Extracting Economic Issues from News Data: With the Help of Generative AI

Younghwan Lee

Bank of Korea

Disclaimer: The views expressed in this talk do not reflect necessarily those of Bank of Korea

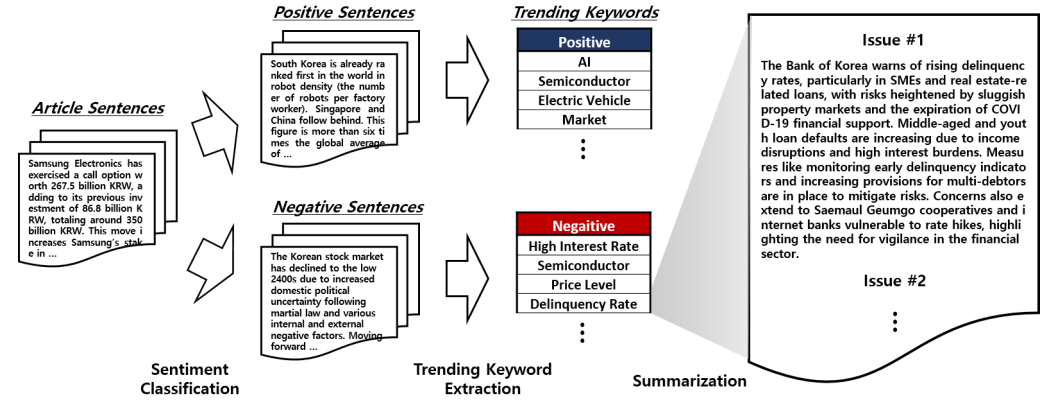
Motivation

- Traditional economic monitoring frameworks that rely on economic variables such as GDP and CSI might only be a second-best method.
  - Doctors treat diseases, not symptoms.
  - Similarly, policymakers must address economic issues, not just economic variables.

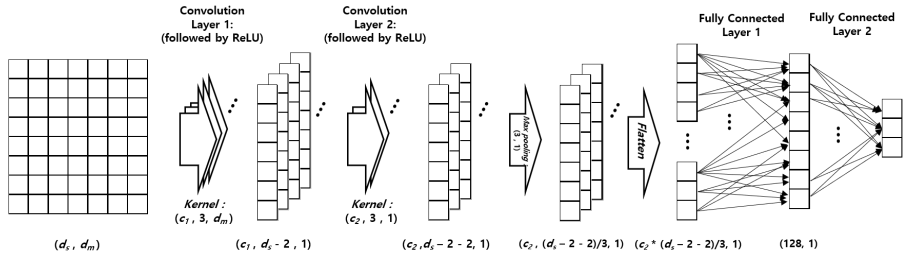
- However, directly monitoring economic issues is challenging in practice, as they are often linguistic rather than numeric.

- In this study, I present a methodology to incorporate economic issues into the monitoring framework that is:
  i) Verifiable,
  ii) Flexible,
  iii) **Easy to implement**.

Snapshot

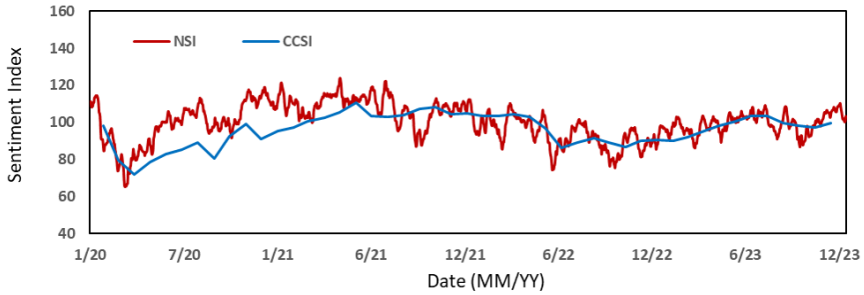

Article Sentences
Samsung Electronics has exercised a call option worth 267.5 billion KRW, adding to its previous investment of 86.8 billion KRW, totaling around 350 billion KRW. This move increases Samsung's stake in ...
Positive Sentences
South Korea is already ranked first in the world in robot density (the number of robots per factory worker). Singapore and China follow behind. This figure is more than six times the global average of ...
Negative Sentences
The Korean stock market has declined to the low 2400s due to increased domestic political uncertainty following martial law and various internal and external negative factors. Moving forward ...
Sentiment Classification
Trending Keywords
Positive
AI
Semiconductor
Electric Vehicle
Market
Negaitive
High Interest Rate
Semiconductor
Price Level
Delinquency Rate
Trending Keyword Extraction
Issue #1
The Bank of Korea warns of rising delinquency rates, particularly in SMEs and real estate-related loans, with risks heightened by sluggish property markets and the expiration of COVID-19 financial support. Middle-aged and youth loan defaults are increasing due to income disruptions and high interest burdens. Measures like monitoring early delinquency indicators and increasing provisions for multi-debtors are in place to mitigate risks. Concerns also extend to Saemaul Geumgo cooperatives and internet banks vulnerable to rate hikes, highlighting the need for vigilance in the financial sector.
Issue #2
Summarization

Classification: Methodology



- CNN (Convolutional Neural Network) is applied to classify sentiment into positive, neutral, and negative categories.
  - i) Tokens are generated using **spaCy** and embedded into 128-dimensional vectors via **Word2Vec**.
  - ii) Token vectors are concatenated into a 100×128 matrix, with zero-padding added if needed for consistent dimensions.

