



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0120404
(43) 공개일자 2024년08월07일

- (51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) G06N 3/0475 (2023.01)
- (52) CPC특허분류
G06N 20/00 (2021.08)
G06N 3/0475 (2023.01)
- (21) 출원번호 10-2023-0013029
- (22) 출원일자 2023년01월31일
심사청구일자 2023년01월31일
- (71) 출원인
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
주식회사 큐빅
서울특별시 서대문구 북아현로 150, 이화여자대학교 산학협력관 124에이호(신촌동)
- (72) 발명자
배호
서울특별시 서초구 남부순환로 2311-12, 101동 801호(방배동, 래미안방배아트힐)
- (74) 대리인
특허법인(유한)아이시스

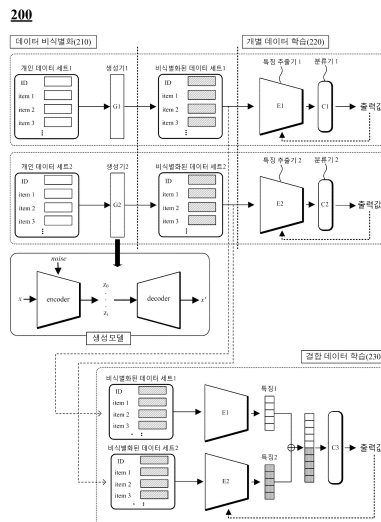
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 이중 데이터 세트를 처리하는 인공지능 모델 구축 방법 및 분석장치

(57) 요약

이중 데이터 세트를 처리하는 인공지능 모델 구축 방법은 학습장치가 개인 정보를 포함하는 이중 데이터 세트를 입력받는 단계, 상기 학습장치가 생성모델을 이용하여 상기 이중 데이터 세트에 포함된 개인 데이터 세트들 각각을 비식별화하여 비식별화된 개인 데이터 세트들을 생성하는 단계, 상기 학습장치가 상기 비식별화된 개인 데이터 세트들 각각을 처리하는 개별 인공지능 모델들을 학습하는 단계 및 상기 학습장치가 상기 비식별화된 개인 데이터 세트들을 이용하여 최종 인공지능 모델을 학습하는 단계를 포함한다.

대표도 - 도2



명세서

청구범위

청구항 1

학습장치가 개인 정보를 포함하는 이중 데이터 세트를 입력받는 단계;

상기 학습장치가 생성모델을 이용하여 상기 이중 데이터 세트에 포함된 개인 데이터 세트들 각각을 비식별화하여 비식별화된 개인 데이터 세트들을 생성하는 단계;

상기 학습장치가 상기 비식별화된 개인 데이터 세트들 각각을 처리하는 개별 인공지능 모델들을 학습하는 단계; 및

상기 학습장치가 상기 비식별화된 개인 데이터 세트들을 이용하여 최종 인공지능 모델을 학습하는 단계를 포함하되,

상기 최종 인공지능 모델은 상기 개별 인공지능 모델들의 계층 중 입력 데이터의 특징을 추출하는 특징 추출 계층들을 포함하고, 상기 특징 추출 계층들은 상기 개별 인공지능 모델들을 학습하는 단계에서 학습이 완료된 구성이고,

상기 최종 인공지능 모델을 학습하는 단계에서 상기 학습장치는 상기 비식별화된 개인 데이터 세트들을 각각을 상기 특징 추출 계층들에 입력하고, 상기 특징 추출 계층들이 출력하는 특징 벡터들을 결합하여 다음 계층에 전달하는 이중 데이터 세트를 처리하는 인공지능 모델 구축 방법.

청구항 2

제1항에 있어서,

상기 개별 인공지능 모델들은 각각 특징 추출기 및 분류기를 포함하고,

상기 최종 인공지능 모델은 상기 개별 인공지능 모델들에서 학습된 특징 추출기들 및 상기 결합된 비식별화된 개인 데이터 세트를 입력받아 분류를 수행하는 별도의 분류기를 포함하는 이중 데이터 세트를 처리하는 인공지능 모델 구축 방법.

청구항 3

제1항에 있어서,

상기 개인 데이터 세트들 중 어느 하나의 개인 데이터 세트를 비식별화하는 상기 생성모델은 상기 어느 하나의 개인 데이터 세트 및 잡음을 입력받아 잠재 벡터를 생성하고, 상기 잠재 벡터로부터 상기 어느 하나의 개인 데이터 세트에 대응하는 비식별화된 개인 데이터 세트를 산출하는 이중 데이터 세트를 처리하는 인공지능 모델 구축 방법.

청구항 4

분석 대상인 이중 데이터 세트를 입력받는 인터페이스 장치;

개인 정보를 비식별화하는 생성모델 및 인공지능 모델을 저장하는 저장장치; 및

상기 생성모델을 이용하여 상기 이중 데이터 세트에 포함된 개인 데이터 세트들 각각을 비식별화하고, 상기 비식별화된 개인 데이터 세트들 각각을 상기 인공지능 모델의 특징 추출 계층들에 입력하고, 상기 특징 추출 계층들이 출력하는 특징 벡터들을 결합하여 상기 인공지능 모델의 분류 계층에 입력하고, 상기 분류 계층의 출력값을 기준으로 상기 이중 데이터 세트에 대한 분류 결과를 산출하는 연산장치를 포함하되,

상기 특징 추출 계층들은 학습과정에서 상기 비식별화된 개인 데이터 세트들 각각을 처리하는 개별 인공지능 모델들의 학습된 특징 추출 계층들을 이용하는 이중 데이터 세트를 분석하는 분석장치.

