Compositional distributional semantics

Ann Copestake

Computer Laboratory
University of Cambridge

November 2017

Outline

Slides marked (A) from Aurelie Herbelot, (G) from Guy Emerson

Outline.

Background / recap

Distributional semantics for phrases

Distributional semantics for individuals

Functional distributional semantics

Compositional semantics

- Compositional semantics is about providing a meaning representation for an entire sentence.
- Classically (e.g., Montague) based on syntax and morphology, meaning expressed as in logic:

  every white cat is asleep

  $\forall x[[white'(x) \wedge cat'(x)] \rightarrow asleep'(x)]$

- Structures built deterministically from a rich syntactic analysis (quantifier scope possibly underspecified).
- Useful in applications like database interfaces where predicates can be grounded.
- Also used for RTE etc with inference rules.
- Can be automatically induced for limited domains.

Distributional semantics

- Classically, distributional semantics is about providing a meaning representation for words.
- But most approaches capture relatedness rather than genuine similarity: e.g., astronomer is related to telescope.
- Based on context in corpora, always automatically induced.
- Little or no morphology or syntax in most approaches.
- Used in lots of practical applications, starting with IR.
- Word embeddings used in neural models are a form of distributional semantics.

Compositional distributional semantics

Various strands of work:

- Distributional semantics for phrases (most work is really about this).
  - doc2vec: not usually described as compositional distributional semantics but clearly related.
  - Non-compositionality: multiword expressions.
- Theoretical accounts that combine compositional semantics and distributional semantics (not so relevant for this course).
- Distributional semantics for individuals (Herbelot).
- Also 'Functional distributional semantics' (Emerson) and work by Erk, Boleda and others.

Outline.

Background / recap

Distributional semantics for phrases

Distributional semantics for individuals

Functional distributional semantics

Distributional semantics for phrases

- Problem specification?
  - Combine vectors to allow for sparse data: metal spoon vs ebony tripod
  - Capturing ordering effects: man bites dog vs dog bites man
  - Semantics for full sentences: quantifiers, truth?
- Methods: addition, multiplication, more complex functions. Higher-order tensors: e.g., transitive verbs are third order tensors.
- Evaluation
  - Predicting actual phrase distributions
  - Similarity judgements on phrasal test sets
  - Extrinsic evaluation

Formal semantics of adjectives

- white cat: white′$(x) \wedge$ cat′(x): set intersection
- tall tree: tall′$(x, \lambda y[P(y)]) \wedge$ tree′(x)
  tall with respect to some contextually defined set of entities, tall with respect to trees, tall with respect to trees in Cambridge etc
  set intersection, but modified set (also gradable)
- fake gun: fake′(gun′(x))
  may or may not be a gun (but note fake watch, plastic aardvark)

Formal semantics and vector operations

- Denotation: white cat: white$'(x) \wedge$ cat$'(x)$: set intersection. Intuitively corresponds to vector multiplication.
- Properties: a white cat has both the properties of being white and the properties of being a cat. Corresponds to vector addition.
- white cat vs *cat white
  man bites dog vs dog bites man
  So: order sensitive?
  BUT: order matters for English much more than some other languages
  how much syntax do we want to (re)do with tensor operations?

Alternative models

- Ganesalingam and Herbelot, unpublished (2013) `www.cl.cam.ac.uk/~ah433/` for an overview of the mathematical properties
- Zanzotto et al (2015) also discuss different operations
- BUT: phrasal similarity datasets are too small to allow statistically valid distinctions between models

Some results

- vector addition is usually a very difficult baseline to beat experimentally
- Grefenstette (2013) demonstrates impossibility of properly modelling quantifiers with the tensor models
- semantic deviance spicy donkey: Vecchi et al (2017)
- Also: semi-compositionality (compound nouns, *heavy table* vs *heavy rain* vs *heavy taxation* etc).

Outline.

Building distributions (A)

- ... saw the cat's ears twitch ...

Building distributions (A)

- ... saw the cat’s ears twitch ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...

Building distributions (A)

- ... saw the cat’s ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat’s whiskers ...

