

# MERIT Internship Report

Department of Applied physics, Sagawa group D2  
MERIT 10th Takuya Kamijima (37-227110)

## Internship Overview

Recipient: Mizuho-DL Financial Technology

Period: 4/17/2023 - 5/19/2023

## Theme 1 Analysis of timely disclosure using Sentence-BERT

Timely disclosure is important corporate information that listed companies are required to disclose for the purpose of fair share pricing and investor protection. This information is used to identify potential corporate defaults or signs of deteriorating performance and to make portfolio changes, but it is difficult for the human eye to evaluate the lengthy text that is updated on a daily basis. The theme of this project is to investigate the possibility of using machine learning, especially natural language processing, which is currently having a great impact on society, to automatically analyze and evaluate corporate risk.

In natural language processing, sentences are not used as they are, but are converted into sentence vectors before learning. A sentence vector is an embedded representation of a sentence and is represented using real numbers of several hundred dimensions ( $d=768$ ). To convert the sentences for timely disclosure, we used Sentence-BERT [1], which is an improved version of the BERT model with enhanced suitability for sentences. In actual application, it is necessary to confirm that the sentences are converted appropriately, and the accuracy and precision of the conversion should be verified. However, since we focused on unsupervised learning this time, labels such as the context of the input sentences were not available. Therefore, we visualized the sentence vector embedded in a lower dimension ( $d=2$ ) using a dimensionality reduction method called UMAP [2], a method that preserves and transforms the local and global structure of a high-dimensional space.

Figure 1 shows the distribution of all documents of timely disclosures published in 2020 after embedding. It can be seen that the overall distribution is divided into (1) small clusters at the periphery, (2) large clusters at the center, and (3) isolated dots at the outer edge. In (1), documents are separated by their main meaning, such as "symbols" (i.e. misacquired documents) or "dividends". In (2), the clusters are combined, and it is confirmed that the "financial statements" documents are widely distributed. We confirmed that the sentence vectors generated by Sentence-BERT form clusters according to the context to some extent, but more detailed labeled data is needed to verify the accuracy of the clusters.

We also checked the consistency with other document classification methods and developed an anomaly detection method using UMAP, but I will not go into the details.

## Theme 2 An Empirical Analysis of International Diversified Investment Strategies in a Stochastically Fluctuating Investment Environment: Application of the Regime-switching Model

Asset prices traded in the market are constantly fluctuating, and those that hold positions are exposed to the risk of future price fluctuations. The magnitude of price fluctuations, or volatility, is also known to vary over time. It has been observed that during recessions, such as the Lehman Shock, prices fall and volatility rises. Although international diversification was said to be an excellent strategy for reducing risk, its superiority has been questioned since it was confirmed that asset correlations become stronger during recessions. In this time we examine international diversification strategies using a stochastic model called the regime-switching model and develop a superior investment strategy.

The regime-switching model [3,4] is a model that assumes that an invisible regime exists in the market and changes over time, and that asset prices fluctuate according to this regime. The unobservable regimes (in this case, boom and bust) evolve over time in a Markov chain, and asset prices (or more precisely, log returns) evolve in an AR or diffusive manner. The parameters of the model are determined by maximum likelihood estimation from historical data. Fitting with weekly U.S. stock futures data confirms that the model can be separated into a "boom" regime with positive average returns and low volatility and a "bust" regime with negative average returns and high volatility (see Figure 2). The "bust" regime includes the Lehman Shock and the Corona Shock, etc., and is found to be generally consistent with the actual experience.

We also developed investment strategies using the regime-switching model. The model was developed using a regime-switching model, which is to invest more in stocks during a boom and more in bonds during a recession. We have confirmed that this strategy is superior to a simple mean-variance strategy in terms of the Sharpe ratio, but I will not go into details.

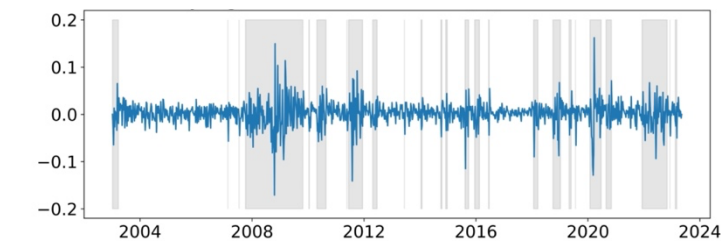
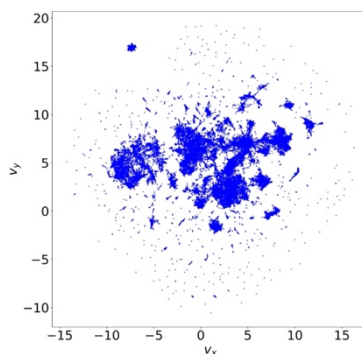


Fig.1 (left) : The distribution of sentence vectors embedded in the low dimension.

Fig.2 (right) : Time variation of the log return of the US stock return. The shaded region represents the "bust" regime judged by the model.

[1] Qiita : 【日本語モデル付き】 2020 年に自然言語処理をする人にお勧めしたい文ベクトルモデル

[2] arXiv:1802.03426

[3] International Asset Allocation with Regime Shifts, Andrew Ang and Geert Bekaert

[4] マルコフスイッチングモデルのマクロ経済・ファイナンスへの応用 沖本竜義

## **Acknowledgments**

I would like to express my sincere appreciation to Mr. Serizawa, who took care of the administrative procedures from the time I applied for the internship, and to Mizuho–DL Financial Technology Co. In particular, I would like to thank Mr. Kaneko of the Risk Management Technology Development Department and Mr. Chen and Mr. Sasaki of the Financial & Public Corporations Advisory Department for their assistance. Thank you very much. I enjoyed the opportunity to gain a better understanding and resolution of the financial industry, and to experience new technologies and knowledge, and to apply them to practical work to create value. Finally, I would like to express my gratitude to my supervisor, Prof. Sagawa, and my associate supervisor, Prof. Arita, who encouraged me to participate in the long-term internship program, and to the MERIT program.