

Companion Cognitive Systems: A step towards human-level AI

Kenneth D. Forbus and Thomas R. Hinrichs

Qualitative Reasoning Group, Northwestern University
1890 Maple Avenue, Evanston, IL, 60201, USA

Abstract

We are developing *Companion Cognitive Systems*, a new kind of software that can be effectively treated as a collaborator. Aside from their potential utility, we believe this effort is important because it focuses on three key problems that must be solved to achieve human-level AI: *Robust reasoning and learning, performance and longevity, and interactivity*. We describe the ideas we are using to develop the first architecture for Companions: Analogical processing, grounded in cognitive science for reasoning and learning, a distributed agent architecture hosted on a cluster to achieve performance and longevity, and sketching and concept maps to provide interactivity.

Introduction

We believe that this is an exciting time for research in artificial intelligence and cognitive science. What is known about cognition has grown significantly, and communication between fields in cognitive science has catalyzed all of them. Large-scale representational resources, such as WordNet, FrameNet, and Cyc, have become available so that one can build large knowledge systems without starting from scratch. CPUs have become fast enough and memories large enough to tackle systems that could only be dreamed about previously. The confluence of these three factors suggest to us that the time is right for more ambitious projects, building integrated systems using the best available results from cognitive science.

The effort we have just embarked on to create *Companion Cognitive Systems* represents one such project. Let us start with our practical goals for Companions. The problems we face are growing more complex, but we are not becoming any smarter. Software can help, but often it becomes part of the problem by adding new layers of complexity. We need to bring software closer to us, improving conceptual bandwidth and having it adapt to us, rather than the other way around. Our vision is this: Companions will be software aide-de-camps, collaborators with their users. 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.

It is useful to distinguish the Companions vision from the DARPA PAL (Perceptive Agents that Learn) Program. While similar in many respects, our ambitions are somewhat more modest. PALs are intended to be like Radar O’Riley: socially aware, for example, and capable of participating in multi-person spoken dialogues. Companions are intended to be, for lack of a better term, nerd sidekicks – assistants who are heavily focused on a couple of areas of interest to you, and nothing else. This lets us factor out several problems, including natural language understanding, vision, and robotics. Companions are, in essence, the classic “mind in a box” model of an artificial intelligence.

Even within these limitations, creating Companions is forcing us to tackle several problems that are crucial for building human-level AIs:

1. *Robust reasoning and learning*. Companions will have to learn about their domains, their users, and themselves.
2. *Longevity*. Companions will need to operate continuously over weeks and months at a time.
3. *Interactivity*. Companions must be capable of high-bandwidth interaction with their human partners. This includes being capable of taking advice.

Our Approach

Here are the ideas we are using to create Companions:

Robust reasoning and learning

Our working hypothesis is that the flexibility and breadth of human common sense reasoning and learning arises from analogical reasoning and learning from experience [10]. 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 [6,9], MAC/FAC for similarity-based retrieval [11], and SEQL for generalization [20]) form the core components for learning and reasoning. These components are based on Gentner's structure-mapping theory of analogy and similarity [16], a psychological account with considerable evidence behind it (cf. [17]).

These simulations have been successfully used to model a large number of psychological findings, and both SME and SEQL have been used to make new psychological predictions which, in turn, have been borne out in experiments with human subjects (cf. [21]). SME and MAC/FAC have already been used in performance systems, involving several different large knowledge bases (cf. [13,25]). One of the key hypotheses we are testing in Companions is the idea that most learning and reasoning can be handled via analogical processing. In other words, it's structure-mapping all the way down.

Longevity and performance

Companions will require a combination of intense interaction, deep reasoning, and continuous learning. We plan to 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 [12], which uses KQML [23] 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.

Interactivity

People communicate with each other in a variety of ways. Natural language is certainly one method, but we also sketch, and people who communicate with each other frequently evolve their own shared signals for communication. We are not tackling natural language in this project, since by itself it harbors several equally ambitious projects in it. Instead, we are building on our work on sketch understanding [14,15] to provide high bandwidth interaction. Our nuSketch approach is uniquely suited for Companions, since it does not restrict us to narrow domains.

Not everything can be sketched, of course. For other kinds of interactions, we are building on the idea of concept maps [27]. Our VModel qualitative modeling system, which uses concept maps founded on a QP theory semantics, has been successfully used by middle-school students in classroom experiments in the Chicago Public School system [8]. This experience, along with SRI's SHAKEN system [29] and the UWF work on concept maps [2], leads us to believe that we can make concept map systems that professionals will find useful.

