J Vis (2020) 23:1113–1127
https://doi.org/10.1007/s12650-020-00671-w

REGULAR PAPER

Di Peng · Wei Tian · Min Zhu · Yukun Ren · Xiaojian Lin · Mingzhao Li

TargetingVis: visual exploration and analysis of targeted advertising data

Received: 28 October 2019 / Revised: 11 May 2020 / Accepted: 22 May 2020 / Published online: 10 July 2020

Abstract Targeted advertising is a dominant form of online advertising. It considers advertisers' major concern of their customers, including the consumers' certain traits, interests and individual preferences. To promote the effectiveness of advertisement delivery, advertising analysts need to understand advertiser delivery behavior and problems in targeting structure. However, statistical methods cannot meet analytical requirements completely, and analysts have to spend a lot of time reading countless data reports. Concretely, there is no efficient tool accomplishing analysis tasks such as exploring targeting usage at different levels, discovering useful or abnormal targeting combination patterns, finding competition from user behavior. In this paper, we design and implement an interactive visual analytics system named TargetingVis to visualize targeted advertising delivery data to face the challenges. After conducting a detailed requirements analysis with the domain experts from Tencent Inc., we design TargetingVis with four linked views: a novel chord diagram for cross-level exploration of targeting relations, a view for delving into the analysis of targeting combination patterns, an auxiliary view for displaying data indicators and a view to help gain insights into the behavior of advertisers. Finally, we evaluate the usability and efficiency through experiments based on real-world datasets.

Keywords Targeted advertising · Visual analytics · Relations in hierarchical data · User behavior analysis

D. Peng · W. Tian · M. Zhu (✉)
College of Computer Science, Sichuan University, Chengdu, China
E-mail: zhumin@scu.edu.cn

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

W. Tian
E-mail: tianwei@stu.scu.edu.cn

Y. Ren · X. Lin
Tencent Inc., Shenzhen, China
E-mail: rickyren@tencent.com

X. Lin
E-mail: karllin@tencent.com

M. Li
RMIT University, Melbourne, Australia
E-mail: mingzhao.li@rmit.edu.au

1 Introduction

With the rapid development of the Internet, online social networking platforms have become the gateway to the Internet for billions of users. These platforms accumulate rich user data which enable themselves to deliver online advertising efficiently. Nowadays, targeted advertising, which is based on products' or people's certain traits, accounts for the largest share of the online advertising market. This phenomenon indicates that advertisers are willing to spend more money on targeted advertising. However, some advertisers fail to use targeted advertising effectively due to incorrect audience identification. Hojjat et al. (2017) suggested that online advertising companies wish to know about their audiences (e.g., how many unique individuals were exposed to the advertisement) and understand how well their targeted strategies perform (e.g., how many times, on average, each audience was exposed to the advertisement). Rich audiences' usage data traced by companies enable them to achieve the above goals. To attract and satisfy more advertisers, domain experts need a better way to analyze the targeting comprehensively and effectively.

In targeted advertising, advertisement delivery is often driven by inferencing, that is, the process of using big data to infer beliefs about demographics and preferences (Bergemann and Bonatti 2011). When an advertiser delivers one advertisement, they could choose a series of targetings (just like labels) to search audiences for the advertisement. Generally, a targeted advertising platform can generate millions of advertisement delivery data per day. Thus, analysts face difficulties including the complexity of targeting structure, a large number of indicators and the difference of advertiser behavior in targeted advertising platforms. Failure to provide a satisfying experience would easily lead to a quick loss of advertisers. Understanding advertiser delivery behavior and problems in targeting structure to promote the effectiveness of advertisement delivery is crucial to the research and development of targeted advertising platforms. However, traditional data analysis methods depend on statistics usually fail by automatic analysis in exploring advertiser delivery behavior (Hojjat et al. 2017). The behavior is complex while a normal behavior may be abnormal in some specific scenarios. Without analysts' participation, it is difficult to define a specific behavior automatically. Besides, analysts need to spend a lot of time reading data reports generated by statistical methods.

Due to the lack of efficient interactive exploration tools, it is quite difficult and time-consuming to complete tasks like cross-level analysis of targeting relations and combinations, e.g., advertising analysts want to know whether the targeting A of second level and targeting B of fourth level are often used together. Visual analytics can help analysts gain insights into targeting advertising data and replace reports lacking interaction. However, to the best of our knowledge, few visual analytics works have researched targeting relations and advertiser behavior patterns of advertisement delivery.

In light of the above, we introduce a highly interactive visual analytics system named TargetingVis for domain experts to visually analyze the targeted advertising data, and promote the effectiveness of advertisement delivery. The system includes the following views: Firstly, the Relation View which combines chord diagram and sunburst chart provides an overview of the targeting structure and co-occurrence information with different hierarchies. Secondly, the targeting combination patterns of advertisements is demonstrated in the Combination View to discover valuable advertisers behavior patterns by combining machine and human intelligence. Thirdly, to show a large number of indicators simultaneously, the Indicator View uses parallel coordinates to analyze the specific effects of each targeting combination pattern. Lastly, the Advertiser Portrait View is built on a grouped bar chart and a normal bar chart, which can be used to explore the overall situation using multiple indicators and dimensions progressively. It is also can be used to obtain original information of advertisements and advertisers. The major contributions of this paper are as follows:

- To the best of our knowledge, this is the first interactive visual analytics system to help advertising domain experts explore the behavior patterns of advertisers, discover how to make their advertising more effective and finally make reasonable data-driven decisions.
- We proposed novel visualization designs with four linked views to complete analysis interactively. Particularly, we combine the chord diagram and sunburst chart to the Relation View to discover relations in hierarchical data. It enables the display of relations and the switching between different data hierarchies.
- We present case studies and domain expert interview based on real-world targeted advertising datasets to verify the usability and effectiveness of the proposed system.

Fig. 1 The user interface of TargetingVis. **a** Control Panel allows users to adjust the structure of targeting or filter the data to change other linked views, **b** Relation View summarizes the relations in hierarchical data to discover targeting usage and disclose problems in the design of targeting structure, **c** Combination View shows targeting combination patterns generated by advertiser delivery behavior so that users can find useful combinations and avoid abnormal combinations, **d** Indicator View demonstrates the overview and detail indicators corresponding to the targeting combination pattern to help users determine the quality of a combination, **e** Advertiser Portrait View displays the concrete performance of advertiser delivery behavior in different dimensions and indicators

2 Related work

In this section, the most relevant studies are surveyed, covering three streams of research: targeted advertising data analysis in the advertising industry, visualization of relations in hierarchical data and visual analysis of behavior pattern exploration.

