Extended Abstract: Kernel Corresponding Projections for Orphan Targets

Katrin Ullrich and Thomas Gärtner

katrin.ullrich@iais.fraunhofer.de, thomas.gaertner@iai.uni-bonn.de
University of Bonn and Fraunhofer IAIS, Sankt Augustin, Germany

Introduction: We consider the problem of learning hypotheses for multiple targets simultaneously: targets with training data (*supervised* targets) and targets without training data (*orphan* targets). With respect to the target without training data we solve an unsupervised learning problem. For that, we propose the *kernel corresponding projections* algorithm which transfers hypotheses learned by any supervised kernel method from supervised to orphan targets. In contrast to common problems like multi-task learning, domain adaption, or inductive transfer, for the orphan target no labelled examples are available during training. Nevertheless, additional to the kernel in the instance space (i.e. the domain of the hypotheses) we assume a further target kernel is given. So-called orphan screening is an important real-world application from drug discovery for the kernel corresponding projections learning scenario introduced above. For this ligand prediction task related work can be found in Geppert et al. [2009] and Jacob and Vert [2008].

Setting: Let $\mathcal{T}$ and $\mathcal{H}$ be sets of targets and hypotheses, where $\mathcal{T}$ and $\mathcal{H}$ form reproducing kernel Hilbert spaces with kernels $k_{\mathcal{T}}$ and $k_{\mathcal{H}}$. We denote $T \subset \mathcal{T}$ and $H \subset \mathcal{H}$ the subsets of supervised targets and its hypotheses, respectively. We assume that for every $t \in T$ a hypothesis can be learned with a supervised method and the corresponding training data. This assignment is represented via the function $g : T \to H$. Furthermore, let $T_o \subseteq \mathcal{T} \setminus T$ be the orphan targets. Our aim is to find $f : T_o \cup T \to \mathcal{H}$, such that the restriction of f on T equals g. For any orphan target $t_o \in T_o$ we can then obtain its hypothesis via $f(t_o) = h_o$.

Kernel Corresponding Projections: To start with, at first we consider linear hypotheses $\mathcal{H} = \mathbb{R}^d$. We want the geometry in $\mathcal{T}$ and $\mathcal{H}$ to be similar. Therefore, the function f should meet the condition

$$\frac{\langle f(t_o), g(t)\rangle}{\|g(t)\|} \approx \frac{k_{\mathcal{T}}^{t_o,t}}{\left(k_{\mathcal{T}}^{t,t}\right)^{1/2}}$$

for any orphan target $t_o \in T_o$ and all $t \in T$. For reasons of simplicity we use the notation $k_{\mathcal{T}}^{t,t'} = k_{\mathcal{T}}(t,t')$. Choosing a least squares approach

$$h_o = f(t_o) = \operatorname*{argmin}_{h \in H'} \sum_{t \in T} \left[\left\langle h, g(t)\right\rangle \left(k_{\mathcal{T}}^{t,t}\right)^{1/2} - k_{\mathcal{T}}^{t_o,t}\|g(t)\|\right]^2$$

and exploiting the convexity of the optimization problem we obtain

$$f(t_o) = \Big[\sum_{t\in T} g(t)k_{\mathcal{T}}^{t,t}g(t)^T\Big]^{\dagger}\Big[\sum_{t\in T} g(t)\|g(t)\|\big(k_{\mathcal{T}}^{t,t}\big)^{1/2}k_{\mathcal{T}}^{t_o,t}\Big]$$

as solution of the *linear corresponding projections* problem. If κ denotes the upper bound for calculating $k_{\mathcal{T}}$ the overall cost for linear corresponding projections is $\mathcal{O}(|T|d^2\kappa)$.

Suppose that additionally $\|g(t)\| = 1$ and $\langle g(t), g(t')\rangle = 0$ hold true for all $t \in T$. This assumption can be justified well for orphan screening utilizing *molecular fingerprints* (vectorial representation of molecules, which are often binary and sparse). Then we can calculate

$$f(t_o) = \big([GDG^T]^T[GDG^T]\big)^{-1}[GDG^T][G\tilde{D}1] = \sum_{i=1}^{n} g(t_i)\frac{k_{\mathcal{T}}^{t_o,t}}{\big(k_{\mathcal{T}}^{t,t}\big)^{1/2}}$$

for $G^T = (g(t_1)|\ldots|g(t_n))$ and D as well as $\tilde{D}$ appropriate diagonal matrices. This simplified version of linear corresponding projections is also faster with running time in $\mathcal{O}(|T|d\kappa)$.

In a more general scenario we investigate an arbitrary reproducing kernel Hilbert space $\mathcal{H}$. Let K be the gram matrix $[K]_{i,j} = k_{\mathcal{H}}(x_i, x_j), x_i, x_j \in X$, with respect to the unified training examples X of all supervised targets. In case the supervised kernel method for the determination of hyptheses from T fulfills the preconditions of the Representer Theorem, both the hypotheses h_o of orphan targets t_o and every learned hypothesis in $H = g(T)$ can be written with coefficients $\alpha_o \in \mathbb{R}^{|X|}$ or $\alpha_i \in \mathbb{R}^{|X|}$, respectively. Hence, $f(T_o \cup T) = \text{span}(H)$ follows by construction. Utilizing this properties the regularized version of corresponding projections' optimization problem in general reproducing kernel Hilbert spaces is

$$f(t_o) = \underset{\alpha_0\in\mathbb{R}^{|X|}}{\text{argmin}}\ \nu\alpha_o^T K\alpha_o + \sum_{i=1}^{n}\Big[\alpha_o^T K\alpha_i\big(k_{\mathcal{T}}^{t_i,t_i}\big)^{1/2} - k_{\mathcal{T}}^{t_o,t_i}\big(\alpha_i^T K\alpha_i\big)^{1/2}\Big]^2$$

for $t_o \in T$ and $T = \{t_1, \ldots, t_n\}$. Its solution can be deduced anologously to the linear case as

$$f(t_o) = \Big[\nu K + \sum_{i=1}^{n} K\alpha_i k_{\mathcal{T}}^{t_i,t_i}\alpha_i^T K\Big]^{-1}\Big[\sum_{i=1}^{n} K\alpha_i\big(\alpha_i^T K\alpha_i\big)^{1/2}\big(k_{\mathcal{T}}^{t_i,t_i}\big)^{1/2}k_{\mathcal{T}}^{t_o,t_i}\Big].$$

References

[2008] Jacob, L., Vert, J.-P.: Protein-ligand interaction prediction: an improved chemogenomics approach. Bioinformatics (Oxford, England), 24, 2149–56 (2008)

[2009] Geppert, H., Humrich, J., Stumpfe, D., Gärtner, T., Bajorath, J.: Ligand Prediction from Protein Sequence and Small Molecule Information Using Support Vector Machines and Fingerprint Descriptors. J. Chem. Inf. Model., Vol. 49, 767–779 (2009)