Implementation of an Adaptive Questionnaire

Petri Nokelainen
Markku Niemivirta
Jaakko Kurhila
Miikka Miettinen
Henry Tirri

Complex Systems Computation Group
Department of Computer Science
University of Helsinki

Abstract: An adaptive on-line questionnaire, named EDUFORM[1], is based on Bayesian statistical techniques that both optimize the number of propositions presented to each respondent and create an individual learner profile. Adaptive graphical user interface is generated partially (propositions of the questionnaire, collaborative actions and links to resources) and computational part totally with Bayesian computational techniques. The preliminary results show that after reducing dramatically the number of propositions (from 64% to 54%) we may still moderately control the error ratio (from 12% to 22 %). The profiling information of respondent is in most cases obtained after asking just one third of propositions.

Keywords: adaptive questionnaire, Bayesian modeling, motivational profiling

Major goal

The major goal of this paper is to describe design and implementation of a software module, named EDUFORM, which allows for an adaptive and dynamic optimization of questionnaire propositions and profiling of learners on-line. This tool is based on probabilistic Bayesian modeling (Bernardo and Smith 2000). Many of the features used in this restricted evaluation task can be directly applied to wider context of modern computer-based learning environments (Dillenbourg 1999). The educational problems considering this study are two-fold:

1. A great number of questionnaires, both on paper and electronic form, are designed with 'one size fits all' principle. Equipped with numerous propositions, usually around one hundred, along with some inadequate propositions related to the theory or model beneath, they prolong the answering process decreasing internal, external and contextual validity.

2. Learning environments are not effectively profiling learners and thus utilizing the possibilities to promote collaborative and cooperative learning, provide adaptive user interfaces, personalized contents of substance and hints to additional resources.

Instrumentation

The instruction data set (1800 students of a Finnish educational institution.) was collected in December 2000 with both traditional and Bayesian optimized MSLQ (Pintrich 2000). The same organization with partly the same respondents (sub-sample of 460 students) is our target on the next measurement in March 2001.

Motivational profiling in this study is based on the Motivated Strategies for Learning questionnaire (MSLQ), which is developed on the basis of motivational expectancy model (Garcia and Pintrich 1994). MSLQ measures both motivational factors and learning strategies and has been adapted to the research field of Finnish vocational education (Ruohotie 2000). The motivation section (A) of MSLQ consists of 28 items that were used to assess students' value for a course, their beliefs about their skill to succeed in the course, and their anxiety about tests in the course. The learning strategy section (B) includes 40 items regarding student's use of different cognitive, metacognitive and resource management strategies. A 5-point Likert-scale ranging from 1 ("Not at all true of me") to 5 ("Very true of me") was used for all items. The initial order of items was randomized.

Bayes methodology

Since we are in the questionnaire context, it is quite natural to model the problem domain by (m) discrete variables $X_1,\ldots,X_m$ (we can assume that continuous values, if needed, are discretized), and that a data d is sampled from the joint distribution of these variables. In finite mixtures we now make an additional modeling assumption that the data D can be viewed as if it were generated by K different mechanisms, all of which can have a distribution of their own. Furthermore, it is assumed that each data vector originates from exactly one of these mechanisms. In this case model family is only a language in which we can express the constraints in data. From these assumptions it follows that the data vector space is divided into K local regions usually called *clusters* or *profiles*, each of which consists of the data vectors generated by the corresponding mechanism.

¨Paper presented in ED-MEDIA 2001, Tampere, Finland, June 25-30, 2001.

The underlying intuitive idea is that a set of data vectors can be modeled by describing a set of profiles, and then describing the data vectors using these profile descriptions. Each description gives the distribution of the variables $X_1,\ldots,X_m$ conditioned that the data vector belongs to the cluster. The cluster descriptions should be chosen in such a way that the information required to describe data vectors in the cluster could be significantly reduced because they are similar to the "prototype" described by the profile. In such a "profile language" a data set D can be described by first giving the profile index for each data vector, and then by describing the differences between the observed and expected values. Construction of mixture models from a given data set D by using the Bayesian approach is described in (Kontkanen et al. 1996) and (Tirri et al. 1996).

Implementation of EDUFORM

The EDUFORM user interface consists of various elements (Figure 1). The meanings of the likert-scale answers are given on the top part of the screen. Because of the simple interface, there is no separate help screen. The instructions on the different buttons in EDUFORM are given in tool tips. In Figure 1, the tooltip is "Profile" since the pointer is on the button visualizing the profile of the user. On the bottom row, user navigation bar with different buttons are presented. The actual propositions are on the middle part of the screen.

