生産システム部門研究発表講演会 2021 講演論文集['21.3.8,9,オンライン]
Copyright © 2021 一般社団法人 日本機械学会

シミュレーションと機械学習の組み合わせによる高速近似解析手法構築の能動学習を用いた効率化に関する研究

Efficiency of Fast Approximation Analysis by Simulation and Machine Learning Using Active Learning

峯田 龍志・大阪大学　岩田 剛治・大阪大学　若松 栄史・大阪大学

Ryushi MINETA, Osaka University　Yoshiharu IWATA, Osaka University　Hidefumi WAKAMATU, Osaka University

松本 侑哉・大阪大学　川村 俊貴・大阪大学

Yuya MATSUMOTO, Osaka University　Toshiki KAWAMURA, Osaka University

論文要旨

In recent years, AI has been adopted in the manufacturing industry, and machine learning-based approximator systems are expected to be used for shortening the development time of products. Although it is generally accepted that big data is essential for training data, it takes a long time to collect the data obtained from simulations, so that a high-performance approximator must be built with a minimum amount of training data.Although active learning has been studied in various fields to reduce the data collection cost, most of the research is on the classification problem, and its adaptation to the regression problem is not advanced. In this study, we construct a simulation approximator for this regression problem with a minimal amount of training data, using expressive neural networks to approximate unknown functions.

Key Words: *Active learning, Machine learning, Neural network, Simulation, Approximator, Small amount data, Improving quality of data*

1. 緒言

近年、AI は製造業において取り入れられ、長時間を要する数値シミュレーションの代わりに、サロゲートモデルとして機械学習による近似器システムを利用することが、製品開発の時短化に期待されている。現状の機械学習技術では、非常に多くの学習データ(ビッグデータ)が不可欠であると一般的に言われているが、数値シミュレーションの結果の収集には長い時間を要する [1)]。よって、最低限の学習データ量で高性能な近似器を構築することが重要になる。

データの収集コストという課題を解決するため、能動学習が様々な分野で研究されているが、分類問題に関する研究が多くを占めており、回帰問題に適応した研究はほとんど行われていない [2-3)]。本研究では、この回帰問題の近似に対応するため、未知の関数を近似するための表現力に富むニューラルネットワークを用い、最低限の学習データ量でシミュレーション近似器を構築する。ベースとなる近似器の予測結果から精度向上に効果的なシミュレーションモデルデータを推定し、そのデータをニューラルネットワークの学習データに対して追加して再度学習を行う。

この行程を図 1 のような能動学習的システムとして捉え、最適な追加サンプルを繰り返し追加していくことにより、可能な限り少ない学習データ量で効率的に学習を行い、近似器モデル精度の向上を目指す。

2. ケーススタディとニューラルネットワークへの適応

本研究のケーススタディとして取り扱うデータは、実際の機械部品の製品設計に用いられる設計変数の値と、その機械部品のシミュレーションによる応力計算値である。表 1 に示すように、6 つの独立した設計変数が持つ値は、それぞれとりうる値の範囲の最大値、最小値、中間値のうち 1 つである。非線形な出力に対応するためには，各設計変数で最低限 3 水準の学習が必要であるため、この 3 水準値を 6 設計変数で総組み合わせを行い、全 729 個(=3^6)のサンプルにおいてシミュレーション解析出力値を得ている。以降このデータをニューラルネットワークの学習データの基本形とする。また、テストデータは、実験計画法の一種であるラテン超方格法により、6 設計変数にランダムな値を持つサンプルを 100 個生成する。

シミュレーション出力値は正規化により無次元化されており、シミュレーション出力値に対する近似器予測値の誤差の標準偏差 σ の 2 倍が 1 以内、すなわち σ≦0.5 に収束することを近似器の目標精度とした。

以上をニューラルネットワークに適応した結果、学習データでは σ=0.28 となったが、テストデータでは σ=0.54 となり仕様を満たさない。そのため，誤差が大きいサンプルの誤差を小さくすることによる近似精度の向上を考えた。よって、誤差の絶対値>0.5 のものを誤差が大きいサンプル、誤差の絶対値<0.5 のものを誤差が小さいサンプルと定義すると、誤差の絶対値が 0.5 を超えるサンプルは学習データ 729 個中 15 個、テストデータ 100 個中 39 個存在しており、誤差改善の余地が十分に残された状態であった。次章ではニューラルネットワークによる予測結果から誤差要因について分析し、誤差改善方法を検討した。

