Natural Language Inference: for Humans and Machines

Valeria de Paiva
Topos Berkeley Seminar

January 2022

Personal stories

Mathematicians are told over and over again that Natural Language is ambiguous, messy and imprecise.
That one should study artificial languages, instead.

Some of us beg to differ.

Personal stories

Manning talking about NLP/NLU/NLI and 'The Deep Learning Tsunami' Computational Linguistics and Deep Learning, 2015 reported that "NLP is kind of like a rabbit in the headlights of the Deep Learning machine, waiting to be flattened."
Hinton 2015: "I will be disappointed if in five years' time we do not have something that can watch a YouTube video and tell a story about what happened."

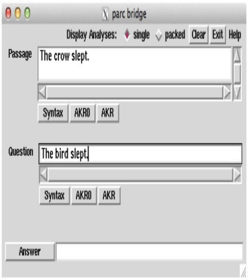
[not totally flattened, yet?]
https://www.youtube.com/watch?v=bZMKhQSERA4 (2019)

Personal stories

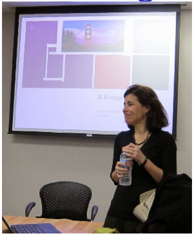
https://www.youtube.com/watch?v=bZMKhQSERA4
Fireside chat with Susan Dumais, MS (2019)
Manning and Schuetze's book (1999) describes statistical learning as complementary/alternative to traditional/pipeline way of doing NLP.
by now statistical NLP is the only way of doing it, 'an appealing drug' says Manning. Another huge change by middle 2010's: deep learning

Personal stories

PARC, XLE and Bridge



Natural Language Inference (NLI)



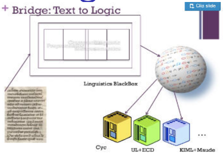
- A shock when the work of almost a decade at PARC was out of reach when I left in 2008
- I gave a talk at SRI proposing to redo it all, open source (de Paiva 2010 Bridges)
- Pleased to report that (almost) all of it now available open-source
- Mostly work by Katerina Kalouli, then PhD student at Konstanz, now faculty at LMU, Munich

Natural Language Inference: what?

Examples from SNLI dataset at Stanford

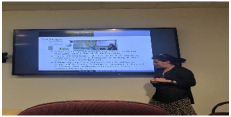
Text	Judgments	Hypothesis
A man inspects the uniform of a figure in some East Asian country.	contradiction C C C C C	The man is sleeping
An older and younger man smiling.	neutral N N E N N	Two men are smiling and laughing at the cats playing on the floor.
A black race car starts up in front of a crowd of people.	contradiction C C C C C	A man is driving down a lonely road.
A soccer game with multiple males playing.	entailment E E E E E	Some men are playing a sport.
A smiling costumed woman is holding an umbrella.	neutral N N E C N	A happy woman in a fairy costume holds an umbrella.

Old Bridges




- Bridges from Language to Logic: Concepts, Contexts and Ontologies. ENTCS, 2011.
- Entailment, intensionality and text understanding. C Condoravdi et al. HLT-NAACL workshop, 2003
- PARC's bridge and question answering system. DG Bobrow et al. Grammar Engineering Across Frameworks, 46-66, 2007.
- Textual Inference Logic: Take Two. DG Bobrow et al, CONTEXT 2007.
- A Basic Logic for Textual inference. D. G. Bobrow et al, AAAI Inference for Textual QA, 2005.

New Bridges



- Kalouli, A.-L., R. Crouch and V. de Paiva. 2020. Hy-NLI: a Hybrid system for Natural Language Inference.
- Kalouli, A.-L., et al. 2020. XplaiNLI: Explainable Natural Language Inference through Visual Analytics.
- Kalouli, A.-L., R. Crouch and V. de Paiva. 2019. **GKR: Bridging the gap between symbolic/structural and distributional meaning representations**. @ACL 2019.
- Crouch, R. and A.-L. Kalouli. Named Graphs for Semantic Representations. Proceedings of *SEM 2018.
- Kalouli, A.-L. and R. Crouch. GKR: Graphical Knowledge Representation for semantic parsing. @NAACL 2018.

