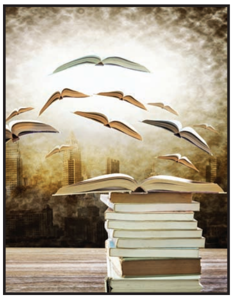

The Next Step for Learning Analytics

Jinan Fiaidhi, *Lakehead University*

Learning analytics is an emergent field of research and practice that aims to use data analysis to inform decisions made on every tier of the educational system.[1] It helps analysts decipher trends and patterns from educational big data—that is, huge sets of student-related data—to further the advancement of personalized, supportive higher education applications.

Learning Analytics Initiatives

Learning analytics is the third wave of developments in instructional technology, which began with the advent of the learning management system (LMS) in 1991. The second wave integrated the LMS into the wider educational enterprise by involving learners on social networks (also known as the Web 2.0 wave). During this third wave, learning analytics as a term has been significantly popularized by the Educause International Conferences on Learning Analytics and Knowledge (LAK),[2] which started in 2011 (https://tekri.athabascau.ca/analytics).

Learning analytics focuses on collecting and analyzing data from a variety of sources to provide information on what works (and what doesn't) with respect to teaching and learning.[3,4] This helps educational institutions improve their quality of learning and overall competitiveness. Consequently, many research communities have developed a variety of promising initiatives, models, and applications to improve learner success. For example, Santa Monica College's Glass Classroom initiative,[5] introduced in December 2012, aims to enhance student and teacher performance by collecting and analyzing large amounts of data. Using real-time feedback of the student's performance, Glass Classroom adjusts the courseware to meet educational objectives.

Another example is at the University of Wisconsin-Madison. Since May 2012, the university has been working to develop a data-driven "early-warning" system that faculty and advisors can use to support student academic success.[6] The system will help identify academically at-risk students, using nontraditional indicators that can be gathered early in a student's career, even at the beginning of a semester. The system aims to intervene early, improve students' academic success, and bolster the campus's retention and graduation rates.

Furthermore, many research groups and societies are providing excellent networks for researchers who are exploring the impact of analytics on teaching, learning, training, and development (see the "Related Learning Analytics Research" sidebar). In particular, these groups and societies are promoting several learning analytics models, which have been

Related Learning Analytics Research

See the following for more information on learning analytics research:

- the Society for Learning Analytics Research (SoLAR)—http://solaresearch.org/,
- the Learning Analytics focus group at the University of Amsterdam—http://learning-analytics.uva.nl,
- the University of Melbourne Learning Analytics research group—https://le.unimelb.edu.au/elearning/larg.html,
- the Social Media Lab at Dalhousie University—http://socialmedialab.ca, and
- Stanford University's Transformative Learning Technologies Lab—https://tltl.stanford.edu/project/multimodal-learning-analytics).

Published by the IEEE Computer Society

Table 1. Notable learning analytics applications.

Application	Purpose and data	Institution & URL
jPoll	This ubiquitous classroom polling and student-response system engages students using a range of interactive teaching situations. Originally developed as a replacement for clicker-type technologies, it produces data based on student opinions of teaching events.	Griffith University http://navigator.nmc.org/project/jpoll
Moodog	This course management system provides a log analysis tool to track students' online learning activities. It provides instructors with insight about how students interact with online course materials, and it lets students compare their own progress with others in the class. Moodog also provides automatic email reminders to students, encouraging them to view available materials that they haven't yet accessed.	University of California, Santa Barbara (An article with more information appears elsewhere.[7])
Equella	Equella provides evidence of appropriate curriculum coverage, student engagement, and equity while students are on clinical placement (which provides a clinical education at an external facility).	University of Wollongong www.pearsonlearningsolutions.com/equella
E2Coach	This computer-based coaching system provides a model for an intervention engine, capable of dealing with actionable information for thousands of students. It's currently being used for a variety of applications in public health—from cessation of smoking to losing weight.	University of Michigan http://sitemaker.umich.edu/ecoach/about_ecoach
Signals	A system that takes course management system (CMS) data (Blackboard data, in this case) into account in predicting whether a student will succeed in a given course.	Purdue University www.itap.purdue.edu/studio/signals
Sherpa	This course recommender system uses a service-oriented architecture to guide students during registration to pick a substitute class if their first choice is full (Phase 1). The second phase provides administrators tools to send "nudges" to students via portal messages, emails, SMS messages, and text-to-speech phone calls. The third phase revamps the student portal with a to-do list, news feed, and calendar, pre-populated with enrollments and important dates.	South Orange County Community College www.socccd.edu/sherpa

developed to identify student risk levels or success factors in real time to increase a student's likelihood for success and improve learning.

