

マシンのためのビッグデータ可視化技術：特徴ベクトルの自動選択とデータ解析法の野蛮化

Visualization technology for machines to understand Bigdata: Automatic selection of feature vectors and barbarization of data analysis methods

*樋口 知之¹

*Tomoyuki Higuchi¹

1. 統計数理研究所

1. The Institute of Statistical Mathematics

スパースモデリングの典型的利用法は、線形回帰モデルの説明変数の自動選択である。説明変数を記述子、また説明変数のセットを特徴ベクトルと呼ぶことも多い。これまで特徴ベクトルの構成法は、ほぼ人間の知的作業に委ねられており、実は、この構成法が、予測・判別性能といったパフォーマンスをほとんど決めていると言っても過言ではなかった。つまり、特徴ベクトル構成法は、機械学習の『匠の技』と言え、機械学習にもかかわらず人間の判断が最も性能向上に大切という、羊頭狗肉の面もあった。深層学習では、“生”データを直接入力とすることで、この特徴ベクトルの構成作業を省略しながらも大きな成果を上げている。深層学習を使えば、ユーザは特徴ベクトルの選択・決定に悩まずに最高の予測・判別性能を得ることができる。まさに、「深層学習は『匠の技』の習得からユーザを解放した」と喧伝される所以である。この方針をさらに進め、前処理などをすべて省いた生データから、最終的な目的まで直接、深層ネットワークをつないだEnd-to-Endという研究も加速している。さらには、この力を逆手に取り、生データを深考無しにとりあえず可視化および画像化し、あわせて、データ処理に関わる諸作業の様子を撮像するなどにより、大量の画像や動画を作成すれば、特徴ベクトルの選択問題を回避しつつ目的を達成することも原理的には可能である。これまでのデータ解析においては、生データの持つ特性と解析目的に沿った解析法の慎重かつ適切な選択が成功の肝であった。しかしながらこれからは、上述したようなすべて画像化・音声化するといった、データ解析の目的からすると遠回りに見える、“野蛮”なデータ解析が普通感覚になっていくかも知れない。

キーワード：可視化技術、深層学習、エンド2エンド学習、転移学習

Keywords: Visualization technique, Deep learning, End-to-end learning, Translational learning