Analysis of Relationship Between Non-Identifiable TV Viewing History Data and Web Search Trends

Hiroki Matsuda[†,‡], Taichi Sakakibara[†], Daiki Mayumi[‡], Yuki Matsuda[‡], and Keiichi Yasumoto[‡]
[†]*Yomiuri Telecasting Corporation*, Osaka, Japan, {hiroki.matsuda, taichi.sakakibara}@ytv.co.jp
[‡]*Nara Institute of Science and Technology*, Nara, Japan, {mayumi.daiki.mb9, yukimat, yasumoto}@is.naist.jp

***Abstract*—Non-identifiable viewing history data can be collected from internet-connected TVs, and it is possible to grasp information such as IP addresses and viewing times without identifying individuals. Yomiuri Telecasting Corporation has collected data from approximately 3.5 million TV units, but it has been considered challenging to find new value from these data alone. Therefore, in this study, we combined non-identifiable viewing history data with internet search data, such as Google Trends, to analyze the impact of TV commercials on viewers' internet search behavior. As a result, we found that as the number of commercial broadcasts increased, search behavior also increased, and we were able to identify the broadcast times of commercials that efficiently led to search behavior.**

***Index Terms*—TV, Television, TV viewing data, Big data, Viewer behavior, Visualization, Internet search, IoT**

I. Introduction

Non-identified viewing history data refers to the data collected by broadcasters in a format that does not identify individuals, and it is expected to be utilized as big data to enhance advertising value. However, although this data is not personal information, it can capture individuals' preferences and interests, making privacy considerations crucial [1]. In comparison to internet advertising, TV advertising is said to lag behind in terms of effect visualization and analysis, as well as the ability to conduct targeted advertising.

In related studies, analysis of viewing data using TVs manufactured by Toshiba was conducted to determine "what types of viewers select which programs" [2], and clustering based on viewing patterns were performed [3]. Additionally, studies have been conducted on the relationship between the frequency of exposure to TV commercials and their effectiveness in driving website visits [4].

In this study, we focused on the post-viewing internet behavior of TV commercial (CM) viewers and analyzed three types of data: non-identified viewing history data, CM broadcasting records, and internet search data. The results demonstrated that the broadcasting of TV commercials influences viewers' internet search behavior. Furthermore, it was revealed that the impact of viewership numbers on internet search trends is relatively minimal, and specific periods that efficiently connect viewers to internet search activities were identified.

II. Datasets

Here, we describe three different datasets for data analysis in this paper. The relationship between those three datasets (A), (B) and (C) are illustrated in Fig. 1.

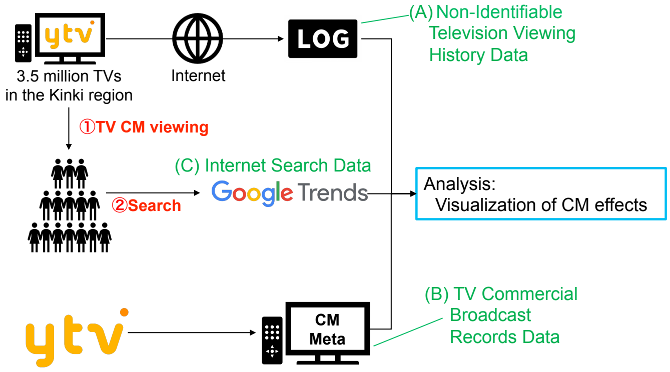

Fig. 1. Analysis Summary Chart

A. Non-Identifiable Television Viewing History Data

In this study, we utilize non-identifiable television viewing history data collected by Yomiuri Telecasting Corporation (hereafter YTV). This dataset consists of approximately 3.5 million television sets connected to the internet in the Kansai region of Japan, excluding those that have opted out. It includes information such as IP addresses, TV device IDs, postal codes, viewing start and end times, and TV manufacturer IDs. The data spans from April 2020 to March 2021.

B. TV Commercial (CM) Broadcast Records Data

For our analysis, we employ actual CM broadcast records provided by YTV. These records are outputted in CSV format from YTV's sales broadcasting system. The dataset contains information such as sponsor names, sponsor industries, CM names, broadcast start times, and CM durations. The data covers the period from April 2020 to March 2021.

C. Internet Search Data

One of the main user behaviors that we focus on in this study is internet search activities. We utilize Google Trends, a service provided by Google, which allows us to analyze the search volume for specified regions and search terms within a specific time period. The search volume represents the relative amount of searching on the Google platform. In this paper, we set Japan as the search region and specify "{sponsor name}," "{sponsor name} + CM," and, if applicable, "{product name}" as the search terms. The data collection period spans one week before the start of the CM broadcast to one week after the broadcast's end.

