Disengagement Detection in Online Learning: Validation Studies and Perspectives

Mihaela Cocea and Stephan Weibelzahl

Abstract—Learning environments aim to deliver efficacious instruction, but rarely take into consideration the motivational factors involved in the learning process. However, motivational aspects like engagement play an important role in effective learning—engaged learners gain more. E-Learning systems could be improved by tracking students' disengagement that, in turn, would allow personalized interventions at appropriate times in order to reengage students. This idea has been exploited several times for Intelligent Tutoring Systems, but not yet in other types of learning environments that are less structured. To address this gap, our research looks at online learning-content-delivery systems using educational data mining techniques. Previously, several attributes relevant for disengagement prediction were identified by means of log-file analysis on HTML-Tutor, a web-based learning environment. In this paper, we investigate the extendibility of our approach to other systems by studying the relevance of these attributes for predicting disengagement in a different e-learning system. To this end, two validation studies were conducted indicating that the previously identified attributes are pertinent for disengagement prediction, and two new meta-attributes derived from log-data observations improve prediction and may potentially be used for automatic log-file annotation.

Index Terms—e-Learning, educational data mining, disengagement prediction, log-file analysis.

✦

---

1 INTRODUCTION

EDUCATIONAL software strives to meet the learners' needs and preferences in order to make learning more efficient; the complexity is considerable and many aspects are taken into consideration. However, most systems do not consider the learner's motivation for tailoring teaching strategies and content, despite its great impact on learning being generally acknowledged. A lack of motivation is clearly correlated with learning rate decrease (e.g., [1]).

A number of attempts have been undertaken to accommodate the learner's motivational states, mostly by means of design. E-learning systems attempted to motivate students through an attractive design by using multimedia materials or including game features that have great potential [2] and have been proved successful in a number of cases (e.g., [3]). Despite these efforts, students are not always focused on learning and even try to game the systems (attempting to succeed in an educational environment by exploiting properties of the system's help and feedback rather than by attempting to learn the material) [1].

Learner's self-assessment has been used for a long time in classroom context, and recently also in e-learning, where it has been proved to be reliable, and a valuable and accurate source of motivational information [4].

However, to effectively address the motivational factors that influence learning they need to be assessed for each individual to allow personalized interventions based on this assessment. To do this efficiently, automatic analysis is necessary.

The learner's actions preserved in log files have been relatively recently discovered as a valuable source of information and several approaches to motivation detection and intervention have used log-file analysis. An important advantage of log-file analysis over self-assessment approaches is the unobtrusiveness of the assessment process, similar to the classroom situation where a teacher observes that a learner is not motivated without interrupting his/her activities.

Several efforts to detect motivational aspects from learners' actions are reported in the literature [1], [5], [6], [7], [8], [9], [10], [11]. However, all these efforts are concentrated on Intelligent Tutoring Systems or problem-solving environments. As online content-delivery systems are increasingly used in formal education, there is a need to extend this research to encompass this type of systems as well. The interaction in these systems is less constrained and structured compared with problem-solving environments, posing several difficulties to an automatic analysis of learners' activity.

To address this challenge, we restricted our research to one motivational aspect, disengagement, and looked at identifying the relevant information from learners' actions to be used for its prediction. Being able to automatically detect disengaged learners would offer the opportunity to make online learning more efficient, enabling tutors and systems to target disengaged learners, to reengage them, and thus, to reduce attrition.

Analyzing data from a web-based interactive environment, HTML-Tutor, we identified six relevant attributes by means of educational data mining techniques [12] to predict whether a learner is disengaged. In this paper, we investigate the extendibility of our approach to other systems by studying the relevance of these attributes for predicting disengagement in a different e-learning system.

- *M. Cocea is with the London Knowledge Lab, Department of Computer Science and Information Systems, Birkbeck College, University of London, 23-29 Emerald Street, WC1N 3QS London. E-mail: mihaela@dcs.bbk.ac.uk.*
- *S. Weibelzahl is with the School of Computing, National College of Ireland, Mayor Street, IFSC, Dublin 1, Ireland. E-mail: sweibelzahl@ncirl.ie.*

We demonstrate that the same attributes can be used for disengagement prediction in the second system, yielding similar information gain.

The rest of the paper is structured as follows: In Section 2, previous work related to motivation and engagement prediction is presented. Section 3 briefly presents the log-file analysis performed on HTML-Tutor data by which the relevant attributes for disengagement prediction were identified. Section 4 includes the two validation studies conducted on iHelp data, and Section 5 discusses the results and implications of the validation studies and relates our outcomes with the previous approaches to engagement prediction. Section 6 discusses several perspectives on the outcomes of this research and its possible impact, and concludes the paper.

2 RELATED RESEARCH

Before presenting related research on detection of motivational aspects, a brief outline is given on how engagement is related to other motivational concepts.

Motivational research [13] makes uses of several concepts, besides motivation itself: engagement, interest, effort, focus of attention, self-efficacy, confidence, etc. The research presented in this paper focuses on engagement, or rather on disengagement, as an undesirable motivation state. For our purposes, a student is considered to be engaged if she/he is focused on the current learning activity and disengaged otherwise. A number of concepts in motivational research such as interest, effort, focus of attention, and motivation are related, though not identical, to engagement (see, e.g., [13]):

1. Engagement can be influenced by interest, as people tend to be more engaged in activities they are interested in; thus, interest is a determinant of engagement.
2. Effort is closely related to interest in the same way: more effort is invested if the person has interest in the activity. The relation between engagement and effort can be resumed by: engagement can be present with or without effort; if the activity is pleasant (and/or easy), engagement is possible without effort; in the case of more unpleasant (and/or difficult) activities, effort may be required to stay engaged.
3. The difference between engagement and focus of attention, as used in research, is that focus of attention refers to attention through a specific sensorial channel (e.g., visual focus), while engagement refers to the entire mental activity (involving at the same time perception, attention, reasoning, volition, and emotions).
4. Engagement is just one aspect indicating that for a reason or another, the person is motivated to do the activity she/he is engaged in, or, on the contrary, if the person is disengaged, that she/he may not be motivated to do the activity. In other words, engagement is an indicator of motivation.

Although there are several approaches to motivational issues in e-learning, we restrict our review to those that are related to detection of motivational aspects in general and engagement in particular, by means of using learners' actions.

Several approaches for motivation detection from learner's interactions with the e-learning system have been proposed ranging from rule-based approaches to Bayesian networks.

A rule-based approach based on the Attention, Relevance, Confidence, and Satisfaction (ARCS) Model [14] has been developed [5] to infer motivational states from the learners' behavior using a 10-question quiz. A number of 85 inference rules were produced by the participants who had access to replays of the learners' interactions with the system and to the learners' motivational traits.

