ESANN 2018 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2018, i6doc.com publ., ISBN 978-287587047-6.
Available from http://www.i6doc.com/en/.

Active Learning based on Transfer Learning Techniques for Image Classification

Daniela Onita and Adriana Birlutiu [*]

Faculty of Science - 1 Decembrie 1918 University of Alba Iulia
Gabriel Bethlen Nr.5, 510009, Alba Iulia - Romania
adriana.birlutiu@uab.ro

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 Introduction

In today's world utilizing huge datasets for solving problems with Machine Learning is natural and many recent algorithms, such as Deep Learning, require tones of data to be trained properly. However, in many applications even though it is easy to find sources of unlabelled data, annotations are still hard and expensive to obtain. This is the case in many medical imaging tasks and other domains that use computer vision techniques. In these domains, training data are generally difficult to acquire because the manual labeling is a complex and time-consuming activity.

The motivation of this work is to make use of our resources as efficient as possible for a image classification. Working in cases of labelled data shortage can be approached with Active Learning (AL) and Transfer Learning (TL) techniques. AL and TL have demonstrated their efficiency and ability to train accurate models with significantly reduced amount of training data in many real-life applications. Even though these methods are widely used, there are still some limitations of the current methods and very few works investigate the combination of AL and TL.

In this paper we investigate the combination of AL and TL for efficient learning in case of data shortage for image classification. We investigate a criterion

[*] This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS/CCCDI UEFISCDI, project number PN-III-P2-2.1-BG-2016-0333, within PNCDI III.

for AL which makes use of data from other learning scenarios similar to the way transfer learning techniques are working.

There are a few recent works that investigate the combination of active and transfer learning in different learning settings [1, 2, 8, 9]. In [1] it is proposed an alternative for the standard criteria in active learning which actively chooses queries by making use of the available preference data from other users in a preference learning setting. In [2] a hierarchical Bayesian model for active transfer learning for activity recognition is proposed. In [8] active transfer learning under model shift is being investigated.

2 Active Transfer Learning

Active Learning [7] represents a range of techniques that can be applied in situations in which labeling points is difficult, time-consuming, and expensive. The idea behind AL is that by optimal selection of the training points a better performance can be achieved instead of random selection.

A tendency is to improve the performance of the AL methods by combining them with heuristics designed either for the context in which they are applied or by the models they use, e.g., making use of the unlabeled data available, exploiting the clusters in the data, diversifying the set of hypotheses, or adapting the AL to other learning techniques such as Gaussian processes.

2.1 Uncertainty Sampling Criterion

Uncertainty sampling criterion [7] is an AL strategy in which an active learner chooses for labeling the example for which the model's predictions are most uncertain. The uncertainty of the predictions can be measured, for example, using Shannon entropy

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

where x represents the point that is to be labelled and y represents the possible label of x. For a binary classifier this strategy reduces to querying points whose prediction probabilities are close to 0.5. Intuitively this strategy aims at finding as fast as possible the decision boundary since this is indicated by the regions where the model is most uncertain.

2.2 Active Transfer Criterion

Transfer Learning [5] is a technique used for transferring knowledge from a source task to a target task. It is inspired by the research on transfer of learning in psychology, more specifically on the dependency of human learning on prior experience. The psychological theory of transfer of learning implies the similarity between tasks so TL algorithms are used when training data for the target task is similar, but not identical with that of the source task. In the context of learning algorithms, TL can be implemented by transplanting the learned features and parameters from one algorithm to initialize another.

We propose here a criterion for AL, which we call Active Transfer (AT), specifically design to make use of 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 the knowledge acquired with a new data point.

We will use the following notation for the predictive probability corresponding to other models

$$p_m(y|x) \equiv p(y|x, M_m)\,. \tag{2}$$

where $M_1, ..., M_M$ represents the data sets specific for each task. Inspired by [4], 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:

$$\mathrm{AT(x)} = \sum_{m=1}^{M} \frac{1}{M}\,\mathrm{KL}[\,\overline{p}(\cdot|x)\|p_m(\cdot|x)\,]\,, \tag{3}$$

with $\overline{p}(\cdot|a)$ the average predictive probability of the entire committee.

