Domain Adaptation in the Context of Sport Video Action Recognition

N. FarajiDavar T. deCampos D. Windridge J. Kittler W. Christmas
CVSSP, University of Surrey*
Guildford, GU2 7XH, UK
t.decampos@surrey.ac.uk

Abstract

We apply domain adaptation to the problem of recognizing common actions between differing court-game sport videos (in particular tennis and badminton games). Actions are characterized in terms of HOG3D features extracted at the bounding box of each detected player, and thus have large intrinsic dimensionality. The techniques evaluated here for domain adaptation are based on estimating linear transformations to adapt the source domain features in order to maximize the similarity between posterior PDFs for each class in the source domain and the expected posterior PDF for each class in the target domain. As such, the problem scales linearly with feature dimensionality, making the video-environment domain adaptation problem tractable on reasonable time scales and resilient to over-fitting. We thus demonstrate that significant performance improvement can be achieved by applying domain adaptation in this context.

1 Introduction

In domain adaptation, a class probability joint distribution $P(\mathbf{Y}, \mathbf{X}^{src})$ over pattern vectors $\mathbf{X}$ and classes $\mathbf{Y}$ in the source domain is assumed to be related to that of a target domain joint distribution $P(\mathbf{Y}, \mathbf{X}^{trg})$. We seek to obtain the conditional class probabilities $P(\mathbf{Y}|\mathbf{X}^{trg})$ using a set of labelled samples distributed according to $P(\mathbf{Y}|\mathbf{X}^{src})$.

Within the definition of domain adaptation there are many approaches possible that can be adapted to existing techniques; for instance, Dai et al. [1] modify AdaBoost to preferentially re-weight misclassified target domain instances iteratively. Broadly, however, we can split domain adaptation into two distinct areas: the transformative and the non-transformative. In the transformative case (c.f. e.g. [2]), the idea is to find some transformation G of $\mathbf{X}$, such that $P(\mathbf{Y}, G(\mathbf{X}^{src}))$ and $P(\mathbf{Y}, \mathbf{X}^{trg})$ can be assumed identical, and the domain adaptation problem becomes a straightforward problem of classification. An example of this approach is that of Satpal and Sarawagi [3], who use feature selection to match source and target distributions. In the non-transformative case, the alternative is to amalgamate labelled source data and unlabeled target data together and treat the problem as one of semi-supervised learning (c.f. [4]), on the assumption of reasonable similarity between $P(\mathbf{Y}, \mathbf{X}^{src})$ and $P(\mathbf{Y}, \mathbf{X}^{trg})$.

In the current paper, we employ a transformative approach in which linear transformations are used to adapt the source domain feature distributions to the target domain. This linearity is particularly advantageous in the video environment given the very large feature dimensionalities possible, both in terms of speeding-up the transformation calculation and also in reducing the danger of over-fitting (a covariance-based transformation would, in contrast, scale with the square of feature dimensionality).

*http://www.ee.surrey.ac.uk/CVSSP/

This project was sponsored by the EPSRC/UK grant EP/F069421/1 (ACASVA) and EU PASCAL2.
This is the preprint of a paper that appears in **NIPS Domain Adaptation Workshop**, Dec 2011.

The chosen domain adaptation problem is thus that of identifying actions in court-game sport videos, with learning transfer between domains characterized by distinct set of rules; specifically, *tennis* and *badminton* (singles/doubles). Actions here are characterized by HOG3D features [5] extracted at the bounding box of each detected player, generating a problem of large intrinsic dimensionality. To our knowledge, only two domain adaptation approaches have been applied in a similar domain; Arnold et al.'s transductive transfer learning method based on a maximum entropy model [6] and our method [7]. Both were proposed as Transductive Transfer Learning techniques, which is a category of transfer learning that shares the assumptions with domain adaptation [8]. In [7], we evaluated these two methods on a small dataset. Here we present further experiments with different modalities of sports.

