

為替レート予測における機械学習システムのモニタリング手法の検討

A Monitoring Method of Machine Learning System in Exchange Rate Prediction

株式会社 日本総合研究所

The Japan Research Institute, Limited

○北野 健太 石川 冬樹¹⁾○Kenta Kitano Fuyuki Ishikawa¹⁾**Abstract**

Systems build with machine learning (ML) techniques expose different characteristics from traditional software systems. The behavior is not directly controllable in a sense it is derived from training data, not from rules human engineers designed. This characteristic makes it difficult to guarantee the quality of a component (trained model) built by ML. It is therefore significant to design the entire system architecture that monitors and modifies the output from the ML-based component. This study presents such an architecture and a monitoring method for prediction of exchange rates in planning financial operations. The proposed monitoring method observes the runtime situation and decide whether to rely on the prediction output or not.

1. はじめに

機械学習では、従来のシステム開発のようにプログラムを直接的に記述する方法とは異なり、訓練データから帰納的に振る舞いが生成される。そのため、できることとできないことの境界を明確に把握することが容易ではない。また、機械学習は 100%の精度を出すことができないため、誤った答えを出す可能性を踏まえた考慮が必要である。

このような特性により、機械学習のモデル単体での品質保証は難しく、機械学習システム全体としての観点から、機械学習モデルからの出力を運用時に監視や修正をするアーキテクチャ等についてはその必要性が言及されている。例えばドメイン知識として許されない範囲の出力を修正するといったことが考えられる[1]。本論文ではこれに対し、機械学習モデルの実証評価から「(不)得意な入力」に関する情報を得て、その情報を運用時に活用することを考えている。[2]では、正しい推論を行った入力と推論を誤った入力との特徴量差異を用いることで、効果的な再訓練を行うことを目指している。このように実証評価から得た「(不)得意な入力」については、理論的保証がないもののヒューリスティックとしてはうまくはたらく可能性がある。本論文においてもこのように実証評価から得た情報を活用している。また、機械学習モデルを適用可能な範囲である Applicability Domain を見積もるためには、一般的には、入力データと訓練データの関係性、例えば訓練データ集合からの距離などを基に、ある入力データに対するモデルの適合性を判断する[3]。すなわち、「学んだデータに何かの意味で近い」かどうかだけを見ている。本研究では、学んだデータのうち正解率が高い(あるいは低い)領域をより踏み込んでみている。これにより、問題が難しく、様々な入力データに対して一律に高い精度が出ないような場合でも、モデルの正答率が高いような適切な入力領域を判断することを目指している。

株式会社 日本総合研究所

The Japan Research Institute, Limited

東京都品川区東五反田 2-18-1 Tel: 03-6627-0417 e-mail:kitano.kenta@jri.co.jp

2-18-1, Higashi-Gotanda, Shinagawa-ku, Tokyo Japan

1)国立情報学研究所 National Institute of Informatics

【キーワード：】 機械学習, モニタリング手法, 品質保証, システム運用

本研究は、外国為替カバー取引という特定の金融業務を題材に、過去の為替レートを元に将来のレートを予測する機械学習モデルを組み込んだシステムのモニタリングの手法を検討し、実験によりその効果を確認したものである。学習時の結果を基に得意・不得意な入力データの基準を作成し、本番時にはその基準に従い得意な入力データを選別し、後続の処理では選別された得意な入力データの推論結果のみ利用することで、システム全体の品質向上を目指している。

2. 実施概要

2.1 監視システムの概要

①学習時の結果を分析することで、得られた機械学習モデルの得意な入力データを特定し監視基準を作成する。②次に本番時には、入力データをモニタリングし、監視基準に従い得意な入力データを選別する。③為替取引を行う後続処理では、正解する期待値が高いと考えられる得意な入力データのみ処理を限定する。このような仕組み(図1)により、得意と考えられるデータに取引を限定することで収益性が向上し、不得意な入力データを処理対象外とすることで損失リスクを低減できることから、システム全体の品質向上に寄与することが期待できる。

