

非特定テレビ視聴履歴データの 視聴時間帯特徴に基づく視聴行動クラスタリング分析

林 虎太朗[†] 松田 悠斗[†] 松田 裕貴^{††} 池尻 拓巨^{††} 横田 哲弥^{††}
松田 裕貴[†]

[†]岡山大学 〒700-0082 岡山県岡山市北区津島中3丁目1-1

^{††}読売テレビ放送株式会社 〒540-8510 大阪市中央区城見1丁目3番50号

E-mail: [†]{kotaro.hayashi, yuto.matsuda}@cocolab.jp, yukimat@okayama-u.ac.jp,
[‡]{hiroki.matsuda, takumi.ikejiri, tetsuya.yokota}@ytv.co.jp

あらまし 近年、インターネット接続テレビの普及により、大規模な非特定テレビ視聴履歴データを用いた視聴行動分析が可能となっている。本稿では、在阪4局の非特定テレビ視聴履歴データを用い、朝時間帯の視聴行動と他の時間帯のチャンネル視聴行動の関連性を分析した。2022年10月から2023年3月までに取得されたデータ（各月約145万台分）を対象とし、平日・土曜日・日曜日の3区分に分け、朝の視聴データから抽出した特微量を用いてクラスタリング分析を行った。その結果、朝に特定チャンネルを中心とするクラスタでは、他の時間帯においても同一チャンネルの視聴割合が高い傾向が確認された。この結果は、朝時間帯の視聴習慣が1日を通じたチャンネル選択傾向に影響を与える可能性を示唆している。

キーワード 非特定テレビ視聴履歴データ、クラスタリング、テレビ視聴行動分析

Clustering Analysis of Audience Behavior Based on Viewing Time Patterns Using Non-Identifiable TV Viewing History Data

Kotaro HAYASHI[†], Yuto MATSUDA[†], Hiroki MATSUDA^{††}, Takumi IKEJIRI^{††}, Tetsuya YOKOTA^{††},
and Yuki MATSUDA[†]

[†]Okayama University Tsushima-naka 3-1-1, Kita-ku, Okayama, Okayama 700-0082, Japan

^{††}Yomiuri Telecasting Corporation Shiromi 1-3-50, Chuo-ku, Osaka 540-8510, Japan

E-mail: [†]{kotaro.hayashi, yuto.matsuda}@cocolab.jp, yukimat@okayama-u.ac.jp,
[‡]{hiroki.matsuda, takumi.ikejiri, tetsuya.yokota}@ytv.co.jp

Abstract In recent years, the widespread adoption of internet-connected television has enabled viewing behavior analysis using large-scale non-identifiable TV viewing history data. This study analyzes the relationship between morning viewing behavior and channel viewing patterns during other time periods using data from four broadcast stations in the Osaka area. The dataset, collected between October 2022 and March 2023, consists of approximately 1.45 million TV sets per month. The data were divided into three categories: weekdays, Saturdays, Sundays, and clustering analysis was conducted using features extracted solely from morning viewing data. The results show that clusters characterized by concentrated viewing of specific channels in the morning also exhibit higher viewing proportions for the same channels during other time periods. These findings suggest that morning viewing habits may influence channel selection tendencies throughout the day.

Key words non-identifiable TV viewing history data, clustering, TV viewing behavior analysis

1. はじめに

伴い、大規模な視聴データを活用した視聴行動分析が可能となっている。これにより、視聴者の行動パターンを詳細に分析

近年、インターネット接続テレビ（Connected TV）の普及に

し、番組編成や広告配信の最適化に資する知見を得ることが期待されている。在阪地区の地上波放送局では、視聴者の個人情報を含まず、オプトアウト方式で収集されるデータである「非特定テレビ視聴履歴データ」が収集されている。このデータは視聴開始・終了時刻、テレビ ID、視聴チャンネルの情報のみを含み、視聴者の属性情報などは取得されないことが特徴である。また、一般社団法人放送セキュリティセンター（SARC）のプラクティス[1]に基づき、第三者提供の禁止や保存期間の制限が設けられている。この方式の利点は、視聴者の負担なく大規模データを収集できる点にあるが、一方で個人属性が不明であるため、視聴者層の特性把握が困難という課題がある。

