Indefinite Topological Kernels
Kernel Topologici Non Definiti

Tullia Padellini and Pierpaolo Brutti

Abstract Topological Data Analysis (TDA) is a recent and growing branch of statistics devoted to the study of the shape of the data. Motivated by the complexity of the object summarizing the topology of data, we introduce a new topological kernel that allows to extend the TDA toolbox to supervised learning. Exploiting the geodesic structure of the space of Persistence Diagrams, we define a geodesic kernel for Persistence Diagrams, we characterize it, and we show with an application that, despite not being positive semi–definite, it can be successfully used in regression tasks.
Abstract Topological Data Analysis (TDA) è una branca della statistica volta allo studio della "forma" dei dati. Data la complessità delle summaries topologiche, introduciamo una nuova famiglia di kernels per estendere TDA a problemi di apprendimento supervisionato. Sfruttando la geodesica dello spazio delle summaries topologiche, in questo lavoro definiamo un kernel geodesico, lo caratterizziamo e mostriamo con un applicazione le sue performance in un problema di regressione.

Key words: Topological Data Analysis, Kernel Methods, Indefinite Kernels

1 Introduction to Topological Data Analysis (TDA)

Topological Data Analysis TDA is a new area of research in statistics consisting of techniques aimed at recovering the topological structure of data [1, 3]. The interest in the topological structure of data stems from the immediate interpretation of the characterization provided by topological invariants: 0-dimensional features represent connected components, 1-dimensional features represent loops, 2-dimensional are voids and so on. These are all quantity of interest in statistical analysis, as for

Tullia Padellini, Pierpaolo Brutti
Sapienza Universit di Roma, Piazzale Aldo Moro, 5, 00185 Roma,
e-mail: {tullia.padellini, pierpaolo brutti}@uniroma1.it

example connected components correspond to clusters [2] and loops represent periodic structures [7].

If data come in the form of point–cloud $\mathbb{X}_n = \{X_1, \dots, X_n\}$, however, it is not possible to compute such invariants directly. A point–cloud $\mathbb{X}_n$, in fact, has a trivial topological structure, as it is composed of as many connected components as there are observations and no higher dimensional features. `TDA` provides a framework for estimating the topological structure of $\mathbb{X}_n$ by enriching it, encoding data into a filtration. The most common way to do so is to grow each observation point X_i into a ball

$$B(X_i, \varepsilon) = \left\{ x \,\middle|\, d_{\mathbb{X}}(x, X_i) \leq \varepsilon \right\},$$

of fixed radius ε.

As the resolution ε changes, the topological structure of the *cover*

$$\mathbb{X}_n^{\varepsilon} = \bigcup_{i=1}^{n} B(X_i, \varepsilon),$$

changes as well. When ε is very small, $\mathbb{X}_n^{\varepsilon}$ is topologically equivalent to $\mathbb{X}_n$; as ε grows, however, balls of the cover start to intersect, "giving birth" to loops, voids and other topologically interesting structures. At some point, when connected components merge, loops are filled and so on, these structures start to "die". Eventually when ε reaches the diameter of the data $\mathbb{X}_n$, $\mathbb{X}_n^{\varepsilon}$ is topologically equivalent to a point ball, and again retains no information. The "lifetime" of the generic i-th feature can be represented by a "birth–time" b_i representing the first value ε for which the i-th feature appears in the data, and a "death–time" d_i corresponding to when the feature disappear, i.e. the first value ε for which $\mathbb{X}_n^{\varepsilon}$ does not retain the i-th feature anymore. Birth and death times for all the features in the cover are then summarized in the *Persistence Diagram* $D = \{(b_i, d_i), i = 1, \dots m\}$.

Points that are close to the diagonal have a "short life", in the sense that the features they represent appear and disappear almost immediately and hence may be neglected; on the other hand the "longer" a feature lives, the more important it is in characterizing the structure of $\mathbb{X}_n$.

The space of Persistence Diagram $\mathscr{D}$ is a metric space when endowed with the Wasserstein distance.

