Cognitive States: Belief State Inference via Deep Learning

Tyler Osborne
CSOM & MCAS 2023
CSCI 4961 Honors Thesis
5 May 2023

Part 0: Acknowledgements

No man is an island, and I am no exception...

Agenda

I. Background: Who Cares?
II. Language Understanding (LU) Corpus
III. Factbank Corpus & Unified DB Model
IV. Data Transformations
V. LU Experiments, Results & Insights
VI. Future Work

Part I: Who Cares?

Why is this an important topic in NLP?

Who Cares?

- When we speak, we convey information, but not all of that information is objective
- Often, what we convey is wrapped up in a *belief*
    - "John said Mary is coming to dinner."
    - To what degree does John believe in the factuality of his utterance?

Who Cares?

- "John said Mary is coming to dinner."
- To us, it is abundantly clear that John fully believes that his utterance is true; we want AI to have the same ability
- "John guessed that Mary might come to dinner."

Who Cares?

- This sort of analysis brings us closer to capturing the full *private state* or *cognitive state* of someone in a text
  - Set of sentiments & beliefs towards what they say
- For our purposes, the people are *sources*, the beliefs they express *targets*, and the degree to which the source believes in the factuality of their utterance, the *label*

Part II: Language Understanding (LU) Corpus

Language Understanding (LU) Corpus

- Corpus: Collection of text

- LU is an annotated corpus
  - Humans have noted source-target pairs in the text and assigned each one a label
  - The author of a sentence itself is the default source
    - "Iraq clears visit by Ohio official."

LU Corpus

- LU's labels are:
  - *CB* for committed belief
    - "I am certain that..."
  - *NCB* for non-committed belief
    - "I am not sure but think that..."
    - "I hope that..."
  - *NA* for not applicable
    - No belief expressed

LU Corpus

"He <u>did not speak</u> to reporters in Jordan, but he <u>told</u> the Associated Press before leaving the United States that he hopes to '<u>separate</u> the <u>humanitarian work</u> from the <u>political issues</u>.'"

Issues with LU

- LU is not a huge corpus (<7000 english words)

- Other corpora with source-target-label annotations exist, but combining them natively is next to impossible
  - Why?

Issues with LU

- However, if we could somehow port each individual corpus into a single, unified format, then we could combine them!
    - Prof. Aviram and I developed such a representation, among other tasks we contributed to the project

Part III: Factbank & Unified Database Model

Factbank: A Natively Relational Corpus

- Factbank is another belief state annotation corpus falling under the source-target-label paradigm
  - Different label scheme

Table 1: Factuality values

VALUE	DESCRIPTOR	USE
		Committed Values
CT+	Certainly positive	According to the source, it is **certainly** the case that X.
PR+:	Probably positive	According to the source, it is **probably** the case that X.
PS+	Possibly positive	According to the source, it is **possibly** the case that X.
CT-	Certainly negative	According to the source, it is **certainly not** the case that X.
PR-	Probably negative	According to the source it is **probably not** the case that X.
PS-	possibly negative	According to the source it is **possibly not** the case that X.

Factbank: A Natively Relational Corpus

Sentence	Target Head	Source Text	Label
… for an economy that many experts thought was once <u>invincible</u>.	invincible	Author	CT+
… for an economy that many *experts* thought was once <u>invincible</u>.	invincible	experts_Author	CT+

LU vs. Factbank

LU

- Three labels
- Bona fide flat files (XML)
- Author-only annotations

Factbank

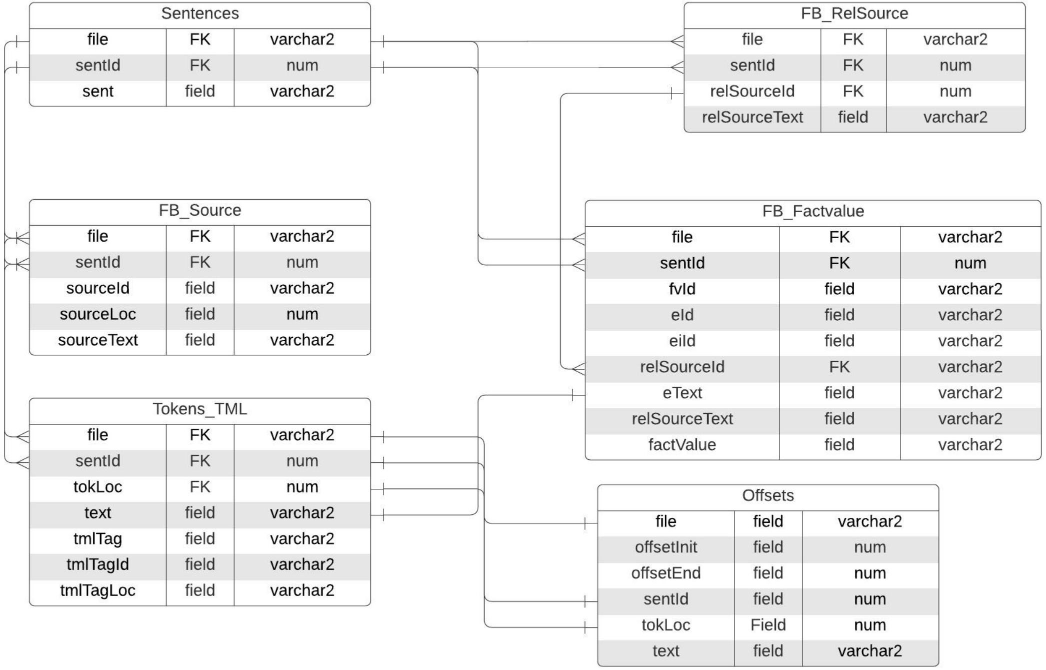
- Six labels
- Relational data stored as flat files
- Author & nested source annotations

