Fast and Robust Rank Aggregation

Robust Rank Aggregation against Model Misspecification

Yuangang Pan

November 19, 2019

Centre for Artificial Intelligence (CAI)
Faculty of Engineering and Information Technology
University of Technology Sydney, Australia

Table of contents

Background

Massive Open Online Course(MOOC)

Table 1: Course-Level Statistics of HarvardX and MITx courses on edX.

	CS[1]	STEM	HHRDE	GHSS
Courses	30	91	94	75
Participants	1,527,300	1,081,995	822,026	1,017,713
Participants per course	21,040	7,905	4,606	10,213

How to know you are the top-10?

Note that: the last teacher who tried to rank the whole participants himself was already dead!

[1] CS = Computer Science; STEM = Science, Technology, Engineering, Mathematics; HHRDE = Humanities, History, Religion, Design, Education; GHSS = Government, Health, Social Sciences

Rank aggregation arise everywhere

Assume you are a go player,

How to calculate your rank position?

Alpha Go only wins a few competitions, why does it rank first in the whole world.

The Olympic Games 2020 is coming,

The Chinese football team does not need to go to Tokyo, since people know they would definitely not be the Champion.

Do you think whether the decision has a bias, since the football teams from Brazil and Argentina even have not won us this year?

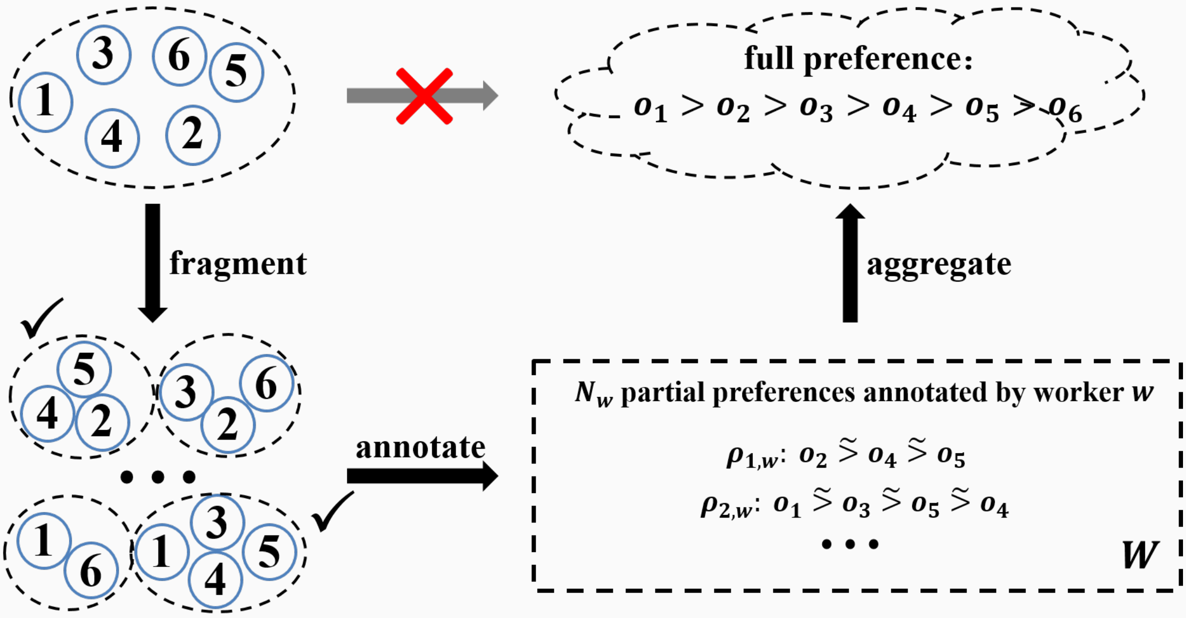
I want to find the smartest student in our school in one day.

any suggestions from you?

~~Assume~~ you submit your paper to a conference,

....

Rank Aggregation



Rank Aggregation:

- to reach a consensus full ranking by aggregating a large volume of ranking lists annotated by amateur workers

Model misspecification:

Homogeneity assumption:

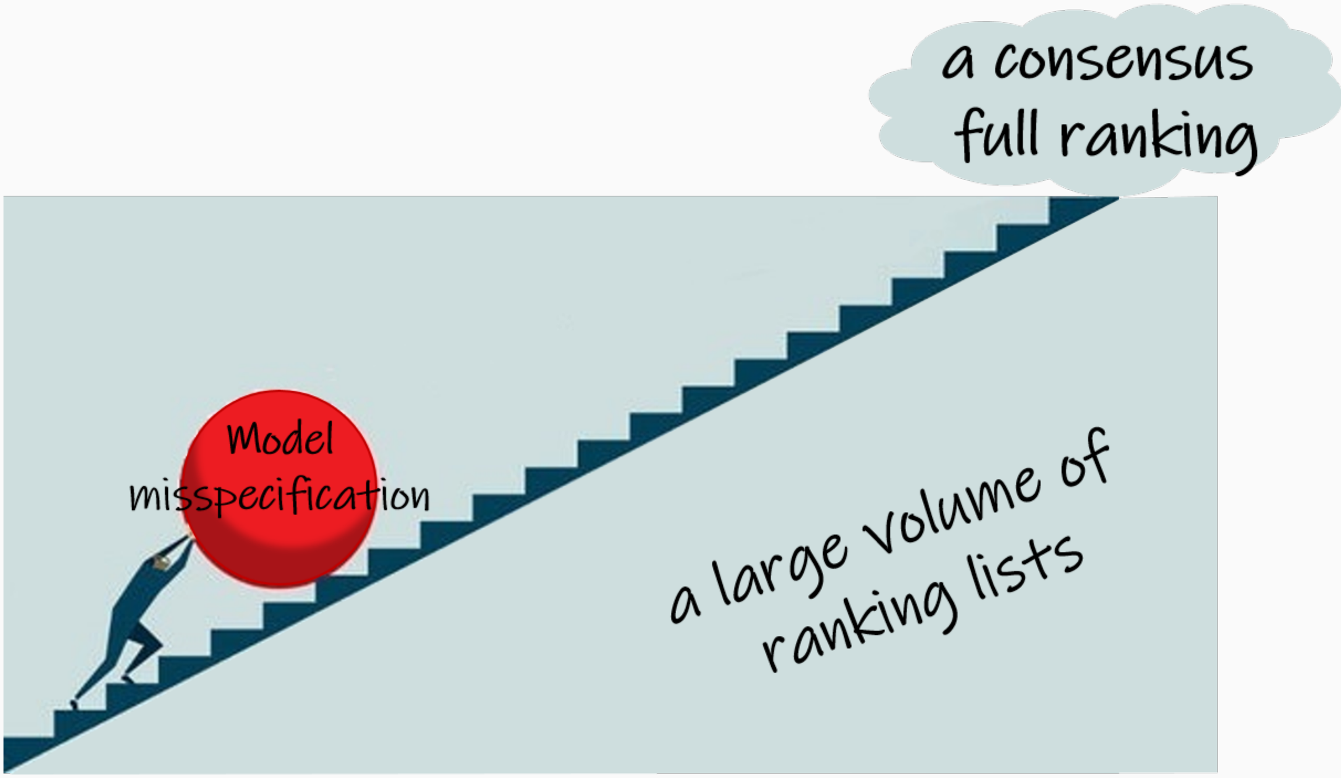
All preferences are provided by homogeneous users, sharing the same annotation accuracy and agreeing with the single ground truth ranking

