

サロゲートモデル構築プロセス検討結果

作成日時：2023 年 11 月 28 日

作成者：JMAAB PMWG

更新履歴

NO	内容	日付	変更者
1	初版発行	2023/11/28	PMWG
2			
3			
4			

目次

更新履歴	2
1. はじめに	5
1.1 背景・目的	5
1.2 内容	5
1.3 対象ユーザー	5
1.4 記載範囲	5
1.5 バージョン&ツールボックス	5
1.6 参照ドキュメント	6
1.7 定義（用語、略語）	6
2. サロゲートモデルとは	8
2.1 目的・用途	8
2.2 構築プロセス概要	10
3. 構築プロセス詳細	12
3.1 構築プロセス詳細一覧	12
3.2 学習用データ取得	13
3.2.1 詳細試験計画	14
3.2.2 シミュレーション	15
3.3 前処理	16
3.3.1 データ特性可視化	18
3.3.2 特徴量検討	19
3.3.3 欠損・外れ・異常値処理	20
3.3.4 フィルタリング	21
3.3.5 スケーリング	22
3.4 学習	23
3.4.1 学習用データ分割	24
3.4.2 アルゴリズム選定	25
3.4.3 ハイパーパラメータ設定	26
3.4.4 学習	27
3.5 合否判定	28
3.5.1 推定器の性能評価	29
3.6 シミュレーション	30
3.6.1 Simulink モデルの作成	31
3.6.2 推定	32
4. まとめ	33

1. はじめに

1.1 背景・目的

ILSVR という大規模画像認識コンペティションが 2012 年に開催されて以来、深層学習技術は注目を集め、画像認識だけでなく様々な分野で研究されている。特に、数値シミュレーションの分野では、計算コストの高い CAE 領域で深層学習が活用されている。

今回 PMWG では、プラントモデルの高速化に向けて、深層学習を活用するためのプロセスを整理した。具体的には、プラントモデルをサロゲートモデル化するための手順を提示した。

1.2 内容

深層学習技術を用いて、プラントモデルをサロゲートモデル化するためのプロセスと、各プロセスにおける要点について記載する。

1.3 対象ユーザー

- ・ MATLAB を用いてプラントモデルを作成した経験があるモデル開発者
- ・ 深層学習を用いたモデル作成の初学者

1.4 記載範囲

縮退化手法には、MAP、統計モデル、伝達関数モデル、機械学習モデル、深層学習モデルなどがある。その中でも本書では深層学習、特に MATLAB Deep Learning Toolbox の Stateful Predict ブロックを用いたサロゲートモデル化プロセスについて示す。なお、実機データではなく、物理モデル（シミュレーション）に限定する。

1.5 バージョン&ツールボックス

本検討で利用した MATLAB のバージョンと、各構築プロセス利用可能な推奨ツールボックスを整理する。

- ・ MATLAB R2022b
 - MATLAB
 - Simulink
 - Control System Toolbox
 - Curve Fitting Toolbox
 - DSP System Toolbox
 - Deep Learning Toolbox
 - Embedded Coder
 - GPU Coder

- Global Optimization Toolbox
- Model-Based Calibration Toolbox
- Optimization Toolbox
- Parallel Computing Toolbox
- Signal Processing Toolbox
- Simscape
- Simscape Battery
- Simscape Driveline
- Simscape Electrical
- Simscape Fluids
- Simscape Multibody
- Simulink Coder
- Statistics and Machine Learning Toolbox
- Symbolic Math Toolbox
- System Identification Toolbox
- Wavelet Toolbox

1.6 参照ドキュメント

[1] Simscape による プラントモデリング入門 JMAAB プラントモデルワークショップ
作成 (2017 年 6 月) <https://jmaab.jp/>

[2] Mathworks 機械学習の Q&A : モデル検証のすべて
<https://jp.mathworks.com/campaigns/offers/next/all-about-model-validation.html>

1.7 定義 (用語、略語)

用語・略語	正式表記	解説
縮退化	—	コンピューター モデルの計算量やストレージ要件を軽減しつつ、期待される忠実度を制御された誤差の範囲内で維持する手法
ROM	Reduced Order Model	縮退化と同じ意味で使用される
機械学習	—	人間が設定した特徴量をもとに、大量のデータからコンピューターがルールやパターンを学習する技術
深層学習	—	大量のデータからコンピューターが特徴量も探しながら、ルールやパターンを学習する技術

