STOCHASTIC, REWARD-BASED LEARNING OF HIDDEN STRUCTURE IN PHONOLOGY

GAJA JAROSZ

Hidden structure, including prosodic structure and underlying representations, is a major outstanding problem in phonological learning. Most existing approaches to hidden structure build on error-driven learning algorithms such as the Gradual Learning Algorithm (Boersma 1997) and Error-Driven Constraint Demotion (Tesar 1995). This paper tackles the hidden structure problem from a different perspective, an underexplored alternative relying on reward-based learning strategies rooted in the machine learning literature (Jarosz 2006). A novel grammatical representation, Stochastic Partial Orders (SPOs), enables efficient estimation of reward-based learning updates, and two novel learning algorithms are presented and tested on two types of hidden structure. The learners are successful, with performance on the metrical test set surpassing the best-performing error-driven learners.

(1) illustrates the challenge of learning in the context of hidden structure. Upon observing [σˈσσ], the error-driven learner uses its current grammar to generate candidate (d) as an output for /σσσ/. This is an error so the learner must compare its output to the winner to determine how the constraint ranking should be updated, but the datum is structurally ambiguous between multiple winners, each with different ranking implications. To deal with hidden structure, error-driven learners must be equipped with some mechanism for parsing overt forms so that their full structural descriptions can be compared with the learner's outputs (Tesar & Smolensky 2000; Boersma & Pater to appear; Jarosz 2013; Biro 2012).

(1) Ambiguity created by hidden structure, after observing learning datum: [σˈσσ]

	/σσσ/	ALLFEETRIGHT	IAMBIC	TROCHAIC	ALLFEETLEFT
	a. (ˈσσ)σ	*	*		
Winner?	b. (σˈσ)σ	*		*	
Winner?	c. σ(ˈσσ)		*		*
Output	d. σ(σˈσ)			*	*

Reward-based learning does not require parsing of ambiguous learning data. Instead, learning relies on a single strategy for both ambiguous and unambiguous data: for each datum the learner determines which relative rankings are favored and disfavored by the datum and makes an update to reward rankings favored by the datum. Specifically, a SPO grammar is represented in terms of pairwise ranking probabilities (e.g. P(A≫B), P(A≫C), etc.), and it is these parameters that are updated during learning. Formally, preference for a pairwise ranking A≫B is defined as the conditional probability of A≫B given the overt datum *d* and current grammar *g*, as shown in (2). Intuitively, this expression indicates how strongly the datum favors one relative ranking over the other. The proposed algorithms rely on Bayes' Law to transform this expression into terms that are easily calculated: the pairwise ranking probabilities and conditional likelihood:

(2) $$\Pr(\mathrm{A} \gg \mathrm{B}|d, g) = \frac{\Pr(d|\mathrm{A} \gg \mathrm{B},\ g)\Pr(\mathrm{A} \gg \mathrm{B}|\ g)}{\Pr(d|\mathrm{A} \gg \mathrm{B},\ g)\Pr(\mathrm{A} \gg \mathrm{B}|\ g) + \Pr(d|\mathrm{B} \gg \mathrm{A},\ g)\Pr(\mathrm{B} \gg \mathrm{A}|\ g)}$$

The paper proposes two stochastic learning algorithms, one batch and one online, which use the same method to estimate the quantity in (2). The pairwise ranking probabilities – $\Pr(\mathrm{B} \gg \mathrm{A}|\ g)$ and $\Pr(\mathrm{A} \gg \mathrm{B}|\ g)$ – are simply the parameters of the SPO grammar and can be accessed directly. To estimate the conditional likelihood terms – $\Pr(d|\mathrm{A} \gg \mathrm{B}, g)$ and $\Pr(d|\mathrm{B} \gg \mathrm{A}, g)$ – the algorithms use a constrained sampling approach. Specifically, the algorithms generate a small sample to estimate the probability with which the observed overt form is generated using the current grammar *except with A ≫ B temporarily fixed to 1*. Intuitively, this yields the likelihood of observing this overt form if A is (categorically) ranked above B but the other constraints are ranked stochastically according to the current grammar. The opposite relative ranking (B ≫ A) is then temporarily fixed to estimate the likelihood for the reverse parameter. This is done independently for all pairs of constraints to determine the conditional likelihood given each parameter setting. Once these terms are estimated, the batch update to the grammar is trivial: the new parameters are proportional to a frequency-weighted sum of (2) over the data. The online version makes small updates in this direction after processing each datum individually.

One set of simulations investigates the algorithms' performance on a large test set of metrical structure that has been used to test a number of error-driven approaches to hidden structure (Tesar & Smolensky 2000; Boersma & Pater to appear; Jarosz 2013; Biro 2012). The batch and online versions achieve success rates of 96.9% and 95.3%, respectively, surpassing the best-performing error-driven learners on this test set. The second simulation demonstrates the generality of the reward-based approach by extending the learning algorithms to underlying representations and demonstrating their success on a constructed language with contrastive stress and length (Tesar 2006).

It is worth highlighting how these reward-based learners deal with hidden structure. Note that none of the calculations performed during learning require parsing of the learning data. In fact, hidden structure is irrelevant to these learners since the estimation of (2) is concerned only with the likelihood of generating the observed *overt* forms. Learning is driven entirely by the overt patterns in the learning data: no special mechanisms are needed to deal with hidden structure. While error-driven learning requires specialized approaches to deal with each kind of hidden structure (structural ambiguity, underlying representations, serial derivations), reward-based learning offers a general approach.

SELECT REFERENCES

Biró, Tamás. to appear. Towards a Robuster Interpretive Parsing, Learning from overt forms in Optimality Theory. *Journal of Logic, Language and Information.*

Boersma, Paul. 1997. How we learn variation, optionality, and probability. *IFA Proceedings* 21. University of Amsterdam. 43-58.

Boersma, Paul and Pater, Joe. to appear. Convergence Properties of a Gradual Learning Algorithm for Harmonic Grammar. In John McCarthy and Joe Pater (editor.), *Harmonic Grammar and Harmonic Serialism*. London: Equinox Press.

Jarosz, Gaja. 2006. Rich Lexicons and Restrictive Grammars - Maximum Likelihood Learning in Optimality Theory. PhD Dissertation, the Johns Hopkins University, Baltimore, MD.

Jarosz, Gaja. 2013. Learning with Hidden Structure in Optimality Theory and Harmonic Grammar: Beyond Robust Interpretive Parsing. *Phonology* 30(1). Cambridge University Press. 27-71.

Tesar, Bruce. 1995. Computational Optimality Theory. PhD Dissertation, University of Colorado, Boulder, CO.

Tesar, Bruce and Smolensky, Paul. 2000. *Learnability in Optimality Theory*. Cambridge, Massachusetts: MIT Press.