Towards Massively Personal Education through Performance Evaluation Analytics

Fernando Koch and Chaitanya Rao

***Abstract*—We are looking for new ways to instrument classrooms towards the ideals of adaptive learning environments and *massively personal* education. In this work, we are focusing on a framework to provide affordable alternatives for data collection, information services, and Analytic models about the classroom environment. This development advances the state-of-the-art by introducing an alternative to analyse the education performance based on differentiated multi-dimensional data and large data sets of relevant data and information. We designed a proof-of-concept experiment to detect variations of level of attentiveness, activity and task performance. In our initial tests, we could successfully collect and analyse relevant signals in a classroom environment and relate them to education performance.**

***Index Terms*—Massive education, learning analytics, big data, ambient intelligence.**

I. Introduction

In a small but growing number of schools, students watch lectures using digital education material to drive their study. In this context, there is a emerging demand for technologies to collect multi-domain learning data, novel models of Learning Analytics, and adaptive learning technologies, related to the concept "Big Data goes to School" mentioned in [1], whatever form of big data it may encompass.

We are researching and developing a framework that combines affordable instrumentation, information services and Learning Analytic models to compose innovative tools to collect in-classroom data. This development exploits tools from Big Data technology [2] applying off-the-shelf solutions to collect data about in-classroom behaviour and analyse this *corpus* using novel models of Learning Analytics aiming to understand the student behaviour and factors that may have influenced the education performance. We call this environment the Smart Classroom [3]. We notice that SAMSUNG has a proposal for Smart School as presented in [4].

During the development, we considered both regular education data for performance evaluation (*e.g.* results from exams, queries, others) and also *human signals* [5] from the education environment in order to understand the behaviour of individuals and groups. For example, we analyse data like the time spent on a page, clicks, zooming gestures, taps, etc. Based on this information we can deduce individual behaviours such as level of interest and attention, focused thought, and group engagement [6]. We aim to create the intelligence to:

1) Classify and understand individual and group education performance based on analysis of results (exams) and signals during manipulation of education material.
2) Recommend actions to improve the education performance, as for example adjusting the digital education content, personalisation, and others.

The paper is structured as follows. Section II describes the motivation and related work. Section III introduces our proposal and describes a prototype implementation. The paper concludes with an analysis of the results and discussion about future work in Section IV.

II. Motivation and Related Work

A. Education Performance

Education Performance provides information about the achievement of outcomes on key aspects of the education system. This information is an essential part of accountability that enables the organisation to assess and report on education progress of individuals and groups and analyse what is working well and what needs improvement.

Metrics of Education Performance can be derived from: 1) analysis of results from exams and surveys, 2) intentional feedback provided through surveys on the education content, and 3) non-intentional feedback, through the analysis of gestures whilst using the content.

For instance, [7] presents diverse approaches for the measurement, collection, analysis and reporting of data about learners and their contexts. Moreover, the work in [8][9] provides a broad view of the use of Analytics in education environments, and [10] introduces Social Learning Analytics by combining learning analytics and social networks.

Examples of analysis resulting from existing tools for evaluating education performance are [11]:

1) Measuring and reporting student achievement.
2) Spotting outliers for early intervention.
3) Predicting potential based on previous scores.
4) Preventing drop-out.
5) Identifying and developing key attributes of good teachers.
6) Analysing standardised testing to even out performance.

B. Performance Evaluation Analytics

Learning Analytics and Big Data technologies [2] provide the capability to collect and analyse large amounts of data

Manuscript received September 5, 2013; revised November 24.

F. Koch is with the SAMSUNG Research Institute Brazil. By the time he worked on this paper, he was with IBM Research and visiting the faculty of Architecture Building and Planning, The University of Melbourne, Australia (e-mail: fernando.koch@samsung.com).

C. Rao is with IBM Research, in Melbourne, Australia (e-mail: chairao@au1.ibm.com).

and derive new methods for Performance Evaluation Analytics. In this context, we are exploiting methods to apply this technology (*i.e.* off-the-shelf products associated to Big Data) to understand the large-scale education environment and create massively personal education material by tailoring the content to student's abilities and knowledge. In a typical adaptive learning system, students read text and complete exercises e.g. using content through a tablet computer. This data is fed into a database where Analytical software combines the information to calculate education performance and predict how well the students perform in class. Examples of analytical models for regular education data are presented below (for the sake of comparison with the more elaborated models to be presented in the next section).

