(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0039696
(43) 공개일자 2024년03월27일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
(21) 출원번호 10-2022-0118325
(22) 출원일자 2022년09월20일
심사청구일자 2022년09월20일

(71) 출원인
한국기술교육대학교 산학협력단
충청남도 천안시 동남구 병천면 충절로 1600 (한국기술교육대학교내)
(72) 발명자
김원태
충청남도 천안시 서북구 공원로 176 천안불당시티프라디움3차, 306-3102
김한진
충청북도 청주시 청원구 오창읍 오창공원로 96 한신더휴센트럴파크, 103-3504
김영진
충청북도 청주시 청원구 오창읍 오창공원로 96 한신더휴센트럴파크, 103-3504
(74) 대리인
특허법인팬브릿지

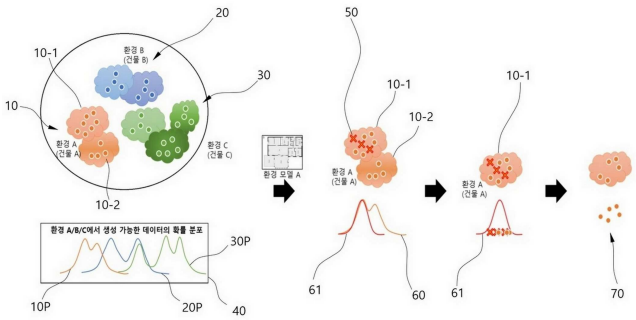
전체 청구항 수 : 총 18 항

(54) 발명의 명칭 **환경 모델을 이용한 합성 데이터 생성 장치 및 방법**

(57) 요 약

본 명세서는 인공지능 모델을 학습시키기 위한 합성 데이터를 생성하는 방법에 관한 것이다. 상기 방법은, 복수의 실제 환경을 각각 모사한 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 독출하는 단계; 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 단계; 상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하는 단계; 및 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하는 단계;를 포함하며, 인공지능 모델을 학습시키는 데 사용할 수 있는 고품질의 학습용 합성 데이터를 생성할 수 있는 효과가 있다.

대 표 도



이 발명을 지원한 국가연구개발사업

과제고유번호	1711160574
과제번호	2022-0-00866-001
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	스마트제조혁신기술개발
연구과제명	(세부1-1) 대규모 확장성 및 고신뢰 분산 시뮬레이션을 지원하는 제조 디지털 트윈 프레임워크 기술 개발
기 여 율	1/1
과제수행기관명	한국전자통신연구원
연구기간	2022.04.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

인공지능 모델을 학습시키기 위한 합성 데이터를 생성하는 방법에 있어서, 복수의 실제 환경을 각각 모사한 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 독출하는 단계;

상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 단계;

상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하는 단계; 및

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하는 단계;를 포함하는

방법.

청구항 2

제1 항에 있어서,

상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 단계는,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출하는 단계인

것을 특징으로 하는 방법.

청구항 3

제2 항에 있어서,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야의 연관 관계는,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 결정되는

것을 특징으로 하는 방법.

청구항 4

제3 항에 있어서,

상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하는 단계는,

상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성 별로 대응되는 파라미터를 상기 환경 특성에 따라 수정하는 단계인

것을 특징으로 하는 방법.

청구항 5

제1 항에 있어서,

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 단계;를 더 포함하는

방법.

청구항 6

제5 항에 있어서,

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 단계는,

상기 타겟 환경에서 획득한 실제 데이터를 수신하는 단계;

상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하는 단계; 및

상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정하는 단계;를 포함하는 것을 특징으로 하는

방법.

청구항 7

인공지능 모델을 학습시키기 위한 합성 데이터 생성 장치에 있어서,

복수의 실제 환경을 각각 모사한 환경 모델들을 저장하는 환경 모델 데이터베이스;

상기 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 상기 환경 모델 데이터베이스로부터 독출하고, 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 환경 특성 분석부; 및

상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하고, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하는 합성 데이터 생성부;를 포함하는

합성 데이터 생성 장치.

청구항 8

제7 항에 있어서, 상기 환경 특성 분석부는,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출하는

것을 특징으로 하는 합성 데이터 생성 장치.

청구항 9

제8 항에 있어서, 상기 환경 특성 분석부는,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 상기 연관 관계를 결정하는

것을 특징으로 하는 합성 데이터 생성 장치.

청구항 10

제9 항에 있어서, 상기 합성 데이터 생성부는,

상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성 별로 대응되는 파라미터를 상기 환경 특성에 따라 수정하는

것을 특징으로 하는 합성 데이터 생성 장치.

청구항 11

제7 항에 있어서,

상기 타겟 환경에서 획득한 실제 데이터를 수집하는 데이터 수집부;를 더 포함하고,

상기 합성 데이터 생성부는,

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 실제 데이터가 속하는 데이터 특성에 더 매칭시키는

것을 특징으로 하는 합성 데이터 생성 장치.

청구항 12

제11 항에 있어서, 상기 합성 데이터 생성부는,

상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하고, 및

상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정하는

것을 특징으로 하는 합성 데이터 생성 장치.

청구항 13

인공지능 모델을 학습시키기 위한 합성 데이터 생성용 전자 장치에 있어서,

통신 회로;

메모리; 및

상기 메모리와 작동적으로(operatively) 연결되는 프로세서를 포함하고,

상기 메모리는, 실행 시, 상기 프로세서가,

복수의 실제 환경을 각각 모사한 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 독출하고,

상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하고,

상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하고,

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하도록 하는 인스트럭션들(instructions)을 저장하는

전자 장치.

청구항 14

제13 항에 있어서,

상기 인스트럭션들은, 상기 프로세서가,

상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하고,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출하도록 하는

것을 특징으로 하는 전자 장치.

청구항 15

제14 항에 있어서,

상기 인스트럭션들은, 상기 프로세서가,

상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 상기 연관 관계를 결정하도록 하는

것을 특징으로 하는 전자 장치.

청구항 16

제15 항에 있어서,

상기 인스트럭션들은, 상기 프로세서가,

상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성 별로 대응되는 파라미터를 상기 환경 특성에 따라 수정하도록 하는

것을 특징으로 하는 전자 장치.

청구항 17

제13 항에 있어서,

상기 인스트럭션들은, 상기 프로세서가,

상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키도록 하는

것을 특징으로 하는 전자 장치.

청구항 18

제17 항에 있어서,

상기 인스트럭션들은, 상기 프로세서가,

상기 타겟 환경에서 획득한 실제 데이터를 수신하고,

상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하고, 및

상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정하도록 하는

것을 특징으로 하는 전자 장치.

발명의 설명

기 술 분 야

[0001] 본 명세서는 인공지능 모델을 학습시키기 위한 데이터의 생성에 관한 것으로, 더욱 상세하게는 환경 모델을 이용하여 고품질의 학습용 합성 데이터를 생성하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 오늘날 인공지능 적용 분야가 다양해지고 그 중요도가 높아짐에 따라 자연스럽게 데이터 확보가 필수적으로 논의되고 있다. 인공지능 시대의 데이터는 3차 산업혁명 시대의 원유로 비유될 정도로 그 중요도가 매우 크며 데이터의 품질이 인공지능의 성능에 직접적인 영향을 미치기 때문에 데이터 확보의 중요성은 점점 더 높아지고 있다. 그러나, 인공지능을 위한 데이터를 수집하는 것은 높은 비용, 시간, 혹은 보안 등의 문제로 인해 쉽지 않으며, 재난과 같은 일부 산업 분야에서는 인공지능 개발에 필요한 데이터를 즉시 측정하거나 수집하는 것이 매우 어려운 실정이다.

[0003] 최근에는 이러한 문제를 해결하기 위해 컴퓨터 시뮬레이션이나 알고리즘을 통해 실 데이터를 모사한 가상의 데이터를 생성하는 합성 데이터(synthetic data) 생성 기법이 주목받고 있다. 합성 데이터는 실제 환경에서 수집되거나 측정된 데이터가 아닌 가상의 환경(예컨대, 시뮬레이션 환경) 및 합성 알고리즘을 통해 또는 전술한 두 방법의 결합을 통해 생성할 수 있다. 합성 데이터 생성 기법은 비용 및 시간 절감, 개인정보보호는 물론 물리적으로 통제되지 않는 가상 환경을 활용한다는 측면에서 데이터의 다양성을 확보할 수 있는 장점이 있다.

[0004] 위 기재된 내용은 오직 본 발명의 기술적 사상들에 대한 배경 기술의 이해를 돕기 위한 것이며, 따라서 그것은 본 발명의 기술 분야의 당업자에게 알려진 선행 기술에 해당하는 내용으로 이해될 수 없다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국공개특허공보 제10-2021-0124811호, 2021.10.15.

발명의 내용

해결하려는 과제

[0006] 합성 데이터를 생성할 때 가장 중요한 것은 생성된 합성 데이터가 얼마나 실제 데이터를 잘 반영할 수 있는 지, 즉 효용성이 높은 지의 여부이다. 효용성이 높다는 것은 실제 데이터의 특성을 잘 반영했다는 것을 의미하며 이는 합성 데이터 품질의 척도가 된다.

[0007] 기존의 합성 데이터 생성 방법은 시간적 혹은 물리적 제약 없이 데이터 생성이 가능하므로 다양한 데이터를 확보할 수 있다는 장점이 있지만, 가상의 환경 또는 통계적 모델 등을 이용하기 때문에 실제 데이터의 특성과는 다소 다른 합성 데이터를 생성할 수 있는 문제가 있다.

[0008] 본 명세서의 실시 예는 전술한 문제점을 해결하기 위한 것으로서, 본 명세서의 일 실시 예는 실제 환경을 모사한 환경 모델을 이용하여 생성된 합성 데이터가 실제 데이터의 특성을 제대로 반영할 수 있도록 하는 것을 목적으로 한다.

[0009] 본 발명이 해결하고자 하는 과제들은 이상에서 언급된 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0010] 본 명세서는 인공지능 모델을 학습시키기 위한 합성 데이터를 생성하는 방법을 제시한다. 상기 방법은, 복수의 실제 환경을 각각 모사한 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 독출하는 단계; 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 단계; 상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하는 단계; 및 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하는 단계;를 포함할 수 있다.

[0011] 상기 인공지능 모델을 학습시키기 위한 합성 데이터를 생성하는 방법 및 그 밖의 실시예는 다음과 같은 특징을 포함할 수 있다.

[0012] 실시 예에 따라, 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 단계는, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출하는 단계일 수 있다.

[0013] 실시 예에 따라, 또한, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야의 연관 관계는, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 결정될 수 있다.

[0014] 실시 예에 따라, 또한, 상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하는 단계는, 상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성 별로 대응되는 파라미터를 상기 환경 특성에 따라 수정하는 단계일 수 있다.

[0015] 실시 예에 따라, 또한, 상기 방법은, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 단계;를 더 포함할 수 있다.

[0016] 실시 예에 따라, 또한, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 단계는, 상기 타겟 환경에서 획득한 실제 데이터를 수신하는 단계; 상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하는 단계; 및 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이

터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정하는 단계;를 포함할 수 있다.

[0017] 한편, 본 명세서는 인공지능 모델을 학습시키기 위한 합성 데이터 생성 장치를 제시한다. 상기 장치는, 복수의 실제 환경을 각각 모사한 환경 모델들을 저장하는 환경 모델 데이터베이스; 상기 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 상기 환경 모델 데이터베이스로부터 독출하고, 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하는 환경 특성 분석부; 및 상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하고, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하는 합성 데이터 생성부;를 포함할 수 있다.

[0018] 상기 합성 데이터 생성 장치 및 그 밖의 실시예는 다음과 같은 특징을 포함할 수 있다.

[0019] 실시 예에 따라, 상기 환경 특성 분석부는, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출할 수 있다.

[0020] 실시 예에 따라, 또한, 상기 환경 특성 분석부는, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 상기 연관 관계를 결정할 수 있다.

[0021] 실시 예에 따라, 또한, 상기 합성 데이터 생성부는, 상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성별로 대응되는 파라미터를 상기 환경 특성에 따라 수정할 수 있다.

[0022] 실시 예에 따라, 또한, 상기 합성 데이터 생성 장치는, 상기 타겟 환경에서 획득한 실제 데이터를 수집하는 데이터 수집부;를 더 포함하고, 상기 합성 데이터 생성부는, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 실제 데이터가 속하는 데이터 특성에 더 매칭시킬 수 있다.

[0023] 실시 예에 따라, 또한, 상기 합성 데이터 생성부는, 상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하고, 및 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정할 수 있다.