サロゲートモデル	—	深層学習等の機械学習技術を活用して詳細プラントモデルを縮退化し、シミュレーション時間とデータ容量を大幅に削減したモデル。
推定器	—	機械学習アルゴリズムを用いて学習させた結果のモデル。
リサンプリング	—	ある時系列でサンプリングされたデータを別の時系列でサンプリングされたデータに変換すること。
アップサンプリング	—	リサンプリングのうち、サンプリング周期を上げる変換を指す。
ダウンサンプリング	—	リサンプリングのうち、サンプリング周期を下げる変換を指す。
FNN	Feedforward Neural Network	ネットワークにループ構造が含まれないタイプのニューラルネットワークを指す。
RNN	Recurrent Neural Network	ネットワークにループ構造が含まれるタイプのニューラルネットワークを指す。
LSTM	Long Short-Term Memory	RNN の一種で、長期的な依存関係を学習可能なネットワークモデル
GRU	Gated Recurrent Unit	LSTM をより簡略化したモデル
K 分割交差検証	—	学習データを K 個に分割し、K-1 個の学習データと 1 個の検証データに分け、K 回学習を繰り返し、モデルの汎化性能を検証する方法。
ハイパーパラメータ	—	機械学習アルゴリズムで使用するパラメータのうち、手動で設定する必要があるパラメータのこと。

2. サロゲートモデルとは

サロゲートモデルとは、深層学習等の機械学習技術を活用して詳細プラントモデルを縮退化し、シミュレーション時間とデータ容量を大幅に削減したモデルである。本章ではサロゲートモデルの目的・用途、構築プロセスの概要を述べる。

2.1 目的・用途

車両開発の現場では主に制御モデル開発プロセスにおいて制御対象となるプラントモデルが構築され、閉ループシミュレーションで利用されている。このプラントモデルは開発する制御モデルの要求に応じて非常に詳細なプラントモデルから簡易なプラントモデルまで様々な粒度のモデルが存在する。これらのプラントモデルのうち詳細プラントモデルは計算負荷が高いため、縮退化（ROM 化）したいという要望がある。また、詳細プラントモデルは専用ソフトが必要となることがあるため、専用ソフトを所有していない部署でも詳細プラントモデルを実行可能できる様に MATLAB/Simulink で実行可能な縮退モデルの要望もある。

詳細プラントモデルを縮退化するときの考え方は大きく 2 つ存在する。1 つはプラントモデル全体を縮退化する方法、もう 1 つはプラントモデルの計算負荷が高くかつブラックボックス化しても問題ない部分を縮退化する方法である。どちらを採用するかは、縮退化の目的に応じて使い分ける必要がある。

従来、縮退化手法として MAP や統計モデルが多く採用されてきた。MAP や統計モデルはモデル作成時に過渡状態のデータを使用しないため、定常状態のみ再現可能なモデルである。近年注目されている深層学習技術を活用したモデルは、モデル作成時に過渡状態のデータも使用するため、過渡状態まで再現可能なモデルと言うことができる。深層学習技術を活用して縮退化したモデルをサロゲートモデルという。参考までに、本書で使用する縮退モデルに関する用語と手法の関係を図 1 に整理した。

プラントモデルを縮退化する場合のデメリットとして、プラントモデルの内部構成やパラメータ設定を変更できないという点が挙げられる。これら構成や設定を変更したい場合は、プラントモデルを修正したのち、再度縮退化する必要がある。制御モデル開発ではプラントモデルの構成や設定を変更することはほとんど無い為、上記デメリットは当てはまら無いと考える。

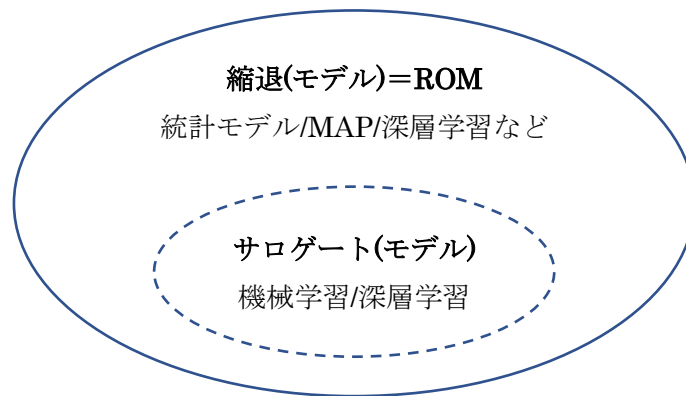


図 1 縮退モデルに関する用語と手法の関係

2.2 構築プロセス概要

本節では深層学習、特に MATLAB Deep Learning Toolbox の Stateful Predict ブロックを用いたサロゲートモデル化プロセスについて示す。まず初めに、本活動で検討したサロゲートモデル化プロセスの全体像を図 2 に示す。

