Improving the Prospects for Educational Data Mining

Steven L. Tanimoto

University of Washington
Dept. of Computer Science and Engineering
Box 352350, Seattle, WA 98195, USA
tanimoto@cs.washington.edu

Abstract. Data mining is an important paradigm for educational assessment. The usual assumption is that mining is performed after educational activity with that activity having been designed without regard for the mining process. This paper discusses how the prospects for successful mining can be improved by imposing constraints or biases on the activities and instruments that generate the data. These biases involve one or more of the following: (a) encouraging, requiring or training students to communicate effectively and often during the course of learning activities, (b) building more instrumentation into the learning environment to enable capturing more kinds of data, including evidence of student attention, (c), enriching the logged expressions themselves so that more inferences from them can be made more easily and with general purpose tools, and (d) seeding the log files with reliable assessment data to help anchor subsequent inferences. A variation on the mining paradigm integrates mining methods into the learning environment itself, so that various forms of "articulated assessment" can become practical. Articulated assessment is the coordination of unobtrusive but less reliable assessment techniques with traditional direct-questioning methods in such a way as to follow a policy that balances the needs for accuracy and unobtrusiveness.

Keywords: educational assessment, data mining, articulated assessment, unobtrusive assessment, online learning environments, intelligent tutoring systems, student modeling.

1 Introduction

Data mining is an important data analysis methodology that has been successfully employed in many domains, and which has become especially popular after the World Wide Web made large volumes of data on many topics widely available. It has been used to analyze byproducts of intelligent tutoring system sessions and other educational activities for purposed of evaluating the activity, the systems, or building models of students or their interactions with systems. Data mining has also been considered as a methodology for extracting shorter-term educational assessment data in order to fill out components of student models such as average time on task, attention span, etc.

The arguments put forth in this paper are intended to help reach the goals of greater accuracy in the results of mining, wider latitude in the scope of questions that can be effectively answered by mining, and greater transparency in the inference processes.

2 Supporting Unobtrusive Assessment

The University of Washington's project on intensive, unobtrusive assessment seeks to harness the full power of computers in making useful assessments of student learning "behind the scenes." The keystone in this project is a system called INFACT that facilitates the creation and capture of evidence of student learning while students engage in problem-solving and construction activities. Before giving a brief description of INFACT, here is the motivation for unobtrusive assessment.

2.1 Goals of Unobtrusive Assessment

Foremost in our motivation is the desire to improve student learning in the context of problem solving and artifact construction. The assessment involves diagnosing student misconceptions and problematical habits. The results are used by teachers and systems to make opportune suggestions to students, select assignments and make pedagogical decisions. The need for unobtrusiveness is a reflection of the cost of interventions with tests, in terms of student motivation to learn and satisfaction with the activity. Another reason to develop unobtrusive assessment is to take advantage of interaction data and records of student communication that are already captured by the computer-based learning environment. Yet one more hope for unobtrusive assessment is that it can be continuous, so that needless gaps in the system's knowledge of the student's cognitive state can be avoided. While unobtrusive techniques might never fully replace traditional testing methods, they may provide new options to teachers and learners that can be used to adapt the pedagogical environment to their needs.

2.2 The INFACT Online Learning Environment

When we created INFACT, we set out to "computerize" the facet-based teaching approach successfully developed for physics (Minstrell, 1992). In this method, students are challenged to predict or explain phenomena that go against their intuitions. In their discussions, they reveal their preconceptions. Their ideas are diagnosed by the instructor, using a catalog of previously observed misconceptions as a guide. Then the teacher presents them with special examples that confront their misconceptions.

The original purpose of INFACT was to host these discussions (and thus obtain a record of them) and to facilitate the diagnosis by also hosting the catalog of misconceptions and providing a database facility for recording the diagnoses (Tanimoto et al, 2000). INFACT stands for Integrated, Networked, Facet-based Assessment Capture Tool. Unlike the DIAGNOSER tool, which makes facet diagnoses according to the results of multiple-choice testing (Levidow et al, 1991), INFACT was designed to support the inference of facets directly from the records of student discussions.

