

# 언어모델 기반 교통사고 실황조사서 자동 생성 연구

서동준\*, 정경용\*\*

## A Study on Automatic Generation of Traffic Accident Investigation Reports based on Language Models

Dongjoon Suh\*, Kyungyong Chung\*\*

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2020R1A6A1A03040583). Additionally, this work was also supported by Kyonggi University's Graduate Research Assistantship (2025)

---

### 요약

본 연구는 자연어 처리 모델인 BERT를 활용하여 교통사고 실황조사서를 자동으로 생성하는 방법론을 제안한다. 기존 사고 조사는 조사관의 주관적 판단에 의존하며, 파편화된 현장 데이터 분석에 많은 시간과 노력이 소요된다. 이러한 문제를 해결하기 위해, 본 방법론은 CCTV, EDR, 운전자 진술 등 다중 소스 데이터를 심층적으로 분석한다. 제안 모델은 먼저 파편화된 정보를 5W1H(육하원칙) 기반으로 추출하여 객관적인 사실 관계를 구조화한다. 이후 시계열 관계 마이닝을 통해 구조화된 정보들 간의 동적인 인과관계를 추론하여 사고의 핵심 서사를 생성한다. 본 방법론은 조사의 신속성과 정확성을 높여 2차 사고 예방 및 사회적 비용 절감에 기여할 것으로 기대된다.

### Abstract

This study proposes a methodology for automatically generating traffic accident investigation reports using the BERT language model. Conventional accident investigation relies heavily on the subjective judgment of investigators, and the process of analyzing fragmented field data is time-consuming and labor-intensive. To address these challenges, this methodology deeply analyzes multi-source data, including CCTV footage, EDR records, and driver statements. The proposed model first structures objective factual relationships by extracting fragmented information based on the 5W1H (Who, What, When, Where, Why, How) principle. Subsequently, it generates the core accident narrative by inferring dynamic causal relationships among the structured information through time-series relationship mining. This methodology is expected to enhance the speed and accuracy of investigations, contributing to the prevention of secondary accidents and the reduction of social costs.

### Keywords

traffic accident analysis, language model, BERT, automatic report generation, deep learning

---

\* 경기대학교 컴퓨터과학과 석사과정  
- ORCID: <https://orcid.org/0009-0008-5412-382X>  
\*\* 경기대학교 AI컴퓨터공학부 교수(교신저자)  
- ORCID: <https://orcid.org/0000-0002-6439-9992>

• Received: Aug. 11, 2025, Revised: Sep. 16, 2025, Accepted: Sep. 19, 2025  
• Corresponding Author: Kyungyong Chung  
Division of AI Computer Science and Engineering, Kyonggi University, South Korea  
Tel.: +82-32-249-1382, Email: [dragonhci@gmail.com](mailto:dragonhci@gmail.com)

## 1. 서 론

대한민국은 정부와 관련 기관들의 다각적인 노력으로 지난 10년간 교통사고 사망자 수가 크게 감소했으나, 인구 10만 명당 사망자 수는 여전히 OECD 평균보다 25% 높은 수준이다[1]. 특히 최근 5년간 사고 사망자 수가 다시 증가세로 돌아선 점은 현행 교통사고 예방 대책의 근본적인 재검토가 필요함을 시사한다.

효과적인 교통사고 예방 대책은 정확하고 신속한 사고 원인 규명에서 출발하지만, 현재 대한민국의 교통사고 조사 체계는 여러 구조적인 한계에 직면해 있다. 첫째, 현장 조사는 교통 정체와 2차 사고 위험을 유발하며, 복잡한 증거 수집 과정은 많은 시간과 노력을 요구한다. 둘째, 자료 분석 및 보존의 제한성이 객관적인 사고 분석을 어렵게 만든다. 교통사고 분석은 CCTV, 차량 블랙박스, 사고기록장치(EDR), 목격자 진술 등 파편화된 여러 데이터를 종합하여 이루어진다. 그러나 이러한 데이터들은 종종 서로 상충되거나 정보가 부족한 경우가 많아, 분석 과정에서 조사관의 주관적인 경험이나 해석이 개입될 여지가 크다. 또한, 수집된 방대한 양의 비정형 데이터를 체계적으로 관리하는 방법론이 미비하여, 과거 데이터를 활용한 심층적인 패턴 분석이나 예방 대책 수립에 어려움을 겪고 있다.

