

# ニュースキュレーションメディアデータを用いた ユーザのイデオロギーの特定とそのユーザ閲覧行動傾向分析方式

中西 崇文<sup>†</sup> 関 喜史<sup>‡</sup> 茨木 瞬<sup>††</sup> 阿山 晴取<sup>‡‡</sup> 羅 旋<sup>††</sup> 加藤 創太<sup>††</sup>

<sup>†</sup> 武蔵野大学データサイエンス学部 〒135-8181 東京都江東区有明 3 丁目 3-3

<sup>‡</sup> 株式会社 Gunosy 〒107-6012 東京都港区赤坂 1-12-32 アーク森ビル

<sup>††</sup> 東京財団政策研究所 〒106-6234 東京都港区六本木三丁目二番一号 六本木グランドタワー34 階

<sup>‡‡</sup> 株式会社フィリップス・ジャパン 〒108-8507 東京都港区港南 2-13-37 フィリップスビル

E-mail: <sup>†</sup> takafumi.nakanishi@ds.musashino-u.ac.jp, <sup>‡</sup> yoshifumi.seki@gunosy.com,

<sup>††</sup> {ibaragi, luo, s-kato}@tkfd.or.jp, <sup>‡‡</sup> apid1993@gmail.com

**あらまし** 本稿では、大規模なニュース配信サービスのデータを対象として、左右のイデオロギーを持つユーザのニュース閲覧行動の傾向の分析手法について示す。メディアの報道とユーザのイデオロギーとの関係 (Media effect) に関する研究は特に政治学分野において多くの研究がされているが、特定のユーザのメディア接触に関する研究の多くは実験やアンケート調査であり、実社会におけるメディアの閲覧行動を用いた研究は少ない。本方式は、特定の政治イベントに関する記事の内容から記事のイデオロギーを推定し、左右のイデオロギーを持つ記事に対するユーザの反応からユーザのイデオロギーを特定することで、左右のイデオロギーを持つユーザのニュース閲覧行動の傾向を示すことを可能とする。本方式によって、記事へのクリックなどユーザのニュース閲覧行動により、イデオロギーを持つユーザが自身のイデオロギーに近い記事を多く閲覧するという「選択的接触」の傾向があることを明らかにした。

**キーワード** ユーザ行動分析, アクセスログ, イデオロギー推定, 閲覧行動傾向

## 1. はじめに

コロナ禍である現在において、Go To トラベルなど、経済を優先した政策をすべきか、または緊急事態宣言などの人々の健康を優先した政策をすべきか、といった議論が多くあるが、文献[1]が示すように、各政策への支持は個々人のイデオロギー (政治思想) によって違いがあると考えられる。例えば、保守寄り (右寄り) のイデオロギーを持つ人々は主に経済優先の政策を支持し、リベラル寄り (左寄り) の人々は主に健康を優先した政策を支持する。このように、個々人のイデオロギーは政策への支持や個々人の行動に影響を与えることである。

ユーザ (または有権者) のイデオロギーに影響を与える要素として、メディアの報道が挙げられる。報道するメディアのバイアス (Media slant) に関する研究が政治学において多くなされており、新聞の記事内容から Media slant を測る研究[2][3]や Twitter などの SNS を用い、アカウントを持つメディアや政治ジャーナリストの slant を測定した研究[4][5][6]などが知られている。

また、メディアの報道とユーザのイデオロギーとの因果関係を示す研究も多くあり、メディアの報道がユーザのイデオロギー位置に影響を与えるとする「メディア効果 (Media effect)」に関する研究[7][8]がある一方、ユーザが自身のイデオロギーに近いメディアのみを選択する「選択的接触 (selective exposure)」に関する

