

銀行の自動審査ロジックの動的化

Dynamic Credit Model of Bank

小野 潔^{*1}
Kiyoshi Ono

松澤 一徳^{*1}
Kazunori Matsuzawa

^{*1} (株)インテック 金融ソリューション開発センター
Financial Solutions Development Center,
Financial Solutions Service Division,
INTEC Inc.

Abstract: Japanese banks begin to develop new credit models using data of Japan Credit Information Reference Center Corp(JICC). We need data of credit information at least three years to develop a credit model. Without sufficient information, the model we made is insecure and not be able to forecast bad debts. And we must select appropriate methods to develop a "stable" credit model. In this paper, we introduce a case of selecting a combination of two methods (1) Random Forest (2) Dynamic Initial Credit Model.

1. はじめに

1.1 背景

最近、銀行の与信モデルを見直す動きが始まっている。銀行が消費者金融業界の日本信用情報機構(JICC)に2012年から加盟できるようになったためである。JICC 信用情報を用いれば、従来のモデルよりも高精度の与信モデルを構築できる。しかし蓄積データは3年を満たさないため、デフォルト件数が著しく少ない。そのため与信モデルを従来の手法で開発したのでは、安定性に問題が生じる。

そのため本稿では初期与信モデルに二つの対策を講じた。①分類器にランダムフォレストを採用し、データの無駄を少なくし、②運用中の実績デフォルト率が悪化する緊急事態に対処するため、自動審査の動的化を初期与信モデルに組込んだ。

1.2 銀行の JICC 加盟への広がり

日本の個人信用情報センターは、銀行、信販会社、消費者金融会社の業種ごとに3社が存在する。信用情報センターは業種別に設立した経緯もあり、異業種の加盟を認めなかったが、貸金業法の総量規制(総借入を年収の1/3に制限)が2010年施行されるに当たり、銀行はJICCに加盟が認められた。2014年には82銀行(全銀行の65%)がJICCに加盟した。

	内 容	登録期間
本人特定情報	氏名、生年月日、性別、住所、電話番号、勤務先、勤務先電話番号、運転免許証等の記号番号等	契約内容に関する情報等が登録されている期間
契約内容	登録会員名、契約の種類、契約日、貸付日、契約金額、貸付金額、保証額等	契約継続中及び完済日から5年を超えない期間
返済状況	入金日、入金予定日、残高金額、完済日、延滞等	契約継続中及び完済日から5年を超えない期間 (ただし、延滞情報については延滞継続中、延滞解消の事実に係る情報については当該事実の発生日から1年を超えない期間)
取引情報	債権回収、債務整理、保証履行、強制解約、破産申立、債権譲渡等	当該事実の発生日から5年を超えない期間 (ただし、債権譲渡の事実に係る情報については当該事実の発生日から1年を超えない期間)
申込み情報	本人を特定する情報(氏名、生年月日、電話番号及び運転免許証等の記号番号等)、並びに申込日及び申込商品種別等	申込日から6ヵ月を超えない期間
.....

表1 (参考)JICC 信用情報の登録内容

2. 与信モデル構築アプローチ

2.1 自動審査のフロー

与信モデルは住宅ローン、無担保ローン、クレジットカード、マイカーローン等の融資判定を行う。与信モデルは自動審査システムに実装され、与信サーバーに構築される。自動審査システムは案件の受付、行内情報の収集、信用情報機構との通信、お客様への通知等の機能を有する。

自動審査システムのロジック(図1)では、最初に案件は否認基準でネガティブ・チェックし、信用情報センターへお客様の信用情報の登録と問合せを行い、借入総件数や借入総額による否認基準に抵触していないかを照合する。次に申請者の情報がすべて与信サーバーに集められ、与信モデルが起動する。与信モデルは倒産スコア値を算出し、該当するモデル格付(≒リスクセグメント)を確定させる。最後に自動審査システムは判定マトリックスで AVR 領域を判定し、融資案件を処理する。

