The Marchitecture: A Cognitive Architecture for a Robot Baby

Marc Pickett I and **Tim Oates**
Cognition Robotics and Learning
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250
marc@coral-lab.org, 410-455-8790
(See also `http://www.coral-lab.org/~marc/pickettProposal.pdf`)

Abstract

The Marchitecture is a cognitive architecture for autonomous development of representations. The goals of The Marchitecture are domain independence, operating in the absence of knowledge engineering, learning an ontology of parameterized relational concepts, and elegance of design. To this end, The Marchitecture integrates classification, parsing, reasoning, and explanation. The Marchitecture assumes an ample amount of raw data to develop its representations, and it is therefore appropriate for long lived agents.

Introduction

Traditional approaches to Artificial Intelligence focus on selecting an application and then constructing representations for that domain. These approaches are problematic in that they require much labor intensive knowledge engineering. Furthermore, these systems tend to be brittle, often failing when they encounter unanticipated situations. An alternate approach is to have the computer develop its representations autonomously. In this alternate approach, the robot is viewed as a "robot baby" (Cohen *et al.* 2002). The robot is provided a minimal amount of knowledge (implicit or otherwise) about the world and is expected to learn and develop a conceptual structure from large amounts of raw sensor data over a long period of time. This approach is attractive because it requires little knowledge engineering and is robust because the agent learns to adapt to unanticipated situations.

If such an agent is to acquire human level intelligence, it will need to be able to represent and learn *relational* concepts (e.g., "cousin", "above", or "enemy"). Development of a cognitive architecture is necessary for the solution to the problem of Artificial Intelligence. Many cognitive architectures have been proposed (Sun 2004), but, to our knowledge, none focus on domain independent autonomous development of representations from raw relational data.

Related Work

Several cognitive architectures have been proposed in the past. Earlier examples include SOAR (Rosenbloom, Laird, & Newell 1993), and ACT-R (Anderson 1983). For overviews of these and other architectures see (Sun 2004) and (Langley & Laird 2002). Most of the architectures discussed in these overviews focus on action and planning given a human-provided domain model. They contain learning as an afterthought, if at all, and none focus on the development of representations. These architectures often assume that they start with a cache of domain specific, human engineered representations. However, The Marchitecture will be able to use ideas from these papers for its planning and reasoning.

The Marchitecture uses a top-down/bottom-up approach for reasoning, planning, data segmentation, explanation, and classification. This approach is an elaboration on ideas developed in a model of the human neocortex (Hawkins & Blakeslee 2004). This earlier model, however, is unable to handle relational data. For concept formation, The Marchitecture uses ideas from our earlier work (Pickett & Oates 2005), which develops a conceptual structure from *non-relational* data. The SUBDUE algorithm (Holder, Cook, & Djoko 1994) develops a conceptual structure from relational data, but is not a full cognitive architecture because it includes neither planning, reasoning, or explanation, and has only rudimentary methods for data segmentation and classification, which fail to include the top-down/bottom-up approach of (Hawkins & Blakeslee 2004).

The Marchitecture

An overview of The Marchitecture is shown in Figure 1. It has a feed of raw, unsegmented "sensor" data, though the data can be from any discretizable domain. The only constraint is that the data must be represented in our relational representation framework, which is what The Marchitecture uses to describe both its raw data and its entire conceptual structure. Since The Marchitecture uses the same framework for representing both raw data and higher level abstractions, it can use the same algorithm to develop abstractions of abstractions, thus forming a concept heterarchy (which, in our case, is a concept hierarchy except that a concept can have multiple parent nodes). Furthermore, the framework allows for parameterized concepts and can thus make grammatical constructs such as a "blue pen" from the concepts of "blue" and "pen". A full description of this representation framework is given in the URL under the authors' address.

The **Main Loop** then calls the **Parsing/Explanation** module, which tries to segment, classify, and explain the

Copyright © 2007, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

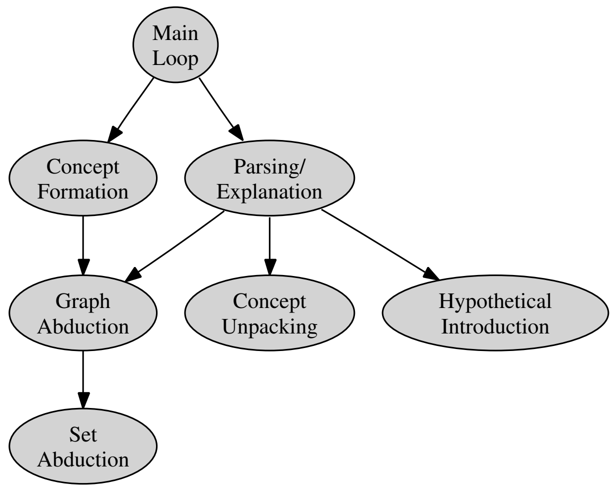

Figure 1: **Overview of The Marchitecture**

new data using concepts that it has already formed. Failing that, the **Main Loop** uses the **Concept Formation** module to find analogies in the data and use them to form new concepts with which to recharacterize the data. This process is akin to the assimilation/accommodation process described by developmental psychologists (Piaget 1954).

