

# 与信モデルのハイブリッド AI 化

## Hybrid AI for Credit Models

小野 潔

Kiyoshi Ono

コムチュア 株式会社 デジタルイノベーション本部

Digital Innovation Dept., COMTURE Corporation

Hybrid AI credit models refer to advanced AI credit systems that combine different types of AI and DB techniques, including machine learning, Bayesian statistics, machine learning of imbalanced and alternative data. We will discuss the use of hybrid AI for simultaneous tuning of initial and developing credit models, machine learning of imbalanced data, factoring models, and the status of hybrid AI in the US.

### 1. はじめに

与信モデルは、顧客の信用情報や金融取引履歴、収入や支出などの情報を基に、顧客の信用力や返済能力を予測し、融資の可否や金利などを決定することができる。ハイブリッド AI 化された与信モデルは、異種の機械学習、統計学、深層学習、自然言語処理、またはオルタナティブデータなど、様々な種類の AI 技術や理論や DB 技術を組み合わせにより高度な処理や分析を行う AI 与信システムを指す。

90年代初頭、米国に機械学習による与信モデルが登場し、95年頃には日本でも導入された。当初は、個人向け住宅ローンやキャッシングに適用され、成功を収めた。2000年頃には、法人向けの中小企業モデルも開発されたが、収益性の高い短期融資の与信モデルの開発は、ハードルが高かった。

近年では、機械学習とベイズ統計学に基づいた複合モデルを使った複雑な与信モデルが登場した。また、オルタナティブデータの活用が進み、同時に中小企業モデル開発に必要な不均衡データの機械学習が進歩したことで、従来はモデル化できなかった分野まで、与信ビジネスが発展してきた。

当初の分類器は決定木、ロジスティック回帰が使われたがその後、アンサンブル学習などの高性能な分類器が使われた。最近では精度よりも XAI (Explainable AI) を重要視する傾向がある。なお分類器については別途論じるため、本稿では言及しない。

本稿では、筆者が関わった与信モデルのハイブリッド AI 化および与信モデルの潮流について論じる。

### 2. 与信モデル概論

与信モデルは、顧客の信用情報、金融取引履歴、収支などの情報を基に、顧客の信用力や返済能力を予測し、融資の可否や金利、融資額を算出できる。与信モデルには初期与信モデルと途上与信モデルの 2 種類がある。初期与信モデルは、新規融資において顧客の信用力を評価する。顧客の信用情報、返済能力、個人属性、金融取引履歴などを説明変数として用いて、初回の融資の可否や融資限度額を算出する。金融機関では、算出されたデフォルト率に基づいて顧客ランク(通常 10~8 クラス)を定義し、融資限度額の決定だけでなく、その後の借入れ枠の増減や利率の変更などにも利用する。

途上与信モデルは、既存の顧客に対して追加融資を行う場合に、初期与信モデルからの信用度の変動を反映して再評価し、融資限度額を調整する。例えば、すでに借入れをしている顧客が新たな融資を申し込む場合、与信モデルは顧客の返済履歴や事業状況、業界動向などを基に、再度顧客の信用力を評価し、適切な融資限度額を設定できる。

留意すべき点は、途上与信モデルの開発において、初期与信モデルと同じ構成で開発すると、運用上失敗が起こりやすくなる。例えば、途上与信モデルに初期与信モデルと同じ説明変数を用いる場合、経済状況の変化に応じて、顧客のデフォルト率が乱高下する。その結果、貸付限度額も頻繁に変更されることになり、運用現場で混乱が生じる。そのため、途上与信モデルでは前回の顧客ランクも考慮したモデル構築になる。

個人向けキャッシングや中小企業向け融資など、与信モデルは様々な業務に適用される。個人向けキャッシングの場合、個人信用情報(借入件数、借入金額、3 カ月間の照会件数等)のみで、モデルの 70~80% 以上の説明力がある。一方、中小企業向け融資の場合、中長期融資では財務指標や業種、金融取引などが説明変数として用いられるが、数カ月の短期融資では財務データから予測モデルを構築することが難しいため、外部企業情報センター(帝国データバンク、商工リサーチ)の情報とマージしたデータが利用される。短期融資の場合、企業情報センターへの照会件数や評点などが重要な説明変数となる。照会件数は1年間に以内に企業情報センターに該当企業の調査を依頼された件数で、回数が多いほど信用リスクが高まる。しかし照会件数は現在未公開であるため、企業情報センターが提供する他のデータを補って信用リスクを推測することになる。なお評点は調査員の恣意が入っており、しかも疑似正規分布であるため、審査基準に使えても倒産確率には使えない。

