

거대 언어모델과 신경망 기반 협업 필터링을 통한 추출-생성 통합된 정보 보존형 요약

Information-Preserving Extractive-Abstractive Summarization via Large Language Models and Neural Collaborative Filtering

요약

정량적 수치, 핵심 용어, 정책적 문장 등 정보 손실이 치명적인 문서 요약 태스크에서는, 단순한 생성형 요약 방식만으로는 정보 보존과 사실 일치성을 확보하기 어렵다. 본 연구에서는 정보 보존형 요약을 달성하기 위해, 신경망 기반 협업 필터링(Neural Collaborative Filtering) 기반의 핵심 문장 추출 과정과 거대 언어모델 (Large Language Model) 기반의 자연어 생성 과정을 통합한 추출-생성 결합된 요약 기법을 제안한다. 제안하는 방법에서는 협업 필터링을 수행하는 추출 신경망을 통해 문서 내 핵심 문장을 선별한 뒤, 이 핵심 문장들과 원문 전체를 함께 입력으로 활용하여 거대 언어 모델이 정보 보존성과 표현력을 갖춘 요약을 생성한다. 단순 추출 요약보다 더 풍부한 맥락을 반영하면서도, 생성형 요약이 흔히 발생시키는 정보 누락이나 환각을 효과적으로 억제할 수 있다. 본 연구는 CNN/DailyMail 벤치마크 데이터셋에 대해 정량적으로 실험하여, 제안하는 방법이 정보 보존성과 표현성 측면 모두에서 기존 추출 또는 생성 단독 방식 대비 BERTScore 1.60%p 우수한 성능을 보임을 확인하였다.

표 1 효과적인 정보 선별과 표현 유연성을 고려한 요약 연구

Index	Data	Approach	Purpose	Description
2020 [1]	CNN/DM	추출	의미적 유사성 기반 요약	문서-후보 간 의미적 유사도 매칭 (Siamese-BERT)으로 추출
2024 [2]	PubMed		문서 구조 반영	지역적 주제와 계층 구조 활용한 문장 중요도 추출
2021 [3]	CNN/DM, NYT	생성	제어 가능한 요약	제한된 마르코프 결정 과정을 활용한 생성 요약 속성 제어
2022 [4]	CNN/DM, XSum		요약의 사실성과 정합성 유지	엔티티 및 관계 정보를 지식 그래프로 구성하여 요약과 원문간 사실적 일관성 유지
2024 [5]	CNN/DM, Reddit, PubMed	추출, 생성	오류 누적 최소화	하나의 구조에서 중요도 마스크를 활용하여 추출과 생성을 함께 수행

현대 사회는 뉴스, 보고서, 논문, 블로그 등 다양한 형태의 텍스트 정보가 실시간으로 생성·유통되는 정보 과잉 환경에 놓여 있다. 이러한 상황에서 모든 문서를 정독하기란 사실상 불가능하며, 핵심 정보를 신속하게 파악할 수 있는 요약 기술의 중요성이 점점 커지고 있다. 그러나 전체 문서로부터 핵심 정보를 정확히 식별하는 작업은 여전히 난이도가 높은 문제로 남아 있다. 문서의 길이가 길어질수록 핵심 내용이외의 불필요하거나 관련성이 낮은 정보가 포함될 가능성이 높아지며, 이로 인해 요약 모델의 성능이 저하될 수 있다.

본 논문은 문서의 핵심 문장을 추출한 뒤, 이를 기반으로 자연스럽게 응집력 있는 요약을 생성하는 추출-재구성 기반 요약 모델을 제안한다. 추출 단계에서는 Neural Collaborative Filtering(NCF)을 활용하여 문서-문장 간 의미적 상호작용을 학습하고, MSE Loss와 Ranking Loss를 병행하여 문장 중요도를 정밀하게 예측한다. 이후, 선택된 문장을 입력으로 대규모 언어 모델(LLM)이 최소한의 재구성을 수행함으로써 표현의 유창성을 향상시킨다. 이러한 접근은 LLM의 표현력을 활용하되, 정보 선별은 NCF 기반 모델이 담당함으로써, 정보 보존성, 처리 효율성, 표현 유연성 간의 균형 도모한다.

