

Tabular-TX: In-Context Learning을 통한 주제-설명 구조 기반 표 요약

곽태윤^{o,1,*}, 김지수^{1,*}, 정기용¹, 이동건², 박희선^{1,†}¹성균관대학교 소프트웨어학과, ²포항공과대학교 인공지능대학원

{njj05043, clrdln, wjdrldyd0213}@g.skku.edu, donggeonlee@postech.ac.kr, hspark20@skku.edu

Tabular-TX: Theme-Explanation Structure-based Table Summarization
via In-Context LearningTaeYoon Kwack^{o,1,*}, Jisoo Kim^{1,*}, Ki Yong Jung¹, DongGeon Lee², Heesun Park^{1,†}¹Department of Computer Science and Engineering, Sungkyunkwan University²Graduate School of Artificial Intelligence, Pohang University of Science and Technology

요약

본 논문은 표 데이터를 효율적으로 처리하기 위한 주제-설명 구조 기반 표 요약 (Theme-Explanation Structure-based Table Summarization; **Tabular-TX**) 파이프라인을 제안한다. Tabular-TX는 표 데이터를 하이라이트 된 셀을 중심으로 전처리한 후, 부사구 형태의 주제 부분(Theme Part)과 서술절 형태의 설명 부분(Explanation Part)이 이어지는 구조로 요약 문장을 생성하는 방식이다. 이 과정에서 표의 구조적 특성과 비교 가능성을 고려하여 맞춤형 분석을 수행한다. 또한, In-Context Learning을 사용하여 미세 조정 없이도 LLM의 분석 능력을 최적화하며, 표 데이터의 구조적 복잡성을 효과적으로 처리한다. 제안하는 Tabular-TX를 적용해 표 기반 요약을 생성한 결과, 데이터셋 크기의 제한에도 불구하고 기존의 미세 조정 기반 방법에 비해 우수한 성능을 나타냈다. 실험 결과, Tabular-TX는 복잡한 표 데이터를 보다 효과적으로 처리할 수 있음을 확인했으며, 특히 자원이 제한된 환경에서 표 기반 질의응답과 요약 작업을 위해 사용할 수 있는 새로운 대안임을 입증하였다.

주제어: 표 이해 (Table Understanding), 초대대 언어 모델 (Large Language Model), In-Context Learning

1 서론

표는 복잡한 데이터를 간결하게 정리하고 비교분석하는 데 중요한 역할을 하는 시각화 수단이다. 그러나 표 데이터는 정보가 압축적으로 담겨 있어 텍스트보다 높은 분석력이 요구되며, 현재의 대규모 언어 모델(Large Language Model; LLM)은 이러한 구조적 복잡성을 효과적으로 처리하는 데 한계를 보인다 [1]. 따라서 LLM이 표 데이터를 효과적으로 처리하고 표 기반 질의응답(Table Question-Answering; TableQA)을 수행할 수 있도록 하는 것은 중요한 과제이다.

이전의 표 기반 질의응답과 관련된 연구들은 주로 LLM을 미세 조정(fine-tuning)하는 방법으로 성능을 개선하고자 했다. 그러나 한국어 표 데이터는 학습에 사용할 수 있는 데이터가 제한적이며, 미세 조정에 필요한 막대한 연산 자원과 시간 역시 문제로 지적된다. 예를 들어, FeTaQA [2]와 같은 표 관련 데이터셋을 사용한 연구들은 미세 조정으로 성능을 향상시켰으나, 막대한 컴퓨팅 자원이 필요하다는 단점이 있었다. 한편, 데이터셋 부족 문제를 해결하기 위해 인공지능을 활용한 데이터셋 합성 기법 [3]이 제시되었지만, 생성된 데이터셋을 반복 학습하면 인공지능의 성능이 크게 저하되는 문제 [4]가 있다.

본 연구에서는 국립국어원 언어정보나눔터에서 제공하는 표

* 공동 저자. (These authors contributed equally.)