Table 1 provides examples of applications that are using such learning analytics models.[7] Higher education institutions have shown increased interest in learning analytics as they face calls for more transparency and greater scrutiny of their student recruitment and retention practices.[8]

Dealing with Unstructured Data

The applications described in Table 1 apply models of structured data, collected from student interactions with the learning system, to answer the following questions:[7]

- When are students ready to move on to the next topic?
- When are students falling behind in a course?
- When is a student at risk for not completing a course?
- What grade is a student likely to get without intervention?
- What is the best next course for a given student?
- Should a student be referred to a counselor for help?

These learning analytics applications use Web analytics, data warehouses, and other data-management tools, measuring Web traffic to assess and improve the effectiveness of the teaching and learning website. (They use clickstream data to record every page, segment, or tag requested by the learner.) Then, they extract knowledge from structured data and store it in databases. You define your learning categories up front, so it's simple to evaluate the status of any given category; track trends in a category; and compare how a category relates to others across time, geography, and so on.

Analyzing unstructured content, however, is different. Unstructured content refers to documents, emails, answers, responses, and other objects (static or dynamic) that are made up of free-flowing text. Analysts, including Gartner and IDC, predict that as much as 80 percent of data is unstructured.[9] This data is a treasure chest of valuable information and insight, but it's notoriously difficult to manage. Gartner defines unstructured data as content that doesn't conform to a specific, predefined data model. It tends to be the human-generated and people-oriented content that doesn't fit neatly into database tables.[10] Text is the classic and most dominant example of unstructured data.

Capturing and analyzing textual data has changed how decisions are made and resources are allocated in businesses, healthcare, government, and many other fields. However, text analysis hasn't yet made the impact on education that it has made in other fields, even though learners interact with instructors, other learners, and with the course materials mainly using text. The proliferation of textual data in education is overwhelming, because such data is being constantly generated via learning blogs, emails, educational websites, learning objects repositories, and so on. Although the amount of textual data is increasing rapidly, learning systems' ability to summarize, understand, and make sense of such data for improved learning remains challenging.

Textual Analytics in Learning

A finer level of understanding of student opinions, answers, and concerns can enable educators, researchers, and university leaders to improve teaching and thus learning.[11] Text analytics typically involves tasks such as text categorization, text clustering, summarization, and concept extraction. Employing learning analytics based on text analytics helps promote automated intervention and reasoning about textual interaction. It also helps archive, filter, search, and classify text.

Thanks to advancements in textual analysis techniques and other machine-learning technologies, systems like the IBM Watson can defeat top human contestants in Jeopardy (see www.ibm.com/smarterplanet/us/en/ibmwatson). According to Seth Grimes,[12] text analytics adds semantics to identify features such as

- named entities (learning objects, learners, and so on);
- pattern-based entities (such as email addresses);
- concepts (abstractions of entities);
- facts and relationships;
- events;
- concrete and abstract attributes; and
- subjectivity in the form of opinions, sentiments, and emotions.

Learning analytics applications enriched with the power of textual analytics might also perform the following smart capabilities and services to empower the learner and the learning system.

Annotate Sensitive Information

In a learning system, you need to communicate clearly and quickly with learners. However, creating screen recordings can be time consuming and frustrating. You need the ability to capture images to get your point across. You also need the ability to add or annotate text and to blur out sensitive information on these images.

Categorize Groups of Documents

This would let applications assign unseen text to one of the predefined categories based on a processing algorithm. There are many kinds of models and strategies for text categorizations, but generally, all algorithms could be divided into the following groups:

- *supervised categorization* uses prior information regarding correct categorization for tested documents;
- *unsupervised categorization* doesn't use any prior or external information, so decisions are made internally based on predefined logic or models; and
- *semisupervised categorization* combines some approaches from the supervised and unsupervised text categorization.

Check for Relevant Live Feeds

Live feeds are different types of feeds, which show all messages that are relevant to a specific user or topic.

Other Text-Based Services

Other potential text-based services include the following:

- classify entities via manual training and automation,
- filter content by metadata and threshold classification,
- generate high-level summaries and detailed reports,
- discover top meta values and related concepts,
- recommend resources,
- analyze feedbacks and opinions,
- build topic models and generate groupings,
- measure and validate results, and

