



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년09월17일
(11) 등록번호 10-2157441
(24) 등록일자 2020년09월11일

(51) 국제특허분류(Int. Cl.)
G06N 3/08 (2006.01) G06N 3/063 (2006.01)
(52) CPC특허분류
G06N 3/08 (2013.01)
G06N 3/063 (2013.01)
(21) 출원번호 10-2019-0064861
(22) 출원일자 2019년05월31일
심사청구일자 2019년05월31일
(56) 선행기술조사문헌
KR1020170134508 A
US20180018553 A1
Sebastian Bach et al. On Pixel-Wise
Explanations for Non-Linear Classifier
Decisions by Layer-Wise Relevance
Propagation. PLoS one, 10(7): e0130140, 2015

(73) 특허권자
이화여자대학교 산학협력단
서울특별시 서대문구 이화여대길 52 (대현동, 이
화여자대학교)
(72) 발명자
강재원
서울특별시 마포구 독막로 145, 104동 904호(창전
동, 서강쌍용에가아파트)
유채화
경상남도 통영시 무전7길 135, 207동 1003호(무전
동, 한진로즈힐비치2차)
김나영
서울특별시 성북구 오패산로3길 17, 104동101호
(하월곡동, 동신아파트)
(74) 대리인
특허법인(유한)아이시스

전체 청구항 수 : 총 10 항

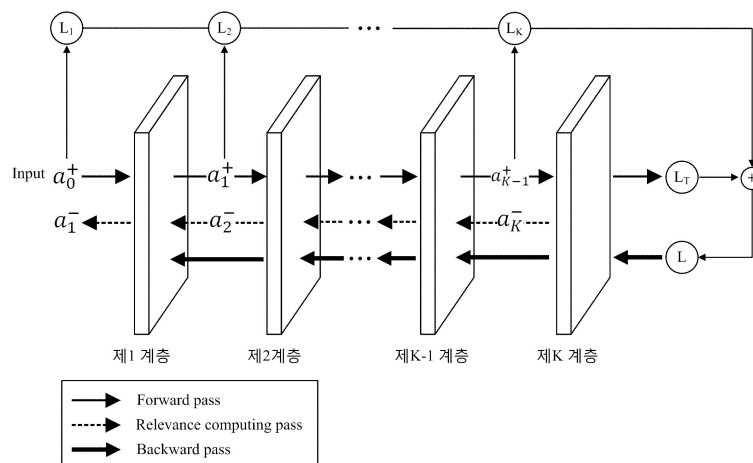
심사관 : 박성수

(54) 발명의 명칭 관련성 전사를 이용한 신경망 학습 방법 및 서비스 장치

(57) 요약

관련성 전사를 이용한 신경망 학습 방법은 학습 장치가 신경망에 대하여 학습 데이터를 이용한 순방향 전사 과정을 수행하여 제1 손실함수를 결정하는 단계, 상기 학습 장치가 상기 신경망에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 계층별 손실함수를 연산하는 단계, 상기 학습 장치가 계층별 손실함수를 합산한 결과로 생성되는 제2 손실함수와 상기 제1 손실함수를 합산하여 최종 손실함수를 결정하는 단계 및 상기 학습 장치가 상기 최종 손실함수를 이용하여 역방향 전사 과정을 수행하면서 상기 각 계층의 가중치를 갱신하는 단계를 포함한다.

대표도



이 발명을 지원한 국가연구개발사업

과제고유번호	1711081006
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	정보통신·방송 연구개발사업-차세대(UHD)방송 서비스 활성화기술개발사업
연구과제명	6DoF지원 초고화질 몰입형 비디오의 압축 및 전송 핵심 기술 개발
기 여 율	1/2
과제수행기관명	이화여자대학교 산학협력단
연구기간	2019.01.01 ~ 2019.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711092198
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	기초연구사업(학술진흥)(2019)-신진연구(연평균연구비 1억원 이내)
연구과제명	고화질 비디오의 장기간 예측과 해석 가능한 딥러닝 모델 기반의 비디오 신호처리 알고리즘 개발
기 여 율	1/2
과제수행기관명	이화여자대학교 산학협력단
연구기간	2019.03.01 ~ 2020.02.29

명세서

청구범위

청구항 1

학습 장치가 신경망에 대하여 학습 데이터를 이용한 순방향 전사 과정을 수행하여 제1 손실함수를 결정하는 단계;

상기 학습 장치가 상기 신경망에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 계층별 손실함수를 연산하는 단계;

상기 학습 장치가 계층별 손실함수를 합산한 결과로 생성되는 제2 손실함수와 상기 제1 손실함수를 합산하여 최종 손실함수를 결정하는 단계; 및

상기 학습 장치가 상기 최종 손실함수를 이용하여 역방향 전사 과정을 수행하면서 상기 각 계층의 가중치를 갱신하는 단계를 포함하는 관련성 전사를 이용한 신경망 학습 방법.

청구항 2

제1항에 있어서,

상기 학습 장치가 학습 데이터를 이용한 순방향 전사 과정과 역방향 전사 과정을 수행하면서 상기 신경망을 초기 학습하는 단계를 더 포함하고,

상기 초기 학습된 신경망의 판단 결과가 기준값보다 높은 정확도를 갖는 경우에 상기 계층별 손실함수를 연산하는 과정을 진행하는 관련성 전사를 이용한 신경망 학습 방법.

청구항 3

제1항에 있어서,

상기 신경망의 i 번째 계층에 대한 상기 계층별 손실함수는 $i-1$ 번째 계층에서 생성한 특징맵과 상기 i 번째 계층에서의 역방향 관련성 사이의 손실 함수인 관련성 전사를 이용한 신경망 학습 방법.

청구항 4

제3항에 있어서,

상기 i 번째 계층에서의 역방향 관련성은 아래의 역방향 활성화 맵으로 결정되는 관련성 전사를 이용한 신경망 학습 방법.

$$a_i^- = f(a_{i-1}^+, \sum_{j=1}^{C_i} w_{i,j} \frac{a_{i+1,j}^-}{a_{i,j}^+})$$

(a_i^- 는 상기 i 번째 계층의 역방향 관련성, C_i 는 상기 i 번째 계층의 채널 크기, w 는 가중치, a_i^+ 는 상기 i 번째 계층의 순방향 액티베이션 맵)

청구항 5

제1항에 있어서,

상기 제2 손실함수는 상기 각 계층 전체의 계층별 손실함수의 평균값인 관련성 전사를 이용한 신경망 학습 방법.

청구항 6

컴퓨터에서 제1항 내지 제5항 중 어느 하나의 항에 기재된 관련성 전사를 이용한 신경망 학습 방법을 실행하기

위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록 매체.

