

The Simple Location Heuristic is Better at Predicting Students' Changes in Error Rate Over Time Compared to the Simple Temporal Heuristic

A.F. NWAIGWE

AMERICAN UNIVERSITY OF NIGERIA, NIGERIA

AND

K.R. KOEDINGER

CARNEGIE MELLON UNIVERSITY, U.S.A

In a previous study on a physics dataset from the Andes tutor, we found that the simple location heuristic was better at making error attribution than the simple temporal heuristic when evaluated on the learning curve standard. In this study, we investigated the generality of performance of the simple location heuristic and the simple temporal heuristic in the math domain to see if previous results generalized to other Intelligent Tutoring System domains. In support of past results, we found that the simple location heuristic provided a better goodness of fit to the learning curve standard, that is, it was better at performing error attribution than the simple temporal heuristic. One observation is that for tutors where the knowledge components can be determined by the interface location in which an action appears, using the simple location heuristic is likely to show better results than the simple temporal heuristic. It is possible that the simple temporal heuristic is better in situations where the different problem subgoals can be associated with a single location. However, our prior results with a physics data set indicated that even in such situations the simple location heuristic may be better. Further research should explore this issue.

Key Words and Phrases: Error attribution methods, Intelligent Tutoring Systems, learning curves, mathematics

1. INTRODUCTION

Increasingly, learning curves have become a standard tool for evaluation of Intelligent Tutoring Systems (ITS) [Anderson, Bellezza & Boyle, 1993; Corbett, Anderson, & O'Brien, 1995; Koedinger & Mathan, 2004; Martin, Mitrovic, Mathan, & Koedinger, 2005; Mathan & Koedinger, 2005; Mitrovic & Ohlsson, 1990] and measurement of students' learning [Anderson, Bellezza & Boyle, 1993; Heathcote, Brown, & Mewhort, 2002]. The slope of learning curves show the rate at which a student learns over time, and reveals how well the tutor's cognitive model fits what the student is learning. However, these learning curves require a method for attributing error to the "knowledge components" (skills or concepts) in the student model that the student is missing. Knowledge components, concepts and skills will be used interchangeably in this paper. In a previous study using data from the Andes Intelligent tutor [VanLehn et al., 2005], four

Authors' addresses: A.F. Nwaigwe, School of Information Technology and Communication, American University of Nigeria, Nigeria; E-mail : adaeze.nwaigwe@aun.edu.ng; K.R. Koedinger, Human Computer Interaction Institute, Carnegie Mellon University, U.S.A.; E-mail : koedinger@cmu.edu

alternative heuristics were evaluated - simple location heuristic (LH), simple temporal heuristic (TH), model-based location heuristic (MLH) and model-based temporal heuristic (MTH) [Nwaigwe et al., 2007]. When evaluated on the learning curve standard, the two location heuristics LH and MLH, outperformed the temporal heuristics, TH and MTH. However, the generality of performance of these heuristics in other ITS subject domains needs to be tested.

In this study conducted in the mathematics domain, we investigated whether the previous performance of the LH and TH generalized to other ITS domains. We specifically asked if the LH was better than the TH at predicting student changes in error rate over time. We used log data from a Cognitive Tutor on a Scatterplot lesson and implemented the learning curves standard using the statistical component of Learning Factors Analysis [Cen, Koedinger & Junker, 2005; Pirolli & Wilson, 1998].

Our intuition is that the LH may be the better choice for error attribution when knowledge components (KCs) can be determined by the interface location where an action occurs. To justify this, imagine that a worker has homes, H_a and H_b in which to perform tasks A and B respectively. The worker goes to home H_a and attempts task A but fails. The worker abandons the failed task A and goes to home H_b , where he/she succeeds at task B. The assumption is that tasks A and B are associated with different KCs. The worker later returns to location H_a , and this time, is successful at task A. The LH will more rationally attribute the initial failed attempt at H_a to the KC associated with task A since its rule is to attribute error to the first successfully implemented KC at the initial error location. The TH will however, wrongfully put blame on the KC associated with task B since its method of error attribution is to blame the KC associated with the first correctly implemented task.

