Mechanisms of Meaning

Autumn 2010

Raquel Fernández

Institute for Logic, Language & Computation
University of Amsterdam

Plan for Today

- Discussion of HW3: http://staff.science.uva.nl/~raquel/teaching/mom2010/homework/mom2010-hw3.pdf
- Compositionality in Distributional Semantic Models:
  - state-of-the-art models, focusing on subject-verb composition
  - two recent papers papers on adjective-noun composition

  Main references:

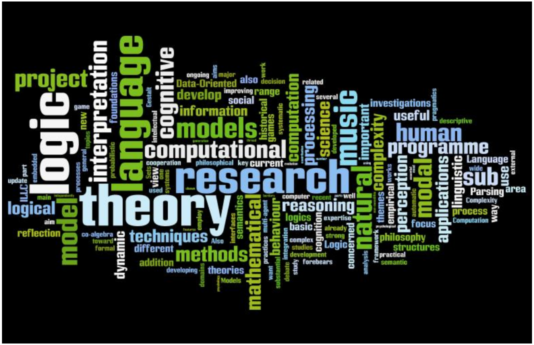
  - Mitchel & Lapata (2008) Vector-based models of semantic composition, *Proceedings of ACL*.
  - Guevara (2010) A Regression Model pf Adjective-Noun Compositionality in Distributional Semantics, *Proceedings of GEMS workshop, ACL*.
  - Baroni & Zamparelli (2010) Nouns are vectors, adjectives are matrices, *Proceedings of EMNLP*.

Aside: a couple of online toys

Two online toys that use word frequencies and distributions:

- Gender Differences in Twitter Messaging:
  http://labs.buradayiz.webfactional.com/gender/query/about
- Nice-looking word clouds: http://www.wordle.net/

  the ILLC as a word cloud:



DSMs and Compositionality

DSMs are interesting candidates for representing meaning, because they are:

- inherently *context-based* and hence *context-dependent*
- inherently *distributed* and *dynamic*
- inherently *quantitative* and *gradual*
- they have been shown to correlate with human linguistic abilities, such as similarity judgements.

However, current DSMs have difficulty accounting for *compositionality*. Can we built compositional distributional models?

Composition Models

General class of *composition models* by Mitchell & Lapata (2008):

$$\mathbf{p} = f(u, v, R, K)$$

- $\mathbf{p}$ denotes the composition of two vectors u and v,
- R stands for the syntactic relation that holds between the constituents represented by u and v, and
- K stands for any additional background knowledge needed.

Most models explored so far: $R =$ subject-verb relation, $K = \emptyset$.

$$\mathbf{p} = f(u, v)$$

Mitchell & Lapata (2008) Vector-based Models of Semantic Composition, *Proceedings of ACL*.

Composition Models

Models differ on the particular function f used for composition:

- *additive* models: $p_i = u_i + v_i$
- *multiplicative* models: $p_i = u_i \cdot v_i$
- symmetry can be relaxed by introducing *weighting constants*, e.g. $p_i = \alpha u_i + \beta v_i$
- more complex models are possible (e.g. tensor product)

Composition Models

Models differ on the particular function f used for composition:

- *additive* models: $p_i = u_i + v_i$
- *multiplicative* models: $p_i = u_i \cdot v_i$
- symmetry can be relaxed by introducing *weighting constants*, e.g. $p_i = \alpha u_i + \beta v_i$
- more complex models are possible (e.g. tensor product)

Hypothetical example from Mitchell & Lapata (2008):

target	*animal*	*stable*	*village*	*gallop*	*jokey*
horse	0	6	2	10	4
run	1	8	4	4	0

- additive model: horse + run = $[1 \;\; 14 \;\; 6 \;\; 14 \;\; 4]$
- multiplicative model: horse $\cdot$ run = $[0 \;\; 48 \;\; 8 \;\; 40 \;\; 0]$
- with weighting constants $\alpha = .4$ and $\beta = .6$:
  horse + run = $[0 \;\; 2.4 \;\; .8 \;\; 4 \;\; 1.6]$ + $[.6 \;\; 4.6 \;\; 2.4 \;\; 2.4 \;\; 0]$ = $[.6 \;\; 5.6 \;\; 3.2 \;\; 6.4 \;\; 1.6]$

Composition Models: Evaluation

Mitchell & Lapata (2008) evaluate several composition models on a *sentence similarity task*:

target sentences	**landmark verbs**
`the horse run`	`gallop`
`the colour run`	`dissolve`

- an appropriate composition model when applied to $\langle \texttt{horse}, \texttt{run} \rangle$ will yield a vector closer to '`gallop`' than to '`dissolve`'.

