Ways of Seeing Student Learning & Metacognition with Machine Learning and Learning Models

Mary Loftus
National University of Ireland, Galway
University Road, Galway
Ireland
+353 876954660
mary.loftus@nuigalway.ie

ABSTRACT

The emerging field of learning analytics is showing promise as a light to shine into the dark corners of individual student experience. By making the richness of the learning process more visible, learners and teachers can access deeper insights into their shared experience. Data and models can provide a mirror for self-reflection and metacognition (Koedinger 2009). As Gašević (2015) reminds us, Learning Analytics are about learning. However, too little attention has been paid to the student's role in data-rich learning environments (Kitto 2016).

This research will use probabilistic machine learning techniques in conjunction with other learning model approaches to produce interactive learning models (Millán 2015) that can be integrated in existing learning analytics systems. One such system will be shared with students in a module of a BSc in Computing degree course and a mixed-methods study of their experience conducted – with students having full control of their data.

Categories

K.3.1 [Computers and Education]: Computer Uses in Education; I.2.7 [Artificial Intelligence]: Machine Learning.

Keywords

Learning Analytics, Open Learning Modelling, Machine Learning, Metacognition

1. Background

As Biesta (2009) notes, central to education's purpose is 'the coming into presence of unique individual beings' and to facilitate this, education spaces must 'open up for uniqueness to come into the world'. He talks about a key part of the education process being the 'individuation' or 'subjectification' of each human being – 'the process of becoming a subject'. Think of Maslow's idea of 'self-actualization', Jung's idea of 'individuation'. This emphasis complements the more usual one in education on 'qualification' and 'socialization'. This is the ontological starting point of this research along with Paulo Freire's (1968) emphasis on the student as an agent of praxis in their learning environment.

LAK 17, March 14, 2017, Vancouver, Canada.

DOI: http://dx.doi.org/...

A key part of this individual development is the role of metacognition. As students encounter learning challenges, they can greatly increase their agency and personal development by learning about their own learning process and engaging in metacognitive activities (Koedinger 2009). Metacognition, or the ability to learn how we learn, is an important skill for the life-long learner. Brookfield (1995) defines it as follows:

> *'a self-conscious awareness of how it is they come to know what they know; an awareness of the reasoning, assumptions, evidence and justifications that underlie our beliefs that something is true.'*

The questions driving this research include whether we can encourage students in their metacognitive awareness and endeavors and provide data and a system interface to assist in this process. If learners could 'see' their learning and their metacognitive processes through learning data, would they be able to control and develop these faculties more effectively?

Metacognitive reflection activities need to be aligned and explicitly designed into a course to ensure student engagement and supportive teacher orientation. How can we facilitate this and use data to shine a light into the spaces between learner, course and teacher?

Systems that require very regular requests for input from students to label their activity are unlikely to be used in the long-term. If the system can categorize activity for the student using machine learning techniques – could this increase adoption?

Could such a system have the added benefit of students overseeing and correcting mis-categorized data as a high-level metacognitive activity which provides clear higher-order learning benefits (Aleven 2002)?

2. GOAL OF THE RESEARCH

Develop and apply machine learning and open learning models to support student metacognition in a pre-existing connected learning analytics systems

This research will seek to bring a number of approaches together to build richer more effective student learning models – while still retaining their accessibility and usability for learners.

It will make clear connections to course learning design and ensure alignment of course learning analytics.

Modelling student learning along with machine learning techniques will be used to make learning more visible to students and facilitate metacognitive reflection on their learning process.

This goal will be achieved through the following objectives:

i. Identify appropriate candidate modelling techniques like Open Learning Modelling (OLM), and similar, to allow students to capture and visualise their metacognitive activities
ii. Classify student learning activity data to build an enhanced model of their learning – particularly in relation to metacognition
iii. Identify, develop and implement appropriate machine learning techniques to use in conjunction with other learning models to allow students to see the nature of their metacognitive activity and to track it over time
iv. Map and visualize these patterns and relations to make them more visible to students
v. Enhance and optimise existing machine learning approaches for future work in student learner modelling

2.1 Ethical framework

These objectives will be grounded in a clear ethical framework for the management and governance of the data involved to ensure the protection of student privacy informed by Prinsloo et al (2013) and Daschler et al (2015).

