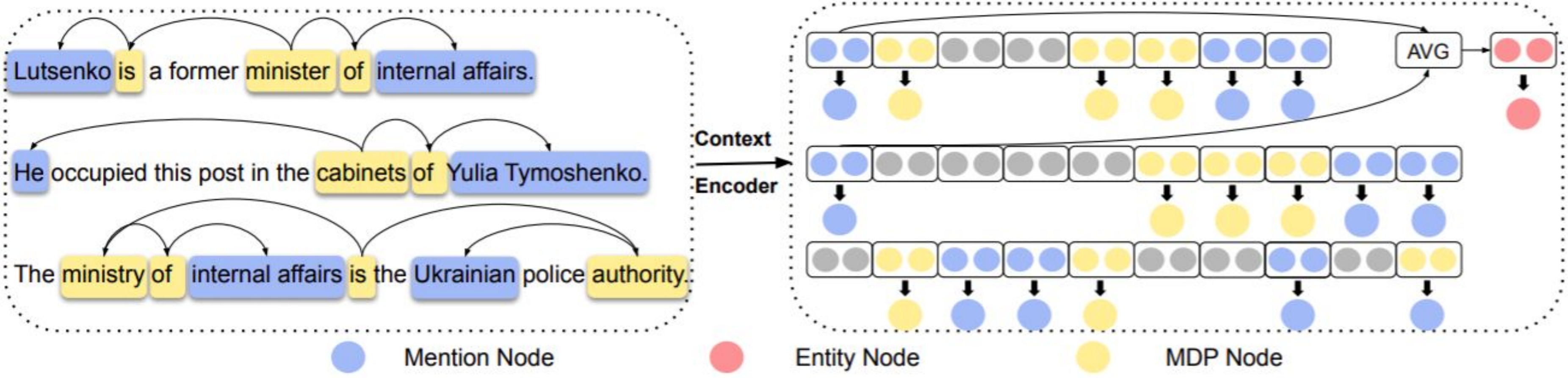
Reasoning with Latent Structure Refinement for Document-Level Relation Extraction

ACL 2020

Motivation

- Document-level relation extraction requires reasoning over the entity and relation mentions across sentences
- Previous graph-based models:
  - Rule based graph construction
  - Co-reference based graph construction
- In this paper:
  - Graph structure is learned end-to-end
  - Three types of nodes are encoded:
    - Mention
    - Entity
    - Tokens on dependency path
  - Graph structure is refined through multiple layers of graph constructions

Node Extraction Overview



Node Extraction

- Each sentence is separately encoded by a sequence encoder (e.g. LSTM or BERT):

$$\overleftarrow{\mathbf{h}_j^i} = \mathbf{LSTM}_l(\overleftarrow{\mathbf{h}^i}_{j+1}, \gamma_j^i)$$
$$\overrightarrow{\mathbf{h}_j^i} = \mathbf{LSTM}_r(\overrightarrow{\mathbf{h}^i}_{j-1}, \gamma_j^i)$$

- Three types of nodes are extracted:
  - Mentions: words in the document referring to entity
  - MDP: Words on the SDP between mentions in sentence
  - Entity: The average of mention representation of an entity

Structure Induction

- Compute scores for each pair of nodes:

$$\mathbf{s}_{ij} = (\tanh(\mathbf{W}_p \mathbf{u}_i))^T \mathbf{W}_b (\tanh(\mathbf{W}_c \mathbf{u}_j))$$

- Compute scores for each node to be root:

$$\mathbf{s}_i^r = \mathbf{W}_r \mathbf{u}_i$$

- Compute weight matrix based on the scores:

$$\mathbf{P}_{ij} = \begin{cases} 0 & \text{if } i = j \\ \exp(\mathbf{s}_{ij}) & \text{otherwise} \end{cases}$$

Structure induction

- Compute laplacian matrix from weight matrix:

$$\mathbf{L}_{ij} = \begin{cases} \sum_{i'=1}^{n} \mathbf{P}_{i'j} & \text{if } i = j \\ -\mathbf{P}_{ij} & \text{otherwise} \end{cases}$$

$$\hat{\mathbf{L}}_{ij} = \begin{cases} \exp(\mathbf{s}_i^r) & \text{if } i = 1 \\ \mathbf{L}_{ij} & \text{if } i > 1 \end{cases}$$

- Compute adjacency matrix from laplacian matrix:

$$\mathbf{A}_{ij} = (1 - \delta_{1,j})\mathbf{P}_{ij}[\hat{\mathbf{L}}^{-1}]_{ij} \\ -(1 - \delta_{i,1})\mathbf{P}_{ij}[\hat{\mathbf{L}}^{-1}]_{ji}$$

