Proceedings of the AAAI-10 Workshop on Visual Representations and Reasoning, 2010

Speculations on Leveraging Graphical Models for Architectural Integration of Visual Representation and Reasoning

Paul S. Rosenbloom

Department of Computer Science and Institute for Creative Technologies
University of Southern California
13274 Fiji Way, Marina del Rey, CA 90292 USA
rosenbloom@usc.edu

Abstract

A new approach to implementing cognitive architectures based on graphical models holds the potential for simpler yet more functional architectures. It also raises the possibility of incorporating visual representation and reasoning into architectures in a manner that is uniformly implemented, and tightly coupled, with both perception and cognition. While much of this is still highly speculative, the core of how it might work is outlined here.

Cognitive architectures are hypotheses about the fixed structure underlying intelligent behavior, whether intended as models of human intelligence and/or implementations of artificial intelligence (Langley, Laird and Rogers 2009). A basic cognitive architecture may comprise memories, decision algorithms, learning mechanisms, and some means of interacting with external environments. More advanced architectures might also have capabilities for things like reflection, theory of mind, motivation and emotion. In a classical architecture, either the external environment with which it interacts is symbolic, making interaction relatively simple, or a continuous perceptuomotor system is required. In the latter case, there is generally a hard wall between the central cognitive system, based on symbol processing, and the peripheral perceptuomotor system, based on signal processing. Results from one side of this wall are thrown over to the other side with an accompanying signal-to-symbol or symbol-to-signal transformation. Sophisticated visual representation and reasoning (VRR) is not common in architectures, although some recent versions do incorporate forms of 2D (Lathrop and Laird, 2009; Cassimatis et al. In press) or 3D (Wintermute and Laird 2007) imagery.

If we were to ask in general how VRR should be integrated into a cognitive architecture, a complete answer would require responses at both the knowledge level and the symbol level (Newell 1982). The knowledge level response must consider what is represented in VRR and how that relates to, and semantically interconverts with, what is represented in any other architectural modules with which it must interact, such as perception, motor control, and central cognition. The symbol level response must consider the structure of these interacting modules and how they can interoperate. Together the two responses effectively determine *what* and *how*.

The speculations here focus only on the symbol level response – i.e., the *how* – starting with the notion of combining symbolic and perceptuomotor information through an intermediary of 3D spatial imagery and reasoning. The starting point is an ongoing effort to reconstruct cognitive architectures from the ground up via graphical models (Koller and Friedman 2009), with the aim of understanding existing architectures better, exploring the overall space of architectures, and developing new and improved architectures (Rosenbloom 2009).

Graphical models provide efficient computation with complex multivariate functions by decomposing them into products of simpler subfunctions. The most familiar class of graphical models within AI and cognitive science is Bayesian networks (Pearl, 1988), which decompose complex joint probability distributions into products of simpler conditional and prior distributions, and then embody this product as a directed graph in which nodes represent random variables, arcs represent dependencies among random variables, and conditional probability tables at nodes represent the distributions of their random variables given the random variables upon which they depend. Markov networks, aka Markov random fields, extend this to undirected graphs that are accompanied by arbitrary functions, called potentials, defined over variable cliques. Factor graphs go one step further, incorporating the potentials from Markov networks into the graphs themselves as functions that are specified by an additional class of factor nodes (Kschischang, Frey and Loeliger 2001). Factor graphs were originally developed in coding theory – where they underlie the “surprisingly effective” performance of turbo codes – to support decomposition of arbitrary multivariate functions.

The intriguing thing about graphical models, and particularly the latter forms of them, for cognitive architecture is how they yield state-of-the-art algorithms across symbol, probability and signal processing from a uniform graphical representation and a single generic message passing algorithm (*summary product*). This includes algorithms for constraint processing and production match (symbols), loopy belief propagation in Bayesian networks (probabilities), and Kalman filters and the forward backward algorithm for hidden Markov models (signals). Graphical models in general have become the standard approach to probabilistic reasoning and perception problems. Their penetration into symbol processing isn't nearly as significant, but they do show potential, with such languages as Alchemy (Domingos et al. 2006) and BLOG (Milch et al. 2007) providing broadly applicable mixtures of statistical and logical/relational processing. Leveraging the broad-yet-uniform state-of-the-art functionality provided by graphical models holds out significant potential for architectures that are simpler and more uniform yet significantly more functional than today's best.

