

企業間取引ネットワークと業種分類情報による 信用リスク評価モデルの構築

Construction of a Credit Risk Evaluation Model Using Corporate Transaction Networks and Industry Classification Information

小林司[†] 成末義哲[†] 森川博之[†]
Tsukasa Kobayashi[†] Yoshiaki Naruse[†] Hiroyuki Morikawa[†]

[†] 東京大学大学院 工学系研究科

[†] Graduate School of Engineering, The University of Tokyo.

要旨

本研究では、金融機関の口座情報から構築した企業間取引ネットワークを活用し、非財務データに基づく新たな信用リスク評価モデルを示す。本手法は、ネットワーク特徴量のみを用いた基本モデルと、これに業種分類情報を追加した業種拡張モデルの2つを構築し、「正常先」と「非正常先」の2クラス分類を試みた。分析の結果、両モデルともに正常先の識別において十分な能力を示し、特に業種拡張モデルでは非正常先の識別能力も向上した。本研究の成果は、財務データを保有していない企業群への信用リスク評価の可能性を示唆しており、金融機関の新規顧客開拓や既存顧客のリスク管理への応用が期待される。

1. はじめに

近年、金融機関の口座情報を活用した研究が急速に発展している。この発展の背景には主に3つの要因が存在する。第一に、国際的なデータ保護規則の導入により、金融機関による顧客データの取り扱いに関する法的枠組みが整備され、顧客データの利活用が進んだことにある。第二に、分散していた顧客データの統合と一元管理が可能になり、顧客データのアクセシビリティが向上したこと、第三に、大規模なデータセットを効率的に処理できる技術の発達により、膨大なデータの分析が可能となったことである。

筆者らは、金融機関が口座情報を保有する企業数が、与信審査を通じて財務データを把握している企業数をはるかに上回るという事実に着目し、非財務データに基づく信用リスク評価モデルの開発を進めている。特に、口座の入出金データから構築可能な企業間取引ネットワークに注目している。先行研究では、このネットワークから得られる特徴量を信用リスク評価モデルに補完的情報として組み込む試みがなされているが、取引ネットワークから得られる特徴量のみを用いた企業の信用リスク評価の可能性については、これまで十分な検討がなされていない[1][2][3][4]。

本稿では、企業間取引ネットワークに業種分類情報を組み込んだ新たな信用リスク評価モデルの有効性を検証する。これは、ネットワーク特徴量のみによる画一的な企業評価に加えて、業種特有の取引パターンを新たな特徴量として活用することで、識別能力の向上が見込めるとの仮説に基づいている。このモデルは、債務者区分において業況が良好で財務内容に特段の問題がないと認められる「正常先」を識別することを目的としている。なお、債務者区分は金融庁のマニュアルに基づき、債務者の財務状況や収益力などから返済能力を判定するもので、本稿では正常先を「正常先」、それ以外の区分を「非正常先」として扱う。具体的には、株式会社山口フィナンシャルグループ (YMFG) の協力のもと、実際の事業性入出金データから抽出した特徴量を用いて機械学習モデルを構築し、「正常先」と「非正常先」の2クラス分類を試みた。分析の結果、「正常先」の識別については基本モデルが再現率 0.83 を示し実用的な識別能力を達成したものの、「非正常先」に関しては今後の改善の余地が残されている。さらに、業種分類情報を特徴量として追加することでモデル全体の予測性能の向上が確認されたものの、「正常先」の識別については再現率で 0.08 ポイント低下した。これらの知見は、財務データに依存しない信用リスク評価手法の確立に向けた第一歩であり、財務データを把握していない企業を対象とした顧客開拓や顧客リスク管理への適用可能性を示唆し、金融機関の経営効率を向上に寄与する可能性を有している。

