IMÁGENES DE PERSISTENCIA

Iris Natalia García Velázquez
natalia17g@outlook.es
Universidad Juárez Autónoma de Tabasco
División Académica de Ciencias Básicas

1. RESUMEN

Al estudiar la *homología persistente* podemos identificar las características de un espacio topológico conforme se van generando sus grupos de homología. A su vez, esta información se codifica en un *diagrama de persistencia* (DP), un multiconjunto de puntos en el plano $\mathbb{R}^2$ y cuya interpretación nos indica el "nacimiento" y "muerte" de algunos elementos de los grupos de homología. Una *imagen de persistencia* (IP) se constituye de una malla sobre el DP, donde cada celda equivale a un pixel.

2. ANTECEDENTES

Estudiar la homología es un paso básico para poder entender a los DP e IP, conceptos como el siguiente son la cimentación con la que podemos permitirnos hablar de IP. [3,4]

Complejo simplicial K

Es una colección finita de *simplejos* que cumplen

1. Las caras de los simplejos de K, pertenecen a K.
2. Para dos simplejos en K, la intersección es una cara para cada simplejo.

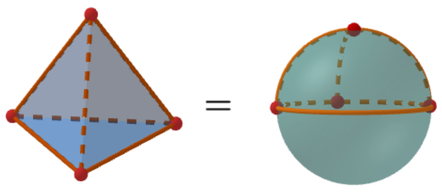


Figura 1: *La esfera es un complejo simplicial.*

Asociados a cada complejo simplicial, tenemos los grupos de homología los cuales miden la cantidad de "agujeros" p-dimensionales que hay en el espacio, [3].

Homología de un complejo

El p-ésimo *grupo de homología* del complejo simplicial K se define como el grupo cociente

$$H_p(A) = \frac{\text{Ker}(\partial_p)}{\text{Im}(\partial_{p+1})},$$

con coeficientes en $\mathbb{Z}_2$.

3. HOMOLOGÍA PERSISTENTE

Para obtener un DP se debe pasar por una *filtración*, existen métodos distintos para realizarla, en este caso daremos a conocer la filtración de Vietoris-Rips. Primero damos una construcción del *complejo de Vietoris-Rips* en un espacio euclideano.

Bola cerrada

Tómese $x \in \mathbb{R}^N$ y $\epsilon > 0$, consideremos la bola cerrada con centro p y radio ϵ

$$B_\epsilon(x) = \{y \in \mathbb{R}^N : \|x - y\| \leq \epsilon\},$$

donde $\|\cdot\|$ es la norma euclidiana.

Complejo de Vietoris-Rips

Sean $\varepsilon > 0$ y $X = \{p_0, \ldots, p_k\}$, con $p_i \in \mathbb{R}^N$. El complejo simplicial de Vietoris-Rips $VR_\varepsilon(X)$ se define como:

(i) Para cada $p_i \in X$, se selecciona un punto $v_i \in \mathbb{R}^k$ de manera que $v_0, \ldots, v_k$ estén en posición general en $\mathbb{R}^k$. Estos nuevos puntos v_i serán los únicos vértices de $VR_\epsilon(X)$.

(ii) Si para los puntos $p_{i_0}, \ldots, p_{i_m} \in X$ y cualesquiera dos de ellos p_{i_s}, p_{i_t}, se satisface que $B_\epsilon(p_{i_s}) \cap B_\epsilon(p_{i_t}) \neq \emptyset$, entonces agregamos un m-simplejo a $VR_\epsilon(X)$ que tenga como vértices a los puntos $v_{i_0}, \ldots, v_{i_m}$ que cumplen la primera condición.

Por construcción, si $0 < \epsilon_1 < \epsilon_2 < \ldots < \epsilon_k$, tenemos una *filtración* en los complejos de Vietoris-Rips, es decir,

$$VR_{\epsilon_1}(X) \subset VR_{\epsilon_2}(X) \subset \ldots \subset VR_{\epsilon_k}(X).$$

