Anatomy of mobile learners: Using learning analytics to unveil learning in presence of mobile devices

Varshita Sher
School of Interactive Arts and Technology
Simon Fraser University, Canada
vsher@sfu.ca

ABSTRACT
Mobile technology has been a focus of research since the early 2000s and has attracted researchers from various disciplines ranging from pedagogical, health-care, technological to app developers. In recent times, there has been a substantial interest in capitalizing on the abundance and the ubiquity of these technologies for their educational use. However, the role of mobile phones in educational setting is still largely under-researched. Similarly, little attention has been paid to the research on the extension of learning analytics to analyze the learning processes and strategies of students adopting mobile platforms. Traditionally, the research into mobile learning has mainly relied upon self-reported data. While literature has evidence that survey data facilitates extraction of invaluable information, it suffers from a significant shortcoming - unreliability due to learner bias and poor recall. Therefore, this paper outlines the doctoral research project that explores the use of mobile technology in educational context using data mining techniques and learning analytics methods to analyze digital trace data and provide insights into how students learn. The results so far have enabled us to categorize students as adopting one of the three technological modality strategies - strategic, minimalist, and intensive - based on how extensive the use of multiple modalities (such as desktops, tablets, smartphones) is for learning activities and their final academic performance. Our results also provide evidence suggesting incorporation of the modality used by the learner, for carrying out an activity, as a viable feature in learner models helps in improving the prediction power of these models.

Keywords
Learning Analytics, Mobile Learning, Trace Analysis

1. INTRODUCTION
Mobile technology has been a focus of research since the early 2000s and has attracted researchers from various disciplines ranging from pedagogical, health-care, educational, technological to app developers. An October 2011 article in *The Economist* posited that, with the number of PCs already surpassing 1 billion in 2008, the number of mobile devices too would reach 10 billion in 2020 [2]. However, even with the proliferation of mobile phones at such an unprecedented rate, their role in the educational setting is still largely under-researched [1, 12]. The challenge for educators and designers, thus, is one of understanding and exploring how best students might use mobile technology to support learning. This challenge is further complicated by the fact that while there exist plethora of learning analytics dashboards (LADs), there is a critical paucity of mobile learning analytics dashboard applications [11] in comparison to their desktop counterpart. The desktop LADs have been extensively researched in terms of their usability, learner strategies, usefulness, effectiveness, and efficiency using complex data mining techniques and learning analytics methods. On the contrary, little attention has been paid to the research on the extension of learning analytics to analyze the learning process of students adopting mobile platforms. The very few extant mobile LADs [5, 10, 7, 4] adopted by students have been analyzed mostly using self-reported data typically collected through questionnaires or think-aloud protocols, which suffer from unreliability issues due to learner bias and poor recall. Moreover, we are still unaware of the impact of the 'source' of log files (from different devices), if any, on the outcome prediction in learner models.

Thus, our research aims to bridge all the previously discussed gaps and explore the use of mobile technology in an educational context. It aims to provide a holistic understanding of learning in presence of mobile devices and its impact on learning - right from the identification of learning strategies employed by mobile learners, to the learning activities benefiting the most from mobile technology (w.r.t. online discussions and course assignments), and to the construction of technological modality-specific learner models for learning outcome prediction. We focus on using advanced data mining techniques and learning analytics methods to analyze digital trace data and provide insights into how students learn using different technological modalities, with main focus on mobile devices.

1.1 Research Question I
The first goal of the proposed research is to explore how mobile devices are used when regulating learning via learning management systems (LMS) in the context of blended courses. To do so, we mine the sequence data from student

logs to examine the extent to which various technological modalities (including mobile devices, laptops and desktop PCs) are either used sequentially and/or simultaneously and assess their potential to influence the overall academic performance and study habits at various learning activities. In order to achieve that, we study the following two research questions:

***RQ1.1**: Can we detect patterns in students' use of multiple modalities that are indicative of their adopted technological modality strategy when using an LMS tool? If so, what kind of strategies emerge?*

***RQ1.2**: Is there an association of the identified strategies with students' performance in online discussions and overall academic performance?*

