

# インテックの与信モデルの特徴と今後の展開

小野 潔

## 概要

2000 年頃から日本の金融機関は業務効率化と人員削減のために与信モデルを積極的に導入した。2012 年から銀行が消費者金融業界の個人信用情報センター（JICC）に加盟できるようになり、近年、銀行の与信モデルを見直す動きが始まった。

与信モデルは統計学や機械学習を利用して融資判定を行う。与信モデルの構築プロセスは一般モデルと同じであるが、金融の特有なロジックを併せ持つ。インテックの与信モデルの特徴は『機械学習を用いた点』、『動的与信機能を自動審査システム（F<sup>3</sup>エフキューブ）に組み込んだ点』である。

インテックの与信モデルは海外でも稼働できるので、アジアへの展開も模索している。また世界ではマイクロファイナンス（数万円の少額ローン）の自動審査が課題になっており、インテックでは AI&IoT 技術を駆使した新たな発展ステージに向かう与信モデルに注視している。

## 1. はじめに

本稿はインテックの与信モデルの技術を紹介する。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統計学や機械学習に基づいて行う。与信モデルは自動審査システム（インテックのF<sup>3</sup>エフキューブ）に組み込まれ、審査担当者の作業をシステム化することにより、業務効率化と人員削減を実現できる。現在ではリテール業務<sup>(1)</sup>を強化するために必要不可欠なシステムである。与信モデル構築法は米国のスコアリング技術<sup>(2)</sup>（1990～95年頃）がベースとなり、データマイニング技術により構築法が確立した [1]。

与信モデルは顧客属性、取引属性、勤務先情報、個人信用情報から融資案件のデフォルト<sup>(3)</sup>を予想する。その中でも融資履歴の信用情報が最もモデルに有効なデータである。与信モデルの構築プロセスは一般モデルと同じであるが、与信モデルの構築には“コンプライアンス”“高精度の分析法”“モデル格付”“判定マトリックス”“AVR 領域”（後述）等の金融の特有な観点やロジックを併せ持つ [2]。

本稿ではインテックの与信モデルについて、最初に与信モデルに最も影響がある個人信用情報の説明を、次に自動審査システムの審査フローや与信モデルで採用した機械学習を解説し、後半は金融特有な構築プロセスと今後の展開に言及する。

(1) リテール業務は預金、振替、住宅ローン、キャッシングなど個人向けのサービスを指す。

(2) スコアリング技術は顧客属性を点数化/数値化し、合計点で融資の可否を判定する技術。

(3) デフォルトは「債務不履行」を指す。具体的にはローンのデフォルトは顧客の破産等の原因により元本や利息の支払いを3ヵ月以上遅延したり、停止する状況を指す。個人信用情報センターでは債務不履行になった場合、5年間記憶が保持される。

## 2. 個人信用情報

与信モデルが広まったのは、全国レベルの個人信用情報センターが創立されたためである。金融機関は顧客の個人信用情報（個人の借入金融商品、借入件数、借入金額、返済履歴、事故情報等）をセンターに照会／登録できる。与信モデルは信用情報を取り込むことで、精度が著しく向上する。特に無担保ローンモデル（キャッシング）は精度の7割以上は信用情報から得られると言っても過言でない。

与信モデルの構築は大手銀行で1997年頃から始まった。地方銀行での導入は2005年頃から始まり、今では大部分の地方銀行にも浸透している。当時の信用情報センターは業種別であり、銀行は消費者金融業界の信用情報センターに加盟できなかった。そのため無担保ローンモデルの精度が低くなり、多くの銀行は消費者金融会社の保証をつけた。ところが2010年の貸金業法の総量規制<sup>(4)</sup>の施行により、2012年から銀行等の金融機関が消費者金融業界の個人信用情報センター（日本信用情報機構：JICC<sup>(5)</sup>）に加盟できるようになった。そのため最近では、既に導入された金融機関の与信モデルを見直す動きが始まった。