A MLE for RA under model misspecification:

$$\max_{P_\theta \in \mathcal{P}, P_o \notin \mathcal{P}} P_\theta(\mathcal{R}_N), \quad \text{where} \quad \mathcal{R}_N = \{\rho_n | \rho_n \sim P_o, n = 1, 2, \ldots, N\}. \quad (1)$$

- $P_\theta \in \mathcal{P}, P_o \notin \mathcal{P}$

Challenges of RA:



Approach:

(1) Model misspecification, a.k.a. learn with a wrong model;
(2) A large volume of ranking lists.

My Solutions:

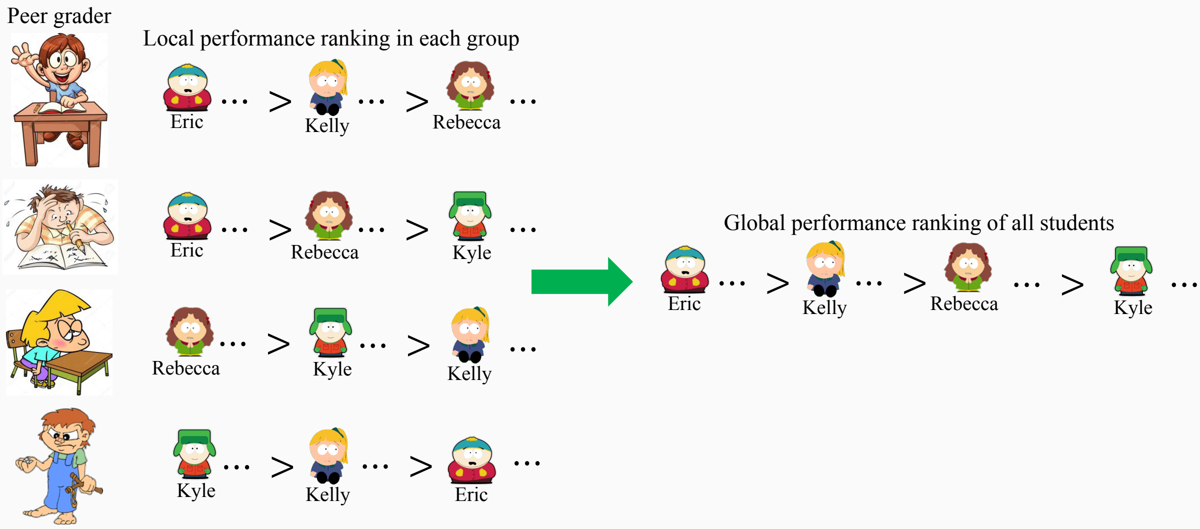
Law of total probability

Convolve the ranking model with specific perturbation mechanisms

$$P_\theta(\rho_n) = \int P(\rho_n, \varrho_n) d\varrho_n = \int P_\theta(\varrho_n) P(\rho_n|\varrho_n) d\varrho_n, \tag{2}$$

- $\rho_n \sim P_o, \ \varrho_n \sim P_\theta$.
- $P_\theta(\rho_n)$ is an invalid likelihood.
- $P(\rho_n|\varrho_n)$ denotes the pre-assumed perturbation mechanism.

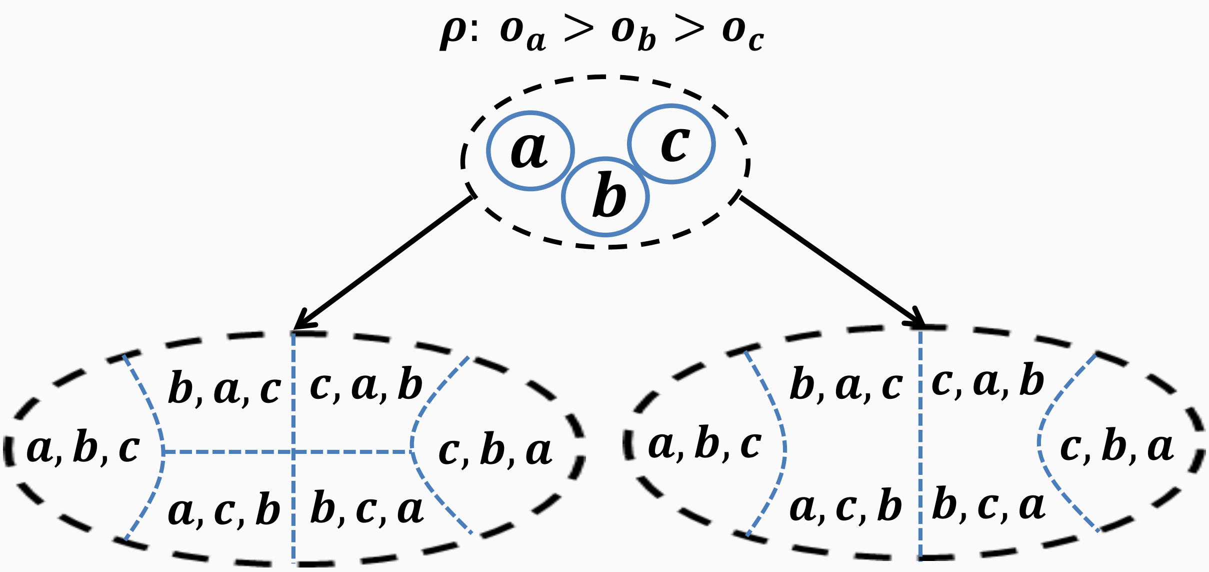
Divide and Conquer



Approach:

(1) divide the worker population into several subgroup;
(2) learn the error pattern for each type of worker.