† 교신 저자. (Corresponding author.)

```
{
  "id": "nikluge-gtps-2023-train",
  "input": {
    "metadata": {
      "title": "4차 산업혁명에 따른 조세환경 변화와 정책 과제",
      "table_title": "4차 산업혁명 관련 조세 부문 주요 의원발의 법률안",
      "date": "2020-06-09",
      "publisher": "국회예산정책처",
      "url": "https://www.nabo.go.kr/Sub/01/Report/01_01_Board.jsp",
      "highlighted_cells": [ [10, 13], [1, 14], [3, 14] ]
    },
    "table": [
      { "value": "조세특례제한법", "is_header": true, "col": 0, "colspan": 4, "row": 0, "rowspan": 1 },
      { "value": "2009580", "is_header": false, "col": 0, "colspan": 1, "row": 1, "rowspan": 1 },
      { "value": "특허 출원-등록비용", "is_header": false, "col": 1, "colspan": 1, "row": 1, "rowspan": 1 },
      { "value": "소득세-법인세 공제", "is_header": false, "col": 1, "colspan": 1, "row": 1, "rowspan": 1 }
      ... 생략 ...
    ]
  },
  "output": [
    "2018년 11월 6일 발의된 부가가치세법의 내용은 부가가치세 과세대상 전자적 용역 범위를 인...생략",
    "부가가치세 과세대상 전자적 용역범위는 인터넷, 클라우드컴퓨팅서비스, 공유경제서비스까지...생략",
    "부가가치세법의 경우 인터넷광고, 클라우드컴퓨팅서비스, 공유경제서비스 등 부가가치세 과세...생략",
    "부가가치세 과세대상 전자적 용역 범위를 인터넷광고, 클라우드컴퓨팅서비스, 공유경제서비스...생략",
    "2018년 11월 6일, 부가가치세법은 인터넷광고, 클라우드컴퓨팅서비스, 공유경제서비스까지 부...생략"
  ]
}
```

그림 1. 표의 일부분에 대한 해석 생성 평가용 말뭉치 [5] 예시

의 일부분에 대한 해석 생성 평가용 말뭉치 [5]를 활용하여, LLM의 표의 일부분에 대한 해석 생성 능력을 향상하는 것을 목표로 한다. 표의 일부분에 대한 해석이란, 그림 1과 같이 표의 메타데이터에서 **highlighted_cells**로 라벨링 된 셀들을 하나의 문장으로 요약하여 설명하는 작업을 의미한다. 이를 위해 본 논문은 주제-설명 구조 기반 표 요약 (Theme-Explanation Structure-based Table Summarization; **Tabular-TX**) 파이프라인을 제안한다.

프라인을 제안한다.

Tabular-TX는 표 데이터의 복잡성을 단순화하고, 데이터의 구조적 특성을 고려한 맞춤형 분석을 수행한다. 구체적으로, Tabular-TX는 표 데이터를 부사구 형태의 주제 부분(Theme Part)과 서술절 형태의 설명 부분(Explanation Part)으로 나누어 요약문을 생성하는 방식을 제시한다. 이를 통해 표의 메타데이터와 셀 데이터를 함께 고려한 종합적인 분석이 가능하도록 설계되었다. 또한, LLM에 기자 페르소나를 부여하여 정보의 객관성과 명확성을 강화하였다. 이로써 본 연구는 표 기반 질의응답과 요약 작업에서의 새로운 접근을 제시한다.

본 연구의 주요 기여는 다음과 같다.

- 주제-설명 구조를 기반으로 하는 표 요약 문장 생성 방법을 제안하여, 표의 일부분에 대한 해석 생성 성능을 크게 향상시킨다.
- 표 데이터의 복잡한 구조적 문제를 해결하여 LLM이 더 잘 이해할 수 있도록 하는 전처리 방법을 제안한다.
- In-Context Learning (ICL)을 통해 미세 조정 없이도 높은 성능을 달성할 수 있음을 입증한다.

2 관련 연구

2.1 표 기반 질의응답(TableQA)

[6]은 표에 대한 복잡한 추론을 수행하기 위해, 기존의 텍스트 기반의 방법인 Chain-of-Thought [7]에서 발전된 표 특화 Chain-of-Table 프레임워크를 제안하였다. 이 방법은 표 내 데이터를 정렬, 추출, 삭제하여 필요한 정보만을 남기고, 표 형태로 통합하여 결과 도출 과정을 단순화한다. 이러한 접근은 표 구조 단순화나 수학적 문제 해결에서 강점을 보이지만, 메타데이터나 배경지식이 필요한 표의 일부분에 대한 해석 생성 작업에는 한계가 있다.

TableLlama [8]은 기존 표 기반 모델들이 특정 작업에만 최적화되는 문제를 해결하기 위해, 14개의 데이터셋을 사용해 하이라이트 된 셀을 활용한 질의응답(Highlighted Cells QA)을 포함한 11개의 작업을 미세조정하였다. 그 결과, 특화 모델과 유사하거나 더 나은 성능을 보였으며, 학습되지 않은 작업에서도 GPT-4 [9]보다 우수한 성과를 기록하였다. 그러나, 모델 학습 과정에 48대의 NVIDIA A100 80GB GPU가 사용되었는데, 이는 미세조정에 상당한 컴퓨팅 자원이 필요함을 의미한다. 따라서 더 적은 자원으로 성능을 높일 방법이 요구된다.