Sometimes, TH might be a better choice for making error attribution. We believe this to be the case when it is necessary to perform a set of tasks in a prescribed sequence. To elaborate, imagine that homeschooler Bella is required to perform two tasks and in the given sequence – eat breakfast (EB), and do schoolwork (DS) and in any of two locations, L1 and L2 on the dining table of the family’s apartment. We again assume that tasks EB and DS are associated with different KCs. Bella decides that she did not like what Mom served for breakfast that morning and goes straight to her schoolwork, DS, at location L1, skipping task EB. However, Bella fails at task DS due to hunger associated distractions. Later, she abandons task DS and revisits and succeeds at task EB at location L2. Bella then goes back location to L1 and completes task DS. In attributing blame, TH will rationally blame the KC associated with task EB. However the LH will wrongfully blame the KC associated with task, DS. These examples imply that it may be better to apply heuristics in making error attribution.

Although an immediate purpose for error attribution is to drive learning curve generation, the assignment of blame problem is more general and affects many aspects of student modeling.

2. ERROR ATTRIBUTION HEURISTICS

A basic assumption of many cognitive models is that knowledge can be decomposed into components, that each component is learned independently of the others and that implementation of a step in the solution of a problem is an attempt to apply one or more knowledge components (KCs). When correct solution steps are generated, either by an expert system or a human expert, the step is often annotated with the KCs that must be applied in order to generate the step. Thus, when a student enters that step, the system can infer that the student is probably (but not necessarily) applying those KCs.

An ITS system can be designed to anticipate and generate some incorrect steps and associated goals, however, it is rare for expert systems or expert authors to anticipate and generate a large number of incorrect steps and corresponding goals. Hence, when the student enters an incorrect step, it is often not clear what KC(s) should have been applied, so the system cannot determine which KC(s) the student is weak on. If the system simply ignores incorrect steps, then it only “sees” successful applications of KCs. It cannot “see” failures of a KC. It may see lots of incorrect steps, but it cannot determine and record what KC(s) to blame for each error [VanLehn et al, 2005] and so, learning curves cannot be generated. This suggests using heuristics.

The tutoring system usually has two clues available: the location of the incorrect step on the user interface and the subsequent steps entered by the student. For instance, if a student makes an error on a step at time 1 and at location A, the student will often attempt to correct it immediately, perhaps with help from the tutor. So if the first correct step, at time 2 is also at location A, and say, that the step is annotated with KC x, then it is likely that the incorrect step at time 1 was a failed attempt to apply KC x. This heuristic allows the system to attribute errors to KCs whenever the system sees a correct step immediately following the target incorrect step, and both steps are in the same location on the user interface.

However, it is not clear how to generalize this heuristic. What if the next correct step is not in the same location? What if there are intervening incorrect steps in different locations? In previous work using data from the Andes Physics Tutor, four automated heuristics for making error attribution (LH, TH, MLH, MTH) were proposed and evaluated guided by whether the heuristic was driven by location or by the temporal order of events [Nwaigwe et al, 2007].

For every error transaction, LH attributes blame to the KC mapped to a subsequent correct entry at the widget location where the error occurred [Anderson, Bellezza & Boyle, 1993; Koedinger & Mathan, 2004; Martin, Mitrovic, Mathan, & Koedinger, 2005] while the TH ascribes blame to the KC that labels the first correct entry in time. When there is no subsequent correct entry with a label of the error location, LH blames the KC with the first correct entry in time, that is, it implements the behavior of TH. When the tutor provides a choice of some KC to blame for an error, the MLH goes with the tutor’s choice otherwise, it simply implements the LH. For an error transaction, MTH also goes with the domain model’s choice if one exists, otherwise it implements the TH.

In this work, we examine the performance of the LH and TH in the mathematics domain. Table I shows sample log transaction from the cognitive tutor for the scatterplot lesson. The table illustrates how the LH and the TH can help resolve the error attribution ambiguity. Columns in table 1 are described thus: “location” column indicates the place on the interface (the interface widget) in which the student made an input; “Outcome” indicates if an input is correct or not, while “Student Model KC” lists the system’s choice of KC which the student should implement.

