

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

Exploration of Television Viewing Trend Clustering Method Using Non-Specific Viewing History Data

由田 翔吾^{1*} 真弓 大輝¹ 榊原 太一²
Shogo Yoshida¹ Daiki Mayumi¹ Taichi Sakakibara²

松田 裕貴^{1,2} 松田 裕貴¹
Hiroki Matsuda^{1,2} Yuki Matsuda¹

¹ 奈良先端科学技術大学院大学
¹ Nara Institute of Science and Technology
² 讀賣テレビ放送株式会社
² Yomiuri Telecasting Corporation

Abstract: In contemporary society, a wide range of entertainment has become prevalent, leading to an increasing complexity in viewers' preferences and tendencies. This research proposes a method for clustering television viewers based on different viewing tendencies, utilizing non-specific television viewing data collected from five local television stations. The aim is to grasp individual viewing tendencies, enabling more effective and targeted approaches in areas such as program production, timing for commercial broadcasts, and business utilization. The goal is to enhance understanding and facilitate more tailored strategies in these aspects.

1 はじめに

近年、急速なテクノロジーの進展により、テレビの視聴環境は大きく変化している。特に、テレビをインターネットに接続し、データ放送コンテンツや動画配信サービスを活用する視聴者が増加している。インターネットに接続されたテレビからは、各視聴者がいつ、どの番組を視聴していたかといった情報を収集することが可能であり、テレビ放送局や、テレビ製造メーカーが視聴履歴データを収集している。これら各視聴者の視聴データは、従来の視聴率等のデータに加えて、番組制作やCM放映タイミングなど営業活動において、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、スポンサーや広告代理店からも活用が期待されている。

本研究で扱う「非特定テレビ視聴履歴データ」とは、視聴者の個人情報を含まず、視聴者の許諾なしにオプトアウト方式で収集されるデータを指す。このデータは、視聴者がテレビをインターネットに接続することで取得され、視聴者はデータ放送画面から提供を拒否

することが可能である。収集される情報には視聴開始・終了時刻やテレビID、視聴チャンネルが含まれ、これらは個人情報や視聴者属性を取得しない。これらの形式で収集されたデータは、一般社団法人放送セキュリティセンター（SARC）が策定した「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」[1]に基づき、放送局以外への第三者への提供の禁止、データの保存期間を定めており、一定の期間が経過したのちに削除される。

オプトアウト方式のデータ収集方法の利点は、多くのデータを能動的な視聴者の負担なく収集できる点が挙げられる。しかし、あらかじめアンケート等で視聴者の個人属性や視聴者属性を調査するオプトイン方式と違い、本手法ではこれらのデータが欠如しているため、視聴者の具体的な趣向や層の分析が難しいという課題がある。そこで本研究では、オプトアウト方式で収集された非特定テレビ視聴履歴データ（以下、非特定テレビ視聴履歴データ）を用いて、視聴者群全体の傾向を分析し、放送コンテンツの改善や新たなサービス開発に資するアプローチを検討する。これにより、個人情報を利用せずとも、有効な視聴者データの活用が可能とし、放送メディアのサービス向上に寄与すること

*連絡先：奈良先端科学技術大学院大学
〒630-0192 奈良県生駒市高山町 8916 番地-5
E-mail: yoshida.shogo.yo4@is.naist.jp

を目的とする。

そこで現在、在阪の5つの民間放送局では、非特定テレビ視聴履歴データの新たな価値創出に向けて、放送局間での非特定テレビ視聴履歴データの連携技術検証と連携データ活用に向けた共同技術実験 [2] を実施している。本研究では、これら実験で収集された非特定テレビ視聴データ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ーリングする手法を提案する。これにより、個別の視聴傾向を把握し、番組制作やCM 放映タイミング、営業活用などにおいて、より効果的でターゲットに合ったアプローチを可能にすることを本研究の目的とする。また、本研究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、2023 WORLD BASEBALL CLASSIC (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つの機関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。

2 関連研究

視聴履歴データの活用については様々な研究が実施されており、本章では東芝製テレビから収集したデータを用いて視聴傾向の分析等を行った研究と、筆者らの研究グループによって研究されている、在阪5局の非特定テレビ視聴履歴データを用いた研究について紹介する。