テレビ業界では、朝時間帯の視聴習慣が視聴者のチャンネル選択行動に重要な影響を与えると考えられてきた。朝に特定のチャンネルを視聴する習慣が確立された視聴者は、他の時間帯でも同じチャンネルを選択する傾向があるという仮説である。近年、地上波放送の視聴履歴データを用いた視聴行動分析は進展しているものの、時間帯間の視聴行動の関連性については十分に検討されていない。この関係性を定量的に示すことができれば、番組編成戦略や投資配分の最適化に向けた具体的な示唆を得ることができる。そこで本稿では、2022年10月から2023年3月までの6か月間における在阪地区の地上波放送局4局の非特定テレビ視聴履歴データ（各月約145万台分）を用い、朝時間帯の視聴行動パターンと他の時間帯の視聴行動との関連性を検証することを目的とし、データ分析を行った。2022年10月の平日の視聴データにクラスタリング分析を行い他の時間帯の視聴傾向との関連性分析を行った結果、朝に特定チャンネルを中心に視聴するクラスタでは、ほかの時間帯においても同一チャンネルの視聴割合が高いことが確認された。これらの結果は、朝時間帯の視聴行動がチャンネル選択傾向と関連していることを示唆する。

2. 関連研究

2.1 ネットワーク対応テレビによる包括的視聴分析

インターネット接続機能を持つネットワーク対応テレビ（Connected TV）のデータを用いた研究では、リアルタイム視聴のみならず、録画視聴やVODを含めた包括的な視聴行動の分析が行われている。菊池ら[2]は、東芝製テレビから収集された全国約24万台の視聴データ（約4年間）を分析し、リアルタイム視聴とタイムシフト視聴の実態を調査した。この研究では、ドラマ全話を通じた視聴動向の変化を追跡するシステムを開発し、広告主やコンテンツ制作者への情報提供を目指している。また、一部ユーザのアンケート回答をもとに機械学習により世帯構成員の性別・年代属性を推定する手法を提案し、F値0.7以上の精度を達成した。水岡ら[3]は、同様のデータを用いて視聴パターンの遷移を自動抽出する手法を開発し、番組ジャンルや曜日ごとの視聴動向の変化を分析した。これらの研究は、視聴行動の長期的变化を捉える手法を確立した点で重要である。しかし、録画やVODを含む分析であるため、地上波リアルタイム視聴における分析には適していない。また、個別世帯の属

性推定を目的としており、視聴行動パターンに基づいて視聴者群を分類する本研究のアプローチとは目的が異なる。

2.2 地上波放送における非特定テレビ視聴履歴データの活用

地上波放送に特化した非特定テレビ視聴履歴データを用いた研究も進展している。筆者らの研究グループでは在阪5局（読売テレビ、毎日放送、朝日放送、関西テレビ、テレビ大阪）が収集する非特定テレビ視聴履歴データを用いて、様々な観点から視聴行動分析を行なってきた。松田らは、複数放送局の視聴履歴データを統合する手法を開発し、局を横断した視聴行動分析を可能にした[4], [5]。さらに、視聴履歴データとGoogle Trendsの検索トレンドを統合することで、テレビCMが視聴者のオンライン行動に与える影響を定量的に分析している[6]。さらに、吉村らは地域別の視聴履歴データを分析し、CMの完視聴率に地域差が存在することを示した[7]。視聴者の分類に関する研究も進められている。由田らは、視聴時間に基づく特微量を用いて視聴者を視聴パターンに応じてクラスタリングした上で、番組カテゴリ別の視聴割合を考慮したサブクラスタリングを実施した[8]。さらに、視聴者のクラスタ遷移に着目し、視聴行動の安定性および変動性を評価している。一方、真弓らは、大規模な非特定テレビ視聴履歴データを用いて、時間帯やコンテンツ特性に基づく視聴傾向の分類モデルを構築し、その結果に基づいて視聴行動をデジタルツインとして再現する枠組みを提案している[9], [10]。この研究では、視聴傾向分類の高い再現性と、行動パターンの仮想的な再現可能性が示されている。また、時間帯を跨いだ視聴行動の関連性に関する研究として、松田らは、就寝前後の視聴行動に着目し、時間帯間の視聴行動の関連性を定量的に検証した[11]。分析の結果、就寝前に視聴していたチャンネルが起床後も継続して視聴される傾向があり、起床後の平均継続視聴時間は約30分、起床後1時間以内の平均合計視聴時間は約50分であることを明らかにした。

2.3 本研究の位置付け

先行研究のように、視聴履歴データを用いた視聴行動の分析の手法が確立してきた。しかし、朝の時間帯という特定の時間帯の視聴パターンが、他の複数時間帯の視聴行動とどのような関連性を持つかは十分に検討されていない。本研究では、朝時間帯の視聴行動を特微量としたクラスタリングにより視聴者群を分類し、各クラスタにおける他時間帯の視聴行動を比較する。これにより、朝時間帯の視聴習慣と他時間帯のチャンネル選択行動との関連性を明らかにし、番組編成戦略や投資配分の最適化に向けた実践的な示唆を提供することを目指す。

