

自然言語分析を用いたメディアバイアスの測定²⁾

黒田 敏史

要約

本稿では黒田（2022）で行ったBERTを用いたテキストデータの定量分析手法を用いて、日本のテレビの地上波放送における軽減税率に関するメディアバイアスを測定し、軽減税率に対する日本の地上波放送局の報道内容を定量評価する。テレビの報道内容が軽減税率への賛成論と反対論のいずれに相対的に近いかを、軽減税率導入前後での変化を分析したところ、軽減税率導入後に報道内容がより賛成論に近くなっていたことが明らかになった。

1. はじめに

本研究は黒田（2022）に引き続き、同様の手法を用いてテレビの地上波放送の字幕データを用いたメディアバイアスの測定を行う。黒田（2022）との主要な違いは、第一に新たに購入したデータを利用して分析期間を延長したこと、第二に軽減税率に関する賛否の参照文章を飯田・荻上チキ（2015）の経済学者による賛成論・反対論にしたこと、第三にBERTの演算にGPUを用いるようにしたこと、第四にメディアバイアスについて、軽減税率対象企業とグループ関係にある民放と、軽減税率の影響を受けないNHKの間で、軽減税率導入後にバイアスが変化したかについての差の差法による分析を行ったこと、である。

2. 利用したデータ

本研究で用いたのは、黒田（2022）同様に2017年から2018年3月まではガラボン社の提供するガラボンTVを用いて自宅のある東京都において受診が可能な放送局の字幕を自動取得したデータ、並びにガラボン社より購入した2019-2022年の全国の地上波放送局のデータである。分析対象とした放送局は東京都において受信可能であったテレビ朝日、テレビ東京、フジテレビ、日テレ、NHK総合である。NHK教育は分析対象となる税制に関する報道がほとんど無かったため、分析対象外としている。TOKYO MXも字幕のカバー率が低いことから分析対象外としている。

BERTによるメディアバイアスを測定するために、黒田（2022）同様に軽減税率に関する賛成論、並びに反対論の文章と、放送波で放送された番組の字幕のうち消費税に関わる単

語の含まれる文章全体、もしくは消費税に関連する単語の近さをコサイン距離を用いて計り、文章全体の意味もしくは消費税に関わる単語・放送局・年月毎に平均値をとった。消費税に関わる単語として用いたのは、「消費増税」「消費税」「税」「軽減税率」の四つの単語である。さらに、賛成論との平均距離から、反対論からの平均距離を引いて、放送局の報道が軽減税率の賛成論に偏っている度合いをバイアスの測定値として用いた。月間のある放送局において関連する番組がない、もしくは単語が登場しなかった場合があり、データには欠損値が生じる。また、バイアスは変動が大きいため、六ヶ月移動平均も計算した。六ヶ月移動平均を計算する際には、放送局・月間毎の欠損値は直前の値で一度置き換えた後、それでもなお残る欠損値は直後の値で置き換える事で、スムーズな変化をみられるようにした。

3. 経済学者アンケートの賛成論・反対論

軽減税率に対する賛成論、反対論のリファレンステキストとして用いたのは、飯田・荻上(2015)のアンケートである。このアンケートは2015年10月に、メールアドレスを公開している日本経済学会会員1596名に対して「「軽減税率」と負担軽減策に関するアンケート」と題して調査表を送り、176名が回答をしている。筆者も10月31日にアンケートを受け取っているが、回答はしていない。同アンケートの設問と選択肢は以下の通りである。

質問1：消費増税時の負担軽減策としての「軽減税率」の導入についてのご意見をお聞かせください。

賛成 場合によっては賛成 反対

質問2：質問1の理由として該当するものを以下からお選び下さい（複数回答可）

- 低所得者対策として必要
- 景気対策として必要
- 必需品の税率は低くあるべき
- 政治的にやむを得ない選択肢だと思う
- 価格機構を歪める
- 税収が減少する
- 事務負担が大きい
- 低所得者対策として効果が薄い
- その他（自由回答）：

質問3：財務省による還付案など、軽減税率以外の方法で低所得者対策や負担軽減策を講

じるといふ提案があります。あなたが考えるよりよい負担軽減策はどのようなものですか。ひとつ選んで下さい。

- 軽減税率導入が望ましい
- 軽減措置は行うべきではない
- マイナンバー等を使った事後還付案
- 低所得者に限定した現金給付
- 所得を限定しない現金給付
- 低所得者に限定したクーポン給付
- 所得を限定しないクーポン給付
- その他（自由回答）：

