Active Learning for Image Classification
Automation of Biological Research
Final Report

Ruijian Wang, Jianhe Luo

December 15, 2015

1 Introduction

Image classification is not a new research area, but it is still a major challenge because of the nature of image. One challenge is a lot of labeled training examples required. It is both time and cost consuming to label a good set of images, especially for medical image. Active learning has attracted a lot of attention since it can select the 'useful' observation by some strategy. So that manual effort can focus on labeling the most helpful examples to build a machine learning model. In this paper, we tried to select the informational examples from different views and apply them in different situations.

2 Background and Related Work

Although many researches have been devoted to image classification, however, only a few of them are related to the medical domain. More and more attention have been paid to medical image classification**citation**. Many machine learning algorithms have been applied to medical image classification, including the large margin classifiers, decision trees and neural networks. Among all of the algorithms, the support vector machine(**SVM**) and kernel logistic regression(**KLR**) appeared to be the most effective methods[2].

2.1 Data Source

The data set is composed of 5120 unlabeled images. All the images are represented by a 26-dimension vector, which are already being processed. There are 8 classes for those images and are represented by 1,2,3,4,5,6,7,8.

2.2 Pool-based learning setup

Pool-based learning task is very common for active learning. Start with a small amount of training observations that are selected randomly, we get a seed set and

a naive machine learning model, based on the naive machine learning model and candidate observation pool, we select new observation(s) along with the query result, adding to the training set to retrain the model. The selection process is iterative such that after each iteration of model's prediction and relevant parameters, we select new observation to improve the model. Every time we call the oracle(query the label from expert), we will calculate the loss value for the entire data set until all queries are consumed. Compare the selection strategy with random selector by plotting the loss value graph.

2.3 Stream-based learning setup

2.4 Difference between pool-based and stream-based

For the pool based learning process, we see all the unlabeled data referred as the active pool. The advantage by seeing all the data is we can rank the data or use the structure of the data [1], and then select the useful observation(s). While for the stream based learning process, the restriction is we only know the current input data, and all the previous data if you cached them. That prevents us from using ranking techniques and we need to make decision based on the current model and the input observation.

3 Methods

3.1 Data distribution

We can get a overview of the distribution of all the labels. From the following graph we can see almost each label has equal amount of observations. Thus we can make a assumption that when we randomly select the seed set data, each label will be covered.

3.2 Pool-based learning

3.2.1 Random learner

The random learner is designed to be compared with some selection strategy. The logic for random learner is quite simple.

Algorithm 1 Random selection for pool based learning

Randomly choose 50 observations as seed set
Train the initial SVM models with seed set
while $oracle_num \leq 462$ **do** Randomly select an observation.
    Add it to the training data and remove it from candidate pool.
    Retrain the 8 binary classifiers.
end while
Plot the loss value

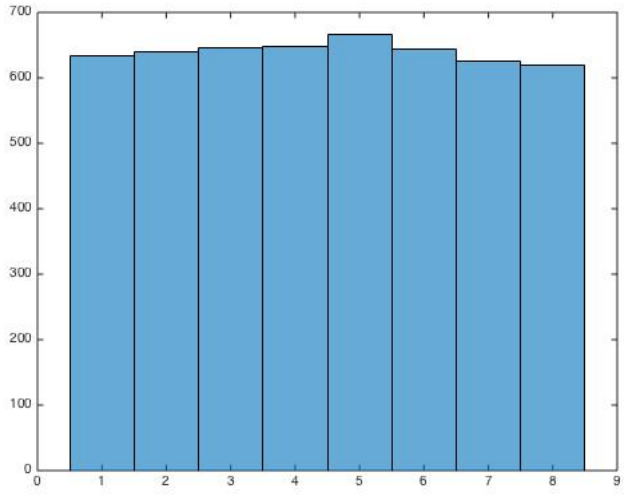

Figure 1: All data's label distribution

3.2.2 Query from committee

The Expected Loss Optimization [5] is a straight-forward strategy. Since we got 8 labels from all the data and we will use 8 binary classifiers to help us select the informative observation, each classifier is a binary **SVM** classification model. For the first binary classifier, we will set data with label 1 as positive observations and choose equal number of data with other labels as negative data. Similarly, for the second and following classifiers, we will set label 2,3...8 as positive observations accordingly, and set observations with other labels as negative observations. By selecting the seed set, we got the initial training data set and get 8 binary classifiers.