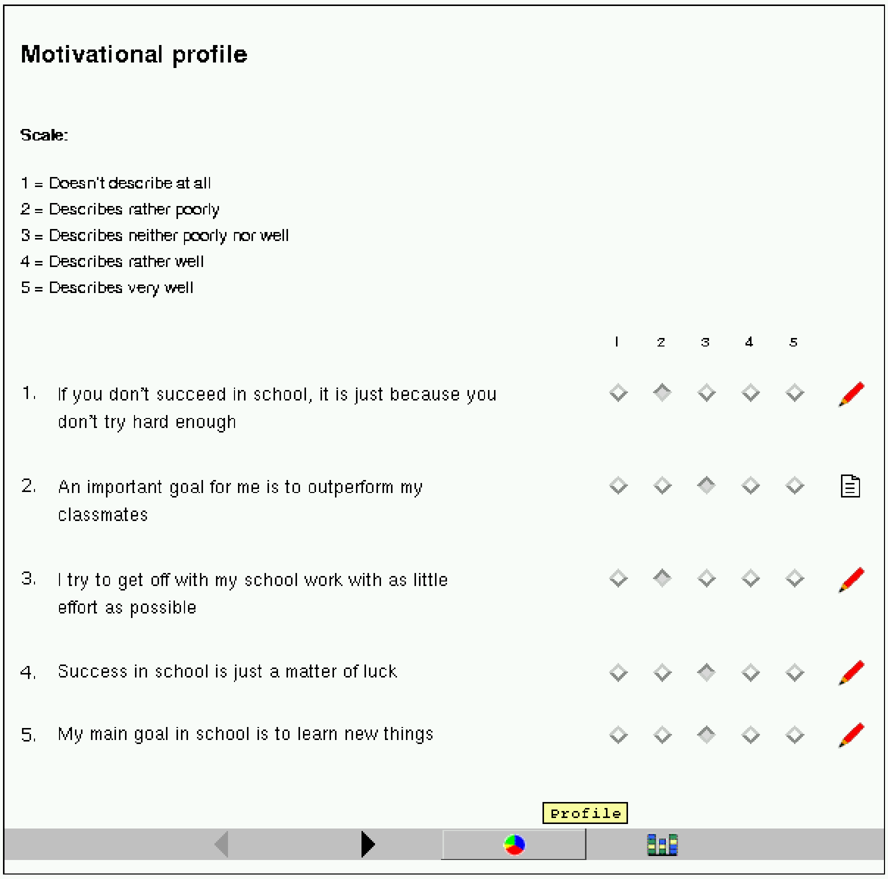

Figure 1.

EDUFORM user interface

In Figure 1, the user has given an answer to all of the first five propositions, and the second proposition has inspired the user to write an open comment regarding the proposition. In Figure 2, the visualization of the state of the questionnaire is presented. This is the visualization shown also to the user when pressing the button "Questionnaire state" (the most right button on the user navigation bar). As seen in Figure 2, the user has answered to the questions on the block number 1. By the adaptation mechanism, EDUFORM has concluded that there is no need to present the next question and the question on the block 2. Figure 2 presents situation where user has actually given 24 responses, EDUFORM has reasoned 24 responses and 37 propositions are at the moment open. It should be noted that this visualization is dynamic, i.e. different at every stage of the fill-in process.

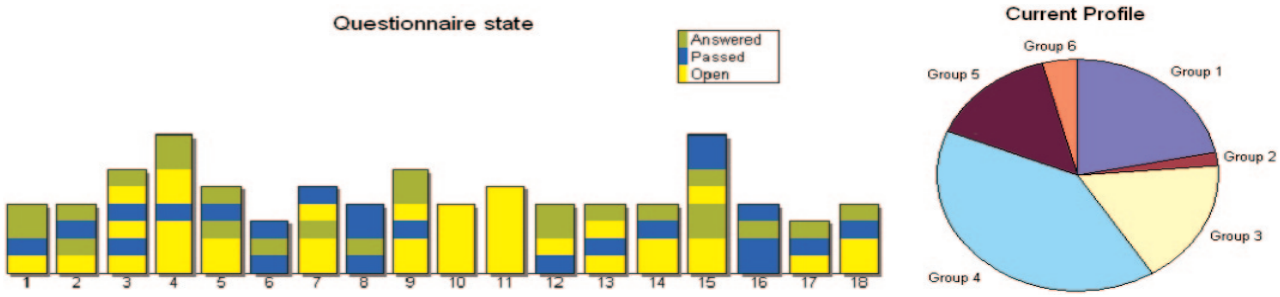

Figure 2.

