Two eras in learning theory: implications for cognitively faithful models of language acquisition and change

Partha Niyogi (niyogi@cs.chicago.edu)
Department of Computer Science & Department of Statistics, 1100 E. 58th Street
Hyde Park, Chicago, IL 60637 USA

Robert C. Berwick (berwick@csail.mit.edu)
Department of EECS and Brain and Cognitive Sciences, MIT, 32D-728, 77 Mass Ave.
Cambridge, MA 02139 USA

Introduction

Formal approaches to language acquisition may be delimited by roughly two twenty-year periods. The first, from the seminal work of Gold to the mid 1980s, focused on language learning using the techniques of recursive function theory. The second, commencing roughly from the work of Valiant on PAC-learning from the mid 1980s to this day, shifted the focus from *effective* to *efficient* computability, echoing the shift in modern computer science from computability to complexity theory, as well as to more cognitively relevant assumption such as inexact learnability and sample size complexity. While these previous approaches have provided useful insights into the possibilities of automated discovery procedures for cognitive learners, they have a key limitation: they have all been formulated in a setting where there is but a *single* target grammar/language that the learner attempts to acquire. Only over the last decade has a new class of learning models been developed that attempts to explicitly confront the more cognitively faithful reality that the learner is situated in a heterogeneous population with potentially many grammars being simultaneously used by the members of a population. In this paper we demonstrate that this explicit shift is *essential* for accurate cognitive modeling, in that it leads to the possibility of dynamical system bifurcations that are otherwise not mathematically possible. We exhibit new results of how to actually estimate the modeling parameters for such systems, grounded on empirical corpus work. If this population view is not adopted, in effect retaining a 'single learner' position, as in all the other simulation-based methods known to us, e.g., Kirby, Dowman, & Griffiths (2007) then these other methods are mathematically reducible to Markov chains that cannot exhibit the observed bifurcations, or more generally the requisite dynamical system complexities.

New Results for Learnability Theories

Until now there have been no previous studies that have actually estimated the parameters of the dynamical systems from historical copora to verify whether the attested patterns of change are indeed those that are predicted by the theoretical account. Over the past year, we have obtained statistics from the Penn Helsinki corpus of Old and Middle English that allow us to estimate historical parameter values. Concretely, we study the competition between two grammatical systems (one primarily verb-final (OV-type) and the other verb-initial (VO-type)) in Middle English.

In this setting there are two grammatical systems with extensional expressions given by L_1 and L_2. Speakers of g_1 produce expressions with probability P_1 over L_1 and g_2 speakers, expressions with probability P_2 over L_2. Define parameter $a = P_1(L_1 \cap L_2)$ and $b = P_2(L_1 \cap L_2)$. Thus a and b are the probabilities with which speakers of pure g_1 and g_2 produce "ambiguous" expressions If x_t is the proportion of g_1-type grammars in the tth generation, then it is possible to show that for a variety of learning algorithms that individual learners might use, $x_{t+1} = \frac{(1-a)x_t}{(1-a)x_t+(1-b)(1-x_t)}$. This has bifurcations as $a - b$ changes continuously. Predictively, we estimate a and b at a single point in time, and then $a - b$ to *predict* which grammatical type will dominate in successive generations. Given data from a mixture distribution $P = xP_1 + (1 - x)P_2$, can we estimate a and b? One can collect data from the Penn-Helsinki corpus by sampling a few individuals at the same point in time. This is nontrivial because we have access only to the surface forms writers' expressions and cannot always uniquely decode the underlying grammatical system. However, it can be overcome by 'tying' parameters in a particular way. Most remarkably, this novel estimation procedure allows us for the first time to test the predictions of a class of models of language change with data from historical corpora, and validates the essential need for a population view of language acquisition, evolution, and change.

References

M. Nowak, N. K., & Niyogi, P. (2001). Evolution of universal grammar. *Science*, *291*, 114-118.

M. Nowak, N. K., & Niyogi, P. (2002). Computational and evolutionary aspects of language. *Nature*, *417*, 611-617.

Niyogi, P. (2006). *The computational nature of language learning and evolution*. Cambridge, MA: MIT Press.

Niyogi, P., & Berwick, R. (1997). A dynamical systems model for language change. *Journal of Complex Systems*, *11*, 161-204.

S. Kirby, M. D., & Griffiths, T. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Science*, *104*, 5241-5245.