Label Distribution Propagation-based Label Completion for Crowdsourcing

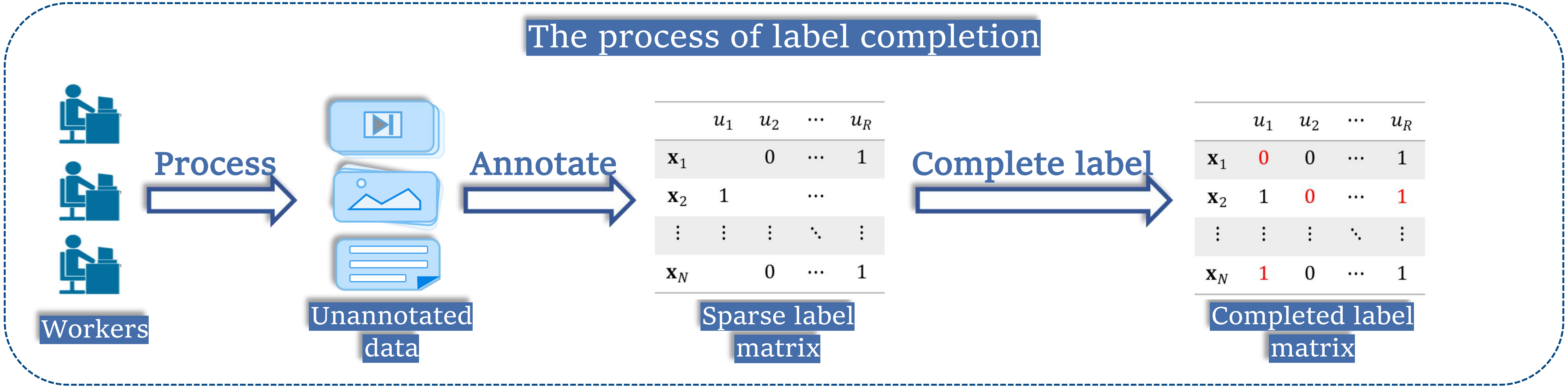
2025 International Conference on Machine Learning

Tong Wu; Liangxiao Jiang; Wenjun Zhang; Chaoqun Li.

Background

- In real-world crowdsourcing scenarios, crowd workers usually annotate a small number of instances only, which results in a very sparse crowdsourcing label matrix, and thus harms the performance of label integration algorithms.

- Thus it is necessary to complete labels before integration:



Sparse label matrix:

	u_1	u_2	$\cdots$	u_R
$\mathbf{x}_1$		0	$\cdots$	1
$\mathbf{x}_2$	1		$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{x}_N$		0	$\cdots$	1

Completed label matrix:

	u_1	u_2	$\cdots$	u_R
$\mathbf{x}_1$	0	0	$\cdots$	1
$\mathbf{x}_2$	1	0	$\cdots$	1
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{x}_N$	1	0	$\cdots$	1

