5th International Workshop on

Personalization Approaches in Learning Environments

PALE 2015

held in conjunction with

23rd International Conference
User Modeling, Adaptation, and Personalization
UMAP 2015

Dublin, Ireland, June 30th, 2015

Proceedings edited by:

Milos Kravcik

Olga C. Santos

Jesus G. Boticario

Maria Bielikova

Tomas Horvath

Published in CEUR workshop proceedings
ISSN 1613-0073

http://ceur-ws.org/

Table of contents

5th International Workshop on Personalization Approaches in Learning Environments (PALE 2015)

Preface

Milos Kravcik[1], Olga C. Santos[2], Jesus G. Boticario[2],
Maria Bielikova[3], Tomas Horvath[4]

[1] RWTH University Aachen, Germany
kravcik@dbis.rwth-aachen.de
[2] aDeNu Research Group, Artificial Intelligence Department,
Computer Science School, UNED, Spain
ocsantos@dia.uned.es – jgb@dia.uned.es
http://adenu.ia.uned.es/
[3] PeWe Research Group. Faculty of Informatics and Information Technologies,
Slovak University of Technology in Bratislava, Slovakia
maria.bielikova@stuba.sk
[4] Pavol Jozef Safarik University, Slovakia
tomas.horvath@upjs.sk

Abstract. Personalization approaches in learning environments are crucial to foster effective, active, efficient, and satisfactory learning. They can be addressed from different perspectives and also in various educational settings, including formal, informal, workplace, lifelong, mobile, contextualized, and self-regulated learning. PALE workshop offers an opportunity to present and discuss a wide spectrum of issues and solutions. In particular, this fifth edition includes 6 papers dealing with adapting the study plan (with highlighting), student's performance (i.e., academic distress), self-regulating learning skills, interoperability in learner modelling by integrating standards (i.e., IMS specification), confusion detection by monitoring mouse movements in a computer game, and knowledge acquisition of mathematical concepts.

1 Introduction

The 5th International Workshop on Personalization Approaches in Learning Environments (PALE)[1] took place on June 30th, 2015 and was held in conjunction with the 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015). Since the topic can be addressed from different and complementary perspectives, PALE workshop aimed to offer a fruitful crossroad where interrelated issues could be

[1] http://adenu.ia.uned.es/workshops/pale2015/

contrasted and discussed. PALE 2015 was a follow-up of the four previous editions of PALE (which took place at UMAP 2011 – 2014).

In order to foster the sharing of knowledge and innovative ideas on these issues, PALE format follows the Learning Cafe methodology[2] to promote discussions on open issues regarding personalization in learning environments. Three Learning Café sessions were set up for this year PALE edition. Each one consisted of brief presentations of the key questions posed by two workshop papers and subsequent small group discussions with participants randomly grouped at tables. Each table was moderated by the presenter of the paper. In the middle of the session, participants changed tables to promote sharing of ideas among the groups. The workshop ended with a summary of the discussions on each paper. In this way, participants attending the workshop could benefit both from interactive presentations, constructive work and knowledge sharing.

The target audience of the PALE workshop includes researchers, developers, and users of personalized and adaptive learning environments. As a long-standing workshop series (for 5 years now, annually run at UMAP) PALE workshop has established itself as a mature channel for disseminating research ideas on personalization of learning environments. This could not be possible without the very much appreciated involvement of the program committee members (many of them supporting PALE all along these years) as well as the active participation of authors who have selected this venue to disseminate and discuss their research. To compile the progress achieved in this field, a special issue on User Modeling to Support Personalization in Enhanced Educational Settings taking into account extended versions of previous contributions to PALE (in addition to papers from an open call) is being guest edited by PALE organizers in the International Journal of Artificial Intelligence in Education[3].

In the following, we introduce PALE 2015 motivation and themes as well as present an overview of the contributions accepted and discussed in the workshop.

2 Motivation and Workshop Themes

Personalization is crucial to foster effective, active, efficient, and satisfactory learning, especially in informal learning scenarios that are being demanded in lifelong learning settings, with more control on the learner side and more sensitivity towards context. Personalization of learning environments is a long-term research area, which evolves as new technological innovations appear.

Previous PALE editions have shown several important issues in this field, such as behavior and embodiment of pedagogic agents, suitable support of self-regulated learning, appropriate balance between learner control and expert guidance, design of personal learning environments, contextual recommendations at various levels of the learning process, tracking affective states of learners, harmonization of educational and technological standards, processing big data for learning purposes, predicting

[2] http://adenu.ia.uned.es/workshops/pale2014/format.htm
[3] http://ijaied.org/journal/cfp/

student outcomes, adaptive learning assessment, and evaluation of personalized learning solutions.

From the past experience, we have identified new research areas of interest to complement the previous ones. Nowadays there are new opportunities for building interoperable personalized learning solutions that consider a wider range of learner situations and interaction features in terms of physiological and context sensors. However, in the current state of the art it is not clear how this enhanced interaction can be supported in a way that positively impacts on the learning process. In this context, suitable user modeling is required to understand the current needs of learners. There are still open issues in this area, which refer to providing open learner models in terms of standards that cover the extended range of available features and allow for interoperability with external learning services as well as taking advantage of the integration of ambient intelligence devices to gather information about the learner interaction in a wider range of learning settings than the classical desktop computer approach.

Therefore, these new features are paving the way to other related topics that are to be considered in the learner modeling, including affective states of the learner as well as changing situations in terms of context, learners' needs and their behavior. Another broad research area addresses personalization strategies and techniques, considering not only the learner model, but the whole context of the learning experience, including the various technological devices that are available in the particular situation.

In this workshop edition we drew attention to sharing and discussing the current research on how user modeling and associated artificial intelligent techniques contextualize the world and provide the personalization support in a wide range of learning environments, which are increasingly more sensitive to the learners and their context, such as: intelligent tutoring systems, learning management systems, personal learning environments, serious games, agent-based learning environments, and others. We are especially interested in the enhanced sensitivity towards learners' interactions (e.g., sensor detection of affect in context) and technological deployment (including web, mobiles, tablets, tabletops), and how this wide range of situations and features may impact on modeling the learner interaction and context. Furthermore, we aim to cover the every time more demanding need of personalized learning at large-scale, such as in massive open online courses (MOOCs).

The higher-level research question addressed in this workshop edition was: “Which approaches can be followed to personalize learning environments?” It is considered in various contexts of interactive, personal, and inclusive learning environments. The topics of the workshop included (but were not limited to) the following:

- Affective computing
- Ambient intelligence
- Personalization of MOOCs
- Learning recommendation
- Learner and context awareness
- Cognitive and meta-cognitive scaffolding
- Social issues in personalized learning environments

- Open-corpus educational systems
- Adaptive mobile learning
- Successful personalization methods and techniques
- Reusability, interoperability, scalability
- Evaluation of adaptive learning environments

3 Contributions

A peer-reviewed process has been carried out to select the workshop papers. Three members of the Program Committee with expertise in the area have reviewed each paper. As a result, 6 submissions (out of 8) were accepted, which discuss ideas and progress on several interesting topics, such as adapting the study plan (with highlighting), student's performance (i.e., academic distress), self-regulated learning skills, interoperability in learner modelling by integrating standards (i.e., IMS specification), confusion detection by monitoring mouse movements in a computer game, and knowledge acquisition of mathematical concepts.

Tintarev et al. [1] focus on the effect of emphasis adaptation in a study plan, which is represented as a workflow with prerequisites. They compare the effectiveness of highlighting when the adaptation was correct (participants responded quicker and more correctly), and when it did not highlight the most relevant tasks (detrimental effect). They found that false statements took longer to process than positive statements (deciding about things that were not in the plan), but also surprisingly had lower error rates than positive statements. In their view, these findings imply that errors in the adaptation are harmful, and may cause students to incorrectly believe that they do not need to do certain tasks.

Greer et al. [2] present SARA, the Student Advice Recommender Agent, which is similar to an early alert system, where predictive models of learners' success combined with incremental data on learners' activity in a course are used to identify students in academic distress. SARA can detect when the student is struggling academically and then provides notifications with a personalized advice how to get back on track. The system represents a scalable advice personalization environment in large university courses and delivers weekly advices. The authors have observed a significant year over year improvement in unadjusted student grades after the SARA's advice recommender was implemented in a 1200-student freshman STEM course.

Tang and Yacef [3] address the challenge of time and environment management. They report on their experience with a leading vocational education provider in Australia (i.e., training of specific skills or trades, often done part time or in personal time over a lengthy period) who is transitioning from classroom-based training to a pilot e-learning system. They present the key lessons learned and the prototype goal-setting and time management interface designed to improve user self-regulation. A growing body of evidence suggests that these self-regulating skills are a key determinant in learning performance and can be improved with computer aided support, increasing engagement and motivation of trainees.

Chacón-Rivas et al. [4] identify open issues when it comes to integrate the information from the learner activity in standards-based learner models, which covers learning styles, competences, affective states, interaction needs, context information and other learner´s characteristics. In particular, there are standards that can be used to cover several of the subjects to be integrated into those models, such as IMS-LIP, IMS-RDCEO, IMS-AFA. Authors present their on-going work in implementing a learner model that aims at providing a holistic user modelling perspective, which is able to hold and collects all relevant information, thus supporting its real-life usage. This approach is expected to facilitate interoperability and sustainability, while still research needs progressing where representation and management is required.

Pentel [5] describes an unobtrusive method for user confusion detection by monitoring mouse movements. A special computer game was designed to collect mouse logs. Users' self-reports and statistical measures were used in order to identify the states of confusion. Mouse movement's rate, full path length to shortest path length ratio, changes in directions and speed were used as features in the training dataset. Support Vector Machines, Logistic Regression, C4.5 and Random Forest were used to build classification models. Those models generated by Support Vector Machine yield to best classification results with fscore 0.946, thus showing that frequent direction changes in mouse movement, are good predictors of confusion.

Silva et al. [6] approach theoretical and implementation issues of a framework aimed at supporting human knowledge acquisition of mathematical concepts. They argue that personalization support can be achieved from problem statement parameters, defined during the creation of Learning Objects and integrated with the skill level of learners and problem solution difficulty. The last two are formally defined as algebraic expressions based on fundamental principles derived from extensive consultations with experts in pedagogy and cognition. Their implemented prototype framework, called ADAPTFARMA, includes a collaborative authoring and learning environment that allows short- and long-term interactions.

4 Conclusions

In this 5th edition of PALE contributions address several gaps identified in the state of the art, such as adapting the study plan (with highlighting), student's performance (i.e., academic distress), self-regulated learning skills, interoperability in learner modelling by integrating standards (i.e., IMS specification), confusion detection by monitoring mouse movements in a computer game, and knowledge acquisition of mathematical concepts..

Nevertheless, other issues remain open such as the integration of ambient intelligence devices to gather information about the learner interaction in a wider range of learning settings than the classical desktop computer approach, aimed to enhance the sensitivity towards learners' interactions through diverse technological deployments (including web, mobiles, tablets, and tabletops), impacting on modeling the learner interaction and context. We expect that future editions in PALE can progress on aforementioned directions.

Acknowledgements

PALE chairs would like to thank the authors for their submissions and the UMAP workshop chairs for their advice and guidance during the PALE workshop organization. Moreover, we also would like to thank the following members of the Program Committee for their reviews (in alphabetical order): Miguel Arevalillo, Mihaela Cocea, Sabine Graf, Peter Henning, Mirjana Ivanovic, Jelena Jovanovic, Iolanda Leite, Noboru Matsuda, Alexander Nussbaumer, Alexandros Paramythis, Lubomir Popelinsky, Elvira Popescu, Sergio Salmeron-Majadas, Natalia Stash, Christoph Trattner, Carsten Ullrich, Stephan Weibelzahl, Michael Wixon.

The organization of the PALE workshop relates and has been partially supported by the following projects: BOOST: Business perfOrmance imprOvement through individual employee Skills Training, LEARNING LAYERS: Scaling up Technologies for Informal Learning in SME Clusters (FP7 ICT-318209), MAMIPEC: Multimodal approaches for Affective Modelling in Inclusive Personalized Educational scenarios in intelligent Contexts (TIN2011-29221-C03-01), MARES: Multimodal and Machine learning techniques to recognize emotions in educational settings (TIN2011-29221-C03-02), Supervised Educational Recommender System (VEGA 1/0475/14), and Virtual Learning Software Lab for Collaborative Task Solving (KEGA 009STU-4/2014).

References

1. Tintarev, N., Green, M., Masthoff, J., Hermens, F. Benefits and risks of emphasis adaptation in study work flows. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 8-15 (2015)
2. Greer, J., Frost, S., Banow, R., Thompson, C., Kuleza, S., Wilson, K., Koehn, G.. The Student Advice Recommender Agent: SARA. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 16-23 (2015)
3. Tang, L.M., Yacef, K. Personalising e-Learning Systems: Lessons learned from a vocational education case study. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 24-30 (2015)
4. Chacón-Rivas, M., Santos, O.C., Boticario, J.G. Modeling Learner information within an Integrated Model on standard-based representations. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User

Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 31-39 (2015)
5. Pentel, A. Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 40-45 (2015)
6. Silva, R.C., Direne, A.I., Marczal, D. Using Problem Statement Parameters and Ranking Solution Difficulty to Support Personalization. In proceedings of the 5th Workshop on Personalization Approaches for Learning Environments (PALE 2015). Kravcik, M., Santos, O.C., Boticario, J.G., Bielikova, M., Horvath, T. (Eds.). 23rd conference on User Modeling, Adaptation, and Personalization (UMAP 2015), CEUR workshop proceedings, this volume, 46-51 (2015)

Benefits and risks of emphasis adaptation in study workflows

Nava Tintarev[1], Matt Green[1], Judith Masthoff[1], and Frouke Hermens[2]

[1] Department of Computing Science, University of Aberdeen,
[2] School of Psychology, University of Lincoln,
n.tintarev@abdn.ac.uk, matt@mjglab.org, j.masthoff@abdn.ac.uk,
frouke.hermens@gmail.com

Abstract. This paper looks at the effect of highlighting in a study plan, represented as a workflow with prerequisites. We compare the effectiveness of highlighting when the adaptation was correct (participants responded quicker and more correctly), and when it did not highlight the most relevant tasks (detrimental effect). False statements took longer to process than positive statements (deciding about things that were not in the plan), but also surprisingly had lower error rates than positive statements. These findings imply that when the system makes errors in the adaptation this is harmful, and may cause students to incorrectly believe that they do not need to do certain tasks.

