



Inferência em Classificação de Séries Temporais

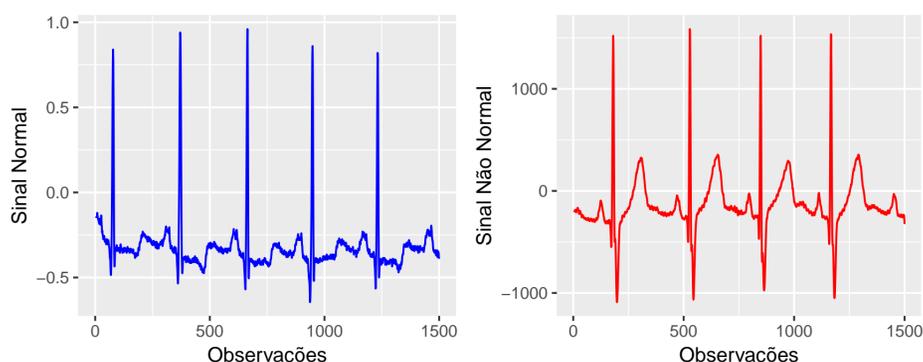
Autora: Mikaela Baldasso

Orientador: Marcio Valk

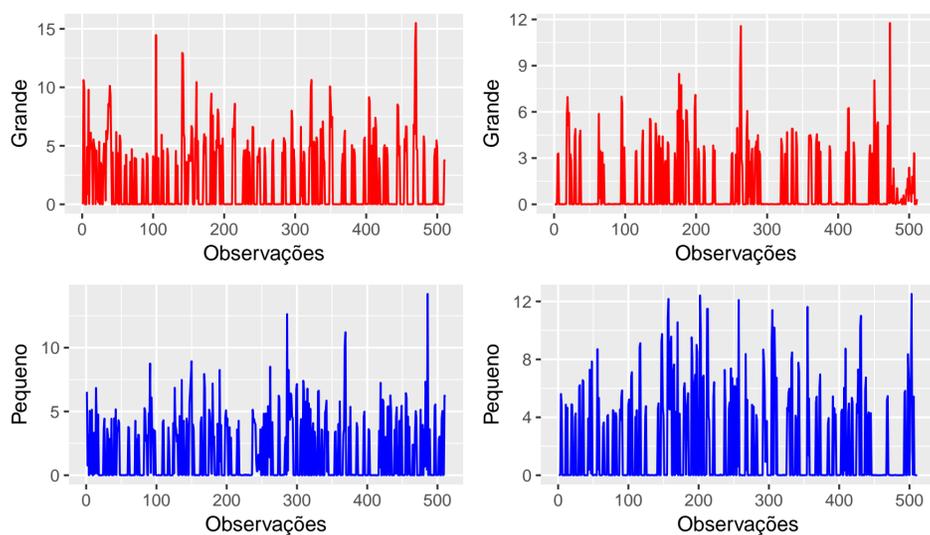
Problemas Reais

Atualmente, nota-se uma crescente demanda para utilização de métodos de classificação e agrupamento de séries temporais, onde é comum a coleta de dados que fornecem importantes informações.

O objetivo desse estudo é abordar problemas complexos de classificação como a identificação de sinais. Dentre esses estão os sinais cardíacos com arritmia e sinais normais, sinais sísmicos que precedem grandes terremotos, entre outros.



Exemplo de um sinal cardíaco normal, em azul, e um sinal cardíaco com arritmia, em vermelho



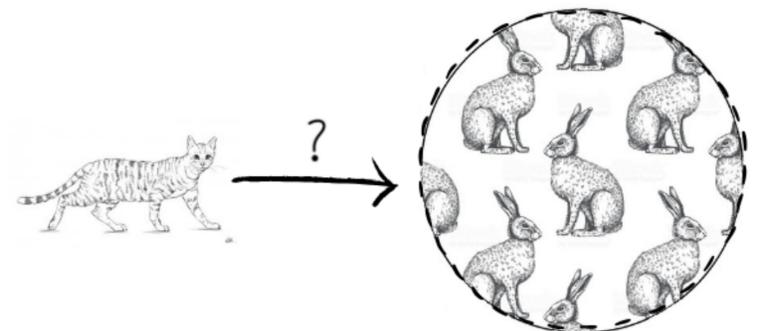
Exemplos de dois sinais que precedem grandes terremotos, em vermelho, e dois sinais que precedem pequenos terremotos, em azul.

Métodos de agrupamento para identificar padrões

O método de *clustering* é um conjunto de técnicas computacionais cujo propósito consiste em separar objetos em grupos distintos de acordo com as características que eles apresentam. De forma geral, a técnica consiste em colocar elementos similares em um mesmo grupo de acordo com algum critério já estipulado.

Metodologia

Dado um grupo de sinais normais, o método de classificação proposto nesse estudo consiste em classificar um novo sinal como pertencente ou não pertencente a ele, por meio do método *Utest*. Ou seja, identificar, via teste de hipótese, se o novo sinal é normal ou não.



Simulações de Monte Carlo

| | Proporção de Rejeição do <i>Utest</i> | | |
|-----------------|---------------------------------------|------------|------------|
| | $n_2 = 10$ e $\phi_2 = 0.4$ | | |
| | $n_1 = 10$ | $n_1 = 20$ | $n_1 = 30$ |
| $\phi_1 = -0.4$ | 1.00 | 1.00 | 1.00 |
| $\phi_1 = -0.6$ | 1.00 | 1.00 | 1.00 |
| $\phi_1 = -0.4$ | 1.00 | 1.00 | 1.00 |
| $\phi_1 = -0.2$ | 0.99 | 1.00 | 1.00 |
| $\phi_1 = 0.0$ | 0.99 | 1.00 | 1.00 |
| $\phi_1 = 0.2$ | 0.18 | 0.48 | 0.54 |
| $\phi_1 = 0.4$ | 0.01 | 0.04 | 0.04 |
| $\phi_1 = 0.6$ | 0.94 | 0.97 | 0.99 |
| $\phi_1 = 0.8$ | 1.00 | 1.00 | 1.00 |

Nesse estudo utilizamos os processos autorregressivos de ordem 1 (AR(1)) para gerar os grupos. O processo é definido por $Y_t = \phi y_{t-1} + \varepsilon_t$, em que o parâmetro ϕ deve satisfazer $|\phi| < 1$ e ε_t é um ruído branco gaussiano. A tabela ao lado, é uma representação da proporção de rejeição do método *Utest*. Nela, a série temporal que compõem o grupo 2 (G_2) é gerada com $n_2 = 1$ e $\phi_2 = 0.4$ e as séries que compõem o grupo 1 (G_1) são geradas a partir de diferentes valores para n_1 e ϕ_1 .

Resultados

Durante a execução do presente trabalho, realizamos simulações de Monte Carlo em um contexto controlado que sugeriram que o método *Utest* pode ser usado para caracterizar sinais com dinâmicas diferentes. Exploramos vários bancos de dados e obtivemos resultados significativos na aplicação do método na classificação de sinais cardíacos: sob H_0 , o teste errou 1 dos 18 sinais, que está dentro dos 5% aceitáveis e, sob H_1 , o teste teve 100% de acertos.

Referências

- [1] Valk, M & Cybis, GB (2019). U-statistical inference for hierarchical clustering.
- [2] Cybis, GB, Valk M and Lopes, SRC (2018). "Clustering and classification problems in genetics through U-statistics." Journal of Statistical Computation and Simulation 88.10: 1882-1902.