

機械学習を用いた倒産予測モデルの研究

片桐 英樹* 平井 裕久* 松丸 正延**

Bankruptcy prediction model using machine learning

Hideki KATAGIRI* Hirohisa HIRAI* Masanobu MATSUMRU**

1. 緒言

不確実性が增大する現代において、企業倒産予測に対するニーズと重要性は益々高まっている。倒産予測に関しては Altman[1]の Multiple Discriminant Analysis (MDA) 研究に代表されるように、これまで統計的手法が数多く用いられてきた。一方で、近年の人工知能 (AI) の発展に伴い、機械学習に注目が集まっている。

本研究では、機械学習アルゴリズムの一つで、多くの実問題に対する有用性が示されている勾配ブースティング[2]を用いた倒産予測モデルに焦点を当てる。具体的には、決定木を用いた勾配ブースティングの一種である eXtreme Gradient Boosting (XGBoost) [3]を用いたモデル構築を行う。日経 NEEDS-Financial QUEST2.0 から収集した財務データを用いた数値実験を行い、他の機械学習モデルの性能と比較することにより、提案モデルの有用性について検証する。

2. 先行研究

従来からよく知られる機械学習アルゴリズムの一つとして、Support Vector Machine(SVM) [4]がある。SVM は多くの判別問題に対して良い精度を示すことが知られており、倒産・非倒産の判別でも MDA と比べて良い性能をもつ[5]。

Wan and Ni [6]は特徴量選択に対してベイジアンハイパーパラメータ最適化を用いた XGBoost を提案し、米国企業の倒産判別モデルを題材した数多くの数値実験を行い、提案モデルの有用性を検証している。

3. 本研究の倒産予測モデルとその特徴

本研究では、日本企業の財務データを用いた XGBoost による倒産予測モデルを開発する。倒産予測において MDA よりも良いと報告されている SVM との比較を行い、XGBoost の優位性を確認する。

モデル開発の際には、変数選択を実施して、特徴集合において予測に寄与する部分集合のみを選択する。また交差検証を行い、訓練データと評価用の検証データを分割して性能を計測することで、特定のデータに拠らない汎化能力の高いモデルを得る方法を用いる。

財務指標の選択については、偏りのない多くの財務指標として 115 財務指標を用いる。期間については 1991 年から 2015 年までの 25 年間のデータを使用した。25 年間という長期間の財務指標データを用いることで、倒産企業のサンプル数が多くなり学習データとして十分な量を確保できる。一方で、経済環境が異なる期間を一緒にしてモデルを構築すると予知率が低くなる可能性がある。そこで、本研究では期間の異なる 5 種類のデータセットを作成して対応する。

一般に、倒産予測研究で扱うデータは、非倒産企業数に比べて倒産企業数が極端に少ないという不均衡データセット (Imbalanced Data Sets) と呼ばれる特徴をもつ。本研究ではこの不均衡データセットに対する解決策の一つとして提案されている SMOTO (Synthetic Minority Over-sampling Technic) 手法を用いている。

XGBoost はパフォーマンスの良い手法であるが、複数のパラメータ調整によるチューニングが必要であり、これがモデル性能に影響することも多い。ただし、本研究では最初の試みとして代表的なパラメータをデフォルト値で設定した。なお、本研究で用いるハイパーパラメータは、学習率、最大葉、最大深度などがある。

4 実証分析

4.1 データセットの説明

日経 NEEDS-Financial QUEST2.0 から財務データを収集した。データセットは倒産が起きた 1 年前のみの財務指標データセット p1, 2 年前の年度のみの財務指標データセット p2, 3 年前の年度のみの財務指標データセット p3, 2 年前のみの財務データセット p2 を追加した財務指標データセット pp1 と、pp1 に更に 3 年前のみの財務指標データセット p3 を追加した財務指標データセット pp2 の 5 種類を作成した。

倒産予測で考慮する財務指標 (モデルに対する入力データ) の種類は 115 とした。表 1, 表 2 に各業種とデータセットの企業 (非倒産企業と倒産企業) の数を示す。

表 1 各業種の企業数

単位:社

	不動産業	建設業	卸売業	サービス業
非倒産企業	135	170	335	449
倒産企業	30	22	12	16

*教授 経営工学科

Professor, Dept. of Industrial Engineering and Management

**客員教授 工学研究所

Visiting Professor, Research Institute for Engineering

表2 各業種のデータセットと企業数

単位:社

業種	企業	p	p1	p2	pp1	pp2
不動産業	非倒産企業	512	512	512	512	512
	倒産企業	30	30	30	60	90
建設業	非倒産企業	667	667	667	667	667
	倒産企業	22	22	22	44	66
卸売業	非倒産企業	1316	1316	1316	1316	1316
	倒産企業	12	12	12	24	36
サービス業	非倒産企業	1691	1691	1691	1691	1691
	倒産企業	16	16	16	32	48

