The Visual Computer
https://doi.org/10.1007/s00371-020-01818-4

ORIGINAL ARTICLE

Multi-level and multi-perspective visual correlation analysis between general courses and program courses

Lianen Ji[1] · Yaming Yuan[1] · Fang Gao[2]

Abstract

Exploring the potential impact of important general courses on program-specific courses in universities can help to improve the entire teaching and learning process for an academic major. However, the large number of courses and multiple factors affecting students' course grades makes it difficult to reveal and analyze the complicated relationship between the two types of courses only from a single perspective or at a single level. Thence, this paper starts with analysis of historical course grades data within an undergraduate program and then presents an interactive visual analytic system, MVCAS, which is designed to demonstrate and explore the various correlations between these two types of courses at different levels and from different perspectives. The major contributions of this work include: (1) a multi-angle preprocessing of course grades data, including decomposition, extraction and conversion; (2) multiple coordinated analysis views which make it possible to effectively explore the overall, categorical and pairwise course correlations and further link courses with students, instructors and semesters together; and (3) a top-down correlation analysis process for general courses and program ones. The effectiveness and usefulness of MVCAS have been preliminarily demonstrated through a case study, in which the field experts use this tool to investigate different levels of correlations between the focused mathematics and program-specific courses in a computer science major comprehensively.

Keywords General and program courses · Multi-perspective · Multi-level · Correlation · Visual analytics

1 Introduction

In today's educational institutions, although there are different undergraduate program settings and different curriculum compositions, especially in different regions of the world, it is still a common and continued pursuit of each major to improve the whole program's quality of teaching and learning. For such a science or engineering degree program, there are usually dozens of courses in its curriculum. These courses can be further classified into different groups or categories according to their roles. For example, in most of Chinese universities, they are commonly divided into two different parts, i.e., the general part and specific part. The former includes general or basic courses like mathematics, college physics, foreign languages and cultures, etc. The latter includes all of the program-specific courses. Except for some free elective courses, most of general courses are scheduled in the first or second year during 4 academic years. Normally students are requested to take these general courses and some program courses in a relative fixed sequence. Against such a background, exploring the potential correlations between these two parts can provide advices from multiple perspectives, such as helping the major-level administrators to analyze and predict student grades in program courses, optimize the curriculum structure and adjust the teaching plan, and offering instructors more references for guiding students' studying process and optimizing course content continuously. Although there have been some studies exploring various factors affecting students' performance in college courses [1–3], the interplay between different types of courses, especially between general parts and specific parts

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00371-020-01818-4) contains supplementary material, which is available to authorized users.

✉ Lianen Ji
jilianen@cup.edu.cn

Yaming Yuan
664602532@qq.com

1 Beijing Key Laboratory of Petroleum Data Mining, Department of Computer Science, China University of Petroleum, Beijing, China

2 Research Institute of China Mobile Communications Company Limited, Beijing, China

Published online: 14 February 2020

of a degree program's curriculum, has not been demonstrated enough.

Traditionally, analyzing course–course correlations mainly relies on some statistical methods, including Pearson's coefficient and regression model, which are performed in an automatic process [4,5]. However, due to the lack of enough interactive control on the analytical process and less effective illustration of analysis results, these methods alone are not enough to support comprehensive analysis of multiple factors influencing students' course grades and also cannot allow for flexible exploration of various correlations between courses or course groups from multiple perspectives. By contrast, visual analytics methods employ various visual presentations, multi-view collaboration and fusion with statistics to establish a more intuitive, sensitive and effective correlation model [6–8]. In this way, analysts can interactively mine the inner association patterns hidden behind the attributes of multivariate course data and further analyze the potential causes.

In recent years, visual analytics has also made some progress in exploring educational data. For example, the exploration of MOOC (Massive Open Online Course) forum data has helped instructors to discover and understand temporal patterns in order to enhance the course outcome and prepare the next release [9,10]. In a previous system called SPVAS, we have made some attempts to study potential grade patterns implied in the traditional course data, reveal the course or student anomalies, and reason the possible causes [11]. However, to fully support university administrators and instructors to perceive and identify correlation patterns between different types of courses, the number of courses and course types required to be analyzed will grow rapidly. Apart from that, a more comprehensive analysis of the association between general courses and program-specific courses also should be made at multiple levels of course scales such as the overall, categorical and pairwise, and from multiple perspectives such as students, courses, instructors and semesters. Therefore, these will pose more and greater challenges to visual mapping design and visual analysis.

For this purpose, this paper conducts a design study to develop a visual analytic system called MVCAS. This system is based on gathered students' grade data of all the program-specific courses and three general math-type courses from a Chinese university's computer science undergraduate program. It offers a set of novel interactive visualizations for supporting users to explore various correlations between two types of courses.

To begin with, the correlation analysis requirements are extracted from three levels of details, and the original course data are preprocessed into multiple record sets from views of students and courses, respectively. Afterward, at the overall level, a series of heatmap matrices are designed to represent the grade distribution of the whole program courses from students' perspective. These heatmaps are further linked with a group of parallel coordinates demonstrating general courses' grade distributions of the corresponding students. Moreover, the pixel bar charts are designed to integrate the statistical distributions of general courses' grades with students' achievements in program courses. At the categorical level, three interactive views, namely the polar scatterplots, sunburst graphs and tag clouds, are designed based on teaching classification, theme classification and association classification, respectively. They support the exploration of correlations between general courses and program courses from multiple perspectives. Furthermore, a set of radar charts and stacked area graphs are designed for associating program courses' instructors and their scheduled semesters with general courses, and identifying the course correlations from the perspectives of instructors and semesters, respectively. At the detailed level, the curve charts and scatterplots are built up to allow for observing and comparing the detailed grade distributions of courses in different contexts. In the end, through view transformation and dynamic interaction, the exploration processes of various correlations and potential patterns between two types of courses on student performance are achieved from top to down and from whole to part, thus offering more intuitive and credible evidences for predicting and improving both teaching and learning outcomes in the focused major.

The contributions of this paper include: (1) a multi-angle preprocessing of course grades data, including decomposition, extraction and conversion; (2) a set of novel visual mappings and coordinated analysis views for effectively exploring the overall, categorical and pairwise course correlations from multiple perspectives of students, instructors and semesters; (3) a top-down hierarchical correlation analysis process for general courses and program ones; and (4) a comprehensive case study to demonstrate the effectiveness and usefulness of MVCAS on real-world datasets.

2 Related work

In this section, we discuss the background and related techniques of this work from two points of view: (1) analysis of course grades and correlations, (2) visual mapping and analytical techniques.

2.1 Analysis of course grades and correlations

Most educational institutions including Chinese universities have deployed online course management systems to support various teaching and learning activities [12,13]. The electronic records captured by these systems can be used to gain richer and deeper insights in course grades, such as the correlations and interplays between courses or course groups.

The statistical methods have been often adopted to analyze course data. For example, Pearson's coefficient was often used to quantify the student performance correlations with prior knowledge [4], other course components and related courses [5] or former students' study traces [14]. Some traditional dimension reduction methods like PCA [15], MDS [16] and clustering in data mining were also applied to simplify the complexity among numerous variables of courses or students so as to reveal the major courses that affect student performance or to focus on the students in need of help [17]. Furthermore, some personalized multi-regression and matrix factorization approaches, initially developed for e-commerce applications, can be used to forecast students' grades in future courses as well as on in-class assessments [18]. However, the lack of sufficient visualization and interactivity limits their ability to comprehensively explore and mine multi-level correlations of courses throughout the overall major curriculum from multiple perspectives.

Visual analytics techniques have also been applied into the education field to analyze course-related data, especially various learning data. For example, Mazza and Dimitrova developed a system, called CourseVis, by use of techniques from information visualization to visualize student tracking data in distance learning classes [19]. But CourseVis aimed to help instructors understand social, behavioral and cognitive aspects related to their students within a single course. Gómez-Aguilar et al. presented a visualization tool to explore the relevance pattern in the temporal frequency of students' learning activities and their performance in eLearning [20]. Although the analysis was performed across different courses, the course–course relevance in students' performance was not addressed specifically. Ritsos and Roberts investigated learning analytics and the potential of visual analytics in technology-enhanced learning (TEL) environments [21]. As they mentioned, most of works in this field focused on the exploration of detailed learning activities and processes, still not extending visual analytics into the high-level analysis of courses, such as between-course correlations and better curriculum mapping. Vieira et al. conducted a more systematic literature review of visual learning analytics of educational data [22]. According to the results they got, little work has been done to bring visual learning analytics tools into traditional classroom settings focused on in our work, and the most common relationship explored was between students' interactions with a given tool and student performance, not between different parts of the curriculum. In particular, some visualization tools and visual analytic systems have already been developed for MOOC data, and Qu et al. gave a survey to summarize these works [9]. More recently, a visual analytics system, called iForum, was further developed for the in-deep exploration of MOOC forum data [10]. In these existing systems, the visualization techniques were also concentrated on analyzing online learning behaviors.

There are some previous works that utilized interactive visualization to explore course–course relationships from a certain perspective [23–26]. Siirtola et al. developed a software tool to visualize the curriculum contents and overlap [23], while Gama et al. created a multi-level visualization to enable the analysis of the interdependence among courses in a university program from the perspective of course semesters [24]. Both of these works did not specifically consider the relevance of students' performance in different courses. Wortman et al. used graphs to explore success, failure and repetition patterns in student performance, but mainly throughout the first three courses in computer science [26]. Raji et al. modeled and visualized the course–course relationships based on the grade similarity, which can support the administrators to explore the student flow patterns through the curriculum of each major [25]. However, they did not further discuss the relationship between different course groups.

Different from these existing researches, MVCAS aims to reveal the between-course correlation patterns from multiple perspectives, such as course types, instructors, students and semesters. Moreover, it allows for the multi-level exploration of student performance correlations between two different parts of the curriculum within an academic major.

2.2 Visual mapping and analytical techniques

To intuitively present distribution patterns of multi-category course datasets and their relationships at multiple levels, more comprehensive and effective visual mapping approaches are needed. These visualization techniques should also be integrated with quantitative statistical analyses and qualitative interactive exploration seamlessly so as to flexibly analyze multi-facet correlations.

At the overall level, MVCAS represents each student's performance in all program courses and general ones. On the one hand, this is a typical issue of high-dimensional data visualization. Two traditional methods are to use 2D heatmaps [27] and parallel coordinate plots (PCPs) [28]. The heatmaps are more intuitive than PCPs in visual presentation, whereas PCPs have more obvious advantages of interactivity. Hence, we adopt the heatmap matrices to show all program courses' grades of each student so as to allow overall observation from multiple perspectives of students and use PCPs to visualize general courses' grades so as to allow interactive exploration of grade distribution patterns. On the other hand, the number of program courses is much more than the number of focused general ones, and it is necessary to connect these two different dimensional subsets together for exploring their correlations. In [29], a multiform visualization technique, termed Domino, was proposed for effectively representing multiple subsets and their relationships. Inspired by this work and other commercial tools such as Tableau [30] and Spotfire [31], we use connecting lines to combine the heatmap matrices with PCPs to form an interactively coordinated layout and further embed histograms and stacked bar charts [32] to represent the dynamics of grade distributions.

