Scaling Up Grounded Representations Hierarchically

Karl Pfleger
Knowledge Systems Laboratory
Computer Science Department,
Stanford University, Stanford, CA 94305 USA
kpfleger@cs.stanford.edu

Abstract

We have been studying the learning of compositional hierarchies in predictive models, an area we feel is significantly underrepresented in machine learning. The aim in learning such models is to scale up automatically from fine-grained to coarser representations by identifying frequently occurring repeated patterns, while retaining the ability to make predictions based on the statistical regularities exhibited by these patterns. Our hierarchical learning begins with data consisting of discrete symbols and can be viewed as a method of grounding high-level concepts in terms of their lower-level parts, which are themselves grounded in raw, environmental signals by other means. This short paper discusses the relationship between hierarchy learning and the learning of low-level grounded representations and also very briefly describes one of our systems for compositional hierarchy learning. A much more detailed discussion, including an extensive literature review, can be found in (Pfleger 2000).

Compositional Hierarchy Learning

Many AI systems operate using representations at levels well above raw environmental data. Whether they are called states, operators, features, rules, schemata, scripts or some other name, the successes or failures of systems have often depended more critically on human choices for instantiating the representational units than on the choices of algorithms or the designs of the representational languages themselves. In order to design highly intelligent autonomous agents with flexibility and aptitude far surpassing that which has been artificially created today (Nilsson 1995; Newell 1990; Hayes-Roth 1993), a vast number of representational units spanning many levels of resolution and abstraction will be needed, and it seems evident that learning must play a role in developing these representations.

We agree with (Sun 2000) that high-level representational units must be grounded in the environment. This includes the implications that both representation and environmental interaction are important. We also agree that a bottom-up learning process is necessary to create the high-level units.[1]

This vague notion of a continuum from low-level to high-level representations conflates two correlated measures of representational level based on two distinct types of structure. *Taxonomic structure* relates entities by `is-a` relationships, and *compositional structure* relates entities by `part-of` relationships. Machine learning and related fields have a long history of producing systems that learn taxonomic hierarchies, but the field has not made corresponding progress at building compositional hierarchies (CHs). In fact, the building of CHs has only surfaced in a few areas and in each case the learning depends on specialized features of the subdiscipline, such as the top-down use of separate domain theories.

We have been attempting to fill this gap by investigating bottom-up, data-driven methods for learning compositional hierarchies. Our methods are domain-independent and action-independent, and they learn in an on-line fashion. The methods operate on unbounded, stationary data (e.g., unbounded sequences) that consist of discrete symbols drawn from a finite alphabet. Learning is unsupervised, and the models are capable of arbitrary statistical inference (prediction). The learning paradigm is analogous to the standard unsupervised prediction paradigm for IID data, adapted to unbounded sequences or higher dimensional analogs. See (Pfleger 2000) for precise details on the learning paradigm.

The general strategy for incrementally building CHs is to repeatedly combine, or *chunk*, frequently occurring patterns into higher-level aggregates, enabling the future combination of these patterns into even larger, higher-level aggregates. Thus, the stance taken here is that representations are learned by noticing and storing patterns exhibited by the environment. Once learned, CHs lead naturally to smooth integration of bottom-up and top-down processing that can mediate lateral in-

[1] We do not, however, agree that decisions about actions are the necessary type of environmental interaction for the creation of *all* forms of high-level concepts. See (Pfleger 2000).

ferences, and the chunks themselves can be useful for memory (Miller 1956; Chase & Simon 1973), communication, planning or reacting, and as features for other learning tasks.

Unlike most learning systems, the learning is both online and structural (the model grows). The underlying hypothesis in this work is that compositional hierarchies are a useful inductive bias for widening models. From a bias/variance perspective, building chunks of frequently occurring patterns by structurally introducing new parameters into the model around the representations of the new patterns allows models to concentrate more parameters exactly where there has been more data.

Symbol Grounding

There is a growing body of work attempting to address the grounding problem of how representational units derive their meanings from the environment by way of low-level data. The thrust of this work is to make a correspondence between available "raw" data, usually continuous, and newly created "symbolic" data, or higher-level representations. For example, (Oates, Schmill, & Cohen 2000) demonstrates clustering of multivariate time series data derived from mobile robotic sensors. (Singer & Tishby 1994) provides a way to map continuous handwritten pen trajectories into a sequence of discrete symbols. Other examples of grounding work include (Coradeschi & Saffiotti 2000) and (Siskind 2000), but these efforts involve richer representations, which presents more difficulty for learning.

The important question is not only how to ground representational units one level above the lowest level, but how the grounding of representational units at all levels can be learned. Rather than tying representations at all levels directly to raw environmental data, it is easier to make the correspondence for high-level units by tying them to low-level representations, which are in turn tied to raw environmental data. The ability to learn hierarchical representations in a bottom-up fashion from the low-level symbolic representations, including specifically the ability to learn hierarchies with compositional structure, is critical for this representational scaling up.

