Prediction Games in Infinitely Rich Worlds

Omid Madani
Yahoo! Research
3333 Empire Ave
Burbank, CA 91504

Abstract

How can an entity acquire millions of complex inter-related categories? The process of playing *prediction games* may make this massive learning possible. This paper briefly describes the games and discusses the basic philosophy underlying the approach.

Introduction

Categorization is fundamental to intelligence (Murphy 2002; Lakoff & Johnson 1999). Without categories, every experience would be new, and one couldn't make sense of one's world. Categorization allows for proper subsequent actions and informs intelligent decisions. Furthermore, an entity needs to acquire large numbers of categories, perhaps millions and beyond, to exhibit increasingly sophisticated intelligent behavior.

A fundamental question is then, *how so many categories may be acquired?* New categories are to join an operational (working) system, and furthermore to increase the system's overall capability. Categories inter-relate in a number of ways and can be complex. For example, categories can be compositions of other categories. Thus, the *face* category in the visual domain is a composition of several subcategories such as the *eye* and the *nose* categories. Categories enjoy various relations (spatial, temporal, functional, ..). These dependencies present opportunities as well as considerable difficulties. For example, such relations may help in easing or accelerating categorization. However, apriori, it is very hard to anticipate and program them. In particular, it is not even clear whether manual encoding of such regularities in order to build an operational system is possible. There are numerous domains that are unfamiliar to humans to varying degrees, and yet obtaining and use of appropriate categories in such domains would be useful.

These considerations motivate exploring unsupervised, long-term, and massive learning approaches, for the goals of achieving the scale of intelligence, in particular sophisticated pattern recognition and generation, exhibited by higher animals. We are exploring a framework in which a *prediction system*, endowed with proper algorithmic processes, engages in a game of prediction with the stream of bits that it is fed from its rich world. Example such worlds include text (pages available on the world wide web) and audio, video, and complex sensory streams.

Prediction Games

The abstraction is a single-person game: a system playing the game in its world (so-called "a game against nature"). The world provides an unlimited stream of bits. The goal of the system by playing the game is to improve its predictions, as we describe below. A picture of the interaction between the system and the world is shown in Figure 1. The world is composed of rich regularities, in terms of categories[1]. Categories are simply recurring patterns (of bits) with structure. The system breaks small segments of its stream into its current categories, and then hides a category and tries to predict what it is from the surrounding categories. We will use prediction over a practically unlimited stream of text (*e.g.*, web pages) as our main source of examples.

In the beginning the system can only detect and predict *low level* (or "wired") categories. In text, these can be single characters. However, processes including *composition* and *grouping* within the system, build larger and more complex categories from existing categories. Thus, short two and three character sequences, and later words and phrases become categories (formed via composition). The *digit* ($\{0, 1, 2, \cdots\}$) and the *day-of-week* ($\{Mon, Tue, \cdots\}$) categories are examples of grouping. *Higher level* categories, such as *phone number* and *resume*, are built from a series of grouping and composition operations. A *phone number* is a composition of digits, and a *resume* is a loose composition of contact-information, education, skills, and so on, and each of these subcategories may in-turn be composite. Thus high level categories exhibit sophisticated recursive structure or regularities.

The Approach

Our conjecture is that the goal of improving predictions can drive the learning processes of composing and grouping to create new categories, and of course, learning prediction weights among the categories. In turn, composing and grouping improve prediction performance. The possible importance of predictions, for instance in the workings

[1] Or the system "sees" it that way: the system repeatedly discretizes or digitizes its world into categories.

