

DNS クエリデータにもとづく  
ソーシャルメディア利用者の行動分析

筑波大学審査学位論文（博士）

2013

三田村 健史

筑波大学大学院  
ビジネス科学研究科 企業科学専攻



# 論文概要

Twitter や Facebook の様なインターネット上で展開される情報メディアであるソーシャルメディアは、身近で手軽なコミュニケーション手段として広く社会に普及してきている。研究領域においても、2006年ごろからソーシャルメディアという用語が使われ始めた後、その利用拡大が進むにつれて注目が集まり、心理学、社会科学、数理社会学、情報学、マーケティングや経営学などでも研究対象として広がってきている。

本研究では、インターネットの基幹技術の1つであり、Webアクセスや電子メールの送受信時などに利用されるDNS (Domain Name System) のクエリデータを活用する。特に、そのデータにおける“種類の数”である“異なり数”を用いて、“投稿者側”の視点からではなく、“閲覧者側”の視点からソーシャルメディアを通して、実社会で起きる現象に対するソーシャルメディア利用者の行動を分析する手法の提案を試みる。

ソーシャルメディア利用者の行動分析に関する従来手法は、実社会で起きる現象に対するソーシャルメディア利用者の行動を、投稿内容を通して投稿者側の視点から分析するものが主流である。しかし、投稿者側の視点からの分析には、やらせ投稿などを行うステルスマーケッターなどの作為的な投稿者による投稿内容の影響を受けやすいという問題が存在している。この問題に対応するために、ソーシャルメディア利用者の行動を閲覧者側の視点から分析する試みは従来から存在していた。先行研究の主な特徴は、以下の3点である。

- 研究の基本は、投稿者から投稿された情報の分析か、登録制のソーシャルメディアの場合であれば設定された個人のプロパティ情報の分析が主流である。
- 一方で、閲覧者の存在の重要性は認識されており、その行動分析が試みられている。
- 閲覧者の分析は、① (登録制のソーシャルメディアの場合のみ) 設定された閲覧者の個人プロパティ情報を分析する、② (登録制のソーシャルメディアの場合のみ) 「閲覧者数 = 登録者数 - 投稿した人数」、として量的分析を行う、③ 「投稿が少ない人」を「閲覧者」と定義して分析する、④ アンケート調査結果を使用する、⑤ 検索エンジン結果を使用する、が主流である。

一方、先行研究に共通する問題としては、以下の3点がある。

- 主流である投稿された情報の分析は、作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”を抱えている。
- 閲覧者の分析は、ソーシャルメディア利用者に調査票を記入させ回収する必要があるため、大規模な調査には不向きであるという“調査規模的問題”を抱えている。
- 閲覧者の分析は、調査票の記入者を選択する際や調査票を記入する際、または検索エンジンを利用して検索結果を得る際に発生し得る、サンプリングバイアスや反応バイアスの影響を受けてしまうという“利用者バイアス問題”を抱えている。

以上の問題を解決するために、本研究では以下の提案を行った。

- 何人の閲覧者がいたかという情報に対応する DNS クエリデータの“異なり数”を用いて、ソーシャルメディアの“閲覧者側”の視点から、利用者の行動を分析する手法（“作為的投稿影響問題”への対応）。
- DNS の中でも各国ごとに割り当てられた ccTLD（country code Top Level Domain）の1つである JP ドメイン名の DNS クエリデータを用いるによる、大規模で即時的なデータ収集の実現（“調査規模的問題”への対応）。
- インターネットの基幹技術である DNS を用いてデータ収集することで、ソーシャルメディア利用者に対する“利用者バイアス問題”への対応

本研究では、DNS クエリデータを用い閲覧者側から分析することで、実社会で起きる現象を分析する際に発生し得る“作為的投稿影響問題”や“利用者バイアス問題”，“調査規模的問題”に対応することを試みた。また、大規模な実 DNS クエリデータを用いて以下の2つの実証検証を行い、提案手法の有効性を確認した。

#### (1) ソーシャルメディアを利用した予測に関する研究

提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、先行研究が抱える問題に対応できることを確認するために映画の観客動員数予測を試みた。その結果、投稿者側からの分析である従来手法（[Mishne 2006]）が、映画公開前で相関係数 0.454～0.542であったのに対し、提案手法は 0.865であった。これにより、DNS クエリデータを用い、投稿者側の視点からではなく閲覧者側の視点から、ソーシャルメディア利用者の

実社会における行動分析が可能であり、ソーシャルメディアを利用した予測に関する研究領域において提案手法の有効性を確認した。

## (2) ソーシャルメディアを利用した計測・モデル化に関する研究

提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、提案手法が大規模で複雑な行動分析も可能であり先行研究の抱える問題に対応できることを確認し、そのソーシャルメディア利用者の行動の時間的推移も明らかにした。

具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における2つの大きな社会現象に対し、DNSクエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用の関連を計測することで、ソーシャルメディア利用者の行動の時間的推移を分析した。

この結果、日本政府が行った2009年から2011年に及ぶアナログ放送の停止とデジタル放送への移行という政策において、家電エコポイント制度が、その目的の1つであった地上デジタル放送対応テレビの普及で、ソーシャルメディア利用者の行動に効果的な影響を与えた経済政策であったことを示した。これにより、長期的で大規模なDNSクエリデータを用いて、閲覧者側の視点からソーシャルメディア利用者の実社会における行動の計測が可能であり、ソーシャルメディアを利用した計測・モデル化に関する研究領域において、提案手法が大規模で複雑な行動分析も可能であり、先行研究の抱える問題に対応できることを確認した。

本研究は、ソーシャルメディア利用者の行動を把握する上で客観的なデータ計測を可能とし、ソーシャルメディア利用者の行動分析に加え、実社会で起きる現象を分析する手法を提案するものである。



# 目次

第1章 序章.....	1
第2章 ソーシャルメディア利用者の行動分析に関する研究範囲と課題.....	5
2.1 ソーシャルメディアの分類軸.....	5
2.1.1 自己管理可能メディアと自己管理不可能メディア.....	5
2.1.2 ソーシャルメディアの分類軸に関する先行研究.....	12
2.1.3 本節のまとめ.....	15
2.2 ソーシャルメディア利用者の分類軸.....	16
2.2.1 投稿者と閲覧者.....	16
2.2.2 ソーシャルメディア利用者の分類軸に関する先行研究.....	18
2.2.3 本節のまとめ.....	22
2.3 ソーシャルメディア利用者の行動分析に関する先行研究.....	23
2.3.1 ソーシャルメディア利用者の行動分析に関する先行研究の分類.....	25
2.3.2 ソーシャルメディアを利用した定義・分類に関する研究.....	27
2.3.3 ソーシャルメディアを利用した計測・モデル化に関する研究.....	32
2.3.4 ソーシャルメディアを利用した予測に関する研究.....	39
2.3.5 ソーシャルメディアを利用した信頼性評価に関する研究.....	44
2.3.6 ソーシャルメディア利用者の行動分析に残された研究課題.....	49
2.3.7 本節のまとめ.....	53
2.4 ソーシャルメディア利用者の行動分析に関する本研究の貢献.....	54
第3章 閲覧者視点からのソーシャルメディア利用者の行動分析手法.....	57
3.1 システム構成.....	58
3.2 延べ数と異なり数.....	60

3.2.1 異なり数を用いた解析技術.....	60
3.2.2 本研究における延べ数と異なり数.....	61
3.3 DNS キャッシュの影響と異なり数.....	63
3.4 閲覧者側視点からの行動分析.....	64
3.5 提案手法の検証.....	66
<b>第4章 ソーシャルメディアを利用した予測に関する研究.....</b>	<b>67</b>
4.1 日本映画の観客動員数予測分析.....	68
4.2 分析の流れとデータ抽出範囲.....	69
4.3 計測データの説明力分析.....	70
4.4 重回帰分析の結果.....	74
4.5 予測力改善の試み.....	77
4.6 実務への適用可能性.....	79
4.7 TTL 値の影響考察.....	83
4.8 本章のまとめ.....	87
<b>第5章 ソーシャルメディアを利用した計測・モデル化に関する研究.....</b>	<b>89</b>
5.1 日本における地上波デジタル放送への移行と家電エコポイント制度.....	89
5.2 分析の概要.....	90
5.3 モデルの構築.....	92
5.4 主成分分析とソーシャルメディア利用者の行動.....	94
5.5 ソーシャルメディア利用者の行動推移.....	101
5.6 考察.....	108
5.7 本章のまとめ.....	109
<b>第6章 まとめ.....</b>	<b>111</b>
6.1 研究成果のまとめ.....	111
6.2 今後の展望.....	113



謝辞 .....	115
参考文献 .....	117
関連業績リスト .....	137



# 第1章 序論

近年,日本におけるインターネット利用環境は,大きく変わってきている.総務省が2013年6月14日に公開した「平成24年通信利用動向調査」<sup>1</sup>によると,13歳~49歳までのインターネット利用率は9割を超えたが,60歳以上の世代においては,なお拡大傾向にある.また,利用情報通信機器もパソコンの利用率は下がり,スマートフォン,タブレット型端末が急速な伸びを示し,Facebook, Twitter, Lineなどのソーシャルメディアの利用動向も個人,企業共に利用者を伸ばしている.たとえば,Lineは2011年6月のサービス開始から,2013年11月25日で登録ユーザが3億人を超えるまでに急成長している<sup>2</sup>.元来ソーシャルメディア(Social Media)という用語は,2006年ごろから使われるようになったとされており<sup>3</sup>,Web 2.0の技術的,思想的背景に基づいたユーザ生成コンテンツ(UGC: User Generated Contents)の流通を媒介するものと定義づけることができるとされている[Kaplan 2010].

上述のとおり,ソーシャルメディアは,身近で手軽なコミュニケーション手段として広く社会に普及してきている.一方で,ソーシャルメディアの手軽さが引き起こす,投稿内容に対する影響の拡大も問題となってきた.たとえば,投稿者の投稿内容に対し,投稿者の予想とは異なる反応を引き起こすことがある.これらは,元の発言に対する多くの非難,誹謗中傷や,多くの投稿者が連携して大量の内容を投稿する現象としてあらわれる.前者は「炎上(flaming)」,後者は「祭り(binge)」と呼ばれ,これらは社会的現象に発展することもある.たとえば,2013年8月2日,日本テレビ系列の「金曜ロードSHOW!」でスタジオジブリの映画「天空の城ラピュタ」が放映され,その終盤で主人公が「バルス!」という滅びの呪文を唱えるのと同時に, Twitter上でも「バルス!」と書き込むことは“バルス祭り”と呼ばれ,投稿数で世界記録を樹立したと報道された<sup>4</sup>.

---

<sup>1</sup> [http://www.soumu.go.jp/johotsusintokei/statistics/data/130614\\_1.pdf](http://www.soumu.go.jp/johotsusintokei/statistics/data/130614_1.pdf) (2013/10/13)

<sup>2</sup> <http://linecorp.com/press/2013/1125636> (2013/12/4)

<sup>3</sup> <http://www.google.com/trends/explore#q=%22social+media%22> (2013/10/13)

<sup>4</sup> 「バルス祭り」でつながる日本人:

<http://www.nikkei.com/article/DGXZZO58092290T00C13A8000000/> (2013/8/5)

この様に年々拡大しているソーシャルメディアは、大きな社会的な影響力を持つようになっており、近年では政治的な影響を持つことも少なくない。例えば、国家や権力を有する機関などに対し、一般大衆が自らの意思や考えを現実化するために、ソーシャルメディアを情報共有手段として利用することもある[Shirky 2011]。Ghoniem は、その様な政治的活動を“**Revolution 2.0**”という概念として述べ、「ソーシャルメディアの活用によって行われる政治的活動は、特別なヒーローがいるわけではなく、みんながヒーローとして少しずつ協力しながら、最終的に世界最大の百科事典を作り上げる Wikipedia のようなもの」であり、デジタル時代の革命だとしている[Ghoniem 2011]。

研究領域においても、2006年ごろからソーシャルメディアという用語が使われ始めた後、その利用拡大が進むにつれて注目が集まり、心理学、社会科学、数理社会学、情報学、マーケティングや経営学などで研究対象として広がってきている。たとえば数理社会学では、ソーシャルメディアで起きる現象を分析の範囲と枠組みで捉えることを試みている。分析範囲として、マイクロレベル、メゾレベル、マクロレベルの3つを定義し、分析枠組みとして、行為 (action) のメカニズムに着眼したモデル、過程 (process) のメカニズムに着眼したモデル、構造 (structure) のメカニズムに着眼した3つのモデルを定義している [荒井 2004]。

本研究では、インターネットの基幹技術の1つである DNS のクエリデータにもとづく“異なり数”を用いて、“投稿者側”の視点からではなく“閲覧者側”の視点から、実社会で起きる現象に対するソーシャルメディア利用者の行動を分析する手法の提案を試みる。以降では、インターネットにおけるソーシャルメディア利用者のことを、単にソーシャルメディア利用者と称する。

従来の研究は、実社会で起きる現象に対するソーシャルメディア利用者の行動を、投稿内容を通して投稿者側の視点から分析するものが主流であった。しかし、投稿者側からの分析には、ステルスマーケティングなど作弄的な投稿者による投稿内容の影響を受けやすいという“作弄的な投稿影響問題”が存在する。たとえば、2012年1月には、日本の大手グルメサイトにおいて口コミを装った“やらせ評価”が行われ、大きな社会的事件に発展している[Phillip 2012]。

この様な問題に対応するために、ソーシャルメディア利用者の行動を閲覧者側の視点から分析する試みとして、調査票によるアンケート結果や検索エンジンの出力結果の分析が行われていた ([Quercia 2012], [Jung 2013], [Dupret 2006], [Baeza-Yates 2006])。し

かし、これらの手法には、ソーシャルメディア利用者が調査票の記入者を選択する時や調査票を記入する時、または検索エンジンを利用して検索結果を得る際に発生し得る各種バイアスの問題（具体的には、標本抽出時に母集団を代表しない特定の性質が紛れ込んでしまうサンプリングバイアスや、アンケート調査などで生じる個人や状況などに特有な反応が紛れ込んでしまう反応バイアス）の影響を受けてしまうという“利用者バイアス問題”が存在する。たとえば[Nielsen 2006b]では、標本数を増やすために報酬を払うことによる影響や、検索エンジンの結果が検索エンジン提供会社のアルゴリズムにより表示され、調査に有用ではない情報を多く含む影響を指摘している。また、ソーシャルメディア利用者に調査票を記入させるため、大規模な調査には不向きであるという“調査規模的問題”も存在する。[Nielsen 2013]では、ソーシャルメディア利用者のアクティビティに、極端に活発な期間とそうでない期間があり、分析結果に大きな違いが出る実験結果を示しており、データ収集時に規模をコントロールする必要性を示している。

本研究では、DNS クエリデータを用いることで、実社会で起きる現象に対するソーシャルメディア利用者の行動を分析する際に発生し得る“作為的投稿影響問題”や“利用者バイアス問題”，“調査規模的問題”に対応できる手法を提案し、その有効性の確認を試みる。

まとめると本研究の目的は、

- ① ステルスマーケティングなど作為的な投稿による影響を受けにくい、閲覧者側の視点からソーシャルメディア利用者の行動分析が可能（“作為的投稿影響問題”への対応）
- ② データ収集時のバイアスを受けにくく、大規模なソーシャルメディア利用者の行動分析が可能（“利用者バイアス問題”，“調査規模的問題”への対応）

である手法の提案を行い、その実証検証を行うことである。

以下本稿では、まず、第2章において本研究で対象とするソーシャルメディアとソーシャルメディア利用者の分類軸を明らかにし、その分類軸を用いてソーシャルメディア利用者の行動分析に関する先行研究を分類する。そして、先行研究の研究範囲と課題を明らかにし、本研究の位置づけと貢献について示す。続いて第3章で、“作為的投稿影響問題”，“利用者バイアス問題”，“調査規模的問題”に対応するためのソーシャルメディア利用者の行動分析手法を提案する。第4章、第5章では、第3章で述べた手法を踏まえ、その有効性の確認を試みる。

第4章では、提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、

提案手法の有効性を確認するために日本映画の観客動員数予測を試みる。続いて第 5 章では提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、提案手法が大規模で複雑な行動分析にも有効であることを確認し、その行動の時間的推移を明らかにする。具体的には、日本政府が 3 年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の 1 つであった“家電エコポイント制度の実施”という実社会における 2 つの大きな社会現象に対し、DNS クエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用の関連を計測することで、ソーシャルメディア利用者の行動の時間的推移の分析を試みる。

最後に第 6 章で、本稿における、ソーシャルメディア利用者の行動分析時に発生し得る“作為的投稿影響問題”や“利用者バイアス問題”，“調査規模的問題”に対応できる手法の提案とその実証検証結果を示し、提案手法の有効性についてまとめ、本研究の今後の展望について述べる。

## 第2章 ソーシャルメディア利用者の行動分析 に関する研究範囲と課題

本研究は、ソーシャルメディアにおける“投稿者側”の視点ではなく、“閲覧者側”の情報を用いて、実社会で起きる現象の分析を試みる。そこで以下では、ソーシャルメディア利用者の行動分析に関する先行研究の範囲を示し、先行研究の抱える課題を明らかにする。まず、本研究で対象とするソーシャルメディアの分類軸（2.1 節）と、ソーシャルメディア利用者の分類軸（2.2 節）を各々の分類軸に関連する先行研究を示しながら明らかにする。続いて、その両分類軸を用いて、ソーシャルメディア利用者の行動分析に関する先行研究を分類し、その研究範囲と抱える課題について明らかにする（2.3 節）。最後に、先行研究の研究範囲と課題を踏まえ、本研究の位置づけと貢献について示す（2.4 節）。

### 2.1. ソーシャルメディアの分類軸

ソーシャルメディアの分類は、主にメディア内容に対する内容を誰がコントロールするかの視点で行われている。本節では、ソーシャルメディアの分類軸と特徴について示した後、その分類軸に関する先行研究について述べる。

#### 2.1.1. 自己管理可能メディアと自己管理不可能メディア

2009年5月、「Multimedia 2.0: From paid media to earned media to owned media and back」<sup>5</sup>というタイトルのブログにおいて、マーケティング・チャネルのメディア分類として、Paid Media, Owned Media, Earned Media という概念が示された。

- Paid Media : テレビ, ラジオ, 雑誌, 新聞, リスティング広告など, 料金を支払って, 商品などを宣伝するために利用するメディアのこと

<sup>5</sup> [http://news.cnet.com/8301-13641\\_3-10237112-44.html](http://news.cnet.com/8301-13641_3-10237112-44.html) (2013/10/13)

Media type	Definition	Example	The role	Benefits	Challenges
Paid Media	Brand pays to leverage a channel	<ul style="list-style-type: none"> <li>● Display ads</li> <li>● Paid search</li> <li>● Sponsorships</li> </ul>	Shift from foundation to a catalyst that feeds owned and creates earned media	<ul style="list-style-type: none"> <li>● In demand</li> <li>● Immediacy</li> <li>● Scale</li> <li>● Control</li> </ul>	<ul style="list-style-type: none"> <li>● Clutter</li> <li>● Declining response rates</li> <li>● Poor credibility</li> </ul>
Owned Media	Channel a Brand controls	<ul style="list-style-type: none"> <li>● Web site</li> <li>● Mobile site</li> <li>● Blog</li> <li>● Twitter account</li> </ul>	Build for longer-term relationships with existing potential customers and earn media	<ul style="list-style-type: none"> <li>● Control</li> <li>● Cost efficiency</li> <li>● Longevity</li> <li>● Versatility</li> <li>● Niche audiences</li> </ul>	<ul style="list-style-type: none"> <li>● No guarantees</li> <li>● Company communication not trusted</li> <li>● Takes time to scale</li> </ul>
Earned Media	When customers become the channel	<ul style="list-style-type: none"> <li>● WOM</li> <li>● Buzz</li> <li>● “Viral”</li> </ul>	Listen and respond – earned media is often the result of well-executed and well-coordinated owned and paid media	<ul style="list-style-type: none"> <li>● Most credible</li> <li>● Key role in most sales</li> <li>● Transparent and lives on</li> </ul>	<ul style="list-style-type: none"> <li>● No control</li> <li>● Can be negative</li> <li>● Scale</li> <li>● Hard to measure</li> </ul>

図 2.1 [Corcoran 2009]におけるソーシャルメディア分類

- Owned Media : 企業の自社パンフレットや広報誌, 自社企業 Web サイトやブランド Web サイトなど, 企業自らが所有し消費者に向けて発信するメディアのこと
- Earned Media : 広告ではなく, 口コミで商品やサービスを紹介するブログや SNS など, 信頼や評判を獲得するメディアのこと

その後, 2009 年 12 月に Corcoran は図 2.1 に示すメディア分類の定義を示している [Corcoran 2009]. この中で Corcoran は Paid Media を, 企業などの情報発信者自らがメディア内容をコントロールできる“自己管理可能メディア”であり, メディアに対して投資を行うことで一時的ではあるが即時性とスケール性を持った効果が得られ, Owned Media や Earned Media に展開していくきっかけとなる特徴を持つメディアであるとしている. Owned Media は, Paid Media と同様に“自己管理可能メディア”であるが, メディアに対して投資を行い, 潜在顧客との長期の関係を構築していくメディアであり, 時間は必要だがブランド・ポータビリティの融通性は高いメディアであるとしている. また, Earned Media は, 一般ユーザーなどの第 3 者がメディア内容をコントロールする“自己管理不可能メディア”であり, 投資を必要としないが良い口コミ情報と悪い口コミ情報の



Media type	Definition	Examples
<b>Paid</b>	Your company pays for media space or for a third party to promote its products	Television commercials, magazine and newspaper ads, billboards, product placement, Web banners, search-engine marketing
<b>Owned</b>	Your company uses or create its own new channels to advertise	Catalogs, Web site, Facebook fan pages, e-mail and customer databases, company-owned retail stores
<b>Earned</b>	Consumers create media and/or share media your company created	Organic search placement, forwarding a popular commercial to friends, consumer ratings and reviews, rankings on community sites
<b>Sold</b>	Your company invites other marketers to place their content on its owned media	An e-commerce retailer selling ad space on its Web site, a consumer marketer creating an online community and selling ad space
<b>Hijacked</b>	Your company 's asset or campaign is taken hostage by those who oppose it	Consumers rallying opposition to a company on Facebook, consumers creating and distributing their own negative versions of ads

図 2.2 [Edelman 2010] におけるソーシャルメディア分類

どちらにも対応する必要があるため、Paid Media や Owned Media と戦略的連携が必要としている。

また Edelman は、図 2.2 に示すように、[Corcoran 2009]の分類に Hijacked と Sold の 2 つのメディア分類を追加している。Hijacked は、収益に対する大きな脅威となり、ネガティブな評価を軽減するために投資を伴うメディアのことであり、Sold は、他社に広告表示を販売することで新しい収益源となるメディアであるとしている[Edelman 2010]。つまり、Hijacked は Earned がネガティブに働く場合（良くない評判が立つなど）のメディアを示し、Sold は自社メディアが Paid Media 並の価値を持つことで、他社の広告機能を提供するメディアとなる場合（Paid となる場合）のことを示している。

本田は、図 2.3<sup>6</sup>に示す[Corcoran 2009]を拡張したフライシュマン・ヒラード(Fleishman Hillard)<sup>7</sup>が提案する PESO (Paid, Earned, Shared, Owned) という 4 つの分類を紹介している[本田 2012]。この中で本田は、[Corcoran 2009]の分類がメディアからの観点が

<sup>6</sup> The Digitization of Research And Measurement In Public Relations : <http://www.socialmediaexplorer.com/online-public-relations/the-digitization-of-research-and-measurement-in-public-relations/> (2013/10/13)

<sup>7</sup> 1946 年に創設され、本社を米国セントルイスに置き、2013 年 9 月時点で世界 28 カ国、80 拠点をもつ戦略コミュニケーション・コンサルティング会社 <http://fleishmanhillard.com/> (2013/9/30)

Media type	Definition	Example	The role	Benefits	Challenges
<b>Paid Media</b>	Brand pays to leverage a channel	<ul style="list-style-type: none"> <li>● Display ads</li> <li>● Paid search</li> <li>● Sponsorships</li> </ul>	Shift from foundation to a catalyst that feeds owned and creates earned media	<ul style="list-style-type: none"> <li>● In demand</li> <li>● Immediacy</li> <li>● Scale</li> <li>● Control</li> </ul>	<ul style="list-style-type: none"> <li>● Clutter</li> <li>● Declining response rates</li> <li>● Poor credibility</li> </ul>
<b>Owned Media</b>	Channel a Brand controls	<ul style="list-style-type: none"> <li>● Web site</li> <li>● Mobile site</li> <li>● Blog</li> <li>● Twitter account</li> <li>● Facebook page</li> <li>● YouTube video</li> </ul>	Build for longer-term relationships with existing potential customers and earn media	<ul style="list-style-type: none"> <li>● Control</li> <li>● Cost efficiency</li> <li>● Longevity</li> <li>● Versatility</li> <li>● Niche audiences</li> </ul>	<ul style="list-style-type: none"> <li>● No guarantees</li> <li>● Company communication not trusted</li> <li>● Takes time to scale</li> </ul>
<b>Earned Media</b>	Brand gains attention of editorial decision-maker	<ul style="list-style-type: none"> <li>● News &amp; feature coverage</li> <li>● Guest by liners</li> <li>● Op-eds/ editorials</li> </ul>	Often drives the online conversation, buzz	<ul style="list-style-type: none"> <li>● Most credible</li> <li>● Catalyzer</li> <li>● Sales/lead generator</li> <li>● Third party endorser</li> </ul>	<ul style="list-style-type: none"> <li>● Less control</li> <li>● Short lived</li> <li>● Limited Journalist bandwidth</li> </ul>
<b>Shared Media</b>	When customers become the channel	<ul style="list-style-type: none"> <li>● WOM</li> <li>● Buzz</li> <li>● Viral</li> <li>● RT, Like &amp; Share</li> </ul>	Shared media is the desired result of well-executed and compelling paid, earned and/or owned media initiatives	<ul style="list-style-type: none"> <li>● Most credible /trusted</li> <li>● Key role in sales/transactions</li> <li>● Transparent and lives on</li> </ul>	<ul style="list-style-type: none"> <li>● No control</li> <li>● Can be negative</li> <li>● Scale</li> <li>● Harder to measure</li> </ul>

図 2.3 [本田 2012] におけるソーシャルメディア分類

強いものに対して、PESO はコンテンツの観点が強いつい分類であると述べている。

**Paid** : 企業にとって購入可能なあらゆるコンテンツ (広告, タイアップ, スポンサーシップなど) のこと

**Owned** : 企業が自己管理できる主に Web 上のプロパティ (自社サイト, ブランドサイト, 公式ブログ, Facebook や Twitter などの公式アカウントなど) のこと

**Earned** : 企業にとって第三者であるマスコミやブロガーなどに影響を与えて生成してもらうコンテンツ (マスコミ記事, ブログ記事など) のこと

**Shared**: 企業が自己管理できず, 消費者が管理するソーシャル・ネットワーク (Facebook, Twitter, 口コミなど) のこと

Type	Definition	Offline Examples	Online Examples
<b>Paid</b>	Media activity related to a company or brand that is generated by the company or its agents	<ul style="list-style-type: none"> <li>● Traditional advertising (e.g., TV, radio, print, outdoor)</li> <li>● Sponsorships</li> <li>● Direct mail</li> </ul>	<ul style="list-style-type: none"> <li>● Display banner advertising</li> <li>● Search advertising (e.g., Google-Ad Words)</li> <li>● Social network advertising (e.g., Facebook ads)</li> <li>● Electronic direct mail (e.g., email advertisements)</li> </ul>
<b>Owned</b>	Media activity related to a company or brand that is generated by the company or its agents in channels it controls	<ul style="list-style-type: none"> <li>● Retail in-store visual merchandising or displays</li> <li>● Brochures</li> <li>● Company press releases</li> </ul>	<ul style="list-style-type: none"> <li>● Company brand website</li> <li>● Company brand blog</li> <li>● Company-owned pages/accounts in online social networks (e.g., Twitter account, Facebook brand page)</li> </ul>
<b>Earned</b>	Media activity related to a company or brand that is not directly generated by the company or its agents, but rather by other entities such as customers or journalists	<ul style="list-style-type: none"> <li>● Traditional publicity mentions in professional media outlets</li> <li>● Ratings and reviews in traditional media outlets (e.g., movie reviews)</li> <li>● Consumer-to-consumer word-of-mouth conversations about products, including advice and referrals</li> <li>● Consumers showing or demonstrating products to each other</li> </ul>	<ul style="list-style-type: none"> <li>● Traditional publicity mentions in digital media outlets (e.g., professional blogs)</li> <li>● Online word-of-mouth referrals (e.g., invitations to join a website)</li> <li>● Posts in online communities or social networks (e.g., status update, tweets)</li> <li>● Online ratings and reviews (e.g., Yelp.com for restaurants, Amazon.com for products)</li> </ul>

図 2.4 [Stephen 2012]におけるソーシャルメディア分類

Stephen は、図 2.4 に示すように、Corcoran の 3 つのメディア分類を更にオフライン・メディア（トラディショナル・メディア）か、オンライン・メディア（ソーシャル・メディア）かで分類している[Stephen 2012]。オフライン・メディアは、従来からあるインターネットなどのネットワークを介さないメディアであり、テレビ、雑誌、カタログなどがある。オンライン・メディアは、ネットワークを介するメディアのことであり、バナー広告、ブランドサイト、口コミサイトなどがある。

以上の先行研究は、ソーシャルメディアが自己管理可能か否かの視点で細かく分類可能であるとしている。図 2.5 に自己管理可能か否かの視点でまとめた分類を示す。図 2.5 において“○”が示されているメディア分類が、各先行研究で定義されているメディアである。各研究において各々の定義があるが、どの研究も Paid Media, Owned Media, Earned Media の 3 つは定義されている。また、自己管理可能メディアか否か、オンライン・メディアか従来からあるテレビやラジオなどの様なオフライン・メディアかの視点で、ソーシャルメディアの分類に関する先行研究を 5 つの研究領域にまとめたものが図 2.6 である。

	自己管理可能メディア			自己管理不可能メディア		
	Paid Media	Owned Media	Sold Media	Earned Media	Shared Media	Hijacked Media
[Corcoran 2009]	○	○		○		
[Stephen 2012]	○	○		○		
[本田 2012] (PESO)	○	○		○	○	
[Edelman 2010]	○	○	○	○		○

図 2.5 自己管理可能メディアと自己管理不可能メディアの分類

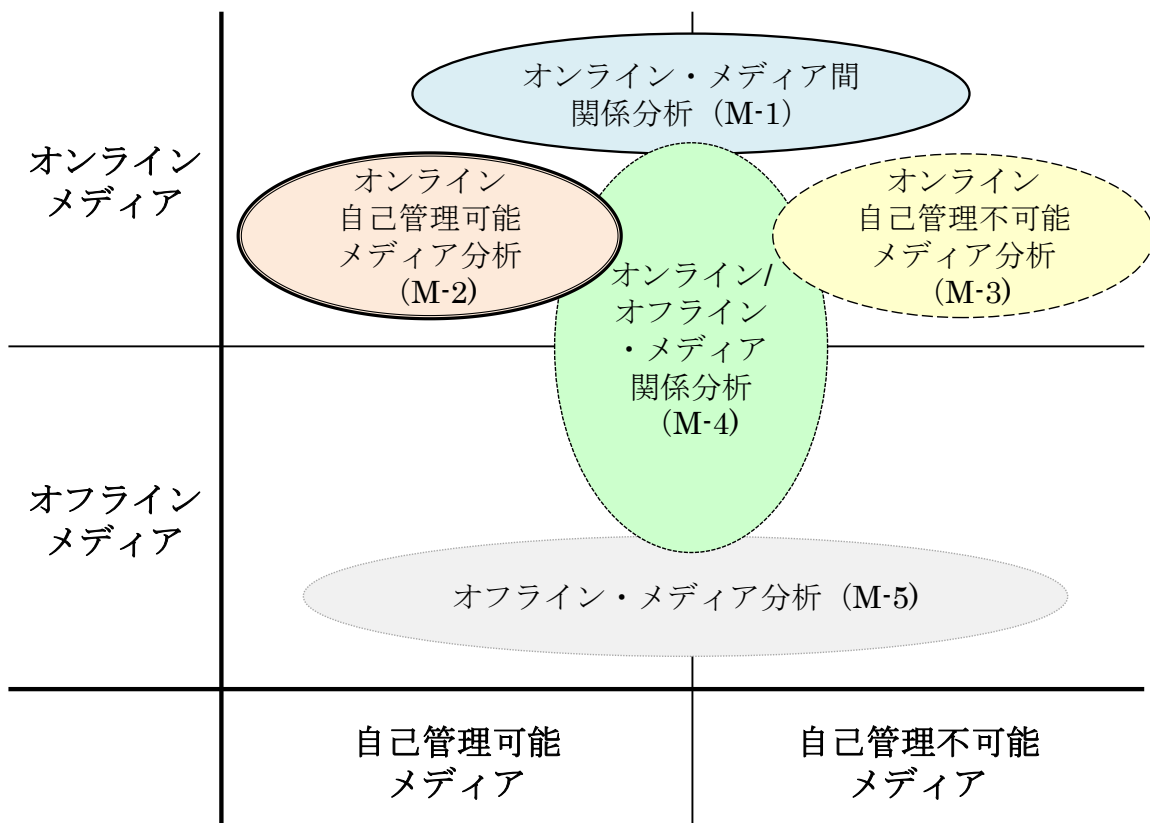


図 2.6 ソーシャルメディアの分類に関する研究領域

#### ■ オンライン・メディア間関係分析 (M-1)

本研究領域は、ソーシャルメディア間の関係に関する研究領域である。たとえば、メディア間の相乗効果に関する研究や企業におけるメディア間の戦略的利用方法に関する研究 ([Burcher 2012], [Brown 2012]) などがある。

#### ■ オンライン自己管理可能メディア分析 (M-2)

本研究領域は、自己管理可能なメディアである Paid Media, Owned Media に関する研究である。たとえば、バナー広告の効果分析や Web サイトのデザインが与える影響分析 ([Pergelova 2010], [Proulx 2012]) などがある。

#### ■ オンライン自己管理不可能メディア分析 (M-3)

本研究領域は、自己管理不可能なメディアである Earned Media などに関する研究である。たとえば、Twitter におけるつぶやき分析や Facebook における繋がり分析 [Hanna 2011] などがある。

#### ■ オンライン/オフライン・メディア関係分析 (M-4)

本研究領域は、トラディショナル・メディアとソーシャルメディアの連携に関する研究である。たとえば、トラディショナル・メディアであるテレビ広告とウェブ広告の効果比較研究や新聞広告とウェブ広告の認知度比較 ([Gallagher 2001], [Rodgers 2005]) などがある。

#### ■ オフライン・メディア分析 (M-5)

本研究領域は、トラディショナル・メディアに関する研究である。メディアと人との関係に関する研究は古くから行われており、たとえば Scott と Starch は心理学的側面から分析している ([Scott 1910], [Starch 1923])。Scott は記憶、感情と情動、本能、暗示、思考・感情・意思の要素から構成される広告効果に対する心理的過程を説明しており、Starch は消費者的要因、媒体的要因、表現的要因などが広告への関心や商品への関心、記憶、購買行為などに与える影響を分析している。

この領域では、広告媒体の違いによる広告効果についての研究が盛んである ([Krugman 1965], [Krugman 1971], [Jacoby 1990])。[Krugman 1965]では、新聞広告とテレビ広

告の広告効果比較を行っている。テレビ広告は、新聞広告の印刷媒体に比べ受動的にメッセージを受け取るため、テレビ広告の視聴で直接消費者が行動に変化をもたらすことはないが、反復的に視聴することで徐々に認知を変化させると述べている。また[Krugman 1971]では、脳波の変化に着目し、雑誌広告とテレビ広告の広告効果の比較を行っている。Jacobson は、雑誌広告とテレビ広告の比較分析において、広告メッセージの理解度について着目し分析を行っている[Jacobson 1990]。結果は、雑誌広告に接した場合の方がテレビ広告に接した場合よりも広告メッセージを正確に理解した人が多かったという結果を示した。

本研究はソーシャルメディアが研究対象なので、「オフライン・メディア分析 (M-5)」は対象外とする。すなわち、「オンライン・メディア間関係分析 (M-1)」、「オンライン自己管理可能メディア分析 (M-2)」、「オンライン自己管理不可能メディア分析 (M-3)」、「オンライン/オフライン・メディア関係分析 (M-4)」における先行研究について述べる。

### 2.1.2. ソーシャルメディアの分類軸に関する先行研究

年々、実社会においてオンラインのソーシャルメディア利用者が増加を続けていることを踏まえ、ソーシャルメディア利用者に対する広告効果や影響に関する研究が数多く行われている。

#### (1) オンライン・メディア間関係分析 (M-1)

多くの研究者が、オンライン・メディアをメディアからの視点、コンテンツからの視点、メディア提供主からの視点などで分類 ([Corcoran 2009], [Stephen 2012], [本田 2012], [Edelman 2010]) し、主なテーマとして、それぞれのオンライン・メディア間における戦略的な連携方法、マーケティング手法などについて述べている ([Burcher 2012], [Brown 2012], [Middleton 2012])。たとえば Middleton は、ソーシャルメディアへの参加者をアクティブ化するためのモデル (Understand Participant, Integrate Channels, Create For Action, Optimize Participation, Connect & Amplify) を提案している。

#### (2) オンライン自己管理可能メディア分析 (M-2)

自己管理可能なメディアである Paid Media や Owned Media, Sold Media に関連して Pergelova は、インターネットが全ての広告効率を上昇させる手助けになるか否かを調べている。具体的には、スペインの自動車産業のサンプルデータを利用してノンパラメトリ

ックな方法で分析した結果、インターネットにおける広告が効率レベルを改善したことを示している[Pergelova 2010].

また Proulx は、トラディショナルな Paid Media であるテレビに対して、ソーシャルメディアとしての Paid Media の1つとして”Social TV”について述べている[Proulx 2012]. ソーシャルメディアは新しく強力な「バック・チャンネル」として捉えることが可能であり、その中でも Social TV は新たなマーケティングの機会を与えると述べている.

### (3) オンライン自己管理不可能メディア分析 (M-3)

オンライン自己管理不可能なメディアについて Brown は、インフルエンサーではなく顧客に注目し、彼らがどのようにブランド志向者からブランド購入者になっていくのかについて分析している. 顧客をブランド購入者に変えるための4つの戦略立案ポイントである“4Ms (Make, Manage, Monitor, Measure)”を定義し、顧客をインフルエンサーに変えるためのリバースエンジニアリング手法について提案している[Brown 2013].

また Hanna は、ソーシャルメディアをマーケティングに活用するための体系整理を行っている. 企業のマーケティング・コミュニケーション戦略における戦略的なソーシャルメディア統合に関連し、複数のソーシャルメディア (Facebook, Twitter, Digg, SCVNGR, YouTube など) を連携させてマーケティングプロモーションを行う重要性を説いている[Hanna 2011].

### (4) オンライン/オフライン・メディア関係分析 (M-4)

1990年代に入りインターネットの浸透が進んだことで、オンライン・メディアと従来からのオフライン・メディア間の関係を対象とした研究が行われ始めている ([Sundar 1998], [Gallagher 2001], [Rodgers 2005], [Dijkstra 2005]). Sundar は、Web と新聞を比較分析対象とし、広告と記事に関する記憶 (認知率と想起率) の比較を行っている. 比較分析の結果、広告においては Web 広告の方が新聞広告よりも認知率、想起率共に低く、記事においては差がないことを示している[Sundar 1998].

Gallagher は、Web とパンフレットを比較分析対象とし、広告と記事に関するブランドの想起、認知、ブランドに対する態度、広告に対する態度についての評価を行っている. 比較分析評価の結果は、広告に対する評価については、Web とパンフレットに明確な差を示さなかったが、記事に対する評価は、パンフレットに対する評価の方が Web に対する評

価よりも良い結果を示したと述べている[Gallagher 2001].

Rodgers は、広告のデザインに着目し、その影響について Web 広告と新聞広告の比較分析を行っている。デザインにおいて自然なものと不自然なものを準備し、それらを見た場合の広告への態度と購買意欲について計測している。結果は、一方の広告媒体が不自然なデザインで、もう一方の広告媒体が自然なデザインであった場合、自然なデザインの広告媒体の方が購買意欲は高くなり、両媒体が自然なデザインであった場合、広告への態度も購買意欲のどちらにおいても明確な差は示さなかったと述べている[Rodgers 2005].

