Computational Models of Phonotactics

Student: Scarlett Xu, Honours Cognitive Science; Supervisor: Timothy O'Donnell

McGill University, Montreal Computational Quantitative Linguistics Lab

Abstract

- The adoption of words is governed by systematic intuitions on the likelihood of different combinations of sounds in a language. For example, between two hypothetical English words "blick" and "bnick", "blick" has a higher likelihood to be accepted as an English word by a native speaker (Chomsky and Halle, 1965). Understanding such constraints on allowable sound sequences is crucial to understanding language productivity and second language acquisition.
- Three classes of models of phonotactics are being compared: maximum entropy models, Bayesian models, and neural-network-based models.
- Given the scores from all three models, we perform statistical analysis, such as computing representativenss, to help us better decide what words are most likely to be discriminated against by all three models.

Maximum Entropy Model (BLICK)

Its grammars consist of constraints that are assigned numerical weights according to the principle of maximum entropy. Possible words are assessed by these grammars based on the weighted sum of their constraint violations.

- The score of a phonological representation x, denoted h(x), is:

$$h(x) = \sum_{i=1}^{k} w_i C_i(x) \quad (1)$$

where $w_i$is the weight of the ith constraint, C_i is the number of times that x violates the ith constraint, and $\Sigma_{i=1}^{k}$ denotes summation over all constraints.

- To get the probability P(x), we need to get the maxent value P*(x) first:

$$P*(x) = exp(-h(x)) \quad (2)$$

$$P(x) = \frac{P*(x)}{Z} \quad (3)$$

Neural Network Models (Ryan)

- Neural sequence to sequence models have been used to capture phonotactic generalizations.
- The distinction between core features and combination features: A core feature refers to a single indicator, such as whether a certain phoneme is present; a combination feature describes a conjunction of features occurring together.

Bayesian Models (Richard)

- Unlike most compu- tational models of phonotactics, this model takes a fully generative approach, modeling a process where forms are built up out of subparts by phonologically-informed structure building operations.
- We learn an inventory of subparts by applying stochastic memoization (Johnson et al., 2006; Goodman et al., 2008) to a generative process for phonemes structured as an and-or graph, based on concepts of feature hierarchy from generative phonology (Clements, 1985; Dresher, 2009)

Representativeness

To choose the stimuli that best characterizes the models, we used a measure of **representativeness**. To compute $R(d, h_i)$ in the presence of more than one alternative hypothesis, we express it in the form:

$$R(d, h_i) = log \frac{P(d|h_i)}{\Sigma_{h_j \in \mathbf{H}} P(d|h_j)P(h_j|h_i)} \quad (4)$$

It shows that d is representative of h_i to the extent that its likelihood under hi exceeds its average likelihood under alternative hypotheses. In particular, since we only have three hypotheses (models) here, the prior probability of three hypotheses are all equal to $\frac{1}{3}$, and the equation (4) can be specified as below for the BLICK model:

$$\log \frac{2P(d, BLICK)}{P(d, Ryan)P(Ryan)P(d, Richard)P(Richard)} \quad (5)$$

Examples of Scores and Representativeness

ARPAbet	BLICK (repr)	Ryan (repr)	Richard (repr)
SH1 W AO1 R T S	9.628 (-10.371)	1.888 (6.526)	-15.564 (-4.553)
K L IH1 NG K S	1.432 (2.226)	1.980 (-0.304)	-15.518 (-5.467)
B R AE1 N T S	3.786 (-2.080)	2.077 (-0.547)	-11.935 (1.826)

Table 1:Improvement DL over random sampling

Top 15 words sorted by representativenss of BLICK, Ryan, Richard model

C	D				
IPA	arpa	IPA	arpa	IPA	arpa
sklɪrkst	S K L IH1 R K S T	kloʊðdz	K L OW1 DH D Z	hændsts	HH AE1 N D S T S
splɪrŋ	S P L IH1 R NG	kloʊðzz	K L OW1 DH Z Z	strændzdz	S T R AE1 N D Z D Z
sklɪrpst	S K L IH1 R P S T	tʃeɪndʒz	CH EY1 N JH Z	frændz	F R AE1 N D Z
splɪrkst	S P L IH1 R K S T	loʊðz	L OW1 DH Z	harldz	HH AA1 R L D Z
splɪrkt	S P L IH1 R K T	loʊðdz	L OW1 DH D Z	barldz	B AA1 R L D Z
splɪrpst	S P L IH1 R P S T	kloʊðzd	K L OW1 DH Z D	streɪtz	S T R EY1 T Z
sklældʒ	S K L AE1 L JH	θorndθs	TH AO1 R N D TH	smʌndz	S M AH1 N D Z
sklɪrkt	S K L IH1 R K T	θwortsz	TH W AO1 R T S Z	krʌndz	K R AH1 N D Z
sklɪrnst	S K L IH1 R N S T	farðz	F AA1 R DH Z	kwandz	K W AA1 N D Z
sprɪrkst	S P R IH1 R K S T	smaɪðz	S M AY1 DH Z	mardz	M AA1 R D Z
sklɪrndʒ	S K L IH1 R N JH	mænðz	M AE1 N DH Z	grændsts	G R AE1 N D S T S
sprɪrkt	S P R IH1 R K T	kloʊðts	K L OW1 DH T S	trændz	T R AE1 N D Z
sprɪrpst	S P R IH1 R P S T	dʒordʒz	JH AO1 R JH Z	swɛlvz	S W EH1 L V Z
splældʒ	S P L AE1 L JH	ʃwortsz	SH1 W AO1 R T S	prɛndz	P R EH1 N D Z
splɪrnst	S P L IH1 R N S T	mjuŋz	M Y UW1 NG Z	prændz	P R AE1 N D Z

Discussion & Future Work

- There might be sets of grammar rules involved human's phonological process when they are deciding how acceptable a form is. If each model consists a set of grammar, their judgements on a stimulus might involve an intersection of the sets of grammar, and it would be a challenge to precisely identify which set of grammars truly accounts for a particular sequence combination. The challenges of the future works will mainly lie in understanding the results in terms of what grammars the models succeed in capturing and what patterns the models fail to account for.

- LSTM LM is not as sensitive to the length, and it is a great improve- ment and some of its highest ranked stimuli have up to 7 phonemes. Neural networks indeed produce the state-of-art performance in modeling phonotactics, but the results are less interpretable compared to other models that draw more inspirations from linguistic theories.

1. Programming of the Experiment: Selecting representative stimulus and building up the behaviour test experiment to collect human judgements on Amazon Mechanical Turk with the jsPsych library.

ACKNOWLEDGEMENT

Thank you to Professor O'Donnell for this opportunity and for his guidance throughout this internship. I am also grateful for all the help and guidance from another lab member, Vanna Willerton. I greatly appreciate the assistance and funding provided to me by the Arts Internship Office.

Related Work

- Futrell, R., Albright, A., Graff, P., and O'Donnell, T. J. (2017). *A generative model of phonotactics.*
- Bruce Hayes and Colin Wilson. 2008. *A maximum entropy model of phonotactics and phonotactic learning.*
- Yoav Goldberg. 2017. *Neural Network Methods for Natural Language Processing Ch. 3, 4, 5 ,9*
- Pimentel, T., Roark, B., AND Cotterell, R. *Phonotactic complexity and its trade-offs.*