Based on the 8 binary classifiers, we can get a 1×8 vector for each observation, after processed, the value in the vector with value 1 means the classifier of this index make a positive prediction. The sum of the vector is the total number of binary classifiers predict positively. Select the observation with most positive results. Since the sum result ranges from 0 to 9, and we have 5120 observations, it will be very likely to have more than one observation with same number of positive results. Among all the observations with most positive predictions, we can apply another selection strategy. For simplicity, we just randomly pick up one of them and add it to the training data.

After each iteration, calculate the loss value and plot the loss value after finishing all the iterations.

Algorithm 2 Query from committee pool based learning

Randomly choose 50 observations as seed set
Train the initial SVM models with seed set
while $oracle_num \leq 462$ **do**
    Predict all the observations and get a $N \times 8$ matrix.
    Sum each observation's positive prediction number.
    Rank based on the number of positive prediction number
    Choose the observation with most positive predictions
    Add it to the training data and remove it from candidate pool.
    Retrain the 8 binary classifiers.
end while
Plot the loss value

3.2.3 Entropy Measure

The entropy measure[4] strategy is select the observation with largest entropy. Unlike the query from committee strategy, we use multi-class Naive Bayes model for this selection strategy since we need the posterior probability for each label. After training the initial model by seed set, estimate the posterior probability of each label for every observation, we can get entropy for each observation, and rank by the entropy.

Algorithm 3 Entropy Measure for pool based learning

Randomly choose 50 observations as seed set
Train the initial multi-calss Naive Bayes model with seed set
while $oracle_num \leq 462$ **do**
    Predict all the observations and get a $N \times 8$ matrix.
    Estimate the posterior probability of each label for each observation.
    Calculate the entropy based on posterior probability.
    Sum each observation's entropy
    Rank based on the sum of entropy.
    Choose the observation with largest entropy.
    Add it to the training data and remove it from candidate pool.
    Retrain the multi-class Naive Bayes classifier.
end while
Plot the loss value

3.3 Stream-based learning

3.3.1 Random learner

There are two kinds of random learner, first random learner only contains a multi-class SVM model, and second random learner only contains 8 binary classifiers.

Algorithm 4 First random learner for stream-based learning

```
Randomly choose 10 observations as seed set
Train the initial multi-calss SVM model with seed set
for i = 1 : oracle_nums do
    Get next image.
    Add the new image to training data set.
    Retrain the model.
    Test error.
end for
Plot error values
```

Algorithm 5 Second random learner for stream-based learning

```
Randomly choose 10 observations as seed set
Construct the training data and train 8 initial binary SVM models
for i = 1 : oracle_nums do
    Get next image.
    Add the new image to training data set.
    Retrain the model.
    Test error.
end for
Plot error values
```

3.3.2 Query from committee

Based on our mid-term report, we make some improvement for our model. Previously, there are only 8 binary classifiers and all prediction came from the 8 binary classifiers predictions.

We add another multi-class classifier, use its predictions as the main result and assisted by the 8 binary classifiers' predictions.

Algorithm 6 Query from committee stream-based learning

Randomly choose 10 observations as seed set
Construct the training data and train 8 initial binary SVM models and one multi-class model.
for **do**$i = 1 : 5120$
  Get prediction vector from binary classifiers referred as binary predictions.
  Get prediction from the multi-class classifier referred as overall prediction.
  if there are more than one positive predictions in binary predictions **then**
    Choose the overall prediction as final result
  else if there are no positive predictions in binary predictions **then**
    Call oracle, add this observation to training data.
    Get Testset Error
    Choose the overall prediction as final result
  else ▷ only one positive
    Choose the positive binary prediction as final result.
  end if
end for
return number of oracle calls, testset error

4 Experiments and Results

4.1 Pool-based learning

For the pool based active learning task, we first randomly select 50 observations as seed set, and at each iteration, select the most useful observation and repeat 462 times since only 512 oracle calls are allowed.

4.1.1 Query from committee selection strategy

To test the assumption we made at 3.1, we plot the label distribution of seed set.