一連の研究[9][10]もある。

しかしながら、ニュースのデータにおいて閲覧行動は見られるが、ユーザのイデオロギーの特定が難しい。特定のユーザのメディア接触に関する研究の多くは実験やアンケート調査であり、実社会におけるメディアの閲覧行動を用いた研究は[11][12]など少数である。文献[11]は Facebook のデータを用いることでとユーザのイデオロギーを特定することが簡単にできるが、閲覧行動を見るのが困難である。加えて、自己申告によりイデオロギー特定しているため、自身のイデオロギーを隠すための虚偽の申告などの可能性がある。

そこで本稿では、株式会社 Gunosy が提供するニュース配信サービスであるグノシーのデータを対象として、ユーザのイデオロギーを特定し、閲覧行動を見ることも可能にするための新しい手法について示す。本方式は、特定の政治イベントに関する記事の内容から記事のイデオロギーを推定し、左右のイデオロギーを持つ記事に対するユーザの反応からユーザのイデオロギーを特定することにより、自己申告などのプロフィールデータを用いることなく左右のイデオロギーを持つユーザのニュース閲覧行動の傾向を示すことを可能とした。また本方式によって、記事へのクリックなどユーザのニュース閲覧行動により、イデオロギーを持つユーザが自身のイデオロギーに近い記事を多く閲覧するという「選択的接触」の傾向があることを明らかにした。

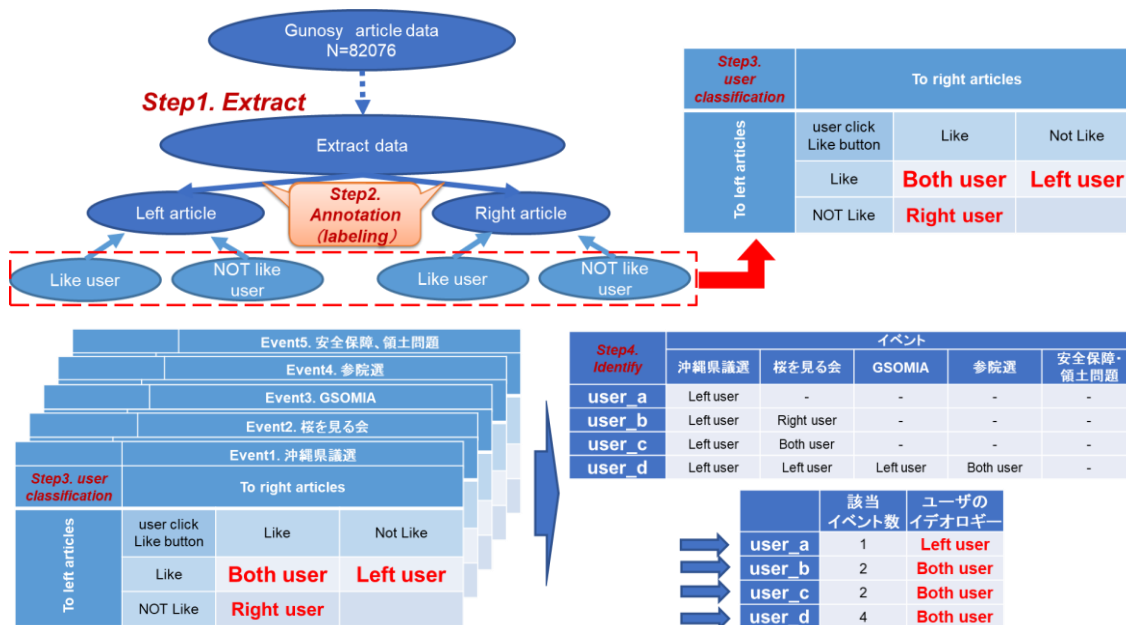


図 1：提案方法の流れ

本方式の貢献は、イデオロギーという特定の嗜好を持つグループの行動を見ることにより、特定のイベントにおける人々の行動の違いを明らかにしたり、推測したりすることが可能になることである。また、イデオロギーの特定により、イデオロギーの強いユーザに反対意見の記事を出すことが出来るなど、推薦システムにおける多様性を担保することが出来ることも貢献である。

