PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses

Qing Chen, Yuanzhe Chen, Dongyu Liu, Conglei Shi, Yingcai Wu, *Member, IEEE*,
and Huamin Qu, *Member, IEEE*

Abstract—Massive open online courses (MOOCs) aim to facilitate open-access and massive-participation education. These courses have attracted millions of learners recently. At present, most MOOC platforms record the web log data of learner interactions with course videos. Such large amounts of multivariate data pose a new challenge in terms of analyzing online learning behaviors. Previous studies have mainly focused on the aggregate behaviors of learners from a summative view; however, few attempts have been made to conduct a detailed analysis of such behaviors. To determine complex learning patterns in MOOC video interactions, this paper introduces a comprehensive visualization system called PeakVizor. This system enables course instructors and education experts to analyze the "peaks" or the video segments that generate numerous clickstreams. The system features three views at different levels: the overview with glyphs to display valuable statistics regarding the peaks detected; the flow view to present spatio-temporal information regarding the peaks; and the correlation view to show the correlation between different learner groups and the peaks. Case studies and interviews conducted with domain experts have demonstrated the usefulness and effectiveness of PeakVizor, and new findings about learning behaviors in MOOC platforms have been reported.

Index Terms—MOOC, online education, visual analytics, clickstream data

✦

1 INTRODUCTION

MASSIVE open online courses (MOOCs) have attracted considerable public attention over the past few years [1], as such media offer a wide range of courses with high temporal and spatial flexibility. Three major MOOC platforms (edX, Coursera and Udacity) were launched in 2012, and since then, Coursera alone has attracted 11.8 million users as of March 2015 [2].

Several MOOC platforms record web log data, including not only basic information about the learners, but also their interaction activities with videos (e.g., clickstreams, click actions on different videos at various times) and in forums (e.g., forum posts and interactions with one another). Thus, a massive amount of data is generated for analyzing learning behaviors at a refined granularity [3]. Several analytics systems have been built to visualize such clickstream data [4], [5], [6]. However, these systems mainly focus on the aggregate behaviors of learners.

We have been collaborating with MOOC instructors in our university since January 2014, and a visual analytic system called VisMOOC [7] has been developed to help assist these instructors in analyzing online learning behaviors based on video clickstream data from Coursera. By observing how instructors use VisMOOC, we note that they spend a considerable amount of time on certain regions with a short period of sudden increase in click actions. We hereafter refer to such periods as the video clickstream "peaks". To clarify why instructors focus on peaks, we conducted multiple discussion sessions with the instructors. We determined that in general, peaks are critical to helping instructors understand how a group of users actively reacted to course videos. For example, if an unexpectedly high occurence of pausing or backward seek activities is observed at a certain segment, then this segment is probably difficult or confusing and thus requires additional time watching and studying. However, if students watch a certain segment without any additional click behavior, then their feelings on that segment are difficult to interpret. Therefore, all of the analytical tasks in this paper focus on analyzing such peaks.

Analyzing the peaks of clickstream data from MOOC videos to clarify e-learning behaviors poses two challenges. First, the multi- attribute clickstream data for each course are large in scale owing to the considerable number of learners watching MOOC videos. Even after filtering by the peaks, millions of click actions remain attributed to each course. The emerging urgent problem involves extracting valuable information from complex and considerable amounts of data. Following several rounds of interviews with the domain experts, we summarize the major issues of concern as follows: a) when analyzing a single peak, the instructors care about the basic statistics of different click actions in addition to those who contributes to the peaks. Therefore, necessary data include the demographic information of these contributors and the temporal information of these clicks. b) Another concern raised by the instructors is that abnormal clicks by learners may generate a strange pattern in the peaks of the clickstream data. Such abnormality may be misleading

- Q. Chen, Y. Chen, D. Liu, and H. Qu are with the Hong Kong University of Science and Technology, Hong Kong.
  E-mail: {qchenah, ychench, dliuae, huamin}@cse.ust.hk.
- C. Shi is with the IBM T.J. Watson Research Center, New York, NY.
  E-mail: conglei.shi@us.ibm.com.
- Y. Wu is with the Zhejiang University, Hangzhou, Zhejinag, China.
  E-mail: ycwu@zju.edu.cn.

when analyzing whether certain video segments are problematic or interesting to the majority learners. c) Instructors can improve as online teachers if they can identify typical learner groups and the correlation among learners in different peaks. Therefore, a comprehensive visualization system should be designed to reveal underlying patterns in complex big data to course instructors and education experts. Such an attempt is challenging, but the results are valuable.

Another challenge is that such a visualization system must be easy to use. An analytical framework [8] has been proposed recently for researchers who have some programming knowledge; these researchers adapted a data model to different MOOC platforms using this framework and integrated the model with mainstream analytic tools, such as MATLAB. However, system end users, general course instructors, and education experts have minimal or even no background knowledge in computer science or visualization; thus, they must be provided with intuitive visual designs that can help alleviate learning burden. As per interviews with the instructors and the education experts we have worked with, a clear and simple system is essential. To this end, the visual designs for specific tasks must be examined carefully by end users.

To address these challenges, we present a comprehensive visualization system called PeakVizor that enables course instructors and education experts to analyze complex online learning patterns in MOOC video interactions. Our system features three views at different levels: a) an overview in which glyphs are used to show basic statistics of the peaks, including the number of clicks, peak duration, and distribution of users; b) a flow view that presents the geographical and temporal origins of click actions as well as the spatio-temporal location of peaks; and c) a correlation view in which parallel coordinates are integrated with a choropleth map to reveal the correlation between different learner groups and the peaks. Furthermore, we carefully design smooth interaction transitions across different views so that end users can explore complex patterns more easily by combining the three views. We follow a user-centered design process and involve domain experts in every stage of iterative development. To evaluate our final system, we conducted case studies with several end users; the results further proved the usefulness and effectiveness of our system.

The main contributions of this paper are as follows:

- A comprehensive visualization system which integrates well-established visualization techniques and several novel visual designs to investigate clickstream peaks.
- A study of the temporal and spatial distributions of clicks in relation to the peaks as well as of the geographic and behavioral distribution of users who contribute to these clicks.
- A novel glyph design to show the multiple attributes of a peak so that users can easily identify specific peaks and their singularity either in the overview or in other co-analysis views.
- Case studies are conducted with real datasets to help instructors and education analysts gain new insights into online learning behaviors.

2 RELATED WORK

In this section, we first summarize existing studies on e-learning analysis. Then, we discuss current works on visual analytics and visualization techniques for MOOCs. Finally, we present several studies that focus on clickstream visualization.

2.1 E-Learning Analysis

Many studies have been conducted on big data analysis of online education. Some studies analyze log data on learning behaviors, such as student access and activity patterns [9], [10], and forum interaction [11]. Student access and activity patterns include what (the type of material), how much (the percentage of the material), how long (how many times), and when (in real time) with respect to viewed content, as well as to dropout behavior [12]. Research on forum activity investigates posting behaviors and interactions among students, forum structure, the sentiment analysis of post content, and topic trends. Other studies have examined learner performance [3], [13], learner demographics [14], [15], [16], and the correlation of these two factors with learner web log data [17]. Current performance measurements rely mainly on the grades from quizzes, assignments, and exams. Meanwhile, many other factors such as due dates [18] may also affect learner performance. Moreover, demographic information covers diverse areas, such as geographic location [19], age group, educational background [20], and objectives for taking a course [21].

With the availability of data on more courses, researchers can conduct analyses across courses. For example, the sequence of emergent courses is studied in [22]. Efforts have also been exerted to determine the social connection and interaction among students [23] as well as adaptive personal support [24]. Most of the studies discussed in this section have helped improve e-learning effectiveness through analyzing various domains, either from a pedagogy mindset, or from the perspective of data mining and machine learning. However, few researchers have addressed e-learning issues from a visual perspective.

2.2 MOOC Visualization

The use of visual aids to explore data has become increasingly helpful in determining patterns, especially given large datasets. In this section, we present recent work on MOOC data visualization.

Coffrin et al. [25] used a series of basic charts, such as bar charts and line charts, to help instructors understand learner behaviors. In [26], user trajectories are first clustered into groups and then visualized by composition. Xu et al. [27] proposed a visual analytic tool that uses scatterplot to analyze user profiles and performances. Wortman and Rheingans [28] demonstrated a set of visualization techniques, including graph layouts, parallel coordinates, color mapping, and interactive selection, for users to explore the data and find student event sequence patterns. Recently, MoocViz [8] was established as a data analytics platform which contains a cross-platform framework, allowing community members to embed additional modules into the system. VisMOOC [7], an interactive visual analytics system with a novel seek diagram design is proposed or analyzing video clickstreams.

TABLE 1
Overview of the Courses Information

Course	Platform	#Events	#Weeks	#Videos	#Watchers	Video Avg Len	Video Max Len	Video Min Len
NCH	Coursera	1,204,947	4	17	11,061	11:33	20:27	3:37
JAVA	Edx	3,197,422	10	122	18,832	5:07	11:26	0:31

In summary, the works mentioned in this section not only use standard graphs (e.g., scatter plots, bar charts, line charts, and pie charts) but also innovative visualizations (e.g., node graphs, 3D diagrams, seek diagrams [7], and parallel coordinates [28]). Moreover, interactive techniques have been explored, including sorting, filtering, zooming, and clustering. Although such approaches enable multi-aspect analysis, they are particularly useful for analyzing aggregate behaviors, thus preventing instructors and education analysts from drawing additional conclusions. Hence, a detailed analytics system must be developed for users to comprehensively understand learning behaviors on MOOC platforms.

