



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

프롬프트 기반 학습을 활용한
텍스트 추상 요약 모델

Prompt-based Abstractive Text
Summarization

2023년 08월

서울대학교 대학원

산업공학과

배지현

프롬프트 기반 학습을 활용한 텍스트 추상 요약 모델

지도 교수 조성준

이 논문을 공학석사 학위논문으로 제출함
2023년 06월

서울대학교 대학원
산업공학과
배지현

배지현의 공학석사 학위논문을 인준함
2023년 08월

위원장 _____ 윤명환 (인)

부위원장 _____ 조성준 (인)

위원 _____ 이성주 (인)

초 록

추상요약은 최근 그 유용성을 크게 인정받고 있다. 추상 요약은 일관성 있는 요약을 생성할 뿐 아니라 저작권 관련 분쟁을 해결할 수 있는 방법으로 입력 텍스트의 의미와 주요 정보를 파악해 새로운 문장을 생성한다. 최근 프롬프트 기반 학습방법은 전체 모델 파인 튜닝과 유사한 성능을 보이며 효과적으로 활용하여 과업 특화 지식을 보유함과 동시에 거대 언어 모델의 사용 가능한 지식 추출하는 데에 사용된다. 본 연구에서는 프롬프트 기반 학습을 통해 추상 요약을 수행하는 방법을 제안한다. 제안 모델이 창의적이고 효율적인 텍스트 요약을 수행할 수 있음을 입증하고, 금융 도메인의 추상 요약 데이터셋을 직접 구축하여 도메인 지식이 필요한 데이터에도 모델이 효율적으로 작동함을 보인다. 타 전체 모델 튜닝 방법 및 요약 특화 모델에 비교해 극히 소량의 파라미터만을 조정함에도 유사한 성능을 보임을 확인하였다.

주요어 : 텍스트 요약, 프롬프트, 프롬프트 튜닝, 언어 생성
학 번 : 2020-24835

목 차

제 1 장 서 론	1
제 2 장 선행연구	8
제 1 절 프롬프트 기반 학습.....	8
제 2 절 텍스트 추상 요약 모델.....	21
제 3 장 제안 방법	26
제 1 절 프롬프트 설계	27
제 2 절 요약 모델 구축.....	31
제 4 장 실험 및 결과	33
제 1 절 실험 데이터.....	33
제 2 절 실험 설계.....	42
제 3 절 실험 결과 및 분석	45
제 5 장 결론	49
참고문헌	51
Abstract	55

표 목차

[표 1.1] 추상 요약 예시 (XSum 데이터셋)	3
[표 2.1] Ding et al.[8] 연구의 프롬프트 학습 적용 과업.....	19
[표 4.1.1.1] Nallapati et al.[14] 연구 결과 예시	35
[표 4.1.2.1] XSum 데이터셋 예시 [15]	37
[표 4.2.1] 데이터 셋 통계	42
[표 4.3.1] CNN/DM 데이터셋 실험 결과	46
[표 4.3.2] XSum 데이터셋 실험 결과	46
[표 4.3.3] Risk10k 데이터셋 실험 결과	47
[표 4.3.4] 실험 결과 예시	48

그림 목차

[그림 2.1] 프롬프트 방법론 개요[1]	9
[그림 2.2] Brown et al.[2] 연구의 실험 결과	10
[그림 2.3] Prefix-tuning의 모델 구조[3]	12
[그림 2.4] P-tuning의 모델 구조[5]	13
[그림 2.5] PTR의 모델 구조[6]	15
[그림 2.6] 그림 2.6 PPT 프레임워크의 조정 방법 프레임워크[7]	16
[그림 2.7] 그림 2.7: Gu et al.[7] 연구의 파인 튜닝 프롬프트 튜닝 성능 비교	17
[그림 2.8] OpenPrompt모델 구조[8]	18
[그림 2.9] PEGASUS의 모델 구조[10]	22
[그림 2.10] SimCLS의 모델 구조[11]	23
[그림 2.11] FactPEGASUS의 모델 구조[19]	24
[그림 3.1.1] 하드 프롬프트 작동 방식	27
[그림 3.1.2] 소프트 프롬프트 작동 방식	27
[그림 3.1.3] Li, X. L., & Liang, P.[3] 연구의 프롬프트 사용 도식	27
[그림 3.1.4] Zhang et al.[12] 연구의 모델 구조	28
[그림 3.2.1] BART의 모델 구조[16]	32
[그림 4.1.3.1] 10-k 보고서 목차 예시	39
[그림 4.1.3.2] 10-K 보고서 item 1A 예시	39
[그림 4.1.3.3] Risk10k 데이터 예시	40

제 1 장 서 론

텍스트 과업 중 추상 요약은 최근 그 유용성을 크게 인정받고 있다. 추상 요약은 입력 텍스트의 의미와 주요 정보를 파악해 새로운 문장을 생성한다. 이는 단순히 원본 텍스트에서 문장을 추출하는 발췌 요약과 달리 더욱 원문 텍스트의 의미를 유지 및 강화한 요약을 제공할 수 있다. 한편 저작권 관련 갈등과 분쟁은 갈수록 증가하는 추세이다. 이로 인해 많은 양의 텍스트 데이터로부터 추출한 인사이트를 타인과 공유할 시 불편함을 느끼는 경우 또한 발생한다. 추상 요약은 이와 같이 텍스트 데이터를 활용할 시 효율성을 더하는 하나의 방안이 될 수 있으며, 저작권 관련 갈등을 피하고 양질의 창작물을 생성할 수 있도록 돕는다. 뿐만 아니라 텍스트 원문과 동일한 단어 및 어절을 사용하지 않기에 추상요약을 통해 더욱 인간 친화적이며 이해가 쉬운 요약문의 생성이 가능하다.

한편, 좋은 질의 추상 요약(abstractive summarization) 결과물을 얻기 위해서는 높은 비용이 요구된다. 특히 추상 요약은 발췌 요약(extractive summarization)에 비해 단순히 문서에서 중요한 단어, 혹은 문장을 추출하여 정리하는 것이 아닌 문서를 이해하고 새로운 어절 혹은 단어를 사용하여 문서를 요약, 정리해야 하기 때문에 생성 모델에 속하며, 동시에 까다로운 과업으로 꼽힌다.

현재 요약 과업에서 많이 사용되는 방식은 발췌 요약으로, 구현이 단순하며 높은 정확도를 보여 현실 세계에서 많이 사용되는 모델이 많다. 그러나 발췌 요약 모델만으로 얻은 결과물은 기존 문서에 사용된 단어 등을 제한적으로 사용해야 한다는 특성으로 인해 독해에 불편함이 있거나, 효율적이지 않은 경우가 많다. 이에 최근 발췌 요약과 추상 요약을 융합하여 질 높은 요약문을 생성하는 과업에 대한 연구가 많아지고 있다. 추상 요약은 발췌 요약에 비교해 문법적으로 읽기 쉽고 일관성이 있으며, 응집력 있는 요약을 생성하는데 도움이 될 뿐 아니라 요약문 자체의 중복 단어 및 문장의 수를 줄여 더욱 인간 친화적인 결과물을 얻을 수 있다. 마지막으로 저작권 관련 문제를 해결하고 새로운 텍스트/창작물을 생성하는 데 있어 강점을 가진다. 표 1.1 에서 추상 요약 과업을 위한 XSum 데이터셋의 데이터, 원문 텍스트와 추상 요약문을 각각 확인할 수 있다.

Original Text	<p>Star Wars fans across the globe have a new reason to rejoice. It was announced yesterday that a new trilogy of Star Wars films is in the works, with the first installment slated to be released in 2026. The trilogy will be directed by acclaimed director Ava DuVernay and written by celebrated science fiction author N.K. Jemisin. The trilogy is expected to take place in a previously unexplored part of the Star Wars universe, and will introduce a host of new characters. The announcement came as a surprise to many fans who had believed that the previous installment, The Rise of Skywalker, would be the last in the long-running series.</p>
Summary	<p>A new trilogy of Star Wars films directed by Ava DuVernay and written by N.K. Jemisin is set to be released in 2026</p>

표 1.1 추상 요약 예시 (XSum 데이터셋)

사전학습 된 언어모델을 특정 과업에 적용하기 위해 파인튜닝 하는 과정은 문장 분류, 감성 분석, 요약을 포함한 모든 NLP 분야 연구의 큰 축이 되었다. 그러나 전체 모델을 특정 데이터, 혹은 특정 과업에 맞게 파인튜닝 하는 것은 언어모델의 크기가 커져감에 따라 부담이 가중되는 작업이다. 또한 전체 모델을 파인튜닝하는 것은 데이터셋이 작은 few-shot 과업 등에 실현이 어렵고, 적은 데이터에 맞게 조정하기 위해서는 과적합을 피하기 위한 신중한 고려가 요구된다. 특히 파인튜닝 방법론을 few-shot 예제에 적용하면 좋지 않은 성능을 보이고, 특히 본 논문이 주제로 하는 추상 요약(abstractive summarization) 등의 생성 과업에서 큰 한계를 보임이 밝혀졌다.

최근 프롬프트 기반 학습은 위 서술한 파인튜닝의 단점을 보완하면서도 좋은 성능을 낼 수 있는 방법으로 많은 관심을 받고 있다. 성능에 대한 이론적 증명이나 근거가 부족함에도 불구하고, 언어 과업에서 프롬프트 엔지니어링은 놀라운 퍼포먼스를 보이며 특히 파인튜닝이 한계를 보였던 생성 과업 등에 유용함이 밝혀졌다.[11] 일반적으로 이러한 프롬프트를 사용한 방법론은 거대 언어모델에서 관련 지식을 뽑아내고, 프롬프트만 튜닝하여 새로운 과업에 빠르게 접목할 수 있도록 함으로써 효과를 검증해왔다. 프롬프트 엔지니어링은 파인튜닝에 비해 훨씬 적은 리소스와 시간을 사용하면서 파인튜닝에 버금가는, 혹은 더 나은 성능을 보이며 적은 learnable 파라미터를 사용한다.

이러한 프롬프트들은 크게 하드 프롬프트와 소프트 프롬프트 두 가지로 구분할 수 있다. 하드 프롬프트는 모델에게 명확하고 명시적인 지침을 제공하는 프롬프트 유형이며, 인간이 이해하기 쉽다. 모델의 동작을 정확히 조정하고 원하는 출력을 생성하는 데 도움을 준다. 반면 소프트 프롬프트는 조금 더 유연성을 가지며 모델이 과업 수행을 위한 기존 입력 텍스트에서의 암묵적인 신호를 파악할 수 있는 프롬프트 유형이다. 주어진 텍스트로부터 암묵적인 힌트를 찾아내어 모델의 동작을 조절하는 데 사용된다. 인간이 이해할 수 없는 형태의 벡터이지만 유연하고 하드 프롬프트에 비해 제약이 적으며 동시에 좋은 성능을 보인다는 장점을 가진다.

최근 프롬프트 튜닝이 프롬프트를 활용한 언어모델의 조정 방법론으로 떠오르고 있다. 최소 백 억 개 이상의 파라미터를 가지는 거대 언어모델에서 프롬프트 튜닝은 단순히 인코더의 인풋에 조정 가능한 소프트 프롬프트를 덧붙여 전체 모델 튜닝의 성능에 버금가는 성능을 보였다.[18] 덧붙인 소프트 프롬프트만을 튜닝하였을 때, 거대 언어모델에서 사용 가능한 지식을 추출하고 과업 및 데이터에 맞게 프롬프트를 조정하는 방식을 통해 전체 모델 튜닝에는 못 미치지만 좋은 성능을 보이며, 큰 발전 가능성을 기대할 수 있음을 해당 논문에서 확인하였다.

이 중 하이브리드 프롬프트 튜닝은 역전파를 통해 임베딩이 학습되는 연속적인 소프트 프롬프트에 하드 프롬프트의 특성을 적용한다. 하드 프롬프트에 비해 소프트 프롬프트는 언어 이해 및 생성 과업에서 좋은 성능을 보여왔다. 그러나 소프트 프롬프트는 하드 프롬프트에 비해 인간이 이해할 수 없다는 단점을 가지고 있으며, 따라서 사용자 제어 가능성이 저하되는 특징을 가진다.[5] 하이브리드 프롬프트 튜닝을 통해 각 프롬프트 방법론의 약점을 보완할 수 있다.

