VUSphere: Visual Analysis of Video Utilization in Online Distance Education

Huan He [1,2*], Qinghua Zheng [1,2 †], and Bo Dong [3,4‡]

[1] Key Laboratory of Intelligent Networks and Network Security, Ministry of Education, Xi'an Jiaotong University
[2] School of Electronic and Information Engineering, Xi'an Jiaotong University
[3] School of Continuing Education, Xi'an Jiaotong University
[4] National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University

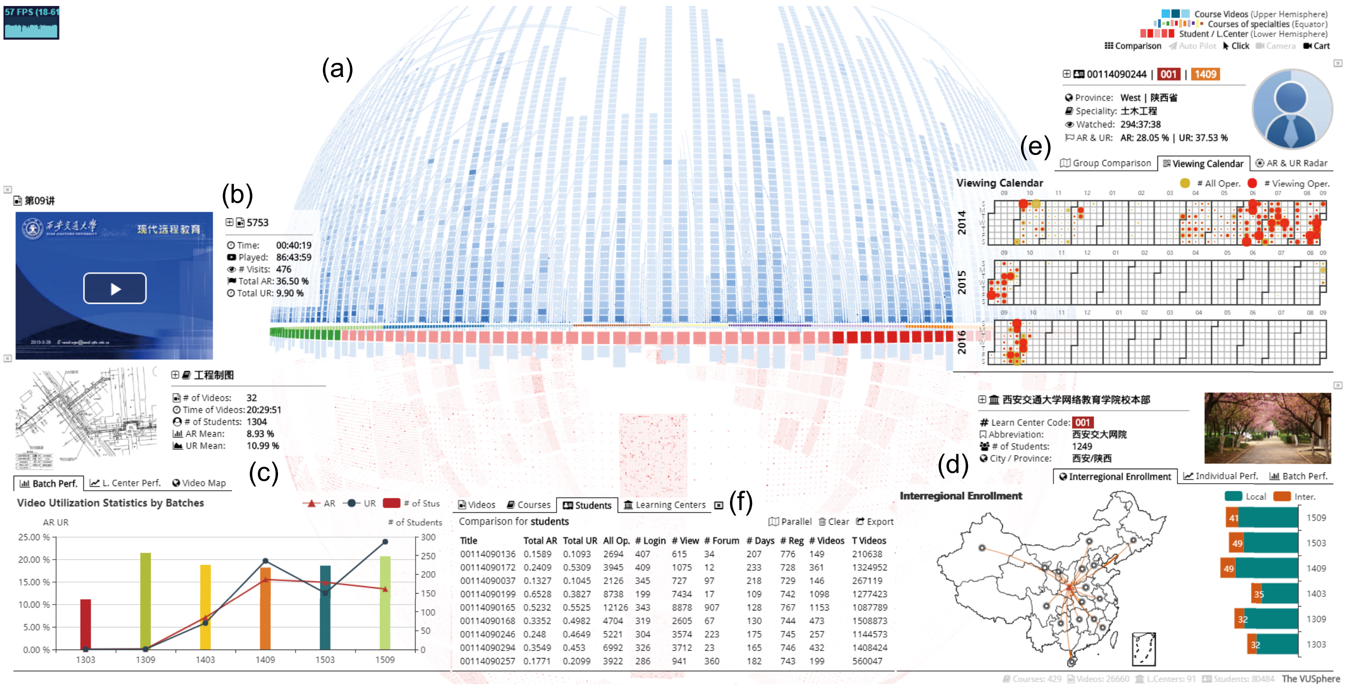

Figure 1: Demonstrating the VUSphere system that shows: (a) an Overview depicting the overall utilization of all videos and students; a Detailed Statistics View including: (b) a video panel for direct access to the video content; (c) a course panel showing the video utilization of a course; (d) a learning center panel showing the geographic distribution and video utilization of the students in a learning center; (e) a student panel revealing the learning process of a student; and (f) a Comparison View examining the differences between individual students.

Abstract

Online Distance Education (ODE) provides massive course videos of various specialties for students across the country to learn professional knowledge anytime and anywhere. Analyzing the utilization of these videos from user log data can help academics better understand the learning process of students, evaluate the quality of service provided by regional learning centers, and improve the quality of program curriculum in the future. However, due to the lack of comparable indicators, it is a great challenge to discover the utilization patterns of massive videos and analyze the learning process of large-scale student population from learning log data. In this paper, we introduce a visual analytics system, called VUSphere, to explore the video utilization from multiple perspectives with two proposed indicators. This system offers three coordinated views: a spherical layout overview to depict the overall utilization distribution of videos, courses, and students; a detailed statistics view with four panels to present video utilization statistics of each element from multiple perspectives; and a comparison view to examine the differences in individual elements. Based on the real dataset from our ODE school, several patterns related to video utilization and enrollment are found in the case study with our domain experts.

Keywords: Video utilization pattern, online distance education, visual analytics.

1 Introduction

With the rapid development of internet technology and the popularity of computing terminals, online learning has become an important supplement to higher education in recent years [38, 39]. Students not only have access to high-quality educational resources to learn professional knowledge and skills without spatial-temporal constrains, but also may obtain corresponding academic certificates to lay the foundation for better career opportunities or financial return [42]. In China, to promote a balanced development of higher education and establish a lifelong learning system, the Chinese Ministry of Education launched the “Modern Distance Education Project” and approved 68 universities to set up schools to provide online distance education (ODE) throughout whole country since 1998 [40]. According to the statistics from the National Bureau of

* hehuan@mail.xjtu.edu.cn
† qhzheng@mail.xjtu.edu.cn
‡ dong.bo@mail.xjtu.edu.cn

Statistics of China, these ODE schools have enrolled more than 10 million students in recent five years [41].

To meet the challenges of producing online courses for many people, ODE schools have established learning management systems (LMSs) to provide a variety of online teaching functions such as video player, live interaction, forums, and quizzes. Based on these functions, massive educational resources can be used by students, such as videos, slides, and textbooks. Among these resources, the educational videos play an important role in the online education [25, 36], and it takes considerable time and effort to produce these videos [27, 37]. Understanding how these video resources are utilized by students may provide support in analyzing student's learning process, evaluating service quality, and improving course design.

Due to the use of LMS, students' learning behaviors have been recorded as learning log. However, there are three challenges in analyzing these log data to gain insight into the video utilization. First, the lack of comparable indicators that measure the video utilization from different perspective. Due to the diversity of the massive courses and videos, commonly used indicators, such as total count of visits and time spent on watching, may cause misunderstanding of video utilization of different courses. Thus, existing indicators are inapplicable to evaluate the video utilization and compare the differences between courses. Second, the learning data is big. Tens of thousands of students are enrolled each year, and each of them may watch hundreds of videos during their long-term studies, which generates a huge amount of log data. In addition, the entire learning process is affected by factors such as curriculum and learning center, which makes data analysis more complex. Last, existing systems have limited capabilities in analyzing large-scale video utilization of massive videos. The analysis of the course resource utilization has been provided in existing LMSs such as Moodle [46] and Blackboard [47] for teachers and related staff. Nevertheless, these analytic systems are designed and applied in a single course or several courses, which are not suitable for direct comparison of a large number of courses.

To address these challenges, we propose two indicators and develop a visual analytics system, called VUSphere (Video Utilization Sphere), to analyze the video utilization from multiple perspectives. The system has three coordinated views: 1) an Overview of spherical layout to depict the overall distribution of video utilization of all videos and students; 2) a Detailed Statistics View that presents four panels with nine tabs to show the statistics of video utilization from multiple perspectives. 3) a Comparison View to examine the differences among multiple elements, such as videos, courses, students and learning centers. In addition, we conduct a case study and an interview with domain experts to evaluate the usefulness and effectiveness of our system.

The major contributions of this paper are as follows:

- An interactive visual analytics system which integrates several visualization techniques to explore the massive learning data, based on a set of domain-specific tasks and design requirements.
- Two indicators for analyzing and comparing the video utilization of different courses, students and learning centers.
- Several patterns related to video utilization are found by our system in the case study with domain experts, based on real datasets from our ODE school.

2 RELATED WORK

In this section, we first review the literature related to online learning visualization. Then, we summarize recent works on educational video analysis. Finally, we discuss the indicators and visual methods related to learning process analysis.

