

News Sentiments and Their Contagion Effects

Overview

This work aims to leverage the power of massive news data and machine learning models to improve the understanding of the influence of news articles on the US stock market.

Objectives

1. Retrieve a large newspaper corpus
2. Given a keyword (“tariff”), extract relevant articles
3. Given articles, compute their sentiment scores
4. Attach sentiment scores to companies of interest
5. Conduct linkage/time series analysis

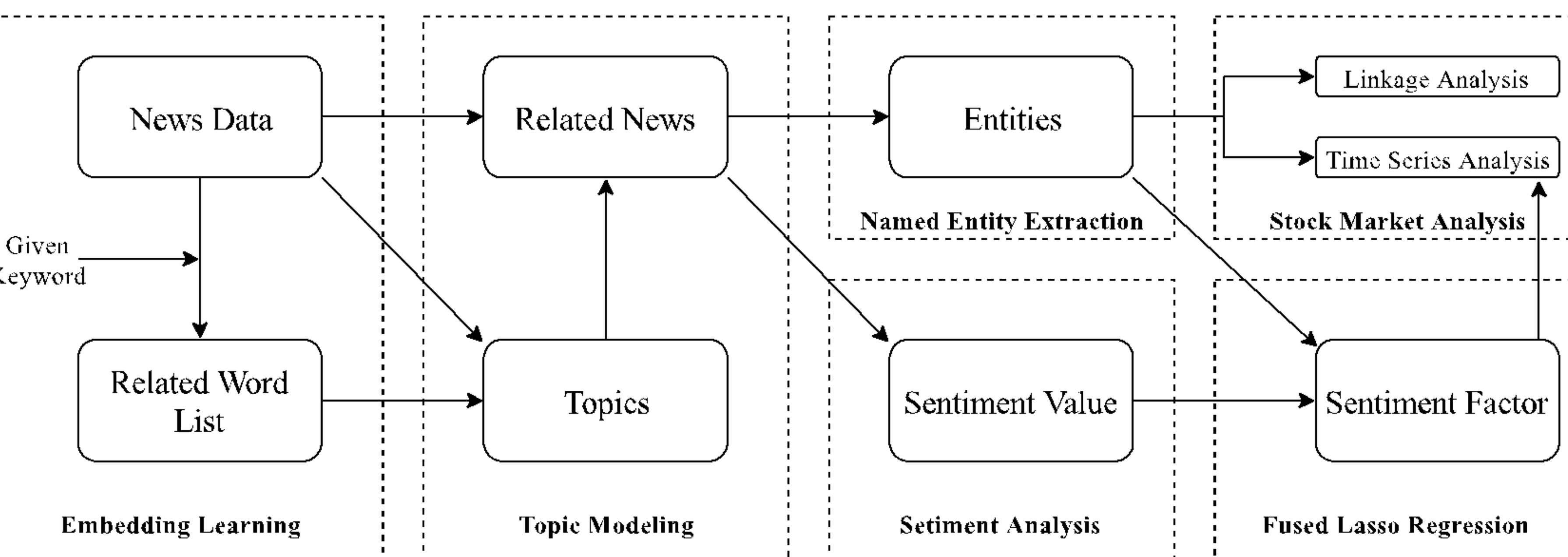


Figure 1. Pipeline of our framework for analyzing the news' effect on stock.

Data Retrieving and Article Extraction

Handmade scrappers for 8+ websites/APIs

Word2Vec -> Semi-supervised LDA [1] for topic learning

“Steel Tariff”	“China Trade War”	“Climate Change”	“Fairness”	“Brexit”
stephen	China	Climate	Woman	Johnson
robinson	Chinese	Said	Men	Brexit
Steel	Trade	Change	Black	Mp
Supplier	US	Year	People	Minister
Contempt	Economy	Energy	White	Would
Inflation	Tariff	Water	Female	Labour
Policy	Economic	World	Sexual	May
Good	World	Emission	Sex	Said
Tariff	Market	Global	Right	Prime
UK	good	new	girl	Parliament

Table 1. Sample extracted topics, first two topics can be identified as related to

References

- [1] Matthew Hoffman, Francis R Bach, and David Blei. Online learning for latent dirichlet allocation. In advances in neural information processing systems, pp. 856–864, 2010.
- [2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

Sentiment Analysis, Named Entity Recognition, Sentiment Factor

We used BERT [2] as the sentiment analysis model, spaCy as the NER tool and fused Lasso regression as the sentiment factor model.

