

ニュースデータによる深層学習モデルを用いた 日経平均株価変化率予測モデルの提案

Proposal for the Forecasting Model of the Rate of Change in the Nikkei Stock Average
using Deep Learning by News Source

マネジメント工学専攻修士課程 秋葉研究室 2491009 猫本遼太

1. 研究の背景

近年,日本では老後 2000 万円問題や低金利といった社会問題により,個人資産形成の必要性が高まっている [1].その主な手段に株式投資等の金融商品の選択がある.そこで本研究では日本経済の動向を測る指標である日経平均株価を予測する.これにより日本経済の先行きを評価し,資産形成に活用すべき金融商品を特定のための指針となる.

2. 研究の目的

先行研究 [2]では LSTM や CNN を用いて過去の株価推移から日経平均株価の予測を行った.この中でパラメータチューニングとミニバッチ学習を用いることで学習速度や精度,安定性の向上が確認された.しかし,急激な株価変動への対応や予測のランダムウォーク抑制,予測精度に課題があった.これらの課題に対して,自然言語処理手法の SS CDV を使用し,株価の急激な変動への対応力の向上を図る.SS CDV は事前に定義された感情辞書を用いることで,トピックと感情の混合クラスタベクトルを生成する.これによりトピックと感情情報の共同使用が可能になる [3].本研究では NewsAPI [4]から取得したニュースデータに SS CDV を適用して得た分散表現と日経平均株価四本値の変化率を共に LSTM や CNN へ学習させ,日経平均株価終値の変化率の予測精度の向上を目指す.

3. 研究方法

本研究の提案手法を説明する.まず,株価とニュースデータを取得する.前者は変化率に変換し,後者は前処理として正規化や形態素解析,ストップワード除去を行う.次に前処理を行ったニュースデータに対して SS CDV を用いて分散表現を獲得し,主成分分析を用いて次元削減を行う.次に株価変化率と統合し,特徴量を生成する.最後に生成した特徴量を訓練と検証,テストデータに分割する.

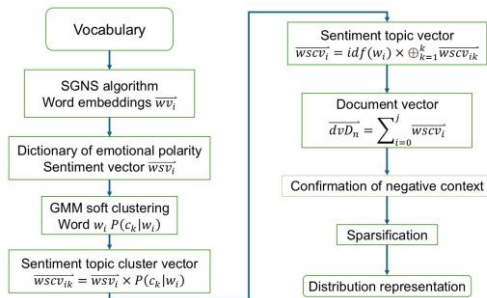


図 1 SS CDV アルゴリズム

本研究の検証では10日分の特徴量を LSTM や CNN に学習させ,11日目の終値を予測する.この時,損失関数と正解率にて予測精度を評価する.図1に SS CDV アルゴリズムの流れを示す.

4. 研究結果

表1に a:パラメータチューニング後, b:ミニバッチ学習導入後における CNN を用いた各最適化アルゴリズムの最高正解率の値を示す.本研究では正解条件として,予測値と正解ラベルの変動方向が一致している場合,正解とみなす.まず,最も高い正解率は Radam の 56.7%となった.次に手法 a と b を比較すると Radam は手法 a, SGD と Adadelata は手法 b の正解率が高くなり,それ以外のアルゴリズムは正解率に変化が見られなかった.また,いずれのアルゴリズムも 50%より高い正解率を示した.しかし,株価のみを用いた場合と比較して正解率を大きく向上させることは叶わなかった.

表 1 CNN を用いた場合の最高正解率

	a	b
SGD	50.3%	53.8%
RMSprop	53.3%	53.6%
Radam	56.7%	51.3%
AdamW	54.9%	53.8%
Adam	53.8%	54.0%
Adadelata	52.4%	54.8%

5. 今後の予定と課題

本研究で生成した特徴量からは汎化性能を持つモデルを作成することができなかった.改善の為に日経平均株価だけでなく Topix や対円米ドル,ダウ平均株価といった別の株価指標を新たに特徴量に加えて共に学習させる手法が考えられる.また,ニュースデータから得られる事実や予測に関する情報だけでなく市場参加者のセンチメントの情報をより多く取り込めるソースを選択することも汎化性能の強化につながる可能性がある.

参考文献

- [1]金融庁,金融審議会,市場ワーキング・グループ報告書「高齢社会における資産形成・管理」
- [2]猫本遼太,深層学習を用いた日経平均株価の予測手法の改良ーミニバッチ学習の導入ー,千葉工業大学2023年度卒業研究(2024)
- [3]上田健太郎,安本慶一,金融市場予測のための Web 金融掲示板投稿文書分散表現の獲得手法の提案と評価,奈良先端科学技術大学院大学 2022 年度修士研究(2023)
- [4]NewsAPI, <https://newsapi.ai/>