J Vis (2021) 24:1287–1301
https://doi.org/10.1007/s12650-021-00771-1

REGULAR PAPER

Shangsong Liu · Di Peng · Haotian Zhu · Xiaolin Wen · Xinyi Zhang · Zhenghao Zhou · Min Zhu

MulUBA: multi-level visual analytics of user behaviors for improving online shopping advertising

Received: 8 July 2021 / Accepted: 18 July 2021 / Published online: 22 August 2021

Abstract The advertising revenue of online shopping platforms comes from the users who click on the advertisements and purchases the advertised goods. Therefore, to accurately advertise and increase revenue, advertising analysts engage in discovering representative groups and their behavior patterns from the data of demographic attributes, shopping behaviors, and advertising click behaviors of a large number of users. Existing methods often represent user behaviors based on single-level user profiles. However, under different community granularity and time scales, user behaviors have different characteristics. In addition, the sequential relationship between advertisement clicks and other shopping behaviors is difficult to be accurately identified by the single-level analysis methods. Therefore, we cooperate with advertising experts and propose a multi-level visual analysis method based on the K-Means algorithm, which can better understand user behaviors from multiple community granularity and multiple time scales. We design two novel visualization diagrams and improve three traditional charts that can help analysts observe user characteristics at the three levels: user groups, user subgroups, and user individuals, as well as can analyze the time-series events such as advertising clicks and product purchases of representative users from multiple time scales. Furthermore, we implement a multi-view interactive prototype system MulUBA to help analysts put targeted advertisements and increase advertising revenue. Finally, we verify the effectiveness and usability of our approach by conducting three case studies and an expert evaluation on a real-world online shopping advertising dataset.

Keywords User behavior analysis · Visual analytics · Online shopping and advertising

S. Liu · D. Peng · H. Zhu · X. Wen · X. Zhang · M. Zhu (✉)
Sichuan University, Chengdu, China
E-mail: zhumin@scu.edu.cn

S. Liu
E-mail: piny@stu.scu.edu.cn

D. Peng
E-mail: dipeng@stu.scu.edu.cn

H. Zhu
E-mail: haotian.zhu@scuvis.org

X. Wen
E-mail: wenxiaolin@stu.scu.edu.cn

X. Zhang
E-mail: zhangxinyi@stu.scu.edu.cn

Z. Zhou
Virginia Polytechnic Institute and State University, Blacksburg, USA
E-mail: zhenghao@vt.edu

1 Introduction

Advertising services promote product sales by putting advertisements to targeted users and charge for advertisers. Choosing audiences based on characteristics of product and user is one of the long-standing problems of advertising services. Different from the advertisements placed on traditional medium, emerging online advertising on the e-commerce platforms (such as Taobao,[1] and Amazon[2]) can obtain user behavior data, such as clicking on advertisements, browsing and purchasing products. By executing data mining and analysis, online advertising can provide more effective product promotion by behavioral targeting (Yan et al. 2009). Online advertising analysts focus on improving the conversion rate of advertising, that is, the proportion of users who click on advertisements and then purchase products. In real-world scenarios, they have to spend a lot of time observing user behaviors and exploring reasonable algorithms through massive data reports, so that correlate the features of users and advertisements to better targeting.

The previous methods on online advertising performance optimization are based on statistical analysis, such as improving audience recognition through targeting technology (O'Donnell and Cramer 2015; Malheiros et al. 2012), and improve the accuracy of CTR (Click Through Rate) prediction models (Gai et al. 2017; Zhou et al. 2019). After user information and behavior data are available, there have been some studies on visualization of user profiles, which are mainly carried out from the perspective of user attributes (Nguyen et al. 2020; Wu et al. 2020) and user behaviors (Xia et al. 2021; Kim et al. 2021). In addition, there are some visual analytics studies of online shopping data, which focus on simple user behavior data (Wei et al. 2012; Xie et al. 2014) or commodity transaction data (Webga and Lu 2015; Fu et al. 2020), but fail to combine advertisement click data to analyze user behavior. After communication with advertising analysts, we find that the existing methods still have two main deficiencies: (1) It is hard to discover the sequential relationship between advertising clicks and shopping behaviors; (2) it is difficult to analyze user characteristics under different community granularity and time scales.

To address the above issues, we present a multi-level visual analysis method based on the K-Means algorithm and conduct a comprehensive analysis of users' demographic attributes, shopping behaviors, and ad click behaviors. The approach clusters users into different groups and helps analysts observe user characteristics at the three levels: user groups, user subgroups, and user individuals, as well as can analyze the time-series events such as advertising clicks and product purchases of representative users from multiple time scales. Under the guide of real-world requirements and design tasks, we design two novel visual diagrams (the Group Behavior View and the Behavior Pattern View), improve three traditional charts (the Subgroup Profile, the User Profile View, and the Cycle Behavior View), along with implementing an interactive multi-view linkage visual analytics prototype system MulUBA (Fig. 1).[3] Through the system, analysts can improve the performance of advertising by observing the distribution of users' behaviors and attributes from different levels intuitively, and analyzing the time-series behavior patterns of representative characters efficiently. Finally, we validate the effectiveness and usefulness of MulUBA through three case studies and domain expert assessments, and give discussion and conclusion of our work.

The contributions of this paper are summarized as follows:

- We summarize the requirements and design tasks for analyzing users' interactions on online shopping behaviors and advertisements, and propose a multi-level visual analytic pipeline to identify and compare user behaviors from multiple community granularity and multiple time scales.
- We design two novel diagrams to overview user groups based on behavior types and the time-series behavior patterns, along with improving three traditional charts for the trade-off between the volume of data and the cognitive ability of the advertising analysts.
- We implement an interactive system **MulUBA** to achieve **mul**tiple-level **u**ser **b**ehavior **a**nalysis on online shopping and advertising dataset, and conduct three case studies and an expert evaluation to validate usefulness and effectiveness of our approach through both qualitative and quantitative assessments.

[1] https://www.taobao.com.

[2] https://www.amazon.com.

[3] https://github.com/Piny-Lyo/MulUBA.

