(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0061190
(43) 공개일자 2025년05월08일

(51) 국제특허분류(Int. Cl.)
*G06V 10/764* (2022.01) *G06V 10/774* (2022.01)
*G06V 10/82* (2022.01)
(52) CPC특허분류
*G06V 10/764* (2023.08)
*G06N 3/09* (2023.01)
(21) 출원번호 10-2023-0145315
(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
**박건희**
서울특별시 강남구 봉은사로51길 40, 102호
(74) 대리인
**두호특허법인**

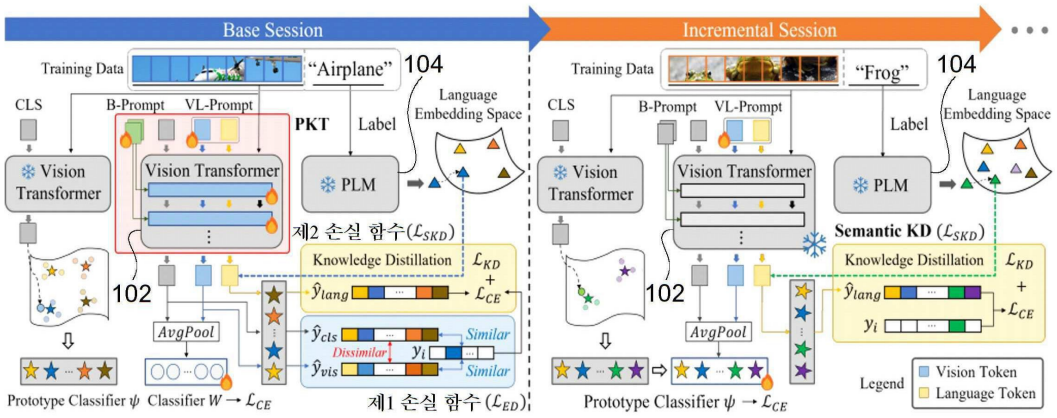
전체 청구항 수 : 총 19 항

(54) 발명의 명칭 **소수 샷 연속 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치**

## (57) 요 약

소수 샷 연속 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치가 개시된다. 개시되는 일 실시예에 따른 방법은, 하나 이상의 프로세서들, 및 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 소수 샷 연속 학습 방법으로서, 소수 샷 연속 학습 방법은, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하는 것으로, 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며, 기본 세션에서, 이미지로 구성되는 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 객체 분류 모델로 입력하는 단계를 포함한다.

## 대 표 도



(52) CPC특허분류
*G06N 3/096* (2023.01)
*G06V 10/774* (2023.08)
*G06V 10/82* (2022.01)

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

# 명 세 서

## 청구범위

### 청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 소수 샷 연속 학습 방법으로서,

상기 소수 샷 연속 학습 방법은, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하는 것으로, 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며,

상기 기본 세션에서, 이미지로 구성되는 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계를 포함하는, 소수 샷 연속 학습 방법.

### 청구항 2

청구항 1에 있어서,

상기 제1 프롬프트 및 상기 제2 프롬프트는,

상기 객체 분류 모델의 신경망이 갖는 가중치와는 별도의 가중치로서 상기 소수 샷 연속 학습에서 학습되는 가중치이고,

제1 프롬프트는 기본 프롬프트(Base-Prompt)이고, 제2 프롬프트는 비주얼 랭귀지 프롬프트(Visual Language-Prompt)인, 소수 샷 연속 학습 방법.

### 청구항 3

청구항 2에 있어서,

상기 제1 프롬프트는, 하나 이상의 토큰을 포함하고,

상기 제2 프롬프트는, 비전 토큰(Vision Token)과 랭귀지 토큰(Language Token)을 포함하며,

상기 객체 분류 모델로 입력하는 단계는,

상기 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계;

상기 학습 데이터에 대한 CLS 토큰(Classification token)을 상기 객체 분류 모델의 입력단으로 입력하는 단계;

상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및

상기 제1 프롬프트를 상기 객체 분류 모델의 일부 레이어로 직접 입력하는 단계를 포함하는, 소수 샷 연속 학습 방법.

### 청구항 4

청구항 3에 있어서,

상기 소수 샷 연속 학습 방법은,

상기 객체 분류 모델의 출력단에서 상기 CLS 토큰의 특징, 상기 비전 토큰의 특징, 및 상기 랭귀지 토큰의 특징을 각각 출력하는 단계;

상기 CLS 토큰의 특징 및 상기 비전 토큰의 특징을 각각 분류기를 통과하도록 하여 CLS 토큰의 출력 분포 및 비

전 토큰의 출력 분포를 생성하는 단계; 및

상기 CLS 토큰의 출력 분포, 상기 비전 토큰의 출력 분포, 및 상기 학습 데이터에 대한 정답 값의 분포에 기초하여 기 설정된 제1 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함하는, 소수 샷 연속 학습 방법.

**청구항 5**

청구항 4에 있어서,

상기 객체 분류 모델을 학습하는 단계는,

상기 CLS 토큰의 출력 분포와 상기 비전 토큰의 출력 분포가 유사하지 않도록 하면서 상기 CLS 토큰의 출력 분포 및 상기 비전 토큰의 출력 분포가 각각 상기 정답 값의 분포와 유사하도록 하는, 소수 샷 연속 학습 방법.

**청구항 6**

청구항 5에 있어서,

상기 제1 손실 함수($L_{ED}$)는 하기 수학식으로 표현되는, 소수 샷 연속 학습 방법.

(수학식)

$$\mathcal{L}_{ED} = -\log\left(\frac{\mathcal{L}_{KL}(\delta(\hat{y_i}^{vis}), \delta(\hat{y_i}^{cls}))}{\mathcal{L}_{CE}(\hat{y_i}^{vis}, y_i) + \mathcal{L}_{CE}(\hat{y_i}^{cls}, y_i) + \mathcal{L}_{KL}(\delta(\hat{y_i}^{vis}), \delta(\hat{y_i}^{cls}))}\right)$$

$L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

$L_{KL}$ : Kullback-Leibler divergence 손실

$\delta(\cdot)$ : 소프트맥스 함수(softmax function)

$\hat{y}_i^{cls}$ : CLS 토큰의 출력 분포

$\hat{y}_i^{vis}$ : 비전 토큰의 출력 분포

$y_i$: 정답 값의 분포

**청구항 7**

청구항 4에 있어서,

상기 소수 샷 연속 학습 방법은,

상기 기본 세션에서, 상기 학습 데이터의 레이블을 사전 학습된 언어 모델에 입력하여 상기 레이블의 워드 임베딩 특징을 추출하는 단계; 및

상기 레이블의 워드 임베딩 특징을 상기 랭귀지 토큰의 특징으로 지식 증류(knowledge distillation)하여 지식을 전달하는 단계를 더 포함하는, 소수 샷 연속 학습 방법.