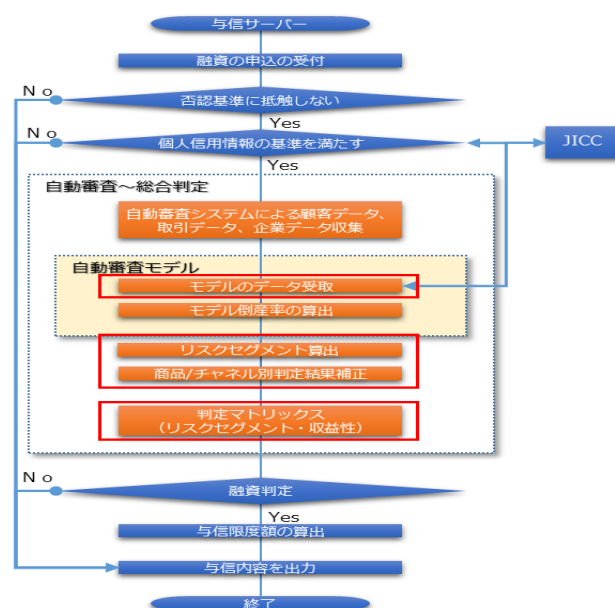


図1 自動審査システム

判定マトリックスは、モデル格付と融資額(または収益ランク)により与信判断を行う意思決定マトリックスである(図2)。AVR

領域とは、A(ACCEPT)領域が自動承認する White、V(REVIEW)領域が審査担当者により判定する Gray、R(REJECT)領域が自動謝絶する Black のことである。

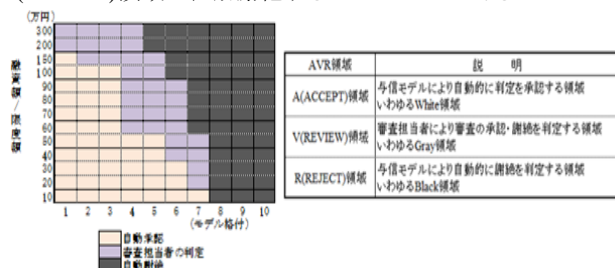


図2 判定マトリックスとAVR領域

3. 与信モデルの分類器

3.1 日本の与信モデル

日本の銀行は単に精度が高いモデルを望んでいるのではなく、審査担当者の理解できるようなモデルを作ることが肝要である。欧米の金融機関ではニューラルネットワークやサポートベクターマシン等の高精度の分類器を採用されることがあるが、両者とも判定理由を明確に示すことができないため、日本の審査担当者は受け入れ難い。決定木は判定理由を明確に示すことができ、さらにランダムフォレストの多数決の原理も、審査担当者には理解しやすい。

日本の金融機関では実務の与信モデルを公開しないため、分類器の正確な数字はわからない。筆者の経験上、ハイブリッドモデルが半数以上、ロジスティック回帰およびプロビット回帰が3割程度、その他少数ながらMBRも実務に採用されているようである。

3.2 ハイブリッドモデル

与信モデルの構築ではデフォルト損失が大きいオーバーフिटtingしない限界までモデルの精度をあげる。ハイブリッドモデルは、2つの分類器を直列に組み合わせる方法である(図3)。第1段階の分類器で得られたスコア値(もしくはデフォルト率)を説明変数に追加する点がミソである。第2段階では第1段階で得られたデフォルト率の寄与度が高いため、第2段階の分類器は微調整を行うことで、高精度が得られる。実務では最初に決定木を行い、次にロジスティック回帰を適用する。決定木のIF-THENルールが現場の審査担当者にわかりやすいためである。

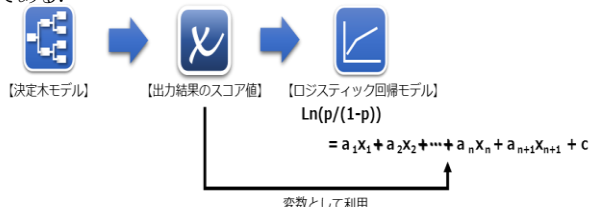


図3 ハイブリッドモデルの考え方

3.3 ランダムフォレスト

アンサンブル学習はブートストラップ法による復元抽出サンプリングを行い、弱い分類器を複数作成し、多数決で判定する方法である。本稿が採用したランダムフォレスト(Random Forest)は、決定木によるアンサンブル学習を改良し、属性もランダムに変化させる(確率的属性選択)分類器である(図4)。