Achievement 1:



Permutation-based Partition
($k!$ partitions)

Kendall-tau Partition
($C_k^2 + 1$ partitions)

- Bo Han*, Yuangang Pan*, Ivor W. Tsang: Robust Plackett-Luce model for k-ary crowdsourced preferences. Machine Learning: 675-702 (2018)

Achievement 2:

Crowdsourced Task $\boldsymbol{\xi} = \{\boldsymbol{O_1}, \boldsymbol{O_3}, \boldsymbol{O_4}, \boldsymbol{O_5}\}$

Annotation Process:

$$\textbf{Stage 1: max}(\boldsymbol{O_1}, \boldsymbol{O_3}, \boldsymbol{O_4}, \boldsymbol{O_5}) \longrightarrow \boldsymbol{O_3} \quad \boldsymbol{\rho}^{(1)}$$

$$\textbf{Stage 2: max}(\boldsymbol{O_1}, \boldsymbol{O_4}, \boldsymbol{O_5}) \longrightarrow \boldsymbol{O_4} \quad \boldsymbol{\rho}^{(2)}$$

$$\textbf{Stage 3: max}(\boldsymbol{O_1}, \boldsymbol{O_5}) \longrightarrow \boldsymbol{O_5} \quad \boldsymbol{\rho}^{(3)}$$

$$\boldsymbol{O_1} \quad \boldsymbol{\rho}^{(4)}$$

k-ary Preference $\quad \boldsymbol{\rho}\text{: } \boldsymbol{O_3} \gtrsim \boldsymbol{O_4} \gtrsim \boldsymbol{O_5} \gtrsim \boldsymbol{O_1}$

- Yuangang Pan, Bo Han, Ivor W. Tsang: Stagewise learning for noisy k-ary preferences. Machine Learning: 1333-1361 (2018)

Coarsening Bayesian

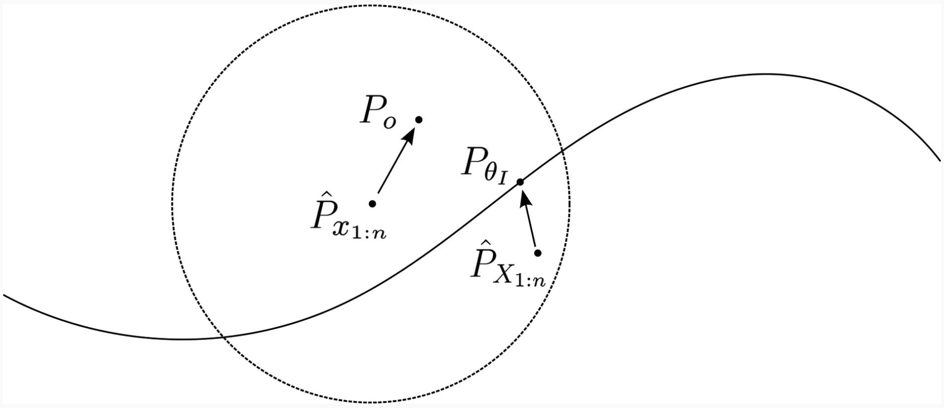


Figure 1: Coarsening Bayesian: inferring over the neighborhood of the observed data.

- $\tilde{P}_{x_{1:n}}$ is the observed data, which converges to the real data distribution P_o.
- $\tilde{P}_{X_{1:n}}$ is the idealized data, which converges to the idealized data distribution P_{θ_I}.
- The ambient space denotes the neighborhood of observed data.

Coarsened Rank Aggregation

performing RA over the neighborhood of the rank lists would enable the vanilla RA against noise agnostic perturbation, a.k.a., distributional robustness.

Let $B(\mathcal{R}_N, \varepsilon)$ denote the neighborhood of the ranking dataset $\mathcal{R}_N$ with size ε. Namely,

$$B(\mathcal{R}_N, \varepsilon) = \{\mathcal{R}'_N | D(\mathcal{R}'_N, \mathcal{R}_N) < \epsilon\}. \tag{3}$$

- $D(\cdot, \cdot)$: distribution-level distance
- $D(\cdot, \cdot)$: divergence of their empirical data distributions
- Kullback-Leibler (KL) divergence, f-divergence and Wasserstein metric

Coarsened Rank Aggregation (CoarsenRank)

CoarsenRank: inferring over the neighborhood of the $\mathcal{R}_N$

$$\max_{\theta \in \Theta} P_\theta(\Re_N), \qquad \text{where} \quad \Re_N \in B(\mathcal{R}_N, \varepsilon). \tag{4}$$

where Θ denotes the parameter space.

Two equivalent (but compact) formulations:

$$\max_{\theta \in \Theta} \mathbb{P}(\theta | B(\mathcal{R}_N, \varepsilon)) \quad \Longleftrightarrow \quad \max_{\theta \in \Theta} \mathbb{P}(\theta | D(\mathcal{R}_N, \Re_N) < \epsilon). \tag{5}$$

CoarsenRank

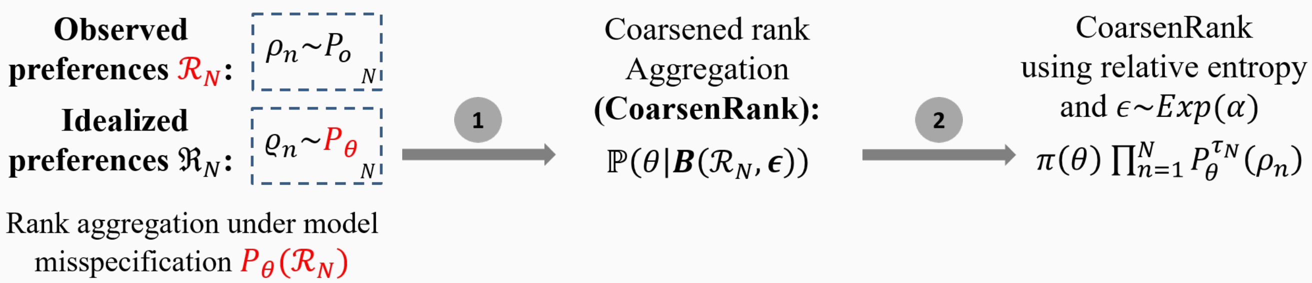


Figure 2: The logic stream of our CoarsenRank. Condition 1: perform rank aggregation over a neighborhood of the collected preferences. Condition 2: adopt relative entropy as the divergence measure and assign an exponential prior for the size of the neighborhood.

