4th International Workshop on Personalization Approaches in Learning Environments (PALE 2014)

held in conjunction with

22nd International Conference User Modeling, Adaptation, and Personalization (UMAP 2014)

Aalborg, Denmark, July 11th, 2014

Proceedings edited by:

Milos Kravcik

Olga C. Santos

Jesus G. Boticario

Published in CEUR workshop proceedings

ISSN 1613-0073

Table of contents

4th International Workshop on Personalization Approaches in Learning Environments (PALE 2014)

Preface

Milos Kravcik[1], Olga C. Santos[2], Jesus G. Boticario[2]

[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/

Abstract. Personalization approaches in learning environments 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 fourth edition includes 8 papers dealing with student's performance, modeling the user profile in a standardize way, computing attributes for learner modeling, detecting affective states to improve the personalized support, and applying user modeling approaches in new contexts, such as MOOCs and gamified environments.

1 Introduction

The 4th International Workshop on Personalization Approaches in Learning Environments (PALE)[1] takes place on July 11th, 2014 and is held in conjunction with the 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014). Since the topic can be addressed from different and complementary perspectives, PALE workshop aims to offer a fruitful crossroad where interrelated issues can be contrasted and discussed. PALE 2014 is a follow-up of the three previous editions of PALE (which took place at UMAP 2011, UMAP 2012 and UMAP 2013).

In order to foster the sharing of knowledge and ideas to research on these issues, PALE format moves away from the classic 'mini-conferences' approach and follows the Learning Cafe methodology[2] to promote discussions on open issues regarding personalization in learning environments. Four Learning Café sessions are set up.

[1] http://adenu.ia.uned.es/workshops/pale2014/
[2] http://adenu.ia.uned.es/workshops/pale2014/format.htm

Each one consists 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 is moderated by the presenter of the paper. In the middle of the session, participants change tables to move swap group discussions and thus, promote sharing of ideas among groups. In this way, participants attending the workshop benefit both from interactive presentations and constructive work.

The target audience of the workshop consists of researchers, developers, and users of personalized and adaptive learning environments. Additionally, the contributions of this workshop have also been disseminated in the Educational Data Mining community at the EDM 2014 conference, which took place on July 4th-7th, 2014 in London.

As a long-standing workshop series (for 4 years now, annually run at UMAP) PALE workshop has established itself as a mature channel for disseminating research ideas on learning environments' personalization. This would have not been 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. As a way to compile the progress achieved in this field, a special issue on User modeling to Support Personalization in Enhanced Educational Settings is being guest edited by PALE organizers in the International Journal of Artificial Intelligence in Education[3]. Papers from PALE editions that presented ideas that have already produced relevant findings have been selected and invited to contribute an extended version of their papers for this special issue. The review process established by the journal is followed to assure that papers finally accepted meet the journal's quality standards.

In the following, we introduce PALE 2014 motivation and themes and 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 student outcomes, adaptive learning assessment, and evaluation of personalized learning solutions.

[3] http://ijaied.org/journal/cfp/

From the past experience, we have identified new research areas of interest in this field 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 needed 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, other related topics are to be considered in the learner modeling, including affective states of the learner, 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 raise the attention to share and discuss the current research on how user modeling and associated artificial intelligent techniques provide personalization support in a wide range of learning environments, which are increasingly more sensitive to 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 can this wide range of situations and features impact on modeling the learner interaction and context. Furthermore, we aim to cover the every time more demanding need of personalized learning in massive open online courses (MOOCs).

The higher-level research question addressed in this workshop edition is: “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
- Learner and context awareness
- Social and educational issues to be addressed
- Open-corpus educational systems
- Adaptive mobile learning
- Successful methods and techniques
- Reusability, interoperability, scalability
- Evaluation of adaptive learning environments

3 Contributions

A blind 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, 8 submissions (out of 10) were accepted, which discuss ideas and progress on several interesting topics: modeling issues such as student's performance, user's profile management in a standardize way (i.e., IMS-LIP), taking care of learner's attributes such as reputation and mind wandering, detecting affective states to improve the personalized support, and applying user modeling in new contexts, such as MOOCs and gamified environments.

Khajah et al. [1] present a unified view of two complementary models of student performance, the Item Response Theory, which allows modeling different student abilities and problem difficulties, and the Knowledge Tracing, which captures skill acquisition and evaluate both models under a common evaluation metric. Results show that both models are equivalent and only differ in their training procedure.

Sawadogo et al. [2] focus on user assistance in an interactive and adaptive system. They proposed a modeling of a scientific user who is a researcher in a personal resource management system. The presented approach assists the users in the consolidated management of their resources and their environment, based on the user's profile. The methodology is based on the IMS-LIP standard extension and the user's trace management.

Lobo et al. [3] introduce how to compute a transferable and domain-independent reputation indicator to support the collaborative behavior and encourage the motivation of students in collaborative learning environments that considers the information extracted from social network analysis, statistical indicators, and opinions received by students in terms of ratings.

Bixler et al. [4] present a proactive personalized learning environment in which learners are provided with materials that would potentially reduce the propensity to mind wander during learning by optimizing learning conditions (e.g., text difficulty and value) for individual learners, and evaluate the performance of such a system by comparing the proposed method to two non-adaptive alternatives.

Arevalillo-Herráez et al. [5] present an intelligent tutoring system that adapts hints for learners to the line of reasoning (i.e. solution scheme) the student is currently following, and discuss some extensions to build a model of the student's most relevant skills aimed at providing a closer behavior to a human expert, by considering both previous interactions and the learner's affective state.

Ocumpaugh et al. [6] propose an extension of the BROMP field observation protocol to take into account behaviors and affective states not previously established during observations of educational multi-user virtual environments, such as disgust and creative meta-narrative. This protocol is used to collect ground truth data for sensor-free models of affect and behavior and to study student engagement in learning environments. The disgust and creative meta-narrative constructs considered in this contribution are not typically coded during field observations of educational software, but they may prove important as virtual worlds are used for educational instruction.

Henning et al. [7] discuss educational and technical challenges for the usage of MOOCs in higher education. In particular, how to make MOOCs more suitable for a greater variability of learning needs by semantically annotating their parts and running them in a semantically enhanced learning platform that provides personalized learning pathways for each learner through didactically meaningful learning object recommendations.

Tang and Kay [8] present their ideas and guidelines for applying gamification as meta-cognitive scaffolds in open learning environments such as MOOCs, and illustrate this approach through examples of how the guidelines can be applied.

4 Conclusions

In this 4th edition of PALE contributions have addressed some of the gaps identified in the state of the art, such as providing open learner models in terms of standards, the modeling of learners' affective and mental states, and the personalization support in new contexts, such as MOOCs and gamified environments.

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, tabletops), impacting on modeling the learner interaction and context. We expect that future editions in PALE can progress on this direction.

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: Miguel Arevalillo, Mária Bieliková, Rafael Calvo, Rosa Carro, Cristina Conati, David Cooper, Dragan Gašević, Sabine Graf, Tomáš Horváth Kinshuk, Iolanda Leite, Noboru Matsuda, Alexander Nussbaumer, Alexandros Paramythis, Diana Perez-Marín, Elvira Popescu and Carsten Ullrich. The organization of the PALE workshop relates and has been partially supported by the following projects: Learning Layers (FP7 ICT-318209) funded by the 7FP of the European Commission and MAMIPEC (TIN2011-29221-C03-01) funded by the Spanish Ministry of Economy and Competence.

References

1. Khajah, M., Huang, Y., Gonzalez-Brenes, J., Mozer, M. and Brusilovsky, P. Integrating Knowledge Tracing and Item Response Theory: A Tale of Two Frameworks. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments

(PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 7-15 (2014)
2. Sawadogo, D., Champagnat, R. and Estraillier, P. User profile modelling for researcher digital resource management systems. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 16-23 (2014)
3. Lobo, J.L., Santos, O.C. and Boticario, J.G. Towards a transferable and domain-independent reputation indicator to group students in the Collaborative Logical Framework approach. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 24-32 (2014)
4. Bixler, R., Kopp, K., D'Mello, S. Evaluation of a Personalized Method for Proactive Mind Wandering Reduction. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 33-41 (2014)
5. Arevalillo-Herráez, M., Arnau, D., Marco-Giménez, L., González-Calero, J.A., Moreno-Picot, S., Moreno-Clari, P., Ayesh, A., Santos, O.C., Boticario, J.G., Saneiro, M., Salmeron-Majadas, S., Cabestrero, R., Quirós, P. Providing Personalized Guidance in Arithmetic Problem Solving. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 42-48 (2014)
6. Ocumpaugh, J., Baker, R.S., Kamarainen, A.M., Metcalf, S.J. Modifying Field Observation Methods on the Fly: Metanarrative and Disgust in an Environmental MUVE. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 49-54 (2014)
7. Henning, P., Heberle, F., Streicher, A., Swertz, C., Bock, J. and Zander, S. Personalized Web Learning: Merging Open Educational Resources into Adaptive Courses for Higher Education. In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 55-62 (2014)
8. Tang, L.M. and Kay, J. Gamification: metacognitive scaffolding towards long term goals? In proceedings of the 4th Workshop on Personalization Approaches for Learning Environments (PALE 2014). Kravcik, M., Santos, O.C. and Boticario, J.G. (Eds.). 22nd conference on User Modeling, Adaptation, and Personalization (UMAP 2014), CEUR workshop proceedings, this volume, 63-68 (2014)

Integrating Knowledge Tracing and Item Response Theory: A Tale of Two Frameworks

*Mohammad M. Khajah[1], *Yun Huang[2], *José P. González-Brenes[3], Michael C. Mozer[1], Peter Brusilovsky[2] *

[1] Department of Computer Science and Institute of Cognitive Science, University of Colorado, Boulder, CO 80309-0430, USA
{mohammad.khajah,mozer}@colorado.edu

[2] Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, USA
{yuh43,peterb}@pitt.edu

[3] Digital Data, Analytics & Adaptive Learning, Pearson Research & Innovation Network, Philadelphia, PA 19103, USA
jose.gonzalez-brenes@pearson.com

Abstract. Traditionally, the assessment and learning science communities rely on different paradigms to model student performance. The assessment community uses Item Response Theory which allows modeling different student abilities and problem difficulties, while the learning science community uses Knowledge Tracing, which captures skill acquisition. These two paradigms are complementary – IRT cannot be used to model student learning, while Knowledge Tracing assumes all students and problems are the same. Recently, two highly related models based on a principled synthesis of IRT and Knowledge Tracing were introduced. However, these two models were evaluated on different data sets, using different evaluation metrics and with different ways of splitting the data into training and testing sets. In this paper we reconcile the models' results by presenting a unified view of the two models, and by evaluating the models under a common evaluation metric. We find that both models are equivalent and only differ in their training procedure. Our results show that the combined IRT and Knowledge Tracing models offer the best of assessment and learning sciences – high prediction accuracy like the IRT model, and the ability to model student learning like Knowledge Tracing.

Keywords: Knowledge Tracing, IRT, Parameter learning

1 Introduction

In many instructional settings, students are graded by their performance on instruments such as exams or homework assignments. Usually, these instruments are made of *items* – questions, problems, parts of questions – which are graded individually. Recent interest in online education, such as Massively Open Online Courses, promises large amounts of data from students solving items over

* The first three authors contributed equally to the paper.

time. The assessment and learning science communities offer two paradigms to model such data. Traditionally, the assessment community relies on *Item Response Theory* (IRT) [12] which infers individual differences amongst students and items, but it does not account for student learning over time. The learning science community uses Knowledge Tracing [2] to estimate skill acquisition as a function of practice opportunities. Although Knowledge Tracing captures student learning, it assumes that students and items do not vary in abilities or difficulties – any two items involving the same skill are assumed to be equivalent.

Empirically we know that neither models' assumptions are correct – these two paradigms are complementary. Earlier attempts towards unifying these two paradigms within the Knowledge Tracing framework only individualize students [9, 10, 16] or items[4, 11, 13] but not both. It is only recently that serious efforts have been made to integrate both student and item effects into Knowledge Tracing. Specifically, two highly related models based on a principled synthesis of Knowledge Tracing and IRT [3, 6] were proposed. The two models were evaluated on different data sets, using different evaluation metrics and with different ways of splitting the data into training and testing sets. In this paper we reconcile the models' results by presenting a unified view of the two models, and by evaluating the models under a common evaluation metric.

The rest of this paper is organized as follows. Section 2 describes the two recent methods that unify Knowledge Tracing and Item Response Theory. Section 3 provides empirical evaluation. Section 4 concludes.

2 Methods

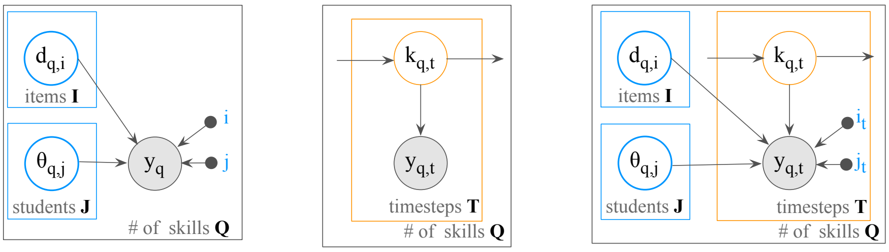
Recently, two models were proposed independently which synthesize Knowledge Tracing and IRT: FAST [3] and LFKT [6]. Although the two models are described in somewhat different terms, they are nearly equivalent, with the key difference being their training method. We now present a unified view of the two models, and then explain their parameter estimate procedures.

2.1 Model Structure

Figure 1 uses plate notation to describe IRT, Knowledge Tracing, and the combined model. In plate notation, the clear nodes represent latent variables; the light gray nodes represent variables that are observed only in training; dark nodes represent variables that are both visible in training and testing; plates represent repetition of variables. We omit drawing the parameters and priors.

Figure 1a shows the plate diagram of the Rasch model, the simplest IRT model. Rasch treats each skill q independently, and can be modeled using logistic regression with binary variables indicators for each student i and each item j. The regression coefficients θ_q and d_q of the binary features can then be interpreted as student ability and item difficulty, respectively. The binary observation variable (y_q) represents whether the student gets an item correct:

$$p(y_q) = \text{logistic}(\theta_{q,i}, d_{q,j}) = \frac{1}{1 + e^{-(\theta_{q,j} + d_{q,i})}} \qquad (1)$$



(a) Item Response Theory (b) Knowledge Tracing (c) Combined IRT and KT

Fig. 1: Plate diagram notation for different student models

Figure 1b describes the Knowledge Tracing model. Knowledge Tracing uses Hidden Markov Models (HMMs)to infer the binary latent student knowledge state ($k_{q,t}$) indicating whether the skill has been mastered at the t^{th} learning opportunity of skill q. The *transition probabilities* between latent states are often referred as learning and forgetting probabilities, and the *emission probabilities* are commonly referred as guess and slip probabilities. The binary variable $y_{q,t}$ represents whether the student gets an item correct:

$$P(y_{q,t}|y_{q,1}...y_{q,t-1}) = \sum_{\substack{l \in \{\text{mastered},\\ \text{not mastered}\}}} \overbrace{P(y_{q,t}|k_t = l)}^{\textit{emission probability}} \cdot P(k_t = l|y_{q,1}...y_{q,t-1}) \qquad (2)$$

Figure 1c shows the combined model. It replaces the emissions with IRT:

$$P(y_{q,t}|y_{q,1}...y_{q,t-1}) = \sum_{\substack{l \in \{\text{mastered},\\ \text{not mastered}\}}} \overbrace{\text{logistic}(d_{q,i_t}, \theta_{q,j_t}, c_{q,l})}^{\text{IRT}} \cdot P(k_t = l|y_{q,1}...y_{q,t-1}) \qquad (3)$$

Here, the logit is parametrized by the difficulty d of the item i, the ability θ of student j and a bias c that is specific to whether the student has mastered the skill. Both Knowledge Tracing and IRT can be recovered from the combined model with different choices of parameter values. For example, when the abilities and difficulties are zero, the combined model is equivalent to Knowledge Tracing. When the bias terms are the same (i.e., $c_{\text{not mastered}} = c_{\text{mastered}}$), we get IRT.

2.2 Parameter Learning

We now briefly review two recent proposals to learn the combined model. A thorough discussion can be found elsewhere [3, 6]. González-Brenes et al. [3] use a recent variant of the EM algorithm [1] that allows learning HMMs with arbitrary features. Although the original framework allows general features to be incorporated into Knowledge Tracing, the model becomes equivalent to our combined model when limiting the features to IRT features only.

Table 1: Basic Dataset Statistics

	Geometry	**Physics**	**Statics**	**QuizJET**
Trials	5,104	110,041	189,297	6,549
Students	59	66	333	110
Problems	139	4,816	1,223	95
Skills	18	652	156	19
Mean Seq. Length	8.0	4.5	6.0	4.7
Mean Correct	75%	83%	77%	60%

Alternatively, Khajah et al. [6] use Bayesian estimation techniques to learn the combined model. They used slice sampling, a MCMC algorithm that generates samples of the joint posterior distribution of the model. This allows using priors on abilities and difficulties which can be used to generalize to unseen students and items. Also, their model allows to fit student ability parameters across different skills – using data from a student's performance on one skill to predict their performance on a different skill.

3 Results

We evaluate our student models by how accurately they predict future student performance. We operationalize predicting future student performance as the classification task of predicting which students solved correctly the items in a held-out set. We evaluate them using a popular machine learning metric, the *Area Under the Curve* (AUC) of the Receiver Operating Characteristic (ROC) curve. The AUC is an overall summary of diagnostic accuracy. AUC equals 0.5 when the ROC curve corresponds to random chance and 1.0 for perfect accuracy. We report the 95% confidence intervals with an implementation of the bootstrap hypothesis test method (`http://subcortex.net/research/code/`), a method that corrects for the non-independence of the points of the ROC.

We use data from four different intelligent tutoring systems: the Geometry Cognitive Tutor [7], the Andes Physics Tutor [15], OLI Statics [14] and QuizJET Java programming tutor [5]. The first three datasets are available on the PSLC Datashop [8]. Table 1 shows a summary of their descriptive statistics.

We divide each dataset into skill-specific subsets consisting of the sequence of trials for each student involving items that require a particular skill. We refer to these sequences as student-skill sequences. If multiple skills are associated with a item, we treat the combination of skills as one unique skill. The last 20% of trials from each sequence were placed in the testing set. Thus, generalization to new skills is not required. For each trial, we compute the prediction (probability of a correct response). Predicted outcomes in the test set are then used to calculate the AUC.

