DOCTORAL CONSORTIUM PAPERS

Dynamic User Modeling within a Game-Based ITS

Erica L. Snow
Arizona State University
Tempe, AZ, 85283
Erica.L.Snow@asu.edu

ABSTRACT
Intelligent tutoring systems are adaptive learning environments designed to support individualized instruction. The adaptation embedded within these systems is often guided by user models that represent one or more aspects of students' domain knowledge, actions, or performance. The proposed project focuses on the development and testing of user models within the iSTART-2 intelligent tutoring system, which will be informed by dynamic methodologies and data mining techniques. My previous work has used post hoc dynamic methodologies to quantify optimal and in-optimal learning behaviors within the game-based system, iSTART-2. I plan to build upon this work by conducting dynamical analyses in real-time to inform the user models embedded within iSTART-2. I am seeking advice and feedback on the statistical methods and feature selection that should included within the new dynamic user model. The implications of this approach for both iSTART-2 and the EDM field are discussed.

Keywords
Intelligent Tutoring Systems, Dynamic Analysis, Adaptation, Log Data, User Models

1. INTRODUCTION
Intelligent tutoring systems (ITSs) are adaptive learning environments that provide customized instruction based on students' individual needs and abilities [1]. ITSs are typically more advanced than traditional computer-assisted training in that they *adapt* to the users' performance and skill levels [2]. The customizable nature of ITSs has resulted in the successful integration of these systems into a variety of settings [3-5].

One hypothesized explanation for the widespread success of these systems is that ITSs provide individualized feedback and adjust content based on the unique characteristics of each student or user. This pedagogical customization allows students to progress through learning tasks at a pace that is appropriate to their individual learning model [6]. It also ensures that students are not only learning at a shallow procedural level, but they are gaining deeper knowledge at an appropriate pace.

One way in which ITSs store and represent information about learners is via *user models*. User models embedded within ITSs incorporate detailed representations of learners' knowledge, affect, and cognitive processes [7]. It is important to note that these models are often continuously updating throughout the students' interaction within the system. Thus, potentially, every student action or decision made within the system contributes to more accurate and holistic user models. Although this concept seems to be intuitive, researchers often struggle to determine what information belongs within the models and how to optimally quantify the dynamic nature of that information.

In prior work, my colleagues and I have proposed that dynamical systems theory and associated analysis techniques are useful tools for examining behavioral patterns and variations within ITSs [8,9]. Indeed, dynamic systems theory affords researchers a unique means of quantifying patterns that emerge from students' interactions and learning behaviors within an ITS. This approach treats time as a critical variable by focusing on the complex and fluid interactions that occur within a given environment rather than treating behavior as static (i.e., set or unchanging), as is customary in many statistical approaches. In the proposed work, I hypothesize that dynamical methodologies have strong potential to inform user models by quantifying changes in students' interactions and learning behaviors across time. This quantification and modeling of behavior can inform decisions about how content and feedback should be presented to each student based on their current learning trajectory. The overall goal of the proposed work is to test the utility of real-time dynamic analyses as a way to inform user models about optimal (and non-optimal) learning behaviors within a game-based ITS.

1.1 iSTART-2
Interactive Strategy Training for Active Reading and Thinking-2 (iSTART-2) is a game-based ITS designed to improve high school students' reading comprehension via self-explanation strategies [10]. In previous studies, iSTART-2, and its predecessors, have been shown to be effective at improving students' self-explanation quality and reading comprehension ability [11, 12].

iSTART-2 consists of two phases: self-explanation training and game-based practice. During training, students watch a series of videos that introduce them to and provide examples of self-explanations strategies. After students view these videos, they transition to practice (see Figure 1 for a screenshot of the game-based practice interface). During practice, students are able to interact with a suite of mini-games, personalizable features, and achievement screens [13]. The game-based practice embedded within iSTART-2 is designed to promote the generation and identification of self-explanation strategies. Within these practice games students are exposed to game mechanics that serve as a form of feedback on their understanding of the self-explanation strategies (see [11] for more details).

The interface of iSTART-2 uniquely affords students substantial agency and control over their learning path by allowing them to choose how they engage with the practice environment [9]. Such freedom also affords researchers with the opportunity to explore and *model* how and when students engage with these features and activities, and to explore the implications of such choices (i.e., how they affect performance and learning).

Figure 1. Screen shot of iSTART-2 Selection Menu

1.2 Current Work

My doctoral research will use educational data mining methods to inform and build dynamic student models within game-based ITSs such as iSTART-2 [13]. Specifically, this study will explore how dynamic techniques such as Hurst exponents and Entropy analysis can be used in real-time to quantify students' behaviors, performance, and cognition while they learn within iSTART-2. Analyses of students' logged choices have been shown to be a *blueprint* regarding successful and unsuccessful behaviors for learning [13, 15]. Therefore, the logged information from iSTART-2 will be used in conjunction with dynamical analysis techniques as a means to quantify various types of in-system behaviors and their impact on learning outcomes in real-time. This information will then be used to adapt the pedagogical content students are exposed to.

2. Proposed Contributions of Current Work

The current work has both *local* and *global* implications. Locally, the development of dynamic user models will improve iSTART-2 pedagogy. Currently, iSTART-2 has limited user models (only guides self-explanation feedback) embedded within the system. Thus, the inclusion of a dynamic user model is expected to improve system feedback and guide the content presentation provided to students. For instance, one research question that arises from this work is how to support optimal learning trajectories for every student. Dynamic user models have the potential to recognize non-optimal learning behaviors and provide feedback or navigate students toward more effective learning behaviors within the practice environment. Thus, it is hypothesized that the implementation of dynamic models will improve the design and generalizability of iSTART-2.

Globally, this project will contribute to the AIED and EDM fields. User models are an important and often crucial aspect of ITS development. However, very few systems (if any) use dynamic data mining techniques to inform their student models. This work will be among the first studies to use techniques such as Hurst exponent analysis in real-time to inform user models that will ultimately be used to adapt the content and feedback presented to students. The methods presented here are generalizable and thus can be used in a variety of settings beyond iSTART-2. Although the goal of the current work is to design user models for the iSTART-2 system, this work is driven by the overarching goal of gaining a better understanding of students' learning processes.

3. Previous Work

My previous research has revealed that dynamic methodologies are useful tools for quantifying students' behavioral patterns within iSTART-2 [8,9,13,14,15]. For instance, Entropy is a dynamical methodology used to measure the amount of predictability within a system or time series [16]. My colleagues and I have employed post hoc Entropy analysis to quantify variations in students' behaviors within iSTART-2 and related them to performance differences. Based on students' choices within games, an Entropy score can be calculated that is indicative of the degree to which students' choice patterns are controlled versus random. In [13], students' Entropy scores were included within a regression analysis to examine how students' choices within the system influenced their self-explanation performance. Students who engaged in more controlled interaction patterns (i.e., strategic and planned out) within iSTART-2 also generated higher quality self-explanations compared to students who acted in more random or impulsive manners.

While Entropy provides an overall view of students' choice patterns within a system, it does not capture fine-grained fluctuations that manifest over time. To address this issue, Hurst exponents have been conducted using iSTART-2 log data. Hurst exponents [17] are similar to Entropy analyses in that they quantify tendencies or fluctuations present within a time series. However, Hurst exponents also act as long-term correlations that can characterize the fluctuations that manifest across time. Hurst exponents classify these fluctuations as persistent, random, or antipersistent [18]. Using this approach, we can identify when students choose to perform the same action(s) repetitively [8]. This technique affords a fine-grained look at students' behaviors across time. Although Entropy and Hurst exponent analyses have shed light upon the effects of students' interactions within an ITS on learning, the analyses thus far have all been conducted post hoc (i.e., using data mining techniques). Thus, the current work seeks to build upon these dynamical analyses and apply dynamic data mining techniques in *real-time* as a means to inform student models within iSTART-2.

4. Advice Sought

For this doctoral consortium, advice is sought regarding two core concerns. First, *what features should be included in dynamic user models?* Currently, I have solely focused on students' behaviors and in-system performance within the game-based practice portion of the system. However, iSTART-2 has powerful logging functionality capable of collecting everything from mouse movements to keystrokes. Thus, in this setting I would benefit from expert opinions or discussions concerning what features should (or could) be included within dynamic user models.

Second, *what other dynamic methodologies and tools are available and relevant to user modeling?* Thus far, I have used random walks, Entropy and Hurst analyses. However each of these measures have one or more weaknesses. For instance, to reliably calculate a Hurst exponent, multiple data points are needed (e.g., over 100), therefore calculating Hurst in real-time may not be practical in all situations (i. e., a single session study). Thus, I would benefit from expert opinion and guidance regarding other dynamic measures or methodologies that could be used in real-time as a way to inform user models within iSTART-2.

5. ACKNOWLEDGMENTS

Prior research was supported in part by the Institute for Educational Sciences (IES R305G020018-02; R305G040046, R305A080589) and National Science Foundation (NSF REC0241144; IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the

IES or NSF. I would also like to thank Danielle McNamara, Rod Roscoe, Tanner Jackson, and Matthew Jacovina for their contributions to this line of work.

6. REFERENCES

[1] Murray, T. 1999. Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, *10,* (1999), 98-129.

[2] Shute, V. J., and Psotka, J. 1994. *Intelligent Tutoring Systems: Past, Present, and Future* (No. AL/HR-TP-1994-0005). ARMSTRONG LAB BROOKS AFB TX HUMAN RESOURCES DIRECTORATE.

[3] Johnson, W. L., and Valente, A. 2009. Tactical Language and Culture Training Systems: using AI to teach foreign languages and cultures. *AI Magazine*, *30*, (2009) 72.

[4] Lynch, C., Ashley, K., Aleven, V., and Pinkwart, N. 2006. Defining ill-defined domains; a literature survey. In Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems (Jhongli, Taiwan, June 26-30, 2006) Lecture Notes in Computer Science / Programming and Software Engineering pp. 1-10.

[5] Siemer, J., and Angelides, M. C. 1995, December. Evaluating intelligent tutoring with gaming-simulations. In *Proceedings of the 27th conference on Winter simulation* (Arlington, VA, December 03 - 06, 1995), IEEE Computer Society, pp. 1376-1383.

[6] Aleven, V. A., and Koedinger, K. R. 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive science*, *26* (2002), 147-179.

[7] Kay, J., Halin, Z., Ottomann, T., and Razak, Z. 1997. Learner know thyself: Student models to give learner control and responsibility. In *Proceedings of International Conference on Computers in Education* (pp. 17-24).

[8] Snow, E. L., Allen L. K., Russell, D. G., and McNamara, D. S. 2014. Who's in control?: Categorizing nuanced patterns of behaviors within a game-based intelligent tutoring system. In J. Stamper, Z. Pardos, M. Mavrikis, and B. M. McLaren (Eds.), Proce*edings of the 7th International Conference on Educational Data Mining* (London, UK July 4 -7, 2014). Heidelberg, Berlin, Germany: Springer, pp. 185-192.

[9] Snow, E. L., Jacovina, M. E., Allen, L. K., Dai, J., and McNamara, D. S. 2014. Entropy: A stealth assessment of agency in learning environments. In J. Stamper, Z. Pardos, M. Mavrikis, B. M. McLaren (eds.), Proce*edings of the 7th International Conference on Educational Data Mining* (London, UK July 4 -7, 2014). Heidelberg, Berlin, Germany: Springer, pp. 185-192.

[10] McNamara, D. S., Levinstein, I. B., and Boonthum, C. 2004. iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers*, 36, (2004), 222-233.

[11] Jackson, G. T., and McNamara, D. S. 2013. Motivation and Performance in a game-based intelligent tutoring system. *Journal of Educational Psychology, 105*, (2013), 1036-1049.

[12] McNamara, D.S., O'Reilly, T., Best, R., and Ozuru, Y. 2006. Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34*, (2006), 147-171

[13] Snow, E. L., Allen, L. K., Jacovina, M. E., and McNamara, D. S. 2015. Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. *Computers and Education, 26*, (2015), 378-392.

[14] Snow, E. L., Likens, A., Jackson, G. T., and McNamara, D. S. 2013. Students' walk through tutoring: Using a random walk analysis to profile students. In S. K. D'Mello, R. A. Calvo, and A. Olney (Eds*.), Proceedings of the 6th International Conference on Educational Data Mining,* (Memphis, Tennessee, July 6 -9, 2013), Springer Berlin Heidelberg, 276-279.

[15] Snow, E. L., Jackson, G. T., and McNamara, D. S. 2014. Emergent behaviors in computer-based learning environments: Computational signals of catching up. *Computers in Human Behavior, 41*, (2014), 62-70.

[16] Grossman, E. R. F. W. 1953. Entropy and choice time: The effect of frequency unbalance on choice-response. Quarterly *Journal of Experimental Psychology, 5*(1953), 41-51.

[17] Hurst, H. E. 1951. Long-term storage capacity of reservoirs. Transactions of the American Society of Civil Engineers, 116, (1951), 770-808.

[18] Mandelbrot, B. B. 1982. The fractal geometry of nature. New York: Freeman.

Use of Time Information in Models behind Adaptive System for Building Fluency in Mathematics

Jiří Řihák
Masaryk University Brno
thran@mail.muni.cz

ABSTRACT

In this work we introduce the system for adaptive practice of foundations of mathematics. Adaptivity of the system is primarily provided by selection of suitable tasks, which uses information from a domain model and a student model. The domain model does not use prerequisites but works with splitting skills to more concrete sub-skills. The student model builds on variation of Elo rating system which provide good accuracy and easy application in online system. The main feature of the student model is use of response times which can carry useful information about mastery.

1. INTRODUCTION

Our aim is to develop a practice system focused on basic mathematics which uses concepts of Computerized adaptive practice [8], i.e. to provide children with tasks that are most useful to them. We focus especially on detecting mastery and fluency using both correctness and timing information about children's responses.

Mathematics is usually associated with procedural knowledge. However, for achieving mastery of advanced topics it is necessary to solve some basic mathematical tasks at the level of fluency and automaticity. Good example of this is multiplication of small numbers which starts as procedural knowledge (child knows that $3 \cdot 5$ is $5 + 5 + 5$ and is able to complete calculation) but ends as declarative knowledge (child knows $3 \cdot 5$ is 15 without further thoughts) [15]. In both cases child gives correct response with high probability and the system is not able to distinguish between these scenarios based only on the correctness of the answer. Thus we want incorporate into our student model the information about response time, which is necessary to detect mastery, the state when the child is correct and fast.

Because our goal is to lead a child to automaticity we want to analyse strengths and weaknesses of the child at the level of individual items. Thus we need to track child's skills in great detail and we treat every item in the system independently. Also the fact that various graphical representations of the same task influence difficulty of the item, highlight need to track their difficulty individually. To estimate correctly difficulties of the items requires a lot of expertise, it is time consuming and is not always reliable. Therefore we do not want to make any assumptions about difficulties of the items and we rather use model which can estimate the difficulty of the solving data from the system. As a consequence we will be able to easily analyse which items are more difficult and why.

Proposed system is called MatMat and is currently available online in beta version at `matmat.cz` for all children (the system is so far implemented only in Czech) and it is free to use. The goal of the system is to provide adaptive practice of arithmetic operations which guide children from basic work with numbers (e.g. counting objects) to mastery of basic mathematical operations.

In contrast with complex intelligent systems for learning mathematics as Carnegie Learning's Tutors [14, 9] or ASSISTments [4] we focus only on small part of learning mathematics and we work only with atomic tasks. Therefore the system does not work with explanations of curriculum or hints and focuses on adaptive selection of tasks and appropriate feedback. Between related systems belongs Dybuster Calcularis [6] which works with basic math especially in context of dyscalculia; Math Garden [8] which has similar focus, works with similar student model and also incorporates time information; or FASTT Math [3] which also focus on building computational fluency.

2. MODELS

In this section we describe working draft of the domain model, which describes how is the content of the system organized, and the student model, which is built on the domain model and provides information about children who interact with the system. We have several requirements for the design of our models. We are in the situation when we use models in online environment and we rely more on collected data instead of expertise or other outside information. Hence we require models which can work on the fly and can quickly adapt to new data in the system. The goal of the student model is to provide estimation of child's abilities which are used for creation of feedback and selection of suitable tasks to practice.

2.1 Domain Model

Mathematics is very complex domain full of diverse components and relationships. Even in our very simplified case, when we considered only basics, situation can still be relatively complicated. One way how to build a domain model for mathematics is based on Knowledge space theory [1]. This approach splits the curriculum to skills a defines relations of prerequisites between them. This oriented graph can then be treated as dynamic Bayesian network [7].

We used different approach which allows us to capture information about very specific abilities, e.g. how good is child in multiplication of 5 and 7. The relations between such concrete skills, are not always prerequisites, e.g. the abilities to compute $5 \cdot 7$ and $5 \cdot 9$ are not one prerequisite to another but they are clearly dependent. We organized the skills into the tree structure (Figure 1) where every node corresponds to skill and its successors to more concrete sub-skills. Similarity of skills then can be expressed as level of the nearest common ancestor. Denote the fact that a skill d is ancestor of a skill c as $d > c$.

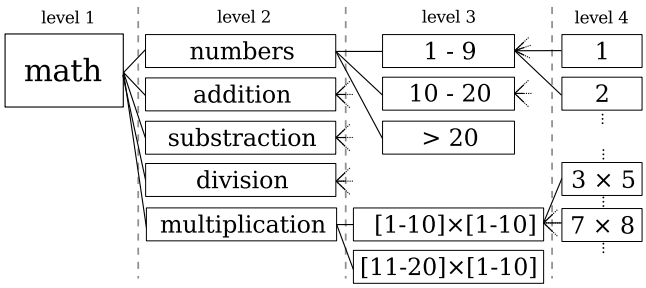

Figure 1: The tree structure of the skills

The root of the tree is a global skill which represents overall knowledge of mathematics. Under that are skills which correspond to basic units in system (level-2) — numbers, addition, subtraction, multiplication and division. In level-3 are sub-skills which represent concepts (inspired by [6]) within parent skill, e.g. under 'numbers' skill are 'numbers in range from 1 to 9', 'numbers in range from 10 to 20', 'numbers greater then 20', ...; or under 'addition' are 'addition in range from 1 to 9 (without bridging to 10)', 'addition in range from 10 to 20 with bridging to 10', ... And finally level-4 skills correspond to the tasks for which mastery on the level of declarative knowledge is expected. Example of these are skills that correspond to numbers (1, 2, 3, ...), simple addition tasks ($1+2$, $5+7$) or multiplication of numbers smaller than 10 ($3 \cdot 5$, $7 \cdot 8$). There are no level-4 skills for more complicated task (e.g. $11 \cdot 13$) for which procedural knowledge is more involved. The items representing these tasks belong typically to more general level-3 skills.

