ISSN 2277-2685
IJESR/Sep. 2023/ Vol-13/Issue-3/1-06
Dr. Rafeeq. MD *et. al.*, / International Journal of Engineering & Science Research

Audience Behavior Mining: Integrating TV Ratings with Multimedia Content

Dr. Rafeeq. MD [1], Mothe Sai Sreya [2], Nimmala Sai Sri Chandan [2], P.Sheeba rani [2], Yeganti Dinesh Goud [2]
Professor,Computer Science and Engineering, CMR Engineering College, medchal, T.S, India,
[2]B.Tech, Computer Science and Engineering, CMR Engineering College, medchal, T.S, India

I. **Abstract:**

To discover relationships between TV ratings and multimedia content, we focus on the change points—that is, the points in time when people first tune in to a particular TV program. Because these points reflect the active intention of TV viewers, they contain valuable information about viewers' interests. We describe these points using visual features extracted from video and using keywords extracted from transcripts. Because the number of such points is huge, we apply various filters. Here, in this extended version of our earlier work, we propose two applications using our framework, demonstrating how the framework can extract valuable knowledge from TVratings data. A general framework for audience behavior mining integrates the analysis of TV ratings and multimedia content. Focusing on change points in TV ratings data, the authors propose applications for interactive mining of audience behavior and for detecting popular news topics.

Keywords: Multimedia Content, TV Ratings, Channel Surfing, Keywords.

II. INTRODUCTION

TV ratings have been investigated for many years, but most work has focused on forecasting the ratings of particular TV programs. The motivation was to estimate the cost of TV advertising, because advertising rates are directly linked to TV ratings. Few works have targeted the mining of ratings data, even though such data contains valuable information. Moreover, researchers have not investigated how to integrate TV ratings with multimedia content— such as with video data from TV programs. Integrating TV

ratings with multimedia content could help identify relationships between audience behaviors and TV program content[2]. This example captures audience behavior for just one event, but analysis of larger amounts of such data could

Furthermore, integrating such data with multimedia content could reveal hidden behaviors. Understanding audience behavior is useful for several reasons. First, it indicates what is of interest to people, which is important for creating TV programs that attract viewers. It's also important in terms of advertising, because identifying patterns that lead to higher ratings can help broadcasters obtain more sponsors. Finally[1], it helps with risk management by revealing how people get information following a disaster, which could help determine how best to convey important information to people in an emergency. Here, we focus on discovering knowledge in TV ratings data by combining such data with multimedia content (videos and transcripts). Our goal is to establish a generic framework for mining TV audience behavior from TV ratings. Although user behavior is being explored by many works, most of this recent work focuses on social networks; little work has addressed the behavior of TV audiences.

II TYPICAL AUDIENCE BEHAVIOR

Our focus here is on the mining of audience behaviors. Before explaining the details of our proposed method, we discuss audience behaviors—that is, how people typically watch TV and switch channels.

Audiences seemingly select channels according to their interests. However, audiences do not always watch the programs that best match their interests, because they only have information about the channel they're watching, they don't have information about the channels they're not watching. Such passive watching behavior does not reflect the audience's intention. The only cases when we can directly observe the intentions of audiences is when they change channels after "zapping" (also known as "channel surfing")[3]. Audiences change channels and start zapping when uninteresting content (such as a CF) starts. Moreover, because audiences get the information of all channels by zapping, they can select the channel of most interest. In other words, such cases show the active intentions of audiences. Our framework extracts such activeaudience intentions by focusing on the change points in TV ratings, which facilitate the discovery of audience interests[4].

A. Framework for Mining Audience Behavior

To automatically find particular patterns or events indicating the interests of people, we focus on the change points in ratings data. In particular, we focus on the micro-level change points, where the per-minute ratings change significantly in a minute. Because the number of viewers increases or decreases suddenly at such points, we assume that these points provide more valuable information than other points, so we detect and analyze them in combination with video content and other metadata, such as captions. We Can better understand the interests of viewers by analyzing the content corresponding to these change points. The meaningful patterns in TV ratings data can be discovered by mining a large number of change points. The pipeline of the proposed framework, shown in Figure 3, has three steps:

- Detect change points: Detect as many micro-level change points as possible from ratings data.
- Describe change points: Extract information for each change point from multimedia data (extract visual features from video data, for example).
- Filter and aggregate: Filter points to reject noise or extract the target, and aggregate filtered points to extract knowledge.

