The Computational Complexity of Distinctive Feature Minimization in Phonology

Hubie Chen and Mans Hulden

https://github.com/mhulden/minphonfeat

University of Colorado **Boulder**

Birkbeck UNIVERSITY OF LONDON

Distinctive Features

- Phonologists treat distinctive features as fundamental atomic primitives in describing a sound or a set of sounds.
- For example, to describe a regressive nasal assimilation rule for vowels ([+syl]), phonologists would postulate something like:
- [+syl] → [+nas] / _ [+nas]
- These feature descriptions are not unique, we could also say the following with the same effect:
- [+syl] → [+nas] / _ [+cons, +son, +voi, -cnt, -hi, -bk]
- As a learning and generalization problem, it is assumed that a learner/phonologist uses the **minimum required number** of features to describe the relevant set (Fant, 1966; Halle, 1962 Archangeli, 1984; Hayes, 2011; Zsiga, 2012).
- If a learner has observed {m,n} being relevant to a generalization (as in the previous), phonologists tacitly assume this minimization problem is automatically solved behind the scenes

Example Feature System

	cons	son	syl	voi	cnt	nas	lat	ant	cor	hi	bk	lo	rd
p	+	-	-	-	-	-	-	+	-	-	-	-	-
t	+	-	-	-	-	-	-	+	+	-	-	-	-
k	+	-	-	-	-	-	-	-	-	+	+	-	-
b	+	-	-	+	-	-	-	+	-	-	-	-	-
d	+	-	-	+	-	-	-	+	+	-	-	-	-
g	+	-	-	+	-	-	-	-	-	+	+	-	-
f	+	-	-	-	+	-	-	+	-	-	-	-	-
s	+	-	-	-	+	-	-	+	+	-	-	-	-
x	+	-	-	-	+	-	-	-	-	+	+	-	-
v	+	-	-	+	+	-	-	+	-	-	-	-	-
ɣ	+	-	-	+	+	-	-	-	-	+	+	-	-
w	-	+	-	+	+	-	-	-	-	+	+	-	+
j	-	+	-	+	+	-	-	-	-	+	-	-	-
l	+	+	-	+	+	-	+	+	+	-	-	-	-
m	+	+	-	+	-	+	-	+	-	-	-	-	-
n	+	+	-	+	-	+	-	+	+	-	-	-	-
a	-	+	+	+	+	-	-	-	-	-	+	+	-
e	-	+	+	+	+	-	-	-	-	-	+	-	-
i	-	+	+	+	+	-	-	-	-	+	-	-	-
o	-	+	+	+	+	-	-	-	-	-	+	-	+
u	-	+	+	+	+	-	-	-	-	+	+	-	+
y	-	+	+	+	+	-	-	-	-	+	-	-	+

Problems

- We analyze the computational complexity of two separate problems related to feature descriptions
  - **The feature description problem**
  - **The feature minimization problem**

• Feature Description

- Given a set of phonemes, what is the complexity of deciding whether that set is a *natural class*? = is definable by some combination of features
- For example, for the feature system on the left **{k,b}** is NOT a *natural class*, while **{k,p}** is.

• Feature Minimization

- Given a set of phonemes Q, what is the complexity of of finding the minimal description for it
- For example, on the right is shown the minimal description of the set **{v,ɣ}** = [-son, +voi, +cnt]

Example minimization

	cons	son	syl	voi	cnt	nas	lat	ant	cor	hi	bk	lo	rd
p	+	-	-	-	-	-	-	+	-	-	-	-	-
t	+	-	-	-	-	-	-	+	+	-	-	-	-
k	+	-	-	-	-	-	-	-	-	+	+	-	-
b	+	-	-	+	-	-	-	+	-	-	-	-	-
d	+	-	-	+	-	-	-	+	+	-	-	-	-
g	+	-	-	+	-	-	-	-	-	+	+	-	-
f	+	-	-	-	+	-	-	+	-	-	-	-	-
s	+	-	-	-	+	-	-	+	+	-	-	-	-
x	+	-	-	-	+	-	-	-	-	+	+	-	-
v	+	-	-	+	+	-	-	+	-	-	-	-	-
ɣ	+	-	-	+	+	-	-	-	-	+	+	-	-
w	-	+	-	+	+	-	-	-	-	+	+	-	+
j	-	+	-	+	+	-	-	-	-	+	-	-	-
l	+	+	-	+	+	-	+	+	+	-	-	-	-
m	+	+	-	+	-	+	-	+	-	-	-	-	-
n	+	+	-	+	-	+	-	+	+	-	-	-	-
a	-	+	+	+	+	-	-	-	-	-	+	+	-
e	-	+	+	+	+	-	-	-	-	-	+	-	-
i	-	+	+	+	+	-	-	-	-	+	-	-	-
o	-	+	+	+	+	-	-	-	-	-	+	-	+
u	-	+	+	+	+	-	-	-	-	+	+	-	+
y	-	+	+	+	+	-	-	-	-	+	-	-	+

Quiz Time!