EDUFORM questionnaire optimization and profiling state

The last additional element, visualization of the learner profile (groups of learners), is shown in Figure 2. The user can see the estimate of his or her profile when pressing the third button from the left (in Figure 1) in the user navigation bar. The users are divided into different groups of learners based on their answers on the questionnaire. The EDUFORM system applies Bayesian reasoning to the profile assessment. The data gathered gives an estimate where the learner profile is most likely to fit into group four or one. Also group one seems still possible, but the groups two and six are very unlikely.

Results

The explicit advantages of applying on-line adaptive questionnaire on educational domain included a) increased task-related participation, b) absence of both second level coding errors, and c) respondents exhaustion effect, d) reduced number of propositions and e) fast feedback. The implicit benefits of EDUFORM covered matters like a) adaptation to respondents learning, cognitive and motivational strategies, b) increased reliability, c) means to implement collaborative actions and d) flexible changeovers made possible between theoretical viewpoints. Preliminary results for optimization are represented in Table 1. Results show that error ratio is reasonable (Part A: 64% / 12%, part B 54% / 22%) and controlled within the instructional data set sub sample (N=230).

Table 1. Preliminary optimization results

	Questions asked (%)	Errors (%)
Part A	58	14
	65	18
	74	11
	57	5
Part B		
	56	29
	44	13
	56	26
	61	33
	50	11

Conclusion

Statistical techniques explored here are one possible solution to provide a intelligent messenger (or agent) to intermediate knowledge between collaborating students (Hoppe and Ploetzner 1999, 147). The architectural question (i.e. platform for software modules like EDUFORM) is difficult to solve globally due to different goals of open and distance learning research teams. Bayesian statistical computation together with proven server architecture, for example “The LispWeb” (Riva and Ramoni 2000), enable implementation of CSLE systems that offer true collaboration between agents with differing domains of expertise. Abductive reasoning as described “inference to the best explanation” (Josphson and Goel 1996, 202) is truly interesting methodological addition to Bayesian reasoning because a large part of knowledge consists of causal understanding.

References

Bernardo, J. & Smith, A. (2000) *Bayesian Theory*. John Wiley & Sons: New York. 2nd ed.

Dillenbourg, P. (Ed). (1999). *Collaborative-learning: Cognitive and Computational Approaches*. Oxford: Elsevier.

Garcia, T. & Pintrich, P. (1994) Regulating Motivation and Cognition in the Classroom: The Role of Self-Schemas and Self-Regulatory Strategies. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-Regulation of Learning and Performance: Issues and Educational Applications*. Hillsdale, N. J.: Erlbaum.

Hoppe, U. and Ploezner, R. (1999) Can Analytic Models Support Learning in Groups? In P. Dillenbourg (Ed). *Collaborative-learning: Cognitive and Computational Approaches*, pages 147-168. Oxford: Elsevier.

Josephson, J. and Goel, A. (1996) Tractable Abduction. In J. Josephson and S. Josephson (Eds.) *Abductive Inference*, pages 202 – 215). New York: Cambridge University Press.

Kontkanen, P., Myllymäki, P. and Tirri, H. (1996) Predictive Data Mining with Finite Mixtures. In *Proceedings of The Second International Conference on Knowledge Discovery and Data Mining*, pages 176-182. Portland, OR, August 1996.

Pintrich, P. (2000) The Role of Motivation in Self-Regulated Learning. In P. Pintrich and P. Ruohotie (Eds.) *Conative Constructs and Self-Regulated Learning*, 31-50. Saarijärvi: Learning and Change Series of Publications.

Ruohotie, P. (2000) Conative Constructs in Learning. In P. Pintrich and P. Ruohotie (eds.) *Conative Constructs and Self-Regulated Learning*, 1-30. Saarijärvi: Learning and Change Series of Publications.

Tirri, H., Kontkanen, P. and Myllymäki, P. (1996) Probabilistic Instance-Based Learning. In L. Saitta, *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 507-515. San Fransicso: Morgan Kaufmann Publishers.

Notes

1. http://eduform.cs.helsinki.fi [Note: Demo in Finnish]
2. http://www.cs.helsinki.fi/research/cosco
3. http://www.tekes.fi/eng/default.asp
4. http://www.helsinki.fi/english

Acknowledgements

The work belongs to educational technology division of an on-going project of Complex Systems Computation Group[2] (CoSCo) named Personalized Adaptive Interfaces (PAI). The partners involved are: National Technology Agency[3], University of Helsinki[4] and various unnamed Finnish Information Technology business partners.