

2016 年度決算を対象とした社是と企業業績の関係 (第一報) -東証二部上場製造業企業を対象とした解析- Relationships between mission statements and profitability in fiscal year 2016 (Preliminary Result) -Analysis for manufacturing companies listed on Tokyo Stock Exchange 2nd Section-

北島 良三^{*1}
Ryozo Kitajima

上村 龍太郎^{*2}
Ryotaro Kamimura

酒井 浩之^{*1}
Hiroyuki Sakai

中川 慧^{*3}
Kei Nakagawa

^{*1}成蹊大学
Seikei University

^{*2}東海大学 情報教育センター
IT Education Center, Tokai University

^{*3}野村アセットマネジメント株式会社
Nomura Asset Management Co., Ltd.

In this paper, we try to analyze relationships between mission statements and profitability. The mission statements are very important messages for companies, because they include founders' spirits, the business policies and so on. Therefore, we consider that mission statements affect profitability. Mission statements and profitability (The Return On Asset (ROA) was used) were gathered from annual securities reports because descriptions are accurate and reports are easy to obtain. As mission statements were written in natural language and data to be analyzed becomes complicated, a neural computational method called 'potential learning' which can interpret internal representations was used. As a result, we found that a generalization performance of the model was 0.6125 (accuracy) and mission statements composed of multiple messages may affect ROA.

1. はじめに

本研究は社是と企業パフォーマンスの関係について解析を試みたものである。社是は、経営上重要視する事項、創業の想い、経営方針、などが記された、企業が掲げる企業行動の最高目標と捉えることができるメッセージである(経営理念や企業理念など様々に称される[鳥羽 84])。企業は日々の業務活動により収益をあげているが、決算の際に1年分の総括が行われ、年間パフォーマンスが発表される。この際、パフォーマンスの高・低といった結果が出てくるが、この結果の根底に社是の作用が存在しているのではないかと我々は考えている。よって、社是と企業パフォーマンスの間に関係性は存在しているのか? 存在しているならばどのような関係なのか? といった事柄が我々の関心事項である。企業のパフォーマンスについて考えると、実際には、例えば市場状況や資本金などの各社を取り巻く様々な事柄が作用するため、社是が企業パフォーマンスに大きく影響を与えることまでは考えられない。しかし、企業の業務姿勢や指針が社是に表されているため企業がこれを無視して業務活動を行うことは考えられず、社是が業績に影響を与えている可能性を無視することもできないと考える。

社是と企業パフォーマンスの関係はこれまでも研究されているが、明確に企業パフォーマンスに影響を与えているという結果は得られておらず、高尾は先行研究レビューを通して「経営理念が企業の収益に寄与できる可能性は十分にあると考えられる」と述べるに留まっている[高尾 10]。また、小田らはテキストマイニングを用いた研究より、経営理念は「従業員に対する経営理念」、「社会貢献を訴える経営理念」、「世界視点を持つ経営理念」の3つに内容によって分類ができること、そしてこの3つの分類の中で従業員に対する経営理念を持つ企業が最も業績が高く、社会貢献を訴える経営理念を持つ企業の業績が最も低いことを報告している[小田 10]。北島らもこれまでテキストマイニングを用いて研究を実施しており、2013年10月時点で東京証券取引所市場第二部(東証二部)上場の製造業

企業を対象とした研究[北島 16a]から「世の中に対して訴求する社是を持つ企業」と「顧客に対して訴求する社是を持つ企業」が収益性がよいことを報告している。

しかし北島らの研究では社是を人手による企業 Web サイト巡回により収集しているため、データ収集に時間的コストを要している。社是は頻繁に変更されるものではないものの、必要に応じて変更されるものである。よって2013年時点の社是を解析し関係性を得られたからと言って、その関係性が未来永劫続くとは考えにくく、社是にはその時代に応じた傾向が存在していると考えられる。しかしこれを追跡し研究するためには長期間の調査が必要となるが、データ収集に時間的コストを要していることは継続した研究は困難である。