표 해석의 대표적인 벤치마크로는 FeTaQA [2]가 있으며, Chain-of-Table과 TableLlama는 각각 ICL [10]과 미세 조정을 해결책으로 제시했다. 그러나 이 방법들은 메타데이터를 적절히 인용하는 데 한계가 있을 뿐만 아니라, 대규모 컴퓨팅 자원이 필요하다는 문제를 안고 있다.

2.2 표 데이터 부족 문제를 해결하기 위한 데이터 합성 및 In-Context Learning 기법

표 데이터 부족 문제를 해결하기 위한 다양한 접근이 제시되어 왔다. [3]은 인공지능을 활용해 표 데이터셋을 합성하는 방법을 제안하였고, [6]은 Zero-shot Prompting과 Few-shot Prompting을 사용하여 소량의 데이터로도 높은 성능을 달성하는 방법을 소개하였다. 그러나 합성 데이터셋을 반복적으로 학습시킬 경우 데이터의 품질과 다양성이 떨어져 성능 저하로 이어질 수 있다는 연구 결과 [4]가 보고된 바 있다.

Few-Shot Prompting은 모델의 파라미터를 직접 업데이트하지 않고, 몇 개의 예제와 설명, 지시 사항을 입력으로 제공하여 답변을 유도하는 방식이다 [10]. 이 방식은 모델 자체를 재훈련할 필요가 없어 데이터셋이 제한적인 환경에서 유용하다. 본 연구에서는 이를 활용하여 제한된 데이터를 기반으로 표의 일부분에 대한 해석을 생성하고자 한다.

3 주제-설명 구조 기반 표 요약 (Theme-Explanation Structure-based Table Summarization; Tabular-TX) 파이프라인

3.1 Theme-Explanation Structure (주제-설명 구조)

본 논문에서는 표의 일부분에 대한 해석 문장을 생성하기 위해 Tabular-TX를 제안한다. 이 방법론을 통해 생성되는 문장은 표 요약에 적합한 ‘주제-설명 구조’라는 일관된 형식을 가지며, 이는 ‘Theme Part (주제 부분)’과 ‘Explanation Part (설명 부분)’로 구성된다.

3.1.1 Theme Part (주제 부분)

Theme Part는 부사구 형태로, 명사구 형태인 표의 제목(table_title)에 출처나 근거를 나타내는 인용 표현이 결합한 구조다. Theme Part가 이러한 형태를 가지는 이유는 표 요약의 중요한 특성 때문이다. 일반적인 텍스트 요약과 달리, 표에서는 셀 값만으로는 충분한 맥락을 제공하기 어렵다. 표의 메타데이터에서 표의 제목은 전체적인 주제를 확인할 수 있는 유일한 부분이므로, 요약 문장에 반드시 포함되어야 한다. 제목을 포함하지 않을 경우, 그림 3과 같이 문맥이 불명확한 문장이 완성될 수 있다.

예를 들어, “2019회계연도 행정안전부 재정운영표 요약에 따르면 재정운영순원가는 전년 대비 9조 4,350억 1,800만 원 증액된 61조 3,014억 6,300만 원이다.”라는 문장에서, Theme Part는 표의 제목인 “2019회계연도 행정안전부 재정운영표 요약”에 인용 표현인 “~에 따르면”이 붙은 부분이다. 이와 같이 Theme Part는 표의 제목을 인용하여 셀 값에 맥락을 부여하는 역할을 하며, 문장 첫머리에 위치하여 뒤에 오는 Explanation Part의 내용을 더 자연스럽게 이해할 수 있도록 한다.

