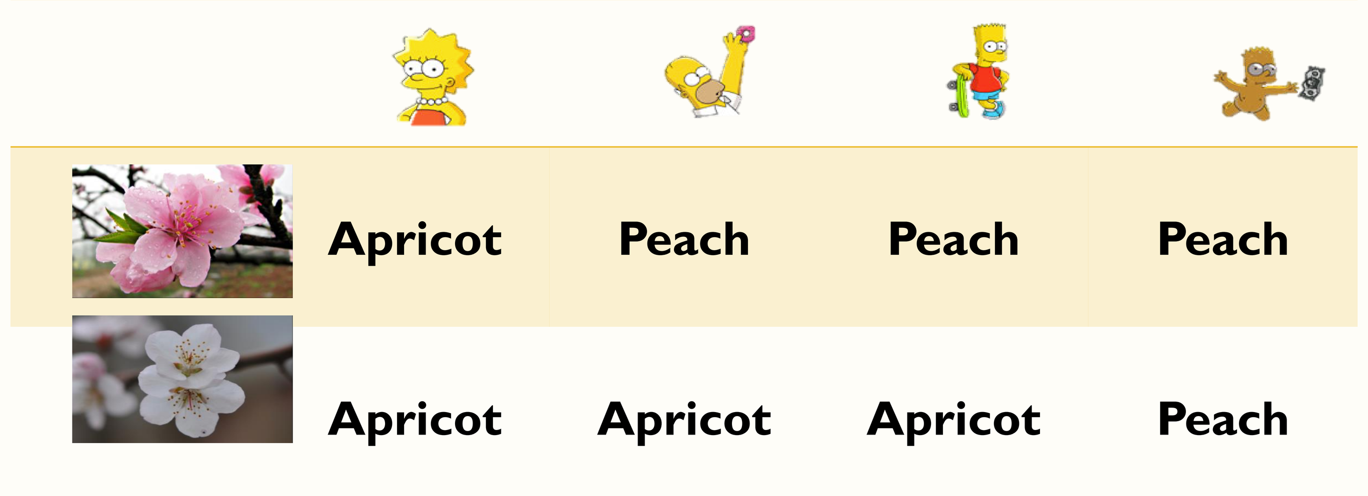
The Twenty-ninth Annual Conference on Neural Information Processing Systems (NIPS) Palais des Congrès de Montréal, Montréal CANADA

Max-Margin Majority Voting for Learning from Crowds

Tian Tian *and* ***Jun Zhu***

from Department of Computer Science and Technology, Tsinghua University.

Objective



Aggregating **noisy** crowdsourcing labels to find **Ground Truths**.

Majority Voting (MV)

Items: $i \in [M]$ Workers: $j \in [N]$ Worker labels: $x_{ij} \in [D]$

Each item have a ground truth: $y_i \in [D]$ $\quad \boldsymbol{x}_i: \{x_{ij}, \forall j\}$

MV: find the most frequent labels

$$\hat{y}_i = \operatorname*{argmax}_{d \in [D]} \sum_{j=1}^{N} \mathbb{I}(x_{ij} = d), \forall i \in [M]$$

Limitations:

1. Workers are equal, lack of discrimination ability
2. Do not consider worker confusability

Constraint Formulation of MV

Expansion Expression

Def: $\boldsymbol{g}(\boldsymbol{x}_i, d) \in \{0,1\}^N$, element j is $\mathbb{I}(x_{ij} = d)$

$\boldsymbol{x}_i$: (1 -1 -1 -1) → $\boldsymbol{g}(\boldsymbol{x}_i, 1)$: (1 0 0 0) ; $\boldsymbol{g}(\boldsymbol{x}_i, -1)$: (0 1 1 1)

Constraint Formulation

MV is equivalent to find $\boldsymbol{y}$ satisfying the constraints:

$$\mathbf{1}_N^\top \boldsymbol{g}(\boldsymbol{x}_i, y_i) - \mathbf{1}_N^\top \boldsymbol{g}(\boldsymbol{x}_i, d) \geq 0, \quad \forall i, d$$

