Discriminative Batch Mode Active Learning

Yuhong Guo and **Dale Schuurmans**
Department of Computing Science
University of Alberta
{yuhong, dale}@cs.ualberta.ca

NIPS 2008

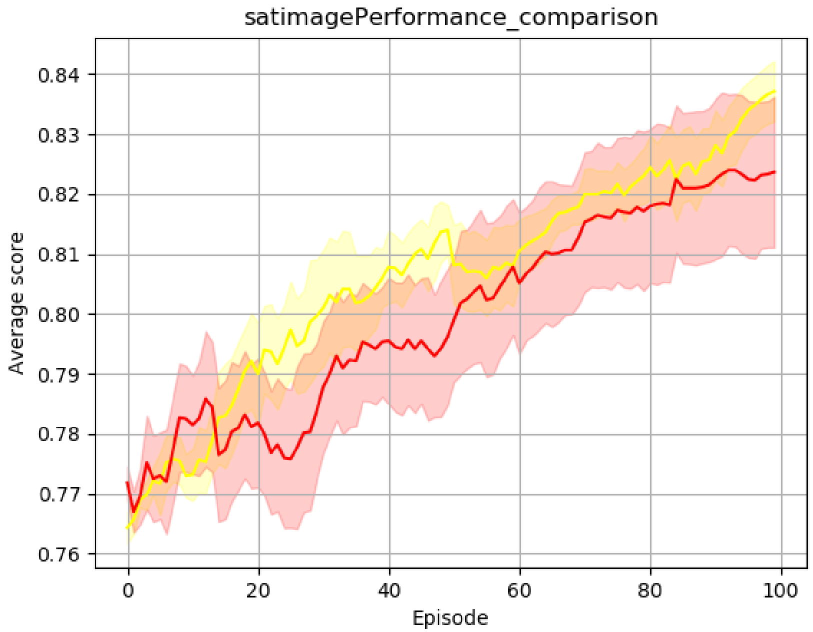
Motivation

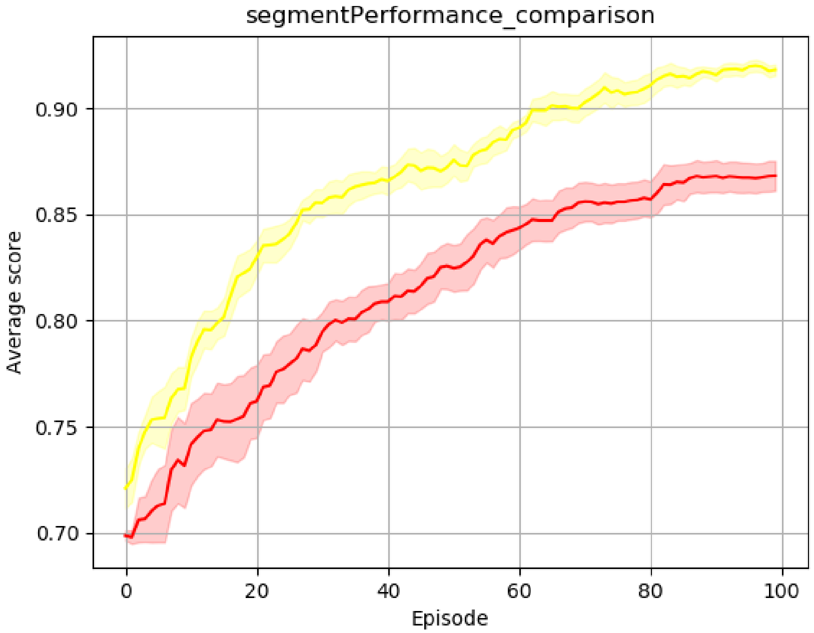
- Why uncertainty works?