청구항 5

제4항에 있어서,

상기 분류 계층은 상기 개별 인공지능 모델들의 학습된 특징 추출 계층들이 출력하는 특징 벡터들을 결합하여 입력받아 목표하는 출력값을 출력하도록 학습되는 이중 데이터 세트를 분석하는 분석장치.

발명의 설명

기술 분야

[0001] 이하 설명하는 기술은 개인 정보 유출 없이 이중 데이터 세트를 분석하는 기법에 관한 것이다. 특히 이하 설명하는 기술은 인공지능 모델을 이용하여 결합된 이중 데이터를 분석하는 기법에 관한 것이다.

배경 기술

[0002] 4차 산업 혁명 시대를 맞이하여 인공지능, 사물 인터넷, 클라우드 컴퓨팅, 빅데이터 등 다양한 정보통신 기술이 활용되고 있다. 특히 인공지능은 대량의 정형 또는 비정형 데이터로부터 가치 있는 정보를 추출하고 분석하는 기술로 주목받고 있다.

[0003] 한편, 개인 정보 보호의 중요성이 대두되면서 인공지능 서비스를 위한 무분별한 수집이 사회적 문제로 대두되고 있다. 다수의 국가에서 개인 정보가 보장되지 않는 데이터는 인공지능 서비스에 활용이 제한될 예정이다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 한국등록특허 제10-2299564호

발명의 내용

해결하려는 과제

[0005] 개인 정보 보호를 위하여 인공지능 모델의 학습에 익명화된 데이터를 이용하기도 한다. 다만 종래 개인 정보 보호 기술은 단일 데이터 세트에 대해서만 활용이 가능하다. 인공지능 모델이 서로 다른 출처 또는 이중의 데이터 세트를 활용하는 경우 데이터의 결합 사용 시점에서는 정보 유출이 되지 않더라도, 학습된 인공지능 모델에 대한 공격을 통해 개인 정보가 유출될 수 있다. 예컨대, 멤버십 추론 공격(Membership Inference Attack)은 인공지능 모델의 학습에 개인 정보 데이터가 활용되었는지 확인할 수 있다. 또는, 모델 전도 공격(Model Inversion Attack)은 학습에 활용된 데이터를 복구하여 개인 정보가 유출될 수도 있다.

[0006] 이하 설명하는 기술은 다수의 데이터 세트를 이용하면서 개인 정보 유출의 문제가 없는 인공지능 모델 내지 모델 학습 방법을 제공하고자 한다.

[0007] 나아가, 이하 설명하는 기술은 서로 다른 특성들을 갖는 이중의 데이터를 이용하면서도 성능 저하 없는 인공지능 모델 내지 모델 학습 방법을 제공하고자 한다.

과제의 해결 수단

[0008] 이중 데이터 세트를 처리하는 인공지능 모델 구축 방법은 학습장치가 개인 정보를 포함하는 이중 데이터 세트를 입력받는 단계, 상기 학습장치가 생성모델을 이용하여 상기 이중 데이터 세트에 포함된 개인 데이터 세트들 각각을 비식별화하여 비식별화된 개인 데이터 세트들을 생성하는 단계, 상기 학습장치가 상기 비식별화된 개인 데이터 세트들 각각을 처리하는 개별 인공지능 모델들을 학습하는 단계 및 상기 학습장치가 상기 비식별화된 개인 데이터 세트들을 이용하여 최종 인공지능 모델을 학습하는 단계를 포함한다.

[0009] 상기 최종 인공지능 모델은 상기 개별 인공지능 모델들의 계층 중 입력 데이터의 특징을 추출하는 특징 추출 계층들을 포함하고, 상기 특징 추출 계층들은 상기 개별 인공지능 모델들을 학습하는 단계에서 학습이 완료된 구성이다.

[0010] 상기 최종 인공지능 모델을 학습하는 단계에서 상기 학습장치는 상기 비식별화된 개인 데이터 세트들을 각각을

상기 특징 추출 계층들에 입력하고, 상기 특징 추출 계층들이 출력하는 특징 벡터들을 결합하여 다음 계층에 전달한다.

[0011] 다른 측면에서 이중 데이터 세트를 분석하는 분석장치는 분석 대상인 이중 데이터 세트를 입력받는 인터페이스 장치, 개인 정보를 비식별화하는 생성모델 및 인공지능 모델을 저장하는 저장장치 및 상기 생성모델을 이용하여 상기 이중 데이터 세트에 포함된 개인 데이터 세트들 각각을 비식별화하고, 상기 비식별화된 개인 데이터 세트들 각각을 상기 인공지능 모델의 특징 추출 계층들에 입력하고, 상기 특징 추출 계층들이 출력하는 특징 벡터들을 결합하여 상기 인공지능 모델의 분류 계층에 입력하고, 상기 분류 계층의 출력값을 기준으로 상기 이중 데이터 세트에 대한 분류 결과를 산출하는 연산장치를 포함한다.

발명의 효과

[0012] 이하 설명하는 기술은 다수의 데이터 세트를 이용하는 인공지능 서비스를 제공하면서도 현존하는 공격 방법에 기인한 개인 정보를 막을 수 있다. 또한, 이하 설명하는 기술은 이중의 데이터 세트를 특성 공간(feature space)에서 결합하여 단일 데이터 세트로 학습된 모델과 같은 성능을 보장할 수 있다.