3. データと分析手法

3.1 使用データとデータ前処理

本研究では、2022年10月3日から2023年3月31日までの6か月間にわたり収集された在阪4局の非特定テレビ視聴履歴データを使用した。データ収集は在阪5局で行われたが、1局については収集データ量が他局と比較して極端に少なかったため、分析対象から除外した。対象データは各月約145万台のテレビから収集された視聴履歴であり、視聴開始・終了時刻、テ

表 1: 分析対象データ数

月	平日	土曜	日曜
2022 年 10 月	20 日 1,355,154 台	4 日 1,311,251 台	4 日 1,318,924 台
2022 年 11 月	20 日 1,350,256 台	4 日 1,305,919 台	4 日 1,320,062 台
2022 年 12 月	17 日 1,345,476 台	3 日 1,281,933 台	3 日 1,296,725 台
2023 年 01 月	11 日 1,323,470 台	2 日 1,236,752 台	2 日 1,244,560 台
2023 年 02 月	19 日 1,336,353 台	4 日 1,290,526 台	4 日 1,295,203 台
2023 年 03 月	22 日 1,356,046 台	4 日 1,306,625 台	4 日 1,309,037 台

表 2: 抽出した特徴量一覧（朝時間帯の定義：5:00～8:00）

特徴量	説明
各チャンネル視聴割合	各チャンネル (A, B, C, D) の視聴割合
総視聴時間対数	総視聴時間の対数

レビ ID, 視聴チャンネルの情報を含む。視聴行動は曜日によって大きく異なるため、平日・土曜日・日曜日の 3 区分で分析を行った。ただし、視聴行動が通常パターンと異なると予想される以下の期間のデータは分析対象から除外した。

- 祝日: 通常日と放送内容や視聴行動傾向が異なることが想定されるため
- 年末年始（2022 年 12 月 24 日～2023 年 1 月 8 日）: 特有な放送内容や視聴行動傾向が生じることが想定されるため
- 2023 年 1 月 9 日～16 日: 非特定テレビ視聴履歴データの収集期間外であったため

さらにデータの前処理として、視聴行動の時系列変化を追跡するため、2022 年 10 月と 2023 年 3 月の両方において視聴ログが存在するテレビのみを対象とした。また、24 時間以上連続して同一チャンネルを視聴しているログはつけっぱなしや記録エラーの可能性があるため除去した。これらの前処理により、元データ（各月約 145 万台）から約 7% が除外された。前処理後のデータを平日・土曜日・日曜日に曜日区分した結果、表 1 に示すとおり、平日で約 135 万台、土曜日・日曜日で約 130 万台が分析対象となった。

3.2 特徴量抽出とクラスタリング

朝時間帯の視聴行動を特徴づけるため、表 2 に示す特徴量を抽出した。具体的には朝時間帯（5:00～8:00）における各チャンネル (A～D) の視聴割合と総視聴時間の対数である。総視聴時間については、分布が右に偏っているため対数変換を施し、特徴量間スケールを調整した。朝時間帯に視聴がないテレビについても視聴時間を 0 として特徴量を算出し、全てのテレビを対象にクラスタリングを行った。

クラスタ数 k の決定には、Inertia, Silhouette Score, Davies-Bouldin Score, Calinski-Harabasz Score の 4 つの評価指標を用いて候補を絞り込んだ後、複数の k 値でクラスタリングを実施し、各クラスタに対して解釈可能な特徴づけができるかを検討した。その結果、平日・土曜日・日曜日のいずれにおいても $k = 6$ が最適であると判断した。各曜日区分（平日・土曜日・日曜日）について、2022 年 10 月のデータでクラスタリングモデルを構築し、11 月以降の各月にこのモデルを適用してクラスタを割り当てた。