2.1 Targeted advertising

In order to make advertisements relevant and attractive to consumers, there are two main approaches by considering user traits, behavior and context in targeted advertising: contextual advertising (CA) and behavioral advertising (BA) (Malheiros et al. 2012; O'Donnell and Cramer 2015). CA shows relevant advertisements based on the content of the website that users viewed. For example, a user who had read news with keyword basketball might receive advertisements about basketball shoes. In contrast, BA traces users online behavior (e.g., visited websites and posted content) to predict user preferences of advertisements. For instance, a user who had shown his or her preference to movie Avengers on social media might receive advertisements about Marvel T-shirt. Comparing these two forms of targeted advertising, BA is more effective than CA (Jin et al. 2016). Malheiros et al. (2012) found that increasing personalization at a certain level could enhance the attractiveness of advertisement. Actually, the perceived advertisement quality of an advertisement is more likely to decrease if the context match is at a high level. Yan et al. (2009) showed that accurate monitoring of the advertisement click-through log collected from commercial search engines could improve the performance of online advertising.

Chen et al. (2009) presented a scalable method of behavioral targeting, based on a linear Poisson regression model that involves granular events as features, such as individual advertisement clicks and search queries. Moreover, Some researchers proposed semantics (Armano et al. 2011; Saia et al. 2015) or recommendation systems (Addis et al. 2010; Vargiu et al. 2013) to improve the advertising effectiveness, but these approaches have not used segmentation techniques to categorize users.

In summary, existing studies in the field of targeted advertising mainly focus on how to improve the accuracy of audience identification. To the best of our knowledge, there is no work focusing on optimizing targeting structure in targeted advertising platforms and exploring advertisers' behavior patterns.

2.2 Visualization of relations in hierarchical data

A lot of researchers in the field of visualization have paid attention to how to visualize hierarchical visualization and relationship visualization. And visualization of relations in hierarchical data is a link between them.

To show relations and hierarchies at once, it would be an ideal design to employ hierarchical edge bundles and tree traditional visualization together (Holten 2006). For example, Hofmann et al. (2017) adopted this approach by improving prior methods to help users explore the change over time and to present different commodity classes simultaneously. Vehlow et al. (2016) developed a technique to support the matrix visualization of relations and hierarchical group structure changing over time. Some radial diagrams have attracted the attention of researchers, and some recent work starts adding relations to diagrams to tackle the challenge that shows relations and hierarchies at once. Zhang et al. (2013) introduced a tool to obtain the medical patient information through the sunburst chart with relations to achieve a more efficient diagnostic conclusion. Meanwhile, the sunburst chart with relations has also been applied to the analysis of pesticide residue data (Chen et al. 2015). TreeNetViz (Gou and Zhang 2011) shows the hierarchical structure by radial, space-filling visualization, displays aggregated networks and reduces visual complexity by an edge bundling technique. ViSeq (Chen et al. 2018) offers a sequence view consisting of a three-level interactive graph including a chord diagram to facilitate users in exploring learning sequences in MOOCs.

For special data characteristics and analysis requirements, the existing visualization approaches are not suitable for analyzing the relations among targetings of advertising, nor is it efficient to display relations in hierarchical data of targeted advertising in a limited space. Inspired by prior work, we design a novel visual view by combining chord diagram and sunburst chart to fulfill the requirements.

2.3 Visual analysis of user behavior

In recent years, the visual analysis of behavior pattern exploration has been developing continuously and prosperously. Although few work is directly related to the focus of our research, some research still deserves attention. For instance, Blascheck et al. (2016) collected rich user data including transcripts, videos, eye movement data and interactive logs. Based on that, a new visual analytics approach of user behavior is presented. Using smartphone usage data, Lu et al. (2016) designed a tool to discover and understand user behavior patterns. Chen et al. (2019b) proposed RelationLines by collecting real-world data from a city with millions of citizens, including taxi GPS data, cell-base mobility data, mobile calling data, microblog data and POI data to support users to discover behavior patterns. Also, there is much related research using data generated by Internet users. VASABI (Nguyen et al. 2020) visualizes users behavior to identify fraudulent activities in the field of network security. For visual analysis of social network, Wang et al. (2016) visualized and analyzed users behavior through unsupervised clickstream clustering. And Chen et al. (2019a) presented a visual analytics system to explore user behavior and diffusion patterns. The visual analysis of online games, MOOCs and financial transactions also emerge in an endless stream (Chen et al. 2017, 2018; Yue et al. 2019).

Despite the fact that many visual analytics systems are available, as far as we know, no prior work can be directly applied to the analysis of targeted advertising data due to the innate data incompatibility. Consequently, under previous work as theoretical supports, we develop a novel visual analytics system to enable rational view designs and rich interactions.

3 Background

In this section, we first introduces the background of targeted advertising data. Thereafter, four analytical requirements are proposed by domain experts.

3.1 Data abstraction

The targeted advertising data mainly consist of two entities: targeting and advertiser. The relation between these two entities is that advertisers will choose a series of targetings to identify audiences according to their products when they deliver an advertisement.

Targetings in advertising systems generally have two indicators: frequency and cost. If the frequency and cost of targeting are high, it indicates that more advertisements use this targeting and more audiences are targeted by this targeting. Moreover, multi-hierarchy is an important character of targeting structure; for example, the gender is a targeting, but gender also has two sub-targetings: male and female. In TargetingVis, there are five different levels in targeting structure. Since there are many targetings in targeting structure, we classify all targetings into five categories such as demographics, user behaviors, devices and so on.

Advertisements advertisements are described by both indicators and non-quantitative attributes in TargetingVis. An advertisement is delivered by an advertiser who generally has industry information. We also record the targeting list of audiences, product type and location of displaying to describe an advertisement. Exactly, an advertisement contains four dimensions including flow, platform, industry and product type. Furthermore, an advertisement owns six quantitative indicators: exposure, click, cost, CTR (click through rate), CPC (cost per click) and eCPM (effective cost per mille).

Targeting combination patterns According to the targeting list of each advertisement, we can generate a number of targeting combination patterns. These patterns are described in TargetingVis with seven indicators: frequency, exposure, click, cost, CTR, CPC and eCPM. The CTR is directly proportional to the click and inversely proportional to the exposure. The higher the value is, the better the advertisement effectiveness is. The CPC determines how much advertisers need to pay when users click the advertisements. The eCPM is the most critical indicator for the quantitative evaluation of revenue, which is used to estimate the revenue of the platform by each thousand advertising exposures. In this research, as the definitions of frequency, exposure, click and cost are in the same way as usual, we only define CTR, CPC and eCPM of targeting combination patterns as follows:

$$\mathrm{CTR}_i = \frac{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.click}}{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.exposure}} \tag{1}$$

$$\mathrm{CPC}_i = \frac{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.cost}}{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.click}} \tag{2}$$

$$\mathrm{eCPM}_i = \frac{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.cost}}{\sum_{\mathrm{ad} \in A_i} \mathrm{ad.exposure}} \times 1000 \tag{3}$$

where i denotes the ith targeting combination pattern and A_i denotes the set of advertisements which matches the ith combination pattern.

3.2 Requirements analysis

Targeted advertising is a form of online advertising, and it is based on the consumers' demographic, purchase history, interests, surrounding environment and so on. Therefore, we get some comprehensive visualization requirements from many stakeholders. Also, they are our target users.

