TIPE : Persistance homologique

Contents

I Définitions

I.1 Géométrie

On appelle enveloppe convexe d'un ensemble de points X l'ensemble des points qui peuvent être écrits comme une combinaison convexe des points de X. Formellement, l'enveloppe convexe de X est définie comme : Soit A une partie de E. L'enveloppe convexe de A est l'intersection de toutes les parties convexes de E qui contiennent A.

On appelle *simplexe* σ la généralisation d'un triangle en dimension quelconque. Par exemple :

- Un simplexe de dimension 0 est un point, noté $\sigma = x$.
- Un simplexe de dimension 1 est une arête, noté $\sigma = x\ y$.
- Un simplexe de dimension 2 est un triangle, noté $\sigma = x\ y\ z$.

Avec x, y et z des sommets.

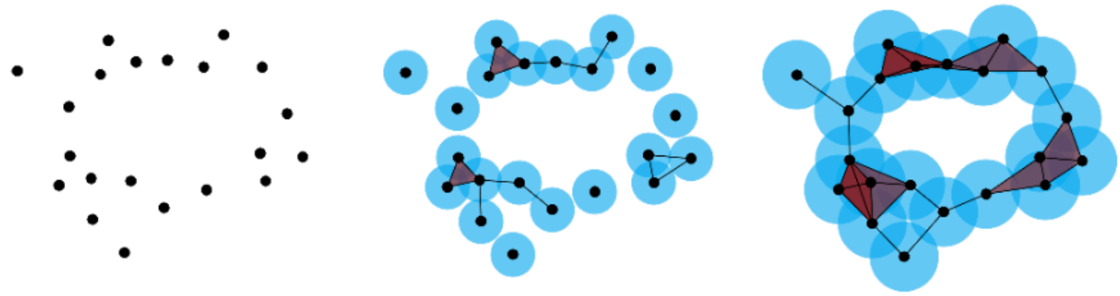

Figure 1: Exemples de simplexes en dimension 0, 1 et 2 (de gauche à droite).

L'orientation d'un simplexe est définie par un ordre sur ses sommets. Par exemple, un triangle $x\ y\ z$ est orienté positivement si x < y < z.

Un *k-simplexe* est alors un simplexe de dimension k. Dès lors, on peut définir un *complexe simplicial* K comme un ensemble de n simplexes d'un espace affine tels que :

- Toutes les faces d'un simplexe de K sont également dans K.
- L'intersection de deux simplexes de K non disjoints doit être une face commune.

On note k la dimension d'un complexe simplicial K si k est la plus grande dimension des simplexes de K.

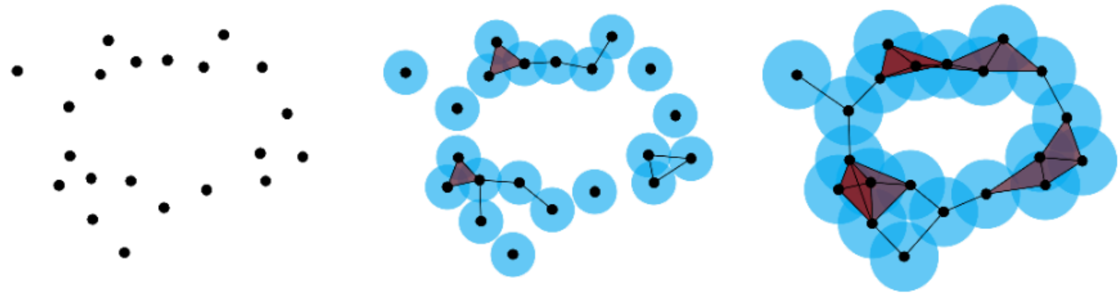

Figure 2: Exemples de complexes simpliciaux de dimension 0 (en rouge), 1 (en bleu, uniquement les simplexes de dimension 1) et 2 (en vert, uniquement les simplexes de dimension 2).

