Learning Analytics: Targeting Instruction, Curricula and Student Support

Craig Bach
Office of the Provost, Drexel University
Philadelphia, PA 19104, USA

ABSTRACT

For several decades, major industries have implemented advanced analytics and decision support structures to advance and support their goals. More recently, institutions of higher education are starting to adapt these methods to target fund raising, inform enrollment decisions, target marketing efforts, improve student support processes, and to better understand retention/persistence patterns. Separately, regional, national, and specialized accreditors, as well as the federal government, are ratcheting up expectations around learning outcomes assessment (e.g., articulation of measurable learning outcomes, assessment of student achievement of those outcomes, and the use of resulting data). Both threads, weaving their way through institutions of higher education, are coming together in the area of *learning analytics* (or, *academic analytics*). This paper outlines a conceptual framework for the development of learning analytics, highlighting lessons learned from industry, limitations of the approach, and important ethical issues involved in the application of these methods to educational contexts.

Keywords: Higher Education, Analytics, Predictive Modeling, Outcomes Assessment, Learning, Decision Support, Knowledge Management

1. INTRODUCTION

Predictive modeling (e.g., logistic regression, neural networks, decision trees, support vector machines, survival analysis) and segmentation modeling techniques (e.g., clustering analysis, categorization analyses) have been used extensively in a range of industries to target resources and support goal achievement. In the insurance industry, for example, these techniques are regularly used to target and customize direct mailing campaigns in order to reduce mailing costs and increase yield, or to predict customer retention [12]. The pharmaceutical industry uses advanced modeling techniques to determine the efficacy of drug interventions and predict patient survival rates.

While the specific techniques differ depending on context and the intended goals of the modeling effort, the general approaches of segmentation and predictive modeling are straightforward. The main goal of segmentation modeling is to separate a population into groups that, in aggregate, behave significantly differently with respect to a desired behavior (i.e., buying a particular insurance policy), or conversely, who look similar based on key demographic, psychographic, or behavioral variables.

Predictive modeling is used to identify the likelihood that members of a population achieve an identified end, make a decision, or take a particular action – attaching a probability to each member of the population. The end user is then able to set appropriate cut levels that determine ranges of actions (i.e., population members with attached 75% probability of responding to a tailored mailing, will get sent the mailing). The predictive model optimally combines the factors in the data set that drive the desired action. Some techniques (e.g., logistic regression, decision trees) specifically identify those combinations of factors and can inform the end user about the relations among variables that support an action. Other techniques (e.g., neural networks, support vector machines) score individuals, but do not provide the relations among variables that determine a score.

In the main, the models are developed based on historical data and then used to score (attach probabilities to) individuals in the decision-targeted population (e.g., potential customers, patients, or policy holders). The resulting scoring provides insight into the new population that helps target resources while improving returns.

2. TRANSFERRING PRACTICE TO EDUCATIONAL CONTEXTS

One of the biggest challenges in using data to inform improvements in learning is the sheer volume of available data. Basically, there is far more data available than can meaningfully be used. Attempts to filter data and focus collection efforts are ongoing challenges and it is often left to end-users to sift through data to find those specific tidbits that are meaningful to them. Support for these decisions is often provided by benchmarking data, highlighting specific results, or running basic tests of significance. But, we need to be much smarter about this effort – here is where learning analytics provides an indispensible tool.

Higher education has been slow to adopt analytics. Initial use in academe was focused on areas most closely aligned to their business counterparts (e.g., marketing, enrollment *qua* customer insight). However, as Campbell and Oblinger point out [3], the application of advanced modeling techniques is being used to support more core academic functions. Currently, most of this effort is focused on better understanding retention/persistence (particularly first-year retention) and enrollment yields.

More broadly, the development of analytic capabilities and the use of these techniques to drive decision-making are most often focused in just a few areas of the institution. Figure 1 displays these areas and their relationships to each other in the context of discussing institutional effectiveness in achieving a mission.