Fig. 1 MulUBA assists advertising analysts in exploring users' online shopping behaviors and attributes from multiple levels with rich interactions. (1) Group-level: The Group Behavior View (**A**) provides an overview of the number of users and behaviors of different groups. (2) Subgroup-level: the Subgroup View (**B**) presents the attribute distribution of user subgroups. (3) Individual-level: The User Profile View (**C**) shows the behaviors and detailed attributes of the selected individuals, the Cycle Behavior View (**D**) visualizes the periodic behavior mode of an individual, and the Behavior Pattern View (**E**) enables the exploration of the detail behavior patterns from multiple time scales

2 Related work

In this session, we survey the traditional methods for optimizing online advertising performance, the visualization of user portraits on attributes and behavior, and visual analytics of online shopping data related to our work.

2.1 Online advertising performance optimization

The existing researches on online advertising performance optimization are mainly manifested in two aspects: (1) improve the effectiveness of audience recognition with targeted technology; (2) improve the accuracy of click through rate (CTR) prediction models.

Personalization aims to make ads more relevant for users and more effective for advertisers (O'Donnell and Cramer 2015). Malheiros et al. (2012) found that increasing personalization to some extent can make advertisements more attractive. Yan et al. (2009) verified that accurate monitoring of advertisement clicks logs collected in commercial search engines can improve the performance of online advertising. Peng et al. (2020) presented a visual analytics system to analyze targeted advertising. Guo et al. (2021) proposed an intelligent system for generating advertising posters to promote product sales.

Regarding the CTR prediction model, Gai et al. (2017) proposed the large-scale piece-wise linear model (LS-PLM) to achieve the CTR prediction of shopping ads. Zhou et al. (2019) proposed the deep interest evolution network (DIEN), which obtains interest with temporal characteristics from the user's historical behavior sequences to predict the CTR. Since the CTR will change due to time-varying factors such as seasons and marketing strategies, Ktena et al. (2019) proposed a training model that can maintain the freshness of the data under the condition of label delay. Yuan et al. (2019) proposed a novel framework for counterfactual CTR prediction to overcome the difficulty of CTR predictions in real-world advertising systems.

However, only using existing methods cannot easily explore the data, analysts need to spend a lot of time browsing the data report to summarize patterns and build effective models.

2.2 Visualization of user profile

Existing researches on user profile visualization can be broadly classified into two categories, one is based on high-dimensional multivariate data visualization methods to depict user properties, and the other is based on time-series data visualization methods to explore behavior patterns.

Bendix et al. (2005) proposed a new visualization method called parallel set to present categorical attribute data. The attribute distributions are presented by stacked histograms (Peng et al. 2018) and bar charts (Zhao et al. 2019). Nguyen et al. (2020) used a topic modeling method to extract tasks from the behavior sequences as user attributes, showing the attributes through a single-axis scatter plot. Wu et al. (2020) model each rally in a game as a sequence of hits and propose a reconfigurable glyph design to help analyze multiple attributes of the hits.

Ogawa and Ma (2010) expressed interactive behaviors between developers in software project evolution through Storyline; Kim et al. (2021) designed a stem structure visualization to express the Git commit behaviors of software developers. Xia et al. (2021) designed a novel Sankey diagram with embedded glyphs to visualize the multi-step problem-solving behaviors of different student groups. Zhao et al. (2020) used ellipses with gradient color to represent the time-varying patterns of different signals. In addition, some studies have mapped time-series data into a two-dimensional plane through the two-step dimensionality reduction method (Fujiwara et al. 2021) or tensor decomposition methods (Liu et al. 2019) to assist in the discovery of behavior patterns.

However, existing high-dimensional multivariate data visualization methods are not well suited to the multi-level user attributes in online shopping scenarios, and those time-series data visualization methods designed for behavior pattern analysis are limited in time scalability.

2.3 Visual analytics of online shopping data

Most of the existing researches on visual analytics of online shopping data revolve around users' behaviors. Wei et al. (2012) proposed a visual analytics system that determines user behavior patterns based on clickstream data of eBay.[4] Xie et al. (2014) proposed a visual analytics system VAET, which supports the exploration of significant behaviors in large-scale transaction data under different time granularities. Dextras-Romagnino and Munzner (2019) proposed a novel visual analytics interface Segmentifier, which supports an iterative process of refining action sequences into meaningful segments.

In addition, there are some researches on visual analytics of online shopping data focus on products. Webga and Lu (2015) designed a real-time visual analytics system to detect ranking fraud in products of online stores. Yu et al. (2020) used the information of users, commodities, and transactions to help merchants make business decisions from the perspective of sales trends. Fu et al. (2020) used IoT technology to track users' eye movements on product pages, using heatmap to visualize consumers' browsing interests and the decision-making processes of online shopping.

However, the analytical granularity of these studies is not comprehensive enough, and most of the analysis tasks only focus on a single group or individual. Moreover, the data are relatively single, most studies do not combine shopping behaviors such as adding products to the shopping cart or wish list, and clicking on advertisements to analyze user behaviors comprehensively.

3 System overview

3.1 Requirements analysis

We interview three domain experts (E1–E3) from the advertising business department of a large Internet company. E1 is the strategic product manager; E2 and E3 are the advertising analysts. After understanding the real-world requirements of domain experts, we list the following three primary design requirements (R1–R3) that help the domain experts to better advertise.

R1. Identify and compare different user groups. Experts need an overview of different user groups so that they can quickly discover and compare user groups with different behavior patterns. For example, they

[4] https://www.ebay.com/.

want to know which group prefers to buy goods or find out user group with high-frequency advertisement clicks for further exploration.

R2. Discover typical subgroups in the group. Due to a large number of users on the online shopping platform, narrowing the scope of data through subgroups can help experts find users with typical characteristics more quickly, especially discover the subgroups with high-frequency advertisement click behaviors to improve the effectiveness of advertising.

R3. Find individuals with special behaviors, and analyze their attribute characteristics and periodic behavior patterns. Observe the attributes and behavior patterns of special individuals, such as users with high-frequency advertisement clicks, can help experts to place advertisements in a more targeted manner. Moreover, experts need to compare the attributes and behaviors of multiple special individuals at the same time. In addition, they need to analyze the daily or weekly periodic behavior patterns of individuals during online shopping. For example, experts want to know on which day of the week the user is more likely to visit the shopping platform and at which times of the day to click on advertisements of specific product types more often.

3.2 Data description

In the online shopping advertising scenarios, user-related data includes the user's demographic attributes and various behaviors including advertising clicks, ad-related data includes product categories, brands, prices, etc. Users will click on specific types of advertisements based on their interests and needs at certain times.