한편, 추상 요약은 추상 요약을 위한 훈련 데이터셋의 부족으로 어려움을 겪는 과업 중 하나이다. 충분한 양과 좋은 질의 훈련 데이터를 수집하는 것은 번거로운 작업이며, 특히 도메인 특화 요약을 위해서는 특정 분야의 전문 지식을 담은 데이터가 필요할 수 있다. 현재 공개된 데이터셋은 CNN/DM, XSUM, Multi-News, GIGAword 등으로 모두 뉴스 기사와 그를

요약한 쌍의 형태로 구축되어 있다. 따라서 본 논문에서는 금융 도메인 특화 데이터셋을 구축하여 데이터셋의 편향을 완화했을 뿐 아니라 오히려 도메인 특화 데이터셋에서 모델 성능이 더욱 높음을 확인하였다. 추상 요약의 장점 중 하나는 특정 분야의 특화된 요약을 생성했을 때 요약의 질이 상대적으로 높다는 것이다. 의학, 법률, 과학 등의 분야에서 도메인 전문지식 활용해 해당 분야의 텍스트를 요약할 수 있는데, 본 논문에서는 과업 수행을 위해 금융 도메인 데이터셋을 구축하였으며 이에 대해 모델이 양질의 요약문을 생성하는 것을 실험을 통해 확인하였다.

본 논문에서는 프롬프트 기반 학습을 통해 추상 요약을 수행하기 위한 방법론을 제안한다. 이 때 소프트 프롬프트와 하드 프롬프트를 융합하여 활용하고, 후술할 하이브리드 방식으로 추상 요약과 같은 언어 생성 과업에서 본 모델이 좋은 성능을 낼 수 있음을 보인다. 뿐만 아니라 요약 과업에 맞게 프롬프트를 텍스트의 단위를 기준으로 두 단계로 나누어 적용하여 프롬프트 기반 학습 방식의 특정 과업 활용에서의 방법과, 그 장점을 소개한다.

본 연구의 공헌은 다음과 같다. 본 논문은 프롬프트 기반 학습 방식을 통해 누구나 접근 가능한 거대 언어모델의 필요 지식을 추출해 특정 과업, 혹은 데이터에 적용할 수 있는 모델을 제안한다. 텍스트 분류, 감성 분석 등의 언어 이해 과업 뿐 아니라 언어 생성 과업에 프롬프트 기반 학습 방식을 적용하였을 때도 높은 성능을 보임을 확인하였다. 특히 추상 요약에 맞는 프롬프트

기반 학습 방식 활용할 시 보다 창의적이고 효율적인 텍스트 요약 수행할 수 있음을 증명한다. 또한 공개 데이터셋이 부족한 추상 요약 과업에 접목하기 위해 본 논문에서 구축한 금융 텍스트 추상 요약 데이터셋을 소개한다. 이를 통해 CNN/DM, XSum 등 추상 요약 학습에 사용되는 데이터셋 뿐 아니라 도메인 지식이 강한 금융 텍스트 등에도 모델이 효율적으로 작동함을 보인다. 마지막으로 본 논문의 방법론은 요약 외 타 언어 생성 과업으로의 확장성이 높다는 점에서 의의가 있다.

본 논문은 5 장으로 구성된다. 이후 제 2 장에서는 추상 요약과 프롬프트 기반 학습 방식에 관한 선행 연구를 소개한다. 제 3 장에서는 본 논문에서 제안하는 프롬프트 기반 추상 요약 모델의 방법론을 제안한다. 제 4 장에서는 실험에 사용된 데이터셋, 새로 구축한 데이터셋 등을 확인하고 실험 결과를 살펴본다. 마지막으로 제 5 장에서는 결론과 본 연구의 한계점 및 향후 연구 방향을 제시한다.

제 2 장 선행연구

제 1 절 프롬프트 기반 학습

프롬프트 기반 학습은 텍스트의 확률을 직접 모델링하는 언어 모델에 기반한 머신러닝의 유형 중 하나이다. 전통적 지도 학습이 입력 값 x 를 받아 출력 값 y 를 $P(x|y)$ 의 형태로 예측하는 모델을 훈련하는 형태였다면, 프롬프트 기반 학습은 입력 값 x 를 특정 템플릿을 사용해 일부 빈 슬롯이 존재하는 텍스트 x' 으로 수정한다. 그 후 거대 언어 모델을 사용하여 빈 정보를 확률적으로 채워 최종 텍스트를 유도하는 형태이다.[1] 즉, 프롬프트 기반 학습은 템플릿을 사용해 입력 값을 수정하고 확률적 방법으로 누락된 정보를 채우는 것으로, 예측 작업을 수행하기 위해 언어 모델을 사용하는 것이다. 이러한 접근 방식은 텍스트 분류, QA 및 언어 생성 등 다양한 자연어 처리 과업에 효과적인 것으로 입증되었다.

Liu et al.[1]의 연구에 따르면 프롬프트 방법론의 연구를 그림 2.1 과 같이 도식화 할 수 있다.

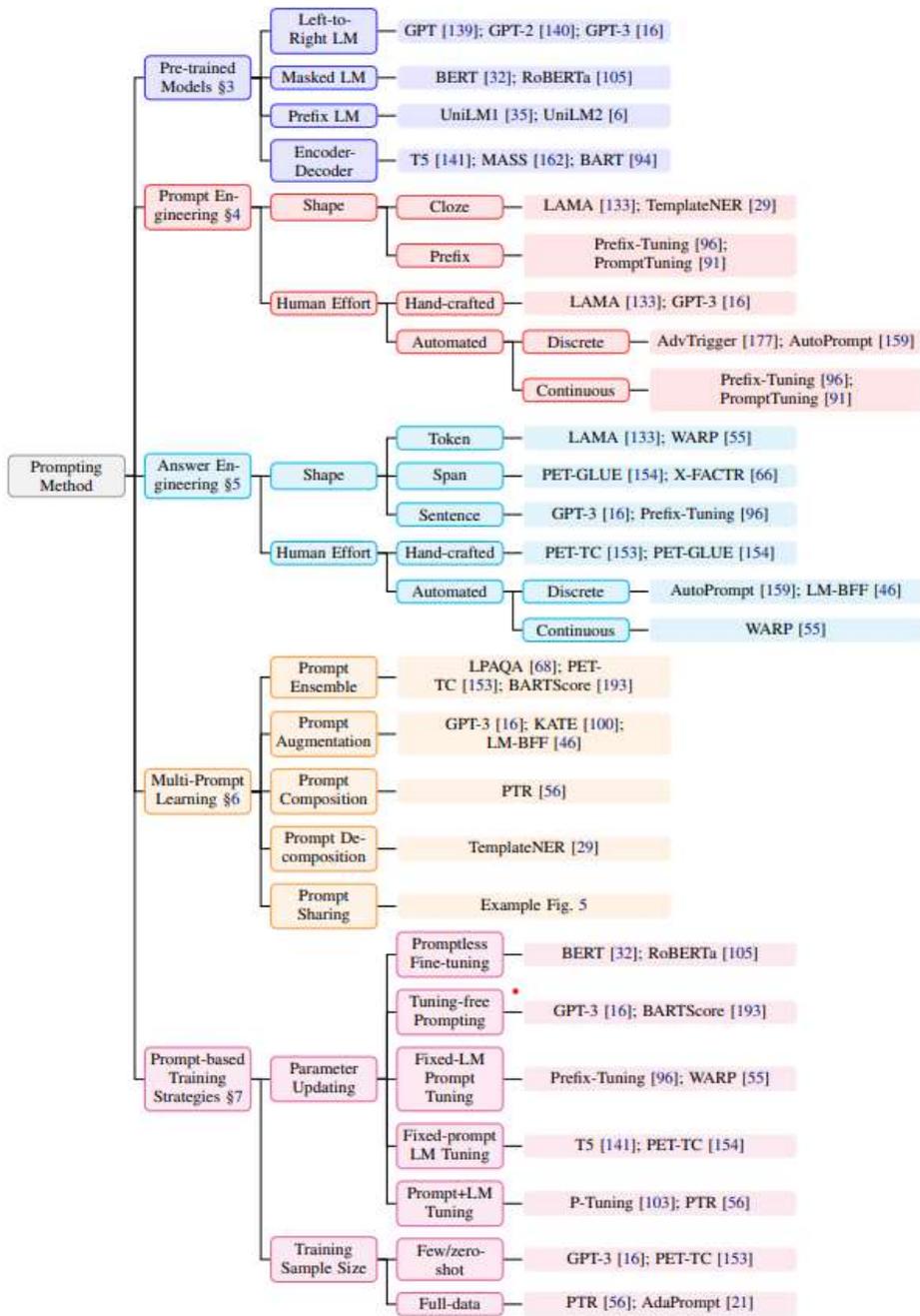


그림 2.1 프롬프트 방법론 개요 [1]

프롬프트 기반 언어 모델에 대한 관심 증가와 추후 연구는 GPT-3 모델의 발표 이후 활성화되었다. 해당 모델을 발표한 Brown et al.[2]에 따르면 모델이 수행해야 하는 특정 과업에 대한 자연어 설명은 모델의 성능을 향상시킨다. 그림 2.2 를 통해 이를 확인할 수 있다.

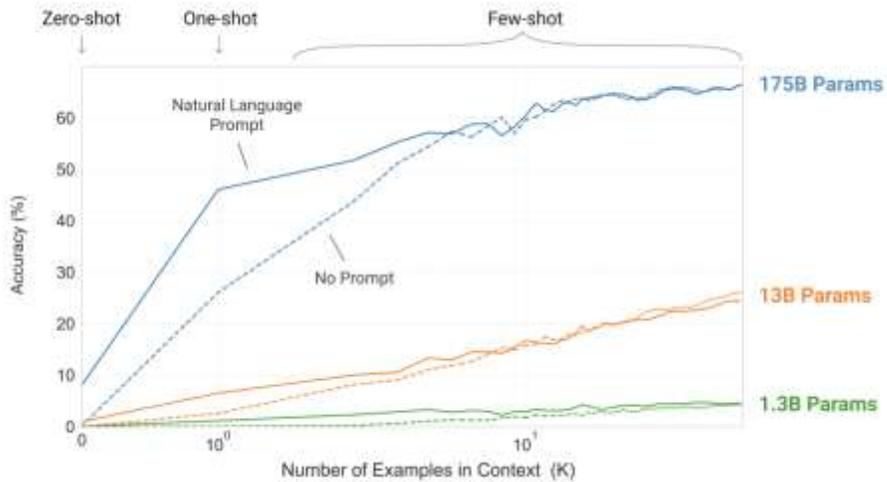


그림 2.2 Brown et al.[2] 연구의 실험 결과

Brown et al.[2]는 zero-shot 셋팅에서의 실험을 통해 prompt 의 성능을 증명하고 추후 프롬프트 기반 연구의 시작점이 되었다는 데 큰 기여점을 가지지만, 논문 발표 이전 NLP 분야의 연구에서 문제되어 왔던 많은 한계를 돌파했다는 의의 또한 가진다. (1) 새로운 과업에 언어 모델을 적용해야 할 때마다 레이블링 된 데이터가 필요하다는 한계, (2) 사전학습 단계에서 거대한 데이터 코퍼스의 지식을 흡수했음에도 특정 과업, 혹은

데이터에 적용하기 위한 파인 튜닝 단계에서 다운스트림 과업의 분포를 따르게 되며 과적합이 발생하거나 일반화 성능을 상실할 수 있다는 한계, (3) 적은 데이터로 task 를 수행할 수 없다는 한계를 일정 부분 극복했다.