Definition 1 (Wasserstein Distance between Persistence Diagrams). Given a metric d, called *ground distance*, the Wasserstein distance between two persistence diagrams D and D' is defined as

$$W_{d,p}(D, D') = \left[\inf_{\gamma} \sum_{x \in D} d\big(x, \gamma(x)\big)^p \right]^{\frac{1}{p}},$$

where the infimum is taken over all bijections $\gamma : D \mapsto D'$.

Depending on the choice of the ground distance d, Definition 1 defines a family of metrics; we focus on the L^2–norm, especially in the case $p = 2$, for which [10] proved that $W_{L^2,2}$ is a geodesic on the space of persistence diagrams.

1.1 Geodesic Topological Kernels

As most statistical learning tools are defined for inner product spaces, the metric structure of the space of persistence diagrams may be limiting, we thus turn to a kernel approach. Roughly speaking a kernel K on a space $\mathscr{M}$ is a symmetric binary function $K : \mathscr{M} \times \mathscr{M} \mapsto \mathbb{R}^+$ which represent a measure of similarity between two elements of $\mathscr{M}$. As every kernel is associated to an inner product space [9], we can use them to implicitely define an inner product space in which it is possible to perform most statistical tasks, from classification to testing, through regression.

Previous attempts in this direction (such as [8]) built kernels on persistence diagrams by considering each point of the diagram individually, thus loosing the structure of the object. In order to consider the diagram as a whole, we propose a kernel which, being based on the Wasserstein distance, preserves information about how points in the diagram are related to each others.

One popular family of kernels for a geodesic metric space $(\mathbb{X}, d)$ is the *exponential kernel*

$$k(x,y) = \exp\left\{d(x,y)^p/h\right\} \qquad p,h > 0$$

where $h > 0$ is the bandwidth parameter; for $p = 1$ this is the Laplacian kernel and for $p = 2$ this is the Gaussian kernel. It is straightforward to use this class to define a *Topological kernel* to be used for statistical learning.

Definition 2 (Geodesic Topological Kernel). Let $\mathscr{D}$ be the space of persistence diagrams, and let $h > 0$, then the Geodesic Gaussian Topological (GGT) kernel $K_{\mathrm{GG}} : \mathscr{D} \times \mathscr{D} \mapsto \mathbb{R}^+$ is defined as

$$K_{\mathrm{GG}}(D,D') = \exp\left\{\frac{1}{h} W_{L^2,2}(D,D')^2\right\} \qquad \forall D,D' \in \mathscr{D}.$$

Analogously, the Geodesic Laplacian Topological Kernel (GLT), K_{GL} is defined as:

$$K_{\mathrm{GL}}(D,D') = \exp\left\{\frac{1}{h} W_{L^2,2}(D,D')\right\} \qquad \forall D,D' \in \mathscr{D}.$$

As opposed to their euclidean counterparts, the Geodesic Laplacian and Gaussian kernels are not necessarily positive definite; as shown in [4], in fact, a Geodesic Gaussian kernel on a metric space is positive definite only if the space is flat, but this is not the case for the space of Persistence Diagram, which has been proved to be curved [10].

2 Application - Fullerenes

Buckyballs fullerenes are spherical pure carbon molecules whose main trait is that atoms' linkage can form either pentagons or hexagons. We will show that our topological kernel can be exploited to predict the Total Strain Energy (measured in Ev) of a molecule from the shape of the molecule, as our Topological Kernel allows us to use Persistence Diagrams as covariates. given a sample $\{X_1,\dots,X_n\}$ of Fullerenes we model their Total Strain Energy, Y as a function of their Persistence Diagrams $\{D_1,\dots,D_n\}$:

$$Y_i = m(D_i) + \varepsilon_i$$

where ε_i is the usual 0–mean random error. As in standard nonparametric regression, we can estimate the regression function $m(\cdot)$ with the Nadaraya–Watson estimator[5], which does not require a positive definite kernel.