In current model every item in the system is mapped to exactly one skill (typically a leaf skill). So under a skill are multiple items. In case of the more general level-3 skills it can be tens or hundreds. In case of the level-4 skills there are from 2 to 10 items which are various forms of the task ($5+7$ and $7+5$) and different graphical representations of task (numbers, objects, number line ...).

2.2 Student Model

Rather then the discrete representation of ability (known or unknown) we used the continues representation, which is more suitable for our situation when we need to track abilities also for relatively general skills. The relation between these abilities and expected probability of correct answer is defined by a logistic function.

For the skill from s and the child c model estimates the value v_{sc} which represents difference of ability relative to parent skill. Overall value of ability is then $\theta_{sc} = \sum_{s \leq \bar{s}} v_{\bar{s}c}$. This approach allows to capture relations between leaf skills. Information obtained from observation about one ability can be naturally propagated to other related abilities. This is especially important for new children in the system with small number of responses (relatively to large number of abilities). The model also estimates the difficulties β_i of the items i, which can be interpreted as a required ability to have 50% chance of solving item correctly. Expected response is then $e_{ci} = \frac{1}{1+e^{\beta_i - \theta_{sc}}}$.

To estimate abilities and difficulties we used a model based on Elo rating system [2] and PFA [12] which is inspired by models which have been successfully used in other projects [8, 11]. The main idea is to update all related abilities and item difficulty based on unexpectedness of response after every answer. To empathize the fact that the correct answer (even repetitive) does not mean mastery we need to take into account the response time t_{ci}. This can be achieved by extension of discrete response r_{ci} (correct or incorrect) to continuous one where values between 0 and 1 mean the correct answer but with longer time than the targeted time τ_i. Example of this extension is decay of the response value exponentially relatively to the ratio of t_{ci} and τ_i (Figure 2).

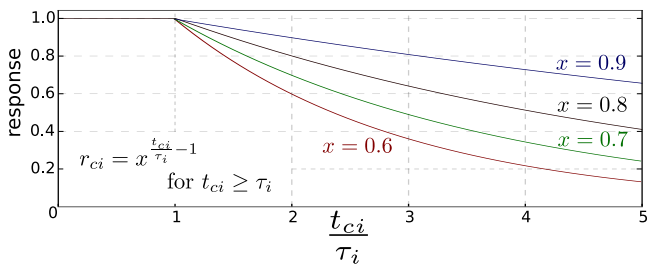

Figure 2: The response value for the correct answers

After the answer, all abilities θ_{sc} belonging to ancestors' skills s are updated. Updates of abilities are performed sequentially from the root of the skill tree. If the answer is the first answer of the child to the item, the difficulty of item β_i is also updated.

$$\beta_i = \beta_i + \frac{\alpha}{1 + \beta \cdot n_i} \cdot (e_{ci} - r_{ci}),$$
$$\theta_{sc} = \theta_{sc} + \gamma_s \cdot K_r \cdot (r_{ci} - e_{ci}).$$

The parameters α and β define shape of the decay function [13] which prevents excessive influence of recent responses. The decay function takes argument n_i – number of previous updates of that difficulty. Parameter K_r corresponds to PFA updates and depends on correctness of answer. Parameter $\gamma_s \in [0, 1]$ tells how much response to the item testifies about

a skill and consequently, how much information obtained from response is propagated to sub-skills. Reasonable values of γ_s are near 1 for the most concrete skills and near 0 for the global skill.

2.3 Item Selection

The selection of an appropriate item that suits the ability of a child is a key feature of the system and has to balance several aspects. The system should not select the same or similar item in a short time, it should select diverse items for better exploration of child's abilities and, foremost, the system should select items with appropriate difficulty – not already mastered (high probability of success) and not too difficult (small probability of success). Currently used algorithm is very similar to the one described in [11]. Only difference is in bringing into account also similarity of items (e.g. $5 + 7$ is similar with $7 + 5$).

3. FUTURE WORK

Most of adaptive educational systems currently work only with correctness of responses. Our goal is to find out if this classical approach can be robustly extended by taking into account timing information and if this extension can be useful in building fluency in the basic mathematical tasks. To target this questions we proposed the system described in this work. This system is still in testing phase but the first analysis of 28 thousand collected answers, show that the ability and difficulty values estimated by the student model make intuitive sense, the system can adapt quickly and the item selection algorithm works reasonably. However, there is a lot of space for improvement.

The domain model can be enriched with prerequisites which can be useful for both ability estimation and for item selection. The current choice of the skills used in the domain model should be reviewed by a domain expert or compared with automatic methods which use collected data [10]. The proposed student model is incorporating response time but current approach is quite simplified and explicitly does not distinguish between accuracy and speed, which can be modeled separately. Also it is not clear how to set, or rather automatically estimate, targeted response times τ_i. Next characteristic of the model is propagation of information about abilities across all skills, which is useful in first phases but later can be undesirable. The propagation is closely connected to parameters γ_s and their influence to the model behaviour should be investigated.

To evaluate our approach the proposed models will be compared to alternative models (e.g. Bayesian network model [7] which works with prerequisites) or simpler versions of Elo model (e.g. model which uses only one global skill and independent local skills [11]). The comparison of the models can be done offline with respect to the quality of predictions or online by comparison of an improvement rate or behaviour of children groups using different models and item selection strategies. These comparisons should bring some light an whether the proposed methods are useful.

Acknowledgements

Author thanks Juraj Nižnan for cooperation on the described system, Radek Pelánek for guidance and useful comments and Jan Papoušek for fruitful discussions.

4. REFERENCES

[1] Jean-Paul Doignon and Jean-Claude Falmagne. Spaces for the assessment of knowledge. *International journal of man-machine studies*, 23(2):175–196, 1985.

[2] A. E. Elo. *The rating of chess players, past and present, volume 3*. Batsford London, 1978.

[3] Ted S Hasselbring, Alan C Lott, and Janet M Zydney. Technology-supported math instruction for students with disabilities: Two decades of research and development. *Retrieved December*, 12:2005, 2005.

[4] Neil T Heffernan and Cristina Lindquist Heffernan. The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4):470–497, 2014.

[5] Tanja Käser, Gian-Marco Baschera, Juliane Kohn, Karin Kucian, Verena Richtmann, Ursina Grond, Markus Gross, and Michael von Aster. Design and evaluation of the computer-based training program calcularis for enhancing numerical cognition. *Frontiers in psychology*, 4, 2013.

[6] Tanja Käser, Alberto Giovanni Busetto, Gian-Marco Baschera, Juliane Kohn, Karin Kucian, Michael von Aster, and Markus Gross. Modelling and optimizing the process of learning mathematics. In *Intelligent Tutoring Systems*, pages 389–398. Springer, 2012.

[7] S Klinkenberg, M Straatemeier, and HLJ Van der Maas. Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education*, 57(2):1813–1824, 2011.

[8] Kenneth R Koeclinger, Albert T Corbett, and Steven Ritter. Carnegie learning's cognitive tutor: Summary research results. 2000.

[9] Juraj Nižnan, Radek Pelánek, and Jiří Řihák. Mapping problems to skills combining expert opinion and student data. In *Mathematical and Engineering Methods in Computer Science*, pages 113–124. Springer, 2014.

[10] Jan Papoušek, Radek Pelánek, and Vít Stanislav. Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining (EDM)*, pages 6–13, 2014.

[11] Philip I. Pavlik, Hao Cen, and Kenneth R. Koedinger. Performance factors analysis - a new alternative to knowledge tracing. In *Proc. of Artificial Intelligence in Education (AIED)*, volume 200 of *Frontiers in Artificial Intelligence and Applications*, pages 531–538. IOS Press, 2009.

[12] Radek Pelánek. Time decay functions and elo system in student modeling. In *Educational Data Mining (EDM)*, pages 21–27, 2014.

[13] Steve Ritter. The research behind the carnegie learning math series. *Online]. Disponible en: http://www. carnegielearning. com*, 2011.

[14] Duane F Shell, David W Brooks, Guy Trainin, Kathleen M Wilson, Douglas F Kauffman, and Lynne M Herr. *The unified learning model: How motivational, cognitive, and neurobiological sciences inform best teaching practices*. Springer Science & Business Media, 2009.

Doctoral Consortium: Integrating Learning Styles into adaptive e-learning system.

Huong May Truong
Eduworks Initial Training Network
Corvinno Technology Transfer Centre, Budapest, Hungary
mtruong@corvinno.com

ABSTRACT
This paper provides an overview and update on my PhD research project which focuses on integrating learning styles into adaptive e-learning system. The project, firstly, aims to develop a system to classify students' learning styles through their online learning behaviour. This will be followed by a study on the complex relationship between learning styles, learning supports and learning outcomes. The findings can contribute significantly to the area that is still left with several unanswered questions. In addition, based on the results, meaningful recommendations and suitable online adaptation can also be made to a wide range of stakeholders of the education system.

Keywords
Learning Styles; E-learning; Adaptive learning system; Data mining; Learning analytics.

1. INTRODUCTION
Learning styles which are defined as students' preferred ways to learn can play an important role in the development of the e-learning system. With the knowledge of different styles, the system can offer insights and advices to a wide range of stakeholders such as students and teachers to effectively organise their learning materials and studying activities to optimise the learning paths. For example, under Felder-Silverman's learning styles frameworks [5], students may prefer to process information actively or reflectively. For "active" students, they perform better through interaction with other students compared to the traditional classroom. Thus, it is advisable for teachers to provide such group the opportunity to interact and discuss the learning topic [5] A recent report by Thalmann [17] surveying e-learning system developers even suggested that learning styles were the most useful personalization sources among other factors such as background knowledge and user history. In addition, there are clear potential benefits for both fields of learning styles research and e-learning system development. On one hand, the integration can help to improve the e-learning experience, providing means to build rules for personalising resources. On the other hand, the e-learning system which allows data mining and computerized algorithms can offer opportunity to observe, analyse and gain further information into students' learning styles throughout the whole learning process which could not easily be done in traditional learning styles theories research.

Nevertheless, integrating the traditional theories which have the base in psychology, pedagogy and cognitive research into the online environment is not a straight forward task. Measurement methods provided by traditional theories are mostly based on long self-judgment questionnaires [4] and thus, do not provide sufficient means fitting to the e-learning system. Furthermore, scholars still do not agree on how to optimize the matching process between learning styles and learning supports [4, 16] which leaves places for further exploration.

With the motivation to address these research problems of integrating learning styles into adaptive e-learning system, this paper contains my proposals as well as the current research progress.

2. PROBLEM STATEMENT AND PROPOSED CONTRIBUTIONS

2.1 Research Questions
In a more comprehensive way, learning styles, according to Keefe [11], can be defined as: "The composite of characteristic cognitive, affective, and physiological factors that serve as relatively stable indicators of how a learner perceives, interacts with, and responds to the learning environment". On the traditional theories side, which is mainly based on psychological, pedagogical and cognitive research, the review by Coffield, Moseley, Hall, and Ecclestone, [4] has identified over 70 theories and models. While there are no theories that outperform others [4], theories that consider the flexibility and changes of styles overtime appear to be more popular in e-learning application. Notable theories in this group include: Felder-Silverman's learning styles theory [5] which divides learners based on their: information input, information process, perception, and understanding, Kolb's Learning styles inventory [12] and Honey and Mumford's Learning styles [10] which both divide styles based on their proposed learning cycles.

The theories undoubtedly provide an essential foundation for learning style research. Nevertheless, there are several unsettled issues when applying to the online environment. In this proposal, with the aim to integrate learning styles into e-learning systems, I focus on two main ones: a) learning style classification system in e-learning and b) the relationship between learning styles, learning support and learning outcomes.

2.1.1 Learning Styles Classification
In terms of learning styles measurements, a review by [4] shows that almost all of the theories are assessed by questionnaires or surveys, requiring learners to evaluate or rank their own styles and behaviours. This type of qualitative measurement suffers many downsides. Firstly, it relies on students' self-judgments which can be bias. Secondly, although learning styles, according to many theories, can change over time, surveys and questionnaires only measure styles at one point in time. Several surveys are, in addition,

questioned by critics in terms of validity and reliability [4]. It is time consuming as there are surveys that can reach over 40 - question long (e.g. [16, 23]), and as a result, they may not be updated easily. Hence, these disadvantages of a long, time consuming, and self-judgement-based measurement create several difficulties when it comes to the adaptive e-learning system development.

In recent years, the application of machine learning which allows computerized algorithms to quickly analyse and mine huge online behaviour dataset provides the opportunity to develop new measurement methods that overcome the current drawbacks. As a result, it has opened a call for integrating learning styles with e-learning system using machine learning application [1, 14].

With the area is still at its early stage, there is still only a few proper peer-reviewed researches that attempt to tackle this theories integration issue [1]. Numerous problems remain unanswered. Firstly, several learning styles predictors can be traced in previous literature which show a complex relationship between learning styles and online behaviour. For example, to measure learning styles under Felder-Silverman's framework, while [6] used attributes related to forums, chats, exam revision etc., [20] measured using variables related to assessment such as questions answering time, performance on the test, questions checking time etc. Nevertheless, through my literature review of 51 previous papers [18], none of the papers has managed to compare the power of different predictors. The results of such comparisons will very interesting and valuable as it can act as guidelines for future developers and contribute significantly in improving the performance and efficiency of classification models.

Secondly, in terms of machine learning classification algorithms, among 51 papers reviewed [18], the most popular method identified is Bayesian networks (and Naïve Bayes – a special case of Bayesian network) (e.g.[6, 7],) which has the base in Bayes theorem. This type of approach has shown positive results in a number of researches so far. Nevertheless, for Bayes theorem to work, it requires a number of conditional probabilities and the relation network to be identified which are not always straight forward tasks. Another popular branch of methods is rules based (e.g. [7, 20]). This group of methods is interpretable, however, it relies heavily on how well the researchers "translate" the theory into the online world. For example, Graf et al., [8] based on the description of learning styles from Felder and Silverman's to obtain "rules" e.g. If a student used exercise more frequently, he is more likely to prefer active learning style. The remaining group of researchers still focuses mainly on single supervised methods which left places for the application of other advanced machine learning methods such as hybrid and ensemble machine learning that combine different machine learning algorithms together. Such advanced methods have shown significant higher performance than single algorithm in other applications such as medical and finance ([3, 19]).

Finally, current models also lack generalisation ([2, 15]). Researches are still employed to only one particular context. Akbult and Cardak [1] pointed out that the research population for almost all of the researches is still limited to undergraduate students. Thus, it raises the question if such models can be applied to a different situation from their own.

These open gaps for a better classification model found in learning styles research field have led to the following research questions:

- How can we incorporate machine learning and traditional learning styles theories? How can we measure learning styles through online behaviour?
- Which predictors are the most meaningful in predicting learning styles in online environment? What is the relationship between online behaviour and learning styles?
- What is a more effective way for learning styles classification compared to current approaches?
- Is it possible to generalize the measurement method?

2.1.2 The relationship between learning styles, learning support and learning outcomes.

The second issue relates to the relationship between learning styles, learning supports methods and learning outcomes. On one hand, students with different learning styles prefer to study in different ways. On the other hand, researchers still do not agree on how to optimise this matching process between learning styles and learning supports and interventions ([4, 16],). At the same time, the relationship between learning styles and learning outcomes is still unclear [1]. Pashler, McDaniel, Rohrer and Bjork [13] reported that previous researches still show flaws in their methodology, which as the result, fail to persuasively show the effect of learning instruments on students with different learning styles. There are also several contradictory results. For example, Ford and Chen (2001 cited in [4]) suggested that matching students learning styles with their preferred teaching style is associated with better learning results. However, Holodnaya [9] found that it will be beneficial to study under a mismatched condition. Consequently, to be able to provide reliable feedback to different stakeholders of the education system, it is essential to revisit the issue. The following research questions have been raised:

- How can we match learning supports to learning styles to improve learning outcomes?
- Under the same condition, are learning styles making any differences to learning outcomes? Are there any styles that are more preferable under certain circumstances?

2.2 Potential contributions

Overall, the area of integrating learning styles theories into e-learning systems has gained interest over the past years, yet there are still many questions that are underexplored. This research, thus, firstly, will address a number of research gaps in the field such as the relationship and influence of different online attributes on learning styles. Interesting patterns between different styles and behaviours can, as the consequence, be identified. Secondly, it aims to advance in the methods for learning styles classification which will improve the accuracy and efficiency. Lastly, it will reconfirm the debate in terms of the relationship between learning styles, learning outcomes and learning supports that can contribute significantly in helping the students to excel in their study. In addition, the findings can also work as guidelines and contribute for future e-learning development research.

3. PROPOSED METHODS AND CURRENT PROGRESS

The research will be carried out in 2 phases that each dedicates to a problem mentioned in section 2. At the current stage, I focus on phase 1 which is to develop a learning styles classification system. Thus, this section will centre mainly on phase 1's method and updates.

To develop a classification method, the following process will be carried out: it will start off with learning style theories selection, then attributes selection and finally, classification methods development and evaluation.

Firstly, while the learning styles classification field is crowed [4], through careful review in terms of theories reliability, validity, usefulness in recommendation, in this study, I chose to follow Felder-Silverman theory which is one of the most popular theories implemented in e-learning system [1]. Hence, it will also provide the opportunity for performance benchmarking.

In terms of attributes selection, I have carried out a literature-based survey [18] focusing on not only previous personalization system development researches, but also papers studying the relationship of learning styles and online behaviour. The result is a long list of potential attributes (over 80 items) which can be divided into three main sources including static data such as user background, ethnics, major etc., online behaviour e.g. time spent on certain activities and other personalization sources e.g. intelligence, memory capacity.

The data for different attributes is currently being programed and collecting for classification methods development using a learning system developed at Corvinno called STUDIO. Felder-Silverman's ILS survey has also been carried out as it is still the base line for online modelling evaluation that has been used in almost all of the previous papers. Over 250 undergraduate students are being observed with the plan of collecting data on the second group of students for model generalisation evaluation ability in the next school term in September.