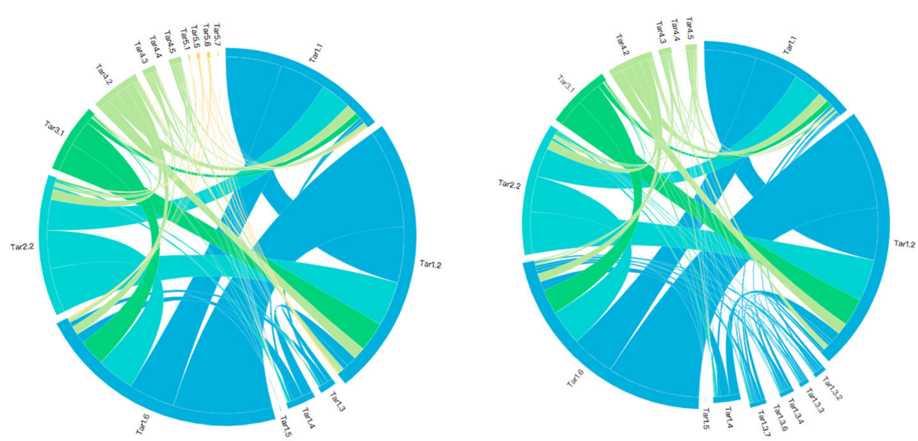
In this study, we cooperated with four experts from Tencent Inc. including an advertising algorithm expert (E1), a senior product manager (E2) and two advertising analysts (E3 and E4). It is their basic demand to optimize the targeting structure by observing the advertisement delivery record so as to improve the effectiveness of advertising delivery. E1 is an expert in advertising algorithms. He has been working in the Internet advertising industry for many years and hoping to find a better way to gain insights into advertiser delivery behavior. E2, who is a product manager with visualization research experience, expects to see more details that traditional methods fail to find through visual analytics. She needs a novel approach to help analysts provide a better service to advertisers. E3 and E4 are familiar with statistical methods to generate reports regularly. They would like a new way to improve the design of targeting structure. With the perspectives contributed by experts, during the past six months, we worked with our expert teams to come up with user requirements. The detailed list of requirements is as follows:

R1: comprehensive overview for targeting usage In general, advertising analysts analyze and report on targeting usage through data reports on a regular basis. The key aspects in the report include the frequency of targeting use, the increase or decrease in cost and some other indicators. Nevertheless, the data reports cannot display the relations between targetings, nor can they show the hierarchy of targeting. E1 said the statistical methods and reports lack interaction and it is burdensome to display results comprehensively. E3 and E4 would like a more convenient way to fully understand the advantages and disadvantages of targeting structure with flexible interactions such as select and filter.

R2: concise display and comparison of useful and abnormal targeting combination patterns When the advertisers deliver the advertisement, he will choose a series of targetings to form a certain targeting sequence pattern. Due to countless kinds of targeting combination patterns, it is difficult for advertising analysts to find meaningful or abnormal targeting combination patterns correctly and quickly. More importantly, different users may expect to see different targeting combination patterns (E3 and E4). For example, product managers in a particular industry would like to see whether a newly designed targeting is often used by advertisers and what targeting combination they usually use. Consequently, it is essential to design a visualization tool to help our users detect the popular targeting combination patterns and also compare different targeting strategies. Our experts also show their concerns about indicators mentioned in Sect. 3.1. Considering the fact that a combination contains multidimensional indicators, these indicators should be presented as thoroughly as possible in our system to provide insights into it.

R3: quick finding competition between advertisers E2 shows her interest in understanding whether there is a competitive relationship in the process of advertisement delivery. Her focus is on how to maximize return on investment of each advertiser. To achieve this goal, the targeted advertising platform must try to minimize the competition between advertisers. In nature, the behavior of advertisers is complex and inexplicable, and E1 believes that existing statistical methods are difficult to observe it from different perspectives. Only when analysts have a comprehensive understanding of a variety of advertiser delivery behavior, is it possible to find the problems and make interpretable decisions based on these results.

R4: rich interactions and features to simulate the advertising analysis process In the visual analytics system, flexible interactions and features design are fundamental. Users can gain insights into data through interactions such as filtering, sorting and other basic interactive operations. Advertising analysts mainly focus on seven indicators and four dimensions (Sect. 3.1) in targeting advertising data under different business scenarios. Also, these dimensions contain numerous subcategories. Constrained by the report length and form, some indicators and dimensional information of each advertisement are usually discarded in the data processing. Therefore, E3 and E4 wish they could complete the whole advertising analysis process like daily work by visual analytics and solve the problems that statistical methods fail to solve effectively. For example, they believe there are some abnormal behaviors. However, they could not define all abnormal behaviors in advance and could only observe the data report with great effort. Therefore, they expect to find such behaviors more intuitively and efficiently.



(a) (b)

Fig. 2 **a** The Relation View of initialization, **b** drilling down Tar1.3 in the Relation View and removing Category5 by the Control Panel

4 Visual design and feature

On the basis of the requirements proposed by us and experts (Sect. 3.2), we design and implement a visual analytics system consisting of five components, namely Control Panel (Fig. 1a), Relation View (Fig. 1b), Combination View (Fig. 1c), Indicator View (Fig. 1d) and Advertiser Portrait View (Fig. 1e). The Control Panel allows the user to switch between "targeting structure" and "filter" function. The tab of targeting structure is shown in Fig. 1a, while the tab of filter supports dimension selection and indicator filtering. To maintain a consistent visual perception, we put to use uniform colors matching for color encoding. The structure of the targeting is a tree. We call the sub-nodes of the root node as targeting categories because they are used to generate categories of targeting. Each targeting category is given a unique color, and all sub-targetings of the same targeting categories use the same color.

4.1 Relation View

The Relation View is the predominant view of our system, and it provides a compact overview of the targeting structure for users. Based on **R1**, this view can help users understand the frequency or cost of cross-level targetings and can also help users quickly discover the relation between targetings. More crucially, it is designed to concatenate other views that users can use to get an in-depth understanding of targeting usage and structure.

Conventionally, the chord diagram is a visualization approach to display the relationship between data in a matrix. The data of nodes are arranged radially along the circumference and linked by weighted arcs. In addition to demonstrating targeting relations at the same level, our design is supposed to display targetings relations at different levels. For this reason, after referring to a lot of literature, we find that the sunburst chart can be well combined with the chord diagram to meet the requirement (Fig. 1b).

The view is drawn from first level targetings (Fig. 2a), whose colors are determined by the categories to which they belong, and the color legend of categories is placed on the upper-right corner of this view. We arrange the targetings that belong to the same level of targetings together, and the targeting names are placed along the center of the arc to make it easier for users to find the targeting they are concerned about. The arc length represents the sum of co-occurrences of a certain targeting and other related associations, and the arc width represents the frequency or cost of the targeting. The width of ribbon represents co-occurrences of the two targetings. Although co-occurrence is equal, the color of ribbon depends on who at the ends of ribbon accounts for a larger proportion of its arc length. In this view, we can clearly compare the frequency or cost of targetings at different levels (or we can find out which targetings are generally used with a particular targeting) and how many times they are used together.