Bar heights in figure 2 are the AUC values for each model and error bars represent 95% confidence intervals. We evaluated the Bayesian and Maximum Likelihood versions of knowledge tracing, IRT, and the combined model. On the smallest dataset, geometry all models perform similarly during testing. On the next two larger datasets, QuizJET and physics, IRT and the combined models perform similarly, beating knowledge tracing. Neither IRT or the combined models emerge as a clear winner. However, on the largest dataset, statics, the Bayesian combined model outperforms all other models significantly. In this dataset, IRT trained using Maximum Likelihood beats the Bayesian version, which might be due to the effects of strong priors on student abilities and item difficulties in the MCMC-trained version. This would also explain why the Bayesian IRT would gain advantage in smaller datasets where the priors influence the most.

In three datasets, the Bayesian version of Knowledge Tracing beats Maximum Likelihood. Our hypothesis is that MCMC training used in Bayesian estimation is more effective at avoiding local optima. We do not think it is due to the use of priors, because we used uninformative priors.

We hypothesize that the reason why the combined model does not outperform IRT is because of the order in which items are presented to students. Specifically, if the items are presented in a relatively deterministic order, the item's position in the sequence of trials is confounded with the item's identity. IRT can exploit such a confound to implicitly infer performance levels as a function of experience, and therefore would have the same capabilities as the combined model which performs explicit inference of student knowledge state. To investigate this, we compute the mean order in which items are presented to students. In Figure 3, the horizontal axis ranks item indices by their mean presentation order, and the vertical axis is the mean order in which items are shown to students. Flat horizontal lines in this plot suggest random item ordering but they may also confound the case where students are assigned to only one item from a set of items. On geometry, we don't see any flat sections which suggests fixed item ordering. Next, the QuizJET dataset exhibits periodic flat sections, but these could be due to students being assigned to single specific items out of sets of items. On the physics and statics datasets, we see a flat line followed by a curve which suggests an initial random assignment of items followed by more structured item ordering. So, there is less information overlap between student learning and item difficulties in the physics and statics datasets, thereby allowing the combined model to put its extra parameters to good use. We plan to carry out more rigours tests of this conclusion in the future.

One of the goals of intelligent tutoring systems is to personalize learning. This requires models that accurately estimate the student's knowledge over time, which is possible with Knowledge Tracing and the combined model, but not with IRT. However, Knowledge Tracing assumes that all items within a skill are equally difficult. It also assumes that all students within a skill share the same initial level of knowledge, learning rate, guessing and slipping probabilities. This may lead to inaccurate student learning estimates which reduces the efficacy of an intelligent tutoring system. To investigate, we calculate the estimated learning

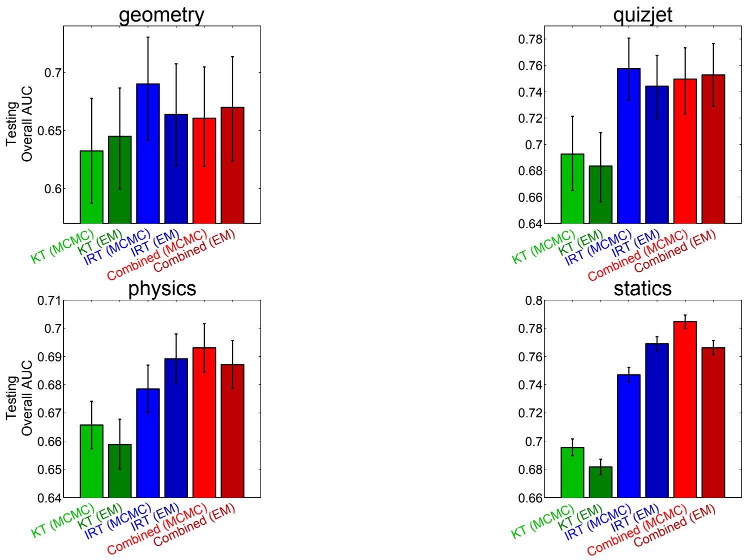


Fig. 2: Test set AUC scores of six models on four datasets. Higher values indicate better performance. Error bars correspond to 95% confidence intervals.

of a student as the probability of mastery at the last practice opportunity minus the probability of mastery at the first practice opportunity for each student-skill sequence ($p(k_T = \text{master}) - p(k_0 = \text{master})$). A value of 0 indicates no difference whilst a value of 1 indicates maximum difference. Figure 4 shows the mean difference in the estimated student learning over all student-skill sequences. For clarity, we omit to draw the estimate of IRT, which assumes no learning occurs. The combined model generally gives higher estimates of student learning than Knowledge Tracing. This suggests that item and student effects are not zero within a skill, which violates the Knowledge Tracing assumptions. Hence, the learning estimates produced by the combined model are more trustworthy than Knowledge Tracing.

4 Conclusion

In this paper we investigate two recent alternatives that integrate Knowledge Tracing and IRT. We discover that both models are in fact equivalent and differ only in their training procedure – using either Maximum Likelihood or Bayesian Estimation. We compare both training procedures, Maximum Likelihood and Bayesian estimation, using the same four datasets, cross validation splits and evaluation metric. We find out that both training methods have similar performance, with a small advantage to the Bayesian method in the largest dataset we used. Future work may investigate why this is the case. The combined model

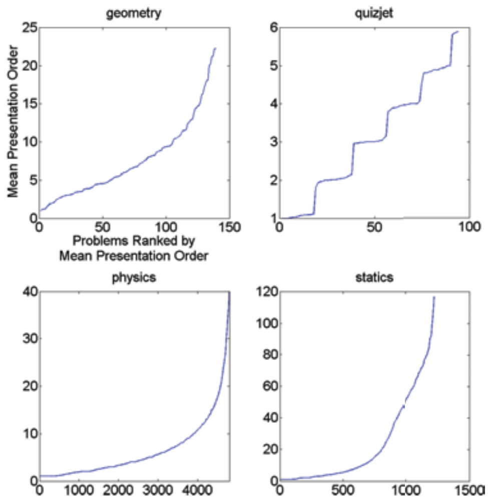


Fig. 3: Mean presentation order for each item in all four datasets (plates). The horizontal axis ranks item indices by their mean presentation order. The vertical axis is the mean order in which items are shown to students.

only outperforms IRT in one dataset. In future work we will investigate whether the lack of improvement is due to a confound of item identity and position in the sequence of trials when nearly deterministic trial sequences are presented

We find that the combined method persistently outperforms Knowledge Tracing, and unlike IRT, it is able to model student learning. Future work may evaluate how useful the combined model is for personalizing learning in an intelligent tutoring system.

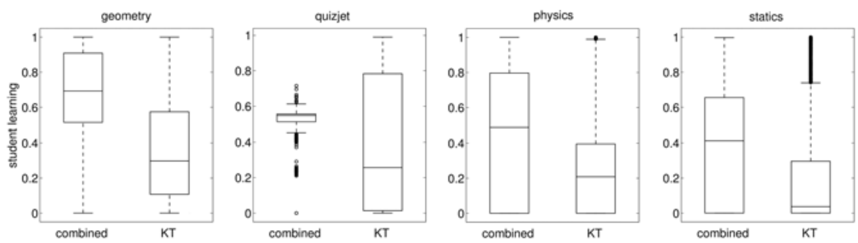


Fig. 4: Boxplot of the estimated student learning of Knowledge Tracing and the combined model. We omit IRT, because it always assumes no learning.

References

[1] Taylor Berg-Kirkpatrick, Alexandre Bouchard-Côté, John DeNero, and Dan Klein. Painless unsupervised learning with features. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 582–590. Association for Computational Linguistics, 2010.

[2] AlbertT. Corbett and JohnR. Anderson. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278, 1995. ISSN 0924-1868. doi: 10.1007/BF01099821. URL `http://dx.doi.org/10.1007/BF01099821`.

[3] J.P. González-Brenes, Y. Huang, and P. Brusilovsky. General features in knowledge tracing: Applications to multiple subskills, temporal item response theory, and expert knowledge. In *Submitted to the 7th International Conference on Educational Data Mining (2014)*.

[4] Sujith M. Gowda, Jonathan P. Rowe, Ryan Shaun Joazeiro de Baker, Min Chi, and Kenneth R. Koedinger. Improving models of slipping, guessing, and moment-by-moment learning with estimates of skill difficulty. In Mykola Pechenizkiy, Toon Calders, Cristina Conati, Sebastián Ventura, Cristóbal Romero, and John C. Stamper, editors, *EDM*, pages 199–208. www.educationaldatamining.org, 2011. ISBN 978-90-386-2537-9.

[5] I-Han Hsiao, Sergey Sosnovsky, and Peter Brusilovsky. Adaptive navigation support for parameterized questions in object-oriented programming. In *Learning in the Synergy of Multiple Disciplines*, pages 88–98. Springer, 2009.

[6] M. Khajah, R. M. Wing, R. V. Lindsey, and M. C. Mozer. Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *In submission*, 2014.

[7] Ken Koedinger. Geometry area (1996-97), February 2014. URL `https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=76`.

[8] K.R. Koedinger, R.S.J.d. Baker, K. Cunningham, A. Skogsholm, B. Leber, and J. Stamper. A data respository for the EDM community: The pslc datashop. In C. Romero, S. Ventura, M. Pechenizkiy, and R.S.J.d. Baker, editors, *Handbook of Educational Data Mining*, 2010. http://pslcdatashop.org.

[9] Jung In Lee and Emma Brunskill. The impact on individualizing student models on necessary practice opportunities. In Kalina Yacef, Osmar R. Zaïane, Arnon Hershkovitz, Michael Yudelson, and John C. Stamper, editors, *Proceedings of the 5th International Conference on Educational Data Mining*, pages 118–125, Chania, Greece, 2012. www.educationaldatamining.org. URL `http://educationaldatamining.org/EDM2012/uploads/procs/Full_Papers/edm2012_full_11.pdf`.

[10] Zachary A. Pardos and Neil T. Heffernan. Modeling individualization in a bayesian networks implementation of knowledge tracing. In *Proceedings of the 18th international conference on User Modeling, Adaptation, and Personalization*, UMAP'10, pages 255–266, Berlin, Heidelberg, 2010. Springer-Verlag. ISBN 3-642-13469-6, 978-3-642-13469-2. URL `http://dx.doi.org/10.1007/978-3-642-13470-8_24`.

[11] ZacharyA. Pardos and NeilT. Heffernan. Kt-idem: Introducing item difficulty to the knowledge tracing model. In JosephA. Konstan, Ricardo Conejo, Jos L. Marzo, and Nuria Oliver, editors, *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 243–254. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-22361-7. doi: 10.1007/978-3-642-22362-4_21. URL http://dx.doi.org/10.1007/978-3-642-22362-4_21.

[12] G. Rasch. Probabilistic models for some intelligence and attainment tests. In *Paedagogike Institut, Copenhagen.*, 1960.

[13] Sarah Schultz and Trenton Tabor. Revisiting and extending the item difficulty effect model. In *In Proceedings of the 1st Workshop on Massive Open Online Courses at the 16th Annual Conference on Artificial Intelligence in Education*, pages 33–40, Memphis, TN., 2013.

[14] Paul Steif and Norman Bier. Oli engineering statics - fall 2011, February 2014. URL https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=507.

[15] Kurt VanLehn. Usna physics fall 2006, February 2014. URL https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=126.

[16] Michael Yudelson, Kenneth R. Koedinger, and Geoffrey J. Gordon. Individualized bayesian knowledge tracing models. In *Artificial Intelligence in Education - 16th International Conference (AIED 2013)*, pages 171–180, Memphis, TN, 2013. Springer.

User Profile Modelling for Digital Resource Management Systems

Daouda Sawadogo, Ronan Champagnat, and Pascal Estraillier

L3i Laboratory, University of La Rochelle, France
{daouda.sawadogo,ronan.champagnat,pascal.estraillier}@univ-lr.fr
http://l3i.univ-lr.fr

Abstract. In this paper, we focus on user assistance in an interactive and adaptive system. The increase in production of digital data, these last two years has raised several issues regarding the management of heterogeneous and multiple source data in the user's environment. One distinctive feature of user assistance is the user model that represents essential information about each user. We propose a modeling of a scientific user, who is a researcher, in a personal resource management system. Our methodology is based on the IMS-LIP standard extension and the user's trace management. Our work can assist users in the consolidated management of their resources and their environment, based on the user's profile. Experimental results obtained by the emperical evaluation in our laboratory are presented.

Keywords: User modelling, adaptation, profile management, trace-based system, research resource management system

1 Introduction

Currently, user profiles have a very important role in all digital environments [1] [2]. The use of profiles is one of the means that enables systems to adapt to the specificities of its users in the digital environments. Every information system, created as a service, ought to support mainly the digital resources in order to respect the logic of usage required by the system. In this logic of usage, the digital user is an essential actor in the landscape of digital information, he acts and interacts with others; making use of, transforming and producing information. Each individual composes their own digital identity, thus generating the e-reputation that characterises them and for which they have to take care [3]. Our general objective was to design an application that would provide a consolidated management of the users' digital resources and environment. We opted for an application that would be used by researchers as several studies had shown that the problem of researcher profiling poses many challenging issues including : the automatic extraction of researcher profiles from distributed sources (homepages, indexing authorities, etc.) [4], the consistency and completeness of information and the resolution of ambiguity [5]. These issues are further aggravated by the explosion of information of research artifacts with the research community. Researchers'

Facets of user assistance	Characteristics to act on this facet
Resources management	Characterises digital resources in order to calculate the relevance of these resources for users in the systems
Process management	Characterises resources management rules and user's process management interaction to adapt the usage
Collaboration management	Criterions for building a user trust community [7], resource sharing and recommendation

Table 1. Facets of assistance in PRISE

information is mostly scattered and is represented in a syntactic manner [4] (not meaningfully described) thus minimizing the interoperability between heterogeneous knowledge and information sources [6]. In our application we first decided to design a consistent researcher profile based on IMS-LIP extension and a trace-based system which can be used to complete the researcher profile. Indeed, the user's profiles provide the system with pertinent information to assist the user in aspects defined in Table1.
Our contribution concerns the first facet of this table, i.e. resources management. In this paper we proposed :

- an extension of the IMS-LIP model for research;
- criterions to characterise the relevance of a digital resource according to the user profile.

2 State of the art

The user model is a representation of information about an individual user that is essential for an adaptive system to provide the adaptation effect. i.e. to behave differently for different users[2]. The researcher is like a learner in the system, who uses resources to acquire knowledge and produce scientific results.

Usually, five most popular and useful features are found when viewing the user as an individual: the user's knowledge, interests, goals, background and individual traits. In [2] The authors discuss modelling the context of a user's work. In our work we wanted to modelize the user in order to assist him in his resource management process and to **characterise his interaction in the environment**.
Some existing learner models in the literature are as follows :

1. IEEE PAPI-Learner [8] (Public and Private Information-Learner specification) is a standard developed within the IEE P1484.2 Learner Model Working Group. Its objective is to specify the semantics and the syntax of a Learner Model, which characterises a learner and his knowledge. It includes elements such as knowledge, skills, abilities, learning styles, records, and personal information. PAPI Learner was initially developed for learning technology

applications but can easily be extended to other types of human related information such as medical and financial applications. PAPI is one of the first standards which provides a framework that organizes learner data. There is lot of learner data that this standard does not take into account [9] and which can be exchangeable between various e-learning systems. This explains why this proposal has been extended by IMS in its new standard IMS-LIP[10].

2. IMS-LIP[11] is based on a data model that describes those characteristics of a learner that are needed for the general purposes of: Recording and managing learning-related history, goals and accomplishments; Engaging a learner in a learning experience; Discovering learning opportunities for learners. The specification supports the exchange of learner information among learning management systems, human resource systems, student information systems, enterprise e-learning systems, knowledge management systems and other systems used in the learning process. We note that IMS LIP provided valuable extensions compared to the PAPI model, but it does not meet all the systems' needs in terms of user data completeness and management, which explains its adaptation within application profiles.
3. SERPOLET [10] offers solutions to the issue of learner data interoperability between e-learning systems. This model is based on IMS-LIP with extension to the need of learner data interoparability.

From these propositions, we chose to extend the IMS-LIP standard for many reasons. Firstly, IMS-LIP is the interoperability standard chosen by CEN/ISSS [12]. Secondly, it defines a user data model as a set of 11 categories to be imported or exported between systems. The IMS-LIP extension in Fig. 1 that we defined provides relevant information about the researcher and his activities. The trace-based system maintains a consistent profile and makes it more complete.

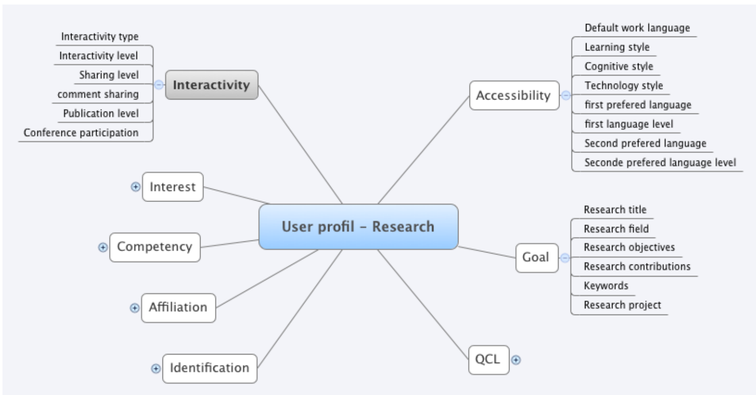


Fig. 1. IMS-LIP extension application profile for researcher

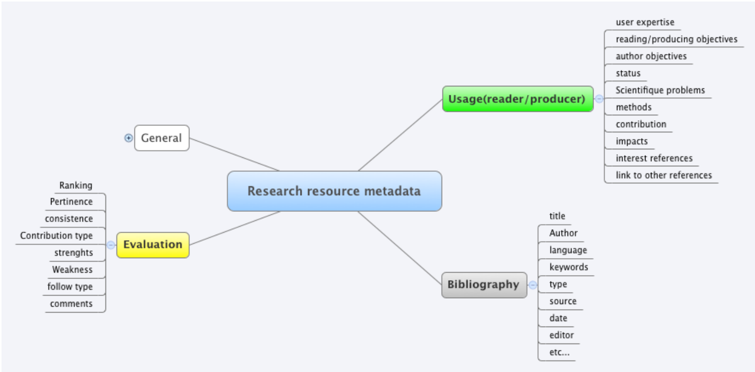


Fig. 2. LOM application profil for researcher resource

We need to understand the interactivity of the user in order to assist him. In [13] the authors proposed an ontology-Based User modelling for Knowledge Management Systems using IMS-LIP. They defined the behaviour category that extended the IMS-LIP model. In our work, we consider that the behaviour is the dynamic information of profile, so we use interactivity to capture the model of user interaction with the resources which can characterise the user model.

