An Optimization Based Framework for Dynamic Batch Mode Active Learning

Workshop on Optimization for Machine Learning

Shayok Chakraborty
Vineeth Balasubramanian
Sethuraman Panchanathan
Agenda

- Batch Mode Active Learning (BMAL)
- Existing Approaches
- Optimization Framework for Dynamic BMAL
- Experiments and Results
- Conclusion
Agenda

- Batch Mode Active Learning (BMAL)
- Existing Approaches
- Optimization Framework for Dynamic BMAL
- Experiments and Results
- Conclusion
Why Active Learning: A General Perspective

Cancer or non-cancer image ??

Experienced doctors are rare and busy

Face Recognition

Text Mining

Optical Character Recognition

Cancer or non-cancer image ??

Experienced doctors are rare and busy

Face Recognition

Text Mining

Optical Character Recognition
Why Active Learning: A General Perspective

Cancer or non-cancer image??

Experienced doctors

Useful when:

- A large amount of unlabeled data is readily available
- Labeling the data is time consuming and expensive

Text Mining

Optical Character Recognition
Batch Mode Active Learning

Unlabeled pool of points

Human Annotator

Classifier update module (e.g., Face recognition classifier)

Training Set
Usefulness of BMAL in Biometrics

- Inherent redundancy in biometric data
- Significant challenge to select the representative images to update a classification model

Data from a video stream containing redundant images
Usefulness of BMAL in Biometrics

- Inherent redundancy in biometric data
- Significant challenge to select the representative images to update a classification model

Select a batch of promising images to learn from
Batch Mode Active Learning (BMAL)

Existing Approaches

Optimization Framework for Dynamic BMAL

Experiments and Results

Conclusion
Existing approaches of BMAL

- Mostly based on greedy heuristics to select a batch of points

- Require the batch size to be specified in advance – impractical for applications like biometrics
Existing approaches of BMAL

Illustration:

Subject 1
(known to the learner)

Unlabeled Video 1

Subject 2
(unknown to the learner)

Unlabeled Video 2
Existing approaches of BMAL

Illustration:

Subject 1  
(known to the learner)

Subject 2  
(unknown to the learner)

Unlabeled Video 1

Unlabeled Video 2

Batch size should be larger for the second video
Existing approaches of BMAL

Illustration:

- Subject 1 (known to the learner)
- Subject 2 (unknown to the learner)
- Unlabeled Video 1
- Unlabeled Video 2

Strong need for dynamic batch selection in BMAL
Agenda

- Batch Mode Active Learning (BMAL)
- Existing Approaches
- Optimization Framework for Dynamic BMAL
- Experiments and Results
- Conclusion
Problem Statement

**Given**: A training set $L_t$ and an unlabeled set $U_t$ at time $t$

**Goal**: Select a batch of unlabeled points $B$ (of unknown size) so as to maximize the performance of the future learner

**Intuition**: 
- Select the batch such that the modified classifier has low uncertainty on the unselected points
- Also select points from the low density regions of the unlabeled set
Optimization Formulation

Current training set

Current unlabeled pool
Let $B$ denote the set of points to be selected.
Let $B$ denote the set of points to be selected.
Let $B$ denote the set of points to be selected. Define an objective function $f(B)$ to be maximized.

$$f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1})$$
Let $B$ denote the set of points to be selected. Define an objective function $f(B)$ to be maximized.

$$f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1})$$

Select points from low density regions

Entropy of the unselected points
Optimization Formulation

\[ f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) \]

Obvious Solution: Select all points in the unlabeled set

Incurs huge labeling cost, defeats basic purpose of active learning
Optimization Formulation

\[ f(B) = \sum_{j\in B} \rho_j - \lambda_1 \sum_{j\in U_t - B} S(y|x_j, w^{t+1}) \]

Obvious Solution: Select all points in the unlabeled set

Incurs huge labeling cost, defeats basic purpose of active learning.
Optimization Formulation

\[ f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) \]

Enforce a penalty on the unknown batch size

\[ \tilde{f}(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m \]
Optimization Formulation

\[
f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1})
\]

Enforce a penalty on the unknown batch size

\[
\tilde{f}(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m
\]

Penalty term
Optimization Formulation

\[ f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) \]

- Enforce a penalty on the unknown batch size
- Exhaustive search is not feasible
- Penalty term
Define a binary vector $\mathbf{M}$ such that if unlabeled point $i$ is selected, $\mathbf{M}(i) = 1$, else $\mathbf{M}(i) = 0$. This gives an equivalent formulation:

$$\frac{\tilde{f}(B)}{\sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m}$$