Several alternatives were considered, discussed and attempted by us, such as the node-link tree diagram or arc diagram. However, we regret to find that neither of these visualization approaches can meet the requirement well. Although the former approach can express relations in hierarchical data, our design could simultaneously encode two types of information (sum of co-occurrence, frequency or cost) by using the length and width of the arc, while the node-link tree diagram can only encode one type of information through the radius of a circle. When the user keeps drilling down to observe a three or deeper level fully expanding, the tree becomes complex and is less space-efficient, which can be circumvented by radial layout. More importantly, the node-link tree diagram cannot show relations between targetings commendably. As for the latter approach, the line width of this design is uniform and cannot be used to encode the co-occurrences of targeting. If we add hierarchical information into this design, it will make its connections seem messy. Moreover, it can only use the radius of a circle to encode information like the former. All things considered, the current design is chosen to achieve the requirement finally.

User interactions and functional features Some interactions and features are assigned to the Relation View to provide navigation during exploring process (**R4**). First, highlighting should be introduced to assist users in discovering the frequency, cost, co-occurrence and numerical information corresponding to the ends of a ribbon. The highlighting of our system can give users prompts through the tooltip and breadcrumb trail when mouse hovers as well as link the Combination View and the Advertiser Portrait View. In addition, our system allows users to lock highlighted state, making it easy to see the details of data exactly. Next, we need to mention the drilling down and drilling up, which are two operations that will be linked with other views. Users can press down a specific key and click on the arc or name of targeting they need to drill down, so that the targeting can be expanded to the next level. By pressing down another specific key and clicking on the arc or name of the sub-targeting, the targeting can shrink back to the previous level. Then, due to the problems of data itself and the limitation of space, users may find that some elements in this view are too small to interact and to observe details. Our system provides users with zooming-in, zooming-out and full-screen features. Users can interact with this view flexibly to fulfill their needs. Finally, our system supports users to choose encoding of the arc width by frequency or cost of targeting.

4.2 Combination View

To realize **R2**, we display the information of targeting combination patterns in the Combination View (Fig. 1c). Taking into account the effectiveness and practicability of this view, experts (E2, E3 and E4) suggest that only the top 50 targeting combination patterns need to be displayed on the ground of their daily experience. Therefore, our system presents a specified number of targeting combination patterns in a form similar to the heatmap. Simply put, this view provides users with a summarization of the targeting combination pattern that facilitates users handy comparison. Users can identify the combination easily. It provides a series of features and supports a more detailed visual analysis of specific combination patterns.

As Fig. 1c illustrates, at the top of this view is the name of targeting corresponding to the Relation View one by one. A row of bars stands for a combination pattern, and a column of bars stands for a kind of targeting. These combinations are distributed from high to low according to the user-specified ranking indicator. If a targeting is used, the colored rectangle in which it is located lights up; otherwise, the colored rectangle in which it is located is transparent. With these rational designs, users can easily get answers to some questions (e.g., which are commonly used together, which are very popular) with the specific ranking indicator and the number of combinations.

Standard visual approaches such as the pie chart and stacked bar chart were also considered before we chose the current design. However, we found that these alternative charts have to draw multiple analogous graphs when displaying combination patterns, and cannot directly compare the combinations. What is more, it would induce visual clutter in this way. Meanwhile, we use visual channels similar to mosaic plot to encode information, which can support users to find patterns of interest or exception easily.

User interactions and functional features For the sake of understanding the targeting combination patterns, our system provides adequate interactions and features for users (**R4**). Like the Relation View, highlighting is an essential interaction. When the mouse hovers over a row of bars in the Combination View or clicks directly to lock it, all the other views will be linked to this operation. The Indicator View is a little bit special as it will highlight a corresponding line and gradually conceal other irrelevant lines. When users move their eyes to the control area of this view, they will find that the necessary features in the process of exploring several views have been provided. Users can choose a variety of targeting combination patterns that they focus on. After clicking the button of multi-selection, other views will also change to respond to this operation. If users want to find a specific targeting combination pattern, they can click the search button. Here, the system provides users with two kinds of logical computation for search. The first is called "and" operation, which means that the display of targeting combination patterns must contain some given targetings. The second operation is called "or" operation, which means that the display of targeting combination patterns must contain at least one of some given targetings. Additionally, users can also conduct competition analysis. After switching tabs in the Control Panel, selecting a specific industry and clicking the button, all views will be reloaded. With the help of the Advertiser Portrait View, users can find out which advertisers in some industries are competing with the target industry. In this way, product managers can work with operators to optimize existing targetings and make advertisements more thriving. What is more, our system enable users to select a sorting indicator and specify the number of targeting combination patterns. It is worth mentioning that all of the above interactions or features are designed for **R2**.

4.3 Indicator View

The Indicator View is mainly used to assist the analysis of Combination View, which is supplemented by data indicators. As we mentioned earlier, there are seven indicators in targeting advertising that can be used to evaluate the quality of an advertisement (**R2**). Thus, we can also use these indicators to evaluate the specific performance of a targeting combination pattern.

We can find that in Fig. 1d, multidimensional data are rendered through parallel coordinates. Unlike traditional designs, we adopt a distinctive method related to the category of targeting combination patterns to implement color encoding of lines. In brief, we add the number of targetings belonging to the same category of one row in the Combination View and then use the color of the category with the largest number of targetings as the color of the line in parallel coordinates. In this way, users can quickly discover which categories are in common use. The category of a line is calculated by Algorithm 1:

Algorithm 1 Finding the category of a line

Input: The line corresponding to one row in the Combination View, i;
Output: The category corresponding to a line with the largest number of targetings in a targeting combination pattern in the Combination View, $LineCategory_i$;

```
1: Initialize the sum of each category, such that the k − th category as Category_k = 0;
2: for the j − th column in i do
3:     Get the category of Targeting_ij as m;
4:     if m = k then
5:         Category_k + +;
6:     end if
7: end for
8: Compute the category of a line LineCategory_i = max(Category_k);
```

We experimented with several candidate designs for illustrating the targeting combination pattern. The radar chart and scatterplots (Chen et al. 2014) were two options that we had considered before. Due to the limitation of screen space and data, these two visualization approaches usually present too dense lines or spots to comprehend. Notably, the radar chart can display multivariate data well, and our target users are familiar with it. However, we need to effectively support the comparison of more than 50 target combination patterns, and the radar chart is difficult to meet the requirement. The Indicator View based on parallel coordinates could make better use of limited screen space. And colored lines correspond to each line of the Combination View one by one to help users quickly discover the combinations they are looking for.

User interactions and functional features As before, we can highlight or click on a line in the Indicator View to affect the presentation of other views. Automatic linking can be triggered among different views by brushing. Our system allows users to drag and swap axes of parallel coordinates to better explore data. In particular, it allows users to select a targeting combination pattern in the Combination View. Once a combination pattern is selected, users can switch the Indicator View from overview to detail and observe the distribution of multiple advertisements in a targeting combination pattern on parallel coordinates (**R4**).