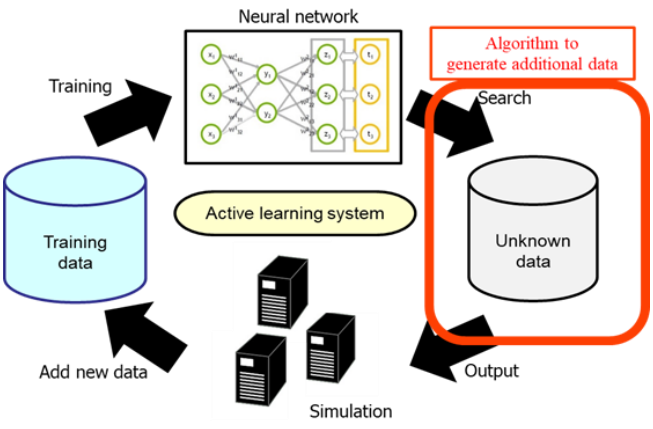


Fig.1 Image of active leraning system

Table.1 Sample of training data

Sample No.	Simulation analysis	Input 1	Input 2	Input 3	Input 4	Input 5	Input 6
1	-1.6	1.2	0	-1.4	1.2	-1.4	-1.4
2	-2.1	1.2	0	-1.4	1.2	-1.4	0
3	-2.4	1.2	0	-1.4	1.2	-1.4	1.4
…	…	…	…	…	…	…	…
727	-1.3	1.2	0	-1.4	0	1.4	-1.4
728	-2.6	1.2	0	-1.4	0	1.4	0
729	-2.4	1.2	-1.4	-1.4	-1.2	1.4	1.4

3. 誤差改善のための追加サンプルの推定手法提案

回帰問題に対し、学習データ数が少ない場合に発生する誤差要因として、設計変数の値が隣接するサンプル間のシミュレーション出力値（以下出力値と記述）の変化が大きい場合、ニューラルネットワークの学習の追従が困難になり、誤差が大きくなる場合がある[4]。本研究においても同様の誤差要因があると仮定し検証を行ったところ、同様の傾向を得た。

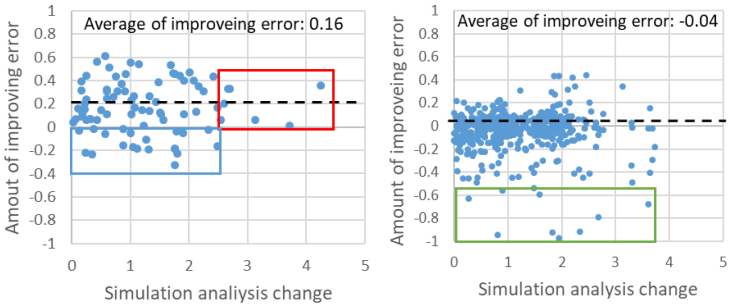
以上より、仮説として、出力値の変化の大きいサンプル間に新たな学習点を追加して隣接サンプル間の変化量が小さくすることで、ニューラルネットワークの変化に対する追従精度が向上し、誤差が改善されると考えられる。

次に、学習データ数をさらに抑制するために、多数ある追加サンプルの候補から有効性の高いものの厳選を試みる。誤差の大きい複数のサンプル（以降サンプル群と呼称する）における設計変数の値の共通性に着目すると、誤差の大きいサンプル群において、ある設計変数の値が頻繁に含まれる場合、その設計変数の値周辺で誤差が大きく出る確率が高いと考えられる。そこで、サンプル群で共通性の高い設計変数の値は、対象サンプルと同じ値もしくは隣接水準値との中間値を与え、それ以外の設計変数では対象サンプルと同じ値を与える。この考え方を追加サンプル推定手法とし、厳選された追加サンプル群を作成する。

