## Avoiding Your Teacher's Mistakes: Training Neural Networks with Controlled Weak Supervision

Making use of weak or noisy signals, like the output of heuristic methods or user click through data for training deep neural networks is increasing, in particular for the tasks where an adequate amount of data with true labels is not available. In a semi-supervised setting, we can use a large set of data with weak labels to pretrain a neural network and fine tune the parameters with a small amount of data with true labels. However, these two independent stages do not leverage the full capacity of clean information from true labels during pretraining. In this paper, we propose a semi-supervised learning method where we train two neural networks in a multitask fashion: a target network and a confidence network. The target network is optimized to perform a given task and is trained using a large set of unlabeled data that are weakly annotated. We propose to weight the gradient updates to the target network using the scores provided by the second confidence network, which is trained on a small amount of supervised data. Thus we avoid that the weight updates computed from noisy labels harm the quality of the target network model. We evaluate our learning strategy on two different tasks: document ranking and sentiment classification. The results demonstrate that our approach not only enhances the performance compared to the baselines but also speeds up the learning process from weak labels.

Author: Mr ALIAKSEI, Severyn (Google Research) Presenter: Mr ALIAKSEI, Severyn (Google Research)