Eliciting Motivation Knowledge from Log Files towards Motivation Diagnosis for Adaptive Systems

Mihaela Cocea, Stephan Weibelzahl

National College of Ireland, School of Informatics,
Mayor Street 1, Dublin 1, Ireland
{mcocea, sweibelzahl}@ncirl.ie

Abstract. Motivation is well-known for its importance in learning and its influence on cognitive processes. Adaptive systems would greatly benefit from having a user model of the learner's motivation, especially if integrated with information about knowledge. In this paper a log file analysis for eliciting motivation knowledge is presented, as a first step towards a user model for motivation. Several data mining techniques are used in order to find the best method and the best indicators for disengagement prediction. Results show a very good level of prediction: around 87% correctly predicted instances of all three levels of engagement and 93% correctly predicted instances of disengagement. Data sets with reduced attribute sets show similar results, indicating that engagement level can be predicted from information like reading pages and taking tests, which are common to most e-Learning systems.

Keywords: e-Learning, motivation, log files analysis, data mining, adaptive systems, user modeling

1 Introduction

Although motivation is a key component of learning, the main focus in adaptive educational systems is on cognitive processes. There is a general agreement about the importance of motivation, but little research is done in this area. Most e-Learning systems, including adaptive systems are focused on cognitive processing and on knowledge acquisition. If motivation is considered when building a system, it only covers aspects of system design, in terms of how the content is structured and presented. Nevertheless, the influence of motivation on cognitive processes explains why some users achieve high performance while others perform poorly or even drop out [18].

Adaptive systems work with user models of goals, knowledge and preferences in order to deliver personalized content and make learning more efficient. Given the close relation between cognitive processes and motivational states, a user model that would integrate information about knowledge and motivational states would lead to a more personalized and more efficient adaptation. Thus, we are interested in motivation diagnosis and in building a user model of the learner's motivation.

In this paper we present the results on the first step of our research project, which is focused on eliciting motivation knowledge from log files. The paper is organized as follows. Section 2 discusses previous work related to motivation in e-Learning; it includes a description of our approach and of the common and the different points in terms of theoretical base and methodology; previous work related to the use of log files analysis in education is also presented, with a particular interest in approaches to motivation. The analysis of log files is discussed in Section 3, and Section 4 concludes the paper with a summary and a brief description of further work.

2 Previous and Current Research

We will refer here only to research on motivation diagnosis, presenting a few relevant works for our approach. We also present our research and stress the communalities and the differences with previous approaches.

2.1 Previous Research on Motivation Diagnosis

Human tutors usually infer motivation from observational cues like mimics, posture, gesture, conversational cues etc. These are difficult to be processed by adaptive systems (e.g. [7], [8], [14]). Moreover, in regards to e-Learning, the amount and type of information that is available to humans and computers is quite limited. Previous approaches have focused on motivation diagnosis from cues that can be easily processed automatically, e.g. learners' interactions with the system, time spent on a task, their statements about their motivation etc..

Three of these approaches are of particular interest for our research. All of them are related to Keller's ARCS model [16] which stands for Attention, Relevance, Confidence and Satisfaction. First, a rule-based approach has been suggested, inferring motivational states from two sources: the *interactions* of the students with the tutoring system and their *motivational traits* [6]. A second approach infers three aspects of motivation: *confidence*, *confusion* and *effort*, from several sources: the learner's focus of attention, the current task and expected time to perform the task [21]. A third approach used factorial analysis to group user's actions that predict *attention* and *confidence* [29].

Our Perspective on Motivation Diagnosis

These approaches target a motivation diagnosis exclusively from the user's interactions with the systems, without involving him/her in this process. We suggest that a motivational diagnosis based only on the interactions with the system is incomplete despite the obvious advantages of unobtrusive diagnosis.

