

Apresentação Resultados das simulações

METODO MAD-STEC

- Forma mais robusta do método STEC proposto em **Assunção e Correia (2009)**.
- Se baseia em um método iterativo onde, ao encontrar um cluster, reduz o número de observações dentro do cluster e aplica novamente o método.
- O número de observações reduzidas é igual ao número de casos observados menos o número de casos esperados dentro do cluster.
- Os dados retirados são escolhidos de forma aleatória. Caso um novo cluster seja detectado e ele tenha interseção com o primeiro cluster, a interseção é retirada e o modelo aplicado novamente.
- Esse processo é feito quantas vezes necessárias para achar o número de clusters presentes na amostra.

NOSSO METODO - I

1. Divida o banco em treino ($1, \dots, k$) e teste ($k+1, \dots, N$). Nosso método irá trabalhar com uma amostra pequena de treino (15%).
2. Em cada b iteração, com $b = 1, \dots, B$, os tempos serão embaralhados. Crie as features e treine o modelo para cada b reembaralhamento. As features são:
3. Para cada uma das observações ($k+1, \dots, N$), crie as features com base nos dados originais e predize como sendo anomalia ou não para cada um dos B treinamentos:
 - Caso seja classificada como anomalia (pelo menos 90% dos treinamentos), ela pertencerá ao cluster.
 - caso não seja, adicione-a ao banco de dados treino (Testamos com e sem essa adição).
4. Caso na observação j ($j > k$) seja observado que já existem um número A de observações classificadas como cluster, e essas observações estão contidas no mesmo cilindro, acione o alarme e pare o processo.

NOSSO METODO - II

As features utilizadas são:

- **Var1:** Distância até a última ocorrência. Caso tenha duas observações com tempo igual, pega a de menor distância.
- **Var2:** Diferença de tempo até a observação mais recente do cilindro de raio ρ e comprimento *time_window*
- **Var3:** Número de observações dentro do cilindro.

Perceba que para essas variáveis, excluimos as observações dentro do intervalo de tempo de t_1 (primeiro tempo) a *time_window*, como uma espécie de queima ou aquecimento, já que as observações no intervalo iniciarão com valores discrepantes do resto.

NOSSO METODO - III

Algumas mudanças foram consideradas ao algoritmo original:

- Foi testado de duas formas, com acrescimo das classificadas como não cluster ao treino (modelo 2) e sem o acrescimo (modelo 1).
- No modelo 2, os treinos eram refeitos apenas quando 10 observacoes fossem classificadas como não sendo anomalias, de modo a não treinar novamente para cada observacao e reduzir o custo computacional.
- O alarme foi alterado para 3 observações ao invés de 5, número de repetições monte carlo para 10, iterações para 30.
- Quando um alarme soava, as observações que estavam dentro do cilindro para ativação do alarme eram retiradas, para que não soasse o mesmo alarme repetidamente.
- Como as observações são preditas uma a uma, foi também observado a diferença de tempo entre o primeiro alarme correto e o tempo de inicio do cluster.

TIPOS DE ESTUDO E MODELOS (Atigo Marquinhos)

Cada um dos dois modelos foi testado segundo:

Estudo 1: Treinamento do modelo em um conjunto de dados gerados sem nenhum cluster, 300 observações.

Estudo 2: Treinamento do modelo em um conjunto de dados gerados com 1 único cluster. 350 observações no total, coordenadas do cluster: $x = [0.95, 1]$, $y = [0.95, 1]$ e $t = [0.95, 1]$

Estudo 3: Treinamento do modelo em um conjunto de dados gerados com 2 clusters em tempos e locais diferentes. 400 observações no total, coordenadas cluster 1: $x = [0.95, 1]$, $y = [0.95, 1]$ e $t = [0.95, 1]$, coordenadas cluster 2: $x = [0.50, 0.55]$, $y = [0.70, 0.75]$ e $t = [0.60, 0.65]$

Os estudos foram feitos segundo simulação Monte Carlo onde $M = 10$, as coordenadas variam entre 0 e 1, tanto (x, y) e t .

TERMOS UTILIZADOS

Falso Cluster - Se a observação foi classificada como anomalia mas não estava presente no intervalo do cluster.

Cluster Correto - Se a observação foi classificada como anomalia e estava presente no intervalo do cluster.

Não Cluster Correto - Se não foi classificada como anomalia e não estava presente no intervalo do cluster.