また Dijkstra は、インターネット・コマーシャル、テレビコマーシャル、印刷広告を比較分析対象とし、認知的な反応や情緒的な反応、意欲的な反応について測定している。結果、認知的な反応について、インターネット・コマーシャルだけを見せた場合は、他のメディアを見せた場合よりも低い反応を示したが、テレビコマーシャルだけを見せた場合は、他のメディアを見せた場合よりも高い反応を示したと述べている。しかし、情緒的な反応と意欲的な反応については、明らかな差異はなかったとしている[Dijkstra 2005].

2006 年ごろからソーシャルメディアが出現した後、オンライン・メディアとしてのソーシャルメディアとトラディショナル・メディア間の関係を対象とした、広告効果、認知度評価、消費者への影響評価などの研究が行われている。たとえば広告効果に関する研究には、[Naik 2009], [Chowdhury 2007]などがある。Naik は、オンラインとオフラインの新しい階層的広告モデルを提案している。このモデルは、メディア内の相乗効果とメディア間の相乗効果の相互作用を可能にするとしている。モデルの実証検証として、自動車ブランドのオフライン広告（テレビ、印刷、ラジオ）とオンライン広告（バナー広告、検索広告）の相乗効果を計測している。

メディアの認知度評価に関する研究には、[Dijkstra 2005], [Voorveld 2011], [Wang 2006]などがある。たとえば Dijkstra は、テレビコマーシャル、印刷広告、インターネット広告を用いて認知的反応を計測している[Dijkstra 2005]. 実験の結果、認知的反応のうち広告再生について、複数メディアの提示グループと単一メディアの提示グループの間で差がみられたとしている。インターネット広告のみを提示した場合は複数メディアを提示した場合よりも低い再生であったが、テレビコマーシャルのみを提示した場合は複数メディアを提示した場合よりも高い再生であったと述べている。

また Voorveld は、219 人に対し、4 つのメディア（テレビコマーシャルーWeb サイト、Web サイトーテレビコマーシャル、Web サイトーWeb サイト、テレビコマーシャルーテ



レビコマーシャル)を視聴した場合の認識過程に関する分析(どのような精神的過程を経て、キャンペーンに寄与したかの分析)を行っている。認識過程は、①フォワードエンコーディング(最初の広告が次の広告に対し、どれだけ関心を引くか)、②イメージ移動(最初の広告のイメージを、次の広告再生まで持っていたか)、③複数の情報ソースを認識できるか、の3段階を設定している。フォワードエンコーディングと複数の情報ソースの認識については効果が確認できたが、イメージ移動については効果が確認できなかったと述べている。

メディアの消費者への影響評価に関する研究には、[Voorveld 2012]、[Danaher 2011]、[chang 2004]、[Stammerjohan 2005]などがある。たとえば Voorveld は、複数のメディア間におけるキャンペーンへの消費者反応を明らかにし、メディア連鎖の役割を調査している。2つの異なる製品カテゴリーを用いて、テレビコマーシャルと Web サイト(“テレビコマーシャル-Web サイト” vs “Web サイト-テレビコマーシャル”)の組合せ効果を測定している。結果は、2つの異なる製品に関するテレビコマーシャル-Web サイトの効果がある間は、“Web サイト-テレビコマーシャル”は高い影響を与えたと述べている。

また Danaher は、消費者とビジネスマンについてソーシャルメディアとトラディショナル・メディアに対する反応比較を行っている。電子メールと SMS を利用する「技術に精通した」若者は、テレビ、ラジオ、新聞などのトラディショナル・メディアに対しても情報の信頼性を示すが、ビジネスマンは、電子メールによるマーケティング活動をより受け入れていると述べている。

### 2.1.3. 本節のまとめ

本節では、ソーシャルメディアの内容を誰がコントロールするのかの分類軸で、一般的な広告サイトやスポンサーサイトなど情報発信者自らがメディア内容をコントロールする“自己管理可能メディア”と、SNS や掲示板など一般ユーザーなどの第3者がメディア内容をコントロールする“自己管理不可能メディア”の大きく2つに分類できることを示した。2.3 節では、この分類軸を用いてソーシャルメディア利用者の行動分析に関わる先行研究を整理する。

ソーシャルメディアの分類軸に関する先行研究の特徴としては、広告効果や認知度評価などソーシャルメディア利用者への影響分析が数多く行われている。特に、マーケティングに関わる広告系の研究が盛んに行われている研究領域であるということから、本研究でも第4章においてソーシャルメディアである広告サイトを取り上げて実証検証を行った。

## 2.2. ソーシャルメディア利用者の分類軸

ソーシャルメディア利用者の分類は、主にメディアに対する利用者の関わり度合いの視点で行われている。本節では、前節のソーシャルメディアの分類軸を踏まえ、その利用者の分類軸と特徴について示した後、その分類軸に関する先行研究について述べる。

### 2.2.1. 投稿者と閲覧者

マーケティングへの活用が注目されているソーシャルメディアであるが、その利用者の特徴の1つに「参加の不均衡性 (Participation inequality)」がある。Nielsen は、インターネットコミュニティには必ず「参加の不均衡」があると述べている[Nielsen 2006a]。

Nielsen は、図 2.7 に示すとおり、ソーシャルメディア利用者の特徴を 3 つに分類している。それぞれの利用者を“Lurkers”, “Intermittent Contributors”, “Heavy Contributors” とし、ソーシャルメディアに対する利用割合が順に 90%, 9%, 1%になる特徴を「90-9-1 ルール」と呼んでいる[Nielsen 2006a]。本ルールでは、ニュースグループなどの利用者参加型メディアの利用割合は、閲覧者“Lurkers”が 90%に対し、投稿者“Intermittent Contributors”が 10%, その“Intermittent Contributors”の更に 10% (全体の 1%) が非常に活発な投稿者“Heavy Contributors”であるとしている。

ブログメディアの場合には、この比率がさらに低下して、“Lurkers”：“Intermittent Contributors”：“Heavy Contributors” = 95：5：0.1 となり、Wikipedia では 99.8：0.2：0.003 になることを示している。同様に Arthur は、YouTube ではコンテンツの投稿者は全体の 0.5%程度であり、Wikipedia では全体の 70%程度の記事投稿は、1.8%の投稿者によって行われていると述べている[Arthur 2006]。

「参加の不均衡性 (Participation inequality)」についての研究は他にもある。Horowitz は、ソーシャルメディア利用者の分類を“Creators”, “Synthesizes”, “Consumers” とし、その利用割合が順に 1%, 9%, 90%であるとしている[Horowitz 2006]。また Speath は、オープンソース開発コミュニティ利用者の分類を“Developer”, “Contributor”, “Lurker” とし、3つのオープンソース開発コミュニティへの活動 (Reputation, Control over technology, Learning Opportunities) に対する関わり方を明らかにしている[Speath 2008]。

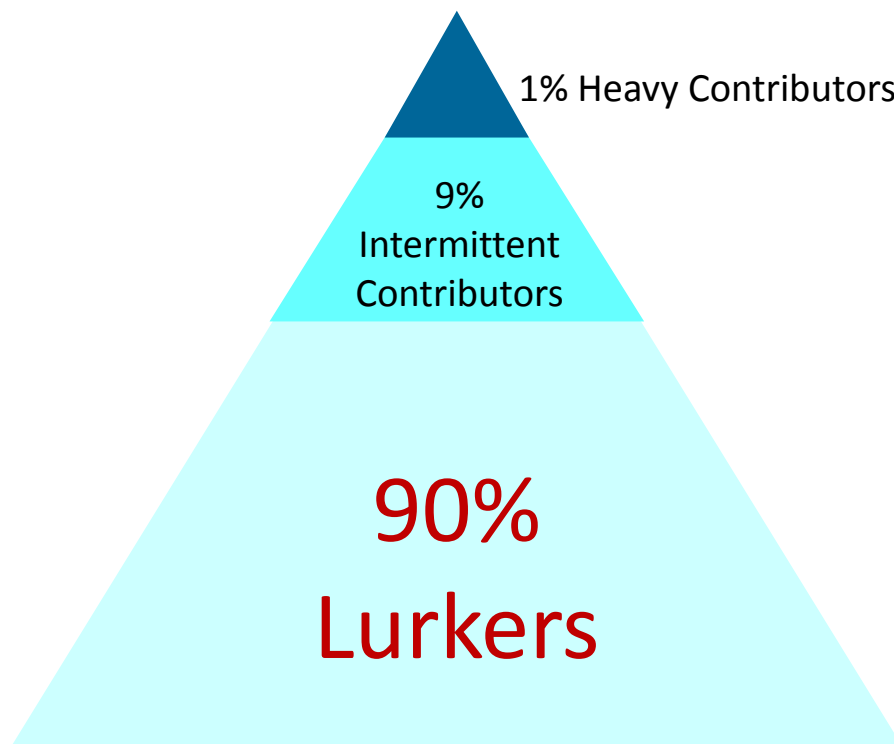


図 2.7 [Nielsen 2006a] におけるソーシャルメディア利用者分類

Stewart は、企業内でのクラウドソーシング<sup>8</sup>利用者を“Super Contributor”，“Contributor”，“Outlier”の3つに分類し、その利用割合を順に33%，66%，1%としている[Stewart 2010]。また Brandtzaeg は、ノルウェーの4つの主要な SNS 利用者 5,233 人からの調査データを分析し、5つの異なったタイプ (Sporadics, Lurkers, Socialisers, Debaters, Actives) に分類している[Brandtzaeg 2011]。

以上をまとめるとソーシャルメディア利用者の分類は、投稿などの貢献を行った者“投稿者”と、投稿などの貢献を行わず閲覧だけを行った者“閲覧者”の大きく2つに分類でき、先行研究におけるソーシャルメディア利用者の分類の定義を整理すると図 2.8 のとおりとなる。ソーシャルメディア利用者の分類は各々あるが、どの先行研究も“投稿者”は複数の定義に分かれるが、“閲覧者”は1つの定義になっている。また、ソーシャルメディア利用者は、投稿者よりも閲覧者の方が圧倒的に多いことが多くの研究で示されている。

<sup>8</sup> ここでのクラウドソーシングは、企業内で進められる共同プロジェクトなどで利用されるツールのことを示している。

[Nielsen 2006a]	[Horowitz 2006]	[Speath 2008]	[Stewart 2010]	[Brandtzaeg 2011]
Heavy Contributors	Creators	Developer	Super Contributor	Actives
Intermittent Contributors	Synthesizes	Contributor	Contributor	Debaters
<b>Lurkers</b>	<b>Consumers</b>	<b>Lurker</b>	<b>Outlier</b>	Socialisers
				Sporadics
				<b>Lurkers</b>

投稿者

閲覧者

図 2.8 ソーシャルメディア利用者の分類

### 2.2.2. ソーシャルメディア利用者の分類軸に関する先行研究

ソーシャルメディア利用者の分類軸に着目した研究は数多くある。それらの研究は、前項で述べた利用者の分類における「投稿者視点での分析 (C-1)」、 「投稿者・閲覧者視点での分析 (C-2)」、 「閲覧者視点での分析 (C-3)」の3つに整理できる。

- 投稿者視点での分析 (C-1)
- 投稿者・閲覧者視点での分析 (C-2)
- 閲覧者視点での分析 (C-3)

「投稿者視点での分析 (C-1)」は、投稿者側からの視点で、ソーシャルメディア利用者を分析している研究領域である。たとえば、ソーシャルメディアに対する投稿内容から影響を分析する研究、投稿者のソーシャルメディア利用理由を分析する研究などがある。「投稿者・閲覧者視点での分析 (C-2)」は、投稿者、閲覧者の両者の視点でソーシャルメディ

ア利用者を分析する研究領域である。たとえば、投稿者と閲覧者の利用比率を分析する研究、投稿者と閲覧者のソーシャルメディア利用理由を分析する研究などがある。「閲覧者視点での分析 (C-3)」は、主に閲覧者側からの視点でソーシャルメディア利用者を分析している研究領域である。たとえば、ソーシャルメディアに対して閲覧者が投稿しない理由調査や、閲覧者の行動分類を試みる研究などがある。

以上を踏まえ、「投稿者視点での分析 (C-1)」、「投稿者・閲覧者視点での分析 (C-2)」、「閲覧者視点での分析 (C-3)」における先行研究について述べる。

### (1) 投稿者視点での分析 (C-1)

投稿内容を用いて投稿者側からソーシャルメディア利用者の分析を行った研究には、[Rafaeli 2004], [Brandtzaeg 2008], [Cobb 2010], [Parvanta 2013], [Bernstein 2013], [Lakhani 2003], [Haklay 2008], [Neis 2012], [Richter 2013], [Wise 2013]などがある。

[Rafaeli 2004], [Brandtzaeg 2008], [Cobb 2010], [Parvanta 2013], [Bernstein 2013]は、オンラインコミュニティを分析対象として、投稿内容を用いてソーシャルメディア利用者の分析を行っている。[Rafaeli 2004]は、初めて投稿した者を投稿者 (De-Luker) として着目し、閲覧者が投稿者 (De-Luker) になるには **Social Capital** が関わっていると述べている。[Brandtzaeg 2008]は、投稿者を対象にして、なぜコミュニティ・ロイヤリティは下がるのかについて、インターネット一般利用者とコミュニティ利用者にアンケートを行い 200 の回答をもとに分析している。コミュニティ・ロイヤリティが下がる理由を、①出席者への不満 (友人の不足など)、②低品質なコンテンツ、③低品質なユーザービリティ、④嫌がらせといじめ、⑤孤立、⑥低い信用性、⑦過度な商業化、⑧議長への不満、⑨不明確な進行、の 9 つに分類している。また[Bernstein 2013]は、Facebook の 1 か月間 222,000 投稿ログを調べた結果、コンテンツの閲覧範囲は実際の 27%程度の大きさであると思われており、過小評価されていると述べている。実際には Facebook 利用者の各投稿で 35%、1 か月以上の期間で計測すると 61%の友人にリーチできる範囲であるとしている。

[Cobb 2010]は、QuitNet<sup>9</sup> (オンラインコミュニティの 1 つ) のネットワーク構造を分析して、その特徴を他の SNS と比較している。データは、7,569 人の利用者間の 103,592 個のリンクを対象とし、SNS の構造は、“女性”、“初期からの利用者”、“長い利用者”、“たばこを吸わない人”の増加と関係があり、メールの送受信は「90-9-1 ルール」の傾向があ

<sup>9</sup> <http://www.quitnet.com/> (2013/10/13)

ったと述べている。[Parvanta 2013]は、ヘルス・コミュニケーション・コミュニティを成功させるためのクラウドソーシング利用条件の分析を行っている。コミュニティが成功するためには、4つのF（Fun, Feeling good (fulfillment), Fame, Fortune）が重要であり、それらを満たすことで閲覧者が大部分を占める「90-9-1ルール」のコミュニティであっても成功すると述べている。

[Lakhani 2003], [Haklay 2008], [Neis 2012]は、オープンソース開発コミュニティを対象に分析を行っている。[Lakhani 2003]は、オープンソース・ソフトウェアの1つであるApacheの開発サイト<sup>10</sup>で、開発者がサイト利用者からの質問に対してオンラインヘルプで回答する行為について、どの程度時間を割いてボランティアで回答をするのか、について分析している。オープンソース開発者の開発時間確保に関する調査で、企業に雇用されている開発者は、「1週間あたり2営業日以上を確保」しており、またボランティア貢献者は「1週間あたり1日以上を確保」していると述べている。

また[Haklay 2008], [Neis 2012]は、どちらもオープンソース開発コミュニティの1つであるOpenStreetMapプロジェクト<sup>11</sup>での分析を行っている。[Haklay 2008]は、投稿せず地図の改善に少しの貢献もしない利用者が99.8パーセント存在していたことを示している。また[Neis 2012]は、登録されたメンバーの38%（192,000人）が少なくとも1つ以上の編集をデータベースで行い、全メンバーの5%（24,000人）がよりプロジェクトへ活発に関与したことを示している。

[Richter 2013]は、企業内コミュニティを分析対象とし、ビジネスツールであるESN(Enterprise Social Networks)の導入を成功させる要因と導入実績の関係を分析している。成功には7つの要因（Search, Edit, Rate, Label, Clarify, Notify, Share）があり、コンテンツを投稿している利用者は、利用開始後1か月で74%程度、半年後で46%程度であったとしている。

[Wise 2013]は、学生向け電子フォーラムにおける利用者を分析している。3週間で96人のオンライン議論参加者が、10から13のグループで課題を解く過程のデータを利用し、投稿者は議論に参加可能な全学習者の約4分の1を占めたとしている。また利用者特性は、①表層的なリスナー/断続的な話し手、②中心となるリスナー/中心となる話し手、③寛容なリスナー/熟考する話し手の3つに分類できるとしている。

---

<sup>10</sup> <http://www.apache.org/> (2013/10/13)

<sup>11</sup> <http://www.openstreetmap.org/> (2013/10/13)

## (2) 投稿者・閲覧者視点での分析 (C-2)

投稿者、閲覧者の両者の視点でソーシャルメディア利用者を分析する研究には、[Awan 2007], [Neylon 2009], [Lampe 2010], [Muller 2010], [Muller 2012]などがある。

[Awan 2007]は、ネットフォーラムを分析し 87%が投稿せず 13%が少なくとも 1 度の投稿、5%が 500 回以上投稿したことを示している。[Neylon 2009]は、論文評価コミュニティにおける投稿者・閲覧者の投稿規模についての研究である。論文に対しコメントを行い、改善を行うことは論文評価の過程として重要であるが、実際には閲覧のみが多く、コメント数は少ないことを示している（例：生物学の論文では、20,000 以上の閲覧があつたにも関わらずコメント数が 17 であつた）。これは、論文評価コミュニティにおける投稿者規模問題でもあると述べている。コミュニティには「90-9-1 ルール」があるとする、100 人に読まれる論文評価コミュニティの規模では、実際にコメントできる人は 1 人程度になってしまうという問題を抱えてしまい、論文評価に影響を与えることになるとしている。

[Lampe 2010]は、オンラインコミュニティの 1 つである Everything2.com<sup>12</sup>の閲覧者である匿名利用者 295 人と、投稿者である登録利用者 304 人をサーベイ対象とし、ソーシャルメディア利用理由の調査をしている。オンラインコミュニティへの参加を継続する理由は、最初の動機とは別に存在し、帰属意識が重要であると述べている。

[Muller 2010], [Muller 2012]は、企業内ツールとしてのコミュニティを分析対象として、投稿内容を用いて投稿者と閲覧者の比較分析を行っている。[Muller 2010]は、企業ファイル共有サービスで、ファイルをアップロードした利用者 (Uploders) とファイルを作成した利用者 (Contributors) の行動を比較している。Uploders は、ファイルをグループ単位でダウンロードし、あまりファイル検索を行わずファイル共有アプリケーションを直接操作する傾向であつたが、Contributors は、その逆の傾向であつたと述べている（検索を多く使い、リモートからのアプリケーションを使用）。また[Muller 2012]は、投稿者 (Contributor) と閲覧者 (Lurker) の行動を比較分析している。IBM の企業内コミュニティ：8,711 個、利用者：224,232 人、投稿者：22,949 人の投稿内容を対象とし、あるコミュニティで閲覧者だった人は、他のコミュニティで投稿者でもあることを示している。この現象は一般的な Lurker の理論とは異なっているが、個別のトピックやタスク、社会的グループに対する個人の性質が、人間関係に一部依存しているためであると理由づけしている。

---

<sup>12</sup> <http://everything2.com/> (2013/10/13)

### (3) 閲覧者視点での分析 (C-3)

閲覧者側からの視点でソーシャルメディア利用者を分析している研究には、[Nagel 2009], [Nonnecke 2000], [Nonnecke 2004], [Nonnecke 2006]などがある。

[Nagel 2009]は、学習コミュニティへの登録参加している学生のうち、学習の観点から特に情報を読むことしかしない閲覧者の行動分析を行っている。その特徴は、①他人の議論を読むだけで慎重に結果を得る、②価値を提供しない、③単位時間に効果的に議論ログを管理し最大の利益を得るが、低いコネクティビティや高いコストは支払わない、④理由なくファシリテーターと交流を続ける、⑤読まず利用もしないことに明確な理由はなく、学習コミュニティの利益を共有しない、という 5 つに分類できると述べている。また [Nonnecke 2000]では、閲覧者が投稿するか否かは、個性、ニーズ、満足感、トピックなどが関係していると述べている。同様に[Nonnecke 2004]では、Microsoft Network (MSN) オンラインコミュニティの 375 人を対象に、ソーシャルメディアの利用目的について分析を行っている。閲覧者の 13.2%は最初から閲覧だけが目的であり、29.7%はまだコミュニティについて学習中であるとし、半数以上 (53.9%) が閲覧するだけで十分であることを示している。[Nonnecke 2006]でも同様の結果を示している。

#### 2.2.3. 本節のまとめ

本節ではソーシャルメディア利用者が、そのメディアに対してどの様に関わるのかの分類軸で、投稿などの貢献を行った“投稿者”と、投稿などの貢献を行わず閲覧だけを行った“閲覧者”の大きく 2 つに分類できることを示した。2.3 節では、2.1 節の軸に加え、本節の分類軸も用いてソーシャルメディア利用者の行動分析に関わる先行研究を整理する。

図 2.9 にソーシャルメディア利用者の分類軸に関する先行研究の特徴を示す。多くの先行研究で「参加者の不均一性」が確認されており、その研究は「投稿者視点での分析 (C-1)」、 「投稿者・閲覧者視点での分析 (C-2)」、 「閲覧者視点での分析 (C-3)」の 3 つに整理できる。特に、ソーシャルメディア利用者数が少ないとされる投稿者 (Heavy Contributors, Intermittent Contributors) の視点の研究である「投稿者視点での分析 (C-1)」、 「投稿者・閲覧者視点での分析 (C-2)」において数多くの事例が存在しており、ソーシャルメディア利用者の大部分を占めるとされる閲覧者 (Lurkers) の視点の研究である「閲覧者視点での分析 (C-3)」においての事例は多くはない。本研究では、先行研究であまり多く行われていない閲覧者の視点で分析を行う。



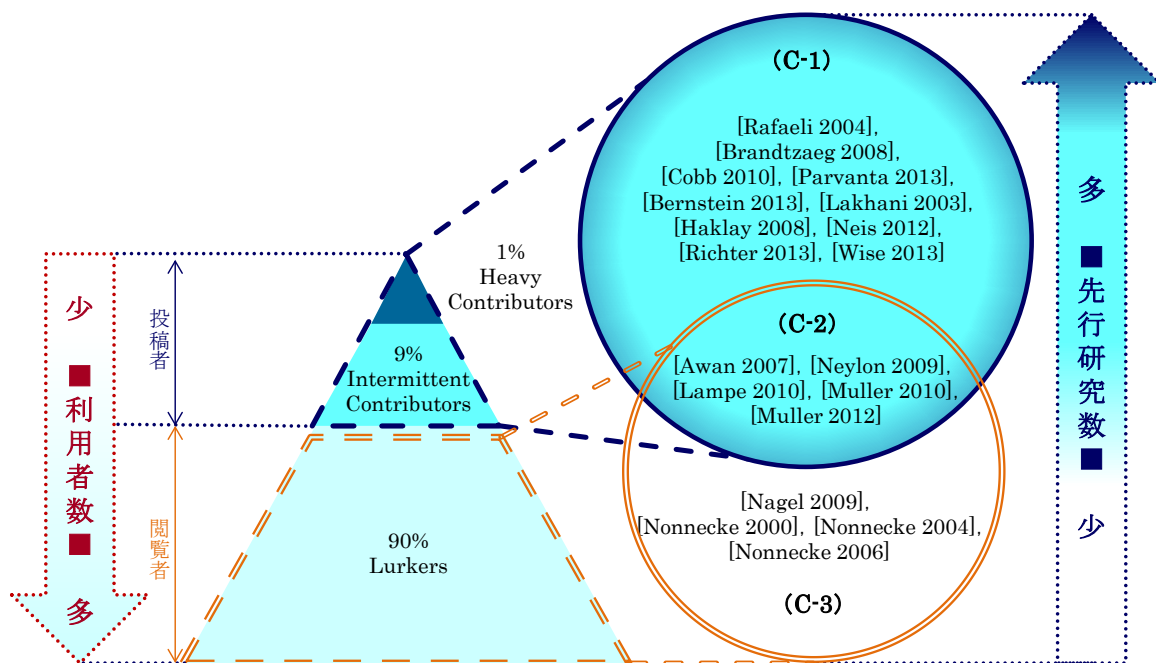


図 2.9 ソーシャルメディア利用者の分類に関する先行研究の特徴

### 2.3. ソーシャルメディア利用者の行動分析に関する先行研究

前節で示したとおり、ソーシャルメディア利用者の分類において大部分を占める閲覧者の行動把握は、年々増加しているソーシャルメディア利用者についての行動分析において重要な要素である。閲覧者の行動把握は、2.2 節で示した先行研究「投稿者・閲覧者視点での分析 (C-2)」, 「閲覧者視点での分析 (C-3)」の研究領域において、閲覧者側の視点からソーシャルメディア利用者の分類と特徴の把握が試みられている。

本節では、2.1 節、2.2 節で示した分類軸を用いて、ソーシャルメディア利用者の行動分析に関する先行研究の分類を行い、その研究範囲と先行研究の抱える研究課題について明らかにする。2.1 節で示したソーシャルメディアの分類軸（オンライン・メディアとオフライン・メディア、自己管理可能メディアと自己管理不可能メディア）と、2.2 節で述べたソーシャルメディア利用者の分類軸（投稿者と閲覧者）で研究領域を整理したものが図 2.10 である。

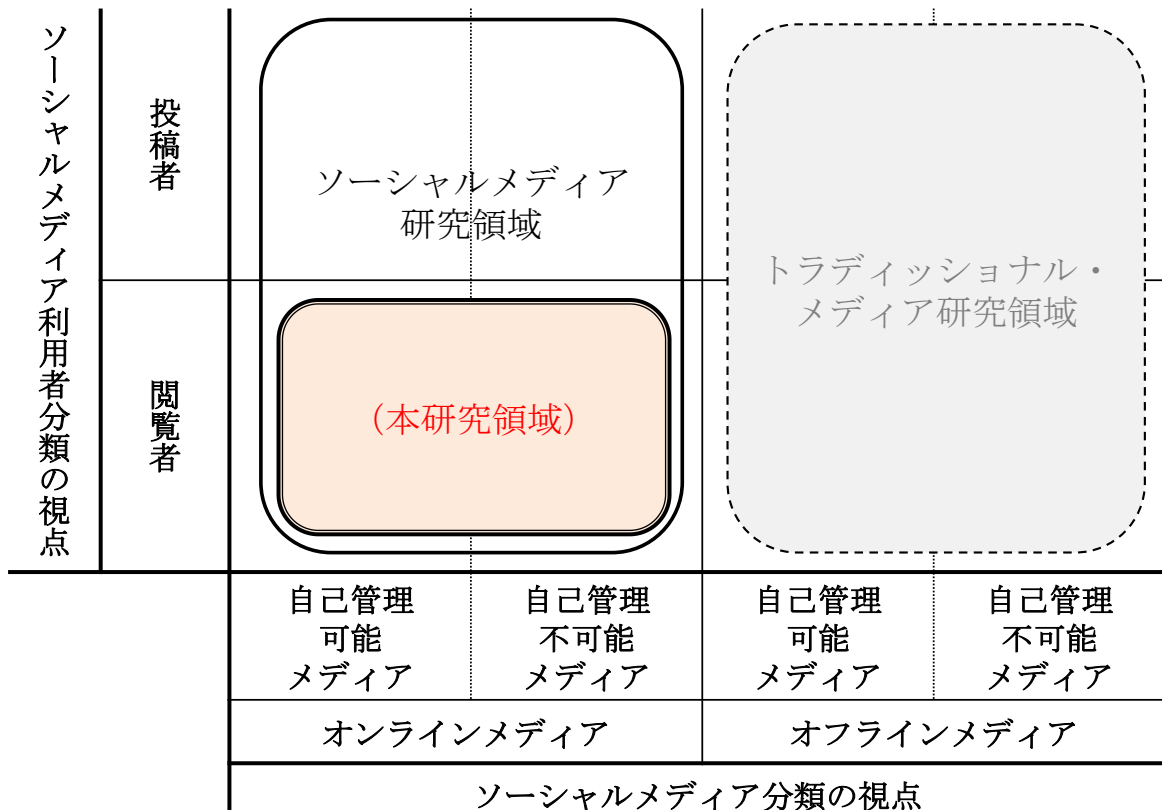


図 2.10 ソーシャルメディアとトラディショナル・メディアの研究領域

本研究においてソーシャルメディアとは、[Stephen 2012]の定義を踏まえ、オンライン・メディア＝ソーシャルメディア、オフライン・メディア＝トラディショナル・メディアであると定義する。また本研究の対象は、図 2.10 に示すソーシャルメディア研究領域であり、特にソーシャルメディア利用者分類における閲覧者視点の研究領域である。

以降、2.3.1 でソーシャルメディア利用者の行動分析における先行研究の分類を示し、2.3.2～2.3.5 で、その分類に応じた先行研究の特徴を述べ、2.3.6 で先行研究の抱える課題を明らかにする。

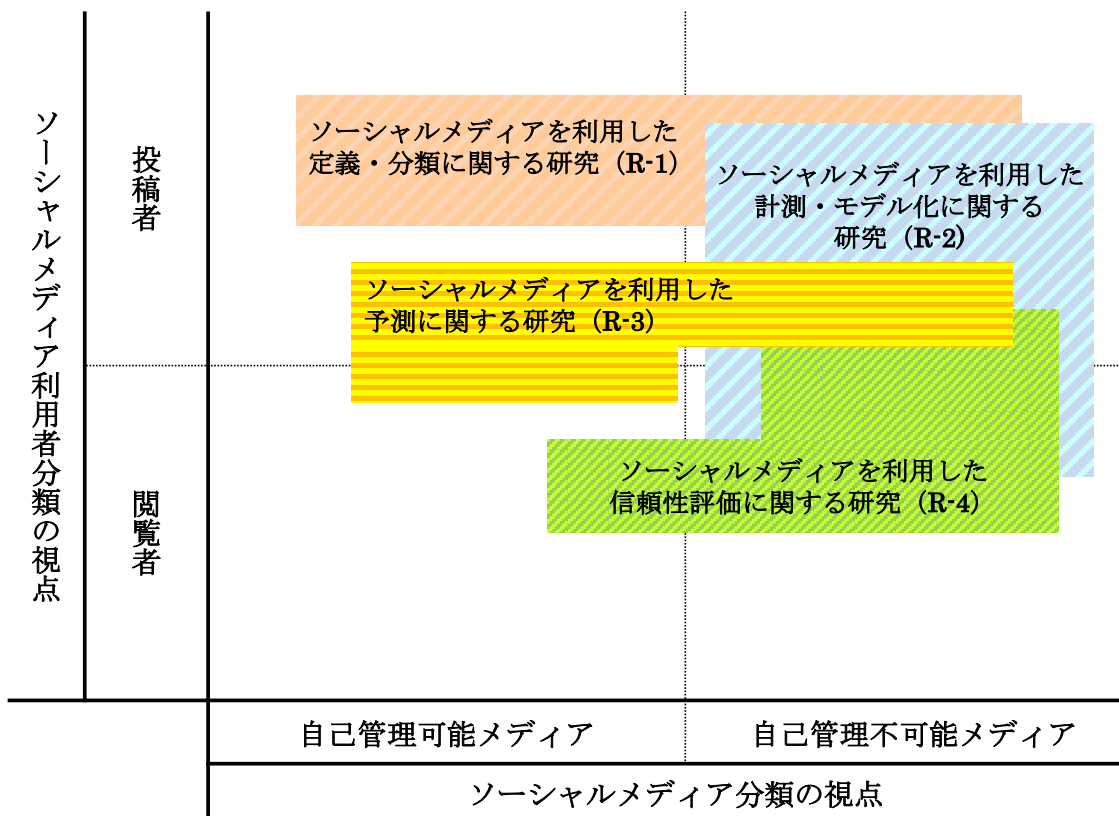
### 2.3.1. ソーシャルメディア利用者の行動分析に関する先行研究の分類

ソーシャルメディア利用者の行動分析に関する先行研究を整理すると、以下の4つに分類できる。

- ソーシャルメディアを利用した定義・分類に関する研究 (R-1)
- ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)
- ソーシャルメディアを利用した予測に関する研究 (R-3)
- ソーシャルメディアを利用した信頼性評価に関する研究 (R-4)

「ソーシャルメディアを利用した定義・分類に関する研究 (R-1)」は、ソーシャルメディア利用者が投稿した内容を手掛かりにして、その内容や利用者の行動様式などを自己管理可能メディアか自己管理不可能メディアかを問わず定義・分類する研究領域である。「ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)」は、ソーシャルメディア利用者が投稿もしくは閲覧した内容を基礎として、行動の変化や影響の広がりなどを主に自己管理不可能メディアにおいて計測・モデル化する研究領域である。「ソーシャルメディアを利用した予測に関する研究 (R-3)」は、ソーシャルメディア利用者が投稿もしくは閲覧した結果を用いて、実社会における人々の行動や現象を自己管理可能メディアか自己管理不可能メディアかを問わず予測する研究領域である。「ソーシャルメディアを利用した信頼性評価に関する研究 (R-4)」は、ソーシャルメディア利用者が投稿、もしくは閲覧した内容を手掛かりに内容の信頼性を自己管理可能メディアか自己管理不可能メディアかを問わず評価する研究領域である。

以上のソーシャルメディア利用者の行動分析に関する先行研究（「ソーシャルメディアを利用した定義・分類に関する研究 (R-1)」、「ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)」、「ソーシャルメディアを利用した予測に関する研究 (R-3)」、「ソーシャルメディアを利用した信頼性評価に関する研究 (R-4)」）を、2.1 節、2.2 節で述べたソーシャルメディア分類の視点（自己管理可能メディア、自己管理不可能メディア）と、ソーシャルメディア利用者分類の視点（投稿者、閲覧者）で分類したものが図 2.11 である。以下では、図 2.11 に示す 4 つの研究領域毎に、その先行研究について示す。



- R-1 : [Golder 2011], [Grinev 2009], [Cohn 2004], [Hollenstein 2013], [Smith 2012], [Liu 2010], [Honey 2009], [Longueville 2009], [Weng 2010], [Kwak 2010], [Kaplan 2010], [Kietzmann 2011], [Berthon 2012], [Kietzmann 2012], [Laroche 2012]
- R-2 : [Lerman 2010], [Cha 2010], [Akioka 2010], [Ye 2011], [Bakshy 2009], [Goh 2013], [Burke 2011], [Boyd 2010], [Vieweg 2010], [Susaria 2012], [Gruhl 2004], [Romero 2011], [Cheong 2009], [Li 2008], [Kurashima 2006], [Shamma 2009], [Uehara 2005], [上原 2008], [Cha 2009], [Wilson 2008], [Brown 2002], [Muniz 2001], [Krishnamurthy 2008], [Java 2007], [Adamic 2008], [Mislove 2008], [Leskovec 2008], [Ellison 2007], [Valenzuela 2009], [Hampton 2011], [Lewis 2008], [Kanai 2012], [Quercia 2012], [Jung 2013], [Steinfeld 2008]
- R-3 : [Liu 2007], [Gruhl 2005], [Asur 2010], [Duan 2008], [Goel 2010], [Dupret 2006], [Baeza-Yates 2006], [Tumasjan 2010], [Bollen 2011], [Tumarkin 2001], [Kholodilin 2009], [Schmidt 2011], [Choi 2012], [McLaren 2011], [Askitas 2009], [Acquisti 2009], [Gilbert 2009], [Gilbert 2012], [Backstrom 2011], [Wang 2011], [Yang 2011], [Leskovec 2010], [Schifanella 2010], [Liu 2008], [Golbeck 2009], [Golbeck 2011], [Suzuki 2010], [Hong 2011], [Szabo 2010], [Gomez 2010], [Ginsberg 2008], [Polgreen 2008], [Johnson 2004]
- R-4 : [Resnick 2000], [Jindal 2008], [Guo 2011], [Ott 2011], [Benevenuto 2010], [Lim 2010], [Kolari 2006], [Lin 2007], [Guzella 2009], [Jindal 2010], [Fogg 2001], [Wathen 2002], [Dwyer 2007], [Fogel 2009], [DuBoi 2011], [Denning 2005], [Adler 2006], [Vuong 2008], [Kittur 2006], [Kittur 2008], [Agichtein 2008], [Suryanto 2009], [Bian 2008], [Mendoza 2010], [Castillo 2011]

図 2.11 先行研究の分類

### 2.3.2. ソーシャルメディアを利用した定義・分類に関する研究 (R-1)

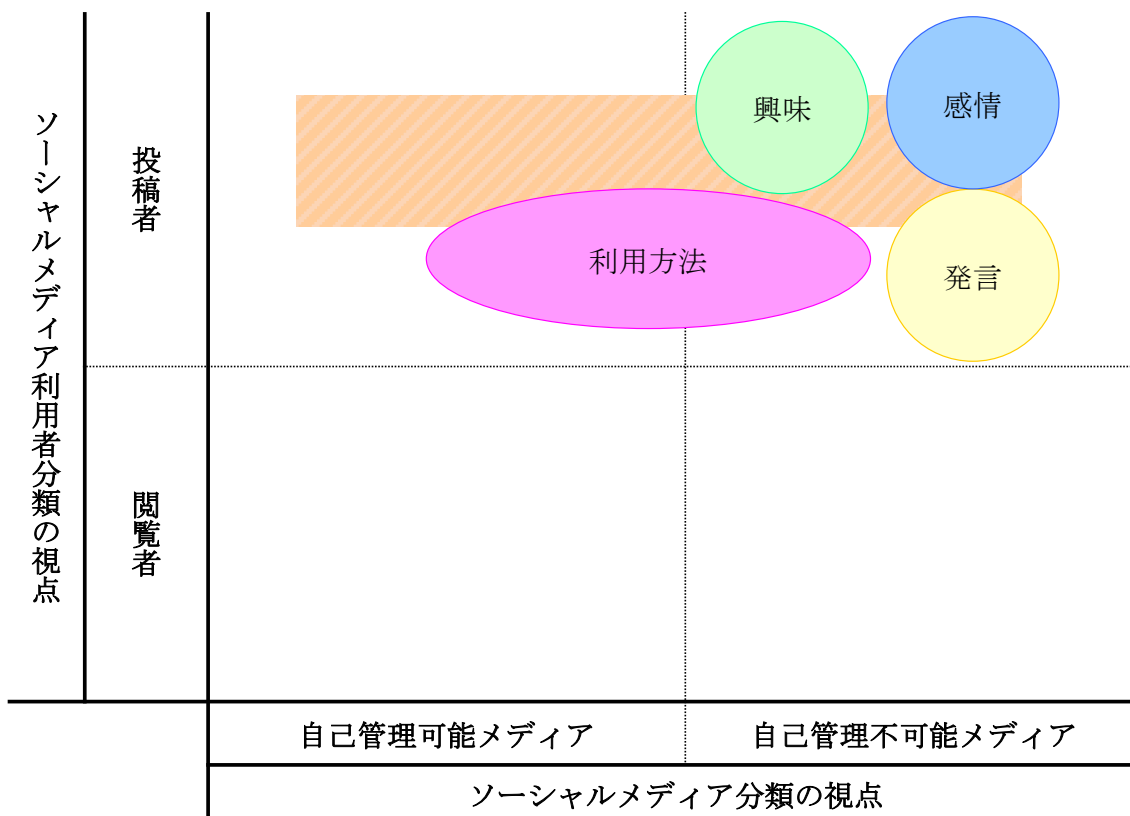
ソーシャルメディアを利用した定義・分類に関する先行研究は、「感情に関わる定義・分類に関する研究」、「興味に関わる定義・分類に関する研究」、「発言に関わる定義・分類に関する研究」、「利用方法に関わる定義・分類に関する研究」の4つに分類できる。その分類を先行研究の分類にマッピングしたものが図 2.12 である。主に投稿者側の視点から、自己管理可能メディア、自己管理不可能メディアの各々でソーシャルメディアの定義・分類に関して分析している。

#### (1) 感情に関わる定義・分類に関する研究

感情に関わる定義・分類に関する研究には、[Golder 2011], [Grinev 2009], [Cohn 2004] などがある。図 2.12 に示すように、主に投稿者側の視点から、自己管理不可能メディアに対する投稿内容をもとに、ソーシャルメディア利用者の感情の分類を試みている。

Golder は、84 か国、240 万人の Twitter への投稿内容をもとに、人の気分や感情の流れを分析している。投稿内容から感情を示す単語を抽出し、肯定的な意味の単語と否定的な意味の単語にグループ分けを行っている。一日の中で一番ポジティブな感情を示したのは朝であるが時間と共に減少していき、逆に日中はネガティブな感情を示す単語が増加すると述べている。また平日よりも休日、冬よりも夏にポジティブな単語は増加したことを示している[Golder 2011]。

同様に Grinev は、Twitter の投稿内容を用いて、どの程度ブランドに言及するかの調査を行っている。ブランドに対する言及の 20%に感情を表す情報が付いており 50%以上がポジティブな情報内容であり、33%がネガティブな情報内容であったと述べている[Grinev 2009]。Cohn は、同様に Twitter への投稿内容を分析し、2001 年 9 月 11 日に発生したアメリカ同時多発テロにおける人々の感情の変化を明らかにしている[Cohn 2004]。



感情 : [Golder 2011], [Grinev 2009], [Cohn 2004]  
 興味 : [Hollenstein 2013], [Smith 2012], [Liu 2010]  
 発言 : [Honey 2009], [Longueville 2009], [Weng 2010], [Kwak 2010]  
 利用方法 : [Kaplan 2010], [Kietzmann 2011], [Berthon 2012], [Kietzmann 2012],  
 [Laroche 2012]

図 2.12 ソーシャルメディアを利用した定義・分類に関する研究 (R-1)

(2) 興味に関わる定義・分類に関する研究

興味に関わる定義・分類に関する研究には, [Hollenstein 2013], [Smith 2012], [Liu 2010]などがある. 図 2.12 に示すように, 主に投稿者側の視点から, 自己管理不可能メディアに対する投稿内容をもとに, 人間の興味に関する分類を試みている.

