Adaptive Transductive Transfer Machines

Nazli Farajidavar
http://personal.ee.surrey.ac.uk/Personal/N.Farajidavar
Teofilo deCampos
http://personal.ee.surrey.ac.uk/Personal/T.Decampos
Josef Kittler
http://www.surrey.ac.uk/cvssp/people/josef_kittler

Center for vision, speech and signal processing
University of Surrey
Guildford, GU2 7XH
UK

Transductive transfer learning methods can potentially improve a very wide range of classification tasks, as it is often the case that a domain change happens between training and application of algorithms, and it is also very common that unlabelled samples are available in the target domain.

In this paper, we propose Adaptive Transductive Transfer Machine (ATTM) which combines methods that adapt the marginal and the conditional distribution of the samples, so that source and target datasets become more similar, facilitating classification (TTM). We further introduce two unsupervised dissimilarity measures which are the backbones of our classifier adaptation approach. ATTM uses these measures to select the best classifier and to further optimise its parameters for a new target domain. We show that our method obtains state-of-the-art results in cross-domain vision datasets using naïve features, with a significant gain in computational efficiency in comparison to related methods.

We propose the following TTM pipeline:

(a) A global linear transformation G^1 is applied to $\mathbf{X}^{src}$ and $\mathbf{X}^{trg}$ such that the marginal $P(G^1(\mathbf{X}^{src}))$ becomes more similar to $P(G^1(\mathbf{X}^{trg}))$. Following [2, 3, 4, 5] we adopt the Maximum Mean Discrepancy (MMD) for defining a projection matrix which aims to minimise the distance between the sample means of the source and target domains.

(b) With the same objective, a local transformation is applied to each transformed source domain sample $G_i^2(G^1(\mathrm{x}_{src}^i))$.

$$G_i^2(G^1(\mathrm{x}_{src}^i)) = G^1(\mathrm{x}_{src}^i) + \gamma \mathrm{b}^i \,, \tag{1}$$

Modeling the unlabelled target data, by a mixture of Gaussian probability density functions (GMM), we can formulate the problem of finding an optimal translation parameters b as one of maximising the likelihood of the translated source sample measured in the target domain.

$$\mathrm{b}^i = \frac{\sum_{k=1}^K P(\mathrm{x}^i + \mathrm{b}_0^i|\lambda_k)\Sigma_k^{-1}(\mathrm{x}^i - \mu_k)}{\sum_{k=1}^K P(\mathrm{x}^i + \mathrm{b}_0^i|\lambda_k)\Sigma_k^{-1}} \,, \tag{2}$$

where b_0^i is an initial value of b^i, which is set to a vector of zeros. In our experiments, we ran (2) only once, though one can iterate it further.

(c) Finally, aiming to reduce the difference between the conditional distributions in source and target spaces, a class-based transformation is applied to each of the transformed source samples $G_{y^i}^3(G_i^2(G^1(\mathrm{x}_{src}^i)))$ following the TST transformation of [1].

Figure 1 illustrates the effect of the three steps of the TTM pipeline.

In the Adaptive TTM we have an extra *classifier Selection and learning parameters adaptation* step where we introduce two unsupervised dissimilarity measures for selecting a proper classifier and for adapting its parameters. More specifically, when both dissimilarity measures indicate that the cross-domain datasets are very different, we suggest that it is better to use a non-parametric classifier, like Nearest Neighbour, so no optimisation is employed at training. When the two domains are similar at both global and cluster levels, it is sensible to use a classifier such as KDA, whose parameters optimised on the source domain have a better chance of working on the target space. And finally when two domains are similar at global levels but the clusters distribution in the two domains are different we propose to use the KDA but adapt the lengthscale σ of the RBF kernel using a linear function of the cluster dissimilarity measure.

Figure 2 demonstrates the full ATTM pipeline.