이러한 문제점을 해결하기 위해, 본 연구에서는 BERT 모델 기반의 교통사고 실황조사서 자동 생성 방법론을 제안한다. 본 방법론은 1) 다중 소스 데이터를 텍스트로 정규화하고, 2) KLUE-BERT-base 모델을 파인튜닝하여 5W1H 정보를 추출하며, 3) 시계열 관계 마이닝으로 사고의 서사를 재구성한다. 제안 방법론은 조사관의 주관성을 배제하여 객관성을 확보하고, 반복 업무를 자동화하여 효율성을 높이는 장점이 있다. 반면, 입력 데이터의 품질에 결과가 좌우되는 데이터 의존성은 한계로 작용할 수 있다. 따라서 본 연구의 핵심 기여(Contribution)는 기존의 정보 '탐지' 연구를 넘어, BERT의 깊이 있는 '문맥 이해' 능력과 '관계 마이닝'을 결합하여 파편화된 사실로부터 논리적 인과관계를 추론하고, 완전한 서사를 갖춘 보고서를 생성하는 최초의 프레임워크를

제안했다는 점에 있다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구 동향을, 3장에서는 제안 방법론의 세부 방법론을 설명한다. 4장에서는 실험을 통한 성능 평가 결과를 기술하고, 5장에서 결론을 맺는다.

## II. 관련 연구

### 2.1 교통사고 조사 지원 방법론

전통적인 교통사고 조사 및 분석은 물리 법칙에 기반한 시뮬레이션 프로그램을 통해 상당 부분 지원되어 왔다. 이 분야에서 가장 널리 알려진 프로그램은 오스트리아에서 개발된 PC-Crash이다. PC-Crash는 조사관이 입력한 차량의 제원, 충돌 전 속도 및 각도, 노면 상태, 제동 흔적 등의 데이터를 바탕으로 충돌 시뮬레이션을 수행한다 [2]. 네덜란드에서 개발된 MADYMO(Mathematical Dynamic MOdels) 역시 중요한 분석 도구 중 하나로, 주로 인체 상해 분석에 특화되어 있으며 정교한 인체 모델을 사용하여 충돌 시 충격량을 계산한다. 다음 그림 1은 전통적인 교통사고 조사 및 분석 과정을 도식화한 것이다.

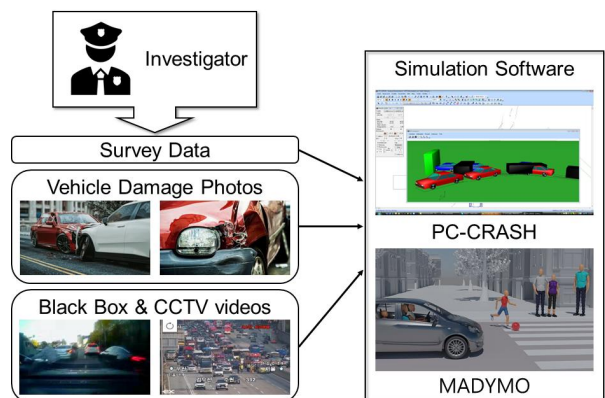


그림 1. 기존 교통사고 조사 및 시뮬레이션 분석 체계  
Fig. 1. Framework for existing traffic accident investigation and simulation analysis

그림 1에서 볼 수 있듯이, 기존 방식은 조사관이 수집한 측량 데이터, 차량 파손 사진, 블랙박스 및 CCTV 영상 등의 다양한 데이터를 시뮬레이션 소프트웨어

트웨어에 수동으로 입력하여 3D 재구성을 하는 복잡한 과정을 거친다. 그러나 이러한 전통적인 시물레이션 기반 방법론들은 'Garbage In, Garbage Out' 원칙에 따라 입력 데이터의 정확성에 결과가 크게 좌우된다는 명확한 한계를 가진다.