Under the guidance and advice of experts, we use the open dataset Ali_Display_Ad_Click[5] offered by Alibaba Cloud and Alimama due to its availability and various data types. The desensitized dataset is about online shopping and advertisement click behaviors of users on the website of Taobao, and it contains four datasheets: user information, user shopping behavior, advertisement information, and advertisement click behavior. Specifically, user information contains user ID, gender, age, consumption level, shopping depth, city level, and whether he or she is a college student; user shopping behavior contains user ID, behavior type (browse products, add to the shopping cart, add to the wish list, and buy products), timestamp, product category ID, and brand ID; advertisement information contains advertisement ID, product category ID and brand ID; advertisement click behavior contains user ID, advertisement ID, timestamp, and whether clicked. The raw data include 723 million records of shopping behaviors and more than 26 million records of advertisement click behaviors of about one million users.

Before being used, the raw data are pre-processed based on our actual requirements. We eliminate the redundancy and noise of the raw data. After that, we get 374,369 users' 232 million records of shopping behaviors in 3 weeks (2017/04/22-2017/05/12) and about 2 million records of advertisement click behaviors in the last week (2017/05/06-2017/05/12). In addition, we use the K-Means clustering algorithm and elbow method (Syakur et al. 2018) to divide users into K clusters based on their behaviors. The elbow method is expressed by Sum of Squared Error as follows:

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{p \in P_i} |p - m_i|^2, \tag{1}$$

where is P_i the i th cluster, p is the data in each cluster, m_i is the center point in each cluster.

After counting the sum of each user's five behaviors (i.e., four shopping behaviors and advertisement click behavior), respectively, and standardizing the data, we combine the Elbow Method and the advice of advertising experts to determine that the best value of K is 5. Finally, the data processed by the K-Means algorithm is rearranged into the structured JSON format and stored in PostgreSQL[6] for later querying.

3.3 System pipeline

According to the requirements above, we develop MulUBA, an interactive visual analytic system for analyzing the user behaviors and attributes of online shopping. The pipeline of our system is shown in Fig. 2. After cleaning and clustering of the large-scale, high-dimensional, and time-series raw data, the

[5] https://tianchi.aliyun.com/dataset/dataDetail?dataId=56.

[6] https://www.postgresql.org/.

processed data are stored structurally. The visualization module includes multiple coordinated views implemented with D3 (Bostock et al. 2011) and SVG, and it provides user-friendly interactions to help analysts understand and interpret user behaviors from multiple community levels and time scales. More specifically, analysts can start the exploration from the Group Behavior View of group-level to get an overview, then check the Subgroup Profile View of subgroup-level, last focus on the User Profile View, the Cycle Behavior View and the Behavior Pattern View of individual-level for detailed information. The analysis requirements proposed by domain experts guide the design and implementation of the visualizations.

4 Visualization

4.1 Design tasks

Based on requirements **(R1–R3)** from expert interviews and the data feature, we derive the following design tasks **(T1–T5)** that guide the visual design.

T1. Show the overview of user behaviors of different groups (R1). The visual design needs to provide the overall distribution of each group on different behaviors and support the comparison of single behaviors and multiple behaviors between different groups.

T2. Demonstrate the user subgroups with specific attributes intuitively (R2). The visualization should emphasize the subgroups with different attributes and support further analysis of interested users' portraits.

T3. Reveal individuals with special behavior, and support the comparison of different individuals' attributes and behaviors (R3). The visual design should be able to find the unusual individuals in the group and show the user profile with detailed attribute distribution and behavior frequency. And, the system should support multiple selections of interested users for detailed comparison.

T4. Display the periodic behaviors of individuals (R3). The visualization design should intuitively show the weekly periodic behavior patterns of individual users during online shopping.

T5. Track the behavior patterns of individuals on a daily basis (R3). The visual design should be able to track the specific moments of users' certain shopping behaviors in a day so that support the discovery of typical behavior patterns of individuals.

4.2 Visual design

We design a novel visual analytics system MulUBA, which accomplishes the aforementioned design tasks. This system can aid advertising analysts in exploring, analyzing, and understanding the online shopping behaviors of users from a multilevel perspective. As illustrated in Fig. 1, the visual interface of our system consists of five coordinated components: the Group Behavior View for group-level behavior analysis, the Subgroup Profile View for subgroup-level attribute analysis, and the User Profile View, the Cycle Behavior View, and the Behavior Pattern View for individual-level profile and time-series analysis.

4.2.1 Group behavior view

The Group Behavior View adopts a novel visualization method for user groups based on behavior types, and it provides an overview of the number of users and the frequency of online shopping behaviors of different groups **(T1)**. Analysts can quickly find groups of interest and special individuals, and complete a series of comparison and selection operations **(T3)**.

The rows in Group Behavior View represent five groups of users after K-Means clustering, the elements within each row are connected by a gray horizontal line to increase the recognition of the same group. We choose five contrasting colors with high saturation, i.e., blue, cyan, yellow, orange, and purple, as the basic hues to map the five groups. In order to ensure the consistency of visualizations, all designs in the system involving group categories use these five colors by default. The left nested rectangle design shows the linear number ratio of all users, the group users, and the advertisement clicked subgroup (Fig. 3a). The outer rectangle with the basic hue represents all users, the dark gray rectangle in the middle represents users of a certain group, and the light gray inner rectangle represents the subgroup with advertisement click behavior. The length of the rectangles encodes the number of three kinds of users.

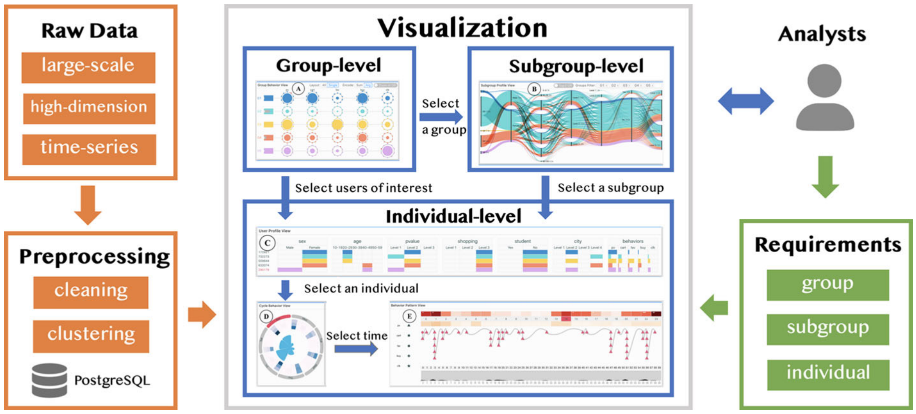

Fig. 2 Pipeline of MulUBA. After data preprocessing, our system supports the exploration of user behaviors and attributes from three levels of the group, subgroup, and individual. Analysts can gain insights from rich multi-view interactions (blue arrows)

