



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0101172
(43) 공개일자 2025년07월04일

(51) 국제특허분류(Int. Cl.) G06N 3/0475 (2023.01) G06F 40/284 (2020.01) G06F 40/49 (2020.01) G06F 40/56 (2020.01) G06N 3/047 (2023.01) G06N 3/096 (2023.01)	(71) 출원인 고려대학교 산학협력단 서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(52) CPC특허분류 G06N 3/0475 (2023.01) G06F 40/284 (2020.01)	(72) 발명자 홍성태 서울특별시 성북구 고려대로24길 3-28, 204호
(21) 출원번호 10-2023-0192543	임희석 경기도 수원시 영통구 영통로 460, 304동 1704호
(22) 출원일자 2023년12월27일 심사청구일자 2023년12월27일	(74) 대리인 김동용

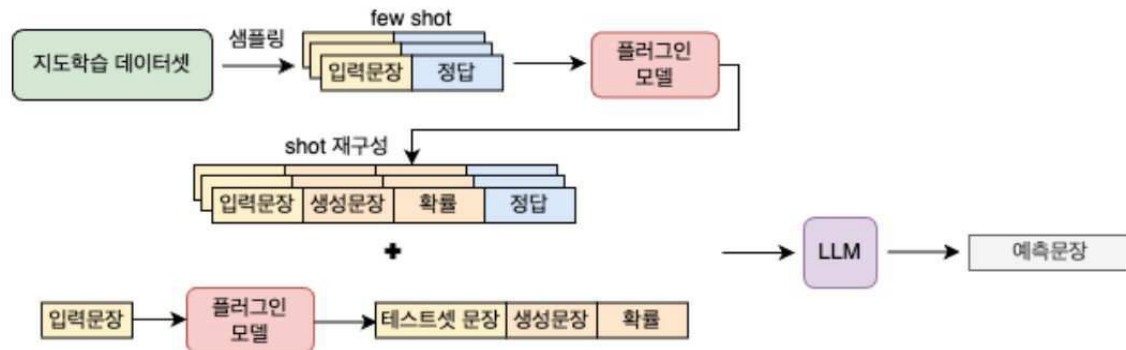
전체 청구항 수 : 총 4 항

(54) 발명의 명칭 Super In-Context Learning을 활용한 생성 장치 및 방법

(57) 요약

SuperICL(Super In-Context Learning)을 활용한 자연어 생성 장치 및 방법이 개시된다. 상기 생성 방법은 프로세서(Processor)를 포함하는 컴퓨팅 장치에 의해 수행되고, 문장들과 각각이 상기 문장들 각각에 대응하는 정답 문장들을 포함하는 지도학습 데이터셋으로부터 미리 정해진 개수의 샷(shot)을 샘플링하는 단계, 샘플링된 샷에 (뒷면에 계속)

대표도 - 도2



대한 미세조정된 언어 모델의 출력을 이용하여, 거대 언어 모델의 제1 입력을 생성하는 단계, 타겟문장에 대한 상기 미세조정된 언어 모델의 출력을 이용하여, 상기 거대 언어 모델의 제2 입력을 생성하는 단계, 및 상기 제1 입력과 상기 제2 입력을 상기 거대 언어 모델에 입력하여 상기 타겟문장에 대한 예측문장을 생성하는 단계를 포함하고, 상기 제1 입력은 상기 샘플링된 샷에 포함되는 입력문장, 상기 입력문장에 대한 상기 미세조정된 언어 모델의 출력인 제1 생성문장, 상기 제1 생성문장의 생성 확률, 및 상기 샘플링된 샷에 포함되는 정답문장을 포함한다.

(52) CPC특허분류

G06F 40/49 (2020.01)

G06F 40/56 (2020.01)

G06N 3/047 (2023.01)

G06N 3/096 (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193711
과제번호	2018-0-01405-006
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	Human-inspired 복합지능 원천기술 연구 및 인력양성
기 여 율	1/2
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193821
과제번호	2020-0-00368-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	차세대인공지능핵심원천기술개발
연구과제명	뉴럴-심볼릭(neural-symbolic) 모델의 지식 학습 및 추론 기술 개발
기 여 율	1/2
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

프로세서(Processor)를 포함하는 컴퓨팅 장치에 의해 수행되는, SuperICL(Super In-Context Learning)을 활용한 자연어 생성 방법에 있어서,

문장들과 각각이 상기 문장들 각각에 대응하는 정답문장들을 포함하는 지도학습 데이터셋으로부터 미리 정해진 개수의 샷(shot)을 샘플링하는 단계;

샘플링된 샷에 대한 미세조정된 언어 모델의 출력을 이용하여, 거대 언어 모델의 제1 입력을 생성하는 단계;

타겟문장에 대한 상기 미세조정된 언어 모델의 출력을 이용하여, 상기 거대 언어 모델의 제2 입력을 생성하는 단계; 및

상기 제1 입력과 상기 제2 입력을 상기 거대 언어 모델에 입력하여 상기 타겟문장에 대한 예측문장을 생성하는 단계를 포함하고,

상기 제1 입력은 상기 샘플링된 샷에 포함되는 입력문장, 상기 입력문장에 대한 상기 미세조정된 언어 모델의 출력인 제1 생성문장, 상기 제1 생성문장의 생성 확률, 및 상기 샘플링된 샷에 포함되는 정답문장을 포함하는,

자연어 생성 방법.