Key words: Visualization · Plan presentation · Study workflows · User-centered evaluation · Highlighting · Emphasis adaptation

1 Introduction

In adaptive learning systems, methods such as link annotation and hiding have been used to help learners navigate learning materials [1]. One of the challenges has been to consider pre-requisites for learning modules, guiding students and supporting them in identifying which materials they should study next. One such approach is the traffic light metaphor ([2, 3]) which indicates differences between recommended reading and material the student is not yet ready for.

The approaches used in such systems (e.g., ISIS-tutor [4], ELM-ART [2], KnowledgeSea [5]) are often non-sequential (e.g., they jump between subjects) and for this reason may not give users an overview of, and an understanding of the pre-requisites, in the study plan. The visual information seeking mantra states: "Overview first, zoom and filter, then details-on-demand." [6]. Supplying an overview may help students to plan their study, and such overviews have been found to improve the efficiency of hypertext [7–9].

For this reason, this paper investigates the presentation of study plans. A study plan can be seen as a workflow with each step representing a study task, and the edges between these tasks representing the transition that occurs once each task is complete. At times several tasks, or prerequisites, must be completed before proceeding to the next step. The path through the workflow can be personalized for each student, and adapted as their goals change.

Previous work on visualizing plans has looked at filtering graphs by content [10], and applying fish-eye views to grow or shrink parts of a graph [11]. There is also research on verbalizing and explaining plans generated by A.I. planning systems [12, 13].

This paper studies the use of emphasis of relevant paths through a workflow as a means to improve the effectiveness of information presentation. This personalized path emphasizes all of the relevant tasks, including all prerequisites.

2 Experiment

In previous (unpublished) studies we found no significant difference in cognitive load (measured in a dual-task paradigm) between adaptations that included highlighting and those that did not. It is possible that the type of adaptation of plans is simply not effective. The current experiment investigates if an emphasis of dependent tasks, using border highlighting, affects participant performance. Since an adaptive system may sometimes adapt to an incorrect inferred goal, we also investigate the effect of such 'unhelpful' highlighting as well, in relation to correct adaptation in 'helpful' highlighting.

We investigate a) whether highlighting had an effect on errors and response times; and b) if so, whether performance was improved by the mere presence of highlighting or if there was a difference when highlighting was for a different path through the plan than for the current learning goal (unhelpful highlighting). In the current experiment we compare the performance (response time and accuracy) for plans with no highlighting, with helpful and unhelpful highlighting.

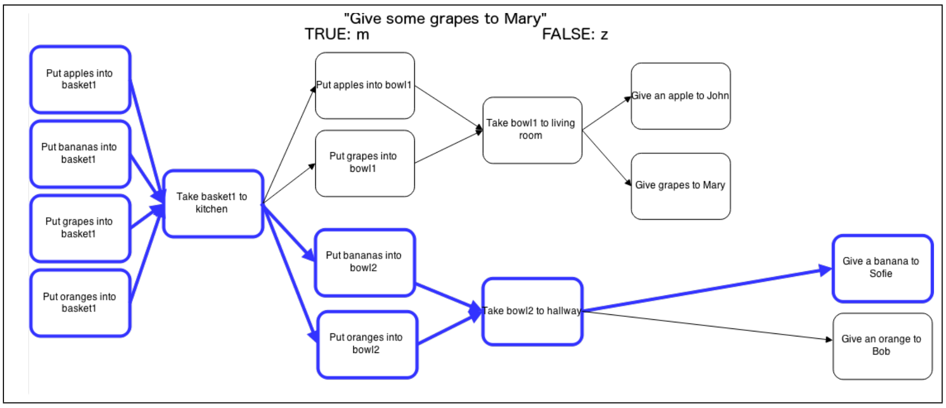

Fig. 1: Material from one experimental trial: plan and statement. The highlighting is *unhelpful* for a statement about grapes, while the highlighting is for bananas. The statement ("Give some grapes to Mary") is *true* since the step with grapes nevertheless is present in the plan.

2.1 Experimental design

The experiment employs a full within-participants design, with all of the participants seeing all of the variants, in randomized order.

The independent variables are: i) **htype** - whether the components of the plan that are highlighted constitute no highlighting, helpful highlighting, or unhelpful highlighting; and ii) **true value** - whether the statement (e.g., "You should study course x" or "Give some grapes to Mary") is true or false in relation to the plan. The dependent variables are: a) **Response time** - the time taken to respond to the statement about the plan; and b) **Errors** - the proportion of incorrect responses.

In the introduction screen participants were given the following instructions: *"On each screen you will be shown a plan and statement about the plan. For now, press any key to start a short practice session. This experiment studies different ways of presenting sequences of actions, or plans. You will be asked to press [true_key] if the statement is true and [false_key] if the statement is false."*

In each trial participants saw a statement and a plan (see Figure 1), and pressed a key to respond whether the statement was true or false for that plan. The keys for true/false were randomly assigned to either 'm' or 'z'. After each statement, participants were given quick feedback as a red or green dot with feedback text (either "correct" or "incorrect") before going on to the next trial.

Participants first completed a practice session (6 trials) before going on to the experimental trials (144). In addition to the independent variables we also included 6 different categories of items (farm, groceries, sports, stationery, furniture (filler), tableware (filler)), with 4 items in each (e.g., apple, grape, banana and orange). This gave a total of 144 trials: 6 categories * 4 items * 3 types of highlighting * 2 truth values. A break was inserted half way through to avoid participant fatigue.

2.2 Materials

Plans. The experiment uses an algorithm introduced and implemented in [14] that selects which steps to highlight, including prerequisite, or intermediate tasks that are required to reach an outcome. Given a study concept, the algorithm first selects all tasks that are related to a learning outcome. The algorithm then finds all paths between each pair of the selected tasks. All tasks on these paths are then added into the list of selected tasks. Lastly, the algorithm inspects all the selected tasks and checks if any of them require completion of other tasks.

While the system supports filtering by multiple items (e.g., apple, and banana) or object types (e.g., fruit), in this experiment it is applied to filtering by one object at a time (e.g., apple). The algorithm selects all the steps an item is directly involved in, as well as any prerequisite steps that may be required to achieve the final learning goal.

The plans were all of the same shape as Figure 1, and thus balanced in terms of width and number of steps, with only the names of the tasks replaced.

The categories used in the experimental trials were: farm, groceries, sports, stationery, furniture (filler), tableware (filler). For each trial and plan four objects were described, for example in the fruit category plans the following items were described: apple, pear, grapes, and banana. The range of domains was selected to minimize the effects of prior knowledge, and to ensure the generalizability of results.

Statements. The statements used in the experiment had four properties: category (e.g., fruit), item (e.g., apple), and the type of highlighting they were associated with (e.g., helpful, unhelpful, no highlighting) a truth value for the statement (i.e., whether or not the statement is true according to the plan). Figure 1 gives an example of a statement for the fruit category. The plan is highlighted for bananas, but the statement is about grapes, so this is unhelpful highlighting. The statement and its truth value are *true*; this is in the plan, but not for the current learning goal.

2.3 Hypotheses

H1: Helpful highlighting stimuli lead to faster response times than the no highlighting and unhelpful highlighting conditions.
H2: Helpful highlighting stimuli lead to fewer errors than the no highlighting and unhelpful highlighting conditions.
H3: True statements will lead to faster response times than false statements.
H4: True statements will lead to fewer errors than the false statements.

2.4 Results

The statistical analyses reported below were carried out in the mixed effects regression framework using the R package *lme4* [15]. This method is well suited for studying repeated measures (several trials per participant), it also allows us to model individual variations between subjects as might be expected by variation in working visual memory [16]. [17] and [18] describe the analysis method and its relationship to ANOVA. Items in the filler categories were excluded from analysis.

Participants. Participants were thirty-seven psychology undergraduate students, participating in a psychology experiment as part of their coursework. Data from two participants were removed because their average response times or error rates were more than 3 SDs away from the mean across participants.

H1: Helpful highlighting stimuli lead to faster response times than the no highlighting and unhelpful highlighting conditions. Table 1 summarizes the results, means are calculated by participant and response times were log normalized. The trend is for helpful highlighting to result in quicker response

htype	times	times.sd	errors	errors.sd
unhelpful	8.00	0.29	0.08	0.10
no	8.02	0.27	0.05	0.07
helpful	7.86	0.33	0.05	0.08

Table 1: Response times in log(ms), and error rates by subject average.

times than both unhelpful and no highlighting, as predicted by H1. Three models were built for complete two-way comparisons: helpful-unhelpful (Table 2), no-helpful (Table 3), no-unhelpful (Table 4) highlighting. There is a significant difference between helpful highlighting and the other two conditions ($p <= 0.01$), but no significant difference between unhelpful and no highlighting[3]. *H1 is supported - helpful highlighting decreases response times.*

	Estimate	Std. Error	t value	Pr($> \|t\|$)
(Intercept)	9.08	0.05	169.81	0.00
htype	-0.14	0.04	-3.28	0.01
true value	-0.17	0.03	-4.84	0.00
htype*true value	-0.01	0.05	-0.27	0.79

Table 2: Model for response times in log(ms) comparing unhelpful and helpful highlighting.

	Estimate	Std. Error	t value	Pr($> \|t\|$)
(Intercept)	9.08	0.05	190.68	0.00
htype	-0.13	0.04	-3.16	0.01
true value	-0.12	0.03	-3.55	0.00
htype*true value	-0.06	0.05	-1.28	0.20

Table 3: Model for response times in log(ms) comparing no and helpful highlighting.

H2: Helpful highlighting stimuli lead to fewer errors than the no highlighting and unhelpful highlighting conditions. Table 1 also summarizes the mean error rates. Overall, the error rates are very low, with only 5-8% errors on average. There are most errors in the unhelpful condition. Three models were built for complete two-way comparisons: helpful-unhelpful (Table 6), no-helpful highlighting (Table 7), no-unhelpful (Table 8). There is a significant difference

[3] Significance levels given using R package lmerTest, `http://cran.r-project.org/web/packages/lmerTest/index.html`, retrieved April 2015

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	9.08	0.05	189.15	0.00
htype	0.00	0.06	0.03	0.98
true value	-0.12	0.03	-3.51	0.00
htype*true value	-0.05	0.05	-0.98	0.33

Table 4: Model for response times in log(ms) comparing no and unhelpful highlighting.

between the helpful highlighting and the other two conditions ($p <= 0.01$), but not between the no and unhelpful highlighting conditions. *H2 is supported, relevant highlighting leads to fewer errors.*

H3: True statements will lead to faster response times than false statements. Table 5 summarizes the response times for true and false statements, with faster responses for true trials compared to false ones. In Tables 2, 3, and 4 we also see a significant difference for each type of highlighting ($p << 0.01$). *H3 is supported: response times are reliably faster for true statements compared to false statements.*

true value	times	times.sd	errors	errors.sd
false	8.04	0.31	0.05	0.08
true	7.88	0.27	0.07	0.09

Table 5: Response times as log(ms) and error rates by true value.

H4: True statements will lead to fewer errors than the false statements. Table 5 summarizes the error rates for true and false statements, with *more* errors for true statements. Tables 6, 7, and 8 show that this difference is significant at $p << 0.01$ for all types of highlighting. Further, we found a significant interaction between type of highlighting and truth value in the comparison between unhelpful and no highlighting (p < 0.01). *H4 is not supported: statements that are true led to **more** errors compared to false statements.*

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	9.08	0.05	169.81	0.00
htype	-0.14	0.04	-3.28	0.01
true value	-0.17	0.03	-4.84	0.00
htype*true value	-0.01	0.05	-0.27	0.79

Table 6: Model for errors comparing unhelpful and helpful highlighting.

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	9.08	0.05	190.68	0.00
htype	-0.13	0.04	-3.16	0.01
true value	-0.12	0.03	-3.55	0.00
htype*true value	-0.06	0.05	-1.28	0.20

Table 7: Model for errors comparing no and helpful highlighting.

	Estimate	Std. Error	t value	Pr(> $\|t\|$)
(Intercept)	1.94	0.02	121.22	0.00
htype	0.00	0.01	0.29	0.77
true value	-0.05	0.01	-3.46	0.00
htype*true value	0.05	0.02	2.53	0.01

Table 8: Model for errors comparing no and unhelpful highlighting

2.5 Discussion

As predicted we found the unhelpful highlighting increased errors and response times compared to helpful highlighting (or to even no highlighting at all). However, contrary to expectations (H4), we found that statements that are true led to *more errors* compared to false statements even if these evaluations were quicker. This suggests that participants "learn" to rely on the highlighting and anticipate the relevant parts of the plan to be highlighted, when in fact this is only true some of the time. This is further corroborated by a significant interaction between type of highlighting and truth value in the comparison between unhelpful and no highlighting. That is, participants made most errors when the statement was true, but the highlighting of the plan was unhelpful. If participants learned to rely on the highlighting this could also explain the longer response times for false statements, as participants may first look for confirmation in the highlighted parts of the plan before performing a more thorough search.

3 Conclusion and future work

Border highlighting of prerequisite steps is an automatic adaptation in the system we are currently designing. The study described in this paper identified this adaptation as helpful, and confirmed the importance of getting the adaptation right: incorrect highlighting decreased effectiveness. We also found that creating a reliance on highlighting could have particularly adverse effects when learners are trying to answer statements that are true, but the highlighting is incorrect. These findings imply that when the system makes errors in the adaptation this is harmful, and may cause students to incorrectly believe that they do not need to do certain tasks.

The next step in this research is to compare *hiding* with *highlighting*, and investigate if individual differences in visual working memory affect which of the adaptations is more effective. We also plan to study the value of highlighting adaptation in other visual representations of educational content such as graphs.