2. 関連研究

企業の信用リスクを評価する上で企業間取引の重要性は認識されているが、金融機関の口座情報から取引ネットワークを構築した研究が見られるようになったのは近年のことである。既存研究においては、財務データに基づく信用リスク評価モデルに中心性・入出次数・取引額などのネットワーク特徴量を付加的情報として組み込むことにより予測性能の向上を図っている。Atan et al. [1]は、製造業向けモデルに取引先情報を組み込み、ジニ係数が 1% 向上したことを報告している。Vinciotti et al. [2]は入出次数などのネットワーク情報の補強、Martínez et al. [3]は中心性の一つである PageRank やトポロジカル指標の追加により予測性能の向上を実証している。Berloco et al. [4]においても、ネットワーク特徴量を付加したモデルの予測性能向上を確認しており、特に入金額、出金額および隣接企業の最大売上高が重要な特徴量であることを示している。しかしながら、本稿で取り組むようなネットワーク特徴量のみを用いた信用リスク評価モデルの研究はいまだ見られない。

業種分類を信用リスク評価に活用した既存研究として、Atan et al. [1]は取引先の業種数を特徴量として導入し、Platt and Platt [5]は財務比率を業界平均で正規化した業界相対モデルを提案している。また、Minakawa et al. [6]は業種分類情報をテキストデータとしてモデルに組み込む手法を示している。しかしながら、業種分類を直接的な特徴量として扱う研究は限定的であり、その有効性については十分な検証がなされていない。

3. 提案手法

本稿では、金融機関の口座情報から構築した企業間取引ネットワークを基に、各企業のネットワーク特徴量及び業種分類情報を用いて、与信審査における債務者区分の識別モデルを提案する。具体的には、ランダムフォレストアルゴリズムを採用し、実際に債務者区分が付与されている企業データを用いて学習を行い、「正常先」と「非正常先」の 2 クラス分類を試みる。ここで「非正常先」には要注意先や破綻先などが含まれる。一般的な信用リスク評価モデルが債務不履行の予測を目的とするのに対し、本稿では債務不履行サンプルの希少性と、顧客開拓への応用可能性を考慮し、この 2 クラス分類設定を採用した。

3.1. データ

本稿で使用する企業間取引情報は、株式会社山口フィナンシャルグループ (YMFG) が保有する口座情報から構築したデータに基づいている。具体的には、2021 年 11 月を基準時点とし、そこから遡って過去 12 ヶ月間の事業性入出金額上位 10 社のデータを抽出した。抽出したデータから、企業間の資金移動関係、および取引額の期間合計を整理し取引データセットを作成した。このデータセットに基づき、企業をノード、取引の方向をエッジ、取引額を重みとする有向グラフを構築した。図 1 にノードとエッジの向きの設定を示す。構築した企業間取引ネットワークは、81,680 ノード 253,789 エッジから構成される大規模な有向グラフである。

3.2. 特徴量

識別モデルに使用する特徴量は、取引活動を示す入出金データとネットワーク特性に基づく指標の 2 つに大別される。取引活動を示す指標として、各企業の口座データから抽出した事業性入金額、事業性入金件数、事業性出金額、事業性出金件数を採用した。また、企業間取引ネットワークにおける各ノードの構造的特性を表す指標として、入次数、出次数、媒介中心性、固有ベクトル中心性、PageRank を使用した。これらの特徴量の詳細な定義を表 1 に示す。