[0024] 또 다른 한편, 본 명세서는 인공지능 모델을 학습시키기 위한 합성 데이터 생성용 전자 장치를 제시한다. 상기 장치는, 통신 회로; 메모리; 및 상기 메모리와 작동적으로(operatively) 연결되는 프로세서를 포함하고, 상기 메모리는, 실행 시, 상기 프로세서가, 복수의 실제 환경을 각각 모사한 복수의 환경 모델들 중 합성 데이터를 생성하고자 하는 타겟 환경을 모사한 타겟 환경 모델을 독출하고, 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하고, 상기 합성 데이터가 상기 환경 특성을 반영하도록 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성하고, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 상기 합성 데이터를 생성하도록 하는 인스트럭션들(instructions)을 저장할 수 있다.

[0025] 상기 전자 장치 및 그 밖의 실시예는 다음과 같은 특징을 포함할 수 있다.

[0026] 실시 예에 따라, 상기 인스트럭션들은, 상기 프로세서가, 상기 타겟 환경에서 데이터의 생성에 영향을 미치는 환경 특성을 추출하고, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 상기 환경 특성을 추출하도록 할 수 있다.

[0027] 실시 예에 따라, 또한, 상기 인스트럭션들은, 상기 프로세서가, 상기 타겟 환경 및 상기 합성 데이터의 응용 분야에 관련된 온톨로지에 기초하여 상기 연관 관계를 결정하도록 할 수 있다.

[0028] 실시 예에 따라, 또한, 상기 인스트럭션들은, 상기 프로세서가, 상기 합성 데이터 생성 함수의 파라미터 중 상기 환경 특성 별로 대응되는 파라미터를 상기 환경 특성에 따라 수정하도록 할 수 있다.

[0029] 실시 예에 따라, 또한, 상기 인스트럭션들은, 상기 프로세서가, 상기 타겟 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 타겟 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키도록 할 수 있다.

[0030] 실시 예에 따라, 또한, 상기 인스트럭션들은, 상기 프로세서가, 상기 타겟 환경에서 획득한 실제 데이터를 수신하고, 상기 실제 데이터에 기초하여 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분

포 영역을 상기 실제 데이터가 속하는 데이터 분포 영역으로 결정하고, 및 상기 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정하도록 할 수 있다.

발명의 효과

[0031] 본 명세서에 개시된 실시 예에 따르면 실제 데이터의 특성을 잘 나타내는 합성 데이터 셋을 생성함으로써 인공지능 학습을 위한 실제 데이터를 얻는 비용, 시간, 보안 상의 문제를 해결할 수 있는 효과가 있다.

[0032] 또한, 본 명세서에 개시된 실시 예에 따르면 실제 환경에 대한 환경 모델을 기반으로 실제 환경에 대한 학습 데이터의 특성을 제대로 반영하여 고품질의 합성 데이터를 생성할 수 있는 효과가 있다.

[0033] 또한, 본 명세서에 개시된 실시 예에 따르면 인공지능 학습 시 사용하는 데이터의 분포가 인공지능이 적용될 실제 환경의 데이터 분포와 상이한 차이를 보여 인공지능 모델의 성능이 떨어지는 문제를 해결할 수 있는 효과가 있다.

[0034] 또한, 본 명세서에 개시된 실시 예에 따르면 환경 모델을 기반으로 확보한 고품질의 대용량 합성 데이터를 바탕으로 인공지능 모델의 과적합 문제를 완화하여 일반화 성능을 향상시킬 수 있는 효과가 있다.

[0035] 한편, 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0036] 본 명세서에 첨부되는 다음의 도면들은 본 발명의 바람직한 실시예를 예시하는 것이며, 발명을 실시하기 위한 구체적인 내용과 함께 본 발명의 기술사상을 더욱 이해시키는 역할을 하는 것이므로, 본 발명은 그러한 도면에 기재된 사항에만 한정되어 해석되어서는 아니 된다.

도 1은 환경 모델을 이용한 합성 데이터 생성 방법을 설명하는 개념도이다.

도 2는 실시 예에 따른 합성 데이터 생성 장치의 구성을 도시하는 블록도이다.

도 3은 실시 예에 따른 합성 데이터 생성 장치의 구성을 도시하는 블록도이다.

도 4는 실제 환경과 합성 데이터의 응용 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

도 5는 실제 환경과 통신 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

도 6은 실제 환경과 공조 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

도 7은 실제 환경과 화재 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

도 8은 실제 환경과 화재 분야 사이의 연관 관계를 나타내는 온톨로지에서 연관 관계가 추가되는 예를 도시한 도면이다.

도 9는 실시 예들에 따른 합성 데이터 생성 방법을 설명하는 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0037] 본 명세서에 개시된 기술은 인공지능 모델을 학습시키기 위한 합성 데이터를 생성하는 장치 및 방법에 적용될 수 있다. 그러나 본 명세서에 개시된 기술은 이에 한정되지 않고, 상기 기술의 기술적 사상이 적용될 수 있는 모든 장치 및 방법에도 적용될 수 있다.

[0038] 본 명세서에서 사용되는 기술적 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 명세서에 개시된 기술의 사상을 한정하려는 의도가 아님을 유의해야 한다. 또한, 본 명세서에서 사용되는 기술적 용어는 본 명세서에서 특별히 다른 의미로 정의되지 않는 한, 본 명세서에 개시된 기술이 속하는 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 의미로 해석되어야 하며, 과도하게 포괄적인 의미로 해석되거나, 과도하게 축소된 의미로 해석되지 않아야 한다. 또한, 본 명세서에서 사용되는 기술적인 용어가 본 명세서에 개시된 기술의 사상을 정확하게 표현하지 못하는 잘못된 기술적 용어일 때에는, 본 명세서에 개시된 기술이 속하는 분야에서 통상의 지식을 가진 자가 올바르게 이해할 수 있는 기술적 용어로 대체되어 이해되어야 할 것이다. 또한, 본

명세서에서 사용되는 일반적인 용어는 사전에 정의되어 있는 바에 따라, 또는 전후 문맥 상에 따라 해석되어야 하며, 과도하게 축소된 의미로 해석되지 않아야 한다.

[0039] 본 명세서에서 사용되는 제1, 제2 등과 같이 서수를 포함하는 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.

[0040] 이하, 첨부된 도면을 참조하여 본 명세서에 개시된 실시 예들을 상세히 설명하되, 도면 부호에 관계없이 동일하거나 유사한 구성 요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.

[0041] 또한, 본 명세서에 개시된 기술을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 명세서에 개시된 기술의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 첨부된 도면은 본 명세서에 개시된 기술의 사상을 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 그 기술의 사상이 제한되는 것으로 해석되어서는 아니 됨을 유의해야 한다.

[0043] 도 1은 환경 모델을 이용한 합성 데이터 생성 방법을 설명하는 개념도이다.

[0044] 일반적으로 특정 환경에서 발생되는 데이터는 실제 환경의 특성에 따라 서로 다른 데이터 분포를 가질 수 있다. 본 발명은 실제 환경을 모사한 환경 모델을 이용하여 실제 환경에 대한 특성, 즉 실제 환경에서 데이터의 생성에 영향을 미칠 수 있는 환경의 특성을 분석하여 도출하고, 환경 특성을 기반으로 생성될 합성 데이터의 분포를 실제 데이터가 속할 데이터의 분포에 더욱 가깝게 매칭시킴으로써 고품질의 합성 데이터를 생성할 수 있다. 여기에서 환경 모델은 목적으로 하는 실제 환경에 대한 디지털 트윈, 건물의 모형 또는 빌딩 정보 모델링(Building Information Modeling: BIM) 모델일 수 있다.

[0045] 도 1을 참조하면, 환경 A, 환경 B, 및 환경 C의 세 가지 실제 환경이 있다고 가정한다면, 각 환경에 해당하는 건물에서 생성되는 데이터(예컨대, 송신기와 수신기 사이의 통신 데이터, 화재 발생 시 화염 확상 데이터 등)는 서로 다른 환경적인 특성에 따라 환경 A에서 생성된 데이터 분포(10), 환경 B에서 생성된 데이터 분포(20), 환경 C에서 생성된 데이터 분포(30)와 같이 서로 다른 데이터 분포를 가질 수 있다. 도면부호 40은 환경 A, 환경 B, 및 환경 C에서 생성 가능한 데이터의 확률 분포를 도시한 것으로, 도면부호 10P는 환경 A, 도면부호 20P는 환경 B, 도면부호 30P는 환경 C의 데이터 확률 분포를 나타낸다.

[0046] 데이터는 유사성이 높을수록 서로 가까이 위치하여 분포하게 되는데(예컨대, 잠재 공간(Latent space)에서 샘플 데이터들이 서로 가깝게 위치하는 특성이 있다.), 환경 A, 환경 B, 환경 C가 서로 다른 공간이라는 특성 외에도 공간을 구성하는 물건들의 배치, 벽의 재질, 주변에서 데이터의 성질에 영향을 미치는 요소의 존재 유무 등 다양한 환경적인 특성이 데이터의 생성 결과에 반영되므로, 환경 A의 데이터 분포(10), 환경 B의 데이터 분포(20), 환경 C의 데이터 분포(30)의 동일 분포 그룹 내에서도 더 상세하게 데이터의 분포가 나뉘어질 수 있다. 이들 상세한 데이터의 분포는 예들 들어 환경 A의 데이터 분포(10)의 동일 분포 그룹 내에서도 하나의 덩어리로 뭉쳐져 있는 각각의 구름(10-1, 10-2)의 형태와 같이 표시된다. 이들 구름 형태의 데이터 분포(10-1) 및 구름 형태의 데이터 분포(10-2)는 전술한 잠재 공간(Latent space)이라 불릴 수 있다.

[0047] 여기에 더하여, 실제 환경의 특성을 활용하면 종래의 단순한 시뮬레이션으로 랜덤하게 데이터를 생성하는 방식과 비교하여 환경 모델에 실제 환경의 특성을 반영할 수 있기 때문에 실제 데이터의 데이터 특성(feature)을 쉽게 알아낼 수 있고 이를 통해서 더 고품질의 합성 데이터를 생성할 수 있다. 높은 유사도를 갖는 데이터들은 동일한 데이터 분포에 속한다고 볼 수 있으며, 동일한 데이터 분포에 속하는 데이터들은 유사한 특성을 갖는다고 할 수 있다. 본 명세서에서는 실제 데이터의 유사한 특성에 영향을 미치는 요인을 실제 환경의 특성(환경 특성)이라 하며 환경 모델을 기반으로 실제 환경의 특성을 추출하여 이를 실제 데이터와 더 유사한 데이터의 분포를 찾는데(근사 또는 추정) 활용할 수 있다. 이렇게 도출된 데이터 분포를 활용하여 합성 데이터를 생성하면 실제 환경의 특성을 잘 반영한 고품질의 합성 데이터를 생성할 수 있다.

[0048] 즉, 데이터의 특성은 구조적으로는 동일한 환경이라 할지라도 공간 내에서 데이터 발생 장치의 위치, 특정 시간에 공간을 구성하는 물건의 배치 등에 따라 달라질 수 있으며, 유사한 특성을 지닌 환경에서 수집된 데이터들은 서로 유사한 분포를 가지게 된다. 데이터의 특성에 영향을 미치는 것은 실제 환경적인 특성 외에도 수없이 많기 때문에 랜덤하게 데이터를 생성하는 기존의 합성 데이터 생성 기법들은 실제 데이터 샘플을 활용하여 합성 데이

터를 생성하더라도 합성 데이터의 특성이 실 데이터의 특성을 잘 반영하기 어렵다.

[0049] 예를 들어, 본 명세서에 개시된 합성 데이터 생성 방법은 실제 환경 A에 대한 데이터 생성을 목적으로 할 경우, 다양한 환경 모델들 중, 실제 환경 A에 대응되는 환경 모델 A를 선택한 뒤, 환경 모델 A에서 실제 환경 A와 관련된 환경 특성을 추출하고, 추출된 환경 특성을 환경 모델 A에 대한 합성 데이터 생성 시뮬레이터에 적용하여 데이터를 생성한다. 이 때, 본 방법은 추출된 환경 특성을 이용하여 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수 또는 합성 데이터 생성 로직을 구성한 뒤, 합성 데이터 생성 시뮬레이터에서 목적으로 하는 데이터를 생성한다. 이러한 과정을 거치면 합성 데이터 생성 방법은 환경 A의 데이터 분포(10), 환경 B의 데이터 분포(20), 환경 C의 데이터 분포(30) 중 환경 A의 데이터 분포(10)를 얻을 수 있게 된다. 여기에 더하여, 본 방법은 실제 환경 A 또는 실제 환경 A에 대한 테스트 베드 등에서 샘플링된 실제 데이터(50)를 취득한 뒤, 이를 기초로 환경 A의 데이터 분포(10)에서 생성되는 데이터를 실제 데이터가 생성되는 확률 분포(10-1)에 매칭시킨다. 환경 A의 데이터 확률 분포는 문자 "m"의 모양(60)과 같이 나타나고, 실제 데이터가 존재하는 환경 A의 데이터 확률 분포(10-1)는 정규 확률 분포 곡선(61)과 같이 나타난다. 이렇게 하면, 본 명세서에 의한 합성 데이터 생성 방법은 생성되는 데이터가 실제 데이터가 존재하는 환경 A에 대한 잠재 공간(Latent space)(10-1) 내에서 출력되게 하여 고품질의 합성 데이터(70)를 얻을 수 있게 된다.

