Identifying Students' Characteristic Learning Behaviors in an Intelligent Tutoring System Fostering Self-Regulated Learning

François Bouchet, Roger Azevedo
Department of ECP
SMART Laboratory
McGill University
Montreal, H3A 1Y2, Canada
francois.bouchet
@mcgill.ca

John S. Kinnebrew, Gautam Biswas
Department of EECS/ISIS
Vanderbilt University
Nashville, TN 37203, USA
john.s.kinnebrew
@vanderbilt.edu

ABSTRACT

Identification of student learning behaviors, especially those that characterize or distinguish students, can yield important insights for the design of adaptation and feedback mechanisms in Intelligent Tutoring Systems (ITS). In this paper, we analyze trace data to identify distinguishing patterns of behavior in a study of 51 college students learning about a complex science topic with an agent-based ITS that fosters self-regulated learning (SRL). Preliminary analysis with an Expectation-Maximization clustering algorithm revealed the existence of three distinct groups of students, distinguished by their test and quiz scores (low for the first group, medium for the second group, and high for the third group), their learning gains (low, medium, high), the frequency of their note-taking (rare, frequent, rare) and note-checking (rare, rare, frequent), the proportion of sub-goals attempted (low, low, high), and the time spent reading (high, high, low). In this paper, we extend this analysis to identify characteristic learning behaviors and strategies that distinguish these three groups of students. We employ a differential sequence mining technique to identify differentially frequent activity patterns between the student groups and interpret these patterns in terms of relevant learning behaviors. The results of this analysis reveal that high-performing students tend to be better at quickly identifying the relevance of a page to their subgoal, are more methodical in their exploration of the pedagogical content, rely on system prompts to take notes and summarize, and are more strategic in their preparation for the post-test (*e.g.*, using the end of their session to briefly review pages). These results provide a first step in identifying the group to which a student belongs during the learning session, thus making possible a real-time adaptation of the system.

1. INTRODUCTION

Use of metacognition and self-regulated processes has been identified as a key element for successful learning in general [2; 19; 20; 22]. In the particular context of an intelligent tutoring system (ITS), it means it is crucial to ensure that students are actively using key self-regulated learning (SRL) processes, which can be achieved through prompts, scaffolding, and feedback. A major challenge is to make the ITS more adaptive to individual learning characteristics, such as browsing behavior and initiative in performing appropriate SRL processes.

Using MetaTutor, an agent-based ITS that fosters the use of SRL processes, we have collected a large amount of data from students interacting with the system while they were learning about the human circulatory system. In this paper, our goal is to answer two questions: (1) how can students be grouped according to their performance and their type of interaction with the system? and (2) how do specific learning behaviors of high- and low-performing students differ, in particular regarding their use of SRL processes in MetaTutor?

In this paper, we propose to answer the first question using a clustering approach that groups students with similar performance and scores on other system interaction metrics. For the second question, we analyze members of the three clusters (especially comparing high- and low-performing students) with a differential sequence mining method [11], which identifies statistically significant differences in frequent behaviors between clusters.

This paper is organized as follows. In section 2, we start by discussing related work that combines clustering and pattern mining techniques for analysis of data from computer-based learning environments. In section 3, we introduce the ITS used for data collection, MetaTutor, as well as theoretical grounding of its key features, which encourage learners to perform self-regulation monitoring and strategy as they learn with the system. Section 4 describes the data collected and the relevant events encoded as actions, as well as the clustering performed to distinguish different types of students. Section 5 presents the principles of the method of differential sequence mining, its application to the data, and the results obtained in terms of patterns of actions that distinguish students from different clusters. Section 6 then discusses the practical implications of those findings in terms of potential modifications to the ITS, before concluding in section 7.

2. RELATED WORK

Analysis of trace log data from users' interactions to better understand their learning process and distinguish groups of learners (*e.g.*, efficient versus inefficient ones) has been an

important area of research in educational data mining. For example, Perera *et al.* [15] follow a 2-step methodology like ours, as they start by using a clustering algorithm (k-means) to identify strong groups of students collaborating in a software development task using an open environment (TRAC). The students are first clustered according to a set of attributes extracted a posteriori, and then they use a modified version of the Generalized Sequential Pattern mining algorithm [17] to identify frequent sequences of actions that characterize the most successful groups. In [16], Romero *et al.* also use a combination of clustering and sequential pattern mining to identify different kinds of browsing behavior that students exhibit in their learning environment, "AHA!", in order to provide them links to the most appropriate pages.

With gStudy, Nesbit *et al.* [14] are interested in the use of self-regulation by students learning from multimedia documents. They apply sequential pattern mining to find common subsequences between groups of students, although they do not perform any clustering beforehand. Martinez *et al.* [13] pursue a similar approach and objective, as they aim to discover frequent sequences of actions that distinguish a group of students with high achievements from one with low achievements. They use a combination of pattern mining and clustering techniques to identify the most successful strategies in the context of a collaborative learning tool on a tabletop device. However, they first extract frequent patterns of actions and then cluster them in order to examine clusters of patterns associated with each group. Tang and McCalla [18] also use sequence mining and then clustering in their web learning environment, to facilitate instructional planning and diagnose students behaviors.