They found that *multiplicative models* were superior for this task.

Adjective-Noun Composition

Two very recent papers on adjective-noun composition:

- ∗ Guevara (2010) A Regression Model pf Adjective-Noun Compositionality in Distributional Semantics, *Proceedings of GEMS workshop, ACL*.
- ∗ Baroni & Zamparelli (2010) Nouns are vectors, adjectives are matrices, *Proceedings of EMNLP*.

There are two aspects that make them particularly interesting:

- they go beyond subject-verb composition;
- they use new evaluation methods.

$\Rightarrow$ *For the technical details please look at the papers.*

Compositional Semantics of Adjectives

Adjectives are a complex category with a varied semantics. One way to classify them into semantic classes is to consider their *intersectivity* with the noun they combine with (Partee 1995).

Compositional Semantics of Adjectives

Adjectives are a complex category with a varied semantics. One way to classify them into semantic classes is to consider their *intersectivity* with the noun they combine with (Partee 1995).

- *Intersective*: $[\![\mathrm{AN}]\!] = [\![\mathrm{A}]\!] \cap [\![\mathrm{N}]\!]$ e.g. *'vegetarian'*, *'male'*, ...

$$\texttt{vegetarian\_professor}(x) \rightarrow \texttt{vegetarian}(x) \wedge \texttt{professor}(x)$$

Compositional Semantics of Adjectives

Adjectives are a complex category with a varied semantics. One way to classify them into semantic classes is to consider their *intersectivity* with the noun they combine with (Partee 1995).

- *Intersective*: $[\![\text{AN}]\!] = [\![\text{A}]\!] \cap [\![\text{N}]\!]$ e.g. *'vegetarian'*, *'male'*, ...

  $\texttt{vegetarian\_professor}(x) \rightarrow \texttt{vegetarian}(x) \wedge \texttt{professor}(x)$

- *Subsective*: $[\![\text{AN}]\!] \subset [\![\text{N}]\!]$ most adjectives

  $\texttt{small\_whale}(x) \not\rightarrow \texttt{small}(x) \wedge \texttt{whale}(x)$
  'white face', *'white bread'*, *'white wine'*, ...

  They can exhibit different *manners of composition* (Pustejovsky 1995):
  red *'car'* (outside) / *'watermelon'* (inside) / *'traffic light'* (signal)
  easy *'problem'* (solve) / *'language'* (learn) / *'recipe'* (follow/cook)

Compositional Semantics of Adjectives

Adjectives are a complex category with a varied semantics. One way to classify them into semantic classes is to consider their *intersectivity* with the noun they combine with (Partee 1995).

- *Intersective*: $[\![\mathrm{AN}]\!] = [\![\mathrm{A}]\!] \cap [\![\mathrm{N}]\!]$ e.g. *'vegetarian'*, *'male'*, ...

  $\texttt{vegetarian\_professor}(x) \rightarrow \texttt{vegetarian}(x) \wedge \texttt{professor}(x)$

- *Subsective*: $[\![\mathrm{AN}]\!] \subset [\![\mathrm{N}]\!]$ most adjectives

  $\texttt{small\_whale}(x) \not\rightarrow \texttt{small}(x) \wedge \texttt{whale}(x)$
  'white face', *'white bread'*, *'white wine'*, ...

  They can exhibit different *manners of composition* (Pustejovsky 1995):
  red *'car'* (outside) / *'watermelon'* (inside) / *'traffic light'* (signal)
  easy *'problem'* (solve) / *'language'* (learn) / *'recipe'* (follow/cook)

- *Privative*: the rest (not an homogeneous category)

  $\texttt{alleged\_criminal} \not\rightarrow \texttt{criminal}(x)$
  $\texttt{fake\_gun} \rightarrow \neg\texttt{gun}(x)$

  $\texttt{stone\_lion}(x) \rightarrow \neg\texttt{lion}(x)$

Compositional Semantics of Adjectives

⇒ How can the meaning of Adjective-Noun combinations be represented in distributional semantics?

Guevara's approach

To account for the variety of adjectival semantic classes, Guevara assumes a general multiplicative model with weighting constants:

$$\vec{AN} = \alpha\vec{A} \cdot \beta\vec{N}$$

- The weights α and β are estimated directly from data, which allows flexibility to model different semantic relations.
- He uses all data available: $\vec{A}$, $\vec{N}$ and $\vec{AN}$.
- The weights are estimated with a machine learning algorithm (regression), treating the dimensions in $\vec{AN}$ as dependent variables
  - $*$ supervised method but no annotated data needed.
- The *evaluation* consists in comparing the predictions made by the model with the observed $\vec{AN}$ vector.