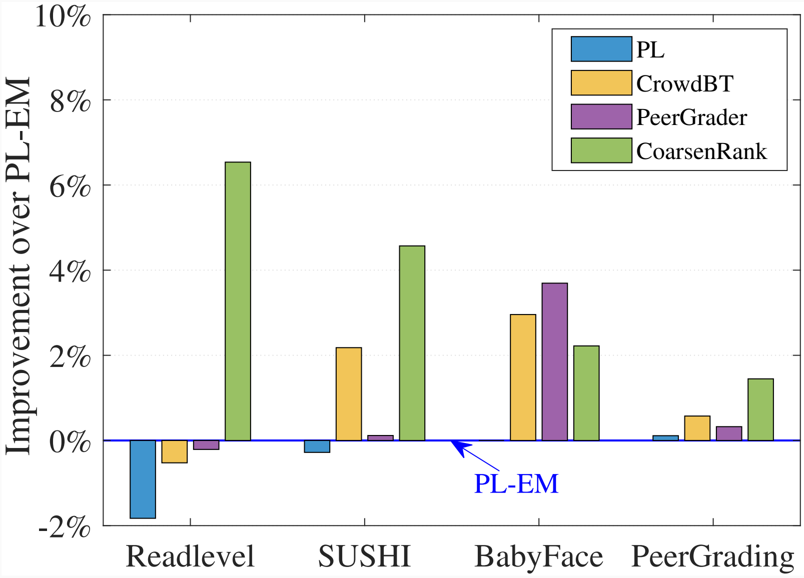
- Yuangang Pan, Weijie Chen, Gang Niu, Ivor W. Tsang, Masashi Sugiyama: Fast and Robust Rank Aggregation against Model Misspecification. Submitted

Three instances of CoarsenRank

$$
\begin{aligned}
\Pi(\theta)\prod_{n=1}^{N} P_{\theta}^{\tau_N}(\rho_n) &\overset{i}{=} \frac{\Pi(\theta)}{(\sqrt{2\pi})^{N\tau_N}}\left[\prod_{n=1}^{N}\int_{-\infty}^{\frac{\triangle\theta_{\rho_n}}{\sqrt{2}}} \exp\left(-\frac{t^2}{2}\right) dt\right]^{\tau_N}, \\
&\overset{ii}{=} \Pi(\theta)\prod_{n=1}^{N}\left[\frac{\theta_{\rho_n^1}}{\theta_{\rho_n^1}+\theta_{\rho_n^2}}\right]^{\tau_N}, \\
&\overset{iii}{=} \Pi(\theta)\prod_{n=1}^{N}\left[\prod_{i=1}^{k-1}\frac{\theta_{\rho_n^i}}{\theta_{\rho_n^i}+\theta_{\rho_n^{i+1}}+\cdots+\theta_{\rho_n^k}}\right]^{\tau_N},
\end{aligned}
$$

- $\tau_N = \frac{1/N}{1/N+1/\alpha}$
- *i* denotes Coarsened Thurstone model.
- *ii* denotes Coarsened Bradley-Terry model.
- *iii* denotes Coarsened Plackett-Luce model

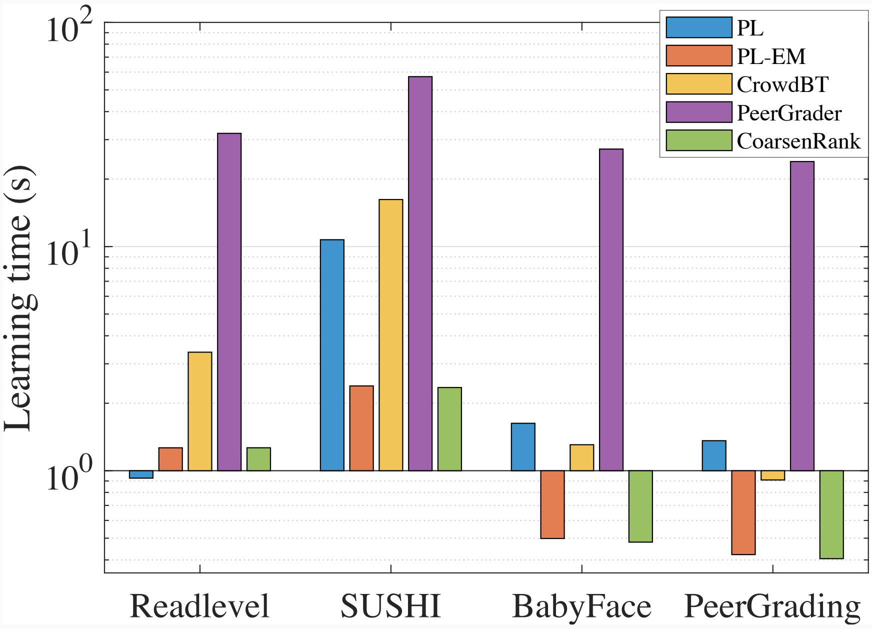
Performance improvement of various methods over PL-EM



Performance

Figure 3: Performance improvement of various methods over PL-EM on four datasets, following $\frac{\tau_* - \tau_0}{\tau_0}$. τ_0 is the accuracy of PL-EM in the Kendall tau distance.

