Challenges of a Holistic Learning Analytics Project

Dirk Ifenthaler

Applied Teaching and Learning Research
University of Potsdam
Karl-Liebknecht-Str. 24-25
14476 Potsdam, Germany
dirk@ifenthaler.info

Abstract: Recently, interest in collecting and mining large sets of educational data on student background and performance to conduct research on learning and instruction has developed as an area generally referred to as learning analytics. Higher education leaders are recognising the value of learning analytics for improving not only learning and teaching but also the entire educational arena. However, theoretical concepts and empirical evidence need to be generated within the fast evolving field of learning analytics. In this paper, we introduce a holistic learning analytics framework. Based on this framework, student, learning, and curriculum profiles have been developed which include relevant static and dynamic parameters for facilitating the learning analytics framework. Based on the theoretical model, an empirical study was conducted to empirically validate the parameters included in the student profile. The paper concludes with practical implications and issues for future research.

1 Introduction

With the increased availability of vast administrative, systems, academic, and personal information within educational settings, educational data management, analysis, and interpretation is becoming complex. Several concepts closely linked to processing such educational information are educational data mining, academic analytics, and learning analytics [LS11]. Educational data mining (EDM) refers to the process of extracting useful information out of a large collection of complex educational datasets [Ro11]. Academic analytics (AA) is the identification of meaningful patterns in educational data in order to inform academic issues (e.g., retention, success rates) and produce actionable strategies (e.g., budgeting, human resources) [CDO10]. Learning analytics (LA) emphasises insights and responses to real-time learning processes based on educational information from digital learning environments, administrative systems and social platforms. Such dynamic educational information is used for real-time interpretation, modelling, prediction, and optimization of learning processes, learning environments, and educational decision-making [If14].

However, learners' needs and their predispositions are multidimensional and quickly change over time [As92]. Numerous approaches for understanding these complex patterns of learning and predicting their future developments for automating instruction have been challenged repeatedly in the past [IPS10]. Applications of LA presupposes a seamless and system-inherent analysis of learner's progression in order to continuously adapt the learning environment [Az05]. Additionally, LA provides the pedagogical and technological background for producing real-time interventions at all times during the learning process.

The purpose of this study was to introduce a holistic model for learning analytics and investigate the validity of student profile parameters influencing the learner's study unit outcome that can guide further empirical studies and instructional design efforts.

2 Learning analytics

2.1 Holistic framework

Figure 1 illustrates a holistic learning analytics framework, linking various types of educational information in a meaningful way.

