

# Applications and Future Prospects of AI for Scoring Models that Allow for Rare Cases

AT/PRC is a Hitachi AI that can accurately predict the probability of events that occur only rarely in actual business practice while also providing an explanation for the prediction. With corporate use of AI having accelerated in recent years, AT/PRC has demonstrated its worth at both Japanese and overseas customers. This article describes the technical features of AT/PRC and presents example applications from Japan and elsewhere. It also describes the issues that arise when deploying AT/PRC and other AIs in customer businesses, what Hitachi is doing to address these issues, and the outlook for the future.

Yohei Uto

Norimasa Nakata

Daisuke Tashiro

Mitsuhiro Akasaki

Takuya Tsuji

## 1. Introduction

The use of artificial intelligence (AI) in various corporate activities has been accelerating in recent years, with the research, development, and deployment of the technology advancing on a daily basis. Hitachi AI Technology/Prediction of Rare Case (AT/PRC) is one of the AIs developed by Hitachi. Its initial uses have included a demonstration project for loan approval (credit scoring) at a Japanese finance company and it has since proven its worth in collaborative creation projects with financial institutions overseas and demonstration projects targeting uses other than credit scoring.

This article describes the technical features of AT/PRC and then presents examples of its use to date. It also describes the issues that have arisen during the deployment of AT/PRC at customer businesses and what Hitachi is doing to address these.

## 2. Technical Features of AT/PRC

The following sections describe two technical features that characterize AT/PRC.

### 2.1

#### Avoiding Overfitting of Rare Cases

Problems with poor accuracy have arisen when AI learning has been applied to events that occur only rarely in practice, such as debtors unable to repay a loan, instances of fraud hidden among large numbers of authentic transactions, or equipment failures. This is due to a phenomenon called “overfitting.” Overfitting occurs when an AI model becomes too attuned to the training data used in its development, such that its accuracy deteriorates when it is applied to new data not included in its training. Existing techniques in common use, when utilized to identify the attributes of rare cases, have ended up learning on noise. To prevent this phenomenon from occurring, AT/PRC adopts a mechanism called signal- and noise-based learning. This incorporates the following steps into the learning process.

- (1) Training data is separated into a number of small groups.
- (2) Model parameters are updated by training on each data group.
- (3) The feature values that have an influence on whether an event occurs are determined for each group. The results are compared and any feature values that are not common to all groups are deemed to be noise that will have a detrimental effect on accuracy. These “noise” feature values are given a lower weighting.

The result of this final step is to diminish the relative weighting of feature values deemed to be noise (see **Figure 1**).

The adoption of this mechanism means that AT/PRC is able to avoid overfitting and the loss of model accuracy when exposed to new data even when trained to recognize rare cases.

## 2.2

### Two Functions for Improving Ability to Explain AI Decisions

It is sometimes necessary, when AI is used for important decisions or predictions, to provide the relevant stakeholders with an explanation as to why it produced the result it did. When AI is used to assess a loan application in a finance business, for example, it must be possible to explain the basis for each conclusion. Enhancing the ability to explain of AI decisions has been one of the trends in the field over recent years and AT/PRC has adopted two techniques that address this requirement.

The first is to devise model formulas that can be interpreted by humans. AT/PRC generates models that are expressed as multiple terms. This means the analyst is able to see how much weight the model places on each explanatory variable.

The second technique involves calculating degrees of influence. This works by analyzing each explanatory variable used in AT/PRC score calculations to quantify how much each one influences the result (see **Figure 2**).

While the multi-term models used by AT/PRC are amenable to interpretation by analysts, they are not so easy for users to understand. Calculating degrees of influence, on the other hand, provides an easier way to understand the basis of decisions in a workplace context.

## 3. Deployment at Customer Businesses

Hitachi has used AT/PRC in collaborative creation projects with customers in Japan and elsewhere. The following sections describe two such projects.

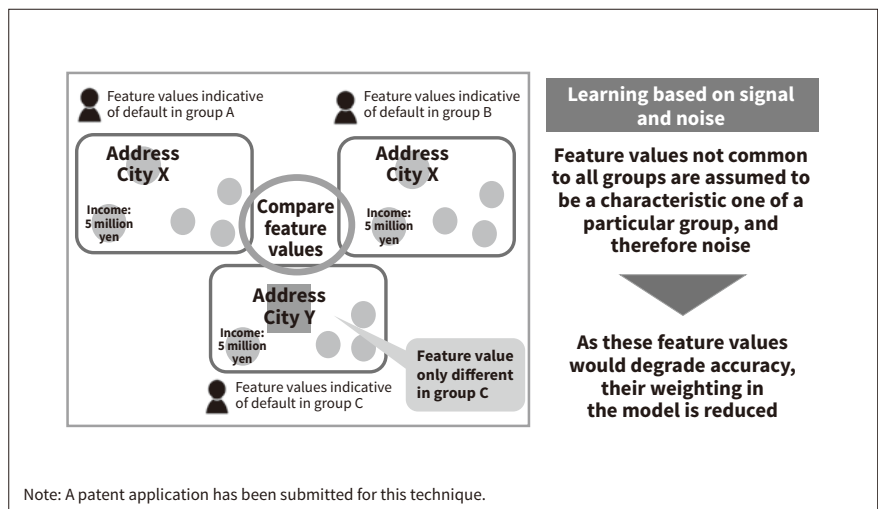
### 3.1

#### Process Improvement Project for Personal Loan Approval

The first example involved the use of AT/PRC to improve credit examination for personal loans offered by the VietCredit Finance Joint Stock Company in Vietnam. The

