Annotated Bibliography Series: Learning Analytics

Last Updated 07 Feb 2016

Department of Geography
1984 West Mall
Vancouver, BC V6T 1Z2

Tel 1.604.822.2663
open.geog.ubc.ca

Recommended Citation:

Liu, Elissa. 2015. *Learning Analytics*. Annotated Bibliography Series. UBC Open Geography: Vancouver, British Columbia. http://open.geog.ubc.ca.

Summary

Learning analytics (LA) is driven by “big data” primarily from learning management systems (e.g., Blackboard). LA can be a powerful tool for understanding learning behaviors and trends for institutions when the correct data and analysis are used. Studies suggest that the general framework for LA include defining the problem, collecting data, analyzing data, and summarizing findings; similar to other statistical analysis. To maximize the potential of LA, several authors mentioned the need for collaboration between LA and other related fields (e.g., academic analytics (AA), educational data science (EDS), and educational data mining (EDM)). In particular, the LA is often discussed in conjunction with EDM. EDS focus collection and processing information tracked by learning management systems (e.g., Blackboard). The general process for data mining include collecting data from learning management systems, preprocessing data, applying data mining techniques, and interpreting results. In the EDM community, the most common methods include predictive models, structure discovery, relationship mining, and discovery with models. WEKA and KEEL are two free applications that are widely used for data mining.

The goal of this annotated bibliography series is to summarize studies that have used LA to assess impacts of flexible learning on where and when students learn. From the literature search, it appears that this research question has not been assessed by the LA community or similar fields. Majority of the studies on LA have focused on using analysis to measure the effectiveness of new programs (e.g., online learning vs classroom). Grades have commonly been used as the assessment variable. This has been one of the biggest drawback across all studies because grades does not reflect all areas of learning. Another drawback is the limitation of Blackboard data, which cannot track activities performed outside the platform (e.g, studying from textbook or notes). It is important for future studies to acknowledge or address these problems.

Annotated Bibliography

Greller, W., and H. Drachsler. 2012. Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Journal of Education Technology & Society* 15 (3):42–57.

Cited by 112

This theoretical paper explores the key dimensions of Learning Analytics (LA), critical problem zones, and potential dangers. The main driver for LA is the desire to improve learning processes by taking advantage of the "data economy", which refers to the vast amount of information available through technologies. The authors proposed a framework when considering LA by reviewing scientific proceedings and presentations, brief literature review of abstracts in the field, and looked at current RTD (the authors didn't define RTD) projects. The authors concluded that stakeholders, internal limitations, external constrains, instruments, objectives, and data are the 6 critical dimensions of LA. The authors then provided detailed descriptions of each dimensions and recommends that all 6 dimensions be considered when designing LA.

This paper is useful for our research because it provides a great overview of LA and maps out the different dimensions that can be for our learning analytics study. The authors also highlights potential risks when using LA, which is important for us to acknowledge in our studies. First of all, LA has the potential to be used abusively to increase manipulative control over students, employees and citizens. The authors recommend that LA development take a bottom-up approach by focusing on the leaner's interest. The second danger is that the impact of LA on education is uncertain. The authors agree that there is a lot of potential for LA but they also point out the importance of developing LA with guiding frameworks. Basing decisions on LA can have unwanted consequences; the authors emphasize the importance of the proposed framework. Lastly, the authors point out that "technologies are not pedagogically neutral". The results is highly dependent on the approach chosen for the analysis.

Siemens, G., and R. S. J. Baker. 2012. Learning Analytics and Educational Data Mining : Towards Communication and Collaboration. *Proceedings of the 2nd international conference on learning analytics and knowledge* : 252–254.