Another approach [8], also based on the ARCS Model, infers three aspects of motivation: confidence, confusion, and effort, from the learner's focus of attention and inputs related to learners' actions: the time to perform the task, the time to read the paragraph related to the task, the time for the learner to decide how to perform the task, the time when the learner starts/finishes the task, the number of tasks the learner has finished with respect to the current plan (progress), the number of unexpected tasks performed by the learner which are not included in the current learning plan, and the number of questions asking for help.

Engagement tracing [6] is an approach based on Item Response Theory that proposes the estimation of the probability of a correct response given a specific response time for modeling disengagement; two methods of generating responses are assumed: blindly guess when the student is disengaged and an answer with a certain probability of being correct when the student is engaged. The model also takes into account individual differences in reading speed and level of knowledge.

A dynamic mixture model combining a hidden Markov model with Item Response Theory was proposed in [9]. The dynamic mixture model takes into account: student proficiency, motivation, evidence of motivation, and a student's response to a problem. The motivation variable can have three values: 1) motivated; 2) unmotivated and exhausting all the hints in order to reach the final one that gives the correct answer: unmotivated-hint; and 3) unmotivated and quickly guessing answers to find the correct answer: unmotivated guess.

A Bayesian Network has been developed [7] from log data in order to infer variables related to learning and attitudes toward the tutor and the system. The log data registered variables like problem-solving time, mistakes, and help requests.

A latent response model [1] was proposed for identifying the students that game the system. Using a pretest-posttest approach, the gaming behavior was classified in two categories: 1) with no impact on learning and 2) with decrease in learning gain. The variables used in the model were: student's actions and probabilistic information about the student's prior skills.

The same problem of gaming behavior was addressed in [10], an approach that combines classroom observations with logged actions in order to detect gaming behavior manifested by guessing and checking or hint/help abuse. In order to prevent this gaming behavior, two active interventions (one for each type of gaming behavior) and a passive strategy have been proposed [11]. When a student was detected to manifest one of the two gaming behaviors, a message was displayed to the student encouraging him/her to try harder, ask the teacher for help or pursue other suitable actions. The passive strategy had no triggering mechanism, but merely provided visual feedback on students' actions and progress.

Fig. 1. Screenshot of HTML-Tutor from XHTML topic.

This was continuously displayed on screen and available for viewing by the student and the teacher.

Besides detection of motivational-affective states from log data, there is a lot of research in the area focusing on a variety of aspects related to the role of motivation and affect in learning. For example, the use of pedagogical agents and their impact on the learners' affective states and learning are investigated in [15], [16], and [17]; intervention strategies to reengage students or change their affective state are designed and tested in [17], [18], and [19]; several cognitive-affective states are measured and their relation to learning is investigated in [16], [19], [20], [21], [22], and [23]. Different sources are used to diagnose motivational and affective states: gross body language [24], physiological data, users' goals and actions and environmental information [25], human observations, test scores, and log data [22]. These aspects are investigated in a variety of learning environments, such as Intelligent Tutoring Systems [16], [17], [18], [20], [24], educational games [21], [20], [25], programming environments [22], simulation problem-solving environments [23], Vygotskyan learning environments [19], and narrative-centered learning environments [26].

3 ENGAGEMENT PREDICTION FROM LOG FILES

Our approach is different from the previous ones in the fact that it envisages prediction of engagement from both main activities encountered in e-learning systems: reading and problem-solving activities. The two models based on IRT presented in the previous section work very well for problem-solving activities, but they have the disadvantage of considering engagement after the learning activity. Tracking engagement while the student is learning, e.g., reading pages, allows intervention at appropriate times and before the self-evaluation of learning (problem solving), when bad performance could be caused by disengagement

in answering the questions, but also by disengagement during learning time.

In previous research [12], we proposed a different approach to engagement prediction that would cover both general learning as well as problem-solving activities typically encountered in e-learning systems. Such an approach would widen the applicability of the detection mechanism from the rather specific problem-solving activities to all types of e-learning systems that involve learning activities such as reading text and answering quizzes. However, we did not consider collaborative learning behavior or learning based on interactive multimedia such as animations as such features were not present in the analyzed systems.

We analyzed log files from HTML-Tutor—a web-based interactive learning environment based on NetCoach [27]. The course content is written in German and is organized in seven high-level topics on HTML, e.g., hyperlinks, layout, XML, etc. In the screenshot displayed in Fig. 1, these topics are listed in the left side of the screen. Each high-level topic includes several subtopics that may contain one or more items. Each component of this hierarchy links to a file that is displayed in the central area of the screen. A navigation bar is also present at the top of this central area. The top of the screen includes a toolbar with several icons linking to: a manual on how to use the system, communication tools, frequently asked questions, preferences on the display of information on the screen, a glossary, a notes tool, and statistics tool about the personal usage of the system (e.g., coverage of topics and performance on tests).

The purpose of the analysis on the HTML-Tutor log data was twofold: 1) to identify attributes that are relevant for prediction and 2) to explore several prediction methods, mainly as a consistency check and second as a way to identify a best performing method (should it be the case). Consequently, three data sets were used to control the

contribution of attributes and eight prediction methods were employed to check the consistency of prediction.

Log files of 48 users were collected. These users spent between one and seven sessions, where a session is marked by login and logout. A pilot study [28] revealed that using sessions as units of analysis leaves no time for intervention to reengage students as disengagement could be detected only after 45 minutes of activity and most disengaged students would log out before that time. To overcome this problem, sessions were divided in sequences of 10 minutes. From this process, 1,015 sequences were obtained: 943 sequences of exactly 10 minutes and 72 sequences varying between 7 and 592 seconds.

From the 14 logged events, a total of 30 different attributes were derived. Two events—reading pages and taking tests—occurred considerably more often than all the others, with a frequency of occurrence of 850 and 458, respectively, out of a total of 1,015 sequences. Two other events—hyperlinks and glossary—were noticeably more frequent than the rest, with a frequency of 245 and 76, respectively, while the remaining 10 events were rare (with an average of 16 occurrences in 1,015 sequences). A few examples of these less frequent events are preferences, search, and statistics. For a complete list of frequencies of all events, see [12].

Based on the frequency of events, three data sets were defined: one that included attributes of all events, one that included the attributes of the four most frequent events, and one that included only the two most frequent events. By doing this, we aimed to identify the relevant features, taking into consideration the sparsity of data at the same time.

Eight methods that were applicable to our data were employed [29], [30]:

1. Bayesian Nets with K2 algorithm and maximum three parent nodes (BNs).
2. Logistic regression (LR).
3. Simple logistic classification (SL).
4. Instance-based classification with IBk algorithm (IBk).
5. Attribute Selected Classification using J48 classifier and Best First search (ASC).
6. Bagging using REP (reduced error pruning) tree classifier (B).
7. Classification via Regression (CvR).
8. Decision Trees with J48 classifier based on Quilan's C4.5 algorithm [21] (DTs).

