Computational Linguistics

Distributional Semantics

Clayton Greenberg & Stefan Thater

FR 4.7 Allgemeine Linguistik (Computerlinguistik)
Universität des Saarlandes
Summer 2015

Goal

Goal:

- treat document clustering and word clustering on the same footing (same semantic space)
- find low dimensional representations

From Frequency to Meaning:
Vector Space Models of semantics

Based on a paper by Turney and Pantel; Journal of Artificial Intelligence Research, 2010, page 141

Advantages

- Derive semantics from corpus
- Good automatic coverage of two main types of lexical properties:
  - attributional similarity, e.g. how similar are “dog” and “cat”?
  - relational similarity, e.g. “dog” : “tail” :: “car” : ?

Similarities

- Document-Document

  -> build a term document matrix

  -> Calculate distance between vectors representing documents directly

  -> Use LSA, …

Similarities

- Document-Document

  -> build a term document matrix

  -> Calculate distance between vectors representing documents directly

  -> Use LSA, …

Alternative: word-context matrix (context being phrase, sentence, paragraph, …)

Assumption: words that occur in similar contexts have similar meaning

Similarity of Relations

Row:

pairs of words (e.g. mason:stone)

Column:

patterns:

X cuts Y

X works with Y

…

Patterns establish relations between words

Measure similarity between patterns

Similarity of Relations

Table 3. The top-20 most similar paths to "X solves Y".

Y is solved by X	Y is resolved in X
X resolves Y	Y is solved through X
X finds a solution to Y	X rectifies Y
X tries to solve Y	X copes with Y
X deals with Y	X overcomes Y
Y is resolved by X	X eases Y
X addresses Y	X tackles Y
X seeks a solution to Y	X alleviates Y
X do something about Y	X corrects Y
X solution to Y	X is a solution to Y

DIRT – Discovery of Inference Rules from Text

Dekang Lin and Patrick Pantel
University of Alberta
Department of Computing Science
Edmonton, Alberta T6H 2E1 Canada
{lindek, ppantel}@cs.ualberta.ca

Similarity of Pairs of Words

Stem:		mason:stone
Choices:	(a)	*teacher:chalk*
	(b)	*carpenter:wood*
	(c)	*soldier:gun*
	(d)	*photograph:camera*
	(e)	*book:word*

Similarity of Semantic Relations

Peter D. Turney *
National Research Council Canada

Similarity of Pairs of Words

Table 10
The sixteen combinations and their cosines. A:B::C:D expresses the analogy "A is to B as C is to D". The third column indicates those combinations for which the cosine is greater than or equal to the cosine of the original analogy, quart:volume::mile:distance.

Word pairs	Cosine	Cosine $\geq$ original pairs
quart:volume::mile:distance	0.525	yes (original pairs)
quart:volume::feet:distance	0.464	
quart:volume::mile:length	0.634	yes
quart:volume::length:distance	0.499	
liter:volume::mile:distance	0.736	yes
liter:volume::feet:distance	0.687	yes
liter:volume::mile:length	0.745	yes
liter:volume::length:distance	0.576	yes
gallon:volume::mile:distance	0.763	yes
gallon:volume::feet:distance	0.710	yes
gallon:volume::mile:length	0.781	yes (highest cosine)
gallon:volume::length:distance	0.615	yes
pumping:volume::mile:distance	0.412	
pumping:volume::feet:distance	0.439	
pumping:volume::mile:length	0.446	
pumping:volume::length:distance	0.491	

Similarity of Semantic Relations

Peter D. Turney *
National Research Council Canada

The big playground

- Pick:
  - Type of context (e.g. document, sentence, pattern, …)
  - Representation (e.g. frequency, tf-idf, …)
  - Way to process the matrix (e.g. original, LSA, NMF, …)
  - Distance metric (Euclidian, cosine, …)

Homework

Gold standard: some human annotations for the similarity of two words:

```
…
tiger	cat	7.35
tiger	jaguar	8.00
tiger	carnivore	7.08
tiger	mammal	6.85
tiger	animal	7.00
tiger	organism	4.77
tiger	fauna	5.62
…
```

- Use Penn-Treebank to build a vector with context words (e.g. one left, one right) containing the frequency (or tf-idf value) for each context word
- Calculate the similarity between the two vectors representing the first and second word (Euclidian or cosine distance)

How good are you able to reproduce the human annotation?

This is a very experimental task!

Vector-based Models of Semantic Composition

Based on a paper by Jeff Mitchell and Mirella Lapata, ACL 2008.

Composition

- Meaning of larger units determined from meaning of smaller units
  - Morphemes
  - Words
  - Phrases
  - Sentences

Approaches

- Logic-based View
  - Write down logical expressions for parts
  - Logical expressions for larger units derived from parts

Limits

- Fully compositional: e.g. “black dog”

vs

- Idioms: “kick the bucket”

Formalizing composition

p=f(u,v,R,K)

What is a good f()?

u: representation of the meaning of the first constituent

v: representation of the meaning of the second constituent

R: syntactic relation

K: Knowledge about the real world

p: meaning of the composition

Vector based approaches

p, u and v are vectors in some semantic space

In particular:

p is in the same space as u and v

Linear functions

- Most general

$$\vec{p} = A\vec{u} + B\vec{v} + \vec{n}$$

with matrices A and B

- Specific versions

$$\vec{p} = \vec{u} + \vec{v}$$ additive

$$\vec{p} = \vec{u} + \vec{v} + \vec{n}$$ "Kintsch"

$$\vec{p} = \alpha\vec{u} + \beta\vec{v}$$ weighted additive

Example

Hypothetical semantic space

	Music	Solution	Economy	Craft	Reasonable
practical	0	6	2	10	4
difficulty	1	8	4	4	0

What would be the “combined semantics” of “practical difficulty” using

- the additive model
- the weighted additive model (using α=0.4 and β=0.6)

Multiplicative combination (bilinear)

- Most general $\vec{p} = C\vec{u}\vec{v}$

C : a rank 3 Tensor

that is $$p_k = \sum_{i,j} C_{k,i,j} u_i v_j$$

- Specific versions

$$p_i = u_i v_i$$ multiplication

$$p_i = \sum_{j=1}^{n} u_j v_{(i-j) \bmod_n}$$ circular convolution

Experiments by Mitchell and Lapata

Collected human similarity ratings for

- adjective-noun
- Noun-noun
- Verb-object

phrases

(e.g. how similar is “professional advice” and “expert opinion”)

Compare to models of compositionality

Results

Model	Adjective-Noun	Noun-Noun	Verb-Object
Additive	0.36	0.39	0.30
Kintsch	0.32	0.22	0.29
Mutliplicative	0.46	0.49	0.37
Tensor Product	0.41	0.36	0.33
Convolution	0.09	0.05	0.10
Weighted Additive	0.44	0.41	0.34
Dilation	0.44	0.41	0.38
Target Unit	0.43	0.34	0.29
Head Only	0.43	0.17	0.24
Humans	0.52	0.49	0.55

Multiplicative combination performs best

But: theoretical foundation needed

Latest trends

- Represent words as matrices (and vectors)
- Combine using neural networks

Summary

The playground of distributional semantics

Compositionality