Graphical Knowledge Representation (Kalouli and Crouch, 2018a)

Division of semantic labour, e.g. Clark and Pulman 2007

- distributional features: conceptual aspect of meanings, lexical aspects, semantic similarity, hypernym/antonym relations
- structural features: function words and Boolean and contextual phenomena, e.g., modals, quantifiers, implicatives, or hypotheticals

Graphical Knowledge Representation (Kalouli, Crouch, de Paiva 2019)

Three broad approaches to combine distributional and symbolic aspects of meaning representations:

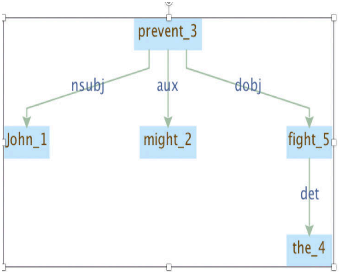
(i) injecting linguistic features into distributional representations
(ii) injecting distributional features into symbolic representation
(iii) combining structural and distributional features in final representation

Here our version of (iii), which you can road test at http://hynli.nlitoolkit.de/

Graphical Knowledge Representation (Kalouli and Crouch, 2018a)

- borrows from the projection architecture of LFG
- borrows from Bridge contexts/concepts logic
- more general: distinct layers/levels/subgraphs of sentence information allows multiple logics and representations alongside one another, i.e., symbolic/structural and distributional
- strict separation of, and controlled interaction between, the conceptual/predicate-argument layer and the contextual/Boolean layer
- rooted, node-labeled, edge-labeled directed graph
- (currently) consists of 6 subgraphs
- produced by our open-source semantic parser written in Java
- particularly suitable for the task of natural language inference (NLI)

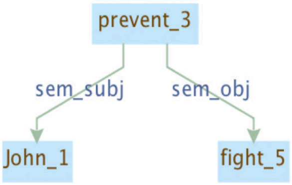
The Dependency subgraph



John might prevent the fight.

- full syntactic parse of the sentence
- output of Stanford CoreNLP
- Stanford Enhanced++ Universal Dependencies
- Stanford graph rewritten to our own dependency graph

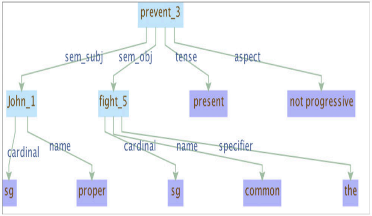
The concepts subgraph



John might prevent the fight.

- central graph of GKR
- propositional content of the sentence: what is talked about
- nodes represent concepts and not individuals
- claims about the existence of the concepts described by these content words
- **NO** claims about the existence of instances of those concepts, graph incomplete but accurate

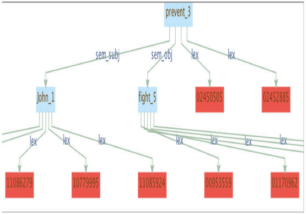
The grammatical properties subgraph



John might prevent the fight.

- on top of conceptual graph
- morpho-syntactic information, e.g., cardinality of nouns, verbal tense and aspect, finiteness of determiners, etc., and quantifiers
- for now: based on our own shallow morphological analysis of the POS tags

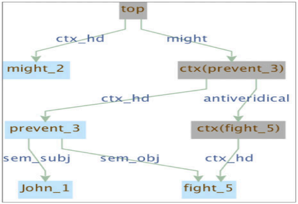
The lexical subgraph



John might prevent the fight.

- on top of conceptual graph
- WordNet senses (ordered with their probability based on the JIGSAW WSD algorithm)
- SUMO concepts
- WordNet 3.0 hyponyms, hypernyms, antonyms, synonyms

The context subgraph



John might prevent the fight.

- on top of conceptual graph
- existential commitments of the sentence
- top context and embedded contexts: each making commitments about its own state of affairs: which concept is instantiated and which isn't in each context
- embedded contexts: negation, disjunction, modals, clausal contexts of belief and knowledge, implicatives and factives, imperatives, questions, conditionals and distributivity

Naming Graphs?

