H-FND: Hierarchical False-Negative Denoising for Distantly Supervision Relation Extraction

Jhin-Wei Chen, Tsu-Jui Fu, Chen-Kang Lee and Wei-Yun Ma

Findings ACL-IJCNLP 2021

Institute of Information Science,
Academia Sinica

Challenges in Information Extraction

- The lack of labeled data
- Labeling data is very expensive and time consuming

Distant supervision was proposed to generate training data by aligning triples in knowledge bases to unannotated sentences.

Noise from Distantly Supervised Relation Extraction Datasets

Knowledge base	Relation	
Steve Jobs, San Francisco	PoB	
Corpus	**Relation**	**Type**
Jobs was born in **San Francisco**	PoB (✓)	TP
Jobs moved back to **San Francisco**	PoB (✗)	FP
Manuela was born in **New York**	NA (✗)	FN

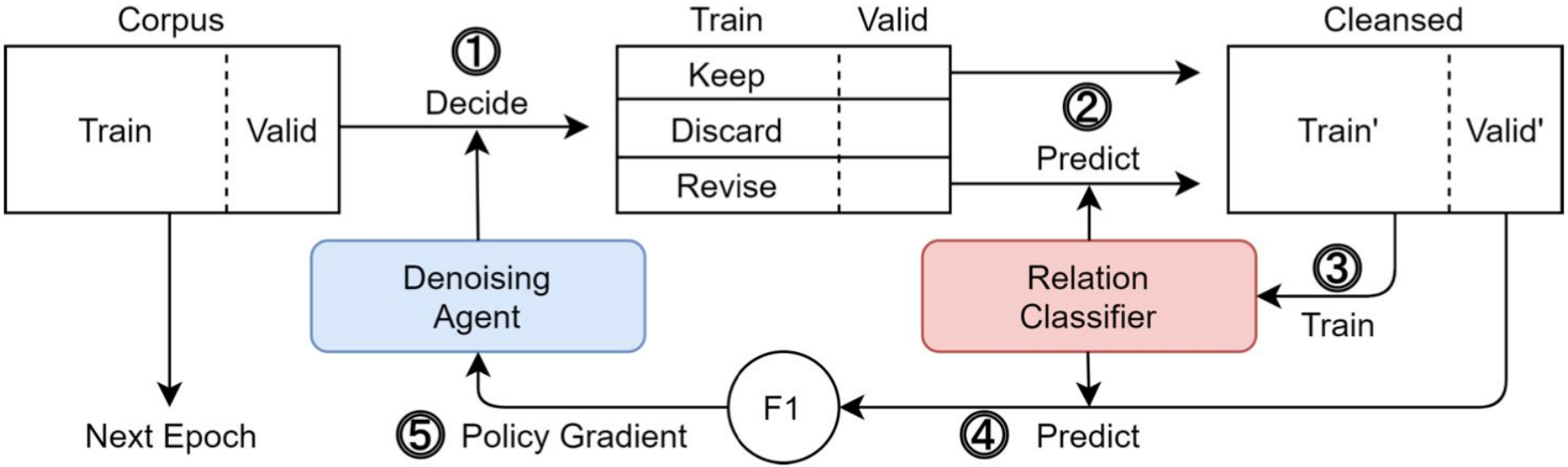
The Ternary Policy of the Agent

To retrieve positive instance from negative samples

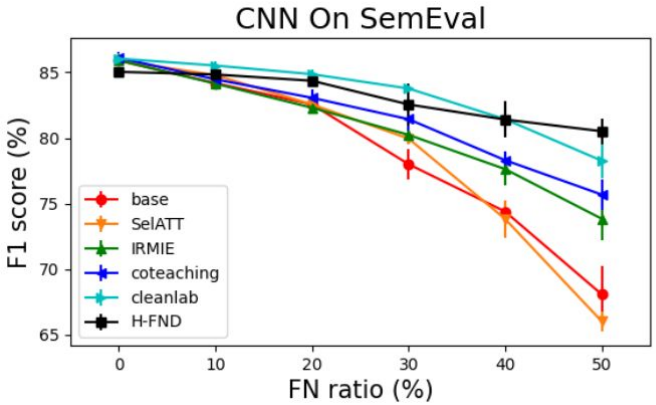
For each negative sample **s** in the dataset:

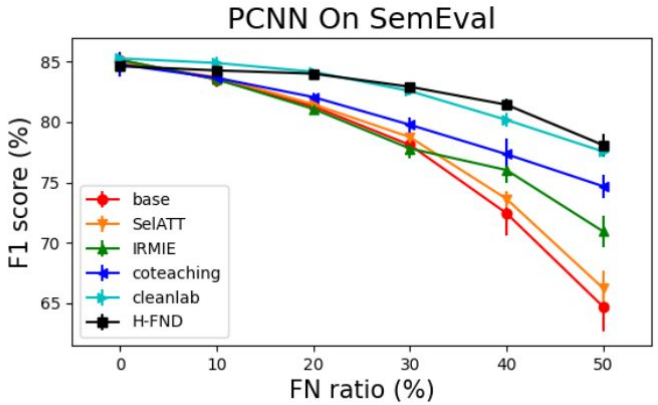
- ***Keep:*** maintain **s** as a negative instance for training/validation
- ***Discard:*** remove **s** to prevent it from misleading the model
- ***Revise:*** predict a new relation type for **s** and treat it as a positive for the following training/validation.

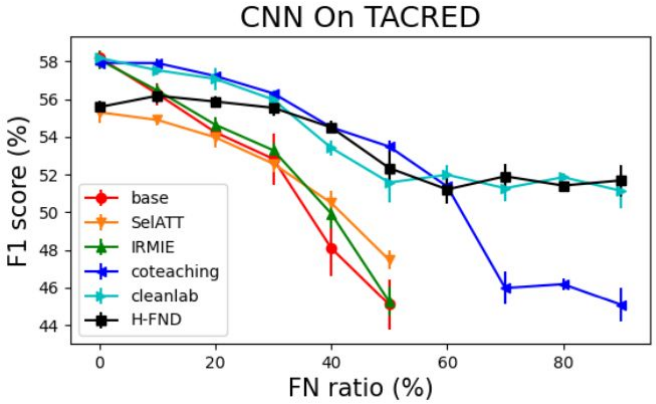
The H-FND Framework

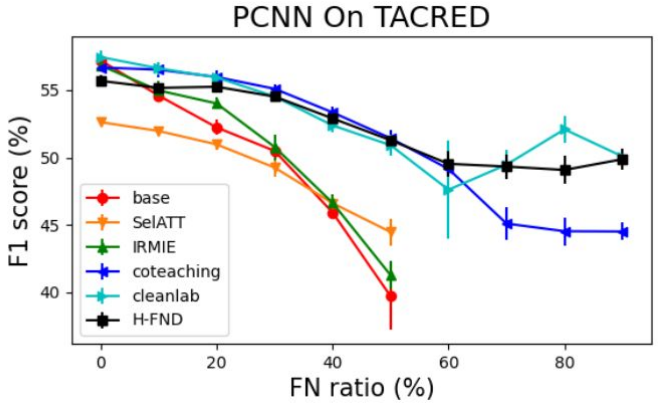


Experiments on Synthetic Noise









Experiment on Distantly Supervised Dataset