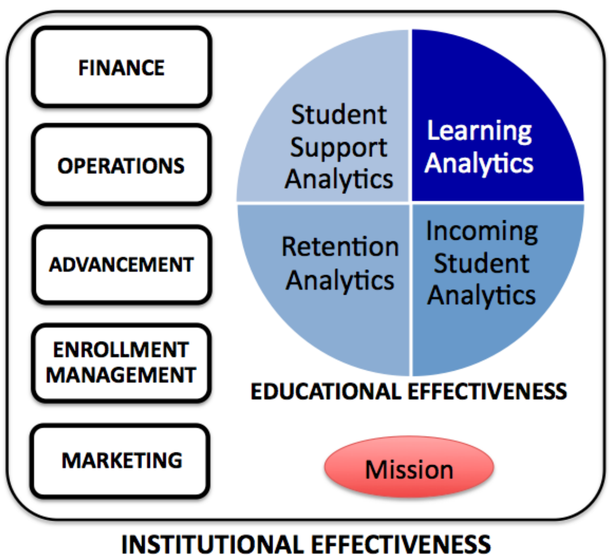

Figure 1. Areas of Analytic Concentration

The areas along the left side of the diagram are the main business areas of the institution with evidenced use of advanced modeling or analytics. In the case of operations, these may be process improvement efforts (e.g., six sigma) that form part of larger modeling work. These areas are usually the drivers of institutional effectiveness and advanced modeling techniques have been used successfully to support and document this effectiveness.

Educational effectiveness is one part (arguably the most important part) of the overall institutional effectiveness of the institution. Historically, educational effectiveness has been measured by broad outputs (graduation rates, mean GPAs, graduate school acceptance rates, employment rates) and inputs (SAT/ACT scores of incoming fist-year students). It is probably not a coincidence that these areas have been the main focus analytic efforts. Less attention has been paid to borrowing analytic methods from operations to improve student-facing processes even though these service and support areas have been shown to have a large impact on student success

However, the current expectation among accreditors is to measure educational effectiveness through learning outcomes assessment. Learning outcomes assessment involves:

- The articulation of measurable learning outcomes,
- The identification of where articulated learning outcomes are supported (e.g., in the curriculum, in other educational activities),
- A method of assessing student achievement of the articulated outcomes, and
- The use of assessment data to make improvements to instruction, curricula and learning.

Learning analytics involves the use of advanced modeling techniques integrated with learning outcomes assessment to better understand student learning and more efficiently and meaningfully target instruction, curricula and support. This area is one of the least developed and most promising areas of analytic work.

3. LEARNING ANALYTICS

Learning analytics is defined as the use of predictive modeling and other advanced analytic techniques to help target instructional, curricular and support resources to support the achievement of specific learning goals. One of the key data items involved in learning analytics is learning outcome achievement data. As such, the success of a learning analytics effort is dependent upon the quality of the learning outcomes assessment data available and the reliability of the measurements used to collect those data[1].

However, learning analytics also make use of a wide range of other kinds of learner characteristic data often used in other modeling efforts. In fact, learning analytics can be seen as the refinement of enrollment, retention, persistence, and graduation models with the introduction of learning outcomes and learning characteristic data.

[1] This is not to say that learning analytics cannot be implemented with less than ideal outcomes data (something I will discuss later in this paper).

Other data types include:

- Test scores (e.g, ACT, SAT)
- Class grades
- Demographic, psychographic data
- Learning styles, characteristics or preferences data
- LMS/CMS activity data
- Survey data

The possible data points are rich and diverse and much of the work is exploratory (identifying new variable groups to develop and run through the process). As with most analytic work, the main goal is to identify the most relevant and actionable drivers of learning outcome(s) achievement.

Application Opportunities

Learning analytics can be applied to address a range of questions and provide insight to a diverse set of learning situations. The following are explanatory examples:

- Predicting Outcome Achievement

Using testing, outcome achievement (from previous courses), survey data, learning styles, demographic data, and LMS activity data to attach a probability (scoring) to student achievement of an individual outcome or set of outcomes prior to entering a program or course. The data could be used to target interventions on areas of greatest challenge.

