Audience Behavior Mining by Integrating TV Ratings with Multimedia Contents

Ryota Hinami and Shin'ichi Satoh

The University of Tokyo, National Institute of Informatics

hinami@nii.ac.jp, satoh@nii.ac.jp

Abstract—TV ratings are a widely used indicator in the TV broadcasting field. While TV ratings are mainly used in advertising, they can also be used as a social sensor that reflects the interests of people. This paper presents a framework for discovering audience behavior through the mining of TV ratings. We have established a framework that enables discovery of numerous patterns of audience behavior from TV ratings. Used along with other multimedia contents such as video and text, it enables various types of knowledge to be semi-automatically found, such as what types of news programs are of most interest and what are the key visual features for acquiring high TV ratings.

The discovery of audience behavior is achieved by focusing on the change points in the rating data, i.e., the points in time where many people switch the channel or turn the television on or off. Rich descriptions that characterize these points are extracted from multimedia contents, and then various filtering techniques are used to extract specific patterns of interest. Several applications of this framework for discovering knowledge demonstrated that it can effectively extract various types of audience behavior. To the best of our knowledge, this work is the first work to analyze the use of ratings data in combination with video and other multimedia data.

I. INTRODUCTION

Television audience ratings (TV ratings), which are used to assess the popularity of TV programs, is a key indicator in the field of TV broadcasting. The TV rating of a program indicates the percentage of all television households tuned in to that program. TV ratings are mainly used by sponsors and broadcasters to measure the reach of the sponsors' advertising. For advertising, Since how many targeted people are watching the advertisement (i.e., commercial films (CFs)) is important to sponsors, the ratings are an important indicator. Broadcasters focus on increasing the ratings of their programs in order to acquire more sponsors.

TV ratings can also be used as sensor to gauge the interests of people. If the contents of a TV program interest people, they tend to tune in to the program, and the ratings consequently increase. Conversely, if people are not interested in the contents, they switch to other channels, and the ratings decrease. Therefore, TV ratings are an indicator of popular topics and social trends, e.g., what types of news are of interest and which performers are currently popular. Broadcasters who discover such knowledge by mining TV ratings data can create TV programs that are needed by many people. Although social media sites such as Twitter and Facebook are also useful media for capturing the interests of people, only a small portion of the users of such media are active users, so the interests of most people i.e., *the silent majority*, are ignored.


Fig. 1. Audience behaviors observed in TV ratings data at the flooding of the Kinugawa River.

TV ratings also include important information for risk management. In an emergency such as a natural disaster, the government needs to swiftly deliver correct information to people in a timely manner. Since television is a key medium for conveying information to many people in real time, how to properly deliver important information by television should be considered. By analyzing TV ratings to determine how many people get information from TV, we can judge whether critical information has been correctly delivered and take measures to improve the spreading of such information.

Although TV ratings have been investigated for many years, most previous work [3] focused on forecasting the ratings of particular TV programs. The motivation was to estimate the cost of TV advertising because the cost of advertising is directly linked to the TV ratings. Few works targeted the mining of ratings data even though such data contains valuable information. Moreover, integration of TV ratings with multimedia contents such as video data for TV programs has not been investigated.

The integration of TV ratings with multimedia contents (e.g., video, speech) facilitates discovering the relationships between audience behaviors and TV program contents. Figure 1 shows the ratings data for two TV station that broadcast live the flooding of the Kinugawa River on Sep. 10, 2015. Thumbnail images of both broadcasts around the time of the flooding are also shown. Both stations were broadcasting live up until the time indicated by the red dotted line. At that point, one station (represented by the green line) started showing CFs. This led to a big transition in viewers from the "green"

station to the "blue" station. Similar behavior from the "blue" station to the "green" station is also observed at gray dotted line. This means that the viewers were greatly interested in the event and thus switched to another station to keep watching the news. This is an example of audience behavior obtained from TV ratings data for only one event. Analysis of larger amounts of such data will enable typical patterns of audience behavior to be found automatically, and more hidden behaviors can be found by integrating TV ratings and multimedia contents.

The motivation for learning about audience behavior has several aspects. First of all, audience behavior indicates what people find of interest, and this knowledge is important for creating TV programs that attract viewers. Understanding user behavior is also important from the aspect of advertising. Obtaining the patterns that increase ratings from the mining of audience behavior will help broadcasters obtain higher ratings for their programs and thereby get more sponsors. Risk management is another aspect. User behavior mining reveals how people get information following a disaster. This information could be used to determine what should be done to better convey important information to many people.

In this paper, we analyze the events or pattern in audience ratings data by using multimedia contents such as video data and transcript data in combination with ratings data. Our goal is to establish a framework that can be used to discover the numerous patterns of user behavior from TV ratings. Although user behavior is being explored by many works [2], [1], most of recent work focus on social networks and little work has addressed the behavior of TV audiences. To the best of our knowledge, this is the first work on the use of audience ratings combined with multimedia data.