3. METATUTOR ENVIRONMENT

3.1 General overview

MetaTutor is a multi-agent, adaptive hypermedia learning environment, which presents challenging human biology science content. The primary goal underlying this environment is to investigate how multi-agent system can adaptively scaffold SRL and metacognition within the context of learning about complex biological content. MetaTutor is grounded in a theory of SRL that views learning as an active, constructive process whereby learners set goals for their learning and then attempt to monitor, regulate, and control their cognitive and metacognitive processes in the service of those goals [6; 3; 2]. More specifically, MetaTutor is based on several theoretical assumptions of SRL that emphasize the role of cognitive, metacognitive (where metacognition is conceptualized as being subsumed under SRL), motivational, and affective processes [19; 22]. Moreover, learners must regulate their cognitive and metacognitive processes in order to integrate multiple informational representations available from the system. While all students have the potential to regulate, few students do so effectively, possibly due to inefficient or a lack of cognitive or metacognitive strategies, knowledge, or control.

As a learning tool, MetaTutor has a multitude of features that embody and foster self-regulated learning (*cf.* Figure 1). These include four pedagogical agents which guide students through the learning session and prompt students to engage in planning, monitoring, and strategic learning behaviors. In addition, the agents can provide feedback and engage in a tutorial dialogue in an attempt to scaffold students' selection of appropriate sub-goals, accuracy of metacognitive judgments, and use of particular learning strategies. The system also uses natural language processing to allow learners to express metacognitive monitoring and control processes. For example, learners can type that they do not understand a paragraph and can also use the interface to summarize a static illustration related to the circulatory system. Additionally, MetaTutor collects information from user interactions with it to provide adaptive feedback on the deployment of students' SRL behaviors. For example, students can be prompted to self-assess their understanding (*i.e.*, system-initiated judgment of learning [JOL]) and are then administered a brief quiz. Results from the self-assessment and quiz allow pedagogical agents to provide adaptive feedback according to the calibration between students' confidence of comprehension and their actual quiz performance.

During learning, MetaTutor is capable of measuring the deployment of self-regulatory processes by allowing us to collect rich, multi-stream data, including: self-report measures of SRL, on-line measures of cognitive and metacognitive processes (*e.g.*, concurrent think-alouds), dialogue moves regarding agent-student interactions, natural language processing of help-seeking behavior, physiological measures of motivation and emotions, emerging patterns of effective problem solving behaviors and strategies, facial data on both basic (*e.g.*, anger) and learning-centered emotions (*e.g.*, boredom), and eye-tracking data regarding the selection, organization, and integration of multiple representations of information (*e.g.*, text, diagrams). The collection of these various data streams is critical to enhancing our understanding of when, how, and why students regulate or do not regulate their learning and adapt their regulatory behaviors. These data are then used to develop computational models designed to detect, track, model, and foster students' SRL processes during learning.

3.2 Self-Regulated Learning with MetaTutor

This paper is theoretically-guided by contemporary models of SRL that emphasize the temporal deployment of these processes during learning [6]. As such, the goal is to use multiple measures to detect, track, and model learners' use of cognitive, affective, and metacognitive (CAM) processes during learning. As such, we use Winne and Hadwin's model [20; 21] because it proposes that learning occurs in four basic phases: (1) task definition, (2) goal-setting and planning, (3) studying tactics, and (4) adaptations to metacognition. Their model emphasizes the role of metacognitive monitoring and control as the central aspects of learners' ability to learn complex material across different instructional contexts (*e.g.*, using a multi-agent system to track and foster SRL) in that information is processed and analyzed within each phase of the model. Recently, Azevedo and colleagues [7; 6; 4; 5; 2] extended this model and provided extensive evidence regarding the role and function of several dozen CAM processes during learning with student-centered learning environments (*e.g.*, multimedia, hypermedia, simulations, intelligent tutoring systems).

