(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0011391
(43) 공개일자 2025년01월21일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2023.01) *G06N 3/0455* (2023.01)
(52) CPC특허분류
G06N 3/082 (2023.01)
G06N 3/0455 (2023.01)
(21) 출원번호 10-2023-0091751
(22) 출원일자 2023년07월14일
심사청구일자 2023년07월14일

(71) 출원인
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이화여자대학교)
(72) 발명자
박형곤
경기도 고양시 일산동구 일산로 241, 110동 204호(마두동,백마마을1단지아파트)
이주홍
서울특별시 은평구 가좌로9길 19-1, 401호(응암동, 다베로)
최다영
서울특별시 서대문구 신촌로9길 47, 303호 (창천동)
(74) 대리인
특허법인(유한)아이시스

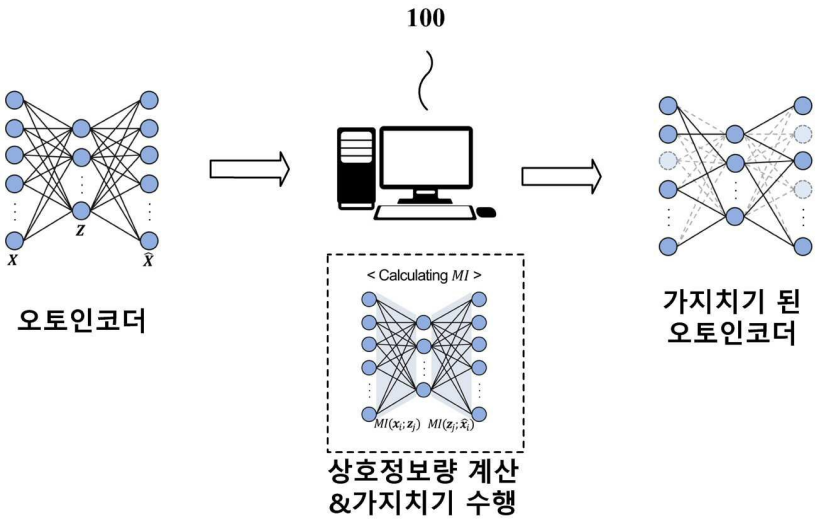
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **상호정보량 기반 오토인코더 경량화 방법 및 장치**

(57) 요 약

상호 정보량 기반 오토인코더 경량화 방법은 경량화 장치가 오토인코더를 입력 받는 단계; 상기 경량화 장치가 상기 오토인코더의 계층 중 가지치기 대상 계층을 선택하는 단계; 상기 경량화 장치가 상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산하는 단계; 상기 경량화 장치가 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기하는 단계; 및 상기 경량화 장치가 상기 가지치기 된 오토인코더를 미세 조정하는 단계; 를 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G06N 3/088 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193461
과제번호	2021-0-00739-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	방송통신산업기술개발(R&D,정보화)
연구과제명	분산/협력 AI 기반 5G+ 네트워크 데이터 분석 기능및 제어 기술 개발
기 여 율	1/1
과제수행기관명	이화여자대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명 세 서

청구범위

청구항 1

경량화 장치가 오토인코더를 입력 받는 단계;

상기 경량화 장치가 상기 오토인코더의 계층 중 가지치기 대상 계층을 선택하는 단계;

상기 경량화 장치가 상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산하는 단계;

상기 경량화 장치가 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기하는 단계; 및

상기 경량화 장치가 상기 가지치기 된 오토인코더를 미세 조정하는 단계; 를 포함하는, 상호정보량 기반 오토인코더 경량화 방법.

청구항 2

제1항에 있어서,

상기 가지치기 대상 계층을 선택하는 것은 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층 중 오토인코더의 잠재 벡터(latent vector)와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것을 포함하는, 상호정보량 기반 오토인코더 경량화 방법.

청구항 3

제1항에 있어서,

상기 경량화 장치가 상기 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단하는 단계; 및

상기 미세 조정 된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되지 못했다고 판단하는 경우, 상기 경량화 장치가 상기 미세 조정된 오토인코더를 대상으로 상기 가지치기 대상 계층을 선택하는 것 상기 링크의 중요도를 계산하는 것, 상기 가지치기 하는 것 및 상기 미세 조정하는 것을 다시 수행하는 단계; 를 더 포함하는, 상호정보량 기반 오토인코더 경량화 방법.

청구항 4

제1항에 있어서,

상기 상호정보량은 하기 수학식1을 통해 계산된 값인, 상호정보량 기반 오토인코더 경량화 방법.

[수학식1]

$$MI(\mathbf{z}_i^{(l-1)};\mathbf{z}_j^{(l)}) = \sum_{s=1}^{d} p(z_{si}^{(l-1)}, z_{sj}^{(l)}) \log \frac{p(z_{si}^{(l-1)}, z_{sj}^{(l)})}{p(z_{si}^{(l-1)})p(z_{sj}^{(l)})}$$

상기 수학식1에서 $z_i^{(l)}$ 는 l번째 계층에서 i번째 노드를 의미한다. 상기 $z_i^{(l)}$ 는 $[z_{11}^{(l)}, z_{21}^{(l)}, z_{31}^{(l)}, \cdots, z_{d1}^{(l)}]$로 표현될 수 있다. 상기 수학식1에서 d는 전체 데이터의 개수를 의미한다. 상기 수학식1에서 $MI(z_i^{(l-1)}; z_j^{(l)})$는 $(l-1)$번째 층의 i번째 노드인 $z_i^{(l-1)}$ 와 l번째 층의 j번째 노드인 $z_j^{(l)}$ 사이의 상호정보량를 의미한다.

청구항 5

제2항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층을 먼저 모두 선택 한 후, 디코더에 속한 계층을 선택하는 순서로 진행되는 것인, 상호정보량 기반 오토인코더 경량화 방법.

청구항 6

제2항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 디코더에 속한 계층을 먼저 모두 선택 한 후, 인코더에 속한 계층을 선택하는 순서로 진행되는 것인, 상호 정보량 기반 오토인코더 경량화 방법.

청구항 7

제2항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층과 디코더에 속한 계층을 번갈아가며 선택하는 순서로 진행되는 것인, 상호정보량 기반 오토인코더 경량화 방법.

청구항 8

제1항에 있어서,

상기 가지치기하는 것은 가지치기 비율에 따라 중요도가 낮은 링크를 제거하는 것인, 상호정보량 기반 오토인코더 경량화 방법.

청구항 9

오토인코더를 입력 받는 입력장치;

오토인코더의 계층 중 가지치기 대상 계층을 선택하고, 상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산하고, 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기하고, 상기 가지치기 된 오토인코더를 미세 조정하는, 연산장치; 및

상기 오토인코더를 저장하는 저장장치;를 포함하는, 상호정보량 기반 오토인코더 경량화 장치.

청구항 10

제9항에 있어서,

상기 가지치기 대상 계층을 선택하는 것은 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층 중 오토인코더의 잠재 벡터(Latent vector)와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것을 포함하는, 상호정보량 기반 오토인코더 경량화 장치

청구항 11

제9항에 있어서,

상기 연산장치는 상기 미세 조정 된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단하고, 상기 미세 조정 된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되지 못했다고 판단하는 경우 상기 경량화 장치가 상기 미세 조정된 오토인코더를 대상으로 상기 가지치기 대상 계층을 선택하는 것, 상기 링크의 중요도를 계산하는 것, 상기 가지치기 하는 것 및 상기 미세 조정 하는 것을 다시 수행하는, 상호 정보량 기반 오토인코더 경량화 장치.

청구항 12

제9항에 있어서,

상기 상호정보량은 하기 수학식1을 통해 계산된 값인, 상호정보량 기반 오토인코더 경량화 장치.

