Title
Two Eras in Learning Theory: Implications for Cognitively Faithful Models of Language Acquisition and Change

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 30(30)

ISSN
1069-7977

Authors
Niyogi, Partha
Berwick, Robert C.

Publication Date
2008

Peer reviewed

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

Abstract

We review recent advances towards more cognitively-faithful models of language acquisition and change that parallel conceptual shifts in computational learning theory, and how these new models can yield improved empirical accounts in actual corpus case studies of English historical language change.

Introduction

Formal approaches to language acquisition fall roughly into two historical periods. The first, dating from Gold to the mid 1980s, focused on language learning using recursive function theory techniques. The second, dating from Valiant's PAC-learning model to this day, shifted the focus from *effective* to *efficient* computability, echoing computer science's shift from computability to complexity theory. While these advances moved to more cognitively faithful assumptions — inexact learnability and learnability relative to sample size complexity — and provided useful insights, they retained a key cognitive limitation: a *single* target grammar/language. Over the last decade, a new class of learning models has been developed (Niyogi & Berwick, 1997) explicitly embracing the cognitive reality that learners are situated in heterogeneous populations, with potentially many grammars. This viewpoint, "Social Learning," embraces the more fully Darwinian picture of variation across *both* parental and offspring generations. However, if one restricts oneself instead to a narrower single parent–single learner setting, as in many simulation-based methods e.g., the "Iterated Learning" model of Kirby, Dowman, & Griffiths (2007), the resulting systems reduce to Markov chains. These frameworks cannot exhibit certain empirically observed phase transitions, which demand nonlinear dynamics.

New Results for Learnability Theories

Importantly, modeling based on this shift to a more cognitively-faithful picture yields improved empirical predictions. First, historically attested phase-transitions in the evolution of English, as outlined in Lightfoot (1999), are better described. Furthermore, until now there have been no previous studies that have actually estimated from historical corpora the parameters of the dynamical systems corresponding to such models, in order to verify whether the attested patterns of change are indeed those predicted by the theoretical accounts. We have now obtained statistics from the Penn-Helsinki corpus of Old and Middle English that permits estimation of historical parameter values. Specifically, we have analyzed the competition between two grammatical systems in Middle English, (one primarily verb-final (OV-type) and the other verb-initial (VO-type)).

In this setting we assume two grammars with corresponding languages L_1 and L_2. g_1 speakers produce expressions with probability P_1 over L_1 and g_2 speakers, probability P_2 over L_2. Parameter $a = P_1(L_1 \cap L_2)$ and $b = P_2(L_1 \cap L_2)$. 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: $x_{t+1} = \frac{(1-a)x_t}{(1-a)x_t+(1-b)(1-x_t)}$. This has bifurcations as $a-b$ continuously changes. We estimate a and b at a single time point, using $a-b$ to *predict* which grammatical type dominates in successive generations. However, given data from a mixture distribution $P = xP_1 + (1-x)P_2$, can we even estimate a and b? Yes: we collect data from the Penn-Helsinki corpus by sampling a few individuals at the same time point. This is nontrivial, because only surface forms of writers' expressions are available; one cannot always uniquely decode underlying grammars. We overcome this by "tying" parameters in a novel way. Importantly, this new estimation procedure permits empirical tests of this class of models for language change using data from historical corpora for the first time, and again validates the need for a fullly *population* view of language acquisition, evolution, and change.

References

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

Lightfoot, D. (1999). *The development of language*. Malden, MA: Blackwells Publishers.

Niyogi, P. (2004). Phase transitions in language evolution. In L. Jenkins (Ed.), *Variation and universals in biolinguistics.* New York: Elsevier.

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.