Improving Entity Linking by Modeling Latent Relations between Mentions

Phong Le[1] and Ivan Titov[1,2]

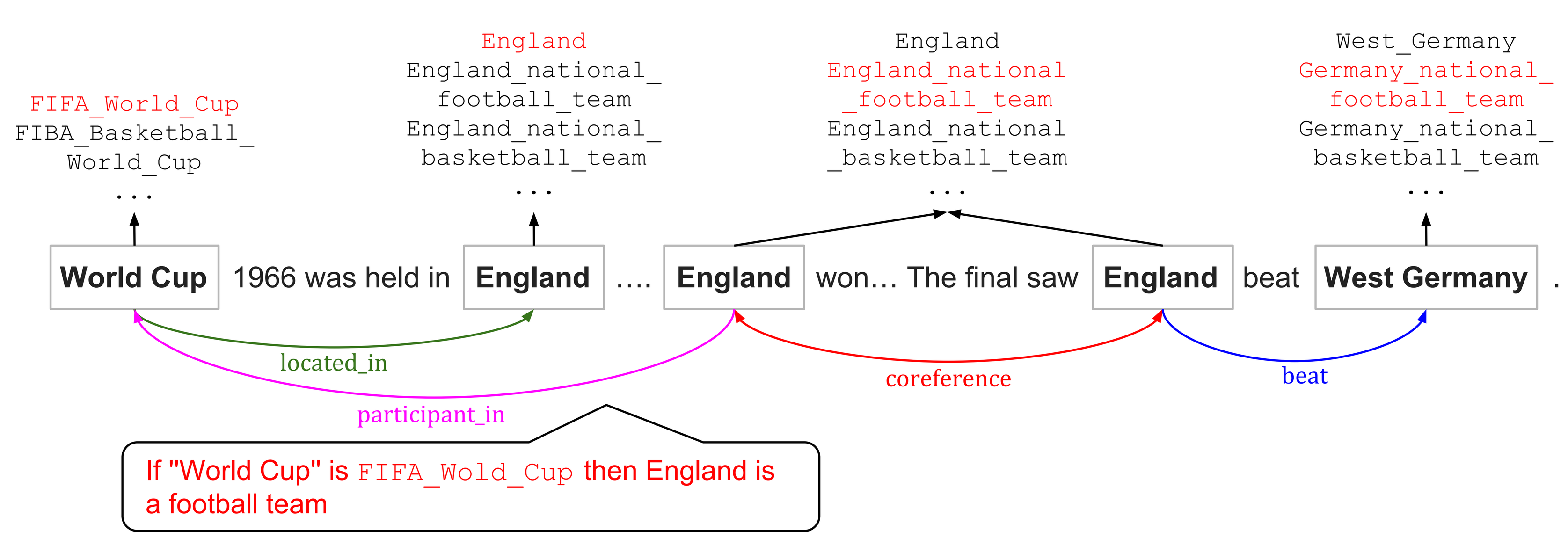
[1]University of Edinburgh [2]University of Amsterdam

Introduction

- Intuition: Relations between entities in a document help to link these entities to a knowledge base
- Previous work: Preprocess text to produce relations (e.g., using a co-reference system) [CR13]

Our work

We treat relations between entities as *latent variables* and induce them in such way as to help entity linking