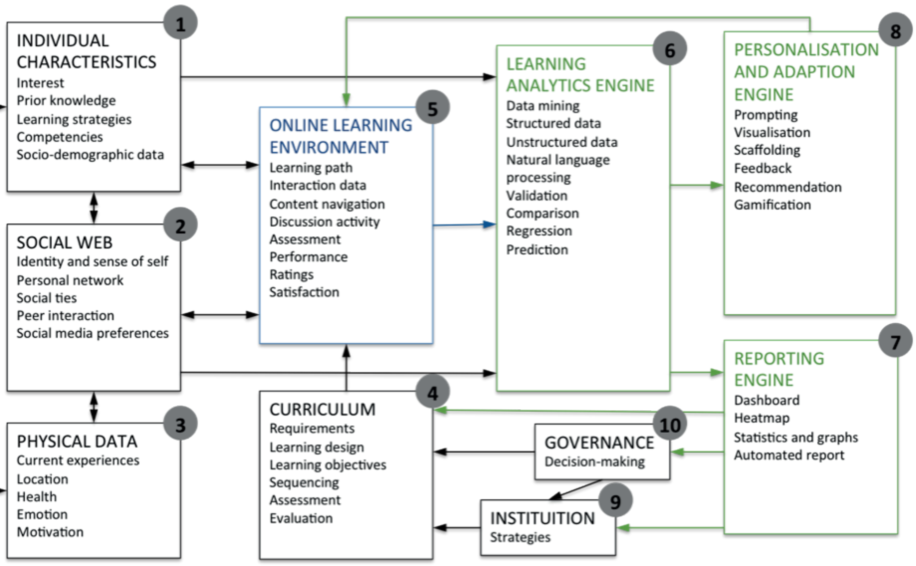

Figure 1: Holistic learning analytics framework

Information about the learners' individual characteristics (1) include socio-demographic information, personal preferences and interests, responses to standardised inventories (e.g., learning strategies, achievement motivation, personality), skills and competencies (e.g., computer literacy), prior knowledge and academic performance, as well as institutional transcript data (e.g., pass rates, enrolment, dropout, special needs). Information from the social web (2) are preferences of social media tools (e.g., Twitter, Facebook, LinkedIn) and social network activities (e.g., linked resources, friendships, peer groups, web identity). Physical data (3) includes information about the learner's location, sensor data (e.g., movement), affective states (e.g., motivation, emotion), and current conditions (e.g., health, stress, commitments). Rich information is available from learners' activities in the online learning environment (5) (i.e., learning management system, personal learning environment, learning blog). These mostly numeric data refer to logging on/off, viewing and/or posting discussions, navigation patterns, learning paths, content retrieval (i.e., learner produced data trails), results on assessment tasks, responses to ratings and surveys. More importantly, rich semantic and context specific information are available from discussion forums as well as from complex learning tasks (e.g., written essays, Wikis, Blogs). Additionally, interactions of facilitators with students and the online learning environment (OLE) are tracked. Closely linked to the information available from the OLE is the curriculum information (4) which includes meta data of the OLE. This data reflects the learning design (e.g., sequencing of materials, tasks and assessments), learning objectives as well as expected learning outcomes. Formative and summative evaluation data are directly linked to specific curricula, facilitators and/or student cohorts.

The learning analytics engine (6) is based on pedagogical theories and methodological and mathematical approaches. Rich information from various sources (i.e., structured and unstructured data) is processed using specific algorithms (e.g., Bayesian networks, neural networks, natural language processing, survival analysis, hierarchical linear modelling) which are closely linked to the underpinnings of applied pedagogical theories. The results of the data mining process are validated before further analysis are computed for real-time comparisons, identification of patterns as well as predictive modelling.

The reporting engine (7) uses the results of the learning analytics engine and automatically produces useful information in forms of interactive dashboards, heatmaps, statistics and graphs as well as automated reports. These automated reports are utilised for specific stakeholders such as the governance level (8; e.g., for cross-institutional comparisons), single institutions (9; e.g., for internal comparisons, optimisation of sequence of operations), governance (10), as well as curriculum level (4) including insights and reports for learning designers and facilitators for analysing instructional processes and students' pathways. The personalisation and adaption engine (8) feeds back the results of the learning analytics engine to the OLE. Interactive elements include simple learning prompts and recommendations (e.g., reminder of deadlines, links to further study, social interaction), rich visualisations (e.g., learning paths) as well as informative scaffolds for specific learning activities and assessment tasks. An optimal implementation of such a holistic learning analytics framework uses a real-time data collection, processing and feedback mechanism. It also allows all stakeholders to personalise the learning analytics process in order to meet their individual requirements.

2.2 Profiles

Based on the holistic learning analytics framework, three profiles have been identified: (1) student profile, (2) learning profile, and (3) curriculum profile (see Figure 2).

The student profile includes static and dynamic parameters. Static parameters include gender, age, education level and history, work experience, current employment status, etc. Dynamic parameters include interest, motivation, response to reactive inventories (e.g., learning strategies, achievement motivation, emotions), computer and social media skills, enrolments, drop outs, pass-fail rate, average performance rate, etc. The learning profile includes variables reflecting the current performance within the learning environment (e.g., learning management system). Dynamic parameters include time specific information such as time spent on learning environment, time per session, time on task, time on assessment. Other parameters of the learning profile include login frequency, task completion rate, assessment activity, assessment outcome, learning material activity (upload/download), discussion activity, support access, ratings of learning material, assessment, support, effort, etc. The curriculum profile includes parameters reflecting the expected and required performance defined by the learning designer and course creator. Static parameters include course information such as facilitator, title, level of study, and prerequisites. Individual learning outcomes are defined including information about knowledge type (e.g., content, procedural, causal, meta cognitive), sequencing of materials and assessments, as well as required and expected learning activities. The available data from all profiles are analyzed using pre-defined analytic models allowing summative, real-time, and predictive comparisons. The results of the comparisons are used for specifically designed interventions which are returned to the corresponding profiles. The automated interventions include reports, dashboards, prompts, and scaffolds. Additionally, stakeholders receive customized messages for following up with critical incidents (e.g., students at risk, assessments not passed, satisfaction not acceptable, etc.).

