

Clusterização de texturas de Broadtz no plano HC

Alejandro Frery¹, Osvaldo Rosso¹, Eduarda Chagas²

¹ Laboratório de Computação Científica e Análise Numérica (LaCCAN)
Universidade Federal de Alagoas (UFAL) – Maceio, AL – Brazil

²Departamento de Ciência da Computação
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte, MG – Brazil

eduardachagas48@laccan.ufal.br

1. Introdução

Em análises anteriores foi levantada a hipótese de que ao realizar a caracterização das texturas de Broadtz no plano Complexidade-Entropia alguns clusters eram formados, dividindo os dados em subconjuntos.

Foram analisadas um total de 45 texturas e seus respectivos pontos no plano, para todas as combinações entre o conjunto de dimensões $D = \{2, 3, 4, 5, 6\}$ e delay $\tau = \{1, 2, 3, 4, 5\}$. Como podemos verificar na Figura 1, os planos $(D = 5, \tau = 1)$, $(D = 5, \tau = 2)$, $(D = 5, \tau = 3)$ e $(D = 5, \tau = 4)$ separaram visualmente os dados em três subconjuntos. Entretanto, nenhuma análise experimental havia sido feito para comprovar tais suposições.

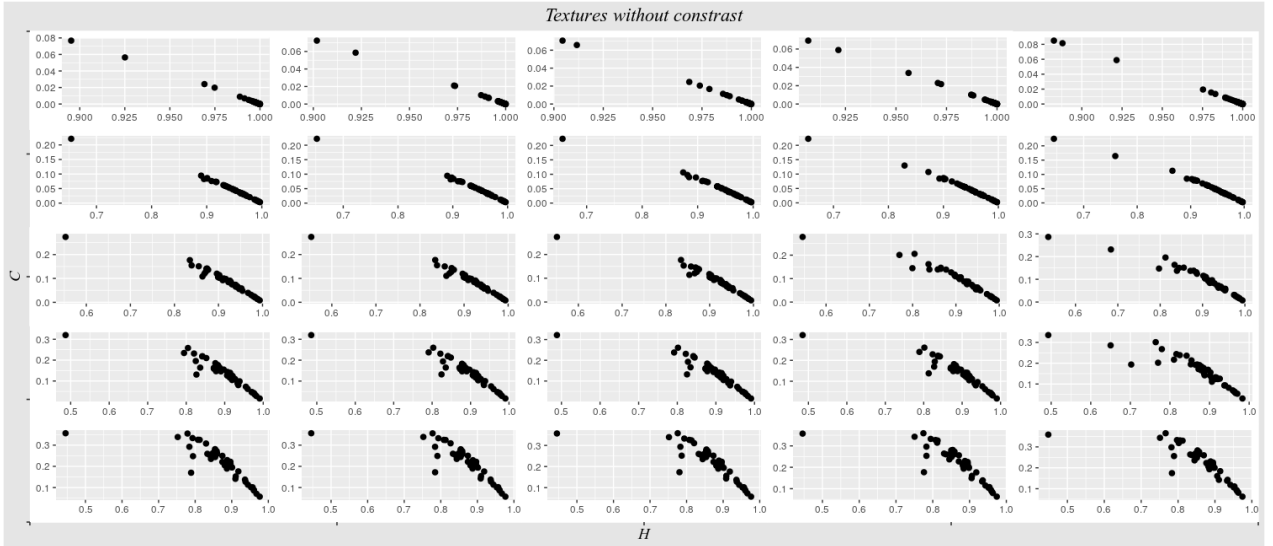


Figura 1. Texturas de Broadtz mapeadas no plano $\mathcal{H} \times \mathcal{C}$

Desse modo, descreveremos brevemente como foi realizado esse estudo e quais resultados foram adquiridos com o mesmo.