Lastly, the classification methods development is still in the early stage. As most of previous researches still use single classification methods, I see an opportunity to apply more advanced techniques such as ensemble machine learning which combines different single algorithms to improve the performance. This branch of methods has shown to outperform single methods in other applications such as medical and finance.

4. FUTURE DIRECTION AND ADVICES SOUGHT

The research is still at the early stage and thus, there are a number of challenges ahead that I hope the consortium can provide advices on or sharing similar experiments and insights related to:

- Attributes comparison in the case with huge number of attributes and algorithms tested.
- While I will focus on ensemble and hybrid methods, I am also interested in if there is any other method, especially in the area of sequence mining.
- Generalisation: Is this necessary/possible to generalise the detection models? What are the conditions that we have to test for generalisation? Is testing on different populations enough?

5. ACKNOWLEDGMENT

This work is being carried out under the framework of Eduworks Initial Training Network, Marie Skłodowska-Curie actions (MSCA) FP7 of the European Commissions.

6. REFERENCES

[1] Akbulut, Y. and Cardak, C.S. 2012. Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011. *Computers & Education*. 58, 2 (2012), 835–842.

[2] D Baker, R.S. 2010. Mining data for student models. *Advances in intelligent tutoring systems*. Springer. 323–337.

[3] Bhattacharyya, S., Jha, S., Tharakunnel, K. and Westland, J.C. 2011. Data mining for credit card fraud: A comparative study. *Decision Support Systems*. 50, 3 (2011), 602–613.

[4] Coffield, F., Moseley, D., Hall, E. and Ecclestone, K. 2004. Learning styles and pedagogy in post-16 learning: A systematic and critical review. *London: Learning and Skills Research Centre.* (2004).

[5] Felder, R.M. and Silverman, L.K. 1988. Learning and teaching styles in engineering education. *Engineering education*. 78, 7 (1988), 674–681.

[6] García, P., Amandi, A., Schiaffino, S. and Campo, M. 2007. Evaluating Bayesian networks' precision for detecting students' learning styles. *Computers & Education*. 49, 3 (2007), 794–808.

[7] Graf, S., Kinshuk, K.D. and Liu, T.-C. 2008. Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour. *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on* (2008), 482–486.

[8] Graf, S., Kinshuk, K.D. and Liu, T.-C. 2009. Supporting Teachers in Identifying Students' Learning Styles in Learning Management Systems: An Automatic Student Modelling Approach. *Educational Technology & Society*. 12, 4 (2009), 3–14.

[9] Holodnaya, M.A. 2002. Cognitive styles: on the nature of individual mind. *Per Se, Moscow*. (2002).

[10] Honey, P. and Mumford, A. 1986. *Using your learning styles*. Peter Honey Maidenhead,, UK.

[11] Keefe, J.W. 1979. Learning style: An overview. *Student learning styles: Diagnosing and prescribing programs*. (1979), 1–17.

[12] Kolb, D.A. 1981. Learning styles and disciplinary differences. *The modern American college*. (1981), 232–255.

[13] Pashler, H., McDaniel, M., Rohrer, D. and Bjork, R. 2008. Learning styles concepts and evidence. *Psychological science in the public interest*. 9, 3 (2008), 105–119.

[14] Romero, C., López, M.-I., Luna, J.-M. and Ventura, S. 2013. Predicting students' final performance from participation in on-line discussion forums. *Computers & Education.* 68, (2013), 458–472.

[15] Scheuer, O. and McLaren, B.M. 2012. Educational data mining. *Encyclopedia of the Sciences of Learning*. Springer. 1075–1079.

[16] Stash, N., Cristea, A.I. and De Bra, P. 2006. Adaptation to learning styles in e-learning: Approach evaluation. (Honolulu, Hawaii, 2006).

[17] Thalmann, S. 2014. Adaptation criteria for the personalised delivery of learning materials: A multi-stage empirical investigation. *Australasian Journal of Educational Technology*. 30, 1 (2014).

[18] Truong, H.M. 2015. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in Human Behavior*. (2015), doi: 10.1016/j.chb.2015.02.014 .

[19] Wang, G., Hao, J., Ma, J. and Jiang, H. 2011. A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications*. 38, 1 (2011), 223–230.

[20] Wen, D., Graf, S., Lan, C.H., Anderson, T. and Kinshuk, K.D. 2007. Supporting web-based learning through adaptive assessment. *FormaMente Journal*. 2, 1-2 (2007), 45–79.

Modeling Speed-Accuracy Tradeoff in Adaptive System for Practicing Estimation

Juraj Nižnan
Masaryk University Brno
niznan@mail.muni.cz

ABSTRACT

Estimation is useful in situations where an exact answer is not as important as a quick answer that is good enough. A web-based adaptive system for practicing estimates is currently being developed. We propose a simple model for estimating student's latent skill of estimation. This model combines a continuous measure of correctness and response-times. The advantage of the model is its simple update method which makes it directly applicable in the developed adaptive system.

1. INTRODUCTION

Estimation is a very useful skill to possess. Particularly in situations where an exact answer is not as important as being able to quickly come up with an answer that is good enough (e.g., total amount on a bill in a restaurant, number of people in a room, total of the coins in a wallet, number of cans of paint needed for painting a room, converting between metric and imperial units). It was shown that estimation ability correlates with the ability to solve computational problems [2, 9, 8]. Because estimation is so useful, we have decided to develop a computerized adaptive system that will let its users practice estimating by solving various tasks.

The adaptive system will include exercises for practicing numerical estimation (results of basic arithmetic operations, converting between imperial and metric units, converting between temperature units, currencies and exchange rates) and visual estimation (counting the number of objects in a scene).

In order to provide adaptive behavior of the system, we need a way of inferring student's ability of estimation. In our setting, the binary-valued correctness-based modeling approach is not suitable. We do not expect the users to input exact responses, we expect them to input their best estimates. So our model should work with some measure of the quality of an answer. Another important point is the speed-accuracy tradeoff. Figure 1A shows a hypothetical tradeoff curve for one user with fixed ability. User can answer a task very quickly but it will probably be a very rough estimate. Or he/she can decide to spend more time on the task and respond with a more precise answer. Therefore, response-time should be a vital part of our model.

The system should be able to detect prior skill (i.e., how good the user was at estimation before he started using the system) which can be deduced from the first interactions of the user with the system. The goal of the developed system is to enable the user to get better at estimating. Therefore, the proposed model should also take into account user's improvement (or learning) over time. Figure 1B illustrates answers of several users on one task as red dots. Ideally, the system will help its users to learn to perform near the green mark, to be fast and accurate.

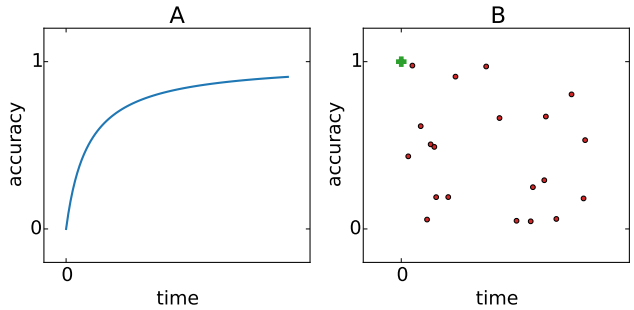

Figure 1: A) hypothetical speed-accuracy tradeoff curve, B) goal of the system

The value of the system will also be in the data that will be collected. It can be used to answer some interesting research questions. Does the speed-accuracy tradeoff curve have the same shape for converting between EUR and USD as for estimating the number of displayed objects? How do the learning curves look? Can estimation tasks in one area be learned more quickly than in another area? How close to the perfect mark can users push their performance? What is the influence of a countdown timer on user's performance? What is the appropriate level of challenge that motivates the users? The last question was addressed in [3], where the authors were trying to validate the *Inverted-U Hypothesis* (i.e., we most enjoy challenges that are neither too easy, neither too hard) on data collected from online estimation game called *Battleship Numberline*. They found out that the

easier the game was, the longer users played the game.

2. MODELS

In this section, we present a few existing models for combining correctness and response-times in Item Response Theory (IRT) and a model for tracking learning currently used in our other adaptive practice system. We then propose a simple model that could be used in the system for practicing estimates. The described models use a logistic function $\sigma(z) = (1 + e^{-z})^{-1}$. Users of the system (or students) are indexed by j. The items (or tasks, problems, questions) that the users solve are indexed by i.

2.1 Models from IRT

A typical example of an approach to the modeling of both correctness and response-times in Item Response Theory is from van der Linden [10]. The approach uses two models, one for correctness (binary) and the other one for response-times (distributed lognormally). The probability of success of a student j on item i can expressed by the 3PL model:

$$p_{ij} = c_i + (1 - c_i) \cdot \sigma(a_i(\theta_j - b_i))$$

where parameter θ_j is the skill of student j and a_i, b_i, c_i are the discrimination, difficulty and pseudo-guessing parameters for the item i. The logarithm of a response-time t_{ij} can be predicted by:

$$\ln \hat{t}_{ij} = \beta_i - \tau_j \tag{1}$$

where β_i represents the amount of labor required to solve item i and τ_j the speed of student j. The disadvantage of this model is that it does not model the speed-accuracy tradeoff explicitly.

An example of a model that directly combines binary correctness with response-time is Roskam's model [7]:

$$p_{ij} = \sigma(\theta_j + \ln t_{ij} - b_i)$$

Here, an increase in item difficulty (or decrease in student's ability) can be always compensated by spending more time on a problem. This tradeoff is called an increasing conditional accuracy function.

2.2 Model for factual knowledge

Here, we present a model that is currently used in a popular adaptive system for practicing geographical facts [4]. This model consists of two parts, one (Elo) estimates the prior knowledge of a student and the second one (PFAE) models student learning. A big advantage of this model is that it uses fast online methods of parameter estimation which makes it suitable for use in an interactive adaptive practice system.

The prior knowledge of a student is modeled by the Rasch (1PL) model. The probability that a student j answers item i correctly is modeled by the likelihood $p_{ij} = \sigma(\theta_j - b_i)$. The parameters are estimated using Elo rating system [1]. Elo was originally developed for rating chess players, but the process of student answering an item can be interpreted as a "match" between the student and the item. After each "match", the parameters are updated as follows:

$$\begin{aligned} \theta_j &:= \theta_j + U(n_j) \cdot (correct - p_{ij}) \\ b_i &:= b_i + U(n_i) \cdot (p_{ij} - correct) \end{aligned}$$

where $U(n)$ is the uncertainty function $U(n) = \frac{\alpha}{1+\beta n}$ and n is the number of updates of the parameter and α and β are metaparameters. The variable *correct* takes value 1 if the student has answered correctly and value 0 otherwise. This model is used for predicting– and trained on–first responses.

After the first interaction of a student j with item i has been observed, we can set student's skill in that particular item to $\theta_{ij} = \theta_j - b_i$. An extended version of Performance Factors Analysis [5] called PFAE is used to model learning and predicting the following interactions of the student with the item. Likelihood of a correct answer is $p_{ij} = \sigma(\theta_{ij})$. The update to student's knowledge of item θ_{ij} after observation is:

$$\theta_{ij} := \begin{cases} \theta_{ij} + \gamma \cdot (1 - p_{ij}) & \text{if the answer was correct} \\ \theta_{ij} + \delta \cdot p_{ij} & \text{if the answer was incorrect} \end{cases}$$

where γ and δ are metaparemeters. The reason for two different metaparameters is that the student learns also during an incorrect response.

2.3 Proposed model for estimates

Here, we propose a model that can be used in the adaptive practice system for estimates. The model combines Roskam's model and the update scheme from Elo and PFAE.

A simple extension of the correctness-based modeling to the setting of practicing estimates is to use a measure of correctness, or a *score* – a rational number ranging from 0 to 1. The way of scoring of an answer could be based on the domain being practiced by the user. For example, for the scenario where the user is estimating the number of objects in a scene, the exact answer would get a score of 1, deviating by one object a score of 0.8, etc.

The model assumes the same parameters and relationship as Roskam's model, but instead of expressing a probability of a correct answer it specifies the expected score:

$$\hat{s_{ij}} = \sigma(\theta_j + \ln t_{ij} - b_i)$$

Figure 2 shows how the score changes as a function of time for different values of user's skill θ_j (with fixed $b_i = 0$). It nicely demonstrates the speed-accuracy tradeoff.

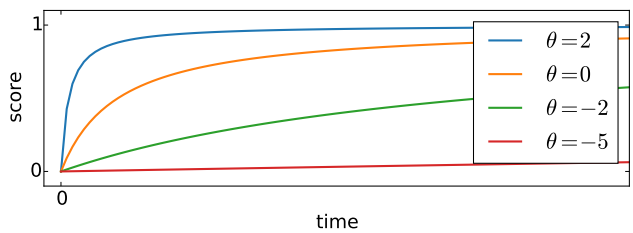

Figure 2: Score function for different values of skill

After observing score s_{ij} that user j obtained for answering item i and response-time t_{ij}, we can update model's beliefs

in the parameters:

$$\begin{aligned} \theta_j &:= \begin{cases} \theta_j + \gamma \cdot (s_{ij} - \hat{s_{ij}}) & \text{if } s_{ij} \geq \hat{s_{ij}} \\ \theta_j + \delta \cdot (\hat{s_{ij}} - s_{ij}) & \text{if } s_{ij} < \hat{s_{ij}} \end{cases} \\ b_i &:= b_i + U(n_i) \cdot (\hat{s_{ij}} - s_{ij}) \end{aligned}$$

Note, that the model uses a single parameter θ_j for the student. This is different from the approach taken in PFAE, where the student has a parameter for each item θ_{ij}. While that approach is suitable for modeling the knowledge of facts – where it is reasonable to assume that the knowledge of one fact is independent of the knowledge of another – it is not suitable here. Student's ability to convert 2 miles to kilometers is surely dependent on his ability to convert 3 miles to kilometers.

We propose using separate model for each concept (e.g., estimating the number of objects, conversion lb to kg, conversion EUR to USD). It is true that student's ability to estimate items corresponding to one concept tells us something about his ability to estimate the other concepts. However, if the user does not know the conversion rate from EUR to USD then being able to estimate well the other concepts will not help him.

The model can be easily extended by adding a discrimination parameter a or a guessing parameter c (similarly to the IRT model): $\hat{s_{ij}} = c + (1-c) \cdot \sigma(a(\theta_j + \ln t_{ij} - b_i))$. These added parameters could be either metaparameters of the model or parameters of the item i. The guessing parameter may be useful for the scenario where the user has to select a value on a numberline.

As we mentioned earlier, this model suffers from the issue that increasing the time spent on an item increases the expected score. This may hold true for the instance where the user knows the underlying concept (e.g., the conversion rate from EUR to USD) but it does not hold when he does not know it. But the model uses the logarithm of response-time and the time a student is willing to spend on an item is limited. Therefore, the model should have reasonable behavior for the time interval of interest, as is demonstrated in Figure 2 by the curve corresponding to $\theta_j = -5$.

3. DISCUSSION

The model works with the response-time as a parameter. Therefore, it cannot be used for predicting response-times directly. A model similar to (1) can be used for that. Predicted time and score can be used for item selection (i.e., which item to offer the user next). This can be done by setting a target score and recommending an item with predicted score close to the target.

Does the model perform better than a simple 1PL model that does not use response-times at all? Does it make sense to add more parameters to the model? How does the model fare against more complicated models? To be able to answer these questions, we need to somehow evaluate the performance of the model. The choice of metric is interesting because a model can predict both score and response-time. When considering only the predicted score, a standard metric like RMSE can be used [6]. When we have a measure of performance, we can explore if the model is well-calibrated with respect to response-times or if the model works similarly well for all the domains (concepts).

Other question that we could ask is how well does the speed-accuracy tradeoff curve that the model assumes correspond to reality.

Acknowledgements

We thank Radek Pelánek (for guidance and useful suggestions) and Roman Orlíček (for actively working on developing the application).

4. REFERENCES

[1] A. E. Elo. *The rating of chessplayers, past and present*, volume 3. Batsford London, 1978.

[2] S. A. Hanson and T. P. Hogan. Computational estimation skill of college students. *Journal for Research in Mathematics Education*, pages 483–499, 2000.

[3] D. Lomas, K. Patel, J. L. Forlizzi, and K. R. Koedinger. Optimizing challenge in an educational game using large-scale design experiments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 89–98. ACM, 2013.

[4] J. Papoušek, R. Pelánek, and V. Stanislav. Adaptive practice of facts in domains with varied prior knowledge. In *Educational Data Mining (EDM)*, pages 6–13, 2014.

[5] P. I. Pavlik, H. Cen, and K. R. Koedinger. Performance factors analysis-a new alternative to knowledge tracing. In *Proc. of Artificial Intelligence in Education (AIED)*, volume 200 of *Frontiers in Artificial Intelligence and Applications*, pages 531–538. IOS Press, 2009.

[6] R. Pelánek. A brief overview of metrics for evaluation of student models. In *Approaching Twenty Years of Knowledge Tracing Workshop*, 2014.

[7] E. Roskam. Toward a psychometric theory of intelligence. *Progress in mathematical psychology*, 1:151–174, 1987.

[8] P. M. Seethaler and L. S. Fuchs. The cognitive correlates of computational estimation skill among third-grade students. *Learning Disabilities Research & Practice*, 21(4):233–243, 2006.

[9] R. S. Siegler, C. A. Thompson, and M. Schneider. An integrated theory of whole number and fractions development. *Cognitive psychology*, 62(4):273–296, 2011.

[10] W. J. Van Der Linden. Conceptual issues in response-time modeling. *Journal of Educational Measurement*, 46(3):247–272, 2009.

Reimagining Khan Analytics for Student Coaches

Jim Cunningham
Arizona State University
Tempe, AZ
1-480-703-7394
jim.cunningham@asu.edu

ABSTRACT
In this paper, I describe preliminary work on a new research project in learning analytics at Arizona State University. In conjunction with an innovative remedial mathematics course using Khan Academy and student coaches, this study seeks to measure the effectiveness of visualized data in assisting student coaches as they help remedial math students achieve success in an online math class.

Keywords
Learning analytics, data visualization, remedial mathematics.

1. INTRODUCTION
With 77,000 students, Arizona State University has become one of the largest institution of higher learning in the United States second only to the University of Phoenix. A certain slice of newly enrolled students at Arizona State find themselves in a dilemma, they have been admitted to the university, but they have failed to meet the minimum math requirement that would allow them to start taking undergraduate classes.

