Amazon at MRP 2019: Parsing Meaning Representations with Lexical and Phrasal Anchoring

Amazon AI · AMAZON LEX

Jie Cao[†*] , Yi Zhang[‡] , Adel Youssef[‡] , Vivek Srikumar[†]

[†]School of Computing, University of Utah
[‡]AWS AI, Amazon

THE UNIVERSITY OF UTAH · UNLP

* Work done when Jie Cao was an intern at AWS AI – Amazon Lex Team.

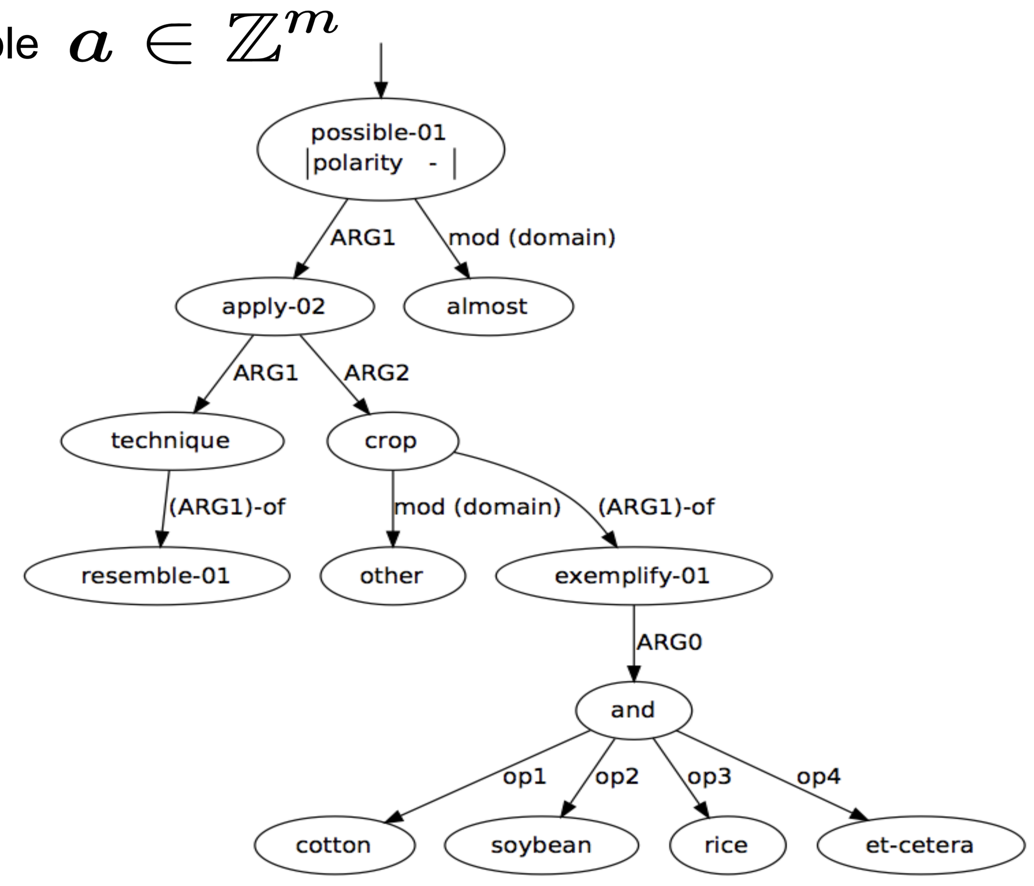
Introduction

- We study the parsing of five meaning representations by jointly modeling node/edge prediction with two types of anchoring:
  - Lexical-Anchoring and Phrasal-Anchoring(LAPA)
- Our graph-based model with latent-alignment mechanism can support both explicit and implicit lexical anchoring.
  - it ranks 1st place in AMR subtask, and 6th in PSD, 7th in DM
- Our constituent tree parsing model handles the phrasal anchoring in UCCA.
  - Equipped with self-attentive encoder and ELMo, our model achieved 5th in post-evaluation phase.