2.2 Critical analysis of learning analytics approaches

A keystone of this research will be a critical analysis of how we 'do' learning analytics and how that impacts learning environments and learners. Perrotta (2016) notes that learning analytics are not objective and neutral. Embedded in them are societal and political power structures and we need to critically reflect on how our analytics-informed interventions impact learners and teachers at those levels. Learners should not be mere data.

3. Research Questions

The primary questions posed in this research are summarized as follows:

i. Can a Learning Analytics system provide an interface for students to engage in metacognitive activities around their own learning, thereby improving individual learning experience and supporting the student's own development goals?
ii. Can we retool an existing learning analytics system using machine learning modelling and classifiers to provide this metacognitive interface to students?
iii. Can such a system help students visualize, track and reflect on their own learning and development goals and help them to improve performance?

4. Current knowledge of the problem

Modelling student learning is an attempt to make visible what goes on in the learning process. It tries to map and model the states and stop-off points as a learner makes their way on their learning journey from start to destination. This process can mirror student activity, provide maps and suggestions for students to guide their metacognitive and other learning processes.

There are many modelling approaches (Chrysafiadi & Virvou 2013) but not all are accessible to the student themselves and not all lend themselves to effective reasoning approaches. Bayesian Networks are simple constructs in some ways but have been proven to be powerful in student modelling (Millán 2010). Bayesian Networks were first described by Judea Pearl (1985) as 'directed acyclic graphs' which can be used to model causal dependencies between variables. The paper was initially presented at a cognitive science conference which may hint at the original motivation behind Bayesian networks. The applications of Bayesian networks are many and varied - they are a widely applicable approach to reasoning using probability.

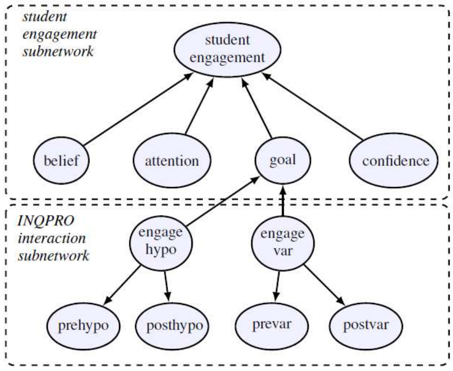

Fig 1. Bayesian Network Model of Student Engagement Ting et al (2013)

In education settings, Bayesian networks have seen a particular application in what was termed Bayesian Knowledge Tracing by Corbett & Anderson (1994). This was an approach where a learner progressing through a given learning path was modelled and this model used to predict whether the learner would successfully negotiate the next step in the learning path.

A particular advantage of Bayesian Networks is that they are white-box algorithms and can be relatively easily understood by humans and represented visually to inform – rather than obfuscate (Xing 2015).

Hidden Markov Models are a class of Dynamic Bayesian network and they allow for the possibility of progression along a path of state transitions. They have been applied to learning contexts in order to model student progress through a learning path of defined states. A HMM is a state transition model showing unobservable states and their corresponding observable indicators (Rabiner 1986)

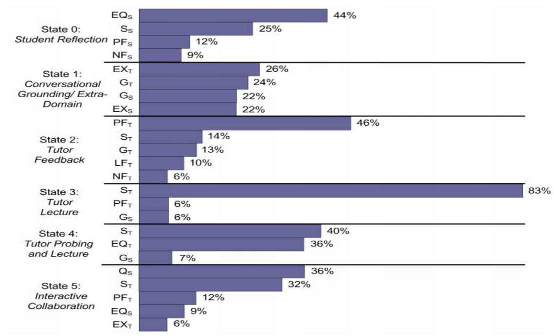

Fig 2 - A HMM state transition table which shows progression of student learning using categorization of learner dialogue Boyer et al (2009)

Open Learning Modelling (OLM) techniques are different to many learning analytics approaches in that they are so firmly rooted in the student experience and are designed with the express purpose of having students interact with them to make their learning more visible (Bull & Kay 2010). Interactive Open Learning Modelling (IOLM) (Dimitrova 2016, Bull et al. 2016) builds an interactive element into this approach. Segedy et al (2013) apply related thinking to propose an assessment approach they call 'model-driven assessments'- applied to Open Ended Learning Environments (OELEs).

