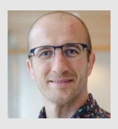
FACULTY OF ENGINEERING AND ARCHITECTURE

IDLab INTERNET & DATA LAB

INTERNET TECHNOLOGY AND DATA SCIENCE LAB

 **Klim Zaporojets**  **Johannes Deleu**  **Yiwei Jiang**  **Thomas Demeester**  **Chris Develder**

TOWARDS CONSISTENT DOCUMENT-LEVEL ENTITY LINKING: JOINT MODEL FOR ENTITY LINKING AND COREFERENCE RESOLUTION

INTRODUCTION

Entity Linking

- Discover named entity mentions (e.g., "NATO", "Alliance", etc.)
- Link the discovered mentions to Knowledge Base (e.g., Wikipedia) entries with the goal to *disambiguate*

Challenges

- Consistent decisions for coreferent entity mentions over the full document
- Coverage of candidate entities (e.g., "Alliance" without correct candidate in Figure 1)

Classical approaches

- Mention-dependent candidates, each mention limited to its own candidate table
- Focus on document-level coherent EL, but not enforced on structural (e.g., coreference) level
- Coreference-level coherence using inefficient Markov Logic-based models

Consistent Document-Level Entity Linking

General idea

- Enforce a single entity link for all the clustered (coreferent) mentions together
- Join entity linking candidates of all the clustered mentions, increasing thus the coverage

Proposed Algorithm

- Structured prediction task over directed trees
- Use globally normalized model (single joint loss)
- Using of Kirchhoff's Matrix-Tree Theorem algorithm ([1, 2]) to efficiently calculate the loss

MODEL

Local Model

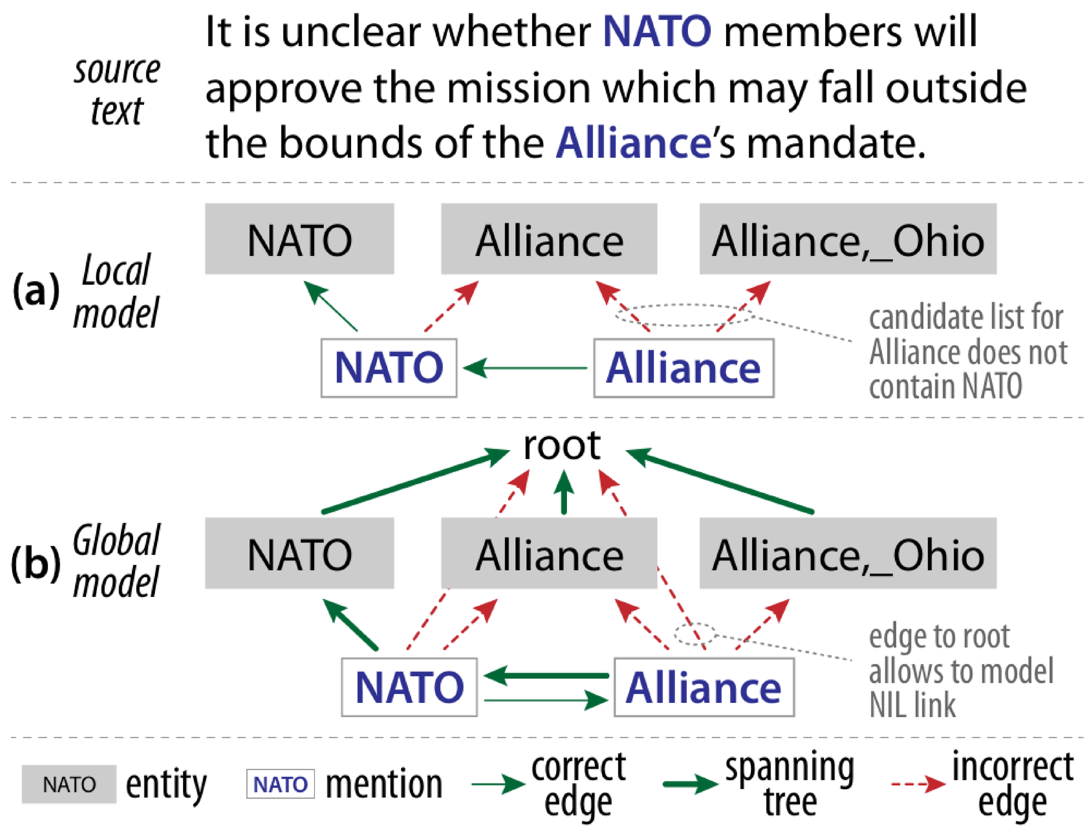
- Optimize marginalized probability of coreferent antecedents of each mention span as in [3, 4]
- Extend marginalization with span's candidate entity links