2.3 Clickstream Visualization

Web log data include multifaceted user activity information; thus, various methods have been proposed to visualize different types of event data, such as news event streams and web clickstreams [29]. Clickstreams are a typical type of web log data that include navigation paths and interactions with web content. Previous work concentrated on general video engagement [30] and e-commerce, such as webpage browsing behavior and clickstream sequences in online shopping [31]. For instance, Agnuair et al. [4] visualized a user-generated clickstream dataset to facilitate intensive understanding of the website visitors and predict user engagement by video streams. Mimic [5] presented a novel application that analyzes micro-interactions to assist interaction designers in improving the usability of web designs. In the MOOC scenario, a few recent attempts have been made to visualize clickstreams and help analysts understand how learners watch course videos. VisMOOC [7] uses a content-based design to show how clickstreams correlate with video content. In [6], area charts were employed to show clickstream statistics collected by MOOCs, and interesting patterns were identified. Inspired by these state-of-the-art studies, we have developed several novel designs to visually represent clickstream data and enable users to analyze them extensively.

2.4 Glyph Visualization

Glyphs, are important graphical entities, that have long been used in the visualization field. A well-designed glyph-based visualization process strongly benefits end users because this small independent entity conveys multiple data attributes through visual channels, such as position, size (length, area, and volume), shape, color, angle, orientation, curvature, and dynamics (motion speed and direction). Therefore, glyphs are constantly used to visualize complex, and multivariate datasets. Research on glyph-based visualization is vast, and a number of surveys have been conducted [32], [33], [34], [35]. Researchers have

summarized many facets of glyph-based design, including technical frameworks, classifications of the visual mapping and layout options of glyphs, discussions on mapping bias in terms of perception, proximity, and grouping, as well as evaluations of different glyphs for various analytic tasks. Despite the many advantages of glyph-based design, it has limitations that may be grouped into two major problems. One is the accuracy problem, which is a result of the high dependence of glyph information on the human visual perceptual system; the other is the scalability problem, which is derived from the difficulty, or even impossibility, of displaying hundreds of glyphs on a single screen. We are inspired by the previous studies mentioned in this section to carefully design several glyph visualization alternatives for end users.

3 PROBLEM CHARACTERIZATION

In this section, we first introduce the data extracted from the databases of two MOOCs and the automatic peak selection method used. Subsequently, we summarize the analytic tasks identified from our interviews with two MOOC instructors who offered MOOCs on the Coursera and edX platforms and an education analyst with a MOOC research background.

3.1 Data Abstraction and Peak Detection

We extracted learner profile and learner web log data from the databases of Coursera and edX. The generated learner profile included the nationality, grade and other information regarding the learners. Web log data included clickstreams on videos and posts in forums. Video clickstream data include different types of click actions such as play, pause, and seek. Table 1 shows the general statistics for two courses we used later for the case studies.

The raw data of clickstream data were noisy; thus, we cleaned them prior to embedding them into the system. Then, a peak detection algorithm was applied. Peak detection algorithms have been proposed for various analytical tasks, such as biomedical engineering, computer vision, and signal processing [36], [37], [38]. Other peak detection methods have also been proposed such as social media anomaly detection [39]. In the present study, we adopted the method proposed in [6]; this method was also used to detect peaks in MOOC clickstream data. It was originally utilized to detect events on Twitter [40] and was extended by Kim et al. [6]. The detection pipeline follows three major steps. First, the clickstream data of each video are binned into a histogram, in our case, by seconds. Then, the histogram is smoothed by a kernel-based smoother. Finally, weighted moving average and variance are calculated by a sliding window to detect
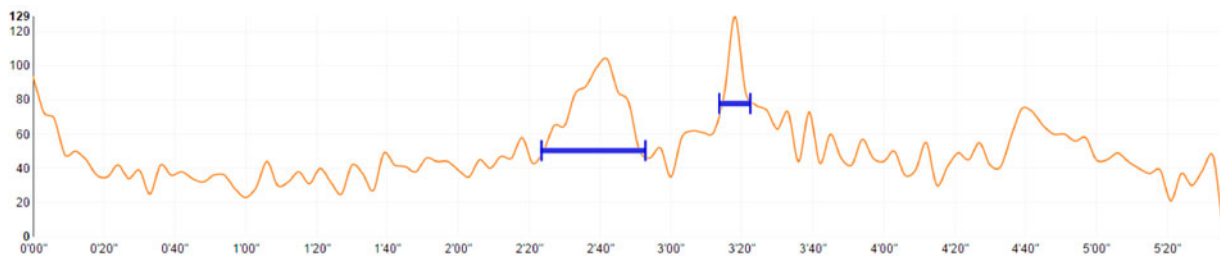
Fig. 1. Peak detection example: The orange line indicates the histogram after clickstream data are binned and the histogram is smoothed, whereas the blue line refers to the detected peaks in this histogram. The detected peaks usually correspond to the regions with large click numbers.

unusually large numbers of clicks in the histogram; this number is defined and can be detected using Eqn.(1):

$$\left| \frac{C_i - mean}{meandev} \right| > Th \qquad (1)$$

where C_i is the ith bin of the histogram, and $C_i > C_{i-1}$. *Mean* and *meandev* are the weighted mean and variance of all the previous bins of current bin i, respectively; and *Th* is the detection threshold. In our work, the parameters, such as the sliding window length, the detection threshold, and the weight function for the weighted mean and variance were manually selected to best fit the time scale of our analysis.

An example of the detected peaks is shown in Fig. 1. The orange line indicates the histogram generated by binning and smoothing the clickstream data, and the blue line shows the peaks detected in this histogram. In general, the detected peaks correspond to the regions with a sudden click increase in the clickstream. More specifically, as shown in Eqn. (1), the detection criterion is based on the weighted mean and variance. Therefore, the shape of the detected peaks is affected by the historical mean and variance. One possible situation is that less steady clickstreams will make the variance larger, which leads that the peaks following

the fluctuated clickstreams should be much sharper to be detected. For example, in Fig. 2, three peaks are detected while the one above (E) is not considered as a peak because the clickstreams before it are quite unsteady and fluctuate a lot. The running time for the peak detection algorithm is O (N) where N is the number of bins in the histogram. After peaks are detected, they are then manually checked by the end users and confirmed to be accurate.

We also extracted several valuable learner statistics directly from clickstream and the forum data; these statistics are as shown in Table 2. In Section 4, we illustrate these attributes further.

3.2 Task Analysis

The research problem was characterized through our collaborations with domain experts. Since early 2013, our university has been offering courses on MOOC platforms, such as Coursera and edX; over the past two years, several instructors who have taught MOOCs and e-learning education experts have been provided us with feedback and advice on problems they wish to study.

We followed the user-centered design process proposed in [41] throughout the task analysis and prototyping stages. This process involved three experts, namely, two university
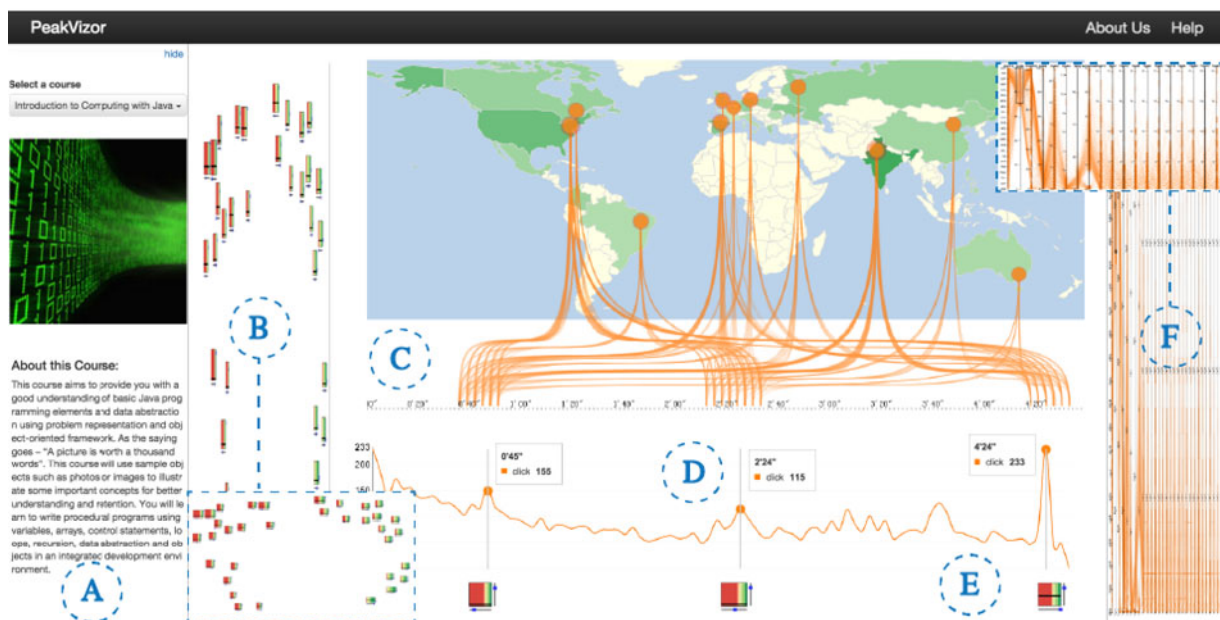

Fig. 2. PeakVizor system. (A) Basic information of the selected course from the course list. (B) The overview lays out all the peaks as glyphs according to the learning behaviors. (C) During the analysis of a selected peak in a video, the clickstreams flow from various countries to the corresponding spots on the video timeline. (D) A click event chart shows the changes in the click actions along the timeline. (E) Each glyph on the bottom represents a peak in the video. (F) The correlation view aims to analyze the correlation between different learner groups and peaks.

professors who offer their own courses on MOOC platforms and an education analyst who has studied e-learning related issues and has worked with MOOC instructors. The three domain experts had also collaborated with us previously. Moreover, we conducted several field studies in which we observed three experts as they explored our previous system, VisMOOC. We then interviewed them, asked for their feedback on the system, and inquired about the problems they encountered in online teaching. The experts mentioned analytic tasks based on popular research topics in their areas of expertise as well. We were able to obtain data on learner interaction data with videos on MOOC platforms, which differs from traditional face-to-face teaching, as well as on the frequent interaction of play, pause, and seek events that can indicate active participation by learners. When certain parts of a video generate considerable attention and interaction from the learners, the resultant data can reflect peaks in the video timeline. Based on the pilot interviews and given their previous experience with MOOCs, the end users were interested in participating our study and volunteered to analyze the peaks in the clickstream data.

