



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0100100
(43) 공개일자 2024년07월01일

(51) 국제특허분류(Int. Cl.)
G06N 3/063 (2023.01) G06N 3/048 (2023.01)
H04L 9/00 (2022.01)
(52) CPC특허분류
G06N 3/063 (2013.01)
G06N 3/048 (2023.01)
(21) 출원번호 10-2022-0182216
(22) 출원일자 2022년12월22일
심사청구일자 없음

(71) 출원인
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
서울대학교산학협력단
서울특별시 관악구 관악로 1 (신림동)
(뒷면에 계속)
(72) 발명자
노종선
서울특별시 동작구 사당로13길 31, 104동 2103호
(사당동, 두산위브트레지움아파트)
김용준
대구광역시 달성군 현풍읍 테크노중앙대로 333,
402동 502호(비슬빌리지)
(뒷면에 계속)
(74) 대리인
특허법인 무한

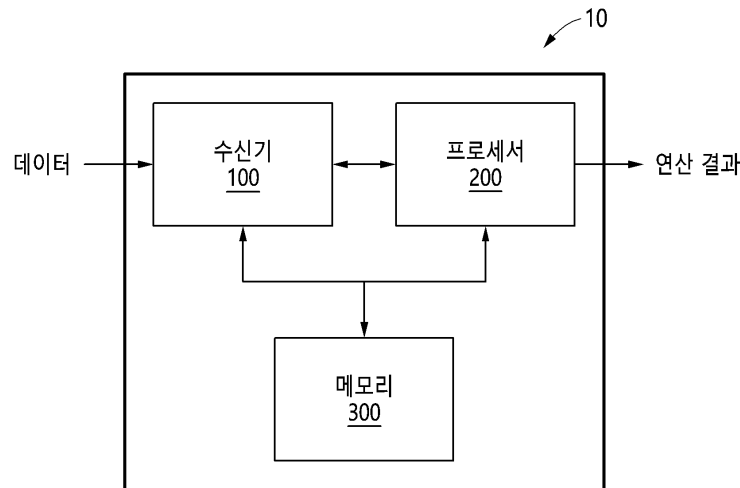
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 동형 암호화된 데이터의 뉴럴 네트워크 연산 장치 및 방법

(57) 요약

동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 장치 및 방법이 개시된다. 동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 장치에 있어서, 일 실시예에 따른 연산 장치는, 상기 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 상기 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신하는 수신기와, 상기 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산하고, 상기 파라미터, 상기 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 상기 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성하고, 상기 근사 다항식에 기초하여 상기 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성하는 프로세서를 포함한다.

대표도 - 도1



(52) CPC특허분류

H04L 9/008 (2013.01)

(71) 출원인

재단법인대구경북과학기술원

대구 달성군 현풍읍 테크노중앙대로 333

조선대학교산학협력단

광주광역시 동구 조선대길 126(서석동,
산학협력단)

(72) 발명자

이은상

서울특별시 송파구 양산로4길 16, 506동 306호(거
여동, 거여5단지아파트)

이정현

서울특별시 관악구 인현 16길 25-10, 302호(
봉천동, 우영아트빌)

김영식

광주광역시 서구 풍암2로 66, 201동 1005호(
풍암동, 금호타운)

이준우

서울특별시 관악구 인현 12다길 31, 302호(봉천동)

명세서

청구범위

청구항 1

동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 장치에 있어서,
 상기 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 상기 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신하는 수신기; 및
 상기 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산하고,
 상기 파라미터, 상기 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 상기 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성하고,
 상기 근사 다항식에 기초하여 상기 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성하는 프로세서
 를 포함하는 연산 장치.

청구항 2

제1항에 있어서,
 상기 파라미터는,
 상기 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함하는,
 연산 장치.

청구항 3

제2항에 있어서,
 상기 프로세서는,
 상기 보정 상수와 상기 표준 편차를 곱함으로써 상기 표준 편차를 보정하는,
 연산 장치.

청구항 4

제1항에 있어서,
 상기 뉴럴 네트워크 연산은,
 정류 선형 유닛(rectified linear unit)을 포함하는,
 연산 장치.

청구항 5

제4항에 있어서,
 상기 프로세서는,
 상기 정류 선형 유닛의 입력 데이터에 기초하여 상기 평균 및 상기 표준 편차를 계산하는,

연산 장치.

청구항 6

제2항에 있어서,

상기 프로세서는,

상기 평균, 상기 표준 편차 및 상기 보정 상수에 기초하여 상기 입력 데이터의 확률 밀도 함수를 설정하고,

상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성하는,

연산 장치.

청구항 7

제6항에 있어서,

상기 프로세서는,

상기 확률 밀도 함수에 기초한 평균 제공 오차를 계산하고,

상기 근사 다항식의 차수 및 상기 확률 밀도 함수에 기초하여 상기 평균 제공 오차가 최소가 되는 다항식을 상기 근사 다항식으로 생성하는,

연산 장치.

청구항 8

제7항에 있어서,

상기 프로세서는,

정류 선형 유닛과 상기 근사 다항식의 차이의 제곱 및 상기 확률 밀도 함수의 곱셈에 기초하여 상기 평균 제공 오차를 계산하는,

연산 장치.

청구항 9

제1항에 있어서,

상기 프로세서는,

상기 근사 다항식의 차수 상기 평균 및 상기 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산하고,

상기 제1 계수 및 상기 제2 계수의 곱셈에 기초하여 상기 근사 다항식을 구성하는 타겟 계수를 계산함으로써 상기 근사 다항식을 생성하는,

연산 장치.

청구항 10

제9항에 있어서,

상기 프로세서는,

상기 평균을 상기 표준 편차로 나눈 값에 기초하여 상기 제1 계수 및 상기 제2 계수를 계산하는, 연산 장치.

청구항 11

동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 방법에 있어서,

상기 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 상기 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신하는 단계;

상기 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산하는 단계;

상기 파라미터, 상기 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 상기 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성하는 단계; 및

상기 근사 다항식에 기초하여 상기 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성하는 단계를 포함하는 연산 방법.

청구항 12

제11항에 있어서,

상기 파라미터는,

상기 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함하는, 연산 방법.

청구항 13

제12항에 있어서,

상기 근사 다항식을 생성하는 단계는,

상기 보정 상수와 상기 표준 편차를 곱함으로써 상기 표준 편차를 보정하는 단계를 포함하는 연산 방법.

청구항 14

제11항에 있어서,

상기 뉴럴 네트워크 연산은,

정류 선형 유닛(rectified linear unit)을 포함하는,

연산 방법.

청구항 15

제14항에 있어서,

상기 평균 및 표준 편차를 계산하는 단계는,

상기 정류 선형 유닛의 입력 데이터에 기초하여 상기 평균 및 상기 표준 편차를 계산하는 단계
를 포함하는 연산 방법.

