HETEROGENEOUS DOMAIN ADAPTATION USING PREVIOUSLY LEARNED CLASSIFIER FOR OBJECT DETECTION PROBLEM

Azadeh Sadat Mozafari, Mansour Jamzad

Computer Engineering Department, Sharif University of Technology, Tehran, Iran
amozaffari@ce.sharif.edu, jamzad@sharif.edu

ABSTRACT

When a trained classifier on specific domain (source domain) is applied in a different domain (target domain) the accuracy is degraded significantly. The main reason for this degradation is the distribution difference between the source and target domains. Domain adaptation aims to lessen this accuracy degradation. In this paper, we focus on adaptation for heterogeneous domains (where the source and target domain may have different feature spaces) and propose a novel algorithm which uses the pre-learned source classifier to adapt a trained target classifier. In this method, a max-margin classifier is trained on the target data and is adapted using the offset of the source classifier. The main strength of this adaptation is its low complexity and high speed which makes it a proper adaptation choice for problems with large-size datasets such as object detection. We test our method on human detection datasets and the experimental results show the significant improvement in accuracy, in comparison to several baselines.

Index Terms**—** Domain adaptation, heterogeneous domains, pre-learned classifier, object detection

1. INTRODUCTION

Object detection is a primitive task in many computer vision applications such as human detection in smart cars, face detection in visual surveillance, car detection in traffic control systems and so on. Object detection problems are specified by the big training dataset and wide range of variation in non-object class samples. The research direction, in the field of object detection, is toward finding appropriate features and more efficient classifiers. The promising results are achieved by now, but there is a limiting assumption here: To get the best detection performance, the test data should have the same marginal distribution as training data; otherwise the detection performance may drop significantly. In real world application scenarios, it is very common that the test data differs in view point, lightening, background and resolution from the train data. These visual differences lead to marginal distribution discrepancy between train and test data and consequences the degradation in detection rate.

To overcome this challenge, one solution is to gather the same distributed training data for every different test domain. Training the classifier with only a limited number of labeled samples are usually not robust for object detection; on the other hand collecting enough labeled data is an expensive and time-consuming process which is not applicable most of the time. The other solution that recently has received considerable attention is a new machine learning strategy which is called domain adaptation. Domain adaptation aims to adapt the trained classifier on the first (source) domain in a way that the classifier operates well in the second (target) domain. In recent years, several vision adaptation algorithms [1, 2, 3, 4, 7, 10, 11] are proposed which use different approaches for adaptation according to different assumptions. The assumptions are mostly about accessibility of the source and target labeled data as well as the similarity of the feature type and size in both domains.

In one of the adaptation approaches, pre-learned classifier is used instead of the source data for adaptation [1, 2, 3, 4]. This approach has been proposed for problems in which the source labeled data are not accessible and just the previously trained classifier of the source domain is available, instead. These algorithms are faster than the methods which use the whole source data for adaptation. So, they can be proper adaptation solutions for object detection problems which have large training datasets and also for the object detection toolboxes that have only pre-learned classifier from the source domain[1]. But there is a weakness here; they are not designed for heterogeneous domain adaptation problem where the size and type of features can be different in both domains.

In this paper, we focus on pre-learned classifier adaptation for heterogeneous domains. We propose an algorithm which learns a max-margin classifier on the target domain and tries to adapt it using the source pre-learned classifier offset. We call our method Heterogeneous Max-Margin Classifier Adaptation or HMCA for short. We evaluate HMCA on human detection datasets and the results

[1] For instance, OpenCV 2.4.x library has a pre-learned classifier for human and face detection.

show improvement in comparison to several state-of-the-art baselines.

2. RELATED WORK

In the last few years, computer vision community has devoted a growing interest to domain adaptation problem. Various approaches are suggested based on different assumptions about the source and target domains. One approach is *supervised domain adaptation*, which assumes that the labeled source samples and a few labeled target samples are available. Considering this assumption, Saenko et al. [5] propose a metric learning algorithm which tries to learn a symmetric regularized transformation for mapping between two same-sized domains. Kulis et al. [6] advance Saenko's work and introduce an asymmetric transformation instead of the symmetric one which can adapt the heterogeneous domains. Hoffman et al. [7] introduce an algorithm which learns a transformation matrix and a classifier jointly. They map the target domain with different feature size to the source domain and then train a classifier on them. Xu et al. [8] use a low rank transformation matrix to map the features from the source and target domains into a common space and train a discriminative domain-invariant classifier on them.

