FedSIR: Spectral Client Identification and Relabeling for Federated Learning with Noisy Labels

Supplementary Material

7. Ablation Study

To better understand the contribution of each component of FedSIR, we perform an ablation study under symmetric label noise with Dirichlet heterogeneity parameter $\alpha = 1$. In this setting, three clients are identified as clean and used to construct the spectral reference model. We evaluate several variants of the proposed framework by removing one component at a time: the spectral relabeling mechanism, LA, KD, and DaAgg. The results are summarized in Table 2.

The full FedSIR model consistently achieves the best performance across noise levels exceeding 50%. Removing the spectral relabeling component leads to the most degradation, particularly under high noise rates. For example, at 90% noise, the accuracy drops from 84.51% to 80.00%, indicating that the relabeling mechanism plays a critical role in correcting corrupted supervision on noisy clients.

LA mainly compensates for class imbalance across clients. Removing this component results in a moderate but consistent decrease in accuracy, confirming its significance for stabilizing local optimization under heterogeneous label distributions.

KD also contributes to performance stability by providing soft guidance from the global model. When KD is removed, performance decreases slightly across most noise levels, suggesting that KD helps regularize the training of noisy clients and prevents overfitting to unreliable labels.

Finally, removing the DaAgg step slightly improves performance at lower noise rates but degrades robustness at higher noise levels. This behavior suggests that DaAgg primarily acts as a safeguard against highly corrupted client updates rather than improving average-case performance when noise is moderate.

Overall, the ablation results demonstrate that each component contributes to the robustness of the proposed framework, with spectral relabeling playing the critical role in handling severe label corruption.

8. Results on CIFAR-100

We further evaluate our method on CIFAR-100 under symmetric label noise in a federated setting with 10 clients. Compared with CIFAR-10, CIFAR-100 contains a significantly larger number of classes, which makes the learning problem more challenging under both label noise and non-IID data. In particular, under strong non-IID settings (small α), each client tends to contain few classes, requiring more clean clients to ensure sufficient class coverage for reliable spectral reference construction. To simulate non-IID data, this dataset is partitioned using a Dirichlet distribution with $\alpha \in \{0.3, 0.5, 2\}$.

Table 3 reports the classification accuracy under symmetric noise rates ranging from 30% to 90%. Despite the increased difficulty of CIFAR-100, our method consistently achieves the best performance across all noise levels and heterogeneity settings. In particular, our approach maintains stable performance even under extremely high noise levels (e.g., 80%–90%).

These results further demonstrate that the proposed spectral client identification and relabeling strategy scales well to more challenging multi-class settings and remains robust to both severe label noise and client heterogeneity.

9. Relabeling Strategy

To analyze the role of the proposed relabeling rule, we compare three variants of the spectral correction mechanism used in Stage II:

- $S^{(r)}$: labels are reassigned according to the dominant-direction alignment score:

$$\hat{y}_i^{(r)} = \arg\max_c S^{(r)}(i, c).$$

- $S^{(n)}$: labels are determined using the residual-subspace projection score:

$$\hat{y}_i^{(n)} = \arg\min_c S^{(n)}(i, c).$$

- Agreement rule: a relabel is accepted only when the scores from $S^{(r)}$ and $S^{(n)}$ coincide: $\hat{y}_i^{(r)} = \hat{y}_i^{(n)}$

Figure 3 reports the average reduction in label noise achieved by these strategies.

To evaluate the impact of these strategies on the final model performance, Table 4 reports the test accuracy obtained using each relabeling variant.

Table 2. Ablation study of FedSIR on CIFAR-10 with symmetric label noise and Dirichlet parameter $\alpha = 1$. Removing each component shows its contribution to the overall framework. **Bold** indicates the best result and underline indicates the second-best result.