### 3. 個人向け与信モデルのハイブリッド AI 化

#### 3.1 初期与信と途上与信モデルの同時チューニング

個人向けの初期与信と途上与信においては、業種により異なる与信戦略が採用される。例えば、キャッシング与信の場合、銀行では初期与信を厳しくし、途上与信をやや緩くする傾向がある。途上与信において融資ストップや限度額の減額が発生すると、コールセンターに顧客の苦情が殺到し、銀行の信用問題につながりかねないためである。一方、消費者金融では逆の戦略を採用する傾向がある。消費者金融の顧客は既に他社から借り

入れをしていることが多く、苦情トラブルに発展しづらいためである。

初期与信と途上与信は相互補完的をしているため、両者の最適なモデルを組み合わせても、必ずしも運用上の最適なモデルとは限らない。従来は、与信に関する時系列データが不足していたため、初期与信と途上与信モデルのハイパーパラメータや顧客ランクの閾値などを同時に調整することはできなかった。しかし、最近ではデータレイクが整備されたことにより、このようなデータの不足問題が解消され、初期与信と途上与信の同時チューニングが可能になった。

筆者は初期与信と途上与信の同時チューニングに、マーケティング戦略も取り入れるため、半自動チューニングを採用する。半自動チューニングは、運用現場のフィードバックを反映しながら、最適な初期与信・途上与信モデルを見つけることができる。また、マーケティング拡大戦略に基づいた顧客ランクの閾値、ハイパーパラメータ、平均デフォルト率を外部パラメータにするため、金融機関は迅速に市場変化に対応できる。ただし、このような自動化や半自動化は、モデルの適用範囲や限界を理解し、運用現場での効果を常に検証することが重要である。

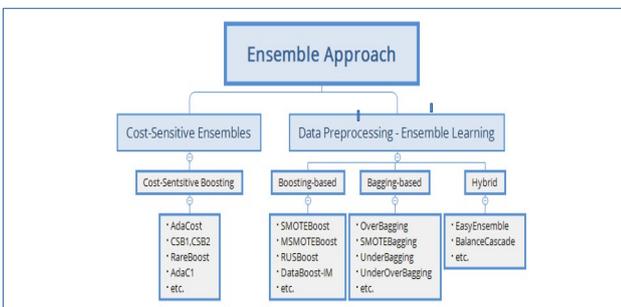
## 4. 中小企業向け与信モデルのハイブリッド AI 化

### 4.1 不均衡データの機械学習の発達

不均衡データは、ターゲット変数の倒産が極端に少ないデータが含まれるデータセットのことである。不均衡データをそのまま学習すると、多数派のクラスに偏った学習が行われ、少数派のクラスに対する予測精度が著しく低下する。これを防ぐには、サンプリング方法を変更するなど、データセットのバランスを調整する必要がある。

個人向け与信モデルでは信用情報の情報量が大きいため、影響が少なかったが、中小企業の与信モデル構築には不均衡データの機械学習の分析が必須になる。不均衡データの理論は無数にあり、下記の SMOTE は 80 種類以上も存在する。不均衡データの分析は統計学から発祥し、機械学習も含めて体系化されたのは、最近である。

モデル評価は少数派クラスの正解を重視する F2 値、PF 曲線、Recall 値を利用し、F1 値や AUC 値は用いない。不均衡データを扱う機械学習は、学習の偏り、モデルの性能低下、モデル評価指数の対策を用いることで、信用リスクの高いクラスに対する予測精度を向上させることができる。不均衡データの理論は ① Data-Level, ② Ensemble Learning(Fig1), ③ Machine Learning, ④ Cost-sensitive Learning, ⑤ Deep Learning Anomaly の Approach に大別される。



[Fig1. Ensemble Learning Approach に含まれる学習の事例]

