Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 8, n. 1, 2021.

Trabalho apresentado no XL CNMAC, Evento Virtual - Co-organizado pela Universidade do Mato Grosso do Sul (UFMS).

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Homologia Persistente: Um breve estudo da distância Bottleneck no agrupamento de imagens

Richard G. dos Santos[1]
Universidade Estadual Paulista - UNESP, Rio Claro
Thiago de Melo[2]
Universidade Estadual Paulista - UNESP, Rio Claro

Resumo. Neste trabalho apresentamos homologia persistente e usamos como uma das ferramentas para agrupamento de imagens, alguns diagramas, conhecidos como barcodes.

Palavras-chave. Diagrama de Persistência, Homologia Persistente, TDA.

1 Introdução

Agrupar, armazenar e analisar dados é de fundamental importância em todas as áreas da Ciência. Quase todo tipo de informação pode ser transformada em um conjunto de dados. Nesse contexto a Análise Topológica de Dados é uma área que vem ganhando bastante relevância, buscando adaptar métodos topológicos ao estudo de dados de alta dimensão.

A Topologia de Persistência consiste no estudo de propriedades de espaços topológicos com filtrações. Representamos conjuntos de dados por funções contínuas $\varphi\colon X \to \mathbb{R}^m$, onde X é um espaço topológico. Cada tal função é chamada *função filtração*, a qual induz uma filtração em X dada pelos conjuntos dos subníveis de φ. A Topologia de Persistência permite analisar os dados representados por uma função filtração e examinar como as propriedades de seus conjuntos de subníveis persistem quando se percorre a filtração, descrevendo o surgimento e desaparecimento de 'buracos de dimensão k' de acordo com variações na filtração do espaço X.

Durante o estudo da homologia persistente, um dos recursos que utilizamos é o chamado *diagrama de persistência.* Um diagrama pode ser pensado como uma persistência, análoga ao número de Betti. A variação do parâmetro na filtração fornece um conjunto de pontos, provenientes das bases dos grupos de homologia calculados, que podem se alterar em função deste parâmetro. Um diagrama é uma representação gráfica do grupo de homologia, na qual podemos ver rapidamente a persistência de certas propriedades homológicas. Por fim, metrizamos o espaço dos diagramas de persistência utilizando a métrica Bottleneck visando comparar dados via a proximidade de seus diagramas.

2 Agrupando imagens via Homologia Persistente

Para esse estudo, criamos um conjunto de 80 imagens, em quatro grupos distintos, de dimensão 100×100 pixels, sendo elas: borboletas, humanos, peixes e espirais. Para cada imagem, obtemos um conjunto de pontos de $\mathbb{Z}^2$, referentes às coordenadas dos pixels das imagens. A partir disso,

[1]richard.guilherme@unesp.br.
[2]thiago.melo@unesp.br.

calculamos a filtração e os respectivos diagramas de persistência da nuvem de pontos de cada imagem, através dos complexos de Vietoris-Rips.

Para classificação das imagens, utilizamos os diagramas de persistência referentes à dimensão 1 e verificamos sua proximidade via métrica Bottleneck. Assim, elaboramos a matriz de distâncias do nosso conjunto, podendo observar a relação de diagramas e suas respectivas imagens, via mapa de calor, apontando uma relação direta entre a proximidade dos diagramas com as suas referentes imagens de mesmo grupo. A partir disso, representamos tal relação via escalonamento multidimensional, ou seja, obtendo um conjunto de pontos de $\mathbb{R}^2$ que melhor aproxime as distâncias entre os diagramas.

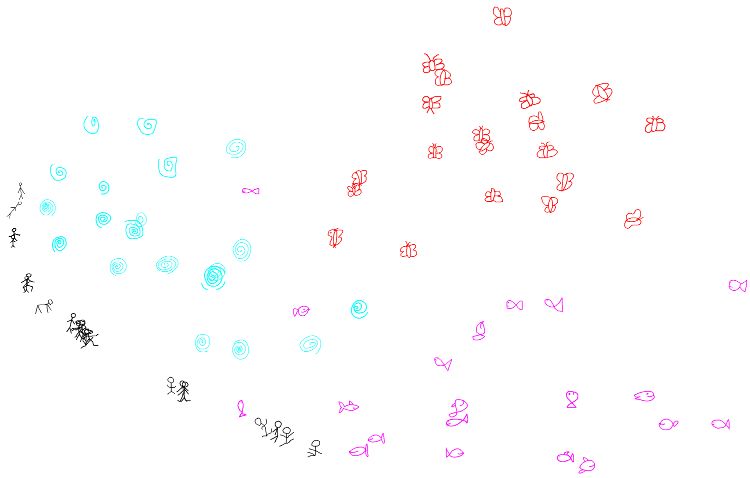

Figura 1: Imagens organizadas através dos pontos obtidos via escalonamento multidimensional.

3 Conclusão

Podemos realizar um agrupamento inicial de imagens através da Homologia Persistente e, baseando nas classes de homologia de dimensão 1, obter um bom agrupamento para o nosso conjunto. Além disso, tais métodos permitem a visualização do agrupamento realizado através de dendrogramas e do método de escalonamento multidimensional, facilitando uma análise inicial da qualidade do agrupamento.

4 Agradecimentos

Os autores agradecem a FAPESP, processo n°2020/08829-8, pelo auxílio financeiro.

Referências

[1] Chazal, F.; Michel, B. *An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists.* arXiv:1710.04019v1, 2017.

[2] Edelsbrunner, H.; Harer, J. *Computational Topology: An Introduction.* Providence: American Mathematical Society, 2010.

[3] Munkres, J.R. *Elements of algebraic topology.* California: Addison-Wesley Publishing Company, 1984.