Joint Generative-Discriminative Aggregation Model for Multi-Option Crowd Labels

Kamran Ghasedi Dizaji[†], Yanhua Yang[‡], Heng Huang[†]

kag221@pitt.edu, yanhyang@xidian.edu.cn, heng.huang@pitt.edu

[†]Electrical and Computer Engineering Department, University of Pittsburgh, USA [‡]School of Electronic Engineering, Xidian University, China

Motivations

- Crowd labels are often *noisy* and unreliable, since crowd workers are usually *inexpert* in the assigned tasks.
- A *crowdsourcing aggregation model* is required to estimate the true labels by aggregating the redundant crowd labels.
- Recent studies have shown that crowd workers cannot completely convey their *non-deterministic* beliefs with the *single-option* crowd labels.
- The standard aggregation models are often incompatible with the *multi-option* crowd labels, and they are only able to handle single-option crowd data.

Contributions

- Proposing a new *discriminative* aggregation model with *convex* problem, and deriving an efficient *optimization* algorithm to solve the corresponding problem.
- Introducing a novel *joint generative-discriminative* aggregation model, coupled via a probabilistic framework.
- Achieving superior or competitive results on *single-option* and *multi-option* crowdsourcing datasets with much *faster running speed* compared to alternative models.

Discriminative Aggregation Model

Objective function:

$$\min_{\mathbf{w},\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} \sum_{i=1}^{N} \|\mathbf{X}_i\mathbf{w}-\mathbf{y}_i\|_1 + \lambda_w\|\mathbf{w}\|_2^2$$

Approximation:

$$\min_{\mathbf{w},\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} \sum_{i=1}^{N} \mathbf{w}^T(\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i+\lambda_w\mathbf{I})\mathbf{w} - 2\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w} + \mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i$$

Update w:

$$\min_{\mathbf{w}} \sum_{i=1}^{N} \mathbf{w}^T(\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i+\lambda_w\mathbf{I})\mathbf{w} - 2\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w}$$

Update Y:

$$\min_{\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} \sum_{i=1}^{N} \mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i - 2\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w}$$

Algorithm 1: CWMV$_{\ell 1}$ aggregation model

Initialize $\mathbf{Y}$ by majority voting
while *not converged* **do**
$\mathbf{w} = \left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i + \lambda_w\mathbf{I}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{U}_i\mathbf{y}_i\right)$
$\min_{\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} \mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i - 2\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w} \quad \forall i\in\{1,\dots,N\}$
end

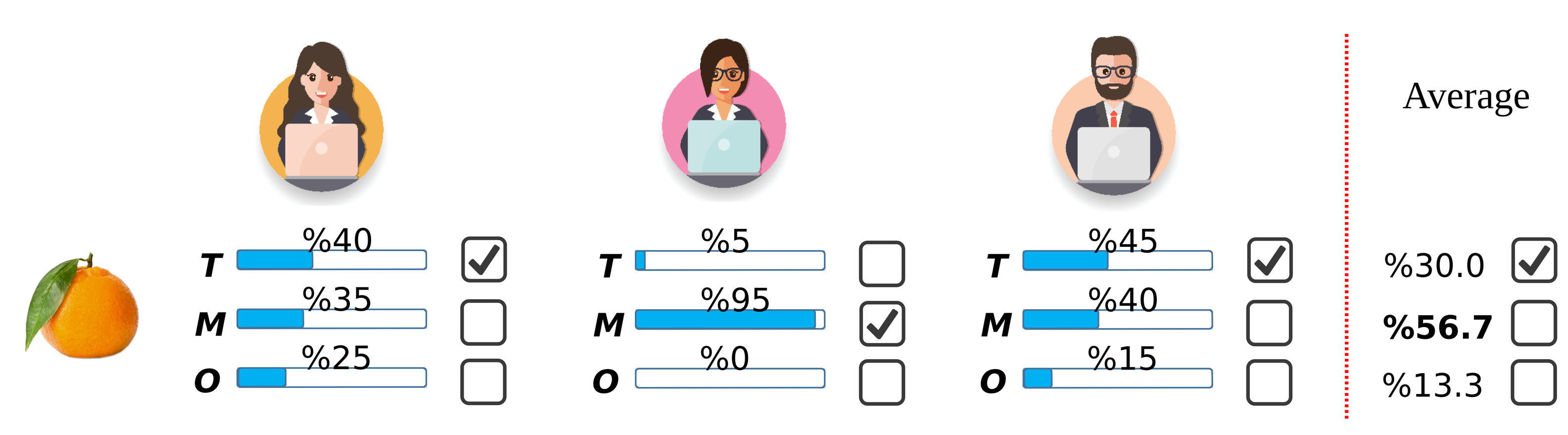

Figure 1: Three crowd workers are asked to classify a figure as tangor (T), mandarin (M) or orange (O). Their single-option and multi-option crowd labels are shown with checked boxes and confidence bars respectively. The average score of multi-option labels correctly shows higher chance for mandarin, while the majority of single-option labels incorrectly suggests tangor as the truth.