**청구항 8**

청구항 7에 있어서,

상기 소수 샷 연속 학습 방법은,

상기 기본 세션에서, 상기 지식 증류된 랭귀지 토큰의 특징을 분류기에 통과시켜 랭귀지 토큰의 출력 분포를 생성하는 단계; 및

상기 랭귀지 토큰의 출력 분포와 상기 정답 값의 분포 간의 차이가 최소화 되는 제2 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함하는, 소수 샷 연속 학습 방법.

**청구항 9**

청구항 8에 있어서,

상기 제2 손실 함수($L_{SKD}$)는 하기 수학식으로 표현되는, 소수 샷 연속 학습 방법.

(수학식)

$$\mathcal{L}_{SKD} = \mathcal{L}_{KD}(f_i^{lang}, w_{c(y_i)}) + \lambda \cdot \mathcal{L}_{CE}(\hat{y}_i^{lang}, y_i)$$

$L_{KD}$ : 지식 증류(Knowledge Distillation)를 위한 손실

$L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

$w_{c(yi)}$ : 레이블의 워드 임베딩 특징

$f_i^{lang}$ : 랭귀지 토큰의 특징

λ : 기 설정된 하이퍼 파라미터

$\hat{y}_i^{lang}$ : 랭귀지 토큰의 출력 분포

$y_i$: 정답 값의 분포

**청구항 10**

청구항 8에 있어서,

상기 기본 세션은, 하기 수학식에 따른 총 손실 함수($L_{Total}$)를 갖도록 상기 객체 분류 모델을 학습하는, 소수 샷 연속 학습 방법.

(수학식)

$$\mathcal{L}_{Total} = \mathcal{L}_{CE}(\hat{y}_i, y_i) + \alpha \cdot \mathcal{L}_{ED} + \beta \cdot \mathcal{L}_{SKD}$$

$L_{ED}$ : 제1 손실 함수

$L_{SKD}$ : 제2 손실 함수

$\mathcal{L}_{CE}(\hat{y}_i, y_i)$: 객체 분류에 따른 제3 손실 함수

α : 제1 손실 함수에 대한 스케일링 팩터

β : 제2 손실 함수에 대한 스케일링 팩터

**청구항 11**

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, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하되, 상기 소수 샷 연속 학습은 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며,

상기 하나 이상의 프로그램들은,

상기 기본 세션에서, 이미지로 구성되는 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한 명령을 포함하는, 컴퓨팅 장치.

**청구항 12**

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 소수 샷 연속 학습 방법으로서,

상기 소수 샷 연속 학습 방법은, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하는 것으로, 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며,

상기 기본 세션에서, 이미지로 구성되는 제1 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계; 및

상기 증분 세션에서, 이미지로 구성되는 제2 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계를 포함하는, 소수 샷 연속 학습 방법.

**청구항 13**

청구항 12에 있어서,

상기 제1 프롬프트 및 상기 제2 프롬프트는,

상기 객체 분류 모델의 신경망이 갖는 가중치와는 별도의 가중치로서 상기 소수 샷 연속 학습에서 학습되는 가중치이고,

제1 프롬프트는 기본 프롬프트(Base-Prompt)이고, 제2 프롬프트는 비주얼 랭귀지 프롬프트(Visual Language-Prompt)인, 소수 샷 연속 학습 방법.

**청구항 14**

청구항 13에 있어서,

상기 제1 프롬프트는, 하나 이상의 토큰을 포함하고,

상기 제2 프롬프트는, 비전 토큰(Vision Token)과 랭귀지 토큰(Language Token)을 포함하며,

상기 기본 세션에서, 상기 객체 분류 모델로 입력하는 단계는,

상기 제1 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계;

상기 제1 학습 데이터에 대한 CLS 토큰(Classification token)을 상기 객체 분류 모델의 입력단으로 입력하는

단계;

상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및

상기 제1 프롬프트를 상기 객체 분류 모델의 일부 레이어로 직접 입력하는 단계를 포함하고,

상기 증분 세션에서, 상기 객체 분류 모델로 입력하는 단계는,

상기 제2 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계;

상기 제2 학습 데이터에 대한 CLS 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및

상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계를 포함하는, 소수 샷 연속 학습 방법.

**청구항 15**

청구항 14에 있어서,

상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계는,

상기 제1 프롬프트 및 상기 제1 프롬프트가 직접 입력되는 상기 일부 레이어를 동결한 상태에서, 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는, 소수 샷 연속 학습 방법.

**청구항 16**

청구항 14에 있어서,

상기 소수 샷 연속 학습 방법은,

상기 증분 세션에서, 상기 객체 분류 모델의 출력단에서 상기 랭귀지 토큰의 특징을 출력하는 단계;

상기 증분 세션에서, 상기 제2 학습 데이터의 레이블을 사전 학습된 언어 모델에 입력하여 상기 레이블의 워드 임베딩 특징을 추출하는 단계; 및

상기 레이블의 워드 임베딩 특징을 상기 랭귀지 토큰의 특징으로 지식 증류(knowledge distillation)하여 지식을 전달하는 단계를 더 포함하는, 소수 샷 연속 학습 방법.

**청구항 17**

청구항 16에 있어서,

상기 소수 샷 연속 학습 방법은,

상기 증분 세션에서, 상기 지식 증류된 랭귀지 토큰의 특징을 분류기에 통과시켜 랭귀지 토큰의 출력 분포를 생성하는 단계; 및

상기 랭귀지 토큰의 출력 분포와 상기 제2 학습 데이터에 대한 정답 값의 분포 간의 차이가 최소화 되는 제2 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함하는, 소수 샷 연속 학습 방법.

**청구항 18**

청구항 17에 있어서,

상기 증분 세션은, 하기 수학식에 따른 총 손실 함수($L_{Total}$)를 갖도록 상기 객체 분류 모델을 학습하는, 소수 샷 연속 학습 방법.