These students are desirable for many reasons. They can help the university meet goals for diversity and social justice. Often these students are first generation college students. However, this population also poses challenges to the university. Being unable to meet the minimum score on the mathematics placement test not only points to gaps in a student's math background, it often points to larger issues of academic readiness.

Overcoming the hurdle of meeting the minimum math requirement has been challenging. Online remedial math classes using an adaptive learning model have historically had a pass rate of around 50%. These pass rates have remained stubbornly low despite various efforts to improve them.

2. A NEW APPROACH
In the summer of 2014, EdPlus (ASU's online education arm) decided to launch a new version of this remedial math class built around Khan Academy and undergraduate peer coaches. One of the big reasons for making this change was data. Khan, because it is a non-profit, was open to sharing the data generated by students and KA's strategies for student success. Arizona State wanted their remedial math classes to become their most data driven offering. Working with Khan had other advantages as well. Because KA has over 10 million unique users every month and over 2 billion problems worked, Khan is able to deploy and adapt its instruction at scale.

3. CHALLENGES WITH KHAN
While the strategy of building a remedial math class around Khan Academy had many strengths, there were also significant challenges. The first challenge was in the form of a problem that many schools face when working with Khan Academy. While seeking to build a comprehensive universe of math instruction, KA has developed explicit pathways to math success. Khan calls these pathways "missions." However, ASU's end goal: passing an exam that is meant to reflect the many math concepts that a student should know before entering college, cut across many Khan missions. In addition, Khan's powerful analytical tools that are meant to aid instructors in following a student's progress are tied to these missions. When math skills are being served up to students *a la carte*, as they are to accomplish ASU's remedial math program, the analytics are unraveled.

A second major challenge facing the Khan Academy program was the same challenge facing the original remedial math program. Many of these students were failing to pass the minimum math requirement to enter Arizona State University because school in general has been challenging for them for a very long time. Putting these students by themselves in an online math course of any kind could be a recipe for failure. They needed additional support. The kind of personal attention these students need is very expensive. ASU decided to control that cost by employing a system of student "coaches." Coming from a variety of majors and backgrounds, these student coaches were handpicked and given responsibility for 20-25 online students each. Their job was to monitor, guide, tutor, and encourage these students to the end goal of having their coachees pass an exam that was meant to reflect their readiness to take college level math.

In order to be effective, these coaches needed access to the data in Khan Academy about their students' progress, but because ASU's exam at the end of the remedial math course measured math skills that spanned several Khan math missions, the state-of-the-art Khan analytics that are tied to those missions were unavailable to the coaches. After a lot of work, (much of it spearheaded by the student coaches themselves) a spreadsheet was developed that was populated by weekly downloads of Khan data. It showed which math skills were practiced and which skills were mastered and matched these up to a rough metric that told the coaches whether their students were on track to successfully master all the math skills they needed before they had to take the exam.

4. RESEACH GOALS
The goal of my research is create custom data visualizations that fit ASU's mission for this remedial math class and then measure the effectiveness of these analytics in assisting the student coaches in their work of creating student success. These analytics are specifically aimed at enabling the student coaches to visualize the

large amounts of data generated through Khan Academy. Khan stores data on each student's attempt to solve a math problem related to a particular skill. There is also data on how many times a student views a Khan video on a math concept or asks for hints when attempting a math problem. Because we are re-envisioning the analytics from the ground up, we have an opportunity to create analytics that are similar to the ones that Khan has created for its missions yet improve these analytics for ASU's specific purposes and create dashboards that visualize the data in other ways that may be even more useful for the student coaches. Because the coaches only had access to Khan data through a spreadsheet the first semester the new remedial math class was taught, there is an opportunity to compare the success of coaches assisting their students with the spreadsheet data versus those using the more sophisticated data visualizations produced from the student's actions within Khan Academy.

In order to achieve this goal, ASU is teaming up with Blue Canary, a learning analytics company headquartered in Chandler, Arizona. Blue Canary and I are working directly with Khan Academy to address data flow issues including creating API's that will automatically access data from Khan databases that will be feeding the dashboards and graphics created for the student coaches. I am also going to be working with Blue Canary to create dashboards and data visualization tools for the Khan online math class in Tableau. These dashboards are directly aimed at assisting student coaches while they help their math coachees achieve success.

5. RESEARCH QUESTIONS

Once the dashboards are created and the student coaches start using them to assist their math students, we can start to address this research question: Do data visualization tools enable student coaches to better assist remedial math students entering Arizona State University achieve success?

6. RESEARCH DESIGN

The preliminary design of this study is to compare data generated by two cohorts of remedial math students. The first cohort has been guided by student coaches who have been accessing the Khan data on through a spreadsheet created to keep track of skill practice and mastery. The second cohort will be guided by student coaches who have access to the data visualization tools and dashboards created by myself and Blue Canary in Tableau. The ultimate measure of coach success will be the pass rate of their students at the end of the course. In addition, there will be many other metrics to measure, as well, such as student engagement and persistence.

This research is in the early stages. Mike Sharkey from Blue Canary and myself have been meeting with student coaches and instructional designers of the remedial math program to assess the needs of the student coaches and talk about possible data visualizations that may be helpful. API's are being designed pull data from Khan for the analytics and dashboard layouts. While we are working on this, data is being generated by students in Khan Academy who are working with coaches that are relying on the spreadsheet to access data about the progress of their students in Khan. I am currently in second year of a four year PhD program, so we have some time to make adjustments and work out problems as they arise.

Data Analysis Tools and Methods for Improving the Interaction Design in e-Learning

Paul Stefan Popescu
University of Craiova
Department of Computer Science
Craiova, Romania
+40724571133
sppopescu@gmail.com

ABSTRACT
In this digital era, learning from data gathered from different software systems may have a great impact on the quality of the interaction experience. There are two main directions that come to enhance this emerging research domain, Intelligent Data Analysis (IDA) and Human Computer Interaction (HCI). HCI specific research methodologies can be used to present the user what IDA brings after learning and analyzing user‘s behavior. This research plan aims to investigate how techniques and mechanisms available in both research areas can be used in order to improve learners' experiences and overall effectiveness of the e-Learning environment. The foreseen contributions relate to three levels. First is the design and implementation of new algorithms for IDA. The next level is related to design and implementation of a generic leaning analytic engine that can accommodate educational data in attempt to model data (i.e., users, assets, etc.) and provide input for the presentation layer. Last and top level is represented by the presentation layer where the output of the underlying levels adapts the user interface for students and professors.

Keywords:

Learning analytics, intelligent data analysis, interaction design, user modeling

1. INTRODUCTION
Standard books or their digital versions (eBooks) or standard e-Learning environments are usually just a simple presenting method of the learning material. In this digital era our day by day devices must became proactive to our needs, i.e. they have to know what we need before we even have to ask them. Considering the field of e-Learning, in order to find user's needs and to improve his learning experience we can log various activity related data as a first step in a data driven analytic engine. These actions may define learners' behavior in e-Learning environments providing IDA with raw data to be analyzed. Based on this data IDA creates a data model which is based on user's performed actions. A sample output of the IDA process may be represented by a user model that is aimed to directly influence the user interface.

Learning using on-line educational environments is getting more and more popular but the effectiveness of interaction between students or students and professors is usually poorer than the interaction in physical educational environments. Improving the interaction design process in e-Learning platforms may have a direct impact on the effectiveness of the learning and be achieved by following a data driven approach. The proposed approach is related to several prerequisites and the learning resource that needs to be well structured and presented. Others are related to the interaction between students and the links that can be created between them, proper data visualization techniques, interpretation of results, adequate data analysis processes with specific goals regarding interface adaptation.

2. RELATED RESEARCH IN I.D.A.
Learning analytics and Machine Learning[2] is still one of the most interesting parts of the IDA research area. One research area of this domain is related to the classification procedures. Some of them are related to the usage of classification on text[1] and some of them are regarding to usage of classification as an user analyzing method[4].

Analysis of students' activities in the online educational systems with the goal of improving their skills and experience through the learning process has been an important area of research in educational data mining. Most of the techniques are trying to predict student's performances[5,6,7,12] based on their actions.

The work in this domain started in the year of 2005 with a workshop referred to as 'Educational Data Mining' AAAI'05-EDM in Pittsburg, USA[8] which was followed by several related workshops and the establishment of an annual international conference first held in 2008 in Montreal[9] . Before of EDM, user modeling domain was the one that was encapsulating this research area.

Several papers, journals and surveys have been written but only two books were published: the first is "Data mining in E-learning"[10] which has 17 chapters oriented to Web-based educational environments and the second is "Handbook of Educational Data Mining"[11] which has 36 chapters about different types of educational settings.

In this research proposal the goal is to combine HCI with IDA and educational research in order to improve the learners experience in digital educational environments. This domain is also related to Intelligent Interfaces research area.

3. RESEARCH AND DEVELOPMENT STATUS
As research status two papers have been written so far.
I am a co-author of the paper Advanced Messaging System for On-Line Educational Environments[3]. This paper presents a method of using a classification procedure for retrieving a set of recommended messages that might be interesting to students.

The second paper is entitled „Building an Advanced Dense Classifier"[4], which has already been published at IDAIR 2014 and won the best paper award. This paper presents a classifier that implements several extra functionalities which can lead to better results. Its goal is to build a Decision Tree classifier that accommodates data (instances). This new data structure extends the functionality of a Decision Tree and is called DenseJ48. This new classifier implements efficiently several extra functionalities besides the core ones that may be used when dealing with data.

Based on this paper, as development background a Weka package which implements the classifier's functionalities is under development. I am also a contributor (http://apps.software.ucv.ro/Tesys/pages/development.php) of Tesys[13], an e-Learning platform used in several faculties from Craiova, mainly focusing on the eLeTK (e-Learning Enhancer Toolkit)[14] module. This is how I found out about Intelligent Data Analysis and Information Retrieval, and the benefits these research areas can bring to the online educational environments.

As relevant training in September 2013 I applied for and obtained a scholarship for attending the 9th European Summer School in Information Retrieval, which took place in Granada, Spain. Being part of this event helped me improve my knowledge in the domain of Information Retrieval – the presentations covered most of this research area, from basics to evaluation techniques and Natural Language Processing. Later I attended Research Methods in Human-Computer Interaction between 25th and 31th of July 2014 in Tallinn, Estonia.(http://idlab.tlu.ee/rmhci) in order to deepen my knowledge of HCI research methodologies.

4. RESEARCH PROBLEMS FROM PHD PROPOSAL

Problems related to this research can be structured in a three layer representation. There is a certain need for improving the interaction between the users (students, professors, etc.) and the system that provide them the learning experience. The research problems are related to closing the gap between classical and digital learning paradigms.

Development of new tools is fundamentally based on functionality provided by a generic learning analytic engine, among which there are: generic representation of learning analytics data of users, integration of various implementations of IDA algorithms, custom integration of interaction design process artifacts. All these three layers build up a learning analytics engine that is designed to run as a service along e-Learning environments in an attempt to improve the quality of the on-line educational system.

4.1 Layers description

4.1.1 Data Representation Layer

First layer is related to the representation of the raw data that can be gathered from the log files and the database. Our desire is to find what data (features, parameters, ranges, etc) is relevant for online learning environments. Based on this data we have to extract features that can define learning resources or those features that enable us to obtain a user representation.

4.1.2 Learning Analytics Layer

Based on the data gathered it is possible to employ different IDA algorithms in order to obtain custom built data pipelines. Experimenting at this level with different algorithms and different feature sets can lead to obtaining output information for solving different problems. Data aggregation and pipelining are the mainly used processes. The purpose of this layer is to offer to the next one data in a structured format which can be presented on the interface.

4.1.3 Presentation Layer

The presentation of the learning material is very important, leaving a mark on the mental model created by the learning resources. In this layer the HCI component of this proposal is employed.

Taking into consideration these aspects related to both domains we can say that there is a need for new tools that could be integrated within the digital learning environments in order to provide an improved learning experience that fulfills the user's needs.

4.2 Research questions & Proposed Approach

The questions that have to be addressed when we talk about research in e-Learning environments are related to the main actors that are using the on-line educational environments. Therefore, learners, teachers and administrators (which can do the data analyst job), by the generic meaning, are the ones we focus on because they are the main users of these systems. Secretaries of the learning environments only concur to configure the e-Learning environment.

The presented questions are from the business goal perspective. Answering these questions needs a close discussion about the presented underlying levels, which are the same regardless of the tackled issue, that define data driven process.

- *How IDA can be efficiently and effectively used for an on-line educational context?*

Proper usage and integration of IDA techniques can create a framework which data analysts and developers can employ for further work.

- *How can e-Learning resources be managed/aggregated in an IDA context?*

There are various types of resources that exist in on-line educational environments. Depending on how they are managed and aggregated, application developers can benefit from them.

- *Which are the common (general purpose) functionalities when dealing with educational data pipelines?*

Several functionalities exist in dealing with data but not all of them are feasible for working with educational data. In this particular case we need to find the most effective ones and adapt them to this particular case.

- *How can the student know his place among his colleagues and be motivated to study harder?*

This question is highly important from the student's perspective. Without knowing his place among his colleagues and

without having an explicit learning path, the learner will not have the indication of his final result and will not have the motivation to maximize his potential. In e-Learning environments, students do not participate together in courses, like in a regular environment, so they are unaware of their colleagues' knowledge level. In a traditional classroom, there is always a certain level of competitiveness, so each student is constantly motivated to improve himself. Therefore, an important goal is to achieve a similar scenario in the online educational environments, although it is not the only one. Besides being competitive, the students must also be engaged in helping others and in turn receive help when they are having difficulties understanding something.

- *How can the professors know where exactly do the students have problems, so they can adapt the course material?*

From the professor's point of view, being aware of his students' progress and the difficulties they encounter in understanding the material is possibly the most important requirement. Although each student is different and has his own learning curve, common points can be found and an overall perception can be formed. The professor must be able to build a mental model regarding the overall performance of his students. By doing so, he can modify and perfect in time the content of the course. Also, taking into consideration the fact that the difficulty level of the final evaluation must be consistent with the students' level of understanding of the course, the professor needs to be aware of that level so he can make the proper adjustments.

- *Which data should be logged in order to extract relevant information about the students?*

Any e-Learning environment whose goal is to integrate an intelligent component should be able to log the necessary data and extract the values of the features. Logging the needed data is a prerequisite to the data analysis process. Logging too much can create a useless load of the server but logging not enough will make impossible the features extraction.

Features are very important in IDA because they define the entity that will be analyzed. Choosing the right features are crucial in different IDA processes. A comprehensive list of features (with proper data types, range values and significance) should be available for further analysis.

5. CLOSING REMARKS

On-line educational environments are here from a long enough time. This aspect brings in front of the scientists many opportunities for improving the learning process and to lower the distance from the classical educational environments to the online ones. Many research areas concur to improve the learning process but the most relevant are the user centered ones.

There are 3 different research areas that concur to bring learners several improvements. IDA is the first one bringing data mining and machine learning algorithms and generate user models, followed by HCI, which is used to optimize the interfaces and create friendly interaction environments and finally the Educational research area is where we put in practice this work.

6. REFERENCES

[1] Vitomir Kovanovic, Srecko Joksimovic, Dragan Gasevic, Marek Hatala .:Automated Cognitive Presence Detection in Online Discussion Transcripts, LAK Workshops, 2014.

[2] Dragan Gasevic, Carolyn Penstein Rosé, George Siemens, Annika Wolff, Zdenek Zdráhal .: Learning analytics and machine learning, LAK Conference, 2014..

[3] Mihai Mocanu, Paul-Stefan Popescu, Dumitru Dan Burdescu, Marian Cristian Mihaescu,: Advanced Messaging System for On-Line Educational Environments Sesimbra, Portugal, 26-28 June 2013.

[4] Paul Stefan Popescu, Cristian Mihaescu, Mihai Mocanu, Dumitru Dan Burdescu, Building an Advanced Dense Classifier, 5th International Conference on Information, Intelligence, Systems and Applications, Chania, Crete, July 2014.

[5] Minaei-Bidgoli, B., Kashy, D.A., Kortemeyer, G.: Predicting student performance: an application of data mining methods with an educational Web-based system. In: Frontiers in Education, 2003. FIE 2003 33rd Annual (Volume:1), pp. T2A--13. Westminster (2003).

[6] Baradwaj, B.K., Pal, S.: Mining Educational Data to Analyze Students' Performance. (IJACSA) International Journal of Advanced Computer Science and Applications, Volume 2, No. 6, 63--69 (2011).

[7] Márquez-Vera, C., Cano, A., Romero, C., Ventura, S.: Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Applied Intelligence, Volume 38, Issue 3, 315--330 (2013)

[8] Beck, J, Proceedings of AAAI2005 workshop on Educational Data Mining, (2005).

[9] Baker, R.S.J.d., Barnes, T., Beck, J.E. (Eds.) Educational Data Mining 2008: 1st International Conference on Educational Data Mining, Proceedings. Montreal, Quebec, Canada. June 20-21, 2008.

[10] C. Romero and S. Ventura, Data Mining in E-Learning, WIT Press, 2006

[11] C. Romero and S. Ventura, Handbook of Educational Data Mining, CRC Press, 2010

[12] B. Minaei-Bidgoli, D. A. Kashy, G. Kortemeyer and, W. F. Punch, "Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA" In Proceedings of ASEE/IEEE Frontiers in Education Conference, Boulder, CO: IEEE, 2003..

[13] Marian Cristian Mihaescu: Software Architectures of Applications Used for Enhancing On-Line Educational Environments, International Journal of Computer Science and Applications (IJCSA), vol. 10, ISSUE 1 MARCH 2013

[14] Burdescu, D.D., Mihaescu ,M.C.: Tesys: e-Learning Application Built on a Web Platform. In: Proceedings of International Joint Conference on e-Business and Telecommunications, pp. 315-318. Setubal, Portugal (2006)

Assessing the Roles of Student Engagement and Academic Emotions within Middle School Computer-Based Learning in College-Going Pathways

Maria Ofelia Z. San Pedro
Teachers College Columbia University
525 W 120th St.
New York, NY 10027
mzs2106@tc.columbia.edu

ABSTRACT

This dissertation research focuses on assessing student behavior, academic emotions, and knowledge from a middle school online learning environment, and analyzing their potential effects on decisions about going to college. Using students' longitudinal data ranging from their middle school, to high school, to postsecondary years, I leverage quantitative methodologies to investigate antecedents to college-going outcomes that can occur as early as middle school. The research first looks at whether assessments of learning, emotions and engagement from middle school computer-based curriculum are predictive at all of college-going outcomes years later. I then investigate how these middle school factors can be associated with college-going interests formed in high school, using the same assessments during middle school, together with self-report measures of interests in college when they were in high school. My dissertation then culminates in developing an overall model that examines how student interests in high school can possibly mediate between the educational experiences students have during middle school technology-enhanced learning and their eventual college-going choices. This gives a richer picture of the cognitive and motivational mechanisms that students experience throughout varied phases in their years in school.

