

Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory

Risto Miikkulainen

(University of Texas, Austin)

Cambridge, MA: The MIT Press
(Neural Network Modeling and
Connectionism Series, edited by Jeffrey
L. Elman), 1993, xii + 391 pp.
Hardbound, ISBN 0-262-13290-7, \$45.00

Reviewed by

Lawrence A. Bookman

Sun Microsystems Laboratories, Inc.

Much connectionist research in natural language processing has been concerned with isolated aspects of understanding language. Very few researchers have attempted to build comprehensive computational models that are biologically and psychologically plausible and that incorporate the components necessary for modeling and testing various complex high-level cognitive phenomena. Miikkulainen's book is an exception to this trend. Using script understanding as a testbed, he shows how script-based inferences can be learned from experience on the basis of the statistical correlations implicit in the example data. He also shows how episodic memory organization can be automatically formed on the basis of these statistical regularities, and how word semantics can be learned from actual use. In an attempt to overcome some of the limitations of traditional AI symbolic approaches to this problem—the processing architecture, mechanisms, and knowledge are hand-coded, and inferences are based on handcrafted rules—he constructs a distributed neural network model composed solely of artificial neural network components.

An important aspect of his system is its ability to address such questions as where performance errors, such as dyslexic errors and semantic slips, come from, how memory can become overloaded, and why certain types of memory confusions can occur in such situations. Constructs such as topological and hierarchical feature maps are introduced to address such issues. Topological maps have the property that complex similarity relationships of some high-dimensional input space become visible on the map. In addition, the maps can be formed by an unsupervised learning process. The hierarchical nature of these maps makes it possible to characterize the input from several graded perspectives: from gross high-level classifications to more specific instantiations of data. Thus a story about Bill eating a lobster pizza appetizer at Biba in Boston could be grossly characterized as a story about a restaurant, or more specifically, a fancy restaurant, or even more specifically about Bill eating lobster pizza.

1. Outline of the Book

The book is subdivided into four parts. Part I provides an overview of both the book and the model DISCERN. Part II describes the basic processing mechanisms and building blocks of DISCERN. Miikkulainen shows how using an extended form of back-propagation he can construct an integrated architecture, composed of separate back-propagation networks, that can automatically parse the input story and answer

questions about the stories in its memory. What is unique about this is the ability to answer ambiguous, misleading, and incorrect questions. These capabilities result from the special properties inherent in distributed representations. Part III discusses the memory mechanisms of DISCERN. Memory consists of both an episodic and a lexical and semantic memory (the lexicon). Episodic memory stores input story traces that can later be retrieved with a partial cue. These traces are implemented as topological feature maps and as such exhibit a recency preference (if similar items are stored on a map, the later traces are more likely to be retrieved), and show effects of retroactive interference and unlearning effects. The lexicon forms the interface between the model's internal world and its external environment. It maps symbols to their meanings, and vice versa. It consists of a semantic and a lexical memory connected via associative connections; the latter memory contains representations for orthographic symbols and their mapping onto the representations of concepts in semantic memory. Words whose orthographic forms are similar (e.g., BALL, DOLL) are represented by nearby units on the lexical map, and words with similar semantic content (e.g., predator, prey) are mapped near each other on the semantic map. This dual representation can account for errors similar to those observed in human aphasia and dyslexia. Part IV exhaustively evaluates DISCERN's performance and describes how well it succeeds in explaining various phenomena in human language processing and memory, such as memory confusions and memory overloading, incomplete input, and slips of the tongue. Many extensions to the basic model are also discussed.

2. Critical Commentary

One of the major goals of the DISCERN project as stated by the author was to demonstrate that a complete natural language processing system can be built from distributed artificial neural networks. As Miikkulainen states, symbolic systems cannot use the statistical properties of the data to enhance processing—they can only achieve what they are programmed for. Since the acquisition of human knowledge is a dynamic process that undergoes constant modification and refinement, the representation of such knowledge requires flexible structures that should not be preset, fixed, or all-or-none. Instead, these representations should somehow emerge automatically from the regularities inherent in the experience. One of the major contributions of this work is to demonstrate how this can be done automatically in a manner that is grossly consistent with known biological and psychological data. Natural language processing requires such dynamic representations and the use of automatic techniques if it is to scale up to the level of human performance. This view is in sharp contrast to the view taken in *Cyc* (Lenat et al., 1990). There, handcrafted techniques in conjunction with fixed representations are used to construct large knowledge bases. These two views thus represent two extremes of the spectrum. One open question is how far we can push the view proposed in this book—can we extract the regularities inherent in our experience using modular extended back-propagation networks or one of its future incarnations? Or must we use some combination of methods, first seeded with some handcrafted linguistic knowledge or templates?

For script-based processing, Miikkulainen's approach represents a plausible beginning. However, some difficult questions remain unanswered. Consider the author's explanation of this story:

John went to MaMaison. John asked the waiter for lobster. John left a big tip.

