Learning with Dataset Bias in Latent Subcategory Models

Dimitris Stamos[1], Samuele Martelli[2], Moin Nabi[2], Andrew McDonald[1], Vittorio Murino[2,3] and Massimiliano Pontil[1]
[1]Dept. of Computer Science, University College London [2]PAVIS, Istituto Italiano di Tecnologia [3]Dept. of Computer Science, University of Verona

Latent subcategory models (LSMs) offer significant improvements over training flat classifiers such as linear SVMs. Training LSMs is a challenging task due to the presence of many local optima in the objective function and the increased model complexity which requires large training set sizes. An obvious way to have larger training set sizes is to merge datasets from different sources. However, it has been observed by [3] that training from combined datasets needs to be done with care. Although we would expect training a classifier from all available data to be beneficial, it may in fact result in decreased performance because standard machine learning methods do not take into account the bias inherent in each dataset.

The principal contribution of this paper is to extend LSMs to deal with multiple biased datasets. We address this problem from a multitask learning perspective [1]. Specifically, we simultaneously learn a set of biased LSMs as well as a compound LSM (visual world model) which is constrained to perform well on a concatenation of all datasets. We describe a training procedure for our method and provide experimental analysis, which indicates that the method offers a significant improvement over both simply training a latent subcategory model from the concatenation of all datasets as well as the undoing bias method of [2]. Hence, our approach achieves the best of both worlds.

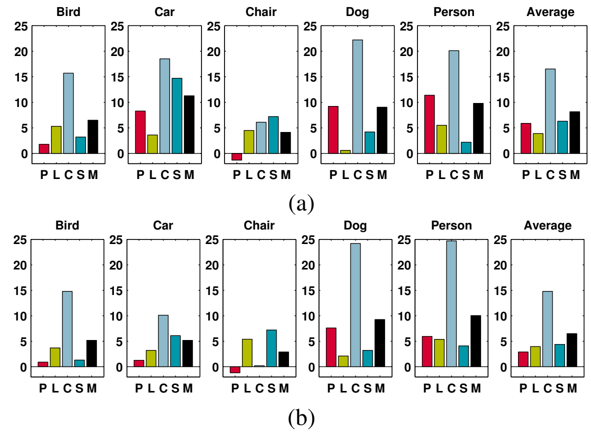

Figure 1: (a) Relative improvement of undoing dataset bias LSM vs. the baseline LSM trained on all datasets at once (P: PASCAL, L: LabelMe, C: Caltech101, S: SUN09, M: mean). (b) Relative improvement of undoing dataset bias LSM vs. undoing bias SVM [2].

1 Learning from Multiple Biased Datasets

Following [2] we assume that we have several datasets pertaining to the same object classification task. Each dataset is collected under specific conditions and so it provides a biased view of the object class. For example, if the task is people classification one dataset may be obtained by labelling indoor images as people / not people, whereas another dataset may be compiled outdoors, and other datasets may be generated by crawling images from internet, etc. In the sequel, we let T be the number of datasets and for $t \in \{1,\dots,T\}$, we let m_t be the sample size in training dataset t and let $\mathcal{D}_t = \{(x_{t1},y_{t1}),\dots,(x_{tm_t},y_{tm_t})\} \subset \mathbb{R}^d \times \{-1,1\}$ be the corresponding data examples.

Undoing Bias SVM: In [2], the authors learn a set of linear max-margin classifiers, represented by weight vectors $\mathbf{w}_t \in \mathbb{R}^d$ for each dataset, under the assumption that the weights are related by the equation $\mathbf{w}_t = \mathbf{w}_0 + \mathbf{v}_t$, where $\mathbf{w}_0$ is a compound weight vector and $\mathbf{v}_t$ captures the bias of the t-th dataset. The weights $\mathbf{w}_0$ and $\mathbf{v}_1,\dots,\mathbf{v}_T$ are then learned by minimizing a regularized objective function which leverages the error of the biased vectors on the corresponding dataset, the error of the visual world vector on the concatenation of the datasets and a regularization term which encourages small norm of all the weight vectors.