이러한 한계를 극복하기 위해 최근에는 인공지능(AI), 특히 컴퓨터 비전 기술을 교통사고 분석에 접목하려는 연구가 활발하게 진행되고 있다. 예를 들어, 사고 차량의 파손 부위 사진을 딥러닝 모델로 분석하여 수리 비용을 견적하거나, 블랙박스 영상으로 법규 위반을 탐지하는 기술들이 개발되고 있다. 하지만 이러한 AI 기반 연구들은 대부분 영상 분석을 통한 특정 정보의 인식(Recognition) 또는 탐지(Detection)에 초점을 맞추고 있다. 즉, 사고와 관련된 개별적인 '사실'들을 추출하는 데는 효과적이지만, 이 사실들을 유기적으로 연결하여 '어떤 원인으로 인해 어떤 과정으로 사고가 발생했으며, 그 결과는 무엇인가'라는 서사(Narrative)를 구성하는 데는 한계가 있다.

## 2.2 대규모 언어 모델(LLM) 및 BERT

최근 딥러닝 기술, 특히 컴퓨터 비전 분야의 발전은 교통사고 분석에 새로운 가능성을 제시했다. 다수의 연구는 CNN(Convolutional Neural Networks)이나 LSTM(Long Short-Term Memory)과 같은 모델을 사용하여 CCTV 영상에서 사고 상황을 직접 감지하거나 [3], 사고 위험을 예측하는 데 초점을 맞추었다. 이러한 연구들은 시각적 데이터로부터 특정 이벤트를 탐지하는 데 높은 성능을 보였지만, 사고의 전후 맥락과 원인을 설명하는 텍스트 기반의 보고서를 생성하는 데는 명확한 한계를 가진다.

한편, 자연어 처리 분야에서는 BERT 이후 GPT, T5, BART와 같은 대규모 언어 모델(LLM)이 등장하며 문서 요약 기술이 크게 발전하였다 [4][5]. 최근에는 GPT-4, Gemini 등 다양한 LLM의 성능을 비교 평가하는 연구가 활발히 진행되고 있으며 [6], BERT 모델을 사용자 리뷰의 감정 분석에 활용하여 추천 시스템의 정확도를 높이는 연구도 이루어지고 있다 [7]. 그러나 일반적인 문서 요약은 텍스트의

표면적인 중요도를 기반으로 문장을 추출하거나 재구성하는 반면, 교통사고 실황조사서는 파편화된 정보들 속에서 명확한 인과관계를 찾아내 논리적인 순서로 재구성해야 한다는 특수성이 있다. 다음 그림 2는 본 연구에서 제안하는 언어 모델(LM)이 어떻게 사고의 핵심 요인을 종합적으로 분석하는지에 대한 개념을 나타낸다.

그림 2는 언어 모델이 '빙판길', '차선 변경', '주요 교차로', '오전 8시 30분' 등과 같은 개별적인 요인들을 단순 나열하는 것이 아니라, 이들 간의 관계를 종합적으로 이해하여 사고의 본질을 파악하는 과정을 상징적으로 보여준다. 본 연구는 바로 이 지점에서 차별성을 가진다. 단순 요약을 넘어, BERT의 강력한 문맥 이해 능력을 바탕으로 구조화된 정보를 추출하고, 관계 마이닝을 통해 사고의 동적인 인과관계를 밝혀냄으로써 '설명 가능한' 보고서를 생성하는 것을 목표로 한다.

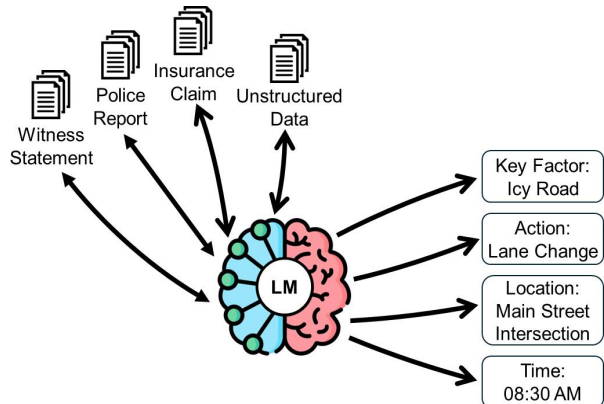


그림 2. LM을 활용한 교통사고 핵심 요인 분석 개념도  
Fig. 2. Conceptual diagram for core traffic accident factor analysis using an LM

