PAPER CODE

Distantly Supervised Relation Extraction with Sentence Reconstruction and Knowledge Base Priors

Fenia Christopoulou, Makoto Miwa, Sophia Ananiadou

efstathia.christopoulou@manchester.ac.uk

The University of Manchester 1824 · NaCTeM The National Centre for Text Mining · National Institute of Advanced Industrial Science and Technology AIST · AIRC · Toyota Technological Institute 豊田工大

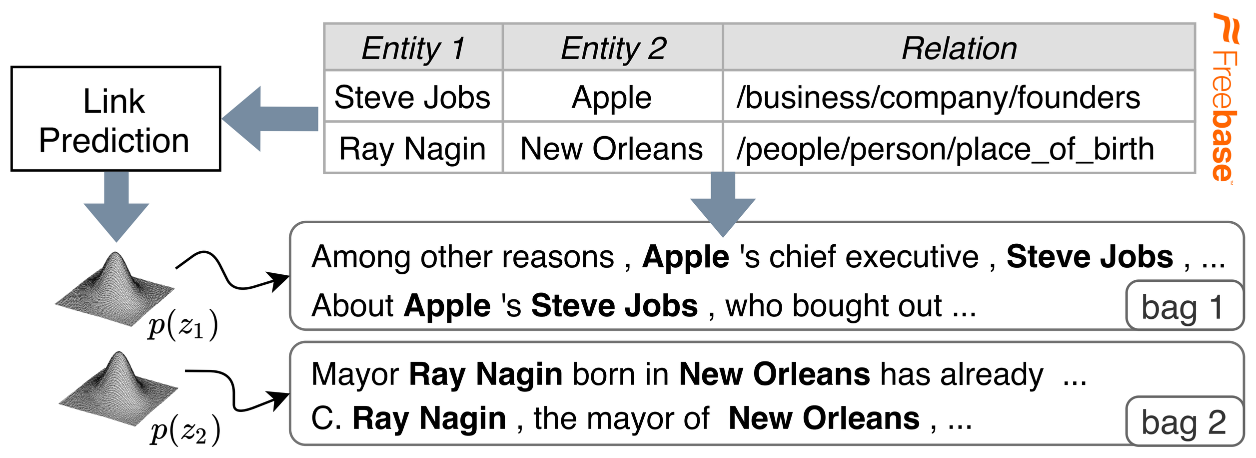
Introduction

- Using KB information in DSRE is beneficial for identifying factual knowledge.
- Prior methods:
  - Require explicit usage of KB embeddings into the model.
  - Force same predictions between KB and sentence pairs.

$\rightarrow$ *Inject signals from the KB to the target task via a probabilistic approach*

- A few methods follow a rather flawed evaluation setting, where several test pairs are included in the training set $\rightarrow$ exaggerated generalisability.

