南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# BatchRank: A Novel Batch Mode Active Learning Framework for Hierarchical Classification

Shayok Chakraborty, Vineeth Balasubramanian, Adepu Ravi Sankar,Sethuraman Panchanathan and Jieping Ye

Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 99-108.

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Introduction

-hierarchical tree structure：
the leaf nodes as outputs and the internal nodes as clusters of outputs

-Structure
First propose a novel BMAL algorithm;
Convex relaxation;
A bound on the solution quality of the relaxation;
Experiment.

This is the first research effort to derive a concrete mathematical guarantee on the solution quality of batch mode active learning in hierarchical classification.

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Model

Model: hierarchical Logistic Regression

C(i): the set of all children of node i

$$f(x) = \begin{cases} \textbf{initialize} : i = 0 \\ \textbf{while} \quad \textbf{C(i)} \quad \textbf{is} \quad \textbf{not} \quad \textbf{empty} \\ \qquad i = \arg\max_{j \in C(i)} w_j^T x \\ \textbf{return } i \end{cases}$$

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Batch Selection Criterion

-valuable information C ( $c \in \Re^{|U_t| \times 1}$ )

yl: the set of labels in level L of the label tree

$$S(y_l|x_i) = -\sum_{y \in y_l} P(y|x_i) \log P(y|x_i)$$

$$c(i) = \sum_{l=1}^{d} S(y_l|x_i)$$

-minimal redundancy

kernelized distance $R(i,j) = \phi(x_i, x_j)$

# Active Batch Selection Framework

$$D(i,j) = \begin{cases} R(i,j), & \text{if } i \neq j \\ \lambda c(i), & \text{if } i = j \end{cases}$$

Integer quadratic programming (IQP) problem

$$\max_{m} m^T D m$$

$$\text{s.t.} \quad m_i \in \{0,1\}, \forall i \quad \text{and} \quad \sum_{i=1}^{|U_t|} m_i = k$$

a binary vector m
mi = 1 : included in the batch

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# An Efficient Convex Relaxation

Integer Linear Programming (ILP) problem

$$\max_{m,Z} \sum_{i,j} d_{ij} z_{ij}$$

$$\text{s.t.} \quad -m_i - m_j + 2z_{ij} \leq 0, \forall i, j$$

$$and \quad \sum_{i=1}^{|U_t|} m_i = k, m_i, z_{ij} \in \{0,1\}, \forall i, j$$

$$\max_{m} m^T D m$$

$$\max_{m,Z} \sum_{i,j} d_{ij} z_{ij}$$

$$\text{s.t.} \quad z_{ij} = m_i m_j, \quad \sum_{i=1}^{|U_t|} m_i = k, \quad and \quad m_i \in \{0,1\}, \forall i$$

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# An Efficient Convex Relaxation

LP relaxation

$$\max_{m} \frac{1}{2}\sum_{i,j} d_{ij}(m_i + m_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{|U_t|} m_i = k \quad d \quad m_i \in [0,1], \forall i$$

$$\sum_{i,j} d_{ij} m_i + \sum_{i,j} d_{ij} m_j$$

$$\max_{m,Z} \sum_{i,j} d_{ij} z_{ij}$$

$$\text{s.t.} \quad -m_i - m_j + 2z_{ij} \le 0, \forall i,j$$

$$and \quad \sum_{i=1}^{|U_t|} m_i = k, m_i, z_{ij} \in \{0,1\}, \forall i,j$$

$$z_{ij} = \frac{m_i + m_j}{2}$$

BatchRank

# Iterative Truncated Power Algorithm

**Algorithm 1** BatchRank algorithm for Batch Mode Active Learning in Hierarchical Classification

**Require:** Training set $L_t$, Unlabeled set $U_t$ and batch size $k$

1: Train a classifier $w^t$ on the training set $L_t$
2: Compute information vector $c$ (Equation 4) and the divergence matrix $R$ (Equation 5)
3: Compute the matrix $D$, as described in Equation 9
4: Compute a vector $v \in \Re^{|U_t| \times 1}$ containing the column sums of $D$
5: Identify the $k$ largest entries in $v$ and derive the initial solution $x_0$
6: t = 1
7: **repeat**
8: Compute $x_t^{'} = D.x_{t-1}$
9: Identify $F_t$ as the index set of $x_t^{'}$ with top $k$ values
10: Set $x_t$ to be 1 on the index set $F_t$ and 0 otherwise
11: t = t + 1
12: **until** Convergence
13: Select a batch of $k$ unlabeled samples based on the final solution $x_t$

At each time step t, the vector $x_{t-1}$ is multiplied by the weight matrix D and then the entries are truncated to zeros except for the k largest entries, which becomes the new solution xt. This process is repeated until convergence.