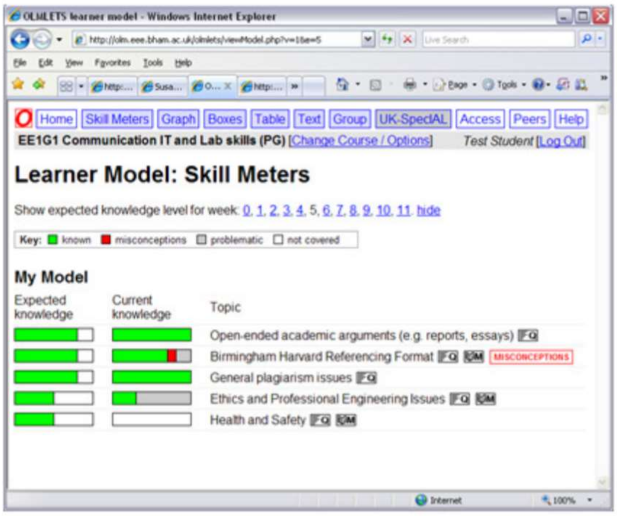

Fig 3. Open Learning Modelling artifacts Bull & Kay (2013)

5. How is this solution different, new or better than existing approaches?

5.1.1 Grounded in the Student perspective

This research seeks to add to the work being done from the student point-of-view and will help develop approaches that are more student-centric. It will incorporate work from Shum & Crick (2012) on learning dispositions and feature a mixed methods approach grounded directly in a third-level classroom environment. It is situated within a machine learning research group but grounded within learning theory and learning design.

5.1.2 Students as owners of their learning data

Students in this study will be in full control of their own data – and it is expected this will improve their data literacy skills and agency. It will also place them in the learning analytics driving seat and they will be empowered to select the modelling and learning analytics parameters that they believe most useful for them. The emphasis then is not on big-data external to the student – but data that is inherently subjective to each students view of her/his internal and external world. Of course, teachers could coordinate students to select the same parameters for the purposes of particular learning scenarios – but the default is to leave this decision to students.

5.1.3 Links learning analytics to learning design

This research emphasizes the link between learning analytics and learning design echoing Mor et al (2015) and Lockyer et al (2013) and draws on Bakharia's 'Learning Analytics For Learning Design Conceptual Framework' (2016).

5.1.4 Emphasis on Connected & Networked learning

In the era of social media, students are not limited to learning with peers in their physical campus spaces. They can learn with others who may be physically located in other countries and even other continents. Personal Learning Networks (PLNs) are a growing feature of academic development for both teachers and learners in modern academic settings and Connected Learning enables new learning spaces to open up (Cronin 2016).

To potentially address this aspect of modern, connected learning, one of the candidate learning analytics platforms that this research could build on is the open source Connected Learning Analytics Toolkit (CLA Toolkit) (Kitto et al 2016). This is an Experience API (xAPI) based data-store with a web front-end of visualization tools for learning analytics. It already has one machine learning powered element – a Naïve Bayes classifier of student activity.

xAPI otherwise known as the Tin Can API is a triple-based data standard which is designed for logging student data in VLE/LMS – and is a successor to the SCORM data standard. This opens up exciting possibilities around modelling for Connected and Networked Learning and Social Learning Analytics.

5.1.5 Machine Learning with an emphasis on modelling and visibility as well as prediction

Rather than use machine learning techniques solely for prediction, the emphasis with this research is to use these techniques to see into learning data and make processes and patterns visible – where possible - as seen and selected by students themselves. Prediction will have a role but it will be tailored to the human nature of learning.

5.1.6 Data literacy capacity building for students

A side-effect of working with students through this research process will be presenting opportunities for students themselves to build their awareness of the ever-increasing role that data plays in their lives and encouraging them to be pro-active agents in this process – rather than unwitting participants.