Our main focus is on steps 2 and 3—that is, describing, filtering, and aggregating the change points using multimedia content. Various features are extracted from oneminute data for each point to attach rich descriptions that characterize the points. These characterizations are used to filter and aggregate the points, and the filtered points are then visualized. Visualizing the statistics of change points helps us better understand patterns in ratings increases[5]. We can interactively add or change filters to visualize more detailed cases and find meaningful patterns. We use the following multimedia data to describe the change points:

- video—that is, broadcast videos corresponding to the TV ratings;
- captions—that is, text transcripts of TV programs; and
- Electric program guide (EPG) information— that is, information on TV programs, including title, category, and description.

We can describe the content at each change point with rich information using this data. Filtering and aggregating described points enables a precise analysis of the relationship between the ratings and the multimedia content [6].

III.FUNCTIONS

Here, we review the functions of the proposed framework.

A. Detecting Change Points

We adopt a simple approach to detect change points—we identify the points where the ratings increase (or decrease) above a predetermined threshold in a minute. The rate of increase (or decrease) for each minute is calculated as the difference between the previous and current per-minute ratings. A certain number of households start watching a particular TV program at change points, either by turning on the TV or switching from other programs[7].

B. Describing Change Points

We use several features to describe change points.
Visual Features: We use several visual features to characterize images. We use color and texture as low-level visual features, and we use object and emotion features as mid-level features. The details of each feature are described elsewhere.
Object Features: We generate two object category features based on the ImageNet large-scale visual recognition competition (ILSVRC) classification score. First, we use the

scores for the nine top-level categories of ImageNet, calculated by aggregating the classification scores for 1,000 object categories[8]. Second, we define new object categories suitable for TV content by clustering the features of Convolutional Neural Networks. On the basis of these categories, binary feature are obtained for each frame, where each dimension indicates whether the frame belongs to each cluster. Each generated cluster represents certain objects or scenes that frequently appear in TV content[9].

C. Keywords

Keywords are extracted from the transcripts at each change point. The words that characterize the TV content are selected as keywords. We first extract nouns from caption and then remove words that are not important to characterize the content. The details of keyword extraction are described elsewhere[10].

D. Other Information

We obtain basic information, such as the broadcast time and category of TV programming, from the EPG[11].

IV.FILTERING TECHNIQUES

Here, we describe the filtering techniques we use.

TV Program Boundaries: Because many people switch the channel when a new TV program starts, TV ratings change greatly at the boundary of a TV program. The change points at the boundary of a TV program reflect the audience’s interest in the TV program itself—not their interest in the content being shown at the time of change. This means that the points at boundaries are sometimes noise obscuring, making it difficult to analyze the relationship between the content and TV rating. We therefore implement filters to exclude points at the boundaries—that is, the points within five minutes of a TV program beginning[12].

Commercial Films: Many change points occur around CFs, because a certain number of people change the channel when a CF starts. We detect CFs from all broadcast data beforehand, because we sometimes want to deal with TV content separate from the CFs—for example, to exclude transitions made during a CF. We detected CFs using a method based on frequent sequence mining. We used two filters to reject change points. The first filter rejects change points that occur during a CF and within two minutes after the CF. This helps remove changes motivated by a CF (instead of by user interests). The second filter extracts change points corresponding to the transition from a channel showing a CF to another channel, which is used to focus on people’s interest in the program content. At such points— that is, those corresponding to when a TV station starts a CF—some viewers start zapping and select a TV program of most interest [13].

Visual Features And Their Thresholds: As noted earlier, visual features are used to filter out change points. The type of visual feature and the threshold of its value determine the filter[14].
Other Filters: In addition to the filters just introduced, we use four additional filters: time period, term, TV program category, and keyword.