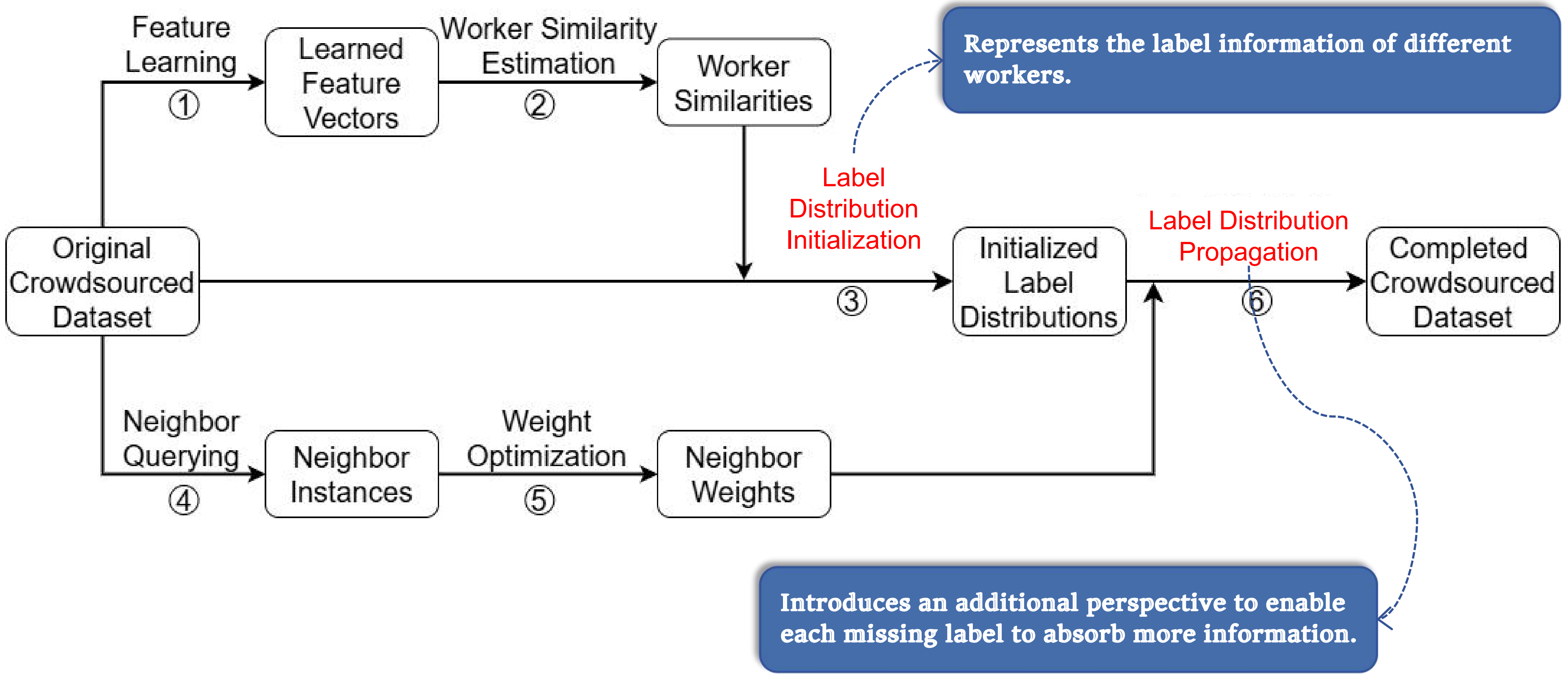
Motivations

CUG-Miner

- Recent proposed label completion algorithm *worker similarity-based label completion* (WSLC) offers a useful way to complete missing labels.
- However, it considers solely the correlation of the labels annotated by different workers on the same instance while totally ignoring the correlation of the labels annotated by different workers among similar instances.
- To overcome this limitation, we propose a *label distribution propagation-based label completion* (LDPLC) algorithm.

Method

Overall framework of LDPLC



Method

CUG-Miner

Label Distribution Initilization

- We first use Pearson correlation to learn a feature vector for each worker and then use cosine similarity to estimate worker similarity for each pair of workers. Finally, we initialize a label distribution for each missing label based on the original crowdsourced dataset and the estimated worker similarity.

Feature Learning

$$v_{rm} = \begin{cases} \sum_{l_r} p(\boldsymbol{C}_r = l_r)\, cor(\boldsymbol{C}_{br}, \boldsymbol{A}_{rm}), \\ \qquad\qquad if\ \boldsymbol{A}_{rm}\ is\ a\ continuous\ variable \\ \sum_{l_r} \sum_{a_{rm}} p(\boldsymbol{C}_r = l_r, \boldsymbol{A}_{rm} = a_{rm})\, cor(\boldsymbol{C}_{br}, \boldsymbol{A}_{brm}), \\ \qquad\qquad otherwise \end{cases}$$

Worker Similarity Estimation

$$cos(\boldsymbol{V}_r, \boldsymbol{V}_{r'}) = \frac{\boldsymbol{V}_r \cdot \boldsymbol{V}_{r'}}{|\boldsymbol{V}_r||\boldsymbol{V}_{r'}|} = \frac{\sum_{m=1}^{M} v_{rm} v_{r'm}}{\sqrt{\sum_{m=1}^{M} v_{rm}^2}\sqrt{\sum_{m=1}^{M} v_{r'm}^2}}$$

$$s(u_r, u_{r'}) = \frac{cos(\boldsymbol{V}_r, \boldsymbol{V}_{r'}) - (-1)}{1 - (-1)}$$

Label Distribution Initialization

$$p_{irq} = \frac{\sum_{r'=1}^{R} \delta(l_{ir'}, c_q) s(u_r, u_{r'})}{\sum_{q=1}^{Q} \sum_{r'=1}^{R} \delta(l_{ir'}, c_q) s(u_r, u_{r'})}$$

Method

CUG-Miner

Label Distribution Propagation

- We first query neighbors for each instance and then use local linear embedding to optimize the neighbors' weights. Next, we propagate the initialized label distribution from weighted neighbors to each missing label of each instance and finally complete each missing label based on its converged label distribution.