Out of the total of 30 attributes, we list only those that refer to the two most frequent events, i.e., accessing pages and taking tests: number of pages, average time spent on pages, number of tests, average time spent on tests, number of correctly answered tests, and number of incorrectly answered tests. Attributes that refer to other events are similar, typically including the frequency of access and the average time. For a complete list of attributes, see [12].

Each sequence was labeled as engaged, disengaged, or neutral. Three human experts (designated as raters) were involved: rater 1 labeled all sequences, while raters 2 and 3 participated in a coding reliability study (more details in [6]). All raters used the unprocessed log files divided in sequences of 10 minutes containing all events. The output of the reliability study was a 92 percent agreement between raters, a Cohen's kappa [31] measurement of agreement of 0.83 ($p < 0.01$), and a Krippendorff's alpha [32] of 0.84,

suggesting the annotation of sequences was conducted in a reliable fashion [33].

The raters considered a learner to be engaged when the logged data showed that users were focused on reading pages, taking tests or both, as well as performing other actions such as searching, looking at statistics, or consulting the glossary, and spending a reasonable time on these actions. A learner was considered to be disengaged when they were browsing quickly through pages or when spending a long time on the same page or test. The neutral label covered situations when the raters could not choose between engaged and disengaged such as when the learner seemed to be engaged for half the time and disengaged for the other half.

Despite these efforts to achieve high reliability and validity of the ratings, the nature of this study implies that the ratings may not always reflect the actual engagement of the learners as raters did not get the opportunity to observe learners' facial expression, gesture, or posture and had to base their judgement purely on behavior records.

The results showed small variation of prediction values across methods and between the three data sets. Two indicators were especially considered: accuracy (the percentage of correct predictions), as an indication of the quality of prediction across all classes (engaged, disengaged, and neutral), and true positive (TP) rate for the disengaged class as an indication of the extent of correct identification of disengaged learners. To give a complete picture and a grasp of the real meaning of the data, other indicators are included: the false positives (FPs) rate for disengaged class, the precision indicator ($TP/(TP + FP)$) for disengaged class, and the mean absolute error. In our context, TP rate is more important than precision because it indicates the correct percentage from actual instances of a class, while precision indicates the correct percentage from predicted instances of that class.

Waikato Environment for Knowledge Analysis (WEKA) [29] was used to perform the analysis. Only sequences of exactly 10 minutes were used and from the 943 entries, 679 (72 percent) were used for training and 264 (28 percent) for testing. The distribution of students within the sets was controlled to avoid having sequences from the same user both in training and testing sets, which could have introduced a positive bias to the results.

Across methods, the prediction values varied between 84.85 percent (using IBk on third data set) and 92.80 percent (using CvR on first data set) accuracy. The variation of the true positive rate for the disengaged class was even smaller: between 0.91 and 0.96 (across all data sets and methods). Using the average across methods, the three data sets were compared: the first data set performed best, with an average of 0.90 percent better accuracy than the second data set and an average of 1.38 percent better than the third data set; the second data set performed better than the third data set by 0.48 percent. The average variation of the true positive rate across data sets was negligible—less than 0.005. Given these relatively small variations and taking into consideration factors like sparsity of data and computational complexity, the attributes of the smallest data set were considered the most relevant for a prediction model of disengagement. The results of the experiments for the smallest data set are presented in Table 1.

To summarize, relevant attributes for disengagement prediction were identified for HTML-Tutor. No method

TABLE 1
Html-Tutor: Experiment Results Summary

	BN	LR	SL	IBk	ASC	B	CvR	DT
%correct	89.77	86.36	87.12	84.85	90.91	89.77	90.91	90.15
TP rate	0.94	0.94	0.95	0.93	0.94	0.93	0.94	0.93
FP rate	0.18	0.29	0.28	0.24	0.16	0.17	0.16	0.16
Precision	0.91	0.86	0.87	0.89	0.92	0.92	0.92	0.92
Error	0.12	0.14	0.14	0.14	0.11	0.13	0.12	0.11

significantly outperformed the others, indicating consistency of prediction and allowing several possibilities for usage of the prediction methods, as discussed in Section 5.

The next step was to investigate whether this approach worked on a different system and more specifically, if the attributes identified as being relevant for HTML-Tutor would be relevant for another system, and therefore, produce acceptable levels of prediction. Two validation studies were conducted for this purpose, which are presented in the next section.

4 VALIDATION STUDIES

In order to validate our approach for engagement prediction presented above, we analyzed data from iHelp, a web-based learning system developed and deployed at the University of Saskatchewan. This system includes two web-based applications designed to support both learners and instructors throughout the learning process: the iHelp Discussion system and iHelp Learning Content Management System (also called iHelp Courses). The former allows communication among students and instructors, while the latter is designed to deliver online courses to students working at a distance, providing course content (text and multimedia) and quizzes. The content is organized in packages that contain a hierarchy of activities. A single package is displayed at one time on the left of the screen, as illustrated in Fig. 2. Besides the structure of the package, on the left, there are two menus, one related to *course actions*, such as preferences or search, and one related to other actions, such as logout. Each activity from the package is linked to a file that is displayed in the main area of the screen. At the top of this area, a navigation bar allows moving back and forward. Collaboration tools—chat and discussion forum—are available in the lower part of the screen.

The same type of data about the interactions was selected from registered information to perform the same type of analysis as the one performed on HTML-Tutor data. An HTML course was chosen to control the domain variable, and therefore, prevent differences in results caused by differences in subject matter.

Two studies were conducted with iHelp data. In the first study, logged data from 11 students were used, comprising a total of 108 sessions and 450 sequences (341 of exactly 10 minutes and 109 less than 10 minutes). The second study included logged data from 21 students (all the students studying that course), comprising a total of 218 sessions and 735 sequences (513 of exactly 10 minutes and 222 less than 10 minutes).

4.1 Study 1

In the analysis, several attributes mainly related to reading pages and taking quizzes were used. These attributes are presented in Table 2. The terms tests and quizzes will be used interchangeably; they refer to the same type of problem-solving activity, except that in HTML-Tutor, they are called tests, while in iHelp, they are referred to as quizzes.

Given the smaller number of instances, sequences of less than 10 minutes were included in the analysis to see if the number of instances has an influence on prediction. As a consequence, to distinguish between these sequences and the ones of exactly 10 minutes, the total time of a sequence was included as an attribute. Compared to the analysis of HTML-Tutor logs, in the first study, for iHelp, there are

Fig. 2. Screenshot of iHelp on XHTML content.

