Are Noisy Sentences Useless for Distant Supervised Relation Extraction ?

汇报人：张佳雯 学号：51205901101 汇报时间：2020/12/10

Motivation

- Relation Extraction is a task that extracts structured relations from unstructured text in NLP.
  - Challenge: lack of large-scale manually labeled data
- Distant supervision is used to automatically construct training data.
  - Assumption: if two entities *(e1, e2)* have a relationship *r* in knowledge graph, then any sentence that mentions the two entities might express the relation *r*.
  - Problem: too strong, the noisy labeling problem

	Sentence	Bag Label	Noise?	Correct Label
Bag	#1: **Barack Obama** was born in the **United States**.	president of	Yes	born in
	#2: **Barack Obama** was the first African American to be elected to the president of the **United States**.		No	president of
	#3: **Barack Obama** served as the 44th president of the **United States** from 2009 to 2017.		No	president of

Table 1: An example of sentence-bag annotated by distant supervision. "Yes" and "No" indicate whether or not each sentence is a noisy sentence. "Correct Label" means the true relationship between the entity pair expressed in each sentence.

Motivation

- Previous studies: Multi-Instance Learning (MIL) framework
  - Categories: the soft decision methods & the hard decision methods
  - Problem: ignore the essential cause of the noisy labeling problem —— the lack of correct relation labels
- Method: utilize noisy sentences
  - The negative influence of noisy sentences is reduced.
  - The number of useful training data is increased.

Contributions

➢In this paper, we propose a novel **D**eep **C**lustering based **R**elation **E**xtraction model, named DCRE.

- Try to convert the noisy sentences as useful training data
- It is the first work to apply unsupervised deep clustering to obtain more appropriate relation labels for noisy sentences.
- Extensive experiments show that our model outperforms the state-of-the-art baselines, and can effectively alleviate the noisy labeling problem.

Method

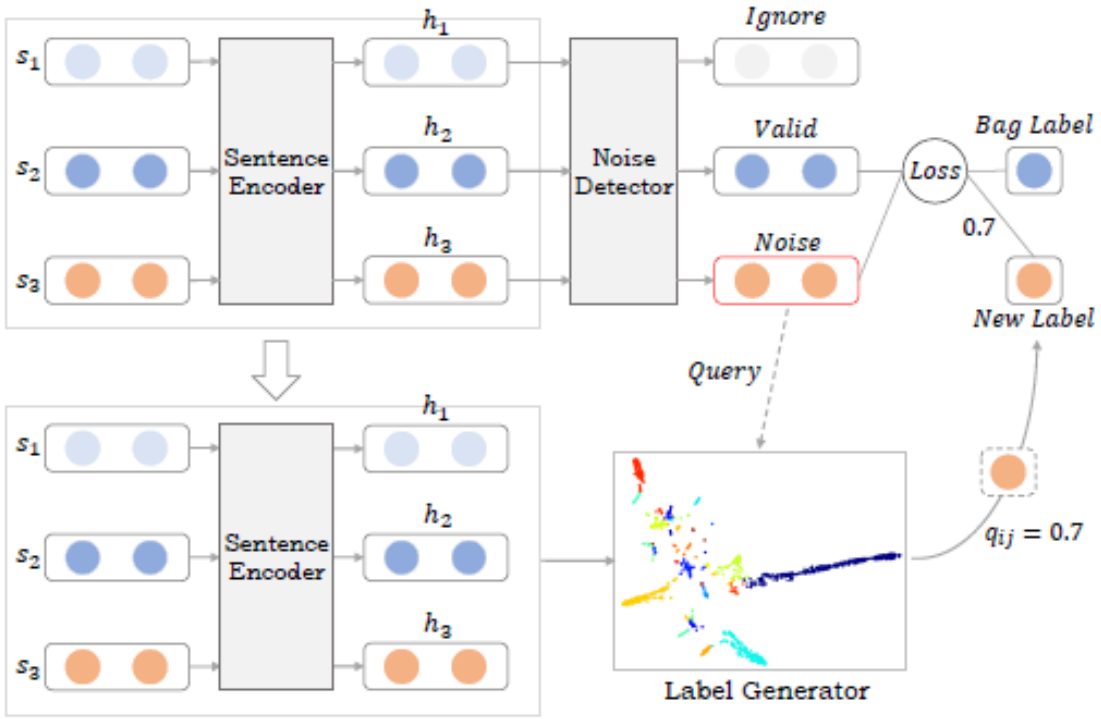


Figure 1: The architecture of DCRE, illustrating the procedure of handling one sentence-bag which contains three sentences.