To discover the relationships between the ratings data and multimedia contents, we focus on the change points, i.e., the points in time when many people newly tune in to a particular TV program. These points are considered to have valuable information, particularly the interests of the viewers. We describe these points using visual features extracted from video and keywords extracted from transcripts. Since the number of such points is huge, we accurately and flexibly apply filtering and aggregation in accordance with the analysis target. Experiments demonstrated that this framework can discover various types of valuable information. A specific application demonstrates a system that can be used to detect and analyze the types of news in which people are interested.

Contributions. Our contributions are as follows: 1) we initiated a study of mining that targets TV ratings, demonstrating that valuable knowledge can be obtained from TV ratings and that TV ratings are an important source as a target of mining. This work is the first to integrate ratings data with various multimedia contents including video and text transcripts as well as the first of investigate the mining of TV ratings. 2) We established a framework for analyzing audience behavior by analyzing change points. It uses various features and filtering techniques introduced from other fields such as a computational aesthetic measure. This framework can discover various types of valuable knowledge, as shown in the application presented in Sec. VI.

II. TV Ratings in Japan

The TV ratings most commonly used in Japan are the ones provided by Video Research Ltd., which started audience measurement in 1962. It has been the only one major firm providing audience measurement since the Nielsen Corporation left Japan in 2000. Its ratings are now the standard source of TV audience measurement information in Japan. We use their ratings in our work.

The ratings of Video Research are based on samples from households randomly selected from a certain area. The ratings are surveyed for each of the 27 broadcast areas. There are 600 households in each of the three main areas of Japan (Kanto, Kansai, and Nagoya) and 200 in each of the other 24 areas, for a total of 6,600 households. The sampling method is based on systematic random sampling: first selecting a random starting point and then picking up every nth households from the list of all households in a certain area. The interval n is the total number of households divided by the number of households to be sampled.

In this work, we used the TV ratings for the Kanto area, i.e., the Greater Tokyo Area, which has a population of approximately 40.6 million constituting 18.2 million households with a television. In the Kanto area, the ratings are calculated on the basis of data collected from 600 sampled households for which the data are gathered using a people meter. The household rating for a particular TV program is defined as the percentage of households that watched that program, which is calculated from the sampled viewing records. The household ratings contain the ratings minute by minute, which we call the per-minute ratings. We used the per-minute ratings for seven TV stations during 2015.

III. Typical Audience Behavior

Before explaining the details of our proposed method, we discuss audience behavior, i.e., how people typically watch television and switch channels. Although some people turn on the television or switch channels simply because they feel like it without any specific reason, most people tend to follow typical viewing patterns. Our analysis of TV rating data revealed that people tend to turn on the television or switch channels in the following cases:

(a). **Boundary of TV show:** Turn on television or switch the channel when the target TV program starts and turn it off when it ends.
(b). **Extrinsic factor:** Turn on television due to non-television factor such as an earthquake.
(c). **Transition:** Switch to another program due to loss of interest caused by the advertisement, topic change, etc.

These cases are clearly evident from TV ratings data. Figure 2 shows examples of these three cases extracted from actual TV ratings data. Case (a) is the simplest: tune in to a particular TV program when it starts. Since many people tend to watch certain programs out of habit, they normally tune in to those programs when they begin without switching through other channels. In this case, similar patterns are observed every week because many people watch the same program every week.
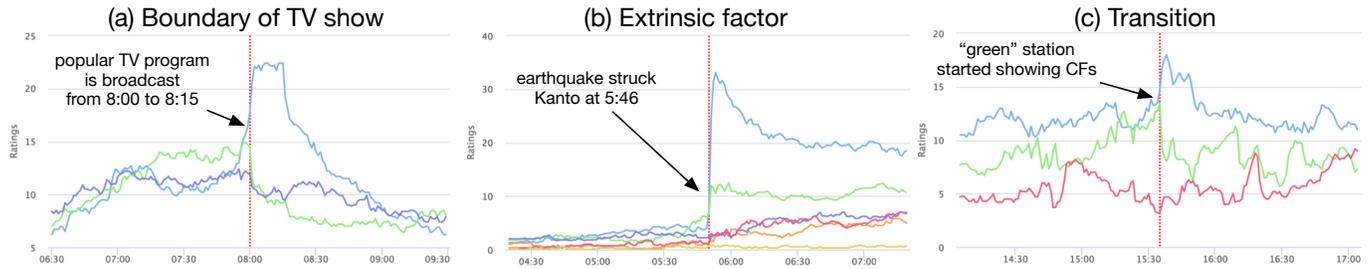
Fig. 2. Examples of three cases observed in audience ratings. (a) Boundary of TV show: the rating increased when the programs began. (b) Extrinsic factor: the rating suddenly increased when an earthquake struck. (c) Transition: the ratings of two programs suddenly increased or decreased due to viewers switching from one program to the other.

Case (b) is when people turn on the television due to an external stimulus. For example, when an earthquake strikes hard enough for people to feel the shake, many of them turn on the television to get related information. Another example is that of buzz on a social networking site such as Twitter. In this instance, the ratings increase is not directly related to the program contents.