TABLE 2
The Attributes Used for Analysis

Code	Attribute description
NoPages	Number of pages read
AvgTimeP	Average time spent reading
NoQuestions	Number of questions from quizzes
AvgTimeQ	Average time spent on quizzes
Total time	Total time of a sequence

TABLE 3
Study 1: Experiment Results

Dataset	Results	BN	LR	SL	IBk	ASC	B	CvR	DT
DS1_S1	%correct	76.51	80.72	81.33	83.73	81.93	82.53	83.13	81.93
	TP rate	0.64	0.82	0.81	0.81	0.79	0.78	0.81	0.79
	FP rate	0.06	0.21	0.19	0.13	0.14	0.11	0.14	0.14
	Precision	0.94	0.84	0.86	0.90	0.88	0.90	0.89	0.88
	Error	0.28	0.27	0.32	0.23	0.25	0.26	0.25	0.25
DS2_S1	%correct	83.62	83.62	83.62	85.34	81.90	83.62	76.72	81.90
	TP rate	0.90	0.92	0.92	0.92	0.90	0.92	1.00	0.90
	FP rate	0.22	0.23	0.23	0.20	0.25	0.23	0.42	0.25
	Precision	0.77	0.76	0.76	0.78	0.74	0.76	0.65	0.74
	Error	0.21	0.24	0.31	0.23	0.27	0.27	0.26	0.27

fewer attributes related to quizzes. Thus, information on number of questions attempted and on time spent on them is included, but information about the correctness or incorrectness of answers given by users was not available at the time of data retrieval.

For each 10 minute sequence, the level of engagement was rated by an expert using the same approach as for HTML-Tutor that was briefly presented in Section 3. With HTML-Tutor, three levels of engagement were used: engaged, disengaged, and neutral. Neutral was used for situations when raters found it hard to decide whether the user was engaged or disengaged. With iHelp, this difficulty was not encountered. The rating consistency was verified on HTML-Tutor data by measuring intercoding reliability. However, with iHelp, only one rater classified the level of engagement for all sequences.

Two data sets were used in the analysis: DS1_S1 that included all sequences and DS2_S1 that included only sequences of exactly 10 minutes (S1 denotes Study 1). The same environment, WEKA, and the same eight methods were used for analysis. For DS1_S1, 67 percent of the sequences were used for training and 33 percent for testing; and for DS2_S1, 63 percent of the sequences were used for training and 37 percent for testing. Like in the experiments on HTML-Tutor data, the distribution of students within the two sets was controlled to avoid having sequences from the same user both in training and testing sets. The results are displayed in Table 3.

Compared to the results obtained on HTML-Tutor data, the prediction values are lower, for both the accuracy and the true positive rates. Also, the results are better for DS2_S1, especially for the true positives rate; however, the same data set has high rates of false positives, meaning that learners are classified as disengaged when in reality they are not. The overall prediction, however, is accurate, on average, more than 82 percent of the time and disengagement is still predicted correctly, on average, more than 85 percent of the time. Therefore, we can conclude that the attributes used for prediction are relevant for iHelp as well.

Two differences between HTML-Tutor data and iHelp data may account for the lower accuracy and true positive rates on the latter: the smaller number of instances and the missing information about the correctness of answers on quizzes. To investigate their influence, another study was needed.

During the labeling process of the iHelp data, a similarity was noticed with HTML-Tutor data in the patterns that disengaged students seemed to follow. Thus, some disengaged students spent a long time on the same page or test, while other students browsed very fast through content seemingly without reading. Based on these

observations, we decided to include two attributes that reflected these aspects and investigate their potential role for an improved prediction.

Therefore, a second study was conducted to address the previously mentioned aspects—the role of more data, of data on the performance on quizzes, and of the two new attributes. The next section described this study and its results.

4.2 Study 2

To address the issue related to the number of instances, more data were processed and labeled, adding up to 735 sequences, of which 513 were of exactly 10 minutes, while 222 were less than 10 minutes.

The initially unavailable information on correctness of answers to quizzes became available later, leading to the addition of a new attribute, i.e., *score* that reflected the performance on all quizzes. Unlike the two attributes in the HTML-Tutor—number of correct and incorrect answers, the score attribute aggregates this information in one indicator (this is how it is logged in iHelp).

We also looked for two attributes to reflect the two types of disengagement behavior identified. As they seemed to be related to time, we intended to use the average time spent on each page across all users, as suggested by [34]. However, data analysis revealed that some pages are accessed by very small numbers of users, sometimes only one—a problem that was encountered in other research as well [35]. Consequently, we decided to use the average reading speed known to be in between 200 and 250 words per minute [36], [37]. According to this reading speed, the majority of the pages would require less than 100 seconds (see Table 4) with only five pages exceeding 400 seconds.

Some pages included images and videos that could increase the time needed to read/view the information displayed. However, only four of the 21 students attempted to watch videos and the number of attempts and their corresponding times per attempt and per student are displayed in Table 5.

Taking into account the aforementioned information about iHelp pages distribution, we defined a lower threshold of five seconds and an upper threshold of 420 seconds (7 minutes). The five seconds threshold for the minimal time to read a page seems to be a "standard" in the literature (e.g., [35]). The 420 seconds threshold, even if

TABLE 4
Time Intervals for Reading and the
Number of Pages in Each Interval

Time interval	No of pages
500-550	3
400-500	2
300-400	5
200-300	41
100-200	145
<100	468

TABLE 5
Number of Attempts and Time Spent
Watching Videos Grouped by User

Subject	No of attempts	Time (sec.)
S1	1	3.47
S2	1	162
S3	9	1.16
	2	2.31
	1	94.91
S4	8	1.16
	2	2.31

TABLE 6
Data Sets Used in the Second Experiment

Dataset	Sequences	Attributes
DS1_S2	All	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score, Total time
DS2_S2	10 min.	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score,Total time
DS3_S2	All	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score, Total time, NoPpP, NoPpM
DS4_S2	10 min.	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score,Total time, NoPpP, NoPpM

TABLE 7
Study 2: Experiment Results

Dataset	Results	BN	LR	SL	IBk	ASC	B	CvR	DT
DS1_S2	%correct	80.43	79.15	79.57	78.72	82.55	78.72	82.13	82.55
	TP rate	0.72	0.62	0.67	0.65	0.69	0.68	0.73	0.73
	FP rate	0.05	0.10	0.10	0.11	0.00	0.08	0.08	0.11
	Precision	0.80	0.85	0.82	0.81	0.88	0.79	0.83	0.84
	Error	0.28	0.27	0.36	0.27	0.28	0.29	0.27	0.27
DS2_S2	%correct	85.62	85.92	85.56	85.44	84.77	85.80	85.37	85.07
	TP rate	0.77	0.77	0.76	0.78	0.76	0.77	0.75	0.76
	FP rate	0.06	0.05	0.05	0.07	0.06	0.05	0.04	0.07
	Precision	0.93	0.93	0.94	0.91	0.92	0.93	0.94	0.92
	Error	0.23	0.21	0.22	0.20	0.24	0.23	0.24	0.23
DS3_S2	%correct	88.09	88.09	87.66	82.13	82.98	88.94	89.36	87.23
	TP rate	0.90	0.86	0.85	0.73	0.62	0.86	0.86	0.85
	FP rate	0.13	0.08	0.11	0.11	0.07	0.13	0.11	0.10
	Precision	0.92	0.87	0.87	0.85	1.00	0.89	0.90	0.86
	Error	0.16	0.20	0.22	0.19	0.25	0.15	0.15	0.17
DS4_S2	%correct	97.50	97.93	97.99	97.87	97.38	97.50	97.44	97.75
	TP rate	0.97	0.97	0.97	0.96	0.95	0.96	0.96	0.97
	FP rate	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.01
	Precision	0.98	0.99	0.99	0.99	1.00	0.99	0.99	0.99
	Error	0.04	0.03	0.04	0.03	0.05	0.04	0.04	0.03

somehow arbitrary, balances the factors involved in our particular case, namely:

1. Most pages, i.e., more than 99 percent, require less than 400 seconds to be read. Moreover, 70 percent of the pages require less than 100 seconds and only five pages, i.e., less than 1 percent, are left out.
2. Very few students watched videos (that could be longer than 5 or even 10 minutes, which would considerably affect the way to establish engagement level for a 10-minutes sequence).
3. There may be individual differences in reading speed, and by allowing a rather loose upper threshold, slow speed is taken into account. However, fast speed is not covered.
4. Some learners go through the material more than once, leading to an at least doubled time needed for reading.

Based on this analysis, the following two meta-attributes were defined: 1) NoPpP: the number of pages above the threshold established for maximum time required to read a page (420 seconds) and 2) NoPpM: the number of pages below the threshold established for minimum time to read a page (5 seconds). These two attributes were added for each sequence. We call them meta-attributes because they are derived from the raw data.

To account for the contribution of more instances and the score attribute on the one hand, and the contribution of the two new attributes (NoPpM and NoPpP) on the other hand, four data sets were defined. These are described in Table 6. By comparing data sets DS1_S2 and DS2_S2 with data sets DS1_S1 and DS2_S1 from study 1, the contribution of more instances and the score attribute can be assessed; also this enables a more realistic comparison with the results from HTML-Tutor data. The results on data sets DS3_S2 and DS4_S2 will establish the influence of the two new attributes.

In the experiments, 68 percent of the sequences were used for training and 32 percent were used for testing. Also, like in the previous studies, the distribution of students was controlled to avoid having sequences from the same user both in training and testing sets.

For the data sets including all 735 sequences (DS1_S2 and DS2_S2), 500 were used for training and 235 for testing. For the data sets with 10 minutes sequences only (DS3_S2 and DS4_S2), from the 513 instances, 348 were used for training and 165 for testing. The results are presented in Table 7.

Comparing the results from DS1_S2 and DS2_S2 with the results from Study 1 (DS1_S1 and DS1_S2), an average decrease of accuracy of 1 percent and an average increase of 2.9 percent, respectively, are noticed. The true positive rate has decreased in Study 2 by 0.09 and 0.15, respectively. Therefore, we can conclude that more data and the additional score attribute did not significantly improve the prediction results.

The results for DS1_S2 and DS2_S2 (the data sets without the new attributes) are lower compared to the results from

TABLE 8
The Confusion Matrix for Simple Logistic

		Predicted		
		Disengaged	Engaged	Total
	Disengaged	79	2	81
Actual	Engaged	0	84	84
	Total	79	86	165

the other two data sets (DS3_S2 and DS4_S2), indicating a positive influence of the two new attributes and a significant information gain. The accuracy varies between 78 and 86 percent, while true positive rates have values between 0.62 and 0.78. Precision values range from 0.79 to 0.94; mean absolute error varies between 0.20 and 0.36.

The results for DS3_S2 and DS4_S2 (the data sets with the new attributes) presented in Table 7 show very good levels of prediction for all methods, with a correct prediction varying between approximately 82 and 98 percent. The results are similar for the true positive rates of the disengaged class, with most values varying between 0.85 and 0.97. However, there are two deviant cases: for DS1_S2, the results obtained with IBk and ASC for the true positive rate are considerably lower, 0.73 and 0.62, respectively. Precision varies between 0.85 and 1.00 and error between 0.03 and 0.25.

As in the case of HTML-Tutor, the very similar results obtained from different methods and trials show consistency of prediction and the attributes used for prediction.

The highest percentage of correctly predicted instances was obtained using Simple Logistic classification on DS4_S2: 97.99 percent. The confusion matrix is presented in Table 8.

Focusing on the disengaged learners only, Simple Logistic classification also performs best (on equal level with three other methods) on this data set: 0.97 true positives rate. The confusion matrix indicates that, on the one hand, none of the engaged learners are classified as disengaged and, on the other hand, two disengaged learners are classified as engaged. Possible implications are that in a real setting, engaged learners will not be interrupted for an intervention that is not required and some disengaged learners will not

be identified as such, and therefore, will not receive an intervention that would be required and beneficial.

Investigating the information gain of each attribute used in the analysis, the following ranking resulted from attribute ranking with information gain ranking filter as attribute evaluator (starting with the highest gain): NoPpP, NoPages, AvgTimeP, NoPpM, AvgTimeQ, Score, and NoQuestions.

The information gain brought by NoPpP is also reflected in the decision tree graph displayed in Fig. 3, where NoPpP is the attribute with the highest information gain, being the root of the tree. NoPpM also brings more information gain than attributes like Score and number of questions (NoQuestions).

The ranking clearly indicates that attributes related to reading are more important than the ones related to taking quizzes. This is consistent with the structure of the learning environment that provides more material for reading than for testing. The two new attributes contribute with metainformation that improves the prediction results.

5 DISCUSSION

The two validation studies on iHelp data indicate that the attributes identified in the studies on HTML-Tutor data are relevant for the new system as well.

Paired t-tests were used to investigate the statistical significance of the differences in the distribution of accuracy and true positive rates across the eighth methods between the two studies on iHelp data, on the one hand, and between the second iHelp study and the HTML-Tutor study, on the other hand. The mean for each data set and the significance of the t-test are displayed in Table 9. All accuracy and TP rates on all data sets were tested and proved to follow a normal distribution.

When comparing the results of two iHelp studies, we can see that the difference is statistically significant with one exception, i.e., the difference between the accuracy distribution for the data sets with sequences of only 10 minutes (DS1_S1 and DS1_S2). As there was some significant increase and some significant decrease as well, we can conclude that the amount of data and the new *score* attribute did not contribute to better predictions.