Moreover, in order to take into account the group structure naturally induced by the isomers, we also considered a model with a fixed group intercept, i.e:

$$Y_{ij} = \alpha_j + m(D_{ij}) + \varepsilon_{ij},$$

where D_{ij} denotes the persistence diagram of the $i^{\texttt{th}}$ isomer of the $j^{\texttt{th}}$ molecule. We fit the resulting partially linear model using Robinson's trimmed estimator, as detailed in [6]. We fit the models using data from $n = 535$ molecules of 10 different types of Fullerenes. For each molecule, the data (freely available at http://www.nanotube.msu.edu/fullerene/fullerene-isomers.html consists of the coordinates of the atoms taken from Yoshida's Fullerene Library and then re–optimized with a Dreiding–like forcefield. We focus on features of dimension 1, which recover the structure of the molecule; as we can see from Figure 1, loops in the diagrams are, in fact, clearly clustered around two centers, which represent the pentagons and the hexagons formed by the carbon atoms. Interestingly enough, the Wasserstein distance and, hence, both the geodesic kernels, fully recover the class structure induced by the isomers, as we can see in Figure 3.

	Geodesic Gaussian Kernel	Geodesic Laplacian Kernel
Nonparametric regression	339.89	342.14
Semiparametric regression	1049.02	331.04

Table 1 Residual Sum of Squares.

After choosing the bandwidth h via Leave–One–Out cross validation, we compare the different models in terms of Residual Sum of Squares (RSS). As we can see from Table 1, the two kernels yield similar results when used in a fully nonparametric estimator, while the Laplacian kernel performs better when adding the group intercept to the model. This can be understood by looking at the kernel matrices (Figure 3); the Gaussian Kernel has a sharper block structure than the Laplace Kernel, which makes it better at discriminating the 10 molecule classes. However, when

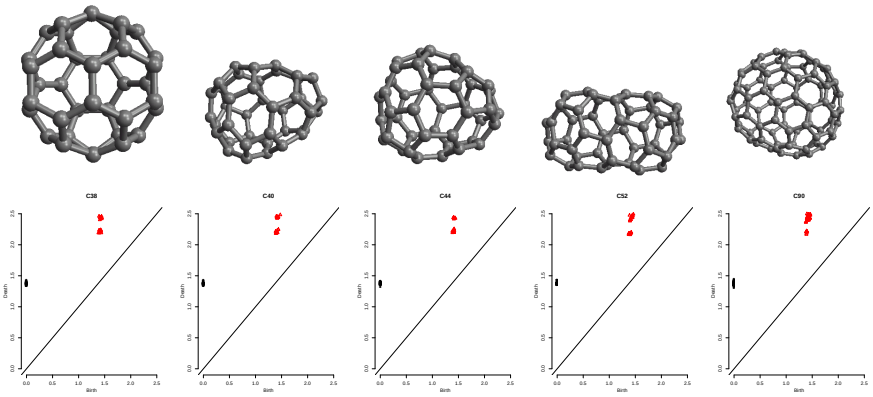

Fig. 1 Topological configurations of some fullerenes (top) and corresponding persistence diagrams (bottom). From left to right: C38(C2v), C40(C1), C44(C1), C52(C2), C90(C1).

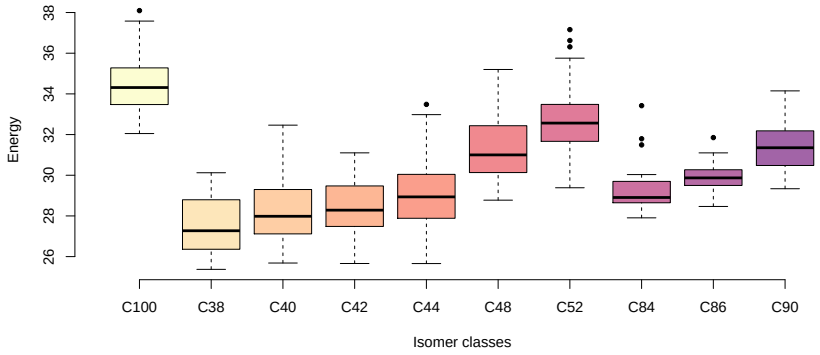

Fig. 2 Energies for the 10 different classes of isomers. It is worth noticing that Fullerenes with higher numbers of atoms do not necessarily have higher energy.

