Iterative Classroom Teaching

Louis Faucon
louis.faucon@epfl.ch

Laboratory for Information and Inference Systems (LIONS)
École Polytechnique Fédérale de Lausanne (EPFL)
Switzerland

Joint work with
Teresa Yeo, Parameswaran Kamalaruban, Adish Singla, Arpit Merchant, Thibault Asselborn, Pierre Dillenbourg, and Volkan Cevher

lions@epfl

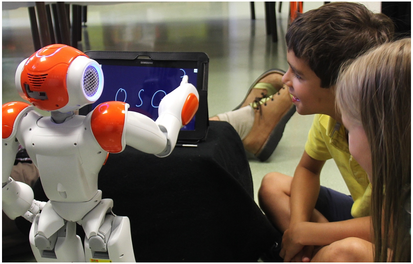
MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

FNSNF
Fonds national suisse
Schweizerischer Nationalfonds
Fondo nazionale svizzero
Swiss National Science Foundation

EPFL
École Polytechnique Fédérale de Lausanne

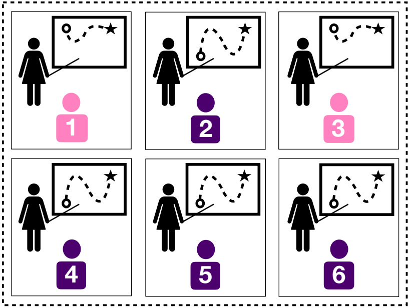
Motivation

Teaching a class of children how to write:



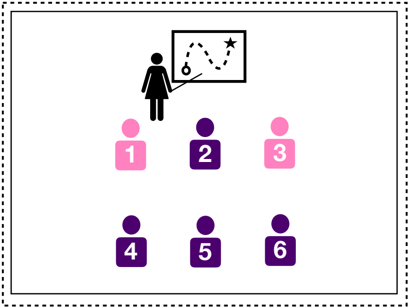
Tension Between Orchestration Cost and Students' Workload

Individual Teaching (IT)



Orchestration cost ↑
Students' workload ↓

Classroom Teaching (CT)

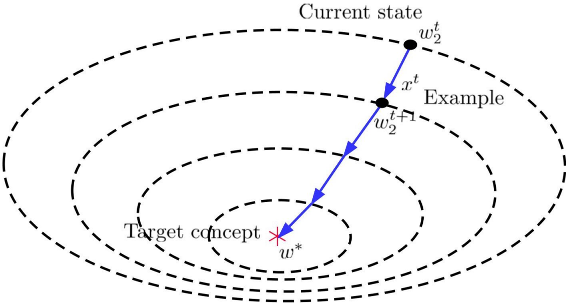


Orchestration cost ↓
Students' workload ↑

Iterative Machine Teaching[1]

Each student is characterized by:

1. Prior knowledge (initial state)
2. Learning ability



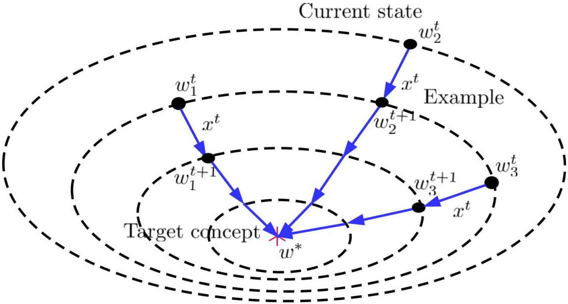
$$w^{t+1} = \texttt{Proj}_{\mathcal{W}}\left(w^t - \eta^t \frac{\partial \ell\left(\pi(w^t, x^t), y^t\right)}{\partial w^t}\right); \quad (x^t, y^t) \in \mathcal{X} \times \mathcal{Y}$$

[1] Liu, W. et al. Iterative machine teaching. ICML 2017

Stylized Model of Classroom

Diversity of the classroom depends on:

1. Prior knowledge (initial state)
2. Learning ability



$$w_j^{t+1} = \texttt{Proj}_{\mathcal{W}} \left(w_j^t - \eta_j^t \frac{\partial \ell \left(\pi(w_j^t, x^t), y^t \right)}{\partial w_j^t} \right); \quad j \in [N], (x^t, y^t) \in \mathcal{X} \times \mathcal{Y}$$

