Developing a Log-based Motivation Measuring Tool

Arnon Hershkovitz and Rafi Nachmias[1]
{arnonher, nachmias}@post.tau.ac.il
[1] Knowledge Technology Lab, School of Education, Tel Aviv University, Israel

Abstract. The purpose of this study is to develop a conceptual framework and a tool for measuring motivation of online learners. The study was carried out in three phases, the first of which was the construction of a framework, based on an extensive literature review. Phase two consisted of identifying variables computable from log file data, associate with the framework and compatible with previous empirical research. For this purpose, an empirical study was designed and a specific learning environment focusing on vocabulary for adults was chosen. Log files of a large population (N=2,162) were collected and variables were identified using *Learnograms*, a visual representation of learning variables over time. This phase resulted in seven explicitly defined variables, along with a mechanism to calculate them from the raw log files. The third phase included preprocessing of the dataset (reducing it to 674 cases) and application of hierarchical clustering of the variables. This phase resulted in three clusters interpreted to fit the three dimensions of motivation defined in the framework. A discussion of this study and further research is provided.

1 Introduction

Assessment of learners' motivation in online environments has been a challenge for both researchers and instructors, and the reason for it is twofold: motivation is an important factor affecting the learning process and explaining individual differences, however it is a factor difficult to evaluate without direct contact with the learner. This gap may be bridged with log file analysis, which makes it possible to learn about the online learner by means of automatically and continuously collected digital traces. Log file analysis in education research is an emerging field, however, only little research was done regarding motivation appraisal using this method.

The purpose of this study is to construct a conceptual framework and to develop a tool for measuring motivation of online learners. The motivation is built upon three dimensions to be measured based on data solely from log files: Engagement, Energization and Source. Learning variables, which describe patterns of learners' behavior, will be identified using the technique of analyzing *Learnogram*, a visual representation of learning variables over time [13]. These variables should be computable from the log files, associate with the framework, and compatible with previous empirical research. For clustering the variables, in order to classify them to the motivation dimensions, an empirical study was planned and conducted, and will be reported in this article. Thus, the main research question being addressed is: how can motivation be measured using only data from log files?

Continuing previous work done in this field, the main contributions of this study is in constructing an empirical-based automated tool for motivation detection from log files. A second contribution is the procedure in itself, since it might be transferred for developing measuring tools for other learner characteristics (e.g., anxiety, self-regulation).

2 Background

Motivation has been suggested as a factor explaining individual differences in intensity and direction of behavior [10]. It is generally accepted that motivation is "an internal state or condition that serves to activate or energize behavior and give it direction" [11]. The sources of motivation can be either internal (e.g., interestingness, enjoyment) or external (e.g., wishing for high grades, fear of parental sanctions) to the person [6].

Motivational patterns, in addition to ability, may influence the way people learn: whether they seek or avoid challenges, persist or withdraw upon difficulties, or whether they use and develop their skills effectively [7]. Different motivational patterns relate to different aspects of the learning process, e.g., achievement goals (performance or mastery), time spent on tasks, performance [1, 8, 12, 16].

Unlike configurations in which the instructor sees the students and might infer their motivation level from facial expressions, tonality - online learning supposedly disables motivation assessment. However, previous research has suggested several methods for tackling this challenge (see [4]). This research is focused on motivation measuring using information available only through log files. Table 1 summarizes the motivation-related terms and variables from five studies that mainly used learner-computer interaction data.

Following the definition given above, and based on the reviewed literature, we suggest a motivation measuring framework which considers three dimensions: a) Engagement - relates to the motivation intensity. (although using the same term, by Engagement we mean a more generalized idea than in [2, 3].); b) Energization, which refers to the way motivation is preserved and directed; c) Source of motivation (internal or external).

Table 1. Previous research on motivation recognition based on learner-computer interaction

Research	Motivation-Related Terms	Learning Variables Calculated from Logs
Beck [2]	Engagement (defined by the author)	Question response time; answer correctness
Cocea & Weibelzahl [3]	Engagement (defined by the authors)	# of pages read; time spent reading pages; # of tests/quizzes; time spent on test/quizzes
Qu & Johnson [14]	Confidence, confusion, effort	Reading time; decision time (before perform the task); task duration; # of finished tasks; # of tasks performed not from learning "plan"
de Vicente & Pain [5][1]	Control, challenge, independence, fantasy; confidence, sensory/cognitive interest, effort, satisfaction.	Quality, speed (of performance), give up
Zhang, Cheng, He, & Huang [17]	Attention, confidence	# of non-error compilations; ratio of working time and class' average; # of hints; # of executions; time until typing in editor

[1] The research was based on captured screen activity.