Linkage Analysis Result

We calculated the number of news articles that target entities were co-mentioned together. Then we conducted panel OLS regression to discover the relationship between linkage indicator and stock return. The result shows that the coefficient for linkage indicator with a high level of statistical significance. It indicates that the stock return of company is correlated with linkage indicator, which can be considered as prior stock return of its connected companies.

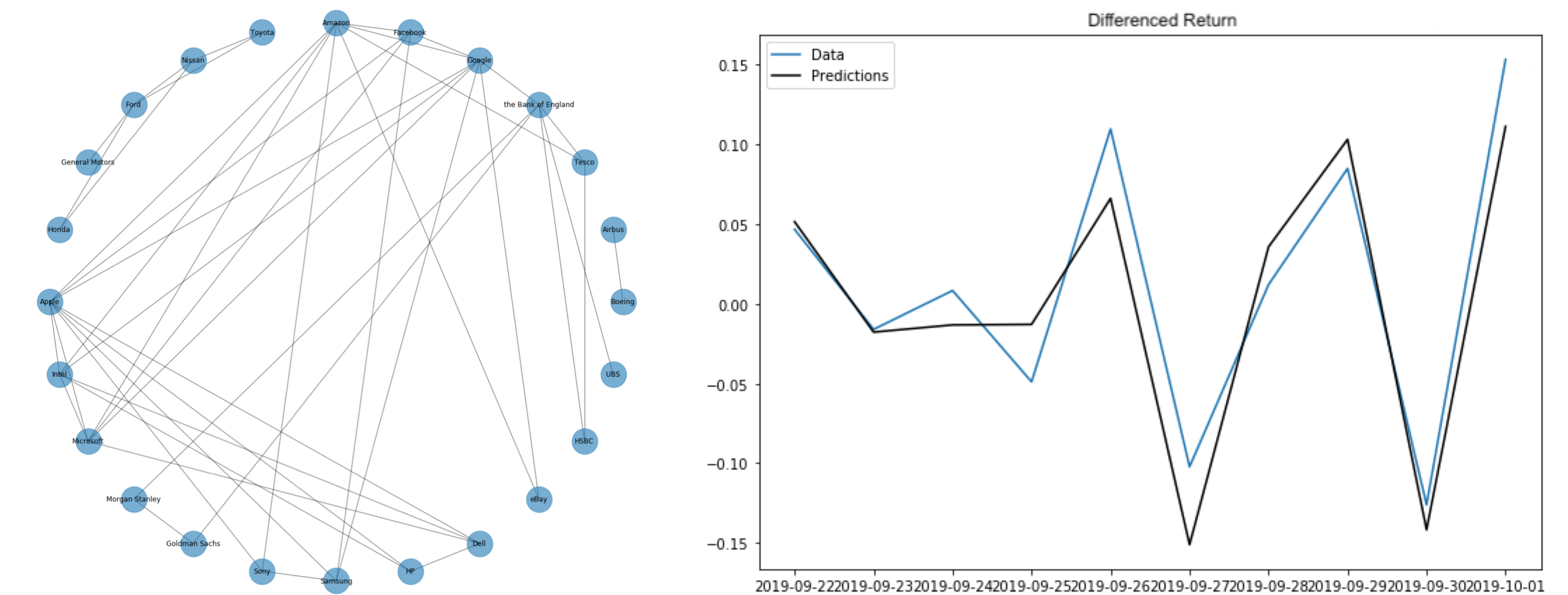


Figure 3. Results of linkage analysis and time series analysis.

Time Series Analysis Result

We have tried VAR and ARIMAX model, and tested the sentiment factors statistical properties. The sentiment factor is significantly related with the stock return based on Cointegration test, with either positive correlation (majority) or negative correlation.

In both cases, adding sentiment factor improves the models' performances for most of the companies we observed.

Conclusion

We arrive at the conclusion that media sentiment indeed affects stocks performance, or to be more precise, there exists a correlation between these two with strong statistical significance.

In this work, we constructed a pipeline with SOTA models (BERT, NER) and customized models (fused Lasso, VAR, ARIMAX, OLS Panel) such that given any keyword, we are able to analyze the effect of its corresponding articles on the companies of interest.

Acknowledgments We thank Adam for his advice, Stephen, Brennan, Chuqi and Karly for their support and insights during this collaboration.