To explain what INFACT is, let's consider the services it provides. At the heart of INFACT is the "Forum" which is a group-oriented written discussion area that uses a threaded-newsgroup format. INFACT has special features for controlling the visibility of student messages (Tanimoto et al, 2002). Closely associated with the forum is a graphical communication tool called INFACT-Sketch that supports "conversational sketching." Around this core of communication tools are computation and construction tools for students, such as a programming facility and an image processing system. Teachers have access to administration and assessment tools that include a markup (annotation) facility for making free assessments and facet-based assessments, an editor for facet catalogs, editors and application monitors for rule-based and Bayes-net based automatic assessment, and an editor and administration facility for traditional multiple-choice testing. Facilities are also included for file sharing by students and teachers, and visualization of assessment data by teachers. Additional details are given in (Tanimoto et al, 2005).

2.3 The Relationship between Data Mining and Unobtrusive Assessment

Post-logging data mining by itself is unobtrusive on one level, because the session is over, the student has gone home, and is not bothered by the system when the inferences are made. Data gathering and logging, on the other hand, may be obtrusive or unobtrusive (depending on how the data is generated and collected), but data mining is philosophically attuned to unobtrusive assessment because of the decoupling of inference from logging. Nonetheless, we can consider some degree of coupling. We do this not to make data mining an intrusive process, but in order to help it make better inferences. The remainder of this paper discusses several approaches.

3 Helping Students Communicate More Readily and Clearly

One important avenue for "enriching the ore" for mining is to increase the amount and the quality of student communication during sessions, and to incorporate this into the log data stream (David, 2005; Mostow, 2004). With INFACT instructors can easily require such communication in assignments. In addition, students can be taught to communicate more openly through specific training exercises. Graphical communication with INFACT-Sketch is a case in point. We use two particular play activities to get students used to conversational sketching. One activity is "Graphical Telephone," in which one student draws an object and passes the sketch to another, who makes a minor modification, after which it is passed to another, etc. It is particularly amusing when the starting subject is the face of a groupmate. The other activity, "Collaborative Comics," is a group storytelling one, in which the first team member creates the first frame of a comic strip, and passes the sketch on to the next team member, etc. These activities help build habits of group interaction via graphics.

At this time, INFACT does not capture audio or video of students during sessions. However, a written equivalent of "think-aloud" activity ("thinking in the fingers") is a viable methodology for capturing more evidence about student cognition. Such behavior can be encouraged through credit-awarding schemes (participation points, etc). A related idea is to engage students at two levels during their problem solving discussions. While they are direct participants in the discussion, they can also be tasked with evaluating the contributions of their classmates through rating mechanisms. Such ratings can serve as extra hints to data mining methods that particular messages or excerpts are worthy of extra attention, or that they may serve to ground inferences from messages that relate to them.

4 "Capturing" Student Attention

Here we really mean capturing *evidence* of student attention. Unless eye-tracking systems are incorporated into the learning environment, it is difficult to know whether a student is actively reading something on the screen or simply daydreaming or tuned out. One approach to better capturing this information is to redesign the interface so that some amount of additional interaction is encouraged and/or required for the reading. This means rethinking observation processes, transforming them from relatively passive activities to explicitly active processes. To turn the experience of reading a page from simply an eye-moving activity to a combined eye-moving and mouse-moving/clicking activity requires two changes: a change of the widget that renders the page, and a change of the information structure to make it hierarchical. A two-level hypertext structure may be sufficient to achieve the goal. This "active observation" strategy thus encourages active observation and it makes passive observation more difficult. The extra level of activity required of students should not be so much as to be a burden. If it easily leads to repetitive stress injury or a much slower rate of reading, then it has gone too far. It is particularly

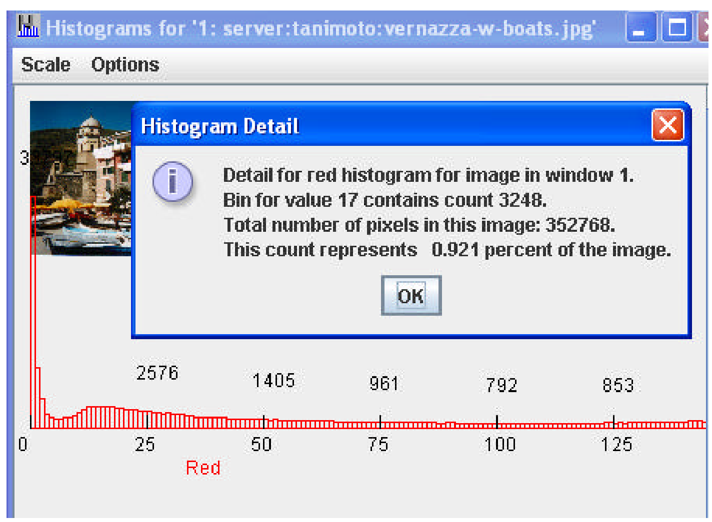

Figure 1. Illustration of an observation tool that encourages but does not require active observation. When the student clicks on a histogram bin, detailed information about the bin is provided in the dialog box. A log file record corresponding to this instance of active observation is created.