Thus, discrete symbols from a finite alphabet can be seen as an entirely reasonable format of input data for autonomous intelligent agents since work such as (Oates, Schmill, & Cohen 2000) and (Singer & Tishby 1994) provide explanations of how these symbols can arise. At the same time, the type of bottom-up hierarchical learning described here provides a plausible means for representations of increasingly expansive signal patterns to be grounded. These two processes need not happen in isolation. Sophisticated combinations of these ideas will allow for interesting interplay between the mechanisms. Note that these issues span a diverse set of domains. Another example is provided by the word learning domain where infants must generate discrete, symbolic representations such as syllables from raw acoustical data, and must also, if you believe models such as those of (Brent & Cartwright 1996) and (Saffran, Aslin, & Newport 1996), build higher-level structures out of these symbols to form representations of words that are thereby also grounded.

Temporally Extended Graphical Models

The two main issues for learning predictive CHs are (a) how to embody a CH in a representation that allows prediction, and (b) how to incrementally grow the CH in response to new data. Our work is directly inspired by the Interactive Activation Model of context effects in letter perception, or IAM (McClelland & Rumelhart 1981), which structurally encoded a CH of individual letters and 4-letter words into a symmetric, recurrent (relaxation-style) neural network. The basic behavior of the network was to make inferences about the letters through bottom-up and top-down flow of activation along part-whole links, but there was no learning.

There have been numerous temporally extended graphical models, such as Hidden Markov Models, dynamic belief networks (Dean & Kanazawa 1989), or Boltzmann chains (Saul & Jordan 1994), which can represent dependencies of varying width and can make predictions from partial information. Currently, such models cannot incrementally grow new hidden variables, however, and thus cannot utilize long-term learning to scale up from low-level to high-level representations. Our work utilizes a form of temporally extended graphical model that creates new hidden variables within structures that mimic CHs (like IAM).

Specifically, our network is an undirected graphical model based on Boltzmann machines (Hinton & Sejnowski 1986). It is thus a symmetric, recurrent neural network with very similar operation to IAM, and can be viewed as a direct descendent of IAM addressing the lack of learning in that system. The chunking mechanism is based on correlations indicated by large Hebbian weights and is similar in spirit to the mechanism in Ring's first system (Ring 1995). Our system is very briefly described in (Pfleger 1999). A detailed paper on this system is in preparation.

References

Brent, M., and Cartwright, T. 1996. Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition* 61:93–125.

Chase, W., and Simon, H. A. 1973. Perception in chess. *Cognitive Psychology* 4.

Coradeschi, S., and Saffiotti, A. 2000. Anchoring symbols to sensor data: Preliminary report. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence.*

Dean, T., and Kanazawa, K. 1989. A model for reasoning about persistence and causation. *Computational Intelligence* 5(3):142–150.

Hayes-Roth, B. 1993. On building integrated cognitive agents: A review of Allen Newell's *unified theories of cognition*. *Artificial Intelligence* 59:329–341.

Hinton, G. E., and Sejnowski, T. J. 1986. Learning and relearning in Boltzmann machines. In *Parallel Distributed Processing: Explorations in the Microstructures of Cognition, Volume 1: Foundations*. MIT Press. chapter 7.

McClelland, J. L., and Rumelhart, D. E. 1981. An interactive activation model of context effects in letter perception: Part 1. an account of basic findings. *Psychological Review* 88:375–407.

Miller, G. A. 1956. The magic number seven, plus or minus two: Some limits of our capacity for processing information. *Psychological Review* 63.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, Mass.: Harvard University Press.

Nilsson, N. J. 1995. Eye on the prize. *AI Magazine* 16(2):9–17.

Oates, T.; Schmill, M.; and Cohen, P. 2000. A method for clustering the experiences of a mobile robot that accords with human judgements. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*.

Pfleger, K. 1999. Learning compositional hierarchies by data-driven chunking. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence*. (abstract).

Pfleger, K. 2000. Learning predictive compositional hierarchies. Technical report, KSL, Stanford CS. See www-cs-students.stanford.edu/~kpfleger/publications/.

Ring, M. 1995. *Continual Learning in Reinforcement Environments*. München, Wien: R. Oldenbourg Verlag.

Saffran, J.; Aslin, R.; and Newport, E. 1996. Statistical learning by 8-month-old infants. *Science* 274:1926–1928.

Saul, L. K., and Jordan, M. I. 1994. Boltzmann chains and hidden Markov models. In *Advances in Neural Information Processing Systems 7*. MIT Press.

Singer, Y., and Tishby, N. 1994. Dynamical encoding of cursive handwriting. *Biological Cybernetics* 71.

Siskind, J. M. 2000. Visual event classification via force dynamics. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence*.

Sun, R. 2000. Symbol grounding: A new look at an old idea. *Philosophical Psychology* 13(2).