さらに、企業属性情報として日本標準産業分類業種コードの大分類を採用した。表 2 に示す業種分類をワンホットエンコーディングにより表現し、特徴量として使用した。

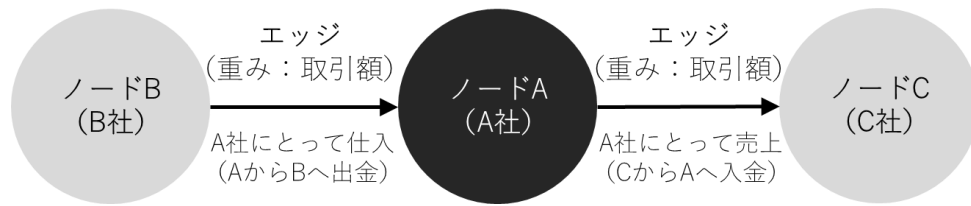


図1 ノードとエッジの向きの設定

表1 特徴量の定義

特徴量名	定義
事業性入金額	基準時点直前の決算年月から遡って過去12ヶ月間の事業性入金額の合計
事業性入金件数	基準時点直前の決算年月から遡って過去12ヶ月間の事業性入金取引件数の合計
事業性出金額	基準時点直前の決算年月から遡って過去12ヶ月間の事業性出金額の合計
事業性出金件数	基準時点直前の決算年月から遡って過去12ヶ月間の事業性出金取引件数の合計
入次数	隣接ノードから当該ノードが製品またはサービスを調達する取引関係の総数
出次数	当該ノードから隣接ノードへ製品またはサービスを提供する取引関係の総数
媒介中心性	ネットワーク内の任意の2ノード間の最短経路上に当該ノードが存在する割合。ネットワーク内の情報や資源の流れにおける当該ノードの重要性を示す指標。
固有ベクトル中心性	隣接するノードの重要性に基づいて当該ノードの重要性を再帰的に評価した指標。重要なノードとの接続性の高さを表す。
PageRank	固有ベクトル中心性を発展させた指標。隣接するノードの重要性をそのノードの出次数で正規化して算出。多数のリンクを持つノードの影響力の過大評価を防ぐ機能を持つ。

表2 日本標準産業分類業種コード

大分類	業種名	大分類	業種名
A	農業、林業	K	不動産業、物品賃貸業
B	漁業	L	学術研究、専門・技術サービス業
C	鉱業、採石業、砂利採取業	M	宿泊業、飲食サービス業
D	建設業	N	生活関連サービス業、娯楽業
E	製造業	O	教育、学習支援業
F	電気・ガス・熱供給・水道業	P	医療、福祉
G	情報通信業	Q	複合サービス事業
H	運輸業、郵便業	R	サービス業（他に分類されないもの）
I	卸売業・小売業	S	公務
J	金融業・保険業	T	分類不能の産業

3.3. ランダムフォレスト

本稿では、分類手法としてランダムフォレストを採用した。この選択は、結果の解釈が容易であること、特徴量の重要度を評価できること、およびアンサンブル学習による汎化性能の向上が期待できる

ことに基づいている。正常先と非正常先のクラス分類において、ランダムサンプリングを活用し複数の決定木を用いたアンサンブル学習を行うことで、過学習のリスクを低減しつつ全体として安定した分類精度の獲得を目指している。

モデルの構築に際しては、以下のハイパーパラメータを設定し最適化を行う：

1. `n_estimators`：決定木の数
2. `min_samples_split`：ノード分割の最小サンプル数
3. `min_samples_leaf`：リーフノードの最小サンプル数
4. `max_depth`：決定木の最大深さ

これらのパラメータの調整により、モデルの性能と汎化能力の最適化を図る。

4. 実験

提案手法の有効性を検証するため、ネットワーク特徴量のみを用いた基本モデルと、これに業種分類情報を追加した業種拡張モデルの比較実験を実施した。分類タスクとして正常先と非正常先の2クラス分類を設定し、企業間取引ネットワーク内で債務者区分が付与されている15,886ノードをデータセットとして用いた。

識別モデルの構築にあたり、まずネットワーク特徴量の正規化を行った。続いて、業種分類を表す変数に対してワンホットエンコーディングを適用し、各業種を二値変数として表現した。その後、データセットを訓練データとテストデータに8:2の比率で分割した。モデルの学習には、ランダムフォレストアルゴリズムを使用し、ハイパーパラメータの最適化にはランダムサーチによる交差検証を採用した。また、正常先と非正常先のクラス不均衡に対処するため、`class_weight`パラメータを用いてクラスの重み付けを行った。

