Phone2Vec

Michael Hammond. U. of Arizona

A vector space model learns major phonetic dimensions, but exhibits marginal performance with other dimensions.

Overview

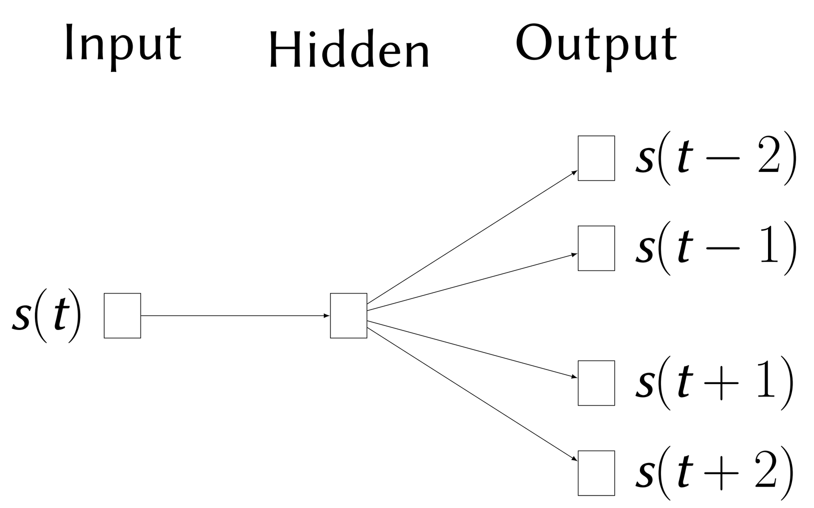
- Can a generic word embedding model (Mikolov et al., 2013a,b,c) discover phonological features?
- Phonotactic restrictions can be expressed in featural terms. For example, while an obstruent consonant can follow a sonorant consonant at the end of an English word, the opposite cannot occur. Thus:

	Possible	Impossible
stamp	[stæmp]	*[...pm]
tent	[tʰɛnt]	*[...tn]
tank	[tʰæŋk]	*[...kŋ]
barf	[barf]	*[...fr]
tilt	[tʰɪlt]	*[...tl]

- Are such phonotactic distributions are *sufficient* to learn the featural distinctions of a language like English?

Methods

- We build vector models from the CMU pronouncing dictionary (Weide, 1998) and then test them with *vector similarity* and *vector dissimilarity* tasks.
- We use the *skipgram* architecture of Mikolov et al. (2013a), Mikolov et al. (2013b), and Mikolov et al. (2013c). This is an older, efficient, and simple system that is implemented in the python `gensim` module (Řehůřek and Sojka, 2010).
- We train a neural net to predict adjacent sounds in a word from each sound (2 sounds on each side). This procedure creates an internal numerical representation of each sound that reflects the contexts it occurs in.



- Embeddings are created at the bottleneck hidden layer. We decide how big those embeddings are and, in our experiments, we make them very small, only eight numbers each.
- Neural nets start with random initial values and results vary depending on these. We therefore run our model 100 times to control for this.

Tests

- With those 100 models, we test *vector similarity* and *vector dissimilarity* for contrastive phonetic dimensions of English.
- Similarity example for vowels: [á, è, i, ó, ù]: across all models, the segment most similar is [ì] (highest average cosine similarity across all models).
- Dissimilarity example for consonants: [p, m, s, a]: the least similar is [a] in all 100 models.

Major class and consonant results

Vowel results in full paper; errors here indicated with question marks.

Dissimilarity

Category/category	Class	Best	Models	
consonants/vowel	[p, m, s, a]	a	100	
vowels/consonant	[a, o, e, s]	s	100	
primaries/secondary	[í, æ̀, ɔ́j, ʊ́, áj]	æ̀	100	
secondaries/primary	[ì, ǽ, ɔ̀j, ʊ̀, àj]	ǽ	100	
secondaries/stressless	[ì, o, ɔ̀j, ʊ̀, àj]	o	100	
stressless/secondary	[i, æ̀, o, ʊ, ə]	æ̀	99	
obstruants/nasal	[p, s, d, v, m]	v	91	?
obstruants/liquid	[p, s, d, v, l]	v	90	?
obstruants/glide	[p, s, d, v, j]	j	100	
nasals/obstruant	[m, n, ŋ, v]	ŋ	100	?
nasals/liquid	[m, n, ŋ, l]	ŋ	100	?
nasals/glide	[m, n, ŋ, j]	j	94	
liquids/obstruant	[l, r, z]	z	83	
liquids/nasal	[l, r, n]	r	80	?
liquids/glide	[l, r, j]	j	100	
glides/obstruant	[w, j, z]	j	99	?
glides/nasal	[w, j, n]	j	98	?
glides/liquid	[w, j, l]	j	73	?
stops/fricative	[p, t, d, g, f]	f	51	!
stops/affricate	[p, t, d, g, č]	č	100	
fricatives/stop	[f, v, s, ð, t]	ð	99	?
fricatives/affricate	[f, v, s, θ, č]	v	79	?
affricates/stop	[č, ǰ, k]	k	95	
affricates/fricative	[č, ǰ, v]	v	95	
voiceless/voiced	[p, t, s, θ, g]	s	77	?
voiced/voiceless	[b, d, z, ð, k]	ð	98	?

Full paper with vowel results, references, and code

`https://github.com/hammondm/phone2vec`

Similarity

Category	Class	Best	
vowels	[á, è, i, ó, ù]	ì	
consonants	[p, m, s, b, l]	d	
primary	[í, ǽ, ɔ́j, ʊ́, áj]	ɛ́	
secondary	[ì, æ̀, ɔ̀j, ʊ̀, àj]	aw	!
stressless	[i, ə, ʊ]	ɛ	!
obstruants	[p, d, g, s, č]	t	
nasals	[m, n]	s	?
	[m, ŋ]	l	?
	[n, ŋ]	l	?
liquids	[l]	s	?
	[r]	l	
glides	[j]	š	?
	[w]	f	?
stops	[p, t, b, d, g]	m	?
fricatives	[f, v, s, θ, ð, š, ž]	d	?
affricates	[č]	ǰ	
voiceless	[p, t, s, θ, š]	d	?
voiced	[b, d, z, ð, ž]	g	

Discussion

- Stress and consonant/vowel distinctions emerge, but other distinctions are less clear.
- We conclude that these other distinctions are either not visible in the phonotactics or are not *sufficiently* visible there.
- The distinction between consonants and vowels is essential to the distribution of stress, so this may contribute to its visibility.
- Another factor that might play a role is the size of the context window: two segments on each side. A context like that is necessary for stress distinctions, but might be a distraction for segmental phonotactics.
- Phonotactic support for the other distinctions is clear in English (Hammond, 1999), but statistically quite infrequent.
- There has been previous work in this area, e.g. Goldsmith and Xanthos (2009), Silfverberg et al. (2018), Mayer (2018), and Kolachina and Magyar (2019). None of these look specifically at which classes in English are learnable in terms of what we know from traditional phonology.
- We also tried GloVe embeddings (Pennington et al., 2014), but these did not perform as well.
- One might consider BERT-style embeddings (Devlin et al., 2018), but these are context-dependent and thus not suitable for our tests.