4.2 実証分析結果と検討・考察

4.2.1 手法の違いによる実証分析結果と検討・考察

XGBoost と SVM の手法の違いによる結果から、XGBoost の偽陽性率(False Positive Rate)は SVM のそれよりも低い数値を示した。実際には倒産企業を非倒産と判別した企業数の割合が偽陽性率であり、倒産判別においては非常に重要な性能評価基準の一つである。この偽陽性率に関して、XGBoost は SVM よりも低い値を示しており、XGBoost が優れていることが示された。

4.2.2 多期間のデータセットを用いた実証分析結果と検討・考察

XGBoost を用いた場合、不動産業、建設業、卸売業およびサービス業の全てにおいて、倒産の1年前のみのデータセット p1 を用いた場合の予知精度が p2 と p3 を用いた場合よりも低い値を示した。

1年前のみのデータセット p1 を用いた場合が良い予知精度を示し、先行研究と同じ結果となった。すなわち、倒産年度に近いデータセットを用いたほうが、倒産年度から離れたデータセットを用いるよりも、倒産予知精度が高いことを意味している。

一方、単年度データセットを用いた場合と多期間の複数の年度にわたるデータセットを用いた場合を比較すると、多期間の複数年度にわたるデータセットを用いた場合のほうが高い倒産予知率を示した。

4.2.3 変数選択による予知精度の検討・考察

XGBoost の特徴の一つとして、回帰や判別の結果に大きく寄与する変数を特定できる機能がある。本研究において、倒産予知に寄与する変数を業種ごとにまとめた結果を表3に示す。

表3 倒産予知に重要な財務指標の結果

業種	1位	2位	3位
不動産業	売上高営業外費用比率	純資産増加率	総資本内部留保利益率
建設業	総資本内部留保利益率	有利子負債平均金利 インスタントカバレッジ	自己資本比率
卸売業	金利	有利子負債平均金利	総資本内部留保利益率
サービス業	有利子負債	総資産運転資本比率	総資本営業外収益

重要な財務指標の結果をしてみると、業種の特徴が表れている。

総資本内部留保率は不動産業で第3位、建設業で第1位、卸売業で第3位と3業種で重要な財務指標となっている。総資本内部留保率は、留保利益を総資本で割った値である。したがって留保利益が総資本に対して小さくなると倒産しやすくなる。換言すれば倒産を避けるには、留保利益の減少には注意をする必要があり、留保利益を厚くしておくことが安全性につながるという結果が導き出されてい

る。

また、有利子負債平均金利については、建設業と卸売業の第2位に位置する重要変数となっている。有利子負債平均金利は小さいほうが良く、有利子負債金利の増大には気をつける必要がある。

さらに、売上高営業外費用比率が不動産業では第一にランクされている。売上高営業外費用比率は支払利息を売上高で割った値であるので、売上高の減少と支払利息の増大に注意を払わなければならない。

以上より貸借対照表では内部留保額の減少、有利子負債額の増大、損益計算書では、売上高の減少、支払利息の増大に気を付ける必要性が示唆される。

また サービス業では第2位に総資産運転資本比率がランクされた。この比率は運転資本を総資産で割った値である。運転資本の構成要素のうち影響の大きいのは棚卸資産であり、棚卸資産が増大すると経営が悪化する。サービス業の特徴が表れており、妥当な結果が得られたと考える。

5. 結言

本研究では、機械学習を用いた日本企業の倒産予知モデルを提案し、企業の財務データを用いた数値実験により提案モデルの有用性を検証した。勾配ブースティング木の一種である XGBoost は偽陽性率において低い数値を示し、SVM よりも優れた結果となった。予知に用いるデータセットについては、単年度よりも多期間、古いデータよりも新しいデータを用いたほうが倒産予知の精度が高まることも示された。また、業種ごとで倒産予知に重要な財務指標についても明らかにした。

今後の研究として、企業間の関係など財務データ以外のデータを用いた倒産予知モデルを検討している。

参考文献

- [1] E. I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance*, 23(4), 589-609(1968)
- [2] J.H. Friedman, Greedy function approximation: A gradient boosting machine, *The Annals of Statistics*, 29(5), 1189-1232(2001)
- [3] T. Chen and C. Guestrin, XGBoost: A Scalable Tree boosting System, *KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining August(2016)*, 785-794, <https://doi.org/10.1145/2939672.2939785> Cornell University, (2016)
- [4] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning*, 20(3), 273-297(1995)
- [5] M. Matsumaru, T. Kawanaka, H. Katagiri and S. Kaneko, Bankruptcy prediction for Japanese corporations using support vector machine, artificial neural network, and multivariate discriminant analysis, *International Journal of Industrial Engineering and Operations Management (IJIEOM)*, 1(1), (2019).
- [6] Y. Wang and X.S. Ni, A XGBoost risk model via feature selection Bayesian hyper-parameter optimization, *International Journal of Database Management Systems (IJDBMS)*, 11(1), 1-17(2019)