

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 Bandt & Pompe (2002) 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 destes 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 $\Pi_b = (0, 2, 1, 3)$, $\Pi_c = (3, 0, 1, 2)$ 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

Usamos como fonte de dados uma imagem SAR tirada do 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](https://drive.google.com/file/d/1-tBmid6Lz_ps_L3OpVVnoR64cENGzR10/view?usp=sharing). Nossos resultados foram baseados em um pequeno conjunto de amostras correspondentes à banda HHHH desta imagem SAR, podendo ser acessado por: https://drive.google.com/file/d/1-tBmid6Lz_ps_L3OpVVnoR64cENGzR10/view?usp=sharing.

Para aplicar as técnicas aqui definidas, consideramos as seguintes configurações nos dados selecionados:

- Foram retiradas amostras com dimensão 200x200;
- Ao total utilizamos oito regiões, assim definidas:
 - Quatro regiões de regiões florestais na Guatemala;
 - Uma região correspondente a regiões de cultivo na Guatemala;
 - Três regiões representam regiões terrestres caracterizadas por apresentarem um comportamento não uniforme.

Cada amostra é então redimensionada para um conjunto de matrizes obtidas de partições deslizantes da imagem. Assim, testamos o conjunto de valores (2, 3, 4, 5, 6) para as dimensões D_x e D_y das partições geradas. Com o delay usamos os valores (1, 2, 3, 4, 5). Para cada partição o processo de simbolização de Bandt-Pompe é realizado, 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 a adquirir a distribuição de probabilidade de Bandt-Pompe chamamos 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 de coloração cinza;
- Regiões de cultivo – Losangos de coloração azul;
- Regiões não uniforme – Círculos de coloração lilás.

Como queremos quantificar e discriminar as características de diferentes tipos de texturas no plano Complexidade-Entropia, os melhores resultados obtidos após essa breve avaliação foram as combinações das seguintes configurações ao aplicar o processo de simbolização de Bandt-Pompe:

- $D = 4 \text{ e } \tau = 5$;
- $D = 6 \text{ e } \tau = 1$.

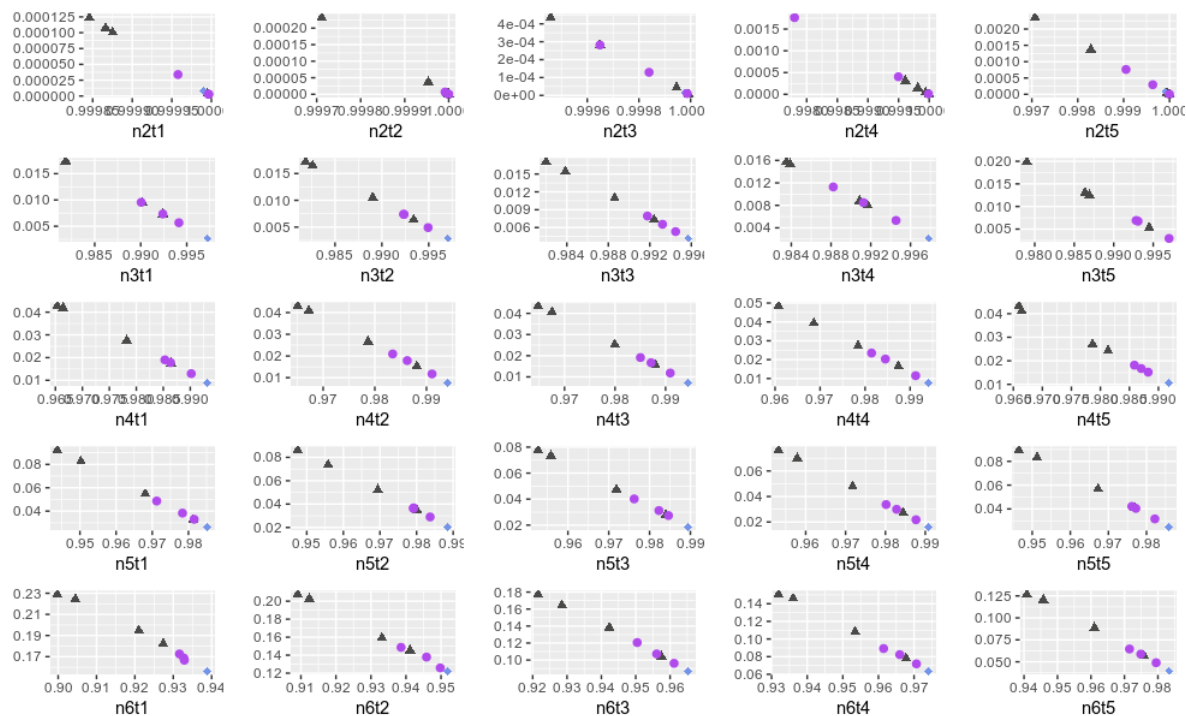


Figura 1. Plano Complexidade-Entropia aplicado as amostras correspondentes à banda HHHH de uma imagem SAR. Verticalmente temos as variações dos valores de delay e horizontalmente dos diferentes valores de dimensão aplicados.

Referências

Bandt, C. & Pompe, B. (2002), 'Permutation entropy: A natural complexity measure for time series', *Physical review letters* **88**, 174102.