Batch Mode Sparse Active Learning

Lixin Shi, Yuhang Zhao
Tsinghua University

Our work

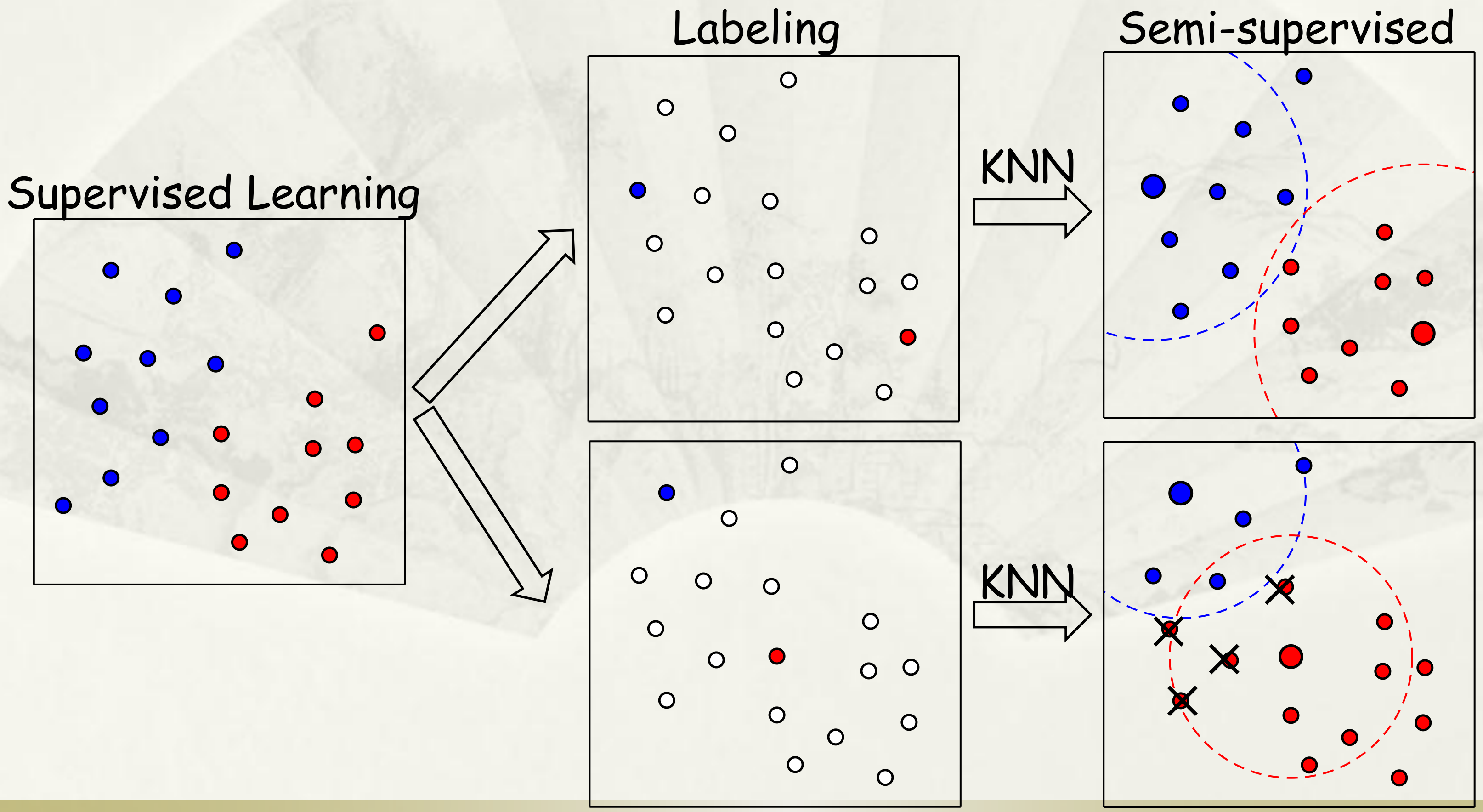
- Propose an unified framework of batch mode active learning
- Instantiate the framework using classifiers based on sparse representation (BMSAL)
- Explore the reliability of BMSAL in different data sets

