Foundations of Language Science and Technology

Semantics 3

Manfred Pinkal
Saarland University

The Story

- Modelling natural-language inference as deduction in a framework of truth-conditionally interpreted logic appears intuitive and straightforward.
- But: Logical methods are expensive and lack robustness and coverage.
- Corpus-based statistical methods for modelling inference are inexpensive and have no coverage problem.

Basic idea: Approximating inference by similarity between H and P

Shallow Inference Checking

Logical Entailment: ✓

$\forall d\,(dolphin(d) \rightarrow mammal(d) \wedge \neg fish(d)) \models \forall d\,(dolphin(d) \rightarrow mammal(d))$

Semantic Interpretation

Dolphins are mammals, not fish. They are warm blooded like man, and give birth to one baby called a calf at a time. At birth a bottlenose dolphin calf ... ⊨ Dolphins are mammals.

Textual Inference

Word Overlap

String match:

P: *Dolphins are mammals, not fish.*

H: *Dolphins are mammals.*

Word Overlap:

P: *William H. Seward served as Secretary of State under President Abraham Lincoln.*

H: *William H. Seward was Lincoln's Secretary of State*

Word Overlap

String match:

P: *Dolphins are mammals, not fish.*

H: *Dolphins are mammals.*

Word Overlap:

P: *William H. Seward served as Secretary of State under President Abraham Lincoln.*

H: *William H. Seward was Lincoln's Secretary of State*

P-H-relatedness: $\frac{\text{\# of words in H occurring in P}}{\text{length of H}}$

Word Overlap: Problem 1

- Entailment may be due to semantic similarity rather than identity of words: Concept overlap.

  P: *Aki Kaurismäki directed his first full-time feature*

  H: *Aki Kaurismäki directed a film*

  P: *Several airlines polled saw costs grow more than expected, even after adjusting for inflation*

  H: *Some companies reported cost increases*

- The degree of semantic similarity between pairs of related words should go into the computation of the P/H-Relation.

WN: Synsets, glosses, examples

- S: (n) **feature**, feature film (the principal (full-length) film in a program at a movie theater) *"the feature tonight is `Casablanca'"*
  - *direct hyponym* / *full hyponym*
  - ***direct hypernym*** / *inherited hypernym* / *sister term*
    - S: (n) movie, film, picture, moving picture, moving-picture show, motion picture, motion-picture show, picture show, pic, flick (a form of entertainment that enacts a story by sound and a sequence of images giving the illusion of continuous movement) *"they went to a movie every Saturday night"; "the film was shot on location"*

WordNet Similarity

- Based on WordNet, quantitative measures of distance / similarity between two concepts or word senses can be defined:
- A simple distance measure: Path length $dist_{WN} = pathlength(s_1,s_2)$
- A simple similarity measure: Inverse of path length $sim_{WN} = \frac{1}{pathlength(s_1,s_2)}$
- Normalisation, e.g., by path length from root to lowest common hypernym
- More complex corpus-related WN measures are based on the ratio between the informativity of the compared senses and the informativity of their lowest common hypernym. (Informativity of s can be measured as the negative log value of the probability that a content word in the corpus is a hypomyn of s.)

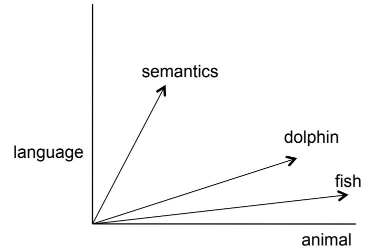
Distributional Similarity

- Distributional hypothesis:
  Two words are semantically similar to the extent that they occur in similar contexts.
- Context of a word w:
  - The document/ paragraph/ sentence, in which w occurs, or:
  - A window containing n (5, 10, 30, ...) words before and after an occurrence of w.
- Distributional “meaning representation” of w:
  - We count (in the simplest case) the number of occurrences of all words/ all content words/ the k (100, 1000, 10000, ...) most frequent words across all contexts of w (in a corpus).
  - The meaning of w is represented as a function from the considered context words to integers (frequencies), in other words:
  - As a vector in a multi-dimensional space (the “word space”).

Simple Example

- Frequencies of ‘animal’ and ‘language’ in the context of ‘dolphin’, ‘fish’, and ‘semantics’.

	dolphin	semantics	fish
animal	55	15	70
language	15	45	5



- The table and its graphical representation indicate the affinity of ‘dolphin’ and ‘fish’ to the domains of zoology, and of ‘semantics’ to language.
- They also indicate that ‘dolphin’ and ‘fish’ are more similar to each other than to ‘semantics’.