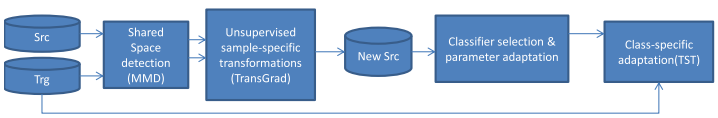
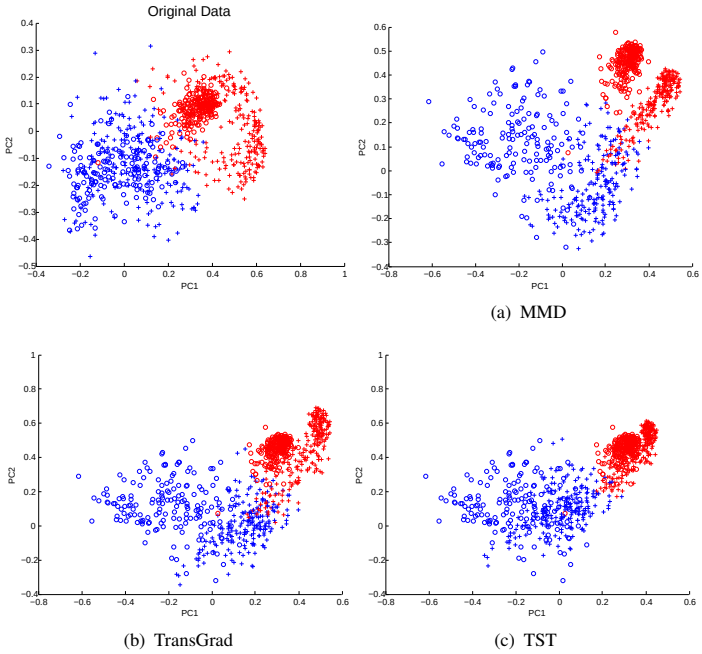


Figure 2: Adaptive Transductive Transfer Machine (ATTM).



(a) MMD (b) TransGrad (c) TST

Figure 1: The effect of different steps of our pipeline on digits 1 and 2 of the MNIST→USPS datasets, visualised in 2D through PCA. The source dataset (MNIST) is indicated by stars, the target dataset (USPS) is indicated by circles, red indicates samples of digit 1 and blue indicates digit 2 (better viewed on the screen).

Comprehensive experiments on MNIST, USPS, COIL20 and Caltech+ Office datasets show that our proposed TTM pipeline leverages the averaged performance by 1.32% compared to the best performing state-of-the-art of approach, JDA [3]. We have further tested our proposed *classifier Selection and learning parameters adaptation* on both JDA and TTM algorithms as AJDA and ATTM. The AJDA performance shows that the model adaptation drastically enhances the final classifier. The performance gains of **4.59** and **4.29** in ATTM and AJDA respectively validates the proposed dissimilarity measures for model selection and adaptation.

It is worth pointing out that ATTM is a general framework with applicability beyond image classification and could be easily applied to other domains, even outside Computer Vision. For future work, we suggest studying combinations of our method with instance reweighting methods and multi-source transfer learning.

[1] N. FarajiDavar, T. deCampos, J. Kittler, and F. Yang. Transductive transfer learning for action recognition in tennis games. In *ICCV,VECTaR workshop*, 2011.

[2] A. Gretton, K. Borgwardt, M. Rasch, B. Scholkopf, and A. Smola. A kernel method for the two sample problem. In *ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS 19*, pages 513–520. MIT Press, 2007.

[3] M. Long, J. Wang, G. Ding, and P. Yu. Transfer learning with joint distribution adaptation. In *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[4] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1187–1192, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[5] Q. Sun, R. Chattopadhyay, S. Panchanathan, and J. Ye. A two-stage weighting framework for multi-source domain adaptation. In *NIPS*, pages 505–513, 2011. URL `http://dblp.uni-trier.de/db/conf/nips/nips2011.html#SunCPY11`.