L'Homologie persistante

Appliquée à la reconnaissance de formes

Mathieu Vallée

Contexte:

On a des données sous la forme de nuage de points, on veut trouver sur quelles courbes, surfaces, formes elles se répartissent.

Contexte:

On a des données sous la forme de nuage de points, on veut trouver sur quelles courbes, surfaces, formes elles se répartissent.



Figure 1: Graphique en 3D représentant le salaire actuel en fonction du premier salaire et du niveau d'éducation de différents individus.

Contexte:

On a des données sous la forme de nuage de points, on veut trouver sur quelles courbes, surfaces, formes elles se répartissent.



Figure 1: Graphique en 3D représentant le salaire actuel en fonction du premier salaire et du niveau d'éducation de différents individus.



Figure 2: Un tore en 3D

Problème

Trouver une méthode permettant de savoir sur quelles formes reposent ces nuages de points

Problème

Trouver une méthode permettant de savoir sur quelles formes reposent ces nuages de points

Objectif

Programmer un algorithme permettant de déterminer la forme d'un nuage de points

La topologie algébrique

Outil mathématique permettant de trier les formes selon des classes d'équivalence.

La topologie algébrique

Outil mathématique permettant de trier les formes selon des classes d'équivalence. Deux formes appartiennent à la même classe d'équivalence si l'on peut passer continuement de l'une à l'autre (ssi elles sont *homéomorphes*).

La topologie algébrique

Outil mathématique permettant de trier les formes selon des classes d'équivalence. Deux formes appartiennent à la même classe d'équivalence si l'on peut passer continuement de l'une à l'autre (ssi elles sont *homéomorphes*).



Figure 3: On peut passer continuement de la surface d'un cube à celle d'une sphère

La topologie algébrique

Outil mathématique permettant de trier les formes selon des classes d'équivalence. Deux formes appartiennent à la même classe d'équivalence si l'on peut passer continuement de l'une à l'autre (ssi elles sont *homéomorphes*).





Figure 3: On peut passer continuement de la surface d'un cube à celle d'une sphère Figure 4: On ne peut pas passer continuement de la surface d'un cube à la surface d'un tore

Objectifs

On veut créer des invariants tels que 2 objets homéomorphes aient les mêmes invariants.

Objectifs

On veut créer des invariants tels que 2 objets homéomorphes aient les mêmes invariants.

Problème:

Jusqu'à maintenant on a seulement raisonné sur des formes continues: on ne peut pas implémenter d'algorithme gérant ce genre d'objet sans passer par une discrétisation.

Objectifs

On veut créer des invariants tels que 2 objets homéomorphes aient les mêmes invariants.

Problème:

Jusqu'à maintenant on a seulement raisonné sur des formes continues: on ne peut pas implémenter d'algorithme gérant ce genre d'objet sans passer par une discrétisation.

Idée

Utiliser ce qu'on appelle un complexe simplicial pour représenter un objet continu.

Objectifs

On veut créer des invariants tels que 2 objets homéomorphes aient les mêmes invariants.

Problème:

Jusqu'à maintenant on a seulement raisonné sur des formes continues: on ne peut pas implémenter d'algorithme gérant ce genre d'objet sans passer par une discrétisation.

Idée

Utiliser ce qu'on appelle un complexe simplicial pour représenter un objet continu.

Proposition

Une forme est homéomorphe à un complexe simplicial.

Objectifs

On veut créer des invariants tels que 2 objets homéomorphes aient les mêmes invariants.

Problème:

Jusqu'à maintenant on a seulement raisonné sur des formes continues: on ne peut pas implémenter d'algorithme gérant ce genre d'objet sans passer par une discrétisation.

Idée

Utiliser ce qu'on appelle un complexe simplicial pour représenter un objet continu.

Proposition

Une forme est homéomorphe à un complexe simplicial.

Ainsi si on définit des invariants sur les complexes simpliciaux, on les aura définit exactement pour les formes.

Définition (Simplexes)

Soit $n \in \mathbb{N}$, soit $i \leq n$, un i-simplexe est l'enveloppe convexe de i+1 points de \mathbb{R}^n . Si α est un i-simplexe alors dim $\alpha = i$

Définition (Simplexes)

Soit $n \in \mathbb{N}$, soit $i \leq n$, un i-simplexe est l'enveloppe convexe de i+1 points de \mathbb{R}^n . Si α est un i-simplexe alors dim $\alpha = i$



Figure 5: De gauche à droite un *0-simplexe* (un point), un *1-simplexe* (une arête), un *2-simplexe* (un triangle) et enfin un *3-simplexe* (un tétrahèdre).