[0050] 여기에서 합성 데이터는 단지 몇 개의 실제 데이터 샘플을 가지고 데이터를 생성하는 통계적/베이지안 방법, ML 모델 등의 알고리즘을 통해 실제 환경에서 측정될 수 있는 데이터와 유사한 특성을 갖는 데이터를 의미한다.

[0051] 도 2는 실시 예에 따른 합성 데이터 생성 장치의 구성을 도시하는 블록도이다.

[0052] 도 2를 참조하면, 본 명세서에서 제안하는 방법 또는 기능을 구현하기 위한 합성 데이터 생성 장치(1000)는 제어부(100), 저장부(300), 데이터 전송을 위한 버스(미도시) 또는 외부와 통신을 수행하기 위한 통신부(200), 출력부(400) 및 입력부(500)를 포함할 수 있다. 도시된 구성 요소들은 필수적인 것은 아니어서, 그 보다 많은 구성 요소들을 갖거나, 그보다 적은 구성 요소들을 갖는 합성 데이터 생성 장치(1000)가 구현될 수도 있다. 또한, 이러한 구성 요소는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합을 통해서 구현될 수 있다.

[0053] 제어부(100)는 프로세서(processor)와 같이 데이터를 처리할 수 있는 모든 종류의 프로세싱 장치들을 의미할 수 있다. 제어부(100)는 본 명세서에서 설명되는 다양한 동작들과 단계들을 수행하기 위한 프로세싱 로직을 실행하도록 구성될 수 있다. 여기서, '프로세서'는, 예를 들어 프로그램 내에 포함된 코드 또는 명령으로 표현된 기능을 수행하기 위해 물리적으로 구조화된 회로를 갖는, 하드웨어에 내장된 데이터 처리 장치를 의미할 수 있다. 이와 같이 하드웨어에 내장된 데이터 처리 장치의 일 예로써, 마이크로프로세서(microprocessor), 중앙 처리 장치(central processing unit: CPU), 프로세서 코어(processor core), 멀티 프로세서(multiprocessor), ASIC(application-specific integrated circuit), FPGA(field programmable gate array) 등의 처리 장치를 망라할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

[0054] 통신부(200)는 유선 및/또는 무선 통신 모듈로 구성될 수 있다. 예를 들어, 통신부(200)는, 와이파이(Wireless Fidelity, Wi-Fi), 블루투스(Bluetooth), 지그비(Zigbee), 엔에프씨(near field communication: NFC), 와이브로(Wireless Broadband Internet: Wibro) 등의 무선 통신모듈과 이더넷(Ethernet) 등의 유선 랜(LAN)과 같은 유선 통신 모듈 등을 포함할 수 있다. 통신부(200)는 네트워크를 통해 다른 장치와 유/무선 통신을 수행하여 데이터를 송수신할 수 있다.

[0055] 본 명세서에서 개시되는 네트워크는 예를 들어, 무선 네트워크, 유선 네트워크, 인터넷과 같은 공용 네트워크, 사설 네트워크, 모바일 통신 네트워크용 광역 시스템(Global System for Mobile communication network; GSM) 네트워크, 범용 패킷 무선 네트워크(General Packet Radio Network; GPRN), 근거리 네트워크(Local Area Network; LAN), 광역 네트워크(Wide Area Network; WAN), 거대도시 네트워크(Metropolitan Area Network; MAN), 셀룰러 네트워크, 공중 전화 교환 네트워크(Public Switched Telephone Network; PSTN), 개인 네트워크(Personal Area Network), 블루투스, Wi-Fi 다이렉트(Wi-Fi Direct), 근거리장 통신(Near Field communication), 초광대역(Ultra-Wide band), 이들의 조합, 또는 임의의 다른 네트워크일 수 있지만 이들로 한정되는 것은 아니다.

[0056] 저장부(300)는 자기 저장 매체(magnetic storage media) 또는 플래시 저장 매체(flash storage media)를 포함할 수 있으나, 본 발명의 범위가 이에 한정되는 것은 아니다.

[0057] 출력부(400)는 시각, 청각 또는 촉각 등과 관련된 출력을 발생시키기 위한 것으로, 예를 들어, 디스플레이부

(450)를 포함할 수 있다. 디스플레이부(450)는 터치 센서와 상호 레이어 구조를 이루거나 일체형으로 형성됨으로써, 터치 스크린을 구현할 수 있다. 이러한 터치 스크린은, 장치와 사용자 사이의 입력 인터페이스를 제공하는 사용자 입력부로써 기능함과 동시에, 장치와 사용자 사이의 출력 인터페이스를 제공할 수 있다.

[0058] 입력부(500)는 버스에 연결된 입출력 인터페이스를 통해 제어부(100)에 접속될 수 있다. 상기 입력부(500)는 키보드, 포인팅 장치, 마이크로폰, 조이스틱, 터치패드, 스캐너 등을 포함할 수 있다. 입출력 인터페이스는 직렬 포트 인터페이스, PS/2 인터페이스, 병렬 포트 인터페이스, USB 인터페이스, IEEE 1394 인터페이스와 같은 매우 다양한 인터페이스 중 어느 하나를 포함하거나, 다른 인터페이스의 조합까지도 논리적으로 나타낼 수 있다.

[0059] 제어부(100)는 후술할 합성 데이터 생성 장치(1000)의 기능, 즉 환경 모델을 활용하여 인공지능 모델을 학습시키기 위한 고품질의 합성 데이터를 생성하는 방법의 과정 및 절차들을 수행할 수 있다.

[0060] 도 3은 실시 예에 따른 합성 데이터 생성 장치의 구성을 도시하는 블록도이다.

[0061] 도 3을 참조하면, 학습 데이터 생성 장치(1000)는, 환경 특성 분석부(110), 합성 데이터 생성부(120), 데이터 수집부(130), 및 합성 데이터 품질 평가부(140)를 포함하고, 이들 구성요소는 프로그램으로 구현되어 저장부(300)에 저장되고, 제어부(100)(프로세서)에 의해 실행될 수 있고, 또는 하드웨어와 소프트웨어의 조합에 의해 구현되어 동작할 수 있다. 학습 데이터 생성 장치(1000)는 환경 모델 데이터베이스(350)를 더 포함할 수 있으며, 환경 모델 데이터베이스(350)는 상기 저장부(300)에 저장되거나 외부의 별도의 저장 장치에 저장될 수도 있다. 도시된 구성요소들은 필수적인 것은 아니어서, 그 보다 많은 구성요소들을 갖거나, 그보다 적은 구성요소들을 갖는 학습 데이터 생성 장치(1000)가 구현될 수도 있다. 이러한 구성요소는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합을 통해서 구현될 수 있다.

[0062] 환경 특성 분석부(110)는 사용자가 합성 데이터를 생성하고자 하는 실제 환경을 모사한 환경 모델을 선택하면, 선택된 환경 모델을 환경 모델 데이터베이스(350)로부터 읽어온 후, 실제 환경의 특성을 분석하여 환경 모델에서 데이터의 생성에 영향을 미치는 환경 특성을 추출한다. 환경 특성 분석부(110)는 실제 환경과 합성 데이터의 응용 분야 사이의 연관 관계에 기초하여 환경 특성을 추출할 수 있다.

[0063] 환경 모델의 환경 특성은 데이터가 생성되는 실제 환경 및 생성되는 합성 데이터의 응용 분야에 따라서 서로 다를 수 있다. 예를 들어, 건물에서 통신 분야의 합성 데이터를 생성한다고 하면, 건물 내의 금속 물품의 유무, 공간 내 안테나의 위치 등은 중요한 환경 특성이 되지만, 건물 내 공기의 흐름 등은 상대적으로 중요한 환경 특성이 아니게 된다. 반면, 건물에서 화재 분야의 합성 데이터를 생성한다고 하면, 건물 내에서 공기의 흐름에 관련된 데이터는 통신 분야에서의 중요 환경 특성인 안테나의 위치에 비하여 매우 중요한 환경 특성이 된다.

[0064] 환경 특성 분석부(110)는 환경 특성을 추출함에 있어 실제 환경과 합성 데이터의 응용 분야 사이의 연관 관계로서 온톨로지(Ontology)를 활용할 수 있다. 환경 특성 분석부(110)는 온톨로지를 활용하여 환경 모델이 포함하고 있는 정보들과 합성 데이터를 생성하고자 하는 응용 분야와의 연관성을 도출할 수 있으며, 도출된 연관성을 이용하여 합성 데이터를 생성하고자 하는 응용 분야에 적합한 환경 특성을 쉽게 추출할 수 있다.

[0065] 도 4는 실제 환경에서 데이터의 생성에 영향을 미치는 환경 특성들의 관계를 나타내는 예를 도시한 도면이다.

[0066] 도 4를 참조하면, 건물이라는 실제 환경에서 데이터 생성에 영상을 미치는 환경 특성은 건물의 유형, 건물의 구조 및 건물의 재질 등으로 구분될 수 있으며, 건물의 유형은 다시 스마트 홈, 공장, 오피스 등으로 구분될 수 있고, 건물의 구조는 다시 창문 위치, 배관, 안테나 위치, 건물 층고, 건물 내 공기 흐름 여부 등으로 구분될 수 있고, 건물의 재질은 세부적으로 시설(건물) 내의 금속 물품 유무, 외벽의 재질 등으로 구분될 수 있다.

[0067] 이하 도 5 내지 도 7을 참조하여, 환경 특성의 추출에 사용되는 온톨로지의 예들을 설명한다.

[0068] 도 5는 실제 환경과 통신 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

[0069] 도 5를 참조하면, 합성 데이터의 응용 분야(501) 중 통신 분야의 데이터 생성에 영향을 미치는 건물의 환경 특성은 건물의 '구조'에서는 '안테나 위치', 건물의 '재질'에서는 '시설 내 금속 물품의 유무' 및 '외벽 재질'과 연관 관계가 있으므로, 환경 특성 분석부(110)는 통신 분야의 합성 데이터를 생성할 때의 환경 특성으로 '안테나 위치', '시설 내 금속 물품의 유무' 및 '외벽 재질'로 추출할 수 있다.

[0070] 도 6은 실제 환경과 공조 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

[0071] 도 6을 참조하면, 합성 데이터의 응용 분야(601) 중 공조(air conditioning) 분야의 데이터 생성에 영향을 미치는 건물의 환경 특성은 건물의 '구조'에서는 '창문의 위치', '배관', '건물 층고' 및 '건물 내 공기의 흐름',

건물의 '재질'에서는 '외벽 재질'과 연관 관계가 있으므로, 환경 특성 분석부(110)는 공조 분야의 합성 데이터를 생성할 때의 환경 특성으로 '창문의 위치', '배관', '건물 층고' 및 '외벽 재질'로 추출할 수 있다.

[0072] 도 7은 실제 환경과 화재 분야 사이의 연관 관계를 나타내는 온톨로지의 예를 도시한 도면이다.

[0073] 도 7을 참조하면, 합성 데이터의 응용 분야(701) 중 화재 분야의 데이터 생성에 영향을 미치는 건물의 환경 특성은 건물의 '구조'에서는 '창문의 위치' 및 '건물 내 공기의 흐름', 건물의 '재질'에서는 '외벽 재질'과 연관 관계가 있으므로, 환경 특성 분석부(110)는 화재 분야의 합성 데이터를 생성할 때의 환경 특성으로 '창문의 위치', '건물 내 공기의 흐름' 및 '외벽 재질'로 추출할 수 있다.

[0074] 실제 환경에 대한 특성들 중 합성 데이터의 응용 분야에 연관된 정보를 환경 모델에서 도출하기 위해서는 실제 환경을 분석하여 어떤 정보들이 응용 분야에 연관되는가를 판단해야 한다. 연관 정보의 판단은 응용 분야 전문가에 의해 수동으로 도출된 것을 환경 특성 분석부(110)가 사용할 수도 있고 인공지능이 수행할 수도 있다. 응용 분야 전문가는 자신의 경험/지식을 바탕으로 생성하고자 하는 데이터와 연관성이 큰 실제 환경의 환경 특성들을 직접 환경 모델에서 도출할 수 있다.