3 Our methods

It should be remembered that the objective of the profile is to provide relevant information about the user and provide mechanisms to assist a user in the consolidated management of his resources and environment. Firstly, we proposed an extension of the IMS-LIP model in Fig. 1 to take into account the user's interaction. In digital resources management system it is very important to characterise the user's interactivity type and interactivity level to adapt the digital resource usage process.
Secondly, our methods involved determining what criterions are necessary to characterise the user's digital resources relevance and how to use it the latter. We identified two levels of use of a user's profile:

- the static level: we can directly use the informations, of the user, stored in the database.
- the dynamic level: we can calculate the information of the user from information stored and his context in the system.

The first level requires less effort in the semantic analysis. The informations that we deemed relevant for the profile are : research domain, research field, user's prefered language, user's technology style, qualification level, keywords and research interests
The second level requires more effort in the semantic analysis. The informations

needed are : the user intention [14] [15], the user objective the user interactivity type, the user interactivity level and the scientific problems. These informations are context-aware, so they require a real-time analysis and they characterise the user's information needs. We have chosen these criterions because they are importants for a researcher looking for informations and in the objectives of our study. Our contribution consisted in combining these two types of data on the user, in order to determine the relevance of the resource and to adapt it to the usage needs.
We characterised the resource using all the metadata associated with it. The Fig. 2 shows the main elements of the research resource metadata in our environment. Despite the fact that quite often the metadata are not the main elements that characterise a document. In the context of this work we put forth the hypothesis that the metadata provide a good characterisation of the resource. In our experimental environment 3, we characterised these metadata as mainly: bibliography data [16], usage (reader/producer) data and evaluation data. After that we compared the matching of the profile information 4 with the resource metadata. We experiment as well the use of TF-IDF [17] to calculate these levels of relevancy. However this aspect is not yet ready enough to be presented.

Our methodology for modelling the research user and managing his profile consists in explaining how the data in the user's profiles is collected, how it is fomalized and how it is used to satisfy system assistance needs.

There are two possibilities for collecting the necessary user data [18] :

1. Explicit collection of the data: users' preferences are found explicitly, by asking them to submit the necessary information manually before any personalisation is provided. Explicitly entered profile information is considered to be high quality, but users generally dislike having to spend time and effort submitting data to a system, especially when the benefits may not be immediately obvious. This can make the explicit collection of sufficient profile data difficult[18]. This type of collection is achieved through the fields that we propose in the IMS LIP model extension for the researcher.
2. Implicit collection of data: users' preferences are inferred from their normal interactions with the system. The advantage of collecting profile data this way is that the user is relieved of the burden of having to supply and keep up-to-date the necessary information. But the implicit measures of interest are generally thought to be lower quality than explicitly gathered preferences [19].

The purpose of trace-based management system is to identify the trace elements necessary to propose a relevant user profile and system for the automatic update of user's data.

We considered the activity category in the profile because it is data that we mainly use in our research. There are traces in the past of the users. Furthermore, there are many interactions in an interactive environment and during its execution, actors can generate traces [20]. We combined the user profile and the trace-based system to construct our model. These traces were collected from the interaction of the user with the environment. The users' traces were used by carrying out three steps consecutively. Firstly, a function was defined to collect all traces that were produced by the interaction between the user and the environment. We propose then a step of traces' filtering. This step aims to filter all the traces in order to get the pertinent traces. Finally, we present the filtered traces and integrate them in the user's profile.

4 Experimentation and discussion

To experiment our proposal we created an environment that allows a researcher to manage their digital resources. The system offers a set of tools that make it more effective in the production of scientific results. Our research environment PRISE (PeRsonal Interactive research Smart Environment) includes several tools : digital resources management (Fig. 3), social network including our model of the researcher's profile (Fig. 4) and events management.

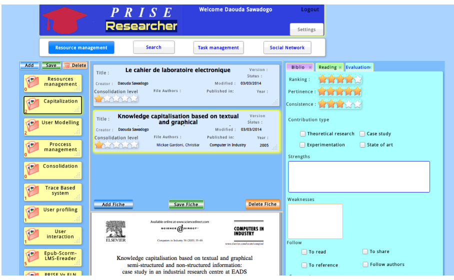


Fig. 3. Digital Resource management in PRISE

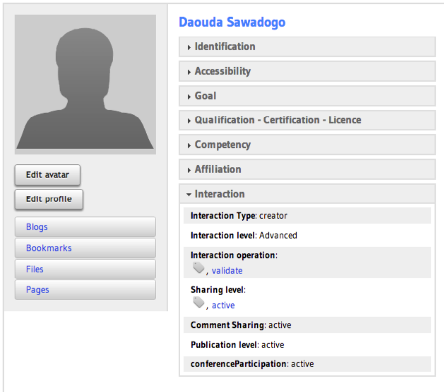


Fig. 4. example of IMS-LIP extension profile for researcher

Through our experimentation in PRISE, we used the elements of the user's profile necessary to achieve the first facets proposed in the Table 1.
Table 2 shows these profile criterions in order to characterise the resource management relevance in the system. These criterions were used to assist the user resources management process, resources relevant for the user and resource recommendation criterions based on the user profile.

Our experimental system implemented our LOM [21] application profile for research. The resource metadata was stored in a NoSQL database using JSON API to manipulate it. NoSQL has the quality required to be a document store database. The user profile model was stored in SQL database for reuse needs and we used REST web services to retrieve the profile information.

Facets of user assistance	Item criterions used in the profil
Resources management	Prefered language; Language level; Research field; Keywords; Research field level; Qualification level; Research interests; interactivity type; interactivity level

Table 2. User profil criterions for user assistance in PRISE

We have discussed our work in terms of the research described in [22] [23]. These authors made an important contribution by automatically characterizing the resource quality using the machine learning. Our contribution completes these works using the user profile and resource metadata. We also found that our user model was more complete and provided relevant information about the researcher in the system. The models we compared, in the main lack information on the objectives of the researcher, their scientific problems, their preferences, as well as entire elements on their activities as well. Our model could also serve as a reference for the social research networks like Mendeley, Academia and ResearchGate etc.

5 Conclusion and future works

In this paper we presented work in progress on the researcher profile modelling in a personal research system. We have identified and used the relevant characteristics to adapt and assist the researcher with his digital resources, to achieve relevant management of his resources. The main contribution was in consistent researcher profile modelling based on IMS-LIP extension and the approach of trace-based management to fill the user profile model use this model to assist user's in research resources management systems.
Future work will firstly consist in improving the dynamic usage of the user profile information and the TF-IDF method to calculate the resource relevance. Secondly, it will provide some mechanisms for a dynamic usage of user's resource process based on the user profile, in order to maintain his environment's consistency and assist him in the consolidated management of his resources.

References

1. Brusilovsky, P.: Methods and techniques of adaptive hypermedia. User Model. User-Adapt. Interact. **6**(2-3) (1996) 87–129
2. Brusilovsky, P., Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In: The Adaptive Web. (2007) 3–53
3. Roose, P.: From reuse to adaptability. Hdr, University of Pau and Adour Country (2008)
4. Oberle, D., Spyns, P.: The knowledge portal "ontoweb". In: Handbook on Ontologies. Springer (2004) 499–516

5. Tang, J., Zhang, D., Yao, L.: Social network extraction of academic researchers. In: ICDM. (2007) 292–301
6. Adnan, S., Tahir, A., Basharat, A., de Cesare, S.: Semantic agent oriented architecture for researcher profiling and association (semora). In: Web Intelligence/IAT Workshops. (2009) 559–562
7. Nepal, S., Sherchan, W., Paris, C.: Building trust communities using social trust. In: Advances in User Modeling. Springer (2012) 243–255
8. IEEE LTSC: Ltsc public and private information for learners, http://www.cen-ltso.net/main.aspx?put=230
9. Paramythis, A., Loidl-Reisinger, S.: Adaptive learning environments and e-learning standards. In: Second European Conference on e-Learning. (2003) 369–379
10. Oubahssi, L., Grandbastien, M.: From learner information packages to student models: Which continuum? In: Intelligent Tutoring Systems, Springer (2006) 288–297
11. IMS GLC: IMS Learner Information Packaging Information Model Specification
12. Ehlers, U.D., Pawlowski, J.M.: Handbook on quality and standardisation in e-learning [electronic resource]. Springer (2006)
13. Razmerita, L., Angehrn, A.A., Maedche, A.: Ontology-based user modeling for knowledge management systems. In: User Modeling. (2003) 213–217
14. Chen, Y.C., Shang, R.A., Li, M.J.: The effects of perceived relevance of travel blogs' content on the behavioral intention to visit a tourist destination. Computers in Human Behavior **30** (2014) 787–799
15. Djamasbi, S., Strong, D.M.: The effect of positive mood on intention to use computerized decision aids. Information & Management **45**(1) (2008) 43–51
16. Lipinski, M., Yao, K., Breitinger, C., Beel, J., Gipp, B.: Evaluation of header metadata extraction approaches and tools for scientific pdf documents. In: JCDL. (2013) 385–386
17. Wu, H.C., Luk, R.W.P., Wong, K.F., Kwok, K.L.: Interpreting tf-idf term weights as making relevance decisions. ACM Transactions on Information Systems (TOIS) **26**(3) (2008) 13
18. Keenoy, K., de Freitas, S., Levene, M., Jones, A., Brasher, A., Waycott, J., Kaszas, P., Turcsanyi-Szabo, M., Montandon, L.: Personalised trails and learner profiling within e-learning environments. Personalised and Collaborative Trails of Digital and Non-Digital Learning Objects **22**(4.1) (2004) 1–43
19. Nichols, D.M.: Implicit rating and filtering. In: Proceedings of the Fifth DELOS Workshop on Filtering and Collaborative Filtering, ERCIM (1997) 31–36
20. Laflaquière, J., Settouti, L.S., Prié, Y., Mille, A.: Trace-based framework for experience management and engineering. In: Proceedings of the 10th International Conference on Knowledge-Based Intelligent Information and Engineering Systems - Volume Part I. KES'06, Springer-Verlag (2006) 1171–1178
21. Barker, P.: What is ieee learning object metadata/ims learning resource metadata? Cetis standards briefings series (2005)
22. Bethard, S., Wetzer, P., Butcher, K., Martin, J.H., Sumner, T.: Automatically characterizing resource quality for educational digital libraries. In: Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries. JCDL '09, ACM (2009) 221–230
23. Wetzler, P., Bethard, S., Leary, H., Butcher, K., Bahreini, S.D., Zhao, J., Martin, J.H., Sumner, T.: Characterizing and predicting the multifaceted nature of quality in educational web resources. ACM Trans. Interact. Intell. Syst. **3**(3) (October 2013) 15:1–15:25

Towards a transferable and domain-independent reputation indicator to group students in the Collaborative Logical Framework approach

Jesus L. Lobo[1], Olga C. Santos[2] and Jesus G. Boticario[2]

[1]Tecnalia Research & Innovation, Bilbao, Spain
jesus.lopez@tecnalia.com
[2]aDeNu Research Group, Artificial Intelligence Dept. Computer Science Sch. UNED, Spain
{ocsantos,jgb}@dia.uned.es

Abstract. Collaborative indicators derived from quantitative statistical indicators of students' interactions in forums can be used by e-learning systems in order to support the collaborative behaviour and motivation of students. The main objective of this research is to achieve a transferable and domain-independent reputation indicator, considering the information extracted from social network analysis, statistical indicators, and opinions received by students in terms of ratings. This paper describes how to consider the reputation indicator in a collaborative environment in order to group students (distributing the most prominent students into different groups) aimed to improve the collaborative indicators (such as initiative, activity, regularity).

Keywords: Collaborative indicators, Reputation indicator, Forum interactions, Grouping students, Social Network Analysis, Collaborative platform

1 Introduction

Providing personalised recommendations to students in order to foster their participation and increase their level of engagement in a collaborative environment is a relevant field that every e-learning system should take into consideration [1]. Collaborative learning environments have been successfully used to support student learning [2]. Using collaborative indicators derived from students' interactions might help an e-learning system in deciding whether to 1) speed up in order to reveal new educational content, 2) slow down in order to go into content in depth , 3) introduce new conversations or messages in order to stimulate new debates and a better understanding of the content, and 4) identify recommendations opportunities that guide students in performing specific actions intended to help their mates on a given task, encouraging participation and improving team work [3].

From previous research [3, 4, 5] carried out by aDeNu research group on collaborative indicators for e-learning environments, statistical indicators (such as number of threads started, number of messages sent, number of replies, etc.) have been proposed as relevant to evaluate the collaboration process. These statistical indicators were collected by a collaborative platform that implemented the CLF (Collaborative Logical Framework) approach on top of dotLRN Learning Management System (LMS) [6] and show their capacity to reveal students' collaboration quality in terms of several students characteristics such as initiative, activity, reputation and regularity [5]. The CLF has been proposed to provide real collaboration in the Logical Framework Approach, and is aimed to facilitate an efficient collaboration among students, grouping them in small clusters for effective collaboration (typically, 4 students per group). Under this framework, three stages have been defined [7]: 1) *Individual stage*: each student works individually to produce his contribution on a given problem; 2) *Collaboration stage*: students have access to the solutions of their mates and must comment (by answering the corresponding forum thread), and rate them; and 3) *Agreement stage*: taking into account the interactions in the two previous stages, a moderator is selected for the group, who is responsible for providing the agreed solution of the group based on the best rated works of the group.

In [3], it was suggested that from forum interactions analysis, those students whose messages receive more replies indicated more interest by fellow students, and this, can be considered a proof of acknowledgment, and thus, of student's reputation. The reputation is a relevant measure of the degree of prominence of an actor in a social network. In turn, [8, 9] showed that reputation was one of the most important attributes for predicting final student performance on the basis of the use of data from online discussion forums.

In this context, this research work aims to complete previous reputation indicator in terms of three different types of analytic data, which are based on forum activity: 1) quantitative information that uses statistical indicators (number of received messages in the threads started by a student and the number of received answers in messages sent by a student), 2) qualitative information that uses the average score of opinions received by the rest of students (rating), and 3) social network information (SNA) and hyperlink analysis [10, 11] that uses the ratio of students' in-links (when a student receives a response from another student). Using the reputation indicator as a reference to form collaborative groups in courses, an e-learning platform that keeps track of the collaboration process and the students' behaviours in terms of the collaborative indicators, could group the most prominent students with those less prominent with the intention of fostering engagement and improving the students' performance. In this way, the collaboration process is expected to be improved [12, 13], and thus, the statistical indicators that reflect student's collaborative characteristics (initiative, activity, regularity).

The work carried out in this research also aims to prove the transferability and domain-independence of the proposal. For this, the CLF approach will be deployed in another e-learning platform (Moodle) showing the transferable characteristic of the collaborative indicators, and also their domain-independence when free-content interaction variables are computed in the same way using the specific interaction data gathered in each environment.

The paper is structured as follows. First, a way to compute the reputation indicator from statistical indicators, rating of students, and social network information is presented. Next, the focus is put on describing how the CLF runs on Moodle Finally, ongoing works are outlined.

2 Reputation basis

As [3] suggested, a reputation indicator should provide information on target student collaboration. Although previous researches [3, 4, 5] took into consideration the reputation indicator from a statistical point of view (*N_r_thrd* as the number of replies to threads started by a student, and *N_r_msg* as the number of replies to messages sent by a student), *N_r_thrd* could be further investigated as one of the most significant indicators to assess student collaboration [3]. For this reason, and being aware of the importance of the reputation in collaboration processes [14], it is of interest to consider a richer definition of this indicator. Additionally, taking into account the results obtained in [9], this research proposes to explore the extension of previous reputation indicator in terms of three different types of analytic data. Grouping students according to this extended reputation indicator could improve the collaboration process, which is expected to improve the computation of the statistical indicators on which initiative, activity and regularity indicators are based. Following a similar approach as [15], which took into consideration several sources of information to define the reputation in terms of a social and scientific scores, the proposed reputation indicator has been composed of three different sources of information: 1) statistical indicators (SI) as quantitative information, 2) rating information (RI) as qualitative information, and 3) information provided by SNA (SNI). Following a similar methodology [3, 5], each of these sources can be normalized between 0 and 1 [9], and computed using a metric to assign a reputation value (Rep) to each student. Different weights (a, b, c) can be used when combining the three sources in the case of correcting some deviations or subjective connotations. A machine learning method, such as linear regression, could learn these weights and automatically compute their relevance:

$$\mathrm{Rep} = \frac{\mathrm{aSI} + \mathrm{bRI} + \mathrm{cSNI}}{\mathrm{a} + \mathrm{b} + \mathrm{c}} \tag{1}$$

For the experiment carried out (see section 4), initially weights used are *a=b=c=1*, as tentative value to start this first experiment.

Reputation has allowed to group students according to its value, pursuing an improvement of the collaborative indicators, and if the experiment shows the expected importance of the reputation indicator, it could be another relevant source of data for the e-learning systems to suggest tailored recommendations and favoring the engagement. The reputation indicator could reflect popularity connotations, above all when one of its three sources (SNI) is based on students' networks and interactions. But the reputation indicator is composed by two other elements (SI and RI) in order to be able to balance the final score in this respect.

2.1 The statistical indicators as quantitative information

The evaluation of information gathered in previous pilot experiences [4] showed that some indicators might have overlapped the description of others, and it was considered the possibility of setting up a range of three values for labelling each indicator instead of using its absolute label. In particular, this research proposes three values to rank initiative, activity, regularity and reputation, namely: improvable, moderate and notable.

The statistical indicators for activity, initiative, regularity and reputation (based on forum conversations started, forum messages sent and replies to student interactions) are calculated following the results of previous works carried out by aDeNu [5]. In the case of reputation and as anticipated above, in [3] two indicators were proposed: the number of replies to threads started by a student (*N_r_thrd*) with respect to the total replies to threads started (*Total_r_thrd*), and the number of replies to messages sent by a student (*N_r_msg*) with respect to the total replies to messages sent (*Total_r_msg*). This work hypothesised that more replies indicated more interest by fellow students, which is proof of acknowledgement. The statistical indicators (SI) can be calculated as follows:

$$\mathrm{SI} = \frac{\mathrm{N\_r\_thrd} + \mathrm{N\_r\_msg}}{Total\_r\_thrd + Total\_r\_msg} \tag{2}$$

2.2 The rating as qualitative information

The instructor is faced with the difficulty of interpreting and evaluating the quality of the participation reflected through students' contributions, considering that current e-learning systems do not provide explicitly many indicators regarding this qualitative information. A reasonable information source to tackle this issue can be to use a rating system, in which students are able to grade the messages of the rest of students according to different values [9]. Each student can set an evaluation or score for the usefulness of each message: non-relevant, interesting, or totally relevant. Following a similar method for computing reputation from the rating point of view [16], but giving different importance to each type of opinion, it can be calculated the rating information (RI) by taking into account the number of non-relevant opinions (NR), the number of interesting opinions (I), and the number of totally relevant opinions (TR). The relevance of the opinions can be weighted by giving 1 point to *NR*, 2 to *I* and 3 to *RT*. The rating information is calculated as follows:

$$\mathrm{RI} = \frac{\mathrm{NR} + 2\mathrm{I} + 3\mathrm{RT}}{3\mathrm{r}} \tag{3}$$

where *r* is the total number of opinions received by the student.

