Quantum Curriculum Learning

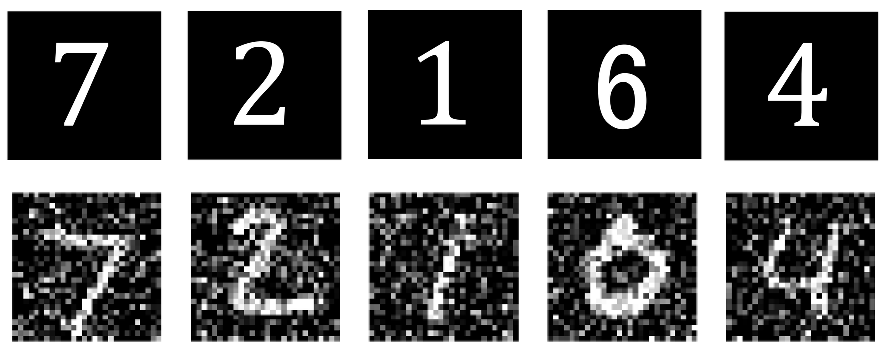
Quoc Hoan Tran, Yasuhiro Endo, Hirotaka Oshima

Quantum Laboratory, Fujitsu Research, Fujitsu Limited

[1] Q. H. Tran et al., **arXiv**:2407.02419 (2024)

Curriculum Learning (CurL) for quantum data

Mimic human intelligence in learning concepts



Easy task and data

(Y. Bengio et al., ICML 2019)

Difficult task and data

Aim to increase the generalization and reliability of the prediction

We propose two methods of CurL with quantum data (Q-CurL)

✓ **Task-based Q-CurL**: use for training main task with limited resource in but having auxiliary tasks with high training resource

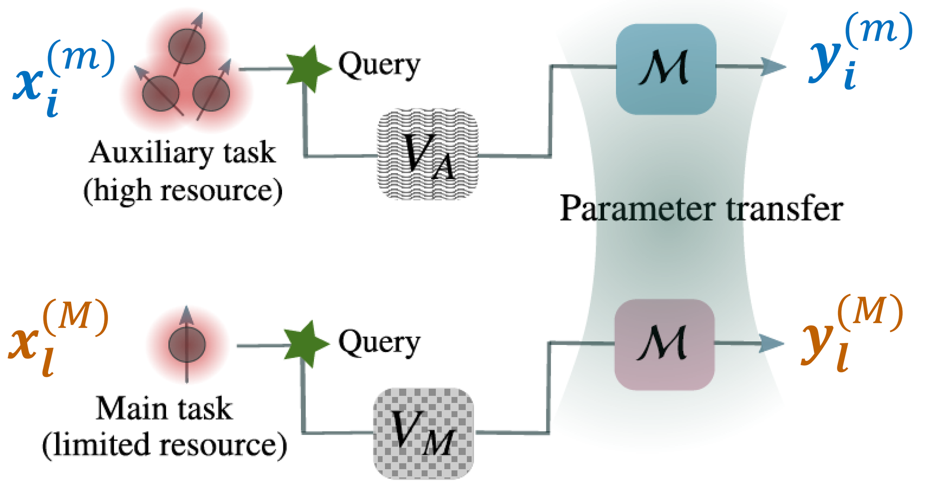
⟹ **Derive a criteria to select the most beneficial auxiliary task**

✓ **Data-based Q-CurL**: use for training with noisy data or corrupted labelling

⟹ **Introduce the data dynamics to the loss function**

Task-based Q-CurL

Given a main task $\mathcal{T}_M$ and a set of auxiliary tasks $\mathcal{T}_m$, task-based **Q-CurL assigns a curriculum weight** $c(\mathcal{T}_M, \mathcal{T}_m)$ to select the most beneficial auxiliary task



Assume that all tasks share the same hypothesis h, **the expected risk** of the main task is transformed as

$$R_{\mathcal{T}_M}(h) = \mathbb{E}_{(x,y)\sim P^{(M)}}[\ell(h(x),y)] = \mathbb{E}_{(x,y)\sim P^{(m)}}\left[\frac{p^{(M)}(x,y)}{p^{(m)}(x,y)}\ell(h(x),y)\right]$$