2.1 Online Learning Visualization

The extensive use of LMS provides fine-grained data for analyzing students' online learning behaviors. Especially in the massive open online courses (MOOCs), the learning process of worldwide students has produced large-scale, multi-dimensional log data. To discover learning patterns from these data, many visual analytics systems have been developed. Dernoncourt et al. [28] introduced the MoocViz, which is an analytics platform that helps researchers analyze log data from multiple MOOC platforms and share analysis scripts. Pardos and Kao [29] developed an open source analysis tool called moocRP to enable reuse and replication of analytics results. Qu and Chen [30] introduced multiple interfaces of visual analysis in the VisMOOC system.

Recently, some studies have focused on applying visual analytics on data of online forums. Wu et al. [31] presented the NetworkSeer, which is a visual system for visualizing interactions in forum and exploring behavioral differences between student groups. Fu et al. [32] developed a visual analytics system, iForum, to examine the topic changes in forum and understand behavioral patterns of students. Wong et al. [33] introduced a visualization tool called ForumGraph to detect topic in the discussion of forum. In addition, visualization techniques are also used in other aspects related to analyzing educational data. For example, a visual system called DropoutSeer was developed to uncover learners' detailed learning activities and identify the factors related to dropout behavior [34]. Another study applied map-based visual analysis method in exploring students' learning behavior patterns in mobile learning environment. [35].

In summary, these studies demonstrate multiple innovative visualizations, such as standard graphs, thread river [32], node-link diagram [33], calendar-based heatmap [30], parallel coordinate [31] and novel glyphs [24] to analyze the behavior patterns of large-scale student population in a single course or several courses. The design of VUSphere has been inspired by the above systems and techniques. Although several visual designs in VUSphere are similar to the techniques used in existing systems, our visual analytics system focuses on comparing the video utilization distribution of many courses by integrating existing techniques in coordinated views. Moreover, we propose a spherical surface layout method for visualizing the overall distribution of the video utilization of massive videos and students.

2.2 Educational Video Analysis

Educational videos are widely used in online courses to transfer knowledge without limitation of time and space [5]. Therefore, there have been many studies on educational videos from several aspects.

Some studies analyze log data of viewing behaviors to explore the effects of various video properties on students' learning process. Guo et al. [7] found that shorter videos, informal talking-head videos and Khan-style videos are more engaging, i.e., students may spend more time watching such videos. Chen and Wu [6] explored how three commonly used video styles affect the sustained attention, emotion, cognitive load, and learning performance. In addition, the complexity of videos also has impact on the time students spend on watching a video [8]. Longer video and tutorial video may lead to higher dropout rate [26]. Other studies have examined the video visits [9], video release strategy [10], and demographics of course participants [11].

Benefiting from advances in web development technologies, new interactive features have been added to video player to improve the learning experience, which raises new research questions. For example, Kovacs [12] analyzed how learners interact with in-video quizzes and the influence of in-video quizzes on learners' viewing behavior. Zhao et al. [13] proposed an approach of reusing past discussion data in video to help learners better understand the materials. Ruip´erez-Valiente et al. [14] developed an algorithm for identifying a specific online cheating strategy from log data of video viewing and answering questions.

Previous studies have used a variety of methods to expand the understanding of educational video analysis in content design, teaching techniques, and learning behaviors. Differently, our system visually reveals the video utilization patterns of massive video resources. Further, our system is able to analyze the video utilization from multiple perspectives, i.e., courses, students, and learning centers in the ODE context.

2.3 Analyzing Online Learning Process

To analyze students' online learning process, many studies proposed indicators which measure the interaction between students and learning resources [15, 16, 17]. Commonly used indicators include: time spent on various learning materials [18], number of sessions [19], and total count of interactions [20].

The use of detailed log data of clickstream during video interaction has been helpful for researchers in analyzing various aspects of the learning process. For example, Brinton and Chiang [21] designed an algorithm to predict student's performance on first attempt in answering questions with summary quantities of video-watching clickstream data. Uchidiuno et al. [22] found that the pausing and slowed play rate were heavily used in viewing video portions without visual aids by English language learners. In recent studies, visual analytics is applied to the analysis of clickstream data. Shi et al. [23] introduced the use of VisMOOC in helping instructors analyze learning behaviors with three linked views, which provide temporal information of click actions from different levels. Chen et al. [24] designed a new glyph to show multiple attributes of peaks and developed a visualization system called PeakVisor to investigate the peaks in clickstreams.

These studies have provided a variety of indicators related to course video for analyzing the learning patterns. However, existing indicators may not be suitable for analyzing and comparing a large number of diverse courses. On the basis of existing indicators, our indicators are designed with information of videos and courses to expand the scope of analysis, which is essential for discovering patterns of video utilization in massive courses and students.

3 PROBLEM CHARACTERIZATION

In this section, we first describe the learning process in ODE school. Then, we introduce the indicators designed for evaluating the video utilization of student and course. Finally, we summarize the analytic tasks identified with three domain experts and derive the design requirements in designing the system.

3.1 The ODE Learning Process

The entire learning process in ODE school, i.e., from entry to graduation, is similar to the undergraduate programs provided by remote learning universities (e.g., the Open University in the UK and the Athabasca University in Canada), but it is different from other forms of online learning, such as Khan Academy and MOOCs.

On one hand, the curriculum is different. The programs provided in ODE school are specialty-based education. For each specialty, students must take approximately 22 courses, including foundation courses (e.g., English and computer basics) and specialized courses which are designated in the course list of each specialty. There are about 1,348 videos of 760 hours on average for each student to watch in at least 2 years (4 semesters). Due to the number of courses, some students may spend 2 to 5 years to complete their studies.

On the other hand, the management of student registration and service is different. First, the enrollment is divided into two batches each year, usually in March and September, namely the spring batch and the autumn batch (e.g., batch 1403 represents spring batch of 2014). Students are asked to register in one of these batches. Second, due to the wide geographical distribution of students and differences in regional enrollment policies, ODE schools usually commission learning centers located in various regions to recruit students and provide several supplementary teaching services, such as organizing offline practice sessions and examinations (typically, there may be several leaning centers in a city, depending on local population, policy and educational needs). However, the quality of services at learning centers may affect the learning process of their students. Some learning centers suspended due to violation of regional policy, resulting in learning interruptions.

Table 1: General statistics of ODE dataset

Type	Items	Information	# of items
Teaching Data	Learning center	Location	91
	Specialty	Course list	11
	Course	Video list	419
	Video	Length	26,660
Learning Data	Student learning profile	Enrollment batch, specialty, learning center, course list	80,484
	Log record	Student ID, location, course, video, timestamp, etc.	100,359,271

In addition to the above differences, the learning process in ODE school is similar to other online learning. Students can access various learning resources on LMS, including videos, PPTs, textbooks, and other materials. Among these learning resources, the video is the most important one. On one hand, most of the knowledge is taught through the video. On the other hand, it takes considerable effort to produce these videos, including organizing teachers and teaching assistants to record and update videos, storage management, and purchasing network bandwidth [37].

Due the importance of video resources in online learning, the managers of ODE school need to find out whether these resources are fully utilized. Since the video utilization is a relatively objective criterion, it may serve as an important reference in the evaluation of students' learning process, video quality, curriculum design and learning center service.

3.2 Data Abstraction and Indicators of Video Utilization

We collected teaching data from the database of the teaching management system (TMS) and the learning data from the database and logging system of the LMS. The teaching data includes basic information of learning centers, the curriculum of each specialty, course data and video information related to each course. The learning data includes student's learning profiles and log data. To avoid disclosure of student privacy, the learning profiles do not include any identification information and only use a unique ID number to represent each student. The log data includes students' learning records of operations on all pages of LMS, such as actions of play, drag, pause and stop in viewing video. The student's location information used in the analysis is obtained by parsing the IP address in the log. Table 1 shows the general statistics of the data used in our case study.

As described in Section 2.3, the number of views and viewing time of videos were often used as indicators to evaluate student's learning process in previous studies [19, 20]. These indicators are applicable within a single course or some courses with similar number and duration of videos. However, the number and length of videos in different courses are significantly different. Simply using these indicators may cause misunderstanding of actual video utilization. Take the number of views as an example, there are 10 videos in course A and 40 in course B. When the video views of both courses are 20, all the videos of course A may be fully utilized, while at least half of the videos of course B are still not viewed.