α	# Clean Clients	Method	Symmetric Noise 30%	40%	50%	60%	70%	80%	90%
		w/o relabeling	85.13±0.73	84.81±0.73	84.29±0.70	83.90±0.76	82.85±0.76	82.06±0.75	80.00±0.77
		w/o LA	84.93±0.69	84.60±0.76	84.63±0.70	84.38±0.75	84.15±0.75	84.04±0.71	84.14±0.72
1	3	w/o KD	85.49±0.69	85.16±0.68	85.03±0.70	84.63±0.66	84.43±0.69	84.16±0.73	84.37±0.68
		w/o DaAgg	**85.65±0.70**	**85.33±0.73**	**85.25±0.68**	84.41±0.70	83.90±0.71	83.95±0.73	83.73±0.72
		Ours	85.21±0.70	85.28±0.68	84.86±0.70	**84.68±0.71**	**84.54±0.73**	**84.35±0.74**	**84.51±0.72**
	3	Pruning				81.78±0.80			
	10	FedAvg				87.23±0.66			

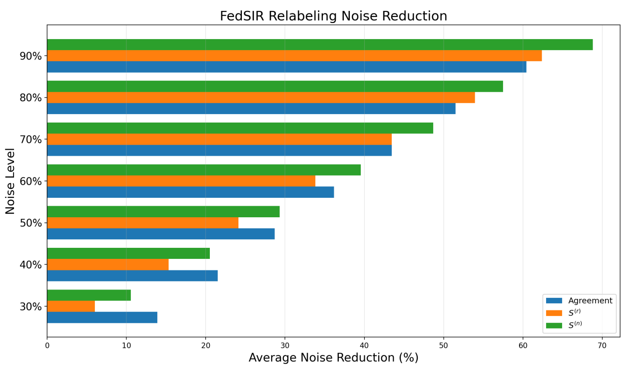

Figure 3. Average relabeling noise reduction under different symmetric noise levels for three relabeling variants: using only the dominant-direction score $S^{(r)}$, using only the residual-subspace score $S^{(n)}$, and the proposed agreement-based strategy.

Table 3. Results on CIFAR-100 under symmetric label noise. **Bold** indicates the best result and underline indicates the second-best result.

