

# 博士論文

## 非特定テレビ視聴履歴データの放送局間統合および 視聴行動分析に関する研究

松田 裕貴

奈良先端科学技術大学院大学  
先端科学技術研究科  
情報理工学プログラム

主指導教員: 安本 慶一  
ユビキタスコンピューティングシステム研究室 (情報科学領域)

2024年7月30日提出

本論文は奈良先端科学技術大学院大学先端科学技術研究科に  
博士(工学)授与の要件として提出した博士論文である。

松田 裕貴

審査委員：

主査	安本 慶一	(情報科学領域)	教授)
	荒牧 英治	(情報科学領域)	教授)
	諏訪 博彦	(情報科学領域)	准教授)
	松田 裕貴	(情報科学領域)	客員准教授)

# 非特定テレビ視聴履歴データの放送局間統合および 視聴行動分析に関する研究\*

松田 裕貴

## 内容梗概

近年、地上波テレビ広告費はインターネット広告費に追い越され、さらに減少が続いている。テレビのCMはたくさんの視聴者に広告を届けることが可能と言われているが、広告価値の可視化は株式会社ビデオリサーチが集計している視聴率とよばれる統計データによる推計でしか行われていない。そこで、各テレビ放送局において、個人を特定しない形式で、インターネット接続されたテレビから視聴開始時刻や視聴終了時刻等を含む非特定テレビ視聴履歴データを収集し、活用する取り組みが進められている。しかし、各放送局で非特定テレビ視聴履歴データの収集方式やデータ粒度は異なっており、各局が蓄積したデータを統合し、利用することができていないため、非特定テレビ視聴履歴データ利活用によるテレビ広告価値の可視化に至っていない。そこで、本研究では、各局が独自の方式で取得している非特定テレビ視聴履歴データを放送局間で統合する手法を提案し、評価するためのシミュレータ設計と実装を行い、提案手法の評価を行う。また、在阪の4つの放送局の統合視聴データを用いてテレビ広告価値向上や番組制作へのデータ活用を目指した視聴行動分析を行う。非特定テレビ視聴履歴データ統合手法では、各局の視聴履歴データのうち、共通しているIPアドレス・郵便番号・メーカーID・ブラウザメジャーバージョン・ブラウザマイナーバージョンの5項目でテレビ受像機を分離処理する。そして、分離された中でこれらの5項目が一致するテレビのうち、さらにチャンネル遷移時刻が一致するテレビを同一テレビと推定する。また、視聴者行動を再現するシミュレータを設計し、そのシミュレー

---

\*奈良先端科学技術大学院大学 先端科学技術研究科 博士論文, 2024年7月30日.

タから合成された視聴履歴データに対して、本手法を適用した結果、生成された250万台分のデータのうち約241万台のテレビIDのマッチングに成功し、再現率96.5%であることを示した。視聴行動分析では、讀賣テレビ放送株式会社の非特定テレビ視聴履歴データを用いた分析、および在阪の4つの民放局の統合された非特定テレビ視聴履歴データを用いた分析を実施した。まず、CM放送やCM視聴者数がインターネット検索に与える影響を可視化することで、テレビのCMが視聴者に与える影響の分析を行った結果、CM放送回数の増加と共に検索行動の増加や検索行動に効率よく結びつくCM放送時間帯を確認できた。次に、CMブロックや番組コーナー単位でチャンネル離脱行動に至ったテレビ端末に着目し、地域別の差異について分析した。その結果、CMになっても離脱行動が少ない地域特性や番組カテゴリ、また番組カテゴリごとに出演者と離脱行動分析をも行い、番組カテゴリ単位で完視聴率が高くなる傾向のある属性や視聴者の好みの推測を行った。最後に、視聴行動にのみ着目し、視聴者の視聴習慣や視聴する番組カテゴリの傾向に基づいた視聴履歴データをクラスタリングする手法を提案する。分析の結果、テレビの視聴時間、視聴番組カテゴリ、視聴時間帯の3つの観点で視聴傾向の分類が可能であることを示した。

本研究を通して、統合視聴データの有用性、更に視聴データから視聴者行動分析を行うことでCMの付加価値を高める可能性を示すことができた。

## キーワード

テレビ、視聴履歴データ、ビッグデータ、IoT、クロスデバイスマッチング、クロスデバイストラッキング、視聴者行動、可視化、インターネット検索、シミュレータ、地域性

# **A study on the integration of non-specific TV viewing history data among broadcasters and analysis of viewing behavior\***

Hiroki Matsuda

## **Abstract**

In recent years, terrestrial television advertising expenditures have been overtaken by Internet advertising expenditures, which have continued to decline. Although it is said that TV commercials can deliver advertisements to a large number of viewers, the value of advertisements has only been visualized through estimates based on statistical data called viewership ratings compiled by Video Research, Inc. Therefore, TV broadcasters are working to collect and utilize non-specific viewing history data, including viewing start and end times, from TVs connected to the Internet in a format that does not identify individual viewers. However, the collection methods and data granularity of non-specific viewing history data differ among broadcasters, and the data accumulated by each broadcaster has not yet been integrated and utilized. In this study, we propose a method for integrating non-specific viewing history data that each broadcaster collects in its own way, design and implement a simulator for evaluating the proposed method, and evaluate the proposed method. We also analyze the viewing behavior of four broadcasters in Osaka, using the integrated viewing data to improve the value of TV advertisements and to utilize the data for program production. In the non-specific viewing history data integration method, TV receivers are separated by the five common items in each station's viewing history data: IP address, zip code,

---

\*Doctoral Dissertation, Graduate School of Science and Technology, Nara Institute of Science and Technology, July 30, 2024.

manufacturer ID, browser major version, and browser minor version. Of the separated TVs that match these five items, the TVs with the same channel transition time are presumed to be the same TV. In addition, we designed a simulator that reproduces viewer behavior, and applied this method to the viewing history data synthesized from the simulator. In the analysis of viewing behavior, we conducted an analysis using the non-specific viewing history data of Yomiuri Telecasting Corporation and the integrated non-specific viewing history data of four commercial TV stations in Osaka. First, by visualizing the impact of CM broadcasts and the number of CM viewers on Internet searches, we analyzed the impact of TV commercials on viewers. As a result, we were able to confirm the increase in search behavior with the increase in the number of CM broadcasts and the time slots of CM broadcasts that are efficiently linked to search behavior. Next, we focused on the TV terminals that led to channel withdrawal behavior by CM block and program corner, and analyzed the differences by region. As a result, we analyzed the regional characteristics and program categories in which the withdrawal behavior was low even after commercials were broadcast, as well as the performers and withdrawal behavior for each program category, to infer the attributes and viewer preferences that tend to increase the complete viewing rate for each program category. Finally, focusing only on viewing behavior, we propose a method for clustering viewing history data based on viewers' viewing habits and trends in the program categories they watch. The results of the analysis show that it is possible to classify viewing tendencies in terms of TV viewing time, program categories viewed, and viewing time period.

Through this study, we were able to demonstrate the usefulness of integrated viewing data and the possibility of adding value to TV commercials by analyzing viewer behavior from viewing data.

**Keywords:**

TV, TV viewing log data, Big data, IoT, Cross-device matching, Cross-device tracking, TV viewer behavior, Data visualization, Internet search, simulator, locality

# 目次

<b>1. 序論</b>	<b>1</b>
1.1 はじめに	1
1.2 日本における視聴履歴データの分類	4
1.2.1 オプトイン型特定視聴履歴データ	4
1.2.2 オプトイン型非特定テレビ視聴履歴データ	5
1.2.3 オプトアウト型非特定テレビ視聴履歴データ	5
1.3 非特定テレビ視聴履歴データが抱える現状の問題	6
1.4 本研究の目的	7
1.5 本論文の構成	7
<b>2. 非特定テレビ視聴履歴データの放送局間統合手法</b>	<b>9</b>
2.1 はじめに	9
2.2 関連研究	9
2.2.1 非特定テレビ視聴履歴データの連携・統合技術実証	9
2.2.2 デバイスやブラウザを跨るデータの統合に関する研究	10
2.2.3 欠損値データの補完に関する研究	11
2.2.4 本章の位置づけ	12
2.3 提案手法（放送局横断テレビ視聴履歴データ統合手法）	12
2.3.1 NNTM アルゴリズム概要	13
2.3.2 NNTM アルゴリズム適用ステップ	13
2.4 本アルゴリズムの実践	18
2.4.1 在阪局のオプトアウト型非特定テレビ視聴履歴データ	18
2.4.2 実視聴履歴データへの適用事例プラクティス	20
2.5 テレビ視聴行動シミュレータ	27
2.5.1 シミュレータの全体構成	27
2.5.2 視聴者モデルの設計	28
2.5.3 テレビ受像機モデルの設計	33
2.5.4 放送局モデルの設計	34

2.6	シミュレータを用いたNNTMアルゴリズムの評価・考察 . . . . .	35
2.6.1	評価手法と結果 . . . . .	35
2.6.2	議論と考察 . . . . .	36
2.6.3	提案シミュレータの制限 . . . . .	38
2.7	まとめ . . . . .	38
<b>3.</b>	<b>非特定テレビ視聴履歴データの分析</b>	<b>40</b>
3.1	はじめに . . . . .	40
3.2	関連研究 . . . . .	40
3.2.1	オプトイン型非特定テレビ視聴履歴データに関する研究 . . . . .	41
3.2.2	テレビのCM視聴に関する研究 . . . . .	41
3.2.3	報道時間と視聴率の関係 . . . . .	42
3.2.4	本研究の位置付け . . . . .	43
3.3	分析1: テレビのCM視聴とインターネット検索データの関係性分析	43
3.3.1	データセット . . . . .	44
3.3.2	分析 . . . . .	45
3.3.3	まとめ . . . . .	55
3.4	分析2: CM視聴傾向の分析 . . . . .	55
3.4.1	データセット . . . . .	55
3.4.2	データの処理方法 . . . . .	56
3.4.3	分析 . . . . .	60
3.4.4	まとめ . . . . .	69
3.5	分析3: 番組コーナー視聴傾向の分析 . . . . .	70
3.5.1	データセット . . . . .	70
3.5.2	データの処理方法 . . . . .	71
3.5.3	分析 . . . . .	75
3.5.4	まとめ . . . . .	83
3.6	分析4: クラスターリング分析 . . . . .	84
3.6.1	データセット . . . . .	84
3.6.2	提案手法 . . . . .	85

3.6.3	データの処理 . . . . .	86
3.6.4	分析と評価 . . . . .	87
3.6.5	まとめ . . . . .	96
3.7	まとめ . . . . .	96
<b>4.</b>	<b>結論</b>	<b>98</b>
	謝辞	<b>100</b>
	参考文献	<b>101</b>
	研究業績	<b>104</b>

## 目 次

1	日本の広告費	2
2	本論文の構成	8
3	あるテレビにおけるマッチングイメージ	13
4	NNTM アルゴリズムの手順	14
5	テレビ分離のイメージ	15
6	一意特定処理の適用例	17
7	4局推定処理の適用例	18
8	項目別のテレビ分離割合	22
9	A局1台に対してB局2台紐づいたパターン	23
10	期間別のマッチング数	26
11	シミュレータ処理の流れ	28
12	分譲集合住宅の戸数分布（東京都）[1]	31
13	賃貸集合住宅の戸数分布（東京都）[1]	31
14	各放送局の状態遷移図（平日18時30分～59分平均）	32
15	IPアドレス変化回数の相対度数分布	33
16	IPアドレス変化回数の相対度数分布と変化時間（1回以上を抜粋）	34
17	期間別の再現率（Step別）	37
18	CM有無別の24時間以内の検索量取得方法	45
19	連続的な3時間においてCM放送回数が計数される例	46
20	24時間以内のCM有無別の検索量（清涼飲料水）	47
21	3時間以内のCM回数別の検索量（アルコール飲料）	48
22	3時間以内のCM回数別の検索量（清涼飲料水）	49
23	3時間以内のCM回数別の検索量（ネット保険）	49
24	CM放送前後の検索量推移	50
25	CM放送時間帯（1時間単位）と視聴台数の分布	53
26	CM放送時間帯（1時間単位）と検索量の分布	53
27	視聴台数10,000台あたりの検索量とCM放送時間帯の分布	54
28	CM視聴UD率の定義	57

29	コーナー時間と完視聴率の平均の変化 . . . . .	74
30	t-SNE によるクラスタリング結果の可視化 . . . . .	88
31	番組カテゴリの対応色 . . . . .	89
32	ヘビー層/朝・夜/偏り少 . . . . .	90
33	ミドル層/朝/バラエティ少 . . . . .	91
34	ミドル層/偏り少/バラエティ少 . . . . .	91
35	ライト層/深夜・早朝/アニメ・音楽多 . . . . .	92
36	ライト層/朝/バラエティ少 . . . . .	92
37	ヘビー層/朝夜/音楽・バラエティ多 . . . . .	94
38	ミドル層/朝夜/スポーツ多 . . . . .	94
39	ミドル層/夜/音楽バラエティ多 . . . . .	95
40	ライト層/全時間帯/アニメ多 . . . . .	95
41	ライト層/朝夜/スポーツ多 . . . . .	96

## 表目次

1	視聴履歴データの分類 . . . . .	4
2	各放送局の時刻データ正確性 . . . . .	16
3	各放送局の取得方式と時刻データ正確性 . . . . .	19
4	在阪4局の視聴履歴データ比較 . . . . .	19
5	データ処理後の在阪4局視聴履歴データ比較 . . . . .	20
6	許容誤差別の放送局1対1マッチング数 . . . . .	24
7	許容誤差別の放送局1対多マッチング数 . . . . .	24
8	一意特定処理数 . . . . .	25
9	期間別のマッチング数 . . . . .	26
10	共同住宅別のIPアドレス共用数(アンケート) . . . . .	29
11	2018年住宅の建て方別住宅数(全国) . . . . .	29
12	2013年マンション総数(東京) . . . . .	29
13	住宅種類別の割合 . . . . .	30

14	テレビ視聴行動シミュレータを用いて合成した在阪4局の合成視聴履歴データの概要	35
15	期間別のマッチング数と再現率	36
16	3時間以内の放送回数別のサンプル数 (CM別)	48
17	CM放送時間と検索量計測期間の例	52
18	視聴UD率の基本統計量	61
19	ch切替回数の基本統計量	61
20	統計データ各項目と1/3以上視聴UD率との相関係数 ( $r$ ) の分布	62
21	統計データ各項目と2/3以上視聴UD率との相関係数 ( $r$ ) の分布	62
22	統計データ各項目と完視聴UD率との相関係数 ( $r$ ) の分布	62
23	統計データ各項目とUDあたりch切替回数との相関係数 ( $r$ ) の分布	62
24	1/3以上視聴UD率の平均と正の相関がある項目	64
25	1/3以上視聴UD率の平均と負の相関がある項目	64
26	2/3以上視聴UD率の平均と正の相関がある項目	64
27	2/3以上視聴UD率の平均と負の相関がある項目	64
28	完視聴UD率の平均と正の相関がある項目	65
29	完視聴UD率の平均と負の相関がある項目	65
30	ch切替回数の平均と正の相関がある項目	66
31	ch切替回数の平均と負の相関がある項目	66
32	1/3以上視聴UD率の標準偏差と正の相関がある項目	66
33	1/3以上視聴UD率の標準偏差と負の相関がある項目	66
34	2/3以上視聴UD率の標準偏差と正の相関がある項目	67
35	2/3以上視聴UD率の標準偏差と負の相関がある項目	67
36	完視聴UD率の標準偏差と正の相関がある項目	67
37	完視聴UD率の標準偏差と負の相関がある項目	67
38	ch切替回数の標準偏差と正の相関がある項目	68
39	ch切替回数の標準偏差と負の相関がある項目	68
40	1/3以上視聴率の平均と正の相関がある項目	76
41	1/3以上視聴率の平均と負の相関がある項目	76

42	2/3 以上視聴率の平均と正の相関がある項目	77
43	2/3 以上視聴率の平均と負の相関がある項目	77
44	完視聴率の平均と正の相関がある項目	77
45	完視聴率の平均と負の相関がある項目	77
46	コーナーのジャンルリストとその出現数(2021年12月)	79
47	ジャンル「情報/ワイドショー」の完視聴率と正の相関があるタレント統計項目	79
48	ジャンル「情報/ワイドショー」の完視聴率と負の相関があるタレント統計項目	79
49	ジャンル「バラエティー」の完視聴率と正の相関があるタレント統計項目	80
50	ジャンル「バラエティー」の完視聴率と負の相関があるタレント統計項目	80
51	ジャンル「ニュース/報道」の完視聴率と正の相関があるタレント統計項目	81
52	ジャンル「ニュース/報道」の完視聴率と負の相関があるタレント統計項目	81
53	ジャンル「ドラマ」の完視聴率と正の相関があるタレント統計項目	81
54	ジャンル「ドラマ」の完視聴率と負の相関があるタレント統計項目	81
55	ジャンル「音楽」の完視聴率と正の相関があるタレント統計項目	82
56	ジャンル「音楽」の完視聴率と負の相関があるタレント統計項目	82
57	ジャンル「スポーツ」の完視聴率と正の相関があるタレント統計項目	83
58	ジャンル「スポーツ」の完視聴率と負の相関があるタレント統計項目	83
59	テレビ番組のカテゴリ一覧	86

# 1. 序論

## 1.1 はじめに

近年、見逃し配信や動画配信サービス（VOD: Video On Demand）の視聴者の増加に伴い、Connected TV(CTV)と呼ばれるインターネットに接続されたテレビが増加しており、2023年には59.6%とのデータとなっている [2]。インターネット接続されたテレビについては、各テレビがどの番組を視聴しているか把握可能になっており、放送局やテレビ受像機メーカー等が視聴履歴データを収集・蓄積している。視聴履歴データは、本人許諾や個人情報の有無により、オプトイン型特定視聴履歴データ、オプトイン型非特定テレビ視聴履歴データ、オプトアウト型非特定テレビ視聴履歴データに分類される。

オプトアウト型非特定テレビ視聴履歴データは、個人を特定しない形式で収集されるデータを指し、視聴者がデータ取得を拒否しない場合は自動的に収集されるため、オプトイン型と比較すると収集規模が圧倒的に大きい。そのため、新たな価値を生み出すビッグデータとして、放送局のみならず、スポンサや広告代理店からも利活用が期待されている。オプトアウト型非特定テレビ視聴履歴データの利活用については、2019年に一般社団法人放送セキュリティセンターにより公表された「オプトアウト方式で取得する非特定視聴履歴の取扱いに関するプラクティス [3]」を基に、各民間放送局により収集が開始され、利活用に向けた取り組みが進んでいる。しかし、プラクティスにより、本データを放送局同士で交換し突合することは認められているが、本データを放送局以外の団体に渡すことや、他の個人情報データと突合することは禁止されているため、各放送局が個別に収集している本データ単体を分析したとしても、単一の放送局におけるテレビ視聴状況しか把握出来ないため、新たな価値を生み出すには至っていないのが現状である。

また、電通 [4] の報告によると、2019年にインターネット広告費が地上波テレビ広告費を上回ったが（図 1）、依然としてテレビ広告は数千万人以上にリーチすることが可能な巨大市場（約 2 兆円規模）を有している。

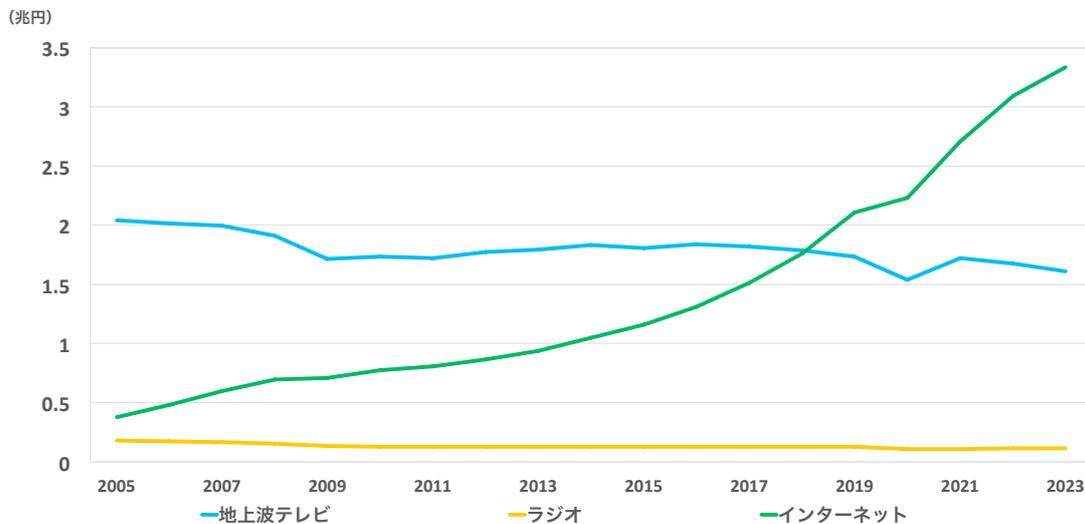


図 1: 日本の広告費

現在、テレビ広告の価値を測る指標としては株式会社ビデオリサーチが提供する視聴率が存在しているが、標本調査による統計データとなっているため、テレビのCMや番組コンテンツをより深く分析するには不十分な場合がある。例えば、株式会社ビデオリサーチのデータを用いた年齢や性別による視聴傾向分析が行われているが、最近の視聴者は趣味趣向が個別化しているため、年齢や性別による単純なカテゴリ分類では議論することが難しい。より詳細な分析を行うためには、同一の番組出演者や同一番組ジャンル等を視聴した視聴者がどのような視聴傾向を持っているか等、視聴者の行動を起点とした視聴傾向を把握することが重要だと考えられる。これらの視聴者行動は、抽出条件を増やすほど視聴者行動をグループ化することが可能となるが、視聴率データは関西地区の1,200世帯を対象としており標本数が少ないため、同様の分析が難しい。

この課題に対して、関西地区だけでも100万台以上のテレビから各放送局が個別に収集している非特定テレビ視聴履歴データを連携・統合することによる解決手法の検討が進んでいる。在阪の4つの民間放送局においても、放送局間での視聴履歴データの連携が可能か議論および、共同技術検証実験[5][6]を実施している。しかし、非特定テレビ視聴履歴データは、各放送局が取得にあたりコスト検討や仕様検討し、各放送局の経営判断により取得を開始しているため、放送局ご

とに方式やフォーマットが異なっており、放送局間のデータを統合することは難しい。

そこで本研究では、各放送局が収集した非特定テレビ視聴履歴データについて、視聴者のチャンネル切替えを把握できるように、放送局間の非特定テレビ視聴履歴データを掛け合わせることで各放送局が各テレビ端末を特定するために割り振っているID（以下、テレビID<sup>1</sup>）をマッチングさせることで非特定テレビ視聴履歴データを統合する手法を提案し、実データへの実践を行った上で、シミュレータを用いて評価する。

提案手法では、IPアドレスをキーにして、放送局を超えてテレビIDをマッチングさせる。本論文では、マッチングとは同一テレビ端末に各放送局が独自で割り振ったテレビID同士を同一推定することを指す。しかし、IPアドレスはマンション等の集合住宅では1つのIPアドレスを複数宅で共用していることも多く、更に一定期間で変更されていくため、IPアドレスだけで同一テレビを推定することは難しい。そこで、IPアドレス以外に視聴履歴データ収集時に取得している郵便番号、メーカID、ブラウザメジャーバージョン、ブラウザマイナーバージョンの5項目すべてが一致するテレビのうち、更にチャンネル遷移タイミングが一致するテレビを同一テレビと推定する。提案手法は、実データのマッチング数では実際の視聴者行動との相関が分からず、評価が難しい。そこで、テレビ視聴行動を再現するエージェントシミュレータを用いて、エージェントの視聴行動からテレビ視聴履歴データを合成し、その合成視聴履歴データに対して提案手法を適用することでテレビIDマッチングの再現率を評価した。その結果、シミュレータから合成された250万台のうち、約241万台のテレビIDのマッチングに成功し、再現率96.5%であることを示した。

また、これらの統合された非特定テレビ視聴履歴データを中心に実践的なデータ分析事例を示す。1つ目の事例として、テレビのCMの価値を可視化するために、CM視聴後のインターネット検索行動と非特定テレビ視聴履歴データの関係性分析を示す。2つ目と3つ目の事例では、CMブロック単位・番組ジャンル別コーナー単位での地域別視聴傾向の分析や、出演タレントと視聴の相関分析を

---

<sup>1</sup>放送局ごとに呼び方は異なるが、本論文ではそれらを「テレビID」と総称する。

行うことで、CMや番組コーナーが良く見られている地域や、継続視聴に寄与している要因などを示した。最後の事例では、テレビ所有者の属性情報を一切持たない非特定テレビ視聴履歴データの視聴番組ジャンル履歴から特徴量を作成し、K-meansを用いてテレビ視聴者を視聴傾向に基づいてクラスタリングする手法を提案する。

## 1.2 日本における視聴履歴データの分類

本節では、視聴履歴データの分類および現状の問題についてそれぞれ述べる。

日本国内において視聴履歴データは、表1のように、本人許諾や個人情報の有無により、オプトイン型特定視聴履歴データ、オプトイン型非特定テレビ視聴履歴データ、オプトアウト型非特定テレビ視聴履歴データの三種類に分類される。各データの特徴を表1に示すとともに、以降で詳述する。

表 1: 視聴履歴データの分類

種類	データの特徴				本論文で扱う データセット
	本人許諾	個人情報	データの多様性	データの数	
オプトイン型 特定視聴履歴データ	有り	有り	低い	少ない	-
オプトイン型 非特定テレビ視聴履歴データ	有り	無し	低い	少ない	-
オプトアウト型 非特定テレビ視聴履歴データ	無し	無し	高い	多い	○

### 1.2.1 オプトイン型特定視聴履歴データ

オプトイン型特定視聴履歴データは、視聴者から許諾を得た上で収集される、メールアドレス等の本人特定が可能な個人情報と紐づけられた視聴履歴データである。放送局は、視聴者がパソコンやスマートフォン等からウェブページを通して会員登録を行い、インターネット接続したテレビのデータ放送画面からログインすることで取得可能な状態となる。本人許諾を得る際に、多様な個人情報を取得できるため、視聴履歴データの分析が容易に行える。しかし、視聴者自身の手

で、事前にテレビ端末以外から会員登録を行い、更にテレビ端末からログインする必要があるため、負担が大きく、多くのデータを集めることは難しい。また、会員登録を行う視聴者層にも偏りが大きいと言われており、収集されるデータの多様性が低い。

### 1.2.2 オプトイン型非特定テレビ視聴履歴データ

オプトイン型非特定テレビ視聴履歴データは、視聴者から許諾を得た上で収集されるが、個人情報を含まない視聴履歴データである。

放送局は、視聴者がインターネットに接続されたテレビのデータ放送画面を用いて許諾を行うことで取得可能な状態となる。この方法では、許諾時にアンケート形式で性別、生年月等の視聴者属性に回答して貰うことで、個人情報とはいかないまでも、視聴者属性を持った視聴履歴データを取得可能である。オプトイン型特定視聴履歴データとは違い、テレビ端末のみで許諾が得られるため、視聴者の負担は軽減されるが、この手法においても、視聴者が能動的に取得のための手続きを行う必要があるため、参加者層の多様性が低く、多くのデータを収集することも難しい。

### 1.2.3 オプトアウト型非特定テレビ視聴履歴データ

オプトアウト型非特定テレビ視聴履歴データは、本人の許諾無しに収集される、個人情報を含まない視聴履歴データである。放送局は、視聴者がテレビをインターネットに接続するだけで取得可能な状態になる。視聴者がデータを提供したくない場合は、データ放送画面からデータ提供を拒否(オプトアウト)できる。視聴者の能動的な会員登録や許諾を必要とせず、個人情報や視聴者属性を取得することはできないため、視聴開始/終了時刻や、テレビに登録されている郵便番号(住所ではない)、テレビIDなどの個人が特定できないデータのみが収集可能である。オプトイン型と異なり、会員登録や許諾を必要としないため、視聴者の負担は無く、多様な視聴者から多くのデータを収集することができる。しかし、個人情報や視聴者属性を用いることができないため、各放送局で蓄積されている自局

のデータのみでは、簡単な分析しか行えない。なお、本論文で使用するデータはオプトアウト型非特定テレビ視聴履歴データ（表1参照）である。

### 1.3 非特定テレビ視聴履歴データが抱える現状の問題

放送局が収集する視聴履歴データは、視聴者がインターネットに接続されたテレビで、特定の放送局にチャンネルを設定したタイミングで各放送局で提供しているデータ放送プログラムを利用して収集される。つまり、各放送局は、自局を視聴中のテレビから、独自の項目および粒度で視聴履歴データを収集している。しかし、視聴履歴データの更なる活用を進めていくためには放送局を横断する視聴履歴データを生成し、今以上に視聴者を理解する必要がある。また、横断的な視聴履歴データ生成は、テレビ広告価値を可視化するための手段として、スポンサや広告代理店からの期待は大きい。例えば、現在のテレビ広告は主にそのCMを視聴している視聴者の性別・年齢帯で価値が決まることが多い。しかし、現在は生活者の趣味嗜好は多様化しており、性別・年齢帯によるターゲティングでは効果を疑問視するスポンサも出てきており、生活者の趣味嗜好や興味関心に対してターゲティングを実施したいニーズが生まれている。そのニーズに合わせていくためには、例えば横断的な視聴履歴データから視聴者の視聴行動のみに重点を置き、視聴している番組ジャンルから興味関心を持っているセグメント分けを行い、興味関心によるターゲティング等を実施していく必要がある。そのためには、単一放送局では放送している番組ジャンルや出演者の数に限りがあるため、放送局のデータを横断的に利用することが望ましい。

そして、横断的な視聴履歴データを生成するためには、少なくとも複数の放送局を跨いだデータの統合が必要である。データの統合については、2018年度から継続的に、総務省により視聴履歴データに関わる実証事業が行われており、在京放送局が非特定テレビ視聴履歴データの連携・統合に向けて技術実証を実施している。

具体的には、テレビ受像機に内蔵されているNVRAMと呼ばれる不揮発性メモリに割り当てられた放送事業者共通保存領域に共通サーバから発番された共通IDを書き込むことにより各社のテレビIDをマッチングさせることでデータを統

