Validating the Detection of a Student's Motivational State

Angel de Vicente
angelv@iac.es
Instituto de Astrofísica de Canarias (Spain)

Helen Pain
helen@inf.ed.ac.uk
The University of Edinburgh (UK)

Abstract

Tutoring systems could greatly benefit from incorporating a model that describes the motivational state of its student. But detecting the motivational state of a person is not a simple issue, as the knowledge to infer this state is not easily elicited. de Vicente and Pain (2002) elicited a large number of knowledge rules relating to the detection of the motivational state of a student by a tutoring system. In order to do so, they designed a study in which a tutor was asked to infer a student's motivational state while watching the pre-recorded interaction of a student with an instructional system. In this paper we describe an empirical study performed in order to validate this set of rules.

The validation study took the form of a simple questionnaire which was administered on-line in which participants were presented with an instructional interaction description and were asked to rate the motivation detection rules previously elicited. This study helped to filter out the rules described by de Vicente and Pain (2002) that were not widely accepted by participants, and provided a validated body of knowledge about the detection of the motivational state of a student that could be readily incorporated into an Intelligent Tutoring System (ITS).

1 Introduction

The issue of how human teachers detect their students' motivation has been virtually unexplored in AI-ED research (although some work has been done in this area, for example: Issroff and del Soldato (1996); Malone and Lepper (1987); del Soldato (1994)). Many cues that help us detect other people's motivation are perceived unconsciously (Davis, 1976), which makes it difficult to elicit motivation detection knowledge. In an effort to make explicit this knowledge, de Vicente and Pain (2002) designed a study in order to extract and formalise tutors' knowledge about motivation detection. In this study, a participant would see exclusively the screen interaction of a student with an instructional system. Thanks to this restricted setting, the knowledge inferred referred only to the type of information readily available to an instructional system (such as time of interaction with the system, mouse movements, etc.), and avoided information difficult to incorporate into an instructional system (such as eye movements, gestures, etc.).

In order to perform this study, de Vicente and Pain (2002) developed a system which could be used to replay the actions of a previous student's interaction with an instructional system. The interaction with this system can be summarised as follows:

1. The participant was presented with information about certain trait characteristics of a student.
2. Then she was shown a replay of the student's interaction with an instructional system.
3. Throughout the interaction, the participant was encouraged to give verbal comments (which were recorded for analysis) on the student's motivational state and the possible factors affecting it in as concrete terms as possible.

As a result of this study de Vicente and Pain (2002) gathered 61 rules about the detection of a student's motivational state, some examples of which can be seen in table 1. The columns in the groups *Performance* and *Teaching Materials* (except those starting with *pre*) refer to characteristics of the current instructional unit. The input factors starting with *pre* refer to characteristics of the instructional unit immediately before the current one. The output of the rules refers to the detection of the given motivational factor under the circumstances given by each of the rules

Rule	Quality	Quantity	Speed	pre(Give up)	Give up	Hesitation	Performed in order	Difficulty	Output
	PERFORMANCE							*TEACHING MATERIALS*	
IE1					No	High			Inc
IE2		Very high							High
IE3		High			No				High
IE4							Yes		Inc
IE5				Yes					High
IE6		High	Avg					High	High
IE7	High		Slow						High

Table 1: Increase Effort diagnosis rules

For example, rule IE3 can be expressed as: "If the student tries to perform most of the exercises in this lesson and he does not give up, then we can infer that his effort is high". Similar rules concerning other motivation factors (such as satisfaction, confidence, cognitive interest and sensory interest) were gathered, which provide an empirically motivated approach to the issue of motivation diagnosis in an ITS. More importantly, these rules are based on very concrete aspects of the interaction, such as mouse movements, quality of performance, etc., which can be easily detected in a tutoring system.