質問 4 軽減税率案または消費増税について、重要だと思われる論点がありますでしょうか。ご自由に記入下さい。

「週刊 SPA !」2015 年 12 月 8 日号に結果、並びに自由記述回答が掲載されている。紙面では、それぞれの回答毎に寄せられた自由回答が掲載されていたため、問 1 に対して「賛成」「場合によっては賛成」として寄せられた五つの賛成理由を賛成論の参照文章とした。また、「反対」として寄せられた六つの反対理由を反対論の参照文章とした。

参照文章における軽減税率への賛成論は、増税によるマクロ経済への影響、他の国で導入されている、エビデンスが存在していない、である。また、軽減税率への反対論が述べているのは高所得者優遇であること、レントシーキング活動を招くこと、である。

4. GPU の利用方法

Hugging Face 社による Python 用ライブラリ Transformers では、PyTorch ライブラリを用いて、モデルを GPU に送る事で、容易に GPU による演算を行う事ができる。BERT モデルによるリファレンステキストの評価では、CPU として Ryzen Threadripper 3970X を用いた場合に 11 秒を要したところ、GPU として GeForce RTX 3070 を用いたところ、5.05 秒と 2 倍程度の高速化が達成された。一方、テレビのニュース番組のうち、税制に関する文章だけを保存したファイルのサイズが最も大きかった 2022 年 12 月の字幕データを同一環境・同一プログラムで BERT モデルによって評価したところ、CPU 処理では 328.7 秒を要したが、GPU 処理では 314.8 秒であり、GPU の優位さは僅かであった。同一のプログラムでデータ量が違うのみであるにも関わらず、データ量が多い場合に余り差が出なかった原因は不明である。また、CPU 処理したファイルに比べ GPU 処理したファイルは 1.2% ほど

自然言語分析を用いたメディアバイアスの測定 2

ファイルサイズが大きかった。

GPU はデータサイズが小さいときに高い優位性を持ち、データサイズが大きい場合であってもわずであるが優位さを持つ。さらにプロセッサの導入費用は GPU の方が低い。2019 年 11 月発売の Ryzen Threadripper 3970X は 2023 年 10 月時点では新品は流通していないが、中古市場で 17 万円程度で取引されている。一方、2020 年 10 月発売の GeForce RTX 3070 は 2023 年 10 月時点で新品が 6 万円程度で入手できる。また、Ryzen Threadripper のようなハイエンドデスクトップ PC 用 CPU はそれぞれ専用のマザーボード等の利用が必要になり、CPU のみを交換することは容易ではないが、GPU は大きなサイズのデスクトップ PC では事後的に内蔵することができ、サイズに小さい PC やノート PC で GPU ボックスを後付けすることができる。より性能のプロセッサがより低い価格で入手でき、導入できる環境も制限しないことから、機械学習では GPU 処理のコストパフォーマンスが良いことがわかる。

5. メディアバイアスの測定結果

図 1 は、2017 年 1 月から 2022 年 12 月までの各放送局の消費税に関わる単語が登場した番組におけるメディアバイアスの六ヶ月移動平均（当該月から五ヶ月後までの平均値）をプロットしたものである。放送局の中では NHK がほぼ全期間において最も軽減税率肯定論に近い報道を行っていることがわかる。また、TBS は他の放送局に比べて相対的に当初軽減税率肯定論から遠い言説を行っていたのに対して、2019 年中頃より急激に軽減税率肯定論に近づいていることがわかる。図 2 から図 5 までは消費税に関わる単語についてのバイアスである。税以外については欠損値の多さと変動の大きさから、視覚的に傾向を把握することは困難であるが、BERT を用いることで単一の単語であっても様々な意味を持って用いられていることがわかる。