In brief, our model makes the following assumptions: (1) successful learning involves having learners monitor and control (regulate) key CAM processes during learning; (2) SRL is context-specific and therefore successful learning may re-
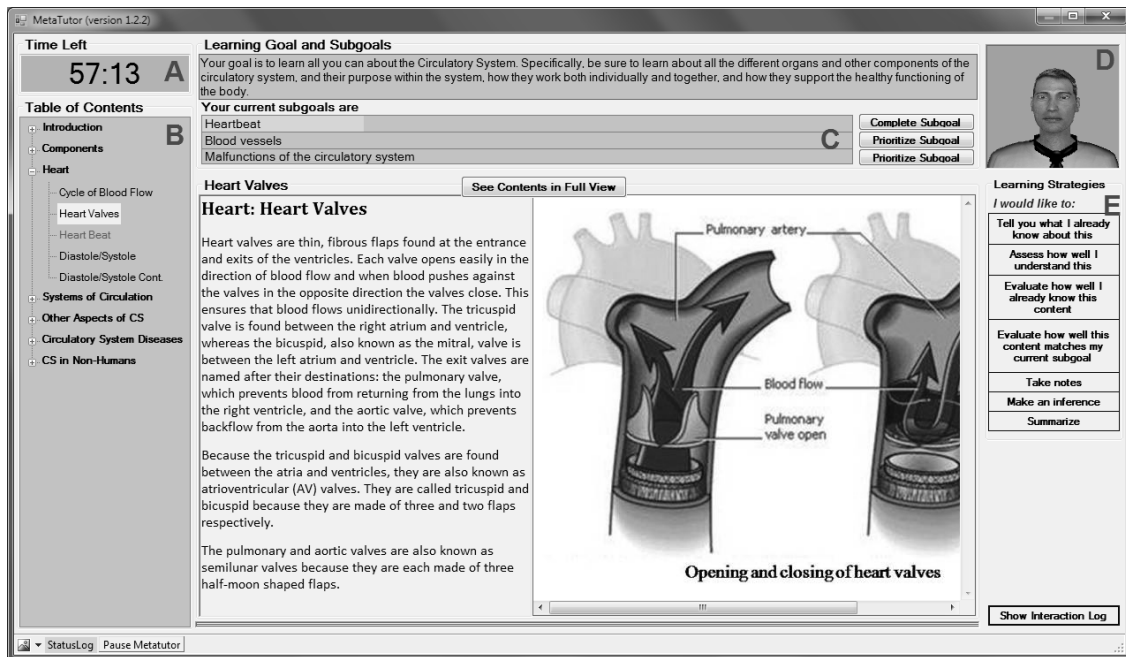
Figure 1: Annotated screenshot of MetaTutor (A: time remaining in the session, B: table of contents, C: current subgoals and progression, D: embodied pedagogical agent, E: palette of monitoring and strategy actions)

quire a learner to increase/decrease the use of certain key SRL processes at different points in time during learning; (3) a learner's ability to monitor and control both internal (*e.g.*, prior knowledge) and external factors (*e.g.*, changing dynamics of the learning environment; relative utility of an agent's prompt) are crucial in successful learning; (4) a learner's ability to make adaptive, real-time adjustments to internal and external conditions, based on accurate judgments of their use of CAM processes, is fundamental to successful learning; and; (5) certain CAM processes (*e.g.*, interest, self-efficacy, task value) are necessary to motivate a learner to engage and deploy appropriate CAM processes during learning and problem solving. This model is best suited for this project since it deals specifically with the person-in-context perspective and postulates that CAM processes occur during learning with a multi-agent system, which will be useful in examining when and how learners will regulate their learning about the human circulatory system. As such, the macro-level processes used in this paper are reading, metacognitive monitoring, and learning strategies. Reading behavior is critical since it is the most important activity related to acquiring, comprehending, and using content knowledge related to the science topic. During reading, learners need to monitor and regulate several key processes such as: (1) selecting relevant content (*i.e.*, text and diagrams) based on their current sub-goal; (2) spending appropriate amounts of time on each page, depending on their relevance regarding their current sub-goal; (3) deciding when to switch or create a new sub-goal; (4) making accurate assessments of their emerging understanding; (5) conceptually connecting content with prior knowledge; (6) adaptively selecting, using, and assessing the effective use of several learning strategies including re-reading, coordinating informational sources, summarizing, making inferences, in order to comprehend the material at various levels (*i.e.*, declara-

tive, procedural, and conceptual knowledge); and, (7) making adaptive changes to behavior based on a variety of external (*e.g.*, quiz scores, quality and timing of agents' prompts and feedback) and internal sources (*e.g.*, affective experiences including both positive and negative affective states, perception of task difficulty). In sum, SRL involves the continuous monitoring and regulation of CAM processes during learning with MetaTutor.

3.3 Participants and data collection

While data has been collected over a sample of 148 undergraduate students from two large public universities in North America, we consider for this study only a sub-sample of 51 participants from the experimental condition that included the most prompts from the pedagogical agents to perform SRL actions and in which students were given some adaptive feedback after having performed those actions. Participants from other conditions did not receive a similar experience with the system, and therefore the values of the variables considered (*cf.* section *4.2.1*) were completely different for them (*e.g.* they took less quizzes as they were not prompted to self-regulate their learning). Considered logs contained an average of 1072 events per session ($\sigma = 255$).

4. PRELIMINARY STEPS

4.1 Data preparation, coding and extraction

For the analysis performed here, as justified in section 3.2, we abstracted the set of collected interactions into three broad categories: reading, monitoring, and strategy (*cf.* Table 1 for The detailed list of actions extracted from the data).

4.1.1 Reading

A reading action (*Read*) is coded each time the student clicks to display a new page of content to read. They can be split

according to two combinatorial criteria, r and t, written as $Read_t^r$, where:

- r stands for the relevance of the page with regard to the student's current subgoal ($+$ for a relevant page, $-$ for an irrelevant page, $\varnothing$ if no subgoal is currently set and relevance can't be determined);

- t stands for the time the student spent reading the page (s if they remain less than 15 seconds, threshold under which no SRL prompt can be triggered, l otherwise).