Não Cluster Incorreto - Se não foi classificada como anomalia e estava presente no intervalo do cluster.

Alarme incorreto - Se o alarme soou fora do tempo e coordenadas do cluster.

Alarme correto - Se o alarme soou dentro do tempo e coordenadas do cluster.

RESULTADO RESUMIDO - I

Aqui é apresentado o resultado médio para cada modelo e estudo

Estudo 1							
	Não Cluster Correto	Falsos Cluster	Alarme Incorreto				
Modelo 1	230.6	9.4	0				
Modelo 2	232.6	7.4	0				
Estudo 2							
	Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto
Modelo 1	207.1	2.1	48.0	22.8	0	45.1	0.0064927
Modelo 2	211.5	1.6	48.4	18.5	0	45.6	0.0052209

RESULTADO RESUMIDO - II

Estudo 3

	Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto (cluster 1)	Tempo até o primeiro alarme correto (cluster 2)
Modelo 1	194.2	4.1	95.9	28.0	0.1	90.6	0.0048949	0.0060397
Modelo 2	193.9	3.3	96.8	27.9	0.2	91.4	0.0060411	0.0039639

RESULTADOS ESTUDO 1 - Modelo 1

Não Cluster Correto	Falsos Cluster	Alarme Incorreto
235	5	0
228	12	0
231	9	0
235	5	0
234	6	0
219	21	0
227	13	0
239	1	0
221	19	0
237	3	0

RESULTADOS ESTUDO 1 - Modelo 2

Não Cluster Correto	Falsos Cluster	Alarme Incorreto
235	5	0
233	7	0
236	4	0
226	14	0
229	11	0
234	6	0
232	8	0
233	7	0
236	4	0
232	8	0

RESULTADOS ESTUDO 2 - Modelo 1

Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto
212	4	46	18	0	43	0.0071561
210	3	47	20	0	44	0.0051367
202	2	48	28	0	46	0.0059994
200	2	48	30	0	45	0.0057512
192	1	49	38	0	45	0.0086899
215	2	48	15	0	46	0.0032095
213	1	49	17	0	45	0.0053110
204	2	49	25	0	45	0.0094279
207	3	47	23	0	45	0.0096931
216	1	49	14	0	47	0.0045524

RESULTADOS ESTUDO 2 - Modelo 2

Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto
214	3	47	16	0	43	0.0054835
197	1	49	33	0	47	0.0040784
213	0	50	17	0	47	0.0008157
218	2	48	12	0	44	0.0042950
217	2	48	13	0	46	0.0071690
210	2	48	20	0	45	0.0050911
215	2	48	15	0	46	0.0062121
206	1	49	24	0	46	0.0036521
212	2	48	18	0	46	0.0062300
213	1	49	17	0	46	0.0091823

RESULTADOS ESTUDO 3 - Modelo 1

Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto 1	Tempo até o primeiro alarme correto 2
197	5	95	26	1	90	0.0039346	0.0062439
203	4	96	19	0	90	0.0081673	0.0023576
188	4	96	34	0	90	0.0051677	0.0062143
203	5	95	20	0	89	0.0068832	0.0084238
192	2	98	29	0	93	0.0029185	0.0071212
188	6	94	35	0	90	0.0048550	0.0057559
199	3	97	23	0	90	0.0054058	0.0063596
189	4	96	33	0	91	0.0035949	0.0046183
192	3	97	30	0	92	0.0048618	0.0024198
191	5	95	31	0	91	0.0031602	0.0108825

RESULTADOS ESTUDO 3 - Modelo 2

Não Cluster Correto	Não Cluster Incorreto	Falsos Cluster	Cluster Correto	Alarme Incorreto	Alarme Correto	Tempo até o primeiro alarme correto 1	Tempo até o primeiro alarme correto 2
198	4	96	24	0	91	0.0061262	0.0037418
200	2	98	21	0	92	0.0034332	0.0040009
201	5	96	22	1	90	0.0071655	0.0032496
185	4	96	37	0	91	0.0054671	0.0025294
194	5	95	28	0	89	0.0041409	0.0056643
196	4	96	27	0	91	0.0093588	0.0017366
177	2	98	44	0	91	0.0042704	0.0056314
198	2	98	23	0	93	0.0042904	0.0025392
201	2	98	20	0	93	0.0081410	0.0067591
189	3	97	33	1	93	0.0080172	0.0037869