4.4 Advertiser Portrait View

By default, the display form of Advertiser Portrait View is a grouped bar chart (Fig. 1e). At the same time, users are allowed to adjust this view to a normal bar chart to analyze the overall situation. This view provides seven indicators and four dimensions for users to switch to gain insights into the data (**R3**). The original information of advertisements and advertisers can also be obtained in this view. Moreover, sorting by cost is used.

The same color coding as the Relation View and Combination View is also used in the grouped bar chart. And the color coding of the bar chart follows the formula of the Indicator View. As shown in Fig. 1e, the x-axis represents the name of objects contained in the specific dimension selected by the user and the y-axis represents the value.

Initially, we intend to reuse the Combination View to the Advertiser Portrait View. In previous designs, a row corresponds to an item in the indicators or dimensions. Besides, we adopted the colored rectangle with different shades of color to indicate the number of advertisement delivery. However, this design may make it difficult for users to distinguish the number of advertisement delivery. Currently, we support users to select the bar chart or grouped bar chart freely, as well as view details of advertisements and advertisers.

User interactions and functional features There is no doubt that highlighting is also supported in this view. However, other views will be linked only in the grouped bar chart. In addition to switching charts, this view also supports users in selecting the indicator or dimension that they need to analyze deeply (**R4**). In the grouped bar chart, users can switch the dimension or brush to focus a bar. In the bar chart, users can switch indicators and dimensions at the same time. If there are too many bars, our system will automatically provide users with the scrollbar to slide and explore, providing users with a better experience.

5 Evaluation

This section presents case studies of three real scenarios completed by the experts with our assistance and valuable interviews. Because of the confidentiality of commercial data, we have carried out data processing simply. For example, Tar1.1 represents the first targeting of the first category.

5.1 Exploring the accessibility of system

E1 believes that many users will be interested in targeting usage and combination patterns (**R1** and **R2**); he intends to learn how to use the system according to the usual analysis process (**R4**) and looks forward to seeing both of them.

In this scenario, E1 gives attention to the Relation View (Fig. 2a) and focuses on the analysis within individual view. He discovers that targetings of Category1 take up most of the space, which means that most advertisers choose this category of targetings in their advertisement delivery. Among these, Tar1.2 and Tar1.6 are especially prominent because they not only have more co-occurrences (the arc length is longer), but also more frequency than other targetings (the arc width is wider). To compare these two targetings, he hovers over Tar1.2 and Tar1.6 to highlight them. The sum of co-occurrences of Tar1.2 is greater than that of Tar1.6, but there is no much difference between these two in terms of frequency; that is to say, Tar1.6 is the targeting that is most frequently used independently.

Then, he continues to highlight Tar1.6. With the help of linking, he begins to look at the targeting combination patterns that contain Tar1.6 in the Combination View. He is somewhat surprised to find that the combination used independently by Tar1.6 ranks second in the Combination View (Fig. 3). He switches to different indicators in the Combination View and then chooses exposure as the sorting indicator. He finds that the new combination pattern is still in the second place. However, after selecting cost as the indicator, this combination pattern falls to the fifth place. To explain this phenomenon, he locks the combination and then checks the specific values through the Indicator View (Fig. 4a). To his surprise, the CPC and eCPM of this combination are very low. For advertisers, the choice of this combination pattern can achieve a better result at a lower cost. Out of curiosity, he switches the dimension to the industry in the Advertiser Portrait View and finds that advertisers in the industry Ind15 prefer to use this combination pattern (Fig. 4b). After that, we verify that the occurrence of this situation is in line with data reports.

5.2 Discovering abnormal combination patterns

E1 brings together other experts to discuss his findings. Due to the lack of handy tools, examining meaningful, especially abnormal targeting combination patterns thoroughly has been a challenging task for E3 and E4 (**R2**). With our system, E3 wonders whether some targetings are misused.

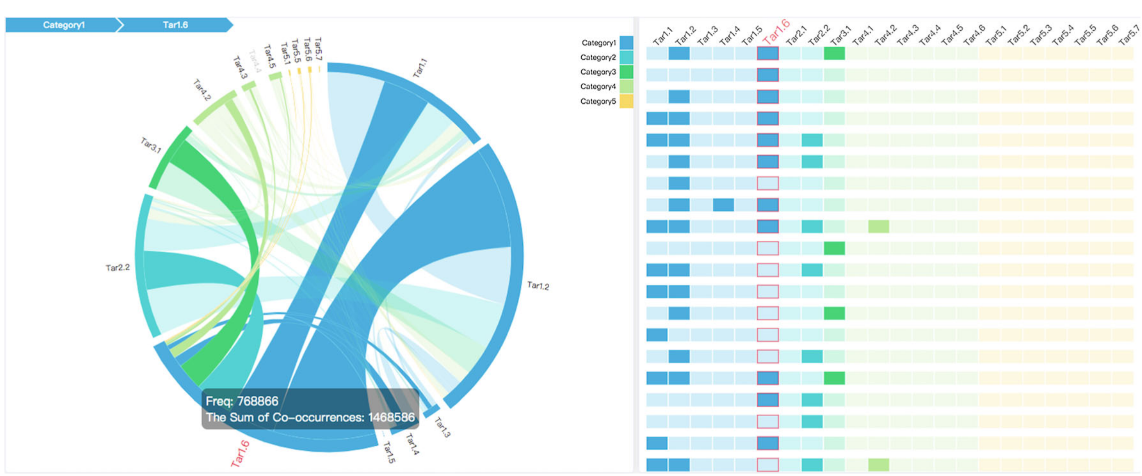
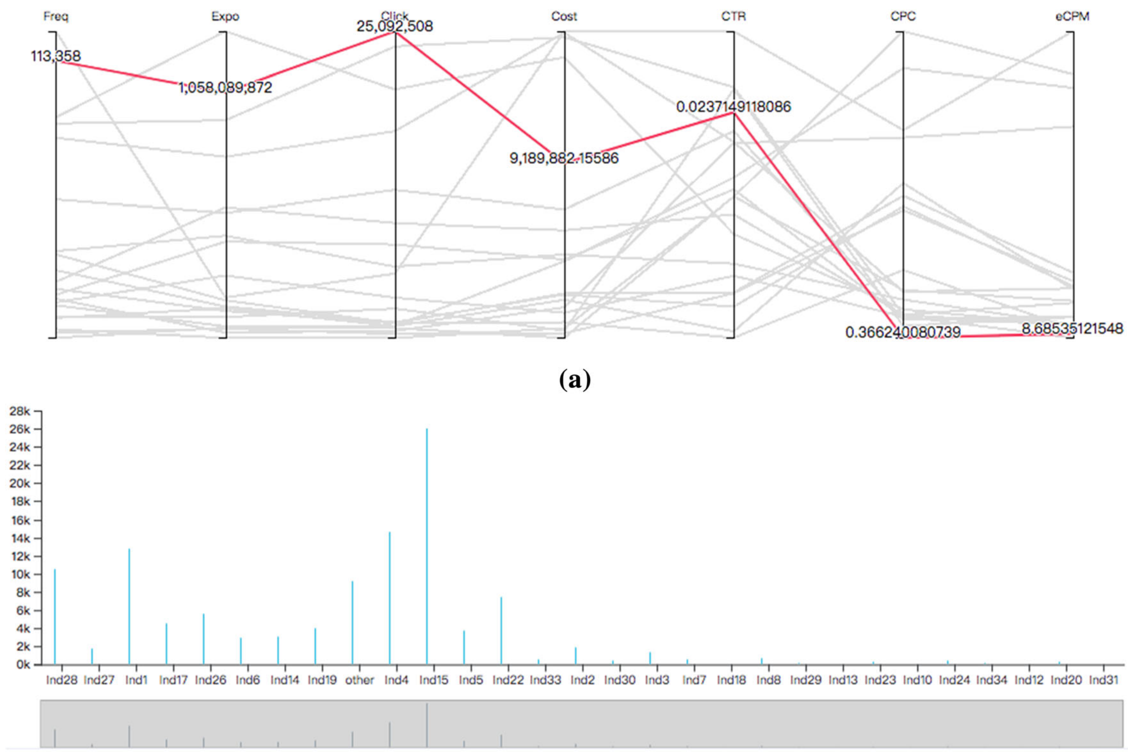