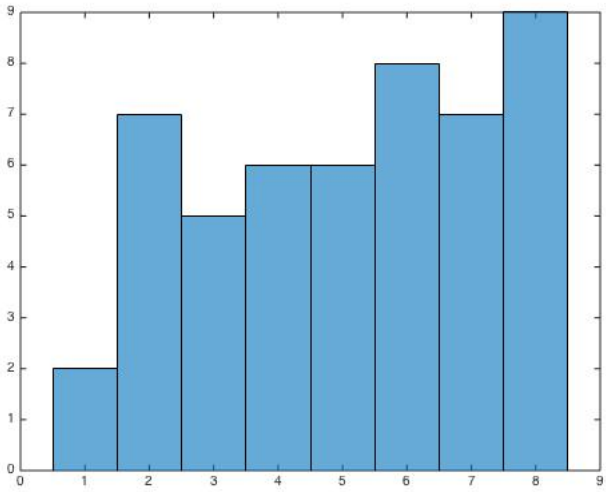

Figure 2: Seed set's label distribution

Run two different active learning process, query form committee and random selection, plot the result function.

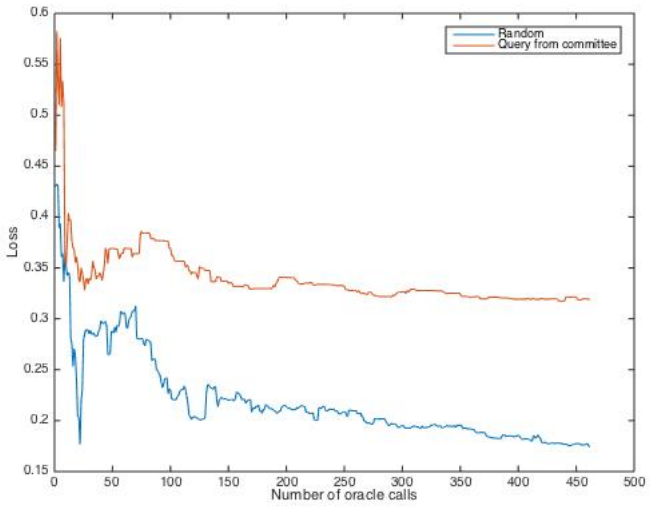

Figure 3: Loss Optimize and Random Learner

From the graph we can see the loss optimize selection strategy does not perform

well. Although it converges at last but the loss value is almost twice as random learner.
To test my guess, I even select the observation the other way around, always select the one with minimum positive predictions, it turns out it also converges at similar loss value.

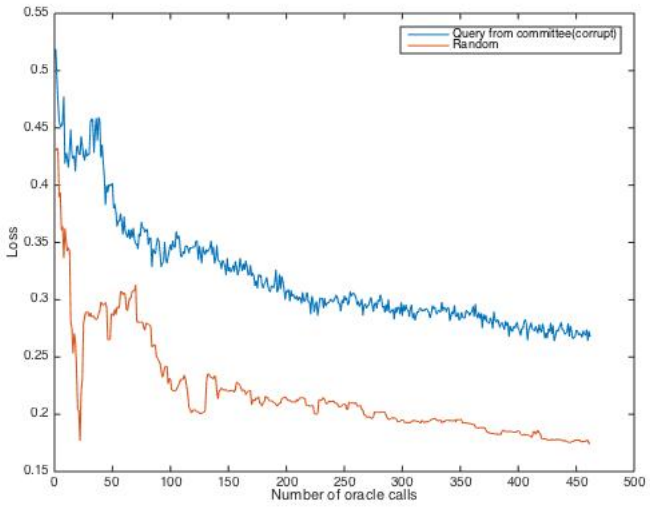

Figure 4: Corrupt Loss Optimize and Random Learner

From the graph we can see, the loss value still goes down but it will keep fluctuating.Those two graphs proves that this strategy is not a good active learning strategy.

4.1.2 Entropy Measure selection strategy

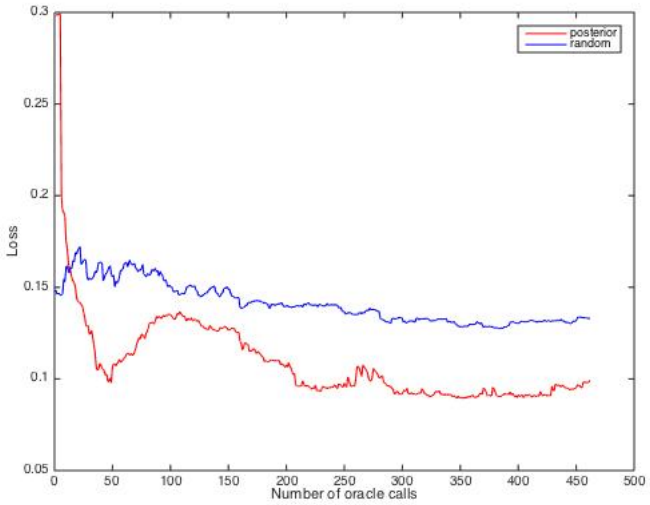

Figure 5: Loss value of entropy measure and random learner

From the graph we can see, the random learner has a lower loss value at the very beginning, however, the seed set is selected randomly, the beginning loss value means nothing. After around 10 iteration, the query from committee selection strategy has reached a lower loss value and keep reducing quickly. And finally reached around 0.1 loss value, lower than the random learner loss value, which is around 0.14.

