



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0061319
(43) 공개일자 2025년05월08일

(51) 국제특허분류(Int. Cl.)
G06N 20/00 (2019.01) G06F 18/24 (2023.01)
(52) CPC특허분류
G06N 20/00 (2021.08)
G06F 18/24 (2023.01)
(21) 출원번호 10-2023-0145630
(22) 출원일자 2023년10월27일
심사청구일자 2023년10월27일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
박경문
경기도 화성시 삼성1로 209, 517호 (석우동)
문준영
강원특별자치도 춘천시 안마산로 244, 905동 1103호 (퇴계동, 퇴계9단지주공아파트)
박진희
서울특별시 강남구 봉은사로51길 40, 102호 (논현동, 캐럿스테이트)
(74) 대리인
두호특허법인

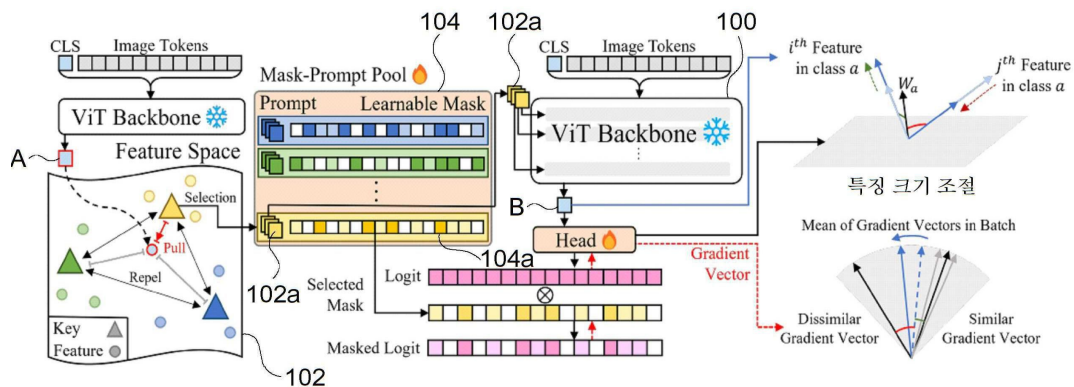
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 인공지능 기반 분류 모델의 온라인 점진 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치

(57) 요약

인공지능 기반 분류 모델의 온라인 점진 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치가 개시된다. 개시되는 일 실시예에 따른 학습 방법은, 하나 이상의 프로세서들, 및 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, 인공지능 기반의 분류 모델을 학습하기 위한 방법으로서, 학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하는 단계, 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하는 단계, 제1 타입의 데이터와 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하는 단계, 및 생성한 학습 데이터를 분류 모델에 입력하여 분류 모델을 학습시키는 단계를 포함한다.

대표도



이 발명을 지원한 국가연구개발사업

과제고유번호	1711198845
과제번호	00258649
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	거대 AI 서비스 지원 클라우드 컨티뉴엄 핵심 기술개발
기 여 율	50/100
과제수행기관명	경희대학교산학협력단
연구기간	2023.07.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711191911
과제번호	2021R1G1A1094379
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	부정맥의 효과적인 조기 진단을 위한 환자 맞춤형 General Intelligence 연구
기 여 율	30/100
과제수행기관명	경희대학교(국제캠퍼스)
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179316
과제번호	00155911
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합혁신인재양성
연구과제명	인공지능융합혁신인재양성(경희대학교)
기 여 율	10/100
과제수행기관명	경희대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193622
과제번호	2021-0-02068-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능 혁신 허브 연구 개발
기 여 율	10/100
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, 인공지능 기반 분류 모델의 온라인 점진 학습 방법으로서,

학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하는 단계;

상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하는 단계;

상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하는 단계; 및

상기 생성한 학습 데이터를 분류 모델에 입력하여 상기 분류 모델을 학습시키는 단계를 포함하는, 분류 모델의 학습 방법.

청구항 2

청구항 1에 있어서,

상기 제1 타입의 데이터는, 해당 데이터가 여러 개의 태스크에 걸쳐서 등장하는 데이터이고,

상기 제2 타입의 데이터는, 해당 데이터가 하나의 태스크에만 등장하는 데이터인, 분류 모델의 학습 방법.

청구항 3

청구항 2에 있어서,

상기 제2 비율은,

각 태스크에서 제1 타입의 데이터가 메이저 클래스(major class)로 분류되는 데이터와 마이너 클래스(minor class)로 분류되는 데이터 간의 비율인, 분류 모델의 학습 방법.

청구항 4

청구항 1에 있어서,

상기 분류 모델을 학습시키는 단계는,

상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하는 단계를 포함하는, 분류 모델의 학습 방법.

청구항 5

청구항 4에 있어서,

상기 프롬프트를 선택하는 단계는,

상기 추출한 특징과 상기 프롬프트 풀에 저장된 키 벡터들 간의 유사도에 기초하여 하나 이상의 키 벡터를 선택

하고, 상기 선택된 키 벡터에 대응하는 프롬프트를 추출하는, 분류 모델의 학습 방법.

청구항 6

청구항 5에 있어서,

상기 분류 모델을 학습시키는 단계는,

상기 프롬프트 풀의 키 벡터들이 학습이 될수록 특징 공간(feature space)에서 움직임이 둔해지도록 하면서 키 벡터들 간의 거리가 유지되도록 상기 프롬프트 풀을 학습하는 단계를 더 포함하는, 분류 모델의 학습 방법.

청구항 7

청구항 6에 있어서,

상기 프롬프트 풀의 학습은 하기의 수학식에 따른 제1 손실 함수($\mathcal{L}_{CVPT}$)에 기초하여 수행되는, 분류 모델의 학습 방법.

(수학식)

$$\mathcal{L}_{CVPT} = -\log \frac{s_n}{s_p + s_n}$$

$$s_n = \sum_{p=1}^P \sum_{q=1}^P \exp(\delta(\mathbf{k}_p, \mathbf{k}_q) / (\mathcal{C}_p + 1))$$

$$s_p = \sum_{p=1}^P \sum_{q=1}^B \exp(\delta(\mathbf{k}_p, \mathbf{q}_q) / (\mathcal{C}_p + 1))$$

P : 프롬프트 풀의 사이즈

B : 학습 데이터의 개수

δ : 코사인 거리

$\mathbf{k}_p$: p번째 프롬프트에 대응하는 키 벡터

$\mathbf{q}_q$: q번째 학습 데이터의 특징

$\mathcal{C}_p$: p번째 프롬프트가 선택되는 횟수

청구항 8

청구항 4에 있어서,

상기 분류 모델을 학습시키는 단계는,

상기 학습 데이터의 토큰과 함께 상기 학습 데이터에 대한 CLS(Classification) 토큰을 상기 분류 모델의 입력 단으로 입력하는 단계; 및

상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 직접 입력하는 단계를 더 포함하는, 분류 모델의 학습 방법.