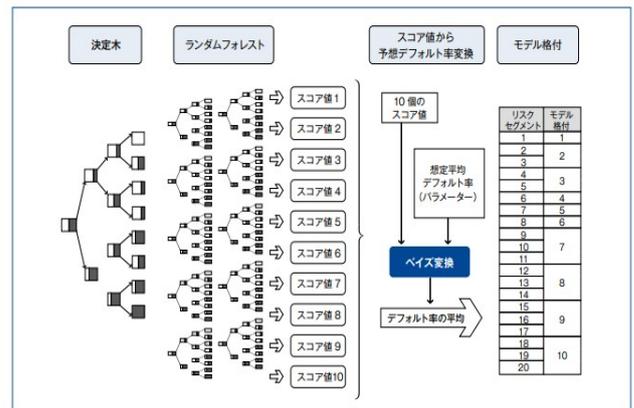
①は Over-sampling(SMOTE, ADASYN)と Under-sampling(Tomik Links)に分けられる。②は Bagging(Random Forests)と

Boosting(AdaBoost)に分けられる。④は少数クラスを誤判別した場合に、高コストのペナルティを発生させる。⑤は深層学習を利用した異常検知や少数データの生成を利用する。実務では Over-Sampling の SMOTE や Cost-sensitive Learning がよく用いられる。 [小野 2022]

### 4.2 機械学習とベイズ統計のハイブリッド AI 化

中小企業モデルでは最小限 300 個以上の負事例(倒産数)がないと、安定モデルを作れない。例えばあるファクタリング会社の中小企業の年間倒産率がわずか 0.1~0.5%である。倒産率 0.1%、負事例 300 件から母集団を推定すると、総融資件数は 30 万件になり、実務ではありえない件数になる。また倒産の定義を変更して件数を増やしても、安定的なモデルを作るには難しく、不均衡データを扱う工夫が必要になる。

筆者の中小企業の与信モデルは、① Under-Sampling でターゲットを正事例:負事例=1:1のサンプリングを行い、②分類器は Random-Forest を採用し、③倒産確率はモデルの確信度を事前分布でベイズ推定して求める(Fig2)。ベイズ推定には母数の平均倒産率と機械学習の確信度が用いられるが、この平均倒産率は全国の中小企業倒産率と相関がある。そこで全国の倒産率を予測することで、ベイズ推定により、個々の中小企業へ反映させることができる。これにより、経済状況の急激な変動が発生した場合、データ蓄積するまでの対処であるが、モデル再構築なしに景気変動を個々の企業の倒産確率に反映できる。 [小野 2014]



[Fig2. 中小企業の与信モデル構築の事例]

### 4.3 建築業のファクタリングモデルのハイブリッド AI 化

ファクタリングとは、企業が取引先からの支払いを受け取るまでに時間がかかる場合に、ファクタリング会社に売掛債権を売却することで即座に現金を手に入れることができる方法である。ファクタリングは、銀行融資に比べて手続きが簡単で、迅速に融資を受けられるメリットがある。

建築業のファクタリングは、中小企業モデルのみでは複雑な要因を取り込むことができない。筆者が関わったファクタリングモデルは中小企業モデル、建築物のスケジュール遅延モデル、連鎖倒産モデルから構成される複合モデルである。中小企業モデルは、前述の外部企業情報のデータから構築され、スケジュール遅延モデルには労務安全データ(グリーンサイトデータ)を使用する。また、建築業の親会社(債権元)が倒産すると、子会社や孫会社にまで広く影響を受けるため、連鎖倒産分析も必要になる。そこで、連鎖倒産モデルは系列構造をとり込むことができる GNN(Graph Neural Network)を検討している。

## 5. 新しい潮流

### 5.1 オルタナティブデータの活用

近年、オルタナティブデータの活用が進み、顧客の情報や購買履歴、SNSの投稿履歴などから顧客の特性を把握することができるようになった。これらのデータを与信モデルに取り込むことで、より精度の高い顧客の返済能力予測が可能になる。前述の建築業の労務安全データ(グリーンサイトデータ)を利用したスケジュール遅延モデルも、その例である。また、米国の信用情報のない難民向けの少額融資において、Facebookを一度だけ金融機関に開放させ、その内容から判定する与信モデルもある。

### 5.2 ブロックチェーン技術の導入

近年、ブロックチェーン技術の進展により、個人情報の漏洩や改ざんを防止することができ、また顧客の情報の分散管理が期待されている。これにより、与信モデルにおいて顧客の信用情報をより正確かつ効率的に取得できるようになり、与信モデルの精度向上が期待される。

### 5.3 米国 POS(Point of Sale) ファイナンスの拡大

オンラインでの自動与信においては、手続きを迅速かつ簡素化することが求められる。具体的には、自動入力機能の強化や必要なクリック回数の削減が課題となる。米国の信用情報機関 Experian は、顧客の申し込みにあたり、データベースから個人情報収集し、自動入力するサービスを提供する。この結果、借入の手入力作業を7割以上の削減に成功した。

