

マルチタスク学習と多次元回帰に関する考察 ～ポケモン画像を例に～

春山 裕亮

要旨

スマートフォンやドローンなどの小型カメラを内包したデバイスの普及も契機となり、近年、多くの画像認識システムが開発、提案されている。これらのシステムの中には、例えば料理画像を与えると、料理種別に加え、カロリー量や塩分量、調理時間など、複数の値（属性値）を同時に推定するものも少なくない。本研究では、深層学習技術、特に畳み込みニューラルネットワーク（Convolutional Neural Network; CNN）を対象に、画像から複数の属性値を推定する問題に焦点を当てる。CNNを用いる場合、様々なモデル構成で複数属性値の推定を実現することが可能である。例えば単純に(1)1次元の出力を行うCNN（単次元CNN）を属性数分準備する、(2)属性数次元の出力を行うCNN（多次元CNN）を一つ準備するといった構成が考えられる。加えて、マルチタスク学習において一般的に用いられる(3)ネットワークの一部を共有する複数のCNN（マルチタスクCNN）を準備するという構成も考えられる。しかし現状、どのようなモデル構成が予測精度の向上に寄与するかは必ずしも明らかではない。これらのことを背景に、本研究では、予測精度向上に寄与するモデル構成に関する知見を獲得するため、様々な構成のモデルを構築し、その精度を実験的に比較、検証する。具体的には、ポケモン画像からタイプとステータス6種を予測する問題を題材とし、単純なモデル構成に加え、属性の意味や統計情報に着目したモデル構成を対象に比較を行った。その結果、組み合わせるCNNの数が多いほど精度が向上する傾向があることや、単一の多次元CNNよりマルチタスク単次元CNNの方が精度が良い傾向があることなど、一定の知見を得ることに成功した。