



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년01월30일
(11) 등록번호 10-2071582
(24) 등록일자 2020년01월22일

- (51) 국제특허분류(Int. Cl.)
G06F 16/00 (2019.01) G06N 3/04 (2006.01)
G06N 3/08 (2006.01)
- (52) CPC특허분류
G06F 16/3347 (2019.01)
G06N 3/0454 (2013.01)
- (21) 출원번호 10-2018-0055651
- (22) 출원일자 2018년05월15일
심사청구일자 2018년05월15일
- (65) 공개번호 10-2018-0125905
- (43) 공개일자 2018년11월26일
- (30) 우선권주장
62/506,724 2017년05월16일 미국(US)
- (56) 선행기술조사문헌
이동건 외 2인, 한국과학기술원 전산학부, RNN을
이용한 한국어 문장 간 구문 유사도 측정
(2016.06.04. 공개)*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
삼성전자주식회사
경기도 수원시 영통구 삼성로 129 (매탄동)
- (72) 발명자
송희준
서울특별시 영등포구 도림로145길 28, 101동 160
1호(문래동4가, 리버뷰 신안인스빌)
쿨카르니 니레시
경기도 수원시 영통구 매영로 207(원천동)
- (74) 대리인
리앤목특허법인

전체 청구항 수 : 총 18 항

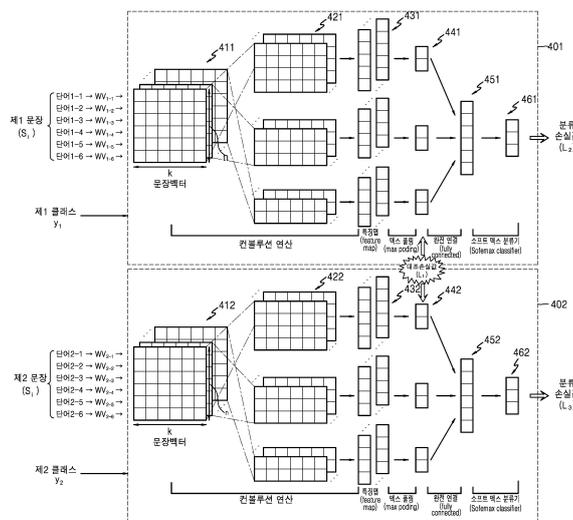
심사관 : 고재용

(54) 발명의 명칭 딥 뉴럴 네트워크(Deep Neural Network)를 이용하여 문장이 속하는 클래스(class)를 분류하는 방법 및 장치

(57) 요약

딥 뉴럴 네트워크(Deep Neural Network)를 이용하여 문장이 속하는 클래스(class)를 분류(classification)하는 방법 및 장치를 제공한다. 본 개시의 일 실시예에 따르면, 제1 뉴럴 네트워크(neural network) 및 제2 뉴럴 네트워크를 통해 제1 문장 및 제2 문장을 각각 학습하고, 학습의 출력 데이터로 생성된 제1 특징 벡터 및 제2 특징 벡터와 제1 문장과 제2 문장이 속하는 클래스의 동일 여부에 기초하여 대조 손실값(contrastive loss)을 획득하고, 대조 손실값이 최대가 되도록 학습을 반복하는 방법 및 장치를 제공한다.

대표도



(52) CPC특허분류
G06N 3/08 (2013.01)

명세서

청구범위

청구항 1

딥 뉴럴 네트워크(Deep Neural Network)를 이용하여 문장이 속하는 클래스(class)를 분류(classification)하는 방법에 있어서,

적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크(neural network)를 통해 제1 특징 벡터(feature vector)를 학습(training)하는 단계;

제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하는 단계;

상기 제1 특징 벡터와 상기 제2 특징 벡터의 내적(dot product) 및 상기 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 계산하는 단계; 및

상기 계산된 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하는 단계;

를 포함하는, 방법.

청구항 2

제1 항에 있어서,

사용자로부터 발화를 입력 받는 단계;

입력 받은 상기 발화를 문장으로 인식하는 단계; 및

인식된 상기 문장 내에 포함된 적어도 하나의 단어를 추출하고, 상기 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환하는 단계; 를 더 포함하고,

상기 제1 특징 벡터를 학습하는 단계는,

상기 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하는 단계; 및

상기 문장 벡터를 상기 제1 뉴럴 네트워크에 입력 데이터로 입력시켜 상기 제1 특징 벡터를 학습하는, 방법.

청구항 3

제1 항에 있어서,

복수의 문장과 상기 복수의 문장 각각이 속하는 복수의 클래스가 데이터베이스에 저장되어 있고,

상기 제2 문장 및 상기 제2 클래스는 상기 데이터베이스 상에서 무작위(random)로 추출되는 것인, 방법,

청구항 4

삭제

청구항 5

제1 항에 있어서,

상기 수식은, 상기 제1 클래스와 상기 제2 클래스가 동일한 경우 1을 출력하고, 상기 제1 클래스와 상기 제2 클래스가 동일하지 않은 경우 0을 출력하는, 방법.

청구항 6

제1 항에 있어서,

제1 특징 벡터를 학습하는 단계는,

상기 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환하는 단계;

변환된 매트릭스를 컨볼루션 뉴럴 네트워크에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성하는 단계; 및

상기 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 상기 제1 특징 벡터를 추출하는 단계;

를 포함하는, 방법.

청구항 7

제6 항에 있어서,

상기 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터값으로 변환하는 단계; 및

상기 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 클래스로 분류되는 확률 분포를 나타내는 제1 분류 예측값을 획득하는 단계;

를 더 포함하는, 방법.

청구항 8

제7 항에 있어서,

상기 제1 분류 예측값과 상기 제1 클래스 간의 차이값인 제1 분류 손실값을 획득하는 단계;

상기 제2 뉴럴 네트워크를 통해 제2 문장이 제2 클래스로 분류되는 확률 분포를 나타내는 제2 분류 예측값을 획득하고, 상기 제2 분류 예측값과 상기 제2 클래스 간의 차이값인 제2 분류 손실값을 획득하는 단계; 및

상기 제1 분류 손실값, 상기 제2 분류 손실값 및 상기 대조 손실값을 합산하여 최종 손실값을 계산하고, 계산된 상기 최종 손실값을 기초로 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절하는 단계;

를 더 포함하는, 방법.

청구항 9

제1 항에 있어서,

상기 제1 뉴럴 네트워크를 통한 학습 단계 및 상기 제2 뉴럴 네트워크를 통한 학습 단계는 동시에 수행되는, 방법.

청구항 10

딥 뉴럴 네트워크(Deep Neural Network)를 이용하여 문장이 속하는 클래스(class)를 분류하는 전자 장치

(electronic device)에 있어서,

뉴럴 네트워크(neural network)를 이용하여 학습(training)을 수행하는 프로세서를 포함하고,

상기 프로세서는, 적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크를 통해 제1 특징 벡터(feature vector)를 학습하고, 제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하고, 상기 제1 특징 벡터와 상기 제2 특징 벡터의 내적(dot product), 및 상기 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 계산하고, 상기 계산된 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하여 수행하는, 전자 장치.

청구항 11

제10 항에 있어서,

사용자로부터 발화를 입력 받는 발화 입력부; 를 더 포함하고,

상기 프로세서는 입력받은 상기 발화를 문장으로 인식하고, 인식된 상기 문장 내에 포함된 적어도 하나의 단어를 추출하고, 상기 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환하는, 전자 장치.

청구항 12

제11 항에 있어서,

상기 프로세서는 상기 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하고, 상기 문장 벡터를 상기 제1 뉴럴 네트워크에 입력 데이터로 입력시켜 상기 제1 특징 벡터를 학습하는, 전자 장치.

청구항 13

제10 항에 있어서,

복수의 문장과 상기 복수의 문장 각각이 속하는 복수의 클래스를 저장하는 데이터베이스;

를 더 포함하고,

상기 프로세서는 상기 데이터베이스로부터 상기 제2 문장 및 상기 제2 클래스는 상기 데이터베이스 상에서 무작위(random)로 추출하여 상기 제2 뉴럴 네트워크에 입력 데이터로 입력하는, 전자 장치.

청구항 14

삭제

청구항 15

제10 항에 있어서,

상기 수식은, 상기 제1 클래스와 상기 제2 클래스가 동일한 경우 1을 출력하고, 상기 제1 클래스와 상기 제2 클래스가 동일하지 않은 경우 0을 출력하는, 전자 장치.

청구항 16

제10 항에 있어서,

상기 프로세서는, 상기 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환하고, 변환된 매트릭스를 컨볼루션 뉴럴 네트워크에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성하고, 상기 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 상기 제1 특징 벡터를 추출하는, 전자 장치.

청구항 17

제16 항에 있어서,

상기 프로세서는, 상기 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터값으로 변환하고, 상기 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 클래스로 분류되는 확률 분포를 나타내는 제1 분류 예측값을 획득하는, 전자 장치.

청구항 18

제17 항에 있어서,

상기 프로세서는, 상기 제1 분류 예측값과 상기 제1 클래스 간의 차이값인 제1 분류 손실값을 획득하고, 상기 제2 뉴럴 네트워크를 통해 제2 문장이 제2 클래스로 분류되는 확률 분포를 나타내는 제2 분류 예측값을 획득하고, 상기 제2 분류 예측값과 상기 제2 클래스 간의 차이값인 제2 분류 손실값을 획득하고, 상기 제1 분류 손실값, 상기 제2 분류 손실값 및 상기 대조 손실값을 합산하여 최종 손실값을 계산하고, 계산된 상기 최종 손실값을 기초로 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절하는, 전자 장치.

청구항 19

제10 항에 있어서,

상기 프로세서는, 상기 제1 뉴럴 네트워크를 통한 학습 및 상기 제2 뉴럴 네트워크를 통한 학습을 동시에 수행하는, 전자 장치.

