로 도시하고 있으나, 이에 한정되는 것은 아니며 시스템 구현 방식이나 환경 등에 따라 생략될 수 있고 혹은 전체 또는 일부의 데이터베이스가 별개의 다른 시스템 상에 구축된 외부 데이터베이스로서 존재하는 것 또한 가능하다.
- [0096] 프로세서(1010)는 기본적인 산술, 로직 및 시공간 모델링 시스템(1000)의 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령은 메모리(1040) 또는 네트워크 인터페이스(1030)에 의해, 그리고 버스(1020)를 통해 프로세서(1010)로 제공될 수 있다. 프로세서(1010)는 데이터 수집부(1011), 분석부(1012), 학습부(1013) 및 시공간 모델링부(1014)를 위한 프로그램 코드를 실행하도록 구성될 수 있다. 이러한 프로그램 코드는 메모리(1040)와 같은 기록 장치에 저장될 수 있다.
- [0097] 데이터 수집부(1011), 분석부(1012), 학습부(1013) 및 시공간 모델링부(1014)는 도 2의 단계들(210~250)을 수행하기 위해 구성될 수 있다.
- [0098] 시공간 모델링 시스템(1000)은 데이터 수집부(1011), 분석부(1012), 학습부(1013) 및 시공간 모델링부(1014)를 포함할 수 있다.
- [0099] 본 발명의 실시예에 따른 데이터 수집부(1011)는 시공간 모델링을 위해 종속 데이터 및 독립 데이터를 수집한다.
- [0100] 본 발명의 실시예에 따른 데이터 수집부(1011)는 상기 시공간 모델링을 수행하기 위해 위치를 포함하는 종속 데이터와 환경적 요인을 포함하는 독립 데이터를 수집하고, 상기 수집된 종속 데이터 및 독립 데이터는 훈련 데이터 및 테스트 데이터를 위한 두 가지 하위 세트를 생성하고, 테스트 데이터는 모델을 훈련하는 데 사용되지 않고, 평가 및 예측 결과를 실제 값과 비교하는 데 사용된다.
- [0101] 본 발명의 실시예에 따른 분석부(1012)는 입력 데이터 모델링에 대해 통계 분석을 수행한다.
- [0102] 본 발명의 실시예에 따른 분석부(1012)는 독립 변수 간의 상관관계를 결정하는 다중 공선성 분석(multicollinearity analysis), 모델링에서 이기종 데이터를 결합하기 위해 입력 변수의 모호성을 처리하는 CF(Certainty Factor) 및 데이터 세트의 표본 중 랜덤하게 표본을 선택하고, 각 매개변수의 유의 정도는 선택한 표본과 인근 두 표본 간의 차이에 따라 분석하는 RRelief(Regression Relief)를 분석한다.
- [0103] 본 발명의 실시예에 따른 학습부(1013)는 GA 알고리즘 및 TLBO 알고리즘을 사용하여 SVR 모델의 최적 매개변수를 획득한다.
- [0104] 본 발명의 실시예에 따른 학습부(1013)는 상기 GA 알고리즘을 수행하고, 상기 상기 GA 알고리즘은 적합도에 따라 적절한 수의 염색체를 선택하고, 교차 연산자는 부모 염색체에 작용하여 새로운 염색체를 결합하고 생성하며, 돌연변이 연산자는 교차 연산에 따른 염색체에 돌연변이를 수행하고, 상기 염색체들의 유전자를 수정함으로써 새로운 정보를 입력한 후 최신 모집단을 선택한다.
- [0105] 본 발명의 실시예에 따른 학습부(1013)는 상기 TLBO 알고리즘을 수행하고, 변수의 상한과 하한, 생산 대수, 적합성 함수를 입력 데이터로 하여 변수에 따라 클래스 멤버가 생성되고, 적합도 함수에 따른 최적 솔루션이 교사로 선택되며, 새로운 솔루션 업데이트 여부에 따라 적합도 함수를 개선한다.
