Tutorial: The CLARION Cognitive Architecture

Nicholas Wilson (wilson3@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute,
110 Eighth Street, Troy, NY 12180 USA
Telephone: (518) 276-2692 | Fax: (518) 276-3017

Michael Lynch (lynchm2@rpi.edu)
Department of Language, Literature and Communication
Ron Sun (rsun@rpi.edu)
Cognitive Science Department
Rensselaer Polytechnic Institute, Troy, NY 12180 USA

The half-day tutorial introduces participants to the CLARION cognitive architecture and presents a detailed description, as well as simulation examples, advanced topics, and demonstrations. It will combine conceptual (psychological), theoretical, and implementation aspects of the architecture. Participants should have some prior exposure to cognitive architectures and artificial neural networks. Preferably, participants should also have some experience with programming languages (in particular Java). However, prior understanding of these areas can be limited, as both basic and advanced topics related to cognitive modeling using CLARION will be covered.

Tutorial Outline

A General Overview of CLARION (15 min.)

In this section, an introduction to cognitive architectures in general, and CLARION in particular, will be presented. CLARION will be compared to various other architectures and a brief discussion of some past and current applications of CLARION will be presented along with cognitive justifications and implications.

CLARION is a unified, comprehensive theory of the mind based on two basic theoretical assumptions: representational differences and learning differences of two different types of knowledge --- implicit vs. explicit (Sun, Merrill, & Peterson, 2001; Sun, Slusarz, & Terry, 2005), among other essential assumptions/hypotheses (Sun, 2003).

The first assumption, the representational difference between these two types of knowledge, relates to accessibility. In each subsystem of CLARION, the top level contains easily accessible explicit knowledge whereas the bottom level contains less accessible implicit knowledge.

The second assumption of CLARION concerns the different learning processes in the top and bottom levels of each subsystem (Sun et al., 2001, 2005). In the bottom level, implicit associations are learned through gradual trial-and-error learning. In contrast, learning of explicit knowledge is one-shot and captures its abrupt availability. The emphasis on bottom-up learning (i.e., the transformation of implicit knowledge into explicit knowledge) is, in part, what distinguishes CLARION from other cognitive architectures (although top-down learning is also a capability of CLARION).

In addition to the aforementioned theoretical assumptions, CLARION is a cognitive architecture composed of four main subsystems: the Action-Centered Subsystem, the Non-Action-Centered Subsystem, the Motivational Subsystem, and the Meta-Cognitive Subsystem.

The Action-Centered Subsystem (60 min.)

In this section, the Action-Centered Subsystem (ACS) will be defined in detail. The structure and design of the various aspects of the ACS, along with the learning mechanisms and the properties of the model, will be presented. Finally, a series of simulation examples related to the operations within the ACS will be presented.

The Action-Centered Subsystem is used mainly for action decision-making. In the ACS, the top level generally contains simple "State → Action" rules, while the bottom level uses multi-layer perceptrons to associate states and actions. Reinforcement learning algorithms (usually with back-propagation) are used in the bottom level while rule learning in the top level is mostly "one-shot" and can be performed bottom-up (via "explicitation") or independently (e.g., through linguistic acquisition).

The ACS has been used to model anything from navigation in minefields (Sun et al., 2001) to Towers of Hanoi, etc. In addition, because CLARION focuses on the dichotomy between explicit and implicit knowledge, benchmark psychological tasks used to demonstrate implicit learning have also been successfully modeled and explained (Sun et al., 2005).

The Non-Action-Centered Subsystem (45 min.)

Similar to the section on the ACS, this section will detail the Non-Action-Centered Subsystem (NACS). The structure and design of the various aspects of the NACS, along with the learning mechanisms and the theorems describing the properties of the model, will be presented. In addition, as with the section on the ACS, a series of simulation examples demonstrating the operations within the NACS will be presented.

The Non-Action-Centered Subsystem is used to store declarative ("semantic" and episodic) knowledge and is responsible for reasoning in CLARION. In the NACS, the top level contains simple associations while the bottom level involves a nonlinear neural network. Associative learning algorithms (e.g., backpropagation or contrastive Hebbian) are generally used in the bottom level whereas associations in the top level are mostly learned "one-shot" (similar to the ACS).

