中国科学院软件研究所学术年会'2021 暨计算机科学国家重点实验室开放周

学术论文

知识增强的特定领域小样本关系分类方法

Knowledge-Enhanced Domain Adaptation in Few-Shot Relation Classification

Jiawen Zhang, Jiaqi Zhu*, Yi Yang, Wandong Shi, Congcong Zhang and Hongan Wang

To appear in *the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2021), Research Track*

Contact: Jiaqi Zhu, zhujq@ios.ac.cn, 13683257241

Motivation

- **Challenges** of **Relation Classification (RC)**
  - The emergence of new relation types
  - Domain-specific annotated data are hard to access
  - The distant supervision brings a lot of noises

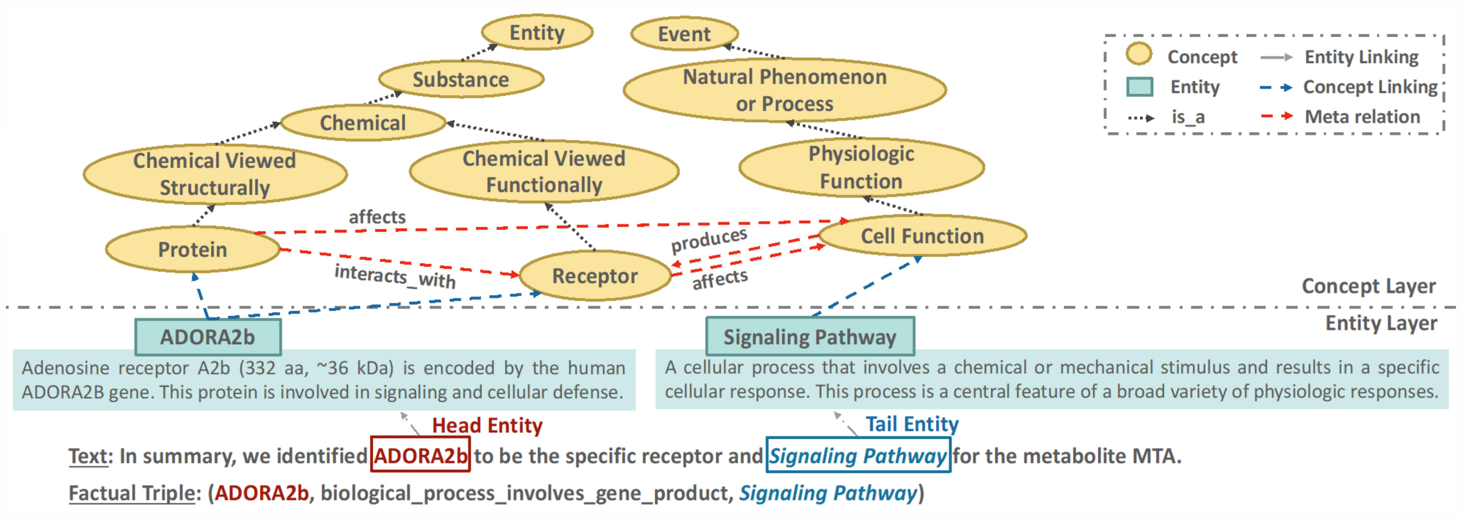
Regarding the RC task as a Few-Shot Learning problem
- Promising results in the general domain
- Poor domain adaptability

