



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO

Previsão de crimes em Valência: Uma abordagem multi-rótulo

Introdução

- Aumento da atividades criminosas
- Alguns fatores determinantes:
 - *Áreas com alta concentração populacional*
 - *Agentes socioeconômicos*
 - *Fatores espaciais e entre outros...*
- Recentes pesquisas:
 - *HOSSAIN (2020)*
 - *BOGOMOLOV (2014)*
 - *RODRIGUES (2010)*
- Classificação multirrótulo
- Dados temporais e espaciais

Metodologia

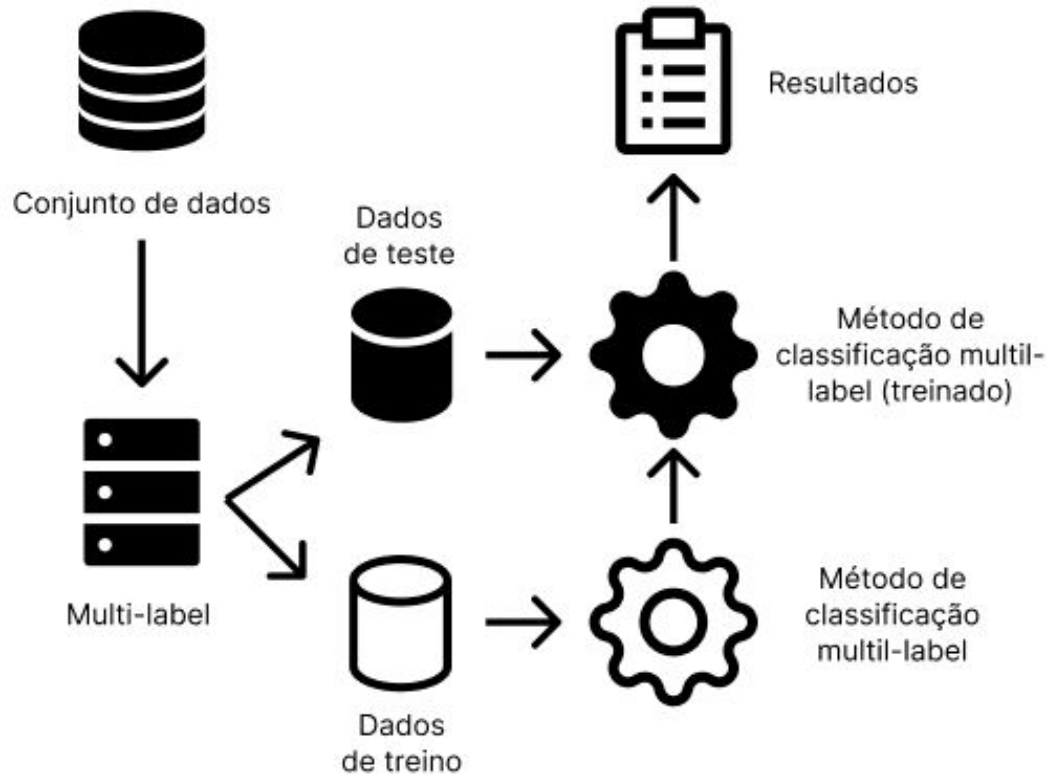


Figura 1 - Fluxograma com visão global da metodologia proposta.

Conjunto de dados

- Conjunto de dados de atividades criminosas
- Cidade de Valência, Espanha.
- Entre 2010 a 2020.
- Atributos:
 - *Dados históricos*
 - *Latitude e Longitude*
 - *Distância até pontos de acesso*
- Dados meteorológicos

Multirrótulo

- Dados classificados em duas ou mais classes.
- Prever propriedades de uma instâncias que não são mutuamente exclusivas.
- Ex.: Previsão de tópicos importantes de um documento (política, economia, educação e entre outros...)