- [0106] 본 발명의 실시예에 따른 학습부(1013)는 SVR 모델의 최적 매개변수는 GA 알고리즘 및 TLBO 알고리즘을 통해 획득된다.
- [0107] 본 발명의 실시예에 따른 시공간 모델링부(1014)는 GA 알고리즘, TLBO 알고리즘 및 SVR 모델을 통해 획득한 최적 매개변수를 이용하여 시공간 모델링을 위한 지도를 맵핑한다.
- [0108] 본 발명의 실시예에 따른 시공간 모델링부(1014)는 상기 맵핑된 지도에 대하여 메트릭 지표를 사용하여 시공간 모델링 결과를 평가한다.
- [0109] 본 발명의 실시예에 따른 시공간 모델링부(1014)는 RMSE(root mean square error), MAE(Mean absolute error), ROC(Receiver operating characteristic) 지수를 종합하여 시공간 모델링 결과를 평가한다.

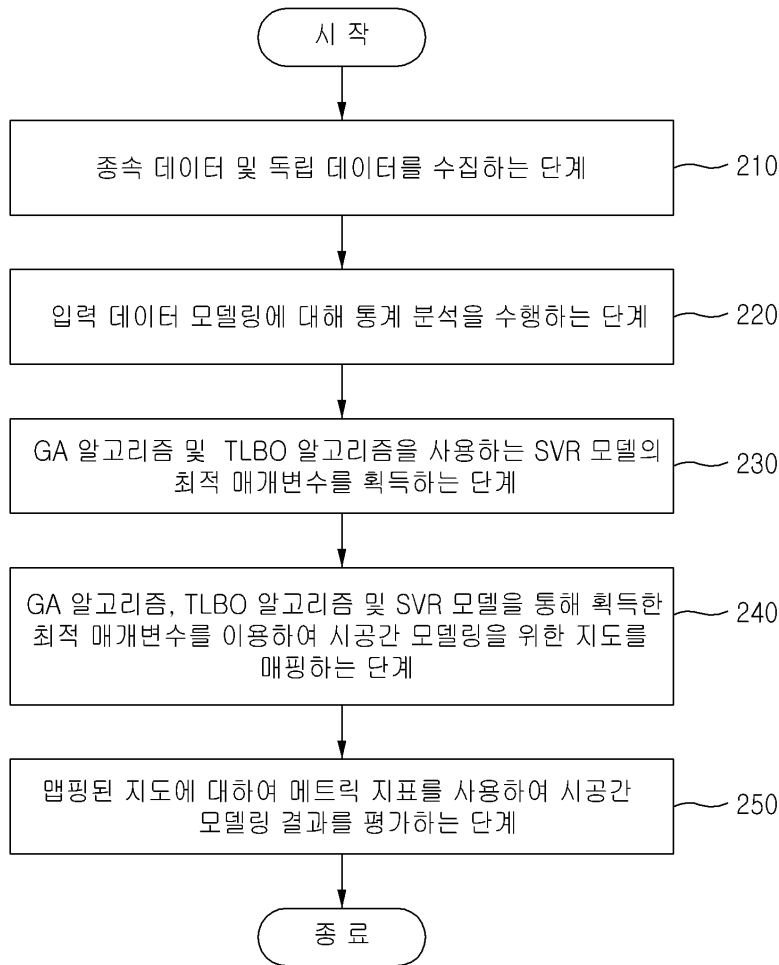
- [0111] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0112] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치에 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0113] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0114] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.