청구항 2

제1항에 있어서,

상기 제2 입력은 상기 타겟문장, 상기 타겟문장에 대한 상기 미세조정된 언어 모델의 출력인 제2 생성문장, 및 상기 제2 생성문장의 생성 확률을 포함하는,

자연어 생성 방법.

청구항 3

제2항에 있어서,

상기 미세조정된 언어 모델은, 학습 데이터를 이용하여 상기 미세조정된 언어 모델이 임의의 태스크를 수행하도록 사전학습된 언어 모델을 미세조정하여 생성되고,

상기 임의의 태스크는 입력에 대한 요약문을 생성하는 태스크, 입력에 대한 번역문을 생성하는 태스크, 및 대화 이력을 포함하는 입력에 대한 응답을 생성하는 태스크 중 어느 하나인,

자연어 생성 방법.

청구항 4

제3항에 있어서,

상기 확률은 상기 제1 생성문장 또는 상기 제2 생성문장에서 각 토큰의 생성 확률이 평균인,

자연어 생성 방법.

발명의 설명

기술 분야

본 발명은 자연어 처리와 기계 학습 방법론에 관한 것으로, 특히 SuperICL(Super In-Context Learning)을 활용한 자연어 생성 모델을 생성하는 장치 및 방법과 생성된 자연어 생성 모델을 이용한 자연어 생성 장치 및 방법에 관한 것이다.

[0001]

배경 기술

[0002] 거대 언어 모델(Large Language Models, LLM)이 등장하면서 기계 번역, 요약 및 대화 등의 다양한 자연어 처리 TASK에서 최고 성능을 달성했다. OpenAI의 GPT-4는 자연어 이해 및 생성 작업에서 인간 수준의 능력을 보여준다. 이러한 뛰어난 성과들로 인해 거대 언어 모델은 현대 인공지능 연구의 중요한 주제 중 하나로 각광받고 있다. 그러나 거대 언어 모델은 몇가지 중요한 문제와 한계를 가지고 있다. 먼저, 대규모 학습 데이터와 막대한 수의 모델 파라미터를 필요로 한다. 이는 학습과 배포 과정에서 상당한 컴퓨팅 리소스와 비용을 요구하기 때문에 모델을 사용하는데 제약사항이 많이 존재한다. 또한, 거대 언어 모델은 다양한 도메인에 알맞게 미세 조정하기 어렵다는 문제도 있다. 특히 새로운 도메인의 전문지식을 필요로하거나 새로운 작업에 적응하는 것은 매우 어려운 과제이다.

[0003] 이러한 문제를 해결하기 위한 방안으로 ICL(In-Context Learning) 방법론이 자연어 처리 분야에서 주목을 받고 있다. ICL은 주어진 데이터의 컨텍스트 정보를 활용하여 거대 언어 모델을 더 효과적으로 조정하고 최적화하는 방법론이다. 이 방법론은 실시간으로 새로운 정보나 패턴을 학습하는 방식이기 때문에 매우 유연하고, 적은 양의 데이터 셋에서도 효율적인 새로운 지식을 습득할 수 있어 효율성이 높다. 그러나 ICL의 몇 가지 한계점이 존재한다. 첫째, 컨텍스트로 주어지는 샷의 개수와 순서에 매우 민감하며, 형식이나 패턴을 완벽하게 따르지 않을 수 있다는 점이다. 둘째, 언어 모델의 최대 입력 토큰 길이이다. 많은 컨텍스트 정보를 제공하기 위해서는 많은 토큰이 필요하며, 언어 모델의 최대 입력 토큰 길이로 인해 정보를 모두 담기에 한계가 있다. 이러한 제한을 극복하기 위해 본 발명에서는 SuperICL을 생성 TASK에 적용함으로써 Super In-Context Learning for Generation(SuperICL4Gen)를 제안한다. SuperICL4Gen은 SuperICL을 생성 TASK로 확장하여 성능을 개선한 방법론이다. 본 발명은 실험을 통해 플러그인 모델의 교체를 통해 다양한 작업에 적응할 수 있는 가능성을 확인하고, 자연어 생성 작업에서 우수한 성능을 평가하였다. 실험 결과에서는 BLEU, ROUGE의 평가 지표를 사용하여 SuperICL4Gen 방법론을 사용한 모델을 기존 모델과 비교했다. 평가 지표를 통해 제안 모델이 기존 모델을 상당히 능가하는 결과를 얻었으며, 또한 선호도 평가에서도 모델의 효과성을 확인했다. 이로써 SuperICL4Gen 방법론은 자연어 생성 작업에서의 혁신적인 성과를 입증하고, 자연어 처리 분야에 새로운 가능성을 제시하는 연구임을 보여준다.

