Learning to Denoise Distantly-Labeled Data for Entity Typing
Yasumasa Onoe and Greg Durrett, {yasumasa, gdurrett}@cs.utexas.edu

1. Distantly-Labeled Data
Pros: Useful to scale up training data for data-hungry statistical models such as neural networks.
Cons: Typically noisy and that noise can vary with the distant labeling technique.
Examples:
(a) Wrong labels
No matter whom they buy from, users blame [Amazon].
(b) Missing labels
The Minnesota Lynx play their home games at Target Center in [Minneapolis].

How to fix these noisy labels produced by distant supervision?

2. Our Framework
Filtering Model: Take a labeled example and swap its types with another example’s types. Learn a binary classifier to identify these swapped exs.
Relabeling Model: Take a labeled example and randomly drop 70% of its types. Learn a model to recover the true types.

3. Model
The figure shows the Relabeling Model. Our final entity typing model uses the mention & context vector only, which follows Choi et al. (2018).

We encode the mention in context as well as its noisy observed types, and predict the true type set based on these signals.

4. Experiments
Dataset: Ultra-Fine Entity Typing (Choi et al. 2018)

Does denoising help?
- We compare the performance of our model without data augmentation and with/without denoising. Also, we compare with the current SOTA model (Choi et al. 2018).
- Adding the distantly-labeled data lowers the performance, but denoising makes the distantly-labeled data useful.

*BERT: BERT without data augmentation achieves 40.2 F1. BERT performs well on OntoNotes but not as well as our model with augmented training data. Using the distant data in BERT was challenging due to instability in training.

Comparing Denoising Models
- We compare with simple denoising heuristics: Pair: look at type pair cooccurrences and use a heuristic to add types. Overlap: use a model trained on the manually-labeled data to predict types, treat those types as gold when they overlap with the noisy types.
- Our learned denoising techniques outperform heuristic baselines.

See our paper for results on the OntoNotes dataset.