A su vez, estas inclusiones inducen inclusiones en los grupos de homología

$$i_j : H_p(VR_{\epsilon_j}(X)) \to H_p(VR_{\epsilon_{j+1}}(X)).$$

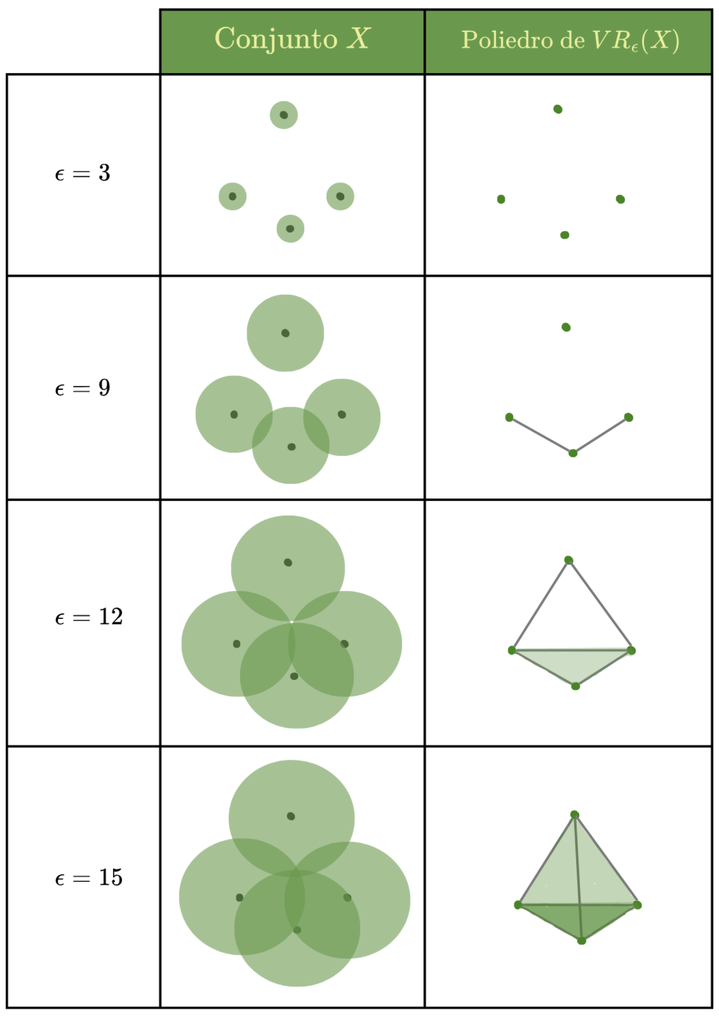


Figura 2: *La tabla muestra una construcción del complejo $VR_\epsilon(X)$ conforme el parámetro ϵ crece.*

Decimos que un elemento en $VR_{\epsilon_{j+1}}(X)$ "nace" cuando dicho elemento está en $H_p(VR_{\epsilon_{j+1}}(X))$, pero no está en $\text{Im}(i_j)$. Mientras que un elemento "muere" en $VR_{\epsilon_{j+1}}(X)$ si está $H_p(VR_{\epsilon_j}(X))$ y es igual a un elemento de $H_p(VR_{\epsilon_{j+1}}(X))$ nacido antes que él, [4].

Finalmente, volvemos a tener una secuencia

$$\ldots \xrightarrow{\Phi_{j-1}} H_p(VR_{\epsilon_j}(X)) \xrightarrow{\Phi_j} H_p(VR_{\epsilon_{j+1}}(X)) \xrightarrow{\Phi_{j+1}} \ldots$$

por lo que la homología persistente realiza un seguimiento del nacimiento y muerte de los generadores de los grupos de homología y la manera en que se visualiza este seguimiento será por medio de diagramas de persistencia. Así que, un diagrama de persistencia es un multiconjunto de puntos de la forma (b, d) en $[0, \infty) \times [0, \infty)$, donde b es el momento en que nace el generador y d es el momento en que muere.

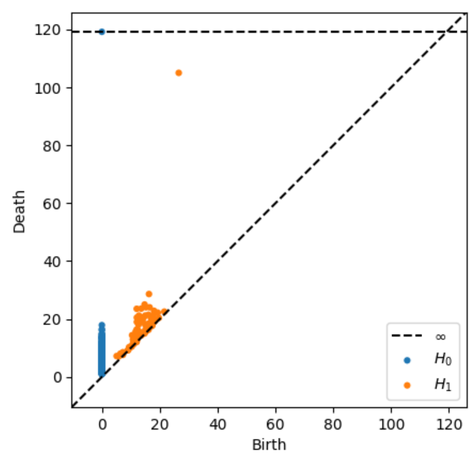


Figura 3: *Un ejemplo de un diagrama de persistencia.*

4. IMÁGENES DE PERSISTENCIA

En la topología computacional, se ha popularizado el uso de *machine learning* (ML) para mejorar la comprensión de los espacios. Estás técnicas se pueden aplicar a un DP al asociársele una métrica que nos garantice su estabilidad, la métrica de *Wasserstein* es una de ellas. No obstante, las técnicas de ML que se le pueden aplicar a un DP son limitadas, ya que consideran a los DP como nubes de puntos, mientras que las demás técnicas requieren más que una métrica.

Como alternativa a este problema limitante, tenemos a las IP que permiten vectorizar al DP para implementar otros métodos de ML, como por ejemplo, el Support Vector Machine. Se ha visto recientemente que las IP llegan a ser iguales o mejores que los DP al aplicar técnicas de ML, más aún, que las IP son más sencillas de usar, [1,2]

Superficie de persistencia

Sea B un DP, la *superficie de persistencia* (SP) es la función $\rho_B : \mathbb{R}^2 \to \mathbb{R}$ definida por

$$\rho_B(z) = \sum_{u \in T(B)} f(u)\phi_u(z).$$

En la definición anterior, $T(B)$ es la transformación $T(x, y) = (x, y - x)$ de puntos en B, $\phi_u : \mathbb{R}^2 \to \mathbb{R}$ es una distribución de probabilidad diferenciable con media $u \in \mathbb{R}^2$ y $f : \mathbb{R}^2 \to \mathbb{R}$ una función peso fija con la característica de que es continua, diferenciable por partes y es cero a lo largo del eje horizontal.