Instância	Atr1	Atr2	Atr3	Atr4	L1	L2	L3	L4	L5
x_1	25	58	24	57	1	0	1	1	0
x_2	43	38	38	781	1	1	1	0	0
x_3	8	73	24	70	0	1	0	1	0
x_4	79	9	65	63	1	0	0	0	1
x_5	100	61	5	48	1	0	1	1	0

Figura 2 - Exemplo do problema multirrótulo

Multirrótulo

- Matriz de dias por rua
- Dias como instâncias e ruas como rótulos
- Rua mais próxima do crime
- 3.652 amostras (dias) e 2078 rótulos (ruas)
- XMC - (*Extreme Multi Label Classification*)

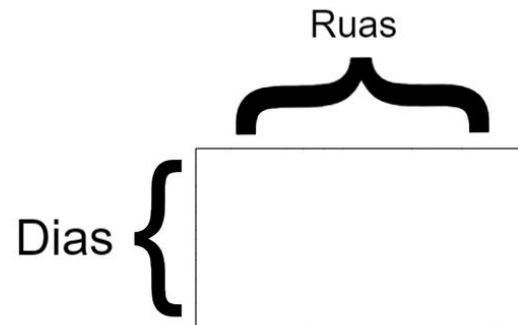


Figura 3 - Exemplo do conjunto de dados resultante

Algoritmo Bonsai

- Generaliza o conceito de representação de rótulos
- Árvores rasas com número de filhos variada
- Parabel (PRABHU et al., 2018)
- Como forma de comparação, Binary Relevance (BR)

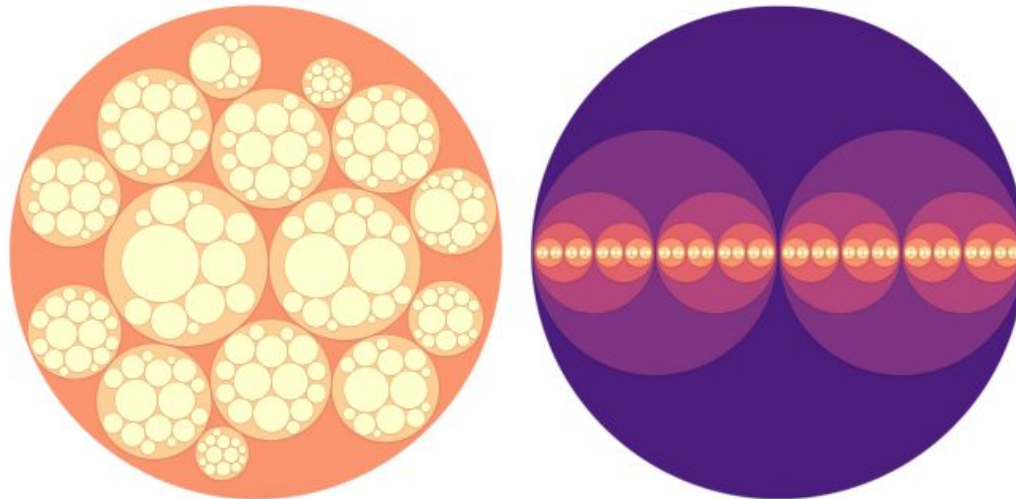


Figura 4 - Diferença na geração de árvores nos métodos Bonsai (esquerda) e Parabel (direita)

Métrica de avaliação

- Precisão@K (Precisão em K)
- Amplamente utilizada:
 - *JASINSKA-KOBUS et al. (2020)*
 - *PRABHU et al. (2018)*
 - *MITTAL et al. (2022)*

- $k \in \{1, 3, 5\}$

- $Precis\tilde{a}o@k(\hat{y}, y) := \frac{1}{k} \sum_{l \in \text{rank}_k(\hat{y})} y_l$

Experimentos

- 3 conjuntos de dados
 - *Dados históricos*
 - *Dados meteorológicos*
 - *Dados espaciais*
- 3 métodos estabelecidos anteriormente
- 70% para treino e 30% para teste
- Base de dados sequenciais

Resultados

- Resultados
 - $P@1 = 29.51$
 - $P@3 = 23.59$
 - $P@5 = 21.37$
- 3 métodos estabelecidos anteriormente
- 70% para treino e 30% para teste
- Base de dados sequenciais