도면의 간단한 설명

[0013] 도 1은 이중의 데이터 세트를 분석하는 인공지능 모델을 이용한 서비스 시스템에 대한 예이다.

도 2는 이중의 데이터 세트를 이용하는 인공지능 모델의 학습 과정의 예이다.

도 3은 이중 데이터 세트를 이용하여 구축된 인공지능 모델의 추론 과정의 예이다.

도 4는 인공지능 모델을 이용하여 이중 데이터 세트를 분석하는 분석장치의 예이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다.

[0015] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 해당 구성요소들은 상기 용어들에 의해 한정되지는 않으며, 단지 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 이하 설명하는 기술의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. 및/또는 이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.

[0016] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 설명된 특징, 개수, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 의미하는 것이지, 하나 또는 그 이상의 다른 특징들이나 개수, 단계 동작 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 배제하지 않는 것으로 이해되어야 한다.

[0017] 도면에 대한 상세한 설명을 하기에 앞서, 본 명세서에서의 구성부들에 대한 구분은 각 구성부가 담당하는 주기능 별로 구분한 것에 불과함을 명확히 하고자 한다. 즉, 이하에서 설명할 2개 이상의 구성부가 하나의 구성부로 합쳐지거나 또는 하나의 구성부가 보다 세분화된 기능별로 2개 이상으로 분화되어 구비될 수도 있다. 그리고 이하에서 설명할 구성부 각각은 자신이 담당하는 주기능 이외에도 다른 구성부가 담당하는 기능 중 일부 또는 전부의 기능을 추가적으로 수행할 수도 있으며, 구성부 각각이 담당하는 주기능 중 일부 기능이 다른 구성부에 의해 전담되어 수행될 수도 있음은 물론이다.

[0018] 또, 방법 또는 동작 방법을 수행함에 있어서, 상기 방법을 이루는 각 과정들은 문맥상 명백하게 특정 순서를 기재하지 않은 이상 명기된 순서와 다르게 일어날 수 있다. 즉, 각 과정들은 명기된 순서와 동일하게 일어날 수도 있고 실질적으로 동시에 수행될 수도 있으며 반대의 순서대로 수행될 수도 있다.

[0020] 이하 설명하는 기술은 인공지능 모델에 대한 학습 기법, 해당 학습 기법으로 구축된 인공지능 모델 및 해당 인공지능 모델을 이용한 추론 기법에 관한 것이다.

- [0021] 인공지능 모델은 다양한 유형의 모델들을 포함하는 개념이다. 딥러닝(deep learning network, DNN)은 대표적인 인공지능 모델이다. DNN은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공신경망 모델이다. DNN은 일반적인 인공신경망과 마찬가지로 복잡한 비선형 관계(non-linear relationship)들을 모델링할 수 있다. DNN은 다양한 유형의 모델이 연구되었다. 예컨대, CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), RBM(Restricted Boltzmann Machine), DBN(Deep Belief Network), GAN(Generative Adversarial Network), RL(Relation Networks) 등이 있다. 이하 설명의 편의를 위하여 DNN을 기준으로 설명한다.
