

新聞記事を対象とした Sentence BERT を用いた経済政策不確実性の分類

桑名祥平 佐々木稔

茨城大学 理工学研究科 情報工学専攻

23nm717n@vc.ibaraki.ac.jp minoru.sasaki.01@vc.ibaraki.ac.jp

概要

本稿では、Sentence BERT のファインチューニングを用いて、経済政策不確実性に言及のある新聞記事を分類する手法の考案を行った。BERT や RoBERTa のファインチューニングを用いた従来手法のモデルと、本稿のモデルを F 値で比較することにより、提案手法の有効性を検証する。また、経済政策不確実性の有無に関するラベル付けを行う際、従来の単語マッチングによるラベル付けを行ったデータセットと、人手でラベル付けを行ったデータセットの両方を使用し、モデルへの影響度の検証を同時に行う。

1 はじめに

近年、新聞記事や SNS などのテキストデータを経済分野に応用する研究が行われている。本論文では、2016 年に発表された経済指標の経済政策不確実性指数 [1] に着目する。経済政策不確実性 (EPU) とは、特定の国や地域において、将来の経済政策に関する不確実な事象や事例のことであり、これらの不確実性に言及した新聞記事から算出されるのが経済政策不確実性指数である。しかしながら、指標の算出で用いられる新聞記事の収集は人手で行われており、効率的な収集方法が求められる。そこで本稿では、EPU に言及のある新聞記事を、機械学習を用いて自動で分類する手法を提案する。

2 関連研究・関連手法

機械学習手法を用いて、経済に関するテキストを分類する研究は多く行われてきた。関ら [2] は、景気ウォッチャー調査¹の五段階評価と回答理由の文

¹ 内閣府 景気ウォッチャー調査

https://www5.cao.go.jp/keizai3/watcher/watcher_menu.html

章を訓練データとして、経済に関する文章を、BERT を用いて五段階に分類し、その数値を基に新たな経済指標を構築している。Godbole ら [3] は、新聞記事に対して、単語マッチングによる EPU のラベル付けを行ったデータセットを訓練データとして、BERT や RoBERTa, DistilBERT などの複数のモデルでファインチューニングを行うことで、新聞記事の分類を行っている。Chen ら [4] は、EPU のラベル付けを行う際の単語マッチングによる指数への影響を考慮し、ニューラルネットワークによるノイズ除去を行っている。Godbole らの結果を応用して、新聞記事を類似度別に分割して作成したデータセットを用いた場合のモデルへの影響を調査した研究も過去に行った。結果としては、ラベル間の類似度が高い場合は分類がうまくできず、反対に類似度が低い場合は正しく分類できるということが分かった。しかしながら、データセットを作成する段階では、経済政策不確実性に関する単語のキーワードマッチングによるラベル付けを行っていた。これは、従来行われてきた手法であり、定量的なラベル付けであるものの、信憑性に欠けるという難点がある。本稿では、実際に新聞記事に目を通し、ラベル付けを行うことによるモデルの精度への影響を調査すると同時に、Sentence BERT を用いた分類手法を評価する。

3 提案手法

3.1 概要

本稿で提案する分類手法の概略図を図 1 に示す。

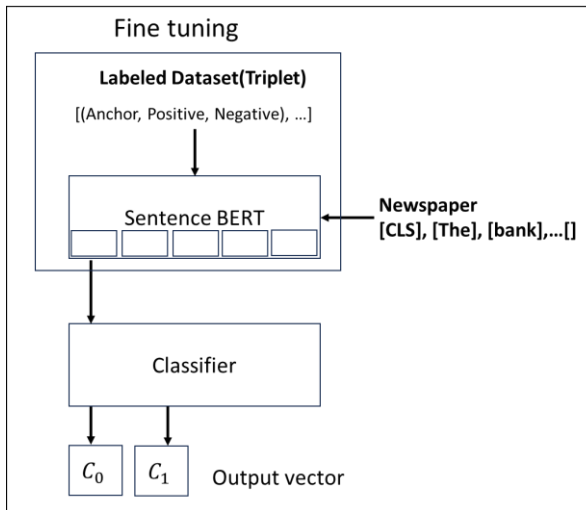


図 1：提案手法の概略図

具体的な処理の流れは、Sentence BERT をファインチューニングした後、Sentence BERT の出力層に分類する層を追加することにより、入力された文新聞記事に対して経済政策不確実性に関する二値分類を行う。