Building distributions (A)

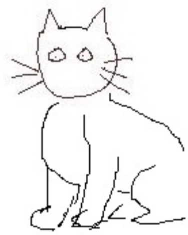
- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...

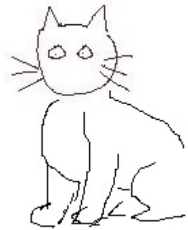
Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...



Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...



Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions (A)

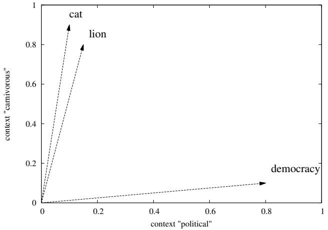
- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

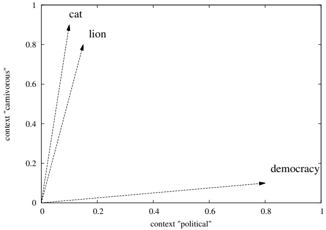
Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...

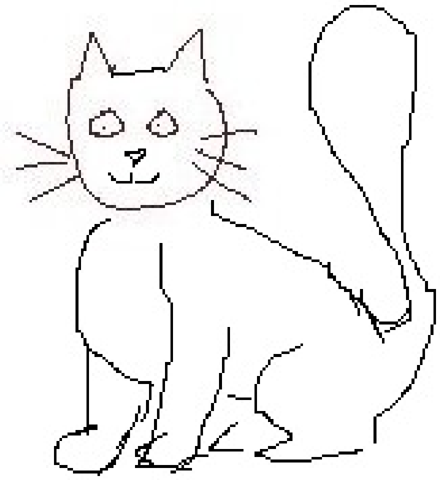


Building distributions (A)

- ... saw the cat's ears twitch ...
- ... the big cat turned his head ...
- ... that the cat had dark green eyes ...
- ... paint with cat's whiskers ...
- ... the cat stretched his legs ...
- ... the cat's enormous fluffy tail ...
- ... he is the cutest cat you'll ever ...



A distributional cat (the theory) (A)



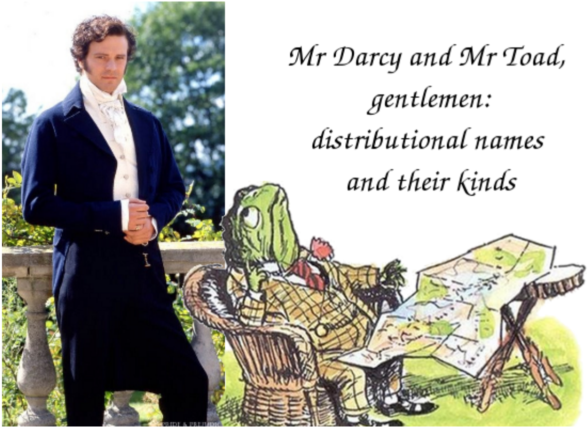
A distributional cat (the reality) (A)

Why does that cat look so bad? (A)

- Distributions model generic information.

Only 7% of NPs are references to kind

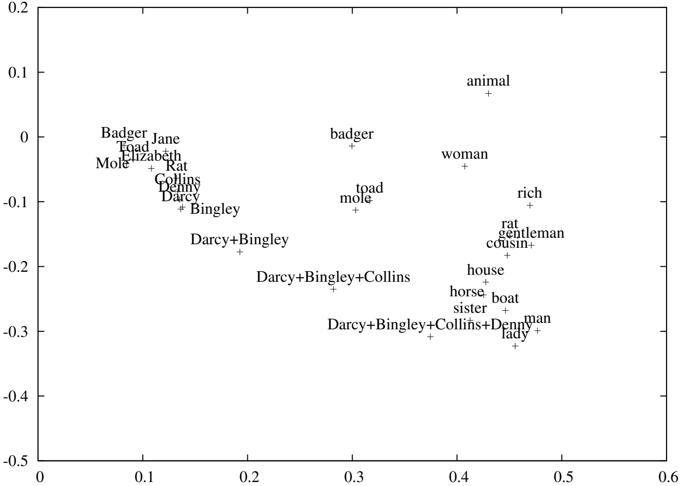
... an entirely black cat, like ...
... she owned a big ginger cat ...
... the cat was striped ...
... two long-haired white cats ...
... was a small grey cat ...
... cats are mammals ...