Définition (Simplexes)

Soit $n \in \mathbb{N}$, soit $i \leq n$, un i-simplexe est l'enveloppe convexe de i+1 points de \mathbb{R}^n . Si α est un i-simplexe alors dim $\alpha = i$



Figure 5: De gauche à droite un *0-simplexe* (un point), un *1-simplexe* (une arête), un *2-simplexe* (un triangle) et enfin un *3-simplexe* (un tétrahèdre).



Définition (Complexe simplicial)

Un complexe simplicial \mathcal{K} est un ensemble de simplexes vérifiant la propriété suivante:

 $\forall \alpha \in \mathcal{K}, \forall \beta \subset \alpha, \beta \in \mathcal{K}$

Définition (Complexe simplicial)

Un complexe simplicial \mathcal{K} est un ensemble de simplexes vérifiant la propriété suivante:

$\forall \alpha \in \mathcal{K}, \forall \beta \subset \alpha, \beta \in \mathcal{K}$



Figure 6: Un complexe simplicial de dimension 2.

Soit \mathcal{K} un complexe simplicial, on construit des espaces vectoriels $\mathcal{H}_0, \mathcal{H}_1, \ldots$ appelés groupes d'homologie.

Soit $\mathcal K$ un complexe simplicial, on construit des espaces vectoriels $\mathcal H_0, \mathcal H_1, \ldots$ appelés groupes d'homologie.

Définition (Nombre de Betti)

Soit \mathcal{K} un complexe simplicial, on appelle n-ème nombre de Betti et on note $\beta_n = \dim \mathcal{H}_n$. Les nombres de Betti représentent le nombre de n-classe d'homologie:

Soit $\mathcal K$ un complexe simplicial, on construit des espaces vectoriels $\mathcal H_0, \mathcal H_1, \ldots$ appelés groupes d'homologie.

Définition (Nombre de Betti)

Soit \mathcal{K} un complexe simplicial, on appelle n-ème nombre de Betti et on note $\beta_n = \dim \mathcal{H}_n$. Les nombres de Betti représentent le nombre de n-classe d'homologie:

• β_0 : nombre de composantes connexes

Soit $\mathcal K$ un complexe simplicial, on construit des espaces vectoriels $\mathcal H_0, \mathcal H_1, \ldots$ appelés groupes d'homologie.

Définition (Nombre de Betti)

Soit \mathcal{K} un complexe simplicial, on appelle n-ème nombre de Betti et on note $\beta_n = \dim \mathcal{H}_n$. Les nombres de Betti représentent le nombre de n-classe d'homologie:

- β_0 : nombre de composantes connexes
- β_1 : nombre de courbes fermées non bouchées par une surface

Soit $\mathcal K$ un complexe simplicial, on construit des espaces vectoriels $\mathcal H_0, \mathcal H_1, \ldots$ appelés groupes d'homologie.

Définition (Nombre de Betti)

Soit \mathcal{K} un complexe simplicial, on appelle n-ème nombre de Betti et on note $\beta_n = \dim \mathcal{H}_n$. Les nombres de Betti représentent le nombre de n-classe d'homologie:

- β_0 : nombre de composantes connexes
- β_1 : nombre de courbes fermées non bouchées par une surface
- β_n , (n > 1) même chose mais en dimension supérieure...



Figure 7: Pour ce complexe simplicial de dimension 2: $\beta_0 = 2$ et $\beta_1 = 1$





Figure 7: Pour ce complexe simplicial de dimension 2: $\beta_0 = 2$ et $\beta_1 = 1$

Figure 8: Pour le tore: $\beta_0 = 1$, $\beta_1 = 2$ et $\beta_2 = 1$

Problème:

Pour pouvoir utiliser l'homologie simpliciale sur notre nuage de points, on doit lui attribuer un complexe simplicial.

Problème:

Pour pouvoir utiliser l'homologie simpliciale sur notre nuage de points, on doit lui attribuer un complexe simplicial.

Idée:

Créer à la main les simplexes à partir du nuage de points.

Problème:

Pour pouvoir utiliser l'homologie simpliciale sur notre nuage de points, on doit lui attribuer un complexe simplicial.

Idée:

Créer à la main les simplexes à partir du nuage de points.

Problème:

Comment procéder pour construire ces simplexes?