The video sequences are divided into point-based clips, such that each sequence provides several samples of actions performed by the same players under similar conditions (so that e.g. the camera set-up and video coding method stay constant). Within each sequence, low level descriptors ought thus to share features that relate to the style of the players, to their appearance and to the appearance of the background and illumination. However, all of these features are subject to variation from one sequence to another and often these variations jeopardize the balance between generalization and discriminative power of classifiers. Further domain disparity is brought about by rule changes (e.g. players are distributed differently in singles and doubles matches). There is thus a clear requirement for an appropriately-tailored domain adaptation approach in order to solve the *sports video annotation problem.*

2 Methodology

Let $\mathbf{X} = \{\mathbf{x}^1, \ldots, \mathbf{x}^i, \ldots, \mathbf{x}^N\} \in I\!R^D$ be a set measurement feature vectors $\mathbf{x}^i = (x_1^i, \ldots, x_j^i, \ldots, x_D^i)$. The goal is to find a transformation $G(\mathbf{X})$ such that the PDF of $P(\mathbf{Y}, G(\mathbf{X}^{src})) \approx P(\mathbf{Y}, X^{trg})$. In other words, we wish to transform the samples in the source domain so that they become more similar to those observed in the unlabeled target domain. As discussed before, several methods can be applied. We follow the methods proposed in [6] and [7] because they are based on linear transformations applied to each individual feature. The transformation is thus estimated and applied for each feature and each class.

Arnold et al.'s transductive method [6] basically consists in defining G using the ratio between the source expectation $E^{src}[x_j, y]$ and an estimate of the target expectation obtained using a classification model trained on the source data set $E^{trg}_{\Lambda_{src}}[x_j, y]$:

$$G(x_j^i) = x_j^i E^{trg}_{\Lambda_{src}}[x_j, y_i] / E^{src}[x_j, y_i], \ \forall i = 1{:}\, N^{src}_{train} \ , \tag{1}$$

$$\text{where} \quad E^{src}[x_j, y] = [\textstyle\sum_{i=1}^{N^{train}_{src}} x_j^i \mathbb{1}_{[y]}(y_i)] / [\sum_{i=1}^{N^{train}_{src}} \mathbb{1}_{[y]}(y_i)] \ , \tag{2}$$

$$E^{trg}[x_j, y] \approx E^{trg}_{\Lambda_{src}}[x_j, y] = \frac{\sum_{i=1}^{N^{trg}} x_j^i P_{\Lambda_{src}}(y|\mathbf{x}_i)}{\sum_{i=1}^{N^{trg}} P_{\Lambda_{src}}(y|\mathbf{x}_i)} \ , \tag{3}$$

and $\mathbb{1}_{[y]}(y_i)$ is an indicator function[1]. The effect is to re-scale x_j, giving more weight to features that occur frequently in the target but rarely in the source. This method is thus dubbed *reweight* in the rest of this paper.

FarajiDavar et al. [7] used a geometric interpretation of the problem and proposed to use a transformation based on translating and scaling (abbreviated as *trans+scale*) features by adjusting their means and standard deviations:

$$G(x_j^i) = \frac{x_j^i - E^{src}[x_j, y_i]}{\sigma^{src}_{j,y_i}} \sigma^{trg}_{j,y_i} + E^{trg}_{\Lambda_{src}}[x_j, y_i] \ , \forall i = 1{:}\, N^{src}_{train} \ , \tag{4}$$

where σ^{src}_{j,y_i} is the standard deviation of feature x_j of the source samples labeled as y_i and

$$\sigma^{trg}_{j,y_i} = \sqrt{\frac{\sum_{k=1}^{N^{trg}} (x_j^k - E^{trg}_{\Lambda_{src}}[x_j, y_i])^2 P_{\Lambda_{src}}(y_i|\mathbf{x}_k)}{\sum_{k=1}^{N^{trg}} P_{\Lambda_{src}}(y_i|\mathbf{x}_k)}} \ . \tag{5}$$

[1] Equations (2) and (3) presented here follow the rectifications suggested in [7].

