Learning with Noise: Improving Distantly-Supervised Fine-grained Entity Typing via Automatic Relabeling

Haoyu Zhang1, Dingkun Long2, Guangwei Xu2, Muhua Zhu2, Pengjun Xie2, Fei Huang2, Ji Wang1
1HPCL, College of Computer, National University of Defense Technology, China; 2Alibaba Group, China;
Correspondence to: zhanghaoyu10@nudt.edu.cn

Introduction
Fine-grained entity typing (FET) is a fundamental task for various entity-leveraging applications. Although great success has been made, existing systems still have challenges in handling noisy samples in training data introduced by distant supervision method. In this paper, we propose a probabilistic automatic relabeling method which treats all training samples uniformly. Our method aims to estimate the pseudo-truth label distribution of each sample, and the pseudo-truth distribution will be treated as part of trainable parameters which are jointly updated during the training process. The proposed approach does not rely on any prerequisite or extra supervision, making it effective on real applications. Experiments on several benchmarks show that our method outperforms previous competitive approaches and indeed alleviates the noisy labeling problem.

Motivation
To address the issue of noisy labeling, most of previous studies try to model the samples with only one label and samples with multiple labels separately, or to detect and weight out noise based on the assumption that the distantly-supervised label set must contain the correct type. The weaknesses of these studies are:
• Previous work rely on overly strong assumptions.
• Existing methods cannot deal with false-positive one type “clean” samples.

Method
A probabilistic automatic relabeling (AR) method which handles the above two limitations simultaneously. As the ground-truth label distribution is not available, our method aims at estimating the pseudo-truth label distribution during the training process. In detail, each sample is assigned a continuous label distribution \( \hat{y} \) over all candidate labels, and \( \hat{y} \) is jointly updated as trainable parameters through the back-propagation algorithm.

Ablation study of AR
• Self-learning (KL divergence) is the key component.
• Noise information (Noisy label initialization) is also important.

Auxiliary Experiments
How does the three phrase training impacts the performance?

Overall Results
• Three datasets Wiki, OntoNotes, BBN. (two shown in poster)
• Several competitive baselines, including noisy learning based FET methods (NFETC-CLSC, NDP).
• Methods with \( \text{hier} \) denotes for variants with hierarchical loss proposed by NFETC.

Conclusion
• A probabilistic automatic relabeling method, verified on fine-grained entity typing.
• Do not rely on extra prerequisite or supervision.
• Backbone and task agnostic, making it general and flexible.