Case (c), which is our main analysis target, is when viewers switch to another program due to a loss of interest in the program being watched. Switching channels when CFs start to be shown is a typical pattern of user behavior. A certain number of people usually change the channel when a CF starts because most people are not interested in CFs. They typically switching through the other channels and select the one with the most interesting program. After a minutes or so, some of them return to the original channel, anticipating that the CFs have ended, while the others continue watching the new channel. A change in the topic of a news program or a switch in the activity on a variety program can also trigger such behavior. For example, some people are interested in certain news topics and continue to switch channels until they find the topic of interest.

In short, viewers select channels in accordance with certain factors, and typical audience behavior can be observed in the ratings data. By analyzing such data, we can learn about audience behaviors, e.g., the things people find interesting. This is the target of this work. We propose a framework for learning about audience behavior through the integration of TV ratings data and multimedia contents.

IV. FRAMEWORK FOR AUDIENCE BEHAVIOR MINING

In this paper, we target the mining of audience behavior from TV ratings data. Our objective is to develop a framework for automatically finding particular patterns or events indicating the interests of people. To this end, we focus on the change points in rating data. In particular, we focus on the micro-level change points, where the per-minute ratings change significantly within a few minutes, which means that the number of viewers increases or decreases suddenly. It is assumed that something happens at these points, so the information they provide is more valuable than that provided by other points. We detect these points and analyze them in combination with video contents and other meta-data such as captions. By analyzing the contents corresponding to the

change points, we can better understand the things that most interest viewers.

By mining a large number of change points, we can discover the events and patterns revealed by the ratings data. The flow of the proposed framework is diagrammed in Figure 3.

1) **Detection of change points:** Detect large number of micro-level change points from ratings data.
2) **Description of change points:** Extract information for each change point from multimedia data (e.g., extract visual features from video data).
3) **Filtering and aggregation:** Filter points to reject noise or extract target, and aggregate filtered points to extract knowledge from them.

In the first step, we find as many change points as possible using a simple rule, relying on later stages to filtering them. Our main focus is on steps 2 and 3, i.e., description, filtering, and aggregation of change points using multimedia contents.

For each detected point, various features are extracted from one-minute multimedia data at the time of each point to provide rich descriptions that characterize the points. These characterization are used to filter and aggregate the points. Filtering is used to reject noise or extract the points that we want to analyze. Finally, the filtered points are aggregated and visualized. For example, visualization of the change point statistics facilitates understanding of ratings increase pattern. User can interactively add or change the filters to visualize more detailed cases, which accelerates the pattern discovery.

In addition to TV ratings data, we use several types of multimedia data to describe and filter the change points.

Video: Broadcast video corresponding to the TV ratings.
Captions (text transcripts): Captions for a program facilitate determination of the topic, which can be difficult when simply watching the program video.
Electric program guide (EPG): Information on TV programs including title, category, and description.

By using these data, we can describe the contents at each change point with rich information. Filtering and aggregation of these data enable precise analysis of the relationship between the ratings and the multimedia contents. In Section V, we explain the functions of the proposed framework, i.e., detection, description, and filtering of change points. In Section VI, we present two example applications of the proposed framework.
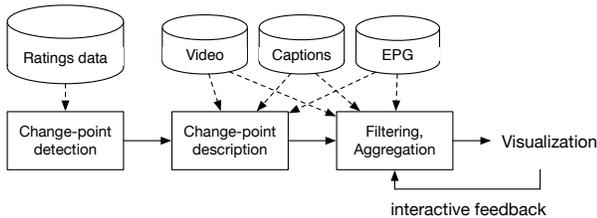
Fig. 3. Flow of proposed framework.

TABLE I
VISUAL FEATURES FOR EACH FRAME. THE FEATURES OF ONE-MINUTE
VIDEO ARE COMPUTED AS MEAN OVER 60 FRAMES (ONE FRAME PER
SECOND). n INDICATES NUMBER OF DIMENSIONS FOR EACH FEATURE.

Feature	n
Hue, Saturation, and Brightness (HSV)[9]	3
Pleasure, Arousal, and Dominance[9]	3
Color Names[9]	11
GLCM-feature (entropy, dissimilarity, energy, homogeneity, and contrast)[9]	5
Object (top-level categories of ImageNet)	9
Object (re-generated 100 object clusters)	100
Emotion	8

V. FUNCTIONS

A. Detection of Change Points

We use a simple approach to detect change points. The rate of increase (or decrease) for each minute is calculated as the difference between the previous and current per-minute ratings, and the rates for the increase points are compared to a pre-determined threshold. Those points for which the rate is above the threshold are taken as change points. Each change point corresponds to a time when a certain number of households began watching a particular TV program, either by turning on the television or switching from other programs. For example, a rating increases of 1%, in the Kanto area, where there are 18.2 million households with a television, means that over 182,000 households tuned in to the program.

While change-point detection has been well studied in such fields as data mining, and various algorithms have been proposed, we adopted the simple approach described above because we want to extract as many change points as possible and our main focus is the later stages, i.e., filtering and aggregating these points by using multimedia contents. Moreover, the threshold should be set in accordance with the application. Using the rate of increase as the threshold enables users, such as broadcasters and advertisers, to intuitively determine and adjust the threshold.

