Distributional Models for Lexical Semantics

Lecture 1: Don't be afraid of vectors

Denis Paperno TyLex – Voronovo 2017

Non-technical definition

- Distributional semantic models
- are ways of creating lexical semantic representations
- through learning by association
- on a large scale

Distributional Semantic Models are vector spaces built from distributional information

Vector: $(a_1, a_2, ..., a_n)$, where each a_i is an element of *F*.

(where F is a field with multiplication and addition operations which generalize to respective operat

ions on vectors via standard axioms)

Origins

- It may be presumed that any two morphemes A and B having different meanings, also differ somewhere in distribution: there are some environments in which one occurs and the other does not
 (Harris 1951)
- "You shall know a word by the company it keeps." (Firth 1957)

Example

he curtains open and the moon shining in on the barely ars and the cold , close moon " . And neither of the w rough the night with the moon shining so brightly , it made in the light of the moon . It all boils down , wr surely under a crescent moon , thrilled by ice-white sun, the seasons of the moon ? Home, alone, Jay pla m is dazzling snow , the moon has risen full and cold un and the temple of the moon , driving out of the hug in the dark and now the moon rises , full and amber a bird on the shape of the moon over the trees in front But I could n't see the moon or the stars , only the rning , with a sliver of moon hanging among the stars they love the sun , the moon and the stars . None of the light of an enormous moon . The plash of flowing w man 's first step on the moon ; various exhibits , aer the inevitable piece of moon rock . Housing The Airsh oud obscured part of the moon . The Allied guns behind

Co-occurrence vector space

- Simplest distributional model: just count!
 - Vectors represent words
 - Dimensions represent contexts

	bark	walk	talk	tail	bag
dog	246	72	78	71	1
cat	5	15	25	32	0
man	0	57	133	0	1
woman	2	203	407	5	18

Vectors as word representations



Vector similarity measures

$$\cos(a,b) = \frac{\sum_{i} a_i * b_i}{\sqrt{\sum_{i} a_i^2} * (\sum_{i} b_i^2)}$$

$$euc(a,b) = \sqrt{\sum_i (a_i - b_i)^2}$$

Vectors as word representations



Cosine values

- 1 for identical vectors
- 0 for orthogonal vectors
- Negative values rare for linguistic vectors







 $a_i * b_i$ $\cos(a,b) =$ $\sum_{i} a_i^2 * \left(\sum_{i} b_i^2 \right)$

cos(cat,man)=

5*0+15*57+25*133+32*0+0*1

$$\sqrt{(5^2+15^2+25^2+32^2+0^2)(0^2+57^2+133^2+0^2+1^2)}$$

cat	5	15	25	32	0
man	0	57	133	0	1



 $a_i * b_i$ $\cos(a,b) =$ $\sum_{i} a_i^2 * \left(\sum_{i} b_i^2 \right)$

• cos(cat,man)= 5*0+15*57+25*133+32*0+0*1

$$\sqrt{(5^2+15^2+25^2+32^2+0^2)(0^2+57^2+133^2+0^2+1^2)}$$

=4180/√1899*20939≈.66

cat	5	15	25	32	0
man	0	57	133	0	1

Application: similarity and relatedness

Examples from WordSim-353:

Word pair	relatedness	cosine	
money~cash	9.15	.98	
tiger~zoo	5.87	.42	
stock~phone	1.62	.04	

Similarity and relatedness datasets exist of other languages, including Russian (Panchenko et al. 2016)



the, (comma)beownerwalkdog517646283195245237

Parameters of DSM

weighting

$$PMI(w,c) = \log \frac{\hat{P}(w,c)}{\hat{P}(w) \cdot \hat{P}_{\alpha}(c)}$$

$$PPMI(w,c) = \max(\log \frac{P(w,c)}{P(w)P(c)}, 0)$$

$SPPMI_{k}(w,c) = \max(PMI(w,c) - \log k, 0)$

PMI Weighting

	the	, (comma)	be	owner	walk
dog	5176	4628	3195	245	237

	the	, (comma)	be	owner	walk
dog	1.6	1.52	1.56	3.05	2.73

Application: Free Association

• What associations do you have when you hear the word PEN?

Free association

Human associations:

- PEN
 - PENCIL
 - INK
 - PAPER
 - WRITE

Free association

Human associations: Distributional model:

- PEN
 - PENCIL
 - INK
 - PAPER
 - WRITE

- PEN
 - PENCIL
 - FOUNTAIN
 - INK
 - PAPER
 - WRITE

Example from Griffiths et al. 2007, p.223

Application: semantic proportions

- man:king=woman:x
 x≈king+woman-man
- Additive method: (cos(x,king)+cos(x,woman)-cos(x,man))
- Multiplicative method: cos(x,king)*cos(x,woman)/cos(x,man) (Levy and Goldberg 2014)

Application: semantic proportions

- Works well for some relations (capital~country, gender) Russia:Moscow=Latvia:x
 ^{1. рига 0.67}
- Less well for others (currency~country, adverb~adjective)
 Russia:ruble=Latvia:x

Illustrations from rusvectores.ru

1. рига 0.67
2. вильнюс 0.64
3. таллин 0.60
4. даугавпилс 0.58
5. каунас 0.56

1. копейка 0.51
2. руб 0.50
3. лат 0.48
4. злотый 0.47
5. риксдалер 0.46

Application: pragmatic alternatives

Sentences of the form

- This is not an X, it is a Y.
- There is no X here but there is a Y.

Examples:

- This is not an alligator, this is a crocodile.
- There is no garlic here but there is a vampire.

Cosine (garlic,vampire) has high correlation with the plausibility of the sentence:

- .86 for THIS IS,
- .89 for THERE IS (Kruszewski et al. 2016)

Toward a new application

- Assume similarity and relatedness solved; a general semantic task that further tests semantic models?
- Idea: test for semantic differences
- Semeval 2018 Task 10 https://competitions.codalab.org/competitions/17326

Word 1	Word 2	feature	difference
dolphin	seal	(has) fins	YES
dolphin	seal	(eats) fish	NO

Have fun

- Visualize related English words http://www.serelex.org/
 by Aleksandr Panchenko
- Vectors for Russian: http://rusvectores.org by Andrei Kutuzov and Elizaveta Kuzmenko