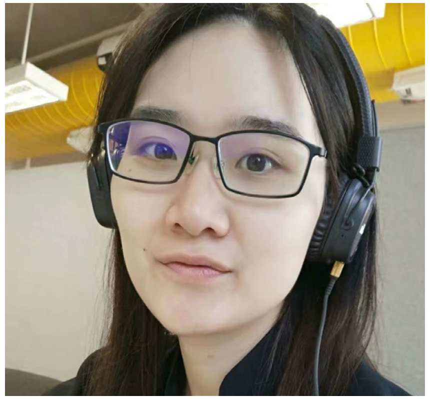
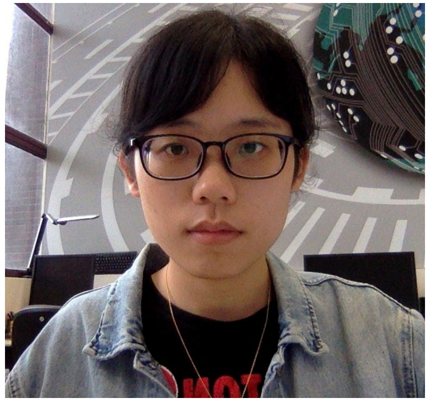
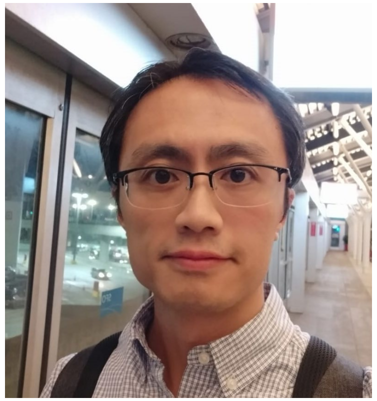
MiSC: Mixed Strategies Crowdsourcing

Ching-Yun Ko, **Rui Lin^**, Shu Li, Ngai Wong

★ **arXiv:** https://arxiv.org/abs/1905.07394



A joint research by

^ Presenter

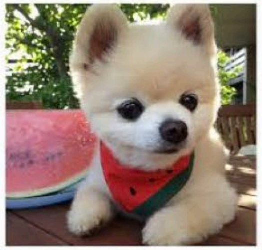
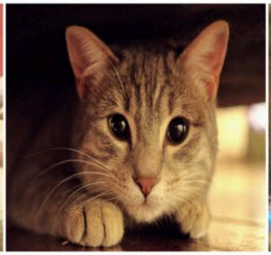
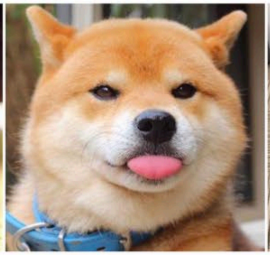
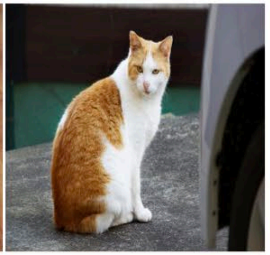
Why do we need crowdsourcing?

Acquiring label data from domain experts or well-trained workers is usually expensive and time-consuming.

Crowdsourcing Task

Obtaining label data from crowd workers is usually cheap and easy*. However, some of the data can be unreliable.

* Amazon Mechanical Turk (AMT), CrowdFlower, and Guru etc.…

item / worker								
Worker 1	2	3		1	1			2
Worker 2	1		2	1		2	3	
Worker 3		3	2		1		1	2
Worker 4	1	2		1		2		3
Worker 5	3		1		2	2	1	

Class 1: Dogs; Class 2: Cats; Class 3: Pigs.

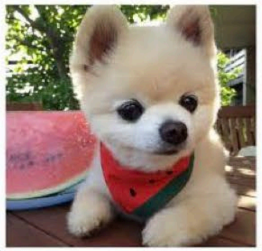
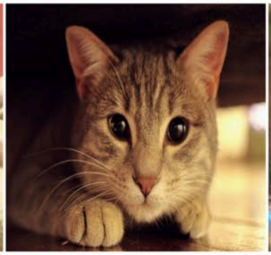
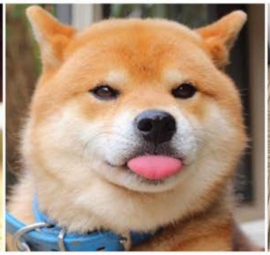
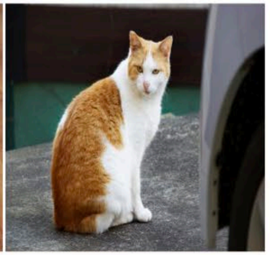
Crowdsourcing Task

vs.

Infer the true labels from a large sum of noisy labels.

Label Aggregation Benchmark Algorithm

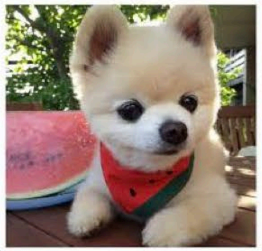
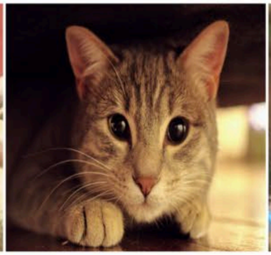
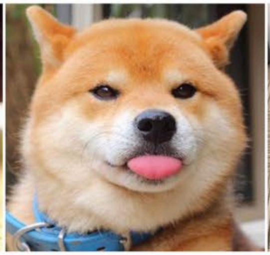
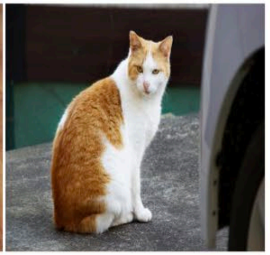
MV	Majority voting
DS-EM	Dawid-Skene model + Expectation Maximization
DS-MF	DS-Mean Field
MMCE(C)	Categorical Minimax Conditional Entropy
MMCE(O)	Ordinal Minimax Conditional Entropy

item / worker								
Worker 1	2	3		1	1			2
Worker 2	1		2	1		2	3	
Worker 3		3	2		1		1	2
☹ Worker 4	1	2		1		2		3
☹ Worker 5	3		1		2	2	1	

Class 1: Dogs; Class 2: Cats; Class 3: Pigs.