Li, X. L., & Liang, P.[3]은 프롬프트 기반 학습 방식에서 영감을 받아 모델에 특정 프롬프트를 제공하여 텍스트를 생성하는 방법을 제안한다. 기존의 파인 튜닝 방식이 많은 비용이 들 뿐 아니라 레이블링 된 데이터가 충분해야 한다는 점을 지적하며, 연속적인 과업 특정 벡터라고 정의할 수 있는 prefix 를 최적화하는 경량화 된 대안을 제안한다. 해당 논문은 벤치마크 모델의 0.1%에 해당하는 매개변수만을 대상으로 한 학습으로 prefix-tuning 이 투입 리소스 및 데이터 대비 우수한 성능을 보인다고 주장한다. 또한 prefix-tuning 은 데이터셋이 적은 환경에서 파인 튜닝보다 우수한 성능 발휘하고, 나아가 학습 중 접하지 못한 주제의 텍스트에 대해서도 더 나은 결과를 보인다는 것을 증명한다. Prefix-tuning 은 그림 2.3 으로 간단히 도식화 할 수 있다.

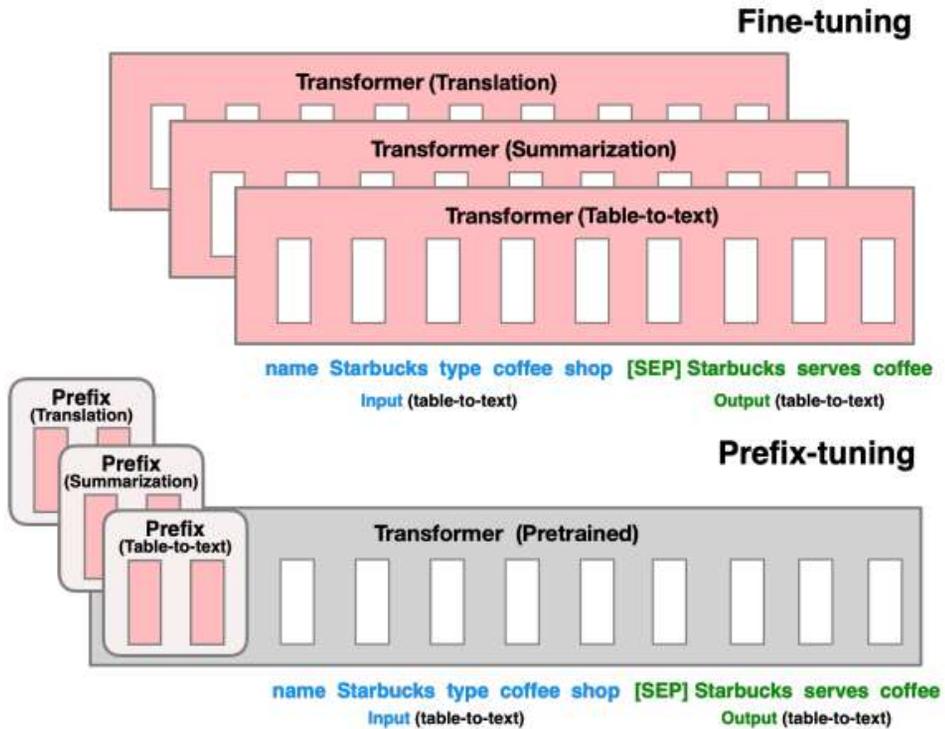


그림 2.3 Prefix-tuning 의 모델 구조 [3]

이전 연구에서 소개된 기존 프롬프트와 달리 prefix 는 실제 토큰을 포함하지 않는 파라미터들로 구성되며, prefix 만을 최적화하는 형태로 학습이 진행되기 때문에 효율적 학습이 가능하다. 해당 논문은 Brown et al.[2]에서 제안한 매뉴얼하게 디자인된 프롬프트를 적용하는 과업마다 생성해야 하는 in-context 학습 방식의 한계를 지적한다.

Prefix-tuning [3]과 같은 방법론을 사전 훈련된 거대 언어모델을 고정하고 다운스트림 과업의 학습 데이터에서 얻은 신호를 사용해 프롬프트의 매개변수만 업데이트하는 언어 모델

고정 상태의 프롬프트 튜닝이라고 정의할 수 있다.[1] 이러한 방식은 few-shot 환경에 적합하지만 zero-shot 환경에는 적합하지 않다. 이와 달리 기존의 사전 학습 및 파인튜닝과 유사하게 모델 자체의 매개변수를 조정하지만, 고정된 매개변수를 가지는 프롬프트를 사용해 모델 동작을 지정하는 방법론도 존재한다. 이 또한 프롬프트 기반 학습의 한 유형으로, 프롬프트를 사용해 모델의 동작을 지정하고 입력값을 수정해 출력값을 예측한다.[4]

프롬프트 기반 학습 방식으로 대표적인 방법론 중 하나로 P-tuning 을 꼽을 수 있다.[5] 언어모델의 입력값에 프롬프트를 인코더를 배치하고 이를 통해 나온 출력값을 프롬프트의 토큰 임베딩으로 사용한다. 여기에 과업과 관련된 고정 토큰을 추가해 성능을 개선하는 방식으로, 모델 구조는 그림 2.4 를 통해 확인할 수 있다.

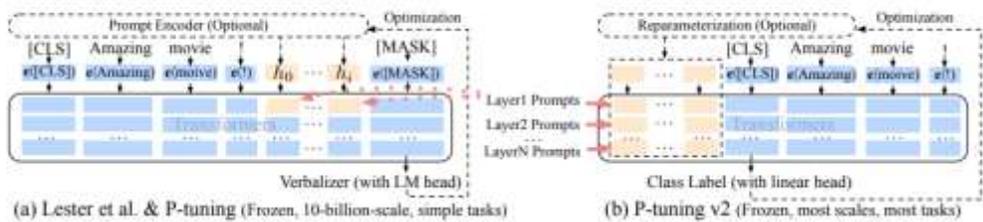


그림 2.4 P-tuning 의 모델 구조[5]

그림 2.4 는 두 가지 모델의 구조를 표현하고 있는데, (a)의 경우는 단순히 학습 가능한 연속적 임베딩 벡터를 입력값의 시퀀스에 포함시켜 파라미터를 조정하는 방식이다. 이러한 모델 구조의 변경은 파인 튜닝 방식에 비교해 특히 단순 텍스트 분류 등의 문제에서 유의미한 성능을 보였다.[5] 그러나 해당 방식은 과업, 혹은 매개변수의 크기에 따라 보편성이 크게 영향을 받는다는 한계점이 있다. 이에 Liu et al.[5]에서는 한계를 보완하기 위한 깊은 프롬프트 튜닝 방식을 제안한다. 이는 그림 2.4 의 (b)에서 확인할 수 있다. 프롬프트들은 각 레이어에 prefix 토큰의 형태로 더해진다. 이 방식은 조정 가능한 특정 과업에서 동작하는 매개변수의 수를 증가시켜 보다 효과적이고 과업에 적합한 결과값을 얻을 수 있도록 한다. 이 때 프롬프트의 길이는 단순 분류 과업의 경우 20, 생성 등의 복잡한 과업은 100 정도의 길이를 제안한다.

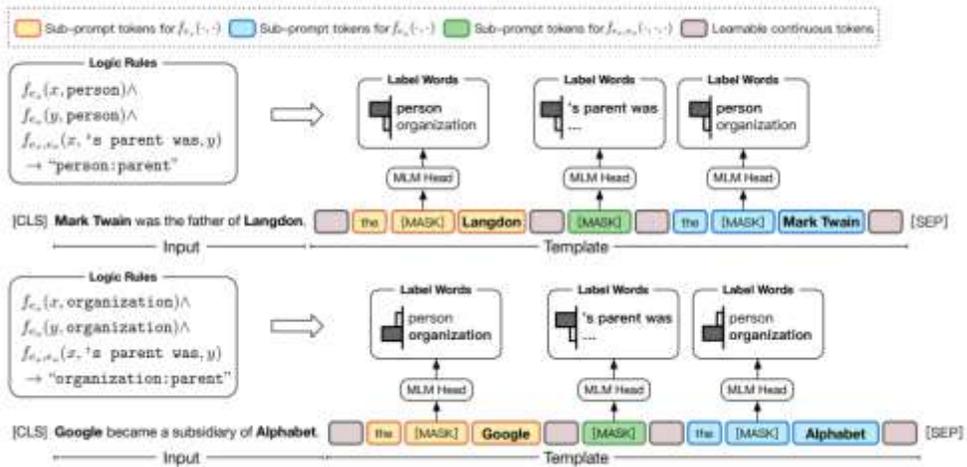


그림 2.5 PTR의 모델 구조[6]

Han et al.[6]은 규칙 기반 프롬프트 튜닝(PTR)이라는 새로운 방법을 제안한다. 해당 논문은 인간의 지식을 포함시켜 거대 언어 모델을 다운스트림 과업에 보다 효과적으로 적용할 수 있다고 주장한다. 추가 모델 레이어나 데이터 레이블링, 데이터 증강 등의 기법을 쓰지 않고 기존 SOTA 모델을 넘어서는 결과를 보여준다. PTR은 간단한 하위 프롬프트들을 조합해 특정 과제에 적합한 프롬프트를 구성하고, 이를 논리 규칙에 따라 조합한다. PTR의 구조는 그림 2.5를 통해 확인할 수 있다.

PTR[6]은 각각 하나의 특정 기능을 수행하는 하위 프롬프트를 설계하여 인간의 논리 추론 과정을 모방하고자 하였으며, 여러 프롬프트를 설계하고 융합하는 방식은 추후 본 논문에서 다른 프롬프트의 설계에 참고하였다. PTR은 텍스트 분류 과업에서 효과적인 성능을 보였으며, 하위 프롬프트를 개체

기반으로 설계하여 해당 과업의 최적화에 알맞게 동작하도록 유도하였다.

Gu et al.[7]의 논문은 기존 프롬프트 튜닝을 개선하기 위한 방법론으로, 사전 학습 단계에서 소프트 프롬프트를 추가해 프롬프트 튜닝의 성능을 개선한다. 이를 구현하기 위해 해당 논문에서 제안하는 프레임워크인 Pre-trained Prompt Tuning(PPT)는 먼저 다운스트림 과업을 몇 가지 형식으로 통합한다. 그 후 각 형식에 대해 자기 지도학습 방식의 사전학습 과업을 설계하고, 각 과업에서 프롬프트를 사전 학습하는 과정을 거친다. 마지막으로, 사전 학습의 결과를 기반으로 파라미터를 초기화하여 다운스트림 과업에서 프롬프트 튜닝을 수행한다. 즉, PPT는 프롬프트 튜닝의 초기값 설정 방식을 개선해 few-shot 학습 성능을 향상시켰으며, 궁극적으로 거대 언어모델과 다운스트림 과업 간 간극을 줄이기 위한 프롬프트를 설계하고 프롬프트 튜닝의 유용성을 증명한다.