[수학식1]

$$MI(\mathbf{z}_i^{(l-1)};\mathbf{z}_j^{(l)}) = \sum_{s=1}^{d} p(z_{si}^{(l-1)}, z_{sj}^{(l)}) \log \frac{p(z_{si}^{(l-1)}, z_{sj}^{(l)})}{p(z_{si}^{(l-1)})p(z_{sj}^{(l)})}$$

상기 수학식1에서 $z_i^{(l)}$ 는 l번째 계층에서 i번째 노드를 의미한다. 상기 $z_i^{(l)}$ 는 [$z_{11}^{(l)}, z_{21}^{(l)}, z_{31}^{(l)}, \cdots, z_{d1}^{(l)}$]로 표현될 수 있다. 상기 수학식1에서 d는 전체 데이터의 개수를 의미한다. 상기 수학식1에서 $MI(z_i^{(l-1)}; z_j^{(l)})$는 $(l-1)$ 번째 층의 i번째 노드인 $z_i^{(l-1)}$ 와 l번째 층의 j번째 노드인 $z_j^{(l)}$ 사이의 상호정보량를 의미한다.

청구항 13

제10항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층을 먼저 모두 선택 한 후, 디코더에 속한 계층을 선택하는 순서로 진행되는 것인, 상호정보량 기반 오토인코더 경량화 장치.

청구항 14

제10항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 디코더에 속한 계층을 먼저 모두 선택 한 후, 인코더에 속한 계층을 선택하는 순서로 진행되는 것인, 상호 정보량 기반 오토인코더 경량화 장치.

청구항 15

제10항에 있어서,

상기 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층과 디코더에 속한 계층을 번갈아가며 선택하는 순서로 진행되는 것인, 상호정보량 기반 오토인코더 경량화 장치.

청구항 16

제9항에 있어서,

상기 가지치기하는 것은 가지치기 비율에 따라 중요도가 낮은 링크를 제거하는 것인, 상호정보량 기반 오토인코더 경량화 장치.

청구항 17

제1항에 기재된 상호정보량 기반 오토인코더 경량화 방법을 실행하기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

청구항 18

이상 탐지 장치가 경량화된 오토인코더에 분석 대상 데이터를 입력하는 단계; 및

상기 이상 탐지 장치가 상기 경량화된 오토인코더의 출력값을 기반으로 이상이 있는지 탐지 하는 단계;를 포함하되,

상기 경량화된 오토인코더는 정상 데이터만 학습 데이터로 하여 학습된 후, 제1 항에 기재된 상호정보량 기반 오토인코더 경량화 방법을 이용해 경량화된 오토인코더인, 경량화된 오토인코더를 이용해 이상을 탐지 하는 방법.

청구항 19

제18항에 있어서,

상기 오토인코더의 출력값을 기반으로 이상이 있는지 탐지하는 것은 이상치 점수(anomaly score)를 계산 한 뒤, 상기 이상치 점수가 미리 설정된 임계값 이상이라면 이상이 있는 경우로 판단하는 것인, 경량화된 오토인코더를 이용해 이상을 탐지 하는 방법.

청구항 20

제1항에 기재된 상호정보량 기반 오토인코더 경량화 방법을 통해 경량화된 오토인코더를 이용해 데이터를 압축, 데이터를 분류, 데이터 생성, 데이터 변환 또는 노이즈 제거하는 방법.

발명의 설명

기 술 분 야

[0001] 이하 설명하는 기술은 상호정보량을 기반으로 오토인코더를 경량화 하는 방법에 대한 것이다.

배 경 기 술

[0002] 심층 신경망(Deep Neural Network, DNN) 구조의 딥러닝 모델은 선형 연산만을 수행하는 모델에 비해 높은 정확도를 가진다. 오토인코더(AutoEncoder)는 딥러닝 모델 중에 하나이다. 오토인코더는 비지도학습 기반 모델 중 하나이다. 오토인코더는 입력 데이터를 압축한 뒤 이를 다시 복원하도록 학습된다. 오토인코더는 이미지를 압축하거나 잡음을 제거하거나 이미지를 재구축하거나 이상을 탐지하는 등에 활용된다.

선행기술문헌

특허문헌

[0003] (특허문헌 0001) 공개특허 10-2022-0099875

발명의 내용

해결하려는 과제

[0004] 오토인코더를 포함한 딥러닝 모델은 많은 자원이 필요하다. 따라서 딥러닝 모델은 연산 자원(computational resource), 저장 공간 및 사용 가능한 에너지양이 제한적인 IoT(Internet of Things) 기기나 모바일 장치에 탑재되는데 어려움이 있다. 이를 해결하기 위하여 딥러닝을 경량화 하는 방법들이 개발되고 있다. 딥러닝 모델을 경량화 하는 방법으로는 모델을 양자화 하거나 불필요한 가중치를 제거하는 가지치기(pruning) 방법들이 있다.

[0005] 이하 설명하는 기술은 상호정보량을 기반으로 딥러닝 모델을 경량화 하는 방법을 제공하고자 한다. 특히 딥러닝 모델 중 오토인코더에 특화된 경량화 방법을 제공하고자 한다.

과제의 해결 수단

[0006] 이하 설명하는 기술은 전술한 과제를 해결하기 위해 상호정보량 기반 오토인코더 경량화 방법을 개시한다.

[0007] 상호 정보량 기반 오토인코더 경량화 방법은 경량화 장치가 오토인코더를 입력 받는 단계; 상기 경량화 장치가 상기 오토인코더의 계층 중 가지치기 대상 계층을 선택하는 단계; 상기 경량화 장치가 상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산하는 단계; 상기 경량화 장치가 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기하는 단계; 및 상기 경량화 장치가 상기 가지치기 된 오토인코더를 미세 조정하는 단계; 를 포함한다.

[0008] 상기 가지치기 대상 계층을 선택하는 것은 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층 중 오토인코더의 잠재 벡터(latent vector)와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것을 포함한다.

발명의 효과

[0009] 이하 설명하는 기술을 이용하면 오토인코더를 경량화 시킬 수 있다. 특히 딥러닝 모델 중 오토인코더에 특화된

경량화 방법을 통해, 오토인코더를 효율적으로 경량화 시킬수 있다. 이를 통해 연산자원과 저장공간 및 사용 가능한 에너지양이 제한적인 기기에서도 오토인코더를 작동시킬 수 있다.

[0010] 이하 설명하는 기술을 이용하면 경량화된 오토인코더를 활용할 수 있다. 구체적으로 경량화된 오토인코더를 이용해 이상을 탐지할 수 있다.

[0011] 이하 설명하는 기술을 이용하면 네트워크 트래픽의 이상을 탐지할 수 있다. 이에 적은 연산량을 요구하는 IoT기기나 모바일 기기에서도 경량화된 오토인코더를 작동시킴으로써 낮은 탐지 복잡도로 우수한 성능의 이상을 탐지할 수 있다.

도면의 간단한 설명

[0012] 도1은 경량화 장치(100)가 오토인코더를 경량화하는 전체적인 과정을 보여준다.

도2는 오토인코더 경량화 방법의 실시예 중 하나의 순서도(200)이다.

도3은 오토인코더의 실시예 중 하나이다.

도4 및 도5는 미리 정해둔 순서의 실시예 중 하나를 보여준다.

도6은 경량화된 오토인코더를 이용해 이상을 탐지하는 방법의 실시예 중 하나의 순서도(300)이다.

도7은 경량화 되지 않은 오토인코더와 50% 경량화된 오토인코더의 이상치 점수(Anomaly score) 히스토그램에 대한 것이다.

도8은 가지치기 비율(Pruning Ratio)에 따른 오토인코더의 이상탐지 성능을 평가한 결과이다.