Fig. 3 Hovering over Tar1.6 in the Relation View. The Combination View shows all the targeting combination patterns containing Tar1.6



(a)

(b)

Fig. 4 The corresponding indicators and distribution of the combination pattern that contains Tar1.6 only. **a** This combination has high frequency, exposure and cost as well as low CPC and eCPM, **b** in the dimension of the industry, advertisement delivery shows an uneven phenomenon

In order to explore whether sub-targetings belonging to Tar1.3 have been misused, E3 delves into a further study about the usage of these targetings. He first drills down the Tar1.3, and then the Relation View and Combination View show the corresponding targeting usage and combination patterns. Because it does not matter whether these targetings are used with Category5, he removes all targetings under this category in the Control Panel, as shown in Fig. 2b. From this, he observes that Tar1.3.1 has never been used and Tar1.3.7 is used frequently, which reflects the problem to other experts and designers.

To get rid of the interference caused by the advertisements of less cost, he filters through the Control Panel, leaving only advertisements whose cost is more than 1000. He turns his attention to the Combination View. Among the targeting combination patterns that used the frequency sorting in the top 20, he fails to find out a combination pattern that used the sub-targetings of Tar1.3. As a result, he operates this view to show the top 50 targeting combinations. Two unusual combination patterns catch his attention. According to the original design, Tar1.3.2, Tar1.3.4 and Tar1.3.7 should not have been used simultaneously. After searching, he confirms that only the two patterns contain these three targetings. In the Indicator View, he finds that the two combination patterns are actually not commonly used. Next, he selects them (Fig. 5) and switches to detail view in Indicator View, hoping to know which advertisers are using such combination patterns and check the specific situation of advertisement delivery. In the Advertiser Portrait View, it shows that they will only be delivered on Site7. Finally, he looks at the information of the advertisement and the advertiser in this view (Fig. 6) and provides feedback to the operator responsible for relevant advertisers.

5.3 Analysis of competition between advertisers

One of the experts, E2, who has been working in the field of Internet advertising for years, mentions that she is interested in the competitive relationship between advertisers (**R3**).

As the head of Ind29 industry, she switches the dimension of the Advertiser Portrait View to industry and then filters the industry in the filter box in the Control Panel, so that the system can only display the industry-related information she is responsible for. At this point, there is only one industry left in the

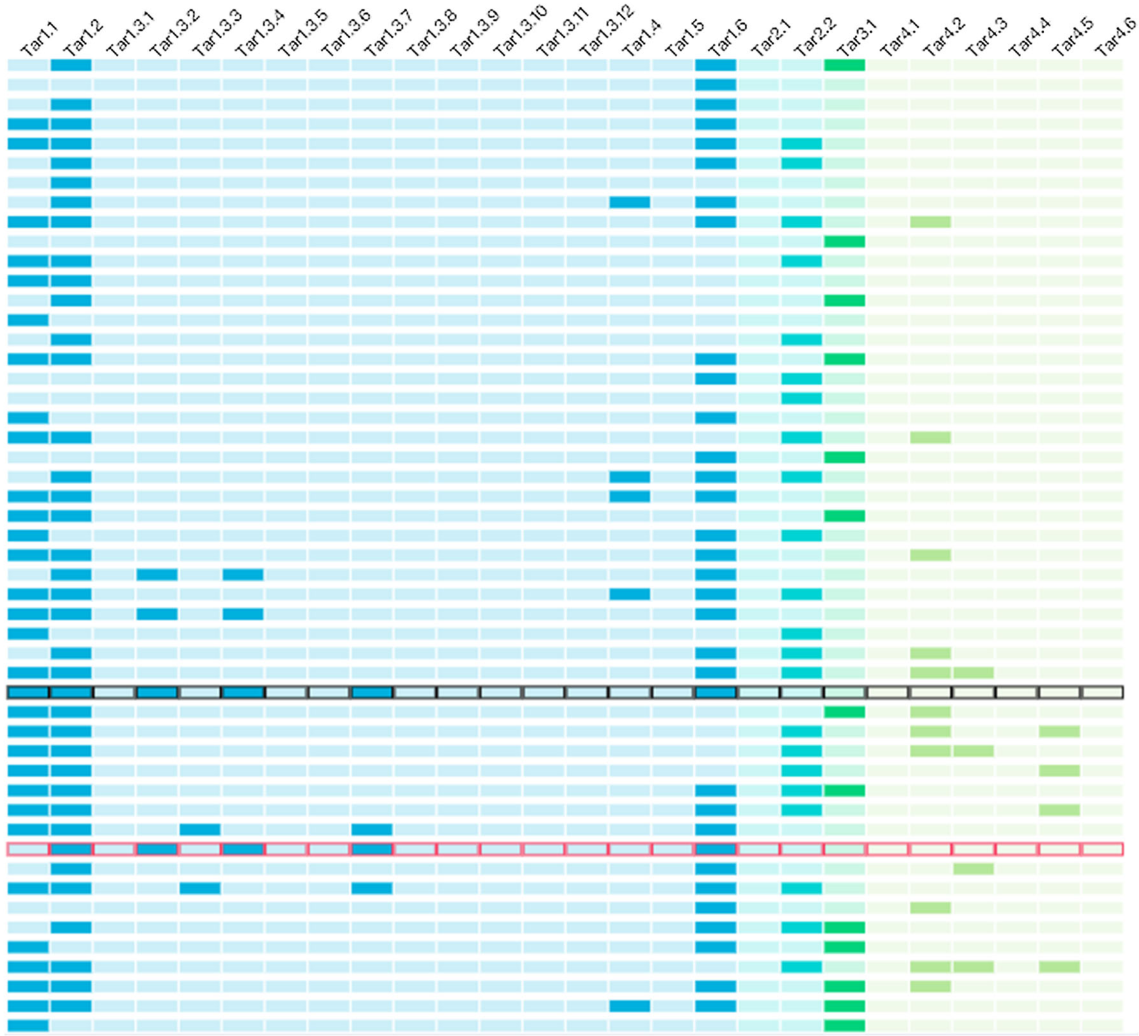


Fig. 5 Selecting multiple targeting combination patterns that contain Tar 1.3.2, Tar 1.3.4 and Tar 1.3.7