Cited by 160

This conference proceeding outlines similarities and differences between the Education Data Mining (EDM) and the Learning Analytics and Knowledge (LAK) communities. The authors first introduces several organizations and literature studies that have recently surfaced from both communities. Generally, these communities are driven by the emergence of "big data" and improving education through data intensive processes. Despite of having the same goals, LAK and EDM have several differences. One of the key differences is the use of automated discovery vs human judgement. LAK leverage human judgement and use automated discovery as a tool whereas EDM is the complete opposite (automated discovery is key, human judgement is tool). It is unclear from the article the meaning of automated discovery. The second biggest difference is that LAK models are more focused on informing and empowering instructors and learners while EDM is more focused on automated adaptation. Lastly EDM is more focused on individual components while LAK takes a holistic approach and attempts to understand the system as a whole. The authors conclude that collaboration of the two communities is necessary in learning how to best exploit "big data" to improve education.

This paper may not be helpful for our research; however, it does help point us to other key research words on learning analytics. It is important that we acknowledge data mining as a field parallel to learning analytics. This paper does provide a list of important literature studies and workshops that have been developed in the interest of using "big data". These references can be used to help us grasp current research topics in this field.

Romero, C., and S. Ventura. 2015. J. A. Larusson, B. White (eds): Learning Analytics: From Research to Practice. *Technology, Knowledge and Learning* 20 (3):357–360.

Book Cited by 2

This book review suggests that "Learning Analytics: From Research to Practice" is a compilation of current research and theories on learning analytics (LA). Learning analytics is related to academic analytics (AA), educational data science (EDS), and educational data mining (EDM). In this book, these fields are all considered LA. The main purpose of the book is to provide institutional stakeholders with an overview of learning analytics by collecting and summarizing current information on LA methods and results. The book is divided into three main sections: 1) LA potential and methods (e.g., abundance of electronic data); 2) Impacts of LA (e.g., case study of how LA informs academic advisors); and 3) LA for teachers and learners (e.g., data-assisted approach to help instructors manually perform pedagogical interventions). The book acknowledges that more case studies; in particular for the last two sections.

For our research, this book is worth reading with focus on section 1 (LA potential and methods). As the reviewers mentioned, the book is a global first that looks at LA. The book also covers various topics in the LA field, from classroom examples to institutional examples. To build on the knowledge obtained from the book, the authors suggest readers to read the EDM Handbook Romero et al. 2010) to get deeper understanding of the area. The book reviewers do mention that editors of the book lack of experience although several literature studies from well-established researchers (e.g., Ryan Baker and Abelardo Pardo) are showcased in the book. Researchers such as George Siemens and Dragan Gasevic are missing. With that said, the book does offer a comprehensive overview of the LA field.

Dyckhoff, A. L., D. Zielke, M. Bültmann, M. A. Chatti, and S. Ulrik. 2015. Design and Implementation of a Learning Analytics Toolkit for Teachers. *Journal of Educational Technology & Society* 15 (3):58–76.

This article gives an overview of the a learning analytics tool created for RWTH Aachen University in Germay. The authors provide a theoretical background, design, implementation, and evaluation details of the tool that has been designed to support the univeristy’s goal of enhancing Virtual Learning Environments (VLE). The tool that has been developed is called eLAT and the main purpose of the tool is to help teachers reflect and improve online teaching methods by visually answering specific questions. Once a question has been estabilished, the learning analytics process begins with data gathering. This step involves addressing privacy issues. The second step is using data mining techniques to process the data. This second step may involve the use of preexisting widgets. The last step is draw conclusions based on the data. Based on literature review and informal conversations, the desired characteristics of eLAT include usability, usefulness, interoperability, extensibility, reusability, real-time operation, and data privacy.

This paper may be as useful for our research because it outlines general steps of learning analytics that we can use a reference. The study also emphaizes some of the problems that need to be addressed through each of the steps (e.g., privacy issues. However, this paper does focus on implementing learning analytics for all purposes; whereas our research question is focused on learning space. The biggest strength of the paper is that the authors gives detailed description of eLAT, which can be used a reference for developing other learning analytic methods. However, the paper does lack versatility because it does focus on the need of one particular univeristy. The authors acknowledge that one of challenges is attempting to make a tool that is versatile and applicable across different learning environments. The authors also specifies that the study does not provide comprehensive field-tests with learning analytics today

Elias, T. 2011. Learning Analytics : Definitions , Processes and Potential. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf.