Tensor Completion Algorithm

LRTC （Liu et. al., 2013）
TenALS (Jain and Oh, 2014)
Tucker (Tucker, 1966; De Lathauwer et al., 2000; Kim & Choi, 2007)

item / worker								
Worker 1	2	3	2	1	1	1	3	2
Worker 2	1	2	2	1	2	2	3	2
Worker 3	2	3	2	3	1	3	1	2
Worker 4	1	2	3	1	1	2	1	3
Worker 5	3	1	1	2	2	2	1	2

Class 1: Dogs; Class 2: Cats; Class 3: Pigs.

MiSC Strategy

Tensor Completion Algorithms
LRTC
TenALS
Tucker

Label Aggregation Benchmark Algorithms
MV
DS-EM
DS-MF
$\text{MMCE}_{(C)}$
$\text{MMCE}_{(O)}$

A versatile complete-aggregate two-step looping structure.

From Label matrix to Binary 3-way Tensor

$$A = \begin{pmatrix} 1 & 0 & 4 \\ 1 & 3 & 0 \end{pmatrix}$$

$A(2,2) = 3$: The 2nd worker thinks the 2nd item belongs to class 3.

$$\mathcal{A}(1,:,:) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{cc} & \text{Class 3} \\ \mathcal{A}(2,:,:) = & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \; \text{2nd item} \\ \text{2nd worker} & \end{array}$$