청구항 20

컴퓨터로 읽을 수 있는 매체에 저장된 컴퓨터 프로그램에 있어서,

상기 매체는,

적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크(neural network)를 통해 제1 특징 벡터(feature vector)를 학습(training)하는 단계;

제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하는 단계;

상기 제1 특징 벡터와 상기 제2 특징 벡터의 내적(dot product) 및 상기 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 계산하는 단계; 및

상기 계산된 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하는 단계;

를 수행하는 명령어들(instructions)을 포함하는, 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 개시는 딥 뉴럴 네트워크(Deep Neural Network)를 이용하여 질문 문장을 구조 분석하여, 문장이 속하는 클래스(class)를 분류하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 인공 지능(Artificial Intelligence, AI) 시스템은 인간 수준의 지능을 구현하는 컴퓨터 시스템이며, 기존 Rule 기반 스마트 시스템과 달리 기계가 스스로 학습하고 판단하며 똑똑해지는 시스템이다. 인공지능 시스템은 사용할수록 인식이 향상되고 사용자 취향을 보다 정확하게 이해할 수 있게 되어, 기존 Rule 기반 스마트 시스템은 점차 딥 뉴럴 네트워크 기반 인공지능 시스템으로 대체되고 있다.

[0003] 인공지능 기술은 기계학습 및 기계학습을 활용한 요소 기술들로 구성된다.

[0004] 기계학습은 입력 데이터들의 특징을 스스로 분류/학습하는 알고리즘 기술이며, 요소기술은 뉴럴 네트워크 등의 기계학습 알고리즘을 활용하여 인간 두뇌의 인지, 판단 등의 기능을 모사하는 기술로서, 언어적 이해, 시각적 이해, 추론/예측, 지식 표현, 동작 제어 등의 기술 분야로 구성된다.

[0005] 인공지능 기술은 인간의 언어/문자를 인식하고 응용/처리할 수 있고, 자연어 처리, 기계 번역, 대화 시스템, 질의 응답, 음성 인식/합성 등에도 활용된다. 인공지능 기술을 이용한 질의 응답 시스템에서는 사용자의 질문 문장을 구조 분석하여, 답의 종류(answer type), 의도(intent), 주어/동사 분석 등을 실시하고, 이를 가지고 데이터베이스에서 관련 답변을 찾는다. 사용자의 명령을 수행하는 질의 응답 시스템의 경우에는 사용자 입력 발화를 분류하여 의도(intent)를 분석하고, 독립 개체(entity)를 발견하여 명령을 처리한다.

[0006] 최근에는 인공지능 기술을 활용하여 사용자의 문제점을 분석하고 적절한 답변을 제공하는 고객 지원용 챗봇(Customer Care Chatbot)이 활용되고 있다. 고객 지원용 챗봇 등에서는 사용자의 발화를 분석하여 사용자가 답변을 받고자 하는 카테고리(category)를 분석하는 것이 중요하다. 기 저장되어 있는 질문의 양이 많지 않은 경우 사용자의 발화가 사용자의 의도와는 달리 다른 카테고리로 잘못 분류될 수도 있다. 이 경우, 사용자는 원하는 답변을 제공받지 못하는 문제점이 발생된다.

발명의 내용

해결하려는 과제

[0007] 본 개시는 제1 뉴럴 네트워크 모델을 이용하여 제1 문장이 속하는 클래스를 분류(classifying)하는데 있어서 별도의 제2 뉴럴 네트워크 모델을 추가로 이용하여 제1 문장의 분류 정확도를 높이는 방법 및 장치를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0008] 상술한 기술적 과제를 해결하기 위하여, 본 개시의 일 실시예는 적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크(neural network)를 통해 제1 특징 벡터(feature vector)를 학습(training)하는 단계; 제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하는 단계; 상기 제1 특징 벡터, 상기 제2 특징 벡터, 및 상기 제1 클래스와 제2 클래스의 동일성 여부에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 획득하는 단계; 및 상기 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하는 단계를 포함하는, 딥 뉴럴 네트워크(Deep neural network)를 이용하여 문장이 속하는 클래스(class)를 분류(classification)하는 방법을 제공한다.

[0009] 예를 들어, 상기 방법은 사용자로부터 발화를 입력 받는 단계; 입력 받은 상기 발화를 문장으로 인식하는 단계; 및 인식된 상기 문장 내에 포함된 적어도 하나의 단어를 추출하고, 상기 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환하는 단계를 더 포함하고, 상기 제1 특징 벡터를 학습하는 단계는 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하는 단계; 및 상기 문장 벡터를 상기 제1 뉴럴 네트

워크에 입력 데이터로 입력시켜 상기 제1 특징 벡터를 학습할 수 있다.

- [0010] 예를 들어, 복수의 문장과 상기 복수의 문장 각각이 속하는 복수의 클래스가 데이터베이스에 저장되어 있고, 상기 제2 문장 및 상기 제2 클래스는 상기 데이터베이스 상에서 무작위(random)로 추출될 수 있다.
- [0011] 예를 들어, 상기 대조 손실값을 획득하는 단계는 제1 특징 벡터와 제2 특징 벡터의 내적(dot product) 및 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식을 통해 상기 대조 손실값을 계산할 수 있다.
- [0012] 예를 들어, 상기 수식은 상기 제1 클래스와 상기 제2 클래스가 동일한 경우 1을 출력하고, 상기 제1 클래스와 상기 제2 클래스가 동일하지 않은 경우 0을 출력할 수 있다.
- [0013] 예를 들어, 제1 특징 벡터를 학습하는 단계는 상기 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환하는 단계; 변환된 매트릭스를 컨볼루션 뉴럴 네트워크에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성하는 단계; 및 상기 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 상기 제1 특징 벡터를 추출하는 단계를 포함할 수 있다.
- [0014] 예를 들어, 상기 방법은 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터값으로 변환하는 단계; 및 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 클래스로 분류되는 확률 분포를 나타내는 제1 분류 예측값을 획득하는 단계를 더 포함할 수 있다.
- [0015] 예를 들어, 상기 방법은 상기 제1 분류 예측값과 상기 제1 클래스 간의 차이값인 제1 분류 손실값을 획득하는 단계; 상기 제2 뉴럴 네트워크를 통해 제2 문장이 제2 클래스로 분류되는 확률 분포를 나타내는 제2 분류 예측값을 획득하고, 상기 제2 분류 예측값과 상기 제2 클래스 간의 차이값인 제2 분류 손실값을 획득하는 단계; 및 상기 제1 분류 손실값, 상기 제2 분류 손실값 및 상기 대조 손실값을 합산하여 최종 손실값을 계산하고, 계산된 상기 최종 손실값을 기초로 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절하는 단계를 더 포함할 수 있다.
- [0016] 예를 들어, 상기 제1 뉴럴 네트워크를 통한 학습 단계 및 상기 제2 뉴럴 네트워크를 통한 학습 단계는 동시에 수행될 수 있다.
- [0017] 상술한 기술적 과제를 해결하기 위하여, 본 개시의 일 실시예는 딥 뉴럴 네트워크(deep neural network)를 이용하여 문장이 속하는 클래스(class)를 분류하는 전자 장치(electronic device)를 제공할 수 있다. 상기 전자 장치는 뉴럴 네트워크(neural network)를 이용하여 학습(training)을 수행하는 프로세서를 포함하고, 상기 프로세서는 적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크를 통해 제1 특징 벡터(feature vector)를 학습하고, 제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하고, 상기 제1 특징 벡터, 상기 제2 특징 벡터, 및 상기 제1 클래스와 제2 클래스의 동일성 여부에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 획득하고, 상기 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하여 수행할 수 있다.
- [0018] 예를 들어, 상기 전자 장치는 사용자로부터 발화를 입력 받는 발화 입력부를 더 포함하고, 상기 프로세서는 입력받은 상기 발화를 문장으로 인식하고, 인식된 상기 문장 내에 포함된 적어도 하나의 단어를 추출하고, 상기 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환할 수 있다.
- [0019] 예를 들어, 상기 프로세서는 상기 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하고, 상기 문장 벡터를 상기 제1 뉴럴 네트워크에 입력 데이터로 입력시켜 상기 제1 특징 벡터를 학습할 수 있다.
- [0020] 예를 들어, 상기 전자 장치는 복수의 문장과 상기 복수의 문장 각각이 속하는 복수의 클래스를 저장하는 데이터베이스를 더 포함하고, 상기 프로세서는 상기 데이터베이스로부터 상기 제2 문장 및 상기 제2 클래스는 상기 데이터베이스 상에서 무작위(random)로 추출하여 상기 제2 뉴럴 네트워크에 입력 데이터로 입력할 수 있다.
- [0021] 예를 들어, 상기 프로세서는 상기 제1 특징 벡터와 상기 제2 특징 벡터의 내적(dot product) 및 상기 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식을 통해 상기 대조 손실값을 계산할 수 있다.
- [0022] 예를 들어, 상기 수식은 상기 제1 클래스와 상기 제2 클래스가 동일한 경우 1을 출력하고, 상기 제1 클래스와 상기 제2 클래스가 동일하지 않은 경우 0을 출력할 수 있다.
- [0023] 예를 들어, 상기 프로세서는 상기 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환하고,

변환된 매트릭스를 컨볼루션 뉴럴 네트워크에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성하고, 상기 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 상기 제1 특징 벡터를 추출할 수 있다.

[0024] 예를 들어, 상기 프로세서는 상기 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터값으로 변환하고, 상기 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 클래스로 분류되는 확률 분포를 나타내는 제1 분류 예측값을 획득할 수 있다.

[0025] 예를 들어, 상기 프로세서는 상기 제1 분류 예측값과 상기 제1 클래스 간의 차이값인 제1 분류 손실값을 획득하고, 상기 제2 뉴럴 네트워크를 통해 제2 문장이 제2 클래스로 분류되는 확률 분포를 나타내는 제2 분류 예측값을 획득하고, 상기 제2 분류 예측값과 상기 제2 클래스 간의 차이값인 제2 분류 손실값을 획득하고, 상기 제1 분류 손실값, 상기 제2 분류 손실값 및 상기 대조 손실값을 합산하여 최종 손실값을 계산하고, 계산된 상기 최종 손실값을 기초로 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절할 수 있다.

[0026] 예를 들어, 상기 프로세서는 상기 제1 뉴럴 네트워크를 통한 학습 및 상기 제2 뉴럴 네트워크를 통한 학습을 동시에 수행할 수 있다.