合する方式である。本方式を使えば、確実に放送局間の ID をマッチングさせることができるため、優れた手法ではあるが、実施のハードルは高い。まず、在京放送局が主導して、NVRAM を共同利用する放送事業者は共同で一般社団法人放送サービス高度化推進協会（A-PAB）に申請をするルールを作っているが、そのルール自体は在京以外の放送局の利用は想定されていない。また、在京方式では共通 ID を発番するための第三者機関が前提となっており、この第三者機関はデータの取扱いに対して高いレベルでのガバナンスが求められているため、実施放送局の管理下にある事業者（放送局が資金面、人事面で強い影響を持つ等）と指定されている。共通 ID 発番のためだけに在阪放送局やその他地域の放送局が地域単位で合同出資法人を立ち上げることは難しい。

## 1.4 本研究の目的

本研究では、放送局のデータ活用推進に向けて、各放送局が独自で取得している視聴履歴データの統合を目指す。具体的には、NVRAM の事業者共通保存領域を利用しない手法を提案することで、日本全国のローカル放送局でも横断的な視聴履歴データの活用が可能となることを示す。

また、統合された非特定テレビ視聴履歴データを用いることで、従来では把握できなかった放送局を跨いだ視聴者行動を可能となった。そこで本研究では、効果的な CM を放送する、さらに視聴者数を維持・向上させるキャスティングすることを目的に非特定テレビ視聴履歴データの分析を行う。

## 1.5 本論文の構成

本論文の構成を図 2 に示す。

第 2 章では、放送局間の視聴データが統合できないという課題に対して、各放送局が独自で発行している ID をマッチングさせることで、非特定テレビ視聴履歴データを統合する手法を提案し、実際の非特定テレビ視聴履歴データへの適用プラクティス紹介・シミュレータ構築と統合手法の評価を実施する。

第3章では、視聴データによる広告価値の可視化や放送局内での業務利用が進んでいないという課題に対して、統合された非特定テレビ視聴履歴データを中心に、共同研究先である讀賣テレビ放送株式会社での非特定テレビ視聴履歴データの営業・制作・編成利用に向けた分析結果を紹介する。

第4章では、提案手法と分析結果についてまとめ、今後の課題等を示す。

課題1  
放送局間の  
視聴データが統合できない

課題2  
視聴データによる  
広告価値可視化に至っていない

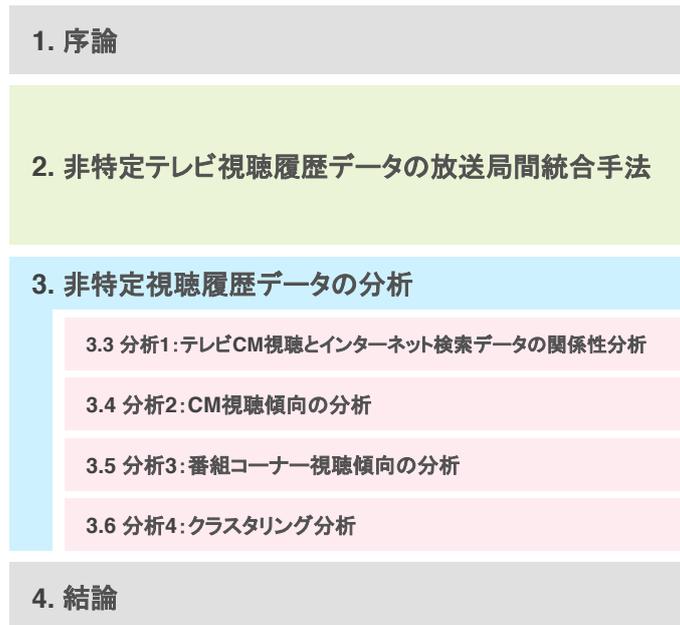


図 2: 本論文の構成

## 2. 非特定テレビ視聴履歴データの放送局間統合手法

### 2.1 はじめに

民放各局が収集を開始している非特定テレビ視聴履歴データではあるが、データの統合が進んでおらず、現状は単一放送局の視聴履歴データのみでの分析に留まっている。また、非特定テレビ視聴履歴データは同一放送エリアの放送局間で共同利用することで利用価値が高まると考えられているが、現状は共同利用には至っていない。

共同利用に至っていない理由は、1.3節で前述の通りだが、本研究では共同利用のボトルネックになっている「NVRAM 放送事業者共通領域」を一切使用しない形での非特定テレビ視聴履歴データの統合手法を提案する。提案方式を利用すれば、「NVRAM 放送事業者共通領域」の利用に必要とされている A-PAB 申請だけでなく、共通 ID 発番を実施する法人の設立などが不要となり、人的・金銭的に余裕のないローカル放送局でも非特定テレビ視聴履歴データの同一放送エリア内での共同利用を実現できる。

本章では、NVRAM 共通領域を使用しない非特定視聴データ統合手法を提案し、大阪にある4つの民放局の実データを利用した実践結果とシミュレータを利用した提案手法の評価を行う。

### 2.2 関連研究

本節では、在京局による非特定テレビ視聴履歴データの統合技術実証やクロスデバイストラッキング技術に関する関連研究を提示した後、本研究との差異や位置づけについて述べる。

#### 2.2.1 非特定テレビ視聴履歴データの連携・統合技術実証

横断的な視聴履歴データを生成するためには、少なくとも複数の放送局を跨いだデータのマッチングが必要である。データのマッチングについては、1.3節で前述した通り2018年度と2019年度に、総務省により視聴履歴データに関わる実

証事業が行われており、在京放送局が非特定テレビ視聴履歴データの連携・統合に向けて技術実証を実施している。具体的には、テレビ受像機に内蔵されているNVRAMと呼ばれる不揮発性メモリに割り当てられた放送事業者共通保存領域に共通IDを書き込むことによりデータをマッチングさせる方式である。本方式を使えば、確実に放送局間のIDをマッチングさせることができるため、優れた手法ではあるが、3点の課題がある。

まず、1点目として、放送事業者共通保存領域は民間地上放送事業者だけでなく、個人情報を大量に保有するNHKや有料放送事業者も読み込むことができる。共通IDは改正個人情報保護法において、個人関連情報と定義されているため、個人関連情報をあらゆる放送事業者が参照できる場所に保管することに対して、個人情報保護委員会から疑義がでている。

2点目として、放送事業者共通保存領域を民間放送局の利益のために使用するためには、テレビ受像機の規格改定が必要となる。在京局が主導して、NVRAMを共同利用する放送事業者は共同で一般社団法人放送サービス高度化推進協会(A-PAB)に申請するルールを制定しているが、そのルール自体が在京局以外が使うことが想定されておらず、在京局以外がNVRAMを共同利用することは難しい。

3点目として、在京局が想定している共同利用方式では、共通IDを発番するための第三者機関が前提となっており、この第三者機関はデータの取扱いに対して高いレベルでのガバナンスが求められている。そのため、共通ID利用放送局の管理下(放送局が資金面、人事面で強い影響を持つ等)にある事業者が条件となっており、共通ID発番のためだけに在阪放送局やその他地域の放送局が地域単位で合同出資法人を立ち上げる必要があるが、金銭的・要員的な余裕がない放送局も多いため、現実的な手段にはなっていない。

### 2.2.2 デバイスやブラウザを跨るデータの統合に関する研究

インターネット広告では広告効果の計測を試みるために、同一の利用者が使用している複数のパソコンやモバイルといった端末を紐づける手法であるクロスデバイストラッキング技術に関する研究が進められてきた。この技術では、パソコン

とモバイルが別サーバへアクセスしたデータを別個に採取し、サーバで取得できる IP アドレスやタイムスタンプ、広告 ID、cookie 等を用いて紐づけを行う。Boda ら [7] は、異なるブラウザ間において、ブラウザで取得可能な IP アドレス、フォントセット、タイムゾーン、画像解像度から利用者を識別する手法を提案し、利用者を一意に識別可能であることに言及した。また、齋藤ら [8] は、深層学習とランダムフォレストをそれぞれ用いて、モバイルアプリとブラウザからそれぞれ得られるデータを用いて紐づけを実施している。データは、IP アドレス、UserAgent 情報、タイムスタンプおよび cookie を用いており、大規模サンプルにおいて、高精度な結果を得ている。

しかし、テレビ端末では広告 ID や cookie 等の第三者が発行しているデータはなく、更に同一の利用者の同一テレビ端末を紐づけるのに、これらの手法を適用することは難しい。

### 2.2.3 欠損値データの補完に関する研究

蓄積されるデータの欠損部分を推定し、適切な値で補う欠損値データ補完として、いくつかの手法が提案されている。欠損値データ補完とは、蓄積データ内で欠落している部分のデータを、一番もっともらしい値で補うことを指す。放送局を跨る非特定テレビ視聴履歴データの放送局間統合は、単局の視点から見ると他局データを補完していると考えられる。データマイニングにおける欠損値データの補完と分析については、様々な手法が存在する。まず、分析に用いるデータから欠損値を含むデータを対象から削除する手法（リストワイズ削除）である。しかしながら、単一放送局のデータ視点だと、すべてのテレビ ID が欠損データの扱いとなってしまうため、利用できない。また、平均値や中央値といった統計量で補完する方法が挙げられる [9]。とある期間のデータが欠損していた場合、時系列毎に過去データの平均値、中央値を用いて補完を行う。しかし、本手法では例えばテレビの特別編成番組などのデータが欠損している場合は、大きく外れたデータを補完してしまう。他には、似た個体の値で補完する方法もある [10]。これは、同様の視聴傾向にあるテレビ端末のデータを補完データとして利用する方法である。本手法であれば、大きな外れ値を補完する可能性は減少するが、テレ

ビは家族などの複数人で利用されることが多く、各利用者の生活リズムやその瞬間の興味関心のも視聴番組に影響が出ると考えると完全な補完は難しい。これら以外にもベイズ推定、機械学習などの予測モデルを用いた補完法[11]も提案されているが、本研究で取組む単一放送局から残り3つの放送局のデータを補完することは、欠損値データが大きいため難しい。

#### 2.2.4 本章の位置づけ

本章では、在阪の4つの放送局（読売テレビ・毎日放送・朝日放送テレビ・関西テレビ）がそれぞれ独自に収集したオプトアウト型非特定テレビ視聴履歴データを用いて、前述の在京局による実証とは違うNVRAMの放送事業者共通保存領域を使用しないデータ統合の手法の提案および実践を示す。更に、テレビのオプトアウト型非特定テレビ視聴履歴データの分析などに基づいて視聴者・テレビ受信機をモデル化することによって、テレビ視聴行動シミュレータを設計・構築するとともに、提案手法の性能評価を行うことを目的とする。

### 2.3 提案手法（放送局横断テレビ視聴履歴データ統合手法）

本節では、NVRAMの放送事業者共通保存領域を使用しないテレビ視聴履歴データ統合手法である、Non-NVRAM TIME マッチングアルゴリズム（以下、NNTMアルゴリズム）を提案する。本手法は、データ収集を行っているデータ放送プログラムは変更せず、各局各様で収集しているオプトアウト型非特定テレビ視聴履歴データをそのまま利用することを前提とするが、オプトアウト型非特定テレビ視聴履歴データ以外への適用も可能である。

著者らはこれまでに、放送局間の視聴履歴データの連携に関する技術検証を目的として実施された「テレビ視聴データ連携に関する共同技術検証実験 [5][6]」（以下、在阪視聴データ連携技術実験）において、在阪の4つの民間放送局でそれぞれ独自方式によって収集した実際の非特定テレビ視聴履歴データを対象に、NVRAMの放送事業者共通保存領域を使用せずに、放送局間でデータを統合する手法について実践的な取り組みを行ってきた。しかしながら、この取り組みでは

統合を行った経験の報告にとどまっております。テレビ視聴履歴データ統合手法の確からしさについては検証できていない。そこで本論文では、テレビのオプトアウト型非特定テレビ視聴履歴データの分析などに基づいて視聴者・テレビ受像機をモデル化することによって、テレビ視聴行動シミュレータを設計・構築するとともに、提案手法の性能評価を行うことを目的とする。

### 2.3.1 NNTM アルゴリズム概要

NNTM アルゴリズムでは、非特定テレビ視聴履歴データの利用する。非特定テレビ視聴履歴データは、各局で収集された際に、テレビ ID のような独自の情報が付与されるが、データ放送技術を用いて収集される情報自体は規格化されているため、放送局を跨いだとしても同様の処理を行うことが可能である。本アルゴリズムでは、その特性を利用し、テレビ端末を各放送局が独自に付与しているテレビ ID 単位で IP アドレスを含む複数のデータで整理する。この処理をテレビ ID の分離処理と呼ぶ。この分離されたテレビのうち、更に視聴終了時刻と視聴開始時刻に一定の猶予を見た時刻が一致したものを同一テレビと推定する。NNTM アルゴリズムのマッチングイメージを図 3 に示す。

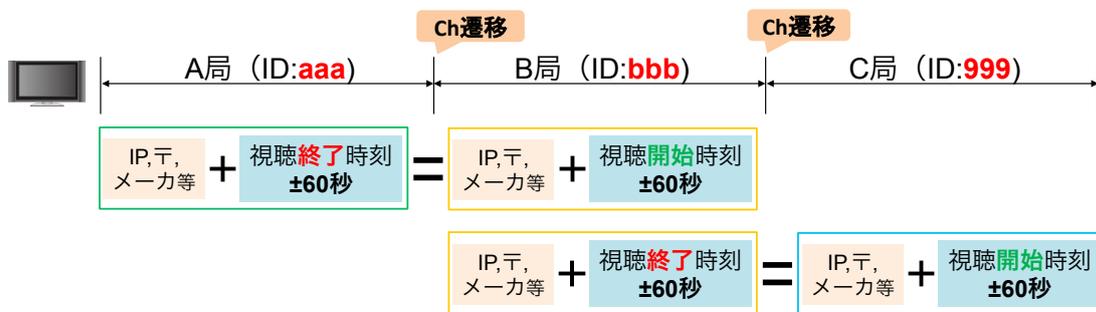


図 3: あるテレビにおけるマッチングイメージ

### 2.3.2 NNTM アルゴリズム適用ステップ

NNTM アルゴリズムを適用するステップを図 4 に示す。本項ではステップ毎にアルゴリズムを詳細に説明する。

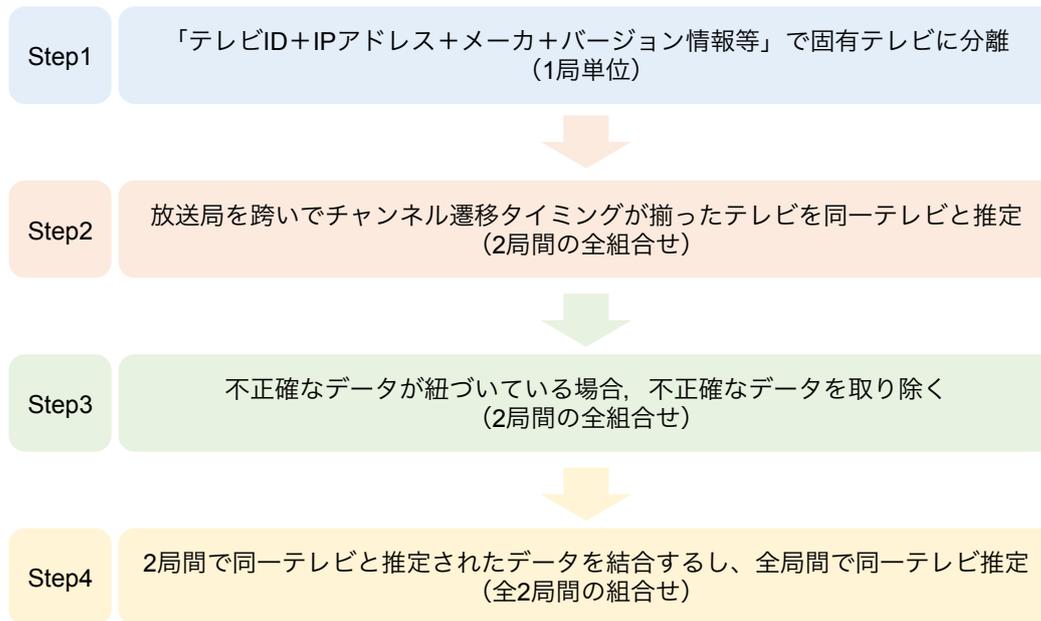


図 4: NNTM アルゴリズムの手順

**Step1: テレビ端末の分離** NNTM アルゴリズムは、クロスデバイストラッキングの一手法である IP アドレスマッチングを基本とする。しかし、IP アドレスはマンション等の集合住宅において 1 つの IP アドレスを複数宅で共用していることも多く、IP アドレスだけではテレビを一意に特定することは難しい。また、IP アドレスは一定期間で変更されていくが、どの程度の期間で変更されるかはプロバイダに依存しており、データを観測すると数分～数日で変更されるものが多い。

そこで、図 4 の Step1 では、各局が同一のテレビ端末から視聴履歴データを集めているという特徴を利用する。図 5 に示す通り、IP アドレスだけでテレビを分離するのではなく、その他にデータ放送プログラムを用いて同時に収集している郵便番号、メーカー ID（各放送局がテレビ受像機メーカーを区別するために割り振っている ID）、テレビ受像機を起動させているブラウザ<sup>2</sup>のメジャーバージョン情報、マイナーバージョン情報を使ってテレビの分離を行う。

<sup>2</sup>テレビ端末がデータ放送を起動する場合、専用のブラウザを利用している。メーカーやテレビ型番により利用されているブラウザは異なる。

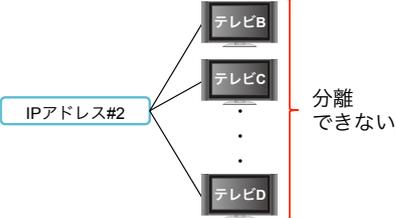
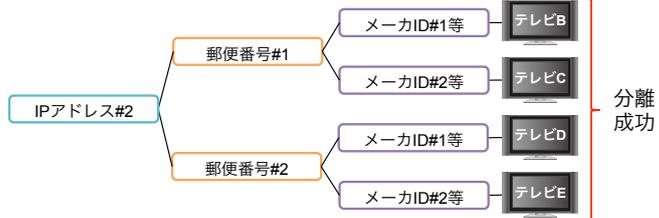
IPアドレスだけでテレビ分離	IPアドレス・郵便番号・メーカーID・ブラウザメジャーバージョン・ブラウザマイナーバージョンでテレビを分離
<p>IPアドレスに1台しか紐づいていない場合は分離可能</p> 	<p>IPアドレスに1台しか紐づいていない場合は分離可能</p> 
<p>IPアドレスに2台以上が紐づいている場合は分離ができない</p> 	<p>IPアドレスが1つであっても郵便番号などの場合は分離できる可能性が高い</p> 

図 5: テレビ分離のイメージ

**Step2: 放送局を跨いだテレビ端末の同一推定** Step2として、各放送局のテレビIDをIPアドレスや郵便番号等で分離した後にマッチングさせていくが、マッチング精度を高めるために更に「視聴時刻」をマッチング条件に使うこととする。具体的には、図3で示す通り、ある一台のテレビがA局・B局・C局とチャンネル遷移をしながら継続視聴している場合を想定する。A局・B局・C局は同一テレビに対して、それぞれ独自のテレビIDを付与している。これらのデータを持ち寄り、前述の通り「IPアドレス+郵便番号+メーカーID+ブラウザメジャーバージョン+ブラウザマイナーバージョン」が全て一致するテレビのうち、更に視聴終了時刻と視聴開始時刻に一定の猶予を見た時刻が一致したものを同一テレビと推定する。猶予時間は、理論的にはマッチングにおける視聴終了時刻と視聴開始時刻は、まったく同一時刻であることが望ましい。しかし、実際はチャンネル遷移後にデータ放送プログラムが起動してデータ収集を開始するタイムラグやデータの正確性に差が存在する。そこで猶予時間を設定することで、より正確なマッチングを目指す。ただし、この猶予時間を取り過ぎると、同じ集合住宅に住んでいる同一メーカー同一機種を所有している世帯間において、偶然チャンネル遷移のタイミングが一致する可能性がある。一致した場合、1対1マッチング数が減少し、1

対多のマッチング数が増加する。そこで、各局の視聴履歴データ特性から最適な猶予時間を設定する必要がある。この最適な猶予時間については、表 2 にある通り、視聴時刻は放送局の仕様により正確性に差があるため、そこから導くことができる。本アルゴリズムでは、誤差が最も大きい A 局ともマッチングすることを目指し、猶予時間を ± 60 秒に設定する。

表 2: 各放送局の時刻データ正確性

放送局	A 局	B 局	C 局	D 局
開始時刻の正確性	正確	正確	正確	正確
終了時刻の正確性	60 秒以内の誤差	15 秒以内の誤差	正確	正確

**Step3: 不正確データの処理** Step3 として、Step2 でも誤マッチングするテレビを完全に除去することは難しいため、不正確なデータが同一テレビと推定されている場合の処理を行う。具体的には、明らかに誤ったテレビが紐づいている場合、そのマッチングを除去する処理を実施する。Step2 の記載の通り、偶然チャンネル遷移のタイミングが一致すると、1 台のテレビに複数台のテレビが多数紐づくことが想定される。また、使用するデータ期間を増やせば増やすほど、例えば 100 回は同じテレビが 1 対 1 で紐づいているが、1 回だけ偶然別テレビが紐づいている等のデータが増えていってしまう。このように複数台がマッチングしたデータを全て捨ててしまうと使用するデータ期間を増やせば増やすほどマッチング数が減少することが想定される。そこで、1 台のテレビに複数台のテレビが紐づいているものに対して、式 (1) に示す条件によるフィルタリングを行うことで、明らかに不正確な紐づきが生じていると考えられるデータを除外し、マッチングデータが 1 対 1 となるように処理する（以降、一意特定と呼称する）。この条件式では、2 局間のテレビ ID の組合せのマッチング数が最も多いものに対し、次に多いマッチング数が半分未満となる場合を一意特定の閾値として設定している。

$$\begin{aligned}
 & \text{1 位のマッチング数} - \text{2 位のマッチング数} \\
 & > \text{全マッチング数} \times \frac{1}{3} \qquad (1)
 \end{aligned}$$

具体的な適用イメージを図6に示す。適用例では、A局とB局の間でA局のテレビ1台に対して、B局のテレビIDが複数マッチングした場合を想定している。この時、A局とB局での総マッチング数は4となり、1番マッチングが多かったIDの組合せである「yyyと222」のマッチング数は3となる。また、チャンネル遷移タイミングが一致することで誤マッチングした組合せが1回あったとする。この場合は条件式(1)を満たしているため、「yyyと222」の組合せを一意特定とし、それ以外のマッチングデータを消去する。

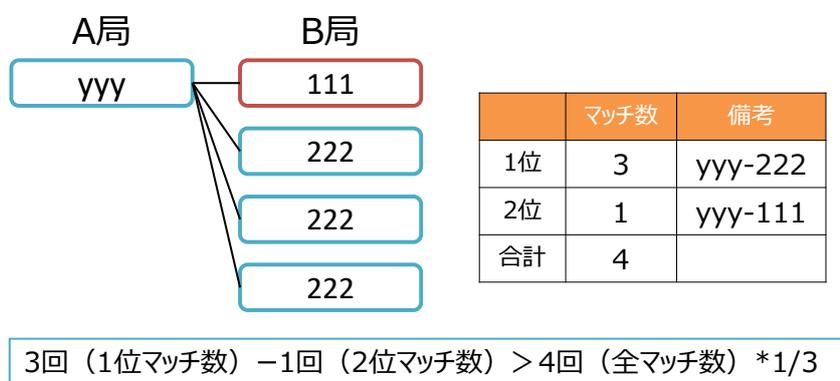


図6: 一意特定処理の適用例

なお、IPアドレスは時間と共に変化するが、本方式では個社が独自で割り振っているテレビIDをキーとして使い、さらに同一時間帯のチャンネル遷移タイミングでマッチングしているため、同じテレビIDに対して紐づくIPアドレスが途中で変更になったとしてもマッチング数への影響は少ないと想定される。

**Step4: 全2局間における同一テレビ推定処理** 最後に Step4 として、2局間で同一テレビ推定されたデータ同士を結合することで全2局間における同一テレビ推定処理を実施する。Step4の適用イメージを図7に示す。この例では、A局とB局、B局とC局、A局とD局で同一テレビと推定されたデータに基づいて4局全ての同一テレビ推定を行っている様子を示している。なお、適用放送局が増えた場合には、2局間の組み合わせ増加に伴い、ある特定の局に対する2局間同一テレビ推定の期待値が高まるため、結果としてNNTMアルゴリズムの性能向上が見

込める。

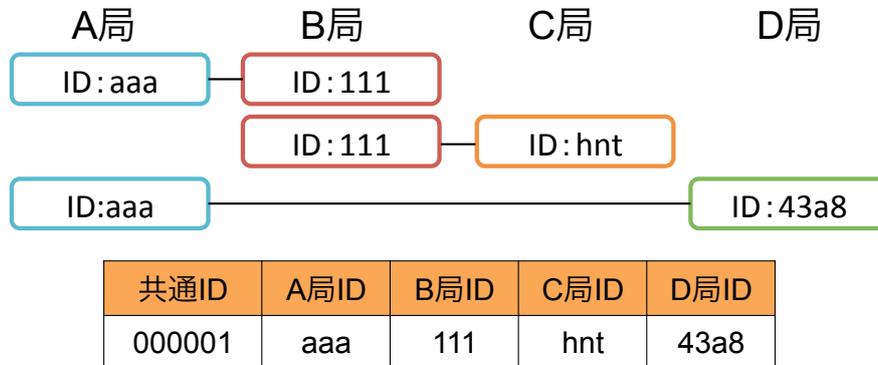


図 7: 4局推定処理の適用例

## 2.4 本アルゴリズムの実践

本節では、在阪視聴データ連携技術実験にて収集した実データと実際に適用したマッチング事例について述べる。

### 2.4.1 在阪局のオプトアウト型非特定テレビ視聴履歴データ

在阪の4つの放送局が取得している非特定テレビ視聴履歴データの特徴について説明する。今回の検証で使用したデータは、在阪4局にて実施した在阪視聴データ連携技術実験で取得・交換したデータを用いる。データ期間は、2021年10月4日～2022年1月5日である。

データ取得方式と取得対象メーカー データ放送プログラムを利用して収集する方法には、ビーコン方式と From-To 方式の2種類の方式があり、表3のように、放送局間で方式や収集条件が異なる。

ビーコン方式は、視聴者が視聴している時に一定間隔で視聴履歴データをサーバに送信する方法である。リアルタイム性には優れているが、送信間隔次第で視聴終了時刻が正確では無くなってしまいうという特徴がある。また、ビーコンの送

信間隔は放送局によって違いがあり、表3のように、A局は60秒、B局は15秒を採用している。

From-To方式は、視聴者が視聴し始めた時刻をNVRAMに保存しておき、また視聴終了時にも時刻をNVRAMに保存しておき、視聴者が当該チャンネルに戻ってきたタイミングで前回視聴開始・終了時刻をサーバに送信する方法である。From-To方式は、データ送信タイミングが視聴者行動に依存しているため、リアルタイム性はないが視聴開始・終了時刻を正確に記録できる特徴がある。

データを取得する対象メーカーも放送局によって違いがあり、A局とB局は全てのメーカーを対象に取得している。しかし、C局は一部の主要メーカーを取得対象としておらず、D局は国内シェアの大きいメーカーのみを対象としている。

表3: 各放送局の取得方式と時刻データ正確性

放送局	A局	B局	C局	D局
方式	ビーコン	ビーコン	From-To	From-To
ビーコン間隔	60秒	15秒	-	-
開始時刻の正確性	正確	正確	正確	正確
終了時刻の正確性	60秒以内の誤差	15秒以内の誤差	正確	正確
対象メーカー	全て	全て	一部未取得	主要メーカー

データ比較 在阪視聴データ連携技術実験で交換した2021年10月4日～2022年1月5日のデータ比較結果を表4に示す。

表4: 在阪4局の視聴履歴データ比較

放送局	テレビID数	データ量	行数
A局	4,639,071	404GB	11億行
B局	4,482,199	13.5TB	645億行
C局	3,614,554	64GB	6.7億行
D局	3,008,024	56GB	5.8億行

表4にあるようにデータ取得の対象が全てのメーカーになっているA局とB局は

テレビ ID 数が多いことが確認できる。また、ビーコン方式で取得している A 局と B 局は行数も多く、From-To 方式で取得している C 局と D 局は行数が少ないことが確認できる。A 局はビーコン方式でデータ取得を行っているが、データベースへ格納するタイミングで一部のデータを From-To 方式のデータ形式に変更することでデータ量と行数の削減を図っているため、同じビーコン方式の B 局よりも少なくなっている。

#### 2.4.2 実視聴履歴データへの適用事例プラクティス

在阪視聴データ連携技術実験の実データを利用したマッチング事例について述べる。

実データの前処理 前提条件を揃えるためにすべてのデータを C 局と D 局が対象としているメーカーに絞る。また、NNTM アルゴリズムを適用するにあたり、全局のデータテーブルを統一し、データを From-To 形式に変換することでデータ量の削減に努めた。データ処理後のテレビ ID 数、データ量、レコード数を表 5 に示す。テレビ ID 数は、全メーカーを対象にデータ取得を行っていた A 局と B 局では約 17-19 % 減少している。また、全放送局のデータのうち、各放送局が独自で取得しているデータを削除し、共通項目だけを共通テーブルとすることでデータ量・レコード数ともに削減できた。

表 5: データ処理後の在阪 4 局視聴履歴データ比較

放送局	テレビ ID 数	データ量	行数
A 局	3,760,849	72GB	4.0 億行
B 局	3,727,169	104GB	5.7 億行
C 局	3,214,689	60GB	6.2 億行
D 局	2,604,325	64GB	5.0 億行