本稿は以下のように構成される。2 節では、本方式で用いたデータセットについて述べる。3 節では、ユーザのイデオロギーを特定するための方式を示し、4 節では本方式を用いた実験結果を示す。5 節では、特定したユーザのイデオロギーを用いたユーザの閲覧行動の分析について示す。6 節で、これらの研究成果についてまとめる。

## 2. 分析対象データ

本分析では、株式会社 Gunosy が提供する情報キュレーションアプリ「グノシー」のデータを用いる。

本方式で「グノシー」を用いる利点は 3 点ある。1 点目は 2021 年 1 月時点でダウンロード数が 3,700 万を超える大規模なニュースサービスであり、特定のカテゴリのデータであってもユーザの閲覧行動を分析するための十分なサンプルを確保することが出来る点である。2 点目はユーザが積極的に意向を発信している SNS のサービスと異なり、受動的な体験を分析できる点である。自身のイデオロギーに近い記事と遠い記事の双方を閲覧できることによって、本方式にてイデオロギーの推定が可能となる。3 点目はクリックだけではなく

Like を用いることが可能な点である。「グノシー」ではユーザは記事のタイトルを見てクリックし、内容を見て Like をクリックする、という手順で記事内容を閲覧する。この Like を用いることでイデオロギーを特定し、クリックを用いて閲覧行動を分析することが可能となる。

## 3. イデオロギー推定方式

### 3.1. イデオロギー推定方式全体像

本方式は、特定の政治イベントに関する記事の内容から記事のイデオロギーを推定し、左右のイデオロギーを持つ記事に対するユーザの反応からユーザのイデオロギーを特定することで、左右のイデオロギーを持つユーザのニュース閲覧行動の傾向を示すことを可能としている。

本方式の実現全体像を示したが図 1 である。図 1 に示すように、本方式は 4 つの Step (Extract, Annotation, User classification, Identify) から構成される。以下において、4 つの Step について詳しく説明する。

### 3.2. Step1: Extract

Step1: Extract では、イデオロギーの判別が可能と思われるイベントの記事を抽出する。人気記事など、多くのイベントが混在するとイデオロギーの判別が難しい。先行研究においても日本の新聞記事における社説を用

表 1: イベント記事抽出キーワード一覧

イベント	沖縄県議選	桜を見る会	GSOMIA	参院選	安全保障、領土問題
キ ー ワ ー ド	沖縄県知事選 or 沖縄知事選 or 普天間 or 辺野古	桜を見る会	ホワイト国 or GSOMIA or 軍事情報包括保護協定	参議院議員通常選挙 or 参議院議員選挙 or 参議院選挙 or 参院選 and 9条（九条） or 憲法改正 or 改憲 or 護憲 or アベノミクス or 増税	日米安保 or 安全保障 or 竹島 or 尖閣

いてイデオロギー推定をしている[3]はトピック抽出を用いてイベントごとに記事を分けてイデオロギーを推定している。そこで本方式では日本で起きたイデオロギーの判別が可能と思われるイベントとして、選挙、外交、国会での話題からイベントを抽出する<sup>1</sup>。

### 3.3. Step2: Annotation

Step2 では、Step1 で抽出したイベントの記事のイデオロギーについて、政治学の研究者である著者がタイトルおよび本文の内容から Annotation<sup>2</sup>を行い、①リベラル寄りの記事を「Left article」、②保守寄りの記事を「Right article」、③中道または両方の内容を含む記事を「Other article」の3つに分類する。この Step により、まずは記事のイデオロギーを特定する。

### 3.4. Step3. User classification

Step2 での Annotation により、左右に分類された記事に対し、各ユーザが Like を行ったか否かにより、ユーザを①リベラル寄り記事のみに Like をしたユーザ「Left user」、②保守寄りの記事のみに Like したユーザ「Right user」、③左右両方の記事に Like をしたユーザ「Both user」の3つに分類する。