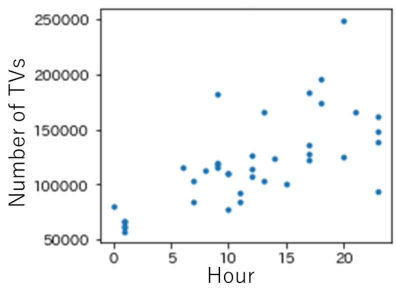

Fig. 2. Distribution of CM broadcast time (per hour) and number of TV viewers

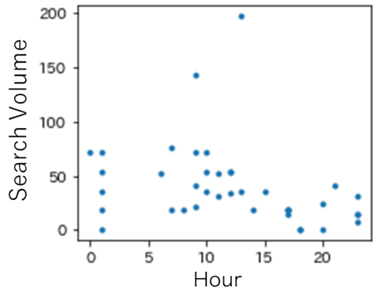

Fig. 3. Distribution of CM broadcast time (per hour) and search volume

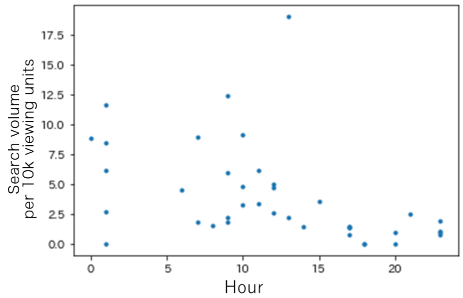

Fig. 4. Distribution of search volume per 10,000 viewing units and CM broadcast time period

TABLE I
EXAMPLE OF COMMERCIAL BROADCAST TIME AND SEARCH VOLUME MEASUREMENT PERIOD

No	CM Broadcast Date	Number of TV viewers	Measurement period
1	Jan 1st, 12:05:30	100,000	Jan 1st, 12:00 – 12:59
2	Jan 1st, 20:59:04	180,000	Jan 1st, 20:00 – 20:59
3	Jan 2nd, 01:03:57	60,000	Jan 2nd, 01:00 – 01:59

III. ANALYSIS

We analyze datasets: Non-Identified Viewing History Data, CM Broadcasting Records Data, and Internet Search Data, to investigate the effects of CM on internet search activities. The overall picture of the analysis is shown in Fig. 1.

A. Methodology

Firstly, we investigated the distribution of CM viewing numbers in relation to the CM broadcasting time and the number of TVs tuned in, focusing on hourly CM broadcasting time and the corresponding viewing numbers in the Kansai region. Next, to understand how internet search volumes vary based on CM viewing numbers, we investigated the distribution of search volumes in the vicinity of each CM, taking into account the viewing numbers for each CM. The viewing numbers were measured during the CM broadcasting time, while the search volumes were measured at the rounded-down minute of the CM's broadcasting date and time. Table I provides an example of the measurement periods. Lastly, to identify the broadcasting time slots of TV commercials that efficiently lead to internet searches, we categorized each CM's broadcasting time into one-hour intervals and examined the distribution of search volumes per 10,000 viewing numbers. Similar to before, the viewing numbers were measured during the CM broadcasting time, while the search volumes were measured at the rounded-down minute of the CM's broadcasting date and time.

B. Results

Fig. 2 shows the distribution of CM broadcasting times and corresponding viewing numbers for a food manufacturer's CM. It confirms that viewing numbers are lower during late-night hours and increase from morning to evening. It is evident that *prime time* (The period that highly anticipated TV programs will be broadcasted. In Japan, it is between 19:00 and 22:00.) is the most efficient period to reach a large number of viewers when seeking broad recognition for CM broadcasts. Fig. 3 shows the distribution of CM broadcasting times and the volume of internet searches for the same manufacturer's products. Unlike in Fig. 2, it can be observed that *prime time* does not necessarily result in higher search volumes. Specifically, even during late-night hours, which are considered to have lower CM value, the search volumes are comparable to or higher than those during prime time. Finally, Fig. 4 shows the distribution of search volumes per 10,000 viewing numbers by categorizing each CM's broadcasting time into one-hour intervals. Similar to before, it confirms that late-night hours and morning/daytime periods with lower viewing numbers are more efficient in leading to search actions compared to prime time.

IV. DISCUSSION & CONCLUSION

As shown in Fig. 2, it is evident that CMs broadcast during *prime time* receive higher viewing numbers, indicating a larger reach. However, as demonstrated in Figures 3 and 4, the time slots following TV CM viewing that lead to search actions are not necessarily during *prime time* but rather during late-night hours and morning/daytime periods. This can be attributed to the fact that late-night hours and morning/daytime are more conducive to "second-screen viewing" when watching TV. These findings indicate that even during time slots considered to have relatively lower advertising value, such as late-night hours and morning/daytime, viewers' internet search behavior is more efficiently connected to CMs than during *prime time*. In future work, we will incorporate multiple telecasting corporations' data for more comprehensive analysis.

REFERENCES

[1] Secure Broadcasting Authorization and Research Center, "Practices for Handling Non-Identifiable Television Viewing History Data Obtained by the Opt-Out Method (ver. 2.2)," https://www.sarc.or.jp/documents/www/NEWS/hogo/2022/optout_practice_ver2.2.pdf.pdf, 2023. (In Japanese)

[2] M. Kikuchi, and T. Tsuboi, K. Nakata, "Program Viewing Analysis Using Large-Scale Television Viewing Data," IPSJ Transactions on digital practices, vol.7, No.4, pp. 352–360, 2016. (In Japanese)

[3] Y. Mizuno, and K. Nakata, R. Orihara, "Analysis of Viewing Pattern Transitions using Large-Scale Television Viewing Data," JSAI2018, pp. 1P203–1P203, 2018. (In Japanese)

[4] T. Kato, and K. Tsuda, "Validation of Three-Hit Theory on Television Advertisement based on Life Log," SIG-KSN-023-03, Vol. KSN-023, pp. P03–P07, 2018.