

Universidade Federal do Rio de Janeiro
Instituto Alberto Luiz Coimbra de
Pós-Graduação e Pesquisa de Engenharia



Programa de Engenharia de Sistemas e
Computação

CPS863 - Aprendizado de Máquina
Prof. Dr. Edmundo de Souza e Silva
(PESC/COPPE/UFRJ)

Classificação de embarcações utilizando
técnicas de Visão Computacional

Luiz Henrique Souza Caldas
email: lhscaldas@cos.ufrj.br

14 de outubro de 2024

Questão 1

Este exercício foi feito em classe no dia 10/Out/2024. A lista contém as questões resolvidas e ainda alguns itens a mais. Faça a lista e complete as questões que faltaram.

Este exercício é motivado pelo trabalho em <https://ieeexplore.ieee.org/document/9006548>, Seção H (Leveraging spatio-temporal correlation across homes). O problema foi simplificado neste exercício.

Imagine que dispomos de um classificador implementado em roteadores residenciais de um provedor de Internet (ISP). A cada janela de tempo (por exemplo a cada 5 minutos) o classificador do roteador i fornece como saída uma dentre 2 possibilidades: (a) existe um ataque DDoS acontecendo a partir da residência do roteador i , nesta janela de tempo; (b) não há ataque acontecendo a partir da residência do roteador i nesta janela.

A cada 5 minutos o ISP amostra o resultado de M roteadores escolhidos de forma aleatória dentre todos os roteadores da sua base que, para todos os efeitos deste problema, pode ser considerada como muito grande (infinita). O objetivo do ISP é determinar, a partir das M amostras coletadas, se um ataque aconteceu ou não durante a janela de tempo amostrada. Em outras palavras, o ISP quer determinar a possibilidade de uma das seguintes hipóteses serem verdadeiras: h_a (há um ataque DDoS acontecendo na rede do ISP na janela amostrada) ou h_b que é a hipótese complementar.

O ISP conhece o classificador usado em cada roteador residencial, e sabe que o resultado não é 100% confiável. Portanto, ele usará correlação espacial conforme sugerido no artigo acima e explicado em classe.

No que se segue usaremos algumas definições comuns que podem ser encontradas em https://en.wikipedia.org/wiki/Confusion_matrix (ver também a figura em https://en.wikipedia.org/wiki/Confusion_matrix).

Notação:

- M : número de roteadores amostrados (em uma janela de tempo);
- Inf: variável aleatória (va) indicando se a residência é um “bot”, isto é, está ou não infectada. $P[\text{Inf}]$ ($P[\overline{\text{Inf}}]$) é a probabilidade de uma residência estar infectada (não estar infectada);
- TPR: (true positive rate ou hit rate) taxa de acerto do classificador, ou probabilidade do classificador corretamente sinalizar um ataque, dado que um ataque está acontecendo no ISP (Nota: obviamente somente residências infectadas podem gerar um ataque quando ele ocorre);
- FPR: (false positive rate) ou probabilidade do classificador do roteador residencial erradamente sinalizar um ataque a partir da residência;
- L : variável aleatória indicadora $L = 1$ se o roteador alarma, $L = 0$, caso contrário;
- $P[h_a]$: probabilidade de ocorrer um ataque DDoS no ISP em uma janela de tempo. (Se você tem algum conhecimento prévio sobre ataques, talvez possa estimar o $P[h_a]$).

Suponha que, em uma determinada janela de tempo, das M amostras coletadas, V roteadores sinalizaram que um ataque estava ocorrendo na janela (e então $M - V$ roteadores sinalizaram que tudo estava normal nas suas respectivas residências). Suponha ainda que um ataque ocorre em um intervalo independentemente das infecções nas residências.

Para os seus cálculos, suponha que: $TPR = 0.8$, $FPR = 0.1$, $P[\text{Inf}] = 0.2$. Como você não tem conhecimento prévio sobre $P[h_a]$, suponha inicialmente que $P[h_a] = P[h_b] = 0.5$, $V = 20$, $M = 200$.

