

# 希少事象を考慮したスコアリングモデルを実現するAI技術の活用事例と今後の展望

Hitachi AI Technology/Prediction of Rare Case (AT/PRC) は、実ビジネスにおいて発生頻度の低い事象の起こる確度を高精度で予測し、その根拠を提示する日立のAI技術である。近年、企業活動にAIを活用する動きが加速している状況下において、AT/PRCは日本国内だけでなく国外の顧客に対しても成果を示してきた。

本稿では、AT/PRCの技術的特長や国内外の活用事例について紹介し、AT/PRCをはじめとするAI技術を顧客業務に適用するうえでの課題と、それに対する取り組みや今後の展望について述べる。

宇都 陽平 | Uto Yohei

仲田 智将 | Nakata Norimasa

田代 大輔 | Tashiro Daisuke

赤崎 充洋 | Akasaki Mitsuhiro

辻 琢矢 | Tsuji Takuya

## 1. はじめに

昨今、さまざまな企業活動にAI(Artificial Intelligence)を活用する動きが加速しており、日々新しいAI技術の研究開発と社会適用が進められている。日立が開発したAI技術の一つであるHitachi AI Technology/Prediction of Rare Case (以下、「AT/PRC」と記す。)は、当初は国内金融機関のローン審査業務での実証実験などで活用されてきたが、海外金融機関との協創案件や、ローン審査以外の顧客業務を対象とした実証実験においても成果を上げてきた。

本稿では、AT/PRCが持つ技術的特長について述べたうえで、過去の適用事例について紹介する。また、AT/PRCを顧客業務に適用する中で見えてきた課題とそれに対する日立の取り組みについても紹介する。

## 2. AT/PRCの技術的特長

AT/PRCが持つ二種類の技術的特長について紹介する。

### 2.1

#### 「まれな事象」に対する過学習抑止の仕組み

実ビジネスにおいて「借入金が返済不能になる」、「膨大な数の中に埋もれた不正」、「機械が故障する」といったまれにしか発生しない事象をAIに学習させようとしたとき、いわゆる過学習の影響で精度が十分でないという問題があった。過学習とは、AIのモデルを作成する過程において、モデルが学習データに過剰に適合してしまい、学習に使っていない未知データでは精度が落ちてしまう現象である。従来の一般的な手法では「まれな事象」の特徴を取り出そうとして、ノイズまで学習してしまっ

ていた。AT/PRCはこの過学習の発生を抑止する「シグナル&ノイズに基づく学習」という仕組みを持つ。これは学習過程において以下の処理を行う。

- (1) 学習データをいくつかの小さなグループに分割する。
- (2) グループごとに学習処理を実行し、モデルのパラメータを更新する。
- (3) グループごとにモデルが「事象の発生に影響する」と捉えた特徴量を比較し、これが共通ではなかった場合、その特徴量を予測精度を下げる要因となるノイズと判断する。ノイズと判断された特徴量の重みを下げる処理を実施する。

(3)の処理によって、ノイズと判断された特徴量は相対的に重要視されなくなる(図1参照)。

この仕組みによって、AT/PRCは「まれな事象」の発生を学習する場合であっても、過学習が発生して未知データに対するモデルの予測精度が下がることを抑止している。

## 2.2

### AIの説明可能性を高める二つの機能

重要な判断や予測にAIを活用する際には、業務に関わるステークホルダーに対して、AIの判断根拠に関する説明を求められるケースがある。例えば、金融分野においてローン審査にAIを適用する場合、個別の審査結果の判断根拠を説明できることが要求される。こういった要求に対して、AIの説明可能性を高めるための取り組みは、近年のAI関連のトレンドの一つでもあり、AT/PRCは二つのアプローチによりこの要求に応える。

一つは、解釈可能なモデルを構築するというものである。AT/PRCは多項式で表現されるモデルを構築する。これにより、分析者はモデルがどの説明変数をどの程度重視しているかを解釈することが可能となる。