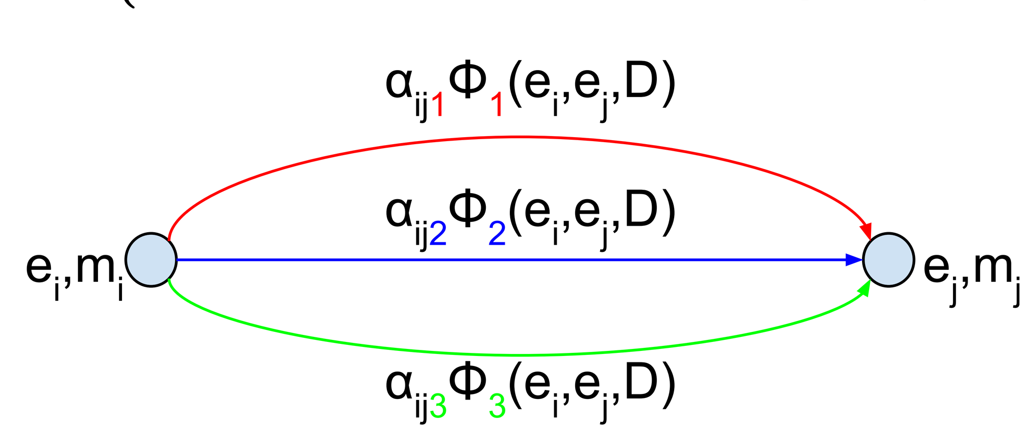


Our General Model: A CRF

We assign each mention m_i an entity e_i. Let $D = \{m_1, ..., m_n\}$ be a document and $E = \{e_1, ..., e_n\}$

- Our CRF:

$$q(E|D) \propto \exp\left\{ \sum_{i=1}^{n} \underbrace{\Psi(e_i, m_i)}_{\text{local score}} + \underbrace{\sum_{i \neq j} \sum_{k} \underbrace{\alpha_{ijk}}_{\substack{\text{how likely the relation}\\ \text{holds for the pair}}} \underbrace{\Phi_k(e_i, e_j, D)}_{\text{pair-wise score}}}_{\text{compatibility score}} \right\}$$

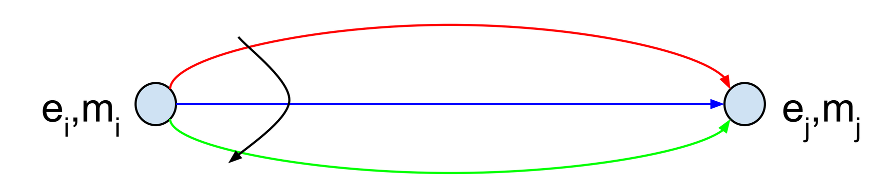


- Pair-wise score: $\Phi_k(e_i, e_j, D) = \mathbf{e}_i^T \underbrace{\mathbf{R}_k}_{\substack{\text{Relation embedding}\\ \text{A diagonal matrix}}} \underbrace{\mathbf{e}_j}_{\text{Entity embedding}}$

Relation weights: Two versions

- **Rel-norm** (Relation-wise normalization):

$$\alpha_{ijk} = \frac{\exp\left\{f^T(m_i)\mathbf{D}_k f(m_j)\right\}}{\sum_{k'} \exp\left\{f^T(m_i)\mathbf{D}_{k'} f(m_j)\right\}}$$

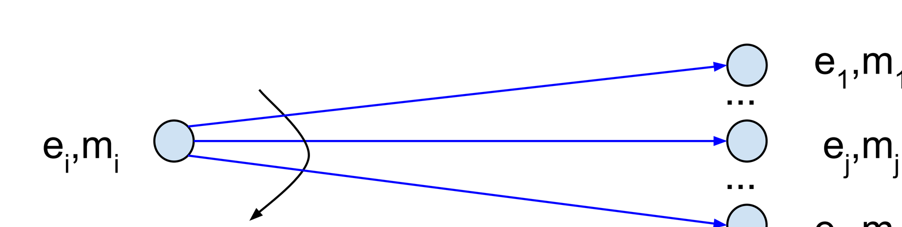


normalize over relations: $\alpha_{ij1} + \alpha_{ij2} + \alpha_{ij3} = 1$

Intuitively, α_{ijk} is the probability of assigning a k-th relation to a mention pair (m_i, m_j).