4.1.2 Monitoring

A monitoring action (Mon) is coded when the student performs, or is prompted to perform, a monitoring action with respect to their learning. This monitoring action could be a judgment of learning (JOL) about what they have just read, a feeling of knowing (FOK) regarding the content of the page, an evaluation of the content (CE) relevance with respect to their current subgoal, or an assessment of their progress towards their current subgoal (MPTG). They can also be split according to two combinatorial criteria, e and i, written as Mon_i^e, where:

- $e \in \{+, -, \varnothing\}$ stands for the correctness of the monitoring evaluation performed by the student ($+$ if the evaluation is right, $-$ if it is wrong, $\varnothing$ if no direct evaluation is possible for the monitoring process);

- $i \in \{u, a\}$ stands for the initiator of the action (u for the user, a for the agent).

Following FOKs and JOLs, as well as when the student claims to have finished a subgoal, students are asked to answer a short quiz (of 3 to 10 questions). Those actions, coded as $Quiz$, can be split along one dimension and are then written $Quiz^s$, where $s \in \{+, -\}$ stands for the success or failure to pass the test ($+$ if the student obtained more than 66% of correct answers, $-$ otherwise).

4.1.3 Strategy

A strategy action (Str) is coded when the student uses a strategy to self-regulate their learning, including when the strategy is prompted by the agent, as well as when the user independently decides to perform the action. Strategy actions include a summarization (SUMM) of the page, a coordination of information sources (COIS) by viewing a related image, an inference (INF) regarding the reading material, a re-reading (RR) of a paragraph that was not well understood, or notes taken about the reading material. This action can also be split depending on the initiator of the action, and is then written Str_i, where $i \in \{u, a\}$ as defined in 4.1.2. Moreover, we distinguish a particular strategy consisting of taking or checking notes in the embedded note interface or using the electronic paper-based notepad provided next to the workstation. These note actions are coded as $Notes$.

4.2 User clustering

4.2.1 Methodology

In a previous study [8], we ran a cluster analysis over a subset of 13 variables extracted from the interaction log after the end of the student's learning session: pretest and posttest score, number of subgoal and page quizzes, mean

Table 2: Synthesis of clusters differences (italic means clusters weren't significantly different from one another according to that variable when using an ANOVA with $p < 0.05$)

Variables	Score for each cluster		
	0	1	2
Pretest score	M	L	H
Posttest score	M	L	H
Session duration	*M*	*M*	*M*
Reading duration	H	H	L
Proportion of subgoals attempted	L	L	H
Number of subgoals changes	M	L	H
Number of subgoals quizzes	*M*	*M*	*M*
Mean first score in subgoal quizzes	M	L	H
Number of page quizzes	*M*	*M*	*M*
Mean first score in page quizzes	M	L	H
Number of note taking	H	L	L
Number of note checking	L	L	H
Time spent taking notes	H	L	L

first score in subgoal and page quizzes, proportion of subgoals attempted among the 7 possible, number of subgoals changes, total session duration, time spent reading content, number of times the student took notes and checked notes, and the duration of the note-taking episodes. This analysis empoyed the Expectation-Maximization (EM) algorithm as implemented in the Weka data mining package [10]. The number of categories to find being undetermined a priori, we used a 10-fold cross-validation, during which we incremented the number of clusters (starting with 1) as long as the loglikelihood averaged over the 10 folds was increasing (*i.e.* we stopped as soon as the loglikelihood with N+1 clusters was lower than with N clusters). We used 1000 different initialization seeds for the EM algorithm, in order to compensate for its tendency to get stuck into local optima, and selected, among the 1000 partitions of students generated, the most frequent one among the most frequently obtained number of clusters (3).

4.2.2 Results

Three clusters were obtained, which characteristics are summarized in Table 2, where clusters 0, 1 and 2 are made of 21, 14 and 16 students, respectively. Generally, students from cluster 2 scored high on pretest, posttest and intermediary quizzes, spent less time than others reading while attempting more subgoals, and took less notes and less time taking them. In contrast, students from cluster 1 scored low on pretest, posttest and intermediary quizzes, attempted less subgoals and took few notes and less time to take them. Students from cluster 0 occupied generally a intermediate position in terms of performance and subgoal uses, but took overall more notes and more time to take them. When using a formula derived from [9] to evaluate learning gains (*cf.* [8] for more details), we also found that students from cluster 2 had the most significant knowledge acquisition, as opposed to those in cluster 1. For all those reasons, cluster 1 will