Progress to date in exploring this potential has focused on the implementation, via factor graphs, of a hybrid (discrete and continuous) mixed (Boolean and Bayesian) memory architecture that provides the kinds of memories embodied in two leading cognitive architectures – ACT-R (Anderson 2007) and Soar 9 (Laird 2008) – while also going beyond them in significant ways. This memory architecture defines a working memory along with multiple long-term memories: classical procedural (rule) and declarative (semantic and episodic) memories plus a constraint memory. The long-term memory structures all compile into graphs; albeit with some variations in their details, such as in the directionality of message passing used by the summary product algorithm – rules use unidirectional message passing while declarative memories use bidirectional message passing. Working memory maps onto *evidence* in graphical models, and as such is encoded as a fixed function within one or more factor nodes.

In these graphs, factor functions, along with the inter-node messages that drive the summary product algorithm, are represented as N dimensional continuous functions, which are in turn approximated as piecewise linear functions over rectilinear regions. The domains of these continuous functions can be discretized to represent probability distributions, and the ranges of the resulting discrete functions can further be Booleanized to represent symbols. The underlying representation remains continuous, enabling all three forms of knowledge to be processed uniformly by the summary product algorithm. But at the same time it provides a broad-spectrum representation, capable of encoding continuous signals and discrete symbols, as well as points intermediate between them. This approach bears a resemblance to Barsalou's (1999) proposal for *perceptual symbol systems*, in which symbols are attentionally extracted patterns of perceptual activity, although it does not presently enforce his taboo on amodal symbols; i.e., symbols abstracted away from any perceptual context.

This memory architecture still falls far short of the full capabilities required in a cognitive architecture – missing, for example, decision making, reflection, perception, motor control, motivations and emotions – but it does point in an interesting direction. The core hypothesis explored in the remainder of this paper is that the layers of the full behavioral hierarchy to be supported by such an architecture – from perception and motor control at the bottom, up through 3D spatial imagery, to symbolic reasoning at the top – along with the bidirectional flow of information among these layers, are uniformly implementable and integratable via these kinds of graphical models.

Let's start with perception. Vision involves processing of 2D arrays of numerically valued pixels. State-of-the-art techniques use Markov random fields and their close cousin conditional random fields. Speech processing occurs via hidden Markov models. All three of these techniques are variants of graphical models. The main speculative aspect here concerns whether comparable capability can be had from the kind of factor graph just described. For example, will a piecewise linear approximation be sufficient to handle these signals with their embedded noise, or will it be necessary to directly implement Gaussian functions? Also, will summary product be sufficiently efficient, or will some other message-passing (or sampling) algorithm be required that is more optimized for such data?

At the other end of the hierarchy, graphical models have not been so well explored for symbol processing. There has been significant work in the context of constraint processing, including work on mixed hybrid variants (Gogate and Dechter 2005), along with the work previously mentioned on mixed languages. My own work has also shown rule match in factor graphs that has better worst-case complexity than the state-of-the-art Rete algorithm (Forgy 1982). Thus, although symbol processing in graphical models is not an extremely well explored area overall, the potential is there to justify a speculation that graphical models of the kind considered here will be sufficient for this aspect.

The most speculative aspect, because it is the most novel, is basing the intermediate layer – a 3D spatial memory with associated reasoning processes – on graphical models. Initial optimism stems from the fact that piecewise linear functions already support a superset of the functionality of a *voxel* approach to 3D representation. But a range of other alternatives are also possible, such as generalizing from rectilinear regions to ones based on convex polytopes (N dimensional polygons) or replacing linear functions with Gaussians to better handle noise and uncertainty. Assuming any of these representational approaches proves adequate, the remaining question at this layer is whether the necessary operations on such a memory – such as object addition, deletion, translation and rotation – are feasible through standard graphical processing methods.