청구항 9

청구항 8에 있어서,

상기 분류 모델을 학습시키는 단계는,

상기 분류 모델의 출력단에서 상기 CLS 토큰의 특징을 출력하고, 상기 CLS 토큰의 특징을 분류기에 통과시켜 CLS 토큰의 출력 분포를 생성하는 단계;

상기 선택된 프롬프트와 대응되는 마스크를 추출하는 단계; 및

상기 CLS 토큰의 출력 분포에 상기 마스크를 적용하여 상기 CLS 토큰의 마스크 된 출력 분포를 생성하는 단계를 더 포함하는, 분류 모델의 학습 방법.

청구항 10

청구항 8에 있어서,

상기 분류 모델을 학습시키는 단계는,

상기 분류 모델의 상기 학습 데이터 전체에 대한 평균 그래디언트 벡터(average gradient vector)와 상기 분류 모델의 각 학습 데이터의 그래디언트 벡터에 기초하여 각 학습 데이터의 무시 스코어(ignorance score)를 산출하는 단계; 및

상기 무시 스코어에 기초하여 소수(minor) 분류로 분류된 학습 데이터에 대한 손실을 강조하는 단계를 더 포함하는, 분류 모델의 학습 방법.

청구항 11

청구항 10에 있어서,

상기 무시 스코어는, 상기 분류 모델의 학습 중 다른 학습 데이터가 해당 학습 데이터를 무시하는 정도를 나타낸 것이고,

상기 무시 스코어는, 상기 학습 데이터 전체에 대한 평균 그래디언트 벡터(average gradient vector)와 각 학습 데이터의 그래디언트 벡터 간의 유사도에 기초하여 산출되는, 분류 모델의 학습 방법.

청구항 12

청구항 11에 있어서,

상기 소수 분류로 분류된 학습 데이터에 대한 손실을 강조하는 단계는,

하기 수학적식에 따른 제2 손실 함수($\mathcal{L}_{GSF}$)에 기초하여 수행되는, 분류 모델의 학습 방법.

(수학적식)

$$\mathcal{L}_{GSF} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \left(\text{Score}_i^{\text{ign}} \right)^{\gamma} \cdot \mathcal{L}_{CE}(\hat{y}_i, y_i)$$

B : 배치 당 학습 데이터의 개수

$\text{Score}_i^{\text{ign}}$: i번째 학습 데이터에 대한 무시 스코어

L_{CE} : 교차 엔트로피 손실

y_i : i번째 학습 데이터의 분류에 대한 정답 값

$\hat{y}_i$: i번째 학습 데이터의 분류 모델의 분류에 대한 예측 값

γ : 기 설정되는 하이퍼 파라미터

청구항 13

청구항 8에 있어서,

상기 분류 모델을 학습시키는 단계는,

각 학습 데이터에 대해 마진 이익 스코어(marginal benefit score)를 산출하는 단계; 및

상기 마진 이익 스코어에 기초하여 해당 학습 데이터의 CLS 토큰의 특징 크기를 조절하는 단계를 더 포함하는, 분류 모델의 학습 방법.

청구항 14

청구항 13에 있어서,

상기 마진 이익 스코어는, 각 학습 데이터가 상기 분류 모델의 학습에 얼마나 이익이 되는지를 나타내는 정도이고,

상기 마진 이익 스코어는, 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대한 분류기의 가중치에 기초하여 산출되는, 분류 모델의 학습 방법.

청구항 15

청구항 14에 있어서,

상기 마진 이익 스코어(Score^{MB})는 하기 수학식에 의해 산출되는, 분류 모델의 학습 방법.

(수학식)

$$\text{Score}_i^{MB} = \delta(f(\mathbf{x}_i), W_{y_i}) + m$$

δ : 코사인 거리(cosine distance)

W_{y_i} : i번째 학습 데이터의 정답 값에 대응하는 분류기의 가중치

$\mathbf{x}_i$: i번째 학습 데이터

$f(\mathbf{x}_i)$: i번째 학습 데이터의 CLS 토큰의 특징

m : 기 설정되는 마진(margin)

청구항 16

청구항 13에 있어서,

상기 CLS 토큰의 특징 크기를 조절하는 단계는,

상기 마진 이익 스코어가 클수록 해당 학습 데이터의 상기 CLS 토큰의 특징 크기가 작도록 조절하고, 상기 마진 이익 스코어가 작을수록 해당 학습 데이터의 CLS 토큰의 특징 크기가 크도록 조절하는, 분류 모델의 학습 방법.

청구항 17

하나 이상의 프로세서들;

메모리; 및

하나 이상의 프로그램들을 포함하고,

상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성되며,

상기 하나 이상의 프로그램들은,

학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하기 위한 명령;

상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하기 위한 명령;

상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하기 위한 명령; 및

상기 생성한 학습 데이터를 분류 모델에 입력하여 상기 분류 모델을 학습시키기 위한 명령을 포함하는, 컴퓨팅 장치.

청구항 18

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, 인공지능 기반의 분류 모델을 학습하기 위한 방법으로서,

학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하는 단계;

상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하는 단계;

상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하는 단계;

상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하는 단계; 및

상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 입력하여 상기 학습 데이터와 함께 상기 분류 모델을 학습시키는 단계를 포함하는, 분류 모델의 학습 방법.

청구항 19

하나 이상의 프로세서들;

메모리; 및

하나 이상의 프로그램들을 포함하고,

상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성되며,

상기 하나 이상의 프로그램들은,

학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하기 위한 명령;

상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하기 위한 명령;

상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하기 위한 명령;

상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하기 위한 명령; 및

상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 입력하여 상기 학습 데이터와 함께 상기 분류 모델을 학습시키기 위한 명령을 포함하는, 컴퓨팅 장치.

발명의 설명

기술 분야

[0001] 본 발명의 실시예는 인공지능 기반의 분류 모델을 학습하기 위한 기술과 관련된다.

배경 기술

[0003] 인공지능 모델을 학습할 때 사용하는 많은 데이터들은 대부분 균등하게 잘 정제되어 있다. 그러나, 현실의 모든 데이터들이 인공지능 학습에 적합하도록 잘 정제되어 있는 것은 아니다. 인공지능의 규모와 컴퓨터 연산 성능이 기하급수적으로 성장하면서, 인공지능의 학습에 필요한 데이터의 양도 점차 많은 양이 요구되고 있다. 이에 비해, 학습할 데이터를 정제하는 것은 여전히 인간의 능력을 요구하여 한계가 있으며, 정제된 데이터는 항상 연속적으로 증가하고 그 양이 불균일한 특성을 가진다. 인공지능 모델은 확률 기반의 모델이기 때문에 학습 데이터가 연속적으로 증가하며 데이터의 양에 있어 불균일한 특성을 갖는 환경에서는 이전에 학습한 분포를 망각하는 파괴적 망각(catastrophic forgetting)이라고 부르는 현상이 발생하게 된다. 또한, 한 작업에서 데이터의 수량이 균등하게 정제되어 있지 않기 때문에 학습 불균형의 문제가 발생한다.