Moreover, we based our approach on a different theory of motivation: Social Cognitive Theory (SCT) [2] using concepts like self-efficacy (SE) and self-regulation (SR). SE is generally described by Bandura [2] as the confidence that the individual has in his/her ability to control his/her thoughts, feelings and actions; more specifically, it refers to a person's belief/expectancy in his/her capacity to successfully complete a task. SR refers to a person's ability to control his/her actions, in our case learning [24]. SCT is a sound theoretical base for motivation diagnosis as it is a well established construct in the literature. There is broad evidence that this theory has good application in classroom ([23], [25]), as well as in online learning ([12], [13]) and blended learning ([26]). The theory offers a variety of possibilities to intervene in order to motivate the learner, a framework for influencing the learner's subjective control of the task through *motivational beliefs* (SE) and *cognitive learning strategies* (SR/ self-monitoring) and also fits very well in other current research directions in e-Learning (e.g. personalization, adaptivity, affective computing, collaborative learning) [5].

We propose a two-step approach for diagnosing motivation: First, the system would unobtrusively monitor the learners diagnose their disengagement based on log files. Then, learners identified to be disengaged will be engaged in a dialog in order to assess their self-efficacy, self-regulation and other related motivation concepts. In this paper we present results related only to the first step (i.e., analysis of log files).

Communalities and Differences

Unlike previous works, our approach exploits the interaction log only as a first step towards motivation diagnosis. Our purpose is to distinguish between engaged and disengaged learners in order to focus further on the disengaged. For this purpose we analyse the data from log files.

In order to establish the user's level of engagement we used an approach similar to the one used in [6]. In this study human tutors were asked to rate several motivational characteristics (e.g. confidence, effort, cognitive/sensory interest etc.) from replays of users interactions with a system. In our research, we use the actions and the timestamps registered in log files in order to rate only one motivational aspect: the engagement level of the user.

2.2 Log Files in Research

Logging the users' interactions in educational systems gives the possibility to track their actions at a refined level of detail. Log files are easy to record for a large number of users, they can capture a large variety of information and they can also be presented in an understandable form. Thus, these data are a potentially valuable source of information to be analysed and used in educational settings. Automatic analysis of log data is frequently used to detect regularities and deviations in groups of users, in order to provide more information to tutors about the learners, and to offer suggestions for further actions, in particular for the "deviation" cases.

Log Files in Educational Research

Automatic analysis of interaction data is used in research areas such as educational systems, data mining and machine learning. Educational systems can benefit from data mining and machine learning techniques by giving meaning to click-through data and associating these data with educational information.

Log file analysis has been used for a variety of purposes: provide information to tutors to facilitate and make more accurate the feedback given to learners [19], monitor group activity [15], identify benefits and solve difficulties related to log data analysis [11], use response times to model student disengagement [3], infer attitudes about the system used, attitudes that affect learning [1], developing tools to facilitate interpretation of log files data [20].

Log Files in Motivation Research
Previous research in this area includes a few interesting approaches. A model for detecting learners' engagement [3] was proposed in order to detect whether a student is engaged in answering questions based on item response theory (IRT). The input of the model was: difficulty of the question, how long the student took to respond and whether the response was correct. The output (obtained from the modified IRT formula) was the probability that a student was actively engaged in trying to answer a question. A second approach [9] related to user interests and motivation inferred from server log files, argued that time spent on pages is more important than simple "hits". Usually, the way of determining user's interest is to log the number of "hits" received per page. The author argues that this is inadequate because the browser will log "hits" not only for the page of interest, but also for every page the user visited to get there. He argues that a path independent measure of user interest is needed and that a time-based measure would be such a measure.

3 Log File Analysis

For the analysis presented here, we created several data sets from existing log files. The level of the learner's engagement was rated by an expert. Eight different data mining methods were applied predicting the engagement level from the log data.

3.1 Log File Description

In our analysis we used log files from a system called HTML tutor, which is a web interactive learning environment based on NetCoach [27]. It offers an introduction to HTML and publishing on the Web; it is online and can be accessed freely, based on a login and a password. We don't have any information about the users except the data from the log files. They could be of any age and using the system for different purposes. Table 1 presents the events registered in the log files and the attributes for each event that were included in the log file analysis.

In a previous experiment [4] with a limited number of data, using the total time spent on a session (i.e., between login and logout) as attribute, the analysis showed that it is possible to judge whether a learner was engaged or disengaged only after 45 minutes; the same analysis showed that most of the disengaged users left the system before that time. In order to overcome this problem, we decided to use for the following experiments sequences of 10 minutes instead of complete sessions. Thus, we split the sessions into sequences of 10 minutes; 943 sequences of 10 minutes and 72 sequences varying between 7 and 592 seconds resulted from this process.