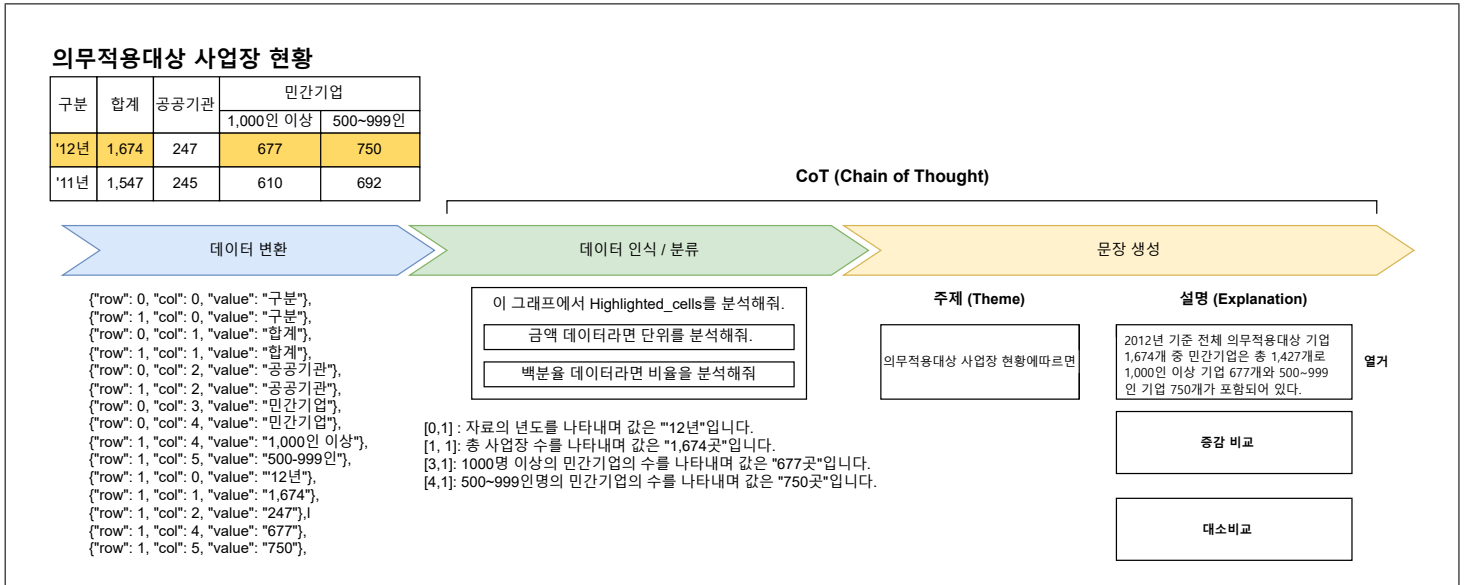


그림 2. 주제-설명 기반 생성(Theme-Explanation Structure-based Table Summarization; Tabular-TX) 파이프라인

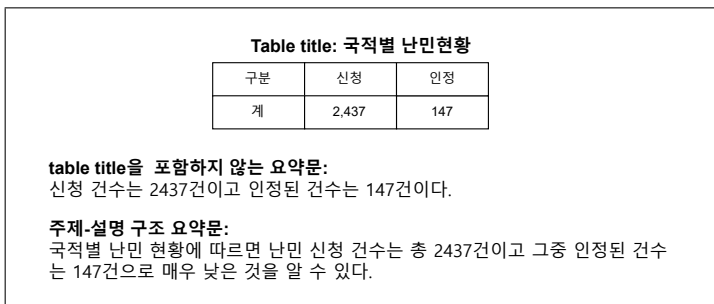


그림 3. Table title을 포함한 문장이 더 문맥을 정확하게 전달하는 모습

3.1.2 Explanation Part (설명 부분)

Theme Part의 뒤에 오는 Explanation Part는 하이라이트된 셀에 대한 분석을 제공하는 서술절로, 요약 문장의 핵심적인 내용을 담고 있다.

이 부분에서는 하이라이트된 셀의 데이터 유형을 고려하여, 정보 열거, 대소 비교, 증감 비교 등의 분석 방법 중 가장 적절한 것을 선택해 사용한다. 이때, 적절한 분석 방법은 셀 간의 비교 가능성을 파악하는 단계에서 결정된다.

예를 들어, “2019회계연도 행정안전부 재정운영표 요약에 따르면 재정운영순원가는 전년 대비 9조 4,350억 1,800만 원 증액된 61조 3,014억 6,300만 원이다.”라는 문장에서, Explanation Part는 “재정운영순원가는 전년 대비 9조 4,350억 1,800만 원 증액된 61조 3,014억 6,300만 원이다.”이다. 이 경우, 비교 가능성을 확인한 후 두 셀 간의 증감 비교를 통해 Explanation Part를 완성하였다.

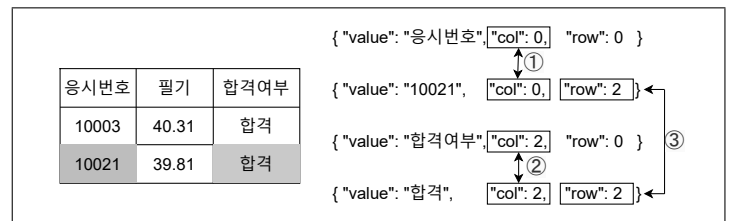


그림 4. 같은 행 혹은 열을 공유하는 데이터 사이의 연관성을 추론하는 모습. ①에서의 추론을 통해 '10021'이 '응시번호'임을 추론. ②에서의 추론을 통해 '합격'의 의미를 추론. ③에서의 추론을 통해 '응시번호'가 '10021'인 수험생이 '합격'함을 추론.