Table 1: List of actions extracted from MetaTutor interaction logs

Category	Action name	Description
Read	$Read_s^+$	Student skims through a page relevant for their current subgoal for less than 15s
	$Read_s^-$	Student skims through a page irrelevant for their current subgoal for less than 15s
	$Read_s^\varnothing$	Student skims through a page without having a subgoal set for less than 15s
	$Read_l^+$	Student reads a page relevant for their current subgoal for more than 15s
	$Read_l^-$	Student reads a page irrelevant for their current subgoal for more than 15s
	$Read_l^\varnothing$	Student reads a page without having a subgoal set for more than 15s
Monitoring	Mon_a^+	Student is prompted to evaluate their knowledge, learning or the relevance of the content they are reading, and evaluates correctly
	Mon_a^-	Student is prompted to evaluate their knowledge, learning or the relevance of the content they are reading, and is wrong in their evaluation
	$Mon_a^\varnothing$	Student is prompted to perform a monitoring action that doesn't require an evaluation
	Mon_u^+	Student takes the initiative of evaluating their knowledge, learning or the relevance of the content they are reading, and evaluates correctly
	Mon_u^-	Student takes the initiative of evaluating their knowledge, learning or the relevance of the content they are reading, and is wrong in their evaluation
	$Mon_u^\varnothing$	Student takes the initiative of performing a monitoring action that doesn't require an evaluation
	$Quiz^+$	Student passes a page or subgoal quiz (more than 66% of correct answers)
	$Quiz^-$	Student fails a page or subgoal quiz (less than 66% of correct answers)
Strategy	Str_a	Student is prompted to deploy a strategy to self-regulate
	Str_u	Student takes the initiative of using a strategy to self-regulate
	$Notes$	Student takes or checks notes using the embedded interface or a paper-based electronic notepad

be referred to as cluster L (for low), cluster 2 as cluster H (for high) and cluster 0 as cluster M (for medium). The fact that exactly three (as opposed to any other number) clusters were extracted might sound unsurprising, but comes from the fact that it was the best partition of the subjects in the 13-dimension space considered.

5. DIFFERENTIAL SEQUENCE MINING

5.1 Method principles

To identify important activity patterns in a comparison between student clusters, we employ a differential sequence mining technique [11]. This technique uses sequence mining and two different measures of pattern frequency to identify *differentially frequent* patterns between two sets of action sequences. Differential sequence mining combines frequency measures and techniques from sequential pattern mining [1], which determines the most frequent action patterns across a set of action sequences, and episode mining [12], which determines the most frequently used action patterns within a given sequence.

The sequential pattern mining frequency measure (*i.e.*, how many sequences/students exhibit the given pattern) is used to identify patterns common to a group of students. We refer to this as the "sequence support" (*s-support*) of the pattern, and we call patterns meeting a given s-support threshold *s-frequent*. In this analysis, we employ an s-support threshold of 0.5 to focus on patterns exhibited by at least half of a given group of students. The episode mining frequency (*i.e.*, the frequency with which the pattern is repeated *within* an action sequence) is important for assessing the extent to which a student relies on a particular pattern of activities. For a given student, we refer to this as the "instance support" (*i-support*), and we call patterns meeting a given i-support threshold *i-frequent*. To calculate the i-support of a pattern for a group of students, we use the mean of the pattern's i-support values across all traces in the group.

The differential sequence mining technique first uses a sequential pattern mining algorithm to identify the patterns that meet a minimum s-support constraint within each group [11]. To compare the identified frequent patterns across groups, we calculate the i-support of each pattern for each student (in each group). Using a t-test, we filter the s-frequent patterns to identify those for which there is a statistically significant difference in i-support values between groups. Comparing the mean i-support value for each pattern between groups then allows us to focus the comparison on patterns that are employed significantly more often by one group than the other.

This comparison produces four distinct categories of frequent patterns: two categories where the patterns are s-frequent in only one group, illustrating patterns primarily employed by the respective groups, and two categories where the patterns are common to both groups but used significantly more often in one group than the other. The patterns in each of these qualitatively distinct categories are (separately) sorted by the difference in mean group i-support[1] to focus the analysis on the most differentially frequent patterns [11].

5.2 Application to the data

In order to identify patterns more closely related to changes in students' knowledge and understanding, we decided to focus mainly on clusters H and L, as defined in section 4.2.2. Moreover, to further identify the patterns most characteristic of students in cluster H (*resp.* L) we identified differentially frequent patterns with respect to the other two

---

[1]Even though a pattern may not be s-frequent in a group of action sequences, it can still occur in some sequences in the group, so an i-support value can be calculated (or the i-support is 0 if the pattern does not occur in any trace in the group).

clusters M and L (*resp.* M and H) in a secondary analysis. We employed an s-support threshold of 50% in this analysis, to consider all the patterns that were exhibited by at least half of the students in a given cluster, and a standard value of 0.05 for the t-test cutoff p value. We tried to preliminarily group sequences of identical actions together, but the results obtained were not very different from the ones without grouping, as the data extracted do not display long sequences of similar actions – therefore, those results are not reported here. Similarly, although we also considered the possibility of using gaps of one or more actions when identifying patterns, we discarded this analysis because the frequency of events collected in the log is low, which means that even a gap of only one action could mean that two actions of a pattern are actually separated by a rather long period of inactivity.