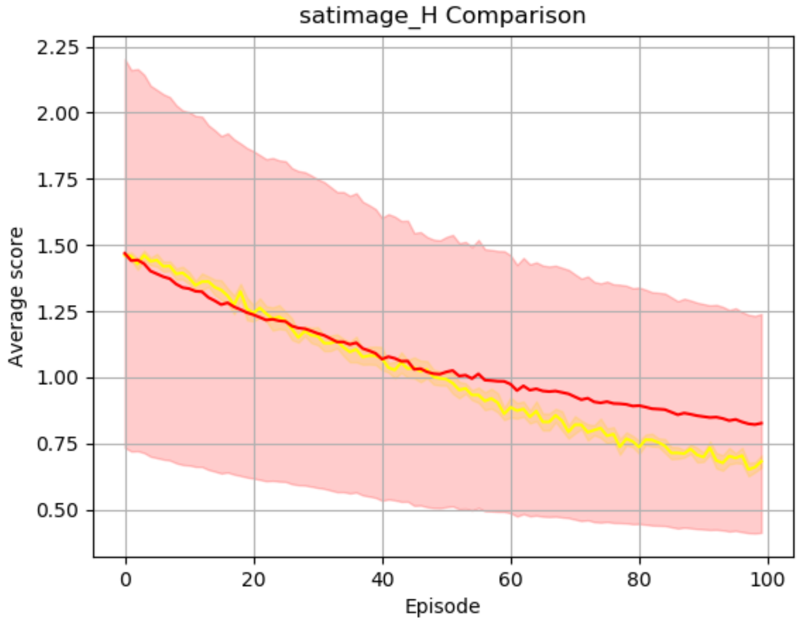
US

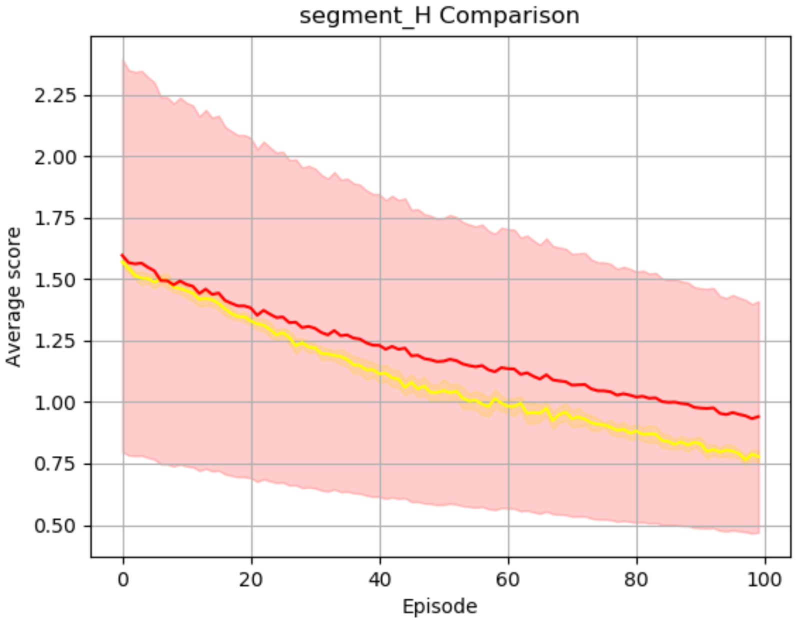
$$x = \max_{x^* \in U} H(y \mid x^*, w) \Longleftrightarrow x = \min_{x^*} \sum_{x_i \in U \setminus x^*} H(y_i \mid x_i, w)$$

从这个角度出发，能否将uncertainty看成在挑选样本后，使得未标记池的不确定性的期望降低，也就是说uncertainty是降低不确定性期望的一种贪心的算法。


satimagePerformance_comparison
Average score
Episode
0.76
0.77
0.78
0.79
0.80
0.81
0.82
0.83
0.84
0
20
40
60
80
100


segmentPerformance_comparison
Average score
Episode
0.70
0.75
0.80
0.85
0.90
0
20
40
60
80
100


satimage_H Comparison
Average score
Episode
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
0
20
40
60
80
100


segment_H Comparison
Average score
Episode
0.50
0.75
1.00
1.25
1.50
1.75
2.00
2.25
0
20
40
60
80
100