The contextual subgraph

Named Graphs (Carroll et al., 2005) (RDF extension) associate an extra identifier (name) with a set of triples, e.g.,:

```
Fred believes John does not like Mary
      :g1 {:john :like :mary }
      :g2 :not :g1
      :fred :believe :g2
```

= Crouch and Kalouli, 2018b

contexts = names   concepts (and their children) = "triples"

```
John might prevent the fight.
   :g1 {fight}
   :g2 {:john :prevent :g1}
   :john :might :g2
```

➔ factoring out hard compositionality phenomena allows us to combine symbolic/structural and distributional approaches

All together?

- well, only look at the graph that interests you
- No more McCarthy style contexts istrue(c, Φ)
- plenty of opportunities to expand (time, conditionals,...)

How distributional gets in?
1. more symbolic: extend the lexical graph. instead of connecting to WordNet/SUMO, use (contextualized) BERT embeddings for concepts and try some learning of the matching process e.g The dog is catching a black frisbee/The dog is biting a black frisbe (still working on it!)
2. more distributional: this paper!

The distributional extension of GKR

- Given an NLI pair, find inference relation (entailment, contradiction and neutral)
- process each sentence of the pair with GKR
- apply the "naming" technique on each sentence: for each concept being a context head and all of its children, compute (whatever) distributional representation. This representation is now associated to a specific context through the context head and thus the representation has a specific instantiability (veridical, antiveridical, averidical)
- match distributional representations across sentences based on their similarity; look up instantiabilities and percolate them, if required
- "trick" we factored out the hard composionality into the contexts, so basic predicate-argument structure compositionality can be achieved in any (distributional) way desired – there are plenty around, e.g InferSent (Conneau et al., 2017)

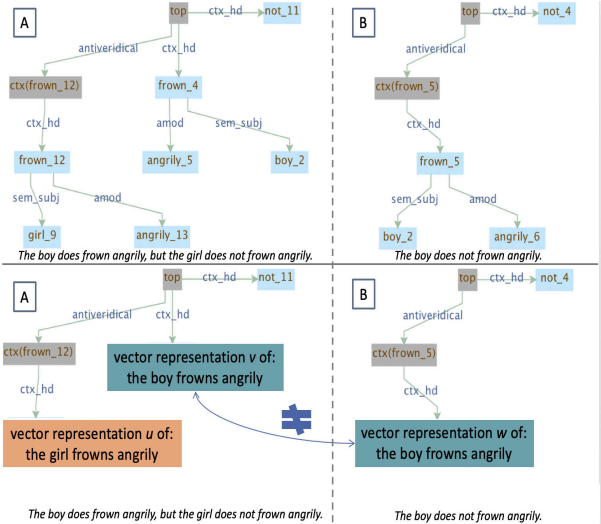
Experimental Work

- Dasgupta et al, 2018 (DS)
  NLI test sets with hard compositionality phenomena, e.g., negation, coordination, etc.
- classifier on the SNLI (Bowman et al, 2015) corpus using the state-of-the-art InferSent (Conneau et al, 2017) embeddings
- results: across sets around 50% accuracy

The experiment

- 2 sets of DS of a total of 4800 NLI pairs
- sentence A involves a conjunction of a positive sentence with a negative and sentence B contains one of the conjunct sentences either in its positive or its negative version.
  A= The boy does frown angrily, but the girl does not frown angrily. B= The boy does not frown angrily. Oops!
- DS report an accuracy of 53.2% and 53.8% for the two sets

Dasgupta et al comparison



Results of Experiment

- 99.5% accuracy on the 2 test sets
- error analysis: wrong output of Stanford Parser → wrong dependency graph → wrong conceptual graph → wrong contextual graph
- more cases of faulty Parser output but computation still succeeds if:
  - the conceptual graph is matched to a valid context graph
  - the matching between distr. representations is good enough

  due to:
  - the precision of the (symbolic) inference computation based on the instantiabilities found in the context graph
  - the robustness of distr. representations that should allow similar ones to match even if they encode partly wrong conceptual graphs

Preliminary Conclusions

- division of semantic labor beneficial both for symbolic/structural and distributional approaches
- GKR fulfills this role: strict separation of conceptual and contextual structures and separation of the sentence information in layers
- concrete proposal for injecting distributionality in GKR: promising results (in 2019)

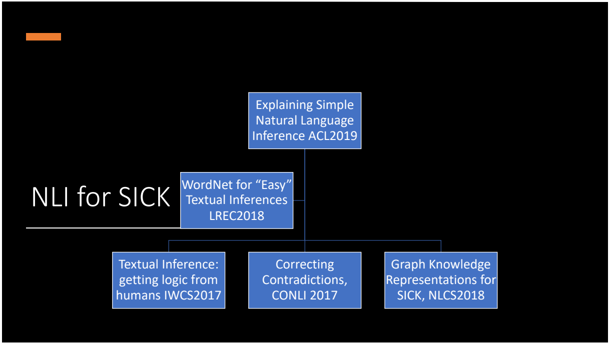
Natural Language Inference: why?

