Distributed word representations: Matrix designs

Christopher Potts

Stanford Linguistics

CS224u: Natural language understanding

word x word

	:)	:/	:D	:|	;p	abandon	abc	ability	able	···
:)	74	1	0	0	0	1	0	2	2	
:/	1	306	0	0	0	0	0	0	17	
:D	0	0	16	0	0	0	6	1	1	
:|	0	0	0	120	0	0	0	1	9	
;p	0	0	0	0	516286	0	0	0	0	···
abandon	1	0	0	0	0	370	24	65	235	
abc	0	0	6	0	0	24	7948	77	291	
ability	2	0	1	1	0	65	77	4820	1807	
able	2	17	1	9	0	235	291	1807	14328	
⋮					⋮					

word x document

	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
against	0	0	0	1	0	0	3	2	3	0
age	0	0	0	1	0	3	1	0	4	0
agent	0	0	0	0	0	0	0	0	0	0
ages	0	0	0	0	0	2	0	0	0	0
ago	0	0	0	2	0	0	0	0	3	0
agree	0	1	0	0	0	0	0	0	0	0
ahead	0	0	0	1	0	0	0	0	0	0
ain't	0	0	0	0	0	0	0	0	0	0
air	0	0	0	0	0	0	0	0	0	0
aka	0	0	0	1	0	0	0	0	0	0

word x discourse context

Upper left corner of an interjection × dialog-act tag matrix derived from the Switchboard Dialog Act Corpus:

	Reject-part	Hedge	Completion	Tag question	Hold	Quotation	Accept	…
absolutely	0	2	0	0	0	0	95	
actually	17	12	0	0	1	0	4	
anyway	23	14	0	0	0	0	0	
boy	5	3	1	0	5	2	1	
bye	0	1	0	0	0	0	0	
bye-bye	0	0	0	0	0	0	0	…
dear	0	0	0	0	1	0	0	
definitely	0	2	0	0	0	0	56	
exactly	2	6	1	0	0	0	294	
gee	0	3	0	0	2	1	1	
goodness	1	0	0	0	2	0	0	
⋮				⋮				

Other designs

- adj. × modified noun
- word × syntactic context
- word × search query
- person × product
- word × person
- word × word × pattern
- verb × subject × object

⋮

Feature representations of data

- *the movie was horrible* becomes $[4, 0, 1/4]$.
- The complex, real-world response of an experimental subject to a particular example becomes $[0, 1]$ or $[118, 1]$.
- A human is modeled as a vector $[24, 140, 5, 12]$.
- A continuous, noisy speech stream is reduced to a restricted set of acoustic features.

Windows and scaling: What is a co-occurrence?

Windows and scaling: What is a co-occurrence?

from	swerve	of	shore	to	bend	of	bay	,	brings
4	3	2	1	0	1	2	3	4	5

Windows and scaling: What is a co-occurrence?

from	swerve	of	shore	to	bend	of	bay	,	brings
4	3	2	1	0	1	2	3	4	5

	from	swerve	of	shore	to	bend	of	bay	,	brings
Window: 3	4	3	2	1	0	1	2	3	4	5
Scaling: flat	0	1	1	1	1	1	1	1	0	0

Windows and scaling: What is a co-occurrence?

from	swerve	of	shore	to	bend	of	bay	,	brings
4	3	2	1	0	1	2	3	4	5

	from	swerve	of	shore	to	bend	of	bay	,	brings
Window: 3	4	3	2	1	0	1	2	3	4	5
Scaling: flat	0	1	1	1	1	1	1	1	0	0
Scaling: $\frac{1}{n}$	0	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{1}$	1	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	0	0

Windows and scaling: What is a co-occurrence?

from	swerve	of	shore	to	bend	of	bay	,	brings
4	3	2	1	0	1	2	3	4	5

- Larger, flatter windows capture more semantic information.
- Small, more scaled windows capture more syntactic (collocational) information.
- Textual boundaries can be separately controlled; core unit as the sentence/paragraph/document will have major consequences.

Code snippets

```
import os
import pandas as pd

DATA_HOME = os.path.join('data', 'vsmdata')

# Yelp: Window size = 5; scaling = 1/n
yelp5 = pd.read_csv(
    os.path.join(DATA_HOME, 'yelp_window5-scaled.csv.gz'), index_col=0)

# Yelp: Window size = 20; scaling = flat
yelp20 = pd.read_csv(
    os.path.join(DATA_HOME, 'yelp_window20-flat.csv.gz'), index_col=0)

# Gigaword: Window size = 5; scaling = 1/n
giga5 = pd.read_csv(
    os.path.join(DATA_HOME, 'giga_window5-scaled.csv.gz'), index_col=0)

# Gigaword: Window size = 20; scaling = flat
giga20 = pd.read_csv(
    os.path.join(DATA_HOME, 'giga_window20-flat.csv.gz'), index_col=0)
```