청구항 16

제12항에 있어서,

상기 근사 다항식을 생성하는 단계는,

상기 평균, 상기 표준 편차 및 상기 보정 상수에 기초하여 상기 입력 데이터의 확률 밀도 함수를 설정하는 단계; 및

상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성하는 단계

를 더 포함하는 연산 방법.

청구항 17

제16항에 있어서,

상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성하는 단계는,

상기 확률 밀도 함수에 기초한 평균 제곱 오차를 계산하는 단계; 및

상기 근사 다항식의 차수 및 상기 확률 밀도 함수에 기초하여 상기 평균 제곱 오차가 최소가 되는 다항식을 상기 근사 다항식으로 생성하는 단계

를 포함하는 연산 방법.

청구항 18

제17항에 있어서,

상기 평균 제곱 오차를 계산하는 단계는,

정류 선형 유닛과 상기 근사 다항식의 차이의 제곱 및 상기 확률 밀도 함수의 곱셈에 기초하여 상기 평균 제곱 오차를 계산하는 단계

를 포함하는 연산 방법.

청구항 19

제1항에 있어서,

상기 근사 다항식을 생성하는 단계는,

상기 근사 다항식의 차수 상기 평균 및 상기 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산하는 단계; 및

상기 제1 계수 및 상기 제2 계수의 곱셈에 기초하여 상기 근사 다항식을 구성하는 타겟 계수를 계산함으로써 상기 근사 다항식을 생성하는 단계

를 포함하는 연산 방법.

청구항 20

제19항에 있어서,

상기 제1 계수 및 제2 계수를 계산하는 단계는,

상기 평균을 상기 표준 편차로 나눈 값에 기초하여 상기 제1 계수 및 상기 제2 계수를 계산하는 단계를 포함하는 연산 방법.

발명의 설명

기술 분야

[0001] 아래 실시예들은 동형 암호화된 데이터의 뉴럴 네트워크 연산 장치 및 방법에 관한 것이다.

배경 기술

[0002] 동형 암호(homomorphic encryption)는 암호화된 데이터 사이에서 임의의 연산을 가능하게 하는 유망한 암호화 방법이다. 동형 암호를 활용하면 암호화된 데이터를 복호화하지 않고, 암호화된 상태에서 임의의 연산을 수행할 수 있을 뿐만 아니라, 격자를 기반(lattice-based)으로 하여 양자 알고리즘에 내성(resistant)이 있어 안전하다.

[0003] 완전 동형 암호화된 데이터를 위한 뉴럴 네트워크 연산의 수행을 위해서는 활성화 함수로 낮은 차수의 다항식이 사용되었다. 하지만, 뉴럴 네트워크 연산 중에서 정류 선형 유닛을 다항식으로 정확하게 근사하기 위해서는 높은 차수의 다항식이 요구된다.

[0004] 완전 동형 암호화된 데이터를 이용한 뉴럴 네트워크 연산의 수행을 위해서 낮은 차수의 다항식을 사용할 경우에 뉴럴 네트워크의 레이어가 깊이 쌓일 수 없고 높은 성능을 달성하기가 어렵다.

[0005] 완전 동형 암호화된 데이터를 위한 뉴럴 네트워크 연산에 있어서, 정류 선형 유닛 함수와 같이 범용적으로 사용되는 활성화 함수를 사용할 수 없기 때문에, 기학습된 뉴럴 네트워크를 그대로 가져와 사용할 수 없고 사용자는 모델을 새롭게 학습시켜야만 했다.

[0006] 또한, 정류 선형 유닛을 다항식으로 정확하게 근사하게 위해서는 높은 차수의 다항식이 요구되기 때문에, 깊이가 깊은 뉴럴 네트워크에서 완전 동형 암호화 연산을 구현하기 위해 많은 부트스트래핑(bootstrapping)이 요구되어 과도한 시간 소요가 필요했다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0007] 동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 장치에 있어서, 일 실시예에 따른 연산 장치는, 상기 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 상기 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신하는 수신기와, 상기 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산하고, 상기 파라미터, 상기 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 상기 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성하고, 상기 근사 다항식에 기초하여 상기 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성하는 프로세서를 포함한다.

[0008] 상기 파라미터는, 상기 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함할 수 있다.

[0009] 상기 프로세서는, 상기 보정 상수와 상기 표준 편차를 곱함으로써 상기 표준 편차를 보정할 수 있다.

[0010] 상기 뉴럴 네트워크 연산은, 정류 선형 유닛(rectified linear unit)을 포함할 수 있다.

[0011] 상기 프로세서는, 상기 정류 선형 유닛의 입력 데이터에 기초하여 상기 평균 및 상기 표준 편차를 계산할 수 있다.

[0012] 상기 프로세서는, 상기 평균, 상기 표준 편차 및 상기 보정 상수에 기초하여 상기 입력 데이터의 확률 밀도 함수를 설정하고, 상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성할 수 있다.