선행기술문헌

특허문헌

[0004] (특허문헌 0001) 대한민국 공개특허 제2023-0118006호 (2023.08.10. 공개)
(특허문헌 0002) 대한민국 공개특허 제2022-0011979호 (2022.02.03. 공개)

발명의 내용

해결하려는 과제

[0005] 본 발명이 이루고자 하는 기술적인 과제는 SuperICL(Super In-Context Learning)을 활용한 자연어 생성 장치 및 방법을 제공하는 것이다.

과제의 해결 수단

[0006] 본 발명의 일 실시예에 따른, SuperICL(Super In-Context Learning)을 활용한 자연어 생성 방법은 프로세서(Processor)를 포함하는 컴퓨팅 장치에 의해 수행되고, 문장들과 각각이 상기 문장들 각각에 대응하는 정답문장들을 포함하는 지도학습 데이터셋으로부터 미리 정해진 개수의 샷(shot)을 샘플링하는 단계, 샘플링된 샷에 대한 미세조정된 언어 모델의 출력을 이용하여, 거대 언어 모델의 제1 입력을 생성하는 단계, 타겟문장에 대한 상기 미세조정된 언어 모델의 출력을 이용하여, 상기 거대 언어 모델의 제2 입력을 생성하는 단계, 및 상기 제1 입력과 상기 제2 입력을 상기 거대 언어 모델에 입력하여 상기 타겟문장에 대한 예측문장을 생성하는 단계를 포함하고, 상기 제1 입력은 상기 샘플링된 샷에 포함되는 입력문장, 상기 입력문장에 대한 상기 미세조정된 언어 모델의 출력인 제1 생성문장, 상기 제1 생성문장의 생성 확률, 및 상기 샘플링된 샷에 포함되는 정답문장을 포함한다.

발명의 효과

- [0007] 본 발명의 실시예에 따른, SuperICL(Super In-Context Learning)을 활용한 자연어 생성 장치 및 방법에 의할 경우, 자연어 생성 모델의 정량적인 성능과 정성적인 성능을 향상시킴으로써 다양한 자연어 생성 태스크에서 우수한 결과를 얻을 수 있다.
- [0008] 또한, 기학습된 미세 조정된 모델과 거대 언어 모델의 결합을 통해, 새로운 도메인 또는 다른 태스크에 알맞은 플러그인 모델이 교체를 통해 적응성을 높일 수 있다.
- [0009] 또한, 입력 재구성을 통해, 예측 문장을 생성할 때 불필요한 정보 없이 정확하고 신뢰할 수 있는 예측을 달성할 수 있다.

도면의 간단한 설명

- [0010] 본 발명의 상세한 설명에서 인용되는 도면을 보다 충분히 이해하기 위하여 각 도면의 상세한 설명이 제공된다.
- 도 1은 ICL을 이용한 컨텍스트 구성 및 학습 방법을 설명하기 위한 도면이다.
- 도 2는 본 발명의 일 실시예에 따른 컨텍스트 구성 및 학습 방법을 설명하기 위한 도면이다.
- 도 3은 본 발명의 일 실시예에 따른 제안 기법의 선호도 평가 결과를 도시한다.
- 도 4는 본 발명의 일 실시예에 따른 SuperICL을 활용한 생성 방법을 설명하기 위한 흐름도이다.

발명을 실시하기 위한 구체적인 내용

- [0011] 본 명세서에 개시되어 있는 본 발명의 개념에 따른 실시예들에 대해서 특정한 구조적 또는 기능적 설명들은 단지 본 발명의 개념에 따른 실시예들을 설명하기 위한 목적으로 예시된 것으로서, 본 발명의 개념에 따른 실시예들은 다양한 형태들로 실시될 수 있으며 본 명세서에 설명된 실시예들에 한정되지 않는다.
- [0012] 본 발명의 개념에 따른 실시예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.
- [0013] 제1 또는 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 목적으로만, 예컨대 본 발명의 개념에 따른 권리 범위로부터 벗어나지 않은 채, 제1 구성 요소는 제2 구성 요소로 명명될 수 있고 유사하게 제2 구성 요소는 제1 구성 요소로도 명명될 수 있다.
- [0014] 어떤 구성 요소가 다른 구성 요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성 요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성 요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성 요소가 다른 구성 요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는 중간에 다른 구성 요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성 요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0015] 본 명세서에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0016] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.