4.2 Stream-based learning

For stream based learning, after finish the initial machine learning model, execute 5120 times for-loop, record the number of oracle calls and select the useful observation for each loop. Each time call the oracle, calculate the test set error. After execution, run random learners based on number of oracle calls, calculate the test set error, plot the test set error to compare performance.

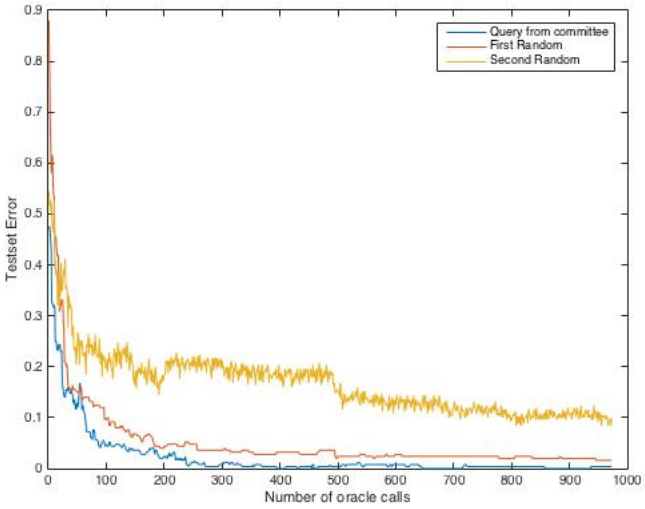

Figure 6: Stream based learner and two random learner

From the graph we can see our active learning strategy do achieve better performance than two random learner. The first random learner, which only contains 8 binary classifiers has higher error. It is consistent with the pool-based query from committee results.

5 Discussion

The result of query from committee in pool based is not a good selection strategy, 4.2 shows that the second random leaner has significantly lower error value, and the second one only contains one multi-class classifier, which proves that our one-versus-all ensemble classifiers are not a good selection strategy.

Also the steam based task converges faster than pool based task, it usually takes 100 iterations for the former one and 300 iterations for the latter one. It is also a interesting result. One possible reason might be pool based learning task selects the observation without replacement. And iteration of stream based might select duplicated observation. Some already selected observation might still be helpful to improve the model, while in pool based non-replacement selection strategy, it cannot select the same observation. So the model can only choose the second most helpful observation to improve the model, which causes the model converges slower.

6 Conclusions and Future Work

In this paper, we proposed different active learning strategy for both pool-based learning and stream-based learning. Through the results of those active learning model, it proves that active learning can reduce notable number of required training examples and achieve better performance at the same time.

For the future, we may try more complex ensemble model since the pool-based query from committee strategy does not work well and one of the potential problems might be that we just use linear number of binary classifiers and also the way we construct training data after seed set are selected. Because the one-versus-one approach [3] shows good classification performance. In the future, we might change the one versus all to one versus one method to test the performance.

Also there are too many oracle calls in steam based task, we tried Linear and Quadratic Discriminant Analysis, this model will call almost twice times oracle as SVM. It forces us to change the strategy to call oracles for future improvement.

References

[1] Sanjoy Dasgupta and Daniel Hsu. Hierarchical sampling for active learning. In *Proceedings of the 25th international conference on Machine learning*, pages 208–215. ACM, 2008.

[2] Steven CH Hoi, Rong Jin, Jianke Zhu, and Michael R Lyu. Batch mode active learning and its application to medical image classification. In *Proceedings of the 23rd international conference on Machine learning*, pages 417–424. ACM, 2006.

[3] Chih-Wei Hsu and Chih-Jen Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415–425, 2002.

[4] Ajay J Joshi, Fatih Porikli, and Nikolaos Papanikolopoulos. Multi-class active learning for image classification. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2372–2379. IEEE, 2009.

[5] Bo Long, Olivier Chapelle, Ya Zhang, Yi Chang, Zhaohui Zheng, and Belle Tseng. Active learning for ranking through expected loss optimization. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 267–274. ACM, 2010.