ランダムフォレストの特徴は精度が高く、説明変数が数百でも効率的に動き、説明変数の重要度の推定、外れ値の特定あるいは欠損値を持つデータでも有効に作動し、ノードの個体数を制御できる点である。

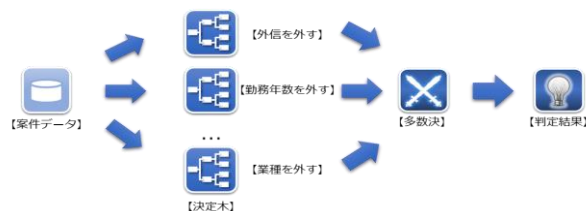


図4 ランダムフォレストの考え方

3.4 システム開発からみたフォレストの利点

与信モデルを実装した自動審査システムの開発作業は、コンサルタントではないエンジニア(コンサルタントとは別会社や別組織に所属する者が多い)により実施される。

この時ランダムフォレストは、IF-THENの集合体であることから、ロジスティック回帰等の統計手法に比べエンジニアに仕様の連携がしやすい。このことは限られた予算の中でモデルを構築する際、その精度を決定付ける顧客データの分析やヒアリング等のコンサルティング業務に、開発作業と比して大きな時間や多数の人員を投入できることに繋がり、結果的に他の分析手法を用いた時よりも深い分析に基づいたモデルを実装できる。

また金融機関の与信システムは損益に直結することからバグの混入が許されず、100%の信頼性が求められる。このためテストパターンが多岐にわたるが、ランダムフォレストのシステムテストはIF-THENの境界値や欠損値(本来入るべき値が欠損値になるイレギュラーケース)を中心にテストが行われる。テストパターン数はたかだか有限でしかもテストパターンをプログラム化できる。本稿の無担保ローンモデル(10個の決定木から構成)でおよそ3千件のテストでその動作を確認した。

他方、統計手法を利用する場合、そのテストの網羅性を金融機関が検証できず、十分なシステムの信頼性を担保できないのが通例である(理論の問題ではなく、システムの信頼性が100%に断定できない)。

4. 初期与信の動的化

4.1 背景

住宅ローンやマイカーローンには途上与信がなく、初期与信モデルがすべての融資可否を決定する。もし初期与信モデルの理論値と運用中の実績値と乖離した場合、従来は銀行がモデル格付を引き下げる与信規定の改訂で、融資審査を厳しくする。しかし銀行の与信規定改訂は、金融庁の監査対象と成りうるため、容易なことではない。例えば半年間の集計等の裏付が必要であり、その間も不良債権が増え続ける。

そこで本稿の自動審査ロジックでは、モデル格付範囲や与信モデル自体を変更せずに、バイズ更新を利用して、実測デフォルト率に追従する動的化機能を自動審査システムに組み込み、予想デフォルト率を変更する。

4.2 初期与信の動的化

本稿の自動審査システムでは、銀行の担当者が目標デフォルト率(例えば運用中の実績デフォルト率)を登録することにより、モデルから算出されるデフォルト率の分布を変更できる(図5)。

従来の与信モデルではコンサルタントにより目標デフォルト値が固定化されており、モデル完成後は変更できない。本稿の自

動審査システムは、登録インターフェースを有し、銀行担当者が目標デフォルト率を設定変更できる。その結果、銀行担当者が審査基準の制御をできるようになる(文献[5])。