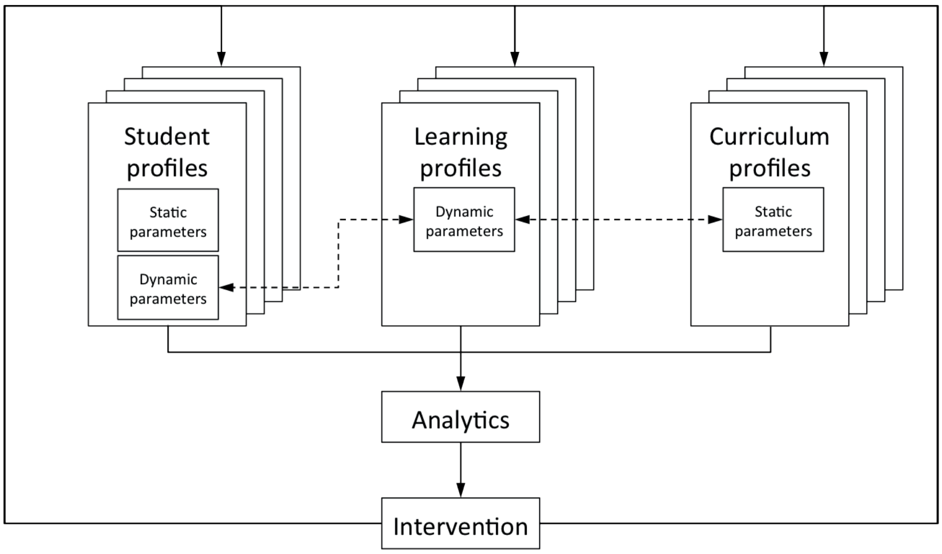

Figure 2: Student, learning, and curriculum profiles

2.3 The present study

Not all educational data is relevant and equivalent. Therefore, the theoretical and empirical validity of underlying models and its reliable analyses is critical for generating useful summative, real-time, and predictive insights from LA. The initial investigation of the present study intended to use large existing datasets to validate the holistic learning analytics framework and its derived profiles. In particular, the present study focused on the student profile which leads to the first research question: Which student related variables do influence study unit outcomes in online higher education?

Well accepted studies identified variables directly linked to the above described student profile, such as age, gender, education background, work hours, etc., as critical factors for study success [Ti97]. Accordingly, we assume that student profile parameters which are associated with study unit outcomes can be identified which at least explain 40% of variance using a linear regression model and a support vector regression (Hypothesis 1).

Further, a major benefit expected from learning analytics is detecting students at risk for early interventions as well as for facilitating their on-going learning progression [AG12]. This leads to our second research question: Do student related variables contain sufficient information for predicting study unit outcomes with acceptable accuracy? We assume that a student profile model can be identified which predicts study unit outcomes (defined as pass/fail of a unit) with at least 80% accuracy (Hypothesis 2).

3 Method

3.1 Participants

The sample consisted of N = 146,001 students (54073 male; 91,928 female) enrolled in 1,509 unique study units with a major online higher education provider in Australia. Their mean age was 33.06 years (SD = 9.90). 85% of the participants reported that they completed secondary school. 5% of the students reported having a disability. 94% studied at undergraduate levels and 6% at postgraduate levels.

3.2 Data models

Table 1 shows the data models which were implemented for the student profile. The first model includes parameters referring to the students' background and demographic data. Parameters of student background include first language spoken, country of residence, and citizenship. Parameters of demographic data include gender, age, socio-economic status, and disability. The second model includes the parameters of model 1 plus parameters referring to the student's and family's historical education background such as completion of secondary school, highest education level of the student, and highest education level of the parents. The third model includes the parameters of model 2 plus parameters referring to information related to the study unit. Parameters of study unit include undergraduate and postgraduate level study, study area, enrolment mode, delivery method, and study support utilized. The forth model includes the parameters of model 3 plus student's historical education record with the institution such as time since last unit, study load, dropped and swapped study units. The fifth model includes the parameters of model 4 plus the historical study performance of the student, i.e., average grade. The sixth and final model includes the most important parameters identified from previous models.