valuable when the material is so dense and rich that students would normally spend a lot of time on it with little explicit interaction, leaving assessment and mining processes in the dark about what they were thinking. An example of an observation instrument for students that takes the unobtrusive approach is a histogram display for images shown in Fig. 1. When the student asks for the display, three full histograms for an image (one histogram for each of the red, green and blue components of the color image) are shown. However, by clicking on individual bins of the histogram, the student can get an exact count of the number of pixels with that value as well as a percentage value for the fraction of the image represented by that bin. When the student clicks, an observation event is registered in INFACT, and the log ends up containing a representation of this observation made by the student.

The image processing system PixelMath, hosted by INFACT, allows students to inspect image pixels in a somewhat unusual way. They can zoom and unzoom as with many image programs, but when they zoom in far enough, they see the numeric pixel values superimposed on the colored pixel squares. In any activity that requires students to work closely with the numeric values of particular pixels, the zooming and unzooming event records represent the student's focus of attention fairly well. Sometimes, students get lost in the details of an image. Navigation can be difficult because of the size and complexity of the image. Dead ends in the navigation can lead to log files polluted with events that do not necessarily reflect an investment of hope by students in their relevance to a task. However, these events do represent the navigation trouble. A smart analysis system needs to be able to distinguish between such navigation problems and intended observations. This is a possible challenge for the incorporation of active observation mechanisms in learning environments.

5 Making Log Files More Expressive

While the enrichment methods described in the preceding two sections involve changes to the student experience, another method is not dependent on making such changes. Instead, it involves altering the representation of events in the log file. Foremost in this approach is overcoming a historical tendency to make log files cryptic in order to save file space. The changing economics of disk space should make us adopt the most robust representation techniques, not the absolutely cheapest ones. Four approaches are these: (a) representing each event completely, (b) using English words, (c) using English grammar, and (d) using standard log-file forms. The first of these means making each log-event record almost self-contained. Distribution of meaning in log-file records among code words and corresponding dictionaries elsewhere adds to the challenges of interpreting the records and inferring patterns from them. Another way of describing this approach is having the system that administers the educational experience "connect the dots" within the log to reduce the likelihood of post-activity inference errors.

The use of English words in log file entries can facilitate (a) the use of general data-mining tools, and (b) a human's configuration of such tools. Many tools are designed to process natural-language text. In order to apply them, their target data must be in the form of text, not encoded binary data. Data mining tools are often exploratory pattern analysis tools that benefit from human-expert configuration or guidance. The use of English terms is likely to help these experts keep track of the meanings of record components and apply common-sense reasoning to the task of configuring the data-mining algorithms and evaluating their results. The use of natural language grammar is an extension of the idea of using English terms. The one caveat here is that English grammar admits a wide variety of forms, and it is a good idea to use a small number of simple, standard forms to avoid the need for parsing or ambiguity resolution during data mining. One approach toward standardization is the creation of a language or metalanguage for expressing the format of log files (Iksal and Choquet, 2005). It may be too soon to try to standardize log files, because they require agreement at the level of ontologies, not just formats. However, if log files from multiple environments are to be integrated by data mining systems, it would help if they adhered to standards.