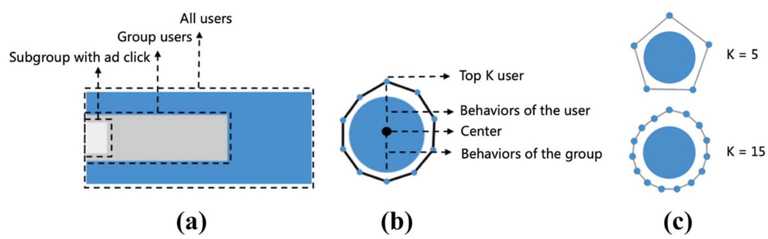

Fig. 3 Visual coding and alternative design choices (c) of the Group Behavior View

The right five columns of circles from left to right represent the five behaviors of the groups: browsing products (**p**age **v**iew), adding to the shopping **cart**, adding to the wish list (**fav**orite), **buy**ing products, and **click**ing on ads, respectively. The size of the circle is proportional to the total or the average number of the group behaviors. The larger the circle, the more times the group performs such behavior. When adopting the multi-behavior comparison layout (Fig. 1A), the comparison base of the circle size is the maximum value of all the behaviors, and the group behaviors can be compared globally from the horizontal and vertical directions. However, when there is a large gap among different behavior data, this layout will cause the relatively small data to be indistinguishable points. To solve this problem, we design the single-behavior comparison layout as shown in Fig. 1A1. When this layout is adopted, the comparison base of the circle size is the maximum value of a certain behavior, and only the vertical direction comparison is supported. We use the following square root scale to map the radius:

$$behR(\mathrm{num}) = \frac{(behR_{\max} - behR_{\min})^2}{behN_{\max}} \cdot \sqrt{num} + behR_{\min}, \tag{2}$$

where $behR$ is the radius of the circle, num is the total or the average number of the group behaviors according to the data mapping option, $behN_{\max}$ is the largest num of all behaviors or one single behavior according to the layout option, $behR_{\max}$ and $behR_{\min}$ are the predefined maximum and minimum radius of the circles, respectively.

In addition, to find individuals with special behavior, we use dots in the radial direction of the circle to represent the top K individuals with the most behaviors (Fig. 3b) in the single-behavior comparison layout. According to different K values, we have designed alternative choices as shown in Fig. 3c, under the advice of experts, we finally adopt K = 10 because it is more intuitive and readable. Starting from zero o'clock, the dots are arranged in a clockwise direction from the largest value, and the larger the number of behavior, the farther the dot is from the center of the circle. The dots are connected by a broken line, which can show the trend of the top K users' behaviors, and find individuals with unusual performance.

In terms of interaction design, we use the highlight, tooltip, and partial magnification to reduce the cognitive burden of analysts. When a single/multiple behavior layout switch or a total/average data mapping switch occurs, a reconfiguration can be triggered to redraw the Group Behavior View. When clicking on the left rectangles, the Subgroup Profile View will display the corresponding subgroups in linkage; when clicking on a dot outer the circles, the User Profile View will display the attributes and behaviors of the user.

4.2.2 Subgroup Profile View

To further present the distribution of group attributes and discover typical user subgroups **(T2)**, the Subgroup Profile View uses an improved parallel set to efficiently display the high-dimensional and multivariate attributes of the large-scale user groups (Fig. 1B). Different from the parallel coordinate diagrams that usually display continuous values, parallel sets can represent categorical high-dimensional multivariate data more effectively (Bendix et al. 2005). Parallel sets use data intervals to represent discrete attribute values and aggregate the broken lines in parallel coordinates into ribbons. While the attribute values are usually messy, the traditional parallel sets are easy to cause visual confusion. We improve the traditional layout algorithm by reordering the attribute values of each column so that the edges are arranged in an orderly manner, which is more readable even in the case of more attributes. As shown in Fig. 1B, the columns from left to right represent group, gender, age, consumption level, shopping depth, college student or not, and city level, respectively. The width of the interval represents the number of people who have the corresponding attribute value. The color scheme is consistent with the Group Behavior View, in which the five colors represent the five groups, and subgroups are divided according to different attributes.

As to the interactive designs, the Subgroup Profile View supports filtering and highlighting the top three subgroups by the number of users, which can help analysts focus on the subgroups of interest. Furthermore, when clicking on a certain subgroup, users of the subgroup will be present in the User Profile View sorted by the total number of behaviors simultaneously.

4.2.3 User Profile View

The User Profile View uses a mosaic chart with an even layout (Fig. 1C) to represent the attribute distribution of multiple individuals **(T3)**. Existing studies mostly use tables to show the distribution of high-dimensional attributes (Peng et al. 2018; Zhao et al. 2019). Although the table layout is simple and easy to understand, it lacks visual intuitiveness. The Mosaic chart displays the distribution of attributes using a combination of rectangles, but the traditional mosaic chart has low space utilization, and the uneven layout method makes it difficult for users to understand (Friendly 1994). To solve these problems, we use even mosaic rectangles to encode user attributes, and histograms to encode user behaviors. This design conforms to users' cognition and takes advantage of the intuitiveness of visualization. Meanwhile, the requirement of canvas height is extremely low, which can make good use of screen space.

As shown in Fig. 1C, the User Profile View consists of three layers of rectangles: the first layer is the name of the attributes or behaviors; the second layer is the specific type of the attributes or behaviors; the third layer is the value of the attributes or behaviors. When visualizing the attributes, fill the corresponding rectangle according to the attribute value with the basic hue of the user's group. When visualizing the behavior, the same group color is used to draw horizontal histograms. The width of the bar is proportional to the number of user behavior, and a logarithmic scale is used to eliminate visual confusion caused by excessive data gaps. When clicking on an ID on the left, the Cycle Behavior View will show the information of the corresponding user.

4.2.4 Cycle Behavior View

The Cycle Behavior View uses the improved circular heatmap (Fig. 1D), from which analysts can discover the periodicity of individual users' different behaviors **(T4)**. The circular heat map has a significant advantage in revealing the periodicity of time-series data on a small canvas. The traditional circular heat map can only encode one type of data (Best et al. 2010), but we need to encode five types of behaviors at the same time. In addition, the four online shopping behaviors in the dataset cover three weeks, while the advertisement click behavior data only covers the last one week.