(수학식)

$$L_{Total} = L_{CE}(\hat{y}_i, y_i) + \beta \cdot L_{SKD}$$

$L_{SKD}$ : 제2 손실 함수

$\mathcal{L}_{CE}(\hat{y}_i, y_i)$: 객체 분류에 따른 제3 손실 함수

β : 제2 손실 함수에 대한 스케일링 팩터

**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, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하되, 상기 소수 샷 연속 학습은 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며,

상기 하나 이상의 프로그램들은,

상기 기본 세션에서, 이미지로 구성되는 제1 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한 명령; 및

상기 증분 세션에서, 이미지로 구성되는 제2 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한 명령을 포함하는, 컴퓨팅 장치.

## 발명의 설명

### 기 술 분 야

[0001] 본 발명의 실시예는 소수 샷 연속 학습 기술과 관련된다.

### 배 경 기 술

[0003] 사람은 기존에 가지고 있는 배경지식을 활용하여 적은 양의 데이터를 통해 이전에 경험하지 못한 새로운 개념을 빠르게 학습할 수 있다. 소수 샷(Few-Shot) 연속 학습은 이러한 사람의 학습 방식을 모방하여 적은 양의 데이터를 통해 효과적인 학습을 연속적으로 수행할 수 있도록 하는 것을 목적으로 한다. 소수 샷 연속 학습 방식은 기본 세션과 증분 세션으로 이루어진다. 기본 세션에서 인공지능 모델은 다수의 카테고리를 충분한 데이터를 기반으로 학습하게 되며, 이후 연속적으로 학습하게 되는 증분 세션에서 인공지능 모델은 적은 수의 카테고리를 소량의 데이터만을 통해 학습하게 된다.

[0004] 소수 샷 연속 학습은 데이터를 확보하기 어렵고 장기간 데이터를 저장하는 데 어려움이 있는 의료 분야 또는 군사 분야 등에서 효과적으로 인공지능 모델을 학습할 수 있는 학습 방식이다. 그러나, 소수 샷 연속 학습은 인공지능 모델이 연속적으로 학습되는 과정에서 이전에 학습한 지식을 망각하는 파괴적 망각(catastrophic forgetting) 현상과 인공지능 모델의 방대한 가중치가 소량의 데이터에 과적합(overfitting) 되는 문제를 해결하기 위한 기술을 필요로 한다.

## 선행기술문헌

### 특허문헌

[0006] (특허문헌 0001) 한국등록특허공보 제10-2316678호(2021.10.25)

## 발명의 내용

### 해결하려는 과제

[0007] 본 발명의 실시예는 파과적 망각 현상 및 과적합 문제를 해결할 수 있는 소수 샷 연속 학습 방법 및 이를 수행하기 위한 컴퓨팅 장치를 제공하기 위한 것이다.

### 과제의 해결 수단

[0009] 개시되는 일 실시예에 따른 소수 샷 연속 학습 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 소수 샷 연속 학습 방법으로서, 상기 소수 샷 연속 학습 방법은, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하는 것으로, 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며, 상기 기본 세션에서, 이미지로 구성되는 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계를 포함한다.

[0010] 상기 제1 프롬프트 및 상기 제2 프롬프트는, 상기 객체 분류 모델의 신경망이 갖는 가중치와는 별도의 가중치로서 상기 소수 샷 연속 학습에서 학습되는 가중치이고, 제1 프롬프트는 기본 프롬프트(Base-Prompt)이고, 제2 프롬프트는 비주얼 랭귀지 프롬프트(Visual Language- Prompt)일 수 있다.

[0011] 상기 제1 프롬프트는, 하나 이상의 토큰을 포함하고, 상기 제2 프롬프트는, 비전 토큰(Vision Token)과 랭귀지 토큰(Language Token)을 포함하며, 상기 객체 분류 모델로 입력하는 단계는, 상기 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계; 상기 학습 데이터에 대한 CLS 토큰(Classification token)을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및 상기 제1 프롬프트를 상기 객체 분류 모델의 일부 레이어로 직접 입력하는 단계를 포함할 수 있다.

[0012] 상기 소수 샷 연속 학습 방법은, 상기 객체 분류 모델의 출력단에서 상기 CLS 토큰의 특징, 상기 비전 토큰의 특징, 및 상기 랭귀지 토큰의 특징을 각각 출력하는 단계; 상기 CLS 토큰의 특징 및 상기 비전 토큰의 특징을 각각 분류기를 통과하도록 하여 CLS 토큰의 출력 분포 및 비전 토큰의 출력 분포를 생성하는 단계; 및 상기 CLS 토큰의 출력 분포, 상기 비전 토큰의 출력 분포, 및 상기 학습 데이터에 대한 정답 값의 분포에 기초하여 기 설정된 제1 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함할 수 있다.

[0013] 상기 객체 분류 모델을 학습하는 단계는, 상기 CLS 토큰의 출력 분포와 상기 비전 토큰의 출력 분포가 유사하지 않도록 하면서 상기 CLS 토큰의 출력 분포 및 상기 비전 토큰의 출력 분포가 각각 상기 정답 값의 분포와 유사하도록 할 수 있다.

[0014] 상기 제1 손실 함수($L_{ED}$)는 하기 수학식으로 표현될 수 있다.

[0015] (수학식)

$$\mathcal{L}_{ED} = -\log\left(\frac{\mathcal{L}_{KL}(\delta(\hat{y_i}^{vis}), \delta(\hat{y_i}^{cls}))}{\mathcal{L}_{CE}(\hat{y_i}^{vis}, y_i) + \mathcal{L}_{CE}(\hat{y_i}^{cls}, y_i) + \mathcal{L}_{KL}(\delta(\hat{y_i}^{vis}), \delta(\hat{y_i}^{cls}))}\right)$$

[0016]

[0017] $L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

[0018] $L_{KL}$ : Kullback-Leibler divergence 손실

[0019] $\delta(\cdot)$ : 소프트맥스 함수(softmax function)

[0020] $\hat{y}_i^{cls}$ : CLS 토큰의 출력 분포

[0021] $\hat{y}_i^{vis}$ : 비전 토큰의 출력 분포

[0022] $y_i$: 정답 값의 분포

[0023] 상기 소수 샷 연속 학습 방법은, 상기 기본 세션에서, 상기 학습 데이터의 레이블을 사전 학습된 언어 모델에 입력하여 상기 레이블의 워드 임베딩 특징을 추출하는 단계; 및 상기 레이블의 워드 임베딩 특징을 상기 랭귀지 토큰의 특징으로 지식 증류(knowledge distillation)하여 지식을 전달하는 단계를 더 포함할 수 있다.