本方式ではユーザ分類においてクリックではなく Like を用いる理由は以下の通りである。先述のように、「グノシー」におけるユーザの閲覧行動の流れとして、一覧で表示されているタイトル見てクリックし、クリックした記事の本文の内容を見て Like をする、という2段階がある。文字数制限があるタイトルでは記事のイデオロギーをユーザが判断しにくい一方、本文の内容を見て判断する Like であればユーザのイデオロギーをより強く反映していると考え、ユーザ分類においては Like を使用している。

### 3.5. Step4. Identify

Step3 においてイベントにおけるユーザのイデオロギ

ーを推定しているが、1つのイベントにおけるイデオロギーであり、そのイベントのみの結果かもしれない。そこで Step4 では、Step3 で分類した各イベントにおけるユーザのイデオロギーについて、他のイベントで Like をした記事もリベラル寄りまたは保守寄りなのかを分析し、本提案方式におけるリベラル寄りのユーザ「Left user」、保守寄りのユーザ「Right user」を特定する。

図1が示すように、本方式では該当するイベント数に関係なく、Left user にのみ該当するユーザを本方式における「Left user」、Right user にのみ該当するユーザを本方式における「Right user」として特定する。一方で、該当するイベント数に関係なく、Left user, Right user, Both user のうち2つ以上に該当するユーザを本方式における「Both user」として特定する。

例えば、図1における user a は1つのイベントのみにしか該当していないが、そのイベントでは「Left user」であったため、本方式における「Left user」と特定するが、user d のように4つのイベントに該当しており、内3つのイベントが「Left user」であっても残りのイベントが「Both user」であったため、user d は本方式における「Both user」と特定している。

## 4. 実験

### 4.1. 実験方法

4節の方式を用いて実際にユーザのイデオロギーを特定する。対象は2018年8月1日～2019年12月27日までの政治カテゴリに属する記事82,076件と、それらの記事に対するユーザのクリック、及び「いいね」を示す Like データを用いる。

### 4.2. Step1: Extract

Step1: Extract では対象期間におけるイデオロギーの判

<sup>1</sup> 遠藤・ジョウ 2019 [14] は日本の有権者のイデオロギーを規定しているのは外交および安全保障政策としている。

<sup>2</sup> 著者らは既に自然言語処理を用いた記事分類およ

びユーザ分類を行っているが、精度が高くないことや教師データの信頼性の問題から、Annotation を選択した。

表 2：イベント別対象記事データ数

イベント	沖縄県議選	桜を見る会	GSOMIA	参院選	安全保障、領土問題	合計
抽出データ数	4,456	1,582	479	2,548	4,025	13,090
Like数10以上	177	191	132	143	254	897

表 3：イベント別記事のイデオロギーおよび割合（左右のみ）

※赤字は各イベントにおいて左右のカテゴリのうち、該当割合が多いカテゴリ

Event / article	Left_article	Right_article	Other_article	N	N (LR)	Left%	Right%
沖縄県議選	62	31	84	177	93	<b>66.67%</b>	33.33%
桜を見る会	80	10	101	191	90	<b>88.89%</b>	11.11%
GSOMIA	13	46	73	132	59	22.03%	<b>77.97%</b>
参院選	63	29	51	143	92	<b>68.48%</b>	31.52%
安全保障・領土問題	25	111	118	254	136	18.38%	<b>81.62%</b>
総計	243	227	427	897	470	51.70%	48.30%
ユニーク数	211	190	367	768	401	52.62%	47.38%

別が可能と思われるイベントとして、沖縄県議会議員選挙（以下、沖縄県議選）、桜を見る会に関する問題（以下、桜を見る会）、日韓貿易からの GSOMIA 破棄問題（以下、GSOMIA）、2019 年参議院議員通常選挙（以下、参院選）、安全保障・領土問題の 5 つのイベントに限定した。