- [0115] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

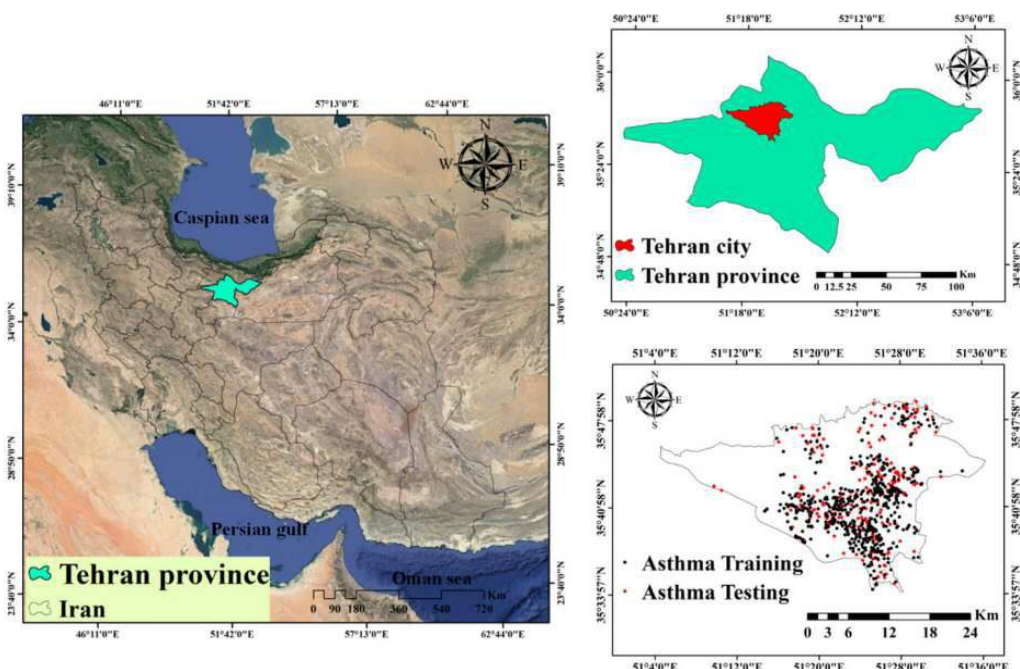
도면1



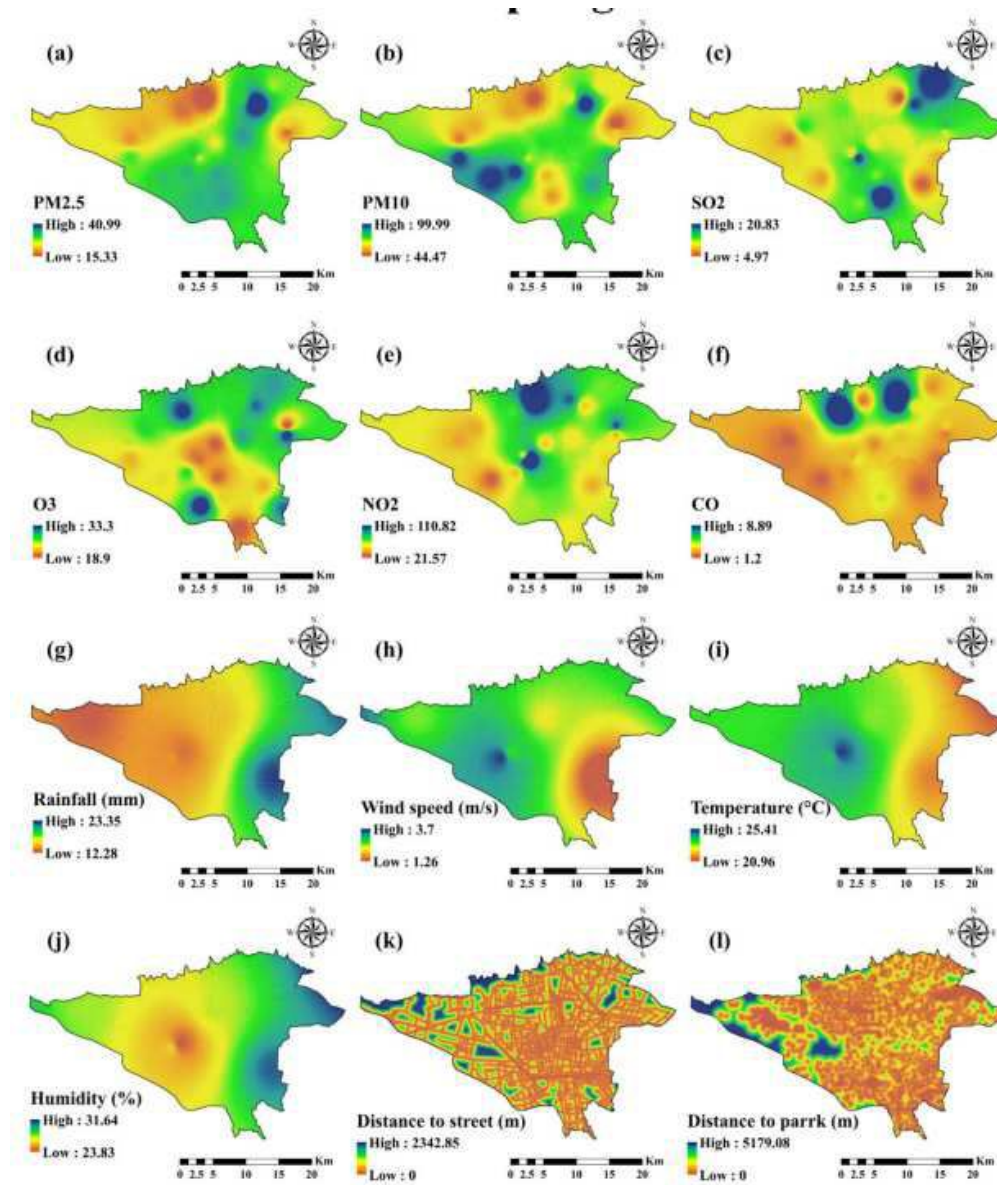
