

GPT-2 を用いた回答文生成ならびに LSTM による景気変動予測のテキスト分析

Language generation by GPT-2 and text analysis of economic fluctuations forecast through LSTM

西良浩^{1*} 菅愛子¹ 高橋大志¹

Yoshihiro Nishi¹, Aiko Suge¹, and Hiroshi Takahashi¹

¹ 慶應義塾大学大学院経営管理研究科

¹ Graduate School of Business Administration, Keio University

Abstract: Economic fluctuations have a great impact on business management. In this study, we explore the method to improve the accuracy of the prediction model of the economic fluctuation using the economic watcher survey reported by the Cabinet Office, the Government of Japan. The economic watcher survey includes interviewees' answers in the text and the economic business trend tag for each answer. The classification was attempted through the LSTM model. As a result of the analysis, we found the prediction accuracy could be improved by using texts generated by GPT-2. Further examination of the classification model will be planned.

1 はじめに

企業経営において景気は重要な経済的環境要因の一つである。景気変動に関する分析は数多く報告されており、それらの一つに景気ウォッチャー調査を対象とした分析があげられる[1][2][3]。景気ウォッチャー調査とは、内閣府が 2000 年 1 月から毎月実施しているサーベイ調査であり、収集された回答文は日本の経済と物価の情報を多く含んでいる[1]。回答文は有用な情報を含む一方、回答文の取得には制約がありデータ量は限定的となる傾向がある。

これらを背景とし、本研究では大規模な言語生成モデルを用いて回答文を生成し、分類分析の精度を向上させることを目的とする。

2 関連研究

ニュースと株価変動の関連性を分析した研究は数多く行われている。例えば、ニュース記事を用いて株価変動をテキストマイニングにより分析した研究では、ニュース記事に含まれるファンダメンタルおよびセンチメントに関する情報が株価に反映されている可能性が報告されている[5][6][10]。

景気ウォッチャー調査の回答文を用いた研究では機械学習を用いてテキストマイニングを行った取り組み[1]、主成分分析を通じ、景気の現状判断と単語の関連性を分析した取り組み[2]、深層学習を用いた景気判断モデルの構築を行った取り組み[3]など、いくつかの取り組みが報告されている。

自然言語処理の分野において、文書生成技術に関する研究は活発に行われている[14]。質問に対する回答文の生成を行った取り組み[4]、ニュース記事の生成を行った取り組み[11]、発話に対する回答文の生成を行った取り組み[15]など、いくつかの取り組みが報告されている。

3 データ

景気ウォッチャー調査とは、内閣府が 2000 年 1 月から毎月実施している日本全国を対象としたサーベイ調査である。回答者は約 2,000 人であり、景気の現状および景気の先行きに対して「◎良」、「○やや良」、「□不変」、「▲やや悪」、「×悪」の 5 段階で評価し、景況感の判断理由を短い文書で回答している。景気施策に関する専門的な内容から、回答者が所属する企業の売上の動向や求人状況まで幅広い

* 連絡先: 慶應義塾大学大学院 経営管理研究科 経営管理専攻
〒223-8526 神奈川県横浜市 港北区 日吉 4-1-1
E-mail: nishi_yoshihiro@keio.jp

内容が含まれている[1].

本研究では、2000年1月から2019年5月までに実施された景気ウォッチャー調査のデータを分析の対象とした。具体的には、当該期間の回答文と5段階評価の結果を用いた。表1は、全データの内訳を示したものである。収集した全データの総件数は244,353件であった。

表2は、ランダムサンプリングの結果を示したものである。全データの中から、「◎良」の回答文を5段階のうち最もPositiveな回答、「×悪」の回答文を5段階のうち最もNegativeな回答とし、各1,000件ずつ、計2,000件をランダムサンプリングし、分析に用いた。