[0004] 또한, 현실에서는 작업이 변화하는 경계를 명확하게 알 수 없는 경우가 많다. 그리고, 일부 데이터는 그 특성상 특정 기간에만 접근할 수 있는 환경이 많이 존재한다. 따라서, 인공지능 모델에서 파괴적 망각 현상을 방지하면서 불명확한 작업 경계에 적용 가능할 수 있는 방안이 요구된다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국공개특허공보 제10-2023-0068107호(2023.05.17)

발명의 내용

해결하려는 과제

[0007] 본 발명의 실시예는 작업 경계가 불분명하고 불균일한 데이터들에 대해 분류를 수행할 수 있는 인공지능 기반 분류 모델의 온라인 점진 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치를 제공하기 위한 것이다.

과제의 해결 수단

[0009] 개시되는 일 실시예에 따른 분류 모델의 학습 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서

들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, 인공지능 기반의 분류 모델을 학습하기 위한 방법으로서, 학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하는 단계; 상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하는 단계; 상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하는 단계; 및 상기 생성한 학습 데이터를 분류 모델에 입력하여 상기 분류 모델을 학습시키는 단계를 포함한다.

[0010] 상기 제1 타입의 데이터는, 해당 데이터가 여러 개의 태스크에 걸쳐서 등장하는 데이터이고, 상기 제2 타입의 데이터는, 해당 데이터가 하나의 태스크에만 등장하는 데이터일 수 있다.

[0011] 상기 제2 비율은, 각 태스크에서 제1 타입의 데이터가 메이저 클래스(major class)로 분류되는 데이터와 마이너 클래스(minor class)로 분류되는 데이터 간의 비율일 수 있다.

[0012] 상기 분류 모델을 학습시키는 단계는, 상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하는 단계를 포함할 수 있다.

[0013] 상기 프롬프트를 선택하는 단계는, 상기 추출한 특징과 상기 프롬프트 풀에 저장된 키 벡터들 간의 유사도에 기초하여 하나 이상의 키 벡터를 선택하고, 상기 선택된 키 벡터에 대응하는 프롬프트를 추출할 수 있다.

[0014] 상기 분류 모델을 학습시키는 단계는, 상기 프롬프트 풀의 키 벡터들이 학습이 될수록 특징 공간(feature space)에서 움직임이 둔해지도록 하면서 키 벡터들 간의 거리가 유지되도록 상기 프롬프트 풀을 학습하는 단계를 더 포함할 수 있다.

[0015] 상기 프롬프트 풀의 학습은 하기의 수학식에 따른 제1 손실 함수($\mathcal{L}_{CVPT}$)에 기초하여 수행될 수 있다.

[0016] (수학식)

$$\mathcal{L}_{CVPT} = -\log \frac{s_n}{s_p + s_n}$$

$$s_n = \sum_{p=1}^P \sum_{q=1}^P \exp(\delta(\mathbf{k}_p, \mathbf{k}_q) / (C_p + 1))$$

$$s_p = \sum_{p=1}^P \sum_{q=1}^B \exp(\delta(\mathbf{k}_p, \mathbf{q}_q) / (C_p + 1))$$

[0019] P : 프롬프트 풀의 사이즈

[0020] B : 학습 데이터의 개수

[0021] δ : 코사인 거리

[0022] $\mathbf{k}_p$: p번째 프롬프트에 대응하는 키 벡터

[0023] $\mathbf{q}_q$: q번째 학습 데이터의 특징

[0024] C_p : p번째 프롬프트가 선택되는 횟수

[0025] 상기 분류 모델을 학습시키는 단계는, 상기 학습 데이터의 토큰과 함께 상기 학습 데이터에 대한 CLS(Classification) 토큰을 상기 분류 모델의 입력단으로 입력하는 단계; 및 상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 직접 입력하는 단계를 더 포함할 수 있다.

[0026] 상기 분류 모델을 학습시키는 단계는, 상기 분류 모델의 출력단에서 상기 CLS 토큰의 특징을 출력하고, 상기 CLS 토큰의 특징을 분류기에 통과시켜 CLS 토큰의 출력 분포를 생성하는 단계; 상기 선택된 프롬프트와 대응되는 마스크를 추출하는 단계; 및 상기 CLS 토큰의 출력 분포에 상기 마스크를 적용하여 상기 CLS 토큰의 마스크

된 출력 분포를 생성하는 단계를 더 포함할 수 있다.

[0028] 상기 분류 모델을 학습시키는 단계는, 상기 분류 모델의 상기 학습 데이터 전체에 대한 평균 그래디언트 벡터(average gradient vector)와 상기 분류 모델의 각 학습 데이터의 그래디언트 벡터에 기초하여 각 학습 데이터의 무시 스코어(ignorance score)를 산출하는 단계; 및 상기 무시 스코어에 기초하여 소수(minor) 분류로 분류된 학습 데이터에 대한 손실을 강조하는 단계를 더 포함할 수 있다.

[0029] 상기 무시 스코어는, 상기 분류 모델의 학습 중 다른 학습 데이터가 해당 학습 데이터를 무시하는 정도를 나타낸 것이고, 상기 무시 스코어는, 상기 학습 데이터 전체에 대한 평균 그래디언트 벡터(average gradient vector)와 각 학습 데이터의 그래디언트 벡터 간의 유사도에 기초하여 산출될 수 있다.

[0030] 상기 소수 분류로 분류된 학습 데이터에 대한 손실을 강조하는 단계는, 하기 수학식에 따른 제2 손실 함수(L_{GSF})에 기초하여 수행될 수 있다.

[0031] (수학식)

$$\mathcal{L}_{GSF} = \frac{1}{|B|} \sum_{i=1}^{|B|} \left(\text{Score}_i^{\text{ign}} \right)^{\gamma} \cdot \mathcal{L}_{CE}(\hat{y}_i, y_i)$$

[0032]

[0033] B : 배치 당 학습 데이터의 개수

[0034] $\text{Score}_i^{\text{ign}}$: i번째 학습 데이터에 대한 무시 스코어

[0035] L_{CE} : 교차 엔트로피 손실

[0036] y_i : i번째 학습 데이터의 분류에 대한 정답 값

[0037] $\hat{y}_i$: i번째 학습 데이터의 분류 모델의 분류에 대한 예측 값

[0038] γ : 기 설정되는 하이퍼 파라미터

[0039] 상기 분류 모델을 학습시키는 단계는, 각 학습 데이터에 대해 마진 이익 스코어(marginal benefit score)를 산출하는 단계; 및 상기 마진 이익 스코어에 기초하여 해당 학습 데이터의 CLS 토큰의 특징 크기를 조절하는 단계를 더 포함할 수 있다.

[0040] 상기 마진 이익 스코어는, 각 학습 데이터가 상기 분류 모델의 학습에 얼마나 이익이 되는지를 나타내는 정도이고, 상기 마진 이익 스코어는, 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대한 분류기의 가중치에 기초하여 산출될 수 있다.

[0041] 상기 마진 이익 스코어(Score^{MB})는 하기 수학식에 의해 산출될 수 있다.