도9는 여러 가지치기 방법을 적용해 오토인코더를 경량화 한 후 성능을 평가한 결과이다.

도10은 경량화 장치의 실시예 중 하나의 구성이다.

발명을 실시하기 위한 구체적인 내용

[0013] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있다. 명세서의 도면에 이하 설명하는 기술의 특정 실시 형태가 기재될 수 있다. 그러나, 이는 이하 설명하는 기술의 설명을 위한 것이며 이하 설명하는 기술을 특정한 실시 형태에 대해 한정하려는 것이 아니다. 따라서 이하 설명하는 기술의 사상 및 기술 범위에 포함되는 모든 변경 물, 균등 물 내지 대체 물이 이하 설명하는 기술에 포함하는 것으로 이해되어야 한다.

[0014] 이하 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 기재된 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.

[0015] 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 구성 부 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.

[0016] 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.

[0018] 이하 오토인코더 경량화 장치(이하, 경량화 장치)가 오토인코더를 경량화 하는 전체적인 과정을 설명한다.

[0019] 도1은 경량화 장치(100)가 오토인코더를 경량화하는 전체적인 과정을 보여준다.

[0020] 경량화 장치(100)는 오토인코더를 입력 받을 수 있다. 경량화 장치(100)는 오토인코더의 계층 중 가지치기 대상

계층을 선택할 수 있다. 경량화 장치(100)는 상호정보량을 기반으로 가지치기 대상 계층과 연결된 링크의 중요도를 계산할 수 있다. 경량화 장치(100)는 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기할 수 있다. 경량화 장치(100)는 상기 가지치기된 오토인코더를 미세 조정할 수 있다.

[0021] 경량화 장치(100)는 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단할 수 있다. 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되지 못했다고 판단하는 경우, 경량화 장치(100)는 미세 조정된 오토인코더를 대상으로 상기 가지치기 대상 계층을 선택하는 것, 상기 링크의 중요도를 계산하는 것, 상기 가지치기 하는 것 및 상기 미세 조정 하는 것을 다시 수행할 수 있다.

[0022] 가지치기 대상 계층을 선택하는 것은 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층 중 오토인코더의 잠재 벡터(latent vector)와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것을 포함할 수 있다.

[0024] 이하 오토인코더 경량화 방법을 구체적으로 설명한다.

[0025] 도2는 오토인코더 경량화 방법의 실시예 중 하나의 순서도(200)이다.

[0026] 경량화 장치는 오토인코더를 입력 받을 수 있다(210).

[0027] 오토인코더의 실시예 중 하나는 도3에 나타내었다.

[0028] 오토인코더는 입력 계층(input layer), 은닉 계층(hidden layer) 및 출력 계층(output layer)을 포함할 수 있다. 입력 계층은 입력 데이터($\boldsymbol{X}$)를 받는 계층이다. 은닉 계층은 입력 계층과 출력 계층을 연결하는 계층일 수 있다. 출력 계층은 출력 데이터($\hat{\boldsymbol{X}}$)를 출력하는 계층일 수 있다. 도3의 오토인코더는 인코더의 L개의 계층, 디코더의 L개의 계층이 있어, 총 2L개의 계층을 가진다.

[0029] 오토인코더의 각 계층에는 노드(node)들이 포함되어 있다. 노드들은 이전 계층의 노드 및 이후 계층의 노드와 연결되어 있을 수 있다. 링크(link)는 노드 사이를 연결할 수 있다. 링크는 서로 다른 가중치(w_j^i)를 가질 수 있다

[0030] 오토인코더는 비지도 학습(unsupervised learning) 기반 모델일 수 있다. 오토인코더는 입력값과 최대한 비슷한 출력값을 출력하도록 학습될 수 있다. 따라서 입력층의 차원이 N이라면 출력층의 차원도 N이다.

[0031] 오토인코더는 입력값을 압축한 뒤, 압축된 값을 기반으로 입력값과 최대한 비슷하게 복원하여 출력값을 만들 수 수 있다. 이에 은닉층은 입력층 및 출력층의 차원보다 낮은 차원을 가질 수 있다.

[0032] 오토인코더는 인코더(encoder) 및 디코더(decoder)로 이루어질 수 있다. 인코더는 입력값을 잠재 벡터(latent vector, $\boldsymbol{Z}^{(L)}$) 로 압축할 수 있다. 디코더는 잠재 벡터를 출력값으로 복원할 수 있다.

[0033] 잠재 벡터는 입력 데이터의 특징이 압축된 정보를 포함할 수 있다. 오토인코더는 입력 데이터의 특징이 잘 압축된 잠재 벡터를 획득하는 것을 목적으로 한다.

[0034] 경량화 장치는 상기 오토인코더의 계층 중 가지치기 대상 계층을 선택할 수 있다(220).

[0035] 경량화 장치는 한번에 하나의 계층에 대해서만 가지치기를 수행한 후 미세 조정을 한다. 즉 가지치기 대상 계층은 경량화 장치가 한번의 가지치기를 수행할 때 가지치기가 되는 계층을 의미할 수 있다. 한번의 가지치기시 하나의 계층에 대해서만 수행하기 때문에 오토인코더의 모든 계층에 대해 가지치기를 하기 위해선 가지치기는 반복해서 수행될 필요가 있다.

[0036] 가지치기 대상을 선택하는 것은 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층 중 오토인코더의 잠재 벡터(latent vector)와 가장 먼 계층을 선택하는 단계를 포함할 수 있다. 잠재 벡터는 오토인코더의 입력 데이터를 압축한 것으로, 입력 데이터의 특징이 가장 많이 압축되어 있다. 따라서 잠재 벡터와 먼 거리에 있는 계층부터 가지치기를 수행함으로써, 중요한 정보가 포함된 잠재 벡터의 정보를 최대한 유지한 채 가지치기를 수행할 수 있다.

[0037] 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택할 때 미리 정해둔 순서에 따라 가지치

기 대상 계층을 선택할 수 있다.

[0038] 도4 및 도5는 미리 정해둔 순서의 실시예 중 하나를 보여준다.

[0039] 일 실시예로 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층을 먼저 모두 선택 한 후, 디코더에 속한 계층을 선택하는 순서로 진행되는 것일 수 있다. 즉 도4의 (a) 처럼 인코더에 속한 계층 중 잠재 벡터와 가장 먼 거리에 있는 계층을 차례로 선택하여 경량화를 먼저 진행할 수 있다. 이후 인코더에 속한 계층이 모두 선택되면 디코더에 속한 계층 중 잠재 벡터와 가장 먼거리에 있는 계층을 차례로 선택하여 경량화를 진행할 수 있다.

[0040] 일 실시예로 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 디코더에 속한 계층을 먼저 모두 선택 한 후, 인코더에 속한 계층을 선택하는 순서로 진행되는 것일 수 있다. 즉 도4의 (b) 처럼 디코더에 속한 계층 중 잠재 벡터와 가장 먼 거리에 있는 계층을 차례로 선택하여 경량화를 먼저 진행할 수 있다. 이후 디코더에 속한 계층이 모두 선택되면 인코더에 속한 계층 중 잠재 벡터와 가장 먼거리에 있는 계층을 차례로 선택하여 경량화를 진행할 수 있다.

[0041] 일 실시예로 오토인코더의 잠재 벡터와 가장 먼 계층을 가지치기 대상 계층으로 선택하는 것은 인코더에 속한 계층과 디코더에 속한 계층을 번갈아가며 선택하는 순서로 진행되는 것일 수 있다. 즉 도5의 (a) 처럼 인코더에 속한 계층을 먼저 선택하고 이후 디코더에 속한 계층을 선택하는 것을 반복하면서, 인코더와 디코더를 번갈아가면서 선택할 수 있다. 또는 도5의 (b) 처럼 디코더에 속한 계층을 먼저 선택하고 이후 인코더에 속한 계층을 선택하는 것을 반복하면서, 인코더와 디코더를 번갈아가면서 선택할 수 있다.

