Medidas de capacidad y análisis topológico de datos*

Alba Lirón-León[1 [0009−0005−7969−6809]**], Inmaculada Gutiérrez[2,4 [0000−0002−7103−393X]] J.Tinguaro Rodríguez[1,3 [0000−0003−4989−2348]], and Daniel Gómez[2,4 [0000−0001−9548−5781]]

[1] Facultad de CC. Matemáticas, UCM
[2] Facultad de Estudios Estadísticos, UCM
[3] Instituto de Matemática Interdisciplinar, UCM
[4] Instituto Universitario de Estadística y Ciencia de Datos, UCM

Abstract. Se presenta una investigación en curso que relaciona el análisis topológico de datos y la teoría de medidas de capacidad. Un resultado preliminar permite obtener filtraciones de complejos simpliciales a partir de medidas de capacidad. Se ilustra este resultado mediante ejemplos.

Keywords: Medidas de Capacidad · Complejos Simpliciales · Análisis Topológico de Datos

1 Introducción

El análisis topológico de datos (ATD, [1]) es un campo de las matemáticas cuyo objetivo es determinar, mediante técnicas topológicas, propiedades cualitativas de conjuntos de datos. Recientemente ha cobrado una gran relevancia en el estudio de problemas en que los datos presentan estructuras subyacentes complejas cuya comprensión, representación y tratamiento es crucial para la obtención de soluciones adecuadas. Por otro lado, las medidas de capacidad (MC, [5]) son una herramienta que permite representar información y relaciones de diverso orden (como importancia, coste, interacción, dependencia o pertenencia colectiva) sobre los elementos de un conjunto.

Este trabajo presenta e ilustra un resultado preliminar de una investigación en curso orientada a establecer vínculos entre las MC y el ATD. Este resultado establece un mecanismo para construir una filtración de complejos simpliciales a partir de una medida de capacidad, y hace posible representar una MC mediante un código de barras de homología persistente, una herramienta de ATD que proporciona una descripción multiescala de la estructura topológica (componentes conexas, ciclos, cavidades, etc.) de un objeto o conjunto de datos a partir de

* Este resultado es parte del proyecto de I+D+i PID2024-155289NB-I00 financiado por MICIU/AEI/10.13039/501100011033 y FEDER/UE. La primera autora ha sido contratada mediante la ayuda PREP2024-003189 financiada por MICIU/AEI/10.13039/501100011033 y "FSE+".

** Autor de Correspondencia: aliron@ucm.es

una filtración del mismo (ver e.g. [2]). Así pues, este enfoque permite estudiar la estructura topológica inducida por una MC.

2 Preliminares

Definición 1. Sea X un conjunto finito. Un **complejo simplicial abstracto** [6] L sobre X es una familia de subconjuntos no vacíos de X tales que, si $\sigma \in L$ y $\tau \subset \sigma$ es no vacío, entonces $\tau \in L$. En particular, los **vértices** de L son los elementos del conjunto X y un **símplice abstracto** σ es cualquier elemento de L. Su dimensión se define como $\dim(\sigma) = \#\sigma - 1$.

Definición 2. Una **filtración** [3] en un complejo simplicial K consiste en una familia de subcomplejos $\mathcal{F} = \{K_i\}_{i \in I}$ indexada sobre un conjunto totalmente ordenado I tal que si $i \leq j$, entonces $K_i \subset K_j$. Además, $\emptyset, K \in \mathcal{F}$.

Dada una filtración de complejos simpliciales, la **homología simplicial** permite estudiar cómo evolucionan sus características topológicas al variar el parámetro de filtración. En particular, los grupos de homología $H_k(K_t; \mathbb{F})$ describen componentes conexas, ciclos, cavidades y estructuras de dimensión superior presentes en cada complejo K_t. La **homología persistente** [2] compara estos grupos a lo largo de la filtración y resume la información mediante un código de barras, cuyos intervalos representan el nacimiento y la muerte de clases homológicas al aumentar el parámetro t.

Definición 3. Sea X un conjunto finito. Una **medida de capacidad** [5] sobre $(X, \mathcal{P}(X))$ es una aplicación $\mu : \mathcal{P}(X) \longrightarrow [0, 1]$ que cumple (1) Condiciones de borde: $\mu(\emptyset) = 0$ y $\mu(X) = 1$; (2) Monotonía: si $A \subseteq B$, entonces $\mu(A) \leq \mu(B)$.

Las medidas de capacidad pueden representarse mediante la transformada de Möbius [4], obteniéndose una descomposición en coeficientes.