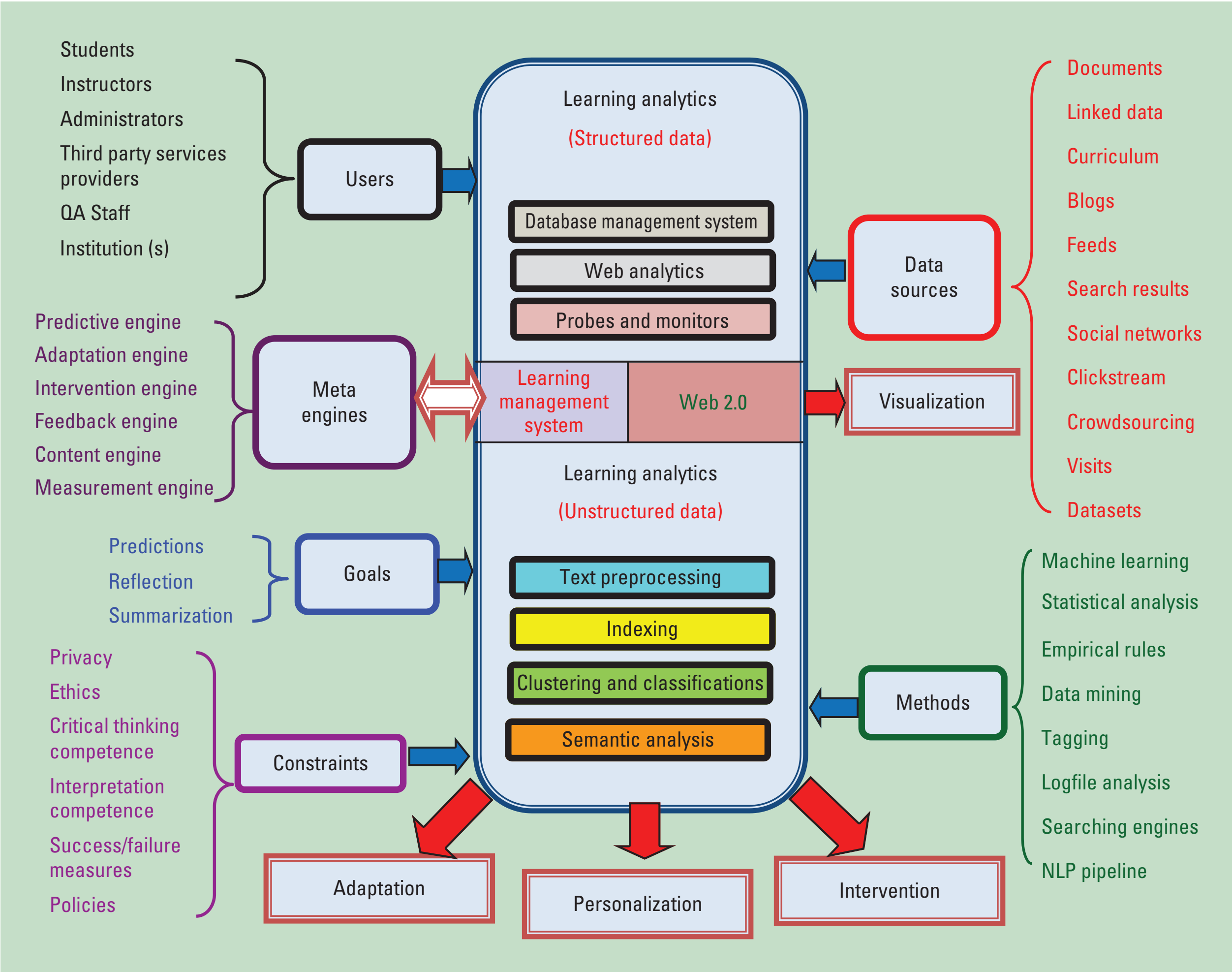

Figure 1. A comprehensive learning analytics architecture.

- identify similar documents and responses (possible plagiarism).

Text analytics applies a variety of natural language processing analysis techniques along with linguistics, statistical, and data-mining techniques to extract concepts and patterns that can be applied to categorize and classify textual documents. It also attempts to transform the unstructured information into data that can be used with more traditional learning analytics techniques. Finally, it helps identify meaning and relationships in large volumes of information. However, there is no single method appropriate for all text analysis tasks.

Learning analytics approaches must take several different perspectives and accommodate different data sources. The ideal vision for learning analytics is to integrate analytics for both structured and unstructured data (mainly of textual nature). Figure 1 illustrates our bird's eye view of a comprehensive learning analytics architecture.

The main feature of this architecture is the component that deals with unstructured textual data (and natural language processing). Adding this to the existing components (that is, to the learning management system, Web 2.0, social networking, and the learning analytics for structured data) will strengthen the power of the "meta engines" to reliably assess students skills and provide students with formative feedback based on their learning processes (their actual cognitive and intellectual development while performing a learning activity). By adding a component that deals with textual data, the meta engines will be able to capture detailed information about teaching and learning. For example, the *predictive engine* will be able to record the progress of certain critical-thinking processes and predict the outcome.