そこで本研究では社是収集源として有価証券報告書を用いることとした。有価証券報告書を使用する利点はその入手性と情報の正確性であり、これにより社是収集の時間的コスト圧縮が期待される。本研究では有価証券報告書を用いて2016年度決算時点の社是と収益性の関係について研究を行った。

2. 研究手法

2.1 研究の流れ

本研究は以下の流れで実施した。

1. 社是と企業パフォーマンスを収集する。
2. 収集された社是に対して形態素解析を実施し、解析用データを作成する。
3. 解析用データを入力とし、知的情報処理手法を用いて、企業パフォーマンスが高い企業と低い企業に分類する分類器を作成する。
4. 分類器の汎化能力を調べ、分類結果を解釈する。

以下、上記手順に沿って詳細を述べる。

2.2 社是と企業パフォーマンスの収集

本研究では社是と企業パフォーマンスデータの収集源として有価証券報告書(有報)を用いることとした。有価証券報告

連絡先: 北島 良三, 成蹊大学理工学部情報科学科,
〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1,
r-kitajima@st.seikei.ac.jp

書は金融商品取引法で定められている金融庁への提出が義務づけられている書類である。データ収集源として有報を採用したのは、多くの企業情報を網羅できること（提出義務のある企業は複数定められているが、その一つに、金融商品取引所に上場されている有価証券の発行者が定められている）、正確な情報であること（虚偽記載は違法であり、また、重大な誤謬が発見された場合は訂正報告書の提出が必要となる）、入手が容易であること（金融庁の「金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム」（Electronic Disclosure for Investors' NETwork (EDINET) よりダウンロード可能である）、などが理由としてあげられる。

社はおよび企業パフォーマンスデータは有報以外からでも、例えば企業 Web サイトを閲覧することにより入手することも可能である。しかし、総務省統計局発行の「日本の統計 2018」によると、4,098,284 社の企業が日本には存在しており膨大な企業数である。また企業を証券取引所上場企業に絞ったとしても、東京証券取引所（日本最大の金融商品取引所である）には 2019 年 1 月 22 日現在で 3,653 社の企業が上場しており、こちらもかなりの数である。また、Web ページ上のどの部分に記載があるのかも各企業で異なっており、検索の必要がある。このため、1 企業ずつ Web サイトを巡回しデータを収集することは時間面で高コストである。さらに、企業 Web サイトは情報発信が未来方向に一方通行であり、例えばデータ収集日の前日に社が更新されている場合に収集側がそれを把握することは困難であるなど、問題点が存在している。

有価証券報告書は金融庁の「金融商品取引法に基づく有価証券報告書等の開示書類に関する電子開示システム」（Electronic Disclosure for Investors' NETwork (EDINET)）により、PDF ファイル形式と XBRL ファイル形式の 2 形式で公開されている。PDF 形式の有報は報告書としてそのまま読めるものであるが、XBRL 形式の有報は eXtensible Business Reporting Language (XBRL) というコンピュータ言語で記述されているため、そのままの状態では可読性が低いものである。しかしコンピュータ言語で記述されているためコンピュータで処理しやすく、本研究では XBRL ファイル形式の有報から XBRL タグを手掛かりに、社はと企業パフォーマンスデータを得た。

有価証券報告書には「経営方針、経営環境及び対処すべき課題等」という項目があり、ここに記載されている文を解析対象候補として入手した。一方企業パフォーマンスであるが、これには様々な指標が存在している。本研究では収益性に着目し、Return On Asset (ROA) を解析対象とした。これの算出には総資産と経常利益を用い、また、その企業単体のパフォーマンスを解析対象としたい目的より、個別決算の値を取得した。

本研究では解析対象企業として東京証券取引所市場第二部に上場している製造業の企業を対象に解析を実施することとし、239 社の有報を収集し「経営方針、経営環境及び対処すべき課題等」を抽出した。そして抽出した文より目視にて社に該当する文を選別したところ、197 社の社是を得ることができた。