Moreover, to represent the statistical distribution of general courses' grades and at the same time show the detailed distribution of program courses' grades under different statistical segments, we use pixel bar charts [33] to integrate these two types of information and visualize them comprehensively. The basic idea is to use the histogram for partitioning general courses' grades into different segments (i.e., bars) and fill in the corresponding bars with the color pixels encoding the grade values of program courses.

At the categorical level, MVCAS visualizes different groups of program courses and reveals their correlations with general courses. First, under teaching classification, all program courses are grouped by multidimensional attributes. Their categories, grade characteristics and correlations with general courses are represented simultaneously. Scatterplot matrices [34] can make use of human perception of spatial position and proximity to help users recognize the distribution patterns. In order to reduce the space occupied by rectangular scatterplots, radial visualizations that map data in a circular fashion are becoming an increasingly popular method in high-dimensional information visualization research [35]. Although this kind of radial layout can be used to visualize the distribution of program courses under multiple teaching categories, it is difficult to present more than two attributes of a single course simultaneously only by the use of an individual data point. Hence, we use the circles with different colors and sizes to replace the points in polar scatterplots. This visual mapping allows for visualizing multiple statistical characteristics of program courses and their correlation strength with general courses and interactively exploring their details.

Second, under theme classification and association classification, program courses and their correlations with general courses are shown in the clustering style. Tag clouds [36] can intuitively display the distributions of multidimensional datasets categorized by keywords or topics, while sunburst graphs [37] can interactively display hierarchical datasets and their groupings. MVCAS uses tag clouds to cluster program courses based on their contents and topics and uses sunburst graphs to present the hierarchical and abstract structure of program courses based on students' specialty knowledge and skill factors derived from PCA.

In addition, to detect courses' correlation patterns from the perspectives of instructors or semesters, MVCAS needs to divide program courses into multiple groups according to different instructors or different semester and then map their distributions and characteristics onto an easily comparable layout. The "small multiples" approach is known in the visualization research field [38]. It uses the compact visualization to support the comparison of multiple items with multi-dimensional attributes. Therefore, we design two novel layouts based on radar charts [39] and stacked area charts [40], respectively. Radar charts map the distribution of program courses grouped by each instructor, and stacked area charts present the distribution of program courses along different semesters.

At the detailed level, MVCAS represents the correlation between a single program course and general one, shows intragroup correlations of program courses and displays the detailed grade distribution of a single course or a group of courses. Scatterplots [34] can help users intuitively recognize the pairwise correlation, while node-link diagrams [41] can quantify the correlations using the line width. The information conveyed by node-link diagrams is not so intuitive as scatterplots [42], but the latter also has a few disadvantages such as larger space occupation and weaker interaction. Therefore, these visual methods are often combined with each other in practice [43]. In this work, we comprehensively apply these two techniques to map the pairwise and intragroup correlations of courses onto the detailed views.

3 Task extraction and data processing

In this section, we extract correlation analysis tasks at multiple levels, present the corresponding course data processing from multiple perspectives and then summarize the derived design rationales.

3.1 Multi-level tasks extraction

In this study, we took computer science major of our university as an example and chose "Advanced Mathematics (i.e. Calculus)," "Linear Algebra" and "Probability & Statistics" as the focused general courses. During the whole developing process of MVCAS, we worked closely with two field experts from the major. One of them is the department chair who is responsible for assigning, monitoring and evaluating all program courses' teaching activities. His considerations focus on understanding the whole correlations between two parts, seeking the references for revising the curriculum and course schedules and proposing the advices to instructors and students for improving the performance of program courses. Another one is a faculty member and undergraduate committee member in this major who has taught several different program courses for more than ten years. Therefore, he is also familiar with the structure of curriculum. He is interested in how students' performance in mathematics courses affects different students and different program courses.

From the concept design to prototype completion, we went through three stages to collect and refine analysis processes and requirements from field experts. In the first stage, we organized a formal interview session with all experts

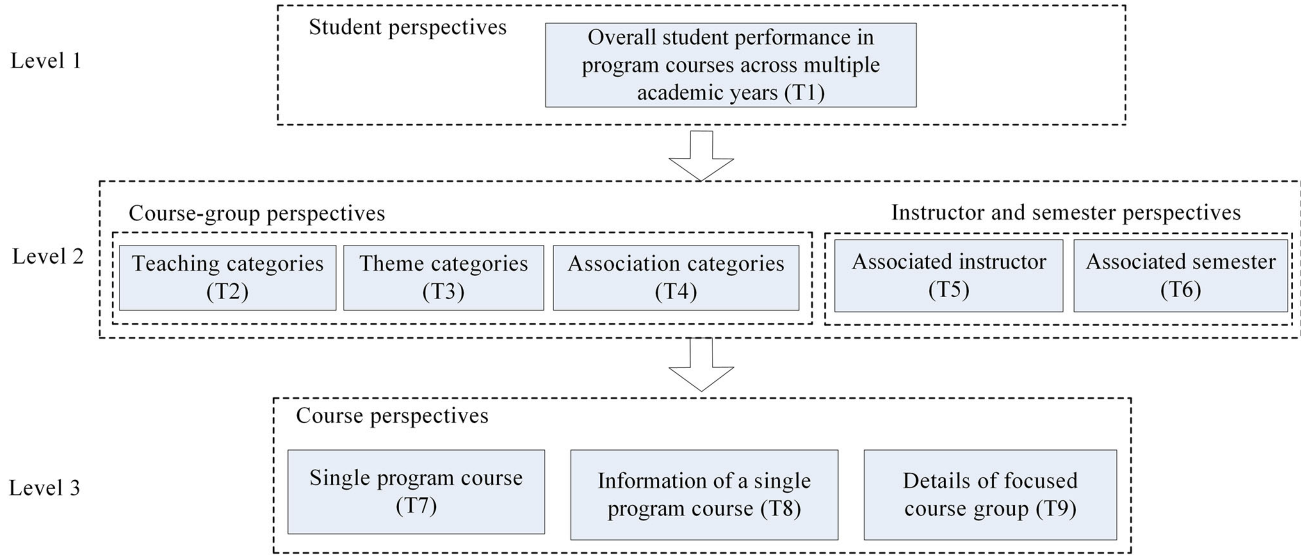


Fig. 1 Correlation analysis tasks at three levels of abstraction with multiple perspectives

to understand the overview structure and classification of program courses and to gather user requirements of analyzing course correlations. After this interview, we also invited or visited experts to have several informal discussions on course details. In the second stage, we continuously presented a series of analysis components or working prototypes to experts and wrote down their feedbacks and suggestions. In the last stage, we completed a full version of MVCAS and had another interview with experts to derive a concrete use-case scenario of exploring various correlations between two course types.

Based on the above interviews with experts, we extracted analysis tasks hierarchically from top to bottom and from the whole to details, as shown in Fig. 1.

Level 1 For an academic major, at first, the department administrators want to know the whole grade relevance between two types of courses from the perspectives of students across several years (**T1**). So it is necessary to demonstrate the distribution of program courses' performance under different combinations of students' mathematics achievements. The obtained course correlation patterns can be used to predict students' overall performance of program courses according to their performance in mathematics and further to make some early necessary interventions, such as improving teaching process and optimizing course schedules.

Level 2 In view of different categories of program courses, the experts hope to further identify the influence of mathematics courses on different types of program courses and to explore the correlation between specific mathematics courses and different groups of program courses. According to the experts' suggestion, program courses are classified from three perspectives of teaching categories (**T2**), theme characteristics (**T3**) and statistical association (based on PCA processing) (**T4**) so as to obtain various subgroups of program courses and their correlations with mathematics courses. These correlation insights can be used for reminding students to pay more attention to certain program courses and meanwhile reminding the corresponding instructors to pay more attention to some connections between program-specific knowledges and general mathematic ones. Moreover, they can also provide direct evidences for administrators to optimize and improve some particular parts of the major curriculum.

In addition, by comparing and analyzing the correlation between different instructors and mathematics courses (**T5**), as well as between different semesters and mathematics courses (**T6**), the department will more clearly know which faculty members should pay more attention to mathematics, and which types of mathematics should be attached more importance in different academic periods to yield better outcome for some program courses.

Level 3 In order to provide more concrete advices for a specific program course or a subgroup, it is also necessary to further identify the relevance between individual program courses and mathematics courses (**T7**) and to demonstrate more details of a single course (**T8**) or focused subgroup of courses (**T9**). Such specific course information should allow users for conveniently digging out the grade distribution characteristics and pairwise correlations of courses in the detailed level.

3.2 Multi-perspective data processing

The course score data of computer science major for visual analysis are derived from the original student-course grade tables in the university's educational database. Each line