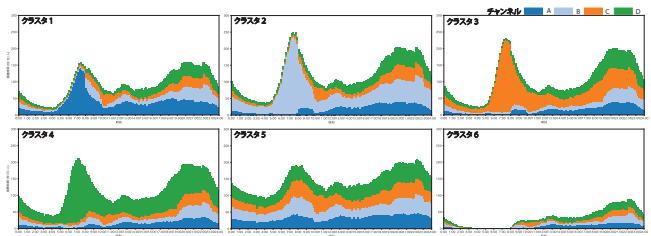


図 1: 平日クラスタリング結果

4. 分析結果・考察

本章では、分析・考察の代表例として 2022 年 10 月の平日データを用いる。図 1 に 6 つのクラスタの時間帯別、チャンネル別視聴時間を示す。 $k = 6$ のクラスタリングの結果、朝時間帯に各チャンネルの視聴傾向が強いクラスタ（4 つ）、朝時間帯の各チャンネルの視聴傾向が均等なクラスタ、朝の視聴がほとんどないクラスタに分類することができた。各クラスタの特徴は以下の通りである。

- クラスタ 1 【朝チャンネル A 中心型】
朝の時間帯にチャンネル A を中心に視聴するクラスタ
- クラスタ 2 【朝チャンネル B 中心型】
朝の時間帯にチャンネル B を中心に視聴するクラスタ
- クラスタ 3 【朝チャンネル C 中心型】
朝の時間帯にチャンネル C を中心に視聴するクラスタ
- クラスタ 4 【朝チャンネル D 中心型】
朝の時間帯にチャンネル D を中心に視聴するクラスタ
- クラスタ 5 【朝均等型】
朝の時間帯に各チャンネル分散して視聴するクラスタ
- クラスタ 6 【朝低視聴型】
朝の時間帯にテレビをほとんど視聴しないクラスタ

4.1 朝視聴行動と他時間帯視聴行動の関連性

図 1 より、朝時間帯に特定のチャンネルを中心に視聴するクラスタでは、他の時間帯においても同一チャンネルの視聴割合が高い傾向がみられる。本研究では、朝時間帯（5:00～8:00）の視聴行動に基づくクラスタリングが、他時間帯の視聴割合にも実質的な差をもたらすか検証した。本研究で使用したデータセットはサンプルサイズが非常に大きいため、Welch の t 検定等の帰無仮説検定では、実務上無視できる微細な差であっても統計的有意差 (p 値 < 0.05) が生じやすい。そのため、本稿では統計的有意差のみならず、実質的な差に焦点を当てた。効果量は、当該時間帯において視聴が発生したテレビ（総視聴時間 > 0）に限定し、クラスタ内における対象チャンネルの平均視聴割合 p_{in} と、クラスタ外における同様の割合 p_{out} との差 ($\Delta = p_{in} - p_{out}$) として定義した。この効果量の不確実性を評価するため、ユーザ単位の視聴割合に対して非パラメトリックブートストラップ法を 1000 回実施し、95% 信頼区間を推定した。ブートストラップ法を採用した理由は、本研究で用いる視聴割合が、区間 [0,1] に制約され、0 および 1 付近に値が集中する歪んだ分布を示すため、正規近似に基づく信頼区間の妥当

性が保証されないためである。実質的な差の判定には、Region of Practical Equivalence (ROPE) を導入した。視聴割合の差が ± 0.15 以内であれば実務上無視できる差、 ± 0.20 を超える場合は特に大きな差と定義した。この閾値は、仮に当該時間帯に120分の総視聴時間があった場合、15%の差は18分、20%の差は24分に相当することに基づく。これだけの差があれば、偶然やザッピングの影響を超えて、視聴者が意図的にそのチャンネルを選択していると判断できるとの現場協議により設定した。95%信頼区間がROPEの外側に完全に位置する場合、実質的に意味のある差が存在すると判断した。

推定結果を表3、表4、表5、表6に示す。例として表6(チャンネルD・クラスタ4)に注目すると、朝時間帯(5:00~8:00)では特に顕著な効果量(0.775)を示しており、95%信頼区間がROPE(± 0.20)を大きく超えている。これは、クラスタリングによって朝の時間帯にチャンネルDを視聴する割合が高いテレビを明確に分類できていることを示している。朝時間帯以外では、8:00~10:00の効果量が最も高く(0.450)、95%信頼区間もROPE(± 0.20)の外側に位置した。これは、朝時間帯に視聴していたテレビが引き続き同じチャンネルを視聴する傾向が高いためと考えられる。また、特微量として用いた時間帯から離れた日中の時間帯(10:00~18:00)においても、効果量は0.222~0.294の範囲にあり、いずれも95%信頼区間もROPE(± 0.20)の外側に位置し、実質的な差が確認された。一方で、ゴールデンタイム(19:00~22:00)およびプライムタイム(19:00~23:00)では、他の時間帯と比較して視聴割合差は小さくなっている(19:00~21:00で0.188、21:00~23:00で0.144)。この要因として、この時間帯には各チャンネルが特に特徴的な番組編成を行うことが多く、視聴者がチャンネルそのものではなく番組内容や出演者に基づいて視聴を選択する傾向が強まる可能性が考えられる。他のクラスタ(表3、表4、表5)においても、表6と概ね同様の傾向が確認された。また、各クラスタで抽出された、朝帯特化チャンネルについては、他のいずれの時間帯においても効果量が負となるケースは確認されなかった。この結果は、朝時間帯における特化視聴が、他時間帯の視聴行動とも一定程度関連している可能性を示唆している。