- [0017] 이하, 본 명세서에 첨부된 도면들을 참조하여 본 발명의 실시예들을 상세히 설명한다. 그러나, 특허출원의 범위가 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0018] 우선, ICL과 SuperICL에 관해 상세히 설명한다.
- [0019] **ICL(In-Context Learning)**
- [0020] ICL(In-Context Learning)은 처음으로 GPT-3 논문(T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell et al., "Language models are few-shot learners," Advances in neural information processing systems, Vol. 33, pp. 1877-1901, 2020.)에서 소개된 개념으로, 모델의 수정이나 업데이트 없이 거대 언어 모델을 새로운 태스크에 활용할 수 있게 도와주는 새로운 접근 방식이다. 일반적으로 특정 문맥에서 정보나 기술을 습득하는 학습 방법을 의미한다. ICL은 거대 언어 모델이 방법을 찾고 예측하는 학습을 가능하게 하기 위해 테스트 입력 앞에 몇 가지 학습 샘플을 컨텍스트로 추가하는 방법을 사용한다.
- [0021] 해당 방법론은 기계 번역 및 데이터 생성과 같은 태스크에서 널리 활용되고 있다. 이는 실시간으로 새로운 정보나 패턴을 학습하는 방식으로, 데이터 양이 적어도 효율적으로 새로운 주어진 컨텍스트를 충분히 이해하고 따라하는 능력을 가질 수 있다. 그러나 ICL에는 몇 가지 제한 사항이 있다. 첫째, 컨텍스트로 주어지는 샷의 개수와 순서에 민감하며, 완벽하게 따르지 않을 수 있다. 둘째, 언어 모델의 최대 입력 토큰 길이로 인해 많은 컨텍스트 정보를 담는 데 제한이 있다. 셋째, 거대 언어 모델이 충분한 컨텍스트 이해를 위해 필요한 샷의 품질이 매우 중요하며, 이를 수작업으로 샘플링하는 것은 번거로운 작업이다. 이러한 단점들을 극복하기 위한 여러 연구들이 진행되고 있다.
- [0022] **SuperICL(Super In-Context Learning)**
- [0023] SuperICL(Super In-Context Learning)(C. Xu, Y. Xu, S. Wang, Y. Liu, C. Zhu, and J. McAuley, "Small models are valuable plug-ins for large language models," arXiv preprint arXiv:2305.08848, 2023.)은 거대 언어 모델과 미세 조정된 모델을 결합하여 지도학습 작업의 성능을 향상시키는 방식을 제안하고 있다.
- [0024] 이 접근 방식은 두가지 단계로 수행되는데, 먼저 SuperICL의 첫 번째는 사전 학습 모델을 작업별 레이블이 지정된 데이터에서 미세 조정하는 것이다. 이 모델들은 플러그인 역할을 하며 분류 태스크에 대한 지식과 예측 결과를 제공한다. 그 다음 플러그인 모델이 제공한 지식을 활용하기 위해 거대 언어 모델에 대한 컨텍스트를 재구성한다. 재구성된 컨텍스트에는 예측된 레이블과 그에 따른 확률값이 모두 포함된다. 추가적인 정보를 컨텍스트에 포함시켜 거대 언어 모델은 이를 고려하여 예측을 생성한다. SuperICL은 분류 태스크에서 미세 조정된 베이스라인 모델 및 ICL 모델에 비해 우수한 성능을 달성하고 불안정성 문제를 해결한다. 또한, 플러그인 모델이 정보를 흡수하도록 레이블을 예측하도록 하고 거대 언어 모델은 더 일반적인 언어 이해에 중점을 둘 수 있도록 하며 사전 학습 모델을 거대 언어 모델에 통합하여 지도 학습을 수행할 수 있는 방법을 보여주고 데이터 셋에서 적절한 샷을 선택해야하는 노고를 줄여준다. 이러한 결과와 함께 SuperICL은 거대 언어 모델 시대의 지도 학습에 대한 희망적인 새로운 패러다임을 제시한다.
- [0025] 제안 기법(SuperICL4Gen)과 SuperICL의 주요 차이점은 추가되는 정보에 각 분류 레이블 확률 대신 각 예측 문장의 토큰 생성 확률의 평균 값을 활용하여 기존의 확률값을 대체한다. 동일한 방법론을 생성 작업을 포함한 다양한 자연어 태스크에 보다 적합하게 활용될 수 있도록 변경하고 범위를 확장한 것이다.
- [0026] 이하에서는, 본 발명에서 제안하는 모델에 대해 설명한다.
- [0027] **1. 플러그인 모델 미세 조정**
- [0028] 본 발명에서 제안 모델은 지도 학습 데이터셋을 활용하여 미세 조정된 생성 모델과 거대 언어 모델을 결합한다. 이때, "플러그인 모델"이라는 용어는 다양한 자연어 생성 태스크에 대해 학습된 모델로서, 필요한 작업에 따라 다른 모델로 대체할 수 있는 기능을 갖춘 모델을 말하며 이러한 플러그인 모델은 SuperICL4Gen의 핵심 요소로 작용한다. 미세 조정된 모델은 언어 생성 작업에 필요한 특정 지식과 능력을 제공하며, 성능을 향상시키기 위해 함께 작동한다. 필요 시 다른 태스크를 수행하는 모델로 대체하여 다양한 자연어 생성 작업에 활용할 수 있다. 예를 들어, 요약 작업에 특화된 플러그인 모델을 사용하거나, 번역 작업을 위해 또 다른 플러그인 모델을 적용할 수 있다. 이는 SuperICL4Gen이 각 작업에 맞게 특정 모델을 유연하게 선택하여 활용하는 다목적 모델 역할을 수행할 수 있음을 말한다.

[0029] **2. 거대 언어 모델을 위한 입력 재구성**

[0030] ICL을 이용한 컨텍스트 구성 및 학습 방법을 도시하는 도 1을 참조하면, 기존의 ICL은 특정 작업에 관련된 컨텍스트를 입력으로 사용하는데 주로 데이터셋에서 샘플링한 예제들로 구성된다. 이러한 예제들은 입력 텍스트와 해당 작업의 정답을 사용하여 해당 태스크에 대한 컨텍스트를 구성한다. 즉, ICL은 데이터셋에서 가져온 예제들을 활용하여 문맥을 형성하고 특정 작업에 활용하는 방식을 채택하고 있다.