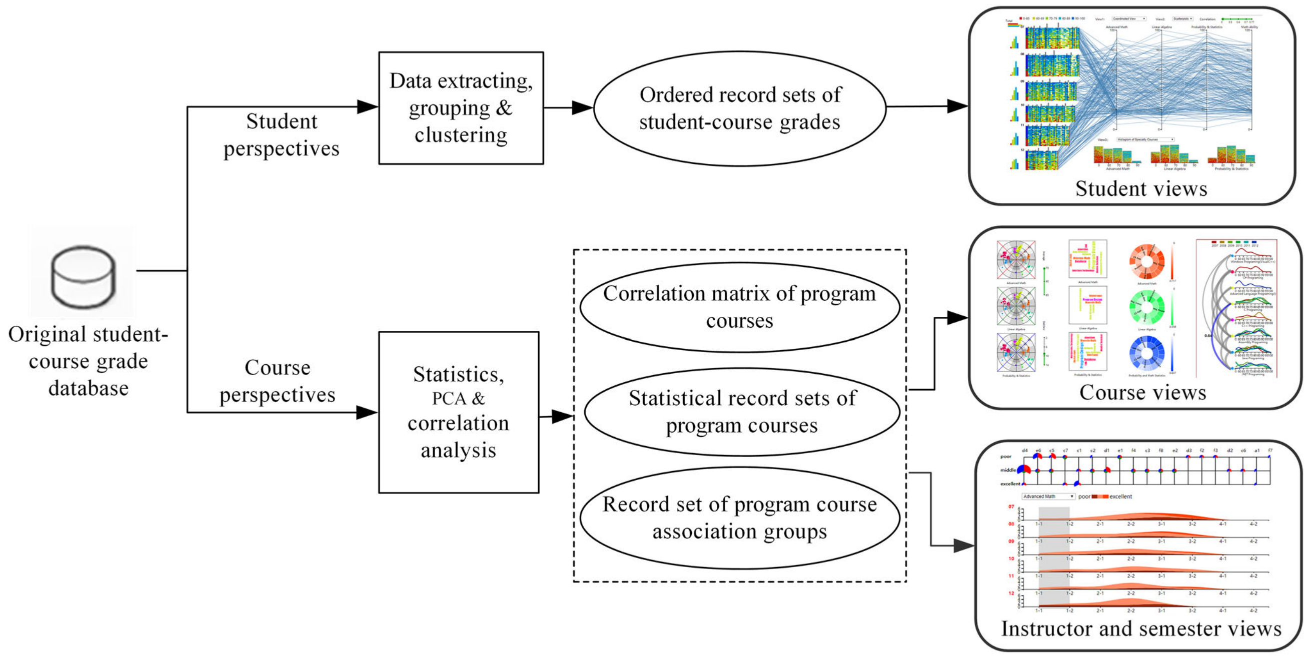


Fig. 2 Data preprocessing from the student and course perspectives

in these tables contains a student's performance (i.e., a hundred-mark score) in a certain class as well as other relative attributes such as the course name, number, academic semester, lecturer, requirement mode, etc. In this work, a class is the actual instantiation of a course. For each course, we only extract students' grades in the class taken by them for the first time so as to eliminate the impact of retaking the course. To meet user requirements for multi-level and multi-perspective course correlation analyses, we conduct a series of data processing. First, all grade records of both general courses and program courses are extracted and clustered from students' perspective. Then, various course data decomposition, PCA-based statistics and correlation computation between courses are conducted according to different classifications of program courses, as shown in Fig. 2.

Student perspectives In order of their entry time, students' scores of program courses in all semesters are extracted and grouped by each student to form original record sets of student-course grades. Then, according to their overall performance in program courses, all students' course grades are clustered into three categories: "Excellent," "Middle" and "Poor." In the first category, students' grades in most of courses are above 80 and none of courses taken by them is failed. In the second category, students' course grades are relatively lower, but all are above the passing mark, namely which are from 60 to 80. In the last category, students' grades in some courses are below the passing mark. In this way, these clustered student-course grades form a record set of program courses sorted by students, thus paving the way for users to make further analysis. In addition, students' scores in three mathematics courses are also extracted, and meanwhile, the weighted average of every student's mathematics scores is calculated to represent his or her comprehensive grade in the mathematics. The credit of each mathematics course is used as the weight. All of these mathematical score data form the corresponding students' mathematics grade record set.

Course perspectives First, the Pearson's coefficients for each pair of program courses in student grades are calculated to form a set of correlation matrices that reflect the interrelation strength among program courses. Then, several attributes closely related to each course, such as instructors, semesters and course types, are extracted and then used to divide program courses into different groups. Specially, according to the usual teaching classification, course types are divided into three dimensions: compulsory or elective mode, theoretical or practical type, and high or low credit. As a result, all program courses are classified into eight categories. Meanwhile, the average score and standard deviation of each course as well as its correlation coefficients with three mathematics courses are calculated, respectively, to form several statistical record sets of program courses. Finally, we extract student grades of all compulsory program courses and those elective program courses which are taken by more than half of all students, and conduct the dimensionality reduction process with PCA to obtain students' specialty knowledge and skill factors. These factors reflect the main differences of students' performance in program courses. Furthermore, the Pearson's coefficients between these specialty factors and mathematics courses are also calculated to form a record set of program course association groups.

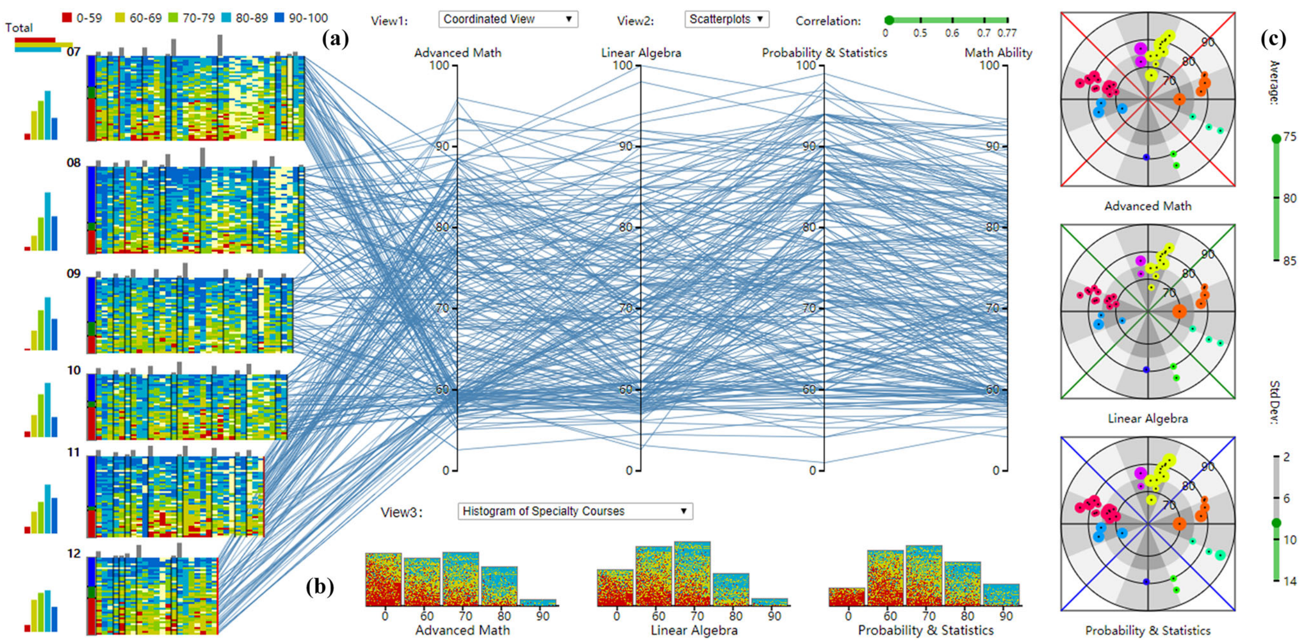


Fig. 3 The main interface of MVCAS. **a** The heatmap matrices coordinated with PCPs present overall correlations of course grade distributions. **b** The pixel bar charts demonstrate statistical distribution correlations. **c** The polar scatterplot shows course grouping correlations in different teaching categories

3.3 Design rationales

According to the above-extracted analysis tasks and course datasets, we derive the following design rationales to guide our design of the MVCAS system:

- **R1:** To reveal overall grade distribution correlations of the whole program courses relative to general courses from multiple perspectives of students (**T1**).
- **R2:** To compare students' grade correlation patterns of program courses in different teaching categories with general courses (**T2**).
- **R3:** To examine correlation patterns of general courses and program courses in different theme categories (**T3**).
- **R4:** To demonstrate specialty knowledge and skill factors of the curriculum and their correlations with general courses (**T4**).
- **R5:** To identify correlation patterns between general courses and program courses taught by different instructors (**T5**).
- **R6:** To unfold correlation patterns between general courses and program courses scheduled in different semesters (**T6**).
- **R7:** To depict the detailed grade distributions and statistical characteristics of individual courses or course groups (**T7, T8, T9**).

4 System design

In this section, we introduce the overall visual design and main components of the MVCAS interface and then describe their visual design details.

4.1 Overview

MVCAS mainly contains three kinds of views, which are student views, course views, instructor and semester views. These views assist the multi-level exploration of correlations between general courses and program courses from the perspectives of students, courses, instructors and semesters, respectively. The layout of the main interface is shown in Fig. 3. Student views are further divided into three parts, which are heatmap matrices (Fig. 3a), PCPs and pixel bar charts (Fig. 3b). The first two parts can be interactively integrated together. Course views include polar scatterplots, tag clouds and sunburst graphs, all of which share a same region in the screen (Fig. 3c) and are switched from one to another. Instructor and semester views include radar charts and stacked area graphs. They share the same region with the PCPs (Fig. 3a). The views and interaction with the system are demonstrated in the accompanying video.

The user interface of MVCAS is developed with D3.js [44] and fully web-based. The backend components use Python to implement all of the data processing. Currently, considering the security of the university management system, all the exported course data and derived data are stored in the off-line CSV files. From the initial concept design to the whole prototype completion, we also followed several iterations of the nested model for visualization creation [45], and the development procedure lasted over a year.

4.2 Student perspective

4.2.1 Heatmaps coordinated with PCPs: presenting overall distribution correlations

To unfold the overall correlation patterns of program courses and general courses (**R1**), we develop the coordinated view of heatmap matrices and PCPs to demonstrate and compare students' grade distributions in these two types of courses.

As shown on the left side of Fig. 3a, a series of heatmap matrices linked with the PCPs are used to display the whole program courses' grades of all students. Students are grouped according to their entry time. The entry time is labeled with a two-digit year, for example, the number "07" at the top left corner of the heatmap indicates the year of 2007. The color mapping scheme of grade values is based on five segments, i.e., below 60, 60–69, 70–79, 80–89 and 90–100. In such a heatmap matrix, each colored rectangular cell represents a student's grade in one course. Each row represents a student's grades in all program courses, which are arranged in order of academic semesters. From top to bottom, all rows of each matrix are arranged in order of decreasing students' overall performance, which are clustered into three categories, as described in Sect. 3.2. Each column represents all students' grades in the same course, while the courses in different semesters are separated by the vertical black line segments, and the top gray bar represents the number of courses within the corresponding semester.

