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
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:

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

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

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. $A = [\beta_1, \beta_2, \ldots, \beta_n]$  
Sparse Representation  
A given Sample (without noise)  
\[ A \cdot \alpha = x \]

$\alpha$ 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: \( \alpha^* \) is sparsest \( \Leftrightarrow \alpha^* = \arg \min \| \alpha \|_0 \) \( \Leftrightarrow \alpha^* = \arg \min \| \alpha \|_1 \)
    - Approx \( \Rightarrow \) L1 norm

* L1 classifier select class \( c^* \) that minimizes:

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

All entries are 0, except that entries corresponding to the bases of class \( c \) are same with \( \alpha^* \)

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

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

$x_c$ 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

\[
f_c(x) = \| A \cdot x_c - x \|_1 \approx \| Ab_c - x \|_1
\]

\( b_c \) 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 \( x \)**
### Corresponding Objective functions

<table>
<thead>
<tr>
<th>Kind</th>
<th>Sparse Classifier</th>
<th>BMSAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>( f_c(x) = | A \cdot \delta_c(\alpha^*) - x |_1 )</td>
<td>( g(D) = \sum_{x \in S} \min { | \alpha |_1 : D\alpha = x } )</td>
</tr>
<tr>
<td>NS</td>
<td>( f_c(x) = | Ab_c - x |_1 )</td>
<td>( g(D) = \sum_{x \in S} | x - DD^+ x |_2^2 )</td>
</tr>
<tr>
<td>NN</td>
<td>( f_c(x) = | Ax_c - x |_1 )</td>
<td>( g(D) = \sum_{x \in S} \min_{b \in D} | x - b |_2^2 )</td>
</tr>
</tbody>
</table>

**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 $\sim (1-1/e)$

* 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

<table>
<thead>
<tr>
<th>Method</th>
<th>Average Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN-BMSAL + NN</td>
<td>56%</td>
</tr>
<tr>
<td>NS-BMSAL + NS</td>
<td>52%</td>
</tr>
<tr>
<td>L1-BMSAL + L1</td>
<td>98%</td>
</tr>
</tbody>
</table>

They are all crowded here!
NO samples to be labeled here!
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