Bayesian Modeling Approach to Implement an Adaptive Questionnaire

Petri Nokelainen
Markku Niemivirta
Henry Tirri
Miikka Miettinen
Jaakko Kurhila
Tomi Silander

Complex Systems Computation Group
University of Helsinki

Abstract: An adaptive questionnaire, named EDUFORM, is based on Bayesian statistical techniques that both optimize the number of propositions presented to each respondent and create an individual learner profile. The preliminary results show that reducing the number of propositions we may still moderately control the error ratio. The respondents’ profiling information is in most cases obtained after one third of propositions.

Major goal
This paper describes design and implementation of a software module, named EDUFORM, which allows for dynamic optimization of questionnaire propositions and profiling of learners on-line. This tool is based on the probabilistic Bayesian modeling (Bernardo and Smith 2000), and many of the features used in this restricted evaluation task can be directly used in the wider context of modern computer based learning environments (Dillenbourg 1999).

Bayes methodology
In the questionnaire context it is quite natural to model the problem domain by (m) discrete variables $X_1, \ldots, X_m$ 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. 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. 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 Tirri et al. (1996).

Implementation of EDUFORM
The instruction data set (1800 students of a Finnish educational institution.) was collected in December 2000 with both traditional and Bayesian optimized questionnaire. Motivational profiling in this study is based on the Motivated Strategies for Learning Questionnaire, which is developed on the basis of motivational expectancy model (Pintrich 2000).

The EDUFORM user interface is shown in Figure 1. The propositions are on the middle part of the screen, and as seen, the seventh proposition has inspired the user to write an open comment regarding the proposition.

The left side of Figure 2 presents a dynamic situation where user has actually given 24 responses; EDUFORM has reasoned 24 responses and 37 propositions are open. Visualization of the current learner profile (groups of learners), is shown on the right side. The users are divided into different groups of learners based on their answers on the questionnaire.