청구항 7

입력 데이터를 입력받는 입력 장치;

관련성 전사를 이용하여 학습된 신경망 모델을 저장하는 저장 장치; 및

상기 입력 데이터를 상기 신경망 모델에 입력하고, 상기 신경망 모델이 출력하는 결과를 이용하여 특정한 서비스 정보를 생성하는 연산 장치를 포함하되,

상기 신경망 모델은 역방향 관련성 전사 과정에서 생성되는 계층별 손실함수를 이용하여 생성되는 최종 손실함수를 기준으로 학습되는 관련성 전사를 이용하여 학습된 신경망을 이용한 서비스 장치.

청구항 8

제7항에 있어서,

상기 최종 손실함수는

학습 데이터를 기준으로 상기 신경망 모델에 대하여 순방향 전사 과정을 수행하여 결정되는 제1 손실함수 및 상기 신경망 모델에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 결정되는 상기 계층별 손실함수를 이용하여 생성되는 제2 손실함수를 합산한 손실함수인 관련성 전사를 이용하여 학습된 신경망을 이용한 서비스 장치.

청구항 9

제7항에 있어서,

상기 신경망의 i 번째 계층에 대한 상기 계층별 손실함수는 i-1 번째 계층에서 생성한 특징맵과 상기 i 번째 계층에서의 역방향 관련성 사이의 손실 함수인 관련성 전사를 이용하여 학습된 신경망을 이용한 서비스 장치.

청구항 10

제9항에 있어서,

상기 i 번째 계층에서의 역방향 관련성은 아래의 역방향 활성화 맵으로 결정되는 관련성 전사를 이용하여 학습된 신경망을 이용한 서비스 장치.

$$a_i^- = f(a_{i-1}^+, \sum_{j=1}^{C_i} w_{i,j} \frac{a_{i+1,j}^-}{a_{i,j}^+})$$

(a_i^- 는 상기 i 번째 계층의 역방향 관련성, C_i 는 상기 i 번째 계층의 채널 크기, w 는 가중치, a_i^+ 는 상기 i 번째 계층의 순방향 액티베이션 맵)

발명의 설명

기술 분야

[0001] 이하 설명하는 기술은 인공 신경망 모델을 학습하는 기법에 관한 것이다.

배경 기술

[0002] 인공신경망은 높은 정확도를 바탕으로 다양한 산업에 적용되고 있다. 그러나 인공신경망 모델의 동작이 복잡한 블랙 박스 (black-box) 모델로서 네트워크의 특정 예측에 대한 정보를 제공하지 않는다. 인공신경망이 내부에서 어떤 기준으로 의사 결정을 하는지 알기 어렵다. 이러한 특징 때문에 네트워크의 판단에 대한 근거가 반드시 필요한 의료, 자율 주행, 군사 등의 산업에 쉽게 사용할 수 없는 한계가 있다. 이러한 한계를 극복하고자 신경망을 해석하려는 연구가 진행되고 있다.

선행기술문헌

비특허문헌

- [0003] (비특허문헌 0001) Sebastian Bach, Alexander Binder, Gregoire Montavon, Frederick Klauschen, Klaus-Robert Muller, Wojciech Samek, "On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation", PLoS one, 10(7): e0130140, 2015

발명의 내용

해결하려는 과제

- [0004] 이하 설명하는 기술은 인공지능망을 해석하는 기법을 응용한 새로운 신경망 학습 방법을 제공하고자 한다. 이하 설명하는 기술은 계층적 관련성 전사의 결과를 이용한 신경망 학습 방법을 제공하고자 한다.

과제의 해결 수단

- [0005] 관련성 전사를 이용한 신경망 학습 방법은 학습 장치가 신경망에 대하여 학습 데이터를 이용한 순방향 전사 과정을 수행하여 제1 손실함수를 결정하는 단계, 상기 학습 장치가 상기 신경망에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 계층별 손실함수를 연산하는 단계, 상기 학습 장치가 계층별 손실함수를 합산한 결과로 생성되는 제2 손실함수와 상기 제1 손실함수를 합산하여 최종 손실함수를 결정하는 단계 및 상기 학습 장치가 상기 최종 손실함수를 이용하여 역방향 전사 과정을 수행하면서 상기 각 계층의 가중치를 갱신하는 단계를 포함한다.
- [0006] 관련성 전사를 이용하여 학습된 신경망을 이용한 서비스 장치는 입력 데이터를 입력받는 입력 장치, 관련성 전사를 이용하여 학습된 신경망 모델을 저장하는 저장 장치 및 상기 입력 데이터를 상기 신경망 모델에 입력하고, 상기 신경망 모델이 출력하는 결과를 이용하여 특정한 서비스 정보를 생성하는 연산 장치를 포함한다. 상기 신경망 모델은 역방향 관련성 전사 과정에서 생성되는 계층별 손실함수를 이용하여 생성되는 최종 손실함수를 기준으로 학습된다.

발명의 효과

- [0007] 이하 설명하는 기술은 신경망 내부 계층에서의 판단 정보가 신경망의 추론에 어떤 결과를 가져오는지를 고려하여 신경망을 학습한다. 따라서 이하 설명하는 기술은 주어진 문제에 대한 성능을 높이면서도 특징 맵의 발현 위치가 특정되도록 가중치가 갱신된다. 나아가 이하 설명하는 기술은 일반적인 딥러닝 모델에 적용 가능한 범용적 기술이다.

도면의 간단한 설명

- [0008] 도 1은 DNN 전사 과정에 대한 예이다.
 도 2는 DNN 역전사 과정에 대한 예이다.
 도 3은 계층적 관련성 전사 과정에 대한 예이다.
 도 4는 DNN 학습 과정에 대한 예이다.
 도 5는 DNN 학습 과정에 대한 순서도의 예이다.
 도 6은 DNN 학습 과정에 대한 순서도의 다른 예이다.
 도 7은 DNN을 포함하는 서비스 장치에 대한 예이다.

발명을 실시하기 위한 구체적인 내용

- [0009] 이하 설명하는 기술은 다양한 변경을 가할 수 있고 여러 가지 실시례를 가질 수 있는 바, 특정 실시례들을 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 포함하

는 것으로 이해되어야 한다.