The Marchitecture forms concepts by finding large and/or frequent subgraphs. We refer to this process as analogy discovery. Although the subgraph isomorphism problem is intractable in the worst case, the most useful concepts are the most common and therefore the most likely to be discovered. The Marchitecture employs a number of tricks to find frequent subgraphs, such as an approximate canonical form. In practice, finding subgraph isomorphisms is usually feasible (McKay 1981). These tricks are encapsulated in the **Graph Abduction** and **Set Abduction** modules, which find supergraphs of a given subgraph. We use Minimum Description Length as a measure for which subgraphs should be turned into concepts. That is, we keep the concepts that allow us to most concisely characterize the data. It has been argued that this may be one of the core purposes of concepts (Wolff 2003).

The **Parsing/Explanation** module uses a top-down/bottom-up algorithm similar to that described by (Hawkins & Blakeslee 2004). This module uses the **Abduction** module to find concepts that are supergraphs of a given set of data. Once a set of concepts are proposed, the **Parsing/Explanation** module uses the **Hypothetical Introduction** module to search different "parses" or segmentations of the data, choosing the parsing that results in the shortest description length. **Hypothetical Introduction** also methodically posits unbound parameters for concepts proposed by **Abduction**. Concepts can be "unpacked" to perform reasoning. That is, a concept can be expanded and the resulting statements are entailments of that concept. The **Parsing/Explanation** can thus explain data either by classification or by explaining the data using a series of concept unpacking (which amounts to forward chaining). Combinations of graph abduction, concept unpacking, and hypothetical introduction can be used for prediction, planning, reasoning, and explanation.

There are some open issues with the current design: Minimum Description Length might not be the best metric for a model. For example, a short model of Euclidean Geometry would simply be the 5 postulates and a set of derivation rules. This model would be complete, but it might be better if some useful lemmas were cached. Thus, sometimes a faster model is preferable to a smaller model.

Conclusion

The Marchitecture tightly integrates several aspects of cognition. The strength of The Marchitecture lies in its simplicity and in its focus on development of representations. So far, we have implemented the **Concept Formation** module of The Marchitecture (described in detail in the URL below the authors' address), **Set Abduction**, and **Graph Abduction**, and tested these on a variety of domains (also described in that paper). To guard against domain dependence, we have a set of disparate domains on which to test The Marchitecture. Our goal is to apply our algorithm to RISK, Conway's Life, robot sonar data, and a traffic simulation domain.

References

Anderson, J. R. 1983. *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.

Cohen, P.; Oates, T.; Beal, C.; and Adams, N. 2002. Contentful mental states for robot baby. In *Proceedings of the 18th National Conference on Artificial Intelligence*.

Hawkins, J., and Blakeslee, S. 2004. *On Intelligence*. Times Books.

Holder, L.; Cook, D.; and Djoko, S. 1994. Substructure discovery in the subdue system. In *Proceedings of the Workshop on Knowledge Discovery in Databases*.

Langley, P., and Laird, J. 2002. *Cognitive architectures: research issues and challenges*. Manuscript.

McKay, B. 1981. Practical graph isomorphism. *Congressus Numerantium* 30:45–87.

Piaget, J. 1954. *The Construction of Reality in the Child*. Basic Books.

Pickett, M., and Oates, T. 2005. The cruncher: Automatic concept formation using minimum description length. In *proceedings of the 6th International Symposium on Abstraction, Reformulation and Approximation (SARA 2005)*, Lecture Notes in Artificial Intelligence. Springer Verlag.

Rosenbloom, P.; Laird, J.; and Newell, A. 1993. *The SOAR papers: research on integrated intelligence*. Cambridge, MA: MIT Press.

Sun, R. 2004. Desiderata for cognitive architectures. *Philosophical Psychology* 17(3):341–373.

Wolff, J. G. 2003. Information compression by multiple alignment, unification and search as a unifying principle in computing and cognition. *Artif. Intell. Rev.* 19(3):193–230.