- [0013] 상기 프로세서는, 상기 확률 밀도 함수에 기초한 평균 제곱 오차를 계산하고, 상기 근사 다항식의 차수 및 상기 확률 밀도 함수에 기초하여 상기 평균 제곱 오차가 최소가 되는 다항식을 상기 근사 다항식으로 생성할 수 있다.
- [0014] 상기 프로세서는, 정류 선형 유닛과 상기 근사 다항식의 차이의 제곱 및 상기 확률 밀도 함수의 곱셈에 기초하여 상기 평균 제곱 오차를 계산할 수 있다.
- [0015] 상기 프로세서는, 상기 근사 다항식의 차수 상기 평균 및 상기 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산하고, 상기 제1 계수 및 상기 제2 계수의 곱셈에 기초하여 상기 근사 다항식을 구성하는 타겟 계수를 계산함으로써 상기 근사 다항식을 생성할 수 있다.
- [0016] 상기 프로세서는, 상기 평균을 상기 표준 편차로 나눈 값에 기초하여 상기 제1 계수 및 상기 제2 계수를 계산할 수 있다.
- [0017] 동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행하는 연산 방법에 있어서, 일 실시예에 따른 연산 방법은, 상기 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 상기 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신하는 단계와, 상기 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산하는 단계와, 상기 파라미터, 상기 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 상기 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성하는 단계와, 상기 근사 다항식에 기초하여 상기 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성하는 단계를 포함한다.
- [0018] 상기 파라미터는, 상기 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함할 수 있다.
- [0019] 상기 근사 다항식을 생성하는 단계는, 상기 보정 상수와 상기 표준 편차를 곱함으로써 상기 표준 편차를 보정하는 단계를 포함할 수 있다.
- [0020] 상기 뉴럴 네트워크 연산은, 정류 선형 유닛(rectified linear unit)을 포함할 수 있다.
- [0021] 상기 평균 및 표준 편차를 계산하는 단계는, 상기 정류 선형 유닛의 입력 데이터에 기초하여 상기 평균 및 상기 표준 편차를 계산하는 단계를 포함할 수 있다.
- [0022] 상기 근사 다항식을 생성하는 단계는, 상기 평균, 상기 표준 편차 및 상기 보정 상수에 기초하여 상기 입력 데이터의 확률 밀도 함수를 설정하는 단계와, 상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성하는 단계를 더 포함할 수 있다.
- [0023] 상기 확률 밀도 함수에 기초하여 상기 근사 다항식을 생성하는 단계는, 상기 확률 밀도 함수에 기초한 평균 제곱 오차를 계산하는 단계와, 상기 근사 다항식의 차수 및 상기 확률 밀도 함수에 기초하여 상기 평균 제곱 오차가 최소가 되는 다항식을 상기 근사 다항식으로 생성하는 단계를 포함할 수 있다.
- [0024] 상기 평균 제곱 오차를 계산하는 단계는, 정류 선형 유닛과 상기 근사 다항식의 차이의 제곱 및 상기 확률 밀도 함수의 곱셈에 기초하여 상기 평균 제곱 오차를 계산하는 단계를 포함할 수 있다.
- [0025] 상기 근사 다항식을 생성하는 단계는, 상기 근사 다항식의 차수 상기 평균 및 상기 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산하는 단계와, 상기 제1 계수 및 상기 제2 계수의 곱셈에 기초하여 상기 근사 다항식을 구성하는 타겟 계수를 계산함으로써 상기 근사 다항식을 생성하는 단계를 포함할 수 있다.
- [0026] 상기 제1 계수 및 제2 계수를 계산하는 단계는, 상기 평균을 상기 표준 편차로 나눈 값에 기초하여 상기 제1 계수 및 상기 제2 계수를 계산하는 단계를 포함할 수 있다.

도면의 간단한 설명

- [0027] 도 1은 일 실시예에 따른 뉴럴 네트워크 연산 장치의 개략적인 블록도를 나타낸다.
- 도 2는 도 1에 도시된 뉴럴 네트워크 연산 장치가 근사 다항식을 생성하는 과정을 설명하기 위한 도면이다.
- 도 3은 입력 데이터의 평균 및 표준 편차를 계산하는 과정을 설명하기 위한 도면이다.
- 도 4는 근사 다항식을 생성하는 과정의 흐름도를 나타낸다.
- 도 5는 근사 다항식의 계수를 계산하는 과정을 나타낸다.
- 도 6은 표준 편차의 보정을 위한 보정 상수를 설명하기 위한 도면이다.

도 7은 도 1에 도시된 뉴럴 네트워크 연산 장치의 동작의 흐름도를 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0028] 이하에서, 첨부된 도면을 참조하여 실시예들을 상세하게 설명한다. 그러나, 실시예들에는 다양한 변경이 가해질 수 있어서 특허출원의 권리 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 실시예들에 대한 모든 변경, 균등물 내지 대체물이 권리 범위에 포함되는 것으로 이해되어야 한다.
- [0029] 실시예에서 사용한 용어는 단지 설명을 목적으로 사용된 것으로, 한정하려는 의도로 해석되어서는 안된다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 명세서 상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0030] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0031] 또한, 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조부호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다. 실시예를 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 실시예의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.
- [0032] 또한, 실시 예의 구성 요소를 설명하는 데 있어서, 제 1, 제 2, A, B, (a), (b) 등의 용어를 사용할 수 있다. 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등이 한정되지 않는다. 어떤 구성 요소가 다른 구성요소에 "연결", "결합" 또는 "접속"된다고 기재된 경우, 그 구성 요소는 그 다른 구성요소에 직접적으로 연결되거나 접속될 수 있지만, 각 구성 요소 사이에 또 다른 구성 요소가 "연결", "결합" 또는 "접속"될 수도 있다고 이해되어야 할 것이다.
- [0033] 어느 하나의 실시 예에 포함된 구성요소와, 공통적인 기능을 포함하는 구성요소는, 다른 실시 예에서 동일한 명칭을 사용하여 설명하기로 한다. 반대되는 기재가 없는 이상, 어느 하나의 실시 예에 기재한 설명은 다른 실시 예에도 적용될 수 있으며, 중복되는 범위에서 구체적인 설명은 생략하기로 한다.
- [0035] 도 1은 일 실시예에 따른 뉴럴 네트워크 연산 장치의 개략적인 블록도를 나타낸다.
- [0036] 도 1을 참조하면, 뉴럴 네트워크 연산 장치(10)는 뉴럴 네트워크 연산을 수행할 수 있다. 뉴럴 네트워크 연산은 뉴럴 네트워크를 이용한 학습 또는 추론(inference)을 수행하기 위해 요구되는 연산을 포함할 수 있다.
- [0037] 뉴럴 네트워크 연산 장치(10)는 동형 암호화(homomorphic encrypted)된 데이터의 뉴럴 네트워크 연산을 수행할 수 있다. 동형 암호는 데이터를 암호화한 상태에서 다양한 연산을 수행할 수 있도록 구성된 암호화 방식을 의미할 수 있다. 동형 암호에서 암호문(ciphertext)들을 이용한 연산의 결과는 새로운 암호문이 되고, 암호문을 복호화함으로써 획득된 평문(plaintext)은 암호화하기 전의 원래 데이터의 연산 결과와 동일할 수 있다.
- [0038] 뉴럴 네트워크는 시냅스의 결합으로 네트워크를 형성한 인공 뉴런(노드)이 학습을 통해 시냅스의 결합 세기를 변화시켜, 문제 해결 능력을 가지는 모델 전반을 의미할 수 있다.
- [0039] 뉴럴 네트워크의 뉴런은 가중치 또는 바이어스의 조합을 포함할 수 있다. 뉴럴 네트워크는 하나 이상의 뉴런 또는 노드로 구성된 하나 이상의 레이어(layer)를 포함할 수 있다. 뉴럴 네트워크는 뉴런의 가중치를 학습을 통해 변화시킴으로써 임의의 입력으로부터 예측하고자 하는 결과를 추론할 수 있다.
- [0040] 뉴럴 네트워크는 심층 뉴럴 네트워크 (Deep Neural Network)를 포함할 수 있다. 뉴럴 네트워크는 CNN(Convolutional Neural Network), RNN(Recurrent Neural Network), 퍼셉트론(perceptron), 다층 퍼셉트론(multilayer perceptron), FF(Feed Forward), RBF(Radial Basis Network), DFF(Deep Feed Forward), LSTM(Long Short Term Memory), GRU(Gated Recurrent Unit), AE(Auto Encoder), VAE(Variational Auto Encoder), DAE(Denoising Auto Encoder), SAE(Sparse Auto Encoder), MC(Markov Chain), HN(Hopfield

Network), BM(Boltzmann Machine), RBM(Restricted Boltzmann Machine), DBN(Depp Belief Network), DCN(Deep Convolutional Network), DN(Deconvolutional Network), DCIGN(Deep Convolutional Inverse Graphics Network), GAN(Generative Adversarial Network), LSM(Liquid State Machine), ELM(Extreme Learning Machine), ESN(Echo State Network), DRN(Deep Residual Network), DNC(Differentiable Neural Computer), NTM(Neural Turning Machine), CN(Capsule Network), KN(Kohonen Network), BNN(Binarized Neural Network) 및 AN(Attention Network)를 포함할 수 있다.