$R = \{r_1, r_2, \ldots, r_k\}$

$S_b = \{s_1, s_2, \ldots, s_b\}$ (e_1, e_2)

Predict the relation r_i of sentence-bag according to the entity pair (e_1, e_2)

Model

➢Sentence Encoder: transform a sentence into low-dimension vectors

- Word Representation

$$x_i = [v_i; p_i^{e_1}; p_i^{e_2}]$$
$$X = \{x_1, x_2, \dots\dots, x_{n_l}\}$$

- PCNN
  - One dimension convolution

$$m_i = W^T x_{i-w+1:\, i}$$
$$M = \{m_1, m_2, \dots, m_{n_l}\}$$

  - One dimension max-pooling

$$M_i = \{M_{i1}, M_{i2}, M_{i3}\}$$
$$h = [p_{i1}, p_{i2}, p_{i3}], \quad p_{ik} = \max(M_{ik}), k \in \{1,2,3\}$$

Model

➢Noise Detector: select noisy sentences from sentence-bags

- Calculate the coupling coefficient
  - $H_b = \{h_1, h_2, \dots, h_b\}, H_b \epsilon R^{b \times d_s}$
  - $L = \{l_1, l_2, \dots l_k\}, L \epsilon R^{k \times d_s}$
  - $a_i = h_i l_j^T$
- Normalize the coupling coefficient
  - $a_i = \frac{\exp(a_i)}{\sum_b \exp(a_i)}$
- Set threshold ∅
- Notes: use the currently deterministic sentences

Model

➢Label Generator: provide high-confidence relation labels for noisy sentences

- Project sentence representations into relation feature space
  - $H = \{h_1, h_2, \dots, h_n\}$
  - $L = \{l_1, l_2, \dots, l_k\}$
  - $C = HL^T + b$
- Clustering
  - $\{\mu_i\}_{i=1}^{n_c}$
  - $q_{ij} = \frac{(1+\|c_i - \mu_j\|^2)^{-1}}{\sum_j (1+\|c_i - \mu_j\|^2)^{-1}}$
  - $\mathcal{L} = KL(P||Q) = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}}$
  - $p_{ij} = \frac{q_{ij}^2 / \sum_i q_{ij}}{\sum_j (q_{ij}^2 / \sum_i q_{ij})}$
- Notes: only generate new label for positive samples

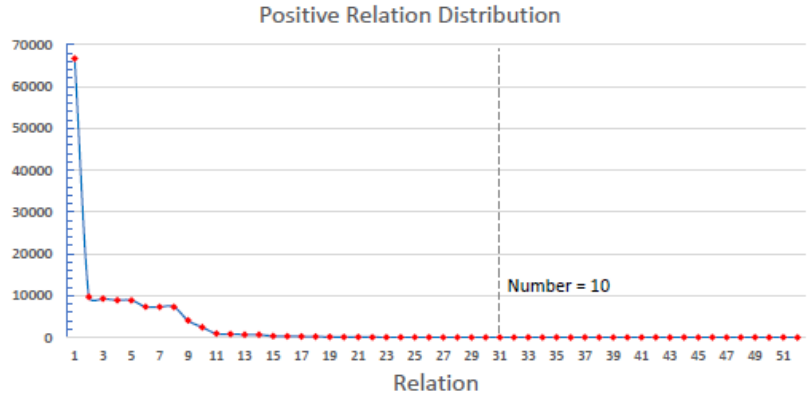


Figure 2: The distribution of 52 positive relations (exclude *NA*) in NYT-10 dataset. The horizontal axis shows different relations sorted by the number of occurrence. The vertical axis shows the number of sentences in training set. The vertical line indicates that the relation whose id is 31 appears 10 times in the training set.

Model

➢Scaled Loss Function

$$\mathcal{T}(\theta) = -\sum_{(x_i,y_i)\epsilon\mathbb{V}} \log p(y_i|x_i;\Theta) - \lambda \sum_{(x_i,y_i)\epsilon\mathbb{N}} q_{ij} \log p(y_j|x_i;\Theta)$$

Experiments

- Dataset: NYT-10
  - Training 2005-2006
  - Including 522611 sentences, 281270 entities pairs, 18252 relational facts
  - Test: 2007
  - Including 172448 sentences, 96678 entity pairs and 1950 relational facts
  - 53 unique relations including NA
- Experiment Setup

Setting	Number
Kernel size	3
Feature maps	230
Word embedding dimension	50
Position embedding dimension	5
Pre-train learning rate	0.4
Clustering learning rate	0.004
Model learning rate	0.1
Threshold ϕ	0.1
Dropout	0.5
Coefficient λ	0.6