Weight Optimization

$$L(\boldsymbol{w}_i) = \sum_{k_1,k_2:\boldsymbol{x}_{k_1},\boldsymbol{x}_{k_2}\in\mathcal{N}_i} w_{ik_1}(\boldsymbol{x}_i-\boldsymbol{x}_{k_1})^T(\boldsymbol{x}_i-\boldsymbol{x}_{k_2})w_{ik_2}$$

$$\min_{\boldsymbol{w}_i} L(\boldsymbol{w}_i)$$

$$s.t. \begin{cases} \sum_{k=1}^{K} w_{ik} = 1 \\ \forall w_{ik} \in \boldsymbol{w}_i, w_{ik} \geq 0 \end{cases}$$

Label Distribution Propogatation

$$\boldsymbol{P}_{ir}^{t+1} = \frac{\sum_{k:x_{ik}\in\mathcal{N}_i} w_{ik}\boldsymbol{P}_{irk}^{t} + \boldsymbol{P}_{ir}}{2}$$

Label Completion

$$l_{ir} = \underset{c_q\in\{c_1,c_2,\ldots,c_Q\}}{arg\,max} \boldsymbol{P}_{ir}^{*}$$

Method

Convergence Analysis

- At the end of label distribution propagation, the distribution of each worker across the whole dataset will converge to a fixed matrix.

Proof of Convergence

$$\mathcal{P}_r^{t+1} = \frac{\mathcal{W}}{2}\mathcal{P}_r^t + \frac{\mathcal{P}_r}{2} \Longrightarrow \mathcal{P}_r^t = (\frac{\mathcal{W}}{2})^t\mathcal{P}_r + \sum_{i=0}^{t-1}(\frac{\mathcal{W}}{2})^i\frac{\mathcal{P}_r}{2} \Longrightarrow \lim_{t\to\infty}(\frac{\mathcal{W}}{2})^t = \mathbf{0}$$

$$\Longrightarrow \lim_{t\to\infty}\sum_{i=0}^{t-1}(\frac{\mathcal{W}}{2})^i = \lim_{t\to\infty}\frac{(\frac{\mathcal{W}}{2})^0 - (\frac{\mathcal{W}}{2})^t}{1-\frac{\mathcal{W}}{2}} = (1-\frac{\mathcal{W}}{2})^{-1} \Longrightarrow \lim_{t\to\infty}\mathcal{P}_r^t = (1-\frac{\mathcal{W}}{2})^{-1}\frac{\mathcal{P}_r}{2}$$

Experiments and Results

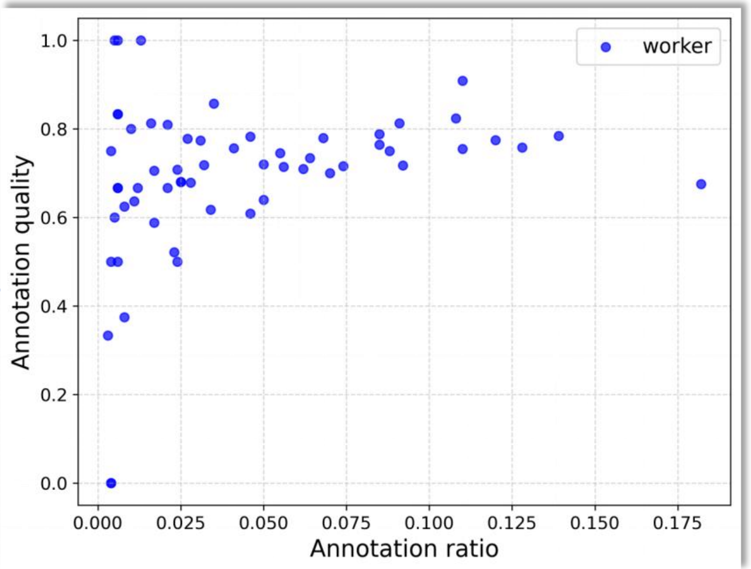
CUG-Miner

◆ Experiments on a real-world dataset:

Dataset	#Instances	#Classes	#Workers	#Labels	Annotation quality	Annotation ratio
LabelMe	1000	8	59	2547	0.75	0.04

◆ Worker distribution of the real-world dataset “LabelMe”:

Most workers’ annotation qualities are in the interval [0.6, 0.9].