Initially, a continuous linear color mapping was performed for the grade interval between 60 and 100. As shown in Fig. 4a, the color confusion was too serious to distinguish the distribution of different grade segments, thereby making it difficult to identify the students' overall performance. To improve perception of differences in colors, the continuous color mapping is optimized to the discrete one that encodes the score values in the same grade segment with more similar color than among different grade regions. Moreover, five grade segments are further aggregated into three levels, namely being the poor level (below 60 points), middle level (60–69 and 70–79) and excellent level (80–89 and 90–100). Similarly, the difference of colors used inside a same level is less than between levels. By doing so, MVCAS can demonstrate students' grade distribution and clustered performance much more distinctly within the same heatmap matrix. In addition, the course columns are further reordered by the course–course correlations so as to aggregate those highly

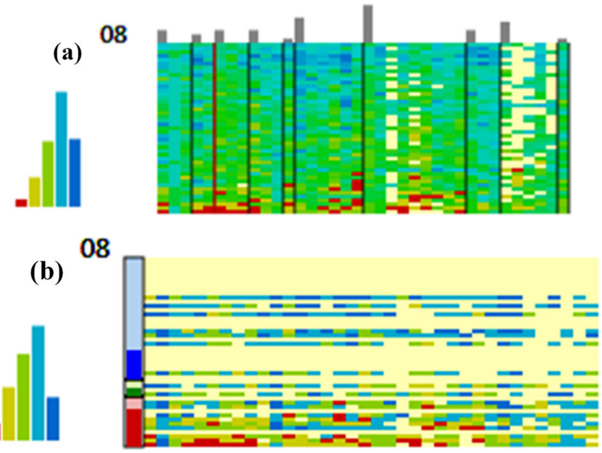


Fig. 4 The heatmap matrix for showing all program course grades of students with a same entry time (the year of 2008). **a** The visual effect before optimizing color mapping and sorting method. **b** Highlighted student records after interacting with parallel coordinates

correlated course columns in space as well as to reduce the visual clutter in color.

The parallel coordinates in Fig. 3a are used to display the student grades in three mathematics courses and their comprehensive performance in mathematics. To link students' mathematic records with their own program course records, the left point of each polyline of parallel coordinates is extended to the rightmost point of one row of rectangular pixels in heatmap matrices. Thus, a series of coordinated demonstrations and interactions will be applied to both of views when users brush each axis of parallel coordinates. This visual layout allows any combination of grade segments of different mathematics courses to construct the query conditions for the grade correlation exploration. One heatmap matrix after interaction is shown in Fig. 4b, where the pixel rows of selected students are highlighted. Through observing the positions and colors of the changed pixel rows, which are related to the selected ranges of mathematic grades in parallel coordinates, users can intuitively judge, which cluster the corresponding students belong to on the whole and how their overall performance in program courses goes.

To illustrate the overall distributions of selected students' grades in all program courses as user filtering in parallel coordinates, a statistical histogram is shown on the left side of each heatmap matrix. Every bar represents the volumes of students' grade records in one grade segment, as shown in Fig. 4b. Meanwhile, a stacked bar chart that dynamically reflects different proportions of selected students in three grade clusters, is also added next to the first left column of each heatmap matrix. Through observing the change of this statistical information when brushing in parallel coordinates, users will have a more accurate identification of the students' grade correlation patterns between mathematics and program course. In general, such a coordinated layout can support well the correlation exploration between two groups of multivariate data with great difference in the number of attribute dimensions.

4.2.2 Pixel bar charts: demonstrating statistical distribution correlations

To further reveal the statistical correlation between all students' mathematic grades and their program course grades (**R1**), another combined layout, namely the pixel bar chart [33], is designed to depict the integrated grade distributions of two types of courses. As shown in Fig. 3b, three groups of histograms represent the distribution of all students' grades in three mathematics courses, respectively. The grades are still divided into five segments in the same way as described in Sect. 4.2.1. Every student's grades in all program courses are mapped onto a colored pixel row that fills into the corresponding bar of histogram regions where his or her mathematic grades belongs to.

In order to reduce subjective influences on student grades like instructor rating, all the program course scores have been converted from hundred-mark rating system to excellent percentile rating system before visual mapping [13]. The percentile value of excellent rate S is calculated as:

$$S = 1 - \frac{i-1}{n} \tag{1}$$

where n registers the number of students who take the course examination, and $i = 1, 2, \ldots, n$ means the ranking order of original scores. For those students who have vacancies in some courses, for example they did not take these courses, their vacancy scores of these courses are filled with the weighted average scores of those program courses taken by them so as to keep the same length of pixel rows for different students. To reduce visual clutter, the percentile values of excellent rate are clustered into three intervals that are [0,0.3), [0.3,0.7) and [0.7,1], and the number of corresponding pixel colors decreases from five to three. This makes it easy to observe students' program course performance relative to five mathematical score segments through the color distribution of pixels. In addition, the space arrangement of student records (i.e., pixel rows) within each histogram is based on a radix sort that means ranking the number of courses in different excellent rate intervals into which current student performance falls. As a result, most of pixels are visually distributed from blue to red in vertical direction. However, the previous version only arranges those who belong to the same grade cluster (as described in Sect. 4.2.1) into a group without reordering them within the cluster. Therefore, the color distributions of pixel rows are less legible than the current version, as shown in Fig. 5a.

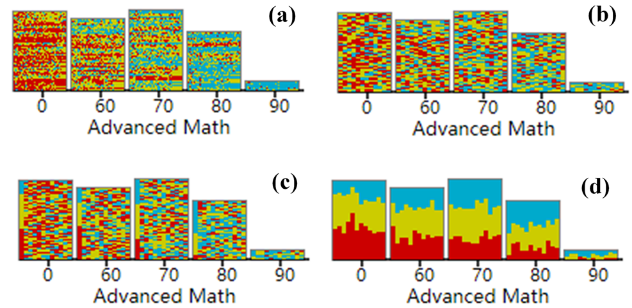


Fig. 5 The design of pixel bar charts. **a** Before optimizing. **b** Original arrangement of all specialty factors. **c** After sorting a single column of factors. **d** After sorting all factors

Additionally, each student's corresponding pixel row can also be replaced by the scores of student's specialty knowledge and skill factors (i.e., principal components of program course score datasets) calculated by the PCA (see Sect. 3.2). These specialty factors of all students are arranged from left to right in accordance with their information strength reflecting students' performance differences in program courses (Fig. 5b). Moreover, to observe the correlation between a certain specialty ability and mathematics course, the pixel rows of students can be rearranged according to the scores of this factor from high to low vertically, as shown in Fig. 5c. However, if the pixel rows of bar charts are still arranged in order of the overall performance of students, it is difficult to provide users a clear insight of distribution patterns. The reason is that the PCA processing has converted the original score variables of highly correlated program courses into one comprehensive variable. Therefore, through reordering the factor scores of each column, respectively, a new layout is provided, as shown in Fig. 5d, in which the statistical correlation between each specialty factor and mathematics courses can be more clearly observed through the distribution of colored pixels.

4.3 Course perspective

4.3.1 Polar scatter: investigating course grouping correlations of teaching categories

To compare the correlation patterns between mathematics courses and program courses in different teaching categories (**R2**), three scatterplots based on polar coordinates are designed to demonstrate the integrated multidimensional course information, as shown in Fig. 3 c. This information includes the categorized correlations between the two types of courses and some statistical characteristics of program course scores. Each scatterplot corresponds to one of mathematics courses.

As shown in Fig. 6, the inner space of each big circle in scatterplots is divided into eight sectors equally according to

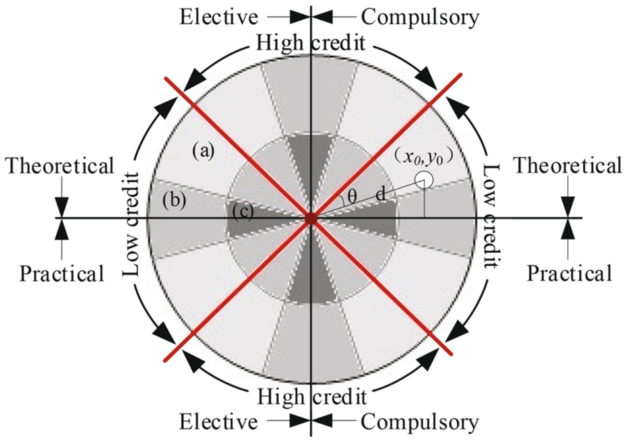


Fig. 6 The mapping rule of scatterplot based on the polar coordinate

three attributes of courses, which are the requirement mode (compulsory or elective), the course content (theoretical or practical) and the number of credit (low or high). In doing so, each sector represents one of course categories, i.e., a particular combination of three attributes (such as the compulsory, theoretical and high credit). The color of diagonal lines (and center space of polar coordinates) encodes the corresponding mathematics course. In each sector, there are some small circles (called course circles) that represent individual program courses belonging to the current category. The filling colors of these circles encode course types. Their radius sizes correspond to the grade correlation coefficients between program courses and current mathematics course. Besides, the positions of circles in polar coordinates are determined by the course average scores *avg0* and standard deviation *sd0*. Assuming that the radius of the big circle is R, the distance d between the center of the course circle (represented by the black dot) and the center of the polar coordinate is calculated by Eq. 2. Meanwhile, the corresponding radial angle θ satisfies Eq. 3, and the Cartesian coordinates of the course circle center are determined by Eq. 4.

$$d = \frac{avg0 - 60}{\max(avg) - 60} R \tag{2}$$

$$\tan\theta = \frac{sd0 - \min(sd)}{\max(sd) - \min(sd)} \tag{3}$$

$$\begin{cases} x0 = \dfrac{\sqrt{2}}{2} R \dfrac{\tan\theta + 1}{\sqrt{\tan\theta^2 + 1}} \\ y0 = \dfrac{\sqrt{2}}{2} R \dfrac{1 - tan\theta}{\sqrt{\tan\theta^2 + 1}} \end{cases} \tag{4}$$

According to the above visual mapping scheme, all program courses are grouped and allocated into scatterplots. When a course performance is better, that is, the average score is higher and the standard deviation is smaller, the corresponding circle will be farther from the center of polar coordinate and its radial angle will be smaller. Three gray areas in each sector can be used to distinguish the statistical characteristics of course performance, which are the excellent, middle and poor, respectively, as marked by (a), (b), (c) in Fig. 6. Therefore, through observing the position and size of course circles inside different sectors, users can intuitively judge the grade distribution characteristics of program courses in any teaching categories and analyze their correlations with the mathematics courses. Furthermore, there is a slider on the top right of main interface for adjusting the correlation threshold so as to filter program courses according to their correlation strengths (Fig. 3c). Also, the other two sliders on the right side are used for adjusting thresholds of the average score and the standard deviation, respectively, so as to change the visualized course sets based on students' performance interactively. What's more, the interaction results are also coordinated with other views, such as the course theme tag clouds, instructor radar charts and semester stacked diagrams, which will be discussed in the following part.

4.3.2 Tag clouds: exhibiting course grouping correlations of theme categories

To discover the topics and contents of program courses highly correlated to three mathematics courses (**R3**), we design three course tag clouds to represent these theme categories and their characteristics.

As shown in Fig. 7a, each text tag represents a course subset. The topic of every and each course is extracted by the field experts according to its theme in the degree program's curriculum, such as "Programming Design," "Database," etc. Thus, courses like "C++," "Java" and ".Net" are clustered into the group of "Programming Design." The size of each tag encodes the number of courses falling in the current subset, in which the correlation coefficients of all courses with mathematics courses exceed the given threshold (as described in Section 4.3.1). For each course subset, the tag color is determined by the teaching category containing the largest number of courses and keeps consistent with the color of the corresponding course circles in the polar scatterplots.