NNTM アルゴリズムにおける検討事項 2.3 節にて提案した手法について、実データを参照しながら提案手法の詳細条件を確定させていく。なお、参照するデータ

は在阪視聴データ連携技術実験で得られたデータのうち、2021年10月4日～10月17日のデータとする。

**検討事項1：マッチング回数** NNTMアルゴリズムでマッチングした異なる放送局のテレビは1回のマッチングで同一テレビと推定しても良いのか検討する必要がある。その為には、IPアドレス以外の要素を用いることでテレビそのものをどれだけユニークと識別できるのか調査する。具体的には、IPアドレスが変更されないと想定される短期間における「IPアドレス+郵便番号+メーカーID+ブラウザメジャーバージョン+ブラウザマイナーバージョン」で分離できるテレビの割合を調べることで、インターネットに結線されたテレビとIPアドレスやその他データの関係性を調査する。2.3.2項のStep2に前述の通り、同一集合住宅内で同一メーカー同型式テレビを保有する世帯が多くなるほどテレビを分離できなくなり、チャンネル遷移タイミングが一致するテレビが多数生じる可能性が高いため、一定期間内における最低マッチング回数などを検討する必要がある。そこで、実データを用いて「IPアドレス+郵便番号+メーカーID+ブラウザメジャーバージョン+ブラウザマイナーバージョン」で分離できるテレビの割合を調査し、マッチング回数の検討と決定を行う。なお、本調査は2021年10月5日19時～21時のデータに対して実施した。

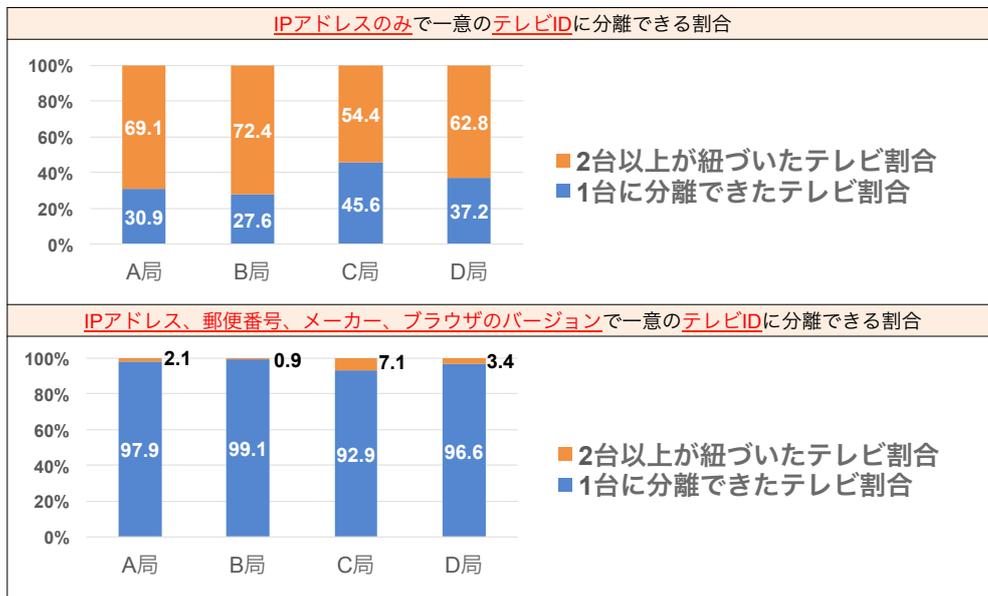


図 8: 項目別のテレビ分離割合

調査結果を図 8 に示す。図 8 の上部は「IP アドレス」のみでテレビを分離した結果を示している。その結果、IP アドレスとテレビ ID が 1 対 1 になるテレビは 54-72 % であった。そして、同一のデータに対して「IP アドレス+郵便番号+メーカー ID+ブラウザメジャーバージョン+ブラウザマイナーバージョン」で分離した結果を図 8 下部に示している。その結果、93-99 % のテレビが一意的に分離されることを確認できた。本データより「IP アドレス+郵便番号+メーカー ID+ブラウザメジャーバージョン+ブラウザマイナーバージョン」で分離をすることでほとんどのテレビを一意的に特定することができ、同一集合住宅において 1 つの IP アドレスを複数宅で共用していることにより起こる問題も解決できる。更に NNTM アルゴリズムでは、これらの分離したデータに対して更にチャンネル遷移タイミングをマッチング条件に加えるため、4 局データにおいてはマッチング数を 1 回に設定すれば良いという知見が得られた。

**検討事項 2：猶予時間** タイムマッチングにおける視聴終了時刻と視聴開始時刻の猶予時間を検討する必要がある。その為に、全放送局の組み合わせにおいて、複数の猶予時間でのマッチング数を調査する。条件として、視聴終了時刻と視聴

開始時刻が同一時刻であることが望ましいが、実データではプログラム起動ラグや、表3記載の通りビーコン方式にはデータ取得間隔があり、データの正確性に差が存在するため、猶予時間を設ける必要がある。

そこで猶予時間を設定することで、より正確なマッチングを目指す。ただし、この猶予時間を取り過ぎると、図9のように、同じ集合住宅に住んでいる同一メーカー同一機種を所有している世帯間において、偶然チャンネル遷移のタイミングが一致する誤マッチングが増えてしまうため、各社の視聴履歴データ特性から最適な猶予時間を検討し、実データから最適な猶予時間を確認する。

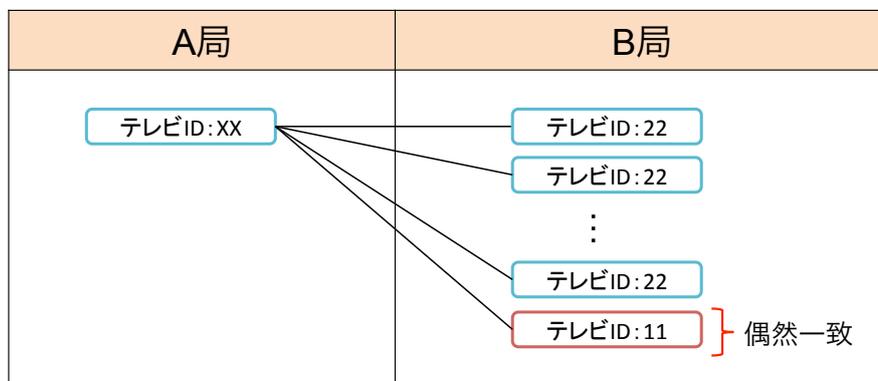


図9: A局1台に対してB局2台紐づいたパターン

表3の通り、A局からその他局へチャンネル遷移する組合せの相性が悪いため、許容誤差は±60秒が最適だと推測される。そこで実データを用いて、各局データのマッチング数を0秒、15秒、30秒、60秒、120秒で分析を行った結果を表6に示す。

表 6: 許容誤差別の放送局 1 対 1 マッチング数

組合せ	許容誤差				
	±0 秒	±15 秒	±30 秒	±60 秒	±120 秒
A 局 ⇄ B 局	3,194	911,032	1,087,265	1,187,329	1,219,404
A 局 ⇄ C 局	2,847	1,122,730	1,242,122	1,300,492	1,310,827
A 局 ⇄ D 局	3,523	1,222,884	1,267,231	1,302,679	1,304,329
B 局 ⇄ C 局	4,880	1,368,043	1,422,740	1,429,611	1,425,659
B 局 ⇄ D 局	8,047	1,236,956	1,296,801	1,309,869	1,310,638
C 局 ⇄ D 局	3,732	1,351,503	1,389,753	1,382,557	1,364,098

この結果から、事前推測していた許容誤差± 60 秒でマッチング数の増加率が鈍化していることが確認できる。また、表 7 の通り、1 対多のマッチング数は許容誤差を拡大すると単調に増加していくことが確認できる。

表 7: 許容誤差別の放送局 1 対多マッチング数

組合せ	許容誤差				
	±0 秒	±15 秒	±30 秒	±60 秒	±120 秒
A 局 ⇄ B 局	475	83,328	136,388	193,984	246,993
A 局 ⇄ C 局	873	130,394	194,810	266,340	343,817
A 局 ⇄ D 局	513	130,243	176,600	236,494	301,756
B 局 ⇄ C 局	1,374	107,988	162,100	225,706	298,432
B 局 ⇄ D 局	817	82,319	126,824	181,890	246,738
C 局 ⇄ D 局	1,652	141,589	211,027	303,482	423,896

以上の結果より、4 局データにおいては猶予時間を± 60 秒と設定すると良いという知見が得られた。

**検討事項 3：一意特定処理** 2.3.2 項の Step3 に前述の通り、マッチングするために使うデータ期間を延ばすほど、1 台に対して複数台が紐づいてしまうことが想定される。そこで条件式 (1) を一意特定の閾値として設定した。

本項では、実データへ条件式 (1) を適用し、一意特定処理を実施することで、一意特定に成功するテレビ ID 数を調査する。

実データへの適用した結果を表 8 に示す。

表 8: 一意特定処理数

組合せ	特定台数	1 対多台数	特定率
A 局 ⇔ B 局	37,410	193,984	19.3 %
A 局 ⇔ C 局	57,939	266,340	21.8 %
A 局 ⇔ D 局	57,152	236,494	24.2 %
B 局 ⇔ C 局	44,204	225,706	19.6 %
B 局 ⇔ D 局	36,792	181,890	20.2 %
C 局 ⇔ D 局	91,653	303,482	30.2 %

表 8 の通り、どの組合せにおいても約 20–30 % の一意特定に成功した。

なお、IP アドレスは時間と共に変化するが、本方式では個社が独自で割り振っているテレビ ID をキーとして使い、さらに同一時間帯のチャンネル遷移タイミングでマッチングしているため、同じテレビ ID に対して紐づく IP アドレスが途中で変更になったとしてもマッチング数への影響は少ないと想定される。

**NNTM アルゴリズム適用結果** 在阪視聴データ連携技術実験で交換した 2021 年 10 月 4 日～2022 年 1 月 5 日までの全ての実データを利用して、NNTM アルゴリズムを適用した結果を表 9、図 10 に示す。結果は、全ての局を特定できた「4 局マッチング数」と 1 つの局だけ特定できなかった「3 局マッチング数」と 2 つだけ特定ができた「2 局マッチング数」で示す。また、適用期間別のマッチング数を確認することで、最低限マッチングに必要な期間を調査する。

表 9: 期間別のマッチング数

期間	4局 マッチ数	3局 マッチ数	2局 マッチ数	合計
Day1	305,241	316,252	447,669	1,069,162
Day7	1,027,334	410,474	409,591	1,847,399
Day10	1,107,861	421,441	402,491	1,931,793
Day14	1,207,100	437,224	391,053	2,035,377
Day20	1,270,870	455,894	393,352	2,120,116
Day30	1,361,738	484,211	390,990	2,236,939
Day40	1,410,476	503,561	399,475	2,313,512
Day50	1,453,540	520,780	407,639	2,381,959
Day60	1,488,446	534,054	416,394	2,438,894
Day70	1,518,303	550,592	428,481	2,497,376
Day80	1,539,263	570,155	441,695	2,551,113
Day90	1,579,022	605,566	455,274	2,639,862
Day94	1,592,786	617,408	458,872	2,669,066

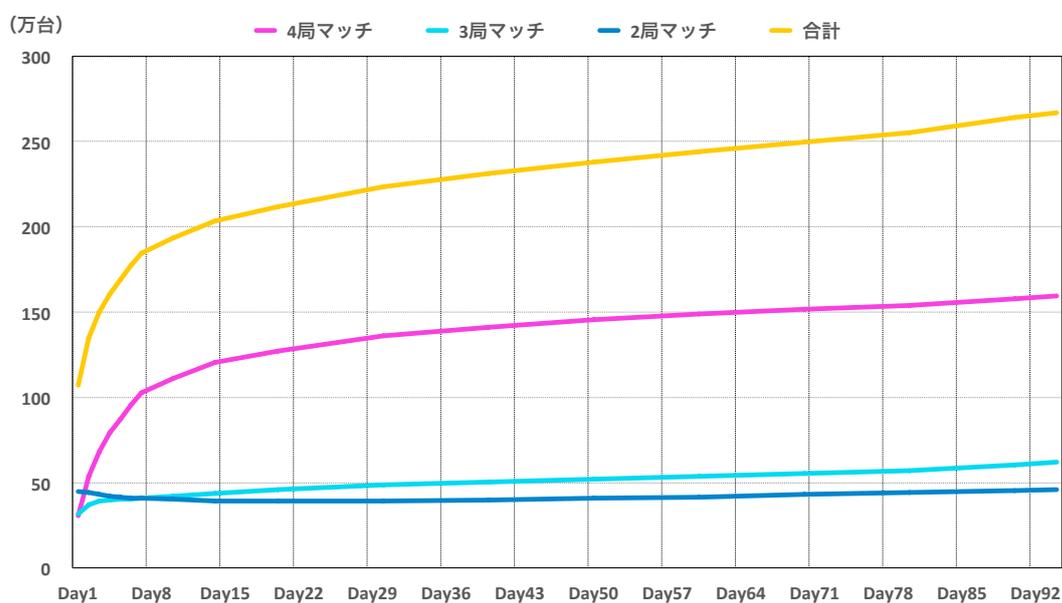


図 10: 期間別のマッチング数

94日分のデータを突合することで、約159万台の4局特定、約267万台の2局以上のマッチングに成功した。在阪4局のテレビID数は表5の通りだが、これらのテレビID数は各放送局単位で計測できているID数ではあり、各テレビ受像機が4つの放送局をすべて視聴しているのかは分からない。そのため、4局特定に至った約159万台の再現率は算出できない。2局以上の特定に成功した約267万台は、近畿2府4県の約922万世帯（R2国勢調査）と2人以上世帯の平均テレビ保有台数2,076台から推定される近畿に存在する約1,915万台の約13.9%に該当する。分析目的により、4局特定IDもしくは2局以上特定IDを利用するかは変わると考えられるが、これらは放送局が放送サービス向上などのためにマーケティング利用するには十分な台数である。また、図10のグラフよりマッチング台数の伸び率が減少し、伸び率が一定に近づくためには30日程度のデータ突合が必要であることがわかる。

大阪には5つの民間放送局が存在しているが、今回の在阪視聴データ連携技術実験は放送エリアが揃っている4つの民間放送局で実施されている。また、本取組みの中で、各局が記録しているIPアドレスの取得タイミングも視聴開始時、視聴終了時、次回視聴開始時など揃っていないことが判明した。

次節以降で、本アルゴリズムの性能評価を実施する。

## 2.5 テレビ視聴行動シミュレータ

NNTMアルゴリズムの性能評価を実施するため、本節ではテレビ視聴行動を再現し視聴履歴データを合成するための「テレビ視聴行動シミュレータ」を設計・実装する。

### 2.5.1 シミュレータの全体構成

本性能評価を実施するためのシミュレータの全体構成を図11に示す。エージェントモデルは、視聴者モデルとテレビ受像機モデルから構成される。視聴者モデルは、非特定テレビ視聴履歴データと国や地方自治体による住宅等の調査結果から設計し、テレビ受像機モデルは非特定テレビ視聴履歴データから設計する。シ

ミュレータはエージェント生成後、視聴履歴データによって設計されている視聴状態の遷移確率に沿って視聴行動を生成する。次に各エージェントの視聴行動から、オブザーバモデルが視聴放送局の特性に応じた視聴履歴データを合成する。

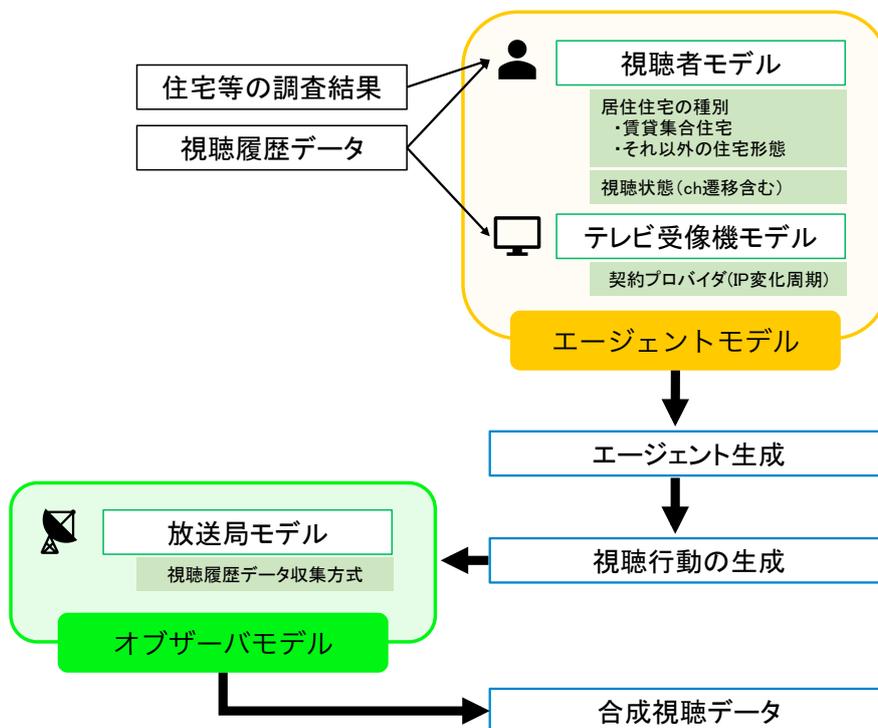


図 11: シミュレータ処理の流れ

## 2.5.2 視聴者モデルの設計

視聴者モデルを設計するにあたり、テレビ受信機の状態遷移確率と住宅等の調査結果を用いた。

■ 居住実態に基づく IP アドレスの割当 テレビ視聴履歴データは、インターネット結線されたテレビ受像機を対象にデータを収集している。一般的にテレビ受像機は家庭のインターネット回線を経由して結線されており、テレビ視聴履歴データではグローバル IP アドレスを取得している。グローバル IP アドレスは、定期的に変更されていき、そして集合住宅では 1 つのグローバル IP アドレスを複数の

住居で共用していることが多い。そこでエージェント生成時に、各エージェントが賃貸集合住宅に居住しているのか、もしくは戸建住宅や分譲集合住宅等に居住しているのか割り振る。また、共同住宅（分譲）と共同住宅（賃貸）でそれぞれの程度の割合で IP アドレス共用しているのかを調査するため、Yahoo クラウドソーシング<sup>3</sup>を用いて、分譲・賃貸住宅に居住している各 1,000 名を対象としたアンケート調査を実施した。調査結果を表 10 に示す。

表 10: 共同住宅別の IP アドレス共用数（アンケート）

	個別契約	大家・住宅会社等による契約
共同住宅（分譲）	700	300
共同住宅（賃貸）	682	318

次にエージェントを割り振るにあたり、総務省統計局による住宅・土地統計調査 [12] から共同住宅の割合、東京都都市整備局によるマンション実態調査 [1] から共同住宅における分譲と賃貸の割合を表 11、表 12 の通り算出した。

表 11: 2018 年住宅の建て方別住宅数（全国）

	一戸建	長屋建	共同住宅	計
戸数（万）	2,876	141	2,334	5,351
割合	53.8%	2.6%	43.6%	100%

表 12: 2013 年マンション総数（東京）

	分譲マンション	賃貸マンション	計
棟数	53,213	79,975	133,188
割合	40.0%	60.0%	100%

表 11 の共同住宅を表 12 の割合で割り振り、更に表 10 の結果を各共同住宅に適用することで簡易的に一戸建、長屋建、分譲共同住宅、賃貸共同住宅のうち、

<sup>3</sup><https://crowdsourcing.yahoo.co.jp/>

IP アドレス共用世帯の割合を表 13 の通り算出した。その結果，分譲共同住宅と賃貸共同住宅の IP アドレス共用世帯合計である 13.5%は IP アドレスを共用している集合住宅と推計できるため，エージェント生成時はこの割合に準じて住居区分を割り振ることとする。

表 13: 住宅種類別の割合

	戸数（万）	割合
一戸建	2,876	53.7%
長屋建	141	2.6%
分譲共同住宅 IP 非共用	652.4	12.2%
分譲共同住宅 IP 共用	279.6	<b>5.2%</b>
賃貸共同住宅 IP 非共用	956.2	17.9%
賃貸共同住宅 IP 共用	445.8	<b>8.3%</b>
合計	5351	100%

次に，集合住宅が IP アドレスをどの程度の戸数で共用しているのかを推計する。東京都都市整備局によるマンション実態調査 [1] から分譲と賃貸集合住宅それぞれの戸数分布が公表されており，図 12, 図 13 の通りとなっている。IP アドレスの共用も図 12, 図 13 の分布に従って，ランダムに割り振りを実施した。ただし，4戸未満の共同住宅はほとんど存在しないことを考慮し，1~20戸では5戸以上を条件にしている。また，IP アドレスの共用も視聴履歴データから 100戸を超えることが稀であることから，100戸以上については全て 100戸として計算した。

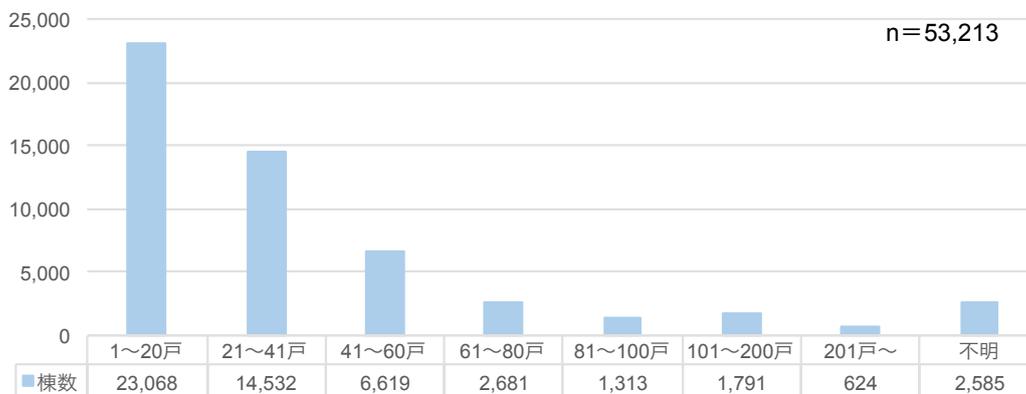


図 12: 分譲集合住宅の戸数分布（東京都）[1]

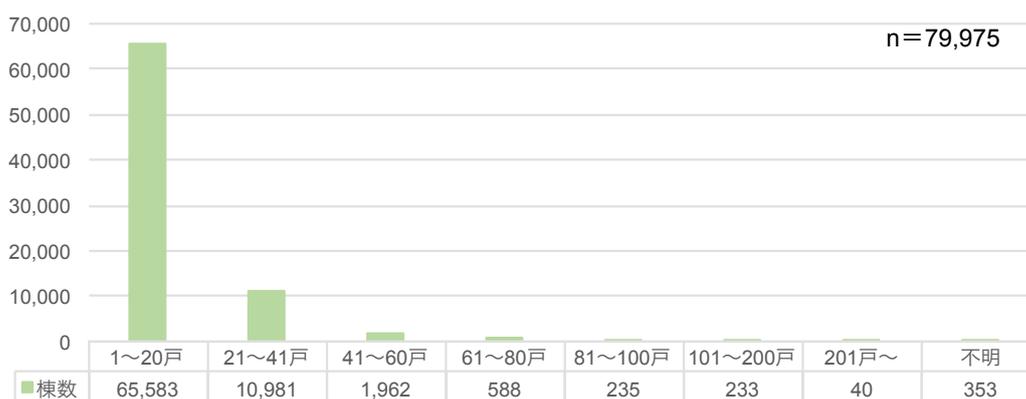


図 13: 賃貸集合住宅の戸数分布（東京都）[1]

■ **テレビ受信機の状態遷移確率** 在阪4局のテレビ視聴履歴データを分析することで各テレビ受信機が次の状態にどのような確率で遷移するのか分析を行った。本分析では、2022年11月7日～2022年12月4日までの4週間分のテレビ視聴履歴データから毎時00分～29分と毎時30分～59分までの30分単位に区切り、各30分から次の30分間にそれぞれのテレビ受信機がどの状態に遷移しているのか分析し、それを確率として算出した。その4週間分のデータを平日と土日でそれぞれ平均化した。また、その分析を4局のそれぞれにおいて実施した。

図 14 は、平日 18 時 30 分～59 分（30 分間）に視聴していたテレビがその後 4

時間の中でどのように状態遷移したかの状態遷移図をそれぞれの状態を起点にして1つの図に示したものである<sup>4</sup>.

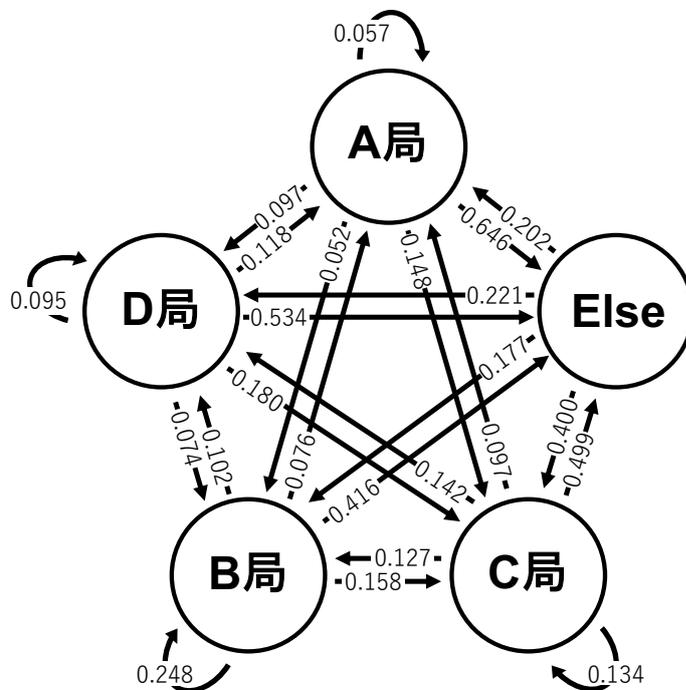


図 14: 各放送局の状態遷移図 (平日 18 時 30 分～59 分平均)

A局からの状態遷移は、4時間の間にA局を見続けている割合は約5.7%であり、4局以外の視聴であったり電源OFF状態であるElseへ遷移するテレビが約64.6%となっているということがわかる。各局からの遷移を比較すると、各局の状態遷移の傾向は類似しているが、差異が存在している。例えば、B局については視聴継続が多く、A局とD局は視聴継続が少ないことが分かる。シミュレータには、4週間分のデータから平日と土日でそれぞれ平均化した30分単位の状態遷移確率を使う。

<sup>4</sup>Elseは放送局からの離脱先・流入元を示す状態であるため、自己ループするデータは存在しない。

### 2.5.3 テレビ受像機モデルの設計

前述の通り、グローバル IP アドレスは定期的に変更されていく。実際の視聴履歴データを分析した結果でも、同一テレビ受信機の IP アドレスは一定ではないことが分かっている。これは視聴者の契約しているプロバイダなどの諸条件により、様々な周期で変化している。

テレビ受像機モデルでは、IP アドレスの変化周期を実際のテレビ視聴履歴データから調査分析した結果を実装する。分析データは、在阪4局のテレビ視聴履歴データから 12,000 台を無作為抽出したものを利用し、調査期間は4週間とする。

まず、4週間の IP アドレス変化回数の相対度数分布を図 15 に示す。全体の約 70%のテレビ受像機において IP アドレスは変わっていないことがわかる。

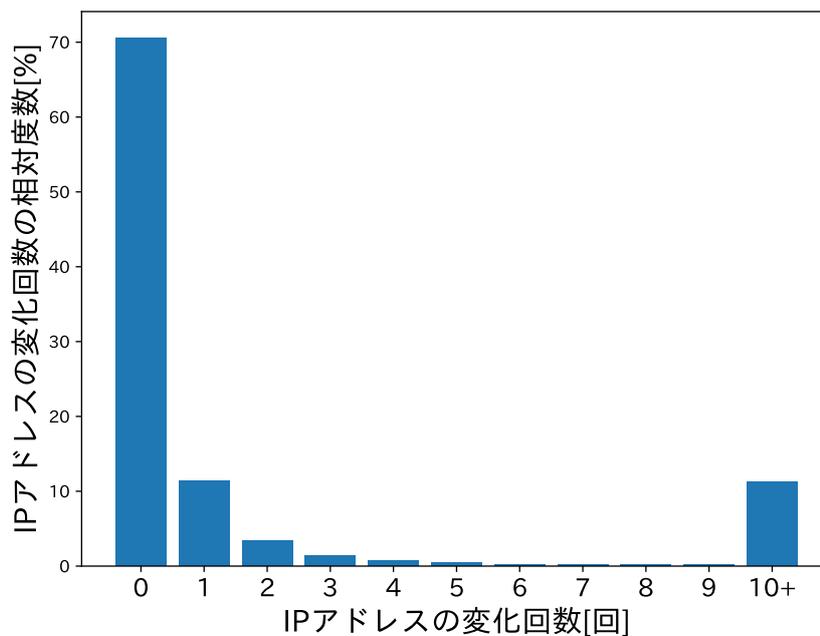


図 15: IP アドレス変化回数の相対度数分布

次に IP アドレスが変わらなかった端末と 10 回以上変化した端末を除いたグラフを図 16 に示す。また、図 16 には変化回数別の変化までの IP アドレス継続時間の平均値と中央値も示している。

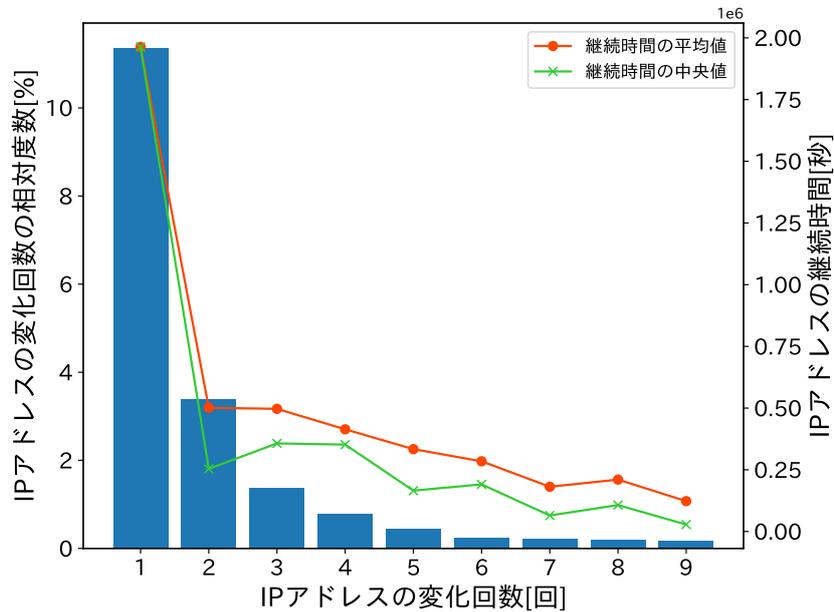


図 16: IP アドレス変化回数の相対度数分布と変化時間（1 回以上を抜粋）

テレビ受像機モデルでは、本分析によって得られた IP アドレス変化回数の分布に沿ってテレビエージェントを生成し、各 IP アドレス変化回数ごとの平均継続時間で IP アドレスを変化させていく。