V.EXPERIMENTS

Our data includes the audience ratings for seven TV stations. We also use video data, transcripts, and EPG data from the NII-TVRECS Video Archive System.

A. User Behavior Mining System

Based on our framework, we developed an interactive system for discovering user behavior. The system detects change points using an "increase" threshold of 0.7 percent and a "decrease" threshold of 0.9 percent; 67,728 and 41,367 points were detected, respectively. We computed all of the visual features listed the detected points, using the features to filter and aggregate points. We then rejected the change points based on user-specified filters and visualized the aggregated result. Analyzing the increase and decrease points revealed valuable knowledge, such as which types of programs were of interest to people, which program features resulted in high TV ratings, and how people behaved following a certain type of event.

Filtering: The user first specifies the filters in accordance with the analysis target. The filters that can be used. Here, because we want to analyze the news, we select the news category filter.

Statistical Analysis: The points after filtering are aggregated and visualized, which provides a clue to find the meaningful pattern of TV rating increases/decreases. That is, the visualized results display information that should help identify which feature is related to the TV ratings. For each visual feature, the differences in the means of the feature values over the increased/decreased change points. This information indicates which features are correlated with TV ratings. Features with a large difference seem to contribute to an increase in the TV rating, and vice versa.

Feature Details: The system then shows, in the form of a graph, details for the features selected by the user. The graph shows the distribution of the selected features in terms of increase and decrease points. It also shows a breakdown by dominant regenerated objects for the change points with the 200 highest values. Thumbnails of the points corresponding to each object. These charts reveal that blue is the most significant color because of the weather reports, indicating that the viewers were highly interested in such reports. In addition to these charts, the system shows lists of points with thumbnails and related information, sorted by the feature value. The system also provides more detail for each point in the form of a "ratings graph" along with the programguide information.

B. Refiltering With Interactive Feedback

We can interactively add or change filters to discover additional patterns. In the example the object corresponding to weather reports(ID = 85) can be filtered out by adding a filter for object features. This filter lets us analyze popular new topics after removing the noise from the weather report. In this way, numerous patterns of audience behavior can be discovered by combining various types of filtering with interactive feedback. In addition to the program category filter, we used the filters that extract transitions made during a CF to focus on people's interest in the program content.

C. News Event Detection and Analysis

We developed an application to detect news stories of interest by analyzing the set of change points in news programs. We use increase points observed in two categories news and informational programs—which reflect viewers' interest in news stories. In addition to finding popular news topics, our analysis revealed other valuable knowledge. For example, by comparing broadcasting times and ratings increases, we could analyze whether the intention of TV stations matched the interests of viewers. In addition, analyzing audience behaviors when watching news about a disaster can help broadcasters and others determine how best to disseminate information regarding disaster preparedness to maximize who sees, retains, and acts on such information

VI.CONCLUSION

TV audience ratings (TV ratings) are a key indicator in the field of TV broadcasting; they're used to assess the popularity of TV programs. A program's TV rating indicates the percentage of all TV households tuned in to that program. TV ratings are a standard measure used to determine the impact of advertising—

that is, how many targeted people are watching the advertisement. Broadcasters focus on increasing the ratings of their programs to acquire more sponsors. TV ratings can also be used as sensor to gauge the interests of people. Programs that capture people's interests get high ratings. Conversely, if people are not interested in a program's content, they switch to another channel, and the ratings decrease. TV ratings can thus act as social sensors that indicate popular topics and social trends— such as what types of news are of interest and which performers are currently popular. Discovering such knowledge from TV ratings data can help broadcasters create TV programs that attract more people. TV ratings also include important information for risk management. In an emergency, such as a natural disaster, the government must deliver correct information to people in a timely manner. TV is a key medium for conveying information to many people in real time. By analyzing TV ratings to determine how people get information from TV, we can judge whether critical information has been correctly delivered and can take measures to improve the dissemination of such information.