The author claims that from *lobster* and *big tip*, you can infer that this is a fancy restaurant and that the food is good. However, eating lobster should not necessarily allow one to conclude this is a fancy food restaurant, as lobster itself does not necessarily represent fancy food. The situation is a lot more complex. Lobster can be eaten in many restaurants, many of which would not be considered fancy. In general, there are too many other possibilities and some of them are equally likely (and with a comparable level of probability). While this system is a vast improvement over Schankian script theory, it is still far from dealing with the full complexity of real-world situations. One potential problem with this approach is its reliance on “blending” or averaging of different alternative values. For example, what would happen if there were many alternative values for each case slot? Would the retrieval mechanism still work, given that it takes an average of the different alternatives? If not, then the question-answering mechanism would also not work as well. In many cases, blending might be the right thing to do, but if the system were to process thousands of different stories, I can see where this mechanism will certainly fail.

Finally, is this architecture capable of representing information for texts that are not as structured as scripts, that might include such nonregularized knowledge as ad hoc categories (Barsalou 1983)? As Miikkulainen states, parallel distributed systems generalize fairly well to new situations, if these situations are similar to the ones originally trained on. However, these systems perform poorly when it comes to handling exceptions and novel situations. Thus for tasks for which there is little or no regularity, these systems will be somewhat limited unless there is further development of new techniques.

3. What This Work Offers the Linguistics Community

Miikkulainen’s take on meaning represents a fundamental departure from traditional linguistic theories, but it has much in common with the current trends in corpus-based linguistics. In DISCERN, the meaning of a word is extracted from its use, that is, from all the contexts where the word has been encountered. In other words, meaning is derived from similarity of use. If there is enough difference between uses, the process described by Miikkulainen will develop different representations for the words. These representations are developed by looking at the correct case-role assignments trained from example data.

As implied above, the inferences that DISCERN generates are based on a very simplified view of the world. In DISCERN, there exist three kinds of scripts, with each script having three different tracks. For example, the basic restaurant script has a fancy-food track, a coffee-shop track, and a fast-food track. One question to ask is whether a more sophisticated analysis based on a finer-grained linguistic analysis can improve its script-based inferencing capabilities. That is, are statistical techniques sufficient, or must they be guided by theoretical linguistic tools? For example, instead of generating a word’s representation from case-role assignments, one might try generating these representations from its qualia structure (Pustejovsky 1991), or use some combination of the two.¹ Thus for a noun such as *tape*, its qualia structure might state that a tape is an “information container” that is also a two-dimensional physical object, where the information is written onto the object. A question for both the AI and linguistics

¹ The qualia structure specifies the different aspects of a word’s meaning through the use of subtyping. These include the subtypes *constitutive* (the relation between the object and its constituent parts), *formal* (that which distinguishes it within a larger domain), *telic* (its purpose and function), and *agentive* (factors involved in its origin).

community to explore is this: can linguistic analysis improve the sophistication and performance of computational models such as the one described by this book, and can the methods from AI (i.e., the natural language neural net community) allow one to better evaluate linguistic models? I believe the answer to both questions is *yes*, and that this book can serve as a useful bridge between these two communities.

4. Conclusions

The advantages of neural networks—adaptability (they can extract complex relationships from data), generalization (they can correctly process data that is broadly similar to the originally trained data), massive parallelism (they contain identical independent operations), noise tolerance (they can provide sensible answers in noisy environments), and graceful degradation (they degrade gradually in proportion to the amount of system damage)—make them very attractive for applications that need to cope with noise, imprecision, and complexity. What is unique about DISCERN is that Miikkulainen has cleverly figured out a way of bringing the benefits of the parallel distributed approach to a task that was previously thought to be more suitable to symbolic processes. Moreover, the research is backed up by careful experimentation, in which both the strengths and limitations of his system are honestly discussed. Despite the open questions left by this approach, Miikkulainen has written an outstanding book that should be of great interest to the computational linguistics community.

References

- Barsalou, L. (1983). "Ad hoc categories." *Memory and Cognition*, 11(3), 211–227.
- Lenat, D.; Guha, R.; Pittman, K.; Pratt, D., and Shepherd, M. (1990). "Cyc: Toward programs with common sense." *Communications of the ACM*, 33(8), 30–49.
- Pustejovsky, J. (1991). "The generative lexicon." *Computational Linguistics*, 17(4), 409–441.

Lawrence A. Bookman is a research scientist at Sun Microsystems Laboratories, Inc. His books *A Two-Tier Model of Semantic Memory for Text Comprehension* and *Computational Architectures Integrating Neural and Symbolic Processes* (co-edited with Ron Sun) will be published by Kluwer Academic Publishers in 1994. Bookman's address is Sun Microsystems Laboratories, Inc., Two Elizabeth Drive, Chelmsford, MA 01824-4195; e-mail: lbookman@east.sun.com.