II. 관련 연구

최근 요약 연구는 문서 내 핵심 정보를 식별하는 정보 선별과 이를 자연스럽게 표현하는 표현 생성의 두 가지 접근으로 구분되며, 이러한 연구 흐름을 표 1에 정리하였다. 문서와 요약 후보 문장 간의 의미적 유사성을 정량화하여 문장을 추출하는

방식을 통해 정보 선별의 정밀도를 향상시키려는 시도가 있었으며 [1], 긴 문서의 지역적 주제 흐름과 계층 구조를 반영하여 문장 중요도를 평가함으로써 요약의 구조적 타당성을 높이려는 접근도 제안되었다 [2]. 한편, 표현의 유연성과 사실성 향상을 위한 생성 기반 접근도 제안되고 있다. 제한된 마르코프 결정 과정을 통한 요약 속성 제어 [3] 및 엔티티와 관계 정보를 활용한 사실적 일관성 유지 [4] 방식으로 정합성과 논리적 흐름을 강화하고자 하였다. 최근에는 추출과 생성을 단일 프레임워크 내에서 결합하려는 시도도 등장하고 있다. 인코더의 중요도 마스크를 활용하여 추출과 생성을 단일 프레임워크에 통합함으로써, 두 단계를 순차적으로 수행할 때 발생하는 오류 누적 문제를 완화하려는 시도도 이루어졌다 [5].

이처럼 기존 연구들은 정보의 정확한 식별, 문서 구조 반영, 그리고 표현의 자연스러움 사이의 균형을 달성하기 위한 다양한 전략들을 제시해왔으며, 본 연구는 이러한 요소들을 통합적으로 고려하되, 핵심 문장 예측과 최소한의 재구성을 설계함으로써 보다 정밀하고 효율적인 요약을 실현하고자 한다.

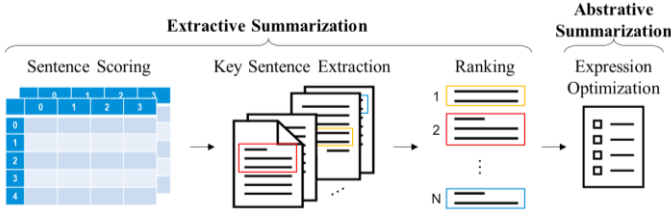


그림 1 핵심 문장 추출-재구성 요약 프로세스

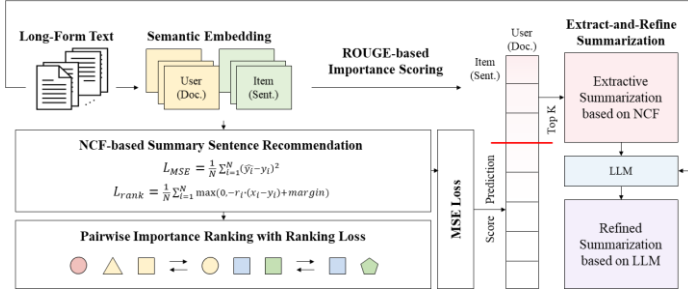


그림 2 NCF 추천 기반 핵심 문장 추출과 LLM 재구성을 통한 요약 시스템 구조도

III. 방법

본 연구에서는 문서로부터 핵심 정보를 효과적으로 요약하기 위해, 핵심 문장 추출과 요약 재구성의 두 단계를 순차적으로 수행하는 요약 프로세스를 설계하였다. 그림 1과 같이, 먼저 원문 내 문장들 중 요약에 기여할 가능성이 높은 핵심 문장을 선별하고, 이후 이들을 바탕으로 거대 언어 모델(LLM)을 활용하여 유창하고 응집력 있는 최종 요약을 생성한다. 이와 같은 방식은 추출 요약의 정보 보존성과 생성 요약의 표현력이라는 장점을 결합하고자 한다.