- [0041] 뉴럴 네트워크 연산 장치(10)는 PC(personal computer), 데이터 서버, 또는 휴대용 장치 내에 구현될 수 있다.
- [0042] 휴대용 장치는 랩탑(laptop) 컴퓨터, 이동 전화기, 스마트 폰(smart phone), 태블릿(tablet) PC, 모바일 인터넷 디바이스(mobile internet device(MID)), PDA(personal digital assistant), EDA(enterprise digital assistant), 디지털 스틸 카메라(digital still camera), 디지털 비디오 카메라(digital video camera), PMP(portable multimedia player), PND(personal navigation device 또는 portable navigation device), 휴대용 게임 콘솔(handheld game console), e-북(e-book), 또는 스마트 디바이스(smart device)로 구현될 수 있다. 스마트 디바이스는 스마트 워치(smart watch), 스마트 밴드(smart band), 또는 스마트 링(smart ring)으로 구현될 수 있다.
- [0043] 뉴럴 네트워크 연산 가속 장치(10)는 가속기를 이용하여 뉴럴 네트워크 연산을 수행할 수 있다. 뉴럴 네트워크 연산 가속 장치(10)는 가속기를 이용하여 뉴럴 네트워크 연산을 수행할 수 있다. 뉴럴 네트워크 연산 가속 장치(10)는 가속기 내부 또는 외부에 구현될 수 있다.
- [0044] 가속기는 NPU(Neural Processing Unit), GPU(Graphics Processing Unit), FPGA(Field Programmable Gate Array), ASIC(Application Specific Integrated Circuit) 또는 AP(Application Processor)를 포함할 수 있다. 또한, 가속기는 가상 머신(Virtual Machine)과 같이 소프트웨어 컴퓨팅 환경으로 구현될 수도 있다.
- [0045] 뉴럴 네트워크 연산 장치(10)는 수신기(100) 및 프로세서(200)를 포함한다. 뉴럴 네트워크 연산 장치(10)는 메모리(300)를 더 포함할 수 있다.
- [0046] 수신기(100)는 수신 인터페이스를 포함할 수 있다. 수신기(100)는 데이터를 수신할 수 있다. 수신기(100)는 외부 또는 메모리(300)로부터 데이터를 수신할 수 있다. 수신기(100)는 수신한 데이터를 프로세서(200)로 출력할 수 있다.
- [0047] 수신기(100)는 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신할 수 있다.
- [0048] 뉴럴 네트워크 연산을 수행하기 위한 데이터는 뉴럴 네트워크 또는 뉴럴 네트워크를 구성하는 레이어에 입력되는 입력 데이터를 포함할 수 있다. 근사 다항식을 생성하기 위한 파라미터는 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함할 수 있다. 뉴럴 네트워크 연산은 정류 선형 유닛(Rectified Linear Unit(ReLU))을 포함할 수 있다.
- [0049] 프로세서(200)는 메모리(300)에 저장된 데이터를 처리할 수 있다. 프로세서(200)는 메모리(300)에 저장된 컴퓨터로 읽을 수 있는 코드(예를 들어, 소프트웨어) 및 프로세서(200)에 의해 유발된 인스트럭션(instruction)들을 실행할 수 있다.
- [0050] "프로세서(200)"는 목적하는 동작들(desired operations)을 실행시키기 위한 물리적인 구조를 갖는 회로를 가지는 하드웨어로 구현된 데이터 처리 장치일 수 있다. 예를 들어, 목적하는 동작들은 프로그램에 포함된 코드(code) 또는 인스트럭션들(instructions)을 포함할 수 있다.
- [0051] 예를 들어, 하드웨어로 구현된 데이터 처리 장치는 마이크로프로세서(microprocessor), 중앙 처리 장치(central processing unit), 프로세서 코어(processor core), 멀티-코어 프로세서(multi-core processor), 멀티프로세서(multiprocessor), ASIC(Application-Specific Integrated Circuit), FPGA(Field Programmable Gate Array)를 포함할 수 있다.
- [0052] 프로세서(200)는 뉴럴 네트워크 연산을 수행하기 위한 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산할 수 있다. 프로세서(200)는 정류 선형 유닛의 입력 데이터에 기초하여 평균 및 표준 편차를 계산할 수 있다. 프로세서(200)는 보정 상수와 표준 편차를 곱함으로써 표준 편차를 보정할 수 있다.