To address this question, we propose the attendance rate (AR), utilization rate (UR) as supplements to existing indicators to measure student's utilization of course videos with combination of the number and duration of videos. The basic objects measured with AR and UR are the usages between a single student and a single video. The formal definition of student's video AR is as follows:

$$ar_{s,v} = \begin{cases} 1 & viewed \\ 0 & not\ viewed \end{cases} \quad (1)$$

where $ar_{s,v}$ is the attendance rate of student s with video v. If student s viewed video v, the $ar_{s,v}=1$, otherwise $ar_{s,v}=0$. The AR is similar to the concept of attendance in traditional classrooms. Each video is similar to a lesson. If a student views a video, the student "attends" the class of a lesson.

The student's video UR is as follows:

$$ur_{s,v} = \frac{wt_{s,v}}{vt_v} \quad (2)$$

where $ur_{s,v}$ is the utilization rate of student s with video v; vt_v is the length of video v, and $wt_{s,v}$ is the length of the video v watched by student s. The UR reflects the amount of video viewing. The longer students view the video, the higher their UR. Based on the $ar_{s,v}$ and $ur_{s,v}$, the AR and UR between a student and a course can are defined.

The student's course AR is defined as follows:

$$ar_{s,c} = \frac{|W_{s,c}|}{|V_c|}, W_{s,c} = \{ar_{s,v} | ar_{s,v} = 1, v \in V_c\} \quad (3)$$

where $ar_{s,c}$ is the attendance rate of student s in course c; V_c is the collection of courseware videos of course c, and $W_{s,c}$ is the collection of distinct videos watched by student s in course c. According to this definition, multiple watching of the same video is regarded only once. For example, if a course has a total of 10 videos and a student watches one of the videos 5 times, then the student's course AR is 1/10 = 0.1, instead of 5/10 = 0.5.

The student's course UR is defined as follows:

$$ur_{s,c} = \frac{\sum_{wt \in WT_{s,c}} wt}{\sum_{vt \in VT_c} vt} \quad (4)$$

where $ur_{s,c}$ is the utilization rate of student s in course c; VT_c is the collection of course video time vt of course c, and $WT_{s,c}$ is the collection of the each duration of watching videos by student s in course c. Unlike AR, the duration of repeatedly watching the same video is cumulative in UR. For example, if a course has a total duration of 10 hours and a student repeatedly watches one video 5 times for 0.5 hours each time, then the student's course UR is 0.5 x 5 / 10 = 0.25, instead of 0.5/10 = 0.05. As a result, the value of a student's course UR may be greater than 1, indicating that the student may have been watching the videos for a longer time.

With the above equations, we can obtain the video utilization of a single student in a single video or a single course. Moreover, by calculating the mean values of a set of AR and UR from the perspective of students or courses at different levels, we can further obtain the video utilization of different specialties, batches, or learning centers. For example, with the course list of a student, the total AR and UR can be calculated by $ar_s = \sum ar_{s,c}/|L_s|$ and $ur_s = \sum ur_{s,c}/|L_s|$ respectively, where L_s is the course list of student s. Moreover, the total AR and UR of a learning center can also be calculated by $ar_l = \sum ar_s/|S_l|$ and $ur_l = \sum ur_s/|S_l|$, where S_l is the list of students enrolled in learning center l and $s \in S_l$.

The combination of AR and UR not only reflects the utilization of course videos, but also reflects student engagement with the video material. As described in Section 6.1 and 6.2, some patterns can be observed from the AR distribution of videos and viewing histories of students. However, using only these two indicators does not fully explain the actual learning behaviors behind the values. We will explore the relationship between these indicators and learning behaviors and performance in the future studies.

3.3 Task Analysis

To characterize the research problem, we have collaborated with three domain experts from the ODE school of our university for a long period. One of them is an associate dean of the ODE school, who is responsible for supervising the learning centers, curriculum resource development, and student services (EA). The other two are engineers of development center. One of them is responsible for the functional development of the TMS and LMS platforms (EB), and the other one is responsible for video resource management (EC). The users of our system are the managers and staff in development center of the ODE school, including the above three experts. Their high-level goal is to understand the video utilization patterns in the ODE school, and to provide data supports to teachers and other departments in the ODE school.

In the spring of 2013, the ODE school deployed the upgraded logging module to record detailed online video viewing in LMS. Since then, a large amount of learning log data has been accumulated. Then, we designed a prototype system based on commonly used indicators (e.g., video visits, viewing time, etc.) to demonstrate the utilization of video resources from the perspective of courses. The prototype system was tested to provide basic visualization for statistical results. During this period, our domain experts found several distribution characteristics of the video viewing and factors that influence the distribution. Based on these findings, we found that the existing indicators were not suitable for comparing courses that were different in terms of number of videos, duration, and number of students, etc. Therefore, we proposed the AR and UR indicators described in Section 3.2.

Based on existing prototype system and several patterns found from log data, we conducted field studies and interviews with EA for the feedback on the system. EA proposed further analysis tasks and provided authorization to access more data (i.e., the access databases in TMS and LMS). After several rounds of interviews, we summarized the following task requirements based on feedback from our domain experts and literature review.

T.1 What is the overall utilization of all videos? Users need to get an overall intuitive impression of the video utilization, and which videos of which courses are highly utilized. Since there are many videos for each specialty and each course, it is helpful to quickly find highly utilized or characteristic videos from all videos for further analysis.

T.2 How is the video utilization of each course? Videos are organized according to courses, and both the content and order of each video are determined by the course design. Users not only need to understand the utilization of all videos in a course from different perspectives to improve the quality of video services, but also need to provide teachers with this information to update video contents and adjust course design.

T.3 How is the video utilization of each student? Domain experts are particularly interested in how students watch the videos, such as the video utilization of each course, the arrangement of learning time and how each student compares to others. They will and provide this information to teachers and teaching assistants for improving teaching and better student services.

T.4 How does each learning center perform? As described in Section 3.1, learning centers play an important role in the learning process. Therefore, users are seriously concerned about the enrollment and student services of learning centers, such as the geographical distribution of enrollments, students' video utilization,

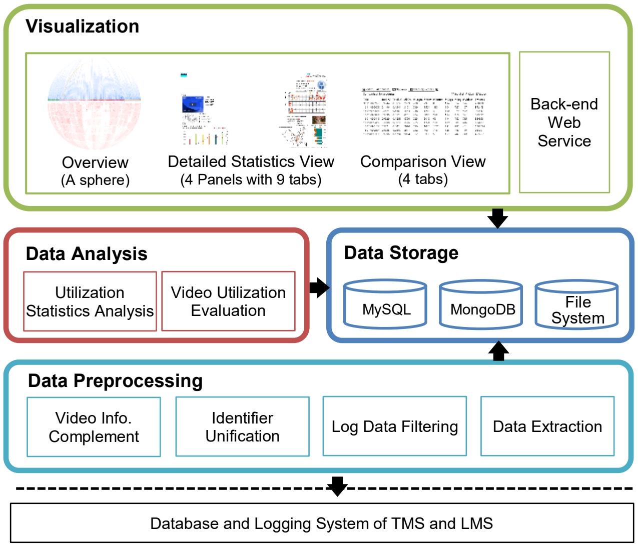

Figure 2: The architecture of VUSphere system

to identify risky learning centers that may be in violation of regional enrollment policy.

3.4 Design Requirements

Based on the tasks addressed above, we derive the following design rationales to guide our design.

R.1 Multi-scale exploration. The massive video viewing behaviors show a variety of patterns in different scales, from the overall video utilization all videos (T.1), courses and students to each one of them. To explore between different scales, multiple visualizations of the global and different scales are need.

R.2 Multi-perspective presentation. There are four types of interrelated element in the ODE learning process, including video, course, student, and learning center. Each element has different video utilization from multiple perspectives (T.2, T.3, and T.4). Therefore, our system needs to provide multiple visualization to facilitate iterative exploration via different perspective.

R.3 Comparative Analysis. With the indicators or AR and UR proposed in Section 3.2, the video utilization can be compared across multiple courses, students or learning centers. Users need to compare multiple elements at the same time to find differences for further investigation (T.2, T.3, and T.4).