Subsequently, we met regularly with the domain experts to obtain constant feedback on a series of prototypes of our new system and to determine further requirements. After additional rounds of interviews, we gained an improved understanding of the problems, and summarized major requirements as follows. The requirements are based on the feedback from the experts, as well as the knowledge gleaned from the literature review.

T.1 What are the general statistics for peaks?
   To obtain valuable information from the detected peaks, instructors must know the general peak statistics, such as the number of peaks, peak duration, and peak timestamp. The peak timestamp indicates where the peak appears in the video and when the video is released. An overview that summarizes all the peaks is therefore beneficial as it can provide users with a clear first impression of the peaks.

T.2 What is the correlation between learner groups and peaks?
   Owing to the open access to MOOC platforms, learners worldwide can interact with course videos, thereby generating millions of various clickstream records. However, different people may have varying goals for taking courses. Even individuals with the same goal may exhibit different learning behaviors, and these dissimilar behaviors result in various clicking patterns. For example, some learners click on most of the peaks detected, whereas others click only on a very small number of these peaks. Thus, learners can be grouped according to their clicking behaviors. Furthermore, end users can identify the similarities between peaks by exploring the correlation between learner groups and peaks. Furthermore, they can and determine whether or not different learner groups report specific demographics and grading and viewing patterns.

T.3 When and where do the clicks for each peak originate?
   Instructors are particularly interested in the spatio-temporal in- formation derived from learner clickstreams; such information deepens the understanding of dynamic learning behaviors. During pilot interviews, we used an

animation to show clickstreams over time on a world map. Interesting patterns were revealed in the process. However, animation itself is not a suitable approach to handle analytic tasks. The current study agrees with previous studies, such as [42]; given the significant cognitive load of animations, an animated dataset may induce several simultaneous changes that complicate comparisons and render them inefficient. Therefore, the domain experts require an appropriate method to reduce cognitive load and help analyze spatio-temporal behaviors.

T.4 Do abnormal peaks exist and who causes these peaks?
   In some cases, a peak is induced by a very small number of learners without any reasonable intention or even by a system error. Such actions may result in abnormal peaks, which do not represent the general regions of interest (ROIs) of learners; therefore, we should consider these peaks to be "outliers" in a visual analysis. However, certain abnormal peaks can still be investigated if those particular learners are expending actual time and effort on such peaks to understand certain concepts and gain knowledge. Automatic methods cannot identify abnormal peaks; thus, an intuitive design that can visually highlight such peaks and the learners who cause them is valuable for detailed analysis.

3.3 Design Requirements
The design process adopted for this work draws from the nested model in [41]. The system is human-centered given that we conducted interviews with end users at different design and implementation stages; consequently, we made adjustments according to their requirements. Furthermore, we identified the following key design rationales for designing PeakVizor.

R.1 Multi-scale and dual-facet exploration.
   An overview can provide instructors a clear first impression of all the peaks. Instructors are likely to determine interesting patterns and particular information efficiently through visual representations. As a result, instructors can conduct detailed analysis of their ROIs. Given the multivariate nature of clickstream data, different views should be presented to enable end users to analyze the peaks from different perspectives. Coordinated analysis is also important to domain experts; such a process facilitates simultaneous iterative exploration via different perspectives.

R.2 Intuitive visual design.
   Intuitive representation is an essential element of numerous visualization systems, especially when end users are not yet adept in visualization. Familiar visual metaphors can reduce the cognitive burden and enhance understanding. In the developed system, the design in the flow view features metaphors that are drawn from daily living and are thus easy to accept and understand. This feature was validated by the instructors and education analysts interviewed. Thus, intuitive visual designs support visual analytics by considerably exploiting user intuition and experience.

R.3 Consistent encoding.
   The consistent use of visual encoding techniques across different data dimensions and scales can reduce visual

complexity by preserving the mental map of a user to enhance understanding and simplify the exploration process. According to object constancy and change blindness theories, abrupt visual design changes should be avoided [41]. Therefore, a consistent color scheme and design should be applied to facilitate the intuitive understanding of learner information and behaviors.

R.4 Visual clutter reduction.

Although analyzing large and complex amounts of data is the key to performing important analytic tasks, visual clutter problems are aggravated by increases in the amount and complexity of data. The count of click-stream events and the number of learners can be massive; for example, the JAVA course (Table 1) logs more than 3 million click events and approximately 19,000 learners. Thus, we must consider the scalability issue when designing visualization systems.

4   VISUALIZATION DESIGN

In this section, we first provide an overview of the PeakVizor system. Then, we illustrate the visual encoding process of each view in detail and discuss alternative designs. Finally, we discuss the interactive exploration of PeakVizor.

4.1   System Overview

Fig. 2 presents an overview of the PeakVizor system. The main interface consists of three views: the overview includes a peak graph with a glyph as a node and provides course instructors and education analysts with a general idea of the peak attributes and similarities across different peaks. In the flow view, end users can explore the spatio-temporal information of each peak. To facilitate a detailed analysis of the correlation between learner groups and peaks, the correlation view highlights whether or not learners form particular groups according to their clickstream behaviors and how learners differ in terms of country, grade, and viewing attributes; these attributes include click activeness, loyalty, delay, review, dropout time, and forum activeness. In addition, PeakVizor supports multiple interactions, such as filtering and selecting in different views. Upon selecting a course from the pop-up window, a column appears on the left side of the screen that shows video screenshots corresponding to each peak and its basic information. In each view, users can use the snapshot function on the top right corner of the screen to record their analysis at any time. Combining all the views enables users to comprehensively analyze learning behaviors in peaks at different levels of granularity.

4.2   Visual Encoding

To solve the analytic tasks summarized from the previous interviews with domain experts, we propose several novel designs and different views in accordance with the design rationales in Section 3.3.

4.2.1   Glyph Design

According to design requirement R.1, end users should be provided with an overview of all the peaks. Furthermore, the visual design should reveal the important features of each individual peak; to this end, a glyph design is
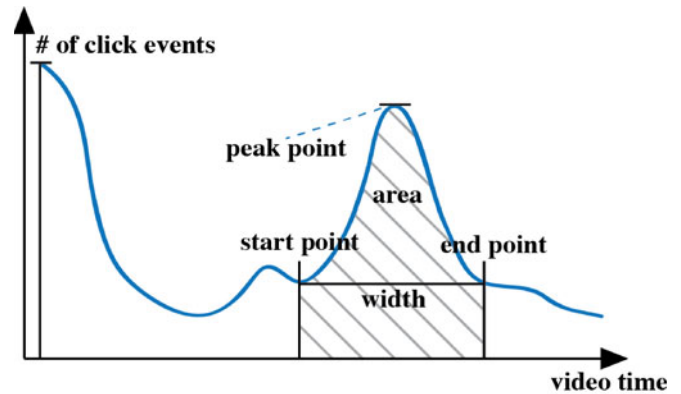

Fig. 3. Click event chart and major peak features. A click event chart shows how the number of click events change along the video timeline. The starting point, area, and width are encoded in the glyph as well.

proposed. This glyph aims to help end users identify different peaks and their singularities. Following our discussions with course instructors and education analysts, we identify the key information that must be encoded into the glyph.

The major features of a peak are shown in Fig. 3; the blue line shows the number of clicks along the video timeline. Specifically, the three key features of a peak are depicted, and the width of a peak is defined as the peak duration. According to [6], different peak widths may refer to different viewing behaviors. The area under a peak indicates the total number of click events observed during the peak duration, whereas the starting point indicates the onset of the peak. In other words, this point indicates when the peak appears in the video. The instructors also point out two other valuable peak features: grading distribution and the statistic that detects potentially abnormal peaks caused by a small number of learners. We then design our glyph to encode these features.

Visual encoding parameters, including size, inner line position, and color, should be combined to convey the key attributes of peaks. We use a rectangle to represent each glyph: the height and the width encode the total number of clickstreams for the chosen click type and peak duration, respectively. The two timelines that surround the glyph are designed to help end users identify the peaks quickly. The dots on the horizontal and vertical lines indicate the onset of the peak in the corresponding video and when the video is released, respectively. For example, the dots on the two surrounding timelines in the design sketch presented in Fig. 4a indicate that the peak appears near the beginning of the video and that the video is released at the beginning of the course.

As discussed in Section 3.2, one of the end users' main aims is to discover any abnormal peaks (T.4). We then encode peak anomaly as follows: first, we identify the learners whose clicks exceed the threshold, th. Then, we calculate the contribution of these particular learners to the total number of click actions in the peak. Finally, we exhibit the percentage as the position of the inner line in a glyph. To choose a proper threshold, we test various values and determine whether or not we can identify the predefined potentially abnormal peaks. We then set the value of th to 10 for our case studies. For example, the inner line in 4(a) shows that learners whose clicks exceed the threshold contribute to approximately half of the total number of clicks in the corresponding peak. Thus, the potentially abnormal peaks
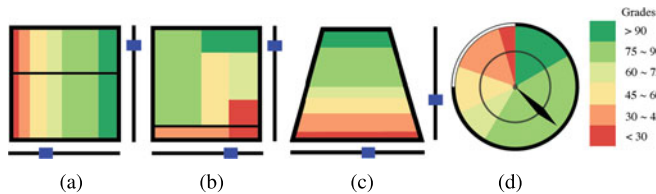
Fig. 4. Glyph design helps end users identify specific peaks and their sin-gularities. (a) Glyph design used in PeakVizor. (b) Treemap-based glyph design. (c) Glyph design that uses shapes to represent the anomaly in peaks. (d) Clock glyph design.