[0024] 상기 소수 샷 연속 학습 방법은, 상기 기본 세션에서, 상기 지식 증류된 랭귀지 토큰의 특징을 분류기에 통과시켜 랭귀지 토큰의 출력 분포를 생성하는 단계; 및 상기 랭귀지 토큰의 출력 분포와 상기 정답 값의 분포 간의 차이가 최소화 되는 제2 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함할 수 있다.

[0025] 상기 제2 손실 함수($L_{SKD}$)는 하기 수학식으로 표현될 수 있다.

[0026] (수학식)

[0027] $$\mathcal{L}_{SKD} = \mathcal{L}_{KD}(f_i^{lang}, w_{c(y_i)}) + \lambda \cdot \mathcal{L}_{CE}(\hat{y}_i^{lang}, y_i)$$

[0028] $L_{KD}$ : 지식 증류(Knowledge Distillation)를 위한 손실

[0029] $L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

[0030] $w_{c(yi)}$ : 레이블의 워드 임베딩 특징

[0031] $f_i^{lang}$ : 랭귀지 토큰의 특징

[0032] $\lambda$ : 기 설정된 하이퍼 파라미터

[0033] $\hat{y}_i^{lang}$ : 랭귀지 토큰의 출력 분포

[0034] $y_i$: 정답 값의 분포

[0035] 상기 기본 세션은, 하기 수학식에 따른 총 손실 함수($L_{Total}$)를 갖도록 상기 객체 분류 모델을 학습할 수 있다.

[0036] (수학식)

[0037] $$\mathcal{L}_{Total} = \mathcal{L}_{CE}(\hat{y}_i, y_i) + \alpha \cdot \mathcal{L}_{ED} + \beta \cdot \mathcal{L}_{SKD}$$

[0038] $L_{ED}$ : 제1 손실 함수

[0039] $L_{SKD}$ : 제2 손실 함수

[0040] $\mathcal{L}_{CE}(\hat{y}_i, y_i)$: 객체 분류에 따른 제3 손실 함수

[0041] $\alpha$ : 제1 손실 함수에 대한 스케일링 팩터

[0042] $\beta$ : 제2 손실 함수에 대한 스케일링 팩터

[0043] 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, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하되, 상기 소수 샷 연속 학습은 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며, 상기 하나 이상의 프로그램들은, 상기 기본 세션에서, 이미지로 구성되

는 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한 명령을 포함한다.

[0044] 개시되는 다른 실시예에 따른 소수 샷 연속 학습 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되는 소수 샷 연속 학습 방법으로서, 상기 소수 샷 연속 학습 방법은, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하는 것으로, 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며, 상기 기본 세션에서, 이미지로 구성되는 제1 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계; 및 상기 증분 세션에서, 이미지로 구성되는 제2 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하는 단계를 포함한다.

[0045] 상기 제1 프롬프트 및 상기 제2 프롬프트는, 상기 객체 분류 모델의 신경망이 갖는 가중치와는 별도의 가중치로서 상기 소수 샷 연속 학습에서 학습되는 가중치이고, 제1 프롬프트는 기본 프롬프트(Base-Prompt)이고, 제2 프롬프트는 비주얼 랭귀지 프롬프트(Visual Language- Prompt)일 수 있다.

[0046] 상기 제1 프롬프트는, 하나 이상의 토큰을 포함하고, 상기 제2 프롬프트는, 비전 토큰(Vision Token)과 랭귀지 토큰(Language Token)을 포함하며, 상기 기본 세션에서, 상기 객체 분류 모델로 입력하는 단계는, 상기 제1 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계; 상기 제1 학습 데이터에 대한 CLS 토큰(Classification token)을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및 상기 제1 프롬프트를 상기 객체 분류 모델의 일부 레이어로 직접 입력하는 단계를 포함하고, 상기 증분 세션에서, 상기 객체 분류 모델로 입력하는 단계는, 상기 제2 학습 데이터를 토큰화 하여 상기 객체 분류 모델의 입력단으로 입력하는 단계; 상기 제2 학습 데이터에 대한 CLS 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계; 및 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계를 포함할 수 있다.

[0047] 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력하는 단계는, 상기 제1 프롬프트 및 상기 제1 프롬프트가 직접 입력되는 상기 일부 레이어를 동결한 상태에서, 상기 비전 토큰 및 랭귀지 토큰을 상기 객체 분류 모델의 입력단으로 입력할 수 있다.

[0048] 상기 소수 샷 연속 학습 방법은, 상기 증분 세션에서, 상기 객체 분류 모델의 출력단에서 상기 랭귀지 토큰의 특징을 출력하는 단계; 상기 증분 세션에서, 상기 제2 학습 데이터의 레이블을 사전 학습된 언어 모델에 입력하여 상기 레이블의 워드 임베딩 특징을 추출하는 단계; 및 상기 레이블의 워드 임베딩 특징을 상기 랭귀지 토큰의 특징으로 지식 증류(knowledge distillation)하여 지식을 전달하는 단계를 더 포함할 수 있다.

[0049] 상기 소수 샷 연속 학습 방법은, 상기 증분 세션에서, 상기 지식 증류된 랭귀지 토큰의 특징을 분류기에 통과시켜 랭귀지 토큰의 출력 분포를 생성하는 단계; 및 상기 랭귀지 토큰의 출력 분포와 상기 제2 학습 데이터에 대한 정답 값의 분포 간의 차이가 최소화 되는 제2 손실 함수를 갖도록 상기 객체 분류 모델을 학습하는 단계를 더 포함할 수 있다.

[0050] 상기 증분 세션은, 하기 수학식에 따른 총 손실 함수($L_{Total}$)를 갖도록 상기 객체 분류 모델을 학습할 수 있다.

[0051] (수학식)

[0052] $$L_{Total} = L_{CE}(\hat{y}_i, y_i) + \beta \cdot L_{SKD}$$

[0053] $L_{SKD}$ : 제2 손실 함수

[0054] $\mathcal{L}_{CE}(\hat{y}_i, y_i)$: 객체 분류에 따른 제3 손실 함수

[0055] β : 제2 손실 함수에 대한 스케일링 팩터

[0056] 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, 사전 학습된 비전 트랜스포머 기반의 객체 분류 모델을 이용하여 소수 샷 연속 학습을 수행하되, 상기 소수 샷 연속 학습은 기본 세션(base session) 및 증분 세션(incremental session)으로 이루어지며, 상기 하나 이상의 프로그램들은, 상기 기본 세션에서, 이미지로 구성되는 제1 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한

명령; 및 상기 증분 세션에서, 이미지로 구성되는 제2 학습 데이터와 함께 기 설정된 제1 프롬프트 및 제2 프롬프트를 상기 객체 분류 모델로 입력하기 위한 명령을 포함한다.

