Supplementary Materials for Data Poisoning Attacks on Crowdsourcing Learning

Anonymous Authors[1]

In this supplementary document, we provide the following details to support the main text:

Section A: we provide the procedure of weighted majority voting.

Section B: we provide the details of aggregating the crowd labels using weighted majority voting on the motivating example.

A. The Procedure of Weight Majority Voting.

Algorithm 1 Weight Majority Voting.

Input: Noisy Labels $\mathbf{Y} = (y_{ij})_{N\times M}$
Initialize the ability of worker $\gamma_j = 1.0$
repeat
$\hat{y}_i \leftarrow \underset{k}{\operatorname{argmax}} \sum_{j=1}^{M} \tau_i \mathrm{I}\left(y_{ij} = k\right), \quad \forall i \in [N]$
$\hat{\gamma}_j \leftarrow \frac{\sum_{i=1}^{N} \mathrm{I}(y_{ij}=\hat{y}_i)}{\sum_{i=1}^{N} T_{ij}}, \quad \forall j \in [M]$
$\tau_j \leftarrow L\hat{\gamma}_j - 1, \quad \forall j \in [N]$
until converges or reaches S iterations.
Output: Aggregated label $\hat{y}_i$

This strategy can potentially improve the performance of majority voting and result in a better estimate for the ability of workers, which further improves the quality of the weights and iterate. In weight majority voting, $\mathbf{Y} = (y_{ij})_{N\times M}$ denotes the label matrix from workers $\{u_j\}_{j=1}^{M}$ to instances $\{\mathbf{x}_i\}_{i=1}^{N}$, $\gamma_j = p(y_{ij} = z_j)$ denotes the ability u_i exhibits, namely the probability of the answers provided by u_i being correct and τ_i denotes the weight associated with alternative label k when worker u_j provides label y_{ij} to instance $\mathbf{x}_i$. L denotes the number of alternative labels of instances. $T = \{T_{ij}\}_{N\times M}$ denotes the indicator matrix where $T_{ij} = 1$ indicates that worker u_j has provided a label to instance $\mathbf{x}_i$, or otherwise. Weight majority voting is a two-step iteration algorithm, which mainly contains two steps as follows.

- In step 1, the weight of each label is computed as $L\hat{\gamma}_j - 1$, and the candidate label which receives the highest weights is computed as the aggregated label of the instance.
- In step 2, the ability of a worker is estimated based on the aggregated label $\hat{y}_i$ of each instance.

B. Label Aggregation Process of Motivating Example

In the *motivating example*, weighted majority voting is applied to aggregate the noisy crowd labels from workers $\{u_1, u_2, u_3, u_4, u_5\}$ to be accurate ones for instances used to train a classifier. In each iteration, the aggregated label of each instance and the estimated ability of each worker is list in Table 1 and Table 2, respectively.

Table 1. Aggregated labels $\hat{z}_i$ of instances computed by weighted majority voting on the motivating example.

ROUND	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$
ROUND 1	B	B	B	B	B	B	B
ROUND 2	B	B	A	B	B	B	B
ROUND 3	B	B	A	B	B	B	B
...	...	...	...	...	...	...	...
CONVERGENCE	B	B	**A**	B	B	B	B

Table 2. Estimated ability of workers γ_j computed by weighted majority voting on the motivating example.

ROUND	u_1	u_2	u_3	u_4	u_5
ROUND 1	0.571	0.571	1.000	**0.857**	**0.857**
ROUND 2	0.429	0.429	0.857	**1.000**	**1.000**
ROUND 3	0.429	0.429	0.857	**1.000**	**1.000**
...	...	...	...	...	...
HLINE CONVERGENCE	0.429	0.429	0.857	**1.000**	**1.000**