Open Hierarchical Relation Extraction

Kai Zhang[1], Yuan Yao[1], Ruobing Xie[2], Xu Han[1], Zhiyuan Liu[1], Fen Lin[2], Leyu Lin[2] and Maosong Sun[1]
[1]Tsinghua University
[2]WeChat Search Application Department, Tencent

NAACL-HLT 2021

Open Relation Extraction (OpenRE)

- Discover new relations from open-domain corpus

Head entity Tail entity

1. Jinnah left after his daughter Dina Wadia was born.
2. She married Polyctor, son of Aegyptus and Caliadne.

……

Father

1. His favorite film was Lawrence Kasdan's < French Kiss > (1995).
2. von Haselberg had a role in the Woody Allen film < Irrational Man >.

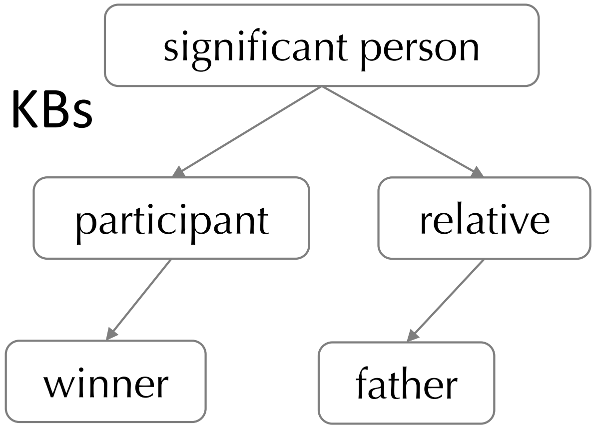
……

Director

- OpenRE is a clustering task

Relation Hierarchy

- In most modern KBs, relations are organized as hierarchies
- 1. crucial in establishing the relation schemes of KBs
- 2. better for users to understand
- 3. easy to utilize in downstream applications



A toy hierarchy

Open Hierarchical Relation Extraction

- **OpenRE is inherently in close connection with relation hierarchies**
- 1. Integrating information from hierarchies into OpenRE models:
  - better model relations' inter-dependencies
  - better extract new relations
- 2. Automatically adding novel relations to relation hierarchies:
  - less time-consuming
  - do not need expert knowledge

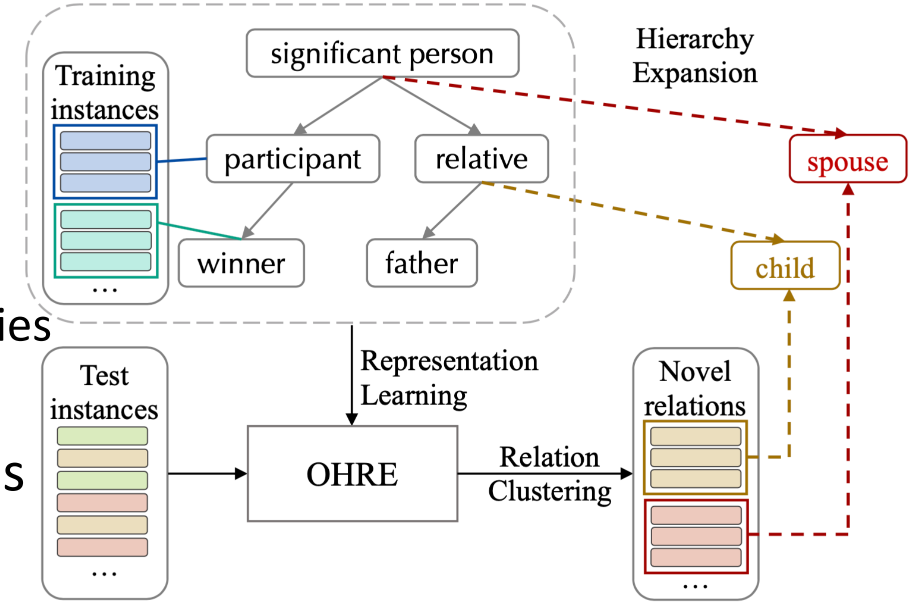


Figure 1: The workflow of OHRE framework. Trained with relation hierarchy and labeled instances, OHRE extracts novel relations from open-domain corpora and adds them into the existing hierarchy.

