

Bazele Tehnologiei Informației

Curs 1

Facultatea de Cibernetică, Statistică și Informatică Economică
(CSIE) – ASE București

Răzvan Daniel ZOTA

zota@ase.ro

<https://zota.ase.ro/bti>

Elemente de teoria (transmisiei) informației

TI este știința care se ocupă cu studiul legilor referitoare la obținerea, transmiterea, prelucrarea și păstrarea informației.

Datorită caracterului aleator al fenomenelor care constituie obiectul de studiu al teoriei informației, aceasta (TI) poate fi considerată ca o ramură a teoriei probabilităților.

Pentru transmiterea informației (oricare ar fi aceasta) este necesară codificarea ei, adică transmiterea informației într-un limbaj constituit din simboluri și semnale (semnalele putând fi oscilații sonore sau luminoase, impulsuri electrice, etc.).

O problemă de bază în TI este aceea a găsirii metodelor celor mai economice de codificare, adică a găsirii probabilităților de transmitere a informației cu ajutorul unui volum minim de simboluri.

Elemente de teoria transmisiei informației

O altă problemă tipică pentru TI este aceea a determinării numărului și capacității optime a canalelor de transmitere a informației a.î. să nu aibă loc pierderi sau denaturări pe drumul de la sursa de informații la receptorul de informație.

Entropia - măsură a gradului de nedeterminare a stării sistemelor fizice

Orice informație constă în totalitatea datelor cu privire la un sistem fizic oarecare. Astfel, informația nu ar avea sens dacă starea sistemului fizic ar fi complet determinată.

Elemente de teoria transmisiei informației

Interesante sunt cazurile în care sistemul fizic considerat se poate afla întâmplător într-o anumită stare dintr-o mulțime de stări posibile, adică în situațiile în care sistemul comportă un anumit grad de incertitudine.

De asemenea, informațiile obținute cu privire la un sistem fizic sunt cu atât mai complete și cu atât mai consistente, cu cât nedeterminarea sistemului este mai mare până la obținerea acestor informații. De aici vine logic întrebarea: ce trebuie să înțelegem printr-un grad mare sau mic de nedeterminare și prin ce trebuie măsurat gradul de nedeterminare al unui sistem fizic dat?

Elemente de teoria transmisiei informației

Exemple simple:

- 1) Sistem fizic: **moneda** (ca rezultat al aruncării poate fi într-una din cele două stări: stemă sau marcă)
- 2) Sistem fizic: **zarul** (ca rezultat al unei aruncări se poate afla într-una din cele șase stări posibile)

? În care dintre cele două cazuri avem un grad de nedeterminare mai mare?

Elemente de teoria transmisiei informației

Să considerăm acum următorul exemplu:

- 1) Sistem fizic: **moneda** (ca rezultat al aruncării poate fi într-una din cele două stări: stemă sau marcă)
- 2) Sistem fizic: **un computer** (în stare de funcționare cu probabilitatea de 99% și în stare defectă cu probabilitatea de 1%)

? **Doar numărul stărilor caracterizează gradul de nedeterminare al unui sistem fizic?**

Elemente de teoria transmisiei informației

1. Entropia informațională

Informația = mesaj care aduce o precizare într-o problemă cu un anumit grad de incertitudine.

Incetitudinea (nedeterminarea) **scade** o data cu apariția (creșterea) informației

Fie sistemul fizic (experimentul) X , cu repartiția probabilistică:

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}$$

Stările sistemului sunt x_i , iar probabilitățile de realizare p_i
Sistemul de evenimente considerat este un **sistem complet**.

Formula lui Shannon

<https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf>

$$H(p_1, p_2, \dots, p_n) = - \sum_{i=1}^n p_i \log_2 p_i$$

H se numește entropie informațională. Logaritmi în bază doi sunt negativi având în vedere valorile posibile pentru p_i , deci se justifică luarea acestei sume cu semn schimbat.