Cited by 110

This report reviews a number of existing literature in the field on learning analytics to “define learning analytics, its processes, and its potential to advance teaching in online education”. The ideas from the literature are explained individually and then summarized collectively. Although the author only used literature studies in the report, the report provides diverse perspectives and examples of learning analytics being utilized in real life applications. When defining analytics, the author notes that “analytic tools are closely tied to business intelligence, web analytics, academic analytics, educational data mining, and action analytics”. Learning analytics involve improving learning and teaching through the development and integration of new processes and tools. The author then summarizes and compares the analytics frameworks and models presented by different scholars: Knowledge continuum, five steps of analytics, web analytics objectives, collective application model, processes of learning analytics (Table 1). Lastly, the author concluded that learning analytics tools and resources include: computers, theory (analytics-related knowledge and good practice accumulated in other fields), people (knowledge, skills, and abilities of humans to ensure effective operation), and organization (collective intelligence).

It is worth pointing out that social network analysis (SNA) has been gaining popularity. Another visualization tool that is gaining popularity is the social network analysis (SNA). SNA evaluates “network properties such as density, centrality, connectivity, betweenness, and degrees”.

This paper would be useful for our research because the author presents a comprehensive overview of learning analytics. The general ideas within the field has been identified through the author's literature review. In particular, the author targets the three main questions in the field of learning analytics, definition, methods, and potential. The biggest strength of the paper is that the author summarizes a diverse range of ideas from the literature which decreases bias.

Baker, R. S., P. S. Inventado, K. Analytics, N. Manouselis, H. Drachsler, E. Duval, E. Wagner, P. Ice, and D. Clow. 2012. Data changes everything: Delivering on the promise of learning analytics in higher education. *EDUCAUSE Review* 47 (3):61–75.

Cited by 43

Education data mining (EDM) is a sister community of learning analytics. In this book section, the authors review key methods of EDM from literature studies. In the EDM community, the most common methods include predictive models, structure discovery, relationship mining, and discovery with models. The prediction model, similar to dependent variables, infers one aspect from several predictor variables. Structure discovery attempts to find structure within the data (e.g., clustering data). Relationship mining refers to discovering relationships between certain variables (e.g., association rule mining). Discovery with models is not commonly used in data mining but more in the computational science domains.

This book section can be used for our research to understand the different methods in data mining. For our purpose, the most relevant data mining model would be relationship mining. The four types of relationship mining include association rule mining, sequential pattern mining, correlation mining, and casual data mining. Out of methods, correlation mining and casual data mining appear to be most applicable for our research. Correlation mining can be used to find the relationship between different type variables (learning style and student behaviour). Casual data mining is used to predict whether one event is caused by another event. The biggest strength of the paper is that the author describe the purpose of each methods, along with the specific analysis. One of the weakness of the paper is that the lack of descriptions for the method analysis due to the large number of methods that the book section coverd.

1. Romero, C., S. Ventura, and E. García. 2008. Data mining in course management systems: Moodle case study and tutorial. Computers and Education 51 (1):368–384.

Cited by 570

This scientific paper examines the application of data mining in learning management systems. The authors uses Moodle as a case study to describe the process of data mining in e-learning, data processing steps, and applications of data mining techniques. Like most learning management systems, Moodle tracks a wide range of data but requires processing for the data to be comprehendible. The general process for data mining include collecting data, preprocessing data, applying data mining, and interpreting results. Tools that the authors suggest to accomplish these steps include WEKA and KEEL, both free software that have Java language and easily obtainable dataset. The most common data mining techniques in e-learning systems include statistics, visualization, clustering, classification, and association rule mining. The authors outline how these techniques are applied to Moodle. The method is sufficient because the paper provides theoretical and empirical evidence for data mining.

