

# AI を用いた時系列予測における 苦手パターン自動追加学習ツール開発

赤銅 知沙

## 1. まえがき

近年、画像の分類や時系列データによる予測など、様々な分野でAI (Artificial Intelligence, 人工知能) の研究が活発である。また、AI 技術の発展により、作業現場へのAI 導入による省人化・作業効率化の期待が高まっている。筆者は、これらの社会的な期待に対応すべく、入社以来、AI の活用や最適化を行う開発に携わっている。

AI の精度向上には、繰り返し追加学習を行う必要があり時間や手間がかかる。そのため、今回AI に学習させる作業を自動化するツールを作成し、改善効果が確認できた。

本論文では、作成したツールの概要とその中で用いた手法、改善の効果について報告する。

## 2. ツールの開発

### 2.1 AI の学習

AI の学習では、学習モデルに学習データを入力し、学習モデルから出力された予測値が期待値に近づくように学習パラメーターを調整することで、学習を行い、精度を向上させる。

学習後に学習モデルの性能を評価するには、評価データを学習モデルに入力して得られた予測値と、評価データに対する期待値を比較することで性能を評価する。評価データは、学習に使用していない、かつ学習データと同じフォーマットであるデータを用いる。

AI を現場適用する際には、あらかじめ収集した学習データで学習した学習モデルを使うことになるが、実際に運用すると、AI 作成時には想定されなかった未学習の入力データによる誤判定が起こる場合がある。このような場合、AI の精度を上げるためには、誤判定した評価データ（以降、苦手パターンと定義する）を次回評価時には誤判定させないように追加学習させる。

この作業を人の手で行う場合、誤判定がなくなるまで、評価が終わるごとに出力された予測値を確認し、苦手パターンを分析し、追加学習データを作成・学習する作業が必要となる。これらの作業を行うのは非常に手間であり時間がかかる。

### 2.2 目的

学習モデルを評価する場合、評価データを学習モデルに入力して得られた予測値が期待値に対して、所望の誤差範囲内であるかを確認する作業が必要となる。

誤差が許容範囲外であった場合、追加学習を行う。評価データのうち、誤差が許容範囲外となる評価データ（苦手パターン）を抽出し、それに類似した追加学習データを作成する。

今回の作成においては、使用する入力データの特性を考慮して、時間方向へのシフト量やノイズ付加レベルなどを決定し、データ拡張することでデータ生成を行い追加学習データに用いる。

以上の作業を1回の学習が終わる度に行う必要があり、終了条件を満たすまで繰り返し行うため、確認する評価データ数が多い、あるいは生成するデータ数が多い場合は、追加学習の作業に膨大な時間を要する。