## 3. 自動審査システムの概略フロー

自動ローンの審査システムは与信サーバーに構築される[3]。図1に審査システムの概略フローを示す。お客様はWeb上からローン申込を行い、その情報に基づき、自動審査システムのサーバーは個人信用情報センターへ個人信用情報の照会を行う。申込情報や個人信用情報等<sup>(6)</sup>は与信サーバーに集められ、与信モデルが融資判定を行う。審査結果の大半は自動判定（承認または謝絶<sup>(7)</sup>）されるが、明確に自動判定できないイレギュラーな案件が発生する。後者は審査担当者による判定を行われるが、その案件数を極力少なくすることが肝要である。最終的な審査結果は、自動審査システムからメールでお客様へ伝えられる。

## 4. 与信モデルの分析法

米国の与信モデルでは判別力が高い機械学習（サポートベクターマシンやニューラルネットワーク等）が採用されるが、日本の金融機関では原因を探る決定木やロジスティック回帰を採用する傾向がある。ロジスティック回帰は統計学の手法であり、精度が高いが審査担当者にはやや理解しづらい[4]。決定

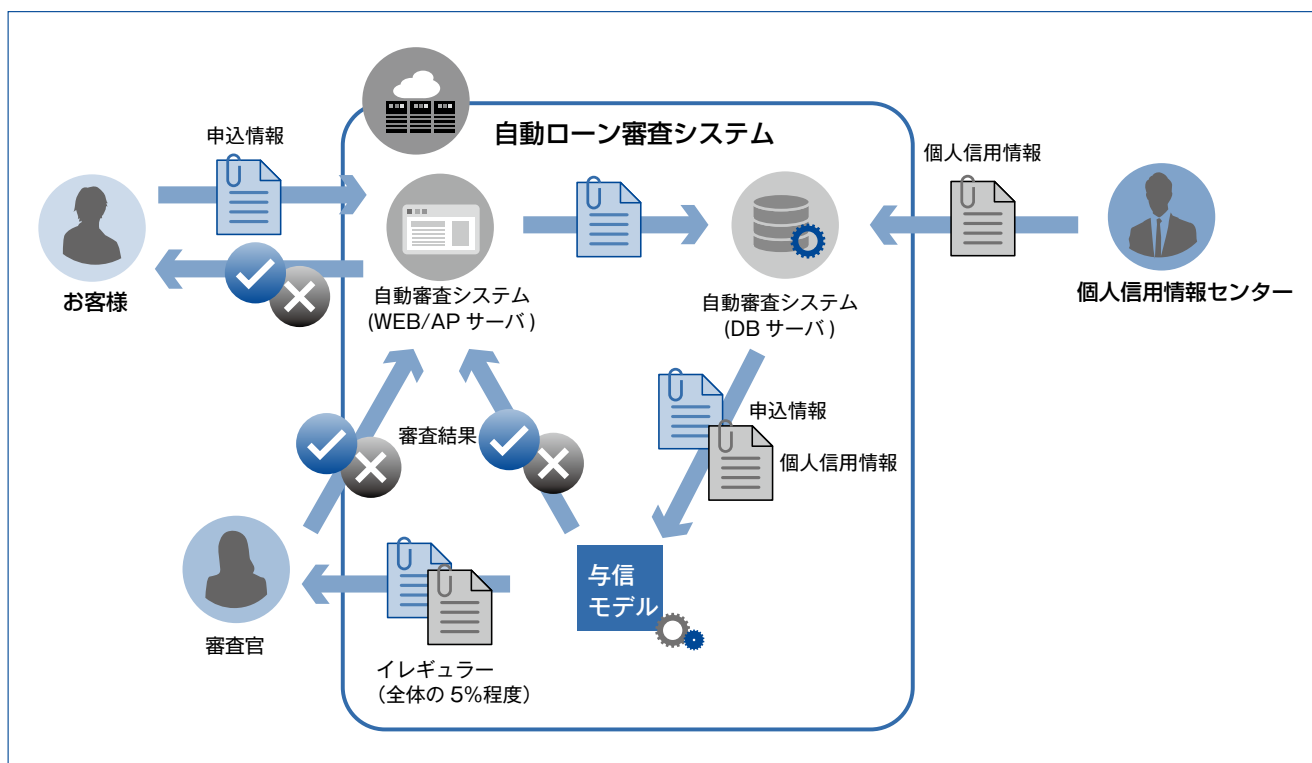


図1 自動審査システムの概略フロー

(4) 総量規制は多重債務問題を解決するために、年収の3分の1を超える借入を規制する法律。

(5) JICCは全国の消費者金融会社が加盟しており、リアルタイムで無担保ローン（キャッシング）の個人信用情報を入手できる。2016年4月時点では87銀行（全銀行の約65%）がJICCに加盟し、ノンバンク、消費者金融などの個人の借入情報を照会している。