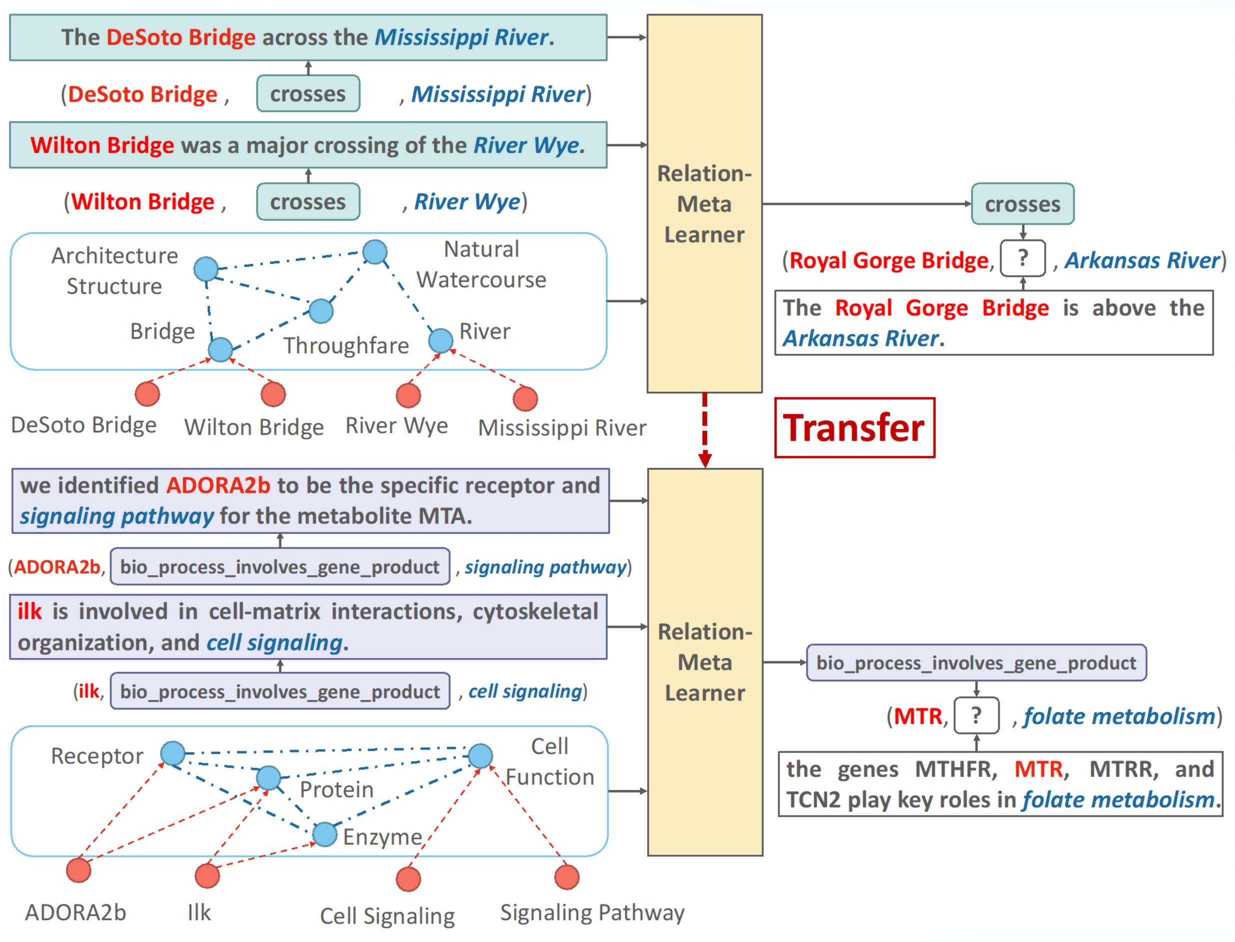
Table 1. Examples of 2-way 1-shot RC tasks

Training Task (Collected from Wikipedia)		
Support Set	(A) crosses	The **DeSoto Bridge** across the *Mississippi River*.
	(B) part_of	**Herm** is one of the *Channel Islands* in the English Channel.
Query		Jingkou District is one of three districts of **Zhenjiang** , *Jiangsu* province, China.
Testing task (Collected from Biomedical Literature)		
Support Set	(A) classified_as	These **tumors** are the most common *non-epithelial neoplasms* of gastric wall.
	(B) occurs_in	Aniridia is a rare **congenital** ocular disorder of complete or partial *iris hypoplasia*.
Query		The lateral lesions and **dental cysts**, especially *radicular cysts*, are compared.