このため、学習改善作業の効率化を図ることを目的として、次章で述べる自動化ツールを作成した。

### 2.3 自動追加学習ツール

学習モデルの性能改善作業を繰り返し行う、自動追加学習ツールを作成した。ツールの概要を図1に示す。

自動追加学習ツールは、終了条件を満たすまで以下の

(1)～(4)を繰り返す。

- (1) 評価
- (2) 苦手パターンの判定
- (3) 追加学習データの生成
- (4) 追加学習

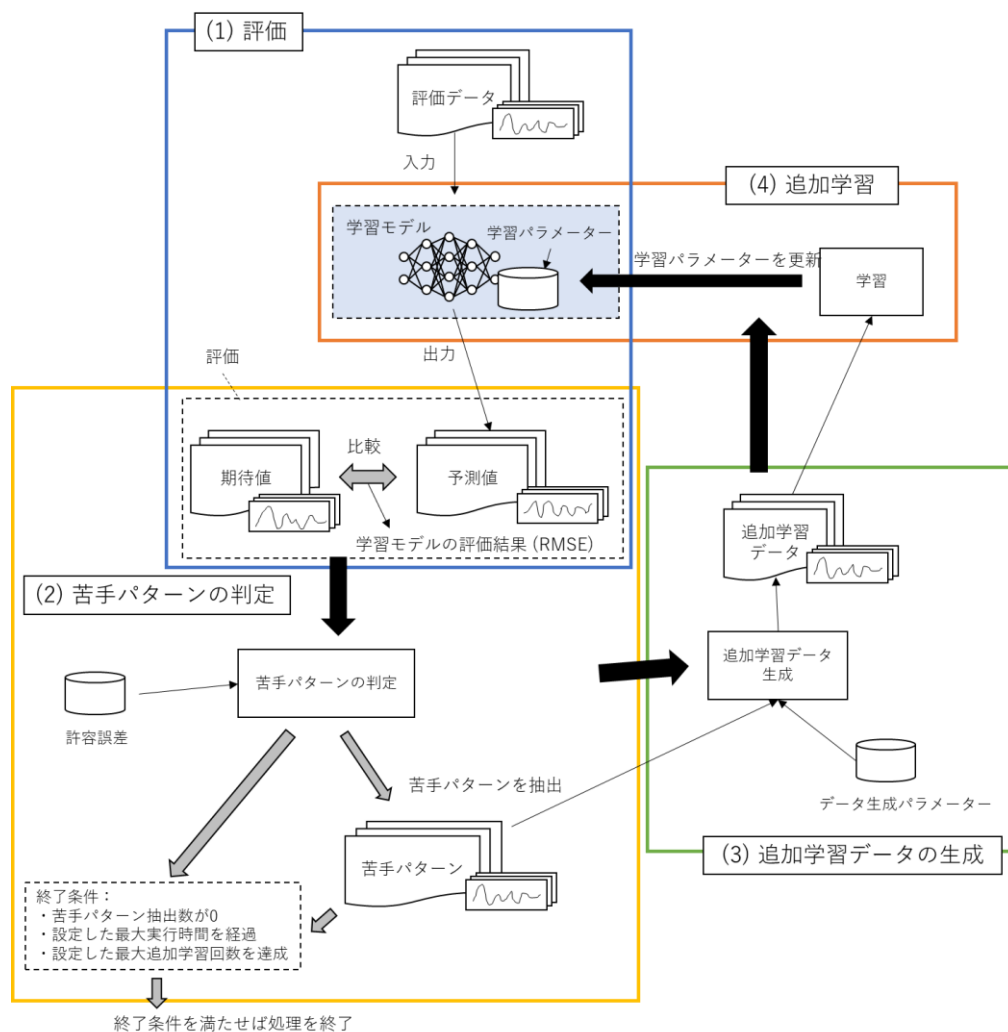


図1. 自動追加学習ツールの概要

### 2.3.1 評価

評価では、評価データを学習モデルに入力し、得られた予測値と評価データに対する期待値を比較して学習モデルの性能を評価する。

今回対象とする学習モデルは、時系列データ（複数）を入力するとその応答となる時系列データ（1つ）を予測する。初期の学習モデルは、シミュレーションで生成したデータで学習をした。

評価データ及び期待値は、対象となるシステム（実機）のセンサー出力値を測定して得られたデータを用いた。これには、物理シミュレーションデータで学習した学習モデルを未学習の実機データでチューニングするねらいがある。評価データには1ファイル300サンプル前後の波形データを32ファイル用意した。

学習モデルの性能評価には、RMSE（平均平方二乗誤差, root-mean-square error）を使用する。RMSEが0に近いほど、正解と予測の誤差が小さくなり、精度が良いと評価する。

### 2.3.2 苦手パターンの判定

苦手パターンの判定では、期待値と予測値の比較を行い、苦手パターンの抽出を行う。

具体的には、時系列の評価データのサンプル中において、期待値と予測値の差が許容誤差以上となる場合に、評価データを苦手パターンとして抽出する。

判定後に以下のいずれかの条件を満たす場合には、処理を中断し追加学習を終了する。

- (1) 苦手パターン抽出数が0
- (2) 設定した最大実行時間を経過
- (3) 設定した最大追加学習回数を達成

最大実行時間及び最大追加学習回数はツールで設定する。

### 2.3.3 追加学習データの生成

学習データ数に比べて抽出した苦手パターンが非常に少ない場合、これを追加学習データとして学習しても、精度の改善は期待できない。そこで、追加学習データの生成

では、抽出した苦手パターンをデータ拡張し、追加学習データを生成する。データの拡張方法は、想定される範囲内で生成する必要があるため、扱うデータの特性によって検討しなければならない。

今回は、測定した時系列データの特性から、想定される微小ノイズを付加し、時間方向の摂動を与えて、追加学習データの苦手パターンに類似するバリエーションを増やした。

追加学習データ生成パラメーターは、以下とした。

- (1) ノイズ生成パラメーター（平均，分散）
- (2) 時間シフト量（最小値，最大値）

### 2.3.4 追加学習

追加学習では、生成した追加学習データのみで学習モデルの学習を行い、学習パラメーターを更新する。

学習方法として、例えば特徴を抽出する箇所の学習パラメーターなど一部を固定し、特徴抽出された出力を用いて最終判定を行う学習パラメーターを更新させることもあるが、今回は学習可能な全ての学習パラメーターを更新させた。