Table 1. Logs events registered and the attributes analysed per session / sequence respectively

Events	Parameters/ Attributes
Goal	The selected goal (from a list of 12 goals)
Preferences	Number; Time spent selecting them
Reading pages	Number of pages; average time reading pages
Pre-tests	Number of pre-tests; average time; number of correct answers; number of incorrect answers
Tests	Number of tests; average time; number of correct answers, number of incorrect answers
Hyperlink, Manual, Help, Glossary, Communication, Search, Remarks, Statistics, Feedback	For each of these: Number of times accessed; average time

Besides the attributes related to events, the data set contains a few more fields: a user ID, a session ID, a sequence ID and total time of the sequence. The number of entries in the data set is 1015, obtained from 48 subjects who spent on HTML tutor between 1 and 7 sessions, each session varying between 1 and 92 sequences. The events/attributes frequencies are displayed in Table 2.

Table 2. Frequency of events registered in log files

Events/attributes	Frequency of appearances (in 1015 sequences)
Goal	59
Preferences	7
Reading pages	850
Pre-tests	14
Tests	458
Hyperlinks	245
Manual	7
Help	11
Glossary	76
Communication	6
Search	27
Remarks	6
Statistics	8
Feedback	4

3.2 Expert ratings on Level of Engagement

For each sequence of 10 minutes a value/code was assigned: engaged (e), neutral (n) and disengaged (d). In the previous experiment [4] we had only 2 categories: engaged and disengaged. Because we introduced the 10 minutes sequences, in some cases it was hard to decide whether overall the learner was engaged or disengaged. Thus, we introduced a third category: neutral. A detailed presentation of the criteria used for this rating is presented in Table 3, which contains the instructions given to a second coder in order to verify the reliability of the ratings.

The investigation conducted in order to verify the coding reliability included two steps: 1) *Informal assessment*, conducted using only 10 sequences; the ratings based on the given instructions were discussed to prevent different results due to instruction vagueness or suggestibility; the percent agreement was 80% (only 2 different ratings from 10); the kappa measurement of agreement was .60 (p=.038) and the Krippendorff's alpha [10] was .60 as well; 2) *Second expert rating.* A second rater coded 100 sequences randomly sampled from the 1015 entries in the data set; the instructions used for the informal assessment were expanded with typical situations/ patterns for each case. Table 3 includes the instructions given to the second rater (instructions used also for coding all sequences).

Table 3. Instructions for level of engagement rating.

Timeframes for HTML Tutor - Necessary time for reading a page: varies from 30 sec. to a maximum of 4-5 minutes. - Necessary time for a test: varies from just a few seconds to a maximum of 3-4 minutes.		
Engaged (e)	**Disengaged (d)**	**Neutral (n)**
Spending reasonable time on pages and tests given the characteristics of HTML Tutor	Spending too much time on pages/tests Moving fast through pages/tests	Hard to decide if overall (for the 10 minutes) the person is engages or disengaged
Examples of patters: - people focused reading – spend most of the time reading and less on other tasks - people focused on taking tests - spend most of the time taking tests and less on other tasks - people that read and take tests - spend most of the time reading and taking tests	Automatic logouts Examples of patterns: - spend more than reasonable time on just one or a few tasks - move fast though the same / different tasks	E.g.: for approximately half of the time the person seems engaged and for the other half seems disengaged E.g.: can't decide if overall the person is moving too fast through pages or the amount of time spent on pages is reasonable

The second expert rating resulted in a rater agreement of 92% (only eight different ratings from 100; in further discussion between the raters the eight disagreements were resolved) with a kappa measurement of agreement of .826 (p<.01) and Krippendorff's alpha of .8449. Although the percent agreement is high, we can see that kappa and Krippendorff's alpha have lower values. The percent agreement is not always the best indicator for agreement as it tends to be too liberal, while Cohen's Kappa and Krippendorff's alpha are known to be more conservative [17]. Thus, overall, the values indicate high inter-coder reliability.

