Conclusion

Colin Lockard, Prashant Shiralkar, Xin Luna Dong, Hannaneh Hajishirzi

Four Challenges

1. Diversity of data
2. Multiple modalities of text
3. Lack of training data
4. Unknown unknowns

Can we build a single extractor to find **consistent signals** across these diverse elements of data **across all modalities of text**?

Key Intuitions

- Diversity: Identifying consistent patterns
  - Leverage consistency in model/representation
  - Combining information from multiple modalities can give more consistent signals
- Lack of training data: Learning with limited labels
  - Find automated ways to label data
  - Employ weak or semi-supervision in limited labeled data settings
- Unknown unknowns: Stay open--Sacrificing granularity of knowledge representation allows for easier scaling

Unstructured Text: Short Answers

- **Consistency**
  - Model problem as text span classification and relationships between spans
  - Word embedding models help capture text semantics
- **Training data**
  - Weak supervision gives cheap training data
- **OpenIE**
  - Discovery of new types and relationships

Semi-Structured Text: Short Answers

- **Consistency**
  - Leverage general key-value pair consistency universal in templates
  - Leverage site-level consistency in layout and presentation
- **Training data**
  - Use distant supervision to generate cheap, but noisy training data
- **OpenIE**
  - Discover new relations by label propagation

Tabular text - Short Answers

- **Subject column detection**
  - Leverage generic features of subject entities such as value uniqueness, string type, number of characters and words
- **Column class detection**
  - Leverage external data -- web extracted triples, knowledge graph
- **Relation extraction between column pair**
  - Measure similarity between a column and entities of a type in a knowledge base

Multi-modal extraction: Short answers

- **Diversity**
  - Textual, layout, and visual signals can combine to form consistent patterns
- **Training data**
  - Multi-modal signals allow for accurate and easy creation of training data with Data Programming
- **OpenIE**
  - Visual semantics help make OpenIE extractions from semi-structured documents without prior knowledge of the subject domain

Future Directions - Unstructured text

- Full document understanding
  - Relation extraction beyond single sentence/paragraph
- Faster embedding models for scalability
- Non-English languages

Future Directions - Semi-structured text

- N-ary relations
- Relations not involving page topic

Future Directions - Tabular text

- Direct extraction (not relying on existing knowledge)

Future Directions - Multi-modal extraction

- Combine all signals from a document
- Make use of images
- Operate from jpgs, scanned pdfs
- Pre-training webpage representations
- Automated ontology construction
- Reproducible research
  - Webpage visual features depend on browser, CSS/JS availability, etc.

Outline

- Introduction (40 minutes)
- Part 1a: Unstructured Text (25 minutes)
- Part 1b: Unstructured Text: Methods (10 minutes)
- Live Q&A (15 minutes)
- Break (30 minutes)
- Part 2: Semi-structured and Tabular Text (40 minutes)
- Part 3: Multi-modal Extraction and Conclusion (35 minutes)
- **Live Q&A (15 minutes)**

Thank you!

https://sites.google.com/view/
acl-2020-multi-modal-ie

Please join us in the Zoom Chat!

___