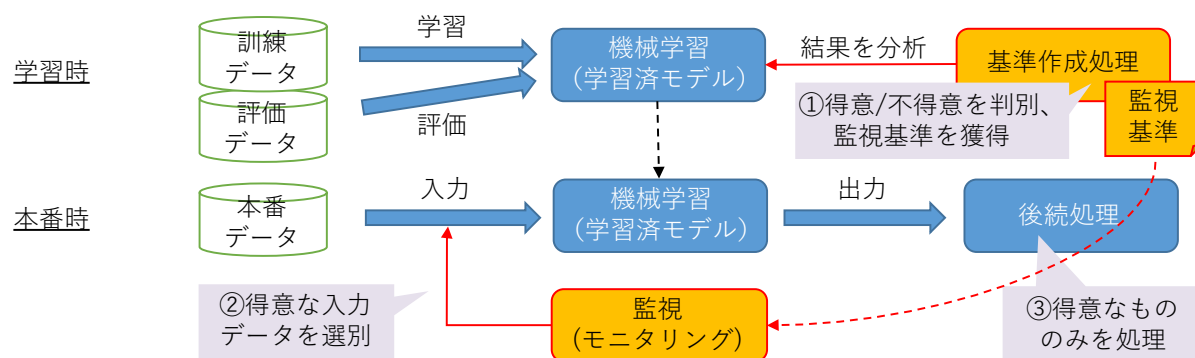


図1 モニタリングを含んだシステム構成

2.2 想定する金融業務 (外国為替カバー取引)

外国為替のカバー取引とは、端的に言えば、金融機関において、一般の顧客からの売り買い注文に相当する外貨をインターバンクより調達（または売却）する業務である。レートの動きを予想し、調達が有利なレートで実行することで利益を得られる可能性がある。例えば、インターバンク市場より、1ドル100円で調達した外貨を顧客に101円で販売できれば利益が得られる。取引においては、買値と売値が存在し通常買値の方が売値よりも高い。インターバンク市場での為替取引のレートをインターバンクレートといい、対顧客相場のレートをカスタマーレートという。通常、カスタマーレートはインターバンクレートを加味した金融機関にとって有利なレートが設定され、また買値と売値の差（スプレッドという）も金融機関の利益となることから、基本的には即時カバー取引を行うことでリスク回避を優先し、大きく勝てる可能性がある場合のみ将来のレート変動を加味した取引を行えば良い。その為、勝てる可能性が大きいパターンが判別できれば業務上有用である。

