A Bayesian approach to modelling users' information display preferences

Beate Grawemeyer and Richard Cox

Representation & Cognition Group
Department of Informatics, University of Sussex, Falmer, Brighton BN1 9QH, UK
{beateg, richc}@sussex.ac.uk

Abstract. This paper describes the process by which we constructed a user model for ERST - an External Representation Selection Tutor - which recommends external representations (ERs) for particular database query task types based upon individual preferences, in order to enhance ER reasoning performance. The user model is based on experimental studies which examined the effect of background knowledge of ERs upon performance and preferences over different types of tasks.

1 Introduction

Successful use of external representations (ERs) depends upon skillful matching of a particular representation with the demands of the task. [1] and [2] provide numerous examples of how a good fit between a task's demands and particular representations can facilitate search and read-off of information. [3] provides a review of studies that show that tasks involving perceiving relationships in data or making associations are best supported by graphs, whereas 'point value' read-off is better facilitated by tabular representations. But people differ in their representational expertise and in their individual ER preferences for particular task types.

We describe the development of a user model for ERST - an External Representation Selection Tutor. This has been constructed on the basis of empirical data gathered from two psychological experiments. The study reported earlier in [4] investigated the representation selection and reasoning behaviour of participants who were offered a choice of information-equivalent data representations (e.g. tables, bar charts, etc.) upon various types of database query tasks. Following that earlier study, the aim of the experiment reported here was to investigate the degree to which task types are more representation-specific[1] than others, with respect to reasoning performance and response latency. For both experiments, a prototype automatic information visualization engine (AIVE) was used to present a series of questions about the information in a database. The results of the experiments indicated that ERST needs to take into account a) individual differences (like user's ER preferences), b) their level of experience and c) the domain task characteristics, in order to provide effective ERs that reflect

[1] These are tasks for which only a few, specialised, representational forms are useful.

individual needs and therefore enhance ER reasoning performance. A Bayesian network for ERST has been constructed based on the experimental results.

2 Experiment

The aim of the study reported here was to investigate the degree to which some task types are more representation-specific than others in terms of reasoning performance and response latency. We were interested to discover what task types can be answered successfully with a variety of different representations and which tasks were more constrained in terms of useful representations. We define representation-specificity as follows: For highly representation-specific tasks only a few, specialised, representational forms are useful. Whereas for low representation-specific tasks a range of different types of representations can be used successfully to solve the problem.

This study builds on previous work [4] in the following ways: The AIVE system has been changed from a Java Applet to a stand alone Java Application in order to produce more accurate timing data[2]. The system has been extended with a new set of more representation-specific critical task types. We also employ a new approach to the assessment of subjects 'graphical literacy' [5].

Our hypotheses were, that different degrees of representation-specific tasks types influence participants' performance on a) their ER selection *skill*; b) time to answer the database query (latency); and c) the correctness of their response on the database query task.

Procedure

Twenty participants were recruited (5 software engineers, 1 graphic designer, 1 html programmer, 2 IT business managers, 7 postgraduate students, and 4 research officers/fellows). Each participant completed 4 pre-experimental ER tasks followed by the AIVE database query problem solving session.

The ER pre-tasks [5] assessed the visual recognition of particular ERs requiring real/fake decisions, ER categorisation, functional knowledge of ERs, and specific naming. This represents an information processing approach to the assessment of 'graphical literacy' [5] and these ER tasks were employed as pre-tests of ER knowledge. Participants then performed the AIVE database query tasks using the same procedure as that used in [4]. Participants were asked to make judgments and comparisons between cars and car features based on database information. The database contained information about 10 cars: manufacturer, model, CO^2 emission, engine size, *etc*. Each subject responded to 30 database questions, of which there were 6 types: identify; correlate ; quantifier-set; locate; cluster; compare negative. Participants were informed that to help them answer the questions, the system (AIVE) would supply the appropriate data from the database. AIVE also offered participants a choice of representations of the data. They could choose between various types of ERs, *eg.* set diagram, scatter plot,

[2] The timing data in the previous experiment lacked precision because of time delays caused by the internet connection.

bar chart, sector graph, pie chart and table (the full range of representations were offered by the system on all queries). The options were presented in the form of an array of buttons each with an icon depicting - in stylised form - an ER type (table, scatterplot *etc*). Participants were told that they were free to choose any ER, but that they should select a form of display they thought was most likely to be helpful for answering the question. Participants then proceeded to the first question, read it and selected a representation. The spatial layout of the representation selection buttons was randomized across the 30 query tasks in order to prevent participants from developing a set pattern of selection. Based on the literature (eg.[1]) a single 'optimal' ER for each task was identified[3]. After the participant made his/her representation choice, AIVE generated and displayed the representation instantiated with the data required for answering the question. Participants then answered the question using the chosen display. Participants were not permitted to select a different representation following their initial selection. This constraint was imposed in order to encourage participants to carefully consider which representation was best matched to the task. Following a completed response, participants were presented with the next task and the sequence was repeated. The following data were recorded: (1) the randomized position of each representation icon from trial to trial; (2) the user's representation choices; (3) time to read question and select representation; (4) time to answer the question using chosen ER; and (5) participants' responses to questions.

