Main Findings & Future Challenges

Hinrich Schütze

Center for Information and Language Processing, LMU Munich

2015-09-18

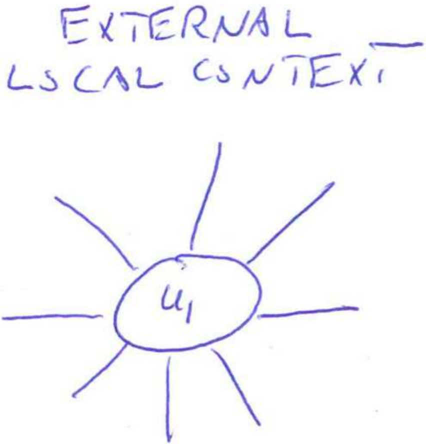
Overview

Outline

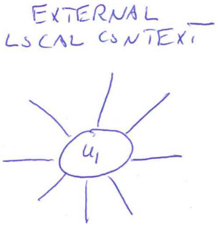
Types of context

- External local context:
  linguistic material to left/right of linguistic unit u_1
- Globally accumulated external local contexts of u_1
- Globally accumulated internal local contexts of a set U
  - U is an equivalence class,
    e.g., a set of propositions $\{u_1, u_2, u_3\}$ with same/similar meaning
- No context

External local context



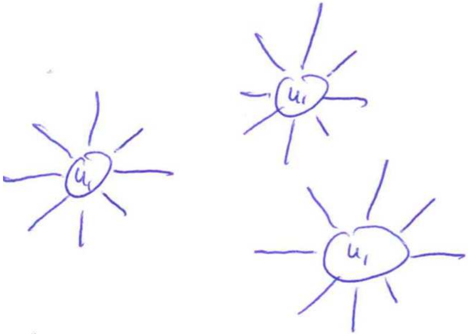
External local context: Example



- This is the standard definition of context.
- For example, disambiguation of an ambiguous word like “suit” is based on external local context.
- In the sentence “he wears a suit except for Fridays”, the external local context of “suit” is “he wears a” + “except for Fridays”.

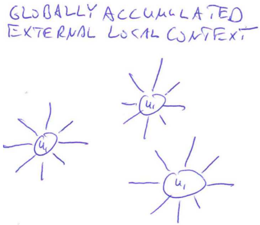
Globally accumulated external local context

GLOBALLY ACCUMULATED
EXTERNAL LOCAL CONTEXT



Globally accumulated external local context

- This is the standard model for learning word embeddings.
- The word embedding of a word like "car" is derived from all external local contexts of "car" in a corpus.

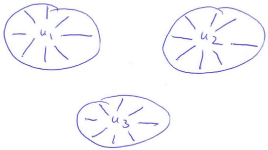


Globally accumulated internal local context

Globally accumulated internal local context

- This is used for consolidation in Ido's proposition graph.
- Proposition 1: Standard IQ tests are the most commonly used means of identifying gifted children.
- Proposition 2: Gifted children were often classified via IQ tests.
- Assumption: We have a good procedure for clustering sentences in the corpus into clusters of sentences with closely related meaning. Cluster in this case: { P1, P2 }
- The sentences in this cluster provide context for each other, e.g., we can infer "IQ tests" = "Standard IQ tests" & "classify" = "identify" (equivalences that don't hold in general).

GLOBALLY ACCUMULATED INTERNAL LOCAL CONTEXT



No context

- Classic compositional semantics
- Builds the meaning of the whole from the meaning of the parts
- Syntax-semantics homomorphism: each syntactic composition operation has a corresponding semantic composition operation
- The same operation, independent of context

no
context

Outline

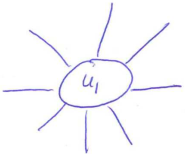
DRT: External local context

EXTERNAL LOCAL CONTEXT

John, the teacher, and Mary, the student, entered the room.
He lectured for two hours.