Outline

* Active Learning
* BMSAL
* Experiments
* Future Work

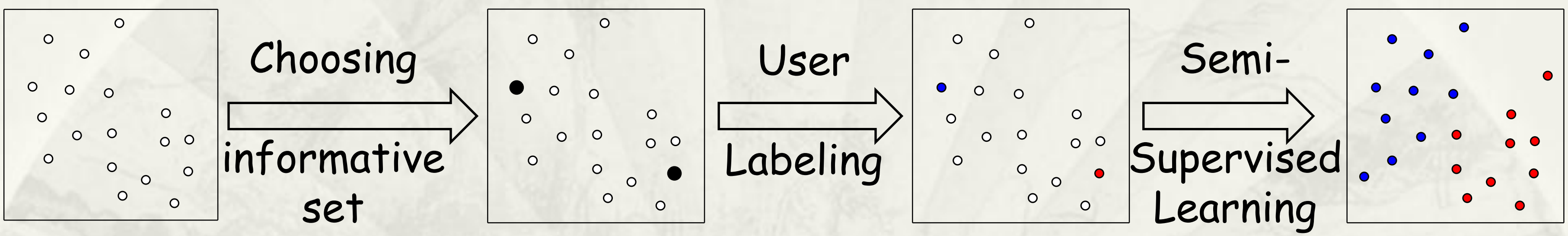
Why active learning

- Labeling is Expensive
- Which to be labeled is curial



Framework

- To Reduce the unreliability of random labeling
- BMAL(Batch Mode Active Learning) Framework



Given a set *S* of almost unlabeled samples and desired size *K*, find a set of *K* samples which are most informative

How?

- Existing Heuristics
  - Most uncertainty
  - Closest to SVM decision boundary
  - Maximizing Fisher Information Matrix
  - …
- But
  - Is heuristics *reliable*?
  - Are there any *unified* framework?

Classifiers: a review

* Classifiers are well-founded and well-learned
  * SVM, KNN, ……
* They could be restated as: *given an objective function f*, we want to find class *c**, s.t.

$$c^* = \arg\min_{c} f_c(s)$$

where $s \in S$ is the sample to be classified

Correspondence

- Correspondence:

  BMAL is to choose the sample set best minimizing the corresponding classifier function *f* for any possible labeling

$$\arg\min_{|D|=k} \left\{ \mathop{\mathbb{E}}_{D\text{'s label}} \left(\sum_{s \in S} \mathop{\mathbb{E}}_{\text{class } c} f_c(s) \right) \right\}$$

  If distribution is not available:

$$\arg\min_{|D|=k} \left\{ \max_{D\text{'s label}} \left(\sum_{s \in S} \min_{\text{class } c} f_c(s) \right) \right\}$$

Outline

- Active Learning
- BMSAL
- Experiments
- Future Work

BMSAL

- BMSAL is an instance of BMAL corresponding to sparse classifiers

Linear Subspace Assumption

- Samples in the same class forms a linear subspace with very small dimension
- Different classes forms disjoint subspaces
- Sparse Representation

Columns are bases of these subspaces, i.e. $A=[\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n]$	Sparse Representation	A given Sample (without noise)

$$A \cdot \alpha = x$$

α is the sparsest solution: Non-zero entries are only those correspond to the bases of the class

Sparse Classifiers (I)

- L1 (l1-minimization)
  - Approximation

L0 norm

L1 norm

$$\alpha^* \text{ is sparsest} \Leftrightarrow \alpha^* = \arg\min \|\alpha\|_0 \overset{\text{Aprox}}{\Leftrightarrow} \alpha^* = \arg\min \|\alpha\|_1$$

  - L1 classifier select class c* that minimizes:

$$f_c(x) = \left\| A \cdot \delta_c(\alpha^*) - x \right\|_1, \text{where } \alpha^* = \arg\min_{\alpha} \{\| \alpha \|_1 : x = A\alpha\}$$

All entries are 0, except that entries corresponding to the bases of class c are same with α*

L1 classifier finds the class that minimizes the error when representing x using the sparsest solution

Sparse Classifiers (II)

- NS (Nearest Subspace)
  - Approximation to L1: the sparsest solution of *x* has the same projection as *x* itself onto the subspace of the class that *x* belongs to
  - NS selects class c* that minimizes

$$f_c(x) = \left\| A \cdot \delta_c(\alpha^*) - x \right\|_1 \approx \| A x_c - x \|_1$$

x_c is projection of x onto subspace of class c

NS classifier finds the class whose subspace is nearest to x

Sparse Classifiers (III)

