Introduction to Learning Analytics
Role, benefits, and challenges

Srecko Joksimovic, Vitomir Kovanovic
School of Education and Teaching Innovation Unit
University of South Australia

Srecco.Joksimovic@unisa.edu.au Vitomir.Kovanovic@unisa.edu.au
#s_joksimovic #vkovanovic
Outline

- History and Definition
- Key Dimensions of Learning Analytics
- Data Sources and Methods
- Learning Analytics Research
- Learning Analytics Tools
- Challenges and Way Forward
- Hands-on Session
History & Definition
A Brief History

1920s - "Early Intelligent Tutoring Systems" (Pressey, 1927)
    The term ITS was coined much later (Sleeman & Brown, 1982)
1930s - Psychometric society founded
1950s - Cognitive revolution, SAKI
    1956 - SAKI, the first Adaptive Teaching System
1960s - Computer Assisted Instruction for Teaching and Learning (Skinner 1968)
1970s - CAI/CAT incorporation of AI-techniques
1980s - Learning sciences
1990s - First LMS (FirstClass by SoftArc)
2000s - The Rise of Online Learning
2011 - The EDM Society established
2011 - The First LAK Conference
2012 - SoLAR established (http://solaresearch.org - unisa-2019)
2016 - LAK conference welcomed more than 450 attendees
2017 - The 1st HLA published
Drivers

Pursuit for personalized and adaptive learning
How can we extract value from these big sets of (learning-related) data?
Education is no different

Huge investments in analytics

Ease of access to learner data

Increased adoption of personal technologies
Political (Economic) concerns

Increasing demand for educational institutions to measure, demonstrate, and improve performance.

Stakeholders

Governments,

Educational institutions, and

Teachers/Learners
Learning analytics as a solution

"...is the *measurement, collection, analysis* and *reporting* of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs."
"What do the learning sciences have to do with learning analytics?"

- "Just about everything!"

Paul Kirschner, LAK’16 keynote. Available at
https://www.youtube.com/watch?v=8QjmnOIiIKI&index=2&list=PLOF7tBP24IAf2uyB6SEZ3_foM51rULkSR
Consolidated Model

Theory
- Adoption of theory
- Contribute to the theory

Design
- Interaction & Visualization design
- Learning Design
- Study Design

Data Science
- Collection, measurement, analysis, and reporting

Driving Impact

Increase retention

Proactively drive success

Improve content & course quality

Cost efficient allocation
Key Dimensions
LA Research & Practice

Data & algorithm issues

Connection with learning & teaching

Purpose

Stakeholder management

Scalability & capacity

Ethics & privacy

(2017 SoLAR Institutional Brief)
Data & Algorithms

"the technology infrastructure of high-need classrooms (so that adoption is not limited to affluent schools)"

"consequences of certain data being available (e.g., interventions and their outcomes)"

"the availability of data in a form which is readily manipulated"

"the harmonisation of data coming from different systems on different platforms"

"establishing common metrics and terms and ultimately a common interpretation of the results"

"the effects that growing capabilities of artificial intelligence algorithms have on EdTech"

"that adopted algorithms and metrics allow actionable measurements"

"the potential replication of the norm by using indicators that are chosen by humans with potential biases"

"consequences of certain data being available (e.g., interventions and their outcomes)"
What data we collect?

Student Information System
- Student profile/Demographics
- Lesson planning
- Scheduling
- Enrolment
- Assignments
- Campus attendance
- Academic data
- Trace data
- Simulation data
- Assessment
- Social interaction
- Content interaction
- Intelligent tutoring system
- Educational context data

Learning Environment
- Library attendance
- Lesson Library use (loan reports)
- Library helpdesk

Library Management System

Instruments
- Video
- Audio
- Gesture
- Gaze
- Psychophysiological data
  - EEG
  - fMRI

Multimodal

Social Media
- Twitter
- Facebook
- Blog
- e-book
- Journals
- Applications

Third-party

Survey
- Questionnaire
- Interview
- Focus groups
- Ethnography
A wide spectrum of algorithms