ただし、8 万以上の記事を 1 つずつ確認してイベントを抽出することは困難である。そこで、5 イベントを抽出するためのキーワードをリスト化し、当てはまる記事を抽出した。5 イベントを抽出するためのキーワードは表 1 の通りである。選挙がテーマである沖縄県議選および参院選は争点を中心にキーワードとしたが、参院選について、「憲法改正」や「アベノミクス」は選挙以外でも話題となっているため、「参議院議員選挙」などの名称と掛け合わせて抽出するようにした。

キーワードに沿って 5 イベントに該当するデータを抽出した結果を示したものが表 2 である。抽出時点では 13,090 記事であるが、ユーザのイデオロギー特定（Step3）にて記事に対する Like 数によってユーザのイデオロギーを推定するため、抽出したデータにおいて、さらに「Like 数が 10 以上」に限定し抽出した（表 2「Like 数 10 以上」）。その結果、本分析に使用するデータ数は 897 記事（ユニーク数は 768 記事）<sup>3</sup>となった。

### 4.3. Step2: Annotation

Step1 で抽出した 5 イベントの記事のイデオロギーについて、政治学の研究者である著者がタイトルおよび

本文の内容から「Left article」、「Right article」、「Other article」の 3 つに分類した結果を示したものが表 3 である。5 イベントのうち、沖縄県議選、桜を見る会、参院選においてはリベラル寄りの記事が多く、GSOMIA および安全保障・領土問題については保守寄りの記事が多いことが示されている。この結果の要因として、政権与党である自由民主党が保守寄りのイデオロギーを持つ政党であることが挙げられる。与党への疑惑の追及など、選挙や桜を見る会は反与党の記事が多く、一方で外交問題である GSOMIA や安全保障・領土問題は、日韓貿易や尖閣諸島などの領土問題に対する相手国の対応に対する反発により、保守寄りの記事が多く見られたと考えられる。

### 4.4. Step3. User classification

Step2 での Annotation により、左右に分類された記事に対するユーザ Like の有無により、ユーザを「Left user」、「Right user」、「Both user」の 3 つに分類した。

ユーザ分類の結果を示したものが表 4<sup>4</sup>である。各イベントの該当割合は記事（表 3）と同じく沖縄県議選、桜を見る会、参院選がリベラル寄り、GSOMIA および安全保障・領土問題が保守寄りのユーザが多いが、全体的に保守寄りのユーザの該当割合が記事の該当割合と比べて多く、5 イベントの総数で比較すると、記事（表 3）が 52:48 とほぼ同数だったのに対し、ユーザ（表 4）は保守より（Right%）が 62.03%と比較的多いことが確認された。この結果は「グノシー」において、保守寄りである Right user に分類されるユーザの方が

<sup>3</sup> イベント間で重複している記事もあるため、ユニーク数も表示している。

<sup>4</sup> イベント間で重複しているユーザがいるため、ユニークユーザ数は総計の 10,669 名ではなく、7,348 名である。



表 4：イベント別ユーザのイデオロギーおよび割合（左右のみ）

※赤字は各イベントにおいて左右のカテゴリのうち、該当割合が多いカテゴリ

Event / user	Left_user	Right_user	Both_user	N	N (LR)	Left%	Right%
沖縄県議選	1,284	898	99	2,281	2,182	<b>58.85%</b>	41.15%
桜を見る会	796	120	18	934	916	<b>86.90%</b>	13.10%
GSOMIA	156	1,865	75	2,096	2,021	7.72%	<b>92.28%</b>
参院選	1,266	590	50	1,906	1,856	<b>68.21%</b>	31.79%
安全保障・領土問題	414	2,924	114	3,452	3,338	12.40%	<b>87.60%</b>
総計	3,916	6,397	356	10,669	10,313	37.97%	<b>62.03%</b>