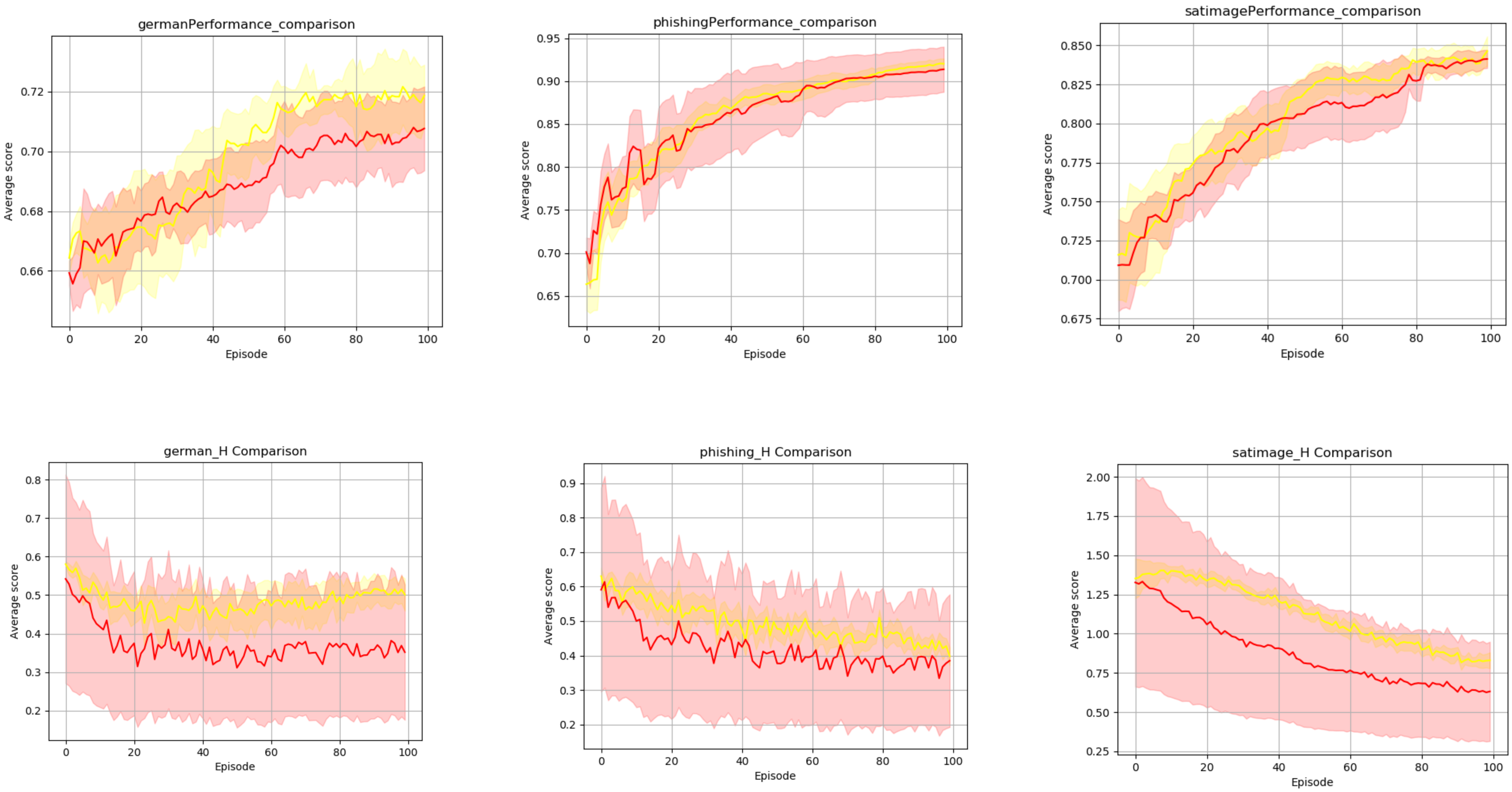
Motivation

- Why uncertainty works?

$$x = \max_{x^* \in U} H(y \mid x^*, w) \quad \overset{\text{US}}{\Longleftrightarrow} \quad x = \min_{x^*} \sum_{x_i \in U \setminus x^*} H(y_i \mid x_i, w)$$

Min H

$$x = \min_{x^*} \sum_{x_i \in U \setminus x^*} \left(H(y \mid x_i, w^{+<x^*,+1>}) + H(y \mid x_i, w^{+<x^*,-1>}) \right)$$


germanPerformance_comparison
phishingPerformance_comparison
satimagePerformance_comparison
german_H Comparison
phishing_H Comparison
satimage_H Comparison
Average score
Episode

Contents

- Methods
- Experiments

Methods

- Semi-supervised Learning by Entropy Minimization (NIPS, 2005)

$$\sum_{i \in L} \log P(y_i|\mathbf{x}_i, \mathbf{w}) + \alpha \sum_{j \in U} \sum_{y=\pm 1} P(y|\mathbf{x}_j, \mathbf{w}) \log P(y|\mathbf{x}_j, \mathbf{w})$$