[0027] 상술한 기술적 과제를 해결하기 위하여, 본 개시의 일 실시예는 컴퓨터로 읽을 수 있는 저장 매체를 포함하는 컴퓨터 프로그램 제품을 제공하고, 상기 저장 매체는 적어도 하나의 단어를 포함하는 제1 문장과 상기 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크(neural network)를 통해 제1 특징 벡터(feature vector)를 학습(training)하는 단계; 제2 문장과 상기 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습하는 단계; 상기 제1 특징 벡터, 상기 제2 특징 벡터, 및 상기 제1 클래스와 제2 클래스의 동일성 여부에 기초하여 상기 제1 문장 및 상기 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 획득하는 단계; 및 상기 대조 손실값이 최대가 되도록 상기 제1 뉴럴 네트워크 및 상기 제2 뉴럴 네트워크를 이용한 학습을 반복하는 단계를 수행하는 명령어들을 포함할 수 있다.

도면의 간단한 설명

[0028] 본 개시는, 다음의 자세한 설명과 그에 수반되는 도면들의 결합으로 쉽게 이해될 수 있으며, 참조 번호(reference numerals)들은 구조적 구성요소(structural elements)를 의미한다.

도 1은 본 개시의 일 실시예에 따른 뉴럴 네트워크 모델에 문장 벡터와 클래스를 입력하여 학습(training)함으로써, 문장이 속하는 클래스의 분류 예측값을 획득하는 실시예를 설명하기 위한 개념도이다.

도 2는 본 개시의 일 실시예에 따른 전자 장치의 구성 요소를 도시한 블록도이다.

도 3은 본 개시의 일 실시예에 따른 전자 장치가 문장이 속하는 클래스를 분류하는 방법을 도시한 흐름도이다.

도 4는 본 개시의 일 실시예에 따른 전자 장치가 컨볼루션 뉴럴 네트워크를 이용하여 문장이 속하는 클래스를 분류하는 방법을 설명하기 위한 도면이다.

도 5는 본 개시의 일 실시예에 따른 전자 장치가 제1 문장이 속하는 클래스로 분류되는 확률값인 분류 예측값을 획득하는 방법을 도시한 흐름도이다.

도 6은 본 개시의 일 실시예에 따른 전자 장치가 뉴럴 네트워크 모델을 통해 획득된 손실값을 기초로 하여 뉴럴 네트워크 모델에 적용된 가중치(weight)를 조절하는 학습 방법을 도시한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0029] 본 명세서의 실시예들에서 사용되는 용어는 본 개시의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 실시예의 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.

[0030] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에 기재된 "...부", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또

는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.

- [0031] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다.
- [0032] 이하에서는 도면을 참조하여 본 개시의 실시예들을 상세하게 설명한다.
- [0033] 도 1은 본 개시의 일 실시예에 따른 뉴럴 네트워크 모델(100, 110)에 문장 벡터(S_i, S_j)와 클래스(y_1, y_2)를 입력하여 학습(training)함으로써, 문장이 속하는 클래스의 분류 예측값(y_1', y_2')을 획득하는 실시예를 설명하기 위한 개념도이다.
- [0034] 딥 뉴럴 네트워크(Deep Neural Network) 등을 포함하는 인공지능(AI) 알고리즘은 인공 뉴럴 네트워크(Artificial Neural Network, ANN)에 입력 데이터를 입력시키고, 컨볼루션 등의 연산을 통해 출력 데이터를 학습하는 것을 특징으로 한다. 인공 뉴럴 네트워크는 생물학적 뇌를 모델링한 컴퓨터 과학적 아키텍처(Computational Architecture)를 의미할 수 있다. 인공 뉴럴 네트워크 내에서, 뇌의 뉴런들에 해당되는 노드들은 서로 연결되어 있고, 입력 데이터를 처리하기 위하여 집합적으로 동작한다. 피드-포워드(feed-forward) 뉴럴 네트워크에서, 뉴럴 네트워크의 뉴런들은 다른 뉴런들과의 연결들(links)을 갖는다. 이와 같은 연결들은 뉴럴 네트워크를 통해, 한 방향으로, 예를 들어 순방향(forward direction)으로 확장될 수 있다.
- [0035] 도 1을 참조하면, 제1 뉴럴 네트워크 모델(100)에 제1 문장 벡터(S_i) 및 제1 클래스(y_1)가 입력 데이터로서 입력되고, 제1 뉴럴 네트워크 모델(100)을 통한 학습(training)을 통해 제1 분류 예측값(y_1')이 출력될 수 있다. 또한, 제2 뉴럴 네트워크 모델(110)에 제2 문장 벡터(S_j) 및 제2 클래스(y_2)가 입력 데이터로서 입력되고, 제2 뉴럴 네트워크 모델(110)을 통한 학습을 통해 제2 분류 예측값(y_2')이 출력될 수 있다.
- [0036] 도 1에 도시된 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)은 컨볼루션 뉴럴 네트워크(Convolution Neural Network, CNN)으로 구현될 수 있지만, 이에 한정되는 것은 아니다. 일 실시예에서, 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)은 회귀 뉴럴 네트워크(Recurrent Neural Network, RNN), 딥 빌리프 네트워크(Deep Belief Network, DBN), 제한된 볼츠만 기계(Restricted Boltzman Machine, RBM) 방식 등의 인공 뉴럴 네트워크 모델로 구현되거나, 또는 서포트 벡터 머신(Support Vector Machine, SVM)과 같은 머신 러닝 모델로도 구현될 수 있다.
- [0037] 제1 문장 벡터(S_i) 및 제2 문장 벡터(S_j)는 자연어 처리 기술을 통해 사용자가 입력한 문장 또는 발화에 포함된 적어도 하나의 단어를 파싱(parsing)하여 추출되고, 추출된 단어를 벡터로 변환하여 생성될 수 있다. 일 실시예에서, 제1 문장 벡터(S_i) 및 제2 문장 벡터(S_j)는 word2vec, GloVe, onehot 과 같은 단어를 벡터로 임베딩(embedding)하는 기계 학습 모델을 통해 생성될 수 있으나, 이에 한정되는 것은 아니다. 제1 문장 벡터(S_i) 및 제2 문장 벡터(S_j)는 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열함으로써 생성될 수 있다.
- [0038] 제1 클래스(y_1) 및 제2 클래스(y_2)은 각각 제1 문장 벡터(S_i) 및 제2 문장 벡터(S_j)가 속하는 클래스를 정의하는 벡터값일 수 있다. 여기서, 클래스(class)란 계층을 의미하는 것이 아니고, 문장이 속하는 카테고리 분류, 예컨대 정치, 사회, 경제, 문화, 연예, IT 등을 의미할 수 있다.
- [0039] 제1 분류 예측값(y_1')은 제1 뉴럴 네트워크 모델(100)을 통한 학습으로 인하여 출력된 결과 데이터로서, 제1 문장 벡터(S_i)가 제1 뉴럴 네트워크 모델(100)을 통해 학습되어 제1 클래스(y_1)로 분류될 수 있는 확률값을 의미할 수 있다. 예를 들어, '정치' 카테고리의 해당되는 제1 클래스(y_1) 값이 (1, 0, 0)인데, 제1 분류 예측값(y_1')이 (0.9, 0.05, 0.05)인 경우 제1 문장 벡터(S_i)에 대응되는 제1 문장은 '정치'에 관련된 글로 분류될 수 있다. 제2 분류 예측값(y_2')은 제2 뉴럴 네트워크 모델(110)을 통해 출력된 결과값으로서, 제2 문장 벡터(S_j)가 제2 뉴럴 네트워크 모델(110)을 통해 학습되어 제2 클래스(y_2)로 분류될 수 있는 확률값을 의미할 수 있다. 제1 분류 예측값(y_1')과 제1 클래스(y_1)의 차이값을 계산하여 제1 분류 손실값(Classification Loss)을 획득할 수 있다. 마찬가지로, 제2 분류 예측값(y_2')과 제2 클래스(y_2)의 차이값을 계산하여 제2 분류 손실값을 획득할 수 있다.