もう一つは、「影響度算出技術」の活用である。本技術はAT/PRCのスコア算出値を説明変数ごとに分解し、どの説明変数の値がどれだけスコアを増減させたのか定量化するというものである(図2参照)。

AT/PRCのモデルは多項式で分析者が解釈可能だが、業務担当者にとっては理解が難しいため、影響度算出技術を使うことで業務観点での判断根拠の理解を容易にする。

図1 | シグナル&ノイズに基づく学習の概要

モデルは「住所」を貸し倒れ発生に影響する特徴量と捉えているが、学習データのグループごとに比較するとその値は共通ではなかった。このような場合に、「住所」の重みを下げることでモデルへの影響度を弱めている。

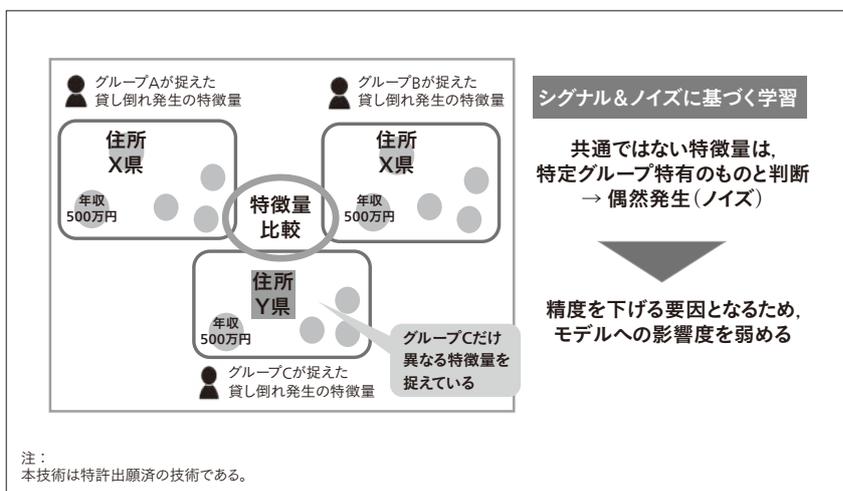
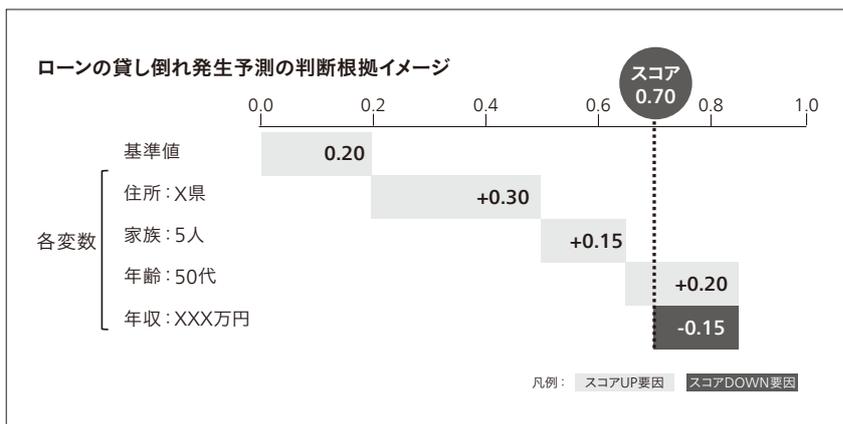


図2 | 影響度算出技術の活用イメージ

モデルはある顧客のローンの貸し倒れ発生リスクを0.7と予測している。影響度算出技術を活用することで、この顧客の居住地や家族構成、年齢がスコアを押し上げ、年収がスコアを押し下げていることが分かり、根拠の説明が可能となる。



### 3. 顧客業務への適用事例

日立はAT/PRCを用いて、国内外の顧客と協創に取り組んできた。その中から二つの事例を紹介する。

#### 3.1

##### 個人向けローンの審査業務改善事例

一つ目は、ベトナムの金融機関VietCredit社の個人向けローンサービスの与信審査をAT/PRCによって改善した事例である。近年、ベトナムでは人口の増加や都市化の急速な進展、中間層の拡大などを背景に消費者ローン市場が拡大し、無理な貸出による不良債権増加が課題となっており、消費者保護の観点で公正かつ精緻な審査システムが求められている。