## 발명의 효과

[0058] 개시되는 실시예에 의하면, 사전 학습된 인공지능 모델 및 프롬프트를 활용하여 기본 세션에서 충분한 지식을 학습하고 학습한 지식을 전이하여 증분 세션에서 연속적인 학습을 수행함으로써, 소수 샷 연속 학습에서 큰 규모의 인공지능 모델을 활용할 수 있고 파괴적 망각 현상 및 과적합 문제를 완화할 수 있게 된다.

## 도면의 간단한 설명

[0060] 도 1은 본 발명의 일 실시예에 따른 소수 샷 연속 학습 기법의 전체 과정을 나타낸 도면

도 2는 본 발명의 일 실시예에서 제1 프롬프트 및 제2 프롬프트를 이용하는 기본 세션과 증분 세션의 학습 과정을 개략적으로 나타낸 도면

도 3은 본 발명의 일 실시예에서 제1 손실 함수($L_{ED}$)를 갖도록 학습되는 상태를 나타낸 도면

도 4는 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

## 발명을 실시하기 위한 구체적인 내용

[0061] 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.

[0062] 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.

[0063] 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.

[0065] 도 1은 본 발명의 일 실시예에 따른 소수 샷 연속 학습 기법의 전체 과정을 나타낸 도면이다. 소수 샷 연속 학습 기법은 기본 세션(base session)과 증분 세션(incremental session)으로 이루어진다. 기본 세션의 학습이 이루어진 후 증분 세션의 학습이 이루어질 수 있다. 기본 세션에서는 많은 클래스들이 충분히 많은 학습 데이터를 통해 학습이 이루어질 수 있다. 증분 세션에서는 이전 세션에서 학습한 데이터 없이 현재 세션에 속하는 적은 클래스들이 적은 학습 데이터를 통해 학습이 이루어질 수 있다.

[0066] 개시되는 실시예에서는, 사전 학습된 인공지능 모델(102)을 이용하여 소수 샷 연속 학습을 수행할 수 있다. 즉, 사전 학습된 인공지능 모델(102)을 대상으로 소수 샷 연속 학습이 이루어지게 된다. 여기서, 사전 학습된 인공지능 모델은 이미지에서 객체를 분류하거나 검출하도록 기 학습된 모델일 수 있다. 예를 들어, 사전 학습된 인공지능 모델(102)은 비전 트랜스포머(vision transformer) 기반의 객체 분류 모델일 수 있으나, 이에 한정되는

것은 아니다.

[0067] 소수 샷 연속 학습에서는 기본 세션에서 주어지는 충분한 데이터를 통해 전이할 수 있는 지식을 학습하고 이를 증분 세션에서 활용하는 것이 중요하다. 이를 위해, 개시되는 실시예에서는, 기본 세션 학습 시 제1 프롬프트(prompt)와 제2 프롬프트를 사전 학습된 인공지능 모델(102)의 일부 레이어와 함께 학습할 수 있다. 제1 프롬프트는 기본 프롬프트(Base-Prompt)이고, 제2 프롬프트는 비주얼 랭귀지 프롬프트(Visual Language-Prompt)일 수 있다.

[0068] 도 2는 본 발명의 일 실시예에서 제1 프롬프트(prompt) 및 제2 프롬프트를 이용하는 기본 세션과 증분 세션의 학습 과정을 개략적으로 나타낸 도면이다. 도 2를 참조하면, 학습 데이터(training data)가 사전 학습된 인공지능 모델(102)에 입력될 때, 제1 프롬프트(B-Prompt)와 제2 프롬프트(VL-Prompt)가 인공지능 모델(102)에 함께 입력될 수 있다. 학습 데이터는 하나 이상의 객체를 포함하는 이미지일 수 있다.

[0069] 여기서, 제1 프롬프트(B-Prompt)와 제2 프롬프트(VL-Prompt)는 소수 샷 연속 학습에서 학습되는 별도의 가중치(weight)일 수 있다. 즉, 제1 프롬프트(B-Prompt)와 제2 프롬프트(VL-Prompt)는 인공지능 모델(102)의 신경망이 갖는 가중치와는 별도의 가중치로서 소수 샷 연속 학습을 통해 학습되는 가중치일 수 있다. 제1 프롬프트(B-Prompt)와 제2 프롬프트(VL-Prompt)는 학습 초기에 랜덤하게 생성될 수 있으며, 학습이 진행되면서 업데이트 될 수 있다.

[0070] 제1 프롬프트(B-Prompt)는 인공지능 모델(102)의 일부 레이어로 직접 입력될 수 있다. 일 실시예에서, 인공지능 모델(102)의 일부 레이어는 인공지능 모델(102)의 첫 번째 레이어 및 두 번째 레이어일 수 있으나, 이에 한정되는 것은 아니다. 제1 프롬프트(B-Prompt) 및 인공지능 모델(102)의 일부 레이어는 기본 세션에서만 학습이 되고, 증분 세션에서는 학습되지 않을 수 있다. 제1 프롬프트(B-Prompt)는 기본 세션의 학습을 통해 인공지능 모델(102)의 일부 레이어를 튜닝(tunning)하게 된다. 이때, 제1 프롬프트(B-Prompt)는 인공지능 모델(102)의 일부 레이어에 직접 입력되어 인공지능 모델(102)의 일부 레이어를 튜닝하게 된다.

[0071] 만약, 증분 세션에서도 학습을 수행하게 되면, 제1 프롬프트(B-Prompt) 및 인공지능 모델(102)의 일부 레이어에서 망각(forgetting)이 발생할 수 있기 때문에, 기본 세션의 학습에 따른 지식을 보존하기 위해 증분 세션에서는 이에 대해 학습을 진행하지 않을 수 있다. 이때, 기본 세션의 학습에 따른 지식을 증분 세션에 전달해주게 된다.

[0072] 제1 프롬프트(B-Prompt)는 사전 학습된 인공지능 모델(102)이 기본 세션에서 학습 시 반영하지 못하는 추가적인 정보를 학습하도록 하며, 증분 세션에서는 제1 프롬프트(B-Prompt) 및 인공지능 모델(102)의 일부 레이어를 동결하여 기본 세션에서 학습한 지식을 전이(transfer)하는 것을 목적으로 한다. 제1 프롬프트(B-Prompt)는 하나 이상의 토큰을 포함하도록 생성될 수 있다.