## 2.4 開発環境

ツールの開発環境には、MATLAB®を用いた。

MATLAB®は、行列計算、機械学習、ディープラーニング及びGUI作成などの様々なライブラリーを有しており、C言語などの従来のプログラミング言語よりも短時間で簡単に科学技術計算を行うことができる。また、CコードやNVIDIA GPUのCUDAコードに変換ができ、将来は生成した学習モデルをCコードで組み込むことを想定しているため、MATLAB®を使用した。

## 3. ツールの評価

自動追加学習ツールを実行して、実行結果や作業時間を評価する。評価する学習モデル及び評価データについては2.3.1に記載している。

### 3.1 パラメーターの設定

表1に、設定したパラメーターを示す。

表1. 設定パラメーター一覧

パラメーター名	値	
最大実行時間	20 h	
最大追加学習回数	最大実行時間を優先するため、設定値なし	
閾値（許容誤差）	公称値の2%	
ノイズ生成パラメーター	平均	0
	分散	1
時間シフト量	最小値	1 サンプル
	最大値	20 サンプル

### 3.2 実行結果

ツールを実行した結果を示す。

2.3.2記載の終了条件の苦手パターン抽出数が0を満たすまでに実行した追加学習回数は48回であり、その所要時間は約8hであった。作業時間の内訳は、データ生成及び評価に2h、学習に6hであった。

図2に、苦手パターン抽出数の増減のグラフを示す。横軸は評価の試行番号、縦軸は苦手パターン数を示す。追加学習の効果により、苦手パターン抽出数が減少していることが分かる。

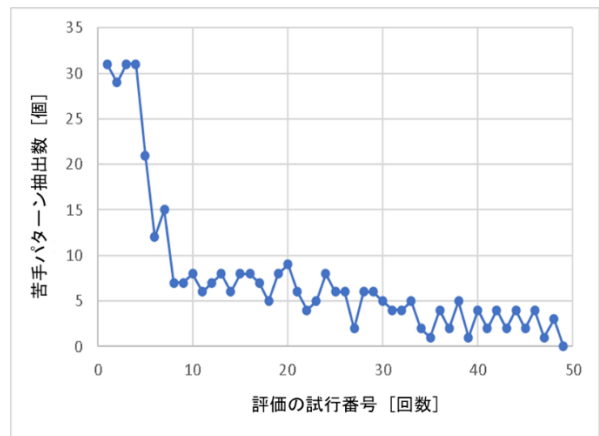


図2. 苦手パターン抽出数の増減

図3、図4に、評価データ2つの追加学習前後の推定結果を示す。

横軸はサンプル（時間）、縦軸はセンサーの値を示す。また、青線が期待値、オレンジの線が予測値を示す。図の黒いサンプル（黒い×点）は、期待値と予測値の誤差が公称値の2%（許容誤差）以上の箇所を表す。

追加学習前後で、誤判定していた期待値が許容誤差以内で推定できるようになっていることが分かる。

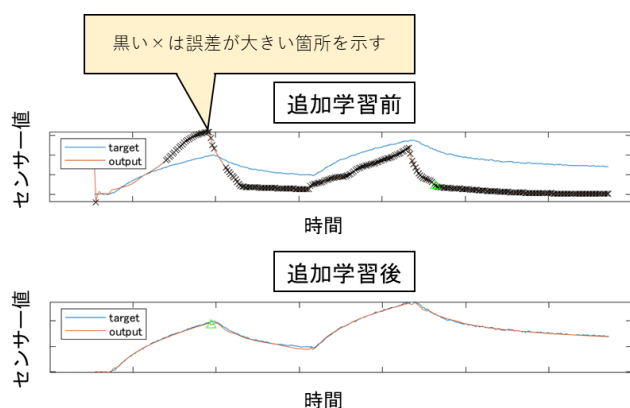


図3. 学習結果1

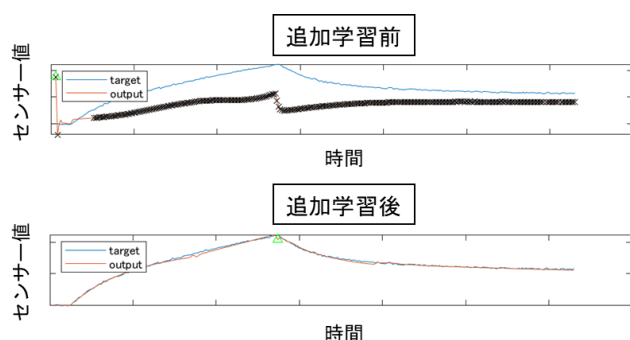


図4. 学習結果2

また、学習モデルの評価結果として、図5に、RMSE (Root Mean Square Error:期待値と予測値との差) の推移を示す。横軸は評価の試行番号、縦軸はRMSEを示す。