Hollenstein は, Flickr のタグを使った旅行時の推薦情報に対する興味度合いについて分類を行い, モバイル利用の場合と比較を行っている [Hollenstein 2013]. Smith は, Twitter,

Facebook, YouTube におけるブランド関連のユーザー生成のコンテンツ (UGC) の 600 の投稿内容を分析し, 2 つの小売衣類ブランドへの興味について分析している [Smith 2012]. Liu は, 投稿内容を “事実” と “意見” に分け, 更に “意見” について, 興味度合いが “ポジティブ” か “ネガティブ” かに分けることを試みている [Liu 2010].

### (3) 発言に関わる定義・分類に関する研究

発言に関わる定義・分類に関する研究には, [Honey 2009], [Longueville 2009], [Weng 2010], [Kwak 2010] などがある. 図 2.12 に示すように, 主に投稿者側の視点から, 自己管理不可能メディアに対する投稿内容をもとに, ソーシャルメディア利用者の発言に関する分類を試みている.

Honey と Longueville は, Twitter 上での発言内容に現れるフレーズに着目している ([Honey 2009], [Longueville 2009]). Honey は, Twitter 上での投稿内容に “@UserName” が含まれるかどうかに着目し, 約 30% 程度含まれていることを示している [Honey 2009]. Longueville は, 2009 年にフランスで発生した森林火災に関して Twitter 上での投稿内容を分析し, 投稿者と投稿内容で引用した URL の参照内容について分類している [Longueville 2009].

Weng と Kwak は, Twitter 上での投稿内容の広がりに着目している ([Weng 2010], [Kwak 2010]). Weng は, Twitter 上での投稿の広がりや全投稿者の中で, 他の投稿者に影響がある投稿をした投稿者の分類を試みている [Weng 2010]. ここでの影響がある投稿とは, 広い範囲に伝搬する投稿のことである. 分類の結果, 投稿者の “フォロー” 関係には趣向に対する類似性があることを示している.

Kwak は, Twitter の 4170 万のユーザープロフィール, 14 億 7000 万のつながり, 4,262 個のトピックと 1 億 600 万のつぶやきを対象にした大規模な分析を行っている [Kwak 2010]. インフルエンサーを特定するために, フォロワー数とページランクで投稿者のランク付けを行い, 2 つのランキングが類似していると述べている. しかし Retweets によるランキングは前の 2 つのランキングと異なり, フォロワー数やページランクとは関係がなかったと述べている. またトピックの分類も行い, トピックの大多数 (85% 以上) がヘッドライン・ニュースまたは永続的なニュースであったことを示している.

#### (4) 利用方法に関わる定義・分類に関する研究

利用方法に関わる定義・分類に関する研究には、[Kaplan 2010], [Kietzmann 2011], [Berthon 2012], [Kietzmann 2012], [Laroche 2012]などがある。図 2.12 に示すように、主に投稿者側の視点から、自己管理不可能メディアに対する活用方法に関する分類を試みている。

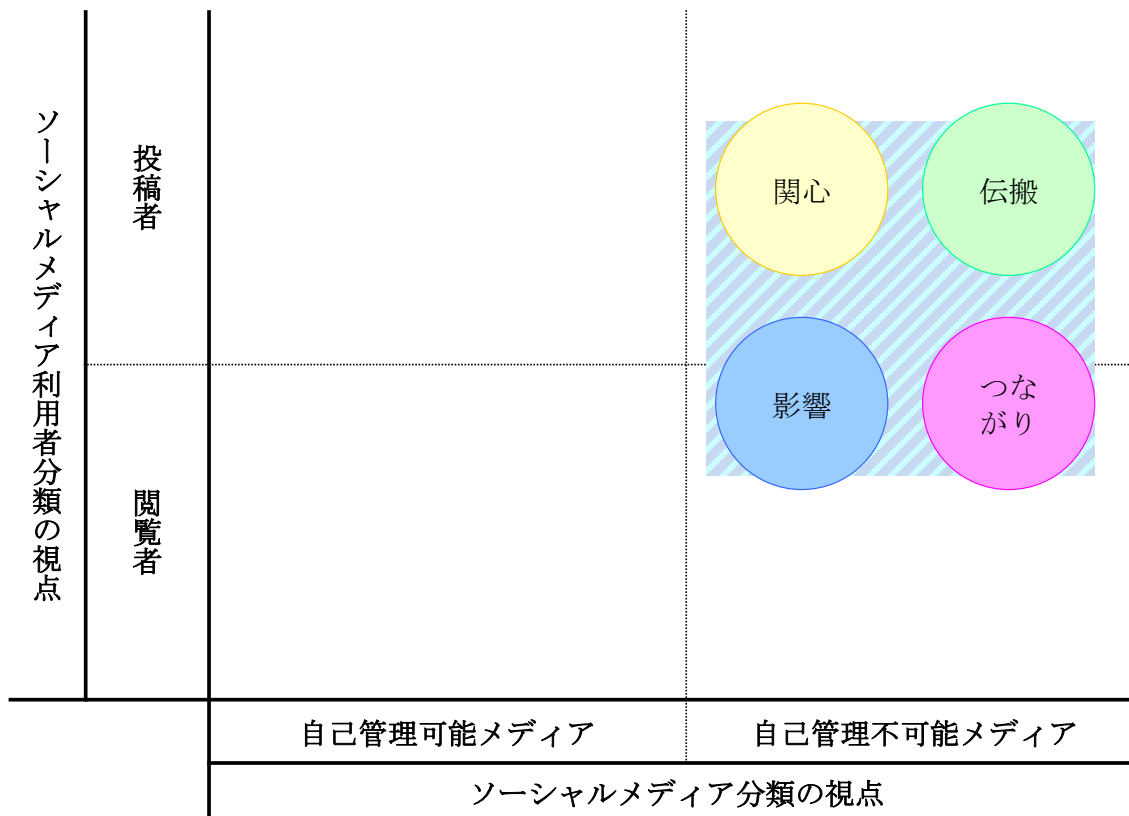
Kaplan は、ソーシャルメディアの利用方法は、社会的存在感(Social Presence), メディアリッチネス(Media Richness), 自己呈示(Self-presentation)/自己開示(Self-disclosure)という指標で分類可能であると述べている[Kaplan 2010]。例えばブログは、自己呈示/自己開示性が高いが社会的存在感、メディアリッチネスは低く、SNS はブログ同様に自己呈示/自己開示性は高いが、社会的存在感/メディアリッチネスはブログよりも高いとしている。

Kietzmann は、同様にソーシャルメディアを利用方法からの視点で分類し、“Identity”, “Conversations”, “Sharing”, “Presence”, “Relationships”, “Reputation”, “Group” と分類している[Kietzmann 2011]。Berthon は、同様の視点で①ソーシャルメディアは、常に技術、文化、特有の国または背景を持った機能である、②ローカル・イベントは、めったにローカルのままではない、③世界的なイベントはローカルに（再）解釈される可能性がある、④創造的なコンシューマの活動と創造物は、技術、文化、政治にも依存している、⑤技術は、歴史に依存している、という 5 つの特徴に分類している[Berthon 2012]。[Kietzmann 2012], [Laroche 2012]も同様の視点で分類している。

#### (5) ソーシャルメディアを利用した定義・分類に関する研究 (R-1) の抱える課題

本先行研究領域は、主に投稿者側の視点からソーシャルメディアへの投稿内容を「感情に関わる定義・分類に関する研究」、「興味に関わる定義・分類に関する研究」、「発言に関わる定義・分類に関する研究」、「利用方法に関わる定義・分類に関する研究」の 4 領域で分析している。しかし、投稿内容を分析しているため、作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”を抱えている。





影響 : [Lerman 2010], [Cha 2010], [Akioka 2010], [Ye 2011], [Bakshy 2009], [Goh 2013], [Burke 2011]

伝搬 : [Boyd 2010], [Vieweg 2010], [Susaria 2012], [Gruhl 2004], [Romero 2011]

関心 : [Cheong 2009], [Li 2008], [Kurashima 2006], [Shamma 2009], [Uehara 2005], [上原 2008], [Cha 2009], [Wilson 2008], [Brown 2002], [Muniz 2001]

つながり : [Krishnamurthy 2008], [Java 2007], [Adamic 2008], [Mislove 2008], [Leskovec 2008], [Ellison 2007], [Valenzuela 2009], [Hampton 2011], [Lewis 2008], [Kanai 2012], [Quercia 2012], [Jung 2013], [Steinfeld 2008]

図 2.13 ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)

### 2.3.3. ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)

ソーシャルメディアを利用した計測・モデル化に関する先行研究は、「影響に関わる計測・モデル化に関する研究」、「伝搬に関わる計測・モデル化に関する研究」、「関心に関わる計測・モデル化に関する研究」、「つながりに関わる計測・モデル化に関する研究」の4つに分類できる。その分類を、先行研究の分類にマッピングしたものが図 2.13 である。本研究領域には、投稿者側の視点と閲覧者側の視点の両視点で、自己管理不可能メディアに対して分析を試みた研究事例がある。特に、「伝搬に関わる計測・モデル化に関する研究」、「関心に関わる計測・モデル化に関する研究」の研究領域は、主に投稿者の視点で分析を試みているが、「影響に関わる計測・モデル化に関する研究」、「つながりに関わる計測・モデル化に関する研究」の2つの研究領域については、投稿者側の視点、閲覧者側の視点の両視点で分析を試みている。

#### (1) 影響に関わる計測・モデル化に関する研究

影響に関わる計測・モデル化に関する研究には、[Lerman 2010], [Cha 2010], [Akioka 2010], [Ye 2011], [Bakshy 2009], [Goh 2013], [Burke 2011]などがある。図 2.13 に示すように、投稿者側の視点では投稿内容を用い、閲覧者側の視点では主にアンケート結果を用い分析を試みている。

「投稿者—自己管理不可能メディア」領域において Lerman は、Digg と Twitter 上のアクティブな投稿者のソーシャル・ネットワークを抽出し、ニュースへの関心がそれらの間でどの様に広がるかを計測しモデル化している[Lerman 2010]。抽出したソーシャル・ネットワークは、ニュースの普及において重要な役割を果たし、そのネットワーク構造が情報の流れのダイナミクスに影響を与えることを示している。

Cha と Akioka は、Twitter 上での投稿内容における影響計測を行っている([Cha 2010], [Akioka 2010])。Cha は、Twitter の入次数 (indegree) と Retweets に着目し、影響に関する 3 つの特徴を述べている[Cha 2010]。各特徴は、①高い indegree を持つ投稿者が Retweets または Mentions<sup>13</sup>を行うことに関して必ずしも影響力があるわけではない、②最も影響力がある投稿者は、いろいろなトピックに対する大きな影響力を持つことができる、③影響力は自然に偶然に増えるのではなく、一つのトピックに絞ったツイートの様な協調したやり取りを通じて増える、としている。Akioka は、日本における Twitter 投稿者

<sup>13</sup> 特定の「@ユーザー名」を含むツイート

の特徴と、出版物やテレビ番組などのメディアが Twitter のコミュニティに与える影響について分析を行っている[Akioka 2010]. 同様に, Ye は旅行サイトへの投稿内容の影響を, Bakshy は second life<sup>14</sup>内での投稿内容における投稿者間での影響を調査している ([Ye 2011], [Bakshy 2009]).

「閲覧者—自己管理不可能メディア」領域において Goh と Burke は, Facebook 上での投稿内容における影響を, 閲覧者側からの視点で計測しモデル化している ([Goh 2013], [Burke 2011]). Goh は, Facebook での UGC(User-Generated Content) と MGC (Marketer-Generated Content)の消費者購買行為への影響を定量化して計測し, UGC は消費者購買行為において MGC より強い影響を示したとしている. Burke は, Facebook 利用者に対し 8 か月間の調査前後で広告に対するアンケートを実施し, Facebook の利用でどんな影響があるかを計測している[Burke 2011]. 結果は, ダイレクトコミュニケーションの流入量は影響力があり Social Capital の増加と相関関係があったが, 流出量には関係がなく受動消費やブロードキャストにも有意な影響はなかったと述べている. また, コミュニケーションスキルの高い人と低い人という 2 つのグループに回答者を分け, それぞれが Social Capital にどんな影響をあたえるかを計測し, ダイレクトコミュニケーションは両グループに強い影響を与え, 受動消費はコミュニケーションスキルの高いグループに影響しないが低いグループには大きな影響があると述べている.

## (2) 伝搬に関わる計測・モデル化に関する研究

伝搬に関わる計測・モデル化に関する研究には, [Boyd 2010], [Vieweg 2010], [Susaria 2012], [Gruhl 2004], [Romero 2011]などがある. 図 2.13 に示すように, 主に投稿者側の視点から自己管理不可能メディアへの投稿内容を用いて分析を試みている.

Boyd, Vieweg は Twitter 上で, Susaria は YouTube 上で, Gruhl はブログ上で, 投稿された情報がどのように伝搬されるかを調査している ([Boyd 2010], [Vieweg 2010], [Susaria 2012], [Gruhl 2004]). たとえば Vieweg は, 2009 年に発生したオクラホマの火事やレッドリバーでの洪水における情報がどのように伝搬していったのかを分析している. また Gruhl は, ブログ間の関係においてトピックがどのように伝搬されるかをモデル化している.

Romero は, ソーシャルメディア内で情報を広く伝搬させる条件について分析を行って

---

<sup>14</sup> <http://secondlife.com/> (2013/10/13)

いる[Romero 2011]. 広く伝搬させるには、各個人が他のメンバーに情報を転送する必要があり、積極的に伝搬に関与する必要があるとしている。また自分とつながっている人やそのフォロワーたちが自らの投稿を読み、それを他者と共有し、そこに書かれたことに基づいて行動していなければ自分の投稿は意味を持たず、他の投稿との違いを生み出してはいないと述べている。

### (3) 関心に関わる計測・モデル化に関する研究

関心に関わる計測・モデル化に関する研究には、[Cheong 2009], [Li 2008], [Kurashima 2006], [Shamma 2009], [Uehara 2005], [上原 2008], [Cha 2009], [Wilson 2008], [Brown 2002], [Muniz 2001]などがある。図 2.13 に示すように、主に投稿者側の視点から自己管理不可能メディアへの投稿内容を用いて分析を試みている。

Cheong と Li は、投稿内容からソーシャルメディア利用者の関心を可視化する手法を提案している ([Cheong 2009], [Li 2008], [Kurashima 2006])。Cheong は、地域、性別、趣味など関心属性別にマイクロブログのコンテンツを分析している [Cheong 2009]。属性ごとにコンテンツの内容が大きく異なること示し、コンテンツ内容の可視化手法を提案している。Li は、ソーシャルメディアにおける利用者タグの頻繁な共起パターンと利用者の関心を特徴づける、インターネットにおける社会的関心を可視化するシステムを開発したと述べている [Li 2008]。Kurashima は、ブログから街の話題に関する言葉を集め、その街における人々の体験として集約し、地図上へのマッピングを試みている [Kurashima 2006]。

Shamma は、テレビ放送で中継されているアメリカ大統領選挙中の党首討論の流れと同時期の Twitter の議論の流れを解析している [Shamma 2009]。Twitter の投稿から動画のアノテーションを試み、議論に関する関心の流れを明らかにしている。同様の研究に [Uehara 2005], [上原 2008] がある。Uehara は、インターネット上の視聴者コミュニティで展開される対話文をもとにして、テレビドラマ番組に対する視聴者コミュニティの関心状態の特徴パターンを番組進行と同期した時間軸上にマッピングした「注目状態グラフ」を生成する手法を提案している。また上原は、インターネット掲示板を用い、投稿内容からテレビ CM に関する語彙群を抽出し、その時系列データから現実社会での CM における関心度を計測している。

Cha と Wilson は、関心がどの様に移り変わるかを分析している ([Cha 2009], [Wilson 2008])。Cha は、Flickr に投稿された 1100 万の写真、250 万人の利用者を分析しロコミ

による関心がどの様に移り変わるのかを明らかにしている[Cha 2009]. Wilson は、ターゲットをアメリカの年配層に絞った SNS であるブログスフィアへの投稿内容を分析し、投稿の位相的パターンを調べることによって関心の移り変わりの位相的構造に関する情報について述べている[Wilson 2008].

Brown と Muniz は、投稿内容から消費者の購買行動に関する関心を計測している ([Brown 2002], [Muniz 2001]). Brown は、仮想コミュニティのユーザーと非ユーザーを比較しオンライン・ショッピングサイトの利用率やリピート率を計測し、その違いを示している. Muniz は、仮想コミュニティの存在と消費者の購買行動におけるブランド力の関係を計測している.

#### (4) つながりに関わる計測・モデル化に関する研究

つながりに関わる計測・モデル化に関する研究には、[Krishnamurthy 2008], [Java 2007], [Adamic 2008], [Mislove 2008], [Leskovec 2008], [Ellison 2007], [Valenzuela 2009], [Hampton 2011], [Lewis 2008], [Kanai 2012], [Quercia 2012], [Jung 2013], [Steinfeld 2008]などがある. 図 2.13 に示すように、投稿者側の視点でも閲覧者側の視点でも分析は行われており、特に閲覧者側の視点からの分析では、主にアンケート結果を用いた分析を試みている.

「投稿者—自己管理不可能メディア」領域において Krishnamurthy と Java は、Twitter 上でのつながりを、投稿者側からの視点で計測しモデル化している ([Krishnamurthy 2008], [Java 2007]). Krishnamurthy は、2008 年 1 月時点での Twitter の投稿者のつながりについての分析し、情報を提供する “broadcasters”, 知人とのコミュニケーションを中心行う “acquaintances”, 閲覧を中心とした “miscreants or evangelists” の 3 グループがあると述べている. また友人グループの特徴は Tweet 数が増えるほど followed と following がより相関関係に近づくとしている. Java は、Twitter 利用者のつながり分析し、アジア、ヨーロッパのコミュニティは北米よりも内部の結びつきは強く、日本語、スペイン語の利用者は同一言語利用者と結びついている可能性が高いと述べている [Java 2007].

Adamic は Yahoo Answers, Mislove は Flickr, Leskovec はマイクロソフト・インスタントメッセージの各つながりについて計測しモデル化している ([Adamic 2008], [Mislove 2008], [Leskovec 2008]). Adamic は、Yahoo Answers を分析し、つながり方

を主に「知識共有」,「相談」,「議論」の3つにモデル化している[Adamic 2008]. Mislove は, Flickr の3ヵ月間, 950,143人の利用者, 970万以上のリンクを含んだデータを用いてリンク構造プロセスをモデル化している[Mislove 2008]. 結果, ソースへのリンクをすでに多く持つ利用者が, 親しい人とリンクを作ることで, 更に他の利用者へリンクを提供する傾向があると示している. Leskovec は, マイクロソフト・インスタントメッセージにおける1か月間, 2億4000万人, 300億のメッセージを分析し6.6次の隔たりがあったと述べている[Leskovec 2008].

「閲覧者—自己管理不可能メディア」領域において Ellison と Valenzuela は, Facebook における Social Capital に着目し, 閲覧者側からの視点で学生のつながりを分析している([Ellison 2007], [Valenzuela 2009]). Ellison は, 2006年時点でのミシガン州立大学の学生が Facebook をどのように利用しているのかについて調査している[Ellison 2007]. 結果, ソーシャルメディアは Social Capital を強化し人間関係を補強すること, 自尊心及び生活満足度が低い人ほどソーシャルメディアを利用して Social Capital を作ろうとする傾向にあることを示している. Valenzuela は, Facebook の利用と個人の Social Capital を高める態度や行動の間に関連が見られるかを検証している[Valenzuela 2009]. テキサス州の学生2,603名を対象に Web 調査を行い, Facebook の利用度と学生の生活満足, 社会的信頼, 市民参加, 政治参加との間に正の相関があったと述べている.

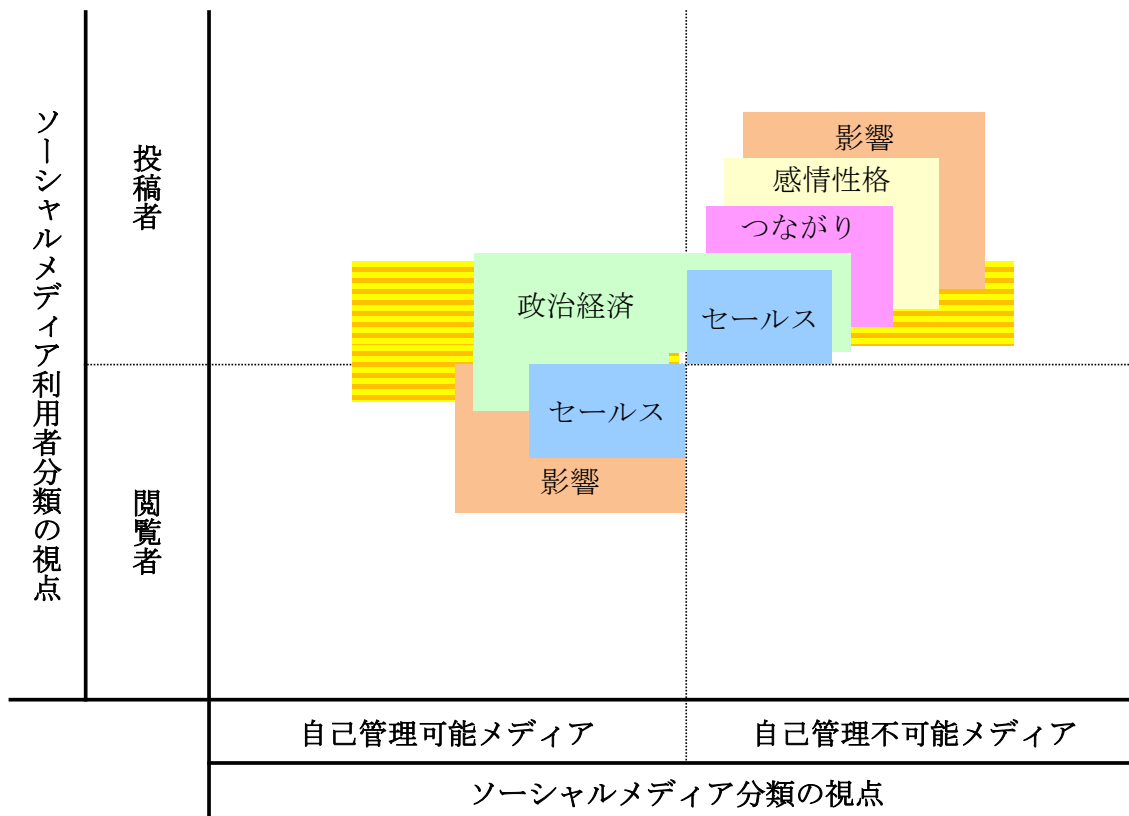
Hampton と Lewis は, Facebook 上での友人関係と実社会での友人関係に着目し, つながり関係を計測しモデル化している([Hampton 2011], [Lewis 2008]). Hampton は, 2010年10月20日から11月28日にかけて米国内に住む18歳以上の Facebook 利用者にインタビューを実施し, 利用者の平均友人数は229名で, その友人の中で一度も会ったことのない人は7%, 一度しか会ったことがない人は3%で, 残りの90%は2回以上会ったことのある人であったと述べている[Hampton 2011]. また, どんな人と Facebook 上でつながっているのかについての調査も行い, 高校からの知り合いが22%, 親せき12%, 同僚が10%, 大学からの友人が9%, 家族が8%, 課外活動での友人が7%, 近所の人2%であったとしている. Lewis は, Facebook の友人関係について調査し, お互いの写真を交換している友人関係は, 交換していない友人に比べ映画, 音楽, 本などの他の好みも非常に共通しており, つながりが強いことを明らかにしている[Lewis 2008].

Kanai と Quercia, Jung, Steinfield は, Facebook におけるつながりと精神的な関係を計測している([Kanai 2012], [Quercia 2012], [Jung 2013], [Steinfeld 2008]). Kanai

は、Facebook を利用しているイギリスの大学生 125 人の脳を MRI で調べ、アカウント数や送信メッセージ数、友人数などを調査している[Kanai 2012]。結果、脳灰白質密度と現実世界の友人関係の広さ、オンライン上の友人関係の広さのどちらにも相関関係があったことを示している。同様に Quercia と Jung もアンケートを用いて、Facebook 内でのつながり方と現実世界のつながり方の関係について調査をしている ([Quercia 2012], [Jung 2013])。Steinfeld は、Facebook の利用度と精神的な健康状態(自尊心と人生の満足度)と Social Capital との関係性を調査している[Steinfeld 2008]。対象はアメリカの大学生約 800 名、この内 18 名の男女は質問表調査の他にインタビューも行ない、自尊心が高い学生よりも低い学生の方が Facebook の使用から得られる利益が高く、自尊心が低い若者は人と接触する際の困難を緩和する役割を持っていると述べている。

#### (5) ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2) の抱える課題

本研究領域は数多くの研究が行われているが、主に投稿者の投稿内容を分析している。また閲覧者側の視点での研究も存在するが、主に数百人規模くらいまでのアンケート調査結果を用いて分析を行っている。投稿内容を分析しているため作為的な投稿者の投稿内容の影響を受けやすいという“作為的投稿影響問題”や[Nielsen 2006b]や[Nielsen 2013]で示されているように、アンケート実施時の“利用者バイアス問題”と“調査規模的問題”に対して注意深い扱いを必要とする。



セールス : [Liu 2007], [Gruhl 2005], [Asur 2010], [Duan 2008], [Goel 2010], [Dupret 2006], [Baeza-Yates 2006]

政治・経済 : [Tumasjan 2010], [Bollen 2011], [Tumarkin 2001], [Kholodilin 2009], [Schmidt 2011], [Choi 2012], [McLaren 2011], [Askitas 2009], [Acquisti 2009]

つながり : [Gilbert 2009], [Gilbert 2012], [Backstrom 2011], [Wang 2011], [Yang 2011], [Leskove 2010], [Schifanella 2010]

感情・性格 : [Liu 2008], [Golbeck 2009], [Golbeck 2011]

影響 : [Suzuki 2010], [Hong 2011], [Szabo 2010], [Gomez 2010], [Ginsberg 2008], [Polgreen 2008], [Johnson 2004]

図 2.14 ソーシャルメディアを利用した予測に関する研究 (R-3)



### 2.3.4. ソーシャルメディアを利用した予測に関する研究 (R-3)

ソーシャルメディアを利用した予測に関する先行研究は、「セールスに関わる予測に関する研究」、「政治・経済に関わる予測に関する研究」、「つながりに関わる予測に関する研究」、「感情・性格に関わる予測に関する研究」、「影響に関わる予測に関する研究」の5つに分類できる。その分類を、先行研究の分類にマッピングしたものが図 2.14 である。本研究領域は、投稿者側の視点でも、閲覧者側の視点でもソーシャルメディア利用者の予測について分析が行われている。投稿者側の視点では投稿内容を用いているが、閲覧者側の視点では、主に検索エンジンの出力結果を用いた分析を行っている。特に、「セールスに関わる予測に関する研究」、「政治・経済に関わる予測に関する研究」の2つの研究領域については、投稿者側の視点、閲覧者側の視点の両視点で分析を試みている。

#### (1) セールスに関わる予測に関する研究

セールスに関わる予測に関する研究には、[Liu 2007], [Gruhl 2005], [Asur 2010], [Duan 2008], [Goel 2010], [Dupret 2006], [Baeza-Yates 2006]などがある。図 2.14 に示すように、主に投稿者側の視点では、自己管理可能メディア、自己管理可能メディアの両メディアに対する投稿内容を用い、閲覧者側の視点では主に自己管理可能メディアにおける検索エンジンの出力結果を用いて予測分析を試みている。

「投稿者—自己管理不可能メディア」領域において Liu と Gruhl は、ブログへの投稿を製品セールス予測に利用するためのモデルを構築している ([Liu 2007], [Gruhl 2005])。Liu は、ブログへの投稿から感情認識モデルを構築し、感情情報を利用した場合、そうでない場合よりも精度よく売上を予測できる事を示している [Liu 2007]。Gruhl は、書籍に言及しているブログへの書き込み数の時系列推移が、当該書籍の電子商取引サイトでの売り上げランキングへ及ぼす影響を分析している [Gruhl 2005]。Asur と Duan は、映画興行予測を行っている ([Asur 2010], [Duan 2008])。Asur は、映画が公開される前の Twitter への投稿内容を使って公開一週間目のボックスオフィス売上を予測している [Asur 2010]。Tweet 数で 97%の精度で予測し、Hollywood Stock Exchange (HSX)<sup>15</sup>での取引を用いた予測より正確で、Tweet の内容がネガティブかポジティブかの感情情報を分析に加えるとさらに正確になることを示している。Duan は、Yahoo! Movies<sup>16</sup>, Variety.com<sup>17</sup>,

<sup>15</sup> Hollywood Stock Exchange (<http://www.hsx.com/>) (2013/10/13)

<sup>16</sup> Yahoo! Movies(<http://www.movies.yahoo.com>) (2013/10/13)

BoxOfficeMojo.com<sup>18</sup>への投稿内容から興行予測を行っている。

「閲覧者—自己管理可能メディア」領域において Goel は、Web の検索でのヒット数が映画やビデオゲーム、音楽などのセールスと関係があり時間的変化まで類似していると述べている[Goel 2010]。またエンタテインメント分野だけでなくインフルエンザの流行と、関係するブログへの投稿数の一致なども時系列的なデータで提示している。Dupret と Baeza-Yates は、クリックスルーログから利用者の関心を予測するモデルを提案している ([Dupret 2006], [Baeza-Yates 2006])。Dupret は、Attractiveness(誘引性) や Relevance(妥当性) を推定する方法の提案し、Baeza-Yates は自動的に利用者の興味を示すフレームを提示する方法を提案している。

## (2) 政治・経済に関わる予測に関する研究

政治・経済に関わる予測に関する研究には、[Tumasjan 2010], [Bollen 2011], [Tumarkin 2001], [Kholodilin 2009], [Schmidt 2011], [Choi 2012], [McLaren 2011], [Askitas 2009], [Acquisti 2009]などがある。図 2.14 に示すように、主に投稿者側の視点から自己管理不可能メディアに対する投稿内容を用いて予測分析を行っている。また閲覧者側の視点では、主に検索エンジンの出力結果を用いた予測分析を試みている。

「投稿者—自己管理不可能メディア」領域において Tumasjan は、Twitter の投稿数からドイツ議会選挙の選挙結果を予測し、投稿内容の感情を分析することで当時の政治状況の考察も行っている[Tumasjan 2010]。Bollen は、Twitter の投稿内容を利用し、ダウ平均株価 (DJIA) のアップダウンを 87.6% の精度で予測できたと述べている[Bollen 2011]。2008 年の 10 ヶ月間、利用者数 270 万人、Tweet 数 980 万回のデータから、感情を測定する 2 つのツール (OpinionFinder, Google-Profile of Mood States 「GPOMS」<sup>19</sup>) を使用し、測定結果と DJIA の終値とを比較することで予測できたとしている[Bollen 2011]。同様に Kumarkin も、個人投資家の掲示板への書き込み頻度から、株式市場の収益率や取扱高の時系列の推移を予測している[Tumarkin 2001]。また、「投稿者—自己管理可能メディア」領域において Acquisti は、住所と誕生日に関する情報があれば、その人物の社会保障番号 (SSN) が予測できると述べている[Acquisti 2009]。

---

<sup>17</sup> Variety.com (Variety: <http://www.variety.com>) (2013/10/13)

<sup>18</sup> BoxOfficeMojo.com (Mojo: <http://www.boxofficemojo.com>) (2013/10/13)

<sup>19</sup> 社会全体の気分の尺度を平穏、警戒、確信、活気、善意、幸福に分類

「投稿者－自己管理不可能メディア」と「閲覧者－自己管理可能メディア」の領域において Kholodilin と Schmidt は、閲覧者側からの視点で検索エンジンの出力結果を用いて個人消費の予測を行っている ([Kholodilin 2009], [Schmidt 2011])。Kholodilin は、検索データを説明変数に用いたモデルの方が、消費者コンフィデンス指標を説明変数に用いたモデルより予測力が高く、特に平常時よりリーマン・ショックの様な緊急時の方が、予測力が高いことを示している。Schmidt も同様の結果を示し、検索データの方がマクロ経済指標（所得、金利、株価）を説明変数に加えたモデルよりも予測力が高かったと述べている。

Choi と McLaren, Askitas は、投稿内容を分析し経済指標の予測を行っている ([Choi 2012], [McLaren 2011], [Askitas 2009])。Choi は、小売販売額、自動車販売、住宅販売、旅行に関する経済指標の予測を行い、単純な時系列モデルに関連する検索データを説明変数に加えると、特に自動車販売や住宅販売については予測力が高くなると述べている。McLaren も同様に、住宅価格指数や失業者数の予測において関連する検索データを説明変数に加えると予測力が高くなると述べている [McLaren 2011]。Askitas は、就職支援サイト名の検索データから失業率の予測を行い高い予測力を示している [Askitas 2009]。

### (3) つながりに関わる予測に関する研究

つながりに関わる予測に関する研究には、[Gilbert 2009], [Gilbert 2012], [Backstrom 2011], [Wang 2011], [Yang 2011], [Leskove 2010], [Schifanella 2010]などがある。図 2.14 に示すように、主に投稿者側の視点から、自己管理不可能メディアに対する投稿内容を用いて予測分析を行っている。

Gilbert は、つながりの強さを予測している ([Gilbert 2009], [Gilbert 2012])。ソーシャルメディアにおける 2,000 以上のつながりのデータセットを使用し、85%以上の精度でつながりの強弱を予測し、更に予測できなかったつながりをインタビューで量的調査結果を補っている。また [Gilbert 2012]では、Twitter のつながりから Facebook のつながりを We Meddle<sup>20</sup>という Twitter アプリケーションを用いて、52 カ国、200,000 以上のつながりを予測している。

Backstrom, Yang, Leskove は、リンク予測を行っている ([Backstrom 2011], [Yang 2011], [Leskove 2010])。Backstrom は、Facebook のリンク予測をノードとエッジの特

<sup>20</sup> <http://wemeddle.com/> (2013/10/13)

微量（ノード：性別，年齢，趣味，エッジ：どういう関係か，同じ写真に何度登場したか）を用いて行っている．Yang は，「Yahoo! Plus」において，関心ネットワーク（利用者とサービス間接続）と友好ネットワーク（利用者間接続）に含まれる情報が相関していることを示し，リンク予測を行っている．Leskove は，Epinions，Slashdot，Wikipedia の局所的なリンク情報を特徴量としてエッジのポジティブ/ネガティブを予測している．

Wang は，600 万人のモバイル電話ユーザーの軌跡とコミュニケーション記録を調査することによって，モバイルアクセスログからのコミュニケーション予測を行っている [Wang 2011]．Schifanella は，ソーシャルメディアの友人関係から Last.fm の提案より正確に好む音楽を予測できたと述べている [Schifanella 2010]．

#### (4) 感情・性格に関わる予測に関する研究

感情・性格に関わる予測に関する研究には，[Liu 2008]，[Golbeck 2009]，[Golbeck 2011] などがある．図 2.14 に示すように，主に投稿者側の視点から，自己管理不可能メディアに対する投稿内容を用いて予測分析を試みている．

Liu と Golbeck は，満足度を予測している ([Liu 2008]，[Golbeck 2009])．Liu は，Yahoo Answer の投稿された質問と回答の組み合わせから質問者満足感をモデル化してベストアンサーを予測している [Liu 2008]．Golbeck は，FilmTrust（映画評価のソーシャルメディア）のデータを用いて映画の満足度予測を行っている [Golbeck 2009]．評価の総類似度だけを説明変数に用いるより，プロフィール特性の方がより正確に予測可能だと述べている [Golbeck 2009]．

また Golbeck は，Facebook のプロフィールから性格を予測することを試みている [Golbeck 2011]．従来，正確に個性を計測するためには，性格テストを受ける必要があったが，Facebook のプロフィール上で公的に入手可能な情報により予測することができる と述べている．

#### (5) 影響に関わる予測に関する研究

影響に関わる予測に関する研究には，[Suzuki 2010]，[Hong 2011]，[Szabo 2010]，[Gomez 2010]，[Ginsberg 2008]，[Polgreen 2008]，[Johnson 2004] などがある．図 2.14 に示すように，主に投稿者側の視点から，自己管理不可能メディアに対する投稿内容を用いて予測分析を試みている．また閲覧者側の視点では，主に検索エンジンの出力結果を用

いた予測分析を試みている。

「投稿者—自己管理不可能メディア」領域において Suzuki は、Twitter で地震、台風などの発生地を推測している[Suzuki 2010]。Twitter 利用者を Social Sensors と見なし、その複数の投稿内容が影響を与えているターゲットイベントか否かを判断し、その大きさを計測することで発生場所を推測している。

Hong は、将来の Retweet 数でメッセージの人気度を予測している[Hong 2011]。メッセージ、時制の情報、メッセージのメタデータと利用者情報に基づいて何千もの Retweet を引きつけるメッセージ予測することができると述べている。

Szabo と Gomez は、投稿内容がどの様に影響し広がっていくかを予測している([Szabo 2010], [Gomez 2010])。Szabo は、digg.com と Google の YouTube に投稿されたコンテンツが受け取るトラフィックを予測している[Szabo 2010]。digg.com の投稿票と閲覧数は、夜間や週末よりもトラフィックがピークの時間帯の方が多く、YouTube の投稿映像は 1 カ月後に閲覧数が増える傾向があったと述べている。Gomez は、1 億 7000 万のブログを分析し、トップ 1,000 のマスコミ・サイトによるニュースの拡散が、残りのウェブに情報を普及させると述べている[Gomez 2010]。

「閲覧者—自己管理可能メディア」領域において Ginsberg, Polgreen, Johnson は、閲覧者側からの視点で検索エンジンの出力結果を用いてインフルエンザの流行予測を行っている([Ginsberg 2008], [Polgreen 2008])。Ginsberg は、インフルエンザ流行と関連のある検索クエリ（相関係数の高い上位 50 語）を調査し、アメリカの CDC (Centers for Disease Control and Prevention) 報告との相関係数が、0.97 (min=0.92;max=0.99) という高い値を示している。同様に Polgreen は Yahoo! のクエリを用いて予測を行っている。

## (6) ソーシャルメディアを利用した予測に関する研究 (R-3) の抱える課題

本先行研究領域では、数多くの研究が行われているが、主に投稿者の投稿内容を「セー  
ルスに関わる予測に関する研究」、「政治・経済に関わる予測に関する研究」、「つながりに  
関わる予測に関する研究」、「感情・性格に関わる予測に関する研究」、「影響に関わる予測  
に関する研究」の 5 領域で分析している。閲覧者側の視点でも予測に関する分析は行われ  
ているが、[Ginsberg 2008], [Polgreen 2008] など検索エンジンの出力結果を用いた分析  
を例外として、数百人規模くらいまでのアンケート調査結果を用いた分析が主である。故  
に、作為的な投稿者の投稿内容の影響を受けやすいという“作為的投稿影響問題”や

[Nielsen 2006b]や[Nielsen 2013]で示されているように、アンケート実施時や検索エンジンの出力結果利用時の“利用者バイアス問題”と“調査規模的問題”に対して注意深い扱いを必要とする。