日立がVietCredit社と2019年10月より実施した実証実験では、個人向けローン申し込み情報から将来的に返済不能となる確度を予測するスコアリングモデルをAT/PRCにより構築し、その検証と評価を行った。本実証実験の結果、AT/PRCのモデルが統計的手法を用いた従来モデルを上回る予測精度となり、借入後の返済不能発生リスクの低減が期待できる成果を得た。また、AT/PRCの予測影響度算出技術を使った判断根拠の提示結果には、審査業務を行っている顧客の経験と共通するものがあり、「納得感、安心感がある」として高い評価を得た。本実証実験はAT/PRCの日本国外初の実績であり、海外においてもAT/PRCの有効性を示すことができた。

本実証実験の成果により、AT/PRCは2021年3月より顧客のローン審査システムにおいて本格導入されている。実運用では、実証実験で分析対象としたローン商品に加えて、従来モデルでは予測が難しかった個人事業主

向けローン商品に対してもAT/PRCのモデルが活用された。AT/PRCと日立のデータサイエンティストが持つノウハウによって構築した本モデルは高い精度を実現し、VietCredit社のビジネス拡大に向けて貢献できた。

今後は、ベトナムでの実績を基に東南アジア諸国へのさらなる展開をめざしていく。

#### 3.2

##### ビジネスマッチング業務の業務効率化事例

二つ目は、金融機関におけるビジネスマッチング業務の効率化を試みた事例である。ビジネスマッチングとは、販売先・仕入先の紹介や業務提携パートナーの紹介など、各企業のニーズに合わせて企業どうしをつなげるサービスである。金融機関はマッチングを仲介する事業を行っており、相談を受けた企業へマッチングが見込まれる候補先企業の紹介を行っているが、候補先企業の選定が業務担当者の知識や経験に依存するという問題や、人手で選定可能な企業数は限られるという問題がある。

日立は、マッチング候補の選定にAIを活用し、企業選定の属人性排除と網羅性向上により、マッチングの成約が期待できる企業の組み合わせをレコメンドするサービスの実証実験を複数行ってきた。実証実験で利用可能なデータには、カテゴリ変数や数値変数といった構造化データ（企業の業種や地域、売上高など）だけでなく、業務担当者が企業からヒアリングした内容を自由入力で登録したテキストといった非構造化データ [企業／商品PR (Public Relations) や相手企業に求める条件など] もあった。非構造化データには、構造化データには含まれていない定性的な情報が含まれていたため、過去の実証実験では、自然言語処理のAIを活用し非構造化データの分析も行った。これにより、構造化データである企業情報や過去のマッチング成約実績に基づいてAT/PRCが学

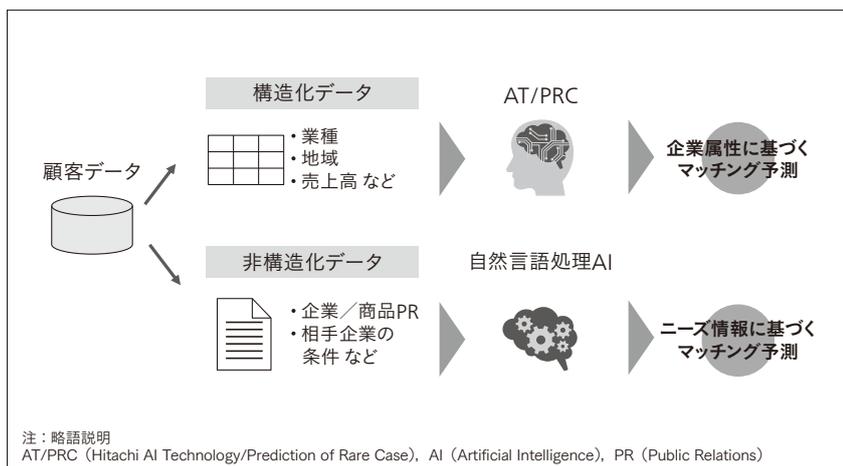


図3| 構造化データと非構造化データを使い分けたビジネスマッチング業務へのAI適用

AT/PRCと自然言語処理AIとで異なる形式のデータを入力とした。特性の異なるデータから得られた予測結果を比較・検証することで、よりの確かなマッチング候補先のレコメンドを行った。