This paper is useful for our research because it focuses on data mining in learning management systems. In terms of flexible learning, learning management systems are likely the most common type of platform for obtaining data. Using this paper as a reference, we can gather specific methods of to gather and analyze data for spatial distributions. The authors also provide several different data mining techniques for different purpose.

Romero, C., and S. Ventura. 2007. Educational data mining: A survey from 1995 to 2005*. Expert Systems with Applications* 33 (1):135–146.

Cited by 816

This paper reviews literature studies on the educational data mining from 1995 to 2006. Drawing on these studies, the authors summarizes the purpose of data mining in education, methods for processing the data, and data mining applications. Distant learning data mining uses server log files (performing data), client log file (interaction between student and system), proxy log file (client and web servers). The types of data collected depends on the types of web-based courses. For courseware using standard HTML, data types can include content, structure, and usage and user profile. For learning content management systems, student activities are monitored and recorded. Raw data taken from servers need to be prepossessed to make sure that the data is relevant and correct. After the data has been preprocessed, the data mining techniques can be applied: "statistics and visualization; clustering, classification and outlier detection; association rule mining and pattern mining; and text mining". Statics and visualization focuses on student usage (e.g., time and frequency of assess), while other techniques focus on patterns and interactions of students (e.g., detecting common behaviours). These patterns guide system improvements and applications.

This is a paper that provides detailed methods for data mining in education and thus will be useful for our research. The methods outlines in the paper on collecting and processing data are particular useful for our research because it covers a wide spectrum of literature studies. Also, this paper provides an extensive list of references on data mining that can be used for our research. However, it is important to be mindful that the references and methods provided by the authors are from literature studies before 2005. The biggest strength of this paper is that it covers a wide range of data mining techniques for different purposes. One of the paper's drawbacks is that the authors do not provide any empirical evidence of these data mining techniques.

Edwards, R., and J. Clarke. 2002. Flexible Learning, Spatiality and Identity. *Studies in Continuing Education* 24 (2):153–165.

Cited by 56

This theoretical paper that highlights spatial concept in relation to flexible learning. Generally, flexibility refers to the process of "teaching and learning can be liberated from constraints of time and place". Drawing on existing literature, this paper analyzes the concept of social space and modularisation. The authors then use their own research to explain the concept of spatially through stories from distance learning students. The authors define space for two types of social space: disciplinary societies and societies of control. Disciplinary societies is where "separate spaces of enclosure is experienced" as students move from one enclosed area to another. On the other hand, societies of control does not have clear transitions between enclosed spaces. Flexible learning relates more to societies of control. Building on the idea of social space, modularising space refers to creating learning that is non-linear and reordering of space-time. In their literature review, one of the studies found that physics is an area that exercise traditional linear pattern while management exercise non-linear, module and flexible. The consequence of the different learning environments yield different skill sets; caused by the learning style rather than the subject. Following the literature search, the authors interview students taking online learning. It was found that education was seen primarily as spaces of enclosure.

This paper offers very interesting insight into the spatial concept of flexible learning which can be used as a reference for our research. In particular, the authors offer clear definitions for different types of space that may exist in learning. One of the paper's weakness is that the author interviewed 2 students are attending distance education to complete their degree. The results would be strengthened with the increase number of participants. It is important to note that the authors mentioned that flexible in terms of space has not been widely studied in relation to learning.