[0042] 경량화 장치는 상호정보량(Mutual Information, MI)을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산할 수 있다(230).

[0043] 상호정보량은 일반적인 정보 이론에서 사용되는 개념일 수 있다. 상호정보량은 두 개의 확률 변수 사이의 상호 의존성을 측정하는데 사용될 수 있다. 즉 상호정보량은 두 변수가 얼마나 상관관계가 있는지, 얼마나 함께 변화하는지를 나타낼 수 있다.

[0044] 오토인코더에서 상호정보량은 각 계층을 통해 신호가 전파될 때, 이전 계층의 노드들이 다음 계층의 노드들에 미치는 영향을 정량화 한 값일 수 있다. 즉 상호정보량이 클수록 계층 간 노드를 연결하는 가중치의 중요도가 높을 수 있다.

[0045] 수학식1은 상호정보량을 계산할 때 이용하는 식이다.

수학식 1

[0046] $$MI(\mathbf{z}_i^{(l-1)};\mathbf{z}_j^{(l)}) = \sum_{s=1}^{d} p(z_{si}^{(l-1)}, z_{sj}^{(l)}) \log \frac{p(z_{si}^{(l-1)}, z_{sj}^{(l)})}{p(z_{si}^{(l-1)})p(z_{sj}^{(l)})}$$

[0047] 수학식1에서 $z_i^{(l)}$ 는 l번째 계층에서 i번째 노드를 의미한다. $z_i^{(l)}$ 는 $[z_{11}^{(l)}, z_{21}^{(l)}, z_{31}^{(l)}, \cdots, z_{d1}^{(l)}]$로 표현될 수 있다. 수학식1에서 d는 전체 데이터의 개수를 의미한다. 수학식1에서 $MI(z_i^{(l-1)} ; z_j^{(l)})$는 $(l-1)$번째 층의 i번째 노드인 $z_i^{(l-1)}$와 l번째 층의 j번째 노드인 $z_j^{(l)}$ 사이의 상호정보량를 의미한다.

[0048] 경량화 장치는 가지치기 대상 계층에 대해서만 링크의 중요도를 계산할 수 있다. 즉 한번의 미세 조정이 끝난 후, 다시 링크의 중요도 등을 계산할 때 이전에 가지치기 대상 계층으로 선택된 계층에 대해서는 링크의 중요도를 계산하지 아니한다.

[0049] 경량화 장치는 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기할 수 있다(240).

[0050] 가지치기하는 것은 링크의 가중치 값을 0으로 대체하는 것을 의미할 수 있다. 이를 통해 해당 경로로 신호가 전파되는 것을 비활성화 할 수 있다.

[0051] 경량화 장치는 가지치기 비율에 따라 중요도가 낮은 링크를 제거할 수 있다. 가지치기 비율은 링크를 얼마만큼 가지치기할 것인지 결정할 때 이용되는 비율일 수 있다. 가지치기 비율은 목표하는 정확도에 따라 임의로 설정될 수 있다. 가지치기 비율은 0 내지 1 사이의 값을 가질 수 있다. 가지치기 비율은 모든 링크를 가지치기하는 경우 1의 값을 가지며, 모든 링크를 가지치기 하지 않는 경우 0의 값을 가질 수 있다. 예를 들어 가지치기 비율이 0.7인 경우 계산된 중요도 중 70% 이하의 중요도를 가지는 링크를 제거할 수 있다. 따라서 가지치기 비율에 따라 상호정보량 값이 큰 가중치를 가지는 링크만을 활성화할 수 있다.

[0052] 경량화 장치는 가지치기 대상 계층이 인코더 및 디코더에 속하느냐에 따라 가지치기 대상이 될 링크를 선택할 수 있다.

[0053] 일 실시예로 가지치기 대상 계층이 인코더에 속하는 계층인 경우 가지치기 대상 계층과 이후 계층을 연결하는 링크가 가지치기의 대상이 될 수 있다. 또는 가지치기 대상 계층이 디코더에 속하는 계층인 경우 가지치기 대상 계층과 이전 계층을 연결하는 링크가 가지치기 대상이 될 수 있다.

[0054] 경량화 장치는 상기 가지치기 된 오토인코더를 미세 조정(fine-tuning)할 수 있다(250).

[0055] 미세 조정은 가지치기된 부분의 가중치는 0으로 유지하면서, 오토인코더의 파라미터를 최적화하는 것을 의미할 수 있다. 구체적으로 미세 조정은 이전 단계에서 가지치기되어 가중치가 0으로 된 부분의 파라미터는 0으로 그대로 유지하되, 학습데이터를 이용해 가지치기 되지 아니하여 가중치가 0이 아닌 나머지 부분의 파라미터를 조정하는 것을 의미할 수 있다. 예를 들어 2번째 계층의 2번 노드 및 3번재 계층의 3번째 노드와 연결된 링크가 이전 단계에서 가지치기 된 경우, 해당 링크의 가중치를 0으로 유지하면서, 학습데이터를 통해 오토인코더의 나머지 파라미터를 조정할 수 있다.

[0056] 경량화 장치는 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단할 수 있다(260).

[0057] 전술한 바와 같이 경량화 장치는 한번에 하나의 가지치기 대상 계층을 선택하여 가지치기와 미세 조정을 수행한다. 이에 오토인코더의 모든 계층이 가지치기되기 위해선 위 과정을 반복할 필요가 있다. 따라서 경량화 장치는 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단할 수 있다. 이때 잠재 벡터와 관련된 계층은 가지치기 대상 계층으로 선택되었는지 판단하지 않을 수도 있다. 잠재벡터와 연결된 링크가 이전에 가지치기 되었기 때문이다.

[0058] 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되지 못했다고 판단하는 경우, 경량화 장치는 미세 조정된 오토인코더를 대상으로 이전에 가지치기 대상으로 선택된 적 없는 계층과 연결된 상기 링크의 중요도를 계산하는 것, 상기 가지치기 대상 계층을 선택하는 것, 상기 가지치기 하는 것 및 상기 미세 조정 하는 것을 다시 수행할 수 있다(270).

[0059] 즉 미세 조정된 오토인코더를 기반으로 전술한 단계(220 내지 250) 다시 수행할 수 있다. 이때 전술한 바와 같이 가지치기 대상 계층을 선택할 때에(230)는 이전에 가지치기 대상 계층으로 선택된 계층을 제외한 나머지 계층을 가지치기 대상 계층으로 선택할 수 있다. 즉 한번 가지치기 대상 계층으로 선택된 경우 다시 가지치기 대상 계층으로 선택되지 아니한다.

[0061] 전술한 방법을 이용하면 오토인코더를 경량화시킬 수 있다. 경량화된 오토인코더는 데이터의 분류, 데이터 압축, 데이터 생성, 데이터 변환, 노이즈 제거, 이상 탐지 등 다양한 분야에 활용될 수 있다.

[0062] 이하 경량화된 오토인코더를 이용하는 예시 중 하나로, 이상을 탐지 하는 방법을 구체적으로 살펴본다.

[0063] 도6은 경량화된 오토인코더를 이용해 이상을 탐지하는 방법의 실시예 중 하나의 순서도(300)이다.

[0064] 이상 탐지 장치는 정상 데이터를 학습 데이터로 하여 오토인코더를 학습할 수 있다(310).

[0065] 정상 데이터는 이상이 없는 상태에서 측정된 데이터일 수 있다. 즉 오토인코더는 이상이 없는 상태에서 측정된 데이터만을 기반으로 학습될 수 있다.