- [0022] 이하 인공지능 모델은 이종의 데이터 세트를 이용하여 학습되는 모델을 가정한다. 이종의 데이터 세트는 이중 데이터들로 구성될 수 있다. 예컨대, 이종의 데이터 세트는 의료 데이터, 생체 데이터, 금융 데이터 등과 같이 데이터의 소스가 이질적인 데이터들로 구성될 수 있다. 또는, 이종의 데이터 세트는 영상 데이터, 음성 데이터, 텍스트 데이터 등과 같이 모달리티(modality)가 서로 다른 데이터들로 구성될 수도 있다. 한편, 이종의 데이터 세트는 개인 정보를 포함한다고 가정한다.
- [0023] 이하 학습장치가 인공지능 모델의 학습을 수행한다고 설명한다. 학습장치는 데이터 전처리, 학습 과정에서 필요한 연산 및 학습 과정 수행이 가능한 컴퓨터 장치를 의미한다. 학습 장치는 서버, PC, 스마트기기, 프로그램이 임베디드된 칩 등의 형태로 구현될 수 있다.
- [0024] 이하 분석장치가 학습된 인공지능 모델을 이용하여 일정한 추론을 수행한다고 설명한다. 분석장치는 입력 데이터 전처리, 인공지능 모델의 구동, 인공지능 모델의 출력값 분석 등이 가능한 컴퓨터 장치를 의미한다. 분석장치는 서버, PC, 스마트기기, 프로그램이 임베디드된 칩 등의 형태로 구현될 수 있다.
- [0026] 도 1은 이종의 데이터 세트를 분석하는 인공지능 모델을 이용한 서비스 시스템(100)에 대한 예이다. 도 1은 분석장치(150)로 서버 및 PC를 예로 도시한다.
- [0027] 이중 데이터 세트는 서로 다른 소스를 통하여 사전에 수집된 상태이다. 도 1은 n개의 이중 데이터들을 예로 도시하였다. 데이터 세트 a, 데이터 세트 b, ..., 데이터 세트 n은 서로 다른 유형 내지 모달리티를 갖는 이중 데이터 세트를 구성한다.
- [0028] 이중 데이터 세트는 동일 사용자에게 대하여 데이터 소스 내지 유형이 상이한 의료 데이터, 금융 데이터, 소비 유형 데이터, 라이프로그 데이터 등을 포함할 수 있다. 또는, 이중 데이터 세트는 동일 대상에 대하여 모달리티가 상이한 영상 데이터, 음성 데이터, 텍스트 데이터 등을 포함할 수도 있다.
- [0029] 분석장치(150)는 데이터 수집 장치로부터 각각 데이터 세트 a, 데이터 세트 b, ..., 데이터 세트 n을 수신할 수 있다. 또한, 도 1과 같이 분석장치(150)는 각각의 데이터베이스(110a, 110b, ..., 110n)로부터 데이터 세트 a, 데이터 세트 b, ..., 데이터 세트 n을 수신할 수도 있다.
- [0030] 분석장치(150)는 사전에 구축한 인공지능 모델을 이용하여 이중 데이터 세트를 분석한다. 분석장치(150)는 이중 데이터들을 각각 서로 다른 입력 계층(특징 추출 계층)에 입력하여 특징을 추출하고, 이중 데이터들의 특징을 결합한 특징값을 분류 계층에 입력하여 일정한 출력값을 얻는다. 분석장치(150)는 인공지능 모델에서 출력하는 출력값을 기준으로 일정한 정보를 추출할 수 있다. 나아가, 분석장치(150)는 추출한 정보를 기준으로 사용자에게 대한 맞춤형 서비스를 제공할 수도 있다.
- [0031] 분석장치(150)는 분석 결과 내지 서비스를 사용자에게 제공할 수 있다. 이때 사용자는 사용자 단말(180)을 통해 분석 결과 내지 서비스를 제공받을 수 있다. 한편, 분석장치(150)가 PC와 같은 사용자 단말인 경우 사용자는 분석장치를 통해 분석 결과나 서비스를 제공받을 수 있다.
- [0033] 도 2는 이종의 데이터 세트를 이용하는 인공지능 모델의 학습 과정(200)의 예이다.
- [0034] 도 2는 두 가지 종류의 데이터 세트를 이용한 예이다. 도 2에서 이중 데이터 세트는 개인 데이터 세트 1 및 개인 데이터 세트 2를 포함한다. 학습 데이터는 동일 유형에 대하여 다수의 데이터들로 구성된다.
- [0035] 개인 데이터 세트 1은 사용자 식별자(ID), 데이터 항목 1(item 1), 데이터 항목 2(item 2), 데이터 항목 3(item 3) 등을 포함한다. 개인 데이터 세트 1은 의료 데이터일 수 있다. 이 경우, 개인 데이터 세트 1은 주민