Complexe de Vietoris-Ripps

Idée:

Créer les *n*-simplexes ($n \in \mathbb{N}$) si tout les points leurs appartenant sont à une distance ϵ les uns des autres inférieure à un certain rayon caractéristique ϵ_0 .

Idée:

Créer les *n*-simplexes ($n \in \mathbb{N}$) si tout les points leurs appartenant sont à une distance ϵ les uns des autres inférieure à un certain rayon caractéristique ϵ_0 .
Idée:

Créer les *n*-simplexes ($n \in \mathbb{N}$) si tout les points leurs appartenant sont à une distance ϵ les uns des autres inférieure à un certain rayon caractéristique ϵ_0 .

Définition (Complexe de Vietoris-Ripps)

Soit $n \in \mathbb{N}$, soit \mathcal{P} un nuage de points de \mathbb{R}^n , soit $\epsilon \in \mathbb{R}_+$. On définit le complexe de Vietoris-Ripps $\mathcal{K}_{\epsilon}^{V-R}$ comme tel: Pour tout $i \in \{2, ..., n+1\}$, pour toute famille $\{p_0, ..., p_i\}$ de points 2 à 2 distincts de \mathcal{P} , si $\forall j, k \in \{0, ..., i\}, j \neq k, d(p_j, p_k) < \epsilon$, alors $\{p_0, ..., p_i\} \in \mathcal{K}_{\epsilon}^{V-R}$

Idée:

Créer les *n*-simplexes ($n \in \mathbb{N}$) si tout les points leurs appartenant sont à une distance ϵ les uns des autres inférieure à un certain rayon caractéristique ϵ_0 .

Définition (Complexe de Vietoris-Ripps)

Soit $n \in \mathbb{N}$, soit \mathcal{P} un nuage de points de \mathbb{R}^n , soit $\epsilon \in \mathbb{R}_+$. On définit le complexe de Vietoris-Ripps $\mathcal{K}_{\epsilon}^{V-R}$ comme tel: Pour tout $i \in \{2, ..., n+1\}$, pour toute famille $\{p_0, ..., p_i\}$ de points 2 à 2 distincts de \mathcal{P} , si $\forall j, k \in \{0, ..., i\}, j \neq k, d(p_j, p_k) < \epsilon$, alors $\{p_0, ..., p_i\} \in \mathcal{K}_{\epsilon}^{V-R}$

En fonction de la densité de points et pour une certaine distance, on pourrait se rapprocher d'un complexe simplicial ressemblant à cette dite forme.

Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre de courbes fermées non bouchées par une surface$

étape
$$\beta_0 \beta_1$$

Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre de courbes fermées non bouchées par une surface$

$$\begin{array}{c|c} \acute{\text{e}tape} & \beta_0 & \beta_1 \\ \hline 1 & 1 & 0 \end{array}$$

а

Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre \ de \ courbes \ fermées \ non \ bouchées \ par une \ surface$



Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre \ de \ courbes \ fermées \ non \ bouchées \ par une \ surface$



Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre \ de \ courbes \ fermées \ non \ bouchées \ par une \ surface$



Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre de courbes fermées non bouchées par une surface$



Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre de courbes fermées non bouchées par une surface$



Rappel

- $\cdot \beta_0 = \dim \mathcal{H}_0 = nombre \ de \ les \ composantes \ connexes$
- $\beta_1 = \dim \mathcal{H}_1 = nombre de courbes fermées non bouchées par une surface$



Problème:

Trouver la bonne distance caractéristique.

Problème:

Trouver la bonne distance caractéristique.



Figure 10: Rayon choisi trop faible.

Problème:

Trouver la bonne distance caractéristique.



Figure 10: Rayon choisi trop faible.



Figure 11: Rayon choisi trop grand.

L'homologie persistante

Principe de l'homologie persistante:

• Construire les $(\mathcal{K}_{\epsilon}^{V-R})_{\epsilon>0}$

Principe de l'homologie persistante:

- Construire les $(\mathcal{K}_{\epsilon}^{\mathsf{V}-\mathsf{R}})_{\epsilon>0}$
- Calculer l'homologie de chaque complexe de cette famille

Principe de l'homologie persistante:

- Construire les $(\mathcal{K}_{\epsilon}^{V-R})_{\epsilon>0}$
- Calculer l'homologie de chaque complexe de cette famille
- On observera alors l'apparition et la disparition de classes d'homologie quand ϵ augmente

L'homologie persistante

L'algorithme:

 \cdot On part du nuage de points avec $\epsilon = 0$

L'algorithme:

- On part du nuage de points avec $\epsilon = 0$
- On augmente ϵ et on ajoute au fur et à mesure dans une liste splx les simplexes créés dans le complexe de Ripps

L'algorithme:

- On part du nuage de points avec $\epsilon = 0$
- On augmente ϵ et on ajoute au fur et à mesure dans une liste splx les simplexes créés dans le complexe de Ripps
- On crée la matrice D_{∂} de l'application bord définie comme telle:

L'homologie persistante

L'algorithme:

- On part du nuage de points avec $\epsilon = 0$
- On augmente ϵ et on ajoute au fur et à mesure dans une liste splx les simplexes créés dans le complexe de Ripps
- On crée la matrice D_{∂} de l'application bord définie comme telle:

$$\mathbf{D}_{\partial}[j,i] := \begin{cases} 1 & \text{si le simplexe } j \text{ est une face du simplexe } i \\ 0 & \text{sinon} \end{cases}$$

Où i représente le i-ème simplexe de la liste splx

L'algorithme:

- On part du nuage de points avec $\epsilon = 0$
- On augmente ϵ et on ajoute au fur et à mesure dans une liste splx les simplexes créés dans le complexe de Ripps
- $\cdot\,$ On crée la matrice ${\it D}_\partial$ de l'application bord définie comme telle:

$$D_{\partial}[j, i] := \begin{cases} 1 & \text{si le simplexe } j \text{ est une face du simplexe } i \\ 0 & \text{sinon} \end{cases}$$

Où *i* représente le *i*-ème simplexe de la liste splx

 On agit ensuite sur la matrice D en suivant un algorithme va qui permettre d'apparier les simplexes qui créent une classe d'homologie avec ceux qui la détruisent -> construire les "paires de persistance" Résultats



Figure 12: Nuage de 200 points représentant un disque.



Figure 12: Nuage de 200 points représentant un disque.



Figure 13: Diagramme de persistance de ce nuage de points.



Figure 14: Nuage de 200 points représentant un annulus.



Figure 14: Nuage de 200 points représentant un annulus.



Figure 15: Diagramme de persistance de ce nuage de points.



Figure 16: Nuage de 200 points représentant deux annulus.



Figure 16: Nuage de 200 points représentant deux annulus.



Figure 17: Diagramme de persistance de ce nuage de points.



Figure 18: Nuage de 50 points représentant un disque.



Figure 18: Nuage de 50 points représentant un disque.



Figure 19: Diagramme de persistance de ce nuage de points.



Figure 20: Nuage de 50 points représentant un annulus.



Figure 20: Nuage de 50 points représentant un annulus.



Figure 21: Diagramme de persistance de ce nuage de points.



Figure 22: Nuage de 50 points représentant deux annulus.



Figure 22: Nuage de 50 points représentant deux annulus.



Figure 23: Diagramme de persistance de ce nuage de points.



Figure 24: Nuage de 500 points représentant la surface d'une sphère.




Figure 24: Nuage de 500 points représentant la surface d'une sphère.

Figure 25: Diagramme de persistance de ce nuage de points.

21



Figure 26: Nuage de 500 points représentant la surface d'un tore.





Figure 26: Nuage de 500 points représentant la surface d'un tore.

Figure 27: Diagramme de persistance de ce nuage de points.

Conclusion

· Différencier deux formes discrétisées en nuage de points en 2D

- · Différencier deux formes discrétisées en nuage de points en 2D
- Rapidité d'exécution

- · Différencier deux formes discrétisées en nuage de points en 2D
- Rapidité d'exécution

Limites:

- · Différencier deux formes discrétisées en nuage de points en 2D
- Rapidité d'exécution

Limites:

• Résultats peu concluants en 3D

- · Différencier deux formes discrétisées en nuage de points en 2D
- Rapidité d'exécution

Limites:

- Résultats peu concluants en 3D
- Liées à la théorie mathématique

- · Différencier deux formes discrétisées en nuage de points en 2D
- Rapidité d'exécution

Limites:

- Résultats peu concluants en 3D
- Liées à la théorie mathématique
- Limite nombre de points, trop long à s'exécuter en non pré-compilé.

Questions?