In row 1, the student makes an error at the location labeled, “var-Oval-1”. The system however does not indicate the KC the student ought to be practicing. To resolve this ambiguity, the LH uses the KC that labels a subsequent correct entry in the same location – see row 5. That is, it chooses “choose variable”. On the other hand, the TH chooses the KC that labels the first correct entry in time, irrespective of interface location. Its choice is “label x-axis”. In row #2, the domain model blames the KC “choose variable” for the student’s error. LH chooses “choose variable” since it is the first correctly implemented KC at the location “var-Oval-1”. TH blames the KC “label x-axis” in this case.

In the prior study, the cognitive model generated by the LH was found to outperform that of the TH and also, the tutor’s original model according to the learning curve

standard. In other words, the LH was better at making error attributions than the other two cognitive models. Compared to the TH, we also found that the error attribution method of the LH was more like that made by human coders. In this work, we conduct our analysis in the math domain and compare the performance of the LH to that of the TH based on the learning curve standard. Our goal is to see if the previous performance of the LH and TH can be generalized to other intelligent tutoring system domains.

Table I. Table illustrating different error attributions made by the 2 methods

#	Location	Outcome	Student Model KC	KC Error Attributions methods	
				LH	TH
1	var-0val-1	incorrect		choose variable	label x-axis
2	var-0val-1	incorrect	choose variable	choose variable	label x-axis
3	var-1val-1	correct	label x-axis	label x-axis	label x-axis
4	var-0val-1	incorrect		choose variable	choose variable
5	var-0val-1	correct	choose variable	choose variable	choose variable

3. LEARNING CURVES

Learning curves plot the performance of students with respect to some measure of their ability over time [Anderson, Bellezza & Boyle, 1993; Corbett, Anderson, O'Brien, 1995; Koedinger & Mathan, 2004; Martin, Mitrovic, Mathan, & Koedinger, 2005; Mathan & Koedinger, 2005; Mitrovic & Ohlsson, 1990]. For ITSs, the standard approach is to measure the proportion of knowledge components in the domain model that have been "incorrectly" applied by the student. This is also known as the "error rate". Other alternatives exist, such as the number of attempts taken to correct a particular type of error. Time is generally represented as the number of opportunities to practice a KC or skill. This in turn may be determined in different ways: for instance, it may represent each new step a student attempts that is relevant to the skill, on the basis that repeated attempts at the KC are benefiting from the student having been given feedback and as-needed instruction about that particular skill and hence may improve from one attempt to the next. If the student is learning the KC or skill being measured, the learning curve will follow a so-called "power law of practice" [Mathan & Koedinger, 2005]. If such a curve exists, it presents evidence that the student is learning the skill being measured or conversely, that the skill represents what the student is learning.

3.1. THE LEARNING CURVES STANDARD

The power law applies to individual skills and does not take into account student effects. The statistical component of Learning Factors Analysis (LFA) extends the power law to a logistic regression model which accommodates student effects for a cognitive model incorporating multiple knowledge components and multiple students [Cen, Koedinger, & Junker, 2005], see equation 1. The following are the assumptions on which equation 1 is based:

1. Different students may know more or less initially. An intercept parameter of this model reflects each student's initial knowledge.

2. Students learn at the same rate. Thus, slope parameters do not depend on the student. Slope parameters reflect the learning rate of each KC which the student model encompasses and are independent of student effect. This assumption made so as to reduce the number of parameters in equation 1 and is further justified since equation 1 is focused on refining a cognitive model rather than on evaluating students' knowledge growth [Draney, Pirolli & Wilson, 1995].
3. Some KCs are more likely known than others. An intercept parameter for each KC captures initial difficulty of the skill.
4. Since some KCs are easier to learn than other, the model of equation 1 uses a slope parameter to reflect this for each skill. Larger values for initial difficulty reflect tougher skills.

$$\ln[p/(1-p)] = \sum \alpha_i X_i + \sum \beta_j Y_j + \sum \gamma_j Y_j T_j. \quad (1)$$

where p – the probability of success at a step performed by student i that requires knowledge component j ; X_i and Y_j – the dummy variable vectors for students and knowledge components respectively; T_j – the number of practice opportunities student i has had on knowledge component j ; α_i – the coefficient that models student i 's initial knowledge; β_j – the coefficient that reflects the initial difficulty of knowledge component j where larger values of initial difficulty reflect tougher skills; γ_j – the coefficient that reflects the learning rate of knowledge component j , given its practice opportunity.