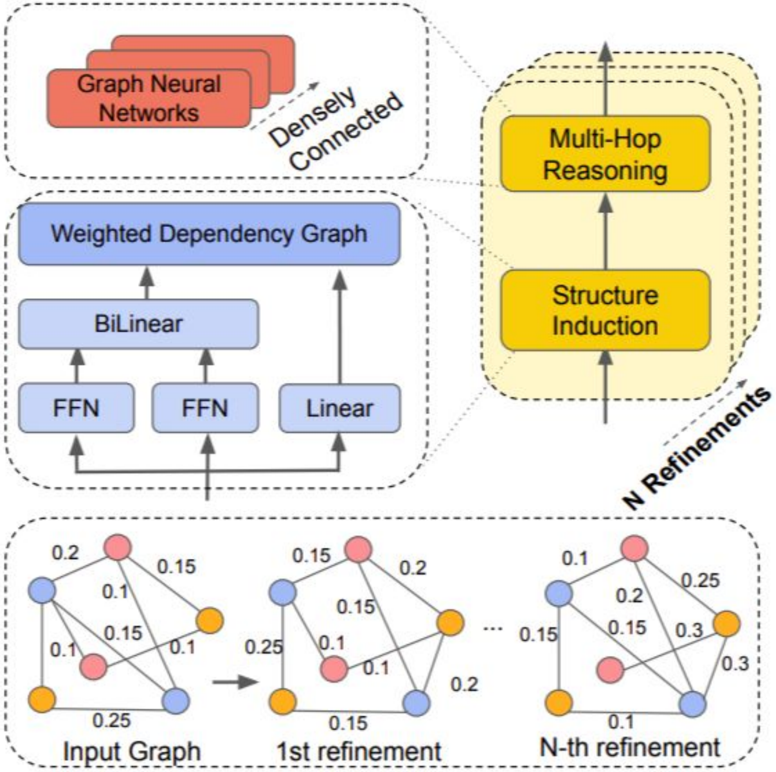
Graph Reasoning and Graph Refinement

- GCN encoder is employed to propagate node information in the induced graph

$$\mathbf{u}_i^l = \sigma\left(\sum_{j=1}^{n} \mathbf{A}_{ij}\mathbf{W}^l\mathbf{u}_i^{l-1} + \mathbf{b}^l\right)$$

- Dense connection is employed in GCN:
    - The input to each layer is combination of outputs of multiple previous layers
- Iterative refinement:
    - Stacking N layers of structure induction and graph reasoning
    - Early layers capture shallow dependencies in the graph and deeper layers extract more abstract connections

Mode Overview



Classification

- To classify the relation between two entities the representation of the corresponding nodes are used:

$$P(r|\mathbf{e}_i, \mathbf{e}_j) = \sigma(\mathbf{e}_i^T \mathbf{W}_\mathbf{e} \mathbf{e}_j + \mathbf{b}_e)_r$$

- Experiments on:
    - DocRed
    - CDR (biomedical domain)
    - GDA (biomedical domain)

Model	Dev				Test	
	Ign $F1$	$F1$	Intra-$F1$	Inter-$F1$	Ign $F1$	$F1$
CNN (Yao et al., 2019)	41.58	43.45	51.87*	37.58*	40.33	42.26
LSTM (Yao et al., 2019)	48.44	50.68	56.57*	41.47*	47.71	50.07
BiLSTM (Yao et al., 2019)	48.87	50.94	57.05*	43.49*	48.78	51.06
ContexAware (Yao et al., 2019)	**48.94**	51.09	56.74*	42.26*	48.40	50.70
GCNN ♣ (Sahu et al., 2019)	46.22	51.52	57.78	44.11	49.59	51.62
EoG ♣ (Christopoulou et al., 2019)	45.94	52.15	58.90	44.60	49.48	51.82
GAT ♣ (Veličković et al., 2018)	45.17	51.44	58.14	43.94	47.36	49.51
AGGCN ♣ (Guo et al., 2019a)	46.29	52.47	58.76	45.45	48.89	51.45
GloVe+LSR	48.82	**55.17**	**60.83**	**48.35**	**52.15**	**54.18**
BERT (Wang et al., 2019)	-	54.16	61.61*	47.15*	-	53.20
Two-Phase BERT (Wang et al., 2019)	-	54.42	61.80*	47.28*	-	53.92
BERT+LSR	**52.43**	**59.00**	**65.26**	**52.05**	**56.97**	**59.05**

Model	$F1$	Intra-$F1$	Inter-$F1$
Gu et al. (2017)	61.3	57.2	11.7
Nguyen and Verspoor (2018)	62.3	-	-
Verga et al. (2018)	62.1	-	-
Sahu et al. (2019)	58.6	-	-
Christopoulou et al. (2019)	63.6	68.2	50.9
LSR	61.2	66.2	50.3
LSR w/o MDP Nodes	**64.8**	**68.9**	**53.1**
Peng et al. (2016)	63.1	-	-
Li et al. (2016b)	67.3	58.9	-
Panyam et al. (2018)	60.3	65.1	45.7
Zheng et al. (2018)	61.5	-	-

Model	$F1$	Intra-$F1$	Inter-$F1$
NoInf (Christopoulou et al., 2019)	74.6	79.1	49.3
Full (Christopoulou et al., 2019)	80.8	84.1	**54.7**
EoG (Christopoulou et al., 2019)	81.5	85.2	50.0
LSR	79.6	83.1	49.6
LSR w/o MDP Nodes	**82.2**	**85.4**	51.1