(6) 詳しくは行内の取引情報やコンプライアンス情報、さらに住宅ローンモデルでは帝国データバンク社の企業概要データ（業種、売上、所在地、資本金、評点等）が使われる。企業概要データは約150万社の企業情報を収録したデータベースであり、勤務先の企業の継続性や安定性の判定に利用される。

(7) 金融機関ではローンやクレジットをお断りすることを「謝絶」と言う。

木は機械学習であり、精度はロジスティック回帰よりもやや低い、審査担当者には理解しやすい[5]。また実務分析ではデフォルトデータが充分にないケースやデータに偏りがあるケースが多く、前述の分析を単純に適用してもモデルの信頼性が劣る。インテックの与信モデルでは課題に対応するために、ランダムフォレストモデル（後述）を採用した。

決定木はデフォルトしやすいグループをデータ属性から分類判別する。結果はツリー構造で表現され「もし…ならば～である」というIF-THENルールを導出できる。決定木の特徴はIF-THENルールに直すことで、専門家の知識を抽出できる点にある。信頼区間などの統計指標は得られないが、分析結果は審査担当者にわかりやすい。ただ6～8階層に達すると、ルールの矛盾が発生する。ツリー構造の分割の優先順位は属性の分割基準値に基づくため、上層ほどデフォルトに影響ある変数で分割される。分割基準値はルールが目的属性値（デフォルト）の分布に与える影響度合いを数値化したものである[6]。

## 5. アンサンブル学習とランダムフォレスト

決定木の弱点は、分析データセットに偏りがあると、モデルが大きく変わることである。そのためインテックの与信モデルでは複数モデルから構成されるアンサンブル学習を採用した。アンサンブル学習は機械学習の手法で、高精度・安定（ロバスト）性に強いモデルを得られる。分類器を複数組み合わせ、その結果を統合することにより、個々の分類器よりも精度を向上させる。構築法は復元抽出サンプリング<sup>(8)</sup>を行い、“あまり精度を高めない分類器<sup>(9)</sup>（弱い分類器）”を複数作成し、多数決で判定する。実装では浅い層の決定木を5個～100個作成し、各決定木の予測値の平均値を代表値とする。図2に、決定木のアンサンブル学習であるランダムフォレスト<sup>(10)</sup>を示す。ランダムフォレストは2001年 Breiman [7] によって提案された学習法であり、サンプリングだけでなく従属属性（説明変数）もランダムに変化させる（確率的属性選択）[8]。例えるならば、異なる専門分野の審査担当者が多角的に分析し、最終判定を多数決で決めるようなものである。図3に分類器の数が一定数に達すると、誤差が収束することを示す。一般に与信モデルでは5個から20個以内に誤差が収束する。

インテックの与信モデルがランダムフォレストを採用した理由の一つは、決定木のIF-THENルールに非線形の関数を使用しないため、エラーが発生しづらい点にある。また決定木は欠損