Results of Exams. Let us consider that exams E are learning objects in material M. Student *s* utilises *M* producing a set of results from exams $RS(s, E)$. There is a function $d(RS(s, E), M)$ returning a value in the range [0, 1] representing the distance of provided answers from correct answers (which are known for E). Then:

- $p(s, M) = 1 - d(RS(s, E), M)$ calculates the student's performance based on the exams' responses.
- $pc(C, M) = (\sum_{s \in C} p(s, M)) / n$ calculates the absolute classroom's performance based on the received responses, where $C = s_1, \ldots, s_n$ is a classroom with students.

Results of Surveys. Let us consider that surveys $q \in Q$ are learning objects in material M, $Q \subseteq M$. Student *s* utilises *Q* producing a set of surveys' results $RQ(s, Q)$. There is a function $q(RQ(s, E), M)$ returning a value in the range [0, 1] to calculate the quality of the material based on the students' responses. Then

$$qc(C, M) = ((\sum_{s \in C} q(RQ(s, E), M)) / n)$$

computes the absolute classroom's evaluation of the quality of material based on the surveys' responses.

Motivated by these concepts we seek an integrated solution to a) collect differentiated multi-dimensional data from the classroom environment, b) curate data sets of relevant data and information, and c) develop new models of Learning Analytics to measure education performance. Next, we outline our proposal to promote adaptive learning environment and improve the quality of education material.

III. Proposal

A. Solution Overview

As mentioned in [12], the concept of *adaptive learning material* refers to a computerised-learning interface that constantly customise material for the students". To make this concept work, developers must map the connections among every concept in a piece of learning material. That is, every time the student utilises the education material, data on students' performance and the signals generated during this interaction must be captured and sent to a server. Then the algorithm takes over classifying, understanding and comparing the performance of individual students and groups. Patterns should emerge that will allow to identify what works well for that type of student.

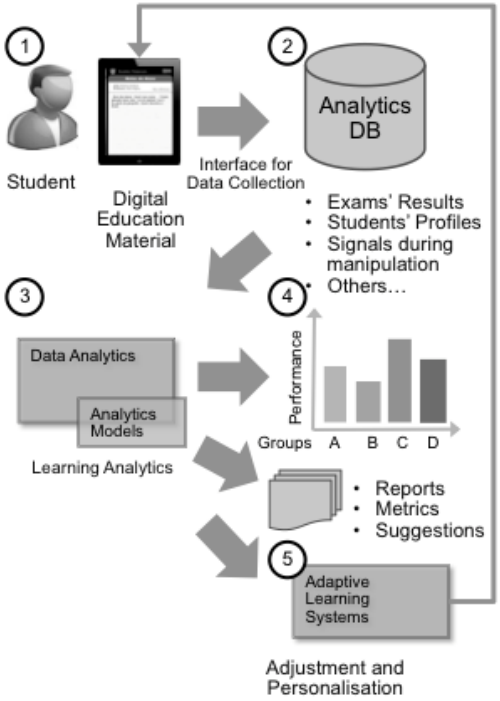

Fig. 1. Solution overview.

Fig. 1 depicts the elements and interactions involved in this process, as detailed below:

1) Students utilise the digital education material and generate data by answering to tests and surveys, and manipulating the material (*e.g.* page, clicks, zooming gestures, taps, etc.).
2) Analytics Database stores this data along with profiles of the students containing, *e.g.* demographics, analysis from historic data, preferences, and others.
3) Data Analytics implement models of Learning Analytics as the "measurement, collection, analysis and reporting of data about learners and their contexts" [8]. These are both statistical models, focusing in trend analysis and correlation, and behavioural models, focusing on segmentation and individual (and group) behaviour analysis. They are directed to understand and optimise learning and the education environment.
4) The Analytic Models yield reports, metrics, and recommendations that can be directly applied by educators and administrators to understand and act upon the learning environment.
5) The Adaptive Learning Systems receives information form the analytic models and implement the logic to *e.g.* recommend adaptations in the education environment, and adjustments in the digital education material and teaching strategy.