### 2.3.5. ソーシャルメディアを利用した信頼性評価に関する研究 (R-4)

ソーシャルメディアの投稿内容は、必ずしも信頼できるデータだけではないという考えから信頼性評価に関する研究も試みられている。ソーシャルメディアを利用した信頼性評価に関する先行研究は、「レビューに関わる信頼性評価に関する研究」、「スパム判定に関わる信頼性評価に関する研究」、「Web サイトに関わる信頼性評価に関する研究」、「情報に関わる信頼性評価に関する研究」の4つに分類できる。その分類を、先行研究の分類にマッピングしたものが図 2.15 である。本研究領域は、投稿者側の視点と閲覧者側の視点の両視点で、自己管理可能メディアと自己管理不可能メディアの両メディアに対して、ソーシャルメディアを利用した信頼性評価について分析を試みている。閲覧者側の視点では、主にアンケート結果を用いた信頼性評価について分析を行っている。

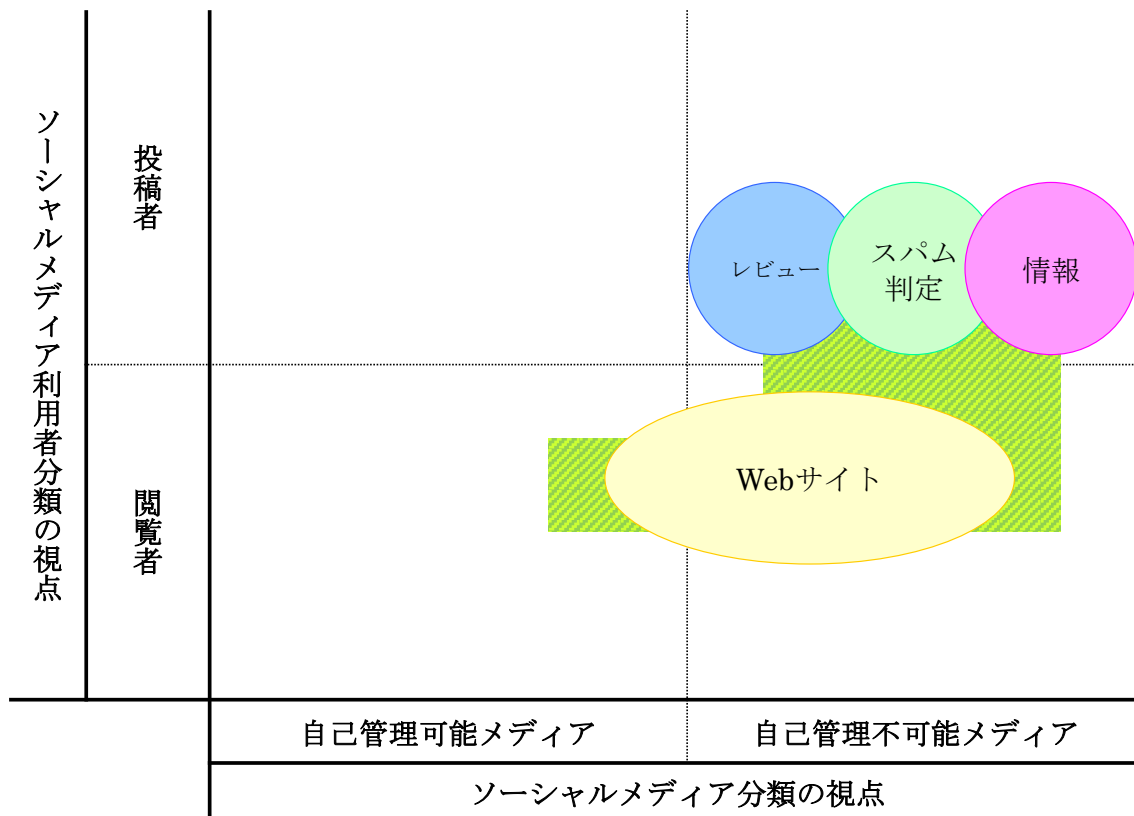
#### (1) レビューに関わる信頼性評価に関する研究

レビューに関わる信頼性評価に関する研究には、[Resnick 2000], [Jindal 2008], [Guo 2011], [Ott 2011]などがある。図 2.15 に示すように、主に投稿者側の視点から、自己管理不可能メディアに対する投稿内容を用いて信頼性評価を行っている。

Resnick, Jindal, Guo は、EC サイトにおける投稿された評価内容に対する信頼性評価を行っている ([Resnick 2000], [Jindal 2008], [Guo 2011])。Resnick は、eBay (1999 年)の商品評価システムを実証分析している。結果は、商品評価に対するフィードバックは半数以上、大部分が肯定的な書き込みであり、購入者と販売者のフィードバックの間に高い相関性があることを示している[Resnick 2000]。Jindal は、Amazon における商品評価内容に対する信頼性評価方法を提案している[Jindal 2008]。評価が高い製品レビューの方がスパムは少なく、評価が低い製品の方がスパムは多かったとしている。同様に Guo は、EC サイトの1つである TAOBAO<sup>21</sup>の商品レビューの信頼性評価を行っている[Guo 2011]。

---

<sup>21</sup> <http://www.taobao.com/> (2013/10/13)



レビュー : [Resnick 2000], [Jindal 2008], [Guo 2011], [Ott 2011]  
 スパム判定 : [Benevenuto 2010], [Lim 2010], [Kolari 2006], [Lin 2007], [Guzella 2009], [Jindal 2010]  
 web サイト : [Fogg 2001] , [Wathen 2002], [Dwyer 2007], [Fogel 2009], [DuBoi 2011]  
 情報 : [Denning 2005], [Adler 2006], [Vuong 2008], [Kittur 2006], [Kittur 2008], [Agichtein 2008], [Suryanto 2009], [Bian 2008], [Mendoza 2010], [Castillo 2011]

図 2.15 ソーシャルメディアを利用した信頼性評価に関する研究 (R-4)

Ott は、ホテルの「やらせレビュー」400件と「本物レビュー」400件を比較し、やらせレビュー中の文言や文脈パターンに着目して分析を試みている[Ott 2011]. 結果、①やら

せレビューは記述内容が大雑把であるが、本物レビューは詳細な記述であること、②やらせレビューは動詞が多いのに対し、本物レビューは名詞が多いこと、という特徴を示している。

## (2) スпам判定に関わる信頼性評価に関する研究

スパム判定に関わる信頼性評価に関する研究には、[Benevenuto 2010], [Lim 2010], [Kolari 2006], [Lin 2007], [Guzella 2009], [Jindal 2010]などがある。図 2.15 に示すように、主に投稿者側の視点から自己管理不可能メディアに対する投稿内容を用いて信頼性評価を行っている。

Benevenuto, Lim, Kolari, Lin, Guzella, Jindal は、投稿内容を分析しスパマーを発見するための信頼性評価を行っている ([Benevenuto 2010], [Lim 2010], [Kolari 2006], [Lin 2007], [Guzella 2009], [Jindal 2010])。

Benevenuto, Lim は、スパムの分類手法を提案している ([Benevenuto 2010], [Lim 2010])。Benevenuto は、Twitter の 5400 万人以上の投稿者、約 19 億のリンク数と約 18 億の投稿数を用いて信頼性評価を行っている [Benevenuto 2010]。スパマーはトレンドな話題と URL を含んだつぶやきを行い、短縮 URL で関係ない Web サイトに誘導し混乱させる傾向があり、提案手法を用いてスパマーの約 70%と非スパマーの約 96%を正しく分類できたと述べている。Lim は、Amazon の商品レビューを分析し、スパマーの分類手法を提案している [Lim 2010]。スパマーは影響を最大にするために、特定の製品グループをターゲットとし、製品評価で他のレビュー者とは異なる傾向を示す特徴があると述べている。

Kolari, Lin, Guzella, Jindal は、スパムの特徴分析を試みている ([Kolari 2006], [Lin 2007], [Guzella 2009], [Jindal 2010])。Kolari は、20 日間の約 1500 万ブログ更新データを調査した結果、約 75%がスパムブログによる更新データであり、それは全英語ブログサイトの約 88%を占めると述べている [Kolari 2006]。また Lin は、スパムブログによる悪影響として(a)情報検索品質の低下、(b)ネットワークとストレージの浪費、が特徴であることを示している [Lin 2007]。同様に、Guzella と Jindal もスパムの特徴分析を試みている ([Guzella 2009], [Jindal 2010])。

## (3) Web サイトに関わる信頼性評価に関する研究

Web サイトに関わる信頼性評価に関する研究には、[Fogg 2001], [Wathen 2002],



[Dwyer 2007], [Fogel 2009], [DuBoi 2011]などがある。図 2.15 に示すように、閲覧者側の視点から、自己管理可能メディアと自己管理不可能メディアの両メディアに対して、主にアンケート結果を用いて信頼性評価を試みている。

「閲覧者－自己管理可能メディア」領域において Fogg と Wathen は、Web サイトの信頼性に影響を与える要素の調査分析を行っている ([Fogg 2001], [Wathen 2002])。Fogg は、1992 年 12 月に 7 つの選択肢、51 項目の Web 調査を行い、Web サイトの信頼性に影響を与える要素を“Real-World Feel”, “Ease of Use”, “Expertise”, “Trustworthiness”, “Tailoring”, “Commercial Implication”, “Amateurism” の 7 つに分類している [Fogg 2001]。Wathen は、「表面的な魅力」や「表面的なデザイン」などの技術的性質が信頼性に影響を与えるとしている [Wathen 2002]。

「閲覧者－自己管理不可能メディア」領域において Dwyer と Figel, DuBoi は、ソーシャルメディア間の信頼性評価を行っている ([Dwyer 2007], [Fogel 2009], [DuBoi 2011])。Dwyer は、Facebook 利用者が Facebook サイトとその利用者に対し、より強い信頼を示し情報共有することに意欲的な傾向を示し、MySpace が新しい人と出会うための Web サイトとして用いている傾向にあることを示している [Dwyer 2007]。Fogel も、大学生 205 人に対し、Facebook と MySpace におけるつながり方の信頼性評価を行っている [Fogel 2009]。結果、女性より男性の方が危険性を持ったつながり方をし、MySpace より Facebook の方が高い信頼度を示したと述べている。DuBoi は、同様に Wikipedia, Slashdot<sup>22</sup>, Epinions<sup>23</sup>のポジティブ/ネガティブの信頼性評価を行い、Wikipedia, Epinion, Slashdot の順に信頼性が高かったと述べている [DuBoi 2011]。

#### (4) 情報に関わる信頼性評価に関する研究

情報に関わる信頼性評価に関する研究には、[Denning 2005], [Adler 2006], [Vuong 2008], [Kittur 2006], [Kittur 2008], [Agichtein 2008], [Suryanto 2009], [Bian 2008], [Mendoza 2010], [Castillo 2011]などがある。図 2.15 に示すように、主に投稿者側の視点から、自己管理不可能メディアに対する投稿内容を用いて信頼性評価を行っている。

Denning, Adler, Vuong, Kittur は、Wikipedia の記載内容に対する信頼性評価を行っている ([Denning 2005], [Adler 2006], [Vuong 2008], [Kittur 2006], [Kittur 2008])。

---

<sup>22</sup> <http://slashdot.org/> (2013/10/13)

<sup>23</sup> <http://www.epinions.com/> (2013/10/13)

Denning は、Wikipedia の記載情報の信頼性評価を行い、①.正確性 (Accuracy), ②動機 (Motives), ③不確実な専門性 (Uncertain Expertise), ④不安定性・変動性 (Volatility), ⑤対象範囲の偏り (Coverage), ⑥出典 (Source) の 6 つのリスクがあると述べている [Denning 2005]. Adler は、Wikipedia の編集履歴に着目し、編集されずに残っていた期間を信頼性の指標にしている [Adler 2006]. Vuong は、Wikipedia 上で発生している信頼性に関する論争を自動的に検出する手法を提案 [Vuong 2008] し、Kittur は、Wikipedia の記載情報の中で矛盾があると考えられる文の可視化と影響評価を行っている ([Kittur 2006], [Kittur 2008]).

Agichtein, Suryanto, Bian は、QA サイトでの質問者と回答者のやり取りに対して信頼性評価を行っている ([Agichtein 2008], [Suryanto 2009], [Bian 2008]). Agichtein は、Q&A サイトでの利用者のやり取りを 2 部グラフで表現し、利用者スコア, hub スコア (多く回答している利用者のスコア) を求め信頼性評価を試みている [Agichtein 2008]. Suryanto は、回答者の専門性の高さを推測し、回答の質の面から信頼性評価を行っている [Suryanto 2009]. また Bian も回答の質の面から信頼性評価を行っており、質問の回答内容, 利用者間のやり取り, 回答に対するフィードバック情報に着目している [Bian 2008].

Mendoza と Castillo は、Twitter への投稿内容に対する信頼性評価を行っている ([Mendoza 2010], [Castillo 2011]). Mendoza は、2010 年のチリ地震における Twitter 利用者の投稿内容について、災害における情報源としてのデマの発信, 拡散についての分類を行っている [Mendoza 2010]. Castillo は、“Twitter の投稿/再投稿(RT)のふるまい”, “投稿内容”, “外部の引用情報”, を用い適合率と再現率が 70~80%の精度で、信頼できる内容と信頼できない内容を自動分類できるとしている [Castillo 2011].

#### (5) ソーシャルメディアを利用した信頼性評価に関する研究 (R-4) の抱える課題

本先行研究領域では、投稿者側の視点と閲覧者側の視点の両視点から「レビューに関わる信頼性評価に関する研究」、「スパム判定に関わる信頼性評価に関する研究」、「Web サイトに関わる信頼性評価に関する研究」、「情報に関わる信頼性評価に関する研究」の 4 領域で分析している。投稿者側の視点では投稿内容を分析しているが、これは本研究領域の目的である投稿内容の信頼性評価のためであるため、作為的な投稿者の投稿内容の影響を受けやすいという“作為的投稿影響問題”は発生しないと考えられる。閲覧者側の視点では、少数の例外を除き、主に数百人規模くらいまでのアンケート調査結果を用いて信頼性評価

を行っている。故に、[Nielsen 2006b]や[Nielsen 2013]で示されているように、アンケート実施時や検索エンジンの出力結果利用時の“利用者バイアス問題”、“調査規模的問題”に対して注意深い扱いを必要とする。

### 2.3.6. ソーシャルメディア利用者の行動分析に残された研究課題

本研究はソーシャルメディア利用者の行動分析であるが、前述に示したようにソーシャルメディアを取り巻く先行研究は数多く存在している。先行研究サーベイの結果、ソーシャルメディア利用者の行動分析に関する研究領域には、以下の特徴があるといえる。

#### < 先行研究の特徴 >

- A) 研究の基本は、投稿者から投稿された情報の分析か、登録制のソーシャルメディアの場合であれば設定された個人のプロパティ情報の分析が主である。以下の3つの研究領域は、投稿者から投稿された情報を分析し、定義・分離や計測・モデル化、予測などを行っている（ソーシャルメディアを利用した信頼性評価に関する研究（R-4）は、投稿された内容の信頼性評価であるためこの特徴からは除外している）。
- ソーシャルメディアを利用した定義・分類に関する研究（R-1）
  - ソーシャルメディアを利用した計測・モデル化に関する研究（R-2）
  - ソーシャルメディアを利用した予測に関する研究（R-3）
- B) 一方で、閲覧者の存在の重要性は認識されており、その行動分析が試みられている。
- C) 閲覧者の分析は、①（登録制の場合のみ）設定された閲覧者の個人プロパティ情報を分析する、②（登録制の場合のみ）「閲覧者数 = 登録者数 - 投稿した人数」、として量的分析を行う、③「投稿が少ない人」を「閲覧者」と定義して分析する、④アンケート調査を行う、⑤検索エンジンの出力結果を使用する、が主流である。以下の3つの研究領域は、主にアンケート調査や検索エンジンの出力結果を用いて閲覧者の分析を行っている。
- ソーシャルメディアを利用した計測・モデル化に関する研究（R-2）
  - ソーシャルメディアを利用した予測に関する研究（R-3）
  - ソーシャルメディアを利用した信頼性評価に関する研究（R-4）

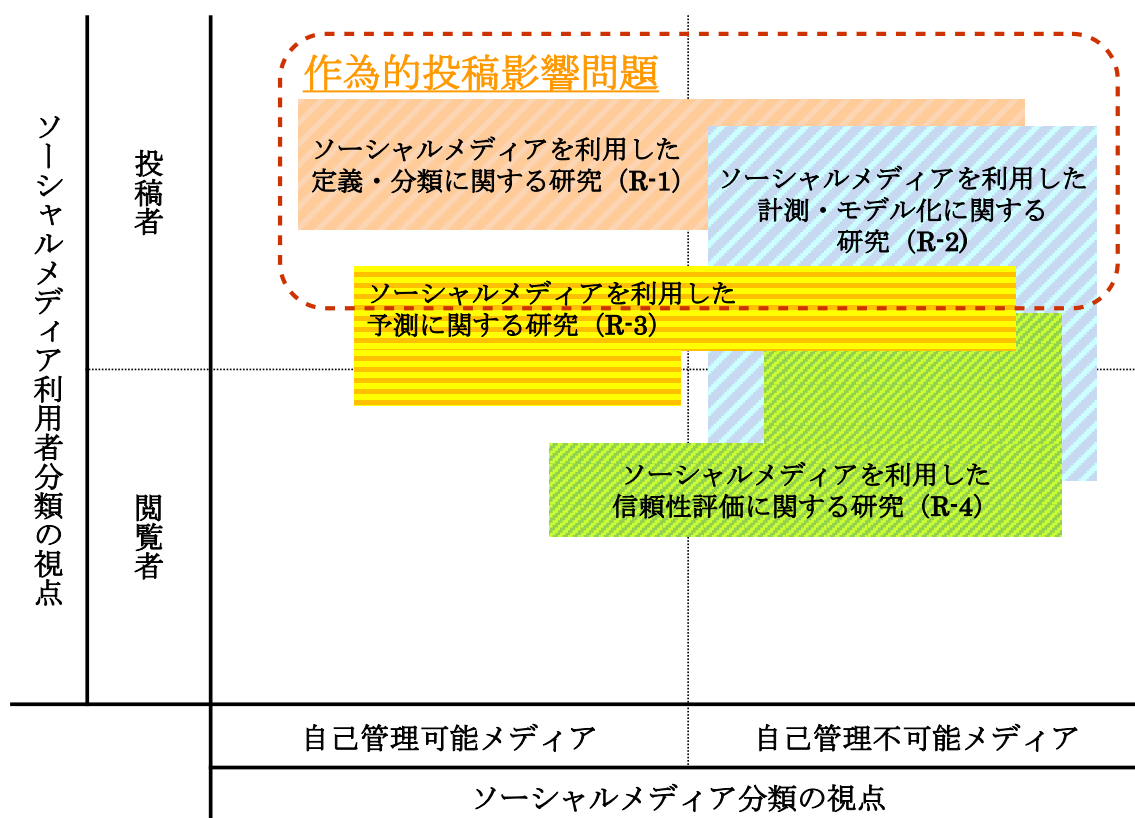


図 2.16 行動分析に残された研究課題（作為的投稿問題）

### (1) 投稿者側からの分析による影響

ソーシャルメディアの利用拡大により、ソーシャルメディアは新たな販売チャンネルとして確立されてきている。同時に、ステルスマーケティングと言われる販売手法も行われるようになり社会的な問題になってきている。たとえば、2011年1月に起きた不正業者による大手口コミグルメサイト「食べログ」への口コミ代行等の“やらせ投稿<sup>24</sup>”，2012年1月に起きた大手QAサイト「Yahoo!知恵袋」への口コミ代行等の“やらせ投稿<sup>25</sup>”，2012年12月に起き刑事事件に発展したオークションサイト「ペニーオークション」での“芸能人やらせブログ投稿<sup>26</sup>”など、社会的事件に発展した問題が数多く発生している。

<sup>24</sup> <http://corporate.kakaku.com/press/release/20120105.html> (2013/10/13)

<sup>25</sup> <http://www.jalux.com/cgi-bin/news/download.cgi?file=0779.pdf> (2013/10/13)

<sup>26</sup>

<http://ja.wikipedia.org/wiki/%E3%83%9A%E3%83%8B%E3%83%BC%E3%82%AA%E3%83%BC%E3%82%AF%E3%82%B7%E3%83%A7%E3%83%B3%E8%A9%90%E6%AC%BA%E4%BA%8B%E4%BB%B6> (2013/12/04)

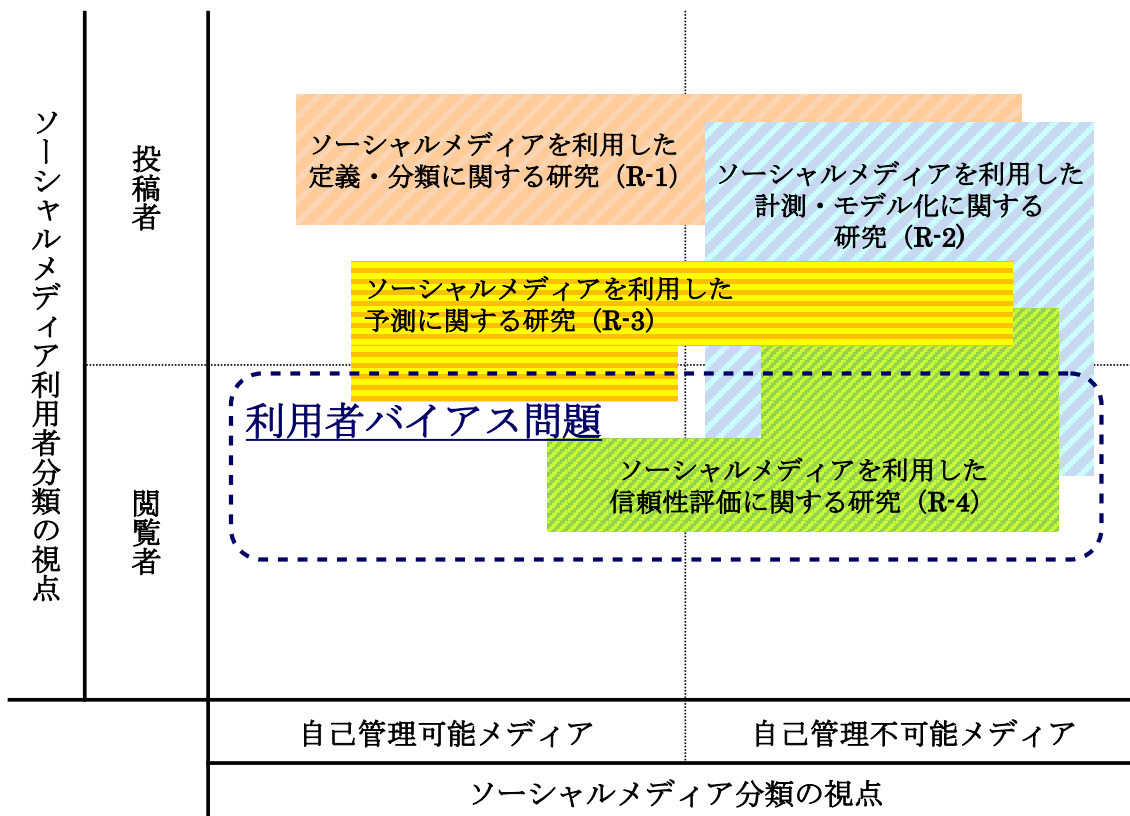


図 2.17 行動分析に残された研究課題 (利用者バイアス問題)

欧米においては消費者保護の観点からステルスマーケティングは違法であるとされ、日本においても消費者庁から「インターネット消費者取引に係る広告表示に関する景品表示法上の問題点及び留意事項」<sup>27</sup>ガイドラインが出されている。

この様な背景において、従来の研究の主流である、ソーシャルメディアへの投稿内容にもとづく投稿者側からのソーシャルメディア利用者の行動分析では、作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”を抱えている。この問題を抱えている主な研究領域は、図 2.16 に示すように、ソーシャルメディアを利用した定義・分類に関する研究 (R-1)、ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)、ソーシャルメディアを利用した予測に関する研究 (R-3) である。この研究領域は課題解決のために、投稿者側からの視点ではなく閲覧者側からの視点でソーシャルメディア利用者の行動分析を行う手法が必要になってくる。

<sup>27</sup> [http://www.caa.go.jp/representation/pdf/120509premiums\\_1.pdf](http://www.caa.go.jp/representation/pdf/120509premiums_1.pdf) (2013/10/13)

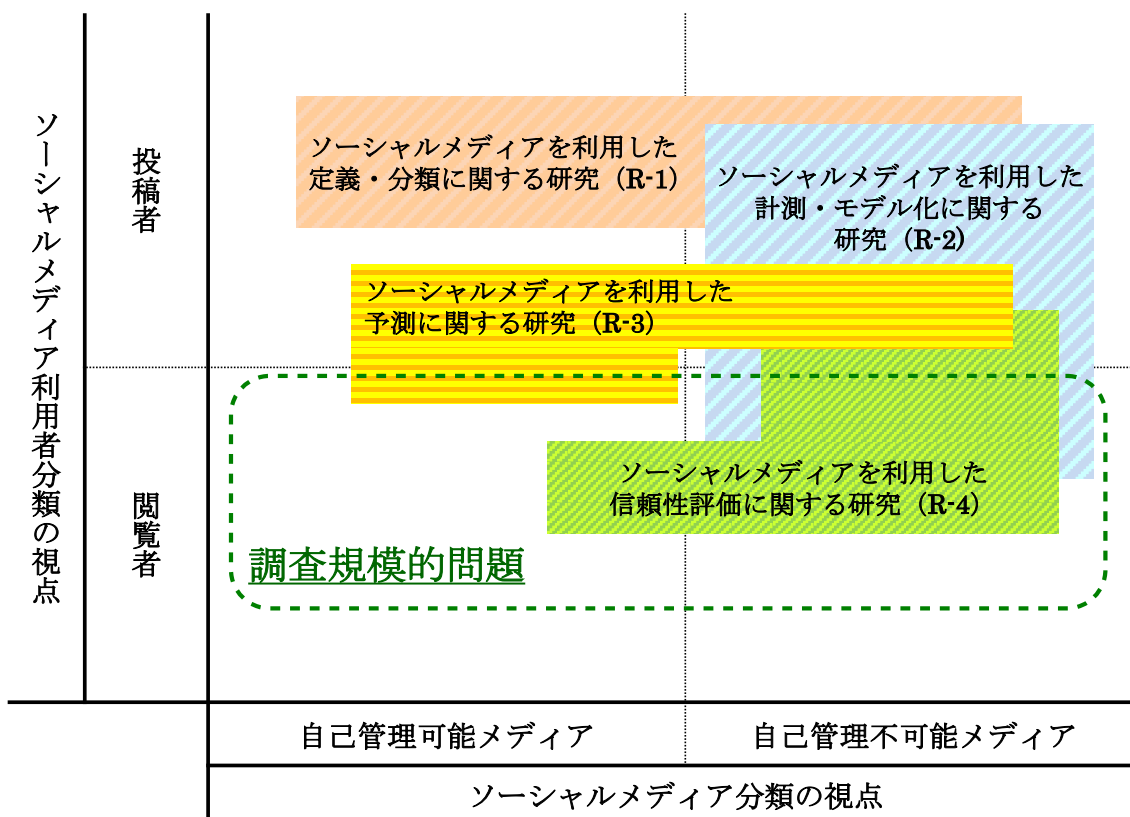


図 2.18 行動分析に残された研究課題（調査規模的問題）

## (2) データ収集方法による影響

前述において示したとおり，従来の先行研究におけるデータ収集方法では，ソーシャルメディア利用者の大部分を占めるとされる閲覧者の分析において，調査票を用いたデータ収集か，検索エンジンの出力結果によるデータ収集が主流である．しかし，調査票を用いたデータ収集では調査票の記入者を選択する時や調査票を記入する時，検索エンジンを用いたデータ収集では検索結果を得る時，に発生し得るサンプリングバイアスや反応バイアスの影響を受けてしまうという“利用者バイアス問題”を抱えている．また，調査票を用いたデータ収集では，ソーシャルメディア利用者に対し，ある一定期間にわたり調査票を記入してもらいデータ収集する必要があるため，大規模な調査には不向きであるという“調査規模的問題”を抱えている．

検索エンジンを使ったデータ収集では，検索エンジン提供者による，独自の結果出力アルゴリズムや，広告効果を狙った SEO (Search Engine Optimization) などの施策で出力結果がコントロールされている．つまり，検索エンジンの出力結果は，利用者が少な

らず検索アルゴリズムの影響を受けてしまうという“利用者バイアス問題”を抱えている。

たとえば[Lampe 2010]は、自らの調査票を用いた研究の限界として、サンプリングバイアスと反応バイアスの影響を受けている可能性について言及している。また Nielsen はサンプリングバイアスと反応バイアスについて、標本数を増やすために報酬を払うことによる影響や、検索エンジンの結果が検索エンジン提供会社のアルゴリズムにより表示され、調査に有用ではない情報を多く含む影響を指摘している[Nielsen 2006b]。[Nielsen 2013]では“調査規模的問題”について、ソーシャルメディア利用者のアクティビティには極端に活発な期間とそうでない期間があり、分析結果に大きな違いが出る実験結果を示しており、データ収集時に調査規模をコントロールする必要性を示している。

これらの問題を抱えている主な研究領域は、図 2.17、図 2.18 に示すように、ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)、ソーシャルメディアを利用した予測に関する研究 (R-3) である。課題解決のためには、データ収集時のバイアスを受けにくい大規模で即時的なデータ収集方法が必要になってくる。

### 2.3.7. 本節のまとめ

先行研究は、投稿者側の視点から投稿内容を分析する場合、作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”を抱えている。また、この問題に対応するために、ソーシャルメディア利用者の行動を閲覧者側の視点から分析する試みもあるが、サンプリングバイアスや反応バイアスの影響を受けてしまうという“利用者バイアス問題”や大規模な調査には不向きであるという“調査規模的問題”を抱えている。これらの先行研究の課題と、先行研究の分類の関係を示したものが図 2.19 である。図 2.19 では、2.3.1 で示した先行研究の分類に対し、2.3.2～2.3.5 で示した先行研究の特徴を踏まえ、3つの問題（“作為的投稿影響問題”，“利用者バイアス問題”，“調査規模的問題”）がどの研究領域に関わっていて、本研究の対応範囲がどこになるのかを示している。本研究は、図 2.19 に示すように3つの問題すべての影響を抱える研究領域である「ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)」, 「ソーシャルメディアを利用した予測に関する研究 (R-3)」に対応する手法を提案する。

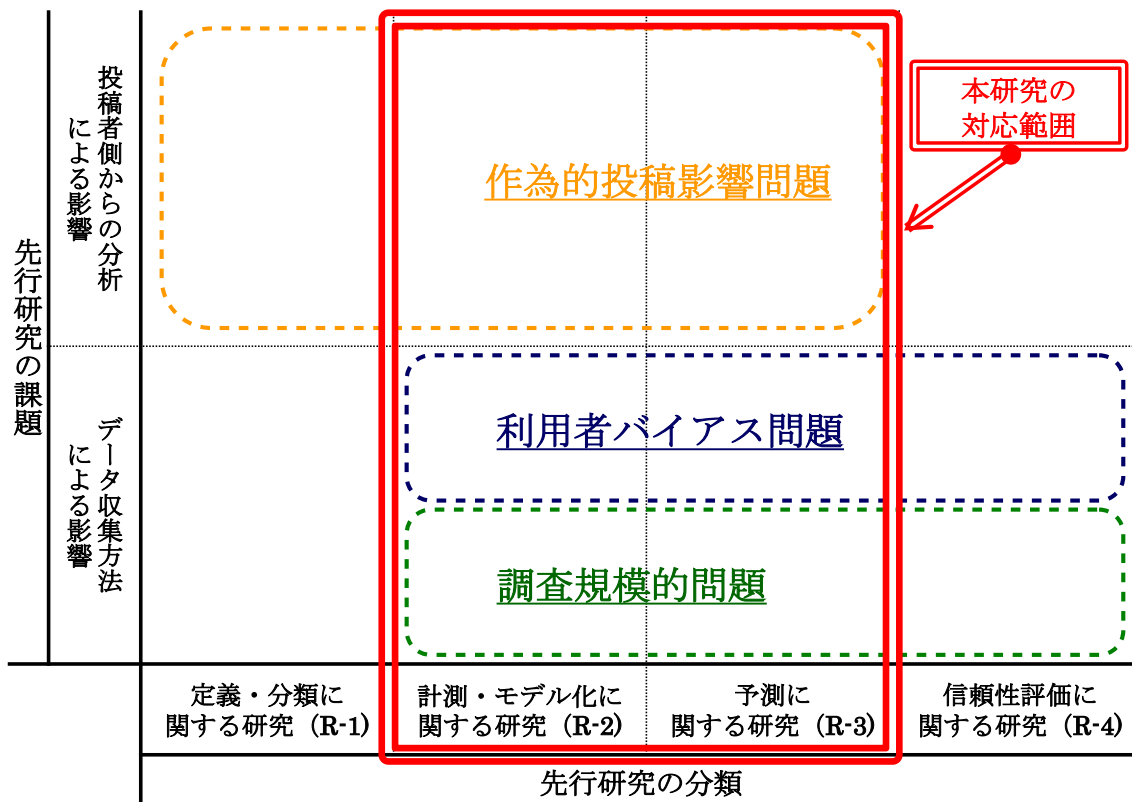


図 2.19 本研究で対応する先行研究の課題

## 2.4. ソーシャルメディア利用者の行動分析に関する 本研究の貢献

2.3 節で示した研究課題を踏まえ、本研究では DNS クエリデータにもとづく、投稿者側からではなく閲覧者側からソーシャルメディア利用者の行動分析を試みる。DNS はインターネットの基盤技術の 1 つであり、インターネット利用者が意識することなく、ソーシャルメディア閲覧を含むインターネット上でのサービス利用時には、即時的に使用される機能の 1 つである。特に本研究では、各国ごとに割り当てられた ccTLD（国別 Top-Level-Domain）<sup>28</sup>の 1 つで、日本に割り当てられた JP ドメイン名の DNS クエリデータを用いて大規模で即時的なデータ収集を行う。これらの対応により、2.3 節で述べた“作為的投稿影響問題”や“利用者バイアス問題”，“調査規模的問題”の解決を試みる。

<sup>28</sup> <https://www.nic.ad.jp/ja/dom/types.html> (2013/10/13)



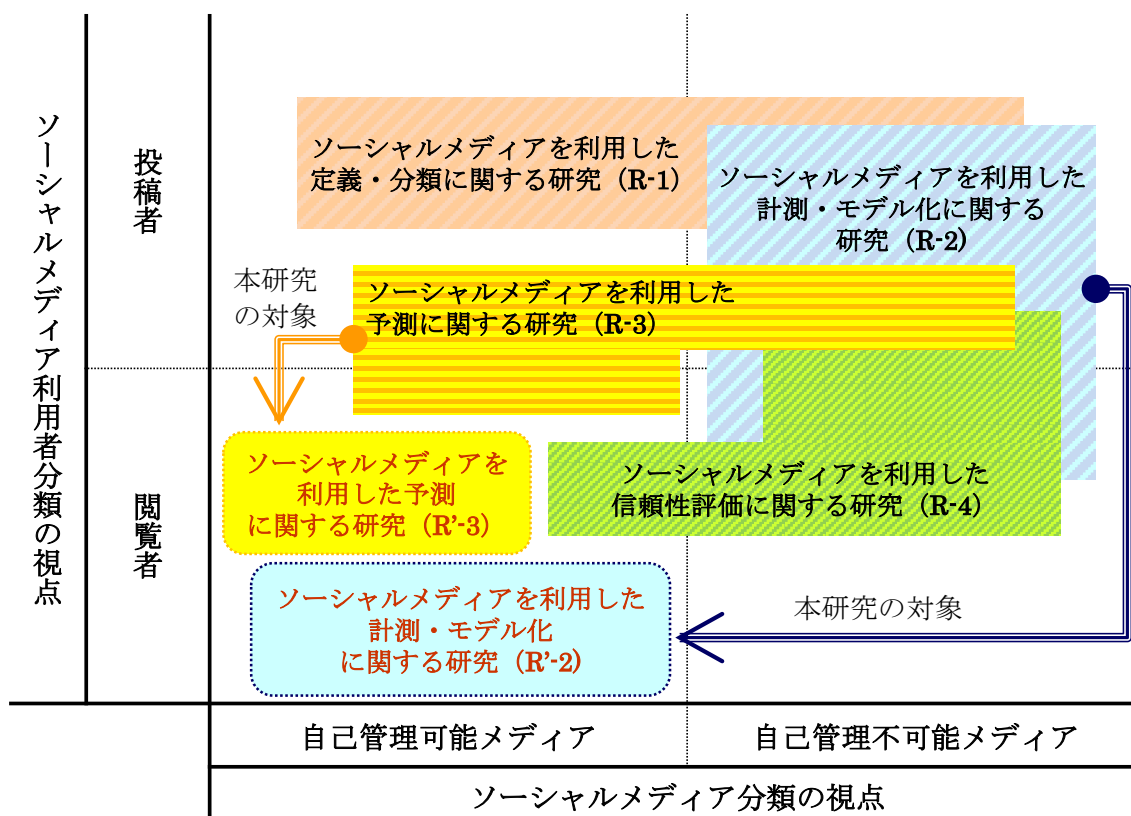


図 2.20 本研究の分析対象範囲

図 2.20 に本研究における分析対象範囲を示す。本研究では、3つの問題（“作画的投稿影響問題”，“利用者バイアス問題”，“調査規模の問題”）すべての影響を抱える「ソーシャルメディアを利用した計測・モデル化に関する研究 (R-2)」, 「ソーシャルメディアを利用した予測に関する研究 (R-3)」の2つの研究領域を分析対象とする。前者については、第4章で、提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、ソーシャルメディアを利用した予測に関する研究領域（図 2.20 中の「ソーシャルメディアを利用した計測・モデル化に関する研究 (R'-2)」）において、提案手法の有効性を確認するために日本映画の観客動員数予測を試みる。後者については、第5章で提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、ソーシャルメディアを利用した計測・モデル化に関する研究領域（図 2.20 中の「ソーシャルメディアを利用した予測に関する研究 (R'-3)」）において、提案手法が大規模で複雑な

行動分析も可能とすることを確認する。具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における2つの大きな社会現象に対し、DNSクエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用との関連を計測することで、ソーシャルメディア利用者の行動の時間的推移の分析を試みる。

# 第3章 閲覧者視点からのソーシャルメディア 利用者の行動分析手法

前章で述べたとおり、従来、ソーシャルメディア利用者の行動分析を試みる研究では、投稿された情報（投稿内容、登録設定情報）やアンケート結果、または検索エンジンの出力結果から得られたデータにもとづき、ソーシャルメディア利用者の行動を分析したり、実社会で起きる現象を予測したりする投稿者側の視点からの研究が主流であった。ゆえに、①作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”，②ソーシャルメディア利用者が、調査票の記入者を選択する時や調査票を記入する時、または検索エンジンを利用して検索結果を得る時などに発生し得る、サンプリングバイアスや反応バイアスの影響を受けてしまうという“利用者バイアス問題”，③ソーシャルメディア利用者に調査票を記入させるため、大規模な調査には不向きであるという“調査規模的問題”を抱えている。

本稿では、インターネットの基幹技術の1つであるDNSのクエリデータにおける“異なり数”を用いて、ソーシャルメディア利用者の“投稿者側”の視点からではなく、“閲覧者側”の視点から、実社会で起きる現象を分析する手法を提案する。特に本研究は、DNSの中でも各国毎に割り当てられたccTLD<sup>29</sup>の1つであるJPドメイン名のDNSクエリログを用いることで、大規模で即時的なデータ<sup>30</sup>収集を特色とする。

DNSクエリの結果は一意性があるため、検索する人によって結果が変わる様なことはなく、検索エンジンを利用して検索結果を得る際に発生し得る“利用者バイアス問題”は発生しない。また、ソーシャルメディアへのアクセス時などで用いられるインターネットの基幹技術であるDNSクエリログを用いて調査を行うため、ソーシャルメディア利用者に意識させることなくデータ収集が可能である。しかし、本提案手法は、解決しなければならない“DNSキャッシュの影響”を抱えている。

---

<sup>29</sup> 各国/地域に割り当てられたTop Level Domainで255存在している(2012年3月時点)。ccTLDは、ISO3166で規定されている2文字の国コードを原則として使用している。

<sup>30</sup> 2013年12月時点でのJPドメイン名のDNSクエリ数は、約16億クエリ/日

以下では、まず本提案手法におけるシステム構成と手法の詳細を示した後、提案手法が抱える“DNS キャッシュの影響”について述べ、その改善方法について示す<sup>31</sup>。