Ad ID	Advertiser ID
2137**4258	617**437
2127**7348	791**408
2188**1258	333**451
2140**6848	535**451
2188**4158	524**458

Fig. 6 The advertisements and advertisers information. After E3 selecting two combinations that contain Tar 1.3.2, Tar 1.3.4 and Tar 1.3.7, through the feature of checking the original information provided by the Advertiser Portrait View, the content he cares about is presented

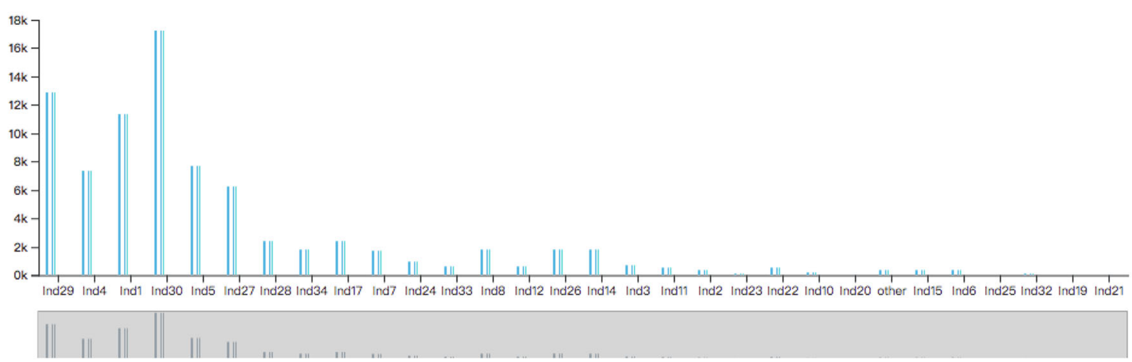
Advertiser Portrait View. In this industry, Tar1.1 and Tar1.2 have become two most widely used targetings. Besides, Tar4.2 is on the top 5 in terms of the frequency of use, and the arc length of other targetings belonging to the same category of Category4 is also significantly longer. According to E2's work experience, she thinks this makes sense because this category is closely related to the type of business of Ind29.

Then, in the Combination View, she sorts the targeting combination patterns by cost and finds the first combination pattern. After selecting it and clicking the button of competitive analysis, the Advertiser Portrait View is updated displaying other industries that compete with Ind29 (Fig. 7a). For ease of observation, E2 converts the view into the bar chart (Fig. 7b) and found that Ind29's biggest competitor is Ind 30. E2 also concludes that Ind1, Ind5 and Ind4 are potential competitors for Ind29. After switching the indicator to cost, it is found that Ind29 has the largest cost in this combination pattern, followed by Ind4. Although Ind30 has a lot of advertising, it does not have much cost. Aided by this finding from the Advertiser Portrait View, E2 determines the goal of targeting optimization in the future.

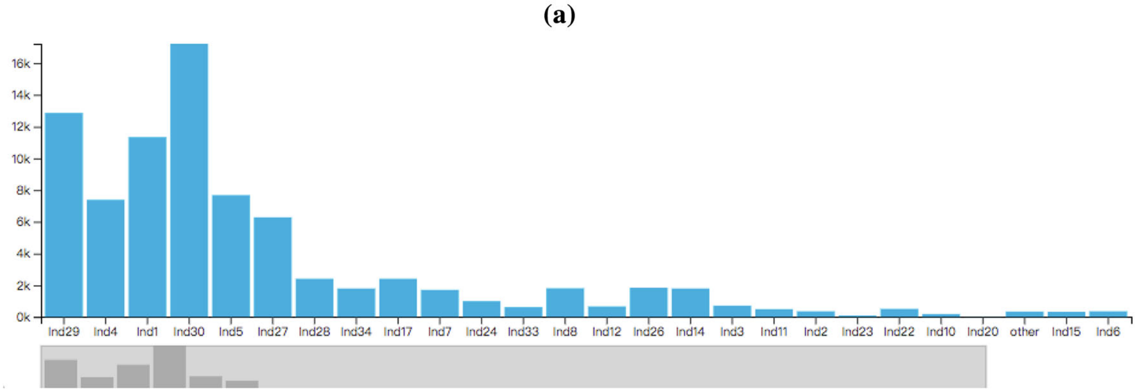
5.4 Interview and discussion

In the interview, we invite the aforementioned four experts and two other advertising analysts (E5 and E6) to share their impressions of the system such as the benefits and shortcomings of TargetingVis. Experts believe that TargetingVis is easy to use and to evaluate problems in the design of targeting structure. Moreover, it can explain the advertiser delivery behavior efficiently and complete analysis tasks comprehensively. Straightforward visualizations and rich interactions along with appropriate features bring them a novel analytical experience significantly. Also, this system can help them glean information and understand how to fundamentally improve the effectiveness of advertisement delivery (Sect. 5.1). In particular, E2 is keen on the feature of competitive analysis, which can enable users to grasp the competition among advertisers and increase the return on investment of advertisers (Sect. 5.3). E3 found this system can assist him to determine whether a combination is valuable or abnormal to ensure advertisers use correctly (Sect. 5.2).

However, there are still some aspects where the system can be improved. E1 and E5 pointed out that a possible improvement might be supporting the analysis of time-series data. They expected to dive into the evolution of advertiser delivery behavior over time, rather than just statistics. Furthermore, E4 and E6 mentioned that they expected to save all applied operations so that the views could be restored to previous states quickly without repeating the same steps. With this feature, it will be easier to save time and show



(a)



(b)

Fig. 7 **a** In the Combination View, the cost is selected as the ranking indicator, and after competitive analysis of the first pattern, the Advertiser Portrait View demonstrates the corresponding bars, **b** to facilitate observation, E2 switches the display mode of this view to the normal bar chart results to others. According to E3, in some special cases such as when there are more than 100 targetings are expanded concurrently, the Relation View and the linked Combination View might cause visual clutter and make users confused. One possible improvement could be to aggregate sub-targetings that belong to the same parent targeting.

At last, we discuss with the experts about the feasibility of analyzing similar data in other fields. They believe the system is extremely extensible. With some modification of the available attributes and features, it can be applied to visualize other data with the same structure. All of them agree that the accessibility of the system is closely related to the knowledge background of the target user. Therefore, they recommend us to add more prompts of the system to help users get started faster.

6 Conclusion and future work

Exploring and understanding targeting usage, targeting combination patterns, and identifying competition between advertisers in massive data is a challenging task especially when the data have multiple indicators and dimensions. In this work, with the assistance of domain experts, we propose an interactive visual analytics system named TargetingVis to help users delve into the targeted advertising data to promote the effectiveness of advertising delivery. Our system provides multiple coordinated views, which support a series of easy-to-use interactions and features to fulfill all requirements while resolving existing problems. Case studies and interviews with target users prove that our system is effective and useful.