このテレビ受像機特有の特徴を再現するテレビ受像機モデルを構築し、前述の視聴者モデルと組み合わせることでエージェントモデルを構築することで、エージェント生成を行うことが可能となる。

#### 2.5.4 放送局モデルの設計

本項では生成された視聴者行動から精緻な視聴履歴データを合成するために、放送局モデルの検討を実施する。放送局が収集する視聴履歴データは、データ放送プログラムを利用して収集しており、ビーコン方式と From-To 方式に分類され、表 3 のようになっている。ビーコン方式では、実際の視聴行動とそこから生成される視聴履歴データが一致しない場合がある。ビーコン方式は、視聴開始からビーコン間隔までの秒間のうち、ランダムでビーコン送信が開始される。これは、番組開始タイミングでサーバへの負荷を減らすために行っているが、ビーコン送信

が開始されるまでにチャンネル遷移が行われると視聴履歴データが生成されないことになる。このような放送局毎の特徴を考慮したモデルを構築することで、精緻な視聴履歴データを合成する。

## 2.6 シミュレータを用いた NNTM アルゴリズムの評価・考察

前節で設計・構築したテレビ視聴行動シミュレータから合成された視聴履歴データを用いて、本アルゴリズムの性能検証を実施する。

### 2.6.1 評価手法と結果

まず、テレビ視聴行動シミュレータを用いて 28 日分、つまり 4 週間の 250 万台分データを合成する。そして、その合成された視聴履歴データに対して、提案手法である NNTM アルゴリズムを適用し、4 局すべてのテレビ ID がマッチングした数を計上し、250 万台に対して特定できた割合を算出し、アルゴリズムを評価する。合成したデータは表 14 の通りとなっている。

表 14: テレビ視聴行動シミュレータを用いて合成した在阪 4 局の合成視聴履歴データの概要

放送局	行数
A 局	153,382,209 行
B 局	185,946,694 行
C 局	277,956,299 行
D 局	186,268,032 行

評価指標は、4 局すべてのテレビ ID 特定に成功した割合を再現率として用いる。具体的には、マッチング可能上限数である 250 万台に対して、本アルゴリズムで特定したマッチング数の割合を示している。まずは、28 日分、つまり 4 週間分のデータに対し NNTM アルゴリズムを適用し、再現率を調査する。また、図 4 の Step3 が ID 特定にどの程度寄与しているのか把握するために、Step2 終了時点

のデータのうち、1対1にIDが紐づいているデータのみを抽出してNNTMアルゴリズムを適用させ、再現率を調べる（Step3未実施）。その結果を表15、図17に示す。

表 15: 期間別のマッチング数と再現率

期間	提案手法（Step3未実施）		提案手法	
	マッチング数 <sup>1</sup>	再現率	マッチング数	再現率
Day1	792,397	31.7%	798,506	31.9%
Day2	1,620,850	64.8%	1,643,101	65.7%
Day3	1,980,176	79.2%	2,029,090	81.2%
Day4	2,103,395	84.1%	2,187,874	87.5%
Day5	2,142,410	85.7%	2,264,447	90.6%
Day6	2,154,163	86.2%	2,309,515	92.4%
Day7	2,157,772	86.3%	2,338,442	93.5%
Day14	2,159,041	86.4%	2,396,111	93.5%
Day21	2,159,024	86.4%	2,406,955	96.3%
Day28	2,159,022	86.4%	2,412,918	96.5%

<sup>1</sup> Step2を適用後に、1対1で紐づいているもののみを抽出した場合

表15の通り、再現率は4週間分データでStep3をスキップして実施した場合（Step3未実施）に86.4%、Step3も含めて全て実施した場合に96.5%を達成しており、図4のStep3が再現率向上に大きく寄与している。また、図17のグラフより再現率の伸び率が減少し、伸び率が一定に近づくためには7日程度のデータ突合が必要であることがわかる。

## 2.6.2 議論と考察

本方式で実施している図4のStepのうち、Step3の効果を計測するためにStep3を実施せず、1対多にIDが紐づいているデータも全て利用した場合の影響を分析

する。

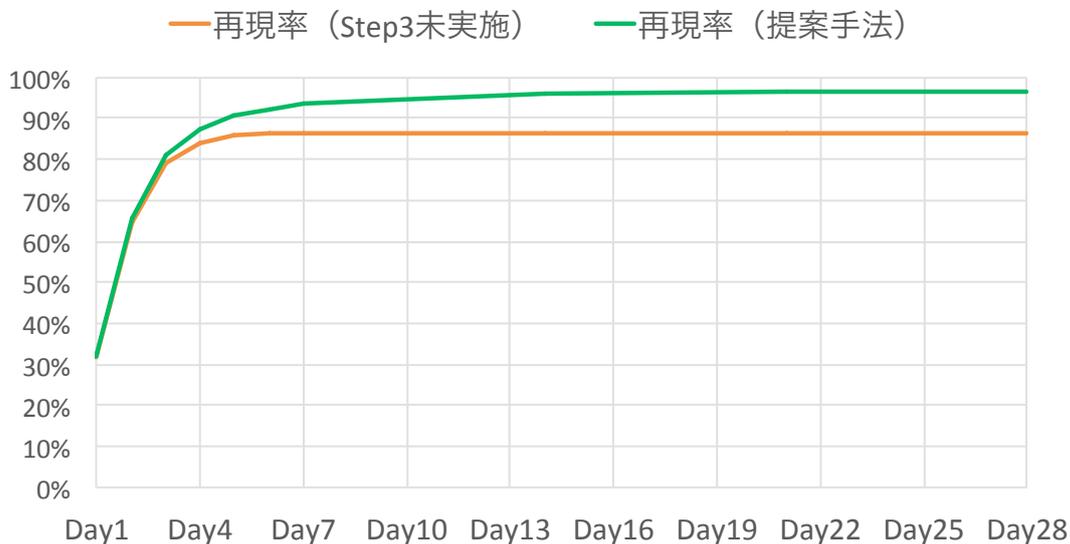


図 17: 期間別の再現率 (Step 別)

Step3 を実施せずに、1 対多に紐づいた状態で NNTM アルゴリズムを適用すると、組合せが非常に多くなるので、4 週間分のデータでは 4 局マッチング数は 19,360,441,222 台となり、適合率は 0.01%、再現率 100% となる。つまり、実際には正しい組み合わせではない多数のテレビ ID 同士がマッチすることとなるため、実用性が極めて低いといえる。そのため、1 対多に紐づいた ID は全て利用せず、1 対 1 に紐づいた ID のみを利用する条件を設定することも考えられる (=Step2 のみの場合)。この場合、表 15 の通り、4 週間分のデータでの 4 局マッチング数は、2,159,022 台となり、適合率 100% となるものの、再現率は 86.4% まで減少する。

これに対し、Step3 を適用することによって、Day28 における 4 局マッチング数の適合率 100% を維持しつつも、再現率 96.5% (Step2 のみと比較し 10.1 ポイント増加) を達成できることを確認した。このように本方式では、Step3 の処理を追加することでマッチング確度を維持した状態でのマッチング数の向上に成功している。

また、条件式 (1) は社会実装を想定し、適合率 100% を維持できるように厳しい条件を設定している。右辺の 1/3 を 0 に近づけていくことで一意特定台数は増え

るが、誤マッチングした組合せも増えてしまい、適合率が低下する。逆に1に近づけていくことで、適合率100%は維持できるが、再現率が低下すると考えられる。そのため、今回は経験的に1/3を設定した。今後は、条件を変更した評価が必要と考えられる。Step3で条件式(1)を利用せず、「1番多い組合せのみを利用する」という条件を採用することも可能だが、全マッチング数が少ない場合は、誤ったマッチングを採用してしまう可能性があり、適合率100%を維持できない可能性がある。

### 2.6.3 提案シミュレータの制限

提案シミュレータでは、エージェントの住居区分割当において、東京都が公開している住宅等に関する調査結果を利用しているため、都市圏を想定したシミュレータとなっている。都市圏以外の地域にも適用するためには、その地域の住宅区分データが必要となる。

また、シミュレータに実装する放送局数を4局としている。この放送局数は、シミュレータの入力データであるテレビ受信機の状態遷移確率データに依存する。局数が増える場合は、Else状態への遷移確率が減少し、局数が減る場合には増加することが想定される。しかし、Else状態への状態遷移確率が増減したとしても、残された放送局間での状態遷移確率は算出可能なため、シミュレータの構築には問題はない。

最後に、現在のシミュレータでは1世帯につき、インターネットに結線されたテレビが1台という想定となっている。インターネット結線率の向上とともに1世帯での複数台結線も増加することが考えられるので、今後のシミュレータでは、複数台結線を導入することを検討する。

## 2.7 まとめ

本章では、非特定視聴データ共同利用のボトルネックになっている「NVRAM放送事業者共通領域」を一切使用しない形での非特定テレビ視聴履歴データの統合手法を提案し、大阪にある4つの民放局の実データを利用した実践結果とシミュ

レータを利用した提案手法の評価を実施した。実践結果として、約3ヶ月間の非特定視聴データを用いることで在阪の4つの放送局テレビIDを約159万台マッチングさせることに成功した。また、マッチング数やマッチング精度の妥当性評価のため、テレビ視聴行動シミュレータを設計・実装することで本提案方式の評価を実施した。その結果、250万台のテレビの非特定視聴データを28日分合成し、このデータに提案方式を適用することで約241万台の4局のテレビIDをマッチングさせることができ、更に適合率100%、再現率96.5%であることを示した。本手法は、2.2.2項で研究されている異なるデバイスやブラウザを利用している同一人物を推定する分野等への応用も想定される。

現在、在阪民放局では本方式の社会実装に向けて、実験ではなく、定常的な非特定テレビ視聴履歴データの交換を視野に入れて検討が始まっている。しかし、社会実装され定常的にデータ交換を実施する場合には、本提案手法の実施頻度と利用データ期間が課題となるため、引き続き検討を進めていく必要がある。また、複数放送局の非特定テレビ視聴履歴データが統合されることで、k-匿名性の観点ではプライバシーリスクの増加が見込まれる。そのため、統合されたデータの取り扱いについては、総務省や個人情報保護委員会への説明が必要だと考えている。

## 3. 非特定テレビ視聴履歴データの分析

### 3.1 はじめに

非特定テレビ視聴履歴データは、視聴者の同意を取得している個人情報ではなく、個人関連情報に該当するため、他の個人情報を突合することは出来ない。また、SARCのプラクティスにより、非特定テレビ視聴履歴データを外部事業者に渡して分析を外部委託することが禁止されており、更に個人情報ではない他データとの突合にも制限が設けられている。放送局では、データ分析や調査会社や広告代理店に依存してきた歴史があり、自社で分析する技術や体制が整っていないのが現実である。そのため、非特定テレビ視聴履歴データの収集を開始しているが、広告価値の可視化や番組制作への活用には至っていない。

そこで本章では、効果的なCMを放送する、さらに視聴者数を維持・向上させるキャストイングすることを目的に非特定テレビ視聴履歴データの分析を行う。具体的には、目的達成に向けた研究課題（リサーチクエスチョン）を設定し、解決を目指し4つの分析を実施する。

- 分析1：テレビのCMに広告効果はあるのか
- 分析2：CM効果に地域差は存在するのか
- 分析3：番組出演者は視聴台数にどの程度影響があるのか
- 分析4：番組ジャンル毎にどのような興味関心を持った方が視聴しているのか

### 3.2 関連研究

本研究では、テレビ視聴行動と視聴後のインターネット行動について分析する。そこで、テレビの報道内容・時間と視聴行動の関係に触れた関連研究について紹介し、本研究との差異や位置付けについて論じる。

### 3.2.1 オプトイン型非特定テレビ視聴履歴データに関する研究

視聴者のテレビ視聴行動に関する分析は従来より関心が高く、テレビメーカーや各テレビ局、第三民間企業等における分析が進められてきた。

菊池ら [13] は、東芝製ネットワーク対応テレビから得られた全国 24 万台分のオプトイン型非特定テレビ視聴履歴データを用いた分析を実施している。菊池らは、「番組をどう視聴しているか」「どんな人々が視聴しているか」という観点に着目している。1つ目の「番組をどう視聴しているか」について、番組ジャンルによる視聴傾向の違い・連続ドラマの視聴傾向・番組の視聴者ロイヤリティ性を明らかにしている。具体的には、ドラマ視聴者は長時間視聴する傾向が強いが、音楽番組は短時間視聴の傾向が強く、更に番組中の視聴者の入れ替わりも激しい。2つ目の「どんな人々が視聴しているか」については、限られたアンケートによる視聴者属性情報から機械学習を用いた視聴世帯属性の推定を行っている。水岡ら [14] は、同データを用いたテレビ視聴パターンの推移を分析している。各テレビから曜日（月曜～金曜）と時間帯（1時間単位）ごとにライブ視聴時間の割合を求めて特徴量とし、クラスタリングを行っている。その結果、代表的な視聴パターン判定を実現しており、番組やCM制作への活用可能性を示した。

これらの先行研究では、視聴者がどのようにテレビを視聴しているのか、またその視聴者はどのような特徴を持っているのかを明らかにすることで視聴行動の可視化を実現している。

### 3.2.2 テレビのCM視聴に関する研究

テレビのCMが購買行動や製品に対する印象に与える影響は、宣伝効果の向上やスポンサ契約の拡大のためマスメディアの発展とともに研究が進められている。例えば Marc らの研究 [15] では、ユーモアを取り入れた広告がそうでない広告と比べて視聴者の興味関心を引くことができ、広告の認知度や好感度を高めることができること示している。牧野ら [16] は、親しみやすさが強く視聴者に肯定的気分を喚起させる要素が強いほど、CMへの好感度や購買意欲を呼び起こすと述べている。Campbell ら [17] は、馴染みのあるブランドについてのCMとそうでないCMを繰り返し放送した場合の効果について検証した結果、馴染みのあるブラン

ドのCMを繰り返したときに、視聴者を退屈させずメッセージの有効性が減少しづらいことを示した。さらに、同研究では広告の影響を調査する際に使われる、架空の商品について実際の商品と結果が変わる可能性を示唆した。

また、テレビのCMとwebサイトへのアクセスとの関連を調べた研究が複数行われている。Lukášら[18]は、1年間分のとあるECサイトのアクセス数とそのサイトに関するCMをテレビで放送した場合の関係を調べた。結果、CM放送後1分程度でアクセス数が増加すること、放送されたCMのうち1/3程度はアクセスに影響を及ぼさなかったことを示した。また、同研究では時間帯や曜日によるWebサイトのアクセス数の増減を調査しており、時間帯では夜21時頃にアクセス数が最大に深夜2時頃に最小となり、曜日に関しては大きな差が見られなかったと報告している。加藤ら[19]は、メーカーWebサイトへの誘引に効果的なテレビ広告の接触回数を評価した。その結果、3-10回の接触がターゲットとする顧客の誘導に最も効果的であり、それ以上の接触を行ってもターゲットに態度変容を促すことは難しいことが判明した。

また、永星[20,21]は、視聴者のザッピング（チャンネルの切り替え）行動を基に、現実の視聴者の視聴行動のモデル化を行い、各チャンネルの番組内容が同じである場合と異なる場合のそれぞれについて、視聴率を失わずにCMを放映できる時間帯を検討している。

また、横井ら[22]は、CMの挿入タイミングが心的状態に及ぼす影響について、脳活動の計測により得られる生理指標を用いて検証している。このようにCM視聴に着目した研究は存在するが実際のCM視聴データを用いた分析は行われていない。これは、非特定テレビ視聴履歴データの収集が始まってから日が浅く、また、基本的にデータが公開されていないためだと考えられる。

### 3.2.3 報道時間と視聴率の関係

報道時間と視聴率の関係について、岸本ら[23]が研究を行っている。岸本らは、特に新型コロナウイルス感染拡大状況におけるコロナ関連報道と視聴率の関係について分析を行なっている。その結果、2020年1月4日～2020年12月31日では、新型コロナウイルス感染拡大期（1月～4月）において、コロナ関連報

道時間の増加と視聴率の増加が同時に起こっていることを明らかにしている。しかし、5月以降においてはコロナ関連報道時間と視聴率に相関関係がなかったことを明らかにしている。テレビ局はより高い視聴率を得ることが一つのモチベーションであることから、視聴者の関心や行動を刺激するようなテレビ番組編成の工夫がなされていると想定される。岸本らの調査では、この工夫の一定の効果が示された一方、視聴者の関心は変化するため番組編成と視聴率の関係は経時で変化することも示された。テレビ局は高視聴率獲得のために色々と工夫はしているが、視聴者行動を把握しているわけではない。それは番組に限らず、CM視聴においても同様である。しかし、テレビ局はCMを放送して収益をあげているので番組やCM視聴者の行動を少しでも把握することでメディアとしてのテレビの価値向上を目指したい。

### 3.2.4 本研究の位置付け

先行研究では、メーカーが取得している視聴履歴データを用いて、視聴者がどのようにテレビを視聴しているのか、またその視聴者はどのような特徴を持っているのかを明らかにすることで視聴行動の可視化を実現している。次に、CM接触回数がWebサイト誘因に与える影響について最適な接触回数を示している。最後に、テレビ視聴ログの分析により視聴者を分類しどのような視聴行動を取っているか分析し、テレビ視聴者の関心が高い内容を報道することで視聴率が向上することを示した。

これに対し、本研究では、放送局が取得する非特定テレビ視聴履歴データを活用し、テレビのCM価値可視化や番組制作・編成業務への活用を目的に、視聴者行動分析を実施する。

## 3.3 分析1：テレビのCM視聴とインターネット検索データの関係性分析

本節では、テレビのCM視聴が視聴者の興味・関心を喚起し、実行動に結びつくことを明らかにする目的で、CM放送前後とインターネット検索行動量の推移

やインターネット検索行動量に影響の大きいCM放送時間帯を調査する。

### 3.3.1 データセット

本節では、非特定テレビ視聴履歴データ、テレビのCM放送実績データ、およびインターネット検索データの3つのデータを分析対象とする。以下では各データセットについての詳細について述べる。

■ **テレビの非特定テレビ視聴履歴データ** 本節では、読売テレビが独自で収集している非特定テレビ視聴履歴データを使用する。読売テレビが収集している非特定テレビ視聴履歴データは、サービス圏内である近畿二府四県にあるインターネット接続されたテレビのうち、オプトアウトしていないテレビが対象となっており、約350万台の規模となる。データにはIPアドレス、テレビデバイスID、郵便番号、視聴開始時刻、視聴終了時刻、テレビメーカーIDなどが含まれている。データ期間は2020年4月～2021年3月である。

■ **テレビのCM放送実績データ** 本節で使用するテレビのCM放送実績データは、実際にCM放送を実施している読売テレビから提供されたものである。これは読売テレビが放送実績を管理している営業放送システムからcsv形式で出力されたものである。データにはスポンサ名、スポンサ業種、CM名、放送開始時刻、CM尺などが含まれている。データ期間は2020年4月～2021年3月である。

■ **インターネット検索データ** ユーザがインターネット上で行う主要な行動の一つとして、本研究ではインターネット検索データに着目する。本論文では、Google社が提供しているサービスである、Googleトレンド<sup>5</sup>を利用した。Googleトレンドは、指定した地域と検索ワードにおける特定期間内の検索数を相対的に表示するサービスである。検索数の絶対値を示すものではないため、本論文ではGoogleトレンドから表示された値を以降では「検索量」と表記する。また、今回は検索地域を日本とし、検索ワードとして「スポンサ名」「スポンサ名+CM」、また商

---

<sup>5</sup><https://trends.google.co.jp/>

品名がある場合は「商品名」を指定している。データ収集期間は、テレビのCM放送開始日の1週間前からCM放送終了1週間後である。

### 3.3.2 分析

前項で述べた3つのデータセットを用い、テレビ視聴者のCM視聴後のインターネット行動を調査するための2つの関係分析を行った：【分析1】CM放送実績データとインターネット検索データの関係分析、【分析2】非特定テレビ視聴履歴データとCM放送実績・インターネット検索データの関係分析。以下では、各分析方法および分析結果について述べる。

■ CM放送実績データとインターネット検索データの関係分析 近畿でテレビ放送されたCM放送実績データと日本全国におけるインターネット検索データの関係进行分析した。

分析A：単発のCM放送の影響分析 まず最初に、単発のCM放送がインターネット検索行動に与える影響を把握するために、図18に示すように、CMが放送された後の24時間における検索量とCM放送がなかった時の24時間の検索量を比較する。比較には、対象期間におけるCM商品名でのGoogleトレンドの数値を1時間単位で取得し、その平均値を用いる。



図 18: CM 有無別の 24 時間以内の検索量取得方法

**分析 B：複数回の CM 放送の影響分析** 次に、短期間における連続的な CM 放送がインターネット検索行動に及ぼす影響を把握するために、過去 3 時間以内に複数回の同じ CM 放送が実施された場合のインターネット検索量を比較する。本論文で取り扱う CM 放送回数の判定・計数方法は次のとおりである。(1) 任意の連続的な 3 時間を切り出し、(2) 1 時間単位で当該 CM が 1 回以上放送されている場合に「CM が 1 回放送された」と判定、(3) CM ありと判定された数を計数する。図 19 に、3 回分の CM 判定例を示す。比較には、対象期間の 3 時間のうち最後の 1 時間における CM 商品名での Google トレンドの数値を取得しその値を用いる。

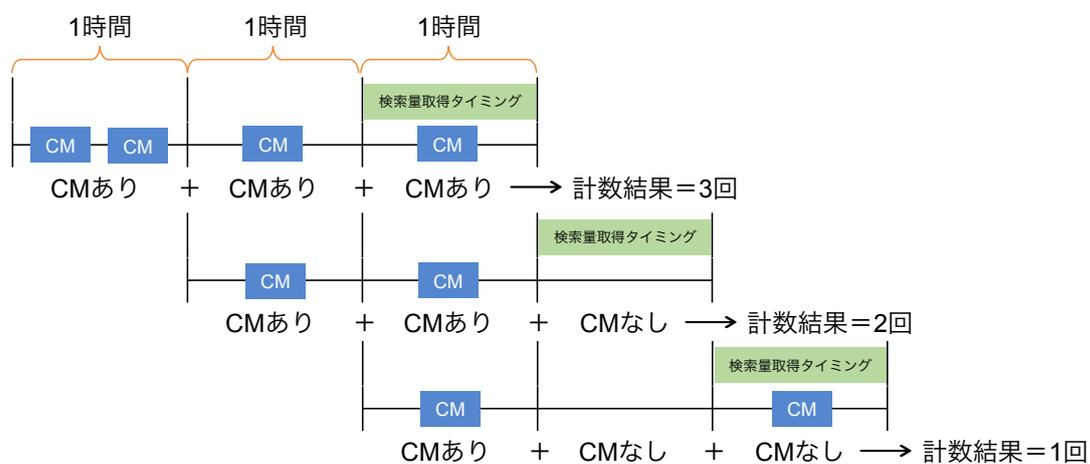


図 19: 連続的な 3 時間において CM 放送回数が計数される例

**分析 C：CM 放送期間・期間外の影響分析** 最後に、テレビでの CM 放送期間前・放送期間・放送終了後の 3 つの期間において、インターネット検索量を比較する。加えて、比較対象として CM で取り上げているサービス名とその運営会社の検索量についても比較する。本分析では、サービス名と運営会社名における Google トレンドの数値を 1 時間単位で算出し、それを 1 日毎の平均にまとめ、期間ごとに比較する。

**分析結果** 分析 A・B・C における分析結果について述べる。なお、分析 A・B・C のすべてのデータ取得で十分なサンプルを得られる対象 CM が本分析対象期間になかったため（放送回数が多く分析 B のデータ取得は可能だが、分析 A のデータ

取得においてCM放送がなかった日のサンプルが極端に少なくなってしまう、など)、それぞれの分析において別の対象CMを用いて分析を実施している。

**分析結果 A：単発のCM放送の影響分析** まず、単発のCM放送の有無による検索量の変化について結果を示す。図20は、24時間において、とある清涼飲料水のCMが放送されなかった場合と放送された場合の検索量の分布と平均検索量(mean)を示している。本分布は横軸にGoogleトレンドの検索量、縦軸にその頻度を示したヒストグラムであり、分布形状が右に偏れば偏るほど検索が通常と比較してより多くなされていることを示す。なお、取得したサンプル数はCMなしが2,141件、CMありが1,852件となった。図20の結果から、CMが放送されることで分布が右に移動し、平均値が23.9%増加していることが明らかとなった。

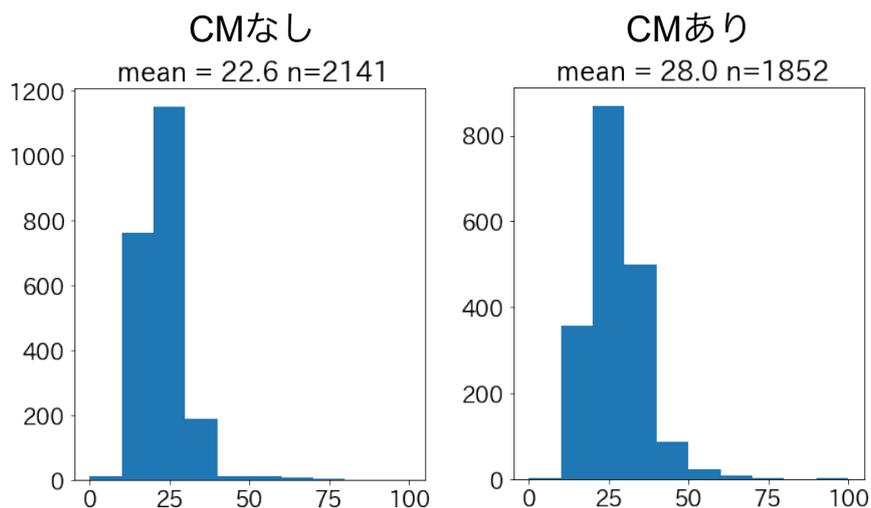


図 20: 24 時間以内の CM 有無別の検索量 (清涼飲料水)

**分析結果 B：複数回のCM放送の影響分析** 次に、短期間における連続的なCM放送による検索量の変化について結果を示す。図21, 22, 23は、アルコール飲料、清涼飲料水、ネット保険のCMにおける、過去3時間以内のCM放送回別の検索量の分布と平均検索量を示している。また、各CM回数別のサンプル数を表16に示す。図22, 23における放送3回はサンプル数が少なく、誤差が大きくなってしまったため、放送2回までを分析対象とする。アルコール飲料(図21)、清涼

飲料水（図 22）については，CM 回数が増加するとともに検索量の平均が増加していることが確認でき，放送 0 回から放送 1 回では 8.1%，17.9% の増加，放送 0 回から放送 2 回では 9.0%，31.9% の増加，アルコール飲料における放送 0 回から放送 3 回では 14.5% の増加が確認できた．また，ネット保険（図 23）は放送 0 回から放送 1 回では 2.7% の増加，放送 0 回から放送 2 回では 0.9% と増加となっているものの単価が低い飲料と比較すると CM 放送の影響が小さい結果となった．

表 16: 3 時間以内の放送回数別のサンプル数（CM 別）

CM 種類	0 回	1 回	2 回	3 回
アルコール飲料	3,681	454	178	56
清涼飲料水	3,385	505	97	6
ネット保険	3,520	792	54	3

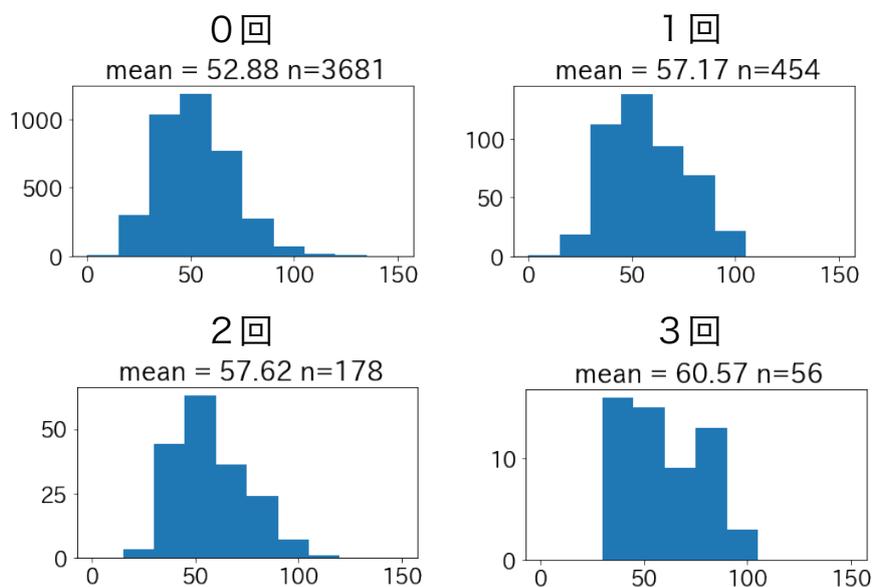


図 21: 3 時間以内の CM 回数別の検索量（アルコール飲料）

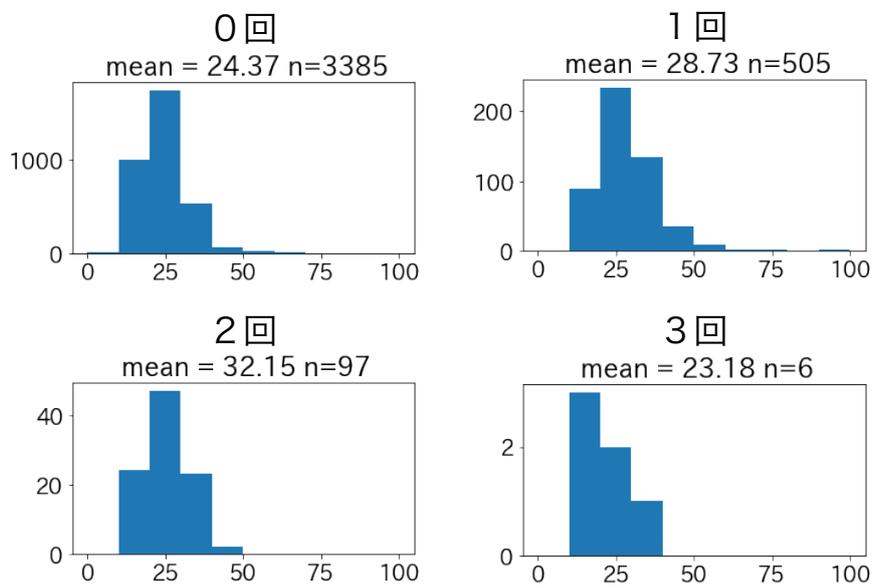


図 22: 3 時間以内の CM 回数別の検索量 (清涼飲料水)