- The new active learning approach:

$$f(S) = \sum_{i \in L^t \cup S} \log P(y_i|\mathbf{x}_i, \mathbf{w}^{t+1}) - \alpha \sum_{j \in U^t \setminus S} H(y|\mathbf{x}_j, \mathbf{w}^{t+1})$$

where $$H(y|\mathbf{x}_j, \mathbf{w}^{t+1}) = -\sum_{y=\pm 1} P(y|\mathbf{x}_j, \mathbf{w}^{t+1}) \log P(y|\mathbf{x}_j, \mathbf{w}^{t+1})$$

Methods

- The new active learning approach:

$$f(S) = \sum_{i \in L^t \cup S} \log P(y_i | \mathbf{x}_i, \mathbf{w}^{t+1}) - \alpha \sum_{j \in U^t \setminus S} H(y | \mathbf{x}_j, \mathbf{w}^{t+1})$$

- Active learning strategy:

$$S^* = \max_S f(S)$$

Methods

- One typical solution:

$$\mathbf{E}[f(S)] = \sum_{\mathbf{y}_S} P(\mathbf{y}_S | \mathbf{x}_S, \mathbf{w}^t) f(S)$$

- An optimistic strategy:

$$f(S) = \max_{\mathbf{y}_S} \sum_{i \in L^t \cup S} \log P(y_i | \mathbf{x}_i, \mathbf{w}^{t+1}) - \alpha \sum_{j \in U^t \setminus S} H(y | \mathbf{x}_j, \mathbf{w}^{t+1})$$

Our opproach will be exactly equivalent to picking the most uncertain instance (when |S| = 1).