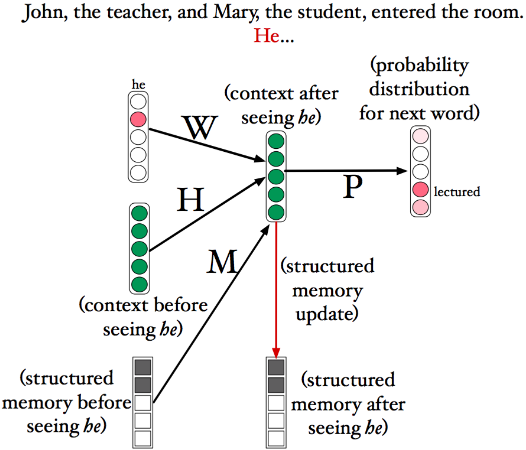
x, y, z
John(x) teacher(x) Mary(y) student(y) room(z) enter(x,z) enter(y,z)

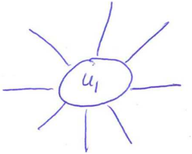
x, y, z, u
John(x) teacher(x) Mary(y) student(y) room(z) enter(x,z) enter(y,z) lecturing(u) u=z



Neural MT: External local context

EXTERNAL LOCAL CONTEXT





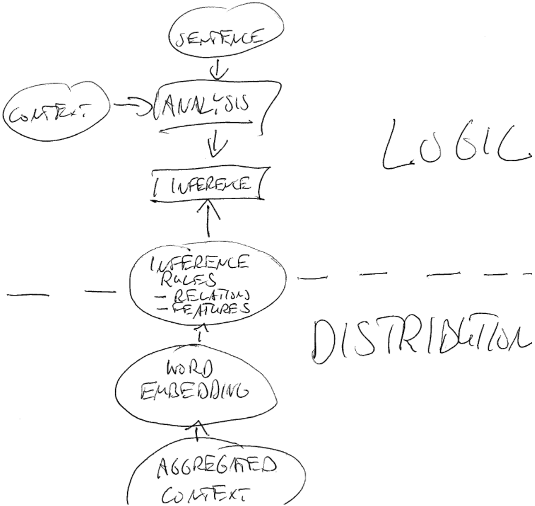
Words/Sentences × Static/Operators

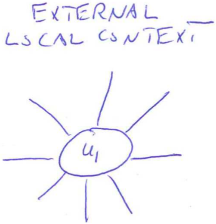
	context-dependent	context-independent
static representation	sense-embedding sentence1	word-embedding1 sentence2
operator C→C	RNN1	DRT, CCG, RNN2 word-embedding2 sentence3

In many formalisms (CCG, RNNs), both words (word-embedding2) and sentences (sentence3) can be viewed as operators that map from context to context. More commonly, words are considered to have static meaning representations (word-embedding1) and this can be extended to sentences (sentence2). Maybe a case of sentence1 would be a static formula with variables (for referents) and these are then substituted for depending on context.

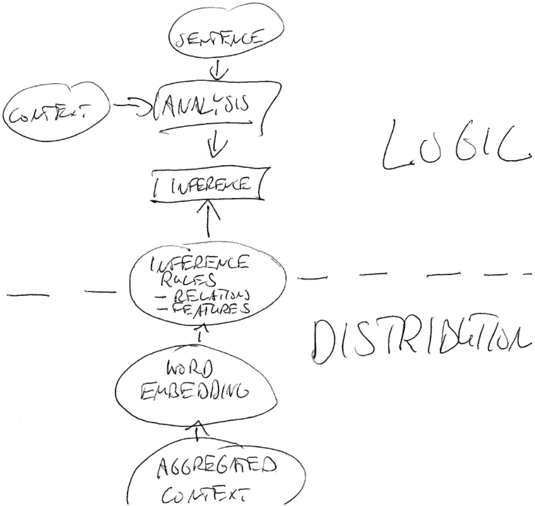
Outline

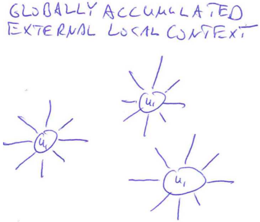
Analysis depends on context: External local context





