Fontes de Informação

Teoria da Informação - AULA 06

Profª. Verusca Severo

Universidade de Pernambuco
Escola Politécnica de Pernambuco

02 de julho de 2021

Introdução

- Para o estudo de teoria da informação é imprescindível que se tenha uma descrição matemática do mecanismo de geração de informação.

- Considerando que a informação é gerada por uma fonte de informação, o passo seguinte é caracterizar matematicamente fontes de informação.

Fonte de informação discreta

- Este modelo de fonte de informação é completamente descrito pelo alfabeto U da fonte e pela distribuição de probabilidade de ocorrência dos símbolos.

- Uma fonte discreta emite sequências de símbolos de um alfabeto fixo $U = \{u_1, u_2, \ldots, u_K\}$, sendo tais símbolos gerados obedecendo a uma distribuição de probabilidade $P(u_1), P(u_2), \ldots, P(u_K)$.

Fonte discreta sem memória

- Os símbolos emitidos por uma **fonte discreta sem memória** U são estatisticamente independentes.

- Ou seja, não há correlação entre os símbolos emitidos pela fonte (não se pode falar de entropia condicional). Logo, a probabilidade conjunta

$$P(u_1, u_2, \ldots, u_K) = P(u_1)P(u_2)\ldots P(u_K).$$

- A entropia (ou incerteza média) dos símbolos de uma fonte discreta sem memória U é dada por:

$$H(U) = \sum_{i=1}^{K} P(u_i) \log_2 \left(\frac{1}{P(u_i)} \right)$$

Fonte discreta sem memória

1. **Exemplo 1:** Seja U uma fonte discreta sem memória com alfabeto $U = \{u_1, u_2\}$, em que tais símbolos são gerados obedecendo a uma distribuição de probabilidade $P(u_1) = \frac{1}{3}$ e $P(u_2) = \frac{2}{3}$. Calcule a entropia da fonte U.

$$H(U) = \sum_{i=1}^{2} P(u_i) \log_2 \left(\frac{1}{P(u_i)} \right)$$

$$H(U) = \frac{1}{3} \log_2 \left(\frac{1}{\frac{1}{3}} \right) + \frac{2}{3} \log_2 \left(\frac{1}{\frac{2}{3}} \right)$$

$$H(U) = 0,918 \; bit/\text{simbolo}$$

Fonte discreta sem memória - Extensão

- Em muitas situações práticas, como por exemplo na cifragem ou na codificação de dados, há interesse em trabalhar-se com blocos de símbolos de uma fonte, em vez de trabalhar-se com símbolos individuais.

- Consideremos blocos de tamanho fixo igual a n símbolos. O número n é chamado de ordem da extensão da fonte.

- Matematicamente este caso é tratado considerando uma nova fonte U^n, cujo alfabeto de saída consiste de todos os blocos distintos formados pela concatenação de n símbolos da fonte U.

Fonte discreta sem memória - Extensão

- O alfabeto da nova fonte U^n consiste de todos os blocos distintos formados pela concatenação de n símbolos da fonte $U = \{u_1, u_2, \ldots, u_K\}$. O número de blocos distintos é:

$$\underset{1}{\underline{\;K\;}}\quad \underset{2}{\underline{\;K\;}}\quad \underset{3}{\underline{\;K\;}}\quad \bullet\bullet\bullet\quad \underset{n}{\underline{\;K\;}} = K^n$$

$$\begin{aligned}
\text{Bloco } 1 &\Rightarrow \sigma_1 &&= u_1 u_1 u_1 \ldots u_1 \\
\text{Bloco } 2 &\Rightarrow \sigma_2 &&= u_1 u_1 u_1 \ldots u_2 \\
&\vdots && \quad\vdots \\
\text{Bloco } K^n &\Rightarrow \sigma_{K^n} &&= u_n u_n u_n \ldots u_n
\end{aligned}$$

- Logo, o alfabeto de U^n é $\{\sigma_1, \sigma_2, \ldots, \sigma_{K^n}\}$.

Extensão de fonte discreta sem memória

- **Definição:** Seja U uma fonte discreta sem memória, cujo alfabeto consiste dos símbolos $\{u_1, u_2, \ldots, u_K\}$ e a probabilidade da fonte emitir o símbolo u_i é $P(u_i)$. A extensão de ordem n de U, denotada por U^n, é a fonte discreta sem memória com K^n símbolos $\{\sigma_1, \sigma_2, \ldots, \sigma_{K^n}\}$, na qual cada σ_i corresponde a uma sequencia específica de n dos símbolos de U. Supondo que σ_i corresponde a sequência $u_{i1}, u_{i2}, \ldots, u_{in}$, a probabilidade $P(\sigma_i)$ do símbolo σ_i é dada por

$$P(\sigma_i) = P(u_{i1}, u_{i2}, \ldots, u_{in}) = P(u_{i1})P(u_{i2})\ldots P(u_{in})$$

2. **Exemplo 2:** Seja U uma fonte discreta sem memória com alfabeto $U = \{u_1, u_2\}$. A extensão de terceira ordem de U, denotada por U^3, possui o seguinte alfabeto:

$$
\begin{aligned}
\sigma_1 &= \quad u_1 u_1 u_1 \\
\sigma_2 &= \quad u_1 u_1 u_2 \\
\sigma_3 &= \quad u_1 u_2 u_1 \\
\sigma_4 &= \quad u_1 u_2 u_2 \\
\sigma_5 &= \quad u_2 u_1 u_1 \\
\sigma_6 &= \quad u_2 u_1 u_2 \\
\sigma_7 &= \quad u_2 u_2 u_1 \\
\sigma_8 &= \quad u_2 u_2 u_2
\end{aligned}
$$