번호와 같은 ID, 성별, 나이, 문진 항목, 검사 결과, 특정 질환 여부 등으로 구성될 수 있다.

- [0036] 개인 데이터 세트 2는 사용자 식별자(ID), 데이터 항목 1(item 1), 데이터 항목 2(item 2), 데이터 항목 3(item 3) 등을 포함한다. 개인 데이터 세트 1은 금융 데이터일 수도 있다. 이 경우, 개인 데이터 세트 1은 주민번호와 같은 ID, 수입, 자산, 채무, 보험 정보 등으로 구성될 수 있다.
- [0037] 학습장치는 개인 데이터 세트들을 비식별화한다(210).
- [0038] 학습장치는 생성기(generator)를 이용하여 개인 데이터 세트를 비식별화할 수 있다. 학습장치는 개인 데이터 세트별로 서로 다른 생성기를 이용하여 데이터 세트를 비식별화할 수 있다. 예컨대, 학습장치는 개인 데이터 세트 1을 생성기 1에 입력하여 비식별화된 데이터 세트 1을 산출할 수 있다. 또한, 학습장치는 개인 데이터 세트 2를 생성기 2에 입력하여 비식별화된 데이터 세트 2를 산출할 수 있다.
- [0039] 생성기는 생성모델(generative model)일 수 있다. 생성모델은 VAE(Variational AutoEncoder), GAN, DPGAN(Differentially Private Generative Adversarial Networks) 등과 같은 모델일 수 있다. 예컨대, 생성모델은 입력 데이터 x 를 입력받아 특징을 추출하는 인코더 및 인코더가 출력하는 잠재 벡터 $z_0, \dots, z_t$ 를 입력받아 비식별화된 데이터 x' 를 출력하는 디코더(GAN 모델의 경우 생성기)를 포함할 수 있다. 이때 인코더는 입력 데이터와 함께 일정한 잡음(noise)을 입력받아 입력 데이터의 개인 정보에 잡음을 삽입한 데이터로부터 잠재 벡터를 생성할 수 있다. 디코더(또는 생성기)는 잠재 벡터를 입력받아 데이터 세트를 복원하는데 잡음의 영향으로 개인 정보가 비식별화된 데이터 세트를 산출한다.
- [0040] 도 2에서 학습장치는 개별 개인 데이터 세트에 대하여 개별 생성기를 사용하였다. 개인 데이터 세트들이 동일한 크기 내지 규격을 갖는다면, 학습장치는 하나의 생성기를 이용하여 개인 데이터 세트들 각각을 비식별화할 수도 있다.
- [0041] 나아가, 학습장치는 다른 유형의 학습 모델을 이용하여 개인 데이터 세트를 비식별화할 수 있다. 또한, 학습장치는 학습모델이 아닌 다른 비식별화 기법을 이용하여 개인 데이터 세트를 비식별화할 수도 있다.
- [0042] 비식별화된 개인 데이터 세트는 ID 및 개인 데이터 세트에 있는 항목들에 대하여 비식별화된 데이터들을 포함할 수 있다. 비식별화된 개인 데이터 세트에서 사용하는 ID는 개인 데이터 세트에 있는 ID와는 다른 내부 ID이다. 즉, 개인 데이터 세트의 ID는 주민번호와 같은 정보라면, 비식별화된 개인 데이터 세트의 ID는 일련의 번호일 수 있다. 한편, 인공지능 모델의 애플리케이션에 따라서 비식별화된 개인 데이터 세트는 내부 ID를 포함하지 않을 수도 있다.
- [0043] 학습장치는 비식별화된 데이터 세트들 각각을 이용하여 인공지능 모델에 대한 초기 학습을 수행한다(220). 도 2의 인공지능 모델은 특징 추출기(feature extractor) 및 분류기(classifier)로 구성된다. 인공지능 모델의 구조는 입력 데이터의 유형, 서비스 애플리케이션 및 개발자의 구현 양태에 따라 다양할 수 있다.
- [0044] 특징 추출기 1(E1)은 비식별화된 데이터 세트 1을 입력받아 특징 벡터를 출력하고, 분류기 1(C1)은 특징 추출기 1(E1)로부터 특징 벡터를 입력받아 일정한 출력값을 출력한다. 출력값은 분석하는 정보 또는 서비스 유형에 따라 다양한 값일 수 있다. 예컨대, 이진 분류라면 출력값은 $[0,1]$ 범위의 확률값일 수 있다.
- [0045] 특징 추출기 2(E2)는 비식별화된 데이터 세트 2를 입력받아 특징 벡터를 출력하고, 분류기 2(C2)는 특징 추출기 2(E2)로부터 특징 벡터를 입력받아 일정한 출력값을 출력한다.
- [0046] 학습장치는 다수의 학습 데이터를 이용하여 인공지능 모델의 학습 과정을 반복한다. 학습과정은 학습장치가 분류기의 출력값과 정답값(라벨값)을 비교하여 특징 추출기 및/또는 분류기의 가중치를 업데이트하는 과정이다. 학습장치는 개인 데이터 세트 1에 대하여 분류기 1(C1)이 임계값 이상의 성능을 보일 때까지 특징 추출기 1(E1) 및/또는 분류기 1(C1)의 학습 과정을 반복한다. 학습장치는 개인 데이터 세트 2에 대하여 분류기 2(C2)가 임계값 이상의 성능을 보일 때까지 특징 추출기 2(E2) 및/또는 분류기 2(C2)의 학습 과정을 반복한다.
- [0047] 학습장치는 개별 개인 데이터 세트를 이용한 개별 모델의 학습 과정이 완료되면, 결합 데이터를 이용한 최종 모델 학습 과정을 수행한다(230).
- [0048] 최종 인공지능 모델은 개별 데이터 학습 과정(220)에서 학습이 완료된 특징 추출기 1(E1) 및 특징 추출기 2(E2)를 이용한다. 최종 인공지능 모델은 특징 추출기 1(E1), 특징 추출기 2(E2) 및 분류기 3(C3)을 포함한다. 분류기 3(C3)은 새로운 분류기이다. 분류기 3(C3)은 분류기 1(C1) 및/또는 분류기 2(C2)와 동일한 분류를 수행할 수 있다. 또한, 분류기 3(C3)은 분류기 1(C1) 및/또는 분류기 2(C2)와 다른 목적으로 학습될 수도 있다.

예컨대, 분류기 3(C3)은 개인 데이터 세트 1 및 개인 데이터 세트 2를 이용한 새로운 목적의 분류기일 수도 있다.