Modeling in Companions

We expect that several kinds of models will play important integrative roles in Companions. Models will be used to mediate and coordinate between interaction modalities, summarize internal state, and facilitate user interaction. For a Companion to provide understandable explanations, persuasive arguments, or behave robustly, we believe it must have concise, accessible representations of its current state, beliefs about the user, and dynamic context. We see four kinds of specialized models as being crucial: (1) *Situation* and *domain models* capture the current problem and relevant knowledge about it. (2) *Task* and *dialogue models* describe the shared task and where the human/computer partnership is in working on it. (3) *User models* capture the idiosyncratic preferences, habits, and utilities of the human partner(s). (4) *Self-models* provide the Companion's own understanding of its operations, abilities, and preferences (cf. [24]). Exploring how systems can learn, adapt, and exploit these models over the long term to support interaction and problem solving is one of the major goals of this project. We discuss each kind of model briefly in turn.

Situation and Domain models: Our experimental domains, described below, leverage our earlier research on qualitative process theory [7], qualitative mechanics [19,26], compositional modeling [4], and sketch understanding. We are also drawing upon a subset of the Cyc KB contents, including much of the knowledge associated with the upper ontology and the DARPA-generated materials about military topics and international relations generated in the HPKB, CPOF, and RKF

programs. This knowledge forms the starting endowment for Companions.

Task and Dialogue models: Our task models will be expressed via a planning formalism, probably based on HTN's [4]. For dialogue models, we will use the representations pioneered by Allen's group [1]. This will require developing new models for knowledge capture and joint problem-solving tasks.

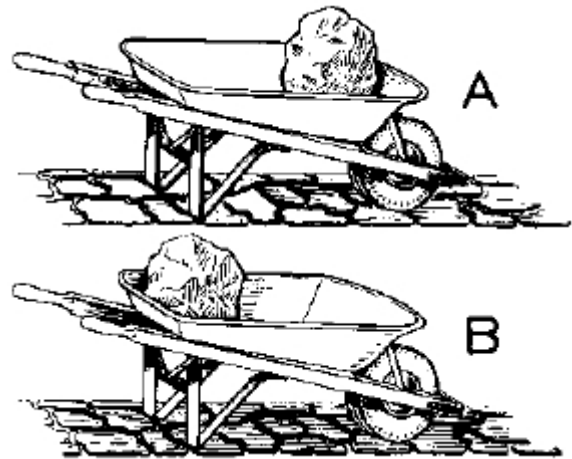
User models: By keeping track of decisions and choices made by the user(s) in the logs, a statistical model of user preferences and conventions will be built up [18]. Some of this information will clearly be about a specific user, e.g., preferred order of tackling sub-problems. However, some of it will be more widely shared conventions (e.g., when drawing a wheelbarrow, the glyph representing the axle of the wheel provides a connection that leaves the wheel free to rotate).

As a user interacts with a Companion, at least one agent will be watching the interaction with standing goals to acquire knowledge about the user's intent, satisfaction with results, and abstract patterns of interaction. It will be fleshing out candidate task and preference models to learn, for example, that when the user is characterizing a situation, they typically want both positive and negative precedents, though perhaps not simultaneously.

Self Models: Our starting point for generating self-models will be the machine-readable logs of user interactions and internal operations, described above. In addition to providing checkpoints/rollback for KB contents and hot-swapping, they will serve as grist for reflection and for learning experiments. Our idea is to construct episodic memories out of summaries of these logs, which could then be consulted for self-predictive purposes. Some kinds of knowledge that it should glean from these logs include expectations about the difficulty of the current task and how long it might take, whether or not it is thrashing (obviously not computable with 100% accuracy), and a context for referring to objects and tasks across sessions. Identifying the precursors to agent failure that can be used to trigger hot-swapping (e.g., memory becoming full) is also important.

Experimental Domains

To ensure generality, we are working from the beginning in several domains. We have used two criteria in selecting these domains: (1) *Sources of complexity*. They must be broad, in contrast with the narrow domains systems operate comfortably in now. (2) *Clear progress metrics*. Reasonably objective methods of measuring performance are important. When an externally validated performance standard is available, that is even better. We discuss each domain in turn.



Domain 1: Everyday physical reasoning

No existing AI system handles the breadth and flexibility of the kinds of reasoning that people do about the everyday physical world around them. While progress in qualitative reasoning has provided a powerful collection of representations to work with [22], the reasoning techniques developed by that community, including ourselves, do not appear to be psychologically plausible [10]. The motto of "structure-mapping all the way down" for Companions encourages an approach that we think will be more plausible: within-domain analogical reasoning.