[0031] 반면, SuperICL에서는 플러그인 모델의 생성 결과를 다시 거대 언어 모델의 입력으로 추가적인 정보를 제공한다. 각 분류 레이블 확률 값이 컨텍스트의 정보로 추가되는 SuperICL과 달리, 본 발명의 일 실시예에 따른 컨텍스트 구성 및 학습 방법을 도시하는 도 2를 참조하면, SuperICL4Gen은 플러그인 모델을 통해 입력에 대한 예측 문장을 생성하고, 이 생성된 예측 문장을 활용하여 각 토큰 생성 확률의 평균 값을 계산하고 이를 레이블 확률 값과 대체한다. 이 과정에서 학습 데이터셋에서 무작위로 샘플링된 샷을 사용하며, 이로써 다양한 문맥과 예제를 포함하므로 모델이 여러 상황에서 작동하고 적응할 수 있도록 도와준다. 각각의 샷은 다음과 같은 정보를 포함한다.

- [0032] - 입력 문장: 특정 작업에 대한 입력 문장
- [0033] - 생성 문장: 플러그인 모델에 의해 생성된 결과 문장
- [0034] - 문장 생성 확률: 플러그인 모델이 예측한 결과에 대한 생성 확률
- [0035] - 정답: 해당 샷에 대한 실제 정답 문장

[0036] 테스트 입력 문장 또한 마찬가지로, 플러그인 모델을 사용하여 입력문에 대한 출력과 확률을 고려하여 입력을 재구성한다. 이러한 방식을 통해 SuperICL4Gen에서 거대 언어 모델은 플러그인 모델의 예측 결과와 생성 확률을 고려하여 최종 예측을 생성한다. 불필요한 정보가 주어진 경우 예측 문장을 참고하지 않고 새로운 생성 결과를 재생성함으로써, 정확하고 신뢰할 수 있는 예측 달성이 가능하다.

[0037] **실험 환경**

[0038] SuperICL4Gen 방법론을 이용하여 요약문 생성 태스크에 대한 성능을 평가하기 위해 실험을 수행한다. 본 실험에서 플러그인 모델로 KE-T5 Base를 기반으로 사용한다. KE-T5 Base(K. AIRC, "Ke-t5: Korean english t5," Mar. 2021. [Online]. Available: <https://github.com/AIRC-KETI/ke-t5>)은 Raffel 등의 모델(C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," The Journal of Machine Learning Research, Vol. 21, No. 1, pp. 5485-5551, 2020.)을 한국어와 영어 코퍼스를 이용하여 사전학습한 모델이다. 미세조정을 위해 사용된 데이터셋은 다양한 주제에 대한 요약문 데이터를 포함하고있는 AIHub 문서요약 텍스트 데이터셋을 사용하였으며, 학습데이터는 23,505개, 검증 데이터는 6,715개를 사용하였다. 제안 모델의 평가를 위해 테스트 데이터 100개를 사용하였으며 랜덤으로 추출된 3개의 샷을 컨텍스트로 제공하며 각 모델마다 주어진 샷은 동일하다.

[0039] **1. 학습 설정**

[0040] 플러그인 모델인 KE-T5-Base는 멀티헤드, 레이어 수는 각 12개이고 768차원의 벡터를 가지며 사전 단어의 크기는 64,128 개이다. 실험을 위한 모델의 하이퍼 파라미터는 배치 사이즈 128, 학습률 $1e-4$, 입력 최대 길이 256, 출력 최대 길이 64, 조기 학습 종료 조건 5로 설정한다. 거대 언어 모델로 사용한 모델은 GPT-3.5-TURBO이다. GPT-3를 개선하여 자연어를 이해하고 생성할 수 있으며 최대 토큰 4,097개를 입력으로 사용 가능한 모델이다.

[0041] **2. 평가 방법**

[0042] **양적 평가 방법** 본 발명에서는 생성된 자연어 요약문의 성능을 평가하기 위해 BLEU와 ROUGE 두 가지 주요 평가 메트릭을 사용한다. BLEU와 ROUGE는 자연어 처리 연구에서 널리 인정받는 평가 지표이다. BLEU는 기계 번역 및 요약 생성과 같은 작업에서 널리 사용되는 평가 메트릭 중 하나이다. BLEU는 생성된 텍스트와 참조 텍스트 간의 단어 및 구문 일치율을 측정하여 평가한다. ROUGE는 주로 요약 생성 및 문서 비교 작업에서 사용되는 평가 방법으로, 생성된 요약문과 참조 요약문 간의 일치 정도를 측정한다. 이 두 평가 메트릭을 사용함으로써, 제안 모델의 자연어 요약문 생성 성능을 객관적으로 평가하고 비교할 수 있다.