モデルの評価指標としてAUC、および各クラスの適合率(Precision)、再現率(Recall)、F1スコアを使用した。さらに、特徴量の重要度評価には、予測結果に対する各特徴量の影響を定量的に評価するPermutation Importanceの手法を用いた。これにより、各特徴量の識別モデルへの寄与度を分析した。

5. 結果と考察

実験結果を表3に示す。基本モデルと比較して、業種拡張モデルはAUCが0.68から0.73へと向上した。特に非正常先の識別において、F1スコアが0.46から0.58へ改善が見られた。一方、正常先の識別能力については、再現率が0.83から0.75へと若干低下したものの、適合率が0.68から0.73へ向上し、F1スコアは同程度の水準を維持した。

特徴量の重要度分析の結果を図2に示す。両モデルともに事業性入金額と出金額、出金件数が上位を占めており、取引活動の規模が信用リスク評価において重要な指標となることが示唆された。業種拡張モデルでは、不動産業、医療・福祉、宿泊業・飲食サービス業といった特定の業種が比較的高い重要度を示した。

表3 実験結果：モデルの性能比較

基本モデル				業種拡張モデル			
AUC : 0.68				AUC : 0.73			
	Precision	Recall	F1-score		Precision	Recall	F1-score
正常先	0.68	0.83	0.75	正常先	0.73	0.75	0.74
非正常先	0.59	0.38	0.46	非正常先	0.59	0.57	0.58

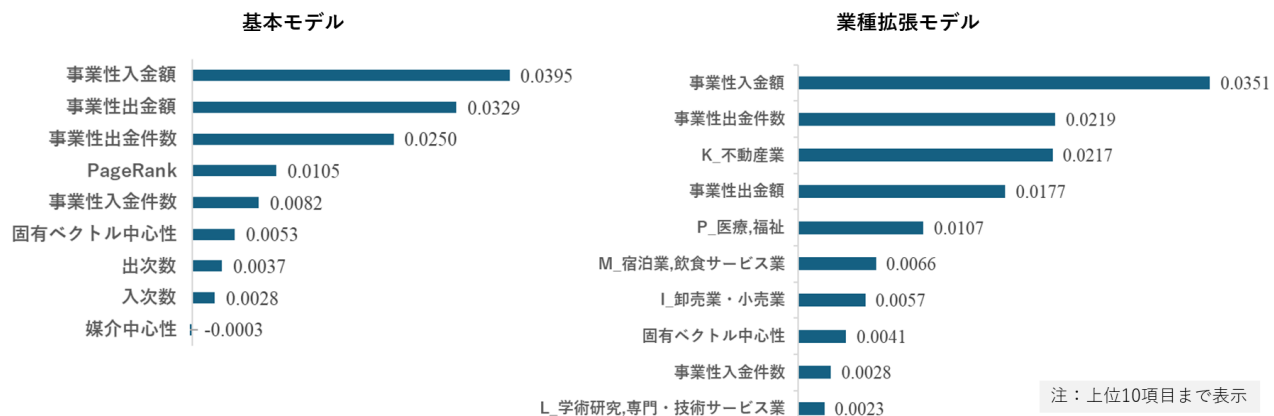


図2 実験結果：特徴量の重要度

表4 ハイパーパラメータの最適化結果

基本モデル		業種拡張モデル	
n_estimators	500	n_estimators	500
min_sample_split	2	min_sample_split	10
min_sample_leaf	1	min_sample_leaf	2
max_depth	None	max_depth	None
class_weight	非正常先：3, 正常先：1	class_weight	balanced

これらの業種は、資金の入出金パターンや取引関係の構造が業種固有の特徴を持つと考えられる。例えば、不動産業では大口の資金移動が存在する、医療・福祉分野では保険診療報酬による安定的な入金が見られるなど、業種特有の取引パターンが識別能力の向上に寄与している可能性がある。