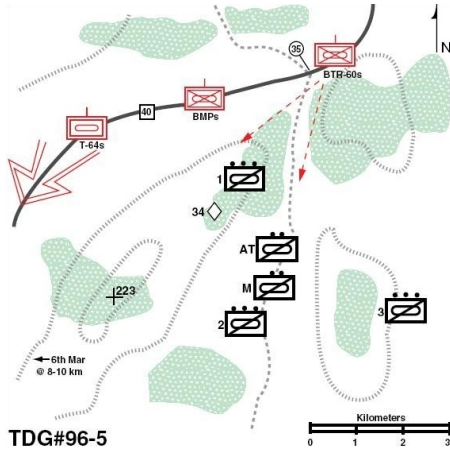
To provide a clear progress metric, we are starting with the kinds of problems found on the Bennett Mechanical Comprehension test. Above is an example of the kind of problem that one sees on this exam: Which way will it be easier to carry the rock in the wheelbarrow? (If equal, mark C). The Bennett Mechanical Comprehension Test has been administered for over 50 years, and is used to evaluate candidates for jobs involving mechanical skills. It is also commonly used by cognitive psychologists as an independent means to measure spatial abilities in experiments. It is an extremely broad test, covering topics such as statics, dynamics, heat flow, and electricity. However, the reasoning is mostly qualitative in nature, using comparative analysis [30]. Each of the two forms of the BMC has 68 questions.

In working with these problems, we will use our *sketching Knowledge Entry Associate* (sKEA) [14] as part of the Companion's interface. We are sketching the pairs of situations in two dimensions, which is sufficient for all but a handful of the problems. A formal language, in terms of everyday relationships, will be used to express the text of the question. The material to be learned includes (a) visual/conceptual mappings and conventions for depicting everyday objects, (b) Modeling assumptions (e.g., how parts of everyday objects map into qualitative mechanics formalisms), (c) Causal models (e.g., that tall things with narrow bases tend to be less stable).

Several kinds of experiments are planned. For example, an analysis of one of the exam forms (the other has been sequestered for future testing purposes) has been carried out to identify the kinds of entities that are involved. Without reference to the questions, knowledge entry sessions will be used to enter basic descriptions of these objects. Then we will build up the Companion's knowledge through problem-solving sessions, with the goal being a suitably high score on the exam. Then the other form of the test will be used as a gauge of breadth. Another kind of an experiment is searching for the minimal set of cases that suffice to give a particular level of performance on the exam via analogical reasoning, with and without generalization.

Domain 2: Tactical Decision Games

Tactical decision games are scenarios used by military personnel to hone their command skills [28]. A scenario is provided in a page or two of text describing the situation,



accompanied by a sketch map illustrating the terrain and what is known about force layouts. These scenarios are quite complex: Typically several answers are equally reasonable, given different assumptions about what the enemy is doing. In keeping with the aide-de-camp nature of Companions, we are not attempting to solve TDGs, but instead to provide several kinds of advice about the scenarios they embody. These include estimating enemy intent (i.e., what are the bad guys up to?), identifying risks (i.e., how might they prevent you from accomplishing your mission?), and identifying opportunities (i.e., how might you go about carrying out your mission?).

We have obtained a corpus of TDGs courtesy of the *USMC Gazette* containing both a set of problems and several solutions to each problem. To encode them, we will include *nuSketch Battlespace* [15], our other sketch understanding system, as a component in the Companion interface. The learning experiments we plan to do are similar to what we plan for everyday physical reasoning: Measure how learned cases and/or generalization improves performance incrementally over a set of problems, how that learned knowledge suffices for a set-aside body of problems, etc.

Current Status

At this writing (5/04) a 10 CPU cluster has been installed to host the first Companion. The agent shells have been implemented, including logging, and distributed reasoning is being debugged. The first-cut interface is being implemented, as well as the starting endowment of representations for the two initial domains. The first round of knowledge acquisition sessions and shakedown of the architecture and interface will start in September, with the first learning experiments being carried out in the fall. Next year we will add a third domain, based on a strategy game such as *FreeCiv*, both to broaden the range of tasks for the system and to stress the system over longer periods of time. Being able to use lessons learned from Tactical Decision Games in other strategy games is an important goal, something that people clearly do, but today's learning systems cannot.

Future Work

The Companions project is only just getting underway, and most of the work lies ahead. Already it is raising a variety of interesting questions, including:

- *Self-awareness*. What contents, level of detail, and organization of logs are needed to support hot-swapping of components and learning better self-models?
- *Encoding*. Some initial encoding of situations is needed to "prime the pump" for analogical retrieval. Learning good encoding strategies seems to be a key learning problem: Psychological results suggest that differences in encoding can account for many novice/expert differences [2], for example.
- *Non-linguistic multimodal communication*. How far can we go with the combination of sketching and concept maps? How do these modalities impact dialogue models?

We hope that by finding answers to these, and many other questions, we can help bring the field closer to achieving human-level AI.

Acknowledgements

This research is supported by DARPA IPTO.