- [0053] 프로세서(200)는 파라미터, 평균 및 표준 편차에 기초한 가중 최소 제곱법(weighted least square)을 이용하여 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성할 수 있다. 예를 들어, 프로세서(200)는 정류 선형 유닛을 근사하는 근사 다항식을 생성할 수 있다.
- [0054] 프로세서(200)는 입력 데이터의 평균, 표준 편차 및 보정 상수에 기초하여 입력 데이터의 확률 밀도 함수를 설정할 수 있다. 프로세서(200)는 확률 밀도 함수에 기초하여 근사 다항식을 생성할 수 있다.
- [0055] 프로세서(200)는 확률 밀도 함수에 기초한 평균 제곱 오차를 계산할 수 있다. 프로세서(200)는 근사 다항식의 차수 및 확률 밀도 함수에 기초하여 평균 제곱 오차가 최소가 되는 다항식을 근사 다항식으로 생성할 수 있다. 프로세서(200)는 정류 선형 유닛과 근사 다항식의 차이의 제곱 및 확률 밀도 함수의 곱셈에 기초하여 평균 제곱 오차를 계산할 수 있다.
- [0056] 프로세서(200)는 근사 다항식의 차수 평균 및 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산할 수 있다. 프로세서(200)는 제1 계수 및 제2 계수의 곱셈에 기초하여 근사 다항식을 구성하는 타겟 계수를 계산함으로써 근사 다항식을 생성할 수 있다. 프로세서(200)는 평균을 표준 편차로 나눈 값에 기초하여 제1 계수 및 제2 계수를 계산할 수 있다.
- [0057] 프로세서(200)는 근사 다항식에 기초하여 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성할 수 있다.
- [0058] 메모리(300)는 프로세서(200)에 의해 실행가능한 인스트럭션들(또는 프로그램)을 저장할 수 있다. 예를 들어, 인스트럭션들은 프로세서(200)의 동작 및/또는 프로세서(200)의 각 구성의 동작을 실행하기 위한 인스트럭션들을 포함할 수 있다.
- [0059] 메모리(300)는 휘발성 메모리 장치 또는 불휘발성 메모리 장치로 구현될 수 있다.
- [0060] 휘발성 메모리 장치는 DRAM(dynamic random access memory), SRAM(static random access memory), T-RAM(thyristor RAM), Z-RAM(zero capacitor RAM), 또는 TTRAM(Twin Transistor RAM)으로 구현될 수 있다.
- [0061] 불휘발성 메모리 장치는 EEPROM(Electrically Erasable Programmable Read-Only Memory), 플래시(flash) 메모리, MRAM(Magnetic RAM), 스핀전달토크 MRAM(Spin-Transfer Torque(STT)-MRAM), Conductive Bridging RAM(CBRAM), FeRAM(Ferroelectric RAM), PRAM(Phase change RAM), 저항 메모리(Resistive RAM(RRAM)), 나노 튜브 RRAM(Nanotube RRAM), 폴리머 RAM(Polymer RAM(PoRAM)), 나노 부유 게이트 메모리(Nano Floating Gate Memory(NFGM)), 홀로그래픽 메모리(holographic memory), 분자 전자 메모리 소자(Molecular Electronic Memory Device), 또는 절연 저항 변화 메모리(Insulator Resistance Change Memory)로 구현될 수 있다.
- [0063] 도 2는 도 1에 도시된 뉴럴 네트워크 연산 장치가 근사 다항식을 생성하는 과정을 설명하기 위한 도면이고, 도 3은 입력 데이터의 평균 및 표준 편차를 계산하는 과정을 설명하기 위한 도면이다.
- [0064] 도 2 및 도 3을 참조하면, 프로세서(예: 도 1의 프로세서(200))는 완전 동형 암호화된 데이터의 뉴럴 네트워크 연산을 수행할 수 있다. 프로세서(200)는 뉴럴 네트워크 연산에 포함된 정류 선형 유닛(또는, 정류 선형 유닛 함수)을 근사할 수 있다. 프로세서(200)는 입력 데이터의 분포를 고려하여 효과적으로 정류 선형 유닛을 다항식으로 근사할 수 있다. 정류 선형 유닛은 $ReLU(x) = \max\{x, 0\}$ 와 같이 나타낼 수 있다.
- [0065] 프로세서(200)는 가중 최소 제곱법을 이용하여 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성할 수 있다. 프로세서(200)는 기학습된 딥러닝 모델(210)(또는, 뉴럴 네트워크(310)) 및 학습용 데이터 셋의 표본(220)에 기초하여 정류 선형 유닛 함수의 입력 값의 평균 및 표준 편차(230)를 계산할 수 있다. 데이터 셋의 표본의 크기는 사용자에게 의해 정의될 수 있다. 근사 다항식의 차수 n은 자연수일 수 있다.
- [0066] 프로세서(200)는 뉴럴 네트워크(310)를 구성하는 복수의 레이어에 포함된 정류 선형 유닛의 입력 데이터의 분포에 대한 정보를 획득할 수 있다. 입력 데이터의 분포에 대한 정보는 입력 데이터의 평균 및 표준 편차를 포함할 수 있다. 프로세서(200)는 획득한 평균 및 표준 편차를 이용하여 평균 제곱 오차(mean square error)를 최소화하는 근사 다항식을 생성할 수 있다.
- [0067] 프로세서(200)는 입력 데이터 값에 따라 다항식의 근사를 효과적으로 수행할 수 있다. 프로세서(200)는 낮은 다항식 차수를 사용하고도 높은 정확도를 갖는 완전 동형 암호화된 데이터를 위한 딥 러닝 모델을 설계할 수 있다.

[0068] 프로세서(200)는 입력 데이터의 분포를 획득할 수 있다. 프로세서(200)는 입력 데이터의 평균, 분산 및/또는 표준 편차를 계산할 수 있다. 프로세서(200)는 기 학습된 딥 러닝 모델(210) 상에서 학습용 데이터 셋을 분석함으로써 복수의 레이어의 정류 선형 유닛의 입력 데이터의 평균, 분산 및/또는 표준 편차를 계산할 수 있다.

[0069] 프로세서(200)는 평균 및 표준 편차(230), 근사 다항식의 차수(240) 및 표준 편차의 보정 상수(250)에 기초하여 뉴럴 네트워크 연산을 근사하는 근사 다항식(260)을 생성할 수 있다.

[0070] 뉴럴 네트워크(310)에 포함된 i 번째 레이어의 정류 선형 유닛의 입력 데이터의 평균 및 표준 편차는 각각 μ_i 및 σ_i 로 나타낼 수 있다. 프로세서(200)는 평균 및 표준 편차에 기초하여 확률 밀도 함수를 계산할 수 있다. i 번째 레이어의 표준 편차의 보정 상수가 k_i 일 때 확률 밀도 함수는 수학식 1과 같이 나타낼 수 있다.

수학식 1

$$p(x) = \frac{1}{\sqrt{2\pi(k_i\sigma_i)^2}} \exp\left(-\frac{(x-\mu_i)^2}{2(k_i\sigma_i)^2}\right)$$

[0074] 프로세서(200)는 확률 밀도 함수에 기초하여 평균 제곱 오차를 계산할 수 있다. 평균 제곱 오차는 수학식 2와 같이 나타낼 수 있다.

수학식 2

$$\int_{-\infty}^{\infty} p(x)(ReLU(x) - r(x))^2 dx$$

[0078] 프로세서(200)는 평균 제곱 오차를 최소화하는 n 차 다항식 $r(x)$ 를 근사 다항식(260)으로 생성할 수 있다.

[0080] 도 4는 근사 다항식을 생성하는 과정의 흐름도를 나타낸다.

[0081] 도 4를 참조하면, 프로세서(예: 도 1의 프로세서(200))는 학습용 데이터셋의 표준, 근사할 다항식의 차수 및 표준 편차의 보정 상수를 수신할 수 있다(410).

[0082] 프로세서(200)는 포워드패스(forward pass)를 통해 정류 선형 유닛의 입력 데이터의 평균 및 표준 편차를 획득할 수 있다(430).

[0083] 프로세서(200)는 평균 및 표준 편차에 기초하여 가중 최소 제곱법을 적용할 수 있다(450). 프로세서(200)는 가중 최소 제곱의 평균 제곱 오차가 최소가 되는 다항식을 근사 다항식으로 출력할 수 있다(470).

[0085] 도 5는 근사 다항식의 계수를 계산하는 과정을 나타낸다.