Ratio of data density $r(x,y) = \frac{p^{(M)}(x,y)}{p^{(m)}(x,y)}$ evaluates the contribution of data in $\mathcal{T}_m$ to $\mathcal{T}_M$. Approximate by linear model:

$$r(x,y) \approx r_\alpha(x,y) = \alpha^T \phi(x,y) = \sum_{l=1}^{N_M} \alpha_l \phi_l(x,y)$$

The **fidelity-based kernel** $\phi_l(x,y) = \mathrm{Tr}\left[x x_l^{(M)}\right] \mathrm{Tr}\left[y y_l^{(M)}\right]$

Find α via the minimization of $\frac{1}{2}\iint [\hat{r}_\alpha(x,y) - r(x,y)]^2 p^{(m)}(x,y) dx dy$

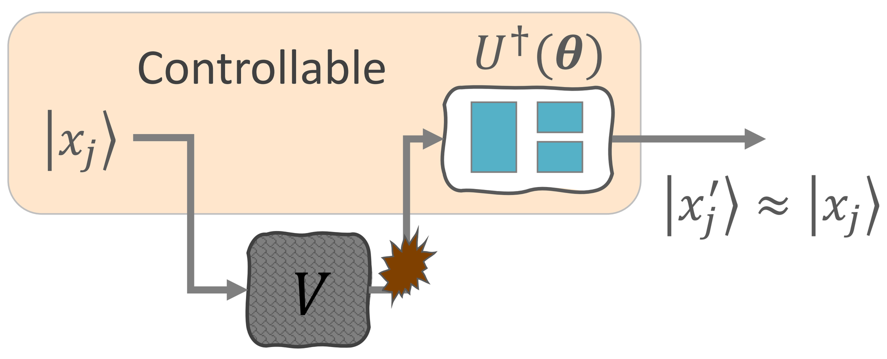
The curriculum weight (higher weight means better curriculum)

$$c(\mathcal{T}_M, \mathcal{T}_m) = \frac{1}{N_m}\sum_{i=1}^{N_m} \hat{r}_\alpha\left(x_i^{(m)}, y_i^{(m)}\right)$$

Application: Learning Unitary Dynamics

Task: learning an unknown unitary V dynamics with N query

$$\mathcal{D}_Q(N) = \{(|x_j\rangle, |y_j\rangle)\}_{j=1}^N$$



Find a decoder $U^\dagger(\theta)$ to minimize

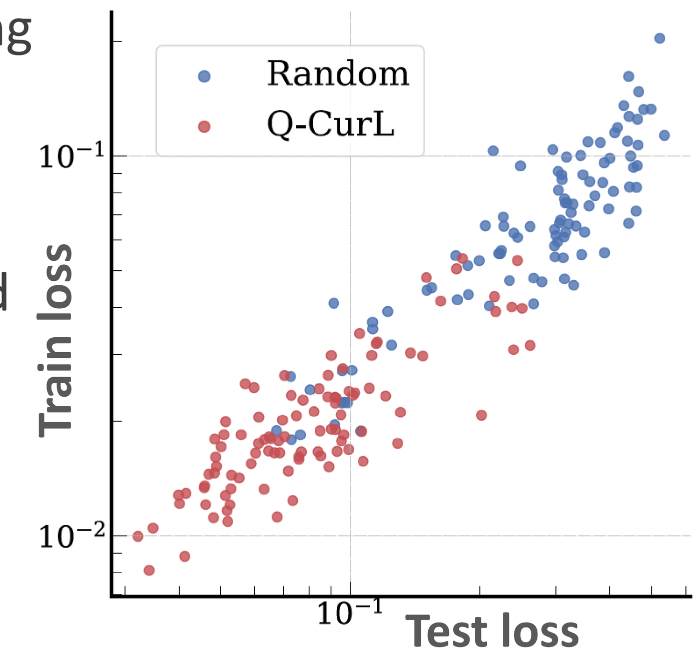
$$C_{\mathcal{D}_Q(N)}(\theta) = 1 - \frac{1}{N}\sum_{j=1}^{N} \left|\langle y_j | U(\theta) | x_j\rangle\right|^2$$