References

1. Brusilovsky, P.: Adaptive navigation support: From adaptive hypermedia to the adaptive web and beyond. PsychNology Journal **2** (2004) 7–23
2. Brusilovsky, P., Schwarz, E., Weber, G.: ELM-ART: An intelligent tutoring system on world wide web. In: Intelligent Tutoring Systems. (1996)
3. Weber, G., Kuhl, H.C., Weibelzahl, S.: Developing Adaptive Internet Based Courses with the Authoring System NetCoach. In: Hypermedia: Openness, Structural Awareness, and Adaptivity. Springer Berlin Heidelberg (2002) 226–238
4. Brusilovsky, P., Pesin, L.: Adaptive navigation support in educational hypermedia: An evaluation of the isis-tutor. Journal of computing and Information Technology **6(1)** (1998) 27–38
5. Jae-Kyung, K., Farzan, R., Brusilovsky., P.: Social navigation and annotation for electronic books. In: Research advances in large digital book repositories. (2008) pp. 25–28
6. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In: Symposium on Visual Languages. (1996) 336–343
7. Chen, C., Rada, R.: Interacting with hypertext: A meta-analysis of experimental studies. Human-Computer Interaction **11** (1996) 125–156
8. McDonald, S., Stevenson, R.J.: Disorientation in hypertext: the effects of three text structures on navigation performance. Applied Ergonomics **27** (1996) 61–68
9. Monk, A.F., Walsh, P., Dix, A.J.: A Comparison of Hypertext, Scrolling and Folding as Mechanisms for Program Browsing. In: People and Computers IV. Cambridge University Press (1988) 421–435
10. Henry, T.R.: Interactive graph layout: The exploration of large graphs. PhD thesis, The University of Arizona (1992)
11. Sarkar, M., Brown, M.H.: Graphical fisheye views of graphs. In: Proceedings of the SIGCHI conference on Human factors in computing systems, ACM (1992) 83–91
12. Bercher, P., Biundo, S., Geier, T., Hoernie, T., Ritcher, F., Schattenberg, B., Nothdurft, F.: Plan, repair, execute, explain - how planning helps to assemble your home theatre. In: AAAI. (2014)
13. Mellish, C., Evans, R.: Natural language generation from plans. Computational Linguistics **15** (1989) 233–249
14. Tintarev, N., Kutlak, R., Masthoff, J., van Deemter, K., Oren, N., Vasconcelos, W.: Adaptive visualization of plans. In: UMAP'14 demo track. (2014)
15. Bates, D., Maechler, M., Bolker, B., Walker, S.: lme4: Linear mixed-effects models using Eigen and S4. (2013) R package version 1.0-4.
16. Conati, C., Merten, C.: Eye-tracking for user modeling in exploratory learning environments: An empirical evaluation. Know.-Based Syst. **20(6)** (2007) 557–574
17. Jaeger, T.F.: Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of memory and language **59**(4) (2008) 434–446
18. Baayen, R., Davidson, D., Bates, D.: Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language **59**(4) (2008) 390–412 Special Issue: Emerging Data Analysis.

The Student Advice Recommender Agent: SARA

Jim Greer, Stephanie Frost, Ryan Banow, Craig Thompson, Sara Kuleza, Ken Wilson, and Gina Koehn

University of Saskatchewan, Saskatoon, Canada
{firstname}.{lastname}@usask.ca

Abstract: SARA, the Student Advice Recommender Agent is a system somewhat like an early alert system, where predictive models of learners' success combined with incremental data on learners' activity in a course can be used to identify students in academic distress. With SARA, rather than give alerts to academic advisors or professors, we provide personalized advice directly to students. An advice string – "A note from SARA" is prepared for each student every week in a semester-long course. The system attempts to direct students to appropriate learning supports and resources according to their individual needs. We have observed a significant year over year improvement in unadjusted student grades after the SARA's advice recommender was implemented in a 1200-student freshman STEM course.

Keywords: early alert, personalized advice, persona, recommender agent

1 Introduction

Early alert systems for students at academic risk have been in use for several years. In such systems, students who seem to be struggling in a course, as evidenced by lower term grades, minimal engagement in learning management system (LMS) activity, or low attendance may be issued warnings or alerts [1]. In most systems, instructors are involved in directing the delivery of alert messages. In some systems, these alerts are also issued to academic advisors (as in Starfish Early Alert or Ellucian Student Success) so that follow up appointments with an advisor or learning specialist can be booked if the advisor so wishes. For the most part, students who seem to be minimally engaged or who are falling behind in coursework, or who are failing intra-term assessments are targeted for additional interventions.

We have taken a different approach to a somewhat similar problem. The problem we are trying to address is how to best assist and support learners during a course when the benefits of big data can be put to work. That is, if we know about the students' academic history, personal history (including demographics), and current activity (such as progress in a course and other related activity pertinent to academic success), what could we do to help? Help would not be for only the struggling student, but for the successful and exceptional students too. The approach we have taken is to construct individualized, personalized advice for students in a large courses on the basis of their academic, personal, and activity profiles (including current progress in the course). We have developed and implemented the Student Advice Recommender Agent (SARA), which generates and delivers an "advice string" to each student each week throughout the term.

Predictive models of student success in the course are computed based on past academic performance and demographic student data. Advice string templates are constructed by instructional experts, focusing on available supports and resources, words of encouragement, and content specific matters. These advice templates are personalized (adjusted/adapted) based upon combinations of student demographic and student activity data. The engineering of advice strings and conditional adaptation is aided by focusing on personas of students who are predicted to fail, pass or excel (as mapped out in [2]). The advice strings are then delivered as learning alerts to each and every student in a course. The advice directs students toward help resources, help or advisory personnel, supplementary course materials, or enrichment activities, as is appropriate.

2 Enhanced Demographics

Beginning in the fall of 2013, the University of Saskatchewan initiated a project to gather enhanced demographic data about incoming freshmen. A 75-item census-style survey is issued annually to students at the time of their first registration to the University. Students are invited to disclose personal information including: goals and aspirations, anticipated and "disappointment threshold" grade point averages; living arrangements; family history of university studies; sources of financial support; expected hours to be spent working for pay, volunteering, studying, engaging in extracurricular activity; disabilities; whether they are supporting dependents; expected major; anticipated advanced or professional degrees sought; level of comfort and connectedness with campus. In addition some short standardized instruments including: a shortened version of Biggs' Study Process Questionnaire [4], the Motivated Strategies for Learning Questionnaire [5] and GRIT-S [3]. Response rates around 60% have been achieved in each of the first two years of this survey.

These data are merged into the University's student data warehouse where information is consolidated from student admissions and recruitment, student grades and academic records, access to academic support services, and learning management system activity.

This rich data repository offers an opportunity to develop comprehensive and relatively accurate predictive models of student academic achievement. The rich data repository also opens possible avenues for inappropriate and prejudicial decisions about students. Data are carefully guarded, identities are encrypted, and personally identifying variables are kept separate from other data. Strict ethical guidelines are followed in making use of the data for student modeling and advising.

3 Temporally Improving Predictive Models

After various data mining attempts using decision trees, regression models, Bayesian networks and naïve Bayes algorithms, predictive models of student academic success in specific courses, overall GPA, retention, and degree completion were derived using student data over the past 5 years. We took a closer look at a face-to-face introductory Biology class that reaches more than a thousand students (mostly freshmen) per year. A sequence of predictive models for final course grades was developed, one model for each week of the Biology course, using the data that would be known as of that week. Models were built with half the students in the 2013 Biology cohort and validated with the other half of the students. We found that log-linear regression models based upon selected demographic data and high school grades could result in good correlations with 2013 course grades (r=0.61). We also found that if LMS activity data and term grades were added in, the grade prediction improves even more as the term progresses. After the course midterm examinations, regression models correlated very highly (r=0.92) with final 2013 course grades.

When these models were applied to the 2014 students, we discovered that the models correlated very well with the 2014 students' grades (r=0.62 at the beginning of term and r=0.82 after the midterm exam). Because of the interventions associated with the introduction of SARA in 2014, we expected correlations between model and final grades might be reduced somewhat – the predictive model applied did not account for changes in 2014. This will be explained in the evaluation section below.

This predictive modeling methodology provides a temporally improving predictive model of student academic achievement in a single course. The model gives a relatively accurate estimate of student success. Factors in the model that offer the greatest degree of predictive power include: actual assessment grades, High school GPA and Biology grades, whether the student was intrinsically or extrinsically motivated, whether the student was a deep versus surface learner, and whether or not the student was the first in their family to attend university.

The methodology also identifies which demographic variables may be considered as risk factors in student retention and success. Using risk factor variables we have constructed a number of personas of canonical successful or less successful students (cf. Brooks & Greer, 2014). Figure 1 shows three sample personas of students.

Student 1	Student 2	Student 3
- rural high school - average admission GPA - first in family at univ - living in univ. residence - high grit score - surface learner tendency predicted to receive D grade	- mature student - has dependents at home - 20 hours per week job - returning - 5 yrs away - deep learner tendency predicted to receive a C grade in course	- exceptional admission GPA - attended top high school - recently settled immigrant - living at home with family - surface learning tendency - aspirations for grad school predicted to receive a B grade in the course

Figure 1: Sample personas of learners

4 CREATION OF ADVICE STRINGS

The personas help our instructional designers, instructors, and academic support specialists construct advice templates that could be tailored for individuals in a particular persona group. Engineering advice strings has turned out to be a fairly difficult activity. Good quality advice for students is highly contextualized – dependent on the time in the term, what is going on in the course, the content being presented, the supports and resources available outside the course, upcoming events and opportunities, news and current events. Good advice also should reflect the student's situation interests and needs, the academic risks they may face, and their determination to survive or excel. For some students useful advice is a message that somebody cares about their success. For others such a message may be a threat and lead to discouragement.

Another challenge with generating advice strings is that predictions may be wrong. Some students predicted to earn an A will not do so. Some students predicted to fail may surprise everyone. We have paid a lot of attention to ensure that our advice to students "does no harm". Advice is framed as a set of positive suggestions, raising opportunities, making reference to resources and supports while offering a supportive and caring tone.

Figure 2 shows the rule for generation of advice in week 3 of our Biology course. The advice strings are written by an subject area learning specialist and they are crafted for stereotypical students who have certain attributes or whose persona has certain features, specified as constraints. The advice string constraints are then interpreted for the attributes of students in the class and a unique advice string is produced for each student. The advice constraints in Figure 2 yield 24 different advice messages. A student will receive the one that best fits their persona. Some weeks there might be only half a dozen different advice messages and some weeks there may be hundreds of different messages. The approximately 1200 students in this Biology course each received a weekly message, tailored as much as possible for their individual context.

We see the messages from SARA as a type of mass personalization. Over the course of a full semester, the collection of weekly advice messages for each of our ~1200 Biology students is nearly unique. That is, the cumulative advice for an individual is likely to be distinct from the cumulative advice given to any other individual in the course. The largest group of individuals who received identical cumulative messages over the term was of size 10. These 10 students were students who were average in every way and for whom we had no enriched demographic data (they did not complete the entry census).

Our students use the BBLearn LMS for access to course materials, lecture recordings, presentation slides, and online quizzes. In order to be sure that students see the weekly advice from SARA (knowing that students tend not to read email), we added an iframe to the course home page where each student sees their weekly "A Note From SARA" immediately as they connect to BBLearn. In addition, an LTI component has been inserted into BBLearn, where all their advice strings for the term so far are available for review, where the advice that SARA gives to others can be browsed (identities hidden), and an opportunity is available to rate the usefulness of or comment upon SARA's most recent piece of advice.