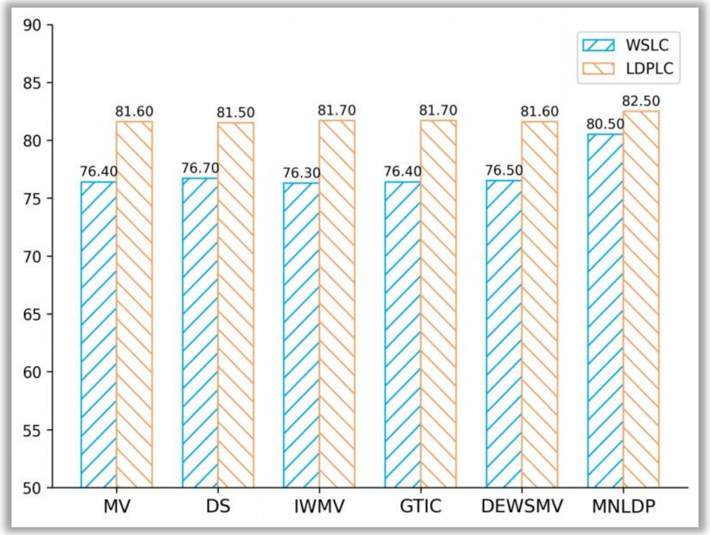


Most workers’ annotation ratios are in the interval [0, 0.1].

Experiments and Results

CUG-Miner

◆ The detailed experimental results on the real-world dataset “LabelMe”:



◆ The integration accuracies of MV (81.60%) and others after label completion by LDPLC are much higher than those of MV (76.40%) and others after label completion by WSLC, respectively.

◆ All these experimental results demonstrate the effectiveness of LDPLC.

Experiments and Results

CUG-Miner

- **Experiments on Simulated Datasets:** 34 datasets published on the CEKA platform
- **Experimental settings:**
  - the number of crowd workers: 40
  - the annotation ratio of each crowd worker: uniform distribution [0, 0.1]
  - the annotation quality of each crowd worker: uniform distribution [0.6, 0.9]

(Here corresponds to the annotation ratio and annotation quality of "LabelMe")

Dataset	#Instances	#Features	#Classes	Missing	Feature type
anneal	898	38	6	yes	hybrid
audiology	226	69	24	yes	nominal
autos	205	25	7	yes	hybrid
balance-scale	625	4	3	no	numeric
biodeg	1055	41	2	no	numeric
breast-cancer	286	9	2	yes	nominal
breast-w	699	9	2	yes	numeric
car	1728	6	4	no	nominal
credit-a	690	15	2	yes	hybrid
credit-g	1000	20	2	no	hybrid
diabetes	768	8	2	no	numeric
heart-c	303	13	5	yes	hybrid
heart-h	294	13	5	yes	hybrid
heart-statlog	270	13	2	no	numeric
hepatitis	155	19	2	yes	hybrid
horse-colic	368	22	2	yes	hybrid
hypothyroid	3772	29	4	yes	hybrid
ionosphere	351	34	2	no	numeric
iris	150	4	3	no	numeric
kr-vs-kp	3196	36	2	no	nominal
labor	57	16	2	yes	hybrid
letter	20000	16	26	no	numeric
lymph	148	18	4	no	hybrid
mushroom	8124	22	2	yes	nominal
segment	2310	19	7	no	numeric
sick	3772	29	2	yes	hybrid
sonar	208	60	2	no	numeric
spambase	4601	57	2	no	numeric
tic-tac-toe	958	9	2	no	nominal
vehicle	846	18	4	no	numeric
vote	435	16	2	yes	nominal
vowel	990	13	11	no	hybrid
waveform	5000	40	3	no	numeric
zoo	101	17	7	no	hybrid