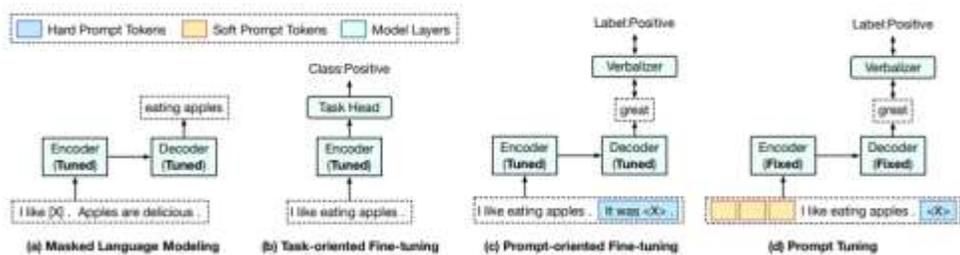
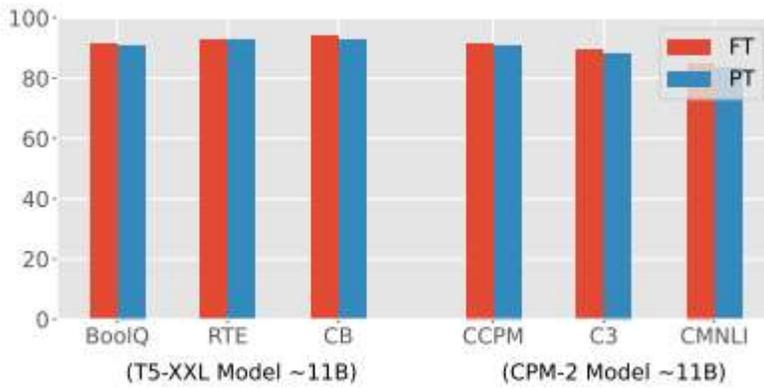
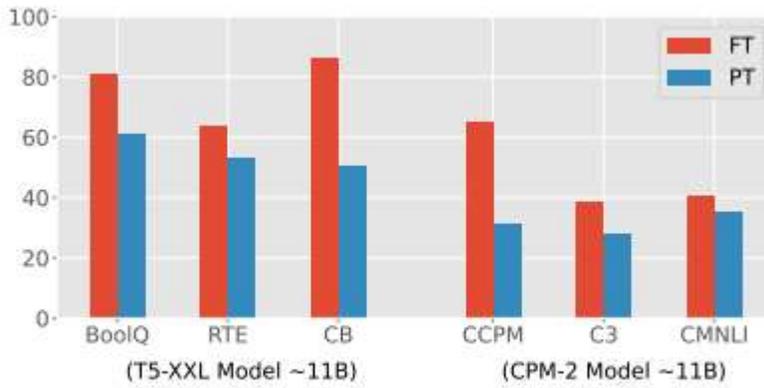


그림 2.6 PPT 프레임워크의 조정 방법 프레임워크 [7]



(a) Full-Data



(b) Few-Shot

그림 2.7 Gu et al.[7] 연구의 파인 튜닝 프롬프트 튜닝 성능 비교

Gu et al.[7]의 연구에서는 그림 2.7 을 통해 프롬프트 튜닝의 유의함도 확인할 수 있다. 100 개의 학습 가능한 임베딩으로 이루어진 토큰을 프롬프트로 설계하여 학습을 진행했을 때, 프롬프트 튜닝 방식의 경우 파인 튜닝이 110 억개의 파라미터를 최적화해야 하는 것에 비교해 단지 41 만개의 파라미터만을 학습함에도 전체 데이터셋 환경에서 성능의 큰 차이를 보이지 않는 것을 살펴볼 수 있다.

프롬프트 학습의 일반적 프레임워크를 제안한 Ding et al.[8]은 기존 프롬프트 학습을 구현하는 데 발생하는 문제들을 해결하기 위한 오픈소스 프레임워크와 코드를 제안한다.

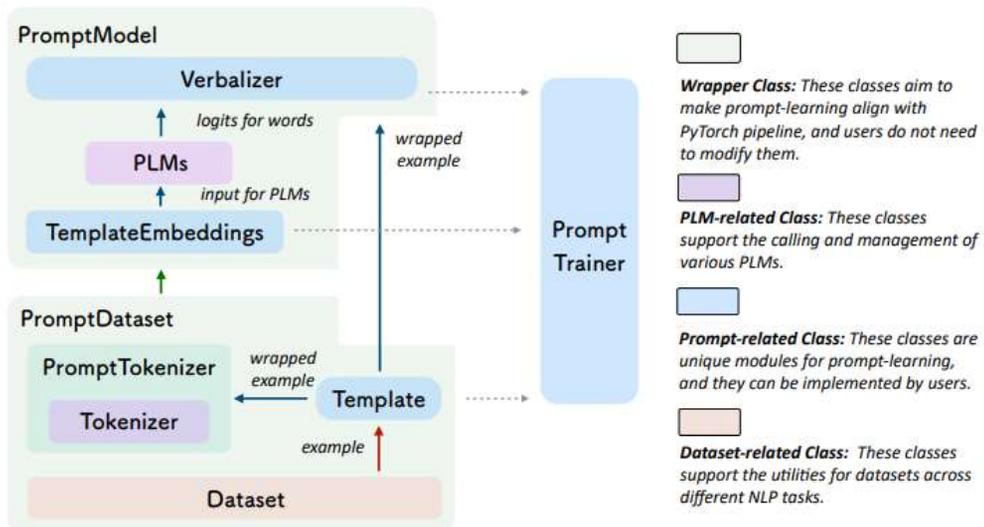


그림 2.8 OpenPrompt 모델 구조 [8]

그림 2.8 을 통해 살펴볼 수 있듯, Ding et al.[8]는 프롬프트 학습의 전체 라이프 사이클을 제안한다. 해당 프롬프트 프레임워크는 과업, 거대 언어모델, 템플릿과 verbalizer 를 선택하여 사용자의 기호에 맞게 모델을 구현할 수 있게 지원한다. 또한 표 2.1 을 통해 제안하는 프레임워크의 실제 사용 예제와 설정 값 등을 확인할 수 있다. 해당 논문은 프롬프트 기반 학습에 대한 관심의 증가를 증명하지만 간이 실험을 통해 특정 다운스트림 과업에 맞게 설계된 프롬프트 기반 학습 방법론 Han et al.[6]의 연구의 성과에는 미치지 못함을 확인할 수 있었다.

Example	PLM	Template	Verbalizer	Task	Reference
Naive TC	MLM & Seq2Seq	M. text	M. One-Many	Text Classification	-
Naive KP	LM & Seq2Seq	M. text	-	Knowledge Probing	-
Naive FET	MLM	M. text (meta info)	M. One-Many	Entity Typing	(Ding et al., 2021a)
PTR	MLM	M. text (complex)	M. One-One	Relation Extracrtion	(Han et al., 2021b)
P-tuning	LM	Soft tokens	M. One-One	Text Classification	(Liu et al., 2021b)
Prefix-tuning	LM, Seq2Seq	Soft tokens	-	Text Generation	(Li and Liang, 2021)
LM-BFF	MLM	A. text	M. One-Many	Text Classification	(Gao et al., 2021)

표 2.1 Ding et al.[8]연구의 프롬프트 학습 적용 과업

Wu et al.[20]의 연구는 자연어 이해 분야에서 소프트 프롬프트 튜닝을 위한 정보 이론적인 프레임워크인 InfoPrompt 방법론을 제안한다. InfoPrompt 는 기존 프롬프트 튜닝 방법보다 더 나은 속도와 성능을 보이는데, 이를 통해 사전 전문 지식 없이도 InfoPrompt 를 사용해 프롬프트 튜닝 수행이 가능하다. 두 가지의 새로운 상호 정보기반 손실 함수를 개발 및 사용하는 방법을 제안하는데, 첫 번째로 다운스트림 과업에 적합한 프롬프트 초기값을 찾고 프롬프트 토큰에서 충분한 과업 관련 정보를 학습한 다음, 두 번째로 사전 훈련 된 언어 모델의 결과값이

학습된 프롬프트에서 포착된 과업 관련 정보를 더욱 잘 인지하도록 한다. 해당 논문은 여러 실험을 통해 InfoPrompt 가 프롬프트 튜닝의 성능을 증가함을 보이고 있다.

제 2 절 텍스트 추상 요약 모델

추상 요약 과업에 관련하여 가장 유명한 논문 중 하나는 Zhang et al.[10]으로, 해당 논문을 통해 PEGASUS 라는 모델을 제안하였다. 사전 훈련 및 파인 튜닝 프레임워크를 사용해 추상적인 텍스트 요약에 대한 새로운 접근 방식을 사용했으며, 간결하고 일관성 있는 요약 텍스트를 생성하는 데 좋은 성능을 보인다.

PEGASUS [10]는 비지도 사전학습 방식과 지도 학습을 사용한 파인 튜닝 두 단계 과정을 거쳐 과업을 해결한다. 사전 훈련 단계에서 PEGASUS 는 뉴스 기사에서 누락된 문장을 예측하는 방법을 학습한다. 입력 텍스트에서 임의로 가려진 문장을 의미하는 'Gap 문장'을 활용한다. 사전 훈련 단계를 통해 해당 모델은 주어진 텍스트의 맥락과 의미를 이해하는 데 유용한 모델이 된다. 이 후 텍스트-요약문 쌍으로 구성된 레이블링 데이터에 대한 파인 튜닝 작업을 수행한다. 모델의 아키텍처는 그림 2.9 를 통해 살펴볼 수 있다.

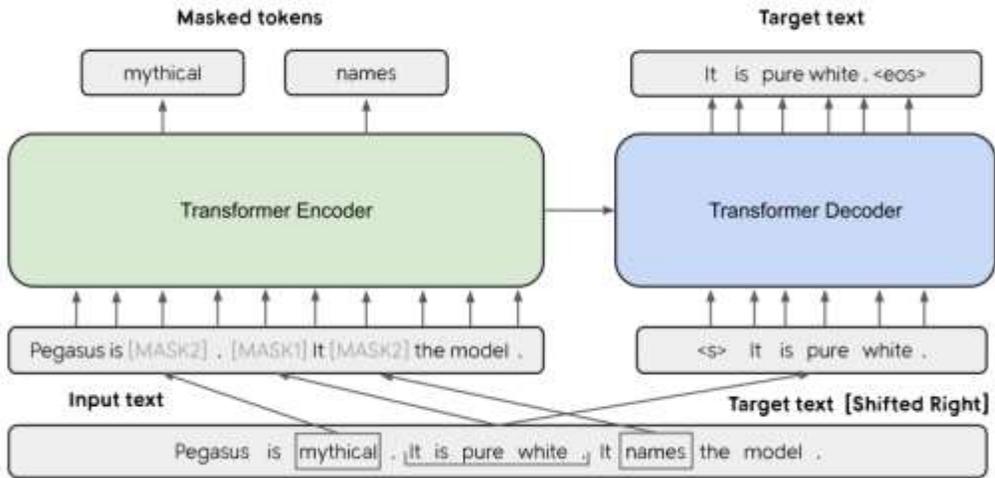


그림 2.9 PEGASUS 의 모델 구조[10]

이 모델은 seq2seq 모델을 기반으로 하며, 파인튜닝 단계에 사용되는 인코더가 입력 텍스트를 처리하고 디코더는 인코딩 된 표현을 기반으로 요약물을 생성한다. 논문이 발표될 당시 요약 모델 중 가장 좋은 성능을 보였으며, 특히 과업에 맞는 특별한 파인튜닝 방법론을 정립해 사용했다는 의의를 가진다. 즉, 해당 논문은 추상 요약이라는 특정 과업을 위해 Gap 문장 생성이라는 새로운 사전학습을 수행하였으며, 이 후 파인튜닝 과정에서의 적은 리소스 비용으로 대부분의 데이터셋에서 가장 좋은 성능을 보였다.

SimCLS 는 추상 요약 과업에 대조 학습 기법을 적용해 우수한 성능을 보이는 모델이다.[11]. 우선 seq2seq 모델을 이용한 추상 요약 학습 방식에 대한 두 가지 한계를 제시한다. 첫째로 목적 함수와 평가 지표 간의 차이 존재로 인해 최대의 성능을 끌어내지 못한다. 두 번째로, teacher forcing 학습 방식으로 인한 노출 편향이라는 한계가 존재한다. 모델 훈련 시 디코더에 입력 값으로 요약문의 정답과 이전 결과값이 주어지는 반면 테스트 시에는 이전 결과값만이 주어지기 때문에 훈련과 테스트 단계의 차이를 초래하게 된다. 이러한 노출 편향이 존재하면 테스트 시 이전 단계에서의 오류가 계속 누적되기 때문에 성능 저하의 원인이 될 수 있다.

