

Caracterização de imagens PolSAR utilizando Bandt-Pompe PDF e Teoria da Informação

Danilo Fernandes¹, Eduarda Chagas², , Roger Almeida¹

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. Processo de simbolização de Bandt-Pompe para padrões bidimensionais

Para aplicarmos a simbolização de dados bidimensionais seguindo a metodologia proposta por [1] devemos considerar, em ambas dimensões, os parâmetros utilizados no algoritmo original. Para fins didáticos, iremos assumir como exemplo uma matriz de tamanho 3 x 3, definida a seguir.

$$X = \begin{bmatrix} 3 & 4 & 8 \\ 5 & 6 & 7 \\ 2 & 8 & 9 \end{bmatrix}$$

O primeiro passo é definir as submatrizes deslizantes e para isso quatro parâmetros são necessários: As dimensões $D_x, D_y \geq 2$, que são o número de elementos que iram formar os padrões ordinais em ambas dimensões e os delays τ_x e τ_y , que informam o quão separados espacialmente estão os símbolos nas duas direções. Neste exemplo, assumiremos $D_x = D_y = 2$ e $\tau_x = \tau_y = 1$, obtendo os seguintes quatro particionamentos:

$$\left[A = \begin{bmatrix} 3 & 4 \\ 5 & 6 \end{bmatrix}, B = \begin{bmatrix} 4 & 8 \\ 6 & 7 \end{bmatrix}, C = \begin{bmatrix} 5 & 6 \\ 2 & 8 \end{bmatrix}, D = \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} \right]$$

Após realizar esta subdivisão, devemos investigar quais padrões aparecem dentro dos elementos das submatrizes. Para isto, iremos analisar os elementos das destas conjuntos linha por linha, assim $\Pi_a = (0, 1, 2, 3)$, pois ao permutar ordenadamente os elementos teremos $a_1 < a_2 < a_3 < a_4$. Logo, vamos ter $\Pi_b = (0, 2, 3, 1)$, $\Pi_c = (2, 0, 1, 3)$ e $\Pi_d = (0, 1, 2, 3)$.

Para todos os padrões ordinais associados a X nós calculamos a distribuição de probabilidade e assim podemos calcular os descritores causais citados anteriormente.

2. Simulação numérica

Nossos resultados são baseados em amostras correspondentes à banda HHHH de um conjunto de imagens SAR. Para isso, fizemos um estudo das seguintes regiões capturadas em dados SAR:

- Parque Nacional Sierra del Lacandon, Guatemala (adquirido em 10 de abril de 2015), disponível em https://uavsar.jpl.nasa.gov/cgi-bin/product.pl?jobName=Lacand_30202_15043_006_150410_L090_CX_01#dados;
- Regiões oceânicas do Cabo Canaveral (adquirido em 22 de setembro de 2016);
- Área urbana da cidade de Munique, na Alemanha (adquirido em 5 de junho de 2015).

Para aplicar as técnicas aqui definidas, retiramos amostras de dimensão 200×200 e no total utilizamos 20 regiões, assim definidas:

- Quatro regiões florestais da Guatemala;
- Uma região de cultivo da Guatemala;
- Três regiões terrestres da Guatemala caracterizadas por apresentarem um comportamento não uniforme;
- Oito regiões do Cabo de Canaveral, possuindo características de dois comportamentos diferentes;
- Quatro regiões urbanas da cidade de Munique.

Cada amostra foi redimensionada para um conjunto de matrizes obtidas de partições deslizantes da imagem onde testamos o conjunto de valores $(2, 3, 4, 5, 6)$ para as dimensões D_x e D_y das partições geradas. Para o delay usamos os valores $(1, 2, 3, 4, 5)$. O processo de simbolização de Bandt-Pompe é realizado para cada partição, sendo importante salientar que cada dimensão D_x e D_y leva a $(D_x D_y)!$ possíveis padrões ordinais.

Enfatizamos que, o uso de rotinas otimizadas implementadas na linguagem C melhoraram notavelmente o tempo de processamento do experimento, quando comparado com as tradicionais rotinas implementadas anteriormente em R. Para adquirir a distribuição de probabilidade de Bandt-Pompe é necessário chamar por meio da interface `.Call()` uma função inteiramente escrita em C. Esta função em C recebe como parâmetros uma matriz contendo as amostras já redimensionadas em particionamentos, a quantidade de colunas que a matriz possui, o que equivale à dimensão D , e a quantidade de linhas, que representa a quantidade de casos a serem analisados.

