

Roteiro de atividade

Projeto Final

Asa Branca Aerospace - IA

October 2020

1 Descrição do problema

Você está encarregado do monitoramento de uma faixa de mar próxima à costa. Recentemente, um navio ainda não identificado derramou um material que imagina que estivesse sendo desenvolvido dentro do próprio navio. Essa modificação do óleo tem densidade menor que o óleo normal e suas partículas tendem a se agrupar. Dessa forma uma mancha provinda de um vazamento consegue se mover como um corpo só, sendo arrastada pela corrente até chegar à areia. A água é contaminada por partículas da superfície da mancha que se desprendem enquanto o óleo está em movimento. Com dados de vazamentos prévios e imagens aéreas obtidas num intervalo de 3 dias(30s em tempo de simulação), seus superiores querem que você estime a área afetada pelo vazamento em m^2 . Você irá utilizar técnicas de processamento digital de imagens para encontrar o contorno da mancha, encontrar a menor elipse que o englobe e guardar seus parâmetros a cada frame, encontrar o seu centroide em cada frame, registrar a distância entre o ponto inicial e final dos centroides(as posições na primeira e última imagens, respectivamente) e a variação no eixo y(referida como Δy) desses mesmos pontos. Ao final da simulação, desenhe em uma imagem base que será fornecida, todas as posições dos centroides encontradas durante a simulação em uma cor bem distinta do plano de fundo junto com as elipses cujos parâmetros foram registrados durante a simulação, de forma a representar a trajetória da mancha durante os três dias de monitoramento. Além disso, encontre uma estimativa da área contaminada como descrito a seguir.

2 Dados fornecidos

Você tem duas opções: aproximar a função de área afetada utilizando uma dataset de 2500000 *samples* com uma rede neural ou utilizar regressão numa dataset de 70000 *samples*. Note que, o primeiro método, se feito de forma correta, encontra com mais precisão o resultado desejado. Os dados contêm a área da menor elipse que engloba todo o contorno da mancha, a distância euclidiana entre o centroide da mancha entra a última e a primeira imagem, e

a variação no eixo y da posição da mancha(centroide). A coluna de targets, é a que contém a área contaminada. Os valores acima estão medidos em pixels(px), na distância e no delta y, e pixels ao quadrado(px2) no caso da área da elipse.

3 Definições

Serão realizados 30s de simulação, durante os quais se obterão o conjunto F de n imagens.

$$F = \{I_1, I_2, \dots, I_n\} \quad (1)$$

n será o mesmo em qualquer um das definições abaixo(Note que n é a cardinalidade de F). Para cada imagem em F , a posição do centroide do contorno da mancha, obtida através de thresholding e momentos de imagem é elemento de um conjunto M de centroides μ :

$$M = \{\mu_1, \mu_2, \dots, \mu_n\} \quad (2)$$

$$\mu_i = (x_i, y_i) \quad (3)$$

A distância entre as posições iniciais e finais do centroide da mancha em I_1 e I_n , é dada por:

$$d(\mu_1, \mu_n) = \sqrt{(\mu_{n1} - \mu_{11})^2 + (\mu_{n2} - \mu_{12})^2} \quad (4)$$

Já a variação destes no eixo y é dado por:

$$\Delta y = y_f - y_0 \quad (5)$$

onde $y_0 = \mu_{12}$ e $y_f = \mu_{n2}$. Além disso, o conjunto Γ contendo os tensores de parâmetros da menor elipse a englobar a extensão do contorno da mancha em cada imagem é descrito como:

$$\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\} \quad (6)$$

$$\gamma_i = (u_i, v_i, r_i, R_i, \theta_i) \quad (7)$$

onde u e v são a posição da origem da elipse, r_i e R_i são os raios da elipse e θ , seu ângulo de rotação. Por fim, devido a variação da iluminação parte do contorno pode deixar de ser registrado, por isso encontre a área média das elipses em Γ baseado na formula:

$$A_m = \frac{1}{n} \sum_{i=1}^n 2\pi\gamma_{i3}\gamma_{i4} \quad (8)$$

os índices 3 e 4 representam a posição dos raios em cada tensor. Baseado nos resultados obtidos em (8), (4) e (5), o algoritmo de aprendizagem de máquina, seja regressão linear ou um aproximador de função baseado em ANN, deve estimar A_c (Área contaminada), de modo em que o algoritmo utilizado seja considerado preciso o suficiente de acordo com a avaliação estabelecida no Anexo

I. A resposta final para o projeto será um arquivo de texto, de acordo com o modelo. Junto a isso, utilize os elementos de Γ e a imagem 'graph.png' para construir a representação da trajetórias da mancha. Siga as instruções dentro do modelo de arquivo correspondente ao método de aprendizagem de máquina utilizado.

Anexo I

Método de avaliação

1 Avaliação de aproximador baseado em MLP

Existe uma função a partir da qual os dados foram gerados que determina a real medida de área contaminada em px^2 . A saída do seu modelo será referida como $h(A_e, \Delta y, d)$. Essas duas funções geram o conjunto T de targets(gerado por h), e P de previsões(gerado pelo seu modelo). Após isso, os valores de cada conjunto forma convertidos para m^2 . Assim, temos:

$$T = \{t_1, t_2, \dots, t_n\}$$

$$P = \{p_1, p_2, \dots, p_n\}$$

Define-se então um coeficiente de pertencimento a margem de erro, que indica se determinado p está dentro de uma margem de até 3 metros de erro em relação a t .

$$\text{coef}(p, t) = \begin{cases} 1 & , |p - t| \leq 3 \\ 0 & , \text{caso contrário} \end{cases}$$

E por fim, obtém-se o índice de estimativa correta da seguinte forma:

$$i_{ec} = \frac{1}{n} \sum_{i=1}^n \text{coef}(p_i, t_i) \quad , n = |P| = |T|$$

O modelo é aprovado se $i_{ec} \geq 0.5$.

2 Avaliação de regressão linear

Os dados fornecidos para regressão foram obtidos da seguinte forma:

$$y = \theta_1 A_e + \theta_2 d(\mu_1, \mu_n) + \theta_3 \Delta y$$

O modelo é aprovado se cada θ for previstos corretamente.