Table 1: Datasets and number of samples per class

label	sport	gender	number	competition	year	non-hit	hit	serve
TWSA03	Tennis	Women	Singles	Australian	2003	944	214	72
TMSA03	Tennis	Men	Singles	Australian	2003	1881	469	123
TWDA09	Tennis	Women	Doubles	Australian	2009	1064	135	36
BMSB08	Badminton	Men	Singles	Beijing	2008	706	458	8

Table 2: Baseline results and results with two methods for DA: *reweight*|*trans+scale*, in ‰.

	source	target		accuracy per class (‰)			macro
		adaptation	test	non-hit	hit	serve	average
a	TWSA03	–	TWDA09	996	149	571	572
b	TWSA03	test set	TWDA09	939|939	418|433	857|886	738|752
c	TWDA09	–	TWSA03	978	305	986	756
d	TWDA09	test set	TWSA03	870|912	676|634	972|972	839|839
e	TWSA03	–	TMSA03	981	248	549	592
f	TWSA03	test set	TMSA03	975|973	427|442	852|902	751|772
g	BMSB08	–	TMSA03	359	779	0	**379**
h	BMSB08	test set	TMSA03	327|393	886|852	0|0	404|415
i	BMSB08+TWSA03	–	TMSA03	940	357	500	599
j	BMSB08+TWSA03	test set	TMSA03	975|917	427|547	852|942	767|802
k	BMSB08	TWSA03	TMSA03	245|330	983|908	0|0	394|**413**

Since the target expectation $E^{trg}_{\Lambda_{src}}[x_j, y_i]$ is only an approximation based on the posterior function rather than the labels (which are not available in the target set), there is a danger that samples that would be miss-classified can lead to negative transfer. To alleviate this, we follow Arnold et al.'s suggestion and use this *smoothing function*:

$$G'(x^i_j) = (1-\theta)x^i_j + \theta G(x^i_j)\ . \tag{6}$$

3 Experiments and Results

We follow FarajiDavar et al. [7] and use HOG3D feature vectors [5] extracted at the bounding box of each player, with a buffer of 24 frames around the key moment (e.g. when the player hits the ball). We used the parameters optimized for the KTH dataset, which give 960 dimensional vectors, as described in [5].

The datasets consist in videos of tennis and badminton, as summarized in Table 1. The TWDA09 is coded in NTSC and the others are in PAL. TWSA03 and TWDA09 are the videos that were used in [7], but in that paper a fully automatic player detection method was used, which in some instances resulted in merging and miss-labeling of bounding boxes. In this paper we focus on the action classification task and allowed for some manual correction of bounding boxes and their labels for training and evaluation of results. This explains why there are less samples per class in comparison to [7], which included false positive player detections. For this reason, we repeat the experiments of [7]. The results are shown in rows (a–d) of Table 2, where rows (a) and (c) are the baseline (without domain adaptation, DA) and rows (b) and (d) show DA results with both methods reviewed in Section 2. We used experiments (b) and (d) to evaluate a range of θ values for Equation (6). Although high values of θ lead to better results the performance nearly plateaued or worsened after $\theta = 0.6$. Following the observations in [7], we used a more conservative value, $\theta = 0.5$ for the experiments in Table 2.

We further present experiments with a more challenging change of domain: from a badminton game (BMSB08) to a tennis game (TMSA03). All the results shown confirm that when DA is used, there is an improvement in performance. All of them also show that the *trans+scale* method gives better results than *reweight*, but this is not consistent across all class labels.

In most experiments presented, we show results using the target set to adapt the parameters and obtain G', which is then applied to the test set, i.e., $\mathbf{X}^{trg} = \mathbf{X}^{test}$. This is the same scenario

evaluated in [7]; it considers that all the unlabeled test set samples are available at once, which is a reasonable assumption for batch processing methods. For the experiment in line (k), we consider the scenario that samples from one domain (badminton, BMSB08) are used for training and unlabeled samples from a second domain (tennis women's, TWSA03) are available to compute G'. Then one wishes to use the same adaptation parameters G' for another set of test samples that come from a similar domain (tennis men's, TMSA03). The result shows an improvement over the baseline which is in line (g), even though the adaptation set is different from the test set $\mathbf{X}^{trg} \neq \mathbf{X}^{test}$. Even more notable is that the result of *trans+scale* in (k) approximates that in (h), which used $\mathbf{X}^{trg} \neq \mathbf{X}^{test}$.