Experiments and Results

◆ The detailed experimental results on 34 simulated datasets:

	MV		DS		IWMV		GTIC		DEWSMV		MNLDP	
Dataset	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC
anneal	70.45	86.31 •	68.21	86.24 •	70.41	86.29 •	69.82	86.26 •	70.35	86.31 •	74.70	86.38 •
audiology	69.25	81.19 •	69.20	81.24 •	69.11	81.33 •	68.98	80.44 •	69.12	81.19 •	71.20	80.62 •
autos	71.76	82.68 •	71.61	82.68 •	71.81	82.63 •	71.85	81.95 •	71.66	82.68 •	73.95	81.17 •
balance-scale	75.15	83.46 •	72.34	83.01 •	74.83	83.34 •	73.46	83.34 •	74.99	83.44 •	79.49	84.16 •
biodeg	77.95	81.42 •	77.72	81.31 •	77.92	81.39 •	77.82	81.44 •	77.92	81.44 •	79.99	81.75 •
breast-cancer	78.39	78.15	77.10	78.14	78.04	78.14	77.80	78.11	78.50	78.32	79.27	77.72
breast-w	78.48	85.81 •	78.37	85.64 •	78.44	85.72 •	78.43	85.84 •	78.44	85.82 •	82.47	88.58 •
car	78.40	86.24 •	75.71	86.25 •	78.41	86.25 •	77.64	86.25 •	78.41	86.25 •	83.81	86.49 •
credit-a	74.59	81.06 •	74.75	81.03 •	74.65	81.05 •	74.64	81.06 •	74.59	81.06 •	76.42	81.32 •
credit-g	77.00	77.80	75.95	77.72	76.76	77.79	76.61	77.79	76.93	77.82	78.44	77.55
diabetes	76.63	78.31	76.05	78.09	76.56	78.15	76.39	78.27	76.69	78.27	78.16	78.18
heart-c	77.33	84.29 •	76.27	84.23 •	77.16	84.19 •	75.58	84.33 •	77.43	84.33 •	80.69	84.16 •
heart-h	77.18	83.74 •	76.67	83.67 •	77.07	83.70 •	76.33	83.64 •	77.14	83.70 •	80.54	83.23
heart-statlog	74.92	78.52 •	74.48	78.59 •	74.44	78.81 •	74.37	78.41 •	74.96	78.56 •	74.26	78.85 •
hepatitis	69.36	83.03 •	69.74	82.84 •	71.29	83.03 •	70.26	82.84 •	69.29	83.03 •	67.55	83.94 •
horse-colic	72.31	77.69 •	72.69	77.58 •	72.39	77.58 •	72.39	77.72 •	72.28	77.69 •	72.94	77.34 •
hypothyroid	83.18	88.79 •	79.58	88.72 •	83.12	88.79 •	82.33	88.79 •	83.18	88.80 •	88.05	89.02
ionosphere	72.59	83.90 •	73.85	83.88 •	73.62	83.82 •	73.59	83.96 •	72.56	83.96 •	75.13	85.02 •
iris	72.00	86.87 •	72.27	87.13 •	71.87	87.13 •	71.80	87.07 •	72.27	86.67 •	75.33	90.93 •
kr-vs-kp	76.04	85.33 •	76.08	85.32 •	76.06	85.32 •	76.02	85.32 •	75.98	85.32 •	78.10	85.61 •
labor	66.14	80.17 •	67.37	82.98 •	63.33	80.35 •	68.95	79.47 •	65.26	80.70 •	50.35	81.75 •
letter	71.18	90.58 •	72.75	91.20 •	73.32	91.19 •	71.20	90.57 •	71.24	90.57 •	78.99	91.32 •
lymph	69.26	84.80 •	69.19	84.80 •	69.53	84.80 •	69.66	84.80 •	69.19	84.80 •	70.14	85.81 •
mushroom	76.66	88.02 •	76.67	88.02 •	76.69	88.02 •	76.67	88.02 •	76.70	88.02 •	79.41	88.37 •
segment	69.64	87.63 •	69.33	87.51 •	69.65	87.53 •	69.67	87.64 •	69.58	87.61 •	75.80	88.46 •
sick	81.26	84.56 •	77.50	84.30 •	81.20	84.54 •	80.50	84.50 •	81.27	84.58 •	84.93	85.12
sonar	71.54	78.56 •	72.40	79.04 •	72.07	78.94 •	72.36	78.80 •	71.54	78.85 •	69.76	79.42 •
spambase	78.36	83.40 •	78.28	83.39 •	78.36	83.39 •	78.33	83.40 •	78.36	83.40 •	80.12	83.64 •
tic-tac-toe	76.75	77.83	75.96	77.80	76.70	77.82	76.54	77.80	76.86	77.83	78.66	76.75 ◦
vehicle	72.67	82.45 •	72.71	82.37 •	72.71	82.38 •	72.72	82.46 •	72.73	82.42 •	76.76	82.35 •
vote	76.57	83.10 •	76.39	83.11 •	76.50	83.11 •	76.46	83.08 •	76.69	83.11 •	78.80	84.09 •
vowel	73.49	91.53 •	73.48	91.53 •	73.47	91.53 •	73.48	91.53 •	73.55	91.53 •	78.23	91.79 •
waveform	74.53	84.84 •	74.01	84.51 •	74.45	84.73 •	74.49	84.84 •	74.55	84.86 •	79.05	85.73 •
zoo	74.06	87.43 •	73.96	87.43 •	73.57	87.53 •	73.47	87.43 •	74.46	87.43 •	79.80	88.62 •
Mean	74.56	83.51	74.08	83.57	74.57	83.54	74.43	83.45	74.55	83.54	76.80	83.99
W/T/L	-	30/4/0	-	30/4/0	-	30/4/0	-	30/4/0	-	30/4/0	-	27/6/1