4.3.3 Sunburst graphs: demonstrating course grouping correlations of association categories

As described in Sect. 3.2, each of students' specialty knowledge and skill factors corresponds to several original program courses. To unfold the correlations between these factors and mathematics courses (**R4**), we map these factors and program courses into three sunburst graphs to present their hierarchical structure.

As shown in Fig. 7b, each sector of the inner ring corresponds to a specialty factor, and its angle encodes the factor's strength of information on program courses' grade differ-

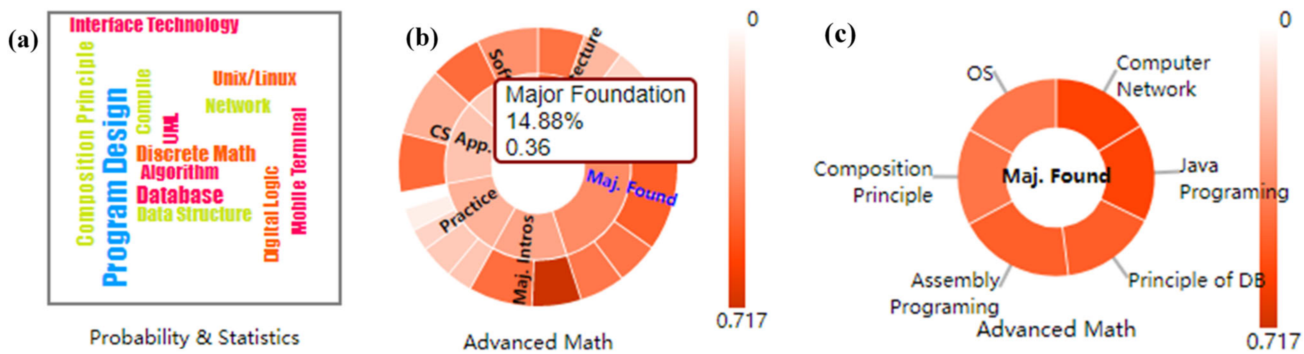


Fig. 7 Course grouping views. **a** Tag cloud for theme categories. **b** Sunburst graphs for association categories. **c** After unfolding a specialty knowledge and skill factor

ences (i.e., the contribution rate of variation), whereas each sector of the outside ring corresponds to an original program course, and the angle represents this course's contribution to current PCA factor. The lightness of the sector's color corresponds to the correlation coefficient between the specialty factor (or program course) and current mathematics course. The sectors are arranged in clockwise according to their correlation coefficient value. Through unfolding a node in the inner layer, this view also allows users to inspect the main courses involved in this specialty factor and their correlation with mathematics courses interactively, as shown in Fig. 7c.

4.3.4 Curve charts: depicting detailed grade distributions

To uncover the grade characteristics and pairwise correlations of courses at the detailed level (**R7**), a set of course curve charts are designed to show the detailed distributions of student grades in a single program course or in a category (or group) of program courses. When users click on a course keyword in tag clouds, or a specialty factor in sunburst graphs or an instructor/semester tag (described in Sect. 4.4), the probability density plots of student scores in all corresponding classes will appear on a popup view, as shown in Fig. 8a. Each group of curve charts reflects the distributions of student grades in current course across multiple years, while each curve represents student grades in a specific class that was taken by students with the same entry time. The curves are color-coded by the year when students enroll into the university.

In addition, the colored course circles corresponding to ones in the polar scatterplots are added to the left side of each curve chart. If the correlation coefficient between two program courses is greater than or equal to a certain threshold (e.g., 0.5), there will be an arc linking these two circles and its width indicates the correlation coefficient value. Using this node-link diagram, users can quickly find the main course types within the current course group and identify the course–course correlations.

Moreover, when a single program course is chosen, more detailed statistical information (e.g., score average, standard deviation and correlation coefficient) will be further displayed, as shown in Fig. 8b. To compare student grades between this course and a certain mathematics course, a scatterplot will also be integrated into this view. Through dragging the mouse on the scatterplot to interactively inspect different areas, users can study more detailed pattern of student grades.

4.4 Instructor and semester perspectives

4.4.1 Radar charts: exploring association with instructors

Instructor radar charts are used to intuitively show the degree of association between program courses taught by different instructors and mathematics courses (**R5**), as shown in Fig. 9a. First, the overall performance of each course is classified into three categories: excellent, middle and poor based on their distribution in scatterplots. Thus, each radar chart is composed of three colored sectors that represent three subsets of program courses related to three mathematics courses, respectively, and the radius of each sector encodes the number of courses in the corresponding course subset, as shown in Fig. 9b. Then, based on the principle of small multiples [38], all radar charts are arranged into a matrix layout. The row represents the same performance category of courses taught by different instructors; the column indicates three different performance groups of program courses taught by the same instructor. Only those program courses whose correlation with mathematics exceeds the given threshold are chosen. The threshold is the same as used in scatterplots described in Sect. 4.3.1.

In addition, all columns indicating different instructors are ranked from left to right according to the total number of courses highly correlated with mathematics. This instructor-course style layout can support multi-faceted observation, exploration and comparative analysis of program course per-

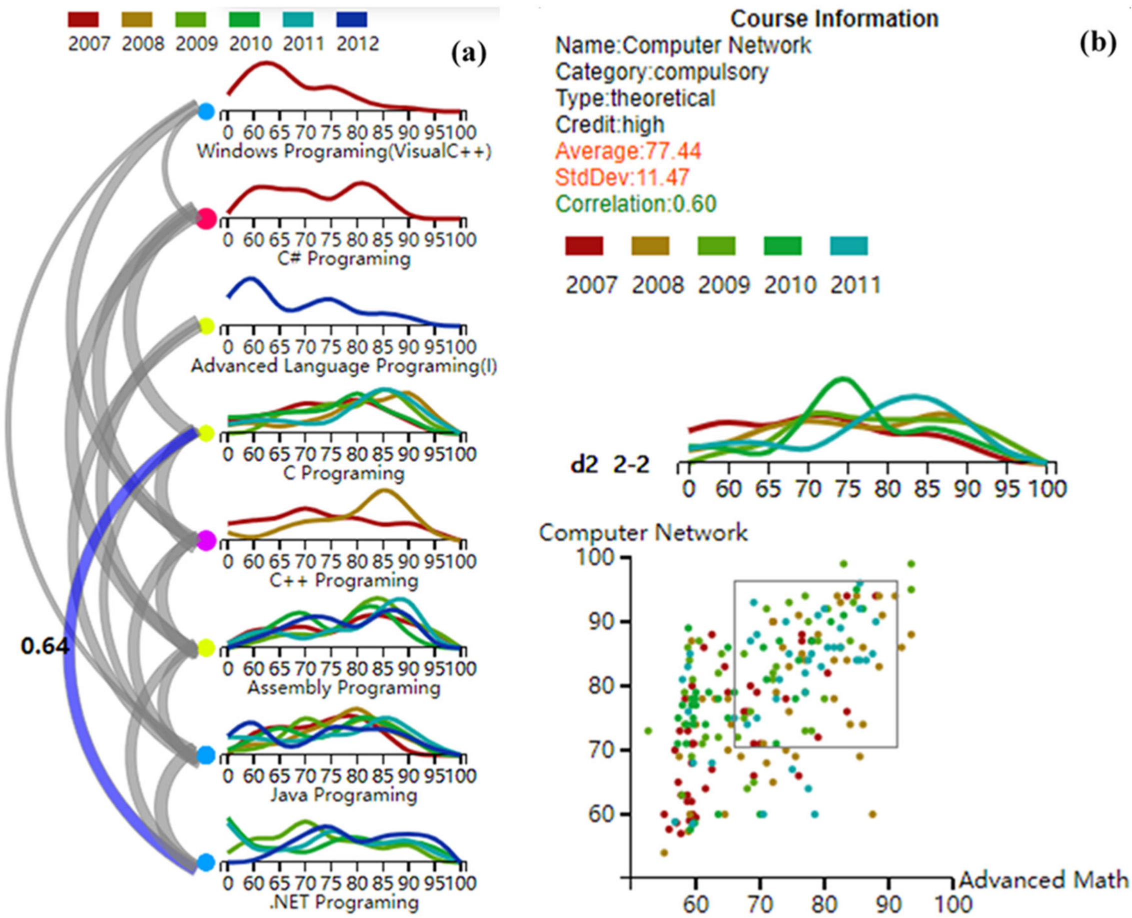


Fig. 8 Course curve charts. **a** Score distributions of a group of courses. **b** Details and score distributions of a single course

formance characteristics from the perspectives of different instructors and different mathematics courses. For example, the administrators of the academic major can easily find which instructors should pay more attentions to the influence of mathematics, and which mathematics course should be paid more attentions by different instructors.

4.4.2 Stacked area graphs: revealing association with semesters

To unfold correlation patterns between general courses and program courses from the perspective of different semesters (**R6**), stacked area graphs are designed to visualize the distributions of program courses along all eight semesters during four academic years, as shown in Fig. 10a. These courses' correlations with mathematics courses are above the given threshold. Horizontally, each stacked graph corresponds to a group of students enrolled in a same year, which is labeled with a two-digit number on the left. The height of the curve at each semester (labeled with "1–1," "1–2" and so on) represents the number of courses, and the color brightness distinguishes three course categories of excellent, middle and poor performance. Vertically, there is a gray band used to mark the semester in which current mathematics course is taken. The course "Advanced Mathematics" lasts two semesters ("1–1" and "1–2"), and the other two mathematics courses last one semester. Thus, when examining the stacked graph horizontally, users can explore the semester distribution of program courses that are highly correlated to mathematics during four academic years. While doing inspection vertically, users can compare these courses' performance among different classes.

Meanwhile, to help users make comparison among three mathematics courses conveniently, a global view including three groups of stacked area graphs can be interactively chosen, as shown in Fig. 10b. Different colors of curves correspond to different mathematics courses. In contrast, to provide more detailed distribution of courses, a "focus + context" interaction technique is adopted, with which the stacked graph of a focused class can be zoomed in, as shown in Fig. 10c.