[0042] (수학식)

$$\text{Score}_i^{\text{MB}} = \delta(f(\mathbf{x}_i), W_{y_i}) + m$$

[0043]

[0044] δ : 코사인 거리(cosine distance)

[0045] W_{y_i} : i번째 학습 데이터의 정답 값에 대응하는 분류기의 가중치

[0046] x_i : i번째 학습 데이터

[0047] $f(x_i)$: i번째 학습 데이터의 CLS 토큰의 특징

[0048] m : 기 설정되는 마진(margin)

[0049] 상기 CLS 토큰의 특징 크기를 조절하는 단계는, 상기 마진 이익 스코어가 클수록 해당 학습 데이터의 상기 CLS 토큰의 특징 크기가 작도록 조절하고, 상기 마진 이익 스코어가 작을수록 해당 학습 데이터의 CLS 토큰의 특징

크기가 크도록 조절할 수 있다.

[0050] 개시되는 일 실시예에 따른 컴퓨팅 장치는, 하나 이상의 프로세서들; 메모리; 및 하나 이상의 프로그램들을 포함하고, 상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성되며, 상기 하나 이상의 프로그램들은, 학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하기 위한 명령; 상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하기 위한 명령; 상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하기 위한 명령; 및 상기 생성한 학습 데이터를 분류 모델에 입력하여 상기 분류 모델을 학습시키기 위한 명령을 포함한다.

[0051] 개시되는 다른 실시예에 따른 분류 모델의 학습 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, 인공지능 기반의 분류 모델을 학습하기 위한 방법으로서, 학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하는 단계; 상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하는 단계; 상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하는 단계; 상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하는 단계; 및 상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 입력하여 상기 학습 데이터와 함께 상기 분류 모델을 학습시키는 단계를 포함한다.

[0052] 개시되는 다른 실시예에 따른 컴퓨팅 장치는, 하나 이상의 프로세서들; 메모리; 및 하나 이상의 프로그램들을 포함하고, 상기 하나 이상의 프로그램들은 상기 메모리에 저장되고, 상기 하나 이상의 프로세서들에 의해 실행되도록 구성되며, 상기 하나 이상의 프로그램들은, 학습 데이터 셋 내의 모든 데이터들을 기 설정된 제1 비율에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분하기 위한 명령; 상기 제1 타입의 데이터들에 대해 기 설정된 제2 비율에 따라 각 태스크에 나누어 분배하기 위한 명령; 상기 제1 타입의 데이터와 상기 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터를 생성하기 위한 명령; 상기 학습 데이터에서 특징을 추출하고, 상기 추출한 특징에 기초하여 기 설정된 프롬프트 풀에서 하나 이상의 프롬프트를 선택하기 위한 명령; 및 상기 선택한 프롬프트를 상기 분류 모델의 일부 레이어로 입력하여 상기 학습 데이터와 함께 상기 분류 모델을 학습시키기 위한 명령을 포함한다.

발명의 효과

[0054] 개시되는 실시예에 의하면, 작업의 경계가 불분명하고 불균일한 특성을 갖는 현실 세계의 데이터에 대해 분류를 효과적으로 수행할 수 있게 된다. 또한, 분류 모델의 학습 시 치명적 망각 현상을 방지하고, 분류의 불균형 문제를 방지하여 각 학습 데이터를 균형적으로 학습할 수 있게 된다.

도면의 간단한 설명

[0056] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

도 2는 본 발명의 일 실시예에서 분류 모델을 학습하기 위한 학습 데이터 셋을 생성하는 상태를 나타낸 도면

도 3은 본 발명의 일 실시예에서 인공지능 기반의 분류 모델을 학습하는 과정을 나타낸 도면

도 4는 본 발명의 일 실시예에 따른 프롬프트 풀(102)에서 키 벡터들이 선택됨에 따라 특징 공간에서 표류(floating)하는 상태를 나타낸 도면

도 5는 본 발명의 일 실시예에서 제1 손실 함수(L_{CPT})에 따라 학습된 이후의 프롬프트 풀에서 키 벡터의 특징 공간에서의 위치를 나타낸 도면

발명을 실시하기 위한 구체적인 내용

[0057] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며

본 발명은 이에 제한되지 않는다.

- [0058] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.
- [0059] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.
- [0061] 도 1은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술된 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0062] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 인공지능 기반의 분류 모델을 학습하기 위한 장치일 수 있다. 여기서, 분류 모델은 특정 태스크가 주어지는 경우, 태스크의 데이터들의 클래스를 분류하기 위한 모델일 수 있다.
- [0063] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0064] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0065] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0066] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.