3 De medidas de capacidad a complejos simpliciales

El siguiente resultado, cuya demostración es directa, permite obtener una filtración de complejos simpliciales a partir de una medida de capacidad.

Teorema 1. *Sea el conjunto $X = \{x_1, \ldots, x_n\}$ y sea μ una capacidad. Para cada $t \in \mathbb{R}$, definimos $K_t^\mu = \{\emptyset\} \cup \{\sigma \subseteq X : \sigma \neq \emptyset \text{ y } \mu(\sigma) \leq t\}$. Entonces la familia $\mathcal{K}^\mu = \{K_t^\mu\}_{t \in \mathbb{R}}$ es una filtración creciente de complejos simpliciales abstractos. Es decir: (1) Para todo $t \in \mathbb{R}$, K_t^μ es un complejo simplicial abstracto. (2) Si $s \leq t$, entonces $K_s^\mu \subseteq K_t^\mu$.*

Por tanto, a partir del Teorema 1, toda medida de capacidad μ definida sobre un conjunto finito X induce una filtración de complejos simpliciales y, en consecuencia, un código de barras asociado a μ. Dicho código de barras describe la evolución de la estructura topológica de los complejos K_t^μ a medida que aumenta el parámetro de escala t.

Observación 1 (Tiempo de nacimiento de un símplice). El valor $\mu(\sigma)$ puede interpretarse como el valor de filtración, o tiempo de nacimiento, del símplice σ, i.e., $\text{birth}(\sigma) = \mu(\sigma)$. La monotonía de μ implica que si $\tau \subseteq \sigma$, entonces $\mu(\tau) \leq \mu(\sigma)$. Por tanto, toda cara τ de un símplice σ aparece no más tarde que σ. Esta es precisamente la condición de compatibilidad necesaria para obtener un complejo simplicial filtrado válido.

Ejemplo 1 (Capacidad aditiva uniforme). Sea $X = \{a, b, c\}$ y definamos $\mu_{\text{unif}}(S) = \frac{|S|}{3}$. Nótese que μ_{unif} es una capacidad aditiva, sin interacciones por pares ni de orden superior.

La filtración inducida por μ_{unif} es $K_t^{\mu_{\text{unif}}} = \left\{ S \subseteq X : \frac{|S|}{3} \leq t \right\}$. Los valores críticos son 1/3, 2/3 y 1. Concretamente, $K_t^{\mu_{\text{unif}}} = \{\emptyset\}$ si $0 \leq t < \frac{1}{3}$; $K_t^{\mu_{\text{unif}}} = \{\emptyset, \{a\}, \{b\}, \{c\}\}$ si $\frac{1}{3} \leq t < \frac{2}{3}$; $K_t^{\mu_{\text{unif}}} = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}\}$ si $\frac{2}{3} \leq t < 1$; $K_1^{\mu_{\text{unif}}} = \mathcal{P}(X)$.

Para $\frac{2}{3} \leq t < 1$, el complejo es la frontera de un triángulo. En ese intervalo aparece una clase de homología en dimensión 1, que muere cuando aparece el 2-simplex $\{a, b, c\}$ en $t = 1$. En homología ordinaria, los códigos de barra son $\mathcal{B}_0 = \left\{ \left[\frac{1}{3}, \frac{2}{3}\right)^{\times 2}, \left[\frac{1}{3}, \infty\right) \right\}$, $\mathcal{B}_1 = \left\{ \left[\frac{2}{3}, 1\right) \right\}$, donde el exponente es la multiplicidad del intervalo (ver Figura 1). Se ilustra así que, en el caso uniforme, la filtración depende únicamente del cardinal de los subconjuntos.

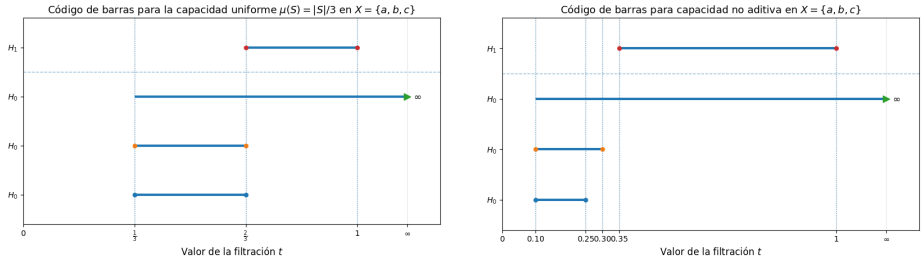
Ejemplo 2 (Capacidad no aditiva). Sea $X = \{a, b, c\}$. Se define la capacidad μ mediante sus coeficientes de Möbius: $m_\mu(\{a\}) = m_\mu(\{b\}) = m_\mu(\{c\}) = 0.10$; $m_\mu(\{a, b\}) = 0.05$; $m_\mu(\{a, c\}) = 0.15$, $m_\mu(\{b, c\}) = 0.10$; $m_\mu(\{a, b, c\}) = 0.40$. Además, tomamos $m_\mu(\emptyset) = 0$. La suma total de todos los coeficientes es 1. Como todos los coeficientes son no negativos, la función $\mu(S) = \sum_{A \subseteq S} m_\mu(A)$ es monótona. Por tanto, μ es una capacidad.