Keywords

College Choices, Academic Emotions, Behavior, Knowledge, Social Cognitive Career Theory, Interests

1. Introduction

College enrollment and completion are key steps towards career success for many learners. However, well before this point, many students effectively drop out of the pipeline towards college quite early. According to Social Cognitive Career Theory (SCCT) [10], academic and career choices are shaped throughout middle school and high school by environment supports and barriers, where higher levels of interest emerge within contexts in which the individual has higher self-efficacy and outcome expectations, and these interests lead to the development of intentions or goals for further exposure and engagement with the activity [10]. Traditional studies also show that family background, financial resources, and prior family academic achievement have significant impacts on where students find themselves after high school. All of these factors, however, are fairly strong displays of disengagement. By the time these indicators are commonplace, students may be in such a precarious situation that many interventions may fail. In general, current models about successful access to postsecondary education may be insufficient to help educators identify which students are on track and which need further support [11]. Fine-grained assessments of student behaviors and academic emotions (emotions that students experience during learning and classroom instruction) have been found to influence learning outcomes [12, 13]. Hence, there is an argument to be made that engagement and academic emotions in middle school play an essential early role in the processes described in SCCT. In SCCT, students' initial vocational interests are modified by their self-efficacy, attitudes, and goals towards career development (i.e. college enrollment, career interest), which are themselves influenced by the student's learning and engagement when encountering the increasingly challenging content in middle school [1, 12] – as poor learning reduces self-efficacy whereas successful learning increases self-efficacy [cf. 2]. As such, student academic emotions, learning, and engagement during middle school may be indicative of their developing interests in career domains which may in turn influence their choice to attend college [6, 9].

For the reasons aforementioned, my research attempts to answer Bowers' [5] call to identify much early, less acute signals of disengagement, the sort that occur when students' engagement is still malleable enough for interventions to succeed. Specifically, I investigate antecedents to college attendance that occur during middle school, using assessments of engagement and disengagement to better understand how these factors interact so that I can develop possible paths to re-engagement before students develop more serious academic problems. The models I create and the analyses I conduct involve the context of an online learning environment, and hence, this work provides both a new perspective on the efficacy of the system and an opportunity to explore how the system and its data can be used to predict long-term educational outcomes – in the case of my dissertation research, intervention and support in keeping students on track towards the pathway to college.

2. Data and Related Methodologies

My dissertation leverages data acquired from both traditional research methods as well as methodologies from machine learning and student modeling in assessing the constructs I analyze in my data, which I then use in developing the outcome models I propose. For middle school measures, I use the ASSISTment system (ASSISTments) as my source for middle school interaction data, and assessed measures of student knowledge, academic emotions, and behavior by using individual models developed to infer them. ASSISTments is a free web-based tutoring system for middle school mathematics that assesses a student's knowledge while assisting them in learning, providing teachers with detailed reports on the skills each student knows [14]. Interaction data from the ASSISTment system were obtained for a population of middle school students who used the system at various school years, from 2004-2005 to 2008-2009. These students are drawn from urban and suburban districts who used the ASSISTment system systematically during the year. I assessed

a range of constructs from interaction data in ASSISTments, which include student knowledge estimates, student academic emotions (boredom, engaged concentration, confusion), student disengaged behaviors (off-task, gaming the system, carelessness), and other information of student usage. These form the features in our final model of college-going outcomes. Aside from educational software data, I also use survey data from the same students who used the system in middle school, consisting of information about the their attitude about the subject (mathematics) and about the system itself. These survey data were acquired around the same time they used the software in middle school.

For my high school measures of interest, students who used the system during their middle school years and who are now in high school, were administered with two surveys: the first is a short questionnaire that asked the highest level of math and science courses that the student completed in high school and asks the student what his/her educational and career plans are upon graduation. The second survey is the an CAPA survey, designed by Fred Borgen and Nancy Betz [4]. It is an online survey with Likert scale inputs from students that gauges their interest and confidence on certain domains and skills, and then assesses their overall self-efficacy and vocational interests using existing instruments.

A subset of our student sample who were expected to be in postsecondary stage of education by the time of data collection were identified for their postsecondary education status. For their college enrollment information, records were requested from the National Student Clearinghouse, with information such as whether they were enrolled in a college or not, the name of the university, date of enrollment, and college major enlisted if available. We supplemented this data with college selectivity classification of the said postsecondary institutions, taken from the Barron's College Selectivity Rating which classifies colleges into ten categories [7, 16], from most selective or 'Most Competitive' to 'Special' which consist of specialty institutions such as schools of music, culinary schools, art schools, etc. Another source of data includes survey data about post-high school academic and career achievements that was administered to this subset of students.

3. Preliminary Work

In developing an overall integrated model, I initially tested the predictive power of the middle school factors on separate postsecondary outcomes. First, I applied fine-grained models of student knowledge, student academic emotions (boredom, engaged concentration, confusion, frustration) and behavior on middle school interaction data to understand how student learning and engagement during this phase of learning can predict college enrollment. A logistic regression model was developed and can distinguish a student who will enroll in college (68.6% of the time, an above average performance for models created from "discovery with models"). In particular, boredom, confusion, and slip/carelessness are significant predictors of college enrollment both by themselves and contribute to the overall model of college enrollment. The relationships seen between boredom and college enrollment, and gaming the system and college enrollment indicate that relatively weak indicators of disengagement are associated with lower probability of college enrollment. Success within middle school mathematics is positively associated with college enrollment, a finding that aligns with studies that conceptualize high performance as a sign of college readiness [15] and models that suggest that developing aptitude predicts college attendance [8].

Next, I also modeled whether students will attend a selective college, combining data from students who used the ASSISTment system with data on college enrollment, and ratings from Barron's on college selectivity. These were used to model another logistic regression model that could distinguish between a student who will attend a selective college and a student who will not attend a selective college 76% of the time when applied to data from new students. This model indicated that the following factors are associated with lower chance of attending a selective college: gaming the system, boredom, confusion, frustration, less engaged concentration, lower knowledge, and carelessness.

I finally looked at college major classification based on middle school student learning and engagement, specifically whether the major belonged to a STEM (Science, Technology, Engineering, Mathematics) or Non-STEM category. The logistic regression model developed could distinguish between a student who took a STEM college major and a student who took an non-STEM college major 66% of the time when applied to data from new students This model indicated that the following factors are associated with lower chance of enrolling in a STEM college major: gaming the system, lower knowledge, and carelessness.

4. Proposed Work

The initial individual models above support existing theories about indicators of successful entry to postsecondary education (academic achievement, grades). It sheds light on behavioral factors a student may experience in classrooms – which are more frequently and in many ways more actionable than the behaviors which result in disciplinary referrals – and how they can be predictive and be associated with long-term student outcomes.

With middle school assessments, I investigate at how student learning, academic emotions, and behavior as early as middle school may contribute as causal factors to a particular postsecondary decision (a in Figure 1 below) – an individual choice that is composed of answering the following questions: 1) Does the student decide to attend college?; 2) Does the student attend a selective college?; 3) What type of major does the student enroll in? I employ multivariate analysis on this part of my research work, for a richer and more realistic view of our postsecondary outcome, which is more than just one dependent variable of interest. Also this type of analysis allows us for causality to be deduced, as well as the inherent or underlying structure that can describe the data in a simpler fashion – in terms of latent variables. I also investigate interaction of features and how it affects our multivariate model via logistic regression, factor analysis and other appropriate statistical and machine learning algorithms that can be employed in our data to further understand the research problem.

In this phase of my dissertation research, I am starting to test the hypothesis of the possible existence of a mediating or indirect effect of high school college (and career) interests in predicting the multivariate postsecondary outcome based on middle school factors. I will establish this by looking at the causal influence of middle school factors to high school data (b in Figure 1 below). By integrating student data of their previous middle school interaction data, interests during their high school years, up to their postsecondary information, I will look at the possible causality of middle school factors to high school factors, as well as causality of high school factors to their postsecondary information. Like in previous analysis, I employ appropriate statistical and machine learning algorithms in trying to establish the indirect effect of high school factors (for our overall mediated

model later on). First, I look at how the middle school measures of student learning, engagement and academic emotions are predictive of the high school questionnaire responses, through multinomial logistic or decision tree algorithms. Then, I explore the association between the high school questionnaire responses with the multivariate postsecondary outcomes using structural equation modeling (factor analysis, regression, or path analysis).

Finally, by integrating emergent relationships and causal effects of middle school and high school factors on postsecondary outcomes conducted in the previous analyses, I will develop a multivariate predictive mediated model (c in Figure 1 below). Using student data that have complete information from middle school, to high school, to postsecondary years, I conduct causal modeling by fitting a mediational pathway model and evaluate how each of the variables influence one another over time [3]. In particular, using structural equation modeling (SEM), I develop a pathway starting from the middle school factors to the postsecondary outcomes, with high school factors as intervening or mediating factors. With significant zero-order correlations between the constructs (middle school factors, high school factors, postsecondary outcomes) established from the previous analyses, I employ a multiple regression analysis predicting postsecondary outcomes from both middle school and high school factors. It is expected that any partial effect (indirect effect) of high school factors (controlling for middle school factors) to be significant, decreasing the direct effect of middle school factors on postsecondary outcomes. Other SEM variants, such as factor analysis and path analysis are expected to be used as well for this analysis phase, to test the mediation model. This causal modeling has been used in educational research modeling motivational phenomena over time [3].

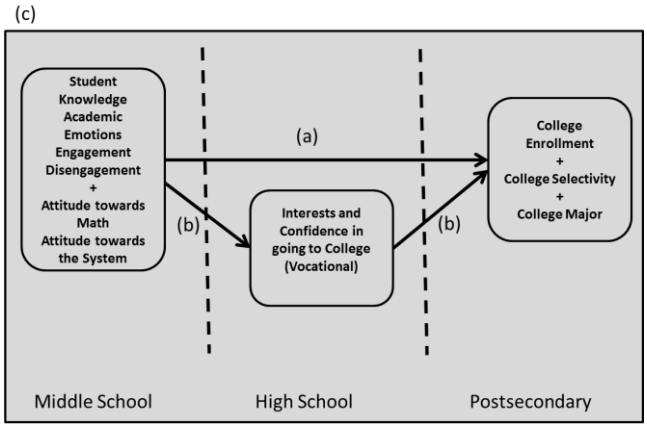

Figure 1. Modeling Postsecondary Outcomes from Middle School and High School factors: (a) Middle school factors predicting postsecondary outcomes; (b) Middle school factors predicting high school factors, High school factors predicting postsecondary outcomes; (c) Overall mediation model.

5. References

[1] Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., and Koedinger, K. 2008. Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19, 2, 185-224.

[2] Bandura, A. 1997. *Self efficacy: The exercise of control*. New York,. NY: W. H. Freeman & Company.

[3] Blackwell, L., Trzesniewski, K., and Dweck, C. 2007. Implicit Theories of Intelligence Predict Achievement Across an Adolescent Transition: A Longitudinal Study and an Intervention. *Child Development*, 78, 1, 246-263.

[4] Borgen, F. and Betz, N. 2008. Career self-efficacy and personality: Linking the Career Confidence Inventory and the Healthy Personality Inventory. *Journal of Career Assessment*, 16, 22-43.

[5] Bowers, A. J. 2010. Grades and graduation: A longitudinal risk perspective to identify student dropouts. *The Journal of Educational Research*, 103, 3, 191-207.

[6] Chen, X. 2009. Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education. Stats in Brief. *NCES 2009-161*. National Center for Education Statistics.

[7] College Division of Barron's Education Series (Ed.). 2012. *Barron's profiles of American colleges (30th ed.)*. Hauppauge, NY: Barron's Educational Series, Inc.

[8] Eccles, J. S., Vida, M. N., and Barber, B. 2004. The relation of early adolescents' college plans and both academic ability and task-value beliefs to subsequent college enrollment. *Journal of Early Adolescence*, 24, 63-77.

[9] Griffith, A. L. 2010. Persistence of women and minorities in STEM field majors: Is it the school that matters? *Economics of Education Review*, 29(6), 911-922.

[10] Lent, R. W., Brown, S. D., and Hackett, G. 1994. Toward a unifying social cognitive theory of career and academic interest, choice and performance. *Journal of Vocational Behavior*, 45, 1, 79–122.

[11] Lent, R. W., Lopez Jr, A. M., Lopez, F. G., and Sheu, H. B. 2008. Social cognitive career theory and the prediction of interests and choice goals in the computing disciplines. *Journal of Vocational Behavior*, 73, 1, 52-62.

[12] Mcquiggan, S. W., Mott, B. W., and Lester, J. C. 2008. Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User modeling and user-adapted interaction*, 18, 81-123.

[13] Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., and Perry, R. P. 2010. Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102, 3, 531.

[14] Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., and Rasmussen, K. P. 2005. The Assistment project: Blending assessment and assisting. In *Proceedings of the 12th Annual Conference on Artificial Intelligence in Education*, 555-562.

[15] Roderick, M., Nagaoka, J., and Coca, V. 2009. College readiness for all: The challenge for urban high schools. *The Future of Children*, 19, 1, 185-210.

[16] Schmidt, W., Burroughs, N., Cogan, L., and Houang, R. 2011. Are College Rankings an Indicator of Quality Education? In *Forum on Public Policy Online*, 2011, 3. Oxford Round Table.

Who Do You Think I Am? Modeling Individual Differences for More Adaptive and Effective Instruction

Laura K. Allen
Arizona State University
Tempe, AZ, 85283
LauraKAllen@asu.edu

ABSTRACT
The purpose of intelligent tutoring systems is to provide students with personalized instruction and feedback. The focus of these systems typically rests in the adaptability of the feedback provided to students, which relies on automated assessments of performance in the system. A large focus of my previous work has been to determine how natural language processing (NLP)_ techniques can be used to model individual differences based on students' natural language input. My proposed research will build on this work by using NLP techniques to develop stealth assessments of students' individual differences and to provide more fine-grained information about the cognitive processes in which these students are engaged throughout the learning task. Ultimately, my aim will be to combine this linguistic data with on-line system data in order to develop more robust student models within ITSs for ill-defined domains.

Keywords
Intelligent Tutoring Systems, Natural Language Processing, Writing, Feedback, User Models

1. INTRODUCTION
The purpose of intelligent tutoring systems (ITSs) is to provide students with personalized instruction and feedback based on their performance, as well as other relevant individual characteristics [1]. The focus of these systems typically rests in the adaptability of the feedback provided to student users, which relies on automated assessments of students' performance in the system. Despite this adaptive feedback, however, many ITSs lack the ability to provide adaptive *instruction* and *higher-level feedback*, particularly when providing tutoring for ill-defined domains. This shortcoming is largely due to the increased difficulties associated with accurately and reliably assessing student characteristics and performance when the learning tasks are not "clear cut." In mathematics tutors, for instance, it can be relatively straightforward to determine when a student is struggling in specific areas; thus, these systems can provide adaptive instruction and feedback accordingly. For ITSs focused on ill-defined domains (such as writing and reading), on the other hand, this process can be more complicated. In particular, students' *open-ended* and *natural language* responses to these systems present unique assessment challenges. Rather than identifying a set of "correct" answers, the system must identify and analyze characteristics related to students' responses in order to determine the quality of their performance as well as the areas in which they are struggling.

Natural language processing (NLP) techniques have been proposed as a means to target this assessment problem in adaptive systems. In particular, NLP provides detailed information about the characteristics of students' natural language responses within these systems [2] and subsequently helps to model students' particular areas of strengths and weaknesses [3]. NLP has begun to be incorporated within ITSs more frequently [4-5] because it allows systems to automatically evaluate the quality and content of students' responses [6-7]. Additionally, these assessments afford systems the opportunity to model students' learning throughout training and subsequently improve models of their performance [8]. Previous research suggests that these NLP techniques can increase the efficacy of computer-based learning systems. In particular, NLP helps to promote greater interactivity in the system and, consequently, leads to increased learning gains when compared to non-interactive training tasks (e.g., reading books, watching videos, listening to lectures [5, 9].

In my previous research, my colleagues and I have proposed that NLP techniques can be used to determine much more than simply the *quality* of a particular response in the system. Specifically, NLP can serve as a powerful methodology for modeling individual differences among students, as well as for examining the specific processes in which these students are engaging [3, 8]. In this overview, I suggest that, when combined with *on-line* interaction data, these NLP techniques can provide critical information that can be used to enhance the adaptability of ITSs, particular those focused on ill-defined domains. Thus, the aim of my research is to investigate how the linguistic characteristics of students' language can provide a window into their cognitive and affective processes. This information will then be combined with system data to promote more personalized learning experiences for the student users in these systems.

1.1 Writing Pal
The Writing Pal (W-Pal) is a tutoring system that was designed for the purpose of increasing students' writing proficiency through explicit strategy instruction, deliberate practice, and automated feedback [10]. In the W-Pal system, students are provided explicit

strategy instruction and deliberate practice throughout eight instructional modules, which contain strategy lesson videos and educational mini-games. The instruction in these modules covers specific topics in the three main phases of the writing process—prewriting (*Freewriting*, *Planning*), drafting (*Introduction Building*, *Body Building*, *Conclusion Building*), and revising (*Paraphrasing*, *Cohesion Building*, *Revising*).

Animated pedagogical agents narrate the W-Pal lesson videos by providing explicit descriptions of the strategies and examples of how these strategies can be used while writing (see Figure 1 for screenshots). The content covered in these videos can be practiced in one or more of the mini-games contained within each module. The purpose of these mini-games is to offer students the opportunity to practice the individual writing strategies without having to compose an entire essay.