When comparing the results of the second iHelp study without the new attributes (DS2_S2) with the HTML-Tutor data, significantly lower accuracy and true positive rates
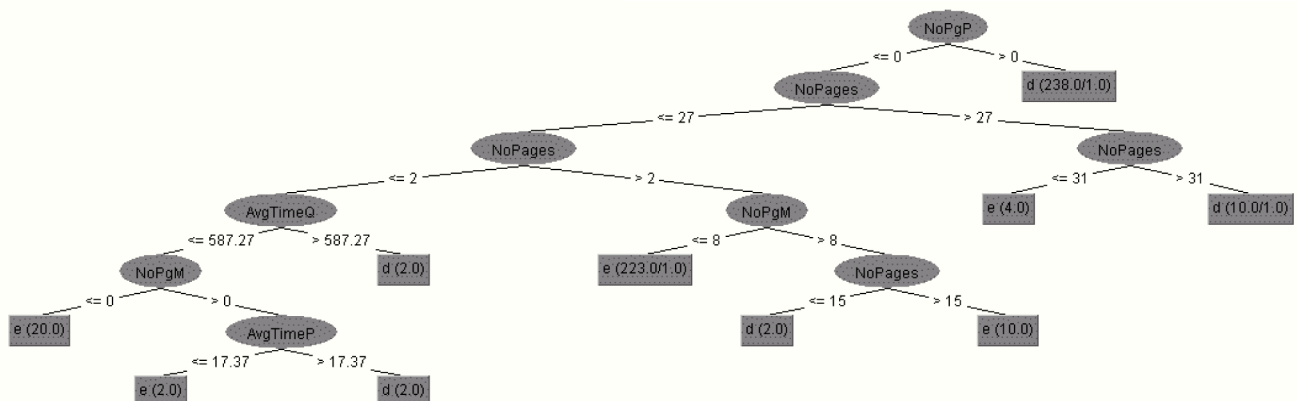

Fig. 3. Decision tree for data set DS4_S2.

TABLE 9
Pair T-Test Results

		Mean	Significance
Compared results	DS1_S1_accuracy	81.48	0.344
	DS1_S2_accuracy	80.48	
	DS1_S1_TP	0.78	0.015
	DS1_S2_TP	0.69	
	DS2_S1_accuracy	82.54	0.013
	DS2_S2_accuracy	85.44	
	DS2_S1_TP	0.92	0.000
	DS2_S2_TP	0.77	
	DS1_S2_accuracy	80.48	0.001
	DS3_S2_accuracy	86.81	
	DS1_S2_TP	0.69	0.006
	DS3_S2_TP	0.82	
	DS2_S2_accuracy-	85.44	0.000
	DS4_S2_accuracy	97.67	
	DS2_S2_TP	0.77	0.000
	DS4_S2_TP	0.96	
	HTML_Tutor_accuracy	88.73	0.007
	DS2_S2_accuracy	85.44	
	HTML_Tutor_TP	0.94	0.000
	DS2_S2_TP	.77	
	HTML_Tutor_accuracy	88.73	0.000
	DS4_S2_accuracy	97.67	
	HTML_Tutor_TP	0.94	0.000
	DS4_S2_TP	0.96	

are noticed for the iHelp data. The difference may be accounted for by the different ways the two systems are used. While HTML-Tutor is freely accessible on the web, iHelp is used in a formal educational setting. This may account for the different percentage of disengaged instances in the two lots of data: 65 percent for HTML-Tutor and 49 percent for the iHelp.

The relatively low contribution of the *score* attribute came as a surprise, as intuitively, such information seems relevant for the prediction of engagement or disengagement. This is even more surprising when considering that such information is essential in related research focused on problem-solving activities. Nevertheless, this may indicate an important difference between problem-solving environments and content delivering systems such as HTML-Tutor and iHelp where students engage in problem-solving activities usually after having studied the related material. To look deeper into this issue, the ranking of attributes in HTML-Tutor and iHelp could be used to give us more information on the importance of such attributes in both systems. Before looking into this, we discuss the contribution of the two new attributes introduced in the second iHelp study: NoPpP (number of pages above the threshold of maximal reading time) and NoPpM (number of pages below the threshold of minimal reading time).

Comparing the DS4_S2 data set from the second iHelp study (last from Table 7) containing the two new attributes with the HTML-Tutor results from Table 1, we notice an average increase of accuracy of 8.9 percent and an average

increase of true positive rate of 0.02. This improvement is most likely accounted for by the two new attributes: NoPpP and NoPpM. The increase in the true positive rate may not seem like a big improvement when directly compared with the HTML-Tutor results, but it is a statistically significant difference, as shown in Table 9. In the second iHelp study, when comparing the data sets with (DS3_S2 and DS4_S2) and without (DS1_S2 and DS2_S2) the new attributes, the results in Table 9 indicate a significant difference too. Therefore, the two new attributes significantly improve the prediction results.

To asses the contribution to prediction of the attributes in each system, three attribute evaluation methods with ranking as search method for attribute selection were used: chi-square, information gain, and OneR [29]. For HTML-Tutor, according to chi-square and information gain ranking, the most valuable attribute is *average time spent on pages*, followed by the *number of pages*, *number of tests*, *average time spent on tests*, *number of correctly answered tests*, and *number of incorrectly answered tests*. OneR ranking differs only in the position of the last two attributes: *number of incorrectly answered tests* comes before *number of correctly answered tests*.

The attribute ranking using information gain filter for iHelp attributes delivered the following ranking: NoPpP, NoPages, AvgTimeP, NoPpM, AvgTimeQ, Score, and NoQuestions. Chi-square evaluator produces the same ranking, except that the positions of the last two attributes are reversed, i.e., NoQuestions contributes a higher gain than Score. OneR evaluator produces a different ranking compared to the other two, even if the main trend is preserved (attributes related to reading come before the ones for quizzes): NoPpP, AvgTimeP, NoPages, NoPpM, NoQuestions, AvgTimeQ, and Score. The comparison in Table 10 is based on information gain evaluator.

The attribute ranking results show that for both HTML-Tutor and iHelp, the attributes related to reading are more important than the ones related to tests. The iHelp *score* attribute and its two correspondent attributes from HTML-Tutor (*number of currently answered tests* and *number of incorrectly answered tests*) are among the least important ones.

Table 10 summarizes the similarities and dissimilarities between the findings from iHelp and HTML-Tutor studies. Although some differences exist, the main fact is that a good level of prediction obtained using similar attributes on data sets from two different systems and applying the same methods indicates that disengagement prediction is possible using information related to events like reading pages and taking tests (solving problems), i.e., using information logged by most e-learning systems.