Conclusion:

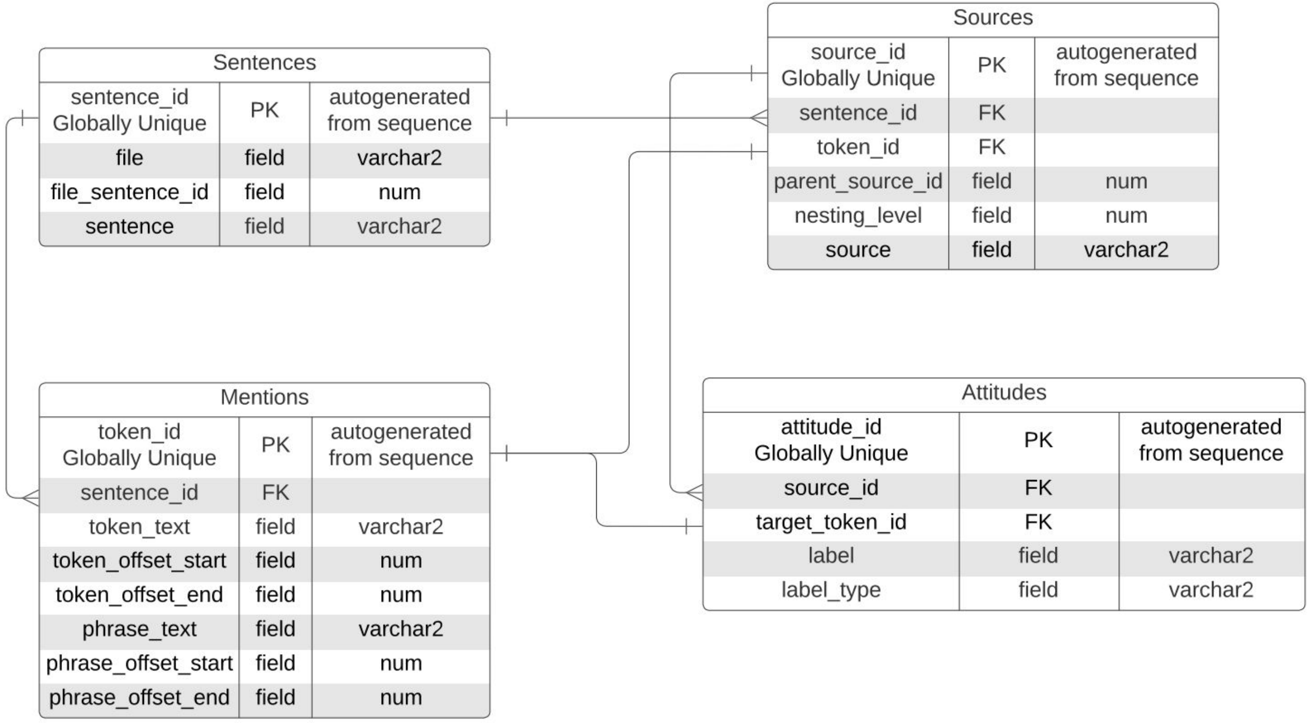
Factbank much more complex; impossible to (natively) combine!

Factbank: A Natively Relational Corpus

FactBank Native ERD (Unused Tables Ommitted)



Unified Database Model: Entity-Relation Diagram



Part IV: Data Transformations

How can we unify data representations across corpora?

