

令和 4 年 6 月 21 日現在

機関番号：32665

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K01843

研究課題名（和文）機械学習による倒産企業予測モデルの研究

研究課題名（英文）Business failures prediction by the machine learning

研究代表者

大槻 明（OTSUKI, Akira）

日本大学・経済学部・教授

研究者番号：30527833

交付決定額（研究期間全体）：（直接経費） 1,700,000円

研究成果の概要（和文）：機械学習のアプローチから企業の財務諸表データを分析することで企業倒産予測を行うモデルについて研究開発した。先行研究では、「倒産する、しない」の2値しか予測できなかったが、本提案モデルではこの2値に加え、倒産の原因になり得る財務状態についても明らかにすることができるようになるため、企業にとって有意な知見が得られるようになると思われる。

研究成果の学術的意義や社会的意義

従来の倒産企業予測モデルでは、判別分析やロジスティック回帰分析を用いて財務諸表データを分析するものが多く、つまりは過去の財務諸表データを分析して「倒産する、しない」の2値を予測することしかできなかった。しかし、「何が原因で倒産しそうなのか？」という理由が分からなければ企業は対策の打ちようが無い。ゆえに、本研究では、この2値に加え、倒産の原因になり得る財務状態についても明らかにすることができるモデルを研究開発した。倒産する、しない、の予測だけでなく、倒産の原因になり得る財務状態についても明らかにすることができるようになるため、企業にとって有用な知見が得られるようになると思われる。

研究成果の概要（英文）：This study did developed the model of company bankrupt by machine learning model using chronological order financial statement data. Previous studies are for predicting binary and therefore, they can predict only binary, it will go bankrupt or will not, by machine learning from the concerned financial statement data. However, if the business organization does not understand what is the cause of going bankrupt, it cannot find any measures to take in the first place. Accordingly, in this study research and development was made about the models to clarify the financial statement data which can be the cause of bankruptcy in addition to just the prediction that it will go bankrupt or will not.

研究分野：機械学習

キーワード：機械学習 企業倒産予測

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

田坂[1]によると、財務諸表を分析することで企業の信用性を明らかにする「信用分析」の研究が19世紀後半に行われたことに端を発し、世界大恐慌をきっかけに、1930年代から本格的にアメリカにおいて倒産企業を予測する研究が進められてきたといわれている。最近では、判別手法やロジスティック回帰を用いた倒産予測モデルが多く提案されている[2-9]。これらの分析手法は2値を予測するものであるため、つまりは、当該財務諸表データから機械学習を行い、倒産する、しない、の2値を予測するものが最近の先行研究の傾向であると考えられる。

2. 研究の目的

しかし、前項の先行研究を受けて「何が原因で倒産しそうなのか？」という理由が分からなければそもそも企業は対策の打ちようが無い。ゆえに、本研究では、従来の倒産企業予測モデルを発展させ、事前処理として機械学習により倒産企業の財務諸表データを分析し、倒産企業を倒産原因ごとに分類(クラスタリング)する。そして、非倒産企業の財務諸表データを同様に分析してこれら各倒産企業クラスタと比較することで、倒産する、しない、の予測だけでなく、倒産の原因になり得る財務状態についても明らかにできる倒産企業予測モデルについて明らかにする。

3. 研究の方法

本研究で明らかにする内容は、下記3-1~3-3に示すとおりである。

3-1 倒産企業の財務諸表データベース(DB)を構築する

3-2 機械学習を用いて倒産企業の財務諸表データを分析することで、倒産企業を倒産原因ごとに分類(クラスタリング)するシステムを構築する

3-3 非倒産企業の財務諸表データを分析し、前項で構築した各倒産企業クラスタと比較することで、倒産予測を行うシステムを構築する

まず、本研究では、倒産の定義を帝国データバンク[10]から引用する。そして、3-1では、上場廃止企業サイト[11]から倒産企業を調査し、調査した倒産企業の財務諸表データを株主プロサイト[12]などから取得する。また、3-2では、これらの財務諸表データを機械学習(クラスタリング)することにより、「倒産原因」による倒産企業の分類を行うモデルを研究する。