Faced with these problems, we make improvements to the traditional circular heat map, enable it to encode multiple behaviors on the circles and support highlighting one of the most important behaviors. As shown in Fig. 1D1, the outer ring is evenly divided into seven arcs representing Monday to Sunday, the three-layer circles in the middle layer represent the first week, the second week, and the third week of the dataset from outside to inside. The four columns of rectangles from left to right represent four behaviors of browsing products, adding to the shopping cart, adding to the wish list, and buying products. The darker the rectangle, the more times the behaviors. To effectively distinguish the advertisement click behaviors from the other four behaviors, we use the inner Nightingale rose chart to indicate the number of advertisement clicks of the user in a day, the more advertisement clicks, the larger the fan size. When an analyst clicks on a certain day in this view, the Behavior Pattern View will link to show the user's sequential behaviors during the day.

4.2.5 Behavior Pattern View

Analysts can discover the behavior patterns of individual users at different times from the Behavior Pattern View to place targeted advertisements more effectively **(T5)**. Existing visualization methods for analyzing user behavior patterns usually present different behaviors on a single time axis (Ogawa and Ma 2010; Kim et al. 2021; Bendix et al. 2005), or multiple horizontal time axes (Nguyen et al. 2020). Although these design methods can present all behaviors at once and help analysts complete the analysis process without interruption, in actual application scenarios, there are still some shortcomings: (1) The screen space utilization is low; (2) it is easy to cause visual confusion when the scale of data is large; (3) in the limited screen space can only analyze behavioral patterns with large time granularity.

The Behavior Pattern View adopts the novel-designed visualization method of individual user behavior patterns (Fig. 1E). It provides analysts with two time-selectors at the hour and minute granularity, respectively, and maps the five behaviors into different symbols to analyze user behaviors at every moment. At the same time, interactive operations such as timeline brush are added to help analysts explore user behavior patterns in detail. The layout and visual coding of the Behavior Pattern View are shown in Fig. 4. The view adopts the dual time axis at the top and bottom, the top is the heatmap timeline at hour granularity, and the bottom is the area graph timeline at minute granularity. The left legend represents the five behaviors with five symbols, and in the middle is the behavior pattern evolution module with connected lines and symbols.

In terms of visual coding, the top timeline is composed of two rows and 24 columns of rectangles, the red rectangles on the top represent the sum of the user's five behaviors per hour in a week, and the orange rectangles on the bottom represent the sum of the user's five behaviors per hour during the selected day. The darker the rectangle, the larger number of behaviors in that hour. To highlight the advertisement click behaviors, small white dots are placed in the rectangle to represent the ad-click times. The bottom timeline uses a fluctuating area graph to encode the number of user behaviors per minute in a certain hour, and the larger the area, the more number of behaviors in that minute. The left legend shows the corresponding relationship between symbols and behavior, in which the triangle represents browsing products, the square represents adding to the shopping cart, the circle represents adding to the wish list, the plus sign represents buying products, and the star represents clicking on ads. The behavior pattern evolution module in the middle displays the specific behavior within the selected time range. Each behavior is represented by a symbol, and the behaviors in the same minute are vertically connected by a straight line, and the tail and head behavior between two adjacent minutes are connected with a cubic Bézier curve with four control points. This design scheme displays user behaviors in the order of time and has neat and clear connections

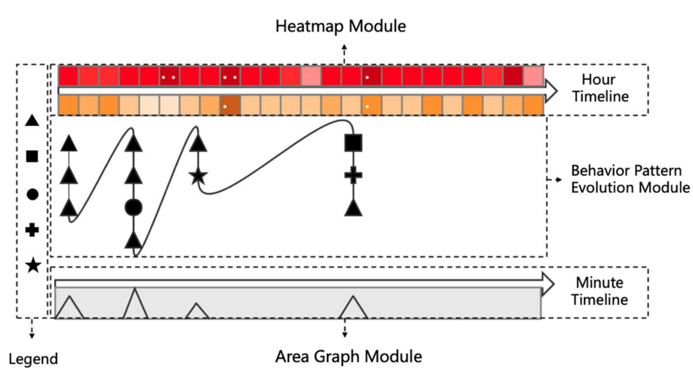

Fig. 4 Layout and visual design of the Behavior Pattern View

without overlaps. Through the distinct types and number of symbols, analysts can quickly discover the user's behavior patterns such as common behaviors and active time ranges.

As to interaction designs, analysts can select and highlight the top timeline at hour granularity, or brush the bottom timeline at minute granularity. When the mouse is hovering over the symbols in the central behavior pattern module, a tooltip will pop up to display the specific date and time, behavior type, and product category information. These interaction designs can avoid visual confusion, provide supplementary information for better understanding, and help analysts focus on the moment they are interested in.

5 Evaluations

5.1 Case study

To further evaluate the usability and usefulness of the MulUBA, we provide the following case studies demonstrating its capability in analyzing behavior patterns and characteristics of the users on Taobao.

5.1.1 Overview of user group characteristics

Analyzing the demographics of online shopping platforms is a key requirement in making effective marketing decisions. In this case, we describe how our system helps experts understand the distribution of user behaviors and attributes.

Example 1: Overview of group behaviors. When the Group Behavior View is initialized, the size of the circle represents the total number of behavior (Fig. 1A). We can find that among the five behaviors, the number of browsing products is much more than that of the other four behaviors. This phenomenon is consistent with the Purchase Funnel model (Hoban and Bucklin 2015). We can also clearly notice that group G2 has the largest number of users because it has the longest dark gray rectangle on the left. Further, use the partial magnification function and tooltip to explore the distribution of the behaviors in each group. In the bottom left of Fig. 1A, we can find the values of advertisement click users (CU), users of the group (GU), and platform users (PU) are equal in G5, which means that all users of this group have advertisement click behavior, so G5 is a group worthy of further analysis.

Example 2: Overview of group attributes. From the Subgroup Profile View (Fig. 1B), we can get the overview of attributes distribution of users on the shopping platform: G2 has the largest number of users; there are more women than men, more non-student users than students; most users aged 30–39; most users' consumption grade is level 2, shopping depths is level 3, and city grade is level 2. To separately observe the group G5 with the high advertisement click rate, we perform group filtering operations in the Subgroup Profile View and highlight the top three subgroups with the largest number of users (Fig. 5). We can get the profile of these subgroups: 30–49 years old female with consumption level 2, shopping depth level 3, and city level 2 or 3. Therefore, people with these characteristics are highly likely to click on advertisements when shopping, which deserves the attention of advertising analysts.