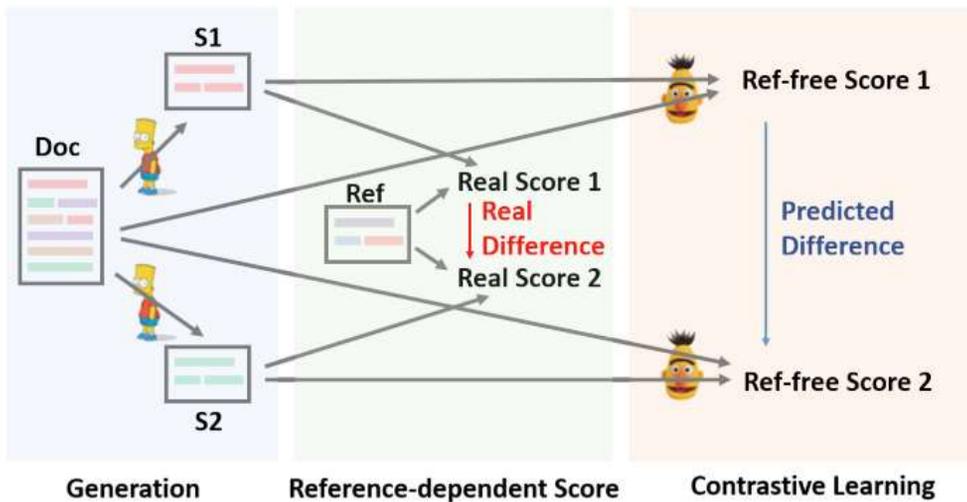


그림 2.10 SimCLS 의 모델 구조[11]

앞서 서술한 한계점을 극복하기 위해 해당 논문에서는 두 단계로 이루어진 프레임워크를 제안한다. 첫 번째로 seq2seq 모델이 후보 요약문을 생성하기 위해 훈련된다. 두 번째 단계는 대조학습을 통해 생성된 후보 요약문에 순위를 매기는 평가 단계로 구성된다. 그림 2.10 통해 두 단계로 구성된 모델의 전체적인 도식도를 확인할 수 있다.

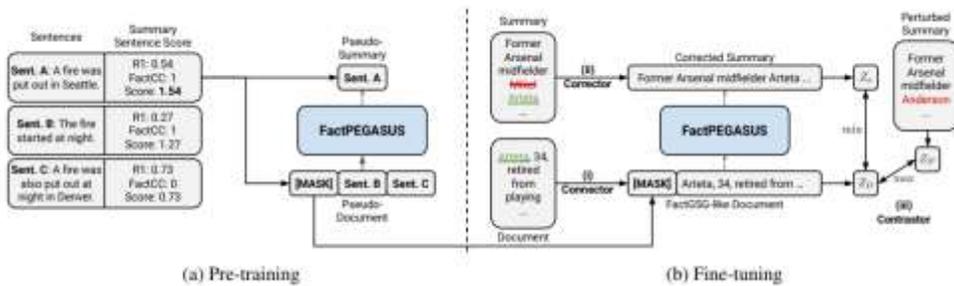


그림 2.11 FactPEGASUS의 모델 구조[19]

Wan et al.[19]의 연구는 사전 훈련 및 파인 튜닝 동안 사실성 (factuality)의 문제를 해결하는 추상 요약 모델 FactPEGASUS를 제안한다. 기존 PEGASUS 모델의 문장 선택 작업을 강화하여 보다 사실적인 요약을 만드는데 집중하였다. 또한 파인 튜닝 작업 개선을 위한 세 가지 구성요소를 제시한다. 첫 번째로 수정자(corrector)는 기존 요약 정답문에 있는 환각을 제거한다. 대조자(contrastor)는 비사실적 요약과 사실적 요약의 구별을 더욱 확실하게 하기 위해 존재한다. 마지막으로 연결자(connector)는 지식을 기존 모델보다 더 잘 전달하기 위해 사전학습과 파인튜닝 간 격차를 해소하는 데 집중한다. 해당 연구에서 소개한 모델은 사실적이고 지식이 강화된 요약문을

생성한다. 그림 2.11 은 FactPEGASUS 의 모델 구조를 도식화한 그림이다.

제 3 장 제안 방법

제 2 장에서 살펴본 선행 연구를 통해 프롬프트 기반의 학습 방법은 텍스트 분류 과업 등의 언어 이해 과업에 주로 적용되어 왔음을 확인할 수 있다. 또한 추상 요약 과업에의 딥러닝 적용 연구 등을 살펴보았다. 이에 본 연구에서는 프롬프트 기반 학습을 추상 요약 과업에 접목하여 모델 성능을 개선하고 모델 구축 및 학습의 비용을 절감할 수 있는 방법론을 제안한다. 이 때 세 가지의 프롬프트를 사용하며, 프롬프트의 초기값을 알맞게 설정하기 위해 자기 지도 방식을 이용한다. 마지막으로 해당 입력 값을 요약 모델에 넣는 과정으로 진행된다.

프롬프트 기반 학습 방식을 추상 요약 과업에 접목하기 위해 기존 프롬프트 기반 학습 방식을 부분 변경하여 진행하였다. 전체 프레임워크를 요약하면 아래의 그림 3.1 과 같다.

다음 절에서는 프롬프트 설계 방식과 초기값 설정을 위한 자기 지도 선행 훈련 모델, 요약 모델로 나누어 서술한다.

제 1 절 프롬프트 설계

언어 이해 과업과 달리 언어 생성 과업에 특화된 프롬프트를 생성하기 위해, [3]의 연구에서 제안하는 prefix-tuning 의 방식에서 차용한 소프트 프롬프트를 텍스트 원문 앞에 배치한 후 인코더에 넣는다. 이 방법을 통해 거대 언어모델의 파라미터를 고정할 때 프롬프트만을 최적화하며 리소스가 크게 줄어든다. 그림 3.1.1, 3.1.2 를 통해 소프트 프롬프트와 하드 프롬프트를 모델에 적용하는 방식을 확인할 수 있다.

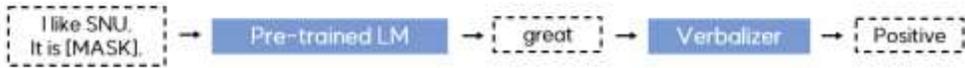


그림 3.1.1 하드 프롬프트 작동 방식

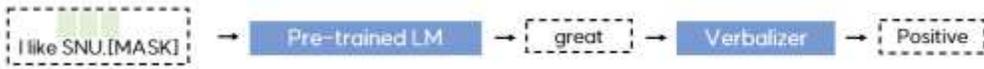


그림 3.1.2 소프트 프롬프트 작동 방식



그림 3.1.3 Li, X. L., & Liang, P.[3] 연구의 프롬프트 사용 도식

그림 3.1.3 은 [3]의 프롬프트 사용 예제를 그림으로 나타낸 것이다. 이와 같이 본 연구에서는 요약하기 전의 텍스트 앞에 소프트 프롬프트를 붙여 프롬프트만 조정하도록 설정했다.

소프트 프롬프트와 하드 프롬프트의 장점을 모두 취할 수 있는 하이브리드 모델의 구축을 위해 요약문 길이를 제한하는 형태의 하드 프롬프트를 구성하였다. Zhang et al.[12]는 제어가능한 요약 모델의 구축을 위해 여러가지 하드 프롬프트를 사용하였는데, 제어 변수와 프롬프트 예시는 다음과 같다. 해당 논문에서는 프롬프트를 직접 구축하여 사용하였다.

'요약문 제어 변수: 프롬프트 예시'의 형태로 살펴보고자 한다. (1)요약문에 반드시 포함되어야 할 키워드: keywords (a,b,c) summarize, (2) abstractiveness : summarize with abstractiveness 30, (3)요약문의 길이: summarize with length 20, (4)여러가지 제어 변수의 융합: summarize with length 20, and abstractiveness 30 등으로 구분하였다. 본 연구는 이러한

형태의 프롬프트에서 조정가능한 변수로 길이를 선택하여 원문 텍스트의 하드 프롬프트로 사용할 수 있도록 모델을 구축하였다. 이 때 길이의 경우, 원문 텍스트와 요약문으로 구성된 데이터 쌍에서 요약문의 토큰 개수를 계산하여 프롬프트에 포함할 수 있도록 설계했다. 해당 방법론의 출처가 된 [12]의 모델 프레임워크는 그림 3.1.4 에서 확인할 수 있다.

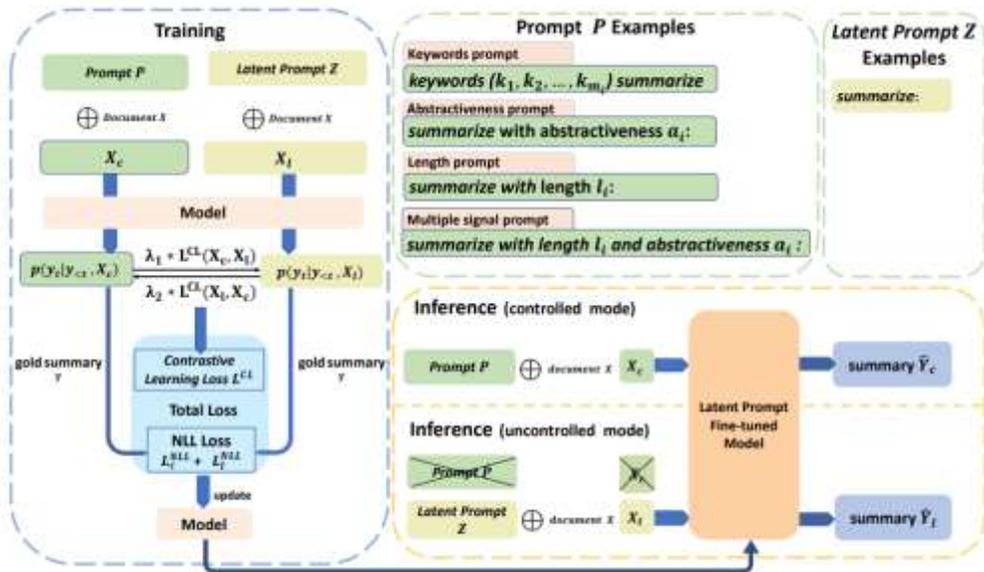


그림 3.1.4 Zhang et al.[12]연구의 모델 구조

즉, 본 논문은 연속적 벡터로 구성된 P_{soft} 와 이해 가능한 언어의 형태를 그대로 가지고 있는 P_{hard} , 그리고 텍스트 데이터를 concatenation 하여 모델의 입력값으로 사용하였다. P_{soft} 와 P_{hard} 이 합쳐진 프롬프트를 primary 프롬프트라고 정의하고, $P_{primary}$ 로 표현한다. 기존 데이터셋의 원문 텍스트 데이터를 X , 요약문을 Y 라고 한다면, 데이터셋의 형태는 궁극적으로 $(P_{primary} \oplus X, Y)$ 의 형태를 띄게 된다.

이 후 선행연구에서 다루었던 PSP[13]모델을 참고하여 elaborate 프롬프트를 설계하였다. [13]은 원문 텍스트의 보다 깊은 이해와 좋은 질의 요약 텍스트를 생성하기 위해 내부 프롬프트를 구성하였다. e 를 원문 텍스트를 특정 단위로 나눈 후 임베딩한 값, 내부 프롬프트를 p 라고 표현하면 새로운 내부 프롬프트만을 포함한 데이터셋의 입력값은 다음과 같이 구성된다.[13]

$$X'_{in} = \{e_1^1 + p_{in}^1, e_2^1 + p_{in}^1, \dots, e_j^i + p_{in}^i, \dots, e_m^n + p_{in}^n\} \quad (1)$$

[13]은 고정 길이를 기준으로 원문 텍스트를 나눈 후 내부 프롬프트를 추가하였다. 본 연구는 해당 방법론을 차용하되 원문 텍스트를 문장 단위로 나눠 내부 프롬프트, 본 연구에서 사용되는 용어로는 elaborate 프롬프트를 추가하였다. 이유로는 elaborate 프롬프트 외 primary 프롬프트 또한 존재하기 때문에 입력값의 크기를 줄이기 위함과 모델의 설명력을 갖추기 위함이다.