そして、3-3では、非倒産企業の財務諸表データを帝国データバンクの「企業財務データベース COSMOS1[13]」などから取得したうえで倒産企業と同様に分析し、3-2で構築した各倒産企業クラスタと比較することで倒産予測を行う。

4. 研究成果

4.1. 倒産・非倒産企業データベースの構築

倒産企業の財務諸表データを株主プロサイト[12]及び及び帝国データバンクの企業財務データベース COSMOS1[13]から取得し、また、非倒産企業の財務諸表データについては、日本取引所グループの業種別分類表の大分類10項目をベースに各大分類につき10社ずつをランダムに取得した。それぞれ取得した件数は次のとおりである。これらのデータを次節以降の企業倒産予測モデル及びアプリのためのデータベースとして構築した。

- ・ 倒産企業の財務諸表データ数 : 124件
- ・ 非倒産企業の財務諸表データ数 : 140件

なお、倒産企業データはそれぞれの倒産年度から直近5年分の財務諸表データを取得しており、非倒産企業データについては2020年度を最終年度として直近5年分のデータを取得した。

4.2. 企業倒産予測モデルの構築

本研究成果として構築したモデルは4.2.1節以下に示す通りである。

4.2.1. 倒産予測に用いる指標

先行研究で用いられている指標を参考に、本研究では倒産予測に用いる指標として次の 8 指標を選定した。

「流動比率」、「営業キャッシュフロー」、「投資キャッシュフロー」、「営業キャッシュフロー/流動負債」、「棚卸資産回転率(売上高/棚卸資産)」、「営業キャッシュフロー/売上高」、「自己資本利益率(当期純利益/自己資本)」、「自己資本/総負債」

4.2.2. 時系列財務データからの特徴量抽出

前節の 8 指標数の 5 年間の推移から特徴量を抽出するモデルについて述べる。時系列データを対象とした場合は、対数変化率(対数収益率)[14-15]を用いることが一般的であるが、自然対数を求める際にマイナス値が含まれていた場合は対数変化率を求めることができない。ゆえに、本研究では対数でなく次式 1 に示すように 5 年間の財務情報の変化率を求めることとした。式 1 では、前年度 $t-1$ と当年度 t の変化率を見る式 (Y_t) であり、 X は前節で評価指標に選定した 8 指標を個別に算出する。

$$Y_t = \frac{(X_t - X_{t-1})}{X_{t-1}} \quad (1)$$

次に、5 年間の推移 ($Y_t, Y_{t+1}, \dots, Y_{t+4}$) から代表値を抽出する手法について述べる。本研究では下記 ~ の 5 パターンを実データで比較検証した結果、前節の各指標のうち減少率を上手く取得できていたのは 負の変化率の「絶対最大値」であったため、この値を代表値 { $FV(Feature Value)$ } として採用することとした。

負の変化率の「算術平均」：「 $(Y_{t-4} + Y_{t-3} + Y_{t-2} + Y_{t-1} + Y_t) / 5$ 」というように 5 年分の変化率の平均値(1 年換算の平均)を代表値とする

負の変化率の「絶対最小値」：「 $\min(|Y_t|, |Y_{t-1}|, \dots, |Y_{t-4}|)$ 」というように 5 年分の負の変化率の絶対最小値を代表値とする

負の変化率の「絶対最大値」：「 $\max(|Y_t|, |Y_{t-1}|, \dots, |Y_{t-4}|)$ 」というように 5 年分の負の変化率の絶対最大値を代表値とする

負の変化率の「合計」：「 $\text{sum}(Y_t, Y_{t-1}, \dots, Y_{t-4})$ 」というように 5 年分の変化率の合計値を代表値とする