- [0040] 일 실시예에서, 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)은 컨볼루션 뉴럴 네트워크(CNN)으로 구성될 수 있다. 제1 문장 벡터(S_1) 및 제2 문장 벡터(S_2)는 각각 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)을 통해 서로 다른 폭(width)을 갖는 복수의 필터를 통해 컨볼루션 연산되고, 각각 제1 특징 벡터 및 제2 특징 벡터가 학습될 수 있다. 제1 뉴럴 네트워크 모델(100)을 통해 학습된 제1 특징 벡터, 제2 뉴럴 네트워크 모델(110)을 통해 학습된 제2 특징 벡터, 및 제1 클래스(y_1)와 제2 클래스(y_2)의 동일성 여부에 기초하여 대조 손실값(contrastive loss)(L_1)이 획득될 수 있다. 일 실시예에서, 대조 손실값(L_1)은 제1 특징 벡터와 제2 특징 벡터의 내적(dot product) 및 제1 클래스와 제2 클래스의 동일 여부를 숫자로 나타내는 수식을 통해 계산될 수 있다. 대조 손실값(L_1)에 대해서는 도 4의 설명 부분에서 상세하게 설명하기로 한다.
- [0041] 일 실시예에서, 대조 손실값(L_1)은 -1 이상 1 이하의 범위의 값을 가질 수 있다. 본 개시의 일 실시예에서, 대조 손실값(L_1)이 최대가 되는 방향으로 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)을 통한 학습이 반복될 수 있다. 여기서, 학습의 반복은 제1 뉴럴 네트워크 모델(100) 및 제2 뉴럴 네트워크 모델(110)에 적용되는 가중치(weight)를 조절하는 것을 의미할 수 있다.
- [0042] 일반적인 딥 뉴럴 네트워크(Deep Neural Network) 기반의 문장 분류(text classification)에서는 학습될 문장의 레이블(Label)을 기초로 하나의 뉴럴 네트워크 모델을 통해 학습하여 하나의 손실 함수(Loss function)를 만들어 분류를 수행한다. 따라서, 분류될 발화가 분류 모델의 어떤 클래스에도 해당되지 않는 경우 많은 오분류가 일어날 수 있다. 또한, 라벨을 위주로 클래스를 분류하는바, 표현이 유사하면 다른 클래스라도 오분류되는 경우가 다수 존재한다. 예를 들어, 사용자 입력 발화가 "'XXX'에게 '카카오톡' 보내줘"인 경우, "'XXX'에게 '문자' 보내줘"로 분류되는 경우가 발생될 수 있다.
- [0043] 본 개시의 일 실시예는, 제1 문장이 속하는 제1 클래스를 분류하는데 있어서 제1 뉴럴 네트워크 모델(100)을 사용하여 학습할 뿐만 아니라, 제2 클래스에 속하는 제2 문장을 제2 뉴럴 네트워크 모델(110)을 통해 학습하고, 제1 뉴럴 네트워크 모델(100)에서 학습된 제1 특징 벡터, 제2 뉴럴 네트워크 모델(110)에서 학습된 제2 특징 벡터, 및 제1 클래스(y_1)와 제2 클래스(y_2)의 동일성에 기초하여 대조 손실값(L_1)을 계산함으로써, 제1 문장과 제2 문장 사이의 표현 상의 유사 정도까지 구분할 수 있는 방법 및 장치를 제공한다. 따라서, 본 개시의 일 실시예에 따른 방법 및 장치는 문장 또는 발화의 레이블 뿐만 아니라 의미적인 유사도를 함께 사용하여, 유사하지만 다른 클래스에 속하는 문장에 대한 분류 정확도를 향상시킬 수 있다.
- [0044] 도 2는 본 개시의 일 실시예에 따른 전자 장치(200)의 구성 요소를 도시한 블록도이다. 전자 장치(200)는 뉴럴 네트워크 모델을 이용하여 문장이 속하는 클래스를 분류(classification)하는 학습(training)을 수행하는 장치일 수 있다. 전자 장치(200)는 컴퓨터 장치로 구현되는 고정형 단말이거나 이동형 단말일 수 있다. 전자 장치(200)는 예를 들어, 스마트폰(smart phone), 휴대폰, 내비게이션, 컴퓨터, 노트북, 디지털방송용 단말, PDA(Personal Digital Assistants), PMP(Portable Multimedia Player), 및 태블릿 PC 중 적어도 하나일 수 있으나, 이에 한정되지 않는다. 전자 장치(200)는 무선 또는 유선 통신 방식을 이용하여 네트워크를 통해 다른 전자 장치 및/또는 서버와 통신할 수 있다.
- [0045] 도 2를 참조하면, 전자 장치(200)는 프로세서(210), 메모리(220), 및 발화 입력부(230)를 포함할 수 있다.
- [0046] 프로세서(210)는 컨볼루션 연산과 같은 산술, 로직 및 입출력 연산을 수행함으로써, 컴퓨터 프로그램의 명령을 처리하도록 구성될 수 있다. 명령(instruction)은 메모리(220)에 의해 프로세서(210)에 제공될 수 있다. 일 실시예에서, 프로세서(210)는 메모리(220)와 같은 기록 장치에 저장된 프로그램 코드에 따라 수신되는 명령을 실행하도록 구성될 수 있다. 프로세서(210)는 예를 들어, 중앙 처리 장치(Central Processing Unit), 마이크로 프로세서(microprocessor) 및 그래픽 프로세서(Graphic Processing Unit) 중 적어도 하나로 구성될 수 있으나, 이에 제한되는 것은 아니다. 일 실시예에서, 전자 장치(200)가 스마트폰, 태블릿 PC 등과 같은 모바일 디바이스인 경우, 프로세서(210)는 애플리케이션을 실행시키는 애플리케이션 프로세서(Application Processor, AP)일 수 있다.
- [0047] 프로세서(210)는 뉴럴 네트워크 모델(neural network model)과 같은 딥 뉴럴 네트워크 기반의 범용 인공 지능 알고리즘을 통해 학습(training)을 수행할 수 있다.
- [0048] 프로세서(210)는 사용자의 발화, 질문 문장 등에서 단어를 추출하고, 추출된 단어를 단어 벡터로 변환하여 문장 벡터를 생성하는 등의 자연어 처리(Natural Language Processing, NLP)를 수행할 수도 있다. 프로세서(210)는

문장의 객체화를 통해 단어 객체를 파싱하고 정지어 처리(관사 등 필터링) 및 토큰 생성(시제, 복수형통일 등) 처리를 거친 이후, 출현 빈도에 기반하여 연관도 높은 키워드를 추출하고 이를 독립 개체(entity)로 관리할 수 있다.

- [0049] 일 실시예에서, 프로세서(210)는 적어도 하나의 단어를 포함하는 제1 문장과 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크를 통해 제1 특징 벡터(feature vector)를 학습하고, 제2 문장과 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습할 수 있다. 여기서, 제1 문장은 사용자가 입력한 문장 또는 발화이고, 제2 문장은 서버 또는 데이터베이스에 저장되어 있는 복수의 문장 중 무작위(random)로 추출된 문장일 수 있다.
- [0050] 프로세서(210)는 제1 특징 벡터, 제2 특징 벡터, 및 제1 클래스와 제2 클래스의 동일성 여부에 기초하여 제1 문장 및 제2 문장의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)을 획득할 수 있다. 일 실시예에서, 프로세서(210)는 제1 특징 벡터와 제2 특징 벡터의 내적(dot product) 및 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식을 통해 대조 손실값을 계산할 수 있다. 대조 손실값을 계산하는 방법은 도 4에 관한 설명 부분에서 상세하게 설명하기로 한다.
- [0051] 일 실시예에서, 대조 손실값은 -1 이상 1 이하의 범위 내의 값을 갖는데, 프로세서(210)는 획득한 대조 손실값이 최대가 되도록 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크를 이용한 학습을 반복하여 수행할 수 있다.
- [0052] 일 실시예에서, 프로세서(210)는 제1 뉴럴 네트워크를 통한 학습 및 제2 뉴럴 네트워크를 통한 학습을 동시에 수행할 수 있다.
- [0053] 전자 장치(200)는 사용자로부터 발화 또는 문장을 입력받는 발화 입력부(230)를 더 포함할 수 있다. 발화 입력부(230)는 사용자의 음성을 인식할 수 있는 음성 인식 모듈을 포함할 수 있으나, 이에 제한되는 것은 아니다. 발화 입력부(230)는 예를 들어, 키패드(key pad), 마우스, 터치 패드, 터치스크린, 조그 스위치 등 사용자의 문장을 입력받을 수 있는 하드웨어 모듈로 구성될 수 있다. 프로세서(210)는 발화 입력부(230)를 통해 입력받은 발화를 문장으로 인식하고, 인식된 문장 내에 포함된 적어도 하나의 단어를 파싱(parsing)하여 추출하고, 추출된 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환할 수 있다. 일 실시예에서, 프로세서(210)는 word2vec, GloVe, onehot 등과 같은 기계 학습 모델을 이용하여 단어를 벡터로 임베딩(embedding)할 수 있으나, 이에 한정되는 것은 아니다. 프로세서(210)는 상기 기계 학습 모델을 이용하여 단어 표현을 벡터 공간 상에 나타낼 수 있는 벡터 값으로 변환할 수 있다.
- [0054] 프로세서(210)는 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하고, 문장 벡터를 제1 뉴럴 네트워크에 입력 데이터로 입력시켜 제1 특징 벡터를 학습할 수 있다. 일 실시예에서, 프로세서(210)는 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환하고, 변환된 매트릭스를 컨볼루션 뉴럴 네트워크에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성하고, 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 제1 특징 벡터를 추출할 수 있다.
- [0055] 일 실시예에서, 프로세서(210)는 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터값으로 변환하고, 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 클래스로 분류되는 확률 분포를 나타내는 제1 분류 예측값을 획득할 수 있다. 마찬가지로, 프로세서(210)는 제2 특징 벡터를 학습하여 추출하고, 제2 특징 벡터를 완전 연결 레이어에 입력시켜 1차원 벡터값으로 변환하고, 1차원 벡터값을 소프트 맥스 분류기에 입력시켜 제2 클래스로 분류되는 확률 분포를 나타내는 제2 분류 예측값을 획득할 수 있다. 이에 대한 상세한 설명은 도 4의 설명 부분에서 후술하기로 한다.
- [0056] 프로세서(210)는 제1 분류 예측값과 제1 클래스 간의 차이값인 제1 분류 손실값을 획득하고, 제2 분류 예측값과 제2 클래스 간의 차이값인 제2 분류 손실값을 획득할 수 있다. 프로세서(210)는 제1 분류 손실값, 제2 분류 손실값 및 대조 손실값을 합산하여 최종 손실값을 계산하고, 계산된 최종 손실값을 기초로 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절하는 학습을 반복할 수 있다.
- [0057] 메모리(220)는 컴퓨터에서 판독 가능한 기록 매체로서, RAM(random access memory), ROM(read only memory) 및 디스크 드라이브와 같은 비소멸성 대용량 기록장치(permanent mass storage device)를 포함할 수 있다. 또한, 메모리(220)에는 운영 체제(Operating System, OS)나 적어도 하나의 컴퓨터 프로그램 코드(예를 들어, 프로세서(210)에서 수행되는 뉴럴 네트워크를 통한 학습 프로그램 등을 위한 코드)가 저장될 수 있다. 이와 같은 컴퓨터 프로그램 코드는 메모리(220)에 저장될 수도 있지만, 별도의 컴퓨터에서 판독 가능한 기록 매체 또는 컴퓨터 프로그램 제품(Computer Program Product)로부터 로딩될 수 있다. 컴퓨터에서 판독 가능한 기록 매체는 플로피 드

라이브, 디스크, 테이프, DVD/CD-ROM 드라이브, 메모리카드 등의 컴퓨터에서 판독 가능한 기록 매체를 포함할 수 있다. 일 실시예에서, 컴퓨터 프로그램 코드는 서버에서부터 네트워크를 통해 제공하는 파일들에 의해 전자 장치(200)에 설치되고, 메모리(220)에서 로딩될 수 있다.

[0058] 도면에는 도시되지 않았지만, 전자 장치(200)는 데이터베이스를 포함할 수 있다. 데이터베이스는 복수의 문장과 복수의 문장 각각이 속하는 복수의 클래스를 저장할 수 있다. 일 실시예에서, 데이터베이스는 전자 장치(200) 내에 구성 요소로 포함될 수 있지만, 이에 한정되는 것은 아니다. 데이터베이스는 전자 장치(200)의 외부에 배치되는 서버 형태로 구성될 수도 있다. 일 실시예에서, 프로세서(210)는 데이터베이스로부터 제2 문장 및 제2 클래스를 무작위(random)로 추출하여 제2 뉴럴 네트워크에 입력 데이터로 입력시켜 학습할 수 있다.

[0059] 도 3은 본 개시의 일 실시예에 따른 전자 장치가 문장이 속하는 클래스를 분류하는 방법을 도시한 흐름도이다.