도면2



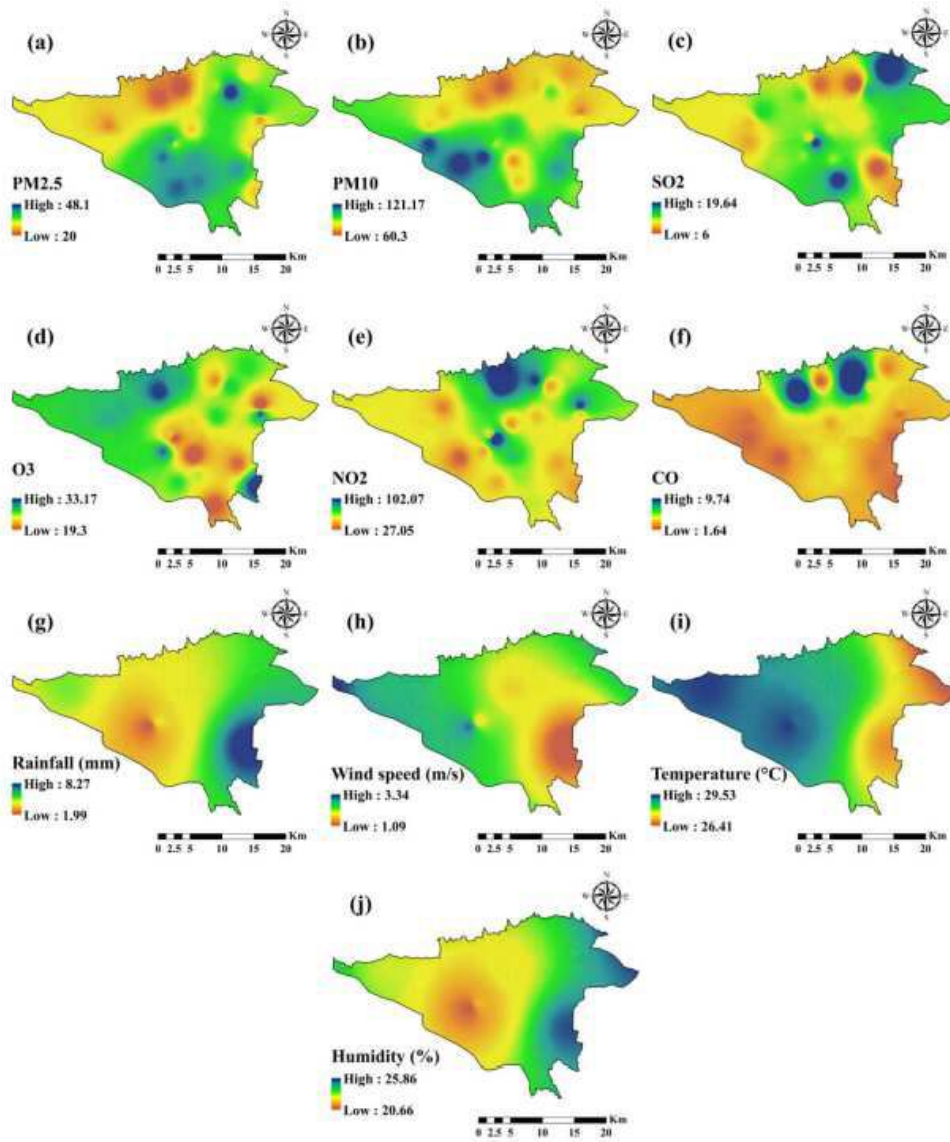
도면3



