THOUGHTS ABOUT THE DEVELOPMENT OF TOOLS FOR COGNITIVE DIAGNOSIS OF STUDENTS' WRITINGS IN AN E-LEARNING ENVIRONMENT

IMBOS, Tjaart
Maastricht University
The Netherlands

Students can be stimulated to become active learners using a tool for active writing. In our university we developed such a tool: POLARIS. Active writings of students about statistical concepts are valuable for the students and the teacher. In their writings students show their understanding of statistical topics. The problem then is how to interpret the writings of students in relation to their proficiency in statistics. Advances in cognitive psychology have extended our understanding of students' learning and broadened the range of performances that can be used to acquire evidence about the developing abilities of the students. Furthermore advanced technology has made it possible to capture complex performances of students in assessment settings. In this paper the advances in both domains will be explored in order to propose a system to monitor and diagnose students' on going learning.

INTRODUCTION

Students can be stimulated to become active learners using a tool to support collaborative learning, working on statistical problems and tasks in small groups. In our university we developed such a tool: POLARIS, an acronym for Problem Oriented Leaning and Retrieving Information System. Experiences in using this program in statistics courses have been reported at ICOTS7. Active writings of students about statistical concepts are valuable for the students but also for the teacher. In their writings students show their understanding of statistical topics. The problem then is how to interpret the writings of students in relation to their proficiency of statistics.

The problem can be rephrased as how to make sense of complex data like the writings of students in discussion boards as they are used in our learning tool, POLARIS. Advances in cognitive psychology have extended our understanding of students' learning and broadened the range of performances that can be used to acquire evidence about the developing abilities of the students. Furthermore advanced technology has made it possible to capture students' complex performances in assessment settings. Cognitive psychology and artificial intelligence have developed tools for describing and representing knowledge and the use of knowledge. Firstly, the methodology for representing knowledge and comparing the differences between the knowledge of experts and novices in a domain will be described.

Next, two technological advances will be explored as tools for solving the problem to understand and diagnose the knowledge base demonstrated in the students' writings. (Ericsson & Smith, 1991; Ericsson & Simon, 1980). The first one is called Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998). The second is Bayesian Nets (BN's) (Jensen, 2001). LSA is a model that induces representations of meaning of words by analyzing the relations between words and passages in texts. The method used by LSA to capture the essence of semantic information is dimension reduction. It can be used as automatic scoring of texts. Based on this scoring of a text a group of students or an individual student could be diagnosed on a certain level of statistical proficiency.

The second method to be examined for usability in the assessment of progress in knowledge of statistical domains is the use of BN's in complex situations. It is well known that item response theory (IRT) has improved testing and assessing. A problem in using IRT is the limited scope of the tasks that can be used without violating the basic assumptions underlying IRT models. Graphical modelling (GM) provides methods for working with such complex situations where multivariate dependencies are inevitable. Using it in a predictive framework, as is needed in our situation, GM is referred to as BN's. Assessment models for complex situations as we have in mind here, could be built around some central ideas: defining unobservable variables to explain patterns of observable responses; assembling tasks so that some sources of variation accumulate and others do not, and using probability-based inference to deal with accumulating information about the latent student competency variables as assessment, here the writings of the students, proceeds. In the paper we will elaborate on these topics and propose a system for ongoing assessment of student writings in an e-learning environment. Because it's work in progress, empirical results can not be presented during the paper session.

WRITING TO LEARN STATISTICS

Statistics is a complex domain and difficult to master for most students in the social and health sciences. It may help and stimulate to learn together in small groups. Individual knowledge develops through interactions with others. Collaborative learning situations elicit discussion, argumentation and explanation and stimulate verbalization and explicit formulation of concepts and processes under discussion (Van der Linden, Erkens, Schmidt, & Renshaw, 2000). Bereiter and Scardamalia are strong advocates of student communities working together to become proficient in fields of knowledge (Bereiter, 2002; Scardamalia & Bereiter, 1994) These authors introduced the concept of knowledge-building communities, where students learn to work with theoretical and practical concepts as objects. They advocate strongly that students become knowledge-builders and participate in the knowledge-building discourse. The focus is on: 1. problems and depth of understanding; 2. decentralized, open knowledge environments for collective understanding; 3. productive interaction within broadly conceived knowledge-building communities.(Bereiter & Scardamalia, 2003). In order to stimulate these learning processes we developed a program, based on modern insights in learning and cognition, called POLARIS (Ronteltap, Koehorst, & Imbos, 2007). The features of this program and its use in the domain of statistics, is described elsewhere (Imbos, Koehorst, & Vesseur, 2006).