<table>
<tr><td>Week: 3</td><td>Order: 0</td><td>Condition: Predicted GPA < 60%</td></tr>
<tr><td colspan="3">Finding balance is important and is often one aspect of academic success that doesn't receive much focus. Regular practice and review of course materials have been proven to help retention of information. This is why completing weekly quizzes, preparing for your labs, and regularly reviewing your textbook and lecture notes will be beneficial to your learning in Biol120! Once you've addressed your academic responsibilities then you can take time for other responsibilities and personal time. Check out this video for additional tips on how to find balance this term.</td></tr>
<tr><td>Week: 3</td><td>Order: 0</td><td>Condition: Predicted GPA 60% - 80%</td></tr>
<tr><td colspan="3">Finding balance is very important and often one aspect of academic success that doesn't receive much focus. Self-awareness, knowing your goals, and self-reflection: these things will be helpful guides in your journey to finding balance. Try to schedule your time so that you are accounting for time to study, time for your other responsibilities, and personal time. Also keep in mind that this schedule will need to be flexible to account for preparation before important deadlines and exams. For additional tips on how to find balance, check out this video.</td></tr>
<tr><td>Week: 3</td><td>Order: 0</td><td>Condition: Predicted GPA > 80%</td></tr>
<tr><td colspan="3">One aspect of academic success that doesn't receive much focus is how to find balance between the time spent on your academics and the time dedicated to your other life responsibilities. Self-awareness, knowing your goals, and self-reflection: these things will be helpful guides in your journey to finding balance. Scheduling time to study is important, but so too is taking time for enjoying other things in life. Check out this video for additional helpful tips on how to find balance this term.</td></tr>
<tr><td>Week: 3</td><td>Order: 1</td><td>Condition: Low # hours spent studying (according to survey response and course load)</td></tr>
<tr><td colspan="3">You indicated on the entry census that you intend to allocate a <i>below average</i> amount of study time on your University courses. Making time to review and study the material is a very important aspect of your overall academic success at University. Setting academic goals and constructing a plan to achieve those goals can be very helpful in guiding how much time you need to prioritize to your studies. Check out these helpful guides to Goal Setting and
 Time Management.</td></tr>
<tr><td>Week: 3</td><td>Order: 1</td><td>Condition: Medium # hours spent studying (according to survey response and course load)</td></tr>
<tr><td colspan="3">You indicated on the entry census that you intend to allocate an average amount of time to studying your University courses. This is a great step towards helping you find balance while also achieving your academic goals! Understanding more about your learning process may help you to use your study time more efficiently - check out this resource on Learning Styles for further information.</td></tr>
<tr><td>Week: 3</td><td>Order: 1</td><td>Condition: High # hours spent studying (survey response and course load)</td></tr>
<tr><td colspan="3">You indicated on the entry census that you would be spending <i>more than</i> an average amount of time studying for your University courses. Did you know that often it's not how long you study, but how efficiently you study, that makes the biggest difference? Using a variety of study methods and taking frequent breaks can help to increase your retention of the material you are studying. To help you use the time you study most efficiently, check out this resource on Learning Styles that might help you understand more about your learning process.</td></tr>
<tr><td>Week: 3</td><td>Order: 2</td><td>Condition: Working 13-20 hours per week or More and Predicted GPA < 80%</td></tr>
<tr><td colspan="3">Working a lot of hours can place extra stress on you, especially as exam times draw near. Be sure you can find a balance between work and your academics. This may mean asking for some time off during heavier study periods. Keep in mind that dropping a few shifts may be more cost and time effective than having to re-take a course!</td></tr>
<tr><td>Week: 3</td><td>Order: 3</td><td>Condition: Predicted GPA < 60</td></tr>
<tr><td colspan="3">Volunteering can provide enriching experiences that can be beneficial to your academic experience and your future career. However, making sure that you are achieving your academic goals needs to be your first priority. After you receive your mid-term grades, if you're on track with success, you may want to consider getting involved with fun and interesting volunteer experiences.</td></tr>
<tr><td>Week: 3</td><td>Order: 3</td><td>Condition: Predicted GPA 60% - 80%</td></tr>
<tr><td colspan="3">One way to achieve balance is to get involved in interesting volunteer opportunities. Volunteering can provide enriching experiences that can be beneficial to your academic experience and your future career. There are lots of ways to get involved on and off campus. To find volunteer positions, check out these volunteering opportunities. If you're interested in gaining valuable experience in a Biology related field, find out about possible volunteering positions by speaking to your Instructor or TA, or contacting the undergraduate's Biology Club.</td></tr>
<tr><td>Week: 3</td><td>Order: 3</td><td>Condition: Predicted GPA 60% - 80%</td></tr>
<tr><td colspan="3">Volunteering can provide enriching experiences that can be beneficial to your academic experience and your future career. Find volunteer opportunities that might be of interest to you by checking out these volunteering opportunities. In addition, if you're interested in gaining valuable experience in a Biology related field, find out about possible volunteering positions by speaking to your Instructor or TA, or contacting the undergraduate's Biology Club.</td></tr>
</table>

Figure 2. Different Advice Strings generated by SARA in week 3

5 Evaluation of SARA

As with many educational interventions initiated in a large course, controlled experimentation is difficult to implement. In the freshman Biology course of 2014 many factors remained the same as in prior offerings. The instructional team, the course objectives and evaluation rubric, the laboratories, and the types of students remained much the same. Key differences in 2014 included a new requirement of weekly online quizzes, the introduction of SARA's advice, and the addition of optional-attendance peer-led study groups.

A remarkable difference in achievement was detected when comparing unadjusted final grades in 2014 against the two previous years. The mean course grade increased by 2.57 percentage points (t-test $p<.00001$). More remarkably, the number of D and F grades in 2014 decreased by 25.3% over the previous year and the number of A grades increased by 28.3% over the previous year. In the ~1200 student course, 100 fewer students scored D and F grades in 2014 than in previous years, indicating a potential boost in student retention.

Students' predicted grades for 2014 were also compared against their observed grades. The predicted model at the beginning of term expected a class average 3.8 percentage points lower than the observed average. After the midterm, the improved predictor expected a class average that turned out to be 1.8 percentage points lower than what was observed. These significantly lower predictions may indicate that the predictive model was lacking. Of course we hoped that the model would underpredict grades if there was indeed a positive effect due to the new teaching interventions.

The fact that correlations between predicted and observed grades remained comparable to the prior year indicates that there was a general lift to grades in 2014 (no significant change in slope of the correlation line). There was clearly an across the board increase in grades in 2014. Instructors believed this was due primarily to the introduction of the weekly quizzes.

An end of term questionnaire was given to determine students' reactions to changes made in the class. Forty four percent of the students completed the survey but nearly half of those students chose to remain anonymous so their responses could not be linked with grades and demographics. Students who completed the survey and gave their identities tended to have somewhat higher grades than those who did not give identities but the patterns of responses among those who did and did not reveal identities were not significantly different. Among the survey items students were asked whether they read the weekly advice from SARA or ignored it and whether or not they appreciated SARA's advice. Only 1/3 of students said they paid attention to SARA's advice, while 40% of the students said they appreciated receiving advice from SARA. Neither students who paid attention to SARA's advice or students who appreciated receiving advice from SARA showed differences in unadjusted final grades or in differences between expected (pre-midterm) and actual grades as compared to students who ignored SARA. Similarly there was no difference in grades between those who chose to participate in the peer-led study groups and those who did not.

Perhaps in an ideal world every student should embrace and act upon SARA's advice. But it is important to realize that even if only a few students listen to and are helped by SARA's advice, and that the advice is helpful, a significant shift in achievement can (and did) occur.

One important measure related to SARA's advice did show promise. As stated above, the post-midterm predictor, correlated with unadjusted final grades at 0.82. We examined more closely the error in this predictor (the difference between actual unadjusted final grade and the post-midterm prediction). Students who regularly read the weekly advice from SARA scored significantly higher than the predictor (4.6 percentage points), while students who did not read SARA's advice scored very near the predicted grade. This could be taken to mean that student's who regularly read SARA's advice had achievement levels higher than was expected.

Given our study and its limitations, we cannot directly attribute the overall achievement improvements in this offering of the course to SARA alone. Improvements could have been due to the weekly quizzes, which every student was required to complete, or the study groups, or SARA or some combination. The Biology instructors were very happy with the outcomes in their course. They were convinced that more learning occurred than in past years and that the combination of interventions was a great success. Yet to more fully understand the impact of SARA, we are contemplating dropping the SARA advice from the Biology course next year and watching to see if the increase in level of achievement persists when only mandatory quizzes and optional study groups are in play. In order to continue our research into SARA we hope to expand SARA's reach into other large freshman STEM courses in the upcoming year (perhaps Physics, Chemistry or Engineering).

6 Conclusions

SARA provides a scalable advice personalization environment in large university courses. In our first offering of a large course using SARA, student achievement improved over previous years and students on average achieved significantly higher grades than our predictive model (based on prior years' features) would have expected. There is some evidence that the improvement could be caused by SARA's weekly advice, but further research is needed to confirm such a claim.

One of the persistent dangers that comes along with predictive modeling of learners is the possibility of prejudicial treatment that may bring negative consequences to some learners. When instructors come to know, for example, that a particular student has a very low probability of successfully completing a course, the instructor may decide to minimize help and support for that learner, turning more attention toward those with higher likelihood of success. Likewise an instructor may (consciously or subconsciously) privilege the students who would most likely pursue advanced courses or graduate programs. Of even greater concern, predictive models may be used to shape enrolment, streaming, and admission policies. It is important for instructors and especially admissions officers to understand that probabilities associated with predictive estimates make it highly inappropriate to assume the predictive tendencies of a group will apply to any particular individual in the group.

Finally, it is important to remember that the goal of improved learning is to make gains in academic achievement. If using predictive models of achievement based on past cohorts to inform decisions about future cohorts, one must be prepared to accept that the predictive models may be a little less accurate than one might like. This paradox associated with modeling the state of learners as they consciously and steadily try to move beyond their current learning state is not new. The temporal dimension of predictive models that takes into account innovations or interventions in teaching and academic support is vital to our growing understanding of the learning process for distinct individual learners.

Acknowledgements

We would like to recognize the contributions, creativity, and willingness to flex of the three faculty members who taught this Biology course in 2014: Susan Kaminskyj, Ken Wilson and Jorge Chedrese.

References

1. Arnold, K. E., & Pistilli, M. D. (2012, April). Course Signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 267-270). ACM.
2. Brooks, C., & Greer, J. (2014, March). Explaining predictive models to learning specialists using personas. In *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge* (pp. 26-30). ACM.
3. Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT–S). *Journal of personality assessment*, *91*(2), 166-174.
4. Fox, R. A., McManus, I. C., & Winder, B. C. (2001). The shortened Study Process Questionnaire: An investigation of its structure and longitudinal stability using confirmatory factor analysis. *British Journal of Educational Psychology*, *71*(4), 511-530.
5. Pintrich, P. R., Smith, D. A., García, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and psychological measurement*, *53*(3), 801-813.

Personalising e-Learning Systems: Lessons learned from a vocational education case study

Lie Ming Tang and Kalina Yacef

Computer Human Adapted Interation Reserach Group
School of Information Technology
University Of Sydney
http://chai.it.usyd.edu.au
ltan8012@uni.sydney.edu.au

Abstract. Vocational education refers to the training of specific skills or trades. It is often done part time or in personal time over a lengthy period (months to years). As such, it requires persistence, self motivation and self regulatory skills including goal setting, planning and time management. A growing body of evidence suggests that these self-regulating skills are a key determinant in learning performance and can be improved with support. We report in this paper our experience with a leading vocational education provider in Australia who is transitioning from classroom-based training to a pilot e-learning system. We present the key lessons learned and the prototype interface we designed to improve user self-regulation in planning and time management.

Keywords: Vocational Education, Personalization, Self-Regulated Learning, e-learning

1 Introduction

Vocational education, which refers to the training of specific skills or trades, differs from academic education as the focus lies in skills and knowledge for specific industries or job roles and often requires proof of practical competency to complete. However, similarly to academic learning, self-regulated learning [11] and the ability to direct one's own learning is seen as an essential part of success in vocational education [5, 9]. Important self-regulation skills include setting appropriate proximal goals [8], resource management (e.g., allocation of resource, time management, managing ones learning environment), self-monitoring and adjusting behaviour based on performance. There is a growing body of evidence that suggests personalised and computer aided support for self-regulated learning improves student engagement and performance in e-learning systems [2]. However, such studies has generally focused on academic learning. Self-regulation skills are very relevant in vocational education and help in making students aware of their own role in education and developing their learning are key components [5].

It is important to understand the key challenges for e-learning systems aimed at vocational education and how their personalisation and scaffolding features can support self-regulation. We performed a study with an Australian leading vocational education provider who transitioning from a classroom focused training model to an online and self-managed training model using a pilot e-learning system. We analysed student usage patterns over a 6 month period to understand how this e-learning system is used and interviewed trainers and trainees to get qualitative feedback on how students managed their learning in the system.

In this paper, we present the key lessons learned from our study and a prototype with user interface designed to demonstrate the key features for future e-learning systems to personalise and support for self-regulation of learning.

2 Vocational Education & Training

Vocational education and training represents a critical sector of education where skills for a particular industry, trade and career are created. Over 11 percent of the Australian population between 15 and 64 undertake vocational education and training and the age spread is wide [7]. Moreover, the percentage of adults with professional education as the highest qualification is also very high (between 5 and 34 percent [1]). Learning topics are broad and include industry accreditation and certification for fields in health care, accounting, engineering, law and information technology and many others. Vocational education is usually competence based: where training and assessments are assessed on whether a participant is skilled and competent in a particular job or trade rather than measuring theoretical knowledge alone. This may involve collecting evidence and reports over many weeks or months as part of the assessment. For example, flight training requires that trainee pilots log the number of hours of actual and simulated flying. Similarly to academic learning, vocational education can span a lengthy period of time from months to years. Both require students to maintain self-motivation and persistence. However, over 88 percent of vocational learners are part time in Australia. This suggests a higher competing presence of other priorities for most vocational learners compared to academic students, with the unavoidable challenges in attention focusing, organisation and time management skills.

3 Case Study: e-learning system

We worked with a leading vocational education provider in Australia who has operated a nationally accredited certification program for their employees for over 10 years. Recently, they started transitioning from a classroom training model to a pilot e-learning training program requiring self directed learning and assessments. A key motivating factor is motivate and encourage students to regulate their own learning and to reduce the contacts needed with trainers.

The pilot e-learning system is used for accessing online learning materials, perform assessments and uploading evidence for practical experience and competence. In the pilot program, participants first read or view online learning material, gain practical experience and reinforce their learning and then physically attend a classroom learning workshop. They then complete online assessments via the e-learning system. While the online learning material is not mandatory, they contain the knowledge needed to pass the online assessments. The workshop offers an opportunity for trainers to reinforce the online learning material, provide discussion and simulated practical experience to prepare learners for their formal online assessment.

We performed a 6 month study to analyse student usage patterns based on data from the e-learning system for over 600 trainees. We also interviewed 3 participants and 3 trainers to gain an qualitative view of the challenges with self-regulation and performance.

3.1 Planning & Time Management

Planning and time management is a significant challenge for many students. A commonly cited problem by students, including the trainees in our study, is the lack of time. However, interviews with trainers indicate that this is not the case because trainees are allocated time or are getting paid for their time spent on learning. Rather, the key issues cited by trainers are attention focusing, planning, and time management rather than time constraints. In many cases, work priorities conflict with the planned learning times and students do not adjust their planning or they forget.

While the majority of trainees complete their learning on their own, trainers needed to organise separate workshops specifically for certain groups of trainees to concentrate and complete their online learning away from their workplace which can be busy and not conducive for learning. Trainers found that the trainees' lack of self-regulation skills in time management, planning, prioritisation, and remembering to perform tasks were key challenges. The difficulty level of the learning material was rarely an issue in this program.

3.2 Environment management

According to the social cognitive theory, the social and work environment are a key determinant of behaviour [3]. As part of the program, most trainees are paired with a coach who helps and supports their learning progress. Feedback from both trainers and trainees were very positive in terms of the support provided. Trainer feedback suggests that when coaching support is not very strong, the trainee is less motivated and requires more trainer engagement. Trainees also highlighted that pairing with a study partner provided mutual support and improved their motivation and persistence.

