

(別紙) 検証成果の詳細

①機械学習に必要となる学習データの品質向上

金融取引における正常／不正を識別する AI モデルを構築する際には、正常に取引が行われた学習データ（以下「負例データ」）と、不正が行われた学習データ（以下「正例データ」）が必要です。

しかし、不正の事例は実際にはほとんど存在しないことから、正例データは少量しか取得できません。そこで、この問題への対処法の一種として、機械学習のフローの前工程「データ取得」において、実際の正例データを基に「実在しないが確からしい」正例データを大量に生成する操作（オーバーサンプリング）を行うことがあります。（図 1）

本検証では、量子アニーリングは規則性のない数値を生み出せるという特性を活用することで統計的に確からしい正例データ生成器を開発（注 1）し、これを用いてオーバーサンプリングを行いました。加えて、不正取引を識別する AI モデル（主にデータの分類に用いられる決定木と呼ばれるモデル）の学習に、オーバーサンプリングしたデータを適用した結果、従来手法（「ランダム」および「SMOTE」）と比較して再現率（注 2）の向上に寄与することも確認しました（図 2）。

【検証概要】

実施期間：2020 年 9 月～12 月

検証環境：D-Wave Systems Inc. の量子コンピュータ等

利用データ：海外クレジットカード会社の実際の取引履歴から作られた公表データ

検証内容：学習データを使って不正取引を識別する AI モデルを構築し、オーバーサンプリングの手法による有用性を評価

オーバーサンプリングの各手法：

- ・量子アニーリングを用いた手法

正例データをボルツマン分布と呼ばれる確率分布で再現する手法。ボルツマン分布は確からしいデータを生成するために利用される。今回、正例データの生成時に量子アニーリングを適用している。

- ・ランダム（従来手法）

複数存在する正例データからいくつかの正例データをランダムに選び、それらをそのまま複製する形で新たなデータを生成し、正例データに加える手法

- ・SMOTE (Synthetic Minority Over-sampling TEchnique)（従来手法）

複数存在する正例データから類似した正例データを 2 件選び、それらの中間的な特徴を持つ新たなデータを生成し、正例データに加える手法

結果：不正取引の再現率を比較した結果、同一の AI モデルであっても学習データの品質相違から、本手法を用いた再現率がランダムより 6～15%、SMOTE より 3～6% 程度向上することを確認。

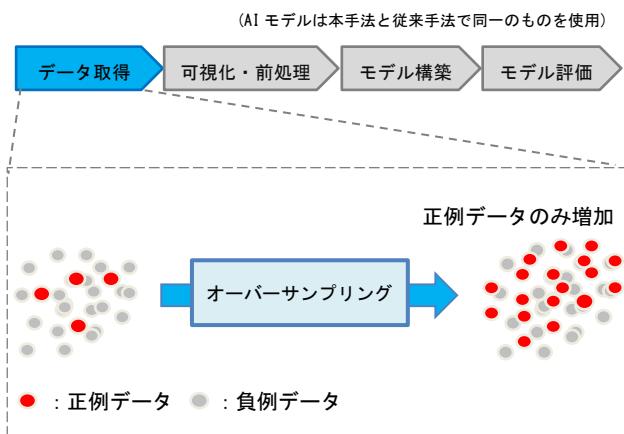


図 1：オーバーサンプリングのイメージ

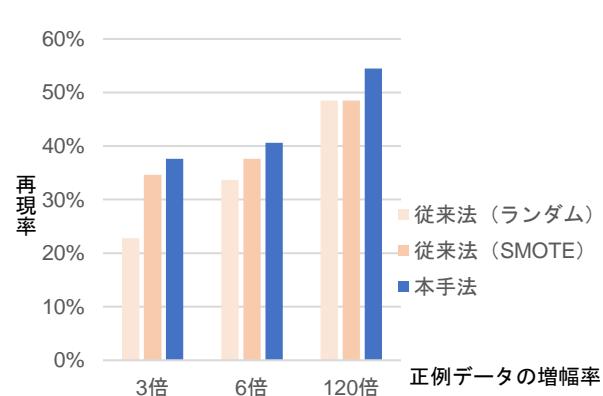


図 2：不正検知の再現率の向上効果

②ストレステストのシナリオ策定における経済指標の調整

SMBC グループの業務戦略策定に際しては、重要なリスクを織り込んだシナリオに基づいてストレステストを行い、資本の健全性を検証しています。

このシナリオに含まれる経済指標は数理モデルを活用したソフトウェアによって推計しますが、専門家等の人的判断による推定も勘案しています。そのため、推定した一部の指標とその他の関連指標が整合的になるように手作業で調整を行う必要があり、従来はこの作業に多くの時間が費やされていました。

本検証では、この調整を最適化問題として定式化し、量子アニーリングの手法を用いて解くことで、調整作業を効率化しました。その結果、実務で使用可能な精度の解を得るために必要な時間が、従来の方法に比べて約 6 分の 1 に短縮されることを確認しました。

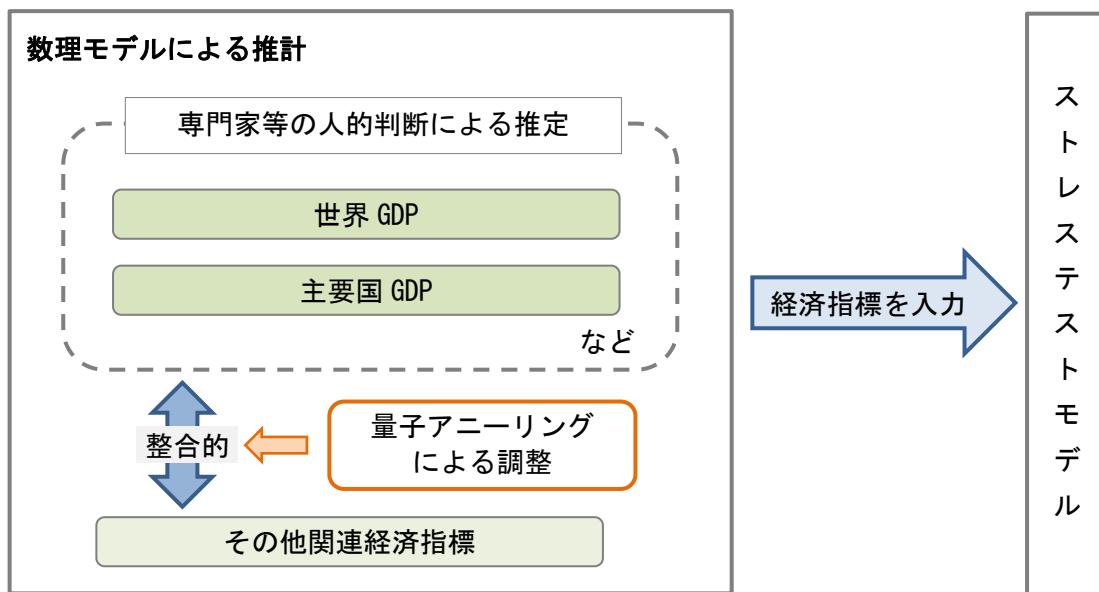


図 3 : 量子アニーリングによる調整イメージ

(注 1)

本手法を構成する技術の一部について、現在特許出願中

(注 2)

検出できた不正取引の割合。

例えば、再現率が 50%から 55%に向上すると、100 件の不正利用のうち 50 件を検知できていた AI モデルが 55 件検知できるようになる。

(参考)

NEC 量子コンピューティングへの取り組みについて

https://jpn.nec.com/quantum_annealing/index.html