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Solution Bound of BatchRank

$$f(m) = ||D||_1 - m^T Dm$$

$$\max_m m^T Dm$$

$$\max_{m,Z} \sum_{i,j} d_{ij} z_{ij}$$

$$\max_m \frac{1}{2} \sum_{i,j} d_{ij}(m_i + m_j)$$

THEOREM 1. *Let $m^*$ and $\widehat{m}$ be the optimal solutions of the original NP-hard IQP in Equation (10) and the convex relaxation in Equation (14) respectively. Then,*

$$f(\widehat{m}) \leq 2f(m^*)$$

$$m^{*T} Dm^* \leq \frac{1}{2} \sum_{i,j} d_{ij}(\widehat{m_i} + \widehat{m_j})$$

$$= \frac{1}{2} \sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} + \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij}$$

# Solution Bound of BatchRank

$$
\begin{aligned}
||D||_1 &= \sum_{ij} d_{ij} \\
&= \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij} + \sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} + \sum_{i,j:\widehat{m_i}+\widehat{m_j}=0} d_{ij} \\
&\geq \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij} + \sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij}
\end{aligned}
$$

Thus,

$$
\sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} \leq ||D||_1 - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij} \qquad (17)
$$

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Solution Bound of BatchRank

Combining the above two, we have

$$f(m^*) = ||D||_1 - m^{*T} D m^*$$

$$\geq ||D||_1 - \frac{1}{2} \sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij}$$

$$\geq ||D||_1 - \frac{1}{2} \left( ||D||_1 - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij} \right) - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij}$$

$$= \frac{1}{2} \left( ||D||_1 - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij} \right) = \frac{1}{2} f(\widehat{m})$$

$$m^{*T} D m^* \leq \frac{1}{2} \sum_{i,j} d_{ij} (\widehat{m_i} + \widehat{m_j})$$

$$= \frac{1}{2} \sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} + \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij}$$

$$\sum_{i,j:\widehat{m_i}+\widehat{m_j}=1} d_{ij} \leq ||D||_1 - \sum_{i,j:\widehat{m_i}+\widehat{m_j}=2} d_{ij}$$

$$f(m) = ||D||_1 - m^T D m$$

# Experiment

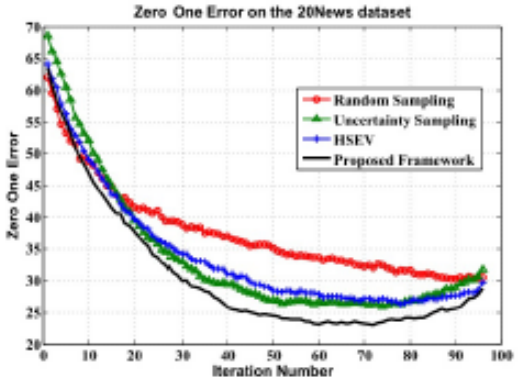
## Evaluation Metrics

zero-one error (error rate)

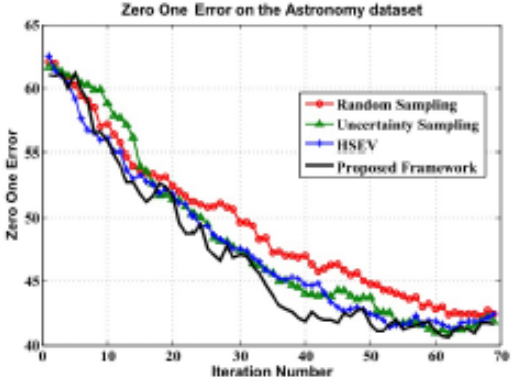
the tree-loss error (the graph distance between the predicted and actual categories)

