Determining order restrictions in recursive modification using compositionality in distributional semantics

Eva Maria Vecchi   Roberto Zamparelli   Marco Baroni

Center for Mind/Brain Sciences
Unversity of Trento

MIXEUR 2013

Distributional semantics

Landauer and Dumais 1997, Turney and Pantel 2010, . . .

```
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
```

Distributional Semantic Models (DSMs): Computational models of meaning based on their distribution: their pattern of cooccurrences within a specified context

- i.e., within a document, within a window of [content] words before and after, etc.

	shadow	shine	planet	night
moon	16	29	10	22
sun	15	45	14	10
dog	10	0	0	4

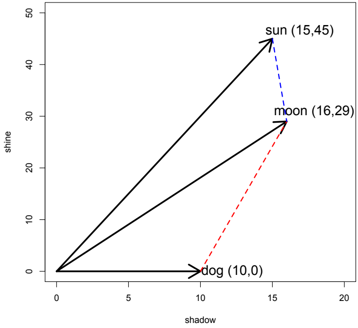
Distributional semantics

The geometry of meaning

Distributional Semantic Models (DSMs): Computational models of meaning based on their distribution: their pattern of cooccurrences within a specified context

- i.e., within a document, within a window of [content] words before and after, etc.

	shadow	shine	planet	night
moon	16	29	10	22
sun	15	45	14	10
dog	10	0	0	4



- semantic similarity approximated by geometric distance of vectors (angle)
- successful in tasks that concern content words: detecting synonyms, lexical entailment, ... (see Turney & Pantel, 2010; Baroni & Lenci, 2010)

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red moon	10.9	2.8	1.1	5.5	0.2	0.3

Compositionality in DSMs

Mitchell and Lapata 2008, 2009, 2010

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red moon	10.9	2.8	1.1	5.5	0.2	0.3
red+moon	39.6	18.9	22.3	27.3	20.3	20.7
red*moon	371.8	56.2	44.2	72.9	22.9	10.1

1. **Additive** (*add*): $\vec{c} = \vec{a} + \vec{n}$
2. **Weighted Additive** (*w.add*): $\vec{c} = \alpha\vec{a} + \beta\vec{n}$
3. **Multiplicative** (*mult*): $\vec{c}_i = \vec{a}_i\vec{n}_i$
4. **Dilation** (*dl*): $\vec{c} = (\vec{a} \cdot \vec{a})\vec{n} + (\lambda - 1)(\vec{a} \cdot \vec{n})\vec{a}$

Compositionality in DSMs

Guevara 2010, Baroni and Zamparelli 2010

5. **Full Additive** (*f.add*): $\vec{c} = \mathbf{W}_1\vec{a} + \mathbf{W}_2\vec{n}$

6. **Lexical Function Model** (*lfm*):

$$\vec{c} = \mathbf{A}\vec{n}$$

$$\begin{pmatrix} c_1 \\ c_2 \\ \dots \\ c_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \times \begin{pmatrix} n_1 \\ n_2 \\ \dots \\ n_m \end{pmatrix}$$

- Success with adjective-noun (AN) phrases in a number of semantic tasks
  - approximation of observed ANs (Baroni & Zamparelli, 2010; Guevara 2010, etc)
  - distinguishing types of modification: intersective, subsective, intensional (Boleda et al., 2012)
  - detecting semantic deviance in novel AN phrases (Vecchi et al., 2011, 2013)

[Recursive] Compositionality in DSMs

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red moon	10.9	2.8	1.1	5.5	0.2	0.3

[Recursive] Compositionality in DSMs

	planet	night	space	color	blood	brown
red	15.3	3.7	2.2	24.3	19.1	20.2
moon	24.3	15.2	20.1	3.0	1.2	0.5
red moon	10.9	2.8	1.1	5.5	0.2	0.3
big red moon	3.2	0.9	1.4	4.7	0.1	0.5

[Recursive] Compositionality in DSMs

1. **Weighted Additive** (*w.add*): $\vec{c} = \alpha\vec{a}_x + \beta(\alpha\vec{a}_y + \beta\vec{n})$
2. **Multiplicative** (*mult*): $\vec{c}_i = \vec{a}_{xi}\vec{a}_{yi}\vec{n}_i$
3. **Full Additive** (*f.add*): $\vec{c} = \mathbf{W}_1\vec{a}_x + \mathbf{W}_1(\mathbf{W}_1\vec{a}_y + \mathbf{W}_2\vec{n})$
4. **Lexical Function Model** (*lfm*): $\vec{c} = \mathbf{A}_x\mathbf{A}_y\vec{n}$

