

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

松田 裕貴^{†,††} 榎原 太一^{††} 木俣 雄太[†] 鳥羽 望海[†] 真弓 大輝[†]

松田 裕貴[†] 安本 慶一[†]

[†] 奈良先端科学技術大学院大学 〒630-0101 奈良県生駒市高山町 8916-5

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

E-mail: [†]{hiroki.matsuda,taichi.sakakibara}@ytv.co.jp,

^{††}{kimata.yuta.kt7,toba.nozomi.th1,mayumi.daiki.mb9,yukimat,yasumoto}@is.naist.jp

あらまし 非特定視聴履歴データは、インターネットに接続されたテレビであれば収集可能で、個人を特定しない形式でIPアドレスや視聴時刻等を把握することができ、読売テレビは近畿一円のテレビから約350万台分のデータを収集している。また、民放各社でも収集が進んでいるが、非特定視聴履歴データ単体での新たな価値の創出には至っていない。本研究では、テレビ視聴者が実際にテレビCMを視聴し、その後の行動に結びついているか等を把握することを目的とする。具体的には、非特定視聴履歴データとGoogleトレンドに代表されるインターネット検索データ等を組み合わせることで、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CMが視聴者に与える影響の分析を行った。その結果、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。

キーワード テレビ, 視聴データ, ビッグデータ, 視聴者行動, 可視化, インターネット検索, IoT

1 はじめに

近年、テレビをインターネットに接続して、データ放送コンテンツや動画配信サービスを利用する視聴者が増えている。このようにインターネットに接続されたテレビであれば、視聴者がいつ、どの番組が視聴されていたのか把握することが技術的に可能となっていることから、放送局やテレビ製造メーカーがテレビの視聴履歴データを収集している。その中でも非特定視聴履歴データは、放送局が個人を特定しない形式で収集しているデータを指し、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利活用が期待されている。

しかし、非特定視聴履歴データは個人情報ではないものの個人の趣味趣向を把握しうるデータとなるため、プライバシーへの配慮が求められており、一般社団法人放送セキュリティセンターは、「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]を2019年5月に公開し、2度のバージョンアップを行い、個人情報を保有している第三者への提供の禁止や放送局が保有している個人情報と突合しないようにルール整備が行われた。その結果、民放各社での非特定視聴履歴データの収集が進んでいる。読売テレビにおいても2019年12月から非特定視聴履歴データの収集を開始している。前述の通り、非特定視聴履歴データは、インターネットに接続されたテレビであれば収集可能で、個人を特定しない形式でいつ、

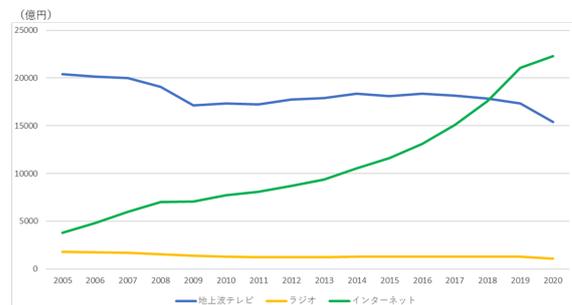


図1 日本の広告費

どの番組が視聴されていたのかを把握することができる。しかし、本データ単体を分析してもテレビ視聴状況把握しか出来ず、新たな価値を生み出すには至っていない。

また、電通の調査[2]によると図1のように2019年に地上波テレビ広告費がインターネット広告費に追い抜かれたとの結果も出ている。その要因の1つにテレビ広告はインターネット広告のように、広告効果の可視化・分析が進んでいないからだと指摘もある。また、インターネット広告では、ユーザが過去に閲覧したWebページなどの情報をもとに広告を最適化するターゲティング広告など、個人に合わせた広告戦略を取ることが可能であるが、一般的なテレビ広告(CM)は視聴者個人に合わせた広告の出し分けが出来ないという制約がある。

