Passing Parser Uncertainty to the Transformer
Labeled Dependency Distributions for NMT

Dongqi Liu

dongqi.me@gmail.com

Khalil Sima'an

k.simaan@uva.nl

University of Amsterdam

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Problem Statement:

Existing syntax-enriched NMT models have three main deficiencies:

- **Only** incorporate the **1-best (highest probability)** syntax tree.

- **Only** considered the **unlabeled** relation between words.

- **Other** methods (**data manipulation**, **linearization** or **embeddings) do not** help the Transformer to exploit the actual syntactic knowledge.

Hypothesis:
*Parser output is more useful if it conveys to the NMT model also its **remaining uncertainty**, rather than a mere forest or 1-best tree.*

Contribution:

A new method to infuse a labeled dependency distribution into the Transformer.

Findings:

- Improvement in both **lexical choice and word order** also for **long sentences.**

- **Highest results** on BLEU-4, compare with the 1-best tree, and baseline Transformer.

- Our approach may **guide** better attention weight computation and **improve** the **attention alignment** quality.

- **Outperforms** the baseline Transformer even when simply incorporating **labeled dependency distributions** as attention matrices.