### 3.1. システム構成

図 3.1 に、一般的な DNS クエリの仕組みと、本研究におけるソーシャルメディア利用者の行動を計測するシステムの構成を示す。DNS は、階層構造を持つ世界中に存在する多数のサーバが協調しあって動作する分散型のデータベースである。この事例では、まず Web の閲覧者によるブラウザの操作により、閲覧者が契約する ISP (Internet Service Provider) の DNS サーバに閲覧先 Web の URL である `www.example.jp` と IP アドレスの変換が要求される。これに対し ISP の DNS サーバでは、階層的に配置された分散データベースの最上位階層にあるルートサーバ (たとえば `m.root-servers.net`) から、1 つ下の階層である `.jp` のアドレス情報を持つ JP-DNS サーバ (たとえば `a.dns.jp`) の IP アドレスを取得し、次にその JP-DNS サーバ (`a.dns.jp`) から `example.jp` のアドレス情報を持つ DNS サーバ (たとえば `dns.example.ad.jp`) の IP アドレスを取得する。最後にその DNS サーバ (`dns.example.ad.jp`) から閲覧先 Web の URL (`www.example.jp`) の IP アドレスを取得し、最終的にブラウザが動作する PC にその IP アドレス (`www.example.jp` の IP アドレス) を通知する。

このように再帰的な検索を行う URL と IP アドレスの変換手続きは、「再帰検索」または「再帰問い合わせ」と呼ばれる。また、検索業務を行いドメイン名と IP アドレスの対応結果をキャッシュする DNS サーバ (図 3.1 では、ISP の DNS) は、「DNS キャッシュサーバ」、または「フルリゾルバ」と呼ばれる。これに対しドメイン名と IP アドレスの対応付けを管理する DNS (図 3.1 では `dns.example.ad.jp`) は、「DNS コンテンツサーバ」と呼ばれる。

本研究で利用する JP-DNS は、インターネットドメイン名の 1 つである JP ドメイン名のための DNS である。JP ドメイン名は、2013/12/01 現在、総合計 1,353,969 の登録数があり、国内における主要な企業や政府機関、地方公共団体、ソーシャルメディアなどの

---

<sup>31</sup> 本章の内容は[Mitamura 2013]に報告した内容にもとづく。

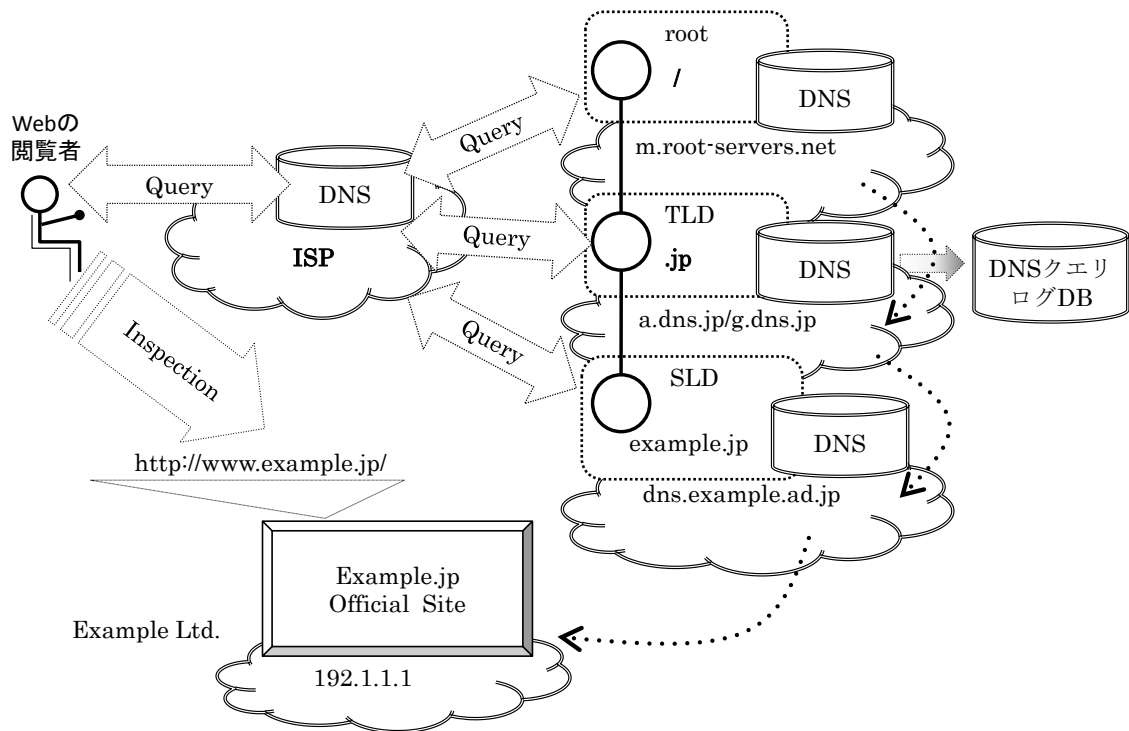


図 3.1 システム構成

ネットワークサービスなどに使われている。つまり、JP-DNS は、日本におけるインターネットトラフィックの基幹システムとして多くのサービスに関わっている。本研究における分析データの取得は、その JP-DNS (統計情報を公開している代表的な JP-DNS サーバである図 3.1 の a.dns.jp/g.dns.jp) で行った。a.dns.jp<sup>32</sup>は、IPv4/IPv6 アドレスで IP anycast 技術を用い東京/大阪の複数拠点で運用<sup>33</sup>されている。g.dns.jp は、IPv4 アドレスで unicast 技術を用い東京拠点で運用されている。

a.dns.jp, g.dns.jp の各々で取得されたログデータは、1つのログデータとして統合した。統合したログデータは、1日1回のバッチ処理で、図 3.1 に示すソーシャルメディア利用者の行動分析のための研究環境の1つである DNS クエリログデータ DB に転送・蓄積した後、行動分析に使用した。DNS クエリログ DB から行動分析を行うデータ処理フローについては、3.4 節で述べる。

<sup>32</sup> <http://jprs.jp/tech/jp-dns-info/a-dns-jp-ja.html> (2013/10/13)

<sup>33</sup> <http://jprs.jp/tech/jp-dns-info/2008-10-06-jp-dns-servers.html> (2013/10/11)

## 3.2. 延べ数と異なり数

本研究では、ソーシャルメディア利用者の行動を分析するために、DNS クエリデータとして延べ数と異なり数という 2 種類の変数を用いる。延べ数は、Web サーバへのアクセス回数（累積）といった一般的な指数であるが、異なり数は、たとえば Web サーバへアクセスする IP アドレスのユニーク数（種類数）である。本研究の重要な特徴である「異なり数」については、以下、3.2.1, 3.2.2 で説明する。

### 3.2.1. 異なり数を用いた解析技術

異なり数を用いた研究は、主にネットワーク管理の技術領域で、解析技術への応用、計測手法に関する研究、集計高速化に関する研究などが行われている。ここでは、本研究に関わる解析技術への応用分野での先行研究について述べる。

ネットワーク管理の技術領域における異なり数を用いた解析技術への応用では、通常時と異常時の変化検出に異なり数を用いられることが主である。1つは通常時と異常時をトラフィックの流れの変化から検出しようと試みる研究であり、もう1つは通常時と異常時をホストの振る舞いの変化から検出しようと試みる研究である。前者の研究には[Kim 2004], [Peng 2004], [Lakhina 2005], [Shomura 2008]があり、後者の研究には[Keys 2005], [Mori 2005], [Zhao 2005]がある。

[Kim 2004], [Peng 2004]は、DDOS (Distributed Denial of Service) 攻撃が行われたような異常時の特徴から異なり数を用いてホストの検出を行う提案をしている。DDOS 攻撃の場合、多くのホストを踏み台にして攻撃先のホストへパケットを送るため、攻撃対象となるホストでは送信元 IP アドレスの異なり数（種類数）が大きくなる、という特徴から異常トラフィックの検出を試みる研究である。[Lakhina 2005], [Shomura 2008]も同様に IP アドレスやポート番号のエントロピーを計算することで異常トラフィックの検出を試みている。

[Keys 2005], [Mori 2005], [Zhao 2005]は、攻撃を起こすために脆弱ホストを探そうとしているホストの振る舞いから、異常ホストの検出を試みる研究である。これらの研究は、脆弱ホストを探す振る舞いとして多くの異なるホスト先への接続を試みる必要があるため、送信先 IP アドレスの異なり数が多くなる傾向にある、という特徴を用いて異常ホストの検出を試みている。

本研究では、単位時間におけるソーシャルメディア利用者の送信元アドレスに関する異なり数の大きさは、その単位時間に起きたソーシャルメディア利用者の特徴的な行動の現れであると捉え、その異なり数を用いてソーシャルメディア利用者の行動分析を試みる。

### 3.2.2. 本研究における延べ数と異なり数

本研究における変数である延べ数と異なり数の概要を図 3.2 に示す。本研究における延べ数は、図 3.2 において当該 Web サイト (example.jp) への DNS クエリが何回発生したかという累計を示す数である。また異なり数は、当該 Web サイト (example.jp) への DNS クエリが、何カ所から発生したかという種類を示す数である。本研究における異なり数では、Web サイトへのクエリ送出元 IP アドレスと Web サイトごとのクエリ送出元の国数の 2 種類の変数を用いる。前者は何台の機器から DNS クエリが発生したかを示し、後者は何カ国から DNS クエリが発生したかを示す。

延べ数と異なり数の具体的な事例について、国際的な賞を数多く受賞した日本映画の公式サイトへの DNS クエリデータを用いて、その特徴を説明する。図 3.2 に示す 3 つの変数 (Web サイトごとのクエリ数、Web サイトへのクエリ送出元の IP アドレス数、Web サイトごとのクエリ送出元の国数) を時系列グラフとしたものが図 3.3 である。図 3.3 のクエリ数が図 3.2 の延べ数であり、図 3.3 の IP アドレス数、国数が図 3.2 に示す異なり数である。図 3.3 の縦軸は変数値 (左軸 : Web サイトごとのクエリ数 / Web サイトへのクエリ送出元の IP アドレス数、右軸 : Web サイトごとのクエリ送出元の国数)、横軸は当該日本映画の公式サイトへの DNS クエリが初観測された日から公開前日までの日数である。

図 3.3 には、当該日本映画に関係するイベント (完成披露記者会見、大手メーカーとキャンペーン開始、国際映画賞のグランプリ受賞) を示したが、3 つの変数すべてにおいて、イベント発生時に大きく増加する傾向が見られる。また、Web サイトごとのクエリ数 (延べ数) と Web サイトへのクエリ送出元の IP アドレス数 (異なり数) は、類似した時系列の推移傾向を示し、映画公開前 100 日前後 (図 3.3 中の横軸 211 辺り) からいくつかのイベント発生と共に徐々に増加している。また、この 2 つの変数は映画公開前 50 日程度 (図 3.3 中の横軸 247 辺り) まで同じ傾向で推移するが、それ以降はクエリ数 (延べ数) の方が強い増加傾向を示している。それに対し Web サイトごとのクエリ送出元の国数 (異なり数) は、初クエリ観測 (図 3.3 中の横軸 37 辺り) からすぐに立ち上がり、イベント発生以外にも上下変動を繰り返しながら増加傾向を示している。

分類	延べ数	異なり数	
概念図			
説明	Webサイトごとのクエリ数	Webサイトへのクエリ送出元のIPアドレス数	Webサイトごとのクエリ送出元の国数

図 3.2 延べ数と異なり数

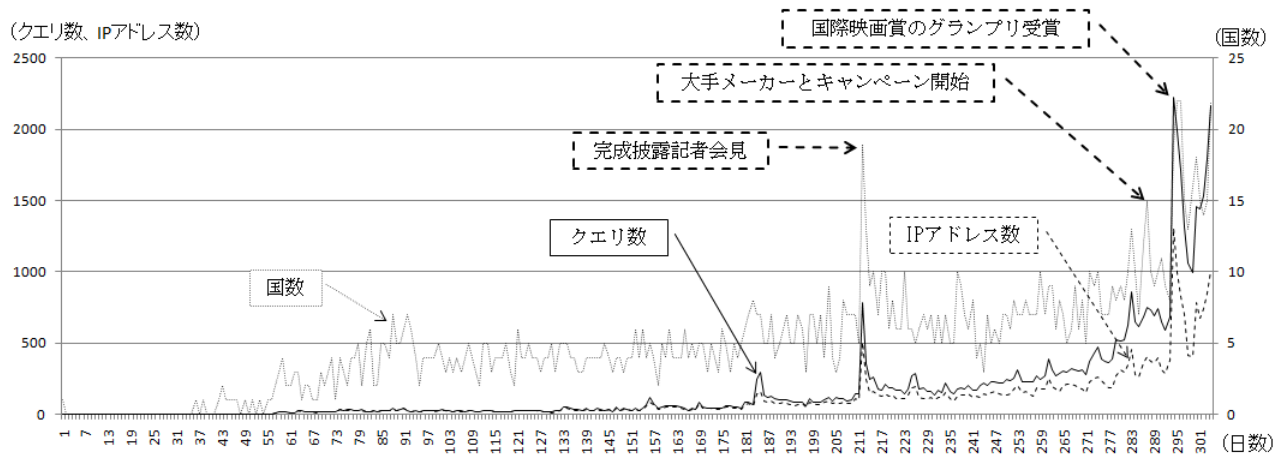


図 3.3 延べ数，異なり数の実際の時系列グラフ



### 3.3. DNS キャッシュの影響と異なり数

DNS はキャッシュという機能を持っている。これは、ドメイン名単位に決められた TTL<sup>34</sup>という時間の間、検索されたドメイン名と IP アドレスの対応データを保持し、TTL 満了までの間に、再度同じドメイン名に対する検索を受けた場合に、保持しているドメイン名と IP アドレスの対応データを用いて代理応答する機能である。

この機能はキャッシュサーバで提供され、図 3.1 の例では ISP の DNS サーバ<sup>35</sup>で実行される。3.1 節で示したように、本提案手法におけるデータ取得は JP-DNS サーバ (図 3.1 の a.dns.jp/g.dns.jp) で行い、そこへのクエリは ISP の DNS サーバなどキャッシュサーバから送信される。そのため、ブラウザからの DNS クエリに対し、キャッシュサーバによる代理応答が行われる可能性があり、JP-DNS サーバでは必ずしもすべてのクエリを受信できているとは言えない。つまり、JP-DNS サーバにおける単純な DNS クエリの延べ数の計測では、キャッシュによる影響を受け、正しくソーシャルメディア利用者の行動を計測できない可能性がある。しかし、JP-DNS サーバへの DNS クエリ送出元の IP アドレス数や国数などの異なり数は、種類の数 (異なるドメイン名から導かれる IP アドレスの種類数や国の種類数) であるため、同じドメイン名の検索で機能する DNS キャッシュの影響を受けにくいと言える。

上述の背景を踏まえ提案手法では、当該 Web サーバに何回アクセスされたかを表す延べ数より、何か所からアクセスされたかを表す異なり数の方が、DNS キャッシュの影響を受けることなくソーシャルメディア利用者の行動を計測できると考え、単純な DNS クエリの延べ数ではなく DNS クエリから抽出される異なり数を用いて、ソーシャルメディア利用者の行動を分析する。この延べ数と異なり数を用いたソーシャルメディア利用者の行動分析に対する影響については第 4 章において示し、本分析における異なり数を用いることの妥当性を検証する。

---

<sup>34</sup> ドメイン名の階層構造は、ゾーンに分割され分散管理されている。これらの分割されたゾーンには、ドメイン名の運用方法を記述した SOA (Start Of Authority) と呼ばれるリソースレコードがあり、そこで TTL (Time To Live) を指定する。

<sup>35</sup> DNS サーバには、管理しているドメイン名の情報を有する権威サーバ (Authoritative Server)、権威サーバへの反復問い合わせや問い合わせた内容を一時保存するキャッシュサーバ (Cache Server)、PC 等の検索プログラムであるスタブリゾルバ (Stub Resolver) がある。

## 3.4. 閲覧者側視点からの行動分析

本研究では, DNS クエリデータの異なり数を用いて閲覧者側の視点からソーシャルメディア利用者の行動分析を試みる. 図 3.6 に, そのデータ処理フローを示す.

### ■Step1 : Web サイトの登録

ソーシャルメディア利用者の行動分析を行う対象との関係が一意に特定できる Web サイト (たとえば, SNS サイトや製品サイトなど) を選択し, 対象 Web サイト DB に登録する.

### ■Step2 : DNS クエリデータ処理

DNS クエリログ(本研究では図 3.1 に示す `a.dns.jp/g.dns.jp` から転送されたログデータ) を格納するデータベースに対し, 対象 Web サイト DB に登録された Web サイトのドメイン名について検索を行う. 検索結果から参照元のアドレス情報(ドメイン名, IP アドレス, AS 番号など) を取り出し, そのアドレスが登録/管理された国名/地域名との対応変換 DB を検索する.

### ■Step3 : 異なり数算出

Step2 の結果を踏まえ, 3.2 節で示した異なり数を時系列データとして計測し, 異なり数 DB に登録する.

### ■Step4 : 分析対象データの実社会データ抽出

Step1 で登録した Web サイトと, 関連があると考えられる分析対象の実社会における観測データ (たとえば, 商品の売上高, 個人や組織・団体などの活動記録など) を対象実社会データ DB に登録する.

### ■Step5 : 行動分析モデルの構築

Step3, Step4 の結果を用いて, 主成分分析, および回帰分析を行い行動分析モデルの算出式を立てる. 主成分分析における主成分は Step3 の結果から抽出し, 回帰分析における目的変数は Step4, 説明変数は Step3 の結果, または主成分分析で抽出した主成分を用いる. 主成分分析結果は, 主成分 DB に登録する.

### ■Step6 : 利用者行動分析

Step5 で抽出した主成分の適合性を確認し, ソーシャルメディア利用者の行動を明らかにするために主成分間の関係性を分析する.

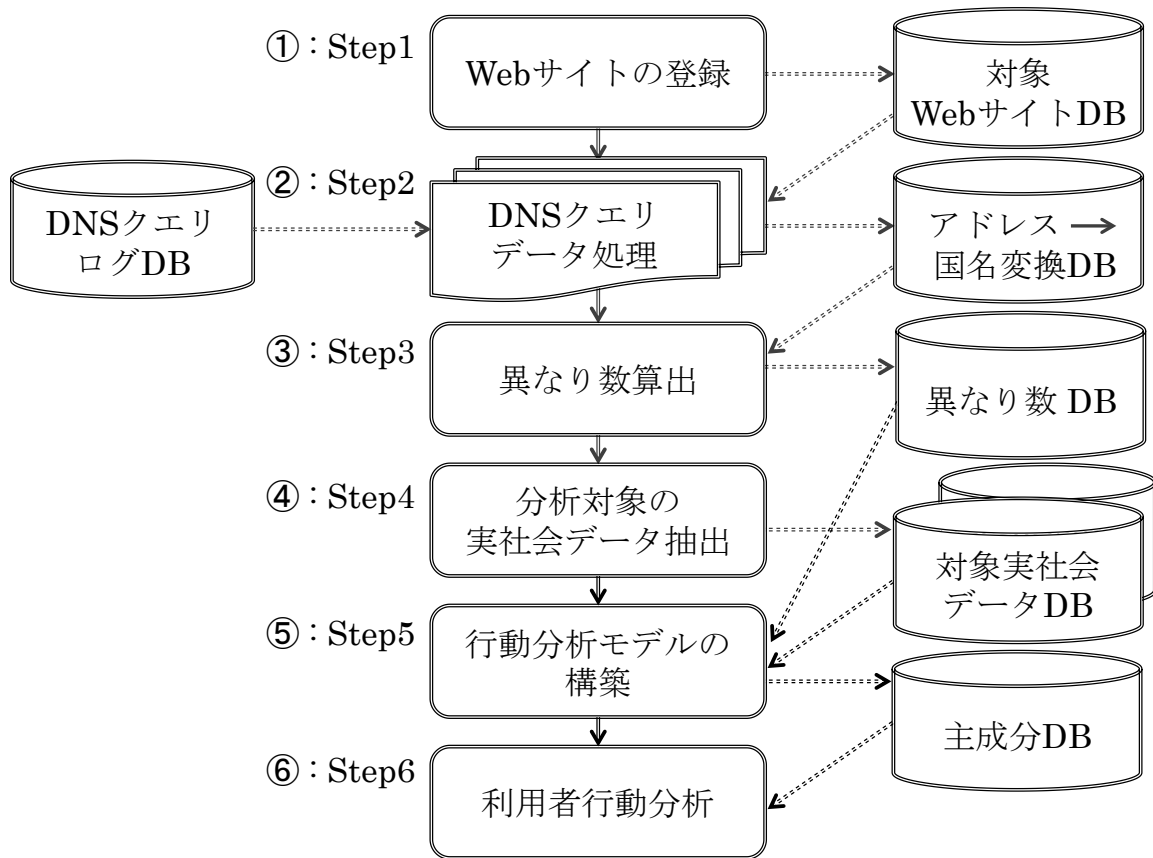


図 3.6 データ処理フロー

### 3.5. 提案手法の検証

本章では、提案手法の基本的な設計が、

- 閲覧者側の視点からソーシャルメディア利用者の行動分析が可能（“作為的投稿影響問題”への対応）
  - データ収集時のバイアスを受けにくいソーシャルメディア利用者の行動分析が可能（“利用者バイアス問題”への対応）
  - 大規模なソーシャルメディア利用者の行動分析が可能（“調査規模的問題”への対応）
- であることを示した。

次章からは、提案手法を用いて具体的な実証検証を行う。第4章では、提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、ソーシャルメディアを利用した予測に関する研究領域において、提案手法の有効性を確認するために日本映画の観客動員数予測を試みる。第5章では提案手法を応用し、より大規模で複雑な実社会における行動分析を行うことで、ソーシャルメディアを利用した計測・モデル化に関する研究領域において、提案手法が大規模で複雑な行動分析にも有効であることを確認し、その利用者の行動の時間的推移を明らかにする。具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における2つの大きな社会現象に対し、DNSクエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用の関連を計測することで、ソーシャルメディア利用者の行動の時間的推移の分析を試みる。

本提案手法は、ソーシャルメディア利用者の行動を把握する上で客観的なデータ計測を可能とし、実社会で起きる現象を分析する手法を提案するものである。

## 第4章 ソーシャルメディアを利用した予測に関する研究

前章で，“作為的投稿影響問題”，“利用者バイアス問題”，“調査規模的問題”に対応できるソーシャルメディア利用者の行動分析手法を提案した。本章では，提案手法をソーシャルメディア利用者の実社会における行動分析に適用し，ソーシャルメディアを利用した予測に関する研究領域において提案手法の有効性を確認するために，映画の公式サイトへの閲覧を示す DNS クエリログから，日本映画の観客動員数の予測を試みる。分析対象である日本映画は，2007年に閣議決定された「文化芸術の振興に関する基本的な方針（第2次基本方針）<sup>36</sup>」や「国立メディア芸術総合センター（仮称）構想<sup>37</sup>」で取り組むべき重要コンテンツの1つとして取り上げられており，その予測研究の重要性も高まっている。

具体的なソーシャルメディア利用者の行動分析対象として選択した映画の興行成績予測研究は，主に米国においてマーケティング的な観点から実施されている。たとえば，当初の潜在顧客規模( $N_0$ )，そのうち実際に観客となる割合( $P_0$ )，公開前の口コミ効果( $\sigma$ )という3つの変数から興行収入のモデル化を試みる研究[Castro 2006]，インターネット掲示板への投稿量と評論家のレビュー数が映画興行の初期興行成績に対して与える影響の研究[Liu 2006]，映画の興行収入の時系列推移において，映画ジャンルや公開時期などの変数に加え，ブログでの映画に対する評価コメントが興行収入に及ぼす影響の研究[Mishne 2006]などがある。[Mishne 2006]は，ブログへの投稿内容と興行収益の相関が，公開前で0.454～0.542，公開後で0.478～0.614程度であることを示している。

第2章で示したように，行動分析に関する先行研究は，“作為的投稿影響問題”や“利用者バイアス問題”を抱えており，[Mishne 2006]もこの問題を抱えている。そこで本章では，先行研究が抱える問題に対応する提案手法を用いて映画の観客動員数予測を行い，提案手法の有効性を確認する<sup>38</sup>。

<sup>36</sup> [http://www.bunka.go.jp/bunka\\_gyousei/housin/kihon\\_housin\\_2ji.html](http://www.bunka.go.jp/bunka_gyousei/housin/kihon_housin_2ji.html) (2013/10/16)

<sup>37</sup> [http://www.bunka.go.jp/oshirase\\_other/2009/mediageijutsu\\_090514.html](http://www.bunka.go.jp/oshirase_other/2009/mediageijutsu_090514.html) (2013/9/1)

<sup>38</sup> 本章の内容は[三田村 2010]に報告した内容にもとづく。

## 4.1. 日本映画の観客動員数予測分析

表 4.1 に観客動員数予測分析対象を示す。ソーシャルメディアを利用した予測分析において、映画公開初週とした理由は次の 2 つである。

- (a) 映画の公開初週の観客動員数（興行成績）を決定できれば、高い精度で全興行収入を予測できる[Sawhney 1996]ことや、全興行成績の 25%は公開初週の 2 週間の成績が占める[Litman 1998]など、興行成績における公開初週の重要性が示されているため。
- (b) 映画公開前では主な情報源は映画の公式サイトとなるが、映画公開後になるとその情報源は広がるのが推測され、興行成績に影響を与える要素が多くなりノイズが増加すると考えられるため。

表 4.1 分析対象

映画	
作品	.jp ドメイン名の公式サイトを持つ日本映画
公開時期	2008 年 6 月～2009 年 11 月
作品数	50 作品
観客動員数	公開初週の動員数
DNS のクエリデータ	
分析期間	2007 年 7 月～2009 年 11 月
データ取得	JP-DNS のクエリ (a.dns.jp, g.dns.jp)

## 4.2. 分析の流れとデータ抽出範囲

3.4 節で示した閲覧者側視点からの行動分析提案手法による分析を以下の流れで行った。

### ■ Step1 : Web サイトの登録

2008 年 6 月から 2009 年 11 月の間に公開された、.jp ドメイン名の公式サイトを持つ日本映画の Web サイトを対象 Web サイト DB に登録する (50 作品が対象)。

### ■ Step2 : DNS クエリデータ処理

Step1 で登録した Web サイトのドメイン名を DNS クエリログ DB から検索し、クエリ数の計算と参照元のアドレス情報 (ドメイン名, IP アドレス) を取り出し, IP アドレスから国名に変換する DB を用いて IP アドレスと国名の変換を行う。

### ■ Step3 : 異なり数算出

Step2 の結果から, IP アドレスと国名の日次, 月次の異なり数を計算し, 異なり数 DB に登録する。

### ■ Step4 : 分析対象データの実社会データ抽出

Step1 で登録した日本映画の公開初週の観客動員数を, 対象実社会データ DB に登録する。

### ■ Step5 : 行動分析モデルの構築

#### ✓ 分析 1 : 単独変数による説明力分析

観客動員数とクエリ数, IP アドレス数, 国数の 3 変数を用いて単回帰分析を行う。ここでは, 延べ数, 異なり数の説明力と DNS キャッシュによる影響について分析を行う。

#### ✓ 分析 2 : 変数の組み合わせによる重回帰分析

DNS キャッシュの影響を考慮し, IP アドレス数, 国数の組み合わせで重回帰分析を行う。

#### ✓ 分析 3 : 回帰式の説明力の改善

分析 2 で実施した重回帰分析結果において, 実務への適用を考慮し事例固有の説明力改善を行う。

本分析におけるデータ抽出範囲を表 4.2 に示す。

### 4.3. 計測データの説明力分析

まず、クエリ数、IP アドレス数、国数の各々の説明力を分析するために、各変数を用いた単回帰分析を行った。単回帰分析の結果を表 4.3 に示す。

相関係数は、IP アドレス数 > 国数 > クエリ数となった。このことから、延べ数（クエリ数）よりも異なり数（IP アドレス数と国数）の方が高い説明力を示すことがわかった。更に、表 4.4 に示すように、クエリ数と IP アドレス数は相関係数が高く、重回帰分析時に多重共線性を起こす可能性がある。

また、観客動員数の実測と予測の差（残差）がマイナス（表 4.5 で▲と表示）となった作品数は、延べ数よりも異なり数の方が多く（クエリ数：28 作品、IP アドレス：26 作品、国数：24 作品）、残差の分散も大きい傾向となった。

これらの結果から、次のことが明らかとなった。

(a) 映画の公開初週の観客動員数は、異なり数と正の相関がある。

Web サイトへの DNS クエリが何回あったかという延べ数と、どこからあったかという異なり数を比較した場合、観客動員数の増加に結び付く情報は、その映画の情報が公開前にどの程度広がったのか、という影響範囲が重要になる傾向が見られる。

(b) IP アドレス数とクエリ数は、相関係数が高い。

(c) 異なり数は延べ数よりも説明力が高い。この事は、DNS キャッシュの影響を考慮した提案手法の妥当性を示唆している。

以上を踏まえ観客動員数の予測式の変数には、延べ数ではなく異なり数を用いる。



表 4.2 抽出範囲

クエリ数	クエリ初観測日から映画公開前日までに発生したクエリ累計数
IP アドレス数	クエリ初観測日から映画公開前日までの各日にアクセスしてきた IP アドレスの種類
国数	クエリ初観測日から映画公開前日までの各日にアクセスしてきた国の種類

表 4.3 単回帰分析結果

	延べ数	異なり数	
	クエリ数	IP アドレス数	国数
相関係数	0.644	0.776	0.695
標準誤差	139843.3	115335.9	131508.6
平均	87607.4	45320.0	83.3

表 4.4 変数間の相関係数

	クエリ数	IP アドレス数	国数
クエリ数	1	-	-
IP アドレス数	0.763	1	-
国数	0.570	0.535	1

表 4.5 単回帰分析結果（作品別観客動員数の残差比較）

映画作品	残差（クエリ数）	残差（IP アドレス数）	残差（国数）
movie01	372,372.6	271,067.6	161,690.4
movie02	201,119.4	153,967.5	99,625.2
movie03	166,641.6	165,954.5	74,988.1
movie04	▲ 185,543.4	▲ 295,470.7	▲ 61,184.0
movie05	46,164.6	33,086.1	103,179.0
movie06	112,003.6	91,172.8	▲ 54,705.1
movie07	▲ 56,985.3	▲ 41,864.9	10,611.6
movie08	37,695.6	11,490.6	▲ 42,783.3
movie09	21,883.6	▲ 7,765.9	50,751.6
movie10	23,882.2	25,452.1	▲ 43,682.5
movie11	6,862.0	18,784.8	▲ 104,417.3
movie12	▲ 55,475.2	▲ 84,527.4	▲ 50,569.7
movie13	▲ 22,863.3	257.2	27,896.2
movie14	▲ 144,616.0	▲ 133,837.0	▲ 1,365.7
movie15	▲ 233,732.0	▲ 269,307.8	▲ 84,971.9
movie16	▲ 32,846.3	▲ 22,824.0	9,684.9
movie17	25,988.2	23,977.2	129,186.6
movie18	65,944.8	81,678.6	106,180.6
movie19	▲ 90,474.0	▲ 83,808.1	27,007.3
movie20	▲ 68,453.1	▲ 61,621.8	5,608.6
movie21	473.7	▲ 14,976.6	14,115.4
movie22	▲ 15,163.2	▲ 13,396.0	▲ 48,942.5
movie23	▲ 120,510.4	▲ 179,318.6	▲ 78,421.7
movie24	▲ 46,138.5	▲ 40,438.6	▲ 82,400.9
movie25	▲ 58,388.6	▲ 118,765.2	▲ 178,102.0

映画作品	残差 (クエリ数)	残差 (IP アドレス数)	残差 (国数)
movie26	▲ 36,771.6	53,023.9	39,692.1
movie27	250,630.7	▲ 29,874.7	188,957.7
movie28	185,312.9	3,085.7	10,268.9
movie29	▲ 81,360.1	63,945.3	59,498.2
movie30	488,742.0	361,777.5	521,606.5
movie31	125,510.6	16,877.5	212,267.5
movie32	24,308.0	▲ 83,971.9	▲ 92,774.8
movie33	▲ 151,204.2	▲ 20,878.1	▲ 24,538.6
movie34	▲ 59,847.1	▲ 159,235.9	137,453.9
movie35	17,651.0	▲ 866.9	7,175.1
movie36	▲ 197,774.5	2,380.1	16,539.8
movie37	▲ 107,498.9	▲ 21,328.2	▲ 184,779.3
movie38	▲ 114,170.8	▲ 43,143.0	51,432.3
movie39	175,887.9	219,893.0	▲ 39,391.5
movie40	20,806.3	1,441.2	42,155.4
movie41	▲ 57,609.5	▲ 38,539.7	▲ 209,442.5
movie42	45,622.4	60,655.6	▲ 70,285.6
movie43	▲ 64,349.8	▲ 14,357.5	▲ 82,780.4
movie44	19,738.8	▲ 27,004.5	137,028.7
movie45	▲ 25,908.7	54,710.7	▲ 41,450.2
movie46	▲ 146,285.4	116,918.9	▲ 160,124.4
movie47	▲ 9,494.1	1,052.7	▲ 265,996.6
movie48	▲ 207,129.6	▲ 55,208.5	▲ 255,852.2
movie49	▲ 8,626.0	36,266.4	41,736.0
movie50	▲ 36,023.1	▲ 6,586.4	▲ 27,375.1

## 4.4. 重回帰分析の結果

4.3 節の結果を踏まえ、説明変数に国数と IP アドレス数（いずれも異なり数）を用いた重回帰分析を行った。結果を図 4.1, 図 4.2, 表 4.6 に示す。

まず「国数+IP アドレス数」の相関係数と自由度調整済み決定係数であるが、図 4.1, 図 4.2 に示すように「クエリ数」, 「IP アドレス数」, 「国数」の単回帰結果よりも共に大きく改善している。また表 4.6 に示すように、説明変数である国数と IP アドレス数の回帰式における各々の影響力は、IP アドレス数 (t 値 : 6.106) の方が国数 (t 値 : 4.222) よりも大きいという結果になった。

予測式は次のとおりである。

$$y = 3457.257x_1 + 3.807515x_2 - 256791$$

ここで、

$y$  : 日本映画の公開初週の観客動員数

$x_1$  : 国数 (異なり数 : 公開前日時点)

$x_2$  : IP アドレス数 (異なり数 : 公開前日時点)

参考までに「国数+IP アドレス数」に「クエリ数」を変数として加えた場合の回帰分析結果も図 4.1, 図 4.2, 表 4.7 に示す。図 4.1 に示すように、相関係数は同程度 (両者とも 0.844) の数値となっているが、自由度調整済み決定係数の数値は、図 4.2 に示すように 0.699 から 0.693 へと下がっており説明力は落ちている。また、表 4.7 に示すようにクエリ数の t 値は小さく多重共線性の疑いが強い。この事もソーシャルメディアを利用した予測分析において、延べ数よりも異なり数を用いた提案手法の有用性を示している。

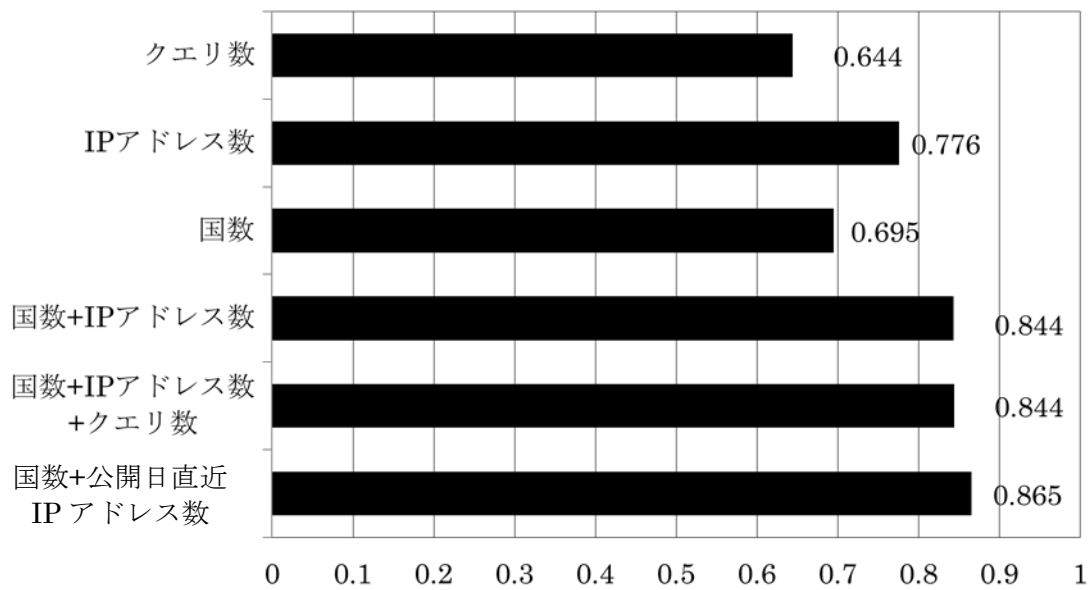


図 4.1 相関係数比較

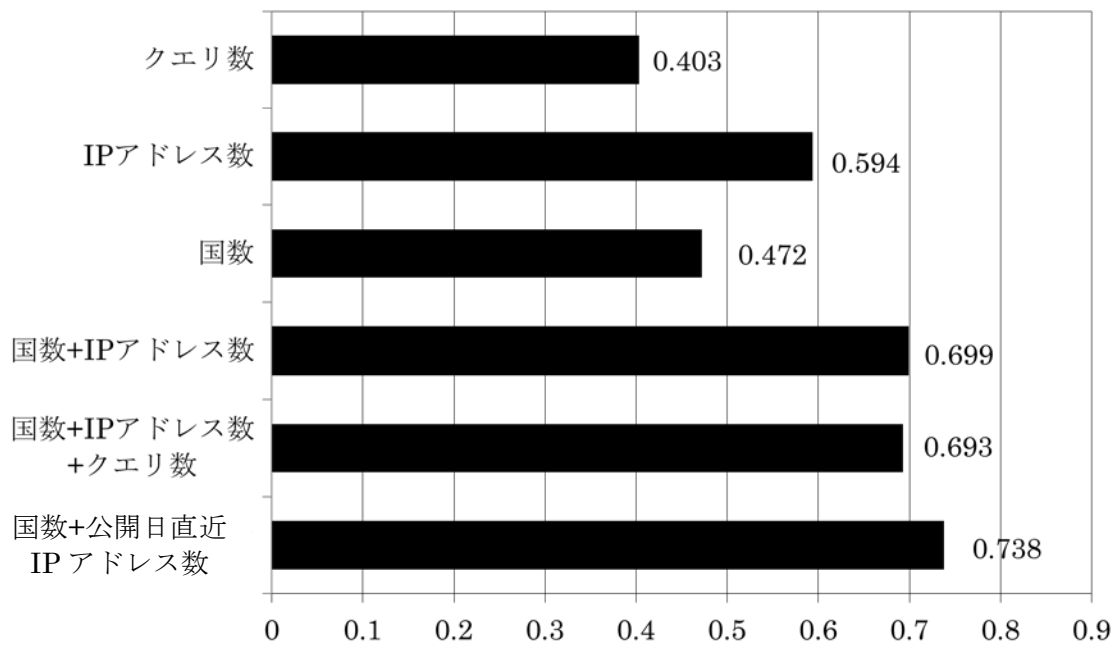


図 4.2 自由度調整済み決定係数比較

表 4.6 回帰係数と t 値

		国数+IP アドレス数		国数+公開日直近期間 IP アドレス	
		回帰係数	t 値	回帰係数	t 値
説明変数	切片	-256791	-4.231	-248298	-4.438
	国数	3457.257	4.222	3120.225	4.031
	IP アドレス数	3.808	6.106	-	-
	公開日直近期間 IP アドレス数	-	-	5.298	7.053

表 4.7 回帰係数と t 値 (変数 : 延べ数追加)

	回帰係数	t 値
切片	-259047	-4.232
国数	3517.432	4.061
IP アドレス数	3.937	4.695
クエリ数	-0.099	-0.234

## 4.5. 予測力改善の試み

本節では実務への適用を考慮し、映画興行予測を行う場合における、予測力の改善を試みる。改善にあたっては、単回帰分析で説明力のもっとも高かった変数である IP アドレス数に着目した。

たとえば図 4.3 のグラフは、ある映画の IP アドレス数の時系列データである。縦軸は IP アドレス数、横軸は DNS クエリの初観測日から映画公開前日までの日数である。この時系列データの推移をみると、公開日に近づくと IP アドレス数は増加しているが、傾向には段階的な特徴が見られる。公開日 5 ヶ月以上前では、ほとんどデータ観測が見られず、5 ヶ月前から 3 ヶ月前の間でいったん緩やかな増加となり、2 ヶ月前から 1 ヶ月前で急激な増加傾向に変わる。

これらの傾向を踏まえ本改善の試みでは、IP アドレス数のデータ抽出において、「公式サイトに対し、より映画公開日に近く、より多くの国からアクセスがある日は、ソーシャルメディア利用における予測分析に有効な情報が、より多く集まっている」と考えて、「より映画公開日に近く、国数が平均よりも多い日が連続している期間」の IP アドレス数の累計（以下、公開日直近期間 IP アドレス数）を計測することで予測力の改善を試みた。