[0075] 전술한 연관 정보의 도출 과정은 수학 모델 또는 인공지능 모델을 통해 자동으로 수행될 수 있다. 예를 들어, 합성 데이터 생성 장치(1000)가 기본적인 연관 정보로 작성된 온톨로지 그래프와 환경 모델을 활용하여 "환경 모델 정보 중 랜덤으로 환경 특성을 선택하고, 선택된 특성을 기반으로 환성 데이터를 생성하고, 생성된 합성 데이터의 품질을 평가"하는 과정을 반복하면서, 특정 환경 특성이 이전보다 더 좋은 품질의 합성 데이터를 생성할 경우, 환경 특성 분석부(110)가 해당 환경 특성의 연관 관계(correlation)를 수학 모델 또는 인공지능 모델에 반영하거나 또는 학습하도록 할 수 있다. 반영 또는 학습된 특성 정보는 온톨로지 그래프에서 링크의 강도 또는 링크의 생성으로 볼 수 있는데 이러한 강도의 변화를 통해 환경 모델에서 특정 응용 분야에 대하여 연관성이 높은 환경 특성을 도출할 수 있다.

[0076] 도 8은 실제 환경과 화재 분야 사이의 연관 관계를 나타내는 온톨로지에서 연관 관계가 추가되는 예를 도시한 도면이다.

[0077] 도 8을 참조하면, 도 7에서 화재 분야의 합성 데이터를 생성할 때의 환경 특성 '창문의 위치', '건물 내 공기의 흐름' 및 '외벽 재질'에 응용 분야 '화재'와 환경 특성 '배관' 사이의 연관 관계(801)를 추가한 것으로, 환경 특성 분석부(110)는 다양한 환경 특성의 조합을 합성 데이터 생성부(120)에 넘겨주어 합성 데이터 생성부(120)로 하여금 이들을 합성 데이터 생성 시뮬레이터에 반영하여 합성 데이터를 생성하도록 할 수 있다. 이러한 과정으로 생성된 합성 데이터의 품질이 이전의 다른 환경 특성들에 따른 합성 데이터의 품질보다 양호하다면, 환경 특성 분석부(110)는 다음의 환경 특성 추출 과정부터는 건물 환경에서 화재 분야의 합성 데이터를 생성하고자 한다면, 환경 특성으로 '배관'을 추가한다.

[0078] 다시, 도 3을 참조하면, 합성 데이터 생성부(120)는 생성될 합성 데이터가 환경 특성 분석부(110)가 추출한 환경 특성을 반영하여 생성되도록 선택된 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수의 구성을 변경(modification)한 뒤, 상기 합성 데이터 생성 시뮬레이터를 이용하여 합성 데이터를 생성할 수 있다. 여기에서, 합성 데이터 생성부(120)는 상기 합성 데이터 생성 함수 또는 합성 데이터 생성 로직의 파라미터들 중에서 환경 특성 별로 대응되는 파라미터를 상기 추출된 환경 특성에 따라 수정하여 합성 데이터 생성 함수의 구성을 변경(modification)할 수 있다. 합성 데이터를 생성하기 위한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수 또는 합성 데이터 생성 로직은 환경 모델의 환경적 특성들을 반영한 데이터가 생성되도록 구성되어 있다. 특히 합성 데이터 생성 함수 또는 합성 데이터 생성 로직의 파라미터 별로 고유한 환경 특성이 대응되도록 구성되어 있으므로, 합성 데이터 생성부(120)는 환경 특성 분석부(110)에 의해 추출된 환경 특성에 대응되는 합성 데이터 생성 함수 또는 합성 데이터 생성 로직의 파라미터를 환경 특성에 맞게 조절함으로써, 실제 환경에 더 가까운 합성 데이터를 생성할 수 있게 된다.

[0079] 여기에 더하여, 합성 데이터 생성부(120)는 목적 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 실제 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시킬 수 있다.

[0080] 데이터 수집부(130)는 실제 환경에서 샘플링하여 수집된 실제 데이터를 저장하고 있으며, 실제 환경 특성 분석부(110)의 요청에 의해 이를 합성 데이터 생성부(120)에 전달한다.

[0081] 전술한 합성 데이터의 데이터 특성을 실제 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 방법을 상세히 설명하면, 합성 데이터 생성부(120)는 먼저 데이터 수집부(130)를 통해 수신한 실제 데이터에 기초하여 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 실제 데이터가 속하는 데

이터 분포 영역으로 결정한다. 다음으로, 합성 데이터 생성부(120)는 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 상기 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 상기 합성 데이터 생성 시뮬레이터를 설정한다. 즉, 합성 데이터 생성부(120)는 실제 데이터가 존재할 잠재 공간(latent space)에서만 합성 데이터가 생성되도록 함 상기 합성 데이터 생성 시뮬레이터를 설정한다.

[0082] 합성 데이터를 생성하는데 있어서 합성 데이터 생성 시뮬레이터는 GAN(Generative Adversarial Networks) 및 VAE(Variational Auto-Encoder)와 같은 생성 모델(Generative model)을 활용할 수 있다. 상기 생성 모델은 데이터의 분포(Distribution)를 학습하여 주어진 데이터가 어떤 분포에 속할 확률이 높은 지를 결정(즉, 실제 데이터 분포와 합성 데이터 분포를 매칭하는 것)하는 모델이라 할 수 있다. GAN은 생성자(Generator)와 판별자(Discriminator)로 구성되는데, 생성자는 판별자가 페이크 데이터(Fake date)(합성 데이터에 해당함)를 식별할 수 없는 확률을 높이는 방향으로 학습하며, 이와 반대로 판별자는 생성자의 페이크 데이터를 실제 데이터와 비교해 이를 식별해 내는 확률을 높이는 방향으로 학습한다. 다시 말해, 생성자는 판별자를 속이기 위해 노력하며, 판별자는 생성자에게 속지 않기 위해 노력한다. 이러한 과정을 통해서 생성 모델로서의 GAN은 실제 데이터의 분포와 합성 데이터의 분포를 매칭시킬 수 있다. VAE(Variational Auto-Encoder)는 잠재 공간(Latent Space)의 확률 분포에 대한 평균과 표준 편차를 학습하여 입력이 만들어지는 확률 분포를 찾는 것이 목적이다. 즉, 합성 데이터를 생성할 잠재 공간의 분포를 학습함으로써 합성 데이터의 분포와 실제 데이터의 분포를 매칭시킬 수 있다.

[0083] 실시 예들은 합성 데이터의 데이터 특성을 실제 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시키는 방법으로서 다양한 생성 모델(GAN 및 VAE로만 제약되지 않음)을 사용할 수 있으며, 생성 모델이 실제 데이터의 분포와 합성 데이터의 분포를 보다 더 잘 매칭시킴으로써 고품질의 합성 데이터를 생성할 수 있다.

[0084] 합성 데이터 품질 평가부(140)는 합성 데이터 생성부(120)에서 생성된 합성 데이터의 품질을 평가하고, 합성 데이터의 품질이 미리 결정된 기준에 미치지 못할 경우, 환경 특성 분석부(110)로 하여금 실제 환경 특성을 다시 분석하게 하여 환경 특성을 추출하도록 한다. 이 때, 환경 특성 분석부(110)는 다시 추출된 환경 특성을 기초로 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수 또는 합성 데이터 생성 로직을 다시 구성한 뒤, 합성 데이터 생성 시뮬레이터에서 목적으로 하는 데이터를 다시 생성하도록 한다.

[0085] 합성 데이터의 품질 평가는 생성된 합성 데이터 세트와 원본 데이터 세트의 직접적인 비교(통계 값 비교, 데이터 분포 비교, 연관성(correlation) 분석 등)를 통해 평가하는 직접 평가 방법과, 생성된 합성 데이터를 기반으로 인공지능 모델을 학습시켜, 인공지능 모델의 성능이 기준치를 달성할 수 있는지 평가하는 간접 평가 방법을 사용할 수 있다.

[0086] 이하에서는 다양한 실시 예들을 통해 본 발명을 상세히 설명한다.

[0087] **실시 예 1: 통신 스케쥴링 AI 모델의 학습을 위한 통신 데이터 생성**

[0088] 본 실시 예에서는 본 발명의 이해를 돕기 위해 학습용 무선 통신 데이터를 생성하는 데에 전술한 학습 데이터 생성 장치 및 방법을 사용하는 것을 예로 들어 설명한다. 특별히, 본 실시 예에서는 무선 통신을 수행하는 장치들이 생성하는 통신 파형을 합성 데이터로 생성하는 예를 설명한다.

[0089] 이 실시 예에서 생성된 합성 데이터는 인공지능 모델이 최적의 통신 스케쥴링을 제시할 수 있도록 특정 환경에서 발생되는 통신 데이터의 예시로 사용될 수 있다.

[0090] 통신 파형에 대한 실제 데이터는 파형이 송신 또는 수신되는 실제 환경에 따라 서로 다른 데이터 분포를 가질 수 있다. 도 1을 참조하면, 환경 A를 스마트 홈, 환경 B를 스마트 오피스, 환경 C를 스마트 팩토리로 가정한다면, 각 환경에 해당하는 건물에서 전송되는 무선 통신 데이터들은 서로 다른 환경적인 특성에 따라 환경 A의 데이터 분포(10), 환경 B의 데이터 분포(20), 환경 C의 데이터 분포(30)와 같이 서로 다른 데이터 분포를 가질 수 있다.

[0091] 학습 데이터 생성 장치(1000)는 환경 A, 환경 B, 환경 C 중 데이터를 얻고자 하는 환경에 대응되는 환경 모델을 선택한 뒤, 환경 모델을 기반으로 대응되는 실제 환경을 분석하여 공간을 구성하는 물건들의 배치, 배치된 물건들의 전자파 흡수 정도, 벽의 재질, 주변에 통신 노드들의 존재 유무 등 다양한 환경적인 특성을 도출한다. 이때, 학습 데이터 생성 장치(1000)는 전술한 통신 분야와 건물의 환경 특성과의 연관 관계를 나타낸 온톨로지에 기초하여 환경 특성들을 추출한다.

[0092] 학습 데이터 생성 장치(1000)는 생성될 합성 데이터가 추출된 환경 특성을 반영하여 생성되도록, 선택된 환경

모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수의 구성을 변경(modification)한 뒤, 상기 합성 데이터 생성 시뮬레이터를 이용하여 합성 데이터를 생성한다. 한편, 학습 데이터 생성 장치(1000)는 실제 환경에서 샘플링된 소수의 통신 데이터를 취득한 뒤, 이를 이용하여 생성되는 합성 데이터의 분포가 상기 실제 데이터의 잠재 공간(latent space)에 더 가깝도록 매칭한다. 마지막으로, 학습 데이터 생성 장치(1000)는 통신 데이터 생성용 시뮬레이터를 통해 도출된 데이터 특성에 따라 고품질의 합성 통신 데이터를 생성한다.

[0093] **실시 예 2: 화재 감지 센서에 탑재될 AI 모델의 학습을 위한 화재 데이터 생성**

[0094] 건물에 대한 화재 데이터를 수집함에 있어서, 예를 들어, 동일한 아파트일지라도, 아파트 평면 타입의 종류, 벽재의 소재, 가구의 소재 및 위치 등에 따라 화재의 확산 정도 및 범위가 다를 수 있다. 화재 데이터의 경우, 법적으로 방화가 금지되며, 방화 실험 비용 또한 매우 높아서 실제 화재 데이터를 수집하기 어렵다.

[0095] 따라서, 실시 예에서는 화재 데이터를 얻고자 하는 실제 환경에 대한 환경 모델 및 해당 환경 모델에서의 화재 데이터 밸생 시뮬레이터를 이용하여 합성 화재 데이터를 생성한다.

[0096] 먼저, 학습 데이터 생성 장치(1000)는 화재 데이터 생성을 목적으로 하는 실제 환경에 대한 모델(예, 3차원 건물 설계도, 벽면의 재료 정보, 가구 배치 정보 등을 포함하는 Building Information Modeling 모델)을 선택하고 환경 모델 데이터베이스에서 읽어 온다.

[0097] 학습 데이터 생성 장치(1000)는 선택된 환경 모델을 기반으로 실제 환경에 대한 특성을 분석 및 도출한다. 학습 데이터 생성 장치(1000)는 선택된 환경 모델에서 화재 데이터의 생성에 영향을 주거나 변화를 주는 요인을 분석하여 이를 추출한다. 해당하는 환경 특성은 아파트 평면의 타입, 창문의 위치, 벽의 소재, 건물의 층고 등이 될 수 있다. 이 때, 학습 데이터 생성 장치(1000)는 전술한 화재 분야와 건물의 환경 특성과의 연관 관계를 나타낸 온톨로지에 기초하여 환경 특성들을 추출한다.

[0098] 학습 데이터 생성 장치(1000)는 생성될 합성 데이터가 추출된 환경 특성을 반영하여 생성되도록, 선택된 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수의 구성을 변경(modification)한 뒤, 상기 합성 데이터 생성 시뮬레이터를 이용하여 합성 데이터를 생성한다. 이 때, 환경 모델에서 도출된 특성을 데이터의 생성에 이용할 수 있도록 특성 정보를 가공한다. 즉, 실제 환경에서 도출된 특성을 환경 모델에 이용할 수 있도록 특성 정보를 가공한다.