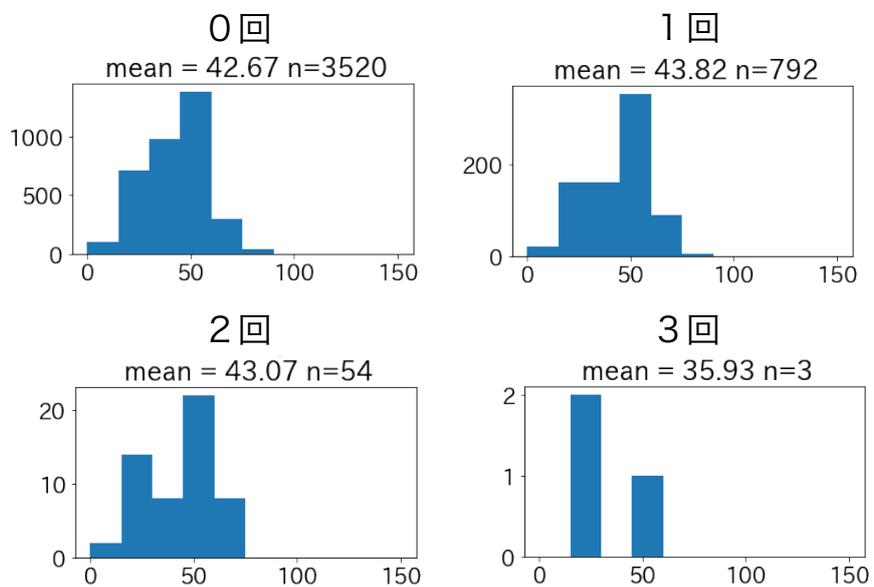


図 23: 3 時間以内の CM 回数別の検索量 (ネット保険)

**分析結果 C：CM 放送期間・期間外の影響分析** 図 24 に、オンラインゲーム運営会社が放送したオンラインゲームの CM 放送前後の検索量 1 日平均の推移を示す。CM 放送開始日から運営会社名の検索量には変化がないが、サービス名であるオンラインゲーム名だと検索数が向上していることが確認できる。また、サービス名では CM 放送終了後であっても高い水準の検索量を保持していることが確認できた。

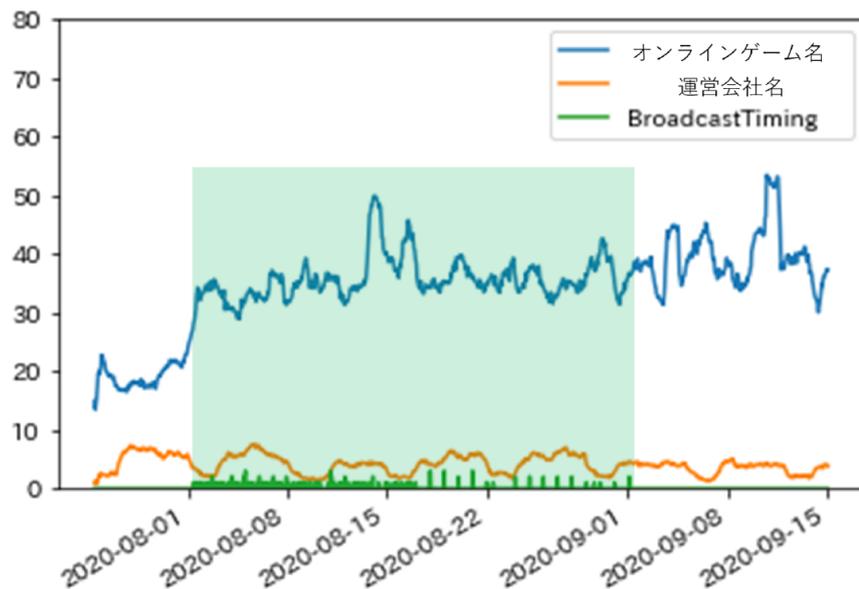


図 24: CM 放送前後の検索量推移

**考察** 図 20 で示した通り、テレビで CM を放送することにより視聴者が興味関心を持ち、インターネット検索行動に結びついていることが確認できた。また、図 21, 22 に示した通り、単価が低い商品の CM であれば、短期間に多数放送することで視聴者がより検索行動を行うことが確認できた。図 23 のネット保険は、単価が高く、購入頻度が少ないため、認知数拡大が検索行動へすぐには結び付かないと予想される。そのため、単価が低い商品と違い、3 時間という短期間においては CM 回数が増えても検索量が増えないことが確認できた。また、図 24 では、視聴者が最も関心を持つであろうサービス名において CM 放送開始後に検索量が増加し、関心が低いと想定される運営会社名では検索量に変化がなく、CM

の視聴が視聴者の検索行動に繋がっていると思われる。CM放送終了後においても、一度上昇した検索量は放送時と同水準を維持していることが確認でき、これはCM放送により認知が広がったためと推測される。これらのことから、本項ではCMの業種や種類によっては、CMを放送することでインターネットにおける検索行動に繋がることを示した。

今後の課題として、CMを商品単価による分類ではなく、広告目的別にCMを分類した分析が出来ていない。同様の低単価商品であっても認知拡大のためのCMと新商品発売告知のためのCMでは、視聴者の検索行動への影響に差異が発生することが想定される。しかし、本分析では読売テレビ単体の視聴履歴データを利用しており、他の放送局影響を排除するために、読売テレビのみがCMを放送したCMが分析対象となっている。そのため、対象となるCM数が少なく、広告目的別の分類による分析に至っていない。今後は、在阪局の統合された視聴データを利用し、分類方法を変えることで、より検索に結び付きやすいCMの種類や時間帯を明らかにしていく必要がある。

また、本分析ではCMの効果としてインターネット検索量を想定しているが、スポンサーの最終目的は商品の購入へ繋げることだと考えられる。そのため、将来的にはスポンサーや広告代理店と連携することで商品購入にまで至るまでの分析を目指したい。

**■ 非特定テレビ視聴履歴データとCM放送実績・インターネット検索データの関係分析** 近畿でテレビのCMを視聴したテレビ台数とCM放送時間帯、日本全国における検索データの関係性を分析した。

**分析方法** まず最初に、CM放送時間によりCM視聴台数がどのように変化するか把握するために、1時間単位のCM放送時間とCM視聴台数の分布を調べた。視聴台数は、当該CM放送時間における近畿の視聴台数となっている。

次にCM視聴台数によってインターネットでの検索量がどのように変化するか把握するために、各CMの視聴台数とそのCMの放送時間近辺における検索量の分布を調べた。本データの視聴台数はCM放送時間で計測しているが、検索量は当該CMが放送された日時の分数を切り下げた時間で計測している。表17に各

計測期間を例示する。

表 17: CM 放送時間と検索量計測期間の例

No	CM 放送日時	視聴台数	検索量計測期間
1	1/1 12:05:30	100,000	1/1 12:00-12:59
2	1/1 20:59:04	180,000	1/1 20:00-20:59
3	1/2 01:03:57	60,000	1/2 01:00-01:59

最後にテレビの CM 視聴が効率良くインターネットでの検索に繋がっている放送時間帯を把握するために、各 CM 放送時間を 1 時間単位で分類し、視聴台数 10,000 台あたりの検索量の分布を調べた。先ほどと同様に、本データの視聴台数は CM 放送時間で計測しているが、検索量は当該 CM が放送された日時の分数を切り下げた時間で計測している。

分析結果 図 25 は、食品メーカーの CM において、どの時間帯に CM 放送が実施され、それぞれの CM でどの程度の視聴台数だったのか分布を示している。深夜の時間帯は視聴台数が少なく、朝から夕方になるにつれて視聴台数が増加していることが確認できる。CM 放送をする際に、とにかくたくさんの視聴者に認知して貰いたい場合は、プライムタイムと呼ばれる 19-22 時が効率よくリーチを獲得できることがわかる。

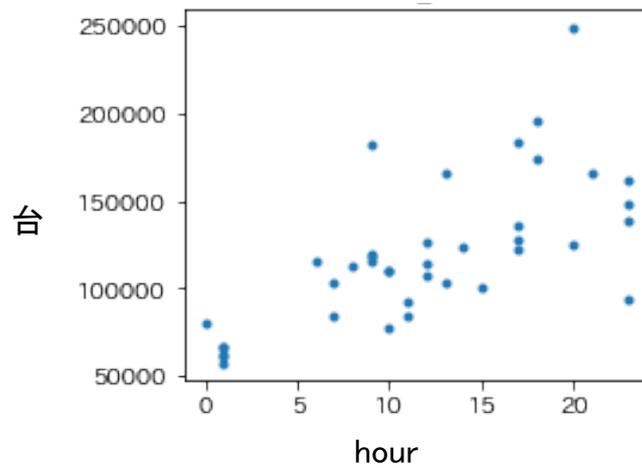


図 25: CM 放送時間帯（1 時間単位）と視聴台数の分布

図 26 は、先ほどと同様のメーカ商品の CM において、どの時間帯に CM が放送されており、それぞれの CM が放送された時間帯でどの程度のインターネット検索が実施されたのか分布を示している。図 25 と違いプライムタイムの検索量が多いわけではないことが確認できる。具体的には、CM 価値が低いといわれている深夜枠であっても検索量はプライムタイムと同等以上にされていることがわかる。

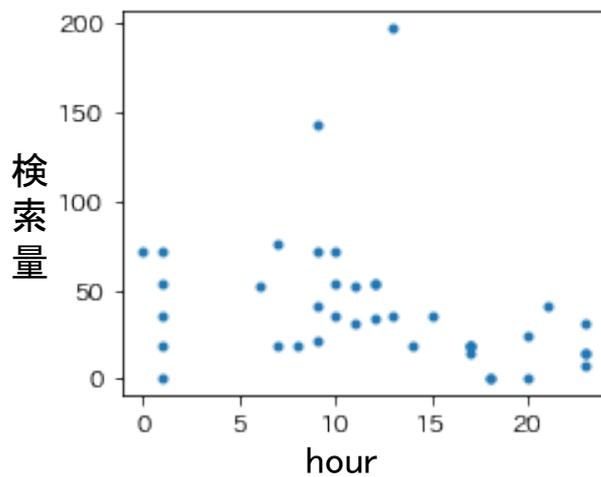


図 26: CM 放送時間帯（1 時間単位）と検索量の分布

最後に図 27 は、各 CM 放送時間を 1 時間単位で分類し、視聴台数 10,000 台あたりの検索量の分布を調べた。先ほどと同様に視聴台数が少ない深夜帯や午前・昼間においてはプライムタイムよりも効率よく検索行動に繋がっていることが確認できる。

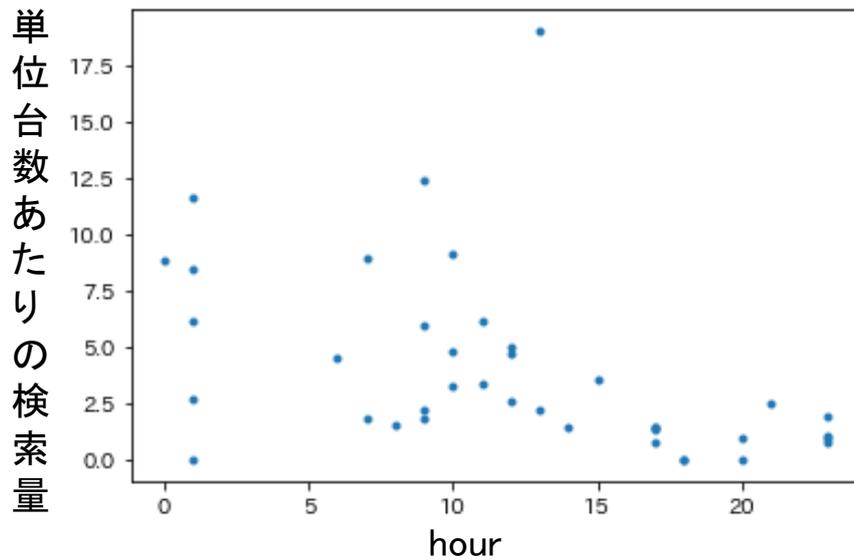


図 27: 視聴台数 10,000 台あたりの検索量と CM 放送時間帯の分布

**考察** 図 25 で示した通り、プライムタイムに放送した CM ほど視聴台数が多く、多くのリーチが獲得できることがわかる。しかし、図 26, 27 で示した通り、CM を視聴した後に検索行動に繋がっている時間帯はプライムタイムではなく、深夜帯や午前・昼間だということがわかる。これはテレビ視聴を行う際に、深夜帯や午前・昼間は「ながら視聴」が行われやすい時間帯だからだと考えられる。これらの通り、比較的広告価値が低いと考えられている深夜帯や午前・昼間であっても視聴者のインターネット検索行動はプライムタイムよりも CM と効率的に繋がっていることが示された。スポンサーは、テレビの CM に対して多種多様なニーズを持っており、その中には自社サイトへの誘引を目的にしていることもある。このようなスポンサーに対しては、本分析手法を使うことで CM 出稿への営業利用や、CM 放送後の効果検証に使うことができる。

### 3.3.3 まとめ

本節では、テレビでのCM視聴が視聴者の興味関心を引き起こし、実際のインターネット検索行動に繋がっているのかをデータ分析を行うことで示した。また、CMの視聴台数がCM放送時間により、実際のインターネット検索行動にどの程度影響を与えるのかを示した。本節は、これまでは困難であったテレビのCM効果を可視化するための一手段になり得るため、今後も様々な業種で同様にデータ分析することでテレビのCM価値の向上が期待できる。今後は、インターネット検索行動に繋がりがやすいCM種別の分析や、CMに限らない番組本編視聴が視聴者に与える影響の分析方法について、検討を進めていく。

## 3.4 分析2：CM視聴傾向の分析

テレビのCMは、民間放送局にとってビジネスの根幹を支える商品であるが、現状、視聴者のCM視聴状況を明確に示す指標は存在しておらず、指標の確立を含め、CM効果の分析・可視化が急務である。

本節では、CM効果の分析・可視化のために、「視聴UD（ユニークデバイス率）」を提案する。「視聴UD率」は、少しでもCM視聴したテレビ台数のうち、一定割合以上視聴したテレビ台数の割合と定義し、視聴UD率を地域ごとに算出する。「完視聴UD率」は、該当するCMにおけるUDのうち、完全に視聴したUDの割合を示し、「1/3以上視聴UD率」、「2/3以上視聴UD率」は、該当するCMを1/3又は2/3以上視聴したUDの割合を示すものと定義する。これに総務省統計局が発表している各種統計データを紐付けることで、CM視聴に対して地域傾向が存在するかを明らかにする。

### 3.4.1 データセット

本項では、本分析で使用するデータセットについて述べる。

データセットは、(1)非特定テレビ視聴履歴データ、(2)番組メタデータ、(3)統計データの3つのデータを分析対象とする。以下では各データセットについての詳細について述べる。

■ **テレビの非特定テレビ視聴履歴データ** 本節では、近畿一円の約350万台を対象として、在阪の4つの放送局（読売テレビ・毎日放送・朝日放送テレビ・関西テレビ）がそれぞれ独自に収集したオプアウト型非特定テレビ視聴履歴データを統合した視聴データを使用する。4つの放送局では取得形式や、振り分けられているテレビIDが異なることから、2.4節で実践された同一のテレビIDとして統合した非特定テレビ視聴履歴データを扱う。

■ **番組メタデータ** 非特定テレビ視聴履歴データは、テレビ受像機ごとの視聴開始と終了時刻を扱う履歴データであるため、番組に関する情報は含まれていない。そのため、読売テレビが保有するコーナーやCMが放送された時間帯などの番組の詳細がまとめられた「番組メタデータ」を用いる。番組メタデータは、各番組ごとの、番組タイトル、番組開始時刻、番組終了時刻、番組時間、番組ジャンル、コーナー開始時刻、コーナー終了時刻、コーナー時間、コーナージャンルなどがまとめられたデータである。本稿では、CMが連続して放送されるひとまとまりについてを対象とし、これを「CMブロック」と定義する。分析には、2021年10月から12月の3ヶ月間のデータを用いる。

■ **統計データ** 本分析では、総務省統計局が公開している「統計でみる都道府県・市区町村のすがた（社会・人口統計体系）」のうち、本研究で対象とする近畿二府四県の全ての市区町村における各種統計データを使用する。これらの統計データは、政府統計総合窓口サイト（e-Stat）<sup>6</sup>のAPI機能を使用して収集を行った。統計データには、割合などの相対値で表されているものと、絶対値で表されているものがある。絶対値で表されている項目に対しては、人口等の影響による偏りを無くすため、標準化した数値を扱う。

### 3.4.2 データの処理方法

本項では、本分析で使用するデータセットの処理方法について述べる。本分析では、CM視聴における傾向を分析するため、CM視聴UD率なる指標を提案す

---

<sup>6</sup><https://www.e-stat.go.jp/>

る。また、CMブロック放送時間内にテレビ画面をどれだけ切り替えたかを確認するための指標として、UDあたりch切替回数を提案する。これらの指標を求めるために行った非特定テレビ視聴履歴データと番組メタデータの処理方法、および、統計データと紐付けるために行った処理方法について述べる。

■ CM視聴UD率の算出 CM視聴UD率についての定義を示す。図28は本指標の概要を示している。

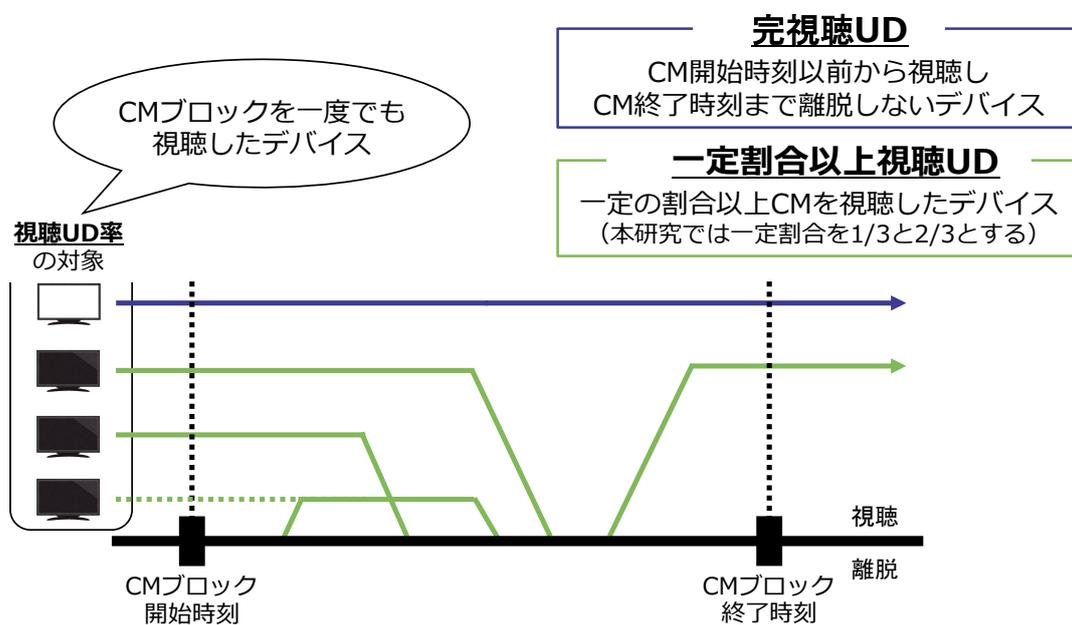


図 28: CM 視聴 UD 率の定義

CMブロック開始時刻、CMブロック終了時刻、視聴開始時刻、視聴終了時刻をそれぞれ  $T_s, T_e, t_s, t_e$  とし、非特定テレビ視聴履歴データにおいて、あるCMブロックにおける視聴レコードが、以下の3条件のいずれかを満たすデバイス（以下、「UD」とする）を対象とする。

$$T_s < t_s \quad \text{かつ} \quad t_e < T_e \quad (2)$$

$$T_s < t_e \quad \text{かつ} \quad t_e < T_e \quad (3)$$

$$t_s \leq T_s \quad \text{かつ} \quad T_e \leq t_e \quad (4)$$

UD に対し、完全に CM を視聴した UD の割合を示すものを「CM 完視聴 UD 率」、1/3 以上視聴した UD の割合を示すものを「CM1/3 以上視聴 UD 率」、2/3 以上視聴した UD の割合を示すものを「CM2/3 以上視聴 UD 率」と定義する。CM 視聴 UD 率を計算するためには、その CM ブロックにおける UD 数と、各 UD が視聴していた時間を抽出する必要がある。前述のように、非特定テレビ視聴履歴データには、CM に関連する情報は含まれていない。そのため、まず番組メタデータから (1) CM 情報を抽出し、そして、各 CM ブロックに対して非特定テレビ視聴履歴データから (2) UD 数 (3) 各 UD 視聴時間 (4) 完 (1/3,2/3 以上) 視聴 UD 数を算出し、CM 完視聴 UD 率、CM1/3 以上視聴 UD 率、CM2/3 以上視聴 UD 率を導き出す。以下で、それぞれの抽出方法について述べる。

**CM 情報** 番組メタデータのうち、コーナージャナルが CM であるレコードのコーナー開始時刻、コーナー終了時刻を CM 情報として抽出する。複数の CM が連続して放送される場合、コーナー開始時刻は最初の CM が放送され始めた時刻、コーナー終了時刻は最後の CM を放送し終えた時刻である。

**UD 数** CM 情報から抽出できる CM ブロック開始時刻及び CM ブロック終了時刻に対して、非特定テレビ視聴履歴データの視聴開始時刻、視聴終了時刻が、式 (2),(3),(4) のいずれかを満たすテレビを抽出する。また、郵便番号を抽出することで、地域別の UD 数を算出する。いずれも 2.3 節で提案されたアルゴリズムによって定められた一意の ID を用いて UD 数を計算した。

**各 UD における CM 視聴時間** 本稿では、CM ブロック開始時刻から CM ブロック終了時刻までにどの程度視聴したかを視聴の尺度としている。即ち、ある CM に対して UD であるとされたテレビのうち、CM ブロック開始時刻から CM ブロック終了時刻までの時間と重なる視聴レコードを抽出し、CM ブロック放送時間と重なる部分の総和を求める。

完（1/3,2/3以上）視聴UD数あるCMブロックに対して求めたUDのうち、CMブロック視聴時間がCMブロック放送時間と等しいものを完視聴UDとする。同様に、CMブロック視聴時間がCMブロック放送時間の1/3(2/3)以上である場合、1/3(2/3)以上視聴UDとする。また、郵便番号を抽出することで、地域別の完（1/3,2/3以上）視聴UD数を算出する。

これらによって、CM完視聴UD率、CM1/3以上視聴UD率、CM2/3以上視聴UD率は以下の式で計算される。

$$CM \text{ 完視聴 } UD \text{ 率} = \frac{\text{完視聴 } UD \text{ 数}}{UD \text{ 数}} \quad (5)$$

$$CM1/3 \text{ 以上視聴 } UD \text{ 率} = \frac{1/3 \text{ 以上視聴 } UD \text{ 数}}{UD \text{ 数}} \quad (6)$$

$$CM2/3 \text{ 以上視聴 } UD \text{ 率} = \frac{2/3 \text{ 以上視聴 } UD \text{ 数}}{UD \text{ 数}} \quad (7)$$

■ UDあたりch切替回数 UDあたりch切替回数は、該当CMブロックの放送時間に、1UDあたりどれだけのch切替回数があるかどうかを示す指標である。ch切替回数は、以下の条件式(8)を満たす視聴レコード数として数えることができる。これを用いて、式(9)によって計算される。

$$T_s \leq t_e \quad \text{かつ} \quad t_e \leq T_e \quad (8)$$

$$UD \text{ あたり } ch \text{ 切替回数} = \frac{ch \text{ 切替回数}}{UD \text{ 数}} \quad (9)$$

■ 統計データと非特定テレビ視聴履歴データの紐付け 統計局の統計データは、JISコードと呼ばれる市町村を一意に識別するコードごとにデータが集約されている。一方で、非特定テレビ視聴履歴データでは、地域情報として郵便番号7桁が収集されており、地域ごとにデータを処理する場合には、郵便番号をキーとする必要がある。そのため、これら2種類のデータを集約するためには、市町村名

と郵便番号による紐付けが必要となる。紐付けについては、郵便番号と市町村の対応を以下の順序で処理することで実現した。

1. 非特定テレビ視聴履歴データ上の郵便番号と、統計データ上の市町村に紐づく郵便番号が1対1に対応する場合、非特定テレビ視聴履歴データ、統計データのいずれもそのまま利用する。
2. 特定視聴履歴データ上の郵便番号に、統計データ上の市町村に紐づく郵便番号が複数対応する場合、該当する市町村の統計データを合算することで対応する。
3. 統計データ上では、市町村に対応する郵便番号が1つしかないため、非特定テレビ視聴履歴データ上の郵便番号に、統計データ上の市町村に紐づく郵便番号が存在しない場合が生じる。この場合、非特定テレビ視聴履歴データを近隣の郵便番号と紐付けて合算し、2種類のデータに漏れがないように調整した。

### 3.4.3 分析

3.4.1 節で述べた3つのデータセットを用い、CM視聴UD率と地域傾向の関係性について分析した。以下では、分析方法、分析結果及びその考察について述べる。

■ 分析方法 CM視聴UD率と地域傾向の関係性を分析する方法について述べる。まず、番組メタデータより、2021年9月から12月の間に4局で放送されたCMブロックの開始時刻と終了時刻を抽出する。次に、非特定テレビ視聴履歴データをJISコードごとに分け、各CMブロックにおける完視聴UD率、1/3以上視聴UD率、2/3以上視聴UD率、UDあたりch切替回数を算出する。これらの指標を、JISコードごとの平均と標準偏差を算出する。そして、JISコードごとに統計データの各項目と算出した指標との相関を確認する。

■ 分析結果 まず、JIS コードごとに算出した完視聴 UD 率、1/3 以上視聴 UD 率、2/3 以上視聴 UD 率、UD あたり ch 切替回数の平均値及び標準偏差における最大値、最小値、第 1 四分位数、第 3 四分位数、分散を表 18、表 19 に示す。

表 18: 視聴 UD 率の基本統計量

	1/3 以上視聴 UD 率 平均	1/3 以上視聴 UD 率 標準偏差	2/3 以上視聴 UD 率 平均	2/3 以上視聴 UD 率 標準偏差	完視聴 UD 率 平均	完視聴 UD 率 標準偏差
最小値	0.9479	0.0246	0.9122	0.0390	0.8785	0.0531
第一四分位数	0.9567	0.0290	0.9254	0.0436	0.8961	0.0576
第三四分位数	0.9597	0.0528	0.9299	0.0718	0.9015	0.0888
最大値	0.9667	0.2145	0.9392	0.2745	0.9140	0.3163
分散	7.564E-06	1.438E-03	1.670E-05	2.217E-03	2.632E-05	2.761E-03

表 19: ch 切替回数の基本統計量

	UD あたり ch 切替回数 平均	UD あたり ch 切替回数 標準偏差
最小値	0.0483	0.0363
第一四分位数	0.0604	0.0407
第三四分位数	0.0645	0.0674
最大値	0.0798	0.2633
分散	1.622E-05	1.971E-03

定義上、自明ではあるが、平均に関しては、1/3 以上視聴 UD 率、2/3 以上視聴 UD 率、完視聴 UD 率の順に割合が低下している。一度その CM ブロックを視聴したユーザは、94%以上がその CM ブロックの 1/3 以上を視聴することが確認できる。これらの平均の分散については、1/3 以上視聴 UD 率、完視聴 UD 率、2/3 以上視聴 UD 率の順に小さい。

次に、完視聴 UD 率、1/3 以上視聴 UD 率、2/3 以上視聴 UD 率、完視聴 UD 率、UD あたり ch 切替回数と 257 項目における統計データとの相関を確認した。表 20、表 21、表 22、表 23 は、1/3 以上視聴 UD 率、2/3 以上視聴 UD 率、完視聴 UD 率、UD あたり ch 切替回数の平均・標準偏差と各統計データの相関係数の分布を示す。平均については、正の相関が見られる項目 ( $0.4 \leq r < 0.7$ ) と負の相関が

見られる項目 ( $-0.7 < r \leq -0.4$ ) が数個確認された。その他の大半の項目は相関が弱い、もしくは相関が見られなかった。標準偏差については、大半の項目において負の相関が確認された。しかし、強い正の相関 ( $0.7 \leq r$ ) と強い負の相関 ( $r \leq -0.7$ ) が見られた項目も数個存在した。

表 20: 統計データ各項目と 1/3 以上視聴 UD 率との相関係数 ( $r$ ) の分布

	$0.7 \leq r$	$0.4 \leq r < 0.7$	$0.2 \leq r < 0.4$	$-0.2 < r < 0.2$	$-0.4 < r \leq -0.2$	$-0.7 < r \leq -0.4$	$r \leq -0.7$
平均	0	10	148	79	10	10	0
標準偏差	7	9	5	22	50	161	3

表 21: 統計データ各項目と 2/3 以上視聴 UD 率との相関係数 ( $r$ ) の分布

	$0.7 \leq r$	$0.4 \leq r < 0.7$	$0.2 \leq r < 0.4$	$-0.2 < r < 0.2$	$-0.4 < r \leq -0.2$	$-0.7 < r \leq -0.4$	$r \leq -0.7$
平均	0	9	150	78	10	10	0
標準偏差	8	8	5	22	52	159	3