[0073] 학습 데이터와 제2 프롬프트(VL-Prompt)는 인공지능 모델(102)의 입력단으로 각각 입력될 수 있다. 학습 데이터와 제2 프롬프트(VL-Prompt)는 인공지능 모델(102)의 입력단으로 입력되어 인공지능 모델(102)의 모든 레이어를 통과할 수 있다. 학습 데이터는 토큰 형태로 인공지능 모델(102)의 입력단에 입력될 수 있다. 이를 위해, 학습 데이터는 토큰화 과정을 거친 후 인공지능 모델(102)에 입력될 수 있다. 이때, 학습 데이터에 대한 CLS 토큰이 학습 데이터의 토큰들과 함께 인공지능 모델(102)에 입력될 수 있다.

[0074] 또한, 제2 프롬프트(VL-Prompt)는 비전 토큰(Vision Token)과 랭귀지 토큰(Language Token)을 포함할 수 있다. 랭귀지 토큰은 학습 데이터에 부여되는 레이블의 언어적 지식에 기초하여 생성될 수 있다. 비전 토큰은 학습 데이터의 이미지로부터 다양한 정보를 학습하도록 생성될 수 있다.

[0075] 제2 프롬프트(VL-Prompt)는 기본 세션뿐만 아니라 증분 세션에서도 학습이 될 수 있다. 즉, 제2 프롬프트(VL-Prompt)는 기본 세션뿐만 아니라 증분 세션에서도 지속적으로 이전에 학습했던 지식들을 다음 번 태스크를 학습하는데 도움이 될 수 있도록 전달해주는 역할을 할 수 있다.

[0076] 인공지능 모델(102)은 출력단에서 학습 데이터에 대한 CLS 토큰의 특징, 비전 토큰의 특징, 및 랭귀지 토큰의 특징을 출력하게 된다. 또한, 인공지능 모델(102)의 출력단에서는 토큰 형태의 학습 데이터 특징을 출력할 수 있다. 여기서, CLS 토큰의 특징 및 비전 토큰의 특징은 각각 분류기(classifier)를 통과하여 CLS 토큰의 출력 분포($\hat{y}^{cls}$) 및 비전 토큰의 출력 분포($\hat{y}^{vis}$)를 생성하게 된다.

[0077] 여기서, CLS 토큰의 출력 분포($\hat{y}^{cls}$)와 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 서로 유사하지 않도록 하면서 CLS 토큰의

출력 분포($\hat{y}^{cls}$) 및 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 각각 정답 값의 분포($y_i$)와의 차이가 최소화 되도록(유사하도록) 학습될 수 있다.

[0078] 즉, CLS 토큰의 출력 분포($\hat{y}^{cls}$)와 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 서로 유사하지 않도록 하여 CLS 토큰과 비전 토큰이 서로 다른 다양한 지식을 학습하도록 하면서, CLS 토큰의 출력 분포($\hat{y}^{cls}$) 및 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 각각 정답 값의 분포($y_i$)와의 차이가 최소화 되도록 하여 CLS 토큰과 비전 토큰이 각각 정답을 예측할 수 있도록 할 수 있다. 이에 따른 제1 손실 함수($L_{ED}$)는 하기 수학식 1로 나타낼 수 있다. 또한, 도 3은 본 발명의 일 실시예에서 제1 손실 함수($L_{ED}$)를 갖도록 학습되는 상태를 나타낸 도면이다.

[0079] (수학식 1)

[0080]
$$\mathcal{L}_{ED} = -\log\left(\frac{\mathcal{L}_{KL}(\delta(\hat{y}_i{}^{vis}), \delta(\hat{y}_i{}^{cls}))}{\mathcal{L}_{CE}(\hat{y}_i{}^{vis}, y_i) + \mathcal{L}_{CE}(\hat{y}_i{}^{cls}, y_i) + \mathcal{L}_{KL}(\delta(\hat{y}_i{}^{vis}), \delta(\hat{y}_i{}^{cls}))}\right)$$

[0081] $L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

[0082] $L_{KL}$ : Kullback-Leibler divergence 손실

[0083] $\delta(\cdot)$ : 소프트맥스 함수(softmax function)

[0084] 여기서, $L_{CE}$는 CLS 토큰의 출력 분포($\hat{y}^{cls}$) 및 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 각각 정답 값의 분포($y_i$)와의 차이가 최소화 되도록 하기 위한 것과 관련된 손실이고, $L_{KL}$는 CLS 토큰의 출력 분포($\hat{y}^{cls}$)와 비전 토큰의 출력 분포($\hat{y}^{vis}$)가 서로 유사하지 않도록 하기 위한 것과 관련된 손실이다.

[0085] 한편, 기본 세션 및 증분 세션 각각에서, 학습 데이터의 레이블(예를 들어, 기본 세션의 경우 "Airplane", 증분 세션의 경우 "Frog")을 사전 학습된 언어 모델(Pre-trained Language Model: PLM)(104)에 입력하여 레이블의 워드 임베딩 특징을 추출할 수 있다. 예를 들어, 사전 학습된 언어 모델(104)은 BERT(Bidirectional Encoder Representation from Transformer)일 수 있으나, 이에 한정되는 것은 아니다. 여기서, 언어 모델(104)은 레이블이 입력되는 경우, 레이블의 워드 임베딩 특징을 추출하는데만 사용되는 모델이며, 소수 샷 연속 학습에서 학습되는 모델이 아니다.

[0086] 추출된 레이블의 워드 임베딩 특징을 사전 학습된 인공지능 모델(102)에서 출력되는 랭귀지 토큰의 특징으로 지식 증류(knowledge distillation)하여 지식을 전달할 수 있다. 이는, 학습 데이터라는 이미지 공간에서의 정보(랭귀지 토큰의 특징)과 레이블이라는 워드 공간에서의 정보(워드 임베드 특징)이 서로 다른 도메인의 정보이기 때문에, 지식 증류 기법을 통해 지식을 전달할 수 있다.

[0087] 또한, 레이블의 워드 임베딩 특징이 지식 증류된 랭귀지 토큰의 특징을 분류기에 통과시켜 랭귀지 토큰의 출력 분포($\hat{y}^{lang}$)를 생성할 수 있다. 이때, 랭귀지 토큰의 출력 분포($\hat{y}^{lang}$)와 정답 값의 분포($y_i$)와의 차이가 최소화 되도록 학습될 수 있다. 이에 따른, 제2 손실 함수($L_{SKD}$)는 하기 수학식 2로 표현할 수 있다.