2.3 SNA as social information

There is a recent line of research on applying social network analysis (SNA) techniques to study the interactions among students in e-learning platforms, for example [17, 18, 19, 20], and it has already been investigated the practicability of SNA in evaluating participation of students [11, 21, 22]. Exploiting SNA techniques it is possible to discover relevant structures in social networks generated from student communications [23]. With visualization of these discovered relevant structures and the automated identification of central and peripheral students, an e-learning system could be provided with better means to assess participation in the online discussions. The practicality of SNA methods in computer supported collaborative learning is demonstrated in [24, 25], using methods for extracting social networks from asynchronous discussion forums, finding appropriate indicators for evaluating participation, and measuring these indicators using social network analysis. A previous work of aDeNu research group [3] suggested the similarity between SNA techniques and the statistical indicators to measure student perceived reputation. As [9] showed, the social network information (SNI) can be calculated as the normalized node in-degree of that student:

$$\mathrm{SNI} = \frac{Z}{p} \tag{4}$$

where Z is the number of in-links and p is the number of students. This research uses Meerkat-ED[1] [26], a specific and practical toolbox for analyzing interactions of students in asynchronous discussion forums of online courses.

3 CLF running on Moodle

The transferable feature of the collaborative indicators emphasized in [3] is demonstrated in this research by deploying the CLF approach on Moodle. Moodle has already been explored as collaborative tool [27, 28], and fits perfectly the purposes of this research. For this, the first step is to see how the CLF functionality can be provided in Moodle. This mapping is compiled in Table 1.

CLF Features	.LRN	MOODLE
Proposing a solution	Survey (for quiz solutions) or file storage area (to upload a solution document) + forum (for discussing the proposed solution)	Q and A forum + assignment + forum in a blog format, to capture students' interactions. Survey for quiz solutions and file storage area are also available.

[1] http://webdocs.cs.ualberta.ca/~rabbanyk/MeerkatED/

CLF Features	.LRN	MOODLE
Management of the CLF stages and timing control	Workflow mechanism	Workflow mechanism
Grouping students	Groups functionality (for manually grouping) and clustering methods provided by Weka data mining suite (for automatic grouping)	Manual groups' functionality, based on reputation. Also an automatic functionality based on the number of groups or number of students per group.
Students' ratings collection	Rating functionality	Rating system based on tailored scales
Reputation estimation	n/a (requires development)	Manual, based on statistical indicators, rating and SNA
Meta-cognitive tools	CLF computed indicators with Weka shown in a customised portlet	Blocks showing information for students. Collaborative information has to be provided manually to be displayed.

Table 1. Comparison between the CLF deployment in dotLRN and Moodle

4 Ongoing work

Previous experiments were carried out by the aDeNu group in 2009, 2012 and 2013, testing the CLF and the collaborative indicators [4]. Now, we are testing the reputation indicator to group participants, looking for an improvement of the collaborative indicators (initiative, activity, regularity).

The research is focused on several aspects, altogether aimed to compute the students' reputation in a domain independent collaborative task called CLF. It is grounded in 1) gathering statistical indicators based on forums interactions, 2) extracting SNA information from the links created among students and 3) considering qualitative data from students' ratings.

An experiment with 23 users was carried out in April with some workers of Tecnalia Research & Innovation[2] centre. They were asked to solve two riddle placed in forums. Previous researches carried out in Madrid Science Week (2009, 2012, 2013) showed the importance of the engagement component in collaborative experiences to get a representative number of participants. For 3 days, the participants collaborated in each stage of the CLF (individual, collaboration, and agreement stage) to find the solution to the first riddle. Next 3 days, they were asked to solve the second riddle. In order to evaluate the benefit of taking into account the reputation indicator in creating

[2] http://www.tecnalia.com/en/

the groups within the CLF, a ‘between- subject’ experiment (i.e., participants were randomly assigned either to the control group, where the CLF grouping was not informed by the reputation indicator and the experimental group, where the CLF grouping considered the reputation indicator by separating the students with higher reputation among the groups, so each group had at least a high reputation participant) was carried out.

Currently, the indicators obtained from the students’ interactions are being analyzed to identify the benefits of taking the reputation indicator into account when making the groups of students. This data analysis can be used to determine required changes in a collaboration process, such as grouping students according to the reputation indicator so as to distribute the students with higher reputation among the groups. This information could also be used by e-learning systems to make tailored recommendations and favouring the engagement, trying to increase the reputation of students less prominent, and improving the collaboration process.

This research also takes the advantage to explore some additional advanced features provided by Moodle, such as learning analytics or the possibility of incorporating meta-cognitive tools [5], automatically calculating the collaborative indicators and displaying the current value of indicators in each stage of the CLF.

Acknowledgement

The authors would like to thank in advance the collaboration of Tecnalia Research & Innovation for the experience that is going to be carried out in Biscay headquarters. They also thank the Spanish Ministry of Economy and Competence for funding the MAMIPEC project (TIN2011-29221-C03-01) in which this research work is framed.

References

1. Kumar, V. S.: Computer-supported collaborative learning: issues for research. In Eighth Annual Graduate Symposium on Computer Science, University of Saskatchewan (1996)
2. Barkley, E. F., Cross, K. P., Major, C. H.: Collaborative learning techniques: A handbook for college faculty. John Wiley and Sons (2004)
3. Anaya, A. R., Boticario, J. G.: A domain-independent, transferable and timely analysis approach to assess student collaboration. International Journal on Artificial Intelligence Tools, 22(04) (2013)
4. Santos, O. C., Boticario, J. G.: Involving Users to Improve the Collaborative Logical Framework. The Scientific World Journal (2014)
5. Anaya, A., Boticario, J.: Content-free collaborative learning modeling using data mining. User Modeling and User-Adapted Interaction, 21(1), 181–216 (2011)
6. Bayón, A., Santos, O. C., Boticario, J. G.: A component to carry out the Logical Framework Approach in dotLRN. In dotLRN. 7th OpenACS/. LRN conference (2008)
7. Santos, O. C., Rodríguez, A., Gaudioso, E., Boticario, J. G. (2003, July). Helping the tutor to manage a collaborative task in a web-based learning environment. In AIED2003 Supplementary Proceedings (Vol. 4, pp. 153-162).

8. Russo, T. C., Koesten, J.: Prestige, centrality, and learning: A social network analysis of an online class. Communication Education, 54(3), 254-261 (2005)
9. Romero, C., López, M.-I., Luna, J.-M., Ventura, S.: Predicting students' final performance from participation in on-line discussion forums. Computers & Education, 68, 458–472 (2013)
10. Memon, N., Xu, J. J., Hicks, D. L., Chen, H.: Data Mining for Social Network Data. Information Systems Journal, 12, 1–7 (2010)
11. Daradoumis, T., Martínez-Monés, A., Xhafa, F.: A layered framework for evaluating online collaborative learning interactions. International Journal of Human-Computer Studies, 64(7), 622-635 (2006)
12. Alfonseca, E., Carro, R. M., Martín, E., Ortigosa, A., Paredes, P.: The impact of learning styles on student grouping for collaborative learning: a case study. User Modeling and User-Adapted Interaction, 16(3-4), 377-401 (2006)
13. Paredes, P., Rodriguez, P.: The application of learning styles in both individual and collaborative learning. In Advanced Learning Technologies. Sixth International Conference on, pp. 1141-1142. IEEE (2006)
14. Sterbini, A., Temperini, M.: Social exchange and collaboration in a reputation-based educational system. In Information Technology Based Higher Education and Training (ITHET), 9th International Conference on (pp. 201-207). IEEE (2010)
15. Clow, D., Makriyannis, E.: iSpot analysed: participatory learning and reputation. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 34-43). ACM (2011, February)
16. Tavakolifard, M., Almeroth, K. C.: Social computing: an intersection of recommender systems, trust/reputation systems, and social networks. Network, IEEE, 26(4), 53-58 (2012)
17. Durugbo, C., Hutabarat, W., Tiwari, A., Alcock, J. R.: Modelling collaboration using complex networks. Information Sciences, 181(15), 3143–3161 (2011)
18. Laat, M., Lally, V., Lipponen, L., Simons, R.-J.: Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. International Journal of Computer-Supported Collaborative Learning, 2(1), 87–103 (2007)
19. Martinez, A., Dimitriadis, Y., Tardajos, J., Velloso, O., Villacorta, M.: Integration of SNA in a mixed evaluation approach for the study of participatory aspects of collaboration. In W5 Moving From Analysis to Design: Social Networks in the CSCW Context), ECSCW'03 Workshop on Social Networks, pp. 14-18. (2003)
20. Yusof, N., Rahman, A. A.: Students' interactions in online asynchronous discussion forum: A Social Network Analysis. In Education Technology and Computer. ICETC'09. International Conference on (pp. 25-29). IEEE (2009)
21. Rabbany, R., Takaffoli, M., Za, O. R.: Social Network Analysis and Mining to Support the Assessment of On-line Student Participation. ACM SIGKDD Explorations Newsletter, 13(2), pp. 20-29 (2012)
22. Cuero, F. V., Collazos, C. A., Zea, N. P., Romo, J. O. Análisis y monitoreo de la interacción en entornos colaborativos, mediante el uso de SNA.
23. Rabbany, R., Takaffoli, M., Za, O. R.: Analyzing Participation of Students in Online Courses Using Social Network Analysis Techniques. In Proceedings of educational data mining (2009)
24. Nurmela, K., Lehtinen, E., Palonen, T.: Evaluating CSCL log files by social network analysis. In Proceedings of the 1999 conference on Computer support for collaborative learning (p. 54). International Society of the Learning Sciences (1999)

25. Cho, H., Gay, G., Davidson, B., Ingraffea, A.: Social networks, communication styles, and learning performance in a CSCL community. Computers and Education, 49(2), 309-329 (2007)
26. Chen, J., Fagnan, J., Goebel, R., Rabbany, R., Sangi, F., Takaffoli, M., Zaiane, O.: Meerkat: Community Mining with Dynamic Social Networks. IEEE International Conference on Data Mining Workshops, (1), 1377–1380 (2010)
27. Dougiamas, M., Taylor, P.: Moodle: Using learning communities to create an open source course management system. In World conference on educational multimedia, hypermedia and telecommunications, vol. 2003, No. 1, pp. 171-178 (2003)
28. Cavus, N., Uzunboylu, H., Ibrahim, D.: Student opinion towards using an open source learning management system together with a collaborative tool. Cypriot Journal of Educational Sciences, 3(2), 120-129 (2009)

Evaluation of a Personalized Method for Proactive Mind Wandering Reduction

Robert Bixler[1], Kristopher Kopp[2], and Sidney D'Mello[1,2]

[1]Deparment of Computer Science and [2]Department of Psychology, University of Notre Dame
{rbixler, kkopp, sdmello}@nd.edu

Abstract. We report on a project with the goal of creating a proactive system that attempts to reduce the propensity to mind wander (MW) by optimizing learning conditions (e.g., text difficulty and value) for individual learners. Our previous work had shown that supervised classification based on individual attributes could be used to detect the learning condition with the lowest MW rates. Here we test the model by comparing MW rates for the predicted optimal conditions to MW rates from a random control condition or in the condition with the overall best MW rate across all learners. Our results suggest that our method is better than these non-adaptive alternatives in certain contexts.

Keywords: engagement, mind wandering, affect, machine learning

1 Introduction

Learner models are at the core of intelligent tutoring systems (ITS). The development of ITSs has been influenced by cognitive learner models [1,2], and in recent years there has been a rise in ITSs that have been informed by affective models [3,4,5]. The cognitive-affective state of engagement is of particular interest for this project. Engagement has been defined as an enjoyable state of involvement in a learning activity or task with focused attention and intense concentration [3]. Engagement is necessary for learning since learners have to attend to information in order to learn. *Mind wandering* (MW) pertains to instances where engagement is disrupted and learners involuntarily shift their attention from their task towards unrelated thoughts. MW can be detrimental to learning [6, 7] because of this lapse in attention. Thus, to facilitate learning, it is important to develop systems that can either sustain engagement by reducing the propensity of MW behaviors or respond when a learner becomes disengaged due to MW. Not all learners exhibit the same MW behaviors when placed in the same learning environments [8]. Some learners experience lower MW rates compared to others depending on the context of the learning activity. For example, in a situation where the learning materials are considered difficult some individuals may be able to sustain attention and remain engaged, while others may disengage as their attention drifts towards thoughts unrelated to the task. With this in mind, we have begun developing a method that adapts the learning environment according to measures of individual attributes in an effort to reduce MW behaviors during a learning session. Our intention is to select optimal learning materials based on these measures, with the purpose of reducing the propensity to MW. For example, learners would be assessed for attributes such as reading comprehension or scholastic aptitude and would then be placed in a learning environment and provided with materials based on those attributes with the goal of reducing the propensity to MW. The goal of this paper is to evaluate the performance of such a system by comparing our method to two non-adaptive alternatives.

1.1 Related Work

A variety of learner models have been employed in ITSs since their inception. Examples of cognitive models include: knowledge tracing models [9, 10], item response theory models [11], and knowledge space models [12]. Recent research of alternatives to cognitive models includes affective models [13, 14], meta-cognitive models [15], and models of disengagement [16] (see [17] for a review of recent models). Advancing the groundwork laid by studies that have investigated the relationship between affect and learning [3], [see 18 for a review, 19], recent research along these lines has led to the development of **Reactive** affect-sensitive ITSs that attempt to sense affective states related to learning and respond accordingly [20, 21, 22]. One example of this type of system is Affective AutoTutor [23]. This system detects the cognitive-affective states of the learner (i.e., boredom, confusion) based on conversational modeling, facial cues, and body language and alters the dynamics of the tutoring session through dialog moves designed to address specific affective states.

Although there are no analogous reactive systems that respond to MW, there have been attempts to develop automatic MW detectors. Drummond and Litman [24] used acoustic-prosodic features extracted from learners' utterances during a spoken learning task to discriminate episodes of low "zoning out" from episodes of high "zoning out", obtaining an accuracy of 64%. With a similar goal in mind, Bixler and D'Mello [25] recently attempted to automatically detect MW during reading on a computer screen using eye movements. They were able to detect MW with an accuracy of 72% (expected = 61%). A similar system, called GazeTutor [4], used an eye tracker to detect when users looked away from the screen for an extended period of time, which was taken to imply attentional disengagement. Although GazeTutor didn't definitively detect instances of MW, it attempted to re-engage learners with interventions when attentional disengagement was detected. Thus, research is steadily moving towards systems that are able to identify and respond appropriately to MW with the goal of sustaining engagement and improve learning.

Conversely, **Proactive** strategies attempt to facilitate affective states that would be beneficial for learning or avoid states that would be detrimental for learning. One example of a system that used a proactive strategy is ConfusionTutor [26], which attempted to induce a state of confusion during learning as there has been evidence that suggests a positive correlation between learning gains and confusion [27].

1.2 The Current Project

We recently took a step towards developing a proactive strategy to reduce MW by selecting learning materials that lead to reduced MW rates for individual learners [8].

MW rates were estimated with learner responses to auditory probes while learners read instructional texts on a computer screen. Each text was either an easy or difficult version and was manipulated to have either low or high value with respect to its weight on a subsequent test. Each learner read a total of four texts: one of each combination of difficulty and value. Supervised learning methods were used to build models that used individual attributes to predict the texts that would result in the lowest MW rate for that learner. Each model was built on data from the other learners (i.e., N – 1) and was then applied to the learner that was held out. The best models were moderately successful, resulting in an accuracy of 64% (expected = 53%). The next step, and the focus of our current research, is to further investigate how effective our method is at personalizing the learning environment in order to reduce MW.

There are many ways to evaluate the effectiveness of a personalized system. Several empirical evaluation methods are mentioned by Chin [28], such as experimental comparisons between systems with and without learner models or evaluating the accuracy of each learner model. Gena [29] covers strategies for evaluating user-adaptive systems, which includes additional strategies such as user-centered evaluation through questionnaires and interviews, observational evaluation through user observation and log files, and predictive evaluations such as expert reviews. Similar evaluation methods are suggested specifically for ITSs by Mark and Greer [30]. Due to the early nature of this project, we opted for a preliminary analysis that takes advantage of existing data in lieu of a more time consuming experimental study.

The present work describes an investigation of the effectiveness of our method to prevent MW [8]. We used existing data which identified the MW rate of each learner for four different learning materials that varied in difficulty and value. To evaluate our method, we then selected a MW rate for each learner based on the model's prediction of the learning materials with the lowest MW rate (i.e., individual best). We then compared these to MW rates derived from two non-adaptive alternative methods. The first alternative was to determine the learning materials with the lowest MW rate on average across all learners and select those learning materials for each learner (i.e., overall best). The second alternative was to simply select learning materials for each learner at random (i.e., random).

2 Data and Methods

What follows is a description of the data collection and analyses for the current project. For a more detailed description of data collection, see [8].

2.1 Data Collection

Undergraduate students (N = 187) from two U.S. universities learned about research methods topics from four texts (i.e., experimenter bias, replication, causality, and dependent variables) presented on a computer screen. Each text contained 1500 words on average (SD = 10) and were split into 30-36 pages. The difficulty and value of each text was manipulated. The difficulty manipulation consisted of presenting either an easy or a difficult version of each text. Value was manipulated based on the weight assigned to each text on a subsequent posttest. Learners read all four texts with one text in each one of the four conditions: 2 (difficulty: easy vs. difficult) × 2 (value: high vs. low). The success of the manipulations was confirmed with self-reports of the perceived difficulty and perceived value of the texts (see [31]). During the task, learners' MW was measured along with several individual attributes.

Mind Wandering was measured through auditory probes (i.e. a beep) on nine pseudorandomly chosen "probe pages" per text, a standard and validated method for collecting online MW reports [6]. The MW rate for each text was then obtained by computing the proportion of "Yes" responses to probes.

Individual Attribute measures were collected for use as features in our models. The following measures were collected: (a) reading comprehension, (b) reading fluency, (c) working memory ability, (d) interest in research methods, (e) general boredom proneness, (f, g) boredom in academic situations (underwhelmed and overwhelmed), (h) scholastic aptitude, and (i) prior knowledge. Scores of all measures were standardized by school to alleviate any large discrepancies due to demographic differences between schools.