B. Description of Change Points

1) Visual Features: Several visual features are used to characterize images as summarized in Table I. Color and texture are used as low-level visual features. Object and emotion features are used as mid-level features. These image features calculated for each of the 60 frames (one frame per second) in a one-minute video. They are then aggregated by taking the mean for each one over the 60 frames, and the aggregated values are used as the features of each change point.

Fig. 4. Examples of clusters created for TV contents.

Low-level Features. We use low-level features introduced from computational aesthetics, which has been shown to be effective in evaluating visual concepts. Following , we implement the color and texture features used in [9]: Color names, GLCM features, HSV statistics, and the Pleasure, Arousal and Dominance values computed from HSV values.

Mid-level Features. In addition to using low-level features, we use higher level features that are more related to the image content. We use the object category classification score and the emotion score in particular.

We first use features on the basis of ImageNet classification [4]. We compute the score for 1000 object categories on the basis of the ILSVRC classification task. We use the AlexNet [7] trained with the ILSVRC 2012 dataset [4] to compute the classification scores (i.e., the probabilities for each category) and use them as features. Since the nature of TV contents differs from the categories in ImageNet, it is difficult to use the scores of original 1000 categories directly. Instead, we use two features computed based on the 1000-class scores. Firstly, we use the scores for the 9 top-level categories of ImageNet (plant, geological, natural, sport, artifact, fungus, person, animal, misc) calculated by aggregating the classification scores for the 1000 object categories.

Secondly, we define object categories suitable for TV contents by clustering the features of convolutional neural networks. We regard the 4096-dimensional fc7 features of AlexNet as semantic features and perform k-means clustering on them (k=100) so that each cluster represents certain objects or scenes that frequently appear in TV contents. Each frame is assigned to the cluster with nearest centroids and 100 dimensional binary feature are obtained for each frame, where each dimension indicates whether the frame belongs to each cluster. Three examples of created clusters (representing sumo, other sports, and weather reports) are shown in Figure 4.

In addition to the features based on the object category, emotional feature is also used. we calculate the scores for eight emotions (amusement, anger, awe, contentment, disgust, excitement, fear, sad) using the method of Machajdik and Hanbury [9]. They showed that emotions evoked by images can be classified using a computational approach. By using the emotion classification scores, we can analyze the effects of the emotions caused by video contents on TV ratings.

2) Keywords: We extract the keywords from the caption for the minute corresponding to each change point. To extract keywords from Japanese text, we first perform morphological analysis to decompose the text into several morphemes because Japanese text has no space between words. We do this using the MeCab morphological analyzer [8], which is the widely used for morphological analysis of Japanese. It outputs the separated text along with part-of-speech information. Only nouns are extracted as candidate keywords.

Next we exclude the stopwords that often appear in speech. In TV contents, there are frequently appeared words for each TV programs such as the title of TV program that are not important to characterize the contents. To exclude these words, we create a model for each of the 24 hours in a broadcast period for each of the 7 major TV stations in the Kanto area and use them to calculate the word appearance rates. Those words that appear over 30% days are excluded. The words (nouns) remaining are used as keywords.

3) Other Information: In addition to visual features and keywords, we use such basic information as the broadcast time and metadata of TV programs obtained from the EPG. Since this information is obtained without special processing, it is explained as required.

C. Filtering Techniques

1) TV Program Boundaries: The boundary of a TV program is one case in which many people switch the channel, and TV ratings change greatly within a few minutes after a TV program begins. The change points at the boundary of a TV program reflect interest in the TV program itself, not the contents at the time of change. This means that the points at boundaries are sometimes noise, obscuring the other points and thus making it difficult to analyze the relationship between the contents and TV rating at the change points.

We therefore use filters to exclude the points at the boundaries. Whether a points is at a boundary can be judged from the information of the EPG data. The filter is set to filter out the points within five minutes of a TV program beginning.

2) Commercial Film: A certain number of people usually change the channel when a CF starts, as mentioned in Sec. III. We sometimes want to deal with CF separately from others, for example, when the bias of transition along with CF want to be excluded. Therefore, we perform CF detection for all the broadcast data beforehand using the method based on frequent sequence mining proposed by Wu and Satoh [10]. It extracts CFs by using a mining procedure and frequently appearing scenes lasting 15 or 30 seconds are detected as CFs.

We use two filters related to the CF to filter out change points. To remove the bias created by the CF, we implemented a filter that filters out change points that occur during a CF and within 2 minutes after the CF. To focus on the pure interest people have in the program contents, we implemented a filter that filters our all change points except for those corresponding to when another TV station start showing a CF. At such points, viewers start switch channels and select a channel showing a program that is of interest.

3) Visual Features: The change points can be filtered using visual features described in Sec. V-B1. The filter is determined by the type of visual features (as listed in Table I) and by the threshold of the value of the feature. For example, we can extract the points that have over 50% black pixels, and filter out the points that have a score for person of more than 0.7.

4) Other Filters: In addition to the filters explained above, following four filters is implemented: time period (e.g., 19:00–23:00), term (e.g., Sep.–Oct.–Nov.), category of TV program (e.g., news), and keyword (e.g., Kinugawa River).