The computational cost compassion of all methods

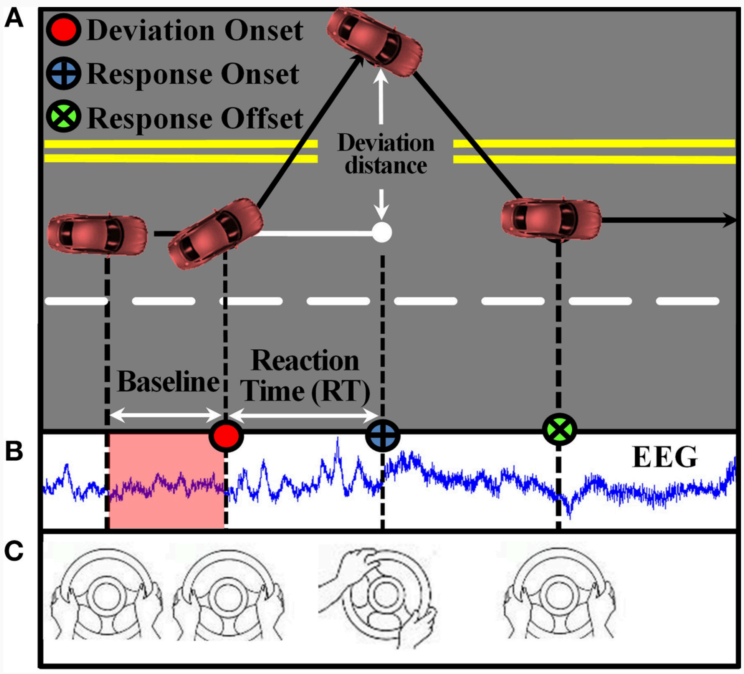


computational cost

Figure 4: The computational cost of all baselines on four datasets, respectively.

Robust RA in NeuroScience

Robust RA in NeuroScience



Mental fatigue monitoring:

Predict the reaction time (RT) to some emergency by aggregating the EEG signal from multiple heterogeneous EEG channels.

RA for mental fatigue monitoring

Revisit mental fatigue monitoring

- Reaction time (RT) $\Longleftrightarrow$ item score
- EEG channel $\Longleftrightarrow$ crowd worker

The mental fatigue monitoring task could then be formulated as RA under model misspecification while involving the EEG signal as the features.

Multi-Channel Regression

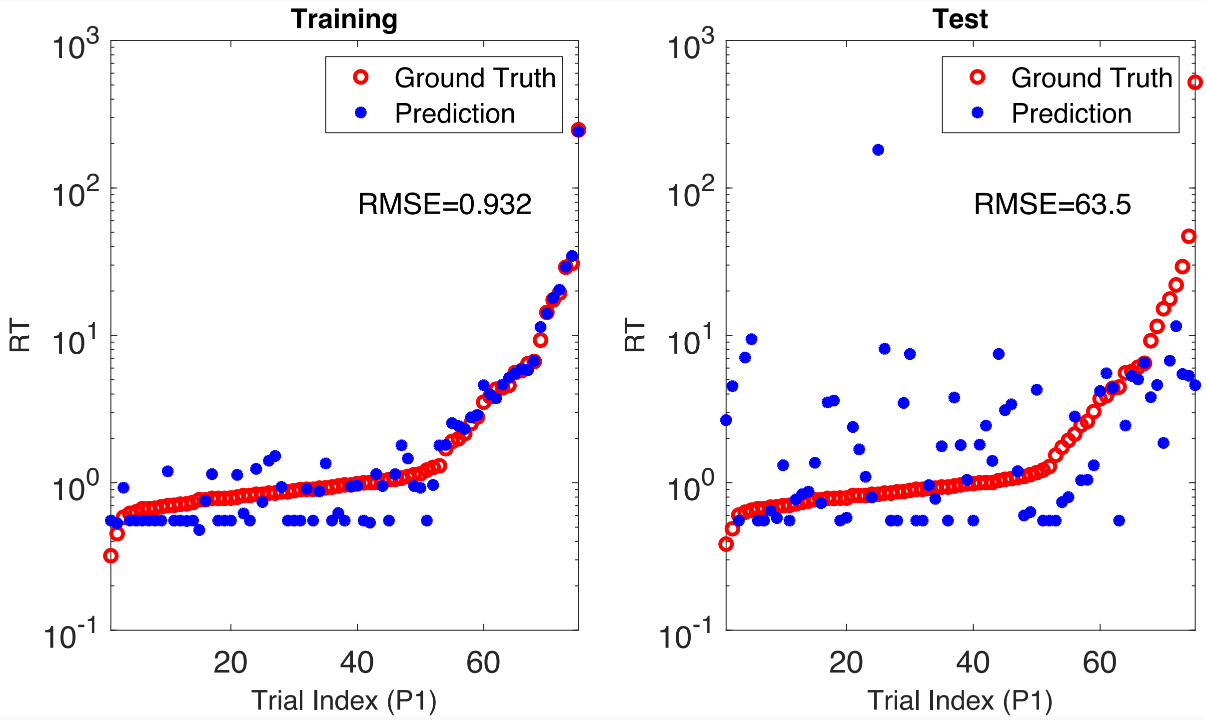


Figure 5: Overfitting of the two-layer deep regression model for mental fatigue monitoring. EEG signals from multiple channels are simply concatenated.