Bag-level Setting

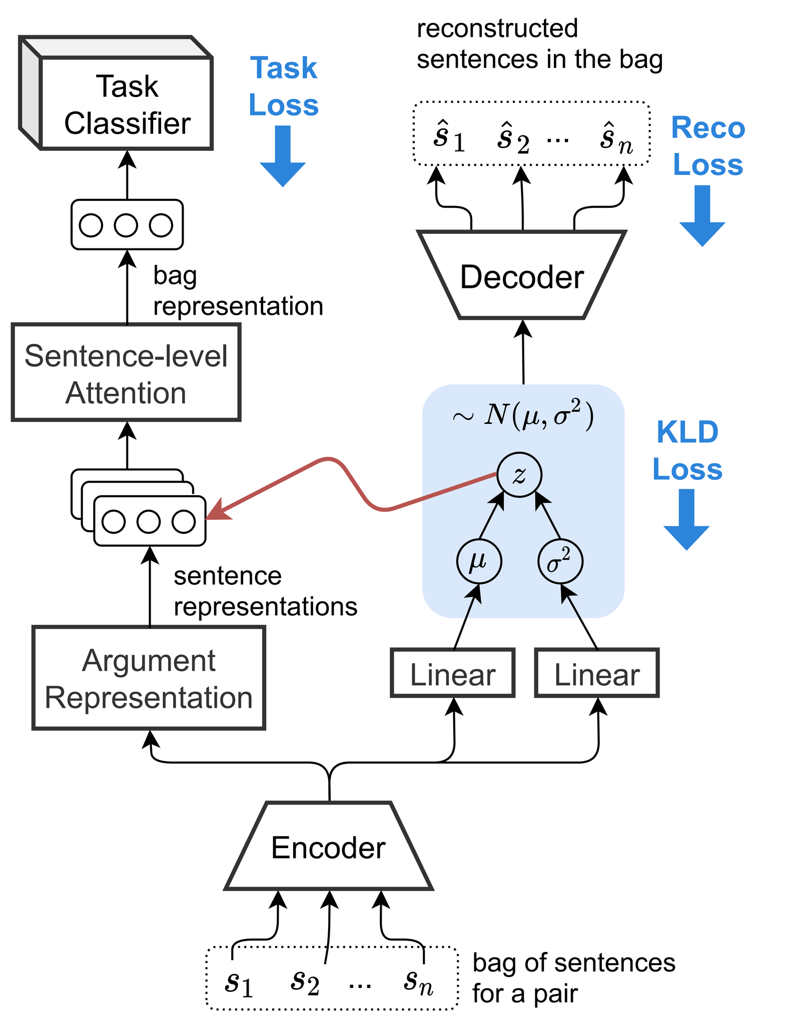


Main Idea: VAE & KB Priors

- Take advantage of two properties:
  - Reconstruction via encoder-decoder networks helps sentence expressivity.
  - Signals from the KB can assist detection of factual relations.
- Combine these two using a VAE together with a bag-level relation classifier.
- Each sentence's latent code can be close to the Normal [2] or to a prior distribution obtained from TransE embeddings:

$$p_\theta(\mathbf{z}) \sim \mathcal{N}(\mu_{\text{KB}}, \mathbf{I}), \text{ with } \mu_{\text{KB}} = \mathbf{e}_h - \mathbf{e}_t$$

Methodology



- The model consists of two branches:
  - Left branch: Bag classification
  - Right branch: Bag reconstruction
- Communication is achieved via the latent vector $\mathbf{z}$, constructed from the last hidden and cell state of the encoder:

$$\mu = \mathbf{W}_\mu[\mathbf{h}; \mathbf{c}] + \mathbf{b}_\mu, \quad \sigma^2 = \mathbf{W}_\sigma[\mathbf{h}; \mathbf{c}] + \mathbf{b}_\sigma$$

$$\mathbf{z} = \mu + \sigma \odot \epsilon, \text{ where, } \epsilon \sim \mathcal{N}(0, \mathbf{I}) \text{ or } \epsilon \sim \mathcal{N}(\mu_{\text{KB}}, \mathbf{I})$$

- Decoder is responsible for reconstructing each sentence in the bag:

$$\mathbf{h}'_0 = \mathbf{W}\mathbf{z} + \mathbf{b}, \quad \mathbf{x}'_t = [\mathbf{w}_t; \mathbf{z}],$$

- Classifier uses sentence representations and attention to form bag representations:

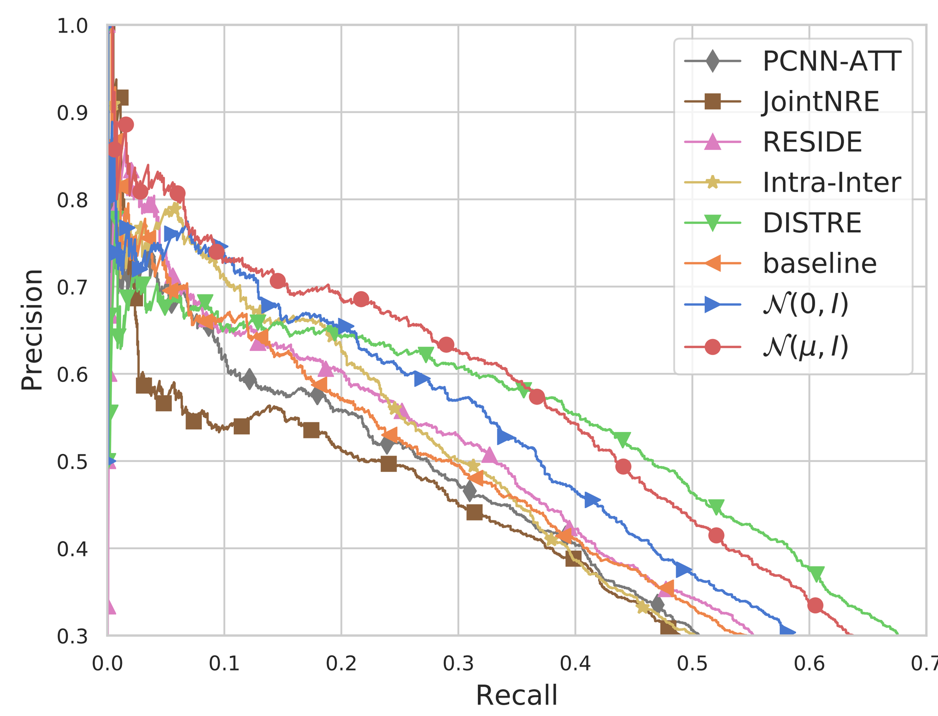
$$\mathbf{e}_i = \frac{1}{|e_i|} \sum_{k \in e_i} \mathbf{o}_k, \quad \mathbf{s} = \mathbf{W}_v[\mathbf{z}; \mathbf{e}_1; \mathbf{e}_2], \quad \mathbf{B}_r = \sum_{i=1}^{|B|} a_r^{(s_i)} \mathbf{s}_i,$$

- Total Loss: $L = \lambda\, L_{\text{BCE}} + (1 - \lambda) L_{\text{ELBO}}$

Results

NYT-10		
Method	520K AUC	570K AUC
Baseline	34.94	43.59
+ $p_\theta(z) \sim \mathcal{N}(0, I)$	38.59	44.64
+ $p_\theta(z) \sim \mathcal{N}(\mu_{\text{KB}}, I)$	**42.89**	**45.52**
PCNN-ATT [5]	32.66	36.25
JOINT NRE [3]	30.62	40.15
RESIDE [6]	35.80	41.60
INTRA-INTER BAG [7]	34.41	42.20
DISTRE [1]	42.20	-

WikiDistant				
Method	AUC	P@100	P@200	P@300
Baseline	28.54	94.0	93.0	88.3
+ $p_\theta(z) \sim \mathcal{N}(0, I)$	30.59	96.0	93.5	89.3
+ $p_\theta(z) \sim \mathcal{N}(\mu_{\text{KB}}, I)$	29.54	92.0	89.0	90.0
PCNN-ATT [4]	22.20	-	-	-
w/o non KB-prior pairs (72% of pairs preserved)				
Baseline	26.16	88.0	85.0	82.6
+ $p_\theta(z) \sim \mathcal{N}(0, I)$	27.46	90.0	88.0	84.6
+ $p_\theta(z) \sim \mathcal{N}(\mu_{\text{KB}}, I)$	28.38	94.0	95.0	89.3