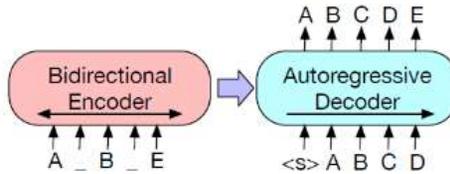
제 2 절 요약 모델 구축

BART 모델을 조정하는 형태로 프롬프트 기반 학습이 이루어졌다. BART 는 Lewis et al.[16]에서 발표한 모델로, 기존 BERT 가 생성 과업에서 사용이 어려운 점을 보완한 모델로 생성 과업의 베이스 모델로 많이 사용된다. BART 는 seq2seq 트랜스포머 구조를 사용했고, BERT 에서 사용했던 ReLU 활성화함수를 GeLUs 로 변경했다. 베이스 모델은 6 개의 레이어, 큰 모델은 12 개의 레이어를 사용했다. 기존 트랜스포머의 디코더 형태와 동일하게 디코더의 각 레이어에서 인코더의 마지막 히든 레이어와 크로스-어텐션 할 수 있도록 설계했다. BERT 의 경우 단어 예측 작업을 위해 feed-forward 레이어를 추가했으나, BART 는 이와 다르게 레이어를 새로 추가하지 않았다. 즉, BART 는 텍스트가 임의의 노이즈 함수를 통해 오염된다. 이를 seq2seq 모델이 복원을 위해 학습되는 방식으로 학습되었다. BART 의 모델 구조는 그림 3.3.1 에서 확인할 수 있다.



(a) BERT: Random tokens are replaced with masks, and the document is encoded bidirectionally. Missing tokens are predicted independently, so BERT cannot easily be used for generation.

(b) GPT: Tokens are predicted auto-regressively, meaning GPT can be used for generation. However words can only condition on leftward context, so it cannot learn bidirectional interactions.



(c) BART: Inputs to the encoder need not be aligned with decoder outputs, allowing arbitrary noise transformations. Here, a document has been corrupted by replacing spans of text with mask symbols. The corrupted document (left) is encoded with a bidirectional model, and then the likelihood of the original document (right) is calculated with an autoregressive decoder. For fine-tuning, an uncorrupted document is input to both the encoder and decoder, and we use representations from the final hidden state of the decoder.

그림 3.2.1 BART 의 모델 구조[16]

본 연구에서 제안하는 추상 요약 모델은 생성 과업에 유리한 BART 를 기반으로 설계되었으며, 기존 파인 튜닝 방법론의 경우 새로운 과업이 생길 때 마다 모든 파라미터를 조정해야 하나, 본 연구의 프롬프트 기반 학습 방법은 입력 값에 소프트 프롬프트를 결합해 추가된 임베딩만을 학습한다.

제 4 장 실험 및 결과

제 1 절 실험 데이터

4.1.1 CNN/DM

CNN/Daily Mail 데이터셋은 IBM 에서 RNN 을 이용한 추상 요약 모델을 구축하며 알려졌다[14]. 총 286,817 학습 데이터 쌍과 13,368 개의 validation 데이터 쌍, 그리고 11,487 개의 테스트 데이터 쌍으로 구성되어 있다. 해당 데이터셋은 평균적으로 29.74 개의 문장으로 이루어진 원문 텍스트를 보유하고 있으며, 요약문의 경우 평균 3.72 개로 이루어져 있다. 데이터셋 명에서 추측할 수 있듯 해당 데이터셋은 대규모 영어 뉴스 기사 데이터셋으로, 뉴스 웹사이트 CNN 과 Daily Mail 에서 수집한 기사와 그 요약문을 포함한다.

해당 데이터셋은 자연어 처리, 특히 요약 과업에 주로 사용되며, 다음과 같은 요소들로 구성된다. (1)원문 텍스트: 각 기사의 본문은 해당 기사의 모든 텍스트를 포함한다. 하나의 문단일 수도 있고, 여러 문단으로 구성될 수도 있다. (2)요약문: 각 기사의 핵심 내용을 요약한 하나 이상의 문장이다. 기사 본문의 중요 정보를 간결히 전달하도록 작성되었다. (3)기사 ID: 각 기사에 고유하게 할당된 식별자이다. CNN/DM 은 뉴스 웹사이트 크롤링 후 요약문 추출, 데이터 정제, 데이터 유형화의 단계로 구축되어 현재 누구나 접근 가능하도록 공개하고 있다.

아래 그림 4.1.1.1 에서 데이터 예시와 해당 논문에서 제안하는 모델의 결과값을 확인할 수 있다. T 가 정답 요약문, O 가 모델의 결과값이다.

<p>Good quality summary output</p> <p>S: a man charged with the murder last year of a british backpacker confessed to the slaying on the night he was charged with her killing , according to police evidence presented at a court hearing tuesday . ian douglas previte , ## , is charged with murdering caroline stuttle , ## , of yorkshire , england</p> <p>T: man charged with british backpacker 's death confessed to crime police officer claims</p> <p>O: man charged with murdering british backpacker confessed to murder</p>
<p>S: following are the leading scorers in the english premier league after saturday 's matches : ## - alan shearer -lrb- newcastle united -rrb- , james beattie .</p> <p>T: leading scorers in english premier league</p> <p>O: english premier league leading scorers</p>
<p>S: volume of transactions at the nigerian stock exchange has continued its decline since last week , a nse official said thursday . the latest statistics showed that a total of ##.### million shares valued at ###.### million naira -lrb- about #.### million us dollars -rrb- were traded on wednesday in , deals .</p> <p>T: transactions dip at nigerian stock exchange</p> <p>O: transactions at nigerian stock exchange down</p>
<p>Poor quality summary output</p> <p>S: broccoli and broccoli sprouts contain a chemical that kills the bacteria responsible for most stomach cancer , say researchers , confirming the dietary advice that moms have been handing out for years . in laboratory tests the chemical , <unk> , killed helicobacter pylori , a bacteria that causes stomach ulcers and often fatal stomach cancers .</p> <p>T: for release at #### <unk> mom was right broccoli is good for you say cancer researchers</p> <p>O: broccoli sprouts contain deadly bacteria</p>
<p>S: norway delivered a diplomatic protest to russia on monday after three norwegian fisheries research expeditions were barred from russian waters . the norwegian research ships were to continue an annual program of charting fish resources shared by the two countries in the barents sea region .</p> <p>T: norway protests russia barring fisheries research ships</p> <p>O: norway grants diplomatic protest to russia</p>
<p>S: j.p. morgan chase 's ability to recover from a slew of recent losses rests largely in the hands of two men , who are both looking to restore tarnished reputations and may be considered for the top job someday . geoffrey <unk> , now the co-head of j.p. morgan 's investment bank , left goldman , sachs & co. more than a decade ago after executives say he lost out in a bid to lead that firm .</p> <p>T: # executives to lead j.p. morgan chase on road to recovery</p> <p>O: j.p. morgan chase may be considered for top job</p>

표 4.1.1.1 Nallapati et al.[14] 연구 결과 예시

4.1.2 XSum

XSum 데이터셋은 Narayan et al.[15]에서 소개된 데이터셋이다. 204,045 개의 학습 데이터쌍, 11,332 개의 validation 데이터 쌍, 11,334 개의 테스트 데이터 쌍으로 구성되어 있다. BBC 뉴스 기사를 토대로 만들어졌으며 원문 텍스트는 평균 19.77 개의 문장으로 이루어진다. 요약문의 경우 1 개 문장으로 이루어져 있으며 평균 23.26 개의 단어 개수를 가진다. XSum 데이터셋은 4.1.1 에서 설명한 CNN/Daily Mail 데이터셋과 같이 원문 본문과 해당 기사의 짧은 요약문으로 구성된다. 그 후 데이터의 일관성과 정확성을 높이기 위해 전처리 과정을 거친 후 공개되었다. 단일 문서 요약 과업을 수행하고 평가, 및 훈련할 때 주로 사용된다. 데이터셋은 아래 그림 4.2 를 통해 예시를 확인할 수 있다.

SUMMARY: *A man and a child have been killed after a light aircraft made an emergency landing on a beach in Portugal.*

DOCUMENT: Authorities said the incident took place on Sao Joao beach in Caparica, south-west of Lisbon.

The National Maritime Authority said a middle-aged man and a young girl died after they were unable to avoid the plane.

[6 sentences with 139 words are abbreviated from here.]

Other reports said the victims had been sunbathing when the plane made its emergency landing.

[Another 4 sentences with 67 words are abbreviated from here.]

Video footage from the scene carried by local broadcasters showed a small recreational plane parked on the sand, apparently intact and surrounded by beachgoers and emergency workers.

[Last 2 sentences with 19 words are abbreviated.]

표 4.1.2.1 XSum 데이터셋 예시 [15]

4.1.3 Risk10k

본 연구의 수행을 위해 뉴스 기사로 구성된 4.1.1 과 4.1.2 의 데이터셋과 다른 데이터셋을 구축하였다. 2008 년 - 2018 년 10-k 리포트 34756 개의 데이터를 수집하였다. 리포트의 형식이 항상 동일함을 활용하여 item 1A 에 해당하는 risk factors 텍스트 데이터를 수집하였다.

Risk factors 는 사업에 위협이 될 상황 및 요소를 문단 형태로 작성하고 이를 간단히 요약한 문장을 볼드체로 작성하는 형태로 구성된다. 기존 텍스트 요약 모델은 주로 뉴스기사 데이터셋을 사용하여 학습했던 반면, 10-k 보고서를 이용해 본 연구에서 구축한 데이터셋의 경우 정부 기준에 맞춘 형태의 리포트에 포함된 문단과, 정확한 요약문 컬럼을 가지고 있어 기존 데이터셋과 다른 특징을 가진다. 해당 데이터셋을 risk10K 라고 정의한다.

Risk10k 의 사용으로 기존 학습에 사용된 데이터의 성격이 다르기에 모델의 범용성을 확대할 수 있고, 금융 도메인 지식을 가지는 텍스트에서도 본 요약모델이 의의를 가진다. 기존 텍스트 요약 모델의 경우 보통 데이터셋으로 뉴스 기사를 사용하는 데 비해 본 데이터셋의 경우 공식적으로 정부에 제출할 포맷에 맞게 작성된 보고서에 포함되는 문단과, 정확한 요약 문을 명시해 기존 데이터셋과 차별점을 가진다. 따라서 본 모델은 공식 레포트와 같은 문서의 추상 요약 과업에 적합하다. 데이터셋은 82905 개의 문단과 요약문으로 구성되었다. 아래 그림 4.1.3.1, 그림

4.1.3.2 는 실제 10-K 보고서의 목차와, item 1A 에 해당하는 risk factors 콘텐츠 예시이다.

TABLE OF CONTENTS		Page
PART I		
Item 1.	Business	1
Item 1A.	Risk Factors	12
Item 1B.	Unresolved Staff Comments	24
Item 2.	Properties	25
Item 3.	Legal Proceedings	26
Item 4.	Mine Safety Disclosures	26
PART II		
Item 5.	Market for Registrant's Common Equity, Related Stockholder Matters and Issuer Purchases of Equity Securities	27
Item 6.	Selected Financial Data	28
Item 7.	Management's Discussion and Analysis of Financial Condition and Results of Operations	29
Item 7A.	Quantitative and Qualitative Disclosures about Market Risk	50
Item 8.	Financial Statements and Supplementary Data	52
Item 9.	Changes in and Disagreements with Accountants on Accounting and Financial Disclosures	93
Item 9A.	Controls and Procedures	93
Item 9B.	Other Information	93
PART III		
Item 10.	Directors, Executive Officers and Corporate Governance	94
Item 11.	Executive Compensation	95
Item 12.	Security Ownership of Certain Beneficial Owners and Management and Related Stockholder Matters	95
Item 13.	Certain Relationships and Related Transactions, and Director Independence	96
Item 14.	Principal Accounting Fees and Services	96
PART IV		
Item 15.	Exhibits, Financial Statement Schedules	97
Item 16.	Form 10-K Summary	97

그림 4.1.3.1 10-k 보고서 목차 예시

We could be adversely affected if any of our existing aircraft are underutilized or we fail to redeploy or deploy aircraft with customers at favorable rates. We could also be adversely affected from the loss of one or more of our aircraft for an extended period of time.

Our operating revenues depend on our ability to effectively deploy the aircraft in our fleet and maintain high utilization of our aircraft at favorable rates. If we have underutilized aircraft, we would seek to redeploy those aircraft in our other lines of business or sell them. If we are unable to successfully redeploy our existing aircraft at favorable rates or sell them on favorable terms, it could have a material adverse effect on our business, results of operations and financial condition. In addition, if one or more of our aircraft are out of service for an extended period of time, our operating revenues would decrease and we may have difficulty fulfilling our obligations under one or more of our existing contracts. The loss of revenue resulting from any such business interruption, and the cost and potentially long lead time and difficulties in sourcing a replacement aircraft, could have a material adverse effect on our business, results of operations and financial condition.