3.3 Analysis and Results

In order to perform the analysis, Waikato Environment for Knowledge Analysis (WEKA) [28] was used. Several methods were experimented to find which one is best for our purpose and to see if results are consistent over several methods. We present here trials used only on a reduced data set of 943 entries obtained from the 1015 entries data set by eliminating the entries with time per sequence shorter than 10 minutes. In order to explore the effect of the number of attributes included, we created three different data sets: 1) all 30 attributes except user ID called DS-30; 2) 10 attributes related to the following events: reading pages, tests, hyperlinks and glossary (DS-10) and 3) six attributes related only to reading pages and tests. The experiment was done using 10-fold stratified cross validation iterated 10 times.

The analysis included eight methods [28]: (a) Bayesian Nets with K2 algorithm and maximum 3 parent nodes (BN); (b) Logistic regression (LR); (c) Simple logistic classification (SL); (d) Instance based classification with IBk algorithm (IBk); (e) Attribute Selected Classification using J48 classifier and Best First search (ASC); (f) Bagging using REP (reduced-error pruning) tree classifier (B); (g) Classification via Regression (CvR) and (h) Decision Trees with J48 classifier based on Quilan's C4.5 algorithm [22] (DT).

The results are displayed in Table 4, which comprises the percentage of correctly classified instances, the true positives (TP) rate, the precision indicator and recall for disengaged class, and the mean absolute error.

Table 4. Experiment results summary

		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30	**%correct**	87.07	86.52	87.33	85.62	87.24	87.41	87.64	86.58
	TP rate	0.93	0.93	0.93	0.92	0.93	0.93	0.92	0.93
	Precision	0.91	0.90	0.90	0.91	0.92	0.92	0.92	0.91
	Recall	0.93	0.93	0.93	0.92	0.93	0.93	0.92	0.93
	Error	0.10	0.12	0.12	0.10	0.10	0.12	0.12	0.11
DS-10	**%correct**	87.18	85.88	85.82	85.13	86.03	86.87	88.07	85.16
	TP rate	0.93	0.93	0.93	0.91	0.92	0.92	0.91	0.91
	Precision	0.91	0.89	0.89	0.92	0.91	0.91	0.92	0.90
	Recall	0.93	0.93	0.93	0.91	0.92	0.92	0.92	0.91
	Error	0.11	0.13	0.14	0.10	0.12	0.13	0.12	0.13
DS-6	**%correct**	86.68	84.15	84.05	83.18	86.95	86.90	87.21	86.20
	TP rate	0.93	0.93	0.93	0.90	0.92	0.92	0.91	0.92
	Precision	0.90	0.87	0.87	0.90	0.92	0.91	0.92	0.91
	Recall	0.93	0.93	0.93	0.90	0.92	0.92	0.91	0.92
	Error	0.12	0.15	0.15	0.12	0.12	0.13	0.13	0.13

Table 4 shows very good prediction for all methods with a correct prediction varying approximately between 84% and 88%. Even better results are shown by the TP rate, precision and recall indicator for disengaged class: values between 87% and 93%. The mean absolute error varies between 0.10 and 0.15. The very similar results obtained from different methods and trials shows consistency of prediction and of the attributes used for prediction.

The highest percentage of correctly predicted instances was obtained using Classification via Regression (CvR) on all data sets, with a maximum for DS-10. This indicates that the attributes that predict the learner's engagement/ disengagement most accurately are the one related to reading pages, taking tests, following hyperlinks and consulting the glossary. The percentage for DS-6 is slightly lower (87.21%), suggesting that hyperlinks and glossary events do not have a big contribution to the prediction model. The confusion matrix for this result is displayed in Table 5.

Table 5. The confusion matrix for data set DS-6 using CvR

		Predicted		
		Disengaged	**Engaged**	**Neutral**
Actual	**Disengaged**	610	56	0
	Engaged	35	218	0
	Neutral	13	11	0

If we focus on the disengaged learners we see that Bayesian Nets (displayed in Fig. 3) have the best performance on all data sets: 93%, even if the percentage of correctly predicted instances for all three classes varies between data sets.