[0088] (수학식 2)

[0089]
$$\mathcal{L}_{SKD} = \mathcal{L}_{KD}(f_i^{lang}, w_{c(y_i)}) + \lambda \cdot \mathcal{L}_{CE}(\hat{y}_i^{lang}, y_i)$$

[0090] $L_{KD}$ : 지식 증류(Knowledge Distillation)를 위한 손실

[0091] $L_{CE}$ : 크로스 엔트로피 손실(cross entropy loss)

[0092] $w_{c(yi)}$ : 레이블의 워드 임베딩 특징

[0093] $f_i^{lang}$ : 랭귀지 토큰의 특징

[0094] λ : 기 설정된 하이퍼 파라미터

[0095] 제2 손실 함수($L_{SKD}$)를 이용하면 학습 데이터에 대한 의미 지식(semantic knowledge)을 워드 임베딩 공간에서 추출하여 워드 임베딩 공간과 다른 도메인을 갖는 시각적 특징 공간으로  의미 지식을 전이하여 증분 세션 동안 추가적인 정보를 제공할 수 있게 된다.

[0096] 한편, 인공지능 모델(102)은 입력되는 학습 데이터(즉, 이미지)에 포함된 객체를 분류하게 되는 바, 객체 분류에 따른 제3 손실 함수($\mathcal{L}_{CE}(\hat{y}_i, y_i)$)을 포함하게 된다. 여기서, 제3 손실 함수는 인공지능 모델(102)의 분류기에서 출력하는 예측 값($\hat{y}_i$)과 정답 값($y_i$) 간의 차이를 최소화 하기 위한 것일 수 있다. 따라서, 사전 학습된 인공지능 모델(102)을 이용한 소수 샷 연속 학습의 총 손실 함수($L_{Total}$)는 하기 수학식 3으로 나타낼 수 있다.

[0097] (수학식 3)

[0098] $$\mathcal{L}_{Total} = \mathcal{L}_{CE}(\hat{y}_i, y_i) + \alpha \cdot \mathcal{L}_{ED} + \beta \cdot \mathcal{L}_{SKD}$$

[0099] 여기서, α 및 β는 각각 제1 손실 함수와 제2 손실 함수에 대한 스케일링 팩터(scaling factor)이다.

[0100] 기본 세션의 학습 시에는 수학식 3의 총 손실 함수($L_{Total}$)를 통해 인공지능 모델(102)을 학습할 수 있다. 반면, 증분 세션의 학습 시에는 제1 손실 함수($L_{ED}$)를 제외한 제2 손실 함수($L_{SKD}$) 및 제3 손실 함수($\mathcal{L}_{CE}(\hat{y}_i, y_i)$)를 통해 인공지능 모델(102)을 학습할 수 있다.

[0101] 개시되는 실시예에 의하면, 사전 학습된 인공지능 모델 및 프롬프트를 활용하여 기본 세션에서 충분한 지식을 학습하고 학습한 지식을 전이하여 증분 세션에서 연속적인 학습을 수행함으로써, 소수 샷 연속 학습에서 큰 규모의 인공지능 모델을 활용할 수 있고 파괴적 망각 현상 및 과적합 문제를 완화할 수 있게 된다.

[0103] 도 4는 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.

[0104] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 소수 샷 연속 학습을 위한 장치일 수 있다. 컴퓨팅 장치(12)는 사전 학습된 인공지능 모델(102)을 이용하여 소수 샷 연속 학습을 수행하기 위한 장치일 수 있다.

[0105] 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.

[0106] 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.

[0107] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0108] 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.

[0110] 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 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