Table 1: Model descriptions for student profile

Model 1	Student background & demographic data
Model 2	Student background & demographic data Student's and parent's historical education background
Model 3	Student background & demographic data Student's and parent's historical education background Study unit related information
Model 4	Student background & demographic data Student's and parent's historical education background Study unit related information Historical education record with institution
Model 5	Student background & demographic data Student's and parent's historical education background Study unit related information Historical education record with institution Average historical grade within institution
Model 6	Most important parameters identified from previous models

3.3 Analytic strategy

As a major analytic strategy, regression models were created with the student's study unit success as the dependent variable. Linear regression algorithm was initially used to identify significant parameters within the model. Subsequently, to identify non-linear and complex parameter relationships, Support Vector Regression (SVR) [CV95] was used as secondary optimization algorithm to create the final regression model. The data models were then used for predicting student success in the study units. The study unit outcome was defined as binary outcome: either passing or failing a subject. To predict study unit outcomes we used Support Vector Machines (SVM) [Dr97], a binary classification technique based on supervised machine learning in the broad area of artificial intelligence. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

4 Results

4.1 Student profile model

For each model of the student profile (see Table 1), we conducted a linear regression analysis as well as a support vector regression analysis to determine whether the student profile parameters were significant predictors for study unit outcome.

Table 2 shows the results of the linear regression analysis yielding an adjusted R^2 of .057 and for the support vector regression R^2–SVR = .059. The final model 6 explained adjusted R^2 = .435 of variance (linear regression) and R^2–SVR = .451 (support vector regression) for predicting the study unit outcome.

Table 2: Student profile model performance comparison

	R^2	Adjusted R^2	R^2-SVR	Predictive accuracy (SVM)
Model 1	0.05671	0.05665***	0.0592	58.63%
Model 2	0.1281	0.128***	0.1296	63.80%
Model 3	0.1867	0.1865***	0.1918	67.50%
Model 4	0.3613	0.3611***	0.424	79.52%
Model 5	0.4408	0.4457***	0.4378	79.69%
Model 6	0.4435	0.4354***	0.4505	80.03%

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

To sum up, the findings suggest that parameters included in the final student profile model 6 explain more than 40% of variance. Table 3 shows the parameter description and relative importance for the final student profile model. Clearly, the most important parameter associated with study unit outcome is the student's average historical grade within the institution.

4.2 Accuracy for study unit outcome

The data was divided in the two classes depending on the study unit outcome with category labels "pass" and "fail". Pass-to-fail ratio of the training data was 1.23. The student data included both first time students which do not have a historical record with the institution. For each model, sufficient amount of total data available from the two categories were randomly chosen to train the SVM classifier. Each classifier was trained with parameters from the models shown in Table 1. We used 5-fold cross validation to analyze prediction performance of the classifier models.

Table 3: Optimal parameter description for student profile model 6

Parameter	Relative importance
Average Historical Grade	43.7%
Historical cumulative fails	18.0%
Highest level of prior education	9.6%
Historical cumulative distinctions	4.2%
Method of payment	4.1%
Historical cumulative higher distinctions	3.7%
Institution of the unit	3.0%
Historical cumulative credit passes	2.7%
Student tutor support	1.7%
Field of study	1.6%
Degree level of the unit taken	1.3%
Gender	0.8%
Number of concurrent subject	0.8%
Source of student enrolment	0.8%
Age	0.7%
If Secondary school completed	0.7%
Historical cumulative Withdrawals	0.7%
Socioeconomic status	0.6%
Time since last time studying	0.4%
Native Australian status	0.4%
If English is the first language	0.4%

The prediction accuracy of each SVM classifier model is reported in Table 2. Classifier with parameters from model 1 predicted correct study unit outcome being 58.63% of the time. The classifier created with parameters from model 6 which were determined as the most significant for the regression models predicted the correct study unit outcome with an accuracy of 80.03%. The training data contained students with no historical record with the institution. Since the historical record is a significant factor, large portion of the misclassifications were first time students. A classifier identical to model 6 and trained with data from students that have taken more than one study unit showed a final prediction accuracy of study unit outcome being 85.16%.