5.1.2 Discover users with special performances

Studying the special behaviors shown by users can help analysts adjust advertising strategies so that put advertisements to target users more accurately. In this case, we describe how MulUBA helps experts explore users with special performances.

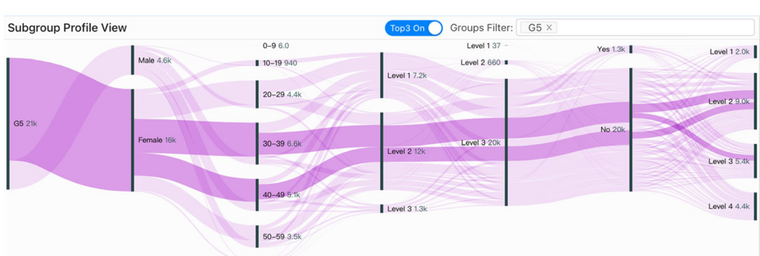

Fig. 5 Filter G5 in the Subgroup Profile View and highlight the top three largest subgroups

Example 1: Discovery of users with special adding to the wish list behavior. After changing the Group Behavior View to a single-behavior comparison layout with vertical comparison, and using the average number to encode the size of the circles (as shown in Fig. 6a), we can find that users of G3 add to the wish list most frequently (the yellow circle is obviously the largest in the third column), and the first dot representing the user who adds to the wish list the most times is protruding outward. Hover the mouse over the first dot, the tooltip shows that the user has added products to the wish list 1055 times (Fig. 6b), and that number of the second-ranked user is 584 with a significant drop (Fig. 6c). Then, select five dots clockwise along the ring to explore the top five users in the G3, and check the portraits of those users from the linked updated User Profile View. From Fig. 6d, we can find that the first five users who enjoy adding goods to the wish list are all women with consumption level 1, shopping depth level 3, city level 2 or 3, and non-student. In addition to adding to the wish list, they also like to browse products, but they have few advertisement click behavior. The biggest difference between the first-ranked user and the rest is that she does not add products to the shopping cart nor click on ads, and we can speculate that she is a middle-aged woman who has much free time and enjoys hanging out on Taobao, but not susceptible to advertisements. Analysts can consider serving few advertisements to such users.

Example 2: Discovery of users with special buying products behavior. Similarly, we also analyze top10 users of buying behavior in G1, because this group has the highest average purchase behavior. From the User Profile View (Fig. 7), we notice that most of them are male and all the shopping depths are level 3. In addition to buying behavior, they also have many browsing behaviors. Surprisingly, these users almost never click on ads. It is likely that these users have clear goals and are not susceptible to other factors. Analysts can consider reducing advertisements served to such users to save advertising costs.

5.1.3 Explore behavior patterns of high ad-click users

By studying the group attributes and behavior patterns of users with high-frequency advertisement clicks, we can reveal when and what advertisements users often click on, what are the behaviors before and after clicking on ads. Once the behavior patterns are summarized according to users' performance, it is possible to deliver targeted advertisements and optimize the performance of advertising.

From Fig. 6a, we can reconfirm that G5 is the group with the highest advertisement click rate, we further select the top10 users of advertisement clicks and analyze the distribution of attributes and behaviors in the User Profile View. As shown in Fig. 8a, the majority of the users belong to the three subgroups recommended for focus in Sect. 5.1.1, Example 2. Looking at the bar chart of the five behavior, we can see that they all prefer to browse products and click on ads, they buy items but not many times, suggesting that such users may have no clear purpose during online shopping. To drill down into the behavior patterns of such users, we select the first user (ID "495482") from the User Profile View to further analyze in the Cycle Behavior View. As shown Fig. 8b, the user has frequent browsing behaviors (the first column of rectangles is darker) during the three weeks (2017/04/22-2017/05/12). In the last week (2017/05/06-2017/05/12), she has advertisement clicks in five days and has clicked on advertisements four times on that Wednesday,

Fig. 6 Case of finding users with special performance in G3

Fig. 7 Profile of top10 buying products users in G1

Thursday, and Saturday. Further analyze the user's behaviors in the Behavior Pattern View (Fig. 8c), from the top red heatmap at week granularity, we find out that the user frequently visits the shopping platform between 11 and 22 o'clock. After exploring the orange heatmap at day granularity, we notice that the two advertisement clicks she has at 22 o'clock both appear on Thursday. Then, we check the user's detailed behavior sequences at 22 p.m. on Thursday and find that she first aimlessly browsed some products, then saw an advertisement with the category ID "6261" at 22:04, after browsing and adding this type of products to the wish list several times, she clicked on such advertisement again about half an hour later. These phenomena indicate that she is a potential purchase user of the merchandise with the category ID "6261."

We have also studied the other nine users in the same way and find that they all have specific periods to visit the shopping platform and are susceptible to advertisements. Such users can be extremely helpful in improving the daily activity of the platform. Based on the user behavior patterns analysts have discovered, they can understand how the advertisements are delivered accurately and make users like to click, further to improve the algorithmic models to optimize advertising effectiveness.

5.2 Expert evaluation

To prove the effectiveness and usability of the system from another aspect, we design an expert evaluation refer to the method used by Li et al. (2021). We invite 10 experts from relevant fields to participate in the evaluation, including the advertising product manager (E1) and advertising analysts E2 and E3 mentioned in Sect. 3.1, other three advertising analysts (E4-E6) from an Internet company, and four visual analytics researchers (E7-E10).

First, we briefly present the background and focus of the study to 10 experts, and explain the requirements and tasks of the visual analytics system. Secondly, we demonstrate the visual coding and functional characteristics of the prototype system and present the cases in Sect. 5.1 to the experts. Then, we invite them to explore the system freely. During their exploration, we have continuously observed and documented the experts' operational processes and oral opinions. Finally, we design a questionnaire with 10 questions (Table 1) and use the 5-point Likert scale ranging from "the most negative" (1) to "the most positive" (5). We also conduct one-on-one interviews with each expert and record the benefits and limitations of our study in detail.