Linear Regression  (Structural) Topic Modeling
Logistic Regression  Bayesian Knowledge Tracing
Decision Tree  Exponential Random Graph Models
SVM  Hidden Markov Models
Naive Bayes
KNN
K-Means
Random Forest
Connection with L&T

"the problem of using tools in such a way as to generate data rather than to solve existing issues in education"

"sound educational theories and learning theories"

"a theory of action reflecting how learning analytics influences teaching and learning that occurs in and out of the classroom"

"how to improve teaching and learning design"

"variations in curricular, instructional, and assessment practices among faculties"

"the need for longitudinal research in which sufficient numbers of students across courses are tracked to understand their progression in studies and into the labour market"

"balancing theory-driven and data-driven work"

"enhancing effective learning"
Purpose

"how learners can benefit from learning analytics"

"what are the big questions or key learning challenges that learning analytics is trying to resolve"

"what has already been done in the applicable field of education research"

"that learners need to be part of the design process"

"the interaction with other teaching innovations"

"what learners want from learning analytics systems"

"what teachers want from learning analytics systems"

"that the learning goals of students may be different from the goals of instructors"
Stakeholder Management

"whose interests are being served by the particular analytics"

"who is excluded from both decision making and implementation and why"

"how to make the case that learning analytics technology is the most effective way limited budget can be spent"

"practical, realistic, sustainable, maintainable and profitable deployment scenarios"

"buy-in from stakeholders at various levels"

"that instructors may feel threatened by learning analytics or see it as an attempt to replace them"

"the cost of implementation and a realistic return-on-investment analysis for a typical educational provider"

"how to communicate the concept and efficacy of learning analytics to various stakeholders"
Scalability & Capacity

"that learning analytics technologies are still in their infancy and thus risky"

"avoiding over-hyping what 'big data' can do, but focusing on the credibility of outcome and claims made about learning analytics"

"the often lengthy timeline required to implement, test, and improve a learning analytics system"

"the various degrees of digital literacy among stakeholders"

"the scalability"

"situations in which people actually have the time and skills to engage in sense making - data for data's sake is not the point"

"which data are needed to improve performance at different levels within the organisation"

"the expertise needed to facilitate learning analytics (e.g., analytics experts, IT professionals, institutional researchers and assessment specialists)"
Ethics & Privacy

"that the ethics committee has been sufficiently informed about the process of data collection and utilisation"

"transparency in collecting, analysing, sharing, and reporting data"

"who can access what data"

"opt-in and opt-out policies"

"the provision of adequate information about data handling policies for all stakeholders"

"the rights of individual learners"

"terms of use, rules and regulations about personal data"

"public perceptions of the nature of student privacy"
Ethics

Use data to benefit learners
Provide accurate and timely data
Ensure accuracy and validity of analyzed results
Offer opportunities to correct data and analysis
Ensure results are comprehensible to end users
Present data/results in a way that supports learning

Gain informed consent
Safeguard individuals' interests and rights
Provide additional safeguards for vulnerable individuals
Publicize mechanisms for complaint and correction of errors
Share insights and findings across digital divides
Comply with the law

Data Protection

Ensure that data collection, usage, and involvement of third parties are transparent

Integrate data from different sources with care

Manage and care for data responsibly

Consider how, and to whom, data will be accessible

Ensure data are held securely

Limit time for which data are held before destruction and for which consent is valid

Clarify ownership of data

Privacy

Anonymize and de-identify individuals

Provide additional safeguards for sensitive data

http://dx.doi.org/10.18608/jla.2016.31.2
Importance vs. Attention received
Data sources and methods of Learning Analytics
LA data sources

Student behavior:
- LMS log data
- Social interactions
- Produced content
- Produced biometric data

Course context:
- Course content
- Course structure

Baseline differences:
- Demographics
- Survey data

Outcomes:
1. Course evaluations
2. Learning outcomes
3. Alumni information
Key aspect of LA: Modeling

Building statistical models of real-world phenomena using learning data

Key goals:

- Predict future
- Improve understanding of:
  - Learners (e.g., self-regulation, motivation, goal-orientation)
  - Course design
  - Instructional interventions
  - Feedback approaches

Some applications:

- Improving retention
- Improving learning outcomes
- Personalization of learning
- Feedback provision
- Course design improvement
- Course materials improvement
How do we model learning?