習した成約企業の特徴に加え、非構造化データである業務担当者のヒアリングによって顕在化した企業のニーズ情報を分析対象とすることで、よりの確な候補先のレコメンドを行った（図3参照）。

## 4. 課題と今後の展望

前章ではAT/PRCの適用事例を紹介したが、企業が持つデータには日々生成・蓄積されるものの有効活用されていない、あるいは活用の手間がかかり活用効率が悪い「ダークデータ」が大量に存在する。ダークデータを適切に活用することができれば、AIによる業務改善の効果を高めることが期待できるが、難易度は高い。例えば、ドキュメントデータのうち請求書や診療明細書、有価証券報告書など非定型ドキュメントは、発行元ごとに表記や様式が異なるため、読み取り・抽出の自動化が難しいケースが多く、活用が難しい。日立は、ダークデータから効率的に価値のあるデータを抽出し分析可能にする「データ抽出ソリューション」を提供している<sup>1)</sup>。本ソリューションは現時点ではドキュメントデータを対象としているが、画像や映像、音声など他の形式のダークデータにも対応するソリューションの実現をめざしている。このデータ抽出ソリューションにAT/PRCのようなAIエンジンを組み合わせることで、AIを適用する業務領域の拡張やAIによる予測結果のさらなる精度向上が期待できる。

また、前章で紹介したVietCredit社との間では、タブレット端末上で申込者とオペレーションセンターをつなぐコミュニケーションツールを活用して受付業務を行うCACM (Compact Automated Contract Machine)をAT/PRCと組み合わせ、ローンの申し込み受付からAIによる即時審査をシームレスに行うソリューションの提供を検討している。

このように、今後はAIと他のデジタルソリューションとを組み合わせることで、顧客のビジネスプロセス全体をカバーするソリューションを構築していく。

## 5. おわりに

本稿では、AT/PRCの技術的特長と適用事例を紹介した。ここで紹介した事例は金融分野のもののみであるが、AT/PRCの技術は業種を問わず適用可能であり、さまざま

まなビジネスプロセスでの活用を検討している。日立が多様な業務への対応で培ってきた知識や知見にAT/PRCの技術を組み合わせることで、企業や社会の課題解決に貢献することをめざしていく。

### 参考文献など

- 1) 日立ニュースリリース、企業内の「ダークデータ」に着目した「データ抽出ソリューション」を提供開始（2021.6）、<https://www.hitachi.co.jp/New/cnews/month/2021/06/0623a.html>

### 執筆者紹介



#### 宇都 陽平

日立製作所 金融ビジネスユニット 金融システム営業統括本部  
事業企画本部 金融イノベーション推進センタ 所属  
現在、金融向けビッグデータ/AIに関わる企画業務に従事



#### 仲田 智将

日立製作所 金融ビジネスユニット 金融システム営業統括本部  
事業企画本部 所属  
現在、金融向けビッグデータ/AIに関わる企画業務に従事



#### 田代 大輔

日立製作所 金融ビジネスユニット 金融システム営業統括本部  
事業企画本部 金融イノベーション推進センタ 所属  
現在、金融向けビッグデータ/AIに関わる企画業務に従事



#### 赤崎 充洋

日立製作所 金融ビジネスユニット 金融第一システム事業部  
金融システム第一本部 第一部 所属  
現在、ベトナムでの金融向けソリューションの提案・推進業務に従事



#### 辻 琢矢

日立製作所 金融ビジネスユニット 金融第二システム事業部  
金融デジタルイノベーション本部 第三部 所属  
現在、金融向けビッグデータ/AIに関わるソリューションの提案・推進業務に従事