Reimplementing a Diagrammatic Reasoning Model in Herbal

Maik B. Friedrich (maik.friedrich@dlr.de)
German Aerospace Center, Braunschweig

Frank E. Ritter (frank.ritter@psu.edu)
College of IST, Penn State

Keywords: Herbal, Diag, Diag-H, reimplementing, cognitive model

Introduction

This paper builds upon a study of how people find faults in a simple device and a corresponding cognitive model (Ritter & Bibby, 2008). This existing model, Diag, was implemented in Soar 6 and is based on the idea that learning consists of procedural, declarative, and episodic learning. Diag was developed to analyze human behavior while solving a simple diagrammatic problem (Ritter & Bibby, 2008), a task with similarities to many important real world problem solving tasks. Because Diag predicted astonishing results and is implemented in a version of Soar that is no longer supported, an implementation in an up-to-date cognitive architecture is necessary to make the model available again and more flexible to future changes.

We maintained Diag's basic structure while reimplementing it in a high-level behavior representation language, Herbal, that generates Soar models and can generate different variants more quickly that in Soar directly. Herbal compiles into Soar 9, which allows not only that the model can be used again for further research with current Soar models but it is also made accessible to more researchers. This newly implemented model, called Diag-H, was validated by comparing its predictions to the existing data. It could be shown that Diag-H creates almost the same results as Diag but also incorporates the advantages of Herbal.

Diag task and results

The Diag task is called fault-finding task (FFT) and builds upon an interface with 4 switches and 7 light that represent an electrical circuit with 7 different components that are connected via switches. The task consists of a combination from interface information and circuit condition to determine which component is faulty.

Diag was implemented with the effort to predict human reaction times and learning behavior while solving the Diag task. The models strategy is based on the energy flow running through the circuit. A light gets selected based on its position in the circuit and tested by the position of the switches and if its lit up or not. On the Problem Space Computational Model (PSCM) (Lehman, Laird, & Rosenbloom, 1996; Newell, Yost, Laird, Rosenbloom, & Altmann, 1991) level, Diag consists of problem spaces that are hierarchically ordered to solve the FFT by testing the components stepwise.

For validating the Diag model, a user study with 10 participants was run. The participants were instructed how the circuit components are connected, how the components are represented on the interface, and what their task is. While solving the FFT the participants had to recall the circuit diagram from memory, combine it with the presented interface constellation, and identify the faulty component. The results showed that the average proportion of variability in problem-solving time per participant was 79%. The task, the study, and the results are described in detail in Ritter and Bibby (2008).

Diag-H

The reimplementation of Diag was done in Herbal (Haynes, Cohen, & Ritter, 2009), a high-level language based on the PSCM that produces models that can run in Soar and Jess. Because of the use of Herbal the reimplementation required an understanding of the PSCM and visual modeling techniques. This serves as an example of how Herbal can provide modelers that have no strong programming background access to the complicated machinery used by cognitive architectures that may traditionally be out of their reach.

Because Diag-H is a reimplementation of the Diag structure, the most important effort was to copy the structure accurately. Diag-H uses the same structure of problem spaces and strategy to solve the FFT. The reimplementation process was supported by Herbal because of the direct implementation of the PSCM. This means Herbal models implement problem spaces directly and assign them hierarchically.

The task knowledge in Diag-H is stored in operators. An operator in Herbal is a combination of generic conditions and actions that can be combined as required. 93 conditions and 56 actions were modeled and combined to 85 operators. Herbal compiles Diag-H into 187 Soar rules.

Diag-H predictions and the existing data

To validate Diag-H, we used the data from Ritter and Bibby (2008). The number of Soar model cycles with learning turned on was used to predict solution times from Diag-H. Using linear regression between Diag-H predictions and the existing user data, an average motor output time (B = 1.42 s) and an average time as slope of decision cycles (0.187 ms) was calculated. To determine how accurate the model predicts individual behavior, the predicted times (as slope of decision cycles * decision cycles + intercept = 0.187 ms * decision cycles + 1.42s) were compared to the observed problem solving times.