[0043] **선호도 평가 방법** 본 발명에서 생성된 두 문장의 선호도 평가를 위해 GPT-3.5-Turbo 모델을 활용하며 온도 (temperature)를 0으로 설정하여 정확한 선호도 평가를 수행한다. 최근 거대 언어 모델을 이용한 평가는 사람의

주석없이도 높은 상관관계를 보인다. 거대 언어 모델은 대량의 다양한 데이터에서 학습되는데 이는 모델이 다양한 언어와 주제를 이해하고 자연스럽게 생성할 수 있도록 도와준다. 학습 데이터의 다양성은 모델이 선호도 평가를 다양한 상황에서 수행할 수 있는 기반을 제공한다. 또, 거대 언어 모델은 동일한 입력에 대해 일관된 예측을 수행하며 주관적인 편견을 가지지 않기 때문에 모델이 사용자의 주관을 반영하지 않고 객관적인 기준에 따라 선호도를 판단한다는 점을 의미하므로 객관성 있는 선호도 평가 결과를 보장한다. 이러한 두 가지 측면은 언어 모델을 선호도 평가에 사용함으로써 더 신뢰할 만한 결과를 얻을 수 있다. 지시어로 사용한 문장은 다음과 같다.

[0044] 1, 2번 중 어느 문장이 요약문 입력에 대해 잘 요약했는지 번호를 알려줘 (단, 비슷한 경우 tie)

[0046] 1번 문장: {1번 문장 입력}

[0047] 2번 문장: {2번 문장 입력}

[0049] 결과:

[0051] 실험 결과

[0052] 1. 양적 실험 결과

[0053] 표 2는 SuperICL4Gen 방법론을 사용한 모델의 요약문 생성 성능을 보여주는 표이고, 표 1은 모델의 결과를 보여주는 예시이다. 요약문 생성 테스트에서 제안 모델인 SuperICL4Gen은 베이스라인 모델보다 BLEU, ROUGE 스코어가 베이스라인 모델에 비해 각각 12.23%, 5.82%, 7.02% 및 1.27% 더 높은 성능을 보인다. 일반적인 ICLearning을 사용한 GPT3.5 ICL 모델과의 비교에서 SuperICL4Gen을 사용한 모델이 GPT-3.5 ICL 모델에 비해 ROUGE-2에서 낮은 결과를 보이지만 BLEU 점수에서 9.9% 더 높은 점수를 기록하고, ROUGE-1, ROUGE-L에서 각각 1.91%, 1.88% 더 높은 결과를 보인다.

[0054] [표 1]

Context	Input: 사우디아라비아의 원유생산시설이 지난 14일(현지시간) 예멘 반군의 드론에 피격돼 하루 570만배럴의 생산 차질을 빚게 됐다는 소식에 주식 시장에서 관련 업종... T5-Base Prediction:사우디의 원유생산시설이 예멘 반군의 드론에 피격돼 하루 570만배럴의 생산 차질을 빚게 됐다는 소식에 주식 시장에서 관련 업종의 희비가 엇갈렸다. Confidence: 0.8450 Gold: 사우디아라비아의 원유시설이 지난 14일 드론에 피격되면서 생산 차질로 인해 관련 업종인 정유·조선 업종은 오르고, 화학 업종의 경우 불안한 흐름을 보이고 있으며, 이로 인해 사우디는 16일까지 피격으로 인해 줄어든 원유 생산량의 30%를 복구할 계획이라고 밝혔다. ...
Test	Input:건축·부동산 전문가, 시민단체 등과 주택정책 토론회 개최 청주시는 29일 오후 4시 상당도서관 다목적실에서 ... T5-Base Prediction:청주시는 29일 오후 4시 상당도서관 다목적실에서 '주택정책 토론회'를 개최하여 청주시 주택시장의 문제점을 공론화하고 합리적인 주택정책 방향을 모색하기 위해 추진되었으며, 이근복 시 공동주택과장은 토론회를 통해 시민을 이해하고 시민과 함께하는 청주시가 되도록 노력하겠다고 말했다. Confidence: 0.8146
Prediction	Ground Truth: 청주시는 29일 '주택정책 토론회'를 개최하여 청주시 주택시장의 문제점을 공론화하고 합리적인 주택정책 방향을 모색하기 위해 노력할 것이라고 밝혔다.

[0055]

[0056] [표 2]

model	BLEU	ROUGE-1	ROUGE-2	ROUGE-L
KE-T5-Base	14.15	0.5845	0.4061	0.4502
GPT-3.5 ICL	14.45	0.6069	0.4470	0.4475
SuperICL4Gen	15.88	0.6185	0.4346	0.4559

[0057]

[0058] 2. 선호도 평가

[0059] 선호도 평가를 위해 무작위로 선택된 테스트 데이터셋 50개에 대해 평가를 수행한다. 선호도 평가는 GPT-3.5-Turbo 모델을 사용하여 진행되었으며, 결과는 도 3에서 확인할 수 있다. 베이스라인 모델인 KE-T5-Base 모델과 비교에서 제안 모델은 압도적인 성능 차이를 보인다. 이 비교에서 SuperICL4Gen은 46번의 승리를 기록하였으며, 무승부 2건을 포함한 48건의 긍정적인 결과를 확인할 수 있었다. 다음으로, GPT-3.5 ICL과의 비교에서도 제안 모델은 높은 성능을 보인다. 11건의 패배가 존재하였으나 38번의 선호도를 차지하는 결과를 확인할 수 있다.