OHRE Framework

- Relation Representation Learning-----Better utilize hierarchical information
  - Relation Embedding Encoder
  - Dynamic hierarchical triplet objective
  - Hierarchical curriculum learning
  - Pair-wise Virtual Adversarial Training
- Relation Hierarchy Expansion---------- Better add novel relation to hierarchies
  - Relation Prototype Learning
  - Top-down hierarchy expansion algorithm

Representation Learning - Relation Embedding Encoder

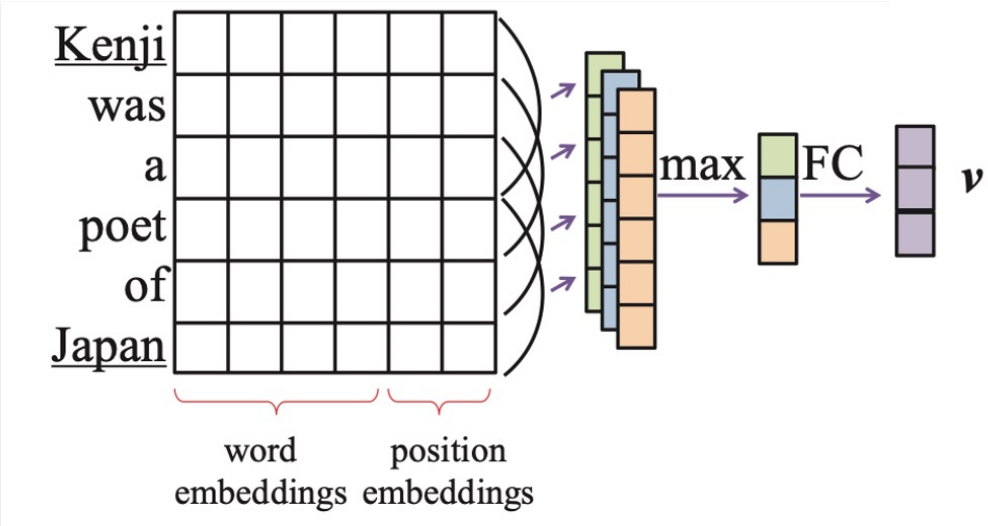


Figure 2: Relation Embedding Encoder

Relation Instance Encoder:

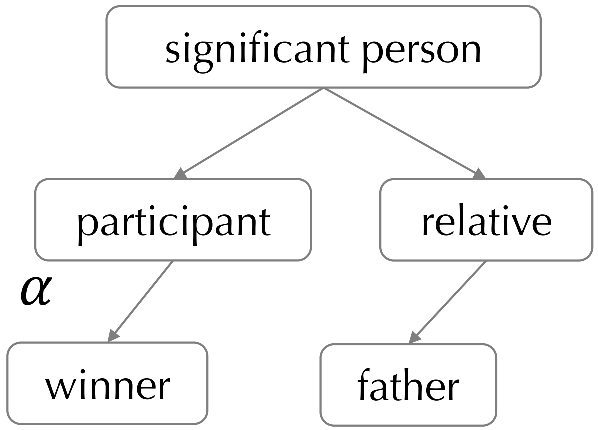
$$\boldsymbol{v} = \mathrm{CNN}(s, e_h, e_t). \quad \|\boldsymbol{v}\|_2 = 1$$

Measure distance between two representations

$$d(\boldsymbol{v}_1, \boldsymbol{v}_2) = \|\boldsymbol{v}_1 - \boldsymbol{v}_2\|_2^2$$

