Adaptive Transductive Transfer Machines

Nazli Farajidavar  
http://personal.ee.surrey.ac.uk/Personal/N.Farajidavar

Teofilo de Campos  
http://personal.ee.surrey.ac.uk/Personal/T.Decampos

Josef Kittler  
http://www.surrey.ac.uk/cvssp/people/josef_kittler

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 naive 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 similar to \( P(G^1(\mathbf{x}^{trg})) \). Following \cite{1,2,3,5,6} 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 \( G^1(G^1(x)_i^{src}) \).

\[
G^2(G^1(x)_i^{src}) = G^2(x'_i^{src}) + \gamma b',
\]

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

\[
b'_i = \frac{\sum_k \frac{P(x'_i + b'_i | \lambda_k) \Sigma_k^{-1} (x'_i - \mu_k)}{\sum_k \frac{P(x'_i + b'_i | \lambda_k) \Sigma_k^{-1}}}}{\sum_k \frac{P(x'_i + b'_i | \lambda_k) \Sigma_k^{-1}}},
\]

where \( b'_i \) is an initial value of \( b' \), which is set to a vector of zeros. In our experiments, we ran \( b' \) 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 \( G^2(G^1(x)_i^{src}) \) following the TST transformation of \cite{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 \( \sigma \) 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).](image-url)