Feedback from both trainer and trainee suggests that those who managed their social and learning environment well had little trouble completing the course. For instance, one successful trainee, who managed his environment by performing his study during his day off, had a study partner in the program. He also ensured that he completed the planned task on schedule through either performing them on time or adjusting his plan. This trainee was able to complete the learning tasks well ahead of schedule.

4 Our Approach

Our hypothesis is that trainees who are struggling in the vocational program can be supported through scaffolding or computer aided support to improve their planning and time management. We built a prototype system, augmenting the existing e-learning platform, with user interface elements designed to promote self regulation. The prototype system lets trainees set time schedules for their learning objectives, monitor their performance and adjust them when necessary. It is also available via mobile application and sends them reminders of upcoming tasks. The prototype system also allows trainers to monitor the performance of their trainees and identify those who need personalised attention.

4.1 Planning & Time Management

To address the feedback above, we made planning and time management a core skill for the prototype system to support and scaffold. There have been very few user interfaces designed to scaffold time management and planning. A previous approaches used Zimmerman's cyclic model of self-regulated learning as the basis to detect and model the learner states [10]. This approach used a calendar-like interface where users define their learning schedules with recommendations and help support. However, we found that the trainees of our vocational program can benefit from an initial engagement to setup a simple schedule and keeping track of their learning task and maintaining their plans. We targeted the more general skill of scheduling of tasks and following through with that plan through monitoring and reminders. In our interface, when users first login to the application, they are first presented with a wizard where they are prompted to set a plan or schedule for when they expect to complete a task, as shown in figure 1. The user can also add this task to their Google calendar and activate an email or SMS notification when their planned task is due. The wizard does this for the first task only and users can access the wizard later if needed. After the wizard exits, they can set schedule for tasks via the plan button (see figure 2). The objective of this wizard is to scaffold users taking control and managing their plan and time.

4.2 Self Monitoring

As part of the interface, the user is allowed to monitor his/her progress in the program. They are also reminded about their upcoming planned tasks, what they have completed and what they still have pending, see figure 3. This allows them to monitor their progress, reflect on their planning and scheduling. In addition, users can monitor their progress compared to their peers for each of the learning objectives in their program, see figure 4. Studies have shown that behaviour can be modified through comparing one's own performance against peers [6], [4].

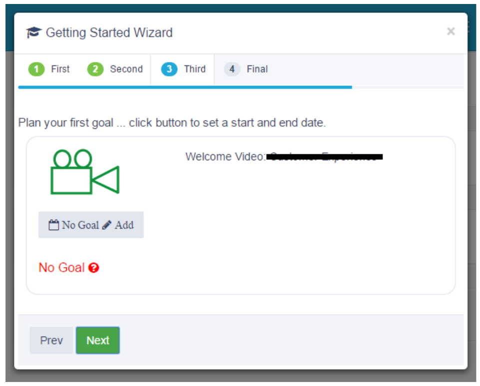

Fig. 1. Wizard to help the user get started with planning their learning schedule. Note: a schedule is referred to as a "goal" in the interface

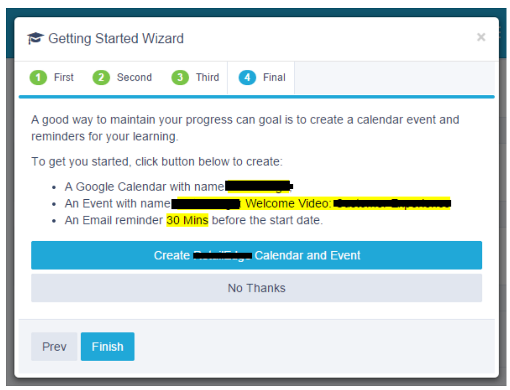

Fig. 2. Add learning schedule to Google calendar and set reminder.

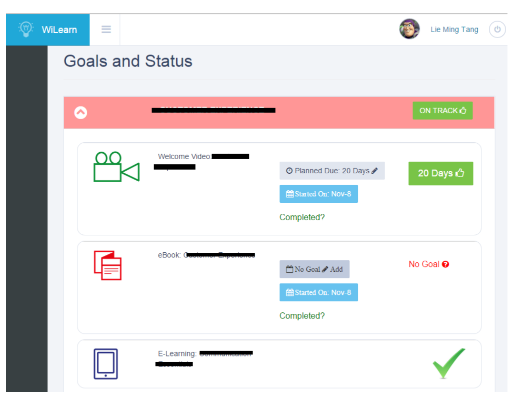

Fig. 3. Learning progress of each learning objective: 1) planned and started but not completed, 2) started but no plan, 3) completed

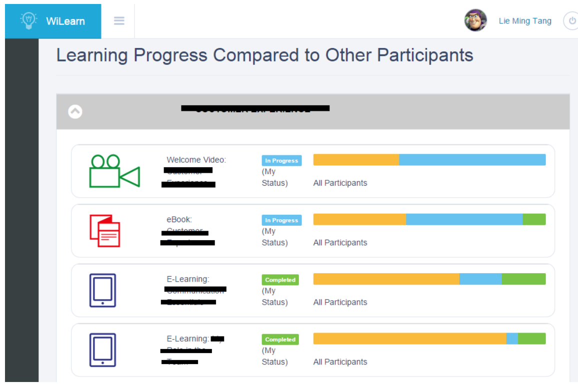

Fig. 4. Monitor learning progress versus other students.

4.3 Monitoring Trainees

We have also provided trainers with the ability to monitor the progress of each of their participants and the status of each e-learning modules i.e., not started, started or completed. This allow trainers to see which students are lagging behind their peers and require personalised attention. See figure 5. Trainers can also see which students have not accessed their e-learning materials so they can send them reminders to maximise the workshop outcomes.

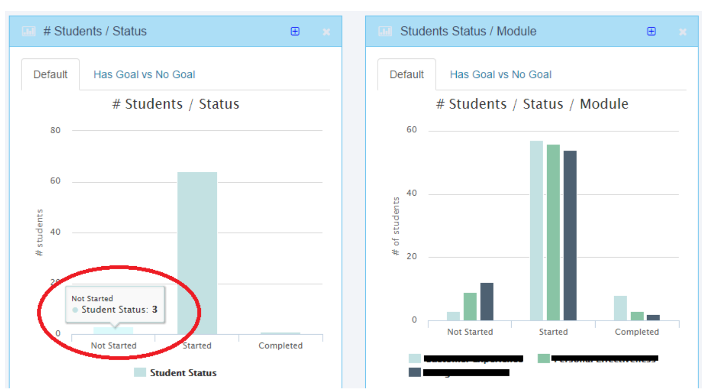

Fig. 5. Trainers see student progress and activity. Further details can be obtained by clicking on the interested status bar drill down (highlighted by red circle).

5 Conclusion & Future Work

We found a key challenge for vocational education students is time and environment management. We designed a prototype system to support students in becoming better planners and time managers. We believe such goal setting and time management interface designs can also be integrated into other e-learning systems where the learning profile is similar to vocational education (e.g., self redirected professional learning, part time academic studies). Peer and trainer engagement and support appears to be important to trainees and trainers and future systems should investigate how scaffolding can be applied. This can potentially reduce withdrawals, increase engagement and motivation for the trainees.

6 Acknowledgement

This work was funded by Smart Services Cooperative Research Centre.

References

1. OECD 2014. *Skills beyond School: Synthesis Report, OECD Reviews of Vocational Education and Trainning.* 2014.
2. Roger Azevedo and Allyson F. Hadwin. Scaffolding Self-regulated Learning and Metacognition Implications for the Design of Computer-based Scaffolds. *Instructional Science*, 33(5-6):367–379, November 2005.
3. Albert Bandura. Human agency in social cognitive theory. *The American psychologist*, 44:1175–1184, 1989.
4. Albert Bandura. Social cognitive theory of self-regulation, 1991.
5. Elly De Bruijn and Yvonne Leeman. Authentic and self-directed learning in vocational education: Challenges to vocational educators. *Teaching and Teacher Education*, 27(4):694–702, 2011.
6. Sten Govaerts, Katrien Verbert, Erik Duval, and Abelardo Pardo. The student activity meter for awareness and self-reflection. In *Proceedings of the 2012 ACM annual conference extended abstracts on Human Factors in Computing Systems Extended Abstracts - CHI EA '12*, page 869, New York, New York, USA, May 2012. ACM Press.
7. Kathrin Hoeckel, Simon Field, Troy R Justesen, and Moonhee Kim. Learning for Jobs. OECD Reviews of Vocational Education and Training. *Education And Training*, (November):220, 2009.
8. Gary P Latham and Edwin a Locke. Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, 50(2):212–247, December 1991.
9. Paul R. Pintrich and Elisabeth V. de Groot. Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1):33–40, 1990.
10. KP Shih and HC Chen. The Development and Implementation of Scaffolding-Based Self-Regulated Learning System for e/m-Learning. *Educational Technology & Society*, 13:80–93, 2010.
11. Barry Zimmerman. Becoming Learner: Self-Regulated Overview. *Theory into Practice*, 41(2), 2002.

Modeling Learner information within an Integrated Model on standard-based representations

Mario Chacón-Rivas[1,*], Olga C. Santos[2], Jesus G. Boticario[2]

[1] TEC Digital, Instituto Tecnológico de Costa Rica, Cartago, Costa Rica
machacon@itcr.ac.cr
[2] aDeNu Research Group, Artificial Intelligence Departament, Computer Science School, UNED C/ Juan del Rosal, 16. Madrid 28040. Spain
{ocsantos,jgb}@dia.uned.es

Abstract. Learner modelling is a process consisting of collecting information explicitly from users and inferring some data from the learner activity. This information is basic for recommending resources as well as to predict performance. There are open issues when it comes to integrate in standards-based user models that information, which covers learning styles, competences, affective states, interaction needs, context information and other learner´s characteristics. In particular, there are standards that can be used to cover several of the subjects to be integrated into those models, such as IMS-LIP, IMS-RDCEO, IMS-AFA. This paper presents a work on implementing a user model that aims at providing a holistic UM perspective, which is able to hold and collects all relevant information, thus supporting its real-life usage. This is expected to facilitate interoperability and sustainability while we are progressing on filling the gaps, where representation and management is required.

Keywords: User modelling, IMS standards, Interoperability of user models, Lifelong Learning User Modelling

1 Introduction

User Models (UM) have been considered as a representation of information on individual users, which is essential for building applications of adaptive systems, intelligent interfaces, intelligent information retrieval and expert systems, among others [1]. Also UM are being used for over the last two decades on implementing personal learning environments, adaptive learning environments and intelligent tutoring systems [2]. Information about UM is usually categorized in terms of personal, affective and cognitive information [2]–[4].

Nowadays there is an increasing interest in taking advantage of new interaction data which cater from learner affection thus requiring integrating into UM affective state indicators [4] [5] [6]. These indicators provide valuable pedagogical pointers, which affect the cognitive process. Actually, learners´ affective modelling is impacting positively on adaptive systems, recommender applications and personalized learning environments [6].

In order to cope with both existing UM information and providing a real life standards-based application this paper introduces existing challenges in terms of the information to be integrated into the model (ie., competences, learning styles, socio-economical data, among others) and the available standards to cope with (e.g., IMS-RDCEO, IMS-LIP, IMS-AFA). The rest of the paper consists of section 2, where UM components and the identification of variability levels are presented, section 3, which summarizes the IMS family international specifications to be integrated into the UM, and last but not least, section 4 where some lines of work in progress are introduced.

2 Identifying UM Components and Information Levels

UM components have been specified in terms of categories or data to be captured to model the learner. Those components could be specified explicitly asking information to the learner or could be specified inferring from the learner interaction with the e-learning platform.

Independently the way to capture the information of the learner, as commented by Brusilovsky and Millán in [2], the interest of information to be modelled in learning environments must allow to identify the user as an individual, thus supporting a feedback process which can be managed by providing recommendations oriented to the meet learners´ needs.

In the context of this research, we identify the UM components and classify them in terms of variability. The variability term is oriented to classify the information depending on the frequency of change, because it will influence any process of recommendation. The UM attributes identified are generated by a methodological process that integrates several sources of information and stakeholders. During the methodology application, the identification of stakeholders designs preliminary security roles.

2.1 UM Components

Several authors defined the UM components oriented to personal information, knowledge and interest. In [5] Bull and Kay enumerate as cognitive, affective and social attributes. Baldiris et. al. [7] present the user model in terms of learning styles, competences and access device preferences, also it includes knowledge level based on six levels of knowledge defined by Bloom´s taxonomy. This proposal also includes a collaboration level based on indicators obtained from learners´interaction in the learning management system.

Based on aforementioned and related literature we are currently defining the UM components information in terms on the following information categories: *personal*, *provenance*, *academic record*, *socio-economical*, *accessibility* or special needs, *psychological*, *learning styles*, *competences, knowledge level and collaborative level*.

Those components can be modelled and organized in terms of standards, such as the IMS family specifications. The use of these international specifications is aimed to support collaboration and systems interoperability and have the advantage of being specifications that are already integrated into dotLRN [7] [8] [9]. IMS-LIP is a collection of information about the learner, which supports data exchange between applications, agents, server and other services concerned about the learners` characteristics [10] IMS-RDCEO is a concise and flexible structure to represent competencies, furthermore this specification is extensible to any competence model [7] [11].

To provide the required standards-based modelling support at TEC, we are following these previous approaches while extending them and filling information gaps when needed. Based on IMS-LIP categories, the element *identification,* is loaded with *personal, provenance, academic, socio-economical* information. The element *Competency* is loaded with competences information. The TEC competences model is based on CEAB model [13] and this competencies information is represented using IMS-RDCEO. *Accessibility* is represented using IMS-AFA. Additional details on IMS family specifications are provided in the section 3.

The UM in our context, called td-um means TEC Digital-User Model. **td-um** is an integrated model because it is gathering together learner information from applications, databases and some indicators collected from learner interaction with the e-learning platform.