6 Seeding Log Files with "Ground Truth"

Another way to make log files easier to interpret is to alter the educational environment and experience somewhat, so that a limited amount of hard-core educational assessment data is captured, analyzed and the results entered into the log file up-front. This may be easy, in intelligent tutoring systems where such assessment may already be performed (Mostow, 2004); but it may require a change in the student experience for constructive tools, such as a dynamic geometry program, a computer software development environment, or a circuit simulation system. Such information would typically have to be obtained with obtrusive interventions involving multiple-choice testing or other highly-directed student tasks. The benefit of such information is that such anchored assessment elements could serve as the seeds that at data-mining time could grow into islands (or a continuum) of highly reliable inferences. An analogy to video representation with the MPEG standard may be instructive: the MPEG data stream includes special video frames that are complete and accurate representations of the video signal at key points in time. The other frames in the video sequence are expressed in terms of these key frames using difference expressions, which tend to be much more compact than full frames. However, the intermediate frames are not completely accurate. The key frames are required to anchor the evolving scene to prevent errors from accumulating too much. With this approach, systems like INFACT need to use a technique we call "articulated assessment." Articulated assessment is a combination of traditional (obtrusive) educational assessment and unobtrusive assessment under the administration of an agent that dynamically optimizes the balance of the two to obtain the requisite accuracy and pedagogical characteristics.

7 Human vs. Automatic Data Mining, and Transparency for Students

Does it matter whether the data mining will be performed by humans or machines? Baker et al found that, as a substitute for video and other high-fidelity recordings, textual descriptions can be devised that serve the main purposes almost as well (Baker et al, 2005) when analysis will be done by human coders. The suggestions we have given for enriching log files should apply no matter whether humans or automated agents are performing the analysis. We can imagine that exploratory data mining will best be done by humans interacting with statistical tools. The log files need to be intelligible to both.

There is a possible side benefit of improving the richness of the log file for data mining. That is allowing the capture and assessment processes within the learning environment to be more transparent. By opening up a view of the log stream to the students, they may get a better understanding of how they are being assessed. Such transparency is consistent with the philosophy of supporting open learner models (Bull, 2004) and is a subject of current research. Intelligibility of the log files to students is then a key factor in the success of the transparency in engendering understanding and trust in the system.

8 Conclusion

There is a variety of ways that activity logs produced by computer-based learning environments can be enriched to make their subsequent analysis more accurate and fruitful. While some of these involve changing the student experience, others have only to do with the way that logged events are formulated. They all involve thinking of the computer-based educational learning environment and the data mining system as parts of a larger, integrated process.

References

1. Baker, R., Corbett, A. T., and Wagner, A. Z. Human classification of low-fidelity replays of student actions. *Proceedings of the Workshop on Educational Data Mining* (held at the 8th International Conference on Intelligent Tutoring Systems -- ITS 2006). Jhongli, Taiwan, pp.29-36.
2. Bull, S. Supporting learning with open learner models. *Proc. 4th Hellenic Conference with International Participation: Information and Communication Technologies in Education*, Athens, 2004.

3. David, J.-P. State of art of tracking and analyzing usage. Kaleidescope Project report D32.3.1 Final. Online at http://telearn.noe-kaleidoscope.org/warehouse/JeanPierre-David-2005.pdf
4. Iksal, S., and Choquet, C. Usage analysis driven by models in a pedagogical context. *Proc. AIED Workshop on Usage Analysis for Learning Systems*, 2005. http://lium-dpuls.iut-laval.univ-lemans.fr/aied-ws/PDFFiles/iksal.pdf.
5. Levidow, B. B., Hunt, E., and McKee, C. 1991. The DIAGNOSER: A HyperCard tool for Building theoretically based tutorials. *Behavior Research Methods, Instruments, and Computers*, Vol. 23, 249-252.
6. Minstrell, J. Facets of students' knowledge and relevant instruction. In Duit, R., Goldberg, F., and Niedderer, H. (eds.), *Res. in Physics Learning: Theo. Iss. & Empir. Stud*. Kiel, Germany: Kiel Univ., Inst. for Sci. Educ., 1992.
7. Mostow, J. Some useful design tactics for mining ITS data. *Proceedings of the ITS2004 Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes*, August, 2004.
8. Tanimoto, S. L., Carlson, A., Hunt, E., Madigan, D., and Minstrell, J. 2000. Computer support for unobtrusive assessment of conceptual knowledge as evidenced by newsgroup postings. *Proc. ED-MEDIA 2000*, Montreal..
9. Tanimoto, S., Carlson, A., Husted, J., Hunt, E., Larsson, J., Madigan, D., and Minstrell, J Text forum features for small group discussions with facet-based pedagogy. *Proc. CSCL 2002*, Boulder, CO (2002).
10. Tanimoto, S. L., Hubbard, S., and Winn, W. 2005. Automatic textual feedback for guided inquiry learning. *Proc. of the 12th Int'l. Conference on Artificial Intelligence in Education* (AIED 2005). Amsterdam.