Procedure. Learners began the task by proceeding through one of two 24 item multiple choice pretests (counterbalanced between pre and posttest across all learners) and several individual attribute measurements. After being given instructions on the learning task, they studied four texts (one at a time) on a page-by-page basis, using the space bar to navigate forward. The title of the text and the corresponding weight of the test questions (value manipulation) were explicitly presented before each text. After learners studied all four texts, they were presented with the remaining 24 item posttest and remaining individual attribute measures.

2.2 Supervised Machine Learning

We used measurements of the individual attributes to predict the learning materials (in terms of difficulty and value) that would result in the least amount of mind wandering using supervised learning. Models were built for 34 machine learning algorithms from the WEKA machine learning software [32]. These included lazy-learners, Bayesian models, decision trees, support vector machines, regression models, etc. There were two additional parameters. The first parameter was the minimum allowable difference (i.e., threshold) between a learner's standardized MW rate for the best and worst materials (i.e., a difference of .0, .25, or .5 standard deviations between the highest and lowest MW rates). The second parameter was the specific classification task. The task was to classify the optimal learning materials between low and high difficulty texts, low and high value conditions, or any of the 4 conditions. Leave-one-person-out cross validation was used to evaluate each data set. Models were built on all learners except for a hold out learner and then tested on the hold out learner; this process was repeated for all learners. This method ensures that the training and testing set for each model are learner-independent. The Kappa statistic was taken as the measure of classifier accuracy. A kappa value of 1 indicates perfect agreement, while a kappa value of 0 indicates agreement was no better than chance.

2.3 Comparison Analysis

The best performing models for each classification task were identified based on the highest kappa. The best model for both the difficulty and difficulty/value classification tasks was built with a decision stump classifier, while the best model for value was built with a simple logistic classifier. These models were then used to assess how our method of assigning materials to learners would perform compared to non-adaptive methods. To illustrate how each MW rate was computed for the comparison, consider a hypothetical situation with 4 learners being compared in the difficulty classification task (Table 1). Individual best MW rates are based on model predictions; in this example, the model predicted that the best materials would be the difficult texts for learners 2 and 3, and the easy texts for learners 1 and 4 (note that the model erred for learners 1 and 2). Overall best MW rates are the MW rates for each learner with the materials that resulted in the lowest MW rate on average across participants; these are the easy texts for this example. Random MW rates are the MW rates for each learner with materials chosen at random; in this example, learner 2 is randomly assigned difficult texts, while learners 1, 3, and 4 are randomly assigned easy texts. Note that in this case, both the overall best and individual best conditions predicted the materials with the lowest MW rate for half the learners, which resulted in comparable average MW rates of about 0.45.

Table 1. MW rates (proportions of yes to total probe responses) for 4 hypothetical learners by classification (easy and difficult) and comparison groups.

Learner	Easy	Difficult	Individual Best	Overall Best	Random
1	0.61	0.39	0.61 – Easy	0.61 – Easy	0.61 – Easy
2	0.28	0.56	0.56 – Difficult	0.28 – Easy	0.56 – Difficult
3	0.67	0.44	0.44 – Difficult	0.67 – Easy	0.67 – Easy
4	0.22	0.61	0.22 – Easy	0.22 – Easy	0.22 – Easy
Average	**0.44**	**0.50**	**0.46**	**0.45**	**0.52**

3 Results

Table 2 lists the average standardized MW rates for each of these conditions based on the complete data set. Our initial step was to assess the accuracies of the classification results when considering all four types of learning materials: difficulty (easy and difficult) × value (low and high). We compared the MW rates of the best performing model (i.e., at the threshold of .25 *sd*'s) which resulted in a kappa of .11 (observed accuracy of 34%, expected accuracy of 26%). The MW rates were significantly lower for the individual best condition compared to the random condition, $t(140) = -2.1$, $p = .04$, but not significantly different from the overall best condition.

We then collapsed across value and then difficulty and conducted similar analyses for each. Value, at the threshold of .25 *sd*'s, resulted in a kappa of .16 (observed accuracy of 59%, expected accuracy of 51%). The MW rates in the individual best condition were not significantly different from either the random or the overall best condition. Difficulty at the threshold of .5 *sd*'s, resulted in a kappa of .24 (observed accuracy of 64%, expected accuracy of 53%). The MW rates in the individual best condition were significantly different from the random condition, $t(97) = -2.4$, $p = .02$, but not different from the overall best condition .

Table 2. Standardized MW rate means by classification task (standard deviations in parentheses). Lower numbers are preferred.

Classification Task	Individual Best	Overall Best	Random	*N*
Difficulty × Value	-.01 (.87)	-.03 (.87)	.13 (.93)	141
Value	-.05 (.79)	-.01 (.80)	-.01 (.81)	141
Difficulty	.07 (.72)	.09 (.75)	.17 (.75)	98

These preliminary results show that the models built on a small suite of individual attributes chose learning materials for each learner that were optimal in terms of resulting in the least amount of MW when compared to placing learners into a random learning condition except when collapsing across value. However, we were unable to choose materials with reported instances of MW that were statistically less than those chosen in the overall best condition across all learners.

We next wanted to take a close look at those individuals whose best model condition was different than the overall best condition to gain further insight into how the mind wandering behaviors differ between these conditions (see Table 3). The analyses described above were repeated after removing learners with the same individual best and overall best condition. For example, if the model predicted a learner should be given low difficulty materials, which is the overall best condition, then that leaner would not be included in the following analysis. For each analysis, the sample size was considerably culled resulting in low power, however, the results of significance are still reported. When considering all four types of learning materials (i.e., difficulty × value value) at the threshold of .25 *sd's*, the MW rates for the individual best condition were higher than the rates for the overall best condition, $t(40) = .799$, $p = .43$. When considering only value at the threshold of .25 *sd's*, the rates for the individual best condition were lower than the overall best condition, $t(51) = -1.5$, $p = .13$. When considering only difficulty at the threshold of .5 *sd's*, the rates for the individual best condition were lower than the overall best condition, $t(18) = -91$, $p = .38$.

Table 3. Standardized MW rate means by classification task for learners that differed on MW rates for the individual best and overall best conditions (standard deviations in parentheses)

Classification Task	Individual Best	Overall Best	Random	*N*
Difficulty × Value	.20 (.78)	.09 (.82)	.28 (.92)	41
Value	.07 (.81)	.17 (.80)	.12 (.82)	52
Difficulty	.14 (.84)	.24 (.97)	.17 (.90)	19

This second analysis shows that when the individual best condition differs from the overall best condition of all learners, there are some drastic differences in the amount of MW rates. When collapsing on value or difficulty (separately), the individual best condition outperforms the overall best condition. However, when considering the difficulty × value classifications, this trend is reversed where the overall best produces the least amounts of mind wandering.

4 Discussion

The goal of this project is to take strides towards creating a personalized learning environment in which a learner is provided with materials that reduce the propensity to MW. While there have been a few encouraging projects that attempt to take such proactive steps toward enhancing the learning experience by adapting the learning environment [see 33 for a review], this project's focus on attempting to proactively sustain engagement by reducing the likelihood that learners would MW based on a rather small number of individual attribute measures is novel. We showed that our method performs either better than or at least as well as two non-adaptive alternatives for choosing learning materials that will lead to a reduced MW rate. This is an initial step towards developing a system sensitive to learners' needs in terms of sustaining engagement. The next step would be to implement an experiment to test the generalizability of the claim that the method described here is, in fact, an effective method to incorporate into a preventative learning environment. Another possibility is to assess an expanded set of individual attribute measures. An exploration of additional measures could determine a specific set of features that are best able to predict a condition with an optimal MW rate.

Two limitations are apparent. First, it is possible that learners reported MW rates incorrectly, which could decrease the accuracy of our method. However, learner self-reports are used extensively in previous studies as there is not currently a good alternative for tracking MW [6]. Second, these findings are based on learners reading texts on research methods in a laboratory setting. Future work could boost claims of generalizability by incorporating different topics and other modes of information delivery.

This research takes a step towards tailoring a learning environment in order to reduce the rate of MW and potentially increase engagement. Systems exist that are sensitive to various states of the learner and take a reactive approach by adapting to the needs of the learner in a variety of contexts [21, 22, 23]. This project takes a proactive approach to addressing the needs of the learner by assessing their attributes and identifying learning materials that would potentially produce the least amount of MW. This method need not be limited to addressing MW behaviors during a leaning session. It would be beneficial for future work to assess how this method could be applied to addressing other cognitive affective states, such as boredom or confusion, which also have an influence on learning.

Acknowledgment. This research was supported by the National Science Foundation (NSF) (ITR 0325428, HCC 0834847, DRL 1235958). Any opinions, findings and

conclusions, or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF.

References

1. Graesser, A. C., Olney, A., Haynes, B. C., & Chipman, P. (2005). AutoTutor: a cognitive system that simulates a tutor that facilitates learning through mixed-initiative dialogue. In C. Forsythe, M. L. Bernard, & T. E. Goldsmith (Eds.), Cognitive systems: Human cognitive models in systems design. Mahwah, NJ: Erlbaum.
2. VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rosé, C. P. (2007). When are tutorial dialogues more effective than reading?. *Cognitive Science*, *31*(1), 3-62.
3. Baker, R., D'Mello, S., Rodrigo, M., & Graesser, A. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies, 68 (4), 223-241.
4. D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies, 70*(5), 377-398.
5. Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology, 4*(3/4), 129-163.
6. Mooneyham, B. W., & Schooler, J. W. (2013). The costs and benefits of mind-wandering: A review. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *67*(1), 11.
7. Szpunar, K. K., Moulton, S. T., & Schacter, D. L. (2013). Mind Wandering and education: from the classroom to online learning. *Frontiers in psychology*, *4*.
8. Kopp, K., Bixler, R., D'Mello, S. K. (In Press). Identifying Learning Conditions that Minimize Mind Wandering by Modeling Individual Attributes. Lecture Notes in Computer Science 8474, Springer-Verlag, pp. 94-103
9. Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)*, *8*, 30-43.
10. Mitrovic, A. (2012). Fifteen years of constraint-based tutors: what we have achieved and where we are going. *User Modeling and User-Adapted Interaction*, *22*(1-2), 39-72
11. Millán, E., & Pérez-De-La-Cruz, J. L. (2002). A Bayesian diagnostic algorithm for student modeling and its evaluation. User Modeling and User-Adapted Interaction, 12(2-3), 281-330.
12. Falmagne, J. C., Cosyn, E., Doignon, J. P., & Thiéry, N. (2006). The assessment of knowledge, in theory and in practice. In *Formal concept analysis* (pp. 61-79). Springer Berlin Heidelberg.
13. Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267-303.
14. D'Mello, S., & Graesser, A. (2012). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Transactions on Interactive Intelligent Systems (TiiS), 2(4), 23.
15. Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence in Education, 16(2), 101-128.

16. Baker, R. S., Mitrović, A., & Mathews, M. (2010). Detecting gaming the system in constraint-based tutors. In User Modeling, Adaptation, and Personalization (pp. 267-278). Springer Berlin Heidelberg.
17. Desmarais, M. C., & d Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. User Modeling and User-Adapted Interaction, 22(1-2), 9-38.
18. D'Mello, S. K. (2013). A Selective Meta-analysis on the Relative Incidence of Discrete Affective States during Learning with Technology, Journal of Educational Psychology.
19. Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004). Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, *29*(3), 241-250.
20. Calvo, R. A., & D'Mello, S. K. (2012). Frontiers of affect-aware learning technologies, IEEE Intelligent Systems, *27(6)*, 86-89
21. Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., ... & Rossi, L. (2012, June). Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. In *Proceedings of the 5th International Conference on Educational Data Mining* (pp. 126-133).
22. Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, *19*(3), 267-303.
23. D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., ... & Graesser, A. (2008, June). AutoTutor detects and responds to learners affective and cognitive states. In *Workshop on Emotional and Cognitive Issues at the International Conference on Intelligent Tutoring Systems*.
24. Drummond, J. & Litman, D. (2010). "In the Zone: Towards Detecting Student Zoning Out Using Supervised Machine Learning," Intelligent Tutoring Systems, Part Ii, Lecture Notes in Computer Science 6095, V. Aleven, et al., eds., Springer-Verlag, pp. 306-308
25. Bixler, R., & D'Mello, S. K. (in press). Toward Fully Automated Person-Independent Detection of Mind Wandering. In User Modeling, Adaptation, and Personalization. Springer Berlin Heidelberg.
26. D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170.
27. D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion*, *25*(7), 1299-1308.
28. Chin, D. N. (2001). Empirical evaluation of user models and user-adapted systems. *User modeling and user-adapted interaction*, *11*(1-2), 181-194.
29. Gena, C. (2005). Methods and techniques for the evaluation of user-adaptive systems. The knowledge engineering review, 20(01), 1-37.
30. Mark, M. A., & Greer, J. E. (1993). Evaluation methodologies for intelligent tutoring systems. *Journal of Artificial Intelligence in Education*, *4*, 129-129.
31. Mills, C., & D'Mello, S. K. (in prep). How Do Extrinsic Value and Difficulty Impact Engagement: An Experimental Approach.
32. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*. 11, 1, 10–18.
33. D'Mello, S. K. & Graesser, A. C. (in press). Feeling, Thinking, and Computing with Affect-Aware Learning Technologies. In Calvo, R. A., D'Mello, S. K., Gratch, J., & Kappas, A. (Eds.) Handbook of Affective Computing. Oxford University Press.

Providing Personalized Guidance in Arithmetic Problem Solving

Miguel Arevalillo-Herráez[1], David Arnau[2], Luis Marco-Giménez[1],
José A. González-Calero[3], Salvador Moreno-Picot[1], Paloma Moreno-Clari[1],
Aladdin Ayesh[4], Olga C. Santos[5], Jesús G. Boticario[5], Mar Saneiro[5],
Sergio Salmeron-Majadas[5], Raúl Cabestrero[6], and Pilar Quirós[6]

[1] Department of Computer Science, University of Valencia, Spain
{miguel.arevalillo,salvador.moreno,paloma.moreno}@uv.es
margi4@alumni.uv.es

[2] Department of Didactics of Mathematics, University of Valencia, Spain
david.arnau@uv.es

[3] Department of Mathematics, University of Castilla-La Mancha, Spain
jose.gonzalezcalero@uclm.es

[4] Faculty of Technology, De Montfort University, UK
aayesh@dmu.ac.uk

[5] aDeNu Research Group. Artificial Intelligence Dept. Computer Science School. UNED, Spain
{ocsantos,jgb,marsaneiro}@dia.uned.es,ssalmeron@bec.uned.es

[6] Department of Basic Psychology, Faculty of Psychology, UNED, Spain
{rcabestrero,pquiros}@psi.uned.es

Abstract. Supervising a student's resolution of an arithmetic word problem is a cumbersome task. Different students may use different lines of reasoning to reach the final solution, and the assistance provided should be consistent with the resolution path that the student has in mind. In addition, further learning gains can be achieved if the previous student's background is also considered in the process. In this paper, we outline a relatively simple method to adapt the hints given by an Intelligent Tutoring System to the line of reasoning that the student is currently following. We also outline possible extensions to build a model of the student's most relevant skills, by tracking user's actions.

Keywords: Personalization, adaptation, Intelligent Tutoring System, Word Problem Solving, Arithmetic teaching

1 Introduction

Developing the students' problem solving skills is a fundamental part of mathematics learning. Word/story problems are commonly used in this context, as a means to promote the student's engagement and provide an adequate framework to practice mathematics skills. The importance of arithmetic word problem is supported by the development of several computer systems focused on this activity, such as HERON [13], Story Problem Solver [12], WORDMATH [10], MathCAL [6] or AnimalWatch [5].

Successful problem solvers construct a model of the situation described in the problem statement, and base their solution plan on this model [8, 14]. We can think of this model as a number of relations between the quantities that explicitly or implicitly appear in the problem statement. When a problem resolution is supervised by a human in a one-to-one situation, direct observation allows the tutor to induce the model that the student has in mind. This allows the tutor to provide contextualized help that is consistent with the student's previous resolution steps. The tutor is also constantly collecting information about the student. This information is generally used to adapt explanations to the student's specific characteristics.

Intelligent Tutoring Systems (ITS) aimed at developing word problem solving skills need also provide personalized guidance. In most cases, the situation described by the problem statement may be modeled in several ways. In this case, the ITS should be able to evaluate the previous user interaction to determine the solution scheme that the student is using, and provide adequate guidance in accordance to this scheme. In this paper, we use a sample problem to illustrate a relatively simple strategy to infer the solution scheme that the student is following.

2 Solution Schemes

When a student reads a problem statement, he/she generally builds a mental scheme of the problem solution. This solution scheme generally includes the student's interpretation of the quantities involved, and a set of relations between these quantities. Lets consider the following problem statement: "A basket contains 60 pieces of fruit, between apples and pears. It has 10 more apples than pears. How may apples are there in the basket?". One possible mental solution scheme (S_1) would be to divide the 60 pieces into two groups of fruits, both with the same number of elements (30); and then mentally transfer half the difference (5) from one group to the other. Another different mental solution scheme (S_2) would consist in mentally setting the 10 extra apples apart; then dividing the remaining pieces of fruit into the two groups; and finally adding the 10 extra apples which were taken apart. Other solution schemes may consider computing the number of apples after computing the number of pears.

Lets suppose that the student has started the resolution by doing the operation $10/2 = 5$, but is finding problems to propose the next operation. If there is a system intervention, it would make little sense that the system suggests the student uses the expression $60 - 10 = 50$. Such a recommendation would very likely cause confusion on the learner. This is because this action belongs to a line of reasoning that is not the one that the student was following. On the contrary, the suggestion $60/2 = 30$ would make more sense, and be in line with the student's solution scheme.

Potentially valid solution schemes can easily be internally represented as a set of quantities and relations between quantities [7]. Table 1 shows such a representation for the solution scheme S_1 above.

Representation	Description	Initial value
Total	Total number of pieces	60
Half_Total	Half the number of pieces	unknown
Excess	Extra number of apples	10
Half_Excess	Half the extra number of apples	unknown
Apples	Number of apples	unknown

Relations
Half_Total=Total/2
Half_Excess=Excess/2
Apples=Half_Total+Half_Excess

Table 1. Quantities (left) and relations (right) in solution scheme S1. Only values for the quantities $Total$ and $Excess$ are known.

This way of representing solution schemes allows any automated system to determine all valid expressions that the student may use. For explanation purposes, lets define an active relation as one that contains a single unknown quantity. In addition, lets define a way to generate a expression from any active relation. This consists in replacing all known quantities by their respective numeric values, and the unknown quantity by the number which makes the resulting expression numerically correct. With these definitions, valid expressions correspond to the ones generated by all active relations, in any solution scheme (and rearranged versions of them).

For example, S_1 has two active relations (the first two relations in Table 1). These generate the expressions $30 = 60/2$ and $5 = 10/2$, respectively. Hence, a student may start solving the problem at hand according to S_1 by using these two expressions (or a rearranged version of them). The use of a different expression may imply a mistake or that the learner is following a different solution scheme.