[0060] 3. 샷의 개수에 따른 성능 분석

[0061] ICL은 샷의 개수에 따라 성능이 큰 폭으로 차이 나고 거대 언어 모델이 패턴을 잘 이해할 수 있게 하려면, 샷 선택 시 높은 퀄리티의 컨텍스트를 수동으로 선택해야 하는 단점이 있다. SuperICL4Gen에서는 적은 샷의 개수에서도 강건하며 컨텍스트에 추가로 제공되는 정보만으로도 거대 언어 모델이 효과적으로 잘 활용하는 능력을 보여준다.

[0062] 표 3은 샷의 개수에 따른 모델 간의 성능 비교이다. 3-shot에서 1-shot으로 샷의 개수를 감소시킬 때 ROUGE-1, ROUGE-2 및 ROUGE-L 점수는 각각 12.14%, 6.38%, 10.97%의 큰 폭의 성능 감소를 보인다. GPT-3.5 ICL은 주어지는 샷의 개수에 따른 큰 폭의 성능 차이를 확인할 수 있다. 그러나 SuperICL4Gen에서는 3-shot에서 1-shot으로 샷의 개수를 감소하더라도 일반적인 ICL에 비해 적은 폭의 차이를 나타내며 강건한 성능을 유지하는 것을 확인할 수 있다. 이는 컨텍스트 내 추가적인 정보만으로도 성능이 향상될 수 있으며, 샷의 개수에 변화에 민감하지 않음을 나타낸다.

[0063] [표 3]

model	n-shot	ROUGE-1	ROUGE-2	ROUGE-L
GPT-3.5 ICL	1-shot	0.5332	0.4185	0.3984
	3-shot	0.6069	0.4470	0.4475
SuperICL4Gen	1-shot	0.5920	0.4224	0.4534
	3-shot	0.6185	0.4346	0.4559

[0064]

[0065] 4. 실험 결과 및 분석

[0066] 양적 실험 결과에서 제안 모델이 문장 수준의 일치도 및 요약 능력에서 강점을 보여주고 있음을 확인할 수 있다. 전반적으로 SuperICL4Gen 방법론을 적용한 모델은 KE-T5-Base와 GPT-3.5 ICL과의 비교 지표에서 일관된 우수한 성능을 보인다. 이는 컨텍스트에 추가적으로 제공하는 정보가 효과적으로 활용되어 요약하는데 도움이 되는 것을 알 수 있다. 선호도 평가 실험에서 단순히 ICL 방법론을 사용하여 입력과 정답으로 구성된 컨텍스트를 제공하는 것보다 더 많은 정보(확률 및 예측 문장)를 제공하여 생성 태스크를 수행하는 방법론이 더 효과적인임을 보여준다.

[0067] 도 4는 본 발명의 일 실시예에 따른 SuperICL을 활용한 생성 방법을 설명하기 위한 흐름도이다.

[0068] 도 4를 참조하면, 생성 방법은 적어도 프로세서(Processor) 및/또는 메모리(Memory)를 포함하는 컴퓨팅 장치에 의해 수행될 수 있다. 다시 말하면, 생성 방법을 구성하는 단계들 중 적어도 일부는 컴퓨팅 장치에 포함되는 프로세서의 동작으로 이해될 수 있고, 이 경우 컴퓨팅 장치는 생성 장치로 명명될 수 있다. 컴퓨팅 장치는 PC(Personal Computer), 서버(server), 랩탑 컴퓨터, 태블릿 PC, 스마트폰(smart phone) 등을 포함할 수 있다. 이하에서, 생성 방법을 설명함에 있어 앞선 기재와 중복되는 내용에 관하여는 그 구체적인 기재를 생략하기로 한다.

[0069] 우선, 플러그인 모델이 획득된다(S110). 플러그인 모델은 임의의 언어 모델(예컨대, KE-T5 Base)을 미세조정함으로써 생성될 수 있다. 실시예에 따라, 미세조정이 완료된 모델이 미리 컴퓨팅 장치에 저장되어 있을 수도 있다. 미세 조정은 원문과 원문에 대한 요약문을 포함하는 학습 데이터를 이용하여 수행될 수 있다.

[0070] 거대 언어 모델의 제1 입력이 생성된다(S120). 지도학습 데이터로부터 미리 정해진 개수의 샷(shot)이 구성될 수 있다. 생성된 샷을 플러그인 모델에 입력함으로써 제1 입력이 생성된다. 제1 입력은 샷에 포함된 입력문장, 입력문장에 대응하는 생성문장, 확률, 및 정답(정답 문장)을 포함할 수 있다.