図 1 文章全体のメディアバイアス

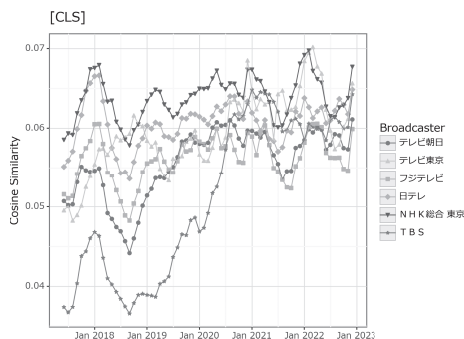


図 2 軽減税率のメディアバイアス

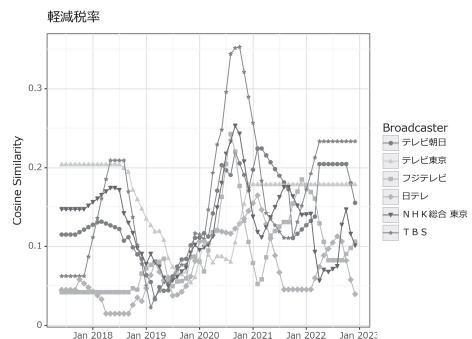


図 3 消費税のメディアバイアス

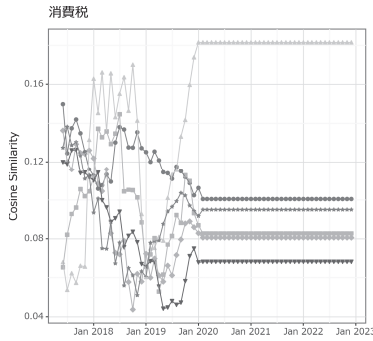


図 4 消費増税のメディアバイアス

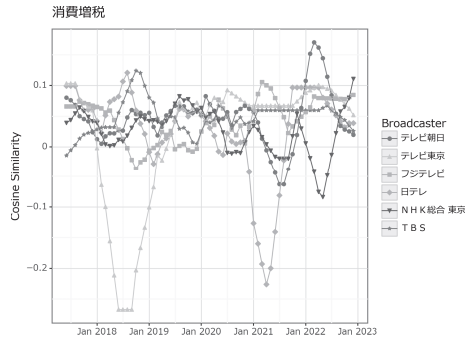
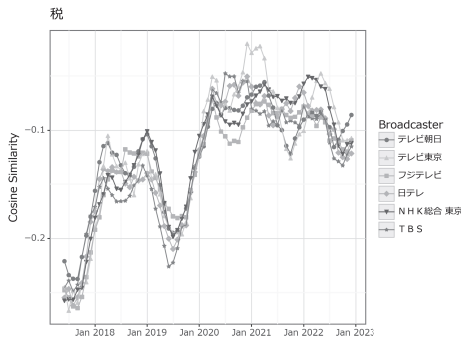


図 5 税のメディアバイアス



6. 差の差法による分析

視覚的にバイアスを把握する事は困難であることから、検定統計量によって定量的に傾向の把握を行うこととする。以下の表 1 はメディアバイアスを放送局ダミー、年月ダミー、軽減税率導入以降ダミーと軽減税率関連企業ダミー（民放で 1, NHK で 0）の交差項に回帰した結果である。なお、この定式化は差の差法と呼ばれている方法と同一であるが、税以外については図から明らかに事前のトレンドが異なっていることから、この推定結果が軽減税率導入による因果効果を捉えていると解釈することはできない。また、合成対照群法による合成コントロールの作成も試みたが、良い合成対照群法がつかれなかったため結果は割愛する。

表 1 差の差法の推定結果

		CLS	消費増税	税	軽減税率
軽減税率導入以降×民放	推定値	0.005	0.017	-0.004	0.011
	標準誤差	0.001	0.024	0.009	0.032
観察数		432	215	430	150
決定係数		0.655	0.379	0.807	0.493

※放送局ダミー・年月ダミーの推定結果は割愛

軽減税率導入以前に比べ、民放は文章全体の意味は軽減税率導入以降により軽減税率肯定論に有意に近くなっているが、各単語については有意な変化はない。第三節で述べたように、軽減税率導入賛成論者の主要な主張の一つは、増税によるマクロ経済への影響である。軽減税率は効率性や再分配効果において望ましくないことが知られているが、軽減税率反対論へ対抗するためには、そのようなミクロ経済学の専門的知識が無ければ理解が難しい性質ではなく、政府支出の増大によって総消費が増加し、各個人の所得が増加する効果があることをまず主張すべきであったかもしれない。

注

- 1) 本研究は、2022年度の東京経済大学個人研究助成費（研究番号 22-06）を受けた研究成果である。

参考文献

- 黒田敏史（2022）「自然言語分析を用いたメディアバイアスの測定：地上波放送テレビの字幕データの例：研究ノート」東京経大会誌 315号
 飯田泰之・萩上チキ（2015）「経済学者に聞いた！ “軽減税率”に関するアンケート」週刊「SPA!」2015年12月8日号