$$\mathcal{A}(2,2,3) = 1$$

From A Label Vector to A Slice

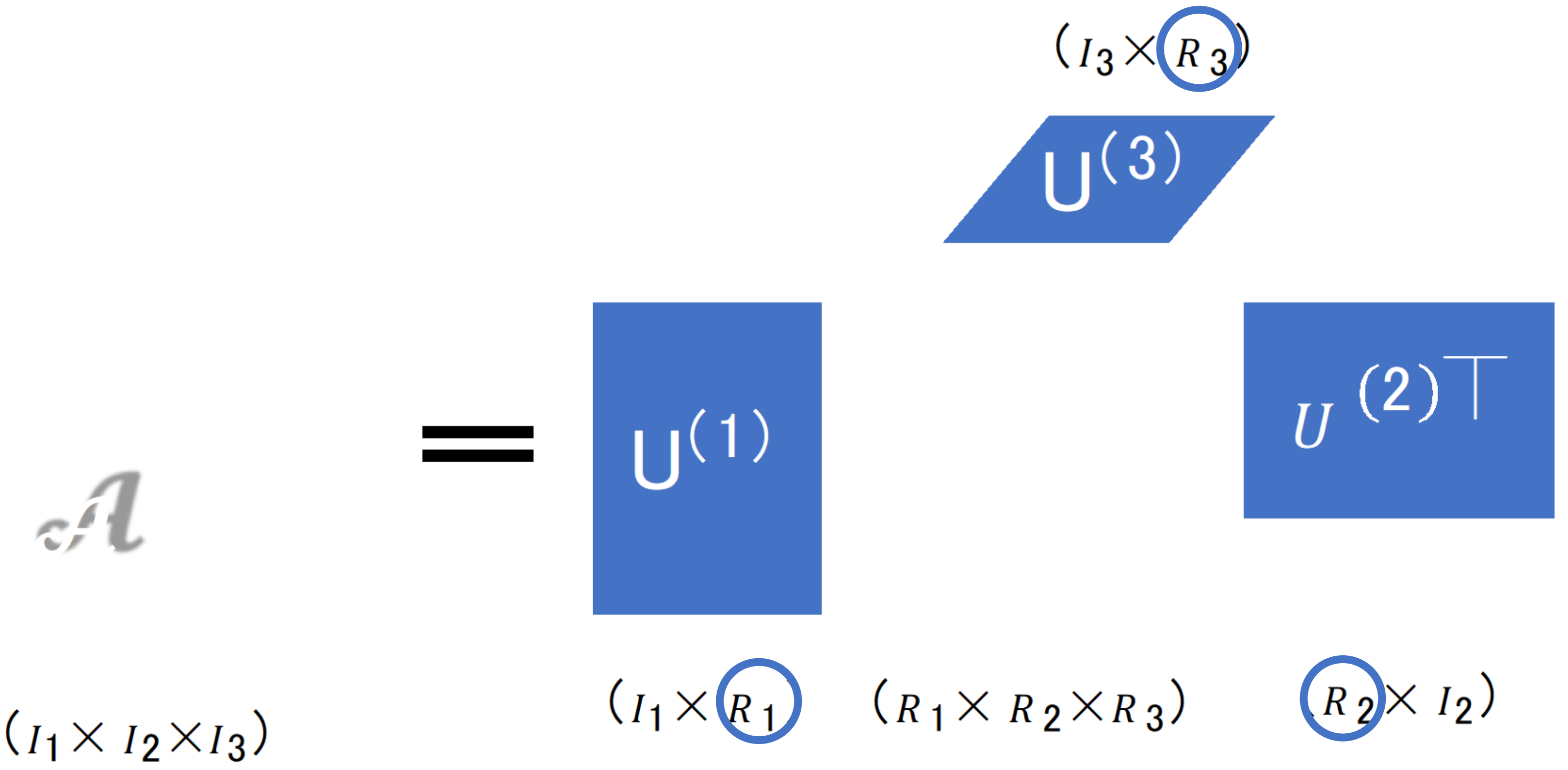
$$s = (1, 3, 2)$$

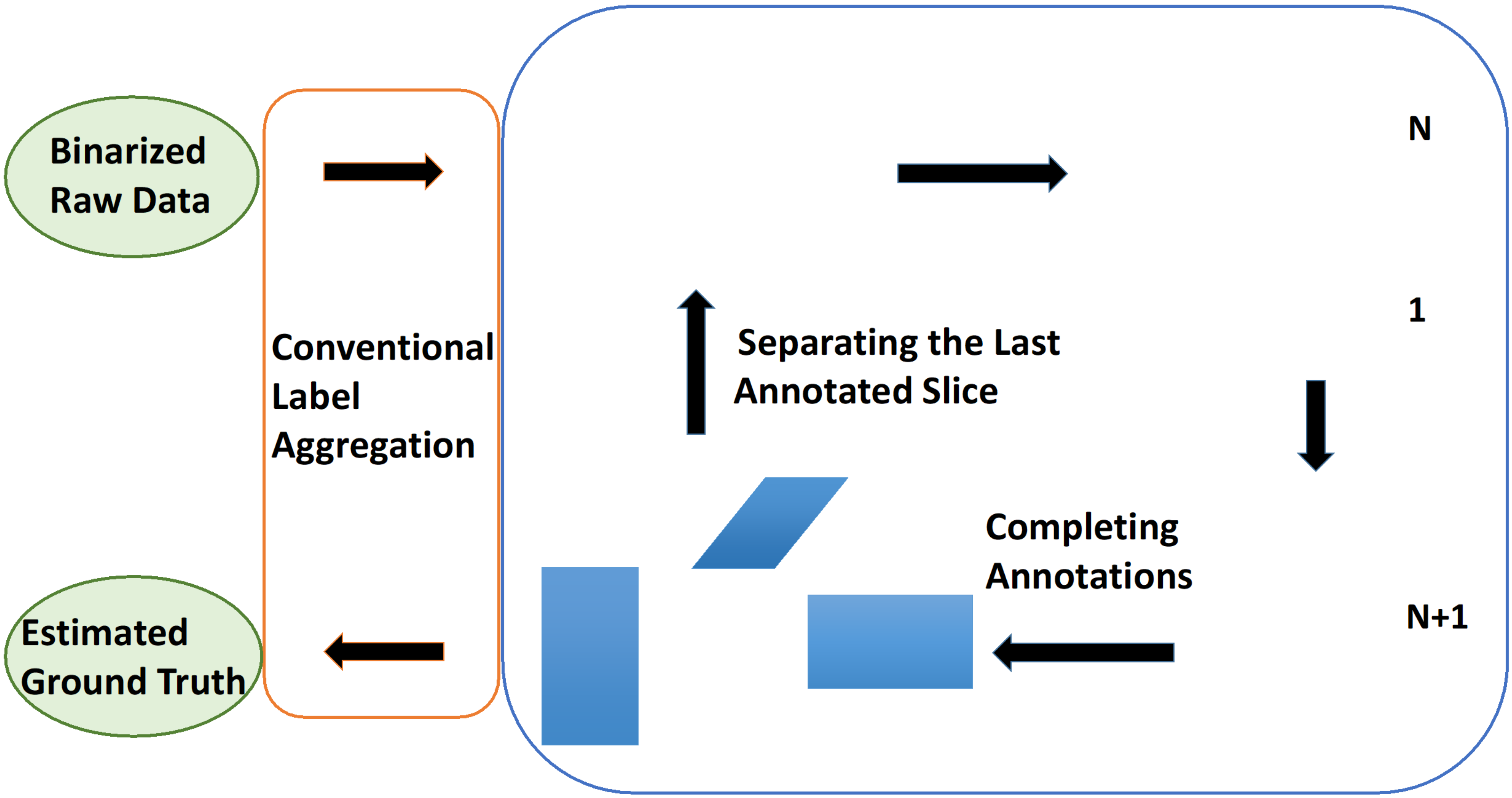
The 3rd item is inferred belonging to class 2.

$$\mathcal{S} = \begin{array}{cc} & \text{Class 2} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} & \begin{array}{l} \\ \\ \text{3rd item} \end{array} \end{array}$$