2.1 東芝製テレビを用いた大規模視聴履歴データの分析

菊池ら [3] は、東芝製のネットワーク対応テレビから収集された全国24万台分の大規模視聴データを用いて、リアルタイム視聴およびタイムシフト視聴を含むテレビ番組の視聴実態と長期間にわたる視聴行動の変化を分析した。菊池らの研究ではドラマ全話を通じた視聴実態など、番組ジャンル別にどのような視聴傾向を持っているかを明らかにし、一部ユーザのアンケート回答を基に機械学習手法を活用してテレビ保有世帯構成員の性年代属性を視聴傾向から推定する手法を提案した。

また、水岡ら [4] は、菊池らと同様に東芝製のテレビ視聴者に限定し、注目する視聴パターンに該当するかどうかを判定する手法、および視聴パターンの推移を

自動抽出する手法を提案し、テレビ視聴データから得られた視聴パターンの推移を自動抽出する手法の妥当性を明らかにした。

2.2 在阪5局非特定テレビ視聴履歴データの分析

放送局が収集している非特定テレビ視聴履歴データについては、筆者らの研究グループによって研究されている。松田ら [5] はテレビCM視聴がその後のインターネット検索行動に与える影響について、Google Trends から1時間ごとの検索クエリのトレンドを取得し、CMのメタ情報と組み合わせることでCMジャンルと検索寄与度の関係を調査している。また吉村ら [6] はCMの完視聴率にどのような地域差が存在するのか分析している。このように視聴履歴データの活用が進められているが、非特定テレビ視聴履歴データを用いて各視聴者の視聴傾向を分類し、その分析結果を番組制作などに活かしていないのが現状である。

3 提案手法

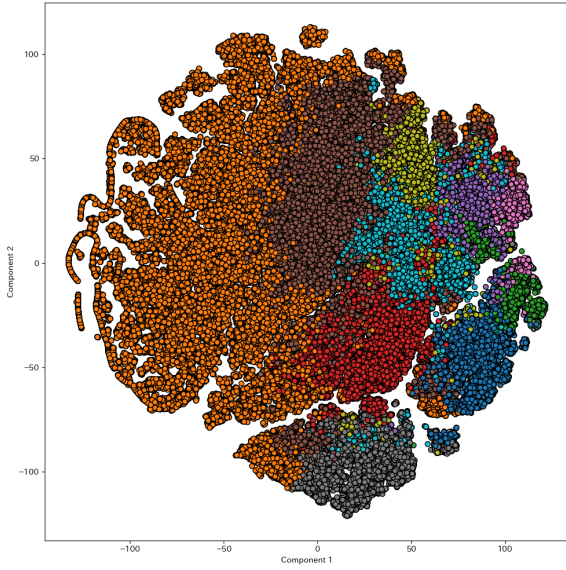
本研究では非特定テレビ視聴履歴データを用いて視聴傾向別に視聴者をクラスターリングする手法を提案する。本節ではまず、クラスターリングに使用する特徴量について述べ、次に視聴傾向の分析方法について説明する。また本稿において、クラスターリング結果に対するラベル付や考察などは読売テレビの専門家の意見を参考にしている。

3.1 特徴量抽出及びクラスターリング

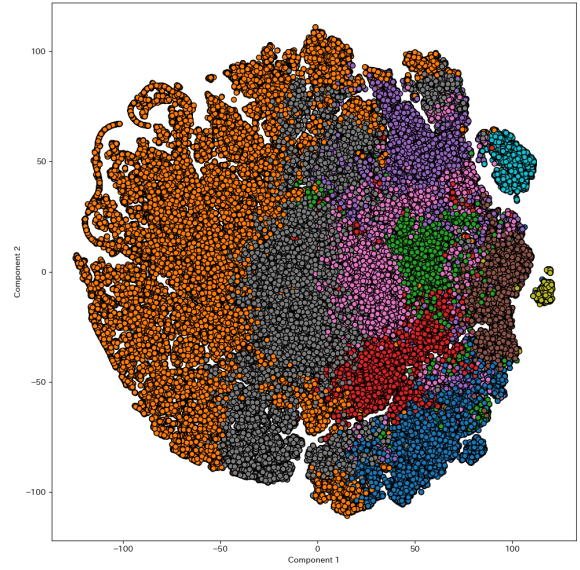
このデータには各視聴者を一意に判別するID、視聴開始時刻と終了時刻、視聴していたチャンネル、視聴した番組カテゴリが含まれている。本研究では分析期間中に放送されたテレビ番組のカテゴリ別の累計視聴時間を視聴者ごとに算出し、これらのデータの特徴量として使用している。分析期間中に含まれる具体的な番組カテゴリの一覧を表1に示す。また、本研究ではクラスターリング手法としてK-meansを使用し、先述した特徴量を用いて分類し、各クラスタごとの傾向を考察する。またクラスタ数は放送休止時刻と放送回数が少ない番組カテゴリを除いた10クラスとした。

