Interactive Visualization based Active Learning

Mohammadreza Babaee[1], Stefanos Tsoukalas[1],
Gerhard Rigoll[1], and Mihai Datcu[2]

[1] Institute for Human-Machine Communication, Technische Universität München
{reza.babaee,rigoll}@tum.de, s.tsoukalas@mytum.de
[2] The Remote Sensing Technology Institute (IMF), German Aerospace Center
mihai.datcu@dlr.de

1 Introduction

Active learning aims to label the most informative data points in order to minimize the cost of labeling [1]. In this work, we introduce a human based approach, namely First Certain Wrong Labeled (FCWL) to select points for labeling. It is based on a ranked list of predictions ordered by confidence, from which the user selects the highest ranking incorrect prediction. The experimental results show the improvement in performance of this method compared to others.

2 Approach

We consider the set of n feature vectors $X = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n\}$ of dimension d and the corresponding class labels $Y = \{y_1, ..., y_n\}$ with a total number of k classes. Additionally let $\mathcal{L}$ denote the set of samples, whose correct label is available to the algorithm.

The proposed algorithm is based on the idea of selecting samples for labeling, that introduce the highest change into the model of a trained classifier. This is achieved by letting the algorithm predict labels during each iteration and correcting a label, that the algorithm is certain about, but predicts incorrectly. As the measure of certainty we use the extension of the margin as suggested in the $\text{SVM}_{\text{Active}}$ algorithm [2] to a multiclass classifier. Given the current weight matrix $W = [\mathbf{w}_1, ..., \mathbf{w}_k]$ of the classifier, we define the margin μ_i of each sample as

$$\mu_i = \max_{l=1,\dots,k} \mathbf{w}_l^{\mathrm{T}} \mathbf{x}_i. \tag{1}$$

and the predicted label $\tilde{y}_i$ of each sample as

$$\tilde{y}_i = \arg\max_{l=1,\dots,k} \mathbf{w}_l^{\mathrm{T}} \mathbf{x}_i. \tag{2}$$

Let $I_{\tilde{l};i}$ denote the image of the unlabeled sample $\mathbf{x}_{\tilde{l};i} \in X \backslash \mathcal{L}$, predicted with label $\tilde{l}$. Then the algorithm during each iteration arranges these images in a table with increasing margins for each class, i.e.

$$i \geq j \Leftrightarrow \mu_{\tilde{l};i} \geq \mu_{\tilde{l};j} \tag{3}$$

as suggested in table 1 and lets the user select the first sample $\mathbf{x}_{\tilde{l};m}$ in a class that is labeled incorrectly, i.e.

$$y_{\tilde{l};m} \neq \tilde{y}_{\tilde{l};m} \text{ and } y_{\tilde{l};i} = \tilde{y}_{\tilde{l};i} \text{ for } i = 1, .., m-1 \tag{4}$$

and relabel it. Since the samples are sorted with decreasing certainty, we can expect to achieve a big correction in the model by selecting the first incorrectly labeled sample. For the next iteration the relabeled sample and the correctly predicted samples before it in the corresponding class are added to the set of labeled samples

$$\mathcal{L} = \mathcal{L} \cup \left\{ \mathbf{x}_{\tilde{l};1}, ..., \mathbf{x}_{\tilde{l};m} \right\}. \tag{5}$$

This is repeated for the desired number of iterations. Since the algorithm lets the user select a sample in the wrong class in each iteration, we call it FCWL. The algorithm is summarized in Algorithm 1.

$\tilde{l} = 1$	$\tilde{l} = 2$	$\cdots$	$\tilde{l} = k$
$I_{\tilde{1};1}$	$I_{\tilde{2};1}$		$I_{\tilde{k};1}$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
$I_{\tilde{1};n_1}$	$I_{\tilde{2};n_2}$		$I_{\tilde{k};n_k}$

Table 1: Predicted samples images presented to the user during each iteration

Algorithm 1 FCWL - Active Learning with incorrect label correction by user

Input: training points X and labels Y
initial set of labeled samples $\mathcal{L}_0$
total number of iterations p

Output: new set of labeled samples $\mathcal{L}$

Algorithm:

for $t = 0, 1, 2, ..., p$ **do**
  Obtain W_t by training SVM classifier on $\mathcal{L}_t$
  Compute margin μ_i for each sample according to (1)
  Predict label $\tilde{y}_i$ for each sample according to (2)
  Present samples to user according to table 1 and equation (3)
  Let user relabel first sample $\mathbf{x}_{\tilde{l};m}$ with incorrect predicted label from one class and set $\mathcal{L}_{t+1} = \mathcal{L}_t \cup \{\mathbf{x}_{\tilde{l};1}, ..., \mathbf{x}_{\tilde{l};m}\}$
end for
return $\mathcal{L}_t$