ASSESSING STUDENTS' WRITINGS ON STATISTICS TOPICS

Students of the School of Health Sciences are assigned to small collaborative learning groups for each course on a regular base. These groups are guided by a tutor. In the academic year 2002-2003 a group of students actively used POLARIS. During a basic course on topics like: statistical testing, regression analysis, analysis of variance and the analysis of cross tabulated data they produced an amount of 167 written documents, using 30 different discussion threads, with a mean length of 5.7 lines of reasoning.

A short part of these discussion threads is shown in figure 1. In figure 1 the left part shows the discussion while the right part shows in interpreted analysis using a program for the qualitative and quantitative analysis of texts. For an overview of such programs see (Popping, 2000) These are nice and convenient methods, but very time consuming, because all the codes used to interpret the text of the students are 'hand made'. There is a need for developing an automatic scoring and classifying system to incorporate or easily combine with our program POLARIS. In the next paragraphs the prerequisites for such a system are analysed.

KNOWLEDGE BASE FOR STATISTICAL PROFICIENCY

Firstly, an extensive knowledge base for all topics relevant for our courses is needed. This base needs to be fed with expert knowledge and with knowledge from other levels: novices, student experts, advanced students and so on. This knowledge base should become the reference against which the writings of the students on a topic can be compared, scored and interpreted. The knowledge base needs to be analysed and described in terms of knowledge elements, explanations, dialogues, strategies and processes. A knowledge base is not simply a container of inert knowledge.

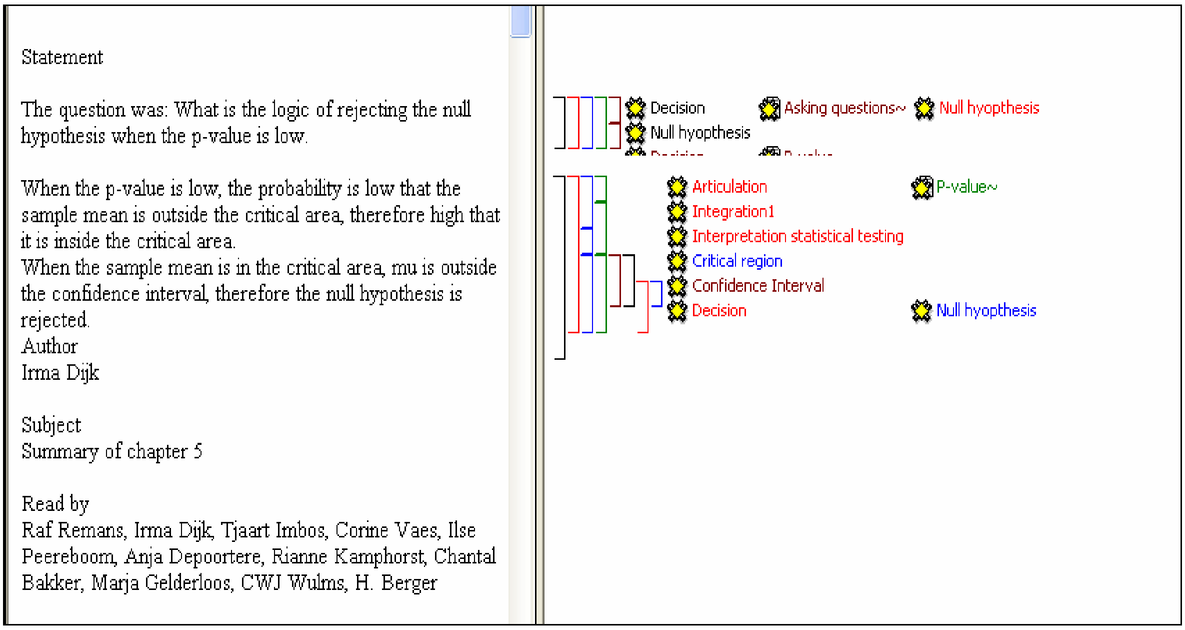

Figure 1. Part of a discussion thread in POLARIS

Knowledge is organised in mental structures. Some properties of these individual, internal mental structures might reflect what a person understands: the richness of the mental structures, the integration of the studied material and additional concepts and links. To measure understanding these properties they have to be preserved during elicitation, next they have to be expressed in external knowledge representations, and subsequently evaluated (Bude, 2007). It is possible to elicit knowledge of experts on statistical topics (Bude, 2007). The results of these elicitations can be analysed using cognitive analysis of discourse as applied in research on expert tutoring (Frederiksen, Donin, & Roy, unpublished manuscript). A description of cognitive processes, reasoning, and knowledge representations can then be made. In order to use this information it needs to be classified and compiled. The result of this is an expert model for statistical topics that are taught. This model is needed to 'understand' the writings of students and to guide them to better and deeper understanding of the knowledge they are learning.