Multi-Channel Classification

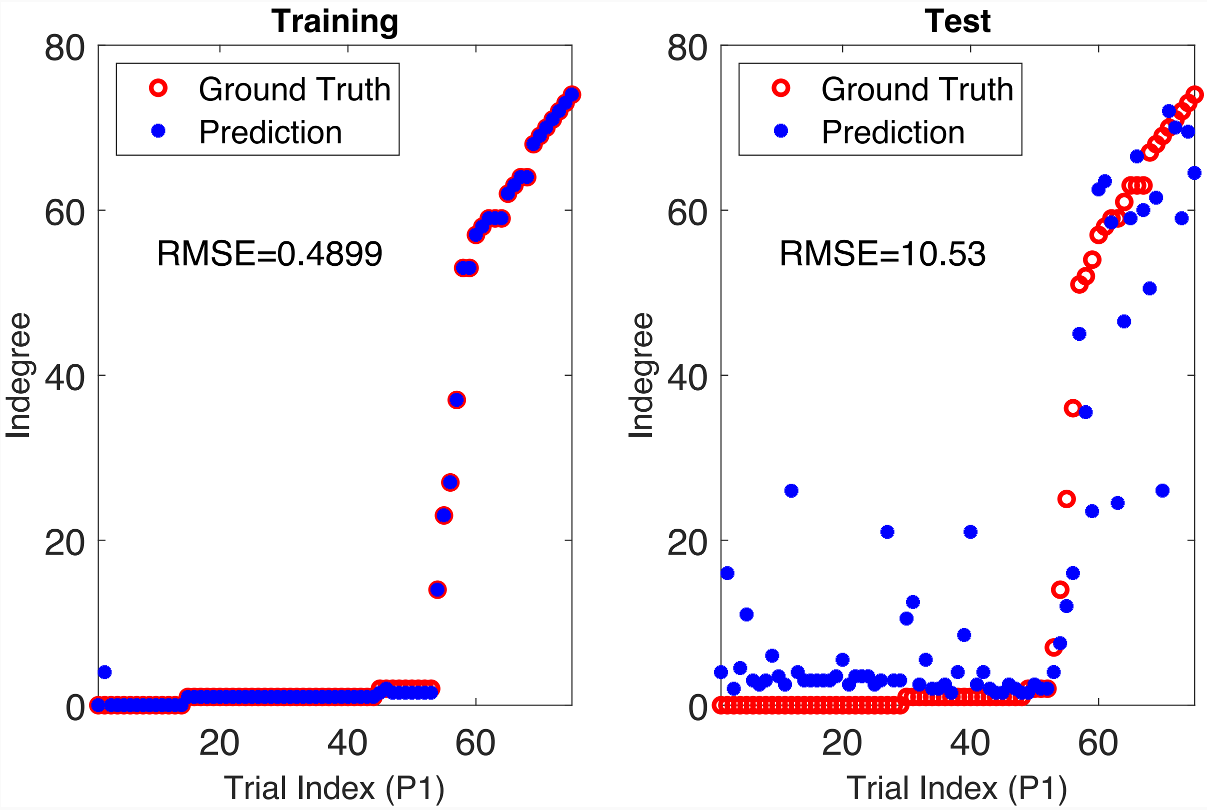
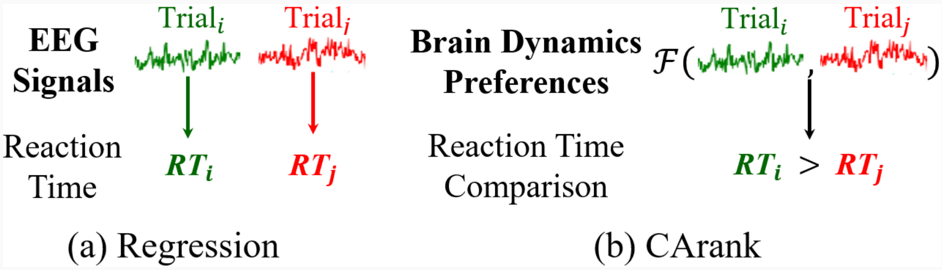


Figure 6: Consistency of the two-layer ordinal regression model using brain dynamics preferences. EEG signals from multiple channels are simply concatenated.

Achievement 1: Channel-reliability Aware Ranking



$$\Pi_n = P(\rho|\rho^{(n)}) = \begin{bmatrix} \pi_n & 0 & (1-\pi_n) \\ 0 & 1 & 0 \\ (1-\pi_n) & 0 & \pi_n \end{bmatrix}, \tag{6}$$

Achievements:

- **Yuangang Pan**, Ivor W. Tsang, Avinash K Singh, Chin-teng Lin, Masashi Sugiyama: Stochastic Multi-Channel Ranking with Brain Dynamics Preferences. Submitted

Ranking Performance

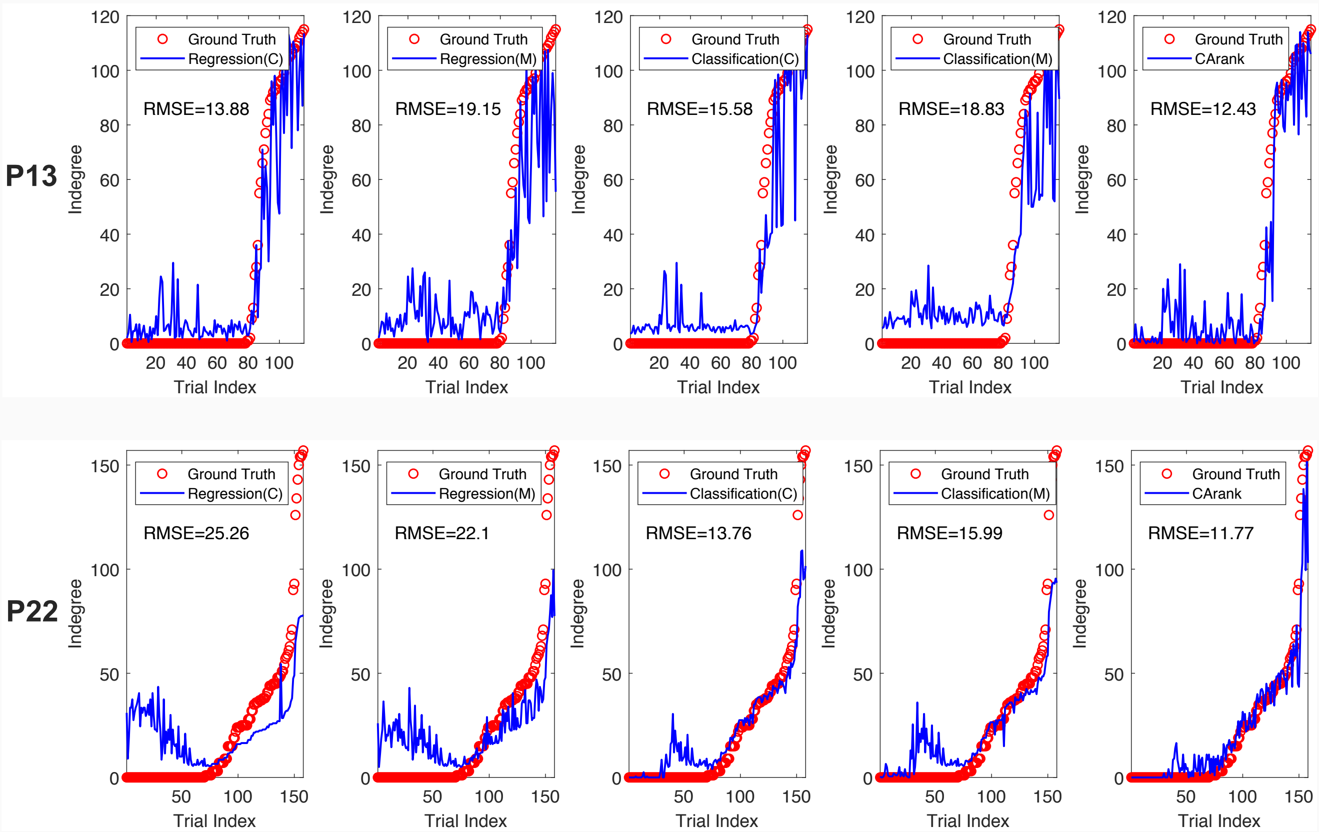


Figure 7: Indegree sequence for CArank and other baselines (closer is better).