As expected, results in (k) are not as good as those in (i) and (j), which used the tennis women's game (TWSA03) and its labels to complement the training set. Line (j) shows an upper bound of performance in this scenario, as both BMSB08 and TWSA03 were used to compute their individual adaptation functions (i.e., one G' for each training set) and their combined adapted sets $G'(\mathbf{X}^{TWSA03})$ and $G'(\mathbf{X}^{BMSB08})$ were used to re-train the classifier. As expected, the result in (j) is significantly better than that in (i).

Note that owing both to the scarcity of serve samples and to the fact that serves in badminton are very similar to backhand hits, this class was never detected when badminton was the only game used for training. This leads the way for future explorations with transfer learning methods to deal with changes of domain in which classes may split into sub-categories (e.g. $hit \rightarrow \{backhand, forehand\}$) or merge into super-categories (e.g. $\{hit, serve\} \rightarrow moving\ arm\}$).

4 Conclusion

We have presented an evaluation of domain adaptation techniques for action classification in court sports. We complemented the experiments presented in [7] using more video sequences and introduced experiments with different sports. We used a video of badminton for training and tennis for testing. By applying domain adaptation, we obtained an improvement in classification results, even if the video used to compute adaptation parameters was not the same as that in the test set.

The court-game sport video environment is an inherently interesting setting for domain adaptation because of the potentiality for cross-modal (audio and video) domain adaptation, and also the possibility of inductive rule transfer; ultimately this will lead us to consider the problem of *simultaneous* rule adaptation and low-level feature adaptation.

References

[1] Wenyuan Dai, Yuqiang Chen, Gui rong Xue, Qiang Yang, and Yong Yu. Translated learning: Transfer learning across different feature spaces. In *Advances in Neural Information Processing Systems (NIPS)*, pages 353–360, 2008.

[2] Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems (NIPS)*, volume 19, pages 137–144. MIT Press, Cambridge, Massachusetts, US, 2006.

[3] Sandeepkumar Satpal and Sunita Sarawagi. Domain adaptation of conditional probability models via feature subsetting. In *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD)*, pages 224–235, Warsaw, Poland, September 2007.

[4] Xiaojin Zhu. Semi-supervised learning literature survey. Technical Report 1530, Computer Sciences, University of Wisconsin-Madison, 2005.

[5] Alexander Kläser, Marcin Marszałek, and Cordelia Schmid. A spatio-temporal descriptor based on 3D-gradients. In *British Machine Vision Conference*, pages 995–1004, sep 2008.

[6] Andrew Arnold, Ramesh Nallapati, and William W. Cohen. A comparative study of methods for transductive transfer learning. In *Proceedings of the Seventh IEEE International Conference on Data Mining Workshops*, ICDMW '07, pages 77–82, Washington, DC, USA, 2007.

[7] N FarajiDavar, T E deCampos, J Kittler, and F Yan. Transductive transfer learning for action recognition in tennis games. In *3rd International Workshop on Video Event Categorization, Tagging and Retrieval for Real-World Applications (VECTaR), in conjunction with ICCV*, 2011. Available from `http://www.ee.surrey.ac.uk/CVSSP/Publications/`.

[8] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22:1345–1359, 2010.

Supplementary Material

Figures 1 and 2 present an update of Figures 3 and 5 of [7] using an improved annotation of player bounding boxes and action labels, as discussed in Section 3. Note that the baseline ($\theta = 0$) is lower in both experiments but the best results with domain adaptation are similar to the best presented in [7]. The best results were obtained with values of θ that are higher than those shown in [7]. Note that $\theta = 0.5$ is a conservative transfer rate value that leads to a significant improvement in both experiments, and for this reason we chose this value for the experiments presented in Table 2.

Another contrast to the results in [7] is that there is a reduction in the difference between the performance of *reweight* and *trans+scale* for the same values of θ. This hints that a more complex transformation may not necessarily lead to better domain adaptation.