In order to validate this set of rules, we performed the validation study which is described in this paper. Cross-participant comparison was not an appropriate way to validate the given set of rules, as the number of rules elicited was not large enough to provide many groups of rules which could be applied under the same conditions. Similarly, comparison with a previously performed self-report study (de Vicente and Pain, 1999) in which participants reported about their own motivational state was not appropriate because there was no reason to believe that the self-report wass necessarily accurate, as 'false' readings can be given under certain circumstances. For example, if the student is too engaged, he would probably forget to report on his motivational state. Also, when asked, it is likely that students will attempt to 'please' the tutoring system by providing artificially positive readings of their motivation (Reeves and Nass, 1998). Therefore, we validated these rules by performing an empirical study in which participants were presented with an instructional interaction context and were asked to rate the rules that could be applied under those conditions.

2 Materials and Methodology

The validation study took the form of a simple questionnaire which was administered on-line. A question was prepared for each of the 61 rules gathered by de Vicente and Pain (2002), and each participant was asked 30 of these questions randomly chosen. Each question took the form of a description of an instructional setting and a question regarding a motivational factor. As an example, we can see figure 1, which represents the question corresponding to the rule IE6 in table 1.

For each question we asked the participants to infer what the value of the motivational factor would be. As possible options we gave them three choices:

- The value given by the corresponding rule.
- The opposite value.
- Don't know.

The *Don't know* option was always the third item in the list. The other two options were ordered in decreasing value. This way the value given by the corresponding rule would be sometimes the first item in the list, and sometimes the second.

As we can see in figure 1, each of the sentences representing the instructional setting correspond to one of the rule's inputs (see table 1). The motivation question referred to the motivational factor **Effort**. In the case of the rule IE6, the output value of the rule is *High*. Thus, the choices for responding to the IE6 question that we gave to

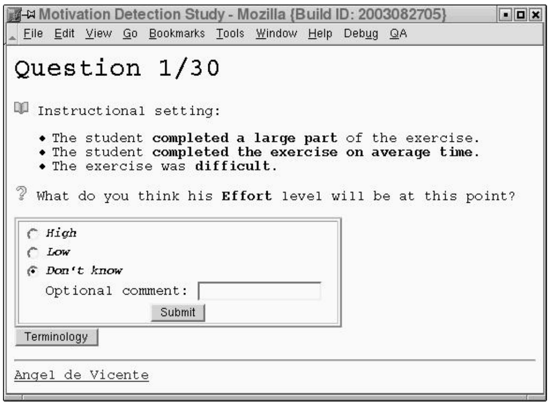

Figure 1: Validity Study Sample question.

the participants were *High*, *Low* (the opposite), and *Don't know*. The participants were asked to choose the option *Don't know* if they thought they didn't have enough information to choose between the other two values. At the same time, they were encouraged to provide any optional comments if they wished to.

To obtain participants for this study, we asked for collaboration in a number of mailing lists dealing with technology and education. The prerequisite for participating in the study was to have at least two years of teaching experience and, as a result of this, 33 participants volunteered to take part in this study. Some very experienced teachers took part in this study. The subject area was not a precondition, and therefore there was an ample range of subjects taught by the participants, amongst others: Programming, Maths, Spanish, Biology, Linguistics, Social Studies, Education and Chemistry.

As mentioned earlier, the study was administered as an on-line questionnaire. After a participant agreed to participate in the study, she was given the web address of the instructions for the study. In order to avoid duplication of answers, the questionnaire web pages were protected by a username and a password, which were valid only once, and which were given to each participant after they had agreed to participate in the study. After the participant had read the instructions and provided her username and password, she was directed to the questionnaire proper.

Each participant was asked 30 questions, chosen randomly amongst the 61 available. For each question they were asked to choose the option that they thought would be applicable under the given instructional setting. If they did not have enough information to make a choice, they were asked to choose the *Don't know* option. Regardless of the option they selected, they were also offered the possibility of giving any extra comments and/or qualifications to their answer. Once they finished answering all the questions, they were given the option to see the results.

3 Results

In total we collected 973 votes (instead of the maximum 990 possible votes), as some of the participants experienced technical difficulties that forced them to finish the questionnaire before they had completed the 30 questions. As the questions were chosen randomly for each participant, not all the questions got the same number of replies, but all of them got at least 15 replies. For an example of the replies obtained see table 2. The columns represent respectively: the name of the rule; the output value of the rules; the number of participants who answered the question corresponding to that rule; the value of the first choice given to participants; the number of participants that selected that choice; and the same for the second and third options.

