2
Teoria da Informação

Neste capítulo apresentamos alguns conceitos básicos sobre Teoria da Informação que utilizaremos durante este trabalho.

2.1
Alfabeto, texto, letras e caracteres

Um alfabeto $\Sigma = (\sigma_1, ..., \sigma_n)$ é um conjunto finito composto de n símbolos distintos σ_i, com $i \in [1, n]$, denominados letras do alfabeto.

Um texto T é uma concatenação de m caracteres $t_1...t_m$, não necessariamente distintos, onde cada caractere é uma letra do alfabeto ($t_i \in \Sigma$, com $i \in [1, m]$).

Por exemplo, o texto 'SHANNON' possui 7 caracteres e 5 letras, com alfabeto $\Sigma = ('S', 'H', 'A', 'N', 'O')$.

2.2
Partição e cadeia

Uma cadeia $T[i..j]$ em T, é uma subsequência de elementos contíguos de T. Uma partição de T é uma divisão de T em cadeias não vazias.

2.3
Fonte de Informação

Uma fonte de informação discreta Γ é aquela que é capaz de emitir informação por meio de símbolos discretos pertencentes a um conjunto finito chamado alfabeto. Uma fonte emite mensagens na forma de textos: $\Gamma = (T_1, T_2, ...)$.

2.4
Entropia

Informação e incerteza descrevem qualquer processo que seleciona um ou mais objetos de um conjunto. Por exemplo, suponha que temos uma fonte que pode emitir três mensagens A, B ou C. A incerteza se caracteriza por não sabermos qual será o próximo símbolo que será produzido. Quando a mensagem

aparece, essa incerteza desaparece e podemos dizer que ganhamos uma certa informação. Assim, informação é uma diminuição da incerteza.

Entropia pode ser pensada como uma medida matemática de informação ou incerteza, calculada como uma função de uma distribuição de probabilidade.

Definição 2.1 *Dada uma variável aleatória* $X = (x_1, x_2, ..., x_n)$ *ocorrendo de acordo com uma distribuição de probabilidade* $p(X)$*, a* ***entropia*** *de* X*, definida por* $H(X)$ *é igual a*

$$H(X) = -\sum_{i=1}^{n} p(x_i).log_2 p(x_i)$$

Shannon (30) apresenta o conceito geral da entropia de uma fonte como sendo a quantidade média de informação contida nesta fonte. Para uma fonte $\Gamma = (T_1, T_2, ...)$, a entropia $H(\Gamma)$, é a média ponderada de todos os símbolos dessa fonte, com pesos iguais às suas probabilidades de ocorrência

$$H(\Gamma) = -\sum_{i=1}^{n} p(T_i).log_2 p(T_i)$$

O cálculo da entropia do texto é feita a partir das frequências de ocorrência f_i das letras no texto

$$H(T) = -\sum_{i=1}^{n} p(t_i).log_2 p(t_i), \text{ onde } p(t_i) = \frac{1}{f_i}$$

A parcela -$log_2 p(t_i)$ é chamada de quantidade de informação de t_i expressa em bits. Então, a Entropia é a soma da probabilidade de cada letra multiplicada pela sua respectiva quantidade de informação. Podemos entender a entropia como sendo a quantidade de informação média, ou seja, um limite inferior para o comprimento médio dos códigos relativos a cada instância gerada pela fonte.

Por exemplo, suponha que T é um texto que contém 3 letras σ_1, σ_2 e σ_3 com probabilidades de ocorrência no texto iguais a $\frac{1}{2}$, $\frac{1}{4}$ e $\frac{1}{4}$, respectivamente. Então o número médio de bits para codificar T é igual a

$$H(T) = -(\tfrac{1}{2}.log_2\tfrac{1}{2} + \tfrac{1}{4}.log_2\tfrac{1}{4} + \tfrac{1}{4}.log_2\tfrac{1}{4}) = 1,5 bits/letra$$

Algumas propriedades da entropia

Propriedade 1

$0 \leq H(\Gamma) \leq log_2 n$

Propriedade 2

$H(\Gamma) = 0 \iff p(T_i) = 1$ *para algum* $i, 1 \leq i \leq n$ *e* $p(T_j) = 0$ *para todo* $j, j \neq i, 1 \leq j \leq n$*;*

Propriedade 3

$H(\Gamma) = log_2 n \iff p(T_i)$ é constante e igual a $\frac{1}{n}$ para todo $i, 1 \leq i \leq n$.

Em um sistema criptográfico, podemos calcular a entropia dos vários componentes. Podemos pensar na chave como sendo uma variável aleatória K que assume valores de acordo com a distribuição de probabilidade $p(k)$ e, consequentemente, podemos calcular a entropia $H(K)$. Analogamente, podemos calcular as entropias $H(T)$ e $H(C)$ associadas com as distribuições de probabilidade do texto original e do texto cifrado.

