

【背景】近年、ソーシャルデータを利用した株式市場の解析が行われているなか、S&P500 構成銘柄に着目して、Wikipedia 閲覧数、出来高、ボラティリティなどが劇増するタイミングの時間的前後関係を統計解析してきた。

【目的・方法】本研究では、確率過程に従う上記3変数の時系列データへの機械学習の適用可否を検討する。具体的には、S&P500 構成銘柄等の Wikipedia 閲覧数と出来高の規格化された変動値に多層パーセプトロンを適用し、得られた予測結果からこれら2つの時系列データ間のタイムラグを検討した。学習環境としては、Google が開発したオープンソフトウェアライブラリ「TensorFlow」とそのラッパーライブラリである「Keras」を用いた。ネットワーク構造としては、入力層、中間層(2層)、出力層の多層パーセプトロンを用いた。

【検討事項】学習させるデータの形による予測精度を検討した。また、Wikipedia 閲覧行動と出来高のピークが一致している場合を除いたデータを学習するなど、入力データとして何が適切なのかを比較・検討した。その他、学習回数、ミニバッチ数の変更による予測精度の比較、損失関数の変化による出来高の予測精度比較、ならびに Wikipedia 閲覧数と出来高の時間的前後関係の分析を行い、多層パーセプトロンでの適切な学習環境を探索した。