Results and Discussion

To recapitulate, each of 20 subjects performed 30 AIVE tasks (600 data points in total). The simple bivariate correlations across all AIVE tasks for display selection accuracy (DSA), database query answering accuracy (DBQA), display selection latency (DSL) and database query answering latency (DBQL) were: DSA correlated significantly and positively with DBQA ($r=.30$, $p<.01$); DSA is significantly negatively correlated with DSL ($r=.-17$, $p<.01$); DSA and DBQL are significantly negatively correlated ($r=-.32$, $p<.01$); There is a significant negative correlation between DBQA and DBQL ($r=-.28$, $p<.01$); DSL and DBQL are significantly positively correlated ($r=.30$, $p<.01$). The results across all AIVE task types show that good display-type selection will lead to better query answering performance. The selection latency results show that a speedy selection of a display type in AIVE is associated with a good display-type choice. Less time spent responding to the database query question is associated with a good display-type choice and correct query response. This suggests that the selection and database query latencies may be used in ERST as predictors of users' DSA and DBQA performance. Looking at the different task types, these results differ extensively in terms of representational specificity. As shown in figure 1, 77% of AIVE *correlate* type queries were answered correctly by participants. Moreover,

[3] However, each AIVE query type could *potentially* be answered with any of the representations offered by the system with the exception of quantifier-set tasks for which the only real effective representation was a set diagram.

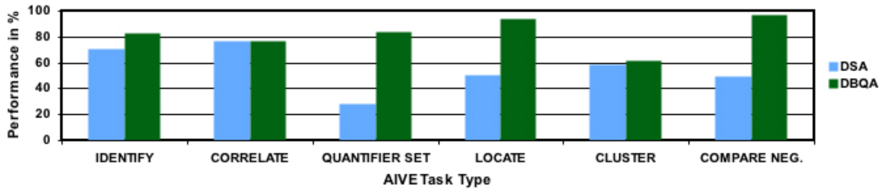

Fig. 1. AIVE task type performance

in 77% of the cases they chose the most appropriate ER display (scatter plot) from the array of display types offered by AIVE. The correlation coefficients for AIVEs' *correlate* tasks were: DSA and DBQA are significantly positively correlated (r=.67 , p<.01); DSL and DBQL are significantly positively correlated (r=.56, p<.01). The results suggest that good display-type selection is associated with accurate query answering performance. Longer display selection latency is associated with longer time spent responding to the database query question. Hence there does not seem to a be a speed/accuracy trade-off in display selection - participants either know which ER to choose and get on with the task, or they don't. In contrast, the locate task could be answered effective with different kinds of data displays. Overall, subjects locate task queries were answered with a high degree of accuracy (94%). However, in only in 51% cases did participants choose the 'right' representation (table or matrix ER). A range of other AIVE display forms were also effective (bar and pie charts, scatterplots). No significant correlations between AIVEs variables were detected. The results show that the correlate task is more representation-specific than the locate task. Therefore in order to predict DSA and DBQA performance ERST needs to include a variety of tasks that differ in terms of their representational specificity.

3 ERST's Bayesian network

ERST's user model needs to track selection accuracy and database query answering performance for various display and response accuracy relationships within and across the various database query task types. ERST will need to be more stringent in its interventions on highly representationally-specific task types such as correlate tasks but will be able to be more lenient on more display-heterogeneous tasks. Various machine learning techniques differ in the advantages and disadvantages they have for particular applications. They also differ in terms of the user data need for the adaptation process. According to our experimental findings, the most appropriate implementation for ERST's user model is a Bayesian network. The network is being constructed and 'seeded' with the empirical data so that it can monitor and predict users' ER selection preference patterns within and across query types, relate query response accuracy and latencies to particular display selections (DS) and contrive query/display option combinations to probe an individual user's degree of graphical literacy. The structure of the Bayesian network based on the experimental data can be seen in Figure 2. For example the arc between DSA and DBQL represents the association that good display selection results in better query performance and the link between DSL and DSA represents the finding that a speedy selection of a display type in AIVE is associated with a good display-type choice. The em-

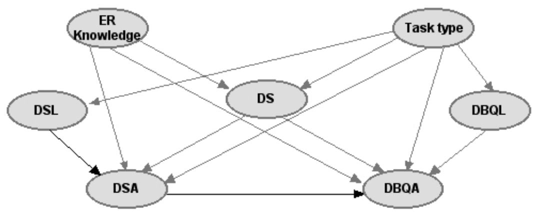

Fig. 2. ERSTs' Bayesian network

pirical data is used to instantiate values in the relevant conditional probability tables (CPTs) at each node of the model.

4 Conclusion and future work

In this paper we described our process of constructing a Bayesian network for modelling user's ER preferences, based on experimental data. The aim of this model is to provide effective ERs that reflect individual needs and therefore enhance ER reasoning performance. The resulting Bayesian network structure is based on empirical findings and the gathered data is reflected in the CPTs. The next step in our research will focus on ERST's adaptation decisions. For example, when and how to recommend ERs and the manner in which ERST utilises data from new users. The user model will need to be evaluated through user studies and will be iteratively refined. We also plan to investigate how well ERST is able to accommodate individual differences in ER selection preference.

References

1. Day, R.: Alternative representations. In Bower, G., ed.: The Psychology of Learning and Motivation **22**, New York, Academic Press (1988) 261–305
2. Norman, D.A., ed: Things that make us smart. Addison-Wesley, MA (1993)
3. Vessey, I.: Cognitive fit: A theory-based analysis of the graphs versus tables literature. Decision Sciences **22**, (1991) 219–241
4. Grawemeyer, B., Cox, R.: The effect of knowledge-of-external-representations upon performance and representational choice in a database query task. In Diagrammatic Representation and Inference: Third International Conference, Diagrams 2004, Berlin, Springer (2004)
5. Cox, R., Romero, P., du Boulay, B., Lutz R.: A Cognitive Processing Perspective on Student Programmers' 'Graphicacy'. In Diagrammatic Representation and Inference: Third International Conference, Diagrams 2004, Berlin, Springer (2004)