「4 年前と最終年の年換算の変化率」：「 $((Y_t - Y_{t-4}) / Y_{t-4}) / 4$ 」というように、4 年前と最終年の年換算の変化率を代表値とする。

具体的に、時系列財務データからの特徴量抽出を行う式は次式のように表される。

$$FV = -\max(|Y_t|, |Y_{t-1}|, \dots, |Y_{t-4}|), \quad \{Y_t, Y_{t-1}, \dots, Y_{t-4}\} < 0 \quad (2)$$

4.2.3. クラスタリング(機械学習)モデル

前節の FV を 8 指標ごとに算出して表 1 に示すような行列データを作成し機械学習(クラスタリング)を行う。なお表 1 の行項目はサンプル数を表し倒産企業と非倒産企業が設定される。

表 1. クラスタリングモデル

企業名	説明変数			
	流動比率	営業キャッシュフロー	...	自己資本/総負債
A	FV	FV	FV	FV
B	FV	FV	FV	FV
...	FV	FV	FV	FV

n	FV	FV	FV	FV
-----	------	------	------	------

表 1 に示したデータは隣接行列ではなく距離行列であるため全ての行サンプル（企業）間に距離データが発生してしまう。ゆえに隣接行列を対象としたクラスタリング手法は使えないことから距離行列を対象としたクラスタリング手法である大槻の手法[16]を用いることとした。大槻のクラスタリング手法は、積寄与率が 90%を超えるまでの主成分得点を用いてユークリッド距離を求めて主成分得点行列を作成する。そして、この主成分得点行列を対象に、式 3 に示すシルエット分析を行い、最もシルエット値が高い時のクラスタ数 K によりクラスタリングを行う。

$$S_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (3)$$

最後に、このクラスタリングにおける「学習モデルの作成」と「の学習モデルを用いた予測」の流れについて次の通り述べる。

モデル作成：学習データ（倒産・非倒産企業混合 184 社） > 正規化 > PCA（主成分分析） > シルエット分析でクラスタ数 K を決定 > K クラスタでクラスタリング > 学習モデル保存

予測：入力データ（非倒産企業 1 社） > 学習モデルで正規化 > 学習モデルで PCA > 学習モデルに当てはめて、どのクラスタに所属するかを予測。予測結果の出力は 1 企業ごとに行い、1 企業につき 1 つの予測結果が出力される。

4.3. 企業倒産予測アプリの構築及び実証結果

前節のモデルを Web アプリとして実装し、4.1 節のデータを分析（クラスタリング）した結果、4 クラスタが形成された。そのメンバリストを表 2 に示す。相関分析の閾値を参考に、「倒産企業割合 (%) >= 0.7」を倒産の可能性が高いクラスタ、「倒産企業割合 (%) <= 0.3」を倒産の可能性が低いクラスタと解釈すると、ある非倒産企業がもしクラスタ 1 に所属した場合は倒産の可能性は低いと解釈できるが、クラスタ 3, 4 に所属した場合は倒産のリスクが高いと解釈できる。

表 2. クラスタリング結果のメンバ数リスト

クラスタ番号	メンバ数	倒産企業数	非倒産企業数	倒産企業割合 (%)
1	100	26	74	0.26
2	36	21	15	0.58
3	41	32	9	0.78
4	7	5	2	0.71

ここで、非倒産企業 1, 2 をサンプルとして取り上げて「予測」を実証した結果を図 1, 2 に示す。示すように、図 1 の非倒産企業 1 はクラスタ 1 に所属して図 2 の非倒産企業 2 はクラスタ 3 に所属していた。