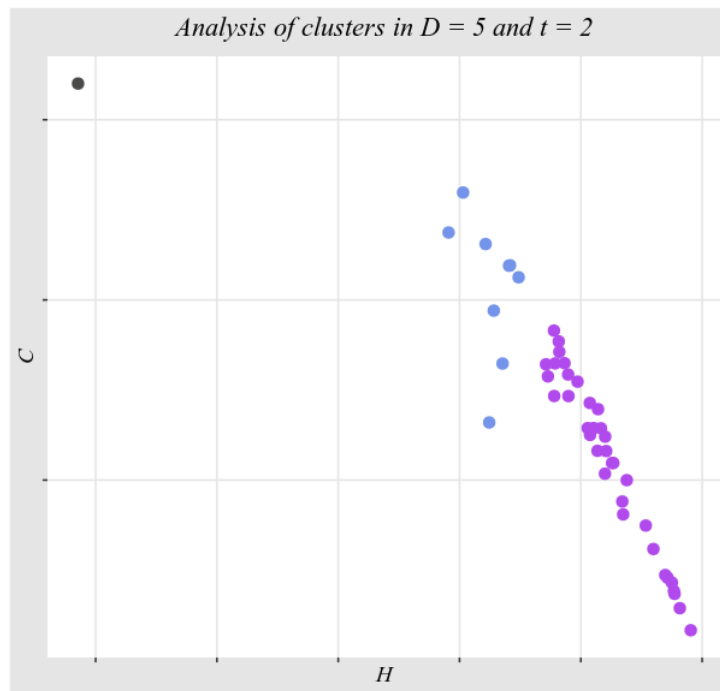


Figura 2. Clusters visualmente formados no plano $\mathcal{H} \times \mathcal{C}$

2. Metodologia

O primeiro passo foi escolher um método de agrupamento que fosse adequado para o problema apresentado. Assim, algumas características e apontamentos foram levados em consideração durante a escolha, foram eles:

- Queremos encontrar uma técnica eficiente e eficaz na detecção de clusters "naturais" e seus arranjos no plano $\mathcal{H} \times \mathcal{C}$;
- Não possuímos nenhuma informação prévia dos grupos que podem ser formados, logo trata-se de um estudo de aprendizagem não supervisionada;
- Os clusters podem ter formatos arbitrários e diferentes tamanhos.

Sabendo disto, optamos por escolher o algoritmo DBSCAN para realizar o agrupamento dos descritores, que se trata de uma método não-paramétrico baseado em densidade. O funcionamento do algoritmo se baseia na ideia de que para cada ponto de um certo cluster, a vizinhança para um dado raio contém, no mínimo, um determinado número de pontos, isto é, a densidade da vizinhança tem que exceder um limiar.

Para isso, utilizamos o pacote `dbscan` para aplicar a técnica nos dados dos descritores e a configuramos, de modo empírico, com os seguintes valores:

- Semente 123;
- Tamanho da vizinhança $\epsilon = 0.035$;
- Número mínimo de pontos na região 2.

3. Resultados e conclusões

Após aplicar o DBSCAN conseguimos verificar a formação de alguns clusters nos planos, como podemos ver na Figura 3.

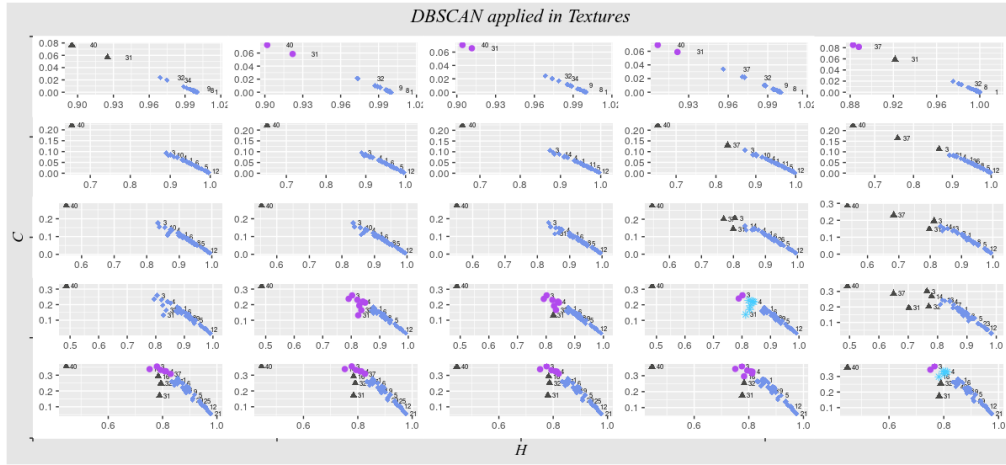


Figura 3. Clusters formados no plano $\mathcal{H} \times \mathcal{C}$ pelo algoritmo DBSCAN

Como resultado, vemos que os planos $\mathcal{H} \times \mathcal{C}$ ($D = 5, \tau = 2$), ($D = 5, \tau = 3$) e ($D = 5, \tau = 4$) nos mostram que os dados podem ser separados em dois ou até três subgrupos de texturas com comportamento similares.

Possuímos agora como trabalhos futuro verificar se os grupos formados com o algoritmo possuem alguma informação semântica similar, nos revelando alguma característica relevante acerca da dinâmica dos seus elementos.