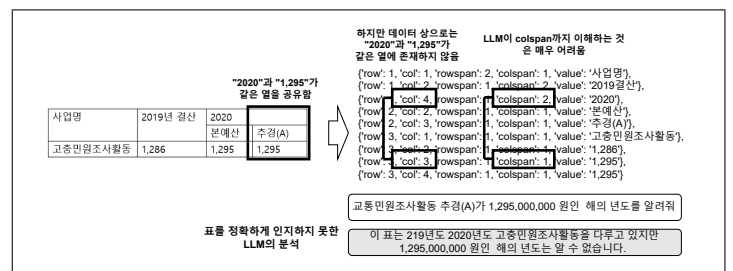


그림 5. 병합된 셀이 표 인식을 저해하는 모습

3.2 문장 생성 방법

주제-설명 구조의 문장을 생성하기까지 Tabular-TX는 여러 과정을 거친다.

3.2.1 데이터 변환

첫 번째 단계는 LLM이 표 데이터를 더 잘 이해할 수 있도록 표 구조를 단순화하는 전처리 과정이다. 먼저, LLM이 표 데이터와 같은 비자연어 데이터를 잘 인식하지 못하는 문제를 해결

하기 위해, 표 데이터를 키-값 쌍의 딕셔너리 형태로 변환한다. 키-값 구조는 자연어 처리에서 자주 사용되므로 표 데이터의 인식률을 크게 향상시킬 수 있었다 [11].

다음으로, 병합된 셀의 처리를 개선하여 표의 구조를 명확히 한다. 병합된 셀은 여러 개의 행 또는 열에 걸쳐 있는 헤더 셀을 의미하며, ‘rowspan’ 및 ‘colspan’으로 그 범위가 지정된다. 그림 4와 같이 LLM은 행 또는 열의 일치를 통해 두 데이터 사이의 연관성을 이해한다. 그러나 병합된 셀의 범위가 제대로 적용되지 않으면 연관성을 잘못 해석하게 된다. 그림 5를 예로 들면, “2020”은 colspan이 2로 지정되어 3열과 4열에 걸쳐 있는 병합된 셀이다. 하지만 LLM은 이를 3열에만 속하는 것으로 인식한다. 이를 해결하기 위해, 병합된 셀을 그 범위만큼 복제하여 LLM이 셀 간의 연관성을 쉽게 찾을 수 있도록 한다.

마지막으로, 변환된 딕셔너리 리스트에서 하이лай트 된 셀과 관련 셀만을 남기고 나머지는 제거한다. 여기서 “관련 셀”이란, 하이лай트 된 셀과 같은 행 또는 열을 공유하는 모든 헤더 셀을 의미한다. 이 과정을 통해 표 데이터의 복잡성을 줄여 LLM의 표 인식을 돕는다.

3.2.2 CoT (Chain-of-Thought)

데이터 변환 후에는 주제-설명 구조에 맞게 표의 일부분에 대한 해석 생성을 시작하는데, 이 작업은 CoT로 구성되어 있어 Compositional Deficiency 문제를 해결한다.

Explanation Part에서 어떤 분석 방법을 사용할지 결정하기 위해 하이лай트 된 셀들이 무엇을 나타내는지를 파악하고, 셀 간의 관계를 고려해야 한다. 이 과정에서 먼저 하이лай트 된 셀을 금액, 백분율, 개수, 문장 등의 유형별로 분류한다. 이 과정을 거치지 않으면, 백분율을 단순 숫자로 파악하는 등의 오류로 인해 비교 분석 과정에서 잘못된 결과가 도출될 수 있다.

다음으로, LLM이 분류된 셀들의 유형에 따라 적절한 능력을 발휘한다. 예를 들어, 금액 데이터의 경우 단위 변환 능력이 필요하고, 백분율 데이터는 소수를 백분율로 변환하는 능력이 요구되며, 문장 데이터는 글을 해석하기 위한 배경 지식이 필요할 수 있다. Tabular-TX에서는 각 셀의 데이터 유형을 인식하여, 데이터 유형에 맞는 분석을 수행하도록 유도하는 프롬프트를 추가하여 LLM의 정확한 인식을 이끌어낸다.