4.2 視聴割合分布の検証

視聴割合の差が一部の外れ値によるものではないか確認するため、視聴割合の分布を可視化した(図2^(注1))。上段のヒストグラム(Density)では、特に月間総視聴時間が短い層(<2h)において0と1に鋭いピークがみられる。これは当該層が「そのチャンネルを全く見ない(0)」か「そのチャンネルしか見ない(1)」という極端な視聴行動をとりやすく、かつ各総視聴時間帯グループごとに積分値が1になるように正規化されているためその傾向が視覚的に強調されていることを示している。クラスタ間の差異(クラスタ4とそれ以外の比較)については以下のようない傾向が読み取れる。上段のグラフをみるとクラスタ4ではクラスタ外に比べ、チャンネルDの視聴割合が0の層が

(注1)：本稿では例として2022年10月平日12:00~14:00のチャンネルDのクラスタ4とそれ以外のクラスタの視聴割合の分布を示す

表3: クラスタ特化チャンネルの効果量推定(チャンネルA・クラスタ1)

時間帯	効果量 ^(注1)	95% CI	P_{in} ^(注2)	P_{out} ^(注3)	n ^(注4)	N ^(注5)
0:00~5:00	0.198*	[0.196, 0.201]	0.393	0.195	87,379	892,670
5:00~8:00	0.785**	[0.785, 0.786]	0.865	0.079	132,654	1,111,944
8:00~10:00	0.294**	[0.292, 0.297]	0.439	0.145	115,080	1,103,170
10:00~12:00	0.186*	[0.184, 0.188]	0.352	0.166	102,030	997,490
12:00~14:00	0.183*	[0.181, 0.185]	0.429	0.246	105,881	1,040,282
14:00~16:00	0.165*	[0.163, 0.168]	0.397	0.232	102,038	1,009,497
16:00~17:00	0.173*	[0.171, 0.176]	0.422	0.249	97,750	967,375
17:00~18:00	0.178*	[0.175, 0.180]	0.411	0.233	105,990	1,052,384
18:00~19:00	0.180*	[0.178, 0.183]	0.378	0.198	114,005	1,140,359
19:00~21:00	0.106	[0.105, 0.108]	0.314	0.207	124,913	1,262,396
21:00~23:00	0.077	[0.075, 0.078]	0.276	0.200	123,804	1,263,046
23:00~24:00	0.109	[0.107, 0.111]	0.302	0.194	104,512	1,081,986

(注1) 効果量 = $P_{in} - P_{out}$ (割合の差)

* 95% 信頼区間が ROPE (± 0.15) 外 (実質的に意味のある差)

** 95% 信頼区間が ROPE (± 0.20) 外 (特に大きな差)

(注2) クラスタ内の予測視聴割合

(注3) クラスタ外の予測視聴割合

(注4) クラスタ内有効テレビ台数 (視聴時間 > 0 秒)

(注5) 時間帶有効テレビ台数 (視聴時間 > 0 秒)

表4: クラスタ特化チャンネルの効果量推定(チャンネルB・クラスタ2)

時間帯	効果量	95% CI	P_{in}	P_{out}	n	N
0:00~5:00	0.213**	[0.211, 0.214]	0.441	0.228	201,006	892,670
5:00~8:00	0.797**	[0.796, 0.797]	0.893	0.096	287,853	1,111,944
8:00~10:00	0.410**	[0.409, 0.412]	0.602	0.192	259,243	1,103,170
10:00~12:00	0.204**	[0.203, 0.206]	0.371	0.167	227,425	997,490
12:00~14:00	0.158*	[0.157, 0.160]	0.336	0.178	235,606	1,040,282
14:00~16:00	0.138	[0.136, 0.139]	0.295	0.157	228,902	1,009,497
16:00~17:00	0.176*	[0.174, 0.177]	0.332	0.156	222,681	967,375
17:00~18:00	0.203**	[0.201, 0.204]	0.352	0.149	242,297	1,052,384
18:00~19:00	0.204**	[0.203, 0.206]	0.349	0.145	258,985	1,140,359
19:00~21:00	0.118	[0.117, 0.119]	0.328	0.210	276,838	1,262,396
21:00~23:00	0.116	[0.115, 0.117]	0.388	0.272	276,179	1,263,046
23:00~24:00	0.163*	[0.161, 0.164]	0.409	0.246	241,431	1,081,986