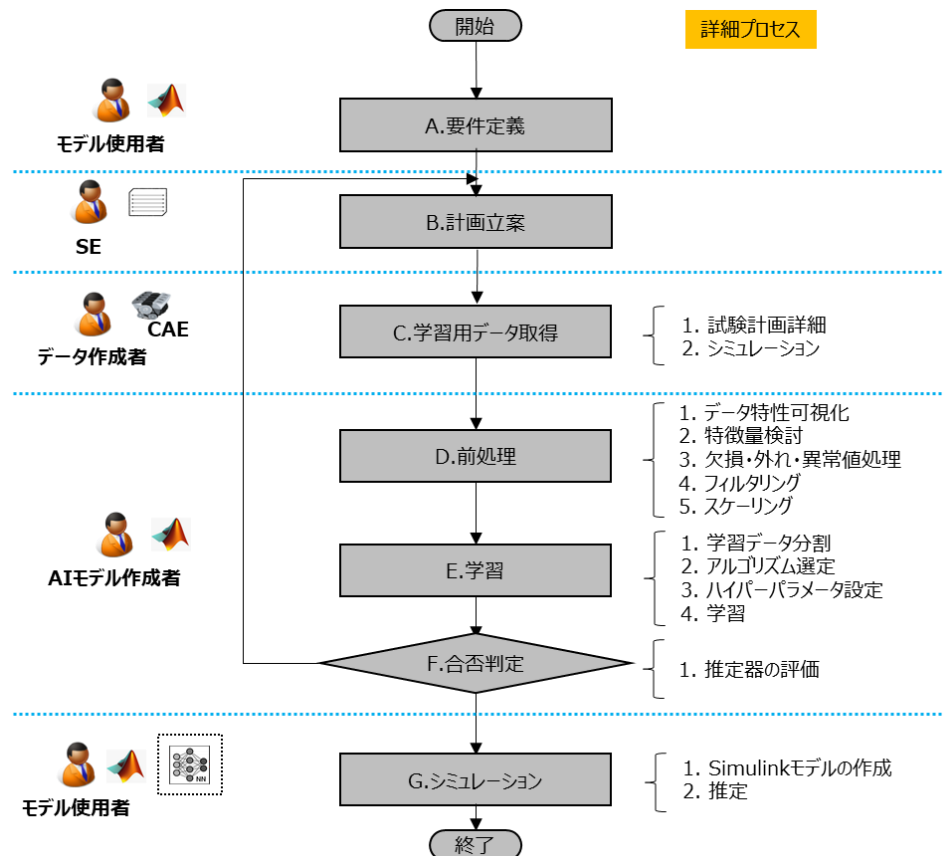


図 2 サロゲートモデル化の全体像

サロゲート化モデルの構築プロセスは、基本的には従来のプラントモデル開発プロセスと同様である。各プロセスの概要は以下の通りである、

「A. 要件定義」では、目標精度やハードウェア制約を考慮し縮退化の要件を定義する。

「B. 計画立案」では、どの縮退化手法が適しているかを判断する。

「C. 学習用データ取得」では、縮退化手法に応じた試験計画を立案し学習用データと検証用データを計測する。

「D. 前処理」では、計測したデータを一次処理し、学習に使える形に整形する。

「E. 学習」では、「B. 計画立案」で決定した縮退化手法のアルゴリズムを用いて、モデルの学習を行う。

「F. 合否判定」では、学習済みの“推論モデル”に検証用データを入力して推論結果を

取得し、縮退化モデルの精度が「A. 要件定義」で定めた基準に合格しているかを判定する。

「G. シミュレーション」では、推論モデルを用いて、もしくは上位システムに組み込んで実務で利用する。

次章では、「C. 学習用データ取得」～「G. シミュレーション」に注目し、詳細な内容を記述する。

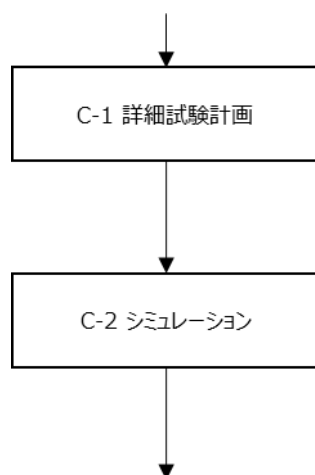
3. 構築プロセス詳細

3.1 構築プロセス詳細一覧

ID	プロセス名	詳細プロセス名	概要
C-1	学習用	詳細試験計画	モデル化対象・範囲の決定
C-2	データ取得	シミュレーション or 試験	データ取得方法の決定
D-1	前処理	データ特性可視化	学習データの特徴把握
D-2		特徴量検討	機械学習モデルの入出力変数を決定する。
D-3		欠損・外れ・異常値処理	欠損・外れ・異常値の除外、補間
D-4		フィルタリング	ノイズ除去もしくはノイズ付加
D-5		スケーリング	特徴量を正規化もしくは標準化する
E-1	学習	学習データ分割	学習用データと検証用データとテストデータを別々に用意する。
E-2		アルゴリズム選定	目的に応じたアルゴリズムを選定する。
E-3		ハイパーパラメータ設定	アルゴリズムに合わせて、ハイパーパラメータを設定する。
E-4		学習	学習用データと検証用データとテストデータを用いて学習を実行する。
F-1	合否判定	推定器の評価	元のモデル(真値)とサロゲートモデルの出力値を統計的な評価指標を用いて比較する
G-1	シミュレーション	Simulink モデル作成	推定器を埋め込む Simulink モデルを作成する。
G-2		推定	上記 Simulink モデルを用いてシミュレーションを実行し、挙動を確認する。

3.2 学習用データ取得

学習用データ取得では主に、詳細試験計画、シミュレーションを実施する。各プロセスをフロー図として整理した。



以下では、各プロセスについて実施内容を簡潔に述べる。

3.2.1 詳細試験計画

実施内容

- サロゲートモデル化対象の選定
 - プラントモデル
- サロゲートモデル化範囲の決定
 - サロゲートモデル化対象の動作パターンを決める
 - 目的変数、説明変数の仮選定
 - サンプルング周期の設定

