Modularity and Plasticity are Compatible

Robert A. Jacobs
Department of Brain and Cognitive Sciences
University of Rochester
Rochester, NY 14627
robbie@bcs.rochester.edu

Introduction

Historically, theorists that have highlighted the importance of modular properties of cognitive processing have tended to suggest that the functions, representations, and procedures of cognitive modules are innately specified (e.g., Fodor, 1983). By doing so, these researchers de-emphasize the importance of cognitive development and experience-dependent adaptation. The converse situation also appears to hold; theorists that have stressed the ubiquity of experience-dependent adaptation have tended to ignore or minimize the importance of modular aspects of cognition (e.g., Piaget, 1955). Thus, modularity and experience-dependent plasticity are often seen as incompatible.

Recently, the view that modularity and plasticity are incompatible has been questioned [e.g., Karmiloff-Smith (1992)]. In particular, several researchers have recently proposed computational systems that are modular learning devices. To understand the benefits of incorporating modularity into a learner, it is useful to distinguish between *divergent* computation and *convergent* computation (Jordan and Jacobs, 1992). Divergent computation involves taking data from a single source and performing different computations on it. This is useful whenever an animal has multiple goals and must utilize the data differently depending on the goal. For a learner, it is often advantageous to use different modules to learn different computations because adaptations that occur when learning to satisfy some goal are decoupled from adaptations that are needed to reach other (presumably different) goals. Convergent computation involves taking data in different channels or formats (such as different sensory modalities) and integrating them into a common channel or format. A learner attempting to discover the structure of its environment may benefit from correlating the outputs of distinct sensory modules that each process data from a different modality. In this way, structure can be found that is not present (or not easily detected) in the data from a single modality.

This abstract briefly reviews two recently proposed modular learning devices, one based on divergent computation and the other on convergent computation.

Mixtures-of-Experts Architecture

The first device is referred to as a "mixtures-of-experts" (ME) architecture and it was originally proposed by Jacobs, Jordan, Nowlan, and Hinton (1991). The architecture is intended as an instantiation of the idea that competition can lead to functional specialization. Analogous to Darwinian evolutionary processes, modules of the architecture compete for the right to learn to perform a set of tasks. Due to the competition, modules specialize; that is, modules that are initially functionally undifferentiated learn over time to perform different tasks. More specifically, an ME architecture is a modular system that learns task decompositions in the sense that it uses different connectionist networks to learn input-output training patterns from different regions of the input space (i.e. the space of all possible inputs). As a result of the competition, different networks learn different training patterns and, thus, learn to compute different functions. The architecture consists of two types of networks: *expert networks* and a *gating* network (see Figure 1). The expert networks compete to learn the training patterns. For each training pattern, feedback information is distributed to the experts on the basis of their relative performance; a network whose response most closely matches the desired response (i.e. the winner of the competition) receives lots of feedback information whereas other networks receive no or little feedback. The gating network weights the outputs of the experts so that, for each input pattern, the expert that is most likely to produce the correct response is weighted more heavily than the other experts.

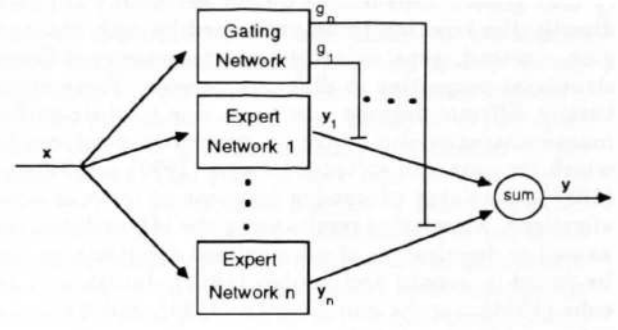

Figure 1: Mixtures-of-Experts Architecture

An interesting feature of the mixtures-of-experts framework is the roles it assigns to nature and nurture in the acquisition of functional specializations. The ME architecture tends to allocate to each task an expert network whose structure is well-matched to that task.

Structural properties of a network, such as its topology, receptive field characteristics, or pattern of connectivity, bias a network so as to make it a particularly good learner for some tasks but a poor learner for other tasks. When expert networks with different structural properties compete to learn the training patterns, each network tends to win the competition for those patterns belonging to the task for which its structure makes it a good learner. Consequently, the architecture is capable of discovering structure-function relationships. The performance of the architecture is consistent with the theory that genetic instructions do not necessarily stipulate directly the function to be performed by each brain region. Instead, genetic instructions may assign different structural properties to different regions. These structurally different regions may then, due to their performance characteristics, take on particular functions for which they are well-suited (cf. Bever (1980) and Kosslyn (1987) for related processing accounts of cerebral lateralization). Simulation results using the ME architecture, as well as descriptions of other related architectures, can be found in Jacobs and Jordan (1993), Jordan and Jacobs (1994), Jacobs and Kosslyn (1994), and Peng, Jacobs, and Tanner (1996).

