Cross-Domain Neural-Kernel Networks[1]

Siamak Mehrkanoon

Department of Data Science and Knowledge Engineering, Maastricht University, The Netherlands

Abstract

A novel cross-domain neural-kernel networks architecture for semi-supervised domain adaption problem is introduced. The proposed model consists of two stream neural-kernel networks corresponding to the source and target domains which are enriched with a coupling term. Each stream neural-kernel networks follows a combination of neural network layer and an explicit feature map constructed by means of random Fourier features. The introduced coupling term aims at enforcing correlations among the output of the intermediate layers of the two stream networks as well as encouraging the two networks to learn shared representation of the data from both source and target domains. Experimental results are given to illustrate the effectiveness of the proposed approaches on real-life datasets.

1 Introduction

Traditional machine learning models assume that the training and test data are drawn from the same distribution. However, in cases where there are distributional, feature and/or dimension mismatch between the two domains (training vs. test), one cannot rely on the models learned using data in one domain to predict the labels of the test data in the other domain. In addition, manual labeling of sufficient training data is costly. To overcome the burden of annotation, domain adaptation and transfer learning methods have been proposed in the literature to exploit and transfer knowledge learned from labeled data available in different but related domains to another. Depending on the availability of the labeled instances in both domains, three scenarios i.e. unsupervised, supervised and semi-supervised domain adaptation can be considered [3]. This paper introduces a novel architecture suitable for homogeneous and heterogeneous domain adaptation problems.

2 Formulation of the method

Consider two training datasets $\mathscr{D}_s = \{x_s^i\}_{i=1}^{n_s}$ with labels $\{y_s^i\}_{i=1}^{n_s}$ and $\mathscr{D}_t = \{x_t^i\}_{i=1}^{n_t}$ corresponding to the source and target domains respectively. In semi-supervised domain adaptation, one assumes that the source domain instances are fully labeled whereas the target domain data consists of both labeled and unlabeled instances ($\mathscr{D}_t = \{x_t^i\}_{i=1}^{n_t^u} \cup \{x_t^i\}_{i=1}^{n_t^l}$). Furthermore, only a small number of labeled target domain data are available $\{y_t^i\}_{i=1}^{n_t^l}$ where ($n_t^l \ll n_s$) which are representative of the entire target dataset. Given training data from source and target domains, the weight matrices $\{W_s^{(i)}\}_{i=1}^2$ and $\{W_t^{(i)}\}_{i=1}^2$, the bias vectors $\{b_s^{(i)}\}_{i=1}^2$ and $\{b_t^{(i)}\}_{i=1}^2$, the source and target stream neural-kernel networks are defined in [2] as follows:

$$h_s^{(1)} = W_s^{(1)} x_s + b_s^{(1)}, \quad h_s^{(2)} = \hat{\varphi}_s(h_s^{(1)}), \quad o_s = W_s^{(2)} h_s^{(2)} + b_s^{(2)},$$

and

$$h_t^{(1)} = W_t^{(1)} x_t + b_t^{(1)}, \quad h_t^{(2)} = \hat{\varphi}_t(h_t^{(1)}), \quad o_t = W_t^{(2)} h_t^{(2)} + b_t^{(2)},$$

respectively. $\hat{\varphi}_s(\cdot)$ and $\hat{\varphi}_t(\cdot)$ denote the explicit feature maps which here are obtained using random Fourier features and furthermore the dimensions of the feature maps are set to be equal. Let us denote

 The full paper has been published in *Pattern Recognition Letters*, Volume 125, pp. 474-480, July 2019.

all the trainable parameters by $\Theta = \left[\{W_s^{(i)}\}_{i=1}^{2}, \{b_s^{(i)}\}_{i=1}^{2}, \{W_t^{(i)}\}_{i=1}^{2}, \{b_t^{(i)}\}_{i=1}^{2}\right]$ and seek their optimal values by minimizing the following optimization problem corresponding to the proposed cross-domain neural-kernel architecture: $J(\Theta) = \gamma\Omega(\Theta) + \sum_{i=1}^{n_s} L_s(x_s^i, y_s^i) + \sum_{i=1}^{n_t^l} L_t(x_t^i, y_t^i) + \mu r(h_s^{(2)}, h_t^{(2)})$. Here $L_s(\cdot,\cdot)$ and $L_t(\cdot,\cdot)$ are the cross-entropy loss over the source and target instances. The $h_s^{(2)}$ and $h_t^{(2)}$ are defined as previously. $\Omega(\Theta)$ denotes the regularization term over the networks parameters for instance imposing l_2 norm on the weights of the networks. The last term in the cost function $J(\Theta)$ is a coupling term that aims at enforcing correlation among the learned score variables of the two stream networks. Here, the Cosine-distance between the projected source and target domains data, i.e. $r(h_s^{(2)}, h_t^{(2)}) = \frac{1}{2}(1 - \frac{h_s^{(2)T} h_t^{(2)}}{\|h_s^{(2)}\| \, \|h_t^{(2)}\|})$ is used as a regularization term that minimizes the cosine distance between projected paired source and target instances in the intermediate layers of the two stream networks (see Fig. 1). The regularization parameters γ and μ control the relative importance given to the regularization terms.