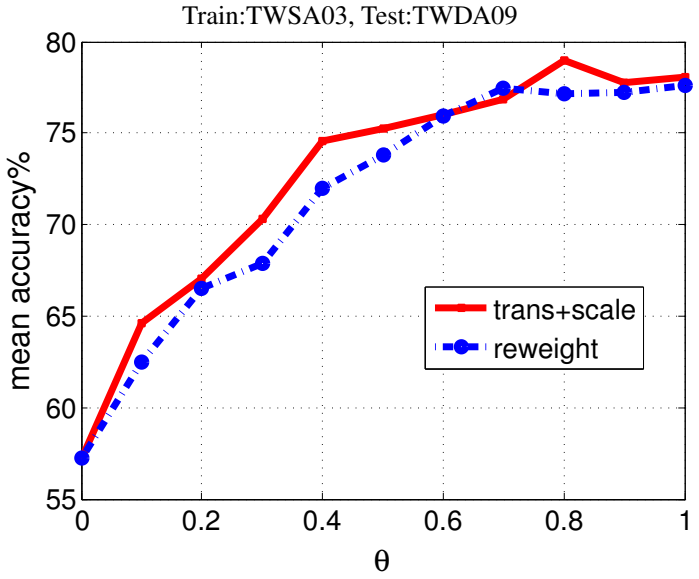

Figure 1: Mean accuracy obtained as a function of the transfer rate θ of Equation (6) for the two adaptation methods discussed in Section 2, using $\mathbf{X}_{train}^{src}$ = TWSA03 and $\mathbf{X}_{test}^{trg}$ = TWDA09.

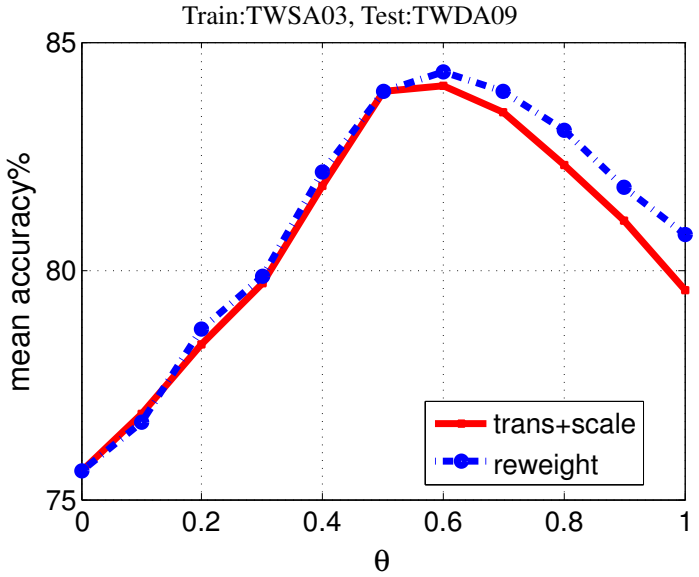

Figure 2: Same as Figure 1, but swapping the sets, i.e., $\mathbf{X}_{train}^{src}$ = TWDA09 and $\mathbf{X}_{test}^{trg}$ = TWSA03.

We also performed experiments swapping the validation (or adaptation) set with the test set. The results, shown in Table 3, present a similar pattern to that of Table 2, i.e., domain adaptation normally leads to performance improvement. One exception is observed in the experiment (g) which shows that only the *trans+scale* method gives a better performance than the baseline.

Table 3: Results (accuracy in ‰) obtained by swapping TMSA03 and TWSA03, i.e., using the men's game for validation or adaptation and the women's game for test. We follow the same format as in Table 2: baseline (top) and *reweight*|*trans+scale* (bottom).

	source	target		accuracy per class (‰)			macro
		adaptation	test	non-hit	hit	serve	average
a	TMSA03	–	TWSA03	971	427	931	776
b	TMSA03	test set	TWSA03	931|917	610|671	972|986	838|858
c	BMSB08	–	TWSA03	391	883	0	**425**
d	BMSB08	test set	TWSA03	362|440	930|873	0|0	431|438
e	BMSB08+TMSA03	–	TWSA03	966	488	887	781
f	BMSB08+TMSA03	test set	TWSA03	921|851	709|803	958|972	862|875
g	BMSB08	TMSA03	TWSA03	300|369	945|939	0|0	416|**436**