## III. LM 기반 실황조사서 자동 생성 방법론

본 연구에서 제안하는 방법론의 전체 아키텍처는 다음 그림 3과 같다. 방법론은 크게 'Step 1: 데이터 수집 및 전처리', 'Step 2: BERT 기반 정보 추출 및 요약', 그리고 'Step 3: 실황조사서 생성'의 3단계로 구성된다. 각 단계는 유기적으로 연결되어 CCTV, 블랙박스, 목격자 진술과 같은 이종(Heterogeneous) 데이터로부터 최종적인 실황조사서를 생성한다.

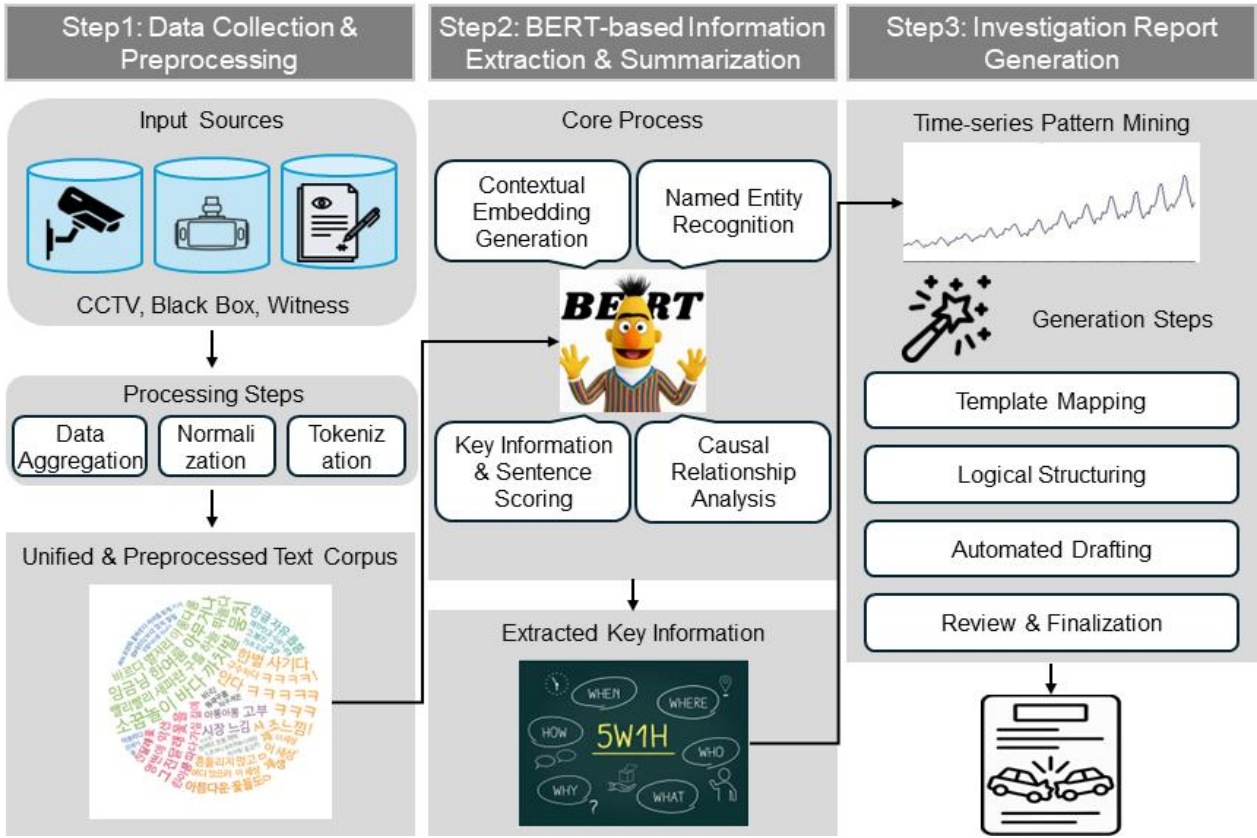


그림 3. 제안 방법론의 전체 아키텍처  
 Fig. 3. Overall architecture of the proposed methodology