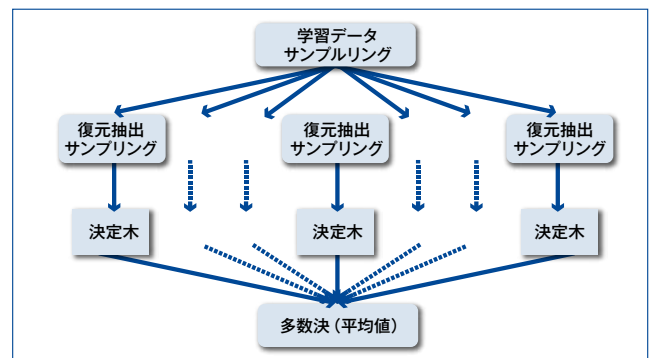


図2 アンサンブル学習の構成

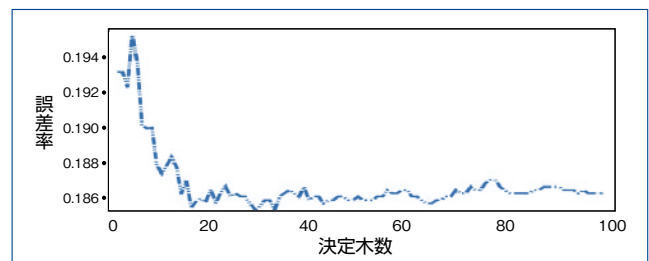


図3 決定木の数と誤差の収束

値が含まれても、欠損値自体を一つのカテゴリーと数えるため、分岐処理のエラーが発生しない。決定木やランダムフォレストは停止することが許されない金融機関のオンラインの与信業務へ適用しやすい手法と言える。

## 6. 与信モデルの構築アプローチ

与信モデルは顧客のデフォルト率を算出する。図4に、決定木からモデル格付の算出までの計算フローを示す。

金融商品のデフォルト率が数%程度のため、デフォルト件数が少ない。このデータ比率でモデル開発すると、データに偏りがあるため、デフォルトの特徴が読み取れない恐れがある。インテックの与信モデルでは、その問題を解決するため【最大のデフォルト件数：正常案件数＝1：1】の復元抽出のサンプリングを数十回行い、ランダムフォレストモデルを構築する。サンプリングデータは現実よりもデフォルト比率が高いため、与信モデルの算出値はスコア値といい、ベイズ変換することで理論上のデフォルト率になる[9]。

ただデフォルト率のままでは、審査担当者には利用しづらいので、デフォルト率に応じたモデル格付<sup>(11)</sup>を設定する。最初にデフォルト率の最大値を20等分したリスクセグメントを作成する。例えばリスクセグメント3に属する案件はリスクセグメント1の約3倍のデフォルト率を有する。次にリスクセグメントに含ま

(8) 復元抽出サンプリングは母集団から標本を抽出するときに毎回もとに戻してから次のものを取り出すサンプリング方法。

(9) 人工知能の機械学習では分析法を分類器と呼ぶ。なおロジスティック回帰はもとより精度が高いため、アンサンブル学習では精度が向上しない。ランダムフォレストではロジスティック回帰と同等以上の精度を得られ、しかもデータの変動にも強い。