Guevara (2010) A Regression Model pf Adjective-Noun Compositionality in Distributional Semantics, *Proceedings of GEMS workshop, ACL*.

Baroni & Zamparelli's approach

Baroni & Zamparelli's approach

In formal semantics, Montague proposed to treat all attributive adjectives homogeneously as *functions* of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$.

$$[\![\text{vegetarian}]\!] = \lambda N \lambda x.[N(x) \wedge \texttt{vegetarian}(x)]$$
$$[\![\text{small}]\!] = \lambda N \lambda x.[N(x) \wedge \texttt{size}(x) < \texttt{size}(\texttt{prototype}(N))]$$
$$[\![\text{fake}]\!] = \lambda N \lambda x.[\neg N(x) \wedge \texttt{looks\_like}(x, \texttt{prototype}(N))]$$

B&Z want to model this intuition with the framework of DSMs.

Baroni & Zamparelli's approach

In formal semantics, Montague proposed to treat all attributive adjectives homogeneously as *functions* of type $\langle\langle e,t\rangle,\langle e,t\rangle\rangle$.

$[\![\text{vegetarian}]\!] = \lambda N \lambda x.[N(x) \wedge \texttt{vegetarian}(x)]$
$[\![\text{small}]\!] = \lambda N \lambda x.[N(x) \wedge \texttt{size}(x) < \texttt{size}(\texttt{prototype}(N)]$
$[\![\text{fake}]\!] = \lambda N \lambda x.[\neg N(x) \wedge \texttt{looks\_like}(x, \texttt{prototype}(N))]$

B&Z want to model this intuition with the framework of DSMs.

- The meaning of an adjective A is taken to be the linear mapping between $\vec{N}$ and $A\vec{N}$. Their model is also multiplicative:

$$\alpha\vec{N} = A\vec{N}$$

  where α is matrix of weights that represents the meaning of the adjective.
  - $\vec{N}$ and $A\vec{N}$ are extracted from corpus data;
  - the adjective vector $\vec{A}$ is not used.

Baroni & Zamparelli (2010) Nouns are vectors, adjectives are matrices, *Proceedings of EMNLP*.

Baroni & Zamparelli's approach

The weights in the matrix α are estimated from data using the same method as Guevara: a form of regression.

- For instance:

  - ∗ *'green'* matrix:
    - large positive weights mapping features of concrete $\vec{N}$s to colour dimensions in $\vec{AN}$;
    - large positive weights mapping features of abstract $\vec{N}$s to political/social dimensions in $\vec{AN}$.
  - ∗ *'sofa'*: near-0 values on the relevant abstract dimensions
  - ∗ *'initiative'*: near-0 values on the relevant concrete dimensions

Evaluation: comparison of the predicted $\vec{AN}$ and the observed $\vec{AN}$.

B&Z claim they get better results than Guevara, but they dataset is different: non-trivial comparison. *See papers for details.*

Extensions?

- Explore vector substraction models? $\vec{AN} - \vec{N}$
- Cluster adjectives into semantic classes? B&Z mention two possible methods:
  - ∗ use average (centroid) of $\vec{AN}$ vectors
  - ∗ use predicted matrices α
- What about adjectives in predicative positions?

Extensions?

- Explore vector substraction models? $\vec{AN} - \vec{N}$
- Cluster adjectives into semantic classes? B&Z mention two possible methods:
  - ∗ use average (centroid) of $\vec{AN}$ vectors
  - ∗ use predicted matrices α
- What about adjectives in predicative positions?

⇒ More on recent developments on distributional semantics and compositionality this Wed 17 Nov 4pm at the CL Seminar:

- ∗ Reinhard Blutner will present work by Stephen Clark on this topic. [room D1.110]

What's next?

Up to now: zooming into *word meaning*

- Firs steps in "modern" lexical semantics: generative lexicon
- Excursion into lexicography: Kilgarriff
- Psychological theories of concepts and word meaning
- Distributional semantics

Coming weeks: zooming out to *meaning in interaction*

- Overview of phenomena that characterise language in dialogue
- The role of interaction in language acquisition and development
- Interaction management: turn-taking

Final papers:

- Make an appointment to speak to me this week.