Recursive adjective modification

Flexible Order (FO)

- phrases where both orders, $A_x A_y N$ and $A_y A_x N$, are frequently attested

estimated total population	*total estimated population*
overall good health	*good overall health*

Rigid Order (RO)

- phrases with *one* order, $A_x A_y N$, frequently attested, and $A_y A_x N$ is unattested.

ancient human remains	∗*human ancient remains*
fine young musician	∗*young fine musician*

Unattested Order (UO)

- phrases in which *neither* order, $A_x A_y N$ or $A_y A_x N$, is attested

∗*significant historic year*	∗*historic significant year*
∗*thin correct material*	∗*correct thin material*

Motivation

1. Compositional Distributional Semantics
   - How distributional composition functions behave when applied recursively
   - Can we use measures extracted from the semantic space to distinguish - or even predict - adjective ordering?
2. Theoretical Linguistics
   - Can distributional methods shed some light on the ways in which adjective ordering is constrained in modification?
3. NLP Tasks
   - determine plausibility of possible parses of recursive phrases
   - suggest the optimal ordering in NLG
   - provide insight to the cognitive processing of phrases
   - opens doors between distributional semantics and syntax

Experimental Design

1. Semantic Space

Source Corpus

- about 2.8 billion tokens
  - Web-derived ukWaC corpus (`http://wacky.sslmit.unibo.it`)
  - a mid-2009 dump of the English Wikipedia (`http://en.wikipedia.org`)
  - British National Corpus (`http://www.natcorp.ox.ac.uk/`)
- tokenized, POS-tagged and lemmatized with the TreeTagger (Schmid 1995)
- co-occurrence statistics extracted at the lemma level, no inflectional information

Source Corpus

- about 2.8 billion tokens
  - Web-derived ukWaC corpus (`http://wacky.sslmit.unibo.it`)
  - a mid-2009 dump of the English Wikipedia (`http://en.wikipedia.org`)
  - British National Corpus (`http://www.natcorp.ox.ac.uk/`)
- tokenized, POS-tagged and lemmatized with the TreeTagger (Schmid 1995)
- co-occurrence statistics extracted at the lemma level, no inflectional information

Semantic Space Matrix

- Rows:
  - 8K most frequent Nouns
  - 4K most frequent Adjectives
  - 179K ANs with frequency > 100 in source corpus
- Contexts:
  - 10K most frequently co-occurring Adjectives, Nouns, Verbs, and Adverbs
  - transform raw counts into Positive Pointwise Mutual Information scores
  - reduce to 300 dimensions by Non-negative Matrix Factorization (NMF)

Adjective-Adjective-Noun (AAN) Phrases

- Generate AANs composed of the 700 most frequently As and the 2.5K most frequent Ns
- Split the AANs into the two types of adjective ordering: FO and RO
- Observed vectors (*obs*) are computed for the FO and attested RO AANs

Gold Standard

- 320 AANs with "good" *obs* vectors
- selected after a qualitative analysis of the nearest neighbors (CrowdFlower experiment)

- **Top neighbors**: A qualitative analysis of the top neighbors in the semantic space of the *obs* and predicted AAN vectors
- **Observed vector approximation**: The distance between a predicted AAN vector and its corpus-extracted observed vector
- **Distance from subparts**: The distance relationship between the AAN vectors and each of its subparts. We consider the distance between the AAN vector and
  - the component N vector
  - the component A vectors (*obs*/*add*) or AN centroids (*lfm*)
  - the component AN vectors

Results

1. Top neighbors

medieval old town	*British naval power*
fascinating town impressive cathedral medieval street	naval war British navy naval power
rural poor people	*contemporary political issue*
poor rural people rural infrastructure rural people	cultural topic contemporary debate contemporary politics
friendly helpful staff	*rapid social change*
near hotel helpful staff quick service	social conflict social transition cultural consequence
national daily newspaper	*fresh organic vegetable*
national newspaper major newspaper daily newspaper	organic vegetable organic fruit organic product
creative new idea	*last live performance*
innovative effort creative design dynamic part	final gig live dvd live release

Examples of top neighbors of the gold standard, both FO (left column) and RO (right column) AANs.

Results

2. Vector Approximation

	Gold	*FO*	*RO*	*sig.*
W.ADD	0.565	0.572	0.558	
F.ADD	0.618	0.622	0.614	
MULT	0.424	0.468	0.384	*
LFM	0.655	0.675	0.637	*

Mean distances between the corpus-extracted and model-generated gold AAN vectors. Based on the results of a *t*.test comparing the means of the two classes of ordering, for all significant results, $p<0.001$.