2.3 解析データの作成

上記のように収集される社はであるが、これらはテキストデータであるためそのままでは解析を実施することができない。そこで社に対して形態素解析を実施し、その結果を基に解析用データを作成する。形態素解析には日本語形態素解析システムである JUMAN を用い、単語の「カテゴリ」を得た。カテゴリは JUMAN の辞書に用意されている情報である。これは単語がどのようなものであるのか詳細を表したもので、例えば「飛行機」という単語は「カテゴリ：人工物-乗り物」と

表 1: 解析用データの変数一覧

番号	カテゴリ名
1	人
2	組織・団体
3	人工物-その他
4	自然物
5	場所-施設
6	場所-その他
7	抽象物
8	数量
9	時間
10	未定義語

いう具合である。

JUMAN には 22 種類のカテゴリが用意されているが、本研究では「未定義語」というカテゴリを新設し、合計 23 種類のカテゴリを社はから得た。未定義語のカテゴリは辞書に未登録の単語を分類するカテゴリである。このカテゴリには未登録単語が蓄積されるため、必要に応じてこのカテゴリの単語を辞書に追加することも可能である。

このように得られるカテゴリであるが、これらは各カテゴリの各社はでの出現頻度 (Term Frequency, TF) を、そのカテゴリの全社はでの出現状態 (文書頻度逆数, (Inverse Document Frequency, IDF)) により重み付けした値 (TF-IDF) として記録し解析用データとした。一方、ROA データであるがこれは ROA をそのまま解析対象としたのではなく、197 社の ROA うち第一四分位数 (2.2813) 以下の企業にターゲットフラグとして「0」を、第三四分位数 (5.8129) 以上である企業に「1」を、それぞれ割り当てて被説明変数とした。各フラグの内訳は、1 が 50 社、0 が 50 社となり合計 100 社が解析対象となった。

以上の手順により入力変数が 23 から成る解析用データが作成されたが、データ中のほとんどの値が 0 である変数が散見されたため、TF-IDF 値が 0 であるものが 100 社中 90% 以上であるものを除去し、最終的に解析用データは 10 変数から成るデータとなった（最終的な解析用データの変数を表 1 に示す）。この作成された解析用データが次節で述べる手法により解析される。

2.4 解析手法

本研究で解析するデータは、社はと企業パフォーマンスから成るデータであり、複雑なデータである。そこで、複雑なデータの解析に定評のあるニューラルネットワークを解析に用いることとした。しかしニューラルネットワークはブラックボックスと称されるほどモデル解釈が困難である。そのためモデル解釈が可能なニューラルネットワークである「潜在学習」[Kamimura 17] を解析手法に採用した。潜在学習はこれまで北島らによって解釈を要する解析に採用され（例えば、スーパーマーケットデータの解析 [北島 16b]）、高いモデル性能、高い解釈性が確認されている。

この手法は図 1 に示されているように自己組織化マップ (Self-Organizing Maps, SOM) と多層パーセプトロン (MultiLayer Perceptron, MLP) が基となっている二段階の学習から構成されている。図 1 中の ① は知識獲得段階と呼ばれる段階で、入力ニューロンの潜在性を算出し、また、SOM に基づいて知識の獲得 (学習) を行う。潜在性とは、ニューロンの多様な状況に対応できる能力と定義されるもので、「潜在性の高いニュー

ロン (多様な状況に対応できるニューロン) は学習で重要な役割を果たすニューロンである」と捉えている。潜在学習ではこの潜在性を学習完了後に解釈することで、どの入力学習に活用されたのか解釈可能である。潜在性算出にあたり入力ニューロンに番号 k ($k = 1, 2, \dots, K$) を割り当てて表すと、 k 番目の入力ニューロンの潜在性 ϕ_k^r は、

