Domain-Specific and Domain-Independent Interactive Behaviors in Andes

Min CHI and Kurt VANLEHN
Learning Research and Development Center,
University of Pittsburgh, Pittsburgh, PA 15260 U.S.A

Abstract. In this study, students studied two different domains in the same Intelligent Tutoring System, Andes. Analysis of 435 log files from 22 subjects indicated that there are two types of interactive behaviours in Andes: domain-independent and domain-specific. We believe the existence of the domain-specific behaviours is one possible reason that similar meta-cognitive behaviors has not been found across domains in a single ITS curriculum [1][2]. This paper describes the study and the potential applications of these findings.

Introduction

In this study, we ported our existing physics tutor Andes [4] to probability by replacing the physics knowledge base with the probability concepts and principles [3]. Apart from the declarative domain knowledge, the two Intelligent Tutoring Systems (ITSs) were identical. We refer to them as Andes-physics and Andes-probability respectively. An important question is what happens when two different task domains exist in the same interface? Would students engage in similar interactive behaviors or different ones? In this paper, we will use the term interactive behaviors refer to the individual actions that the students performed on the ITS.

Previous studies in this area have shown that students do *not* exhibit the same patterns of meta-cognitive behaviors across domains in a single ITS curriculum [1][2]. Baker et al. built a gaming detector by aggregating patterns of students' behavior from logs. They investigated how well it transferred across domains in the same ITS within a single student population. They found that the detector built from data in one domain does well within the domain but cannot successfully predict behavior across domains.

In this study, we used aggregate measurements of interactive behaviors, such as the total time taken or the total number of entries made. Two types of behaviors were defined: domain-specific and domain-independent. An interactive behavior is called domain-specific if its appearance was associated with the domain characteristics and thus its appearance in one domain did *not* predict its appearance in another domain; whereas it is called domain-independent if its appearance was *not* associated with the domain characteristics and its appearance in one domain did predict its appearance in another. We found that both types of behaviors existed in Andes. Next, we will describe this study in a greater detail. Following that we will present the results. Finally, we presented the conclusions.

1. Methods

Data was collected from 23 students during the fall of 2005 and the early spring of 2006. Students were required to have basic knowledge of high-school algebra but to not have taken college-level statistics or physics courses. They studied two domains: probability and physics. Each domain contained ten principles. Students first took a background survey and then went through probability instruction followed by physics instruction. Each instruction consisted of four steps: 1) pre-training, 2) pre-test, 3) training on Andes, and 4) post-test. During the pre-training, the students read a general description of each principle, reviewed some examples, and solved some problems.

During training on Andes, students first watched a video of a problem being solved in Andes. Then they solved twelve probability problems in Andes during the probability instruction and eight physics problems during the physics instruction. Andes gives immediate feedback as soon as an action is performed. Entries are colored green if they are correct and red if incorrect. We distinguish between two types of errors: *tiny errors* and *red errors*. *Tiny errors* are those likely due to lack of attention such as typos. *Red errors* are likely caused by conceptual misunderstandings or a lack of domain knowledge such as incorrect principle applications. When *tiny errors* occurred, Andes popped up an automatic error message. When *red errors* occurred, Andes colored the entry red and gave the students the option to ask for *what's-wrong help* or fix it on their own. Students could ask for *next-step help* at any time if they are not sure what to do next. Both *what's-wrong help* and *next-step help* were provided via a sequence of hints that gradually increased in specificity. The last hint in the sequence, called the *bottom-out hint*, told the student exactly what to do.

2. Results:

One student was eliminated due to a perfect probability pre-test score. For the remaining 22 students, there were 435 Andes log files, 259 for probability and 176 for physics. For each log file, we extracted seven interactive behaviors: the total time and the total number of next-step help, what's-wrong help, bottom-out hint, red errors, tiny errors, correct entries, and total entries. Table 1 shows that students had significantly higher values during the physics training than during the probability training on all behaviors. This suggested that the physics training problems seemed to be more difficult and challenging for students than the probability ones.

