Active Learning based on Transfer Learning Techniques for Image Classification

UNIVERSITATEA "1 DECEMBRIE 1918" ALBA IULIA

Daniela Onita, Adriana Birlutiu*

Faculty of Science - 1 Decembrie 1918 University of Alba Iulia
Gabriel Bethlen Nr.5, 510009, Alba Iulia - Romania
*Corresponding author, e-mail: *adriana.birlutiu@uab.ro*

EXECUTIVE AGENCY FOR HIGHER EDUCATION, RESEARCH, DEVELOPMENT AND INNOVATION FUNDING

Abstract

In many imaging tasks only an expert can annotate the data. Though domain experts are available, their labor is expensive and we would like to avoid querying them whenever possible. Our task is to make use of our resources as efficient as possible for a learning task. There are various ways of working in cases of labelled data shortage. This type of learning problems can be approached with Active and Transfer Learning techniques. Active Learning and Transfer Learning have demonstrated their efficiency and ability to train accurate models with significantly reduced amount of training data in many real-life applications. In this paper we investigate the combination of Active and Transfer Learning for building an efficient algorithm for image classification. The experimental results show that by combining active and transfer learning, we can learn faster with fewer labels on a target domain than by random selection.

1. Active Transfer Learning

The idea behind Active Learning (AL) [1, 5] is that by optimal selection of the training points a better performance can be achieved instead of random selection. AL techniques can be applied in situations in which labeling points is difficult, time-consuming, and expensive.

1.1 Uncertainty Sampling Criterion

Uncertainty sampling criterion [1] is an AL strategy in which an active learner chooses for labeling the example for which the predictions are most uncertain. One choice for measuring the uncertainty of the predictions is Shannon entropy:

$$\text{Uncertainty(x)} = -\sum_{y} p(y|x) \log p(y|x). \tag{1}$$

where x represents the point that is to be labelled and y represents the possible label of x.

1.2 Active Transfer Criterion

Transfer Learning (TL) [2] is a technique used for transferring knowledge from a source task to a target task. TL techniques can be applied in settings in which data is available from multiple domains.

We propose a criterion for AL, which we call Active Transfer (AT), specifically design for the AL and TL settings. The main idea behind the AT criterion is to exploit learning with multiple data sets and use the learned models of other data sets when determining how informative a new data point is.

Inspired by [3], we measure the disagreement by taking the average prediction of the entire committee and compute the average Kullback-Leibler (KL) divergence of the individual predictions from the average:

$$\text{AT(x)} = \sum_{m=1}^{M} \frac{1}{M} \text{KL}[\overline{p}(\cdot|x) \| p_m(\cdot|x)]\,, \tag{2}$$

where

$$p_m(y|x) \equiv p(y|x, M_m)\,. \tag{3}$$

and $M_1, ..., M_M$ represents the data sets specific for each task, and $\overline{p}(\cdot|a)$ the average predictive probability of the entire committee.

The KL divergence for discrete probabilities is defined as

$$\text{KL}[p_1(\cdot|x) \| p_2(\cdot|x)] = \sum_{c} p_1(y|x) \log \frac{p_1(y|x)}{p_2(y|x)}. \tag{4}$$

The KL divergence can be seen as a distance between probabilities, where we abused the notion of distance, since the KL-divergence is not symmetric, i.e., $\text{KL}[p_1 \| p_2] \neq \text{KL}[p_2 \| p_1]$. This drawback of the KL-divergence can be overcome by considering a symmetric measure, for example, $\text{KL}[p_1 \| p_2] + \text{KL}[p_2 \| p_1]$.

2. Experimental evaluation

1. Two data sets were used in the experimental evaluation.
   - Breast Cancer Histopathological Images (BreakHis) [4] consists of microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors (2495 bening and 5276 malign).
   - Porcelain ware image data set, contains 346 images (200 defective and 146 correct), collected from an industrial partner producing porcelain ware.

