Locality Sensitive Teaching

Zhaozhuo Xu*, Beidi Chen†, Chaojian Li*, Weiyang Liu‡§,
Le Song$#, Yingyan Lin* and Anshumali Shrivastava*.

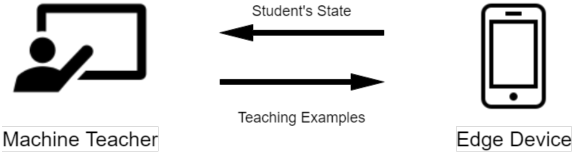
* Rice University, † Stanford University, ‡ University of Cambridge, § MPI-IS Tübingen,
$ Biomap, # MBZUAI

December, 2021

Machine Teaching

- The machine teacher wants the student to know a model w^*.
- The machine teacher cannot directly share w^* because: (1) the student is a human, (2) privacy concerns.
- The machine teacher feeds the examples to the student with the knowledge learned from w^*.

Machine Teaching on Edge Device



- A more interactive and more personalized way of education.
- A increasing demand during the COVID pandemic.

Existing software: Coursera, Duolingo, and EdX.

Existing Iterative Machine Teaching (IMT)

- Given a teacher model with parameter w^*, a teaching set S contain feature-label pairs (x, y) and a loss function $\ell(\langle w^t, x\rangle, y)$.
- Given have a student with model parameter w_0, we perform iterative machine teaching on the student in each time t by:

$$x^t, y^t = \arg\min_{x,y} \eta_t^2 \left\|\nabla l\right\|_2^2 - 2\eta_t \left\langle w^t - w^*, \nabla l \right\rangle,$$

where we have $\nabla l = \frac{\partial \ell(\langle w^t, x\rangle, y)}{\partial w_t}$.

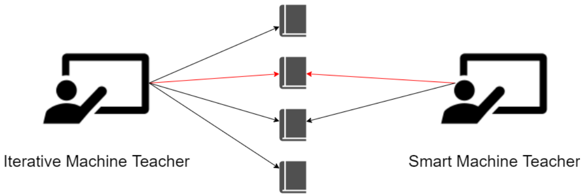
- Clearly, at least a linear scan over the data (or more) to solve the optimization. *Too Slow*

Limitation

The teacher needs to traverse all examples in the dataset.

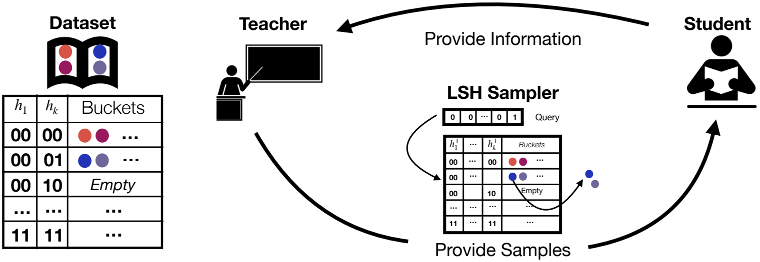
- Extremely time-consuming and expensive.
- Prevents iterative machine teaching from practice.

Intuition



- An inexperienced teacher may go over all materials when preparing the class.
- A experienced teacher would quickly locate the materials.

How can we mimic an experienced teacher efficiently: Locality Sensitive Teaching



Procedures:

- Obtain the student information.
- Look up the example from "carefully tailored" hash table.
  Probabilistic Guarantee of Relevance
- Feed the example to student.

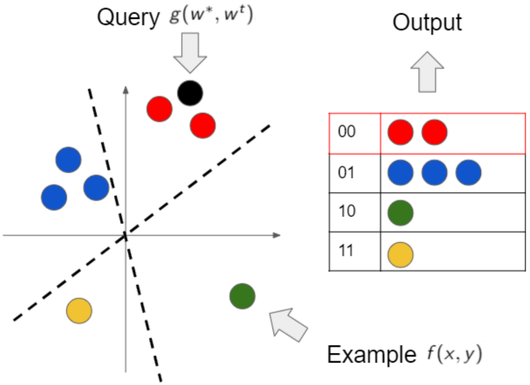
Our Formulation

We reformulate the problem as a inner product sampling problem.
Given a large set S and a query point $q = g(w^*, w^t)$, we aim to sample an example $(x, y) \subset S$ that:

$$x, y = \arg\max_{(x,y)\in S} f(x, y)^\top g(w^*, w^t)$$

- f and g are some transformation functions with the same dimension.
- this formulation is equivalent to previous formulation in linear models.

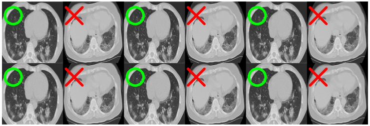
Hashing-based Sampling

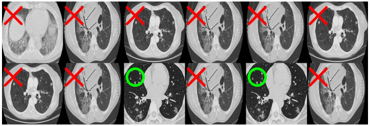


- **Preprocessing:** index the examples via hash table
- **Transform:** generate vector $g(w^*, w)$.
- **Query:** extract the examples from hash table.

First proposed in (Spring & Shrivastava 2017, Chen et. al. 2019)

Advantages of LST





Left: Example selected by IMT. **Right:** Example selected by LST.

- **Task:** teach a machine student to recognize COVID CT image.
- **Visualize:** COVID image: green circle, regular image by red cross.
- **Observation:** LST selects diverse, confusing images while IMT keeps selecting the same 2 images.

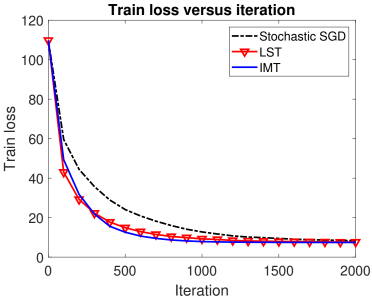
Experiment

LST can preserve teachability while accelerating the speed.

Test on two real dataset.

- Task: regression by ridge regression.
- Teacher: model pre-trained on teaching set.
- Student: model with random initialization
- Paradigm: the teacher feeds example to student in each iteration.
- Evaluation: the speed to convergence (iteration and time).

Accuracy and Efficiency



Observation:

- LST approximates IMT very well
- LST is at most 425x faster than IMT

Energy Saving on Edge Device

Deploying LST on a Nvidia TX2.

Dataset	Energy Savings	Speedups
abalone	99.76%	425.12×
space_ga	99.34%	149.07×
mg	98.73%	117.20×

Observation:

- LST saves up to 99.76% energy over IMT.
- LST enables efficient machine teaching on IoT devices.

Concluding Remarks

We propose an efficient machine teaching algorithm using advances in Efficient Statistical Sampling via LSH (Spring & Shrivastava 2017, Chen et. al. 2019).

Benefits:

- Real-time feedbacks in training labelers.
- Energy efficient machine teaching on IoT or mobile devices.

Thanks!