Figure 1. Screenshots of the W-Pal Lesson Videos

W-Pal contains an AWE component in addition to the eight instructional modules, where students can practice holistic essay writing. This component of W-Pal contains a word processor where students can compose essays and automatically receive summative (i.e., holistic scores) and formative (i.e., actionable, strategy-based) feedback on these essays. The *summative* feedback in W-Pal is calculated using the *W-Pal assessment algorithm*. This algorithm employs linguistic indices from multiple NLP tools to assign essays a score from 1 to 6 (for more information, see 11). The purpose of the *formative* feedback is to teach students about high-quality writing and to provide them with actionable strategies for improving their essays. To deliver this feedback, W-Pal first identifies weaknesses in students' essays (e.g., essays are too short; essays are unorganized). It then provides students with feedback messages that designate specific strategies that can help them to work on the problems. Previous studies have demonstrated that W-Pal is effective at promoting increases in students' essay scores over the course of multiple training sessions [6; 12].

1.2 Current Work

The focus of my doctoral research will be on the use of NLP techniques to develop stealth assessments of students' individual differences and to provide more fine-grained information about the cognitive processes in which these students are engaged throughout the learning task. Ultimately, the aim of this research will be to combine this linguistic data with on-line system data in order to develop more robust student models within ITSs for ill-defined domains, such as W-Pal.

The goal of this specific research project will be to use the linguistic properties of students' essays to model individual differences related to writing performance (e.g., vocabulary knowledge). This data will then be combined with *on-line* process data, such as students' keystrokes while writing, to provide a more complete understanding of their writing processes. Ultimately, this project will aim to determine whether there are specific writing processes (as identified by the *characteristics* of the essays and students' *on-line processes*) that are more or less predictive of successful writing and revision. My final goal will then be to use this information to provide more adaptable instruction and formative feedback to students.

2. Proposed Contributions of Current Work

This proposed research project will contribute to both the W-Pal system, as well as the EDM community more generally. Regarding the W-Pal system, the development of stealth assessments and online student models will significantly enhance the adaptability and, theoretically, the efficacy of the system. The current version of W-Pal does not provide individualized instruction to students and only adapts the feedback based on single (i.e., isolated) essays that they generate. Thus, the system does not consider students' previous interactions with the system when providing feedback, nor the individual characteristics of these student users. Therefore, the proposed work will help to provide a much more robust student model, which should help W-Pal provide more personalized instruction and feedback.

More generally, the results of this project (and future projects) will contribute to the EDM community, as well as to research with natural language data more broadly. Language is pervasive and, here, we propose that it can be used to provide *unique* information about individuals' behaviors, cognitive processes, and affect. By investigating the specific characteristics of students' natural language data, we can glean important insights about their learning processes, beyond information that can be extracted from system log data. By combining NLP with other forms of data, researchers will gain a more complete picture of the students using the system, which should ultimately lead to more effective instruction.

3. Previous Work

A large focus of my previous work has been to determine how NLP techniques can be used to model individual differences based on students' natural language input. Importantly, this input has ranged from more structured language (such as essays) to naturalistic language responses (such as self-explanations). As an example, in one study, my colleagues and I investigated whether we could leverage NLP tools to develop models of students' comprehension ability based on the linguistic properties of their self-explanations [3]. Students (n = 126) interacted with a reading comprehension tutor where they self-explained target sentences from science texts. Coh-Metrix [13] was then used to calculate the linguistic properties of these aggregated self-explanations. The results of this study indicated that the linguistic indices were predictive of students' reading comprehension ability, over and above the current system algorithms (i.e., the self-explanation scores). These results are important, because they suggest that NLP techniques can inform stealth assessments and help to improve student models within ITSs.

In further research projects, we have begun to investigate how these linguistic characteristics change across time, and how these changes relate to individual differences among the students [14].

In particular, we proposed that the *flexibility* of students' writing style could provide important information about their writing proficiency. In one study, we investigated college students' (n = 45) flexibility in their use of cohesion across 16 essays and whether this flexibility related to their writing proficiency. The results suggested that more proficient writers were, indeed, more flexible in their use of cohesion across different writing prompts and that this cohesive flexibility was most strongly related to the unity, or coherence, of students' writing. The results of this study indicated that students might differentially employ specific linguistic devices in different situations in order to achieve coherence among their ideas. Overall, the results of these (and many other) studies provide preliminary evidence that NLP techniques can be used to provide unique information about students' individual differences and learning processes within ITSs.

4. Advice Sought

I am seeking advice for my proposed research regarding two primary questions. First, *what analytical methods should be used to most effectively model individual differences based on linguistic data?* In previous research, my colleagues and I have relied heavily on stepwise regression and discriminant function analysis techniques to model students' essay scores and individual differences. However, this technique can pose particular problems and is not always the most effective regarding large-scale data sets containing many variables, such as these. Thus, I would largely benefit from expert advice regarding the specific modeling techniques that can help to improve this research.

My second question relates to: *what on-line process data can be most effectively tied with this linguistic data – and how?* In previous studies, we have heavily relied on the linguistic properties of students' responses alone to model and understand the learning process. However, these models could be greatly strengthened through the addition of on-line processing data, such as keystrokes or eye tracking. We have begun to implement keystroke logging into the W-Pal system to begin to investigate this question. However, I would greatly benefit from expert advice regarding the best methods for combining this data into a reliable and accurate student model.

5. ACKNOWLEDGMENTS

Prior research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080589 to Arizona State University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. I would also like to thank Danielle McNamara, Scott Crossley, Erica Snow, Jennifer Weston, Rod Roscoe, and Matthew Jacovina for their contributions to this line of work.

6. REFERENCES

[1] Murray, T. 1999. Authoring intelligent tutoring systems: An analysis of the state of the art. International *Journal of Artificial Intelligence in Education,* 10, (1999) 98-129.

[2] Crossley, S. A., Allen, L. K., and McNamara, D. S. 2014. Analyzing discourse processing using a simple natural language processing tool (SiNLP). *Discourse Processes,* 51, (2014) 511-534.

[3] Allen, L. K., Snow, E. L., and McNamara, D. S. in press. Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. *Proceedings of the 5th International Learning Analytics and Knowledge Conference*.

[4] Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A. and Louwerse, M. 2004. AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, 36*, (2004) 180-193.

[5] VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., and Rose, C. P. 2007. When are tutorial dialogues more effective than training? *Cognitive Science, 31*, (2007), 3-62.

[6] Crossley, S. A., Varner, L. K., Roscoe, R. D., and McNamara, D. S. 2013. Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In K. Yacef et al (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED).* Springer, Heidelberg, Berlin, 269-278.

[7] Rus, V., McCarthy, P., Graesser, A. C., and McNamara, D. S. 2009. Identification of sentence-to-sentence relations using a textual entailer. *Research on Language and Computation, 7,* (2009) 209-229.

[8] Varner, L. K., Jackson, G. T., Snow, E. L., and McNamara, D. S. 2013. Are you committed? Investigating interactions among reading commitment, natural language input, and students' learning outcomes. In S. K. D'Mello, R. A., Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining*. Springer, Heidelberg, Berlin, 368-369.

[9] Graesser, A. C., McNamara, D. S., and Rus, V. 2007. Computational modeling of discourse and conversation. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge University Press, Cambridge, UK.

[10] Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., and McNamara, D. S. 2014. The Writing Pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, 34, (2014) 39-59.

[11] McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., and Dai, J. 2015. Hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, (2015), 35-59.

[12] Allen, L. K., Crossley, S. A., Snow, E. L., and McNamara, D. S. 2014. Game-based writing strategy tutoring for second language learners: Game enjoyment as a key to engagement. *Language Learning and Technology*, 18 (2014), 124-150.

[13] McNamara, D. S., Graesser, A. C., McCarthy, P., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* Cambridge: Cambridge University Press.

[14] Allen, L. K., Snow, E.L., and McNamara, D. S. 2014. The long and winding road: Investigating the differential writing patterns of high and low skilled writers. In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 304-307). London, UK.

Developing Self-Regulated Learners Through an Intelligent Tutoring System

Kim Kelly
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
508-461-6386
kkelly@wpi.edu

Neil Heffernan
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
508-831-5569
nth@wpi.edu

1. ABSTRACT

Intelligent tutoring systems have been developed to help students learn independently. However, students who are poor self-regulated learners often struggle to use these systems because they lack the skills necessary to learn independently. The field of psychology has extensively studied self-regulated learning and can provide strategies to improve learning, however few of these include the use of technology. The present proposal reviews three elements of self-regulated learning (motivational beliefs, help-seeking behavior, and meta-cognitive self-monitoring) that are essential to intelligent tutoring systems. Future research is suggested, which address each element in order to develop self-regulated learning strategies in students while they are engaged in learning mathematics within an intelligent tutoring system.

2. KEYWORDS

Intelligent tutoring systems, self-regulated learning, meta-cognition

3. DEFINING THE PROBLEM

Intelligent tutoring systems (ITS) are designed to provide independent learning opportunities for students. Learning occurs through hints, tutoring, scaffolding and correctness feedback. A great body of research exists surrounding types and timing of feedback [6] and tutoring that have been found to improve student outcomes. As a classroom teacher I have used several different ITS with students to help them learn mathematics. Over the years, I have seen many students benefit from these systems. However, I have also witnessed students struggling to use the systems and who fail to learn, despite all of the assistance provided. Addressing this failure serves as the basis of my dissertation. For an ITS to achieve maximum results, the students using the system must be good self-regulated learners. My proposed research attempts to use an ITS to develop self-regulating strategies, while students are learning the desired content.

Zimmerman and Campillo [12], suggest that self-regulated learning is a three-phase process. During the *Forethought Phase*, students engage in a task analysis, which includes goal setting and strategic planning. Self-motivational beliefs, including self-efficacy [11, 4] outcome expectations, task value/interest [10], and goal orientation also play a significant role in this phase as they have been found to positively affect student learning. During the *Performance Phase*, students demonstrate self-control by employing various task strategies and help-seeking behaviors. Self-observation, which includes meta-cognitive self-monitoring, is also crucial. During the final phase, *Self-Reflection*, students engage in self-judgment and self-reaction.

2. PROPOSED SOLUTION

To help develop self-regulated learners, these components must be explicitly taught. However, some aspects are seemingly more relevant than others when interacting with an ITS. Specifically motivational beliefs, help-seeking behavior, and meta-cognitive self-monitoring can all be addressed within the structures of intelligent tutoring systems. The following sections discuss each of these components by presenting relevant literature, sharing results of my previously published studies, and proposing future research components of my dissertation.

2.1 Motivational Beliefs

One aspect of the first phase of self-regulated learning is motivation. Students who are strong self-regulated learners have high self-efficacy. Schunk [11] defines self-efficacy as "an individual's judgment of his or her capabilities to perform given actions." A student's belief that they are capable of learning can be influenced by a growth mindset [4]. Some of my earlier research, using teacher-created motivational videos, attempted to create a growth mindset in students while they were completing math homework inside of an intelligent tutoring system [7]. While the minimal intervention failed to show changes in student self-reports of mindset, there was a significant increase in the perception of task value and homework completion rates as a result of a video inspired by [10]. In addition to improving self-efficacy, increasing task value/interest is important to developing self-regulated learners. The protocol employed in my initial study is promising and a more sophisticated intervention will be explored to further increase motivation.

2.2 Help Seeking Behaviors

Intelligent tutoring systems provide many different structures to support student learning. One such structure that I have explored is correctness-only feedback. I found that this simple support provided by an ITS during a homework assignment was found to improve student learning significantly compared to traditional paper and pencil homework that did not provide immediate feedback [8]. Yet research has shown that many students do not effectively take advantage of these features. Aleven et al. [1] explores ineffective help use in interactive learning environments and suggests that there are system-related factors, student-related factors and interactions between these factors that impact help-seeking behaviors. In one of my recent studies, I found that there

are students who, despite access to the same instructional supports, do not successfully take advantage of them and therefore do not learn [9]. This has resulted in a phenomenon called wheel spinning [3], where students persist without making progress towards learning. I hypothesize that wheel spinning is a result of ineffective help-seeking behaviors. Therefore, I propose a study that would provide direct interventions to teach students the necessary help-seeking behaviors to become self-regulated learners.

2.3 Meta-Cognitive Self-Monitoring

Elements of meta-cognition, are evident in all three phases of self-regulated learning. For example, goal setting is prominent in phase one. Other elements, like self-monitoring, are evident in multiple phases. Self-monitoring involves students becoming aware of their performance and judging their knowledge. This is sometimes referred to as metacognitive knowledge monitoring [5]. In phase two, while students are participating in a learning task, they must monitor what they are learning. Students who are strong self-regulated learners will seek feedback to easily monitor their progress. I surveyed my students to better understand their perception of feedback. High performing students claimed that the immediate feedback provided by an ITS caused frustration, but was also beneficial to their learning [8]. They were able to identify their mistakes and learn from them. To help all students recognize the importance of monitoring their learning, I propose a study where students are provided feedback along with progress monitoring to show the benefits.

Self-monitoring continues into the third phase of self-regulated learning. During this reflection stage, students assess their success or failure. Strong self-regulated learners may challenge themselves in some way to confirm their success. A willingness to seek out challenges ties back into the growth mindset that is addressed in phase one. Students who believe that intelligence is fixed will often shy away from challenges for fear of failure, whereas students with a growth mindset view challenges as opportunities to learn more [4]. Therefore, to encourage all students to seek out challenges as a method to self-monitor, I propose a study where growth mindset messages are embedded in ITS and opportunities for students to choose challenging problems are provided.

3. CONTRIBUTION

Intelligent tutoring systems rely on independent learning practices to effectively teach students. For example, students must use available hints and tutoring to navigate new material. However not all students successfully learn when using an ITS. Some early research suggests that these students are those who struggle with self-regulated learning. The field of psychology has studied self-regulated learning for more than a decade, resulting in many ideas that can improve instruction. Some ITS have incorporated features to help students who lack self-regulated learning strategies, like automatically detecting when a student is frustrated [2] and providing additional assistance when a student is failing. However, little research has explored how technology can actually promote self-regulated learning. By integrating the capabilities of intelligent tutoring systems with the vast knowledge of self-regulated learning, the proposed research seeks to teach students how learn effectively. By addressing specific aspects of self-regulated learning, ITS can actually teach students how to learn while teaching them content.

This paper is part of my dissertation proposal and is being submitted as a doctoral consortium paper to the Artificial Intelligence In Education Conference (2015) and the Educational Data Mining Conference (2015).

4. REFERENCES

[1] Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, 73(3), 277-320.

[2] Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., Koedinger, K. (2008) Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments. *Journal of Interactive Learning Research*, 19 (2), 185-224

[3] Beck, J., & Gong, Y. (2013). Wheel-Spinning: Students Who Fail to Master a Skill. In *Proceedings of the 16th Artificial Intelligence in Education*, Lane, H.C., Yacef, K., Mostow, J., & Pavlik, P. (Eds.) 431-440.

[4] Dweck, C. (2006). *Mindset*. New York: Random House.

[5] Isaacson, R. M., & Fujita, F. (2006). Metacognitive Knowledge Monitoring and Self-Regulated Learning: Academic Success and Reflections on Learning. *Journal of the Scholarship of Teaching and Learning*, 6(1), 39-55.

[6] Kehrer, P., Kelly, K. & Heffernan, N. (2013). Does Immediate Feedback While Doing Homework Improve Learning. In Boonthum-Denecke, Youngblood(Eds) *Proceedings of the Twenty-Sixth International Florida Artificial Intelligence Research Society Conference*, FLAIRS 2013, St. Pete Beach, Florida. May 22-24, 2013. AAAI Press 2013. p 542-545.

[7] Kelly, K., Heffernan, N., D'Mello, S., *Namias*, J., & Strain, A. (2013a). Adding Teacher-Created Motivational Video to an ITS. *Florida Artificial Intelligence Research Society* (FLAIRS 2013). 503-508.

[8] Kelly, K.. Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, J., Soffer Goldstein, D., (2013b). Estimating the Effect of Web-Based Homework. In Lane, Yacef, Motow & Pavlik (Eds) *The Artificial Intelligence in Education Conference*. Springer-Verlag. pp. 824-827.

[9] Kelly, K., Wang, Y., Thompson, T., Heffernan, N. (2015). Defining Mastery: Knowledge Tracing Versus N-Consecutive

Correct Responses. Submitted to Educational Data Mining Conference: Madrid Spain (2015).

[10] Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315-341.

[11] Schunk, D. H. (1996). Self-efficacy and academic motivation. *Educational Psychologist*, 26, 207-231.

[12] Zimmerman, B. J. & Campillo, M. (2002). Motivating self-regulated problem solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The nature of problem solving*. New York: Cambridge University Press.

Data-driven Hint Generation from Peer Debugging Solutions

Zhongxiu Liu
Department of Computer Science
North Carolina State University
Raleigh, NC 27695
zliu24@ncsu.edu

ABSTRACT

Data-driven methods have been a successful approach to generating hints for programming problems. However, the majority of previous studies are focused on procedural hints that aim at moving students to the next closest state to the solution. In this paper, I propose a data-driven method to generate remedy hints for BOTS, a game that teaches programming through a block-moving puzzle. Remedy hints aim to help students out of dead-end states, which are states in the problem from where no student has ever derived a solution. To address this, my proposed work includes designing debugging activities and generating remedy hints from students' solutions to debugging activities.

1. INTRODUCTION

Programming problems are characterized by huge and expanding solution spaces, which cannot be covered by manually designed hints. Previous studies have shown fruitful results in applying data-driven approaches to generate hints for programming problems. Barnes and Stamper [1] designed the Hint Factory, which gives student feedback using previous students' data. The Hint Factory uses a data structure called an interaction network as defined by Eagle et al. [3], in which nodes represent the program states and edges represent the transitions between states. Peddycord et al. [7] applied the Hint Factory in BOTS, a game that teaches programing through block-moving puzzles. This study introduced worldstates, which represent the output of a program, and compared them to codestates, snapshots of the source code. This study found that using interaction networks of worldstates can generate hints for 80% of programming states. Rivers and Koedinger [9] applied the Hint Factory in a solution space where snapshots of students' code (program state) are represented as trees, and trees are matched when the programs they represent are within a threshold of similarity. Piech et al. [8] applied data-driven approach to programs from a MOOC. This work compared the methods in Rivers and Koedinger's[9] and Barnes's[1] studies, together with algorithms that predict the desirable moving direction from a program state and generate hints to push students toward the desirable direction.