- [0049] 학습장치는 데이터 비식별화 과정(210)에서 비식별화한 개인 데이터 세트를 학습 데이터로 이용한다. 최종 인공지능 모델은 기본적으로 비식별화된 데이터를 이용하기에 개인 정보 유출의 위험성이 낮다.
- [0050] 학습장치는 비식별화된 데이터 세트 1을 특징 추출기 1(E1)에 입력하고, 비식별화된 데이터 세트 2를 특징 추출기 2(E2)에 입력한다. 즉, 최종 인공지능 모델은 개별 개인 데이터 세트별로 입력 계층을 달리 갖는다.
- [0051] 학습장치는 특징 추출기 1(E1)이 출력하는 특징 1과 특징 추출기 2(E2)가 출력하는 특징 2를 결합(concatenation)한다. 최종 인공지능 모델은 결합된 데이터 세트를 최종 분류기에서만 이용한다. 비식별화된 데이터 세트의 결합 시에 동일한 ID를 사용하는 데이터의 매칭 과정이 개인 정보 유출의 통로가 될 수 있다. 그러나, 도 2와 같이 결합 데이터의 사용을 분류기에 한정하면 개인 정보 유출의 위험도를 현저하게 줄일 수 있다. 나아가, 이종 데이터 세트는 특성이 상이한 데이터로 구성될 수 있고, 이 경우 인공지능 모델의 성능이 떨어지기 쉽다. 그러나, 도 2와 같이 특성 공간에서 데이터를 결합하면 해당 인공지능 모델은 단일 데이터 세트를 이용한 모델과 같은 성능을 가질 수 있다.
- [0052] 학습장치는 결합된 특징 벡터를 분류기 3(C3)에 입력하고 분류기 3(C3)이 출력하는 값과 정답값을 비교하여 분류기 3(C3)의 가중치를 업데이트한다. 경우에 따라서, 학습장치는 분류기 3(C3)이 출력하는 값과 정답값을 비교하여 특징 추출기 1(E1), 특징 추출기 2(E2) 및 분류기 3(C3) 중 적어도 하나의 계층에 대한 가중치를 업데이트할 수도 있다. 학습장치는 분류기 3(C3)이 임계값 이상의 성능을 보일 때까지 다수의 학습 데이터를 이용하여 학습 과정을 반복한다.
- [0054] 도 3은 이종 데이터 세트를 이용하여 구축된 인공지능 모델의 추론 과정(300)의 예이다. 도 3의 인공지능 모델은 도 2에서 학습된 특징 추출기 1(E1), 특징 추출기 2(E2) 및 분류기 3(C3)을 포함한다.
- [0055] 분석장치는 분석 대상인 이종 데이터 세트를 입력받는다. 도 3에서 이종 데이터 세트는 개인 데이터 세트 1 및 개인 데이터 세트 2로 구성된다.
- [0056] 분석장치는 개인 데이터 세트들을 비식별화한다(310). 비식별화 과정은 도 2에서 설명한 바와 같다. 분석장치는 생성기를 이용하여 개인 데이터 세트를 비식별화할 수 있다. 생성기는 DPGAN과 같은 생성 모델일 수 있다. 분석장치는 개인 데이터 세트별로 서로 다른 생성기를 이용하여 데이터 세트를 비식별화할 수 있다. 분석장치는 개인 데이터 세트 1을 생성기 1에 입력하여 비식별화된 데이터 세트 1을 산출할 수 있다. 또한, 분석장치는 개인 데이터 세트 2를 생성기 2에 입력하여 비식별화된 데이터 세트 2를 산출할 수 있다. 또는 경우에 따라서, 분석장치는 동일한 생성기를 이용하여 개인 데이터 세트 1 및 개인 데이터 세트 2를 각각 비식별화된 데이터 세트로 변환할 수도 있다.
- [0057] 분석장치는 비식별화된 데이터 세트 1을 특징 추출기 1(E1)에 입력하고, 비식별화된 데이터 세트 2를 특징 추출기 2(E2)에 입력한다. 학습장치는 특징 추출기 1(E1)이 출력하는 특징 1과 특징 추출기 2(E2)가 출력하는 특징 2를 결합한다.
- [0058] 분석장치는 결합된 특징 벡터를 분류기 3(C3)에 입력하고 분류기 3(C3)이 출력하는 출력값을 기준으로 일정한 정보를 추출할 수 있다. 나아가, 분석장치는 추출한 정보를 기준으로 사용자에게 대한 맞춤형 서비스 정보를 생성할 수도 있다.
- [0060] 도 4는 인공지능 모델을 이용하여 이종 데이터 세트를 분석하는 분석장치(400)의 예이다. 분석장치(400)는 도 2의 과정을 통해 학습된 인공지능 모델을 이용하여 이종 데이터 세트를 분석하는 장치이다. 분석장치(400)는 도 3의 과정과 같은 동작을 수행한다. 분석장치(400)는 스마트기기, PC, 네트워크상의 서버 등과 같이 다양한 형태 중 어느 하나로 구현될 수 있다.
- [0061] 분석장치(400)는 저장 장치(410), 메모리(420), 연산장치(430), 인터페이스 장치(440), 통신 장치(450) 및 출력장치(460)를 포함할 수 있다.
- [0062] 저장 장치(410)는 도 2의 과정을 통해 학습된 인공지능 모델을 저장한다.