3.2 ユーザ別視聴傾向分析

先述した特徴量を用いてK-meansによるクラスターリングを行い、各視聴者(視聴ID)がどのクラスタに属す



(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 1: t-SNE によるクラスタリング結果の可視化

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、各クラスター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。

4 実験

4.1 使用する非特定テレビ視聴履歴データについて

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 [2] 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。

4.2 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 1 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ー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

表 1: 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



図 2: 番組カテゴリの対応色

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。本研究では、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。

4.2.1 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3~7に示す。ヒストグラムの中の各色は図2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。

1. ヘビー層 (図 3) 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 4) 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ーの割合が少ない。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 5) 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ーの割合が少ない。図4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4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 6) 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。このクラスタでは視聴傾向が早朝や深夜にアニメを視聴しており、またバラエティー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若い層だと予想される。

5. ライト層 (図 7) 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、バラエティーが少ない。比較的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。

4.2.2 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8~12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

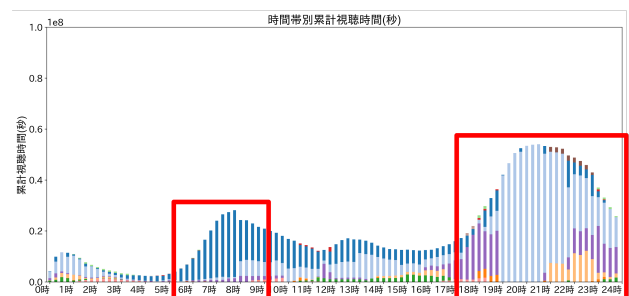


図 3: ヘビー層/朝・夜/偏り少

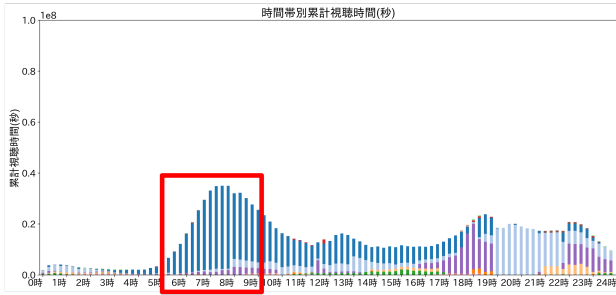


図 4: ミドル層/朝/バラエティ少

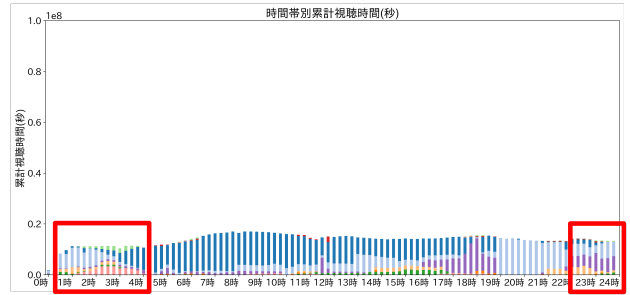


図 6: ライト層/深夜・早朝/アニメ・音楽多

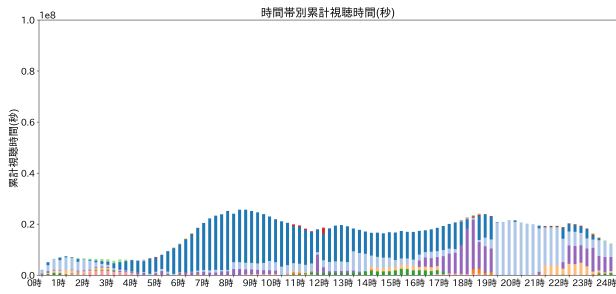


図 5: ミドル層/偏り少/バラエティ少

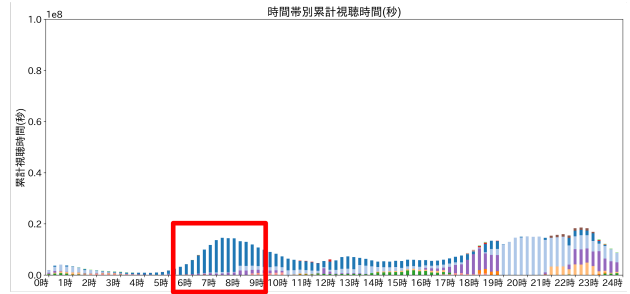


図 7: ライト層/朝/バラエティ少

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 8) 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. このクラスターでは他クラスター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 9) 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の 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 10) 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 11) 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. 他クラスター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 12) 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.