The knowledge base of an expert consists of different types of knowledge. The most well-known are: declarative and procedural knowledge. However, for our purpose a more elaborate system of knowledge qualifications is needed. An experts' knowledge base can also be characterised by different qualities such as: level of knowledge (deep or surface); generality of knowledge; level of atomization of knowledge; modality of knowledge and structure of knowledge. Quite a lot of knowledge terms can be found in the literature. A system to organise different 'knowledge theories' is proposed by de Jong and Ferguson-Hessler (de Jong & Ferguson-Hessler, 1996). This system is interesting and usable because it combines elements of a knowledge base with the function it fulfils in a performance or task. For our students this is an important characteristic. We don't want to just teach knowledge but give knowledge that can be used to solve research problems in the domain of the health and life sciences.

The knowledge matrix these authors propose is characterised by two dimensions: *types of knowledge* and *qualities of knowledge*. By combining these two dimensions, a suited description of a knowledge base relevant to certain types of problems and tasks can be created. The resulting qualities of knowledge are not all independent. Some qualities not only refer to separate types of knowledge but also to larger units of the knowledge base, for instance to schemata.

Using knowledge matrices in this way, it is theoretically possible to describe a complete knowledge base. For a complete statistics curriculum the total knowledge base can easily become large. In order to deal with this problem, knowledge compilation techniques as used

in Artificial Intelligence (AI) can be applied. Although the problem of knowledge compilation is not a simple one, a lot of proposals and techniques are available.
(See: http://citeseer.ist.psu.edu/ArtificialIntelligence/NaturalLanguageProcessing/date.html).

ON THE DIMENSIONALITY OF A STATISTICS KNOWLEDGE BASE

Researchers who use quite a number of variables frequently use data reduction techniques as factor analysis and cluster analysis. In the case of discussion forums in POLARIS, quite a lot of information becomes available in the form of words and sentences related to statistical topics. Here also data reduction is needed. For qualitative data, as students writings, a technique comparable to factor analysis is available: Latent Semantic Analysis (LSA) (Landauer , Foltz & Laham, 1998).

Suppose two students write about their understanding of statistical testing. Even if they both have a good understanding of the topic, compared to that of an expert, their writings will still differ. They can be seen as different but also comparable with both having a good understanding of the subject matter. Therefore the objective is how to explain that the first writing is similar to the second, or how two parts of the texts can possibly be compared quantitatively, indicating reasoning processes of the two students. It seems possible to introduce a theory and methodology for writing tasks based on LSA. This theory addresses issues related to the induction, representation and application of knowledge. Actually, LSA infers knowledge from many weak constraints about statistical topics that are present in the writings of students while they are learning. LSA does not represent a whole *knowledge space* but only the paths students have chosen to find their way in that space. This offers a nice view of how the knowledge space is understood by the students. LSA is a *computational* theory on how students learn to find their way in the knowledge space of statistics and how that space can be described. The features of LSA are: (1) It does not assume independence of writing actions, instead it uses dependencies to infer the structure in the writings; (2) LSA reduces the dimensionality of the space. (3) There are no *a-priori* assumptions about the knowledge space. LSA is self-organising.

DESCRIPTION OF LSA

LSA is a machine-learning model that induces representations of the meaning of words by analysing the relation between words and passages in large bodies of text. LSA is both a method for educational applications as well as a theory of knowledge representation to model comprehension of statistical topics. The method used by LSA captures the essential information in text passages, while ignoring accidental and inessential word usages. The method selects the most important dimensions from a co-occurrence matrix using *single value decomposition*. As a result LSA can be used to assess semantic similarity between two any two samples of text in an automatic way. That is what we need for the problem described in this paper. Using LSA, student writings can be compared with the expert knowledge base as meant in earlier paragraphs. LSA has been used in applied settings with a high degree of success like essay grading, automatic tutoring and in human language acquisition simulations and in modelling comprehension phenomena (Landauer & Dumas, 1997).