However, previous studies mainly focused on generating procedural hints that direct students to the next program state. Data from previous students' work may be insufficient to provide a next-step hint from a "dead-end state". Second, even if a next-step hint could be generated, simply telling students where to move next is not enough. An example of this situation is shown in Figure 2 - if a student follows a path that leads to a dead-end state (marked in blue), then the only hint we are able to offer is to delete all work since the last branching point. This may be a bad advice; just because we have not seen a student solve the problem this way does not mean that the solution is incorrect. Even with a correct solution down this path, we are unlikely to see it since most students solved the problem in a more conventional way, either because they have a better understanding of the problem or because our hints guide them towards the more conventional solution. Thus, students in dead-end states, who may actually have a correct solution in mind, are unable to receive helpful hints.

In this paper, I propose a data-driven method to generate remedy hints in Bots. Remedy hints are hints that help students in dead-end states by telling them why their current state is wrong, and where to move from their current state. To address the problem of insufficient data, I will collect data from debugging activities in BOTS, where students work out solutions from dead-end states and provide explanations. I hypothesize that this study will not only help students who are wheel-spinning on dead-end states, but also the students who are providing debugging solutions.

2. RESEARCH METHODOLOGY

2.1 Designing Debugging Activities

Debugging activities will be designed as bonus challenges for students who successfully complete a level. The content of debugging activities will be the dead-end states from the problem they completed. Given a dead-end state, a student will first be asked to explain the error in the program, and why it led to the dead-end state. The student will then be asked to explain his/her debugging strategy. Lastly, the student will apply his/her debugging strategy and fix the program from its current state to a goal state. In this process, both the student-written explanations and the transitions of program states will be used as hints. A more detailed explanation of these are explained in the following section.

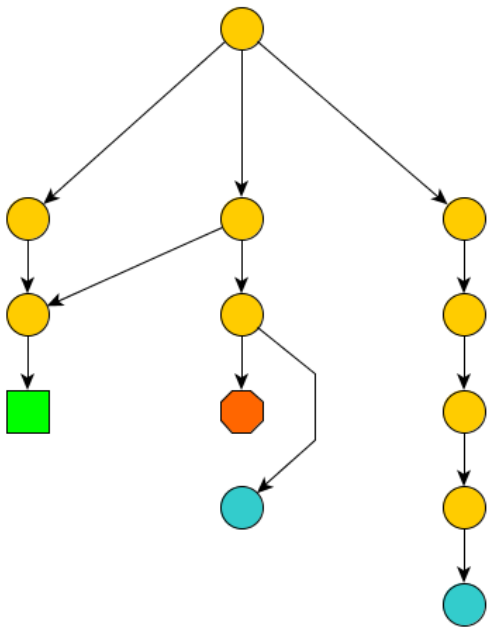

Figure 1: Interaction Network in BOTS.Green is the solution state;orange is an error state (e.g. the robot runs off the stage);blue are the dead-end states; yellow represents the rest states

To encourage students to participate in debugging activities, I will introduce a voting system. Completing debugging activities will earn points or advantages from the game. Currently, BOTS applies a rewarding system for students who solve the puzzle with fewer lines of code, as shown in Figure 2. On the left is the optimal number of lines of code needed to solve the puzzle. On the right is the current player's record for the fewest lines of code. Players earn 4 stars for reaching the optimal solution, 3 stars for being within a certain threshold value, down to one star for merely completing the puzzle. Additionally, clicking the optimal solution shows the name of the first user to reach the optimal solution.

I will design a similar leaderboard to reward students who used fewer steps when debugging for a dead-end state. Encouraging students to use fewer steps will reduce the size of debugging solutions, and the likelihood that a student will delete previous work and start from scratch. Moreover, students will receive rewards for writing good quality explanations on states and debugging strategies. The quality will be measured by a voting mechanism. Students who received a student-written explanation will be able to vote for the hint as "helpful." The more votes an explanation receives, the more points its author will get. Students with the most points will have their names appear in a leaderboard.

2.2 Construct Hint from Debugging Work

Completing a debugging problem is defined as successfully moving from the current state to the final goal state. The debugging process will be treated as a self-contained problem with its own local interaction network. When completed, this local interaction network will be added to the global interaction network for the problem. With a more complete

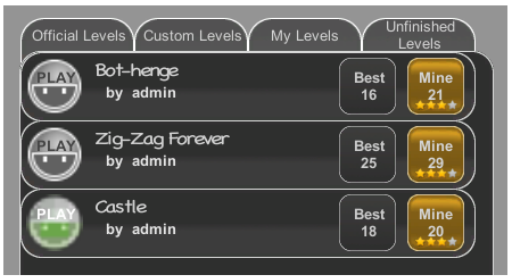

Figure 2: BOTS rewarding system

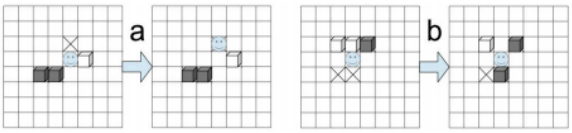

Figure 3: Two generated hints for a simple puzzle. The blue is the robot. The 'X' is a goal. Shaded boxes are boxes placed on goal spot. Not shaded boxes are not on goal spot.

global interaction network, Hint Factory [1] can be applied to generate hints for previously dead-end states.

Student-written explanations will be presented together with hints generated by the Hint Factory. An example hint from the current BOTS system is shown in Fig 3. Before presenting the hints from the Hint Factory, a student in dead-end state will see a student-written explanation on where and why their current program is wrong. This will give students a chance to reflect on their own program. Then, the student can request to see a student-written explanation of the debugging plan for the current state. This will enable the student to solve the problem on their own following a debugging plan, instead of blindly following procedural hints.

When multiple debugging approaches are available for a state, I will experiment with selecting the best debugging solution to generate hints. Ideally, I would select a debugging approach with the shortest solution path. However, there may be situations where students debugged by starting over from the beginning, which may or may not be the best solution. One approach is to evaluate the path that leads toward the current state. Assume there is a failure state in the student's solution; the earlier this failure state occurs in the path, the more likely the solution is wrong from the start and back-to-start is a good solution.

When multiple student-written explanations are available for a debugging solution, I will start by randomly choosing one explanation. As the voting process goes, I will filter out the explanations with significantly lower 'helpful' votes.

3. EVALUATION

My evaluation will focus on the below research questions:

- What percentage of students will participate in the debugging activities, and how many write explanations? Why do students participate or not?

- What is the relationship between students' involvement in debugging and their programming performance? Will students who complete problems with shorter solutions be more involved in debugging?

- Will writing or reading student-written explanations and debugging strategies help learning?

- In the global interaction network, what percentage of the dead-end program states receive hints from student debugging solutions?

Previous BOTS participants are students from after-school programming education activities. In my experiment, I will randomly recruit the same type of students. These students will be separated into a control group where students will use the traditional BOTS system, an experimental group A where students will be given the option to do debugging challenges, and an experimental group B where students must do debugging challenges after completing a level.

To answer the first research question, students from the two experimental groups will do a post survey on their opinions about debugging activities and hints generated from student-written explanations. For experimental group A, I will add survey questions on why students chose to participate or not participate in debugging activities. To answer the second question, students' interaction and compilation data while playing BOTS will be recorded. These data will be used to measure the relationship between involvement in the debugging activities and programming performance. To answer the third research question, students from all groups will do pre and post-tests on basic programming and debugging concepts that are related to BOTS content. Learning gains will be measured as the difference between pre and post-test. To answer the fourth question, the program state space coverage will be compared between the three groups.

4. PROPOSED CONTRIBUTION

My work will generate a new type of hint that may lead to different pedagogical results than the procedural hint, especially for students in dead-end states. My work will demonstrate the feasibility of collecting data from peer students' debugging processes, and generating helpful hints.

My work will design a feature that supports both programming and debugging activities in an educational game. This design will have several pedagogical benefits. First, Kinnunen and Simon's[6] research have shown that novice programmers experienced a range of negative emotions after errors. Practicing debugging will help novice programmers proceed after errors, and enjoy programming experiences. Second, students will make self-explanations on the observed flaw and debugging strategy, and decades of research such as Johnson and Mayer's[5], and Chi et al.[2] have shown that self-explanation is extremely beneficial to learning. Third, students in dead-end states will not only receive help, but also learn what peer students think given the same situation.

5. ADVICE SOUGHT

Johnson and Mayers[5], and Hsu et al. studies[4] have shown that merely adding self-explanation features did not help learning, but students' engagement in self-explaining did. Therefore, I want to seek advice on the design of debugging activities that engage students in debugging and writing explanations, and produce quality work. I also want to seek advice on the evaluation. Given the previous question, how should I measure the level of engagement in debugging and self-explaining?

Moreover, introducing debugging challenges as extra activities will affect other measurements. For example, students who spend a significant amount of time in debugging may complete less problems given the time constraint, and exhaust earlier. How should I address this problem and measure students' performance fairly? Moreover, how to design pre and post-tests to measure learning gains from debugging process? Lastly, what are the potentials, benefits, and risks to expand this work into programming problems using mainstream programming languages?

6. REFERENCES

[1] T. Barnes and S. John. Toward automatic hint generation for logic proof tutoring using historical student data. In *Proceedings of the 6th International Conference on Intelligent Tutoring System*, pages 373–382, 2008.

[2] M. T. Chi, N. Leeuw, M. H. Chiu, and C. LaVancher. Eliciting self–explanations improves understanding. *Cognitive Science*, 18(3):439–477, 1994.

[3] M. Eagle, M. Johnson, and T. Barnes. Interaction networks: Generating high level hints based on network community clusterings. In *Proceedings of the 6th International Conference on Intelligent Tutoring System*, pages 164–167, 2012.

[4] C. Y. Hsu, C. C. Tsai, and H. Y. Wang. Facilitating third graders' acquisition of scientific concepts through digital game–based learning: The effects of self–explanation principles. *The Asia–Pacific Education Researcher*, 21(1):71–82, 2012.

[5] C. I. Johnson and R. E. Mayer. Applying the self-explanation principle to multimedia learning in a computer-based game–like environment. *Computers in Human Behavior*, 26(6):1246–1252, 2010.

[6] P. Kinnunen and B. Simon. Experiencing programming assignments in cs1: the emotional toll. In *Proceedings of the 6th international workshop on Computing education research*, pages 77–86, 2010.

[7] B. Peddycord III, A. Hicks, and T. Barnes. Generating hints for programming problems using intermediate output. In *Proceedings of the 7th International Conference on Educational Data Mining*, pages 92–98, 2014.

[8] C. Piech, M. Sahami, J. Huang, and L. Guibas. Autonomously generating hints by inferring problem solving policies. In *Proceedings of Learning at Scale*, 2015.

[9] K. Rivers and K. R. Koedinger. Automatic generation of programming feedback: A data-driven approach. In *Proceedings of the 1st workshop on AI-supported Education for Computer Science, th 16th International Conference on Artificial Intelligence on Education*, pages 50–59, 2013.

Enhancing Student Motivation and Learning Within Adaptive Tutors

Korinn S. Ostrow
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609
ksostrow@wpi.edu

ABSTRACT
My research is rooted in improving K-12 educational practice using motivational facets made possible through adaptive tutoring systems. In an attempt to isolate best practices within the science of learning, I conduct randomized controlled trials within ASSISTments, an online adaptive tutoring system that provides assistance and assessment to students around the world. My work also incorporates big data analytics through the establishment of data driven learning models that promote the use of finite assessment to optimize student modeling and enhance user motivation. This paper highlights a turning point in my research as I transition into PhD candidacy. My contributions thus far and my research goals are discussed, with consult sought on how to best meld the realms of my work moving forward. An iteration of this work has also been published as a Doctoral Consortium at AIED 2015 [4].

Keywords
Motivation, Learning, Feedback, Choice, Assessment Methodologies, Adaptive Tutoring

1. RESEARCH FOCUS

1.1 Adaptive Tutoring: ASSISTments
The U.S. Department of Education's National Educational Technology Plan supported the idea that technology will play a key role in delivering personalized educational interventions [14]. Yet there remains a severe lack of research regarding the effectiveness of online learning systems for K-12 education [15]. Adaptive tutoring systems offer interactive learning environments that allow students to excel while providing teachers a unique approach to classroom organization and data-driven lesson plans. Before the development of these adaptive platforms, research within classrooms was costly and generally required a longitudinal approach. As such, much of the evidence that supports K-12 educational practice is generalized from studies conducted by psychologists in laboratory settings with college undergraduates.

My research acts on this deficit, by conducting controlled trials using student level randomization within ASSISTments, an online adaptive tutoring system, to isolate best practices for learning outcomes while enriching the user experience. ASSISTments, commonly used for both classwork and homework, presents students with immediate feedback and a variety of rich tutorial strategies. The platform is also a powerful assessment tool, providing teachers with a variety of student and class reports that pinpoint where students are struggling and enhance classroom techniques using real time data. Further, the platform is unique in that it allows educational researchers to design and implement content-based experiments without extensive understanding of computer programming, serving as a shared collaborative tool for the advancement of the science of learning [3].

1.2 Motivational Trinity
Essentially, my work seeks to enhance student motivation and performance by enriching content through optimized feedback delivery, exploring opportunities to make students shareholders in the learning process, and attempting to boost motivation and proper system usage through improved assessment techniques.

1.2.1 Feedback Mediums
Until recently, virtually all feedback within the ASSISTments tutoring platform was provided using text, typically with font color or typeset signifying important variables. However, adaptive tutoring systems offer the opportunity to utilize a variety of hypermedia elements, as outlined by Mayer's multimedia principles for the optimal design of e-Learning environments [1]. These twelve principles, driven by cognitive theory, promote active learning while reducing cognitive load and accounting for the average user's working memory [1]. Educational technologies that employ video tend to do so in a manner that resembles lectures rather than feedback (i.e., Khan Academy). Thus, the introduction of matched content video feedback to the ASSISTments platform through brief 15-30 second YouTube recordings offered a novel approach to investigating hypermedia within an adaptive setting.

1.2.2 Student Choice
While platforms like ASSISTments offer a variety of features, few make students shareholders in the learning process. Despite the fact that users can endlessly customize their experiences with commercial products, student preference is not a key element in the realm of education. Choice is an intrinsically motivating force [11] that has the potential to boost subjective control, or a student's perception of their causal influence over their learning outcomes [12]. Feelings of control are balanced by appraisals of subjective value, or a student's perceived importance of her learning outcome. By providing the student with choices at the start of her assignment, it may be possible to enhance

expectancies regarding her performance and thereby enhance achievement emotions such as motivation [12]. Considering the control-value theory within the realm of an adaptive tutoring system for mathematics content may help to explain and ameliorate female dropout in STEM fields [2]. Feedback medium personalization offers one simple method to examine the motivational effect of choice within these platforms.

1.2.3 Improving Assessment

Adaptive tutoring systems typically function through measures of binary correctness on a student's first attempt or first action within a problem. Within such systems, students who take advantage of tutoring feedback are unduly penalized. This creates an environment in which students are afraid to use the beneficial features of these platforms, or instead, overuse feedback if they have already lost credit (i.e., skipping to the answer rather than reading a series of hints). The establishment of partial credit scoring would help to alleviate these issues, serving to motivate student performance while simultaneously offering teachers a more robust view of student knowledge. Using data mining approaches, partial credit can be defined algorithmically [16] for the purpose of enhancing student modeling. Real time implementation of these data driven models could offer substantial benefits for all parties.

2. PROPOSED CONTRIBUTIONS

Thus far, my work has lead to eight peer reviewed articles already published or in press, as well as a multitude of projects that are in progress. Projects that best highlight my goals as I transition to my PhD work are described in the following subsections.

2.1 Published Works

2.1.1 Video vs. Text Feedback

The ASSISTments platform was used to conduct a randomized controlled trial featuring matched content video and text feedback within the realm of middle school mathematics [7]. Results suggested significant effects of video feedback, showing enhanced learning outcomes on next question performance after receiving adaptive video tutoring, as well as increased efficiency. Further, through self-report it was observed that students perceived video as a positive addition to their assignment. This study was the first of its kind to explore the potential for replacing text feedback, already shown to be successful within ASSISTments [13], with an alternate medium. A scaled-up replication of this study is currently underway. This work inspired an influx of video content into the ASSISTments platform, providing new opportunities to examine the subtleties of video feedback, including a crowd-sourced approach to feedback creation.

2.1.2 Dweckian Motivation

Moving beyond the use of video feedback and into the realm of pedagogical agents, my co-authors and I sought to investigate the motivational effects of Dweckian inspired mindset training within ASSISTments feedback [10]. A six-condition design was used to examine how growth mindset messages promoting the malleability of intelligence delivered with domain based feedback effected motivation and learning outcomes. Conditions differed on elements of audiovisual message delivery, ranging from plain text to an animated pedagogical agent. Although limited by a small sample size and ceiling effects, analyses across five mathematics skills revealed that mindset messages altered student performance as measured by persistence, learning gain, and self-reported enjoyment of the system (trends, $p \approx 0.1$). Trends also pinpointed gender differences in response to messages delivered using the pedagogical agent.

2.1.3 Partial Credit Assessment

By data mining log files from ASSISTments usage spanning the 2012-2013 school year, this work established a simple student modeling technique for the prediction of next problem correctness (time $t + 1$) using algorithmically defined partial credit scores at time t [5]. Although traditional modeling approaches and most adaptive tutors are driven by binary metrics of student correctness, employing partial credit can enhance student motivation and promote proper use of system features such as adaptive feedback, while allowing teachers a more robust understanding of student ability and simultaneously enhancing predictive modeling. Predictions gathered using a tabling approach based on maximum likelihood probabilities were able to compete with standard Knowledge Tracing models in terms of model accuracy, while drastically reducing computational costs [5].

2.2 Works in Press or in Progress

2.2.1 Student Choice

This work served as a pilot study on the addition of student choice into the ASSISTments platform [8]. This line of research examines motivation and learning when students are able to invest in the learning process. Students were randomly assigned to either Choice or No Choice conditions within a problem set on simple fraction multiplication. Those given choice were asked to select their feedback medium, while those without choice were randomly assigned to receive either text or video feedback. Results suggested that even if feedback was not ultimately used, students who were prompted to choose their feedback medium significantly outperformed those who were not. A second iteration of this study is currently underway using a new If-Then navigation infrastructure that was built because of the significant effects observed in the pilot. If previous results are replicated, these findings may be groundbreaking in that the addition of relatively inconsequential choices to adaptive tutoring systems could enhance student motivation and performance.