ポイント

- 模擬したい動作を網羅することが理想
- 実験計画法を用いた効率的な条件選定
- 精度・計測（演算）時間を考慮したサンプルング周期の設定

課題

- 網羅性の判断
- 効率の良い計画の立案

サンプルモデル対応箇所

- —

3.2.2 シミュレーション

実施内容

- 詳細試験計画で立案した内容に従い、モデルを動作させ、データを取得する。

ポイント

- —

課題

- 効率的なデータ取得

サンプルモデル対応箇所

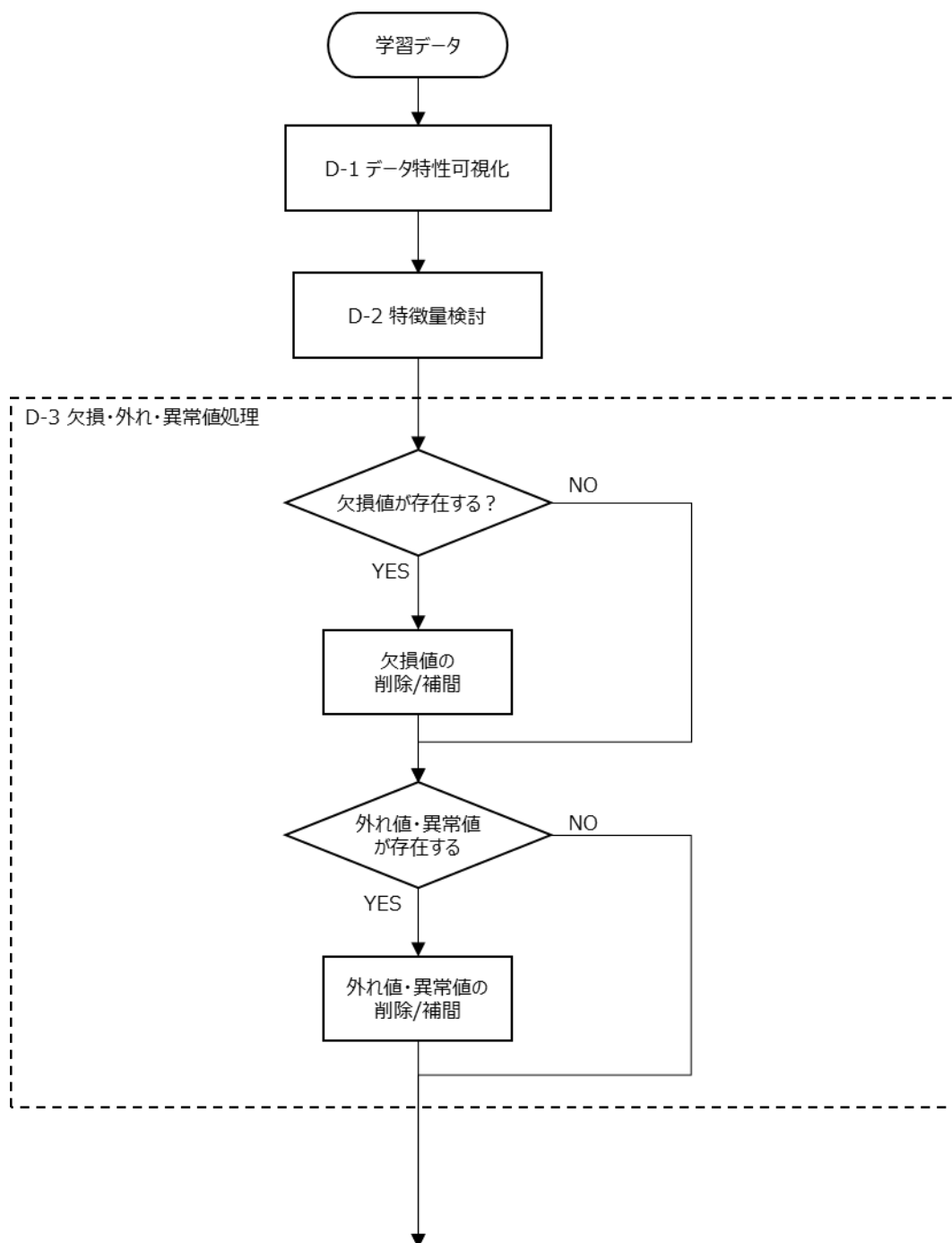
- サンプルモデルでは、下記論文を参照して、
<https://data.mendeley.com/datasets/wykht8y7tg/1> に掲載されているデータ
を使用した