3.2 データセット

本実験で使用した新聞記事は、マイクロソフトが提供している、過去のマイクロソフトニュースのデータセットと、ロイター通信の新聞記事の二つを使用する。いずれも英字の新聞記事で、あらかじめ経済カテゴリの新聞記事のみを抽出している。ファインチューニングにおいて、損失に Triplet Loss を用いるため、データセットでは、Anchor, Positive, Negative からなる Triplet データを用意する必要がある。正負合わせて 100 件を手でラベル付けを行い、組み合わせを変えた Triplet データ約 5000 件をファインチューニング用に使用する。

3.3 ラベル付け

ファインチューニングを行う際には、教師ありデータが必要であるため、新聞記事にラベル付けをする必要がある。Godbole ら[3]は、経済に関する単語、政策に関する単語、不確実性に関する単語それぞれ最低一単語が文章に含まれていた場合に、不確実性ありの”1”、無い場合には”0”のラベルを振っている。これらの単語が含まれていない場合でも、経済政策不確実性に関係する新聞は多くあり、単語のマッ

ングでは抽出することができない。そこで、人手でラベル付けを行うことにより、これらの可能性を排除している。単語のマッチングによるラベル付けを行ったデータセットがどの程度正確なのかを調査するため、人手でラベルを振りなおした時の件数を、以下の表 1 に示す。

表 1：ラベルの件数

	ラベリング前	0	1
ラベリング後		0	1
	0	38	9
	1	12	41

ラベルを振りなおした結果、変更があったのは 100 件中 21 件であった。単語のマッチングによるラベル付けは、約 80% の精度でラベル付けができていたということが分かった。

3.4 Sentence BERT

Sentence BERT[5](図 2)は、2 つの文の埋め込み表現が同時に学習され、これによって文の意味的な関連性がより正確に捉えることができる。これにより、テキストの対比や意味の推論などのタスクで優れた性能を発揮することができ、広く利用されている。処理内容はシンプルで、ある文章と類似する文章を組として学習し、似ている文章同士のベクトルは似たベクトルになるよう BERT をファインチューニングするものである。前述のように、本実験では損失に Triplet Loss を使用するものとする。Triplet Loss では、ある基準となる文章 A を一つ選び、それに類似する文章 P と、そうでない文章 N 二つを組として、特徴量間の距離から損失を計算して学習する。具体的には、アンカーとなる P の文章ベクトルは、P のベクトルに近づけ、N のベクトルとは遠ざけるような処理が行われる。

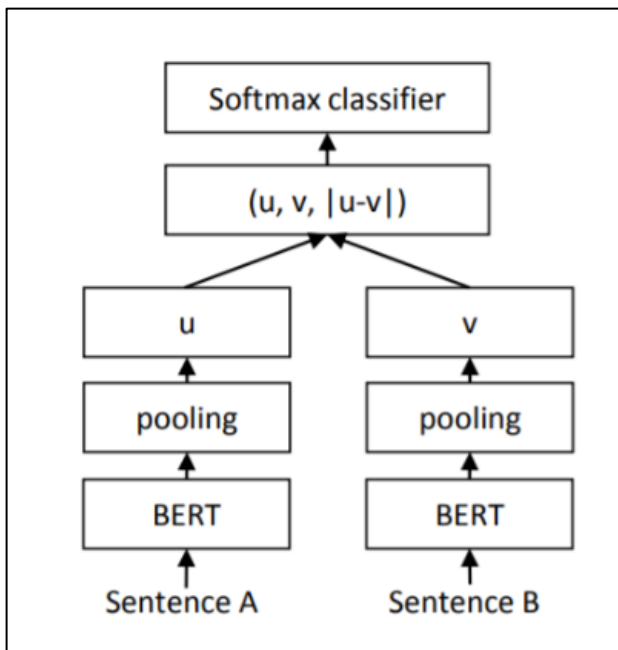


図 2: Sentence BERT²

3.5 分類層

分類器では、ニューラルネットワークの機械学習モデルを使用する。SentenceBERT への入力 is 新聞記事のテキストとなり、出力はベクトルとなる。この出力をニューラルネットワークの分類器に接続し、分類したいクラス分の次元を持つベクトル出力することにより、分類を行う。今回の場合は二値分類であるから、入力層は 768 次元、出力層は 2 次元となる。SentenceBERT から出力されるベクトルは、各トークンのベクトルの平均値である MEAN、各トークンのベクトルの最大値である MAX、系列の最初にあたる[CLS]の三種類存在するが、本実験では[CLS]を分類器への入力としている。分類器の構成は、合計入力層、中間層(100 次元)、出力層の 3 層であり、損失関数はクロスエントロピー損失を用いている。また、中間層の活性化関数には Relu 関数、出力層ではシグモイド関数を用いている。