Distributional Similarity

- One standard measure for distributional similarity is cosine:
- Cosine is 1, if vectors have identical directions ($cos(0^0)=1$), it is 0, if vectors are orthogonal ($cos(90^0)=0$).
- General definition:

$$sim\ (x,y) = \cos(\vec{x},\vec{y}) = \frac{\vec{x}\cdot\vec{y}}{|\vec{x}||\vec{y}|} = \frac{\sum_{i=1}^{n} x_i \times y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}$$

- In our example:

$$sim(semantics, dolphin) = 0.55$$
$$sim(semantics, fish) = 0.38$$
$$sim(fish, dolphin) = 0.98$$

Distributional Similarity and Inference

The general strategy for determining inference through semantic similarity:

- Compute similarity between all word pairs in text and hypothesis.
- Alignment: Find the best match between H and P words, i.e., the match that maximises similarity between H and P.
- Given an alignment, add all similarity values for matching word pairs, normalise with sentence length.
  Example, using the simple WordNet similarity measure:
  P: Aki Kaurismäki directed his first full-time feature
  H: Aki Kaurismäki directed a film
  Alignment: <Aki, Aki>, <Kaurismäki, Kaurismäki>, <directed, directed>, <film, feature>
  sim (H, P) = (1+1+1+0.5)/5 =3.5/5 = 0.7
- For determining entailment, set a threshold s in a way that maximises accuracy.
  P ⊨ H iff sim (H, P) > s

Word Overlap: Problem 2

- Same word may be used with different senses:
  - Word-sense disambiguation: selecting for each word occurrence a word sense from a set of senses given by an external resource (dictionary, WordNet)
  - Word-sense discrimination: Determining whether two word occurrences belong to the same word sense or not.

Word&Concept Overlap: Problem 3

- Bag-of-Words and simple vector-space models of inference are insensitive to word-order, which makes them highly imprecise:
  Man bites dog vs. *Dog bites man.*
- Use a method that penalises matches between distant words, and rewards matches between longer strings.
- But compare:
  P: The main race track in Qatar is located in Shahaniya, on the Dukhan road.
  H: Qatar is located in Shahaniya.

Word&Concept Overlap: Problem 3

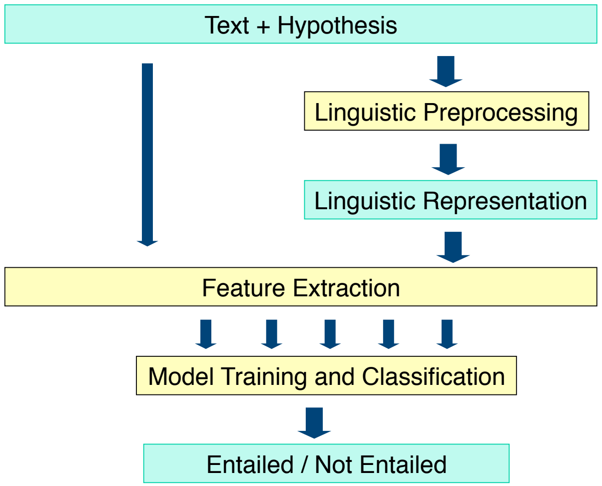
- Bag-of-Words and simple vector-space models of inference are insensitive to word-order, which makes them highly imprecise:
  Man bites dog vs. *Dog bites man.*
- Use a method that penalises matches between distant words, and rewards matches between longer strings.
- But compare:
  P: The main race track in Qatar is located in Shahaniya, on the Dukhan road.
  H: Qatar is located in Shahaniya.
- We need syntactic information in addition (provided by statistical parsers - context-free and dependency parsers).

- How good can shallow, similarity-based methods perform in principle?

Recognizing Textual Entailment

- The RTE Challenge: An annual competition / shared task
- Started in 2004
- Training corpus and test corpus
  - 800 T-H pairs
  - 400 true, 400 false ones
  - formed on the basis of material taken from IR, IE, Q&A, Summarization tasks
  - no domain restriction
  - Training corpus annotated with Yes/No.
- Task: Build a system (using whichever methods and resources) that matches the Y/N annotation of the corpus as close as possible.