Undoing Bias LSM: We now extend the above framework to the latent subcategory setting. We let $\mathbf{w}_t^1,\dots,\mathbf{w}_t^K \in \mathbb{R}^d$ be the weight vectors for the t-th dataset, for $t=1,\dots,T$. For simplicity, we assume that the number of subclassifiers is the same across the datasets, but the general case can be handled similarly. Following [1, 2], we assume that the weight vectors representative of the k-th subcategory across the different datasets are related by the equation

$$\mathbf{w}_t^k = \mathbf{w}_0^k + \mathbf{v}_t^k \tag{1}$$

for $k=1,\dots,K$ and $t=1,\dots,T$. The weights $\mathbf{w}_0^k$ are shared across the datasets and the weights $\mathbf{v}_t^k$ capture the bias of the k-th weight vector in the t-th dataset. We learn all these weights by minimizing the objective function

$$\begin{aligned} & C_1 \sum_{t=1}^{T}\sum_{i=1}^{m_t} L(y_{ti} \max_k \langle \mathbf{w}_0^k + \mathbf{v}_t^k, \mathbf{x}_{ti}\rangle) \qquad (2) \\ + \quad & C_2 \sum_{t=1}^{T}\sum_{i=1}^{m_t} L(y_{ti} \max_k \langle \mathbf{w}_0^k, \mathbf{x}_{ti}\rangle) + \sum_{k=1}^{K}\Big(\|\mathbf{w}_0^k\|^2 + \rho \sum_{t=1}^{T}\|\mathbf{v}_t^k\|^2\Big). \end{aligned}$$

In addition to the number of subclassifiers K, the method depends on 3 other nonnegative hyperparameters, namely C_1, C_2 and ρ, which can be tuned on a validation set. Note that the method reduces to that in [2] if $K=1$ and to the one in [1] if $K=1$ and $C_2=0$. Furthermore our method reduces to training a single LSM on the concatenation of all datasets if $C_1=0$.

As an additional contribution of this paper, we observe that if the positive examples admit a good K-means clustering, and the associated regularisation parameter used in LSM is small relative to the cluster separation, then a good suboptimal solution for the LSM can be obtained by simply clustering the positive class and then training independent SVMs to separate each cluster from the negatives, provided the clustering distortion error is smaller than the regularization parameter used to train the LSM. This result supports a commonly used K-means based heuristic for training subcategory models.

2 Experiments

Following the setting in [2], we employ four datasets: Caltech101, LabelMe, PASCAL2007 and SUN09. We test the methods in two different scenarios, following the "seen dataset" and "unseen dataset" settings. In the first scenario we test on the same datasets used for training. The aim of this experiment is to demonstrate that the visual world model works better than a single model trained on the concatenation of the datasets, and it is competitive with a specific model trained only on the same domain. Furthermore, we show the advantage over setting $K=1$. In the second scenario, we test the model on a new dataset, which does not contribute any training points. Here our aim is to show that the visual world model improves over just training a model on the concatenation of the training datasets as well as the visual world model from [2]. In experiments, we demonstrate that the compound LSM, when tested on the four datasets above in a leave-one-dataset-out fashion, achieves an average improvement of over 6.5% over a previous SVM-based undoing bias approach and an average improvement of over 8.5% over a standard LSM trained on the concatenation of the datasets (see Fig. 1).

[1] T. Evgeniou and M. Pontil. Regularized multi–task learning. In *SIGKDD*, pages 109–117, 2004.

[2] A. Khosla, T. Zhou, T. Malisiewicz, A.A. Efros, and A. Torralba. Undoing the damage of dataset bias. In *ECCV*, pages 158–171. 2012.

[3] A. Torralba and A.A. Efros. Unbiased look at dataset bias. In *CVPR*, pages 1521–1528, 2011.