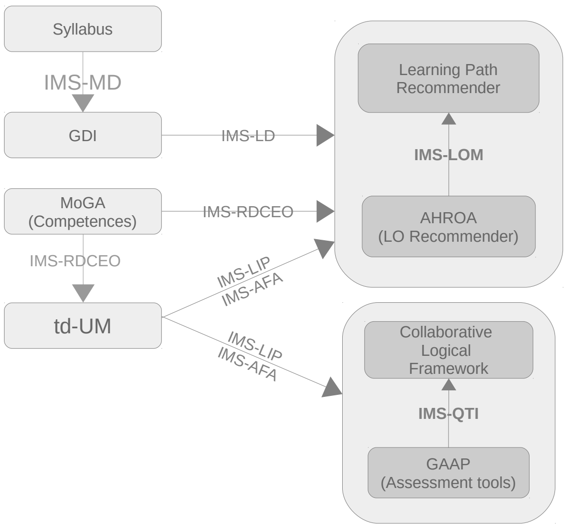

Figure 1: td-UM Integrated model

From that integrated approach and after studying available information from literature, we have detected some gaps in the information to be modelled, these are the following:

- Learner knowledge level: it is been modelled based on specific background of knowledge from each discipline studied by the learner. In case of computing students, the knowledge level is modelled using knowledge areas presented by ACM in [14]. The gap to be resolved is based on several bodies of knowledge from different disciplines. The other disciplines to be modelled in TEC are: Industrial engineering, Electronic engineering, Materials engineering, Electromecanic engineering, Construction engineering, Agricultural engineering, Industrial Maintenance engineering, Occupational Safety and Environmental Hygiene Engineering. This other disciplines have a different body of knowledge, therefore the structures used to model should be sufficiently flexible.

- Academic record attributes: these cover information reflecting the progress in program courses in terms of final grades or qualifications. It is frequently confused with the knowledge level. This is mainly required to preserve historical information.
- Competences attribute: it contains a set of competences that requires reflecting the level of domain of each competence, and evidences used to assess each competence, among others. The work in progress is designing the model which we are adapting to cope with TEC competence model.
- Variability of information: an important issue is to track the progress in competence domain, as well in academic records. This progress tracking represent some level of variability of information that could impact the recommendation and adaptivity of platform.
- Categories and attributes privacy levels: the privacy level of some attributes or for the whole category are not clear in the specifications. For example, in socio-academic attribute, the sub-attributes: *level of sociability*, *esteem*, *motivation*, *coping strategies* contains private information accessible only for department of Psychology and the learner. In this work is needed to define and to model the privacy level by category and attribute based on privacy modelling [15]. The model is contemplating the user roles definition and the integration with dotLRN.

Currently at TEC Digital we have implemented several applications that are using partial learner models, Figure 1 shows the integration architecture. The immediate work is being focused on adapting td-UM, to be used as source of integrated learners´ information, which will be able to support recommendations and assessments. Those applications are:

- Adaptive Learning Paths, which use the learning design of a course and the students´ performance information to recommend learning resources. The recommendation is based on bayesian networks [16].
- Hybrid Agent Recommender of Learning Objects (AHROA), based on the learning design and syllabus of the course, recommends learning objects to learners. The recommendation is prepared using TF-IDF[1] to work the terms relevancy, also uses cosine similarity. The UM will provide information about learner needs to AHROA in order to identify the impact on the quality of recommendations.
- Collaborative Logical Framework, (CLF) implemented by aDeNu [17], it uses collaborative indicators to assess the learners collaboration. An important activity on the CLF is the identification of each group moderator during

[1] More details in http://nlp.stanford.edu/IR-book/html/htmledition/tf-idf-weighting-1.html

the consensus stage, this activity is currently based on the learner interaction in the platform. The UM will impact the CLF integrating specific attributes concerning to the leadership and entrepreneurship. The TEC Digital adapted CLF to improve the indicators information [18] and to analyse the impact of the UM in CLF assessment.

- Learning activities editor application (GAAP) implements learning styles test based on Felder&Soloman theory [19].
- Several test to determine personality and character, leadership, entrepreneurship, communications competences. These tests are defined by the Psychology department and by the team responsible to design the competence model. Some tests are in processes to be patented by TEC. The variability and tracking progress of these competences are very important to be considered in the CLF, GAAP and AHROA.

2.2 UM Variability Information Levels

In order to take advantage of the information being modelled in real-life situations we are particularly interested in taking into consideration the "variability" factor. The levels of information *variability* reflect the frequency of variability or changes on the values of UM components. Authors as Sosnovsky and Dicheva in [4], defined this variability as *long term* and *short term* variability. In our research we are defining these variability levels as *low*, *medium* and *high* as explained bellow:

Low Variability: Some UM components hardly ever vary during the learning period and are used either for managing personal information (e.g., *name*, *birthdate*, *provenance*, *native language*) or academic processes, such as birth date, which can be used to calculate the learner age, the native language and other language domain features that may have an impact on the learning process.

Medium Variability: UM components seldom vary on a daily basis but the change more frequently than those being described as *low*, including periods of relatively stable values; usually are characteristics that are modified during the learner progress in the curricula. It could be presented with the competences category. Improvements on competences are been registered and assessment once at year. Some examples of competencies are *communication skills*, *team work, problem analysis, knowledge base of engineering, ethics and equality,* among others. The impact on medium-level variability changes in the learning process is very important for learners because the progress on some of those components affected by them has a direct and immediate influence in the learning performance.

High Variability: Those components vary almost continuously and hardly ever remain stable, some even could vary daily.

3 Representation based in standards

The use of international specifications, such as the IMS family, is aimed to support integration and interoperability. The LMS used in TEC is based on dotLRN, which supports IMS-LD and IMS-QTI. The model td-um is based on IMS specifications adapting to the specific needs mentioned above as gaps in the models.

Table 1 shows the categories and standards we are using along with the main attributes of the UM.

Table 1: Standards and information category in td-UM

Standard	Category	Attributes[2]
IMS-LIP	personal	name, birthdate, address, phone number, email address, native language, affective, socio-academic needs{ study conditions, study habits, metacognitive study strategies, development study strategies, organizational study strategies, level of sociability, esteem, motivation, coping strategies}
	provenance	place of provenance (based to recognize the social development index)
	socio-economical	scholarship, loan financing
	academic record	based on progress record on each course
	learning style	based on Felder and Solomon learning style test
	knowledge level	based on the body of knowledge of specific disciplines
	collaborative level	based on indicators tracked from the interaction with the platform.
IMS-AFA	accessibility	visual adaptation, hearing adaptation, cognitive adaptation, learning needs {reading-writing, understanding, speaking, math, attention, depression, anxiety, difficulties with peers, family problems, difficulties with teachers}
IMS-RDCEO	competences	knowledge bases of engineering, problem analysis, investigation, design, use of engineering tools, communication skills, professionalism, impact of engineering on environment and society, ethics and equity, economics and project management, lifelong learning, resource utilization

IMS-LIP categories: *Academic record* is being adapted to support historical information. *Knowledge level* is being adapted to cover the body of knowledge of several disciplines. Also the *knowledge level* should model the knowledge area or topic with a level reflecting the expertise in the given domain. This level is going to be described in terms of the Bloom´s Taxonomy, following previous approaches [7].
Personal category is been adapted to model *socio-academic needs*{ study conditions, study habits, metacognitive study strategies, development study strategies, organiza-

[2] This column enumerates only principal attributes.

tional study strategies, level of sociability, esteem, motivation, coping strategies} all captured using a test of 40 questions.

IMS-AFA category of *Accessibility* is being adapted to model *learning needs* {reading-writing, understanding, speaking, math, attention, depression, anxiety, difficulties with peers, family problems, difficulties with teachers}. These learning needs are captured using a test of 66 questions.

IMS-RDCEO category of competences {*Knowledge base of engineering, Problem analysis, Investigation, Design, Use of engineering tools, Individual and team work, Communication skills, Professionalism, Impact of engineering on environment and society, Ethics and equity, Economics and project management, Lifelong learning, Resources utilization*}. Adaptation to be provided require to support the representation of the domain level of each competence, the evidences used to assess each competence and the authority. Another modelling issue is to support the integration and matching of the competence model with those required for an international accreditation process. Accreditation processes are oriented to model the program of courses or careers in universities, while learner models reflects personal and individual information. The competence model for international accreditation is based on statistical samples [20].

4 Works in progress towards an integrated learner model

Once the aforementioned issues are designed, structured and integrated to cope with the information to be represented, the next decision will be the specific way to represent information on each competences, learning styles and affective indicators. The interoperability of those indicators will require an ontological representation that allows to deal with the information to be used in each foreseeable situation.

As an example of the decision to be done, we are planning to apply a test of temperament to identify (1) extroverted – introverted *temperaments*, (2) ways to *capture information*: by intuition-by senses, (3) *ways of making decisions*: by thought- by feeling, (4)*ways to organize time*: judicious-mandatory. This test has 70 questions, the interpretation of the results will give a value for each element to identify. A learner could have a value for the way to capture information of 6-4 (ie, 6: by thought, 4: by feeling). The internal decision about the representation of those values could affect the adaptive process, if the UM stores the 6 value or the pair 6-4. A recommender system could take several considerations concerning the type of resources to recommend. The tests used to capture information are being used by TEC since 2002 and they are bases in [21].

Finally, this research is aimed to fill the gaps beyond current usage of UM in adaptive learning systems thus making it really extensible, sustainable and applicable in any situation.

We are currently progressing on the first stage of this research, which covers: (1) understanding the dimensions of UM and the results of experimental research in learning scenarios, (2) focusing on reviewing the state of the art in UM and its components, (3) identifying a methodological approach to gather the UM attributes in a real learning environment, (4) identifying possible gaps that may come up when integrating UM into real dimensions of learners characteristics that have impact on the learning process.

This contributions of our research are focused in (1) defining a methodology to identify attributes of UM in real learning environment that support personalised and inclusive e-learning scenarios, (2) identify the UM attributes that really impact in recommendation processes using AHROA and CLF, (3) validating if the standards are enough to cope modelling real learning environments supporting relevant recommendations.

The work in progress is done in the context of a PhD thesis research with aDeNu group. This is implemented in the Instituto Tecnológico de Costa Rica (TEC). The implementation is based on dotLRN platform, instantiated by TEC Digital [22]. This research is aimed to provide a model containing learner information to be used in adaptive and recommendation processes, based on interaction indicators computed from large scale setting which corresponds to official courses run in TEC.

5 Acknowledgements

Authors would like to thank the Spanish Ministry of Economy and Competence (MINECO) for funding BIG-AFF project (TIN2014-59641-C2-2-P), where this research is partially supported. Authors would also like to thank the Department of "Orientación y Psicología" (DOP) at TEC, specially to Alejandra Alfaro.

References

1. A. Kobsa, "User modeling: Recent work, prospects and hazards," *Hum. Factors Inf. Technol.*, vol. 10, pp. 111–111, 1993.
2. P. Brusilovsky and E. Millán, "User Models for Adaptive Hypermedia and Adaptive Educational Systems," in *The Adaptive Web*, Berlin, Heidelberg: Springer-Verlag, 2007, pp. 3–53.
3. S. Bull and J. Kay, "Metacognition and open learner models," in *The 3rd Workshop on Meta-Cognition and Self-Regulated Learning in Educational Technologies, at ITS2008*, 2008, pp. 7–20.
4. S. Sosnovsky and D. Dicheva, "Ontological technologies for user modelling," *Int. J. Metadata Semant. Ontol.*, vol. 5, no. 1, pp. 32–71, 2010.
5. S. Bull and J. Kay, "Open learner models," in *Advances in Intelligent Tutoring Systems*, Springer, 2010, pp. 301–322.
6. C. Conati and H. Maclaren, "Empirically building and evaluating a probabilistic model of user affect," *User Model. User-Adapt. Interact.*, vol. 19, no. 3, pp. 267–303, 2009.
7. S. Baldiris, O. C. Santos, C. Barrera, J. Boticario, J. Velez, and R. Fabregat, "Integration of Educational Specifications and Standards to Support Adaptive

Learning Scenarios in ADAPTAPlan.," *Int. J. Comput. Appl.*, vol. 5, no. 1, pp. 88–107, 2008.
8. O. C. Santos and J. G. Boticario, "Requirements for Semantic Educational Recommender Systems in Formal E-Learning Scenarios," *Algorithms*, vol. 4, no. 2, p. 154, 2011.
9. J. Boticario, A. Rodriguez-Ascaso, O. C. Santos, E. Raffenne, L. Montandon, D. Roldán, and F. Buendía, "Accessible Lifelong Learning at Higher Education: Outcomes and Lessons Learned at two Different Pilot Sites in the EU4ALL Project.," *J UCS*, vol. 18, no. 1, pp. 62–85, 2012.
10. "IMS GLC: Learner Information Package Specification." [Online]. Available: http://www.imsglobal.org/profiles/index.html. [Accessed: 14-Apr-2015].
11. "IMS GLC: RDCEO Specification." [Online]. Available: http://www.imsglobal.org/competencies/. [Accessed: 17-Apr-2015].
13. "Accreditation Resources | Engineers Canada." [Online]. Available: https://www.engineerscanada.ca/accreditation-resources. [Accessed: 07-Apr-2015].
14. "Computing Curricula 2005: The Overview Report." [Online]. Available: http://www.acm.org/education/education/curric_vols/CC2005-March06Final.pdf. [Accessed: 28-Jul-2014].
15. Y. Wang and A. Kobsa, "A PLA-based privacy-enhancing user modeling framework and its evaluation," *User Model. User-Adapt. Interact.*, vol. 23, no. 1, pp. 41–82, Mar. 2013.
16. I. Gámez, C. Garita, and M. Chacón-Rivas, "Generación de Sugerencias de Rutas de Aprendizaje Adaptativas en Entornos de e-learning," presented at the Conferencia Latinoamericana en Informática - CLEI 2012, Medellín, Colombia, 2012, pp. 1–10.
17. O. C. Santos and J. G. Boticario, "Involving Users to Improve the Collaborative Logical Framework," *Sci. World J.*, vol. 2014, pp. 1–15, 2014.
18. M. Chacón-Rivas, O. C. Santos, and J. G. Boticario, "Collaborative Logical Framework adapted to instructors and learners," in *Artificial Intelligence in Education-Interactive Events*, Madrid, Spain, 2015.
19. R. M. Felder and B. A. Soloman, "Learning styles and strategies," *N. C. State Univ. Httpwww Ncsu Edufelder-PublicILSdirstyles Htm*, 2000.
20. E. Raffenne, "MIRLO: una ontología para dar soporte a un modelo de estudiante abierto," UNED, Madrid, Spain, 2013.
21. A. Alfaro, "Demandas académicas y afrontamiento en estudiantes con adecuaciones curriculares | Alfaro Barquero | Actualidades en Psicología." [Online]. Available: http://www.revistas.ucr.ac.cr/index.php/actualidades/article/view/36. [Accessed: 04-May-2015].
22. M. Chacon-Rivas and C. Garita, "A Successful OSS Adaptation and Integration in an e-Learning Platform: TEC Digital," in *Open Source Software: Mobile Open Source Technologies*, Springer, 2014, pp. 143–146.

