Generative Calibration of Inconsistent Annotation for Label Distribution Learning
supplementary material

Overview

To demonstrate the adequacy of our work, we need to supplement the experimental results. In this supplementary material, we present the detailed experimental results of the experiments, including recovery experiment, prediction experiment, and ablation experiment.

Experimental Results

Recovery Experiment

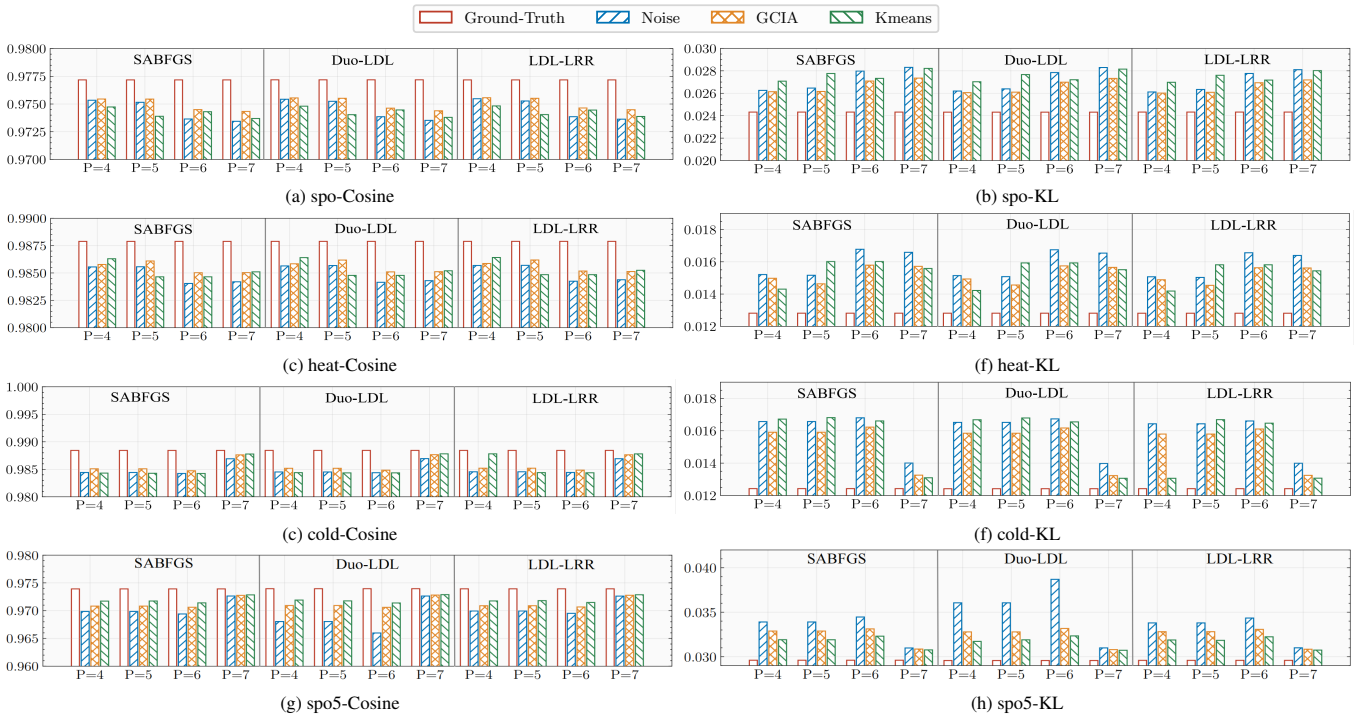
	P=4	GCIA	Kmeans	P=5	GCIA	Kmeans	P=6	GCIA	Kmeans	P=7	GCIA	Kmeans
						ns						
Cheb (↓)	0.1102	0.2525	0.1299	0.1411	0.1681	0.1891	0.1778	0.1852	0.1808	0.1962	0.2003	**0.1945** •
Clark (↓)	0.6897	2.4071	2.2345	0.8367	2.3613	2.3704	0.9834	2.2590	2.2360	1.0860	2.3200	2.0508
Canber (↓)	1.6856	6.5293	5.8222	2.0587	6.2172	6.2971	2.4418	5.9687	5.9669	2.7029	6.2083	5.4264
KL (↓)	0.1370	0.5738	0.1478	0.3224	**0.2853** •	0.3238	0.5617	**0.3636**	**0.3237** •	0.7258	**0.4235** •	**0.4245**
Cosine (↑)	0.9467	0.8302	**0.9491** •	0.9194	0.9243	**0.9016** •	0.8879	**0.8940**	**0.8956** •	0.8703	**0.8775** •	**0.8730**
Intersec (↑)	0.8748	0.5950	0.8406	0.8388	**0.7800** •	**0.7641**	0.7991	0.7781	0.7861	0.7753	0.7567	0.7713
Rho (↑)	0.7389	0.5520	**0.7474** •	0.7002	0.7142	0.7076	0.6535	**0.6892**	**0.7109** •	0.6166	**0.6643**	**0.6911** •
						spo5						
Cheb (↓)	0.3460	**0.2460** •	**0.2475**	0.3654	**0.2590** •	**0.2610**	0.3880	**0.2758** •	**0.2767**	0.2233	**0.2016**	**0.2014** •
Clark (↓)	0.8587	**0.5089** •	**0.5166**	0.9124	**0.5334** •	**0.5427**	0.9620	**0.5622** •	**0.5695**	0.5398	**0.4594**	**0.4589** •
Canber (↓)	1.2803	**0.7920** •	**0.8007**	1.3694	**0.8355** •	**0.8468**	1.4533	**0.8860** •	**0.8932**	0.7896	**0.6840**	**0.6832** •
KL (↓)	1.6647	**0.2163** •	**0.2257**	1.9085	**0.2369** •	**0.2480**	2.1544	**0.2639** •	**0.2741**	0.5860	**0.2093**	**0.2069** •
Cosine (↑)	0.7929	**0.8661** •	**0.8647**	0.7790	**0.8558** •	**0.8542**	0.7590	**0.8406** •	**0.8399**	0.8865	**0.9017**	**0.9018** •
Intersec (↑)	0.6540	**0.7540** •	**0.7525**	0.6346	**0.7410** •	**0.7390**	0.6120	**0.7242** •	**0.7233**	0.7767	**0.7984**	**0.7986** •
Rho (↑)	0.1257	**0.1545** •	**0.1349**	0.1464	**0.1516** •	**0.1512**	0.0909	**0.1070** •	**0.0957**	0.2457	**0.2671**	**0.2717** •