VI. EXPERIMENTS

A. Datasets

TV ratings. We used TV ratings data for Japanese TV provided by Video Research Ltd. The data included the audience ratings for seven TV stations in the Kanto area during 2015, from January 1 to December 31.

Video data. We used the video stream (MPEG4 encoded) for 2015 as broadcast by the TV stations corresponding to the TV ratings data. Since processing and decoding all the video data would have taken too much time, we processed and decoded the data only near the change points.

Transcripts. We used the transcripts of the TV programs provided by the broadcasters. They are generated by the broadcasters in real time. This text data amounted to 2.9 gigabytes in total for the whole year for the seven TV stations.

EPG data. The EPG data (provided by the broadcasters in XML format) included the program start time, duration, title, short description, and category.

The video data, transcripts, and EPG comes from the NII-TVRECS Video Archive System [6].

B. User Behavior Mining System by Visual Feature Analysis

1) System Overview: We now present a system for discovering user behavior by using our framework. In this system, change points are detected and descriptions are added to them in advance. In our experiments, the change points were detected using an increase threshold of 0.7% increase and a decrease threshold of 0.9%; 67,728 and 41,367 points were detected respectively. In our system, the visual features listed in Table I are computed for all detected points for use in filtering and aggregation. The change points are then filtered using the user-specified filter, and the result of filtered point aggregation is visualized. By analyzing the points with a rating increase or decrease exceeding a threshold, we can learn 1) what types of programs are of interest to people, 2) what features of programs result in high TV ratings, and 3) how people behave following a certain type of event.

Figure 5 shows the interface and flow of this system. In the following, we explain its usage by using a toy example of analyzing *in what type of news are users interested?*.

Filtering. The user first specifies the filters in accordance with the analysis target. The filters that can be used were explained in Sec. V-C. Here, since we wanted to analyze the news, we select the news category filter (Fig. 5 (a)).

Statistical analysis. The system then aggregates the filtered points and displays a clue to find the particular pattern of a TV rating increase. That is, it displays information that should help to identify which feature is related to the TV ratings. For each visual features, the differences in the means of the feature values over the increase change points and over the decreased points is shown in Fig. 5 (b). This information indicates which features should be analyzed in detail. If the difference is large, the feature apparently contributes to an increase in the TV rating, and vice versa. A breakdown in the form of pie charts of the change points by dominant color and emotion are also displayed (Fig. 5 (c)). Dominant color means the color with
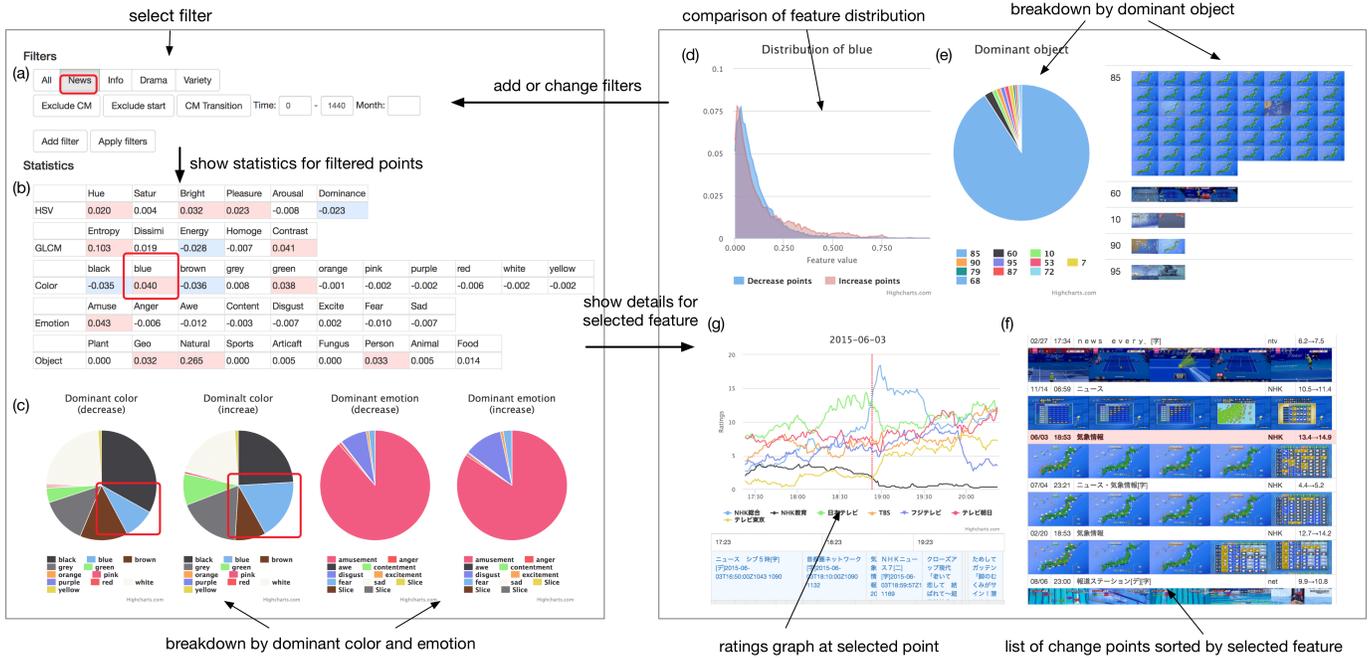
Fig. 5. Interface of user behavior mining tools.

the highest value of the 11 basic colors. The same is true of dominant emotion. On the basis of these charts, the user selects the feature that seems to be important in order to analyze it in detail. In this example, the user may select the feature *blue* because it is significant in both charts.