R.4 Interactive exploration. Videos, courses, students and learning centers are related to each other, our system should enable users to switch back and forth between various elements at the macroscopic level and microscopic level and immediately respond to their operations.

4 System Overview

As shown in Figure 2, the architecture of VUSphere system consists of four components: data storage module, data preprocessing module, data analysis module and visualization module.

To meet the different data storage requirements of each module, we use three different storage methods in the data storage module. First, due to the large amount of disk space occupied by log files extracted from logging system of LMS, local file system is required to save these files on the analysis server. Second, referring to the data models introduced by moocRP [29] and MOOCdb [43], we integrate the data extracted from LMS and TMS and save them in the MySQL database. In addition, temporary data and intermediate results used during data analysis are also saved in this database. Third, to facilitate visualization access and flexible adjustment, the statistical results of video utilization and other indicators related to learning process are saved in the MongoDB database.

The data preprocessing module includes several sub-modules to fetch raw data from LMS and unify the various identifiers used in raw data. The amount of raw log data was voluminous, and some records were not relevant to this study (e.g., web browser type, HTTP headers etc.). Thus, we develop a log data filtering module to clean them and save the cleaned data in MySQL database. The raw teaching data and student data come from multiple systems, which maintained different identifiers for videos, courses, students and detailed information. It is difficult to conduct further analysis without a set of uniform identifiers. For example, the identifier of a course in TMS is the course code, while it is a number in LMS and a different number in log files (e.g., the course “Computer Foundation” is represented as JS001 in TMS, 160 in LMS and 4 in user log files). In addition, the FFmpeg is used to get the duration and screenshots of each course video from the stream media server of LMS. The general statistics of extracted data is listed in Table 1.

In the analysis module, due to the advantages of SQL in complex queries and dynamic debugging, we apply Python scripts that embedded SQL queries to calculate the indicators of video utilization for each video and each student. For example, to obtain the course AR for each student, we first execute the SQL query that counts the number of unique viewed videos of each course for each student in log data table to generates a dataset; then, use Python script to calculate AR on the dataset line by line according to Eqn. (3); finally, save the results into MongoDB in JSON format.

The visualization module is a web-based application, which is developed by following the design rationales described in Section 3.4. This module provides users with three interactively coordinated views to explore the video utilization of courses, students and learning centers from multiple perspectives: 1) The Overview shows the overall distribution utilization of all videos and students. 2) The Detailed Statistics View demonstrates the statistics of each video, course, student and learning center from multiple perspectives. 3) The Comparison View displays multiple elements in table and parallel coordinate to help users inspect the differences. The back-end of this module is based on Flask framework [1], and the front-end of this module implemented using vue.js [2], Three.js [3] and ECharts [4]. In addition, each chart in the Detailed Statistics View is packaged as a separate module that facilitated integration into other systems.

5 VUSphere Design

In this section, we illustrate the visual design of the VUSphere system and the interaction among the views. As shown in Figure 2, there are three coordinated views to show the utilization of video resources from different perspectives: The Overview, the Detailed Statistics View, and the Comparison View.

5.1 The Overview

To provide users with the intuitive impression on overall utilization and distribution of course videos and student population (T.1), we design a fixed radius sphere to present all videos and students on the spherical surface. Users can observe the distribution and utilization of all course videos before selecting certain elements for conducting further analysis (R.1). As discussed in previous section, we use indicators of AR and UR to evaluate the utilization of video resources and the engagement of students. As shown in Figure 1(a), the video elements are distributed on the surface of the upper hemisphere and student elements are on the surface of the lower hemisphere.

As described in Section 3.3, users need to know not only the utilization of the entire course videos, but also the distribution of video utilization in one course (T.2). To address these questions, we take advantage of the specific attribute information of course

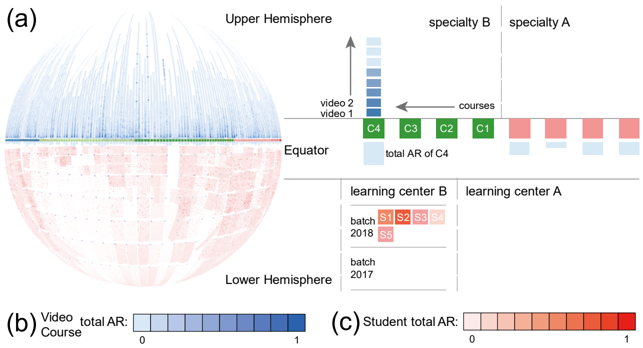

Figure 3: The design of Overview. (a) The spherical layout to place the elements of videos, courses, and students on the sphere surface. (b) The color encoding of total AR for video and course. (c) The color encoding of total AR for student.

video to place each video element on the sphere surface. Since we choose a fixed radius sphere to present elements, only the latitude and longitude of each element need to be calculated according to its own attributes.

First, each course is represented as a square, which is color-coded with the specialty number. Since all course elements are placed on the equator, the latitude of each course element is the same. The longitude is calculated by the following method. 1) All courses are grouped by specialties (one public course group and eleven specialty groups). 2) All groups are sorted by specialty ID, and courses in each group are sorted according to the specialty curriculum. 3) According to the group order and the order within the group, all course elements are evenly filled to the equator so that each course element has a unique longitude (Figure 3(a)). In addition, the video utilization of each course is represented as a bar under the course square and the bar height is mapped to course AR.

Second, each video is represented as a blue fixed-size rectangle, which is color-coded with its total AR (the darker the higher AR, as shown in Figure 3(b)). It is distributed on the surface of the upper hemisphere as the following rules: First, since each video belongs to only one course, the longitude of the course element is used as the longitude of the video element. Second, the video element is distributed on a longitude line according to its teaching sequence in course, from equator to north pole. The sequence number is mapped to the latitude (Figure 3(a)). To observe the overall distribution of video utilization, user can move the camera to the south pole of the sphere by dragging the mouse. When viewing this sphere from the bottom perspective, all spherically distributed video elements on the upper hemisphere will be projected onto the equatorial plane without overlapping (Figure 8(a)).

Third, each student is represented as a red fixed-size rectangle, which is color-coded with his/her total AR (the darker the higher AR, as shown in Figure 3(c)) located on the surface of lower hemisphere (Figure 3(a)). We use the learning center code, enrollment batch and sequence number of registration for calculating coordinates. First, the surface of the lower hemisphere is divided into several longitude bands according to the number of learning centers. Second, according to the chronological order and number of enrollment batches, each longitude band is further divided into several latitude areas from equator to the south pole. Third, in each latitude area, students are arranged one by one from west to east, row by row from north to south in ascending order of the number in their student ID.

We designed the Overview by working with our domain experts and users. In the early prototype system, we arranged all videos according to the specialty and course order (the layout was similar to Figure 8(c)). Users could zoom in and drag the view to see the video of interest. However, they found that the process of exploring

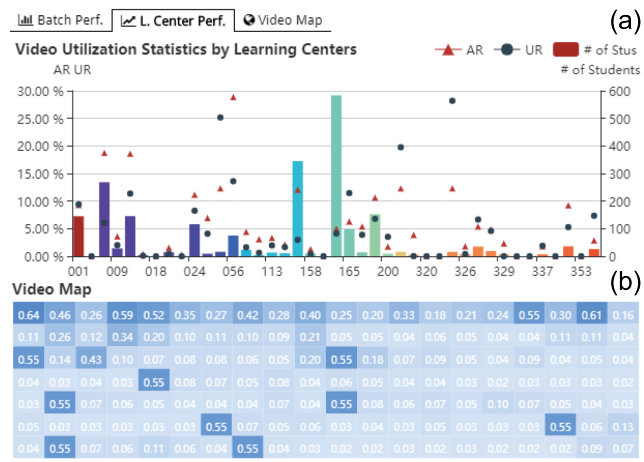

Figure 4: The course panel in the Detailed Statistics View. (a) The learning center performance tab presents the statistics of video utilization of a course by learning center. (b) The video map tab displays the total AR of each video in a course.

videos was a bit tedious. Due to the large number of videos, the size of each video rectangle was very small. Users had to constantly zoom in and out of the view to click different videos. Inspired by "Focus + Context" [44] design idea we developed a spherical surface layout (Figure 3(a)) to organize all videos and students. Although we chose a 3D-like spherical surface layout, all elements were distributed on a limited surface rather than in space, which made this layout more like a 2D layout. In addition, since the size of each video rectangle was fixed, and the focus was fixed in the center of the screen, the distortion of each element on the surface was predictable. Our users stated that operating this sphere was similar to using a globe to view the world map.

