Semi-supervised Learning based on Joint Diffusion of Graph Functions and Laplacians

Kwang In Kim, University of Bath
k.kim@bath.ac.uk

Abstract: In existing anisotropic diffusion-based semi-supervised learning approaches, anisotropic graph Laplacian is estimated based on (potentially noisy) function evaluations. We propose to regularize the graph Laplacian estimates.

We develop a framework that regularizes the Laplace-Beltrami operators on Riemannian manifolds, and discretize it to a regularizer on diffusivity operators on graphs.

Isotropic Laplace-Beltrami operator Δ^g on a Riemannian manifold (M, g) with a metric g is a second-order differential operator:

$$\Delta^g f = \nabla^{g*}\nabla^g f,$$

where ∇^g and ∇^{g*} are the gradient and divergence operators, respectively.

Δ^g generates the diffusion process on M:

$$\frac{\partial f}{\partial t} = -\Delta^g f.$$

Anisotropic Laplace-Beltrami operator Δ^D is defined based on a symmetric positive definite *diffusivity operator* D:

$$\Delta^D f = \nabla^{g*} D \nabla^g f.$$

D controls the strength and direction of diffusion at each point x on M.

Regularizing Δ^D by regularizing D as a surrogate:

1) *Kernel-based Δ^g representation [HAL05]*: A consistent kernel-based estimate $\Delta_h^g f$:

$$[\Delta_h^g f](x) = \frac{1}{h^2}\left(f(x) - \frac{[A_h^g(x)f]}{d_h(x)}\right), \text{ where } [A_h^g(x)f] = \int_M k_h(x,y)f(y)dV(x),$$

$d_h(x) = [A_h^g(x)\mathbf{1}]$, $dV(x) = \sqrt{|\det(\mathbf{g})|}dx$ ($\mathbf{g}$: g's coordinate matrix), and

$$k_h(x,y) = \begin{cases} \frac{1}{h^m}k(\|i(x)-i(y)\|^2_{\mathbb{R}^m}, h^2) & \text{if } \|i(x)-i(y)\|_{\mathbb{R}^m} \leq h \\ 0 & \text{otherwise} \end{cases}$$

with $k(a,b) = \exp(-a/b)$ and i being the embedding of M into $\mathbb{R}^m$.
⇒ The spatial variation of Δ^g is entirely determined by the metric g.

2) *Equivalence of metric and diffusivity operator on manifolds*:

Proposition 1 (KTP15) *The anisotropic Laplacian operator Δ^D on a compact Riemannian manifold (M, g) is equivalent to the Laplace-Beltrami operator $\Delta^{\overline{g}}$ on $(M, \overline{g})$ with a new metric $\overline{g}$ depending on D.*

When the diffusivity operator D is uniformly positive definite, $\overline{g}$ is explicitly obtained as $c(x)\overline{\mathbf{g}}(x) = \mathbf{g}(x)\mathbf{D}^{-1}(x)$, where $\overline{\mathbf{g}}(x)$ and $\mathbf{D}(x)$ are the coordinate matrices of $\overline{g}$ and D at each point x, and $c(x) = \sqrt{\det \mathbf{g}(x)}/\sqrt{\det \overline{\mathbf{g}}(x)}$.

⇒ Anisotropic diffusion on (M, g) is isotropic diffusion on M with a new metric $\overline{g}$.

Discretization: On a weighted graph (X, E, W) with nodes $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_u\}$, edges $\{E_i\} = \{e_{ij}\} \subset X \times X$, non-negative similarities $w_{ij} := w(e_{ij}) \in W$, and the space of functions $H(E_i)$ on E_i, the *local graph diffusivity operator* $D_i : H(E_i) \to H(E_i)$ is defined as:

$$D_i := \sum_{\{j:(j,i)\in E_i\}} q_{ij}\mathbf{b}_{ij} \otimes \mathbf{b}_{ij} \Leftrightarrow [D_i S](e_{ij}) = q_{ij}\mathbf{b}_{ij}\langle \mathbf{b}_{ij}, S\rangle, \forall S \in H(E_i),$$

$\otimes$: the tensor product; basis function $\mathbf{b}_{ij} := \mathbf{1}_{ij} \in H(E)$.

The *anisotropic graph Laplacian* is defined as:

$$[L^D f](\mathbf{x}_i) := [\nabla_i^* D_i \nabla_i f](\mathbf{x}_i) = \left(\frac{1}{d_i}\sum_{j=1}^{u} w_{ij}q_{ij}\right) f(\mathbf{x}_i) - \frac{1}{d_i}\sum_{j=1}^{u} w_{ij}q_{ij}f(\mathbf{x}_j) \quad (1)$$

with $d_i = \sum_j w_{ij}$,

$$w_{ij} = \begin{cases} k\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2, h^2\right) & \text{if } (j,i) \in E_i \text{ or } (i,j) \in E_j \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

$$q_{ij} = k\left(-|f(\mathbf{x}_i) - f(\mathbf{x}_j)|^2, h'\right), \quad (3)$$

and h and h': hyper-parameters.