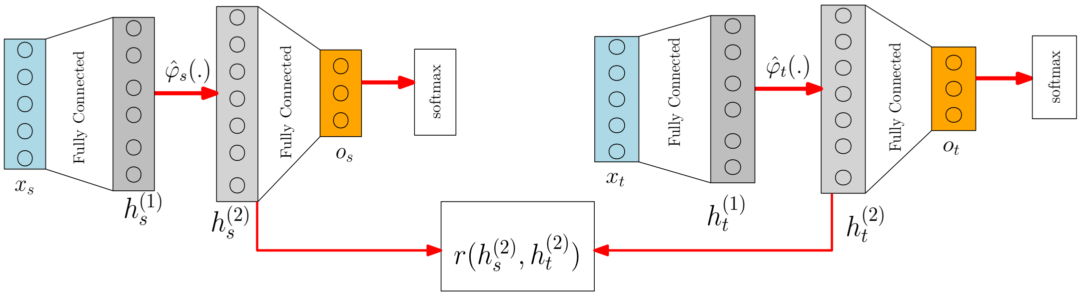

Figure 1: Cross-Domain Neural-Kernel Network architecture introduced in [2].

3 Experimental Results

The applicability of the approach is shown on the Multiple Features Dataset taken from the UCI Machine Learning repository. It contains ten classes, $(0-9)$, of handwritten digits with 200 images per class, thus for a total of 2000 images. These digits are represented in terms of six different types of features (heterogeneous features) with different dimensionality. We first select at random 70% from both source and target data in order to construct the training samples i.e. $\mathcal{D}_s$ and $\mathcal{D}_t$. All the data points of the source domain $\mathcal{D}_s$ are labeled whereas in the target domain $\mathcal{D}_t$ only two instances per class are labeled. The remaining 30% of the target domain is used as test dataset $\mathcal{D}_t^{\text{test}}$. A comparison between the accuracy of the proposed cross neural kernel network (CNKN) and two other models are tabulated in Table 1.

Table 1: The average test accuracy of the proposed CNKN model with those of HFA and RSP-KCCA models over 10 simulation runs.

	RSP-KCCA [3]					**HFA** [1]					**proposed CNKN** [2]				
Source	**fou**	**kar**	**mor**	**pix**	**zer**	**fac**	**kar**	**mor**	**pix**	**zer**	**fac**	**fou**	**mor**	**pix**	**zer**
fac	0.80	0.95	0.70	0.96	0.80	0.54	0.68	0.66	0.83	0.60	0.80	0.95	0.70	0.97	0.81
fou	0.92	0.95	0.70	0.88	0.83	0.73	0.68	0.66	0.82	0.59	0.95	0.95	0.71	0.96	0.82
kar	0.93	0.79	0.70	0.94	0.80	0.78	0.55	0.66	0.79	0.51	0.95	0.79	0.71	0.96	0.79

Acknowledgments. This work was partially supported by the Postdoctoral Fellowship of the Research Foundation-Flanders (FWO: 12Z1318N). Siamak Mehrkanoon is an assistant professor at the Department of Data Science and Knowledge Engineering, Maastricht University, The Netherlands.

References

[1] Wen Li, Lixin Duan, Dong Xu, and Ivor W Tsang. Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 36(6):1134–1148, 2013.

[2] Siamak Mehrkanoon. Cross-domain neural-kernel networks. *Pattern Recognition Letters*, 125:474–480, 2019.

[3] Siamak Mehrkanoon and Johan A.K. Suykens. Regularized semipaired kernel CCA for domain adaptation. *IEEE transactions on neural networks and learning systems*, 29(7):3199–3213, 2017.