Lexical-Anchoring: Graph-based Parsing with Latent Alignment

For **m** words w, to predict concepts **C**, relations **R**, marginalize in the latent alignment **discrete** variable $a \in \mathbb{Z}^m$

$$\begin{aligned} P(C,R|w) &= \sum_a P(a)P(C,R|w,a) \\ &= \sum_a P(a)P(R|w,a,c)P(c|w,a) \\ &= \sum_a P(a)\prod_i^m P(c_i|h_{a_i})\prod_{i,j=1}^m P(r_{ij}|h_{a_i},c_i,h_{a_j},c_j) \end{aligned}$$



For DM, PSD, explicit alignment,
P(a*) = 1.0 and P(a!=a*) = 0.0

For AMR, latent alignment

- estimating posterior alignments model
- Variational Inference into ELBO
- Perturb-and-Max(MAP)
- Gumbel-Softmax

A similar technique is almost **impossible** to **apply** to other crops, **such as** cotton, soybeans and rice.

Preprocessing — Tokenizing, Lemmatizing, MWE Labeling, NER Labeling

Any Sequence Encoder with Any Embedding

Node Identification With decomposed label tuple

- Lemma Classifier (with copy): (NEG, possible, 02, N/A) <possible-02 :polarity ->
- Category Classifier (with copy): (FRAME, apply, 02, N/A) <apply-02>
- POS Classifier: (MWE, exemplify, 02, N/A) <exemplify-02>
- Sense Classifier

…

Two Separate Encoders for Head and Dep Node(with bi-lexicon) Encoding

Edge Identification Multiple Pass Biaffine Attention

<possible-02, apply-02>
<apply-02, possible-02>
<emplify-02, possible-02>
<possible-02, possible-02>

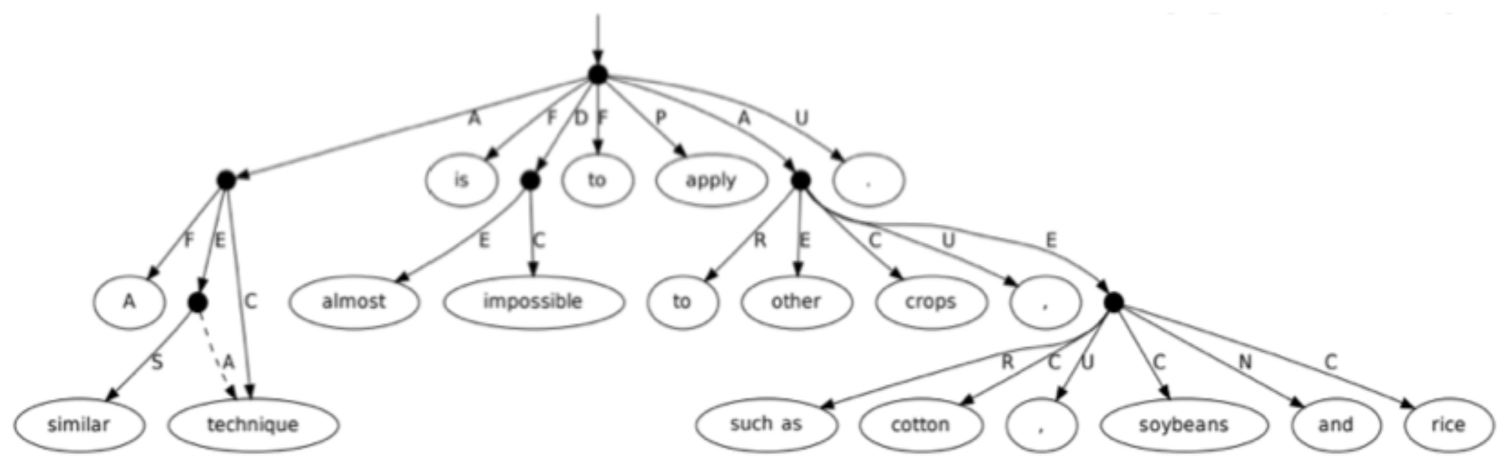
n * n, n is number of nodes → Deep Biaffine Classifier → $E \in \mathbb{R}^{n*n*r}$ (n × n), r, number of edge labels