Proposition 2 (The convergence of L^D to $\Delta^{\overline{g}}$) *Assume that (M, g) is an n-dimensional Riemannian submanifold of $\mathbb{R}^m$ and $(M, \overline{g})$ is a new manifold with an updated metric $\overline{g}$. Let $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_u\}$ be a sample from a compactly supported, uniform distribution p on M and the coefficients $\{q_{ij}\}$ of the graph diffusivity operator D are given as*

$$h^m q_{ij} = k(\|\mathbf{x}_i - \mathbf{x}_j\|^2_{\mathbb{R}^m}, h^2) + k(-\|i^{-1}(\mathbf{x}_i) - i^{-1}(\mathbf{x}_j)\|^2_{\overline{g}^2}, h^2).$$

Then, Δ_D converges to $\Delta_{\overline{g}}$ almost surely as $u \to \infty$, $h \to 0$, and $uh^{m+2}/\log(n) \to \infty$.

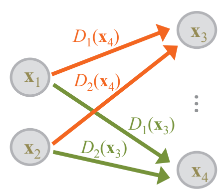
Joint diffusion of f and D (explicit Euler approximation): $\mathbf{f} = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_u)]^\top$

$$\mathbf{f}(t+\delta) = (I - \delta L^D(t))\mathbf{f}(t), \qquad \mathbf{M}(t+\delta) = (I - \delta L^D(t))\mathbf{M}(t). \quad (4)$$

Algorithm *AND*: Semi-supervised learning using combined diffusion of function and Laplacian.

Input: Data points $X = \{\mathbf{x}_1, \ldots, \mathbf{x}_u\} \subset \mathbb{R}^m$;
labels $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_l\} \subset \mathbb{R}^c$; hyper-parameters h, h', T_1 and T_2
Output: Diffused labels $\mathbf{f}$.

```
Build an isotropic graph Laplacian L (Eqs. 1-3 with q_ij = 1)
For t = 1, ..., T_1 do
    Update f(t) (Eq. 4).
    At each 20-th iteration,
        Update L_D(t) (Eqs. 1-3);
        For t' = t, ..., t + T_2 do
            Update L^D(t') (Eq. 4);
        end
    Assign labels to f^t and L^D(t);
    Normalize L^D(t);
end
```



We regularize the Laplacian on manifolds by enforcing the smoothness of the corresponding diffusivity operator. Instantiating this for graphs, our algorithm enforces the smoothness of the graph diffusivity functions $\{D_1, \ldots, D_u\}$: If $\mathbf{x}_1$ and $\mathbf{x}_2$ are close, the corresponding diffusivity functions $D_1(\cdot)$ and $D_2(\cdot)$ should be *similar*, e.g., $D_1(\mathbf{x}_3) \sim D_2(\mathbf{x}_3)$ and $D_1(\mathbf{x}_4) \sim D_2(\mathbf{x}_4)$.

Semi-supervised learning results: The three best results for each dataset are marked with **boldface blue**, *plain green*, and plain orange fonts, respectively. LNP [WaZ06] requires explicitly calculating the Euclidean distances between data points, and so it cannot be directly applied to *MPEG7* and *SwL* (Swedish leaf) datasets. The final Avg. % column shows the mean percentage difference from the best result across all datasets, where 100% would indicate that particular technique was best across all datasets.

LGC: local and global consistency [ZBL03]; $p-L$: powers of Laplacian [ZhA06]; LNP: linear neighborhood propagation [WaZ06]; Iso: isotropic diffusion; AL: anisotropic linear diffusion [SMC08]; AN: anisotropic nonlinear diffusion [KTP15]; ALD and AND: linear and nonlinear diffusion of the diffusivity operators, respectively.

	USPS	COIL	COIL$_2$	PCMAC	Text	ETH	Cal101	MNIST	MPEG7	SwL	Avg. %
LGC	7.03	7.65	1.53	12.48	25.59	10.92	53.41	4.50	2.93	**2.29**	159.72
$p-L$	4.07	9.95	0.80	*9.56*	**22.33**	10.05	53.41	*3.66*	2.90	**2.29**	124.21
LNP	8.70	**6.31**	2.53	13.52	36.30	13.34	65.52	5.44	N/A	N/A	231.72
Iso	7.52	8.24	0.45	12.34	28.72	11.62	53.22	5.08	3.33	2.45	135.11
AL	6.00	8.45	0.79	13.54	27.01	11.27	52.36	4.92	3.11	2.54	140.30
AN	4.35	7.53	*0.39*	13.46	26.87	*10.04*	**52.23**	3.72	2.87	2.45	*115.71*
ALD	*4.09*	7.63	0.76	10.98	23.89	**9.96**	*52.26*	4.72	*2.71*	2.35	123.82
AND	**3.28**	*6.76*	**0.33**	**8.96**	*23.09*	10.11	*52.26*	**3.65**	**2.69**	2.38	**101.60**

Acknowledgement: Kwang In Kim thanks EPSRC EP/M00533X/1.

Reference:

ZBL03 D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, *Learning with local and global consistency*, In Proc. NIPS, 2003.

ZhA06 T. Zhang and R. Ando, *Analysis of spectral kernel design based semi-supervised learning*, In Proc. NIPS, 2006.

SMC08 A. D. Szlam, M. Maggioni, and R. R. Coifman, *Regularization on graphs with function-adapted diffusion processes*, JMLR, 2008.

WaZ06 F. Wang and C. Zhang, *Label propagation through linear neighborhoods*, Proc. ICML, 2006.

KTP15 K. I. Kim, J. Tompkin, H. Pfister, and C. Theobalt, *Context-guided diffusion for label propagation on graphs*, Proc. ICCV, 2015.

HAL05 M. Hein, J.-Y. Audibert, and U. von Luxburg, *From graphs to manifolds - weak and strong pointwise consistency of graph Laplacians*, In Proc. COLT, 2005.