Fig. 5. Overview. The graph is used to show the similarity among peaks. Each node represented by a glyph corresponds to a peak. The layout of the graph is calculated by multidimensional scaling based on peak simi-larity. A short distance indicates high similarity between two peaks.

can be detected based on the position of the inner line, and education analysts can analyze the reasons behind the gen-eration of these peaks further through other views.

To maximize the glyph space, we use stripped bars to encode the grade distribution (Fig. 4a). Green color indi-cates a high score, whereas red represents as a failing score. We categorize the grades into six different segments for the case study. Nevertheless, end users can freely adjust the number of segments and the grading criteria. Fig. 4a depicts the sample distribution of learner grades. Approximately half of the learners received grades higher than 75, and very few learners reported grades lower than 30.

From the perspective of a human perception, we take into account certain design considerations during the estab-lishment of glyph prototypes. One is glyph size. In setting glyph length, we should decide whether to use the total count of the clicks for a period of the same length or to nor-malize the count by assigning the same width to the peak given that the number of clicks is highly related to peak duration. However, the use of a constant width for each glyph may suggest that high glyphs correspond to high click density. This scenario can mislead end users. Another concern involves the application of either a uniform length or a scaled length to the timelines when the height and width of the peaks change. The trade-offs between aes-thetics and precision should be determined for the given tasks. Therefore, we conduct interviews with experts regarding glyph design and have finally chosen to utilize the scaled length according to interviewee feedback.

Several design alternatives have also been implemented and evaluated during the prototyping stage.

- Initially, we overlay a treemap (Fig. 4b) inside the rectangle to encode grade distribution. Treemaps have been well established as a visual representation method; thus, we believed that this approach would be received well by the end users. Although they regarded the treemap as aesthetically pleasing, it complicated the comparison of the percentages of learners marked with the same color at different peaks because the layout varies. The situation depicted in Fig. 4b complicates the analysis of some tasks, such as determining the percentage of learners who have received grades higher than 75. This repre-sentation may also lead to misunderstandings because the grading data do not convey hierarchical information.
- Inspired by DICON [43], we use the shape visual channel to encode the singularity of each peak (Fig. 4c). First, we calculate the average and standard

deviation of the number of times a learner clicks on each peak and draw the distribution for each peak. However, the use of the shape visual channel, in which the "peakness" cue indicates the uniqueness of a certain peak, is too complicated for end users to understand.

- Another option is the clock glyph shown in Fig. 4d. The concept originates from the pie chart which is a typical visual representation for distri-bution. We examine the inner and outer circles to indicate the singularity of each peak and to show the week in which the video was released. The pointer of the clock indicates the point at which the peak appears in the video. Although end users can easily understand the encoding process and agree that the design is aesthetically pleasing, the circular representation is mostly used for peri-odic time, such as a 24-hour time cycle, and may thus induce confusion.

4.2.2 Overview Layout

To provide users with the general peak statistics and the rela-tion among them (T.1), we use the overview layout to present all the peaks in a course so that users can have an initial global glimpse of all the peaks before selecting certain peaks for detailed exploration (R.1). A similarity-based layout is applied in the overview (Fig. 5). We calculate the similarities among different peaks and utilize Multidimensional Scaling (MDS) [44] to compute glyph layout on this basis. MDS is commonly used to cluster the similarities of individual cases in a dataset. The time complexity for MDS is $O(N^3)$ where N is the number of peaks. To eliminate the overlapping among nodes, we employed a well-established method [45].

To measure the similarity between two peaks, we adopted a simple and intuitive approach that involves calculating the number of shared learners. However, this measurement is affected by the total number of clicks of each peak, thereby generating misleading results. To avoid this problem, we computed a feature vector for each peak based on the histogram of users general statis-tics (e.g., demographic information and performance) and then applied the histogram intersection kernel (HIK) [46] to calculate the distance between two peaks. HIK is a histogram-based kernel for measuring feature similari-ties; it is denoted by k_{HI} and is defined as follows:

$$k_{HI}(p_i, p_j) = \sum_{k=1}^{d} \min(p_{ik}, p_{jk}), \tag{2}$$

where $p_i = [D_i, G_i]$ is the feature vector of ith peak. D_i, G_i are the demographic and grade histograms of the learners, respectively, and p_{ik} is the kth bin of the feature vector. k_{HI} increases if learner statistics are closely distributed between two peaks. The time complexity of HIK similarity is $O(N^2 * m)$ where N is the number of peaks and m is the length of the feature vector. The total running time of the MDS layout, and the HIK similarity calculation is less than 0.1 s, which is efficient enough for real-time interaction.

4.2.3   Flow View

This view shows the spatio-temporal distribution of the peaks detected. The choropleth map on top indicates the distribution of click actions by country. The size of the outer circle on the map reflects the number of clicks in each country, whereas the size of the inner circle represents the number of learners by nation. Learners are geocoded through their IP addresses, and we ensure that the flows terminate in capital cities at the country level. The name of a country is shown when the pointer is hovered over its corresponding location on the map. If users click on the country, then the city-level choropleth map is exhibited, with small circles that correspond to the location of each city. At the city level, the flows represent the click actions from different cities in the country specified by an end user. The clickstreams from different countries flow to various timespots on the timeline below. Users may choose from two types of timelines in the system according to dissimilar exploration purposes. The first choice is video level; when a user selects a glyph in the overview and directs it to the flow view, a click event chart with the corresponding peak is shown below the video timeline (Fig. 2). The flow map above indicates the origin of the click actions of the peak in this video. The click event chart (Fig. 3) shows the changes in the number of click events along the video timeline to identify when the peak appears in the video. This chart also reveals the shape of the selected peak in the video to facilitate the detection of potential false peaks. An example is illustrated in Section 5. The second choice is course level. The timeline is correlated with the course period from the day the course opens to the day it is completed. Each bar below the timeline represents the sum of the clickstreams on each date. When a peak is selected, the clickstreams flow from different countries to various dates on the timeline.

To ensure that the view is simple and clear, we design the flow path as follows: first, we split this path into two parts. We then draw a Bezier curve for each part to ensure that both components are smooth, as shown in Fig. 2. Subsequently, we carefully position the intersection point of the two parts. This point (x_{cp}, y_{cp}) is defined as:

$$\begin{aligned} x_{cp} &= 0.5x_{sp} + 0.5x_{ep} \\ y_{cp} &= d + \alpha x_{sp} \end{aligned} \tag{3}$$

where (x_{sp}, y_{sp}) and (x_{ep}, y_{ep}) are the starting and end points of each path, respectively. d is the vertical distance between the starting point and the lower edge of the choropleth map, and α is a parameter that is dynamically adjusted according to the

size of the entire view. Thus, paths from the same country are bundled within the choropleth map to enable end users to view the choropleth map without much interference. The y_{cp} of each country differs as long as longitudes vary; therefore, most path bundles can easily be discriminated. If two countries have similar longitudes, users can check the paths of either country by highlighting their names. We have also designed two modes to avoid the influence of the total number of learners from different countries when comparing the click flows from various nations. One mode shows the number of clicks as in our original design, whereas the other mode shows the number of clicks normalized by the corresponding number of learners. In this way, users can examine two different statistics for their analytic tasks.

Unlike the static flowmap design, we initially produced an animation to present the same information on different dates. Instructors have displayed much interest in the use of animation and were quite impressed by the aesthetically pleasing presentation. Upon exploring further, however, the instructors found it difficult to observe patterns in the animation. Furthermore, watching the entire animation is time consuming. The instructors were required to maintain a high level of attention throughout the duration of the animation viewing. A generally desirable criterion for effective visualization is the capability to preserve a mental map, that is, maintaining the visual attention of the end users significantly while occupying little memory space. In this aspect, static views outweigh animation at this point; thus, we abandoned the animation design. The flow view offers an intuitive means through which the spatio- temporal information of the selected peak can be shown. We also provide a series of snapshot windows captured in flow view for side-by-side comparison, and interesting patterns can be spotted; this finding is explained in additional detail in Section 5.

4.2.4   Correlation View

End users can obtain general information and the spatio-temporal distribution of each peak from the overview and the flow view, respectively. However, end users still cannot deduce whether or not learners form particular groups according to their clickstream behaviors in the peaks detected. Thus, the correlation view is designed to such users to further analyze the correlation between different learner groups and the peaks.

Upon entering correlation view, end users are shown a choro- pleth map on a set of parallel axes to display the distribution of different attributes. The parallel axes can be classified into two categories; the first several axes represent the different attributes of learners, including country, grades, loyalty, delay, dropout time, video activeness, review activeness, and forum activeness. Table 2 shows the definitions of the attributes extracted from web log data. In this study, loyalty refers to the adherence of a learner to different peaks. For instance, if a learner clicks on most of the peaks (9 times out of 10), he or she is a high-loyalty learner (with 90 percent loyalty). Other attributes, such as delay and dropout time, can also help analyze the synchronicity in watching course videos and the persistence of a certain learner group. The second category includes all the peak axes; the coordinate on each axis depicts the number of times a learner has clicked on this peak. End users can filter different axes to explore the attributes

TABLE 2
Definition of Each Learner Attribute

Attribute	Meaning
loyalty	The ratio between the number of clicked peaks and the total number of peaks.
dropout time	The index of the last video which has been watched.
delay	The average time interval between the earliest watching date and the releasing date of each peak.
click activeness	The average click number of each peak.
review activeness	The average review times of each peak.
forum activeness	The total number of posts in the course forum.

they are interested in. The peak axes are rearranged after filtering, and the order is based on the number of learners who satisfy the filtering options in each peak. End users can also move different learner attribute axes and rearrange them according to their interests and domain knowledge. During the interviews, one instructor mentioned that he wishes to select particular users and to check their attributes and watching behaviors at an individual level. Accordingly, we have added a "draw lines" option to overlay parallel coordinates over the choropleth map when end users are interested in exploring individual level behaviors. Each line represents a learner in the parallel coordinates. This function is quite useful because the lines indicate clearer patterns when the number of learners is reduced after multiple instances of filtering. However, the parallel coordinates suffer from a visual clutter problem when too many learners are selected; thus, we merely show the choropleth map to end users by default.