### 3.1 데이터 수집 및 전처리

방법론의 첫 번째 단계에서는 교통사고 현장의 다각적인 실체를 파악하기 위해 여러 이종 데이터 소스로부터 정보를 수집한다. 여기에는 CCTV 및 차량 블랙박스에서 확보되는 시각적 영상 데이터, 사고기록장치(EDR)와 운행기록계(DTG)가 기록한 정밀한 센서 데이터, 그리고 운전자 진술서나 현장 조사관의 초기 보고서와 같은 비정형 텍스트 데이터가 모두 포함된다.

특히, CCTV 및 블랙박스 영상과 같은 시각 데이터는 그대로 언어 모델에 입력될 수 없으므로, 텍스트로 변환하는 과정이 필수적이다. 본 연구에서는 이를 위해 Vision-Language 모델인 BLIP (Bootstrapping Language-Image Pre-training)을 활용한다 [8]. BLIP 모델은 영상의 주요 프레임을 입력받아 '빨간색 쏘나타 차량이 신호를 위반하여 교차로에 진입함과 같이 상황을 묘사하는 캡션을 생성하

여, 시각적 정보를 BERT가 처리할 수 있는 텍스트 데이터로 변환한다. 이렇게 변환된 텍스트와 기존 텍스트 데이터를 통합한 후, 불필요한 특수문자나 중복 정보를 제거하는 정제 과정을 거치고, 문장의 의미 단위의 토큰으로 분리하는 토큰화를 수행하여 최종적으로 모델이 분석할 수 있는 단일 형식의 텍스트 코퍼스를 구축한다.

### 3.2 BERT 기반 정보 추출 및 요약

두 번째 단계는 방법론의 핵심 두뇌 역할을 수행하는 과정으로, 잘 정제된 text corpus를 BERT 모델에 입력하여 사고의 핵심 정보와 인과관계를 심층적으로 분석하고 추출한다. 이러한 기능은 사전 학습된 KLUE-BERT-base 모델을 교통사고 데이터셋에 맞게 파인튜닝(Fine-tuning)하여 구현되었다.

BERT 모델은 입력된 문장에 대해 다음 그림 4와 같이 4가지 주요 과정을 순차적으로 수행한다.

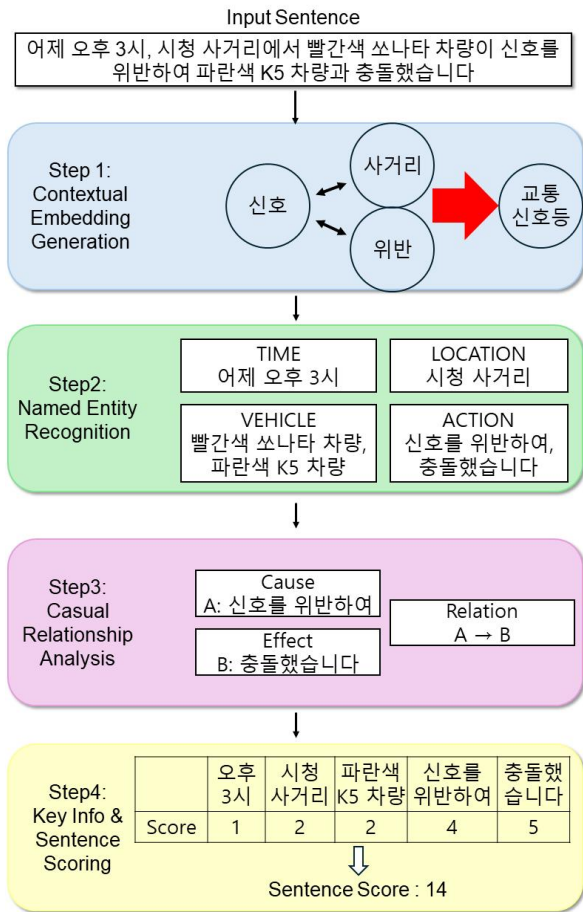


그림 4. BERT 모델의 사고 문장 분석 과정 예시  
Fig. 4. Example of the accident sentence analysis process in the BERT model