表5: クラスタ特化チャンネルの効果量推定(チャンネルC・クラスタ3)

時間帯	効果量	95% CI	P_{in}	P_{out}	n	N
0:00~5:00	0.160*	[0.158, 0.161]	0.359	0.199	185,167	892,670
5:00~8:00	0.799**	[0.798, 0.799]	0.893	0.094	272,243	1,111,944
8:00~10:00	0.465**	[0.463, 0.466]	0.662	0.197	246,647	1,103,170
10:00~12:00	0.269**	[0.267, 0.270]	0.571	0.303	211,303	997,490
12:00~14:00	0.224**	[0.222, 0.226]	0.397	0.173	217,306	1,040,282
14:00~16:00	0.206**	[0.205, 0.208]	0.360	0.154	210,021	1,009,497
16:00~17:00	0.214**	[0.212, 0.216]	0.397	0.183	201,103	967,375
17:00~18:00	0.235**	[0.233, 0.236]	0.419	0.185	221,309	1,052,384
18:00~19:00	0.226**	[0.225, 0.228]	0.431	0.205	241,505	1,140,359
19:00~21:00	0.117	[0.116, 0.118]	0.343	0.226	261,576	1,262,396
21:00~23:00	0.097	[0.096, 0.098]	0.297	0.200	259,661	1,263,046
23:00~24:00	0.109	[0.108, 0.110]	0.277	0.168	226,402	1,081,986

減少し、1の層が増加していることがどの総視聴時間の層においても読み取れる。また、中段のヒストグラム(0と1を除外し見やすくしたもの)においても、クラスタ4の分布は全体的に右側(高視聴割合側)にシフトしている。さらに、下段グラフに示す通り、総視聴時間の長短に関わらず、全ての層においてクラスタ4はクラスタ外より「視聴割合0」が少なく、「視聴割合1」または中間層への比率が高くなっている。以上より算出された効果量は一部の極端なテレビのみに起因するものではなく、視聴時間の長短に関わらずクラスタ4全体としてチャンネルDを選択する傾向が強いことに起因すると結論づけられる。

表 6: クラスタ特化チャンネルの効果量推定(チャンネル D・クラスタ 4)

時間帯	効果量	95% CI	P_{in}	P_{out}	n	N
0:00~5:00	0.216**	[0.214, 0.217]	0.457	0.241	149,804	892,670
5:00~8:00	0.775**	[0.774, 0.776]	0.863	0.088	210,096	1,111,944
8:00~10:00	0.450**	[0.448, 0.452]	0.610	0.160	184,758	1,103,170
10:00~12:00	0.294**	[0.292, 0.296]	0.488	0.194	163,027	997,490
12:00~14:00	0.235**	[0.233, 0.237]	0.499	0.263	167,624	1,040,282
14:00~16:00	0.222**	[0.220, 0.224]	0.552	0.330	162,455	1,009,497
16:00~17:00	0.246**	[0.243, 0.248]	0.515	0.269	156,490	967,375
17:00~18:00	0.281**	[0.279, 0.283]	0.555	0.274	170,165	1,052,384
18:00~19:00	0.288**	[0.286, 0.290]	0.581	0.293	184,623	1,140,359
19:00~21:00	0.188*	[0.187, 0.189]	0.454	0.266	200,328	1,262,396
21:00~23:00	0.144	[0.143, 0.145]	0.397	0.253	199,486	1,263,046
23:00~24:00	0.156*	[0.154, 0.157]	0.453	0.297	174,600	1,081,986