Typical Architecture of RTE Systems



Information used in different RTE Systems

- Word overlap
- Distributional similarity
- WordNet information
- String Matching
- Shallow syntactic Information

- Deep syntactic information
- FrameNet information
- World knowledge
- Logical inference

Challenges

- Distributional similarity does not guarantee semantic similarity: Antonymous words typically have similar distribution
- Distributional similarity is a symmetric relation, entailment/ inference an asymmetric one.
- No intuitively appropriate concept of semantic composition (relating lexical similarity to text-hypothesis relation).
- In particular, the effect of negation, quantification, modality cannot be modelled.

General Tendencies of Results

- “Knowledge-lean” systems relying on shallow information (word overlap, string match, distributional similarity) perform better than naïve baseline of 50%, but only to some degree (60-65%).
- They may provide a good estimate of “aboutness”: Is the Premiss/ text about the issue raised by the hypothesis?
- Systems relying on deep linguistic analysis and logical entailment perform drastically worse than naïve baseline (but are significantly more precise on cases they can treat).

- How can the best of deep and shallow methods be combined?

Negation and polarity

P: *Whooping cough, or pertussis, is a highly contagious bacterial infection characterized by violent coughing ts, gasp for air that resemble 'whoop' sounds, and vomiting*

H: *Pertussis is <u>not</u> very contagious.*

P: *Energy analysts said oil prices <u>could</u> soar as high as $80 a barrel, if damage reports from oil companies bear bad news.*

H: *Oil prices surged.*

"Natural Logic" and Textual Inference

The RTE Example again:

Several airlines polled saw costs grow more than expected.

Some companies reported cost increases.

A simplified version:

Several airlines reported cost increases

Several companies reported cost increases

Textual Inference and Logical Inference

P: *Several airlines reported cost increases*

H: *Several companies reported cost increases*

- H can be obtained from P by a single substitution.
- *airlines* and *companies* stand in hyponymy or "lexical entailment" relation (which we also write as $\sqsubset$).
- From this, it clearly follows that P (logically) entails H - without a full logical analysis of the sentences.

More examples

P: *Several airlines polled reported cost increases*

H: *Several airlines reported cost increases*

- Deletion of modifiers preserves entailment.

P: *Several airlines polled reported cost increases*

H: *Several companies reported cost increases*

- Two entailment-inducing edits ad up to entailment again.

More examples

P: *Several airlines reported cost increases*

H: *Several airlines polled reported cost increases*

- Insertion (of modifiers) causes non-entailment (actually, it causes inverse entailment.

P: *Several airlines reported cost increases*

H: *Several companies polled reported cost increases*

- The combination of edits with opposite entailment effects leads to non-entailment (semantic independence) of P and H.

The Natural Logic Approach to Inference

- Relate P and H through a sequence of atomic edit operations - deletions, insertions, or substitutions.
- Compute the entailment relation between P and H by joining the entailment effects of the atomic edits.

Example

P: *Several airlines polled saw costs grow more than expected.*

H: *Some companies reported cost increases.*

We may obtain H from P through the following edits with their respective entailment effect:

SUB(*several, some*)	→	$\sqsubset$
SUB(*airlines, companies*)	→	$\sqsubset$
DEL(*polled*)	→	$\sqsubset$
SUB(*saw, reported*)	→	$\equiv$?
SUB(*costs, cost*)	→	$\equiv$
SUB(*grow, increases*)	→	$\equiv$?
DEL(*more than expected*)	→	$\sqsubset$

What we need

- A method to find the best or most appropriate alignment/ sequence of edit steps between P and H.
- A sufficiently general definition of the semantic relations induced by atomic substitutions ("lexical entailment relations") for arbitrary lexical items.
- A method to identify the specific lexical entailment relations induced by specific SUB edits; DEL and INS induce $\sqsubset$ and $\sqsupset$, respectively.
- A full specification of the join operation between entailment relations.
- A method to compute the effect of the lexical entailment relations on the logical entailment relation between full sentences - taking the context of the edits into account.

The effect of context

P: *John bought a new convertible.*
H: *John bought a new car.*

P: *John didn't buy a new convertible.*
H: *John didn't buy a new car.*

- In an affirmative standard context, a context with "positive polarity", an "upward monotonic" context, sentence-level entailment is atomic lexical entailment.
- In the context of a negation, a context with "negative polarity", a "downward monotonic" context, atomic lexical entailment is inverted on the sentence level.