On appelle *filtration* une suite de complexes simpliciaux croissante pour l'inclusion.

$$\emptyset = K_0 \subseteq \dots \subseteq K_p = K$$

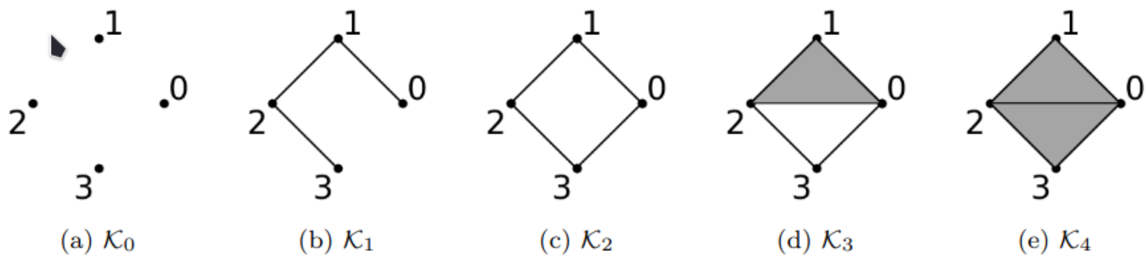

Figure 3: Exemple de filtration, tiré de [1]

I.2 Topologie

L'homologie persistante est une méthode de topologie algébrique qui permet de mesurer la forme des données. Elle est basée sur l'homologie, qui est une notion mathématique permettant de détecter les trous dans un espace topologique.

Cette idée nous permet de définir les groupes d'homologie persistante. Ces groupes sont des invariants topologiques qui permettent de mesurer la forme des données. Ils sont définis à partir d'une filtration de complexes simpliciaux.

Finalement, nous voulons détecter les trous k dimensionnels, soit les k-trous. Pour cela, on va collecter les k-simplexes autour de ces trous. Par exemple :

- un 0-trou est délimité par un sommet dans une composante connexe, ainsi la présence de 0-trou détecte les déconnexions dans K
- un 1-trou est délimité par des arêtes autour de lui, détectant ainsi les boucles dans K.
  Exemple : un triangle sans face.
- un 2-trou est délimité par des triangles autour de lui, détectant ainsi les zones de vide dans K.
  Exemple : l'intérieur d'une pyramide.

Dans la suite on va définir formellement les outils mathématiques pour détecter ces trous.

I.2.1 Homologie

Un espace topologique K est un couple (E, T) où E est un ensemble et T une topologie sur E, c'est à dire un ensemble de parties de E que l'on appelle ouverts de K qui vérifie les propriétés suivantes :

1. E et $\emptyset$ appartiennent à T.
2. Toute union quelconque d'ouverts est un ouvert, c'est à dire si $(O_i)_{i\in I}$ est une famille d'éléments de T, indexée par un ensemble quelconque I alors
$$\cup_{i\in I} O_i \in T$$
3. Toute intersection finie d'ouverts est un ouvert, c'est à dire si $O_1, ..., O_n$ sont des éléments de T alors
$$O_1 \cap ... \cap O_n \in T$$

Un *homomorphisme* est simplement un morphisme de groupes.

Un *homéomorphisme* est une application bijective continue d'un espace topologique et dont l'inverse est continue. Dans ce cas, les deux espaces sont dits *homéomorphes*.

Un *invariant topologique* est une propriété d'un espace topologique qui reste inchangée par homéomorphisme.

Soit K un espace topologique. *L'homologie* de K est un ensemble d'invariants topologiques de K représentés par ses *groupes d'homologie*

$$H_0(K), H_1(K), H_2(K), ...$$

Où le k-ième groupe d'homologie $H_k(K)$ décrit les trous de dimension k dans K.

Une *chaîne complexe* de $C(K)$, ou *complex de chaîne* est une séquence de groupes abéliens ou de modules $C_0, C_1, C_2, ...$ relié par des homomorphismes $\partial_n : C_n \to C_{n-1}$ nommés *opérateurs limite*. On a

$$... \overbrace{\to}^{\partial_{n+1}} C_n \overbrace{\to}^{\partial_n} C_{n-1} \overbrace{\to}^{\partial_{n-1}} ... \to C_2 \overbrace{\to}^{\partial_2} C_1 \overbrace{\to}^{\partial_1} C_0 \overbrace{\to}^{\partial_0} \{0\}$$

Où $\{0\}$ est le groupe trivial et $C_i \equiv \{0\}$ pour $i < 0$. De plus, il faut que $\partial_n o \partial_{n+1} = 0$ l'application nulle pour tout n.