Experiments

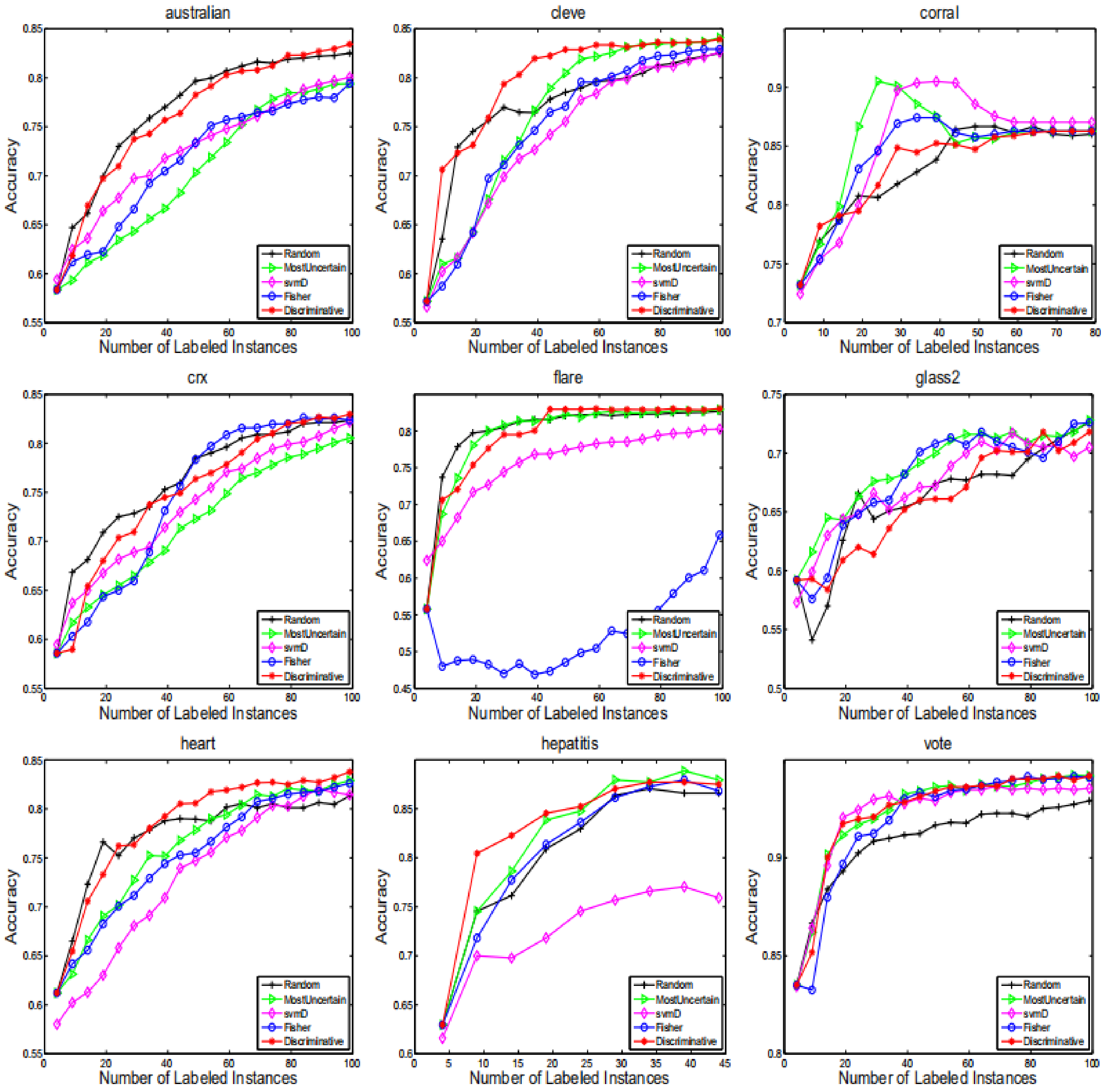


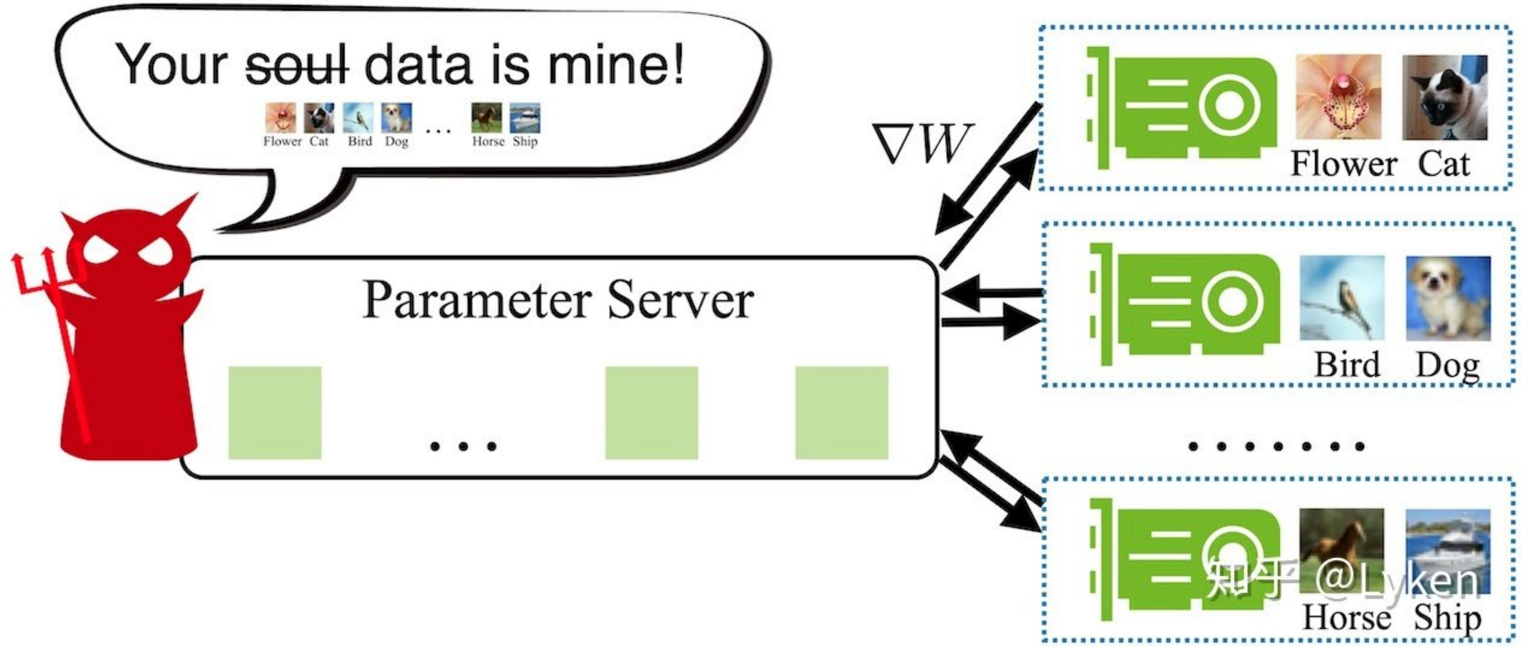
Figure 1: Results on UCI Datasets

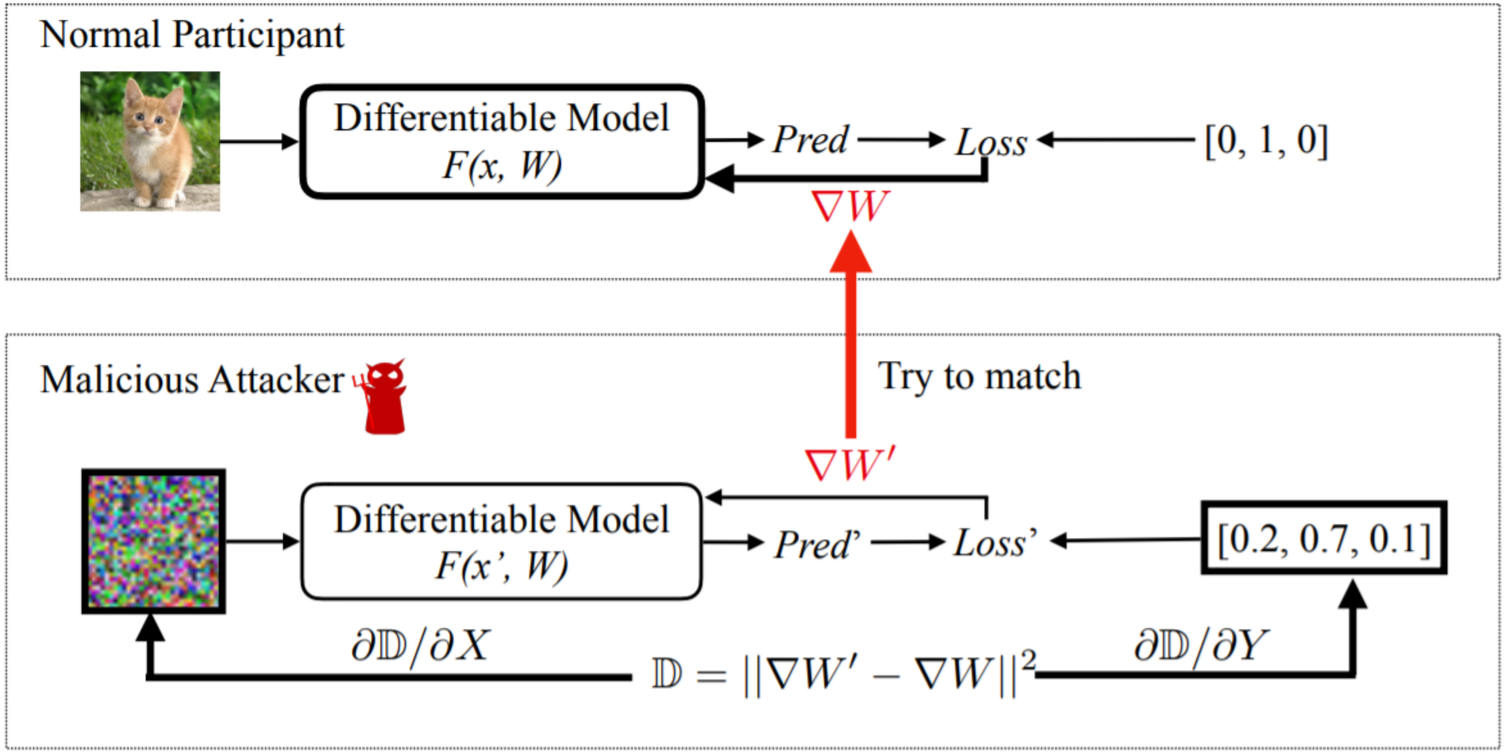
Deep Leakage from Gradients

Ligeng Zhu Zhijian Liu Song Han
Massachusetts Institute of Technology
{ligeng, zhijian, songhan}@mit.edu

NIPS 2019

Methods


Your ~~soul~~ data is mine!
Flower Cat Bird Dog ... Horse Ship
∇W
Parameter Server
...
Flower Cat
Bird Dog
....
Horse Ship
知乎 @Lyken


