Proceedings of the Society for Computation in Linguistics

Volume 2 Article 60

2019

How the Structure of the Constraint Space Enables Learning

Jane Chandlee
Haverford College, jchandlee@haverford.edu

Remi Eyraud
remi.eyraud@lis-lab.fr

Jeffrey Heinz
Stony Brook University, jeffrey.heinz@stonybrook.edu

Adam Jardine
Rutgers University, adam.jardine@rutgers.edu

Jonathan Rawski
jonathan.rawski@stonybrook.edu

How the Structure of the Constraint Space Enables Learning

Jane Chandlee
Haverford College
jchandlee@haverford.edu

Rémi Eyraud
QARMA team, LIS
remi.eyraud@lis-lab.fr

Jeffrey Heinz
Stony Brook University
jeffrey.heinz@stonybrook.edu

Adam Jardine
Rutgers University
adam.jardine@rutgers.edu

Jonathan Rawski
Stony Brook University
jonathan.rawski@stonybrook.edu

Keywords: phonology, learning algorithms, constraint induction, relational learning

Overview

There is significant diversity of opinion among phonologists regarding how constraint-based generalizations can be learned from examples. The maximum entropy approach to constraint induction (Hayes and Wilson 2008) leverages statistical information to select from a flat set of constraints which represent illicit combinations of phonological features. However, recent work on model-theoretic phonology shows that particular phonological representations and relations play a large role in learning properties of well-formed structures (Heinz 2010a,b; Vu et al. 2018). Wilson and Gallagher (2018) advocate a MaxEnt learner, expressing skepticism that learning constraints represented with features can occur without statistics: "Could there be a non-statistical model [like Heinz's] that learns by memorizing feature sequences? The problem confronting such a model is that any given segment sequence has many different featural representations." This is a form of the problem of learning hidden structure: which of the many feature combinations is responsible?

In this paper we answer their question in the affirmative, by presenting a non-statistical learning method that provably identifies the responsible constraints without relying on memorization. Instead, adapting ideas from relational learning (De Raedt 2008), we identify a generality relation that partially orders the space of possible constraints into sets of ideals, which are lower, directed sets. Formally: $x \in L$ implies $y \in L$ for every $y \leq x$, and for all $x, y \in L$ there is some z such that $x \leq z$ and $y \leq z$. The ordering could straightforwardly correspond to the substring relation, but can also encapsulate other relation like the subsequence one, enabling long distance phenomena.

This ordering enables a bottom-up learning mechanism which successively builds more complex structures and checks their validity as constraints against the sample. A boost in efficiency is gained since whenever a structure S is deemed valid as a constraint, the generality relation can be used to prune the remaining space so no structures strictly more complex than S are considered.

An Intuitive Example

Here we present a simple example to illustrate the main idea, using constraints of length one and three binary features: (N)asal, (V)oiced, and (C)onsonantal. These features form a partial order under the *contains* relation: the feature combination [-N,-V,+C] contains [-N,-V], [-N,+C], [-V,+C], [-N], [-V], [+C], and the empty feature set [] (Fig. 1). Suppose the learner observes data that reflect the constraints *[+N,-V] (no voiceless nasals), *[+N,-C] (no nasal vowels), and *[-V,-C] (no voiceless vowels), and that it sees examples containing the combinations [+N,+V,+C] (voiced nasal consonants), [-N,+V,+C] (voiced non-nasal consonants), [-N,-V,+C] (voiceless non-nasal consonants), and [-N,+V,-C] (voiced non-nasal vowels), and no others. Since the observed feature combinations are licit, the feature combinations they contain are also licit and can themselves also be removed from consideration as constraints (shown in blue in Fig. 1). The smallest structures that remain (shown in red in Fig. 1) are then hypothesized as the active constraints — in this example, *[+N,-V], *[+N,-C], and *[-V,-C].

