Leveraging Disagreement for Structured Output Prediction in NLP

Daniil Ignatev

22.02.2024

Introduction to Structured Output Prediction in NLP

- Output is mostly binary tree-like.
- Key tasks: Dependency parsing, Constituency parsing, RST parsing.
- Objective: Improve accuracy and reliability of parsing techniques.

Parsing Techniques: A Comparative Overview

- Linearized trees: Transforming tree structures into a bracketed format.
- Parser actions: Sequential actions to construct parse trees.
- Bottom-up generation with LLMs: Leveraging large language models for labeling natural language spans; building from bottom to top.
- Unaccounted for: array-like linearization strategy.

Challenges in Structured Output Prediction

- Hard to learn from soft labels due to potential misalignment between EDUs (for RST); Using a soft loss for generation can lead to problematic results.
- Alternative ways to integrate disagreeing annotations: perspectivist adaptation / adaptation to data flaws.

Dissecting RST Parsing Complexities

- Disagreement in segmentation and tree structuring: mostly human error.
- Node labeling challenges: Dealing with underspecified relations and vague rules.
- Disagreements on different levels belong are of different kinds.

Preliminary experiments

- Methodology: Adapting Sheng et al. (2008) for RST parsing with mixed annotation approaches.
- Incorporation of special tokens to signify annotation styles.
- Data: Utilization of the RST-DT dataset, enriched with double annotations.

Experiment Metrics

ID	Special Tokens	Double Annot.	Metrics	
			Dev	**Test**
A	No	No	S: 0.9004, N: 0.7981, R: 0.6880, F: 0.6796	S: 0.8958, N: 0.7892, R: 0.6727, F: 0.6640
B	Yes	Yes	S: 0.9058, N: 0.8018, R: 0.6919, F: 0.6809	S: 0.8813, N: 0.7647, R: 0.6464, F: 0.6367

Table: Experiment metrics showing the impact of using special tokens and double annotations on model performance.

Note: Metrics cover Segmentation (S), Nuclearity (N), Relation (R), and Full (F) scores to provide a comprehensive view of parsing accuracy and improvements.

Pending Work

- Training LLM adapters for contrastive / collaborative decoding
- Possible modification: Learn brackets collaboratively and labels contrastively
- Qualitative analysis of experiments

Thank You for Your Attention