Global Model

- Expressive *maximum spanning tree* model that allows bi-directional connections between mentions
- Intractable naïve approach to identify all possible spanning trees, we resort to Kirchhoff's Matrix-Tree Theorem to efficiently solve this

$$\mathcal{L} = -\log \frac{\prod_{c\in \overbrace{\boxed{C}}^{\text{Ground truth clusters}}} \sum_{t\in \overbrace{\boxed{\mathcal{T}_c}}^{\text{Spanning trees in cluster } c}} \exp\left(\boxed{\Phi_{\mathrm{tr}}(t)}\right)}{\sum_{t\in \underbrace{\boxed{\mathcal{T}_{all}}}_{\text{All possible spanning trees}}} \exp\left(\underbrace{\boxed{\Phi_{\mathrm{tr}}(t)}}_{\text{Predicted score for tree } t}\right)} \underbrace{=}_{\text{Matrix-Tree Theorem}} -\log \frac{\prod_{c\in C}\det\left(\overbrace{\boxed{\hat{\boldsymbol{L}}_{r,c}}}^{\text{Modified Laplacian of cluster } c}(\Phi_{\mathrm{edges}})\right)}{\det\left(\underbrace{\boxed{\boldsymbol{L}_{r,r}}}_{\text{Minor of Laplacian in root } r}\left(\underbrace{\boxed{\Phi_{\mathrm{edges}}}}_{\text{Edge scores of the graph}}\right)\right)}$$

Figure 1: Illustration of the proposed graph models



RESULTS

Analysis 1: General Results (F1-score)

	DWIE		AIDA$_a^+$		AIDA$_b^+$	
Setup	Linking	Coref	Linking	Coref	Linking	Coref
Baseline	78.4	94.5	80.7	93.8	74.0	91.5
Local	83.4	94.4	83.1	94.7	75.8	**92.3**
Global	**83.9**	**94.7**	**83.7**	**95.1**	**76.0**	92.2

- Our joint models (*Local* and *Global*) achieve up to +5% F1-score on *entity linking* task compared to the baseline
- Less noticeable improvement on coreference resolution task
- On average, the *Global* joint model achieves the best performance

Analysis 2: Performance (F1-score) on singletons (S) and coreference clusters with multiple mentions (M)

	DWIE		AIDA$_a^+$		AIDA$_b^+$	
Setup	S	M	S	M	S	M
Baseline	80.4	69.5	82.9	70.7	77.0	57.0
Local	**82.6**	78.6	84.9	74.8	**79.8**	61.4
Global	**82.6**	**80.0**	**85.1**	**76.8**	79.3	**63.0**

- *Global* model achieves the most consistent predictions for clusters with multiple mentions (+10.5% F1-score)
- Less noticeable improvement for singletons (S) with similar performance of *Local* and *Global* models

Analysis 3: Performance on mentions without correct candidate

Setup	DWIE	AIDA$_a^+$	AIDA$_b^+$
Baseline	0.0	0.0	0.0
Local	41.7	27.4	26.9
Global	**57.6**	**50.2**	**29.7**

- Mention-based *baseline* cannot solve these cases: no correct candidate in the candidate list
- *Global* model performs best: robust in this challenging corner case.

Conclusions

- Both *joint* (*Global* and *Local*) architectures outperform the baseline on coreference and entity linking tasks
- *Global* model superior for clusters with multiple mentions and mentions without correct candidate entity

CONTACT:
klim.zaporojets@ugent.be –
http://klimzaporojets.github.io/

[1] William Tutte. 1984. Graph theory. Encyclopedia of Mathematics and its Applications, 21
[2] Terry Koo, Amir Globerson, Xavier Carreras, and Michael Collins. 2007. Structured prediction models via the matrix-tree theorem. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL 2007), pages 141–150.
[3] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. 2017. End-to-end neural coreference resolution. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017), pages 188–197.
[4] Kenton Lee, Luheng He, and Luke Zettlemoyer. 2018. Higher-order coreference resolution with coarse-to-fine inference. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018), pages 687–692.

GHENT UNIVERSITY

imec