5.2 Detailed Statistics View

Users can obtain general expression of the utilization distribution of videos from the Overview. However, they need to understand the detailed information of the statistic results on utilization of each video, student and learning center (T.2, T.3, and T.4). Therefore, the Detailed Statistics View is designed for analyzing the detailed video utilization of each element from multiple perspectives (R.2). In this view, four panels are designed and placed on both sides of the Overview (Figure 1(b), (c), (d) and (e)), providing basic information and multi-perspective statistical results of videos, courses, learning centers and students.

5.2.1 Video Panel

The video panel provides basic information and utilization statistics of a video (Figure 1(b)). The AR and UR displayed in this panel is the total utilization of a video. To facilitate viewing the video content directly for users, we add a HTML5 video player on left of this panel, refer to the content-based view design in the VisMOOC [23, 30]. Users can directly view the video of interest, such as videos with higher AR.

5.2.2 Course Panel

The course panel (Figure 1(c)) provides basic course information and statistics of video utilization from multiple perspectives in three tabs (T.2).

First, in the *batch performance tab* (Figure 1(c)), the line chart shows the statistical results of video utilization by all students in each batch, and the bar chart shows the number of students attended in this course. Each bar is color-coded with the batch number.

Second, in the *learning center performance tab*, the statistics of video utilization of a course by regional learning centers is shown

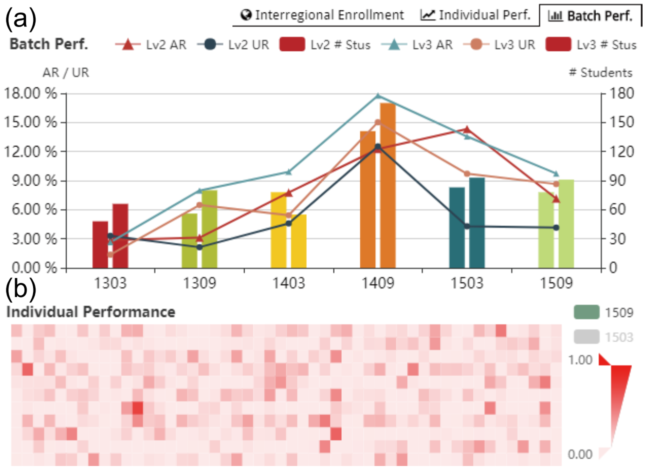

Figure 5: The learning center panel in the Detailed Statistics View. (a) The batch performance tab presents the average video utilization of each batch in a learning center. (b) The individual performance tab displays the total AR of each student in a learning center.

(Figure 4(a)). The bar chart shows the number of students from each learning center, and the marks of triangle and circle show the statistics on AR and UR respectively. Users can explore each learning center by clicking on the data point in this tab.

Third, in the *video map tab* (Figure 4(b)), a video heatmap represents the overall utilization of each video through the same color encoding in the Overview (Figure 3(b)), and the value in each cell represents the total AR of each video. For example, the value 0.64 in first upper-left cell indicate that 64% of students who selected this course have watched this video more or less. While the charts in the first two tabs refer to the domain experts' monthly work report, the visual design of this chart is inspired by the heatmap matrix used in the iForum system [32] and heatmap demos in ECharts [4]. Each cell in this chart represents a video in this course and all cells are arranged in 20 cells per row, from left to right and top to bottom according to the teaching sequence. For instance, the first cell in the first row is the 1st video in a course, and the first cell in the second row is the 21st video.

5.2.3 Learning Center Panel

In the learning center panel, users can evaluate the performance of a learning center from three perspectives (T.4).

First, as shown in figure 5(a), in the *batch performance tab*, a combination of line chart and bar chart represents the statistics results of video utilization of the students in this learning center.

Second, the *individual performance tab* (Figure 5(b)), the video utilization of each student in each batch is presented in the heatmap, in which each cell uses the same color encoding in the Overview (Figure 3(c)). To facilitate in-depth analysis, this heatmap supports AR filter (i.e., showing students within a specified AR range) by selecting sliders on the bottom right corner of this tab. The design considerations for the first two tabs are similar to the tabs in course panel, while the order of cells in this tab is not the same. Each cell in this chart represents a student in a specific batch registered in this learning center, and all cells are arranged in 50 cells per row, from left to right and top to bottom in ascending order of the number in their student ID.

Third, the *Interregional enrollment tab.* Refer to the map visualization techniques used in the MoocViz [28] and VisMOOC [30], we add this tab to count the number and proportion of students come from each province. As shown in Figure 1(d), each circle marker on the map indicates the enrollment of a province. The

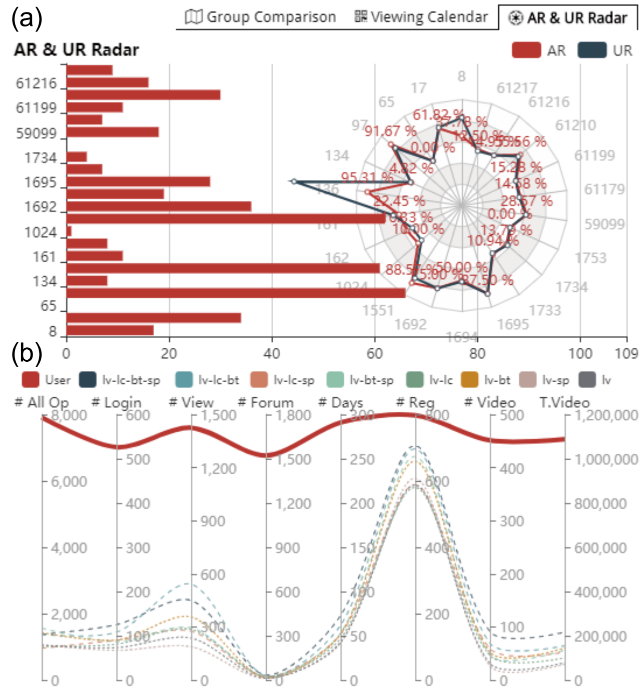

Figure 6: The student panel in the Detailed Statistics View. (a) The AR & UR radar tab shows a student's video utilization of each course. (b) The group comparison tab compares the statistics of a student (the red solid line) with several reference groups.

number of students from this province is mapped to the size of the marker. We place a green triangular marker on the map to represent the location of the learning center, which is connected to other markers with curves to emphasize the enrollment sources. A summary of cross province enrollment is added as a bar chart on right of the map, where the width of the orange bar maps to the total number of migration enrollment.

5.2.4 Student Panel

To achieve this further exploration of individual video utilization (T.3), we design the student panel to present statistics from different aspects. In the upper part of the student panel, student's basic information (student ID, enrollment batch, rough geographic location and specialty) and total statistics of video utilization are listed (Figure 1(e)). In the lower part, three tabs are integrated to show detailed statistics.

First, the *AR & UR radar tab.* As shown in Figure 6(a), the numbers of videos viewed for each course by the student are displayed as a bar chart on the left, where the y-axis is the course ID list, arranged in the learning order. When users hover over a bar of one course, the tooltip will display the course name and detailed information. The AR and UR of each course are shown in a radar-like chart on the right side, which are calculated according to Eqn. (3) and Eqn. (4). We design this chart to compare the differences between AR and UR for each course of the student. Referring to the evaluation of radial visualization solutions [45] and considering that there are only two series (ARs and URs for all courses, respectively), this radar-like design is chosen. However, users stated that this design was not very useful since each student had a lot of courses. In addition, most students had low AR and UR for each course, which made the series in the chart too dense to see.

Second, the *viewing calendar tab.* To analyze the temporal characteristics of student's learning history, we design a calendar-based scatter plot to present the number of student's daily online activities since registered in ODE school (Figure 1(e)). As shown in the viewing calendar tab, each circle marker represents the student's activities on the day in calendar, and the number of activities is mapped to the size of marker. Since we mainly focus on the video utilization of student, two colors are used to encode two types of activities in this diagram: red for video-related activities and yellow for all activities. Obviously, video-related activities are included in all activities, so the size of the yellow marker is larger than red marker. The visual design of this chart is inspired by the calendar view in the VisMOOC [23] which shows the popularity of a selected video day by day and provides events animation to show viewing patterns. Since we need to show the overall distribution of a student's activities in long-term learning process, the date range of the calendar is extended to 3 years (most students will finish their studies in 3 years), and only the total amount of two types of activities are displayed.