3 Adaptive Help

Tracking the state of each solution scheme is a key issue to provide adequate help messages that are consistent with the student's current line of reasoning. To this end, every valid learner's input is simultaneously processed in the context of each solution scheme. This is done by comparing the user's input to the expressions generated by the active relations. If an equivalent expression is found, the value of the corresponding unknown quantity is updated, and the relation is removed from the relations table. For example, the expression $10/2 = 5$ as a first user input would match the relation Half_Excess=Excess/2 in S_1, and yield the dynamic state in Table 2. With this method, unknown quantities are solved one at a time. Hence, the number of remaining expressions in each mental solution scheme is always the same as the number of unknown quantities in the scheme.

This simple tracking mechanism allows one to associate the progress of a mental solution scheme with the percentage of relations that have already been used. This simple measures allows an automated system to easily compute the

Representation	Description	value
Total	Total number of pieces	60
Half_Total	Half the number of pieces	unknown
Excess	Extra number of apples	10
Half_Excess	Half the extra number of apples	5
Apples	Number of apples	unknown

Unused Relations
Half_Total=Total/2
Apples=Half_Total+Half_Excess

Table 2. Representation of the dynamic state of S_1 after processing expression $10/2 = 5$. The quantity $Half\_Excess$ has become known, and the relation has disappeared from the corresponding table.

solution scheme that the learner is most likely following, and provide consistent feedback based on his information. For example, lets suppose that a student has already introduced the expression $10/2 = 5$ and asks for a suggestion. S_1 would be pointed out as the most likely solution scheme. Since the only active relation in this scheme is Half_Total=Total/2 (see Table 2), an automated system would be able to recommend the use of the expression $30 = 60/2$. The way this action is translated into a help message may depend on the specific system and/or other information registered as part of the student model. Foe example, the system may simply recommend the user to compute half the number of pieces by using an appropriate expression, and refine the message further if the user is still not able to set it correctly.

To test the effectiveness of this strategy to determine the mental scheme that the student is following, the method has been integrated into an existing ITS [1, 4]. The results obtained have been reported in [3]. To support the relevance of the research, a first study showed that there are significant differences between the help provided by expert and novice teachers. In many cases, novice teachers gave hints which were not consistent with the observable reasoning, according to the previous learner's calculations. In addition, it is shown that the aids provided by the ITS were similar to the ones offered by experts teachers (in 84% of the cases).

4 Further adaptation

Apart from considering the student's current line of reasoning, it is also possible to build a student model out of his/previous interaction with the system. This model can be used to further adapt help messages to the user's needs. For example, a particular learner may have difficulties at using multiplicative relations, and benefit from additional explanations. We are currently working on the definition of an ontology that allows the system to keep track of the most relevant skills in arithmetic problems solving.

A first attempt in this direction was made in [1], in the context of algebra learning. A labeling scheme for relations allowed the ITS to estimate the learner's skills at detecting and expressing certain type of conceptual schemes. Following with this idea, we are working on the definition of an appropriate labeling for an arithmetic context. The intention is that correct inputs, mistakes and help requests can be linked to concrete skills and tracked by an automated system.

In addition, we are currently considering ways of detecting and using the student's affective state to improve learning. An initial discussion was provided in [2]. As a first experiment in this direction, we have prepared a series of exercises that students will need to solve using the ITS. Some of these exercises seek to elicit concrete emotions. For example, a student may get confused if the ITS repeatedly provides hints based on a solution scheme that he/she is not following; or frustrated if right answers are reported as incorrect and suggestions to use relations in non-natural solution schemes are issued. To capture emotional data of interest, we have prepared a modified version of the existing ITS. This new version uses self-reporting at several stages. Before the student starts solving any exercise, he/she has to fill the Attributional Achievement Motivation Scale presented in [11]. This is a self reporting test based on Weiners attributional theory [15], which is used to explain the attributional causes of the academic achievement on a given subject (arithmetic problem solving in our case). The test is composed of 22 items structured in 5 factors, namely interest, task or capacity, effort, exams and the teachers pedagogical capacity. After completing each exercise, the student has to report about his/her affective state (valence and activation). To this end, we have used Self-Assessment Manikins (SAM) [9]. At the end of the series, the student is asked to fill a self-report. Finally, we have included a descriptive self-report that the student has to fill once the entire series of exercises has been completed. Results from this research will be used to built an ITS that provides emotional support, and measure the performance improvement obtained with respect to the original ITS. A first proposal consists in replacing the current help on demand mechanism by a rule-based system that is able to use interaction data to both provide automatic recommendations and adapt the content of the messages, according to the user's affective state.

5 Conclusions

Teaching arithmetic word problem solving is a complex task. Significant differences in tutoring between expert and non-expert teachers have been identified and reported in [3]. One major factor behind these differences is the ability of the teacher to provide feedback that is consistent with the current student reasoning. In this paper, we have described an strategy that makes it possible to transfer this fundamental skill to an automated system. It could be claimed that the system would not be able to handle solution schemes that the system is not aware of. However, this is also the case in human supervision. A human may interpret as incorrect any action that does not match a valid step in the solution schemes that he/she is able to generate.

We have also outlined future improvements aimed at providing a closer behavior to a human expert, by considering both previous interactions and the learner's affective state. We have also described the design of a new experiment to help the integration of affective support into the ITS.

Acknowledgments. This work has been supported by the Spanish Ministry of Economy and Competitiveness through projects TIN2011-29221-C03-02 and TIN2011-29221-C03-01.

References

1. Arevalillo-Herráez, M., Arnau, D., Marco-Giménez, L.: Domain-specific knowledge Representation and Inference Engine for an Intelligent Tutoring System. Knowledge-Based Systems 49, 97–105 (2013)
2. Arevalillo-Herráez, M., Moreno-Picot, S., Arnau, D., Moreno-Clari, P., Boticario, J., Santos, O.C., Cabestrero, R., Quirós, P., Salmeron-Majadas, S., Riesco, A.M., Saneiro, M.: Towards enriching an its with affective support. In: Berkovsky, S., Herder, E., Lops, P., Santos, O.C. (eds.) UMAP Workshops. CEUR Workshop Proceedings, vol. 997. CEUR-WS.org (2013)
3. Arnau, D., Arevalillo-Herráez, M., Gonzalez-Calero, J.: Emulating human supervision in an intelligent tutoring system for arithmetical problem solving. IEEE Transactions on Learning Technologies To be published (April 2014)
4. Arnau, D., Arevalillo-Herráez, M., Puig, L., González-Calero, J.A.: Fundamentals of the Design and the Operation of an Intelligent Tutoring System for the Learning of the Arithmetical and Algebraic Way of Solving Word Problems. Computers & Education 63, 119–130 (2013)
5. Beal, C., Arroyo, I., Cohen, P., Woolf, B.: Evaluation of AnimalWatch: An Intelligent Tutoring System for Arithmetic and Fractions. Journal of Interactive Online Learning 9(1), 64–77 (2010)
6. Chang, K.E., Sung, Y.T., Lin, S.F.: Computer-Assisted Learning for Mathematical Problem Solving. Computers & Education 46(2), 140–151 (2006)
7. Filloy, E., Rojano, T., Puig, L.: Educational Algebra. A Theoretical and Empirical Approach. Springer, New York (2008)
8. Hegarty, M., Mayer, R.E., Monk, C.A.: Comprehension of arithmetic word problems: A comparison of successful and unsuccessful problem solvers. Journal of Educational Psychology 87(1), 18–32 (1995)
9. Lang, P.: Behavioral treatment and bio-behavioral assessment: computer applications. In: Sidowski, J.B., Johnson, J.H., Williams, T.A. (eds.) Technology in mental health care delivery systems, pp. 119–137. Norwood, NJ: Ablex (1980)
10. Looi, C.K., Tan, B.: WORDMATH: A Computer-Based Environment for Learning Word Problem Solving. In: Díaz de Ilarraza Sánchez, A., Fernández de Castro, I. (eds.) Computer Aided Learning and Instruction in Science and Engineering, Lecture Notes in Computer Science, vol. 1108, pp. 78–86. Springer Berlin Heidelberg (1996)
11. Manassero, M.A., Vazquez, A.: Validación de una escala de motivación de logro. Psicothema 10(2), 333–351 (1998)
12. Marshall, S.P.: Schemas in Problem Solving. Cambridge University Press, New York (1995)
13. Reusser, K.: Tutoring Systems and Pedagogical Theory: Representational Tools for Understanding, Planning, and Reflection in Problem Solving. In: Lajoie, S.P., Derry, S.J. (eds.) Computers as Cognitive Tools, pp. 143–177. Lawrence Erlbaum Associates, Hillsdale, NJ (1993)
14. Riley, M.S., Greeno, J.G., Heller, J.L.: Development of Children's Problem-Solving Ability in Arithmetic. In: Ginsburg, H.P. (ed.) The development of mathematical thinking, pp. 153–196. Academic Press, New York (1983)

15. Weiner, B.: An attribution theory of achievement motivation and emotion. Psychological Review 92, 548—573 (1985)

Modifying Field Observation Methods on the Fly: *Creative Metanarrative* and *Disgust* in an Environmental MUVE

Jaclyn Ocumpaugh[1], Ryan S. Baker[1], Amy M. Kamarainen[2], Shari J. Metcalf[3]

[1] Teachers College Columbia University, New York, New York
jo2424@tc.columbia.edu
[2] NY Hall of Science, New York, New York
[3] Harvard University, Massachusetts

Abstract. Automated detection of constructs associated with student engagement, disengagement, and meta-cognition plays an increasingly prominent part of personalized online education. Often these detectors are trained with ground truth labels obtained from field observations, a method that balances collection speed with label quality. Some behaviors and affective states (e.g., boredom) are regularly modeled across learning environments, but other constructs (e.g., gaming the system) manifest in fewer systems. New environments create the possibility of entirely unexpected constructs. In this paper, we describe how a field observation protocol (already proven effective for affect and behavior detection in several systems) was adapted to provide the flexibility needed to document previously unidentified or rare constructs. Specifically, we describe the in-field modification of the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) to accommodate categories not previously established (e.g., *creative metanarrative*) during observations of an educational multi-user virtual environment (MUVE). We also discuss the importance of developing methods that allow researchers to conduct such explorations while still capturing standard data constructs.

1 Introduction

As educational software has become more advanced, greater emphasis has been placed upon personalizing systems to react sensitively to student needs. Early work to model and adapt to student knowledge in tightly-scaffolded systems has given way to efforts to detect more ill-defined constructs (e.g. student engagement and meta-cognition) in more open-ended systems (e.g. virtual worlds). One approach to determining engagement with educational software is to construct automated detectors of affective states and behaviors, which can then be used both to research affect and learning [6, 8, 11] and to drive automated interventions [1,10].

Automated detectors have been produced from a variety of different data sources. Physical sensors (e.g., webcams, posture sensors and electroencephalograms) can be quite effective, but are often costly and fragile, making implementation difficult, particularly in poorer schools, leading some to develop sensor-free affect detection based on field observations [4, 15]. Recent research has expanded the scope of behavior detection to a wide range of systems, including games and simulations. As new environments are studied, we find that student behaviors differ across environments. Gaming the system is not seen in systems without feedback. WTF behaviors are more common in games than in tightly-constrained systems, and so on. As new systems are designed, fully anticipating relevant constructs may be impossible, particularly if classroom access or resources are limited. Given these concerns, researchers need coding methods that rigorously document known/expected constructs while being robust to unexpected findings is important.

In this paper, we discuss the adaption of the Baker Rodrigo Ocumpaugh Monitoring Protocol, (formerly the Baker-Rodrigo Observation Method Protocol), or BROMP, to address these concerns. BROMP is an established field observation method. It has been used to collect ground truth data for sensor-free models of affect and behavior and to study student engagement in non-technology-mediated learning environments. BROMP has already been successfully used to develop sensor-free affect detection in a variety of systems, including Cognitive Tutor [4] ASSISTments [15], and EcoMUVE [5] (the software described in this study). Here, we describe the expansion of BROMP coding schemes *in situ* to accommodate new affective and behavioral constructs that manifested during observations of EcoMUVE [13]. Currently, these constructs (*disgust* and *creative metanarrative*) are not typically coded for during field observations of educational software, but they will likely prove important as we increasingly rely upon virtual worlds for educational instruction.

2 Quantitative Field Observations (QFOs) using BROMP

The development of BROMP began in 2004 with field observations of students who were supposed to be learning from the Cognitive Tutor but were actually gaming the system [2, 3]. It was extended in 2007 when affective states were added as a second coding scheme [16], and further extended with the addition of teacher behaviors as well as student behaviors in some contexts [9]. In 2012, the method was formalized with the creation of a training manual [14]. New coders must achieve an adequate inter-rater reliability (Cohen's Kappa of 0.6 for both affect and behavior, individually) with a trainer in order to become BROMP-certified. At present, 60 individuals have been certified for coding in the United States, the Philippines, and India.

BROMP works well for collecting ground truth observations of student affect and behavior both because of its simplicity and because the protocol is enforced by an app designed for Android, known as the Human Affect Recording Tool (HART) [4], which streamlines data collection process. At the beginning of each observation session, a coder inputs student login information into the HART application and selects a coding scheme. HART then presents each student's login info back to the coder in the order in which they were entered. The coder then selects the behavioral and affect categories being presented by that student, ignoring the behaviors and affective states of other students except to the degree to which that information is contextually relevant to the student being coded.

3 BROMP Coding Schemes

During BROMP observations, behavior and affective states are coded separately but simultaneously. The coder has up to 20 seconds to categorize each student's behavior and affect, but records only the first thing he or she sees. For example, if a student is throwing a pencil at the teacher at the start of the observation, but then re-engages with the software while the coder is deciding what affective state is presenting, the behavior is recorded as off-task. In situations where a student has left the room, where the affect or behavior do not match any of the categories in the current coding scheme, or when the student can otherwise not be adequately observed, a '?' is recorded and that observation is eliminated from the data used to train automated detectors. This approach is valid when constructs that do not fit the coding scheme are rare, but researchers often need the flexibility to document new constructs.

The first BROMP publication to incorporate affect included seven different affective states and six behavioral categories [16]. These consisted of boredom, confusion, delight, surprise, frustration, flow, and neutral (drawn from [7]) as well as on task, on task conversation, off-task conversation, off-task solitary behavior, inactivity, and gaming the system (drawn from [2, 3]). However, at present, there are 24 coding schemes available, and it is possible to customize HART to a new schema.

The most commonly used BROMP coding schemes were developed for the Pittsburgh Science of Learning Center (PSLC). PSLC affective states include boredom, confusion, engaged concentration, frustration, and ?, while behavior categories include on task, on-task conversation, off-task, gaming the system, and ?. Because these constructs are seen as particularly relevant to educational settings, they are included in most BROMP coding schemes, but each time we work with a new learning environment, we reevaluate to ensure we are documenting all of the constructs relevant to that system and population.

4 Adapting BROMP Coding Schemes to EcoMUVE

When developing a coding scheme for EcoMUVE, expert field observers drew from prior coding schemes, from a qualitative pilot study, and from the EcoMUVE designers' expertise. We extended prior schemes with delight (which is seen substantially more in games than ITS) and sorrow, which is not typically included in educational research on affective states, being seen as rare [8]. We also extended the coding schema to enable us to document any categories that were unanticipated before entering the field, appending 3 different "user defined" categories (2 behavioral categories and 1 affective category) that the expert observer could specify in field.

Very early in the fieldwork, an affective state distinct from the anticipated categories emerged. As students began to explore this virtual world on day one, several reacted strongly EcoMUVE activities that they would have found "icky" in the real world, including tasks involving pond water or discoveries of dead fish. These reactions were coded as *disgust*, labeled as *User Defined 3* in HART. *Disgust* is rare in most learning, including EcoMUVE (0.04%) despite being one of Ekman's core emotions. Still, it was more prevalent than sorrow (0.03%), a category anticipated prior to fieldwork. Despite its negative valence, it indicates a lack of indifference. We do not yet know if it is positively or negatively associated with learning in EcoMUVE, but identifying this construct allows us to study how students respond to it. Anecdotally, students in this study maintained engagement once the *disgust* faded, but it could be an early indicator of later disengagement.

As fieldwork progressed, an unanticipated behavioral category was also identified. This behavior, which we term *creative metanarrative* (CM), was an unusual form of on-task conversation where students constructed their own storyline, often involving rogue police officers and illicit activities that did not reflect EcoMUVE design elements. CM differs from several other constructs that have been previously identified in the literature. While students often discuss content with each other during online learning (what [17] terms metanarrative), these students were transforming the plot of the game into a storyline that was more interesting to their peers. On it's face, this sounds similar to [12]'s transforming the game mechanic, which also includes a social component, or to previously identified WTF behaviors [18], but CM differs from these constructs behavior because it is not clear that it detracted from EcoMUVE's primary learning activities. In fact, the alternative storylines manufactured by these students may have made the software experience more exciting, forestalling the sort of unproductive within-game behaviors documented in [17, 18].

In contrast with the addition of *disgust* (which was coded within HART as soon as it was identified), the observer took more time to begin using the User Defined button in the behavioral coding scheme to code for *creative metanarrative*. This delay was driven by CM's relatively low frequency. Unlike *disgust*, CM did not manifest until the second day of field observation and only comprised 1.2% of the observations. (This is a low rate, but equal to the off-task behavior observed in this study.) Instead, the observer manually recorded this event on paper using the observation number and student number that HART provides as a reference at the top of each observation screen. After careful discussions with other BROMP-certified coders at the end of the second day of fieldwork, the field observer officially began automatically recording CM (using User Defined 1) in the field and the initial (manually recorded) instances were changed from the more generic on-task conversation to CM in the HART files.

5 Adapting BROMP Coding Schemes to EcoMUVE

Educational technology continues to evolve, and as it does researchers must have the tools that allow the agility to accurately and succinctly define relevant affective and behavioral constructs. As virtual worlds and other forms of educational software become more common educational tools, researchers are increasingly recognizing the importance of developing systems that are sensitive to indicators of student engagement. In particular, different systems promote different behavioral and affective responses. The quality and cost-effectiveness of field observation methods like BROMP make them an attractive option for collecting the ground truth labels needed for automated detectors of affect and behavior. In this paper, we discuss rapid, in-field extensions to BROMP (and HART, the software app used to enforce BROMP) that increase our ability to identify new constructs as we study student engagement in new software systems and populations.