[0099] 학습 데이터 생성 장치(1000)는 실제 환경에서 샘플링된 소수의 화재 데이터를 이용하여 생성될 합성 데이터의 분포가 상기 실제 데이터가 속할 데이터 분포(잠재 공간(latent space))에 더욱 가깝게 되도록 매칭시킨다. 본 실시 예에서는, 화재 데이터 생성을 목적으로 하는 실제 환경에 대한 실제 데이터를 샘플링하기 어려우므로, 실제 환경과 유사한 환경에서 수집되어 공개된 실데이터(예, 미국 국립표준기술연구소(National Institute of Standards and Technology)의 실제 화재 데이터)를 활용할 수 있다.

[0100] 마지막으로, 학습 데이터 생성 장치(1000)는 Fire Dynamics Simulator(FDS)와 같은 화재 시뮬레이터를 이용하여 고품질의 합성 데이터를 생성한다.

[0101] **실시 예 3: 건물 공기조화 시스템에 탑재될 AI 모델을 위한 고품질 데이터 생성**

[0102] 공기조화시스템(Heating, Ventilation, & Air Conditioning: HVAC)은 난방, 환기, 냉방을 통합하여 공기조화의 4대요소(온도, 습도, 기류, 청정도)를 실내 또는 자동차의 환경에 알맞은 상태로 조정하여 쾌적한 환경을 조성하는 것이 주 목적이다. 동일한 구조의 실내 공간일지라도 공조기의 위치, 창문의 위치 및 두께, 가구 배치 등에 따라 전혀 다른 특성의 온도, 습도, 기류, 청정도의 변화를 보일 수 있다.

[0103] 공기조화시스템의 실시 예에서는 온도, 습도, 기류, 청정도 4대 요소에 대한 서로 다른 데이터를 생성하는 경우이며, 데이터들은 서로 간에 상관 관계(Correlation)가 있을 수 있다.

[0104] 먼저, 학습 데이터 생성 장치(1000)는 데이터 생성을 목적으로 하는 실제 환경에 대한 모델(예, BIM(Building Information Modeling) 모델 또는 디지털 트윈 모델)을 선택하고 환경 모델 데이터베이스에서 읽어들인다. 환경 모델로서 디지털 트윈 모델은 건물 및 공기조화시스템의 노후 정도를 반영할 수 있으므로, 공기조화기의 노후 정도에 따라 가동 시 실제 환경의 온도, 습도, 기류, 청정도가 다르게 변하는 데이터를 획득할 수 있다. 또한, 환경 모델로서 디지털 트윈 모델은 건물 내 기류의 흐름 정보를 가지고 있으므로 공기조화기의 작동에 따른 정확한 기류 데이터를 획득할 수 있다.

[0105] 다음으로, 학습 데이터 생성 장치(1000)는 데이터 생성을 목적으로 하는 실제 환경에 대한 실 데이터를 샘플링

한다. 실 데이터를 샘플링 할 때 비용 등의 이유로 4대 요소에 대한 4가지 데이터 중 측정 가능한 데이터 유형의 수가 상이할 수 있다. 측정 가능한 데이터 유형의 수(온도, 습도, 기류, 청정도)가 많을수록 데이터 유형 간 상관관계를 잘 반영하기 때문에 보다 고품질의 합성 데이터를 생성할 수 있다.

[0106] 학습 데이터 생성 장치(1000)는 환경 모델을 기반으로 실제 환경에 대한 특성을 분석 및 도출한다. 학습 데이터 생성 장치(1000)는 특정 환경 모델에서 공기조화시스템을 위한 데이터 생성에 영향 또는 변화를 주는 요인(예, 공조기의 위치, 창문 위치 및 두께, 건물 외벽의 모양, 건물 내 기류 흐름 정보, 공조기의 노후 정도 등)을 분석하여 추출한 다음, 고품질 합성 데이터 생성 시 환경 모델에서 도출된 특성을 이용할 수 있도록 특성 정보를 가공한다. 이 때, 학습 데이터 생성 장치(1000)는 전술한 공조 분야와 건물의 환경 특성과의 연관 관계를 나타낸 온톨로지에 기초하여 환경 특성들을 추출한다.

[0107] 학습 데이터 생성 장치(1000)는 생성될 합성 데이터가 추출된 환경 특성을 반영하여 생성되도록, 선택된 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수의 구성을 변경(modification)한 뒤, 상기 합성 데이터 생성 시뮬레이터를 이용하여 합성 데이터를 생성한다. 한편, 학습 데이터 생성 장치(1000)는 실제 환경에서 샘플링된 소수의 공조 데이터를 취득한 뒤, 이를 이용하여 생성되는 합성 데이터의 분포가상기 실제 데이터가 속하는 데이터 분포(잠재 공간(latent space))에 더욱 가깝도록 매칭시켜 고품질의 합성 데이터를 생성한다.

[0108] 이하에서는 전술한 학습 데이터 생성 장치(1000)를 이용하여 합성 데이터 생성 방법을 상세하게 설명한다.

[0109] 도 9는 실시 예들에 따른 합성 데이터 생성 방법을 설명하는 순서도이다.

[0110] 도 3 및 9를 참조하면, 학습 데이터 생성 장치(1000)에 의한 학습 데이터 생성 방법은, 먼저, 사용자가 데이터 생성을 목적으로 하는 실제 환경에 대한 환경 모델, 즉 실제 환경을 모사한 환경 모델을 선택하면, 선택된 환경 모델을 환경 모델 데이터베이스(350)로부터 읽어 온다(S910).

[0111] 다음으로, 학습 데이터 생성 장치(1000)는 선택된 환경 모델을 기반으로 특정 공간이 가지고 있는 데이터 생성에 영향을 미치는 환경 특성 같은 실제 환경에 대한 특성을 분석하여 특성 정보를 추출한다(S920).

[0112] 학습 데이터 생성 장치(1000)는 생성되는 합성 데이터가 상기 환경 특성을 반영하도록 특성 정보를 가공하여 환경 모델에 대한 합성 데이터 생성 시뮬레이터의 합성 데이터 생성 함수를 구성한다(S930).

[0113] 학습 데이터 생성 장치(1000)는 상기 환경 모델에 대한 합성 데이터 생성 시뮬레이터를 이용하여 고품질의 합성 데이터를 생성한다(S940).

[0114] 마지막으로, 학습 데이터 생성 장치(1000)는 생성된 합성 데이터의 품질을 평가한 뒤, 생성된 합성 데이터가 요구되는 데이터 품질을 만족하는 경우에는 합성 데이터의 생성을 종료하며, 만족하지 못할 경우에는 특성 정보를 추출하는 과정(S920)부터 다시 수행한다(S950).

[0115] 여기에 더하여, 학습 데이터 생성 장치(1000)는 더 고품질의 합성 데이터를 생성하기 위하여, 상기 환경 모델에 대한 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 특성을 상기 실제 환경에서 발생되는 실제 데이터가 속하는 데이터 특성에 매칭시킬 수 있다.

[0116] 상기 데이터 특성의 매칭 과정은, 실제 환경에서 획득한 실제 데이터를 수신하는 과정, 수신된 실제 데이터에 기초하여 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터의 데이터 분포 영역을 실제 데이터가 속하는 데이터 분포 영역으로 결정하는 과정 및 합성 데이터 생성 시뮬레이터에서 생성되는 합성 데이터 중 실제 데이터가 속하는 데이터 분포 영역에 포함되는 합성 데이터만 출력되도록 합성 데이터 생성 시뮬레이터를 설정하는 과정을 포함할 수 있다. 이 때, 학습 데이터 생성 장치(1000)는 데이터 수집부(130)를 통해 데이터 생성을 목적으로 하는 실제 환경 또는 실제 환경에 대한 테스트 베드에서 실제 데이터를 샘플링하여 취득한다.

[0117] 상술한 설명에서, 단계들, 과정들 또는 동작들은 본 발명의 구현예에 따라서, 추가적인 단계, 과정 또는 동작으로 더 분할되거나, 더 적은 단계, 과정 또는 동작으로 조합될 수 있다. 또한, 일부 단계, 과정 또는 동작은 필요에 따라 생략될 수도 있고, 단계 또는 동작 간의 순서가 전환될 수도 있다. 또한, 전술한 학습 데이터 생성 방법이 포함하는 각 단계 또는 동작은 컴퓨터프로그램으로 구현되어 컴퓨터로 읽을 수 있는 기록매체에 저장될 수 있으며, 컴퓨터 장치에 의해 각 단계, 과정 또는 동작이 실행될 수도 있다.

[0119] 본 명세서에서 사용된 용어 "부"는(예를 들면, 제어부 등), 예를 들어, 하드웨어, 소프트웨어 또는 펌웨어

(firmware) 중 하나 또는 둘 이상의 조합을 포함하는 단위(unit)를 의미할 수 있다. "부"는, 예를 들어, 유닛(unit), 로직(logic), 논리블록(logical block), 부품(component), 또는 회로(circuit) 등의 용어와 바꾸어 사용(interchangeably use)될 수 있다. "부"는, 일체로 구성된 부품의 최소 단위 또는 그 일부가 될 수 있다. "부"는 하나 또는 그 이상의 기능을 수행하는 최소 단위 또는 그 일부가 될 수도 있다. "부"는 기계적으로 또는 전자적으로 구현될 수 있다. 예를 들어, "부"는, 알려졌거나 앞으로 개발될, 어떤 동작들을 수행하는 ASIC(Application-Specific Integrated Circuit) 칩, FPGAs(Field-Programmable Gate Arrays) 또는 프로그램 가능 논리 장치(programmable-logic device) 중 적어도 하나를 포함할 수 있다.

[0120] 다양한 실시예에 따른 장치(예: 모듈들 또는 그 기능들) 또는 방법(예: 동작들)의 적어도 일부는, 예컨대, 프로그램 모듈의 형태로 컴퓨터로 읽을 수 있는 저장매체(computer-readable storage media)에 저장된 명령어로 구현될 수 있다. 상기 명령어가 프로세서에 의해 실행될 경우, 상기 하나 이상의 프로세서가 상기 명령어에 해당하는 기능을 수행할 수 있다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터로 읽을 수 있는 저장매체/컴퓨터로 판독 가능한 기록 매체는, 하드디스크, 플로피디스크, 마그네틱 매체(magnetic media)(예: 자기테이프), 광기록 매체(optical media)(예: CD-ROM(compact disc read only memory), DVD(digital versatile disc), 자기-광 매체(magneto-optical media)(예: 플롭티컬 디스크(floptical disk)), 하드웨어 장치(예: ROM(read only memory), RAM(random access memory), 또는 플래시 메모리 등) 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함할 수 있다. 또한, 프로그램 명령에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함할 수 있다. 상술한 하드웨어 장치는 다양한 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지다.

[0121] 다양한 실시예에 따른 모듈 또는 프로그램 모듈은 전술된 구성요소들 중 적어도 하나 이상을 포함하거나, 일부가 생략되거나, 또는 추가적인 다른 구성요소를 더 포함할 수 있다. 다양한 실시예에 따른 모듈, 프로그램 모듈 또는 다른 구성요소에 의해 수행되는 동작들은 순차적, 병렬적, 반복적 또는 휴리스틱(heuristic)한 방법으로 실행될 수 있다. 또한, 일부 동작은 다른 순서로 실행되거나, 생략되거나, 또는 다른 동작이 추가될 수 있다.

[0122] 본 명세서에 사용된 용어 "하나"는 하나 또는 하나 이상으로 정의된다. 또한, 청구 범위에서 "적어도 하나" 및 "하나 이상"과 같은 도입 문구를 사용하는 것은, 동일한 청구항에 "적어도 하나" 및 "하나 이상"과 같은 도입 문구 및 "하나" 같은 불명료한 문구가 포함되어 있는 경우라 할지라도, 불명료한 문구 "하나"에 의한 다른 청구항 요소의 도입이 그러한 요소를 하나만을 포함하는 발명에 대해 그렇게 도입된 청구항 요소를 포함하는 임의의 특정 청구항을 제한한다는 것을 의미하는 것으로 해석되어서는 안된다.

[0123] 본 문서에서, "A 또는 B" 또는 "A 및/또는 B 중 적어도 하나" 등의 표현은 함께 나열된 항목들의 모든 가능한 조합을 포함할 수 있다.