In this paper, the model of equation 1 is used to apply the learning curve standard. Bayesian Information Criterion (BIC) [Wasserman, 2004] is used to estimate prediction risk in the model while loglikelihood is used to measure model fit. Lower BIC scores, mean a better balance between model fit and complexity.

4. DATA SOURCE

The data used for this research was collected as part of a study conducted in a set of 5 middle-school classrooms at 2 schools in the suburbs of a medium-sized city in the Northeastern United States. Student ages ranged approximately from 12 to 14 years. The classrooms studied were taking part in the development of a new 3-year cognitive tutor curriculum for middle school mathematics [Baker, 2005; Baker., Corbett, Koedinger & Wagner, 2004]. Data collected was from the study on these classrooms during the course of a short (2 class periods) cognitive tutor unit on scatterplot generation and interpretation. Scatterplots depict the relationship between two quantitative variables in a Cartesian plane, using a point to represent paired values of each variable.

The scatterplot lesson consisted of a set of problems and for each problem, a student was given a data set to generate a graph. The student then had to choose from a list, the variables that were appropriate for use in the scatterplot (see figure 1); those that were quantitative or categorical; and subsequently whether a chosen variable was appropriate for a bar chart.

Next the student was required to label the X and Y-axis (see figure 2), and to choose each axis bound and scale. The student was then required to plot points on the graph by clicking on the desired position on the graph. Finally, the student was required to answer a set of interpretation questions to reason about the graph's trend, outliers, monotonicity, and extrapolation and in comparison with other graphs. In our dataset, students solved a maximum of six problems and a minimum of two in the scatterplot lesson.

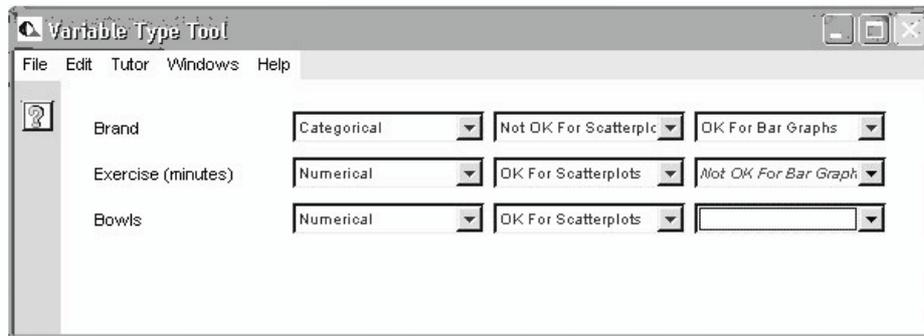


Figure 1 Scatterplot lesson interface for choosing variable type [Baker, 2005]