Normal Participant
Differentiable Model
$F(x, W)$
$Pred$
$Loss$
[0, 1, 0]
∇W
Try to match
Malicious Attacker
$\nabla W'$
Differentiable Model
$F(x', W)$
$Pred'$
$Loss'$
[0.2, 0.7, 0.1]
$\partial \mathbb{D} / \partial X$
$\mathbb{D} = ||\nabla W' - \nabla W||^2$
$\partial \mathbb{D} / \partial Y$

Algorithm

```python
def deep_leakage_from_gradients(model, origin_grad):
  dummy_data = torch.randn(origin_data.size())
  dummy_label =  torch.randn(dummy_label.size())
  optimizer = torch.optim.LBFGS([dummy_data, dummy_label] )

  for iters in range(300):
    def closure():
      optimizer.zero_grad()
      dummy_pred = model(dummy_data)
      dummy_loss = criterion(dummy_pred, dummy_label)
      dummy_grad = grad(dummy_loss, \
                 model.parameters(), create_graph=True)

      grad_diff = sum(((dummy_g - origin_g) ** 2).sum() \
        for dummy_g, origin_g in zip(dummy_grad, origin_grad))
      grad_diff.backward()

      return grad_diff
    optimizer.step(closure)
  return  dummy_data, dummy_label
```

Algorithm

Input ... culated by

training d

Outp

1: **proce**

2: x' ... and labels.

3: **fo**

4: ... gradients.

5:

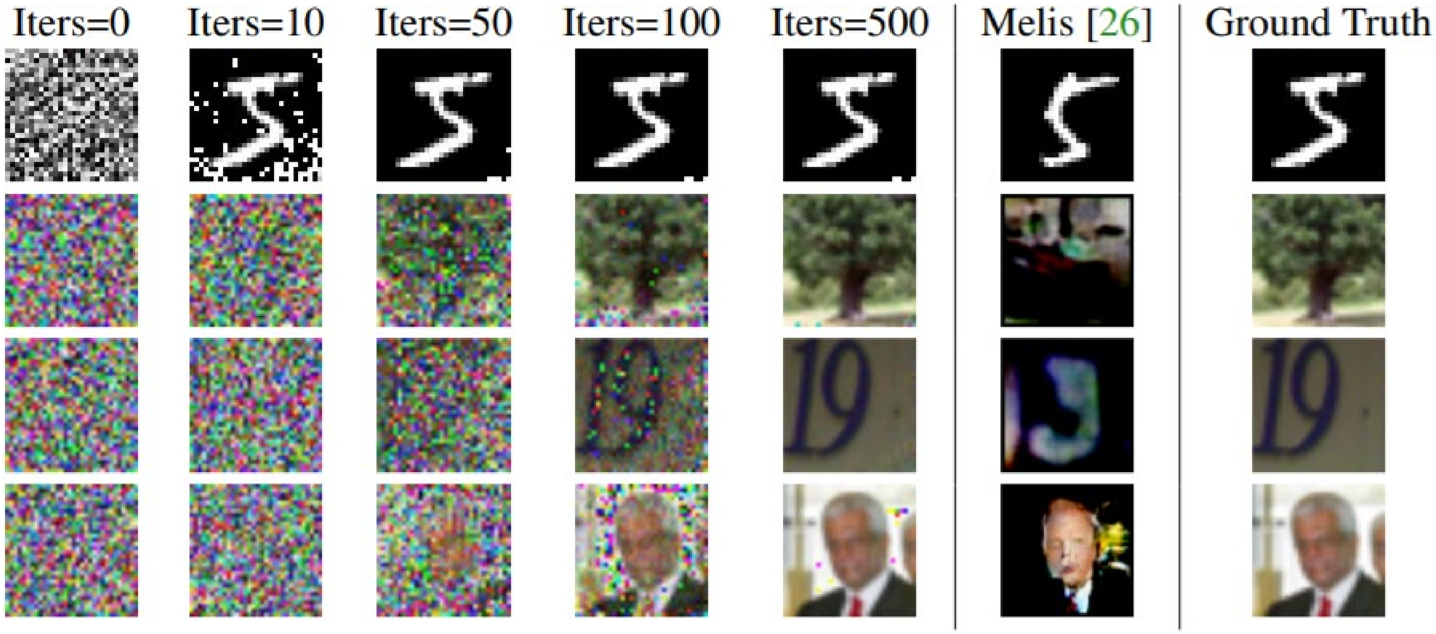
6: ... gradients.

