Learning Latent Forests for Medical Relation Extraction

Zhijiang Guo[1*], Guoshun Nan[1*], Wei Lu[1] and Shay B. Cohen[2]

[1]StatNLP Research Group, Singapore University of Technology and Design

[2]ILCC, School of Informatics, University of Edinburgh

zhijiang_guo@mymail.sutd.edu.sg, guoshun_nan@sutd.edu.sg, luwei@sutd.edu.sg, scohen@inf.ed.ac.uk

Task

- **Input: A sentence from medical literature together with given entities.**

*Crystallographic analysis of the human **phenylalanine hydroxylase** catalytic domain with bound **catechol** inhibitors at 2.0 A resolution.*

- **Output: The relation between given entities**: Down regulator

Motivation

Dependency structures are often used for relation extraction as they are able to **capture long-range relations** that are only implicit in the surface form alone. However, dependency parsing accuracy is **relatively low** in the medical domain, which may downstream the relation extraction pipeline.

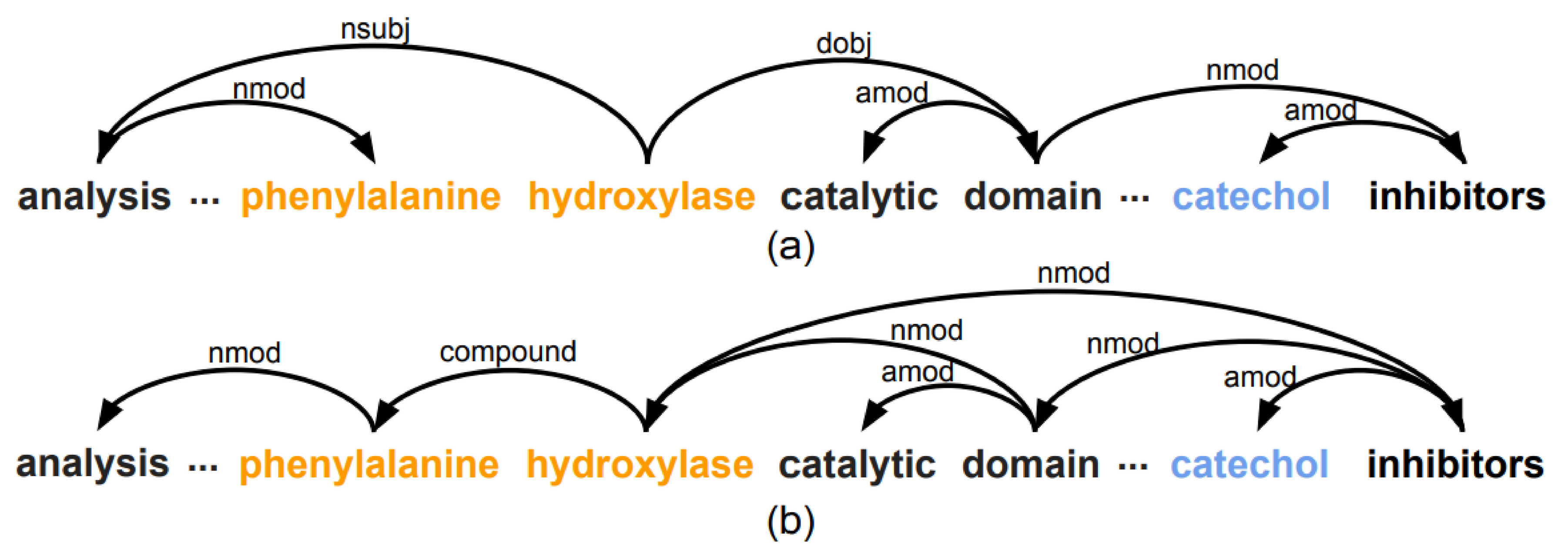
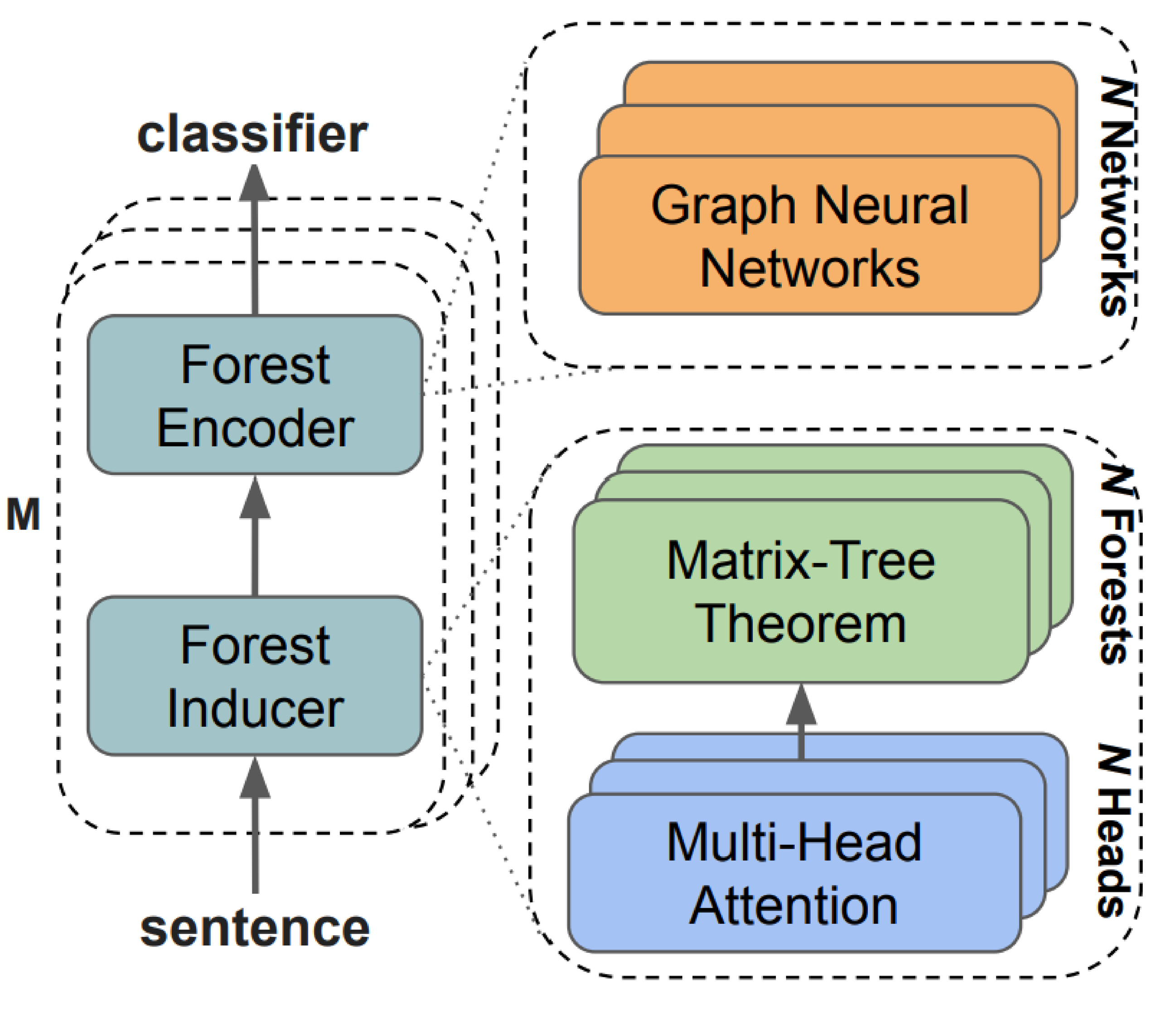


Figure (a) shows the 1-best dependency tree, **(b)** is the manually labeled tree.