Specifically, these extensions increase observers' agility to add unanticipated categories to the coding schemes in field, refining construct validity. While not correlated constructs, the two categories added in this study, *disgust* and CM, share qualities that are notable to educational researchers. Both manifest with rather prominent student displays within the classroom and may have broader impacts than their frequency would otherwise suggest. Both seem likely to reoccur in other virtual environments, suggesting that it may be increasingly important to take these constructs into account. Finally, both seem undesirable at first glance, but are actually indicators of engagement, suggesting that they may have interesting and complicated interactions with student outcomes. As researchers work to improve the sorts of engagement measures that facilitate the personalization of MUVEs, adding *disgust* and *creative metanarrative* to the suites of automated detectors already developed for systems like EcoMUVE [5] could substantially increase our understanding of learning and engagement, leading to greatly enhanced personalization options.

Acknowledgements

This research was supported by the Bill & Melinda Gates Foundation (#OPP1048577) and by the Institute of Education Sciences, U.S. Department of Education (#R305A080514).

References

1. Arroyo, I., Park Woolf, B., Cooper, D., Burleson, W., Muldner, K. (2011). The impact of animated pedagogical agents on girls' & boys' emotions, attitudes, behaviors & learning. *11th Int. Conf. Adv. Learning Tech. (ICALT),* 506-510.
2. Baker, R.S., Corbett, A., Koedinger, K., Wagner, A. (2004). Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System." *Proc. ACM CHI: Computer-Human Interaction*, 383-390.
3. Baker, R.S., Corbett, A.T., Koedinger, K.R. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. *Proc. 7th Int. Conf. Intell. Tutoring Sys*, 531-540.
4. Baker, R.S., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A., Salvi, A., Aleven, V., Kusbit, G., Ocumpaugh, J., Rossi, L. (2012). Towards Sensor-free Affect Detection in Cog Tutor Algebra. *Proc. 5th Int. Conf. Ed. Data Mining*, 126-133.
5. Baker, R.S., Ocumpaugh, J., Gowda, S.M., Kamarainen, A., Metcalf, S. (in press) Extending Log-Based Affect Detection to a Multi-User Virtual Environment for Science. *Proc. 22nd Conf. User Modelling, Adaptation, & Personalization*.

6. Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing,* 1(1),18-37.
7. D'Mello, S., Craig, S., Witherspoon, A., McDaniel, B., Graesser, A. (2005). Integrating affect sensors in an intelligent tutoring system. *Affective Interactions: The Comp. in the Affective Loop Wkshp, Int. Conf. Intell. User Interfaces,* 7-13.
8. D'Mello, S., Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. *Cognition & Emotion* 25(7),1299-1308.
9. Godwin, K.E., Almeda, M.V., Petroccia, M., Baker, R.S., Fisher, A.V. (2013) Classroom activities and off-task behavior in elementary school children. *Proc. Annu. Meet. Cog. Sci. Soc.*, 2428-2433.
10. Lehman, B., D'Mello, S., Strain, A., Millis, C., Gross, M., Dobbins, A., Wallace, P., Millis, K., & Graesser, A. (2013). Inducing and tracking confusion with contradictions during complex learning, *Int. J. Artificial Intell. Ed.*, 22(2),85-105.
11. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R. (2013). Sequences of Frustration & Confusion, & Learning. *Proc. Int. Conf. Ed. Data Mining*, 114-120.
12. Magnussen, R., Misfeldt, M. (2004). Player Transformation of Educational Multiplayer Games. *Proc. Other Players.* Copenhagen, Denmark.
13. Metcalf, S., Kamarainen, A., Tutwiler, M.S., Grotzer, T., Dede C. (2011). Ecosystem science learning via multi-user virtual environments. *Int. J. Gaming & Computer-Mediated Simulations.* 3(1), 86.
14. Ocumpaugh, J., Baker, R.S.J.d., Rodrigo, M.A. (2012). The Baker-Rodrigo Observation Method Protocol (BROMP) 1.0 *Training Manual*.
15. Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. *Proc. 3rd Int.Conf. Learning Analytics & Knowledge*, 117-124.
16. Rodrigo, M.M.T., Baker, R.S.J.d., Lagud, M.C.V., Lim, S.A.L., Macapanpan, A.F., Pascua, S.A.M.S., Santillano, J.Q., Sevilla, L.R.S., Sugay, J.O., Tep, S., Viehland, N.J.B. (2007). Affect and Usage Choices in Simulation Problem Solving Environments. *Proc. Artificial Intell. Ed.* 145-152.
17. Rowe, J. P., Shores, L. R., Mott, B. W., & Lester, J. C. (2010). A framework for narrative adaptation in interactive story-based learning environments. *Proc. of the Intell. Narrative Technologies III Workshop 14*.
18. Wixon, M., Baker, R.S.J.d., Gobert, J., Ocumpaugh, J., Bachmann, M. (2012). WTF? Detecting Students who are Conducting Inquiry Without Thinking Fastidiously. *Proc. Int. Conf. User Modeling, Adapt. & Personalization*, 286-298.

Personalized Web Learning: Merging Open Educational Resources into Adaptive Courses for Higher Education

Peter A. Henning[1], Florian Heberle[1], Alexander Streicher[3], Andrea Zielinski[3], Christian Swertz[4], Jürgen Bock[2], Stefan Zander[2], and the INTUITEL collaboration

[1] Institute for Computers in Education, Karlsruhe University of Applied Sciences, Karlsruhe, Germany
[2] FZI Research Center for Information Technology, Karlsruhe, Germany
[3] Fraunhofer IOSB, Karlsruhe, Germany
[4] University of Vienna, Austria

Abstract. In this paper, educational and technical challenges for applying learning pathways in Massive(ly) Open Online Courses in higher education are outlined. We argue that quality issues and didactical concerns may be overcome by (1) reverting to small Open Educational Resources that are (2) adaptively joined into concise courses by considering (3) predefined learning pathways with proper semantic annotations and (4) the observation of learner behaviour. Such a merger does not only require conceptual work and corresponding support tools, but also a new meta data format and an engine which interprets the semantic annotations as well as the measures of learner's actions. These factors are then turned into didactically meaningful recommendations for the next learning steps, thereby creating a personalized learning pathway for each learner. The EU FP7 project INTUITEL is introduced, which has already contributed to the conceptual work and is currently developing the software to achieve these tasks.

1 Introduction

Massive(ly) Open Online Courses (MOOCs) involving thousands of learners via internet are currently a major topic in technology enhanced learning (TEL). With this new approach, inquisitive learners from all over the world can participate in the lectures of proven experts. As formulated enthusiastically in the New York Times: *"...even in a remote developing country like Mongolia [...] you can find high-school students tuning into courses from American universities like M.I.T., Harvard and Berkeley"* [1]. If one follows the UNESCO [2], Open Educational Resources (OER) could even provide a solution to the world's educational problems.

Although there is a lot of praise, there is also a lot of critique. One aspect that is often discussed concerns the high dropouts rates MOOCs usually suffer from, which according to selected studies (e.g. [3–5]) amount up to 90%. However, this number has to be analyzed critically, as it is questionable whether this is an appropriate measure (cf. [6]). While 10% of thousands of students is still a large number, students also have varying motivations to enlist in a MOOC and some never actually planned to finish a course—and, to our knowledge, one contributing factor is the rigidity of MOOC learning. The relevant question thus is how to give those learners the best support, who actually planned but did not finish a course.

In this context, the different cultural and educational backgrounds of the students make the provision of knowledge in a one-size-fits-all manner questionable. Consider e.g. that *"Chinese classrooms tend to be more structured and authoritarian than classrooms in the West, [...] American schools try to encourage critical thinking skills and student interaction with teachers"* [7]. When applying these different cultural challenges to the creation of MOOCs, two fundamentally different courses will result. If different (culturally motivated) learning styles are integrated into the course directly, the students' time and effort to adjust is reduced, individual learners can be better supported and learning satisfaction is likely to increase. We believe that personalization of learning content is a very promising approach to achieve this. In this paper, we therefore investigate a technical solution given by the EU project INTUITEL[5] on how MOOC learning can be made more individual, human-centered and interactive.

Such a development also appears useful to overcome the problem of interaction between students and teaching staff in MOOCs, which is almost impossible for sheer numerical reasons: a higher learning satisfaction does lead to a lesser demand for personal interaction[6].

MOOCs also have the disadvantage that full-fledged courses with high quality content are expensive to produce, difficult to maintain and almost impossible to adapt to individual needs. Conversely, in the past few years, a large number of "small information pieces" have shown up on the internet, providing excellent free content covering almost any subject. We call these artifacts *Small Open Educational Resources (SOER)* (cf. [8, 9]). The second aspect of this work therefore elaborates on how SOER can be effectively orchestrated along predefined learning pathways in order to create a MOOC-like course.

2 Technical approach of INTUITEL

In the following, we assume that the learning content for a TEL course consists of a set of *knowledge objects* (KOs). They may be accessed separately and in different order according to some predefined sequence we call a *learning pathway* (LP). The desired personalization then consists of selecting an order of the knowledge objects based on considering all the aforementioned aspects for an individual learner - but in contrast to other approaches, INTUITEL avoids *enforcing* such an order. Even more, at any stage the learner is given full freedom to chose his preferred KO. We consider this freedom to be one of the main advantages of self-paced learning, not to be dropped in favour of a more or less "programmed" learning for reasons of efficiency and speed.

---

[5] INTUITEL = Intelligent Tutorial Interface for Technology Enhanced Learning, http://www.intuitel.eu, is funded in the 7th framework programme of the European Union (FP7-ICT-2011.8, Challenge 8.1) under grant no. 318496

[6] This experience has been gained by one of the authors (P.A.H.) in a long standing involvement in the Virtual University of Bavaria in Germany with more than 25.000 enrolled students in the fall of 2013, see http://www.vhb.org

The INTUITEL system then tries to give a non-intrusive guidance, much in the way a caring and responsible teacher would do on the basis of his deep pedagogical knowledge and respecting the fact that all learners are different [10]. This task is addressed for five different leading eLearning platforms (eXact LCMS[7], Clix[8], Crayons[9], ILIAS[10] and Moodle[11]).

While these are typical Learning Management Systems (LMSs) and not MOOC platforms, the underlying concept of personalizing the learning process is identical[12].

The INTUITEL system has been designed in a way that decouples the presentation of content from the provided service to act independently from the used "front-end". This allows it to evaluate the added value in a smaller context before applying it to large scale settings. Expanding the service to MOOC-style courses is then an issue of scalability and optimization rather than a conceptual one. In the following, we introduce the main components and give an overview of the proposed system.

Extension of the hosting platform: The enhanced learning software interacts with INTUITEL via a lightweight web service, which gives access to its data and user interface:

1. General services to, for instance, pre-load metadata for the enhancement of learning material.
2. User score extraction (USE) to acquire learner-specific data.
3. Tutorial guidance (TUG) to exchange information with the learner.
4. Learning object recommendation (LORE) to suggest the most suitable learning material.

The specification is open and can be applied to every type of LMS, furthermore the concrete implementations for ILIAS and Moodle are open source and usable as blueprints for other systems.

Hierarchy of ontologies: A set of static and dynamic ontologies build on one another to represent learner- and course-specific data as well as adaption strategies (cf. user, domain and teaching model [11]). The basis of this hierarchy, the pedagogical ontology (PO), is founded on Meder's web didactics [12] and insights gained from the L3 project [13]. It contains the vocabulary and relations necessary for enhancing learning content with didactical and technical metadata [14]. The Semantic Learning Object Model (SLOM) describes how learning material needs to be enhanced in order to be interpretable by the INTUITEL system. Software to comfortably edit metadata and learning pathways with a graphical user interface is currently in development. In the optimal scenario, teachers will only be required to interrelate content with LPs, while the remaining data is determined automatically. INTUITEL therefore also provides a rather

[7] cf. http://www.exact-learning.com/
[8] cf. http://www.im-c.de/en/
[9] cf. http://www.iosb.fraunhofer.de/servlet/is/4525/#
[10] cf. http://www.ilias.de/
[11] cf. http://moodle.com/
[12] We want to emphasize at this point, that a commercial partner of the INTUITEL project very successfully provides MOOCs to industrial customers and now actively integrates INTUITEL features in their commercial system.

complete tool suite for non-technical target groups, attempting to provide innovation as well.

Back-end: Apart from aggregating the required information, the INTUITEL back-end creates learning recommendations and feedback with a combination of modules using Java and OWL reasoners. For each learning step, the respective data is at first pre-processed in the Learning Progress Model (LPM), then analyzed in the INTUITEL Engine and post-processed in a block called Recommendation Rewriter.

Communication layer: To enable an efficient message exchange, the INTUITEL communication layer (CL) interconnects the previously described components and manages message distribution. Since all exchanged data is based on XML, the data transmission is relatively simple. Two types of messages transmission technologies are available, HTTP and XMPP. Nevertheless, questions of scalability need to be considered.

3 Creating Personalized Learning Recommendations

Within the INTUITEL project the learning process is analyzed by considering the learning pathway of a learner through a course and by gathering additional data. The system may draw these data from four different sources: (i) *the learning content*, i.e. what has to be learned? (ii) *the learner history*, i.e. what has already been learned? (iii) *the learning environment*, i.e. what are the temporal, spatial and physical parameters? (iv) *the learner*, i.e. what are the characteristics of this person?

In the context of INTUITEL, we extract from these sources so called *didactic factors* that are symbolic statements with each of them having a distinct meaning for the learning process. They are defined statically, but calculated for each learner individually. By combining them with the learning pathway information, it is possible to deduce that a certain knowledge object is better suited for the learner than another one. Moreover, it also is possible to state *why* this is the case (e.g. because it is age-appropriate, has a suitable difficulty level, etc). This enables self-reflection of the learners and thus increases their metacognitive skills.

This personalized recommendation and feedback creation process is started at the moment when a learner begins a new learning step. The relevant situational and learner-specific data is requested from the learning platform and also the domain and content information is retrieved from the corresponding SLOM repository. With this and the previously stored data (e.g. past recommendations and beforehand requested information), the most suitable learning pathways and the didactic factors are determined in a first pre-processing step.

INTUITEL takes two approaches for finding optimal learning pathways for a learner, an interactive and a technological one. Firstly, it may carry out an interactive dialogue with each learner. For this case, teachers can add notes and describe for whom a particular pathway is most suitable. This makes it possible for learners to make an informed choice, but one has to keep in mind that self-assessments are commonly qualitatively limited [15]. INTUITEL therefore also implements a data-driven approach that allows evaluating choices algorithmically [16, 17]. With this method, the system can automatically come to conclusions whether the current selection is optimal, or if the learner's behavior indicates that another learning pathway would be more suitable.

The basic definitions of the didactic factors and their value ranges are present as a separate ontology, which is interpreted by the LPM. This allows it to incorporate various soft aspects into eLearning, like e.g. motivation or other emotions [18].

All these data are then forwarded to the INTUITEL Engine. This component is a combination of a set of Java modules and standard OWL-reasoners (like e.g. FaCT++ or HermiT). Its task is to analyze the provided ontologies in order to identify the most suitable knowledge objects with regard to the most suitable learning pathways and the current situation as expressed by the didactic factors. It therefore generates semantic queries and starts the most efficient reasoners for the specific query. INTUITEL thereby builds on the results and insights of the THESEUS project [19] and in particular the HERAKLES Reasoning Broker [20]. Not only does this allow to exchange the reasoner, but the scalability of the reasoning process necessary for a large number of participants is also provided. The output of this procedure is then interpreted in a post-processing step in order to create the final learning recommendations and also generates natural language messages for the learner, if appropriate.

This multi-layered procedure allows a high level of personalization, which is based on sound didactical models. The learning progress of each learner is evaluated gradually in respect to multiple aspects. This not only allows to select the most suitable learning pathway for each student, but also to determine which of the routes on these pathways fits the individual cultural and educational needs of the learner. In this process the didactic factors can furthermore be used to guide learners in regard to fine-granular aspects and thus consider the given individual boundary conditions.

Let us note, that while this recommendation process of course follows the well-known reference model for Adaptive Learning Environments [11], it is rather different from existing implementations of this model by keeping the learner's freedom of choice in every moment and therefore acting as a non-intrusive guide.

4 Personalized Web Learning

Apart from providing a manageable adaptive system, the INTUITEL approach also allows to overcome the second deficiency of MOOCs pointed out in the introduction. This may be attributed to the fact that learning material is extended with SLOM data externally, i.e. the content remains as is. Introducing additional elements in the material is not necessary. Course authors are thus not restricted in their choices of what learning material they provide and in which style they do it. They just need to add further information to it in a subsequently following step—and in principle this material can reside anywhere on the internet. An effectively personalized course, consisting only of the content relevant for a certain learner, but nevertheless following a well-defined didactical model, will be the result.

This approach preserves the high level of freedom for course creation currently demanded by authors, but allows the reuse of their content in a novel way. Given that this is extended to a (possibly decentralized P2P-) network of SLOM repositories, renowned authors from all over the world can link their Small OER via URIs and provide their learners with a huge knowledge space. It is conceivable that such a knowledge space can attract as much learners as one of the current MOOCs—but more flexibly so and with an almost unlimited individuality. We leave it open whether one should call this a "MOOC" then.

A course designer—or many of them—can contribute to this knowledge space not only by adding new learning content. They can also contribute a new Cognitive Content Map (CCM), which defines new learning pathways through this knowledge space. Cultural adaptation is only one of the possibilities such opened. Another possibility is to keep such a course up to date: one may add the actuality of a knowledge object to the set of didactical factors and then automatically receive recommendations to use more recent learning content with higher priority. At the same time, this creates an innovative learning pathway: adding new learning content while keeping the old one also allows learning about the history of a knowledge domain. Last but not least, we mention two further options: (i) creating an international federation of eLearning content providers and (ii) finding similar learning material via its SLOM properties.

5 Summary

In this paper, we outlined a way to make MOOCs more suitable for a greater variability of learning needs, by semantically annotating their parts and running them in a semantically enhanced learning platform. Such a platform is not necessarily a LMS, but could also be a future version of current MOOC platforms. As pointed out above, the INTUITEL project generalizes this semantic approach to be independent of the technical details of the front end, and is currently also integrated into a successful MOOC platform.

INTUITEL therefore contributes to key aspects of MOOCs, e.g. how to create online courses in a didactically meaningful way, how to add semantic interoperability and how learning platforms can assist in that. Such a semantic reconstruction of current MOOCs will, in our estimate, contribute to resolve their current problems.

We furthermore outlined how complex large courses may be constructed from Small OER, thereby resolving the problems maintainability and adaptability of current MOOCs. The INTUITEL system here serves as the "glue" integrating various learning content into a greater knowledge space.

Let us furthermore emphasize again that our approach, while of course implementing the well-known reference model for Adaptive Learning Environments [11], does so in a fashion which is rather different from previous implementations [16, 17]. Preserving the freedom of choice for each learner is targeted to remove the observed rigidity in present MOOC learning.