References

1. James Allen, George Ferguson, and Amanda Stent 2001. An architecture for more realistic conversational systems. *Proceedings of IUI-01*, Santa Fe, NM, January 14-17, 2001.
2. Cañas A. J., Ford K. M., Brennan J., Reichherzer T., Hayes P. 1995. Knowledge Construction and Sharing

- in Quorum. *World Conference on Artificial Intelligence in Education*, Washington DC, July.
3. Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts –and novices. *Cognitive Science*, 5, 121-152.
 4. Erol, K., Nau, D., and Hendler, J. 1994. UMCP: A sound and complete planning procedure for hierarchical task-network planning. In *AIPS-94*, Chicago, June, 1994.
 5. Falkenhainer, B. and Forbus, K. 1991. Compositional Modeling: Finding the Right Model for the Job. *Artificial Intelligence*, 51 (1-3).
 6. Falkenhainer, B., Forbus, K., Gentner, D. 1989. The Structure-Mapping Engine: Algorithm and examples. *Artificial Intelligence*, 41(1-63).
 7. Forbus, K. 1984. Qualitative Process Theory *Artificial Intelligence*, 24(85-168)
 8. Forbus, K., Carney, K., Sherin, B. and Ureel, L. To appear. VModel: A visual qualitative modeling environment for middle-school students. To appear in *Proceedings of IAAI-04*, San Jose, July 2004.
 9. Forbus, K., Ferguson, R. and Gentner, D. 1994. Incremental structure-mapping. *Proceedings of CogSci-04*, August.
 10. Forbus, K. and Gentner, D. 1997. Qualitative mental models: Simulations or memories? *Proceedings of QR-97*, Cortona, Italy.
 11. Forbus, K., Gentner, D. and Law, K. 1995. MAC/FAC: A model of Similarity-based Retrieval. *Cognitive Science*, 19(2), April-June, pp 141-205
 12. Forbus, K. D. & Kuehne, S. E. 1998. RoboTA: An Agent Colony Architecture for Supporting Education. *Proceedings of Agents-98*, Minneapolis/St. Paul, MN, pp. 455-456
 13. Forbus, K., Mostek, T. and Ferguson, R. 2002. An analogy ontology for integrating analogical processing and first-principles reasoning. *Proceedings of IAAI-02*, July 2002.
 14. Forbus, K. and Usher, J. 2002. Sketching for knowledge capture: A progress report. *IUI'02*, January 13-16, 2002, San Francisco, California.
 15. Forbus, K. Usher, J. and Chapman, V. 2003. Qualitative spatial reasoning about sketch maps. *Proceedings of IAAI-03*, Acapulco, Mexico.
 16. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7, 155-170.
 17. Gentner, D. and Markman, A. 1997. Structure Mapping in Analogy and Similarity. *American Psychologist*, January, pp 45-56
 18. Horvitz, E. Kadie, C.M., Paek, T. and Hovel, D. 2003. Models of attention in computing and communication: From principles to applications. *CACM* 46(3):52-59, March
 19. Kim, H. (1993). Qualitative reasoning about fluids and mechanics. Ph.D. dissertation and ILS Technical Report, Northwestern University.
 20. Kuehne, S., Gentner, D. and Forbus, K. 2000. Modeling infant learning via symbolic structural alignment. *Proceedings of CogSci-2000*. Philadelphia, PA.
 21. Kuehne, S., Forbus, K., and Gentner, D. 2000. SEQL: Category learning as incremental abstraction using structure mapping. *Proceedings of CogSci-2000*. Philadelphia, PA.
 22. Kuipers, B. 1994. *Qualitative Reasoning: Modeling and simulation with incomplete knowledge*. MIT Press, Cambridge, MA
 23. Labrou, Y. and Finin, T. 1997. A proposal for a new KQML specification. Technical Report TR CS-97-03, University of Maryland Baltimore County.
 24. McDermott, D. 2001. *Mind and Mechanism*. MIT Press, 2001
 25. Mostek, T., Forbus, K. and Meverden, C. 2000. Dynamic case creation and expansion for analogical reasoning. *Proceedings of AAAI-2000*. Austin, Texas.
 26. Nielsen, P.E. (1988). A qualitative approach to rigid body mechanics. (Tech. Rep. No. UIUCDCS-R-88-1469; UILU-ENG-88-1775). Urbana, Illinois: University of Illinois at Urbana-Champaign, Department of Computer Science.
 27. J.D. Novak and D.B. Gowin, *Learning How To Learn*, New York: Cambridge University Press. 1984.
 28. Schmitt, J.F. 1994. *Mastering Tactics: A Tactical Decision Game Workbook*. Marine Corps Association, Quantico, VA.
 29. Thomere, J. and Rodriguez, A. and Chaudhri, V. and Mishra, S. and Eriksen, M. and Clark, P. and Barker, K. and Porter, B. 2002. A Web-based Ontology Browsing and Editing System. *Proceedings of IAAI-02*.
 30. Weld, D. 1990. *Theories of comparative analysis*. MIT Press, Cambridge, MA.