Table 2: Parameters Setting

Experiments

➢Baselines

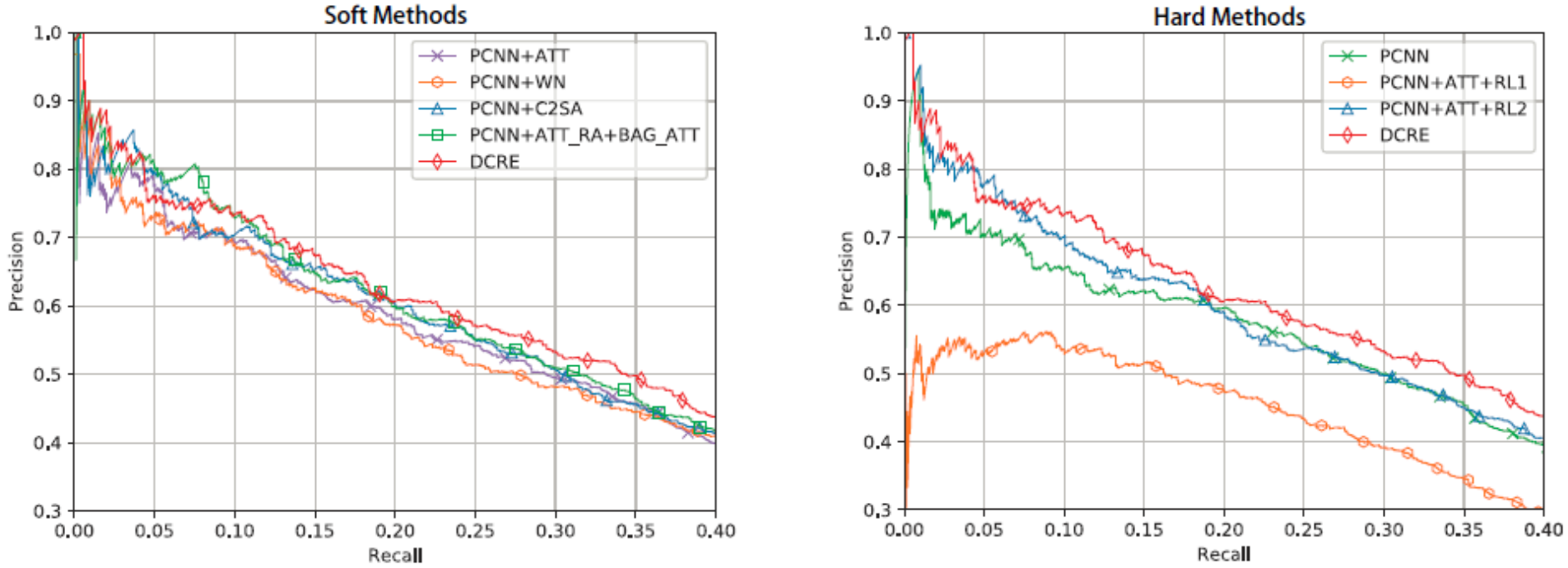


Figure 3: Comparison results with soft methods (left) and hard methods (right).