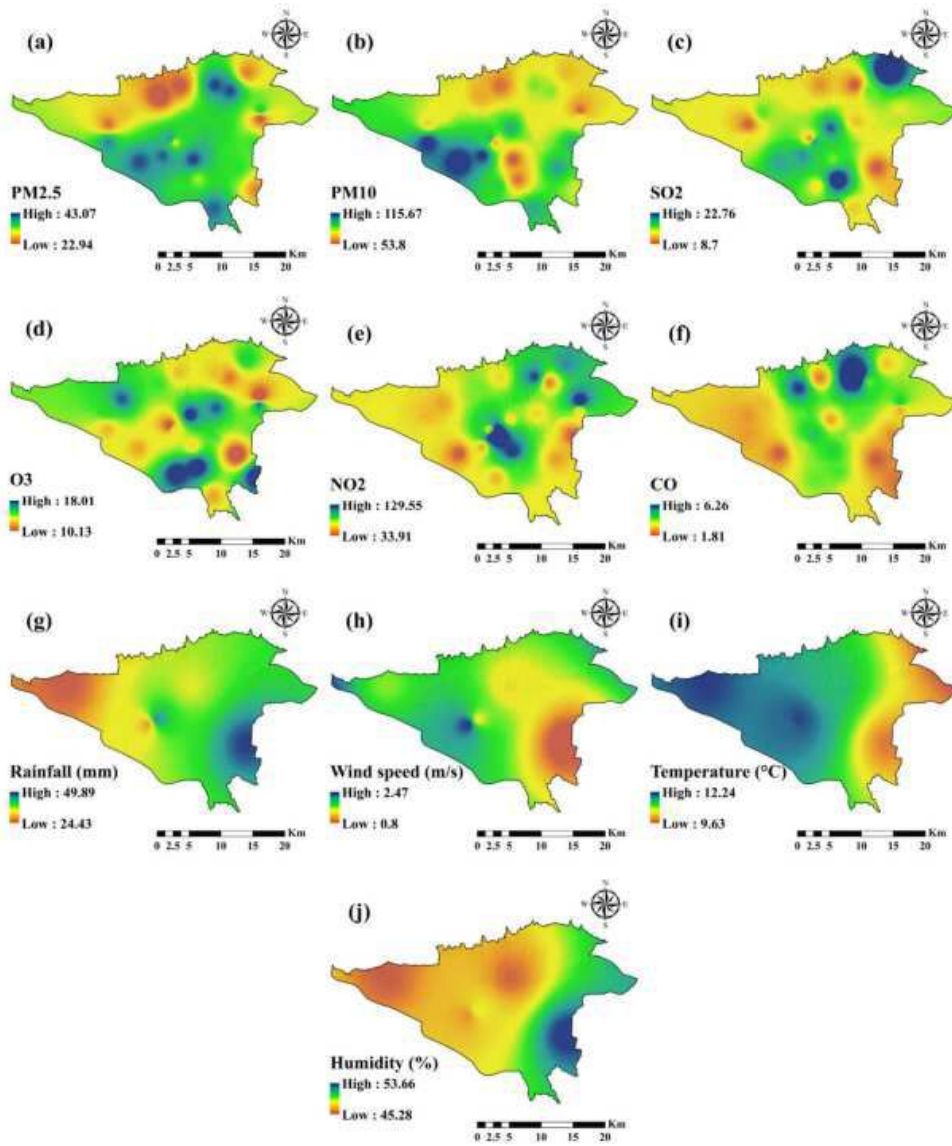
도면4



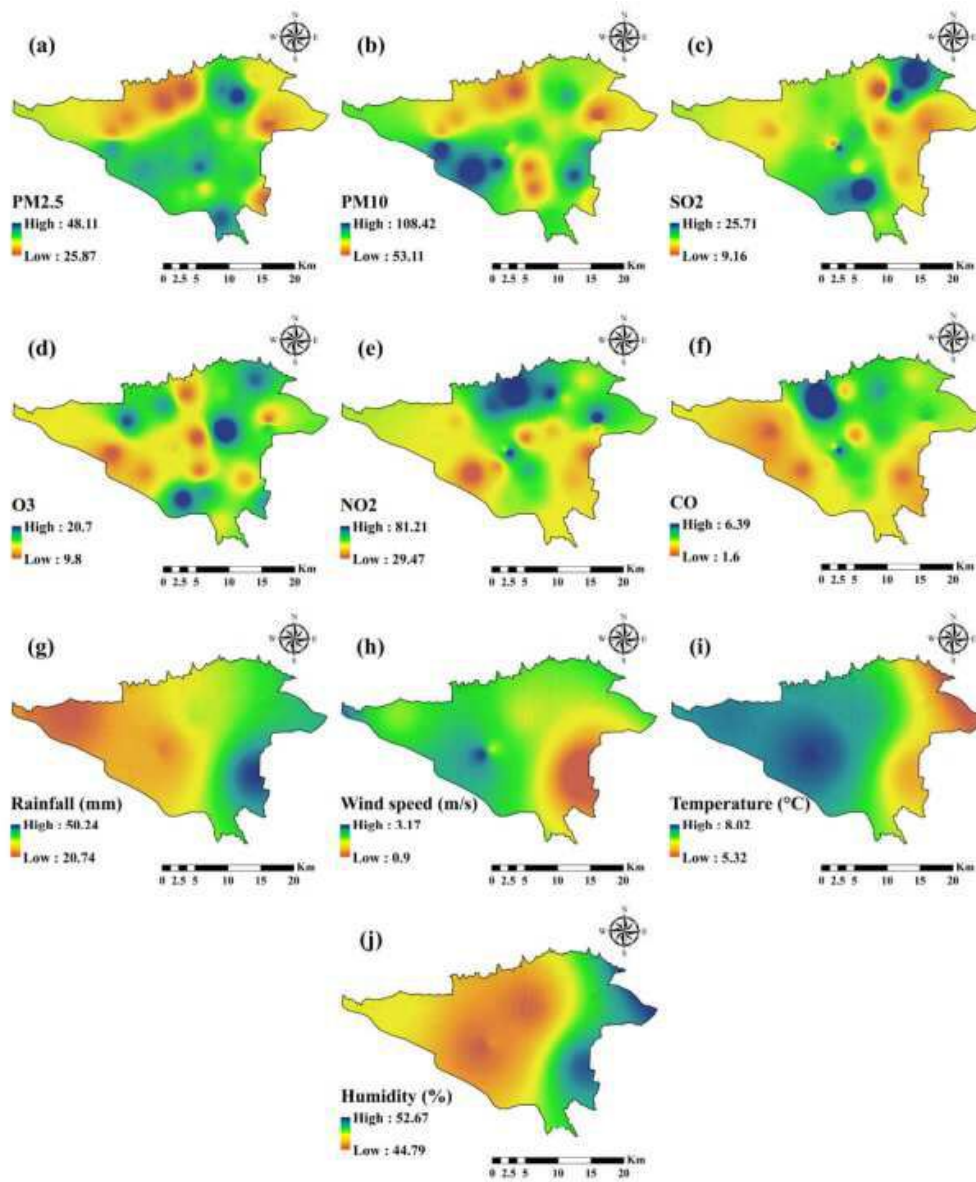