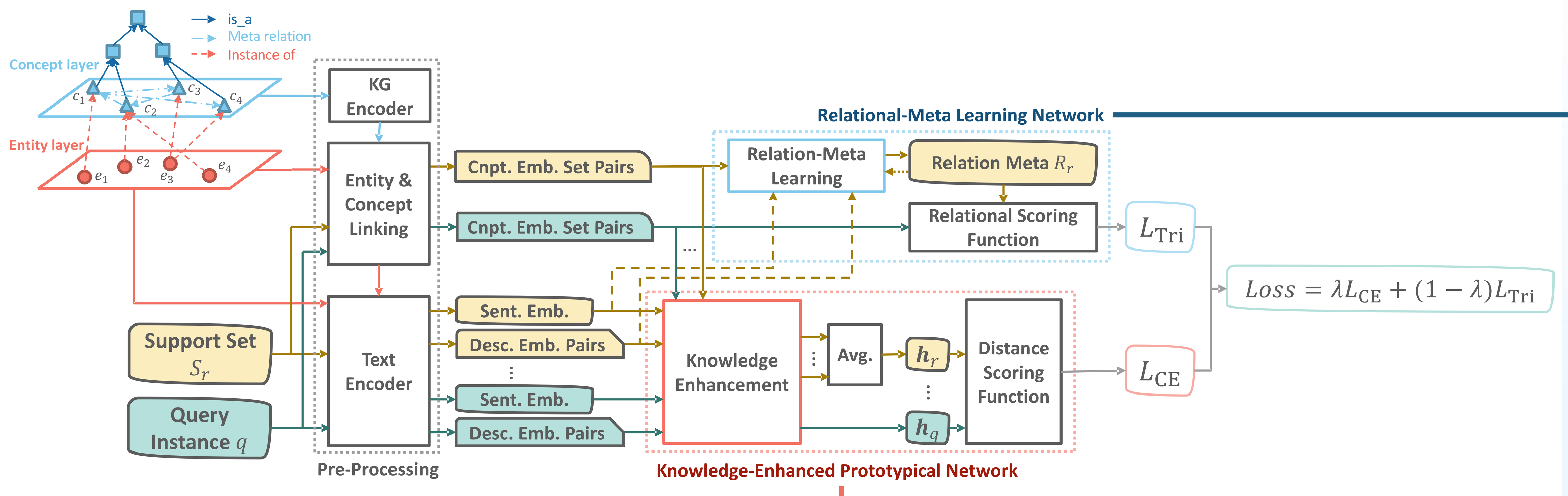
Main Ideas

1. Using open knowledge graphs (KGs) directly in downstream tasks
2. Utilizing the lightweight concept-level instead of the entity-level KGs
3. Summarizing the global semantics of relation types in addition to the instance-level knowledge enhancement
4. Treating the manner of using KGs as a kind of meta-information that can be transferred across tasks, even across domains





Proposed Model KEFDA



Instance Matching

- **Knowledge Enhancement Module**
  - Entity-level enhancement
    - $s = (x, e_1, e_2)$: Instance
    - $\boldsymbol{h}_x$: Contextual sentence representation
    - $\boldsymbol{h}_{e_1}^{\mathrm{des}}$ and $\boldsymbol{h}_{e_2}^{\mathrm{des}}$: Head-tail entity description features

$$\boldsymbol{h}' = \mathrm{FFNN}(\boldsymbol{h}_x \oplus \boldsymbol{h}_{e_1}^{\mathrm{des}} \oplus \boldsymbol{h}_{e_2}^{\mathrm{des}})$$

  - Concept-level enhancement
    - C_i: Concept set for entity e_i through concept linking
    - $\boldsymbol{h}_{C_i}$: Concept feature

$$\boldsymbol{h}_{C_i} = \frac{1}{|C_i|}\sum_{c \in C_i} \boldsymbol{h}_c,\ (i = 1,2)$$

$$\boldsymbol{h}_s = \boldsymbol{h}'' = \mathrm{FFNN}(\boldsymbol{h}' \oplus \boldsymbol{h}_{C_1} \oplus \boldsymbol{h}_{C_2})$$

- **Prototypical Network**[1]
  - $\mathcal{S}_r$: Support set
  - $\boldsymbol{h}_r$: Prototype (knowledge-enhanced relation type embedding)
  - $\boldsymbol{h}_q$: Knowledge-enhanced query instance embedding
  - I_r: Indicator of whether the relation r is the ground-truth result

$$\boldsymbol{h}_r = \frac{1}{|\mathcal{S}_r|}\sum_{s \in \mathcal{S}_r} \boldsymbol{h}_s$$

$$p_\phi(y = r|q) = \frac{\exp(-\mathrm{d}(\boldsymbol{h}_q, \boldsymbol{h}_r))}{\sum_{r' \in \mathcal{R}} \exp(-\mathrm{d}(\boldsymbol{h}_q, \boldsymbol{h}_{r'}))}$$

$$L_{\mathrm{CE}} = -\sum_{(\mathcal{R},\mathcal{S},q)\in\mathcal{T}_{\mathrm{Train}}} \sum_{r \in \mathcal{R}} I_r \log p_\phi(y = r|q)$$

Implicit Relation Matching

- **Relation-Meta Learner**
  - Concept-pair-specific relation meta $\boldsymbol{R}_j$
    - $\mathcal{P}_s = \{(c_1^j, c_2^j) | c_1^j \in C_1 \wedge c_2^j \in C_2\}$: Concept pairs

$$\boldsymbol{R}_j = \mathrm{FFNN}(\boldsymbol{h}_{c_1^j} \oplus \boldsymbol{h}_{c_2^j})$$