그 후, 데이터 간의 관계를 가장 잘 설명할 수 있는 분석 방법을 선택하여 Explanation Part에 적용한다. 예를 들어, 데이터가 비교 가능하지 않고 단순히 나열되어 있어 각 항목이 독립적으로 중요한 경우에는 정보 열거 방식을 사용할 수 있으며, 가장 큰 값을 가지는 셀 값부터 내림차순으로 설명할 때는 대소 관계 비교, 시간에 따라 값이 증가하거나 감소한 경우에는 증감 비교를 선택한다. 이러한 선택 과정은 요약문의 정보 전달력을 크게 좌우하기 때문에 매우 중요하다.

Table title: 보건복지부 관서 운영비 한도

예산과목	용도
운영비(210)	건당 500만원이하의 경비(단, 공과금 및 위원회 참석비, 선거 관련 용품제작·인쇄비용·우편요금, 청사임차료, 기업특별회계의 당해 사업에 직접 소요되는 경비는 금액제한 없음).
여비(220)	전체
업무추진비(240)	건당 500만원 이하의 경비(단, 기업특별회계의 당해 사업에 직접 소요되는 경비는 금액제한 없음)

highlighted_cells : [(0,1), (1,1)]

기자 페르소나 없이 프롬프트 활용

위의 표의 내용을 요약해줘

운영비(210)는 500만 원 이하의 경비와 기업 특별 회계 사업에 직접 소요되는 경비에 사용되며, 후자의 경우 금액 제한이 없습니다.

기자 페르소나를 적용한 프롬프트 활용

너는 한 문장 짜리 기사를 쓰는 신문 기자야. 이 표에서 highlighted_cells들을 모두 언급하여 딱 한 문장으로 기사를 써줘.

보건복지부 관서 운영비 한도에 따르면 운영비(210)는 공과금 및 위원회 참석비, 선거 관련 용품 제작·인쇄비용, 우편요금, 청사 임차료 등 500만 원 이하의 경비에 사용되며, 기업특별회계의 당해 사업에 직접 소요되는 경비에는 금액 제한이 없다.

그림 6. 페르소나를 적용한 프롬프트가 더 정확한 요약문을 생성하는 모습

최종적으로, 프롬프트를 통해 LLM에 기자 페르소나를 부여하여 주제-설명 구조의 문장을 생성한다. 기자 페르소나를 선택한 이유는 표 요약문과 스트레이트 기사의 작성 방식이 서로 유사하기 때문이다. 스트레이트 기사는 뉴스 기사의 한 종류로, 사실에 기반해 객관적이고 간결하게 정보를 전달하는 기사를 말한다. 표 요약문은 셀 값을 왜곡하지 않고, 주관을 배제하여 간단명료하게 작성되어야 하는데, 이는 스트레이트 뉴스 기사의 작성 방법과 유사하다. 따라서, 기자 페르소나를 적용함으로써 LLM이 핵심 정보를 정확히 판단하고, 셀 값을 인용하여 명료하게 정보를 전달하는 효과를 얻을 수 있다.

4 실험 환경

4.1 데이터셋 및 평가 지표

학습 및 평가에는 국립국어원 언어정보나눔터에서 제공하는 표의 일부분에 대한 해석 생성 평가용 말뭉치 [5]를 사용했으며, 데이터셋의 예시는 그림 1과 같다. 데이터셋은 Train 7,170개, Validation 876개, Test 876개의 표로 구성되어 있다. 각 데이터는 메타데이터 아래 문서 제목, 표 제목, 발행일자, 발행기관, 표 출처 URL, 하이лай트 된 셀 정보, 표 데이터, 그리고 해당 표의 하이лай트 된 부분을 요약하는 문장으로 구성된다.

표의 일부분에 대한 해석 생성 성능을 평가하기 위해 ROUGE-1, ROUGE-L, BLEU 세 가지 지표를 사용하여, 요약문이 표의 핵심 내용을 효과적으로 전달하는 동시에 의미론적으로 높은 완성도를 갖는지 종합적으로 판단한다.

모델	ROUGE-1	ROUGE-L	BLEU	평균
kobart-base-v2 - 미세조정	0.37	0.28	0.35	0.33
EXAONE 3.0 7.8B - ICL	0.21	0.14	0.01	0.12
EXAONE 3.0 7.8B - LoRA	0.27	0.21	0.05	0.17
EXAONE 3.0 7.8B - Table-TX	0.51	0.39	0.44	0.45
llama-3-Korean-Bllossom-8B-ICL	0.33	0.25	0.27	0.28
llama-3-Korean-Bllossom-8B-Table-TX	0.48	0.37	0.42	0.43

표 1. 모델 별 평가지표 점수

4.2 모델

4.2.1 Tabular-TX 파이프라인 적용

본 연구에서는 EXAONE 3.0 7.8B [12] 모델 및 llama-3-Korean-Bllossom-8B¹을 기본 모델로 활용하였다. EXAONE 3.0 7.8B는 한국어 표 질의응답이 포함된 KorWikiTableQuestions [13] 벤치마크에서 1위를 달성한 EXAONE-LM-v1.0의 후속 모델로, 한국어 TableQA 능력이 우수하다. 또한 llama-3-Korean-Bllossom-8B 모델의 경우, 한국어 언어모델 다분야 사고력 벤치마크에서 10B 이하 모델 중 1위를 달성하였다. 본 연구에서는 두 모델에 Tabular-TX를 적용하여 다른 기법들과의 성능을 비교하였다.