- NN (Nearest Neighbor)
  - Approximation to NS: The projection of x should be the same with the base closest to x
  - NN selects class c* that minimizes

$$f_c(x) = \|A \cdot x_c - x\|_1 \approx \| A b_c - x \|_1$$

b_c is the base vector of the subspace corresponding to class c and which minimizes the distance to x

NN classifier finds the class whose subspace has a base vector with minimized distance to x

BMSAL

- *Corresponding* Objective functions

Kind	Sparse Classifier	BMSAL
L1	$f_c(x) = \left\| A \cdot \delta_c(\alpha^*) - x \right\|_1$	$g(D) = \sum_{x \in S} \min\{\| \alpha \|_1 : D\alpha = x\}$
NS	$f_c(x) = \| Ab_c - x \|_1$	$g(D) = \sum_{x \in S} \| x - DD^+ x \|_2^2$
NN	$f_c(x) = \| Ax_c - x \|_1$	$g(D) = \sum_{x \in S} \min_{b \in D} \| x - b \|_2^2$

BMSAL: choose columns of D to minimize $g(D)$

BMSAL: Shared Properties

- Monotonic
  - The objective function g decreases as the number of selected samples to be labeled increases
- (Approx) Submodularity
  - The speed that g decreases will get slower (with bounded errors) when number of samples to be labeled increases

Proofs could be found in the paper

BMSAL: Algorithms

- Due to the shard properties, we can get a greedy algorithm, with bounded error rate ~ (1-1/e)
- We further optimize the greedy algorithm for large-scale data sets

Proofs could be found in the paper

Outline

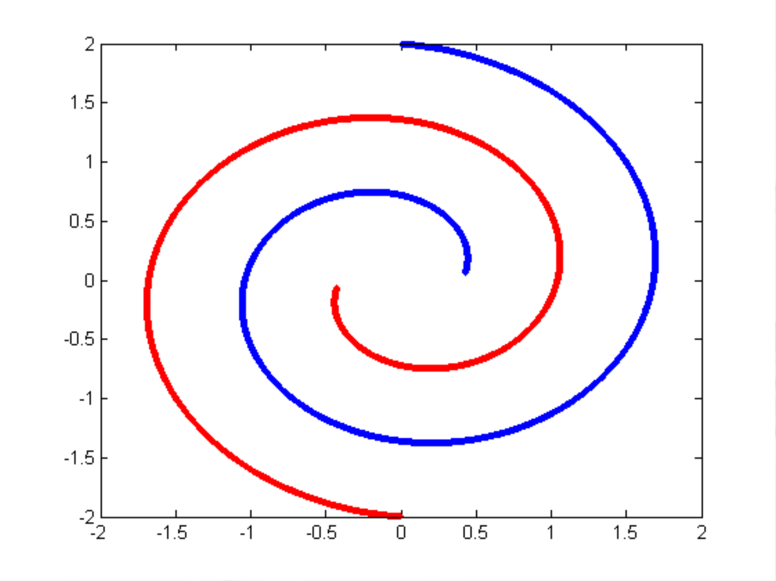
- Active Learning
- BMSAL
- Experiments
- Future Work

Experiments

- Two Goals:
  - Provide empirical evidence about the performance of BMSAL
  - Check the performance of sparse representation based BMSAL in non-linear data sets that does NOT satisfy the linear subspace assumption