- **Ment-norm** (Mention-wise normalization):

$$\alpha_{ijk} = \frac{\exp\left\{f^T(m_i)\mathbf{D}_k f(m_j)\right\}}{\sum_{j'} \exp\left\{f^T(m_i)\mathbf{D}_k f(m_{j'})\right\}}$$



normalize over mentions: $\alpha_{i12} + \alpha_{i22} + ... + \alpha_{ij2} + ... + \alpha_{in2} = 1$

Similar to multi-head attention [VSP+17].

Estimation and Training

- Using Loopy Belief Propagation (LBP) [GH17]:

$$\hat{q}_i(e_i|D) \approx \max_{\substack{e_1,...,e_{i-1}\\ e_{i+1},...,e_n}} q(E|D)$$

The final score for each mention

$$\rho_i(e) = g(\hat{q}_i(e|D), \hat{p}(e|m_i))$$

where g is a 2-layer neural network. $\hat{p}$ is mention-entity hyperlink count statistics from Wikipedia, a large Web corpus and YAGO.

- We minimize the following ranking loss:

$$L(\theta) = \sum_{D \in \mathcal{D}} \sum_{m_i \in D} \sum_{e \in C_i} h(m_i, e)$$

$$h(m_i, e) = \max\left(0, \gamma - \rho_i(e_i^*) + \rho_i(e)\right)$$

Reference

[CR13] Xiao Cheng and Dan Roth, *Relational inference for wikification*, EMNLP, 2013.

[GH17] Octavian-Eugen Ganea and Thomas Hofmann, *Deep joint entity disambiguation with local neural attention*, EMNLP, 2017.

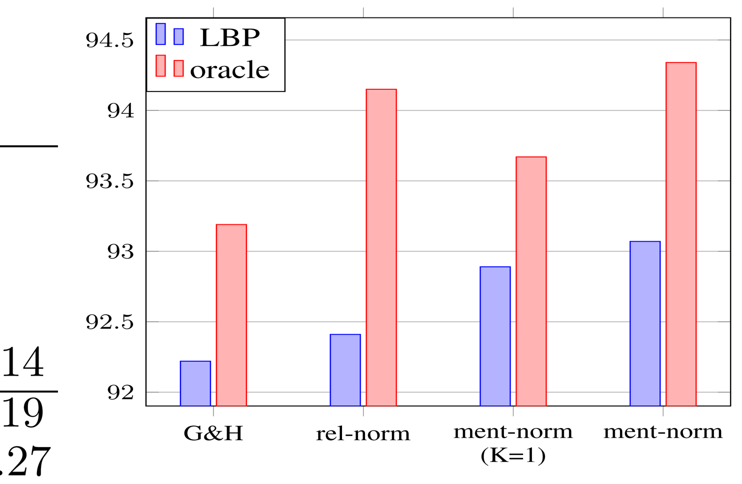
[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin, *Attention is all you need*, NIPS, 2017.

Experiments

F1 micro score. Mean and 95% confidence interval of 5 runs.

- In-domain:

Methods	Aida-B
Chisholm and Hachey (2015)	88.7
Guo and Barbosa (2016)	89.0
Globerson et al. (2016)	91.0
Yamada et al. (2016)	91.5
Ganea and Hofmann (2017)	92.22 ± 0.14
rel-norm	92.41 ± 0.19
ment-norm	**93.07** ± 0.27
ment-norm ($K = 1$)	92.89 ± 0.21