그림 4에 나타난 바와 같이, BERT 모델은 4단계를 거쳐 문장을 분석한다. 먼저 '문맥적 임베딩 생성(Contextual Embedding Generation)' 단계에서는 모델이 문장 전체의 맥락을 양방향으로 파악하여 '신호'가 '교통 신호등'임을 이해하는 등 각 단어의 의미를 벡터로 변환한다. 다음으로 '개체명 인식(NER, Named Entity Recognition)' 단계에서는 이 벡터를 기반으로 '시간', '장소' 등 사전에 정의된 유형으로 주요 개체를 식별하고 분류한다. 이어서 '인과관계 분석(Causal Relationship Analysis)' 단계에서는 식별된 개체와 행위들 간의 논리적 관계를 분석하여 '신호 위반'이 '충돌'을 유발했다는 인과관계를 추론하는데, 이는 두 개체 간의 관계를 분류하는 관계 추출 문제로 정의할 수 있다. 마지막으로 '핵심 정보 및 문장 점수화(Key Information & Sentence Scoring)' 단계에서는 문장이 사고의 핵심 요소를 얼마나 포함하

는지 평가하여 중요도 점수를 부여하고, 높은 점수를 받은 문장을 핵심 정보로 선택한다.

이러한 심층적인 분석 과정을 통해 BERT 모델은 비정형적인 문장을 '누가', '언제', '어디서', '무엇을', '왜', '어떻게'에 해당하는 육하원칙(5W1H) 기반의 구조화된 정보로 변환하게 된다. 이렇게 추출된 5W1H 정보는 실황조사서 자동 생성을 위한 가장 중요한 기반 데이터가 된다.

### 3.3 실황조사서 생성

마지막 단계는 추출된 5W1H 정보들을 '사건 블록'으로 재구성하고, 이들 간의 동적인 인과관계를 밝혀내는 서사 생성 과정이다. 이 단계는 개별적인 사실 데이터를 유기적인 이야기로 전환하는 본 방법론의 핵심적인 추론 기능을 담당한다.

다음 그림 5는 이러한 복합적인 추론 및 생성 과정이 어떻게 이루어지는지를 개념적으로 보여준다.

그림 5에서 볼 수 있듯이, 시스템은 먼저 BERT가 추출한 파편화된 사실 데이터(Who, When, Where 등)를 'A 차량이 차선을 변경했다'와 같이 완전한 '사건 블록(Event Block)'으로 재구성한다.

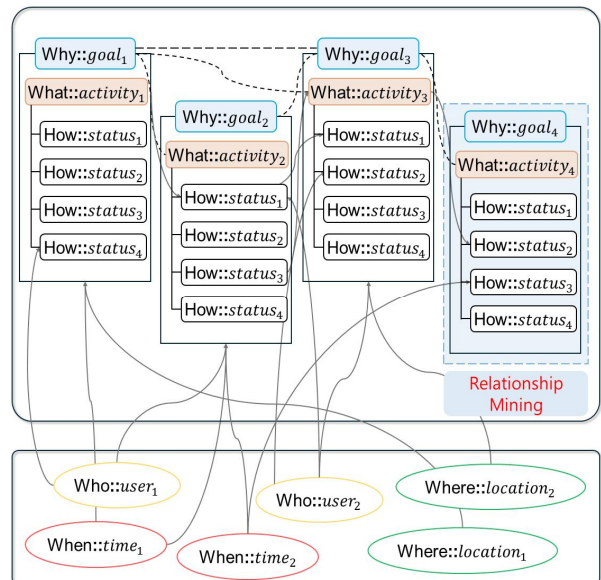


그림 5. 5W1H 기반 시계열 관계 마이닝 및 서사 생성 개념도

Fig. 5. Conceptual diagram of 5W1H-based time-series relationship mining and narrative generation