Third, the *group comparison tab.* As shown in Figure 6(b), to help users compare a student's utilization of videos and system functions with a particular student group (T.3), we choose parallel coordinate to show the multiple indicators. We refer to the design of the feature statistics view in NetworkSeer [31], in which the parallel coordinate is used to query a specific group of students. However, since the groups are already specified by our domain experts during the interview, the number of series on this chart is limited. In addition, the series switcher and brushing function on this chart helps to filter groups that are not of interest. According to the statistics in early prototype system and interviews with our domain experts, we select eight statistical indicators as the parallel axes. The first four axes represent the number of various types of operation, including all operation, login times, video related operation (play, pause, stop, drag and drop progress bar, etc.), and forum related operations (view, write, reply, and delete posts etc.); the next two axes reflect the number of days the student is actually online and the accumulated days since the registration in learning center; the last two axes represent the total number of distinct videos viewed by students and the total cumulative duration.

5.3 Comparison View

To compare the difference between individual students, as well as videos, courses and learning centers side-by-side, we develop the Comparison View (Figure 1(f)) that includes four tabs to display multiple elements with their attributes (R.3).

As shown in Figure 1(f), in the Comparison View, there are four tabs to compare videos, courses, students, and learning centers, respectively. In each tab, there is a table to display the elements that needs to be compared, while each row represents an element, and each column represents an attribute. The columns of each table are decided by the elements to be compared. Users are familiar with this kind of spreadsheet-like presentation, which is consistent with their experience of using office software in daily work. In addition to tables, we also provide parallel coordinates to show the differences of the same elements in each table. However, domain experts and other users found it was difficulty to use, especially in such a small area. Therefore, we add an export button in the upper right corner of each tab for exporting the tabular data in csv format. Users can use other analysis software to make further exploration based on the exported data.

Each element displayed in Comparison View is linked with the same element in the Detailed Statistics View. On one hand, when clicking the "+" button next to the title of element in each panel of the Detailed Statistics View, the clicked element will be added to the corresponding tab in the Comparison View. On the other hand, when clicking the title column of each element in the Comparison View, the relevant tab in the Detailed Statistics View will update its content.

5.4 Interactive Exploration

To enable users to explore the utilization of different elements from multiple aspects, the interaction among panels and views is carefully designed (R.4). First, all visualization views in VUSphere are interactively coordinated by clicking hyperlinks, which enables exploration among videos, courses, students and learning centers from multiple perspectives. When users click any rectangle in the Overview, the corresponding panel in the Detailed Statistics View will be opened and display related information. For instance, when a blue video rectangle in the Overview is clicked, the video panel and related course panel in the Detailed Statistics View will be opened or updated. If users check the utilization of each video in a course by clicking the blue cell in the video map tab of the course panel (Figure 4(b)), the view in the Overview will be automatically adjusted to the corresponding video rectangle and the video panel will be updated. Similar to the interaction of video elements above, the Overview and other panels (course panel, student panel and learning center panel) are also coordinated, allowing users to explore of video utilization in different perspectives.

Second, we provide several keyboard shortcuts and assist functions to help users reduce repetitive operations. For example, a set of layer switcher to remove/add layers of elements or panels; a "Auto Pilot" function that automatically presents information of different elements to help users browse data; and a "Cart" function for adding element in the Overview or other panels to the Comparison View continuously.

6 Case Study

To assess the effectiveness and usefulness of VUSphere system, we conducted case study with the same three domain experts. We deployed the data preprocessing and analysis modules on our private cloud server with 2.67GHz Intel X5650 CPU and 40 GB memory. With the data preprocessing module, we first collected teaching data and learning data to our storage, unified the identifiers, and removed irrelevant attributes and records in log data. The cleaned data was saved in MySQL database and the statistics are shown in the Table 1. Then, we conducted data analysis with EB and EC. During the analysis with VUSphere system, we identified serval patterns of video utilization with our domain experts. These patterns are classified into three categories according to the tasks described in Section 3.3.

6.1 Overall Utilization of Videos and Courses

The overall utilization of all course videos is presented in the Overview, as shown in Figure 7(a). The total AR of most videos is less than 0.1, i.e., these videos are viewed by less than 10% of students who register the course, which indicates that the video resources are not fully utilized by students. With the help of the early prototype system, domain experts have found that the number of video views are significantly different in courses. For example, some videos are rarely viewed, and the public fundamental courses have much higher video views than specialty courses. However, the total AR of each course shows that the video utilization of public fundamental courses is not high and may even be lower.

Take a public fundamental course "English II" as an example, more than 10,000 students registered this course in each batch and the number of views for each video in course was relative higher. However, the total AR of this course was less than 8% in each batch. EB stated that these findings quantified his usual experience: The log statistics of the LMS showed that lots of videos in a course were rarely viewed. EB believed that there were multiple reasons for this result, which may include: the video contents of this course exceed the examination requirements. EB explained that the course videos of each specialty were usually created with reference to existing curricula for full-time students in college, whereas the teaching

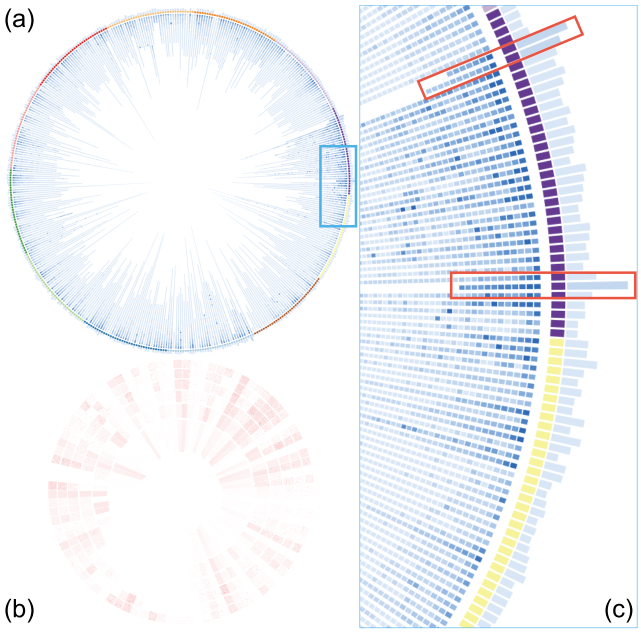

Figure 7: Using the Overview to show all videos and students. (a) The overall distribution of all videos and courses. (b) The overall distribution of all students (the darker the larger enrollment). (c) The magnified area of the blue box in Figure 7(a), in which there are some courses with higher total AR (marked with red rectangle)

goal for ODE students was relatively lower than full-time students. Watching only some of the videos will guarantee students pass the exams and earn credit. The low utilization of video resources was not only a waste of investment, but also may imply that there was a mismatch between current videos and the needs of students. Thus, EB considered providing this visualization to teachers for adjusting the video contents of a course.

In addition, EC found that some videos have higher total AR than others, including: 1) the first several videos in the course (those dark blue rectangles near the outside in Figure 8(c)); 2) most of the videos in some specific courses, such as those marked with red rectangle in (Figure 7(c)).

EC explained that the first finding may be related to student's general learning sequence: students usually view the first several videos according to the outline of the course. Whereas, when there are too many videos, they may only watch the videos about the exam topics (e.g., the videos with total AR>0.5 shown in Figure 4(b)). For the second finding, after examining the basic information and detailed utilization statistics of these course, EC found that these courses are certification training courses, i.e., the special course for passing the certification exams. For example, passing the national unified English examination is a necessary condition for students to obtain a college degree. Therefore, student may pay more attention to this short-term training to pass the certification examination. EC commented that for these videos with such high AR, the CDN services can be further applied to accelerate access and improve the learning experience.

In addition, EA used our system and showed great interest in exploring the Overview and course panel to compare video utilization between courses. Both EB and EC found that the video map tab may be useful for teachers to adjust their course structure and update the video content, and EB considered integrating the video map tab into the dashboard of TMS for teachers. All three experts appreciated the visual design of the Overview and the panels of video and course, they decided to develop a special version without personal identical information for online presentation of video utilization in their office.