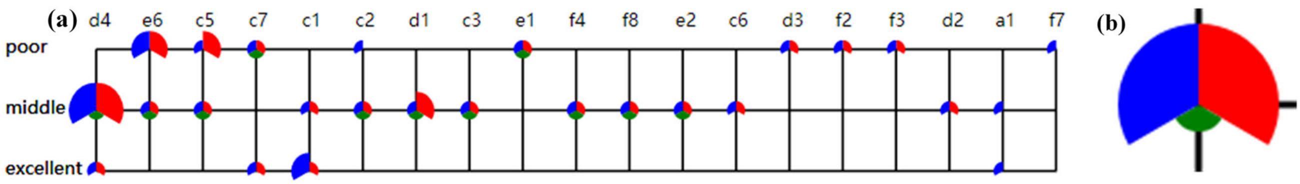
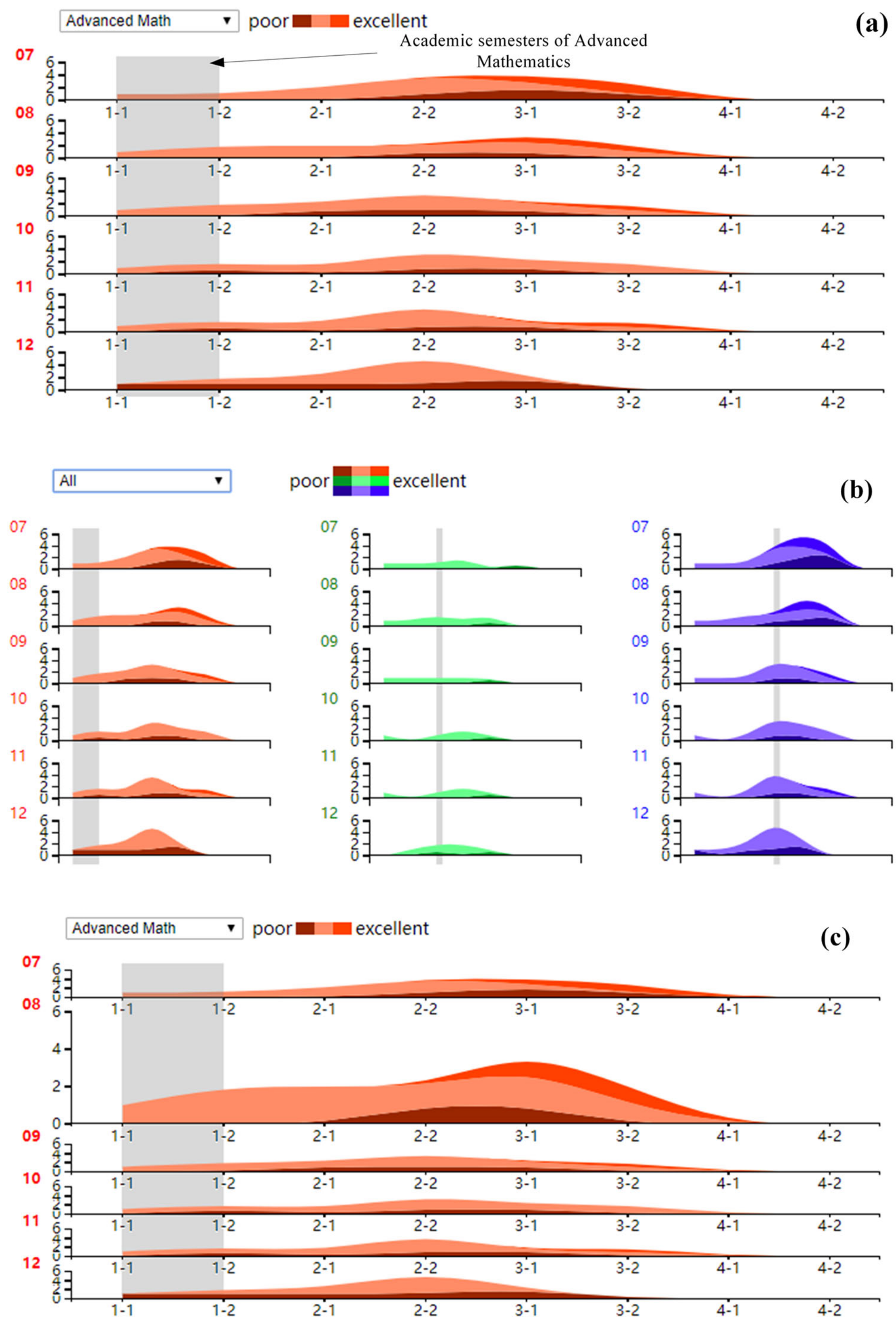


Fig. 9 **a** The overview of radar charts. **b** The layout of a single radar chart

Fig. 10 Stacked area graphs of academic semesters. **a** The view corresponding to one of mathematics courses. **b** The overview integrating three types of mathematics courses together. **c** Zooming in on one of charts



4.5 Multi-view switching

In order to effectively support the multi-perspective and multi-level exploration process, besides those within-view interactions like selecting, filtering and linking, a series of between-view switching interactions are further designed.

As shown in Fig. 3 , the center part of the main interface can be switched from the view of parallel coordinates (the student perspective) to the combined association view of radar charts and stacked area graphs (the perspectives of instructors and semesters). The inner contents of pixel bar charts at the bottom of interface can also be switched among three different perspectives (see Sect. 4.2.2), thus making it easy to observe overall correlations of focused mathematics courses and all program courses, student specialty factors and even any single specialty factor, respectively. The right part of interface can be switched among three course views for allowing the between-course correlation analysis of student performance from the perspectives of teaching categories, course themes and student specialty factors (see Sect. 4.3). Furthermore, the curve charts of students' course grades can also be interactively generated and pop up from the above course-related views, instructor-related views and semester-related views so as to easily support users to observe the detailed student performance in an individual course or multiple courses.

5 Case study

In this paper, the current MVCAS prototype was developed based on a real-world course grade dataset of the computer science major over a period of 8 years (2007~2014) from a Chinese university. The course data contain a total of 21464 original grade records, involving 244 students and 47 program courses. These program courses were taught by 19 instructors. The analyzed general courses are three mathematics courses, which are "Advanced Mathematics," "Linear Algebra" and "Probability & Statistics."

To assess the effectiveness and usefulness of MVCAS, we invited the same two field experts, as described in Sect. 3.1, and two senior students from this major to use the system to examine these course grade data. The expert evaluation was conducted through a semi-structured interview. It took us about 30 minutes, at first, to describe the data, demonstrate the system and instruct them to learn various interactions. Then, they tried to conduct multi-level and multi-perspective exploration in the following 60 minutes. We recorded their informal feedbacks and suggestions during the whole process for later analyses.

5.1 Exploring overall correlations between mathematics and program courses

First, the experts wanted to examine the overall distribution and correlation patterns of students' performance in the two types of courses (**R1**). From the heatmap matrices mapping students' performance in all program courses and the PCPs showing the grade distributions of three mathematics courses (Fig. 3a), they gained the first overview of students' achievements in dozens of courses across multiple academic years and deeper insights into different characteristics about students' grades in these mathematics courses: "It is so obvious that more students got the low achievement in 'Advanced Mathematics' than in the other two mathematics courses! Perhaps, as their first mathematics course, the department should remind students (freshmen) make more efforts in this course."

Through further observing three pixel bar charts below the PCPs (Fig. 3b), they noticed that as mathematic grades gradually increased, the overall percentage of red pixel blocks inside the colored bars gradually decreased from left to right, while the change of blue ones was inverse instead. This indicated that if students' mathematic performance was better, then they had more possibilities to make a better performance in program courses. "It verified a subjective guess that students' scores of three mathematics courses were positively correlated to their scores of program courses in different degrees," the department chair said.

Moreover, the experts observed that student scores in "Advanced Mathematics" and "Probability & Statistics" presented more obvious characteristics of partial distribution than in "Linear Algebra." So they decided to conduct more detailed exploration on individual mathematics courses. Through comparing three bars of the score interval below 60 with one another and interactively examining the coordinated view of heatmap matrices and PCPs, they found that students with lower scores in "Probability & Statistics" also showed poorer performance in the other two mathematics courses and most of program courses. However, if students got lower scores in "Advanced Mathematics" or "Linear Algebra" but relatively higher scores in the other mathematics courses, their program courses' scores were not necessarily lower. The stacked bars on the left of heatmaps also showed that some students still got good or middle performance in a few program courses (Fig. 11a). One expert said to another, "Compared to the other two mathematics courses, 'Probability & Statistics' showed higher correlation with program courses in the computer science major." Two students expressed the similar feeling. Judging from this, the experts considered using students' performance in "Probability & Statistics" to predict their overall performance in program courses in future.

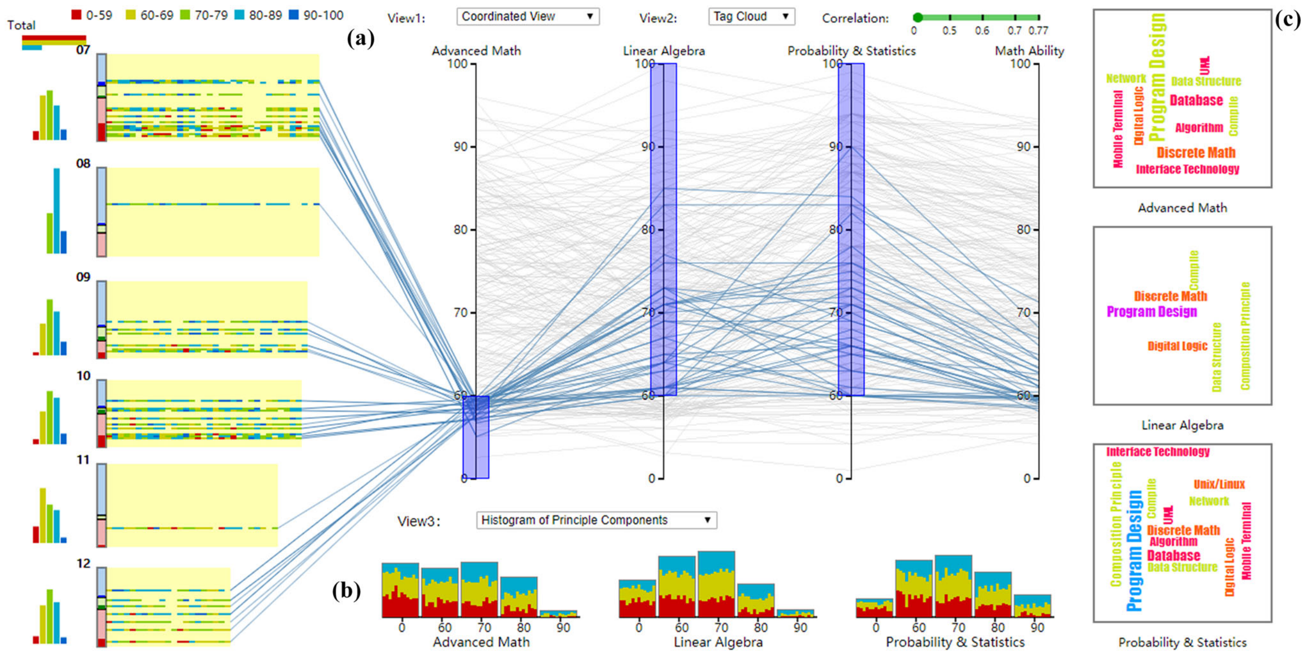


Fig. 11 The view of heatmap matrices when selecting students with poor performance in "Advanced Mathematics" and good performance in two other mathematics scores in parallel coordinates. **b** The view of pixel bar charts for showing all ordered specialty knowledge and skill factors. **c** Tag clouds of program courses

In addition, after the experts switched the pixel bar charts to show the overall distribution of students' specialty knowledge and skill factor scores, they also clearly observed a positive correlation between these factors and mathematics. Specially in the "Probability & Statistics" histogram, the ratio of red pixels to blue ones among different score segments had more obvious changes than in the other two histograms (Fig. 11b), which perhaps indicated that this mathematics course was highly correlated to students' comprehensive specialty abilities once again. When switching the pixel bar charts to the original distribution of specialty factors ordered by each student, they found that overall correlations between these factors and mathematics became rather obscure. But after reordering the pixel rows according to the score of a single factor, as shown in Fig. 5c, they found that there was still a strongly positive correlation between this specialty factor and mathematics courses.