Agudo-Peregrina, Á. F. 2012. Predicting academic performance with learning analytics in virtual learning environments. In International Symposium on Computers in Education, 1–6.
Cited by 13

The main objectives of this scientific article include: 1) developing classifications for types of interactions in Virtual Learning Environments (VLEs); and 2) attempt to identify relationships (if any) between types of interactions and academic performances. Using literature studies, the authors identified three main types of interactions related to VLEs. The first is interactions based on agents involved in the e-learning process (e.g., student-teacher, student-content, student-student interactions). The second is interaction based on the frequency use of activities and features in the VLE (e.g., transmission of content, evaluating students). The last type of interaction is based on participation mode (passive and active). The authors then used empirical study to identify how these interactions relate to academic performance in "real-life". The empirical study took place at the Universidad Politécnica de Madrid with 139 distance learning student. The authors do not explain how different types of interactivity are categorized in relation to the tool. The data was processed using regression and multiple regression methods (independent variable being types of interactions and dependent variable being academic performance). The authors found that that components in agent-based interactivities related most to academic performance. Also, student-teacher, student-student, evaluating students, and active interactions had significant impact on academic performance. It is important to note that the authors emphasized this is an exploratory research and further analysis is needed to confirm the findings.

This paper may not be useful for our research because the authors do not mentioned interaction with space. However, we can learning from the paper’s strength and weakness. The biggest strength of this paper is that it clearly lists out theoretical framework of interactivity in relation to academic performance from the literature and also performed an empirical study. However, this paper only measures academic performance as the dependent variable. Also, the authors do not explain in details how interactivity was measured using the course tool; in particular it is confusing what was classified as each type of interactions. It is important to note that authors mentioned that this paper is an exploratory research and future research is needed to establish standardized method for LA.

Arnold, K. E., and M. D. Pistilli. 2012. Course Signals at Purdue: Using Learning Analytics to Increase Student Success. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (May):267–270.

This conference proceeding introduces the benefits of using Course Signals as a tool for learning analytics. Course Signal (CS) uses data from institutions to provide feedback to students based on predictive models. The goal CS is to identify students that are potentially at risk as indicated by their efforts. CS considers performance to date (grades), effort (e.g.,interaction with Blackboard Vista), previous academic history (e.g., academic preparation, high school GPA), and student characteristics (e.g. age). These components are placed in a predictive student success algorithm (SSA). Based on the results from SSA, a red, yellow, or green signal is displayed for both the students and instructors. The instructors can then intervene and provide assistant to students that require assistance. CS was piloted in 2007 at Purdue University. Generally, academic performance increased 0.59 to 13.8%. The authors also pointed out that the number of students enrolled till graduation increased for those that were enrolled in CS courses. There was also an overall positive perception of the tool; with both students and instructors pointing out that the tool helps motivate students. It is also important to note that the most common issue was the excess of emails, increased dependency, and the lack of best practices with using the tool. The first concern could not be mitigated, the second concern is mitigated based on the higher number of students graduating, and the third concern is addressed by suggesting best practice tips at http://www.itap.purdue.edu/learning/tools/signals.

The biggest strength of the paper is that the authors used more than one component of learning to calculate potential students at risk. However, the actual design of the algorithm is not explained in details in this paper. With the lack of information, it is hard to assess whether the CS tool can effectively identify student at risk. Also, the authors stated that grades increased by 0.59 to 13.8% for students using the CS compared with previous semesters. The increase appears to have a large range, which the authors provide any reasons. The authors also does not assess the amount that grades fluctuate from year-to-year for classes without CS. Despite of certain concerns with the analysis in the paper, the concept of CS can be useful for our research. In particular, there is a lack of literature that actually provides a method for learning analytics. This paper can be a great reference for current tools and methods of learning analytics.

Macfadyen, L. P., and S. Dawson. 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers and Education* 54 (2):588–599.

Cited by 241

This scientific paper analyzes the predictive power of Learning Management System (LMS) in relation to student success. The three main research questions presented in the paper includes: 1) Why use LMS tracking; 2) How to use LMS tracking 3) Can LMS tracking be useful for the development of a student learning community? LMS data are becoming easily obtainable in institutions. The authors propose that the data has been accessed to date but effectively analyzed and interpreted. The paper uses current data from a fully online undergraduate biology course offered at UBC to demonstrate how LMS data can be used. In total, 22 variables from the LMS data were selected (e.g, time spent on an assignment, number of times a student accessed the online tool). The frequency of certain activities and time spent were the two main pieces of information tracked for students that completed the course. Forum discussion data was extracted from one section using SNAPP (Greasemonkey from Firefox browser extension). Scattered plots coupled with simple correlation analysis were employed to identify potential correlations. The variables were plotted against the final grades from students. The authors acknowledge that correlation does not signal causation but it can be an indication. To assess the discussion forum data, the authors created a sociogram that offered a visual interpretation of the amount of discussion. The method used does not take into that marks may be influenced by external variables (outside of LMS); for example previous experience with biology. Also, students that did not participate in the discussion online may have participated in discussions outside of class.