[0060] 단계 S310에서, 전자 장치는 제1 문장과 제1 문장이 속하는 제1 클래스를 입력 데이터로 하여 제1 뉴럴 네트워크(neural network)를 통해 제1 특징 벡터(feature vector)를 학습(training)한다. 일 실시예에서, 전자 장치는 사용자로부터 발화 또는 질문을 입력받고, 입력 받은 발화 또는 질문을 문장으로 인식할 수 있다. 전자 장치는 자연어 처리(NLP) 기술을 사용하여 인식된 문장 내에 포함된 적어도 하나의 단어를 파싱(parsing)하고, 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환할 수 있다. 일 실시예에서, 전자 장치는 word2vec, GloVe, onehot 등과 같은 기계 학습 모델을 사용하여 적어도 하나의 단어를 적어도 하나의 단어 벡터로 임베딩(embedding)할 수 있으나, 이에 한정되는 것은 아니다. 전자 장치는 상기 기계 학습 모델을 이용하여 단어 표현을 벡터 공간 상에 나타낼 수 있는 벡터 값으로 변환할 수 있다.

[0061] 일 실시예에서, 전자 장치는 임베딩된 적어도 하나의 단어 벡터를 매트릭스(matrix) 형태로 배열하여 문장 벡터를 생성하고, 생성된 문장 벡터를 제1 뉴럴 네트워크에 입력 데이터로 입력시켜 제1 클래스로 분류되는 확률 분포를 학습할 수 있다.

[0062] 단계 S320에서, 전자 장치는 제2 문장과 제2 문장이 속하는 제2 클래스를 입력 데이터로 하여 제2 뉴럴 네트워크를 통해 제2 특징 벡터를 학습한다. 일 실시예에서, 제2 문장과 제2 문장이 속하는 제2 클래스는 데이터베이스 형태로 저장되어 있을 수 있다. 전자 장치는 데이터베이스로부터 제2 문장 및 제2 클래스를 무작위(random)로 추출하여 제2 뉴럴 네트워크에 입력 데이터로 입력시켜 학습할 수 있다. 도 3에는 단계 S320가 단계 S310 이후에 수행되는 것으로 도시되었지만, 이에 한정되는 것은 아니다. 전자 장치는 제1 특징 벡터를 학습하는 단계(S310)과 제2 특징 벡터를 학습하는 단계(S320)를 동시에 수행할 수도 있다.

[0063] 단계 S330에서, 전자 장치는 제1 특징 벡터, 제2 특징 벡터, 및 제1 클래스와 제2 클래스의 동일성 여부에 기초하여 대조 손실값(contrastive loss)을 획득한다. 일 실시예에서, 대조 손실값은 제1 특징 벡터와 제2 특징 벡터의 내적(dot product) 및 제1 클래스와 제2 클래스의 동일성 여부를 숫자로 나타내는 수식을 통해 계산될 수 있다. 상기 수식은, 제1 클래스와 제2 클래스가 동일한 경우 1을 출력하고, 제1 클래스와 제2 클래스가 동일하지 않는 경우에는 0을 출력할 수 있다. 대조 손실값을 계산하는 구체적인 실시예에 대해서는 도 4의 설명 부분에서 상세하게 설명하기로 한다.

[0064] 단계 S340에서, 전자 장치는 대조 손실값이 최대가 되도록 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크를 이용한 학습을 반복한다. 일 실시예에서, 대조 손실값은 -1 이상 1 이하의 범위의 값을 가질 수 있다. 전자 장치는 대조 손실값이 최대가 되는 방향으로 제1 뉴럴 네트워크 모델 및 제2 뉴럴 네트워크 모델을 이용하는 학습을 반복할 수 있다. 여기서, 학습의 반복은 제1 뉴럴 네트워크 모델 및 제2 뉴럴 네트워크 모델에 적용되는 가중치(weight)를 조절하는 것을 의미할 수 있다.

[0065] 도 4는 본 개시의 일 실시예에 따른 전자 장치가 컨볼루션 뉴럴 네트워크를 이용하여 문장이 속하는 클래스를 분류하는 방법을 설명하기 위한 도면이다.

[0066] 도 4를 참조하면, 전자 장치는 제1 뉴럴 네트워크(401)에 제1 문장(S_1)과 제1 클래스(y_1)를 입력 데이터로 입력하여 제1 문장(S_1)이 제1 클래스(y_1)로 분류되는 확률 분포를 학습할 수 있다. 또한, 전자 장치는 제2 뉴럴 네트워크(402)에 제2 문장(S_2)과 제2 클래스(y_2)를 입력 데이터로 입력하여 제2 문장(S_2)이 제2 클래스(y_2)로 분류되는 확률 분포를 학습할 수 있다. 여기서, 제1 클래스(y_1) 및 제2 클래스(y_2)은 각각 제1 문장 벡터(S_1) 및 제2 문장 벡터(S_2)가 속하는 클래스를 정의하는 벡터값일 수 있다.

[0067] 도 4에서 제1 뉴럴 네트워크(401) 및 제2 뉴럴 네트워크(402)는 컨볼루션 뉴럴 네트워크 모델(Convolution

Neural Network, CNN)으로 구성될 수 있지만, 이에 한정되는 것은 아니다. 일 실시예에서, 제1 뉴럴 네트워크(401) 및 제2 뉴럴 네트워크(402)는 회귀 뉴럴 네트워크(Recurrent Neural Network, RNN), 딥 빌리프 네트워크(Deep Belief Network, DBN), 제한된 볼츠만 기계(Restricted Boltzman Machine, RBM) 방식 등의 인공 뉴럴 네트워크 모델로 구현되거나, 또는 서포트 벡터 머신(Support Vector Machine, SVM)과 같은 머신 러닝 모델로도 구현될 수 있다.

[0068] 전자 장치는 자연어 처리 기술을 통해 제1 문장(S_i) 내의 복수의 단어(단어 1-1 내지 단어 1-6)를 추출하고, 제1 문장(S_i)을 복수의 단어(단어 1-1 내지 단어 1-6)로 파싱(parsing)하고, 복수의 단어(단어 1-1 내지 단어 1-6)를 추출할 수 있다. 도 4에서 단어는 1-1 내지 1-6으로 총 6개로 도시되었지만, 이는 예시적인 것이고, 제1 문장(S_i)에 속하는 단어가 6개로 한정되는 것은 아니다.

[0069] 전자 장치는 복수의 단어(단어 1-1 내지 단어 1-6)를 각각 복수의 단어 벡터($w_{v_{1-1}}$ 내지 $w_{v_{1-6}}$)로 변환할 수 있다. 일 실시예에서, 전자 장치는 제1 문장 벡터(S_i) 및 제2 문장 벡터(S_j)는 word2vec, GloVe, onehot과 같은 기계 학습 모델을 이용하여 복수의 단어(단어 1-1 내지 단어 1-6)를 복수의 단어 벡터($w_{v_{1-1}}$ 내지 $w_{v_{1-6}}$)로 임베딩(embedding)할 수 있다.

[0070] 전자 장치는 변환된 복수의 단어 벡터($w_{v_{1-1}}$ 내지 $w_{v_{1-6}}$)를 매트릭스(matrix) 형태로 배열하여 문장 벡터(411)를 생성하고, 문장 벡터(411)를 제1 뉴럴 네트워크(401)에 입력시킬 수 있다. 문장 벡터(411)는 n 개의 단어와 k 의 차원(dimension)을 갖는 $n \times k$ 매트릭스일 수 있다.

[0071] 전자 장치는 문장 벡터(411)에 서로 다른 폭(width)을 갖는 복수의 필터(421)를 적용하여 컨볼루션 연산을 수행하고, 이를 통해 특징 맵(feature map)(431)을 생성할 수 있다. 복수의 필터(421)는 서로 다른 가중치(weight)를 갖는 벡터로서, 학습이 진행됨에 따라 가중치의 값은 변경될 수 있다. 전자 장치는 문장 벡터(411)의 벡터 값과 복수의 필터(421)의 가중치 값을 서로 곱하고 더하는 연산을 통해 특징 맵(431)을 생성할 수 있다. 도 4에서 복수의 필터(421)는 2, 3, 및 4의 폭을 갖는 것으로 도시되었지만 이에 한정되는 것은 아니다. 복수의 필터(421)에서 차원(k)은 문장 벡터의 차원(k)과 동일할 수 있다.

[0072] 전자 장치는 특징 맵(431)을 맥스 풀링 레이어(max pooling layer)에 통과시켜 특징 맵(431)을 서브 샘플링(subsampling)하고, 이를 통해 제1 특징 벡터(441)를 생성할 수 있다. 제1 특징 벡터(441)는 맥스 풀링 레이어를 통해 특징 맵(431)에서 최대값을 갖는 벡터 값만을 추출하여 생성된 싱글 특징 벡터(single feature vector)로서, 제1 문장(S_i)의 표현(representation)을 나타내는 표현 벡터로 정의될 수 있다. 도 4에는 맥스 풀링 레이어를 통해 제1 특징 벡터(441)를 생성하는 것으로 도시되었지만, 다른 서브 샘플링 레이어를 대체될 수도 있다. 예를 들어, 전자 장치는 평균값 풀링(average pooling) 또는 L_2 -norm pooling 등을 통해 제1 특징 벡터(441)를 생성할 수도 있다.

[0073] 전자 장치는 제1 특징 벡터(441)를 완전 연결 레이어(fully connected layer)에 입력시켜 연결하고(concatenation), 이를 통해 1차원 벡터(451)를 생성할 수 있다. 전자 장치는 1차원 벡터(451)를 소프트맥스 분류기(softmax classifier)에 입력시켜 제1 분류 예측값 벡터(461)를 생성할 수 있다. 제1 분류 예측값 벡터(461)는 제1 문장(S_i)이 제1 클래스(y_1)로 분류될 수 있는 확률 분포를 나타낼 수 있다. 여기서, 전자 장치는 가중치를 조절하는 과정에서 발생하는 오버피팅(overfitting)의 발생을 방지하기 위하여 드롭 아웃(dropout) 연산을 수행할 수도 있다.

[0074] 전자 장치는 제2 뉴럴 네트워크(402)에 제2 문장(S_j) 및 제2 클래스(y_2)를 입력 데이터로 입력시켜 제2 특징 벡터(442)를 생성하고, 이를 통해 제2 분류 예측값 벡터(462)를 생성할 수 있다. 제2 뉴럴 네트워크(402)를 통한 학습 방법은 입력 데이터와 학습 결과를 제외하면 제1 뉴럴 네트워크(401)를 통한 학습 방법과 동일하므로 중복되는 설명은 생략한다.