Representation Learning - Dynamic hierarchical triplet objective

- Given relation r_i and r_j, randomly sample:
  - anchor instance a, positive instance p from r_i
  - negative instance n from r_j
- $d(\boldsymbol{v}_a, \boldsymbol{v}_p) < d(\boldsymbol{v}_a, \boldsymbol{v}_n)$ at least dynamic margin α
  - margin $\alpha = \frac{l(r_i, r_j)}{1 + l(r_i, r_j)}$



$$\mathcal{L}_t = \sum_{r_i, r_j \sim T} \max[0, d(\boldsymbol{v}_a, \boldsymbol{v}_p) + \lambda_d \frac{l(r_i, r_j)}{1 + l(r_i, r_j)} - d(\boldsymbol{v}_a, \boldsymbol{v}_n)]$$

Representation Learning - Hierarchical curriculum learning

- Train model with relation pair from shallow layer (r_i^1 and r_j^1) to deep layer (r_i^2 and r_j^2).
- Use as a warm-up

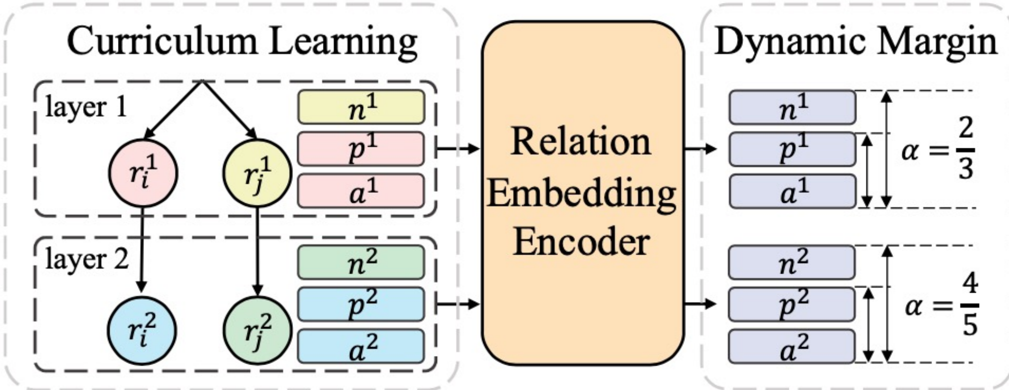


Figure 3: OHRE samples triplets from relations in hierarchy following a shallow-to-deep paradigm.

Representation Learning - Pair-wise Virtual Adversarial Training

- For each randomly sampled instance pair, we add worst-case perturbations and penalize the loss changes as follows:

$$\mathcal{L}_v = \sum_{v_1, v_2} \|d(\boldsymbol{v}_1, \boldsymbol{v}_2) - d(\tilde{\boldsymbol{v}}_1, \tilde{\boldsymbol{v}}_2)\|_2^2$$

- Final loss is defined as:

$$\mathcal{L} = \mathcal{L}_t + \lambda_v \mathcal{L}_v$$

Hierarchy Expansion – Relation Prototype Learning

- Cluster new relations from unsupervised data by Louvain
  - a graph cluster algorithm
  - automatically terminates clustering
- Prototype is represented with instances
  - prototype of a novel relation consists of all its instances
  - prototype of an existing relation contains all instances from itself and all children relations

Hierarchy Expansion – Top-down hierarchy expansion algorithm

- Similarity of two instances

$$w(\boldsymbol{v}_1, \boldsymbol{v}_2) = \max[0, 1 - d(\boldsymbol{v}_1, \boldsymbol{v}_2))]$$

- Similarity of two prototypes:

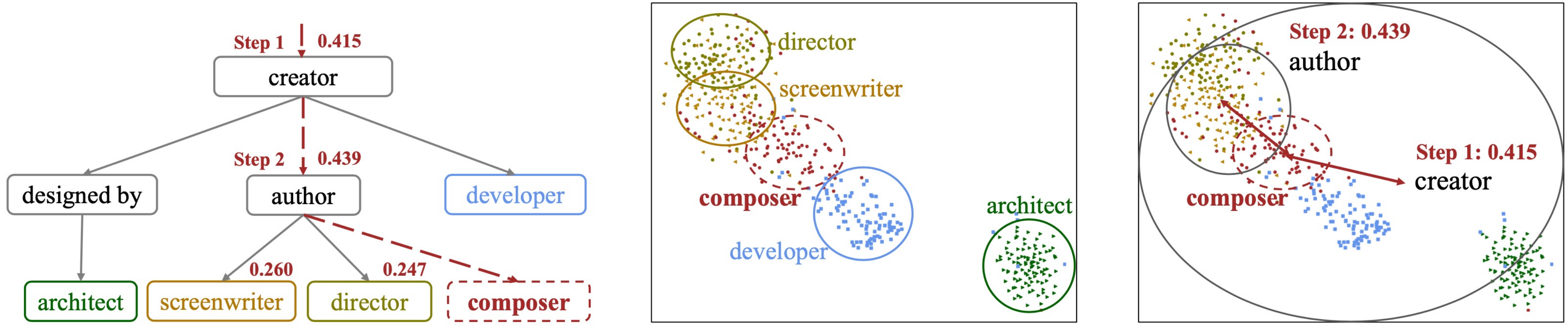
$$S(r_i, r_j) = \frac{\sum_{v_1 \in P_i} \sum_{v_2 \in P_j} w(\boldsymbol{v}_1, \boldsymbol{v}_2)}{|P_i| \cdot |P_j|} \cdot \sqrt{1 + |P_j^s|}$$

Algorithm 1 Top-Down Hierarchy Expansion

Require: r: A novel relation
Require: λ_W: Expansion threshold

Init search candidates C = root relations of trees
Init highest similarity in previous layer W = 0
while C not empty **do**
    Search relation $\hat{c} = \arg\max_{c \in C} S(r, c)$
    if $S(r, \hat{c}) > W$ **then**
        // Move to the next layer
        Update highest similarity $W = S(r, \hat{c})$
        Update search candidates C = children of $\hat{c}$
    else
        Stop searching
if $W \geq \lambda_W$ **then**
    Expand r as child of $\hat{c}$
else
    Cast r as singleton relation

Hierarchy Expansion – Top-down hierarchy expansion algorithm



(a) The expansion recommendation by OHRE on an existing relation hierarchy. The average connection score at each time step is shown.

(b) OHRE first clusters novel relations from open-domain corpora, and learns relation prototypes.

(c) OHRE then expands relation hierarchy based on relation prototypes in a top-down paradigm.

Experiments – Datasets

- FewRel Hierarchy
  - re-split from FewRel
  - a supervised dataset created from Wikipedia and Wikidata.
- NYT-FB Hierarchy
  - re-split from NYT-FB
  - a distantly supervised dataset created from New York Times and Freebase

		Relations	Instances
FewRel Hierarchy	Train	64	44,800
	Dev	16	1,600
	Test		1,600
NYT-FB Hierarchy	Train	212	33,992
	Dev	50	3,835
	Test		3,858

Experiments – Relation Clustering Results

Dataset	Model	B³ F1	B³ Prec.	B³ Rec.	V-measure F1	V-measure Hom.	V-measure Comp.	ARI
FewRel Hierarchy	VAE (Marcheggiani and Titov, 2016)	23.0	14.2	61.4	24.1	17.7	37.9	4.9
	RW-HAC (Elsahar et al., 2017)	32.7	28.0	39.4	39.7	36.0	44.4	12.4
	SelfORE (Hu et al., 2020)	60.6	60.1	61.1	70.1	69.5	70.7	54.6
	RSN-CV (Wu et al., 2019)	63.8	57.4	71.7	72.4	68.9	76.2	54.2
	OHRE	**70.5**	64.5	77.7	**76.7**	73.8	79.9	**64.2**
NYT-FB Hierarchy	VAE (Marcheggiani and Titov, 2016)	25.2	17.6	44.4	35.1	28.2	46.3	10.5
	RW-HAC (Elsahar et al., 2017)	35.0	43.3	29.4	58.9	61.7	56.3	28.3
	SelfORE (Hu et al., 2020)	38.1	42.6	34.5	59.0	60.7	57.5	30.4
	RSN-CV (Wu et al., 2019)	38.9	26.3	74.2	44.1	74.3	55.4	26.2
	OHRE	**43.8**	31.4	72.3	**60.0**	49.9	75.3	**31.9**