The KL divergence for discrete probabilities is defined as

$$\mathrm{KL}[p_1(\cdot|x)\|p_2(\cdot|x)] = \sum_c p_1(y|x)\mathrm{log}\frac{p_1(y|x)}{p_2(y|x)}\,. \tag{4}$$

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., $\mathrm{KL}[p_1\|p_2] \neq \mathrm{KL}[p_2\|p_1]$. This drawback of the KL-divergence can be overcome by considering a symmetric measure, for example, $\mathrm{KL}[p_1\|p_2] + \mathrm{KL}[p_2\|p_1]$.

3 Experimental Evaluation

We set two goals for the experimental evaluation: *i)* to test whether optimally selecting data for labeling using an active strategy achieves higher accuracy than random selection; and *ii)* to test whether optimally selecting data for labeling using the Active Transfer criterion achieves higher accuracy than using Uncertainty sampling criterion.

3.1 Data Sets and Data Preprocessing

Two data sets were used in the experimental evaluation.

Breast Cancer Histopathological Images Classification (BreakHis) [6] is a data set composed of microscopic images of breast tumor tissue collected from 82 patients using different magnifying factors: 40X (1994 images), 100X (2081 images), 200X (2013 images), and 400X (1683 images).

As second data set we used a set containing 346 porcelain ware images, out of which 200 were defective and 146 correct. This data was collected from an industrial partner producing porcelain ware. Figure 1 shows samples of different types of defects in the porcelain image data set.

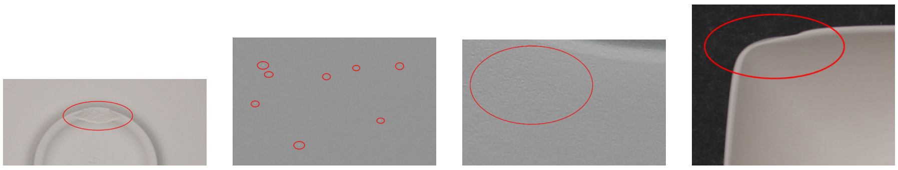

Fig. 1: Different types of defects in porcelain image data set. From left to right: deterioration after pressing, bumps, texture defects, margin deformation.

Each image was converted to gray scale and it was resized to 28x28, thus one image is described by a 784-dimensional feature vector. Furthermore, the images were preprocessed as follows. First, the centering of data around zero was performed: for each image patch, the mean pixel value was computed and subtracted from the data. Second, the whitening of data was performed: i) the data covariance matrix was computed and the SVD factorization of the matrix was obtained; ii) data was decorrelated by rotating and reducing the dimension; iii) the decorrelated data was devided by the eigenvalues.

3.2 Experimental Protocol and Results

The experimental evaluation was performed using Python programming language, and in particular, the Sklearn library [3].

In order to meet the first goal, i.e., to test whether optimally selecting data for labeling using an actively strategy achieves higher accuracy than random selection, we first tested several machine learning algorithms for random selection. The following algorithms were compared: Logistic Regression (tolerance = 0.0001, C parameter = 1.0), Linear Discriminant Analysis, Decision Tree, Naive Bayes, Random Forest (number of trees = 10) and SVM (linear kernel, C parameter=0.1). The results are presented in Table 1. We used accuracy (mean ± standard deviation) as a measure of performance. SVM perform best, thus we use it for the next experiments.

Table 1: Comparison of different learning algorithms dor the two data sets: BreakHis and Porcelain. The mean accuracy ± standard deviation is shown.

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

Next, we compared active selection of training points to random selection. 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. The results were averaged over 20 splittings of data into training, unlabeled and validation sets. All algorithms were learned from 50 randomly and actively selected data points.

Figure 2 compares the accuracy obtained with random and active selection using the Uncertainty sampling criterion for the two data sets used in the experimental evaluation. The results obtained using an active selection are better than the results obtained using random selection of training points.