- In May 2016 Google announced Parsey McParseface, the world's most accurate parser[1]: 94% accuracy
- In 2014 Marelli et al launched the SICK corpus at SemEval 2014: an easy (no named entities, no temporal phenomena, limited vocabulary, etc..), linguist curated corpus to test compositional knowledge
- Can we use SyntaxNet to process SICK with off-the-shelf tools such as WordNet and SUMO?
- It's complicated! Six papers and counting!

[1] ai.googleblog.com/2016/0/announcing-syntaxnet-worlds-most.html

NLI for Humans

- Easier to detect inference than to decide on "good" semantic representations
- Data-driven NLU need large, diverse, high-quality corpora annotated to learn inference relations: entails, contradicts, neutral
- Can we trust the corpora we have?
- Are they really learning logical inferences?
- Are the findings on the big corpora available SNLI, MNLI, SciTail, etc transferable and generalizable? (Plenty of recent work showing no, systems learn biases of the corpora, cannot be redeployed)


NLI for SICK
Explaining Simple Natural Language Inference ACL2019
WordNet for "Easy" Textual Inferences LREC2018
Textual Inference: getting logic from humans IWCS2017
Correcting Contradictions, CONLI 2017
Graph Knowledge Representations for SICK, NLCS2018

NLI for SICK

- Are the annotations in SICK logical? Can we trust them?
- Several problems: lack of guidelines on co-reference, how to annotate contradictions, ungrammatical and non-sensical sentences, noisy data, etc..
- This meant contradictions in SICK are not symmetric and they need to be
- Contradictions require alignment between entities and events, which need to be "close enough"
- how to decide when things are close enough?
- Can we do simpler case where sentences are "one-word-apart" using WordNet?
- More guidelines necessary for SICK annotation?

NLI for SICK

- https://logic-forall.blogspot.com/2020/03/sick-dataset-in-these-trying-times.html

Conclusions

- Working for division of semantic labor between symbolic/structural and distributional approaches
- Have implemented proposal GKR with strict separation of conceptual and contextual structures
- Also concrete proposal for injecting distributionality in GKR: promising results of hybrid system
- Produced a 'correct' SICK, finally
- Submitted paper on annotations and theorem provers, together with this new SICK
- **Further Work**: Hardening system
- Test GKR with further datasets, further distributional architectures (RoBERTa)
- plenty of ideas: new languages, porting to Python, improving resources

More information

GKR source code:

`https://github.com/kkalouli/GKR_semantic_parser`

`https://github.com/kkalouli/GKR4NLI`

`https://github.com/kkalouli/XplaiNLI`

Demos for all bits of system

Screenshot

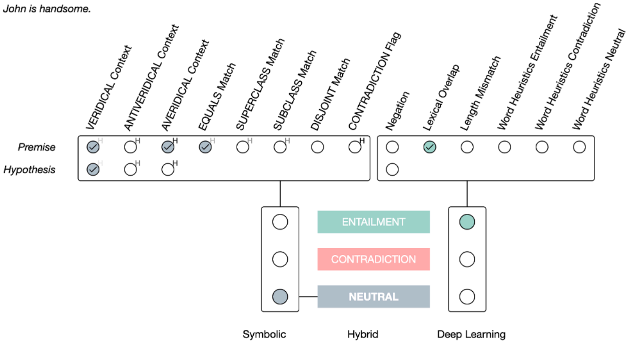
Explanation

After exploring the visualization, click on the inference label that you think is correct for this pair. Thanks for your feedback!

WARNING: The visualization has only been tested on Safari, Firefox and Chrome.

Premise: *Mary believes that John is handsome.*

Hypothesis: *John is handsome.*



Thanks!