6   FUTURE PERSPECTIVES AND CONCLUSIONS

The validation studies suggest that our proposed approach for disengagement detection is potentially system-independent and it could be generalized to other systems. These results provide the blueprint for a component for automatic detection of disengagement that can be integrated into e-learning systems to keep track of the learner's engagement status. Such a component offers the opportunity to intervene when appropriate—either automatically or through a tutor. We argue that disengagement detection represents the first step toward more detailed motivation elicitation. For example, once disengagement has been

COCEA AND WEIBELZAHL: DISENGAGEMENT DETECTION IN ONLINE LEARNING: VALIDATION STUDIES AND PERSPECTIVES

TABLE 10
Similarities and Dissimilarities between iHelp and HTML-Tutor

	iHelp	HTML-Tutor
Prediction based on reading and tests attributes	85-86% with no additional attributes 97-98% with two additional attributes	85-91%
Attribute ranking (information gain)	1'. Number of pages above a threshold (NoPgP) 1. Number of pages accessed (NoPages) 2. Average time reading (AvgTimeP) 2'.Number of pages below a threshold (NoPgM) 3. Average time on quizzes (AvgTimeQ) 4. Score 5. Number of questions (NoQuestions)	1.Average time on pages 2. Number of pages 3. Number of tests 4. Average time on tests 5. Number of correctly answered tests 6. Number of incorrectly answered tests

detected, the system may enter into a dialog with the learner in order to find out more about his/her motivation [38]. Furthermore, this information could be used for more targeted personalized intervention [39].

In both systems, iHelp and HTML-Tutor, two different categories of disengaged learners were distinguished based on their patterns of behavior: 1) disengaged students that click fast through pages without reading them and 2) disengaged students that spend long time on a page, (far) exceeding the needed time for reading that page. Two of the previous approaches mentioned in Section 2 also present some patterns. Thus, we find a similarity between blind guess in [6] and unmotivated guess in [9], on the one hand, and the fast click through pages, on the other hand, as both reflect students' rush and lack of attention. However, we found no correspondent pattern in the literature for the long time spent on the same page. This may be due to the nature of the system, as this pattern is more likely to be displayed while reading rather than problem solving. This pattern also gives rise to problems like not knowing if a learner is still engaged in learning, but not using the system, if she/he is disengaged with regards to the current activity and engaged in other behaviors like chatting with friends, reading e-mail, or using other software in general, or simply took an intentional break and spent the break time on the computer or somewhere else. This could easily be addressed by including in the system "break" and "resume" buttons, for example. As the learners may forget to use these buttons, another approach would be for the system to display a window after some time of inactivity asking the learner whether the elapsed time was a break and if she/he would like some help. The help choice could trigger either a more detailed assessment of their motivation or an intervention strategy.

Despite the problem they may pose, knowledge about the two patterns of disengagement would be useful for a more targeted intervention and in further work, the possibility to predict them will be investigated.

The two observed patterns of disengagement led to the introduction of two meta-attributes. Their usage considerably improved the prediction values. However, another way of using this knowledge would be to derive some rules that could be used for automatic annotations of data. For example, sequences for which the time spent on a page is above the upper threshold (420 seconds) for reading a page could be labeled as disengaged. Similarly, sequences that

have more than two-thirds of the pages below the lower threshold (5 seconds) for reading a page could be labeled as disengaged. This is another direction for future work that we intend to follow.

As already mentioned, previous research addressed disengagement and system gaming behavior [1], [10] (as a type of disengagement) only for problem-solving activities for which information on correctness or incorrectness of answers is very important, if not essential. For our approach, this information has some importance, but it is not indispensable as shown in the first study on iHelp data. Therefore, if the learners are only reading, without doing any problem-solving activities, prediction of disengagement is still possible.

Moreover, the comparison of prediction values across the two validation studies on iHelp data suggests a rather limited impact of the amount of available data on prediction quality. The differences observed were quite small indicating that the data necessary for training (at least for the initial one) are fairly modest, consequently facilitating the introduction of an automatic component for disengagement detection.

ACKNOWLEDGMENTS

REFERENCES

[1] R. Baker, A. Corbett, and K. Koedinger, "Detecting Student Misuse of Intelligent Tutoring Systems," *Proc. Seventh Int'l Conf. Intelligent Tutoring Systems,* pp. 531-540, 2004.

[2] T. Connolly and M. Stansfield, "Using Games-Based eLearning Technologies in Overcoming Difficulties in Teaching Information Systems," *J. Information Technology Education,* vol. 5, pp. 459-476, 2006.

[3] G.D. Chen, G.Y. Shen, K.L. Ou, and B. Liu, "Promoting Motivation and Eliminating Disorientation for Web Based Courses by a Multi-User Game," *Proc. World Conf. Educational Multimedia and Hypermedia and World Conf. Educational Telecomm.,* June 1998.

[4] C.R. Beal, L. Qu, and H. Lee, "Classifying Learner Engagement through Integration of Multiple Data Sources," *Proc. 21st Nat'l Conf. Artificial Intelligence,* pp. 2-8, 2006.

[5] A. de Vicente and H. Pain, "Informing the Detection of the Students' Motivational State: An Empirical Study," *Proc. Sixth Int'l Conf. Intelligent Tutoring Systems,* S.A. Cerri et al., eds., pp. 933-943, 2002.

[6] J. Beck, "Engagement Tracing: Using Response Times to Model Student Disengagement," *Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology,* C. Looi et al., eds., pp. 88-95, IOS Press, 2005.

[7] I. Arroyo and B.P. Woolf, "Inferring Learning and Attitudes from a Bayesian Network of Log File Data," *Artificial Intelligence in Education, Supporting Learning through Intelligent and Socially Informed Technology,* C.K. Looi et al., eds., pp. 33-34, IOS Press, 2005.

[8] L. Qu, N. Wang, and W.L. Johnson, "Detecting the Learner's Motivational States in an Interactive Learning Environment," *Artificial Intelligence in Education,* C.-K. Looi et al., eds., pp. 547-554, IOS Press, 2005.

[9] J. Johns and B. Woolf, "A Dynamic Mixture Model to Detect Student Motivation and Proficiency," *Proc. 21st Nat'l Conf. Artificial Intelligence (AAAI-06),* 2006.

[10] J. Walonoski and N.T. Heffernan, "Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems," *Proc. Eighth Int'l Conf. Intelligent Tutoring Systems,* M. Ikeda, K. Ashley, and T.-W. Chan, eds., pp. 382-391, 2006.

[11] J. Walonoski and N.T. Heffernan, "Prevention of Off-Task Gaming Behaviour within Intelligent Tutoring Systems," *Proc. Eighth Int'l Conf. Intelligent Tutoring Systems,* M. Ikeda, K. Ashley, and T.-W. Chan, eds., pp. 722-724, 2006.

[12] M. Cocea and S. Weibelzahl, "Eliciting Motivation Knowledge from Log Files towards Motivation Diagnosis for Adaptive Systems," *Proc. 11th Int'l Conf. User Modelling (UM '07),* C. Conati, K. McCoy, and G. Paliouras, eds., pp. 197-206, 2007.

[13] P.R. Pintrich and D.H. Schunk, *Motivation in Education: Theory, Research and Applications.* Prentice Hall, 2002.

[14] J.M. Keller, "Development and Use of the ARCS Model of Instructional Design," *J. Instructional Development,* vol. 10, no. 3, pp. 2-10, 1987.