- Course and Program Dashboards

Develop course dashboards[2] informed by model scoring to provide professors, students and advisors with a targeted view of students' "probability to achieve" selected outcomes prior to the start of a class, or use modeling to identify areas of challenge and strength rolled up for the entire course related to a set of course outcomes to help direct faculty to specific instructional techniques or content remediation.

Develop a pre-term program-level dashboard informed by predictive model scoring that identifies areas of potential challenge for students and helps target interventions across a program.

- Curricular Evaluation

Use modeling techniques to identify supporting relations among learning outcomes in order to define data-evidenced pre-requisites. Use the data to refine curricular sequencing to maximize student success.

Develop a complexity index for each outcome or set of outcomes that allows academic teams to focus resources on areas of greatest challenge for students, and so with specific learning and learner data correlated to the achievement of those outcomes (i.e., with the data in hand to be able to target interventions).

- Prioritize Learning Outcomes

Identify learning outcomes or sets of outcomes the achievement of which most strongly correlates to retention, persistence and graduation. Use the data to identify at-risk students and support targeted interventions across the institution that are informed at the level of student learning and student learner characteristics.

- Set Course and Instructional Policies

Identify the data from the learning management systems that are the strongest drivers of student learning and success. Set curricular and instructional policies aligned to these findings and identify targeted points of intervention (e.g., phone calls, meetings) to better support retention and learning.

- Defining Academic Quality

Develop research initiatives to identify those practices, curricular structure and pedagogies that best support achievement of selected learning outcomes. Use findings to set and insure compliance with quality benchmarks.

4. LIMITATIONS

The main limitation of deploying learning analytics is the reliability and validity of the learning outcomes and learner characteristic data used in the models. Or, more simply, the availability of, and appropriate granularity of outcomes data from which models can be developed. Although accreditors have been focusing on learning outcomes assessment for over 10 years, for most institutions the effort is still in its infancy.

One method to address this limitation is to develop disciplinary consortia similar to those developed to support the tuning process in the Bologna paradigm. The consortia could develop a consistent articulation of a focused set of learning outcomes and identify methods for assessing them. The underlying data set would be comprised of data from across all member institutions. In this way, data sets become large enough to become analytically viable.

[2] Dashboards are being implemented by Untra Corporation's *Academic Evaluation, Feedback and Intervention System* (AEFIS) as part of an *Instructional Decision Support Systems* (IDSS) approach. The IDSS is an interactive computer-based information system which links student characteristics, student performance, instructor characteristics, learning outcomes, and instructional methods to inform faculty decisions on the appropriate educational pedagogy and materials to improve student learning.

A second limitation is the need to communicate modeling results in an efficient and meaningful way to end-users who are able to understand and use the data, and the current ability of most institutions to address the problems inherent in developing appropriate technologies.

J. Campbell has made huge strides in this area at Purdue with the *Signals* tool [4]. In addition, the *Academic Evaluation, Feedback and Intervention System* (AEFIS) solution platform with its instructional decision support system approach offers a strong approach to the communication problem as well as the possibility to support disciplinary consortia.

5. ETHICAL CONSIDERATIONS

There are several ethical considerations to be addressed when deploying learning analytics methods, or academic analytics more broadly. These issues include:

- What data is appropriate to collect about students? What data is inappropriate?
- Who should be able to access the data and view results? Which data should be reported anonymously? Which can be tagged to students for educational purposes?
- What is the impact of showing faculty modeling results? Do any of the data bias faculty instruction and evaluation of students?

The answer to these questions most often depends on the culture of the institution and its IRB, but there are overlapping legal concerns around FERPA, USDE policies, federal mandates and privacy laws. Given the nascent nature of learning analytics as a field of inquiry, an institution can expect to address many similar, but unforeseen, questions as the program moves forward. As such, any institution engaging in these kinds of modeling activities should plan on developing a concurrent, and systematic, conversation that addresses *for the institution* how it approaches the ethical and legal issues that might arise along the entire modeling process (from data collection to data usage).