[0086] 도 5를 참조하면, 프로세서(예: 도 1의 프로세서(200))는 프로세서(200)는 입력 데이터의 분포를 이용하여 원하는 차수만큼 정류 선형 유닛을 근사하는 다항식의 계수를 생성함으로써 근사 다항식을 생성할 수 있다.

[0087] 프로세서(200)는 입력 데이터의 평균, 표준 편차 및 근사 다항식의 차수를 수신할 수 있다(511). μ 는 입력

데이터의 평균을 의미하고, σ 는 입력 데이터의 표준 편차를 보정한 값을 의미할 수 있다. 다시 말해, σ 는 보정된 표준 편차 또는 입력 데이터의 표준 편차에 보정 상수를 곱해준 값을 의미할 수 있다.

- [0088] n 은 근사 다항식의 차수를 의미할 수 있다. n 은 근사 다항식은 $p(x) = a_0 + a_1x + \dots + a_nx^n$ 의 형태를 가질 수 있다.
- [0089] 프로세서(200)는 평균 및 보정된 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산할 수 있다. 프로세서(200)는 평균을 보정된 표준 편차로 나눈 값에 기초하여 제1 계수의 초기값을 설정할 수 있다(513).
- [0090] 프로세서(200)는 i 가 n 보다 큰지 여부를 판단할 수 있다(515). i 가 n 보다 크지 않은 경우, 프로세서(200)는 제1 계수를 계산할 수 있다(517). 프로세서는 i 에 1을 더할 수 있다(519). 프로세서(200)는 515의 조건을 만족할 때까지 517, 519의 동작을 반복적으로 수행할 수 있다.
- [0091] 프로세서(200)는 평균을 보정된 표준 편차로 나눈 값에 기초하여 제2 계수의 초기값을 설정할 수 있다(521). 이때, i 는 4로 설정될 수 있다.
- [0092] 프로세서(200)는 i 가 n 보다 큰지 여부를 판단할 수 있다(523). i 가 n 보다 크지 않은 경우, 프로세서(200)는 제2 계수를 계산할 수 있다(525). 프로세서는 i 에 1을 더할 수 있다(527). 프로세서(200)는 523의 조건을 만족할 때까지 525, 527의 동작을 반복적으로 수행할 수 있다.
- [0093] 프로세서(200)는 제1 계수 및 제2 계수에 기초하여 근사 다항식을 구성하는 타겟 계수 a_k 를 생성할 수 있다(519).
- [0095] 도 6은 표준 편차의 보정을 위한 보정 상수를 설명하기 위한 도면이다.
- [0096] 도 6을 참조하면, 프로세서(예: 도 1의 프로세서(200))는 보정 상수를 조절하여 뉴럴 네트워크 연산의 성능을 향상시킬 수 있다. 생성된 근사 다항식을 이용하여 계산한 뉴럴 네트워크 연산 결과를 분석하여 연산 결과에 비정상적으로 큰 값 또는 오버플로우(overflow)가 발생하는 경우, 프로세서(200)는 보정 상수의 값을 증가시켜 새로운 근사 다항식을 생성할 수 있다.
- [0097] 연산 결과에 비정상적으로 큰 값 또는 오버플로우가 많이 발생하지 않는데도 뉴럴 네트워크 연산의 성능 저하가 일어날 경우, 프로세서(200)는 표준 편차의 보정 상수의 값을 감소시킴으로써 새로운 근사 다항식을 생성할 수 있다.
- [0098] 프로세서(200)는 접근 가능한 학습용 데이터의 표본을 사용하여 타겟 성능을 획득할 때까지 표준 편차의 보정 상수를 조절할 수 있다. 도 6의 예시는, 뉴럴 네트워크의 성능(예: 탑-1 정확도(top-1 accuracy))이 최대가 되는 표준 편차의 보정 상수를 나타낸 그래프일 수 있다.
- [0100] 도 7은 도 1에 도시된 뉴럴 네트워크 연산 장치의 동작의 흐름도를 나타낸다.
- [0101] 도 7을 참조하면, 수신기(예: 도 1의 수신기(100))는 뉴럴 네트워크 연산을 수행하기 위한 데이터 및 뉴럴 네트워크 연산에 대응하는 근사 다항식을 생성하기 위한 파라미터를 수신할 수 있다(710).
- [0102] 뉴럴 네트워크 연산을 수행하기 위한 데이터는 뉴럴 네트워크 또는 뉴럴 네트워크를 구성하는 레이어에 입력되는 입력 데이터를 포함할 수 있다. 근사 다항식을 생성하기 위한 파라미터는 근사 다항식의 차수 및 상기 표준 편차를 보정하기 위한 보정 상수를 포함할 수 있다. 뉴럴 네트워크 연산은 정류 선형 유닛을 포함할 수 있다.
- [0103] 프로세서(200)는 뉴럴 네트워크 연산을 수행하기 위한 데이터에 기초하여 뉴럴 네트워크를 구성하는 레이어의 입력 데이터의 평균 및 표준 편차를 계산할 수 있다(730). 프로세서(200)는 정류 선형 유닛의 입력 데이터에 기초하여 평균 및 표준 편차를 계산할 수 있다. 프로세서(200)는 보정 상수와 표준 편차를 곱함으로써 표준 편차를 보정할 수 있다.
- [0104] 프로세서(200)는 파라미터, 평균 및 표준 편차에 기초한 가중 최소 제곱법을 이용하여 뉴럴 네트워크 연산을 근사하는 근사 다항식을 생성할 수 있다(750).
- [0105] 프로세서(200)는 입력 데이터의 평균, 표준 편차 및 보정 상수에 기초하여 입력 데이터의 확률 밀도 함수를 설

정할 수 있다. 프로세서(200)는 확률 밀도 함수에 기초하여 근사 다항식을 생성할 수 있다.

[0106] 프로세서(200)는 확률 밀도 함수에 기초한 평균 제곱 오차를 계산할 수 있다. 프로세서(200)는 근사 다항식의 차수 및 확률 밀도 함수에 기초하여 평균 제곱 오차가 최소가 되는 다항식을 근사 다항식으로 생성할 수 있다. 프로세서(200)는 정류 선형 유닛과 근사 다항식의 차이의 제곱 및 확률 밀도 함수의 곱셈에 기초하여 평균 제곱 오차를 계산할 수 있다.

[0107] 프로세서(200)는 근사 다항식의 차수 평균 및 표준 편차에 기초하여 제1 계수 및 제2 계수를 계산할 수 있다. 프로세서(200)는 제1 계수 및 제2 계수의 곱셈에 기초하여 근사 다항식을 구성하는 타겟 계수를 계산함으로써 근사 다항식을 생성할 수 있다. 프로세서(200)는 평균을 표준 편차로 나눈 값에 기초하여 제1 계수 및 제2 계수를 계산할 수 있다.