Another approach is *unsupervised domain adaptation* where the labeled source and the unlabeled target data are available. Most of works in this field try to close the distribution of the source and target domain by mapping them to a new subspace. Pan et al. [9] use a kernel based approach to find a new subspace in order to reduce the distribution difference between the domains. Gopalan et al. [10] use an incremental learning approach to model two domains as points on Grassmann manifold and take samples of these points along geodesic path to obtain the desired subspace. Gong et al. [11] improve Gopalan work by considering a kernel based approach which uses infinite number of subspaces rather than sampling a finite number of them. Mirrashed et al. [12] use a new approach which considers max margin discriminative property as well as predictability property to map data to a common space. The above mentioned methods use both source and target data for adaptation which can have a high memory and computational complexity especially for large size object detection problem. Also they are not suitable for problems that do not have access to the source data.

In an alternative approach, the previously trained classifier can be used for adaptation, instead of the source data. Yang et al. [1] introduce A-SVM method to adapt a trained classifier on the source data by adding a perturbation function to it. They also, propose Adapt-SVM [2] which uses a new regularizer in SVM's objective function in order to minimize both the target classification error and discrepancy between the target and pre-learned source classifier. Aytar et al. [3] improve the Adapt-SVM algorithm by defining more appropriate regularizer for SVM objective function and call it PMT-SVM. In these methods, the target classifier should have the same size as the source classifier in order to be adapted via pre-learned source classifier. So, these methods cannot work in heterogeneous domains and they are only suitable for homogeneous domains.

In this paper, we address the heterogeneous supervised domain adaptation problem using pre-learned classifier. Our main idea is based on finding a classifier which has the max-margin in the target domain while its offset is close to the source classifier offset. This idea is based on adapting domains using the distribution closeness of the source and target data in a third subspace. After applying the classifiers, the source and target data are mapped to a one dimensional space in which distribution closeness is measured via the distance between source and target classifier offsets.

3. PROPOSED METHOD

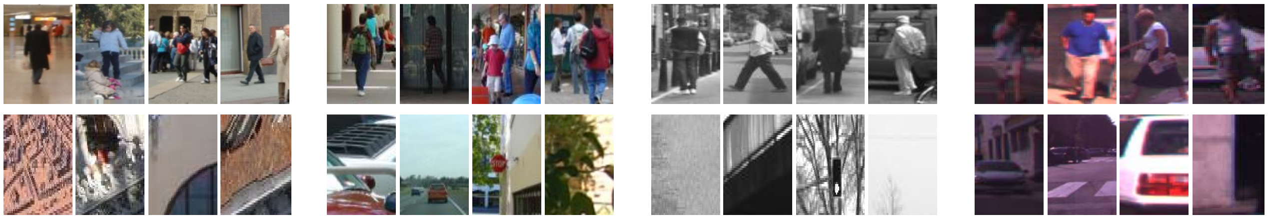
Before going to describe our main idea we should introduce several notations:

3.1. Notations

Let $D_l^t = \{(x_i^t, y_i^t)\}_i^N$ denotes a labeled target dataset which has a few number of samples. x_i^t is the i_{th} data vector from target domain and $y_i^t \in \{-1,1\}$ is its binary label. $x_i^t \in R^m$, where *m* is the target domain feature size. f^s and $b^s$are the pre-learned linear classifier on the source dataset and its offset, respectively.f^s should be a linear SVM classifier in order to have the same mapping process in source and target domain (we discuss about it in 3.2) . $f^s \in R^d$, where *d* is the source feature domain dimension. In our problem *d* and *m* can be equal or different.