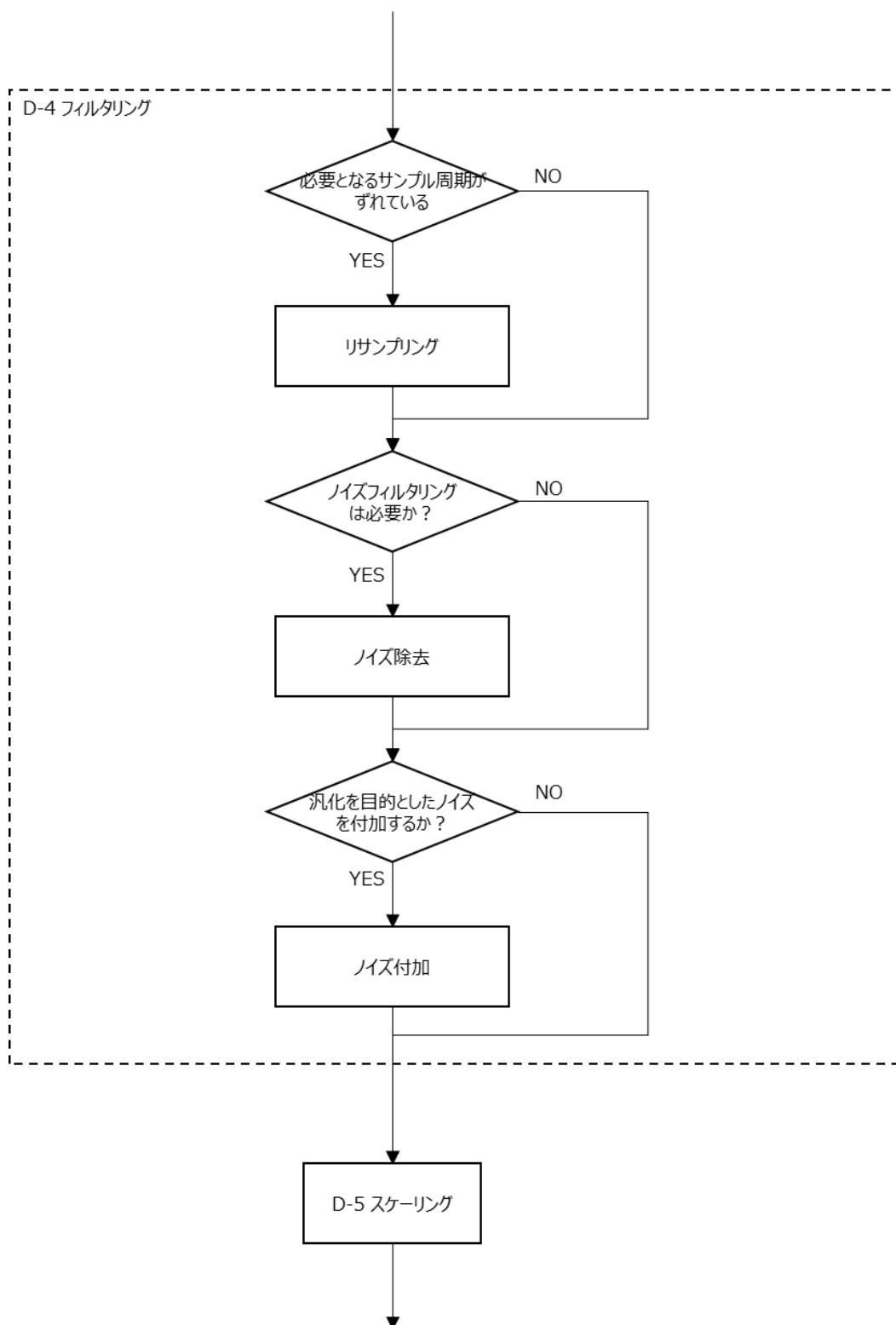
Reference

Kollmeyer, Phillip (2018), “Panasonic 18650PF Li-ion Battery Data”,
Mendeley Data, V1, doi: 10.17632/wykht8y7tg.1

3.3 前処理

前処理では主に、データ特性可視化、欠損・外れ・異常値処理、フィルタリング、スケーリング、特徴量検討を実施する。各プロセスをフロー図として整理した。





以下では、各プロセスについて実施内容を簡潔に述べる。

3.3.1 データ特性可視化

実施内容

- 学習データをグラフ等で可視化し、欠損値や外れ値の有無、周期性や値の範囲を確認する。
- 学習データ間の相関関係を、相関行列等を用いて確認する。

ポイント

- 想定した相関関係が得られない場合は、微分値も確認する。

課題

- —

サンプルモデル対応箇所

- —

3.3.2 特徴量検討

実施内容

- 学習データから、どの項目を特徴量（目的変数、説明変数）として採用するか決定する。

ポイント

- 対象モデルの物理特性を参考に決定する。

課題

- 特徴量を多くしすぎると、学習に必要となるデータ量が指数関数的に増加する恐れがある（次元の呪い）。