We have also considered several alternatives, such as storyline design in which each actor represents a learner and each scene indicates a peak. However, the scalability problem is the foremost issue with this approach given that a single course often caters to more than 10,000 learners. By contrast, state-of-the-art storyline visualization [47] is designed only for dozens or hundreds of actors. Another possible design is the co-occurrence matrix, which is similar to parallel coordinates as each row of the matrix represents a learner and each column represents a peak. The color saturation indicates the corresponding click number. However, the additional visual channel and the color saturation may lead to perceptual ambiguity. More specifically, determining whether a dense cluster is caused by many learners clicking on the same peak or a single user clicking on a peak many times is difficult for users.

4.3 Interactive Exploration

The interaction among the three views is carefully designed to support the exploration of the clickstream data from different aspects and at various levels of detail.

Details-on-demand. Selecting, filtering and highlighting are the three major operations in PeakVizor. When a peak is selected in the overview, a pop-up window displays the detailed information of this peak and the corresponding video screenshot. Thus, users can quickly identify the content of the selected peak. Upon selecting a peak, users can proceed to the flow view, which has a video timeline. Then, the

corresponding video is shown in the flow view, and users can conduct further analysis smoothly. These users can also choose to compare cities from different countries by clicking on various nations in the choropleth map. By filtering various axes in the correlation view, users can explore the behaviors of different learner groups. When such users click on a certain axis, the corresponding glyph and the peak screenshot are highlighted to help users remember the peak they are interested in. The highlighting function in the flow view facilitates convenient viewing of the corresponding glyph and the screenshot of each peak. When users click on an axis in the correlation view (or on a glyph in the overview), the corresponding glyph (or axis) is highlighted along with the corresponding screenshot in the peak list. Thus, a coordinated analysis can be performed efficiently and effectively.

Comparative analysis. To facilitate a side-by-side comparison, a snapshot function is provided in the top right corner of the screen. Moreover, a clipboard is shown below each view to record the current view. While exploring the system and discovering interesting patterns, end users can click the snapshot button to capture images that are then stored in the clipboard. Any captured snapshots are magnified by hovering over a clipboard box, thus enabling users to browse their previous findings easily. After several rounds of exploration and switching among views, end users can open the clipboard that contains all their snapshots for further comparison and analysis.

5 CASE STUDIES

To evaluate the effectiveness and usefulness of PeakVizor, we conducted case studies with MOOC instructors and education analysts. We first performed data preprocessing which included data extraction and peak detection as mentioned in Section 3. In the data preprocessing part, the most time consuming part was to extract the data we need for calculating the clickstreams. The time was dependent on the size of the clickstream data. For example, it took 10 minutes for the NCH course with 1.2 million clickstreams; while for the JAVA course with 6 million clickstreams, it took about an hour. The peak detection algorithm worked efficiently. For a 10-minute video with a histogram of 600 bins, the running time was 60 ms on average. Therefore, for a typical MOOC course, the whole running time can be done in a few seconds, which costs little extra time out of the whole data preprocessing part. Then, we deployed the back-end part of the system into our server with a 2.7 GHz Intel Core i7 CPU, 8 GB memory PC. The data were stored in a local database (in our case, MongoDB) and loaded by the web server. Since PeakVizor is a web-based application, users can access the system through the web browser. During analysis, several patterns were identified, and the underlying insights were explained by the participating end users. We classified the major findings into four categories, which correspond to tasks T.1 to T.4 in Section 3.2.

5.1 General Statistics for Peaks

The general statistics of the peaks are mainly presented in the overview. For example, the JAVA course has four major clusters of peaks, as shown in Fig. 5. The peaks in each cluster clearly share similar grading distributions. The majority of
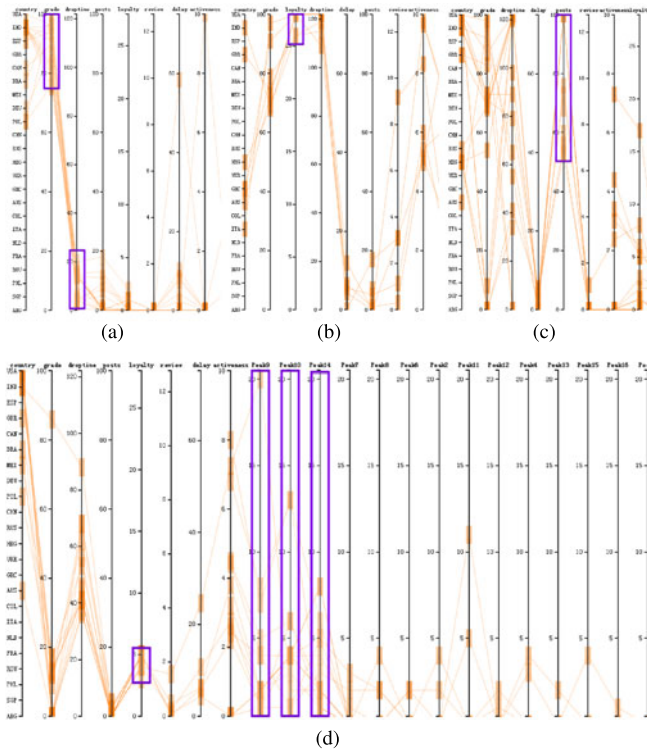
Fig. 6. Patterns detected in the correlation view. (a) The "Indian Phenomenon": learners drop out of the course early but still receive high grades. (b) Learners with high loyalty tend to receive high grades. (c) Learners who are active in the course forum tend to have low watching delay. (d) A small group of learners who focus on peaks #9, #10, #14.

the learners in the two clusters on the left side of the interface received low grades, whereas most of the learners in the bottom-right cluster reported high grades. By checking the inner line on the left of the glyphs, the instructor determined that the peaks at which most learners received higher grades were those peaks in the latter videos. This finding is reasonable given that those who followed the course and watched videos until the last are likely to receive high grades.

A few outliers are located at a slight distance from the center of the cluster, such as the two glyphs marked with red rectangles. A close examination of the two glyphs reveals that the inner lines of both peaks located around the middle are significantly higher than those of the others; this result indicates that a small number of learners contributed to approximately 50 percent of the total number of clickstreams in the peak. The instructors commented that such peaks are unusual and should be explored further.

5.2 Correlation between Learner Groups and Peaks

The correlation between learners and peaks can be analyzed in the correlation view, which enables users to explore how learners correlated with one another according to their behaviors in peaks and other learner attributes, as shown in Table 2. In the following, the findings related to the four different attributes are discussed.

Dropout time. Since courses on MOOC platforms usually report high attrition rate; thus, a large number of learners occasion- ally drop out throughout the course duration. For both the *JAVA* and *NCH* courses, the number of learners logging click actions dropped considerably after the first week. By contrast, learners who finished the first two weeks' videos

are likely to finish the course. In other words, the dropout rate is the highest in the first week and stabilizes after the first two weeks. The report from [20] showed similar results: "registrants who are active after the first week have a fairly low chance of leaving in subsequent weeks." When the instructor of the *JAVA* course explored the dropout time axis in the correlation view, he noted that certain learners who dropped out of the course early but still performed well in the assignments and in the final exam, as shown in Fig. 6a. To examine this phenomenon further, the instructor used the "draw line" function and reported that most of the learners who exhibited such learning patterns were from India. The question was raised of how Indian learners can perform well even without viewing the online course videos and interacting with the course content; the education analyst we consulted called this finding the "Indian Phenomenon"; this analyst believes that although distance education has been quite common in India and MOOC platforms have become popular recently, still a large number of learners from India tend to download the course videos instead of watching them online because of network bandwidth limitations. Therefore, no clickstream data were recorded for such offline video viewing behaviors. This pattern is normally observed in the Chinese learner group as well. For the *JAVA* course, the instructor uploaded the course videos via the YouTube platform; thus, many Chinese learners cannot access them. The use of the YouTube platform may explain the previous pattern for Indian learners as well as the small number of Chinese registrants in this course.

Grades and country. The instructor for the *JAVA* course first selected the learners with high grades and then filtered the results by selecting US learners and Indian learners on the country axis. The peak axes were dynamically ordered; therefore, the instructor could immediately identify the peaks that have high relevance to the selected group of learners. Subsequently, the instructor clicked on the snapshot button to record the filtered results, compared the snapshots and noted that these two groups of learners differed from each other in terms of interests. Specifically, among the top 10 peaks clicked by the high-performing US learners, only two ranked in the top 10 for the Indian learners. To ascertain the reasons behind this difference, the instructor checked the corresponding content of these high-ranked peaks. Upon careful observation, the instructor assumed the following: US learners tend to pause videos at the peaks when the instructor demonstrates a program or explains source codes, whereas Indian learners typically pause at points where the instructor illustrates concepts, such as certain terms. The instructor suggested that we investigate whether or not this finding holds true across courses when data for more courses are available.