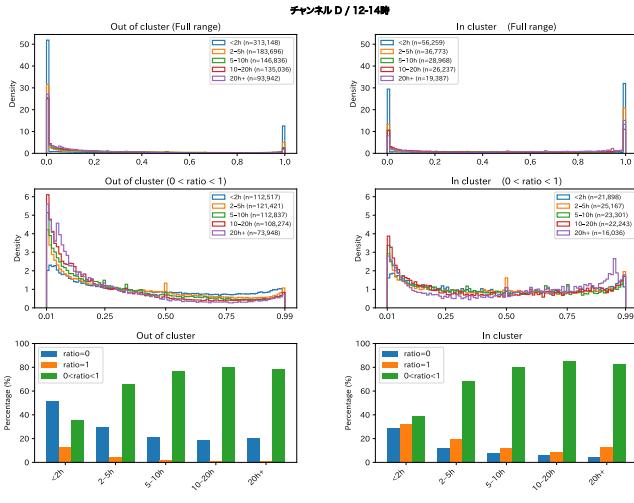


図 2: クラスタ 4 内外の分布の可視化

4.3 個人視聴分析

これまでの分析では、各クラスタの集団として捉え、その平均的な視聴特性を明らかにしてきた。一方で、同一クラスタ内において個々の視聴者が実際にどのような視聴行動をとっているかを確認することは、クラスタの性質を解釈する上で重要である。そこで、各クラスタからランダムに 10 台ずつテレビを抽出し、2022 年 10 月の平日における個人単位の視聴行動を可視化した（図 3）。図右側の各パネルは、横軸を日付、縦軸を時刻（00:00～24:00）とした 2 次元マップであり、色がついている部分はその時間帯において当該色のチャンネルの視聴割合が最も高かったことを示す。可視化の結果、同一クラスタに属していても、その視聴パターンには顕著な個人差が存在することが確認された。具体的には、朝などの特定時間帯のみに明瞭な同一色の横線が見られ、局的に強固な視聴ルーティーンが確立されているテレビもあれば、一日を通して規則正しく視聴パターンが維持されているテレビ、あるいは、視聴が散発的で明確な習慣が見られないテレビなどである。この結果は、本研究のクラスタリングが集団レベルの主要な傾向を適切に捉えている一方で、その背後には多様な生活スタイルや習慣強度が混在していることを示唆している。したがって、今後は本研究のクラスタリングにさらにサブクラスタリングを行うなどして、細粒度の分析を行いさらに時系列分析などを行うことによって、視聴習慣の形成過程や変化の兆候をより詳細に解明できる可能性がある。

5. おわりに

本研究では、朝時間帯の視聴行動に基づくクラスタリングを通じて、視聴者のチャンネル選択傾向が一日の中でどのように持続するかを分析した。その結果、朝に特定チャンネルを中心に視聴するクラスタでは、他時間帯においても同一チャンネルの視聴割合が一貫して高く、多くの時間帯で実質的に意味のある差が確認された。一方で、ゴールデンタイムやプライムタイムではその差が相対的に小さくなり、番組内容主導の視聴行動が強まる可能性が示唆された。個人単位の可視化からは、同一クラスタ内であっても視聴習慣の強度や安定性には大きな個人差が存在することが確認された。今後は、サブクラスタリングや時系列解析を通じて、視聴習慣の形成過程や変化の兆候をより詳細に分析することが課題である。

文 献

- [1] 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 (ver2.1).,” https://www.sarc.or.jp/documents/www/NEWS/hogo/2021/optout_practice_ver2.1.pdf, 2021. Accessed on Feb. 28, 2025.
- [2] 菊池匡晃, 坪井創吾, 中田康太他, “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.
- [3] 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,” 第 32 回人工知能学会全国大会論文集, pp.1P203–1P203, 2018.
- [4] 松田裕貴, 柳原太一, 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.
- [5] 松田裕貴, 柳原太一, 真弓大輝, 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, vol.65, no.10, pp.1488–1500, 2024.
- [6] 松田裕貴, 柳原太一, 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), pp.1–6, 2022.
- [7] 吉村 啓, 水本旭洋, 柳原太一, 松田裕貴, “テレビ視聴時の cm 離脱と地域傾向分析,” 人工知能と知識処理研究会, vol.121, pp.43–48, 2022.
- [8] 由田翔吾, 松田裕貴, 松田裕貴, 横田哲弥, 柳原太一, 安本慶一, “非特定テレビ視聴履歴データに基づくテレビ視聴行動クラスタリングと視聴傾向変化分析,” AIoT 行動変容学会第 9 回研究会 (BTI-9), pp.28–34, 2025.
- [9] 真弓大輝, 松田裕貴, 松田裕貴, 横田哲弥, 柳原太一, 安本慶一, “非特定テレビ視聴履歴データを用いたテレビ視聴行動クラスタリング手法の構築と評価,” AIoT 行動変容学会第 8 回研究会 (BTI-8), pp.23–30, 2024.
- [10] D. Mayumi, H. Matsuda, T. Yokota, T. Sakakibara, Y. Matsuda, and K. Yasumoto, “Clustering tv viewing behavior for digital twin construction using television viewing history data,” IEEE Access, vol.13, pp.192795–192806, 2025.
- [11] 松田悠斗, 林虎太朗, 松田裕貴, 池尻拓巨, 横田哲弥, 松田裕貴, “非特定テレビ視聴履歴データを用いた就寝前後における視聴チャンネルの関係性分析,” 研究報告ユビキタスコンピューティングシステム (UBI), 第 2025-UBI-87 卷, pp.1–8, 2025.