6. CONCLUSIONS

The convergence of two major efforts in higher education – the application of analytics and the development of learning outcomes assessment – holds a great deal of promise for helping educators and institutions better understand student learning and the factors that support student success, and more efficiently target instructional, curricular and support resources to improving student learning. In addition, the strength of the most successful analytics approaches comes from an institution-wide approach to knowledge management that is focused on what the institution strives to achieve in its mission (the goals and purposes that define the institution). In this sense, learning analytics will provide one piece of educational effectiveness in a larger analytics program supporting institutional effectiveness. The real strength of the approach is in the sharing of insights (not to mention data and model equations) from across all sectors of the institution focused on a coherent, informed and collaborative vision. In addition, thinking more broadly, the analytics approach, and specifically learning analytics, offers an opportunity and structure to develop relationships across institutions of higher education focused on student learning. The opportunities offer rich paths for improving higher education both here and abroad.

7. REFERENCES

[1] L. Baer, J. Leonard, L. Pugliese & P. Lefrere, "Action Analytics: Measuring and Improving Performance that Matters in Higher Education", **Educause Review**, 2008.

[2] L. Briggs, "Data-Driven Decision-Making: Data Pioneers", **Campus Technology**, 20(3), 2-30, 2006.

[3] J. Campbell & D. Oblinger, "Academic Analytics", **Educause Quarterly**, (October), 1-20, 2007.

[4] J. Campbell, P. DeBlois, & D. Oblinger, "Academic Analytics: A New Tool for a New Era", **Educause Review**, (July/August), 40-57, 2007.

[5] P. Goldstein, "Academic Analytics: The Uses of Management Information and Technology in Higher Education", **Educause Center for Applied Research**, 2005.

[6] J. Gonzalez & S.L. DesJardins, "Artificial Neural Networks: A New Approach for Predicting Application Behavior", **Proceedings of the Association for Institutional Research (AIR) 41st Annual Forum** (pp. 1-39). Long Beach, CA: Association for Institutional Research, 2001.

[7] A. Karmon, "'Institutional Organization of Knowledge': The Missing Link in Educational Discourse", **Teachers College Record**, 109(3), 603-634, 2007.

[8] J. Kidwell, K. Vander Linde, & S. Johnson, "Applying Corporate Knowledge Management Practices in Higher Education", **Educause Quarterly**, (4), 28-33, 2000.

[9] C. Lee, "Diagnostic, Predictive and Compositional Modeling with Data Mining in Integrated Learning Environments. **Computers & Education**, 49(3), 562-580, 2007.

[10] P.C. Liezl van Dyk, "Creating Business Intelligence from Course Management Systems", **Campus-Wide Information Systems**, 24(2), 120-133, 2007.

[11] D. Norris, J. Leonard, L. Pugliese, L. Baer, & P. Lefrere, "Framing Action Analytics and Putting Them to Work", **Educause Review**, 43(1), 1-8, 2008.

[12] K. Poulin & A. Freeman, "Developing a Marketing Geographic Segmentation System Using SAS® Software", **SAS Users Group International Proceedings**, Seattle, WA, 2003.

[13] G.E. Steele & K.C. Thurmond, "Academic Advising in a Virtual University", **New Directions for Higher Education**, (146), 85-95, 2004.

[14] G. Swan, "Tools for Data-Driven Decision Making in Teacher Education: Designing a Portal to Conduct Field Observation Inquiry", **Journal of Computing in Teacher Education**, 25(3), 107-113, 2009.

[15] H. Tolley & B. Shulruf, "From Data to Knowledge: The Interaction between Data Management Systems in Educational Institutions and the Delivery of Quality Education", **Computers & Education**, 53(4), 1199-1206, 2009.