Replace batch size $m$ as:

$$m = \|\mathbf{M}\|_0 \approx \|\mathbf{M}\|_1 = \sum_j \mathbf{M}_j$$

subject to the constraint:

$$M_j \in [0, 1]$$
Optimization Formulation

\[ f(B) = \sum_{j \in B} \rho_j - \lambda_1 \sum_{j \in U_t - B} S(y|x_j, w^{t+1}) - \lambda_2 m \]

Define a binary vector \( \mathbf{M} \) such that if unlabeled point \( i \) is selected, \( \mathbf{M}(i) = 1 \), else \( \mathbf{M}(i) = 0 \). This gives an equivalent formulation:

\[
\max_{\mathbf{M}, m} \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 m \\
\text{subject to the constraint:} \\
M_j \in [0, 1]
\]

Replace batch size \( m \) as:

\[
m = \|\mathbf{M}\|_0 \approx \|\mathbf{M}\|_1 = \sum_j M_j
\]

subject to the constraint:

NP hard
Define a binary vector $M$ such that if unlabeled point $i$ is selected, $M(i) = 1$, else $M(i) = 0$. This gives an equivalent formulation:

$$\max_M \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 \sum_j M_j$$

subject to the constraint:

$$M_j \in [0, 1]$$

Replace batch size $m$ as:

$$m = \|M\|_0 \approx \|M\|_1 = \sum_j M_j$$

$$\max_M \sum_{j \in U_t} \rho_j M_j - \lambda_1 \sum_{j \in U_t} (1 - M_j) S(y|x_j, w^{t+1}) - \lambda_2 \sum_j M_j$$

subject to the constraint: $0 \leq M_j \leq 1$
Algorithm highlights

✓ Solved using the Quasi Newton method

✓ Step size obtained using backtrack line search – guarantees monotonic convergence

✓ Batch size and selection criteria are combined into a single formulation

✓ Solving a single optimization problem yields the batch size as well as the specific points to be selected
Extensions of the framework

Learning from multiple sources of information

Office setting

Context aware dynamic batch selection

Home setting
## Agenda

<table>
<thead>
<tr>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Batch Mode Active Learning (BMAL)</td>
</tr>
<tr>
<td>Existing Approaches</td>
</tr>
<tr>
<td>Optimization Framework for Dynamic BMAL</td>
</tr>
<tr>
<td>Experiments and Results</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
</tbody>
</table>
Setup

• **Datasets**
  - VidTIMIT and MBGC (Multiple Biometric Grand Challenge) biometric datasets
  - Represent challenging real-world conditions

• **Facial Feature Extraction**
  - Discrete Cosine Transform (DCT) followed by PCA

• **Classification Model**
  - Gaussian Mixture Model (GMM)

Parameters $\lambda_1$ and $\lambda_2$ empirically set to 1 in our experiments
Experiment 1: Dynamic Batch Size Selection

- Unlabeled videos presented to the learner
- Proportion of unknown subjects gradually increased from 0% to 100% in steps of 20%
- Learner unaware about the composition of a video stream
- Predicted batch size noted
Results

**Observation**: Predicted batch size increases with increasing proportions of unknown subjects
Experiment 2: Comparison against heuristic BMAL techniques

1. Given unlabeled video stream
2. Compute the batch size using the framework
3. Decide the specific points to be selected using the framework
4. Supply as input to a heuristic BMAL technique
5. Update model and test on a test set
6. Compare
7. Update model and test on a test set
Experiment 2: Comparison against heuristic BMAL techniques

Heuristic methods of Batch Mode Active Learning

- **Random Selection**: From a pool of unlabeled points, select a batch of points at random.

- **Diversity based selection**: Select a batch of points incrementally such that at each step, the distance of the selected point to the already selected batch is maximal. (Brinker, ICML 2003)

- **Uncertainty based ranked selection**: Based on the current classifier, select the top m uncertain points.
Observation: The accuracy on the test set grows fastest for the proposed framework
Agenda

- Batch Mode Active Learning (BMAL)
- Existing Approaches
- Optimization Framework for Dynamic BMAL
- Experiments and Results
- Conclusion
Conclusions

- A generic optimization framework for dynamic batch selection in BMAL
- Combining the batch size and selection criteria in a single formulation
- Solving a single problem helps us identify the batch size and the points to be selected

Future Work

- Intelligently decide the parameters in the framework
- Improve the scalability of the framework
Thank You !!!

Questions ??..