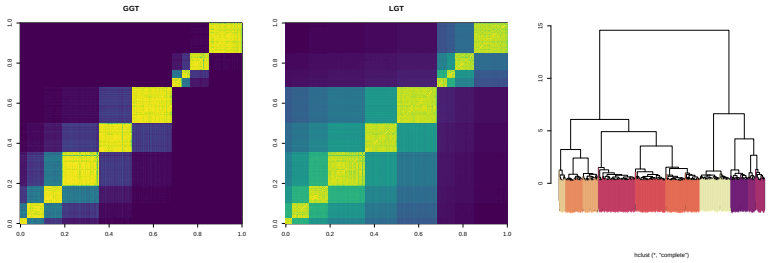

Fig. 3 Kernel Matrix for the Geodesic Gaussian Kernel (left), Geodesic Laplacian Kernel (center), Hierarchical Clustering built from the Wasserstein distance with complete linkage (right). Colors represent the different isomer classes as shown in Figure 2.

the group structure is taken into account by the model itself, this clustered structure leads to worse prediction.

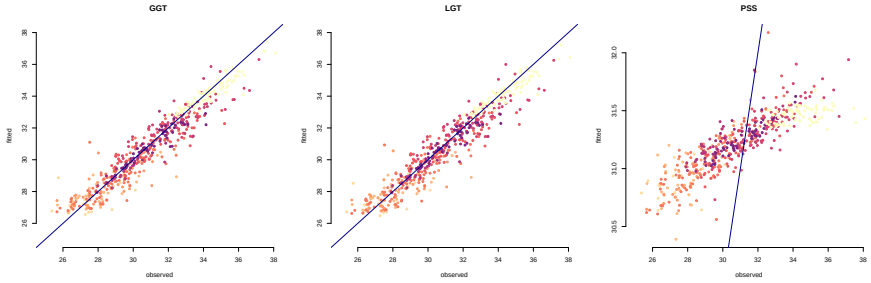

Fig. 4 Observed vs fitted plot for the fully nonparametric model fitted with the Geodesic Gaussian (left), Geodesic Laplacian (center) and the Persistence Scale Space kernel (right). Colors represent the different isomer classes as shown in Figure 2.

Finally, we compare the performance of our geodesic kernels with the Persistence Scale Space kernel K_{PSS}; as we can clearly see from the fitted-vs-observed plots in Figure 4, the positive definiteness of the PSS kernel does not result in more accurate prediction, as both K_{GG} and K_{LG} outperform it.

References

1. Carlsson, G.: Topology and data. Bulletin of the American Mathematical Society **46**(2), 255–308 (2009)
2. Chazal, F., Guibas, L.J., Oudot, S.Y., Skraba, P.: Persistence-based clustering in riemannian manifolds. Journal of the ACM (JACM) **60**(6), 41 (2013)
3. Fasy, B.T., Kim, J., Lecci, F., Maria, C., Rouvreau, V.: Tda: Statistical tools for topological data analysis. Availabl e at https://cran. r-project. org/web/packages/TDA/index. html (2014)
4. Feragen, A., Lauze, F., Hauberg, S.: Geodesic exponential kernels: When curvature and linearity conflict. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3032–3042 (2015)
5. Härdle, W.K., Müller, M., Sperlich, S., Werwatz, A.: Nonparametric and semiparametric models. Springer Science & Business Media (2012)
6. Li, Q., Racine, J.S.: Nonparametric econometrics: theory and practice. Princeton University Press (2007)
7. Perea, J.A., Harer, J.: Sliding windows and persistence: An application of topological methods to signal analysis. Foundations of Computational Mathematics **15**(3), 799–838 (2015)
8. Reininghaus, J., Huber, S., Bauer, U., Kwitt, R.: A stable multi-scale kernel for topological machine learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4741–4748 (2015)
9. Scholkopf, B., Smola, A.J.: Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT press (2001)
10. Turner, K., Mileyko, Y., Mukherjee, S., Harer, J.: Fréchet means for distributions of persistence diagrams. Discrete & Computational Geometry **52**(1), 44–70 (2014)