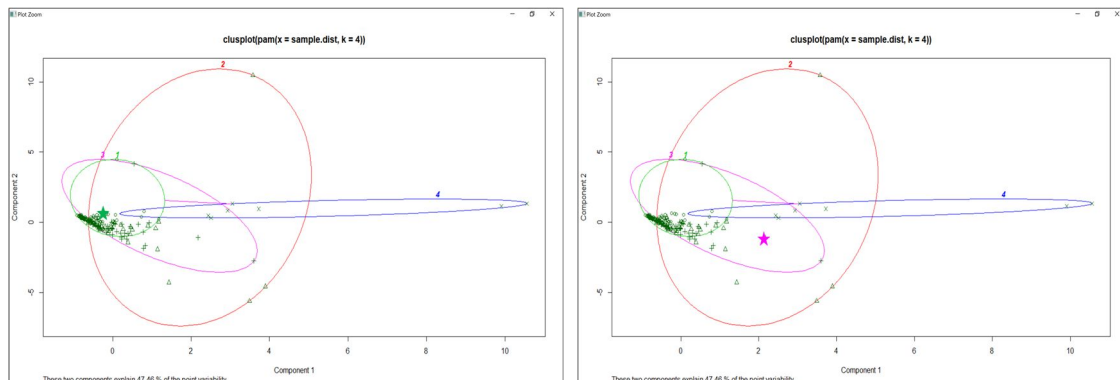


図1. クラスタリングの結果（非倒産企業1） 図2. クラスタリングの結果（非倒産企業2）

非倒産企業1が所属するクラスター1のメンバ特徴量を図3に示す。横軸のNot bankruptedは非倒産企業、Bankruptedは倒産企業を表す。そして縦軸はクラスター1における負の絶対値の最大変化率を、倒産、非倒産毎にグループ化した場合の平均をそれぞれ表す。クラスター1に所属する倒産企業の倒産要因の特徴としては「投資キャッシュフロー(investment_cash_flow_dmax)の減少」が目立つが、クラスター1は表2から倒産企業割合が最も小さいクラスターであるため、このクラスターに所属する非倒産企業1の倒産リスクは低いと解釈できる。

次に、非倒産企業2が所属するクラスター3のメンバ特徴量を図4に示す。クラスター3に所属する倒産企業の倒産要因は「自己資本利益率(return_on_equity_dmax)の減少」が著しく、「投資キャッシュフロー(investment_cash_flow_dmax)の減少」も目立つ。表2からもクラスター3の倒産企業割合は85%であるため、このクラスターに所属する非倒産企業2は、自己資本利益率及び投資キャッシュフローの減少により倒産するリスクがあると解釈できるため、これらの減少に対する対策を検討することが望まれる。

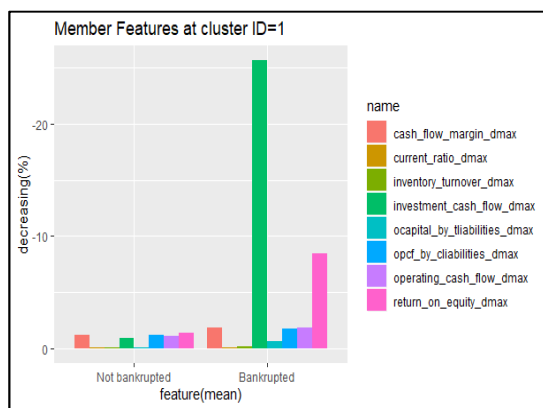


図3. クラスタ1のメンバ特徴量

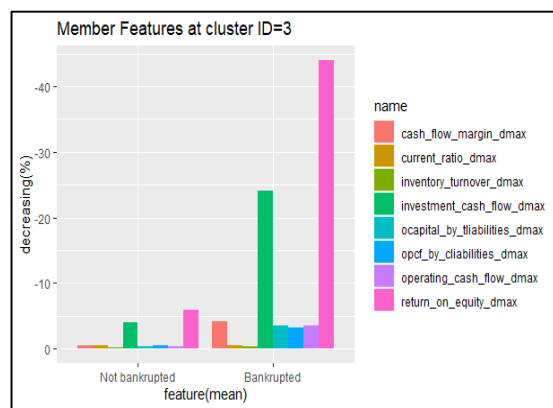


図4. クラスタ3のメンバ特徴量

以上に述べたように、本研究では時系列財務諸表データから特徴量を抽出し、この特徴量を機械学習に適用することで、倒産する、しない、の予測だけでなく、倒産の原因になり得る財務状態についても明らかにできるモデルについて研究開発した。本研究の成果は Web アプリとして公開可能であるため、例えば、ある企業が自社の財務諸表データを本 Web アプリに登録することで、自社の財務状態の現状分析や、倒産しないための示唆を得ることができるようになるため、本研究の成果は、広く社会に貢献できるものであると考えられる。