Loyalty. One of our education analysts was interested in the correlation between loyalty and learning behaviors. The analyst explored this topic over a long period of time in the correlation view and determined that, learners with high grades generally tend to be highly loyal, that is, they click on the peaks more often and interact with the key content in the videos actively. Conversely, learners with low loyalty do not exhibit clear performance patterns. The expert indicated that while successful learners share common characteristics, those who fail usually have multiple reasons for doing so. Thus, the behaviors of the latter are difficult to define.
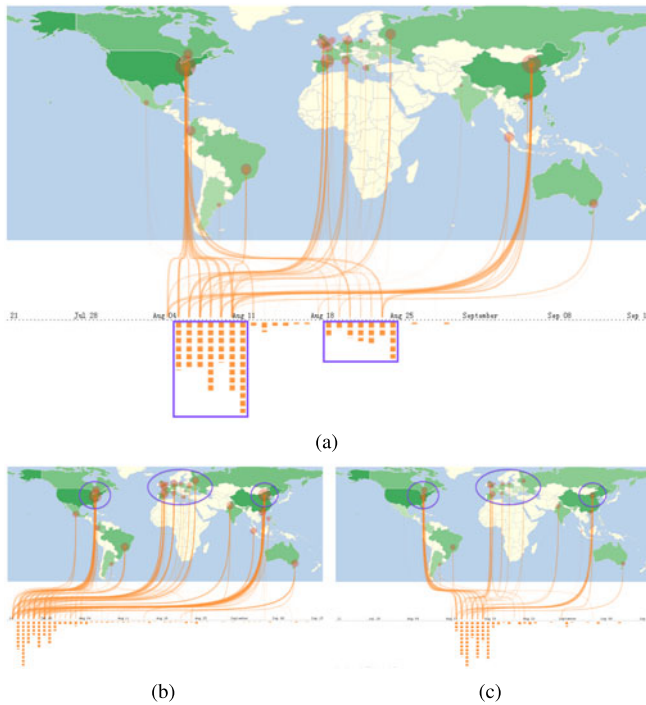
Fig. 7. Two patterns in flow view: two hot spots (marked with purple rectangles) along the timeline and a difference in the persistence between learners from different countries (marked with purple circles). (a) The first hot spot appears a week after the video is released, and the second hot spot appears before the final exam. The second hot spot usually lasts longer for US learners than for Chinese learners. (b) Flow view of a peak in the first video. (c) Flow view of a peak in the video during the third week (four weeks in total).

Forum activeness. We explored the relationship between video watching behaviors and forum activities as well. As shown in Fig. 6c, we filtered the forum activeness axis and determined that learners who were active in the forum (i.e., with numerous posts) generally watched the course videos with low delay. This finding indicated that learners who keep pace with the course schedule are proactive; furthermore, they are likely to discuss problems and share experiences with their fellow learners. Whenever a video is released, they always constitute the first group of learners who watched and interacted actively with one another in the forum.

Another education analyst was interested in whether or not learners could form different groups based on similarities in clicking behaviors; this analyst merely obtained the peak axes. By filtering the different axes through dynamic ordering, this analyst identified a small group of learners whose focus on peaks was distinct from that of the other learners. As per Fig. 6d, a group of learners focused specifically on three peaks but rarely clicked on the other peaks. The education analysts also noticed that most learners in this group performed poorly; however, these analysts have not yet developed a plausible explanation for the unusual focus of these learners. Therefore, the education analysts suggested that we further analyze this issue by exploring more attributes in these users, such as their behaviors in the course forum.

5.3 Spatio-Temporal Information of Peaks
The flow view is of significant interest to the instructors because this view can uncover spatial and temporal

information regarding the peaks for a detailed analysis of learning behaviors. Upon investigating the flow view of the *NCH* course, one instructor immediately detected two major hot spots (marked with purple rectangles in Fig. 7) in the clickstream for most of the peaks. By referring to the course syllabus, we learned that the first hot spot corresponds to the first week after the video is released and the second hot spot to the review week before the exam. By exploring this view further, the instructors also determined that the review habits of US learners and Chinese learners vary. Specifically, US learners review throughout the week, whereas Chinese learners review only on the final two days of the week. Thus, the cramming period of Chinese learners is shorter than that of US learners. With respect to the difference in the total number of learners from different countries, the normalized mode was initiated to ensure that this result was not generated by comparing a small number of learners with a large number of them. The result is unchanged because the total number of learners from the US and from China was similar in this course. Furthermore, a few peaks were preferred over the review hot spot. The instructors suggested that these peaks can reflect the interest of learners because they are irrelevant to the exam. Thus, instructors can accurately understand the genuine interest of learners by examining the content of these peaks.

Another insight presented by an instructor is that the number of clicks decreases significantly over time because MOOCs usually report high dropout rates. Upon comparing the flow maps of different peaks, the instructor noted that the number of clicks decreases gradually for European countries such as the UK and Spain. As shown in Fig. 7b, the flow view of peak #1 from the first video of *NCH* suggests that learners from the US and China (marked with purple circles) contribute most of the clicks in this peak. By contrast, the sizes of the circles from most regions shrink significantly for peak #28 (Fig. 7c), which appears in the video during the third week. Nonetheless, the sizes of the circles from West European countries remain almost the same. This pattern may indicate that the learners from the latter region exhibit persistent viewing behavior.

5.4 Anomaly among Peaks
The anomaly among peaks was discovered by an instructor who performed an in-depth analysis supported by our system. As mentioned in Section 5.1, two unusual peaks were detected by observing the overview. Subsequently, the instructor clicked on a glyph to obtain detailed information regarding the peak. A pop-up window appeared with a screenshot of the corresponding video content along with a click event chart of the selected peak and detailed information regarding the peak. This information includes the video title, video length, peak duration, and action number of the peak. Upon examining the corresponding video content, the instructor did not find confusing or interesting material at this peak; thus, the issue remained unsolved until he realized that this peak may be a false peak caused by a system error or disruptive learners. To check the feasibility or validity of this assumption, he explored this peak further in a detailed view by clicking the "go flow view" button. Fig. 2 depicts the scenario encountered when he entered the flow view. The click event chart of the corresponding video is

shown below the flow map, and the peaks were marked with glyphs. The clickstreams flowed from all over the world to the three peaks in the video; unlike the other two peaks, the shape of the third peak was much "sharper" and was exactly the type of abnormal peak being investigated by the instructor. We checked the original data and learned that the peak was caused by hundreds of clicks from a single user within a two-second period. Therefore, this peak is considered to be a false peak and is probably caused by a system error because a single user cannot possibly click at such a high frequency in two seconds.

6 EXPERT INTERVIEWS

To evaluate our system further, we conducted in-depth interviews with five domain experts and seven teaching assistants. Among them, two MOOC instructors (MIs) and an education analyst (EA1) had collaborated closely with us from the problem characterization phase. The other two experts, namely, one MOOC instructor and one education analyst, did not have prior experience in using our system before the interviews. We also invited another seven teaching assistants who had experience working with MOOC instructors. We denote all the participants here as MI1, MI2, MI3, EA1, EA2, TA1, TA2, TA3, TA4, TA5, TA6, and TA7; they are aged between 23 and 63 years and are not color blind. Among them, three are female and the remaining nine are male.

Procedure. Prior to the interviews, we collected background information from the participants, including their research experience, teaching experience, and background knowledge on MOOC, visualization, and MOOC-related data analytics. All of them had at least one year of experience with both MOOCs and university teaching. Seven out of 12 had at least one year of experience in visualization, whereas the other five had no background knowledge on this subject. Half of the participants had previously observed MOOC-related data analyses, but only two regarded the analysis as significantly useful to them. We then briefly introduced the goals and features of PeakVizor and showed a detailed tutorial video of the system. Afterward, the participants were asked to explore the system and to familiarize themselves with its features and functions via several small tasks we proposed for them to perform.

The following small tasks were designed to help answer the questions in the task analysis part described in Section 3.2:

- Acquire a general overview of the statistics for peaks in the required course and identify typical patterns (T.1).
- Find a specific peak (a potentially abnormal peak) and retrieve the corresponding detailed information about this peak (T.4).
- Observe the flow view and the correlation view to identify different learner groups with peaks (T.2).
- Explore the video-level flow view and determine when and where the clicks of a selected peak originate (T.3).

All the small tasks were based on the JAVA course dataset while the tutorial was mostly based on the NCH course. The participants were also required to answer some questions after finishing each task. Finally, we invited them to select a course (either *JAVA* or *NCH*) and explore the system using the functions they had just learned. Then, they were required to answer some open questions regarding the patterns they observed from each view; to provide their opinions on system usability, visual designs and interaction, and specific functions; to inform us of difficulties they encountered when exploring the system; and to provide suggestion for further improvement. Each interview study lasted approximately 1.5 hours.