- [0068] 컴퓨팅 장치(12)는 인공지능 기반의 분류 모델을 학습하기 위한 학습 데이터 셋을 생성할 수 있다. 도 2는 본 발명의 일 실시예에서 분류 모델을 학습하기 위한 학습 데이터 셋을 생성하는 상태를 나타낸 도면이다.
- [0069] 컴퓨팅 장치(12)는 현실의 데이터에서 발생하는 불분명한 작업의 경계와 불규칙하게 변화하는 데이터들의 분포를 효과적으로 모사하기 위해, 학습 데이터 셋 내의 모든 데이터를 기 설정된 제1 비율(도 2에서는 Disjoint Class Ratio로 표시됨)에 따라 제1 타입의 데이터와 제2 타입의 데이터로 구분할 수 있다. 즉, 제1 비율은 모든 학습 데이터가 제1 타입의 데이터와 제2 타입의 데이터로 분류되는 데이터 간의 비율일 수 있다.
- [0070] 여기서, 제1 타입의 데이터는 작업(태스크)의 경계를 흐리는데 사용할 데이터(Blurry Classes)로서, 여러 개의 태스크에 걸쳐서 등장하게 되는 데이터일 수 있다. 또한, 제2 타입의 데이터는 하나의 태스크에만 등장하는 데이터(Disjoint Classes)로서 다른 태스크에서는 소멸하게 되는 데이터일 수 있다. 컴퓨팅 장치(12)는 제1 타입의 데이터 및 제2 타입의 데이터를 기 설정된 제1 비율에 따라 각 태스크에 무작위로 할당할 수 있다. 제1 타입의 데이터가 할당된 태스크는 Blurry Task(T^B)로 지칭될 수 있고, 제2 타입의 데이터가 할당된 태스크는 Disjoint Task(T^D)로 지칭될 수 있다.
- [0071] 컴퓨팅 장치(12)는 제1 타입의 데이터에 대해 기 설정된 제2 비율에 따라 각 작업에 무작위로 나누어 분배할 수 있다. 즉, 제1 타입의 데이터는 여러 개의 태스크에 걸쳐서 등장하게 되는 데이터이므로, 기 설정된 제2 비율에 따라 각 작업(태스크)에 나누어 분배할 수 있다. 제2 비율은 각 태스크에서 제1 타입의 데이터가 메이저 클래스(major class)로 분류되는 데이터와 마이너 클래스(minor class)로 분류되는 데이터 간의 비율일 수 있다.
- [0072] 컴퓨팅 장치(12)는 제1 타입의 데이터와 제2 타입의 데이터를 합하여 각 태스크에 대한 학습 데이터(T_1^{B+D} , T_2^{B+D} , ..., T_N^{B+D})로 사용할 수 있다. 이러한 학습 데이터는 현실에서 발생하는 불분명한 작업의 경계와 불규칙하게 변화하는 데이터들의 분포를 모사하게 되는 바, 해당 학습 데이터로 학습된 분류 모델은 추론 단계에서 현실의 데이터가 입력되었을 때 이를 효과적으로 분류할 수 있게 된다.
- [0074] 도 3은 본 발명의 일 실시예에서 인공지능 기반의 분류 모델을 학습하는 과정을 나타낸 도면이다.
- [0075] 도 3을 참조하면, 컴퓨팅 장치(12)는 학습 데이터 셋을 이용하여 분류 모델(100)을 학습할 수 있다. 분류 모델(100)은 인공지능 기반의 분류 모델로서, 예를 들어 비전 트랜스포머(Vision Transformer)일 수 있으나, 분류 모델(100)이 이에 한정되는 것은 아니며 그 이외의 다양한 종류 및 형태의 분류 모델이 사용될 수 있음은 물론이다.
- [0076] 이하에서는, 분류 모델(100)이 비전 트랜스포머인 경우를 일 예로 설명하기로 한다. 분류 모델(100)이 비전 트랜스포머인 경우, 학습 데이터는 이미지 일 수 있다. 컴퓨팅 장치(12)는 기 설정된 프롬프트 풀(prompt pool)에서 하나 이상의 프롬프트를 추출하고, 추출한 프롬프트를 분류 모델(100)의 일부 레이어에 입력하여 분류 모델(100)을 학습할 수 있다.
- [0077] 구체적으로, 컴퓨팅 장치(12)는 학습 데이터로부터 특징(feature)(A)을 추출하고, 추출한 특징(A)에 기초하여 프롬프트 풀(102)에서 하나 이상의 프롬프트를 선택할 수 있다. 예를 들어, 컴퓨팅 장치(12)는 학습 데이터를 토큰화(tokenization) 한 후, 분류 모델(100)에 입력하여 특징을 추출할 수 있으나, 이에 한정되는 것은 아니며 학습 데이터로부터 특징을 추출하기 위한 별도의 모델을 이용할 수도 있다.
- [0078] 프롬프트 풀(102)에는 복수 개의 키 벡터(key vector)들이 프롬프트와 각각 매칭되어 저장될 수 있다. 프롬프트(prompt)는 일종의 가중치(weight)로서 분류 모델(100)의 신경망이 갖는 가중치와는 별도의 가중치로서 분류 모델(100)을 학습하는 과정에서 학습되는 가중치일 수 있다.
- [0079] 컴퓨팅 장치(12)는 학습 데이터에서 추출한 특징과 프롬프트 풀(102)에 저장된 키 벡터들 간의 유사도에 기초하여 하나 이상의 키 벡터를 선택하고, 선택된 키 벡터에 대응하는 프롬프트(102a)를 추출할 수 있다.
- [0080] 한편, 학습 데이터에서 추출한 특징과 프롬프트 풀(102)에 저장된 키 벡터 간의 유사도에 기초하여 키 벡터를 선택하게 되면, 학습이 진행될수록 키 벡터들이 특징 공간(feature space)에서 이동하게 되므로, 학습 시 선택된 키 벡터가 다시 선택되지 못하고 망각되는 현상이 발생할 수 있다.
- [0081] 도 4는 본 발명의 일 실시예에 따른 프롬프트 풀(102)에서 키 벡터들이 선택됨에 따라 특징 공간에서 표류

(floating)하는 상태를 나타낸 도면이다. 도 4에서 키 벡터들은 삼각형 모양으로 도시되었다. 도 4에서와 같이, 학습 데이터에서 추출한 특징과 프롬프트 풀(102)에 저장된 키 벡터 간의 유사도에 기초하여 키 벡터를 선택하게 되면, 학습이 진행될수록 키 벡터들이 특징 공간(feature space)에서 이동하게 되는 것을 볼 수 있다.

[0082] 이에, 컴퓨팅 장치(12)는 프롬프트 풀(102)의 키 벡터들이 학습이 될수록 특징 공간에서 움직임이 둔해지도록 하면서 키 벡터들 간의 거리가 유지되도록 학습할 수 있다. 일 실시예에서, 컴퓨팅 장치(12)는 추출된 특징과 키 벡터가 유사한 것은 그 거리가 가까워지도록 하고, 추출된 특징과 키 벡터가 비유사한 것은 상호 간 거리가 멀어지도록 학습할 수 있다.

[0083] 여기서, 추출된 특징과 키 벡터 간의 유사 여부는 특징과 키 벡터 간 코사인 거리(cosine distance)에 의해 판단할 수 있다. 특징과 키 벡터 간 코사인 거리가 기 설정된 임계 값 이상이면 특징과 키 벡터가 비유사한 것으로 판단할 수 있고, 특징과 키 벡터 간 코사인 거리가 기 설정된 임계 값 미만이면 특징과 키 벡터가 유사한 것으로 판단할 수 있다.

[0084] 프롬프트 풀(102)의 키 벡터들이 학습이 될수록 특징 공간에서 움직임이 둔해지도록 하면서 키 벡터들 간의 거리가 유지되도록 하기 위한 제1 손실 함수($\mathcal{L}_{CVPT}$)는 하기 수학적 식 1로 나타낼 수 있다.

[0085] (수학적 식 1)

$$\mathcal{L}_{CVPT} = -\log \frac{s_n}{s_p + s_n}$$

$$s_n = \sum_{p=1}^P \sum_{q=1}^P \exp(\delta(\mathbf{k}_p, \mathbf{k}_q) / (C_p + 1))$$

$$s_p = \sum_{p=1}^P \sum_{q=1}^B \exp(\delta(\mathbf{k}_p, \mathbf{q}_q) / (C_p + 1))$$

[0088]

[0089] P : 프롬프트 풀의 사이즈

[0090] B : 학습 데이터의 개수

[0091] δ : 코사인 거리

[0092] $\mathbf{k}_p$: p번째 프롬프트에 대응하는 키 벡터

[0093] $\mathbf{q}_q$: q번째 특징

[0094] C_p : p번째 프롬프트가 선택되는 횟수

[0095] 도 5는 본 발명의 일 실시예에서 제1 손실 함수($\mathcal{L}_{CVPT}$)에 따라 학습된 이후의 프롬프트 풀(102)에서 키 벡터의 특징 공간에서의 위치를 나타낸 도면이다. 도 5를 참조하면, 키 벡터들이 선택된 이후에도 특징 공간(feature space)에서 골고루 분산되며 거의 이동하지 않는 것을 볼 수 있다. 이에 따라, 키 벡터를 선택하는 과정에서 발생할 수 있는 망각 현상의 문제를 방지할 수 있게 된다.