Joint Generative-Discriminative Aggregation Model

Objective function:

$$\min_{\mathbf{w},\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1,\boldsymbol{\nu}_j} \mathbf{KL}(p^{\nu}\|p_0^{\nu}) - \sum_{ijck} x_{ijk}y_{ic}\log(\nu_{jck}) + \gamma\sum_i \|\mathbf{X}_i\mathbf{w}-\mathbf{y}_i\|_1 + \lambda_w\|\mathbf{w}\|_2^2$$

Approximation:

$$\min_{\mathbf{w},\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1,\boldsymbol{\nu}_j} \mathbf{KL}(p^{\nu}\|p_0^{\nu}) - \sum_{ijck} x_{ijk}y_{ic}\log(\nu_{jck}) + \sum_i \mathbf{w}^T(\gamma\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i+\lambda_w\mathbf{I})\mathbf{w} - 2\gamma\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w} + \gamma\mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i$$

Update Y:

$$\min_{\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} - \sum_{ijck} x_{ijk}y_{ic}\log(\nu_{jck}) + \sum_i \mathbf{w}^T(\gamma\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i+\lambda_w\mathbf{I})\mathbf{w} - 2\gamma\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w} + \gamma\mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i$$

Update $\boldsymbol{\nu}$:

$$\min_{\boldsymbol{\nu}_j} \mathbf{KL}(p^{\nu}\|p_0^{\nu}) - \sum_{ijck} x_{ijk}y_{ic}\log(\nu_{jck})$$

Update w:

$$\min_{\mathbf{w}} \sum_{i=1}^{N} \mathbf{w}^T(\gamma\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i+\lambda_w\mathbf{I})\mathbf{w} - 2\gamma\mathbf{y}_i^T\mathbf{U}_i\mathbf{X}_i\mathbf{w}$$

Algorithm 2: DS-CWMV$_{\ell 1}$ aggregation model

Initialize $\mathbf{Y}$ by majority voting
while *not converged* **do**
$\mathbf{w} = \left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{U}_i\mathbf{X}_i + \frac{\lambda_w}{\gamma}\mathbf{I}\right)^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i^T\mathbf{U}_i\mathbf{y}_i\right)$
$p^{\nu_{jck}} = Dir(\mu + \sum_i x_{ijk}y_{ic}) \quad \forall j\in\{1,\dots,M\}, \forall\{c,k\}\in\{1,\dots,B\}$
$\min_{\mathbf{y}_i\geq 0,\mathbf{1}^T\mathbf{y}_i=1} \mathbf{y}_i^T\mathbf{U}_i\mathbf{y}_i - 2\mathbf{y}_i^T\left(\mathbf{U}_i\mathbf{X}_i\mathbf{w} + \frac{1}{\gamma}\sum_{jk} x_{ijk}\log(\boldsymbol{\nu}_{jk})\right) \quad \forall i\in\{1,\dots,N\}$
end

	Model	*Web Search*	*Age*	*RTE*	*Temp*	*Flowers*	Average
baselines	*MV*	26.90	34.88	10.31	6.39	22.00	28.83
	IWMV	15.04	34.53	8.12	5.84	19.00	17.09
	M^{3}V	12.74	33.33	7.88	6.06	13.50	15.43
	DS	16.92	39.62	7.25	5.84	13.00	18.69
	DS+Prior	13.26	34.53	[7.13]	5.84	13.50	15.80
	GLAD	19.30	35.73	**7.00**	[5.63]	13.50	16.20
	Entropy (M)	11.10	**31.14**	7.50	[5.63]	13.00	14.03
	Entropy (O)	10.40	37.32	-	-	-	17.76
	CrowdSVM	9.42	33.33	7.75	[5.63]	13.50	13.65
	G-CrowdSVM	7.99±0.26	32.98±0.36	7.67±0.19	5.71±0.33	[12.10±1.07]	12.78±0.31
ours	*CWMV$_{\ell 2}$*	10.89	34.43	7.25	[5.63]	16.00	14.65
	CWMV$_{\ell 1}$	10.70	34.23	7.50	[5.63]	13.00	14.43
	DS-CWMV$_{\ell 2}$	[7.58]	32.04	[7.13]	[5.63]	13.00	[12.32]
	DS-CWMV$_{\ell 1}$	**6.78**	[31.54]	**7.00**	**5.41**	**10.00**	**11.65**

Table 1: Error rates (%) of aggregation models on single-option crowdsourcing datasets.