도면5



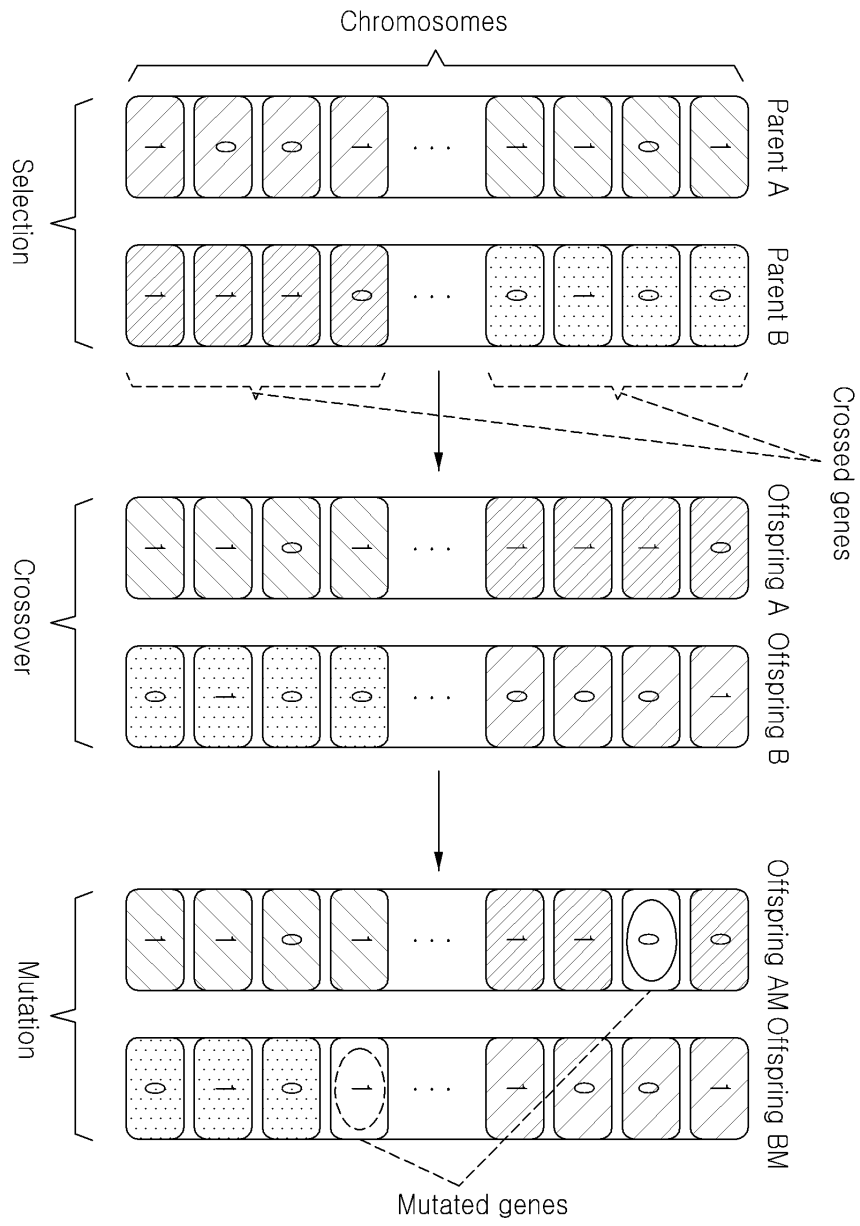
도면6



