

学習モデルグラフ上での仮説検証に基づく 機械学習モデル生成方法の提案と評価

白崎 悠太† 青山 幹雄†

南山大学 大学院 理工学研究科 ソフトウェア工学専攻†

1. 研究背景と研究課題

機械学習を用いたシステム開発の必要性が高まっている。要求認識精度を達成する学習モデル生成にはフィーチャ(特微量)設計の利用が提案されている。しかし、学習へのフィーチャの影響分析は未確立である。

本稿ではプロパティグラフ上での仮説検証に基づく機械学習モデル生成方法を提案する。以上を踏まえ次の3点を研究課題として設定する。

- (1) プロパティグラフによる機械学習モデルのモデル化方法の確立
- (2) プロパティグラフによるフィーチャ分析方法の確立
- (3) 実データに適用し有効性と妥当性を評価

2. 関連研究

2.1. 機械学習モデル生成方法

機械学習モデルの生成方法にはデータのみから生成する方法、表現学習しフィーチャを抽出する方法、転移学習がある[2]。

2.2. フィーチャ工学 (Feature Engineering)

フィーチャ工学とは、機械学習の適用対象となる問題の本質を表現したフィーチャをデータから生成する技術体系である[1]。機械学習の予測精度と学習効率向上を期待できる。

2.3. プロパティグラフ

プロパティグラフとはデータ間の意味定義を表現可能とするセマンティックグラフの拡張である。ノードとエッジに属性の集合をプロパティとして付与可能である[3]。

3. アプローチ

データに対するプロパティグラフ上での仮説検証型の反復プロセスを提案する。反復を通して各プロセスを詳細化する。

学習モデル構造をプロパティグラフとして表現する。学習モデルの学習過程における最終的な認識精度には現れないパラメータの変化に着目する。プロパティグラフ上で学習モデルの差分を分析することによって学習モデルに影響を与えるフィーチャを特定し、それに基づいて学習する。

4. プロパティグラフを用いた仮説検証型機械学習モデル生成方法

4.1. 提案プロセス

提案プロセスを図1に示す。

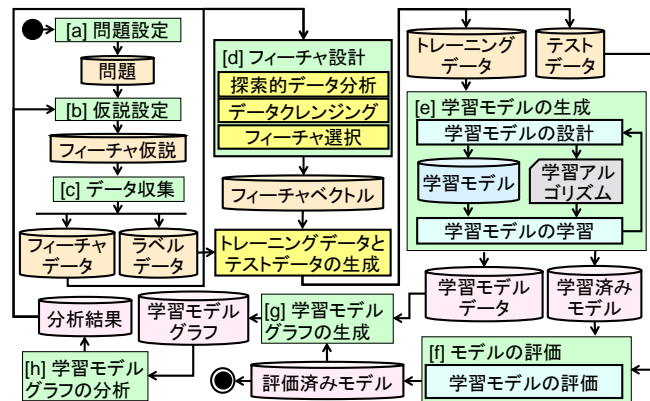


図1 提案プロセス

[a] 問題設定

機械学習モデルが要求を満たすためにモデルが解くべき問題を設定する。プロセスを繰り返すにつれて問題設定は詳細化される。

[b] 仮説設定

問題解決に必要なだと推定されるデータについての仮説を設定する。学習モデルグラフの分析によって分析対象として収集すべきデータを見直す。

[c] データ収集

仮説から分析に必要なデータを取得する。

[d] フィーチャ設計

学習モデルの入力となるフィーチャベクトルを生成する。探索的データ分析、データクレンジング、フィーチャ選択の3ステップからなる。

[e] 学習モデルの生成

学習モデルを設計と学習の2ステップから生成する。

[f] モデルの評価

テストデータを用いて目的関数に基づき学習モデルの汎化誤差を評価する。モデルの評価結果が要求を満たしている場合、提案プロセスを終了する。

[g] 学習モデルグラフの生成

学習モデルデータからプロパティグラフとして差分を分析する学習モデルグラフを生成する。

[h] 学習モデルグラフの分析

生成した学習モデルグラフから重みの変化率を分析することで、フィーチャの影響力を分析する。

4.2. 学習モデルグラフの生成

機械学習モデルの構造をプロパティグラフとして定義したグラフを学習モデルグラフと呼ぶ。本稿ではニューラルネットワークを学習モデルグラフとして定義し分析対象とする。

4.3. 学習モデルグラフの分析

重みの平均変化率の大きい重みと結ばれているフィーチャードを探索することによって学習に影響を与えるフィーチャを特定する。そのために学習数 e から学習数 e' における第 i 層目のフィーチャードに連結する重みの平均変化率を式(1)により定義する。式(1)の評価結果からフィーチャの影響力を評価する。

$$R_j^{(l)} = \frac{1}{n} \sum_{i=1}^n \frac{|w_{i,j,e'}^{(l)} - w_{i,j,e}^{(l)}|}{|w_{i,j,e}^{(l)}|} \quad (1)$$