Imagen de persistencia

Para B un DP, su *imagen de persistencia* es la colección de pixeles

$$I(\rho_B)_q = \iint_q \rho_B dy dx.$$

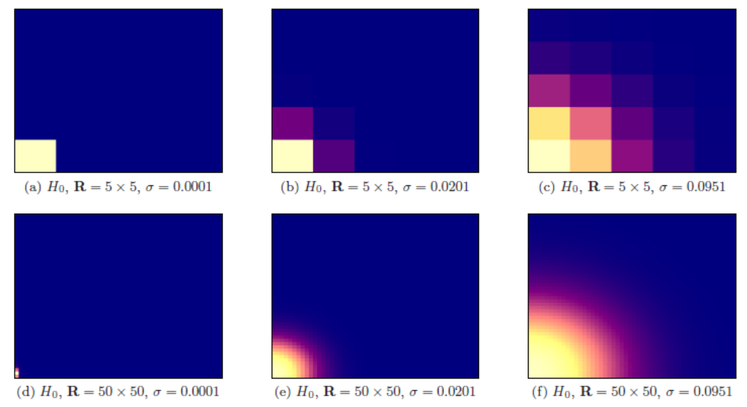


Figura 4: *Ejemplo de una IP sobre el 0-ésimo grupo de homología, donde la malla y varianza de la distribución de probabilidad varían. Créditos a [1].*

Como se mencionó al inicio de la sección, a los DP se les puede asociar la siguiente métrica la cual nos ayuda a tener estabilidad en las SP cuando consideramos a $p = 1$.

Métrica de Wasserstein

Sean B, B' diagramas de persistencia, se define la distancia p de *Wasserstein* entre B y B' como

$$W_p(B, B') = \inf_{\gamma : B \to B'} \left(\sum_{u \in B} \|u - \gamma(u)\|_\infty^p \right)^{1/p},$$

donde $1 \leq p < \infty$ y γ son todas las posibles biyecciones que hay entre B a B'.

Recordemos que la norma infinito es $\|x\|_\infty = \text{máx}_{1 \leq i \leq n}\{|x_i|\}$.

Las SP son estables precisamente porque las distancia de dos SP asociadas a dos DP está acotada por arriba. Con eso, damos el teorema de estabilidad para las IP cuya demostración está en [1], y lo que nos dice es que pequeñas variaciones en los DP producen IP muy similares.

Teorema 4.1 *La IP $I(\rho_B)$ es estable con respecto a la distancia 1-Wasserstein entre DP. Aún más, si A es el área máxima de cualquier pixel en la imagen, A' el área total de la imagen y n el número de pixeles en la imagen, entonces*

$$\|I(\rho_B) - I(\rho_{B'})\|_\infty \leq \sqrt{10}A(\|f\|_\infty |\nabla\phi| + \|\phi\|_\infty |\nabla f|)W_1(B, B')$$
$$\|I(\rho_B) - I(\rho_{B'})\|_1 \leq \sqrt{10}A'(\|f\|_\infty |\nabla\phi| + \|\phi\|_\infty |\nabla f|)W_1(B, B')$$
$$\|I(\rho_B) - I(\rho_{B'})\|_2 \leq \sqrt{10n}A(\|f\|_\infty |\nabla\phi| + \|\phi\|_\infty |\nabla f|)W_1(B, B').$$

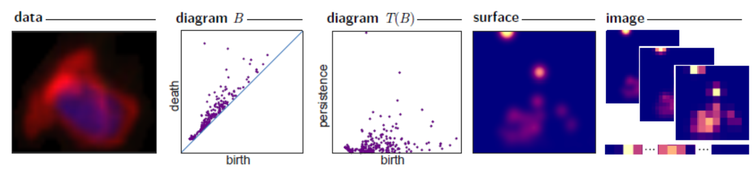


Figura 5: *Ejemplo del procedimiento para obtener una IP de una base de datos, pasando por la transformación $T(B)$. Créditos a [2].*

Referencias

[1] Hepushtanova, S., Emerson, T., Hanson, E., Kirby, M., Motta, F., Neville, R., Peterson, C., Shipman, P., Ziegelmeier, L. (2016). Persistence images: An alternative persistent homology representation. Journal of Machine Learning Research, 18(8), 1–35.

[2] Henry Adams, Timothy Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francesco Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. Journal of Machine Learning Research, 18(8):1–35, 2017.

[3] Herbert Edelsbrunner and John Harer. Persistent homology-a survey. Contemporary mathematics, 453(26):257–282, 2008.

[4] Herbert Edelsbrunner and John Harer. Computational topology: an introduction. American Mathematical Soc., 2010.