5.3 Results

The Table 3 displays the patterns with the highest difference of S-support between clusters H and L (positive value in column 3) as well as between clusters L and H (negative value in column 3), provided that difference is statistically significant (*i.e.* a t-test p value below 0.05 in column 4). It also displays a selection of interesting patterns, which differed in a statistically significant way between the two clusters. Columns 6 to 11 provide the results obtained for that selection of patterns using two different samples of students: first (columns 6 to 8), cluster H alone and a merge of clusters L and M, and then (columns 9 to 11), cluster L alone and a merge of cluster H and M. Columns 5, 8 and 11 show, for the two considered samples, if only one or both of them were having a s-support above 50% for the considered pattern. Values N/A are used when the pattern is non statistically significant for the two considered samples.

The following observations can be made:

– According to *pattern 1*, when prompted to use a strategy (regardless of the one suggested by the agent), students in cluster H reacted by taking notes more often than students in cluster L. We already knew that students in cluster H had received significantly more prompts from the system, and taken less notes overall than those in cluster L (but checked them more often). This pattern seems to suggest that the reason might be that the notes they were taking mainly came from prompts from the agents. Moreover, since when they type a summary, students are offered the possibility to add it to their notes, it appears that students from cluster H must have preferred that strategy, which also would explain why they spent less time with the note-taking interface open (since the summary is typed in a different text box, and the note-taking interface is opened only to add the already typed text). Finally, the fact that the difference for this pattern is significant for cluster H *vs.* L, H *vs.* M&L and H&M *vs.* L indicate that the degree to which one relies on the prompt for notes or summaries to take notes is directly correlated to the belonging to one of the three clusters (*i.e.* this behavior is observed more in cluster H than in M, and more in M than in L).

Similarly, *pattern 3* indicates that after a note-taking event, students from cluster H often moved on to another relevant page, which they read for an extended period. *Pattern 5*, which is a combination of patterns 1 and 3, confirm the idea that students from cluster H had a very methodical approach to navigating through the content: they selected a relevant page, read it until being prompted by the agent to take notes or summarize it, performed that action, and then moved on to a new relevant page. Incidentally, it also indicates their effectiveness in identifying a page relevant to their current subgoal simply from its title (since that is all they can see before opening it). This latter hypothesis is itself reinforced by the observation that *patterns 10 and 11*, relative to a brief visit on an irrelevant page or to a succession of brief visits to irrelevant pages, is characteristic of students from clusters M and L, as opposed to students from cluster H who seem to not even need to open the pages to figure out they are irrelevant to their current subgoal.

– *Pattern 2* simply confirms what we already knew about the tendency of student in cluster H to have answered correctly more often to intermediate quizzes (for a page or a subgoal). It also significantly distinguish members of cluster H from those in cluster M&L considered together.

– *Patterns 4 & 7* are relative to pages viewed when the students did not have any active subgoal set. Pattern 4 indicates that students in the cluster H have visited more pages for a long time without having a subgoal set, which is confirmed by pattern 7 which also indicate an alternation between short and long reads when no subgoals were set. As we also know that students from cluster H attempted more subgoals overall than students in the cluster L, it cannot mean that they have simply refused to set additional subgoals once they had finished their original ones (e.g., in an attempt to get rid from the system prompts and feedback), but rather that: a) they might have spent some time reviewing pages already read before taking the posttest, and/or b) instead of setting a final subgoal when they did not have much time left, they took some time to review the pages they had not yet explored.

This hypothesis can be confirmed by looking at the temporal distribution of those two patterns: for students in cluster H, the median time is of 108 and 112 minutes (for an overall session of approximately 120 minutes), which means that it's during the last 15 minutes of their learning session that students were displaying that kind of browsing behavior, clearly distinct from the ones they had displayed earlier in the session.

– *Pattern 6* indicates that students in cluster H seemed to more often estimate properly their level of understanding of the content or the relevance of the page they were visiting when it was relevant for their current subgoal. While this pattern is only marginally significant when comparing clusters H and L, it is statistically significant when comparing H to M&L, confirming that it is specific of students in cluster H. It tends to show that not only other students had difficulties to identify the relevance of a page from its title, but that even once they had been able to spend some time reading its content, they were less prone to correctly evaluate its relevance or their understanding of it.

This hypothesis seems to be confirmed by the complementary *pattern 8*, which indicates that students from cluster L, when they were on a page irrelevant for their subgoal for a long time and got prompted to evaluate its relevance (the only prompt they can get on a non-relevant page), tended to be wrong in their evaluation.

If we consider again the temporal distribution of those two patterns, we can notice that the median time, for students in cluster L, is of 50 and 45 minutes, *i.e.* less than the median time of the session (60 minutes). We can therefore assume