表 22: 統計データ各項目と完視聴 UD 率との相関係数 ( $r$ ) の分布

	$0.7 \leq r$	$0.4 \leq r < 0.7$	$0.2 \leq r < 0.4$	$-0.2 < r < 0.2$	$-0.4 < r \leq -0.2$	$-0.7 < r \leq -0.4$	$r \leq -0.7$
平均	0	9	150	78	10	10	0
標準偏差	8	8	5	22	56	155	3

表 23: 統計データ各項目と UD あたり ch 切替回数との相関係数 ( $r$ ) の分布

	$0.7 \leq r$	$0.4 \leq r < 0.7$	$0.2 \leq r < 0.4$	$-0.2 < r < 0.2$	$-0.4 < r \leq -0.2$	$-0.7 < r \leq -0.4$	$r \leq -0.7$
平均	0	10	10	55	171	11	0
標準偏差	8	8	5	21	53	159	3

そこで、地域性による違いを確認するため、視聴 UD 率とより相関が見られた項目について確認する。表 24, 表 26, 表 28 は、1/3 以上視聴 UD 率, 2/3 以上視

聴UD率,完視聴UD率との平均と正の相関 ( $0.4 \leq r < 0.7$ ) がある項目を示している。これら3つの指標に共通して登場する項目として、人口増減率、生産年齢人口関連の割合、母子世帯割合が大きい地域では、これらの視聴UD率も大きいことが示された。表25,表27,表29は、1/3以上視聴UD率,2/3以上視聴UD率,完視聴UD率との平均と負の相関 ( $-0.7 < r \leq -0.4$ ) がある項目を示している。主に、65歳以上の人口割合や、第一次産業従業者比率など、高齢者層に関連する項目が共通して登場することが確認できた。なお、「最終学歴人口割合(小・中)」に関しては、80代以上が全体の9割を占めており、戦前の教育制度の違いや、戦後の混乱などが背景にあると考えられるため、高齢者に関連する項目として捉えることができると考えられる。表32,表34,表36は、1/3以上視聴UD率,2/3以上視聴UD率,完視聴UD率との標準偏差と正の相関 ( $0.7 \leq r$ ) がある項目を示している。これらは、平均で負の相関を持った項目と類似しており、65歳以上の人口割合や第1次産業事業所割合、年齢中位数など、年齢が大きくなると値が大きくなる項目が多く確認できた。表33,表35,表37は、1/3以上視聴UD率,2/3以上視聴UD率,完視聴UD率との標準偏差と負の相関 ( $r \leq -0.7$ ) がある項目を示している。18歳から64歳までの人口比率や、人口増減率が相関を持つことが確認できた。人口増減率と負の相関を持つとは、人口が減っている地域では標準偏差が大きくなる傾向にあるという意味であり、この項目も、高齢者が多く存在するという解釈ができるため、高齢層とそれ以外の層での関係性が確認できる。

表 24: 1/3 以上視聴 UD 率の平均と正の相関がある項目

統計項目	相関係数
18～39 歳人口割合	0.509
15～64 歳人口割合	0.508
18～39 歳人口割合 (男)	0.507
0～17 歳人口割合 (女)	0.500
18～39 歳人口割合 (女)	0.497
15 歳未満人口割合	0.492
人口増減率	0.468
0～17 歳人口割合	0.462
母子世帯割合	0.452
0～17 歳人口割合 (男)	0.41

表 25: 1/3 以上視聴 UD 率の平均と負の相関がある項目

統計項目	相関係数
高齢単身世帯割合 (女)	-0.426
第 1 次産業就業者比率	-0.455
高齢夫婦世帯割合	-0.458
高齢夫婦のみの世帯の割合	-0.458
65 歳以上の世帯員のいる核家族世帯割合	-0.472
65 歳以上人口割合 (女)	-0.54
65 歳以上人口割合	-0.559
65 歳以上人口割合	-0.559
65 歳以上人口割合 (男)	-0.568
年齢中位数	-0.597

表 26: 2/3 以上視聴 UD 率の平均と正の相関がある項目

統計項目	相関係数
18～39 歳人口割合	0.523
18～39 歳人口割合 (女)	0.519
18～39 歳人口割合 (男)	0.512
未婚者割合 (15 歳以上人口)	0.506
15～64 歳人口割合	0.504
母子世帯割合	0.466
完全失業率	0.423
人口増減率	0.41

表 27: 2/3 以上視聴 UD 率の平均と負の相関がある項目

統計項目	相関係数
就業者比率	-0.423
最終学歴人口割合 (小・中)	-0.429
65 歳以上の世帯員のいる核家族世帯割合	-0.446
高齢夫婦世帯割合	-0.465
高齢夫婦のみの世帯の割合	-0.465
65 歳以上人口割合 (男)	-0.489
第 1 次産業就業者比率	-0.49
65 歳以上人口割合 (女)	-0.504
65 歳以上人口割合	-0.505
65 歳以上人口割合	-0.505
年齢中位数	-0.549

表 28: 完視聴 UD 率の平均と正の  
相関がある項目

統計項目	相関係数
18～39 歳人口割合	0.499
15～64 歳人口割合	0.497
18～39 歳人口割合 (男)	0.495
18～39 歳人口割合 (女)	0.489
0～17 歳人口割合 (女)	0.483
15 歳未満人口割合	0.474
母子世帯割合	0.465
人口増減率	0.464
0～17 歳人口割合	0.441

表 29: 完視聴 UD 率の平均と負の相関がある  
項目

統計項目	相関係数
高齢単身世帯割合 (女)	-0.408
高齢夫婦世帯割合	-0.445
高齢夫婦のみの世帯の割合	-0.445
65 歳以上の世帯員のいる核家族世帯割合	-0.462
第 1 次産業就業者比率	-0.463
65 歳以上人口割合 (女)	-0.531
65 歳以上人口割合	-0.545
65 歳以上人口割合 (男)	-0.547
年齢中位数	-0.588

視聴 UD 率だけでなく、UD あたり ch 切替回数と統計項目の相関についても確認する。表 30 は、UD あたり ch 切替回数との平均と正の相関 ( $0.4 \leq r < 0.7$ ) がある項目を示している。年齢中位数や高齢層に関する項目、第 1 次産業就業者比率が大きい地域では、ch 切替回数も増加傾向にあることが確認された。表 31 は、UD あたり ch 切替回数との平均と負の相関 ( $-0.7 < r \leq -0.4$ ) がある項目を示している。正の相関とは対照的に、若い世代の人口比率や、母子世帯割合、人口増減率が大きい地域では、ch 切替回数が減少傾向にあること表 38 は、UD あたり ch 切替回数との標準偏差と正の相関 ( $0.7 \leq r$ ) がある項目を示している。標準偏差と正の相関を持つ項目として、65 歳以上の人口割合や第 1 次産業事業所割合、年齢中位数など、年齢が大きくなると値が大きくなる項目が確認された。表 39 は、UD あたり ch 切替回数との標準偏差と負の相関 ( $r \leq -0.7$ ) がある項目を示している。人口増減率と高齢者以外の層に関連する項目が、負の相関を示していることが確認できた。

表 30: ch 切替回数の平均と正の相関がある項目

統計項目	相関係数
年齢中位数	0.625
65 歳以上人口割合 (男)	0.590
65 歳以上人口割合	0.586
65 歳以上人口割合	0.586
65 歳以上人口割合 (女)	0.569
65 歳以上の世帯員のいる核家族世帯割合	0.500
第 1 次産業就業者比率	0.483
高齢夫婦のみの世帯の割合	0.481
高齢夫婦世帯割合	0.481
高齢単身世帯割合 (女)	0.426

表 31: ch 切替回数の平均と負の相関がある項目

統計項目	相関係数
0~17 歳人口割合 (男)	-0.403
未婚者割合 (15 歳以上人口)	-0.427
0~17 歳人口割合	-0.454
母子世帯割合	-0.468
15 歳未満人口割合	-0.489
0~17 歳人口割合 (女)	-0.491
人口増減率	-0.505
18~39 歳人口割合 (女)	-0.531
18~39 歳人口割合 (男)	-0.532
15~64 歳人口割合	-0.539
18~39 歳人口割合	-0.539

表 32: 1/3 以上視聴 UD 率の標準偏差と正の相関がある項目

統計項目	相関係数
65 歳以上人口割合 (女)	0.848
65 歳以上人口割合	0.831
65 歳以上人口割合	0.831
年齢中位数	0.812
65 歳以上人口割合 (男)	0.794
最終学歴人口割合 (小・中)	0.793
高齢単身世帯割合 (女)	0.702

表 33: 1/3 以上視聴 UD 率の標準偏差と負の相関がある項目

統計項目	相関係数
18~39 歳人口割合 (女)	-0.709
人口増減率	-0.746
15~64 歳人口割合	-0.772

表 34: 2/3 以上視聴 UD 率の標準偏差  
と正の相関がある項目

統計項目	相関係数
65 歳以上人口割合 (女)	0.847
65 歳以上人口割合	0.829
65 歳以上人口割合	0.829
年齢中位数	0.809
最終学歴人口割合 (小・中)	0.792
65 歳以上人口割合 (男)	0.791
第 1 次産業事業所割合	0.706
高齢単身世帯割合 (女)	0.704

表 35: 2/3 以上視聴 UD 率の標準偏差  
と負の相関がある項目

統計項目	相関係数
18~39 歳人口割合 (女)	-0.707
人口増減率	-0.745
15~64 歳人口割合	-0.770

表 36: 完視聴 UD 率の標準偏差と正  
の相関がある項目

統計項目	相関係数
65 歳以上人口割合 (女)	0.847
65 歳以上人口割合	0.828
65 歳以上人口割合	0.828
年齢中位数	0.808
最終学歴人口割合 (小・中)	0.792
65 歳以上人口割合 (男)	0.79
第 1 次産業事業所割合	0.709
高齢単身世帯割合 (女)	0.706

表 37: 完視聴 UD 率の標準偏差と負  
の相関がある項目

統計項目	相関係数
18~39 歳人口割合 (女)	-0.706
人口増減率	-0.744
15~64 歳人口割合	-0.770

表 38: ch 切替回数の標準偏差と正の相関がある項目

統計項目	相関係数
65 歳以上人口割合（女）	0.847
65 歳以上人口割合	0.829
65 歳以上人口割合	0.829
年齢中位数	0.809
最終学歴人口割合（小・中）	0.792
65 歳以上人口割合（男）	0.791
第 1 次産業事業所割合	0.706
高齢単身世帯割合（女）	0.704

表 39: ch 切替回数の標準偏差と負の相関がある項目

統計項目	相関係数
18～39 歳人口割合（女）	-0.707
人口増減率	-0.745
15～64 歳人口割合	-0.77

■ 考察 分析の結果を踏まえると、地域性を示す項目は年齢や世帯状況に関連するものに相関が見られていることが分かる。特に、高齢者層に関連する項目に相関が顕著に現れており、対照的な傾向が、若年層で見られている。高齢者層（特に 65 歳以上に関連）の割合が大きい地域では、視聴 UD 率は低くなり、率自体のばらつきは大きくなる。また、高齢者層の割合が大きい地域では、ch 切替回数も大きくなり、回数自体のばらつきも大きくなる。視聴 UD 率が低くなる傾向と、ch 切替回数が増加する傾向が同時に成り立つ関係性は自然であると考えられるため、この結果は妥当であると考えられる。一方、若年層や母子家庭、生産年齢人口の割合が大きい地域では、視聴 UD 率は高くなり、率自体のばらつきは小さくなる傾向にある。また、若年層や母子家庭、生産年齢人口の割合が大きい地域では、ch 切替回数が減少し、回数自体のばらつきも小さくなる傾向にある。高齢者層の視聴傾向と同様に、視聴 UD 率と ch 切替回数がトレードオフとなっていることは自然であると考えられるため、この結果は妥当であると考えられる。更に、高齢者層と、若年齢層に関連する統計項目で正反対の視聴傾向が現れている点も、結果の妥当性を根拠づけていると考察できる。以上より、高齢者の視聴傾向として、CM の期間に ch を切り替えることが多く、UD 視聴率が低下することがある。これは、高齢者が普段からテレビを余暇の時間にのんびりと視聴している直感的印象とは逆で、例えばメインで視聴している番組が CM に入った際、他局でも番組本編を見続けたい意識が高いのではないかと推察できる。そして、若年層や母

子世帯などが多い地域での視聴傾向としては、CMの期間にchを切り替えることが少なく、UD視聴率は高くなる。若年層ではCM放送時間も視聴しているユーザが多いため、この年齢層に合った広告効果が一定見込めると推察できる。母子家庭などでは、家事や育児の際に子どもにテレビを視聴させるケースも多いと考えられ、それに応じてch切替回数も減るのではないかと考えられる。また、若年層に関しては、テレビの機能として備わっている番組表などを上手に活用し、ザッピング回数等も減る傾向にあるのではないかとと思われる。このように、今後はそのような他の視聴行動と掛け合わせることで、さらなる視聴傾向の分析につなげていくことが重要である。

#### 3.4.4 まとめ

本分析では、視聴者のCM視聴を確認するため、4つの民間放送局で収集された非特定テレビ視聴履歴データを用いて、「CM視聴UD率」「ch切替回数」なる指標を提案し、分析した。これらの指標をもとに、地域ごとの視聴傾向を捉えるべく、市町村別の統計データを用いて、「CM視聴UD率」「ch切替回数」と関連のある統計データ項目を算出し、地域性による視聴傾向の違いを分析した。分析の結果、高齢者層が多い地域と、若年層が多い地域で視聴傾向が異なることが確認できた。高齢者層が多い地域では、CMの視聴はされにくく、視聴のばらつきは大きくなる。伴って、ch切替回数も大きくなり、回数自体のばらつきも大きくなる。一方、若年齢層ではCM視聴されやすい傾向にあり、chの切替回数も少ないことが示された。テレビ広告業界では、50歳以上の広告価値は低く、49歳以下の広告価値は高い（非公開の単価設定データ）。本分析結果では、テレビ離れが進んでいると言われている若年層の方がCM完視聴率が高く、ch切替回数が少ないことが示された。本分析結果は、テレビの価値をスポンサーへの説明するデータとして使うことも出来る。しかし、それ以上にスポンサーが特に注力したい市町村に関する分析データを示すことで、スポンサーにとってテレビのCMが魅力的な広告手段であることを説明し、CM出稿の獲得に利用していきたい。

今後は、例えば番組ジャンルにおける視聴傾向の違いや、放送時間による視聴傾向の違い等を確認することで、CMの視聴傾向をより詳細に明らかにし、更な

るテレビ価値拡大に繋げたい。

### 3.5 分析3：番組コーナー視聴傾向の分析

テレビ番組は、民間放送局にとってビジネスの根幹を支えるコンテンツであるが、現状のコンテンツ制作において、視聴率以外に視聴状況を明確に示す指標は存在しておらず、指標の確立を含め、番組視聴の分析・可視化が急務である。

本節では、前節で提案した「視聴UD（ユニークデバイス）率」を番組の各コーナーへ拡張する。「視聴UD率」は、少しでも番組の各コーナーを視聴したテレビ台数のうち、一定割合以上視聴したテレビ台数の割合と定義し、視聴UD率を地域ごとに算出する。「完視聴UD率」は、該当する番組コーナーにおけるUDのうち、完全に視聴したUDの割合を示し、「1/3以上視聴UD率」、「2/3以上視聴UD率」は、該当する番組コーナーを1/3又は2/3以上視聴したUDの割合を示すものと定義する。これに総務省統計局が発表している各種統計データを紐付けることで、番組コーナー視聴に対して地域傾向が存在するかを明らかにする。

#### 3.5.1 データセット

本項では、本分析で使用するデータセットについて述べる。

データセットは、(1)非特定テレビ視聴履歴データ、(2)番組メタデータ、(3)地域統計データ、(4)出演タレント統計データの4つのデータを分析対象とする。以下では各データセットについての詳細について述べる。

■ テレビの非特定テレビ視聴履歴データ 本節では、近畿一円の約350万台を対象として、在阪の4つの放送局（読売テレビ・毎日放送・朝日放送テレビ・関西テレビ）がそれぞれ独自に収集したオプアウト型非特定テレビ視聴履歴データを統合した視聴データを使用する。4つの放送局では取得形式や、振り分けられているテレビIDが異なることから、2.4節で実践された同一のテレビIDとして統合した非特定テレビ視聴履歴データを扱う。

■ **番組メタデータ** 非特定テレビ視聴履歴データは、テレビ受像機ごとの視聴開始と終了時刻を扱う履歴データであるため、番組に関する情報は含まれていない。そのため、読売テレビが保有するコーナーやCMが放送された時間帯などの番組の詳細がまとめられた「番組メタデータ」を用いる。番組メタデータは、各番組ごとの、番組タイトル、番組開始時刻、番組終了時刻、番組時間、番組ジャンル、コーナー開始時刻、コーナー終了時刻、コーナー時間、コーナージャンルなどがまとめられたデータである。分析には、2021年10月から12月の3ヶ月間のデータを用いる。

■ **統計データ** 本分析では、総務省統計局が公開している「統計でみる都道府県・市区町村のすがた（社会・人口統計体系）」のうち、本研究で対象とする近畿二府四県の全ての市区町村における各種統計データを使用する。これらの統計データは、政府統計総合窓口サイト（e-Stat）<sup>7</sup>のAPI機能を使用して収集を行った。統計データには、割合などの相対値で表されているものと、絶対値で表されているものがある。絶対値で表されている項目に対しては、人口等の影響による偏りを無くすため、標準化した数値を扱う。

■ **出演タレント統計データ** 2021年12月の番組の出演者に対して、Yahooクラウドソーシングを用いて性別や出身地、職業、SNSのフォロワーなどのデータを収集した。収集したデータを番組コーナーごとに、各属性の全出演者中の割合や年齢やフォロワーといった数値の最大値、最小値、平均値などを算出し、出演タレント統計データとした。

### 3.5.2 データの処理方法

本項では、本分析で使用するデータセットの処理方法について述べる。本研究では、番組視聴における傾向を分析するため、コーナー完視聴UD率なる指標を提案する。これらの指標を求めるために行った非特定テレビ視聴履歴データと番

---

<sup>7</sup><https://www.e-stat.go.jp/>

組メタデータの処理方法，および，地域統計データ，タレント統計データと紐付けるために行った処理方法について述べる．

■ コーナー完視聴 UD 率の算出 コーナー完視聴率についての定義を示す．コーナー開始時刻，コーナー終了時刻，視聴開始時刻，視聴終了時刻をそれぞれ  $T_s, T_e, t_s, t_e$  とし，非特定テレビ視聴履歴データにおいて，あるコーナーにおける視聴レコードが，以下の3条件のいずれかを満たすデバイス（以下，「UD」とする）を対象とする．

$$T_s < t_s \quad \text{かつ} \quad t_e < T_e \quad (10)$$

$$T_s < t_e \quad \text{かつ} \quad t_e < T_e \quad (11)$$

$$t_s \leq T_s \quad \text{かつ} \quad T_e \leq t_e \quad (12)$$

UD に対し，完全にコーナーを視聴した UD の割合を示すものを「コーナー完視聴 UD 率」，1/3 以上視聴した UD の割合を示すものを「コーナー 1/3 以上視聴 UD 率」，2/3 以上視聴した UD の割合を示すものを「コーナー 1/3 以上視聴 UD 率」と定義する．コーナー完視聴率を計算するためには，そのコーナーにおける UD 数と，各 UD が視聴していた時間を抽出する必要がある．前述のように，非特定テレビ視聴履歴データには，コーナーに関連する情報は含まれていない．そのため，まず番組メタデータから (1) コーナー情報を抽出し，そして，各コーナーに対して非特定テレビ視聴履歴データから (2) UD 数 (3) 各 UD 視聴時間 (4) 完 (1/3, 2/3 以上) 視聴 UD 数を算出し，コーナー完視聴 UD 率，コーナー 1/3 以上視聴 UD 率，コーナー 2/3 以上視聴 UD 率を導き出す．以下で，それぞれの抽出方法について述べる．

コーナー情報 番組メタデータのうち，コーナージャンルが CM でないレコードのコーナー開始時刻，コーナー終了時刻をコーナー情報として抽出する．

UD 数 コーナー情報から抽出できるコーナー開始時刻及びコーナー終了時刻に対して，非特定テレビ視聴履歴データの視聴開始時刻，

視聴終了時刻が，式(2),(3),(4)のいずれかを満たすテレビを抽出する．また，郵便番号を抽出することで，地域別のUD数を算出する．いずれも2.3節で提案されたアルゴリズムによって定められた一意のIDを用いてUD数を計算した．

**各UDにおけるコーナー視聴時間** 本稿では，コーナー開始時刻からコーナー終了時刻までにどの程度視聴したかを視聴の尺度としている．即ち，あるコーナーに対してUDであるとされたテレビのうち，コーナー開始時刻からコーナー終了時刻までの時間と重なる視聴レコードを抽出し，コーナー放送時間と重なる部分の総和を求める．

**完(1/3,2/3以上)視聴UD数** あるコーナーに対して求めたUDのうち，コーナー視聴時間がコーナー放送時間と等しいものを完視聴UDとする．同様に，コーナー視聴時間がコーナー放送時間の1/3(2/3)以上である場合，1/3(2/3)以上視聴UDとする．また，郵便番号を抽出することで，地域別の完(1/3,2/3以上)視聴UD数を算出する．

これらによって，コーナー完視聴UD率，コーナー1/3以上視聴UD率，コーナー2/3以上視聴UD率は以下の式で計算される．

$$\text{コーナー完視聴UD率} = \frac{\text{完視聴UD数}}{\text{UD数}} \quad (13)$$

$$\text{コーナー1/3以上視聴UD率} = \frac{\text{1/3以上視聴UD数}}{\text{UD数}} \quad (14)$$

$$\text{コーナー2/3以上視聴UD率} = \frac{\text{2/3以上視聴UD数}}{\text{UD数}} \quad (15)$$

■ **コーナー完視聴率の算出** 図29に示すように，コーナー完視聴UD率はコーナーそのものの放送時間によって変化する．コーナー時間が長くなればなるほど

コーナー完視聴 UD 率の平均は低下する。このことからコーナー時間によってコーナー完視聴 UD 率を調整する必要がある。

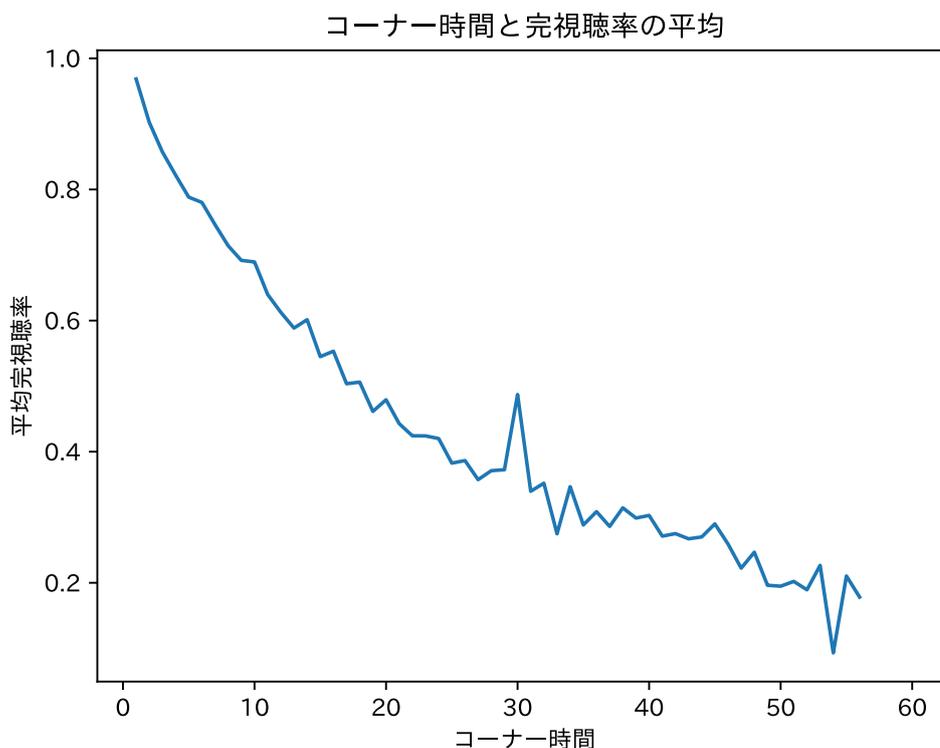


図 29: コーナー時間と完視聴率の平均の変化

コーナー時間  $x$  に対するコーナー完視聴 UD 率の平均  $f(x)$  は以下のように近似できる。

$$f(x) = 0.84459249 \times e^{-0.04342554x} + 0.12258209 \quad (16)$$

よってコーナー完視聴率は以下の式に表すように、コーナー完視聴 UD 率から、そのコーナー時間の平均コーナー完視聴 UD 率の差をとった値とする。

$$\text{コーナー完視聴率} = \text{コーナー完視聴 UD 率} - f(x) \quad (17)$$

■ 統計データと非特定テレビ視聴履歴データの紐付け 統計局の統計データは、JIS コードと呼ばれる市町村を一意に識別するコードごとにデータが集約されて

いる。一方で、非特定テレビ視聴履歴データでは、地域情報として郵便番号7桁が収集されており、地域ごとにデータを処理する場合には、郵便番号をキーとする必要がある。そのため、これら2種類のデータを集約するためには、市町村名と郵便番号による紐付けが必要となる。紐付けについては、郵便番号と市町村の対応を以下の順序で処理することで実現した。

1. 非特定テレビ視聴履歴データ上の郵便番号と、統計データ上の市町村に紐づく郵便番号が1対1に対応する場合、非特定テレビ視聴履歴データ、統計データのいずれもそのまま利用する。
2. 特定視聴履歴データ上の郵便番号に、統計データ上の市町村に紐づく郵便番号が複数対応する場合、該当する市町村の統計データを合算することで対応する。
3. 統計データ上では、市町村に対応する郵便番号が1つしかないため、非特定テレビ視聴履歴データ上の郵便番号に、統計データ上の市町村に紐づく郵便番号が存在しない場合が生じる。この場合、非特定テレビ視聴履歴データを近隣の郵便番号と紐付けて合算し、2種類のデータに漏れがないように調整した。

### 3.5.3 分析

**分析方法1** :コーナー完視聴率と地域性の関係性 コーナー視聴率と地域傾向の関係性を分析する方法について述べる。まず、番組メタデータより、2021年9月から12月の間に4局で放送された番組コーナーの開始時刻と終了時刻を抽出する。次に、非特定テレビ視聴履歴データをJISコードごとに分け、各コーナーにおける完視聴率、1/3以上視聴率、2/3以上視聴率を算出する。これらの指標を、JISコードごとの平均を算出する。そして、JISコードごとに統計データの各項目と算出した指標との相関を確認する。

**分析結果1** まず、完視聴率、1/3以上視聴率、2/3以上視聴率と257項目における統計データとの相関を確認した。地域性による違いを確認するため、視聴率と

より相関が見られた項目について確認する。表 40, 42, 44 は, 1/3 以上視聴率, 2/3 以上視聴率, 完視聴率との平均と正の相関 ( $0.5 \leq r$ ) がある項目を示している。これら 3 つの指標に共通して登場する項目として, 若年者層の割合や人口増減率が大きい地域では, これらの視聴率も大きいことが示された。表 41, 43, 45 は, 高齢者の割合に関わる項目がほとんどで, この項目が大きいほど視聴率が小さくなることが示された。なお, 「最終学歴人口割合 (小学校・中学校)」に関しては, 80 代以上が全体の 9 割を占めており, 戦前の教育制度の違いや, 戦後の混乱などが背景にあると考えられるため, 高齢者に関連する項目として捉えることができると考えられる。

表 40: 1/3 以上視聴率の平均と正の相関がある項目

統計項目	相関係数
15~64 歳人口割合	0.655
18~39 歳人口割合 (女)	0.628
18~39 歳人口割合	0.620
人口増減率	0.617
18~39 歳人口割合 (男)	0.594
0~17 歳人口割合 (女)	0.570
15 歳未満人口割合	0.552
母子世帯割合	0.544
未婚者割合 (15 歳以上人口)	0.529
0~17 歳人口割合	0.518

表 41: 1/3 以上視聴率の平均と負の相関がある項目

統計項目	相関係数
年齢中位数	-0.716
65 歳以上人口割合 (女)	-0.707
65 歳以上人口割合	-0.695
65 歳以上人口割合 (男)	-0.663
最終学歴人口割合 (小学校・中学校)	-0.642
第 1 次産業就業者比率	-0.577
高齢単身世帯割合 (65 歳以上の者 1 人) (女)	-0.568
高齢夫婦のみの世帯の割合	-0.547
高齢夫婦世帯割合 (高齢夫婦のみ)	-0.547
65 歳以上の世帯員のいる核家族世帯割合	-0.512

表 42: 2/3 以上視聴率の平均と正の相関がある項目

統計項目	相関係数
15～64 歳人口割合	0.688
18～39 歳人口割合（女）	0.654
人口増減率	0.648
18～39 歳人口割合	0.643
18～39 歳人口割合（男）	0.613
0～17 歳人口割合（女）	0.591
15 歳未満人口割合	0.575
未婚者割合（15 歳以上人口）	0.559
母子世帯割合	0.548
0～17 歳人口割合	0.547

表 43: 2/3 以上視聴率の平均と負の相関がある項目

統計項目	相関係数
65 歳以上人口割合（女）	-0.743
年齢中位数	-0.738
65 歳以上人口割合	-0.729
65 歳以上人口割合（男）	-0.693
最終学歴人口割合（小学校・中学校）	-0.689
高齢単身世帯割合（65 歳以上の者 1 人）（女）	-0.608
第 1 次産業就業者比率	-0.575
高齢夫婦のみの世帯の割合	-0.568
高齢夫婦世帯割合（高齢夫婦のみ）	-0.568
高齢単身世帯割合（65 歳以上の者 1 人）	-0.530
第 1 次産業従業者割合	-0.523
65 歳以上の世帯員のいる核家族世帯割合	-0.521
第 1 次産業事業所割合	-0.520