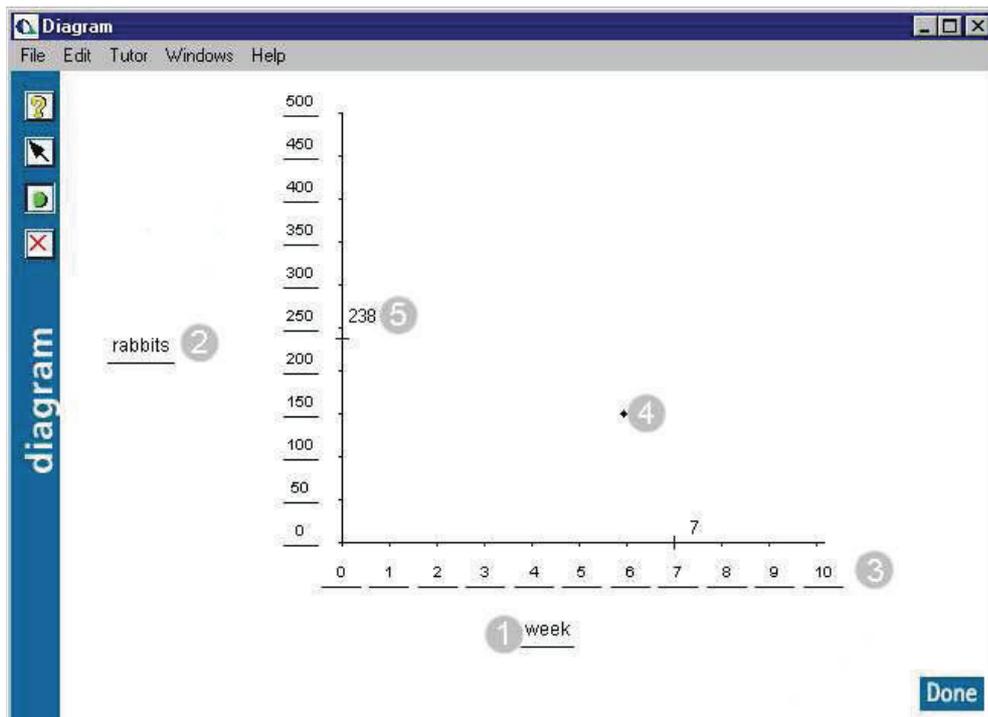


Figure 2 Interface for graph creation in the scatterplot lesson [Baker, 2005]

5. METHODOLOGY

The algorithms for the LH and TH used in this research was implemented in pure java 1.6 and designed to process student log data in MS Excel format. Both algorithms used sequential search. Log data from the cognitive tutor unit on scatterplot generation and interpretation served as input to the programs. The output from each program was the choice of KC codes made by the heuristic being implanted as explained in section 2.

To analyze the cognitive model of each heuristic according to the learning curve standard, the data output from each program was then fit to equation 1 to derive learning behavior. The coefficients of equation 1, initial KC difficulty (β_j), initial student difficulty (α_i) and KC learning rate (γ_j) were used to describe learning behavior for each heuristic. If the intercept of a KC was higher, then, its initial difficulty was lower. Further, if the slope of each KC was higher, then, the faster students learned that skill. For the model of each heuristic, BIC score was used to estimate prediction risk while loglikelihood was used to measure model fit.

6. RESULTS AND DISCUSSION

Table II summarizes the results of the learning curve standard for the student models for both the LH and TH. The results show that the simple location heuristic, LH (BIC score: 7,510.12) shows better fit to the learning curve standard compared to the simple temporal heuristic, TH (BIC scores: 7,703.58). This means that the model of the LH is more reliable and so, a prediction error is more likely to occur if one used the TH model. Loglikelihood score was also better for the LH (-3,370.37) than for the TH (-3,464.93), indicating that the LH model was a better fit to the data than the competing TH model. This shows how the different error attribution methods affect the result.

Table II. Results of the Learning Curve Standard

		TH	LH
logLikelihood		-3,464.93	-3,370.37
BIC		7,703.58	7,510.12
Learning Rate (γ_j)	Mean (Std)	0.09 (0.09)	0.133 (0.11)
Initial KC Difficulty (β_j)	Mean (Std)	-1.81 (0.94)	0.08 (1.10)
Initial Student Difficulty (α_i)	Mean (Std)	2.03 (0.61)	-0.00 (0.63)
# of KCs		17	17
# of transactions across entire scatterplot lesson		16,291	16,291
# of students		52	52

Table III. Knowledge Component Details for the two Heuristics

Knowledge Component (KC)	Simple Temporal Heuristic (TH)			Simple Location Heuristic (LH)		
	Ave Opp	β_j (Initial difficulty)	γ_j (learning rate)	Ave Opp	β_j (Initial difficulty)	γ_j (learning rate)
CHOOSE-VAR-TYPE-CAT	6.6	-1.449	0.076	6.6	0.048	0.244
MMS-VALUING-DETERMINE-SET-MAX	6.9	-0.793	0	6.2	1.275	0
MMS-VALUING-DETERMINE-SET-MIN	6.3	-0.361	0.031	6.2	1.587	0.063
QUANTITATIVE-VALUING-FIRST-BIN	6.1	-2.642	0.159	5.5	-0.565	0.163
QUANTITATIVE-VALUING-SECOND-BIN	5.6	-0.947	0	5.4	0.879	0.052
MMS-VALUING-LABELSUSED	6.9	-2.625	0.049	5.8	-0.942	0.219
CHOOSE-VAR-TYPE-NUM	18.2	-1.044	0.038	16.2	0.799	0.044
MMS-VALUING-DETERMINE-SCALE	53.1	-0.069	0.007	50.2	2.364	0
MMS-VALUING-LABELSUSED-PLUS2	5.9	-1.99	0.063	5.8	-0.805	0.187
TEST-SLOPE	3.3	-2.213	0.131	3.3	0.238	0
CHOOSE-OVERALL-REL	5.0	-3.175	0.257	5.2	-0.865	0.149
EXTRAPOLATE	1.7	-2.093	0	1.5	0.206	0
CHOOSE-OK-BG	11.5	-1.708	0.198	11.4	0.263	0.215
CHOOSE-X-AXIS-QUANTITATIVE	4.3	-2.572	0.274	3.2	-0.719	0.314
CHOOSE-Y-AXIS-QUANTITATIVE	3.7	-2.618	0.018	3.2	-1.884	0.276
MMS-VALUING-DETERMINE-MIN	6.3	-3.053	0.119	5.8	-1.106	0.139
MMS-VALUING-DETERMINE-RANGE	6.3	-1.357	0.147	6.1	0.584	0.196