IMAX Learning Architecture

The IMAX learning architecture is a system that uses convergent computation and it was proposed by Becker and Hinton (1992). The modules of this architecture receive data from different modalities (such as vision and touch) or from the same modality at different times (such as consecutive views of a rotating object) or even spatially adjacent parts of the same visual image. It is assumed that different portions of the perceptual input have common causes in the external world. Modules that look at separate but related portions can discover these common causes by striving to produce outputs that agree with each other. In particular, modules adjust their parameters so as to maximize the mutual information among their outputs. This occurs when the output of each module can be used to predict the outputs of the other modules (see Figure 2). An interesting feature of this architecture is that its learning procedure is entirely unsupervised; there is no external teacher that provides the architecture with training information. Instead, each module acts as a teacher for each of the other modules in the sense that each module compares its output with the outputs of the other modules. Becker and Hinton (1992) showed that when two modules view adjacent patches of two-dimensional visual images, an architecture that has no prior knowledge of the third dimension can discover depth in random dot stereograms of curved surfaces.

Acknowledgments

This work was supported by NIH grant R29-MH54770.

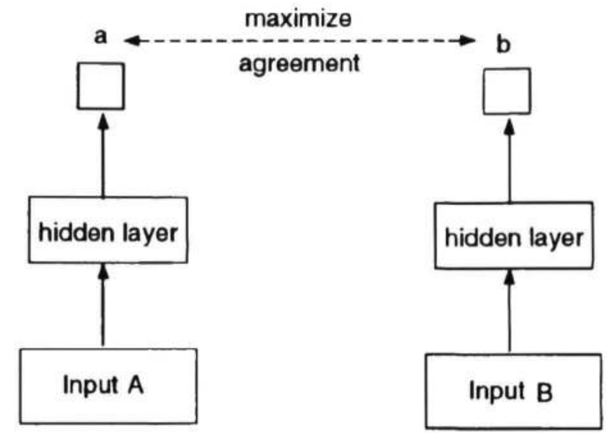

Figure 2: IMAX Learning Architecture (adapted from Becker and Hinton(1992))

References

Becker, S. & Hinton, G.E. (1992) Self-organizing neural network that discovers surfaces in random-dot stereograms. *Nature*, 355, 161-163.

Bever, T.G. (1980) Broca and Lashley were right: Cerebral dominance is an accident of growth. In D. Kaplan & N. Chomsky (Eds.), *Biology and Language*. Cambridge, MA: MIT Press.

Fodor, J.A. (1983) *The Modularity of Mind.* Cambridge, MA: MIT Press.

Jacobs, R.A. & Jordan, M.I. (1993) Learning piecewise control strategies in a modular neural network architecture. *IEEE Transactions on Systems, Man, and Cybernetics*, 23, 337-345.

Jacobs, R.A., Jordan, M.I., Nowlan, S.J., & Hinton, G.E. (1991) Adaptive mixtures of local experts. *Neural Computation*, 3, 79-87.

Jacobs, R.A. & Kosslyn, S.M. (1994) Encoding shape and spatial relations: The role of receptive field size in coordinating complementary representations. *Cognitive Science*, 18, 361-386.

Jordan, M.I. & Jacobs, R.A. (1992) Modularity, unsupervised learning, and supervised learning. In S. Davis (Ed.) *Connectionism: Theory and Practice.* Oxford University Press.

Jordan, M.I. & Jacobs, R.A. (1994) Hierarchical mixtures of experts and the EM algorithm. *Neural Computation*, 6, 181-214.

Karmiloff-Smith, A. (1992) *Beyond Modularity: A Developmental Perspective on Cognitive Science.* Cambridge, MA: MIT Press.

Kosslyn, S.M. (1987) Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, 94, 148-175.

Piaget, J. (1955) *The Child's Construction of Reality.* Routledge and Kegan Paul.

Peng, F., Jacobs, R.A., & Tanner, M.A. (1996) Bayesian inference in mixtures-of-experts and hierarchical mixtures-of-experts models with an application to speech recognition. *Journal of the American Statistical Association*, in press.