3.1. 핵심 문장 추출을 위한 NCF 기반 문장 추천 모델

그림 2와 같이, 본 연구에서는 문서 내 핵심 문장을 예측하기 위해 신경망 기반 협업 필터링(Neural Collaborative Filtering, NCF) 구조를 적용하였다. 이는 기존 추천 시스템의 사용자-아이템 간 상호작용 개념을 확장하여, 문서-문장 쌍 간의 중요도 상호작용을 학습하는 방식으로 설계되었다.

입력으로는 전체 문서의 임베딩 벡터와 각 문장 임베딩 벡터를 구성하여, 문서-문장 간 쌍을 형성한다. 이 임베딩 쌍은 NCF 네트워크에 입력되어, 각 문장이 해당 문서의 요약에 기여할 가능성을 수치적으로 예측한다.

모델 학습에는 두 가지 목적 함수를 병렬로 사용하였다. 첫 번째는 예측 점수와 ROUGE 기반 참조 중요도 간의 MSE를 최소화하는 회귀 손실 함수이며, 다음과 같이 정의된다:

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (1)$$

여기서 \hat{y}_i 는 NCF 모델이 예측한 문장 i 의 중요도 점수이고, y_i 는 참조 요약과의 ROUGE-L F1 기반 정답 중요도 점수이다. 이 손실은 예측의 정량적 정확도를 직접적으로 최적화한다. 두 번째는 문장 간 상대적 중요도 순위를 고려하는 쌍별 랭킹 손실(pairwise ranking loss)로, 다음과 같이 정의된다:

$$L_{rank} = \frac{1}{N} \sum_{i=1}^N \max(0, -r_i(x_i - y_i) + \text{margin}) \quad (2)$$

여기서 x_i 와 y_i 는 각각 쌍을 이루는 문장들의 예측 중요도 점수이며, $r_i \in \{+1, -1\}$ 는 해당 쌍의 정답 중요도 순위를 나타낸다. $r_i = +1$ 이면 문장 x_i 가 y_i 보다 중요하다는 의미이고, 반대의 경우 $r_i = -1$ 이다. 이 손실은 모델이 문장 간 상대적 중요도 관계를 잘 학습하도록 유도하며, 마진(margin)은 구분 기준을 제공한다.

최종적으로, 학습된 NCF 모델은 문서 내 각 문장에 대해 중요도 점수를 예측하고, 이를 바탕으로 핵심 문장 후보를 선택한다. 핵심 문장의 개수 K 는 문서마다 다르게 설정되며, 참조 요약의 단어 수 $|ref|$ 와 평균 문장 길이 avg_words 를 기반으로 다음과 같이 계산된다:

$$K = \min\left(\max\left(1, \left\lfloor \frac{|ref|}{avg_words} \cdot \beta \right\rfloor\right), K_{max}\right) \quad (3)$$

이후 중요도 점수에 따른 단순 정렬이 아닌, MMR(Maximal Marginal Relevance) 알고리즘을 적용하여 문서와의 유사도 및 문장 간 중복을 함께 고려함으로써, 정보 보존성과 다양성을 동시에 확보하였다.

3.2. In-Context Learning을 통한 LLM 기반 재구성 요약

추출된 핵심 문장의 정보는 그대로 유지하면서도 표현의 유창성과 응집력을 개선하기 위해, 본 연구는 거대 언어 모델(LLM)을 활용한 재구성 요약을 수행하였다. 이를 위해 별도의 파인튜닝 없이도 고품질 출력을 유도할 수 있도록, In-Context Learning 기반의 프롬프트 설계 기법들을 적용하였다.

표 2와 같이, 주요 기법은 다음과 같다. Instruction Prompting을 통해 요약 작성 지침을 명확히 전달하고, Act-as Prompting을 활용해 모델에 '전문 요약가'의 역할을 부여하였다. 또한, Retrieval-Augmented Prompting으로 원문 전체를 함께 제공하여 문맥 이해를 도왔고, Schema-Constrained Generation을 통해 출력 형식을 단락 구조로 고정하였다. 마지막으로, Lexical-Constrained Generation을 적용하여 고유명사 및 주요 어휘의 누락을 방지하였다.