- [0063] 저장 장치(410)는 개인 데이터 세트 비식별화를 위한 생성모델을 저장한다.
- [0064] 저장 장치(410)는 분석 대상인 이중 데이터 세트를 저장할 수 있다.
- [0065] 저장 장치(410)는 입력 데이터 전처리, 추론 과정은 인공지능 모델 동작 제어, 추론 과정의 데이터 처리 등을 제어하는 소스 코드 내지 프로그램을 저장할 수 있다.
- [0066] 메모리(420)는 분석장치(400)가 인공지능 모델을 이용하여 이중 데이터 세트를 분석하는 과정에서 발생하는 데이터를 저장할 수 있다.
- [0067] 인터페이스 장치(440)는 일정한 정보 내지 데이터를 입력받는 장치이다. 인터페이스 장치(440)는 외부 객체(외부 저장 장치 등)로부터 이중 데이터 세트를 입력받일 수 있다. 인터페이스 장치(440)는 이중 데이터 세트를 분석한 결과 내지 서비스 정보를 외부로 전달할 수도 있다.
- [0068] 통신 장치(450)는 유선 또는 무선 네트워크를 통해 일정한 정보를 수신하고 전송하는 구성을 의미한다. 통신 장치(450)는 외부 객체로부터 이중 데이터 세트를 수신할 수 있다. 통신 장치(450)는 이중 데이터 세트를 분석한 결과 내지 서비스 정보를 외부 객체에 송신할 수도 있다.
- [0069] 나아가, 인터페이스 장치(440)는 통신 장치(450)를 통해 수신하는 정보 내지 데이터를 분석장치(400) 내부로 전달하는 구성을 지칭할 수도 있다.
- [0070] 연산 장치(430)는 이중 데이터 세트를 비식별화한다. 연산 장치(430)는 도 2에서 설명한 바와 같이 일정한 생성 모델을 이용하여 개인 데이터 세트 각각을 비식별화할 수 있다. 생성 모델은 DPGAN과 같은 모델일 수 있다. 연산 장치(430)는 개인 데이터 세트별로 서로 다른 생성 모델을 이용하여 데이터 세트를 비식별화할 수 있다. 경우에 따라서, 연산 장치(430)는 동일한 생성 모델을 이용하여 개인 데이터 세트들 각각을 각각 비식별화된 데이터 세트로 변환할 수도 있다.
- [0071] 연산 장치(430)는 비식별화된 데이터 세트 각각을 개별 특징 분류기에 입력한다. 예컨대, 도 3과 같이 이중 데이터 세트가 개인 데이터 세트 1 및 개인 데이터 세트 2로 구성된다면, 연산 장치(430)는 비식별화된 데이터 세트 1을 특징 추출기 1(E1)에 입력하고, 비식별화된 데이터 세트 2를 특징 추출기 2(E2)에 입력한다.
- [0072] 연산 장치(430)는 개별 특징 추출기가 출력하는 특징을 결합한다. 예컨대, 도 3과 같이 이중 데이터 세트가 개인 데이터 세트 1 및 개인 데이터 세트 2로 구성된다면, 연산 장치(430)는 특징 추출기 1(E1)이 출력하는 특징 1과 특징 추출기 2(E2)가 출력하는 특징 2를 결합한다.
- [0073] 연산 장치(430)는 결합된 특징 벡터를 분류기(예컨대, 도 3의 C3)에 입력하고 분류기 3(C3)이 출력하는 출력값을 기준으로 일정한 정보를 추출할 수 있다. 나아가, 연산 장치(430)는 추출한 정보를 기준으로 사용자에게 대한 맞춤형 서비스 정보를 생성할 수도 있다.
- [0074] 연산 장치(430)는 데이터를 처리하고, 일정한 연산을 처리하는 프로세서, AP, 프로그램이 임베디드된 칩과 같은 장치일 수 있다.
- [0075] 출력 장치(460)는 이중 데이터 세트 분석 과정에 필요한 인터페이스 화면을 출력할 수 있다. 또한, 출력 장치(460)는 이중 데이터 세트를 분석한 결과 또는 사용자에게 대한 서비스 정보를 출력할 수 있다.
- [0077] 또한, 상술한 바와 같은 다수의 데이터 세트를 이용한 인공지능 모델 학습 방법 또는 해당 인공지능 모델을 이용한 추론 방법은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.
- [0078] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM(read-only memory), PROM(programmable read only memory), EPROM(Erasable PROM, EPROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.
- [0079] 일시적 판독 가능 매체는 스태틱 램(Static RAM, SRAM), 다이내믹 램(Dynamic RAM, DRAM), 싱크로너스 디램

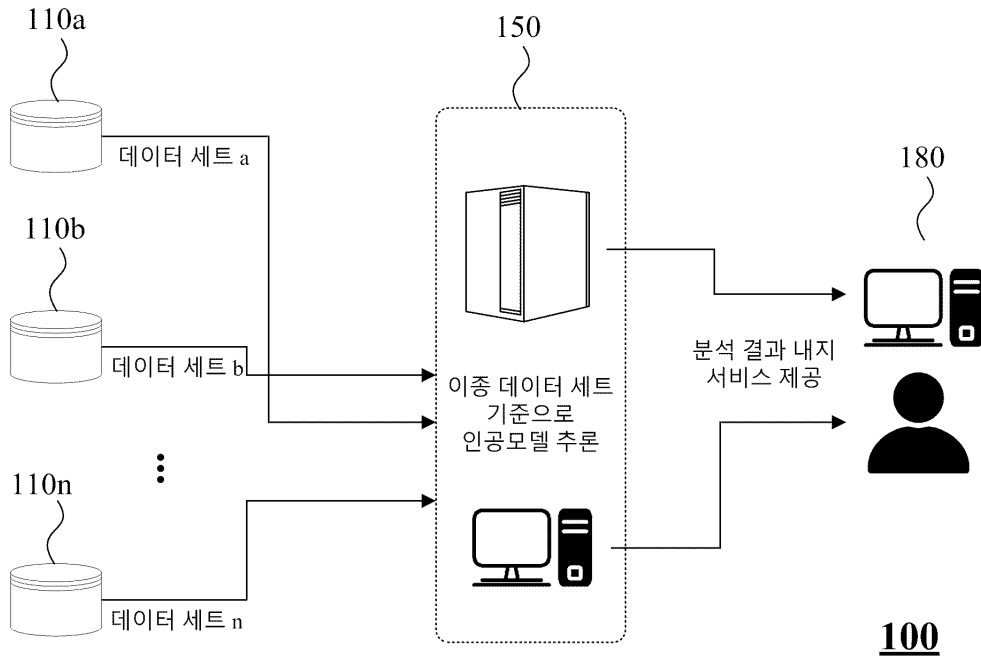
(Synchronous DRAM, SDRAM), 2배속 SDRAM(Double Data Rate SDRAM, DDR SDRAM), 증강형 SDRAM(Enhanced SDRAM, ESDRAM), 동기화 DRAM(SyncLink DRAM, SLDRAM) 및 직접 램버스 램(Direct Rambus RAM, DRRAM) 과 같은 다양한 RAM을 의미한다.

