

Self-Modeling in Companion Cognitive Systems: Current Plans

Kenneth D. Forbus and Thomas R. Hinrichs
Qualitative Reasoning Group, Northwestern University
1890 Maple Avenue, Evanston, IL, 60201, USA

Background: We are developing *Companion Cognitive Systems*, a new architecture for software that can be effectively treated as a collaborator. Here is our vision: Companions will help their users work through complex arguments, automatically retrieving relevant precedents, providing cautions and counter-indications as well as supporting evidence. Companions will be capable of effective operation for weeks and months at a time, assimilating new information, generating and maintaining scenarios and predictions. Companions will continually adapt and learn, about the domains they are working in, their users, and themselves.

The ideas we are using to try to achieve this vision include:

Analogical learning and reasoning: Our working hypothesis is that the flexibility and breadth of human common sense reasoning and learning arises from analogical reasoning and learning from experience. Within-domain analogies provide rapid, robust predictions. Analogies between domains can yield deep new insights and facilitate learning from instruction. First-principles reasoning emerges slowly, as generalizations created from examples incrementally via analogical comparisons. This hypothesis suggests a very different approach to building robust cognitive software than is typically proposed. Reasoning and learning by analogy are central, rather than exotic operations undertaken only rarely. Accumulating and refining examples becomes central to building systems that can learn and adapt. Our cognitive simulations of analogical processing (SME for analogical matching, MAC/FAC for similarity-based retrieval, and SEQL for generalization) form the core components for learning and reasoning.

Distributed agent architecture: Companions will require a combination of intense interaction, deep reasoning, and continuous learning. We believe that we can achieve this by using a distributed agent architecture, hosted on cluster computers, to provide task-level parallelism. The particular distributed agent architecture we are using evolved from our RoboTA distributed coaching system, which uses KQML as a communication medium between agents. A Companion will be made up of a collection of agents, spread across the CPUs of a cluster. We are assuming roughly ten CPUs per Companion, so that, for instance, analogical retrieval of relevant precedents proceeds entirely in parallel with other reasoning processes, such as the visual processing involved in understanding a user's sketched input.

Robustness will be enhanced by making the agents "hot-swappable", i.e., the logs maintained by the agents in operation will enable another copy to pick up (at a very

coarse granularity) where a previous copy left off. This will enable an Agent whose memory is clogging up (or crashes) to be taken off-line, so that its results can be assimilated while another Agent carries on with the task. This scheme will require replicating the knowledge base and case libraries as necessary to minimize communication overhead, and broadcasting working memory state incrementally, using a publish/subscribe model, as well as disk logging. These logs will also be used for adaptation and knowledge reformulation. Just as a dolphin only sleeps with half of its brain at a time, our Companions will use several CPU's to test proposed changes by "rehearsing" them with logged activities, to evaluate the quality and performance payoffs of proposed learned knowledge and skills.

Modeling: We are using the DARPA subset of Cyc KB contents, augmented with our own work on qualitative and spatial representations and reasoning, as the starting point for this effort. Declarative models will be used whenever possible, to enhance opportunities for incremental adaptation and learning. Logs of user interactions will be kept and used for experiments in learning dialogue, user, and task models. Logs of internal operations will be kept for experiments in learning self-models.

Status: Our initial domains for Companion experiments include everyday physical reasoning and tactical decision games. The interface will use our sketching systems, combined with a concept map system, textual inputs, and specialized dialogues. At this writing (1/31/04) a 10 CPU cluster has been installed and the first-cut agent shells are being debugged on it. The bootstrap reasoning facilities and initial domain representations are being assembled from our prior work on projects supported by DARPA, ONR, and NSF. Corpora and examples needed for our domain learning experiments are being selected and formalized.

Issues: Our Companion architecture raises a number of issues for building self-aware systems, including

- How should resource reallocation decisions be made, given changing demands in terms of interaction and background assignments?
- What contents and level of detail are needed in the agent logs to support efficient hot-swapping?
- How should the agent logs be organized to support effective improvement of operations and self-modeling?

Acknowledgements: This research is supported by DARPA IPTO.