[0108] 프로세서(200)는 근사 다항식에 기초하여 뉴럴 네트워크 연산을 수행함으로써 연산 결과를 생성할 수 있다 (770).

[0110] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

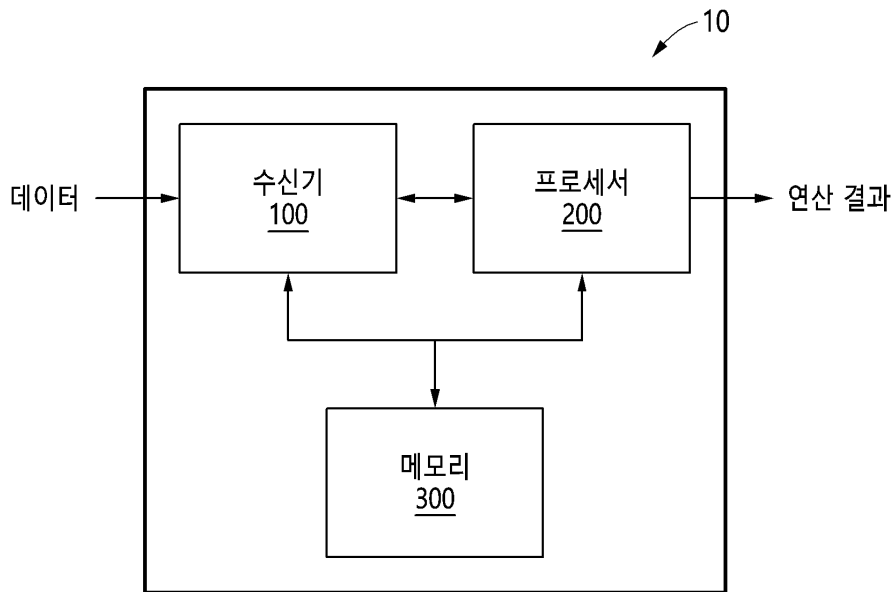
[0112] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 매체에 저장될 수 있다.

[0114] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

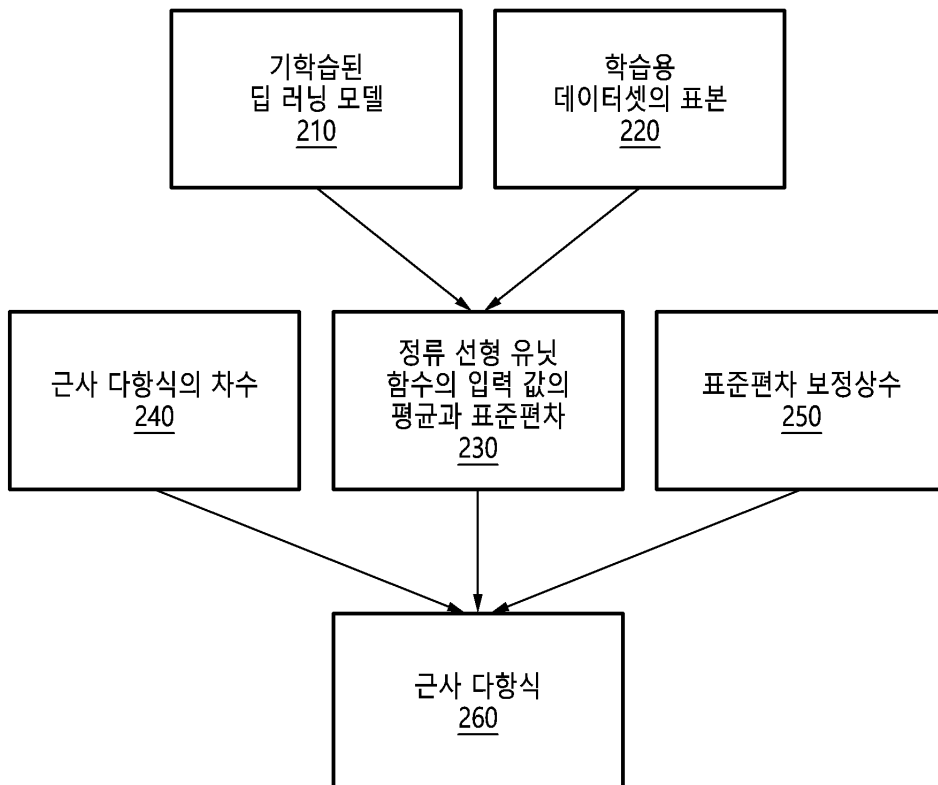
[0115] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 청구범위의 범위에 속한다.