Los valores de nacimiento de los símplices son: $\mu(\{a\}) = \mu(\{b\}) = \mu(\{c\}) = 0.10$; $\mu(\{a, b\}) = 0.25$, $\mu(\{b, c\}) = 0.30$, $\mu(\{a, c\}) = 0.35$, $\mu(\{a, b, c\}) = 1$.

En este ejemplo, los tres vértices nacen simultáneamente en $t = 0.10$. La arista $\{a, b\}$ aparece en $t = 0.25$ y une dos componentes conexas, por lo que una clase de H_0 muere en ese instante. Después, la arista $\{b, c\}$ aparece en $t = 0.30$ y conecta la tercera componente con las anteriores, de modo que otra clase de H_0 muere en $t = 0.30$. La componente conexa restante persiste. Por tanto, $\mathcal{B}_0 = \{[0.10, 0.25), [0.10, 0.30), [0.10, \infty)\}$.

Finalmente, cuando aparece la arista $\{a, c\}$ en $t = 0.35$, se forma el ciclo $a - b - c - a$, que genera una clase de homología en dimensión 1. Esta clase muere en $t = 1$, cuando aparece el 2-símplice $\{a, b, c\}$ y rellena el ciclo. Por tanto, $\mathcal{B}_1 = \{[0.35, 1)\}$. Así, la no aditividad retrasa la aparición del triángulo completo y produce una clase en dimensión 1 que persiste en el intervalo $[0.35, 1)$.

Este ejemplo muestra el papel de los coeficientes de Möbius. El coeficiente ternario $m_\mu(\{a, b, c\}) = 0.40$ no afecta al nacimiento de las aristas, pero sí retrasa la aparición del triángulo completo. Por tanto, la no aditividad de la capacidad introduce información de interacción de orden superior en la filtración.



(a) Capacidad uniforme. **(b)** Capacidad no aditiva.

Fig. 1: Comparación entre los códigos de barras obtenidos a partir de una capacidad aditiva uniforme y una capacidad no aditiva sobre $X = \{a, b, c\}$.

4 Conclusiones y trabajo en curso

Se ha descrito un procedimiento para asociar una filtración de complejos simpliciales a una medida de capacidad μ. La monotonía de μ resulta esencial en este proceso para garantizar que las caras aparecen antes que sus cofaces. Se hace posible así aplicar herramientas de análisis topológico de datos al estudio de medidas de capacidad: los códigos de barras obtenidos proporcionan una descripción topológica de la estructura de interacción codificada por la medida. La transformada de Möbius añade una interpretación algebraica: permite descomponer el nacimiento de cada simplex en contribuciones individuales, binarias y de orden superior.

Como trabajo en curso, se están se están analizando familias particulares de capacidades, como las 2-aditivas y capacidades definidas a partir de distribuciones de probabilidad. También se está investigando la relación con integrales de Choquet, con el objetivo de incorporar señales o pesos adicionales sobre los vértices. Se está desarrollando una implementación computacional para construir estas filtraciones y comparar sus códigos de barras en ejemplos concretos.

Referencias

1. Edelsbrunner, H., Harer, J.: Computational topology: an introduction. American Mathematical Society (2010)
2. Otter, N., Porter, M.A., Tillmann, U., Grindrod, P., Harrington, H.A.: A roadmap for the computation of persistent homology. EPJ data science **6**(1), 17 (2017)
3. Oudot, S.Y.: Persistence theory: from quiver representations to data analysis, vol. 209. American Mathematical Society Providence (2015)
4. Rota, G.C.: On the foundations of combinatorial theory: I. Theory of Möbius functions. In: Classic Papers in Combinatorics, pp. 332–360. Springer (1964)
5. Sugeno, M.: Theory of Fuzzy Integrals and Its Applications. Ph.D. thesis, Tokyo Institute of Technology (1974)
6. Virk, Ž.: Introduction to Persistent Homology. Založba UL FRI (2022)