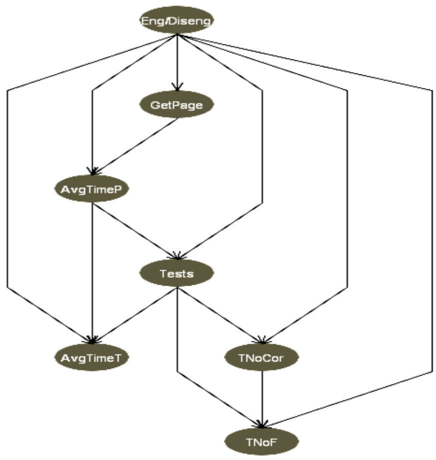

Fig 1. Bayesian Network from data set DS-6

The Bayesian Network from DS-6 has an interesting structure: Number of False (TNoF) and Correct (TNoCor) Answers to Tests feed into the Number of Tests (Tests). Which itself, together with Average Time on Tests (AvgTimeT) feeds into Average Time spend on Pages (AvrTimeP). All of them also feed directly into the Level of Engagement (Eng/Diseng), i.e., the Bayesian Network structured the attributes in a semantically meaningful way.

As good results were obtained for all trials with small differences between them, considering the MDL (minimum description length) principle we argue for the use of a minimum number of attributes: the six attributes used in DS-6. These attributes were ranked first from all attributes using information gain attribute evaluation. There are no particular attributes that give bad performance, but many of them do not contribute to prediction and thus, removing them does not affect the prediction performance, as we can see from the similarity of results from the three data sets. Thus, the attributes related to reading pages and taking tests are most valuable and as they are common to most e-Learning systems and most frequent actions that occur in using such systems, we argue for a prediction model that includes only these attributes.

4 Summary, Implication and Future Perspectives

We presented results of eliciting motivation knowledge from log files. The analysis showed good overall prediction e.g. 87% using classification via regression and even better value for prediction of disengagement e.g. 93% using Bayesian Nets. The analysis included 943 sequences of 10 minutes from 48 users, showing that a general indicator of the motivational level could be predicted from very basic data commonly recorded in log files, such as events related to reading pages and taking tests. A prediction module could be included in educational systems that log learner's actions. Our research plan includes further elicitation of motivation to be included in a user model in order to have a system that adapts to the motivational level of the learner.

Further work includes an external validation of the expert rating and an analysis of log files from a different system in order to compare the results. We also plan a pilot study in order to compare the dialog responses of learners with their responses on questionnaires, to verify the validity and reliability of the measurement using the dialog.

Acknowledgment. Part of this work was supported by the European Commission through the project "Adaptive Learning Spaces" (229714-CP-1-2006-NL-MINERVA-M).

References

1. Arroyo, I., Murray, T., Woolf, B.P.: Inferring unobservable learning variables from students' help seeking behavior. In Proceedings of the workshop Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes at ITS (2004) 29-38
2. Bandura, A.: Social foundations of thought and action: A social cognitive theory. Englewood Cliffs, NJ: Prentice Hall (1986)
3. Beck, J.E.: Using response times to model student disengagement. In Proceedings of the Workshop on Social and Emotional Intelligence in Learning Environments at ITS (2004).
4. Cocea, M., Weibelzahl, S.: Can Log Files Analysis Estimate Learners' Level of Motivation? In Schaaf, M. and Althoff, K-D. LWA 2006. Lernen-Wissensentdeckung – Adaptivität. Universität Hildesheim (2006) 32-35
5. Cocea, M., Weibelzahl, S.: Motivation–Included or Excluded From E-Learning, In Kinshuk, Sampson, D. G., Spector J.M., Isaias, P., Cognition and Exploratory Learning in Digital Age (2006) 435-437.
6. De Vicente, A., Pain, H.: Informing the Detection of the Students' Motivational State: an empirical Study. In Cerri, S. A.; Gouarderes, G.; Paraguau, F: Intelligent Tutoring Systems, 6th International Conference. Berlin: Springer-Verlag., (2002) 933-943
7. D'Mello, S. K., Craig, S. D., Gholson, B., Franklin, S., Picard, R. W., Graesser, A. C.: Integrating Affect Sensors in an Intelligent Tutoring System, In Affective Interactions: The Computer in the Affective Loop Workshop at 2005 International conference on Intelligent User Interfaces New York: AMC Press (2005) 7-13
8. Fernandez, R., Picard, R. W.: Classical and Novel Discriminant Features for Affect Recognition from Speech, Interspeech 2005 - Eurospeech 9th European Conference on Speech Communication and Technology pp. September 4-8, Lisbon Portugal (2005)
9. Fuller, R., de Graaff, J.J.: Measuring User Motivation from Server Log Files. Retrieved on 14/03/06 from http://www.microsoft.com/usability/webconf/fuller/fuller.htm