Responda as seguintes perguntas, mas só substitua os valores no final:

1. Suponha que um ataque esteja ocorrendo. Calcule $P[L = 1|\text{Inf}, h_a]$ e $P[L = 1|\overline{\text{Inf}}, h_a]$, e então $P[L = 1|h_a]$ e $P[L = 0|h_a]$.

Resposta:

Se o ataque está ocorrendo, temos:

$$P[L = 1|\text{Inf}, h_a] = TPR = 0.8$$

e

$$P[L = 1|\overline{\text{Inf}}, h_a] = FPR = 0.1$$

Podemos calcular $P[L = 1|h_a]$ pela lei total da probabilidade:

$$P[L = 1|h_a] = P[L = 1|\text{Inf}, h_a] \cdot P[\text{Inf}] + P[L = 1|\overline{\text{Inf}}, h_a] \cdot P[\overline{\text{Inf}}]$$

Substituindo os valores:

$$P[L = 1|h_a] = 0.8 \cdot 0.2 + 0.1 \cdot 0.8 = 0.16 + 0.08 = 0.24$$

Agora, $P[L = 0|h_a]$ é simplesmente o complementar de $P[L = 1|h_a]$:

$$P[L = 0|h_a] = 1 - P[L = 1|h_a] = 1 - 0.24 = 0.76$$

2. Suponha que um ataque não esteja ocorrendo. Calcule $P[L = 1|\text{Inf}, h_b]$ e $P[L = 1|\overline{\text{Inf}}, h_b]$, e então $P[L = 1|h_b]$ e $P[L = 0|h_b]$.

Resposta:

Se um ataque não está ocorrendo, temos:

$$P[L = 1|\text{Inf}, h_b] = FPR = 0.1$$

e

$$P[L = 1|\overline{\text{Inf}}, h_b] = FPR = 0.1$$

Podemos calcular $P[L = 1|h_b]$ usando a lei total da probabilidade:

$$P[L = 1|h_b] = P[L = 1|\text{Inf}, h_b] \cdot P[\text{Inf}] + P[L = 1|\overline{\text{Inf}}, h_b] \cdot P[\overline{\text{Inf}}]$$

Substituindo os valores:

$$P[L = 1|h_b] = 0.1 \cdot 0.2 + 0.1 \cdot 0.8 = 0.02 + 0.08 = 0.10$$

Agora, $P[L = 0|h_b]$ é o complementar de $P[L = 1|h_b]$:

$$P[L = 0|h_b] = 1 - P[L = 1|h_b] = 1 - 0.10 = 0.90$$

3. Calcule $P[D|h_a]$ em função de V e M (Likelihood).

Resposta:

Sabemos que D representa os dados observados, isto é, V roteadores sinalizando um ataque de um total de M roteadores.

A probabilidade $P[D|h_a]$ pode ser modelada como uma distribuição binomial, onde V roteadores sinalizam um ataque dado que um ataque realmente está ocorrendo (h_a):

$$P[D|h_a] = \binom{M}{V} (P[L = 1|h_a])^V (P[L = 0|h_a])^{M-V}$$

Substituindo $P[L = 1|h_a] = 0.24$ e $P[L = 0|h_a] = 0.76$:

$$P[D|h_a] = \binom{M}{V} (0.24)^V (0.76)^{M-V}$$

4. Calcule $P[D|h_b]$ em função de V e M (Likelihood).

Resposta:

De forma similar à Pergunta 3, modelamos $P[D|h_b]$ como uma distribuição binomial. Agora, h_b indica que não há ataque acontecendo, e portanto utilizamos $P[L = 1|h_b]$ e $P[L = 0|h_b]$:

$$P[D|h_b] = \binom{M}{V} (P[L = 1|h_b])^V (P[L = 0|h_b])^{M-V}$$

Substituindo $P[L = 1|h_b] = 0.1$ e $P[L = 0|h_b] = 0.9$:

$$P[D|h_b] = \binom{M}{V} (0.1)^V (0.9)^{M-V}$$

5. Calcule $P[h_a|D]$ e $P[h_b|D]$ (Posterior).

Resposta:

Usamos o teorema de Bayes para calcular $P[h_a|D]$ e $P[h_b|D]$.