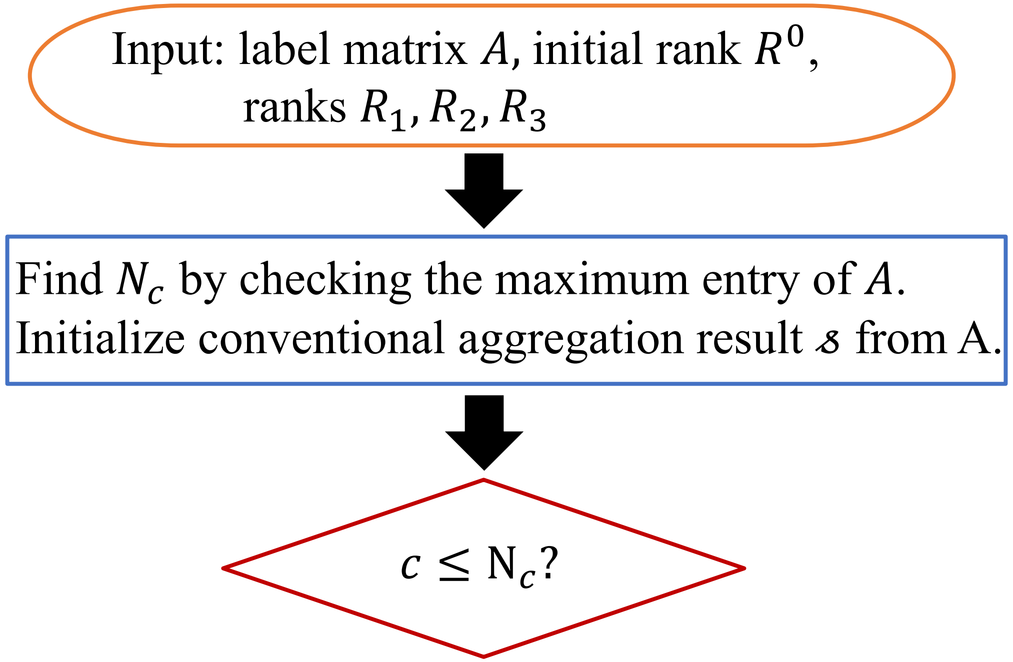
Tucker Decomposition

Tucker decomposition decomposes a tensor into a set of matrices and one small core tensor. (R_1, R_2, R_3) are called tensor ranks, they affect the performance of approximation.

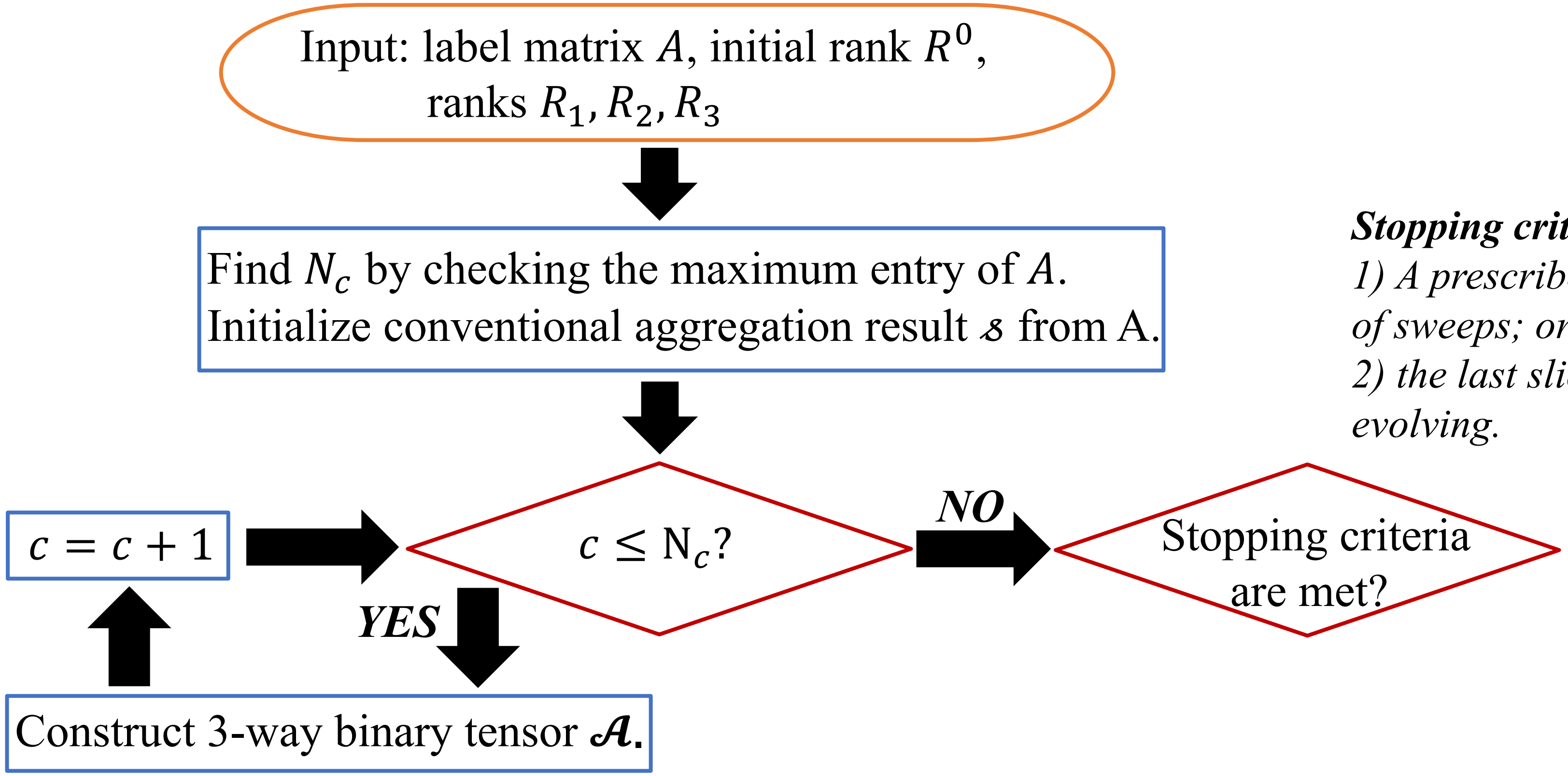



Binarized Raw Data
Conventional Label Aggregation
Estimated Ground Truth
N
1
Separating the Last Annotated Slice
Completing Annotations
N+1