Patterns of Confusion: Using Mouse Logs to Predict User's Emotional State

Avar Pentel

Tallinn University, Institute of Informatics, Tallinn, Estonia

pentel@tlu.ee

Abstract. This paper describes an unobtrusive method for user confusion detection by monitoring mouse movements. A special computer game was designed to collect mouse logs. Users' self-reports and statistical measures were used in order to identify the states of confusion. Mouse movement's rate, full path length to shortest path length ratio, changes in directions and speed were used as features in the training dataset. Support Vector Machines, Logistic Regression, C4.5 and Random Forest were used to build classification models. Models generated by Support Vector Machine yield to best classification results with f-score 0.946.

Keywords: confusion detection, behavioral biometrics, mouse dynamics.

1 Introduction

The ability to recognize, interpret and express emotions plays a key role in human communication and increasingly in HCI. In the context of learning systems, the ability to detect user emotional states, gives promising applications to adaptive recommendations, adaptive interfaces, etc. Usually special equipment is used for emotion detection: electroencephalogram, skin conductance, blood volume pressure [1,2] or gaze and facial data [3,4]. But when it goes to real life application, we can relay no more, than unobtrusive standard computer inputs like mouse or keyboard.

The theory of "embodied cognition" [5] gives a theoretical framework studying mouse movements in order to predict mental states. Barsalou suggests that this bi-directional relationship between mental states and bodily states emerges because the core of social and cognitive information processing lies in the simulation of original information [6]. There are some studies [7,8,9,10] about mouse movement and emotions, which all suggest a link between mouse movement and emotions. Yet, most of these studies are conducted with relatively small samples. Secondly, all these studies are dependent on the specific context of an experiment, and general link between emotions and mouse movements is not investigated.

In the current study, we aim to find a link between confusion and mouse movements and try to avoid both of previously mentioned shortcomings by using larger sample, and avoiding specific context in our experiment.

2 Methodology

2.1 Data Collection Procedure and Sample

A simple computer game was built to collect user mouse data. The idea of the game come from Christmas Calendar chocolate boxes, where the chocolates are hidden behind numbered doors. There are usually numbers from 1 to 24, and in order to make the right door harder to find, numbers are randomly arranged and they look differently. Similarly, we designed a game, which fills screen with randomly arranged buttons labeled with numbers 1 to 24. All buttons are different size and color (Fig.1). User task is to click on all buttons in the right order as fast as possible. To keep up the motivation, the game was installed in school computer class as a part of login system, i.e. in order to log in users were forced to play this game. There was also an option to play it many times. It was also publicly announced, that best performers would be awarded. For every game session mouse activity (movements and clicks) was logged.

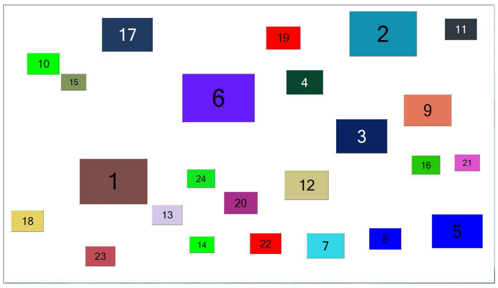

Fig. 1. A Christmas Calendar game built for data collection. The user has to click as fast as possible on all buttons in the right order

Our logging procedure was an event based, which means that mouse position was not recorded in fixed interval, but only if difference in position of mouse occurred. In our case this difference of position was set to 30 pixels. Our mouse logs consisted triples of x and y coordinates and timestamp. We recorded data from 516 game sessions played by 262 individual users. As each game session consisted of 24 searching tasks (to find next number), we had all together 12384 comparable records, each of them presenting mouse movement logs between two button clicks.

2.2 Labeling Data with Emotional state

We also interviewed selected participants (N = 44) right after the game. Reviewing together the whole game session again, we asked to describe the emotions during the game. Initially we asked users to position his/her emotions on Russell's circular model [11], but pre testing revealed, that in the current set of the experiment, users were only able to describe two categories of emotions - the state of confusion and the state of content. Therefore we continued to collect self-report data on a 7 point Likert scale where 1 = content, and 7 = confused. While users were not able to specify the exact time when the state of confusion began or end, we divided the game session to 24 separate searching tasks, and linked those emotion feedback data to a whole task. All together we got 44 x 24 = 1056 tasks labeled with emotion data.

It is intuitively clear, that in such circumstances, confusion and target finding speed are related. While target finding speed differs individually, all these finding scores were standardized session-wise, and then Pearson correlation with confusion self-report data were found. As expected, there was significant correlation between confusion and standardized finding time ($r = 0.86$). Also, all tasks associated with confusion, had standardized finding speed half standard deviation below the mean, and those associated with a feeling of content, half standard deviations above the mean. While our interviews covered only less than 10% of all game sessions, we extended this relation to all other game sessions too.

We suppose, that very quick results may not include confusion at all, i.e. user is aware from the beginning about the location of the target. But in order to minimize possible confusion, which may be present in the beginning of each task, we divided finding time to the half and used only the last half of log data as characterizing non-confusion. Similarly, it is obvious, that in tasks that were characterized as confusing, the state of confusion does not cover the whole time between two button clicks. Obviously, confusion must end in some moment, when user notices the next button. It is reasonable to suppose that confusion ends somewhere in the second half of the searching process. Therefore, we split each of these slower result logs to the half and used only the first half of searching task as characterization of confusion (Fig. 2.).

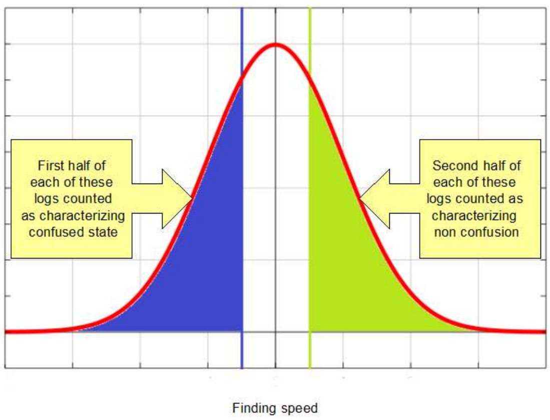

Fig. 2. Separation of mouse logs representing state of confusion and non-confusion.

Out of these two subsets we excluded repeated sessions by the same users, and extreme results. From the remaining data we created balanced training dataset of 2282 records.

2.3 Features

In the current study, we extracted 33 features based on distance, speed, direction and direction change angles (Table 1.). Feature selection procedure with Chi Squared attribute evaluation and SVM attribute evaluation revealed, that strongest features were those of speed based and those of based on relations of shortest distance and actual distance. Best models with those attributes yield to F-score 0.96 width SVM and Logistic regression.

Table 1. Features.

Type	Feature	Explanation
Distance*	Precision	Shortest distance between two button clicks and actual mouse path length ratio.
Speed**	Speed	Actual mouse path length between two button clicks divided by task completion time.
	AdjSpeed	Actual mouse path length between two button clicks divided by shortest path, and then divided by task completion time.
Direction	Direc-tionX	Number of mouse movements in particular direction. We divided movement directions to 8 distinctive segments as north, northeast, east, etc. We counted all movements in particular direction segment, and divided to all movements.
Direction changes	TurnA Turn10, Turn20,... Turn 180	Mouse movements' path was recorded as consecutive straight lines of 30px length. We measured each angle between two consecutive movements and extracted 18 features representing turns from 0 to 180 degrees by 10-degree step. Counted results were normalized by whole number of movements.
	TurnA+	All turns greater than angle A (A counted by 45-degree step).

* Excluded in training feature set of the models titled as "target unknown" in Table 2.

** Excluded in all training feature sets.

For our final model we had to exclude features that were related to speed, because speed was previously used by us for associating tasks with emotional states. Without speed related features, models F-score dropped from 0.96 to 0.946.

As our goal was to identify confusion patterns without knowing the real target, we also excluded the feature that was calculated by using information about shortest distance. All reminded features were based on movement direction and direction changes. Direction based features were number of movements on specific direction divided by mouse path length. Direction changes were measured as the angle between previous and next movement. Within these features strongest features were direction changes closer to 180 degrees, more than 135 degrees and between 160 and 170 degrees.

2.4 Machine Learning Algorithms and Technology

For classification we tested four popular machine-learning algorithms: Logistic regression, Support Vector Machine, Random Forest, and C4.5. Motivation of choosing those algorithms is based on literature [12,13]. The suitability of listed algorithms for given data types and for given binary classification task was also taken in to account. In our task we used Java implementations of listed algorithms that are available in freeware data analysis package Weka [14].

For evaluation, we used 10 fold cross validation. We partitioned our data into 10 even sized and random parts, and then using one part for validation and another 9 as training dataset. We did so 10 times and then averaged validation results.

3 Results

As mentioned before, when excluding all speed-based features, our SVM model with standardized data yield to F-score 0.946. When excluding all distance-based features, results dropped considerably, but all our classifiers still yield to F-scores over 0.8. In following table (Table 2.) are presented results of different classifiers generated with features that are calculated using data about known target (i.e. the shortest path) and without these features.

Table 2. The results of the models trained with different feature sets

Model	**Target known**			**Target unknown**		
	Accuracy	***F-score***	***ROC***	***Accuracy***	***F-score***	***ROC***
SVM (standardized)	94.61%	0.946	0.946	82.38%	0.824	0.825
Logistic Regression	93.49%	0.935	0.978	82.72%	0.827	0.889
Random Forest	92.07%	0.921	0.971	84.47%	0.845	0.825
C4.5	91.96%	0.919	0.937	83.59%	0.835	0.836

4 Discussion and Conclusion

Simple feature set of directions, direction changes and relations between actual and shortest distance proved to be useful in classification confused and non-confused user. As we can see from Table 1, knowing the target makes predictions better, but even without knowing the target, frequent direction changes in mouse movement, are still good predictors of confusion. This might be an indirect confirmation to studies about the correlation between gaze and mouse movements.

However, we have to address the limitations of such set of experiment. Depending on the tasks and page layout, user mouse movements might differ considerably. Our results are applicable in situations, where users have to find something particular on unfamiliar (web) environment, in set of menus, links or graphical elements. But our approach might not work in web page considered for reading. For example, while somebody is used to fallow line with mouse cursor while reading the text, the mouse logs will show frequent changes in directions, which in by our model will predict confusion. Therefore more study is needed in different types of environments.

References

1. G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, “Emotion Assessment : Arousal Evaluation Using EEG’s and Peripheral Physiological Signals,” in Proc. Int. Workshop on Multimedia Content Representation, Classification and Security, pp. 530-537 (2006)
2. E. Leon, G. Clarke, V. Callaghan, and F. Sepulveda, “A user-independent real-time emotion recognition system for software agents in domestic environments,” Engineering Applications of Artificial Intelligence, vol. 20, no. 3, pp. 337–345 (2007)
3. Happy, S.L. et al. Automated Alertness and Emotion Detection for Empathic Feedback During E-Learning, IEEE Digital Library (2013)
4. Jaques, N. et al. Predicting Affect from Gaze Data during Interaction with an Intelligent Tutoring System, Lecture Notes in Computer Science Volume 8474, pp 29-38 (2014)
5. P. M. Niedenthal, "Embodying emotion," Science, vol. 316, pp. 1002-1005, (2007)
6. L. W. Barsalou, "Grounded cognition," Annual Review of Psychology, vol. 59, pp. 617-645, (2008)
7. J. Scheirer, R. Fernandez, J. Klein, and R. W. Picard, "Frustrating the user on purpose: a step toward building an affective computer," Interacting with computers, vol. 14, pp. 93-118, (2002)
8. Zimmermann, P. S. Guttormsen, B. Danuser, and P. Gomez, "Affective computing - a rationale for measuring mood with mouse and keyboard," International journal of occupational safety and ergonomics, vol. 9, pp. 539-551, (2003)
9. Zimmermann, P. "Beyond Usability–Measuring Aspects of User Experience," Doctoral Thesis (2008)
10. Maehr, W. eMotion: Estimation of User's Emotional State by Mouse Motions: VDM Verlag, (2008)
11. Russell, J. A. 1980. A Circumplex Model of Affect. Journal of Personality and Social Psychology. Vol. 39, No. 6. 1161-1178 (1980)
12. Wu, X. et al. Top 10 algorithms in data mining. Knowledge and Information Systems. vol 14, 1–37. Springer (2008)
13. Mihaescu, M. C. Applied Intelligent Data Analysis: Algorithms for Information Retrieval and Educational Data Mining, pp. 64-111. Zip publishing, Columbus, Ohio (2013)
14. Weka. Weka 3: Data Mining Software in Java. Machine Learning Group at the University of Waikato. http://www.cs.waikato.ac.nz/ml/weka/

Using Problem Statement Parameters and Ranking Solution Difficulty to Support Personalization

Rômulo C. Silva[1,2], Alexandre I. Direne[2], and Diego Marczal[3]

[1] Western University of Paraná (UNIOESTE)
[2] Federal University of Paraná
[3] Federal Technological University of Paraná
romulocesarsilva@gmail.com
alexd@inf.ufpr.br
dmarczal@gmail.com

Abstract. The work approaches theoretical and implementation issues of a framework aimed at supporting human knowledge acquisition of mathematical concepts. We argue that personalization support can be achieved from problem statement parameters, defined/set during the creation of Learning Objects (LOs) and integrated with the skill level of learners and problem solution difficulty. The last two are formally defined here as algebraic expressions based on fundamental principles derived from extensive consultations with experts in pedagogy and cognition. Our implemented prototype framework, called ADAPTFARMA, includes a collaborative authoring and learning environment that allows short- and long-term interactions. We present our ongoing research about student modeling to support personalization. Finally, we draw conclusions about the suitability of the claims and briefly direct the reader's attention to future research.