## Datasets

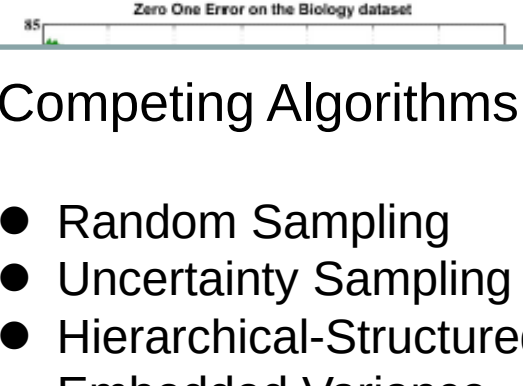
| Dataset | Dimensionality | Label Tree Size | Initial Training | Unlabeled | Testing | Batch Size |
|---|---|---|---|---|---|---|
| 20 Newsgroups | 26,214 | 25 | 100 | 2,900 | 2,000 | 30 |
| Astronomy | 54,632 | 34 | 100 | 700 | 845 | 10 |
| Biology | 148,644 | 99 | 100 | 1,900 | 1,151 | 30 |
| CLEF | 80 | 47 | 100 | 2,900 | 2,000 | 30 |
| Earth Sciences | 71,756 | 52 | 100 | 1,900 | 1,102 | 30 |
| Math | 108,559 | 104 | 100 | 1,900 | 1,862 | 30 |
| OHSUMED | 18,143 | 87 | 100 | 2,900 | 2000 | 30 |
| Reuters | 47,236 | 97 | 100 | 2,900 | 2000 | 30 |
| WIPO | 74,437 | 188 | 100 | 900 | 700 | 20 |




(a) 20 News

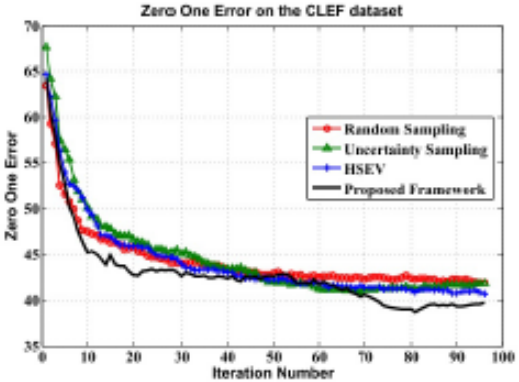



(b) Astronomy

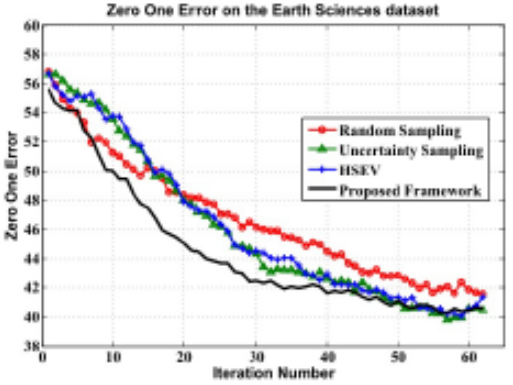



## Competing Algorithms

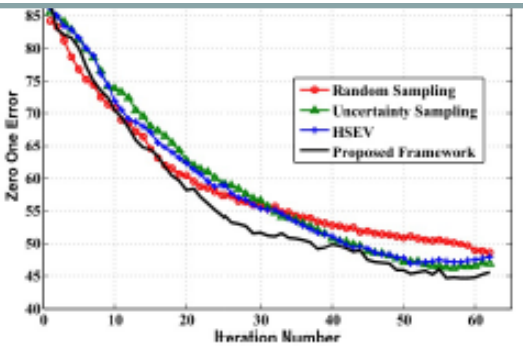
- Random Sampling
- Uncertainty Sampling
- Hierarchical-Structured Embedded Variance (HSEV)