Root Identification — Root Encoder with MLP(with anchoring word)

$$R \in \mathbb{R}^n$$

MSCG Connectivity — From root node, greedily select edges until all nodes are connected, force connecting some wrongly predicted NULL edge

Phrasal-Anchoring: CKY Parsing with Self-Attentive Encoder



- Assign edge label to dep non-terminal node label
- Remove 'remote' edge
- Ignoring discontinuous span
- 8 layers with 9 heads transformer encoder with positional encoding
- Span encoding with CKY

```
(TOP
    (HEAD
        (:A
            (:F  (TOK A))
            (:E  (:S (TOK similar)))
            (:C  (TOK technique)))
        (:F  (TOK is))
        (:D  (:E almost) (:C impossible))
        (:F  (TOK to))
        (:P  (TOK apply))
        (:A  (:R (TOK to))
             (:E (TOK other))
             (:C (TOK crops))
             (:U (TOK ,))
             (:C (TOK soybeans))
             (:N (TOK and))
             (:C (TOK rice))
        (:U  (TOK .))
    )
)
```

Post- Evaluation on UCCA

MR	Ours (P/R/F1)	Top 1/3/5 (F1)
UCCA(5)	80.83/73.42/**76.94**	81.67/77.80/73.22
EDS	N/A	94.47/90.75/89.10

For phrasal-anchoring models:

- Adding ELMo on Self-attentive encoder leads 3 points gaining
- 7-8 points worse than Top1 model on edge predicting.
- After assigning edge label as the label of the child node, how to Involve parent node information for span encoding or even using other span-based biaffine classifier worth to try in the future work.

Official Results and Analysis

All results here are evaluated on an official test set with unified MRP metric

Official Submission on AMR, DM, PSD

MR	Ours (P/R/F1)	Top 1/3/5 (F1)
AMR(1)	**75/71/73.38**	**73.38/71.97/71.72**
PSD(6)	89/89/**88.75**	90.76/89.91/**88.77**
DM(7)	93/92/92.14	94.76/94.32/93.74

Error Breakdown AMR

		data	tops	labels	prop	edges	all
TUPA single		all	63.95	57.20	22.31	36.41	44.73
		lpps	71.96	55.52	26.42	36.38	47.04
TUPA multi		all	61.30	39.80	27.70	27.35	33.75
		lpps	72.63	50.11	20.25	33.12	43.38
Ours(1)		all	65.92	82.86	**77.26**	63.57	**73.38**
		lpps	72.00	78.71	58.93	63.96	71.11
Top 2		all	78.15	82.51	71.33	63.21	72.94
		lpps	83.00	76.24	51.79	60.43	69.03

For lexical-anchoring models:

- Good at node label prediction, consistently better than other AMR model. Almost as good as Top models on DM and PSD.
- Worse on top node and edge prediction

Conclusion and Future work

- With latent alignment mechanism, our unified graph-based method can support both implicit and explicit lexical anchoring.
- Our AMR parser is especially good at predicting node properties, and it consistently performs better than other models on all the subcomponents in the graph, except for top node prediction.
- A multiple task learning method may benefit from the universal framework for DM, PSD, AMR. We leave this for future work.
- Moreover, we believe that multitask learning and pre-trained deep models such as BERT may also boost the performance of our parser in future
- Our span-based CKY parsing can partially resolve the phrasal-anchoring in UCCA. We also noticed that the span-based encoding for predicting edges miss information from the parent information.
- Phrasal-anchoring in EDS still requires more investigation.