Adaptive Transductive Transfer Machines

N. FarajiDavar, T. deCampos , J. Kittler
Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
(n.farajidavar, t.decampos, j.kittler)@surrey.ac.uk

Introduction

Our ultimate goal is to investigate transfer learning methods for annotating image data of varying qualities.
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 .

Amazon Caltech DSLR Webcam

Method

We propose the following TTM pipeline:

(a) A global linear transformation G^1 is applied to X^{src} and X^{trg} such that the marginal $P(G^1(X^{src}))$ becomes more similar to $P(G^1(X^{trg}))$ using **MMD** [2, 3] to minimise the distance between the sample means of the source and target domains.

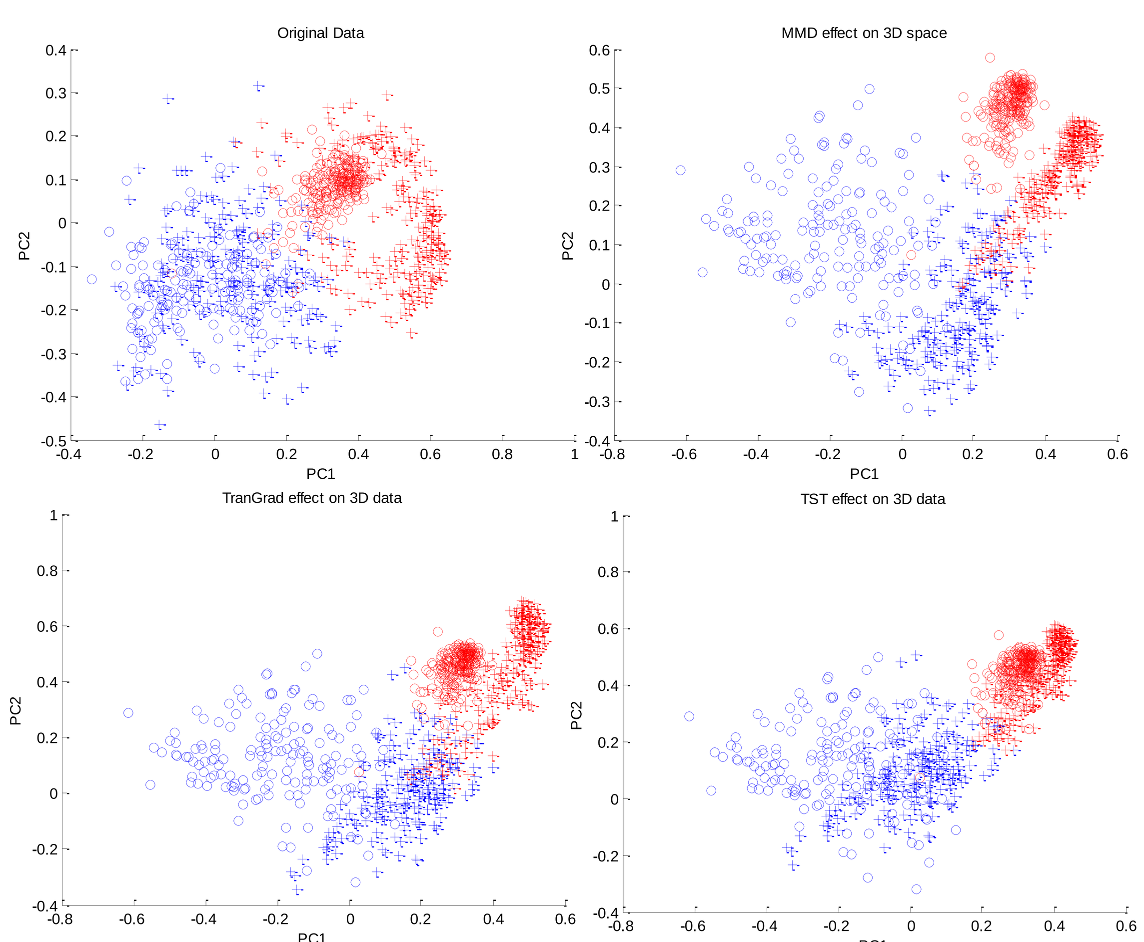
(b) With the same objective, a local transformation is applied to each transformed source domain sample $G^2_i(G^1(x^i_{src}))$ called **TransGrad**,