Channel-reliability Estimation

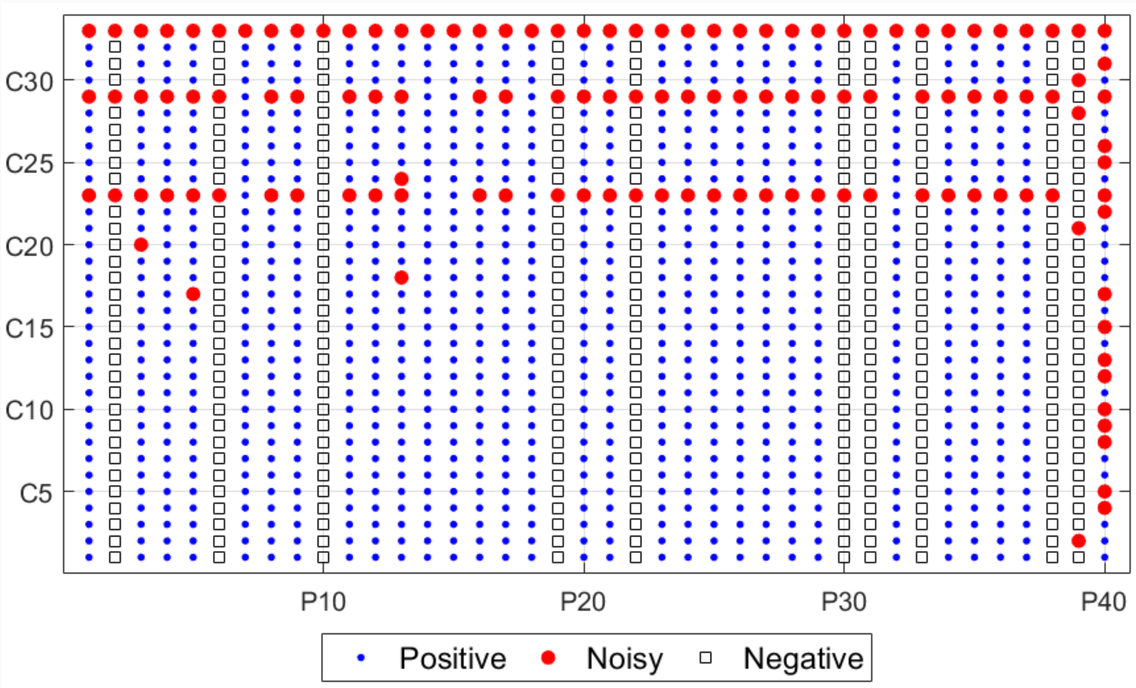


Figure 8: Reliability of different channels for forty participants estimated by CArank. Each column denotes the states of 33 channels for each participant. The channels with estimated reliability $0.15 \le \pi_n \le 0.85$ are considered as noisy channels marked in red.

Achievement 2: Self-Weighted Ordinal Regression Model

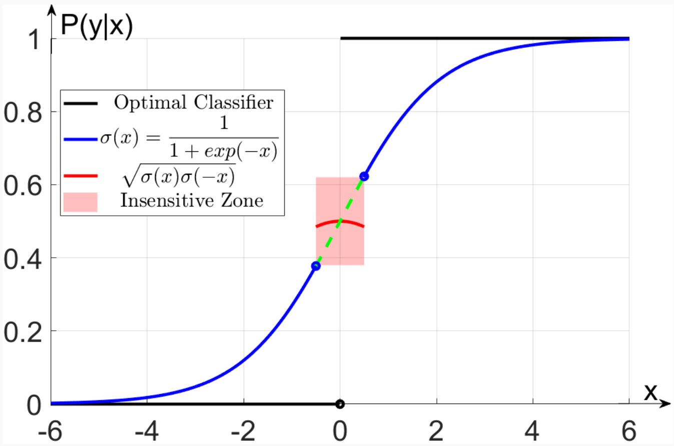


Figure 9: Gradient flattening w.r.t. sigmoid function.

$$P(y|w, \pi_{1:N}, x_0^{1:N}, x_1^{1:N}) = \begin{cases} \prod_{n=1}^{N} \left[\pi_n \sigma(w^T \Delta x_n) + (1-\pi_n)\sigma(-w^T \Delta x_n)\right] & y \in \mathcal{Y}_1, \\ \prod_{n=1}^{N} \sqrt{\sigma(w^T \Delta x_n)\sigma(-w^T \Delta x_n)} & y \in \mathcal{Y}_2. \end{cases}$$

- **Yuangang Pan**, Ivor W. Tsang, Yueming Lyu, Avinash K Singh, Chin-teng Lin: Online Brain Dynamics Ranking with Real-time Monitoring. Submitted

