
COSE474-2024F: Final Project Proposal

“경제뉴스의 사실과 의견 요약 알림”

2022320009 이수현

1. Introduction

경제를 이해하고 공부하려면 경제 기사를 읽는 것부터 시작하라는 말을 많이 들었습니다. 그러나 다양한 언론사가 실시간으로 기사를 보도하다 보니, 어떤 것이 중요한 뉴스인지 파악하기 어렵다는 것을 느꼈습니다.

또한 경제 공부 초보자에게는 경제 이슈가 좋은 것인지, 그리고 이러한 사실이 경제 전반에 어떤 영향을 미칠지를 판단하기 쉽지 않습니다. 뉴스에 언론사의 시각이 들어가 편향적인 정보를 제공하는 점도 객관적인 경제 사실을 파악하는데 혼란을 주기도 합니다.

이러한 문제를 해결하고자, 본 프로젝트는 매일 객관적인 경제 이슈를 제공하고, 이에 대한 다양한 시각의 뉴스를 요약하여 제공하는 서비스를 목표로 합니다.

2. Problem definition & challenges

본 프로젝트는 경제뉴스에서 매일 중요한 경제뉴스를 요약하여 받아볼 수 있는 시스템을 개발하는 것입니다. 이 시스템의 특징은 경제 뉴스를 객관적인 사실과 그에 대한 다양한 관점을 분리해서 제공한다는 것입니다. 매일 같은 시간에, 어제의 많은 뉴스 중 중요한 “객관적인 경제 이슈”를 제공하고 그에 대한 다양한 관점의 기사를 요약하여 제공합니다.

프로젝트 과정 중에 예상되는 어려움은 다음과 같습니다.

첫째, 많은 언론사에서 같은 주제에 대한 기사를 보도하기 때문에 기사들 간의 중복 문제가 발생합니다. 둘째, 뉴스에서 객관적인 사실과 언론사의 관점이 포함된 의견의 구분이 필요합니다. 셋째, 기사의 편향성 분석입니다. 뉴스에는 각 언론사의 시각이 반영되기 때문에 다양한 시각의 뉴스를 제공해야 합니다.

3. Related Works

대형 언어 모델을 사용한 뉴스 요약 기술은 최근 많은 연구에서 주목을 받고 있습니다. PEGASUS는 기사 요약 작업에서 높은 성능을 보인 모델입니다. LLaMA 모델도 이러한 요약을 잘 할 수 있으며, 특히 생성적 요약에서 뛰어난 성능을 보입니다. 사실, 의견 구분에서 대형 언어 모델들은 대량의 데이터를 기반으로 사실적인 정보와 의견을 구분하는데 높은 성능을 발휘할 수 있습니다. .

4. Datasets

같은 객관적인 경제 사실에 대한 다양한 관점의 뉴스를 요약해 제공하기 위해 뉴스별 편향도를 분석해야 합니다. 이 과정에서 아래의 데이터셋을 이용하고 네이버 뉴스 크롤링을 할 예정입니다.

금융 뉴스 문장 감성 분석 데이터셋 :

https://github.com/ukairia777/finance_sentiment_corpus

그 외에도 네이버 뉴스에서 경제 부분을 크롤링 할 예정입니다.

5. State-of-the-art methods and baselines

본 프로젝트는 경제 뉴스 요약, 객관적인 경제 이슈 추출, 뉴스 편향성 분석을 목표로 합니다. 최신 SOTA 모델들인 PEGASUS, BART, BERT, GPT-4는 뉴스 요약 및 편향성 분석에서 뛰어난 성능을 보입니다. LLaMA 3는 계산 효율성과 다양한 시각 분석에서 장점이 있어, 적은 자원으로도 높은 성능을 발휘할 수 있습니다. LLaMA 3는 경제 뉴스 요약 및 분석에 적합한 모델로 평가됩니다.

6. Schedule

- 1주차 : 프로젝트 데이터 수집 및 전처리
- 2주차 : llama3기반 객관적인 이슈 추출
- 3주차 : llama3기반 뉴스 편향성 분석
- 4주차 : llama3기반 뉴스 요약
- 5주차 : 사용자 인터페이스 구현

References

- Hoang, M., Bihorac, O. A., and Rouces, J. Aspect-based sentiment analysis using bert. In Hartmann, M. and Plank, B. (eds.), *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, chapter 1, pp. 187–196. Lawrence Erlbaum Associates, Inc., Turku, Finland, 2019.
- Zhang, J., Zhao, Y., Saleh, M., and Liu, P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization, 2019.