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 2.3.1 項参照。
※サンプルモデルでは、物理モデルに基づき特徴量を決定してからデータを計測した。

3.3.3 欠損・外れ・異常値処理

実施内容

- 欠損値がある場合、欠損値を補完もしくは除外するなどの対策を実施する。
- 外れ値や異常値がある場合、一定の基準に則って外れ値を補正もしくは除外するなどの対策を実施する。

ポイント

- 模擬したい振る舞いは除外しないこと。
- データを可視化して、どのような補完方法が適しているか決定する。

課題

- 必要十分なデータの選別に向けた判定基準の設定方法
 - 異常値（例：上限制約以上の値など）かどうかはモデル化対象の知識に基づいて決定する必要がある。
 - 外れ値（例： 3σ イベントなど）かどうかはモデル化対象の知識に基づいて決定する必要がある。

サンプルモデル対応箇所

- —

3.3.4 フィルタリング

実施内容

- 特徴量を横並びで評価するため、サンプリング時間をそろえる。そのため、必要に応じてリサンプリング（アップサンプリング、ダウンサンプリングなど）を実施する。
- 模擬対象外の特徴（例：高周波成分）を除外するため、必要に応じてノイズ除去フィルタを使用する。
- モデルの汎化性能を向上させるため、必要に応じてノイズ（例：ホワイトノイズ）を付加する。

