How the Constraint Space Structure Enables Learning

Jane Chandlee, Rémi Eyraud, Jeffrey Heinz, Adam Jardine, Jonathan Rawski
Haverford College, Aix-Marseille Université, Stony Brook University, Rutgers University

Overview

How is the hypothesis space of a phonological learner structured, and how does the learner use this structure to generalize from examples? Recent work on model-theoretic phonology shows that particular phonological representations and relations play a large role in learning properties of well-formed structures. Here we:

- Extend current model-theoretic accounts of phonology to unconventional string models incorporating fratalional information;
- Show how features structure the hypothesis space into ideals and filters;
- Describe a non-statistical, non-enumerative learning algorithm that provably learns the most general constraints over features consistent with the data;
- Its efficiency and integration with statistical models is focus of current research.

Model-Theoretic Phonological Representations

A model theory for words defines a class of relational structures. Here we:

1. How is the hypothesis space of a phonological learner structured, and how does the learner use this structure to generalize from examples?
2. Recent work on model-theoretic phonology shows that particular phonological representations and relations play a large role in learning properties of well-formed structures.

Feature Ideals and Grammatical Entailments

Let \( S \) and \( T \) be segments represented as bundles of n-ary features. Then \( T \) is an feature extension of \( S \) for grammar \( G \) (\( S \sqsubset T \)) iff \( T \) is the result of inserting one or more n-ary features of \( G \) in \( S \).

Feature Ideals: If \( T \) is a feature extension of \( S \) for \( G \) and \( G \) generates \( T \), then \( G \) generates \( S \).

Feature Filters: If \( T \) is a feature extension of \( S \) for \( G \) and \( G \) forbids \( S \), then \( G \) forbids \( T \).

Organizing the hypothesis space into sets of ideals and principal filters allows the learner to exploit these grammatical entailments they provide.

Example: Banning Singular Segments

Suppose the learning data consists of:

- \([N,N,V,C]\) (voiced nasal consonants),
- \([N,N,-V,C]\) (voiced nonnal consonants),
- \([-N,N,V,C]\) (voiceless nonnal consonants),
- \([-N,N,-V,C]\) (voiceless vowel consonants).

What constraints ought to be posited?

Positing \([-N,N,V]\) (voiceless nasals), \([-N,N,C]\) (nasal vowels), and \([-N,V,C]\) (voiceless vowels) accounts for the absence of the four unobserved feature combinations with fewer constraints.

Example: Aari Long distance sibilant harmony

In Aari, all sibilants agree in anteriority.

- 1) ha’er ‘he brought’
- 2) ga’er ‘I arrived’
- \( G = \{ [\{'p\}', \{'v\}', \{'s\}'] \} \)

Summary of Learning Guarantees

Given a finite positive data sample, the bottom-up learner finds a constraint grammar \( G \) such that:

1. The largest forbidden substructure is of size \( k \).
2. \( G \) is consistent, i.e. it covers the data:
- \( D \subseteq L(G) \)
3. \( L(G) \) is the smallest language in \( L \) which covers the data:
- For all \( L \in L \) where \( D \subseteq L \), \( L(G) \subseteq L \).
4. \( G \) includes structures \( S \) that are restrictions of structures \( S' \) included in other grammars \( G' \) that also satisfy (1,2,3).
- For all \( S' \subseteq G' \), there exists \( S \in G \) such that \( S \subseteq S' \).

Statistics and Structure

- Structured Hypothesis Spaces allow for correct generalization.
- Hayes and Wilson are right to have a generality relation in their MaxEnt Learner, but why not use the ordering it gives?
- What is the efficiency tradeoff between statistics and structure?
- Is there a constraint learner which can allow this structure?

References