Systems Learning for Complex Pattern Problems

Omid Madani
AI Center, SRI International
Menlo Park, California 94025
madani@ai.sri.com

Abstract

Learning numerous complex patterns, or *concepts*, in higher animals and humans, for instance for the task of visual object recognition, may require extensive exposure and learning experience over a relatively long time period. What are the principles behind such long-term learning and development? We describe our efforts in this direction and discuss the many challenges.

Introduction

Categorization is fundamental to intelligence (Murphy 2002; Lakoff and Johnson 1999). Without categories, every experience would be new, and we couldn't make sense of our world. Categorization allows for proper subsequent actions and informs intelligent decisions.

It is also the case that generally speaking, the higher the intelligence, the more concepts (categories) and interrelations among them need to be acquired. The myriad concepts should be highly efficiently and accurately accessed to be useful in tasks such as visual scene interpretation. How are concepts, from the lowest levels of visual shapes and primitive sound patterns ("round", "hand", "table", "wall"), to the highest abstract ones (such as "democracy", "philosophy", "infinity"), acquired and developed?

There remains much to be discovered regarding the computational processes underlying the learning and development of what may be referred to as the perceptual concepts (Rakison and Oakes 2003). These concepts are utilized for complex pattern problems including human-level visual capability (*i.e.*, complex pattern recognition and generation). Such perceptual concepts are the basis for how we interpret the world as well as the foundation for the more abstract categories.

We have been investigating a framework based on the following premises (Madani 2007b): 1) much of the (perceptual) concepts can be effectively learned, or preprogramming is not required, 2) substantial learning and experience is required (for instance, years in terms of humans), 3) the learning processes are almost entirely unsupervised, in the sense that no explicit teacher (parent, etc.) is required. 4) A *systems* approach is likely needed, *i.e.*, we need several algorithms to achieve the different learning functions and whatever else is required to support the main functions. 5) Due to the long-term and large-scale nature of the task, it's best to look for algorithms that are primarily online and that can handle significant nonstationarity and uncertainty. There are numerous challenges. To start with, even useful formulation of many of the problems should require substantial research. We will next briefly describe our particular approach further, and highlight a few of the challenges.

The Task

In this learning framework, the system continually processes and learns from a stream of bits from a rich world. For instance, in the visual domain, the training signal can come from video streams or many millions of the available online image. We have built an exploratory system on text streams (Madani 2007a), taking as input the corpora of news articles. There is much rich statistical regularities embedded in such data, as we'll explain shortly. The key idea has been that simple *prediction* of the content of the streams can be powerful in providing ample and effective learning. The particular details of the prediction task depends on the domain and applications. Broadly however, the system segments its stream into *concepts* it currently has, and hides one or more of the concepts and attempts to predict them using surrounding concepts as predictors. Concepts correspond to patterns. In text, concepts can begin at the lowest level of single characters, the *primitive* concepts. The primitive concepts are given to the system (hard wired). *Higher level* concepts are more complex. Such concepts are built out of the lower level concepts via *composing* and *grouping* operations. Thus after many episodes of observing "e" immediately followed by "w", the system may compose them to create the concept "ew". And some time later, the concepts "new" and then later "new york" can be created. Groupings are made of concepts that tend to occur in similar contexts ("synonymous" concepts), such as the digits (0,1,...,9), or days of the week. Using both composition and grouping, higher level concepts such as the patterns for phone numbers (a composition of digits and other concepts) should be discovered. The system uses the concepts it currently holds to predict the next concept it will see in its stream. Please see (Madani 2007b) for further description and discussion of the learning task. Can we build systems that can stand on their own for

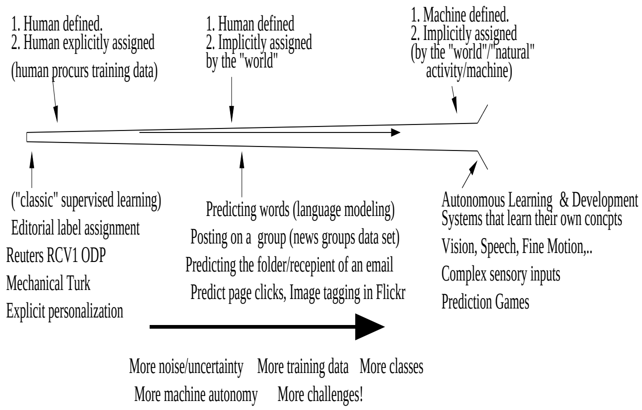

Figure 1: A view of supervised (feedback driven) learning in terms of who defines the classes of interest and who/what provides the class feedback during learning. As we move from left to the right, the noise/uncertainty but also the amount of training data as well as the number of classes can increase substantially. At the right end of the spectrum, we arrive at systems that build and develop their own concepts. The width of the axis enlarges to reflect increasing number of challenges and possibilities.

long time periods and continue to fruitfully learn? We now briefly list a few of the challenges (see (Madani 2007b) and (Madani 2007a) for an extended exploration).