	Model	*A-Flag*	*C-Flag*	*A-Dog*	*C-Dog*	Average
baselines	*(soft) MV*	21.67	20.83	8.98	8.98	12.90
	(soft) DS	22.50	20.83	10.16	9.76	13.70
	(soft) DS+Prior	20.00	19.17	8.98	8.59	12.23
	(soft) Entropy (M)	17.50	16.67	12.89	12.89	14.23
ours	*CWMV$_{\ell 2}$*	16.67	16.67	8.98	8.59	11.30
	CWMV$_{\ell 1}$	[11.67]	[10.83]	[8.59]	[8.20]	[9.31]
	DS-CWMV$_{\ell 2}$	14.17	14.17	9.38	8.59	10.64
	DS-CWMV$_{\ell 1}$	**13.33**	**10.00**	**8.20**	**7.81**	**9.17**

Table 2: Error rates (%) of aggregation models applied on multi-option crowd datasets.

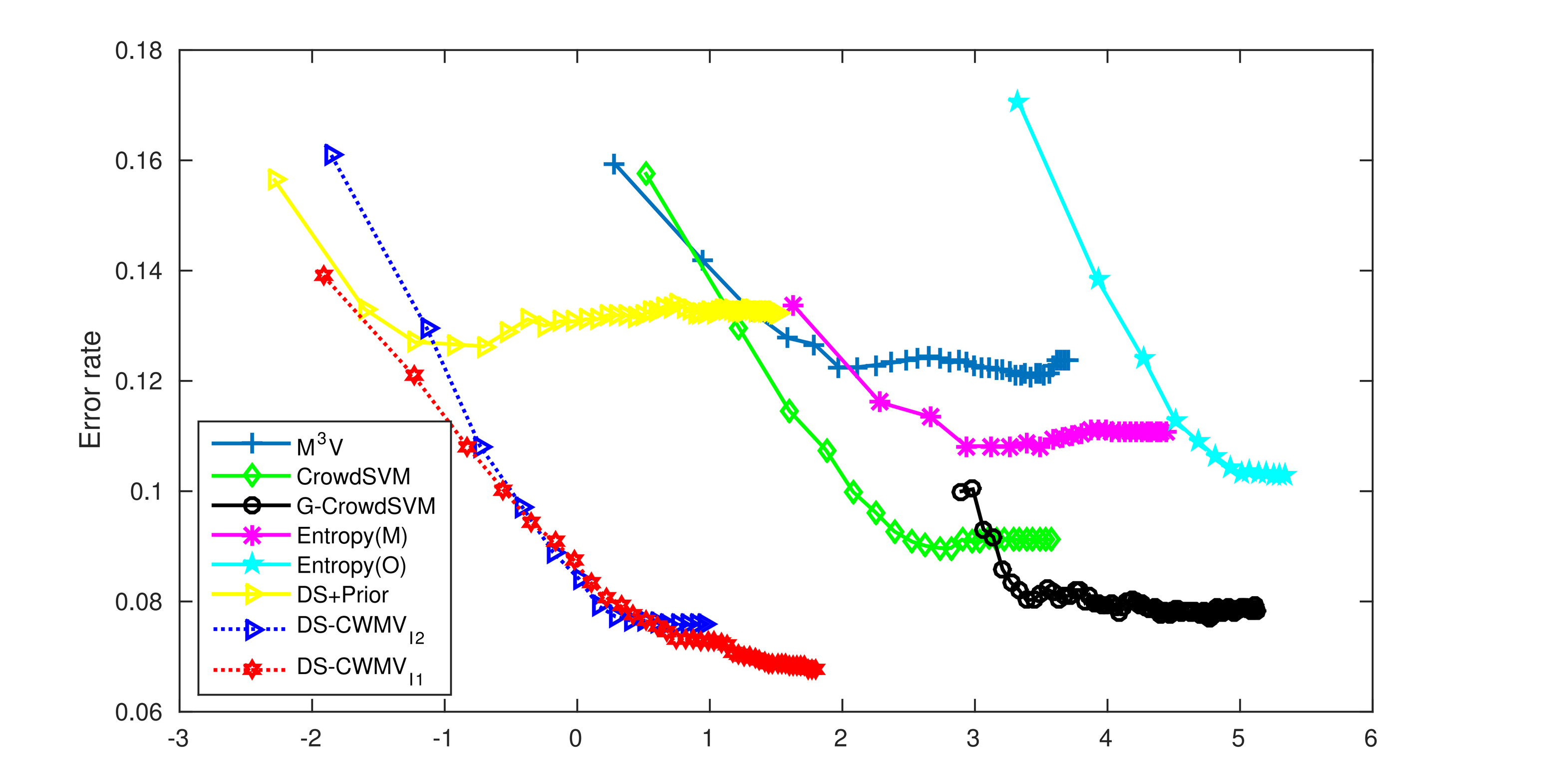

Figure 2: Convergence comparison of aggregation models on *Web Search* dataset.

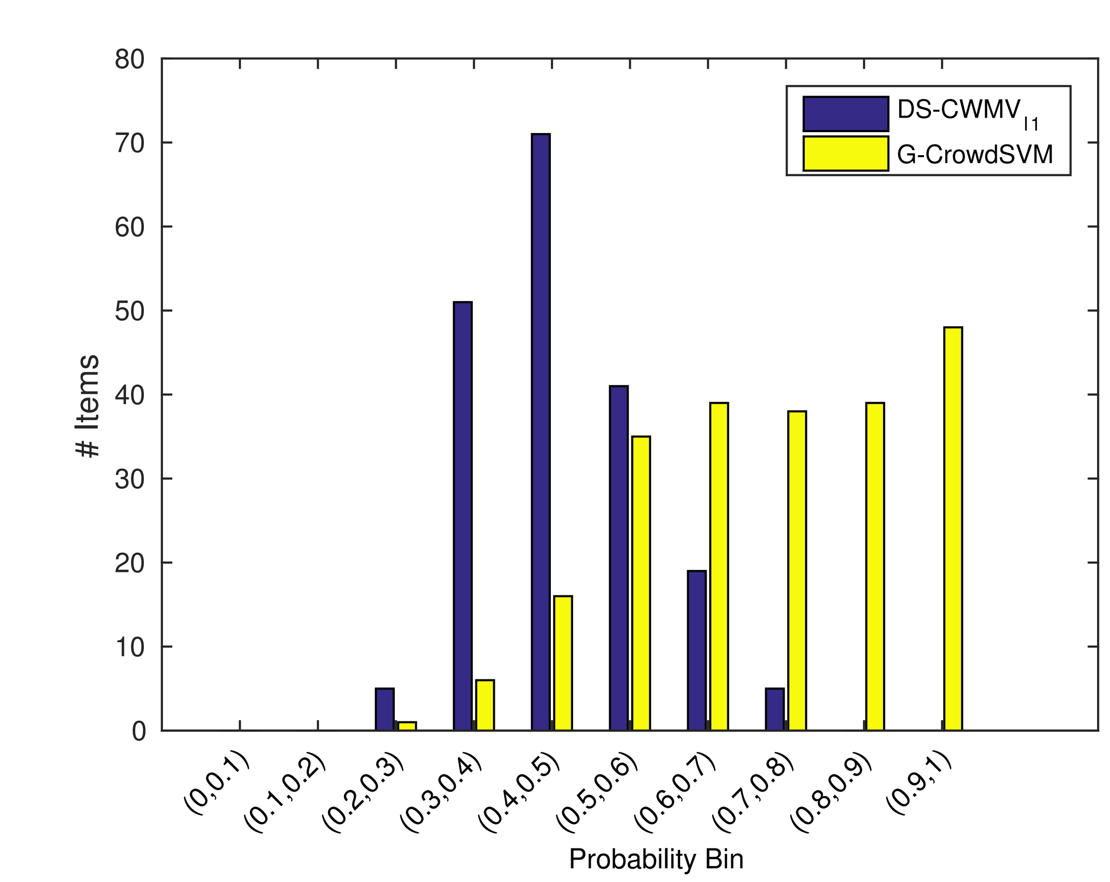

Figure 3: Histogram of the truths for mispredicted items. The results belongs to *DS-*