テレビCMの広告効果を向上させる方策を探るためには、テレビCMがテレビ視聴者の後の行動に及ぼす影響を明らかにす

ることが必要であると考えた。そこで本研究では、テレビ視聴者の CM 視聴後における行動としてインターネット行動に着目し、CM と当該行動の関係性を把握することを目的としている。

本論文では、前述の非特定視聴履歴データ、テレビ 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つのデータを分析することで、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。具体的には、CM 放送実績データとインターネット検索データの相関性を分析することで、テレビ CM の放送が視聴者のインターネット検索行動に繋がっていることを示した。また、非特定視聴履歴データと CM 放送実績・インターネット検索データの相関性を分析することで、視聴台数の多さがインターネット検索トレンドの増加に与える影響は少ないこと、そして視聴台数が少なくとも、視聴者を効率的にインターネット検索行動へ繋げることができる時間帯を明らかにした。

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。

2.1 テレビ視聴ログの分析

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 [3] は、東芝製ネットワーク対応テレビから得られた大規模視聴データ全国 24 万台分の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 [4] 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。しかし、視聴者がテレビ視聴後にテレビ以外でどのような行動変容に繋がったのかは研究がなされていない。

2.2 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 [5] が研究を行っ

てきている。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。

2.3 CM データの分析

テレビ CM との接触回数別の効果検証について、加藤ら [6] が研究を行ってきている。加藤らは、日本の自動車業界を対象に、テレビ視聴データとデジタルデバイスからの Web 閲覧ログデータを用いて、メーカー WEB サイトへの誘因に効果的なテレビ CM の接触回数を評価した。その結果、Three-hit theory のとおり3回の接触が最も効果的であり、ターゲット顧客を Web サイトへ誘導するには、3-10回の接触で十分であり、それ以上接触しても態度変容を起こすことは困難であることを明らかにしている。

2.4 本研究の位置付け

先行研究では、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。また、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。これに対し、本研究では、テレビ視聴が視聴者の興味・関心を喚起し、実行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 データセット

本論文では、非特定視聴履歴データ、テレビ 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。

3.1 テレビの非特定視聴履歴データ

本研究では、読売テレビが独自で収集している非特定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視聴履歴

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、テレビメカ 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。

3.2 テレビ CM 放送実績データ

本研究で使用するテレビ 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ー名、スポンサー業種、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。

3.3 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¹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「スポンサー名」「スポンサー名+CM」、また商品名がある場合は「商品名」を指定している。データ収集期間は、テレビ CM 放送開始日の 1 週間前からテレビ CM 放送終了 1 週間後である。

4 分析

前章で述べた 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。

4.1 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

近畿で放送されたテレビ 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分析した。

4.1.1 分析方法

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 2 に示すように、テレビ 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 2 CM 有無別の 24 時間以内の検索量取得方法

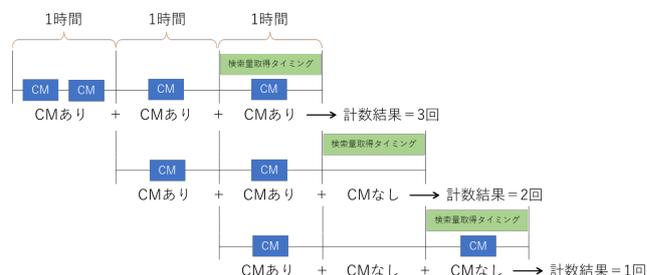


図 3 連続的な 3 時間において CM 放送回数が計数される例

分析 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 3 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。

分析 C：CM 放送期間・期間外の影響分析

最後に、テレビ 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、比較対象としてテレビ 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。

4.1.2 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 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 4 は、24 時間において、とある清涼飲料水のテレビ 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