米国の POS(Point Of Sale, 販売時点)ファイナンスは、前述の信用情報機関のサービスを利用することで、購買意欲や顧客ロイヤルティの向上に成功し、急速に拡大している。POSファイナンスは、クレジットカードや消費者金融の貸付に比べて、簡単かつ迅速に申請・審査・決済ができ、分割払いも可能である。POSファイナンスの貸手は新興フィンテック企業であり、主に若年層がオンラインで商品・サービスを購入する際に利用される。

POSファイナンスと自動与信は相性が良い。日本でPOSファイナンスを実現するには、信用情報機関が高度なサービスを提供することが不可欠である。しかし、残念ながら、日本の個人情報機関は24時間サービスや高度なサービスには対応しない。POSファイナンスは顧客にとっても有益であり、また日本の新しい金融業を育成する上でも重要である。したがって、法律の規制があるとはいえ、早急に対応することが望まれる。

[野村 2022]

### 5.4 自然言語処理の活用

銀行支店、クレジットカード会社の行員や社員は、担当企業との接点が多いため、担当企業のビジネスや経済状況に関する生情報を入手でき、それを営業日誌に記録している。金融機関では、営業日誌をデータベース化し、CRMに活用するまで至ったが、与信モデルには利用されていない。現在、金融機関や企業情報センターでは自然言語処理のBERTを用いて営業日誌を分析し、与信モデルの説明変数に取り込む方法を模索している。

## 6. 結び

本稿では、「与信モデルにおけるハイブリッド AI化」に関する調査結果および筆者のモデルを報告した。合わせて、従来のモデルの課題や解決法、不均衡データの理論について言及した。紹介した多くの分析法は既存技術であるが、ハイブリッド AI化として統合的に使用され始めたのは比較的最近のことである。

米国ではすでにハイブリッド AI化が進んでおり、日本でも同様に加速することが予想される。本稿では、米国の与信モデルにおけるハイブリッド AI化の方向性も紹介した。

今後は、新興フィンテック企業(例:LINE、PayPayなど)のオルタナティブデータや、営業日誌の自然言語処理を取り込んだハイブリッド AI化された与信モデルに注目することが重要である。

## 7. 参考文献

- [Chawla 2002] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. : SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357, 2002
- [Han 2005] Han, H., Wang, W. Y., & Mao, B. H. : Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing* (pp. 878-887). Springer, Berlin, Heidelberg., 2005
- [Bunkhumpornpat 2009] Bunkhumpornpat, C., Sinapiromsaran, K., & Lursinsap, C. Safe-level-smote: Safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 475-482). Springer, Berlin, Heidelberg, 2009
- [Fernández 2018] Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. : SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61, 863-905, 2018
- [Gordy 2019] Gordy, M. B., & Marchenko, M. (2019). Scalable Credit Risk Modeling with Heterogeneous Hazard of Default. *Management Science*, 66(5), 1845-1867, 2019
- [Guérard 2018] Guérard, S., Kharoubi-Rakotomalala, C., & Tétéjo, C. (2018). Economic capital allocation for credit risk using sensitivity-based approach. *Journal of Banking and Finance*, 94, 49-70, 2018
- [Liu 2018] Liu, Z., Li, X., & Liang, H. (2017). An innovative credit risk assessment model for small-and-medium-sized enterprises: Evidence from China. *Journal of Business Research*, 81, 163-174, 2018
- [野村 2022] 野村拓也: “米国における POS ファイナンスの拡大とわが国に求められる信用情報機関の高度化”, *Research Focus No.2022-04*, 日本総研, 2022
- [小野 2001] 小野潔: “データマイニングを利用した融資モデルの現状と課題”, *人工知能学会研究会資料 SIG-J-A004*, pp49-54, 2001
- [小野 2014] 小野潔: “与信モデル構築”, 2014 年日本 SAS ユーザー会研究発表論文集, SAS Institute Japan, 2014
- [小野 2022] 小野潔: “機械学習による不均衡データの理論体系”, 2022 年日本 SAS ユーザー会研究発表論文集, SAS Institute Japan, 2022
- [Mays 2001] Mays, Elizabeth: “クレジットスコアリング”, *シグマベイズキャピタル*, 2001