Weighted MV

We introduce worker weights $\boldsymbol{\eta} \in \mathbb{R}^N$, then the constraint formulation changed into

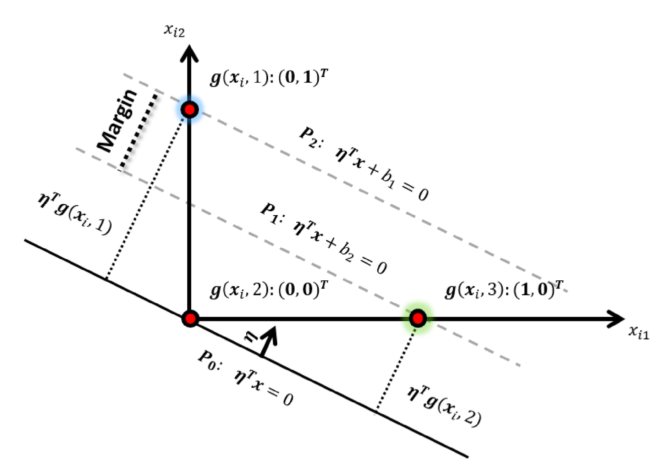
$$\boldsymbol{\eta}^\top \mathbf{g}(\mathbf{x}_i, y_i) - \boldsymbol{\eta}^\top \mathbf{g}(\mathbf{x}_i, d) \geq 0, \forall i, d,$$

The discriminative function for infer ground truth:

$$\hat{y}_i = \operatorname{argmax}_{d \in [D]} \boldsymbol{\eta}^\top \boldsymbol{g}(\boldsymbol{x}_i, d)$$

Max Margin Majority Voting (M³V)

Incorporate max-margin principle to estimate $\boldsymbol{\eta}$



$\boldsymbol{x}_i$: (3 1) → $\boldsymbol{g}(\boldsymbol{x}_i, 1)$: (0 1) ; $\boldsymbol{g}(\boldsymbol{x}_i, 2)$: (0 0) ; $\boldsymbol{g}(\boldsymbol{x}_i, 3)$: (1 0)

$$\inf_{\xi_i \geq 0, \boldsymbol{\eta}, \boldsymbol{y}} \|\boldsymbol{\eta}\|_2^2 + c \sum_i \xi_i$$
$$\text{s.t.}: \boldsymbol{\eta}^\top \boldsymbol{g}_i^\Delta(d) \geq \ell_i^\Delta(d) - \xi_i, \forall i \in [M], d \in [D]$$

where $\boldsymbol{g}_i^\Delta(d) := \boldsymbol{g}(\boldsymbol{x}_i, y_i) - \boldsymbol{g}(\boldsymbol{x}_i, d)$ and $\ell_i^\Delta(d) = \ell \mathbb{I}(y_i \neq d)$.

Here we using soft-margin for robustness.

Solving by iteratively updating $\boldsymbol{\eta}$ and $\boldsymbol{y}$.

Solver for $\boldsymbol{\eta}$:

$$\boldsymbol{\eta} = \sum_{i=1}^{M} \sum_{d=1}^{D} \omega_i^d \boldsymbol{g}_i^\Delta(d)$$

$\boldsymbol{\omega}$ is the solution of the dual problem:

$$\sup_{0 < \omega_i^d < c/2} -\frac{1}{2} \boldsymbol{\eta}^\top \boldsymbol{\eta} + \sum_i \sum_d \omega_i^d \ell_i^\Delta(d),$$

Solver for $\boldsymbol{y}$:

$$\hat{y}_i = \operatorname*{argmax}_{y_i \in [D]} \left(-c \max_{d \in [D]} \left(\ell_i^\Delta(d) - \hat{\boldsymbol{\eta}}^\top \boldsymbol{g}_i^\Delta(d) \right)_+ \right),$$

Dawid-Skene Model (DS)

Define and estimate worker confusion matrices.