7: **er**

8: **re**

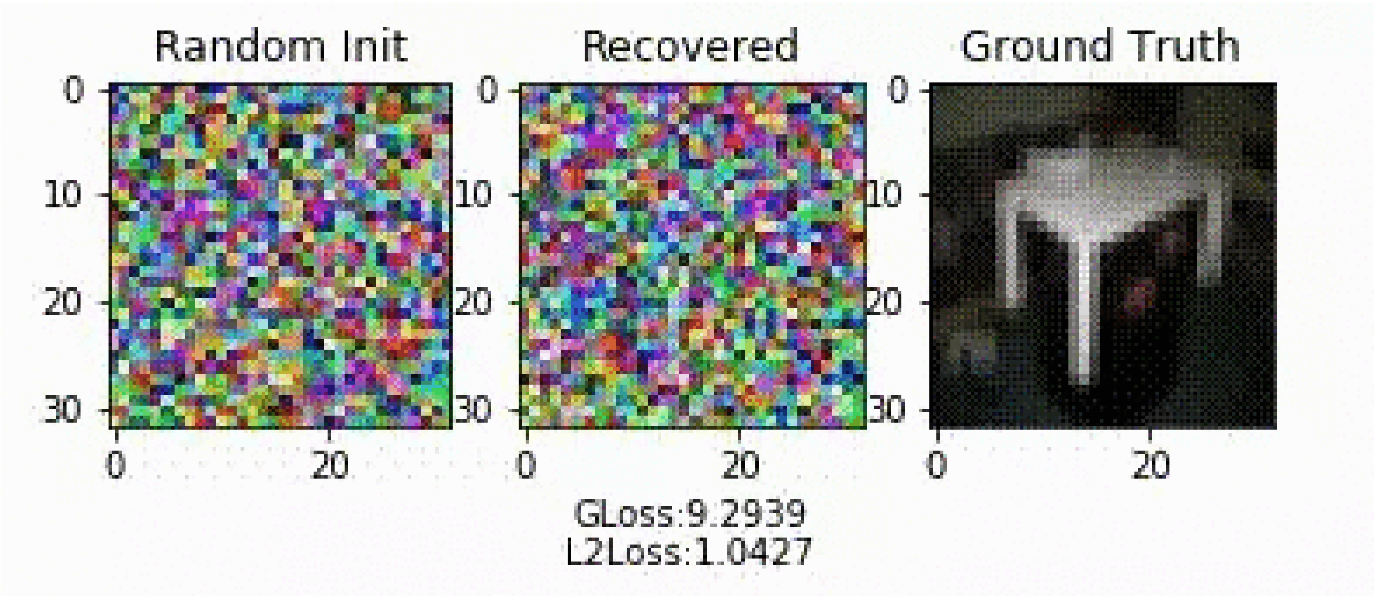
9: **end p**

Experiments



	Example 1	Example 2	Example 3
Initial Sentence	tilting fill given **less word **itude fine **nton overheard living vegas **vac **vation *f forte **dis cerambycidae ellison **don yards marne **kali	toni **enting asbestos cutler km nail **oof **dation **ori righteous **xie lucan **hot **ery at **tle ordered pa **eit smashing proto	[MASK] **ry toppled **wled major relief dive displaced **lice [CLS] us apps _ **face **bet
Iters = 10	tilting fill given **less full solicitor other ligue shrill living vegas rider treatment carry played sculptures lifelong ellison net yards marne **kali	toni **enting asbestos cutter km nail undefeated **dation hole righteous **xie lucan **hot **ery at **tle ordered pa **eit smashing proto	[MASK] **ry toppled identified major relief gin dive displaced **lice doll us apps _ **face space
Iters = 20	registration , volunteer applications , at student travel application open the ; week of played ; child care will be glare .	we welcome proposals for tutor **ials on either core machine denver softly or topics of emerging importance for machine learning .	one **ry toppled hold major ritual ' dive annual conference days 1924 apps novelist dude space
Iters = 30	registration , volunteer applications , and student travel application open the first week of september . child care will be available .	we welcome proposals for tutor **ials on either core machine learning topics or topics of emerging importance for machine learning .	we invite submissions for the thirty - third annual conference on neural information processing systems .
Original Text	Registration, volunteer applications, and student travel application open the first week of September. Child care will be available.	We welcome proposals for tutorials on either core machine learning topics or topics of emerging importance for machine learning.	We invite submissions for the Thirty-Third Annual Conference on Neural Information Processing Systems.

Table 2: The progress of deep leakage on language tasks.



THANKS