Table 1: Relation clustering results on two datasets (%).

Experiments – Relation Clustering Ablation study

1. hierarchical information is helpful in relation clustering.
2. reducing over-fitting with Pair-wise VAT is important.

Model	B^3 F1	V-F1	ARI
RSN-CV	63.8	72.4	54.2
w/o VAT	53.3	65.0	43.2
OHRE	**70.5**	**76.7**	**64.2**
w/o Dynamic Margin	68.9	76.1	63.5
w/o Curriculum Train	68.5	75.7	62.1
w/o Pair-wise VAT	58.3	68.8	49.5

Table 2: Ablation results on FewRel Hierarchy (%).

Experiments – Relation Expansion Results

Dataset	Method	Golden Cluster	Match F1	Match Prec.	Match Rec.	Taxonomy F1	Taxonomy Prec.	Taxonomy Rec.	Arith. F1	Harm. F1
FewRel Hierarchy	RW-HAC		33.2	33.9	37.6	37.5	37.5	37.5	35.3	35.2
	RSN-CV		69.6	63.7	85.8	34.5	38.5	31.3	52.0	46.1
	OHRE		**78.5**	73.6	88.4	**53.3**	57.1	50.0	**65.9**	**63.5**
	RW-HAC	✓	N/A			43.8	43.8	43.8	71.9	60.9
	RSN-CV	✓				37.5	37.5	37.5	68.8	54.5
	OHRE	✓				**57.4**	62.5	53.1	**78.7**	**73.0**
NYT-FB Hierarchy	RW-HAC		29.6	34.3	34.0	10.1	8.7	12.0	19.8	15.0
	RSN-CV		45.1	33.2	83.1	10.5	15.2	8.0	27.8	17.0
	OHRE		**51.7**	42.7	76.2	**22.3**	23.9	21.0	**37.0**	**31.2**
	RW-HAC	✓	N/A			20.0	16.7	25.0	60.0	33.3
	RSN-CV	✓				13.0	16.0	11.0	56.5	23.1
	OHRE	✓				**23.0**	23.0	23.0	**61.5**	**37.4**

Table 3: Hierarchy expansion results. Golden cluster indicates the golden relation clusters are given, in which case matching metric for relation clustering is not applicable. Arith: arithmetic mean, Harm: harmonic mean.

Experiments – Different Relation Topologies Results

- single relation without a parent (6 relations)
- relation with a parent in train set (8 relations)
- relation with a parent in test set (2 relations)

	Relation Clustering			Hierarchy Expansion		
	sgl.	p-trn.	p-tst.	sgl.	p-trn.	p-tst.
RW-HAC	31.6	35.0	42.8	**60.0**	0.0	0.0
RSN-CV	67.1	77.8	**64.4**	58.8	0.0	0.0
OHRE	**75.2**	**84.6**	53.9	58.8	**36.4**	0.0

Table 4: Relation clustering (B^3 F1) and hierarchy expansion (Taxonomy F1) results on relations in different hierarchy topologies. sgl.: relations without a parent, p-trn.: parent in train set, p-tst.: parent in test set.

Future Work

- Design methods to model the **global** interaction between new relations and hierarchy
- Develop end-to-end models to **jointly** optimize clustering and expansion stages for better results.

Thank you!

Code and Data: https://github.com/thunlp/OHRE

Contact: drogozhang@gmail.com