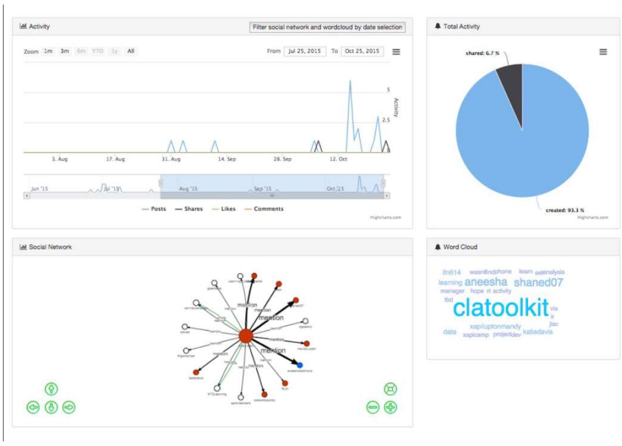

Fig 4 – The Connected Learning Analytics Toolkit (Kitto, 2016)

6. Methodology

Before the research analyses are conducted, there will be a number of system and modelling development tasks to complete:

- Develop and adapt student modelling and machine learning techniques to provide views on metacognitive activities. Initially, these will be relatively simple models with further development over time informed by user feedback.
- Review existing connected learning analytics systems and select one with a good basic analytics feature-set
- Integrate the new/adapted student modelling functionality into the existing learning analytics toolkit
- Introduce students to the system, conduct workshops on metacognition and facilitate students in using the system
- When students get used to the system and show some adeptness in its use, build some assessment requirements into the course where they will use data to reflect on their metacognition activity and reflect on their own learning

The methodological approach for the research is still under consideration but the following are some of the requirements:

- Quantify student metacognitive awareness and ability
- Quantify improvements in that awareness and ability when using the learning analytics metacognition toolset
- Where quantitative analyses reveal patterns of note, subsequent qualitative research will further evaluate students' metacognitive experience and awareness - and improvements when using the system.

Candidate approaches for this qualitative work currently include:

- Action research involving students and teachers
- Grounded theory (Glaser 2009) involving interviews with students and teachers to identify a theoretical basis for students using the system to develop metacognitive awareness.

7. Current Status of the Work

An initial literature review has been completed – which encompasses a wide review of learning analytics approaches and the significant work done over recent years.

Tooling selection and model selection work is ongoing. This includes a review of a wide range of Open Learning Modelling techniques and a variety of machine learning approaches which might be used in concert.

Also, a number of visualization and interface tools are being evaluated for suitability in presenting this work directly to students and allowing them to interact productively with models and analyses. This learning analytics base system will be adapted to incorporate the new modelling and classification functionality.

Research methodologies are currently being considered and will encompass both quantitative and qualitative approaches.

A University Research Ethics Committee application is being prepared to ensure there is a well-defined ethical foundation to the work and clear responsibilities around the use of student data during the lifetime of this research.

8. Acknowledgments

The authors acknowledge the support of Ireland's Higher Education Authority through the IT Investment Fund and ComputerDISC in NUI Galway.

9. References

Aleven, V. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. Cognitive Science, 26(2), 147–179. https://doi.org/10.1016/S0364-0213(02)00061-7

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., … Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics (pp. 329–338). ACM Press. https://doi.org/10.1145/2883851.2883944

Biesta, G. (2009). Good education in an age of measurement: on the need to reconnect with the question of purpose in education. Educational Assessment, Evaluation and Accountability (Formerly: Journal of Personnel Evaluation in Education), 21(1), 33–46. https://doi.org/10.1007/s11092-008-9064-9

Boyer, K. E., Ha, E., Wallis, M. D., Phillips, R., Vouk, M. A., & Lester, J. C. (2009). Discovering Tutorial Dialogue Strategies with Hidden Markov Models. In AIED (pp. 141–148).

Brookfield, S. (1995). Adult learning: An overview. International encyclopedia of education, 10, 375-380.

Bull, S., & Kay, J. (2010). Open Learner Models. In R. Nkambou, J. Bourdeau, & R. Mizoguchi (Eds.), Advances in Intelligent Tutoring Systems (Vol. 308, pp. 301–322). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from http://link.springer.com/10.1007/978-3-642-14363-2_15

Bull, S., Ginon, B., Boscolo, C., & Johnson, M. (2016). Introduction of learning visualisations and metacognitive support in a persuadable open learner model. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (pp. 30–39). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2883853

Chrysafiadi, K., & Virvou, M. (2013). Student modeling approaches: A literature review for the last decade. Expert Systems with Applications, 40(11), 4715–4729. https://doi.org/10.1016/j.eswa.2013.02.007

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. User Modeling and User-Adapted Interaction, 4(4), 253–278.