Extensão de fonte discreta sem memória

3. **Exemplo 3:** Calcule a entropia da fonte U^3, em que $P(u_1) = \frac{1}{3}$ e $P(u_2) = \frac{2}{3}$

$$H(U^3) = \sum_{j=1}^{K^3} P(\sigma_j) \log_2 \left(\frac{1}{P(\sigma_j)} \right) = \sum_{j=1}^{8} P(\sigma_j) \log_2 \left(\frac{1}{P(\sigma_j)} \right),$$

A distribuição de probabilidade do alfabeto de U^3 é:

$$\begin{aligned}
P(\sigma_1) &= P(u_1, u_1, u_1) = P(u_1)P(u_1)P(u_1) = 1/27 \\
P(\sigma_2) &= P(u_1, u_1, u_2) = P(u_1)P(u_1)P(u_2) = 2/27 \\
P(\sigma_3) &= P(u_1, u_2, u_1) = P(u_1)P(u_2)P(u_1) = 2/27 \\
P(\sigma_4) &= P(u_1, u_2, u_2) = P(u_1)P(u_2)P(u_2) = 4/27 \\
P(\sigma_5) &= P(u_2, u_1, u_1) = P(u_2)P(u_1)P(u_1) = 2/27 \\
P(\sigma_6) &= P(u_2, u_1, u_2) = P(u_2)P(u_1)P(u_2) = 4/27 \\
P(\sigma_7) &= P(u_2, u_2, u_1) = P(u_2)P(u_2)P(u_1) = 4/27 \\
P(\sigma_8) &= P(u_2, u_2, u_2) = P(u_2)P(u_2)P(u_2) = 8/27
\end{aligned}$$

3. **Exemplo 3:** Calcule a entropia da fonte U^3, em que $P(u_1) = \frac{1}{3}$ e $P(u_2) = \frac{2}{3}$

   Logo, temos que:

$$H(U^3) = \sum_{j=1}^{8} P(\sigma_j) \log_2 \left(\frac{1}{P(\sigma_j)} \right) = 2,754$$

4. **Exemplo 4:** Determine a extensão de ordem $n = 8$ da fonte U (Exemplo 2) e calcule sua entropia.

- **SOLUÇÃO:** Temos que a extensão de oitava ordem de U, denotada por U^8, possui um alfabeto com $2^8 = 256$ símbolos, em que:

$$
\begin{aligned}
\sigma_1 &= u_1u_1u_1u_1u_1u_1u_1u_1 \\
\sigma_2 &= u_1u_1u_1u_1u_1u_1u_1u_2 \\
\sigma_3 &= u_1u_1u_1u_1u_1u_1u_2u_1 \\
&\vdots \\
\sigma_{256} &= u_2u_2u_2u_2u_2u_2u_2u_2
\end{aligned}
$$

com:

$$
\begin{aligned}
P(\sigma_1) &= P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1) = (1/3)^8 \\
P(\sigma_2) &= P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_2) = (1/3)^7.(2/3) \\
P(\sigma_3) &= P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_1).P(u_2).P(u_1) = (1/3)^7.(2/3) \\
&\vdots \\
P(\sigma_{256}) &= P(u_2).P(u_2).P(u_2).P(u_2).P(u_2).P(u_2).P(u_2).P(u_2) = (2/3)^8
\end{aligned}
$$

- **SOLUÇÃO:** Calculando a entropia, temos que:

$$H(U^8) = \sum_{i=1}^{256} P(\sigma_i) log_2 \left[\frac{1}{P(\sigma_i)} \right]$$

$$H(U^8) = 7,344 \ bits$$

Extensão de fonte discreta sem memória

- No **Exemplo 4**, a entropia da fonte U^8 é igual a $7,344$, que corresponde a uma extensão de ordem 8 da fonte U.

- No **Exemplo 3**, a entropia da fonte U^3 é igual a $2,754$, que corresponde a uma extensão de ordem 3 da fonte U.

- No **Exemplo 1**, a entropia da fonte U é igual a $0,918$

- Perceba que:

$$0,918 \times 8 = 7,344 \quad \text{e que} \quad 0,918 \times 3 = 2,754$$

- Ou seja, $H(U^8) = 8 \times H(U)$ e $H(U^3) = 3 \times H(U)$.

Será que isso é **sempre** verdade?

Extensão de fonte discreta sem memória - ENTROPIA

- Como cada símbolo de U^n corresponde a n símbolos de U, seria razoável esperar que a entropia por símbolo de U^n, denotada por $H(U^n)$, fosse igual a n vezes a entropia por símbolo de U. Como é mostrado a seguir, este resultado de fato é verdadeiro.

- **Entropia da extensão de uma fonte discreta sem memória (Teorema):** A entropia da extensão de ordem n de uma fonte discreta sem memória, denotada por U^n, apresenta entropia

$$H(U^n) = n \times H(U)$$

  - **Prova:** Ver notas de aula!