Experiments

➢Effect of the threshold

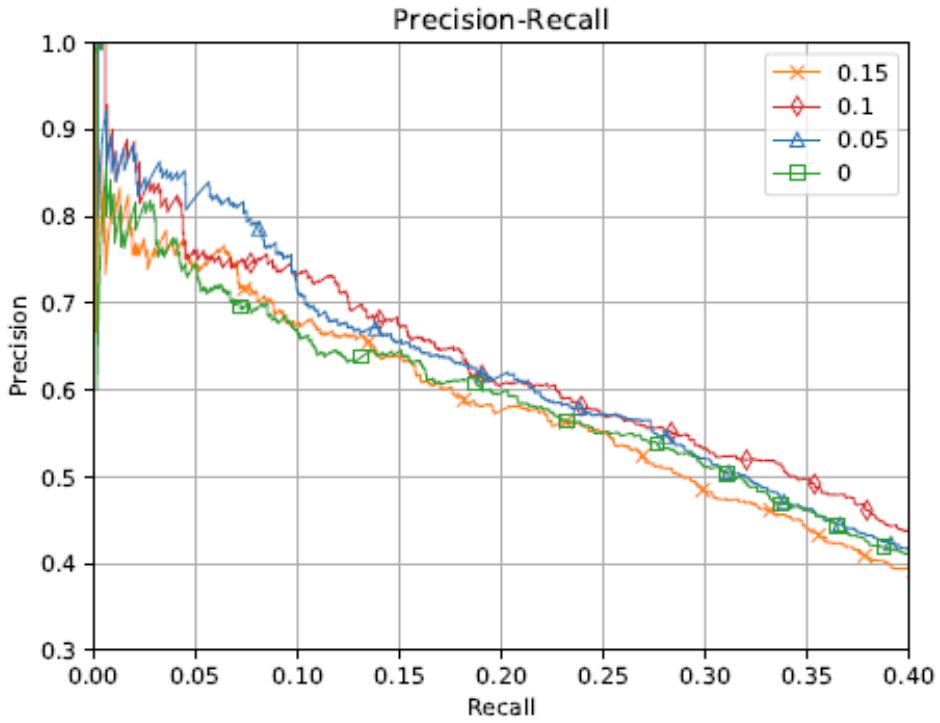
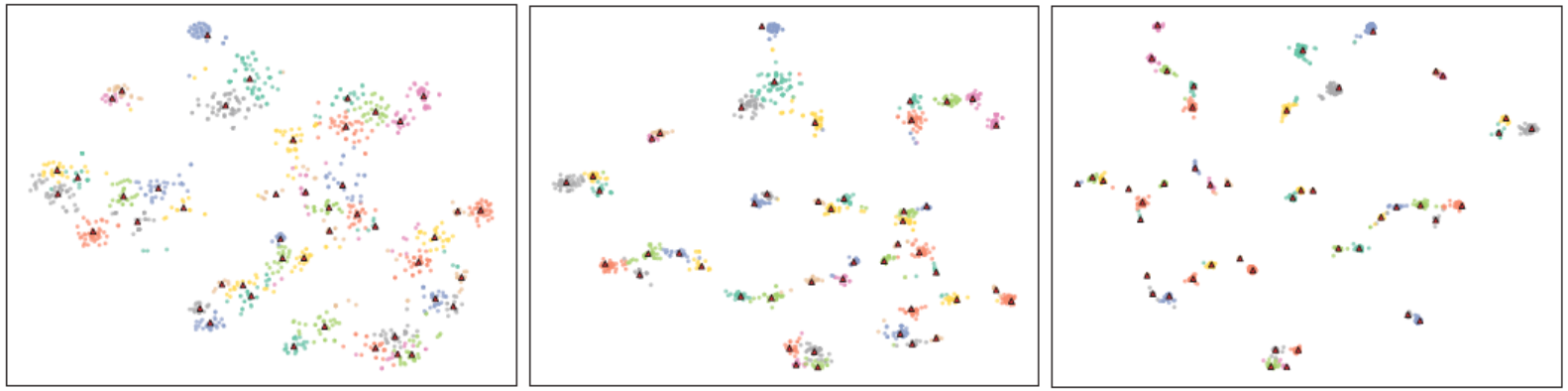


Figure 4: Effect of the threshold ϕ.

Experiments

➢Clustering Result



epoch 1 epoch 5 epoch 10

Figure 5: $t-$SNE visualization of clustering results on subset of NYT-10. The red triangles are cluster centers.

ID	Entity pair	Sentence	Original label	Generated label	Correct?
1	(China,Beijing)	**Beijing** has tried to enlist the support of Uzbekistan in fighting Islamic separatism in **China**'s western region of Xinjiang, while also lining up secure supplies of oil and gas.	/location/location/contains	/location/cn province /capital	No
2	(Italy, Rome)	Mr. Tomassetti's companies are named after L'Aquila, **Italy**, his birthplace 58 miles northeast of **Rome**.	/location/country/capital	/location/location/contains	Yes
3	(Saddam Hussein, Iraq)	As national journal reported in April, it was Senator Roberts who stated as the **Iraq** war began that the U.S. had "human intelligence that indicated the location of **Saddam Hussein**."	/people/deceased person/place of death	/people/person/place lived	Yes
4	(Edith Sitwell, England)	His first book was published privately in his own country and then by a major publisher in **England**, where he had many supporters in the literary world, most notably **Edith Sitwell** and Angus Wilson.	/people/person/nationality	/people/person/place of birth	No
5	(Louisiana, New Orleans)	The book, by a **New Orleans** resident, John M. Barry, describes the history and politics behind a flood that killed 1,000 people and displaced 900,000 from **Louisiana** to Illinois.	/location/location/contains	NA	Yes

Table 3: Five sentences randomly selected in NYT-10 dataset. The text in **bold** represents the entity, the text underline represents the relation label is correct.

Conclusion and Future Work

- **Conclusion**
  - An unsupervised deep clustering based distant supervised relation extraction model
  - Convert noisy sentences into useful training data
  - Outperform than comparable approaches
- **Future Work**
  - Multi-class clustering
  - End-to-end method

谢谢聆听

Thank You