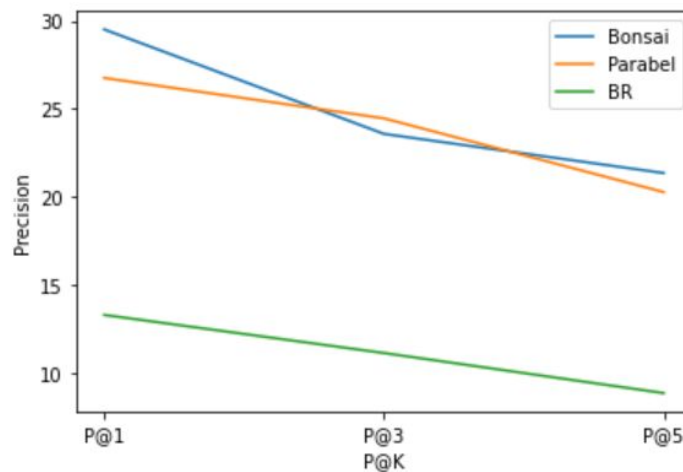


Figura 5 - Gráfico de comparação entre métodos

Resultados



Figura 6 - Região com alta probabilidade de ocorrências criminosas

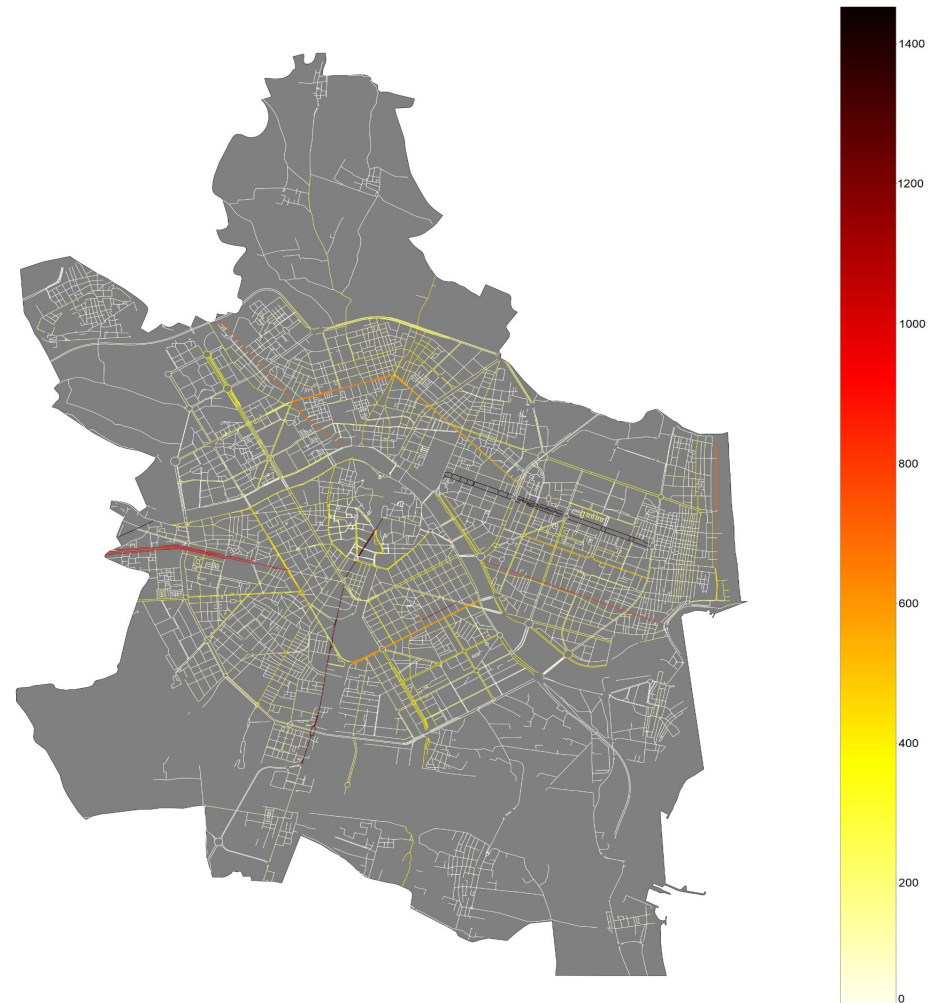


Figura 7 - Frequência de crimes por rua

Considerações Finais

- Aumentar o número de características do modelo
- Utilizar o método em outras localizações

Obrigado !