Cronin, C. (2016). Open, networked and connected learning: Bridging the formal/informal learning divide in higher education. Proceedings of the 10th International Conference on Networked Learning 2016. Retrieved from http://networkedlearningconference.org.uk/abstracts/pdf/S3_Paper2.pdf

Dimitrova, V., & Brna, P. (2016). From Interactive Open Learner Modelling to Intelligent Mentoring: STyLE-OLM and Beyond. International Journal of Artificial Intelligence in Education, 26(1), 332–349. https://doi.org/10.1007/s40593-015-0087-3

Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., Manderveld, J. (2015). Ethical and privacy issues in the application of learning analytics (pp. 390–391). ACM Press. https://doi.org/10.1145/2723576.2723642

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: five approaches (pp. 23–33). Presented at the International Conference on Learning Analytics & Knowledge, Canada. Retrieved from http://oro.open.ac.uk/32910/1/LAK2012-RF-SBS.pdf

Ferguson, R., & Shum, S. B. (2012). Social learning analytics: five approaches (p. 23). ACM Press. http://doi.org/10.1145/2330601.2330616

Freire, P. (1968). Pedagogy of the oppressed (revised). New York: Continuum.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. TechTrends, 59(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Glaser, B. G., & Strauss, A. L. (2009). The Discovery of Grounded Theory: Strategies for Qualitative Research. Transaction Publishers.

Kitto, K., Gašević, D., Siemens, G., Lynch, G., Bakharia, A., Lupton, M., Dawson, S. (2016). The connected learning analytics toolkit (pp. 548–549). ACM Press. https://doi.org/10.1145/2883851.2883881

Koedinger, K. R., Aleven, V., Roll, I., & Baker, R. (2009). In vivo experiments on whether supporting metacognition in intelligent tutoring systems yields robust learning. Handbook of Metacognition in Education, 897–964.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. American Behavioral Scientist, 2764213479367. https://doi.org/10.1177/0002764213479367

Millán, E., Jiménez, G., Belmonte, M.-V., & Pérez-de-la-Cruz, J.-L. (2015). Learning Bayesian Networks for Student Modeling. In C. Conati, N. Heffernan, A. Mitrovic, & M. F. Verdejo (Eds.), Artificial Intelligence in Education (Vol. 9112, pp. 718–721). Cham: Springer International Publishing. Retrieved from http://link.springer.com/10.1007/978-3-319-19773-9_100

Millán, E., Loboda, T., & Pérez-de-la-Cruz, J. L. (2010). Bayesian networks for student model engineering. Computers & Education, 55(4), 1663–1683. https://doi.org/10.1016/j.compedu.2010.07.010

Mor, Y., Ferguson, R., & Wasson, B. (2015). Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action. British Journal of Educational Technology, 46(2), 221–229.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. University of California (Los Angeles). Computer Science Department.

Perrotta, C., & Williamson, B. (2016). The social life of Learning Analytics: cluster analysis and the 'performance' of algorithmic education. Learning, Media and Technology, 0(0), 1–14. https://doi.org/10.1080/17439884.2016.1182927

Rabiner, L., & Juang, B. (1986). An introduction to hidden Markov models. IEEE ASSP Magazine, 3(1), 4–16. https://doi.org/10.1109/MASSP.1986.1165342

Segedy, J. R., Loretz, K. M., & Biswas, G. (2013). Model-driven Assessment of Learners in Open-ended Learning Environments. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 200–204). New York, NY, USA: ACM. https://doi.org/10.1145/2460296.2460336

Shum, S. B., & Crick, R. D. (2012). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics (p. 92). ACM Press. http://doi.org/10.1145/2330601.2330629

Slade, S., & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas. American Behavioural Scientist, 57(10), 1510–1529. http://doi.org/10.1177/0002764213479366

Ting, C.-Y., Cheah, W.-N., & Ho, C. C. (2013). Student Engagement Modeling Using Bayesian Networks (pp. 2939–2944). IEEE. https://doi.org/10.1109/SMC.2013.501

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. Computers in Human Behavior, 47, 168–181. https://doi.org/10.1016/j.chb.2014.09.034