Generally, we observed that, the LH performed better than the TH when the student failed to successfully complete an attempted step and subsequently attempted and succeeded at a different step. As shown in table I, the student unsuccessfully attempted a step at location “var-0val-1” (trn # 1 & 2). The student subsequently went to location “var-1val-1”, attempted and succeeded at the new step. While the TH incorrectly blamed “label x-axis” which is the KC associated with the new step at location “var-1val-1”, the LH more rationally blamed “choose variable” which is the KC that should be associated with the step at location “var-0val-1”. Because the LH uses location for error attribution, it correctly assigns blame to the KC associated with the error. TH however, wrongfully blames the first subsequent KC that the student correctly attempts. Of the 16,291

transactions in our dataset, error transactions recorded were 5,733. Of the latter, the LH and TH differed on 1,583 (36%) transactions with respect to error attribution choices.

We also found that both the LH and the TH had the tendency to yield the same result when the student succeeded at a step, even after multiple attempts, prior to attempting and succeeding at a new step. This was the case 64% of the time.

In table III, average practice opportunity, initial KC difficulties and learning rates are given for KCs and used to describe learning behavior for each heuristic. For example, for the KC “CHOOSE-VAR-TYPE-CAT”, the learning rate (γ_j) for the LH was more than 3 times that of the TH. Judging by KC initial difficulty (β_j), “CHOOSE-VAR-TYPE-CAT” appeared more difficult for the model of the TH (-1.449) than for the model of the LH (0.244). The average practice opportunity measured for that skill (6.6), was the same for each heuristic. The latter means that on the average, each student had approximately 7 opportunities to practice the KC “CHOOSE-VAR-TYPE-CAT”.

From table III, for the most part, KC learning rate was higher for the skills in the cognitive model of the LH compared to that for the TH. The trend for initial KC difficulty was in the opposite direction as seen for KCs such as “MMS-VALUING-DETERMINE-SET-MIN”, “QUANTITATIVE-VALUING-SECOND-BIN”, etc. Generally, KCs in the cognitive model for TH appeared more difficult to students initially, when compared to similar KCs in the cognitive model of the LH.

From table II, the mean learning rate for the LH was 0.133(\pm 0.11) which evaluated higher than that of the TH, 0.09(\pm 0.09). The mean initial KC difficulty for the LH and TH were 0.08(\pm 1.1) and -1.84(\pm 0.94) respectively. The reason for the latter seems to be due to more errors being attributed to later opportunities in the TH than the LH. These results thus illustrate the effects of error attribution.

7. CONCLUSION

In this paper, we investigated the generality of performance of two alternative methods for making error attribution in intelligent tutoring systems - the simple location heuristic and the simple temporal heuristic. Our study was carried out in the mathematics domain using data from a cognitive tutor unit on scatterplot generation and interpretation. In support of previous results obtained in the physics domain, we found that the simple location heuristic was better at predicting students' changes in error rate over time compared to the simple temporal heuristic. This work shows that simpler, easier-to-implement methods can be effective in the process of making error attribution.