A Bottom-Up Learning Algorithm

Given a set of observed data D and a maximum constraint length k, the learner returns a set of constraints G. It builds a set of candidate constraints incrementally by adding either one valued feature, or increasing the size of the sequence domain by one at a time. For example, starting from the empty structure [], the first candidates

Proceedings of the Society for Computation in Linguistics (SCiL) 2019, pages 387-388.
New York City, New York, January 3-6, 2019

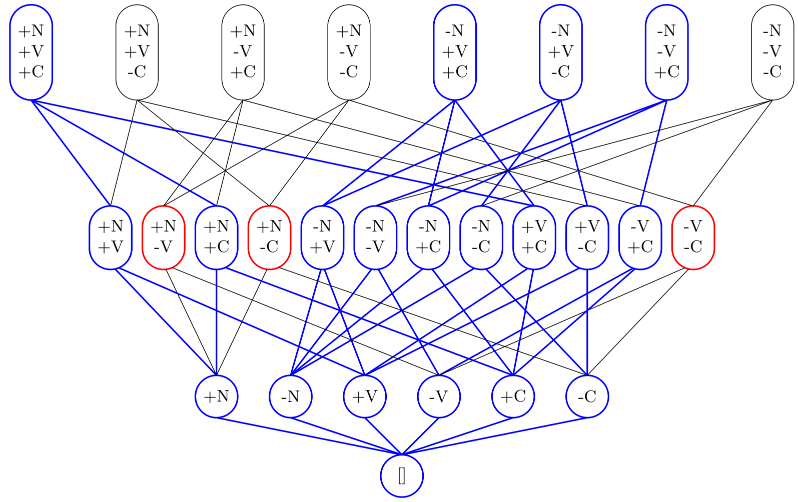

Figure 1: Partially ordered set of ideals

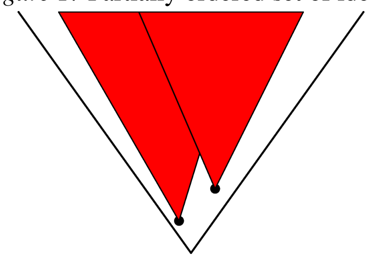

Figure 2: Pruning the constraint space

considered might be [+N], [-N], [+V], [-V], etc. If the candidate constraint is contained by an example in *D*, it is discarded. Otherwise, the candidate is kept and is also used to prune the hypothesis space: any "higher" candidate constraints in the set of ideals (i.e., any candidate constraints that themselves contain the kept constraint, as shown with the red cones in Fig. 2) are NOT considered as the learner explores upward. In this way the learner exploits the *contains* relation to keep the otherwise exponentially massive hypothesis space manageable.

We prove that this algorithm satisfies the following quality criteria (De Raedt 2008):

1. The largest constraint in the returned set of constraints G is of size k;
2. G *covers* the data D, meaning $D \subseteq L(G)$, where $L(G)$ is the language generated by G;
3. G is more specific than all the other hypotheses G' that also cover the data;
4. G forbids structures S that are substructures of structures S' forbidden by other grammars G' that also satisfy (1) and (2), i.e. for all $S' \in G'$, there exists $S \in G$ such that S' contains S.

Conclusion

Feature-based phonological constraints are naturally structured in a way that provides a generality relation which learners can use to effectively prune and search the hypothesis space. Exploiting the structural generalizations present in the hypothesis space enables a learner to identify feature-based constraints without relying on statistical regularities.

Luc De Raedt. *Logical and Relational Learning*. Springer-Verlag Berlin Heidelberg, 2008.

Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39:379–440, 2008.

Jeffrey Heinz. Learning long-distance phonotactics. *Linguistic Inquiry*, 41(4):623–661, 2010a.

Jeffrey Heinz. String extension learning. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 897–906, Uppsala, Sweden, July 2010b. ACL.

Mai Ha Vu, Ashkan Zehfroosh, Kristina Strother-Garcia, Michael Sebok, Herbert G. Tanner, and Jeffrey Heinz. Statistical relational learning with unconventional string models. *Frontiers in Robotics and AI*, July 2018.

Colin Wilson and Gillian Gallagher. Constraint complexity in surface-based phonotactics: A case study of South Bolivian Quechua. *Linguistic Inquiry*, 2018.