<参考文献等リスト>

[1]田坂成吾: 倒産企業に関する実証研究—Beaver, Altmanモデルを踏まえて—, Kwansei Gakuin University NII-Electronic Library Service, pp.71-99.
 [2]河野洋介ほか: 企業倒産予測の予測可能性に関する考察, 桃山学院大学学生論集, 復刊24号, 2009年.
 [3]大久保彩伽: 財務諸表から見える黒字倒産の原因—キャッシュフロー計算書を中心に—, 会津大学短期大学部2010年度卒業研究論文, 2010.
 [4]石川勝, 倪頌詩: キャッシュフロー情報にもとづく企業倒産の研究, 現代経営経済研究, 3(1), 35-58, 2012.
 [5]溝口竜二, 中嶋俊介: 企業の判別分析と倒産確率の推定, 南山大学情報理工学部・数理情報学部2012年度 卒業論文 要旨集, 2012.
 [6]蔣飛鴻: 企業の財務予測とキャッシュ・フロー情報, 明治大学経営学研究論集, 19巻, pp.175-192, 2003.
 [7]地代所混司, 藤田ハミド, 樽松理樹, 羽倉淳: 倒産リスク予測モデルによる日本企業のリスク予測の実験, 岩手県立大学ソフトウェア情報学部卒業論文, 2017.
 [8]増山祐一: 倒産企業分析からみた企業成長の要件, 彦根論叢オンラインジャーナル(滋賀大学経済学会), No.413, pp.16-31, 2017.
 [9]西郷鎮廣, 中野一豊: キャッシュ・フロー会計情報と企業価値に関する基礎的研究, 豊橋創造大学紀要(16), p.25-43, 2012.
 [10] <https://www.tdb.co.jp/tosant/teigi.html>
 [11] <http://delisting.info/index.html>
 [12] 株主プロサイト: <http://www.kabupro.jp/code/9963.htm>, 2019/10/17閲覧済み.
 [13] <https://www.tdb.co.jp/lineup/cosmos1/index.html>
 [14] 祝迫得夫: 株価指数の系列相関と規模別ポートフォリオの相互自己相関, <http://www.ier.hit-u.ac.jp/~iwaisako/research/JRWfinal.pdf> (2018/12/21閲覧済み).
 [15] 対数収益率について 株価リターンの計算 ~ : <http://capitalmarket.jp/post-431/> (2019/10/17閲覧済み).
 [16] 大槻明: 主成分距離行列シルエットクラスタリングによる潜在因子ラベル付けモデル, 情報処理学会論文誌 データベース Vol.13 No.4, pp.1-12, 2020.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 1件）

1. 著者名 Akira Otsuki, Shohei Narumi, Masayoshi Kawamura	4. 巻 22
2. 論文標題 A Study on Machine Learning Prediction Model for Company Bankruptcy using Features in Time Series Financial Data	5. 発行年 2022年
3. 雑誌名 Global Journal of Management and Business Research: Administration and Management	6. 最初と最後の頁 9-17
掲載論文のDOI（デジタルオブジェクト識別子） 10.34257/GJMBRAVOL22IS1PG9	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 大槻明, 鳴海祥平, 川村雅義
2. 発表標題 時系列財務諸表データの特徴量を用いた機械学習による企業倒産予測モデル
3. 学会等名 情報処理学会第83回全国大会講演論文集
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------