The NACS has mostly been used to simulate memory and reasoning. In particular, CLARION was able to capture the effect of mixed rule-based and similarity-based reasoning (e.g., when judging the likelihood of simple deductive forms). In addition, other reasoning phenomena (e.g., inheritance-based reasoning, reasoning from incomplete information, etc) have also been explained using CLARION (e.g., Sun & Zhang, 2006).

The Motivational and Meta-Cognitive Subsystems (30 min.)

In the fourth section, the structure and design of the motivational (MS) and meta-cognitive (MCS) subsystems will be explored in detail. In addition, several past and current simulation examples related to the operations within the MS and the MCS will be presented.

The Motivational Subsystem contains both low-level (physiological) and high-level (social) primary drives that take into account both environmental and internal factors in determining drive strengths. These drive strengths are reported to the Meta-Cognitive Subsystem, which regulates not only goal structures but also other cognitive processes as well (e.g., monitoring, parameter setting, etc). For more details on motivation and meta-cognition see Sun (2003, 2007, 2009).

Simulations using these subsystems, for example, have shown how anxiety-inducing drives can affect the parameters within the ACS in terms of explicit vs. implicit response weighting and overall performance (Wilson et al., 2009). Other simulations have addressed the combination of drives in the MS toward the setting of goals by the MCS. On this basis, models of human personality have been developed.

Introduction to the CLARION Library (30 minutes)

The CLARION implementation (in Java) has recently undergone a number of improvements and enhancements allowing for the simulating of a wide variety of tasks, as well as interfacing with a variety of virtual environments. In the last section of the tutorial, an overview of the CLARION Library will be presented. Participants will be given copies of the newest release of the library and will be shown how it can be used to run new and existing simulations.

Relevance for Cognitive Science

The CLARION cognitive architecture is well established and has been the subject of more than 100 scientific papers and several books. CLARION is particularly relevant to cognitive scientists because of its strong psychological plausibility and the breadth of its application to cognitive modeling and simulation. In CLARION, each structure corresponds to a psychological process/capacity. CLARION-based models have been used to explain data as diverse as implicit learning, cognitive skill acquisition, inductive and deductive reasoning, meta-cognition, motivation, personality, and social simulations (Sun, 2006).

Presentation Details & History

Descriptions and demonstrations during the presentation will be provided using PowerPoint and the Eclipse Java development environment.

Participants in the tutorial are encouraged to ask questions throughout the presentation to clarify any ideas described. The presenters are versed in both the conceptual and implementation details of the CLARION cognitive architecture.

An older variation of the proposed tutorial had been presented at the 30th Annual Meeting of the Cognitive Science Society in Washington D.C. as well as the 2009 International Joint Conference on Neural Networks in Atlanta, GA. In addition, this tutorial has been given as a lecture series on several occasions for various courses in Cognitive Science at Rensselaer Polytechnic Institute.

Sample Materials

- A complete technical specification of CLARION: http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf
- A list of CLARION-related publications: http://www.cogsci.rpi.edu/~rsun/clarion-pub.html
- Current versions of the CLARION Library, slides, etc.: http://www.cogsci.rpi.edu/~rsun/clarion.html

References

Sun, R. (Ed.). (2006). *Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation*. Cambridge, UK: Cambridge University Press.

Sun, R. (2003). A Tutorial on CLARION. Technical report, Cognitive Science Dept., Rensselaer Polytech. Institute. http://www.cogsci.rpi.edu/~rsun/sun.tutorial.pdf

Sun, R. (2007). Motivation and metacognitive control of CLARION. In: W. Gray (ed.). *Modeling Integrated Cognitive Systems*. New York, NY: Oxford University Press.

Sun, R. (2009). Motivational representations within a computational cognitive architecture. *Cognitive Computation*, 1, 91-103.

Sun, R., Merrill, E., & Peterson, T. (2001). From implicit skills to explicit knowledge: A bottom-up model of skill learning. *Cognitive Science, 25*, 203-244.

Sun, R., Slusarz, P., & Terry, C. (2005). The interaction of the explicit and the implicit in skill learning: A dual-process approach. *Psychological Review, 112*, 159-192.

Sun, R. & Zhang, X. (2006). Accounting for a variety of reasoning data within a cognitive architecture. *Journal of Experimental and Theoretical Artificial Intelligence*, *18*, 169-191.

Wilson, N., Sun, R., & Mathews, R. (2009). A Motivationally-based Simulation of Performance Degradation Under Pressure. *Neural Networks*, 22, 502-508.