[0124] 동일한 기능을 달성하기 위한 구성 요소의 배열은 효과적으로 "관련"되어 원하는 기능이 달성된다. 따라서, 특정 기능성을 달성하기 위해 결합된 임의의 2 개의 구성 요소는 구조 또는 중개하는 구성 요소와 관계없이 원하는 기능이 달성되도록 서로 "관련"되는 것으로 간주될 수 있다. 마찬가지로 이와 같이 연관된 두 개의 구성 요소는 원하는 기능을 달성하기 위해 서로 "작동 가능하게 연결"되거나 "작동 가능하게 결합된" 것으로 간주될 수 있다.

[0125] 또한, 통상의 기술자는 전술한 동작들의 기능성 사이의 경계가 단지 예시적인 것임을 인식할 것이다. 복수의 동작들은 단일 동작으로 결합될 수 있고, 단일 동작은 추가 동작들로 분산될 수 있으며, 동작들은 시간적으로 적어도 부분적으로 겹쳐서 실행될 수 있다. 또한, 대안적인 실시예들은 특정 동작에 대한 복수의 인스턴스들을 포함할 수 있고, 동작들의 순서는 다양한 다른 실시예에서 변경될 수 있다. 그러나, 다른 수정, 변형 및 대안이 또한 가능하다. 따라서, 상세한 설명 및 도면은 제한적인 의미가 아니라 예시적인 것으로 간주되어야 한다.

[0126] "X일 수 있다"는 문구는 조건 X가 충족될 수 있음을 나타낸다. 이 문구는 또한 조건 X가 충족되지 않을 수도 있음을 나타낸다. 예를 들어, 특정 구성 요소를 포함하는 시스템에 대한 참조는 시스템이 특정 구성 요소를 포함하지 않는 시나리오도 포함해야 한다. 예를 들어, 특정 동작을 포함하는 방법에 대한 참조는 해당 방법이 특정 구성 요소를 포함하지 않는 시나리오도 포함해야 한다. 그러나 또 다른 예를 들면, 특정 동작을 수행하도록 구성된 시스템에 대한 참조는 시스템이 특정 작업을 수행하도록 구성되지 않은 시나리오도 포함해야 한다.

[0127] 용어 "포함하는", "갖는", "구성된", "이루어진" 및 "본질적으로 이루어진"은 상호 교환적으로 사용된다. 예를

들어, 임의의 방법은 적어도 도면 및/또는 명세서에 포함된 동작을 포함할 수 있으며, 도면 및/또는 명세서에 포함된 동작만을 포함할 수 있다. 또는, "포함하는"이라는 단어는 청구항에 나열된 요소들 또는 동작들의 존재를 배제하지 않는다.

[0128] 통상의 기술자는 논리 블록들 사이의 경계가 단지 예시적인 것이며, 대안적인 실시 예들이 논리 블록들 또는 회로 소자들을 병합하거나 또는 다양한 논리 블록들 또는 회로 소자들 상에 기능의 대체적인 분해를 부과할 수 있음을 인식할 것이다. 따라서, 여기에 도시된 아키텍처는 단지 예시적인 것이며, 사실 동일한 기능을 달성하는 많은 다른 아키텍처가 구현될 수 있다는 것으로 이해되어야 한다.

[0129] 또한, 예를 들어, 일 실시예에서, 도시된 예들은 단일 집적 회로 상에 또는 동일한 장치 내에 위치된 회로로서 구현될 수 있다. 대안적으로, 상기 예들은 임의의 수의 개별적인 집적 회로들 또는 적합한 방식으로 서로 상호접속된 개별 장치들로서 구현될 수 있으며, 다른 변경, 수정, 변형 및 대안들이 또한 가능하다. 따라서, 명세서 및 도면은 제한적인 의미가 아니라 예시적인 것으로 간주되어야 한다.

[0130] 또한, 예를 들어, 전술한 예들 또는 그 일부는, 임의의 적절한 유형의 하드웨어 기술 언어와 같은, 물리적 회로 또는 물리적 회로로 변환 가능한 논리적 표현의 소프트웨어 또는 코드 표현으로서 구현될 수 있다.

[0131] 또한, 본 발명은 비 프로그래머블 하드웨어로 구현된 물리적 장치 또는 유닛으로 제한되지 않지만, 일반적으로 본원에서는 '컴퓨터 시스템'으로 표시되는 메인 프레임, 미니 컴퓨터, 서버, 워크스테이션, 개인용 컴퓨터, 노트패드(notepad), 개인용 디지털 정보 단말기(PDA), 전자 게임(electronic games), 자동차 및 기타 임베디드 시스템, 휴대전화 및 다양한 다른 무선 장치 등과 같은, 적절한 프로그램 코드에 따라 동작함으로써 원하는 장치 기능을 수행할 수 있는 프로그램 가능한 장치 또는 유닛에도 적용될 수 있다.

[0132] 이 명세서에 언급된 시스템, 장치 또는 디바이스는 적어도 하나의 하드웨어 구성 요소를 포함한다.

[0133] 이상에서 본 명세서의 기술에 대한 바람직한 실시 예가 첨부된 도면들을 참조하여 설명되었다. 여기서, 본 명세서 및 청구 범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니되며, 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야 한다. 본 발명의 범위는 본 명세서에 개시된 실시 예들로 한정되지 아니하고, 본 발명은 본 발명의 사상 및 특허청구범위에 기재된 범주 내에서 다양한 형태로 수정, 변경, 또는 개선될 수 있다.

부호의 설명

[0134] 10: 환경 A에서 생성되는 데이터 분포

10-1, 10-2: 환경 A에서 발생되는 데이터 분포의 잠재 공간

20: 환경 B에서 생성되는 데이터 분포

30: 환경 C에서 생성되는 데이터 분포

10P, 20P, 30P, 40: 환경 A, 환경 B, 환경 C의 데이터 확률 분포

50: 실제 데이터

60: 환경 A의 데이터 확률 분포 곡선

61: 환경 A에서 실제 데이터가 발생할 확률 분포 곡선

70: 고품질의 합성 데이터

100: 제어부

110: 환경 특성 분석부

120: 합성 데이터 생성부

130: 데이터 수집부

140: 합성 데이터 품질 평가부

200: 통신부

300: 저장부

350: 환경 모델 데이터베이스

400: 출력부

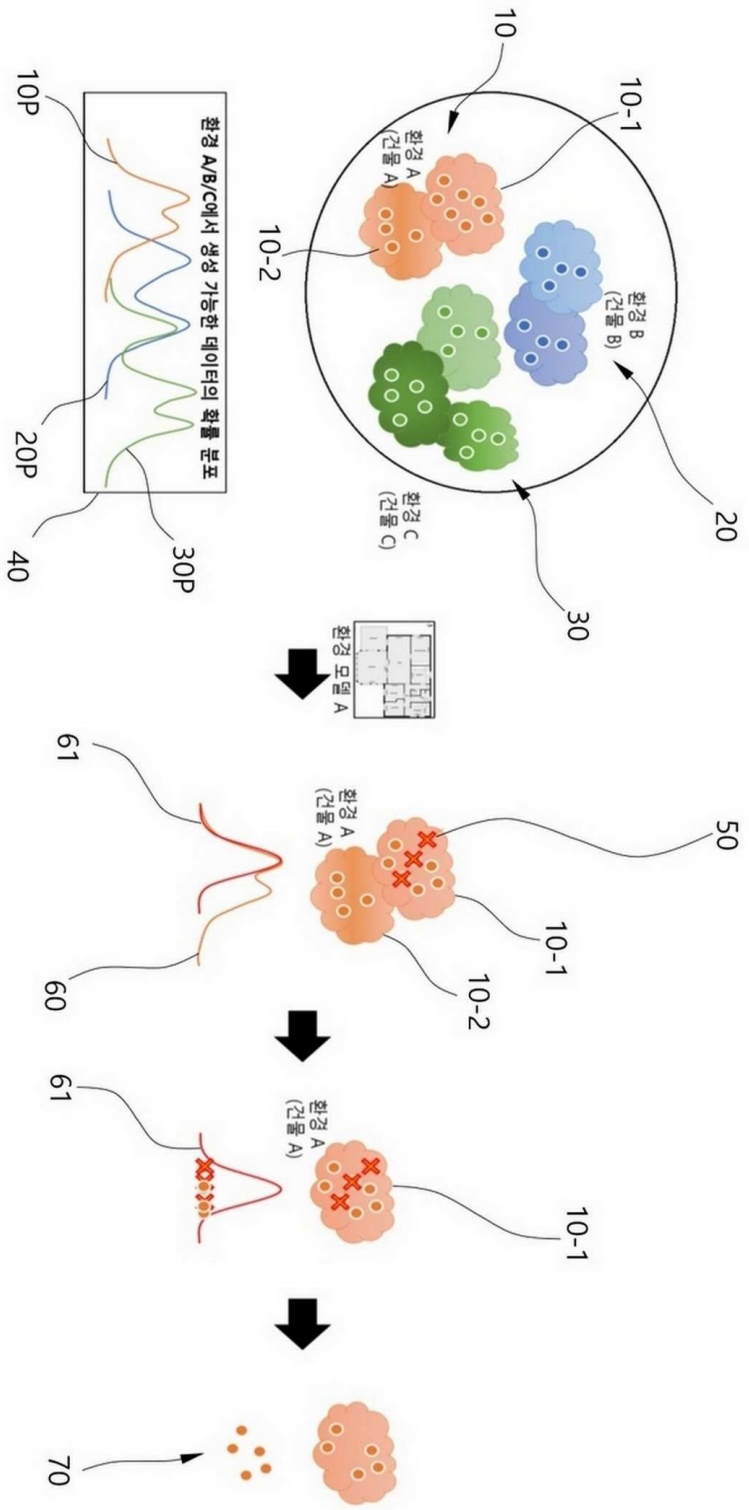
450: 디스플레이부

500: 입력부

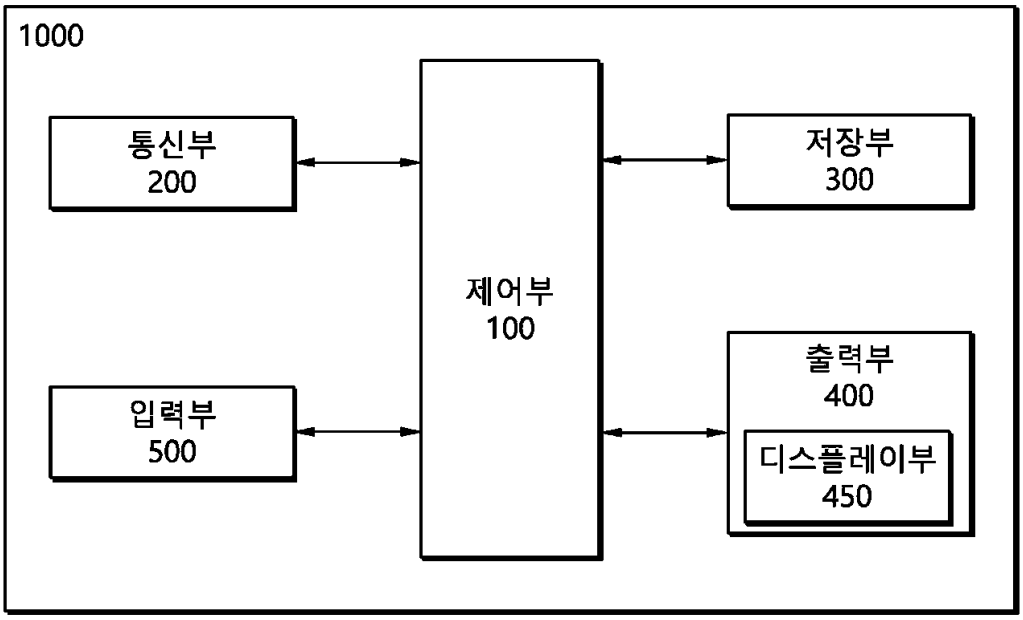
1000: 합성 데이터 생성 장치

도면

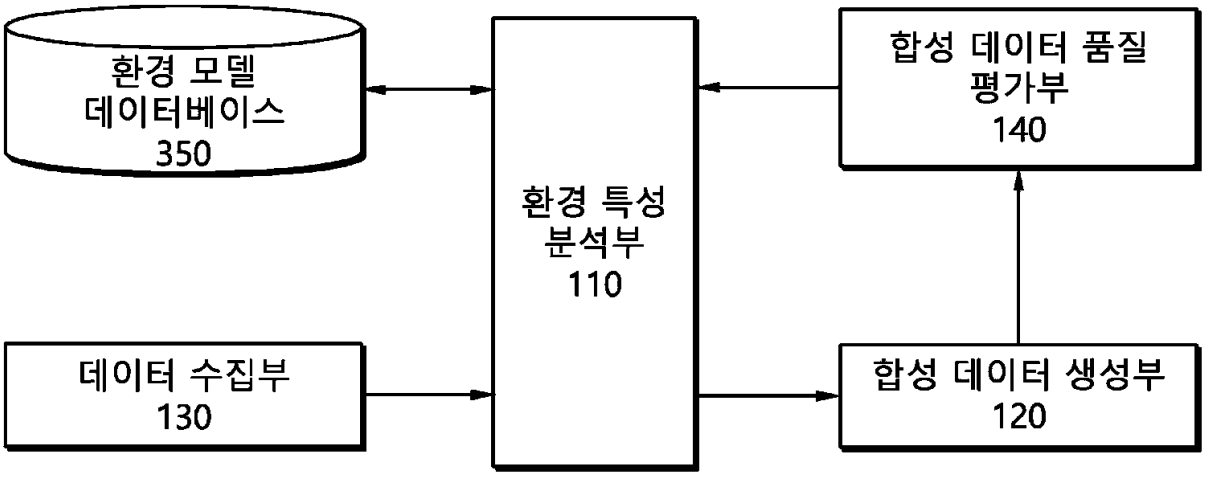
도면1


10
10-1
10-2
환경 A
(건물 A)
환경 B
(건물 B)
환경 C
(건물 C)
20
30
환경 A/B/C에서 생성 가능한 데이터의 확률 분포
10P
20P
30P
40
환경 모델 A
61
60
50
10-1
10-2
환경 A
(건물 A)
61
10-1
환경 A
(건물 A)
70