Synthetic Data Set

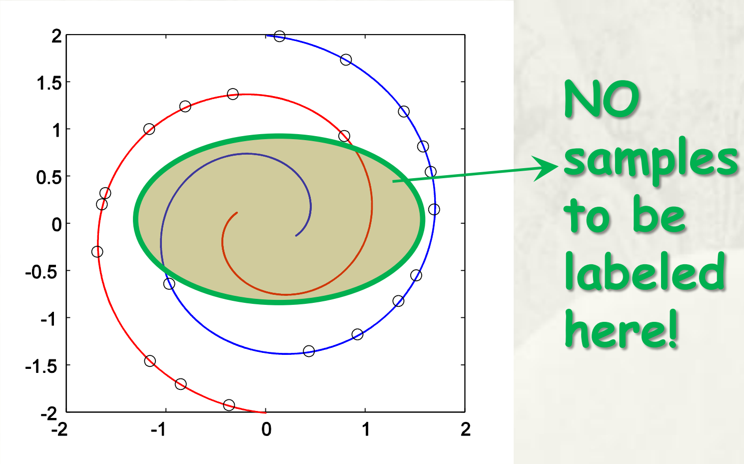
- Setup
  - Binary Classification in the two-spirals data
  - Methods:
    - L1-BMSAL + L1
    - NS-BMSAL + NS
    - NN-BMSAL + NN