表 44: 完視聴率の平均と正の相関がある項目

統計項目	相関係数
15～64 歳人口割合	0.691
18～39 歳人口割合（女）	0.654
18～39 歳人口割合	0.645
人口増減率	0.642
18～39 歳人口割合（男）	0.617
0～17 歳人口割合（女）	0.589
15 歳未満人口割合	0.571
未婚者割合（15 歳以上人口）	0.565
母子世帯割合	0.553
0～17 歳人口割合	0.547

表 45: 完視聴率の平均と負の相関がある項目

統計項目	相関係数
65 歳以上人口割合（女）	-0.740
年齢中位数	-0.736
65 歳以上人口割合	-0.728
65 歳以上人口割合（男）	-0.696
最終学歴人口割合（小学校・中学校）	-0.679
高齢単身世帯割合（65 歳以上の者 1 人）（女）	-0.606
高齢夫婦のみの世帯の割合	-0.570
高齢夫婦世帯割合（高齢夫婦のみ）	-0.570
第 1 次産業就業者比率	-0.558
高齢単身世帯割合（65 歳以上の者 1 人）	-0.529
65 歳以上の世帯員のいる核家族世帯割合	-0.525
第 1 次産業従業者割合	-0.505

**考察1** 分析の結果を踏まえると、地域性を示す項目は年齢や世帯状況に関連するものに相関が見られていることが分かる。特に、高齢者層に関連する項目に相関が顕著に現れており、対照的な傾向が、若年者層で見られている。高齢者層(特に65歳以上に関連)の割合が大きい地域では、完視聴率は低くなっている。これは高齢者が番組のコーナー中にチャンネルを切り替える頻度が高いことを示しており、高齢者が普段からテレビを余暇の時間にのんびりと視聴している直感的印象とは逆の結果となっている。これは同じコーナーの中でも興味がなくなったタイミングで他に面白い番組を探すために他局にチャンネルを切り替えていることが推測できる。

また、若年者層ではコーナーの完視聴率が高い傾向がある。これは、若者がスマートフォンを操作しながらテレビを見ることが増えたことでチャンネル切り替えを行う頻度が少なくなったことや、子供がいる世帯で子供がテレビを占領してチャンネルを家族に切り替えさせないようにしていることが背景として考えられる。また、若年者層に関しては、テレビの機能として備わっている番組表などを上手に活用し、ザッピング回数等も減る傾向にあるのではないと思われる。このように、今後はそのような他の視聴行動と掛け合わせることで、さらなる視聴傾向の分析につなげていくことが重要である。

**分析方法2**:コーナー完視聴率とタレント統計項目の関係性 コーナー視聴率と出演タレント統計データの関係性を分析する方法について述べる。まず、番組メタデータより、2021年12月に4局で放送された番組コーナーの開始時刻と終了時刻を抽出する。次に、非特定テレビ視聴履歴データをJISコードごとに分け、各コーナーにおける完視聴率、1/3以上視聴率、2/3以上視聴率を算出する。番組のジャンルごとの傾向を調べるために、各コーナーごとに、タレント統計情報との相関係数を計算した。また、ジャンルの一覧とその出現数を表46に示す。サンプルの相関係数が0.2だった場合に、母相関係数がゼロではないと統計的に証明するためには194サンプル必要であることから、「情報/ワイドショー」、「バラエティー」、「ニュース/報道」、「ドラマ」、「音楽」、「スポーツ」についてのみ分析を行う。

表 46: コーナーのジャンルリストとその出現数 (2021 年 12 月)

ジャンル	出現回数
情報/ワイドショー	4349
バラエティー	3572
ニュース/報道	1481
ドラマ	606
音楽	291
スポーツ	244
ドキュメンタリー	97
趣味/教育	68
アニメ/特撮	56
映画	26
演劇/公演	22

**分析結果 2 と考察** 各ジャンルごとのコーナー完視聴率とタレント統計項目との相関を確認した。出演者による完視聴率の違いを確認するため、視聴率とより相関が見られた項目について確認する。表 47, 49, 51, 53, 55, 57 は、それぞれのジャンルでのコーナー完視聴率の平均と正の相関がある上位 6 項目を示している。また、表 48, 50, 52, 54, 56, 58 は、それぞれのジャンルでのコーナー完視聴率の平均と負の相関がある上位 6 項目を示している。

表 47: ジャンル「情報/ワイドショー」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
グラビアアイドルの割合	0.150
最年少出演者の年齢	0.145
事務所 A 所属タレントの割合	0.116
YouTube チャンネルを持つ出演者の割合	0.110
事務所 B 所属タレントの割合	0.109
事務所 C 所属タレントの割合	0.109

表 48: ジャンル「情報/ワイドショー」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
アナウンサーの割合	-0.131
東北地方出身者の割合	-0.123
事務所 D 所属タレントの割合	-0.119
出演者数	-0.096
一般人の割合	-0.095
北海道出身者の割合	-0.083

まず、表 47, 48 で示したジャンル「情報/ワイドショー」については、アナウ

ンサーの割合が高いほど完視聴率が低いことが分かっている。一方でグラビアアイドルや YouTube チャンネルを持つ出演者といった堅くない属性の出演者が多いほど完視聴率が高くなっている。事務所 A と事務所 B はバラエティーで高い視聴率傾向を持つお笑い芸人が多数所属する芸能事務所であり、これもまた高い完視聴率を持つ傾向にある。このことから、堅くない出演者が多い番組の方が完視聴率が高くなりやすいことが推測できる。

表 49: ジャンル「バラエティー」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
事務所 A 所属タレントの割合	0.155
芸人の割合	0.139
Instagram のアカウントを持つ出演者の割合	0.107
事務所 E 所属タレントの割合	0.078
関西地方出身者の割合	0.061
事務所 B 所属タレントの割合	0.060

表 50: ジャンル「バラエティー」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
料理人の割合	-0.232
平均年齢	-0.214
最年長出演者の年齢	-0.174
最年少出演者の年齢	-0.130
俳優の割合	-0.072
事務所 F 所属タレントの割合	-0.072

次に、表 49, 50 で示したジャンル「バラエティー」については、お笑い芸人が多く所属している芸能事務所 A,E,B や芸人の割合が高い順位に表れていることから、芸人が多い割合で構成されている番組が高い完視聴率を示している。関西地方出身者の割合も高くなっているが、これ関西圏での視聴データを分析しているので純粋に関西出身の芸人が人気であるのか、同じ地方の出演者が多いことがいい結果に繋がっているのかは判断できない。また、平均年齢、最年長出演者の年齢、最年少出演者の年齢が高いほど完視聴率が低くなることが分かった。これは視聴者が全体的な年齢層が低い番組をバラエティー番組として求めていると考察できる。料理人や俳優といったお笑い番組とは離れたバラエティー番組では相対的に完視聴率が低くなることも分かる。

表 51: ジャンル「ニュース/報道」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
アナウンサーの割合	0.129
中国地方出身者の割合	0.107
事務所 A 所属タレントの割合	0.62
関東地方出身者の割合	0.057
芸人の割合	0.056
X のアカウントを持つ出演者の割合	0.053

表 52: ジャンル「ニュース/報道」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
出演者数	-0.198
元スポーツ選手の割合	-0.165
歌手の割合	-0.158
中部地方出身者の割合	-0.118
最年長出演者の年齢	-0.115
平均年齢	-0.114

表 51, 52 で示したジャンル「ニュース/報道」については、アナウンサーの割合が高いほど完視聴率は高くなっている。ニュース番組ではアナウンサーが一人はいることがほとんどであるので、出演者数が負の相関を持つことと合わせて出演者が少なくアナウンサーの割合が高い番組の方が完視聴率が高くなっていることが分かる。また、中国地方出身者の割合については、人気ニュース番組の一つのメイン MC が中国地方出身者であることから、その番組の完視聴率の高さに引上げられていると推測できる。

表 53: ジャンル「ドラマ」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
X の平均フォロワー数	0.192
バラエティタレントの割合	0.162
X の最大フォロワー数	0.156
東北地方出身者の割合	0.149
事務所 G 所属タレントの割合	0.124
事務所 H 所属タレントの割合	0.106

表 54: ジャンル「ドラマ」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
最年長出演者の年齢	-0.194
平均年齢	-0.173
YouTube の最大登録者数	-0.147
YouTube の平均登録者数	-0.131
事務所 A 所属タレントの割合	-0.124
インスタグラムの最大フォロワー数	-0.107

次に、表 53, 54 で示したジャンル「ドラマ」では、X のフォロワー数やバラエティタレントの割合が上位に来ている。このことから話題性のある出演者がいるドラマの方が完視聴率が高くなることが分かる。事務所 G は俳優・歌手が多く所

属する事務所で事務所 H はアイドルが多く所属する事務所となり、完視聴率が高くなる傾向があることが分かる。また、YouTube やインスタグラムのフォロワー数に関する項目で負の相関が見られることから、完視聴率が低い傾向がある。以上のことから、リアルタイムで感想を共有したりトレンド機能がある X で人気が出るドラマの方が完視聴率が高い可能性がある。事務所 A はお笑い芸人が多く所属する事務所となり、ドラマ出演者に占める割合が高いと完視聴率が低くなることが分かる。

表 55: ジャンル「音楽」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
作曲家の割合	0.131
最年少出演者の年齢	0.124
中国地方出身者の割合	0.061
事務所 H 所属タレントの割合	0.043
事務所 B 所属タレントの割合	0.040
事務所 I 所属タレントの割合	0.034

表 56: ジャンル「音楽」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
男性の割合	-0.349
関東地方出身者の割合	-0.311
インスタグラムのアカウントを持つ出演者の割合	-0.239
インスタグラムの最大フォロワー数	-0.228
俳優の割合	-0.227
YouTube の平均登録者数	-0.212

表 55, 56 で示したジャンル「音楽」では、正の相関と負の相関で相関係数の値に大きな差が存在している。男性の割合が高いほど完視聴率が低くなるということは、逆に女性の割合が高いほど完視聴率が高くなっていると言えるので、音楽番組に関しては、女性アーティストの方が視聴率を維持しやすいことが分かる。事務所 H はアイドルが多数所属しており、音楽番組の完視聴率と高い相関がある。また、SNS に関する項目が負の相関で見られる。これは SNS アカウントを持たない高齢者に人気のアーティストが高い完視聴率を記録していることが分かる。

表 57: ジャンル「スポーツ」の完視聴率と正の相関があるタレント統計項目

タレント統計項目	相関係数
事務所D所属タレントの割合	0.138
バラエティタレントの割合	0.083
元スポーツ選手の割合	0.082
北海道出身者の割合	0.081
事務所J所属タレントの割合	0.061
関東地方出身者の割合	0.34

表 58: ジャンル「スポーツ」の完視聴率と負の相関があるタレント統計項目

タレント統計項目	相関係数
最年長出演者の年齢	-0.232
Xのアカウントを持つ出演者の割合	-0.188
Xの最大フォロワー数	-0.166
中国地方出身者の割合	-0.165
平均年齢	-0.160
歌手の割合	-0.152

最後に表 57, 58 で示したジャンル「スポーツ」については、プレイヤーは出演者には入らず、キャスターやパネラーが出演者として認識される。パネラーの属性として、バラエティタレントや元スポーツ選手の評価が高く、歌手の評価が低いことが分かる。

### 3.5.4 まとめ

本分析では、視聴者の番組コーナー視聴を確認するため、4つの民間放送局で収集された非特定テレビ視聴履歴データから得られるコーナー時間を考慮した番組コーナー完視聴率を用いて分析した。これらの指標をもとに、地域ごとの視聴傾向を捉えるべく、市町村別の統計データを用いて、番組コーナー完視聴率と相関のある統計データ項目を算出し、地域性による視聴傾向の違いを分析した。分析の結果、高齢者層が多い地域ではコーナー完視聴率が低く、若年層が多い地域ではコーナー完視聴率が低いことが確認できた。また、番組ジャンルごとに完視聴率と出演タレントの統計項目の高さの相関から視聴傾向の分析も行った。その結果、番組ジャンルごとに完視聴率が高くなる属性やユーザーの好みを推測することができた。今後は、より細かい粒度でのジャンル分けや放送時間を限定して分析を行い視聴傾向を確認することで、番組コーナーの視聴傾向をより詳細に明らかにし、更なるテレビ価値拡大に繋げたい。

## 3.6 分析4：クラスタリング分析

放送業界の商習慣では、ビデオリサーチ株式会社が調査している視聴率で視聴者規模を計測することでテレビのCM価値証明を実施して、取引がされている。この視聴率には、性別・年齢の属性情報が付与されており、CMを出稿したいスポンサは自社製品・サービス等のターゲットである性別や年齢の方が多く視聴している時間帯へのCM出稿を希望し、さらにCM放送後にターゲットへのリーチ実績の報告を求めている。しかし、近年個人の趣味嗜好や興味関心は多様化しており、例えば「30代の男性」といっても同様の趣味趣向や興味関心を持っているとは限らない。そこで、インターネット広告業界では性別・年齢でのターゲティングだけでなく、個人個人の趣味嗜好や興味関心をもとに効果的にターゲティングを実施している。これはパソコンやスマートフォンの利用履歴から各個人の興味関心を推定することで実現している。

そこで本分析では、視聴者の性別・年齢などの属性情報が不明な非特定テレビ視聴履歴データのテレビ番組の視聴履歴を基に特徴量を作成し、K-meansを用いてテレビ視聴者を異なる視聴傾向に基づいてクラスタリングする手法を提案する。

本分析では、視聴習慣などが大きく変化するような大型イベントがない通常の放送期間として2023年2月20日～26日、WORLD BASEBALL CLASSIC 2023 (WBC) やサッカー日本代表戦、世界フィギュアスケート選手権2023等の大型のスポーツ番組が放送されていたイベント期間として2023年3月20日～26日の2つの期間に対して提案手法を適用し、通常とは異なる番組が放送されている期間にも提案手法が有効であるかを調査した。結果として提案手法を用いて2つの期間において様々な視聴傾向ごとに分類可能であることを確認した。

### 3.6.1 データセット

本項では、本分析で使用するデータセットについて述べる。

データセットは、(1) 非特定テレビ視聴履歴データ、(2) 番組メタデータの2つのデータを分析対象とする。以下では各データセットについての詳細について述べる。

■ **テレビの非特定テレビ視聴履歴データ** 本節では、近畿一円の約350万台を対象として、在阪の4つの放送局（読売テレビ・毎日放送・朝日放送テレビ・関西テレビ）がそれぞれ独自に収集したオプアウト型非特定テレビ視聴履歴データを統合した視聴データを使用する。4つの放送局では取得形式や、振り分けられているテレビIDが異なることから、2.4節で実践された同一のテレビIDとして統合した非特定テレビ視聴履歴データを扱う。分析には、2023年2月20日～26日と2023年3月20日～26日のデータを用いる。

■ **番組メタデータ** 非特定テレビ視聴履歴データは、テレビ受像機ごとの視聴開始と終了時刻を扱う履歴データであるため、番組に関する情報は含まれていない。そのため、読売テレビが保有するコーナーやCMが放送された時間帯などの番組の詳細がまとめられた「番組メタデータ」を用いる。番組メタデータは、各番組ごとの、番組タイトル、番組開始時刻、番組終了時刻、番組時間、番組ジャンル、コーナー開始時刻、コーナー終了時刻、コーナー時間、コーナージャンルなどがまとめられたデータである。分析には、2023年2月20日～26日と2023年3月20日～26日のデータを用いる。

### 3.6.2 提案手法

本分析では非特定テレビ視聴履歴データを用いて視聴傾向別に視聴者をクラスタリングする手法を提案する。本項ではまず、クラスタリングに使用する特徴量について述べ、次に視聴傾向の分析方法について説明する。また本分析におけるクラスタリング結果に対するラベル付や考察などは読売テレビの専門家の意見を参考にしている。

■ **特徴量抽出及びクラスタリング** このデータには各視聴者を一意に判別するID、視聴開始時刻と終了時刻、視聴していたチャンネル、視聴した番組カテゴリが含まれている。本分析では分析期間中に放送されたテレビ番組のカテゴリ別の累計視聴時間を視聴者ごとに算出し、これらのデータを特徴量として使用している。分析期間中に含まれる具体的な番組カテゴリの一覧を表59に示す。また、本

研究ではクラスタリング手法として K-means を使用し、先述した特徴量を用いて分類し、各クラスごとの傾向を考察する。またクラス数は放送休止時刻と放送回数が少ない番組カテゴリを除いた 10 クラスとした。

表 59: テレビ番組のカテゴリ一覧

アニメ/特撮	スポーツ
ドキュメンタリー	ドラマ
ニュース/報道	バラエティー
情報/ワイドショー	放送休止
映画	演劇/公演
趣味/教育	音楽

■ ユーザ別視聴傾向分析 先述した特徴量を用いて K-means によるクラスタリングを行い、各視聴者（視聴 ID）がどのクラスに属するかを決定する。その後、非特定テレビ視聴履歴データに含まれる視聴 ID を基に、非特定視聴データに対して本研究のクラスタリング結果を付与し、各クラスに属する視聴者がどの時間帯にどの番組カテゴリを視聴しているかを明らかにし、各カテゴリ別の視聴時間を積み上げ棒グラフとして本稿では取り扱う。

### 3.6.3 データの処理

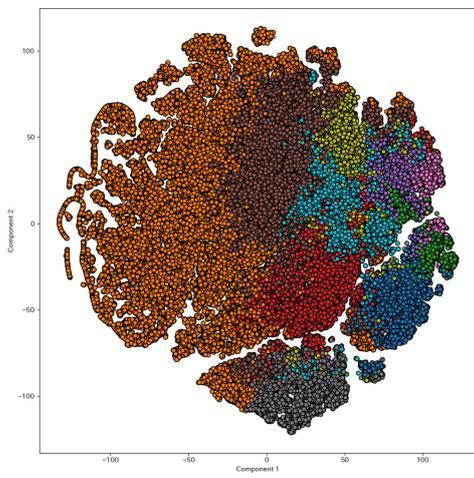
本研究では、各視聴者がどの時間帯にどの番組カテゴリを視聴していたかを調べるために、読売テレビ、毎日放送、朝日放送、関西テレビ放送における非特定テレビ視聴履歴データを用いて実験を行う。本データは、テレビ視聴データ連携に関する共同技術検証実験 [6] により、放送局間で連携されたデータであり、各視聴者ごとの視聴開始/終了時刻、視聴チャンネル、視聴時間が記録されている。本データに対して放送スケジュールを基に各視聴者がどの番組カテゴリを視聴していたかを紐づける実験を行う。

本実験では、以下の 2 つの期間のデータを使用する。1 つ目は視聴習慣などが大きく変化するような大型イベントがない通常の放送期間である 2023 年 2 月 20 日～2023 年 2 月 26 日のデータを、2 つ目は大型イベント (WBC) やサッカー日本

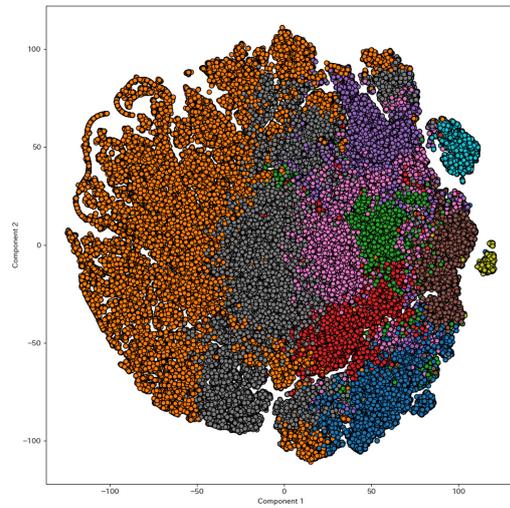
代表選，世界フィギュアスケート選手権 2023 等のスポーツカテゴリに属する番組が放送されていた期間である 2023 年 3 月 20 日～2023 年 3 月 26 日の非特定テレビ視聴履歴データを用いて分析を行い，これら 2 つの期間でどの程度視聴傾向に影響があるかを調査する．また，全件調査を実施するには非常に多くの計算リソースを必要とするため，期間内のデータから全体の 10 % の視聴者を抽出し実験を行っている．各期間におけるデータ数は，通常期間 5,910,986 件，140,748 台，イベント期間 5,972,231 件，141,264 台である．

#### 3.6.4 分析と評価

各期間別にクラスタリングを実施し，各期間別に t-SNE を用いて次元削減を行い，K-means によるクラスタリング結果を可視化したところ図 30 のようになった．通常期間とイベント期間で一部データが混ざっている部分はあるものの，ある程度の傾向別にクラスタリングできていることが確認できた．さらに具体的にクラスタごとの番組カテゴリ別の視聴時間や視聴している時間帯といった視聴傾向を分析したところ，両期間で特定の番組カテゴリを好む層や良く視聴する時間帯別に分類できていることが確認できた．本分析では，専門家の助言に基づき，クラスタごとの視聴時間を基に 1 日のうちに複数時間テレビを視聴していると予想されるヘビー層，不定期にテレビを視聴していると予想されるライト層，これら 2 つの中間に位置するミドル層に分類した．これらに加えて，視聴する番組カテゴリ，視聴している時間帯の 3 つの観点からデータに対してラベル付を行い，そのクラスに属する視聴テレビの傾向，属性について考察する．



(a) 通常期間 (2023 年 2 月 20 日～  
26 日)



(b) イベント期間 (2023 年 3 月 20 日～  
26 日)

図 30: t-SNE によるクラスタリング結果の可視化

■ 通常期間における分析結果 通常期間においてクラスタごとの視聴傾向を可視化した結果の一部を図 32～36 に示す。ヒストグラムの中の各色は図 31 に示す番組カテゴリと対応している。グラフについては縦軸が累計の視聴時間、横軸が 0 時から 24 時までの各時間帯を表し、各目盛りは 15 分ごとに区切られている。

番組カテゴリ	
<span style="color: #FF6666;">■</span>	アニメ/特撮 D
<span style="color: #008000;">■</span>	スポーツ E
<span style="color: #FF8C00;">■</span>	ドキュメンタリー F
<span style="color: #FFA500;">■</span>	ドラマ D
<span style="color: #800080;">■</span>	ニュース/報道 A
<span style="color: #ADD8E6;">■</span>	バラエティー C
<span style="color: #0000FF;">■</span>	情報/ワイドショー B
<span style="color: #C0C0C0;">■</span>	放送休止 Z
<span style="color: #8B4513;">■</span>	映画 D
<span style="color: #A0522D;">■</span>	演劇/公演 F
<span style="color: #FF0000;">■</span>	趣味/教育 F
<span style="color: #32CD32;">■</span>	音楽 C

図 31: 番組カテゴリの対応色

### 1. ヘビー層（図 32）の特徴

ボリュームゾーンは朝と夜，視聴ジャンルに偏りが少ない．このクラスタでは8時付近で一度視聴時間が減少していることから，8時に仕事等で外出する層と，昼間も在宅しテレビを視聴している層の2つが含まれていると予想される．

### 2. ミドル層（図 33）の特徴

ボリュームゾーンは朝，視聴ジャンルとしてはバラエティの割合が少ない．8時から8時15分の区間で一度視聴者が減っているのは同時間帯に放送されているNHKの朝の連続テレビ小説を視聴するためにチャンネルを切り替えていると予想される．同様に19時にも視聴者が減っており，こちらはNHKニュース視聴のための移動と考えられ，この層はNHKを好んで視聴する層だと予想される．

### 3. ミドル層（図 34）の特徴

比較的日中も見えており，特定のボリュームゾーンが存在しない．視聴ジャンルとしてはバラエティの割合が少ない．図33と似た傾向が見られ，朝と夜にはNHKを視聴するためにチャンネルを切り替えていると考えられる．本クラスタは図33と類似する視聴傾向を持っているが，日中は学校や仕事

等で外出せず、比較的日中も在宅している層だと予想される。

#### 4. ライト層（図 35）の特徴

ボリュームゾーンは早朝と深夜、視聴ジャンルとしては音楽やアニメの割合が多い。このクラスでは視聴傾向が早朝や深夜にアニメを視聴しており、またバラエティの割合も他クラスと比較して低いことから、深夜アニメが好きな比較的年轻層だと予想される。

#### 5. ライト層（図 36）の特徴

ボリュームゾーンは朝に偏り、視聴ジャンルとしてはアニメ、音楽、バラエティが少ない。比較朝早くからテレビを見ているものの、他クラスと比較して19時～22時のゴールデンタイムといわれる視聴者が多い時間帯の視聴時間が短く、音楽やアニメの視聴時間が少ないことから高齢者を多く含む層だと予想される。

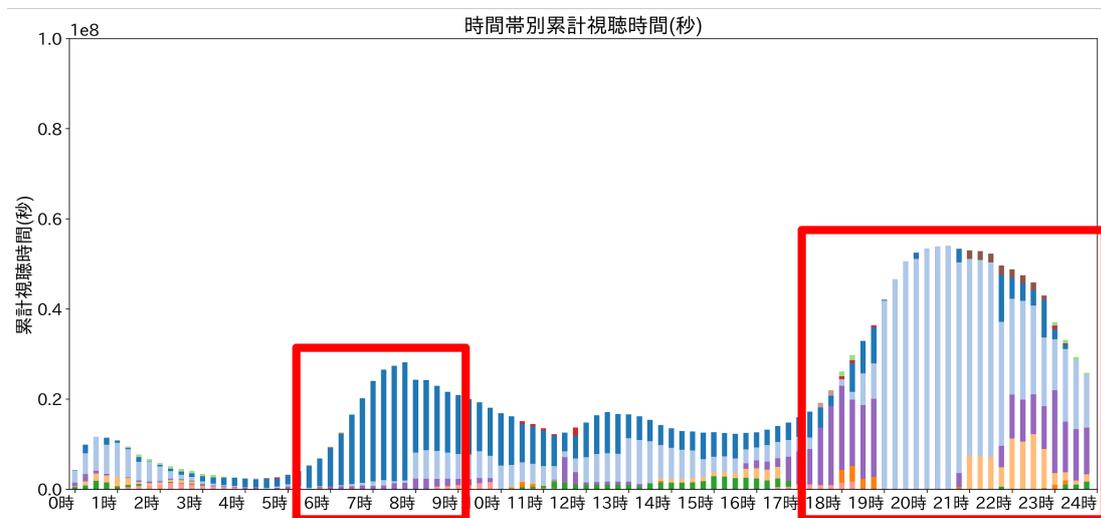


図 32: ヘビー層/朝・夜/偏り少

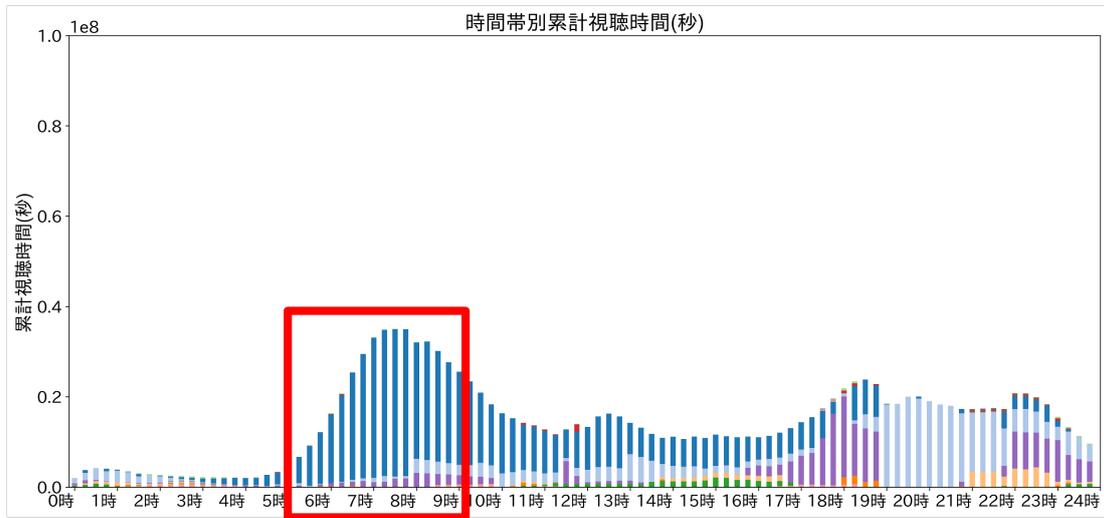


図 33: ミドル層/朝/バラエティ少

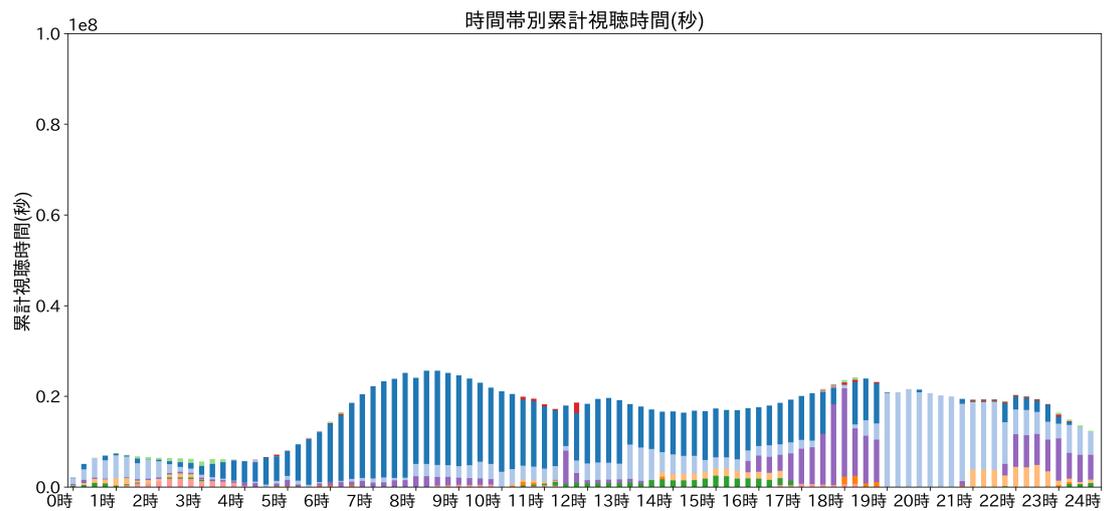


図 34: ミドル層/偏り少/バラエティ少

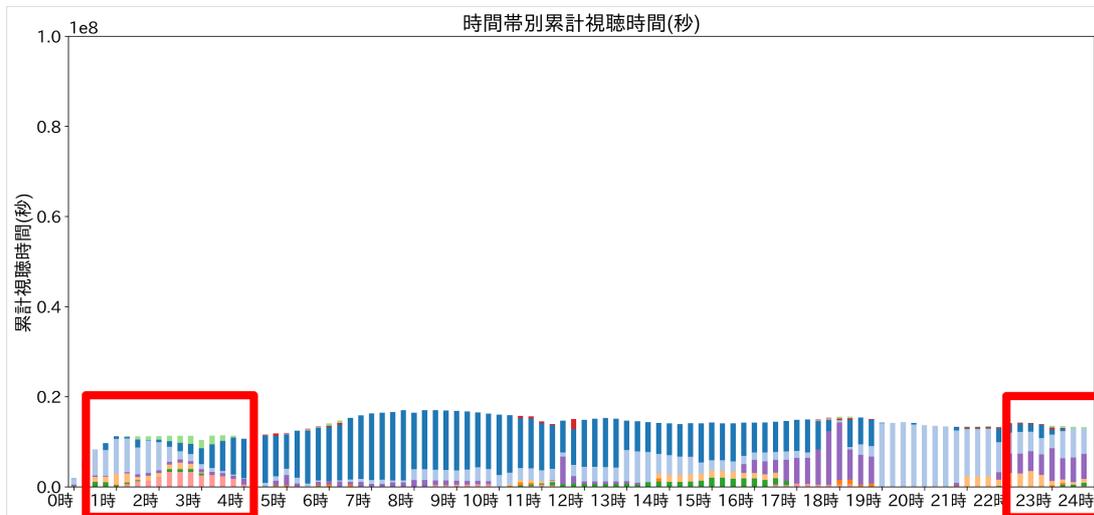


図 35: ライト層/深夜・早朝/アニメ・音楽多

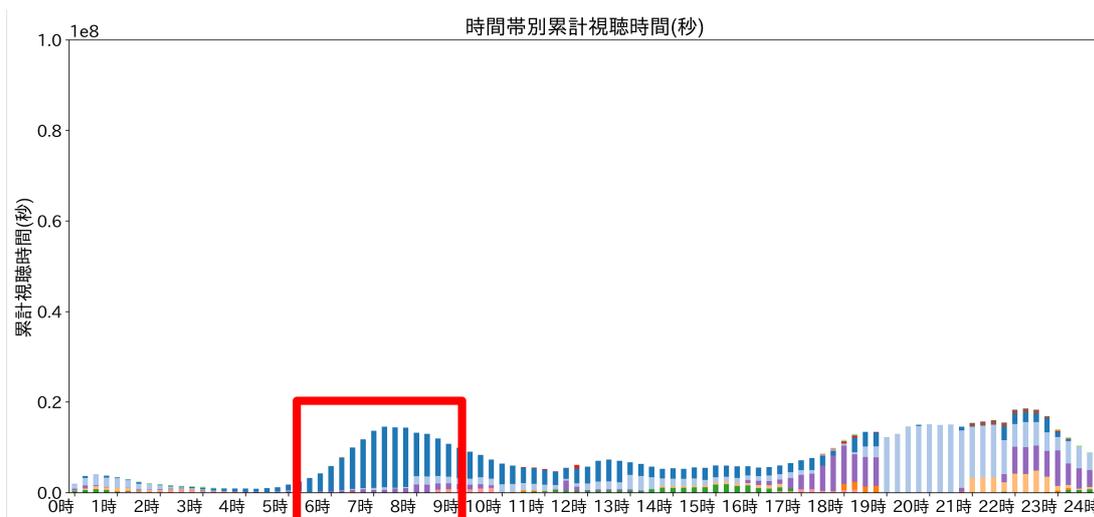


図 36: ライト層/朝/バラエティ少

■ 大型イベント期間における分析結果 通常期間においてクラスタごとの視聴傾向を可視化した結果の一部を図 37~41 に示す。結果としては、大型イベント期間中の、各クラスタの視聴傾向を通常期間と同じ観点から確認したところ似た傾向のクラスタに分類することができた。また大型イベントの影響を受けている