4.2.2 In-Context Learning (ICL) 적용

실험에는 기본적인 ICL 기법을 사용하여 표 데이터와 함께 분석을 수행한 후, Tabular-TX를 적용한 경우와 성능을 비교하였다.

4.2.3 매개변수 효율적 미세 조정(LoRA)

Tabular-TX의 자원 효율성과 성능을 평가하기 위해, EXAONE 3.0 7.8B 모델에 매개변수 효율적 미세 조정 기법 중 하나인 LoRA (Low Rank Adaptation) [14] 기반 미세 조정을 적용하였다. `data_point['metadata']`는 그림 1의 메타데이터와 동일하고, `data_point['output']`은 그림 1의 output과 동일하다.

4.2.4 모델 전체 미세조정

크기가 작은 모델을 미세 조정하였을 때의 성능을 테스트하기 위해 gogamza/kobart-base-v2² 모델의 파라미터 전체를 미세 조정하였다.

5 실험 결과

표 1은 다양한 모델의 성능을 ROUGE-1, ROUGE-L, BLEU로 평가한 결과 및 평균 점수를 보여준다. kobart-base-v2 모델

은 미세 조정 후 평균 0.33점을 기록했고, EXAONE 3.0 7.8B 모델은 ICL 방식에서 0.12점, 미세 조정 후 0.17점, Tabular-TX 방식에서는 0.45점을 달성했다. llama-3-Korean-Bllossom-8B 모델도 평균 0.43점으로 성능이 크게 향상되어, Tabular-TX 방식이 전반적으로 가장 높은 성능을 보였음을 확인하였다.

EXAONE 3.0 7.8B 모델과 kobart-base-v2 모델을 같은 데이터셋으로 미세 조정했음에도 성능 향상 정도에 차이가 있는 이유는 multiplicative joint scaling law [15] 때문이다. 멀티플리케이션 확장 법칙은 학습에 사용된 데이터셋의 크기가 모델 크기에 비해 충분히 크지 않은 경우 성능 향상이 미미함을 의미한다. kobart-base-v2의 크기는 124M지만 EXAONE 3.0 7.8B는 7820M이므로 비슷하거나 더 나은 수준의 성능 향상을 위해서는 약 63배 큰 데이터셋을 학습시켜야 한다.

본 연구는 제한된 Tabular-TX를 통해 미세 조정 없이도 표 데이터 분석 성능을 향상시킬 수 있음을 확인하였다. 특히 Tabular-TX는 미세 조정 모델에 비해 적은 데이터셋에서도 우수한 성능을 보였으며, 이는 제한된 학습 데이터 환경에서도 Tabular-TX가 유용한 대안이 될 수 있음을 시사한다. 또한, Tabular-TX는 일반적인 ICL을 적용하였을 때와 대비하여 약 4배 높은 평균 성능을 기록하였다.

그러나 Tabular-TX는 숫자 계산이 필요한 돈의 단위, 비율, 평균 등의 카테고리에서 성능이 낮았다. '100,000천 원'과 같은 잘못된 단위 표기가 대표적이다. 이는 모델의 수학적 추론 능력이 부족했기 때문으로 분석되며, 이를 해결하기 위해 수학 연산에 특화된 모델의 도입이 필요하다. 이러한 결과는 ICL의 확장 가능성을 강조한다.

6 결론

본 연구는 제한된 표 데이터셋 및 연산 자원에서 표의 일부분에 대한 해석 생성 작업의 성능을 높이기 위한 주제-설명 기반 생성(Theme-Explanation Table Summarization; Tabular-TX) 파이프라인을 제안하였다. 연구의 주요 목표는 ICL을 적용하여 대규모 데이터 학습 없이도 복잡한 표 데이터를 효율적으로 분석하는 방법을 탐색하는 것이다.

