

MERIT 国内インターンシップ報告書

工学系研究科 物理工学専攻 沙川研究室 博士課程 2 年

MERIT 10 期生 上島卓也 (学籍番号:37-227110)

インターンシップ概要

受け入れ先:みずほ第一フィナンシャルテクノロジー

実施期間: 2023 年 4 月 17 日 ~ 2023 年 5 月 19 日

テーマ 1 適時開示の Sentence-BERT 文ベクトルの分析

適時開示とは、公正な株価等の形成および投資者保護を目的として上場会社に開示が義務付けられている重要な会社情報である。この情報から企業のデフォルト可能性や業績悪化の兆候を掴んでポートフォリオの変更などが行われるわけだが、日々更新されている長大な文章を人の目で評価するのは大変である。そこで、現在社会に大きなインパクトを与えている機械学習、特に自然言語処理を用いて、それらを自動で分析し企業リスク評価をすることはできないかというのが前半のテーマである。

自然言語処理においては、文章をそのまま使わずに文ベクトルへの変換が行ってから学習を行う。いわば文ベクトルとは文章の埋め込み表現であり、数百次元 ($d=768$) の実数を用いて表される。適時開示の文章を変換するにあたって、BERT というモデルの文章への適性を高めた改良版である Sentence-BERT [1]を使用した。実際の応用にあたって、文章が適切に変換されているか確認する必要がある、その正確性・精緻性を検証した。ただし、今回は教師なし学習であるため、入力される文章がどのようなコンテキストを持つかなどといったラベルは含まれていない。そこで、UMAP [2]と呼ばれる次元削減手法を用いて文ベクトルを低次元($d=2$)に埋め込んで可視化した。UMAP とは高次元空間の local な構造と global な構造を保持して変換する手法である。

図 1 は 2020 年の公開された適時開示の全文書の埋め込み後の分布である。全体の分布として、(1)周辺部の小クラスタ、(2)中央部の大クラスタ、(3)外縁部の点々に分かれていることがわかる。(1)では文書が「記号」(取得ミス of 文書)や「配当」といった主旨で分離されていることが確認された。(2)ではクラスタが結合しており、「決算」の文書が広範に分布していることが確認された。(3)ではいずれのクラスタにも属さない・共通の主旨が確認できない点がまばらに分布していることが確認された。Sentence-BERT で生成された文ベクトルが、ある程度文脈に従ってクラスタを形成していること

は確認できたが、その精度を検証するためにはより詳細なラベルのあるデータが必要になると考えられる。

この他にも他の文書分類手法との整合性を確認したり、UMAP を利用した異常検知手法の開発なども行ったりしたが、その詳細は割愛させていただく。

テーマ 2 確率的に変動する投資環境における国際分散投資戦略の実証分析：Regime-switching model の活用

市場で取引される資産価格は常に変動しており、ポジションを持つものは将来的な価格変動のリスクにさらされている。価格の揺らぎの大きさ、すなわちボラティリティも時間的に変動することが知られている。リーマンショックなど不況時には価格が下落し、ボラティリティは上昇することが確認された。国際分散投資はリスクを低減する優れた戦略であると言われていたが、不況時には資産相関が強くなることが確認されたためその優位性に疑問が呈されるようになった。後半では、レジームスイッチングモデルという確率モデルを用いて国際分散投資戦略の検証およびより優れた投資戦略の開発を行った。

Regime-switching model [3,4]とは、市場に見えざる状態（レジーム）が存在して時間と共に変遷し、資産価格がそのレジームに応じて変動すると考えるモデルである。観測できないレジーム（ここでは好況と不況）は Markov 連鎖で時間発展し、資産価格（正確には対数リターン）が AR もしくは拡散的な時間発展をすると考える。モデルのパラメータは過去データから最尤推定を行なって決める。米国株式先物の週次データでフィッティングを行ったところ、平均収益が正かつボラティリティの小さい「好況」レジームと、平均収益が負かつボラティリティの大きい「不況」レジームに分離できることが確認できた（図 2 参照）。不況はリーマンショックやコロナショック等を含んでおり、概ね実感と適合することがわかった。

また、Regime-switching model を用いて投資戦略の開発を行なった。好況の場合は株式多め、不況の場合は債券多めといった具合である。単純な平均分散の戦略よりもシャープレシオの意味で優れていることが確認できた。詳細は割愛させていただく。

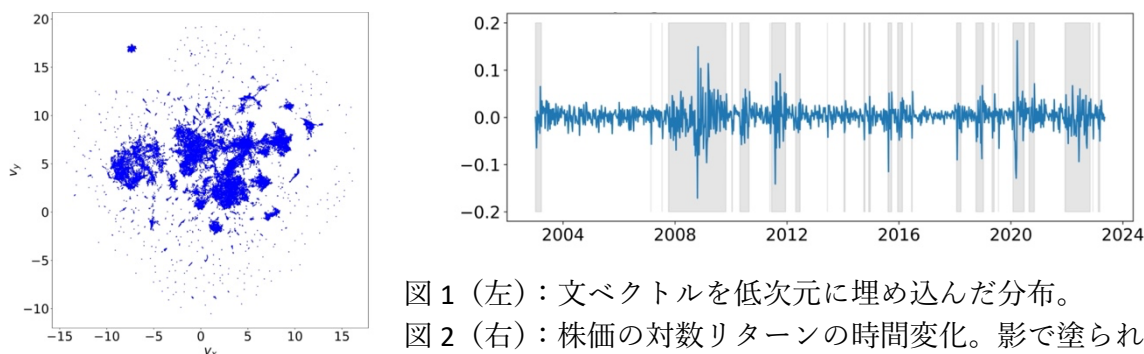


図 1 (左)：文ベクトルを低次元に埋め込んだ分布。

図 2 (右)：株価の対数リターンの時間変化。影で塗られた部分がモデルで不況と判定された領域。

- [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

謝辞

インターンシップ応募時から事務手続きをしてくださった芹澤様をはじめ、みずほ第一フィナンシャルテクノロジー株式会社の皆様に心からお礼申し上げます。特にリスクマネジメント技術開発部の金子様、金融・公共法人アドバイザー部の陳様、佐々木様には大変お世話になりました。誠にありがとうございました。金融に対する理解や解像度が高まったとともに、数々の新しい技術や知見に触れ、それを実務に応用して価値を創出していく体験をすることができて非常に楽しかったです。最後に長期インターンシップへの参加を後押しいただいた指導教官の沙川先生、副指導教員の有田先生、そしてMERITプログラムに感謝いたします。