Within this framework, the two objectives of this study are: a) To identify variables related to the three motivation dimensions, which are computable from data stored in the system log files; b) To cluster these variables, based on empirical data, for classifying them according to the three motivation dimensions.

3 Methodology

3.1 Procedure

The study was carried out in three consecutive phases using different methodologies:

Phase I – Constructing the Research Framework. This literature-based phase was used to conceptualize the terms to be used in our motivation research. Within the framework, three dimensions were defined (Engagement, Energization, Source); see previous section.

Phase II – Identifying Variables. In order to choose and define motivation-related variables, *Learnograms* – visual representations of learning variables over time – were used as the main research tool (as described in [13]). *Learnograms* presenting students' activity (N=5) in an online vocabulary course for adults (see section 3.2) were observed, in order to identify the relevant computable variables. The compatibility of the variable to previous empirical research in this field was taken into consideration, as well as their association to our framework. At the end of this phase, seven variables were identified.

Phase III – Classifying the Variables According to the Motivation Dimensions. An empirical study for the evaluation of the identified variables was conducted (with the same learning environment used in phase I). Log files of a large population (N=2,162) for one month (April 2007) were collected and preprocessed. Students using the researched system belong to different courses (varied by length, intensity, starting date and proximity to the exam), however this logged segment was analyzed regardless the student's learning stage. A filter was applied for keeping students with at least 3 active sessions (N=1,444). Algorithms for calculating the variables were formally written and implemented using Matlab. The dataset was preprocessed and the final set of cases to be analyzed was defined (N=674). Finally, Hierarchical Clustering of the variables was applied using SPSS, with Pearson Correlation Distance as the measure and Between-groups Linkage as the clustering method.

3.2 The Learning Environment

A simple yet very intensive online learning unit was chosen as the research field. This fully-online environment focuses on Hebrew vocabulary and is accessible for students who take a face-to-face preparatory course for the Psychometric Entrance Exam (for Israeli universities). The online system is available for the participants from the beginning of the course and until the exam date (between 3 weeks and 3 months in total).

The system includes a database of around 5,000 words/phrases in Hebrew and offers varied instructional strategies: a) Memorizing, in which the student browses a table of the words/phrases along with their meanings; b) Practicing, in which the student browses the table of the words/phrases without their meaning. The student may ask for a hint or for the explanation for each word/phrase; c) Searching for specific word/phrase; d) Gaming; e) Self testing, in the form of the exam the students will finally take. Throughout the learning process, the student may mark each word/phrase as "well known", "not-well known" or "unknown". This information is stored and being used by the system.

3.3 Log File Description

The researched system logs the students' activity, thus each student is identified by a serial number. Each row in the log file documents a session, initiating by entering the system and ending with closing the application window. For each session, the following attributes are kept: starting date, starting/ending time, number of words marked as "known" at the beginning/end of the session, ordered list of actions and their timestamps.

4 Results

4.1 Motivation-related Variables (Phase II)

The authors have examined the *Learnograms* of a few students (N=5), searching for interesting patterns and irregularities, while considering learning behavior which may be related to motivation. Seven variables were identified and calculated (see Table 2).

4.2 Classifying the Variables (Phase III)

The variable distributions were examined over the 3-sessions filtered dataset (N=1,444), see Figure 1. Three of the variables had a significant 0-values noise: *wordMarkPace*, *examPC*, *gamePC*, thus cases with 0-value in them were cleaned for focusing on the positive-value cases. Since the variables were skewed in the final dataset (N=674), we used transformations of log (*timeOnTaskPC*, *avgSession*, *wordMarkPace*, *examPC*, *gamePC*) and square-root (*avgActPace*, *avgBtwnSessions*).

Table 2. Motivation-related variables

Variable name	Variable description	Unit	Calculation remarks
timeOnTaskPC	Time on task percentage	[%]	Total time of active sessions ([min]) divided by total time frame documented.
avgSession	Average session duration	[min]	
avgActPace	Average pace of activity within sessions	[actions/ min]	Pace of activity per session is the number of actions divided by the session duration
avgBtwnSessions	Average time between sessions	[min]	
wordMarkPace	Pace of word marking	[words/ min]	Changed number of known words from beginning to end (can be negative) divided by total time frame documented
examPC	Exam activities percent	[%]	# of exam actions divided by total # of actions
gamePC	Game activities Percent	[%]	# of game actions divided by total # of actions