Conform acestei definiții, putem spune că unitatea de măsură a informației este bitul. Astfel, un bit (**B**inary **digIT**) este definit ca fiind cantitatea de informație obținută prin precizarea unei variante din două egal probabile.

Exemplu de codificare

Ce înseamnă faptul că un semnal codificat are 1,75 biți/simbol?

= putem converti semnalul original într-un șir de 1 și 0 astfel încât **media** este 1,75 cifre binare pentru fiecare simbol din semnalul original.

Pp. că avem 4 simboluri: A, B, C, D cu probabilitățile:

$$P_A=1/2; P_B=1/4; P_C=1/8; P_D=1/8$$

$$-\log_2 P_A = 1 \text{ bit}, -\log_2 P_B = 2 \text{ biți}, -\log_2 P_C = 3 \text{ biți}, -\log_2 P_D = 3 \text{ biți}$$

Conform formulei lui Shannon, nedeterminarea este:

$$H = \frac{1}{2} \bullet 1 + \frac{1}{4} \bullet 2 + \frac{1}{8} \bullet 3 + \frac{1}{8} \bullet 3 = 1,75 \text{ biti}$$

Exemplu de codificare (cont.)

Dacă folosim reprezentarea binară pentru simbolurile A,B,C,D:

A = 1; B = 01; C = 000; D = 001, atunci șirul ABADCAAB va fi codificat astfel:

10110010001101

(14 cifre binare utilizate pentru codificarea celor 8 simboluri => media este $14/8 = 1,75$)

Obs. Ce se întâmplă dacă folosim o codificare de genul:

A = 00; B = 01; C = 10; D = 11 ?

Principalele proprietăți ale entropiei informaționale

$$P1. H(p_1, p_2, \dots, p_n) \geq 0$$

$$P2. H(p_1, p_2, \dots, p_n) = 0 \Leftrightarrow \exists k \in \{1, 2, \dots, n\} \text{ a.i. } p_k = 1$$

$$P3. H(p_1, p_2, \dots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)$$

Principalele proprietăți ale entropiei informaționale (cont.)

$$P4. H(p_1, p_2, \dots, p_n, 0) = H(p_1, p_2, \dots, p_n)$$

$$P5. H(X_1 \times X_2 \times \dots \times X_n) = H(X_1) + H(X_2) + \dots + H(X_n)$$

$$P6. H(X \times Y) = H(X) + H(Y / X), \forall X, Y$$

$$\text{unde: } H(Y / X) = \sum_{k=1}^n p(x_k) \cdot H(Y / x_k)$$

$$H(Y / x_k) = - \sum_{i=1}^m p(y_i / x_k) \log_2 p(y_i / x_k)$$

Bazele aritmetice ale calculatoarelor

Sistem de numerație = totalitatea regulilor de reprezentare a numerelor cu ajutorul unor simboluri (*cifre*). (a se vedea și <https://web.ma.utexas.edu/users/mks/326K04/what.html>)