Literature review of ways in which learning has been modeled in MOOCs

FIGURE 5. The adopted model of the association between context, engagement, and proximal learning outcome, originally developed by Reschly and Christenson (2012), with indicators specific for learning in nonformal, digital educational settings.

Note. Supplementary Figure S1 (in the online version of the journal) depicts the original model, as proposed by Reschly and Christenson.
Popular LA methods

Model building:

1. Supervised methods:
   a. Regression
   b. Classification
   c. Multivariate analysis & latent variable modelling (SEM)

2. Unsupervised methods:
   a. Clustering
   b. Factor analysis
   c. Topic modeling
   d. Process mining

Data analysis:

1. Natural language processing
2. Video analysis
3. Discourse analysis
4. Writing analysis
5. Social network analysis (SNA)
6. Epistemic Network analysis (ENA)

Data use:

1. Dashboard development
2. Feedback provision
3. Understanding learning
Supervised methods

Learn to predict “outcome” characteristic from a set of input characteristics.

Outcome can be:

1) Categorical (at-risk/not at-risk)
2) Numerical (percentage grade)

Purpose

1) Prediction on new data
2) Increase understanding of the domain
Unsupervised methods

No outcome variable.

Train model to find groups of similar data:

- Patterns in student characteristics (profiling, principal component analysis, factor analysis)
- Patterns in text documents (topic modeling, latent semantic analysis)
- Patterns in action sequences (process mining)
LA research examples
Classifying student video reflections

- **Data:** 4,430 utterances coded as either observations, reflections, or motive statements
- **Input:** 503 different linguistic features
- **Output:** type of utterance (observation, reflection, or motive)
- **Result:** classifier with 75% classification accuracy (Cohen’s kappa .51)

Classifying student video reflections

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>MDG¹</th>
<th>Other</th>
<th>Observation</th>
<th>Goal</th>
<th>Motive</th>
</tr>
</thead>
<tbody>
<tr>
<td>liwc.see</td>
<td>Perceptual processes: seeing (e.g., view, saw, seen)</td>
<td>219.94</td>
<td>1.01 (3.35)</td>
<td>1.64 (4.43)</td>
<td>1.62 (4.57)</td>
<td>1.55 (3.29)</td>
</tr>
<tr>
<td>cm.SMCAUSr</td>
<td>Situation model: ratio of casual particles to causal verbs</td>
<td>201.33</td>
<td>0.12 (0.32)</td>
<td>0.13 (0.34)</td>
<td>0.11 (0.30)</td>
<td>0.37 (0.47)</td>
</tr>
<tr>
<td>cm.DRPVAL</td>
<td>Syntactic pattern density: agentless passive voice density, incidence</td>
<td>183.90</td>
<td>2.46 (12.11)</td>
<td>3.02 (17.57)</td>
<td>2.03 (13.36)</td>
<td>4.66 (20.52)</td>
</tr>
<tr>
<td>liwc.focuspast</td>
<td>Time orientation: focus towards past (e.g., ago, did, talked)</td>
<td>151.83</td>
<td>1.46 (3.38)</td>
<td>4.57 (6.34)</td>
<td>0.80 (2.81)</td>
<td>4.80 (6.66)</td>
</tr>
<tr>
<td>cm.WRDNOUN</td>
<td>Word information: noun incidence</td>
<td>120.80</td>
<td>252.12 (212.42)</td>
<td>186.47 (99.92)</td>
<td>208.13 (127.82)</td>
<td>194.34 (95.04)</td>
</tr>
</tbody>
</table>

Predicting learning outcomes from interactions

How much are Moore’s interaction types predictive of student academic success?