$$\phi_k^r = \left(\frac{V_k}{\max_{k=1, \dots, K} V_k} \right)^r,$$

により定義される ($0 \leq \phi_k^r \leq 1$)。ここで V_k は k 番目の入力ニューロンの分散 (これは j 番目の出力ニューロン ($j = 1, 2, \dots, J$) から k 番目の入力ニューロンへ接続されている重み $w_{j,k}$ から求められる [北島 16b]), r は算出される潜在性を制御するパラメータである。よってパラメータ r が大きい値を持つ場合は大きな分散を持つ入力ニューロンが大きな潜在性を持つようになる。潜在学習はこの潜在性を求めた後、知識獲得段階で学習を行う。知識獲得段階では SOM により学習が行われるが、通常の SOM と異なり、入力ニューロン (k 番目の入力ニューロンを x_k で示す) と j 番目の出力ニューロンの距離 d_j を求める際に潜在性が、

$$d_j = \sqrt{\sum_{k=1}^K \phi_k^r (x_k - w_{j,k})^2},$$

として利用される。すなわち潜在学習の知識獲得段階では入力ニューロンの潜在性で重みを付けられた距離を学習に使用する。その他の学習ロジックは通常の自己組織化マップと同様である。

知識獲得段階での学習が終了すると、続いて予測段階 (図 1 中の ② の処理) での学習が行われる。予測段階は MLP にて、そして、入力層-中間層間の重みの初期値に知識獲得段階で得られた、重みと潜在性より算出した値 (重み $\times$ 潜在性) がセットされ学習が行われる。通常 MLP による学習結果は初期重みに左右されるが、潜在学習ではこの初期重み設定により、獲得された知識に基づいた学習が期待できる。以上が潜在学習の概要である。なお、潜在学習にはいくつかのバリエーションがあり、ネットワーク構造や潜在性の算出方法などに違いがある。

次章ではこの潜在学習を用いて、前節で作成したデータを解析していくが、本研究で使った潜在学習は入力ニューロンの数を 10 個、知識獲得段階における出力ニューロンの数および予測段階における中間ニューロンの数を 48 個、予測段階における出力ニューロンの数を 2 個とした。また予測段階において、中間ニューロンの伝達関数は双曲線正接関数、出力ニューロンの伝達関数はソフトマックス関数とした。中間ニューロンの数であるが、これは解析に使用したソフトウェアの標準個数^{*1}である。また、パラメータ r については 1 から 2 まで 0.1 刻みで変化させ、最も正解率の良いパラメータ値を探索した。さらに、解析時にデータは全サンプル (100 件) の、70% を学習用途、15% を過学習防止用途 (早期停止)、15% を汎化能力試験用途、に用いた。くわえてこの分割比率はそのまま、各用途に使用されるサンプルがランダムに異なるものを 10 パターン作成し、これらを用いて分類器を作成した。これはすなわち 10 個のモデルを作成したことになるが、次章で述べる分類結果などモデルパフォーマンスはこの 10 個の平均値を採用していることをここで述べておく。

*1 本研究で解析に使用したソフトウェア (MATLAB の SOM Toolbox) の設定 (mapsize:normal) による。

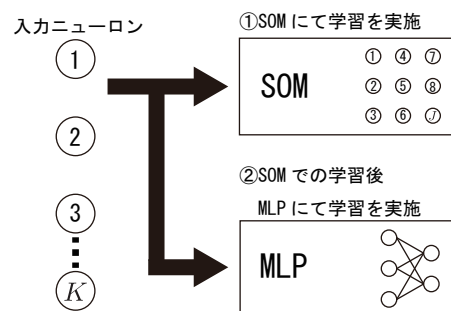


図 1: 潜在学習概要

3. 結果と考察

解析の結果、社是は正解率: 0.6125, 精度: 0.6058, 再現率: 0.6500 で分類ができた。この際パラメータ r の値は 1.9 であった。潜在学習により得られたこれら結果であるが、比較のために実施した、多層パーセプトロン、潜在性を使用しない潜在学習 (ネットワーク構成は潜在学習と同じであるが、学習時に潜在性を加味しないもの)、と比較すると高い結果であることを確認した (表 2 に結果を示す)。このことより、社是と企業パフォーマンスの関係調査には潜在学習が有効であることを確認した。しかし著しく高いパフォーマンスと言える結果でもないため、この結果は社是とパフォーマンスの関係は解析が難しい問題であることを示していると言える。モデルパフォーマンスが上がらなかった原因の一つとして学習データが少ないことが考えられる。解析対象データのうち、社是まで取得できたデータは 197 社あるため、無駄なく使用できればあと 97 社のデータを解析に使用することができる。もちろん 97 社のデータの内訳までも考慮しなければならないため、単純に 97 社分データを増加させられるわけではないが、閾値の調整により少々の増加は期待できる。