5. プロトタイプの実装

プロトタイプのアーキテクチャを図2に示す。

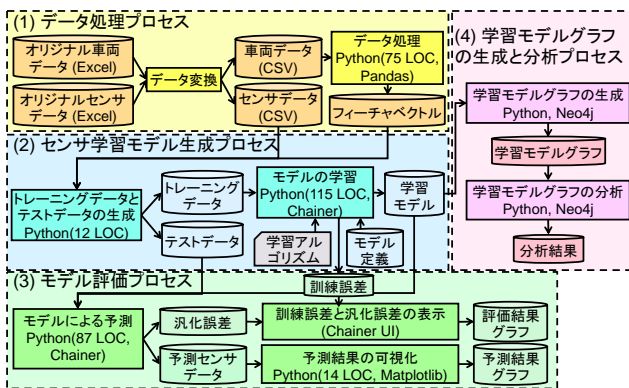


図2 プロトタイプのアーキテクチャ

6. 車載超音波センサデータへの適用

車載超音波センサが受信する路面反射波の電圧データに提案方法を適用し評価した。[a]問題設定プロセスからセンサが利用されている自動車の環境条件データ(搭載位置, 湿度, センシングパターンなど)からセンサが受信する超音波の反射波電圧を予測するモデルを生成する問題として問題設定をした。[b]仮説設定プロセスにより収集すべきデータについての仮説を設定した。[c]データ収集プロセスにより実験データからフィーチャデータと教師データとする反射波電圧のデータをラベルデータとして取得した。[d]フィーチャ設計プロセスにより, 仮説に基づきフィーチャデータから学習モデルの入力とするフィーチャを選択しフィーチャベクトルとして再構成した。[e]学習モデルの生成プロセスにおいて学習モデルに MLP を採用した。その時のモデル設定を表1に示す。

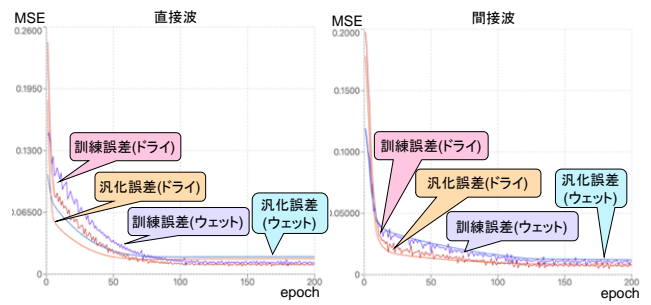
表1 モデル設定

epoch	300
バッチサイズ	1
最適化アルゴリズム	MomentumSGD
活性化関数	シグモイド関数, ReLU
学習率	0.01, 0.001(epoch>150)

評価関数として式(2)の平均二乗誤差を採用し訓練誤差と汎化誤差の評価をした(図3)。 n は入力データ数, f_i は i 番目の入力 x_i に対するモデルの出力, y_i は i 番目の入力 x_i に対する教師データの反射波電圧を表す。

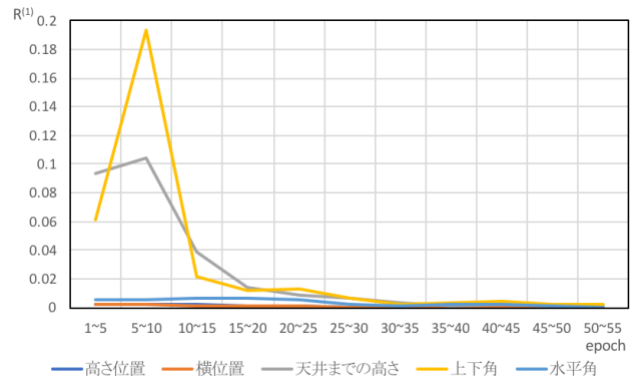
$$MSE = \frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2 \quad (2)$$

図3 学習モデルの認識精度



直接波のドライの学習における第1層のフィーチャードに連結する重みの平均変化率の推移を図4に示す。高さ位置と天井までの高さについての変化率が他よりも大きいことがわかる。したがって, 学習初期では上下角と天井までの高さというフィーチャの影響力が高いことが推定される。

図4 重みの平均変化率



7. 考察

- (1) プロパティグラフによる学習モデルのモデル化
プロパティグラフとして学習モデルをモデル化し表現することにより, 学習モデルをデバックすることが可能になり学習プロセスの効率化が期待できる。
- (2) 学習モデルグラフの分析
学習に影響を与えるフィーチャの特定から選択したフィーチャの検証を行うことが可能になる。フィーチャ選択の効率化によって多様なドメインでの機械学習システムの開発支援が可能である。

8. まとめ

本稿では学習モデル構造を学習モデルグラフとしてモデル化し分析することによるプロパティグラフ上での仮説検証型機械学習モデル生成方法を提案した。学習に影響を与えるフィーチャの特定に基づくフィーチャ選択の仮説検証によって, 多様なドメインでの機械学習モデル生成の支援が期待できる。

謝辞: センサデータをご提供頂いた株式会社デンソーの関係各位に感謝する。

参考文献

- [1] G. Dong, et al., Feature Engineering for Machine Learning and Data Analytics, CRC Press, 2018.
- [2] I. Goodfellow, et al., Deep Learning, MIT Press, 2016.
- [3] I. Robinson, et al., Graph Databases, 2nd ed., O'Reilly, 2015.