(10) ランダムフォレストは多数の決定木（ツリー）モデルから構成され、その名がつけられた。各決定木の出力平均値を代表デフォルト率とすることが多い。

(11) 格付は法人企業の信用リスクに応じて割振るものであるが、モデル格付は個人のデフォルト率に応じて行う。金融機関により相違するが、10段階の格付分類が多い。

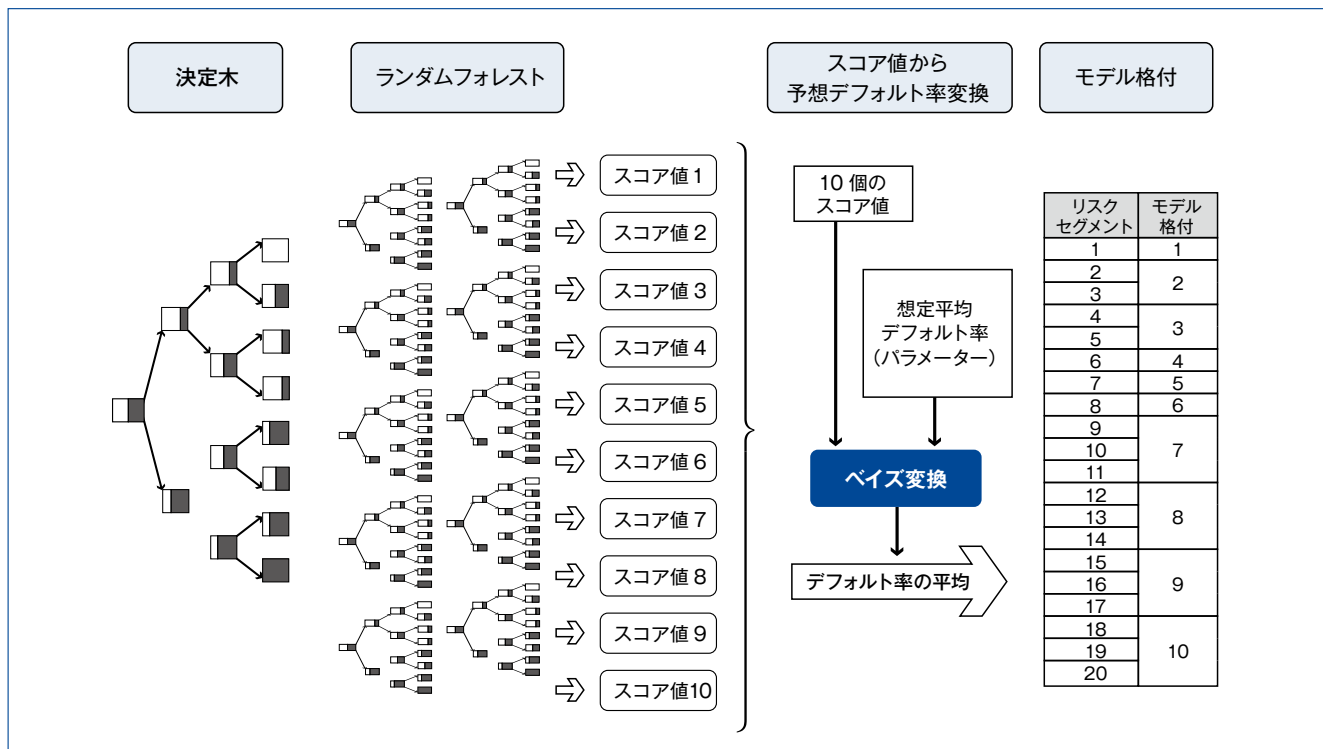


図4 決定木からモデル格付までの計算フロー

れる案件数を考慮に入れて、複数のセグメントを集めてモデル格付を設定する。

判定はデフォルト率と、デフォルトした場合の損失額（≒回収額）や収益を考慮するため、モデル格付と回収金額（あるいは生涯収益）の判定マトリックスを利用する [10]。図5に、モデル格付と回収金額ランクによる与信判断の決定マトリックス上の AVR 領域を示す。AVR 領域は A 領域が承認自動判定、V 領域が審査担当者判定、R 領域が謝絶自動判定を意味する。

モデル格付や AVR 領域は運用に関わるため、金融機関とのミーティングを重ねることで設定できる。運用が始まると、これらの設定を適宜見直す。

## 7. 動的与信機能

新たに構築した与信モデルでも、実務運用中にモデルの想定デフォルト率<sup>(12)</sup>が実測値から乖離する不測の事態が生じる恐れがある。そこでインテックの与信モデルでは、ベイズ統計学に基づく『動的与信機能』<sup>(13)</sup>を自動審査システム（インテックのF<sup>3</sup>エフキューブ）に組み込んでいる [9]。ユーザーは自動審査システムの想定デフォルト率のパラメーターを変更するだけで、運用中の審査判定を厳しくも緩くもできる。その結果、モデルの想定デフォルト率を実測デフォルト率に追従させることができ、最適な判定が可能になる。

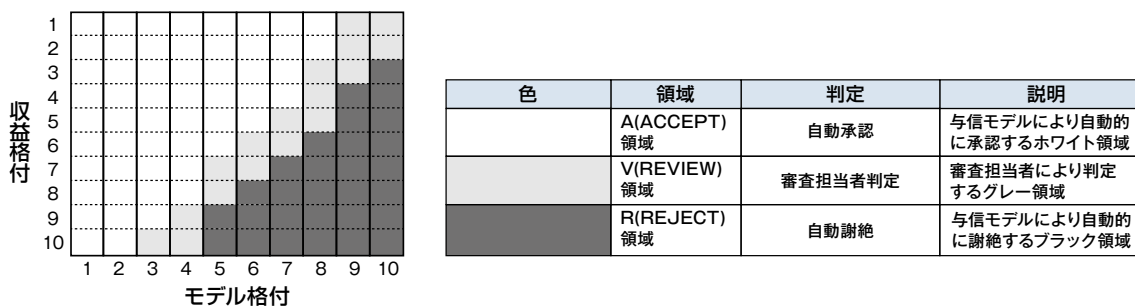


図5 AVR 領域

(12) 想定デフォルト率とはモデル構築時に設定したデフォルト率を指す。一般に過去の年間平均デフォルト率が使われることが多い。

(13) 例えば当初の与信モデルの構築時の平均デフォルト率が1%であったが、経済不況になり実測デフォルト率が2%に上昇したケースを考える。実測デフォルト率が2倍なので、与信モデルが算出したデフォルト率も2倍にすると、例えばデフォルト率60%の案件は120%となり統計学が成り立たない。動的与信機能はベイズ更新を用いてこの問題を解決した。