Para cada matriz de representação, calculamos duas medidas de complexidade: Entropia de permutação normalizada H e a Complexidade estatística C . Em seguida, aplicamos os valores H e C no plano complexidade-entropia que consiste de uma poderosa ferramenta de discriminação e quantificação das diferentes características dos dados.

3. Resultados e conclusões

Sabendo que objetivamos caracterizar as diferentes regiões coletadas em dados SAR, realizamos testes com alguns valores de dimensão e delay. Avaliamos a implicação da modificação destes parâmetros no processo de caracterização destes dados no plano Complexidade-Entropia, como podemos verificar abaixo.

No gráfico gerado com todas as configurações de dimensões e delays, visualmente representamos as diferentes regiões do seguinte modo:

- Regiões florestais – Triângulos cinza;
- Regiões de cultivo – Triângulos azul escuro;

- Regiões não uniforme – Triângulos violeta;
- Regiões oceânicas (Comportamento 1) – Losangos azul;
- Regiões oceânicas (Comportamento 2) – Losangos verdes;
- Regiões urbanas – Círculos laranjas.

Após realizada a caracterização destes dados, algumas questões e pontos despertaram a curiosidade:

- Porque os dados parecem seguir um comportamento linear?
- Os dados associados a regiões urbanas aparecem muito espaçados em todas as configurações de dimensão e delay. Seria isto um comportamento típico deste tipo de dado?
- Regiões de cultivo e Regiões oceânicas (comportamento 1) deram sinais de que podem se encontrar em áreas semelhantes do plano.
- Ao analisar apenas os dados de regiões da Guatemala, a melhor configuração para diferenciar as imagens foram os planos HC ($D = 4, \tau = 5$) e ($D = 6, \tau = 1$). Entretanto, não podemos concluir que estas configurações se aplicam para as novas amostras estudadas.
- Outra característica interessante foi o destaque dos diferentes comportamentos dos dados das regiões oceânicas do Cabo de Canaveral.

Desse modo, desejamos que com a finalização deste trabalho tais questionamentos sejam respondidos.

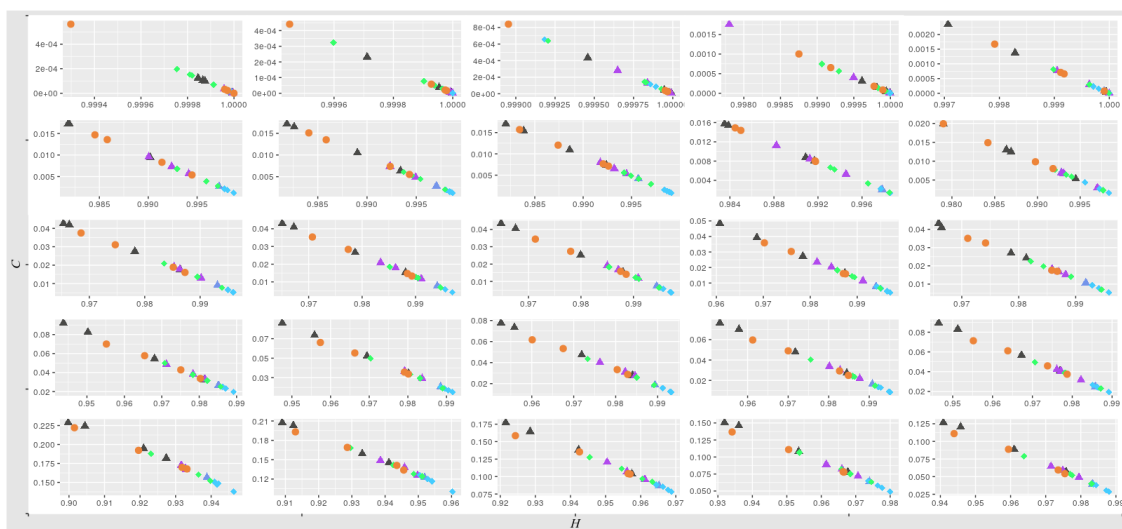


Figura 1. Plano Complexidade-Entropia aplicado as amostras correspondentes à banda HHHH de uma imagem SAR. Verticalmente temos as variações dos valores de delay e horizontalmente dos diferentes valores de dimensão aplicados.

Referências