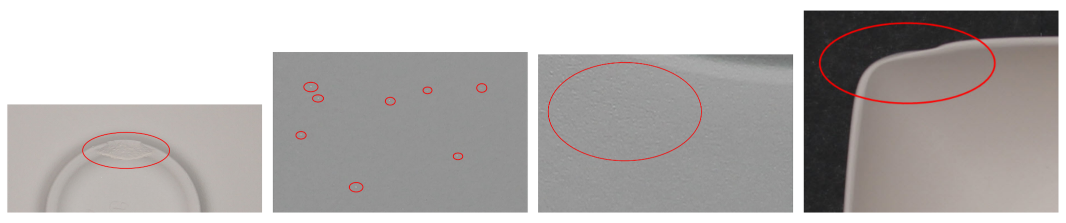

Figure 1: *Different types of defects in porcelain image data set. From left to right: deterioration after pressing, bumps, texture defects, margin deformation.*

2. The images were preprocessed as follows:
   - Converting each image to gray scale.
   - Resizing to 28x28 (784-dimensional feature vector).
   - Centering of data around zero.
   - Whitening the data.
3. Comparison of different learning algorithms for the two data sets. The mean accuracy $\pm$ standard deviation is shown.

Algorithm	BreakHis data set	Porcelain ware data set
SVM	**0.66 $\pm$ 0.00**	**0.84 $\pm$ 0.05**
Logistic Regression	0.66 $\pm$ 0.00	0.63 $\pm$ 0.08
LDA	0.57 $\pm$ 0.00	0.72 $\pm$ 0.05
Decision Tree	0.59 $\pm$ 0.04	0.55 $\pm$ 0.10
NB	0.56 $\pm$ 0.04	0.63 $\pm$ 0.06
Random Forest	0.59 $\pm$ 0.01	0.75 $\pm$ 0.08

4. Comparison of accuracies: Uncertainty sampling selection vs. random selection and Active transfer selection vs. Uncertainty sampling selection (Figure 2):
   - The training data was used as a pool out of which points were selected for labeling either randomly or actively.
   - After selection of a point, either active or random, the point was added to the training data and deleted from unlabeled data.
   - The model was retrained on the new training set and predictions were made on the validation set (50 retraining).
   - Averaged results over 20 splittings of data into training, unlabeled and validation sets.

3. Results

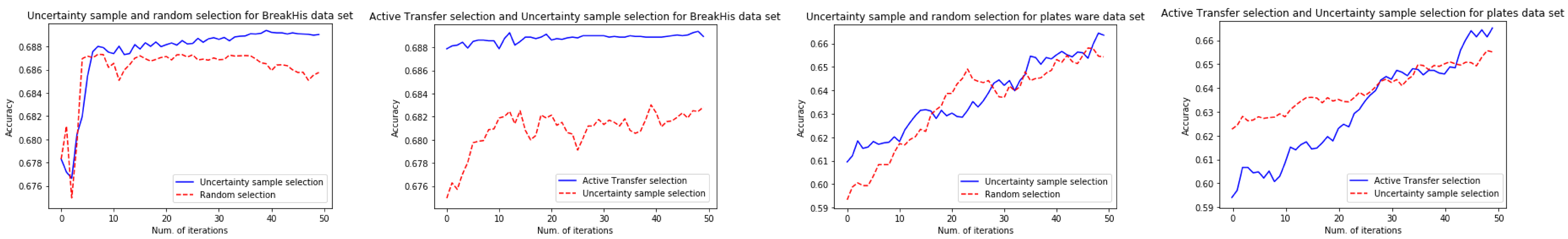

Figure 2: *Uncertainty sampling selection vs. random selection and Active transfer selection vs. Uncertainty sampling selection.*

4. Conclusions

- The motivation of this work was to make use of the available resources as efficient as possible.
- We proposed the Active Transfer criterion which makes use of the models learned on similar tasks to select for labelling those points that give most of the information about the current task.
- The experimental results show that by combining active and transfer learning, we can learn faster and with fewer labels on a target domain than by random selection.

References

[1] B. Settles. Active learning. Morgan Claypool, 2012.

[2] S.J. Pan, Q. Yang. A survey on transfer learning. IEEE Trans. Knowle. Data Eng.2010.

[3] A. McCallum, K. Nigam. Employing EM and pool-based active learning for text classification. Internation Conference on Machine Learning, pp. 350–358, 1998.

[4] F. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte. A Dataset for Breast Cancer Histopathological Image Classification, IEEE Trans. on Biomed. Engineering, 63(7), 2016.

[5] X. Wang, Active Transfer Learning, PhD Thesis. CMU, 2016.

ESANN, 26th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 25-27 April 2018