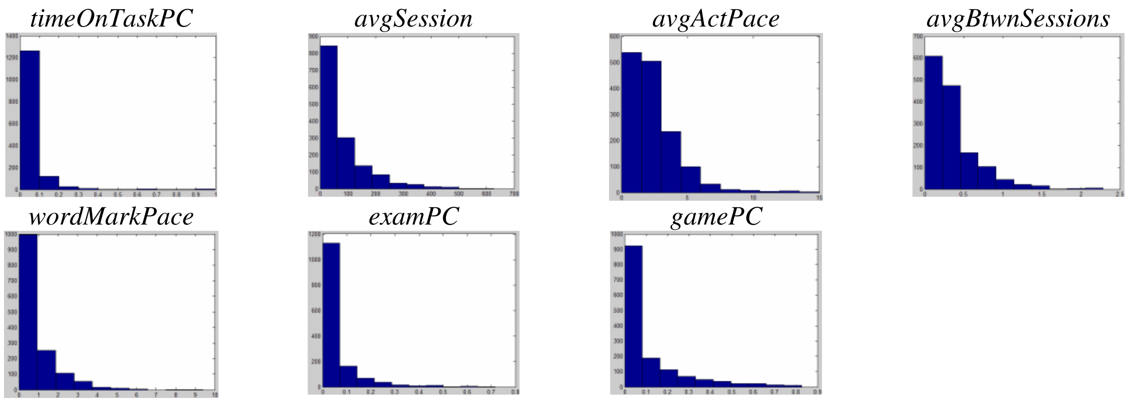

Figure 1. Distribution of the variables before cleaning and transformation were applied (N=1,444)

The clustering process is described by a dendrogram (from the Greek dendron "tree", -gramma "drawing") presented in Figure 2. The vertical lines determine which variables/clusters were grouped together and at which stage of the algorithm (from left to right). For example, the first coupled variables were *timeOnTaskPC* and *avgSession*, and next *examPC* and *gamePC* were grouped. The resulting clusters appear in Table 3, their relation to the motivation dimensions is given in the Discussion below.

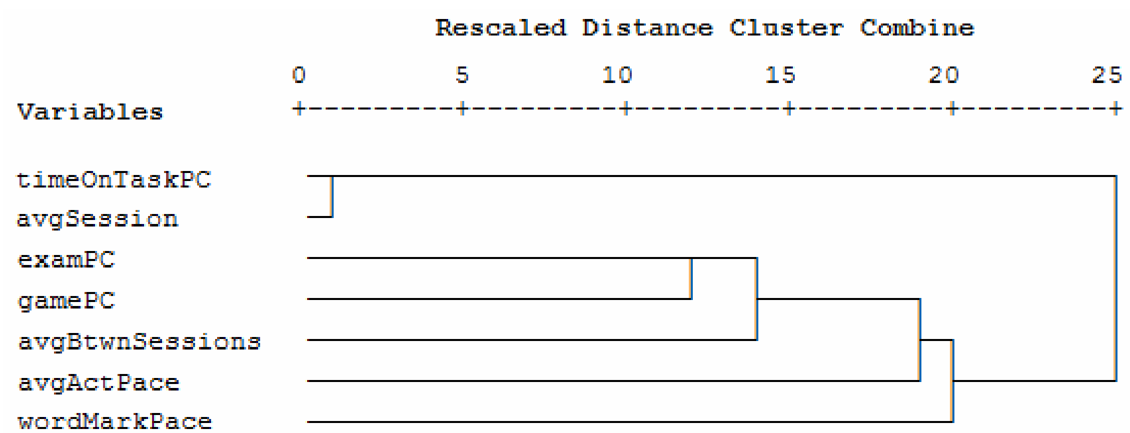

Figure 2. Dendrogram of the hierarchical clustering process

Table 3. The resulted clusters and their mapping to the motivation dimensions

Cluster	1	2	3
Variables	*timeOnTaskPC* *avgSession*	*examPC* *gamePC* *avgBtwnSessions* *avgActPace*	*wordMarkPace*
Motivation dimension	Engagement	Source	Energization

5 Discussion

In this study, an empirically-constructed tool was developed for log-based measuring of online learners' motivation. Motivation is measured by three dimensions – Engagement, Energization, and Source – and by seven computable variables corresponding to these dimensions (see Table 3). The classification of the clustered variables to the three dimensions is based on previous research in this area.