5.2 Exploring the grouping correlations of courses

After having explored the overall correlation patterns, the experts focused on polar scatterplots to investigate the correlations between mathematics and program courses in different teaching categories (**R2**). Through comparing the size of course circles inside three scatterplots (Fig. 3c), they found that whether from the overview or from the view of individual teaching categories, both "Advanced Mathematics" and "Probability & Statistics" had strong correlation with more program courses and more course types than "Linear Algebra." Moreover, "Advanced Mathematics" showed more highly correlated with theoretical program courses, as shown in the upper half part of corresponding scatterplot in Fig. 3c. Perhaps apart from the specific mathematic knowledge, the ability for abstracting and logical thinking developed in this course was useful for learning the computer theory knowledge, they guessed. So students should be suggested to pay more attention to the training of logical thinking ability while learning "Advanced Mathematics" during their first year, thence establishing a more solid foundation for studying the following program courses. By further observing the distribution of each particular type of course circles, they also found that the standard deviation of student scores in compulsory theoretical courses with high credit (visualized by yellow circles) was generally larger than in the other course types. That indicated that students' performance in these important program courses had more significant difference. There may be a potential pattern that "if a student's score in Advanced Mathematics was better, his or her performance in this type of program courses would be also relatively better, and vice versa." Therefore, they thought these program courses' instructors should pay more attention to students with poor performance in "Advanced Mathematics."

Then, the experts wondered what the correlation patterns of program courses in different theme categories might differ

(**R3**). From Fig. 11c, they found that “Advanced Mathematics” and “Probability & Statistics” were highly correlated with the subgroup of “Programming Design.” Through clicking on the “Programming Design” tag in the course tag clouds corresponding to “Probability & Statistics,” the grade curve charts of involved courses were displayed, respectively (**R7**), as shown in Fig. 8a. From these grade distributions, the experts found that most of these courses were also highly correlated to one another, and some courses were similar to “Probability & Statistics” in the grade distribution pattern, while others were inverse to that. Since the grade distribution of “Probability & Statistics” was right skewed, it indicated that “although some students’ grades in ‘Probability & Statistics’ fell in a high score range, their grades in some programming courses might be still in a relatively low score range.” This could offer a reminding that although those students with better performance in “Probability & Statistics” could get the high achievement in most of programming courses, they still needed to pay more attention to their learning of some particular programming courses.

To further analyze the characteristics of program courses in different association groups and detect their correlation with mathematics, the experts chose to visualize the course sunburst graphs (**R4**). Consequently, they found that the main specialty knowledge and skill factors were “Computer Architecture” and “Major foundations.” Of them, the “Major foundations” factor was strongly correlated with all three mathematics courses and meanwhile it also had the strongest ability to represent the overall performance differences of program courses (Fig. 7b). After unfolding this factor, they further found that it was mainly composed of six courses including “(Computer) Composition Principle,” “Operating System (OS),” “Computer Network,” “Java Programming,” “Principle of DB (Database)” and “Assembly Programming.” All of them were highly correlated with mathematics courses (as shown in Fig. 7c). “The math is really so important!”, exclaimed a student. Therefore, the experts thought that this type of courses required more attention from the department and students. It was necessary for the relative instructors to further delve into the intrinsic reasons of high correlation and think about effective methods to improve courses’ teaching outcome from the course-content and course-thinking perspectives.

5.3 Exploring the correlation of a single program course with mathematics

Based on the above findings of course grouping correlation, the experts wanted to know the detailed correlation patterns of mathematics courses with individual program courses, especially with those in the compulsory, theoretical and high-credit category (**R7**). In the polar scatterplot corresponding to “Advanced Mathematics,” they clicked on the course circle of “Computer Network” to open its curve chart and then observe the detailed information. According to the shape of course grade curves, as shown in Fig. 8b, it can be determined that the proportions of both high scores and low scores in this course were relatively large. Furthermore, through selecting different regions of the scatterplot at the bottom of the popup window, they identified that most of students who failed in the class of “Advanced Mathematics” could not pass this course, whereas almost all students with a high grade over 90 in “Advanced Mathematics” had a grade over 80 in this course. “These discoveries are also meaningful to make more specific advices for our instructors and students,” one expert said.

5.4 Exploring the association of mathematics with instructors and semesters

Finally, after the above course–course exploration, the experts shifted their focus to the instructor-course and semester-course views (**R5**, **R6**). As shown in Fig. 9a, they found that the courses taught by the instructors “d4” and “e6” were closely associated with “Advanced Mathematics” and “Probability & Statistics.” What’s more, through inspecting the corresponding course curves, they further determined which courses the instructors should pay more attention to. Also, in the stacked area graphs of semesters, they observed that most of program courses highly correlated to “Advanced Mathematics” were distributed in the semesters of 2–2 or 3–1, mainly following the semester in which “Advanced Mathematics” was taken. But as for “Probability & Statistics,” there are more correlated program courses taken in the same semester, as shown in Fig. 10b. “Although we have known that students’ performance in mathematics courses would have influence on their studying in the following program courses, it is the first time for us to observe these distribution patterns in time series so clearly,” one expert said to another, “These phenomena can also be used to remind our instructors to pay attention to different kinds of mathematic knowledges in different academic stages of students, and to take into consideration the preceding mathematics courses in their teaching.”

6 General feedback and discussion

After the field experts had made several rounds of iterative exploration in this system, we conducted a discussion with them to further analyze the advantages and disadvantages of the system.

6.1 System advantages

The experts showed great interest in using MVCAS and believed that the significant advantage of this prototype was to support the exploration and analysis of the complex correlation patterns between general courses and program courses at multiple levels and from multiple perspectives. First, several types of course-associated information and statistical data, which were decomposed, extracted and transformed from the traditional student-course grade database, were of great meanings and values for studying student performance correlations among different types of courses. Second, various visual mappings and their combination provided novel and effective support for representing and discovering potential grade patterns between courses. Furthermore, several integrated views can help users to establish a more intuitive, sensitive and effective correlation model and support comprehensive analysis of the complex association between these two types of courses from the perspectives of students, courses, instructors and semesters, respectively.

The hierarchical correlation analysis process can also provide rich evidences and clues to students, instructors and administrators so as to jointly enhance the teaching and learning outcomes in program courses. Specifically, the obtained correlation patterns at the overall level can be used to predict the whole performance of the major curriculum and to prepare intervention measures in advance; the specific correlation patterns and phenomena, obtained at the classification and grouping level, were helpful to guide students to make efforts in learning certain types of program courses and helpful to remind the instructors and administrators to attach more importance to some particular groups of students as well as the connection between specialty knowledge and general mathematics knowledge in different semesters; the more detailed features derived from pairwise comparisons can be used to assist instructors to detect the intrinsic reasons for the association between their teaching courses and mathematics from the perspectives of knowledge and thinking.

6.2 System limitations

Although this visual analytics system can effectively support hierarchical analysis tasks from multiple perspectives, it still has some limitations, mainly including the following three aspects.

Analysis methods While MVCAS provides plenty of visual mappings and views to allow for intuitive exploration, there are still a few deficiencies in the fusion with automated data analysis methods in the current prototype. For instance, when exploring the detailed patterns of course–course correlations, users can only make subjective judgments on the differences among the score distributions. Perhaps it would be necessary to introduce a more quantitative measurement of grade similarity for helping users to conduct more efficient analyses between courses. In addition, the experts suggested that, besides the PCA, other statistical methods like MDS and interactive regression modeling [46] could be applied to assist the existing visual exploration.

Visual mapping As shown in Fig. 3a, each heatmap matrix is composed of multiple visual components with different styles. For example, besides the left-hand histogram and top gray bars, which represent some statistical information of students and courses, a stacked bar chart is added into each heatmap matrix to demonstrate the dynamic proportions of students in different grade clusters. Each row of heatmap matrices is further connected with the PCPs at different positions. This design requires bit learning for users who are not familiar with the multiform visualization. Exploring a more concise and straightforward visualization design is necessary to improve the ease of use. In addition, the polar scatterplot integrates multiple dimensional attributes of program courses. It is also a bit difficult to distinguish the exact course categories. Therefore, we plan to present more text hints and visual indicators along with user interaction.

Case study The current use case in MVCAS is mainly focused on the exploration of various correlations between mathematics courses and program courses from a special degree program. The type and number of general courses are still relatively limited. Different types of general courses could have influence on different parts of the curriculum. Moreover, there are different course modules and categories in different degree programs. So it is necessary to extend courses' and students' records to a larger scale and expand academic majors to a more diverse range in order to make further assessment of the system's applicability. In addition, it is also needed to conduct a more quantitative evaluation for further assessing user performance and experience [47].

7 Conclusion

This paper has taken a set of multi-level and multi-perspective approaches to design a visual analytics system named MVCAS. It allows analysts to explore the complex relationship between several important general courses and a wide range of program courses within a college major comprehensively. Through the multi-angle statistical computation, transformation and visualization of the traditional student grade data, MVCAS helps demonstrate and discover various correlation patterns between two types of courses at the overall, categorical and pairwise levels, respectively, and further helps investigate the potential causes. A series of multi-view coordinated layouts, such as the heatmaps combined with PCPs, pixel bar charts, polar scatterplots, etc., are used for representing course correlations from the perspectives of students, courses, instructors and semesters. The evidences and clues derived from the case study have demonstrated the effectiveness and usefulness of MVCAS in helping multiple participants, including administrators, instructors and students, to improve the teaching and learning process of program courses collaboratively. Moreover, this correlation analysis and visualization method can further be applied to similar visual analytics of multivariate data in other fields.