VII. REFERENCES

[1]P.J. Danaher and T.S. Dagger, "Using a Nested Logit Model to Forecast Television Ratings," Int'l J. Forecasting, vol. 28, no. 3, 2012, pp. 607–622.

[2]P. Buddh Bhagwan Sahu and Rajesh Tiwari, "Improving Processor Load by using Token Memory Approach", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 4, Issue 5, May 2016, pp 10004 – 10007, DOI: 10.15680/IJIRCCE.2016. 0405280, ISSN (online): 2320-9801 ISSN (print) : 2320-9798.

[3]H. Yin et al., "A Temporal Context-Aware Model for User Behavior Modeling In Social Media Systems," Proc. 2014 ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2014, pp. 1543–1554.

[4] Rajesh Tiwari, Manisha Sharma and Kamal K. Mehta, "Decentralized Dynamic Load Distribution to Improve Speedup of Multi-Core System using Parallel Models", Journal of Computational Information Systems, Vol. 15, issue 3, June 2019, pp 308 – 315, ISSN : 1553 – 9105.

[5] Santosh Kumar Srivastava Dr. Yogesh Kumar Sharma Dr. Yogesh Kumar SharmaSheo Kumar, " Precision enhancement of Intrusion detection system through outlier detection and feature classification", International Journal of Control and AutomationVol. 12, No. 6, (2019), pp. 820-830, ISSN: 2005-4297 IJCA.

[6]TV Rating Guide Book [in Japanese], Video Research,2016;www.videor.co.jp/rating/wh/rgb201610.pdf.

[7]J. Rajesh Tiwari, Manisha Sharma, Kamal K. Mehta and Mohan Awasthy, "Dynamic Load Distribution to Improve Speedup of Multi-core System using MPI with Virtualization", International Journal of Advanced Science and Technology, Vol. 29, Issue 12s, 2020, pp 931 – 940, ISSN: 2005 – 4238.

[8]A. Krizhevsky, I. Sutskever, and G.E. Hinton, [9]"ImageNet Classification with Deep Convolutional Neural Networks," Proc. Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.

[10] Srihari Rao Nidamanuru, Rajesh Tiwari, Bhaskar Koriginja, Jincy Denny, J Ramesh Babu, "Identifying the Websites that Maintain Operational Standards through Obligation Links to Website-Standards Approval Body", Turkish Journal of Computer and Mathematics Education, Vol. 12, Issue 10, 2021, pp 4456-4461, e-ISSN: 1309-4653.

[11]N. Katayama et al., "Mining Large-Scale Broadcast VideoArchives Towards Inter-Video Structuring," Advances in Multimedia Information Processing— PCM 2004, LNSC 3332, Springer, 2005, pp. 489.

[11] Santosh Kumar Srivastava1a,, Yogesh Kumar Sharma2, Sheo Kumar3, "Characteristics Categorization Dataset KDD cup'99 ", Advances in Basic Science (ICABS 2019)
AIP Conf. Proc. 2142, 110034-1–110034-7; https://doi.org/10.1063/1.5122494.

[12] Missing Kid Identification Proof Framework Utilizing Deep Learning And Multiclass Svm.
[13] Dr. C.N. RAVi , D. Palanivel Rajan, Desa Uma Vishweshwar, Edem Sureshbabu (CMREC)'A Review on Various Cloud-Based Electronic Health Record Maintenance System for COVID-19 Patients'Name of the Journal with ISSN:*Advances in Cognitive Science and Communications*, Cognitive Science and Technology - Springer Nature Singapore Pte Ltd. 2023 (CMREC Conference- ICCCE-2022)Vol. / Issue /PP. No. / Date/Month & Year of Publication:Impact Factor: https://doi.org/10.1007/978-981-19-8086-2_15, April 2023.
[14] Mrs.G.Sumalatha1 , Y.Jaideep Naidu M.Srividya, Karra karthik Reddy , D.Niharika, 'Smart OCR for Document Digitization', JASC: Journal of Applied Science and Computations, ISSN NO: 1076-5131, Volume VIII, Issue III, Macrh/2021.