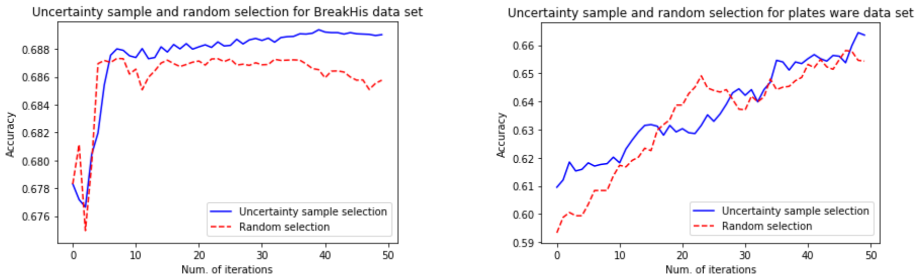

Fig. 2: Comparison of accuracies obtained with random versus active selection of training points. Uncertainty Sampling criterion was used for actively selecting training points. Left: BreakHis data set. Right: porcelain ware data set.

For the uncertainty sampling criterion, because we have a binary classifier, the point for which the prediction probabilities are closest to 0.5 was chosen. For active transfer criterion, we choose one of data set as target. This will be trained using a pre-trained model on a data set formed by the rest of data sets corresponding to different magnification factors for Breast cancer data set or different porcelain models for the porcelain ware data set. The point which was selected was the one on which the other models disagree the most. Figure 3 compares the accuracy obtained with active learning using two criteria of actively selecting points to label: uncertainty sampling criterion and active transfer criterion. The plot shows that the active transfer criterion improves the performance in some cases.

We see from the experiments that the accuracy obtained using random selection is lower than using AL strategies for both data sets. The results in the figures above indicate that the Active Transfer criterion improves performance compared to the Uncertainty Sampling selection strategy.

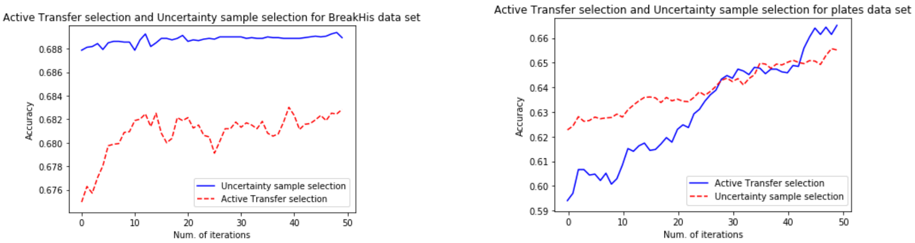

Fig. 3: Comparison of accuracies obtained with two strategies of actively selecting training points: Active Transfer criterion and Uncertainty Sampling. Left: BreakHis data set. Right: porcelain ware data set.

4 Conclusions and Future Work

This work investigated how to obtain an efficient algorithm that can classify images from small data sets by combining Active and Transfer Learning. 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 with fewer labels on a target domain than by random selection

References

[1] Birlutiu A., Groot P., Heskes T. (2013) Efficiently learning the preferences of people. Machine Learning Journal, 90 (1), pp.1-28, Springer, ISSN: 0885-6125.

[2] Diethe, T., Twomey, N., and Flach, P. Active transfer learning for activity recognition. ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, 2016.

[3] Pedregosa, F., et al. Scikit-learn: Machine Learning in Python - ACM Digital Library, 2011.

[4] McCallum, A., Nigam, K. Employing EM and pool-based active learning for text classification. In Proceedings of the 15th international conference on machine learning (pp. 350−358), 1998.

[5] Pan S.J., Yang Q. (2010) A survey on transfer learning. IEEE Trans. Knowle. Data Eng. 2010;vol. 22, pp. 13451359.

[6] Spanhol, F., Oliveira, L. S., Petitjean, C., Heutte, L., A Dataset for Breast Cancer Histopathological Image Classification, IEEE Transactions on Biomedical Engineering (TBME), 63(7):1455−1462, 2016.

[7] Settles, Burr. Active learning. Morgan & Claypool, 2012.

[8] Wang, X., Huang, T.-K., Schneider, J., Active transfer learning under model shift. Proceeding ICML'14 Proceedings of the 31st International Conference on International Conference on Machine Learning - Volume 32, Pages II-1305-II-1313

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

[10] Zhao, L., Jialin Pan, S., Wei Xiang, E., Zhong,E., Lu, Z., and Yang, Q. Active transfer learning for cross-system recommendation. In Proceedings of the 27th AAAI Conference on Artificial Intelligence, 2013.