In the future, we plan to improve our system to support time-series data analysis to help users explore the evolution of advertiser delivery behavior. We will add features to support storing and reading operations in each step that has been taken into consideration so that the results of associated visualizations can be shared among users or recover previous analysis easily. Moreover, we will find the rational way to ensure complete tasks while increasing the innovation of views and reducing the deficiency of design.

Acknowledgements We would like to thank the domain experts from Tencent Inc. and anonymous reviewers for the valuable suggestions. We also thank Dongxin Wei, Yanan Li, Xiaoxiao Xiong, Jing Liang and Qinglai He for their participation. This work was supported by Tencent Rhino-Bird Focused Research Program (18H0487).

References

Addis A, Armano G, Giuliani A, Vargiu E (2010) A recommender system based on a generic contextual advertising approach. In: Proceedings of the 15th IEEE symposium on computers and communications, pp 859–861. https://doi.org/10.1109/ISCC.2010.5546548

Armano G, Giuliani A, Vargiu E (2011) Studying the impact of text summarization on contextual advertising. In: 2011 database and expert systems applications, pp 172–176. https://doi.org/10.1109/DEXA.2011.78

Bergemann D, Bonatti A (2011) Targeting in advertising markets: implications for offline versus online media. RAND J Econ 42(3):417–443. https://doi.org/10.1111/j.1756-2171.2011.00143.x

Blascheck T, Beck F, Baltes S, Ertl T, Weiskopf D (2016) Visual analysis and coding of data-rich user behavior. In: 2016 IEEE conference on visual analytics science and technology, pp 141–150. https://doi.org/10.1109/VAST.2016.7883520

Chen Y, Pavlov D, Canny JF (2009) Large-scale behavioral targeting. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining, pp 209–218. https://doi.org/10.1145/1557019.1557048

Chen H, Chen W, Mei H, Liu Z, Zhou K, Chen W, Gu W, Ma K (2014) Visual abstraction and exploration of multi-class scatterplots. IEEE Trans Vis Comput Graph 20(12):1683–1692. https://doi.org/10.1109/TVCG.2014.2346594

Chen Y, Zhang X, Feng Y, Liang J, Chen H (2015) Sunburst with ordered nodes based on hierarchical clustering: a visual analyzing method for associated hierarchical pesticide residue data. J Vis 18(2):237–254. https://doi.org/10.1007/s12650-014-0269-3

Chen W, Lu J, Kong D, Liu Z, Shen Y, Chen Y, He J, Liu S, Qi Y, Wu Y (2017) GameLifeVis: visual analysis of behavior evolutions in multiplayer online games. J Vis 20(3):651–665. https://doi.org/10.1007/s12650-016-0416-0

Chen Q, Yue X, Plantaz X, Chen Y, Shi C, Pong TC, Qu H (2018) ViSeq: visual analytics of learning sequence in massive open online courses. IEEE Trans Vis Comput Graph. https://doi.org/10.1109/TVCG.2018.2872961

Chen S, Chen S, Wang Z, Liang J, Wu Y, Yuan X (2019a) D-Map+: interactive visual analysis and exploration of ego-centric and event-centric information diffusion patterns in social media. ACM TIST 10(1):11:1–11:26

Chen W, Xia J, Wang X, Wang Y, Chen J, Chang L (2019b) Relationlines: visual reasoning of egocentric relations from heterogeneous urban data. ACM TIST 10(1):2:1–2:21. https://doi.org/10.1145/3200766

Gou L, Zhang XL (2011) Treenetviz: revealing patterns of networks over tree structures. IEEE Trans Vis Comput Graph 17(12):2449–2458. https://doi.org/10.1109/TVCG.2011.247

Hofmann J, Größler M, Rubio-Sánchez M, Pichler P, Lehmann DJ (2017) Visual exploration of global trade networks with time-dependent and weighted hierarchical edge bundles on GPU. Comput Graph Forum 36(3):273–282. https://doi.org/10.1111/cgf.13186

Hojjat SA, Turner J, Cetintas S, Yang J (2017) A unified framework for the scheduling of guaranteed targeted display advertising under reach and frequency requirements. Oper Res 65(2):289–313. https://doi.org/10.1287/opre.2016.1567
Holten D (2006) Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. IEEE Trans Vis Comput Graph 12(5):741–748. https://doi.org/10.1109/TVCG.2006.147
Jin Y, Seipp K, Duval E, Verbert K (2016) Go with the flow: effects of transparency and user control on targeted advertising using flow charts. In: Proceedings of the international working conference on advanced visual interfaces, pp 68–75. https://doi.org/10.1145/2909132.2909269
Lu M, Meng D, Peng Y, Li Y, Zhao Y, Fan X, Zhou F (2016) A visual analytics approach for exploring individual behaviors in smartphone usage data. In: 2016 IEEE pacific visualization symposium, pp 224–228. https://doi.org/10.1109/PACIFICVIS.2016.7465275
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: CHI conference on human factors in computing systems, pp 579–588. https://doi.org/10.1145/2207676.2207758
Nguyen PH, Henkin R, Chen S, Andrienko NV, Andrienko GL, Thonnard O, Turkay C (2020) VASABI: hierarchical user profiles for interactive visual user behaviour analytics. IEEE Trans Vis Comput Graph 26(1):77–86. https://doi.org/10.1109/TVCG.2019.2934609
O'Donnell K, Cramer H (2015) People's perceptions of personalized ads. In: Proceedings of the 24th international conference on world wide web companion, pp 1293–1298. https://doi.org/10.1145/2740908.2742003
Saia R, Boratto L, Carta S (2015) A latent semantic pattern recognition strategy for an untrivial targeted advertising. In: 2015 IEEE international congress on big data, pp 491–498. https://doi.org/10.1109/BigDataCongress.2015.78
Vargiu E, Giuliani A, Armano G (2013) Improving contextual advertising by adopting collaborative filtering. TWEB 7(3):13:1–13:22. https://doi.org/10.1145/2516633.2516635
Vehlow C, Beck F, Weiskopf D (2016) Visualizing dynamic hierarchies in graph sequences. IEEE Trans Vis Comput Graph 22(10):2343–2357. https://doi.org/10.1109/TVCG.2015.2507595
Wang G, Zhang X, Tang S, Zheng H, Zhao BY (2016) Unsupervised clickstream clustering for user behavior analysis. In: Proceedings of the 2016 CHI conference on human factors in computing systems, pp 225–236. https://doi.org/10.1145/2858036.2858107
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, pp 261–270. https://doi.org/10.1145/1526709.1526745
Yue X, Shu X, Zhu X, Du X, Yu Z, Papadopoulos D, Liu S (2019) Bitextract: interactive visualization for extracting bitcoin exchange intelligence. IEEE Trans Vis Comput Graph 25(1):162–171. https://doi.org/10.1109/TVCG.2018.2864814
Zhang Z, Wang B, Ahmed F, Ramakrishnan IV, Zhao R, Viccellio A, Mueller K (2013) The five Ws for information visualization with application to healthcare informatics. IEEE Trans Vis Comput Graph 19(11):1895–1910. https://doi.org/10.1109/TVCG.2013.89