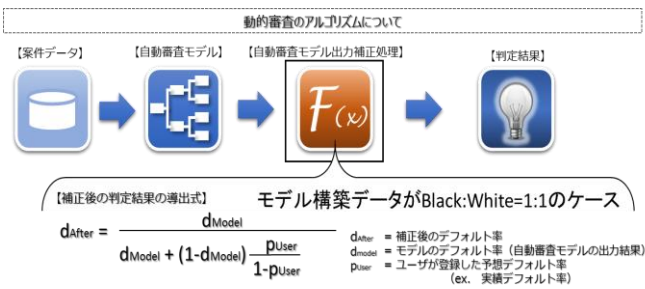


図5 初期与信の動的化

5. 結果

5.1 データ準備

無担保ローンの与信モデル構築には2年間約1万件データを使用した。ただJICC信用信息データの蓄積期間が2年未満と短いため、金融商品のデフォルトがあまり発生せず、デフォルト件数は1%未満と少ない。そのためモデル構築は、すべてのデフォルト案件とランダムサンプリングした正常案件からBlack:White=1:1のデータセットを作成した。ランダムフォレストでは、すべての「デフォルト案件」と「復元抽出を繰り返した数十個の正常案件」をペアにしたデータセットを用意した。

モデル構築は学習データ数:検証データ数=7:3で作成した。無担保ローンモデルの変数は約250個存在し、加工変数を合わせて約700個から採用変数約20個を選択。住宅ローンモデルの変数は約200個、加工変数を合わせて約500個から採用変数約20個を選択。

5.2 信用情報を用いた無担保ローンの決定木

変数選択される重要な信用情報は、異動情報(○事故情報)、借入専業件数(専業=消費者金融会社)、6カ月の照会件数(ローン・クレジット申込回数)、借入総残高等である。「信用情報を利用したモデル」と「利用しないモデル」では、決定木の構造と予想精度に大きな差異が発生する。「信用情報を利用したモデル」の決定木の上位層の分岐は信用情報のみで構成される。また「使わないモデル」よりも10%以上の精度向上が見られた(図7)。

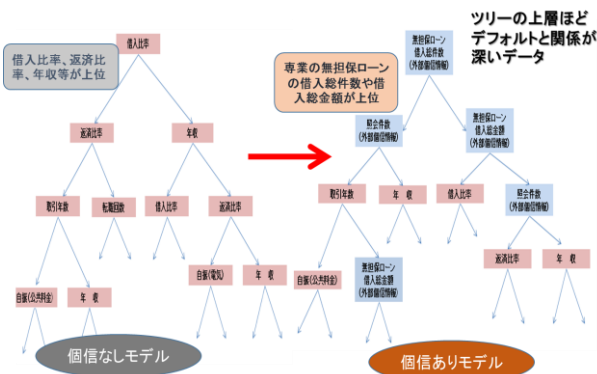


図6 信用情報を利用した無担保ローンの決定木

5.3 ハイブリッドモデルとランダムフォレスト

少数のデフォルトデータにハイブリッドモデルを適用した場合、運用中に精度に大きなぶれが生じ、モデルの安定性に問題が生じる恐れがある。今回の場合、ハイブリッドモデルでは1データセットから開発されるため、大量の正常案件がモデル構築に使用されない。それに対して、ランダムフォレストは複数の決定木から判定され、無駄になる正常データが少ない点から、少なからず安定性がハイブリッドモデルより向上すると期待される。

住宅ローンモデルのAR値は約85%、無担保ローンモデルのAR値約75%となった。ハイブリッドモデルのROC曲線はランダムフォレストとほぼ同じ曲線となり(図8では重なっている)、同期間の検証データだけでは優劣がつかない。本来ならば期間を変えてモデルの安定性を比較するのだが、信用信息の蓄積データが不足のため検証はできなかった。

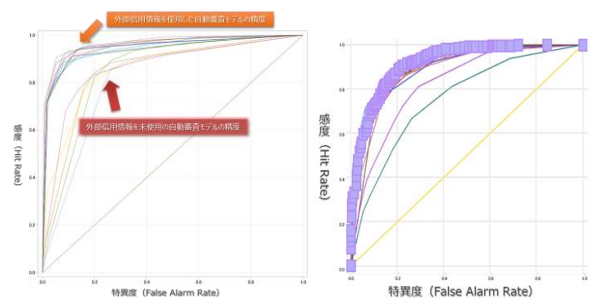


図7 無担保ローン(ROC図) 図8 住宅ローン(ROC図)

6. 今後の課題

銀行がJICC信用信息を入手できるようになったため、今後の与信モデルの高度化は「判別力の向上」よりも「収益の算出」と「ビックデータの利用」に移りつつある。これらについて説明を加える。