Aurelie Herbelot, IWCS 2015

Contextualised individuals in the BNC semantic space

(A)



Outline.

Background / recap

Distributional semantics for phrases

Distributional semantics for individuals

Functional distributional semantics

Emerson and Copestake (2016, 2017)

- Functional Distributional Semantics: functions mapping from points in semantic space (‘pixies’ — corresponding to individuals) to truth values.
- Distinguish between probabilistic truth values and observed text.
- DMRS gives joint distribution between entities.
- Implementation using Cardinality Restricted Boltzmann machine (CaRBM) trained on the Wikiwoods corpus.
- Inference via conditional probabilities, also distributional similarity.
- e.g., *lion*, *stone lion*; *roses*, *plastic roses*, *stone roses*

Functional Distributional Semantics (G)

$$\text{dog} \xleftarrow{\text{ARG1}} \text{chase} \xrightarrow{\text{ARG2}} \text{cat}$$

Functional Distributional Semantics (G)

$$x \xleftarrow{\text{ARG}1} y \xrightarrow{\text{ARG}2} z$$

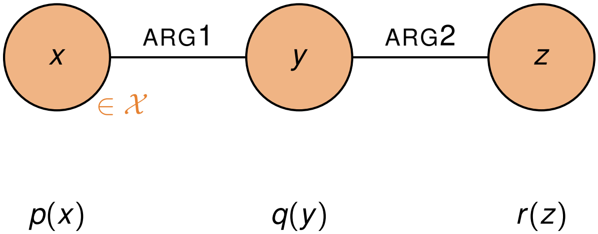
$\text{dog}(x)$ $\text{chase}(y)$ $\text{cat}(z)$

Functional Distributional Semantics (G)

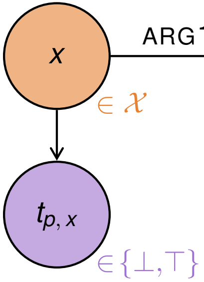
$$x \xleftarrow{\textsc{arg}1} y \xrightarrow{\textsc{arg}2} z$$

$$p(x) \qquad\qquad q(y) \qquad\qquad r(z)$$

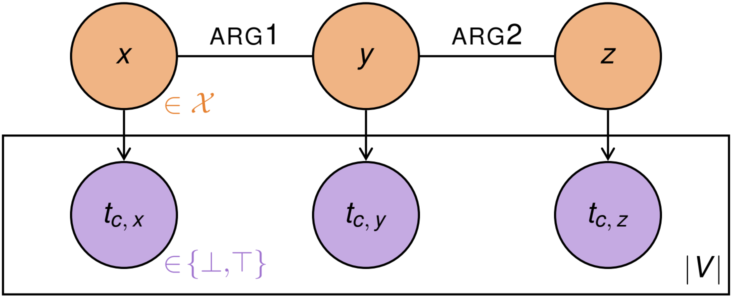
Functional Distributional Semantics (G)



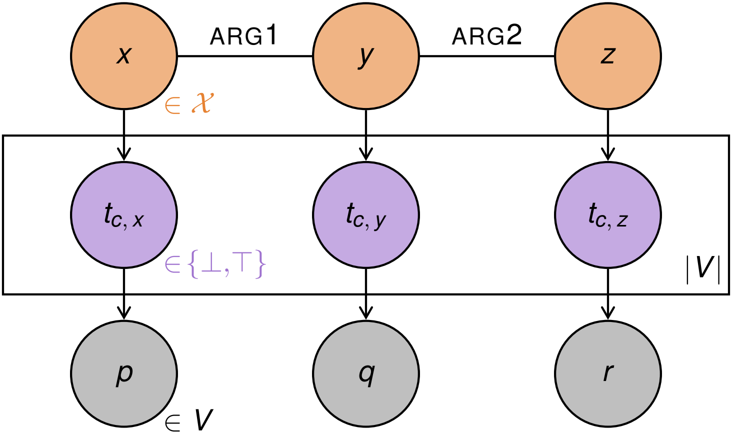
Functional Distributional Semantics (G)