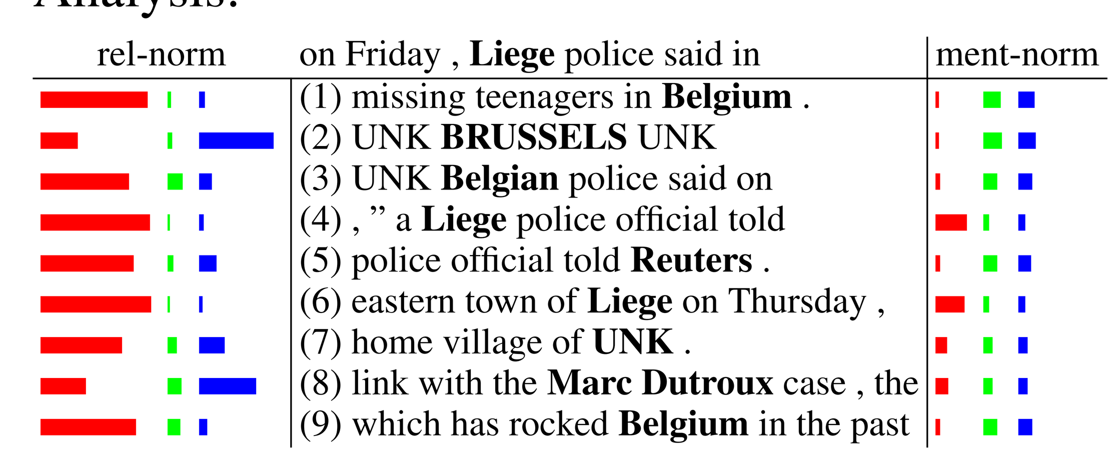


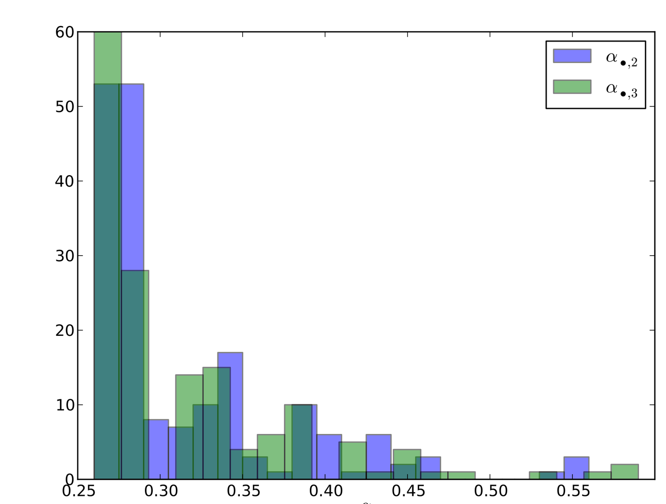
Rel-norm is more sensitive to prediction errors.

- Out-of-domain:

Methods	MSNBC	AQUAINT	ACE2004	CWEB	WIKI	Avg
Milne and Witten (2008)	78	85	81	64.1	81.7	77.96
Hoffart et al. (2011)	79	56	80	58.6	63	67.32
Ratinov et al. (2011)	75	83	82	56.2	67.2	72.68
Cheng and Roth (2013)	90	**90**	86	67.5	73.4	81.38
Guo and Barbosa (2016)	92	87	88	77	**84.5**	**85.7**
Ganea and Hofmann (2017)	93.7 ± 0.1	88.5 ± 0.4	88.5 ± 0.3	**77.9** ± 0.1	77.5 ± 0.1	85.22
rel-norm	92.2 ± 0.3	86.7 ± 0.7	87.9 ± 0.3	75.2 ± 0.5	76.4 ± 0.3	83.67
ment-norm	**93.9** ± 0.2	88.3 ± 0.6	**89.9** ± 0.8	77.5 ± 0.1	$\underline{78.0} \pm 0.1$	$\underline{85.51}$
ment-norm ($K = 1$)	93.2 ± 0.3	88.4 ± 0.4	88.9 ± 1.0	77.0 ± 0.2	77.2 ± 0.1	84.94

- Analysis:





- Hard to interpret relations induced by rel-norm
- Ment-norm: relation 1 is similar to coreference. Relation 2 and relation 3 complement relation 1, but they are quite different.

Conclusions

- Inducing multiple relations between entities is beneficial for entity linking.
- Our system does not use any supervision for relations and uses minimal amount of feature engineering.
- Future work: Injecting linguistic knowledge (discourse and syntax).

Source code

https://github.com/lephong/mulrel-nel