さて分類器が作成でき、性能も明らかになったため、続いては分類に重要な役割を果たした変数の解釈を実施する。一般的にニューラルネットワークはブラックボックスと称され、解析時に活用された変数を解釈することは困難である。しかし潜在学習は潜在性を確認することでこれを可能としている。図 2 は潜在性を示したものである。この図より、第 10 変数の潜在性が最大であることが確認できる。これは未定義語のカテゴリを意味する変数である。続いて潜在性がネットワークで活用された変数であるのかを重みの視点から確認する。図 3 にネットワークの入力層-中間層間の重みを示しているが、図中の四角形は重みを表しており、図形の大きさと重みの大きさを、緑色が正の値を、赤色が負の値を示している。比較のために実施した多層パーセプトロンでは、重みは各ニューロンでバラバラの重みを持っており、重要変数の判断は困難である。潜在性を使用しない潜在学習では多層パーセプトロンの場合よりも規則性が見えるものの、どの入力変数が重要であるかの判断は依然困難である。一方潜在学習の重みは 10 番目の入力ニューロンに重みが集中していることから潜在学習が 10 番目の変数を積極的に活用している状態が確認できた。

それでは未定義語のカテゴリのうちどのような単語が出現しているのだろうか。続いてはそれを調査した。その結果、形態素解析の誤りが大多数を占めていること、これを除くと数値の出現が多いことがわかった。ただしこれはアラビア数字ではなく、例えば ① などと通常とは異なる表記の数字であった。なお社是はビジネスメッセージであるが、未定義語のカテ

表 2: 解析結果

手法	正解率	精度	再現率
多層パーセプトロン	0.5250	0.5411	0.5125
潜在性を用いない潜在学習	0.5438	0.5532	0.5875
潜在学習 ($r = 10$)	0.6125	0.6058	0.6500

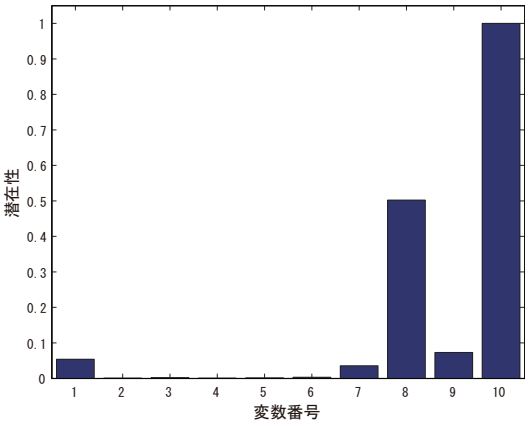


図 2: 入力変数の潜在性 ($r = 10$)

ゴリにはビジネス関連用語はあまり含まれていなかった。最重要変数であるものの、未定義語のカテゴリから特殊表記の数字が得られたことは意外であったため、2 番目に潜在性の高い第 8 変数を調べたが、この変数自体が数量であることから (漢数字の出現が多いことがわかった)、本研究で解析対象とした社是では数字が重要な要因であると判断した。

この結果より解釈を実施したが、社是中でのアラビア数字でない数字は社是を複数発信している場合に用いられるものであった。よって一つの社是で複数訴求メッセージがあることを示している。未定義語に分類された文字を使用して複数メッセージを発信している企業は 9 社あり、うちターゲットフラグ 1 に属する企業数は 4 社、フラグ 0 に属する企業数は 5 社であり、フラグ 0 に属している企業数の方が多いという結果であり、未定義語に限って解釈を実施すると、複数メッセージ発信企業の方が 1 社だけではあるがフラグ 0 に属している数が多いことがわかった。