Table 1. Compare various interactive behaviours across domains

	Next-Step	Bottom-Out	What's-Wrong	Red Errors	Tiny Errors	Correct Entries	Total Entries	Total time
Probability	16.00	3.02	1.05	4.40	2.05	14.94	17.98	793.4
Physics	34.67	7.85	5.53	11.08	4.41	29.1	39.33	1351.

Table 2 shows the correlation of the interactive behaviors across the instruction sessions in Andes. The unit of analysis was the student. Each correlation is the result of 22 2-tuples, one pair for each student. We found that two help-seeking behaviors, *next-step help* and *bottom-out hints*, were correlated across domains while *what's-wrong help* was not. These results make sense: asking for *next-step help* may result from the students' lack of general problem-solving skills rather than any domain-specific characteristics. Since the last hint in the sequence of *next-step help* is a *bottom-out hint*, it was no surprise that *bottom-out hints* were also correlated across domains. Asking for *what's-wrong help*, on the other hand, is more likely to be driven by the domain characteristics. We predict that students' ability or willingness to fix an error on their own is largely dependent on how difficult they think the domain is. Since physics appeared be much harder for them than probability, students were more likely to ask for *what's-wrong help* in physics than in probability.

We interpreted the correlation between the number of *red errors*, correct entries, total entries, and the total time as a sign that students' problem-solving strategies and general competence remained consisted across domains. It indicated that Andes did not teach them anything about general problem solving skills. The lack of correlation of making tiny errors is likely due to domain complexity: for example, students often need to input both a number and a unit in physics but only a number in probability.

Table 2. Correlation of interactive behaviours across Probability and Physics

#	Interactive Behaviours	correlation	#	Interactive Behaviours	correlation
1	Next-Step Help	r= 0.8015, p< 0.0001	**5**	Tiny Errors	-
2	Bottom- Out Hints	r= 0.6684, p=0.002	**6**	Correct Entries	r= 0.719, p=0.001
3	What's- Wrong help	-	**7**	Total entries	r= 0.5629, p=0.012
4	Red Errors	r=0.545, p=0.016	**8**	Total time	r=0.834, p< 0.00001

3. Conclusions and discussions:

Apparently, our findings seem to contradict previous studies [1][2] in that some meta-cognitive behaviors in our study seem stable across domains. However, we argue that our results are consistent with theirs. Baker et al. aggregated multiple meta-cognitive behaviors with distinct patterns of activities whereas we focused on isolated but still interpretable activities. In order for their predictor to function accurately, the relative frequencies of each meta-cognitive behavior would need to remain constant across domains. That is, all of the individual behaviors in their study would need to be domain-independent. However, we have shown that some behaviors that matter to their detector are domain-specific. For example, one crucial gaming criteria in their study was that students who keep making errors do eventually get the correct answer. This would show up in our study as students keep asking for *what's-wrong help* and eventually get the correct answer. We expect that some of behaviors in their study would be domain-specific. In this study, we investigated the interactive behaviors in only two domains: probability and physics. In order to draw a more comprehensive conclusion, we would need to test it in multiple domains. We think it would be worth the attention and time because we expect that a model constructed solely on domain-independent behaviors would exhibit greater prediction accuracy across domains and thus lead to a more accurate student model. We believe that porting an existing ITS to a new domain is not only efficient in terms of time and cost [3], but also may improve learning by allowing more accurate prediction and prevention of help-abusers.

4. References

[1] Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Roll, I. *Generalizing Detection of Gaming the System Across a Tutoring Curriculum*. Proceedings of the 8th ITS (2006). 402-412

[2] Baker, R.S., Corbett, A., Koedinger, K., Roll, I. (2005) *Detecting When Students Game The System, Across Tutor Subjects and Classroom Cohorts* . Proceedings of User Modeling 2005, 220-224.

[3] Chi, M, and VanLehn K, Porting an Intelligent Tutoring System across Domains. 13th International Conference on Artificial Intelligence in Education

[4] VanLehn, K., Lynch, C., Schultz, K., et al. (2005). The Andes physics tutoring system: Lessons learned. *International Journal of Artificial Intelligence and Education, 15*(3), 147-204.