The key element in this process is the methods to classify and understand the data being captured. Our current development focuses on a combination of calculation models in form of mathematical and statistical functions. For instance, let us assume that H(s,A) represents a set of signals generated by student *s* while engaging in activity *A*, where *h*

= <ts, tp, par>, *ts* is the timestamp, *tp* is the type of the signal (*e.g.* turn page, click on link, zoom element, tap on element, selection option, etc) and *par* are descriptive parameters. $C = s_1,\ldots,s_n$ is a group of students (a classroom).

Level of activity while resolving a task. Given a task to read elements and respond to a set of tests $E \subseteq M$, student *s* generates the signals H(s, E). There is a function levAct(H(s, E)) returning a value in the range [0, 1] representing the level of activity ac_s of a student *s* realising the exercise *E* based *e.g.* on the time lapse between actions and the sequence of action types. Then:

$$\text{avgAct(C, E)} = ((\sum_{s \in C} levAct(H(s,E)))/n)$$

calculates the average level of activity *gac* of a group of students *C*.

- Act(ac_s) classifies level of activity where T_{ac} is a threshold (*e.g.* T_{ac}=0.2 in our experiments), as:

Slow activity, if $ac_s \leq gac \times (1\text{-}T_{ac})$

Normal activity, if $gac \times (1+T_{ac}) > ac_S > gac \times (1\text{-}T_{ac})$

High activity, if $ac_s \geq gac \times (1+T_{ac})$

Level of attention while resolving a task. Similar to above, given a task to read elements and respond a set of tests $E \subseteq M$, student *s* generates the signals H(s, E). There is a function levAct(H(s, E)) returning a value in the range [0, 1] representing the level of attention at_s whilst *s* is resolving the task. It is calculated based *e.g.* on the time between actions, time switching in and out of the application (*i.e.* distractions by other applications), and others. Then:

$$\text{avgAtt(C, E)} = ((\sum_{s \in C} levAtt(H(s,E)))/n)$$

calculates the average level of attentiveness *gat* of a group of students C.

- Att(c, I) classifies level of activity, where T_{at} is a threshold (*e.g.* T_{at}=0.5 in our experiments), as:

Inattentive, if $ats \leq gat \times (1\text{-}T_{at})$

Attentive, if $gat \times (1+T_{at}) > at_s > gat \times (1\text{-}T_{at})$

Highly attentive, if $at_s \geq gat \times (1+T_{at})$

Performance resolving a task. Given a task to read elements and respond to a set of tests $E \subseteq M$ there is a set OE(E) of optimal sequence of events to execute the instruction. Student *s* executes the actions A(s, E) resolving the task. There is a function distEvt(OE(E),A(s, E)) returning a value in the range [0, 1] representing the proximity pf_s between the two sets. Then:

$$\text{avgDist(C, E)} = ((\sum_{s \in C} distEvt(OE(E), A(s,E)))/n)$$

calculates the average performance *gpf* of a group of students C.

- *perf*(*c*, *I*) classifies performance, where T_{pf} is a threshold (*e.g.* T_{pf} =0.2 in our experiments), as:

Low performance, if $pf_s \leq gpf \times (1\text{-}T_{pf})$

Normal performance, if $gpf \times (1+T_{pf}) > p_s > gpf \times (1\text{-}T_{pf})$

High performance, if $pf_s \geq gpf \times (1+T_{pf})$.

Based on these (alternative) Learning Analytic models, we can implement experiments to collect data in order to classify individual and group behaviour as demonstrated below.

B. Experiment

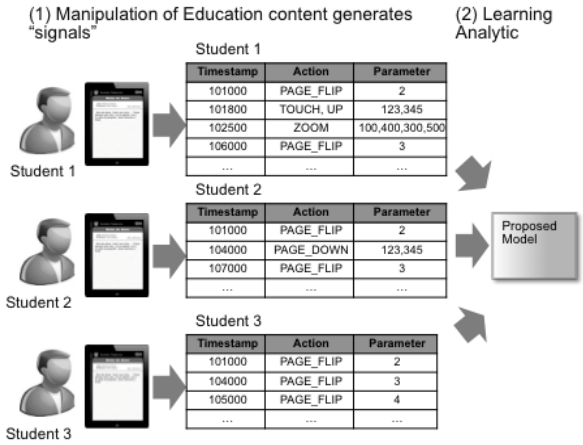

Fig. 2. Experimentation scenario.