Définition théorique de l'homologie simpliciale

Définition (Chaine)

Soit \mathcal{K} un complexe simplicial, notons $(\mathcal{K}_n)_{n \in \mathbb{N}}$ la suite des ensembles des n-simplexes de \mathcal{K} . Soit $i \in \mathbb{N}$, On appelle i-chaîne une combinaison linéaire de i-simplexes à coefficients dans un corps \mathbb{K} . Notons $C_i(\mathcal{K})$ le \mathbb{K} -espace vectoriel généré par \mathcal{K}_i . Si on a une chaîne $c \in C_i(\mathcal{K})$ alors il existe $(c_{\alpha})_{\alpha \in \mathcal{K}_i} \in \mathbb{K}^{\mathcal{K}_i}$ tel que $c = \sum_{\alpha \in \mathcal{K}_i} c_{\alpha} \alpha$.

Définition (Face d'un simplexe)

Soit $n \in \mathbb{N}$, soit $\alpha \in \mathcal{K}_n$, $\alpha = \{x_0, \dots, x_n\}$. On pose pour tout $i \in \{0, \dots, n\}$,

$$\partial_n^i \alpha = \alpha \setminus \{X_i\} = \{X_0, \dots, X_{i-1}, X_{i+1}, \dots, X_n\}$$

On l'appelle la face numéro i de α

Définition (Application bord)

On definit l'application bord $\partial_n : C_n(\mathcal{K}) \to C_{n-1}(\mathcal{K})$ comme tel:

$$\forall \alpha \in C_n(\mathcal{K}), \partial \alpha = \sum_{i=0}^n (-1)^i \partial_n^i \alpha$$

Proposition (admise)

 $\partial_{n-1}\circ\partial_n=0$

Autrement dit

 $\operatorname{Im}(\partial_n) \subset \operatorname{Ker}(\partial_{n-1})$

Définition (Groupe d'homologie)

On définit le n-ème groupe d'homologie \mathcal{H}_n de \mathcal{K} comme tel:

 $\mathcal{H}_n(\mathcal{K}) = un \ supplémentaire \ de \ Im(\partial_{n+1}) \ dans \ Ker(\partial_n)$

Complexité spatiale du complexe de Vietoris-Ripps





Figure 28: Nombre de simplexes en fonction du nombre de points en dimension 2.

Figure 29: Nombre de simplexes en fonction du nombre de points en dimension 3.

Méthode:

Dans la pratique on utilisera un autre type de complexe simplicial.

Définition (Witness Complex)

Soit $\mathcal{P} \subset \mathbb{R}^n$ un nuage de k points et $\mathcal{L} \subset \mathcal{P}$ un sous-ensemble de l points de \mathcal{P} . Soit d une distance sur \mathbb{R}^n . En fixant $v \leq l$, la filtration pour le Witness Complex $\mathcal{F} = \{\mathcal{W}_{\epsilon}^v, \epsilon > 0\}$ de \mathcal{P} est telle que: Pour $k \leq n$, l'ensemble de points $\{p_0, \ldots, p_k\} \subset \mathcal{L}$ est un k-simplexe de \mathcal{W}_{ϵ}^v si il existe $q \in \mathcal{P} \setminus \mathcal{L}$ tel que

 $\max_{0\leq i\leq k}d(p_i,q)\leq \epsilon+\delta_v$

où δ_v est la v-ème plus petite distance entre q et un point de ${\cal L}$

Remarque

$$\epsilon' > \epsilon \Rightarrow \mathcal{W}_{\epsilon}^{\mathsf{V}} \subseteq \mathcal{W}_{\epsilon'}^{\mathsf{V}}$$

Définition (Lazy Witness Complex)

C'est une adaptation du Witness Complex qui va utiliser le fait qu'un simplexe est dans le complexe si ses faces sont toutes dans le complexe. Un couple p_0, p_1 de points décrit une arête du complexe si il existe $q \in \mathcal{P} \setminus \mathcal{L}$ tel que

 $\max\{d(p_0,q),d(p_1,q)\} \leq \epsilon + \delta_v$

Les simplexes de dimension supérieure seront alors dans le complexe si leurs faces le sont.

On definit *low(i)* comme étant le numéro de la ligne avec le dernier 1 dans la colonne i de D_{∂} ou 0 si la colonne est nulle.Au cours de l'algorithme, *low(i)* représentera le simplexe auquel se "réfère" actuellement σ_i .

Algorithm 1 Réduction(D_∂)

- 1: $R = D_{\partial}$
- 2: for $j = 0 \rightarrow N$ do
- 3: while it existe $j_0 < j$ tet que $low(j_0) = low(j)$ do

4:
$$\mathbf{R}[:,j] = \mathbf{R}[:,j] + \mathbf{R}[:,j_0] \mod 2$$

- 5: end while
- 6: end for