表 5：イベントの該当数別ユーザ分類

		Multiple events				
	合計	1	2	3	4	5
unique_user	7,348	4,782	2,018	380	129	39
left_user	2,792	2,175	560	42	15	0
right_user	4,076	2,577	1,292	169	32	6
both_user	480	30	166	169	82	33

比較的多く Like をすることが示唆される。

#### 4.5. Step4. Identify

Step4 において、5 イベントで少なくとも 1 つのイベントで Like をしているユーザのうち、すべての Like が Left article に対してであるユーザを本方式における「Left user」とし、すべての Like が Right article に対してであるユーザを本方式における「Right user」とした。

ユーザのイデオロギー特定の結果を示したものが表 5 である。ユニークユーザ 7,348 名のうち、93.5%にあたる 6,868 名が Left user または Right user に分類された。

本節では、3 節で示した方式により、5 つのイベントの記事およびユーザを 3 つのカテゴリに分類した。その結果、Like10 以上の 5 つのイベント記事データから 211 の記事を Left article とし、190 の記事を Right article とした。また、イデオロギーを特定した記事から 2,792 名のユーザを Left user、4,076 名のユーザを Right user と特定した。

5 節においては主にイデオロギーが特定された 6,868 名のユーザの閲覧行動を分析する。

### 5. ユーザの閲覧行動分析

#### 5.1. 分析内容

本節では、前節の方式でイデオロギーを特定したユーザのデータを用いて、イデオロギーの違いによるユーザ閲覧行動の違いについて分析を行う。具体的には、特定のユーザの Left article または Right article へのク

リック率を分析することで、Left user (または Right user) が Left Article (または Right article) にだけにクリックするかという、選択的接触に関する分析を行う。

#### 5.2. 分析結果

表 6 は特定のユーザごとにイデオロギー分類をした記事に対する平均クリック数を算出したものである。表 6 より、Left user は Right article の約 2 倍 Left article をクリックしていることが示されており、Right user にも同様に Right article の平均クリック数が Left article の約 2 倍となっている。3.4 節で述べたように「グノシー」におけるユーザの閲覧行動の流れとして、タイトル見てクリックし、本文の内容を見て Like をする、という 2 段階があるため、Like によりユーザのイデオロギーを特定している本方式では同じイデオロギーの記事の方が多くクリックされる懸念はあるものの、クリック行動の時点で Left user (または Right user) が Left article (または Right article) にだけクリックするという選択的接触の傾向が示唆されている。また、Both user の平均クリック数が Left article と Right article でほぼ変わらないことも本方式が信頼できる根拠とも言えると考えられる。

表 6：ユーザ別平均クリック数

Click / Article	left_article N=211	right_article N=190	other_article N=367
left_user N=2,792	71.93	39.16	57.39
right_user N=4,076	58.4	114.77	99.25
both_user N=480	27.94	28.26	30.27

## 参 考 文 献

一方で、Left user と Right user は Both user と比べて反対のイデオロギーの平均クリック率が多い。Left user の Right article への平均クリック率は 39.16 であり、Both user の Right article への平均クリック率 (28.26) よりも 1.39 倍平均クリック率が多い。また、Right user の Left article への平均クリック率は 58.40 であり、Both user の Left article への平均クリック率 (27.94) よりも 2.09 倍平均クリック率が多い。この結果により、イデオロギーの強いユーザは反対の意見も多くクリックする傾向が示唆される。

## 6. おわりに

本稿では、株式会社 Gunosy が提供する情報キュレーションアプリ「グノシー」のデータを対象として、ユーザのイデオロギーを特定し閲覧行動を見ることも可能とする新しい手法として、左右のイデオロギーを持つ記事に対するユーザの反応を用いた方式について示した。本方式によって、記事へのクリックなどユーザのニュース閲覧行動により、イデオロギーを持つユーザが自身のイデオロギーに近い記事を多く閲覧するという「選択的接触」の傾向があることを明らかになった。