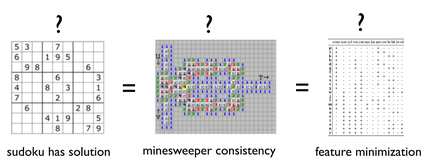
- To appreciate the potential difficulty of these questions, consider this pop quiz designed to take you back mentally to your favorite phonology class:
- (1) Can the set **{b,g}** be described exactly with the feature system below?
- (2) If so, what is the minimal description (using the smallest number of features)?

	cons	son	syl	voi	cnt	nas	lat	ant	cor	hi	bk	lo	rd
p	+	-	-	-	-	-	-	+	-	-	-	-	-
t	+	-	-	-	-	-	-	+	+	-	-	-	-
k	+	-	-	-	-	-	-	-	-	+	+	-	-
b	+	-	-	+	-	-	-	+	-	-	-	-	-
d	+	-	-	+	-	-	-	+	+	-	-	-	-
g	+	-	-	+	-	-	-	-	-	+	+	-	-
f	+	-	-	-	+	-	-	+	-	-	-	-	-
s	+	-	-	-	+	-	-	+	+	-	-	-	-
x	+	-	-	-	+	-	-	-	-	+	+	-	-
v	+	-	-	+	+	-	-	+	-	-	-	-	-
ɣ	+	-	-	+	+	-	-	-	-	+	+	-	-
w	-	+	-	+	+	-	-	-	-	+	+	-	+
j	-	+	-	+	+	-	-	-	-	+	-	-	-
l	+	+	-	+	+	-	+	+	+	-	-	-	-
m	+	+	-	+	-	+	-	+	-	-	-	-	-
n	+	+	-	+	-	+	-	+	+	-	-	-	-
a	-	+	+	+	+	-	-	-	-	-	+	+	-
e	-	+	+	+	+	-	-	-	-	-	+	-	-
i	-	+	+	+	+	-	-	-	-	+	-	-	-
o	-	+	+	+	+	-	-	-	-	-	+	-	+
u	-	+	+	+	+	-	-	-	-	+	+	-	+
y	-	+	+	+	+	-	-	-	-	+	-	-	+

{b,g}?

TL;DR

- The Feature description problem **is in P**
  - Very efficiently checkable whether a set of phonemes Q can be described with some feature system F
- The decision version of the Feature Minimization Problem **is NP-complete**
  - "Decision version" meaning, answer the question: "can phoneme set Q be described with k features or less?"
- The minimization version **is NP-hard**
  - "Minimization" being: find the minimal set of features needed to describe a set of phonemes Q
- We show this by reduction to set cover (Karp, 1972)
- In fact, we can reduce both to and from set covering to show the feature minimization problem is isomorphic to set cover



Main Reduction

- We reduce from set cover
- A set cover is sequence of sets S_1, …, S_m drawn from **S,** such that $S_1 \cup \ldots \cup S_m = U$ (the universe)
- Decide whether there exists a set cover of less than or equal to some k
- Example, **S** = {{a,b,c}, {b,d}, {c,d}, {d,e}}
- Decision Question Example: does a set cover exist if k = 2 (or less)?
- Answer: **yes** (since {a,b,c} ∪ {d,e} = {a,b,c,d,e} = U)
- Main trick: given an instance of set cover, convert sets to a feature system which is the negation of each set, and add a fresh element *x* to each feature
- Solve minphonfeat for the phoneme set Q = {x}
- Reduces to and from set cover without changing k

SETCOVERING

S = {{a,b,c},{b,d},{c,d},{d,e}}, U = {a,b,c,d,e}, k = 2

MINPHONFEAT

P = U ∪ {x} Q = {x} k = 2

F_1 = {d,e,x}
F_2 = {a,c,e,x}
F_3 = {a,b,e,x}
F_4 = {a,b,c,x}

"yes" [+F_1 +F_4] = Q

Discussion

- Assuming minimal feature discovery as part of a phonological learning component or model is problematic
- Could some fast strategy be used which would work on phoneme systems because they exhibit a special structure?
  - We've ruled out a **greedy strategy** (sometimes fails in actual attested feature systems)
  - Branch&Bound type search strategies work but the search space is still huge
- Or should the interaction between observed sounds "phonemes" and learned feature combination be loosened?

Check Quiz Answer Here!

(1) YES, **{b,g}** is a natural class!
(2) The minimal description is:

[+voi, -cnt, -nas, -cor]
or
[-son, +voi, -cnt, -cor]