¹MLP-KTLim/llama-3-Korean-Bllossom-8B

²<https://huggingface.co/gogamza/kobart-base-v2>

연구 결과, Tabular-TX가 표 데이터 분석에 유용한 도구임을 확인하였다. 표 데이터를 체계적으로 요약할 수 있는 새로운 문장 생성 방법인 주제-설명 구조를 제시함으로써, 복잡한 표 데이터 요약문 생성에 기여하였다. 또한, Tabular-TX 파이프라인을 통해 각 표 유형의 특성을 반영한 정교한 분석이 가능함을 보였다. 마지막으로, Tabular-TX는 미세조정 없이도 ICL을 통해 우수한 성능을 보여주었으며, 제한된 학습 환경에서도 탁월한 성능을 입증하여, 향후 모델의 직접적인 학습 없이 표 데이터 분석에 중요한 기여를 할 수 있음을 보여주었다.

그러나, EXAONE 3.0 7.8B 및 llama-3-Korean-Blossom-8B 모델 외 다양한 LLM에 Tabular-TX를 적용한 사례를 충분히 제시하지 못했다는 한계가 있다. 향후 연구에서 다양한 특성을 가진 LLM 모델에 Tabular-TX를 적용하여, 모델별 특성을 고려한 성능 비교와 Tabular-TX의 범용성을 검증할 필요가 있다. 또한, 여러 도메인에서의 적용 가능성에 대한 검토도 이루어져야 할 것이다.

참고문헌

- [1] J. Zhao, J. Tong, Y. Mou, M. Zhang, Q. Zhang, and X. Huang, “Exploring the compositional deficiency of large language models in mathematical reasoning,” *CoRR*, Vol. abs/2405.06680, 2024.
- [2] L. Nan, C. Hsieh, Z. Mao, X. V. Lin, N. Verma, R. Zhang, W. Kryscinski, H. Schoelkopf, R. Kong, X. Tang, M. Mutuma, B. Rosand, I. Trindade, R. Bandaru, J. Cunningham, C. Xiong, and D. R. Radev, “Fetaqa: Free-form table question answering,” *Trans. Assoc. Comput. Linguistics*, Vol. 10, pp. 35–49, 2022.
- [3] Q. Hou, J. Wang, M. Qiao, and L. Tian, “Synthesizing realistic data for table recognition,” *CoRR*, Vol. abs/2404.11100, 2024.
- [4] S. Alemohammad, J. Casco-Rodriguez, L. Luzi, A. I. Humayun, H. Babaei, D. LeJeune, A. Siahkoohi, and R. G. Baraniuk, “Self-consuming generative models go MAD,” *The Twelfth International Conference on Learning Representations*, 2024.
- [5] 언어정보나눔터, “표의 일부분에 대한 해석 생성 평가용 말뭉치,” https://kli.korean.go.kr/corpus/taskOrdtm/taskDownload.do?taskOrdtmId=109&clCd=ING_TASK&subMenuId, Sep 2024.
- [6] Z. Wang, H. Zhang, C.-L. Li, J. M. Eisenschlos, V. Perot, Z. Wang, L. Miculicich, Y. Fujii, J. Shang, C.-Y. Lee, and T. Pfister, “Chain-of-table: Evolving tables in the reasoning chain for table understanding,” *The Twelfth International Conference on Learning Representations*, 2024.
- [7] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, “Chain-of-thought prompting elicits reasoning in large language models,” *Advances in Neural Information Processing Systems* 35, 2022.
- [8] T. Zhang, X. Yue, Y. Li, and H. Sun, “Tablellama: Towards open large generalist models for tables,” *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 6024–6044, 2024.
- [9] OpenAI, “GPT-4 technical report,” *CoRR*, Vol. abs/2303.08774, 2023.
- [10] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language models are few-shot learners,” *Advances in Neural Information Processing Systems*, Vol. 33, pp. 1877–1901, 2020.
- [11] E. Stengel-Eskin, K. W. Murray, S. Zhang, A. S. White, and B. V. Durme, “Joint universal syntactic and semantic parsing,” *Trans. Assoc. Comput. Linguistics*, Vol. 9, pp. 756–773, 2021.
- [12] LG AI Research, “Exaone 3.0 7.8b instruction tuned language model,” *CoRR*, Vol. abs/2408.03541, 2024.
- [13] C. Jun, J. Choi, M. Sim, H. Kim, H. Jang, and K. Min, “Korean-specific dataset for table question answering,” *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pp. 6114–6120, 2022.
- [14] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” *The Tenth International Conference on Learning Representations*, 2022.
- [15] B. Zhang, Z. Liu, C. Cherry, and O. Firat, “When scaling meets LLM finetuning: The effect of data, model and finetuning method,” *The Twelfth International Conference on Learning Representations*, 2024.