Quanto maior a entropia das chaves $H(K)$ e dos textos cifrados $H(C)$, maior é o nível de incerteza para o criptoanalista, pois o procedimento de capturar um texto cifrado e, através de técnica de força bruta, tentar encontrar o texto original torna-se inviável.

2.5 Redundância

Seja H_L a entropia de uma linguagem natural L. Ela mede a informação média por letra na mensagem.

Como primeira aproximação para o cálculo de H_L, podemos usar um idioma com todas as 26 letras equiprováveis. De acordo com a propriedade 3, a entropia nesse caso será igual a $log 26 \approx 4.76$. Para a língua inglesa, podemos usar a distribuição de probabilidades da tabela 2.1 retirada de (12) para um texto típico, e assim teremos $H_L \approx 4,17$.

Uma segunda aproximação para o cálculo de H_L é calcular a entropia da distribuição de probabilidade de todos os digramas e depois dividir por 2.

De forma geral, considere que Y^n é uma variável aleatória que tem em sua distribuição de probabilidades todos os n-gramas do texto, assim como Y corresponde às letras, Y^2 correspondente aos digramas e Y^3 aos trigramas.

Definição 2.2 *Podemos então definir a* ***entropia de*** *L como*

$$H_L = \lim_{n\to\infty} \frac{H(Y^n)}{n}$$

Para a língua inglesa o valor teórico para H_L é

$$1,0 \leq H_L \leq 1,5$$

Definição 2.3 *A redundância de um código C é a diferença entre o comprimento médio das palavras do código e a sua entropia*

$$R_C = \sum_{i=1}^{n} p(c_i).l_i - H(C)$$

Tabela 2.1: Distribuição de probabilidades típica (língua inglesa).

Letra	Probabilidade (%)
A	8,04
B	1,54
C	3,06
D	3,99
E	12,51
F	2,30
G	1,96
H	5,49
I	7,26
J	0,16
K	0,67
L	4,14
M	2,53
N	7,09
O	7,60
P	2,00
Q	0,11
R	6,12
S	6,54
T	9,25
U	2,71
V	0,99
W	1,92
X	0,19
Y	1,73
Z	0,09

Em relação à primeira aproximação, a **redundância de** L é definida por

$$R_L = log_2|Y| - H_L$$

Considerando que o tamanho do alfabeto é $|Y| = 26$ e que um valor médio para $H(L) \approx 1,3$, temos que a redundância de L é igual a $R_L = 4,7 - 1,3 = 3,4$ bits por letra.

2.6 Códigos

Um código binário para um alfabeto Σ pode ser descrito como uma função que associa uma cadeia c_i de bits a cada letra σ_i do alfabeto. Cada uma dessas cadeias de bits constitui uma palavra de código. Podemos utilizar a notação $c_i = C(\sigma_i)$.

Tabela 2.2: Exemplo de códigos binários.

Letra	Código 1	Código 2	Código 3	Código 4	Código 5
A	00	0	00	0	0
B	01	01	10	10	10
C	10	11	11	110	110
D	11	010	110	111	1110

Por exemplo, a tabela 2.2 mostra alguns códigos binários possíveis para um mesmo alfabeto.

O comprimento $|c|$ de uma palavra de código c é dado pela sua quantidade de bits.

Definição 2.4 *Um código é dito de* ***tamanho variável*** *quando contém palavras de código de comprimento distinto e de* ***tamanho fixo*** *quando todas as palavras de código têm o mesmo tamanho.*

Assim, os códigos 2, 3, 4 e 5 da tabela 2.2 são de tamanho variável enquanto que o código 1 é de tamanho fixo.

Definição 2.5 *Um código é dito* ***distinto*** *se suas palavras de código são todas distintas.*

Todos os códigos da tabela 2.2 são distintos.

Definição 2.6 *Uma codificação é dita* ***ambígua*** *se $c(T) = c_1...c_n = d_1...d_m$ e existe i, $1 \leq i \leq min(m, n)$, tal que $c_i \neq d_i$. Ou seja, existem duas partições diferentes de $c(T)$ que levam a codificações válidas para um mesmo código.*

Definição 2.7 *Um código é dito* ***unicamente decifrável*** *se ele não admite codificações ambíguas. Portanto, uma cadeia de bits concatenada de acordo com um código unicamente decifrável permite uma decodificação sem ambigüidade.*

Na tabela 2.2, apenas o código 2 não é unicamente decifrável pois permite diferentes partições, gerando ambigüidade. Por exemplo, considere o texto codificado 010010, onde foi utilizado o código 2. Observe que a partição $010 - 010$ resulta no texto original DD, mas se a partição for $01 - 0 - 010$, o texto original é BAD.