[0096] 프롬프트 풀(102)에서 선택된 키 벡터에 대응하는 프롬프트(102a)는 분류 모델(100)의 일부 레이어로 직접 입력될 수 있다. 즉, 학습 데이터는 토큰화 된 후 분류 모델(100)의 입력단으로 입력될 수 있다. 학습 데이터에 대한 CLS(Classification) 토큰이 학습 데이터의 토큰들과 함께 분류 모델(100)의 입력단으로 입력될 수 있다. 이때, 선택된 프롬프트(102a)가 분류 모델(100)의 일부 레이어로 직접 입력되어 학습될 수 있다. 프롬프트(102a)가 입력되는 레이어는 미리 설정된 레이어일 수 있다. 일 실시예에서, 프롬프트(102a)는 분류 모델(100)의 첫 번째 레이어 및 두 번째 레이어로 직접 입력될 수 있으나, 이에 한정되는 것은 아니다.

[0097] 여기서, 분류 모델(100)은 출력단에서 학습 데이터에 대한 CLS 토큰의 특징(B)을 출력하게 된다. CLS 토큰의 특징(B)은 분류기(Classifier)를 통과하여 CLS 토큰의 출력 분포(Logit)를 생성하게 된다.

- [0098] 컴퓨팅 장치(12)는 기 설정된 마스크-프롬프트 풀(Mask-Prompt Pool)(104)에서 기 선택된 프롬프트(102a)와 대응되는 마스크(104a)를 추출할 수 있다. 즉, 마스크-프롬프트 풀(104)에는 복수 개의 프롬프트와 마스크가 상호 대응하여 저장될 수 있으며, 컴퓨팅 장치(12)는 프롬프트 풀(102)에서 선택된 프롬프트(102a)에 대응하는 마스크(104a)를 마스크-프롬프트 풀(104)에서 추출할 수 있다. 프롬프트 풀(102)과 마스크-프롬프트 풀(104)은 하나의 풀로 이루어질 수 있다. 이 경우, 풀 내에는 키 벡터, 프롬프트, 및 마스크가 각각 매칭되어 저장될 수 있다.
- [0099] 컴퓨팅 장치(12)는 CLS 토큰의 출력 분포와 추출된 마스크(104a)를 아다마르곱하여 CLS 토큰의 마스크 된 출력 분포(Masked Logit)를 생성할 수 있다. 분류 모델(100)은 CLS 토큰의 마스크 된 출력 분포(Masked Logit)에 기초하여 학습 데이터에 대한 분류를 수행할 수 있다.
- [0100] 개시되는 실시예에서, 학습 데이터는 명시적인 작업 경계가 주어지지 않으므로, 인공지능 모델을 학습하는데 주로 사용되는 교차 엔트로피(cross entropy) 손실 함수를 위한 비교 대상이 없어 인공지능 모델에 대해 치명적 망각 현상을 유발하게 된다.
- [0101] 이에, 개시되는 실시예에서는 선택된 프롬프트와 대응하는 마스크(104a)를 CLS 토큰의 출력 분포에 적용함으로써 이러한 망각 현상을 방지할 수 있다. 즉, 프롬프트(102a)는 프롬프트 풀(102)에서 키 벡터의 선택 과정을 통해 선택됨에 따라 특징 공간의 분포를 대표하게 된다. 여기서, 프롬프트(102a)가 대표하는 특징 공간의 분포와 관계 없는 부분이 마스크를 통해 가려지고 프롬프트(102a)와 관계 있는 부분만 학습되도록 함으로써, 학습 데이터의 작업 경계가 명확하지 않아 발생하게 되는 치명적 망각 현상을 방지할 수 있게 된다.
- [0102] 한편, 불분명한 작업 경계에서는 분류 간의 불균형으로 인해 소수 분류에 대한 학습은 다수 분류에 대한 학습에 의해 무시되어 모든 분류에 대한 지식을 효과적으로 학습하기가 어렵다. 이에, 개시되는 실시예에서는 소수 분류에 대한 학습이 무시되는 문제를 해결하기 위해, 미분 유사도를 기반으로 제2 손실 함수(L_{GSF})를 갖도록 분류 모델(102)을 학습할 수 있다.
- [0103] 구체적으로, 컴퓨팅 장치(12)는 학습 데이터 전체에 대한 평균 그래디언트 벡터(average gradient vector)와 각 학습 데이터의 그래디언트 벡터에 기초하여 각 학습 데이터의 무시 스코어(ignorance score)를 산출할 수 있다. 각 학습 데이터의 무시 스코어($Score_i^{ign}$)는 하기 수식식 2를 통해 산출할 수 있다. 무시 스코어($Score_i^{ign}$)는 분류 모델(100)의 학습 중 다른 학습 데이터가 학습 데이터 i 를 무시하는 정도를 나타낸 것일 수 있다.
- [0104] (수식식 2)
- [0105]
$$Score_i^{ign} = \delta(\nabla W_{y_i}(f(\mathbf{x}_i)), \nabla W_{y_i}(f(\mathcal{B})))$$
- [0106] δ : 코사인 거리(cosine distance)
- [0107] $\mathcal{B}$: 학습 데이터 셋
- [0108] $\mathbf{x}_i$: i 번째 학습 데이터
- [0109] $\nabla W_{y_i}(f(\mathcal{B}))$: 학습 데이터 셋의 평균 그래디언트 벡터
- [0110] $\nabla W_{y_i}(f(\mathbf{x}_i))$: i 번째 학습 데이터의 그래디언트 벡터
- [0111] W_{y_i} : i 번째 학습 데이터의 정답 값에 대응하는 분류 모델의 가중치
- [0112] 즉, 무시 스코어($Score_i^{ign}$)는 i 번째 학습 데이터의 그래디언트 벡터와 학습 데이터 셋의 평균 그래디언트 벡터 간의 유사도에 기초하여 산출할 수 있다. 여기서는, 유사도의 산출 시 코사인 거리(cosine distance)를 사용하는 것을 일 예로 하였으나, 이에 한정되는 것은 아니며 다양한 유사도 산출 기법이 적용될 수 있음은 물론이다.
- [0113] i 번째 학습 데이터의 그래디언트 벡터는 i 번째 학습 데이터를 분류 모델(100)로 입력하는 경우, 분류 모델(100)의 i 번째 학습 데이터에 대한 CLS 토큰의 출력 분포에 기초하여 산출할 수 있다.
- [0114] 학습 데이터 셋의 평균 그래디언트 벡터는 분류 모델(100)의 배치 단위로 학습할 때, 하나의 배치 단위의 학습

데이터 셋에 대한 평균 그래디언트 벡터를 의미할 수 있다. 학습 데이터 셋의 평균 그래디언트 벡터는 수학식 3으로 나타낼 수 있다.

[0115] (수학식 3)

$$\nabla W_{y_i}(f(\mathcal{B})) = \frac{1}{|\mathcal{B}|} \sum_{(\mathbf{x}, y) \in \mathcal{B}} \nabla W_{y_i}(f(\mathbf{x}))$$

[0116] 즉, 학습 데이터 셋의 평균 그래디언트 벡터는 하나의 배치 단위의 학습 데이터들을 분류 모델(100)로 입력하는 경우, 분류 모델(100)의 각 학습 데이터에 대한 CLS 토큰의 출력 분포에 기초한 그래디언트 벡터들을 평균한 값일 수 있다.