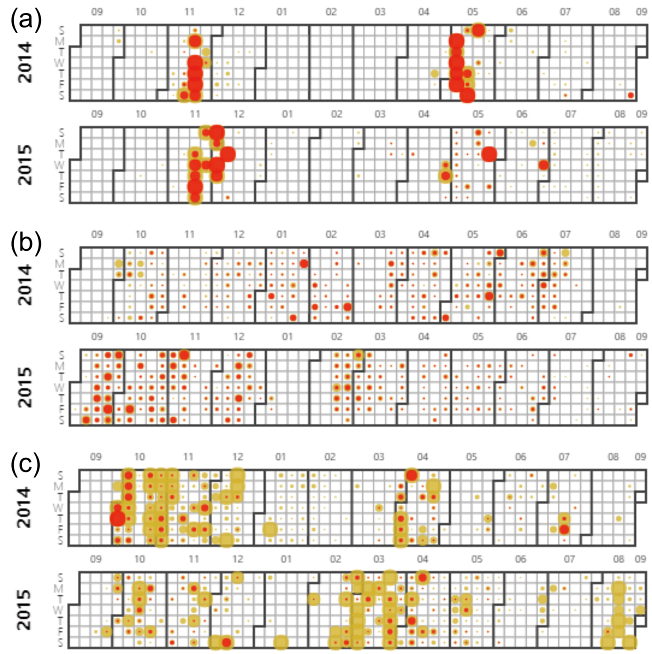

Figure 8: Three cases of typical learning history patterns of students found in the viewing calendar tab. (a) “cram session”. (b) “studying every day”. (c) “forum activist”.

6.2 Student’s Video Utilization

With the Overview and the learning center panel in the Detailed Statistics View, our experts selected some students and used the student panel to further analyze their video utilization. The findings are discussed and summarized from two perspectives as follows.

Video utilization of each course. EC first viewed the overall AR distribution of all students in the Overview and found that most of the students had low AR and there were many students with AR=0 (Figure 7(b)). Then, EC selected some students with low AR and some with high AR to further view their video utilization in each course. EC found that students with high AR usually viewed more videos in most courses, and their UR was greater than AR, while students with low AR had less UR than AR. EC explained that this result may reflect student’s learning process. On one hand, UR<AR indicates that the student may quit study without watching the video completely; on the other hand, UR>AR indicate that the student may repeatedly watch the video. However, among the students with UR>AR, most students’ UR didn’t exceed 2 times that of AR, while some students exceeded 4 times. Both experts were interested in what kinds of learning process and motives may produce such different patterns of AR and UR. In addition, EC commented that AR and UR may help teachers to understand student's online learning process, but further analysis was needed to uncover the reasons behind various patterns of AR and UR.

Arrangement of learning time. While analyzing student's AR and UR of each course, EC also viewed the viewing history of each student in the viewing calendar tab. Several interesting patterns were found including the following typical cases: First, the "cram session" pattern. These students rarely view videos except for the exams. Figure 8(a) shows a typical learning history of student with this pattern (the final exams for the spring and fall semesters are in June and December, respectively). Second, the "studying every day" pattern (Figure 8(b)). These students insist on learning every day and usually have relatively high AR. Third, the "forum activist" pattern (Figure 8(c)). These students prefer to participate in the forum discussions compared to viewing videos. EC commented the process of exploring patterns of learning history was interesting, but manually reviewing each student’s learning history one by one

Figure 9: A case of abnormal learning center. (a) The interregional enrollment (The orange bars represent the number of interregional enrollment). (b) The total AR of each student in this learning center.

was tedious. Especially in the identification of the third pattern, EC had to combine with the statistics in the group comparison tab to find the difference from others (Figure 6(b) shows that the student's forum activities far exceed average), which made the exploration process more complex. Both EB and EC considered integrating the viewing calendar tab into the dashboard of LMS to help students understand their learning progress.

6.3 The Learning Center's Performance

Both EB and EC explored the statistics of interregional enrollment and the distribution of students' video utilization in each learning center one by one. They found that the interregional enrollment in several learning centers were abnormal. Take the learning center shown in Figure 9 as an example, the experts examined the interregional enrollment (Figure 9(a)) and found that a large proportion of students came from the provinces outside the province where the learning center located. Then, they found the low AR of most students in this learning center (Figure 9(b)) and confirmed their suspicions. EC explained that it was normal for a learning center to have students from other provinces (e.g., the learning centers shown in Figure 1(d) and Figure 5(b)), but such a large proportion of interregional enrollment and poor video utilization of all students were abnormal, which may imply that management issues existed in this learning center. The experts reported to the managers of our ODE school about the abnormal learning centers with our visualization in their detailed reports.

EA showed great interests in exploring interregional enrollment and commented that examining the statistics of each learning center required too much time in current system. He decided to develop an independent overview of interregional enrollments over all learning centers to identify abnormal learning centers.

7 Domain Expert Interview

We performed interviews with five experts to further evaluate the effectiveness of VUSphere system. Among them, three experts from our ODE school have been collaborated with us from the early prototype system design to the data analysis based on the current system. Thus, they have prior experience with our system. The other two are education experts from our university and they haven't seen our system before (EE1 and EE2). The interview process took about 60 minutes. In the first 20 minutes, we introduced the research background and demonstrated the system. Afterwards, experts were free to explore the system and ask questions at any time. We collected their feedback and discussed as follows.

First, all experts commented that the Overview was useful in observing the distribution of all videos, but some rectangles were very close and it is difficult to accurately click. Similar problems were also found in the parallel coordinates of the Comparison View and the AR & UR radar tab of the Detailed Statistics View, where elements were too dense to conduct further analysis. To resolve this problem, we plan to add a set of filters in the Overview and the Detailed Statistics View to reduce the elements for visualizing. In addition, we consider improving these visual designs based on detailed evaluation results after the formal usability test.

Second, EE1 and EE2 commented that the process of mining learning behavior patterns manually was cumbersome, especially when the number of students was very large. They suggested to conduct a cluster analysis on students' learning histories to obtain a small amount of "behavioral patterns" for further analysis.

Last, EE1 and EE2 suggested two general directions for further development of our system. First, as described in the case study, the experts identified several temporal patterns of student video utilization from the learning history. EE1 suggested that we can analyze the relationship between these patterns and students' exam scores. Second, EE2 suggested combining the temporal analysis with other indicators (e.g., AR, UR and other commonly used indicators) to further explore other learning behaviors.

8 Conclusion and Future Work

In this paper, we presented VUSphere as a visual analytics system to analyze the video utilization with the proposed two indicators. Based on collaborations with the domain experts from our ODE school, several learning patterns of video utilization related to courses, students and learning centers were identified.

In the future, we plan to extend our system with the following two features. First, we consider exploring the relationship between the video utilization and academic performance to provide reference for learning quality evaluation. Second, since current analysis didn't reflect the change of student's video utilization over time, we would like to conduct temporal pattern analysis on the video utilization. Further, we plan to conduct a formal usability evaluation for improving the visual design of our system.

Acknowledgments

The authors would like to thank Dr. Ling Chen, Dr. Rui Li, Dr. Haipeng Du, Prof. Dehai Di, Prof. Feng Tian and Prof. Chen Li in XJTU for participating this project as domain experts, Prof. Buyue Qian for his constructive suggestion of revising paper and the anonymous reviewers for their valuable comments. This research was partially supported by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" under the National Key Research and Development Program of China with Grant No. 2018YFB1004500, the MOE Innovation Research Team No. IRT17R86, the National Science Foundation of China under Grant Nos. 61721002, 61502379, 61532015, and Project of China Knowledge Centre for Engineering Science and Technology.

References

[1] Flask. http://flask.pocoo.org/.
[2] Vue.js. https://vuejs.org/.
[3] Three.js. https://threejs.org/.
[4] ECharts. http://echarts.baidu.com/.
[5] L. Cagliero, L. Farinetti, M. Mezzalama, E. Venuto, and E. Baralis, "Educational video services in universities: A systematic effectiveness analysis," in *2017 IEEE FIE*, 2017, pp. 1–9.
[6] C.-M. Chen and C.-H. Wu, "Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance," *Computers & Education*, vol. 80, pp. 108–121, 2015.
[7] P. J. Guo, J. Kim, and R. Rubin, "How video production affects student engagement: an empirical study of MOOC videos," in *Proceedings of the First (2014) ACM Conference on Learning @ Scale*, 2014, pp. 41–50.