[0066] 이상 탐지 장치는 학습이 완료된 오토인코더를 경량화 할 수 있다(320).

[0067] 오토인코더를 경량화 하기 위해 앞서 도1 내지 도5등을 통해 설명한 오토인코더 경량화 방법을 이용할 수 있다.

[0068] 이상 탐지 장치는 경량화된 오토인코더에 분석 대상 데이터를 입력할 수 있다(330).

[0069] 분석 대상 데이터는 이상이 있는지 탐지하고자 하는 대상에서 측정한 데이터를 의미할 수 있다. 일 실시예로 네트워크에 이상이 있는지 탐지하기 위해, 네트워크 트래픽 데이터를 측정하여 분석 대상 데이터로 이용할 수 있다.

[0070] 이상 탐지 장치는 오토인코더의 출력값을 기반으로 이상이 있는지 탐지할 수 있다(340).

[0071] 오토인코더는 정상 데이터만을 기반으로 학습되었기 때문에, 이상이 있는 데이터가 입력되면 복원을 잘 하지 못한다. 따라서 오토인코더에 입력된 값과 오토인코더가 출력된 값의 차이가 적다면 정상 데이터, 차이가 크다면 이상 데이터로 판단할 수 있다.

[0072] 일 실시예로 이상 탐지 장치는 오토인코더의 출력값을 기반으로 이상치 점수(anomaly score)을 계산하여 이상이 있는지 탐지할 수 있다. 이상치 점수는 입력값과 출력값과의 차이 정도를 포함할 수 있다. 이상치 점수는 평균 절대 오차(Mean Absolute Error, MAE), 평균 제곱 오차(Mean Squared Error, MSE) 또는 크로스 엔트로피(cross entropy)를 기반으로 계산된 값일 수 있다.

[0073] 수학식2는 평균 절대 오차(MAE) 기반 이상치 점수 $e(d_k)$를 계산할 때 이용하는 식이다.

수학식 2

$$e(d_k) = \frac{1}{n}\sum_{l=1}^{n} |d_k^{(l)} - \hat{d}_k^{(l)}|$$

[0074]

[0075] 수학식2에서 $d_k^{(l)}$는 k번째 입력 데이터의 l번째 차원의 값을 의미한다. 수학식2에서 $\hat{d}_k^{(l)}$는 k번째 입력 데이터를 오토인코더에 입력했을때 출력값의 l번째 차원의 값을 의미한다. 수학식2에서 n은 입력 데이터의 차원수를 의미한다. 수학식2에서 $e(d_k)$는 k번째 입력 데이터가 입력되었을 때의 이상치 점수를 의미한다.

[0076] 수학식2에서 $e(d_k)$가 임계값 보다 크다면 분석 대상 데이터는 이상이 있는 데이터로 분류할 수 있다. 반대로 수학식2에서 $e(d_k)$가 임계값 보다 작다면 분석 대상 데이터는 정상인 데이터로 분류할 수 있다.

[0078] 연구원은 전술한 상호정보량 기반 오토인코더 경량화 방법을 적용하여 오토인코더를 경량화 하였다. 연구원은 경량화된 오토인코더의 성능을 평가 실험을 진행하였다. 실험에 이용된 오토인코더의 인코더 및 디코더는 각각 3개의 계층을 가진다. 이하 실험결과에 대해 살펴본다.

[0080] 도7은 경량화되지 않는 오토인코더와 50% 경량화된 오토인코더의 이상치 점수(anomaly score) 히스토그램에 대한 것이다. 도7(a)는 경량화되지 않는 오토인코더에 대한 것이다. 도7(b)는 50% 경량화된 오토인코더에 대한 것이다.

[0081] 경량화 되지 않은 오토인코더는 입력데이터의 모든 정보를 활용하여 출력데이터를 구성한다. 따라서 정상 데이터(normal data)와 이상 데이터(abnormal data)의 이상치 점수가 쉽게 구분되는 것을 확인할 수 있다.

[0082] 50% 경량화된 오토인코더의 경우 정보의 손실이 일부 발생하였음에도 불구하고 여전히 정상 데이터의 분포와 이상 데이터의 분포가 쉽게 구분되는 것을 확인할 수 있다.

[0083] 즉 전술한 오토인코더 경량화 방법을 적용하더라도 데이터의 이상을 탐지함에 어려움이 없다는 것을 확인할 수 있다. 이는 계층 간 노드 사이의 중요도가 높은 연결은 보존하면서 가지치기 되었기 때문이다. 즉 이상탐지에 불필요한 정보는 제외되고 정상 데이터와 이상 데이터를 구분하는 핵심적인 정보가 유지되었음을 의미한다.

[0085] 도8은 가지치기 비율(pruning ratio)에 따른 오토인코더의 이상탐지 성능을 평가한 결과이다.

[0086] 전술한 상호정보량 기반 오토인코더 경량화 방법을 통해 가지치기한 모델(Proposed)는 가지치기 비율(rruning ratio)가 증가하더라도 정확도(accuracy) 및 F1-socre가 크게 감소하지 않는 것을 확인할 수 있다. 오히려 F1-socre는 가지치기 하기 전 모델(Dense autoencoder)과 비교할 때 높은 것을 확인할 수 있다.

[0087] 반면 전술한 상호정보량 기반 오토인코더 경량화 방법에서 미세 조정(fine-tuning)을 수행하지 않는 모델(Pruned autoencoder without fine-tuning)은 가지치기 하기 전 모델(Dense autoencoder)와 비교할 때 정확도 및 F1-socre의 값이 떨어지는 것을 확인할 수 있다.

[0088] 이를 통해 전술한 상호정보량 기반 오토인코더 경량화 방법을 통해 오토인코더를 경량화 한 경우 가중치의 표현력을 보존하며 우수한 성능으로 이상탐지가 가능한 것을 확인할 수 있다.

[0090] 도9는 여러 가지치기 방법을 적용해 오토인코더를 경량화한 후 성능을 평가한 결과이다.

[0091] 도9의 (a)는 입력 계층으로부터 출력 계층까지 순서대로 가지치기한 경우의 실험결과이다.

[0092] 도9의 (b)는 가지치기 대상 계층을 선택 시, 디코더에 속한 계층을 먼저 선택 한 후 인코더에 속한 계층을 선택한 경우의 실험 결과이다. 즉 도4의 (b)방법이 적용된 경우의 실험 결과이다.

[0093] 도9의 (c)는 가지치기 대상 계층으로 선택 시, 인코더에 속한 계층과 디코더에 속한 계층을 번갈아 가며 선택하는 경우의 실험 결과이다. 즉 도5의 (a)방법이 적용된 경우의 실험 결과이다.

[0094] 도9의 (d)는 별도의 미세 조정 없이 전체 계층을한번에 가지치기 한 경우의 실험결과이다.

[0095] 도9(a)의 실험결과와 도9(b) 및 도9(c)의 실험결과를 비교하면 도9(b) 및 도9(c)의 경우가 가지치기 비율이 증가되더라도 상대적으로 성능 감소 정도가 약한 것을 확인할 수 있다.

[0096] 도9(d)의 실험결과를 살펴보면 별도의 미세 조정 과정이 없는 경우 경량화 되지 않은 오토인코더의 경우보다 성능이 매우 나빠지는 것을 확인할 수 있다.

[0098] 이하 경량화 장치에 대해 설명한다.

[0099] 도10은 경량화 장치의 실시예 중 하나의 구성이다.