[0075] 전자 장치는 제1 특징 벡터(441), 제2 특징 벡터(442), 및 제1 클래스(y_1)와 제2 클래스(y_2)의 동일성 여부에 기초하여 제1 문장(S_i)과 제2 문장(S_j)의 표현 상의 유사 정도를 수치화한 대조 손실값(contrastive loss)(L_1)을 획득할 수 있다. 제1 특징 벡터(441)을 $F(S_i)$ 로 정의하고, 제2 특징 벡터(442)를 $F(S_j)$ 로 정의하면, 대조 손실값(L_1)은 하기의 수학식에 기초하여 계산될 수 있다.

수학식 1

$$L_1 = Y \times |F(S_i) \cdot F(S_j)| - (1 - Y) \times |F(S_i) \cdot F(S_j)|$$

$$\text{where, } Y = \begin{cases} 1, y_1 = y_2 \\ 0, y_1 \neq y_2 \end{cases}$$

[0076]

[0077] 수학식 1을 참조하면, 대조 손실값(L₁)은 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)의 내적(dot product)의 절대값 및 Y를 통해 계산될 수 있다. 수학식 1에서 Y는 제1 클래스(y₁)와 제2 클래스(y₂)의 동일성 여부를 숫자로 변환하는 표기(notation)로서, 제1 클래스(y₁)와 제2 클래스(y₂)가 동일한 경우 1을 출력하고, 제1 클래스(y₁)와 제2 클래스(y₂)가 동일하지 않는 경우 0을 출력할 수 있다.

[0078]

제1 문장(S_i)과 제2 문장(S_j)이 서로 다른 클래스에 속하고, 문장 표현이 유사한 경우, 대조 손실값(L₁)은 -1일 수 있다. 즉, 수학식 1에 대입하면, 제1 클래스(y₁)와 제2 클래스(y₂)가 서로 다른 클래스이기 때문에 Y=0이고, 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)가 유사하여 F(S_i)와 F(S_j)의 내적 절대값이 1에 가까운 값이 되므로, 대조 손실값(L₁)은 0 × |-1| - (1-0) × |-1| = -1 로 계산될 수 있다.

[0079]

제1 문장(S_i)과 제2 문장(S_j)이 서로 다른 클래스에 속하고, 문장 표현도 서로 달라 구별되는 경우, 대조 손실값(L₁)은 0일 수 있다. 즉, 제1 클래스(y₁)와 제2 클래스(y₂)가 서로 다른 클래스이기 때문에 Y=0이고, 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)도 서로 구별되는바 cosine 값이 0에 가까워서 내적 절대값이 0에 근사한 값이 될 수 있다. 따라서, 수학식 1에 대입하면, 대조 손실값(L₁)은 0 × |-0| - (1-0) × |-0| = 0 으로 계산될 수 있다.

[0080]

제1 문장(S_i)과 제2 문장(S_j)이 동일한 클래스에 속하고, 문장 표현이 서로 달라 구별되는 경우, 대조 손실값(L₁)은 0일 수 있다. 즉, 제1 클래스(y₁)와 제2 클래스(y₂)가 동일한 클래스이기 때문에 Y=1이고, 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)가 서로 구별되므로 cosine 값이 0에 가까워서 내적 절대값이 0에 근사한 값이 될 수 있다. 따라서, 수학식 1에 대입하면, 대조 손실값(L₁)은 1 × |-0| - (1-1) × |-0| = 0 으로 계산될 수 있다.

[0081]

제1 문장(S_i)과 제2 문장(S_j)이 동일한 클래스에 속하고, 문장 표현이 유사한 경우, 대조 손실값(L₁)은 1일 수 있다. 즉, 제1 클래스(y₁)와 제2 클래스(y₂)가 동일한 클래스이기 때문에 Y=1이고, 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)가 유사한 클래스이므로, cosine 값이 1에 가까워서 내적 절대값이 1에 근사한 값이 될 수 있다. 따라서, 수학식 1에 대입하면, 대조 손실값(L₁)은 1 × |-1| - (1-1) × |-1| = 1 로 계산될 수 있다.

[0082]

상기 수학식 1을 참조하면, 대조 손실값(L₁)은 제1 특징 벡터 및 제2 특징 벡터가 분류되는 클래스(y₁, y₂)의 동일 여부 뿐만 아니라, 제1 특징 벡터 F(S_i)와 제2 특징 벡터 F(S_j)의 유사 정도도 판단할 수 있다.

[0083]

전자 장치는 대조 손실값(L₁)이 최대가 되는 방향으로 학습할 수 있다. 상기 수학식 1을 참조하면, 대조 손실값(L₁)은 -1 이상 1 이하의 값을 갖는바, -1인 경우 제1 뉴럴 네트워크(401) 및 제2 뉴럴 네트워크(402)의 가중치(weight)를 변경하여 학습의 횟수를 증가시킬 수 있다. 이와 반대로 전자 장치는, 대조 손실값(L₁)이 1인 경우, 제1 뉴럴 네트워크(401) 및 제2 뉴럴 네트워크(402)를 통한 학습의 횟수를 줄일 수 있다. 즉, 전자 장치는 제1 문장(S_i)과 제2 문장(S_j)이 서로 다른 클래스에 속함에도 불구하고 서로 유사한 표현을 갖는 경우, 서로 구별되도록 학습 횟수를 증가시킬 수 있다. 또한, 전자 장치는 제1 문장(S_i)과 제2 문장(S_j)이 동일한 클래스에 속하면 서로 유사한 표현을 갖는 경우에는 상대적으로 학습의 횟수를 증가시키지 않을 수 있다.

[0084]

일 실시예에서, 전자 장치는 제1 뉴럴 네트워크(401)를 통한 학습의 출력 데이터인 제1 분류 예측값 벡터(461)와 제1 클래스(y₁)의 벡터의 차이값인 제1 분류 손실값(classification loss)(L₂)을 획득할 수 있다. 마찬가지로

로, 전자 장치는 제2 뉴럴 네트워크(402)를 통한 학습의 출력 데이터인 제2 분류 예측값 벡터(462)와 제2 클래스(y_2)의 벡터의 차이값인 제2 분류 손실값(L_3)을 획득할 수 있다. 제1 분류 손실값(L_2)과 제2 분류 손실값(L_3)은 각각 제1 문장(S_i)이 제1 클래스(y_1)로 분류되는 정도, 제2 문장(S_j)이 제2 클래스(y_2)로 분류되는 정도를 나타내는 값으로서, 값이 작을수록 분류의 정확도가 높을 수 있다.

[0085] 일 실시예에서, 전자 장치는 하기 수학적 식 2와 같이 대조 손실값(L_1), 제1 분류 손실값(L_2), 및 제2 분류 손실값(L_3)을 합산하여 최종 손실값(total loss)(L)을 계산할 수 있다.

수학적 식 2

$$L = L_1 + L_2 + L_3$$

[0086]

[0087] 전자 장치는 계산된 최종 손실값(L)을 기초로 제1 뉴럴 네트워크(401) 및 제2 뉴럴 네트워크(402)에 적용되는 가중치를 조절하는 학습을 수행할 수 있다.

[0088] 도 4에 도시된 본 개시의 실시예에 따르면, 제1 문장(S_i) 및 제2 문장(S_j)이 각각 제1 클래스(y_1) 및 제2 클래스(y_2)로 분류되는 분류 손실값(L_2 , L_3) 뿐만 아니라, 문장 표현(representation)을 고려하는 대조 손실값(L_1)을 동시에 고려함으로써, 서로 다른 클래스에 속하지만, 단어 표현이 유사한 경우 그 표현이 서로 달라지도록 하여 오분류를 막을 수 있다. 즉, 본 개시의 실시예에 따른 전자 장치는 다른 클래스에 속하는 유사한 발화 또는 문장에 대해서 효과적으로 분리할 수 있는 표현을 학습할 수 있고, 따라서 발화 또는 문장 분류의 정확도를 높일 수 있다.

[0089] 특히, 전자 장치가 Bixby 등과 같은 대화형 로봇(Chatbot) 프로그램을 실행하는 경우, 사용자가 입력한 발화인 제1 문장(S_i)이 제1 클래스(y_1)에 속하더라도 문장 표현이 달라서 다른 클래스로 분류될 수 있다. 이런 경우 사용자는 원치 않는 제1 문장(S_i)에 따른 원하는 답변이 아닌 잘못된 분류된 답변을 받을 수 있다. 이와 같은 경우 전자 장치는 대조 손실값(L_1)을 고려하여 학습함으로써, 사용자의 질문이 속하는 클래스로 분류되는 정확도를 높일 수 있다. 또한, 제1 문장(S_i)이 전자 장치에 기 저장된 클래스 중 어떠한 클래스에도 속하지 않는 경우가 있을 수 있는데, 이 경우 전자 장치는 제1 문장(S_i)을 어느 클래스로도 분류하지 않을 수 있어(reject), 사용자가 원치 않는 답변을 받을 가능성을 줄일 수 있다.

[0090] 도 5는 본 개시의 일 실시예에 따른 전자 장치가 제1 문장이 속하는 클래스로 분류되는 확률값인 분류 예측값을 획득하는 방법을 도시한 흐름도이다.

[0091] 단계 S510에서, 전자 장치는 제1 문장을 적어도 하나의 단어 벡터를 포함하는 매트릭스 형태로 변환한다. 일 실시예에서, 제1 문장은 사용자가 입력한 발화 또는 문장일 수 있다. 전자 장치는 제1 문장에 포함된 적어도 하나의 단어를 추출하고, 적어도 하나의 단어를 적어도 하나의 단어 벡터로 각각 변환할 수 있다. 일 실시예에서, 전자 장치는 word2vec, GloVe, onehot 과 같은 기계 학습 모델을 이용하여 적어도 하나의 단어를 적어도 하나의 단어 벡터로 임베딩(embedding)할 수 있다. 전자 장치는 적어도 하나의 단어 벡터를 매트릭스 형태로 배열하여 제1 문장 벡터를 생성할 수 있다.