Para $P[h_a|D]$:

$$P[h_a|D] = \frac{P[D|h_a] \cdot P[h_a]}{P[D]}$$

A probabilidade total $P[D]$ é dada por:

$$P[D] = P[D|h_a] \cdot P[h_a] + P[D|h_b] \cdot P[h_b]$$

Substituímos os valores de $P[D|h_a]$, $P[D|h_b]$, $P[h_a] = 0.5$ e $P[h_b] = 0.5$ (hipóteses iguais):

$$P[h_a|D] = \frac{P[D|h_a] \cdot 0.5}{P[D|h_a] \cdot 0.5 + P[D|h_b] \cdot 0.5}$$

Similarmente, para $P[h_b|D]$:

$$P[h_b|D] = \frac{P[D|h_b] \cdot 0.5}{P[D|h_a] \cdot 0.5 + P[D|h_b] \cdot 0.5}$$

Substituindo os valores de $P[D|h_a]$ e $P[D|h_b]$ calculados anteriormente, podemos encontrar os valores de $P[h_a|D]$ e $P[h_b|D]$ em função de V e M :

$$P[h_a|D] = \frac{(0.24)^V (0.76)^{M-V}}{(0.24)^V (0.76)^{M-V} + (0.1)^V (0.9)^{M-V}}$$

$$P[h_b|D] = \frac{(0.1)^V (0.9)^{M-V}}{(0.24)^V (0.76)^{M-V} + (0.1)^V (0.9)^{M-V}}$$

6. Qual o mínimo de roteadores que deveriam alarmar (V) para que você tenha confiança de que um ataque ocorreu.

Resposta:

O número mínimo de roteadores V que devem alarmar para termos confiança de que um ataque está ocorrendo pode ser calculado comparando $P[h_a|D]$ e $P[h_b|D]$. Queremos encontrar V tal que:

$$P[h_a|D] > P[h_b|D]$$

Para garantir essa desigualdade, substituímos as expressões que encontramos para $P[h_a|D]$ e $P[h_b|D]$:

$$\frac{(0.24)^V (0.76)^{M-V}}{(0.24)^V (0.76)^{M-V} + (0.1)^V (0.9)^{M-V}} > \frac{(0.1)^V (0.9)^{M-V}}{(0.24)^V (0.76)^{M-V} + (0.1)^V (0.9)^{M-V}}$$

Como o denominador é o mesmo dos dois lados da desigualdade, temos:

$$(0.24)^V (0.76)^{M-V} > (0.1)^V (0.9)^{M-V}$$

Rearranjando os termos, obtemos:

$$\frac{(0.24)^V}{(0.1)^V} > \frac{(0.9)^{M-V}}{(0.76)^{M-V}}$$

Definindo $k_1 = \frac{0.24}{0.1} = 2.4$ e $k_2 = \frac{0.9}{0.76}$:

$$(2.4)^V > (k_2)^{M-V}$$

Tomando o logaritmo em ambos os lados, temos:

$$V \cdot \log(2.4) > (M - V) \cdot \log(k_2)$$

Distribuindo, isso se torna:

$$V \cdot \log(2.4) + V \cdot \log(k_2) > M \cdot \log(k_2)$$

Rearranjando:

$$V \cdot (\log(2.4) + \log(k_2)) > M \cdot \log(k_2)$$

Assim, isolando V :

$$V > \frac{M \cdot \log(k_2)}{\log(2.4) + \log(k_2)}$$

Finalmente, para $M = 200$:

$$V > 32.37$$

Portanto, pelo menos 33 roteadores devem alarmar para garantir que a probabilidade de um ataque esteja ocorrendo seja maior do que a probabilidade de não estar ocorrendo.

7. Caso $P[h_a] = 0.1$, os resultados variam? Trace as curvas $P[h_a|D]$ e $P[h_b|D]$ em função de V e explique as curvas.