2.6.1 Códigos de prefixo

Definição 2.8 *Um código é dito* ***livre de prefixo*** *ou, simplesmente,* ***de prefixo****, quando, considerando todas as palavras do código, nenhuma palavra é prefixo de outra.*

Como conseqüência dessa definição, temos que todo código distinto de tamanho fixo é código livre de prefixo. Na tabela 2.2, os códigos 1, 4 e 5 são livres de prefixo, ao contrário do código 2, onde a palavra de código 0, associada a A, é prefixo da palavra de código 01 associada a B; e do código 3, onde a palavra de código 11, associada a C, é prefixo da palavra de código 110 associada a D.

Todo código de prefixo é unicamente decifrável, portanto, os códigos de prefixo são muito úteis pois, além de permitir que se decodifique uma cadeia de bits sem ambigüidade, a sua partição pode ser feita da esquerda para a direita de tal forma que o bit onde se inicia um novo elemento da partição é determinado sem que seja necessário se olhar para os bits à direita do mesmo. Seja, por exemplo, o seguinte código.

$$C = 00, 01, 10, 110, 11100, 11101, 11110, 11111$$

Pode ser visto que nenhuma palavra de código é prefixo de qualquer outra. Logo C é código livre de prefixo. Além disso, por exemplo, a cadeia 011101110001 só pode ser particionada da seguinte forma: 01-110-11100-01. Isto significa que:

- não há ambigüidade;
- a partição é feita da esquerda para a direita e o lugar de cada ponto é definido sem que se necessite considerar os bits a sua direita.

2.6.2 Comprimento médio do código

Definição 2.9 *Dado o conjunto de frequências* $(f_1, f_2, ..., f_n)$ *e o vetor* $(l_1, l_2, ..., l_n)$*, onde* f_i *é a frequência da i-ésima letra, e* l_i *é o tamanho do código para essa letra, definimos o* ***comprimento médio do código*** *como*

$$\bar{l} = \sum_{i=1}^{n} f_i.l_i$$

No código de prefixo os tamanhos dos códigos satisfazem uma certa desigualdade conhecida como desigualdade de Kraft-McMillan.

Definição 2.10 ***Desigualdade de Kraft-McMillan***

$$\sum_{i=1}^{n} 2^{-l_i} \leq 1$$

Esta propriedade também é válida no sentido oposto, ou seja, se os tamanhos dos códigos satisfazem a desigualdade de Kraft-McMillan então um código de prefixo pode ser construído com esses tamanhos de código.

Definição 2.11 *Um código de prefixo que minimiza o comprimento médio do código é chamado de* ***código de prefixo ótimo****.*

Propriedade 4 *Se H é a entropia de uma fonte discreta então o comprimento médio de um código ótimo de prefixo para essa fonte é limitado por*

$$H \leq \bar{l} < H + 1$$

No próximo capítulo mostramos que os códigos de Huffman são códigos de prefixo ótimos.

2.7 Distância de Unicidade

Uma característica importante de um sistema criptográfico é a medida do número de caracteres de um texto cifrado necessários para determinar uma solução única para a criptoanálise da mensagem.

Shannon definiu distância de unicidade como o valor aproximado do tamanho do criptograma (texto cifrado) a partir do qual somando a entropia do texto em claro associado mais a entropia da chave obtêm-se o número de bits do criptograma.

Shannon mostrou também que criptogramas maiores do que esta distância têm uma probabilidade elevada de ter uma única decifragem legível, enquanto que criptogramas significativamente menores do que a distância de unicidade têm, provavelmente, várias decifragens igualmente válidas. Ou seja, a distância de unicidade corresponde à quantidade de criptograma necessário para determinar unicamente uma chave possível. À medida que aumenta o comprimento do criptograma diminui a distância de unicidade.

Definição 2.12 *A* ***distância de unicidade*** *é definida como uma razão entre a entropia da chave e a redundância da linguagem:*

$$n_0 \approx \frac{H_K}{RL}$$

A distância de unicidade é inversamente proporcional à redundância da linguagem e quando esta tende para zero, mesmo uma cifragem trivial pode tornar-se inquebrável.

Porém, a distância de unicidade é uma medida estatística e, portanto, não fornece resultados determinísticos, mas sim probabilísticos. É apenas uma estimativa da quantidade de criptograma necessária para que haja uma única solução razoável para a criptoanálise. Quando é pequena indica que o algoritmo é pouco seguro, porém não garante segurança mesmo com um valor elevado.

Além disso, em muitos algoritmos, o cálculo da distância de unicidade é muito difícil. Entretanto, constatamos que em cifras inquebráveis ou ideais a distância de unicidade é infinita.