Summary of interviews on system usability. All the participants were satisfied with the system and deemed that the views and features of the system could fulfill most of their requirements. EA2 commented, *"Compared with the current analytic systems, PeakVizor offers various levels of data analysis, which is powerful and will allow us to dig deeper and gain more insights into online learning behaviors."* EA2 had examined previous works on learning behavior analysis in MOOCs and was surprised that a system could facilitate learner group behavior analysis from clickstream data. When exploring the system, MI1 and MI2 focused on analyzing their respective courses and obtained interesting information, which were described in detail in the previous section. Prior to the interview, MI3 had already heard about our system from one of the MOOC instructors who collaborated with us. MI3 himself also offered an MOOC, and after trying the system, he was keen on having us analyze his course as well. He even suggested that PeakVizor can be applied not only to MOOCs but also to other e-learning platforms, such as the online system used in his university. EA2 also put forward a similar idea of implementing the system to analyze clickstream behavior for different video types. The TAs agreed that the system was useful. For instance, TA1 commented, *"I think the layout is intuitive. The overview clusters the students with different grade well. It's useful for teachers to analyze the students learning habit and improve their teaching quality."* However, some TAs complained about the loading time of the correlation view. TA5 stated that *"The interactions are quite good but due to the volume of data or performance of the computer, the fluency of system is not good."* At the same time, TA4 suggested that *"The tutorial is useful, however, for those who are not so familiar with visualization or MOOCs, it might takes more time to learn how to use the system."*

Summary of interviews on visual designs and interactions. The visual designs were well received by all the participants, and they appreciated the various interactive functions. The color and size encodings of the glyph were promptly accepted; however, understanding the inner line consumed some time when it was explained in the tutorial. TA2 commented, *"Since the glyph encodes too much information, it would be better to add some visual encodings in the diagrams to help assist understanding."* EA1 and MI1 agreed that the glyph design not only presents the basic information and singularity of a peak well but also serves as an icon for the glyph so that it is easy to locate at different views for different analytic tasks. MI2 also commented, *"The overview gives me a clear idea about the general peak distribution and the outliers are easy to detect from both the glyph itself and the layout of the glyph."* EA2 appreciated the flow view and considered it *"visually appealing"*; this analyst regarded it as a *"vivid*

representation" given that learners and clickstreams flow from different countries to the corresponding date on the timeline. TA3 commented, *"The correlation view contains a lot of information, I can spend a whole day exploring it."* With regard to the correlation view, all the participants felt that the default choropleth map is easy to understand; however, after they clicked on the "draw lines" button, some of the participants found the parallel coordinates quite difficult to understand. After we explained this concept carefully, they accepted it and considered the coordinates to be clearer than the choropleth map when the number of learners is sufficiently small after several steps of filtering. With regard to interactions, M2 was intrigued by the pop-up window showing the corresponding video content and clickstream curve and said *"When I was considering going back to check what content the peak corresponds to, I suddenly found this pop-up window, and it showed the exact information I wanted to know!"* The filtering and highlighting functions were frequently used when the participants investigated the correlation view. We observed their exploration carefully and noted some differences in their attention. The two major attributes the three MOOC instructors focused on are the country and grade information, which we derived from the databases directly. They also paid much attention to dropout time because they wished to analyze when and why learners give up during the course period. However, they did not explore the delay axis much, whereas the two education analysts filtered this axis few times. When we asked for their reasons, they said they would like to compare the MOOC offerings with traditional courses, in which learners are constrained in terms of when and where to study. Only one user was especially interested in the loyalty axis, and she thought it was an innovative way to encourage loyalty.

Summary of interviews on difficulties and suggestions. In addition to the comments about the usability of the system and the visual designs, the domain experts also provided valuable suggestions on various aspects. TA4 mentioned the scalability problem caused by the considerable number of glyphs. However, when we showed the glyphs on a large screen, the scalability problem was negligible. For projection on a small screen, TA4 suggested, *"You could add a magnifying lens to help observe the glyph clearly."* Most of them appreciated the snapshot and clipboard functions. TA6 and TA7 commented that these features helped them revisit the detected patterns and were useful in comparing the flow views for two peaks from different videos. However, EA2 mentioned, *"I recorded too many snapshots and got confused later. It would be better to support the automatic generation of a pdf file after the whole exploration."* EA1 suggested combining the analysis with forum data to further explore the learner group behaviors, which we plan to focus on in our future work. MI1 also mentioned that knowing the analysis results several days after the video is released is useful so that he could prepare some supplementary materials, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TVCG.2015.2505305, and upload them the next week or place these materials in the forum to help learners understand the course and catch up with the lessons. EA1 suggested that we relate our analysis with the learner surveys on the onset of the course and the final evaluation at the end of the course to uncover more interesting information about the correlation between the motivation of learners and their behaviors and performance in the course.

7 DISCUSSIONS

PeakVizor is designed to help explore and understand the learning behaviors underlying the clickstream peaks in MOOC videos. Three coordinated views, namely, overview, flow view, and correlation view, are used to analyze underlying learning patterns from different perspectives. Case studies are conducted to illustrate the new insights into learning behaviors. The system is well received and highly rated by the end users.

However, the system still has several limitations. Although the choropleth map is integrated into the parallel coordinates to reduce visual clutter, the system still suffers from a scalability problem when the number of users is considerably large. The number of selected peaks should only be up to 30 to 40 to enable users to observe the different attributes clearly. Fortunately, according to the feedback of the participating end users, the scalability of the correlation view is sufficient to handle most analytic tasks. In the data processing part, we use a peak detection algorithm to extract peaks from videos. During the interviews, we learned that automatic peak detection limits end users within the detected peaks. If the ROIs of a particular end user are missed by the detection algorithm, then users cannot analyze these ROIs. Therefore, if the system allows end users to select the peaks themselves when analyzing the data and update the views accordingly in real-time, then the end users can quickly examine the results after they find new interesting content.

As we have only collected data from two courses in this work, education analysts may have difficulty drawing general conclusions that are applicable to different courses. For instance, the lengths of the two courses are significantly different, and their bodies of literature are different as well. Therefore, we cannot draw absolute conclusions as to whether or not video length may result in differences among peaks. Nonetheless, the number of peaks in long videos is indeed higher than that in short videos. This result may be ascribed to the fact that the attention of learners is retained in short videos, thereby resulting in few click action peaks. If data from more courses can be collected and analyzed, then course-level analysis can be performed, and numerous patterns, which the current system can find, can be further examined. Thus, a visualization system that supports course-level analysis is useful and can enable users to investigate data from diverse aspects.

The roles of language and educational culture have been commonly studied in MOOC analyses, and analyzing how their roles can affect the formation of a peak in a learning behavior is valuable. However, we are currently unable to obtain the detailed profile information of learners. We plan to actively collaborate with platform providers to obtain these data.

Combining the video content with peak analysis is also worthy of further study. Previous studies such as [10] reported that videos showing an instructors talking head are more engaging than slides alone. Both of the courses we

use for our system show the instructors talking head most of the time together with slides in the videos; thus, we did not draw the same conclusion. However, as PeakVizor provides the corresponding video screenshots of peaks, end users can easily analyze the data with the video content. By comparing the screenshots of forward seeks, the experts have determined that learners often skip three types of videos: 1) the opening of each video, 2) flashing the same slide or the instructor talking in front for an extended period of time, and 3) the instructor talking about his own opinion on some issues. The final type is surprising to the course instructors; we speculate that the students care more about the results and conclusions than the instructors.

In Section 5.2, we discussed a pattern regarding some learners who download videos and subsequently watch these videos offline. Some previous work regarding offline study, such as [3] and [48], generated results suggesting that students who collaborate offline may perform well in class. However, the problem regarding offline watching behavior has not been discussed and studied. The authors in [10] reported that the viewing activities of students who downloaded videos and watched offline could not be tracked. In the future, we plan to collaborate with a MOOC platform provider to obtain the download records of learners and distribute some questionnaires about the offline study experience of learners.

All our current analysis is based on the learners' clickstreams; therefore, if learners watch a certain segment without any additional click behavior, then interpreting how they feel about that segment is difficult. We admit that "unexpectedly" low activities ("valleys") can also be informative and useful for analysis. Even though we did not examine this issue in this work given that none of our end users expressed interest in this issue during the collaboration, we believe that our visual analytic system can be easily extended to detect and highlight such valleys for the purpose of exploring and understanding unexpectedly low activities.

8 CONCLUSION AND FUTURE WORK

In this paper, a visual analytic system, PeakVizor, is presented to help course instructors and education experts to analyze peaks in video clickstreams in MOOCs from different aspects. Based on our collaborations with end users, we first abstract the analytic tasks and summarize the design requirements accordingly. PeakVizor is then carefully designed to solve the tasks and meet the requirements. We then conduct case studies and interview experts to evaluate this system. Positive feedback and in-depth insights into learning behaviors have confirmed the usefulness and effectiveness of the system.

Nevertheless, PeakVizor can still be improved. In the future, we will add a functionality which allows manual peak selection ac- cording to the interests of users. We also plan to extend our system in the following two directions. First, a visual analysis of MOOC forum data can be added to enhance the understanding of learning patterns. Second, a comparative analysis of different courses is useful for users; therefore, we are considering the extension of our system to support course-level analysis. Moreover, we will design a more structured evaluation with multiple dimensions for users to rate the usability of the system, and methods such as heuristics can be used to evaluate it more systematically.

REFERENCES

[1] K. Williams, "Content analysis of coursera, edx, and udacity course platforms," Ph.D. dissertation, Univ. Prince Edward Island, Charlottetown, PE, Canada, 2014.

[2] Coursera. (2015, last access date: 2015/08/30) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Coursera

[3] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton, "Studying learning in the worldwide classroom: Research into edxs first MOOC," *Res. Practice Assessment*, vol. 8, no. 1, pp. 13–25, 2013.

[4] E. Aguiar, S. Nagrecha, and N. V. Chawla, "Predicting online video engagement using clickstreams," *arXiv preprint arXiv:1405.5147*, 2014.

[5] S. Breslav, A. Khan, and K. Hornbæk, "Mimic: Visual analytics of online micro-interactions," in *Proc. Int. Working Conf. Adv. Visual Interfaces*, 2014, pp. 245–252.

[6] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller, "Understanding in-video dropouts and interaction peaks inonline lecture videos," in *Proc. 1st ACM Conf. Learn.@ Scale Conf.*, 2014, pp. 31–40.

[7] C. Shi, S. Fu, Q. Chen, and H. Qu, "Vismooc: Visualizing video clickstream data from massive open online courses," in *Proc. IEEE Pacific Vis. Symp.*, 2015, pp. 159–164.

[8] F. Dernoncourt, C. Taylor, U.-M. OReilly, K. Veeramachaneni, S. Wu, C. Do, and S. Halawa, "Moocviz: A large scale, open access, collaborative, data analytics platform for MOOCs," in *Proc. NIPS Workshop Data-Driven Edu.* [Online]. Available: Retrieved from http://groups. csail. mit. edu/EVO-Design-Opt/groupWebSite/uploads/Site/MoocViz. pdf, 2013.

[9] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," *IEEE Trans. Syst., Man, Cybern., Part C: Appl. Rev.*, vol. 40, no. 6, pp. 601–618, Nov. 2010.