公開日直近期間 IP アドレス数の抽出領域例を図 4.4 に示す。図 4.4 の左縦軸は IP アドレス数、右縦軸は国数、横軸は初クエリ観測からの日数である。右縦軸 4.8 の横線が国数の平均値を示し、それを超える公開日により近い連続した期間の公開日直近期間 IP アドレス数（図中右端の両矢印で示した期間）が、改善を試みるデータ抽出対象領域である。

このような方針を踏まえ、IP アドレス異なり数のデータ抽出を全分析対象作品に対して行い、「国数+公開日直近期間 IP アドレス数」での回帰分析を行った。この結果も図 4.1、図 4.2、表 4.6 に示す。

まず相関係数と自由度調整済み決定係数の変化は、図 4.1、図 4.2 に示すようにそれぞれ 0.844 から 0.865 へ、0.699 から 0.738 へと共に約 3%改善した。また改善された回帰式において表 4.6 に示す t 値を見ると、国数の t 値は 4.222 から 4.031 に若干減少したが、「公開日直近期間 IP アドレス数」の t 値 (7.053) は「IP アドレス数」の t 値 (6.106) から大きく伸びており、「公式サイトに対し、より映画公開日に近く、より多くの国からアクセスがある日は、ソーシャルメディア利用における予測分析に有効な情報が、より多く集まっている」という考えは、誤りではなかったことを示唆している。

(IP アドレス数)

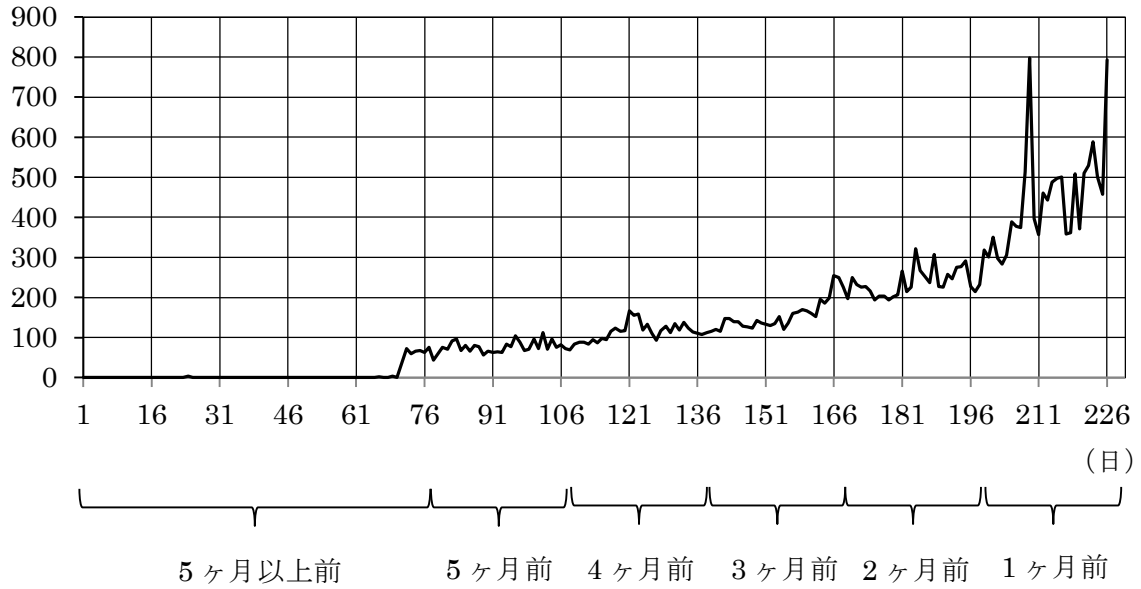


図 4.3 IP アドレス数の時系列データ例

(IP アドレス数)

(国数)

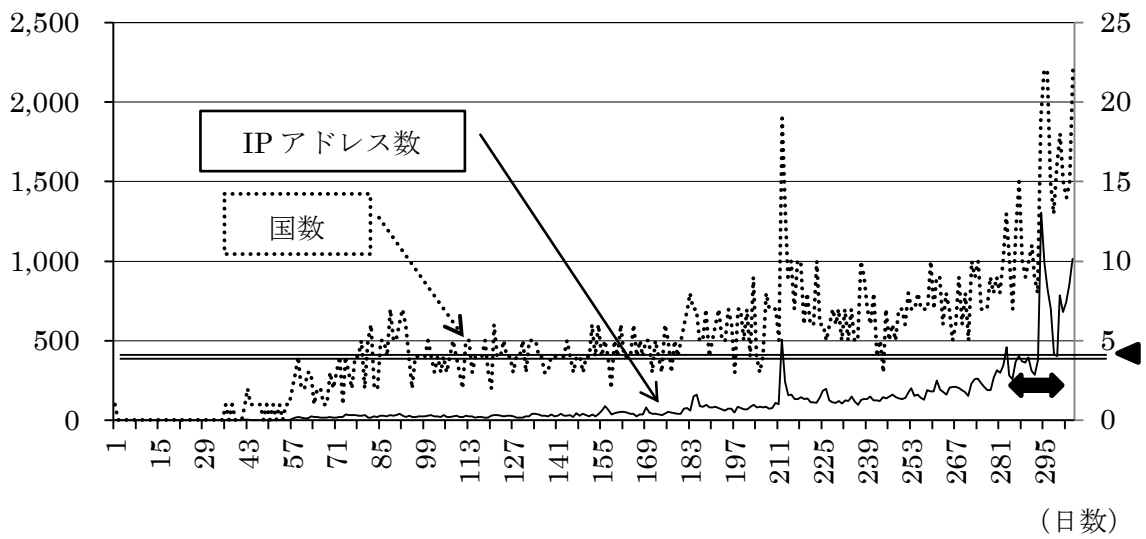


図 4.4 データ抽出領域



改善を行った場合の予測式は次のとおりである。

$$y = 3120.225x_1 + 5.297912x_2 - 248298$$

ここで、

$y$  : 日本映画の公開初週の観客動員数

$x_1$  : 国数 (異なり数 : 公開前日時点)

$x_2$  : 公開日直近期間 IP アドレス数

(ただし、国数が平均値を超えた連続期間の累計)

## 4.6. 実務への適用可能性

前節まで、提案手法を用いて映画公開初週の観客動員数の予測を行い、提案手法の有効性を検証した。投稿者側からの視点による従来手法[Mishne 2006]の予測力が、相関係数で 0.478~0.614 であるのに対して、閲覧者側からの視点による提案手法では 0.865 と同等の結果を得ることができた。

本節では、前節までに示した日本映画の観客動員数予測式の実務への適用可能性について考察する。

本研究の実務適用可能性を探るために、実務家へのインタビューを行った。映画配給会社によっても差異は出ると考えられるが、公開前の主な決定事項と時期は次のような目安であるとのことであった。

- 公開 2 週間程度前まで : シネマコンプレックスの席数決定
- 公開 1 ヶ月程度前まで : メディアへの広告枠決定
- 公開 3 ヶ月程度前まで : 上映館 (施設) の決定

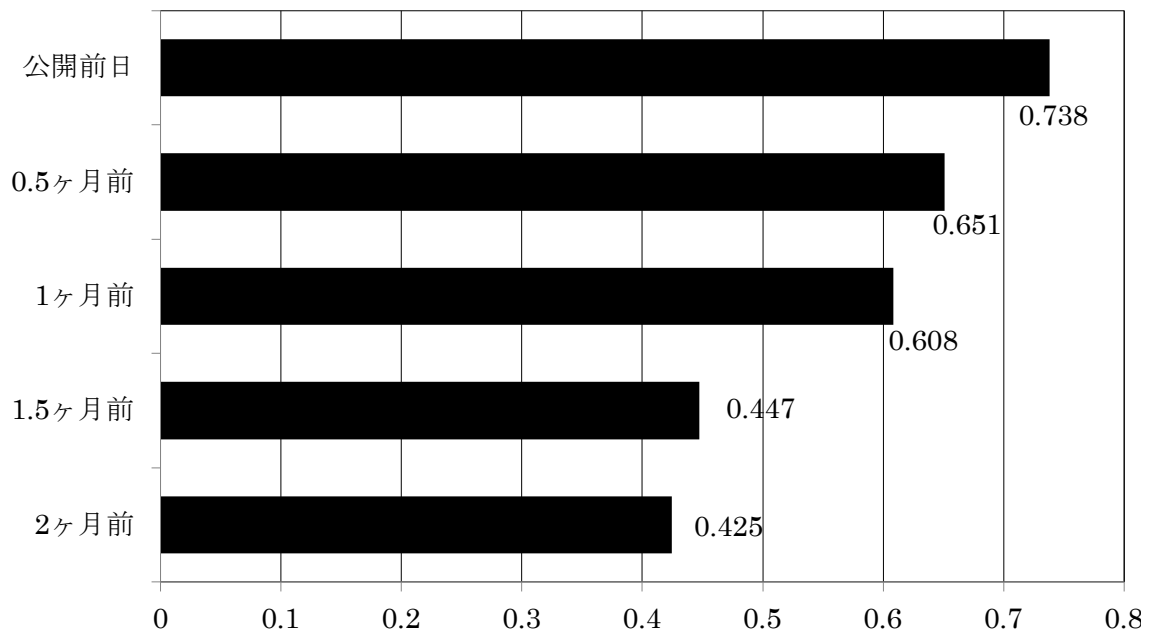


図 4.5 自由度調整済み決定係数の推移

このことから、実務では最低でも映画公開の2週間前、可能であれば1ヶ月から3ヶ月くらい前の時点での観客動員数の予測を求められることがわかった。

実務への適用可能性を検証するため、データ計測時期を映画公開前日から2週間単位で2ヶ月前までさかのぼり、4.5節で示した予測式の予測力がどのように推移するのかを分析した。今回分析対象とした映画作品は、初クエリの発生から映画公開前までの期間において3ヶ月に満たない作品があったため、分析は2ヶ月前までとした。自由度調整済み決定係数の推移を図4.5に、観客動員数予測力の推移を図4.6に示す。

まず、自由度調整済み決定係数の推移であるが、図4.5に示すように映画公開の1ヶ月前で0.608と高い説明力を示す結果となった。観客動員数予測力の推移については、図4.6に示す予測と実測が一致する実線を軸に分布を見ると、0.5ヶ月前の段階で観客動員数の大きな作品から外れ始め、映画公開前日からさかのぼるほど、実測値よりも予測値が低くなる傾向を示す結果となった。図4.5、図4.6から全体的な推移傾向は、公開前日から1ヶ月前までは緩やかに下がり、1ヶ月前と1.5ヶ月前の間で落ち込みが大きくなり、その後再度緩やかに下がることがわかった。

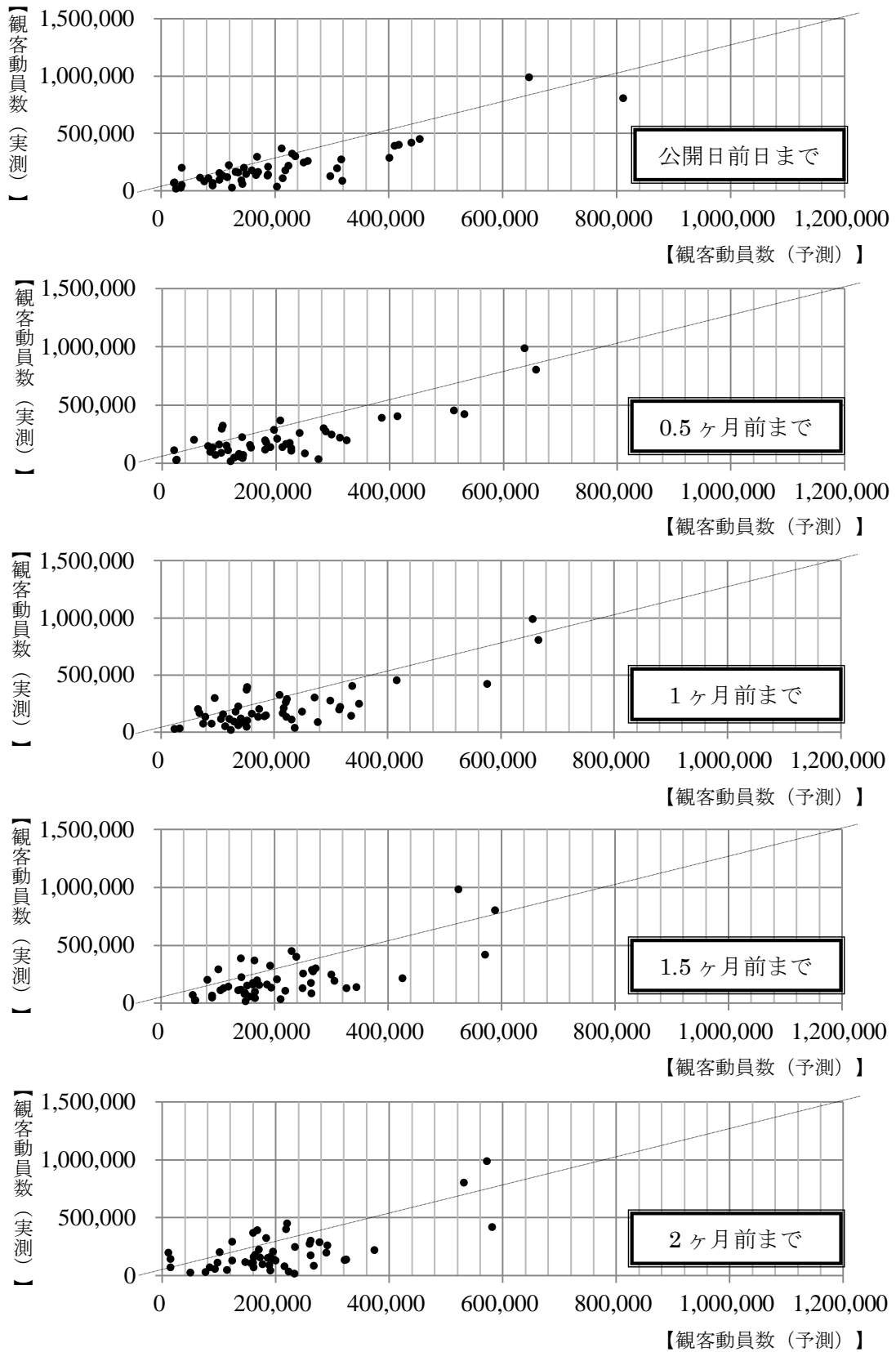


図 4.6 観客動員数予測力の推移

表 4.8 基本統計量（観客動員数）

観客動員数	
平均	203,616.08
標準誤差	25,594.02
中央値	156,934.50
最小	17,480
最大	987,384
標本数	50

次に、もう少し詳細に観客動員数の実測と予測の差異について分析する。分析対象とした映画の観客動員数に関する基本統計量を表 4.8 に示す。

この基本統計量を踏まえ、観客動員数の標準誤差である 2.5 万人単位、および平均値と標準誤差の比率である 12.5%単位の差異を基本として A から F までの 6 つに分類した。

- A：予測と実測の差異が 2.5 万人未満，または 12.5%未満
- B：予測と実測の差異が 2.5 万人以上 5 万人未満，または 25%未満
- C：予測と実測の差異が 5 万人以上 7.5 万人未満，または 37.5%未満
- D：予測と実測の差異が 7.5 万人以上 10 万人未満，または 50%未満
- E：予測と実測の差異が 10 万人以上 12.5 万人未満，または 67.5%未満
- F：予測と実測の差異が 12.5 万人以上，または 67.5%以上

この 6 分類における期間毎の作品数と全体に対する占有率（分類内の作品数/全作品数）が、どのように推移するのかについて示したものが表 4.8 である。この結果から本研究における提案手法では、映画公開の 1 ヶ月前において観客動員数の予実の差異が 7.5 万人未満または 37.5%未満（表 4.9 の A+B+C）に収まるものが 64%で、実務にも有用な精度を持つことがわかった。

表 4.9 作品数と占有率の推移

国数+公開日直近期間 IP アドレス数	作品数					
	A	B	C	D	E	F
公開前日	20	11	7	2	4	6
0.5ヶ月前	16	4	14	6	3	7
1ヶ月前	10	11	11	6	4	8
1.5ヶ月前	15	6	8	2	6	13
2ヶ月前	11	9	6	5	1	18
国数+公開日直近期間 IP アドレス数	占有率					
	A	A+B	A+B+C			
公開前日	0.40	0.62	0.76			
0.5ヶ月前	0.32	0.40	0.68			
1ヶ月前	0.20	0.42	0.64			
1.5ヶ月前	0.30	0.42	0.58			
2ヶ月前	0.22	0.40	0.52			

## 4.7. TTL 値の影響考察

本節では、4.6 節で示した予測力に対し、3.3 節で示した DNS キャッシュの影響と関わりが深い TTL 値が、予測力にどの程度の影響を与えているかを考察する。

まず、本研究で取り上げた 50 作品の映画公式サイトにおける、TTL 値別の作品数割合を図 4.7 に示す。TTL 値は、120 秒～259200 秒まで広い範囲で分布している。

図 4.8 は、観客動員数が同程度で、TTL 値が異なる映画公式サイトにおいて、同一 IP アドレスからの DNS クエリ数を調べた事例である。縦軸は DNS クエリ数であり、横軸は

DNS クエリの初観測日から映画公開前日までクエリ数を 1 日単位で計測し、DNS クエリ数の多い日順に並べたものである。TTL 値は 900 秒、3600 秒、28800 秒、86400 秒で、観客動員数は、各々、260,115 人、247,042 人、302,068 人、288,050 人である。図 4.8 に示すように、観客動員数が同程度であっても、TTL 値の違いによらず、クエリ数は多い日で 50 回程度を記録しグラフの傾向も類似している。このことから、TTL 値が短ければクエリ数は多くなるなど、TTL 値の長短とクエリ数の大小には必ずしも明確な規則性があるとは言えない。

表 4.10 は、ある同一の DNS クエリ送出元 IP アドレスにおける TTL 値別、観客動員数別（100,000 人単位）のクエリ数比較である。観客動員数別に示すクエリ数は、DNS クエリ初観測日から映画公開前日までの期間における 1 日あたりの最大数である。TTL 値と観客動員数の接点（列と行の接点）内に対象となる映画作品が複数存在する場合は、その中からクエリ数が最大となるものを記載した。たとえば TTL 値が 86400 秒で、観客動員数が 100,000 人以下の作品群において、同じ DNS クエリ送出元 IP アドレスから 1 日最大 36 回のクエリが存在した作品があることを示している。表 4.10 から、本研究で取り上げた 50 作品の観客動員数別 1 日当たりのクエリ数と、TTL 値までキャッシュした場合の 1 日当たりの想定クエリ数の間には、明確な規則性は見られない。

以上の考察により、次の 2 点が明らかとなった。

- ① TTL 値の長短とクエリ数の大小には、必ずしも明確な規則性があるとは言えないこと。
- ② 観客動員数別 1 日当たりのクエリ数と、TTL 値までキャッシュした場合の 1 日当たりの想定クエリ数の間には、規則性が見られないこと。

これらの結果から、本研究において TTL 値は、観客動員数に対する予測力に大きな影響を与えているとは言えないと推測できる。

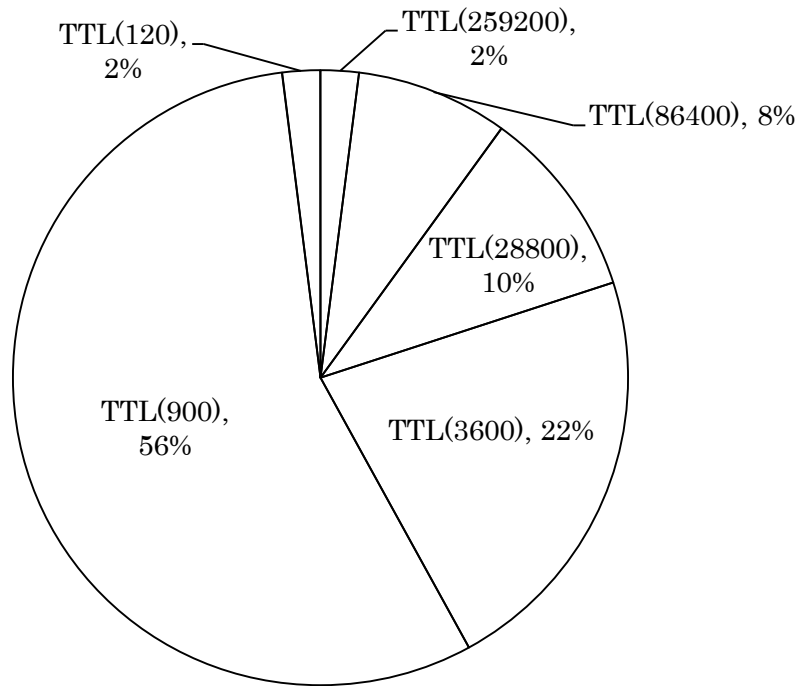


図 4.7 TTL 値の割合

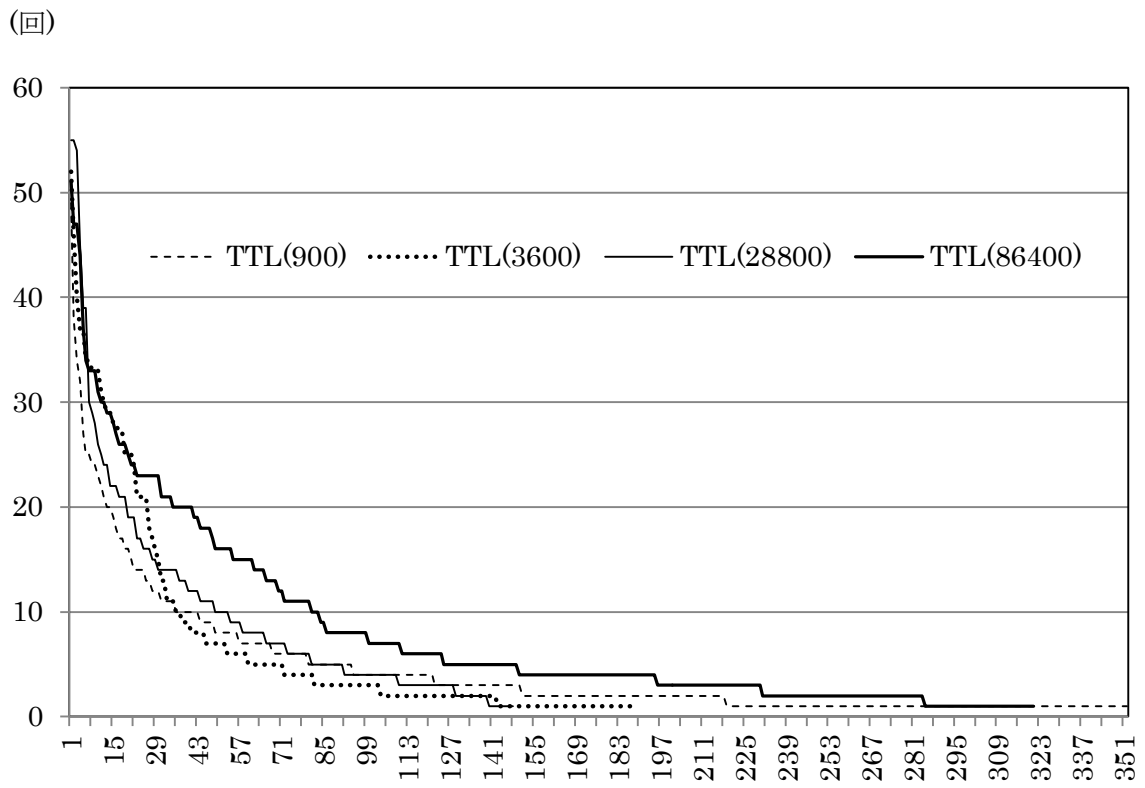


図 4.8 TTL 値とクエリ数 (回)

表 4.10 TTL 値別，観客動員数別クエリ数比較

TTL 値 (秒)		120	900	3600
観客動員数	0-100,000	-	53	74
	100,001-200,000	153	70	52
	200,001-300,000	-	51	52
	300,001-400,000	-	46	-
	400,001-500,000	-	89	-
	800,001-900,000	-	-	88
	900,001-1000,000	-	39	-
TTL 値までキャッシュした場合の 1 日当たりの想定クエリ数		720	96	24
TTL 値 (秒)		28800	86400	259200
観客動員数	0-100,000	-	36	-
	100,001-200,000	21	51	15
	200,001-300,000	21	51	-
	300,001-400,000	78	-	-
	400,001-500,000	-	-	-
	800,001-900,000	-	-	-
	900,001-1000,000	-	-	-
TTL 値までキャッシュした場合の 1 日当たりの想定クエリ数		3	1	0.3



## 4.8. 本章のまとめ

本章では、提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、ソーシャルメディアを利用した予測に関する研究領域において、提案手法の有効性を確認するために日本映画の観客動員数予測を試みた。

行動分析に関する先行研究は、①投稿者の投稿内容を分析する、②調査票を用いたアンケート調査結果を分析する、③検索エンジン結果を分析する、が主な手法であった。しかし、これらの手法は、“作為的投稿影響問題”や“利用者バイアス問題”を抱えている。そこで本章では、先行研究が抱える問題に対応する提案手法を用いて映画の観客動員数予測を行い、提案手法の有効性を確認した。これにより、投稿者側の視点からではなく、DNS クエリデータを用いた閲覧者側の視点から、ソーシャルメディア利用者の実社会における行動予測が可能であり、提案手法が先行研究の抱える問題に対応できることを確認した。提案手法は、DNS クエリログを用いてソーシャルメディア利用者の実社会における行動分析を行う手法であるが、映画の観客動員数の予測結果からは、ソーシャルメディア利用者の行動分析に加え、実社会で起きる現象の動向予測になっていることに注意されたい。

また、実務への有用性については、実務家へのインタビューで最低でも映画公開の2週間前、可能であれば1ヶ月から3ヶ月くらい前の時点で、観客動員数の予測が求められていることがわかった。本提案手法では、映画公開の1ヶ月前においても観客動員数の予実の差異が7.5万人未満または37.5%未満（表4.9のA+B+C）に収まるものが64%で、実務にも有用な精度を持つことを明らかにした。



## 第5章 ソーシャルメディアを利用した計測・モデル化に関する研究

前章では、提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、日本映画の観客動員数を予測することで、提案手法が先行研究の抱える問題に対応できることを確認した。

本章では、提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、提案手法が大規模で複雑な行動分析も可能とすることを確認する。更に、そのソーシャルメディア利用者の行動の時間的推移も明らかにする。具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における複数の社会現象に対し、家電エコポイント数の発行数とDNSクエリデータにもとづくソーシャルメディア利用の関係性を計測することで、ソーシャルメディア利用者の行動の時間的推移を分析する<sup>39</sup>。

### 5.1. 日本における地上波デジタル放送への移行と家電エコポイント制度

2011年7月、日本ではアナログテレビ放送を停止し、地上波デジタル放送への移行を完了させる。日本政府が2006年にアナログテレビ放送からデジタルテレビ放送への移行を発表した後、日本ではこの移行について多くの注目が集まった。日本政府は、2009年から様々なWebサイトを立ち上げ、それらを活用しながら政策を推進した。家電エコポイント制度<sup>40</sup>は、政策推進のための1つの施策であり、①地球温暖化対策の推進、②経済活性化、③地上デジタル放送対応テレビの普及を目的として、統一省エネラベル4☆相当以上のエ

---

<sup>39</sup> 本章の内容は[Mitamura 2013]に報告した内容にもとづく。

<sup>40</sup> <http://www.meti.go.jp/press/2011/06/20110614002/20110614002-2.pdf> (2013/10/13)

アコン・冷蔵庫・地上デジタル放送対応テレビの購入者に対し、2009年5月から2011年3月の購入期限までエコポイントが発行された政策である（平成23年1月以降、エコポイントの発行対象を5☆製品への買い換えの場合に変更）。地上デジタル放送対応テレビ普及の観点では、家電エコポイント制度を実施したことで、地上デジタル放送対応テレビの国内出荷台数累計が約2.2倍に増加したと報告されている<sup>41</sup>。

本研究では、提案手法により地上波デジタル放送への移行と家電エコポイント制度に関するソーシャルメディア利用者の行動分析を行い、政府施策の妥当性の検証も試みる。

## 5.2. 分析の概要

分析対象とする地上波デジタル放送への移行と家電エコポイント制度は、主に2009年～2011年の間に社会的現象が発生しているため、本研究も2009年1月から2011年12月までTLDレベルのDNSクエリログを分析した。具体的には、地上デジタル放送対応テレビと家電エコポイント制度に対するソーシャルメディア利用者の行動を評価するために、関連WebサイトへのDNSクエリ数とそれらの異なり数を取得<sup>42</sup>し、家電エコポイント制度を利用して、ソーシャルメディア利用者が申請したエコポイント数を目的変数とした主成分回帰分析（PCR）を実施した。

表5.1に用いた変数を示す。政府が管理運営を行い、家電エコポイント制度に関するニュース配信やポイント登録などを実施するための総合Webサイト関連の説明変数として、Eco-IPC（IPアドレス異なり数）、Eco-CC（国異なり数）と。専門団体が管理運営を行い、デジタル放送を普及促進するための総合Webサイト関連の説明変数として、D-IPC（IPアドレス異なり数）、D-CC（国異なり数）を用いた。

また、政府は2009年にアナログ放送からデジタルテレビ放送への移行促進のために、2つの関連Webサイトを開設した。本研究では、これらの2つのWebサイトに関連する変数として、Da-IPC（IPアドレス異なり数）、Da-CC（国異なり数）、Db-IPC（IPアドレス異なり数）、Db-CC（国異なり数）も用いた。

<sup>41</sup> <http://www.env.go.jp/council/02policy/y020-60/mat03.pdf> (2013/10/13)

<sup>42</sup> 4つの重要なWebサイトに関するdns.jpのDNSクエリログから、4つの重要なWebサイトのアクセス数と異なり数を取得している。第3.4節のStep1-3に相当

表 5.1 変数

目的変数	
エコポイント発行数	主管庁から月次で公表されるエコポイント数 (地上デジタル放送対応テレビで発行された数)
説明変数	
Eco-IPC:IP アドレス異なり数 (エコポイント関連サイト)	①地球温暖化対策の推進, ②経済活性化, ③地上デジタル放送対応テレビの普及を目的とした家電エコポイント制度のポータルサイトの IP アドレス異なり数と国異なり数
Eco-CC:国異なり数 (エコポイント関連サイト)	
D-IPC : IP アドレス異なり数 (デジタル放送関連サイト)	アナログ放送からデジタル放送への移行に必要な地上波放送/衛星放送デジタルフォーマットの普及を目的とした総合サイトの IP アドレス異なり数と国異なり数
D-CC : 国異なり数 (デジタル放送関連)	
Da-IPC : IP アドレス異なり数 (デジタル放送支援 A サイト)	地上波デジタル放送への移行のためのサポート総合サイトの IP アドレス異なり数と国異なり数
Da-CC : 国異なり数 (デジタル放送支援 A サイト)	
Db-IPC:IP アドレス異なり数 (デジタル放送支援 B サイト)	経済理由のために地球波デジタル放送を視聴できない人などの支援を目的としたサイトの IP アドレス異なり数と国異なり数
Db-CC : 国異なり数 (デジタル放送支援 B サイト)	
Date	日本時間をシリアル値に変換した数値

変数に使用した異なり数データが、3.3 項で示した DNS キャッシュの影響を取り除く効果を持つことを[三田村 2010]は示しており、本研究でも同様に異なり数を変数として用いた。日時 (Date) は、ソーシャルメディア利用者の行動推移を時系列で分析する上で、重要な変数であると考え説明変数の 1 つとして用いた。

目的変数は、家電エコポイント制度の主管庁である政府機関から月次で公表されるエコポイント発行数とした (第 3.4 節の Step4 で用いられる)。

以上、分析に用いられる説明変数はすべて、DNS クエリログから得られる異なり数である。また目的変数は、地上デジタル放送対応テレビの拡販という経済効果を狙った施策に関する 1 経済指標である。つまり本提案手法は、インターネットで発生する現象と実社会で発生する現象を関係付けることを意図している。

### 5.3. モデルの構築

DNS クエリログ (第 3.3 節で述べた提案手法の Step5) にもとづき毎月公表される発行エコポイント数を予測した PCR 分析結果を表 5.2, 図 5.1 に示す。

主成分回帰分析の最初のステップとして、主成分分析 (PCA) を実行し 3 つの主成分 (Component1, 2, 3) を選択した。選択基準は、固有値 (Eigenvalues) が 1 以上の主成分とし、その累積寄与率 (Cumulative Proportion of Variance) は表 5.2 に示すように 0.93851 である。選択した 3 つの主成分を用いて主成分回帰分析を行った結果が図 5.1 である。相関係数  $R$  は 0.971 となり予測力は高い。

以上の結果は、本提案手法 (何人の閲覧者がいたかという情報に対応する DNS クエリデータに基づく予測) が、実社会で起きる現象 (すなわち、実際に発行されるエコポイント数) のモデル化に用いることができることを示している (第 3.3 節, Step5)。言い換えると、本提案手法は、“地上波デジタル放送への移行”と、それを後押しする政策の 1 つであった“家電エコポイント制度の実施”という実社会で起きた複数の大きな社会的現象に対する、ソーシャルメディア利用者の行動分析が可能であるということである。これは本提案手法が、第 4 章の実証検証よりも、より大規模で複雑な実社会で起きる現象の分析も可能であり、“作為的投稿影響問題”や“利用者バイアス問題”への対応も可能であることを示している。

表 5.2 主成分分析結果（月次）

Component	Eigenvalues	Proportion of Variance (%)	Cumulative Proportion of Variance (%)
1	5.895	65.502	65.502
2	1.475	16.394	81.896
3	1.076	11.955	93.851

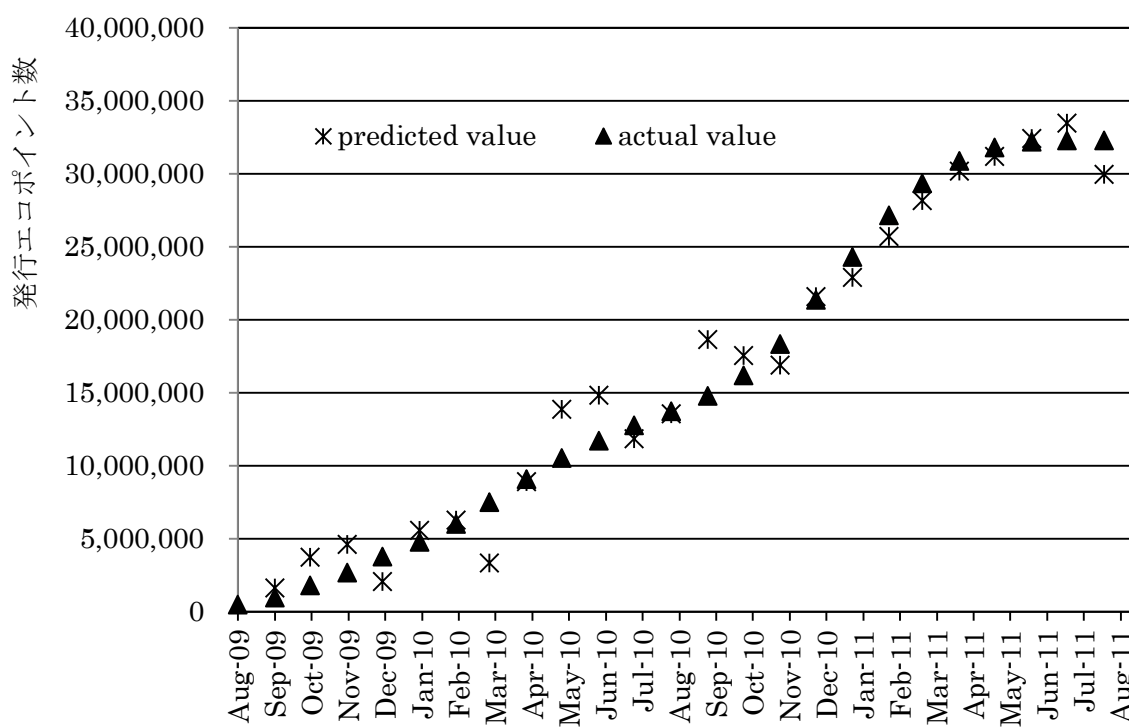


図 5.1 主成分回帰分析結果

## 5.4. 主成分分析とソーシャルメディア利用者の行動

前項で示したように、本提案手法は、ソーシャルメディア利用者の行動を示す DNS クエリログから抽出される主成分を用いて、大規模で複雑な実社会の現象を分析可能なモデルを構築することができる。

本節では、それを踏まえ、ソーシャルメディア利用者の行動を分析（第 3.3 節, Step6）する。利用者の行動を詳細に分析するために、本節では日次ベースの DNS クエリログを分析に用いる。

表 5.3 に主成分分析結果を示す。3 つの主成分（Component1, 2, 3）を選択し、その累積寄与率（Cumulative Proportion of Variance）は 0.85086 である。図 5.2 は、3 つの主成分得点の日次変化傾向を示している。図 5.3 は、3 つの主成分の主成分得点におけるピーク変化傾向を検出するために図 5.2 を基に作成している。図 5.4 を用いて、図 5.2 から図 5.3 の作成について説明する。図 5.4 は、図 5.2 の各主成分をヒストグラム化したものである。図 5.4 に示す「抽出部分」は、上位 2.5%以上の主成分得点と、下位 2.5%以下の主成分得点を表している。その抽出部分のみで作成したものが図 5.3 である。表 5.4 は、図 5.3 の中で使用した識別子（A～O）の当該日に発生した社会現象について説明している。

続いて 3 つの主成分解釈を行う。まず第 1 主成分（Component1）であるが、テレビ放送を示す“TV Broadcasting”に関するソーシャルメディア利用者の行動と解釈した。

### 【第 1 主成分の解釈理由】

- 第 1 主成分の主成分得点は、地上波デジタル放送への移行が行われた期間である 2009 年 1 月から 2011 年 12 月まで増加傾向を示している（図 5.2 を参照）。特に、2011 年 7 月 25 日（アナログ放送終了の翌日）に強いピークを示している。
- 第 1 主成分の主成分負荷量（Principal Component Loadings）は、全変数に対して相対的に大きな値を示す傾向がある（表 5.3 を参照）。これらの傾向は日本がこの期間、テレビ放送全般に関わる大きな社会的イベントが発生した、という事実を反映していると考えられる。その裏付けとして第 1 主成分の顕著なピークが、いろいろなイベントの発表日付と一致している（図 5.3 と表 5.4 を参照）。
  - 2010 年 8 月 31 日（図 5.3 中の A）にアナログ放送が停止され、2011 年 1 月 25 日（図 5.3 中の B）にデジタル放送への移行計画が発表されている。



表 5.3 主成分分析結果（日次）

Component	Eigenvalues	Proportion of Variance (%)	Cumulative Proportion of Variance (%)
1	5.470	60.782	60.782
2	1.239	13.769	74.551
3	0.948	10.535	85.086
Variables	Principal Component Loadings		
	<i>Component1</i>	<i>Component2</i>	<i>Component3</i>
Eco-IPC	0.540	0.335	0.730
D-IPC	0.334	0.861	-0.297
Da-IPC	0.852	0.227	-0.262
Db-IPC	0.905	0.041	0.050
Eco-CC	0.855	-0.242	0.381
D-CC	0.631	0.271	0.083
Da-CC	0.908	-0.182	-0.207
Db-CC	0.906	-0.218	-0.159
Date	0.865	-0.348	-0.188

- 2011年7月11日（図5.3中のC）に東日本大震災発生によるデジタル放送への移行延期が発表され、2011年7月25日（図5.3中のD）に地上波デジタル放送への移行が完了している。
- 2011年9月9日（図5.3中のE）にNHK放送の契約解約数増加に関する発表が行われている。