1.2 Research Question II

Central to the idea of mobile learning is that learning can occur context-free; across different places and at different times and not confined to the formal classroom settings. While location-aware mobile learning systems has been widely studied, albeit using descriptive statistics only, even lesser attention has been given to the temporal aspect of the use of these mobile technologies. That is to say, not much is known of the associations between different modalities (such as desktops, mobiles and tablets) and the time of the day during which the modality-learner interactions take place. We assess the associations of time of the day they learn, not only with the frequency of usage of a modality, but with the sequential patterns of usage of different modalities in a blended course. Knowledge of the personally-negotiated learning time-frames, mediated by different modalities are conducive to timely, personally tailored feedback, reinforcing the 'right information at right time' learning motto. In order to achieve that, we study the following research question:

***RQ2**: Are there any underlying associations between time of the day and the patterns of modality usage based on a learner's modality-use profile?*

1.3 Research Question III

The introduction of mobile technology as a pedagogical tool has witnessed many enthusiastic supporters who successfully incorporate mobility in their everyday learning routine. However, it is still unclear what dictates the students' decision to adopt or resist mobile technology in the first place. The following research questions ascertain whether learners' patterns of modality usage are driven by their inherent student characteristics or the type of activities they must engage in.

***RQ3**: Do students use technological modalities differently when engaging with different types of learning activities (say, Assignments and Online Discussions)?*

1.4 Research Question IV

The research area of analyzing log file trace data to build academic performance prediction models has tremendous potential for pedagogical support. Currently, these learner models are developed from logs that are composed of one intermixed stream of data, treated in the same manner regardless of which platform (mobile, desktops) the data came from. Given that learners use a combination of devices when engaging in learning activities, it is apparent that weighing the logs based on the platform they originate from might generate different (possibly better) models, with varying priority assigned to different model features. For instance, frequency of course material access might be a less powerful indicator of academic performance compared to the frequency of course material access 'from mobile devices', probably due to the benefits associated with ubiquitous any-time access available to mobile learners. Thus, the primary goal of this research question is to bring to light the potential for improvement of prediction power of models after considering the learner's platform of access while generating learner models for predictive analysis.

***RQ4**: To what extent is the predictive strength of LMS features influenced by distinguishing the modality of learner access when predicting course grade??*

2. POTENTIAL CONTRIBUTIONS

The research on mobile learning has primarily focused on studying the *effectiveness* or *design* of the mobile learning systems [13]. There are two major flaws to this. Firstly, as pointed out by [6], *it is the learner that is mobile, rather than the technology.* Hence, while the research focused on designing of specific portable technologies has been useful, it is now time to dig deeper into complex interactions between learners, mobile devices, learning activities, and available learning materials (briefly touched upon by [1]).

Furthermore, assessments of effectiveness of mobile learning systems are generally conducted using overly general, broad surveys and self-reported questionnaire, usually in a lab setting. The traditional surveys have been recognized as highly flawed in the educational research community due to unreliable issues stemming from learner bias and poor recall and the extremely controlled environment in lab settings deters the observation of learners' actions, decisions and learning strategy choices in their natural environment (thereby threatening external validity of experiment). Consequently, the studies have only sufficed in making superficial claims about *trends* in mobile learning, using simple analyses such as aggregates and percentages - *20% people prefer to use mobile phones for participation in the discussion activity.* This is insufficient for the explication of the actual way in which mobile technology is impacting the everyday learning process in authentic educational settings wherein interleaved pattern of usage are observed which up until now have been understudied [8, 9]. This is exactly what my doctoral thesis will cater to. The advantages of analyzing such patterns is three-fold. It supports instructors in blended courses through more refined interpretation of students' actions in the LMS when participating in the learning activity. For learners, this allows for recognition of strategies (comprising modality-action pairs) that maximizes student's learning achievements and helps in refraining from suboptimal learning strategies. It is equally useful for future LMS designers when designing notification systems that are capable of sending reminders for modality-specific learning actions.

3. METHOD AND CURRENT PROGRESS

In my research, I am analyzing the data produced by the second and third year undergraduate students in two programming oriented courses at a Canadian university. Currently, the data has been collected from two semesters (Fall 2017 and Spring 2018) and is being continually logged from subsequent offerings of the two courses. Each course spans over 13 weeks and for the two semesters, a combined enrollment of 121 students (83+38) was observed. The courses use blended delivery, utilizing the university's learning management system (LMS) to support learning activities and students' overall schoolwork. The LMS hosts access to reading material, posted lecture slides, tutorial materials, general course information, weekly or bi-weekly course assignments, assignment submission, grades, and allows participation in online discussion activities. In addition to the web-browser versions of the LMS (desktop/laptop/mobile), students have access to the mobile app version provided by the LMS vendor. Comparison of the features and functionalities offered by the two versions have revealed no apparent differences.