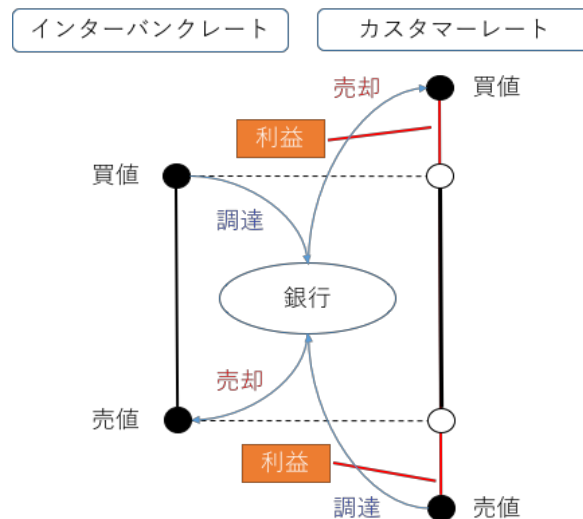


図2 カバー取引における利益

2.3 機械学習モデルの説明

過去 60 秒間の為替の値動きデータを入力とし、5 秒後の為替が上がるか (円安), 下がるか (円高) を予測する機械学習モデルを構築した (図 3). モデルは LSTM (Long Short Term Memory) を活用している. LSTM は, 時系列性の有するデータを学習できるニューラルネットワークであるリカレントニューラルネットワーク (RNN) の一種である. 入力である 60 秒間のレートを時系列データとして扱う. ライブラリは Keras を用いて, 平均二乗誤差を最小化する回帰問題として実装した. 解釈を容易にするため, レート以外のテクニカル指標やニュース記事を使ったセンチメント分析などは用いておらず, 為替の値動きのデータのみで予測するシンプルなモデルとしている. データは, オープンデータより「2016 年 1 月 4 日から 2 月 3 日 (1 ヶ月間)」の「ドル/円」通貨ペア (1 秒足) を取得し, 5 秒後との値動きの差に応じた正解ラベルを付与して利用している.

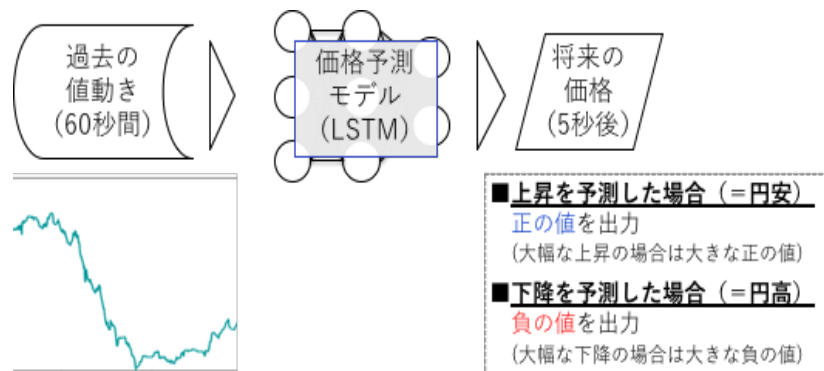


図3 モデル概略

3. 実施結果

3.1 監視基準の設計

得意/不得意の差異が検出可能か, それを識別する監視基準が設計できるか否か, が重要である. 予測の正解/不正解に対応する 60 秒間の入力データの特徴量を分析し活用することで監視基準の作成を試みた. 60 秒間の入力データの①分散値, ②平均値, ③最大値と最小値の差, の 3 パターンで実験している (図 4).

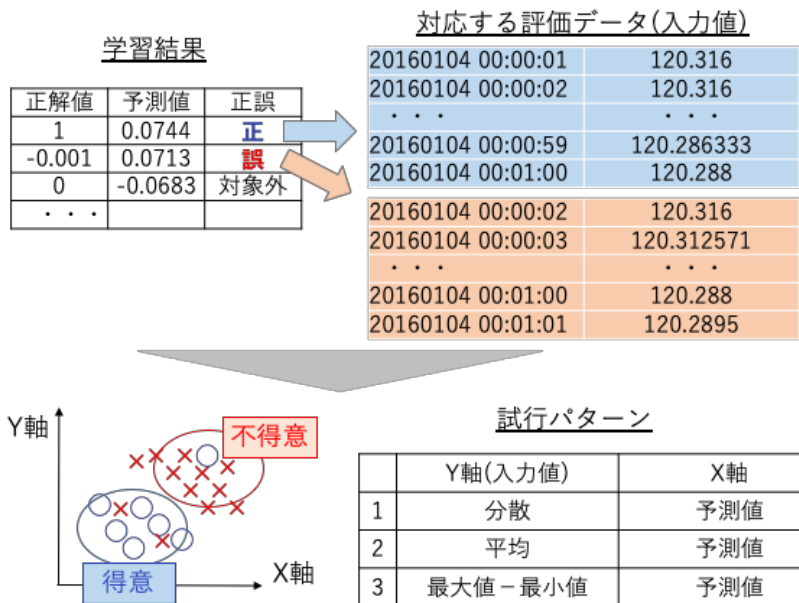


図4 得意・不得意判別への取り組みイメージ

X軸に予測値、Y軸に3つの特徴量をそれぞれプロットし、正解/不正解の偏りがあるか探索した。平均値、最大値と最小値の差の特徴量を用いたケースにおいては、優位な差が見られなかった一方で、表1の通り、分散値の場合は正解した182,621個のデータの平均値および標準偏差ともに不正解のデータより大きく、分散が大きいほど正解する傾向を確認した。分散が大きいデータというのは、60秒間の上昇または下降の変化が大きいものであることから、値動きが大きいものほど予測が成功しやすいモデルであるとの仮説を得た。(図5)

