

QE スクエア



本欄は「会員の声」と同様、個人意見の主張の場であり、営利目的や誹謗等を除き、会員が自由に主張や意見を述べるためのスペースである。

●実験計画法を応用した機械学習モデルのロバスト性向上手法

富士通(株) 井出 勝
富士通(株) 中澤克仁

1. はじめに

タグチメソッドは主に製造業において量産設計をする際の品質向上、いわゆるばらつきの制御設計をすることによって市場に出た製品の不具合を減らし、生産者および消費者にとって特に品質コスト的に価値のある手法であるが¹⁾、今後製品の核としてシステムの主流になっていくと各所で予測されているAIシステムを搭載した製品にはこのタグチメソッドはいまだ応用例が少ない。近年は、製造業以外にこのタグチメソッドを応用した例が出てきており²⁾³⁾、品質工学の適応範囲は広範囲に及ぶことが示されてきた。そのため、この品質工学をAI技術に応用していくことは、市場の品質向上のために十分検討価値があると考えられる。AIシステムは機能性評価をしようにも、例えば入力データは正解のラベリングされた膨大な教師付き画像データを用いる場合、出力のロバスト性であるばらつきを制御するには最小勾配法などによる出力精度の最適化関数がモデル内に存在しており、学習用データで学習済みのモデル構築の際、入力データが一意的に抽出されるモデルの中身に関してはブラックボックス化してしまうという課題がある⁴⁾。つまり、一般的な手法でAIモデルを作成すると内部パラメータが一意的に決められ、ある程度自動で精度の高い最適化がなされてしまうため、ばらつきの少ないモデルを狙って作ることが難しい。一般的にAIモデルはデータが多ければ多いほど、精度の高い出力を出すと言われているが⁵⁾、実際の現場で得られるデータには多くの場合ノイズが混入したデータで学習するため、場合によっては全く収束せず、適切な予測結果の期待は低くなるケースがある。また、大きすぎる

データサイズを学習する際、オフラインで稼働する必要のある現場などではGPUを搭載したハイスペックPCが要求され、コストの面から、AI開発現場で膨大なデータを処理することは現実的ではない。一方で、学習モデル構築の前処理としてクレンジング処理、異常検知処理等の手法⁶⁾により、丁度良いサイズのデータにまとめたとしても、偏り過ぎるデータになりがちであり、過学習を起し、ノイズに弱く、ある限られた範囲の予測しかできなくなってしまう。そもそも、機能を定義して、その機能をばらつきなく出力させる手法をAIシステムは要求仕様として持っていないため、想定外のデータやノイズ混入データが与えられた場合、どのような結果を出力するか、開発者ですら予測が難しいのが現状である。そういった課題の中で、ノイズに強いAIモデルの構築手法を本研究テーマの着眼点とした。

富士キメラ総研の人工知能ビジネス総調査⁷⁾によると、業務効率化を目的としたAIの活用が進み、画像分析による業務支援サポートが行われてきており、近年、業務負担の軽減やAIが組み込まれた監視カメラ等の導入が進むとみられている。こういった背景の中で、われわれは画像処理・分類によって人の活動や操作の代替、もしくは効率化に関する技術分野にフォーカスした。従来手法である生成モデル学習手法⁸⁾は、数増しのために生成されたデータから、ユーザより好ましくないと判断されたデータをはじく手法として提案されている。ただし、学習用画像が増大することを防ぐ際のユーザの判断には時間とコストが必要であり、アナログで1つずつデータを判断することが課題である。また、数増したデータを、「誤る可能性のある人の判断」に委ねられ、誤って必要な訓練データをはじいてしまうことも課題となっている。

2. 近年の開発現場状況

2.1 ノイズ付与データでの訓練例

たとえば予測結果を狂わせる Adversarial