層や受けていない層などを確認することができた。本期間において WBC は 3 月 21 日午前 7 時から 12 時の間準決勝、3 月 22 日 7 時から 12 時 45 分の間に決勝戦を放送していた。また WBC のについては同日 19 時から再放送が行われている。その他にも期間中 3 日間に渡って、19 時から 2,3 時間程度世界フィギュアスケート選手権 2023 が放送されていた。

#### 1. ヘビー層（図 37）の特徴

ボリュームゾーンは朝と夜、視聴ジャンルは音楽やバラエティが多い。このクラスタでは他クラスタと比較するとスポーツの視聴割合が低く、WBC には興味がないがテレビは好きな層だと考えられる。また音楽やバラエティの視聴割合が高いことから女性を多く含んでいると予想される。

#### 2. ミドル層（図 38）の特徴

ボリュームゾーンは朝と夜、視聴ジャンルはスポーツとバラエティが多い。通常期間と同様に、NHK の朝の連続テレビ小説と NHK ニュース視聴のために NHK に移動していると考えられ、この層は基本的に NHK を好んで視聴する層だが、WBC 等の大型イベントにも興味がある層だと予想される。

#### 3. ミドル層（図 39）の特徴

朝や日中はあまりテレビを視聴せず、夜間の視聴時間が長いため、日中は学校・仕事等で外出していると考えられる。番組の視聴カテゴリを見ると音楽やバラエティといった娯楽番組の割合は高いが、スポーツの割合が低いため、WBC にはあまり興味がない比較的若い層だと考えられる。

#### 4. ライト層（図 40）の特徴

日中も見ている、視聴ジャンルとしては深夜アニメの割合が高い。他クラスタと比較してスポーツの視聴割合が低く、深夜アニメの視聴割合が高く、深夜アニメ好きな層は WBC 開催期間であってもスポーツ番組等を見ない層だと予想される。

#### 5. ライト層（図 41）の特徴

ボリュームゾーンは朝と夜、視聴ジャンルはスポーツが多い。全体の視聴

時間は短いながらも、スポーツの割合が大きくなっており、普段あまりテレビは見ないがWBCといったイベントは視聴する層だと予想される。

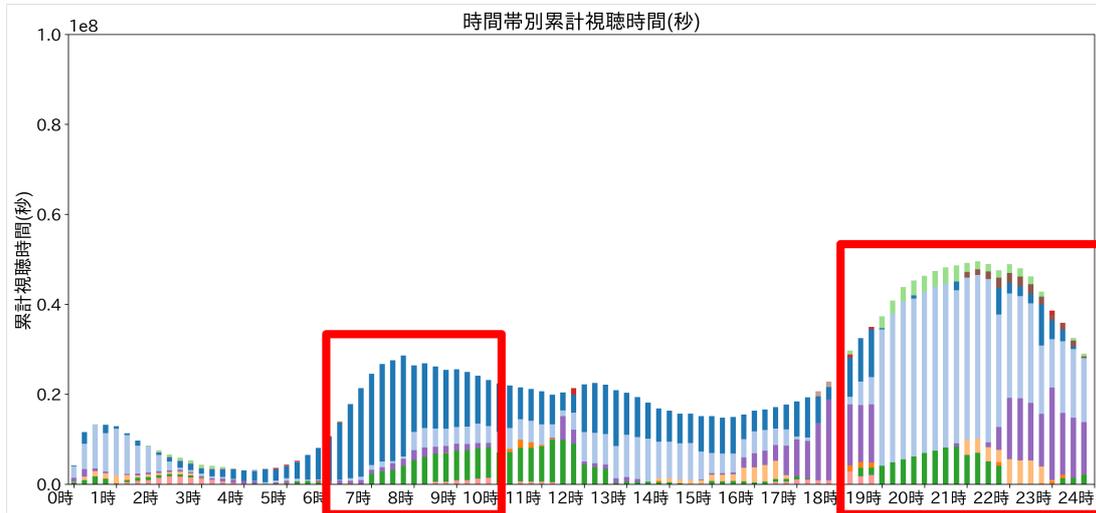


図 37: ヘビー層/朝夜/音楽・バラエティ多

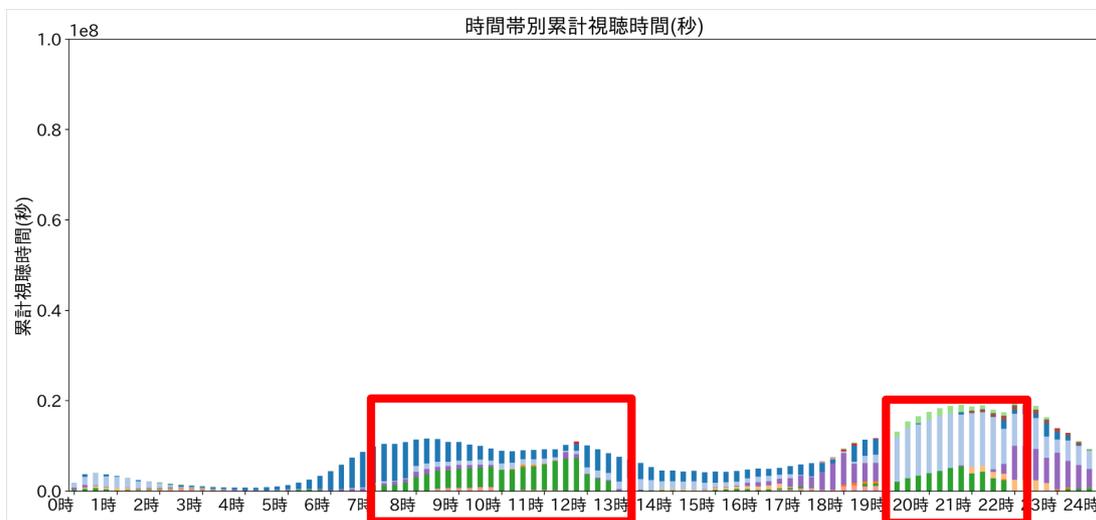


図 38: ミドル層/朝夜/スポーツ多

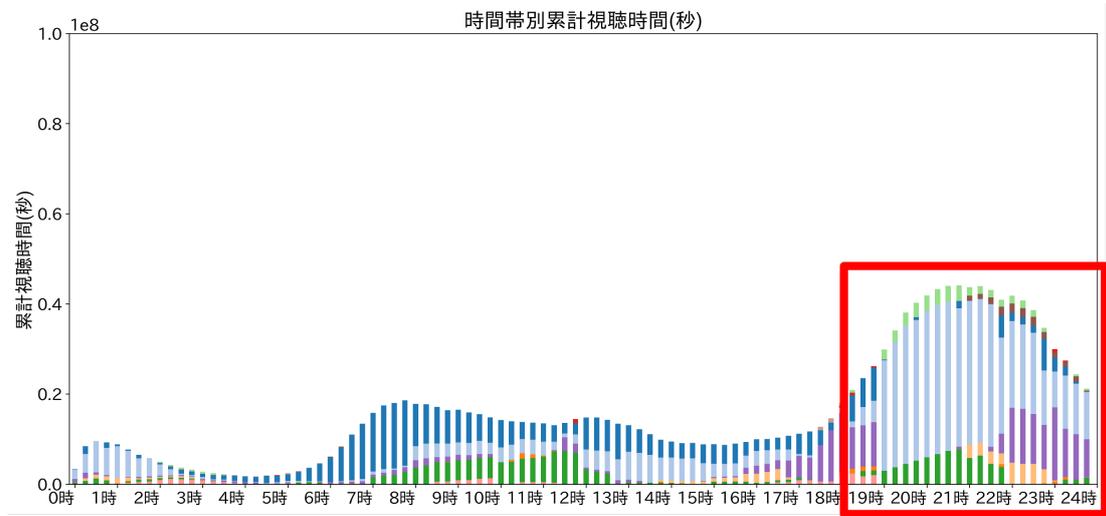


図 39: ミドル層/夜/音楽バラエティ多

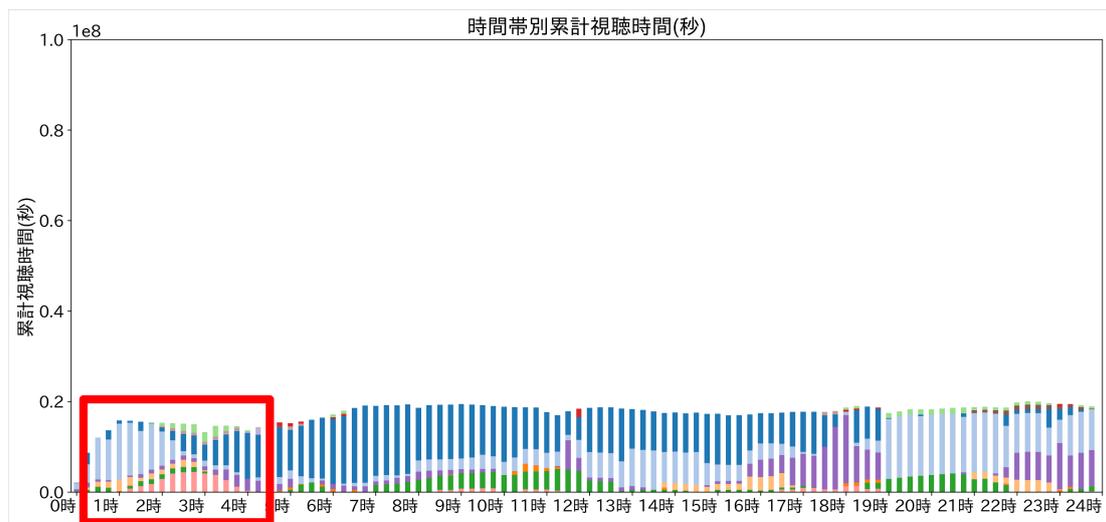


図 40: ライト層/全時間帯/アニメ多

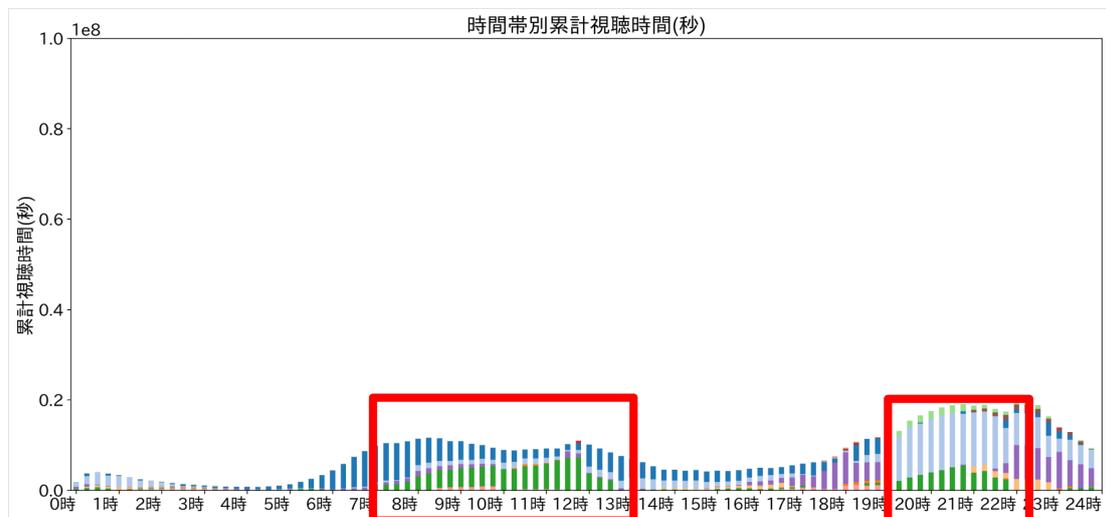


図 41: ライト層/朝夜/スポーツ多

### 3.6.5 まとめ

本分析では、在阪の5つのテレビ局で通常期間(2023年2月20日から2023年2月26日)とWBCが放映されていたイベント期間(2023年3月20日から2023年3月26日)で収集した非特定テレビ視聴履歴データを用いて視聴テレビを異なる視聴傾向に基づいてクラスタリングを行う手法を提案した。結果として、テレビの視聴時間、番組カテゴリ、視聴時間帯の3つの観点で視聴傾向別に分類可能であることが示された。また通常の期間だけでなく、WBCといった特殊なイベントが放送されている期間でも同様に視聴傾向別に分類できることが確認できた。今後は分析するデータ範囲の拡大や、実際に番組制作やCM放映タイミング、営業活用などに活かすことを目標とする。

### 3.7 まとめ

本章では、放送会社における番組制作・CM販売などの実務への利用を想定した分析を実施した。本分析結果の一部はすでに共同研究先である読賣テレビ放送株式会社の業務への活用が始まっている。また、分析結果の一部は、読賣テレビ

放送株式会社が全社員向けに公開しているダッシュボードへの実装もなされており、普段からデータを参照しながら仮説を持ち、判断に活かしていくデータドリブン文化の寄与に大きく貢献している。

## 4. 結論

本論文では、放送業界におけるデータ利活用推進を目指し、筆者が行った研究について記述した。研究内容は、(1) 非特定テレビ視聴履歴データの放送局間におけるデータ統合手法の研究、(2) 非特定テレビ視聴履歴データの分析に関する研究である。以下、得られた成果および今後の課題をまとめる。

非特定テレビ視聴履歴データの放送局間におけるデータ統合手法では、提案手法が企業が採用するだけの性能があることを確認した。本論文ではデータ統合手法として、各社が独自に収集した非特定テレビ視聴履歴データのみを利用して、テレビIDの同一推定を実施した。本提案手法では簡易に同一推定が出来る一方で、テレビ受像機側の利用保存領域やデータを収集するためのデータ放送プログラムにまったくの変更を加えていないため、同一推定精度の向上も限界がある。これは、現在、在京局間や在阪局間で継続して検討が進んでいることから、今後解決されていくと考えられる。また、本研究では限られた期間のデータに対して、提案方式を適用させた。しかし、非特定テレビ視聴履歴データの常時共同利用が開始された場合、本提案方式の実施タイミングと利用する非特定テレビ視聴履歴データの期間の検討が必要となる。

非特定テレビ視聴履歴データの分析では、共同研究先である讀賣テレビ放送株式会社の業務へ活用を目的とした結果、いくつかの分析においては営業や編成・制作業務への活用が実現している。テレビのCM視聴とインターネット検索データの関係性分析では、CM放送がインターネット検索行動に繋がっていることや、インターネット検索行動へ誘導するのに効果的な時間帯を示した。CMと番組コーナー視聴傾向の分析では、テレビ視聴者の『離脱』という強い意思を持った行動に着目し、離脱行動の多寡でCM効果の高い地域や番組・コーナー企画に必要な分析を行った。最後のクラスタリング分析では、視聴者の行動に着目し、好きな放送局や好きな番組ジャンル別にクラスタリングを行うことで、クラス分類できることを示した。今後は、「必ず視聴する番組（番組固定ファン）」「各放送局の視聴時間（放送局ファン）」「天候」なども特徴量に加えていくことで、視聴者の興味関心によるクラス分類を実現したい。今後は、これらの分析結果を讀賣テレビ放送株式会社の業務で利用されるためのシステム化が課題となる。具体的には、

讀賣テレビ放送株式会社が構築している「非特定テレビ視聴履歴データダッシュボードシステム」への本研究結果の実装を検討していく。

## 謝辞

本研究を進めるにあたり、安本慶一教授には、研究全般、論文添削にて多大なるご指導・ご助言を賜りました。また、社会人博士課程として迎え入れていただき、充実した研究環境の整備など、研究活動を手厚くご支援いただきました。感謝の意を表すとともに、心より厚く御礼申し上げます。諏訪博彦准教授には、ご多忙の中、論文審査委員を引き受けてくださった上で、様々なご助言をいただきました。感謝の意を表すとともに、心より厚く御礼申し上げます。千葉工業大学の水本旭洋准教授には、本研究を進めていく中で、深く議論に参加していただき、非常に有益な助言をいただきました。また、研究をはじめとした質問や相談に丁寧に回答してくださいました。感謝の意を表すとともに、心より厚く御礼申し上げます。松田裕貴客員准教授には、本研究を進めていく中で、深く議論に参加していただき、非常に有益な助言をいただきました。提案手法の設計、シミュレーションシステム構築・評価、視聴者のインターネット検索行動分析、クラスタリング手法に関してご指導・ご助言をいただき、研究をはじめとした質問や相談に丁寧に回答してくださいました。感謝の意を表すとともに、心より厚く御礼申し上げます。読売テレビとの共同研究に参加して一緒にプロジェクトを進めてくれた真弓さん、木俣さん、鳥羽さん、由田さん、大阪大学の吉村さん、大河原さん、大野さんにも感謝申し上げます。また、金岡恵事務補佐員、山内奈緒事務補佐員をはじめ、3年間ユビ研でお世話になったスタッフと学生の皆様に感謝申し上げます。荒牧英治教授には、ご多忙の中、論文審査委員を引き受けてくださった上で、貴重なご意見とコメントを賜りました。感謝の意を表すとともに、心より厚く御礼申し上げます。

NNTM アルゴリズムの検証・評価にあたって多大なご協力をいただきました株式会社毎日放送、朝日放送テレビ株式会社、関西テレビ放送株式会社の皆様に、この場を借りて深く感謝申し上げます。また、読売テレビ放送株式会社の皆様に感謝申し上げます。業務と並行しての博士課程でご迷惑をかけることもあったかと思いますが、今後はこの経験を活かして恩返ししていきたいと思っております。

最後に、これまで育ててくれた両親と会社・学生生活の両立を支えてくれた妻、元気に成長してくれた子供に心より感謝します。

## 参考文献

- [1] 東京都都市整備局. マンション実態調査結果. <https://www.mansion-tokyo.metro.tokyo.lg.jp/pdf/03jittai-chousa/03jittai-chousa-01.pdf>, 2013.
- [2] 株式会社ビデオリサーチ. 「コネクテッド tv (ctv) とは？」今さら聞けない!基本の『キ』. <https://www.videor.co.jp/digestplus/market/2023/06/44927.html>, 2024.
- [3] 一般財団法人放送セキュリティセンター視聴関連情報の取扱いに関する協議会. オプトアウト方式で取得する非特定視聴履歴の取扱いに関するプラクティス (ver2.3) . [https://www.sarc.or.jp/documents/www/NEWS/hogo/2024/optout\\_practice\\_ver2.3.pdf](https://www.sarc.or.jp/documents/www/NEWS/hogo/2024/optout_practice_ver2.3.pdf), 2024.
- [4] 株式会社電通. 2023 年日本の広告費. [https://www.dentsu.co.jp/knowledge/ad\\_cost/2023/](https://www.dentsu.co.jp/knowledge/ad_cost/2023/), 2023.
- [5] 読売テレビ放送株式会社. 「テレビ視聴データ連携に関する共同技術検証実験」について. <https://www.ytv.co.jp/privacy/experiments/index.html>, 2021.
- [6] 読売テレビ放送株式会社. 「テレビ視聴データ連携に関する共同技術検証実験 (2022 年度)」について. <https://www.ytv.co.jp/privacy/experiments2022/index.html>, 2022.
- [7] Károly Boda,  F, G Gy Guly, and S Imre. User tracking on the web via cross-browser fingerprinting. In *Proceedings of the 16th Nordic Conference on Information Security Technology for Applications(NordSec'11)*. Springer Verlag, 2012.
- [8] 齋藤祐太, 細谷竜平, 齋藤孝道, 森達哉. クロスアプリケーションフィンガープリンティング-同一端末上のアプリケーション間の紐付け-. コンピュータ

セキュリティシンポジウム 2018 論文集, 第 2 巻, pp. 1251–1258. 情報処理学会, 2018.

- [9] Stijnen T Moons KG Donders AR, van der Heijden GJ. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol*, Vol. 2006;59(10), pp. 1087–91, 2006.
- [10] P Royston. Multiple imputation of missing values: Update. *The Stata Journal*, Vol. 5(2), pp. 188–201, 2005.
- [11] 大野健太. 深層生成モデルによるテーブルデータ生成と仮想人体生成モデルへの応用. *JSBi Bioinformatics Review*, Vol. 5, No. 1, pp. 16–27, 2024.
- [12] 総務省統計局. 平成 30 年住宅・土地統計調査. [https://www.stat.go.jp/data/jyutaku/2018/pdf/g\\_gaiyou.pdf](https://www.stat.go.jp/data/jyutaku/2018/pdf/g_gaiyou.pdf), 2019.
- [13] 菊池匡晃, 坪井創吾, 中田康太. 大規模テレビ視聴データによる番組視聴分析. *デジタルプラクティス*, Vol. 7, No. 4, pp. 352–360, 2016.
- [14] 水岡良彰, 中田康太, 折原良平. 大規模テレビ視聴データによる視聴パターン推移の分析. *人工知能学会全国大会論文集*, pp. 1P203–1P203, 2018.
- [15] Marc G. Weinberger and Charles S. Gulas. The impact of humor in advertising: A review. *Journal of Advertising*, Vol. 21, No. 4, pp. 35–59, 1992.
- [16] 幸志牧野. 広告効果に及ぼすコンテンツ情報の影響に関する研究 (2) : 受け手の気分, CM 内容の印象とテレビ広告の好感度, 商品評価との関連. *経営情報研究 : 摂南大学経営情報学部論集*, Vol. 16, No. 1, pp. 1–11, 2008.
- [17] Margaret C Campbell and Kevin Lane Keller. Brand familiarity and advertising repetition effects. *Journal of consumer research*, Vol. 30, No. 2, pp. 292–304, 2003.
- [18] Lukáš Veverka and Vladimír Holý. The Impact of TV Advertising on Website Traffic. *arXiv:2112.08530*, 2021.

- [19] 加藤拓巳, 津田和彦. ライフログに基づくテレビ広告の three-hit theory の検証. 人工知能学会第 23 回知能流通ネットワーク研究会, Vol. KSN-023, pp. P03–P07, 2018.
- [20] 永星浩一. ザッピングと視聴率を最大化する cm のタイミング. 福岡大学商学論叢, Vol. 55, No. 4, pp. 301–320, mar 2011.
- [21] 永星浩一. テレビ番組の品質と cm のタイミング. 福岡大学商学論叢, Vol. 57, 1・2, pp. 27–49, sep 2012.
- [22] Shinichi YOKOI, Takashi X FUJISAWA, Koji KAZAI, Haruhiro KATAYOSE, and Noriko NAGATA. The effects of the timing of commercial breaks by the measurement of brain activity using fnirs and autonomic nervous activity. *Proc. 13th Korea-Japan*, Vol. 2017, pp. 206–211, 2007.
- [23] 岸本大輝, 井原史渡, 栗原聡. 新型コロナウイルスの感染状況に対するテレビ報道の特徴と報道変容の分析. 人工知能学会全国大会論文集, Vol. JSAI2021, pp. 1J3GS10e02–1J3GS10e02, 2021.

## 研究業績

### 本論文に関する業績

#### 学術論文誌

1. 松田裕貴, 榊原太一, 松田裕貴, 水本旭洋, 安本慶一: 放送局を横断する大規模テレビ視聴履歴データの統合手法の提案と実践, 情報処理学会論文誌デジタルプラクティス (TDP), Vol. 4, No. 1, pp. 34-44, 2023.  
－ 2 章に対応
2. 松田裕貴, 榊原太一, 松田裕貴, 水本旭洋, 安本慶一: 非特定テレビ視聴履歴データの放送局間統合手法, 情報処理学会論文誌ユビキタスコンピューティングシステム (XII) 特集号, (Accepted)  
－ 2 章に対応

#### 国際会議

1. Hiroki Matsuda, Taichi Sakakibara, Daiki Mayumi, Yuki Matsuda, Keiichi Yasumoto: Analysis of Relationship Between Non-Identifiable TV Viewing History Data and Web Search Trends, the 12th Global Conference on Consumer Electronics (GCCE '23), pp. 1109-1110, Nara, Japan, 2023.  
－ 2 章に対応

#### 国内会議

1. 松田裕貴, 榊原太一, 木俣雄太, 鳥羽望海, 真弓大輝, 松田裕貴, 安本慶一: テレビ視聴における非特定視聴履歴データとインターネット検索データの関係性分析, 第 14 回データ工学と情報マネジメントに関するフォーラム (DEIM '22), オンライン, 2022.  
－ 3 章に対応

2. 吉村啓, 大河原一輝, 榊原太一, 松田裕貴, 水本旭洋: テレビ視聴時のCM視聴と地域傾向の分析, ネットワークが創発する知能研究会 (JSST), 社会システムと情報技術研究ウィーク (WSSIT ' 22) , 北海道, 日本, 2022.

－ 3 章に対応

3. 松田裕貴, 榊原太一, 真弓大輝, 松田裕貴, 安本慶一: テレビ視聴行動を再現するエージェントシミュレータの構築に向けた基礎分析, 第 15 回データ工学と情報マネジメントに関するフォーラム (DEIM ' 23) , 岐阜, 日本, 2023.

－ 2 章に対応

4. 木俣雄太, 鳥羽望海, 真弓大輝, 松田裕貴, 榊原太一, 松田裕貴, 安本慶一: テレビCMがネット検索行動に与える影響の調査システムの開発, 社会情報学会 関西支部 (SSI) , 社会システムと情報技術研究ウィーク (WSSIT ' 23) , 北海道, 日本, 2023.

－ 3 章に対応

5. 吉村啓, 榊原太一, 松田裕貴, 水本旭洋: 放送局をまたがる統合視聴データを用いたCM視聴傾向の分析, ネットワークが創発する知能研究会 (JSST), 社会システムと情報技術研究ウィーク (WSSIT ' 23) , 北海道, 日本, 2023.

－ 3 章に対応

6. 由田翔吾, 真弓大輝, 榊原太一, 松田裕貴, 松田裕貴: 非特定視聴履歴データを用いたテレビ視聴傾向クラスタリング手法の検討, ネットワークが創発する知能研究会 (JSST), 社会システムと情報技術研究ウィーク (WSSIT ' 24) , 北海道, 日本, 2024.

－ 3 章に対応

## その他の業績

### 国内会議

1. 松田裕貴, 松田裕貴: 珠算学習支援のための盤面認識に基づくリアルタイム情報提示手法, マルチメディア, 分散, 協調とモバイル (DICOMO 2023) シンポジウム, pp. 63-67, 富山県富山市, 2023.