ポイント

- 模擬したい振る舞いは除外しないこと。
- ノイズ除去フィルタを用いる場合は、モデル化対象の知識に基づいてフィルタ時定数を決定する。

課題

- ノイズ判定基準の設定方法
- サンプル数に比例した学習時間の増加が懸念される

サンプルモデル対応箇所

- —

3.3.5 スケーリング

実施内容

- 学習時間の短縮および推論の安定性向上を目的に、特徴量をスケーリングする。

ポイント

- スケーリング方法には、平均を 0、分散を 1 とする“標準化”と、最小値を 0、最大値を 1 とする“正規化”と特徴ベクトルの二乗和が 1 や総和が 1 になるようにする“規格化”がある。
 - 標準化と正規化は数値処理で、規格化は主に言語処理で用いられることが多い。
- スケーリングを実施する場合は、特徴量の特性を把握し、状況に応じて適切な手法を選択する必要がある。

課題

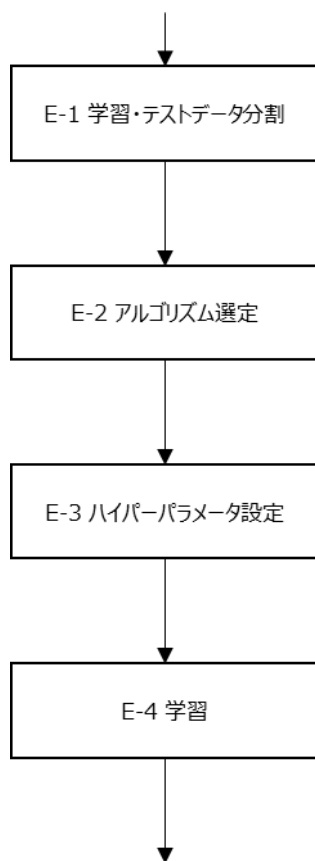
- スケーリング手法の採用基準

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 3.2 節 ①学習条件の設定 を参照

3.4 学習

学習では主に、学習用データ分割、アルゴリズム選定、ハイパーパラメータ設定、学習を実施する。各プロセスをフロー図として整理した。



以下では、各プロセスについて実施内容を簡潔に述べる。

3.4.1 学習用データ分割

実施内容

- 事前に用意した全データを、訓練データ、テストデータ、検証データに分割する[2]
 - 訓練データ：モデルの学習と適合、およびパラメータの決定に使用する
 - 検証データ：モデルのハイパーパラメータを調整しながら、モデルの性能を評価するために使用される
 - テストデータ：訓練データでの最終的なモデルの適合度を不偏的に評価するために使用される

ポイント

- 訓練データ、テストデータ、検証データの分割割合
- 訓練データ、検証データ、テストデータは重複しないこと。

課題

- 一般的に K 分割交差検証が用いられるが、トライ&エラーの要素が強い。

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 3.1 節 ③学習用データ整形 を参照

3.4.2 アルゴリズム選定

実施内容

- 機械学習アルゴリズムの種類や構造を決定する。
 - ニューラルネットワークの場合、FNN や RNN がある。

ポイント

- 状態保持の可否に応じて FNN か RNN を選択する。
- RNN の中でも、応答遅れを含む物理モデルの時、LSTM、GRU などの長期依存を表現できるモデルが有効である。
- 隠れ層の数や活性化関数の種類を選択する

課題

- 学習アルゴリズムに起因する誤差として、バイアス（アルゴリズム構造が単純すぎるために生じる誤差）とバリエーション（アルゴリズム構造が複雑すぎるために生じる誤差（過適合））がある。これらの誤差は背反関係にあるため、学習対象に応じて学習アルゴリズムの複雑さを変更する必要がある。
- アルゴリズム選定の明確な指標がない

サンプルモデル対応箇所

- —

3.4.3 ハイパーパラメータ設定

実施内容

- アルゴリズムに合わせて、ハイパーパラメータ（エポック数、バッチサイズ等）を設定する。
 - ツールによっても設定対象は異なる。
 - ハイパーパラメータの最適化を学習と合わせて行う場合もある。

ポイント

- モデル化対象ごとにハイパーパラメータを調整すること

課題

- 経験則や試行錯誤が必要で、理論的に求めるのが難しい。
- パラメータの値変更毎に再学習が必要なため、最適値探索には時間がかかる

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 3.2 節 ①学習条件の設定 を参照

3.4.4 学習

実施内容

- 事前に準備した、学習データ、テストデータ、検証データを用いて学習を実行する
- 学習結果の確認
 - 評価指標等から、学習用データ、前処理、ハイパーパラメータ、アルゴリズムなど、改良が必要な箇所を検討する。