Feature details. The system then shows in the form of a graph details for the features selected by user (Figure 5 (d)). The graph shows the distribution of the selected features in terms of increase points and decrease points. It also shows a breakdown by dominant re-generated objects for the change points with the top-200 highest values (bluest in this example) (Figure 5 (e)). Thumbnails of the points corresponding to each object are also shown. These charts reveal that blue is the most significant color because of the weather reports, indicating that the viewers were highly interested in weather reports and thus tended to tune in to them.

In addition to these charts, the system show example lists of points with thumbnails and brief information for each point, sorted by feature value (Fig. 5 (f)). It also provides more detail for each point in the form of a ratings graph along with the program guide information (Fig. 5 (g)).

Re-filtering with interactive feedback. Additional patterns can be discovered by adding or changing filters on the basis of the initial results. In this example, the object corresponding to weather reports (id=85) can be filtered out by adding a filter for visual features, as presented below in Sec. VI-B2 Q1. In this way, numerous patterns of audience behavior we can discovered by combining various types of filtering with interactive feedback.

2) Examples: Figure 6 shows examples of questions that can be asked to obtain knowledge from TV ratings data using our framework. The first three questions (Q1–Q3) represent analysis for finding a particular pattern with rating increase

for a specific category. In addition to the program category filter, we used the filters that limit to transition at time of CF, and exclude CF and program boundary in order to focus on the pure interest people have in the program contents. Q1 is the same question as the toy example in Sec. VI-B1. On the basis of the results of the first analysis, we filter out weather report using object filter in order to analyze other factors. The result after filtering indicated that green was the most sensitive color. By selecting green, we obtained more examples and found that sports news is also popular. For Q2, we used the drama filter. Black and dominance showed the most significant difference between the increase points and decrease points. The examples indicate that scenes where the screen is heavily dark tend to be serious scenes, so viewers tend to stay on that channels. For Q3, we used the variety filter. We found that high entropy and amusement tends to increase the TV ratings. The examples show that when the screen has high entropy or amusement, the scene tends to be lively, which matches the needs of people who watch variety shows. The difference of user behavior between in variety shows and dramas can be clearly recognized at a glance from the comparison of the examples produced by our tool.

For Q4, we tested whether animals have ratings, which is empirically believed by many broadcasters in Japan. We first compared the distribution of the scores for an animal between the increase points and the decrease points for some category and found that a difference in distribution can be observed, especially for variety shows. To discover more about how an animal can contribute to TV ratings, we used an animal filter with a threshold score of 0.3. Analyzing decrease points with animal revealed that animals do indeed have ratings, except for ones in water.

Q1: In what type of news are people interested?

→ A: People are intersted in weather reports and sports news.
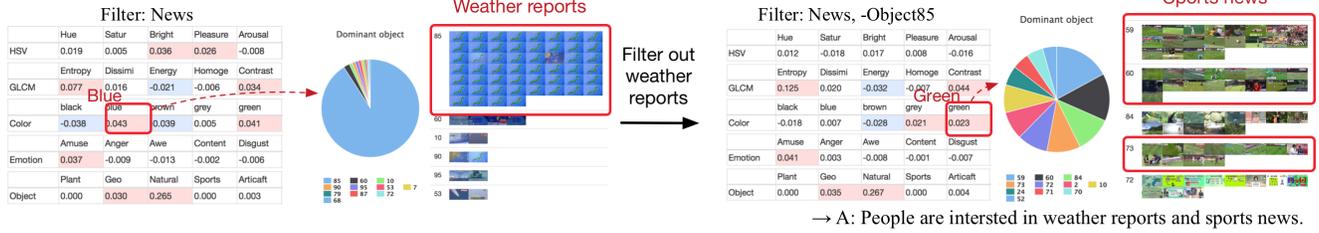
Q2: What makes people watch drama?

→ A: People tend to stay channel in serious scenes where the screen becomes dark.
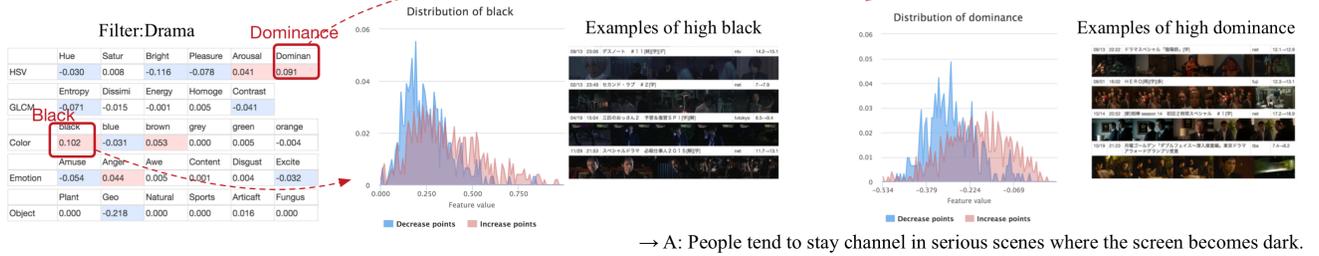
Q3: What is the key indicator for variety shows?

