2A-03

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

大槻　明†　鳴海祥平‡　川村雅義§

日本大学†　NTT データシステム技術株式会社‡　MK 未来ソフトウェア§

1　はじめに

企業経営の実態を財務諸表データ観点から明らかにする手法として，判別手法[1-3]やロジスティック回帰[4]を用いた倒産予測モデルが提案されている．しかしこれらの分析手法は 2 値を予測するものであるため，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 値を予測することしかできない．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，倒産する，しない，の予測だけではなく，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．

2　企業倒産予測モデル

2.1　倒産企業の定義及び分析対象データ

本研究では倒産の定義を帝国データバンクから引用する[1]．そして，上場廃止企業サイト[2]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[3]及び及び帝国データバンクの企業財務データベース　COSMOS1[4]から取得した．また，非倒産企業の財務諸表データいついては，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．

- 倒産企業の財務諸表データ数　：84 件
- 非倒産企業の財務諸表データ数　：100 件

2.2　倒産予測に用いる指標

本提案モデルに必要なデータは前節で示した出典から取得するため，これらのデータから本倒産予測モデルで用いる倒産評価指標（説明変数）を抽出して次の通り設定した．

「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」

2.3　財務情報時系列推移データからの特徴量抽出

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で，前節で設定した財務情報 8 項目の 5 年間の推移から特徴量を抽出する．式 1 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各事項の財務情報を表す．

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

次に，5 年間の推移（Y_t，Y_{t-1}，…，Y_{t-4}）から代表値を抽出する手法について考える．今回は表 1 に示す流動資産のダミーデータを用いて，次に示す 5 パターン（①～⑤）を比較検証した．この結果変化率の差が大きい企業のケースの時には大きい値を，変化率の差が小さい企業のケースの時には小さい値を返していたのは，「③最大変化率」であったため，これを用いることとした．なお，倒産予測を行うためには，負の変化率のみに絞る（つまり正の変化率を除外する）必要があるが，この作業は手動でスクリーニングした．また，負の変化率が無い倒産評価指標（説明変数）については，プラスマイナスの影響が無い変数であると解釈し 0 に置き換えて分析した．

① 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

② 変化率の**「最小変化率」**：「$\min(Y_t, Y_{t-1}, \ldots, Y_{t-4})$」というように 5 年分の変化率の最小値を代表値とする

③ 変化率の**「最大変化率」**：「$\max(Y_t, Y_{t-1}, \ldots, Y_{t-4})$」というように 5 年分の変化率の最大値を代表値とする

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

⑤ **「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．

表 1．5 年間の推移（Y_t，Y_{t-1}，…，Y_{t-4}）から代表値を抽出する手法の比較（単位：千円）

【パターン 1】変化率の差が大きい企業のケースによる比較

		流動資産				
		最終年	1 年前	2 年前	3 年前	4 年前
各年度の変化率		4,977,284	3,921,740	6,208,416	40,249,410	981,290
①	9.77	0.27	−0.37	−0.85	40.02	
②	−0.85					
③	40.02					
④	39.07					
⑤	1.02					

注）【パターン 1，2】の 1 列目の番号はそれぞれ次の通り．①変化率_算術平均，②最小変化率，③最大変化率，④変化率_ＳＵＭ合計，⑤4 年前と最終年の年換算の変化率

【パターン 2】変化率の差が小さい企業のケースによる比較

		流動資産				
		最終年	1 年前	2 年前	3 年前	4 年前
各年度の変化率		22,226,115	36,741,231	42,170,776	34,044,383	42,536,683
①	−0.13	−0.42	−.013	0.24	−0.20	
②	−0.42					
③	0.24					
④	−0.51					
⑤	−0.13					

そして，財務情報の時系列推移の特徴量（*FV(Feature Value)*）として「③最大変化率」を抽出するためのモデルは次式に示す通りである．なお，Y_t を求める式は，上述の式 1 に示す通りである．

$$FV = \max(Y_t, Y_{t-1}, \ldots, Y_{t-4}) \qquad (2)$$

そして，前節で示した 8 つの倒産評価指標（説明変数）ごとに FV を求めて行列データを作成して機械学習を行う．

2.4　クラスタリング（機械学習）モデル

前節の *FV* を 8 つの倒産評価指標ごとに算出して表 2 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 2 の行項目はサンプル数を表し，倒産企業及び非倒産企業が設定される．

[1] https://www.tdb.co.jp/tosan/teigi.html
[2] http://delisting.info/index.html
[3] http://www.kabupro.jp/code/9963.htm
[4] https://www.tdb.co.jp/lineup/cosmos1/index.html

Study about Predict Model of Company Bankrupt By Machine Learning Model using Chronological Order Financial Statement Data
†Akira Otsuki・Nihon University　‡Shohei Narumi・NTT DATA SYSTEM TECHNOLOGIES　§Masayoshi Kawamura・MK Future Software

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

表 2. 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 2 に示したデータは，隣接行列ではなく距離行列であり，つまりは全てのサンプル（企業）間に距離データが発生してしまうため隣接行列を対象としたクラスタリング手法は使えない．ゆえに，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[5]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．