The results of the questionnaire are shown in Fig. 9. In general, both the usability and the effectiveness of our system are highly rated by experts. They agree that MulBUA is convenient to use and efficient for user behavior analysis in online shopping advertising scenarios. They can complete the analysis process through the interactions provided by the system, even experts who have never been exposed to such visual analytic systems find it easy to understand the visual designs. E4 is puzzled at first because he cannot understand the

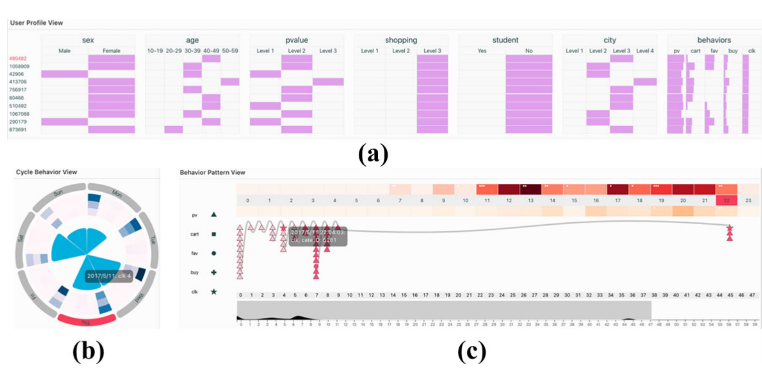

Fig. 8 Case of exploration of behavior patterns of the high-frequency advertisement click user "495482"

Table 1 Questionnaire of expert evaluation

Q1	The system is very easy (difficult) to learn
Q2	The system is very easy (difficult) to use
Q3	It is very easy (difficult) to understand the visual design of the system
Q4	The system can (cannot) improve productivity compared to traditional data reports
Q5	I am very willing (unwilling) to use the system in the advertising analysis scenarios
Q6	It is very easy (difficult) to compare the behavior distribution of different groups in the Group Behavior View
Q7	It is very easy (difficult) to find individual users of interest in the Group Behavior View
Q8	It is very easy (difficult) to discover the attribute distribution of subgroups in the Subgroup Profile View
Q9	It is very easy (difficult) to compare the attributes and behaviors of individuals in the User Profile View
Q10	It is very easy (difficult) to summarize the periodic behaviors of an individual in the Cycle Behavior View
Q11	It is very easy (difficult) to track the time-series behavior patterns of an individual with high-frequency advertisement clicks in the Behavior Patter View

Q1–5 are focusing on assessing the usability of MulUBA and Q6–11 evaluate its effectiveness in facilitating user behavior analysis in online shopping advertising scenarios

meaning of the elements in the Cycle Behavior View immediately, but after explaining its usage, he begins to enjoy using this view. E7 and E8 suggest that some SVG rendered elements can be changed to Canvas to improve the rendering speed and optimize the user experience. E10 suggests that when presenting top-K users in the single-behavior comparison layout of the Group Behavior View, the selection function of the K value can be added to support analysis of 5, 15, or custom number of users. Moreover, most experts agree that the User Profile View preserves the simple form of the table, reduces the learning cost, and visualizes the data in a clear form, meanwhile, it has obvious advantages when comparing both user attributes and behaviors. Experts also think the Cycle Behavior View and the Behavior Pattern View support the flexible track of the behaviors of the selected user, and can effectively address their analytic needs. But E1 suggests that advertising indicators, such as advertising exposure and consumption, can be added in future work so that the analysis results will be more comprehensive and specific.

6 Discussions

Our evaluations demonstrate the usefulness and effectiveness of MulUBA. Nevertheless, there is still space for improvement. Our system uses SVG to render views because it has unique interactive advantages, but this approach responds slowly to large-scale data, and one possible solution is to render elements that do not require interaction with Canvas. Although the customized Cycle Behavior View is useful, its scalability is low due to the peculiar time range of the dataset we use. In addition, the clustering algorithm we use is limited to K-Means, and it is not easy to find significant differences and bases of the classification results from the Group Behavior View and the Subgroup Profile View, and we can further consider supporting multiple clustering algorithms and exploring ways to enhance the interpretability of the results.

In the future, we can focus our research on advertising and work with e-commerce platforms to obtain more complete and rich data on user behaviors and advertising factors on one hand. On the other hand, we can improve the system framework to be more universal and flexible, which supports different datasets, custom clustering algorithms, custom TopK values, custom behavior icons, and apply it to other scenarios that require multi-level user behavior analysis, such as social media, educational platforms, entertainment games, etc.

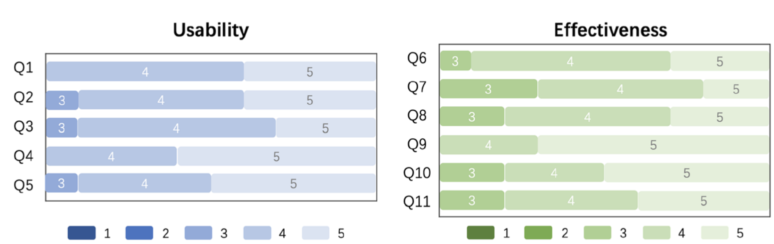

Fig. 9 Results of Q1–11 in our questionnaire, 1–5 represents "the most negative" to "the most positive"

7 Conclusion

In this paper, we propose MulUBA, a visual analytic pipeline that assists advertising analysts in exploring users' online shopping behaviors and attributes from multiple community levels and time scales. Under the requirements and design tasks, we implement the system that integrates the novel designed Group Behavior View and Behavior Pattern View with multiple coordinated views (Subgroup Profile, User Profile View, and Cycle Behavior View) to facilitate the comprehensive analysis. The insights learned from the exploring process can guide analysts to optimize the strategies of advertising. Three case studies and an expert evaluation demonstrate the effectiveness and usability of our approach on an online shopping and advertising dataset.

Acknowledgements This research is partially supported by the School-City Cooperation Special Fund Project (2020CDSN-02). We would like to thank the industry sponsor Alibaba Cloud and Alimama for providing with the data.

