

感情分析モデルを用いた株式併合による 上場廃止企業の有価証券報告書特徴分析

Feature Analysis using an Sentiment Analysis Model on Securities Reports of Companies Delisted Due to Stock Consolidation

2241125 宗像 譲太郎

Jotaro MUNAKATA

指導教員 秋葉 知昭

This study verified whether corporate characteristics could be captured by applying sentiment analysis to text data from securities reports and performing clustering analysis using the resulting sentiment scores as features. The target companies were 107 firms that were delisted due to stock consolidation between April 2022 and March 2025, and for which securities report submissions were confirmed between 2019 and 2024. Two sentiment analysis models with different numbers of polarities were adopted. After determining the appropriate number of clusters using the elbow method, k-means clustering was applied. The results indicated that, under specific conditions, companies in the transportation-related and metal-related industries tended to be classified into relatively small clusters. This suggests sentiment analysis may capture industry-related characteristics to some extent. However, effects stemming from industrial imbalances within the sample companies and the characteristics of the sentiment analysis models were also observed. Further verification of this hypothesis requires more careful consideration of the analytical conditions.

1. 緒言

近年、株式併合による企業は、増加傾向にあると確認できる。また、有価証券報告書に対して自然言語処理技術を用いて特徴分析を図る研究が見られ、テキストデータが企業分析において有用である可能性が示されている[1]。

先行研究[2]では、有価証券報告書内の数値データから株式併合による上場廃止企業の財務的特徴が明らかにされた。一方で、テキストデータを対象とした分析や数値とテキストとの横断的な分析に関してはさらなる研究の余地があると考えた。

本研究では、有価証券報告書内のテキストデータを一定程度定量化するため感情分析によって確率に変換し、クラスタリングを用いて企業ごとの特徴の差異を捉えることが可能か検証する。

2. 分析方法

2.1 データ収集

2022年4月から2025年3月の間に株式併合による上場廃止を行っている点、EDINET[3]上にて2019年から2024年提出分の有価証券報告書ファイルいずれかが確認できる点、銀行法や米国会計基準を採用していない点、以上を満たす107企業を対象とした。

2.2 感情分析

本研究ではテキストデータを一定程度定量化することを目的とし、既存のモデルを用いて感情分析を行った。利用モデルは、HuggingFace上で公開

されている次の2モデルである。

bardsai/Finance-Sentiment-ja-base [4]

c299m/japanese_stock_sentiment [5]

いずれも有価証券報告書内のテキストデータを入力し、bardsaiはpositive,negative,neutralの3つ、c299mはbullish,bearishの2つの極性に対応するlogitが出力される。logitに対し、Python上でsoftmax関数による正規化処理を行うことで確率

分布とした。n次元ベクトル $x = (x_1 x_2 \cdots x_n)^T$ に

対し、softmax関数は次式で求められる。

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}} \quad (i = 1, 2, \dots, n) \quad (1)$$

また、比較可能性を考慮し極性においてbullishをpositive、bearishをnegativeとして整理した。以降の分析では、各モデルのpositive,negativeの確率を使用し、neutralは対象外とした。

2.3 クラスタリング

与えられた確率を特徴量としてk-meansによるクラスタリングを行う。クラスタ数kはエルボー法を用い3と決定した。また、モデルと極性の組み合わせから4つの条件に分割し、各6年分のファイルを1年ごと、計24回のクラスタリングを行った。

3. 結果及び考察

2019年分の有価証券報告書をc299mモデルで分析した結果のうち、positiveの確率に対してクラ

スタリングを行った結果を図1に示す。

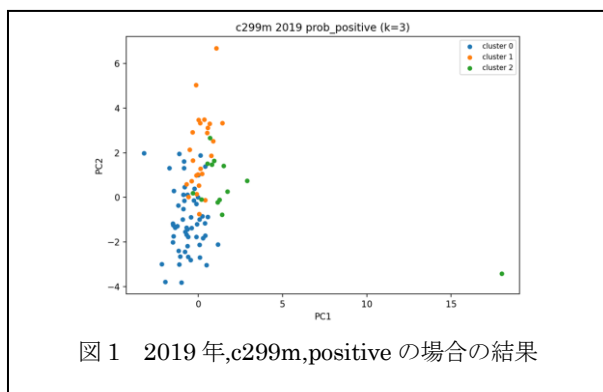


図1 2019年,c299m,positiveの場合の結果

クラスタリングにおいて、小さなクラスタに分類されやすかった企業は、多数派の企業と比較し、何らかの差異が評価されていると考え、着目した。日本取引所グループが運営する東証上場会社情報サービス[6]で確認できる企業ごとの業種に着目し、6年のうち半数以上で最も大きいクラスタ以外に属していた企業について考察した。