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

最後に，このクラスタリングを用いた「①モデル作成」と「②予測」の流れについて次の通り述べる．

① **モデル作成**：学習データ（倒産・非倒産企業混合）＞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

② **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

3 実証結果及び考察

3.1 クラスタリングの結果（①モデル作成）

2.4 節の①のモデルを用いてクラスタリングした結果，6 クラスタが形成された．そのメンバリストを表 3 に示す．次節では，2.4 節の②の予測結果について考察を行うが，もし入力データ（非倒産企業 1 社）がクラスタ 3 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 1，2，5 に所属した場合は倒産のリスクがあると解釈できる．

表 3. 図 1 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(%)
0	48	28	20	58.3
1	20	17	3	85.0
2	10	10	0	100.0
3	73	10	63	13.7
4	27	13	14	48.1
5	6	6	0	100.0

3.2 クラスタリングの結果（②予測）

非倒産企業 1，2 の財務諸表データを用いて「②予測」した結果を図 1，2 に示す．図 1 の非倒産企業 1 はクラスタ 3 に所属していて，図 2 の非倒産企業 2 はクラスタ 1 に所属していた．

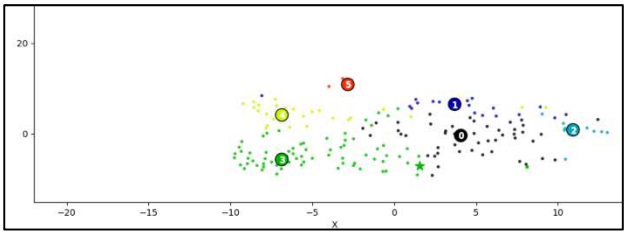

図 1. クラスタリングの結果（非倒産企業 1）

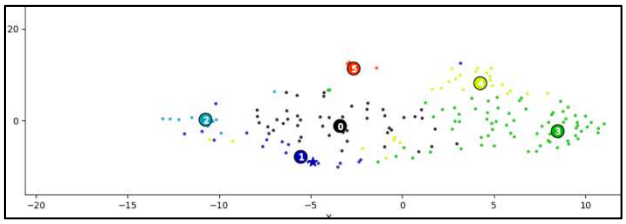

図 2. クラスタリングの結果（非倒産企業 2）

非倒産企業 1 が所属するクラスタ 3 のメンバ特徴量を図 3 に示す．横軸の 0 は非倒産企業，1 は倒産企業の特徴を表す．また，クラスタ 3 に所属する倒産企業の倒産要因は「自己資本利益率の減少」と「投資キャッシュフローの減少」が目立つが，表 3 からクラスタ 3 は倒産企業割合が最も小さいクラスタであるため，このクラスタに所属する非倒産企業 1 の倒産リスクは低いと考えられる．

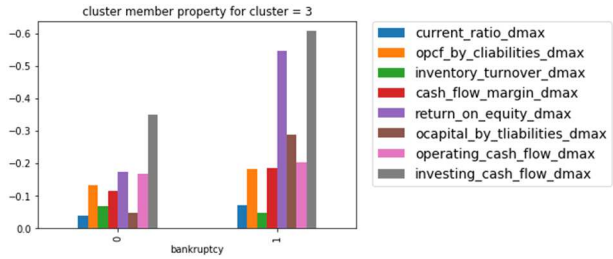

図 3. クラスタ 3 のメンバ特徴量

非倒産企業 2 が所属するクラスタ 1 のメンバ特徴量を図 4 に示す．クラスタ 1 に所属する倒産企業の倒産要因は「投資キャッシュフローの減少」が著しく，表 3 からもクラスタ 1 の倒産企業割合は 85%であるため，このクラスタに所属する非倒産企業 2 は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考えられる．

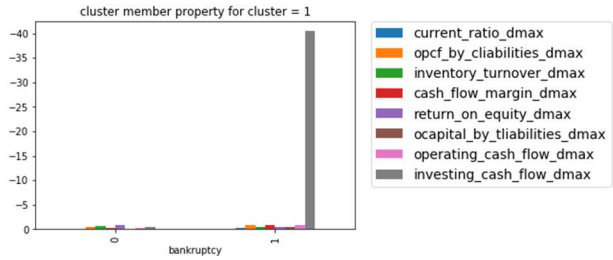

図 4. クラスタ 1 のメンバ特徴量

4 おわりに

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することで，倒産する，しない，の予測だけではなく，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．評価実験の結果，倒産原因ごとに倒産企業をクラスタ分類することができた．また，非倒産企業についても，将来の倒産リスクごとに適切に分類できていることから，財務諸表時系列データから特徴量を抽出しクラスタ分析する本モデルの手法は倒産予測と将来の倒産リスクの解明について有効であることが明らかとなった．

参考文献

[1] 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.
[2] 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.
[3] 蒋飛鴻 ：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.
[4] 地代所滉司, 藤田ハミド, 博松理樹, 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.
[5] 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 TOD，第 87 号, Vol. 13，No. 4，pp. 1-12，2020.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.