I plan to carry out my research in four phases, one for each research question. At the current stage, I am working on phase 2 and 3, having completed and submitted research questions from phase 1 (accepted) and phase 4 (in review). Hence, in this section, I will briefly touch upon the methodology used (or intend to use) for each of these research questions and the results obtained so far.

3.1 RQ1

In order to examine the presence of patterns in students' use of several technological modalities, I encoded each learning session as a sequence of modalities (used to carry out each action in that learning sessions) using a representation format of the TraMineR R package [3]. These sequences were clustered using agglomerative clustering based on Ward's method. The computation of the distance (similarity) between sequences, required for the clustering algorithm, was based on the optimal matching distance metric [3]. The sequence clustering algorithm produced four clusters, i.e. technological-modality profiles - Diverse (use of many different modalities), Mobile-oriented, Short-desktop and Desktop. Next, the students were clustered, based on how many of their sequences belong in each modality profile, using Euclidean metric to compute the distance between vectors. As a result, three student clusters (Strategic, Minimalist and Intensive), representative of their modality strategies, were obtained.

To examine if there was a significant difference between the identified student groups, we performed a multivariate analysis of variance (MANOVA). The student cluster assignment was treated as the single, independent variable along with following dependent variables: overall academic score, counts and time spent on viewing, posting and replying to messages along with the word counts and quality of messages in discussion board. I found a moderate effect size ($\epsilon^2 = 0.12$) of students' adopted strategies on the final course grade. Furthermore, when looking specifically at online discussion engagement and performance, students' adopted technological modality strategies explained a large amount of variance ($\eta^2 = 0.68$) in their engagement and quality of contributions.

3.2 RQ2

For this question, I have encoded learning sessions from all students (as was done in RQ1), followed by categorizing each learning session into one of the four broad TOD (time of the day) categories, intuitively: Morning (5 - 11 a.m.), Afternoon (11 a.m. - 4p.m.), Evening (4 - 7 p.m.) and Night (7 p.m. - 5 a.m.). To examine if there was an overall significant relation between the modality-profile of learning sessions and the time of the day each of these sessions took place, a chi-square test of independence was performed across all learning sessions after summarizing data, composed of each sessions' technological modality profile cluster and the TOD category it belongs to, in a two-way contingency matrix. The analyses was replicated across each of the three modality strategies (obtained in RQ1) and results revealed significant associations for all three strategies. However, there were significant differences with respect to the preferred time of the day for carrying out learning sessions belonging to each modality profile, based on the learner's modality strategy. For instance, mobile-oriented learning sessions were carried out mostly in the afternoon by strategic and intensive learners but in the morning by minimalist learners.

3.3 RQ3

The ongoing activities for this research question are mainly focused on preliminary data analysis. I have encoded the learning sequences, from two main learning activities i.e. assignment and online discussions, as representations of the TraMineR format. For each learning activity, the sequence clustering based on optimal matching metric resulted in four and three technological modality profiles (TMP), respectively. Consequently, two different partitioning of the students, one based on TMP clusters from online discussion activity and the other based on TMP clusters from assignment activity, both resulted in 3 student clusters each.

The initial exploratory analysis comparing the two student clusterings, obtained from the two learning activities, revealed a rand index of 0.48, meaning that the two clustering agree to a very small extent only. Therefore, we posit with some certainty that the learner's patterns of modality usage (or in short their choice of technological modality strategy) is dependent on the learning activity they are engaging in. To further strengthen this claim, we look at the rand indices obtained from the two learning activities when compared to the benchmark student classification (from RQ1). The benchmark classification corresponds to the student strategies (i.e. clusters) that are gauged from their overall engagement with the LMS and reflect the *generic* or habitual patterns of use of different modalities (and thus considered analogous to a student's innate characteristics). We found a large overlap (rand index = 0.85) of strategies in the benchmark classification with the assignment activity but only a small overlap (rand index = 0.51) with online discussion activity. This indicates that, in addition to the strategies from the two learning activities being different, the strategies from one of them (i.e. assignment activity) *more* closely resemble the strategies in the benchmark classification, compared to the other (i.e. discussion activity). As with any speculation, follow up inferential analysis will be required to solidify our claims.