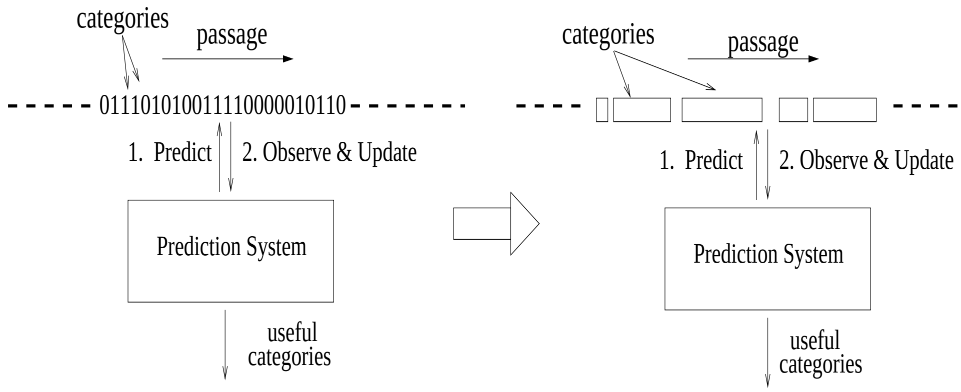

Figure 1: The world as an infinite stream of bits. The prediction system begins by "seeing" the world at low level categories (recurring patterns), 0's and 1's here. With experience (learning), the prediction system sees the world as a stream of "bigger" categories (bigger chunks) as well, and may play the game at multiple levels. Seeing means predicting the next category and then (often) determining whether the prediction was true by observation (matching) and updates.

of the brain, has been much high lighted before (*e.g.*, (Ballard 2000; Hawkins & Blaskeslee 2004)).

We also expect that these functionalities require highly scalable learning, and thus space and time efficient, and noise tolerant algorithms are important. Reasons for requiring scalability include the number of useful categories to acquire as well as the challenges of uncertainty and drift. The learning processes should remain robust for long durations as the learning will be sequential, cumulative, and long-term. Due to these considerations, we expect that a *systems approach* is most suitable in thinking about the solution. The system is composed of multiple algorithmic process that should work in concert. The system will be functioning primarily online, in order to be able to consume the data necessary for the learning.

There are numerous challenges in achieving this functionality (Madani 2007b), but important observations and conjectures that make us optimistic include:

1. Rich worlds provide ample experience sufficient for massive learning (of millions of categories and their relations).
2. Rich worlds enjoy the type of simplifying regularities that allow for efficient learning.

We expect that the hierarchical properties and near decomposability of complex categories (Callebaut & Rasskin-Gutman 2005; Simon 1996) may allow for a type of efficient learning that is in part "bottom up". Much has been learned from machine learning research to date. This should provide some guidance, in terms of what may be feasible, and what task formulations to avoid. However, we also expect that a number of new problems are to be formulated and effectively tackled, in designing and understanding prediction systems. We believe that building and exploring such systems to better expose the issues should be very fruitful at this point. In our recent work, we describe a system that achieves some of the functionality, and further discuss current research directions (Madani 2007a).

Potential applications are numerous: a prediction system, after a sufficient learning period (in order of days to years, in human terms), becomes a highly capable pattern recognition and generation machine. Thus, notable potentials include extending the reach of current language modeling techniques (Rosenfeld 2000), achieving general visual object recognition (Forsyth & Ponce 2003), and generating intericate physical motion. Higher intelligence such as sophisticated reasoning may also rest heavily on repeated use of such inductively-obtained subsystems. We expect that the problems formulated and solved will also shed insight on the learning that is taking place in higher animals during their development.

References

Ballard, D. H. 2000. *An Introduction to Natural Computation*. The MIT Press.

Callebaut, W., and Rasskin-Gutman, D., eds. 2005. *Modularity: Understanding the Development and Evolution of Natural Complex Systems*. The MIT Press.

Forsyth, D. A., and Ponce, J. 2003. *Computer Vision*. Prentice Hall.

Hawkins, J., and Blaskeslee, S. 2004. *On Intelligence: How a New Understanding of the Brain will lead to Truly Intelligent Machines*. Owl Books.

Lakoff, G., and Johnson, M. 1999. *Philosophy in the Flesh: the Embodied Mind and Its Challenge to Western Thought*. Basic Books.

Madani, O. 2007a. Exploring massive learning via a prediction system. In *AAAI Fall Symposium Series*,.

Madani, O. 2007b. Prediction games in infinitely rich worlds. Technical report, Yahoo! Research.

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

Rosenfeld, R. 2000. Two decades of statistical language modeling: Where do we go from here? *IEEE* 88(8).

Simon, H. A. 1996. *The Sciences of the Artificial*. The MIT Press, third edition.