Similarly, the *content engine* will provide more focused information and evaluations using the indexing and content categorizations. The *adaptation engine* will provide more

Editorial Board Changes

Arnold Bragg has retired from *IT Professional*'s Advisory Board. We thank him for his countless contributions to the publication, especially during his years as Editor in Chief.

Thomas Jepsen has retired from *IT Pro*'s Editorial Board and instead will be serving as a member of the Advisory Board. We thank him for his years of service and continued support.

—San Murugesan, Editor in Chief

Call for Issue Proposals

IT Professional invites your theme proposals for its upcoming issues in 2016. If you would like to see a topic of relevance covered or wish to guest edit a theme issue, please send your proposal to San Murugesan, Editor in Chief, at san@computer.org.

customized content delivery for individual students' performance and interests. The *intervention engine* will let instructors and administrators bypass the LMS to directly interact with students. The *feedback engine* will provide evaluations from other meta engines (such as the predictive engine) as feedback for students, instructors, faculty, and administrators. Finally, the *measurement engine* will provide measures such as the level of similarity between the student's solution and that of the instructor.

There are many text analytics tools and APIs that can be used to build the learning analytics component that deals with textual unstructured data (such as LingPipe, RapidMiner, Textanalytics, and OpenText).[13] Incorporating textual analytics into what we have achieved so far in terms of learning analytics research and development will lay the groundwork for redesigning educational institutions to teach 21st century skills.

References

1. *NMC Horizon Report 2013*, The New Media Consortium, 2013; http://net.educause.edu/ir/library/pdf/HR2013.pdf.
2. A. Cooper, "A Brief History of Analytics," *JISC CETIS Analytics Series*, Nov. 2012; http://publications.cetis.ac.uk/wp-content/uploads/2012/12/Analytics-Brief-History-Vol-1-No9.pdf.
3. M. Brown, "Learning Analytics: The Coming Third Wave," *Educause*, Apr. 2011; http://net.educause.edu/ir/library/pdf/ELIB1101.pdf.
4. K.D. Mattingly, M.C. Rice, and Z.L. Berge, "Learning Analytics as a Tool for Closing the Assessment Loop in Higher Education," *Knowledge Management & E-Learning" An Int'l J.*, vol. 4, no. 3, 2012; www.kmel-journal.org/ojs/index.php/online-publication/article/viewArticle/196.
5. L. Johnston, "The Glass Classroom," blog, 1 Dec. 2012; http://glassclassroom.blogspot.ca/2012/12/the-glass-classroom-big-data.html.
6. "Learning Analytics Pilot," University of Wisconsin-Maddison, Spring 2014; www.cio.wisc.edu/learning-analytics-pilot.aspx.
7. Zhang, H. & Almeroth, K. (2010). Moodog: Tracking Student Activity in Online Course Management Systems. Journal of Interactive Learning Research, 21(3), 407–429. Chesapeake, VA: AACE, http://www.editlib.org/p/32307/
8. M. Bienkowski, M. Feng, and B. Means, "Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics," US Department of Education and Office of Educational Technology, Oct. 2012; www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf.
9. "Six Reasons Businesses Need to Pay Attention to Unstructured Data," *IT Business Edge*, 5 Nov. 2013; www.itbusinessedge.com/slideshows/six-reasons-businesses-need-to-pay-attention-to-unstructured-data.html
10. "Gartner Predicts 2013: Business Intelligence and Analytics Need to Scale Up to Support Explosive Growth in Data Sources," Gartner, 2013.
11. *Unstructured Data Analytics: Unlock the Treasures of Your Unstructured Content*, white paper, Filelytics, 2012, https://www.filelytics.com/sites/default/files/downloadable/Unstructured_Content_Whitepaper.pdf.
12. S. Grimes, "An Introduction to Text Analytics: Social, Online and Enterprise," Text and Social Analytics Summit, workshop presentation, 2013; www.slideshare.net/SethGrimes/an-introduction-to-text-analytics-2013-workshop-presentation.
13. "Top 30 Software for Text Analysis Text Mining Text Analytics," *Predictive Analytics Today*, 2013; www.predictiveanalyticstoday.com/top-30-software-for-text-analysis-text-mining-text-analytics.

Jinan Fiaidhi *is a full professor and graduate coordinator with the Department of Computer Science at Lakehead University, Ontario, Canada. She is also an Adjunct Research Professor with the University of Western Ontario. Her research interests are in collaborative learning, learning analytics, and cloud computing. More information about her can be obtained from her website http://flash.lakeheadu.ca/~jfiaidhi. Contact her at jfiaidhi@lakeheadu.ca.*