이러한 제약 기반 프롬프트 설계는 In-Context Learning의 유연성과 사전학습 모델의 표현 능력을 효과적으로 결합하여, 추출 요약의 정보성을 유지하면서도 문장 연결성과 표현 품질을 개선한 결과를 생성할 수 있도록 하였다.

표 2 LLM 기반 요약 재구성을 위한 프롬프트 설계 기법

Technique	Description	Purpose
Act-as Prompting	“전문 요약가입니다”와 같이 모델의 역할을 지정	작업을 역할에 맞게 조정하도록 유도
Instruction Prompting	요약 방식, 포함 정보, 금지 조건 등 지침 명시	모델이 요약 작업의 목적과 기준을 명확히 이해하도록 유도
Retrieval-Augmented Prompting	핵심 문장 외에 원문 전체를 함께 제공	문장 연결성과 사실성 향상, 정보 누락 방지
Schema-Constrained Generation	출력이 한 문단 형식이 되도록 구조 제한	일관된 요약 구조 유지 및 평가 가능성 향상
Lexical-Constrained Generation	고유명사, 수치, 시간 등 핵심 어휘 유지지침 포함	핵심 정보 보존 및 불필요한 일반화·생략 방지

표 3 요약 결과의 표면적 일치도 평가 (ROUGE: F1-Score)

	Extractive			Abstractive			Extractive + Abstractive
	Cosine Top-k	TextRank	NCF	BART	T5-base	LLaMA	NCF-LLaMA (Ours)
ROUGE-1	0.3445	0.3091	0.4140	0.3776	0.3971	0.3864	0.4433
ROUGE-2	0.1525	0.1131	0.1800	0.1786	0.1782	0.1430	0.1891
ROUGE-L	0.2230	0.1943	0.2626	0.2350	0.2789	0.2513	0.2801

표 4 의미 보존 측면의 요약 성능 평가 (BERTScore)

	Extractive			Abstractive			Extractive + Abstractive
	Cosine Top-k	TextRank	NCF	BART	T5-base	LLaMA	NCF-LLaMA (Ours)
Precision	0.5538	0.5415	0.6161	0.5505	0.6537	0.5942	0.6496
Recall	0.6326	0.6120	0.6466	0.6686	0.6304	0.6690	0.6658
F1-Score	0.5895	0.5736	0.6304	0.6028	0.6409	0.6285	0.6569

IV. 실험 및 결론

4.1. 실험 요약 데이터

본 연구에는 CNN/DailyMail 요약 데이터셋을 사용하였으며, 총 약 31만 개의 기사-요약 쌍으로 구성되어 있다. 이 중 CNN 뉴스 기사가 약 9만 개, DailyMail 기사가 약 22만 개로, 두 출처의 뉴스 기사와 해당 기사에 대한 다중 문장 요약(reference summary)이 함께 제공된다.

4.2 요약 결과의 표면적 일치도 평가 (ROUGE: F1-Score)

표 3은 ROUGE-1, ROUGE-2, ROUGE-L의 F1-Score를 기준으로, 각 요약 기법이 참조 요약과 얼마나 높은 수준의 단어 일치도를 보이는지 정량적으로 비교한 결과이다. ROUGE는 문서 요약에서 정보 보존 수준과 표현 정확도를 평가하는 대표 지표로 널리 활용된다.

실험 결과, 제안한 NCF-LLaMA (Ours) 모델은 세 가지 ROUGE 지표 모두에서 기존 추출 방식(NCF, TextRank, Cosine Top-k)과 생성 방식(BART, T5-base, LLaMA)을 일관되게 상회하는 성능을 기록하였다. 특히 ROUGE-1 기준으로 NCF 대비 약 0.03p, ROUGE-L 기준으로 T5-base 대비 약 0.002p 향상되었으며, 이는 정보 선택의 정밀도와 표현 재구성의 유연성을 동시에 확보했음을 의미한다.