Table 3: Significant and most frequent patterns differentiating clusters

#	Pattern	Cluster H vs. L			Cluster H vs. M&L			Cluster H&M vs. L		
		I-Supp. Diff	t-test (p value)	S-Freq. Cluster	I-Supp. Diff	t-test (p value)	S-Freq. Cluster	I-Supp. Diff	t-test (p value)	S-Freq. Cluster
1	$Str_a \prec Notes$	**3.93**	**0.002**	**Both**	**3.28**	**0.005**	**Both**	**2.30**	**0.007**	**Both**
2	$Quiz^+$	**3.10**	**0.036**	**Both**	**2.09**	**0.046**	**Both**	2.30	0.086	Both
3	$Notes \prec Read_l^+$	**2.86**	**0.004**	**Both**	**2.35**	**0.012**	**Both**	**1.71**	**0.012**	**Both**
4	$Read_l^{\varnothing}$	**2.63**	**0.039**	**H**	**2.27**	**0.050**	**H**	1.48	0.107	H&M
5	$Str_a \prec Notes \prec Read_l^+$	**2.38**	**0.001**	**Both**	N/A	N/A	N/A	**1.27**	**0.017**	**Both**
6	$Read_l^+ \prec Mon_a^+$	1.96	0.065	Both	**1.96**	**0.048**	**Both**	0.85	0.304	Both
7	$Read_s^{\varnothing} \prec Read_l^{\varnothing}$	**1.33**	**0.050**	**H**	1.23	0.061	H	N/A	N/A	N/A
. . .										
8	$Read_l^- \prec Mon_a^-$	**-0.54**	**0.039**	**L**	N/A	N/A	N/A	-0.25	0.360	L
9	$Read_s^+ \prec Read_l^-$	**-0.65**	**0.012**	**L**	N/A	N/A	N/A	**-0.53**	**0.038**	**L**
10	$Read_s^- \prec Read_s^-$	N/A	N/A	N/A	**-1.77**	**0.030**	**M&L**	N/A	N/A	N/A
11	$Read_s^-$	-3.49	0.149	Both	**-2.56**	**0.036**	**Both**	-2.39	0.321	Both

that, at least, students from cluster L have been slightly improving their capacity to evaluate their learning and the relevance of a page over time.

– *Pattern 9* confirms the previous observation that students in cluster L really had issues to see the relevance of a page with regard to their subgoal: they did not simply end up going to random pages that were irrelevant to their subgoal, or ignored the subgoal they had set, but instead, they appeared to sometimes skim through a relevant page, miss its relevance, and end up instead spending a long time on a page that wasn't irrelevant to their subgoal. This tendency is shared, to some extent, with students from cluster M, as the results of clusters H *vs.* M&L are also statistically significant.

– A final observation can be made regarding the tendency of a student to obey system prompts: if we run the same analysis without distinguishing the correctness of the evaluation of students monitoring (*i.e.* by considering actions $Mon_a = Mon_a^+ \cup Mon_a^-$ and $Mon_u = Mon_u^+ \cup Mon_u^-$), we observe that the pattern $Mon_a \prec Mon_u$ is significantly more frequent for students in cluster H, which tends to indicate that when prompted to perform an optional monitoring action (most likely, a MPTG, since otherwise there should be a *Quiz* action following the Mon_a), they are more prone to accomplish the suggested action.

6. DISCUSSION

To summarize the results obtained in the previous section, we can conclude that students from cluster H are more inclined to follow the system prompts and to follow the suggestions to take notes or summarize what they have just learned. Further, they are more prone to keep applying the same method for each page they read, are better at identifying a page relevant to their subgoal from its title, and are more strategic in their preparation for the posttest (*e.g.*, they usually use their last 10 to 15 minutes to briefly review various pages). From an ITS design point of view, the fact that these students used system prompts to effectively regulate their learning tends to indicate that the frequency of Strategy prompts should probably not be reduced. However, as they seem good at distinguishing relevant pages from irrelevant ones, they might need less scaffolding regarding the Monitoring processes. On the other hand, students from cluster L appear particularly unable to identify pages rele-

vant to their subgoal, which is probably linked to their lower prior knowledge. For them, it seems that additional scaffolding from the system would certainly be beneficial. However, even when prompted to monitor their learning, they tend to be mistaken in their evaluation. Therefore, it could be necessary to go further than the methods currently employed to suggest ways in which they can better evaluate the relevance of a page.

7. CONCLUSION, FUTURE DIRECTIONS

In this paper, we have presented a two-step analysis of data collected with an ITS designed to foster self-regulated learning. First, the clustering of students using Expectation Maximization has allowed us to distinguish three clusters of students with different prior knowledge on the topic, learning performance, and strategies. We then described a set of actions extracted from the system interaction trace log and employed a sequence mining technique to identify differentially frequent activity patterns. We used the identified patterns to characterize students from different clusters with particular emphasis on those that had the highest and the lowest learning gains. We have been able to identify patterns of actions that suggest students with high prior knowledge and high learning gains tended to be more compliant with system prompts, using them to validate their progression. Further, these students were better at identifying pages relevant to their subgoals from the page title and tended to have a phase at the end of the session during which they reviewed the content in preparation for the posttest.

The analysis performed here will allow us to more accurately identify the group to which a student belongs during their use of MetaTutor and dynamically adapt the scaffolding and feedback mechanisms accordingly. Another future research direction will involve the use of other channels of data collected while students use MetaTutor (eye-tracking information, affective data extracted from video captures, and think-aloud data) in order to enhance our identification and understanding of phases when low-performing students are unable to properly monitor their learning.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] R. Agrawal and R. Srikant. Mining sequential patterns. In *Proceedings of the Eleventh IEEE International Conference on Data Engineering (ICDE)*, pages 3–14, 1995.