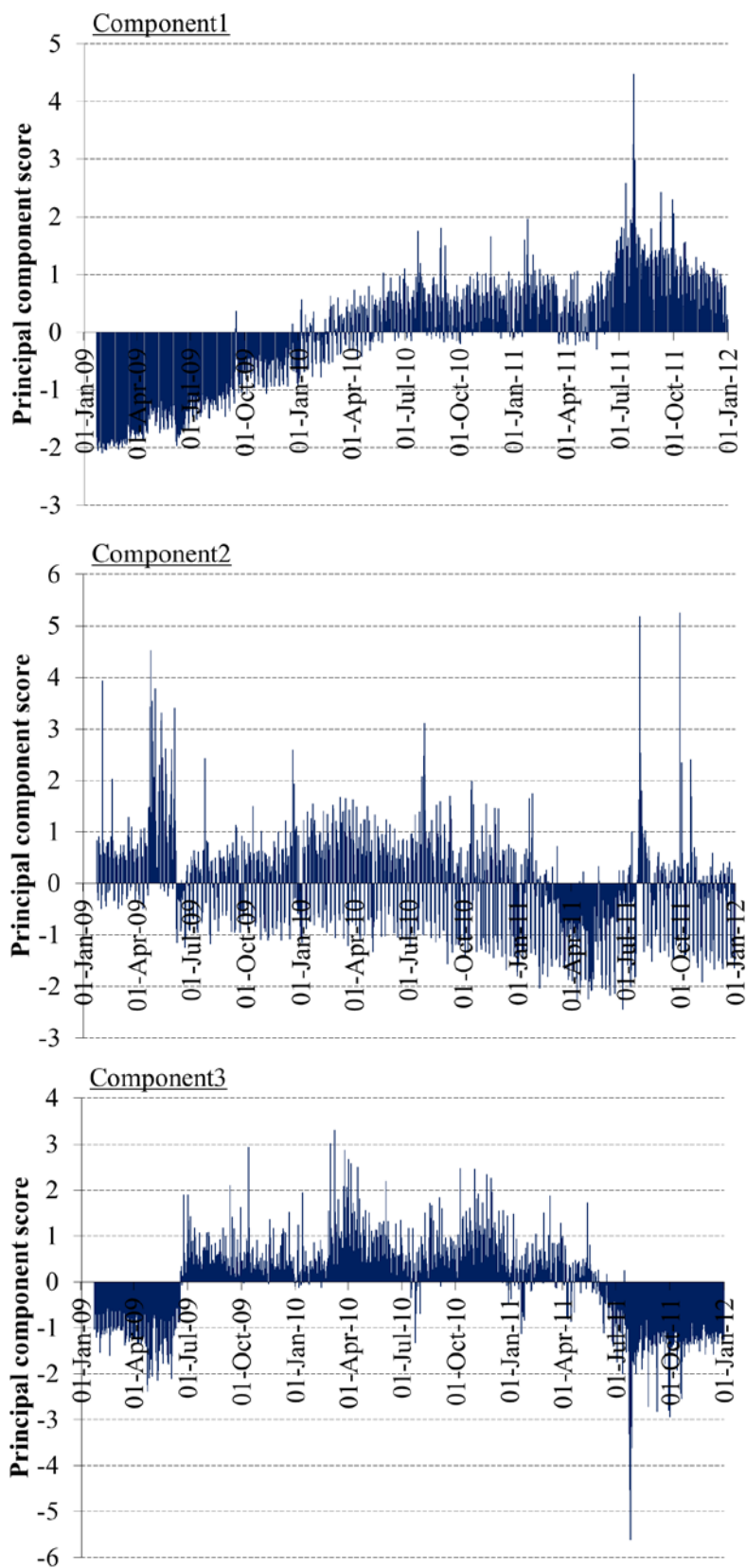


图 5.2 主成分得分

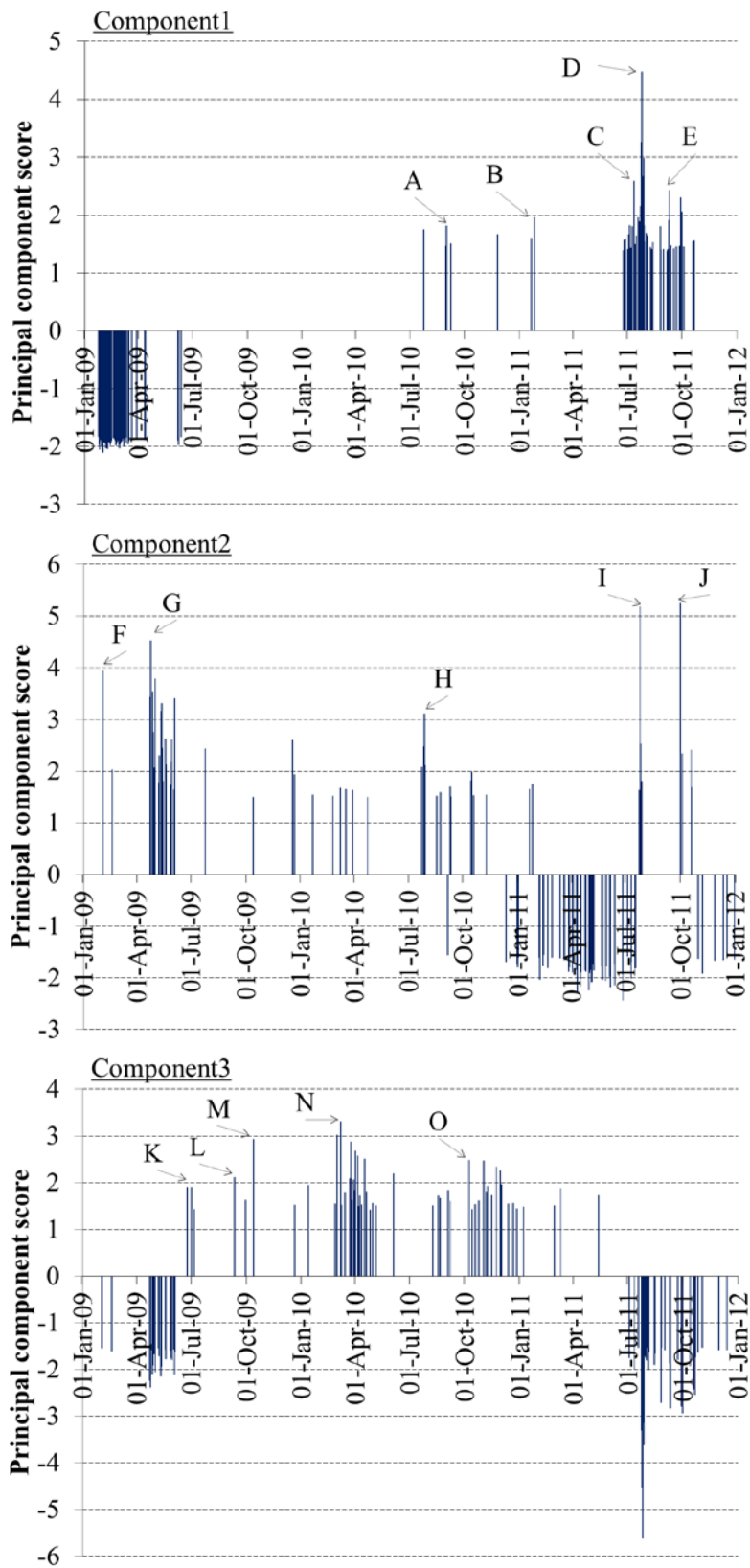


図 5.3 主成分得点ピーク

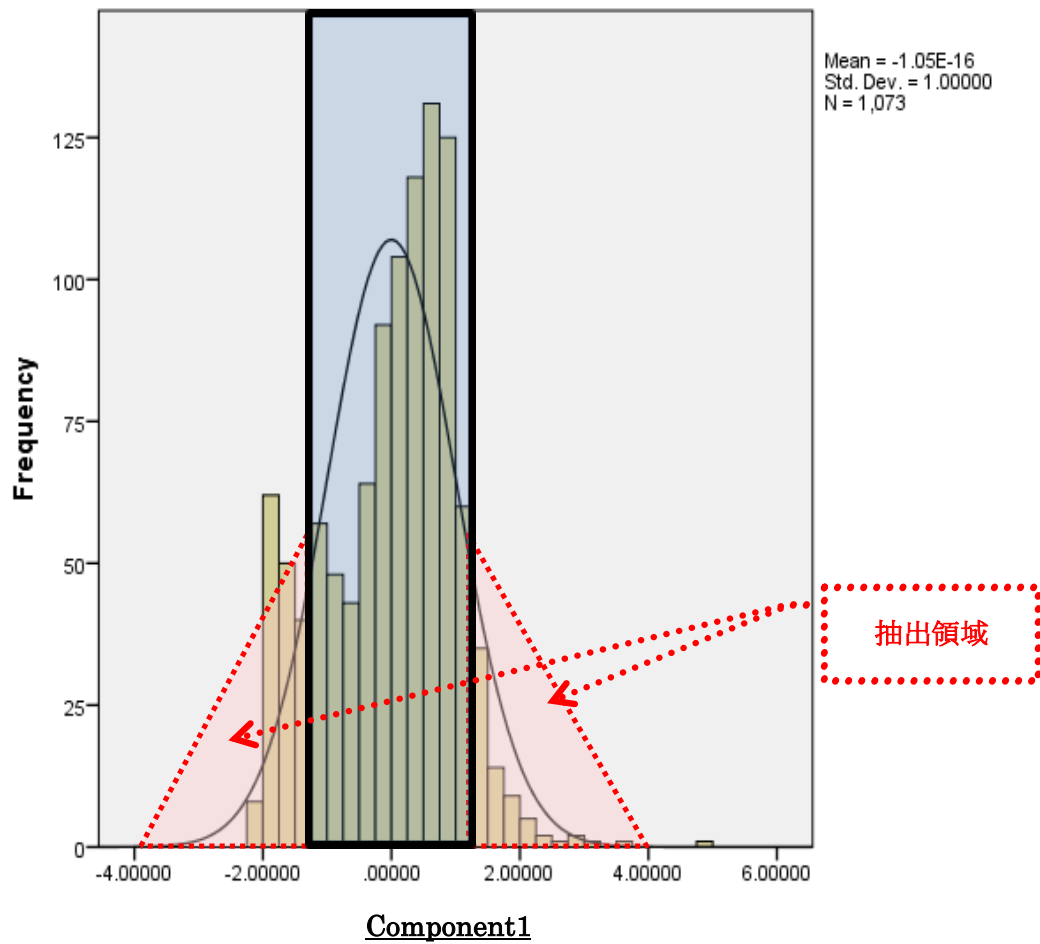


図 5.4 主成分パーセンタイル処理

表 5.4 発生イベント

ID (図 5.3)	発生イベント
A (31 Aug.2010)	政府によるアナログ放送停止のアナウンス (31 Aug. 2010).
B (25 Jan.2011)	政府による「完全デジタル化最終行動計画」及び「完全デジタル化に向けた最終国民運動」のアナウンス (24 Jan. 2011).
C (11 Jul.2011)	政府による東日本大震災による東北 3 県における地上アナログ放送用周波数の使用期限延長の公表 (11 Jul. 2011).
D (25 Jul.2011)	地上波デジタル放送への移行完了 (24 Jul. 2011).
E (9 Sep.2011)	NHK によるアナログ停波での 90000 件の契約解除報告 (8 Sep. 2011).
F (2 Feb.2009)	政府による「総務省テレビ受信者支援センター」が全都道府県拡充のアナウンス (2 Feb. 2009).
G (24 Apr.2009)	政府による地上波デジタル放送への移行を 2011 年 7 月 24 日実施のアナウンス (23 Apr. 2009).
H (27 Jul.2010)	石川県珠洲市でデジタル放送への切り替え実施 (24 Jul. 2010).
I (25 Jul.2011)	地上波デジタル放送への移行実施 (24 Jul. 2010).
J (30 Sep.2011)	BS デジタル放送のマルチチャンネル化実施 (24 Jul. 2011).
K (24 Jun.2009)	政府による家電エコポイント対象の製品とサービスを公表 (24 Jun. 2009).
L (11 Sep.2009)	政府による家電エコポイント対象の製品/サービスの更新公表 (11 Sep. 2009).
M (13 Oct.2009)	家電エコポイント制度継続の政府次年度予算議論の公表 (13 Oct. 2009).
N (25 Mar.2010)	政府による家電エコポイント対象の地上波デジタル TV 対応商品へのポイント増加アナウンス (25 Mar. 2010).
O (8 Oct.2010)	政府による家電エコポイント対象の地上波デジタル TV 対応商品へのポイント減少アナウンス (8 Oct. 2010).

次に第 2 主成分 (Component2) であるが、アナログ放送からデジタル放送への移行を示す “switchover” に関するソーシャルメディア利用者の行動と解釈した。

#### 【第 2 主成分の解釈理由】

- 第 2 主成分の主成分得点は、2009 年 2 月から 2011 年 10 月まで断続的にピークを繰り返している (図 5.2 を参照)。特に、強いピークは 2009 年と 2011 年に現れ、その各ピークはアナログ放送からデジタル放送への移行時期と一致している。
  - 2009 年 4 月 24 日に日本政府は、地上波デジタル放送への移行予定を発表し、2011 年 7 月 25 日にアナログ放送から地上波デジタル放送への移行を実施している。
- 第 2 主成分の主成分負荷量 (Principal Component Loadings) は、デジタル放送に関する変数に対して相対的に大きな値を示す傾向がある (表 5.3 を参照)。これらの傾向は日本がこの期間、デジタル放送に関わる大きな社会的イベントが発生した、という事実を反映していると考えられる。更に第 2 主成分は、2009 年から 2011 年まで、地上波デジタル放送への移行に関するイベントと一致した強いピークを示している (図 5.3 と表 5.4 を参照)。
  - 2009 年 2 月 2 日 (図 5.3 中の F) に日本政府は、地上波デジタルテレビ・サポート・センターを設立し、2009 年 4 月 23 日に地上波デジタル放送への完全移行 (図 5.3 中の G) を公式に発表している。
  - 2010 年 7 月 27 日 (図 5.3 中の H) に石川県珠洲市では、地上波デジタル放送への移行リハーサルを完了し、地上波デジタル放送への移行日である 2011 年 7 月 24 日 (図 5.3 の I) と第 2 主成分のピーク日は一致している。
  - 2011 年 9 月 30 日 (図 5.3 中の J) に BS デジタル放送は、アナログ放送からデジタル放送に移行しており、それも第 2 主成分のピーク日と一致している。

最後に第 3 主成分 (Component3) であるが、家電エコポイント制度を示す “eco-points” に関するソーシャルメディア利用者の行動と解釈した。

### 【第3主成分の解釈理由】

- 家電購入者が、家電エコポイント制度により助成金を受けられることができる期間（2009年5月15日～2011年3月31日）に、その主成分得点はより高い値を示している（図5.2を参照）。
- 第3主成分の主成分負荷量（Principal Component Loadings）は、家電エコポイント制度に関する変数に対して相対的に大きな値を示す傾向がある（表5.3を参照）。これらの傾向は日本がこの期間、家電エコポイント制度に関わる大きな社会的イベントが発生した、という事実を反映していると考えられる。更に第3主成分は、2009年から2011年まで、家電エコポイント制度に関するイベントと一致した強いピークを示している（図5.3と表5.4を参照）。
  - 2009年6月24日と2009年9月11日（図5.3中のK, L）に付与エコポイント数に関わる製品とサービスが発表されている。
  - 2009年10月13日（図5.3中のM）に日本政府による家電エコポイント制度継続に関わる次年度予算策定の発表されている。
  - 2010年3月25日と2010年10月8日（図5.3中のN, O）に家電エコポイント交換レートの変更内容についての日本政府の公式発表が行われている（図5.3と表5.4を参照）。

## 5.5. ソーシャルメディア利用者の行動推移

前節では、ソーシャルメディア利用者のDNSクエリログから抽出された主成分が、ソーシャルメディア利用者の行動に関わる実社会の現象を表す1つの指標として解釈できることを示した。これは言い換えると、抽出された主成分得点の時間的推移を計測すると、ソーシャルメディア利用者の行動の時間的推移を推測することができるということである。以上を踏まえ本節では、ソーシャルメディア利用者の行動の時間的推移を分析する。

図5.5は、5.3節および5.4節で論じた3つの主成分（Component1：“TV Broadcasting”，Component2：“switchover”，Component3：“eco-points”）の推移を表現するために用いる説明図である。

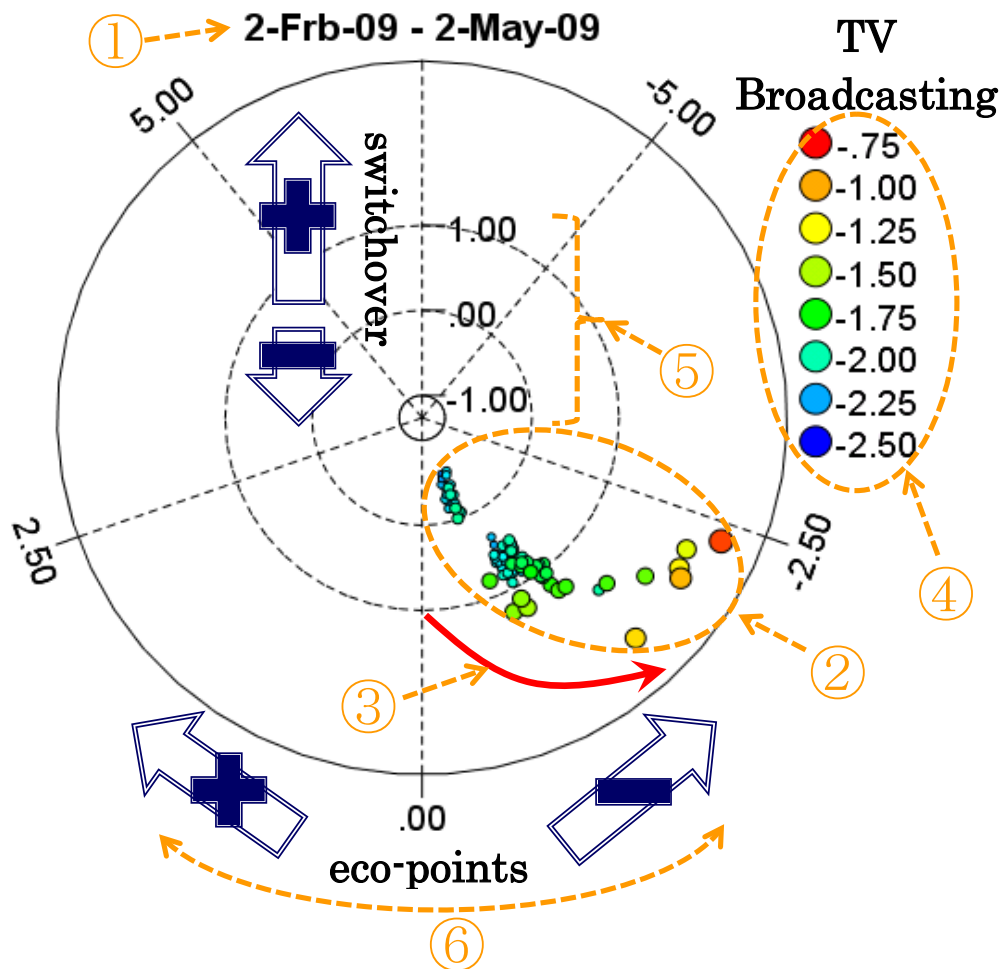


図 5.5 ソーシャルメディア利用者の行動推移サンプル

図中のタイトル（図 5.5 中の①）は、計測対象期間を示している。各バブル（図 5.5 中の②）は、計測対象期間の各日付の主成分得点を示している（言い換えると、各日付のソーシャルメディア利用者の行動を示している）。バブルサイズとバブルカラー（図 5.5 中の④）は、第 1 主成分（“TV Broadcasting” に関わるソーシャルメディア利用者の行動）に対応しており、より大きなバブルサイズは“TV Broadcasting”に対するより強い反応を示し、バブルカラーも青から赤に近づくほど“TV Broadcasting”に対するより強い反応を示している。

中心からの距離（図 5.5 中の⑤）は、第 2 主成分（“switchover” に関するソーシャルメディア利用者の行動）に対応しており、バブルがより外に向かうほど“switchover” に対



するより強い反応を示している。またチャート底から左右に向かう角度（図 5.5 中の⑥）は、第 3 主成分（“eco-points”に関するソーシャルメディア利用者の行動）に対応しており、より左に向かうほど“eco-points”に対するより強い反応を示し、より右に行くほどより弱い反応を示している。バブルは計測対象期間において、特定の推移を示すので、その方向を赤い矢印（図 5.5 中の③）として示している。

図 5.6 に、2009 年から 2011 年におけるソーシャルメディア利用者の行動の時間的推移を示す。本計測では、ソーシャルメディア利用者の行動の時間的推移を 3 か月単位に表現している。これは、①各月ベースの分析では、地上波デジタル放送への移行と家電エコポイント制度に関する日本政府による政策の影響評価を行うには短いこと、②[経済産業省 2011]では、家電エコポイント制度の効果計測を 3 か月間で分析していること、という 2 つの理由からである。

まず 2009 年 2 月 2 日から 2009 年 5 月 2 日までの期間（図 5.6 中の(1)）において、“switchover”のみソーシャルメディア利用者の強い反応を示しているが、“TV Broadcasting”と“eco-points”については、まだその傾向が見られない。これは、アナログ放送を停止しデジタル放送へ完全移行することを日本政府が公式発表（図 5.3 と表 5.4 の G）したことが、“switchover”に対する強い反応を引き起こしていると考えられる。しかし、“TV Broadcasting”については強い反応にはないので、まだソーシャルメディア利用者の行動は、テレビ放送全体に対しての反応としては現れていないと解釈できる。また、“eco-points”についても強い反応にはないが、家電エコポイント制度は 2009 年 5 月に開始されたので、この時点でのソーシャルメディア利用者の行動としては自然な現象であると考えられる。

次に家電エコポイント制度が始まった 2009 年 4 月 23 日から 2009 年 7 月 23 日までの期間（図 5.6 中の(2)）であるが、“TV Broadcasting”、“switchover”についての大きな変化は見られないが、“eco-points”についてはソーシャルメディア利用者の強い反応を示し始めている。この“eco-points”についての反応は、2010 年 1 月 12 日から 2010 年 4 月 12 日の期間（図 5.6 中の(3)）まで継続している。

特徴的なのは、それまで変化を示さなかった“TV Broadcasting”が、2010 年 1 月 12 日から 2010 年 4 月 12 日の期間において強い反応を示してきていることである。これは、地上デジタル放送対応テレビに対する付与エコポイント数が増加（図 5.3 と表 5.4 の G）したためであると考えられる。

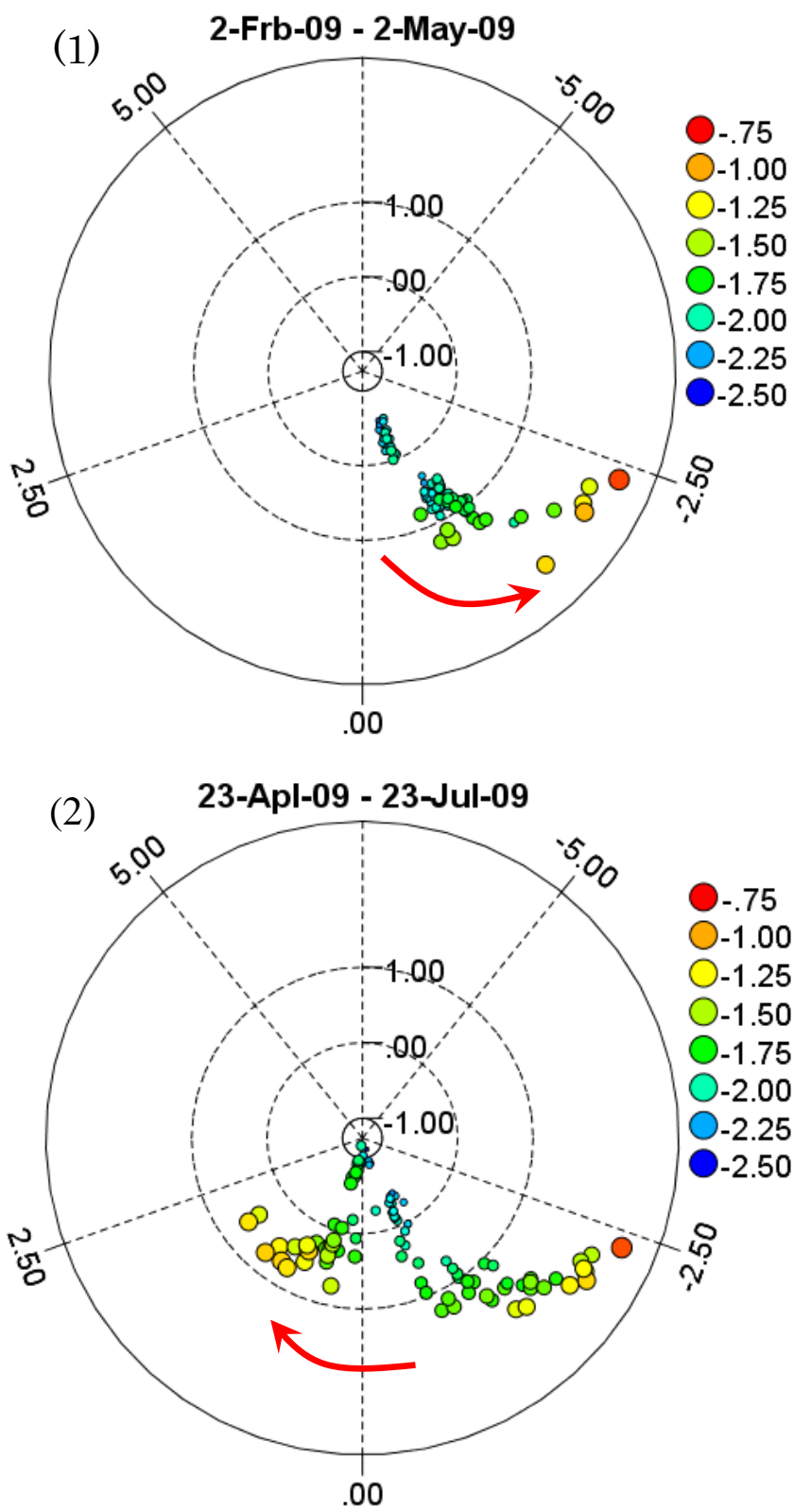


図 5.6 ソーシャルメディア利用者の行動推移-1

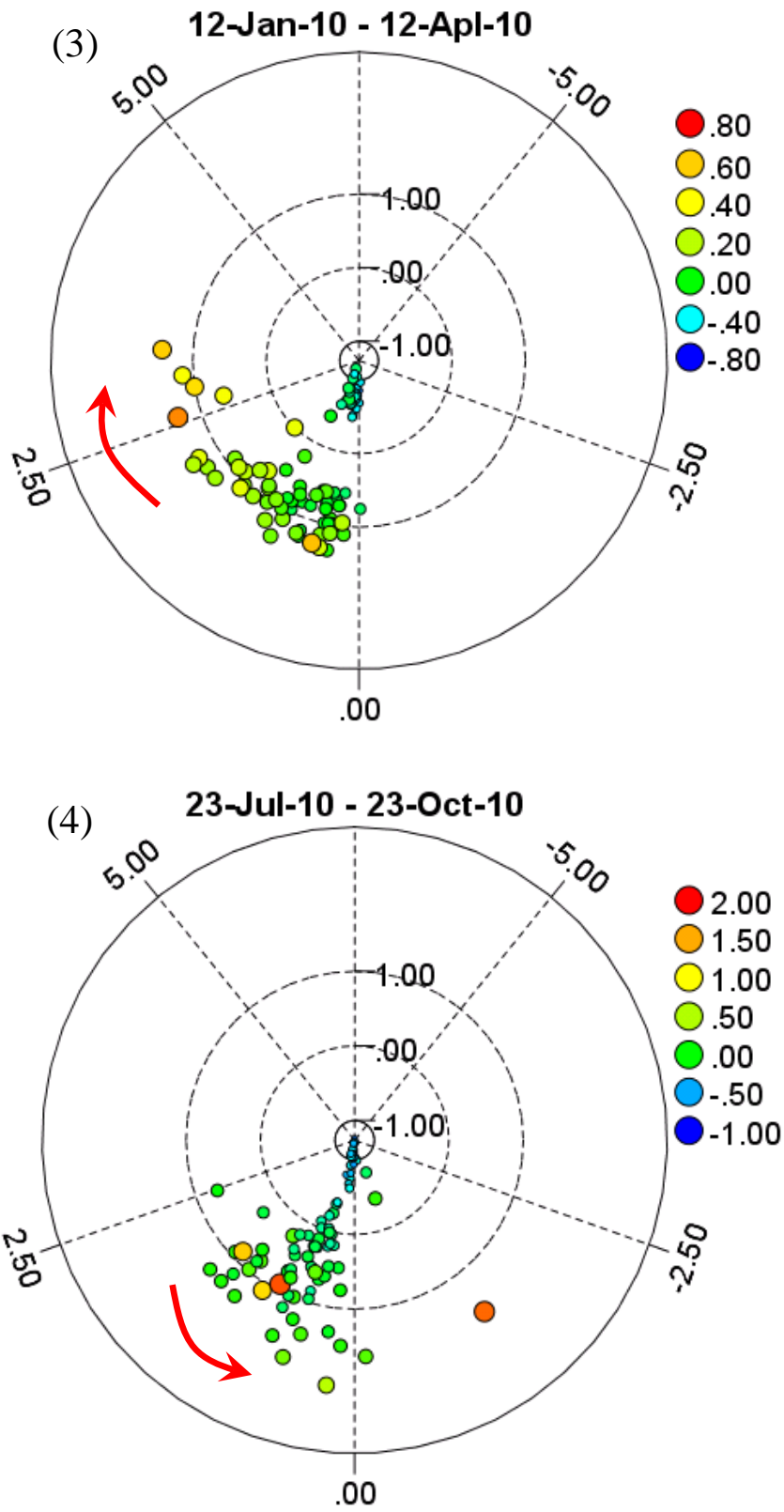


図 5.6 ソーシャルメディア利用者の行動推移-2

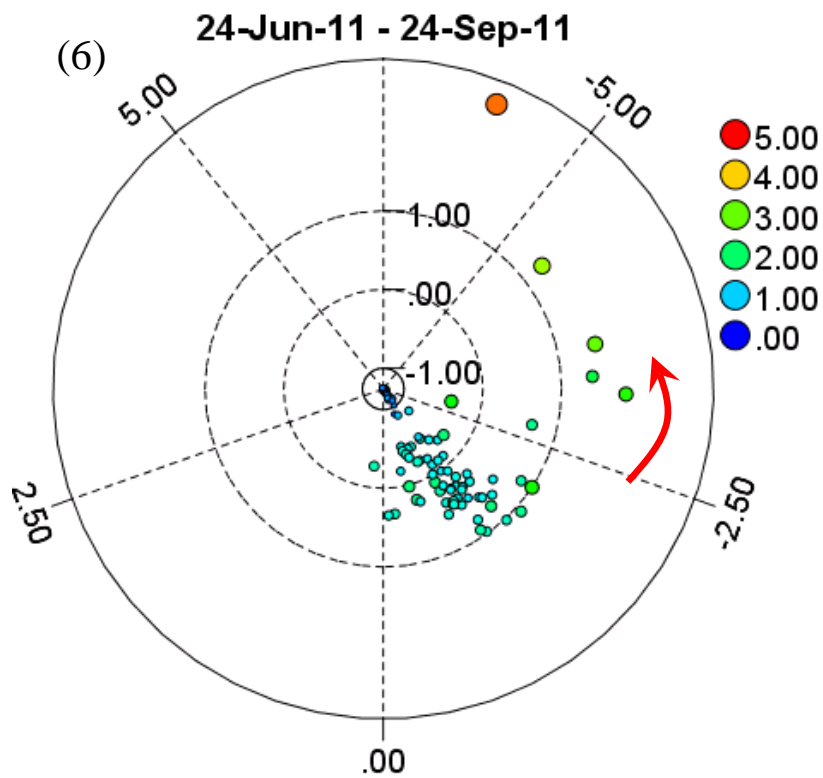
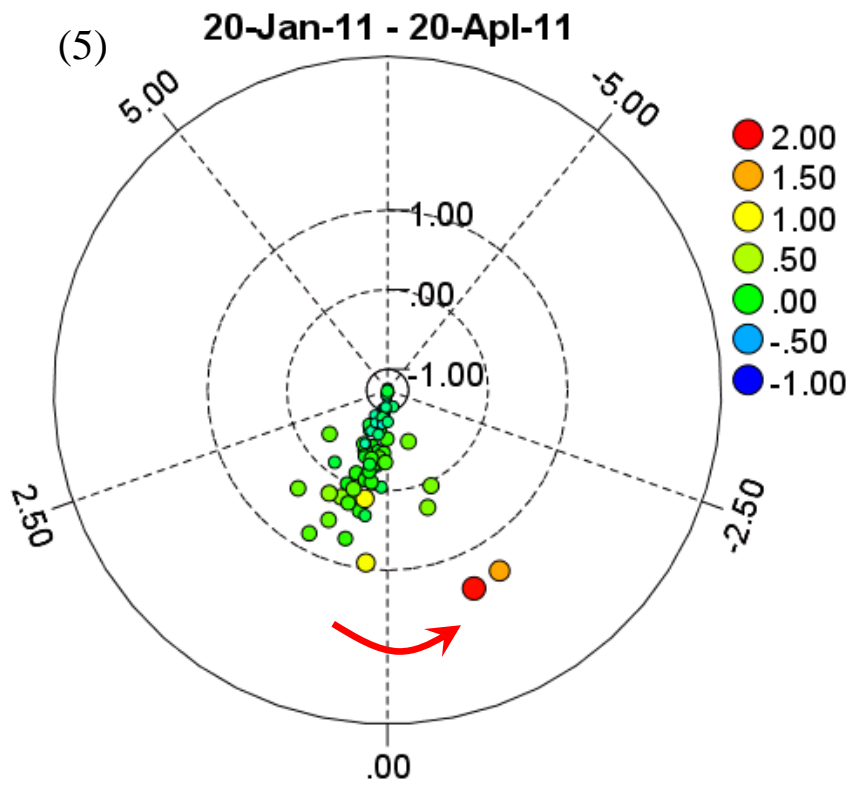


図 5.6 ソーシャルメディア利用者の行動推移-3

同様の理由から、地上デジタル放送対応テレビに対する付与エコポイント数が減少（図 5.3 と表 5.4 の O）した期間である 2010 年 7 月 23 日から 2010 年 10 月 23 日の期間（図 5.6 中の(4)）と、家電エコポイント制度が終了した期間である 2011 年 1 月 20 日から 2011 年 4 月 20 日（図 5.6 中の(5)）、2011 年 6 月 24 日から 2011 年 9 月 24 日（図 5.6 中の(6)）は、ソーシャルメディア利用者の弱い反応が継続している。

2011 年 6 月 24 日から 2011 年 9 月 24 日の期間（図 5.6 中の(6)）では、一部特異的な現象がみられる。アナログ放送が停止されデジタル放送への移行が完了（図 5.3 と表 5.4 の D）したにも関わらず“TV Broadcasting”，“switchover”に対するソーシャルメディア利用者の反応は強くなっている。更なる詳細な分析が必要であるが，“TV Broadcasting”，“switchover”に関する様々な社会的問題が発生したことに起因しているとも考えられる。たとえば NHK は、アナログ放送からデジタル放送への移行上の問題が発生したことから、契約キャンセル数が増加したと発表している（図 5.3 と表 5.4 の E）[NHK 2011]。

以上をまとめると、日本政府が行った 2009 年から 2011 年に及ぶアナログ放送の停止とデジタル放送への移行という政策において、家電エコポイント制度が、その目的の 1 つであった地上デジタル放送対応テレビの普及で、ソーシャルメディア利用者の行動に効果的な影響を与えた経済政策であったことがわかる。具体的には、地上デジタル放送対応テレビに対する付与エコポイント数をコントロールする施策により、ソーシャルメディア利用者の行動において、テレビ放送全体に対しての行動へ広がりを示しており、地上デジタル放送対応への切り替えを加速させる傾向を示した。

本章における実証検証により、DNS クエリデータを用いた閲覧者側の視点から、より大規模で複雑な実社会で起きる分析が可能であることを明らかにした。

## 5.6. 考察

第4章で述べたソーシャルメディア利用者の日本映画への行動予測と、本章で述べた日本における地上波デジタル放送への移行と家電エコポイント制度に関するソーシャルメディア利用者の行動分析について比較考察する。

第4章では、実社会で起きた1つの現象（日本映画に対するソーシャルメディア利用者の行動）を対象にして実証検証を行った。それに対して本章では、実社会で起きたより大規模で複雑な複数の現象（日本全土における3年間の政策であるアナログ放送停止とデジタル放送への移行、それらを支援する家電エコポイント制度に対するソーシャルメディア利用者の行動）を対象にして実証検証を行った。両検証について、分析期間と異なり数の比較を表5.5に示す。

表5.5に示すように、本提案手法が、実社会で起きる単体の現象だけではなく、より大規模で、より複雑な実社会で起きる現象についても分析が可能であり、先行研究の抱える問題にも対応できることを確認した。

表 5.5 アクセス数比較

データ比較（日）	日本映画の 観客動員数予測 （第4章）	地上波デジタル放送への移行 と家電エコポイント制度 （第5章）
計測期間	191.3	1213
平均国異なり数	8.6	38.4
平均 IP アドレス異なり数	225.6	1297.6

## 5.7. 本章のまとめ

本章では提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、提案手法が大規模で複雑な実社会で起きる現象の分析も可能であり、先行研究の抱える問題に対応できることを確認し、実社会で起きる現象の時間的推移も明らかにした。

具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における2つの大きな社会現象に対し、DNSクエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用の関係性を計測することで、その時間的推移を分析した。

この結果、日本政府が行った2009年から2011年に及ぶアナログ放送の停止とデジタル放送への移行という政策において、家電エコポイント制度が、その目的の1つであった地上デジタル放送対応テレビの普及で、ソーシャルメディア利用者の行動に効果的な影響を与えた経済政策であったことを示した。これにより、大規模なDNSクエリデータを用いた閲覧者側の視点から、より大規模で、より複雑な実社会で起きる現象の分析が可能であり、提案手法が先行研究の抱える問題に対応できることを確認した。

本提案手法は、上述のようにソーシャルメディア利用者の行動を客観的に把握することを可能とするが、2011年6月24日から2011年9月24日の期間（図5.6中の(6)）に見られた特異的な現象の原因分析など、更なる詳細な分析を行う必要がある場合は、調査票を用いたアンケートを併用するなど、他手法と組み合わせたより高度な分析手法の検討が必要であると考えられる。





## 第6章 まとめ

### 6.1. 研究成果のまとめ

本研究では、インターネットの基幹技術の1つであるDNSのクエリデータにおける“異なり数”を用いて、ソーシャルメディアの“閲覧者側”の視点から、実社会で起きる現象を分析する手法の提案を試みた。第2章で述べたとおり、ソーシャルメディア利用者の行動分析に関する先行研究は、以下の特徴と問題を持っており、投稿内容の影響を受けやすいという問題に十分に対応できていなかった。具体的には、先行研究の主な特徴は、以下の3点である。

- 研究の基本は、投稿者から投稿された情報の分析か、登録制のソーシャルメディアの場合であれば設定された個人のプロパティ情報の分析が主流である。
- 一方で、閲覧者の存在の重要性は認識されており、その行動分析が試みられている。
- 閲覧者の分析は、①（登録制のソーシャルメディアの場合のみ）設定された閲覧者の個人プロパティ情報を分析する、②（登録制のソーシャルメディアの場合のみ）「閲覧者数 = 登録者数 - 投稿した人数」、として量的分析を行う、③「投稿が少ない人」を「閲覧者」と定義して分析する、④アンケート調査結果を使用する、⑤検索エンジン結果を使用する、が主流である。

一方、先行研究に共通する問題としては、以下の3点がある。

- 主流である投稿された情報の分析では、作為的な投稿者による投稿内容の影響を受けやすいという“作為的投稿影響問題”を抱えている。
- 閲覧者の分析では、ソーシャルメディア利用者に調査票を記入させ回収する必要があるため、大規模な調査には不向きであるという“調査規模的問題”を抱えている。
- 閲覧者の分析では、調査票の記入者を選択する際や調査票を記入する際、または検索エンジンを利用して検索結果を得る際に発生し得る、サンプリングバイアス

や反応バイアスの影響を受けてしまうという“利用者バイアス問題”を抱えている。

以上の問題を解決するために、本研究では以下の提案を行った。

- 何人の閲覧者がいたかという情報に対応する DNS クエリデータの“異なり数”を用いて、ソーシャルメディアの“閲覧者側”の視点から、利用者の行動を分析する手法（“作為的投稿影響問題”への対応）。
- DNS の中でも各国ごとに割り当てられた ccTLD の 1 つである JP ドメイン名の DNS クエリデータを用いるによる、大規模で即時的なデータ収集の実現（“調査規模的問題”への対応）。
- インターネットの基幹技術である DNS を用いてデータ収集することで、ソーシャルメディア利用者に対する“利用者バイアス問題”への対応