The variables *timeOnTask* and *avgSession*, which form the first cluster, might be related to the extent of Engagement, as it was previously suggested that working time might be a measure for attention or engagement [3, 17]. *examPC* and *gamePC* - grouped together in the second cluster - reflect the student's *Source* of motivation; it may be reasonable to hypothesize (inspired by, e.g., [9, 15]) that students who frequently tend to take self exams (related to performance-goal orientation) have extrinsic motivation to learn, while those who tend to game applications (related to learning-goal orientation) are intrinsically motivated. The variables *avgActPace* and *avgBtwnSessions* are also clustered together with the previous two, but their closeness to Source of motivation is yet to be established. The variable *wordMarkPace*, indicating the word marking speed, forms the third cluster. According to a diagnosis rule found in de Vicente and Pain [5], fast speed of activity together with high quality of performance (when staying in similarly-difficulty exercises) suggests increasing motivation. Since an increase in the number of words marked is an indication of the student's perceived knowledge (i.e., a reflection of the performance), *wordMarkPace* might be related to the direction of motivation, i.e., Energization.

The tool developed in this study enables to measure online learners' motivation by using solely information stored in log files. However, there are three limitations to this innovative tool. First, variables were identified based on a specific learning environment; it might be useful for similar systems, but for different environments (varied by, e.g., learning domain, instruction modes available) these variables should be converted, and their clustering should be re-examined. Secondly, the classification to the motivation dimensions within the framework has not been validated yet, as well as their actual scales; it is within the authors' agenda to continue in the direction of validation. Third, the tool might not be complete; we only focused on seven variables, however others might be considered. Identifying these variables based on a segment of the learning makes it possible to employ this tool during the learning process; that way, intervention when needed might be possible, and changes in motivation may be analyzed.

Further to the developed tool, the process used in this study – i.e., constructing a literature-based conceptual framework, using *Learnograms* for identifying variables, clustering and classifying these variables within the framework - is of great importance, since it is a procedure which might be transferable to other domains (e.g., anxiety, self-regulation) for developing measuring tools.

Measuring the online learner's motivation has a major role in the instruction-learning cycle. Monitoring the learner's motivation might enable the instructor to interfere when needed (e.g., when student's motivation is decreasing), and should help in developing of intelligent tutoring systems which react not only to the learner's cognitive behavior but also to her or his affective situation. The overall objective of this underlying approach is to increase the efficiency of the learning process.

References

1. Ames C, Archer J: Achievement goals in the classroom: Students' learning strategies and motivation. Journal of Educational Psychology 80: 260-267 (1988).

2. Beck JE: Using response times to model student disengagement ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments, Maceio, Brazil (2004).

3. Cocea M, Weibelzahl S: Cross-system validation of engagement prediction from log files Second European Conference on Technology Enhanced Learning (EC-TEL 2007), Crete, Greece (2007).

4. de Vicente A, Pain H: Motivation diagnosis in intelligent tutoring systems In: Goettl BP, Halff HM, Redfield CL, Shute VJ (eds) Fourth International Conference on Intelligent Tutoring Systems, pp. 86-95. Springer Berlin, San Antonio, TX (1998).

5. de Vicente A, Pain H: Informing the detection of the students' motivational state: An empirical study The Sixth International Conference on Intelligent Tutoring Systems (ITS 2002), Biarritz, France (2002).

6. Deci EL, Ryan RM: Intrinsic Motivation and Self-Determination in Human Behavior. Plenum, New York, NY (1985).

7. Dweck CS: Motivational processes affecting learning. American Psychologist 41: 1040-1048 (1986).

8. Elliott ES, Dweck CS: Goals: An approach to motivation and achievement. Journal of Personality and Social Psychology 54: 5-12 (1988).

9. Heyman GD, Dweck CS: Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. Motivation and Emotion 16: 231-247 (1992).

10. Humphreys MS, Revelle W: Personality, motivation, and performance: A theory of the relationship between individual differences and information processing. Psychological Review 91: 153-184 (1984).

11. Kleinginna PR, Kleinginna AM: A categorized list of emotion definitions, with suggestions for a consensual definition. Motivation and Emotion 5: 345-378 (1981).

12. Masgoret AM, Gardner RC: Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. Language Learning 23: 123-163 (2003).

13. Nachmias R, Hershkovitz A: Using Web Mining for Understanding the Behavior of the Online Learner International Workshop on Applying Data Mining in e-Learning, at the 2nd European Conference on Technology Enhanced Learning (EC-TEL'07), Crete, Greece (2007).

14. Qu L, Johnson WL: Detecting the learner's motivational states in an interactive learning environment Artificial Intelligence in Education, Amsterdam, The Netherlands (2005).

15. Ryan RM, Deci EL: Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary Educational Psychology 25: 54-67 (2000).

16. Singh K, Granville M, Dika S: Mathematics and science achievement: Effects of motivation, interest, and academic engagement. Journal of Educational Research 95: 323-332 (2002).

17. Zhang G, Cheng Z, He A, Huang T: A WWW-based learner's learning motivation detecting system International Workshop on Research Directions and Challenge Problems in Advanced Information Systems Engineering, Honjo City, Japan (2003).