MiSC: Mixed Strategies Crowdsourcing algorithm



MiSC: Mixed Strategies Crowdsourcing algorithm

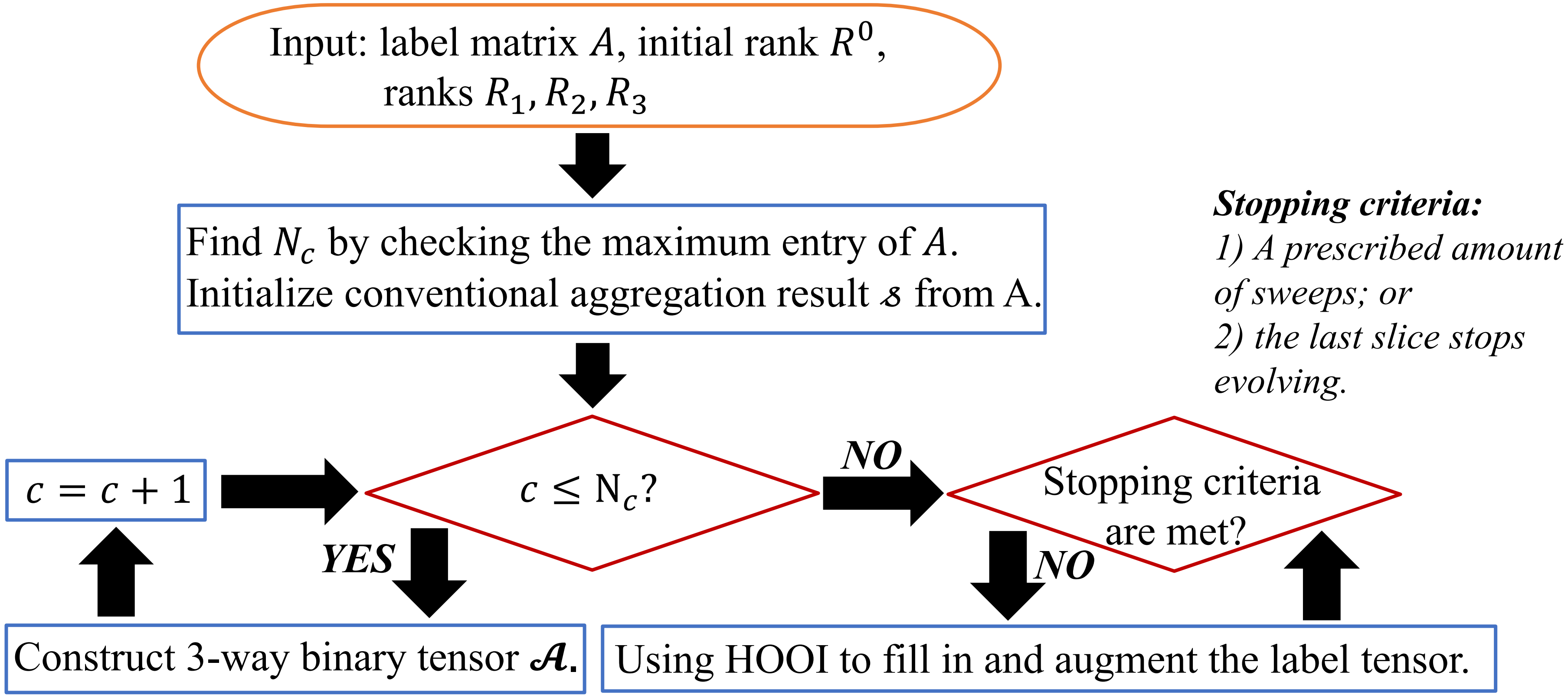


Stopping criteria:

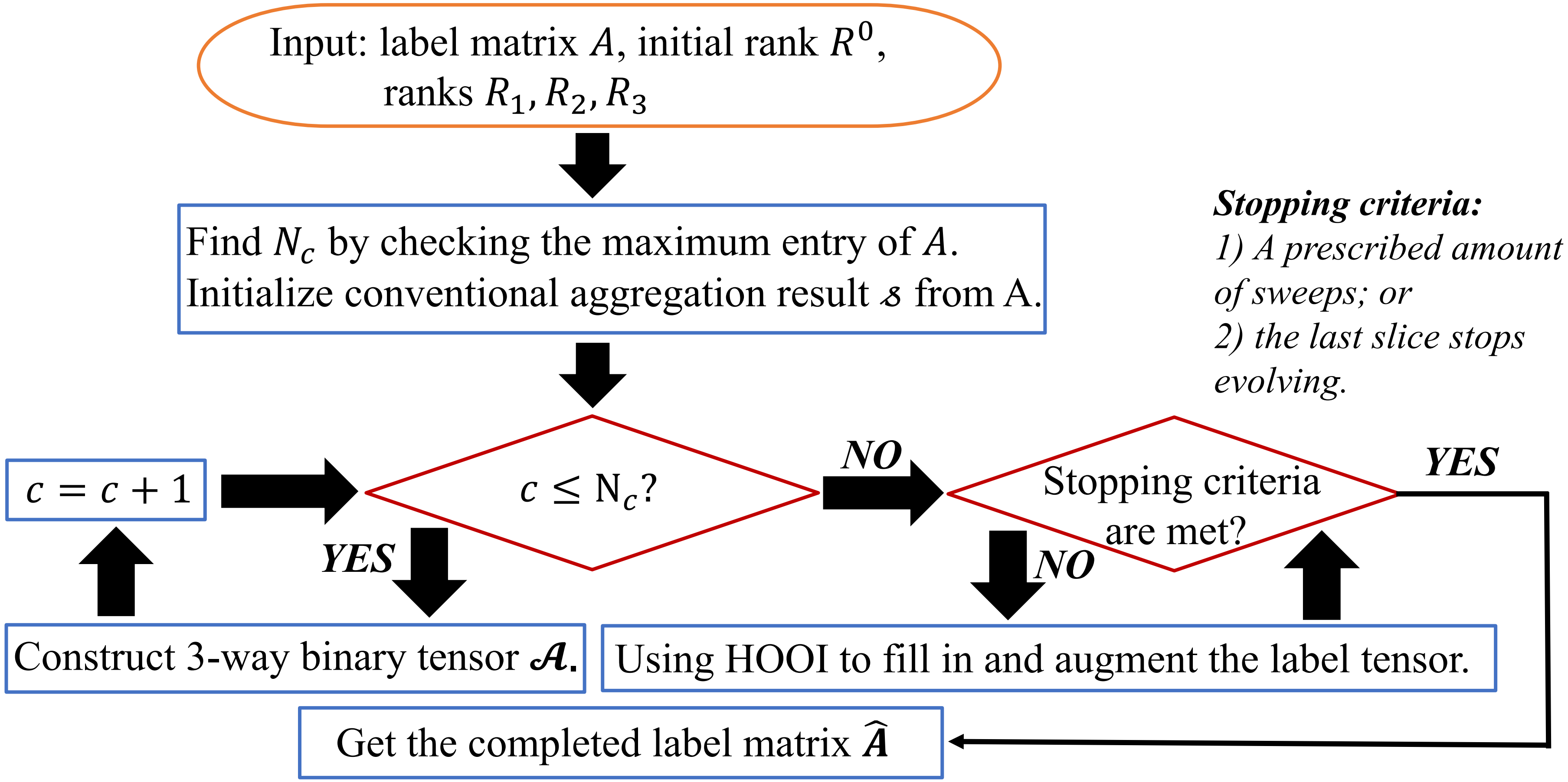
1) A prescribed amount of sweeps; or

2) the last slice stops evolving.

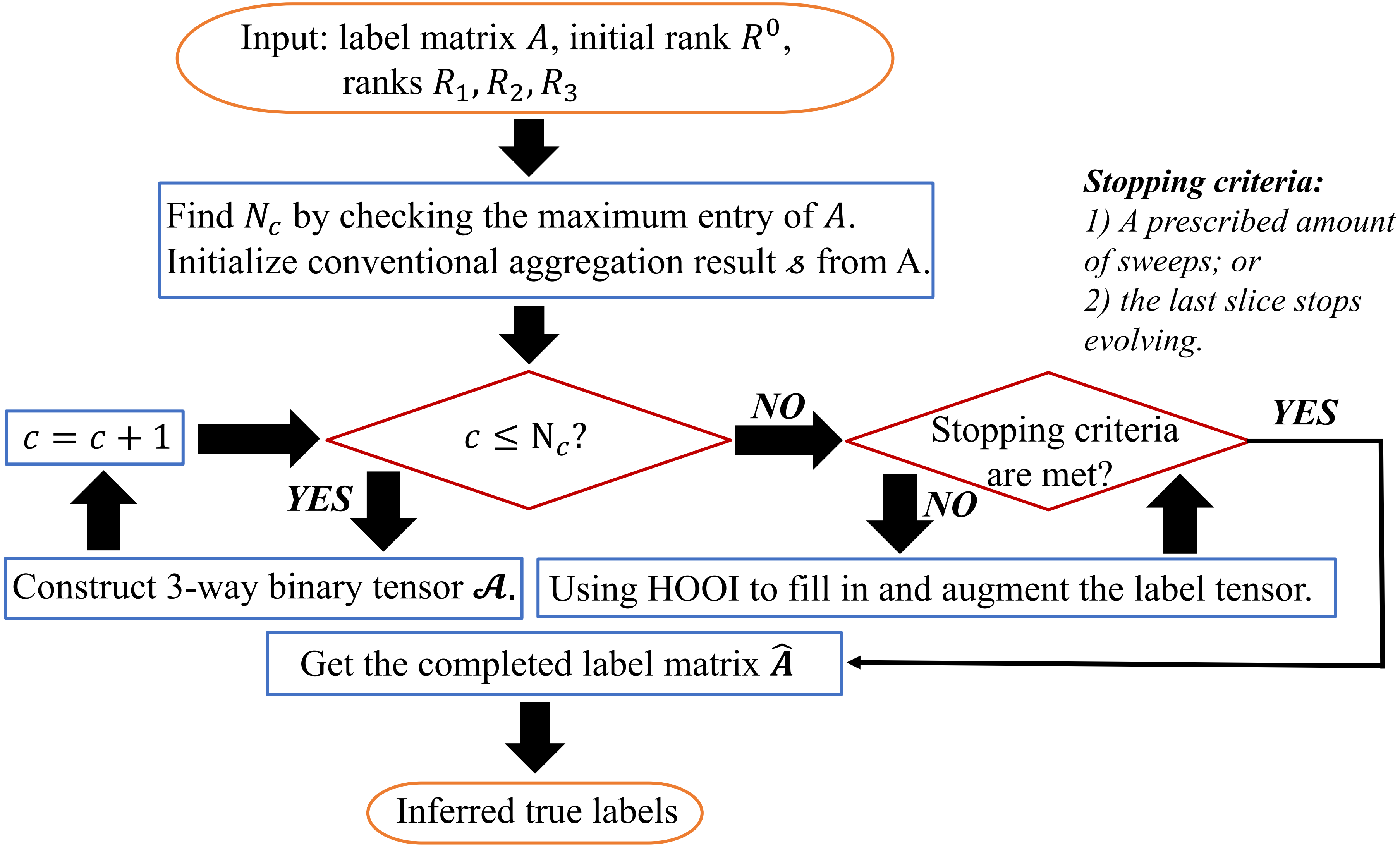
MiSC: Mixed Strategies Crowdsourcing algorithm



MiSC: Mixed Strategies Crowdsourcing algorithm



MiSC: Mixed Strategies Crowdsourcing algorithm



Steps in computing the true labels with an exemplary Tucker completion case.