Unified DB Model: Data Transformations

- **Goal:** Preserve native data while inserting synthetic data where gaps appear
  - Ex: MPQA has a reported belief class, Factbank does not

Unified DB Model: Data Transformations

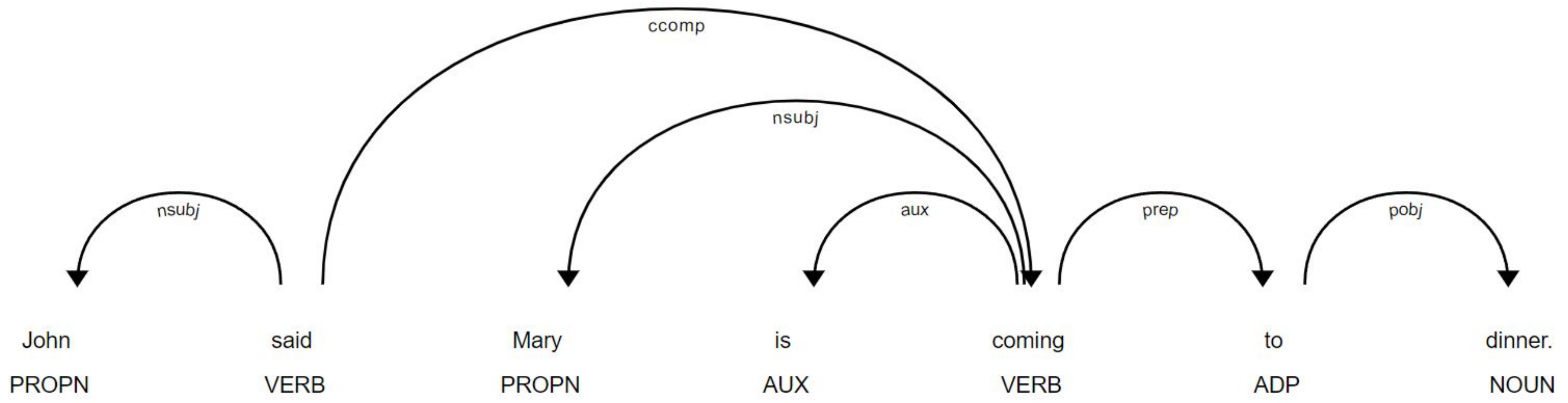
I. Unigram heads ↔ N-gram spans

   A. Parse trees! 🌳

II. Additional Classes (Factbank: ROB, LU: O)

Quick Aside: Dependency Parse Trees

- `spaCy` library in Python (`displayCy` 😁)



Unigram Heads ↔ N-Gram Spans

- **<u>Goal</u>:** Extract embedded proposition containing the target (noun or verb phrase)
    - Parse trees contain noun/verb phrases

- Factbank target head words live inside one of these phrases or may head it

Additional Classes

- Factbank: Reported Belief (ROB)
  - Natively grouped with Uu
- LU: Other (O)
  - Natively unannotated

Part V: LU Experiments, Results & Insights

Using our shiny new dataset for deep learning

Experimental Methods

Task	Model(s)	Paradigm	Example Input	Example Output
End-to-End	Flan-T5	Extraction	John said Mary is coming to dinner.	(coming, CB)
End-to-End	Flan-T5	Annotation	John said Mary is coming to dinner.	John said Mary is [coming|CB] to dinner.
Classification	Flan-T5, BERT	N/A	… Ex 1. coming|||John said Mary is coming to dinner. Ex 2. to|||John said Mary is coming to dinner. …	… Ex 1. CB Ex 2. O …

Experimental Methods

- Five-fold cross validation
- Average of three runs
  - Random seeds 7, 21 & 42
- Prediction Normalization *(Zhang et al.)*
- Python codebase *(Zhang et al., Murzaku et al.)*

Results

End-to-End Extraction Paradigm on Flan-T5

Classifier	Metric	Value
Committed Belief	Macro-F1	0.706
Average of All Labels	Macro-F1	0.405

Results

End-to-End Annotation Paradigm on Flan-T5

Classifier	Metric	Value
Committed Belief	Macro-F1	0.730 ⏫(+0.024)
Average of All Labels	Macro-F1	0.441 ⏫(+0.036)

Results

Classification on Flan-T5

Classifier	Metric	Value
Committed Belief	Macro-F1	0.746
Non-Committed Belief	Macro-F1	0.397
Non-Applicable	Macro-F1	0.611
Other	Macro-F1	0.967
Average of all labels except other	Macro-F1	0.585

Results

Classification on BERT

Classifier	Metric	Value
Committed Belief	Macro-F1	0.742 (-0.004)
Non-Committed Belief	Macro-F1	0.397 ⏫(+0.08)
Non-Applicable	Macro-F1	0.611 ⏫(+0.011)
Other	Macro-F1	0.967 (=)
Average of all labels except other	Macro-F1	0.585 ⏫ (+0.021)

Part VI: Future Work

What's next?

Future Work

I. Improve minority-class performance by leveraging few-shot learning and/or synthesizing additional data via keyword substitutions (WordNet)
II. Port additional corpora to unified format (MPQA, BEST)
III. Investigate performance of LLMs on head ↔ span task versus `spaCy` or some other dependency parsing tool

Thank You!