- The 1-best dependency tree obtained by an off-the-shelf parser contains an error. Specifically, the entity phrase ***phenylalanine hydroxylase*** is broken since ***hydroxylase*** is mistakenly considered as the main verb.
- Unlike previous research efforts that rely on **dependency parsers trained on newswire text**, our model is able to generate **task-specific dependency structures** for capturing non-local interactions.

Model

- Our model consists of two modules: **Forest Inducer** and **Forest Encoder**. The forest inducer has two sub-modules, where the first one computes ***N*** attention matrices and the second sub-module takes the ***N*** attention matrices as inputs to obtain ***N dependency forests***. Then the forest encoder uses graph neural networks to encode the induced forests.

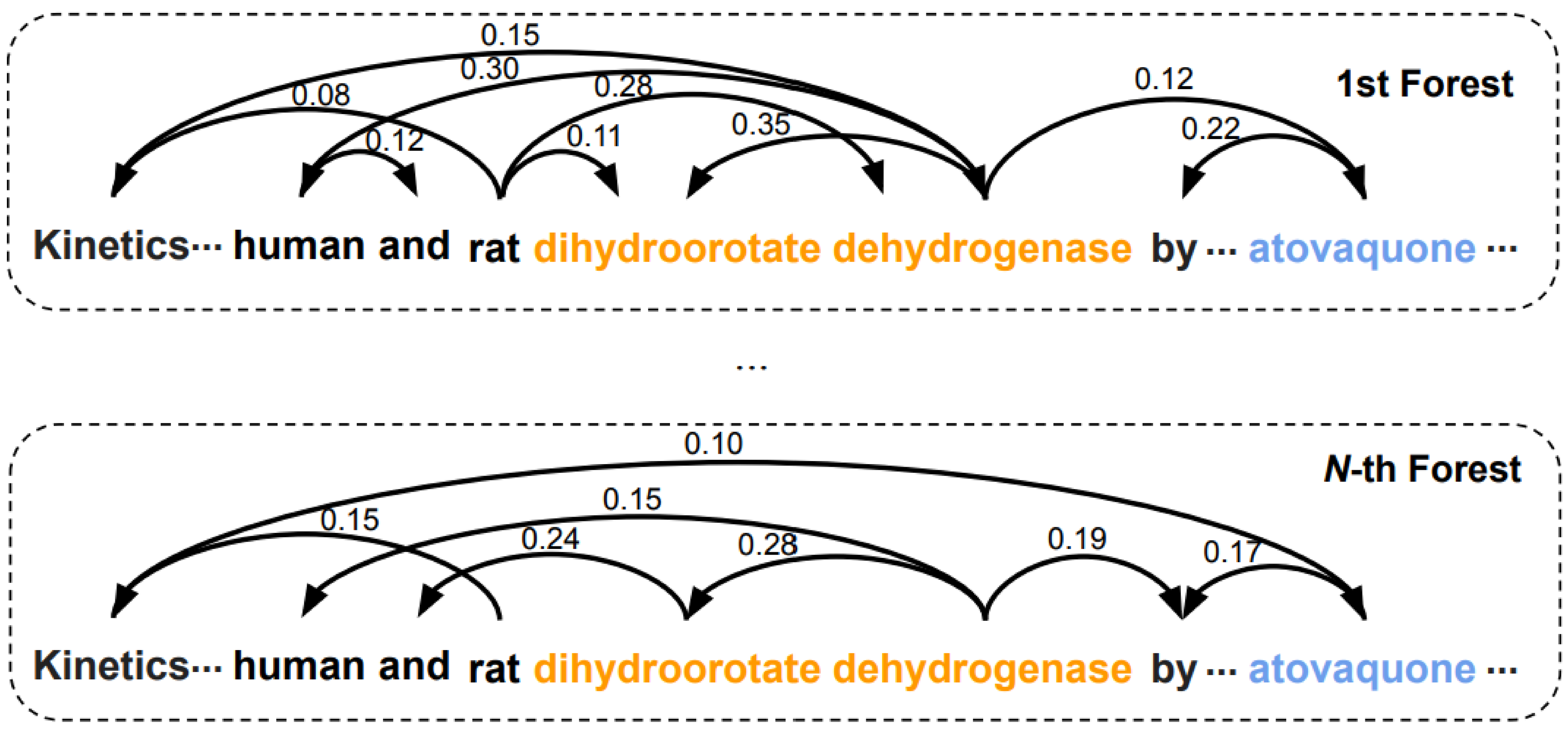


Latent Forests

Inspired by Liu and Lapata (2018), we use a **variant of Kirchhoff's Matrix-Tree Theorem** (MTT; Koo et al., 2007; Smith and Smith, 2007) to induce the latent structure of an input sentence. For the graph ***G***, MTT takes the edge scores and root scores as inputs then generates a latent forest by computing the **marginal probabilities for each edge:**

$$P(z_{ij}^k = 1) = \sum_{\mathbf{y} \in \mathbf{T}^k : (i,j) \in \mathbf{y}} P(\mathbf{y} | \mathbf{h}; \boldsymbol{\theta}^k)$$

Induced latent structure as shown below can be viewed as **multiple full dependency forests,** which efficiently represent all possible dependency trees within a compact and dense structure.



Experiments