Experiment 1: MiSC vs. State-of-the-arts pure strategies

Our method has the best performance on Web dataset, which has low nonzero rate ($\frac{\#worker\ labels}{\#items \times \#workers}$) and high relative error rate ($\frac{\#wrong\ labels}{\#worker\ labels}$).

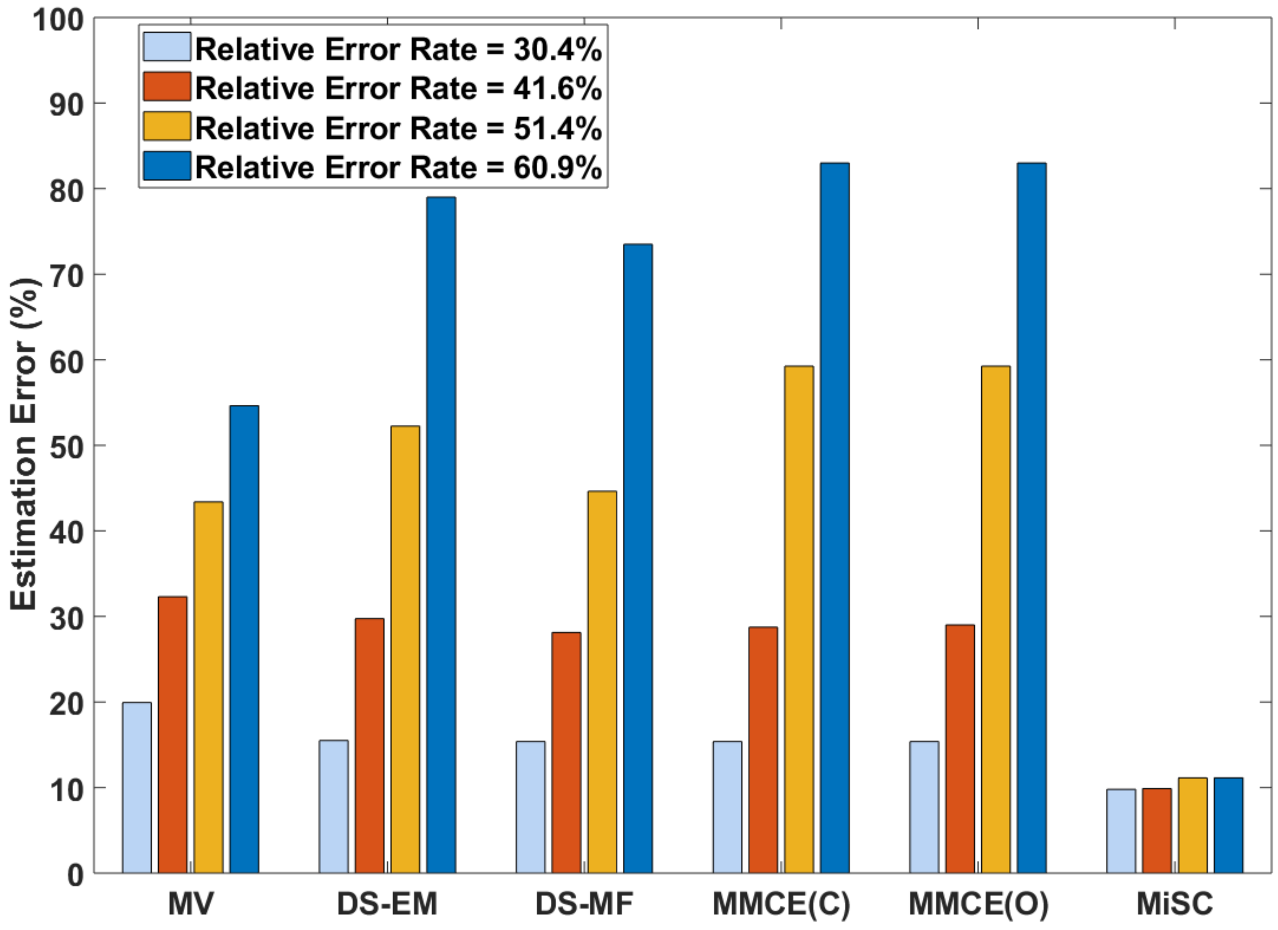
Web(3.3/63.4)	MV	DS-EM	DS-MF	MMCE(C)	MMCE(O)
pure	26.93	16.92	16.10	11.12	10.33
LRTC	26.76	16.55	16.09	11.12	10.33
TenALS	26.93	16.77	15.83	11.12	10.33
Tucker	10.87	5.77	5.73	6.97	5.24

Experiment 1: MiSC vs. State-of-the-arts pure strategies

On other five dataset, MiSC improves the accuracy as well.

Web (3.3/ 63.4)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$	**BM** (6.0/ 31.1)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$
pure	26.93	16.92	16.10	11.12	10.33	pure	30.4	27.60	26.90	27.10	27.10
LRTC	26.76	16.55	16.09	11.12	10.33	LRTC	29.25	27.60	26.90	27.10	27.10
TenALS	26.93	16.77	15.83	11.12	10.33	TenALS	27.60	27.60	26.90	27.10	27.10
Tucker	10.87	5.77	5.73	6.97	**5.24**	Tucker	26.50	27.00	**26.20**	26.40	26.40
RTE (6.1/ 16.3)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$	**Dog** (9.2/ 30.0)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$
pure	10.31	7.25	7.13	7.50	7.50	pure	17.78	15.86	15.61	16.23	16.73
LRTC	9.25	7.25	7.00	7.50	7.50	LRTC	15.61	15.61	15.61	15.61	15.61
TenALS	10.25	7.25	7.13	7.50	7.50	TenALS	15.86	15.74	15.61	15.86	15.86
Tucker	8.38	6.88	**6.75**	7.50	7.50	Tucker	15.61	15.49	**15.37**	15.86	15.86
Temp (13.2/ 15.9)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$	**Bluebirds** (100.0/ 36.4)	MV	DS-EM	DS-MF	$MMCE_{(C)}$	$MMCE_{(O)}$
pure	6.39	5.84	5.84	5.63	5.63	pure	24.07	10.19	10.19	8.33	8.33
LRTC	5.19	5.63	5.63	5.63	5.63	LRTC	20.37	9.26	9.26	6.48	6.48
TenALS	5.41	5.63	5.84	5.63	5.63	TenALS	23.15	9.26	9.26	6.48	6.48
Tucker	5.19	**4.98**	**4.98**	5.41	5.41	Tucker	19.91	8.33	9.26	**4.63**	**4.63**