Results

3. Distance from subparts

	Measure	FO_{mean}	RO_{mean}	*sig.*
OBS	$\cos A_x$	0.342	0.298	*
	$\cos A_y$	0.353	0.438	*
	$\cos N$	0.555	0.457	*
	$\cos A_x N$	0.752	0.661	*
	$\cos A_y N$	0.748	0.724	

Results of a *t*.test for corpus-extracted vectors for the gold items.
For all significant results, $p<0.05$.

FO: *rural* *poor* *people*
RO: *rapid* *social* *change*

Results

3. Distance from subparts

	Measure	FO_{mean}	RO_{mean}	*sig.*
OBS	$\cos A_x$	0.342	0.298	*
	$\cos A_y$	0.353	0.438	*
	$\cos N$	0.555	0.457	*
	$\cos A_x N$	0.752	0.661	*
	$\cos A_y N$	0.748	0.724	

Results of a *t*.test for corpus-extracted vectors for the gold items.
For all significant results, $p<0.05$.

FO:	*rural*	*poor*	*people*
RO:	*rapid*	*social*	*change*

	Measure	FO_{mean}	RO_{mean}	*sig.*
OBS	$\cos A_x$	0.342	0.298	*
	$\cos A_y$	0.353	0.438	*
	$\cos N$	0.555	0.457	*
	$\cos A_x N$	0.752	0.661	*
	$\cos A_y N$	0.748	0.724	

Results of a *t*.test for corpus-extracted vectors for the gold items.
For all significant results, $p<0.05$.

FO: *rural* *poor* *people*
RO: *rapid* *social* *change*

Results

3. Composition models: FO vs RO

	Measure	*Estimate*	*sig*	
W.ADD	cosA_x	-11.142	*	FO>RO
	cosA_y	-0.503		
	cosN	-4.210	*	FO>RO
	cosA_xN	1.357		
	cosA_yN	0.475		
F.ADD	cosA_x	-8.259	*	FO>RO
	cosA_y	2.126		
	cosN	-8.029	*	FO>RO
	cosA_xN	18.140	*	RO>FO
	cosA_yN	-8.496	*	FO>RO
MULT	cosA_x	-3.011	*	FO>RO
	cosA_y	2.328	*	RO>FO
	cosN	-3.538	*	FO>RO
	cosA_xN	-0.149		
	cosA_yN	-1.940		
LFM	cosA_x	-6.148	*	FO>RO
	cosA_y	1.988		
	cosN	-7.272	*	FO>RO
	cosA_xN	0.670		
	cosA_yN	-4.142		

Attested-order RO:	*rapid*	*social*	*change*
Unattested-order RO:	*social*	*rapid*	*change*

Results

3. Composition models: Predicting the correct order

	Measure	*Estimate*	*sig.*	
W.ADD	cosA_x	12.297	*	U>RO
	cosA_y	-8.650	*	RO>U
	cosN	-0.343		
	cosA_xN	10.740	*	U>RO
	cosA_yN	-5.418		
F.ADD	cosA_x	17.497	*	U>RO
	cosA_y	-9.558	*	RO>U
	cosN	0.646		
	cosA_xN	-8.525	*	RO>U
	cosA_yN	6.884		U>RO
MULT	cosA_x	5.966	*	U>RO
	cosA_y	-5.573	*	RO>U
	cosN	-0.624		
	cosA_xN	-2.603		
	cosA_yN	3.701	*	U>RO
LFM	cosA_x	22.802	*	U>RO
	cosA_y	-6.916	*	RO>U
	cosN	6.559	*	U>RO
	cosA_xN	5.739		
	cosA_yN	-0.596		

Conclusions

- Corpus-extracted AAN vectors appear to be meaningful, semantically coherent objects
- Composition functions are able to approximate corpus-extracted AANs quite well
- The distance relationship between the AANs and their subparts provides a clear, significant distinction between FO and RO phrases
- Composition models are able to capture the distinction between attested vs. unattested RO AANs
- Given two adjectives and a noun, at random, we can exploit these unsupervised measures to determine the correct adjective ordering

Next steps...

- A more extensive qualitative analysis of the neighborhoods of the model-generated AAN phrases
  - e.g., compare the nearest neighbors between the attested- and unattesetd-order RO AANs to determine if there is a qualitative difference between the two
- Compare these measures also for *unattested* AANs
- Classify types of adjectives and nouns which are found in each ordering and determine if a pattern exists
  - exploit it to improve prediction of adjective ordering.
- *** Can we use these measures (+ plausibility measures[1]) to improve bracketing of NPs?

[1] Vecchi et al., 2011, 2013

That's all, folks!

http://evavecchi.com