Resposta:

Quando alteramos $P[h_a]$ para 0.1, os cálculos de $P[h_a|D]$ e $P[h_b|D]$ são impactados. Isso ocorre porque o valor anterior $P[h_a] = 0.5$ era uma suposição de hipóteses equiprováveis.

Com $P[h_a] = 0.1$, temos:

$$P[h_a|D] = \frac{P[D|h_a] \cdot 0.1}{P[D|h_a] \cdot 0.1 + P[D|h_b] \cdot 0.9}$$

e

$$P[h_b|D] = \frac{P[D|h_b] \cdot 0.9}{P[D|h_a] \cdot 0.1 + P[D|h_b] \cdot 0.9}$$

As curvas de $P[h_a|D]$ e $P[h_b|D]$ em função de V mudam com essa nova probabilidade a priori, sendo necessário recalcular os valores de $P[D|h_a]$ e $P[D|h_b]$ para diferentes V e observar a diferença. Como $P[h_a] = 0.1$ indica que um ataque é menos provável, $P[h_a|D]$ se torna menor do que no caso anterior para os mesmos valores de V .

8. Caso $P[h_a] = 0.1$, trace a curva $\log(P[h_a|D]/P[h_b|D])$ em função de V .

Resposta:

A curva $\log(P[h_a|D]/P[h_b|D])$ representa a relação logarítmica entre a probabilidade de um ataque estar ocorrendo em relação à probabilidade de não haver ataque, para diferentes valores de V .

Com $P[h_a] = 0.1$, a equação que devemos traçar é:

$$\log\left(\frac{P[h_a|D]}{P[h_b|D]}\right) = \log\left(\frac{P[D|h_a] \cdot 0.1}{P[D|h_b] \cdot 0.9}\right)$$

Substituindo as expressões para $P[D|h_a]$ e $P[D|h_b]$ em função de V , podemos traçar a curva dessa razão logarítmica. Conforme V aumenta, a razão $P[h_a|D]/P[h_b|D]$ também aumenta, indicando uma maior confiança de que um ataque está ocorrendo.

9. Para $TPR = 0.9$ e $FPR = 0.1$, plote, em um mesmo gráfico, a função de probabilidade de massa:

- (a) do número de roteadores que alarmam quando há um ataque;
- (b) do número de roteadores que alarmam quando não há um ataque.

Na implementação do classificador central (aquele que recebe os sinais dos roteadores domésticos, e que são os “sensores” em cada residência), você deve decidir a partir de quantos roteadores residenciais alarmando o classificador central deverá detectar que um ataque está ocorrendo.

- (a) Explique como avaliar o erro da sua decisão.

(b) Estime esse erro para o valor escolhido.

Resposta:

(a) Para o número de roteadores que alarmam quando há um ataque, usamos a distribuição binomial com $TPR = 0.9$. A função de probabilidade de massa (PMF) é dada por:

$$P(V|h_a) = \binom{M}{V} (0.9)^V (0.1)^{M-V}$$

(b) Para o número de roteadores que alarmam quando não há um ataque, usamos a distribuição binomial com $FPR = 0.1$. A função de probabilidade de massa (PMF) é dada por:

$$P(V|h_b) = \binom{M}{V} (0.1)^V (0.9)^{M-V}$$

A partir desses gráficos, podemos escolher um número de roteadores alarmando como o limiar para que o classificador central detecte um ataque.

(i) Explique como avaliar o erro da sua decisão.

O erro pode ser avaliado observando as taxas de falsos positivos e falsos negativos associadas ao limiar escolhido. Ao aumentar o limiar, reduzimos os falsos positivos, mas podemos aumentar os falsos negativos.

(ii) Estime esse erro para o valor escolhido.

Se o limiar for, por exemplo, $V = 15$, podemos calcular as probabilidades de erro com base nas funções de probabilidade de massa. Se muitos roteadores alarmam quando h_b é verdadeiro, teremos falsos positivos. Da mesma forma, se poucos roteadores alarmam quando h_a é verdadeiro, teremos falsos negativos.