Problem

Teaching objective

The class on average converges.

$$\frac{1}{N}\sum_{i=1}^{N}\left\|w_j^T - w^*\right\|^2 \leq \epsilon \quad \text{and} \quad \left\|w_j^T - w^*\right\| \leq \epsilon, \forall j \in [N]$$

Teaching protocol

Pick examples that minimizes average distance between students' internal states and target.

$$\begin{aligned} x^t &= \arg\min_{x \in \mathcal{X}} \frac{1}{N}\sum_{j=1}^{N}\left\|w_j^t - \eta_j \frac{\partial \ell\left(\left\langle w_j^t, x\right\rangle, \langle w^*, x\rangle\right)}{\partial w_j^t} - w^*\right\|^2 \\ y^t &= \left\langle w^*, x^t\right\rangle \end{aligned}$$

Results

Reduced the orchestration cost for CT

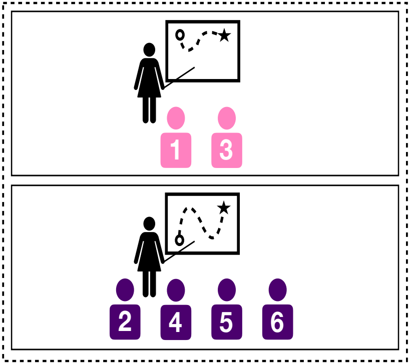
- Classroom of 1000 students, *i.e.*, $N = 1000$
- Each student is represented by a vector of size 100, *i.e.*, $d = 100$
- Orchestration cost for IT: $\mathcal{O}\left(N \log \frac{1}{\epsilon}\right)$
- Orchestration cost for CT: $\mathcal{O}\left(\min\{N, d\} \log \frac{1}{\epsilon}\right)$

Orchestration cost: learning ability vs prior knowledge

Finer partitions for slower learners:

$$\{[\eta_{\text{min}}, 2\eta_{\text{min}}), [2\eta_{\text{min}}, 4\eta_{\text{min}}), \ldots, [2^m\eta_{\text{min}}, 2\eta_{\text{max}})\}\text{, where } m = \left\lfloor \log_2 \frac{\eta_{\text{max}}}{\eta_{\text{min}}} \right\rfloor$$

Classroom Teaching with Partitioning (CTwP)



k Number of groups

$T(k)$ Orchestration cost

$S(k)$ Students' average cost

λ Tradeoff parameter

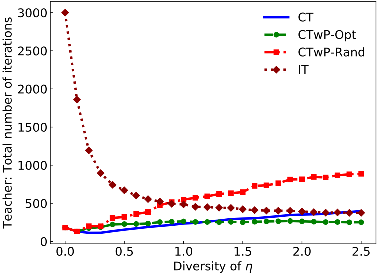
$$\boxed{\text{cost}(k) = T(k) + \lambda S(k)}$$

Homogenous partitioning based on:

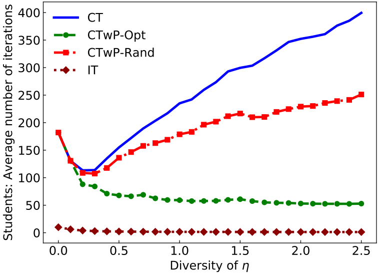
1. Learning ability: Finer partitions for slower learners.
2. Prior knowledge: Via pre-quiz or by demographics.

Experiments - Simulations

Diversity impacts the number of iterations (both students' and teacher's)



(a) Convergence from teacher's perspective



(b) Convergence from students' perspective

Legend:

CT: Classroom teaching

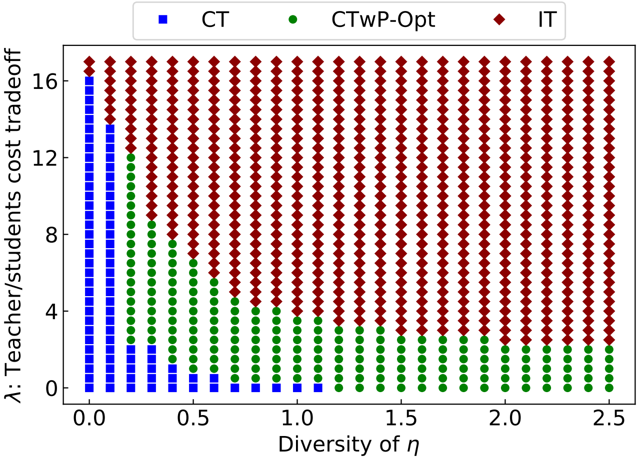
CTwP-Opt: CT with optimal partitioning,

CTwP-Rand: CT with random partitioning,

IT: Inidivdual teaching

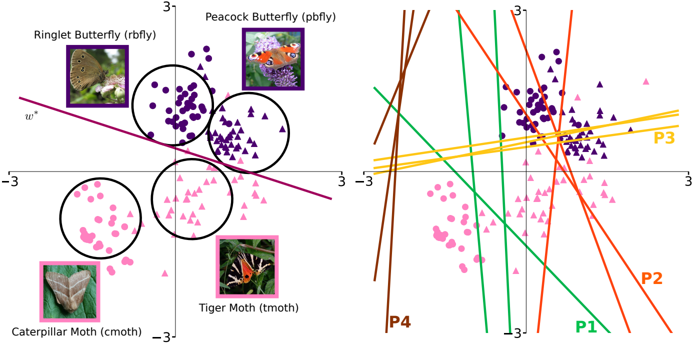
Experiments - Simulations

Tradeoff between teacher's and students' workload.



Experiments - Real World

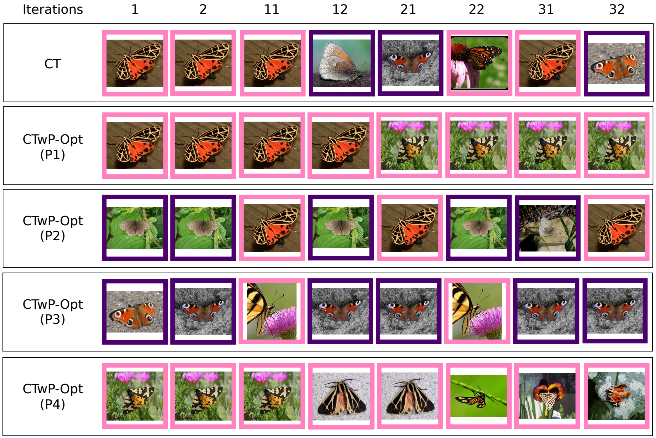
Embeddings of the dataset and target concept.



(a) Initial dataset   (b) Initial learners

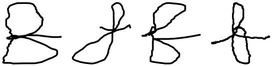
Experiments - Real World

Training examples selected by the teacher in a class with 4 types of students.

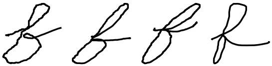


Teaching How to Write

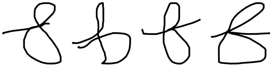
Shaky and distorted handwriting:



Shaky and rotated handwriting:

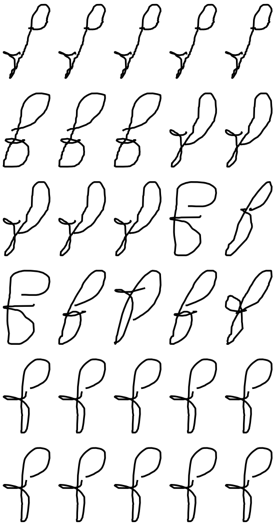


Rotated and distorted handwriting:



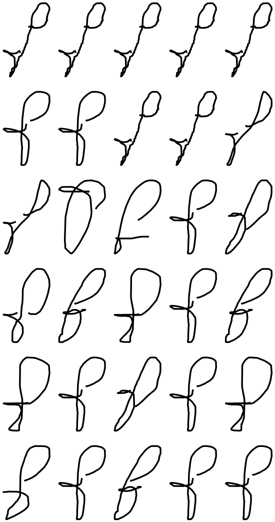
Teaching How to Write

Teaching examples for shaky and rotated handwriting:



Teaching How to Write

Teaching examples for distorted, rotated, and shaky handwriting:



Conclusions

- Trade-off between orchestratrion cost and students' workload
- Learning ability matters more than prior knowledge
- A natural homogeneous partitioning strategy based on learning ability