Rule	Output	n	Option 1		Option 2		Option 3	
DS1	Low	15	High	0	Low	15	DK	0
DC2	Low	15	High	1	Low	10	DK	4
DE1	Low	16	High	6	Low	6	DK	4
ICI4	High	17	High	5	Low	6	DK	6

Table 2: Sample motivation diagnosis rules results.

Rule	n	Accept	Reject	p
DS1	15	15	0	<0.00001
DC2	15	10	5	0.00617
DE1	16	6	10	0.72064
ICI4	17	5	12	0.72844

Table 3: Sample chi-square results.

Was it easy to select an option? Each rule offered the participants three options: the output value of the corresponding rule, its opposite and *Don't know*. A choice of *Don't know* generally meant a lack of information on which to make a choice. By analysing the distribution of the *Don't know* replies amongst the 31 participants that completed the 30 questions, we can see that some of the participants chose *Don't know* for as many as 20 questions, but that was not usual. On average, each participant chose the option *Don't know* for only 6.94 questions, so most of the time participants seemed to have enough contextual information to make a choice between the other two options.

Which rules to accept? In order to decide which rules to accept and which ones to reject, we proceeded in the following manner:

1. For each question we collected the values for the three options (see sample in table 2).
2. One of the options is the same as the one suggested by the corresponding rule (column *Value* in table 2), and we consider this the **Accept** condition. The other two options are considered **Reject** conditions (either the participant inferred the opposite of what the corresponding rule infers, or selected *Don't know*). We added the votes for the two possible *Reject* options (see sample in table 3 for the same rules as in table 2).
3. For each of the rules we performed a chi-square *goodness of fit* test, where the null hypothesis is that there is no preference for either *Accept* or *Reject*. Thus, under the null hypothesis we would expect 33% of the votes to go into the *Accept* option and 66% of the votes to the *Reject* option. The probability of the distribution obtained in our study given the null hypothesis is given in column *p* in table 3.
4. We consider valid rules those in which the participants showed a statistically significant preference ($p<0.01$) for the *accept* category.

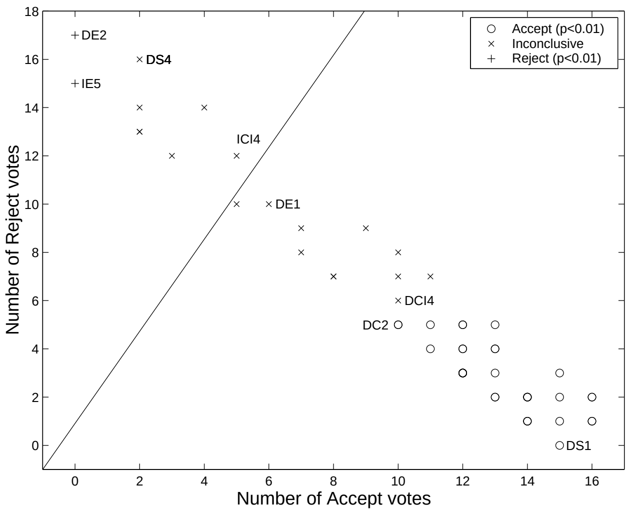

Figure 2: Acceptance distribution of Motivation Diagnosis rules.

In figure 2 we can see the distribution of the rules according to the number of participants that chose the *accept* option. The line represents the expected number of *accept* and *reject* votes if the null hypothesis was true. Each marker represents the actual number of *accept* and *reject* votes for a particular rule[1]. At the same time, we have drawn a circle for the rules that are accepted.

Thus, we can see in figure 2 that the rule DS1 obtained 15 *accept* votes, but 0 *reject* votes. This is clearly a significant preference for *accept*, and as such it is drawn with a circle. Rule DC2 obtained 10 *accept* votes and 5 *reject* votes. This is still statistically significant.

Rule DCI4 obtained 10 *accept* votes and 6 *reject* votes. This is not statistically significant and therefore we will not accept this rule in our final set. In the case of the rules IE5 and DE2, the distribution obtained has a $p<0.01$, but due to the fact that all the votes were *Reject* votes these rules are not accepted.