USING BAYESIAN NETS TO RELATE MODELS IN AN ASSESSMENT SYSTEM

Suppose we are able to build an automatic system using complex knowledge bases and using LSA to reduce the dimensionality of it, then the next problem arises: how to make sense of the complex data that result? How to interpret the writings of students as evidence for their statistical competency? We cannot use the statistical methods and rules-of-thump developed for class room quizzes and standardised tests. To solve that problem two conditions needs to be fulfilled. The first is that we need tools of probability-based reasoning that have proved to be useful in modern test theory and adopt these for more complex situations that result from a system we have in mind. Second, we need more than a scoring system. We also need designing principles for a complex assessment system. Such principles should guide us through questions as: what inferences do we want to make; what observations do we need to ground them; what situations can evoke them, and which reasoning can connect them. In other words we need, a framework for designing assessments, called an *evidence centred* framework (Mislevy, Steinberg, & Almond, 2002).

Fundamental for developing such a framework is to know and describe the building blocks for understanding statistics. We need insights in evidentiary reasoning of domain experts and students. The knowledge base described earlier and how it is used can lead to insight in the principles of reasoning and usage of knowledge in order to solve real research problems. In order to achieve a coherent assessment system three basic models should be present and connected. They are: student models; evidence models, and task models.

The *student model* describes which competencies should be assessed. The model contains student variables to approximate aspects in the domain of statistics. Students are measured and scored on these variables, but actually they are unobservable. The perspective on these variables can be behaviourist, trait, cognitive or situational. In all cases the problem is the same: constructing the student variables from limited evidence. The number and nature of the student variables depend on the purpose of the assessment. It can be one summarising variable or several variables. If there is more than one variable in the student model the empirical or theoretical relations between them can be described for each student at a certain point in time. These relations can be described in terms of a probability distribution that can be updated as new evidence about the student becomes available. In that case the student model takes the form of a Bayesian inference network, or Bayes Net (BN) (Jensen, 2001). BN's offer the methodology to manage knowledge and uncertainty in the complex assessment systems that we have in mind.

The *evidence model* is the heart of evidentiary reasoning in assessment. Here the arguments about why and how the observations are made. The evidence model consists of two parts: the responses or writings of the students, the *students' products*, leading to *observed variables*, i.c. scores on aspects of the students' product. The second part of the evidence model is the *statistical submodel*, which expresses how observed variables in probability depend ,on student variables. Examples of models of these kinds are: classical test theory; item response theory, latent class models, and factor analysis. These models can be expressed as special cases of BN's as an extension of the relation between student variables and observed variables.

Finally, the *task model* constitutes situations to elicit the behaviour used in the evidence model. The task model provides a frame work for constructing and describing the situations in which the students act. The task model is the situation in which the students produce their work products. It is the input for the evidence model.

BAYESIAN NETS CONNECT SUBMODELS OF AN ASSESSMENT SYSTEM

Item response theory (IRT) has advanced educational measurement substantially. Some applications are difficult however without violating the assumptions as conditional independence and the limited capability for dealing with multiple aspects of knowledge or skills. Graphical modelling (GM) (Jensen, 2001) provides methods for working with such complex dependencies. GM's used in a predictive framework are also addressed to as BN's. In IRT an examinees capability is expressed in terms of an unobservable student variable, θ. The responses of a student are assumed to be independent, conditional on both θ and the characteristics of the writing task. For n tasks scored as correct or incorrect the following can be specified:

$$P(\chi_1, \chi_2, \cdots \chi_n) = \prod_{j=1}^{n} P(\chi_j \big| \theta, \beta_j), \qquad 1$$

where, x_j is the response to task j
β_j is the difficulty of task j, and

$$P(\chi_j \big| \theta, \beta_j) = EXP\left[x_j(\theta - \beta_j)\right] \,]/[\, 1 + EXP(\theta - \beta_j)]$$

