

感情分析と時系列モデルを統合したデータ駆動型予測システムの開発 — PBL 科目実践紹介 —

(東京都立産業技術高等専門学校¹)

○下沢 亮太郎¹・蓑手 智紀¹・石垣 雄太郎¹・高田 拓¹・笠原 美左和¹
宮野 智行¹・吉田 嵩¹・福田 恵子¹

キーワード：画像生成, 時系列予測, 感情分析, BERT, LSTM

1. 緒言

都立産技高専荒川キャンパスでは、選抜された3～5年生を対象に、IoT/AI技術を学習する未来工学教育プログラムを実施している。5年生のPBL(Project Based Learning)科目では、学生たちの自由な発想を基に、学習した知見や技術を活用してアイデアを実現する取り組みが行われている。本報告では、PBL科目での取り組みについて紹介する。

社会・経済指標などの時系列予測を精度よく行うためには、社会に影響を与える事象をモデルに定量的に入れる必要がある。しかし、一般によく利用される時系列予測モデルでは、数値を入力とすることが多い半面、社会的な事象の多くは、文章や映像で伝えられるのが事実である。

本研究では、経済指標として株価指数S&P500に属している株式を用いて、機械学習モデルであるLSTM(Long Short Term Memory)により、時系列予測する。また、社会的な事象の評価方法としては、ニュース記事を言語モデルBERT(Bidirectional Encoder Representations from Transformers)を用いて感情分類し、「感情らしさ」として評価する。これらの手法の概要と初期結果について報告する。

2. システムの概要

本研究では、大規模言語モデルで作成した架空の企業ニュースと、感情スコア(失望、楽観、懸念、興奮、安定)を付与したデータセットを用いてBERTの転移学習を実施した。

表1に二つの見出しの例を示した。プラスな印象を受ける見出しとして、「【速報】世界が注目するMVIDIAが決算発表『最終的な利益 前年比7.3倍 2兆3300億円』勢い止まらず」、マイナスな印象を受ける見出しとして「UUスチール買収計画が窮地に 鉄鉄、訴訟も視野」を推論した結果を示した。プラスの見出しでは楽観と「興奮」が高く、マイナスの見出しでは「懸念」と「失望」が高く出たため、モデルは期待通りの傾向を示した。

また、LSTMの説明変数として5つの感情パラメータの他にNASDAQ100指数、S&P500指数、恐怖指数を採用し、目的変数を各株式の一日の終値とした。

図1にシステム全体の概要図を示す。

表 1: 推論結果例

プラスの文章				
失望	楽観	懸念	興奮	安定
0.0534	0.395	0.109	0.231	0.212
マイナスの文章				
失望	楽観	懸念	興奮	安定
0.254	0.110	0.351	0.135	0.150

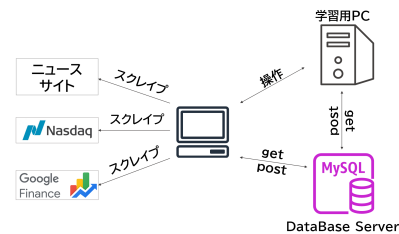


図 1: システム全体の概要図

3. 結果

ウェブスクレイピングにより取得した約500社データを個別に学習したモデルのMAE(Mean Absolute Error)の分布を図2に示す。test_maeの中央値は0.059と良好に見えるが、これはデータ不足による過学習やデータ収集時期が全体的に上昇傾向だったことに起因すると考える。

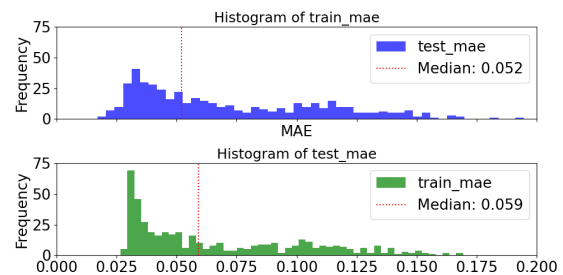


図 2: MAE の分布

4. 結言

本研究では、ニュースから得られる感情情報を活用し、株価推移を予測するシステムを構築した。今後は、評価方法の改善等を通じて、実用化を目指す。

お問い合わせ先

氏名：高田 拓

E-mail：takada@metro-cit.ac.jp