表 1: 景気ウォッチャー調査の回答件数

	景気の先行きに対する判断
◎良	4,441
○やや良	52,333
□不変	117,290
▲やや悪	52,935
×悪	17,354
合計	244,353

表 2: ランダムサンプリングした回答件数

	景気の先行きに対する判断
◎良 (Positive)	1,000
×悪 (Negative)	1,000
合計	2,000

4 分析手法

本研究では「景気の先行きに対する判断」の回答文を景気変動予測に関する回答文として用いた。回答文を英訳し、GPT-2を用いて文書生成した。回答文のベクトル化にWord2vecを用い、LSTMによる分類分析を行った。図1に主な分析の手順を示す。

4.1 LanguageAppを用いた回答文の英訳

LanguageApp¹とは開発者向けに提供されている

¹ <https://developers.google.com/apps-script/reference/language/language-app>

Google Apps Script のクラスである。

2019年9月時点において、GPT-2は日本語の文書生成ができない。そのため、Google Apps ScriptのLanguageAppを用いて、「景気の先行きに対する判断」の「◎良」および「×悪」の回答文を全て自動で英訳し、文書生成を行った。

4.2 GPT-2を用いた回答文の生成

GPT-2とは大規模な言語生成モデルである。本研究において使用するGPT-2のモデルは、800万のWebページ(計40GB)を24層のネットワークで、およそ3億5,000万個のパラメータを用いて学習している[12][13]。Positiveな回答文とNegativeな回答文を元に、ラベルごとに回答文を生成した。生成する回答文のトークン数(length)は500とした。Positiveな回答とNegativeな回答を1,000件ずつ生成し、計2,000件の回答文の生成を行った。

4.3 Word2vecを用いた回答文のベクトル化

回答文の単語ベクトルの学習にはWord2vecを用いた。Word2vecは文書中の単語の特徴を数値ベクトルで表現できるモデルである[8]。文書中の中心の単語から周辺の単語を予測するモデルをSkip-gramといい、周辺の単語から中心の単語を予測するモデルをCBOWという[9]。本研究では最も広く用いられているWord2vecのSkip-gramモデルを用いた。

4.4 LSTMによる分類分析

LSTM(Long Short-Term Memory)は、時系列データを学習するRNNの一種である。LSTMはRNNを拡張しており、長期的な依存関係の学習を可能としている[7]。分類分析にLSTMモデルを使用し、精度検証には正解率(Accuracy)を用いた。

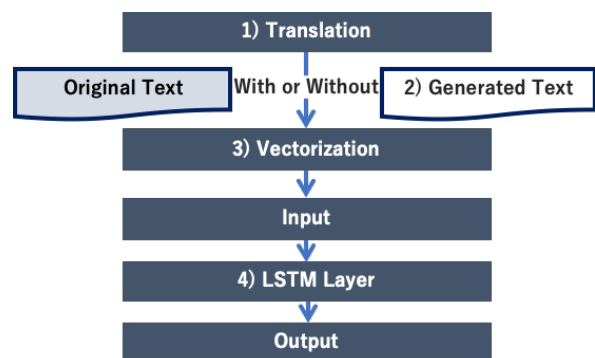


図 1: 分析の手順

5 分析結果

オリジナルの回答文のみをデータセットとして用いたモデルをモデル 1, 生成した回答文を新たなデータとしてデータセットに追加したモデルをモデル 2 とし, LSTM により分類分析を行い, 正解率を比較した.

5.1 回答文の英訳

LanguageApp を用いて「◎良」および「×悪」の回答文を全て自動で英訳した. 例として, 「◎良」の「新商品の発売で通常より販売数が増えると思われる。」という回答文は "The number of units sold is expected to increase as a result of new product launches." と英訳された.

5.2 回答文の生成

表 3 は, 各モデルのデータセットの内訳を示したものである. GPT-2 を用いて Positive な回答文と Negative な回答文を各 1,000 件, 計 2,000 件生成し, 新たなデータとしてモデル 2 のデータセットに追加した.