[0080]

본 실시예 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

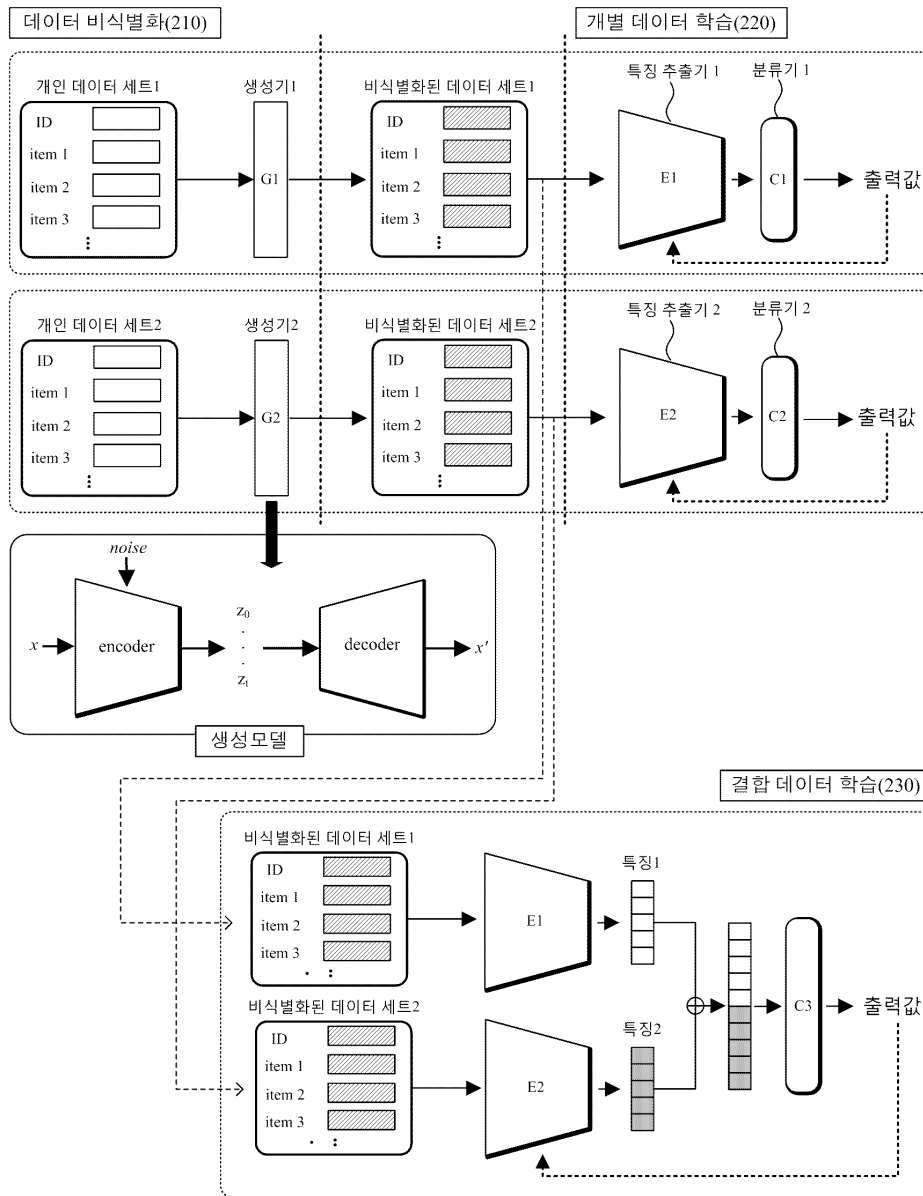
도면

도면1

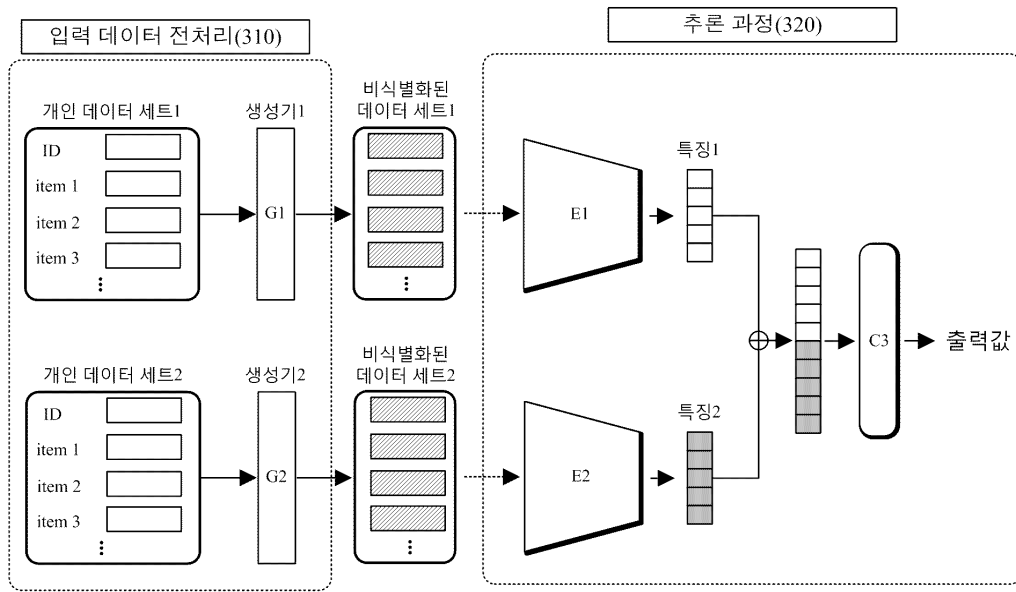


도면2

200



도면3



300

도면4

