

Analysis of Student Retention and Drop-out using Visual Analytics

Jan Géryk

CVT and KD Lab

Faculty of Informatics, Masaryk University

Brno, Czech Republic

geryk@fi.muni.cz

Lubomír Popelínský

KD Lab

Faculty of Informatics, Masaryk University

Brno, Czech Republic

popel@fi.muni.cz

ABSTRACT

Student retention is an important measure for higher education institutions. Exploration and interactive visualization of multivariate data without significant reduction of dimensionality remains a challenge. Visual analytics tools like Motion Charts show changes over time by presenting animations within two-dimensional space and by changing element appearances. In this paper, we present a new visual analytics tool intended for exploratory analysis of educational data. We also utilized the tool for analyzing the data in order to verify the hypothesis concerning student drop-out behavior. The hypothesis assumes the existence of a correlation between the changes of fields of study and the student retention.

Keywords

Student retention, student drop-out, visual analytics, motion charts, animation.

1. INTRODUCTION

Higher education institutions have a major interest in improving the quality and the effectiveness of the education. In [1], hundreds of higher education executives were surveyed on their analytics needs. Authors resulted that the advanced analytics should support better decision-making, studying enrollment trends, and measuring student retention. They also pointed out that management commitment and staff skills are more important than the technology. In [2], authors concluded that the increasing accountability requirements of educational institutions represent a key to unlocking the potential for analytics to effectively enhance student retention and graduation levels. The application of data mining techniques in higher education systems have some specific requirements not present in other areas, as pointed out in [3].

Effective analysis depends on the consistent and high-quality data. Exploration and interactive visualization of multivariate data without fundamental dimensionality reduction remains a challenge. Animations represent a promising approach to facilitate better perception of changing values. In [4], authors pointed out that animations help to keep the viewer's attention. Correctly designed animations significantly improve graphical perception at both the syntactic and the semantic levels, as concluded in [5]. However, visualizations are often engaging and attractive but a naive approach can confuse the analyst. Motion Charts represent an animated data presentation method which shows multiple elements and dimensions on a two dimensional plane, as described in [6].

Motion Charts allow exploring and formulating additional hypotheses, as well as it helps to easily identify hidden patterns and trends in the data. The variable mapping is one of the most

important parts of the exploratory data analysis. Both the data characteristics and the investigative hypothesis should influence the selection of a variable mapping.

In this paper, we briefly describe the motivation and design of the enhanced version of the Visual Analytics (VA) tool EDAIME, firstly introduced in [7]. In the next section, we present several papers concerning with data analysis using Motion Charts. Subsequently, we describe the design of the tool. Then, we make use of the tool for analyzing educational data in order to verify the hypothesis concerning student dropout behavior. Finally, we conclude the paper with future work and summarize the conclusion of the results.

2. RELATED WORK

Number of papers concerning the Motion Charts has increased recently. In [6], authors incorporated examples using recent business and economic data series and illustrated how Motion Charts can tell dynamic stories. For the first analysis, they utilized data about Current Employment Statistics and presented differences between the perception of common static tables and graphs, and the dynamic manner of Motion Charts. They concluded that static presentation style serves well the purpose of relaying accurate and non-biased quantitative data to the analyst. They also emphasized that the benefit of Motion Charts lays in displaying rich multidimensional data through time on a single plane with the dynamic and interactive features. Users are allowed to easily explore, interpret, and analyze information behind the data. They concluded that the Motion Charts are an excellent and interesting way of presenting valuable information that may otherwise be lost in the data.

Beneficial feature for better visual perception of changes in time-series analysis is presented in [8]. Author emphasized both the benefits and the drawbacks of common data visualization methods, namely line chart and bar chart. Then, the author focused on dealing with issues with the time-series analysis. Subsequently, he presented capabilities of Motion Charts which are more suitable for this kind of analysis. Moreover, author stressed that patterns of change through time can take many meaningful forms and introduced new feature, called visual trails, designed for Motion Charts which allows seeing the full path that elements take from one point in time to another. He also demonstrated proposed improvements.

3. THE EDAIME TOOL

Two main challenges are addressed by the presented VA tool EDAIME. The tool enables visualization of multivariate data and the interactive exploration of data with temporal characteristics. Moreover, it is optimized to process educational data.

The main purpose of the tool is to increase the education effectiveness and the quality of the study. The motivation to develop an enhanced version of Motion Charts was to extend their abilities and to improve the expression capability to facilitate analysts to depict each student as the central object of interest. Moreover, the implementation enhances the portfolio of animations that express the student's behavior during their study more precisely. The main technical advantages over other implementations of Motion Charts are its flexibility, the ability to manage many animations simultaneously. Optimizations of the animation process were necessary, since even tens of animated elements significantly reduced the speed and contributed to the distraction of the analyst's visual perception.