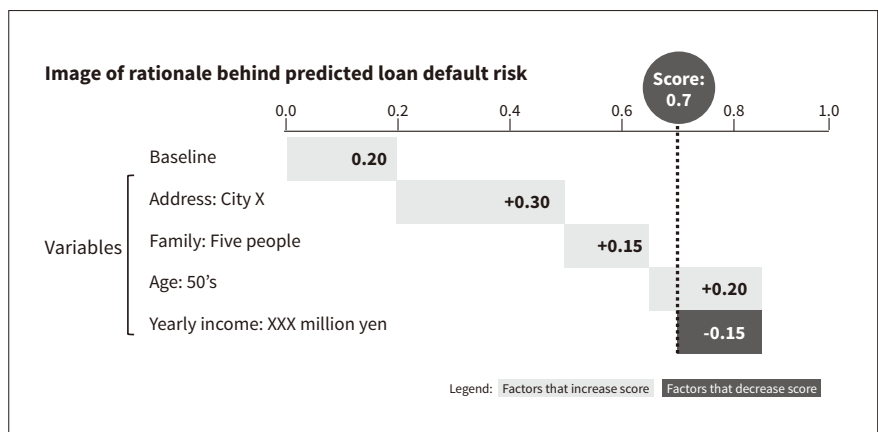
**Figure 1 — Overview of Learning Based on Signal and Noise**

While the model identifies the address as a feature value indicative of loan default, a comparison shows that it is not consistent across the different learning data groups. Accordingly, the weighting of address is reduced to minimize its influence on the model.



**Figure 2 — How Degrees of Influence Calculation is Used**

The model predicts a risk score of 0.7 for a particular customer defaulting on a loan. The degrees of influence calculation can explain the reasons for this result, indicating that the score was pushed up by the customer's address, family structure, and age, and pulled down by their income.



consumer loan market in Vietnam has grown over recent years against a background of rapid population growth and urbanization as well as an expanding middle class. As this has brought issues with bad debts due to inappropriate lending, there is a need for fair and fine-grained applicant screening systems from the viewpoint of consumer protection.

A demonstration project launched by Hitachi in October 2019 in partnership with VietCredit used AT/PRC to devise a scoring model for predicting the likelihood of a loan applicant being unable to repay based on the information provided in their application. The model was trialed in practice and its performance assessed. The results indicated that the AT/PRC model was more accurate than the existing model created using statistical methods and that its use should reduce the risk of borrowers becoming unable to repay their loans. Moreover, positive feedback was received when the basis of decisions produced by AT/PRC's degrees of influence calculation was presented. The finance company's loan screening staff found it to be convincing and inspiring of confidence, being in line with what they would have concluded from their own experience. This project was the first time AT/PRC had been used outside Japan, demonstrating that it can also be useful overseas.

Based on the results of the demonstration project, AT/PRC entered full commercial use in the finance company's loan screening system from March 2021. In addition to the loan categories covered by the demonstration project, this AT/PRC model was practically used on commercial loans to sole proprietors, a category where prediction had been difficult using the previous model. The model developed using AT/PRC in tandem with the expertise of Hitachi data scientists delivered a high level of accuracy and has been contributing to business growth at VietCredit.

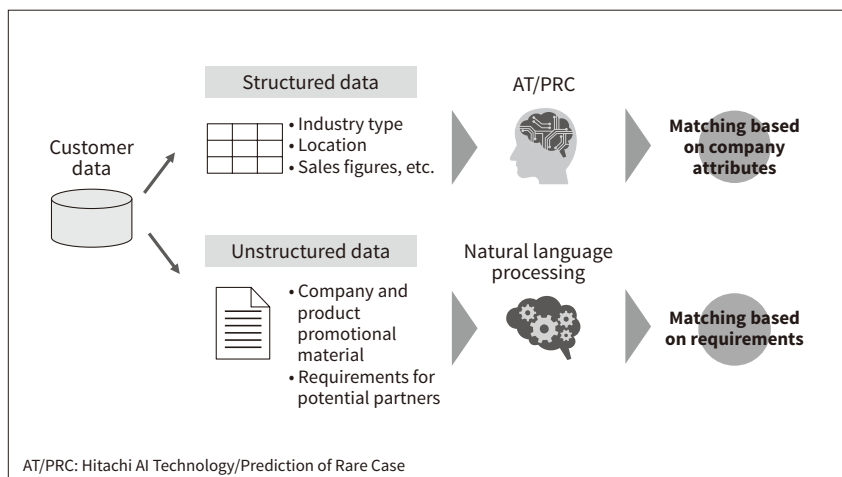
Drawing on this experience in Vietnam, Hitachi is now looking to deploy the technology more widely in other Southeast Asian nations.