各条件において考察対象とした企業で見られた業種を表1にまとめる。なお、2値分類モデルであるc299mは特微量とした2値の確率の和が企業ごとに必ず1となるため、データ空間内の相対的な位置関係がpositive,negative間で同一であり、結果は同一となった。

表1 着目した企業で見られた業種

	positive	negative
bardsai	海運、倉庫・運輸 関連、情報・通信	情報・通信、鉄鋼、倉庫・運輸 関連、卸売、その他製品
c299m	サービス、情報・通信、卸売、金属製品、小売、化学、 不動産、ガラス・土石製品、ゴム製品、機械、食料品、 鉄鋼、電子機器、非鉄金属、陸運、その他製品	

クラスタリング結果をまとめると、多くの業種が見られる。ただしサービス業や情報・通信業など、業種ごとの企業数が多く、相対的には見られやすいとは言えない業種も見られた。それらを除くと、類似する業種をまとめた際に、運輸関連業や金属関連業において小さなクラスタとして分類される傾向にある可能性があった。仮にこれが正しいとすると、業種の系統によってビジネスモデルや業務内容の類似性がある事から、それらを説明するための言い回しや専門用語が評価されているためだと考えられる。

また、企業単位においては複数条件で共通して小さなクラスタとして分類されたものが2社見られたが、関連を見出すことはできなかった。原因として、当該企業数が少数であることに加え、その中でも業種の類似性は見られなかったことが挙げら

れる。

最後に、本研究における複数モデル間の差異について、クラスタリング結果から極性が2つのモデルは3つのモデルと比較しクラスタごとのまとまりがよく、適切に行われたように思われる。しかし、モデル自体の極性の数が違う中、モデルを除き同様の条件下での分析を行った点、モデルc299mにおいて、bullish,bearishという出力結果を解釈に基づいてpositive,negativeと同様に評価している点、感情分析は極性の数が少ないほど偏った結果になりやすい点など、十分に対照的であったとは言えない。そのためモデルごとの有効性の比較について示すことができたとは言えず、極性について同様のモデルを選定するなど、さらなる検証が必要である。

4. 結 言

本研究では株式併合による上場廃止を行った企業に対し、感情分析の結果を用いクラスタリングを行うことで特徴分析を試みた。中でも特定の類似業種において、非財務的特徴を捉える可能性が示唆された。

一方で、対象企業内で業種分布の偏りがあった点や、利用モデルの仕様の違いによって十分に傾向の有無を読み取ることができなかった。そのため今後の課題として、対象企業やモデル選択において、偏りが出ることのないよう再検討することが挙げられる。またその際、クラスタリングにおいてもクラスタ数を再検討し、業種によって分類が行われていることをより正確に明らかにすることが望まれる。

文 献

- [1] 近藤隆史,石光裕：有価証券報告書へのテキストマイニングの適用に関する文献レビュー,京都マネジメント・レビュー, Vol42,pp.93-117 (2023)
- [2] 織田恵子：有価証券報告書に基づく株式上場廃止企業の特徴分析,千葉工業大学社会システム科学部経営情報科学科卒業論文 (2025)
- [3] 金融庁：EDINET, <https://disclosure2.edinet-fsa.go.jp/week0010.aspx> (2025 時点)
- [4] bardesai：bardesai/Finance-Sentiment-ja-base, <https://HuggingFace.co/bardsai/finance-sentiment-ja-base> (2025 時点)
- [5] c299m：c299m/japanese_stock_sentiment, https://HuggingFace.co/c299m/japanese_stock_sentiment (2025 時点)
- [6] 日本取引所グループ 東京証券取引所：東証上場会社情報サービス, <https://www2.jpx.co.jp/tseHpFront/JJK010030Action.do> (2025 時点)