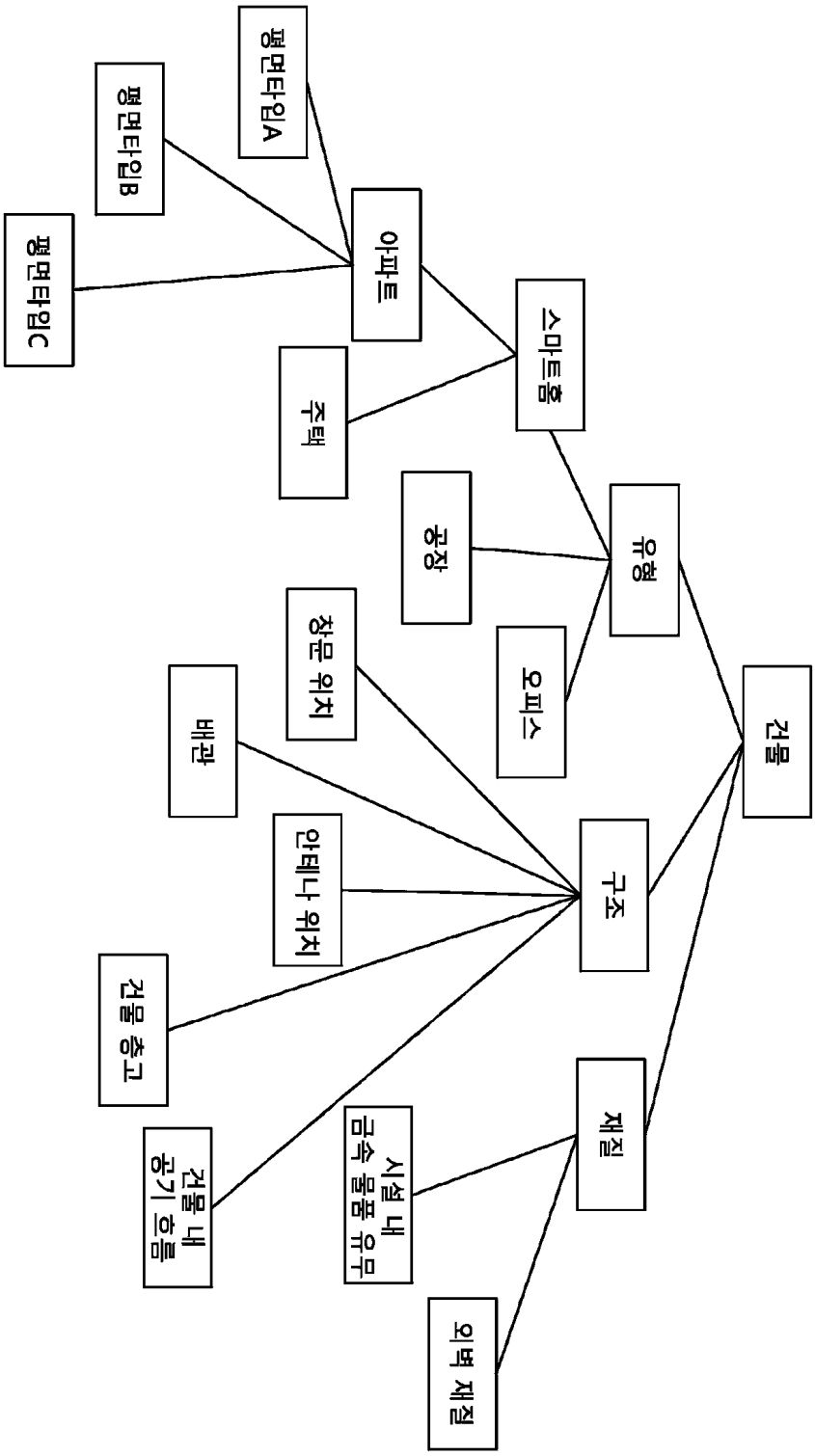
도면2


1000
통신부
200
제어부
100
저장부
300
출력부
400
디스플레이부
450
입력부
500

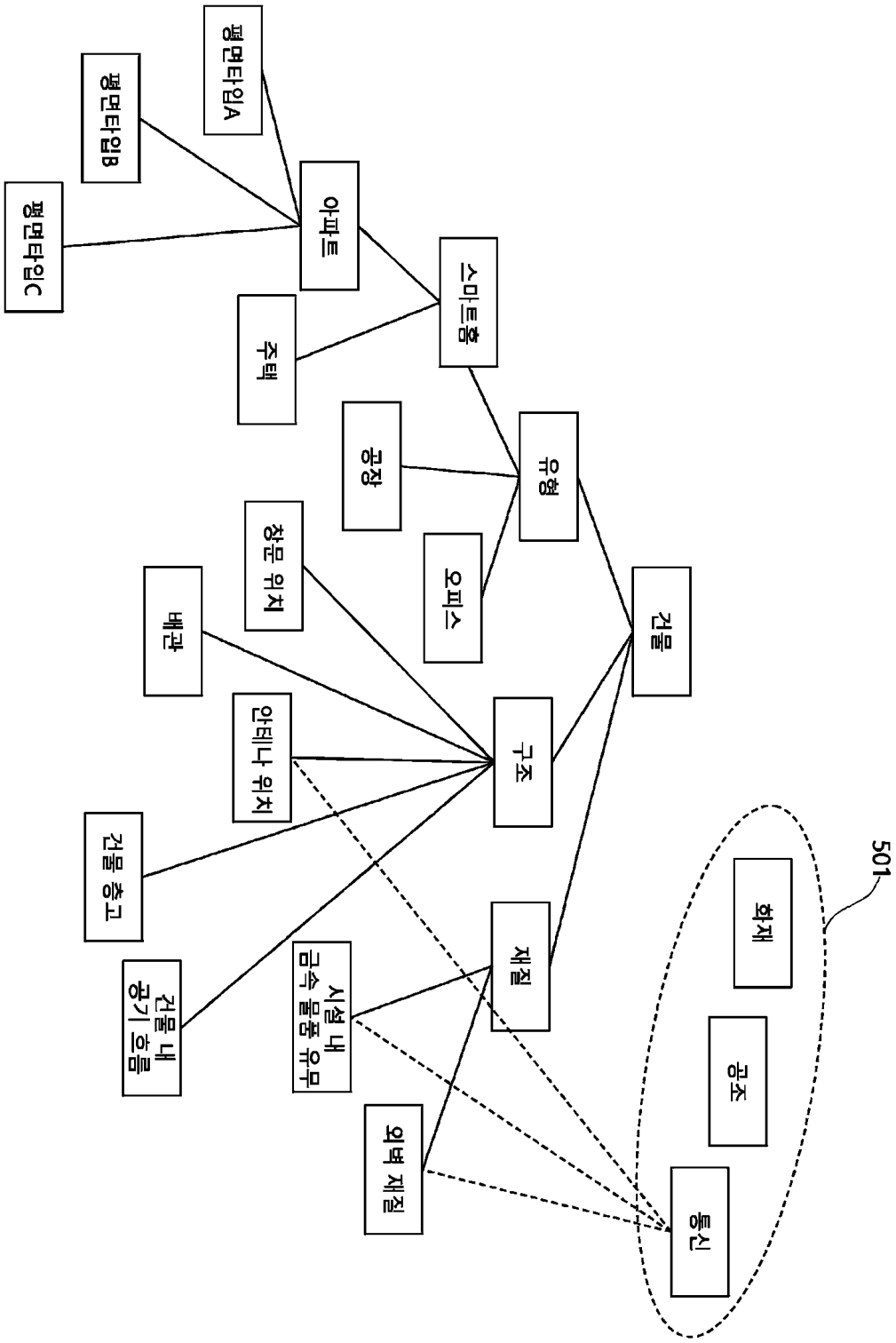
도면3


환경 모델
데이터베이스
350
환경 특성
분석부
110
합성 데이터 품질
평가부
140
데이터 수집부
130
합성 데이터 생성부
120

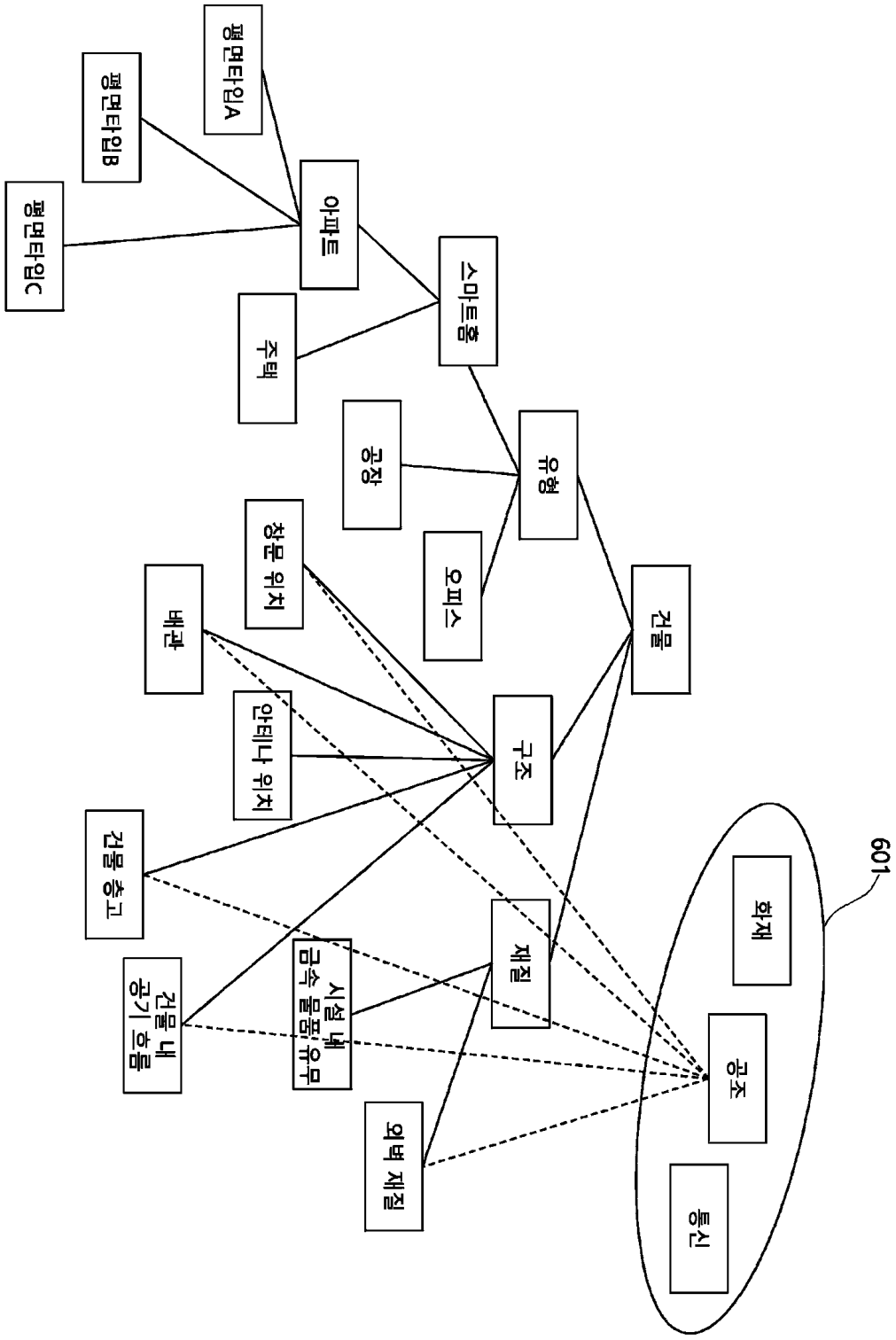
도면4


건물
유형
아파트
평면타입A
평면타입B
평면타입C
스마트홈
주택
공장
오피스
구조
창문 위치
배관
안테나 위치
건물 층고
건물 내 공기 흐름
재질
시설 내 금속 물품 유무
외벽 재질