Enhance the accuracy in learning dynamics of the spin-1/2 XY model (see [1] for more details)

Q-CurL game: Find a proper order of solving all auxiliary tasks $\mathcal{T}_1, \dots, \mathcal{T}_{M-1}$ to obtain the best performance of $\mathcal{T}_M$ $(M = 20)$

Task $\mathcal{T}_m$: m variational layers for Trotterized approximation of $V = \exp(-iH_{XY}t)$

$$H_{XY} = \sum_k h_k Z_k + \sum_k (X_k X_{k+1} + Y_k Y_{k+1})$$



Greedy algorithm with curriculum weight

$$\mathcal{T}_{i_M = M} \leftarrow \mathcal{T}_{i_{M-1}} \leftarrow \mathcal{T}_{i_{M-2}} \leftarrow \cdots \leftarrow \mathcal{T}_{i_1}$$

$$\mathcal{T}_{i_k} \leftarrow \mathcal{T}_{i_{k-1}} \quad s.t.\ i_{k-1} = \underset{j \in 1, M/\{i_M, \dots, i_k\}}{\operatorname{argmax}} c(\mathcal{T}_{i_k}, \mathcal{T}_j)$$

Data-based Q-CurL

Data-based Q-CurL enhances the robustness by dynamically weighting data difficulty

Single loss $\ell_i(\theta) = \ell(h(x_i), y_i)$

Conventional global loss $\mathcal{L}^c(\theta) = \frac{1}{N}\sum_{i=1}^N \ell_i(\theta)$

Dynamical loss function

$$\mathcal{L}^d(\theta, w, \lambda) = \frac{1}{N}\sum_{i=1}^N [\ell_i(\theta) - \eta]\, w_i + \lambda Reg(w_i)$$

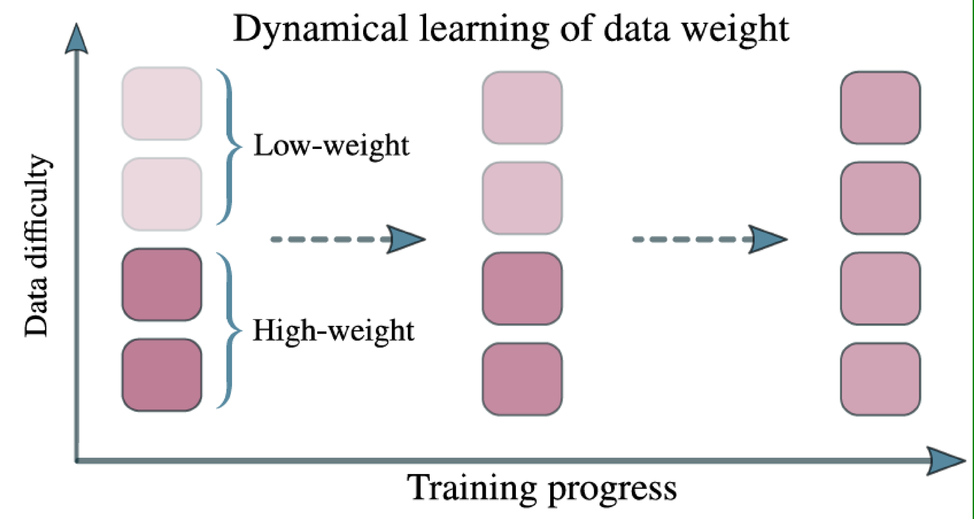
No increasing quantum resource $\min_\theta \min_w \mathcal{L}^d(\theta, w, \lambda)$

Only classical resources

✓ $\ell_i(\theta) \ll \eta \rightarrow w_i$ is big

✓ $\ell_i(\theta) \gg \eta \rightarrow w_i$ is small

✓ $\eta = \mathcal{L}^c(\theta)$ conventional loss from the previous training epoch