→ A: People want lively scenes with high entropy in variety show.
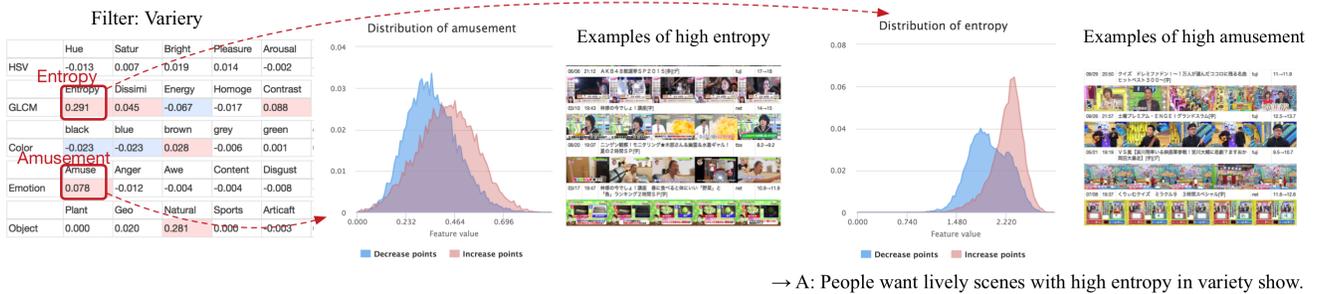
Q4: Is it true that animals have ratings?

→ A: Animals certainly have ratings, especially for variety shows, except for animals in water.
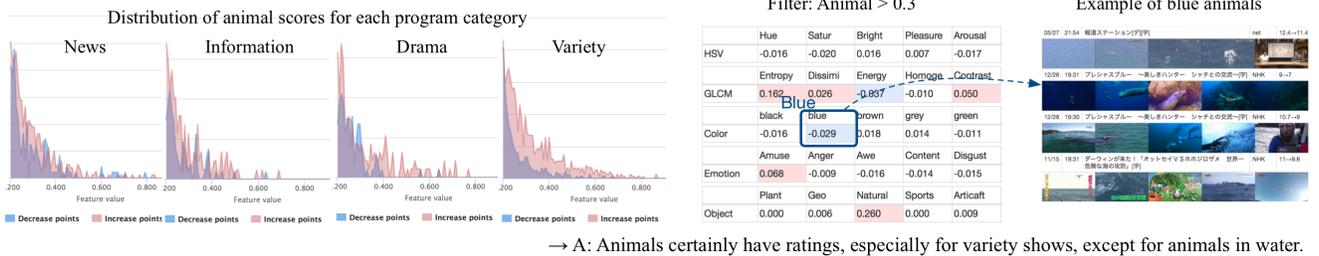
Fig. 6. Examples of analysis of user behavior using proposed mining system.

C. News Event Detection and Analysis

1) Overview and Detection Algorithm: We present the system that targets the analysis of news programs. We detect the news stories of most interest using the set of change points. We use increase points observed in programs of news and information program categories that are expected to reflect the viewers' interest to news stories. Analysis of the popular news reveals valuable knowledge. For example, whether the intention of TV stations is at variance with the actual interests of people can be confirmed by the relationship between the number of increase points and broadcasting time. In addition, the viewing information at disaster will reveal the consciousness of disaster prevention, which is the important knowledge from the aspect of risk management.

We detect news stories using keywords cooccur with increase points. The set of increase points are regarded as the set of graph nodes. First, the nodes that are considered to be the same topic are connected, which is determined by keywords and broadcast dates; two points separated within two days and having more than three common keywords are connected. Connected components of graph are then extracted and each component represents the news story. The number of change points in each news story indicates the degree of people's interest in the news story. Top-ten frequently appeared keywords for each news story are used as the keywords of the news, which is used to extract the broadcasting time for each news. Using this algorithm, news stories of most interest can be detected with no prior knowledge. Although we tested

TABLE II

Correspondence between top-ten news stories in Yomiuri Newspaper and our top-ten detection results. Some keywords detected by our method are also shown in second column and the keywords annotated by hand are shown in brackets for the news stories that cannot be detected by our methods. The Y and CP columns show the rank in Yomiuri Newspaper and our results based on the number of increase points. The time column shows the accumulated airtime for each news story and the ranking of airtime among the 15 news stories in table is also shown in brackets.