3 Experiments

3.1 Datasets

The datasets used in our experiments are: 1) Corel dataset; 2) Caltech dataset; 3) Handwritten Digits; 4) Yale Faces;

- **Corel dataset** contains 1500 images in 15 different groups, where each group contains 100 images represented by the Bag of word model of local SIFT [3] local descriptors and 200 visual words. For each experiment we chose a random subset of 500 images as training set and a different random subset of 500 images as a test set.
- **Caltech dataset** contains 3379 images in 10 different groups with a different number of images for each group. For each experiment we chose a random subset of 800 images as a training set and a different random subset of 800 images as a test set.
- **UPS handwritten digits**[3] contains 8-bit gray-scale images of the size 16x16 of the digits $0 - 9$, with 1100 images per class. For each experiment we chose a random subset of 1000 images as a training set and a different random subset of 1000 images as a test set.
- **Yale faces dataset**[4] contains 165 gray-scale images of 15 individuals. For each experiment we chose a random subset of 80 images as a training set and a different random subset of 80 images as a test set.

3.2 Methods

In addition to the proposed algorithm we applied the following active learning methods on the datasets:

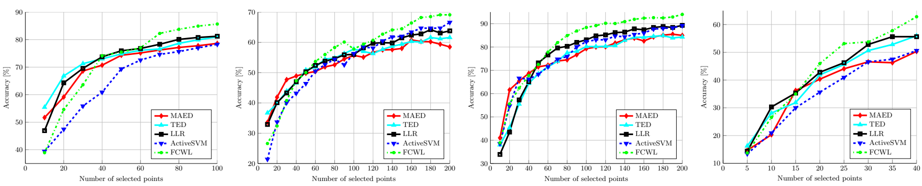
- TED [4], which selects training points by minimizing the covariance of the prediction error of a least squares classifier
- MAED [5], which extends the TED algorithm with a manifold adaptive kernel, in order to incorporate the manifold structure into the selection process.
- $\text{LLR}_{\text{Active}}$ [6], which minimizes the error of reconstructing the whole dataset based on the selected samples and the matrix describing the locally linear embedding.
- $\text{SVM}_{\text{Active}}$ [2], which iteratively adds points closest to the boundary of an SVM.

3.3 Results and Discussion

For each dataset we repeated the experiment 4 times and computed the average results. A screenshot of the FCWL algorithm after 10 iterations on the Corel dataset is presented in figure 1. As can be seen from this figure, the algorithm already predicts many labels correctly and can therefore be trained faster by adding multiple labeled samples during each iteration. The classification results are presented in figure 2. The plots show, that while the proposed algorithm does worse at the beginning, it outperforms the other algorithms for all datasets as the number of points increases. This difference in behavior between label-based algorithms and experimental-design based algorithms is often observed in experiments.

[3] http://www.cs.toronto.edu/ roweis/data.html
[4] http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html

Fig. 1: Screenshot of the images presented by the algorithm for the Corel SIFT dataset. The first row contains images representing each class. The sample images are arranged in the columns, which correspond to the predicted class, with decreasing margin.



(a) (b) (c) (d)

Fig. 2: Classification accuracy for different Active Learning algorithms; (a) Caltech SIFT; (b) Corel SIFT; (c) Handwritten Digits; (d) Yale Faces.

References

1. Settles, B.: Active learning literature survey. Technical report (2010)
2. Tong, S., Chang, E.: Support vector machine active learning for image retrieval. In: Proceedings of the Ninth ACM International Conference on Multimedia. MULTIMEDIA '01, ACM (2001) 107–118
3. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**(2) (2004) 91–110
4. Yu, K., Bi, J., Tresp, V.: Active learning via transductive experimental design. In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06, New York, NY, USA, ACM (2006) 1081–1088
5. Cai, D., He, X.: Manifold adaptive experimental design for text categorization. Knowledge and Data Engineering, IEEE Transactions on **24**(4) (April 2012) 707–719
6. Zhang, L., Chen, C., Bu, J., Cai, D., He, X., Huang, T.S.: Active learning based on locally linear reconstruction. Pattern Analysis and Machine Intelligence, IEEE Transactions on **33**(10) (2011) 2026–2038