  - Instance-specific relation meta $\boldsymbol{R}_s$

$$\boldsymbol{R}_s = \sum_{(c_1^j, c_2^j) \in \mathcal{P}_s} p_j \boldsymbol{R}_j,\ p_j = \mathrm{softmax}(\alpha_j)$$

$$\alpha_j = -\max\left(-\mathrm{d}\left(\boldsymbol{h}_x, \boldsymbol{h}_{c_1^j}^{\mathrm{des}}\right), -\mathrm{d}(\boldsymbol{h}_x, \boldsymbol{h}_{c_2^j}^{\mathrm{des}})\right)$$

  - Relation-specific relation meta $\boldsymbol{R}_r$

$$\boldsymbol{R}_r = \frac{1}{|\mathcal{S}_r|}\sum_{s \in \mathcal{S}_r} \boldsymbol{R}_s$$

- **Relation-Meta Updater** (Inspired by MetaR[2])
  - Use gradient meta to make rapid corrections to the learned relation meta

$$g_r(c_1^j, c_2^j) = h_{c_1^j}^{\top} R_r h_{c_2^j}$$

$$L(\mathcal{S}_r) = \sum_{s \in \mathcal{S}_r} \sum_{(c_1^j, c_2^j) \in \mathcal{P}_s} \max(0, \gamma - g_r(c_1^j, c_2^j) + g_r(c_1^j, c_2'))$$

$$G_r = \nabla_{\boldsymbol{R}_r} L(\mathcal{S}_r),\ \boldsymbol{R}_r' = \boldsymbol{R}_r - \beta G_r$$

  - Calculate the total loss upon query instance

$$L_{\mathrm{Tri}} = \sum_{(\mathcal{R},\mathcal{S},q)\in\mathcal{T}_{\mathrm{Train}}} L(q)$$

Experiments

- **Dataset**: FewRel 2.0 Domain Adaptation (DA) challenge[3]
- **KGs**: WikiData (general domain), UMLS (medical domain)

Few-Shot RC Model	Avg.	5-Way 1-Shot	5-Way 5-Shot	10-Way 1-Shot	10-Way 5-Shot
Proto (CNN)	35.67	35.09	49.37	22.98	35.22
Proto (BERT)	38.75	40.12	51.50	26.45	36.93
Proto-ADV (BERT)	40.35	41.90	54.74	27.36	37.40
Proto-ADV (CNN)	43.54	42.21	58.71	28.91	44.35
BERT-PAIR	66.93	67.41	78.57	54.89	66.85
PAMN	78.98	77.54	90.40	65.98	82.03
DualGraph	81.83	80.11	91.01	73.89	82.34
GTP	82.18	80.04	92.58	69.25	86.88
KEFDA	**88.82**	**87.81**	**95.00**	**81.84**	**90.63**

- As the leading approach in the challenge, KEFDA dramatically improves the classification accuracy for all settings. It raises GTP, the best model so far except ours, by 6.63% on average

Few-Shot RC Model	Avg.	5-Way 1-Shot	5-Way 5-Shot	10-Way 1-Shot	10-Way 5-Shot
ERNIE	54.26	55.24	62.70	47.68	51.43
KEFDA-DistMult (-Desc. -Cnpt. -Meta.)	53.24	58.63	63.08	33.64	57.60
KEFDA-DistMult (-Cnpt. -Meta.)	66.53	72.95	68.58	59.59	64.98
KEFDA-DistMult (-Meta.)	87.52	85.55	93.75	**80.38**	90.40
KEFDA-RotatE	64.69	60.82	76.92	50.82	70.19
KEFDA-TransE	67.48	62.82	80.98	53.69	72.43
KEFDA-ANALOGY	86.85	85.58	94.30	78.84	88.69
KEFDA-DistMult	**87.69**	**86.18**	**94.38**	79.46	**90.77**

- The performance drops with the absence of each feature
- Concept features are most effective and significant
- Simple KG encoder which can handle multi-relational edges is better

Applications

New Knowledge Extraction: Discovering up-to-date knowledge from professional unstructured data which are updated and evolving over time

Knowledge Graph Updating: Updating existing KGs gradually and automatically based on domain-specific texts

[1] Snell Jake et al. "Prototypical Networks for Few-shot Learning". In NeurIPS. 2017, pp. 4077–4087.
[2] Mingyang Chen et al. "Meta Relational Learning for Few-Shot Link Prediction in Knowledge Graphs". In EMNLP-IJCNLP. 2019, pp. 4216–4225.
[3] Tianyu Gao et al. "FewRel 2.0: Towards More Challenging Few-Shot Relation Classification". In EMNLP-IJCNLP. 2019, pp. 6249–6254.