Numărul de simboluri permise s.n. baza (rădăcina) sistemului de numerație

Sisteme de numerație

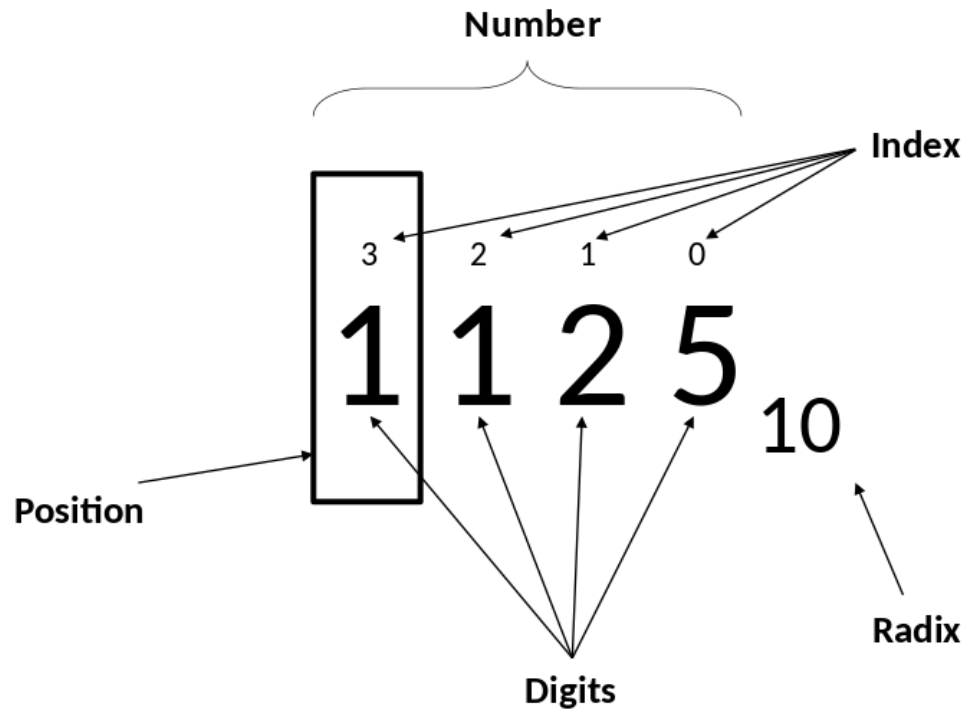
- poziționale
- nepoziționale

Sistemul roman

I	X	C	M	V	L	D
1	10	100	1000	5	50	500

Bazele aritmetice ale calculatoarelor

Sisteme de numerație poziționale



Bazele aritmetice ale calculatoarelor

Atunci când un simbol cu o valoare mai mică este *poziționat după* un simbol cu o valoare mai mare, valorile simbolurilor *se adună*.

III	→	$1 + 1 + 1$	=	3
VIII	→	$5 + 1 + 1 + 1$	=	8
XVIII	→	$10 + 5 + 1 + 1 + 1$	=	18
LXXII	→	$50 + 10 + 10 + 1 + 1$	=	72
CI	→	$100 + 1$	=	101
MMIVII	→	$1000 + 1000 + 5 + 1 + 1$	=	2007
MDC	→	$1000 + 500 + 100$	=	1600

Bazele aritmetice ale calculatoarelor

Atunci când un simbol cu o valoare mai mică este *poziționat înainte* de un simbol cu o valoare mai mare, valoarea mai mică *se scade* din cealaltă.

IV

→ 5 - 1

=

4

XIX

→ 10 + (10 - 1)

=

19

Reprezentarea într-o bază oarecare a unui număr întreg

- ◆ Reprezentarea unui număr întreg

$$N = \overline{a_n a_{n-1} \cdots a_0}$$

$$N = a_n \bullet b^n + a_{n-1} \bullet b^{n-1} + \cdots + a_0 \bullet b^0$$

Este reprezentarea numărului întreg N în baza b . Cifrele numărului N au următoarea proprietate:

$$0 \leq a_i \leq b-1, \forall i \in \overline{0, n}$$

Reprezentarea într-o bază oarecare a unui număr real

- ◆ Reprezentarea unui număr real R

$$R = \overline{a_n a_{n-1} \cdots a_0 a_{-1} \cdots a_{-m}}$$

$$R = a_n \bullet b^n + a_{n-1} \bullet b^{n-1} + \cdots + a_0 \bullet b^0 + a_{-1} \bullet b^{-1} + \cdots + a_{-m} \bullet b^{-m}$$

Este reprezentarea numărului real R în baza b . Cifrele numărului R au următoarea proprietate:

$$0 \leq a_i \leq b-1, \forall i \in \overline{-m, n}$$

nu există k astfel încât $a_k = a_{k-1} = a_{k-2} = \cdots = b-1$

Conversia bazei de numeratie

- ◆ Conversia bazei de numerație (partea întreagă, partea fracționară)
- ◆ Conversia rapidă între baze între care există relația:

$$b_1 = b_2^p, \quad p \in \mathbb{N}$$