[15] W. Burleson and R.W. Picard, "Evidence for Gender Specific Approaches to the Development of Emotionally Intelligent Learning Companions," *IEEE Intelligent Systems,* Special Issue on Intelligent Educational Systems, vol. 22, no. 4, pp. 62-69, 2007.

[16] S. D'Mello, T. Jackson, S. Craig, B. Morgan, P. Chipman, H. White, N. Person, B. Kort, R. el Kaliouby, R.W. Picard, and A. Graesser, "AutoTutor Detects and Responds to Learners Affective and Cognitive States," *Proc. Workshop Emotional and Cognitive Issues at the Int'l Conf. Intelligent Tutoring Systems,* June 2008.

[17] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard, "Affect-Aware Tutors: Recognising and Responding to Student Affect," *Int'l J. Learning Technology,* vol. 4, nos. 3/4, pp. 129-163, 2009.

[18] I. Arroyo, K. Ferguson, J. Johns, T. Dragon, H. Meheranian, D. Fisher, A. Barto, S. Mahadevan, and B.P. Woolf, "Repairing Disengagement with Non-Invasive Interventions," *Proc. 13th Int'l Conf. Artificial Intelligence in Education,* pp. 195-202, 2007.

[19] M.M.T. Rodrigo, G. Rebolledo-Mendez, R.S.J.d. Baker, B. du Boulay, J.O. Sugay, S.A.L. Lim, M.B. Espejo-Lahoz, and R. Luckin, "The Effects of Motivational Modeling on Affect in an Intelligent Tutoring System," *Proc. Int'l Conf. Computers in Education,* 2008.

[20] R. Baker, S. D'Mello, M. Rodrigo, and A. Graesser, "Better to be Frustrated than Bored: The Incidence and Persistence of Affect during Interactions with Three Different Computer-Based Learning Environments," *Int'l J. Human-Computer Studies,* vol. 68, no. 4, pp. 223-241, 2010.

[21] C. Conati and H. Maclaren, "Empirically Building and Evaluating a Probabilistic Model of User Affect," *User Modeling and User-Adapted Interaction,* vol. 19, no. 3, pp. 267-303, 2009.

[22] M.M.T. Rodrigo, R. Baker, M.C. Jadud, A.C.M. Amarra, T. Dy, M.B.V. Espejo-Lahoz, S.A.L. Lim, S.A.M.S. Pascua, J.O. Sugay, and E.S. Tabanao, "Affective and Behavioral Predictors of Novice Programmer Achievement," *Proc. Conf. Innovation and Technology in Computer Science Education (ITiCSE '09),* pp. 156-160, 2009.

[23] M.M.T. Rodrigo, R.S.J.d. Baker, M.C.V. Lagud, S.A.L. Lim, A.F. Macapanpan, S.A.M.S. Pascua, J.Q. Santillano, L.R.S. Sevilla, J.O. Sugay, S. Tep, and N.J. B. Viehland, "Affect and Usage Choices in Simulation Problem-Solving Environments," *Proc. Conf. Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work (AIED '07),* pp. 145-152, 2007.

[24] S. D'Mello and A. Graesser, "Automatic Detection of Learners' Emotions from Gross Body Language," *Applied Artificial Intelligence,* vol. 23, no. 2, pp. 123-150, 2009.

[25] S. Lee, S.W. McQuiggan, and J.C. Lester, "Inducing User Affect Recognition Models for Task-Oriented Environments," *Proc. Int'l Conf. User Modeling,* pp. 380-384, 2007.

[26] J.P. Rowe, S.W. McQuiggan, J.L. Robison, and J.C. Lester, "Off-Task Behavior in Narrative-Centered Learning Environments," *Proc. Conf. Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work (AIED '09),* pp. 99-106, 2009.

[27] G. Weber, H.-C. Kuhl, and S. Weibelzahl, "Developing Adaptive Internet Based Courses with the Authoring System NetCoach2," *Hypermedia: Openness, Structural Awareness, and Adaptivity,* pp. 226-238, Springer, 2001.

[28] M. Cocea and S. Weibelzahl, "Can Log Files Analysis Estimate Learners' Level of Motivation?" *Proc. 14th Workshop Adaptivity and User Modeling in Interactive Systems (ABIS '06),* pp. 32-35, 2006.

[29] I.H. Witten and E. Frank, *Data Mining. Practical Machine Learning Tools and Techniques,* second ed. Morgan Kauffman/Elsevier, 2005.

[30] T.M. Mitchell, *Machine Learning.* McGraw Hill, 1997.

[31] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement,* vol. 20, no. 1, pp. 37-46, 1960.

[32] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology.* Sage, 2004.

[33] M. Lombard, J. Snyder-Duch, and C.C. Bracken, "Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research," http://www.temple.edu/mmc/reliability, 2003.

[34] R. Rafter and B. Smyth, "Passive Profiling from Server Logs in an Online Recruitment Environment," *Proc. IJCAI Workshop Intelligent Techniques for Web Personalization (ITWP '01),* 2001.

[35] R. Farzan and P. Brusilovsky, "Social Navigation Support in E-Learning: What Are Real Footprints," *Proc. Workshop Intelligent Techniques for Web Personalization (IJCAI '05),* pp. 49-56, 2005.

[36] Speed Reading Test, http://www.readingsoft.com, 2007.

[37] TurboRead Speed Reading, http://www.turboread.com, 2007.

[38] M. Cocea, "Assessment of Motivation in Online Learning Environments," *Proc. Fourth Int'l Conf. Adaptive Hypermedia and Adaptive Web-Based Systems,* V. Wade et al., eds., pp. 414-418, 2006.

[39] T. Hurley, "Intervention Strategies to Increase Self-Efficacy and Self-Regulation in Adaptive OnLine Learning," *Proc. Fourth Int'l Conf. Adaptive Hypermedia and Adaptive Web-Based Systems,* V. Wade et al., eds., pp. 440-444, 2006.

Mihaela Cocea received the BSc degree in psychology and education and the BSc degree in computer science from "Al. I. Cuza" University of Iasi in 2002 and 2003, respectively, and the MSc degree in human relations and communication by Research in Learning Technologies from the National College of Ireland in 2007. She is currently working toward the PhD degree at the Department of Computer Science and Information Systems, Birkbeck College, University of London. Her research interests include intelligent learning environments, user modeling, and adaptive feedback.

Stephan Weibelzahl received the PhD degree from the University of Trier, Germany. He holds a lecturer position at the National College of Ireland in Dublin. After heading a research group at the Fraunhofer Institute of Experimental Software Engineering (IESE), Kaiserslautern, Germany, he joined National College of Ireland in 2004. With his background in psychology and computer science, he has long-standing research expertise in developing and evaluating Adaptive e-Learning Systems. His research interests include Adaptive Systems, learning technologies, evaluation, Knowledge Management, and Blended Learning.