- [0010] 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.
- [0011] 본 명세서에서 사용되는 용어에서 단수의 표현은 문맥상 명백하게 다르게 해석되지 않는 한 복수의 표현을 포함하는 것으로 이해되어야 하고, "포함한다" 등의 용어는 설시된 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.
- [0012] 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.
- [0013] 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.
- [0015] 이하 설명하는 기술은 신경망 학습 방법에 관한 것이다.
- [0016] 신경망(neural network) 내지 인공신경망(artificial neural network)은 생물의 신경망을 모방한 통계학적 학습 알고리즘이다. 다양한 신경망 모델이 연구되고 있다. 최근 딥러닝 신경망(deep learning network, DNN)이 주목받고 있다.
- [0017] DNN은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공 신경망 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.
- [0018] 신경망 내지 DNN은 학습 방법도 연구되고 있다. 이하 설명하는 기술은 역전사(Backward propagation) 방식에 기반한 학습 기법에 적용할 수 있다. 역전사 방식에 따른 가중치 조절은 대부분의 신경망 모델에서 이용하는 방식이다. 따라서, 이하 설명하는 기술은 특정한 신경망 모델에 국한되는 것이 아니고, 역전사 기반한 학습 방법을 갖는 다양한 신경망 모델에 적용 가능하다.
- [0019] 신경망 학습 과정은 순방향 전사(Forward propagation) 및 역방향 전사(Backward propagation)를 포함한다. 순방향은 입력층에서 출력층 방향을 의미한다. 역방향 전사는 간략하게 역전사라고도 명명한다.
- [0021] DNN을 기준으로 신경망 학습 과정에 대하여 먼저 설명한다.
- [0022] 도 1은 DNN 전사 과정에 대한 예이다. 신경망은 복수의 계층으로 구성된다. 통상적으로 신경망은 입력층, 복수의 은닉층 및 출력층으로 구성된다. 초기 학습 단계에서 신경망의 각 계층들은 가중치가 랜덤한 값으로 초기화된다.
- [0023] 신경망에서 각 계층은 주어진 입력 데이터와 가중치의 연산을 수행하여 특징 맵을 결정한다. 하나의 계층에서 생성한 특징 맵은 다음 계층의 입력으로 사용된다. 예컨대, 도 1에서 첫 번째 계층(입력층)에서 출력되는 특징 맵은 $a_1^+ = f(a_0^+, w_1)$ 이다. 여기서, f 는 그 계층이 수행하는 함수, a_0^+ 및 w_1 는 각각 그 계층의 입력과 가중치를 의미한다. 신경망의 입력층 및 은닉층이 이와 같은 과정을 반복하고, 마지막으로 출력층은 최종적으로 나온 결과를 바탕으로 비용 함수(cost function)를 계산한다. 출력층은 결정 계층이라고 명명하기도 한다.
- [0024] 비용함수는 학습의 형태 (지도학습, 비지도학습, 강화학습 등), 신경망 모델의 종류, 활성화 함수(activation

function) 같은 요인들을 기준으로 선택될 수 있다. 예컨대, 다중 분류 문제에 대해 지도 학습을 수행할 때 일반적으로 활성화 함수와 비용 함수는 각각 소프트맥스(softmax) 함수와 교차 엔트로피 함수(cross entropy function)로 결정된다. 각 함수의 정의는 도 1에 표시하였다.

[0025] 출력층은 $d_1, d_2, \dots, d_k, \dots, d_N$ 을 결과를 출력한다. $d_1, d_2, \dots, d_k, \dots, d_N$ 에 대하여 활성화 함수를 적용한 값이 $p_1, p_2, \dots, p_k, \dots, p_N$ 이다. 출력층의 결과 중 활성화 함수를 적용한 후 가장 큰 값 p_k 라고 가정한다. p_k 는 신경망이 내리는 판단(prediction)이다. 비용 함수는 신경망이 내린 판단과 실제 기대하는 결과(label) 간의 손실(loss) L 을 계산한다.

[0027] 도 2는 DNN 역전사 과정에 대한 예이다.

[0028] 신경망(DNN)은 많은 비선형 활성화 함수들이 여러 계층을 통해 복잡하게 얹힌 구조이다. 따라서 계층의 가중치를 최적화하는 일은 매우 어려운 논-컨벡스(non-convex)한 문제이다.

[0029] 일반적인 신경망의 가중치는 확률적 경사 하강법(stochastic gradient descent)를 통하여 조정(갱신)된다. 가중치는 도 2에 표시한 확률적 경사 하강법을 통하여 적당한 값까지 수렴시키는 방법으로 갱신된다. 각 계층의 가중치는 전사를 통해 마지막 계층에서 결정된 비용 함수에 각 가중치가 얼마나 영향을 미쳤는지에 따라 갱신된다.

[0030] 신경망은 복잡한 구조 때문에 중간 계층의 가중치의 영향을 바로 계산할 수 없으므로 연쇄 법칙(chain rule)을 이용하여 계산된다. 도 2는 역전사 과정에 대하여 설명한다. 도 2에서 i 번째 계층의 가중치는 w_i , 가중치가 갱신되는 학습 순간을 $t+1$ 이다.

[0031] 출력층(예컨대, n 계층)에서 손실이 계산되면 이 값에 대한 가중치 변화량을 계산한다. 출력층 바로 앞에 있는 계층($n-1$ 계층)의 손실에 대한 가중치 변화량은 출력층(n 계층)에서 나온 손실에 대한 가중치 변화량과 출력층(n 계층)의 가중치에 대한 해당 계층($n-1$ 계층)의 가중치 변화량을 연쇄적으로 곱하여 구할 수 있다. 각 계층의 가중치는 이 변화량에 따라 갱신되고 이 변화량은 계속해서 앞 계층으로 전달된다. 즉, 가중치 갱신은 출력층에서 입력층 방향으로 진행된다. 이와 같은 가중치 갱신 과정을 역전사라고 부른다. 도 2는 역전사 과정 및 계산에 필요한 수식을 도시한다.

[0033] 도 3은 계층적 관련성 전사(Layerwise Relevance Propagation, LRP) 과정에 대한 예이다.

[0034] 히트맵(heatmap)은 배열을 값들을 2차원에 컬러로 표현하는 그래픽 기법이다. 값의 크기에 따라 다른 색을 지정함으로써 배열의 값 분포를 알 수 있다. 인공지능 분야에서는 히트맵은 입력 영상의 픽셀 각각이 신경망의 판단에 기여하는 정도를 표현할 수 있다. 신경망에서 히트맵을 생성하는 대표적인 방법으로 계층적 관련성 전사(Layerwise Relevance Propagation, LRP) 알고리즘이 있다. LRP는 네트워크 내에서 역방향으로 전달되는 신호를 관련성(Relevance)이라고 정의하고, 이 관련성이 계층을 통과해도 같은 값으로 보존된다는 가정으로 진행된다.

[0035] 신경망의 출력층의 관련성은 전사 과정 후 나오는 결과 자체로부터 계산한다. 역전사 과정과 같이 마지막 계층에서부터 관련성이 전달되는데 관련성은 관련성의 총합이 유지되면서 도 3의 수식처럼 통과하는 계층의 학습된 가중치와 입력되는 특징 맵의 가중 합(weighted sum)을 바탕으로 전달된다. 도 3의 수식에서 c_i 는 해당 i 번째 계층의 커널 개수를 의미하고, 함수 f 는 해당 계층의 연산 함수와 동일한 함수이다. w_i 는 i 번째 계층의 가중치이다. 이와 같은 과정을 통하여 각 계층의 입력 특징맵의 픽셀 각각이 네트워크의 판단에 미친 기여도를 알 수 있다. 입력층까지 통과하면 입력 영상과 동일한 크기로 입력 영상의 픽셀이 네트워크 판단에 미친 기여도를 표현하는 히트맵 i 번째 계층의 a_i^- 을 얻을 수 있다.