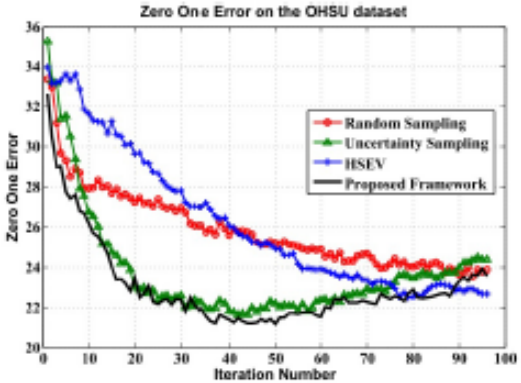

(d) CLEF



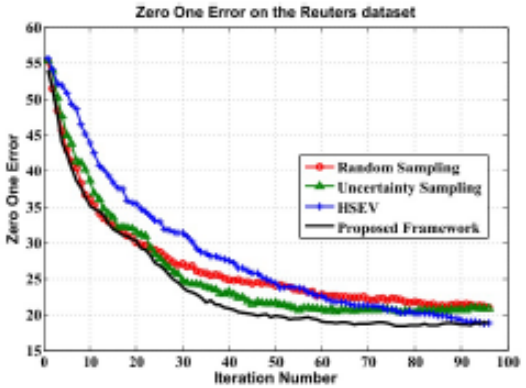

(e) Earth Sciences



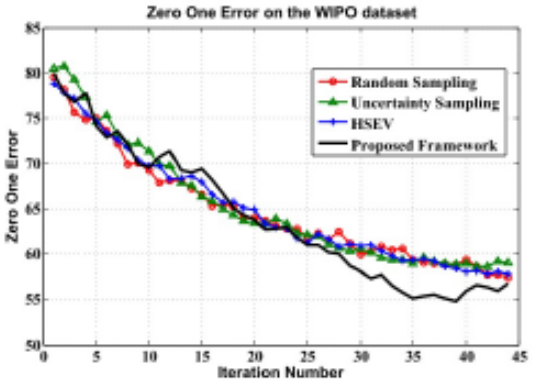

(f) Math



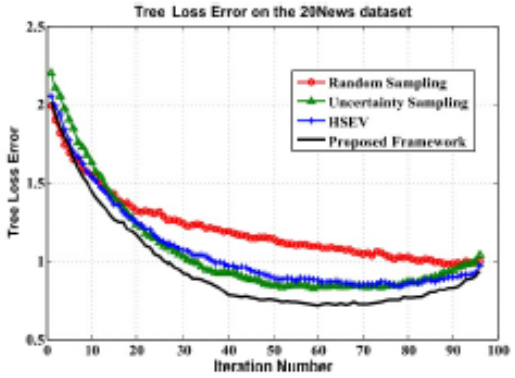

(g) OHSUMED



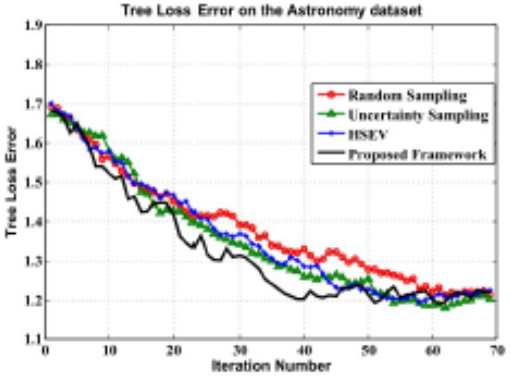

(h) Reuters



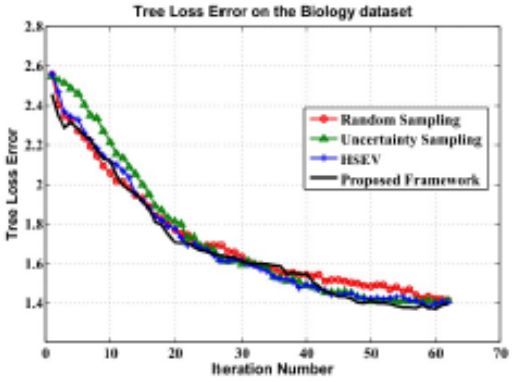

(i) WIPO



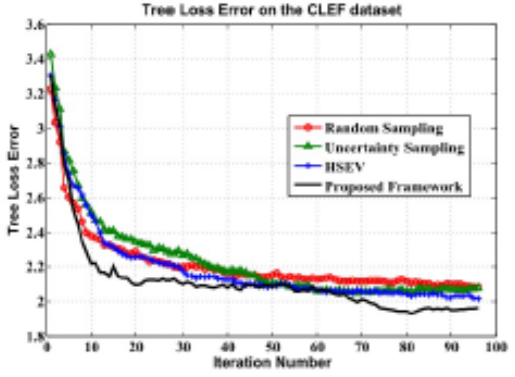

(a) 20 News