By providing the information on the learning process in a suitable format, and by delivering the necessary interfaces, INTUITEL also opens the doors for implementing other technologies like learning analytics and data mining directly into the learning platforms[13]. With the insights that can be gained from a data driven perspective, this could result in new didactical approaches and thus enhance education in general.

[13] cf. APPLYTEL project proposal by the INTUITEL consortium

References

1. Pappano, L.: The Rise of MOOCs, The New York Times Magazine, (Sept. 16, 2013)
2. UNESCO World OER Congress releases 2012 Paris OER Declaration (UNESCO, Paris 2012)
3. Knowledge@Wharton, Koller D.: MOOCs on the Move: How Coursera is Disrupting the Traditional Classroom, Retrieved 13 May 2014 from http://knowledge.wharton.upenn.edu/article/moocs-on-the-move-how-coursera-is-disrupting-the-traditional-classroom/
4. Breslow, L. B., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., Seaton, D. T.: Studying learning in the worldwide classroom: Research into edX's first MOOC. Research & Practice in Assessment, 8 (2013) 13-25.
5. Jordan, K.: Initial trends in enrollment and completion of massive open online courses. The International Review of Research in Open and Distance Learning, 15(1) (2014)
6. Rivard, R.: Measuring the MOOC Dropout Rate (March 8 2013), Retrieved 20 Dec 2013 from http://www.insidehighered.com/news/2013/03/08/researchers-explore-who-taking-moocs-and-why-so-many-drop-out
7. Lockette, T.: UF study: Contrasting teaching styles in U.S.-China classrooms may influence students learning preferences (2006) Retrieved 13 Dec 2013, http://news.ufl.edu/2006/08/22/chinese-learn/
8. Yuan, L., MacNeill S., Kraan W.: Open Educational Resources - Opportunities and challenges for higher education. Joint Information Systems Committee (JISC) CETIS (2008) 1-34
9. Duval, E., Wiley, D.: Guest Editorial: Open Educational Resources. IEEE Transactions on Learning Technologies, 3(2) (2010) 83-84.
10. Koedinger, K., Tanner, M.: 7 Things You Should Know About Intelligent Tutoring Systems, (July 2013) Retrieved 16 Dec 2013, from https://net.educause.edu/ir/library/pdf/ELI7098.pdf
11. De Bra, P., Houben, G., Wu, H.: AHAM: A Dexter-based Reference Model for Adaptive Hypermedia. In: Westbomke, J., Will, U.K., Leggett, J.J., Tochtermann, K., Haake, J.M. (eds): Proceedings of the Tenth ACM Conference on Hypertext and Hypermedia: Returning to Our Diverse Roots. ACM, New York (1999) 147-156
12. Meder, N.: Web-Didaktik: Eine neue Didaktik webbasierten, vernetzten Lernens, (Bertelsmann, Bielefeld 2006)
13. Leidig, T.: L3 -Towards and Open Learning Environment. ACM Journal of Educational Resources in Computing, Vol. 1 No. 1 (2001)
14. Swertz,C., Schmölz, A., Forstner, A., Heberle, F., Henning, P.A., Streicher, A., Bargel, B.A., Bock, J., Zander, S.: A Pedagogical Ontology as a Playground in Adaptive Elearning Environments. In: Horbach, M. (ed): INFORMATIK 2013: Informatik angepasst an Mensch, Organisation und Umwelt, GI-Edition Lecture Notes in Informatics (LNI), P-220 (2013) 1955-1960
15. Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K.E., Perrier, L.: Accuracy of Physician Self-assessment Compared With Observed Measures of Competence. JAMA The Journal of the American Medical Association, 296(9) (2006) 1094-1102
16. Swertz, C., Schmölz, A., Forstner, A., Dambier, N., Heberle, F., Henning, P.A.,Streicher, A. Burghart, C., Bock, J., Badii, A., de la Fuente, L., Parodi, E., Thiemert, D., Gal, E., Ronen, M., Zander, S. 2014: Lernpfadmodellierung mit der Webdidaktik für die adaptive Erweiterung von Lernmanagementsystemen. In: Breiter, A., Rensing, C. (eds.): Proceedings of DeLFI 2013, Bremen 2013, GI-Edition Proceedings Vol. 218 (2013) 277-280

17. Henning, P. A., Heberle, F., Swertz C., Schmölz, A., Burgos, D., de la Fuente Valentin, L., Gal, E., Verdu, E., de Castro, J.P., Parodi, E., Schwertel, U., Steudter, S.: Learning Pathway Recommendation in Intelligent Tutoring Systems, submitted to the DeLFI 2014 conference
18. Erriquez, E., Grasso, F.: Generation of Personalized Advisory Messages: an Ontology Based Approach. In: Puuronen, S., Pechenizkiy, M., Tsymbal, A., Lee, D.J. (eds): Proc. 21st IEEE International Symposium on Computer-Based Medical Systems, IEEE Los Alamitos (2008), 437-442
19. BMWi Federal Ministry of Economics and Technology: The THESEUS Research Program (2008) Retrieved from http://www.bmwi.de/English/Redaktion/Pdf/theseus-research-program,property=pdf,bereich=bmwi2012,sprache=en,rwb=true.pdf
20. Bock, J., Tserendorj, T., Xu, Y., Wissmann, J., Grimm, S.: A Reasoning Broker Framework for OWL. In: Hoekstra, R., Patel-Schneider, P.F. (eds): Proceedings of the 6th International Workshop on OWL: Experiences and Directions (OWLED 2009),CEUR WS proceedings, Chantilly (2009)

Gamification: metacognitive scaffolding towards long term goals?

Lie Ming Tang and Judy Kay

University Of Sydney
http://chai.it.usyd.edu.au
ltan8012@uni.sydney.edu.au

Abstract. The ability to self regulate is a key skill in learning. This is especially relevant for open learning environments such as MOOCs. Metacognitive scaffolding refers to computer based support is used to teach and engage users in metacognition and self regulated learning. These techniques have been found very useful in supporting students in e-learning environments. Studies in game based learning suggests game playing engage players in metacogniton as well as self regulated learning. We see great potential in applying Gamification as a form of metacognitive scaffolding to improve self regulation in learners. Gamification can also provide a framework to personalise self regulated learning support. In this paper, we present our ideas and guidelines for applying Gamification as metacognitive scaffolding. We will illustrate through examples of how they can be applied and discuss how these concepts can be the foundation for future work.

Keywords: metacognition, self regulated learning, gamification, personalisation

1 Introduction

Our research focus on how we can help people better achieve *Sisyphean* goals which demands consistent, repeated effort over long periods of time [13]. We introduce the concept of gamification as a form of metacognitive scaffolding to address these challenges.

Metacognition refers to the knowledge and control an individual has over their thinking and learning activities [3]. It represents a huge body of work grounded in psychology since the 1970s [7]. It includes what people know about their own abilities, what influences their performance and their knowledge of tools and strategies. Self regulated learning refers to setting learning goals, attempt to monitor, regulate and control their cognitive and metacognitive processes in the service of these goals [19]. It is learning guided by metacongition. Metacognitive scaffolding refers to providing scaffolding or computer based support to enhance metacognitive awareness and self regulated learning [1, 17]. Many studies have indicated that people who engage in metacognitive processes and exhibit higher metacognitive awareness, achieve higher performance over the long term than those that do not [16, 1].

Gamification can be described as *'use of game design and game thinking in a non game context'* [6]. The idea is to apply game elements that have proven successful in engaging players and encouraging desired behaviour to applications where entertainment is not the main objective. Studies show game players exhibits a number of metacognitive and self regulated learning behaviours including planning and goals setting, self monitoring, evaluation and strategy use [8]. While we have seen many examples such as fitbit, endomondo[1], [10] using Gamification techniques, the focus and objective of these approaches are on engagement and motivation for a specific desired activity or behaviour (e.g., increased physical activity, regular exercise) rather than developing self regulation skills towards long term goals attainment. While studies found evidence of self regulation and metacognition in game players [14, 8], we have yet to find cases for the combination of solidly grounded theories associated with metacognition and self regulated learning with emerging uses of gamification.

The key distinction and motivation of our position is for learners to be successful, it is important for systems to also engage and develop user in self regulation and metacognition as a skill rather than focus on a particular short term task or activity. Over three decades of research in metacognition and self regulation have shown that such development will lead to better performance and goal attainment over the long term [1]. In this paper, we will present ideas and guidelines for applying gamification as metacognitive scaffolds as a different perspective or focus. We will illustrate our ideas through examples of such scaffolding towards long term goals. Finally, we will offer concepts and ideas for future research in this largely unexplored area.

2 Related Work

Metacognitive self monitoring involves evaluating ones knowledge of cognition including monitoring performance, knowledge and understanding. Self reflection refers to the process of comprehending and reasoning on the result of self monitoring. Planning include in goal setting, activating relevant background knowledge, selecting appropriate strategies, time management and resources allocation. Research suggests that experts in a particular task or domain are more self-regulated compared to novices largely due to effective planning that occurs prior to beginning a task [15]. Self evaluation and assessment refers to appraising the products and regulatory processes of one's learning. This can include performing self tests and assessments.

Studies in games based learning or educational games has examined their effect on a player or learner metacognition and self regulation. A recent study designed to engage students in learning software programming asked students to program virtual characters using Java to compete within a game environment. This study found students actively engage in analysing each others strategies, review, discuss and reflect on game results and performance. They also engage

[1] endomondo.com

in self evaluation and perform drills and practices [8]. The results of an survey on players in StarCraft and online Chess, both online games played by millions, show a large percentage of players engage in metacognitive and self regulated learning processes such as self evaluation, monitoring performance, practising and studying other player's strategies [8]. This indicates a great potential to scaffold metacognitive processes using gamification.

Previous approaches in Gamification focus on motivation and engagement for a particular task. Commercial fitness service providers such as fitbit routinely use achievement badges and challenges as motivation and engagement. However, they have been limited in teaching or fostering self regulation. For example, many systems ask users to set goals but do not focus on improving the quality and user's goal setting skill. MOOC providers such as KhanAcademy [2] adopt Gamification techniques such as badges and points to motivate and engage users to participate in different courses and challenges. These techniques do not focused on teaching or engaging users in self regulated learning or invoking the metacognitive processes.

We propose to design Gamification applications with a view of enhancing metacognition and self regulation skills. Indeed there are gamification examples that can be considered limited metacognitive scaffolding. For example, HealthMonth [3] use the concept of short term (i.e., monthly) achievable goals as a platform for achieving behaviour change and goal attainment. They use Gamification techniques to engage users to set goals, monitor their progress and set new goals. This is a form of metacognitive scaffolding as it engage users to set goals as well as scaffolding them to monitor and self evaluate. Over time, this approach has the potential to improve a user's goal setting and planning ability.

It is important to note here that gamification is a developing field of study and is not without criticism [9]. A number of pitfalls has been highlighted including overuse of extrinsic rewards. The emerging view in the community is these challenges are symptoms of poor design and application which can be overcome [9]. It is then important that we present key guidelines in game design and applying gamification.

Gamification practitioners recommend to design with different player personalities in mind [18]. A common notion is there are four player types within games: explorers (discovery), achievers (winning), socialisers (interaction, social) and killers (dominating others) [2]. Game design should align intended outcome with the personality and profile of the target users. With respect to rewards and motivation, it is recommended to align rewards with three basic motivation or needs grounded in self-determination theory [5]: autonomy (choice, self control), competence (feel effective, challenged) and relatedness (interact and connect with others). It is necessary to take into consideration how gamification design impact these needs. A final concept to lay the foundation is the theory of "Flow" [4]. It posits that we can achieve optimum user engagement, as long as

[2] khanAcademy.org
[3] healthmonth.com

the users are continually challenged by tasks that are not too difficult but still feel challenged.

The MDA (mechanics, dynamics and aesthetics) framework [11] is frequently used in game design [18]as a foundation for understanding games. MDA describes games and their behaviour from three perspective or "views" of the game. They are mechanics, dynamics and aesthetics. ***Mechanics*** are rules and game artefacts that users act on or manipulate such as scores, badges, leader-boards, rewards and levels. ***Dynamic*** refers to how the mechanics act on each other and can be thought of as behaviour or actions users engage in. Examples include sharing, collaborating, competing and cheating. ***Aesthetics*** refers to the resulting user experience from engaging with game mechanics and dynamics.

3 Gamified Metacognitive Scaffolding

In this paper, we will present our ideas for applying gamification as metacognitive scaffolds through a metacognitive "view" of the game mechanics (rules and artefacts) dynamics (interaction, behaviour) and aesthetics (user experience, feel) as described in the MDA framework. We will demonstrate the concepts through a hypothetical user 'Alice'. She is a young professional who commits to self development and learning in her profession through MOOCs and e-learning as well as maintaining long term health and fitness through regular physical activity and exercise.

Self monitoring and reflection. Rewards and reward schedules are powerful techniques that can engage users in self monitoring and reflection. For example, at variable intervals, the system send Alice questions (in the form of a quiz) and she is rewarded based on the *accuracy* of her knowledge in her own activities and performance. E.g., how regular does she participate in a MOOCs course, how well does she compare against her peers. This can encourage her to self monitor more closely, develop a habit and maintain this behaviour over time [18].

Planning and strategy. Game elements can be designed to engage users to practice planning, consider what resources they need and how to apply them, suggest strategies to follow and generally improve these skills. An example of this can be to use challenges and rewards specifically for planning and strategy use. E.g., achievement badges for setting goals and plans and completing within the plan. Rewards for sticking to her planning. Compare her planning with peers and providing feedback on her goal setting and planning abilities. There is opportunity here to personalise the techniques to use. For example, a system could make use of indicators of self efficacy or confidence when analysing Alice's planning. The objective is to scaffold her in the metacognitive processes of planning and goal setting rather than just doing as much as she can to complete a challenge.

We can also scaffold users to develop **Strategy** use. The guideline here is to design rewards that allow and / or highlight different paths to success. An example mechanic can be to highlight alternative strategies e.g., by showing the strategies of other learners or top performers [8]. Reward Alice on trying new and different strategies. E.g., badges that shows the number of learning strategies she used. A key consideration is to avoid the perception where success is defined by innate or natural abilities and confront the users need for competence [5]. For example, if the challenge is for Alice to achieve the four minute mile, which is within the domain of elite athletes, she may develop the perception or belief that there is little chance of success [4]. Instead, provide challenges that is personalised such as relative improvement (e.g., percent increase) or achieving a personal monthly goal.

Self evaluation and assessment. While existing approaches such as fitbit and KhanAcademy offers mechanics such as achievement badges and levels, they are mainly intended to show progress and motivate further activity. When applying gamification as a metacognitive scaffolding for self evaluation and assessment, the objective is the encourage users to engage in these tasks. Examples can be to reward based on frequency of self assessment, apply self evaluation quizzes and use of comparisons.

Collaboration and Group dynamics. Gamification is a powerful tool for engaging users in social dynamics including exploration, collaboration and competition (e.g., foursquare, fitbit). Game dynamics that engage users in team or group related activities have been found very successful in engaging users thus promising to apply toward metacognition. Examples include team score, achievements. Mechanics can be designed to engage users to share and discuss strategies, reflect on their achievements as individuals and as a group, socialize for motivation and encouragement [12].

Game Aesthetics. A key challenge for helping users achieve their long term *Sisyphean* goals is the need to be motivated and be persistent over the long term. We suggest that metacognitive game mechanics and dynamics needs to achieve game aesthetics that convey feelings of autonomy and competence [6]. For example, the game dynamics that encourage Alice to regularly engage in strategy use, planning and monitoring, can foster the feeling of competence as she is made aware of tools and strategies. In addition, by increasing her metacognitive skills through scaffolding we help foster feelings of autonomy and self efficacy term [5].

4 Discussion

When designing a systems that implements these ideas and guidelines, we must also consider the dynamic and aesthetic outcomes of the game as a whole not just from a metacognitive perspective. For example, game mechanics of leaderboards, badges and points may invoke self monitoring and reflection. At the same time these mechanics have the potential to demotivate some type of players [2].

A challenge worth noting is while some behaviour such as goal setting are easier to detect others such as self monitoring, mood, engagement requires more sophisticated measurement techniques. Also, not all MOOCs are the same and differ in instructional design significantly. Future work is needed in this area to identify what gamification design and self regulated learning scaffolding techniques are appropriate for different designs.

References

1. Roger Azevedo, Jennifer G Cromley, and Diane Seibert. Does adaptive scaffolding facilitate students ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, 2004.
2. Richard A Bartle. *Designing virtual worlds*. New Riders, 2004.
3. Ann Brown. Metacognition, executive control, self-regulation and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, 1987.
4. Mihaly Csikszentmihalyi. *Finding flow: The psychology of engagement with everyday life*. Basic Books, 1997.
5. Edward L Deci and Richard M Ryan. *Self-Determination*. Wiley Online Library, 1985.
6. Sebastian Deterding. Gamification : Toward a Definition. 2011.
7. John H. Flavell. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 1979.
8. Stephen R Foster, Sarah Esper, and William G Griswold. From Competition to Metacognition : Designing Diverse , Sustainable Educational Games. 2013.
9. Ian Glover. Play as you learn: gamification as a technique for motivating learners. 2013.
10. Geoff Goehle. Gamification and Web-based Homework. *Primus*, 2013.
11. Robin Hunicke, Marc LeBlanc, and Robert Zubek. MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI*, 2004.
12. Rik Janssen, Huub Prüst, Ir Robbert-Jan Beun, Pieterjan van Wijngaarden Msc, and Tjip de Jong. Persuasive technology to support self-directed learning in adult learning programs. 2012.
13. Judy Kay and Bob Kummerfeld. Mneme: shifting control of pervasive infrastructure to the user for flexible, pervasive support of sisyphean goals. In *Workshop on Pervasive Intelligibility, Pervasive*, 2011.
14. Jennifer Sabourin, Lucy Shores, Bradford Mott, James Lester, and North Carolina. Predicting Student Self-Regulation Strategies in Game-Based Learning Environments. 2011.
15. Gregory Schraw and David Moshman. Metacognitive theories. *Educational psychology review*, 7(4), 1995.
16. H Lee Swanson. Influence of metacognitive knowledge and aptitude on problem solving. *Journal of educational psychology*, 82(2), 1990.
17. Lie Ming Tang and Judy Kay. Lifelong user modeling and meta-cognitive scaffolding: Support self monitoring of long term goals. In *Second Workshop on LifeLong User Modelling, in conjunction with UMAP*, volume 2013, 2013.
18. Gabe Zichermann and Christopher Cunningham. *Gamification by Design: Implementing game mechanics in web and mobile apps*. O'Reilly Media, Inc., 2011.
19. Barry J Zimmerman. Models of self-regulated learning and academic achievement. In *Self-regulated learning and academic achievement*, pages 1–25. Springer, 1989.