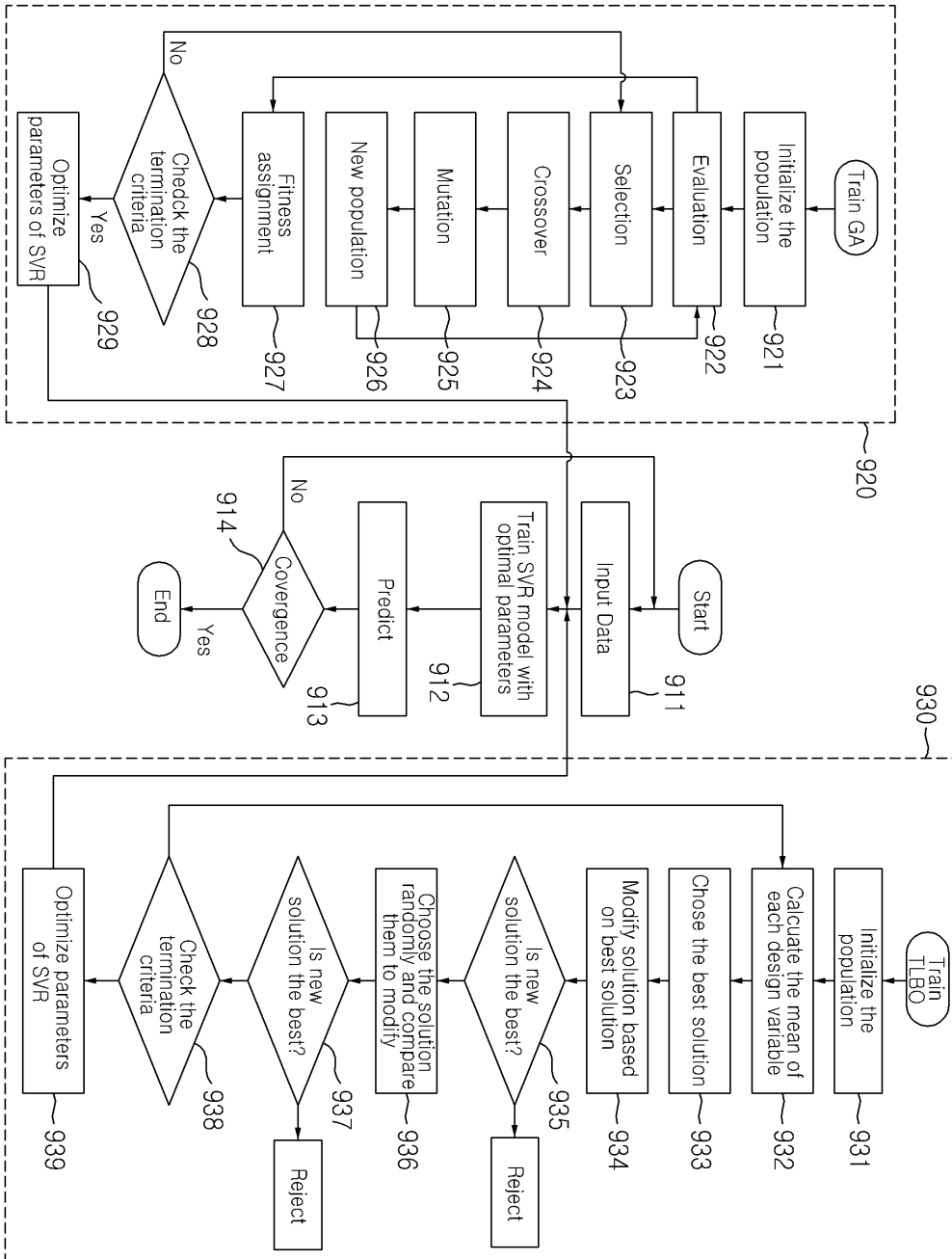
도면7



도면8



도면9



도면10