The symbols • and ◦ in the table indicate that the integration accuracy has a statistically significant improvement or degradation using our proposed LDPLC compared to WSLC, respectively.

These results strongly demonstrates the effectiveness of LDPLC.

Experiments and Results

◆ The annotation quality of each crowd worker: Gaussian distribution $N(0.75, 0.15^2)$

	MV		DS		IWMV		GTIC		DEWSMV		MNLDP	
Dataset	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC	WSLC	LDPLC
anneal	68.30	84.37 ●	65.69	84.16 ●	68.35	84.32 ●	67.59	84.21 ●	68.23	84.36 ●	71.35	84.26 ●
audiology	68.54	81.33 ●	68.41	81.33 ●	68.54	81.33 ●	68.54	79.42 ●	68.63	81.28 ●	69.69	80.44 ●
autos	69.42	81.76 ●	69.27	81.76 ●	69.37	81.80 ●	69.32	81.76 ●	69.27	81.80 ●	71.22	81.12 ●
balance-scale	73.81	83.02 ●	71.31	82.24 ●	73.58	82.82 ●	73.01	82.90 ●	73.73	82.99 ●	78.64	83.74 ●
biodeg	78.40	81.65 ●	78.11	81.56 ●	78.37	81.59 ●	78.26	81.64 ●	78.47	81.66 ●	80.26	81.83
breast-cancer	77.45	77.55	76.61	77.52	77.20	77.52	77.03	77.55	77.31	77.52	79.37	77.59
breast-w	80.04	86.28 ●	79.89	86.15 ●	80.00	86.22 ●	79.88	86.28 ●	80.10	86.32 ●	83.55	89.07 ●
car	79.19	87.03 ●	76.39	87.03 ●	79.23	87.04 ●	78.53	87.04 ●	79.17	87.04 ●	83.89	87.24 ●
credit-a	74.81	80.19 ●	75.41	80.16 ●	75.03	80.22 ●	75.04	80.23 ●	74.83	80.20 ●	76.29	80.67 ●
credit-g	74.48	74.70	73.76	74.63	74.45	74.72	74.43	74.67	74.51	74.70	76.13	74.15
diabetes	76.63	77.88	75.37	77.80	76.42	77.83	76.06	77.75	76.58	77.79	78.11	77.92
heart-c	76.24	84.29 ●	75.25	84.29 ●	75.98	84.29 ●	74.85	84.29 ●	76.17	84.29 ●	80.79	84.49 ●
heart-h	76.12	82.99 ●	74.76	82.99 ●	75.95	82.99 ●	74.59	82.99 ●	76.05	83.02 ●	80.61	83.30
heart-statlog	73.63	78.00	73.70	77.85	73.78	77.82	73.56	78.00	73.67	78.00	74.11	77.78
hepatitis	65.68	79.29 ●	66.58	79.61 ●	67.42	79.74 ●	66.90	79.23 ●	65.68	79.22 ●	61.93	79.93 ●
horse-colic	71.66	78.12 ●	72.06	77.93 ●	71.90	77.93 ●	71.79	78.04 ●	71.66	78.02 ●	72.15	77.47 ●
hypothyroid	83.16	88.49 ●	79.87	88.36 ●	83.13	88.48 ●	82.48	88.47 ●	83.15	88.48 ●	87.74	88.86
ionosphere	68.20	77.95 ●	70.09	78.52 ●	70.20	78.32 ●	70.17	77.89 ●	68.32	77.92 ●	69.54	79.03 ●
iris	71.07	87.13 ●	71.40	87.33 ●	72.40	86.87 ●	72.87	87.67 ●	70.60	87.07 ●	75.53	91.47 ●