[0037] 이하 기술한 히트맵을 생성하는 알고리즘을 응용한 신경망 학습 과정에 대하여 설명한다. 도 4는 DNN 학습 과정에 대한 예이다. 도 4에 도시한 학습 과정은 비용 함수 결정에 종래의 손실 정보와 함께 중간 계층의 특징 맵에 대한 손실 정보를 반영한다. 학습 과정은 크게 3가지 단계로 구성된다. 학습 과정은 순방향 전사, 관련성 전사 및 역방향 전사로 구성된다. 즉, 종래 기법과 달리 관련성 전사가 추가되었다. 관련성 전사의 결과는 최종적인 손실 정보 결정에 사용된다.

[0038] 도 4에서 순방향 전사를 통해 계층의 결과로 나오는 특징 맵을 a^+ 로 표기하였고, LRP 계산은 네트워크의 역방향

으로 진행되므로 계산된 관련성은 a^- 로 표기하였다.

[0039] (i) 먼저 순방향 전사를 통해 신경망이 다루는 문제에 대한 판단과 기대하는 결과 사이의 손실 L_T 을 계산한다.

[0040] (ii) 그 다음 계층별 특징 맵 각각이 네트워크의 판단에 영향을 미치는 정도를 나타내는 관련성을 계산하기 위해 LRP를 진행한다.

[0041] 신경망의 i 번째 계층을 기준으로 설명하면, i 번째 계층에서 계산되는 손실은 해당 계층의 입력 특징 맵인 $(i-1)$ 번째 계층에서 만든 특징 맵 a_{i-1}^+ 과 순방향 전사가 끝난 후 신경망이 내린 판단 정보가 전달된 계층의 관련성 a_i^- 간의 L_1 손실 $L_i = L_1(a_{i-1}^+, a_i^-)$ 이다.

[0043] 이때 가장 첫 계층의 입력 특징 맵 a_0^+ 은 실제 신경망의 입력이라고 정의한다. K 개의 계층이 있는 신경망의 경우 계층마다 이 손실을 계산하여 나온 K 개의 손실과 L_T 를 모두 더한 L 이 신경망의 최종적인 손실이 된다.

[0044] (iii)LRP 과정이 모두 끝나면 최종 손실 L 을 바탕으로 역전사 과정을 수행하여 신경망의 가중치들을 갱신하며 학습을 진행한다.

[0046] 도 4에서 설명한 과정을 수식을 이용하여 정의하면 아래와 같다. CNN을 기준으로 설명한다.

[0047] 순방향 과정에서 i 번째 계층에 대한 순방향 액티베이션 맵(activation map) 세트는 $a_i^+ = [a_{i,1}^+, a_{i,2}^+, \dots, a_{i,C_i}^+]$ 라고 정의한다. a_i^+ 는 아래 수학적 식 1과 같이 정의된다. C_i 는 i 번째 계층의 채널 크기이다.

수학적 식 1

[0049]
$$a_i^+ = f(a_{i-1}^+, w_i)$$

[0050] C_i 는 i 번째 계층의 채널 크기이고, 각 채널의 컨볼루션 커널(kernel)은 $w_i = [w_{i,1}, w_{i,2}, \dots, w_{i,C_i}]$ 이다. f 는 해당 계층의 함수이다.

[0052] 역방향 과정에서 i 번째 계층에 대한 역방향 관련성 세트는 $a_i^- = [a_{i,1}^-, a_{i,2}^-, \dots, a_{i,C_{i-1}}^-]$ 라고 정의한다. a_i^- 는 아래 수학적 식 2와 같이 정의된다.

수학적 식 2

[0054]
$$a_i^- = f(a_{i-1}^+, \sum_{j=1}^{C_i} w_{i,j} \frac{a_{i+1,j}^-}{a_{i,j}^+})$$

[0055] 이때 커널은 a_i^+ 와 a_{i+1}^- 의 비율로 재정규화된다. 도 4에 도시된 바와 같이 a_i^+ 및 a_{i+1}^- 는 각각 i 번째 계층의 내측(inward) 액티베이션 맵과 관련성에 해당한다.

[0057] 각 계층에서 순방향 정보와 역방향 정보를 이용한 새로운 손실 함수를 합산한 평균값은 아래의 수학적 식 3과 같이 정의될 수 있다. 아래 L_{FBI} 는 순방향 특징맵과 역방향 계층의 관련성을 이용한 손실 정보이다.

수학식 3

$$L_{FBI} = \frac{1}{K} \sum_{i=1}^K L_1(\mathbf{a}_{i-1}^+, \mathbf{a}_i^-)$$

[0059]

[0060]

K 는 신경망의 계층 개수이다. $L_1(\cdot)$ 은 두 개의 입력 값(term)의 절대 차이를 연산하는 l_1 손실 함수이다.

[0062]

전체 손실함수는 아래의 수학식 4와 같이 정의할 수 있다.

수학식 4

$$L = (1 - \lambda)L_{TSK} + \lambda L_{FBI}$$

[0064]

[0065]

L_{TSK} 는 신경망의 순방향 과정에서 산출되는 손실함수이다. 즉, 실제 라벨값과 크로스-엔트로피 손실 사이의 손실이다. 도 4에서 L_T 로 표시한 값이다.

[0066]

이러한 과정을 통하여 손실함수가 새롭게 정의된다. 따라서 전술한 과정을 관련성 전사를 이용한 정규화 기법이라고 할 수 있다.

[0068]

이제, 최종적인 손실 함수 L 을 이용하여 신경망을 학습할 수 있다.

[0070]

도 5는 DNN 학습 과정(100)에 대한 순서도의 예이다. 신경망 학습 과정은 입력 데이터 처리가 가능한 연산 장치 내지 컴퓨터 장치가 수행한다. 컴퓨터 장치가 신경망 학습을 수행하므로, 컴퓨터 장치는 학습 장치라고 명명할 수도 있다. 이하 설명의 편의를 위해 컴퓨터 장치가 신경망 학습을 수행한다고 가정한다.

[0071]

컴퓨터 장치는 학습 데이터를 입력받는다(110). 학습 데이터는 영상 데이터 등일 수 있다. 최초 각 계층의 가중치는 랜덤한 값으로 설정될 수 있다.

[0072]

컴퓨터 장치는 학습 데이터를 이용하여 순방향 전사 과정을 수행한다(120). 순방향 과정을 통해 각 계층은 특징 맵을 생성한다. 순방향 과정을 통해 출력층은 특정 문제에 대한 판단을 출력한다. 그리고 출력층은 실제 라벨 값과 출력한 판단을 비교하여 손실 L_T 를 연산한다.