## 8. 金融機関が望む与信モデル

与信モデルはコンプライアンスに抵触することがある。一般に子供数が多いほどデフォルトが高まるため、うっかりすると子供の人数が多い家族ほど住宅ローンを受けづらい与信モデルができてしまう。少子化が問題の日本で、金融機関が住宅ローンモデルに子供数を組込むことは、法律上、問題がなくとも抵抗がある。インテックはたとえ精度が低くなったとしても、コンプライアンスに抵触しないように、金融機関に助言している。

以前は所属業界により、与信モデルは相違していたが、どの金融機関も JICC 個人信用情報が使えるようになると、同じような与信モデルへ収束していく。ただ銀行が消費者金融会社型のモデルを運用しても、不良債権の回収率が異なるため、判定マトリックスが違うものとなり、運用面は同じにならない。今後も所属する金融業界に合わせた与信モデルの再構築が必要である。

与信モデルの構築は、単に精度のよいモデルを開発するだけでない。日本の金融機関は高精度モデルよりも、審査担当者が理解できるような与信モデルを望んでいる。インテックでは、現状の経済・コンプライアンス事情を把握し、金融庁・日本銀行の方針に合わせたモデルを構築する。

## 9. 今後の展開

日本の多くの金融機関は既に与信モデルを導入しており、新規の金融商品に対応した与信モデルの開発は続いている。一方、アジアの自動審査システムは日本・韓国・中国の一部でしか稼働しておらず、日本よりも大きな発展の可能性を秘めている。インテックの与信モデルは全国レベルで信用情報を蓄積している国ならば稼働できるので、アジアへの展開を模索している。

世界に目を向ければ、融資額数万円以下のマイクロファイナンスが重要課題である。少額ファイナンスは移民・難民・貧民層のsmallビジネスの支援につながるため、社会的な意義が大きい。しかし顧客が信用情報センターに登録されていないため、従来の与信モデルではまったく対応できない。米国・ドイツの企業ではこの課題のために AI & IoT 技術の利用を進めている [11]。従来の与信モデルは金融機関の発案であったが、これからはフィンテック (Fintech) の産物であり、鍵を握るのは IT ベンダーである。インテックの最先端技術と与信モデルの融合は、少額ファイナンスの自動審査を新たな発展ステージに引き上げることになる。

## 参考文献

- [1] エリザベス・メイズ (スコアリング研究会訳): クレジットスコアリング, シグマベイスキャピタル, (2001)
- [2] 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)
- [3] 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, pp715-725, (2014)
- [4] 丹後俊郎, 高木晴良, 山岡和枝: ロジスティック回帰分析 (SAS を利用した統計解析の実際), 朝倉書店, (2013)
- [5] J.R. キンラン (古川康一監修): AI によるデータ解析, トッパン, (1995)
- [6] 鈴木義一郎: 情報量基準による統計解析入門, 講談社, (1995)
- [7] Breiman, L.: Random Forests, Machine Learning, 45(1), pp5-32, (2001)
- [8] 下川俊雄, 杉山知之, 後藤昌司: 樹木構造接近法, 共立出版, (2013)
- [9] 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)
- [10] 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/](https://www.boj.or.jp/research/brp/ron_2011/ron111124b.htm/)
- [11] 増井智則: ソーシャルメディアを活用した新たなリテール金融と信管理の可能性, 三菱総合研究所 (2014)  
<http://www.mri.co.jp/opinion/column/uploadfiles/20140929report.pdf>

本論文には他社の社名、商号、商標および登録商標が含まれます。



**小野 潔**

ONO Kiyoshi

- 社会システム戦略事業部  
社会システムプラットフォーム開発部
- データ分析、モデル構築に従事 データサイエンティスト
- 人工知能学会員、日本不動産金融工学会員、  
SAS ユーザー会世話人／論文審査員、  
日本不動産金融工学会評議員