kr-vs-kp	75.78	84.25 ●	75.77	84.24 ●	75.80	84.25 ●	75.75	84.23 ●	75.81	84.23 ●	77.50	84.62 ●
labor	60.88	79.30 ●	62.98	79.65 ●	62.98	79.83 ●	66.32	79.12 ●	60.70	78.95 ●	39.65	78.42 ●
letter	71.44	90.74 ●	67.67	88.55 ●	71.87	90.64 ●	71.42	90.74 ●	71.43	90.73 ●	79.52	91.53 ●
lymph	69.05	83.51 ●	69.05	83.51 ●	69.39	83.45 ●	69.32	83.45 ●	69.12	83.58 ●	71.28	83.85 ●
mushroom	76.96	87.85 ●	76.97	87.85 ●	76.97	87.85 ●	76.96	87.85 ●	76.95	87.85 ●	79.04	88.20 ●
segment	70.75	87.53 ●	70.75	87.50 ●	70.75	87.48 ●	70.74	87.55 ●	70.72	87.53 ●	77.20	88.59 ●
sick	79.42	82.33	75.47	81.96 ●	79.40	82.36	78.50	82.26 ●	79.36	82.32	82.93	82.99
sonar	69.81	78.99 ●	70.15	79.62 ●	69.62	79.52 ●	70.14	79.18 ●	69.66	79.04 ●	67.89	78.94 ●
spambase	76.87	81.63 ●	76.83	81.63 ●	76.87	81.62 ●	76.85	81.64 ●	76.88	81.64 ●	78.43	82.01 ●
tic-tac-toe	76.36	77.47	75.27	77.47	76.36	77.50	76.09	77.50	76.34	77.46	78.53	76.65 ○
vehicle	72.78	81.97 ●	72.86	81.97 ●	72.86	81.93 ●	72.94	81.98 ●	72.80	82.08 ●	76.50	82.17 ●
vote	77.61	84.99 ●	77.49	84.99 ●	77.68	85.01 ●	77.68	84.99 ●	77.75	84.99 ●	79.91	86.35 ●
vowel	71.18	89.94 ●	71.22	89.94 ●	71.23	89.94 ●	71.20	89.94 ●	71.28	89.94 ●	75.86	90.17 ●
waveform	72.65	83.36 ●	72.22	82.98 ●	72.51	83.25 ●	72.59	83.32 ●	72.56	83.35 ●	77.52	84.31 ●
zoo	74.46	86.44 ●	74.56	86.44 ●	73.96	86.44 ●	73.57	86.54 ●	74.46	86.44 ●	79.80	87.43 ●
Mean	73.61	82.72	73.04	82.63	73.81	82.73	73.68	82.66	73.59	82.70	75.66	83.13
W/T/L	-	28/6/0	-	29/5/0	-	28/6/0	-	29/5/0	-	28/6/0	-	25/8/1

These results validate the robustness of LDPLC under different annotation quality distributions.

Conclusions

- We design a worker similarity weighted majority voting algorithm to initialize a label distribution for each missing label to represent the label information of similar workers.

- We design a label distribution propagation algorithm to enable each missing label of each instance to iteratively absorb its neighbors' label distributions.

- We propose a label distribution propagation-based label completion (LDPLC) algorithm and validate its effectiveness on a large number of crowdsourced datasets.

CUG-Miner

Thank you!