本研究により未定義語に含まれる単語により社是がパフォーマンス別に分類可能であることが明らかになった。今後はビジネス関連用語を辞書に登録して未定義語のカテゴリを廃し、また、形態素解析の誤りにも対応し、より詳細な解析を実施していきたい。

4. まとめ

本研究は社是と企業パフォーマンスの関係について調査したものである。企業パフォーマンスには ROA を採用し、これの算出に必要な数値と、そして解析対象である社是は有価証券報告書より入手した。有価証券報告書をデータ源とすることで、信頼性の高いデータを時間的コストをかけずに入手することができた。また ROA は個別決算の数値を用い、これによりグループ企業の影響を排した関係性調査を実施できた。社是と ROA 間の関係は複雑であるため、内部表現解釈が可能なニューラルネットワークである潜在学習を解析手法に採用し

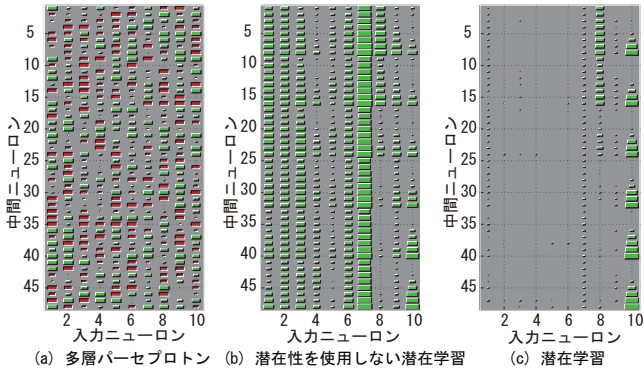


図 3: 入力層-中間層間の重み

た。解析の結果、正解率 0.6125 でパフォーマンス別に企業を分類できるモデルを作成でき、またこの分類には ① といった特殊な表記の数値が影響を及ぼしていることがわかった。比較のため潜在学習の他、多層パーセプトロン、潜在性を使用しない潜在学習、を用いて解析を実施したところ、潜在学習の分類性能が最も良い結果であり、また内部表現は潜在学習以外は明確に解釈できず、潜在学習が社是と ROA の解析に有効な手法であることが確認できた。しかし分類精度は改善の余地があり、今後は精度向上、またビジネス用語を形態素解析の辞書に整備するなどし、さらなる重要変数の解釈を実施していく。

参考文献

[鳥羽 84] 鳥羽欽一郎, 浅野俊光: 戦後日本の経営理念とその変化-経営理念調査を手がかりとして- (1984), 組織科学, Vol. 18, No. 2, pp. 37-51.

[小田 10] 小田恵美子, 三橋平: 経営理念と企業パフォーマンス: テキスト・マイニングを用いた実証研究 (2010), 経営哲学, Vol. 7, No. 2, pp. 22-37.

[高尾 10] 高尾義明: 経営理念は役に立つのか: 経営理念と成果についての先行研究からの考察 (2010), 経営哲学, Vol. 7, No. 2, pp. 38-51.

[北島 16a] 北島良三, 上村龍太郎, 内田理: 社是に含まれる業績に影響を与える単語に関する研究 -TF-IDF と共起語を切り口にした解析- (2016), 第 32 回ファジィ システム シンポジウム講演論文集, pp. 463-468.

[北島 16b] 北島良三, 遠藤啓太, 上村龍太郎: 入力ニューロンの潜在性に着目した小売店店舗の非継続来店顧客検知モデルの作成 (2016), オペレーションズ・リサーチ, Vol. 61, No. 2, pp. 88-96.

[Kamimura 17] Kamimura Ryotaro: Collective mutual information maximization to unify passive and positive approaches for improving interpretation and generalization (2017), Neural Networks, Vol. 90, pp. 56-71.