[0092] 단계 S520에서, 전자 장치는 변환된 매트릭스를 컨볼루션 뉴럴 네트워크(convolution neural network)에 입력 데이터로 입력하고, 복수의 필터(filter)를 적용하여 특징 맵(feature map)을 생성한다. 일 실시예에서, 전자 장치는 서로 다른 폭(width)을 갖는 다중 필터를 적용하여 컨볼루션 연산을 수행할 수 있다. 다중 필터는 서로 다른 가중치(weight)를 갖는 벡터로서, 학습이 진행됨에 따라 가중치의 값은 변경될 수 있다. 다중 필터는 단계 S510에서 생성된 문장 벡터의 차원(dimension)과 동일한 차원을 가질 수 있다.

[0093] 단계 S530에서, 전자 장치는 특징 맵을 맥스 풀링 레이어(max pooling layer)에 통과시켜 제1 특징 벡터를 추출한다. 전자 장치는 맥스 풀링 레이어를 통해 특징 맵에서 최대값을 갖는 벡터 값만을 추출하여 생성된 싱글 특징 벡터(single feature vector)인 제1 특징 벡터를 추출할 수 있다. 다만, 서브 샘플링에 사용되는 레이어가 맥스 풀링 레이어로 한정되는 것은 아니다. 일 실시예에서, 전자 장치는 평균값 풀링(average pooling) 또는

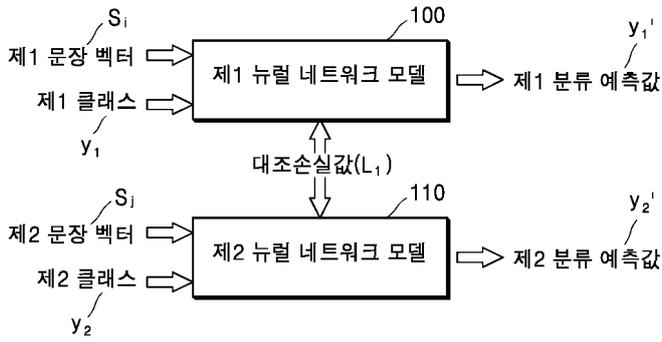
L_2 -norm pooling 등을 통해 제1 특징 벡터를 추출할 수도 있다.

- [0094] 단계 S540에서, 전자 장치는 제1 특징 벡터를 완전 연결 레이어(fully connected layer)에 입력시켜 1차원 벡터 값으로 변환한다. 일 실시예에서, 전자 장치는 서로 다른 폭을 갖는 필터를 이용하여 생성된 복수의 특징 맵들로 구성된 제1 특징 벡터를 하나로 연결(concatenate)하여 1차원 벡터값으로 변환할 수 있다. 단계 S540에서, 제1 특징 벡터를 학습하는 동안 발생하는 오버 피팅(overfitting)을 해결하고, 학습 데이터의 정확도를 높이기 위하여 드롭 아웃(dropout) 연산이 이용될 수도 있다.
- [0095] 단계 S550에서, 전자 장치는 1차원 벡터값을 소프트 맥스 분류기(softmax classifier)에 입력시켜 제1 분류 예측값을 획득한다. 일 실시예에서, 제1 분류 예측값은 제1 문장이 제1 클래스로 분류될 수 있는 확률값을 의미하는 것으로서, 소프트 맥스 분류기를 통과함으로써 생성될 수 있다. 1차원 벡터에 포함되는 벡터 값은 소프트 맥스 분류기를 통과하여 벡터 값의 총 합이 1이 되는 확률값으로 변환될 수 있다.
- [0096] 도 5에는 제1 문장을 컨볼루션 뉴럴 네트워크에 입력시켜 제1 분류 예측값을 획득하는 과정을 도시하였지만, 도시된 단계들은 제2 문장에도 동일하게 적용될 수 있다. 일 실시예에서, 전자 장치는 단계 S510 내지 S550에 따라 제2 문장을 컨볼루션 뉴럴 네트워크에 입력시켜 제2 분류 예측값을 획득할 수 있다. 일 실시예에서, 전자 장치는 제1 분류 예측값을 획득하는 제1 학습 과정과 제2 분류 예측값을 획득하는 제2 학습 과정을 동시에 수행할 수 있다.
- [0097] 도 6은 본 개시의 일 실시예에 따른 전자 장치가 뉴럴 네트워크 모델을 통해 획득된 손실값을 기초로 하여 뉴럴 네트워크 모델에 적용된 가중치(weight)를 조절하는 학습 방법을 도시한 흐름도이다.
- [0098] 단계 S610에서, 전자 장치는 제1 분류 예측값과 제1 클래스 간의 차이값인 제1 분류 손실값을 획득한다. 제1 분류 손실값은 제1 문장이 제1 클래스로 분류될 수 있는 확률값인 제1 분류 예측값과 제1 클래스 벡터의 차이값을 의미할 수 있다.
- [0099] 단계 S620에서, 전자 장치는 제2 분류 예측값과 제2 클래스 간의 차이값인 제2 분류 손실값을 획득한다. 제2 분류 손실값은 제2 문장이 제2 클래스로 분류될 수 있는 확률값인 제2 분류 예측값과 제2 클래스 벡터의 차이값을 의미할 수 있다. 일 실시예에서, 단계 S610과 단계 S620은 동시에 수행될 수 있다.
- [0100] 단계 S630에서, 전자 장치는 제1 분류 손실값, 제2 분류 손실값, 및 대조 손실값을 합산하여 최종 손실값을 획득한다. 대조 손실값을 계산하는 구체적인 방법은 도 4의 설명 부분에서 설명하였는바, 중복되는 설명은 생략한다.
- [0101] 단계 S640에서, 전자 장치는 최종 손실값을 기초로 하여 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크에 적용되는 가중치(weight)를 조절한다. 일 실시예에서, 제1 뉴럴 네트워크 및 제2 뉴럴 네트워크는 복수의 필터를 적용하여 특징 맵을 생성하는 컨볼루션 뉴럴 네트워크로 구성되고, 전자 장치는 최종 손실값의 크기에 따라 컨볼루션 뉴럴 네트워크에 적용되는 복수의 필터의 가중치 값을 조절할 수 있다.
- [0102] 본 명세서에서 설명된 전자 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 개시된 실시예들에서 설명된 전자 장치는 프로세서, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다.
- [0103] 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.
- [0104] 소프트웨어는, 컴퓨터로 읽을 수 있는 저장 매체(computer-readable storage media)에 저장된 명령어를 포함하는 컴퓨터 프로그램으로 구현될 수 있다. 컴퓨터가 읽을 수 있는 기록 매체로는, 예를 들어 마그네틱 저장 매체(예컨대, ROM(read-only memory), RAM(random-access memory), 플로피 디스크, 하드 디스크 등) 및 광학적 판독 매체(예컨대, 시디롬(CD-ROM), 디브이디(DVD: Digital Versatile Disc)) 등이 있다. 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템들에 분산되어, 분산 방식으로 컴퓨터가 판독 가능한 코드가 저장되고 실행될 수 있다. 매체는 컴퓨터에 의해 판독 가능하며, 메모리에 저장되고, 프로세서에서 실행될 수 있다.

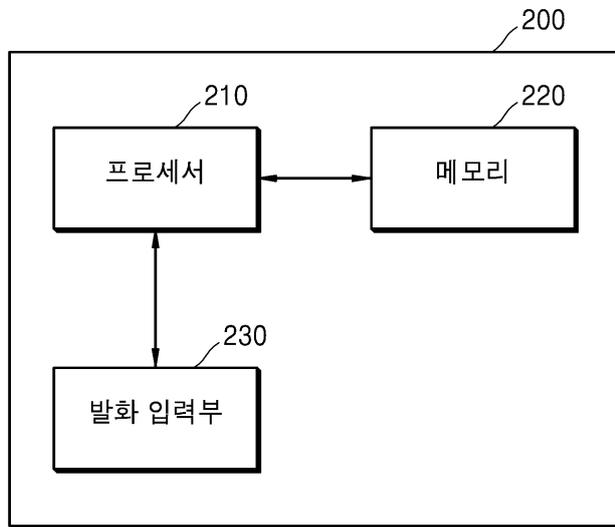
- [0105] 컴퓨터는, 저장 매체로부터 저장된 명령어를 호출하고, 호출된 명령어에 따라 개시된 실시예에 따른 동작이 가능한 장치로서, 개시된 실시예들에 따른 전자 장치를 포함할 수 있다.
- [0106] 컴퓨터로 읽을 수 있는 저장매체는, 비일시적(non-transitory) 저장매체의 형태로 제공될 수 있다. 여기서, '비일시적'은 저장매체가 신호(signal)를 포함하지 않으며 실제(tangible)하다는 것을 의미할 뿐 데이터가 저장매체에 반영구적 또는 임시적으로 저장됨을 구분하지 않는다.
- [0107] 또한, 개시된 실시예들에 따른 전자 장치 또는 방법은 컴퓨터 프로그램 제품(computer program product)에 포함되어 제공될 수 있다. 컴퓨터 프로그램 제품은 상품으로서 판매자 및 구매자 간에 거래될 수 있다.
- [0108] 컴퓨터 프로그램 제품은 소프트웨어 프로그램, 소프트웨어 프로그램이 저장된 컴퓨터로 읽을 수 있는 저장 매체를 포함할 수 있다. 예를 들어, 컴퓨터 프로그램 제품은 전자 장치의 제조사 또는 전자 마켓(예를 들어, 구글 플레이 스토어, 앱 스토어)을 통해 전자적으로 배포되는 소프트웨어 프로그램 형태의 상품(예를 들어, 다운로드 가능한 애플리케이션(downloadable application))을 포함할 수 있다. 전자적 배포를 위하여, 소프트웨어 프로그램의 적어도 일부는 저장 매체에 저장되거나, 임시적으로 생성될 수 있다. 이 경우, 저장 매체는 제조사의 서버, 전자 마켓의 서버, 또는 소프트웨어 프로그램을 임시적으로 저장하는 중계 서버의 저장매체가 될 수 있다.
- [0109] 컴퓨터 프로그램 제품은, 서버 및 단말(예로, 초음파 진단 장치)로 구성되는 시스템에서, 서버의 저장매체 또는 단말의 저장매체를 포함할 수 있다. 또는, 서버 또는 단말과 통신 연결되는 제3 장치(예, 스마트폰)가 존재하는 경우, 컴퓨터 프로그램 제품은 제3 장치의 저장매체를 포함할 수 있다. 또는, 컴퓨터 프로그램 제품은 서버로부터 단말 또는 제3 장치로 전송되거나, 제3 장치로부터 단말로 전송되는 소프트웨어 프로그램 자체를 포함할 수 있다.
- [0110] 이 경우, 서버, 단말 및 제3 장치 중 하나가 컴퓨터 프로그램 제품을 실행하여 개시된 실시예들에 따른 방법을 수행할 수 있다. 또는, 서버, 단말 및 제3 장치 중 둘 이상이 컴퓨터 프로그램 제품을 실행하여 개시된 실시예들에 따른 방법을 분산하여 실시할 수 있다.
- [0111] 예를 들면, 서버(예로, 클라우드 서버 또는 인공 지능 서버 등)가 서버에 저장된 컴퓨터 프로그램 제품을 실행하여, 서버와 통신 연결된 단말이 개시된 실시예들에 따른 방법을 수행하도록 제어할 수 있다.
- [0112] 또 다른 예로, 제3 장치가 컴퓨터 프로그램 제품을 실행하여, 제3 장치와 통신 연결된 단말이 개시된 실시예에 따른 방법을 수행하도록 제어할 수 있다.
- [0113] 제3 장치가 컴퓨터 프로그램 제품을 실행하는 경우, 제3 장치는 서버로부터 컴퓨터 프로그램 제품을 다운로드하고, 다운로드된 컴퓨터 프로그램 제품을 실행할 수 있다. 또는, 제3 장치는 프리로드된 상태로 제공된 컴퓨터 프로그램 제품을 실행하여 개시된 실시예들에 따른 방법을 수행할 수도 있다.
- [0114] 또한, 이상에서는 본 개시의 실시예에 대하여 도시하고 설명하였지만, 본 개시는 전술한 특정의 실시예에 한정되지 아니하며, 청구범위에서 청구하는 본 발명의 요지를 벗어남이 없이 당해 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 다양한 변형실시가 가능한 것은 물론이고, 이러한 변형 실시들은 본 개시의 기술적 사상이나 전망으로부터 개별적으로 이해되어서는 안될 것이다.
- [0115] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 전자 장치, 구조, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