Achievement 2: Self-Weighted Ordinal Regression Model

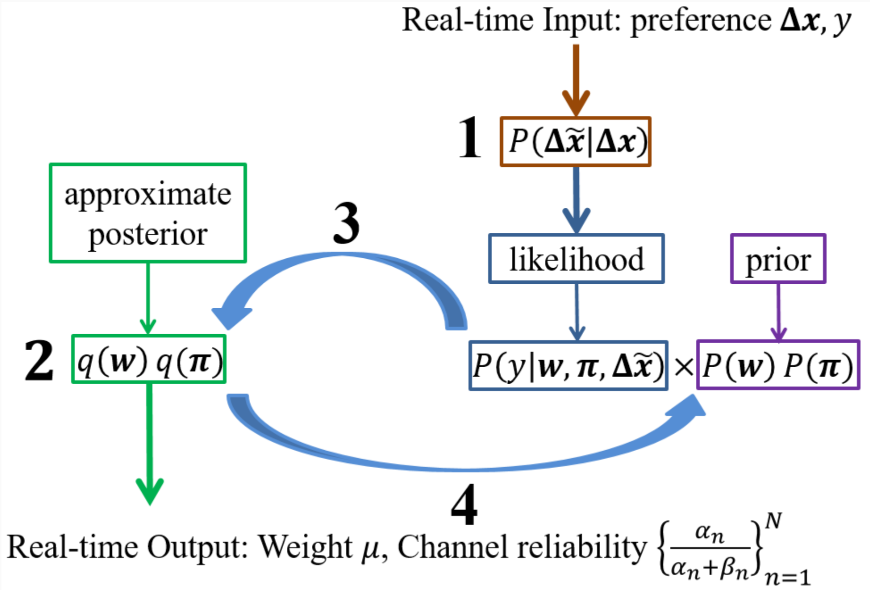


Figure 10: OGMM with Data Augmentation. (1) sample the corrupted EEG signal $\Delta\tilde{x}$ from the predefined corrupting distribution $P(\Delta\tilde{x}|\Delta x)$; (2) define $q(w)q(\pi)$ in the same form as the prior (product of a Normal with Betas); (3) estimate $q(w)q(\pi)$ with generalized Bayesian moment matching; (4) replace prior $P(w)P(\pi)$ with approximate posterior $q(w)q(\pi)$.

Online Evaluation

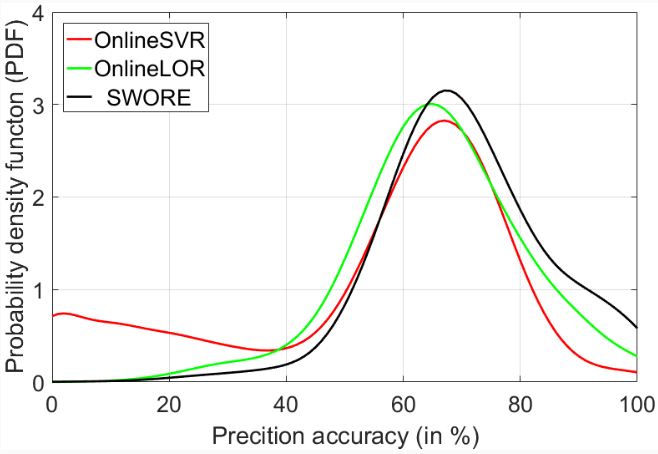


Figure 11: The PDF of prediction accuracy.

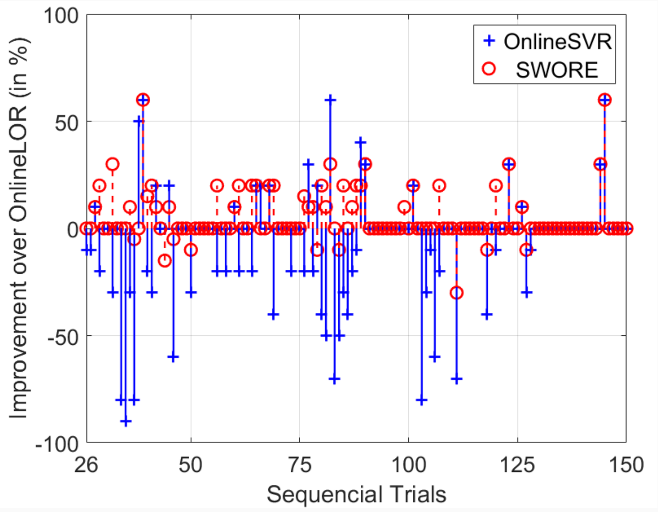


Figure 12: Real-Time Mental Fatigue Evaluation.

Online Evaluation

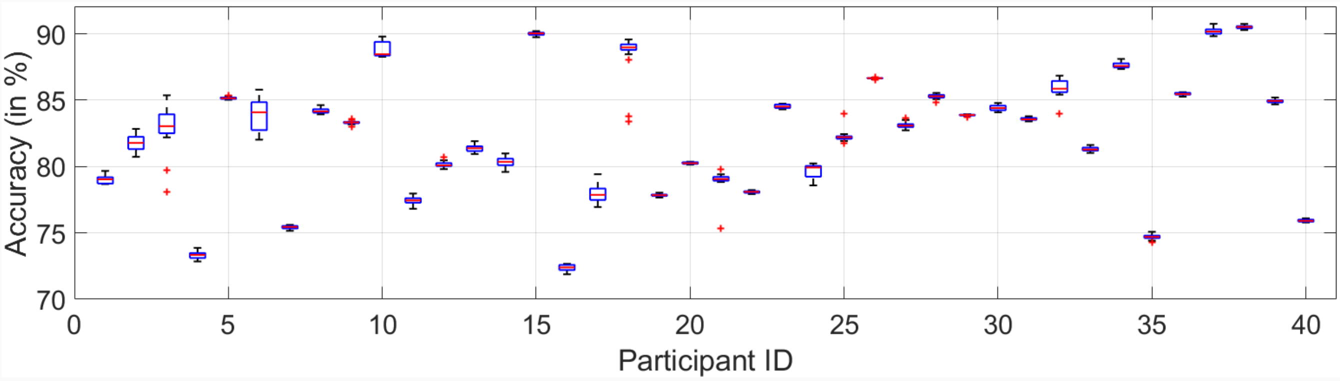


Figure 13: Box plot of the prediction accuracy on the test dataset. The symbol "+" denotes the outliers.

Other Topics:

- Bo Han, Quanming Yao, **Yuangang Pan**, Ivor W. Tsang, Xiaokui Xiao, Qiang Yang, Masashi Sugiyama: Millionaire: a hint-guided approach for crowdsourcing. Machine Learning (2019) Accepted
- Yaxin Shi, Donna Xu, **Yuangang Pan**, Ivor W. Tsang, Shirui Pan: Label Embedding with Partial Heterogeneous Contexts. AAAI (2019) Accepted
- Yinghua Yao, **Yuangang Pan**, Ivor W. Tsang, Xin Yao: Support Matching: A Novel Regularization to Escape from Mode Collapse in GANs. ICONIP (2019) Accepted
- Yaxin Shi, **Yuangang Pan**, Donna Xu, Ivor W. Tsang: Probabilistic CCA with Implicit Distributions. Submitted
- Huiting Hong, Xin Li, **Yuangang Pan**, Ivor W. Tsang: Domain-adversarial Network Alignment. Submitted
- **Yuangang Pan**, Weijie Chen, Ivor W. Tsang, Masashi Sugiyama: Exploring Disentangled Features among Multiple Responses. Submitted

Q&A

Thank you!