From the analysis, the authors concluded that the number of forum postings, mail message sent, and assessments completed can be used to predict the student's grades. Using these variables to predict students at the risk of failure yielded 70.3% accuracy. As mentioned above regarding the issue with the methodology, the problem with using solely online data to predict grades may yield biases. In particular, some students may not use the LMS tools but use external resources (e.g., textbooks, online tutorials). Despite of certain drawbacks, this paper has lots of important details on the importance of LMS tracking and how to conduct a LMS tracking. Like the authors stated in the paper, there are very few tools and methods for effectively using data for improving learning. This paper offers a great reference for learning analytics design that can be built on by adding certain analysis. The findings from this paper is also very interesting for our flexible learning research as interaction appears to be one of the most important component of learning success.

Long, P., and G. Siemens. 2011. FOGPenetrating. EDUCAUSE Review 46 (5):32–40. http://net.educause.edu/ir/library/pdf/ERM1151.pdf

Cited by 355

This journal article discusses how data can help improve decision-making in higher education. With the vast amount of data available today, the concept of "Big Data" has been used to describe datasets that is "beyond the ability of typical database software tools to capture, store, manage, and analyze". Other fields (e.g., clinical) have taken advantage of this new resource but it has not been applied effectivity to learning. The two types of analytics related to learning includes: learning and academic analytics. Academic analytics refer to data being used for institutional purposes while learning analytics focus on learning processes. For both types of analytics, the authors propose that course-level, educational data-mining, intelligent curriculum, adaptive content, and adaptive learning be considered. The authors does not discuss each of the concepts in much detail nor provide examples.

The authors propose that using "Big Data" in the learning realm will be very useful for making decisions. The authors does not actually suggest any current models that can be used but the several cautions are recommended when considering analytics. The first is assuming that all future conditions can be completely based on the past. The second is simplifying models based on misguided assumptions. Despite the challenges, learning analytics is seen as a great tool for learning how to best allocate the limited amount of resources given to institutions today.

Ali, L., M. Hatala, D. Gasevic, and J. Jovanovic. 2012. A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education* 58 (1):470–489.

Cited by 65

This paper evalutates the improvements made to the first version of LOCO-Analyst tool, which was made to provide feedbacks to students and teachers. To generate feedbacks on the tool, the author evaluated feedbacks before (2006) and after (2009) the the improvements were made. In both studies, questionaires were distributed from University of Belgrade (Serbia), Simon Fraser Univeristy (Canada) and Athabasca University (Canada). In 2006, there were 18 participants while there were 22 participants in 2009 with some from a private Canada-based company developing technology for professional trainning. Out of the 22 participants, 6 were the same as those from the 2006 study. The data were then analysed using quantitative (coding) and qualitative methods (content analysis). Both studies were conducted similarly but participants from the 2009 study watched video clips of the tool's new feature. The method was generally sufficient for their analysis but the participant selection process varied slightlt from 2006 to 2009. Comparing the results from 2006 and 2009, the authors concluded that improved visualization increased the percieved value of the feedbacks from the tool.

This study may not be as useful for our research because it is very specific on the tool that it is analyzing but does not provide much information on the algorithms behind the tool development. It also focuses on the end result of learning analytics rather than the processes. Thus, the authors focused on describing the method for analyzing satification with the tool However, it is important to acknowledge the existence of the LOCO-Analyst tool which has been developed and tested locally.