Our financial condition may suffer if we experience unexpected costs as a result of ongoing lawsuits, claims and investigations related to alleged pricing practices or other legal and regulatory matters.

In the United Kingdom, several groups of named claimants have brought suit against British Airways Plc ("British Airways") in connection with alleged improper matters related to the use of fuel surcharges and other cost components for air cargo services and are seeking damages allegedly arising from that conduct. British Airways has filed claims in the lawsuit against Polar Air Cargo LLC ("Old Polar"), formerly Polar Air Cargo, Inc., a consolidated subsidiary, and other carriers for contribution should British Airways be found liable to claimants.

In the Netherlands, Stichting Carel Costenbayer, successor in interest to claims of various shippers, has filed suit in the district court in Amsterdam against British Airways, KLM, Martinair, Air France, Lufthansa and Singapore Airlines seeking recovery for damages purportedly arising from the same pricing practices at issue in the proceeding described above. In response, British Airways, KLM, Martinair, Air France and Lufthansa filed third-party indemnification lawsuits against Old Polar and Polar seeking indemnification in the event the defendants are found to be liable in the main proceedings.

그림 4.1.3.2 10-K 보고서 item 1A 예시

그림 4.1.3.3 은 실제 구축한 데이터셋의 예시이다. 한 문단에 짧은 요약문이 매칭되어 있다. 해당 데이터를 살펴보면, 10k 를 작성한 주체인 회사의 비즈니스가 영화 배급사에 의존해 수익 창출이 이루어지고 있고, 반독점 사건 관련한 법령 등이 영화 배급에 영향을 미치기 때문에 사업에 부정적 영향을 줄 수 있다고 설명한다. 또한 현재 배급사의 사정에 따라 장기 계약을 체결할 수 없어 영화 공급이 보장되지 않으며, 배급사와의 관계가 매출에 중요한 영향을 끼치는 만큼 이러한 배급사와의 관계가 해당 회사의 비즈니스에 부정적 영향을 미칠 수 있다고 서술되어 있다. 또한 이에 해당하는 요약문을 살펴보면 영화 배급사와의 관계 약화는 상업적으로 성공할 영화를 얻게 될 해당 회사의 능력에 부정적 영향을 끼칠 수 있다고 요약되어 있다.

* Our results of operations fluctuate on a seasonal basis. Our results of operations vary from period to period based upon the quantity and quality of the motion pictures that we show in our theatres. The major film distributors generally release the films they anticipate will be most successful during the summer and holiday seasons. Consequently, we typically generate higher revenues during these periods. The timing of releases, however, has become less pronounced as distributors have begun releasing content more evenly throughout the year. In our Latin American markets, while Hollywood content has similar release dates as in the U.S., the local holidays and seasons can vary. The unexpected emergence of a successful film during other periods or the failure of an expected success at a key time could alter this seasonality trend. Due to the dependency on the success of films released from one period to the next, results of operations for one period may not be indicative of the results for the following period or the same period in the following year.

* A deterioration in relationships with film distributors could adversely affect our ability to obtain commercially successful films.

We rely on the film distributors to supply the films shown in our theatres. The film distribution business is highly concentrated, with seven major film distributors accounting for approximately 88% of U.S. box office revenues and 49 of the top 50 grossing films during 2017. Numerous antitrust cases and consent decrees resulting from the antitrust cases impact the distribution of films. Film distributors license films to exhibitors on a theatre-by-theatre and film-by-film basis. Consequently, we cannot guarantee a supply of films by entering into long-term arrangements with major distributors. We are therefore required to negotiate licenses for each film and for each theatre. A deterioration in our relationship with any of the seven major film distributors could adversely affect our ability to obtain commercially successful films and to negotiate favorable licensing terms for such films, both of which could adversely affect our business and operating results.

* We face intense competition for patrons and films which may adversely affect our business. The motion picture exhibition industry is highly competitive. We compete against local, regional, national and international exhibitors in many of our markets. We compete for both patrons and licensing of films. In markets where we do not face nearby competitive theatres, there is a risk of new theatres being built. The degree of competition for patrons is dependent upon such factors as location, theatre capacity, quality of projection and sound equipment, film showtime availability, customer service quality, products and amenities offered, and ticket prices. The principal competitive factors with respect to film licensing include the theatre's location and its demographics, the condition, capacity and grossing potential of each theatre, and licensing terms. We also face competition from new concept theatres such as dine-in theatres and tavern style theatres that open in close proximity to our conventional theatres. If we are unable to attract patrons or to license successful films, our business may be adversely affected.

그림 4.1.3.3 Risk10k 데이터 예시

기존 추상 요약 모델들은 주로 4.1.1 의 CNN/DM 데이터셋[14]과 4.1.2 의 XSum 데이터셋[15]으로 구성되는데, 이들은 서술하였듯 뉴스 기사를 기반으로 한 데이터셋이다. 뉴스 기사의 특징은 초반 2-3 줄에 기사의 핵심 내용이 압축되어 서술된다는 점이다. 따라서 데이터에 편향이 존재하고, 이를 risk10k 데이터셋의 사용으로 상쇄할 수 있다. 뿐만 아니라 상대적으로 길이가 짧은 경우가 많기 때문에 모델이 원문 텍스트의 길이에 구애 받지 않고 요약문을 생성할 수 있는 유연성 또한 확보할 수 있다.

실제 ChatGPT 를 비롯 현존하는 요약 모델의 경우 요약문을 길게 생성하는 편으로, risk10k 데이터셋은 요약문의 길이가 공개되어 있는 데이터셋들에 비해 짧게 구성되었기 때문에 데이터의 편향성을 감소시킨다. 추가적으로 해당 데이터셋은 본 연구와 같이 추상 요약 과업에도 사용될 수 있지만, 원문 텍스트와 요약문 모두 금융 의사결정의 측면에서 보았을 때 부정적 의미를 내포하는 경향이 있기 때문에 금융 도메인 특화 감성 분석 과업 등에도 사용할 수 있다.

제 2 절 실험 설계

실험은 두 가지로 구성하였다. 첫 번째 실험에서 BART를 기 서술한 방법론을 적용해 조정한 후 PEGASUS 모델과 SimCLS 모델, BART를 파인튜닝 한 모델과 비교하였다. 이 때 BART 모델은 base 모델과 large 중 large를 택해 활용하였으며, Adam Optimizer를 사용하였다. 비교 대상이 되는 PEGASUS와 SimCLS는 제 2장 선행연구에 포함되어 있다. 모델과 데이터셋은 모두 Huggingface 사이트에서 쉽게 다운받고 적용할 수 있다. 제 1절에서 서술한 본 실험에 사용된 데이터셋을 표로 표 4.2.1에서 살펴볼 수 있다.

	CNN/DM	XSum	Risk10k
학습 데이터 수	287,226	204,045	70905
테스트 데이터 수	11,490	11,803	7000
검증 데이터 수	13,368	11,513	5000
원문 텍스트 평균 길이	781 단어	399 단어	175 단어
기준 요약문 평균 길이	774 단어	34 단어	30 단어

표 4.2.1 데이터 셋 통계

Primary 프롬프트의 기존 소프 프롬프트는 200 토큰으로 구성하였으며, 프롬프트와 원문 텍스트를 융합하는 과정에서 원문 텍스트의 오른쪽에 소프 프롬프트를 배치하는 것이 왼쪽에 배치하는 파일럿 실험을 돌렸을 때보다 조금 더 나은 성능을 보였기에 오른쪽에 배치하였다.

프롬프트 기반 학습과 파인튜닝의 성능을 비교하기 위해 CNN/DM과 XSum 데이터셋을 대상으로 10 에폭으로 실험하였으며, Risk10k의 경우 상대적으로 용량이 작아 20 에폭으로 실험하였다. 실험 도중 learning rate에 성능이 큰 영향을 받는다는 점을 발견하였고, 이에 착안하여 다양한 learning rate로 파일럿 실험을 수행한 후 XSum 데이터에 대해서는 0.005를, CNN/DM 데이터와 Risk10k 데이터에 대해서는 0.5의 learning rate를 설정하였다.

평가 지표로 ROUGE-1, ROUGE-2, ROUGE-L 점수를 사용하였다. ROUGE(Recall-Oriented Understudy for Gisting Evaluation) 점수는 요약 결과의 품질을 평가하기 위해 사용되는 평가지표이다. ROUGE 점수는 생성된 요약문과 기존 요약문 사이 유사성을 측정하여 요약의 품질을 평가한다. 주로 텍스트 요약 모델의 성능 평가에 사용되는 점수이다. ROUGE-1 점수는 단일 토큰 레벨에서의 유사성을 측정한다.[17] ROUGE-2는 bigram 단에서의 유사성을 측정하며, 연속된 두 단어의 시퀀스를 고려해 기존 요약문과 생성된 요약문 사이 유사성을 측정한다. ROUGE-L은 Longest Common Sequence를 사용해 재현율과 정밀도를 조합한 F1 점수를 계산한다. 단어 순서 및 문맥을 고려한다는 장점을 가진다

두 번째 실험은 적은 데이터 환경에서 모델의 작동을 관찰하기 위해 수행하였다. CNN/DM 데이터셋을 분할해 50개, 100개, 300개, 500개의 데이터가 있을 때 파인튜닝과 프롬프트 기반 학습의 차이를 비교하고자 했다. 각 데이터셋 크기에 대해 10개의 샘플 데이터셋을 추출해 성능을 평균 계산하여 수행했다.

제 3 절 실험 결과 및 분석

본 절은 세 가지의 데이터셋에 BART를 파인튜닝한 모델, PEGASUS, SimCLS, 본 연구에서 제안하는 모델을 접목하여 실험한 결과와 그 분석으로 구성된다. ROUGE-1, ROUGE-2, ROUGE-L 점수를 성능 평가의 지표로 활용하였다. 마지막으로 적은 데이터셋 환경에서의 성능 비교를 위해 데이터셋 크기 별 파인튜닝 된 모델과 프롬프트 기반 학습된 모델의 성능을 비교한다.

먼저 CNN/DM 데이터셋으로 실험한 결과를 표 4.3.1에서 살펴볼 수 있다. 본 연구에서 제안하는 프롬프트 기반 학습방식은 전체 모델 파인튜닝에서 사용하는 파라미터의 0.3%만을 사용했음에도 불구하고 필적할 만한 성능을 보이는 것을 확인할 수 있다. 특히 타 점수에 비해 ROUGE-L 점수가 높는데, 본 연구에서 제안하는 방법론은 elaborate 프롬프트를 설계하여 사용한 만큼 원문 텍스트의 문맥이나 순서가 잘 고려되었음을 확인할 수 있다. 특히 CNN/DM의 경우 본 연구에서 실험에 사용한 데이터셋 중 원문 텍스트와 요약문의 평균 길이가 가장 긴 데이터셋으로, 세계의 데이터셋 중 본 방법론이 가장 적합한 데이터셋이라는 것을 예상할 수 있다.

Model	Param(M)	R-1	R-2	R-L
Finetuned BART	400M	44.07	22.56	38.58
PEGASUS	568M	43.18	22.70	41.23
SimCLS	400M	44.45	22.17	40.09
Ours	0.6M	40.14	19.07	37.81

표 4.3.1 CNN/DM 데이터셋 실험 결과

표 4.3.2를 보면, XSum 데이터셋으로 실험하였을 때에는 비록 PEGASUS와 SimCLS의 결과에 미치지 못하는 못하나 파인튜닝된 BART보다 R-1 점수가 미세하게 더 높은 것을 확인할 수 있다. 조정에 사용된 파라미터의 개수를 고려하여 시간과 비용의 절감 측면을 생각한다면 프롬프트 기반 학습 방식이 과업 수행에 굉장히 적합함을 알 수 있다. 특히 단순 프롬프트 튜닝 방법론 등을 사용하기 보다 프롬프트를 과업에 맞게 변형하여 적용하였을 때, 과업 특화 모델을 빠르고, 저렴하고 쉽게 구축할 수 있음을 짐작할 수 있다.