[0118] 여기서, 무시 스코어($\text{Score}_i^{\text{ign}}$)가 클수록 분류 모델(100)의 학습 중 해당 학습 데이터에 대한 분류가 무시되어 학습에 잘 반영되지 않는 것을 의미하고, 무시 스코어($\text{Score}_i^{\text{ign}}$)가 작을수록 분류 모델(100)의 학습 중 해당 학습 데이터에 대한 분류가 무시되지 않고 학습에 잘 반영되는 것을 의미할 수 있다.

[0119] 이에, 개시되는 실시예에서는, 각 학습 데이터에 대한 무시 스코어($\text{Score}_i^{\text{ign}}$)를 기반으로 소수 분류에 따른 학습 데이터가 무시되지 않도록 제2 손실 함수($\mathcal{L}_{\text{GSF}}$)를 제안한다. 제2 손실 함수($\mathcal{L}_{\text{GSF}}$)는 하기 수학식 4로 나타낼 수 있다.

[0120] (수학식 4)

$$\mathcal{L}_{\text{GSF}} = \frac{1}{|\mathcal{B}|} \sum_{i=1}^{|\mathcal{B}|} \left(\text{Score}_i^{\text{ign}} \right)^\gamma \cdot \mathcal{L}_{\text{CE}}(\hat{y}_i, y_i)$$

[0122] B : 배치 당 학습 데이터의 개수

[0123] $\mathcal{L}_{\text{CE}}$: 교차 엔트로피 손실

[0124] y_i : i번째 학습 데이터의 분류에 대한 정답 값

[0125] $\hat{y}_i$: i번째 학습 데이터의 분류 모델의 분류에 대한 예측 값

[0126] γ : 기 설정되는 하이퍼 파라미터

[0127] 제2 손실 함수($\mathcal{L}_{\text{GSF}}$)에 의하면, 소수 분류로 인해 무시되는 학습 데이터의 손실을 강조하여 소수 분류로 인한 무시 문제를 완화할 수 있게 된다. 여기서, 무시 스코어($\text{Score}_i^{\text{ign}}$)에 따라 교차 엔트로피 손실($\mathcal{L}_{\text{CE}}$)을 동적으로 조정할 수 있다.

[0128] 또한, 인공지능 모델은 현재 작업(태스크)에서 모든 데이터가 현재 작업의 모든 분류 대상에 대해 균등한 수를 갖도록 정제된 것이 아니므로, 다수 분류 및 새로 접하는 분류에 대하여 과적합(overfitting) 되는 문제가 있다. 이를 완화하기 위해, 컴퓨팅 장치(12)는 각 학습 데이터를 학습할 때 해당 학습 데이터의 학습에 따른 이익을 점수로 수치화하여 해당 학습 데이터의 특징의 크기를 조절할 수 있다.

[0129] 구체적으로, 컴퓨팅 장치(12)는 각 학습 데이터에 대해 마진 이익 스코어(marginal benefit score)(Score^{MB})를 산출할 수 있다. 마진 이익 스코어(Score^{MB})는 각 학습 데이터가 분류 모델(100)의 학습에 얼마나 이익이 되는지를 나타내는 정도일 수 있다.

[0130] 컴퓨팅 장치(12)는 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대한 분류기의 가중치에 기초하여 마진 이익 스코어(Score^{MB})를 산출할 수 있다. 각 학습 데이터에 대한 마진 이익 스코어(Score^{MB})는 하기 수학식 5를 통해 산출할 수 있다.

[0131] (수학식 5)

$$\text{Score}_i^{MB} = \delta(f(\mathbf{x}_i), W_{y_i}) + m$$

[0133] δ : 코사인 거리(cosine distance)

[0134] W_{y_i} : i번째 학습 데이터의 정답 값에 대응하는 분류기의 가중치

[0135] x_i : i번째 학습 데이터

[0136] $f(x_i)$: i번째 학습 데이터의 CLS 토큰의 특징

[0137] m : 기 설정되는 마진(margin)

[0138] 즉, 마진 이익 스코어(Score^{MB})는 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대응하는 분류기의 가중치 간의 유사도에 기초하여 산출할 수 있다. 여기서는, 유사도의 산출 시 코사인 거리(cosine distance)를 사용하는 것을 일 예로 하였으나, 이에 한정되는 것은 아니며 다양한 유사도 산출 기법이 적용될 수 있음은 물론이다.

[0139] 여기서, 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대응하는 분류기의 가중치가 유사할수록 마진 이익 스코어(Score^{MB})가 작게 나오고, 각 학습 데이터의 CLS 토큰의 특징과 해당 학습 데이터의 정답 값에 대응하는 분류기의 가중치가 비 유사할수록 마진 이익 스코어(Score^{MB})가 크게 된다.

[0140] 컴퓨팅 장치(12)는 각 학습 데이터에 대한 마진 이익 스코어(Score^{MB})에 따라 해당 학습 데이터의 CLS 토큰의 특징 크기를 수학식 6과 같이 조절할 수 있다.

[0141] (수학식 6)

$$\mathbf{h}_i = \frac{f(\mathbf{x}_i)}{\text{Score}_i^{MB}}$$

[0142] $\mathbf{h}_i$: 크기가 조절된 i번째 학습 데이터의 CLS 토큰의 특징

[0144] 즉, 마진 이익 스코어(Score^{MB})가 클수록 해당 학습 데이터의 CLS 토큰의 특징 크기가 작도록 조절하고, 마진 이익 스코어(Score^{MB})가 작을수록 해당 학습 데이터의 CLS 토큰의 특징 크기가 크도록 조절할 수 있다.

[0145] 한편, 분류 모델(100)은 입력되는 학습 데이터에 대해 분류를 수행하게 되는 바, 분류에 따른 제3 손실 함수(L_{CE})를 포함하게 된다. 제3 손실 함수(L_{CE})는 분류 모델(100)에서 출력하는 예측 값($\hat{y}_i$)과 정답 값(y_i) 간의 차이를 최소화 하기 위한 것일 수 있다. 따라서, 분류 모델(100)을 학습하기 위한 전체 손실 함수(L_{total})는 아래 수학식 7로 나타낼 수 있다.

[0146] (수학식 7)

$$\mathcal{L}_{total} = (1 - \alpha)\mathcal{L}_{CE} + \alpha\mathcal{L}_{GSF} + \mathcal{L}_{CVPT}$$

[0148] α : 기 설정되는 하이퍼 파라미터

[0149] L_{CVPT} : 제1 손실 함수

[0150] L_{GSF} : 제2 손실 함수

[0151] L_{CE} : 제3 손실 함수

[0152] 개시되는 실시예에 의하면, 작업의 경계가 불분명하고 불균일한 특성을 갖는 현실 세계의 데이터에 대해 분류를

효과적으로 수행할 수 있게 된다. 또한, 분류 모델의 학습 시 치명적 망각 현상을 방지하고, 분류의 불균형 문제를 방지하여 각 학습 데이터를 균형적으로 학습할 수 있게 된다.