[2] R. Azevedo, R. Behnagh, M. Duffy, J. M. Harley, and G. J. Trevors. Metacognition and self-regulated learning in student-centered leaning environments. In D. Jonassen and S. Land, editors, *Theoretical foundations of student-center learning environments*, pages 216–260. Erlbaum, Mahwah, NJ, 2nd edition, 2012.

[3] R. Azevedo, J. G. Cromley, D. C. Moos, J. A. Greene, and F. I. Winters. Adaptive content and process scaffolding: A key to facilitating students' self-regulated learning with hypermedia. *Psychological Testing and Assessment Modeling*, 53:106–140, 2011.

[4] R. Azevedo and R. Feyzi-Behnagh. Dysregulated learning with advanced learning technologies. *Journal of e-Learning and Knowledge Society*, 7(2):9–18, 2011.

[5] R. Azevedo, A. Johnson, A. Chauncey, and A. C. Graesser. Use of hypermedia to convey and assess self-regulated learning. In B. Zimmerman and D. Schunk, editors, *Handbook of self-regulation of learning and performance*, pages 102–121. Routledge, New York, 2011.

[6] R. Azevedo, D. C. Moos, A. Johnson, and A. Chauncey. Measuring cognitive and metacognitive regulatory processes used during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4):210–223, 2010.

[7] R. Azevedo and A. M. Witherspoon. Self-regulated learning with hypermedia. In D. J. Hacker, J. Dunlosky, and A. C. Graesser, editors, *Handbook of metacognition in education*, pages 319–339. Routledge, Mahwah, NJ, 2009.

[8] F. Bouchet, J. M. Harley, G. J. Trevors, and R. Azevedo. Clustering and profiling students according to their interactions with an intelligent tutoring system fostering Self-Regulated learning. *Journal of Educational Data Mining*, (submitted), 2012.

[9] S. D'Mello, B. Lehman, J. Sullins, R. Daigle, R. Combs, K. Vogt, L. Perkins, and A. Graesser. A time for emoting: When Affect-Sensitivity is and isn't effective at promoting deep learning. In V. Aleven, J. Kay, and J. Mostow, editors, *Intelligent Tutoring Systems*, volume 6094, pages 245–254. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18, Nov. 2009.

[11] J. S. Kinnebrew, K. M. Loretz, and G. Biswas. A differential sequence mining method contextualized by student performance evolution to derive learning behavior patterns. *Journal of Educational Data Mining*, (submitted), 2012.

[12] H. Mannila, H. Toivonen, and A. Inkeri Verkamo. Discovery of frequent episodes in event sequences. *Data Mining and Knowledge Discovery*, 1(3):259–289, 1997.

[13] R. Martinez, K. Yacef, J. Kay, A. Al-Qaraghuli, and A. Kharrufa. Analysing frequent sequential patterns of collaborative learning activity around an interactive tabletop. In *Proceedings of the Fourth International Conference on Educational Data Mining.*, Eindhoven, Netherlands, 2011.

[14] J. C. Nesbit, Z. Mingming, Y. Xu, and P. H. Winne. Advancing log analysis of student interactions with cognitive tools. In *12th Biennial Conference of the European Association for Research on Learning and Insruction (EARLI)*, Budapest, Hungary, 2007.

[15] D. Perera, J. Kay, I. Koprinska, K. Yacef, and O. R. Zaane. Clustering and sequential pattern mining of online collaborative learning data. *IEEE Trans. on Knowl. and Data Eng.*, 21(6):759–772, June 2009.

[16] C. Romero, S. Ventura, A. Zafra, and P. d. Bra. Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems. *Computers & Education*, 53(3):828–840, Nov. 2009.

[17] R. Srikant and R. Agrawal. Mining sequential patterns: Generalizations and performance improvements. In P. Apers, M. Bouzeghoub, and G. Gardarin, editors, *Advances in Database Technology - EDBT'96*, volume 1057, pages 1–17. Springer-Verlag, Berlin/Heidelberg, 1996.

[18] T. Y. Tang and G. McCalla. Student modeling for a web-based learning environment: a data mining approach. In *Proc. of the 18th national conference on Artificial intelligence*, page 967968, Menlo Park, CA, USA, 2002. American Association for Artificial Intelligence.

[19] P. H. Winne. Self-regulated learning viewed from models of information processing. In B. Zimmerman and D. Schunk, editors, *Self-regulated learning and academic achievement: Theoretical perspectives*, pages 153–189. Erlbaum, Mahwah, NJ, 2001.

[20] P. H. Winne and A. Hadwin. Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, and A. C. Graesser, editors, *Metacognition in educational theory and practice*, pages 277–304. Erlbaum, Hillsdale, NJ, 1998.

[21] P. H. Winne and A. Hadwin. The weave of motivation and self-regulated learning. In D. Schunk and B. Zimmerman, editors, *Motivation and self-regulated learning: Theory, research and applications*, pages 297–314. Routledge, 2008.

[22] B. Zimmerman and D. Schunk, editors. *Handbook of self-regulation of learning and performance*. Routledge, New York, 2011.