表1 分散値の統計量

統計量(60秒間の分散)	正解	不正解
データ数	182,621	181,524
最大値	0.48582	0.48548
最小値	0.0	0.0
平均値	0.00022	0.00017
標準偏差 (SD)	0.00501	0.00307

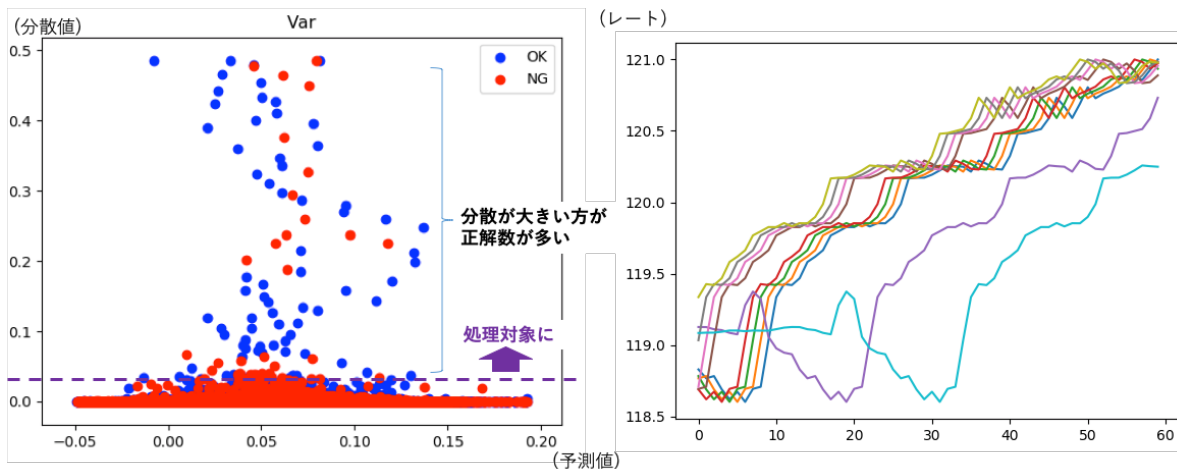


図5 散布図(左)、分散値が大きい入力データの例(右)

表1の正解データの統計量より、平均値と標準偏差を用いて監視基準を①平均(0.00022)、②平均+SD(0.00521)、③平均+SD*2(0.01024)の3パターン設計(表2)し、次節に示すそれぞれの基準を満たす、60秒間の入力データの分散値が大きいものに限定して取引を行う投資シミュレーションを実施しその有用性を確認する。

表2 フィルタリング基準

パターン	特徴量	フィルターしきい値
①	平均値	0.00022
②	平均値+SD	0.00521
③	平均値+SD*2	0.01024

3.2 投資シミュレーション

外国為替のカバー取引を想定した投資シミュレーションの実験を行う。評価指標として、正解率と収益性を用いる。正解率は5秒後に値動きがないものを成否判定の対象から除外し、価格変化の方向を当てることができた割合とする。

その他、投資シミュレーションの前提条件は、以下の通りとした。

- ・1回の取引量は1万ドル
- ・毎秒カバー取引が1回発生。ただし、基準により処理対象外の入力データの場合は発生しない(つまり、処理残は発生しない)
- ・インターバンクレートとカスタマーレートは同レート
- ・予測に応じて、利益が上がるオペレーションを実施(つまり、上がる(円安)予測であれば、売りオペ。下がる(円高)予測であれば買いオペ実施)