本研究では、DNS クエリデータを用い閲覧者側から分析することで、実社会で起きる現象を分析する際に発生し得る“作為的投稿影響問題”や“利用者バイアス問題”，“調査規模的問題”に対応することを試みた。また、大規模な実 DNS クエリデータを用いて以下の 2 つの実証実験を行い、提案手法の有効性を確認した。

#### (1) ソーシャルメディアを利用した予測に関する研究

提案手法をソーシャルメディア利用者の実社会における行動分析に適用し、先行研究が抱える問題に対応することを確認するために映画の観客動員数予測を試みた。その結果、投稿者側からの視点である従来手法（[Mishne 2006]）が映画公開前で相関係数 0.454～0.542 であったのに対し、閲覧者側からの視点である提案手法は 0.865 であることを示した。これにより、DNS クエリデータを用い、投稿者側の視点からではなく閲覧者側の視点から、ソーシャルメディア利用者の実社会における行動分析が可能であり、ソーシャルメディアを利用した予測に関する研究領域において提案手法の有効性を確認した。

#### (2) ソーシャルメディアを利用した計測・モデル化に関する研究

提案手法を応用し、より大規模で複雑なソーシャルメディア利用者の実社会における行動分析を行うことで、提案手法が大規模で複雑な行動分析も可能であり先行研究の抱える

問題に対応できることを確認し、そのソーシャルメディア利用者の行動の時間的推移も明らかにした。

具体的には、日本政府が3年に渡り主導した“地上波デジタル放送への移行”と、それを後押しする政策の1つであった“家電エコポイント制度の実施”という実社会における2つの大きな社会現象に対し、DNSクエリデータにもとづき家電エコポイント数の発行数とソーシャルメディア利用の関連を計測することで、ソーシャルメディア利用者の行動の時間的推移を分析した。

この結果、日本政府が行った2009年から2011年に及ぶアナログ放送の停止とデジタル放送への移行という政策において、家電エコポイント制度が、その目的の1つであった地上デジタル放送対応テレビの普及で、ソーシャルメディア利用者の行動に効果的な影響を与えた経済政策であったことを示した。具体的には、地上デジタル放送対応テレビに対する付与エコポイント数をコントロールする施策により、ソーシャルメディア利用者の行動において、テレビ放送全体に対しての行動へ広がる傾向を示しており、地上デジタル放送対応への切り替えを加速させる傾向を示した。これにより、大規模なDNSクエリデータを用いて、閲覧者側の視点からソーシャルメディア利用者の実社会における行動の計測が可能であり、ソーシャルメディアを利用した計測・モデル化に関する研究領域において、提案手法が大規模で複雑な実社会で起きる現象の分析も可能であり、先行研究の抱える問題に対応できることを確認した。

## 6.2. 今後の展望

本研究では、ソーシャルメディア利用者の行動を把握する上で客観的なデータ計測を可能とし、実社会で起きる現象を分析する手法を提案した。第2章の先行研究で示したように、近年のソーシャルメディア利用者の行動に関する研究は、TwitterやFacebookなどのソーシャルメディア利用者を対象とした研究が盛んである。それらは、インターネット上におけるソーシャルメディアという仮想空間の中で、その利用者がどのような行動をしているかを分析し、実社会で起きる現象との接点を見出そうと試みる研究である。

一方、インターネットの利用方法は多様化が進んでおり、また、インターネットで提供されるサービスも多様化の流れを踏まえ、独自性を打ち出したものも数多く出現してきて

いる中で、ソーシャルメディア利用者の選択肢も広がってきている。その様な潮流の中で、ソーシャルメディア利用者の行動は、インターネット上での単独のサービス内で完結するとは言えない状況になってきている。

本研究で用いた DNS は、インターネットにおいてサービスを利用するには必ず必要となる仕組みである。つまり DNS は、インターネット上において多様化が進む多くのサービスと関わりが深い仕組みであり、ソーシャルメディア利用者に対して、それら多くのサービスを横断的に結び付けていく仕組みである。本稿では、その DNS にもとづく提案手法を、まず広告ブランドサイトや政府関連サイトなどの複数のソーシャルメディアに適用し、その利用者の行動分析から実社会で起きる現象との接点を見出すことを試みた。本研究の今後の展開可能性としては、提案手法を拡張し、競争激化している SNS やオンラインゲームサービス、iTunes や Google Play などのプラットフォームサイト、更には Google や Yahoo!、Bing などの検索エンジンサイトなど利用用途の異なる多様性を持ったインターネットサービス間に跨るソーシャルメディア利用者の行動分析に適用し、実社会で起きるより複雑で多様性を持った現象との関係を明らかにしていくことで、実社会で行動するより多くの人々に対して、インターネット利用によるより豊かな実社会の実現と、インターネットサービスの更なる発展に貢献できると考えている。

# 謝辞

本研究を進めるにあたっては、多くの方々のご指導、ご協力、ご支援を賜りました。ここに、お世話になった皆様への感謝の意を表します。

まず、修士課程から6年間の長きにわたり、主指導教員として、論文指導、研究発表指導から研究テーマの方向性に関する助言に至るまで、弛まぬ指導、助言をいただいた筑波大学大学院ビジネス科学研究科 吉田 健一教授に深く感謝いたします。副指導教員、筑波大学大学院ビジネス科学研究科 山田 雄二教授からは、金融工学の立場から研究テーマの発展の可能性についての様々な助言をいただきました。副指導教員、筑波大学大学院ビジネス科学研究科 倉橋 節也准教授からは、複雑システム論の立場から研究テーマの発展の可能性についての様々な助言をいただきました。両教員に改めて深謝の意を表します。

また、発表会や国際会議前などの演習の場などでは、筑波大学大学院ビジネス科学研究科教員の方々や吉田研究室の皆様には貴重なご意見を頂戴しました。ここに深く感謝致します。

更に、本学在学の長きにわたり、仕事との両立において理解と励まし、支援を継続的にしていただいた株式会社 日本レジストリサービス 佐野晋代表取締役副社長に深く感謝致します。

最後に、就学前の学位取得に挑戦するか否かで悩んでいた時から、遅々として研究が進まず何度も挫折しそうになった時、先が見えてきて研究が形になってきた時など、常に心の支えになり、そして共に喜んでくれた妻 葉子と娘 果林に心より感謝します。



## 参考文献

- [Acquisti 2009] Alessandro Acquisti and Ralph Gross. "Predicting Social Security numbers from public data." *Proceedings of the National academy of sciences* 106.27, 2009: 10975-10980.
- [Adamic 2008] Lada A. Adamic, et al. "Knowledge sharing and yahoo answers: everyone knows something." *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.
- [Adler 2006] Thomas B. Adler and Luca De Alfaro. "A content-driven reputation system for the Wikipedia." *Proceedings of the 16th international conference on World Wide Web*, ACM, 2006.
- [Agichtein 2008] Eugene Agichtein, et al. "Finding high-quality content in social media." *Proceedings of the international conference on Web search and web data mining*. ACM, 2008.
- [Akioka 2010] Sayaka Akioka, et al. "Cross-media impact on twitter in japan." *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM, 2010.
- [荒井 2004] 荒井功, et al. "社会を<モデル>でみる——数理社会学への招待." *社会を<モデル>でみる——数理社会学への招待*, 2004: 128-131.
- [Arthur 2006] Charles Arthur. "What is the 1% rule?." *The Guardian*, 2006.
- [Askitas 2009] Nikos Askitas and Klaus Zimmermann. "Google econometrics and unemployment forecasting." *Appl Econ Q* 55(2), 2009:107–20.
- [Asur 2010] Sitaram Asur and Bernardo A. Huberman. "Predicting the future with social media." *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, IEEE, 2010.

- [Awan 2007] Akil N. Awan. "Virtual jihadist media Function, legitimacy and radicalizing efficacy." *European Journal of Cultural Studies* 10.3, 2007: 389-408.
- [Backstrom 2011] Lars, Backstrom and Jure Leskovec. "Supervised random walks: predicting and recommending links in social networks." *Proceedings of the 4th ACM international conference on Web search and data mining*. ACM, 2011.
- [Baeza-Yates 2006] Ricardo Baeza-Yates, Liliana Calderón-Benavides, and Cristina González-Caro. "The intention behind web queries." *Proceedings of the 13th international conference on String Processing and Information Retrieval*, 2006.
- [Bakshy 2009] Eytan Bakshy, Brian Karrer, and Lada A. Adamic. "Social influence and the diffusion of user-created content." *Proceedings of the 10th ACM conference on Electronic commerce*. ACM, 2009.
- [Benevenuto 2010]0 Fabricio Benevenuto, et al. "Detecting spammers on twitter." *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference*, 2010.
- [Bernstein 2013] Michael S. Bernstein, et al. "Quantifying the invisible audience in social networks." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2013.
- [Berthon 2012] Pierre R. Berthon, et al. "Marketing meets Web 2.0, social media, and creative consumers: Implications for international marketing strategy." *Business Horizons* 55.3, 2012: 261-271.
- [Bian 2008] Jiang Bian, et al. "Finding the right facts in the crowd: Factoid question answering over social media." *Proceedings of the 17th international conference on World Wide Web*. ACM, 2008.
- [Bollen 2011] Johan Bollen, Huina Mao and Xiaojun Zeng. "Twitter mood predicts the stock market." *Journal of Computational Science* 2.1, 2011: 1-8.



- [Boyd 2010] Danah Boyd, Scott Golder, and Gilad Lotan. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." Proceedings of the 43rd Hawaii International Conference on System Sciences, IEEE, 2010.
- [Brandtzaeg 2008] Petter B. Brandtzaeg and Jan Heim. "User loyalty and online communities: why members of online communities are not faithful." Proceedings of the 2nd International Conference on Intelligent Technologies for Interactive Entertainment, ICST, 2008.
- [Brandtzaeg 2011] Petter B. Brandtzaeg and Jan Heim. "A typology of social networking sites users." International Journal of Web Based Communities 7.1, 2011: 28-51.
- [Brown 2012] Eileen Brown. Working the crowd: Social media marketing for business. British Infomatics Society, 2012.
- [Brown 2013] Danny Brown. Influence Marketing: How to Create, Manage, and Measure Brand Influencers in Social Media Marketing. Que Publishing, 2013.
- [Brown 2002] Shona L. Brown, Tilton Andrew, and Dennis M. Woodside. "The case for on-line communities." The McKinsey Quarterly 1.1, 2002.
- [Burcher 2012] Nick Burcher. Paid, Owned, Earned: Maximizing Marketing Returns in a Socially Connected World. Kogan Page Publishers, 2012.
- [Burke 2011] Moira Burke, Robert Kraut, and Cameron Marlow. "Social capital on Facebook: Differentiating uses and users." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2011.
- [Castillo 2011] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. "Information credibility on twitter." Proceedings of the 20th international conference on World Wide Web. ACM, 2011.
- [Castro 2006] Alejandra Castro, and Carlos Rodriguez-Sickert. "The effect of social interactions in the primary consumption life cycle of motion pictures." New Journal of Physics 8.4, 2006: 52.

- [Cha 2009] Meeyoung Cha, Alan Mislove, and Krishna P. Gummadi. "A measurement-driven analysis of information propagation in the flickr social network." Proceedings of the 18th international conference on World Wide Web. ACM, 2009.
- [Cha 2010] Meeyoung Cha, et al. "Measuring User Influence in Twitter: The Million Follower Fallacy." Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, 2010.
- [Chang 2004] Yuhmiin Chang, and Esther Thorson. "Television and web advertising synergies." Journal of Advertising 33.2, 2004: 75-84.
- [Cheong 2009] Marc Cheong and Vincent Lee. "Integrating web-based intelligence retrieval and decision-making from the twitter trends knowledge base." Proceedings of the 2nd ACM workshop on Social web search and mining. ACM, 2009.
- [Choi 2012] Hyunyoung Choi, and Hal Varian. "Predicting the present with google trends." Economic Record 88.s1, 2012: 2-9.
- [Chowdhury 2007] Rafi M.M.I Chowdhury, Adam Finn, and Douglas Olsen. "Investigating the simultaneous presentation of advertising and television programming." Journal of Advertising 36.3, 2007: 85-96.
- [Cobb 2010] Nathan K. Cobb, Amanda L. Graham, and David B. Abrams. "Social network structure of a large online community for smoking cessation." American Journal of Public Health 100.7, 2010: 1282-1289.
- [Cohn 2004] Michael A. Cohn, Matthias R. Mehl, and James W. Pennebaker. "Linguistic markers of psychological change surrounding September 11, 2001." Psychological Science 15.10, 2004: 687-693.
- [Corcoran 2009] Sean Corcoran. "Defining earned, owned and paid media." Forrester Research, 2009.
- [Danaher 2011] Peter J. Danaher, and John R. Rossiter. "Comparing perceptions of marketing communication channels." European Journal of Marketing 45.1/2, 2011: 6-42.

- [Denning 2005] Peter Denning, et al. "Wikipedia risks." *Communications of the ACM* 48.12, 2005: 152-152.
- [Dijkstra 2005] Majorie Dijkstra, Heidi EJJM Buijtels, and Fred van Raaij. "Separate and joint effects of medium type on consumer responses: a comparison of television, print, and the Internet." *Journal of Business Research* 58.3, 2005: 377-386.
- [Duan 2008] Wenjing Duan, Bin Gu, and Andrew B. Whinston. "Do online reviews matter?—An empirical investigation of panel data." *Decision Support Systems* 45.4, 2008: 1007-1016.
- [DuBoi 2011] Thomas DuBois, Jennifer Golbeck, and Aravind Srinivasan. "Predicting trust and distrust in social networks." *Proceedings of the 3rd IEEE International Conference on Social Computing*, IEEE, 2011.
- [Dupret 2006] Georges E. Dupret, and Benjamin Piwowarski. "A user browsing model to predict search engine click data from past observations." *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008.
- [Dwyer 2007] Catherine Dwyer, Starr Roxanne Hiltz, and Katia Passerini. "Trust and Privacy Concern within Social Networking Sites: A Comparison of Facebook and MySpace." *Proceedings of the 13th Americas Conference on Information Systems*. 2007.
- [Edelman 2010] David Edelman, and Brian Salsberg. "Beyond paid media: Marketing's new vocabulary." *McKinsey Quarterly*, 2010: 1-8.
- [Ellison 2007] Nicole B. Ellison, Charles Steinfield, and Cliff Lampe. "The benefits of Facebook "friends:" Social capital and college students' use of online social network sites." *Journal of Computer - Mediated Communication* 12.4, 2007: 1143-1168.
- [Fogel 2009] Joshua Fogel, and Elham Nehmad. "Internet social network communities: Risk taking, trust, and privacy concerns." *Computers in Human Behavior* 25.1, 2009: 153-160.

- [Fogg 2001] B. J. Fogg, et al. "What makes Web sites credible?: a report on a large quantitative study." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2001.
- [Gallagher 2001] Katherine Gallagher, Dale K. Foster, and Jeffrey Parsons. "The medium is not the message: advertising effectiveness and content evaluation in print and on the Web." Journal of Advertising Research 41.4, 2001: 57-70.
- [Ghonim 2011] Wael Ghonim. "Inside the Egyptian revolution." TED, March (2011).
- [Gilbert 2009] Eric Gilbert and Karrie Karahalios. "Predicting tie strength with social media." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2009.
- [Gilbert 2012] Eric Gilbert. "Predicting tie strength in a new medium." Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work. ACM, 2012.
- [Ginsberg 2008] Jeremy Ginsberg, et al. "Detecting influenza epidemics using search engine query data." Nature 457.7232, 2008: 1012-1014.
- [Goel 2010] Sharad Goel, et al. "Predicting consumer behavior with Web search." Proceedings of the National Academy of Sciences 107.41, 2010: 17486-17490.
- [Goh 2013] Khim-Yong Goh, Cheng-Suang Heng, and Zhijie Lin. "Social media brand community and consumer behavior: Quantifying the relative impact of user- and marketer-generated content." Information Systems Research 24.1, 2013: 88-107.
- [Golbeck 2009] Jennifer Golbeck. "Trust and nuanced profile similarity in online social networks." ACM Transactions on the Web (TWEB) 3.4, 2009: 12.
- [Golbeck 2011] Jennifer Golbeck, Robles Cristina, and Turner Karen. "Predicting personality with social media." Proceedings of the 29th ACM Conference on Human Factors in Computing Systems. ACM, 2011.

- [Golder 2011] Scott A. Golder, and Michael W. Macy. "Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures." *Science* 333.6051, 2011: 1878-1881.
- [Gomez 2010] Manuel Gomez-Rodriguez, Jure Leskovec, and Andreas Krause. "Inferring networks of diffusion and influence." Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2010.
- [Grinev 2009] Maxim Grinev, et al. "Sifting micro-blogging stream for events of user interest." Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2009.
- [Gruhl 2004] Daniel Gruhl, et al. "Information diffusion through blogspace." Proceedings of the 13th international conference on World Wide Web. ACM, 2004.
- [Gruhl 2005] Daniel Gruhl, et al. "The predictive power of online chatter." Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM, 2005.
- [Guo 2011] Stephen Guo, Mengqiu Wang, and Jure Leskovec. "The role of social networks in online shopping: information passing, price of trust, and consumer choice." Proceedings of the 12th ACM conference on Electronic commerce. ACM, 2011.
- [Guzella 2009] Thiago S. Guzella, and Walmir M. Caminhas. "A review of machine learning approaches to spam filtering." *Expert Systems with Applications* 36.7, 2009: 10206-10222.
- [Haklay 2008] Mordechai Haklay, and Patrick Weber. "Openstreetmap: User-generated street maps." *Pervasive Computing, IEEE* 7.4, 2008: 12-18.
- [Hampton 2011] Keith Hampton, et al. "Social networking sites and our lives." Technical report, Pew Internet & American Life Project, 2011.

- [Hanna 2011] Richard Hanna, Andrew Rohm, and Victoria L. Crittenden. "We're all connected: The power of the social media ecosystem." *Business Horizons* 54.3, 2011: 265-273.
- [Hollenstein 2013] Livia Hollenstein, and Ross Purves. "Exploring place through user-generated content: Using Flickr tags to describe city cores." *Journal of Spatial Information Science* 1, 2013: 21-48.
- [本田 2012] 本田哲也. ソーシャルインフルエンス 戦略 PR× ソーシャルメディアの設計図. アスキー・メディアワークス, 2012.
- [Honey 2009] Courtenay Honey, and Susan C. Herring. "Beyond microblogging: Conversation and collaboration via Twitter." *Proceedings of the 42nd Hawaii International Conference on System Sciences*. IEEE, 2009.
- [Hong 2011] Liangjie Hong, Ovidiu Dan and Brian D. Davison. "Predicting popular messages in twitter." *Proceedings of the 20th international conference companion on World Wide Web*. ACM, 2011.
- [Horowitz 2006] Bradley Horowitz. "Creators, synthesizers, and consumers." URL: <http://www.elatable.com/blog> (2006).
- [Jacoby 1990] Jacob Jacoby, and Wayne D. Hoyer. "The Miscomprehension of Mass-Media Advertising Claims: A Re-Analysis of Benchmark Data." *Journal of Advertising Research* 30.3, 1990: 9-16.
- [Java 2007] Akshay Java, et al. "Why we twitter: understanding microblogging usage and communities." *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*. ACM, 2007.
- [Jindal 2008] Nitin Jindal, and Bing Liu. "Opinion spam and analysis." *Proceedings of the international conference on Web search and web data mining*. ACM, 2008.
- [Jindal 2010] Nitin Jindal, Bing Liu, and Ee-Peng Lim. "Finding unusual review patterns using unexpected rules." *Proceedings of the 19th ACM international conference on Information and knowledge management*. ACM, 2010.

- [Jung 2013] Yumi Jung, et al. "Favors from facebook friends: unpacking dimensions of social capital." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2013.
- [Kanai 2012] Ryota Kanai, et al. "Online social network size is reflected in human brain structure." Proceedings of the Royal Society B: Biological Sciences 279.1732, 2012: 1327-1334.
- [Kaplan 2010] Andreas M. Kaplan, and Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media." Business horizons 53.1, 2010: 59-68.
- [経済産業省 2011]  
<http://www.meti.go.jp/statistics/toppage/report/bunseki/pdf/h23/h4a1106j2.pdf>
- [Keys 2005] Ken Keys, David Moore, and Cristian Estan. "A robust system for accurate real-time summaries of internet traffic." ACM SIGMETRICS Performance Evaluation Review. Vol. 33. No. 1. ACM, 2005.
- [Kholodilin 2009] Konstantin Kholodilin, et al. "Google searches as a means of improving the nowcasts of key macroeconomic variables." DIW Berlin Discussion Paper No. 949, 2009.
- [Kietzmann 2011] Jan H. Kietzmann, et al. "Social media? Get serious! Understanding the functional building blocks of social media." Business Horizons 54.3, 2011: 241-251.
- [Kietzmann 2012] Jan H. Kietzmann, et al. "Unpacking the social media phenomenon: towards a research agenda." Journal of Public Affairs 12.2, 2012: 109-119.
- [Kim 2004] Myung-Sup Kim, et al. "A flow-based method for abnormal network traffic detection." Network Operations and Management Symposium, 2004. NOMS 2004. IEEE/IFIP. Vol. 1. IEEE, 2004.
- [Kittur 2006] Aniket Kittur, et al. "He says, she says: conflict and coordination in Wikipedia." Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, 2007.

- [Kittur 2008] Aniket Kittur, Bongwon Suh and Ed H. Chi. "Can you ever trust a wiki?: impacting perceived trustworthiness in wikipedia." Proceedings of the 2008 ACM conference on Computer supported cooperative work. ACM, 2008.
- [Kolari 2006] Pranam Kolari, Akshay Java, and Tim Finin. "Characterizing the splogosphere." Proceedings of the 3rd annual workshop on weblogging ecosystem: Aggregation, analysis and dynamics, 15th World Wid Web conference, 2006.
- [Krishnamurthy 2008] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. "A few chirps about twitter." Proceedings of the First Workshop on Online Social Networks.. ACM, 2008.
- [Krugman 1965] Herbert E. Krugman. "The impact of television advertising: Learning without involvement." Public opinion quarterly 29.3 (1965): 349-356.
- [Krugman 1971] Herbert E. Krugman,. "Brain wave measures of media involvement." Journal of Advertising Research 11.1, 1971: 3-9.
- [Kurashima 2006] Takeshi Kurashima, Taro Tezuka, and Katsumi Tanaka. "Mining and visualizing local experiences from blog entries." Database and Expert Systems Applications. Springer Berlin Heidelberg, 2006.
- [Kwak 2010] Haewoon Kwak, et al. "What is Twitter, a social network or a news media?." Proceedings of the 19th international conference on World Wide Web. ACM, 2010.
- [Lakhani 2003] Karim R. Lakhani, and Eric V. Hippel. "How open source software works:“free” user-to-user assistance." Research policy 32.6, 2003: 923-943.
- [Lakhina 2005] Anukool Lakhina, Mark Crovella and Christophe Diot. "Mining anomalies using traffic feature distributions." Proceedings of ACM SIGCOMM Computer Communication Review. Vol. 35. No. 4. ACM, 2005.



- [Lampe 2010] Cliff Lampe, et al. "Motivations to participate in online communities." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010.
- [Laroche 2012] Michel Laroche, et al. "The effects of social media based brand communities on brand community markers, value creation practices, brand trust and brand loyalty." Computers in Human Behavior 28.5, 2012: 1755-1767.
- [Lerman 2010] Kristina Lerman, and Rumi Ghosh. "Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks." Proceedings of the 4th International Conference on Weblogs and Social Media, 2010: 90-97.
- [Leskovec 2008] Jure Leskovec and Eric Horvitz. "Planetary-scale views on a large instant-messaging network." Proceedings of the 17th international conference on World Wide Web. ACM, 2008.
- [Leskovec 2010] Jure Leskovec, Daniel Huttenlocher and Jon Kleinberg. "Predicting positive and negative links in online social networks." Proceedings of the 19th international conference on World Wide Web. ACM, 2010.
- [Lewis 2008] Kevin Lewis, et al. "Tastes, ties, and time: A new social network dataset using Facebook.com." Social networks 30.4, 2008: 330-342.
- [Li 2008] Xin Li, Lei Guo and Yihong E. Zhao. "Tag-based social interest discovery." Proceedings of the 17th international conference on World Wide Web. ACM, 2008.
- [Lim 2010] Ee-Peng Lim, et al. "Detecting product review spammers using rating behaviors." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.
- [Lin 2007] Yu-Ru Lin, et al. "Splog detection using self-similarity analysis on blog temporal dynamics." Proceedings of the 3rd international workshop on Adversarial information retrieval on the web. ACM, 2007.

- [Litman 1998] Barry R. Litman. *The motion picture mega-industry*. Needham Heights, MA: Allyn and Bacon, 1998
- [Liu 2006] Yong Liu. "Word of mouth for movies: Its dynamics and impact on box office revenue." *Journal of marketing*, 2006: 74-89.
- [Liu 2007] Yang Liu, et al. "ARSA: a sentiment-aware model for predicting sales performance using blogs." *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2007.
- [Liu 2008] Yandong Liu, Jiang Bian, and Eugene Agichtein. "Predicting information seeker satisfaction in community question answering." *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 2008.
- [Liu 2010] Bing Liu. "Sentiment analysis and subjectivity." *Handbook of natural language processing 2*, 2010: 568.
- [Longueville 2009] Bertrand De Longueville, Robin S. Smith, and Gianluca Luraschi. "OMG, from here, I can see the flames!: a use case of mining location based social networks to acquire spatio-temporal data on forest fires." *Proceedings of the 2009 International Workshop on Location Based Social Networks*. ACM, 2009.
- [McLaren 2011] Nick McLaren and Rachana Shanbhogue. "Using internet search data as economic indicators." *Bank of England Quarterly Bulletin* 2011.
- [Mendoza 2010] Marcelo Mendoza, Barbara Poblete, and Carlos Castillo. "Twitter Under Crisis: Can we trust what we RT?." *Proceedings of the first workshop on social media analytics*. ACM, 2010.
- [Middleton 2012] Daina Middleton. *Marketing in the Participation Age: A Guide to Motivating People to Join, Share, Take Part, Connect, and Engage*. John Wiley & Sons, 2012.

- [Mishne 2006] Gilad Mishne and Natalie S. Glance. "Predicting Movie Sales from Blogger Sentiment." AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs. 2006.
- [Mislove 2008] Alan Mislove, et al. "Growth of the flickr social network." Proceedings of the First Workshop on Online Social Networks. ACM, 2008.
- [三田村 2010] 三田村健史, and 吉田健一. "DNS クエリデータに基づくコンテンツへの関心度分析 (< 特集> 社会基盤としてのインターネットアーキテクチャ論文)." 電子情報通信学会論文誌. B, 通信 93.10, 2010: 1368-1377.
- [Mitamura 2012] Takeshi Mitamura and Kenichi Yoshida. "Viewers' side analysis of social interests." Proceedings of the 12th IEEE International Conference on Data Mining Workshops, IEEE, 2012.
- [Mitamura 2013] Takeshi Mitamura and Kenichi Yoshida, "Analyzing People's Behavior Using Network Data." Journal of the Japan Society for Management Information 22.3, 2013. [In Printing]
- [Mori 2005] Tatsuya Mori, et al. "Detection of worm-infected hosts by communication pattern analysis." IEIC Technical Report (Institute of Electronics, Information and Communication Engineers) 105.405, 2005: 13-18.
- [Muller 2010] Michael Muller, et al. "We are all lurkers: consuming behaviors among authors and readers in an enterprise file-sharing service." Proceedings of the 16th ACM international conference on supporting group work. ACM, 2010.
- [Muller 2012] Michael Muller. "Lurking as personal trait or situational disposition: lurking and contributing in enterprise social media." Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, 2012.
- [Muniz 2001] Albert M. Muniz Jr, and Thomas C. O'guinn. "Brand community." Journal of consumer research 27.4, 2001: 412-432.

- [Nagel 2009] Lynette Nagel, Seugnet A. Blignaut and Johannes C. Cronjé. "Read-only participants: a case for student communication in online classes." *Interactive Learning Environments* 17.1, 2009: 37-51.
- [Naik 2009] Prasad A. Naik and Kay Peters. "A hierarchical marketing communications model of online and offline media synergies." *Journal of Interactive Marketing* 23.4, 2009: 288-299.
- [Neis 2012] Pascal Neis and Alexander Zipf. "Analyzing the contributor activity of a volunteered geographic information project - The case of OpenStreetMap." *ISPRS International Journal of Geo-Information* 1.2, 2012: 146-165.
- [Neylon 2009] Cameron Neylon and Shirley Wu. "Article-level metrics and the evolution of scientific impact." *PLoS biology* 7.11, 2009: e1000242.
- [NHK 2011] NHK Corporate Info, "On one month since the switchover to digital broadcasting," Summary of Press Conference, Sep. 2011, [http://www.nhk.or.jp/pr/english/toptalk/kaichou\\_e/k\\_e1109.htm](http://www.nhk.or.jp/pr/english/toptalk/kaichou_e/k_e1109.htm) (16 Apr.2012)
- [Nielsen 2006a] Jakob Nielsen. "Participation inequality: Encouraging more users to contribute." *Jakob Nielsen's alertbox* 9, 2006.
- [Nielsen 2006b] Jakob Nielsen. "Participation inequality: lurkers vs. contributors in internet communities." *Jakob Nielsen's Alertbox*, 2006.
- [Nielsen 2013] Jakob Nielsen. "Participation inequality: Internet Activity Bias Causes Lumpy User Behavior." *Jakob Nielsen's Alertbox*, 2013.
- [Nonnecke 2000] Blair Nonnecke and Jenny Preece. "Lurker demographics: Counting the silent." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2000.
- [Nonnecke 2004] Blair Nonnecke, et al. "Online Lurkers Tell Why." *Proceedings of the 10th Americas Conference on Information Systems, ICIS*, 2004.
- [Nonnecke 2006] Blair Nonnecke, Dorine Andrews, and Jenny Preece. "Non-public and public online community participation: Needs, attitudes and behavior." *Electronic Commerce Research* 6.1, 2006: 7-20.

- [Ott 2011] Myle Ott, et al. "Finding deceptive opinion spam by any stretch of the imagination." Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011.
- [Parvanta 2013] Claudia Parvanta, Yannig Roth and Heidi Keller. "Crowdsourcing 101 A Few Basics to Make You the Leader of the Pack." Health promotion practice 14.2, 2013: 163-167.
- [Peng 2004] Tao Peng, Christopher Leckie and Kotagiri Ramamohanarao. "Proactively detecting distributed denial of service attacks using source IP address monitoring." Proceedings of the Third International IFIP-TC6 Networking Conference, 2004: 771-782.
- [Pergelova 2010] Albena Pergelova, Diego Prior and Josep Rialp. "Assessing advertising efficiency." Journal of Advertising 39.3, 2010: 39-54.
- [Phillip 2012] Brador Philip "Rigged online food reviews should come as no surprise", <http://www.japantimes.co.jp/text/fd20120115pb.html> (16 Apr.2012)
- [Polgreen 2008] Philip M. Polgreen, et al. "Using internet searches for influenza surveillance." Clinical infectious diseases 47.11, 2008: 1443-1448.
- [Proulx 2012] Mike Proulx and Stacey Shepatin. "Social TV: how marketers can reach and engage audiences by connecting television to the web, social media, and mobile." Wiley.com, 2012.
- [Quercia 2012] Daniele Quercia, et al. "The personality of popular facebook users." Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. ACM, 2012.
- [Rafaeli 2004] Sheizaf Rafaeli, Gilad Ravid, and Vladimir Soroka. "De-lurking in virtual communities: a social communication network approach to measuring the effects of social and cultural capital." Proceedings of the 37th Annual Hawaii International Conference on. IEEE, 2004.
- [Resnick 2000] Paul Resnick, et al. "Reputation systems." Communications of the ACM 43.12, 2000: 45-48.

- [Richter 2013] Alexander Richter, et al. "Success Measurement of Enterprise Social Networks." *Wirtschaftsinformatik*. 2013.
- [Rodgers 2005] Shelly Rodgers. "Intermedia Effects for Appropriate/ Inappropriate Print and Internet Stimuli." *Advertising, promotion, and new media*, 2005.
- [Romero 2011] Daniel M. Romero, et al. "Influence and passivity in social media." *Machine learning and knowledge discovery in databases*. Springer Berlin Heidelberg, 2011: 18-33.
- [Sawhney 1996] Mohanbir S. Sawhney and Jehoshua Eliashberg. "A parsimonious model for forecasting gross box-office revenues of motion pictures." *Marketing Science* 15.2, 1996: 113-131.
- [Schifanella 2010] Rossano Schifanella, et al. "Folks in folksonomies: social link prediction from shared metadata." *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010.
- [Schmidt 2011] Simeon Vosen, and Torsten Schmidt. "Forecasting private consumption: survey - based indicators vs. Google trends." *Journal of Forecasting* 30.6, 2011: 565-578.
- [Scott 1910] Walter D. Scott. "The psychology of advertising: A simple exposition of the principles of psychology in their relation to successful advertising." Small, Maynard, 1910.
- [Shamma 2009] David A. Shamma, Kennedy Lyndon, and Churchill, Elizabeth F. "Tweet the debates: understanding community annotation of uncollected sources." *Proceedings of the first SIGMM workshop on Social media*. ACM, 2009.
- [Shirky 2011] Clay Shirky. "Political Power of Social Media-Technology, the Public Sphere Sphere, and Political Change, The. *Foreign Aff.* 90, 2011: 28.
- [Shomura 2008] Yusuke Shomura, Yoshinori Watanabe and Kenichi Yoshida. "Analyzing the number of varieties in frequently found flows." *IEICE transactions on communications* 91.6, 2008: 1896-1905.

- [Smith 2012] Andrew N. Smith, Eileen Fischer, and Chen Yongjian. "How does brand-related user-generated content differ across YouTube, Facebook, and Twitter?." *Journal of Interactive Marketing* 26.2, 2012: 102-113.
- [Speath 2008] Sebastian Spaeth, et al. "Communal resources in open source software development." *Information Research* 13.1, 2008: 2.
- [Stammerjohan 2005] Claire Stammerjohan, et al. "An empirical investigation of the interaction between publicity, advertising, and previous brand attitudes and knowledge." *Journal of Advertising* 34.4 (2005): 55-67.
- [Starch 1923] Daniel Starch. "Principles of Advertising, 1923; Testing the effectiveness of advertisements." *Harvard Business Review* 2, 1923: 464-475.
- [Steinfeld 2008] Charles Steinfield, Nicole B. Ellison and Cliff Lampe. "Social capital, self-esteem, and use of online social network sites: A longitudinal analysis." *Journal of Applied Developmental Psychology* 29.6, 2008: 434-445.
- [Stephen 2012] Andrew T. Stephen, and Jeff Galak. "The effects of traditional and social earned media on sales: A study of a microlending marketplace." *Journal of Marketing Research* 49.5, 2012: 624-639.
- [Stewart 2010] Osamuyimen Stewart, David Lubensky and Juan M. Huerta. "Crowdsourcing participation inequality: a SCOUT model for the enterprise domain." *Proceedings of the ACM SIGKDD Workshop on Human Computation*. ACM, 2010.
- [Sundar 1998] Shyam Sundar, et al. "Does web advertising work? Memory for print vs. online media." *Journalism & Mass Communication Quarterly* 75.4, 1998: 822-835.
- [Suryanto 2009] Maggy A. Suryanto, et al. "Quality-aware collaborative question answering: methods and evaluation." *Proceedings of the 2nd ACM International Conference on Web Search and Data Mining*. ACM, 2009.

- [Susaria 2012] Anjana Susarla, Jeong-Ha Oh, and Yong Tan. "Social networks and the diffusion of user-generated content: Evidence from YouTube." *Information Systems Research* 23.1, 2012: 23-41.
- [Suzuki 2010] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. "Earthquake shakes Twitter users: real-time event detection by social sensors." *Proceedings of the 19th international conference on World Wide Web*. ACM, 2010.
- [Szabo 2010] Gabor Szabo, and Bernardo A. Huberman. "Predicting the popularity of online content." *Communications of the ACM* 53.8, 2010: 80-88.
- [Tumarkin 2001] Robert Tumarkin and Robert F. Whitelaw. "News or noise? Internet postings and stock prices." *Financial Analysts Journal*, 2001: 41-51.
- [Tumasjan 2010] Andranik Tumasjan, et al. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment." *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, 2010.
- [上原 2008] 上原宏, 佐藤忠彦, and 吉田健一. "インターネット・コミュニティ・データを使ったテレビ CM の商品イメージ形成効果測定." *人工知能学会論文誌* 23.3, 2008: 205-216.
- [Uehara 2005] Hiroshi Uehara and Kenichi Yoshida. "Annotating TV drama based on viewer dialogue-analysis of viewers' attention generated on an Internet bulletin board." *Proceedings of IEEE/IPSJ international symposium on applications and the internet (SAINT2005)*, 2005.
- [Vakratsas 2005] Demetrios Vakratsas,. "Advertising response models with managerial impact: an agenda for the future." *Applied Stochastic Models in Business and Industry* 21.4 - 5, 2005: 351-361.
- [Valenzuela 2009] Sebastián Valenzuela, Namsu Park and Kerk F. Kee. "Is There Social Capital in a Social Network Site?: Facebook Use and College Students' Life Satisfaction, Trust, and Participation1." *Journal of Computer - Mediated Communication* 14.4, 2009: 875-901.



- [Vieweg 2010] Sarah Vieweg, et al. "Microblogging during two natural hazards events: what twitter may contribute to situational awareness." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010.
- [Voorveld 2011] Hilde A. Voorveld, Peter C. Neijens, and Edith G. Smit. "Opening the black box: Understanding cross-media effects." *Journal of Marketing Communications* 17.02, 2011: 69-85.
- [Voorveld 2012] Hilde A. Voorveld, Peter C. Neijens and Edith G. Smit. "The interacting role of media sequence and product involvement in cross-media campaigns." *Journal of Marketing Communications* 18.3, 2012: 203-216.
- [Vuong 2008] Ba-Quy Vuong, et al. "On ranking controversies in wikipedia: models and evaluation." Proceedings of the international conference on Web search and web data mining. ACM, 2008.
- [Wang 2006] Alex Wang. "When synergy in marketing communication online enhances audience response: the effects of varying advertising and product publicity messages." *Journal of Advertising Research*, 46.2, 2006: 160-170.
- [Wang 2011] Dashun Wang, et al. "Human mobility, social ties, and link prediction." Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011.
- [Wathen 2002] Nadine Wathen and Jacquelyn Burkell. "Believe it or not: Factors influencing credibility on the Web." *Journal of the American society for information science and technology* 53.2, 2002: 134-144.
- [Weng 2010] Jianshu Weng, et al. "Twitterrank: finding topic-sensitive influential twitterers." Proceedings of the 3rd ACM international conference on Web search and data mining. ACM, 2010.
- [Wilson 2008] Marcella Wilson and Charles Nicholas. "Topological analysis of an online social network for older adults." Proceedings of the 2008 ACM workshop on Search in Social Media. ACM, 2008.

- [Wise 2013] Alyssa F. Wise, et al. "Broadening the notion of participation in online discussions: examining patterns in learners' online listening behaviors." *Instructional Science* 41.2, 2013: 323-343.
- [Yang 2011] Shuang-Hong Yang, et al. "Like like alike: joint friendship and interest propagation in social networks." *Proceedings of the 20th international conference on World Wide Web*. ACM, 2011.
- [Ye 2011] Qiang Ye, et al. "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings." *Computers in Human Behavior* 27.2, 2011: 634-639.
- [Zhao 2005] Qi Zhao, Abhishek Kumar and Jun Xu. "Joint data streaming and sampling techniques for detection of super sources and destinations." *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*. USENIX Association, 2005.

## 関連業績リスト

### 1. 学術論文[査読付き]

- (a) 三田村健史, and 吉田健一. "DNS クエリデータに基づくコンテンツへの関心度分析 (< 特集> 社会基盤としてのインターネットアーキテクチャ論文)." 電子情報通信学会論文誌. B, 通信 93.10 (2010): 1368-1377.
- (b) Takeshi Mitamura and Kenichi Yoshida, "Analyzing People's Behavior Using Network Data." Journal of the Japan Society for Management Information 22.3, 2013. [In Printing]

### 2. 国際会議論文[査読付き]

- (a) Takeshi Mitamura and Kenichi Yoshida. "Viewers' side analysis of social interests." Proceedings of the 12th IEEE International Conference on Data Mining Workshops, IEEE, 2012.

### 3. 国際会議発表

- (a) Takeshi Mitamura and Kenichi Yoshida. "Cardinality in Big Data – Examples in L3&L7 Network." IEICE Technical Committee on Internet Architecture, Oct 2013 [招待講演]

### 4. その他発表

- (a) 三田村健史, "インターネットにおける社会的関心の移り変わり分析", 筑波大学開学 40+110 周年記念事業シンポジウム「情報伝搬のメカニズム分析」, 2013-01