본 단계의 핵심인 관계 마이닝은 시간 순서에 따라 정렬된 이 사건 블록들의 시퀀스에서 가장 가능성 높은 인과 사슬을 찾는 알고리즘이다. 이를 위해 각 사건 블록을 노드로 하는 유향 그래프를 구성하고, 두 사건 블록 간의 인과관계 확률을 BERT의 NSP 점수를 변형하여 엣지의 가중치로 사용한다. 최종적으로 다익스트라 알고리즘 등을 이용해 가장 높은 누적 가중치를 가지는 경로를 탐색하여 사고의 핵심 서사로 채택한다. 이렇게 발견된 핵심 서사를 바탕으로, 본 방법론은 최종적으로 실황조사서의 '사고 경위' 부분을 논리적으로 완결성 있게 작성하며, 이를 통해 조사관은 사고 전체를 직관적으로 파악할 수 있는 보고서 초안을 제공받는다.

#### IV. 성능 평가

본 장에서는 제안하는 LM 기반 실황조사서 자동 생성 방법론의 성능을 객관적으로 평가하고, 기존 모델들과의 비교를 통해 기술적 우수성을 입증한다.

##### 4.1 실험 환경 및 데이터셋

실제 교통사고 실황조사서는 개인정보보호 문제로 확보가 어려워, 본 연구에서는 형식과 내용이 유사한 교통사고 관련 법원 판결문 1,000건을 데이터셋으로 활용하였다[9]. 대법원 종합법률정보 사이트에서 공개하는 판결문의 '기초사실' 부분을 정답 요약문으로 간주하고 성능을 측정하였으며, 실험은 PyTorch 환경에서 NVIDIA RTX 3090 GPU를 사용하여 진행한다.

##### 4.2 평가 지표 및 결과

제안 모델의 성능을 그래프 기반의 TextRank와 생성 모델인 BART-ko-news와 비교하였다. 평가지표로는 어휘적 일치도를 측정하는 ROUGE와 의미적 유사성을 비교하는 BERTScore를 함께 사용하였다 [10][11]. 동일한 데이터셋으로 측정한 성능 비교 결과는 표 1과 같다.

표 1. 모델별 요약 성능 비교 결과

Table 1. Comparison of ROUGE and BERTScore Evaluation Metrics

Model	ROUGE-L (Recall)	BERTScore (F1)
TextRank	0.32	0.81
BART-ko-news	0.41	0.88
Proposed model	0.48	0.93

실험 결과, 제안 모델은 모든 측면에서 베이스라인 모델보다 높은 성능을 보였다. 특히 의미적 정확성을 나타내는 BERTScore에서 가장 큰 성능 향상을 보였는데, 이는 제안 모델의 핵심 기능인 인과관계 마이닝이 사고의 논리 구조와 맥락을 정확히 파악하고 있음을 시사한다. TextRank는 문맥이 매끄럽지 못했고, BART는 사실과 다른 내용을 생성하는 한계를 보인 반면, 본 제안 모델은 사실에 기반하여 논리적으로 가장 타당한 서사를 구성함으로써 높은 정확성과 신뢰도를 확보할 수 있었다.

#### V. 결론 및 향후 연구

본 연구에서는 BERT 언어 모델과 시계열 마이닝을 결합하여 교통사고 실황조사서를 자동으로 생성하는 방법론을 제안했다. 제안된 방법론은 다양한 소스로부터 수집된 데이터를 종합적으로 분석하고, BERT 모델의 뛰어난 문맥 이해 능력을 활용하여 사고의 핵심 정보와 인과관계를 추출하며, 시계열 마이닝을 통해 동적인 서사를 구성한다. 본 연구는 BERT 언어 모델을 교통사고 조사라는 특수 도메인에 적용하여, 단순 사실 나열이 아닌 인과관계에 기반한 서사 생성을 성공적으로 수행했다는 점에서 학술적, 실용적 기여를 가진다.