On pose $B_n(X) = \mathrm{Im}(\partial_{n+1})$ l'ensemble des limites et $Z_n(X) = \mathrm{Ker}(\partial_n)$ l'ensemble des cycles.

(Paragraphe sur les groupes normaux)

On peut alors définir le k-ième groupe d'homologie de K comme le quotient des cycles par les limites :

$$H_k(K) = \frac{Z_k(K)}{B_k(K)}$$

Les éléments de $H_k(K)$ sont appelés les *classes d'homologie* de K. Chaque classe d'homologie est une classe d'équivalence sur plusieurs cycles et deux cycles de la même classe d'homologie sont dits homologues.

I.2.2 Homologie simplicial

On considère une k-chaîne, soit est une combinaison linéaire de k-simplexes sur K. On note C_k le groupe abélien libre dont les générateurs sont les simplexes de X orientés en k

dimension. La relation entre les faces induit une notion de bordure pour la k-chaîne : l'application *des limites* $\partial_k : C_k(K) \to C_{k-1}(K)$ est définit sur chaque simplexe $\sigma = [v_0, ..., v_k]$ par

$$\partial_k(\sigma) = \sum_{i=0}^{k} (-1)^i [v_0, ..., v_{i-1}, v_{i+1}, ..., v_k]$$

Considérée comme nulle si $k = 0$. Ce comportement sur les générateurs induit un homomorphisme sur tout C_k, en effet soit $c \in C_k$, l'écrire comme la somme des générateurs

$$c = \sum_{\sigma_i \in K_k} m_i \sigma_i$$

Où K_k est l'ensemble des k-complexes de K et les m_i sont des coefficients de l'anneau C_k (Généralement des entiers). Puis définir :

$$\partial_{k(c)} = \sum_{\sigma_i \in K_k} m_i \partial_k(\sigma_i)$$

La dimension de la k-ième homologie de K s'avère être le nombre de k-trous dans K.

On a $H_k(K) \approxeq \underbrace{\mathbb{Z}_2 \oplus ... \oplus \mathbb{Z}_2}_{\beta_k \text{ fois}}$

Le k-nombre de Betti de K β_k est le rang du k-ième groupe d'homologie de K. Intuitivement il s'agit du nombre de k-trous dans K.

II Modélisation

Soit un espace métrique (M, d) et X un ensemble de points de M.

Similairement à [1], on va utiliser un ensemble de points X pour représenter une ville, chaque point x_i a un poids w_i qui sera le temps d'attente moyen à la station, que l'on etablira via les densités de population environnante aux stations.

On définit pour chaque point x_i, et pour le paramètre de filtration t la fonction *rayon* $r_i(t)$ comme

$$r_i(t) = \begin{cases} -\infty & \text{si } t < w_i \\ t - w_i & \text{sinon} \end{cases}$$

On définit le VR complexe pondéré $V_t = (X, M, d, w)$ à la filtration t comme le complexe simplicial qui contient tout point x_i tel que $w_i < t$ et tout simplexe $\sigma = \left(x_{i_0}, ..., x_{i_k}\right)$ qui satisfait $d\left(x_{i_j}, x_{i_l}\right) + w_{i_j} + w_{i_l} < 2t$

Intuitivement, $\cup_i B(x_i, r_i(t))$ est l'ensemble des points tels que si un individu est en y alors son temps estimé pour prendre le transport (incluant le temps de marche jusque la station et le temps d'attente en station)

On définit la fonction distance comme

$$d(x, y) = \min(t_{\text{marche}}(x, y), t_{\text{voiture}}(x, y))$$

Bibliography

[1] ABIGAIL HICKOK, BENJAMIN JARMAN, MICHAEL JOHNSON, JIAJIE LUO, and MASON A. PORTER, *Persitent Homology for Resource Coverage: A Case Study of Access to Polling Sites*