[0073]

컴퓨터 장치는 전술한 계층적 관련성 전사(LRP)를 수행한다(130). 이 과정에서 각 계층에서 순방향의 특징 맵과 역방향의 관련성 정보에 대한 L_1 손실 함수를 연산한다. 그리고 컴퓨터 장치는 각 계층의 L_1 손실 함수를 모두 합산한 값과 L_T 를 합산하여 최종 손실함수 L 을 결정한다(140). 이때 컴퓨터 장치는 수학식 3에서 설명한 바와 같이 각 계층의 L_1 손실 함수를 모두 합산하여 평균한 값과 L_T 를 합산하여 최종 손실함수 L 을 결정할 수 있다.

[0074]

컴퓨터 장치는 최종 손실함수 L 을 기준으로 역방향 전사를 수행하여 각 계층의 가중치를 갱신할 수 있다(150).

[0076]

도 6은 DNN 학습 과정(200)에 대한 순서도의 다른 예이다.

[0077]

신경망의 학습 초기에 관련성 계산 과정을 수행하면 낮은 정확도의 판단 정보로부터 관련성 정보가 연산된다. 이 경우 부정확한 정보가 가중치 학습에 반영되어 신경망의 가중치 학습이 최적화되지 못 할 위험이 있다. 따라서, 종래 신경망 학습 방법에 따라 학습을 수행한 후 신경망 판단의 정확도가 어느 정도까지 도달한 후 전술한 방법에 따라 신경망의 가중치를 갱신할 수도 있다. 도 6은 이와 같은 과정에 대한 예이다.

[0078]

컴퓨터 장치는 학습 데이터를 입력받는다(210). 학습 데이터는 영상 데이터 등일 수 있다. 최초 각 계층의 가중치는 랜덤한 값으로 설정될 수 있다.

[0079]

컴퓨터 장치는 학습 데이터를 이용하여 순방향 전사 과정을 수행한다(220). 순방향 과정을 통해 각 계층은 특징 맵을 생성한다. 순방향 과정을 통해 출력층은 특정 문제에 대한 판단을 출력한다. 그리고 출력층은 실제 라벨

값과 출력한 판단을 비교하여 손실 L_{T1} 를 연산한다. 컴퓨터 장치는 최종 손실함수 L_{T1} 을 기준으로 역방향 전사를 수행하여 각 계층의 가중치를 갱신할 수 있다(230).

- [0080] 컴퓨터 장치는 실제 라벨값과 현재 신경망 출력층의 결과 사이의 차이가 일정한 기준값 이상인지 판단한다(240). 컴퓨터 장치는 학습 정확도가 기준값 미만인 경우(240의 NO), 순방향 전사와 그 결과를 이용한 역방향 전사 과정을 반복적으로 수행할 수 있다.
- [0081] 컴퓨터 장치는 학습 정확도가 기준값 이상인 경우(240의 YES), 계층적 관련성 전사를 수행할 수 있다(250). 이 시점에서 출력층이 결정한 손실 함수가 L_{TK} 라고 가정한다. 이 과정에서 컴퓨터 장치는 각 계층에서 순방향의 특징 맵과 역방향의 관련성 정보에 대한 L_1 손실 함수를 연산한다.
- [0082] 컴퓨터 장치는 각 계층의 L_1 손실 함수를 모두 합산한 값과 L_{TK} 를 합산하여 최종 손실함수 L 을 결정한다(250). 이때 컴퓨터 장치는 수학적 식 3에서 설명한 바와 같이 각 계층의 L_1 손실 함수를 모두 합산하여 평균한 값과 L_{TK} 를 합산하여 최종 손실함수 L 을 결정할 수 있다.
- [0083] 컴퓨터 장치는 최종 손실함수 L 을 기준으로 역방향 전사를 수행하여 각 계층의 가중치를 갱신할 수 있다(260).
- [0085] 도 7은 DNN을 포함하는 서비스 장치(300)에 대한 예이다. 서비스 장치(300)는 전술한 신경망 학습 방법으로 마련한 신경망 모델을 이용하여 일정한 응용 서비스를 제공하는 장치를 의미한다. 서비스 장치(300)가 신경망을 학습한 컴퓨터 장치일 수도 있다. 또는 서비스 장치(300)는 신경망을 학습한 컴퓨터 장치와 별개의 장치일 수도 있다.
- [0086] 서비스 장치(300)는 저장 장치(310), 메모리(320), 연산장치(330), 인터페이스 장치(340) 및 통신 장치(350)를 포함한다.
- [0087] 저장 장치(310)는 서비스 장치(300)의 동작을 위한 프로그램 내지 코드를 저장한다. 저장 장치(310)는 전술한 학습 방법으로 마련된 신경망 모델을 저장한다. 나아가 저장 장치(310)는 전술한 신경망 학습 과정을 위한 프로그램 내지 코드를 저장할 수도 있다.
- [0088] 메모리(320)는 서비스 장치(300)의 동작 과정에서 생성되는 데이터 및 정보 등을 임시 저장할 수 있다.
- [0089] 인터페이스 장치(340)는 외부로부터 일정한 명령 및 데이터를 입력받는 장치이다. 인터페이스 장치(340)는 물리적으로 연결된 입력 장치 또는 물리적인 인터페이스(키패드, 터치 패널 등)로부터 일정한 정보를 입력받을 수 있다. 인터페이스 장치(340)는 신경망 모델, 신경망 모델 학습을 위한 정보, 학습 데이터 등을 입력받을 수 있다. 인터페이스 장치(340)는 신경망 모델에 입력하기 위한 입력 데이터를 입력받을 수 있다.
- [0090] 통신 장치(350)는 무선 네트워크를 통해 일정한 정보를 송수신한다. 통신 장치(350)는 신경망 모델, 신경망 모델 학습을 위한 정보, 학습 데이터 등을 입력받을 수 있다. 통신 장치(350)는 신경망 모델에 입력하기 위한 입력 데이터를 입력받을 수 있다. 통신 장치(350)는 신경망 모델을 이용한 서비스 결과(판단 결과)를 외부 객체에 전송할 수 있다.
- [0091] 인터페이스 장치(340) 및 통신 장치(350)는 사용자 또는 외부 객체로부터 일정한 정보 및 데이터를 입력받을 수 있다. 따라서 인터페이스 장치(340) 및 통신 장치(350)를 포괄하여 입력 장치라고 명명할 수 있다.
- [0092] 연산 장치(330)는 저장장치(310)에 저장된 프로그램 내지 코드를 이용하여 서비스 장치의 동작을 제어한다. 연산 장치(330)는 학습된 신경망 모델을 이용하여 일정한 판단 과정을 수행할 수 있다. 연산 장치(330)는 입력 데이터를 학습된 신경망 모델에 입력하여, 일정한 판단 결과를 생성할 수 있다. 예컨대, 연산 장치(330)는 영상의 객체 검출, 특정 입력에 대한 판단 결과, 특정 입력에 대한 제어 명령 등을 생성할 수 있다.
- [0093] 나아가 연산 장치(330)는 산출한 결과를 피드백하여 신경망 모델을 갱신할 수 있다. 이 과정에서 연산 장치(330)는 전술한 학습 과정을 통하여 신경망 모델의 가중치를 갱신할 수도 있다. 연산 장치(330)는 데이터를 처리하고, 일정한 연산을 처리하는 프로세서, AP, 프로그램이 임베디드된 칩과 같은 장치일 수 있다.
- [0095] 또한, 상술한 바와 같은 DNN 내지 신경망 학습 방법은 컴퓨터에서 실행될 수 있는 실행가능한 알고리즘을 포함하는 프로그램(또는 어플리케이션)으로 구현될 수 있다. 상기 프로그램은 비일시적 판독 가능 매체(non-transitory computer readable medium)에 저장되어 제공될 수 있다.
- [0096] 비일시적 판독 가능 매체란 레지스터, 캐쉬, 메모리 등과 같이 짧은 순간 동안 데이터를 저장하는 매체가 아니