[8] F. Van der Sluis, J. Ginn, and T. Van der Zee, “Explaining Student Behavior at Scale: The Influence of Video Complexity on Student Dwelling Time,” in *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, New York, NY, USA, 2016, pp. 51–60.
[9] MOOCs@Edinburgh Group, “MOOCs @ Edinburgh 2013 – Report #1,” University of Edinburgh, May 2013.
[10] T. Mullaney and J. Reich, “Staggered Versus All-At-Once Content Release in Massive Open Online Courses: Evaluating a Natural Experiment,” in *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 2015, pp. 185–194.
[11] I. Chuang, “HarvardX and MITx: Four Years of Open Online Courses -- Fall 2012-Summer 2016,” *SSRN Electronic Journal*, 2016.
[12] G. Kovacs, “Effects of In-Video Quizzes on MOOC Lecture Viewing,” in *Proceedings of the Third (2016) ACM Conference on Learning @ Scale*, New York, NY, USA, 2016, pp. 31–40.
[13] Q. Zhao, S. Varma, and J. A. Konstan, “In-Video Reuse of Discussion Threads in MOOCs,” in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, 2017, pp. 153–156.
[14] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, G. Alexandron, and D. E. Pritchard, “Using Machine Learning to Detect ‘Multiple-Account’ Cheating and Analyze the Influence of Student and Problem Features,” *IEEE Trans. on Learning Technologies*, vol. PP, no. 99, pp. 1–1, 2017.
[15] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz, and J. C. Núñez, “Students’ LMS interaction patterns and their relationship with achievement: A case study in higher education,” *Computers & Education*, vol. 96, pp. 42–54, 2016.
[16] L.-Y. Li and C.-C. Tsai, “Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance,” *Computers & Education*, vol. 114, pp. 286–297, Nov. 2017.
[17] A. Peña-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works,” *Expert Systems with Applications*, vol. 41, no. 4, Part 1, pp. 1432–1462, Mar. 2014.
[18] L.-Y. Li and C.-C. Tsai, “Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance,” *Computers & Education*, vol. 114, pp. 286–297, Nov. 2017.
[19] J. W. You, “Identifying significant indicators using LMS data to predict course achievement in online learning,” *The Internet and Higher Education*, vol. 29, pp. 23–30, Apr. 2016.
[20] S. Joksimović, D. Gašević, T. M. Loughin, V. Kovanović, and M. Hatala, “Learning at distance: Effects of interaction traces on academic achievement,” *Computers & Education*, vol. 87, pp. 204–217, Sep. 2015.
[21] C. G. Brinton and M. Chiang, “MOOC performance prediction via clickstream data and social learning networks,” in *2015 IEEE INFOCOM*, 2015, pp. 2299–2307.
[22] J. Uchidiuno, J. Hammer, E. Yarzebinski, K. R. Koedinger, and A. Ogan, “Characterizing ELL Students’ Behavior During MOOC Videos Using Content Type,” in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, 2017, pp. 185–188.
[23] C. Shi, S. Fu, Q. Chen, and H. Qu, “VisMOOC: Visualizing video clickstream data from Massive Open Online Courses,” in *2015 IEEE Pacific Visualization Symposium (PacificVis)*, 2015, pp. 159–166.
[24] Q. Chen, Y. Chen, D. Liu, C. Shi, Y. Wu, and H. Qu, “PeakVizor: Visual Analytics of Peaks in Video Clickstreams from Massive Open Online Courses,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 10, pp. 2315–2330, Oct. 2016.
[25] H. Q. Yu, C. Pedrinaci, S. Dietze, and J. Domingue, “Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning,” *IEEE Transactions on Learning Technologies*, vol. 5, no. 2, pp. 130–142, Apr. 2012.
[26] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller, “Understanding in-video dropouts and interaction peaks in online lecture videos,” in *Proceedings of the First (2014) ACM Conference on Learning @ Scale*, 2014, pp. 31–40.
[27] L. Cagliero, L. Farinetti, M. Mezzalama, E. Venuto, and E. Baralis, “Educational video services in universities: A systematic effectiveness analysis,” in *2017 IEEE Frontiers in Education Conference (FIE)*, 2017, pp. 1–9.
[28] F. Dernoncourt, C. Taylor, K. Veeramachaneni, S. Wu, S. Halawa, and U.-M. O’Reilly, “MoocViz: A Large Scale, Open Access, Collaborative, Data Analytics Platform for MOOCs,” in *NIPS 2013 Education Workshop*, 2013.
[29] Z. A. Pardos and K. Kao, “moocRP: An Open-source Analytics Platform,” in *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 2015, pp. 103–110.
[30] H. Qu and Q. Chen, “Visual Analytics for MOOC Data,” *IEEE CG &A*, vol. 35, no. 6, pp. 69–75, Nov. 2015.
[31] T. Wu, Y. Yao, Y. Duan, X. Fan, and H. Qu, “NetworkSeer: Visual analysis for social network in MOOCs,” in *2016 IEEE Pacific Visualization Symposium (PacificVis)*, 2016, pp. 194–198.
[32] S. Fu, J. Zhao, W. Cui, and H. Qu, “Visual Analysis of MOOC Forums with iForum,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 1, pp. 201–210, Jan. 2017.
[33] G. K. W. Wong, S. Y. K. Li, and E. W. Y. Wong, “Analyzing academic discussion forum data with topic detection and data visualization,” in *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering*, 2016, pp. 109–115.
[34] Y. Chen, Q. Chen, M. Zhao, S. Boyer, K. Veeramachaneni, and H. Qu, “DropoutSeer: Visualizing learning patterns in Massive Open Online Courses for dropout reasoning and prediction,” in *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016, pp. 111–120.
[35] D. Zhou, H. Li, S. Liu, B. Song, and T. X. Hu, “A map-based visual analysis method for patterns discovery of mobile learning in education with big data,” in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 3482–3491.
[36] C. Crook and L. Schofield, “The video lecture,” *The Internet and Higher Education*, vol. 34, pp. 56–64, Jul. 2017.
[37] F. M. Hollands and D. Tirthali, “Resource requirements and costs of developing and delivering MOOCs,” *The International Review of Research in Open and Distributed Learning*, vol. 15, no. 5, Oct. 2014.
[38] X. Ding, J. Niu, and Y. Han, “Research on distance education development in China,” *British Journal of Educational Technology*, vol. 41, no. 4, pp. 582–592, 2010.
[39] K. Lee, “Rethinking the accessibility of online higher education: A historical review,” *The Internet and Higher Education*, vol. 33, pp. 15–23, Apr. 2017.
[40] MOE of PRC, “Action Scheme for Invigorating Education Towards the 21st Century,” 24-Dec-1998. [Online]. Available: http://www.moe.gov.cn/jyb_sjzl/moe_177/tnull_2487.html. [Accessed: 23-Jun-2017].
[41] NBS of PRC, “National Data,” 2016. [Online]. Available: http://data.stats.gov.cn/english/. [Accessed: 28-Jun-2017].
[42] W. R. Doyle and B. T. Skinner, “Estimating the education-earnings equation using geographic variation,” *Economics of Education Review*, vol. 53, pp. 254–267, Aug. 2016.
[43] V. Kalyan, D. Franck, T. Colin, P. Zachary, and O. Una-May, “MOOCdb: Developing Data Standards for MOOC Data Science,” in *16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings*, 2013, p. 17.
[44] Munzner, T. (2015). *Visualization analysis and design*. Boca Raton: CRC Press, Taylor & Francis Group, CRC Press is an imprint of the Taylor & Francis Group, an informa business.
[45] Albo, Y., Lanir, J., Bak, P., & Rafaeli, S. (2016). Off the Radar: Comparative Evaluation of Radial Visualization Solutions for Composite Indicators. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), 569–578.
[46] Moodle Analytics Plugins. https://moodle.org/plugins/?q=analytics
[47] Blackboard Analytics. http://www.blackboard.com/education-analytics/index.html