苦手パターン抽出数が減少するにつれ、RMSE が小さくなり、学習モデルの精度が改善していることが分かる。

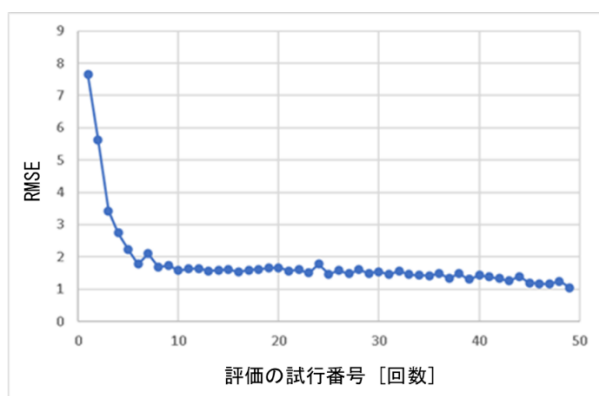


図5. RMSE の推移

また、図6に、ツール使用による時間削減効果を示す。

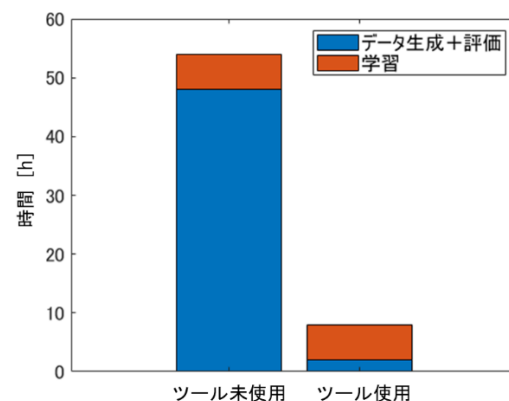


図6. 時間削減効果

ツールを使用しない場合は、1回の試行で、手作業で推定結果の評価を行い、苦手パターンを抽出し、追加データ生成をするため、約1hかかる。今回の学習条件の場合、目標精度達成には48回追加学習が必要であり、手作業〔データ生成+評価〕時間(1h×48回) + 学習時間6h (=1/8h×48回)だと約54hかかるが、自動学習ツール導入により約8h (= [データ生成+評価] 2h + 学習時間6h)で達成でき、作業の効率化が得られた。

手作業で行う場合、追加学習は1日2,3回が限度であるが、自動で追加学習ができると、目標精度まで十分学習を行うことができる。

#### 4. むすび

今回、自動追加学習ツールの導入により、追加学習作業の効率化が実現できた。また、ツールの開発を通じて、苦手パターンの判定方法やデータ拡張方法を検討し、時系列データを扱った学習モデルの精度改善を行う際のフローを確立することができた。

ツールを使用することで、実験室環境で対応できなかったパターンのデータを、現場で苦手パターンとして抽出し、自動で追加学習を行い、実機チューニングすることができた。

今回作成した自動学習ツールでは、時系列データを学習の入力パラメーターとしたが、AIの学習では画像も入力パラメーターとしてよく使用される。

自動学習ツールに画像データの入力を拡張することで、対応できる学習モデルの種類が増え、より精度の高い判定が可能なAIの実現が可能となることから、今後の学習モデルの性能改善に活用していく。

## 謝 辞

本開発遂行にあたり、ご指導賜った三菱電機（株）先端技術総合研究所情報制御プラットフォーム技術部モデルベースシステム技術グループの皆様に、深く感謝の意を表す。

## 商 標

- (1) MATLAB は、The MathWorks, Inc. の登録商標である。
- (2) NVIDIA, CUDA は米国及びその他の国における NVIDIA Corporation の商標又は登録商標である。

## 参考文献

- (1) 巢籠悠輔：詳解ディープラーニング TensorFlow・Keras による時系列データ処理, マイナビ出版(2017)

---

## 執筆者紹介

**赤銅 知沙** (せきどう ちさ)

2017年入社。

AI 開発に関するソフトウェア開発に従事。

現在、第3事業部制御応用技術部制御応用第3課に所属。