[0100] 경량화 장치(400)는 도1에서 설명한 경량화 장치(100)에 해당할 수 있다. 경량화 장치(400)는 전술한 상호정보량 기반 오토인코더 경량화 방법을 수행하는 장치일 수 있다. 더 나아가 경량화 장치(400)는 이상 탐지 장치의 역할을 수행할 수 있다. 즉 경량화 장치(400)는 오토인코더를 경량화 한 뒤, 경량화된 오토인코더를 이용해 이상 탐지 방법을 수행할 수 있다. 즉 하기 장치에 대한 설명은 이상 탐지 장치에도 그대로 적용될 수 있다.

[0101] 경량화 장치(400)는 물리적으로 다양한 형태로 구현될 수 있다. 예를 들어 경량화 장치(400)는 PC, 노트북, 스마트기기, 서버 또는 데이터처리 전용 칩셋 등의 형태를 가질 수 있다.

[0102] 경량화 장치(400)는 입력장치(410), 저장장치(420), 연산장치(430), 출력장치(440), 인터페이스 장치(450) 및 통신장치(460)를 포함할 수 있다.

[0103] 입력장치(410)는 일정한 명령 또는 데이터를 입력 받는 인터페이스 장치(키보드, 마우스, 터치스크린 등)를 포함할 수도 있다. 입력장치(410)는 별도의 저장장치(USB, CD, 하드디스크 등)를 통하여 정보를 입력 받는 구성을 포함할 수도 있다. 입력장치(410)는 입력 받는 데이터를 별도의 측정장치를 통하여 입력 받거나, 별도의 DB를 통하여 입력 받을 수도 있다. 입력장치(410)는 유선 또는 무선 통신을 통해 데이터를 입력 받을 수 있다.

[0104] 입력장치(410)는 상호정보량 기반 오토인코더 경량화 방법에 필요한 정보 및 모델을 입력받을 수 있다. 입력장치(410)는 오토인코더를 입력받을 수 있다. 입력장치(410)는 분석 대상 데이터 및 학습 데이터를 입력받을 수 있다.

[0105] 저장장치(420)는 입력장치(410)를 통해 입력 받은 정보를 저장할 수 있다. 저장장치(420)는 연산장치(430)가 연산하는 과정에서 생성되는 정보를 저장할 수 있다. 즉 저장장치(420)는 메모리를 포함할 수 있다. 저장장치

(420)는 연산장치(430)가 계산한 결과를 저장할 수 있다.

[0106] 저장장치(420)는 상호정보량 기반 오토인코더 경량화 방법에 필요한 정보 및 모델을 저장할 수 있다. 저장장치(420)는 오토인코더를 저장할 수 있다.

[0107] 연산장치(430)는 데이터를 처리하고, 일정한 연산을 처리하는 프로세서, AP, 프로그램이 임베디드된 칩과 같은 장치일 수 있다. 연산장치(430)는 경량화 장치(400)를 제어하는 제어신호를 생성할 수 있다.

[0108] 연산장치(430)는 상호정보량 기반 오토인코더 경량화 방법에 필요한 연산을 수행할 수 있다.

[0109] 연산장치(430)는 상기 오토인코더의 계층 중 가지치기 대상 계층을 선택할 수 있다. 연산장치(430)는 상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산할 수 있다. 연산장치(430)는 상기 가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기할 수 있다. 연산장치(430)는 상기 가지치기 된 오토인코더를 미세 조정할 수 있다. 연산장치(430)는 상기 미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택되었는지 판단할 수 있다. 미세조정 된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한 번씩 선택되지 못했다고 판단하는 경우, 연산장치(430)는 상기 미세 조정된 오토인코더를 대상으로 상기 링크의 중요도를 계산하는 것, 상기 가지치기 대상 계층을 선택하는 것, 상기 가지치기 하는 것 및 상기 미세 조정 하는 것을 다시 수행할 수 있다. 연산장치(430)는 분석 대상 데이터를 경량화된 오토인코더에 입력한 뒤, 오토인코더의 출력값을 기반으로 이상이 있는지 탐지할 수 있다.

[0110] 출력장치(440)는 일정한 정보를 출력하는 장치가 될 수도 있다. 출력장치(440)는 데이터 과정에 필요한 인터페이스, 입력된 데이터, 분석결과 등을 출력할 수도 있다. 출력장치(440)는 디스플레이, 문서를 출력하는 장치, 등과 같이 물리적으로 다양한 형태로 구현될 수도 있다.

[0111] 인터페이스 장치(450)는 외부로부터 일정한 명령 및 데이터를 입력 받는 장치일 수 있다. 인터페이스 장치(450)는 물리적으로 연결된 입력 장치 또는 외부 저장장치로부터 상호정보량 기반 오토인코더 경량화 방법에 필요한 정보 및 모델을 입력받을 수 있다. 인터페이스 장치(450)는 경량화 장치(400)를 제어하기 위한 제어신호를 입력 받을 수 있다. 인터페이스 장치(450)는 경량화 장치(400)가 분석한 결과를 출력할 수 있다.

[0112] 통신장치(460)는 유선 또는 무선 네트워크를 통해 일정한 정보를 수신하고 전송하는 구성을 의미할 수 있다. 통신장치(460)는 경량화 장치(400)를 제어하는데 필요한 제어 신호를 수신할 수 있다. 통신장치(460)는 경량화 장치(400)가 분석한 결과를 전송할 수 있다.

[0114] 전술한 상호정보량 기반 오토인코더 경량화 방법 또는 경량화된 오토인코더를 이용해 이상을 탐지하는 방법은 컴퓨터 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다.

[0115] 상기 프로그램은 일시적 또는 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.

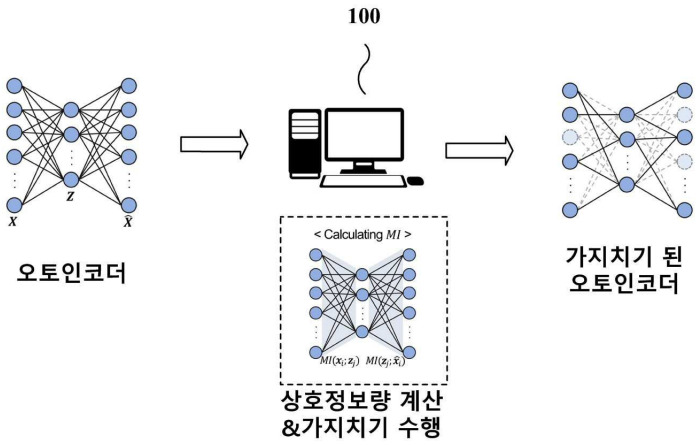
[0116] 비일시적 판독 가능 매체는 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM(Read-Only Memory), PROM(Programmable Read Only Memory), EPROM(Erasable PROM) 또는 EEPROM(Electrically EPROM) 또는 플래시 메모리 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

[0117] 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.