One observation is that for tutors where the KCs can be determined by the interface location (or widget) in which an action appears it is likely that the LH will show better results than the TH. This feature is mostly true of the scatterplot tutor. It is possible that the TH is better in situations where the different problem subgoals can be associated with a single location. However, our prior results with a physics data set indicated that even in such situations the LH may be better. Further research should explore this issue.

We also intend to investigate whether the use of the simple location-based heuristic may improve on-line student modeling and associated future task selection. The availability of datasets from the Pittsburgh Science of Learning Center's 'DataShop' (see <http://learnlab.org>) will facilitate the process of getting appropriate data.

ACKNOWLEDGEMENTS

We would like to thank Hao Cen for his help with the LFA tool.

REFERENCES

- ANDERSON, J. R., BELLEZZA & BOYLE, C. F., 1993. The Geometry Tutor and Skill Acquisition. In Anderson, J. R. (Ed.) *Rules of the Mind*, Chapter 8. Hillsdale, NJ: Erlbaum.
- BAKER R. S., 2005. *Designing Intelligent Tutors that Adapt to when Students game the system*. Doctoral Thesis. Human Computer Interaction. School of Computer Science. Carnegie Mellon University.
- BAKER, R.S., CORBETT, A.T., KOEDINGER, K. R., WAGNER A. Z., 2004. Off-Task Behavior in the Cognitive Tutor Classroom: When Students “Game the System”. *Proceedings of ACM CHI: Computer-Human Interaction*, pps. 383-390.
- CEN, H., KOEDINGER, K. & JUNKER, B., 2005. Automating Cognitive Model Improvement by A* Search and Logistic Regression. In: *Proceedings of AAAI 2005 Educational Data Mining Workshop*.
- CORBETT A.T., ANDERSON, J.R., O'BRIEN A.T., 1995. Student Modelling in the ACT Programming Tutor. In: *Cognitively Diagnostic Assessment*. Hillsdale, NJ: Erlbaum.
- DRANEY, K., PIROLI, P. & WILSON, M. 1995. A Measurement Model for a Complex Cognitive Skill. In *Cognitively Diagnostic Assessment*. Erlbaum, Hillsdale, NJ
- HEATHCOTE, A., BROWN, S., & MEWHORT, D. J. K., 2002. The Power Law revealed: The Case for an Exponential Law of Practice. *Psychonomic Bulletin & Review*, vol. 7, pps. 185-207.
- KOEDINGER, K.R. AND MATHAN, S., 2004. Distinguishing qualitatively different kinds of learning using log files and learning curves. In: *ITS 2004 Log Analysis Workshop*, Maceio, Brazil. pps. 39-46.
- MARTIN, B., MITROVIC, T., MATHAN, S., & KOEDINGER, K.R., 2005. On Using Learning Curves to Evaluate ITS. *Automatic and Semi-Automatic Skill Coding With a View Towards Supporting On-Line Assessment*. In: *Proceedings of the 12th International Conference on Artificial Intelligence in Education*. Amsterdam, IOS Press.
- MATHAN S. & KOEDINGER K., 2005. Fostering the Intelligent Novice: Learning From Errors With Metacognitive Tutoring *Educational Psychologist*. 40(4), pps. 257–265.
- MITROVIC A. OHLSSON S., 1990. Evaluation of a Constraint-Based Tutor for a Database Language. *International Journal of Artificial Intelligence in Education*. vol. 10, pps. 238-256.
- NEWELL, A. & ROSENBLOOM, P., 1981. Mechanisms of Skill Acquisition and the Law of Practice. In: Anderson J., (ed.): *Cognitive Skills and their Acquisition*, Erlbaum Hillsdale NJ.
- NWAIGWE, A., KOEDINGER, K. ET AL. , 2007. Evaluating Alternative Methods for Making Error Attribution in Intelligent Tutoring Systems. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Education*. Los Angeles, CA.
- PIROLI P. & WILSON M. A., 1998. Theory of Measurement of Knowledge Content, Access and Learning, *Psychological Review*. vol. 105, 1, pps. 58-82.
- VANLEHN, K., LYNCH, C., SCHULTZ, K., SHAPIRO, J. A., SHELBY, R. H., TAYLOR, L., ET AL., 2005: The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education*. 15(3), pps.147-204.
- WASSERMAN, L., 2004. *All of Statistics: A Concise Course in Statistical Inference*. Springer.