To sum up, the findings suggest that parameters included in the final student profile model 6 account for 80% accuracy for predicting study unit outcome.

5 Discussion

At the moment, well-established empirical evidence within the emerging field of LA is lacking. As new frameworks for LA are developed across the higher education sector, we argue that they need to be empirically tested with regard to their reliability and validity before they may be implemented at larger scale.

The present study tested two hypotheses. First, we were able to identify a model for the student profile which explained 45% of variance. The most significant parameters of this model included the student's historical grade within the institution, historical cumulative fails as well as their highest level of prior education. Parameters such as age, gender, first language, or socioeconomic status only had a minor significant influence on the model. Second, the SVM analysis showed that our final model was able to predict the study unit outcome with an accuracy of 80%. Accordingly, the current model of the student profile may be accepted as empirically validated. A cross-validation using data from other institutions will be a next logical step. This will be a first step towards the empirically validated benchmarks for LA.

Further, several analytic strategies have been applied for LA projects, such as linear regression analysis, logistic regression analysis, survival analysis, Bayesian network analysis, neural network analysis, etc. Our analytic strategy used a classic linear regression analysis as well as a non-linear analysis strategy based on support vector machine algorithms, i.e., support vector regression (SVR). The non-linear modeling strategy proofed to provide better results than the linear analytic strategy with all models we tested. Accordingly, further analysis of models from the learning and curriculum profiles will provide proof of concepts towards the application of our proposed analytic strategy for LA.

5.1 Implications

The benefits of the holistic learning analytics framework can be associated with four levels of stakeholders: mega-level (governance), macro-level (institution), meso-level (curriculum, teacher/tutor), and micro-level (learner, OLE). An essential prerequisite for LA benefits, however, is the real-time access, analysis and modelling of relevant educational information. The mega-level facilitates cross-institutional analytics by incorporating data from all levels of the learning analytics framework. Such rich datasets enable the identification and validation of patterns within and across institutions and therefore provide valuable insights for informing educational policymaking.

The macro-level enables institution-wide analytics for better understanding learner cohorts for optimising associated processes and allocating critical resources for reducing dropout and increasing retention as well as success rates. The meso-level supports the curriculum and learning design as well as provides detailed insights about learning processes for course facilitators (i.e., teachers, tutors). This information can be used for improving the overall quality of courses (e.g., sequencing of learning processes, alignment with higher level outcomes) as well as enhancing learning materials (e.g., their alignment to anticipated learning outcomes and associated assessments). The micro-level analytics supports the learner through recommendations and help functions implemented in the OLE. Learners benefit from such personalised and adaptive scaffolds and are expected to be more successful in reaching the learning outcomes. Another critical component for improving the benefits of LA is information from the physical environment (e.g., learner's current emotional state) which is not directly linked with the educational data. Accordingly, data may be collected within the OLE through reactive prompts and linked with the available educational information.

Table 4: Matrix of learning analytics benefits

	Perspective		
Stakeholder	Summative	Real-time	Predictive
Governance	• Apply cross-institutional comparisons • Develop benchmarks • Inform policy making • Inform quality assurance processes	• Increase productivity • Apply rapid response to critical incidents • Analyze performance	• Model impact of organizational decision-making • Plan for change management
Institution	• Analyze processes • Optimize resource allocation • Meet institutional standards • Compare units across programs and faculties	• Monitor processes • Evaluate resources • Track enrolments • Analyze churn	• Forecast processes • Project attrition • Model retention rates • Identify gaps
Learning design	• Analyze pedagogical models • Measure impact of interventions • Increase quality of curriculum	• Compare learning designs • Evaluate learning materials • Adjust difficulty levels • Provide resources required by learners	• Identify learning preferences • Plan for future interventions • Model difficulty levels • Model pathways
Facilitator	• Compare learners, cohorts and courses • Analyze teaching practices • Increase quality of teaching	• Monitor learning progression • Create meaningful interventions • Increase interaction • Modify content to meet cohorts' needs	• Identify learners at risk • Forecast learning progression • Plan interventions • Model success rates
Learner	• Understand learning habits • Compare learning paths • Analyze learning outcomes • Track progress towards goals	• Receive automated interventions and scaffolds • Take assessments including just-in-time feedback	• Optimize learning paths • Adapt to recommendations • Increase engagement • Increase success rates

