

インタラクティブGAを用いた 「CNNの汎化性能に必要なとなる訓練画像」の探索

永田研究室 F122043 中島 栞

1. 目的

近年、深層学習の技術を画像認識に特化させたCNNや、CNNを特徴抽出器としてSVMと融合させたAIモデルなどを製品の欠陥検出に応用しようとする試みが数多く報告されており、日々新たな手法が公開されるなど高性能化に向けた進化を続けている。このような目的でCNNモデルを訓練する際は、学習用に画像データが必要となり、検出する欠陥によっては膨大な枚数の画像データが必要な場合もある。しかしながら、この大量のデータセットを訓練向けに最適化するためには画像枚数の調整を経験的に手動で行う必要がある。本研究ではインタラクティブGAを用いてCNNの訓練に必要なパラメータの自動調整に加え、より少ない訓練画像枚数で訓練用データセットを最適化し、訓練に伴う計算コストや学習時間の削減を可能にし、分類精度の高いCNNの構築を目指す。

2. 研究内容

本研究では最大エポック数（整数）、ミニバッチサイズ（整数）、学習率、訓練用画像データセットの中から実際の訓練に用いられた画像の抽出割合X [Image usage rate %]（整数）の4つをCNNの訓練時の探索パラメータとした。これら4つの値をそれぞれ8ビットで二値化し、遺伝子コードに変換した。実験では集団を2つのエリート個体(E=2)を含む12個体(N=12)で構成し、1個体は32(=8×4)ビットの遺伝子長となる。次世代の集団(N=12)を生成するために、図1の探索フローのような操作を行った。まず、トーナメント選択によりエリート(E=2)を除く個体の中からランダムに3個ずつの個体を選択しながら、その中で適応度の最も高い個体を次世代に残した。この操作を全個体数－エリート数(N-E)の回数だけ実行し、次世代にも同数の個体を生成させた。次に、交叉確率50%の一樣交叉による交叉処理を行い、その後、個体のビット長を考慮し、 $1/32 = 3.13\%$ の確率で突然変異の処理を行っている。評価にはキャッピングタスクロボットのキャップ装着音「capping」検出用のCNNを構築する際の画像データセットを用い、ワークの衝突音や落下音などの雑音「noise」、人間の話し声の波形画像「voice」を含む3カテゴリから構成される。訓練画像は100枚×3カテゴリ=300枚、テスト画像は50枚×3カテゴリ=150枚である。

3. 結果

試行錯誤的に手動設定でパラメータ調整を行いながら訓練した場合と、インタラクティブGAによりパラメータを自動調整しながら訓練した場合の性能を比較するため、前述のデータセットを用いてそれぞれ10回ずつの学習を実施し、結果を比較評価した。その結果、GAを用いることにより、より短時間に高い分類精度を発揮できるパラメータを自動的に獲得できることが分かった。また、GAによるパラメータの自動調整の結果により得られた訓練画像データは、900枚から387枚(元のデータセットの43[%])まで削減できており、より少ない画像枚数でも効果的に学習できていたことから、特徴量のリッチな画像が抽出できていたものと思われる。なお、抽出された画像と抽出されなかった画像を目視で比較したところ、画像の特徴の違いや傾向を識別することは困難であり、同様の分別作業を人の手で行うことが非常に難しいことも分かった。この他、提案手法をCIFAR10、芋の葉の画像、ラップフィルムの画像などのデータセットにも適用したところ、同様の有効性が確認された。このようにインタラクティブGAの適用は、人間の目視だけでは難しい作業訓練用データセットの最適化を可能にしているといえる。

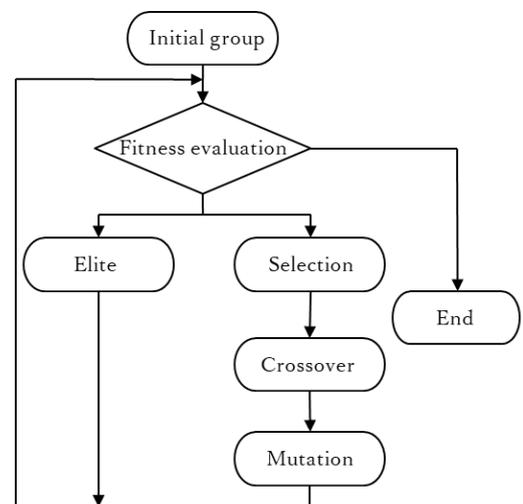


Fig. 1 System flow diagram of the GA used