Classification: Result



- Sample 10,000 sentences daily to derive an index. The results show it is highly correlated with existing sentiment indices.
  - $$X_t = \frac{\#\,\text{Positive Sentences} - \#\,\text{Negative Sentences}}{\#\,\text{Positive Sentences} + \#\,\text{Negative Sentences}}$$
    - Scale and translate X_t to resemble a standardized index (Mean = 100, Std. = 10).

Trending Keyword: Idea

- It is inefficient to directly work with a large set of sentences. We need a method to focus on a few important issues that characterize current economic conditions.
- ***Trending Keywords*** are defined as keywords extracted from news headlines during a specific time period that best describe the key issues of that period.
- Then, how can we capture the ***Trending Keywords***?
- Suppose that you are provided with a collection of news headlines, and you are asked to guess when those headlines were published based on the keywords in the headlines.
- I define ***Trending Keywords*** as a set of keywords that are most useful for us to address the question above.

Trending Keywords: Methodology

- The Neyman-Pearson lemma states that the log-likelihood test is the most powerful test. By utilizing the empirical likelihood ratio of keyword frequency, we can address the question at hand.

- For each keyword $i \in I$ define trending score S_i as follow:

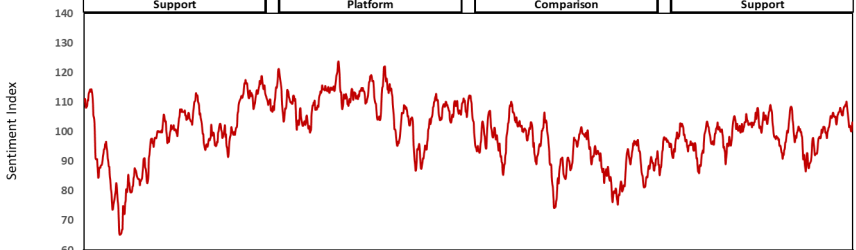
$$S_i = \Pr[i|\text{target period}] \ln \frac{\Pr[i|\text{target period}]}{\Pr[i|\text{base period}]}.$$

- S_i increase when the keyword i
  - i) is encountered more frequently in the headlines during the target period,
  - ii) is relatively uncommon in the base period.
- We can extract ***Trending Keywords*** for the target period by ranking keywords according to their trending score.

Trending Keywords: Result

- ***Trending Keywords*** from 2020 summarize the captures important issues after COVID-19 outbreak.

2020	2021	2022	2023
COVID19	COVID19	ESG	AI
Spread	Vaccine	Electric Vehicle	Semiconductor
Crisis	ESG	LG Energy Solution	Electric Vehicle
Non-face-to-face	Eco-friendly	Price	Market
Support	Platform	Comparison	Support



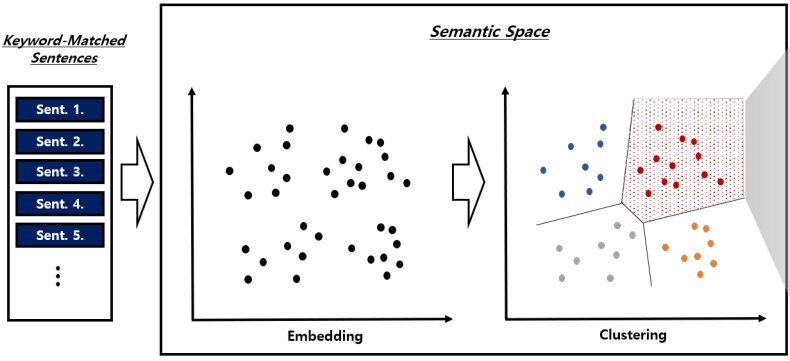
2020	2021	2022	2023
COVID19	COVID19	Ukraine	High Interest Rates
Spread	Vaccine	Inflation	Semiconductor
Crisis	Variant	Russia	Inflation
Cases	Inflation	Interest Rates	Delinquency Rate
Resurgence	Bitcoin	Increase	Real Estate

Summarization: Methodology

- The goal of summarization is to extract information comprehensively and deliver it succinctly.
  - Working with keywords is efficient but often lacks clarity and sufficient context.
  - Reviewing all keyword-matched sentences is inefficient as many convey similar meanings.

- Strategy:
  i) **Embed**: Transform keyword-matched sentences into semantic vectors (**Word2Vec**).
  ii) **Cluster**: Group sentences by semantic similarity using K-means (via **FAISS**).
  iii) **Summarize**: Generate concise summaries for each cluster with a pretrained LLM (**Llama 3.2**).

Summarization: Result



Keyword: High-Interest Rate Issue#1

The prolonged high-interest rate environment is driving up loan delinquency rates, though banks assess the situation as "not alarming." Housing purchases have dropped to 30% on average, and rising labor costs (49.7%) remain a key issue for SMEs. The real estate market continues to stagnate, with capital gains taxes down by 20.7 trillion won compared to the previous year.
High-interest loans are increasing, fueling concerns over the commercial real estate market, where vacancies and related loan delinquencies have reached a 10-year high. Globally, the impact of high rates has led to a 9% decline in net profits for 13,000 publicly listed companies, marking four consecutive quarters of contraction.
In Seoul, significant housing supply overlaps have caused steep declines in rental prices, with jeonse rates in Gangnam falling to 40%. This highlights the severe freeze in real estate transactions driven by the high-interest environment.

Summarization

Concluding Remarks

- This study proposes a verifiable, flexible, and easy-to-implement methodology to extract economic issues from a large volume of news articles.
  - With the help of generative AI and pretrained LLMs, the unstructured nature of economic issues is effectively addressed.
- Future Research:
  - Exploring how to predict next quarter's trending keywords and summarized issues.