4 実験

4.1 実験設定

本実験で調査したいことは以下の二点である。

1. 従来の BERT や RoBERTa のファインチューニングしたモデルと、本実験で使用する Sentence BERT のモデルのどちらが有効かを検証すること
2. 単語マッチングによるラベル付けを行ったデータセットと、人手でラベル付けしたデータセットの両方を使用することで、有効なラベル付けの手法を模索すること

項目 2 を調査するにあたり、ファインチューニングに用いるデータセットを二種類用意する。一つは、従来の単語マッチングによるラベル付けを行ったデータセットであり、もう一つは今回新しく作成した人手でラベル付けを行ったデータセットである。

Sentence BERT のファインチューニングのバッチサイズを 8、エポック数を 5 としている。分類層のニューラルネットワークでは、エポック数は 50、学習率は 0.1 としている。

4.2 評価方法

本実験では、実験結果を比較する指標として、テストデータの分類結果の正解率(Accuracy)、適合率(Precision)、再現率(Recall)、F 値を用いる。

5 実験結果

Sentence BERT のファインチューニングにおいて、人手でラベル付けしたデータセットを用いた場合と、従来の単語マッチングによるラベル付けをしたデータセットを用いた場合の実験結果を以下の表 1 に示す。また、過去に行った BERT や RoBERTa などのモデルを使用した場合の結果を比較として示す(F 値のみ)。

² <https://arxiv.org/abs/1908.10084> より引用

表 1：実験結果

	正解率	再現率	適合率	F 値
人手ラベル	0.66	0.73	0.65	0.69
単語マッチングラベル	0.69	0.69	0.71	0.70
BERT				0.78
RoBERTa				0.74

実験結果から、人手でラベル付けを行ったデータセットと、単語のマッチングによるラベル付けを行ったデータセットでは、モデルの分類精度に大きな影響を与えることはないことが分かった。F 値にあまり差異は見られないものの、正解率や再現率、適合率で微妙な差がみられた。また、今回の実験では、新たに SentenceBERT のファインチューニングを用いたモデルを使用した。これまでのモデルと比較すると精度の良いモデルとは言えない結果となった。

5.1 考察

正しくラベルを振りなおしたデータセットを用いた場合と、従来の単語マッチングによる 80% の精度でラベル付けを行ったデータセットでは、見かけ上、分類精度に大きな差は見られない。つまり、これまで誤ってラベル付けされていたデータで、正しく分類されていたと思われるものが正しくラベル付けされたことで、これまで誤って分類されていた件数が減っていると考えることもできる。しかし、ラベル付けを行う時間とモデルの性能を考慮すると、自動かつ定量的なラベル付けの単語マッチングの手法の精度も低いというわけではないことがわかる。今回新たに、Sentence BERT のファインチューニングを用いたが、他のモデルの分類精度よりも低い結果となった。考えられる理由として、ファインチューニングに用いたデータセットの件数が少なく、100 件の記事の組み合わせを変えただけであるため、似たような損失計算が行われてしまったものと思われる。

6 おわりに

本稿では、Sentence BERT のファインチューニングを用いて、経済政策不確実性に言及のある新聞記事を分類する手法の考案と、ラベル付けの手法

の違いによるモデルの精度への影響を調査した。Sentence BERT のファインチューニングの有効性を示すような結果を得ることはできなかった。一方で、人手でラベル付けを行ったデータセットを用いることで、一定の有効性を示すことができた。今後の方針としては、ハイパーパラメータのチューニングや、充実したデータセットの構築、損失関数の見直しなどによるモデルの改良に加えて、抽出した経済政策不確実性に関する新聞記事を使用し、経済指標を構築することをゴールに進めていく予定である。

参考文献

- [1] Baker, S.R., Nicholas, B., Steven, J., & Davis. (2016). Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4), 1593-1636.
- [2] Seki, K., Ikuta, Y., & Matsubayashi, Y. (2021). News-based Business Sentiment and its Properties as an Economic Index
- [3] Godbole, S., Grubinska, K., & Kelnreiter, O. (2020). Economic Uncertainty Identification Using Transformers – Improving Current Methods., Seminar Information System | WiSe 19/20.
- [4] Chen, C, C., Huang, Y, L., & Yang, F. (2023). Semantics matter: An empirical study on economic policy uncertainty index. In *IREF*.
- [5] N. Reimers, I. Gurevych, "Sentence-BERT: Sentence Embeddings using Siamese BERT – Networks ", Proc. EMNLP 2019, pp. 3982–3992, 2019.