Prior/Posterior Analysis

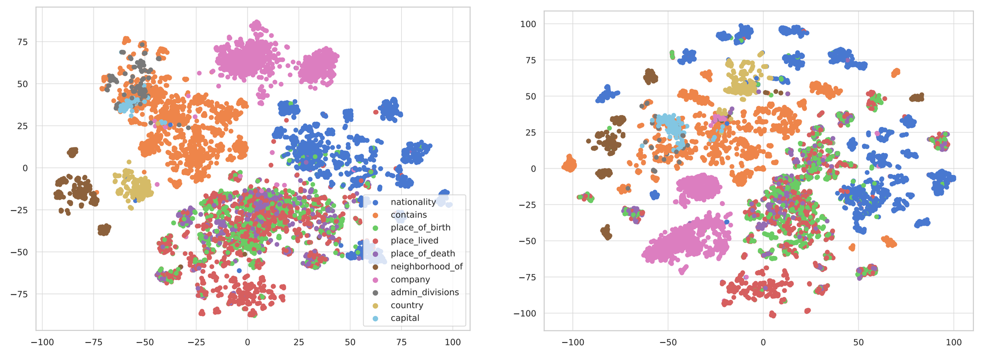


Figure 1: NYT-10 Priors and Posteriors

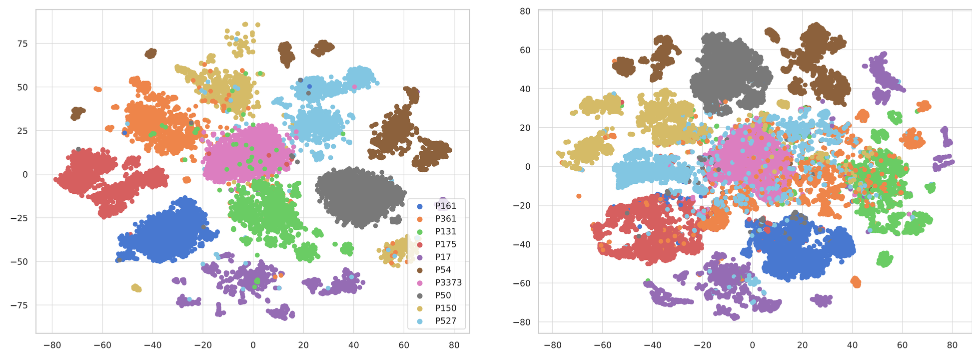


Figure 2: WikiDistant Priors and Posteriors

Sentence Reconstruction

- Sentence reconstruction with KB priors produces quite plausible sentences
- Both the Normal and KB prior distributions preserve topic-related words

Examples from WikiDistant:

	INPUT	ng 's first role was in the # *michael hui* comedy film " *the private eyes* " .
$\mathcal{N}(0, \mathbf{I})$	MEAN	the film was adapted into the # film ' the _ ' , directed by _ .
	SAMPLE	in # , he appeared in ' the _ ' , a # film adaptation of the same name by _ .
$\mathcal{N}(\mu_{\text{KB}}, \mathbf{I})$	MEAN	_ 's first film was ' the _ ' , starring _ and starring _ .
	SAMPLE	_ , who was the first female actress to win the academy award for best actress .

Conclusions/Takeaways

- The proposed approach brings close sentences that contain the same KB pairs.
- Our method does not require any external information during inference time.
- Jointly reconstructing sentences with relation classification is helpful for DSRE and KB priors further boost performance.
- We show that we can manipulate the space of sentences to match the space of KB triples, while reconstruction keeps topic-related terms.
- We are able to surpass a pre-trained GPT-2 model on the NYT10 dataset!

References

[1] C. Alt, M. Hübner, and L. Hennig. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. In *Proceedings of ACL*, pages 1388–1398, 2019.

[2] S. R. Bowman, L. Vilnis, O. Vinyals, A. Dai, R. Jozefowicz, and S. Bengio. Generating sentences from a continuous space. In *Proceedings of SIGNLL*, pages 10–21, 2016.

[3] X. Han, Z. Liu, and M. Sun. Neural knowledge acquisition via mutual attention between knowledge graph and text. In *Proceedings of AAAI*, 2018.

[4] X. Han, T. Gao, and Y. e. a. Lin. More data, more relations, more context and more openness: A review and outlook for relation extraction. In *Proceedings of AACL*, pages 745–758, 2020.

[5] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun. Neural relation extraction with selective attention over instances. In *Proceedings of ACL*, pages 2124–2133, 2016.

[6] S. Vashishth, R. Joshi, and S. S. e. a. Prayaga. RESIDE: Improving distantly-supervised neural relation extraction using side information. In *Proceedings of EMNLP*, pages 1257–1266, 2018.

[7] Z.-X. Ye and Z.-H. Ling. Distant supervision relation extraction with intra-bag and inter-bag attentions. In *Proceedings of NAACL*, pages 2810–2819, 2019.