Experiment 2: MiSC for Sparse and Noisy Annotations



Estimation errors (%) of pure and mixed strategies on highly sparse and severely noisy annotations in RTE dataset.

MiSC is remarkably advantageous, when the data has high sparsity and severe noise.

Conclusions

MiSC has three important advantages:

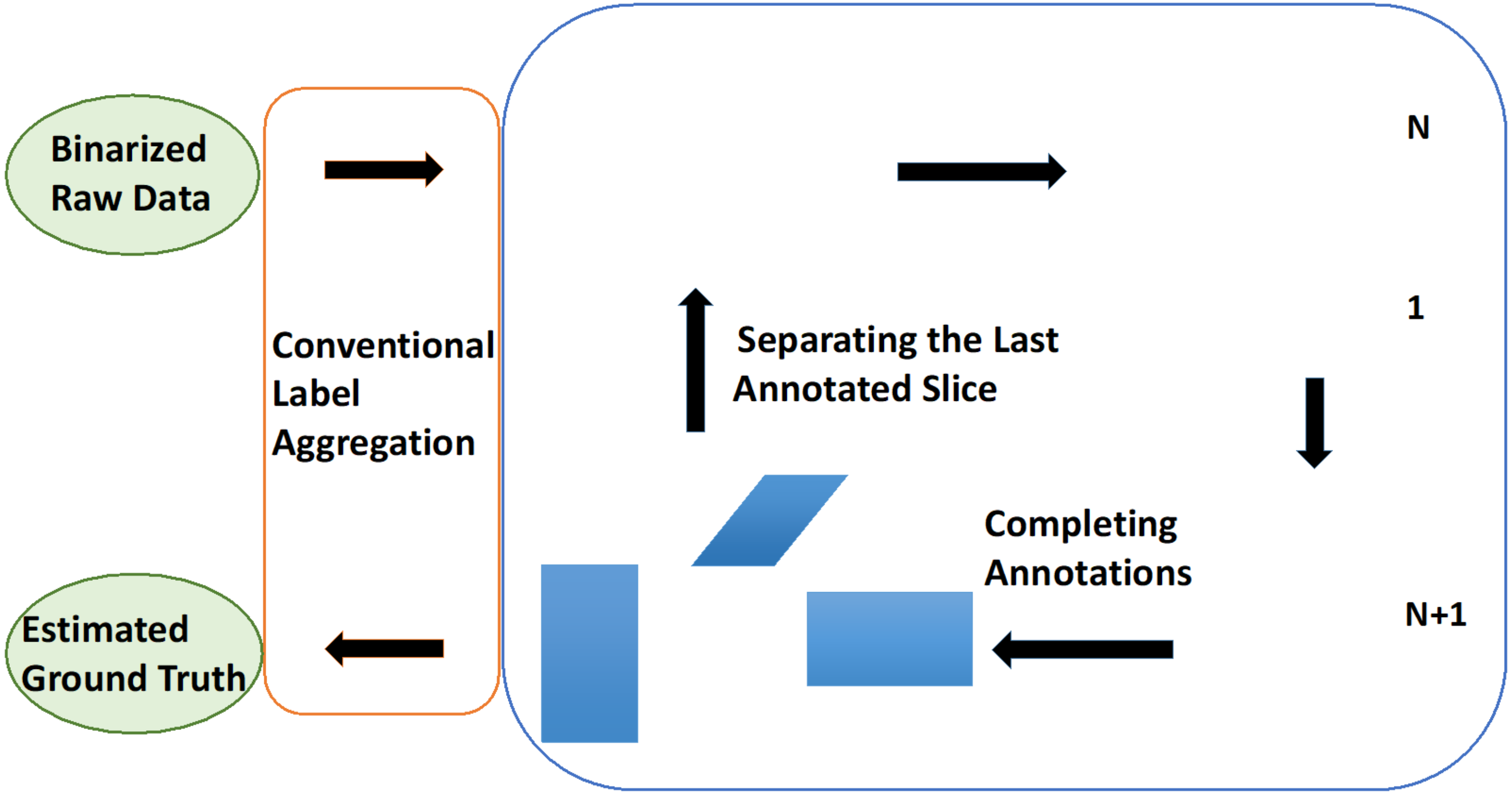
1) ***Novel*** – it is the first work that introduces tensor decomposition methods to exploit the structural information in the label tensor.

2) ***Versatile*** – it is a general framework for crowdsourcing that improves existing methods to achieve higher accuracy.

3) ***Powerful*** – the proposed MiSC algorithm is especially robust to annotation sparsity and noise compared with other benchmarking pure label aggregation approaches.

MiSC: Mixed strategies Crowdsourcing

- ★ **poster:** 15:00 – 16:00 @ 2073-2074
- ★ **arXiv:** https://arxiv.org/abs/1905.07394



Follow our project!