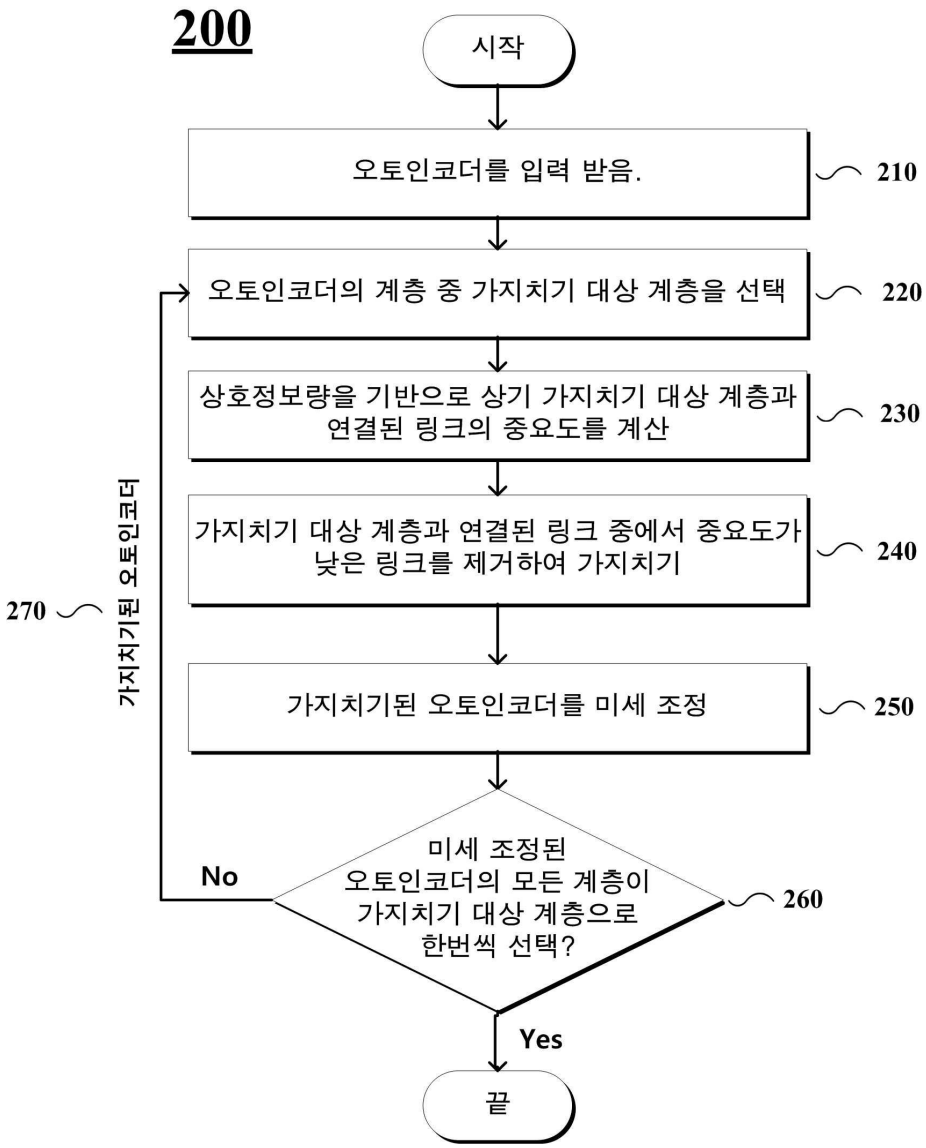
[0118] 본 실시예 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시예는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

도면

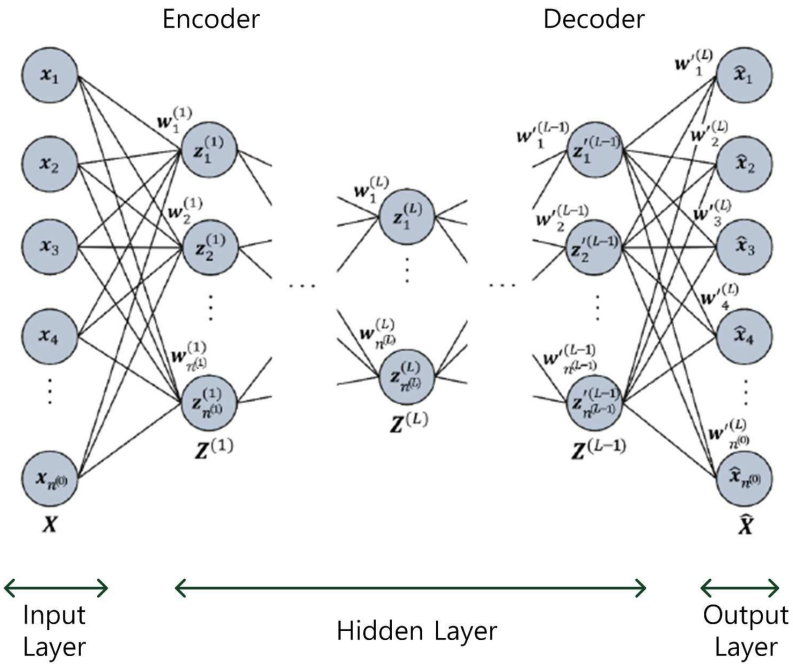
도면1


100
X
Z
$\hat{X}$
오토인코더
< Calculating MI >
$MI(x_i; z_j)$ $MI(z_j; \hat{x}_i)$
상호정보량 계산
&가지치기 수행
가지치기 된
오토인코더

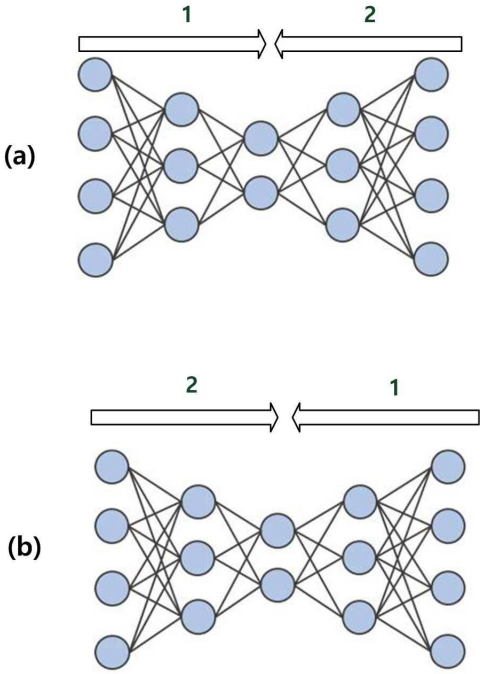
도면2


200
시작
오토인코더를 입력 받음.
210
오토인코더의 계층 중 가지치기 대상 계층을 선택
220
상호정보량을 기반으로 상기 가지치기 대상 계층과 연결된 링크의 중요도를 계산
230
가지치기 대상 계층과 연결된 링크 중에서 중요도가 낮은 링크를 제거하여 가지치기
240
가지치기된 오토인코더를 미세 조정
250
미세 조정된 오토인코더의 모든 계층이 가지치기 대상 계층으로 한번씩 선택?
260
No
Yes
끝
270
가지치기된 오토인코더