ポイント

- —

課題

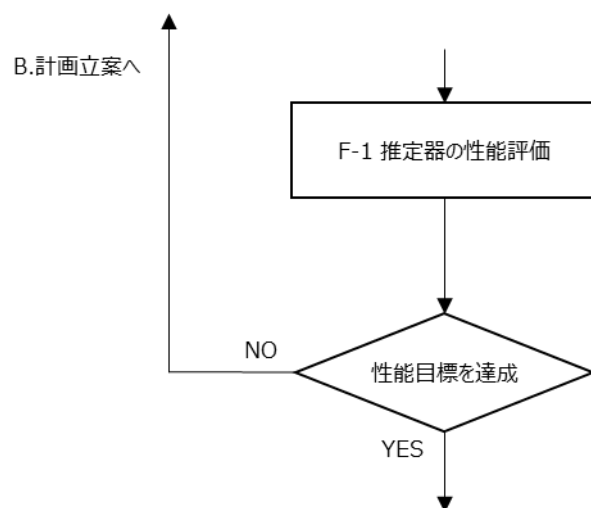
- 教師データの容量によっては何日もかかるが、学習が終わるまで結果が分からない
- 断片的な時系列データを用いて、既存のモデルを追加で学習可能な機能があるとよい。

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 3.2 節 ②ネットワークの学習 を参照

3.5 合否判定

合否判定では主に、学習精度の評価、推論器の評価を実施する。各プロセスをフロー図として整理した。



以下では、各プロセスについて実施内容を簡潔に述べる。

3.5.1 推定器の性能評価

実施内容

- 作成したモデルと元のモデルで挙動を比較
 - 学習データ取得元のモデルや機械のチェックリストがある場合、そのチェックリストも確認し、学習結果の良し悪しを確認する。

ポイント

- 絶対値や応答性が一致しているかを確認
- 学習データとテストデータの **LOSS** 計算結果から、サロゲートモデルが過学習となっていないかを確認する。
- 偏差、**RMSE**、**RMS**、相関係数等の評価指標を用いた判定

課題

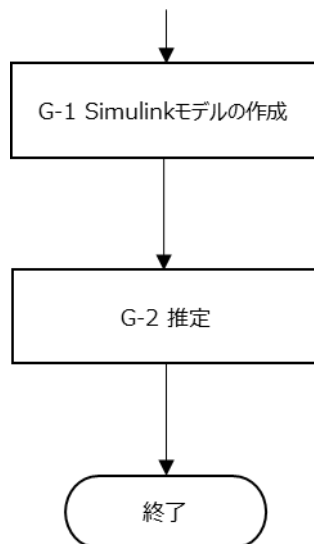
- 評価指標の選定及び、評価指標値の設定
- 評価の自動化
- 評価基準に未達だった場合、原因箇所の特定が難しい。

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 2.3.2 項 検証用モデル を参照
- サロゲートモデルサンプル解説書の 3.2 節 ③学習済ネットワークの評価 & ④学習済ネットワークの評価結果可視化 を参照

3.6 シミュレーション

シミュレーションでは主に、**Simulink** モデルの作成、推定を実施する。各プロセスをフロー図として整理した。



以下では、各プロセスについて実施内容を簡潔に述べる。

3.6.1 Simulink モデルの作成

実施内容

- 推定器を埋め込むモデルを作成する。
 - 推定器と埋め込み先モデルの入出力を合わせること。
- ソルバーの設定なども行う。

ポイント

- サロゲートモデルは学習時と同じ固定ステップサイズで動作させること。

課題

- —

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 2.3.3 項 LSTM 版 HEV モデル を参照

3.6.2 推定

実施内容

- シミュレーションを実行し、挙動を確認する。

ポイント

- —

課題

- 挙動が想定と異なる場合、制御とプラントのどちらに課題があるかの判別が難しい。

サンプルモデル対応箇所

- サロゲートモデルサンプル解説書の 2.3.3 項 LSTM 版 HEV モデル を参照

4. まとめ

JAMMB PMWG モデリングサブ WG では、近年注目されている機械学習/深層学習を用いたプラントモデルのサロゲートモデル化について、サンプルモデルを作成しながら、そのプロセスや手法を明確にすることを目的に活動を実施した。本書はこの活動内容を基に、サロゲートモデル作成のプロセスと各プロセスの実施内容/ポイントを WG メンバの可能な範囲で整理した。本書の内容が、読者の一助となれば幸いである。