α	# Clean Clients	Method	Symmetric Noise						
			30%	40%	50%	60%	70%	80%	90%
0.3	7	FedAvg [17]	57.36±0.96	56.92±0.92	55.93±0.94	55.60±0.95	55.02±0.99	54.07±0.97	53.10±1.00
		FedProx [12]	57.55±0.98	57.05±1.02	56.89±1.01	55.80±0.95	55.66±0.97	54.88±0.95	54.27±1.00
		RoFL [22]	46.77±0.99	45.75±0.96	46.03±0.98	42.82±0.97	41.46±0.95	42.22±0.98	40.72±0.94
		RHFL [4]	18.64±2.01	18.53±2.05	17.76±2.67	17.25±2.50	16.86±3.55	15.51±4.90	13.37±4.94
		FedLSR [8]	47.55±1.06	46.99±1.01	46.60±0.99	46.14±0.99	44.73±1.00	44.74±1.03	43.53±0.98
		FedCorr [21]	50.72±1.04	50.56±1.02	50.23±1.00	48.71±0.94	47.77±0.95	47.11±0.96	47.24±0.99
		FedNed [15]	56.18±0.92	56.46±0.95	56.56±1.03	56.23±1.01	56.58±0.96	56.21±0.98	56.36±0.97
		FedELC[9]	58.65±0.93	58.45±0.95	58.56±0.98	58.14±0.95	58.38±0.97	58.17±0.94	57.99±0.93
		FedNoRo[20]	56.90±0.97	56.90±0.94	56.74±0.98	56.46±0.99	56.41±0.97	56.51±0.97	55.92±0.98
		FedSIR (Ours)	**59.43±0.95**	**59.53±0.96**	**58.94±0.92**	**59.18±0.99**	**58.81±0.95**	**58.78±0.94**	**58.43±0.93**
	7	Pruning				56.74±0.95			
	10	FedAvg				60.67±0.93			
0.5	6	FedAvg [17]	57.38±1.01	57.15±0.97	56.05±0.99	54.95±0.96	54.65±0.94	52.62±0.93	51.25±1.04
		FedProx [12]	57.72±1.00	57.17±0.96	55.95±0.92	55.33±1.00	55.34±0.97	53.58±0.98	52.14±0.95
		RoFL [22]	51.20±0.99	48.76±0.97	48.44±0.96	46.12±0.96	47.02±1.00	44.92±1.01	45.24±0.96
		RHFL [4]	21.47±2.08	21.33±2.00	21.08±2.66	19.23±3.27	17.85±4.20	16.32±5.43	15.39±6.14
		FedLSR [8]	48.50±0.95	47.82±0.98	46.02±1.00	45.53±1.02	44.13±0.99	43.48±1.01	42.38±1.01
		FedCorr [21]	50.85±0.96	50.80±0.97	50.37±1.01	50.34±0.94	50.03±0.97	49.65±1.04	49.52±0.95
		FedNed [15]	56.47±0.97	55.97±0.91	56.04±1.00	55.55±0.98	56.05±0.94	55.70±0.94	55.27±0.94
		FedELC[9]	57.91±0.97	57.85±0.92	57.86±0.97	57.91±0.99	57.64±0.97	57.30±0.96	57.85±0.95
		FedNoRo[20]	58.59±0.98	58.22±0.95	57.74±1.00	57.95±0.97	57.70±0.97	57.10±0.97	57.07±0.98
		FedSIR (Ours)	**59.74±0.94**	**59.73±0.91**	**59.60±0.98**	**59.30±0.94**	**59.08±0.96**	**58.88±0.95**	**58.56±0.94**
	6	Pruning				56.57±0.93			
	10	FedAvg				61.91±0.93			
2	5	FedAvg [17]	58.81±0.96	57.88±0.96	56.77±1.02	55.90±0.97	54.48±1.00	53.49±0.99	51.20±0.97
		FedProx [12]	58.41±0.94	57.15±0.98	56.51±0.96	55.61±0.98	54.51±1.00	53.29±0.97	51.78±1.00
		RoFL [22]	54.62±1.01	53.99±0.98	53.60±0.96	52.65±0.96	50.80±0.99	50.61±0.96	46.69±0.98
		RHFL [4]	29.16±1.69	27.90±2.24	26.05±3.11	23.74±4.22	22.00±5.18	19.18±7.05	16.65±7.85
		FedLSR [8]	49.61±1.03	47.98±1.00	47.52±1.01	46.68±0.96	45.07±1.01	43.73±1.03	41.63±0.99
		FedCorr [21]	53.48±0.94	52.68±0.99	52.30±1.04	51.56±0.95	51.59±0.98	51.22±0.97	50.56±1.04
		FedNed [15]	57.53±0.95	56.68±1.00	57.30±1.00	57.03±0.96	56.50±0.97	56.43±0.98	56.26±0.99
		FedELC[9]	58.02±1.01	58.04±1.00	57.88±0.98	57.49±0.97	57.57±1.00	57.21±0.97	57.05±0.97
		FedNoRo[20]	59.76±1.04	59.61±1.01	58.80±1.00	58.72±0.97	57.88±0.93	57.58±0.99	57.19±0.98
		FedSIR (Ours)	**60.42±0.98**	**59.81±0.95**	**59.44±0.94**	**58.81±0.97**	**58.91±0.94**	**58.27±1.00**	**58.11±1.01**
	5	Pruning				56.87±1.02			
	10	FedAvg				63.05±0.92			

Table 4. Effect of different relabeling strategies on CIFAR-10 under symmetric label noise. **Bold** indicates the best result and underline indicates the second-best result.

α	# Clean Clients	Method	Symmetric Noise						
			30%	40%	50%	60%	70%	80%	90%
1	3	$S^{(r)}$	84.73±0.70	84.64±0.72	84.45±0.71	84.07±0.73	84.31±0.68	83.95±0.70	84.06±0.70
		$S^{(n)}$	84.43±0.70	84.21±0.70	84.15±0.71	84.34±0.72	84.18±0.72	84.25±0.71	83.89±0.72
		Agreement	**85.21±0.70**	**85.28±0.68**	**84.86±0.70**	**84.68±0.71**	**84.54±0.73**	**84.35±0.74**	**84.51±0.72**