-89-14

A position paper for the panel on Future of AI at IJCAI-89, Detroit

9/21/89

FUTURE DIRECTIONS OF AI IN A RESOURCE-LIMITED ENVIRONMENT

Ryszard S. Michalski
David C. Littman
Center for Artificial Intelligence
and
Department of Computer Science
George Mason University
Fairfax, VA

Artificial intelligence (AI) research, as any other human endeavor, operates in a resource-limited environment. It becomes increasingly necessary to make choices regarding the types of research to support and the directions in which to encourage young professionals to develop their interests and expertise.

It is probably self-evident that these directions of AI depend upon the response to three general issues:

Issue 1: *What functionality should future intelligent systems have?*
Issue 2: *What inference capabilities are required to achieve the desired functionality?*
Issue 3: *What engineering considerations should be respected to achieve this functionality?*

In our view, the capabilities of future intelligent systems should include:

- Instructability, i.e., it should be relatively easy to communicate all kinds of knowledge to them as well as to teach them all kinds of skills
- The ability to explain the reasoning behind any conclusion, plan or task execution in human understandable terms and constructs, except in such situations where such an explanation is clearly unimportant. For example, in situations pertaining to human medical treatment, economical advice, or military defense actions, such an explanation would be necessary;

for situations regarding robot manipulation or a delivery task, an explanation may not be important if the task was performed satisfactorily.

- The ability to interact with humans through their own media, e.g., texts, speech and images.

As to inference capabilities, future AI systems should be capable of performing the type of inference that humans would label plausible or commonsense. These systems should be capable of meta-knowledge reasoning, forming descriptions or hypotheses at different levels of abstraction, evaluating and selecting different representational schemes, and constructing and evaluating goals and plans. All these capabilities would engage diverse kinds of knowledge.

Therefore, individual systems would have to possess various parts of technical, as well as intellectual knowledge that humanity has accumulated during its existence. We believe the implication that ultimately these systems would have to be able to acquire knowledge from human texts at least partially by themselves, as well as from human speech.

As for engineering considerations, AI systems would have to be cost-effective, maintainable, reusable and generalizable.

In order to make good use of limited resources and funds, one must evaluate the implications of these three issues in selecting paradigms for AI research and development, and weigh the proper balance of effort.

At present, two general approaches to AI are predominant -- the symbolic paradigm and the connectionist paradigm, and their relative importance is increasingly the subject of discussion. Much of this discussion results from the lack of a clear understanding of their respective capabilities. The symbolic paradigm, which emphasizes explicit, localized concept representation, and symbol manipulation for expressing concept transformations, is well established. It has had many successes but now it faces the growing difficulty of building large-scale knowledge bases and implementing powerful multi-type inference systems.

The fledgling connectionist paradigm, which focuses on distributed non-perspicuous knowledge representations, is becoming increasingly popular. The research in this paradigm is highly experimental, and avoids, at least at the present time, the issues of representing/manipulating large amounts of knowledge and diverse types of knowledge. Because it focuses on systems that are highly data-oriented and relatively knowledge-poor, it is not difficult for one researcher, or a few researchers, to make a series of experiments, and publish a research contribution in a short period of time. The latter factor significantly contributes to the current attractiveness of research in this paradigm.

Research on both, the connectionist paradigm and the symbolic paradigm, may primarily address one of two different yet interrelated goals: 1) a cognitive one, concerned with increasing our understanding of the functions of a human or an animal brain that are responsible for intelligent behavior; or 2) an engineering one, concerned with developing systems that ultimately are able to perform some useful task, or serve as a useful tool for human society. In this paper, we will consider these paradigms from the engineering perspective.

As a step towards dealing with relative merits of two approaches, we have identified five candidate views for guiding future AI development. We also suggest which candidate we feel is most likely to succeed in this goal.

1) *Connectionist hegemony* holds that most, if not all, aspects of intelligent information processing can and should be performed in a substrate, as similar as possible to the low-level structure of the nervous system.

2) *Symbolic hegemony* holds that most, if not all, aspects of intelligent information processing can and should be performed using symbolic representations..

