

枝刈り・量子化・符号化を用いた カプセルネットワークの軽量化

島田 研太

要旨

深層ニューラルネットワークの一種である畳み込みニューラルネットワーク (Convolutional Neural Network; CNN) は、特徴抽出を行う畳み込み層と次元圧縮を行うプーリング層とを複数接続させることで、画像認識において高い精度を実現している。これに対し近年、プーリング層に起因する問題が指摘され、それを回避するための新たな深層ニューラルネットワークのモデルとしてカプセルネットワーク (Capsule Network; CapsNet) が提案されている。CNN など従来の深層学習モデルでは、各ニューロンはパラメタとして重みベクトルを持ち、一つのスカラー値を出力するが、CapsNet では、各ニューロンはパラメタとして重み行列を持ち、カプセルと呼ばれるベクトル値を出力する。そのため、必然的にモデル (パラメタ数) が大規模化し、大量の演算リソース (計算時間とメモリ) が必要とされることが予想される。この問題を軽減するため、本研究では、CapsNet に対する軽量化手法を模索する。具体的には、CNN に対する代表的な軽量化手法である枝刈り、量子化、符号化を基本とするが、スカラー出力をベクトル出力に変更したという CapsNet の大きな特徴を考慮し、各技術の適用単位をスカラーではなく、ベクトルとすることを提案する。提案手法の有効性を確認するため、MNIST データと CIFAR-10 データを用いた計算機実験を行った。単位とするベクトルサイズや量子化粒度 (クラスタ数) を変化させながら、識別精度とモデルサイズの観点から評価を行った結果、一定の効果が確認された。