ϕ_j is the confusion matrix of worker

$\phi_{jkd} = p(x_{ij} = d | y_i = k), \forall i$

Ground Truth \ Worker Label	Apricot	Peach
Apricot	0.8	0.2
Peach	0.4	0.6

CrowdSVM

Consider Majority Voting and confusability in a single model.

M³V:

$$\inf_{\xi_i \geq 0, \boldsymbol{\eta}, \boldsymbol{y}} \|\boldsymbol{\eta}\|_2^2 + c \sum_i \xi_i$$
$$\text{s.t.}: \boldsymbol{\eta}^\top \boldsymbol{g}_i^\Delta(d) \geq \ell_i^\Delta(d) - \xi_i, \forall i \in [M], d \in [D]$$

\+

DS:

$$\inf_{q(\boldsymbol{\Phi}, \boldsymbol{\eta})} \mathcal{L}(q(\boldsymbol{\Phi}, \boldsymbol{\eta}); \boldsymbol{y}),$$
$$\mathcal{L}(q; \boldsymbol{y}) := \mathrm{KL}(q \| p_0(\boldsymbol{\Phi}, \boldsymbol{\eta})) - \mathbb{E}_q[\log p(\boldsymbol{X} | \boldsymbol{\Phi}, \boldsymbol{y})]$$

→

CrowdSVM:

$$\inf_{\xi_i \geq 0, q \in \mathcal{P}, \boldsymbol{y}} \mathcal{L}(q(\boldsymbol{\Phi}, \boldsymbol{\eta}); \boldsymbol{y}) + c \cdot \sum_i \xi_i$$
$$\text{s.t.}: \mathbb{E}_q[\boldsymbol{\eta}^\top \boldsymbol{g}_i^\Delta(d)] \geq \ell_i^\Delta(d) - \xi_i, \forall i \in [M], d \in [D],$$

Variational Inference

regularized Bayesian inference (Zhu et al. 2015)