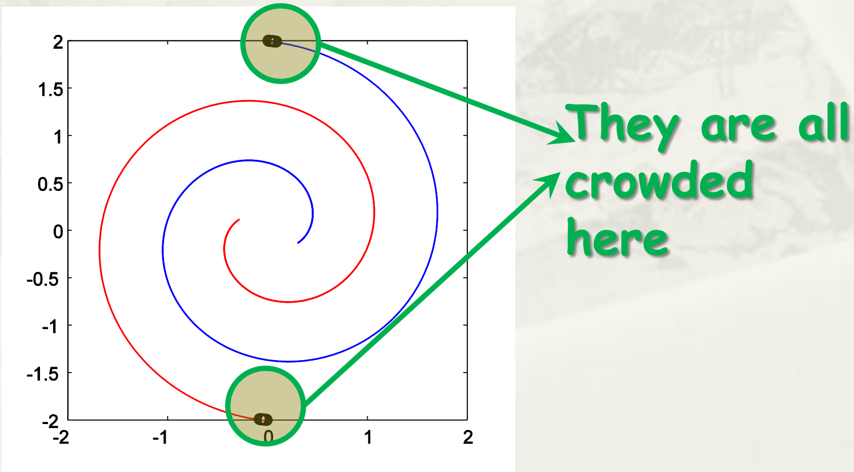
Result in Synthetic Sets

- Precision Result

Method	Average Precision
NN-BMSAL + NN	56%
NS-BMSAL + NS	52%
L1-BMSAL + L1	98%



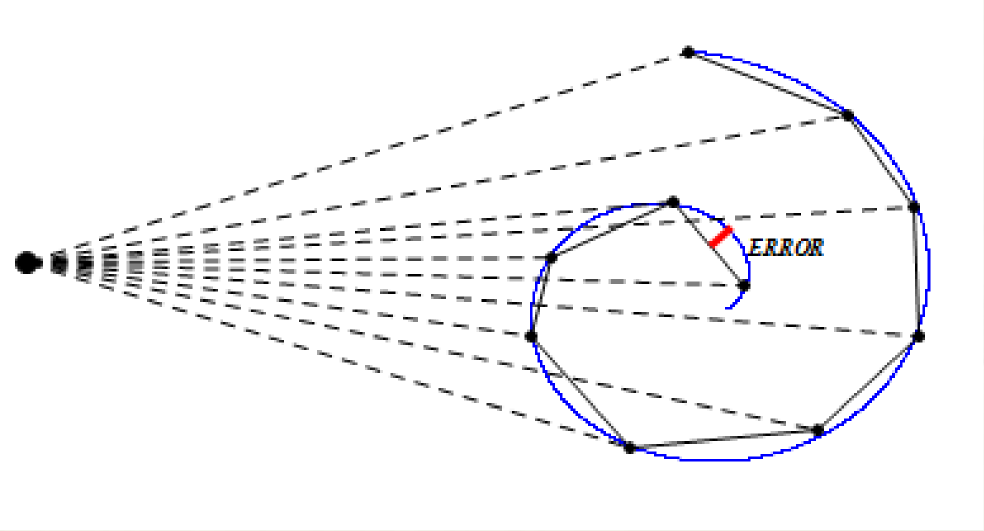
NN-BMSAL



NS-BMSAL

Piece-wise Argument

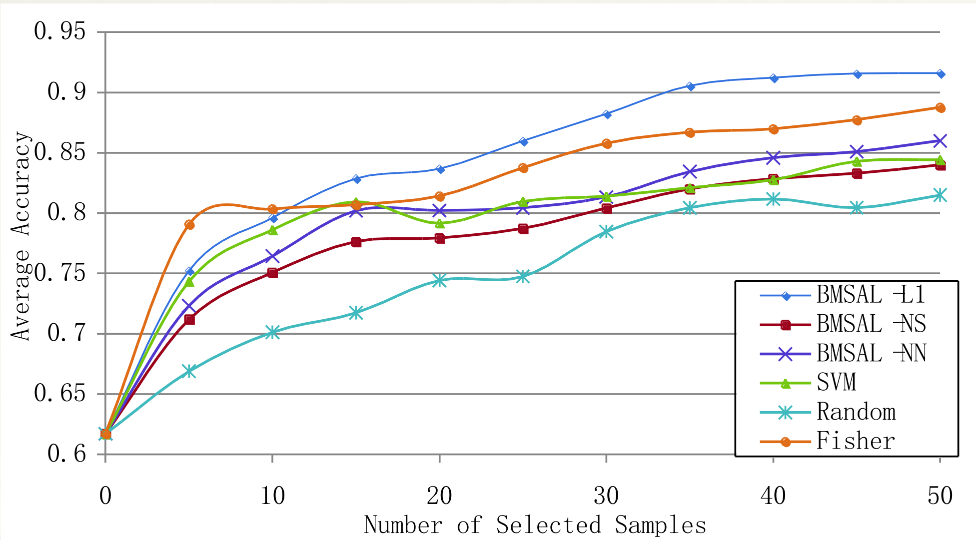
- Assumption: original point is far
- Piece-wise: point in each piece could be approximately viewed as linear combination of the two ends

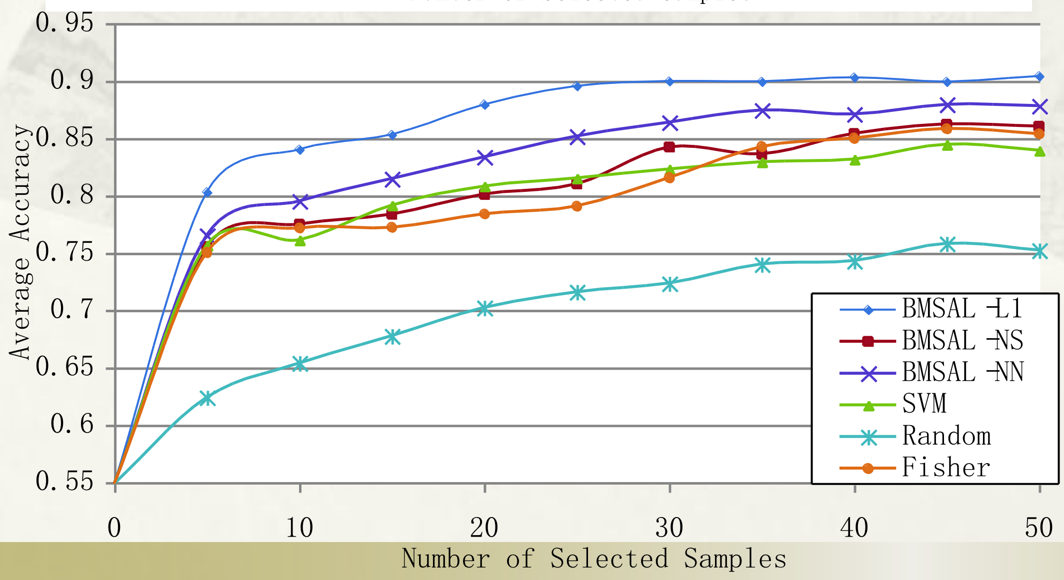


Real-world Data set

- Document Classification sets:
  - UCI 20NewsGroups
  - WebKB
- Baseline
  - Random Choosing
  - Fisher Information based
  - SVM-based BMAL

Result in Real-world Data





- L1-BMSAL+L1 outperforms others
- Extensive experiments show that L1 is also reliable

Outline

- Active Learning
- BMSAL
- Experiments
- Future Work

Future Work

- Reliability of Sparse representation
  - We have only provide logical and empirical evidence
  - Provide theoretical foundations of BMSAL in non-linear application
- Exploit BMAL corresponding with other family of classifiers

- Thank you!
- Q&A