Story	Keywords	Y	CP	Time
Two Japanese awarded Nobel Prize	(Novel Prize, Kajita, Omura)	1	-	11h 57m (11)
Three victories in Rugby World Cup	(Rugby, World Cup, Goroumaru)	2	-	3h 3m (15)
Japanese hostages killed by ISIS	Goto, Jordan, Islam, release, restraint	3	2	115h 25m (1)
Start of My Number system	(My Number, system, individual)	4	-	19h 51m (9)
Heavy rainfall in Kanto and Tohoku areas	outburst, save, levee, heavy rain, Kinugawa	5	1	29h 45m (5)
Enactment of security bills	bill, ruling party, vote, security	6	10	21h 35m (8)
Launch of Hokuriku Shinkansen	(Hokuriku, Shinkansen, Launch)	7	-	21h 53m (7)
Building data falsification	material, falsification, building, construction	8	8	24h 43m (6)
Agreement in principle on TPP	(TPP, agreement, partnership)	9	-	9h 58m (14)
Withdrawal of Tokyo Olympics logo	emblem, design, Olympics, Belgium, logo	10	6	18h 20m (10)
November 2015 Paris attacks	Paris, terrorism, France, multiple	-	3	89h 36m (2)
Osaka junior high school students' murders	Hirata, Yamada, abandonment, Neyagawa	-	4	55h 52m (3)
Kawasaki murder at Tamagawa	Uemura, boy, Kawasaki, bank, Tamagawa	-	5	34h 36m (4)
Plane crash in Chofu	small size, crash, Chofu, airframe, airfield	-	7	10h 10m (13)
Wakayama 11-year-old boy's murder	Nakamura, cutlery, boy, Morita, Kinokawa	-	9	10h 37m (12)

some basic clustering algorithms such as k-means and spectral clustering, they tended to merge unrelated news stories and our simple algorithm produced more accurate results than others.

We also extract the broadcasting time for each detected news story. Topic segmentation, i.e., finding topic boundaries of news programs, is first performed on the basis of the method used in [5]. It detects a topic boundary by finding a point where the keyword distribution has significantly changed. If a segment includes more than three of ten keywords of a certain news story, we regard the segment as that news story. In this way, we can obtain the broadcasting time of a certain news story if its keywords are given.

2) Analysis of News of Interest: We first evaluated whether our method can detect important news stories. We used the *top-ten news stories in Japan for 2015*, as reported by the Yomiuri Newspaper, as the ground truth. Table II shows the correspondence between the ground truth and our top-ten detection results. The CP column shows the rank based on the number of change points in each detected news. This result shows that our method detected five of ten ground truth. The rank in our results reflects the active behavior of people who desire the information about the news, which is different from the nature of just important news. Our results revealed that the terrifying news stories such as a case of murder or disaster tend to make viewers switch the channels to proactively collect information of the stories, while positive news stories such as Nobel Prize and Rugby World Cup do not make viewers switch the channels because viewers passively watched the stories.

We also show the accumulated airtime of each news in Table II (the time column). The murder cases that were ranked in our result but not ranked in Yomiuri (Osaka, Kawasaki, and Wakayama cases) were ranked highly in airtime, while the important news stories that were not detected by ours such as the launch of Hokuriku Shinkansen were also spared much airtime. It indicates TV stations broadcast important news stories equally to some extent regardless of the interest of people, while news stories of public interest are broadcast with some emphasis.

VII. Conclusion

Our proposed framework integrating TV ratings and other multimedia data such video and text transcripts can be used to discover audience behavior. We focused on micro-level change points in TV ratings data that contain valuable information about audience behavior. A system based on our framework can discover the various types of knowledge from a large number of change points by using various filtering and an aggregation technique. The results produced by our mining tools demonstrated that our framework can discover numerous valuable knowledge from TV ratings by combining the filters interactively. Another example application demonstrated that a system based on our framework can detect the news stories of most interest without any prior knowledge.

VIII. Acknowledgment

References

[1] Anagnostopoulos, A., Kumar, R., Mahdian, M.: Influence and correlation in social networks. In: Proc. of ACM SIGKDD (2008)
[2] Benevenuto, F., Rodrigues, T., Cha, M., Almeida, V.: Characterizing user behavior in online social networks. In: Proc. of ACM SIGCOMM (2009)
[3] Danaher, P.J., Dagger, T.S., Smith, M.S.: Forecasting television ratings. International Journal of Forecasting 27(4), 1215–1240 (2011)
[4] Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., Fei-Fei, L.: Imagenet large scale visual recognition competition 2012 (ilsvrc2012) (2012)
[5] Ide, I., Mo, H., Katayama, N., Satoh, S.: Topic threading for structuring a large-scale news video archive. In: Proc. of CIVR (2004)
[6] Katayama, N., Mo, H., Ide, I., Satoh, S.: Mining large-scale broadcast video archives towards inter-video structuring. In: Proc. of PCM (2004)
[7] Krizhevsky, A., Sulskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. In: Proc. of NIPS (2012)
[8] Kudo, T., Yamamoto, K., Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis. In: Proc. of EMNLP (2004)
[9] Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: Proc. of ACMMM (2010)
[10] Wu, X., Satoh, S.: Ultrahigh-speed TV commercial detection, extraction, and matching. IEEE Transactions on Circuits and Systems for Video Technology 23(6), 1054–1069 (2013)