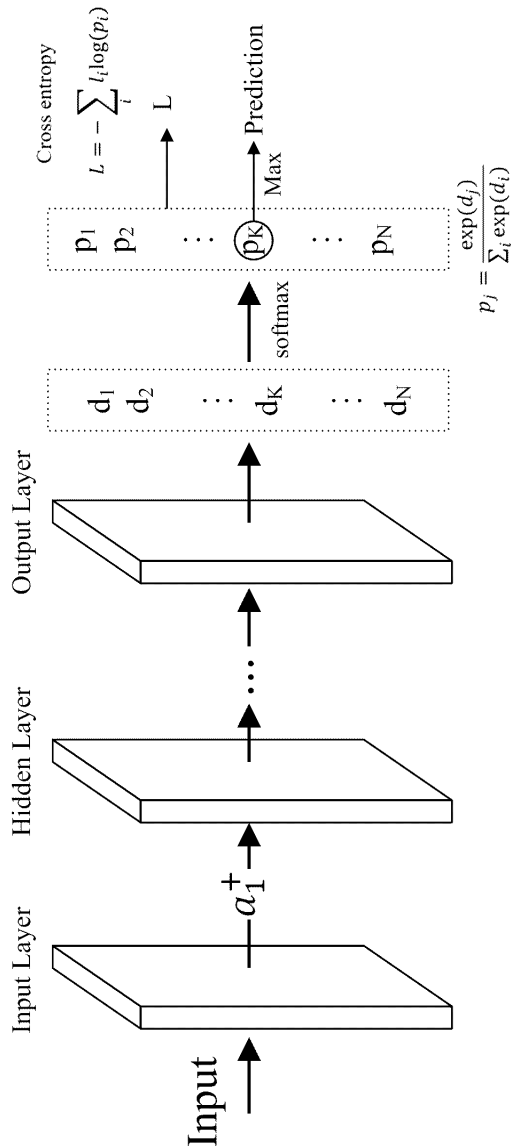
라 반영구적으로 데이터를 저장하며, 기기에 의해 판독(reading)이 가능한 매체를 의미한다. 구체적으로는, 상술한 다양한 어플리케이션 또는 프로그램들은 CD, DVD, 하드 디스크, 블루레이 디스크, USB, 메모리카드, ROM 등과 같은 비일시적 판독 가능 매체에 저장되어 제공될 수 있다.

[0098]

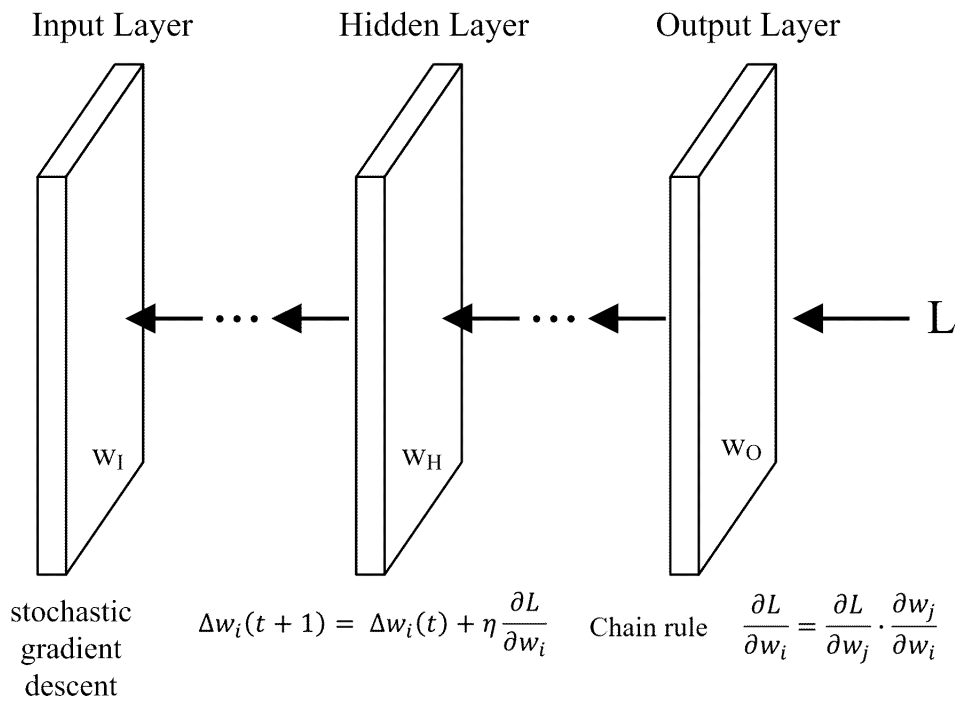
본 실시례 및 본 명세서에 첨부된 도면은 전술한 기술에 포함되는 기술적 사상의 일부를 명확하게 나타내고 있는 것에 불과하며, 전술한 기술의 명세서 및 도면에 포함된 기술적 사상의 범위 내에서 당업자가 용이하게 유추할 수 있는 변형 예와 구체적인 실시례는 모두 전술한 기술의 권리범위에 포함되는 것이 자명하다고 할 것이다.