향후 연구로는 멀티모달(Multi-modal) 학습의 적용을 계획하고 있다. 현재 제안된 방법론은 텍스트 데이터를 중심으로 분석을 수행하지만, 영상이나 센서 데이터에 포함된 비언어적 정보를 직접적으로 활용하는 데는 한계가 있다. 텍스트, 이미지, 수치 데이터를 동시에 입력받아 상호 보완적으로 분석하는 멀티모달 딥러닝 모델을 개발하여, 보다 종합적이고 정확한 사고 상황 인식이 가능하도록 방법론을 고도화할 계획이다. 이를 통해 텍스트만으로는

파악하기 어려운 차량의 동선, 충돌 직전의 미세한 움직임 등 시각적 정보를 분석에 통합하여, 사고 원인 규명의 정확성을 한층 더 높일 수 있을 것이다. 본 연구가 성공적으로 수행된다면, LM 기술을 활용하여 교통사고 조사 분야의 패러다임을 전환하고, 궁극적으로는 더 안전한 교통 환경을 만드는 데 중요한 기여를 할 수 있을 것으로 기대한다.

## References

- [1] Road Traffic Authority (KoROAD), "Statistical Analysis of Traffic Accidents 2023", TAAS Traffic Accident Analysis System, 2024.
- [2] H. Steffan and A. Moser, "The Collision and Trajectory Models of PC-Crash", SAE Technical Paper 960886, Feb. 1996. <http://doi.org/10.4271/960886>.
- [3] X. Wu and T. Li, "A deep learning-based car accident detection approach in video-based traffic surveillance", *Journal of Optics*, Vol. 53, pp. 3383-3391, Jan. 2024. <http://doi.org/10.1007/s12596-023-01581-4>.
- [4] 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", *Journal of Machine Learning Research*, Vol. 21, No. 140, pp. 1-67, Jan. 2020. <http://doi.org/10.5555/3455716.3455856>.
- [5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1 (Long and Short Papers), Minneapolis, Minnesota, pp. 4171-4186, Jun. 2019. <http://doi.org/10.48550/arXiv.1810.04805>.
- [6] D. Lee, "A Comparative Study of LLM-based Agents for Interactive Text-based Game Automation", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 25, No. 3, pp. 1-8, Jun. 2025. <http://doi.org/10.7236/JIIBC.2025.25.3.1>.
- [7] J. B. Kim, Y. G. Kim, and J. M. Park, "Collaborative Filtered Enhanced Recommendation System Using BERT", *The Journal of The Institute of Internet, Broadcasting and Communication*, Vol. 24, No. 5, pp. 61-67, Oct. 2024. <https://doi.org/10.7236/JIIBC.2024.24.5.61>.
- [8] J. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation", *Proc. of the 39th International Conference on Machine Learning (ICML)*, Baltimore, USA, pp. 12888-12900, Jul. 2022.
- [9] Supreme Court of Korea, "Comprehensive Legal Information on Judicial Precedents", <https://portal.scourt.go.kr/pgp/index.on>. [accessed: Sep. 19. 2025]
- [10] C. Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", *Proc. of the ACL-04 Workshop on Text Summarization Branches Out*, Barcelona, Spain, pp. 74-81, Jul. 2004.
- [11] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi, "BERTScore: Evaluating Text Generation with BERT", *International Conference on Learning Representations (ICLR)*, online, Apr. 2020. <http://doi.org/10.48550/arXiv.1904.09675>.

저자소개

서 동 준 (Dongjoon Suh)



2025년 3월 ~ 현재 : 경기대학교  
컴퓨터과학과 석사과정  
관심분야 : 데이터마이닝,  
컴퓨터비전, 딥러닝, 빅데이터,  
인공지능

정 경 용 (Kyungyong Chung)



2000년 2월 : 인하대학교  
전자계산공학과(공학사)  
2002년 2월 : 인하대학교  
전자계산공학과(공학석사)  
2005년 8월 : 인하대학교  
컴퓨터정보공학부(공학박사)  
2006년 3월 ~ 2017년 2월 :

상지대학교 컴퓨터정보공학부 교수

2017년 3월 ~ 현재 : 경기대학교 AI컴퓨터공학부 교수

관심분야 : 데이터 마이닝, 빅데이터, HCI, 인공지능