

APRENDIZADO DE MÁQUINA PARA WEBSENSORS: APLICAÇÕES EM MONITORAMENTO DE QUEIMADAS A PARTIR DE NOTÍCIAS

Thábata Amaral, Marcos Nesso, Ricardo Marcacini

Objetivos

A extração de padrões a partir de dados publicados na web visando monitorar, compreender e prever eventos do dia-a-dia é um tema de pesquisa que tem recebido grande atenção. Modelos computacionais conhecidos como websensors permitem tal atividade via algoritmos de aprendizado de máquina em notícias de um domínio de interesse [1]. Neste trabalho, foi modelado um websensor para monitoramento de queimadas no estado de Mato Grosso do Sul, a partir de notícias, visando analisar o impacto das queimadas no número de internação de crianças (até 9 anos) por problemas do aparelho respiratório.

Métodos/Procedimentos

A base de dados utilizada é composta por notícias de Mato Grosso do Sul. Dessa notícias, 177 notícias foram rotuladas como queimadas no contexto de meio ambiente, denominado conjunto de treinamento. A partir do conjunto de treinamento, foram extraídos os parâmetros do websensor (via aprendizado probabilístico), ou seja, um vetor de características que contém um termo (stem extraído do texto) e a probabilidade de estar relacionado a uma notícia de queimada. O websensor resultante contém 271 parâmetros, em que os seis mais relevantes foram:

Foc (1.0)	Incend (0.9)	Fog (0.7)
Parqu (0.6)	Florest (0.5)	Queim (0.5)

O websensor foi empregado então para monitorar a ocorrência de queimadas nas notícias no período de 2012 a 2014. Quando uma nova notícia de queimada é identificada, é gerado um sinal entre [0,1] considerando a correlação entre os parâmetros do websensor os termos presentes no texto da notícia.

Resultados

Na Figura 1, um monitoramento de queimadas realizado pelo INPE (via imagens de satélite) é apresentado no gráfico em azul. O gráfico em vermelho ilustra o websensor que monitora queimadas a partir de notícias. Foi constatada uma correlação significativa (*Pearson*) entre os indicadores. Foi analisada a relação de picos do websensor com o aumento de internações de crianças devido a problemas do aparelho respiratório (IPR), com a diferença de IPR do mês da queimada e o mês subsequente, conforme dados do SUS-SIH de MS.

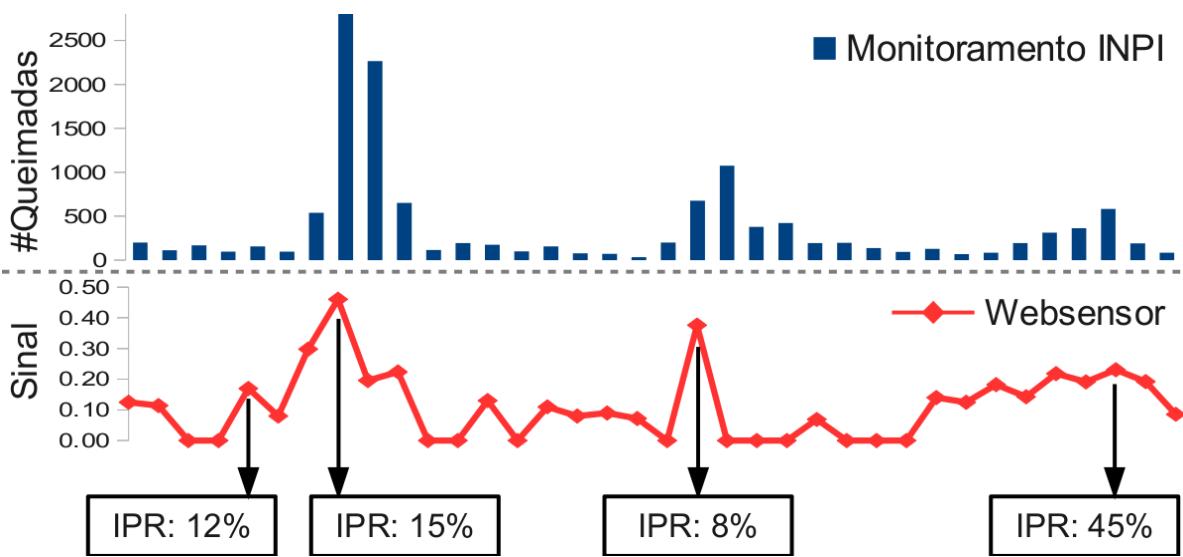


Fig 1: Monitoramento de queimadas via websensor.

Conclusões

Mesmo com dados ruidosos e não estruturados das notícias, foi possível aprender um websensor de queimadas com desempenho similar aos atualmente utilizados. Além disso, os resultados podem ser alinhados com indicadores de saúde pública, permitindo monitorar em tempo real o impacto das queimadas em uma determinada região conforme publicações da mídia e, futuramente, com base em dados de redes sociais.

Referências Bibliográficas

- [1] Marcacini, R., et al. Aprendizado não Supervisionado de Websensors. Kdmile, 2013.