We envisioned a test scenario to correlate level of activity, level of attentiveness, and resulting performance, as depicted in Fig. 2. Our hypothesis is that depending on the students' behaviour it is possible to classify their profiles as inattentive, hyperactive, or normal behaviour and then compare the results from observations based on surveys conduced with these students. We executed the experiment in a control environment, using a content player based on the open-source Cool Reader and instrumented the application to capture the signals such as page change (PAGE_FLIP), touch, zoom gesture, going down and up into the page (PAGE_DOWN, PAGE_UP), and other on-screen gestures. The signals are stored in log files as tuples <ts, tp, pr> where *ts* is the timestamp, *tp* is the type, and *pr* are description parameters. At the end of the class, the applications upload the log files to a server where they are stored and indexed. The experiments were executed in 4 classes with 10-15 students each, running the application loaded with digital education material.

During the class, the students 1) manipulate the education content generating the logs of signals. We are specifically interested in signals related to execution of tasks ordered by the teacher during the class – as we know when the teacher ordered the action, based on the timestamp, we can correlate to the actions in response to that order. Next, we collect the log files at the end of the class and 2) process using the Learning Analytic models aforementioned. Finally, we tabulated the results for analysis presented in Table I.

TABLE I: EXAMPLE OF TEST RESULTS

	Low Activity	Normal Activity	High Activity
Inattentive	Task Low / Exam Low	Task Normal / Exam Low	Task Low / Exam Low
Attentive		Task Normal / Exam Normal	Task High / Exam Normal
High Attentive		Task High / Exam High	Task High / Exam High

From the results, we notice that students classified as *inattentive* whilst utilising the education material attain lower performance for both task execution and exams. We concluded that the students with *low activity* in this group present the characteristics of inattentiveness, whilst the ones with *high activity* tend towards hyperactive – however, we grant that this observation is not conclusive and may not be always the case. During the survey, the students with a known inattentiveness condition reported difficulty to: pay attention to the class, understand what is being discussed in a given moment, and keep attention whilst the tablet computing offers other distractions (*i.e.* applications other than the content player). On the other hand, the students with a known hyperactive condition reported that they needed to feel in control of the tablet computing and player application, so they spent considerable amount of time playing with the configurations. Conversely, students classified as *attentive* and *highly attentive* attain best performance in both metrics. We cannot conclude that high activity in manipulating the education content reflects the identified conditions. During the survey, the average students (*i.e.* the ones with normal activity level) reported that: ``it was easy to use the player application and the interface is friendly". Some of the *highly attentive* users complained that other students were taking too long to complete the tasks, delaying their performance in class. We were fairly satisfied with the results and the experiment demonstrates the feasibility and potential of the proposed technology. We can also envision other forms of Analytics models based on the collected data that we intend to test in the future.

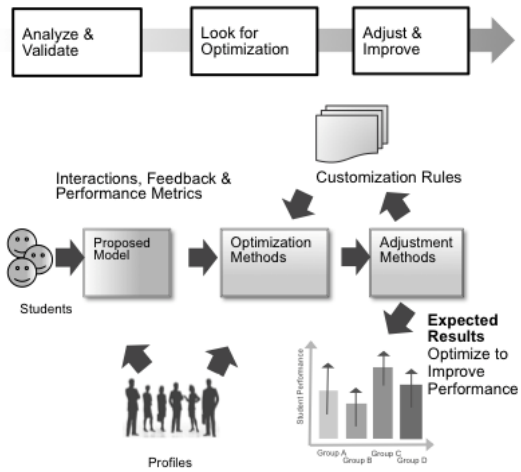

Fig. 3. Future scenario.

IV. Conclusion

This research advances the state-of-the-art by introducing a method to analyse education performance based on patterns in human signals. The techniques of data collection and Learning analytics introduced in this project allow us to 1) classify individual and group behaviour based on human signals in learning environments, and; 2) correlate social behaviour to education performance. We executed a successful experiment in a controlled environment and provided statistical analysis of collected data allowing us to identify categories of student behaviour.

We acknowledge the legislative, ethical, and organisational issues related to the field implementation of this proposal. However, so far we are working on proving the concept and applicability of the solutions.

