

# 어텐션 및 그래프 방법론을 적용한 텍스트 계층적 요약 연구

A study on Hierarchical Text Summary applying structured attention and graph methodology

나운영<sup>1</sup>, 김산<sup>2</sup>, 김재광<sup>3†</sup>  
 Yunyeong Na, San Kim, and Jaekwang Kim<sup>†</sup>

<sup>1</sup>성균관대학교 인공지능융합학과/소셜이노베이션융합전공 석사과정

E-mail: pond24@g.skku.edu

<sup>2</sup>성균관대학교 소프트웨어학과 박사과정

E-mail: saankim@skku.edu

<sup>3</sup>성균관대학교 글로벌융합학부/소셜이노베이션융합전공 부교수

E-mail: linux@skku.edu

## 요약

최근, 긴 텍스트 데이터에서 중요한 내용을 추출하여 독자가 이해하기 쉽도록 텍스트를 요약하는 방법에 관한 연구들이 활발히 진행되고 있다. 하지만 요약이 실제 내용과 다르거나 직관적으로 이해하기 어려울 수 있다.

본 연구는 이러한 한계를 극복하기 위해 어텐션 메커니즘과 그래프의 장점을 결합한 새로운 텍스트 요약 방법론을 제안한다. 주요 단어와 문장의 중요도를 어텐션 메커니즘을 통해 계산하고, 이를 그래프 형태로 변환해 텍스트를 계층적으로 표현 및 요약한다. 이후 노드 정보 종합하여 문서 라벨 예측, 정확도를 비교하여 검증한다. 다양한 카테고리 별 기사로 구성된 NYT 데이터셋의 본문 데이터를 이용해 실험한 결과 단일 방법을 적용하는 경우보다 우수한 성능을 보인다.

**키워드** : 계층적 관계 표현, 텍스트 요약, 어텐션 메커니즘, 그래프 신경망, 자연어 처리

## 1. 서론

텍스트 요약은 주어진 문서로부터 주요 키워드 및 문장을 추출하여 독자가 짧은 시간안에, 더 적은 노력으로도 중요한 내용을 파악할 수 있도록 돕는다. 하지만 요약된 내용이 실제로는 문서의 중심 내용과 다르거나, 문장으로 표현되어 직관적인 이해가 어렵다는 한계가 있다.

이를 위해, 어텐션 메커니즘 기반 모델을 활용해 주어진 텍스트 데이터의 중요한 단어와 문장에 집중하는 계층적 어텐션을 이용한 사례가 있다[Ref. HAN]. 하지만 추출한 키워드 간 관계를 모델링 하는 과정이 없어 관계를 계층적으로 표현하지 못한다는 한계를 지닌다.

그 외에도, 그래프의 노드(단어, 문장 등)와 엣지(단어 간 관계, 문장 간 관계 등)를 활용해 복잡한 상호작용을 학습하고 텍스트를 계층적으로 표현한 사례가 있다[Ref. GCN]. 해당 모델은 텍스트 데이터를 단어 그래프(graph-of-words)로 변환하는 과정에서 단어 동시 등장(co-occurrence) 및 문서-단어 관계를 그래프로 표현한 후 이를 학습하는 방식을 사용한다. 하지만 단어를 선별하는 과정에서 단어의 중요도는 고려하지 않아 요약 결과를 신뢰하기 어렵다.

본 연구에서는 어텐션 메커니즘과 그래프가 가진 장점을 모두 활용한 방법론을 제시하고자 한다. 먼저, 어텐션 메커니즘을 활용해 입력 데이터로부터 중요한 단어와 문장을 찾아 중요도를 계산하고, 그래프의 계층적 표현력을 이용하기 위해 추출한 주요 단어와 문장을 벡터로 변환한다. 단어와 문장은 노드로, 단어 및 문장 간 관계는 엣지로 표현한다. 이러한 과정을 거친 데이터는 계층적 관계로 표현 및 요약된다. 최종적으로 노드 정보를 종합하여 문서 라벨을 예측하고, 학습 데이터의 라벨과 비교하는 검증 과정을 거친다.