[10] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: An empirical study of MOOC videos," in *Proc. 1st ACM Conf. Learn.@ Scale Conf.*, 2014, pp. 41–50.

[11] V. P. Dennen, "Pedagogical lurking: Student engagement in non-posting discussion behavior," *Comput. Human Behavior*, vol. 24, no. 4, pp. 1624–1633, 2008.

[12] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in MOOCs using learner activity features," in *Proc. Eur. MOOC Stakeholder Summit*, 2014.

[13] Y. Belanger and J. Thornton, "Bioelectricity: A quantitative approach," Duke University's First MOOC, Tech. Rep., Duke University, NC, USA, 2013.

[14] G. Christensen, A. Steinmetz, B. Alcorn, A. Bennett, D. Woods, and E. J. Emanuel, "The MOOC phenomenon: Who takes massive open online courses and why?," Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2350964, 2013.

[15] S. O. Nesterko, S. Dotsenko, Q. Han, D. Seaton, J. Reich, I. Chuang, and A. Ho, "Evaluating the geographic data in MOOCs," in *Proc. Neural Inf. Process. Syst.*, 2013.

[16] A. Kasunic, J. Hammer, and A. Ogan, "Cultural relevance in MOOCs: Asking about socioeconomic context," in *Proc. 2nd ACM Conf. Learn.@ Scale.*, 2015, pp. 389–392.

[17] P. J. Guo and K. Reinecke, "Demographic differences in how students navigate through MOOCs," in *Proc. 1st ACM Conf. Learn.@ Scale Conf.*, 2014, pp. 21–30.

[18] S. O. Nesterko, D. Seaton, J. Reich, J. McIntyre, Q. Han, I. Chuang, and A. Ho, "Due dates in MOOCs: Does stricter mean better?," in *Proc. 1st ACM Conf. Learn.@ Scale Conf.*, 2014, pp. 193–194.

[19] A. C. Robinson, J. Kerski, E. C. Long, H. Luo, D. DiBiase, and A. Lee, "Maps and the geospatial revolution: Teaching a massive open online course (MOOC) in geography," *J. Geography Higher Edu.*, vol. 39, no. 1, pp. 65–82, 2015.

[20] A. D. Ho, J. Reich, S. O. Nesterko, D. T. Seaton, T. Mullaney, J. Waldo, and I. Chuang, "Harvardx and mitx: The first year of open online courses, fall 2012-summer 2013," HarvardX and MITx Working Paper No. 1, 2014.

[21] R. F. Kizilcec and E. Schneider, "Motivation as a lens to understand online learners: Toward data-driven design with the olei scale," *ACM Trans. Comput.-Human Interaction*, vol. 22, no. 2, 2015, Art. no. 6.

[22] A. D. Ho, I. Chuang, J. Reich, C. A. Coleman, J. Whitehill, C. G. Northcutt, J. J. Williams, J. D. Hansen, G. Lopez, and R. Petersen, "Harvardx and mitx: Two years of open online courses fall 2012-summer 2014," Available at SSRN http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2586847, 2015.

[23] Y.-C. Lee, W.-C. Lin, F.-Y. Cherng, H.-C. Wang, C.-Y. Sung, and J.-T. King, "Using time-anchored peer comments to enhance social interaction in online educational videos," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 689–698.

[24] S. Kardan and C. Conati, "Providing adaptive support in an interactive simulation for learning: An experimental evaluation," in *Proc. 33rd Annu. ACM Conf. Human Factors Comput. Syst.*, 2015, pp. 3671–3680.

[25] C. Coffrin, L. Corrin, P. de Barba, and G. Kennedy, "Visualizing patterns of student engagement and performance in MOOCs," in *Proc. 4th Int. Conf. Learn Analytics Knowl.*, 2014, pp. 83–92.

[26] D. Trimm, P. Rheingans, and M. desJardins, "Visualizing student histories using clustering and composition," *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2809–2818, Dec. 2012.

[27] Z. Xu, D. Goldwasser, B. B. Bederson, and J. Lin, "Visual analytics of MOOCs at maryland," in *Proc. 1st ACM Conf. Learn.@ Scale Conf.*, 2014, pp. 195–196.

[28] D. Wortman and P. Rheingans, "Visualizing trends in student performance across computer science courses," *ACM SIGCSE Bull.*, vol. 39, no. 1., pp. 430–434, 2007.

[29] J. Wei, Z. Shen, N. Sundaresan, and K.-L. Ma, "Visual cluster exploration of web clickstream data," in *Proc. IEEE Conf. Visual Analytics Sci. Technol.*, 2012, pp. 3–12.

[30] R. Shaw and M. Davis, "Toward emergent representations for video," in *Proc.13th Annu. ACM Int. Conf. Multimedia*, 2005, pp. 431–434.

[31] Z. Shen, J. Wei, N. Sundaresan, and K.-L. Ma, "Visual analysis of massive web session data," in *Proc. IEEE Symp. Large Data Anal. Vis.*, 2012, pp. 65–72.

[32] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based visualization: Foundations, design guidelines, techniques and applications," *Eurographics State Art Rep.*, pp. 39–63, 2013.

[33] J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg, "Evaluation of alternative glyph designs for time series data in a small multiple setting," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2013, pp. 3237–3246.

[34] M. O. Ward, "A taxonomy of glyph placement strategies for multidimen- sional data visualization," *Inf. Vis.*, vol. 1, no. 3–4, pp. 194–210, 2002.

[35] M. O. Ward, "Multivariate data glyphs: Principles and practice," in *Proc. Handbook Data Vis.*, 2008, pp. 179–198.

[36] J. Pan and W. J. Tompkins, "A real-time QRS detection algorithm," *IEEE Trans. Biomedical Eng.*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985.

[37] M. I. Sezan, "A peak detection algorithm and its application to histogram- based image data reduction," *Comput. Vis., Graph., Image Process.*, vol. 49, no. 1, pp. 36–51, 1990.

[38] J. E. Albus, et al., *Syntactic Pattern Recognition, Applications*. New York, NY, USA: Springer, 2012, vol. 14.

[39] A. Guille and C. Favre, "Event detection, tracking, and visualization in twitter: A mention-anomaly-based approach," *Social Netw. Anal. Mining*, vol. 5, no. 1, pp. 1–18, 2015.

[40] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: Aggregating and visualizing microblogs for event exploration," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 227–236.

[41] T. Munzner, "A nested model for visualization design and validation," *IEEE Trans. Vis. Comput. Graph.*, vol. 15, no. 6, pp. 921–928, Nov. 2009.

[42] T. Munzner, "Visualization analysis and design," *J. Geography Higher Edu.*, pp. 132–133, 2014.

[43] N. Cao, D. Gotz, J. Sun, and H. Qu, "DICON: Interactive visual analysis of multidimensional clusters," *IEEE Trans. Vis. Comput. Graph.*, vol. 17, no. 12, pp. 2581–2590, Dec. 2011.

[44] J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," *Psychometrika*, vol. 29, no. 2, pp. 115–129, 1964.

[45] T. Dwyer, K. Marriott, and P. Stuckey, "Fast node overlap removal," in *Proc. 13th Int. Conf. Graph Drawing*, 2006, vol. 3843, pp. 153–164.

[46] S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2008, pp. 1–8.

[47] S. Liu, Y. Wu, E. Wei, M. Liu, and Y. Liu, "Storyflow: Tracking the evolution of stories," *IEEE Trans. Vis. Comput. Graph.*, vol. 19, no. 12, pp. 2436–2445, Dec. 2013.

[48] J. DeBoer, G. S. Stump, D. Seaton, and L. Breslow, "Diversity in MOOC students backgrounds and behaviors in relationship to performance in 6.002 x," in *Proc. 6th Learn. Int. Netw. Consortium Conf.*, vol. 4, 2013.

Qing Chen received the BEng degree in digital media technology from the Department of Computer Science, Zhejiang University, China, in 2013. She is currently working toward the PhD degree in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology (HKUST). Her research interests include information visualization, visual analytics, human-computer interaction, and online education.
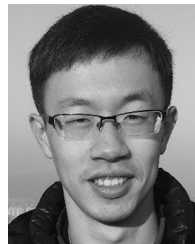
Yuanzhe Chen received the BE and ME degrees in electrical engineering from Shanghai Jiao Tong University, China, in 2014. He is currently working toward the PhD degree in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology. His research interests include information visualization and human-computer interaction.
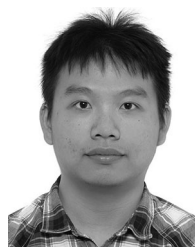
Dongyu Liu received the BS degree in computer science from the Zhejiang University, China in 2014. He is currently working toward the PhD degree in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology. His research interests include graph visualization, crowdsourcing, and e-learning data analysis.

Conglei Shi received the BSc degree from Shanghai Jiao Tong University in major of computer science and the PhD degree in computer science from the Hong Kong University of Science and Technology. He is a postdoctoral research fellow in the IBM T.J. Watson Research Center. His main research interests include information visualization, visual analytics, and human computer interaction.

Yingcai Wu received the PhD degree in computer science from the Hong Kong University of Science and Technology (HKUST). He is an assistant professor at the State Key Lab of CAD CG, Zhejiang University, Hangzhou, China. Prior to his current position, he was a researcher at the Internet Graphics Group in Microsoft Research Asia, Beijing, China. His primary research interests include visual behavior analytics, visual analytics of social media, visual text analytics, uncertainty-aware visual analytics, and information visualization. He is a member of the IEEE.

Huamin Qu received the BS degree in mathematics from Xi'an Jiaotong University, China, and the MS and PhD degree in computer science from the Stony Brook University. He is a professor in the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology. His main research interests include visualization and computer graphics, with focuses on urban informatics, social network analysis, e-learning, and text visualization.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.