도면

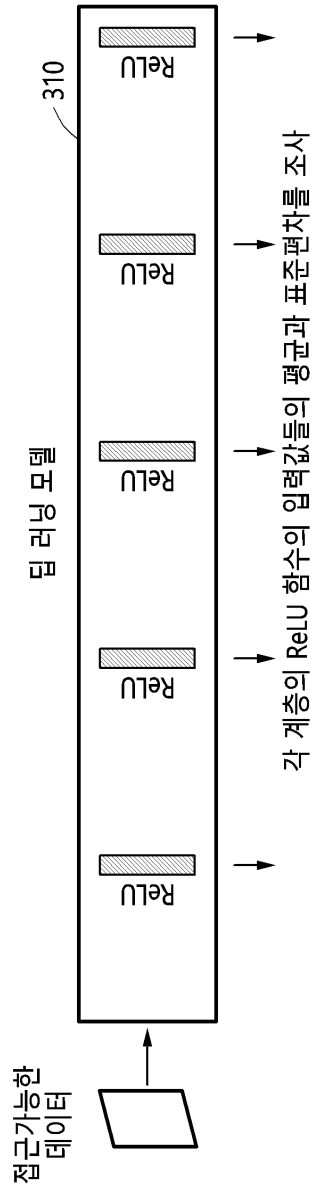
도면1



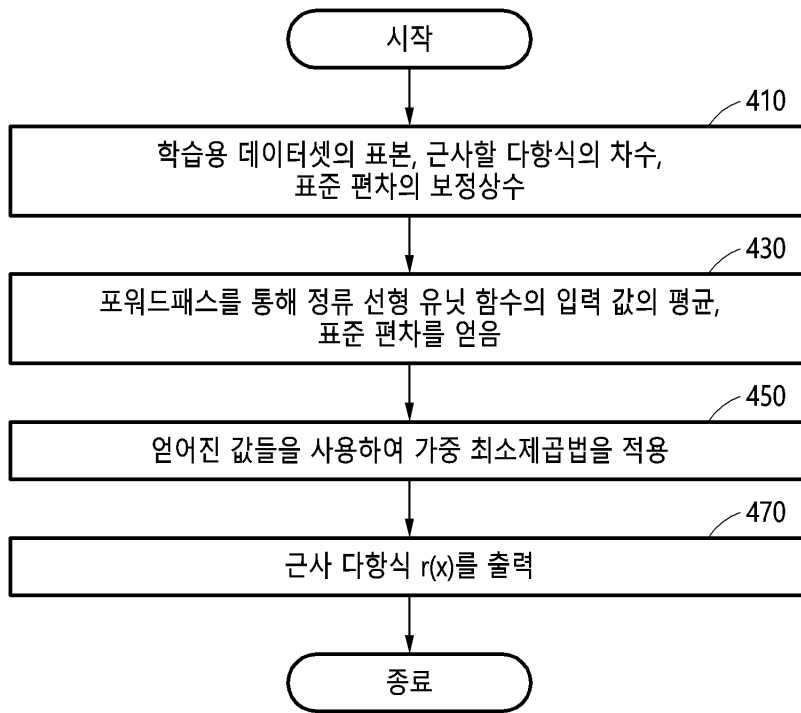
도면2



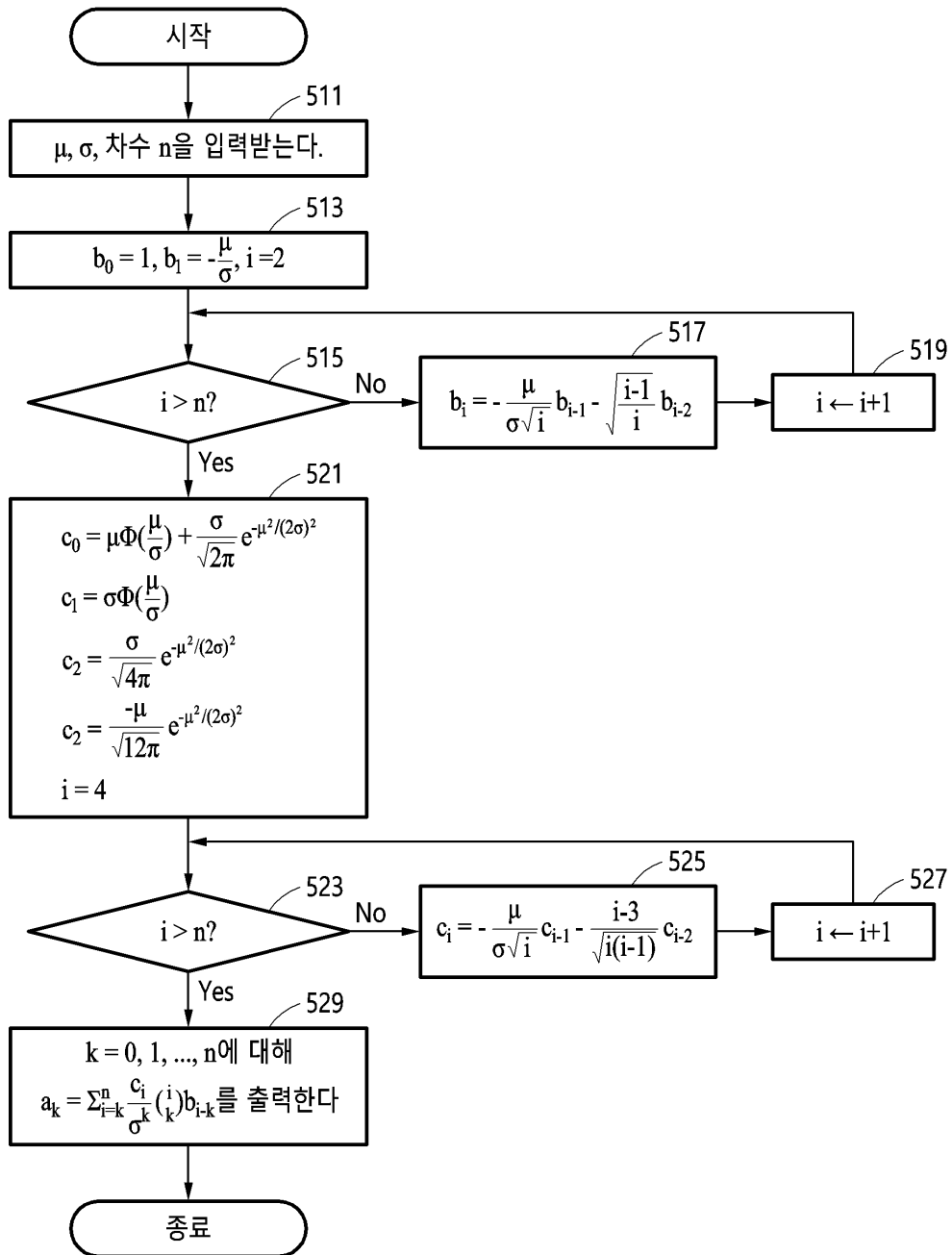
도면3



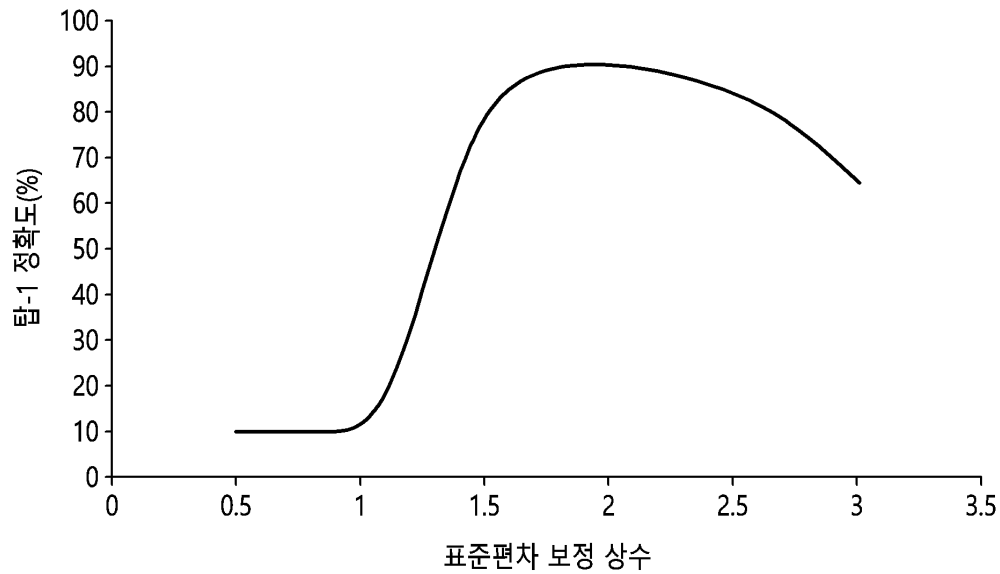
도면4



도면5



도면6



도면7