ハイパーパラメータの最適化結果を表4に示す。業種拡張モデルで `min_samples_split` の値が大きくなっており、これは過学習を抑制する方向に調整されたことを示している。また、`class_weight` パラメータが `balanced` に設定されたことで、クラス不均衡への対応がより適切に行われたと考えられる。

一方で、非正常先の識別能力については、業種拡張モデルによる改善後も F1 スコアが 0.58 に留まっており、さらなる向上の余地が残されている。これは、非財務データのみでは企業の財務状況の悪化や経営上の問題を捉えることに一定の制約がある可能性を示唆している。しかし、この課題に対する改善策として、取引先の取引活動や業種などの取引パターンを特徴量として取り込むことで、非正常先の識別能力向上を目指す。

6. まとめ

本稿では、金融機関の実際の口座情報から得られる企業間の取引データと、業種分類情報を利用し、非財務データによる信用リスク評価モデルの構築を試みた。提案手法は、基本モデル、業種拡張モデルともに正常先の識別において F1 スコア 0.7 以上を達成し、特に基本モデルでは再現率 0.83 という高い識別能力を示した。さらに業種拡張モデルでは AUC が 0.73 に向上し、非正常先の識別能力も改善が見られた。これらの結果は、基本モデルが正常先の識別に優れ実務的な有用性を持つ一方、業種情報の追加が全体的な予測性能の向上に寄与することを示している。非正常先の識別能力には改善の余地を残すものの、財務データを保有していない企業群に対する信用リスク評価の可能性が示されたことは、本研究の重要な成果といえる。これにより、新規顧客開拓や既存顧客のリスク管理など、金融機関の営業活

動への応用可能性が示された。

今後の展望として、さらなるネットワーク特徴量の検討や、隣接ノードにあたる取引先の情報などを組み込むことで、非正常先の識別能力向上が期待される。本研究は、財務データに依存しない新たな信用リスク評価手法の確立に向けた第一歩であり、金融機関の経営効率向上に寄与する可能性を有している。

謝辞

本研究に必要なデータを提供していただいた YMFG に感謝する。

参考文献

- [1] Atan, E. Duymaz, A. Sarisözen, F. Aydın, U. Koraş, M. Akgün, B. and Gönen, M., “Corporate Network Analysis Based on Graph Learning,” in *Proc. International Conference on Machine Learning, Optimization, and Data Science*, Springer Nature Switzerland, Cham, 2022, pp. 268-278.
- [2] Vinciotti, V. Tosetti, E. Moscone, F. and Lycett, M., “The effect of interfirm financial transactions on the credit risk of small and medium-sized enterprises,” *Journal of the Royal Statistical Society Series A: Statistics in Society*, Vol.182, No.4, 2019, pp. 1205-1226.
- [3] Martínez, A. Nin, J. Tomás, E. and Rubio, A., “Graph Convolutional Networks on Customer/Supplier Graph Data to Improve Default Prediction,” in *Proc. Complex Networks X: Proceedings of the 10th Conference on Complex Networks CompleNet 2019 10*, Springer, 2019, pp.135-146.
- [4] Berloco, C. De Francisci Morales, G. Frassinetti, D. Greco, G. Kumarasinghe, H. Lamieri, M. Massaro, E. Miola, A. and Yang, S., “Predicting corporate credit risk: Network contagion via trade credit,” *PLoS One*, Vol.16, No.4, 2021, pp. e0250115.
- [5] Platt, H.D. and Platt, M.B., “A note on the use of industry-relative ratios in bankruptcy prediction,” *Journal of Banking & Finance*, Vol.15, No.6, 1991, pp.1183-1194.
- [6] Minakawa, N. Izumi, K. Sakaji, H. and Sano, H., “Graph representation learning of banking transaction network with edge weight-enhanced attention and textual information,” in *Proc. Companion Proceedings of the Web Conference 2022*, 2022, pp.630-637.