[1] In some cases more than one rule has the same number of *accept* and *reject* votes, so the same dot represents all of them.

Comments by participants. As mentioned earlier, the participants were encouraged to give extra comments for each of the questions if they wanted to qualify their decision. Although we cannot analyze this in detail here, we can just point out a number of things. From the 973 questions answered in total, the participants gave comments to 205 of them. Of these, 130 were given after a *Don't know* vote. Most of the participants made comments to 1 to 5 questions, although two participants gave comments to more than 50% of all the questions.

In many cases the participants commented that there was not enough information in the instructional setting to make an inference for the given motivational factor. But not all comments were of this type: 75 were given after selecting an option other than *Don't know*. These were qualifying comments, which in some cases could help us refine the motivation diagnosis rules. Other types of comments could be categorized as: a) Motivational factor depends on an extra variable; b) Relation between rule inputs and outputs not clear; c) Not enough information; d) A choice is made, but under certain assumptions; e) A choice is made, but certain information is missing ; e) *Don't know* selected, but a choice is actually made; f) Elaborations on choice made.

4 Discussion

With this study we obtained sufficient data to perform a *goodness of fit* test for each individual rule, in order to ascertain whether participants preferred the *accept* option over the *reject* ones. As a result we managed to reduce the original set of rules to 41. At the same time we gathered a large number of comments from the participants, which could help us to further improve some of the rules.

It would be interesting to analyse whether the comments made for the non-accepted rules could help us to add/modify/remove some conditions of the rules, so that in a subsequent study we could incorporate these rules in the accepted set. For example, one of the comments for the rule DS5 was that the inference would depend on "whether there are enough fantasy characteristics for the student". If similar comments were made by other participants about this rule, we could add a condition to test this. In a subsequent study we could test whether adding this condition would make this rule more widely acceptable.

Although we can only speculate on the reasons why some of the rules were not commonly accepted, by performing this study we have managed to filter the rules elicited by de Vicente and Pain (2002) that were not widely accepted by the participants of our study to a smaller set of rules that could be readily incorporated into an Intelligent Tutoring System (ITS).

References

Cerri, S. A., Gouardères, G., and Paraguaçu, F., editors (2002). *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems*, volume 2363 of *Lecture Notes in Computer Science*, Berlin. Heidelberg. Springer.

Davis, F. (1976). *La Comunicación No Verbal*, volume 616 of *El Libro de Bolsillo*. Alianza Editorial, Madrid, Spain. Translated by Lita Mourglier from: "Inside Intuition - What we Knew About Non-Verbal Communication"; McGraw-Hill Book Company, New York.

Issroff, K. and del Soldato, T. (1996). Incorporating motivation into computer-supported collaborative learning. In P. Brna, A. Paiva, and J. Self, editors, *Proceedings of the European Conference on Artificial Intelligence in Education*, pages 284–290, Lisbon. Colibri.

Malone, T. W. and Lepper, M. R. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. In R. E. Snow and M. J. Farr, editors, *Conative and Affective Process Analyses*, volume 3 of *Aptitude, Learning, and Instruction*, chapter 10, pages 223–253. Lawrence Erlbaum Associates, Inc., Hillsdale, New Jersey.

Reeves, B. and Nass, C. I. (1998). *The Media Equation: How People Treat Computers, Television and New Media Like Real People and Places*. Centre for the Study of Language and Information, US, New York.

del Soldato, T. (1994). *Motivation in Tutoring Systems*. Ph.D. thesis, School of Cognitive and Computing Sciences, The University of Sussex, UK. Available as Technical Report CSRP 303.

de Vicente, A. and Pain, H. (1999). Motivation self-report in ITS. In S. P. Lajoie and M. Vivet, editors, *Proceedings of the Ninth World Conference on Artificial Intelligence in Education*, pages 648–650, Amsterdam. IOS Press.

de Vicente, A. and Pain, H. (2002). Informing the detection of the students' motivational state: An empirical study. In Cerri *et al.* (2002), pages 933–943. (*ITS2002 Best Paper Award*).