As a continuation of this work, we are developing advanced Analytic Models by exploiting techniques from the area of topic modelling to detect profile and context. We can use these methods by mapping the discrete elements as: words (features) are equivalent to context conditions and signals; documents (sequence of features) are equivalent to set of collected features from classroom, and; topic are equivalent to student profiles and performances. The questions we are trying to solve with these models are, for instance: 1) what is the probability of a sequence of words (*i.e.* context conditions / signals) resulting in a topic (profile / performance result), and, conversely: 2) What is the probability of topic (profile / performance result) being caused by a (subset of) words (*i.e.* context conditions / signals)?

In the near future, we envisage the structure depicted in Fig. 3 were we deliver variations of digital education material to different groups and measure the performance and other reactions using the Analytic methods (as outlined in this work).

Thus, we curate and exploit large and robust data sets of learning data, which can provide new ways of looking at problems and underpin sound decision-making. Next, we intend to research and develop optimisation methods to recombine the learning objects based on observed performance. In practice, this leads to the creation of adjustment methods to reconfigure the configuration rules used to compose education material. By applying this method, we expect to optimise and improve the education performance across the different groups.

Acknowledgment

This work has been supported and partially funded by FINEP / MCTI, under subcontract no. 03.11.0371.00. Dr Koch was working with IBM Research Brazil and visiting the Faculty of Architecture Building and Planning, The University of Melbourne, Australia, during the time this research was conduced.

References

[1] J. Bartholet, "Big Data goes to school," *Scientific American*, vol. 309, no. 2, pp. 43–45, August 2013.

[2] P. Zikopoulos and C. Eaton, *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, 2011.

[3] F. Koch, M. Ito, A. B. da Silva, S. Borger, and J. Nogima, "Exploiting Human Signals in Learning Environment as an Alternative to Evaluate Education Performance," in *Proc. 3rd International Workshop on Personalization Approaches in Learning Environments (PALE)*, CEUR Workshop Proceedings, 2013, pp. 20–25.

[4] *The Smart School* (White Paper), SAMSUNG, 2013.

[5] A. Pentland, "To Signal is Human," *American Scientist*, vol. 98, no. 3, pp. 204–211, 2010.

[6] A. Pentland, *Honest Signals: How They Shape Our World*, The MIT Press, 2008.

[7] R. Ferguson. "The State of Learning Analytics in 2012: A Review and Future Challenges," Technical Report KMI-2012-01, Knowledge Media Institute, 2012.

[8] G. Siemens. "Learning Analytics: The Emergence of a Discipline," *American Behavioral Scientist*, 2013.

[9] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education," *Educause Review*, vol. 46, no. 5, pp. 30–32, 2011.
[10] S. B. Shum and R. Ferguson. "Social Learning Analytics," *Educational Technology & Society*, vol. 15, no. 3, pp. 3–26, 2012.
[11] *Analytics for Achievement* (White Paper), IBM, 2013.
[12] S. Fletcher, "Learning in the Digital Age–Machine Learning," *Scientific American*, vol. 309, no. 2, pp. 54–60, August 2013.

Fernando Koch is a researcher at Samsung Research Institute Brazil, former Researcher Scientist at IBM Research Brazil (during the time when this study was implemented), and Honorary Senior Fellow at The University of Melbourne. He received the IBM Eminence and Excellence Award and the IBM Outstanding Contributor Award. He obtained the Ph.D in Computer Sciences in 2009, from Utrecht University in collaboration with The University of Melbourne. He obtained the M.Sc. and B.Sc. in Computer Sciences by the Federal University of Santa Catarina, Brazil. He has over 20 years of IT Industry experience, with practice in business and R&D. Dr Koch has over 50 papers in diverse areas, such as Smarter Cities, Health Informatics, Distributed Computing, and Artificial Intelligence.

Chaitanya Rao is a research scientist at IBM Research Australia. He completed Bachelor's degrees in Mathematics and Electrical Engineering at the University of Melbourne, and Masters and Ph.D. degrees in Electrical Engineering at the California Institute of Technology (Caltech). He subsequently worked for NEC Australia in algorithm design and testing for mobile communications, between2007 and 2012. His current research interests are in mathematical modelling and analytics applied to the healthcare, natural resource and education industries.