3.2. Heterogeneous Max-Margin Classifier Adaptation (HMCA) Method

The main idea in most of heterogeneous domain adaptation methods [6, 7, 8, 9, 12] is transforming the source and target data to a common domain where the discrepancy between the source and target data distribution is minimized. This idea is not applicable when only the previously trained classifier from the source domain is available. We get the transformation idea and customized it for pre-learned classifier problem. When a classifier is applied to the data, we can say, it transforms data to a one dimensional space in which the data can be discriminated using the classifier offset. In this new space, the distribution of data is such that the two classes are far from each other and the offset is the point between them. Now for adaptation, we should look for a classifier in target domain that discriminates the target



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

Figure 1: Positive and negative samples of INRIA, NICTA, Daimler and CVC-CER-01 are shown in (a), (b), (c) and (d), respectively. As we can see, these four human detection datasets are different in lightening and resolution.

data as well as transforming them to a one dimensional space in which the target data have close distribution to the mapped source data. We use SVM framework to find the max-margin classifier on the target data and by adding one more regularizer part to the SVM objective function which minimizes the distance between the source and target classifier offsets, we try to close the distribution of the transformed source and target data. We consider the offsets as an indicator for data distributions in one dimensional space. The proposed objective function is shown in Eq.(1).

$$\begin{aligned} &\min_{w,b_t} \frac{1}{2}\|\boldsymbol{w}\|^2 + C_1 \sum_{i=1}^{N} \varepsilon_i + C_2 \|b_s - b_t\|^2 \\ &\boldsymbol{s.t:} \\ &\quad \varepsilon_i \geq 0\,; \\ &\quad y_i(\boldsymbol{w}^{\boldsymbol{T}} x_i + b_t) \geq 1 - \varepsilon_i, \ \ \forall (x_i, y_i) \in D_l^t \end{aligned} \qquad (1)$$

The constant C_1 penalizes the target classification error and C_2 controls the amount of adaptation. Eq.(1) is a simple convex optimization which can be solved using any standard QP optimization package. Additionally, since the number of constraints is just related linearly to the number of labeled target samples, the optimization can be solved efficiently, especially for object detection problems.

4. EXPERIMENTAL RESULTS

We now present our experiments to evaluate the performance of the proposed adaptation model. The model is tested on two scenarios: the homogeneous domain adaptation and heterogeneous domain adaptation. All results are shown for the class of human detection.

4.1. Dataset setting and source-target domain setup

We test and analyze our algorithm on INRIA [13], NICTA[2] [14], CVC-CER-01[3] [15], and Daimler[4] [16] datasets which are standard image datasets for human detection. All of them have positive and negative samples for test and train. Some positive and negative images of these four datasets are shown in Figure 1. In all experiments, INRIA is considered as a source domain and other datasets as target domains. For source domain, we choose 1200 positive images and its mirror and 6000 randomly clipped negative images from training part of INRIA dataset. In each experiment, we make target training data by randomly selecting 100 positive and 300 negative samples from one of the target datasets (NICTA, CVC or Daimler) and test the results on remained images from it.

4.2. Shape and texture descriptor

In the following experiments we extract HOG [17] feature from the source and target domains for homogeneous adaptation and extract HOG from the source and HOT [18] features from the target domain for heterogeneous domain adaptation. For all of the positive and negative examples, we fix their size to 32×64 pixels. Thus, the HOG feature has 756 dimensional and HOT 672 dimensional vector.

4.3. Optimization package

We use CVX [19], a package for specifying and solving convex programs, in all experiments.

4.4. Parameter setting

In HMCA, C_1 and C_2 parameters are chosen 0.3 and 0.01 via cross validation.

4.5. Baselines

In this paper, we focus on object detection problem. Accordingly, we select baselines which can be applied for object detection problem with large dataset. We choose PMT-SVM [3], Hoffman et al. [7], SVM-Source[5] (a linear SVM classifier trained only with source domain samples),

[2] We choose 2000 positive (1000 human image and its mirror) and 11000 negative images.
[3] We choose 2000 positive images (1000 human image and its mirror) and 6000 negative images.
[4] We choose randomly 2400 positive images (1200 human image and its mirror) and 12000 randomly clipped negative images

[5] LibSVM package [20] is used for training linear SVM.

