The EDM Vis Tool

Matthew W. Johnson, Michael John Eagle, Leena Joseph and Dr. Tiffany Barnes
University of North Carolina at Charlotte

We introduce EDM Vis, an information visualization software tool for exploring, navigating and understanding student data logs. EDM Vis loads student logs, entire classes, hundreds of students, at a time, and displays the student behavior of those students as they solved problems using a software tutor. The visualization uses a tree structure to provide an overview of class performance, and interface elements to allow easy navigation and exploration of student behavior.

---

1. INTRODUCTION

"Advancing personalized learning" has been declared a Grand Challenge by the National Academy of Engineers. With increasing use of the web for instructional materials and learning management systems, the amount of data available to help address this challenge is growing rapidly. However, these large datasets on learning can be unwieldy and deciding just how to use them for making learning more effective is also a challenge. The PSLC DataShop, a repository for educational data, has collected logs from over 42,000 students from different tutors with a wide range of topics, from algebra to Chinese [Koedinger et al. 2010]. The DataShop was also used for the 2010 Knowledge Discovery and Data Mining Cup challenge, illustrating that mainstream data mining conferences are realizing the growing need to understand educational data sets.

Anscombe in 1973 identified some key problems in understanding data through statistical measures alone, demonstrating that the same mean and standard deviations can be used to summarize a "quartet" of distinctly different data sets. "A computer should make both calculations and graphs. Both sorts of output should be studied; each will contribute to understanding."[Anscombe 1973]. Visualization can help avoid misinterpretation of data. In 2002, Ben Shneiderman wrote an influential article, *Inventing Discovery Tools: Combining Information Visualization with Data Mining*, highlighting that both communities, InfoVis and Data Mining, should "integrate data mining and information visualization to invent discovery tools."[Shneiderman 2001]. Stuart K. Card defines the purpose of visualization to "amplify cognition" about data [Card et al. 1999], in our case, amplify an educator's cognition about student behavior.

Being able to see and use step-by-step data logs is an important advantage Intelligent tutoring systems can have over traditional homework methods, but this advantage will not been fully leveraged until Visualization systems can be developed to make the data consumable by the masses. In this paper, we report on our attempt to combine InfoVis and data mining to invent a "discovery tool": EDM Vis. We describe some of the design decisions we made when developing EDM Vis and highlight some of the key considerations we made when developing a visualization tool for log data from students performing procedural problem solving.

2. THE EDM VIS TOOL

EDM Vis is a software visualization tool built to provide educators and researchers a graphical representation of tutor log data, to help improve understanding of student problem-solving and to provide insight about how to improve teaching and learning. EDM Vis is intended to be domain-independent and useful interactive visualization for procedural problem domains.

We designed our visualization tool for visualizing procedural domains, domains requiring a series of steps in order to solve a problem so we use a hierarchical graph representation. In the graph, each node represents a state and each edge an action that takes a user from one state to the next. In this way, we can allow users to visualize an entire set of student attempts at once providing an effective overview of the solution domain as created by the students working in the tutor. An additional inherent requirement is that the tutor's data can be expressed in the form of states, typically the values of the variables in the tutor at discrete times. In general, this is reasonable because most logging systems in tutors log their data in this way.

In addition to the required data: student-ID, start-state, action, end-state, we also support loading in additional data which is stored on the student. Future versions will also support storing extra data on states and actions. This functionality allows users to, for example, load test-scores for the students and compare their actions to other students based on their performance on, for example a pre-test.

3. CONCLUSION AND FUTURE WORK

We presented EDM Vis a software visualization tool designed for exploring and interacting with software-tutor log data. Our tool allows users to get a better understanding, over looking strictly at tutor logs, of how students interacted with a software tutor by visualizing the tutor's interaction logs.

A short survey we provided to our users gave us helpful feedback for areas to improve on in future interactions of the tool. EDM Vis is our first steps to providing a domain independent visualization tool for understanding student behavior in software tutors, and our initial results seem promising for the future development of this tool.

Allowing the user to annotate and augment their data in the visualization seems useful, particularly for highlighting an important aspect of the data a user wishes to share. In addition a method of allowing comparison between problems could prove interesting, for both looking at a single student over multiple problems as well as the same problem from multiple tutors.

It is also clear that for each graph and sub-graph we should provide some general statistics to supply the user with concrete numbers, essentially more details on demand. In addition, we must fix some issues with our use of color, making it easier to know which nodes are errors when selected. Another advanced functionality would be to allow selections to be unions, intersections or excludes of sub-selections, as well as a way to check if student-1 performed action-X did, they also perform action-Y. These types of selection and filtering and their extensions seem to be a good direction to pursue for the next iteration of EDM Vis. Lastly, the advice we gathered from the qualitative portion of the survey will certainly be taken into account and implemented in all possible cases.

REFERENCES

ANSCOMBE, F. J. 1973. Graphs in statistical analysis. *The American Statistician 27,* 1, pp. 17–21.

CARD, S. K., MACKINLAY, J. D., AND SHNEIDERMAN, B. 1999. Readings in information visualization. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, Chapter Using vision to think, 579–581.

KOEDINGER, K., BAKER, R., CUNNINGHAM, K., SKOGSHOLM, A., LEBER, B., AND STAMPER, J. 2010. A data repository for the edm community: The pslc datashop.

SHNEIDERMAN, B. 2001. Inventing discovery tools: Combining information visualization with data mining. In *Discovery Science,* K. Jantke and A. Shinohara, Eds. Lecture Notes in Computer Science Series, vol. 2226. Springer Berlin / Heidelberg, 17–28.