表 3: データセット

	モデル 1 (オリジナル)	モデル 2 (文書生成)
Positive	1,000	2,000
Negative	1,000	2,000
合計	2,000	4,000

5.3 回答文のベクトル化と分類分析

Word2vec の Skip-gram モデルを用いて回答文をベクトル化し, LSTM により二値分類を行った. 表 4 は, 各モデルの教師データ数とテストデータ数を示したものである. モデル 1 とモデル 2 のデータはともに scikit-learn の train_test_split 関数を用いて, 教師データとテストデータに分けた. テストサイズはどちらも 0.1 とした.

表 5 は, 分類分析の結果を示したものである. LSTM による分類分析の結果, モデル 1 よりもモデル 2 の方が, 正解率 (Accuracy) が 4 ポイント高いことを確認できる.

表 4: 教師データとテストデータの数

	モデル 1 (オリジナル)	モデル 2 (文書生成)
教師データ	1,800	3,600
テストデータ	200	400
合計	2,000	4,000

表 5: LSTM による分類分析の結果

	モデル 1 (オリジナル)	モデル 2 (文書生成)
Accuracy	0.79	0.83

6 まとめと今後の課題

本研究により, 回答文を生成することで, 分類分析の精度を向上できる可能性を見いだした. データ数の増加, 多項分類を用いた分析等は今後の課題である.

参考文献

- [1] 大高一樹, 菅和聖: 機械学習による景気分析 — 「景気ウォッチャー調査」のテキストマイニング, 日本銀行ワーキングペーパーシリーズ, No. 18, (2018)
- [2] 山澤成康: 計量テキスト分析による景気判断-コーディングルールや主成分を使った時系列分析, ESRI Discussion Paper Series, No. 345, (2018)
- [3] 山本裕樹, 松尾豊: 景気ウォッチャー調査の深層学習を用いた金融レポートの指数化, In Proceedings of the The 30th Annual Conference of the Japanese Society for Artificial Intelligence, pp. 1-4, (2016)
- [4] Aishwarya A., Jiasen L., Stanislaw A., Margaret M. C., Lawrence Z., Dhruv B., Devi P.: VQA: Visual Question Answering, In Proceedings of the International Conference on Computer Vision, (2015)
- [5] Fung G. P. C., Yu J. X., Lam W.: Stock Prediction: Integrating Text Mining Approach using Real-time News, In Proceedings of the IEEE International Conference on Computational Intelligence for Financial Engineering, pp. 395-402, (2003)
- [6] Gidófalvi G.: Using News Articles to Predict Stock Price Movements, Department of Computer Science and Engineering, Technical Report University of California, (2001)
- [7] Hochreiter S., Schmidhuber J.: Long Short-Term Memory,

- Neural Computation, Vol. 9, No. 8, pp. 1735-1780, (1997)
- [8] Mikolov T., Chen K., Corrado G., and Dean J.: Efficient Estimation of Word Representations in Vector Space, In Proceedings of the International Conference on Learning Representations Workshop, (2013)
 - [9] Mikolov T., Sutskever I., Chen K., Corrado G., and Dean J.: Distributed Representations of Words and Phrases and their Compositionality, In Proceedings of the NeurIPS, (2013)
 - [10] Mittermayer M. A.: Forecasting Intraday Stock Price Trends with Text Mining Techniques, In Proceedings of the 37th Hawaii International Conference on System Sciences, (2004)
 - [11] Nishi Y., Suge A., Takahashi H.: Text Analysis on the Stock Market thorough "Fake" News Generated by GPT-2, In Proceedings of the INFORMS Annual Meeting, (2019) (to appear)
 - [12] Radford A., Narasimhan K., Salimans T., and Sutskever I.: Improving Language Understanding by Generative Pre-Training, Technical Report OpenAI, (2018)
 - [13] Radford A., Wu J., Child R., Luan D., Amodei, D., and Sutskever I.: Language Models are Unsupervised Multitask Learners, Technical Report OpenAI, (2019)
 - [14] Reiter E., Dale R.: Building Natural Language Generation Systems, Cambridge University Press, (2000)
 - [15] Zhang R., Guo J., Fan Y., Lan Y., Xu J., Cheng X.: Learning to Control the Specificity in Neural Response Generation, In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, pp. 1108-1117, (2018)