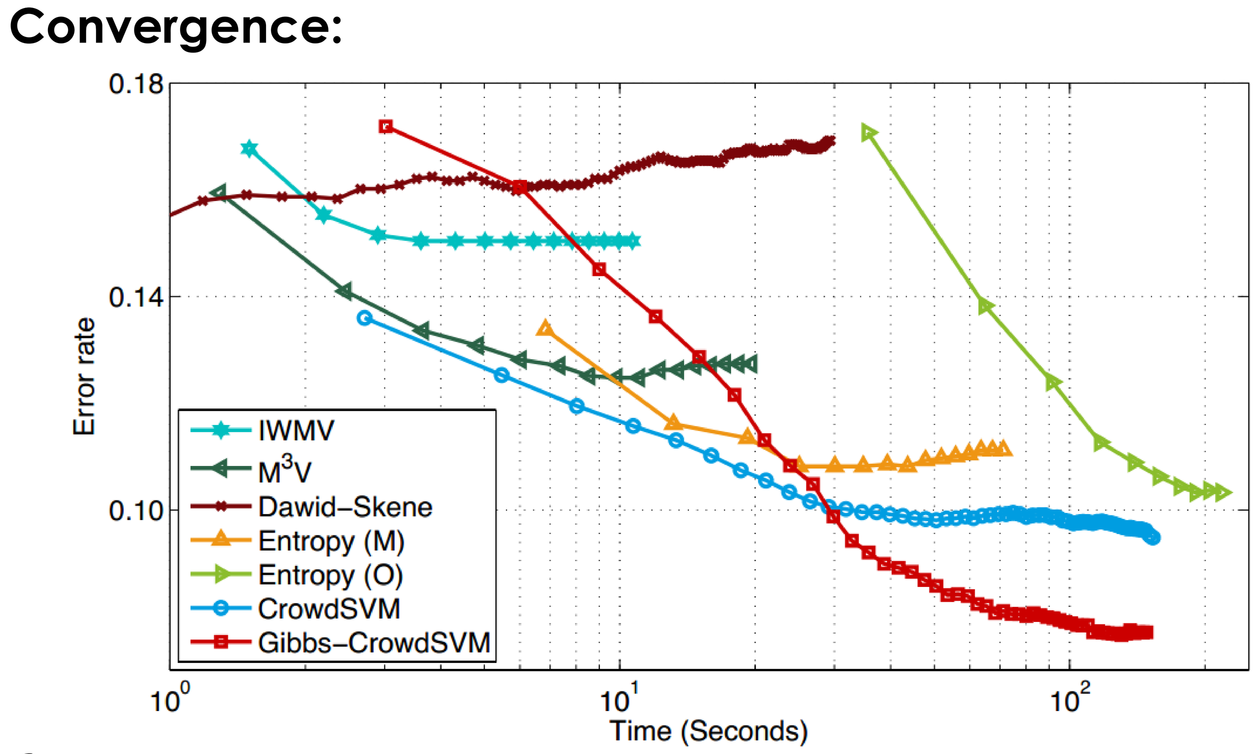
Experimental Results

We test the aggregation error rate on several popular datasets.

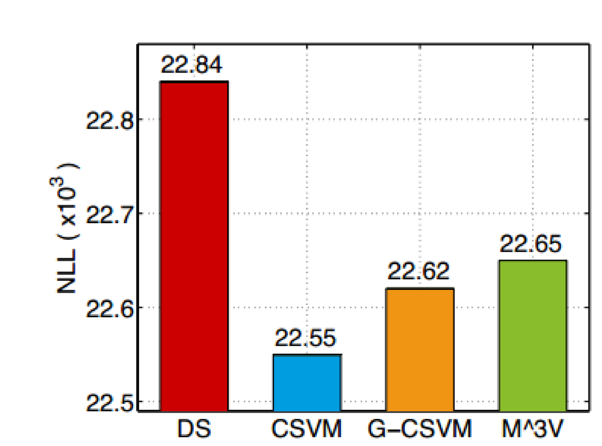
Table 2: Error-rates (%) of different estimators on four datasets.

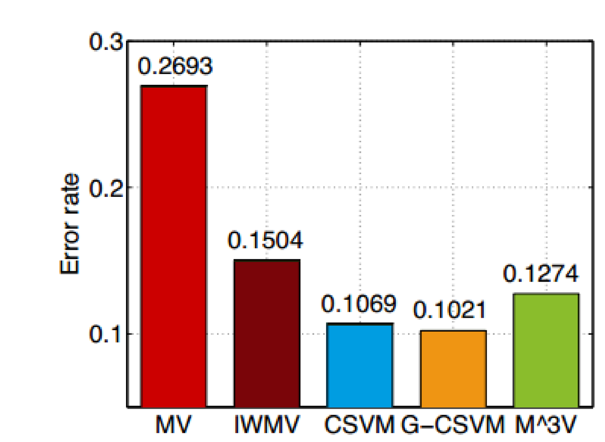
	Method	Web Search	Age	Bluebirds	Flowers
I	MV	26.90	34.88	24.07	22.00
	IWMV	15.04	34.53	27.78	19.00
	M³V	**12.74**	**33.33**	**20.37**	**13.50**
II	DS	16.92	39.62	10.19	**13.00**
	DS+Prior	13.26	34.53	10.19	13.50
	CrowdSVM	**9.42**	**33.33**	10.19	13.50
III	ME	10.40	**31.14**	**8.33**	13.00
	G-CrowdSVM	**7.99 ± 0.26**	32.98 ± 0.36	10.37±0.41	**12.10 ± 1.07**

Convergence:



Generative vs. Discriminative:





Conclusion:

1.Max-margin principle can enhance majority voting.

2.Both generative and discriminative component benefits from the other.

Contact Info

Tian Tian (田天) PhD Student @ THU

Email: tiant13@mails.Tsinghua.edu.cn

Welcome for discussing!