## *부호의 설명*

[0112] 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

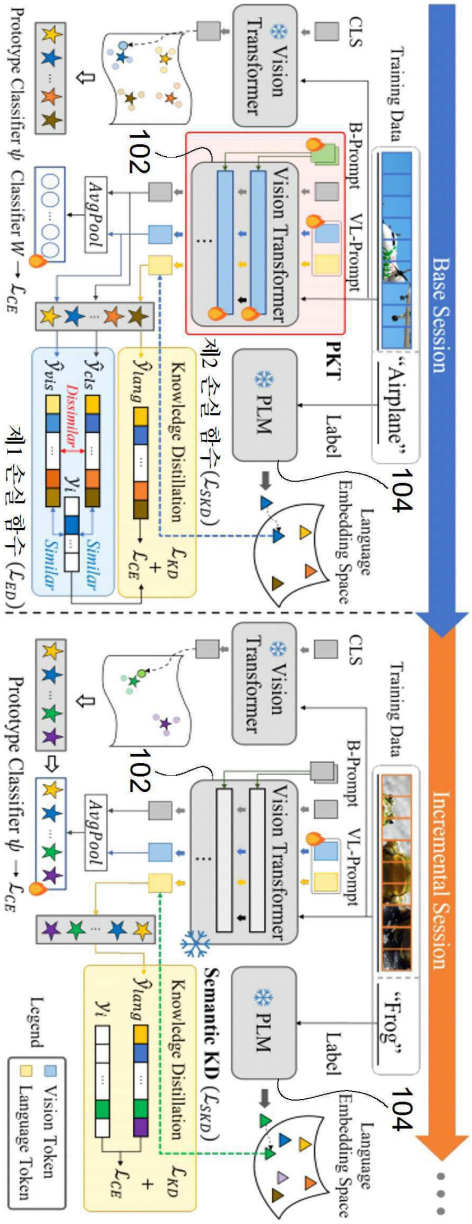
26 : 네트워크 통신 인터페이스

102 : 인공지능 모델

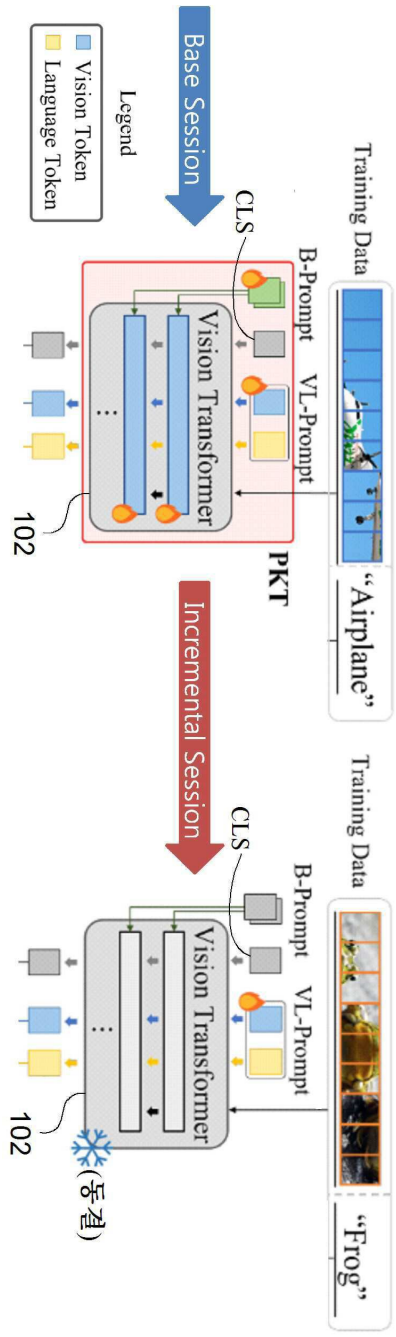
104 : 언어 모델

도면

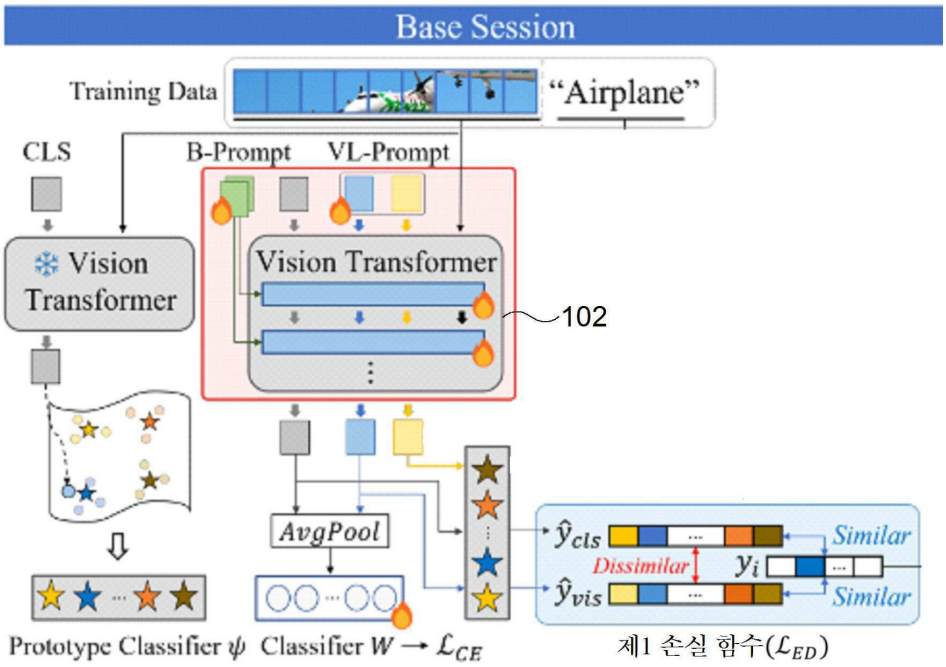
도면1


Base Session
Incremental Session
Training Data
"Airplane"
"Frog"
CLS
Vision Transformer
B-Prompt
VL-Prompt
Vision Transformer
PKT
Label
PLM
Language Embedding Space
104
102
AvgPool
Prototype Classifier $\psi$
Classifier $W$ → $\mathcal{L}_{CE}$
제2 손실 함수 ($\mathcal{L}_{SKD}$)
Knowledge Distillation
$\mathcal{L}_{KD}$
$\mathcal{L}_{CE}$
$\hat{y}_{lang}$
$\hat{y}_{cls}$
$\hat{y}_{vis}$
$y_i$
Dissimilar
Similar
제1 손실 함수 ($\mathcal{L}_{ED}$)
Semantic KD ($\mathcal{L}_{SKD}$)
Legend
Vision Token
Language Token

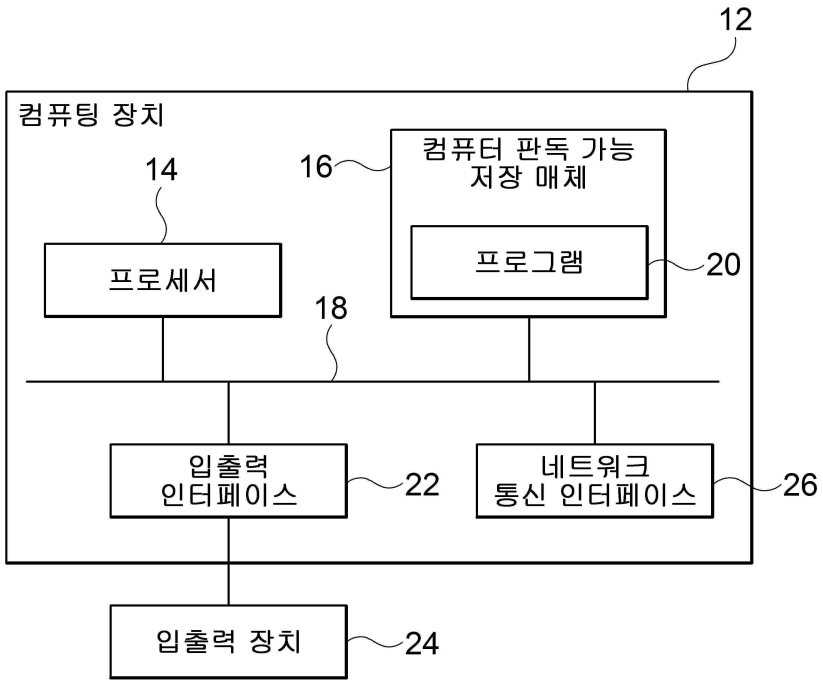
도면2


Base Session
Legend
Vision Token
Language Token
Training Data
CLS
B-Prompt
VL-Prompt
Vision Transformer
102
PKT
"Airplane"
Incremental Session
Training Data
CLS
B-Prompt
VL-Prompt
Vision Transformer
102
(동결)
"Frog"

도면3


Base Session
Training Data
"Airplane"
CLS
B-Prompt
VL-Prompt
Vision Transformer
Vision Transformer
102
AvgPool
Similar
Dissimilar
Similar
Prototype Classifier $\psi$
Classifier $W \rightarrow \mathcal{L}_{CE}$
제1 손실 함수($\mathcal{L}_{ED}$)

도면4


10
12
컴퓨팅 장치
14
16
컴퓨터 판독 가능 저장 매체
프로세서
프로그램
20
18
입출력 인터페이스
22
네트워크 통신 인터페이스
26
입출력 장치
24