モニタリングせず、モデルの予測に従い必ずカバー取引を5秒後に実行する戦略ケース1と、モニタリングを適用し3種類の閾値(フィルタリングの基準)に従い、該当のデータのみ処理するものの計4パターンで比較することとした。検証期間は、期間1(2016年2月4日から2月11日)、期間2(2016年2月12日から2月19日)、期間3(2016年2月22日から2月29日)の1週間ごとに3週分のデータを投入して検証した。フィルタリング基準の選定に用いた評価期間および各検証期間の値動きは、上昇基調や下降基調など様々なトレンドを含むものとなっている。シミュレーションの結果をみると、評価期間ごと取引数のばらつきはあるものの、ケース2に関しては、フィルタリングを実施しないケース1よりも正解率、利益ともに悪化するという結果になったが、より分散が大きい入力データに取引を限定したケース3および4については、正解率、収益性とも良い結果を得られた。特に、ケース4においては、フィルタリングを実施することで、処理回数は大幅に減るものの、正解率が10%以上向上し、1回あたりの収益も100倍以上になっている。モニタリングにより得意と判断した入力データのみ処理することで、正解率と収益ともに向上することを確認できた。(表3)

表3 投資シミュレーションの結果

ケース	取引数	正解数	収益合計(円)
1. モニタリング無	1,107,433	50.04%	-123,793
2. モニタリング有(①)	124,692	49.90%	-219,449
3. モニタリング有(②)	930	51.30%	38,633
4. モニタリング有(③)	304	61.50%	39,503

4. 結論

機械学習システム全体の品質保証を目指した最初の段階として、入力データの得意/不得意の識別に取り組み、得意と思われる入力データに絞って為替取引をすることで正解率および収益が向上すること確認できた。今後、実験量を増やして検証が必要であるが、今回の投資シミュレーションで設定した、 $\text{平均値} + \text{SD} * 2$ という基準があらゆるケースに適用可能な汎用性を備えているとは考えにくい。 $\text{平均値} + \text{SD} * 2$ も一つの選択肢となりうるが、他にも基準となりうる候補をいくつか提示することが可能であって、その可能性のある候補の中から実験的に選定することが妥当と考えている。そうした候補となりうる基準を複数提示することや、その中からどのように選択するかなどをより自動的機械的に行えるようにすることが今後の課題である。また将来的には、入力だけでなく本番時の出力も踏まえたモニタリング手法の検討も進めたい。

5. 謝辞

本論文は、国立情報学研究所の教育プログラム「トップエスイー」の2018年度アドバンス・トップエスイーコースにおいて行った研究をまとめたものです。本研究の機会、環境を与えて下さった関係者の皆様に感謝申し上げます。

6. 参考文献

- [1] 丸山 宏, 機械学習応用システムにおける不変量の保証について, 日本ソフトウェア科学会第35回大会(2018年度)講演論文集
- [2] Shiqing Ma, Yingqi Liu, Wen-Chuan Lee, Xiangyu Zhang, Ananth Grama, MODE: Automated Neural Network Model Debugging via State Differential Analysis and Input Selection, The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE 2018), pp.175-186, 2018
- [3] Joanna Jaworska, Nina Nikolova-Jeliazkova and Tom Aldenberg: "QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review", ATLA 33(5):445-59 · November 2005
- [4] 石川冬樹, 徳本 晋, 機械学習応用システムのテストと検証, 情報処理学会誌 Vol.60 No.1 pp25-33, 2019
- [5] 丸山宏, 城戸隆, 「機械学習工学へのいざない」, 人工知能学会誌, 33 巻 2 号, 2018 年 3 月
- [6] 科学技術振興機構 研究開発戦略センター「AI 応用システムの安全性・信頼性を確保する新世代ソフトウェア工学の確立」, 2018
- [7] David Gunning: "Explainable Artificial Intelligence(XAI)", DARPA Program Update, November 2017.
- [8] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Computation, Vol. 9, No. 8, pp. 1735-1780 (1997)