1 : <https://trends.google.co.jp/>

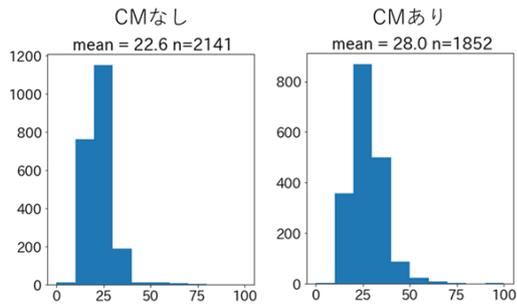


図4 24時間以内のCM有無別の検索量(清涼飲料水)

表1 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

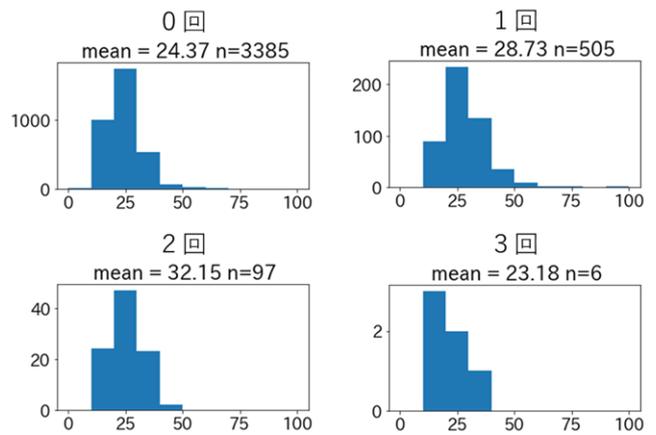


図6 3時間以内のCM回数別の検索量(清涼飲料水)

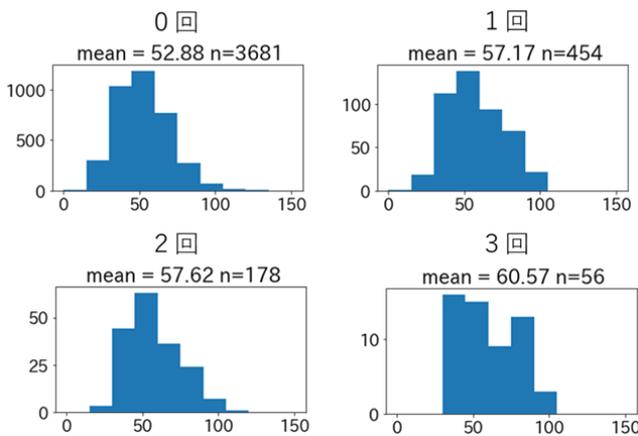


図5 3時間以内のCM回数別の検索量(アルコール飲料)

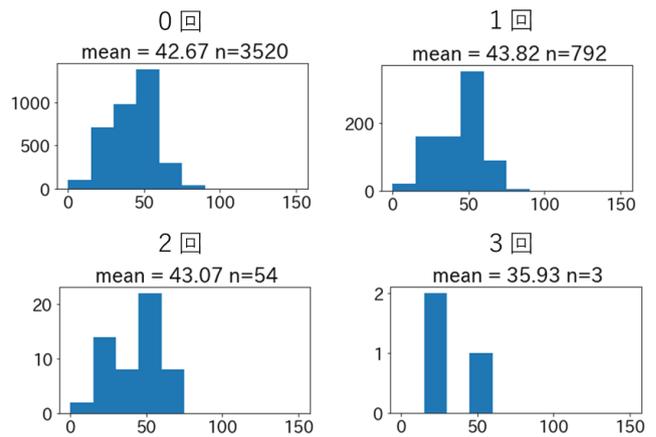


図7 3時間以内のCM回数別の検索量(ネット保険)

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4の結果から、テレビ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。

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放送による検索量の変化について結果を示す。図5, 6, 7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放送回数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1に示す。図6, 7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5)、清涼飲料水(図6)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7)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。