Model	Param(M)	R-1	R-2	R-L
Finetuned BART	400M	42.10	20.77	38.15
PEGASUS	568M	45.70	25.22	40.91
SimCLS	400M	44.75	23.88	39.72
Ours	0.6M	42.14	20.81	35.77

표 4.3.2 XSum 데이터셋 실험 결과

마지막으로 Risk10k 데이터셋을 활용하여 실험한 결과를 표 4.3.3에서 확인할 수 있다. 이 데이터셋의 경우 타 데이터셋에 비해 전반적으로 성능이 높는데, 해당 데이터셋은 모두 동일하게 10-K 보고서에서 크롤링해 구축한 것으로, 요약문을 작성하는 방식이 어느 정도 규격화되어 있기 때문으로 예상된다. 역시 프롬프트 기반 학습이 좋은 성능을 보이며 특히 ROUGE-L 점수가 높음을 확인할 수 있다. 표 4.3.4를 통해 원문 데이터와 기준 요약문, 각 모델의 결과값을 확인할 수 있다.

Model	Param(M)	R-1	R-2	R-L
Finetuned BART	400M	46.89	22.12	40.75
PEGASUS	568M	46.55	21.91	39.44
SimCLS	400M	45.18	22.71	40.15
Ours	0.6M	43.71	21.96	40.01

표 4.3.3 Risk10k 데이터셋 실험 결과

종합하면 프롬프트 기반 학습 방식은 파인튜닝 대비 정확도 측면에서 상대적으로 부족한 성능을 보이나, 사용하는 파라미터 개수의 급격한 차이로 특화 모델 구축의 비용을 현저히 줄일 수 있음을 확인할 수 있다. 본 모델의 경우 프롬프트 기반 학습 활용 시 기존 파인튜닝의 0.15%에 해당하는 파라미터만을 조정한다. 또한 정부 문서, 보고서 등의 규격화 된 문서를 요약하는 데 사용되었을 때 그 유용성이 돋보인다.

Passage	We have greatly expanded our operations over the last decade through targeted worldwide theatre development and acquisitions. We continue to pursue a strategy of expansion that will involve the development of new theatres and may involve acquisitions of existing theatres and theatre circuits both in the U.S. and internationally. There is significant competition for new site locations and for existing theatre and theatre circuit acquisition opportunities. As a result of such competition, we may not be able to acquire attractive site locations, existing theatres or theatre circuits on terms we consider acceptable. The pace of our growth may also be impacted by delays in site development caused by other parties. Acquisitions and expansion opportunities may divert a significant amount of management's time away from the operation of our business. Growth by acquisition also involves risks relating to difficulties in integrating the operations and personnel of acquired companies and the potential loss of key employees of acquired companies. Our expansion strategy may not result in improvements to our business, financial condition, profitability, or cash flows. Further, our expansion programs may require financing above our existing borrowing capacity and operating cash flows. We may not be able to obtain such financing or ensure that such financing will be available to us on acceptable terms or at all.
Reference	We are subject to uncertainties relating to future expansion plans, including our ability to identify suitable acquisition candidates or new theatre site locations, and to obtain financing for such activities on favorable terms or at all.
Finetuned BART	The company has pursued a strategy of expansion over the past decade through worldwide theatre development and acquisitions, but there is significant competition for new site locations and existing theatre circuits, which may result in delays in site development, the diversion of management's time, and risks associated with integrating the operations of acquired companies and financing the expansion programs.
Pegasus	We continue to pursue a strategy of expansion that will involve the development of new theatres and may involve acquisitions of existing theatres. There is significant competition for new site locations and for existing theatre and theatre circuit acquisition opportunities. Our expansion strategy may not result in improvements to our business, financial condition, profitability, or cash flows
SimCLS	Our company has expanded globally through targeted theatre development and acquisitions. However, competition for desirable locations and acquisitions is intense. Delays and integration challenges may impact our growth. Our expansion strategy may not yield desired improvements, and financing may be uncertain.
Ours	Our future expansion plans are subject to risk, including our capacity to locate new theater locations or viable acquisition candidates, as well as our ability to get funding for such activities on favorable terms or even at all.

표 4.3.4 실험 결과 예시

제 5 장 결론

본 논문에서는 언어 생성 과업 중 하나인 추상 요약 과업을 프롬프트 기반 학습 방식을 적용하여 모델을 구축하였다. 특히 기존의 프롬프트 기반 학습 방식과 달리, 목표하는 과업에 적합한 프레임워크의 구축을 위해 primary 프롬프트와 elaborate 프롬프트로 나누어 프롬프트를 설계하였으며, primary 프롬프트를 소프트 프롬프트와 하드 프롬프트를 융합하여 사용해 성능을 개선하였다.

프롬프트 기반 학습 방식을 활용하여 실험한 결과로 전체 모델 파인 튜닝 대비 적은 시간과 비용을 사용해 유의한 성능 향상을 보임을 확인할 수 있었다. 특히 ROUGE-L 점수가 타 점수에 비해 높은 것으로 미루어 보아 본 연구에서 제안하는 모델이 문맥 및 순서를 고려한 요약문 생성에 유리함을 확인하였다.

따라서 본 연구는 프롬프트의 개념을 추상 요약이라는 특정 과업에 접목하여 특화된 모델을 개발했다는 점, 소프트 프롬프트와 하드 프롬프트를 결합한 하이브리드 방식으로 성능을 끌어올렸다는 점, 새로운 금융 도메인 데이터셋을 구축하여 도메인 특화 데이터셋에서도 모델이 유의하게 작동한다는 것을 증명했다는 점에서 의의가 있다. 특히 새로 구축한 Risk10k 데이터셋은 다른 과업 및 연구에도 확장되어 사용할 수 있을 것으로 기대된다.

본 연구에서 진행한 실험의 결과로 미루어 보아 프롬프트

튜닝은 적은 데이터셋, 특정 분야 혹은 특정 과업에 해당하는 데이터셋이 적을 때 활용할 수 있는 효율적인 방법임을 알 수 있다. 타 데이터셋에 비해 규모가 작은 Risk10k 데이터셋의 경우 오히려 가장 좋은 성능을 보였다. 이를 통해 현실 세계에서 직면 가능한 특정 업무를 위한 언어 모델을 구축할 때, 공개된 거대 언어모델에 프롬프트 튜닝을 도입하는 것 만으로 파인튜닝에 비교해 적은 시간과 자원을 활용해 특화된 인공지능 언어 모델을 구축할 수 있음을 예상할 수 있다. 이 때 P-tuning과 같이 이미 공개된 프롬프트 기반 학습 모델을 그대로 차용하는 것도 유용하나, 본 연구와 같이 과업에 알맞은 프롬프트를 추가 구현하여 적용한다면 더 좋은 성능의 모델이 구축될 것으로 기대된다.

본 연구의 한계는 다음과 같다. 첫 번째, 성능 지표로 사용되는 ROUGE-1, ROUGE-2, ROUGE-L 점수의 경우 문맥 이해, 문장의 의미, 인간 친화적 설명력의 부족이라는 비판을 받는 만큼 성능을 평가하는 지표의 부족함을 들 수 있다. 두 번째, 프롬프트 사전 학습을 진행하지 않아 더 큰 성능 향상을 이루지 못했다는 한계가 존재한다.

향후 연구로는 먼저 자기지도 학습방식을 통해 프롬프트의 사전 학습을 수행하여 보다 유의한 결과를 도출할 수 있는 프롬프트의 설계가 필요하다. 또한 본 연구에서 제안하는 모델에 기반하여 다른 생성 과업에 활용할 수 있는 방법론에 대한 논의가 필요할 것이다. 마지막으로 추상 요약에 알맞은 다양한 데이터셋으로 실험을 수행하여 모델의 정합성을 증명하고자 한다.

참고 문헌

- [1] Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9), 1–35.
- [2] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33, 1877–1901.
- [3] Li, X. L., & Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*.
- [4] Schick, T., & Schütze, H. (2020). Exploiting cloze questions for few shot text classification and natural language inference. *arXiv preprint arXiv:2001.07676*.
- [5] Liu, X., Ji, K., Fu, Y., Tam, W., Du, Z., Yang, Z., & Tang, J. (2022, May). P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 61–68).
- [6] Han, X., Zhao, W., Ding, N., Liu, Z., & Sun, M. (2022). Ptr: Prompt tuning with rules for text classification. *AI Open*, 3, 182–192.
- [7] Gu, Y., Han, X., Liu, Z., & Huang, M. (2021). Ppt: Pre-

trained prompt tuning for few-shot learning. arXiv preprint arXiv:2109.04332.

[8] Ding, N., Hu, S., Zhao, W., Chen, Y., Liu, Z., Zheng, H. T., & Sun, M. (2021). Openprompt: An open-source framework for prompt-learning. arXiv preprint arXiv:2111.01998.

[9] Jobson, E., & Gutierrez, A. (2018). Abstractive text summarization using attentive sequence-to-sequence rnns. Stanford Reports

[10] Zhang, J., Zhao, Y., Saleh, M., & Liu, P. (2020, November). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In International Conference on Machine Learning (pp. 11328–11339). PMLR.

[11] Liu, Y., & Liu, P. (2021). Simcls: A simple framework for contrastive learning of abstractive summarization. arXiv preprint arXiv:2106.01890.

[12] Zhang, Y., Zhang, X., Wang, X., Chen, S. Q., & Wei, F. (2022). Latent Prompt Tuning for Text Summarization. arXiv preprint arXiv:2211.01837.

[13] Liu, X., Bai, Y., Li, J., Hu, Y., & Gao, Y. (2022). Psp: Pre-trained soft prompts for few-shot abstractive summarization. arXiv preprint arXiv:2204.04413.

[14] Nallapati, R., Zhou, B., Gulcehre, C., & Xiang, B. (2016). Abstractive text summarization using sequence-to-sequence rnns and beyond. arXiv preprint arXiv:1602.06023.

- [15] Narayan, S., Cohen, S. B., & Lapata, M. (2018). Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. arXiv preprint arXiv:1808.08745.
- [16] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., ... & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. arXiv preprint arXiv:1910.13461.
- [17] Lin, C. (2004). Rouge: A package for automatic evaluation of summaries. In Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain.
- [18] Lester, B., Al-Rfou, R., & Constant, N. (2021). The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691.
- [19] Wan, D., & Bansal, M. (2022). FactPEGASUS: Factuality-aware pre-training and fine-tuning for abstractive summarization. arXiv preprint arXiv:2205.07830.
- [20] Wu, J., Yu, T., Wang, R., Song, Z., Zhang, R., Zhao, H. & Henao, R. (2023). InfoPrompt: Information-Theoretic Soft Prompt Tuning for Natural Language Understanding. arXiv preprint arXiv:2306.04933.
- [21] Wu, H., & Shi, X. (2022, May). Adversarial soft prompt

tuning for cross-domain sentiment analysis. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 2438–2447).

Abstract

Prompt-based Abstractive Text Summarization

Jihyun Bae

Department of Industrial Engineering

The Graduate School

Seoul National University

Recently, prompt-based learning has shown similar performance to full model tuning in large language models by utilizing soft prompts. Prompt-based learning methods are effectively used to incorporate task-specific knowledge while extracting available knowledge from large language models. In this study, we propose a method for generating abstract summaries through prompt-based learning. Abstractive summaries, as one of the language generation tasks, can produce readable and consistent summaries compared to extractive summarization. The proposed model combines soft prompts and hard prompts and achieves good performance with less time and cost compared to fine-tuning the entire model. Additionally, we introduce a two-step approach for applying prompts based on text units. The contributions of this study are as follows: firstly, we propose a prompt-based learning method that can extract knowledge from large language models and apply it to specific tasks or data, demonstrating that this methodology can

perform creative and efficient text summarization not only in language understanding tasks but also in language generation tasks. We verify the efficient functioning of the model on data requiring domain knowledge by directly constructing an abstractive summarization dataset in the financial domain. Experimental results show that the model introduced in this study outperforms other publicly available models and exhibits improved performance compared to existing prompt-based learning methods.

Keywords : Abstractive Summarization, Prompt, Prompt Tuning, Language Generation
Student Number : 2020-24835