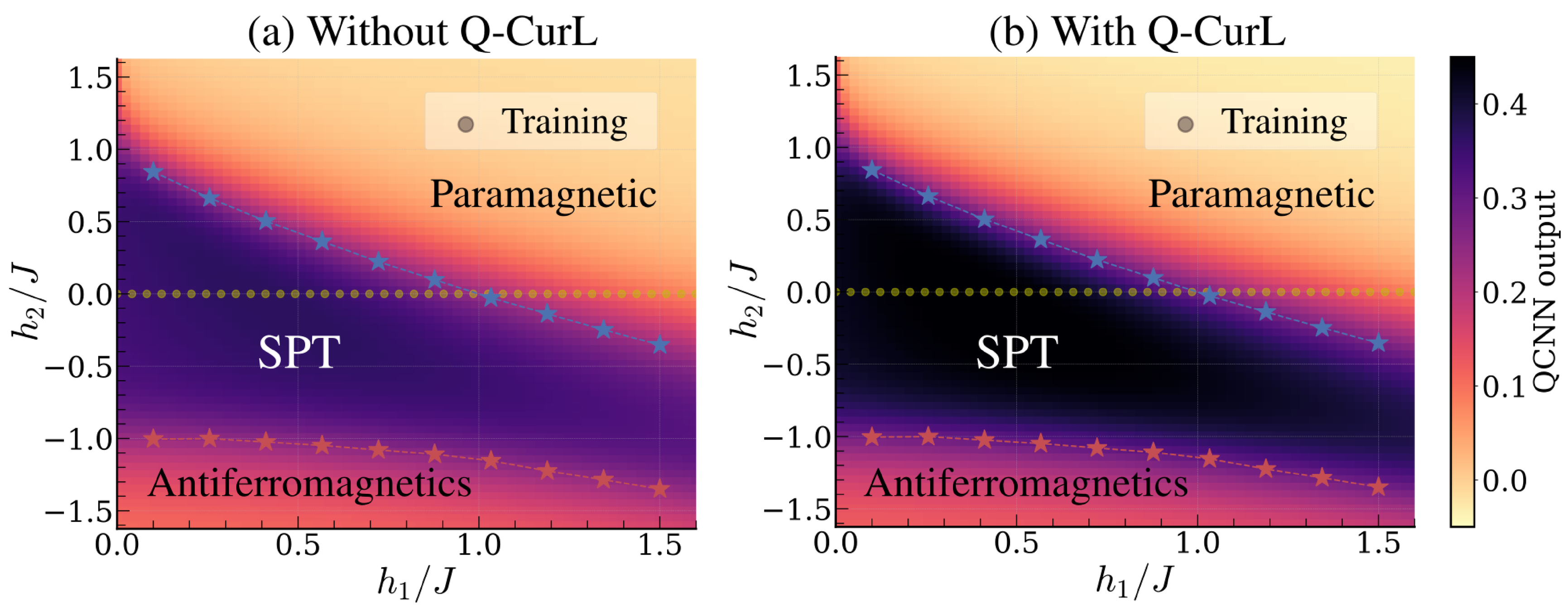
Application: Learning Noisy-labelled Quantum Phase

Task: given ground state data with noisy labeling, train a quantum model that demonstrates robustness and generalizes effectively to unseen data.

Ground states of the one-dimensional cluster Ising model exhibits different phases (SPT, antiferromagnetics, and paramagnetic)

$$H = -J\sum_k Z_k X_{k+1} Z_{k+2} - h_1 \sum_k X_k - h_2 \sum_k X_k X_{k+1}$$

Enhance the robustness of phase detection (see [1] for more details)



QCNN (I. Cong et al., **Nat. Phys.** 15, 1273, 2019) architecture to train the quantum data with label = 1 for SPT and 0 for other phases

Corrupted labelling: flip each training label with a probability p $(= 0.3$ in the above diagrams)