[0154] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

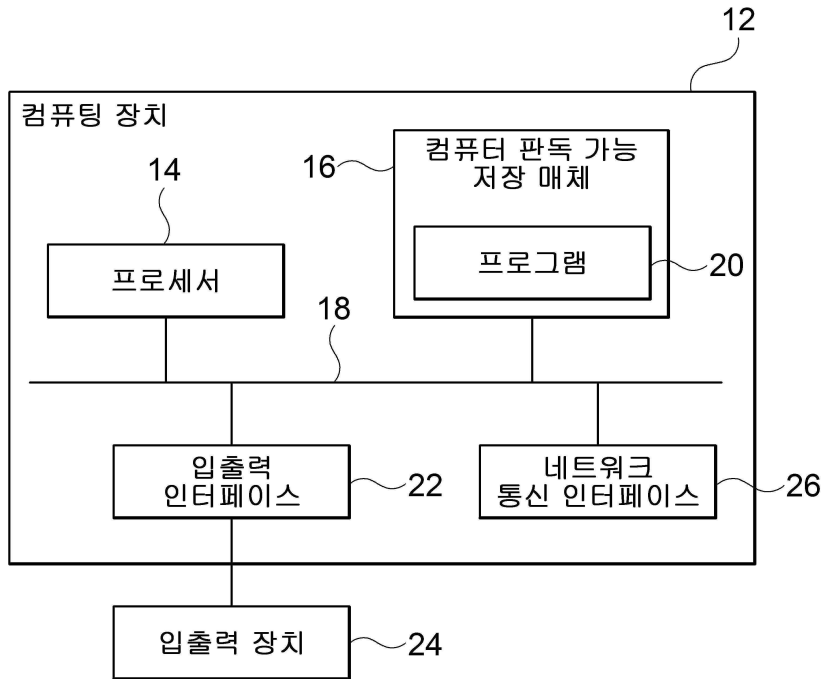
부호의 설명

[0156] 10 : 컴퓨팅 환경
12 : 컴퓨팅 장치
14 : 프로세서
16 : 컴퓨터 판독 가능 저장 매체
18 : 통신 버스
20 : 프로그램
22 : 입출력 인터페이스
24 : 입출력 장치
26 : 네트워크 통신 인터페이스
100 : 분류 모델
102 : 프롬프트 풀
104 : 마스크-프롬프트 풀

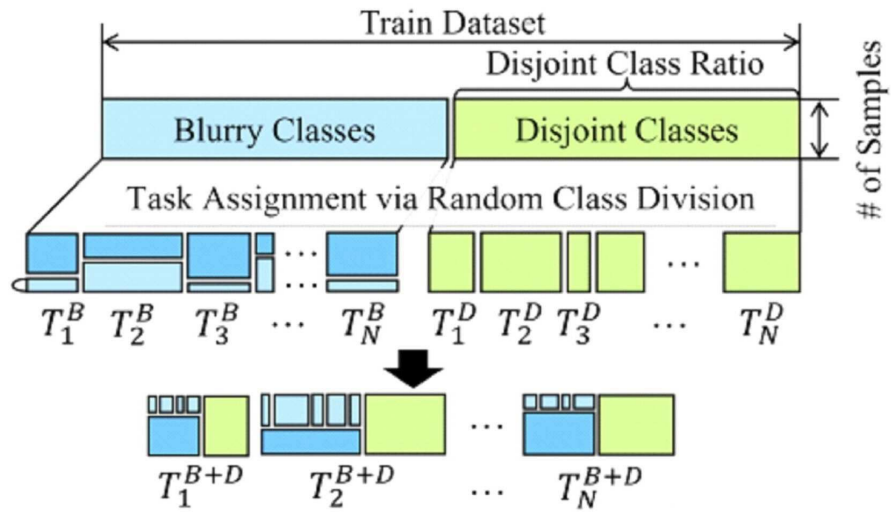
도면

도면1

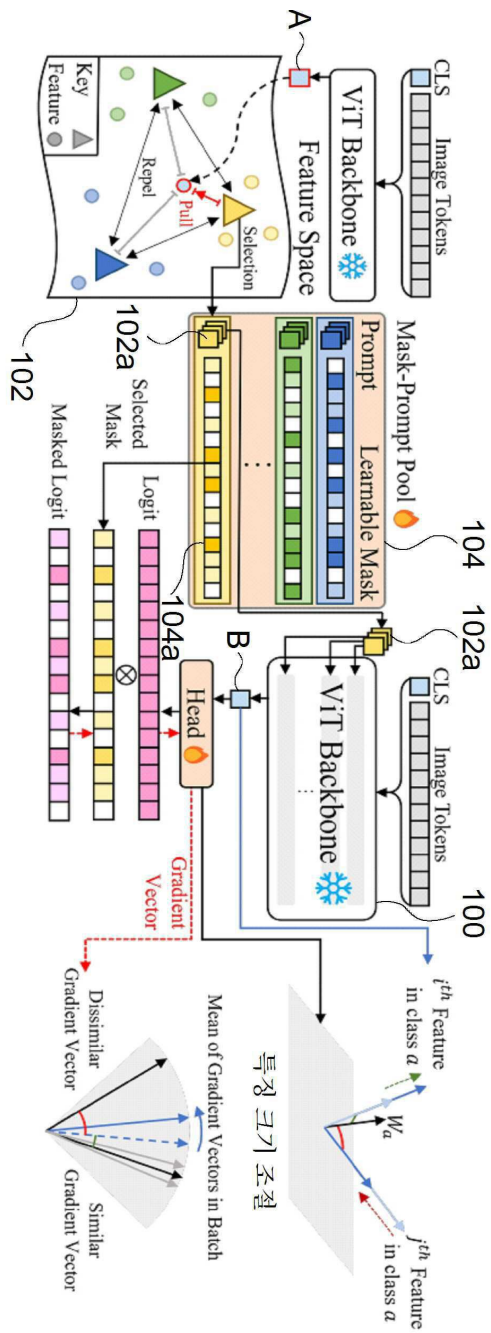
10



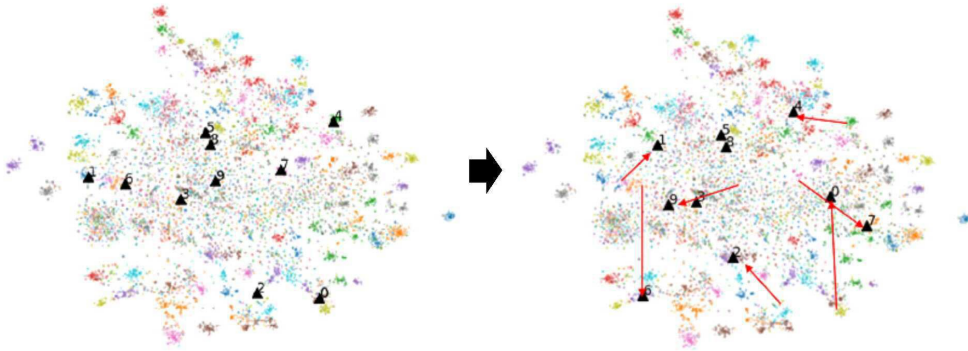
도면2



도면3



도면4



도면5