도면

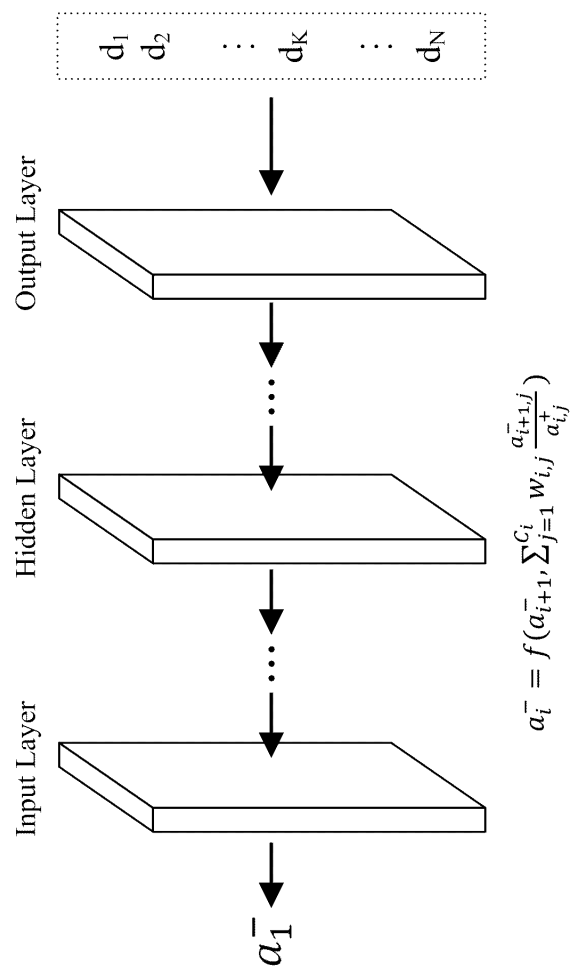
도면1



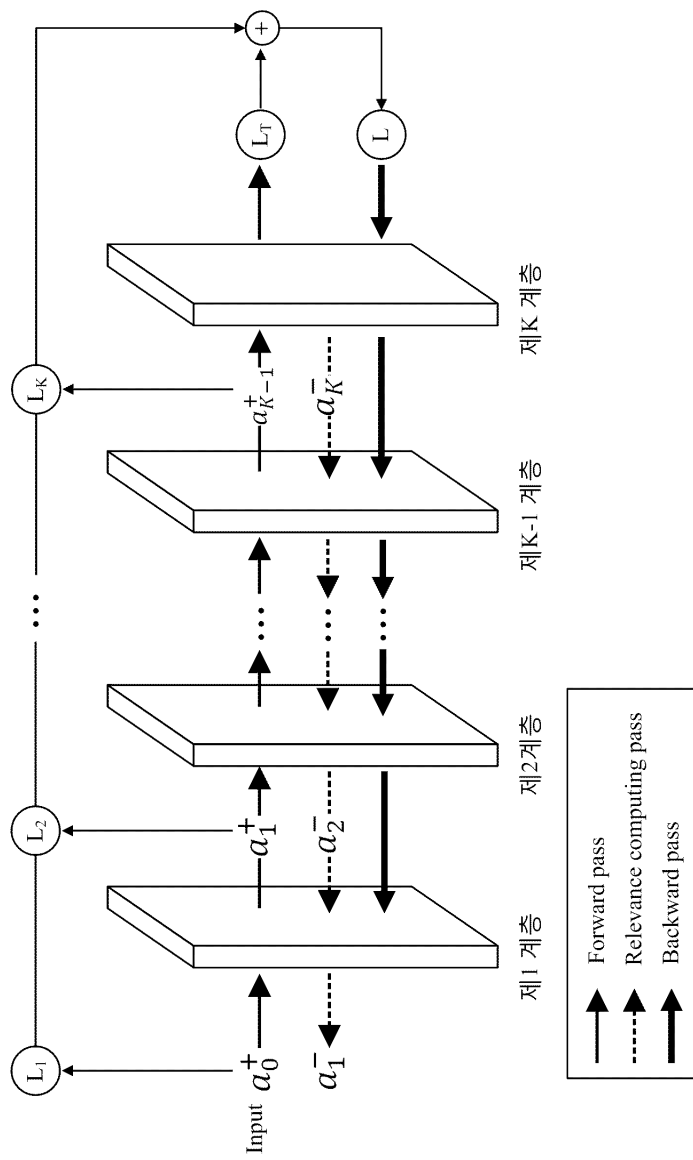
도면2



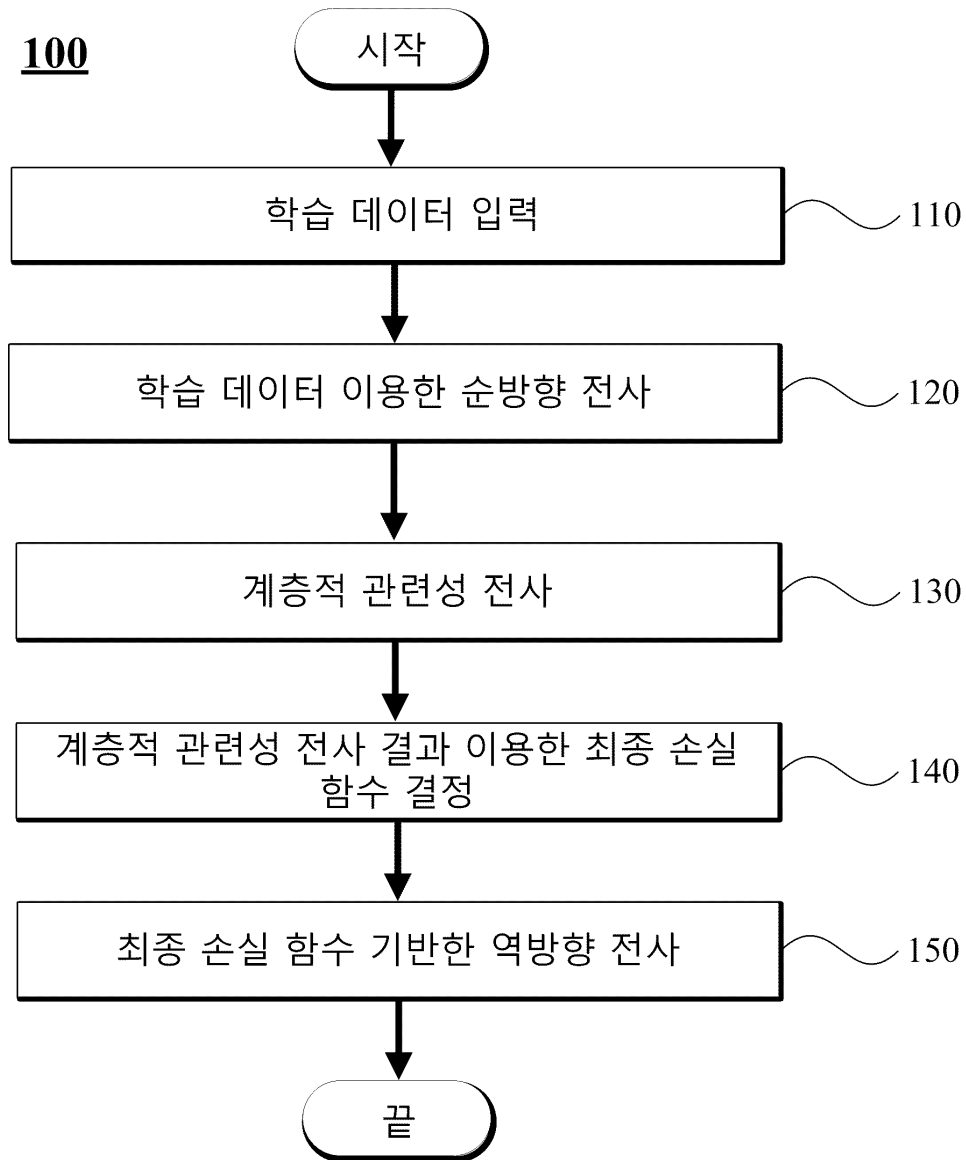
도면3



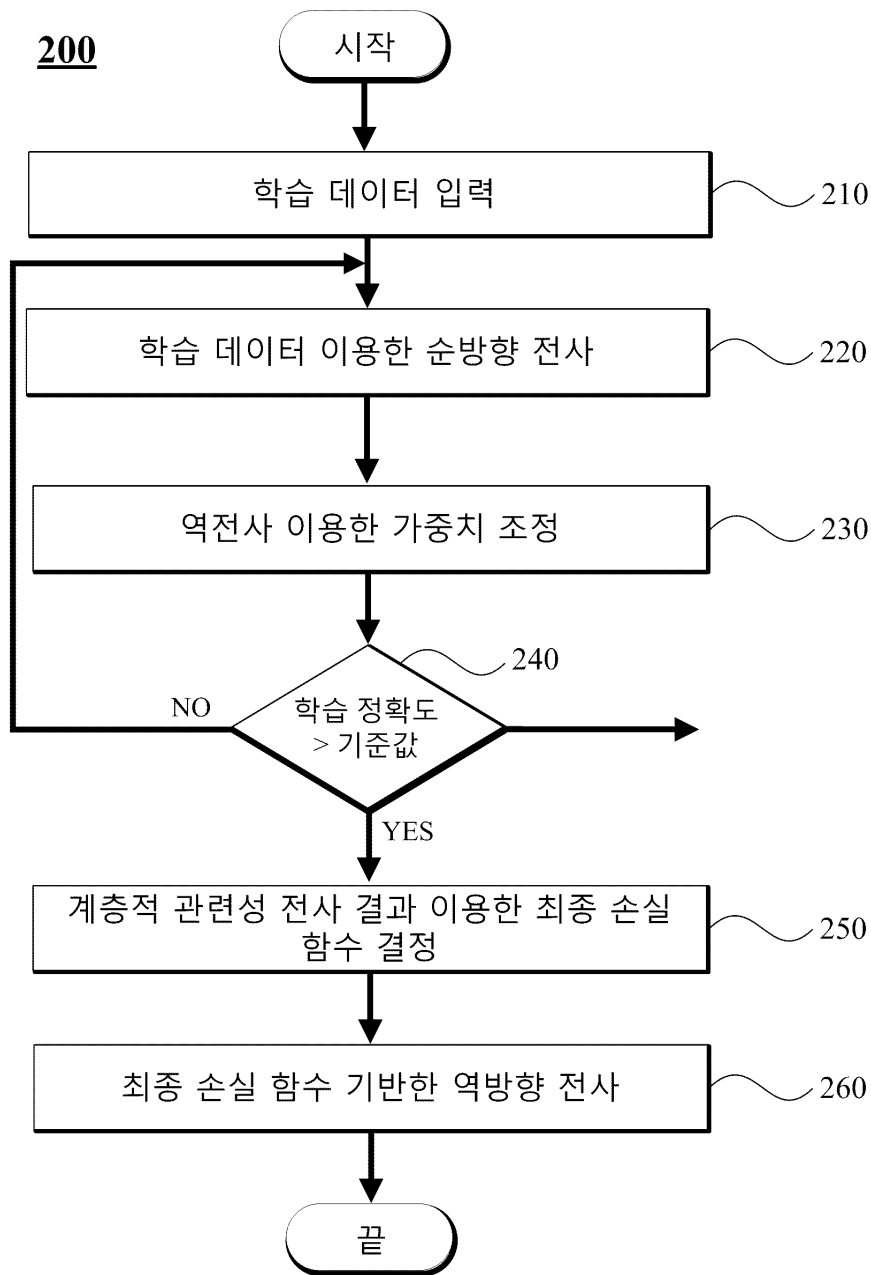
도면4



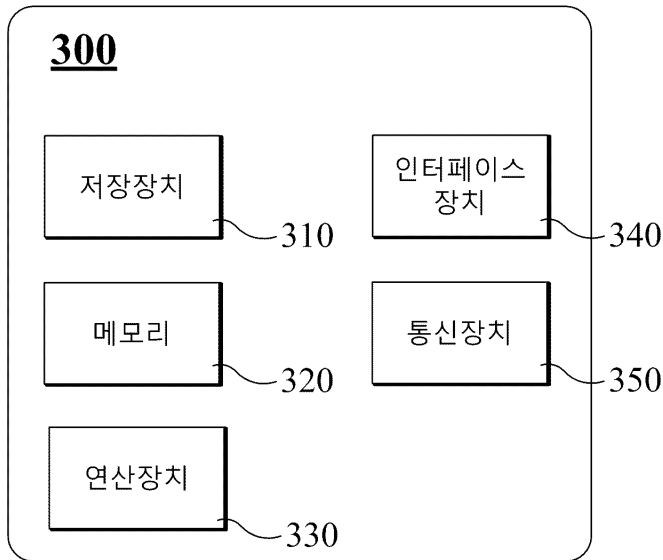
도면5



도면6



도면7



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 청구항 1

【변경전】

학습 장치가 신경망에 대하여 학습 데이터를 이용한 순방향 전사 과정을 수행하여 제1 손실함수를 결정하는 단계;

상기 학습 장치가 상기 신경망에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 계층별 손실함수를 연산하는 단계;

상기 학습 장치가 계층별 손실함수를 합산한 결과로 생성되는 제2 손실함수와 상기 제1 손실함수를 합산하여 최종 손실함수를 결정하는 단계; 및

상기 학습 장치가 상기 최종 손실함수를 이용하여 역방향 전사 과정을 수행하면서 상기 각 계층의 가중치를 갱신하는 단계를 포함하는 관련성 전사를 이용한 신경망 학습 방법.

【변경후】

학습 장치가 신경망에 대하여 학습 데이터를 이용한 순방향 전사 과정을 수행하여 제1 손실함수를 결정하는 단계;

상기 학습 장치가 상기 신경망에 대하여 역방향 관련성 전사 과정에서 각 계층의 특징 맵과 역방향 관련성 정보를 기준으로 계층별 손실함수를 연산하는 단계;

상기 학습 장치가 계층별 손실함수를 합산한 결과로 생성되는 제2 손실함수와 상기 제1 손실함수를 합산하여 최종 손실함수를 결정하는 단계; 및

상기 학습 장치가 상기 최종 손실함수를 이용하여 역방향 전사 과정을 수행하면서 상기 각 계층의 가중치를 갱신하는 단계를 포함하는 관련성 전사를 이용한 신경망 학습 방법.