Keywords: rating, problem difficulty calibration, Intelligent Tutoring Systems

1 Introduction

The personalization in computer-based learning systems can range from simple student preferences to motivational state detection. Besides, the system is expected to adapt to specific learning needs, including different assessment mechanisms. In Algebra, the student's expertise is usually developed by solving problems that require a set of assessed skills. This is done in both conventional education schools and by applying advanced learning technologies, such as Intelligent Tutoring Systems (ITS). Normally, human teachers detect students' misconceptions when marking tests and exercises. Depending on how much the answer of a question departs from its correct version, two students that missed the same question could be scored different grades for that specific question.

Another aspect that can be used to compose the score is how difficult the question is. The difficulty degree of an exercise can be measured by the number of students that have skipped or made a mistake in that exercise. Thus, a student who finds the correct answer of a question that many missed, probably has more skills than others and the score should reflect that. Conversely, a student who makes a mistake in a question that many were successful to answer, might possess fewer skills. A student error can basically be used as a guideline for two actions: simply to assess the student or to detect misconceptions towards a more effective pedagogical practice. In the latter sense, recently, there has been increasing interest for direct use of errors as a source of teaching material, in order to learn more deeply about the content of the domain and thus develop metacognitive skills [5].

Another desirable aspect in ITS is in predicting or prospecting whether a learner will be able to answer a question correctly or not before it is actually showed to him or her, allowing a more effective personalization support. Usually this kind of feature requires that questions be previously calibrated according to their difficulty and matched to the assessed student's skills.

2 Literature review

Segedy *et al.* [11] propose a taxonomy for adaptive scaffolding in computer-based learning environments, named Suggest-Assert-Modify (SAM). Suggestion scaffolds provide information to learners for the purpose of prompting them to engage in a specific behaviour. Assertion scaffolds communicate information to learners as being true that will be integrated with their current understanding. Modification scaffolds change aspects of the learning task itself.

A manner to support personalization is by implementing algorithms that generate different content sequencing according to a learner's needs. In this sense, Champaign and Cohen propose an algorithm [1] for content sequencing that selects the appropriate learning object to present to a student, based on previous learning experiences of like-minded users. The granularity of sequencing is on the LO level, not exercises or issues. Segal *et al.* [10] propose an algorithm for personalizing educational content in e-learning systems to students. It combines collaborative filtering algorithms with social choice theory. Schatten and Schmidt-Thieme [9] present the Vygotski Policy Sequencer (VPS), based on the concept of Zone of Proximal Development devised by Vygotski. It combines matrix factorization (a method for predicting user rating) with a sequencing policy in order to select at each time step the content according to the predicted score.

Ravi and Sosnovsky [8] propose a calibration method for solution difficulty in ITS based on applying data mining techniques to a student's interaction log. Using the classical bayesian Knowledge Tracing (KT) method [2], the probability that a student has acquired a skill is calculated on the basis of a tentative sequence of exercises for which the soluctions involve a given concept. The logged events are grouped by exercises and classified according to the student's skills.

3 Automatic calculation of rating

Rating systems are frequently used in games to measure the players skills and to rank them. Usually, the rating is a number in a range $[minRank, maxRank]$

such that it is very unlikely that a player falls on the extremes. Inspired by game rating systems and taking the performance of other learners, this study proposes Equation 1 to assess iteratively a student's ability.

The following guidelines were adopted: *(1)* each question is scored a difficulty degree with a Real value in the range [0..10] and the student is rated a number in the range [1..10] to express his or her expertise level in the subject matter; *(2)* the easier the question, the greater the likelyhood that student will answer it correctly (in this case, a student's rating should have just a small increase if he or she enters the correct answer and should have a large decrease in the case of failure); *(3)* students that are successful in the first attempt to solve a question are scored a higher increment in their expertise level compared to those who need several attempts; *(4)* skipped questions are considered wrong.

Consider Equation 1. The details of its parameters are as follows:

$$R_J^q = R_J^{q-1} + Ak_1\alpha(10 - \frac{9T_J^q}{T_{med}^q}) - Ek_2\beta \times 10\frac{T_J^q}{T_{med}^q} \tag{1}$$

- R_J^q: student J's rating after answering question q. $R_J^0 = 5.5$ (initial rating);
- $A = 1$ and $E = 0$ for successful in answering q, otherwise $A = 0$ and $E = 1$;
- T_J^q: number of unsuccessful attempts of student J to answer question q;
- T_{med}^q: median of wrong attempts on question q during classroom time;
- $\alpha = \frac{1}{N_a^q}$ and $\beta = \frac{1}{N_e^q}$ are weight factors to increase and decrease the rating respectively (N_a^q and N_e^q are the number of students that were successful and unsuccessful answering question q, respectively);
- k_1 and k_2: multiplier factors of rating increase and decrease, respectively, calculated by $k_1 = 1 - \frac{R_J^{q-1}}{10}$ and $k_2 = \frac{R_J^{q-1}-1}{10}$.

Although there is no limit to the number of attempts a student can make to answer a question, for calculation purposes, 10 trials is considered the maximum. Factors k_1 and k_2 avoid results of the expression in Equation 1 to reach upper and lower bounds of the range [1..10].

Using only the number of attempts, the difficulty degree of a question q can be defined by Equation 2 and its parameters are as follows:

$$D^q = \frac{\sum_{J=0}^{J=n} T_J^q}{N_e^q + N_a^q} \tag{2}$$

- D^q: difficulty degree of the question q after an exercise session;
- T_J^q: number of unsuccessful attempts of student J to answer question q. If the number of attempts is greater than 10 trials, then 10 is taken as T_J^q;
- N_e^q and N_a^q are the same as in Equation 1

4 The ADAPTFARMA environment

The ADAPTFARMA (Adaptive Authoring Tool for Remediation of errors with Mobile Learning) prototype software tool is a modified version of FARMA [6], an authoring shell for building mathematical learning objects. In ADAPTFARMA, a learning object (LO) consists of a sequence of exercises following their introductory concepts.The introduction is the theoretical part of a LO where concepts are defined through text, images, sounds and videos. The implementation was carried out aiming at its use on the web, either through personal computers or mobile devices.

For each question, the teacher-author must set a reference solution, which is the correct response to the question. ADAPTFARMA allows arithmetic and algebraic expressions to be entered as the reference solution. Under the learner's functioning mode, the tool deals automatically with the equivalence between the learners response and the reference solution.

An important feature of ADAPTFARMA is the capability of backtracking the teacher to the exact context in which the learner made a mistake. It allows the teacher to view a learner's complete interaction with the tool in the chronological order by means of a graphical timeline. In addition, he/she can perform a closer monitoring of problem solutions from other classroom students, as long as system permission is given through the collaboration mechanisms. Likewise, learners can backtrack to the context of any of their right or wrong answers in order to reflect about their own solution steps and find new solution hypotheses. Additionally, on the collaborative side, it is possible for the teacher to carry out a review of students' responses and then provide them with non-automatic feedback, which can be done by exchanging remote messages through the system.

5 Algorithm for exercises sequencing

The ADAPTFARMA environment was designed such that different pedagogical strategies can be used and tested. In this study, we propose an algorithm for sequencing exercises, named Adaptive Sequencing Method (ASM), to be shown in ascending order of difficulty, combined with a mechanism similar to numerical interpolation. We carried out an experiment with 149 highschool students, aging fifteen to seventeen, including pre- and post-tests. The results demonstrate that there has been a significant increase between pre- and post-test scores of students that were subject to ASM (p-value = 0.0037). However, there has been no significant difference in student score gains between ASM-determined and teacher-defined sequencing methods.

A minimal sequence of exercises is defined such that it always begins with the easiest exercise and finishes with the most difficult one. The intermediate-level exercises in the minimal sequence are distributed evenly among the easiest and most difficult exercises such that the number of exercises is $\left\lceil \frac{n}{stepsize} \right\rceil$, where n is the total of exercises and the *stepsize*, set by the LO's author, refers to the number of exercises that may be skipped when the student is successful.

Initially, the algorithm presents the exercises in the minimal sequence order. If the number of attempts in an exercise reaches the average number of attempts obtained in the calibration phase,the next exercise presented to the student is of a mid range difficulty, considering the last exercise correctly answered and the current one. Unlike the calibration phase, the student cannot skip exercises and if he/she continually misses the correct answer, the presentation becomes strictly sequential.

6 Ongoing Research

Our ongoing research related to student modeling, including the learner interaction and context, is based on problem statement parameters and ranking solution difficulty in order to support personalization. During the creation of the LO, the teacher-author sets certain parameters that affect the pedagogical strategy, as follow:

- maximum number of retries (attempts) per question;
- tips for each question;
- remediation rules for each question;
- prerequisites for the solution of the exercise, that can be topics, theoretical pages of the LO itself or other LOs in ADAPTFARMA;
- difficulty degree for each question in the range $[1 - 10]$, such that $[1 - 2]$ means very easy, $[3 - 4]$ means easy, $[5 - 6]$ means medium, $[7 - 8]$ means difficult and $[9 - 10]$ means very difficult;
- exercises sequencing strategy, that can be difficulty-biased, teacher-defined or ASM-determined (presented in the previous section).

The student profile is assembled from the previous parameters. By analysing the tips used and relating them to associated prerequisites, the system can provide feedback to both teacher and student on topics that should be further explored or even recommend other complete LOs to be inspected. In addition, the difficulty degree of the questions and the student rating can be updated after each problem solving session has finished.

7 Conclusion and Future Work

The personalization support in learning systems can include adaptive mechanisms of assessment and generation of different content sequencing. We proposed an automatic rating system that can be used as an additional tool to assess students. Depending on the number of attempts and the difficulty degree of a question, different students can get different scores for the same solution. Also, we proposed an algorithm for sequencing exercises using a formalization of the intuitive notion of difficulty degree combined with a mechanism similar to numerical interpolation. All that was implemented in the ADAPTFARMA environment, a web authoring tool for creating and executing LOs.

Future research concentrates in adding new features to ADAPTFARMA in two ways. Firstly, we are working in a deeper approach to user adaptation that includes more dimensions than just the matching between problem difficulty and student skill. One such new feature will be a function for generating problem statements based on teacher-defined problem template parameters as in [4] and [3]. Secondly, on the interface side, more interaction modes will be available to improve collaboration tasks for monitoring student performance progress.

References

1. John Champaign and Robin Cohen. A Model for Content Sequencing in Intelligent Tutorign Systems Based on the Ecological Approach and Its Validation Through Simulated Students. pages 486–491. Association for the Advancement of Artificial Intelligence (AAAI), 2010.

2. Albert T. Corbett and John R. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1994.
3. R.M. Garcia Rioja, S. Gutierrez Santos, A. Pardo, and C.D. Kloos. A parametric exercise base tutoring system. In *Frontiers in Education, 2003. FIE 2003 33rd Annual*, volume 3, pages S1B_20–S1B_26, Nov 2003.
4. Julio Guerra, Shaghayegh Sahebi, Peter Brusilovsky, and Yu-Ru Lin. The Problem Solving Genome: Analyzing Sequential Patterns of Student Work with Parameterized Exercises. In Pardos Z. Mavrikis M. McLaren B.M. Stamper, J., editor, *Proceedings of the 7th International Conference on Educational Data Mining*, pages 153–160, 2014.
5. Seiji Isotani, Deanne Adams, Richard E. Mayer, Kelley Durkin, Bethany Rittle-Hohnson, and Bruce M. McLaren. Can Erroneous Examples Help Middle-School Students Learn Decimals? volume 6964 of *Lecture Notes in Computer Science*, pages 181–195, Palermo, 2011. Springer Berlin Heidelberg.
6. Diego Marczal and Alexandre Direne. Farma: Uma ferramenta de autoria para objetos de aprendizagem de conceitos matemáticos. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 23, 2012.
7. Niels Pinkwart and Frank Loll. Comparing three approaches to assess the quality of students' solutions. In Darina Dicheva, Riichiro Mizoguchi, and Niels Pinkwart, editors, *AIED 2009 Workshops Proceedings Volume 2, SWEL'09: Ontologies and Social Semantic Web for Intelligent Educational Systems Intelligent Educational Games*, pages 81–85, Jul 2009.
8. Gautham Adithya Ravi and Sergey Sosnovsky. Exercise difficulty Calibration Based on Student Log Mining. In F. Mdritscher, V. Luengo, E. Lai-Chong Law, and U. Hoppe, editors, *Proceedings of DAILE'13: Workshop on Data Analysis and Interpretation for Learning Environments*, Villard-de-Lans (France), Janeiro 2013.
9. Carlotta Schatten and Lars Schmidt-Thieme. Adaptive Content Sequencing without Domain Information. *6th International Conference on Computer based Education*, April 2014.
10. Avi Segal, Ziv Katzir, Kobi Gal, Guy Shani, and Bracha Shapira. EduRank: A Collaborative Filtering Approach to Personalization in E-learning. In Pardos Z. Mavrikis M. McLaren B.M. Stamper, J., editor, *Proceedings of the 7th International Conference on Educational Data Mining*, pages 68–75, 2014.
11. James R. Segedy, Kirk M. Loretz, and Gautam Biswas. Suggest-assert-modify: A taxonomy of adaptive scaffolds in computer-based learning environments. In Gautam Biswas, Roger Azevendo, Valerie Shute, and Susan Bull, editors, *AIED 2013 Workshops Proceedings Volume 2: Scaffolding in Open-Ended Learning Environments (OELEs)*, pages 73–80, Jul 2013.