- **Data:** 204 course offerings from 29 different courses
- **Input:** 10 features (S-S count, S-S time, S-T count, S-T time, S-C count, S-C time, S-Sy count, S-Sy time, Course name, Course type)
- **Output:** Percent grade
- **Results:**
  - S-Sy time: consistent and positive effect
  - S-C count: negatively effect

Identifying student profiles from trace data

- Six different student profiles
- Differences in their final grades and cognitive presence

Examining interactions in cMOOCs

Skrypnyk, O., Joksimović, S., Kovanović, V., Gašević, D., & Dawson, S. (2015). Roles of course facilitators, learners, and technology in the flow of information of a cMOOC. The International Review of Research in Open and Distributed Learning, 16(3).
Key themes in MOOC discourse

Data: 4,000 news articles about MOOCs

What are the key themes and how they changed over time?

<table>
<thead>
<tr>
<th>#</th>
<th>Topic Label</th>
<th>N</th>
<th>Distinctive Terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EdX</td>
<td>156</td>
<td>edx, mit, agarwal, join, anant</td>
</tr>
<tr>
<td>2</td>
<td>Coursera</td>
<td>139</td>
<td>coursera, koller, partner, stanford, andrew</td>
</tr>
<tr>
<td>3</td>
<td>FutureLearn</td>
<td>125</td>
<td>futurelearn, british, london, launch, chancellor</td>
</tr>
<tr>
<td>4</td>
<td>Udacity</td>
<td>120</td>
<td>stanford, udacity, thrun, intelligence, artificial</td>
</tr>
<tr>
<td>5</td>
<td>MOOCs in Australia</td>
<td>114</td>
<td>international, australian, australium, chancellor,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tertiary</td>
</tr>
<tr>
<td>6</td>
<td>MOOC accreditation</td>
<td>93</td>
<td>credit, college, council, accept, transfer</td>
</tr>
<tr>
<td>7</td>
<td>Business and management MOOCs</td>
<td>91</td>
<td>business, management, dean, executive, manager</td>
</tr>
<tr>
<td>8</td>
<td>Assesment in MOOCs</td>
<td>90</td>
<td>forum, assignment, video, peer, quiz</td>
</tr>
<tr>
<td>9</td>
<td>MOOCs as community college alternative</td>
<td>90</td>
<td>college, community, tuition, adult, throughout</td>
</tr>
<tr>
<td>10</td>
<td>MOOCs and cuts in education fundings</td>
<td>86</td>
<td>funding, government, budget, cut, fund</td>
</tr>
</tbody>
</table>

Key themes in MOOC discourse

Data: 4,000 news articles about MOOCs

What are the key themes and how they changed over time?

LA tool examples
Course signals

UniSA teaching dashboard
Feedback provision: On-task

www.ontasklearning.org
Academic Writing Analytics
utscic.edu.au/tools/awa

Internal structure of learning power through the aligning application and analysis rich data collection and the ethical need for an integrated approach to research.

CONTRAST: Disagreement, tension, options, inconsistency

TREND: A trend, growth, pattern or tendency

This experience had also highlighted existing framework which could be explored and hopefully resolved through...

An early unpublished study in prisons gave cause for concern about the concept of resilience as represented in the original ELLI structure (Deakin Crick and Salway, 2006) with an emerging hypothesis that to become resilient a person needed to utilise all the learning power dimensions over time.

Roll over sentences with Fkeys for a popup reminding you of their meaning.
LA Challenges & Way forward
Some important challenges

Provision of the data is not enough. Instructors need to know how to use displayed data.

Role of study (course) context is hard to capture -> generalizability of study findings is low

Implementation of learning analytics is a complex adventure that requires more than installing software

Better linking with theory

- Pedagogy
- Assessment
- Visualisation
- Psychology
Ways forward

- More replication studies
- Better reporting of studies to enable reproducible research
- Making data publically available to enable model comparison and improvement
- Pre-registration of LA studies
- Development of techniques and methods for educational data analysis to enable more LA research
- Make LA implementations more actionable
Hands-on Session
DEMO

https://demo.intelliboard.net

- Login as a teacher