Table 1: Recovery performance on 2 datasets. Every three columns from left to right represent one group of experiments, with the same P-value used to generate noise. 'P=4/5/6/7' represents the difference between the unprocessed noise dataset and the ground-truth. 'GCIA/KMeans' represents the difference between the dataset calibrated by the calibrationn method and the ground-truth. Bold indicates that the calibrated dataset is superior to the noise dataset. A dot following the data indicates that the calibrationn method is better.

The experimental results of the recovery experiment were supplemented here. In addition to the 6 datasets mentioned in the main body, the experimental results of 2 additional datasets (No.7, and No.8) mentioned in the article are presented here. The experimental results met our expectations.

Prediction Experiment

In the main body, we presented the results of the 'sj' and 'mov' datasets. Here, we provide additional supplementary experimental results for other datasets.



(a) spo-Cosine (b) spo-KL

(c) heat-Cosine (f) heat-KL

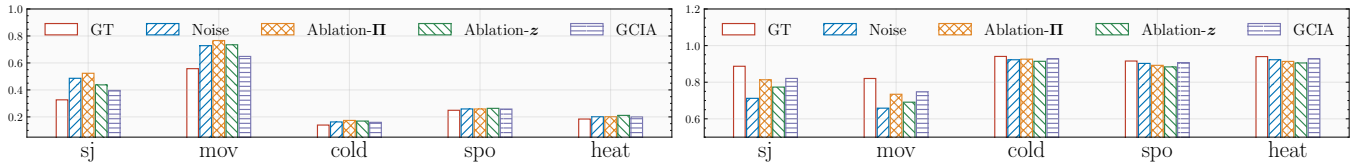
(c) cold-Cosine (f) cold-KL

(g) spo5-Cosine (h) spo5-KL

Figure 1: Prediction performance. "Noise" refers to directly using a dataset that contains noise for training an LDL model. "GCIA/KMeans" represents using GCIA/KMeans as a preprocessing method for the dataset before utilizing it for model training. The processed dataset is then employed in the training process. "P=4/5/6/7" indicates different values of professionalism P used to generate the noisy dataset.

Ablation Experiment

In the main body, we presented experimental results for two metrics. Here, we provide additional supplementary results for the other two metrics in the ablation experiments.



(a) Clark (↓) (b) Intersec (↑)

Figure 2: Ablation study. By setting K=1, we obtained the prediction experiment results without the categorical variable z (using SABFGS as an example). The figures depict the experimental results at $P = 4$. 'Ablation-' indicates the removal of this component. 'GCIA' represents a complete model.