Table 4 provides a matrix outlining the benefits of LA for stakeholders including three perspectives: (1) summative, (2) real-time and (3) predictive. The summative perspective provides detailed insights after completion of a learning phase (e.g., study period, semester, final degree), often compared against previously defined reference points or benchmarks. The real-time perspective uses ongoing information for improving processes through direct interventions. The predictive perspective is applied for forecasting the probability of outcomes in order to plan for future strategies and actions.

Yet, the above described matrix requires empirical research within different educational settings. Accordingly, Table 4 may serve as a collection of open empirical questions for future research in learning analytics.

5.2 Limitations and future work

As with all empirical research, there are limitations to the present study, which need to be addressed. First, while our sample size was large enough to achieve statistically significant results, the explained variance for some of our regression models was moderate. This indicates that besides the tested variables other variables may have influenced the outcomes which were not tested in the reported study. As we only tested the student profile of the holistic learning analytics framework, future research will include rich data from the learning and curriculum profiles which will add substantially towards the explained variance of our models. A preliminary analysis of the learning profile yielding an adjusted R^2 of .889 (support vector regression) and predictive accuracy of 95% (support vector machine). Second, the predictions are only valid for individual study unit outcomes, however, do not reflect higher education outcomes in general. Accordingly, further studies will be needed to cross-validate the initial results of this study.

Future work includes empirical validation of all profiles and a full implementation of the holistic learning analytics framework as a dynamic plug-in for learning management systems. A further iteration of the holistic learning analytics framework will also include a natural language processing (NLP) approach which will be utilized for analyzing discussion forums and providing recommendations of social interaction [Da11] and rich semantic feedback in near real-time [IP11].

Currently, the LA framework is being further developed which will allow us to test the personalisation and adaption engine and therefore move beyond the curent stage of validation. These results will advance the development of the holisitic framework and provide important insights on the effects of the LA project on learning and teaching.

5.3 Challenges and concerns

Serious concerns and challenges are associated with the application of LA: (1) Not all educational data is relevant and equivalent [MD12]. Therefore, the validity of data and its analyses is critical for generating useful summative, real-time and predictive insights. This generates a new interdisciplinary research area for cognitive psychology, educational technology, learning design, psychometrics, data management, artificial intelligence, web development and statistics. The challenges are to investigate the complex processes within LA frameworks and to understand their immediate and long-term effects on learning and teaching processes. (2) Ethical issues are associated with the use of educational data for LA [SP14]. That implies how personal data is collected and stored as well as how it is analysed and presented to different stakeholders. Hence, procedures regulating access and usage of educational data need to come into operation before LA frameworks are implemented. This will also include transparency of applied algorithms and weighting of educational data for predictive modelling. Storing and processing anonymised personal data is only a small step towards a more comprehensive educational data governance structure for LA. (3) Limited access to educational data generates disadvantages for involved stakeholders. For example, invalid forecasts may lead to inefficient decisions and unforseen problems. A misalignment of prior knowledge, learning pathways and learning outcomes could increase churn and the late identification of learners at risk may create dropouts. A definition of threshold standards for LA could prevent vast gaps between educational institutions and provide equal opportunities for all stakeholders. (4) The preparation of stakeholders for applying insights from LA in a meaningful way is vital. Professional development for stakeholders ensures that issues are identified and benefits are transformed into meaningful action. Hence, the increased application of LA requires a new generation of experts with unique interdisciplinary competences. This will also require new infrastructures for administration and research in order to accelerate the understanding of LA. (5) Information from distributed networks and unstructured data cannot be directly linked to educational data collected within an institution's environment. An aggregation of such data and uncontrolled relations to existing educational data increases the chance of critical biases as well as invalid analysis, predictions and decisions. The challenge is to develop mechanisms to filter biased information and warn stakeholders accordingly. (6) An optimal sequence of data collection and economic response times (seconds, minutes, hours, days, weeks) of LA have yet to be determined. This includes the minimum requirements for making valid predictions and creating meaningful interventions. Missing data is a critical challenge for future LA algorithms. (7) Besides the analysis of numerical data (e.g., click streams), a qualitative analysis of semantic rich data (e.g., content of discussion forums, responses to open-ended assessments) enables a better understanding of learners' knowledge and needs. An obvious requirement is the development of automated natural language processing (NLP) capabilities. The major challenge besides the development of real-time NLP is the validation of such algorithms and the link to quantitative educational data.