The Challenges (a Subset)

Handling Many Classes Scalability constraints have a number of implications, but we will focus on one aspect that our research has explored. It is clear that the number of concepts that the system may acquire can quickly grow to exceed tens of thousands, and the system somehow has to scale gracefully with this growth. At any point in time, the system is presented with features of the situation. For instance, in the text domains, these are the character sequences (concepts) that the system could have just seen. The system is to use such cues to quickly and accurately predict the concept (the pattern) that it will see next. This is a *many-to-many mapping problem*. We are investigating (supervised) learning algorithms to build and update sparse indices (bipartite mappings) to address the challenges (*e.g.*, (Madani and Connor 2008; Madani and Huang 2008)). We note that while the algorithms are supervised, the feedback that the system obtains need not come from humans. Figure 1 gives a perspective on supervised learning. On the left end of the spectrum, humans both define the classes of interest and procure training data for the supervised learning algorithm. As we move to the right, the extent of human involvement decreases, and the feedback is primarily provided by the natural environment or certain natural activities (such as simply listening to all available recorded conversations in English). The increasing width of the axis indicates that the challenges and opportunities for research increases as we move to the right.

Systems that Learn their own Concepts Scalable many-class learning can support the tasks of prediction from many to many classes (concepts), *given* the concepts, but we now desire systems that build their own concepts, from the ground up. Concepts that are simply strings of bits (deterministic sequences or layouts) may be learnable (possibly with additional realistic assumptions). However, we seek also flexibility in concepts. Thus the phone number concept should match any sequence of digits and other characters that conforms to the phone pattern. What is the nature of learnable concepts? Unfortunately, much research in learning theory provides many negative results on the learnability of seemingly simple concept classes such as stochastic grammars and probabilistic finite automata. Even more daunting, we seek systems and algorithms for learning and developing *many* concepts, and somewhat in tandem (not one at a time). We assume no help from a teacher (who might segment the input and tag with the appropriate concepts). A key idea, mentioned earlier, is that prediction of unseen bits can serve for validation and adaptation of prediction weights and concepts. However, there remains much uncertainty: the fidelity of the concepts the system uses is always uncertain. Furthermore, the segmentation process that breaks the input stream of bits into concepts (to serve as predictors or targets of prediction) also interacts with concept generation. How can robust but flexible segmentation be learned in an unsupervised manner?

Finally, we raise the issue of nonstationarity. Much work in learning, with the exception of online learning, has been based on the simplifying assumption of a stationary distribution. However, particularly initially in development, there is much nonstationarity especially within the system. And nonstationarities remain everywhere throughout our life times. For example, our brain's face detectors need to adapt gracefully to the changes in faces of all the persons that we can identify (children growing, aging,..). Nonstationarity limits the amount of training available, requires ongoing adaptation, which has costs, and requires an effective way for forgetting parts of the past.

Conclusions The study of systems that develop their own numerous complex concepts (patterns), when situated in rich environments for indefinite periods, offers many challenges as well as possibilities. By striving to build such systems, we seek to gain deeper insights into the underlying problems, and to understand the complex relation between the system and its world, that makes intelligence a possibility.

References

Lakoff, G., and Johnson, M. 1999. *Philosophy in the flesh: the embodied mind and its challenge to western thought*. Basic Books.

Madani, O., and Connor, M. 2008. Large-scale many-class learning. In *SIAM Conference on Data Mining*.

Madani, O., and Huang, J. 2008. On updates that constrain the number of connections of features during learning. In *ACM KDD*.

Madani, O. 2007a. Exploring massive learning via a prediction system. In *AAAI Fall Symposium: Computational Approaches to Representation Change During Learning & Development*.

Madani, O. 2007b. Prediction games in infinitely rich worlds. In *AAAI Fall Symposium*. Yahoo! Research Technical Report.

Murphy, G. L. 2002. *The Big Book of Concepts*. MIT Press.

Rakison, D. H., and Oakes, L. M., eds. 2003. *Early Category and Concept Development*. Oxford University Press.