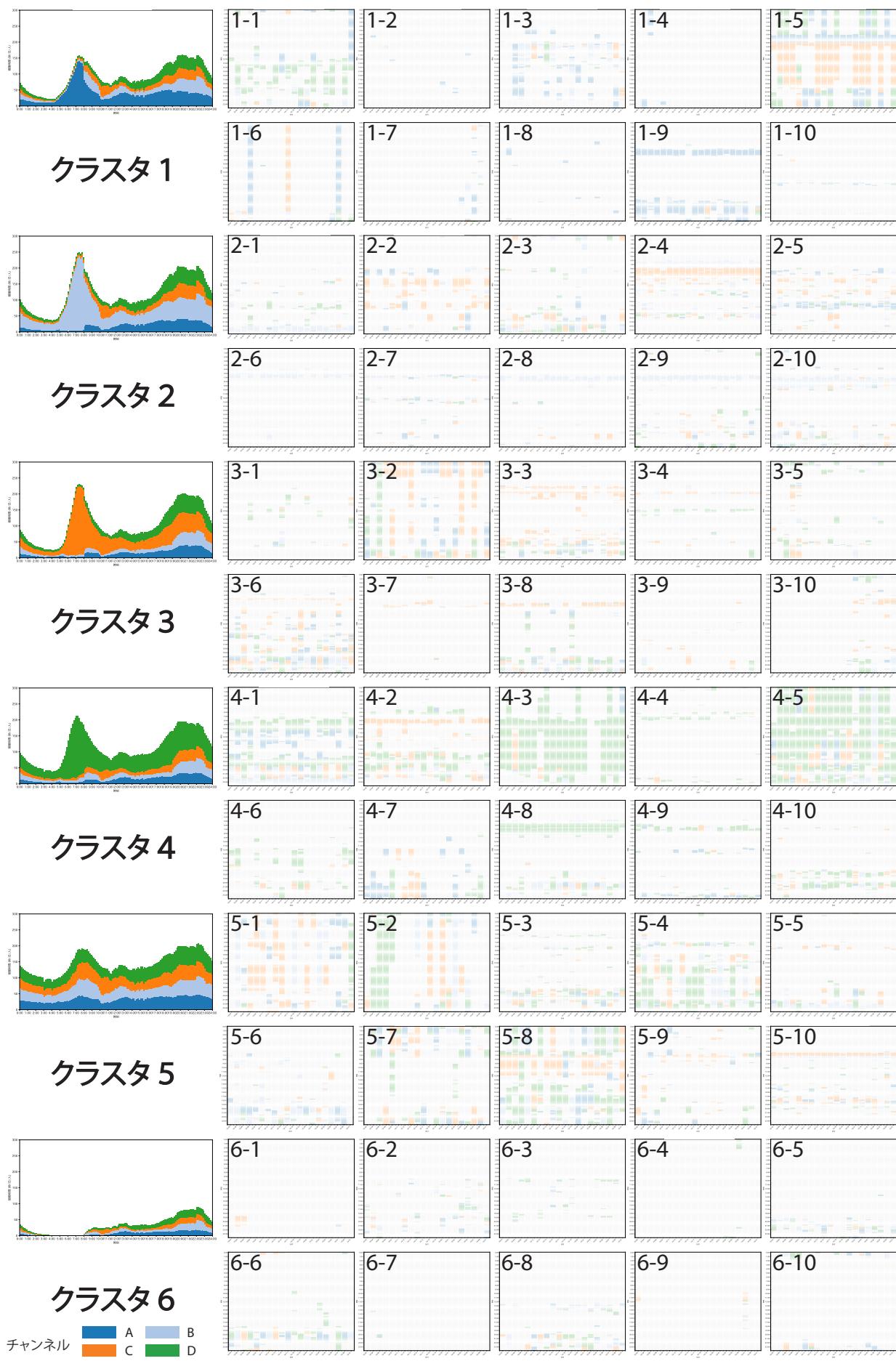


図 3: 個人視聴の可視化