4. 提案手法の誤差改善効果の検証

本章では実際に提案手法の誤差改善効果の検証を行う。まず、全データ 729 個を学習したニューラルネットワークの予測結果において、最も誤差の絶対値が大きいサンプルを誤差改善対象とした。誤差の大きいサンプル群は 15 個のサンプルで構成され、共通の設計変数の値の数の割合の大小を決める閾値 n は 50%とした。これらの実験条件を踏まえ、提案手法により追加サンプルが 23 個生成された。また本検証では、追加サンプルを 1 個のみ学習データに加え、隣接するサンプル全てでの誤差改善効果評価を目的とする。学習データに対するサンプル追加 1 個を 23 パターン行った予測器の学習結果を総合して用い、追加サンプルに隣接するサンプル 546 個（隣接点 2 個の組合せが 273 通り）が抽出された。

図 2 は、誤差の絶対値 0.5 を基準に分けた誤差大サンプル 86 個と誤差小サンプル 460 個における、隣接サンプル間の出力値の変化量と誤差改善量の関係を示している。また、誤差大サンプルに対しては中間点の追加により最大値 0.6、平均値 0.16 の誤差改善が確認できたため、提案手法による誤差改善は可能であると言える。しかし、図 2(a)中の左下枠が示す通り誤差悪化しているケースも存在し、現状の手法では全ての誤差大サンプルに対して誤差改善が行われないことを確認した。また、隣接点間の出力値の変化量という評価軸では、データ追加によって隣接サンプル間の出力値の変化量の大きい点で誤差改善量が大きくなるという仮説を明確に立証できなかったが、図 2(a)中の右上枠に示す出力値の変化量の大きい箇所では常に誤差が改善しているため、更に評価



(a) Large error (n=86)　　(b) small error (n=460)

Fig.2 Improving error with simulation analysis change

軸の検討を行う必要がある。さらに、図 2(b)に示す誤差の小さい点に関しては、誤差改善の平均値が-0.04 と元の誤差とほぼ変化は無いという結果が得られた。しかし、図 2(b)中の下枠に示す通り、誤差が大幅に増加している点が少数ではあるが散見されるため、そのような点の特徴や傾向を調査する必要がある。

5. 追加サンプルの更なる絞り込みに関する考察

現在、追加サンプル絞り込みの新たな指針として、周辺のサンプルとの複数設計変数での傾きの変化の増減によるニューラルネットワークの不追従の可能性を見出している。設計変数の寄与度の大小に関わらず、ニューラルネットワークの学習が難しくなっているのは、ある学習点とその周辺との変化の傾向が他の学習点と異なる場合であると推測される。よって、今後の課題として、各設計変数において誤差が大きくなる変化パターンの組み合わせ等、誤差と誤差要因の関係性を明らかにする評価軸を精査する必要がある。

6. 結言

本研究では、ニューラルネットワークで最低限の学習データを用いた近似器を作成し、誤差が大きい点の特徴から効率的な学習データの推定を行い、効果的な精度向上を確認した。

本稿では、隣接サンプル間の出力値の変化量が大きい箇所にサンプルを追加するアルゴリズムのみに着目してきたが、このアルゴリズムでは学習データ内においてデータが疎な箇所にはサンプルの追加ができないため、近似器の汎化性能としては不安が残る。今後の展望としては、テストデータの予測値と誤差分布を詳細に分析することにより、シミュレーション解析出力値や設計変数の値が疎な箇所にサンプルを追加するアルゴリズムを構築することが目標である。最終的に全てのアルゴリズムから得られた結果を総合評価し、汎化性の高い近似器を構築可能なシステムを目指していく予定である。

参考文献

(1) Ankita Mangal，Nishant Kumar，“Using Big Data to Enhance the Bosch Production Line Performance: A Kaggle Challenge”，IEEE International Conference on Big Data (2016), pp.2-9

(2) Graham Neubig，　森 信介，”能動学習による効率的な情報フィルタリング”，　言語処理学会 第 18 回年次大会 発表論文集(2012), pp 887-890

(3) 桜井 啓介，谷祐 樹，谷川 修，　”競合性ニューラルネットを用いたオンライン準教師付き能動学習法”，電子情報通信学会論文誌. D，情報・システム / 電子情報通信学会編，Vol.90(11)，No.431(2007), pp 3091-3102

(4) 松本 侑哉，“シミュレーション近似器における能動学習の適用に関する研究”，大阪大学卒業論文(2020), pp1-51