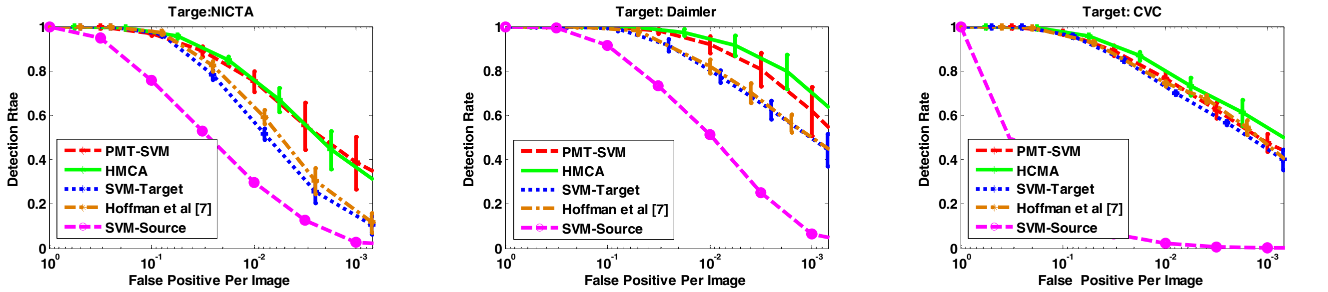


Figure 2: Results of homogeneous domain adaptation on human detection datasets. INRIA dataset is considered as a source domain and the results of adaptation on each target datasets are shown separately. Clearly, HMCA shows better performance in compared to other baselines.

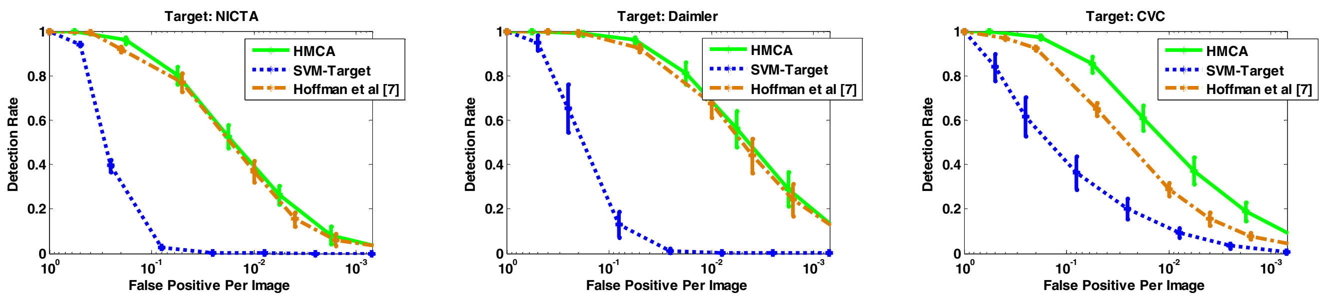


Figure 3: Results of heterogeneous domain adaptation on human detection datasets. INRIA dataset is considered as a source domain and the results of adaptation on each target datasets are shown separately. HMCA achieves better detection rate.

Table 1: Level of complexity for three adaptation methods. Complexity is described by an order on the number of constraints and variables. N and S are the number of target and source samples and d and m are the size of source and target feature vectors, respectively. (In our heterogeneous problem, $N = 400$, $S = 8400$, $d = 756$ and $m = 672$)

Method	O(Variables)	O(Constraints)
HMCA	O(m)	O(N)
Hoffman et al. [7]	O($m\times d$)	O(S+N)
PMT-SVM	O(m)	O(N)

and SVM-Target (a linear SVM classifier trained only with target domain samples) as baselines for homogeneous adaptation. For heterogeneous adaptation, because PMT-SVM and SVM-Source cannot be applied on target domains with different dimension from that of the source; therefore we have to compare HMCA with SVM-Target and Hoffman method.

4.4. Experimental discussion

We report the average detection rate versus false positive per image in each scenario and for every target domain, separately. We create five random train/test splits and average the results across them. The results for homogeneous and heterogeneous domain adaptation are shown in Figure 2 and 3, respectively. As we can see HMCA achieves better detection rate in both scenarios versus all other baselines while it has low complexity. In this approach, complexity is defined by the number of variables and the number of constraints used in optimization. Table 1 shows the level of complexity for three methods of PMT-SVM, HMCA and Hoffman. High complexity can make an adaptation method non-tractable for large scale object detection problems while it could have acceptable accuracy. So, the main strength of HMCA is its low complexity in addition to its high accuracy that makes it a proper choice for object detection problems.