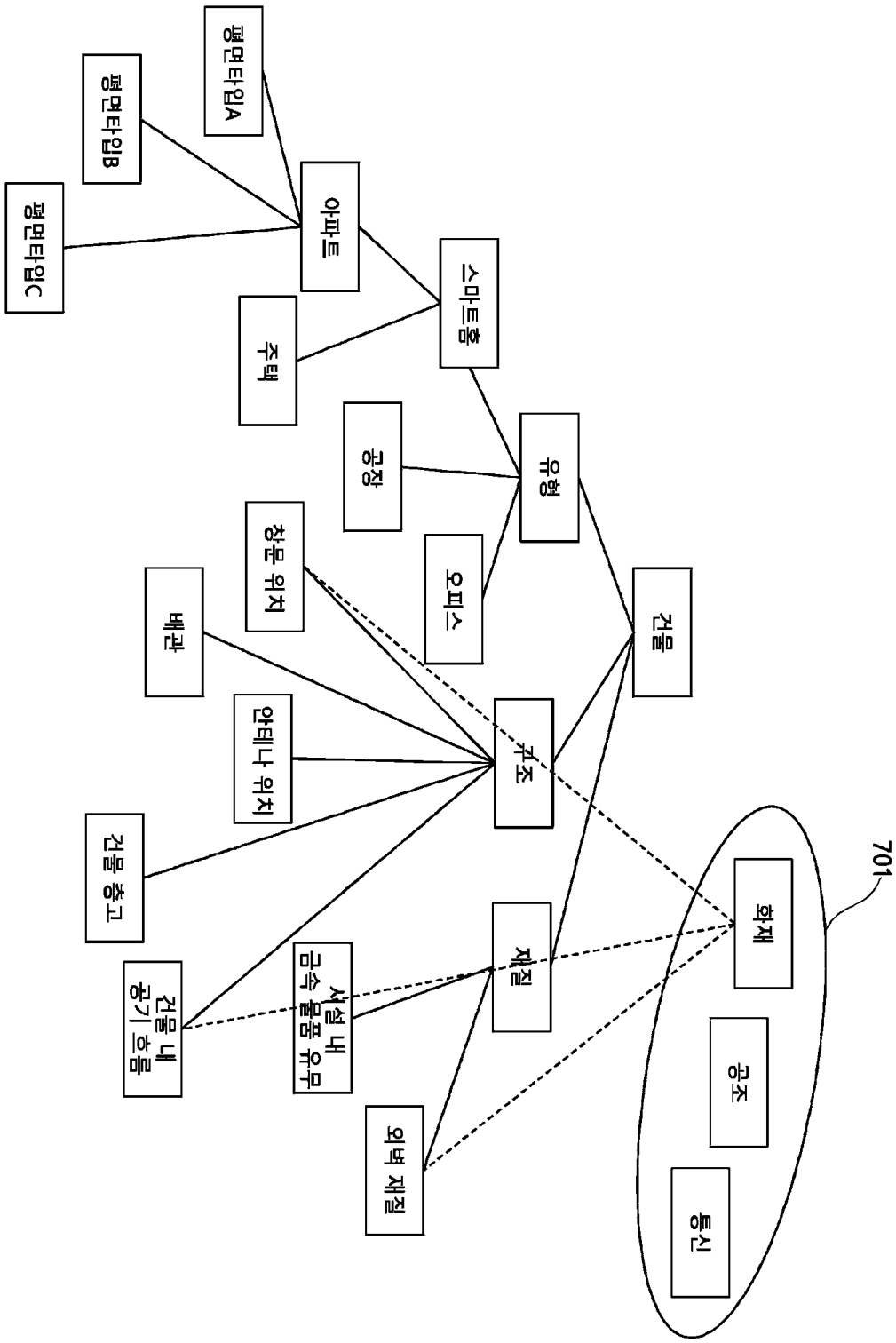
도면5


평면타입A
평면타입B
평면타입C
아파트
주택
스마트홈
공장
오피스
유형
창문 위치
배관
안테나 위치
건물 층고
건물 내 공기 흐름
구조
시설 내 금속 물품 유무
외벽 재질
재질
건물
화재
공조
통신
501

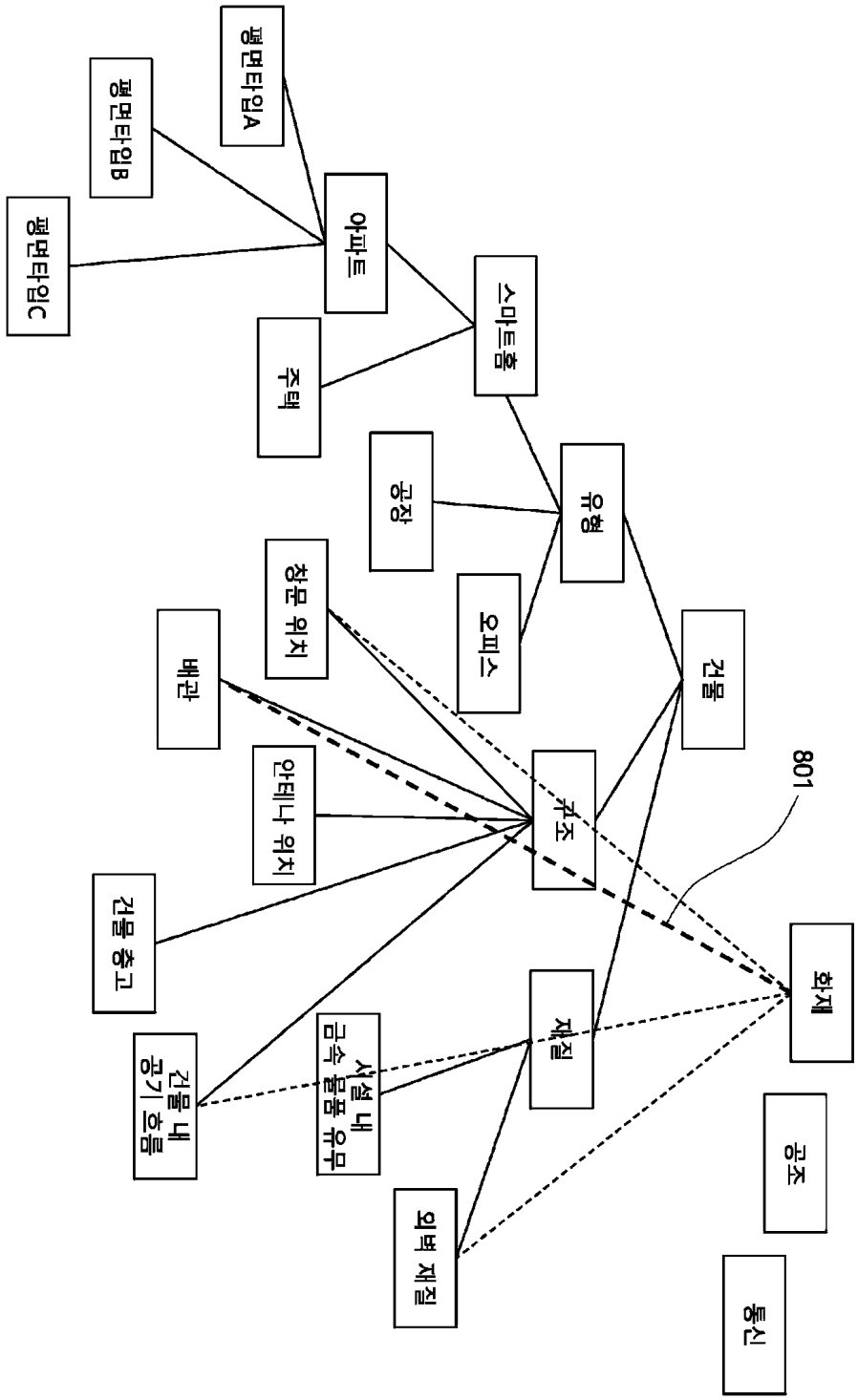
도면6


평면타입A
평면타입B
평면타입C
아파트
주택
스마트홈
공장
오피스
유형
창문 위치
배관
안테나 위치
건물 층고
건물 내 공기 흐름
시설 내 금속 물품 유무
구조
재질
외벽 재질
건물
화재
공조
통신
601

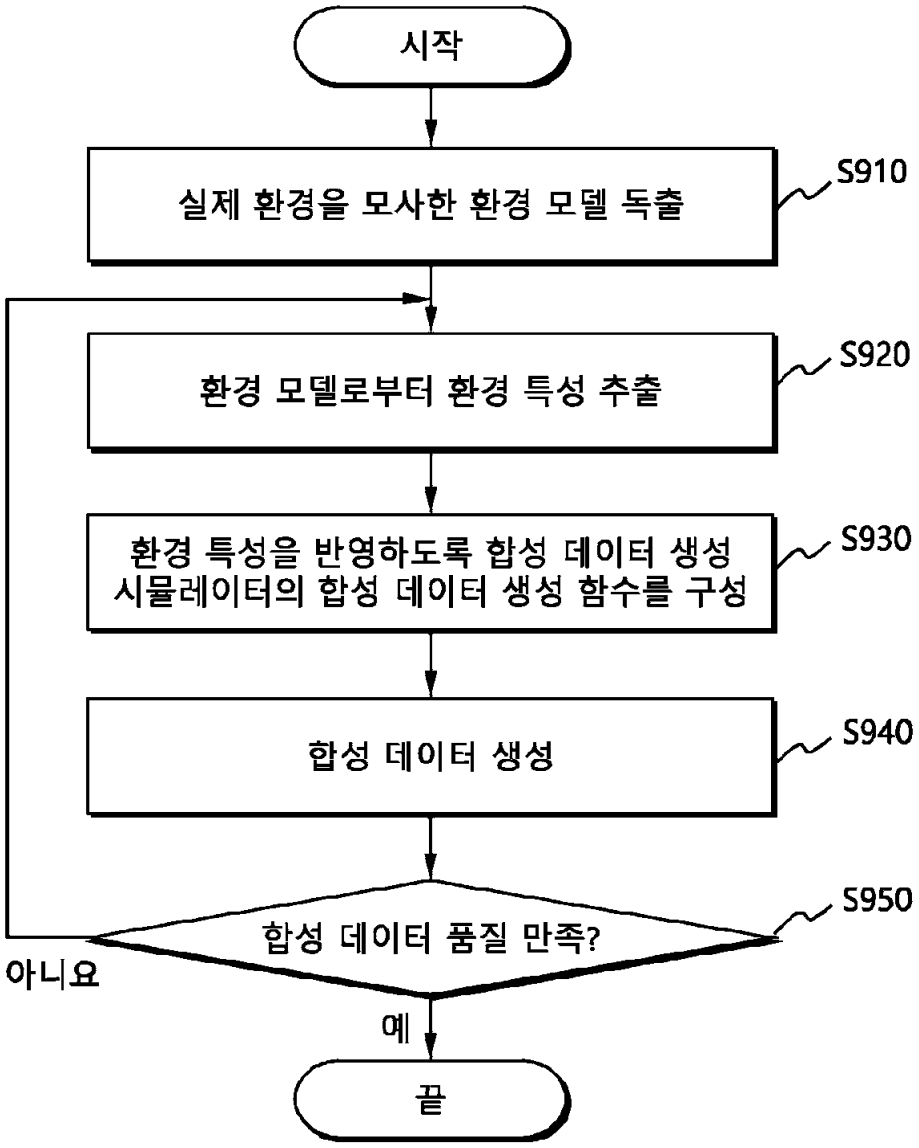
도면7


평면타입A
평면타입B
평면타입C
아파트
주택
스마트홈
공장
오피스
유형
창문 위치
배관
안테나 위치
건물 층고
건물 내 공기 흐름
서설 내 금속 물품 유무
외벽 재질
구조
재질
건물
화재
공조
통신
701

도면8


평면타입A
평면타입B
평면타입C
아파트
주택
스마트홈
공장
오피스
유형
창문 위치
배관
안테나 위치
건물 층고
건물 내 공기 흐름
설치 내 금속 물품 유무
건물
구조
재질
외벽 재질
801
화재
공조
통신

도면9


시작
실제 환경을 모사한 환경 모델 독출
S910
환경 모델로부터 환경 특성 추출
S920
환경 특성을 반영하도록 합성 데이터 생성
시뮬레이터의 합성 데이터 생성 함수를 구성
S930
합성 데이터 생성
S940
합성 데이터 품질 만족?
S950
아니요
예
끝