이러한 결과는 정보의 누락이 치명적인 문서 요약 환경에서, 단순한 생성형 접근으로는 확보하기 어려운 정확한 표현 보존이 가능함을 시사한다. 제안 기법은 핵심 문장 추출과 LLM 기반 생성의 결합을 통해, 정보 보존성과 표현 품질을 동시에 충족시키는 실용적 요약 방식임을 입증하였다.

4.3 의미 보존 측면의 요약 성능 평가 (BERTScore)

표 4는 BERTScore를 기준으로 각 요약 기법의 의미 유사도 성능을 정량적으로 비교한 결과를 나타낸다. BERTScore는 사전학습된 언어 모델의 임베딩을 활용하여 요약과 참조 요약 간의 의미적 일치를 측정하며, 단어 수준의 표현 차이를 넘어 의미 보존 여부를 평가할 수 있는 지표로 널리 활용된다.

실험 결과, 제안하는 NCF-LLaMA (Ours) 모델은 F1-Score 0.6569로 전체 모델 중 가장 높은 점수를 기록하였으며, 이는 기존 생성요약 기법인 T5-base(0.6409)와 LLaMA(0.6285)보다도

우수한 성능이다. 또한, 추출 방식 중 가장 높은 점수를 기록한 NCF(0.6304)보다도 약 0.0265p 향상되어, 정보의 보존뿐 아니라 의미 표현의 정확성 면에서도 경쟁력을 확보했음을 보여준다.

특히, Cosine Top-k나 TextRank와 같은 단순 추출 방식은 BERTScore 기준으로 낮은 성능을 보였으며, 이는 표현이 원문에 충실하더라도 의미 연결성이나 표현 적절성이 부족할 수 있음을 시사한다. 반면, 본 연구의 방식은 정보 선택과 의미 표현 간의 상호 보완을 통해, 정확하고 자연스러운 요약을 생성하는 데 효과적임을 실험적으로 입증하였다.

4.3. 결론 및 향후 연구

본 연구는 정보 보존을 중시하는 요약 과제에 대응하기 위해, 협업 필터링 기반 핵심 문장 추출과 거대 언어 모델 기반 문장 재구성을 결합한 요약 방식을 제안하였다. 실험 결과, 제안한 NCF-LLaMA 모델은 기존의 추출 또는 생성 방식 대비 정보 보존성과 표현 유연성 모두에서 우수한 성능을 입증하였다. 향후 연구로는, 현재의 핵심 문장 추출 과정이 문장 간 상호작용을 고려하지 않고 각 문장을 개별적으로 분석한다는 한계를 극복하기 위해, 문맥 구조와 문장 간 의미 관계를 반영한 추출 모델을 설계함으로써, 보다 정교하고 일관성 있는 핵심 선별이 가능하도록 개선할 필요가 있다.

감사의 글

참고 문헌

- [1] M.Zhong, P.Liu, Y.Chen, D.Wang, X.Qiu and X.Huang, "Extractive summarization as text matching," *arXiv preprint arXiv:2004.08795*, 2020.
- [2] T.Wang, C.Yang, M.Zou, J.Liang, D.Xiang, W.Yang and J.Li, "A study of extractive summarization of long documents incorporating local topic and hierarchical information," *Scientific Reports*, vol.14, p.10140, 2024.
- [3] H.P.Chan, L.Wang and I.King, "Controllable summarization with constrained Markov decision process," *Trans. of the Association for Computational Linguistics*, vol.9, pp.1213-1232, 2021
- [4] Y.Lyu, C.Zhu, T.Xu, Z.Yin, Z and E.Chen, E, "Faithful abstractive summarization via fact-aware consistency-constrained transformer," *In Pro. of the 31st ACM Int. Conf. on Information & Knowledge Management*, pp. 1410-1419, 2022.
- [5] Y.Wu, H.Li, H.Zhu, G.Nenadic, and X.J.Zeng, "Extract-and-Abstract: Unifying Extractive and Abstractive Summarization within Single Encoder-Decoder Framework," *arXiv preprint arXiv:2409.11827*, 2024.