Semi-supervised Learning with Explicit Relationship Regularization

Kwang In Kim[1], James Tompkin[2] Hanspeter Pfister[2] Christian Theobalt[3]
[1]Lancaster University [2]Harvard SEAS [3]MPI for Informatics

The goal of semi-supervised learning is to learn a function f which maps from an input space M to a target space N given a sparse labeling on data points. The lack of labels is compensated for by exploiting unlabeled data points to provide additional information, e.g., on the geometry of and/or probability distribution on M, from which the data are generated. Regularization tries to measure and limit the complexity of proposed f solutions by preferring smaller training errors and placing restrictions on smoothness.

In many applications, the target space N has a structure which may be defined implicitly or, in some applications, explicitly through pair-wise similarity or dissimilarity potentials. However, current regularization methods operate only on the function itself, and do not *explicitly* consider the potentially rich informative structure of N as something which can be used for regularization. In this paper, we explore regularizing this structure of N or the *relationships* between entities in N.

One example that benefits from this principle occurs when *relationship labels* are provided. In semi-supervised or *constrained* spectral clustering [2, 3, 5], the labels are provided not on the underlying cluster assignment function f but on the binary relationships k between the function evaluations, as *must-link* or *cannot-link* labels. These are exploited by applying conventional regularization on f with the condition that the constraints are satisfied. However, in this case, the relationship itself can also be a natural object to regularize (Fig. 1): For instance, if (x_1, x_3) *must link*, i.e., if they belong to the same cluster, then a *relationship function* k on N is defined such that $k(f(x_1), f(x_3))$ is positive. For point x_2, which is close to x_1 in M, we expect the relationship function $k(f(x_2), f(x_3))$ is *similar* to $k(f(x_1), f(x_3))$ and therefore, to be positive also.

In general, the relationship itself is not formally defined or observed; however, in many applications, certain relationships are manifested through a smooth function, where the number of arguments corresponds to the relationship degree, e.g., a distance metric is a function of two arguments. k can be defined either directly from the data or from labels; either way, once the relationship is defined, regularization is independent of the existence of labels and therefore applies generally to any learning problem.

We develop this intuition to a new regularization functional which extends the well-established harmonic energy functional and p-th iterated Laplacian semi-norm [1, 4, 6]. In our framework, a relationship is represented by an n-th order *relationship function* k defined on N^n, where n is application specific. For instance, these relationships can represent similarity between pairs or n-tuples of entities or, in general, any non-metric relationships, e.g., *left of* or *on top of* for generating topographic maps. Specifically, for semi-supervised classification and spectral data embedding, we use a Gaussian similarity relationship function k:

$$k(f(x), f(x')) = \exp\left(-\frac{\|f(x) - f(x')\|^2}{\sigma_f^2}\right) \quad (1)$$

where $\sigma_f^2 > 0$. Our new regularizer on M is then defined as:

$$\mathcal{R}_k(f) = \int_M \int_M \|\nabla f^* h_{x'}(x)\|^2_{T_x^*} dV(x) dV(x'), \quad (2)$$

where $f^* h_{x'}(x) := k(f(x), f(x'))$.

For each fixed x' in the function, $f^* h_{x'}(x)$ encodes the relationship between $f(x)$ and $f(x')$, and since $f^* h_{x'}(x)$ is a function of a single variable $x \in M$, $\nabla f^* h_{x'}(x)$ lies in $T_x^*(M)$. This implies that the inner integral measures the variation of $f^* h_{x'}(x)$ that corresponds to pair-wise relations between the fixed x' and each value of x. In particular, when $k(a,b)$ measures the Euclidean distance between a and b, the inner integral is zero only when the distances between each pair $f(x)$ and $f(x')$ are identical for all $x \in M$. This does not require that k is zero. Then, the outer integral averages x' over the entire M.

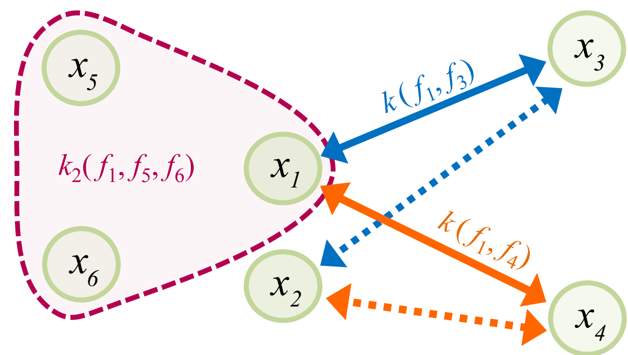

Figure 1: If two data points x_1 and x_2 are close on the domain M of f, then conventional regularizers enforce that the corresponding function values f_1 and f_2 in co-domain N of f are similar ($f_i \equiv f(x_i)$). We assume that relationships between pairs of function evaluations f_i and f_j are represented by smooth functions $k(f_i, f_j)$, e.g., a similarity measure. Our regularizer explicitly enforces that $k(f_1, f_j)$ and $k(f_2, f_j)$ are similar for any j. For instance, if $k(f_1, f_3)$ is large as f_1 and f_3 are similar, but $k(f_1, f_4)$ is small as f_1 and f_4 are dissimilar (solid arrows), then our algorithm enforces that $k(f_2, f_3)$ and $k(f_2, f_4)$ are large and small, respectively (dotted arrows), as x_1 and x_2 are close in M. The same principle applies to high-order relationships: if $k_2(f_1, f_5, f_6)$ represents a ternary relationship, e.g., a third-order correlation, the similarity of $k_2(f_1, f_5, f_6)$ and $k_2(f_2, f_5, f_6)$ is enforced.

In many practical applications, M is not directly observed but indirectly represented as a sampled point cloud $\mathcal{X} = \{x_1, \ldots, x_u\}$ and accordingly, we approximate $\mathcal{R}_k$ based on evaluations of f on $\mathcal{X}$. For a given set of u data points, our final regularization functional is defined as:

$$\widetilde{\mathcal{R}}_k(\mathbf{f}) = tr[K^\top L K], \quad (3)$$

where $tr[\cdot]$ is the trace, $K_{ij} := k(f(x_i), f(x_j))$, and $L(u \times u)$ is the graph Laplacian.

Our new regularizer can easily be combined with existing regularizers on point clouds. In experiments, we demonstrate that our regularizer significantly improves the performance of state-of-the-art algorithms in semi-supervised classification and in spectral data embedding for constrained clustering and dimensionality reduction.

[1] Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2005.

[2] Z. Li, J. Liu, and X. Tang. Constrained clustering via spectral regularization. In *Proc. CVPR*, pages 421–428, 2009.

[3] S. Rangapuram and M. Hein. Constrained 1-spectral clustering. *JMLR W&CP (Proc. AISTATS)*, 22:1143–1151, 2012.

[4] Florian Steinke, Matthias Hein, and Bernhard Schölkopf. Nonparametric regression between general Riemannian manifolds. *SIAM Journal on Imaging Sciences*, 3(3):527–563, 2010.

[5] X. Wang and I. Davidson. Flexible constrained spectral clustering. In *Proc. SIGKDD*, pages 563–572, 2010.

[6] X. Zhou and M. Belkin. Semi-supervised learning by higher order regularization. *JMLR W&CP (Proc. AISTATS)*, pages 892–900, 2011.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.