Initial ideas for solving this problem may be found in related areas such as sequence predication and stereo vision. Sequence prediction hypothesizes what comes next given what has already occurred, with Markov models supplying standard techniques. Stereo vision computes a disparity map between two 2D images, with Markov and conditional random fields providing typical tools. Our task starts with a 3D image plus a set of transformations – serving a role that is analogous to the disparity map in the stereo problem – and must compute a second 3D image. A good place to start will thus be to look at fusing the insights from these two related problems.

Beyond implementing the three layers of the hierarchy, these layers must also be integrated together so as to enable straightforward interoperation. Experience implementing the memory architecture mentioned earlier showed that subtle incompatibilities among multiple capabilities can arise when attempting to implement them in a uniform substrate – e.g., conflicts between closed world and open world assumptions in procedural and declarative memories – but that otherwise the uniform implementation substrate acts as an excellent integration medium across the capabilities. So such interoperability does appear to be within the realm of feasibility for our three hierarchy layers.

Should this general approach and its accompanying speculations pan out, it will yield a cognitive architecture embodying a uniformly implemented and tightly integrated capability for VRR. Such a capability would also hopefully yield improvements in overall VRR capabilities, from either the breadth of functionality provided by graphical models or the tight integration with cognition and perception. However, how this might actually occur is a speculation on speculations that is best left to later, when the present speculations have become reality. Also left for later are speculations on the across-layer semantic interconversion issues that can only be addressed by a knowledge level response to the original question.

Acknowledgements. This effort has been sponsored by the USC Institute for Creative Technologies and the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

Anderson, J. R. 2007. *How Can the Human Mind Occur in the Physical Universe?* Oxford University Press.

Barsalou, L. W. 1999. Perceptual Symbol Systems. *Behavioral and Brain Sciences* 22: 577–660.

Cassimatis, N., Bignoli. P., Bugajska, M., Dugas, S., Kurup, U., Murugesan, A. and Bello, P. In press. An architecture for adaptive algorithmic hybrids. *IEEE Transactions on Systems, Man, and Cybernetics*.

Domingos, P., Kok, S., Poon, H., Richardson, M. and Singla, P. 2006. Unifying logical and statistical AI. In *Proceedings of the 21st National Conference on Artificial Intelligence*, 2-7. AAAI Press.

Forgy, C. L. 1982. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence* 19(1): 17-37.

Gogate, V. and Dechter, R. 2005. Approximate inference algorithms for hybrid Bayesian networks with discrete constraints. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, 209-216.

Koller, D. and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press.

Kschischang, F. R., Frey, B. J., and Loeliger, H. 2001. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory* 47: 498-519.

Laird, J. E. 2008. Extending the Soar cognitive architecture. In *Artificial General Intelligence 2008: Proceedings of the First AGI Conference*. IOS Press.

Langley, P., Laird, J. E., & Rogers, S. 2009. Cognitive architectures: Research issues and challenges. *Cognitive Systems Research* 10: 141-160.

Lathrop. S. D. and Laird, J. E. 2009. Extending Cognitive Architectures with Mental Imagery. In *Proceedings of the 2nd Conference on Artificial General Intelligence*.

Milch, B. Marthi, B., Russell, S. Sontag, D., Ong, D. L., Kolobov, A. 2007. BLOG: Probabilistic models with unknown objects, in L. Getoor, B. Taskar (Eds.), *Introduction to Statistical Relational Learning*, Cambridge, MA: MIT Press.

Newell, A. 1982. The Knowledge Level. *Artificial Intelligence*, 18(1): 87-127.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufman.

Rosenbloom, P. S. 2009. Towards a new cognitive hourglass: Uniform implementation of cognitive architecture via factor graphs. In *Proceedings of the 9th International Conference on Cognitive Modeling*.

Wintermute, S. and Laird, J.E. 2007. Predicate Projection in a Bimodal Spatial Reasoning System. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*.