References

Bendix F, Kosara R, Hauser H (2005) Parallel sets: visual analysis of categorical data. In: IEEE symposium on information visualization, 2005. INFOVIS 2005, pp 133–140

Best DM, Bohn S, Love D, Wynne A, Pike WA (2010) Real-time visualization of network behaviors for situational awareness. In: Proceedings of the seventh international symposium on visualization for cyber security, VizSec '10. Association for Computing Machinery, New York, NY, USA, pp 79–90

Bostock M, Ogievetsky V, Heer J (2011) D^3 data-driven documents. IEEE Trans Vis Comput Graph 17(12):2301–2309

Dextras-Romagnino K, Munzner T (2019) Segmentifier: interactive refinement of clickstream data. Compu Graph Forum 38(3):623–634

Friendly M (1994) Mosaic displays for multi-way contingency tables. J Am Stat Assoc 89(425):190–200

Fu H, Manogaran G, Wu K, Cao M, Jiang S, Yang A (2020) Intelligent decision-making of online shopping behavior based on internet of things. Int J Inf Manag 50:515–525

Fujiwara T, Shilpika F, Sakamoto N, Nonaka J, Yamamoto K, Ma K-L (2021) A visual analytics framework for reviewing multivariate time-series data with dimensionality reduction. IEEE Trans Vis Comput Graph 27(2):1601–1611

Gai K, Zhu X, Li H, Liu K, Wang Z (2017) Learning piece-wise linear models from large scale data for ad click prediction. arXiv:1704.05194

Guo S, Jin Z, Sun F, Li J, Li Z, Shi Y, Cao N(2021) Vinci: an intelligent graphic design system for generating advertising posters. In: Proceedings of the 2021 CHI conference on human factors in computing systems, CHI '21. Association for Computing Machinery, New York, NY, USA

Hoban PR, Bucklin RE (2015) Effects of internet display advertising in the purchase funnel: model-based insights from a randomized field experiment. J Mark Res 52(3):375–393

Kim Y, Kim J, Jeon H, Kim Y-H, Song H, Kim B, Seo J (2021) Githru: visual analytics for understanding software development history through git metadata analysis. IEEE Trans Vis Comput Graph 27(2):656–666

Ktena SI, Tejani A, Theis L, Myana PK, Dilipkumar D, Huszár F, Yoo S, Shi W(2019) Addressing delayed feedback for continuous training with neural networks in ctr prediction. In: Proceedings of the 13th ACM conference on recommender systems, RecSys '19. Association for Computing Machinery, New York, NY, USA, pp 187–195

Li H, Xu M, Wang Y, Wei H, Qu H (2021) A visual analytics approach to facilitate the proctoring of online exams. Association for Computing Machinery, New York

Liu D, Xu P, Ren L (2019) Tpflow: progressive partition and multidimensional pattern extraction for large-scale spatio-temporal data analysis. IEEE Trans Vis Comput Graph 25(1):1–11

Malheiros M, Jennett C, Patel S, Brostoff S, Sasse MA (2012) Too close for comfort: a study of the effectiveness and acceptability of rich-media personalized advertising. In: Proceedings of the SIGCHI conference on human factors in computing systems, CHI '12. Association for Computing Machinery, New York, NY, USA, pp 579–588

Nguyen PH, Henkin R, Chen S, Andrienko N, Andrienko G, Thonnard O, Turkay C (2020) Vasabi: hierarchical user profiles for interactive visual user behaviour analytics. IEEE Trans Vis Comput Graph 26(1):77–86

O'Donnell K, Cramer H (2015) People's perceptions of personalized ads. In: Proceedings of the 24th international conference on World Wide Web, WWW '15 Companion. Association for Computing Machinery, New York, NY, USA, pp 1293–1298

Ogawa M, Ma K-L (2010) Software evolution storylines. In: Proceedings of the 5th international symposium on software visualization, SOFTVIS '10. Association for Computing Machinery, New York, NY, USA, pp 35–42

Peng D, Tian W, Lu B, Zhu M (2018) Dmnevis: a novel visual approach to explore evolution of dynamic multivariate network. In: 2018 IEEE international conference on systems, man, and cybernetics (SMC), pp 4304–4311

Peng D, Tian W, Zhu M, Ren Y, Lin X (2020) Targetingvis: visual exploration and analysis of targeted advertising data. J Vis 23(6):1113–1127

Syakur MA, Khotimah BK, Rochman EMS, Satoto BD (2018) Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In: IOP conference series: materials science and engineering, vol 336, pp 012017

Webga K, Lu A (2015) Discovery of rating fraud with real-time streaming visual analytics. In: 2015 IEEE symposium on visualization for cyber security (VizSec), pp 1–8

Wei J, Shen Z, Sundaresan N, Ma K-L (2012) Visual cluster exploration of web clickstream data. In: 2012 IEEE conference on visual analytics science and technology (VAST), pp 3–12

Wu J, Guo Z, Wang Z, Xu Q, Wu Y (2020) Visual analytics of multivariate event sequence data in racquet sports. In: 2020 IEEE conference on visual analytics science and technology (VAST), pp 36–47

Xia M, Velumani RP, Wang Y, Qu H (2021) Qlens: visual analytics of multi-step problem-solving behaviors for improving question design. IEEE Trans Vis Comput Graph 27(2):870–880

Xie C, Chen W, Huang X, Hu Y, Barlowe S, Yang J (2014) Vaet: a visual analytics approach for e-transactions time-series. IEEE Trans Vis Comput Graph 20(12):1743–1752

Yan J, Liu N, Wang G, Zhang W, Jiang Y, Chen Z (2009) How much can behavioral targeting help online advertising? In: Proceedings of the 18th international conference on World Wide Web, WWW ’09. Association for Computing Machinery, New York, NY, USA, pp 261–270

Yu S, Yang D, Hao Y, Lian M (2020) Visual analysis of merchandise sales trend based on online transaction log. Int J Pattern Recognit Artif Intell 34(11):2059036

Yuan B, Hsia J-Y, Yang M-Y, Zhu H, Chang C-Y, Dong Z, Lin C-J (2019) Improving ad click prediction by considering non-displayed events. In: Proceedings of the 28th ACM international conference on information and knowledge management, CIKM ’19. Association for Computing Machinery, New York, NY, USA, pp 329–338

Zhao X, Wu Y, Lee DL, Cui W (2019) iforest: interpreting random forests via visual analytics. IEEE Trans Vis Comput Graph 25(1):407–416

Zhao Y, Luo X, Lin X, Wang H, Kui X, Zhou F, Wang J, Chen Y, Chen W (2020) Visual analytics for electromagnetic situation awareness in radio monitoring and management. IEEE Trans Vis Comput Graph 26(1):590–600

Zhou G, Mou N, Fan Y, Pi Q, Bian W, Zhou C, Zhu X, Gai K (2019) Deep interest evolution network for click-through rate prediction. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, no 01, pp 5941–5948