The observed response vector **x** $(\chi_1, \chi_2, \cdots \chi_n)$ becomes the likelihood function for θ, $L(\theta|\mathbf{x}, \mathrm{B})$. Bayesian inference is based on the posterior distribution $p(\theta|\mathbf{x}, \mathrm{B}) \propto L(\theta|\mathbf{x}, \mathrm{B})\, p(\theta)$, a summary of which can be given by the posterior mean of θ and the posterior variance $\mathrm{var}(\theta|\mathbf{x}, \mathrm{B})$. An IRT model can be depicted as a GM in the same way as is done in structural equation modeling with θ as a single parent of all writing tasks, graphically depicted as arrows starting from θ to the observed writing response X_j. At the beginning of a discussion thread the full joint distribution $P(\chi_1, \chi_2, \cdots \chi_n, \theta)$ characterizes a student's θ and his future responses to writing tasks. This distribution then can be obtained as the product of the initial distribution of θ, $p(\theta)$ times the conditional distribution of each response to the writing task $P(X_j|\theta)$, given by the IRT model and $p(\theta)$ is the examiner's 'belief' in an examinees θ. Based on new information the examinees θ can be updated leading to a new posterior distribution for θ . All new information leads to new inferences about students'θ. This process continues until the written discussion is terminated. Using GM's as a predictive framework the resulting BN's combine the student model with the task model of the assessment system and visa versa. For a detailed discussion see (Mislevy, 1994). Such a complete system is of course complicated. But estimation procedures are available (Mislevy, 1994; Murphy, 2001).

CONCLUSION

Developing a system to assess the writings of students in an e-learning environment is difficult. Such a system needs a knowledge base elicited from experts. It also needs a scoring system to compare student writings with some standard. Finally a system as intended, needs an instrument to update the assessments in the case of incremented learning. Also not easy but a system as is needed can be developed with the use of the methodology and technology described in this paper. A lot needs to be investigated further, but it seems that a project for the development of a system with automatic scoring and diagnosing students' proficiency of statistics lies ahead. Using cognitive methods for the description of knowledge and its use, latent semantic analysis and combining IRT and Bayesian networks are the main tools in that project.

REFERENCES

Bereiter, C. (2002). *Education and mind in the knowledge age*. Mahwah, NJ.: Lawrence Erlbaum Associates, Inc.

Bereiter, C., & Scardamalia, M. (2003). Learning to work creatively with knowledge. In E. de Corte, L. Verschaffel, N. Entwistle & J. van Merrienboer (Eds.), *Powerful learning environments: Unravelling basis components and dimensions* (pp. 55-68). Oxford: Elsevier Science Ltd.

Bude, L. (2007). *On the improvement of students' condceptual understanding in statistics education*. Maastricht: Maastricht University press.

de Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113.
Ericsson, K. A., & Smith, J. (Eds.). (1991). *Towards a general Theory of Expertise* (1 ed.). Cambridge: Cambridge University Press.
Ericsson, K. A. & Simon., H.A. (1980). Verbal Reports as Data. *Psychological Review, 87*(3), 215-251.
Frederiksen, C. H., Donin, J., & Roy, M. (unpublished manuscript). *Human tutoring as a model for computer tutors: studying human tutoring from a cognitive perspective*.Unpublished manuscript, Montreal.
Imbos, T., Koehorst, A., & Vesseur, A. (2006, 2 - 7 July).
A tool for collaborative knowledge-building in statistics education. Paper presented at the 7th International Conference on Teaching Statistics, Salvador, Bahia, Brazil.
Jensen, F. V. (2001). *Bayesian networks and decision graphs*. New York: Springer.
Landauer, T. K., & Dumas, S. T. (1997). A solution to Plato's problem: The latent semantic theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211-240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.
Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*(4), 439 - 483.
Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3-62.
Murphy, K. P. (2001). *The bayes net toolbox for matlab*.Unpublished manuscript, Berkeley.
Popping, R. (2000). *Computer-assisted text analysis*. London: Sage Publications.
Ronteltap, F., Koehorst, A., & Imbos, T. (2007). Online Knowledge-building Interactions in
Problem-based Learning: The POLARIS Experience. In T. O. Seng (Ed.), *Problem-based learning in eLearning Breakthroughs* (pp. 169-184). Singapore: Thomas Learning.
Scardamalia, M., & Bereiter, C. (1994). Computer Support for Knowledge-building Communities. *The Journal of the Learning Sciences, 3*(3), 265-283.
Van der Linden, J., Erkens, G., Schmidt, H., & Renshaw, P. (2000). Collaborative learning. In R. J. Simons, J. v. d. Linden & T. Duffy (Eds.), *New Learning,* 37-57. Dordrecht/Boston/London: Kluwer Academic Publishers.