Each participant saw a different order of the 20 faults. Figure 1 shows the individual problem-solving time for participant 8 and the predicted times aggregated over this stimulus predicted by Diag-H. This example shows how well the Diag-H predictions fit to the user data.

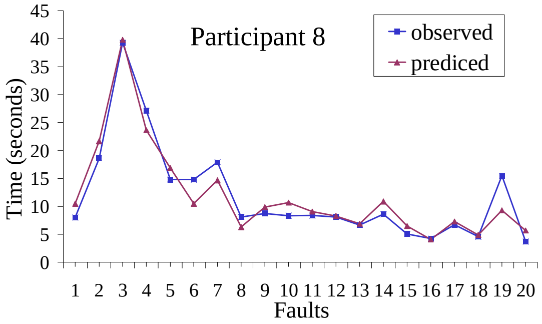

Figure 1: The observed and predicted problem-solving times over 20 trials for participant 8.

To compare the Diag-H predictions further to the user data, each set of model cycle per run was regressed to the problem-solving times for each participant individually. The average proportion of variability in problem-solving time per participant accounted by Diag-H was $r^2 = 72.2\%$. By removing two non significant participants from the analysis the significance reaches $r^2 = 87\%$.

These comparisons showed that Diag-H was able to predict the existing participant performance to a good extent. Similar to Diag, Diag-H also has problems in predicting the performance of participants P5 and P7. However, when comparing the correlations for the predictions per fault, per trial, and per participant Diag-H is constantly 5% less accurate than Diag.

Summary

We have described the use of a high level behavior representation language, Herbal, to reimplement Diag, a model that solves a diagrammatic reasoning task. The reimplementation, Diag-H, was validated by testing whether it creates the same predictions as Diag. Diag-H uses the same strategy and reaches almost the same results by predicting human behavior and combines this with Herbal advantages. A Herbal model can predict similar results to a Soar model but has a shorter implementation time. The generic Herbal structure allows quick adaptations to future requirements and further development of models. These results allow proceeding with research on the Diag task supported by the Diag-H model.

Diag-H offers several new possibilities for research. One aspect is implicated by two participants (P5 & P7) that did not fit either the existing Diag predictions or the updated Diag-H predictions. Because these participants' error rates were not significantly higher than the average, the results suggest that they used a different strategy than Diag-H. Therefore, the development of several strategies will be necessary for a detailed analysis of the performance of these two participants. Through the use of Herbal as implementation language the process of creating new strategies will be simplified. In the future even Herbal compiled ACT-R models will be available (Paik, Kim, & Ritter, 2009).

Acknowledgements

DLR and ONR provided resources for this work and ONR supported the development of Herbal.

References

Bibby, P. A., & Payne, S. J. (1993). Internalisation and the use specificity of device knowledge. *Human-Computer Interaction, 8*, 25-56.

Bibby, P. A., & Payne, S. J. (1996). Instruction and practice in learning to use a device. *Cognitive Science*, *20*(4), 539-578.

Haynes, S. R., Cohen, M. A., & Ritter, F. E. (2009). A design for explaining intelligent agents. *International Journal of Human-Computer Studies, 67*(1), 99-110.

Lehman, Laird, & Rosenbloom. (1996). A gentle introduction to Soar, an architecture for human cognition. In S. Sternberg & D. Scarborough (Eds.), *Invitation to cognitive science* (Vol. 4). Cambridge, MA: MIT Press.

Newell, A., Yost, G. R., Laird, J. E., Rosenbloom, P. S., & Altmann, E. (1991). Formulating the problem space computational model. In R. F. Rashid (Ed.), *Carnegie Mellon Computer Science: A 25-Year commemorative* (pp. 255-293). Reading, MA: ACM-Press (Addison-Wesley).

Ritter, F. E., & Bibby, P. A. (2008). Modeling how, when, and what learning happens in a diagrammatic reasoning task. *Cognitive Science*.