In the future, we plan to further boost the analytical efficiency and ability of MVCAS by introducing more quantitative measurements based on statistics models and machine learning methods. Meanwhile, since each of colleges and majors could have particular characteristics in the relationship between courses, we would like to extend MVCAS to more other academic majors' datasets so as to enhance the scalability and adaptability of visual analytical components.

Acknowledgements This research is partially supported by National Natural Science Foundation of China (No. 60873093). We further thank all anonymous reviewers for their valuable comments.

Compliance with ethical standards

Funding This study was funded by the National Natural Science Foundation of China (Grant No. 60873093).

Conflict of interest Lianen Ji declares that he has no conflict of interest. Yaming Yuan declares that he has no conflict of interest. Fang Gao declares that she has no conflict of interest.

References

1. Goolsby, C.B., Dwinell, P.L., Higbee, J.L., Bretscher, A.S.: Factors affecting mathematics achievement in high risk college students. Res. Teach. Dev. Educ. **4**(2), 18–27 (1988)
2. Bergin, S., Reilly, R.: Programming: factors that influence success. ACM SIGCSE Bull. **37**(1), 411–415 (2005)
3. Tessema, M.T., Ready, K., Yu, W.: Factors affecting college students satisfaction with major curriculum: evidence from nine years of data. Int. J. Humanit. Soc. Sci. **2**(2), 34–44 (2005)
4. Hudson, H.T., Rottmann, R.M.: Correlation between performance in physics and prior mathematics knowledge. J. Res. Sci. Teach. **18**(4), 291–294 (1981)
5. Malapati, A., Murthy, N.L.B.: Performance of students across assessment methods and courses using correlation analysis. In: 2013 IEEE International Conference in MOOC, Innovation and Technology in Education (MITE), pp. 325–328 (2013)
6. Zhang, Z., McDonnell, K.T., Mueller, K.: A network-based interface for the exploration of high-dimensional data spaces. In: 2012 IEEE Pacific Visualization Symposium, pp. 17–24 (2012)
7. Huamin, Qu, Chan, Wing-Yi, Anbang, Xu, Chung, Kai-Lun, Lau, Kai-Hon, Guo, Ping: Visual analysis of the air pollution problem in hong kong. IEEE Trans. Vis. Comput. Graph. **13**(6), 1408–1415 (2007)
8. Sukharev, J., Wang, C., Ma, K., Wittenberg, A.T.: Correlation study of time-varying multivariate climate data sets. In: IEEE Pacific Visualization Symposium, pp. 161–168 (2009)
9. Qu, H., Chen, Q.: Visual analytics for mooc data. IEEE Comput. Gr. Appl. **35**(6), 69–75 (2015)
10. Fu, S., Zhao, J., Cui, W., Qu, H.: Visual analysis of mooc forums with iforum. IEEE Trans. Vis. Comput. Graph. **23**(1), 201–210 (2017)
11. Ji, L.E., Gao, F., Huang, K.H., et al.: Visual exploration and analysis of multi-subject correlation of student performance in college courses. J. Comput. Aided Des. Comput. Gr. **30**(1), 44–56 (2018)
12. Romero, C., Ventura, S., García, E.: Data mining in course management systems: moodle case study and tutorial. Comput. Educ. **51**(1), 368–384 (2008)
13. Ding, S.: The Analyses of Training Quality and Influence Factors for the Undergraduates and Postgraduate. Doctor, University of science and technology of China, Hefei (2009)
14. Knauf, R., Kinshuk, Takada, K., Sakurai, Y., Kawabe, T., Tsuruta, S.: Personalized and adaptive curriculum optimization based on a performance correlation analysis. In: 2012 Eighth International Conference on Signal Image Technology and Internet Based Systems, pp. 655–660 (2012)
15. Pearson, K.: Liii. on lines and planes of closest fit to systems of points in space. Lond. Edinb. Dublin Philos. Mag. J. Sci. **2**(11), 559–572 (1901)
16. Kruskal, J.B.: Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika **29**(1), 1–27 (1964)
17. Singh, I., Sabitha, A.S., Bansal, A.: Student performance analysis using clustering algorithm. In: 6th International Conference—Cloud System and Big Data Engineering (Confluence), pp. 294–299 (2016)
18. Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., Rangwala, H.: Predicting student performance using personalized analytics. Computer **49**(4), 61–69 (2016)
19. Mazza, R., Dimitrova, V.: Coursevis: a graphical student monitoring tool for supporting instructors in web-based distance courses. Int. J. Hum. Comput. Stud. **65**(2), 125–139 (2007)
20. Gómez-Aguilar, D.A., Hernández-García, A., García-Peñalvo, F.J., Therón, R.: Tap into visual analysis of customization of grouping of activities in elearning. Comput. Hum. Behav. **47**, 60–67 (2015)
21. Ritsos, P.D., Roberts, J.C.: Towards more visual analytics in learning analytics. In: Proceedings of the 5th EuroVis Workshop on Visual Analytics, pp. 61–65 (2014)
22. Vieira, C., Parsons, P., Byrd, V.: Visual learning analytics of educational data: a systematic literature review and research agenda. Comput. Educ. **122**, 119–135 (2018)
23. Siirtola, H., Räihä, K., Surakka, V.: Interactive curriculum visualization. In: 17th International Conference on Information Visualisation, pp. 108–117 (2013)
24. Gama, S., Gonçalves, D.: Visualizing large quantities of educational datamining information. In: 18th International Conference on Information Visualisation, pp. 102–107 (2014)
25. Raji, M., Duggan, J., DeCotes, B., Huang, J., Vander Zanden, B.T.: Modeling and visualizing student flow. IEEE Transactions on Big Data(Early Access). (2018). https://doi.org/10.1109/TBDATA.2018.2840986
26. Wortman, D., Rheingans, P., Rheingans, P.: Visualizing trends in student performance across computer science courses. ACM SIGCSE Bull. **39**(1), 430–434 (2007)
27. Pryke, A., Mostaghim, S., Nazemi, A.: Heatmap visualization of population based multi objective algorithms. In: International Conference on Evolutionary Multi-Criterion Optimization, pp. 361–375 (2007)
28. Inselberg, A.: The plane with parallel coordinates. Vis. Computer. **1**(2), 69–91 (1985)
29. Gratzl, S., Gehlenborg, N., Lex, A., Pfister, H., Streit, M.: Domino: extracting, comparing, and manipulating subsets across multiple tabular datasets. IEEE Trans. Vis. Comput. Graph. **20**(12), 2023–2032 (2014)

30. Tableau software. http://www.tableausoftware.com. Accessed 15 May 2019
31. Tibco spotfire. https://www.tibco.com/products/tibco-spotfire. Accessed 15 May 2019
32. Few, S.: Show me the Numbers: Designing Tables and Graphs to Enlighten, 2nd edn. Analytics Press, USA (2012)
33. Keim, D.A., Hao, M.C., Dayal, U.: Hierarchical pixel bar charts. IEEE Trans. Vis. Comput. Graph. **8**(3), 255–269 (2002)
34. Li, J., Martens, J.B., van Wijk, J.J.: Judging correlation from scatterplots and parallel coordinate plots. Inf. Vis. **9**(1), 13–30 (2010)
35. Yan, C.W., Qian, Z., Feng, L., Chi, K.G., Seah, H.S.: Polarviz: a discriminating visualization and visual analytics tool for high-dimensional data. Vis. Comput. **35**(11), 1567–1582 (2019)
36. Viégas, F.B., Wattenberg, M.: Timelines: Tag clouds and the case for vernacular visualization. Interactions **15**(4), 49–52 (2008)
37. Collins, C., Carpendale, S., Penn, G.: Docuburst: Visualizing document content using language structure. Comput. Gr. Forum **28**(3), 1039–1046 (2009)
38. Tufte, E.R.: The Visual Display of Quantitative Information, 2nd edn. Graphics Press, Cheshire (1983)
39. Albo, Y., Lanir, J., Bak, P., Rafaeli, S.: Off the radar: Comparative evaluation of radial visualization solutions for composite indicators. IEEE Trans. Vis. Comput. Graph. **22**(1), 569–578 (2016)
40. Byron, L., Wattenberg, M.: Stacked graphs - geometry aesthetics. IEEE Trans. Vis. Comput. Graph. **14**(6), 1245–1252 (2008)
41. Zhang, Z., McDonnell, K.T., Zadok, E., Mueller, K.: Visual correlation analysis of numerical and categorical data on the correlation map. IEEE Trans. Vis. Comput. Graph. **21**(2), 289–303 (2015)
42. Ghoniem, M., Fekete, J.D., Castagliola, P.: On the readability of graphs using node-link and matrix-based representations: a controlled experiment and statistical analysis. Inf. Vis. **4**(2), 114–135 (2005)
43. Henry, N., Fekete, J.: Matrixexplorer: a dual-representation system to explore social networks. IEEE Trans. Vis. Comput. Graph. **12**(5), 677–684 (2006)
44. Bostock, M., Ogievetsky, V., Heer, J.: D^3 data-driven documents. IEEE Trans. Vis. Comput. Graph. **17**(12), 2301–2309 (2011)
45. Munzner, T.: A nested model for visualization design and validation. IEEE Trans. Vis. Comput. Graph. **15**(6), 921–928 (2009)
46. Mühlbacher, T., Piringer, H.: A partition-based framework for building and validating regression models. IEEE Trans. Vis. Comput. Graph. **19**(12), 1962–1971 (2013)
47. Isenberg, T., Isenberg, P., Chen, J., Sedlmair, M., Möller, T.: A systematic review on the practice of evaluating visualization. IEEE Trans. Vis. Comput. Graph. **19**(12), 2818–2827 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Lianen Ji is an associate professor of computer science in Beijing Key Laboratory of Petroleum Data Mining and College of Information Science and Engineering at China University of Petroleum, Beijing, China. His research interests include visualization and visual analytics of educational data and industrial control process data. Ji has a PhD in computer science from Institute of Software, Chinese Academy of Sciences.

Yaming Yuan received the B.Eng. degree in the Computer Science and Technology from China University of Petroleum, Beijing, China, in 2017. He is currently pursuing the M.Eng. degree in the College of Information Science and Engineering at China University of Petroleum, Beijing, China. His research interests include visual analytics of educational data.

Fang Gao received the B.Eng. degree in the Computer Science and Technology from Dalian University of Technology, Dalian, China, in 2014, and the M.Eng. degree from College of Information Science and Engineering, China University of Petroleum, Beijing, China, in 2017. She works at Research Institute of China Mobile Communications Company Limited, Beijing, China. Her research interests include visual analytics of educational data.