[0071] 지도학습 데이터는 입력문장과 입력문장에 대응하는 정답문장을 포함하고, 이는 컴퓨팅 장치에 미리 저장되어 있을 수 있다. 생성문장은 플러그인 모델의 출력으로써, 미세 조정에 따라 미리 정해진 태스크를 수행하는 플러그인 모델의, 입력문장에 대한 출력을 의미한다. 예시적으로, 요약문 작성 태스크의 경우 생성문장은 입력문장에 대한 요약문일 수 있고, 번역 태스크의 경우 생성문장은 입력문장에 대한 번역문일 수 있고, 대화 태스크의

경우 생성문장은 입력문장(대화이력을 의미할 수 있음)에 대응하는 머신의 응답일 수 있다. 확률은 문장생성 확률을 의미한다. 다시 말하면, 퓨샷 러닝(few-shot learning)을 통해 거대 언어 모델의 제1 입력을 생성할 수 있다.

[0072] 거대 언어 모델의 제2 입력이 생성된다(S130). 제2 입력은 입력문장(타겟문장이나 테스트문장으로 명명될 수도 있음)에 대한 플러그인 모델의 출력을 이용하여 생성될 수 있다. 제2 입력은 타겟문장, 타겟문장에 대응하는 생성문장, 및 확률을 포함할 수 있다.

[0073] 타겟문장에 대한 예측문장이 생성된다(S140). 예측문장은 제2 입력에 포함되어 있는 타겟문장에 대응하는 예측문장으로서, 요약문, 번역문, 발화 등이 될 수 있다. 이때, 예측문장은 생성된 제1 입력과 제2 입력을 거대 언어 모델에 입력함으로써 생성될 수 있다. 즉, Super ICL 기법을 이용하여 타겟문장에 대한 예측문장이 생성될 수 있다.

[0074] 실시예에 따라, 거대 언어 모델은 컴퓨팅 장치에 저장되어 있지 않고, 웹방식을 통해 제공될 수 있다. 이 경우, 제1 입력과 제2 입력을 거대 언어 모델 서비스를 제공하는 서버로 송신하고, 이에 대한 응답을 수신함으로써 예측문장이 생성될 수 있다.

[0075] 이상에서 설명된 장치는 하드웨어 구성 요소, 소프트웨어 구성 요소, 및/또는 하드웨어 구성 요소 및 소프트웨어 구성 요소의 집합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성 요소는, 예를 들어, 프로세서, 컨트롤러, ALU(Arithmetic Logic Unit), 디지털 신호 프로세서(Digital Signal Processor), 마이크로컴퓨터, FPA(Field Programmable 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. 처리 장치는 운영 체제(Operation System, OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술 분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(Processing Element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(Parallel Processor)와 같은, 다른 처리 구성(Processing Configuration)도 가능하다.

[0076] 소프트웨어는 컴퓨터 프로그램(Computer Program), 코드(Code), 명령(Instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(Collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성 요소(Component), 물리적 장치, 가상장치(Virtual Equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(Signal Wave)에 영구적으로, 또는 일시적으로 구체화(Embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

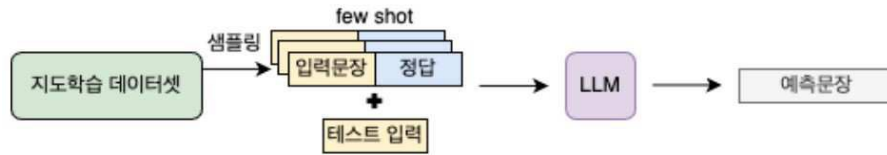
[0077] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(Magnetic Media), CD-ROM, DVD와 같은 광기록 매체(Optical Media), 플롭티컬 디스크(Floptical Disk)와 같은 자기-광 매체(Magneto-optical Media), 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0078] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성 요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성 요소 또는 균등물에 의하여 대치되거나

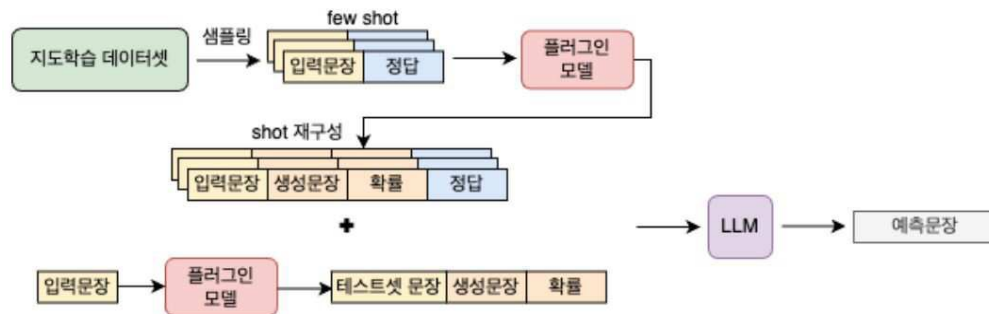
치환되더라도 적절한 결과가 달성될 수 있다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

도면

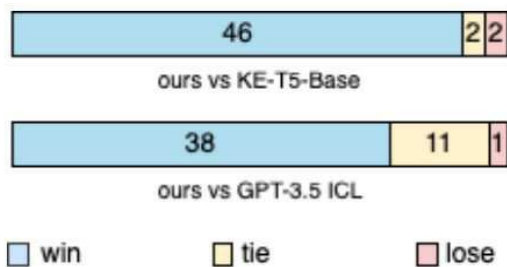
도면1



도면2



도면3



도면4