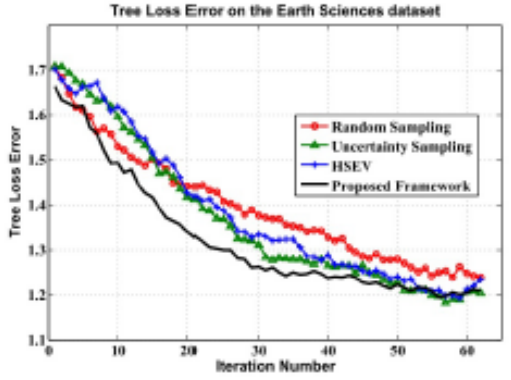

(b) Astronomy



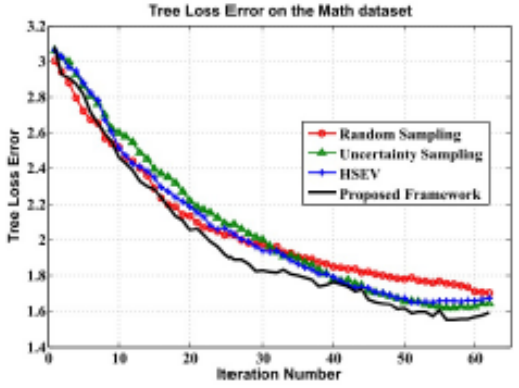

(c) Biology



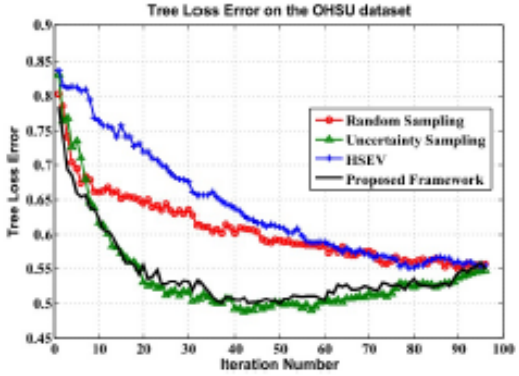

(d) CLEF



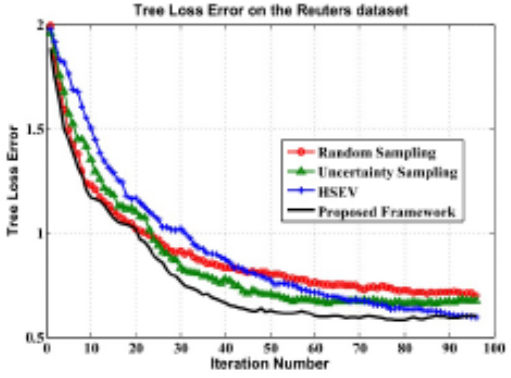

(e) Earth Sciences



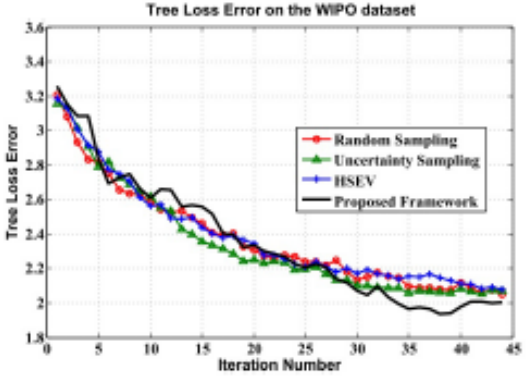

(f) Math




(g) OHSUMED




(h) Reuters




(i) WIPO

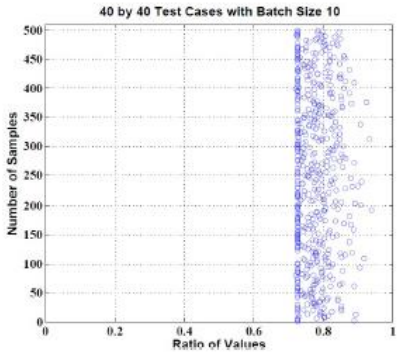
南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