3) *Connectionist/symbolic equality* holds that connectionist and symbolic methods are equally capable for implementing intelligent behavior, and thus, a priori, there is no reason for preferring the connectionist or the symbolic approach for implementing any particular intelligent functionality. A modification of this view is that for different problem classes different approach might most effective, and thus both approaches should be viewed as equally important for AI.

4) *Connectionist preeminence* holds that, while symbolic methods may be useful to solve certain types of problems, such as performing numerical computations and some symbolic manipulations, connectionist methods are more likely than symbolic methods to permit the implementation of the desired functions in future intelligent systems.

5) *Symbolic preeminence* holds that, while certain aspects of intelligent information processing might be most effectively implemented in a connectionist system (e.g., low-level control

functions or background knowledge-limited signal processing, most of the high level functions responsible for intelligent behavior (e.g., context-dependent plausible reasoning, high-level abstraction, goal formulation, conceptual learning, complex planning and determining problem-oriented representations) are most appropriately and cost-effectively performed symbolically.

In the remainder of this paper, we present some reasons that convinced us that symbolic preeminence is the best bet for guiding AI research in the future.

Symbolic processing systems can perform with ease a larger variety of complex information processing operations than can connectionist systems. In contrast to the connectionist approach, the symbolic processing paradigm provides several high-level, powerful and easily implementable functions. These functions include copying, modifying or erasing various parts of knowledge, operating on names representing components of knowledge, and performing "conceptual" inference operations such as generalizing, specializing or similizing selected knowledge segments. We define a "conceptual" inference operation as a transformation of knowledge which is easy to understand, and which can be simply explained in terms of human concepts. For example, generalizing a decision rule by removing a condition is a "conceptual" operation of generalization. On the other hand, generalizing a set of equations by replacing one numerical coefficient by another numerical coefficient, so that the modified set of equations describes a larger set of entities, is not a conceptual generalization.

The basic operation underlying the function of connectionist systems is parameter modification, i.e., a modification of the strength of the connections between various units. Through this modification, structural properties of knowledge represented by a network are modified indirectly. Such an operation is relatively easy to implement in biological systems, namely, through biochemical processes affecting synapses of neurons. Computer systems, however, offer information processing capabilities very different from those of biological systems. Consequently, it is likely that a possible future intelligent machine will be very different from its human counterpart. For example, this machine would be able to copy knowledge to (or from) another machine or to erase, on command, undesirable segments of knowledge from its memory. Both of these operations are not possible for a human brain.

Focusing on a cost-effectiveness argument, let us suppose that the connectionist approach would overcome various important limitations of the current systems (e.g., how to represent complex structural descriptions with quantified variables), and that we could assume that symbolic and connectionist systems are computationally equivalent. This argument, however, does not resolve any issue. For example, the Turing machine is theoretically equivalent to the contemporary computer, but from the pragmatic viewpoint, these two computational devices are vastly different. Because the symbolic paradigm is able to perform effectively more operations, this approach would likely lead to more cost-effective engineering solutions than the connectionist paradigm, given the assumption that these solutions would implement the capabilities stated earlier as being desirable in future intelligent systems.

Finally, let us use an analogy from the history of science. The success in the development of modern aircraft was primarily due to engineering experimentation and the development of general principles of flight independent of biological considerations, rather than the study and imitation of the biological properties that enable birds to fly.

In sum, we believe that there is a clear need to fully determine the capabilities and limitations of the connectionist and symbolic approaches. However, one should not be deceived by believing that there exists a "shortcut" to building an intelligent machine. With any new technology, the old dream comes back that somehow machines can self-organize themselves and learn on their own to do everything for us. The following issues remain central for the development of AI:

- how to implement commonsense and plausible reasoning
- how to develop powerful learning capabilities that are able to take advantage of all kinds of prior knowledge and to explain to humans what was learned
- how to introduce large amounts of human knowledge to machines and how to update that knowledge
- how to recognize objects and concepts from incomplete, variable and context-dependent cues

These issues, which are basic items on the agenda of the symbolic approach, cannot be avoided and should be given a high priority in the development of intelligent systems.