### 3.2

#### Efficiency Improvement Project for Business Matching

The second example project was aimed at improving the efficiency of business matching at a financial institution. Business matching is a service for putting different businesses in touch with one another based on their particular needs, such as introducing potential business partners or new customers and suppliers. This involves the financial institution acting as an intermediary, introducing clients to other businesses that are a potential match. The problem is that the process is reliant on the knowledge and experience of the matchmaker and that this limits the number of candidate companies that can be identified manually.

Hitachi has conducted a number of demonstration projects for a service that recommends pairs of companies with the potential to become matching clients, using an AI for this purpose so as to consider a wider range of candidates and to make the process less dependent on the person doing the matching. These demonstration projects had access not only to structured data in the form of category and numeric variables (data on a company's industry sector, workplace locations, sales, and so on) but also unstructured data in the form of free-form text entered by the matchmaker based on consultations with the company concerned (such as promotional material about the company and its products and what it is looking for in a partner). As the unstructured data included qualitative information, not captured by structured data, natural language processing was also used to analyze the unstructured data in the past demonstration projects. Accordingly, recommendation candidates can be identified with greater accuracy by using not only the attributes of potential matching clients that AT/PRC has learned from structured data about companies and past matching arrangements, but by also augmenting the analysis with unstructured data indicative of what clients are really looked for that has been acquired by the matchmaker through the consultation process (see **Figure 3**).



**Figure 3 — Differing Uses of Structured and Unstructured Data in AI for Business Matching**

AT/PRC and the natural language processing take different input data formats. Crosschecking of the prediction results obtained using the different types of data provides more accurate matching recommendations.

## 4. Challenges and Future Prospects

The previous section described example applications of AT/PRC. However, the data held by companies also includes large quantities of “dark data,” meaning data that is routinely collected and stored without being used, or that is difficult to make use of efficiently. While better use of dark data has the potential to amplify the benefits of AI for business operation improvement, this is easier said than done. For example, it is often difficult to automate the scanning and interpretation of unstructured documents such as invoices, itemized medical bills, or securities reports because they all use different terminology and formats. Hitachi supplies a data extraction solution that can efficiently extract valuable information from such dark data<sup>(1)</sup>. While this solution currently only works on documents, Hitachi intends to develop solutions that can also work with dark data in other forms such as audio, video, or images. When combined with an AI engine such as AT/PRC, these data extraction solutions have the potential to expand the scope of application of AI and further improve its prediction accuracy.

Hitachi is also working with VietCredit, the company discussed earlier in this article, looking at how to deliver a solution that can provide seamless on-the-spot AI screening from loan applications by combining AT/PRC with compact automated contract machines (CACMs) that perform the reception service using tablet-computer-based communication tools for connecting applicants with the company’s operations center.

In this way, Hitachi intends to create solutions that combine AI with other digital solutions to cover all customer business processes.

## 5. Conclusions

This article has described the technical features of AT/PRC and presented example applications. While the applications described here have all been from the finance sector, the technology of AT/PRC is suitable for use in a range of different industries and Hitachi is looking at applying it to a variety of business processes. By combining AT/PRC with the knowledge and expertise it has built up from working in a diverse range of industries, Hitachi aims to contribute to the resolution of challenges facing both the corporate sector and wider society.

### Reference

- 1) Hitachi News Release, “Launch of ‘Data Extraction Solution’ Focusing on ‘Dark Data’ in the Enterprise” (Jun. 2021) in Japanese, <https://www.hitachi.co.jp/New/cnews/month/2021/06/0623a.html>

### Authors



#### Yohei Uto

Financial Innovation Center, Business Planning Unit, Financial Information Systems Sales Management Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research:* Big data/AI-related planning for financial services.



#### Norimasa Nakata

Business Planning Unit, Financial Information Systems Sales Management Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research:* Big data/AI-related planning for financial services.



#### Daisuke Tashiro

Financial Innovation Center, Business Planning Unit, Financial Information Systems Sales Management Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research:* Big data/AI-related planning for financial services.



#### Mitsuhiro Akasaki

Department 1, Financial Information Systems Business Unit 1, Financial Information Systems 1st Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research:* Proposal and promotion of solutions for financial services in Vietnam.



#### Takuya Tsuji

Department 3, Digital Innovation Business Unit, Financial Information Systems 2nd Division, Financial Institutions Business Unit, Hitachi, Ltd. *Current work and research:* Proposal and promotion of big data/AI related solutions for financial services.