Embeddings:
Globally accumulated external local context

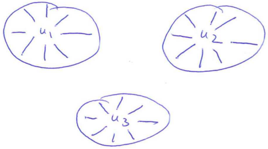




Outline

Merging nodes in proposition graph: Globally accumulated internal local context

GLOBALLY ACCUMULATED
INTERNAL LOCAL CONTEXT



- "Peter fixed the typo." /
  "Peter corrected the type."
- "The drug cured the cancer." /
  "The drug killed the cancer."
- "The drug cured the patient." /
  "The drug killed the patient."

Interpretation depends on (external local) context

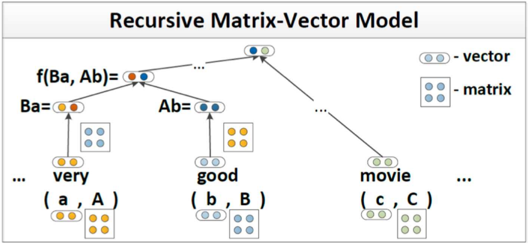
- Extreme example: pragmatic intrusion
- About a planned execution: "He will be lucky to die quickly."
- Interpretation 1: He is unlikely to die quickly, so he will need a lot of luck to die quickly. . . . "Unfortunately, this is a US state that does not have a working execution protocol."
- Interpretation 2: He will die quickly and he is lucky. . . . "Fortunately for him, this is a US state that has a working execution protocol."
- One interpretation is p, the other $\neg p$ where $p =$ "he will die quickly"!
- (example due to Lauri Karttunen)

Intermediate solution vs End game

- "Although standard IQ tests are the most commonly used means of identifying gifted children, other tests of both intelligence and creativity are also used."
- Similar views in deep learning and pragmatic computational linguistics.
- If we cannot handle the hard stuff now ("most commonly used")
- then just ignore this really hard stuff and use context as substitute.
- Interesting similarity between Ido's and Phil's approach.

Outline

Recursively constructing operators: No context



no context

Context vs Sentences-as-vectors: Views

- Non-problem:
  e.g., logical-form representation also is non-contextual
- Learn representations from context of other sentences the same way we learn representations of words from context of other words (Marco)
- Sentences as functions from contexts to contexts
- Maybe the difference between [sentence-as-vector] and [vector-represents-contexts-up-to-this-point] is actually pretty small: most of context-up-to-this-point is the last sentence.
- Keep more structure in the sentence (Zanzotti, Smolensky)

Operator semantics vs Content semantics

- Do we need one representation for operator semantics and one representation for content semantics?
- Can both be built by classic compositionality?
- If we have both operator and content semantics, what to do with them?
- operator(sentence-i+1)(context-up-to-sentence-i) = context-up-to-sentence-i+1
- content(sentence-i+1)

Outline

Things I've learned (1)

- Linguistic/cognitive desideratum
  - Make models modular, structured
  - E.g., modular memory:
    working, semantic, episodic, procedural
- End-to-end learning desideratum
  - Differentiability and practical learnability are key.
  - Easiest way to get there:
    Big-blob kitchen-sink models
  - But these models are hard to analyze.

Things I've learned (2)

- Unappreciated insight from recent research: distribution $\subset$ corpus evidence
- $|$distribution$| \ll |$corpus evidence$|$
- Widespead naive view of distributional semantics: PPMI values, i.e., (weighted) cooccurrence counts
- Not the future?
- Alternative
  - "Deep" analysis of each occurrence of a word
  - Aggregate analysis of contexts from big corpus
  - Increasingly popular approach (Bar-Ilan, Trento, LMU)

Things I've learned (3)

- Different meanings of:
  "I understand what my model is doing."
- The bar for this claim in deep learning seems very low.
- The bar for this claim in linguistics is very high.
- Bridging this gap should be an important goal.

Understanding models: We need to get to the level of what vision does