2.2.2 Content Delivery Patterns

Motivation and learning outcomes can also be improved by making content delivery more adaptive. Recent work within ASSISTments has revealed the benefit of interleaving (or mixing) skill content within homework settings [9]. Serving as a conceptual replication of previous work in the field, our goal was to isolate the interleaving effect within a brief homework assignment, as measured by learning gains on a delayed posttest. Using a randomized controlled trial, a practice session was presented featuring either interleaved or blocked content spanning three math skills. This study was unique in that rather than relying on a formal posttest, a second homework assignment was used to gauge learning gains through average score, hint usage, and attempt count. The use of tutoring feedback during posttest provided additional dependent variables for analysis while allowing students continued learning opportunities. Observations revealed that interleaving can be beneficial in adaptive learning environments, and appears especially significant for low performing students.

2.2.3 Assessment Enhancing Motivation

An extension of the work presented in 2.1.3, this research examined partial credit scoring using a grid search of 441 algorithmically defined models through per hint and per attempt

penalizations [6]. Binary scoring, as utilized by most adaptive tutoring systems, can serve to demotivate students from engaging with tutoring feedback and rich system features that are intended to excel beyond traditional classroom practices. For each of the 441 models examined, tables were established using maximum likelihood probabilities to predict binary next problem correctness (time $t + 1$), given the partial credit score on the current question (time t). Findings suggest that a data driven approach to defining partial credit penalization is possible and that an optimal penalization range can be isolated using model accuracy. Further, findings suggest that within the optimal range, lower penalizations do not differ significantly from higher penalizations, allowing leeway for content developers and teachers to enhance student motivation through reduced penalization.

2.3 Goals & Insight Sought

As I delve into my dissertation I expect my work to grow and meld into a unified construct surrounding the enhancement of student motivation and learning within adaptive tutoring systems. It is clear that the facets discussed here will link the two underlying realms of my research (i.e., randomized controlled trials and data mining), but it is not yet clear how. Through continued investigation of feedback, student choice, and assessment methodologies, I hope to establish a unique line of research that remains broad and yet powerful. Advice on how to drive a broad topic dissertation is sought. Essentially, I hope to gain an external expert's opinion on how to best merge the facets of my research. Advice on future endeavors within individual facets would also be appreciated.

The immediate impact of my research is already evident through continued improvements to the ASSISTments platform. The work presented here has inspired content expansion as well as infrastructure changes to enhance future research design. Within the next three years I expect that my research will continue to refine ASSISTments while increasing intellectual merit in my field. The broader impact of my work will be measured in long-term achievements that affect systemic change in education and promote data driven practices and individualized learning via adaptive tutoring platforms.

3. ACKNOWLEDGEMENTS

Funding for my work has been granted by a U.S. Department of Education GAANN fellowship (P200A120238), an Office of Naval Research grant (N00014-13-C-0127), and a PIMSE fellowship (NSF DGE-0742503). Thanks to S.O. & L.P.B.O.

4. REFERENCES

[1] Clark, R.C. & Mayer, R. E. (2003). e-Learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning. San Francisco, CA: Pfeiffer

[2] Frenzel, A.C., Pekrun, R. & Goetz, T. (2007). Girls and mathematics – A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. Eur. J of Psych of Ed. 22 (4). pp. 497-514.

[3] Heffernan N. & Heffernan C. (2014). The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. Int J Art Intel in Ed.

[4] Ostrow, K. (In Press). Motivating Learning in the Age of the Adaptive Tutor. To be included in Conati, C., Heffernan, N., Mitrovic, A., & Verdejo, M. (Eds.), Proceedings of the 17th Int Conf on AIED.

[5] Ostrow, K., Donnelly, C., Adjei, S., & Heffernan, N. (2015). Improving Student Modeling Through Partial Credit and Problem Difficulty. In Russell, Woolf, & Kiczales (Eds.), Proceedings of the 2nd ACM Conf on L@S. pp. 11-20.

[6] Ostrow, K., Donnelly, C., & Heffernan, N. (In Press). Optimizing Partial Credit Algorithms to Predict Student Performance. To be included in Romero, C., Pechenizkiy, M., Boticario, J.G., & Santos, O.C. (Eds.), Proceedings of the 8th Int Conf on EDM.

[7] Ostrow, K.S. & Heffernan, N.T. (2014). Testing the Multimedia Principle in the Real World: A Comparison of Video vs. Text Feedback in Authentic Middle School Math Assignments. In Stamper, J., et al. (Eds) Proceedings of the 7th Int Conf on EDM. pp. 296-299.

[8] Ostrow, K. & Heffernan, N. (In Press). The Role of Student Choice Within Adaptive Tutoring. To be included in Conati, C., Heffernan, N., Mitrovic, A., & Verdejo, M. (Eds.), Proceedings of the 17th Int Conf on AIED.

[9] Ostrow, K., Heffernan, N., Heffernan, C., Peterson, Z. (In Press). Blocking vs. Interleaving: An Attempt to Replicate the Concept of Comparing Schedule Types Within One Night of Math Homework. To be included in Conati, C., Heffernan, N., Mitrovic, A., & Verdejo, M. (Eds.), Proceedings of the 17th Int Conf on AIED.

[10] Ostrow, K.S., Schultz, S.E. & Arroyo, I. (2014). Promoting Growth Mindset Within Intelligent Tutoring Systems. In CEUR-WS (1183), Gutierrez-Santos, S., & Santos, O.C. (eds) EDM 2014 Extended Proceedings: NCFPAL Workshop. pp. 88-93.

[11] Patall, E.A., Cooper, H., & Robinson, J.C. (2008). The Effects of Choice on Intrinsic Motivation and Related Outcomes: A Meta-Analysis of Research Findings. Psychology Bulletin. 134 (2), pp 270-300.

[12] Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. Educational Psychology Review, 18 (4), pp. 315-341.

[13] Razzaq, L. & Heffernan, N.T. (2006). Scaffolding vs. hints in the ASSISTments system. In Ikeda, Ashley & Chan (Eds). *Proceedings of the Eighth International Conference on Intelligent Tutoring Systems*. Springer-Verlag: Berlin. 635-644.

[14] U.S. Department of Education, Office of Educational Technology. (2010a). Transforming American Education: Learning Powered by Technology. Washington, D.C.

[15] U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. (2010b). Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies. Washington, D.C.

[16] Wang, Y. & Heffernan, N. (2011). The "assistance" model: leveraging how many hints and attempts a student needs. In Proceedings of the *Florida Artificial Intelligence Research Society Conference* (FLAIRS 2011).

Estimating the Local Size and Coverage of Interaction Network Regions

Michael Eagle
North Carolina State University
Department of Computer Science
Raleigh, NC 27695-8206
mjeagle@ncsu.edu

Tiffany Barnes
North Carolina State University
Department of Computer Science
Raleigh, NC 27695-8206
tmbarnes@ncsu.edu

ABSTRACT

Interactive problem solving environments, such as intelligent tutoring systems and educational video games, produce large amounts of transactional data which make it a challenge for both researchers and educators to understand how students work within the environment. Researchers have modeled the student-tutor interactions using complex network representations to automatically derive next-step hints, derive high-level approaches, and create visualizations of student behavior. However, students do not explore the complete problem space. The nonuniform exploration of the problem results in smaller networks and less next-step hints. In this work we explore the possibility of using frequency estimation to uncover locations in the network with differing amounts of student-saturation. Identification of these regions can be used to locate specific problem approaches and strategies that would be most improved by additional student-data.

Keywords

Interaction Networks, Data-Driven, Problem Solving

1. INTRODUCTION

Data-driven methods to provide automatic hints have the potential to vastly reduce the cost associated with developing tutors with personalized feedback. Modeling the student-tutor interactions as a complex network provides a platform for researchers to generate hint-templates and automatically generate next-step hints; the interaction networks also work as useful visualization of student problem-solving, as well as a structure from which to mine high level approaches of student problem-solving approaches. Data-driven approaches require an uncertain amount of data collection before they can produce feedback, and it is not always clear how much is needed for different environments. Eagle et al. explored the structure of these student interaction networks and argued that networks could be interpreted as an empirical sample of student problem solving [4]. This would mean that students who are similar in problem-solving approaches would also be represented in the same parts of the interaction network. This would suggest that students who are more similar would have smaller networks as they explore the same parts of the problem space. We argue that as the expectation is for different populations of students to have different interaction networks and that different domains will require different amounts of student-data, there need to be good metrics for describing the quality of the networks.

In this work, we will make use of Good-Turing frequency estimation on interaction level data to predict the local size and hint-producing capability of interaction network regions. Our estimator makes use of Good-Turing frequency estimation [5]. Good-Turing frequency estimation estimates the probability of encountering an object of a hitherto unseen type, given the current number and frequency of observed objects. It was originally developed by Alan Turing and his assistant I. J. Good for use in cryptography efforts during World War II. In our context, the object types will refer to network-states (vertices,) and observations will refer to the student interactions (edges.)

Creation of adaptive educational programs is expensive, intelligent tutors require content experts and pedagogical experts to work with tutor developers to identify the skills students are applying and the associated feedback to deliver [7]. In order to address the difficulty in authoring intelligent tutoring content, Barnes and Stamper built an approach called the Hint Factory to use student data to build a Markov Decision Process (MDP) of student problem-solving approaches to serve as a domain model for automatic hint generation [12]. Other approaches to automated generation of feedback have attempted to condense similar solutions in order to address sparse data sets. One such approach converts solutions into a canonical form by strictly ordering the dependencies of statements in a program [9]. Another approach compares *linkage graphs* modelling how a program creates and modifies variables, with nested states created when a loop or branch appears in the code [6]. In the Andes physics tutor, students may ask for hints about how to proceed. Similarly to Hint Factory-based approaches, a solution graph representing possible correct solutions to the problem was used, however it was automatically generated rather than being derived from data, and uses plan recognition to decide which of the problem derivations the student is working towards [13].

Interaction networks are scale-free, in that there is a small

subset of the overall network-states which contain the largest number of neighboring states [4]. Eagle et al. argued that this was in part due to the nature of the problem-solving environment, where by students with similar problem solving ability and preferences would travel into similar parts of the network and problem-features would result in some states being more important to the problem than others [4]. With this interpretation as a basis sub-regions of the network corresponding to high-level approaches to the problem were shown to capture problem-solving differences between two experimental groups [3]. A region of the network representing a minority approach, would result in locations of the network that would not produce adequate hints for students taking that approach.

2. INTERACTION NETWORKS

An *Interaction Network* is a complex network representation of all observed student and tutor interactions for a given problem in a game or tutoring system [4]. To construct an Interaction Network for a problem, we collect the set of all solution attempts for that problem. Each solution attempt is defined by a unique user identifier, as well as an ordered sequence of interactions, where an interaction is defined as {initial state, action, resulting state}, from the start of the problem until the user solves the problem or exits the system. The information contained in a *state* is sufficient to precisely recreate the tutor's interface at each step. Similarly, an *action* is any user interaction which changes the state, and is defined as {action name, pre-conditions, result}. Regions of the network can be discovered by applying network clustering methods, such as those used by Eagle et al. for deriving maps high-level student approaches to problems [3].

Stamper and Barnes' Hint Factory approach generates a next-step Hint Policy by modeling student-tutor interactions as a Markov Decision Process [12]. This has been adapted to work with interaction networks by using a value-iteration algorithm [2] on the states [4]. We define a state, S to be *Hintable* if there exists a path on the network to a goal-state starting from S. We define the *Hintable* network to be the induced subset of the interaction network containing only *Hintable* states.

The "cold start problem" is an issue that arises in all data-driven systems where for early users of the system, predictions made are inaccurate or incomplete [11, 10]. Barnes and Stamper [1] approached the question of how much data is needed to get a certain amount of overlap in student solution attempts by incrementally adding student attempts and measuring the step overlap over a large series of trials. This was done with the goal of producing automatically generated hints, and thus solution-attempts that did not reach the goal were excluded. Peddycord et al. [8] performed a similar technique to evaluate differences in overlap between two different interaction network state representations.

2.1 Good-Turing Network Estimation

In this work, we are presenting a new method for estimating the size of the unobserved portion of a partially constructed Interaction Network. Our estimator makes use of Good-Turing frequency estimation [5]. Good-Turing frequency estimation estimates the probability of encountering an object of a hitherto unseen type, given the current number and frequency of observed objects. It was originally developed by Alan Turing and his assistant I. J. Good for use in cryptography efforts during World War II. Gale and Sampson revisited and simplified the implementation [5]. In its original context, given a sample text from a vocabulary, the Good-Turing Estimator will predict the probability that a new word selected from that vocabulary will be one not previously observed.

The Good-Turing method of estimation uses the frequency of frequencies for the sample text in order to estimate the probability that a new word will be of a given frequency. Based on this distribution, we calculate the probability of observing a new word in the vocabulary based on the observed probability of observing a word with frequency 1. Therefore, the expected probability of the next observation being an unseen word P_0 is estimated by:

$$P_0 = \frac{N_1}{N} \tag{1}$$

Where N_1 is the total number of words occurring with frequency 1, and N is the total number of observations. Since N_1 is the largest and best explored group of words, the so-far observed value of N_1 is a reasonable estimate of P_1. To apply this method to an interaction network, we will estimate the probability of encountering a new state, based on the previously seen state frequencies. P_0 can then be used to smooth the estimation proportions of the other states.

Our version of P_0 is the probability of encountering a new state (a state that currently has a frequency of zero,) on a new interaction. We also interpret this as the proportion of the network missing from the sample. We will refer to an interaction with a unobserved state as having *fallen off* of the interaction network. We will use the complement of P_0 as the estimate of *network coverage*, I_C, the probability that a new interaction will remain on the network: $I_C = 1 - P_0$.

The *state space* of the environment is the set of all possible state configurations. For both the BOTS game and the Deep Thought tutor the potential state space is infinite. For example, in the Deep Thought tutor a student can always use the addition rule to add new propositions to the state. However, as argued in Eagle et. al. [4], the actions that reasonable humans perform is only a small subset of the theoretical state space; the actions can also be different for different populations of humans. We will refer to this subset as the *Reasonable State Space*, with *unreasonable* being loosely defined as actions that we would not expect a human to take. An interaction network is an empirical sample of the problem solving behavior from a particular population, and is a subset of the state space of all possible *reasonable* behaviors. Therefore, our metrics P_0 and I_C are estimates of how well the observed interaction network represents the reasonable state space.

3. DISCUSSION

Figure 1 shows the results of a preliminary analysis on an interaction network based on student-log data from a tutoring environment. For each region we calculated values of network coverage, I_C, and have highlighted regions of the network which have values below 90% coverage. Good-

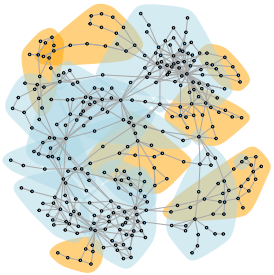

Figure 1: An interaction network with regions of high coverage highlighted in light blue and regions of low coverage highlighted in orange. The low coverage regions of the network require more data before coverage could reach a I_C level above 90%.

Turing Estimation works well in the contexts of interaction networks. Our network coverage metric I_C allows a quick and easy to calculate method of comparing different state representations, as well as quantifying the difference. New methods for improving automatic hint generation can target these areas of the network which have the lowest coverage, such as asking for instructor input on specific regions or by starting advanced students in these regions in order to observe their paths out.

We were also able to interpret this metric as measure of the proportion of the network not yet observed P_0. On a high-level this value alone is a useful metric for the percentage of times a student-interaction is to a not yet observed state. The P_0 score for the hint-able network is likewise a measure for the probability that a student will "fall off" of the network from which we can provide feedback. Therefore, we can use the P_0 metric to predict next-step "fall off" we could estimate the "risk" of different network regions. If we are reasonably sure that the majority of successful paths to the goal have been previously observed then falling off of the network likely means that the student is unlikely to reach the goal.

Region-level coverage also has implications given our previous theories on the network being a sample created from bias (non-random) walks on the problem-space, as the more homogeneous the bias-walkers are, the faster the network will represent the population and smaller total states explored will be. We revisited the results of [3], and have added more description to the effect of hint; students with access to hints explored less overall unique states which implies that the students were more similar to each other in terms of the types of actions and states they visited within the problem.

Future directions for this research include general improvements to the network clustering algorithms which generate the regions. Regions which have low coverage might not be worth separating from their parent region for visualization or high-level hint generation processes. The local and global measures of network coverage can help identify problematic regions in interaction networks which could harm hint production; they also provide a metric to evaluate new, "cold start" problems and make sure that enough data has been collected in order produce hints to multiple problem solving approaches. Finally, exploration of coverage between groups has the potential to uncover differences in problem solving behavior, and improve automatic hinting and understanding of student approaches to problems.

4. REFERENCES

[1] T. Barnes and J. Stamper. Toward automatic hint generation for logic proof tutoring using historical student data. *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS 2008)*, pages 373–382, 2008.

[2] R. Bellman. A markovian decision process. Technical report, DTIC Document, 1957.

[3] M. Eagle and T. Barnes. Exploring differences in problem solving with data-driven approach maps. *Proceedings of the Seventh International Conference on Educational Data Mining*, 2014.

[4] M. Eagle, D. Hicks, P. III, and T. Barnes. Exploring networks of problem-solving interactions. *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (LAK 15)*, 2015.

[5] W. A. Gale and G. Sampson. Good-turing frequency estimation without tears*. *Journal of Quantitative Linguistics*, 2(3):217–237, 1995.

[6] W. Jin, T. Barnes, J. Stamper, M. J. Eagle, M. W. Johnson, and L. Lehmann. Program representation for automatic hint generation for a data-driven novice programming tutor. In *Intelligent Tutoring Systems*, pages 304–309. Springer, 2012.

[7] T. Murray. Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, 10:98–129, 1999.

[8] B. Peddycord III, A. Hicks, and T. Barnes. Generating hints for programming problems using intermediate output.

[9] K. Rivers and K. R. Koedinger. Automating hint generation with solution space path construction. In *Intelligent Tutoring Systems*, pages 329–339. Springer, 2014.

[10] J. B. Schafer, D. Frankowski, J. Herlocker, and S. Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.

[11] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[12] J. Stamper, M. Eagle, T. Barnes, and M. Croy. Experimental evaluation of automatic hint generation for a logic tutor. *International Journal of Artificial Intelligence in Education (IJAIED)*, 22(1):3–18, 2013.

[13] K. Vanlehn, C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill. The andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence in Education*, 15(3):147–204, 2005.