Representation Change in The Marchitecture

Marc Pickett I and **Don Miner**
Cognition, Robotics, and Learning
Department of Computer Science and Electrical Engineering
University of Maryland, Baltimore County
1000 Hilltop Circle, Baltimore, MD 21250

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. Representation change comes about through a pair of primary mechanisms, which we term "assimilation" and "accommodation". Accommodation is the formation of concepts, which builds an ontology, and assimilation amounts to explaining or recasting data in terms of the developed ontology. This model is inspired by Piaget's Constructivist theory of learning. 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, which we call an *ontology*, 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.

We're developing a cognitive architecture, called The Marchitecture, which follows this alternate approach, and focuses on developing representations. The Marchitecture relies on as little human involvement as possible, and is therefore unsupervised. Since the goal isn't as straightforward as it would be for supervised learning, we need a criterion for evaluating whether a representation change is useful. It has been suggested that a core purpose of (unsupervised) intelligence is to concisely characterize a set of data (Wolff 2003), (Hutter 2004). That is, given data, an intelligent agent should generate representations that best compress the data. This is the principle of Minimum Description Length (MDL). It is fundamentally equivalent to Ockham's Razor, which says, in effect, that "The shortest model (that predicts the data) is the best model.". If we assume that the prior probability of a model is inversely proportional to the exponent of its description length, then Ockham's Razor is also fundamentally equivalent to the Bayesian principle that states that "The most probable model is the best model.".

Ockham's Razor and Marctar's Axe

We somewhat agree with these claims. An intelligent agent should be able to build a model that concisely characterizes its sensor data, and it should be able to use this model to answer queries about the data. Such queries might consist of making accurate predictions about given situations. The agent should also be able to generate plans to accomplish goals (or obtain reward). However, the time needed (in terms of steps of computation) to answer these queries should also be taken into account. Thus, it is sometimes useful to occasionally trade memory for time. For example, an intelligent being might cache a result that it has deduced if it expects to use the result again.

To make this concrete, suppose our agent's domain is Euclidean Geometry. In this domain, a huge but finite set of theorems of can be "compressed" down to a model containing just 5 postulates and some rules for inference. Such a model would neither be very useful nor would it work the same way as a person. A professional (human) geometer would likely "cache" useful lemmas thereby speeding up his or her future deductions. It seems true that the same should apply to a generally intelligent being.

Thus, we propose an alternative to Ockham's Razor called Marctar's Axe, which states "The quickest model (that predicts the data) is the best model.". By quickest, we mean the model that takes the fewest steps of computation to get accurate answers to queries. Of course, there's a tradeoff between speed and accuracy, but this can be folded into a single number by setting a parameter that would act as an "exchange rate" between steps of computation and bits of accuracy. Marctar's Axe somewhat overlaps with Ockham's Razor in that fast models tend to be small and tidy so that computation isn't spent searching through disorganized sets of information. Marctar's Axe also addresses the utility of caching: caching the answers to frequent queries (or frequent "way points" in derivations) can yield a faster model.

Relational Representations

If The Marchitecture is to acquire human level intelligence, it will need to be able to represent and learn *relational* concepts (e.g., "cousin", "above", or "enemy"). Relational concepts are a form of invariant representations since they can have ground instances that vary widely. We suspect that such representations are necessary for solving a wide range of problems. For example, when a person dons a pair of green-tinted sunglasses for the first time, they have little trouble adapting to their altered visual input, but this isn't such a trivial task for a (visual) robot. In terms of raw sensor data, a green-tinted scene has very different values from the same scene in its natural color. Representations developed from visual data should also be invariant to translation, rotation, and scaling. These *invariant representations* aren't limited to visual data. A stenographer can hear different speakers say the same phrase in different pitches, volumes, and speeds, yet produce the same transcription.

Piaget's Constructivist Theory of Learning

Whether in terms of Ockham's Razor or Marctar's Axe, the goal of The Marchitecture is to characterize its input data, and forming concepts and building an ontology are means to this end. However, for concepts to be helpful, the data needs to be expressed in terms of these concepts. For example, having developed the concept of "a duck", we should be able to *recognize* an instance of a duck and classify it as such. Doing this prevents us from having to redescribe all the attributes that a duck has every time we see one, so this "shorthand" process is useful for minimizing description length. It's also useful for quickly answering queries because, for example, knowing that ducks lay eggs, we can quickly answer whether a particular duck lays eggs.

Inspired by Piaget's Constructivist theory of learning (Piaget 1954), we call the process of developing concepts "accommodation", and the process of using these concepts to classify or explain data "assimilation". The Marchitecture further follows Piaget's theory in that, when encountering new data, it first attempts to explain the data using the current ontology. If it's unsuccessful, it sets the unexplained data aside and periodically looks for patterns in its set of unexplained data out of which it can form new concepts. When a new concept is formed, The Marchitecture attempts to use this to explain the previously unexplained data.

Some suggest that analogy may be the "core of cognition" (Hofstadter 2001). Analogy allows us to focus on the relations among entities rather than superficial aspects of the entities. For example, we might notice that a red ant killing a black ant and stealing a piece of food it is analogous to a situation in Hamlet where Claudius murders Hamlet's father and usurps the throne of Denmark. In this situation *binding* is important. That is, we must be able to specify that the red ant corresponds to Claudius, the black ant to Hamlet's father, and the piece of food maps to the throne. Analogy is also useful for *knowledge transfer*: if an analogy is found, then conclusions about one domain can map to another domain.

The Marchitecture uses *analogy discovery* to form its concepts in a manner similar to the SUBDUE system (Holder, Cook, & Djoko 1994). The Marchitecture represents its data in a relational format and searches for frequently occurring subgraphs, which become candidates for becoming concepts (depending on its potential to reduce description length or query time). When a subgraph is selected to become a concept, it becomes *parameterized*. For example, the concept of "stealing" might have the parameters "thief", "victim" and "stolen".

These parameterized concepts allow us to unify the processes of classification and explanation (which we call assimilation) because we can these concepts as rules similar to those used in logic programming. The Marchitecture uses a top-down/bottom-up process of explanation which is inspired by the model of the neocortex described by (Hawkins & Blakeslee 2004).

Outlook

By focusing on development of representations, we believe we've addressed problems with acquiring knowledge, and that we've developed the basis for a full cognitive architecture. Using this basis, The Marchitecture tightly integrates other aspects of cognition not discussed here. For example, meta-cognition, planning, and reasoning. The strength of The Marchitecture lies in its simplicity and in its focus on development of representations. 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

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.

Hofstadter, D. R. 2001. Analogy as the core of cognition. *The Analogical Mind: Perspectives from Cognitive Science* 499–538.

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

Hutter, M. 2004. *Universal Artificial Intelligence: Sequential Decisions based on Algorithmic Probability*. Berlin: Springer.

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

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.