도면

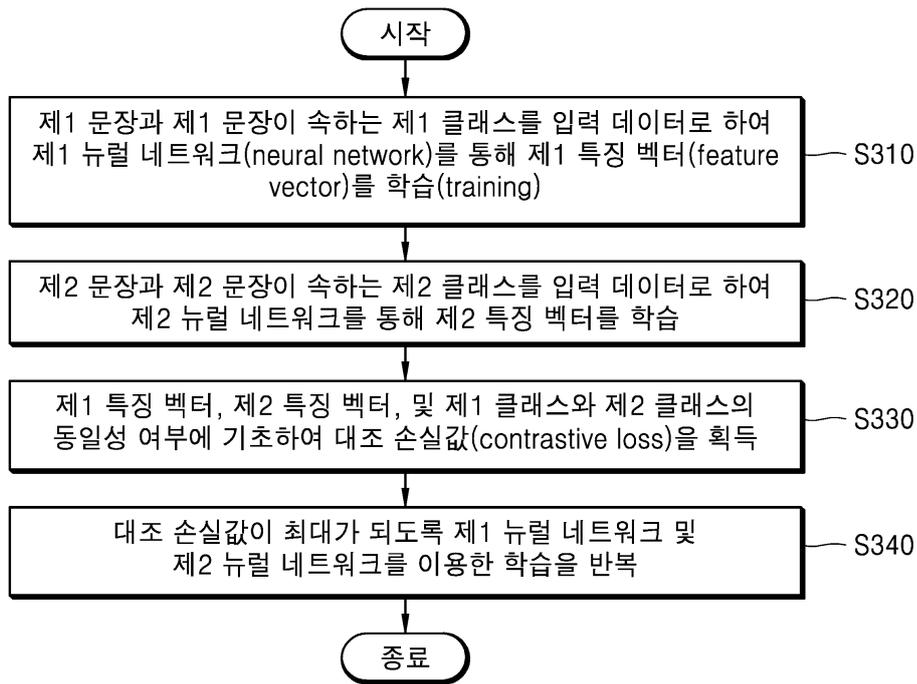
도면1



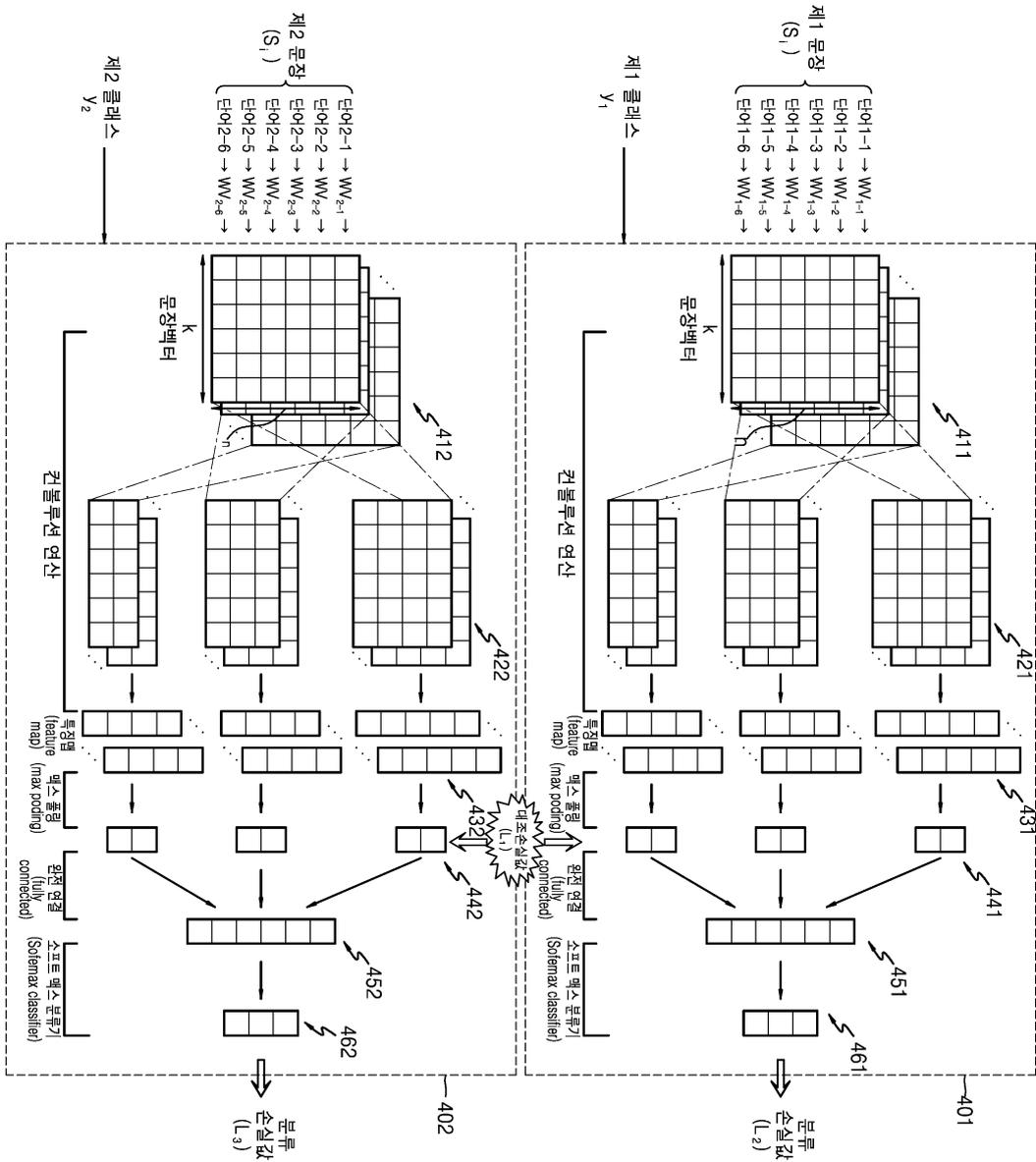
도면2



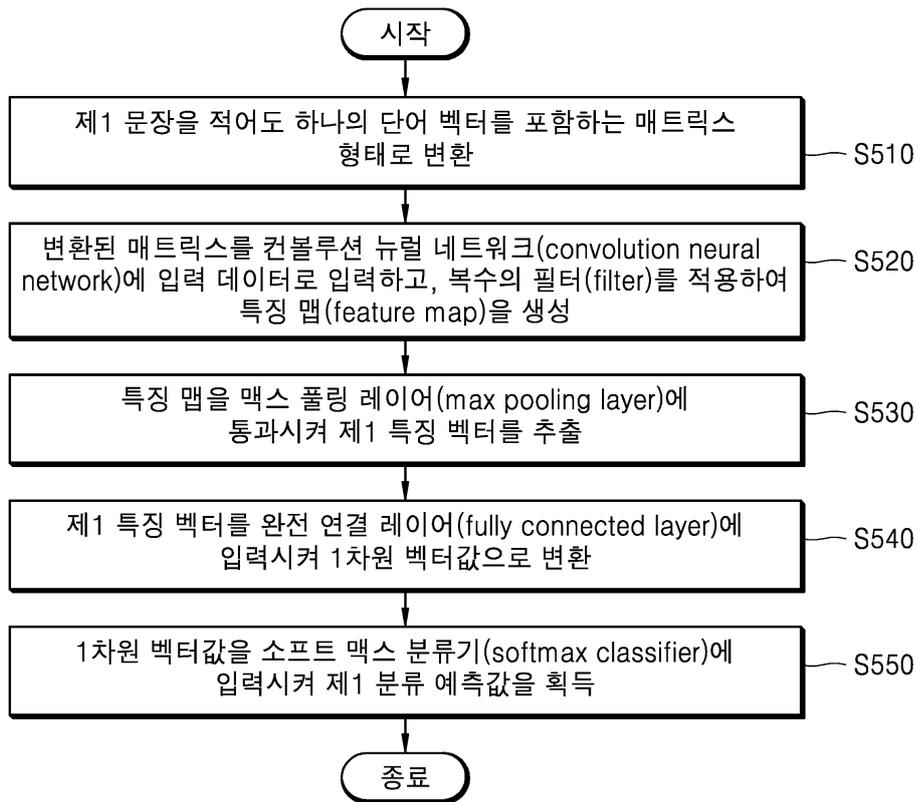
도면3



도면4



도면5



도면6