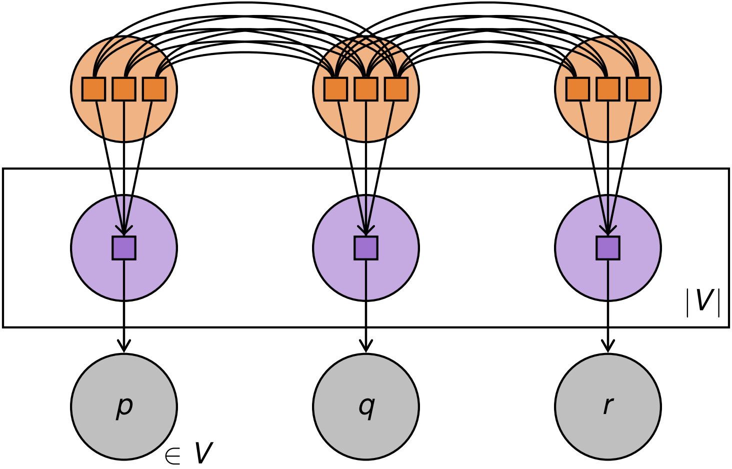
Functional Distributional Semantics (G)



Functional Distributional Semantics (G)



Functional Distributional Semantics (G)



RELPRON: Rimell et al (2016) (G)

telescope	*device that astronomers use*
telescope	*device that detects planets*
saw	*device that cuts wood*
philosopher	*person that defends rationalism*
survivor	*person that helicopter saves*
...	*...*

RELPRON: Rimell et al (2016) (G)

telescope	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	...

RELPRON: Rimell et al (2016) (G)

saw	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	...

RELPRON: Rimell et al (2016) (G)

philosopher	*device that astronomers use*
	device that detects planets
	device that cuts wood
	person that defends rationalism
	person that helicopter saves
	...

RELPRON: Rimell et al (2016) (G)

device that astronomers use
device that detects planets
device that cuts wood
person that defends rationalism
person that helicopter saves
...

Designed to test compositionality — standard distributional model (relatedness) with addition works reasonably well. But . . .

RELPRON confounders (G)

balance	*quality that ear maintains*
account	*document that has balance*

Confounders fool simple similarity / vector addition.

RELPRON confounders (G)

balance	*quality that ear maintains*
account	*document that has balance*

Confounders fool simple similarity / vector addition.

RELPRON confounders (G)

- Test set has 27 confounders
- Word2Vec:
  - 17 confounders in top rank
  - all confounders in top 4 ranks

RELPRON confounders (G)

- Test set has 27 confounders
- Word2Vec:
  - 17 confounders in top rank
  - all confounders in top 4 ranks
- Ensemble (word2vec plus FDS):
  - 9 confounders moved out of top 10 ranks

Next time (last lecture)

November 16 15:00: Q and A

- How can neural models be used as a drop in for previous algorithms and how can the construction of a well-studied problem be modified for "the latest deep learning" to be useful?
- What is your approach to deep learning model selection?
- Evaluating model performance and learning in non-probabilistic models: can you do something equivalent to feature analysis in SVMs in neural models?
- What are your approaches to learning validation?

Next time (last lecture)

November 16 15:00: Q and A

- How can neural models be used as a drop in for previous algorithms and how can the construction of a well-studied problem be modified for "the latest deep learning" to be useful?
- What is your approach to deep learning model selection?
- Evaluating model performance and learning in non-probabilistic models: can you do something equivalent to feature analysis in SVMs in neural models?
- What are your approaches to learning validation?

Vania and Lopez (2017): looks at whether various character-level LSTM models are really capturing morphology.