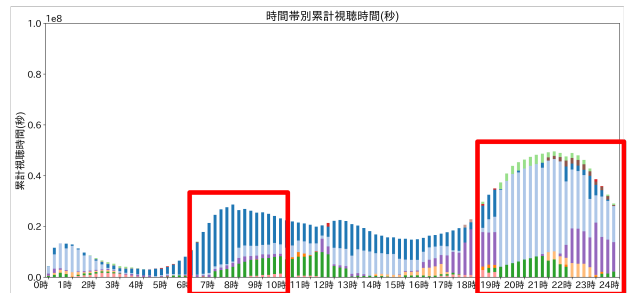


図 8: ヘビー層/朝夜/音楽・バラエティ多

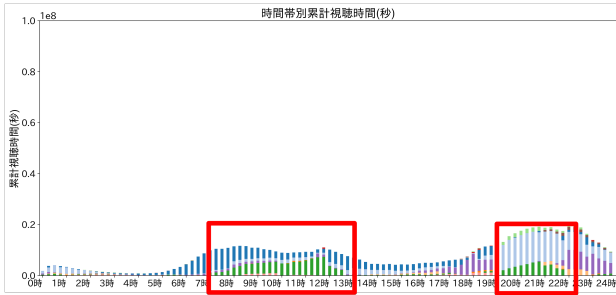


図 9: ミドル層/朝夜/スポーツ多

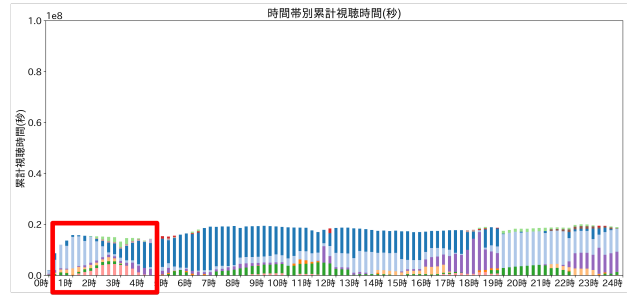


図 11: ライト層/全時間帯/アニメ多

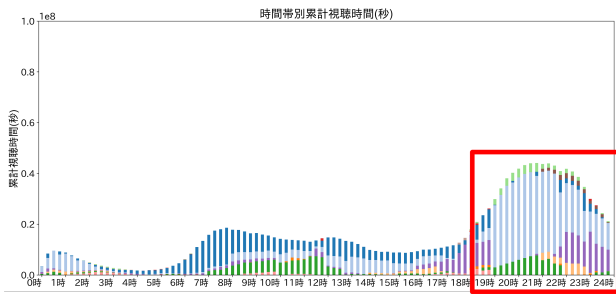


図 10: ミドル層/夜/音楽バラエティ多い

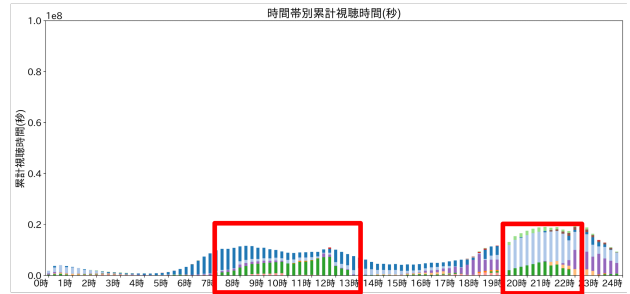


図 12: ライト層/朝夜/スポーツ多

5 おわりに

本研究では、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。

参考文献

- [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.
- [2] 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/experiment_s2022/index.html}, 2022.

- [3] 菊池匡晃, 坪井創吾, 中田康太ほか. 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.
- [4] 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 第32回(2018), pp. 1P203–1P203. 一般社団法人人工知能学会, 2018.
- [5] 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), pp. 1–6, 2022.
- [6] 吉村啓, 水本旭洋, 榊原太一, 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.