5. CONCLUSION

In this paper, we proposed an adaptation method for heterogeneous domains using pre-learned classifier, called HMCA. This method tries to learn a max-margin classifier on the target data while trying to close the distribution of target data to the source data by closing their classifier offset to each other. The introduced algorithm has a convex objective function with a few numbers of linear constraints which makes it a superior adaptation method. We applied it to the human detection problem and the results show it outperforms several baselines. In HMCA, the offset of classifier is used as an indicator of the data distribution in one dimensional space; however it is not accurate estimation. In the future, we would like to advance our method by using the latent information scattered in the classifier line for distribution estimation to overcome the inaccuracy in estimation.

6. REFERENCES

[1] J. Yang, R. Yan, and A.G. Hauptmann, "Cross-domain video concept detection using adaptive SVMs," *In Proceeding of the 15th International Conference on Multimedia*, New York, USA, pp. 188-197, 2007.

[2] J. Yang, R. Yan, and A.G. Hauptmann, "Adapting SVM classifiers to data with shifted distribution," *In Workshops Proceedings of the 7th IEEE International Conference on Data Mining*, Nebraska, USA, pp. 69-76, 2007.

[3] Y. Aytar, and A. Zisserman, "Tabula rasa: model transfer for object category detection," *In Proceeding of the International Conference on Computer Vision*, Washington, USA, pp. 2252-2259, 2011.

[4] L. Bruzzone, and M. Marconcini, "Domain adaptation problems: a DASVM classification technique and a circular validation Strategy," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 770-787, 2010.

[5] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," *In Proceeding of the 11th European Conference on Computer Vision*, Crete, Greece, pp. 213-226, 2010.

[6] B. Kulis, K. Saenko, and T. Darrell, "What you saw is not what you get: domain adaptation using asymmetric kernel transform," *In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, USA, pp. 1785-1792, 2011.

[7] J. Hoffman, E. Rodner, J. Donahue, K. Saenko, and T. Darrell, "Efficient learning of domain-invariant image representations," *Accepted to be Published In Proceeding of International Conference on Learning Representation*, Arizona, USA, 2013.

[8] J. Xu, S. Ramos, X. Hu, D. Vazque, and A.M. Lopez, "Multi-task bilinear classifier for visual domain adaptation," *Accepted to be Published In Proceeding of Neural Information Processing Systems Foundation*, Nevada, USA, 2013.

[9] S.J. Pan, I. Tsang, J.T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transaction on Neural Networks*, vol. 22, pp.199-210, 2011.

[10] R. Gopalan, R.Li, and R. Chellappa, "Domain adaptation for object recognition: an unsupervised approach," *In Proceeding of 13th International Conference on Computer Vision*, Barcelona, Spain, pp. 999-1006, 2011.

[11] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," *In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, RI, USA, pp. 2066-2073, 2012.

[12] F. Mirrashed, and M. Rastegari, "Domain Adaptive Classification," *Accepted to be Published In Proceeding of International Conference on Computer Vision*, Sydney, Australia, 2013.

[13] INRIA dataset http://lear.inrialpes.fr/dat.

[14] G. Overett, L. Petersson, N. Brewer, L. Andersson and N. Pettersson, "A new pedestrian dataset for supervised learning," *In Proceeding of IEEE Intelligent Vehicles Symposium*, Eindhoven, Netherland, pp. 373-378, 2008.

[15] D. Gerónimo, A.D. Sappa, A. López, and D. Ponsa, "Adaptive image sampling and windows classification for on-board pedestrian detection," *In Proceedings of the International Conference on Computer Vision Systems*, Bielefeld, Germany, 2007.

[16] M. Enzweiler, and D.M. Gavrila, "Monocular pedestrian detection: survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2179-2195, 2009.

[17] N. Dalal, and B. Triggs, "Histograms of oriented gradients for human detection," *In Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, CA, USA, pp. 886-893, 2005.

[18] S.Tang, and S. Goto, "Histogram of template for human detection," In Proceeding of IEEE International Conference on Acoustics Speech and Signal Processing, TX, USA, pp. 2186-2189, 2010.

[19] M. Grant, and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, 2013.

[20] C.C. Chang, and C.J. Lin, "LIBSVM: a library for support vector machines," http://www.csie.ntu.edu.tw/cjlin/libsvm, 2001.