10. Hayes, A. F., Krippendorff, K. (in press): Answering the call for a standard reliability measure for coding data. Communication Methods and Measures. Retrieved on 8/11/2006 from http://www.comm.ohio-state.edu/ahayes/
11. Heiner, C., Beck, J., Mostow, J.: Lessons on Using ITS Data to Answer Educational Research Questions. In Proceedings of the workshop Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes at ITS (2004) 1-9
12. Hodges, C. B.: Designing to Motivate: Motivational Techniques to Incorporate in E-Learning Experiences, The Journal of Interactive Online Learning, Vol. 2, No 3 (2004)
13. Irizarry R.: Self-Efficacy & Motivation Effects on Online Psychology Student Retention. USDLA Journal, Vol. 16, No 12. (2002)
14. Kapoor, A., Picard, R.W., Ivanov, Y.: Probabilistic Combination of Multiple Modalities to Detect Interest, International Conference on Pattern Recognition, Cambridge, U.K. (2004)
15. Kay, J., Maisonneuve, N., Yacef, K., Zaïane, O.: Mining patterns of events in students' teamwork data. In Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems, Jhongli, Taiwan (2006) 45-52
16. Keller, J. M.: Development and use of the ARCS model of instructional design. Journal of Instructional Development, 10(3), (1987) 2-10
17. Lombard, M., Snyder-Duch, J., Campanella Bracken, C.: Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research (2003). Retrieved on 06/11/06 from http://www.temple.edu/mmc/reliability
18. Martinez, M.: High Attrition Rates in e-Learning: Challenges, Predictors, and Solutions. The e-Learning Developers Journal, July (2003). Retrieved on 25/02/2006 from http://www.elearningguild.com
19. Merceron, A., Yacef, K.: A Web-Based Tutoring Tool with Mining Facilities to Improve Learning and Teaching. In Proceedings of the 11th International Conference on Artificial Intelligence in Education AIED, IOS Press (2003) 201-208
20. Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner C.: An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell! In Educational Data Mining: Papers from the 2005 AAAI Workshop, ed. Joseph E. Beck, pages 15-22.
21. Qu, L., Wang N., Johnson, W. L.: Detecting the Learner's Motivational States in an Interactive Learning Environment. AIED, C.-K. Looi et al.(Eds.), IOS Press (2005) 547-554
22. Quinlan, R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA (1993)
23. Schraw, G., Brooks, D. W.: Helping Students Self-Regulate in Math and Sciences Courses: Improving the Will and the Skill (2000). Retrieved on 14/03/06 from http://dwb.unl.edu /Chau/SR/Self_Reg.html
24. Schunk, D. H., & Zimmerman, B. J.. Self-regulation of learning and performance: Issues and educational applications. Hillsdale, NJ: Lawrence Erlbaum (1994)
25. Tuckman, W. B.: A Tripartite Model of Motivation for Achievement: Attitude/Drive/ Strategy (1999) Retrieved on 14/03/06 from http://dwb.unl.edu/Chau/CompMod.html
26. Wang, A. Y., Newlin, M. H.: Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class. Computers in Human Behavior, 18 (2002)151-163
27. Weber, G., Kuhl, H.-C., Weibelzahl, S.: Developing adaptive internet based courses with the authoring system NetCoach. In Hypermedia: Openness, Structural Awareness, and Adaptivity (LNAI 2266), Springer, Berlin (2001) 226-238.
28. Witten, I.H., Frank, E.: Data mining. Practical Machine Learning Tools and Techniques. Second Edition, Morgan Kauffman Publishers, Elsevier (2005)
29. Zhang, G., Cheng, Z., He, A., Huang, T. A WWW-based Learner's Learning Motivation Detecting System. Proceedings of International Workshop on "Research Directions and Challenge Problems in Advanced Information Systems Engineering", Honjo City, Japan, (2003). Retrieved on 14/03/06 from http://www.akita-pu.ac.jp/ system/KEST2003