

感情分析と時系列予測を統合したデータ駆動型システムの開発

— 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, 機械学習, 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ます。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ます。本報告では、PBL科目での取り組みについて紹介します。

社会的な事象に影響を与えることが多いのは文字列や映像に関連することが多い一方で、一般的な時系列予測モデルは数値データを入力とする場合がほとんどです。

そこで、本研究ではニュースの文章をBERT(Bidirectional Encoder Representations from Transformers)という自然言語処理モデルを利用して数値化し、「感情らしさ」として活用しました。その上で、時系列予測に特化したLSTM(Long Short Term Memory)を利用して、株価の推移を学習し、予測しました。

2 BERTの転移学習

BERTは、Googleによって開発された自然言語処理モデルであり、転移学習を行うことで特定のタスクに適応させることができます。

本研究では、ChatGPTを用いて作成した架空の企業ニュースと、それに応じた感情スコア(失望、楽観、懸念、興奮、安定の5つのパラメータ)を付与したデータセットを用いて、転移学習を行いました。

転移学習後のBERTモデルの推論例を表1に示します。プラスの文章として『【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」、マイナスの文章として「UUスチール買収計画が窮地に鉄鉄、訴訟も視野」という架空のニュースを用いました。プラスの文章では「楽観」と「興奮」、マイナスの文章では「懸念」と「失望」が高く出ており、期待した傾向を持つモデルを作成できたと考えられます。

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

表 1: 推論結果例

3 実験及び考察

本研究では、S&P500に含まれる約500社の株式データを対象にモデルの学習を行いました。

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ました。今後は、評価方法の改善やモデルのハイパーパラメータ調整を通じて、予測精度の向上を目指します。

お問い合わせ先

氏名：高田 拓

E-mail：takada@metro-cit.ac.jp