- We evaluate our proposed model with four medical relation extraction datasets on two tasks, including cross-sentence *n*-ary relation extraction (Peng et al., 2017) and sentence-level relation extraction, CPR (Krallinger et al., 2017) and PGR (Sousa et al., 2019).

Syntax Type	Model	Binary-class				Multi-class	
		Ternary		Binary		Ternary	Binary
		Single	Cross	Single	Cross	Cross	Cross
Full Tree	DAG LSTM [Peng *et al.*, 2017]	77.9	80.7	74.3	76.5	-	-
	GRN [Song *et al.*, 2018]	80.3	83.2	83.5	83.6	71.7	71.7
	GCN [Zhang *et al.*, 2018]	84.3	84.8	84.2	83.6	77.5	74.3
Pruned Tree	GCN [Zhang *et al.*, 2018]	85.8	85.8	83.8	83.7	78.1	73.6
Forest	AGGCN [Guo *et al.*, 2019a]	87.1	87.0	85.2	85.6	79.7	77.4
	LF-GCN (Ours)	**88.0**	**88.4**	**86.7**	**87.1**	**81.5**	**79.3**

Syntax Type	Model	$F1$
None	BioBERT [Lee *et al.*, 2019]	67.2
Tree	BO-LSTM [Lamurias *et al.*, 2019]	52.3
	GCN [Zhang *et al.*, 2018]	81.3*
	Tree-GRN [Lifeng *et al.*, 2020]	78.9
Forest	Edgewise-GRN [Song *et al.*, 2019]	83.6
	KBest-GRN [Song *et al.*, 2019]	85.7
	AGGCN [Guo *et al.*, 2019a]	89.3*
	ForestFT-DDCNN [Lifeng *et al.*, 2020]	89.3
	LF-GCN (Ours)	**91.9**

Syntax Type	Model	$F1$
None	Random-DDCNN [Lifeng *et al.*, 2020]	45.4
	Att-GRU [Liu *et al.*, 2017]	49.5
	Bran [Verga *et al.*, 2018]	50.8
Tree	GCN [Zhang *et al.*, 2018]	52.2*
	Tree-DDCNN [Lifeng *et al.*, 2020]	50.3
	Tree-GRN [Lifeng *et al.*, 2020]	51.4
Forest	Edgewise-GRN [Song *et al.*, 2019]	53.4
	KBest-GRN [Song *et al.*, 2019]	52.4
	AGGCN [Guo *et al.*, 2019a]	56.7*
	ForestFT-DDCNN [Lifeng *et al.*, 2020]	55.7
	LF-GCN (Ours)	**58.9**

References

- Liu and Lapata. Learning structured text representations. TACL18.
- Koo et al., Structured prediction models via the matrix-tree theorem. EMNLP07.
- Smith et al., Probabilistic models of nonprojective dependency trees. EMNLP07.