6 Conclusion

More educational data does not always make better educational data [GD12]. Hence, LA has its obvious limitations and data collected from various educational sources can have multiple meanings. Empirically validating learning analytics frameworks and corresponding profiles such as presented in this paper may provide evidence for the implementation of intelligent systems which have the capabilities to facilitate learning of individual students, improve instructional practice of teachers and improve the quality of higher education offerings of individual intuitions and across the sector.

7 Acknowledgements

We gratefully acknowledge the Open Universities Australia for making this work possible by resourcing the Holistic Learning Analytics project. We would like to express our deep appreciation to Chat Widanapathirana for his superb work as Learning Analytics Specialist.

References

[AG12] Aflalo, E.; E. Gabay: An information system for dropout prevention. Education and Information Technologies, 2012. 17(2): p. 233-250.

[As92] Ashby, F.G., ed.: Multidimensional models of perception and cognition. 1992, Lawrence Erlbaum Associates: Hillsdale, NJ.

[Az05] Azevedo, R. et al.: Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. Instructional Science, 2005. 33(5-6): p. 381-412.

[CDO10] Campbell, J.P.; P.B. DeBlois; D. Oblinger: Academic analytics: a new tool for a new era. Educause Review, 2010. 42(4): p. 40-57.

[CV95] Cortes, C.; V. Vapnik: Support-vector networks. Machine Learning, 1995. 20(3): p. 273-297.

[Da11] Dawson, S. et al.: Using Social Network Metrics to Assess the Effectiveness of Broad-Based Admission Practices. Australasian Journal of Educational Technology, 2011. 27(1): p. 16-27.

[Dr97] Drucker, H. et al.: Support vector regression machines, in Advances in neural information processing systems 9, M.C. Mozer, M.I. Jordan, and T. Petsche, Editors. 1997, MIT Press: Cambridge, MA. p. 155-161.

[GD12] Greller, W; H. Drachsler: Translating learning into numbers: A generic framework for learning analytics. Educational Technology & Society, 2012. 15(3): p. 42-57.

[If14] Ifenthaler, D. et al.: Challenges for education in a connected world: Digital learning, data rich environments, and computer-based assessment - Introduction to the inaugural special issue of technology, knowledge and learning. Technology, Knowledge and Learning, 2014. 19(1-2): p. 121-126.

[IP11] Ifenthaler, D.; P. Pirnay-Dummer: States and processes of learning communities. Engaging students in meaningful reflection and elaboration, in Social media tools and platforms in learning environments: Present and future, B. White, I. King, and P. Tsang, Editors. 2011, Springer: New York. p. 81-94.

[IPS10] Ifenthaler, D., P. Pirnay-Dummer and N.M. Seel, eds. Computer-based diagnostics and systematic analysis of knowledge. 2010, Springer: New York.

[LS11] Long, P.D.; G. Siemens: Penetrating the fog: Analytics in learning and education. EDUCAUSE Review, 2011. 46(5): p. 31-40.

[MD12] Macfadyen, L.; S. Dawson: Numbers are not enough. Why e-Learning analytics failed to inform an institutional strategic plan. Educational Technology & Society, 2012. 15(3): p. 149-163.

[Ro11] Romero, C. et al., eds.: Handbook of educational data mining. 2011, CRC Press: Boca Raton, FL.

[SP14] Slade, S.; P. Prinsloo: Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 2014.

[Ti97] Tinto, V.: Limits of theory and practice i student attrition. The Journal of Higher Education, 1982. 53(6): p. 687-700.