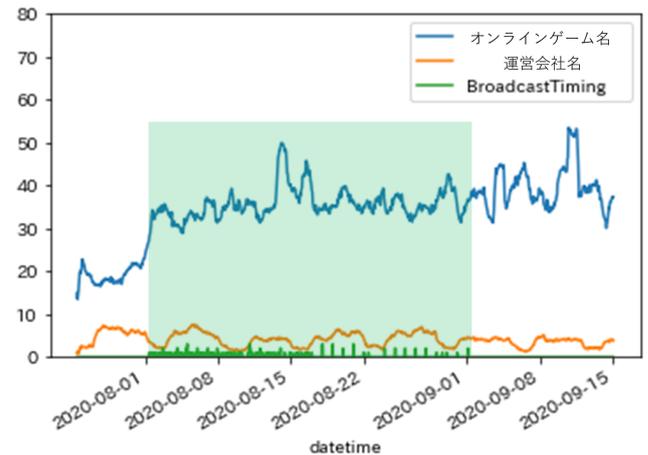


図8 CM放送前後の検索量推移

分析C: CM放送期間・期間外の影響分析

図8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。テレビ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、サービス名ではテレビ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。

表2 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

4.1.3 考 察

図4で示した通り、テレビ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5, 6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7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8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テレビ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、テレビ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。テレビ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、これはテレビ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、テレビ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。

4.2 非特定視聴履歴データと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を分析した。

4.2.1 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2に各計測期間を例示する。

最後にテレビ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。

4.2.2 分析結果

図9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

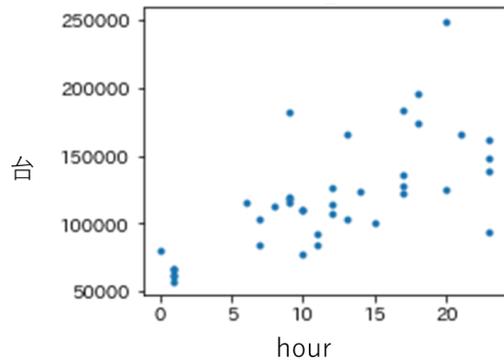


図9 CM放送時間帯(1時間単位)と視聴台数の分布

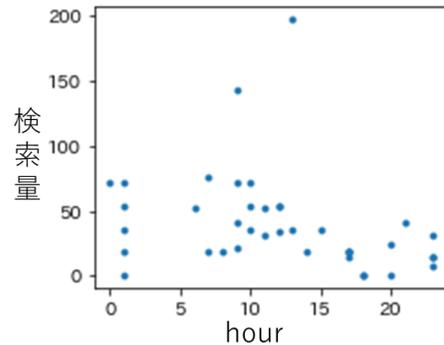


図10 CM放送時間帯(1時間単位)と検索量の分布

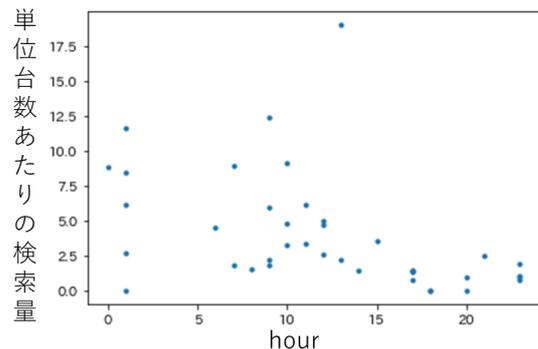


図11 視聴台数10,000台あたりの検索量とCM放送時間帯の分布

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10は、先ほどと同様のメーカー商品の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9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。

最後に図11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。

4.2.3 考 察

図9で示した通り、プライムタイムに放送したテレビ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10, 11で示した通り、テレビ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。

5 おわりに

本研究では、テレビCMの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、テレビ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。本研究は、これまでは困難であったテレビCMの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やすいCM種別の分析や、テレビ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。

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] 株式会社電通. 2020年日本の広告費. https://www.dentsu.co.jp/knowledge/ad_cost/2020/, 2020.
- [3] 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.
- [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, Vol. JSAI2018, pp. 1P203-1P203, 2018.
- [5] 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.
- [6] 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.