The Force Layout component of D3 (<http://d3js.org/>) provides the most of the functionality behind the animations, and collisions utilized in the interactive visualization methods. Linearly interpolated values are calculated for missing and sparse data.

3.1 Analysis of Educational Data

The main aims to improve student retention and graduation levels, are closely connected with analyses of changes of the mode and changes of the field of study. We utilize the EDAIME tool for analyzing educational data in order to verify hypothesis concerning with student dropout behavior. The hypothesis supposes the existence of a correlation between the changes of fields of study and student retention.

The large elements that represent a particular field of study consist of small elements that represent individual students. Therefore, the size of the large elements corresponds to the number of students enrolled in a particular field of study. The size of the small elements corresponds to the number of credits gained in a particular semester of study.

Besides the study progress, animations are also utilized to express the study termination, the change of the mode of study and the change of the field of study. Dropout students turn red and fall down the chart in the semester when they left the studies. The stroke-width of the elements represents the state of the study and the element color represents the attributes of the study.

To verify the aforementioned hypothesis we examined educational data about students admitted to bachelor studies of the Faculty of Informatics Masaryk University between the years of 2006 and 2008. The semester number is mapped to time variable. The grade point average is mapped to x-axis. The average number of credits is mapped to y-axis. The number of gained credits is mapped to element size.

Motion Charts show that the number of students decreases in all fields of study besides applied informatics (BcAP) because it is frequent target of change for the students in the first two semesters. After that, the number of students decreases uniformly for all fields of study. It is visually clear that the majority of students change the field of study to BcAP. More precisely, the highest migration between two fields of study is from computer graphics (GRA) to BcAP. Analysis show that the most student dropouts occur in the freshmen year, but over the time the number of unsuccessful students decreases significantly. Motion Charts also reveal that the ratio of the number of successful students to the number of unsuccessful students is significantly higher for students that changed their field of study. The supposed correlation exists, but a further analysis with a different mapping

is needed to better express the relation between the migration target and the study success.

4. CONCLUSION

Common data visualization methods have limitations in terms of the volume and the complexity of the processed data. Motion Charts are transparent methods that can present a good overview of the complex data and also enable analyst to observe interesting elements while the previous ones are still fresh in his or her memory.

In the paper, we have described the motivation and design of the VA tool EDAIME which is intended for exploratory analysis of educational data. We enhanced the concept of Motion Charts and successfully expanded it to be more suitable for such analyses. We have successfully employed it to verify the suggested hypothesis. A further in-depth analysis with different mapping of variables is needed to quantify the correlations more accurately. Despite the fact that common data visualization methods are quite beneficial, there are types of questions that cannot be examined using them. Since the questions involve quantitative relationship other than change through time.

The additional representation of the data gives the analyst more possibilities in exploring the data, but the additional functionality can also confuses the analyst. To verify user friendliness and usability of the tool, we will carry out a controlled experiment with two groups of users. They will use different VA tool and methods trying to understand the same dataset.

5. ACKNOWLEDGMENTS

We thank Michal Brandejs and Knowledge Discovery Lab for their assistance. This work has been partially supported by Faculty of Informatics, Masaryk University.

6. REFERENCES

- [1] Goldstein, P. J. and Katz, R. N. 2005. Academic Analytics: The Uses of Management Information and Technology in Higher Education, ECAR Research Study Volume 8.
- [2] Campbell, & Oblinger, D. 2007. Academic analytics. Washington, DC: EDUCAUSE Center for Applied Research.
- [3] Delavari, N. and Phon-Amnuaisuk, S. and Beikzadeh, M. R. 2008. Data Mining Application in Higher Learning Institutions. Informatics in Education, 31-54.
- [4] Tversky, B. and Morrison, J. B. and Betrancourt, M. 2002. Animation: Can It Facilitate? International Journal Human-Computer Studies, 247-262. DOI=10.1006/ijhc.2002.1017.
- [5] Heer, J. and Robertson, G. 2007. Animated Transitions in Statistical Data Graphics. IEEE Transactions on Visualization and Computer Graphics, 1240-1247.
- [6] Battista, V. and Cheng, E. 2011. Motion Charts: Telling Stories with Statistics. JSM Proceedings, Statistical Computing Section. Alexandria, VA: American Statistical Association, 4473-4483.
- [7] Géryk, J. 2013. Visual Analytics by Animations in Higher Education. In Proceedings of the 12th European Conference on e-Learning ECEL 2013, 565-572.
- [8] Few, S. 2007. Visualizing Change: An Innovation in Time-Series Analysis. In Visual Business Intelligence Newsletter.