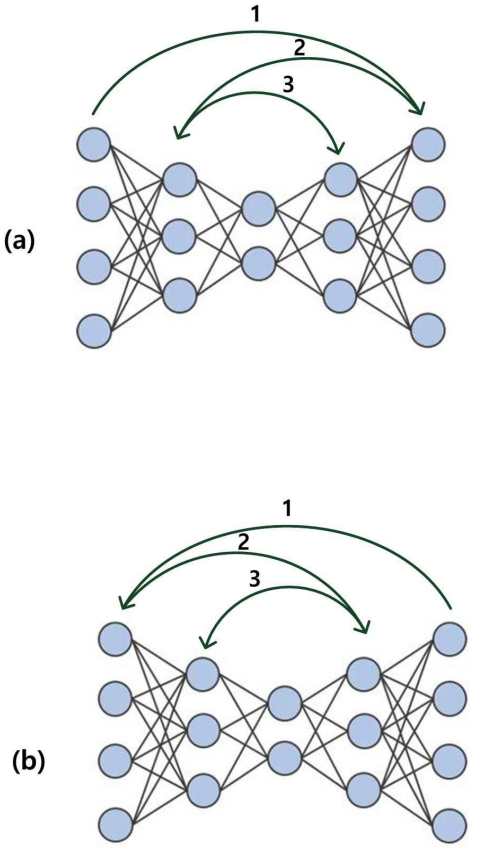
도면3


Encoder
Decoder
Input Layer
Hidden Layer
Output Layer

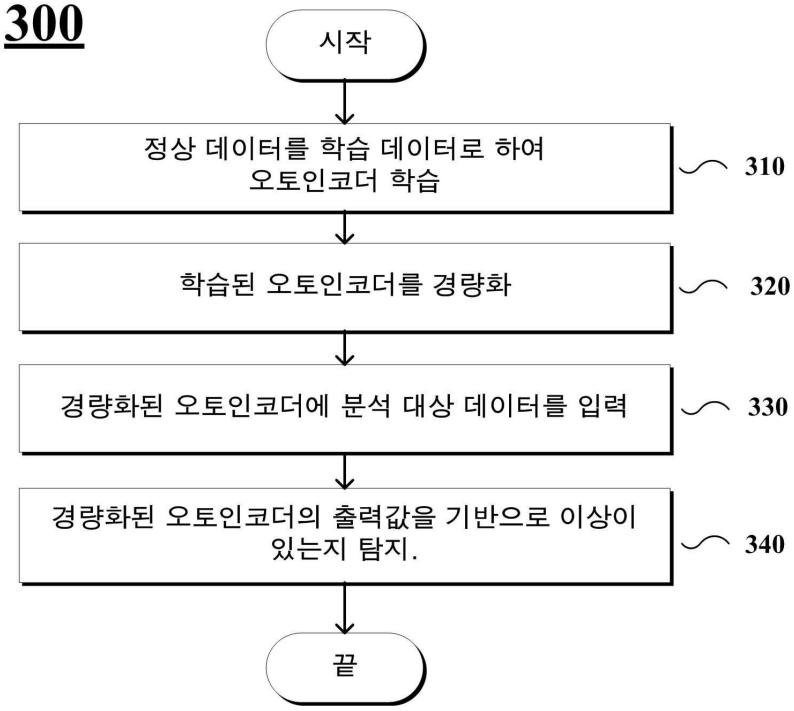
도면4


(a)
1
2
(b)
2
1

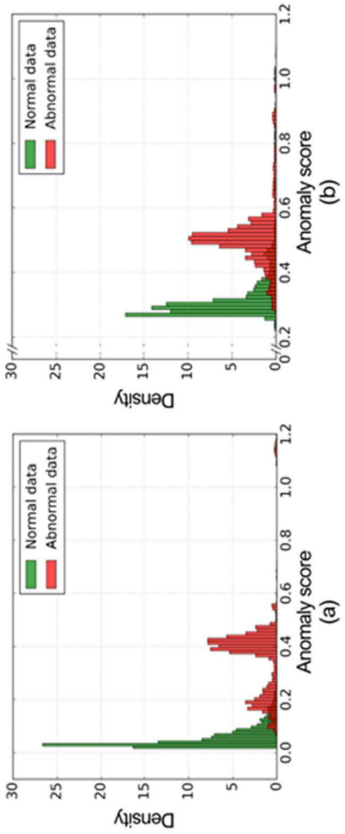
도면5


(a)
1
2
3
(b)
1
2
3

도면6


300
시작
정상 데이터를 학습 데이터로 하여 오토인코더 학습
310
학습된 오토인코더를 경량화
320
경량화된 오토인코더에 분석 대상 데이터를 입력
330
경량화된 오토인코더의 출력값을 기반으로 이상이 있는지 탐지.
340
끝

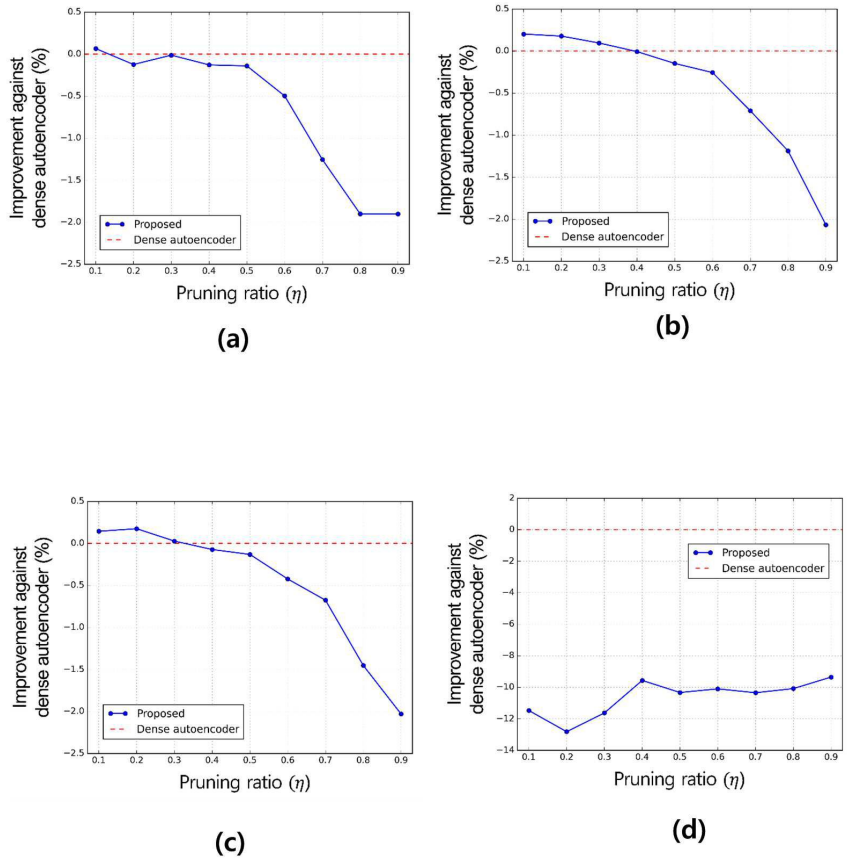
도면7


Normal data
Abnormal data
Density
Anomaly score
(b)
Normal data
Abnormal data
Density
Anomaly score
(a)

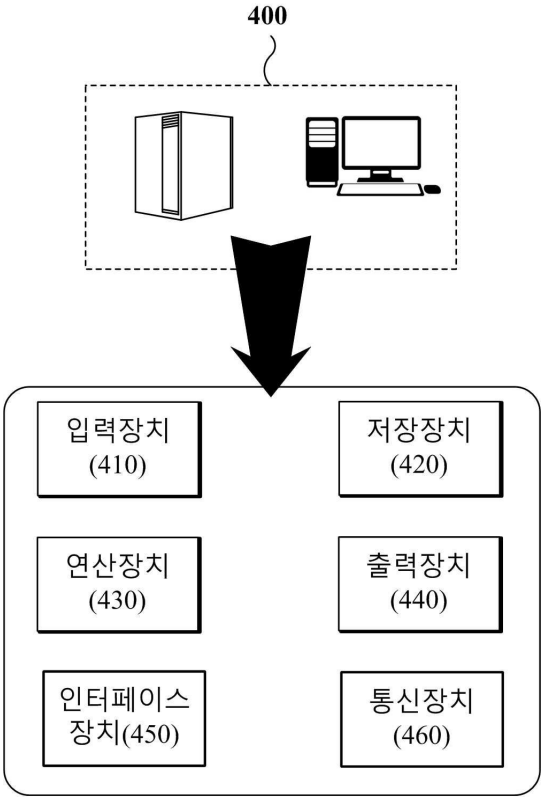
도면8

		Accuracy	Precision	Recall	F1-score
Pruning ratio	0%	0.9608	0.9686	0.9837	0.9761
	30%	0.9410	0.9529	0.9756	0.9641
	50%	0.9379	0.9486	0.9766	0.9624
	70%	0.9266	0.9657	0.9433	0.9543

도면9


Improvement against dense autoencoder (%)
Pruning ratio (η)
Proposed
Dense autoencoder
(a)
(b)
(c)
(d)

도면10


400
입력장치 (410)
저장장치 (420)
연산장치 (430)
출력장치 (440)
인터페이스 장치(450)
통신장치 (460)