$$G^2_i(G^1(x^i_{src})) = G^1(x^i_{src}) + \gamma b^i$$

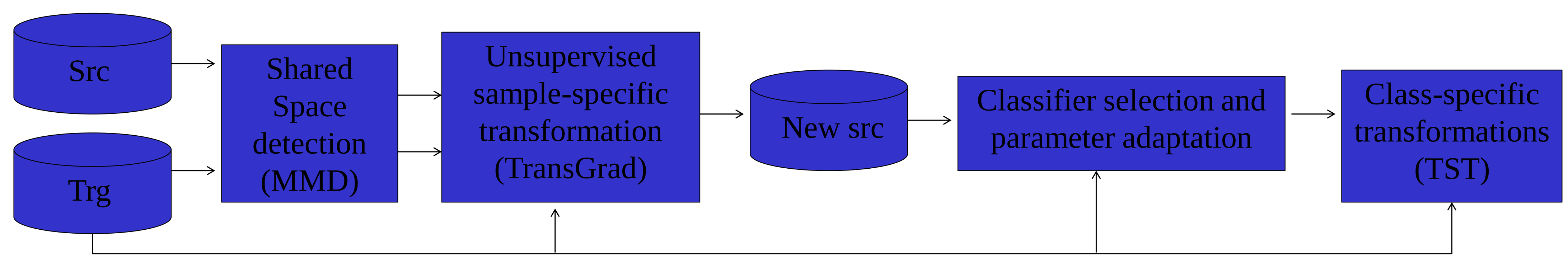
where the unlabelled target data distribution is modelled by a GMM:

$$b^i = \frac{\sum_{k=1}^{K} P(\mathrm{x}^i + b_0^i \mid \lambda_k)\Sigma_k^{-1}(\mathrm{x}^i - \mu_k)}{\sum_{k=1}^{K} P(\mathrm{x}^i + b_0^i \mid \lambda_k)\Sigma_k^{-1}}$$

(c) 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 $G^3_{yi}(G^2_i(G^1(x^i_{src})))$ following the **TST** transformation of [1].



(d) In the **Adapt**ive 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



Results

Table.1: Classifiers' evaluations on individual domains: 5-fold cross validation accuracy.

Classifier	MNIST	USPS	COIL1	COIL2	Caltech	Amazon	Webcam	DSLR
PCA+NN	91.97	93.64	99.02	98.91	38.80	60.59	79.58	**76.95**
LR	86.15	89.22	92.36	92.22	56.27	72.46	80.01	67.49
KDA	**94.05**	94.84	**100.00**	**99.71**	**58.16**	**78.73**	**89.54**	63.94
SVM	91.80	**95.28**	99.72	99.44	57.17	74.86	86.44	75.80

Table.2: Dissimilarity measures and recognition accuracies with datasets abbreviated as M: MNIST, U: USPS, C: Caltech, A: Amazon, W: Webcam, and D: DSLR.

TTL Test	Cluster diss.	Global diss.	NN baseline	GFK (PLS, PCA)	JDA (1NN) [3]	TTM0 (TST + 1NN)	TTM1 (MMD + TTM0)	TTM2 (TransGrad + TTM1)	AJDA (Adapt. JDA)	ATTM (Adapt. TTM2)
M → U	0.034	0.984	65.94	67.22	67.28	75.94	76.61	77.94	67.28	**77.94**
U → M	0.032	0.981	44.70	46.45	59.65	59.79	59.41	61.15	59.65	**61.15**
COIL1 → COIL2	0.026	0.627	83.61	72.50	89.31	88.89	88.75	93.19	**94.31**	92.64
COIL2 → COIL1	0.025	0.556	82.78	74.17	88.47	88.89	88.61	88.75	**92.36**	91.11
C → A	0.032	0.548	23.70	41.40	44.78	39.87	44.25	46.76	58.56	**60.85**
C → D	0.031	0.786	25.48	41.10	45.22	50.31	44.58	47.13	46.86	**50.32**
A → C	0.035	0.604	26.00	37.90	39.36	36.24	35.53	39.62	40.43	**42.92**
A → W	0.035	0.743	29.83	35.7	37.97	37.63	42.37	39.32	49.83	**50.51**
W → C	0.037	0.752	19.86	29.3	31.17	26.99	29.83	30.36	**35.80**	34.02
W → A	0.035	0.717	22.96	35.5	32.78	29.12	30.69	31.11	38.94	**39.67**
D → A	0.034	0.790	28.50	36.1	33.09	31.21	29.75	30.27	37.47	**38.73**
D → W	0.033	0.471	63.39	79.1	89.49	85.05	**90.84**	88.81	89.49	88.81
Average	-	-	43.06	50.00	54.88	54.12	55.10	56.20	59.17	**60.72**

Conclusion

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.

Acknowledgements

This project is sponsored by EPSRC/UK through grant EP/F069421/1 (ACASVA - Adaptive Cognition for Automated Sports Video Annotation). We are also grateful for the support of the EU PASCAL2 network of excellence.

References

[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 NIPS 19, pages 513–520, 2007.
[3] M. Long, J.Wang, G. Ding, and P. Yu. Transfer learning with joint distribution adaptation 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 joint conference on Artificial 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.