3.4 RQ4

To investigate the effect of modality on different types of commonly included learning-related activities (Table 1) and their traces in the online courses, I selected 10 features (5 counts + 5 time spents for each activity) for inclusion in my analyses as predictors of academic success. Variables derived from the LMS trace data include: syllabus, course material (lecture + tutorial slides and instructor provided supplementary material), assignments, feedback on the assignments and calendar. For each student I extracted the number of times (and the time spent on) using a particular feature by aggregating individual operations. I call these variables *LMS features*. Each of these variables was split up further to account for the platform used to access that particular feature. For instance, in addition to having the total number of assignment views for a student, I computed three more variables – mobile views, desktop views and tablet views – which indicate the respective number of assignment views from each of the three main modalities. I call such variables *Modality features*.

Next, I conducted a series of regression analyses with course grade as the outcome variable in each. For each of the ten learning features, two regression models were built using (a) LMS features and (b) Modality features. The results reflected that for each of the ten features, an increase in R^2 from Model 1 to Model 2 was observed. This increase accounts for the percentage of variability in student course grade explained by the Modality features over and above the LMS feature. An ANOVA analysis using F-test of the statistical significance was conducted to ascertain whether the increase was statistically significant and it was found to be significant for more than 50% of the variables tested.

4. ADVICE SOUGHT

My research is at an intermediate stage and I hope the consortium could provide me some guidance and insights in the following areas:

- I am interested to know if there is a method, especially in the area of network analysis that can help me assess the (direction and frequency of) transitions between mobiles, desktops and tablets, and how these transitions differ when learners are engaging in different learning activities.

- For RQ3, while I am focusing on chi-square test of independence to test associations between time of the day and modality patterns, for the next steps I need to know how to identifying/interpret patterns using the time-series analysis. Also, how can I account for the 'random noise' that might be introduced from students' use of different devices for non-educational purposes?

5. ACKNOWLEDGMENTS

This research was supported by Natural Sciences and Engineering Research Council of Canada (NSERC) and Social Sciences and Humanities Research Council of Canada (SSHRC).

6. REFERENCES

[1] N. R. Aljohani and H. C. Davis. Significance of learning analytics in enhancing the mobile and pervasive learning environments. In *Next Generation Mobile Applications, Services and Technologies (NGMAST), 2012 6th International Conference on*, pages 70–74. IEEE, 2012.

[2] T. Economist. Beyond the pc, Oct 2011.

[3] A. Gabadinho, G. Ritschard, N. S. Mueller, and M. Studer. Analyzing and visualizing state sequences in r with traminer. *Journal of Statistical Software*, 40(4):1–37, 2011.

[4] H. Heflin, J. Shewmaker, and J. Nguyen. Impact of mobile technology on student attitudes, engagement, and learning. *Computers & Education*, 107:91–99, 2017.

[5] J. Nakahara, S. Hisamatsu, K. Yaegashi, and Y. Yamauchi. itree: Does the mobile phone encourage learners to be more involved in collaborative learning? In *Proceedings of th 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!*, pages 470–478. International Society of the Learning Sciences, 2005.

[6] M. Sharples, J. Taylor, and G. Vavoula. Towards a theory of mobile learning. In *Proceedings of mLearn*, volume 1, pages 1–9, 2005.

[7] R. Shen, M. Wang, and X. Pan. Increasing interactivity in blended classrooms through a cutting-edge mobile learning system. *British Journal of Educational Technology*, 39(6):1073–1086, 2008.

[8] G. Stockwell. Using mobile phones for vocabulary activities: Examining the effect of platform. 2010.

[9] G. Stockwell. Tracking learner usage of mobile phones for language learning outside of the classroom. *CALICO Journal*, 30(1):118–136, 2013.

[10] B. Tabuenca, M. Kalz, H. Drachsler, and M. Specht. Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, 89:53–74, 2015.

[11] K. Verbert, S. Govaerts, E. Duval, J. L. Santos, F. Van Assche, G. Parra, and J. Klerkx. Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514, 2014.

[12] I. S. H. Wai, S. S. Y. Ng, D. K. Chiu, K. K. Ho, and P. Lo. Exploring undergraduate studentsâĂŹ usage pattern of mobile apps for education. *Journal of Librarianship and Information Science*, page 0961000616662699, 2016.

[13] W.-H. Wu, Y.-C. J. Wu, C.-Y. Chen, H.-Y. Kao, C.-H. Lin, and S.-H. Huang. Review of trends from mobile learning studies: A meta-analysis. *Computers & Education*, 59(2):817–827, 2012.