# Experiment

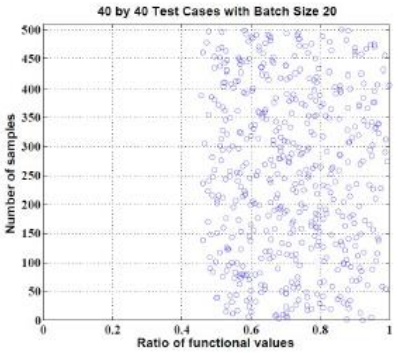
## Solution Quality Analysis

we empirically validate the quality of the solution obtained using the proposed algorithm, for different values of the number of unlabeled samples **n** and the batch size **k**.

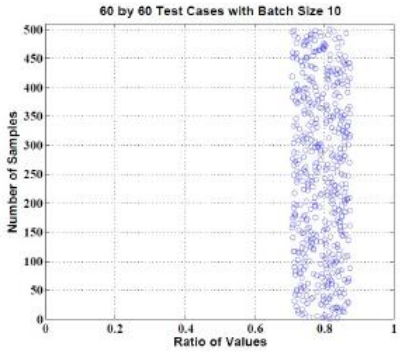
$$\frac{\hat{m}^T D \hat{m}}{m^{*T} D m^*}$$



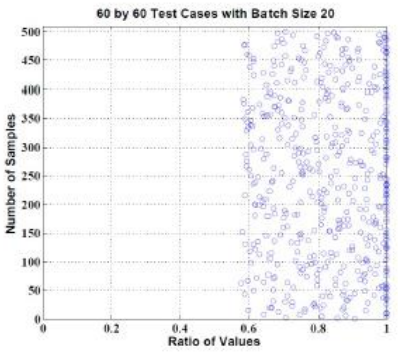

(a) $n = 40, k = 10$




(b) $n = 40, k = 20$




(c) $n = 60, k = 10$




(d) $n = 60, k = 20$

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

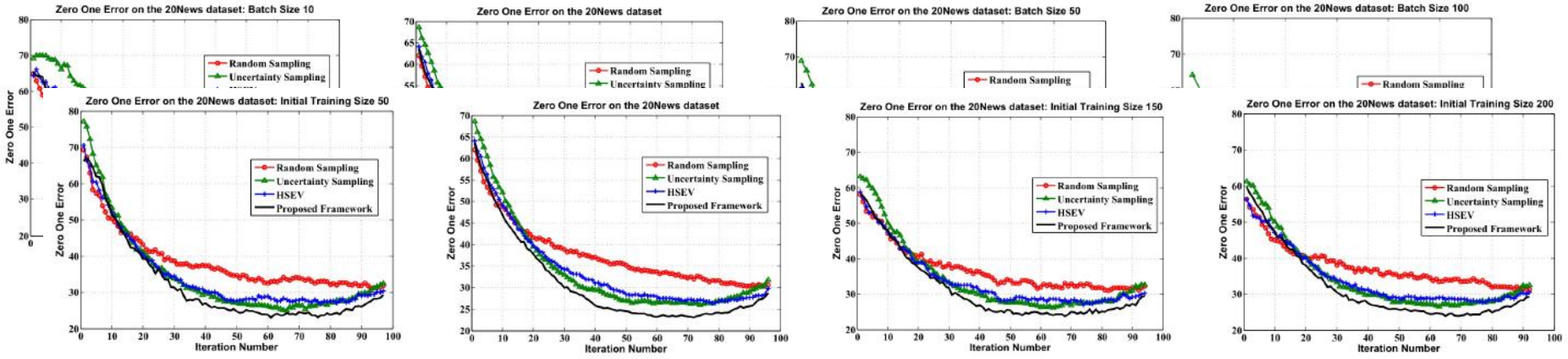
# Experiment

## Parameter Sensitivity

different batch sizes :10, 30, 50,100
different initial training set : 50, 100, 150 , 200

## 20 Newsgroups dataset

zero-one error




(a) Initial Training Size 50 (b) Initial Training Size 100 (c) Initial Training Size 150 (d) Initial Training Size 200

# Thank you