今、解決すべき課題は、収益マトリックス(図2)を構成する住宅ローンの収益である。住宅ローンは返済期限が長期のため、極力リスクを避けるべきであるが、最近の低金利競争で多大なリスクを取っている。そのため新規融資の半数近くが「借換」という銀行も存在する。この事実を裏返せば、5~7年後には住宅ローンの半数が解約するリスクを内蔵している。

金融庁/日銀のガイドラインでは「住宅ローンの収益は単純に金利と残高から求めるだけでなく、中途解約、経年解約、金利変動リスク等も考慮すること」(表2, 文献[3][4])を指導している。事実、金融庁監査では地銀の住宅ローンの収益性を問題視しており、地銀の再編成をあと押しするための指摘事項にもなっている。しかし30年以上の長期わたる住宅ローンの実務データが揃っている銀行は存在しないし、住宅ローンの各リスクの統合モデルは理論的に未完成である。現在、メガバンクと大手地銀が5~10年後の収益予想に取り掛かっているが、まだ与信モデルには利用されていない。

住宅ローン収益の算出リスク	市場リスク	市場金利変動リスク ブリベメントリスク	金利リスク管理
	信用リスク	デフォルトリスク 回収率変動リスク	
	経費	団体信用生命保険料 審査費用 サービシング・コスト (管理回収経費)	生涯収益シミュレーション

表2 住宅ローン収益の高度化

日本の与信モデルにおける「ビックデータの利用」は遅れている。日本では取引データの利用中心に研究されているが、審査官のアイデアの域を出ておらず、ビックデータの利用まで行き着いていない。それに対して特に米国・中国ではソーシャルメディアを活用した小口融資のリテールモデルが 2010 年以降に実用化している。日本でもソーシャルメディアが、信用リスクを予測する重要データに認識が変わりつつあり、筆者も次世代の与信モデルへの活用に検討したい。

7. おわりに

無担保ローンモデルでは JICC 信用情報だけで 7 割以上の説明力を有するため、信用情報の大きな変動がない限り、安定的である。しかし少数のデフォルトデータから開発した「無担保ローンは運用1～2年後から」、「住宅ローンは運用4～6年後から」実績デフォルト率が急速に悪化する可能性を否定できない。銀行のモデル運用の失敗は、巨額の損失が発生するため、本稿で提案した自動審査ロジックの高度化は欠かせない。

参考文献

- [1] T.Hastie, R.Tibshirani, J.Friedman; The Elements of Statistical Learning---Data Mining, Inference, and Prediction Second Edition. Springer, 2009.
- [2] エリザベス・メイズ, “クレジットスコアリング”, シグマベイスキャピタル, 2001.
- [3] 日本銀行金融機構局, “住宅ローンのリスク管理”, 2007.
http://www.boj.or.jp/research/brp/ron_2007/data/ron0703c.pdf
- [4] 日本銀行金融機構局, “住宅ローンのリスク・収益管理の一層の強化に向けて”, 2011.
https://www.boj.or.jp/research/brp/ron_2011/ron111124b.htm/
- [5] 特許出願番号: 2014-238446 出願者: 小野潔 (所属 株式会社インテック)
- [6] 小野潔, 松澤一徳, “与信モデル構築”, SAS ユーザー総会 2014 論文集, pp.715-725.
- [7] 小野潔, 松澤一徳, “銀行の与信モデルの高度化”, 日本不動産金融工学学会: JAREFE 定期大会予稿集, pp.25-33, 2015.
- [8] 小野潔, “ハイブリッド・コンポーネントの構築”, 第 20 回日本 SAS ユーザー会研究発表論文集, pp.269-327, SAS Institute Japan, 2001.
- [9] 大久保豊, 尾藤剛, “ゼロからは始める信用リスク管理”, きんざい, 2011.
- [10] 本田義一郎, 三森仁, “住宅ローンのマネジメントを高める”, きんざい, 2004.
- [11] 下川俊雄, 杉山知之, 後藤昌司, “樹木構造接近法”, 共立出版, 2013.
- [12] J.R.キンラン, “AI によるデータ解析”, トッパン, 1995.