今後は「グノシー」のデータの拡張として、ユーザ数や記事の拡張を行うとともに、タイムスタンプや記事での滞在時間など時間に関するデータセットを用いた分析を行っていきたいと考えている。記事での滞在時間を用い、記事の内容を理解して Like をしているのか、それともあまり理解をせずに Like をしているのかの区別ができるため、よりイデオロギー特定の精度が上がると考えられる。

また、時系列データの分析が可能であるため、メディアの報道とイデオロギーの変化に関する時系列分析やユーザの違いによる特定の記事への閲覧時間も踏まえた分析を行うことが可能である。

また、特定のユーザは政治カテゴリ以外の記事にも多くのクリックをしている可能性があるため、政治カテゴリ以外の記事も含んだ分析を行いたいと考えている。

最後に今後の課題を示す。まず本方式において著者が記事のラベル付け (Annotation) を行っているが、文献 [13] のような複数名によるラベル付けを行い、一致度係数の算出など Annotation の精度を上げていきたいと考えている。

また、本方式により分類した Media slant を教師データとした分析も可能であるため、先述の Annotation の精度を上げた後、本方式を教師データとした機械学習による分類も進めていきたいと考えている。

- [1] Jon Green, Jared Edgerton, Daniel Naftel, Kelsey Shoub, and Skyler J. Cranmer (2020), "Elusive consensus: Polarization in elite communication on the COVID-19 pandemic," *Science Advances* 6(28) EABC2717.
- [2] Gentzkow, Matthew. and Shapiro, Jesse M. (2010), "What drives media slant? Evidence from U.S. daily newspapers," *Econometrica* 78: 35-71.
- [3] Kaneko, Tomoki, Takaaki Asano, and Hirofumi Miwa (2020), "Estimating ideal points of newspapers from editorial texts," *The International Journal of Press/Politics*. doi:10.1177/1940161220935058
- [4] Barberá Pablo, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau (2015), "Tweeting from left to right: Is online political communication more than an echo chamber?," *Psychological Science* 26(10): 1531-1542.
- [5] 三輪洋文 (2017) 「Twitter データによる日本の政治家・言論人・政党・メディアのイデオロギー位置の推定」『選挙研究』 33 (1): 41-56.
- [6] Hassell, Hans J. G., John B. Holbein, and Matthew R. Miles (2020), "There is no liberal media bias in which news stories political journalists choose to cover," *Science Advances* 6(14) EAY9344.
- [7] Della Vigna, S. and Kaplan, E. (2007), "The Fox News Effect: Media Bias and Voting," *The Quarterly Journal of Economics* 122 (3): 1187-1234.
- [8] Martin, Gregory J. and McCrain, Joshua M. (2019), "Local news and national politics," *American Political Science Review* 113(2): 372-384.
- [9] Stroud, Natalie Jomini (2011), *Niche news: The politics of news choice*, Oxford: Oxford University Press.
- [10] 辻大介・北村智 (2018) 「インターネットでのニュース接触と排外主義的態度の極性化—日本とアメリカの比較分析を交えた調査データからの検証—」『情報通信学会誌』 36 (2), 99-109.
- [11] Bakshy, Eytan, Solomon Messing, and Lada A. Adamic (2015), "Exposure to ideologically diverse news and opinion on Facebook," *Science* 348 (6239): 1130-1132.
- [12] 栗本真太郎・関喜史 (2020) 「ニュース閲覧行動を用いた選挙の争点分析」2020 年度 人工知能学会全国大会 (第 34 回) 報告論文.
- [13] Tyler, Matthew (2020), "Getting More out of Human Coders with Statistical Models," American Political Science Association 2020 annual meeting.
- [14] 遠藤・ジョウ (2019) 『イデオロギーと日本政治：世代で異なる「保守」と「革新」』新泉社.
- [15] 谷口将紀 2015. 『政治とマスメディア』東京大学出版会.