



UNIUNEA EUROPEANĂ



GUVERNUL ROMÂNIEI



Instrumente Structurale
2007-2013



Platformă de e-learning și curriculum e-content pentru învățământul superior tehnic

Transmisia datelor multimedia in rețele de calculatoare

3. Modele

Modele

- Un model bun pentru date conduce la algoritmi de compresie eficienți
- Pentru a dezvolta algoritmi care efectuează operații matematice asupra datelor, acestea trebuie modelate matematic
- **Modele fizice**
 - Dacă se cunoaște ceva despre mecanismul de generare a datelor, se poate folosi această informație pentru construirea modelului
 - De exemplu, în aplicațiile referitoare la vorbire, cunoașterea mecanismului de producere a vorbirii poate fi folosit la construirea unui model matematic pentru procesul vorbirii eșantionate.

Modele

- **Modele probabilistice**

- Cel mai simplu model statistic pentru sursă este de a presupune că fiecare mesaj furnizat de sursă este independent de celelalte și fiecare se produce cu aceeași probabilitate
 - Acesta este numit *model de ignoranță* și ar putea fi util când nu se cunoaște nimic despre sursă
- Următorul pas în creșterea complexității modelului este de a păstra presupunerea asupra independenței, dar de a înlătura presupunerea de probabilitate egală pentru mesaje
 - În acest caz se alocă fiecărui mesaj o probabilitate în funcție de frecvența de furnizare a mesajului respectiv.

Modele

- **Modele probabilistice (cont)**

- Pentru o sursă care generează mesaje dintr-un alfabet

$$S = \{s_1, \dots, s_M\}$$

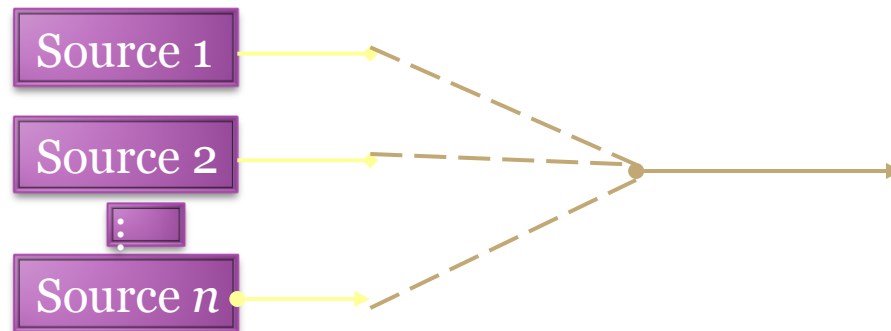
se poate folosi *modelul de probabilitate*

$$P = \{p(s_1), p(s_2), \dots, p(s_M)\}$$

- Cu acest model se poate calcula entropia sursei și pot fi construite coduri eficiente pentru reprezentarea mesajelor din S , adică se poate folosi un număr mediu minim de biți pentru fiecare mesaj
- Dacă se renunță la presupunerea de independență, se pune problema găsirii unui mod de a descrie dependența datelor, așa cum este în cazul modelelor Markov

Modele

- **Model de surse compuse**
- În multe aplicații nu este potrivit a se folosi un singur model pentru a descrie sursa, caz în care se poate defini o *sursă compusă*, care poate fi văzută ca o combinație de diverse surse, din care una este activă la un moment dat
- Fiecare din acestea este descrisă de un model propriu



Modele

- **Model Markov**

- Unul dintre cele mai răspândite moduri de reprezentare a dependenței între date este folosirea modelului Markov
- Pentru modelarea datelor, în compresia fără pierderi se folosește un model particular numit *lanț Markov discret*
- Fie $\{x_n\}$ secvența observată. Aceasta secvență se zice că este un model Markov de ordin k , dacă:

$$P(x_n | x_{n-1}, \dots, x_{n-k}) = P(x_n | x_{n-1}, \dots, x_{n-k}, \dots)$$

adică simbolul x_n depinde numai de ultimele k mesaje x_{n-1}, \dots, x_{n-k}

Modele

- Multimile $\{x_{i-1}, \dots, x_{i-k}\}$, $i=n, n-1, \dots$, reprezinta *starile procesului*
- Dacă mărimea alfabetului sursei este l , *numărul de stări* este l^k
- Cel mai des folosit model Markov este cel de ordinul 1, pentru care:

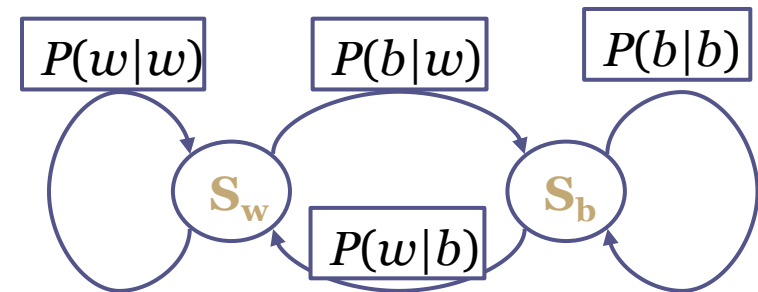
$$P(x_n | x_{n-1}) = P(x_n | x_{n-1}, \dots, x_{n-k}, \dots)$$

Modele

- Folosirea modelului Markov nu necesită prezumția de liniaritate
- De exemplu, fie o imagine binară, care are numai două feluri de pixeli - albi și negri
- Se știe că apariția unui pixel alb la observarea următoare depinde în oarecare măsură dacă pixelul curent este alb sau negru
- Prin urmare, se poate modela succesiunea de pixeli cu un lanț Markov

Modele

- Se defines doua stari: S_b & S_w pentru valoarea pixelului curent
- Se defines probabilitatile:
 - $P(S_b)$ = prob. ca pixelul sa fie negru
 - $P(S_w)$ = prob. Ca pixelul sa fie alb
- Probabilitatile de tranzitie sunt:
 - $P(b|b)$, $P(b|w)$
 - $P(w|b)$, $P(w|w)$



$$H(S_w) = -P(b/w)\log(b/w) - P(w/w)\log(w/w)$$

$$H(S_b) = -P(w/b)\log(b/w) - P(b/b)\log(b/b)$$

$$P(w/w) = 1 - P(b/w), \quad P(b/b) = 1 - P(w/b)$$

$$H = P(S_b)H(S_b) + P(S_w)H(S_w)$$

Exemplu

Fie

$$P(S_w) = 30/31 \quad P(S_b) = 1/31$$

$$P(w/w) = 0.99 \quad P(b/w) = 0.01 \quad P(b/b) = 0.7 \quad P(w/b) = 0.3$$

Pentru modelul *iid* (Independently and Identically Distributed):

$$H_{iid} = -0.8 \log 0.8 - 0.2 \log 0.2 = 0.206$$

Pentru modelul Markov:

$$H(S_b) = -0.3 \log 0.3 - 0.7 \log 0.7 = 0.881$$

$$H(S_w) = -0.01 \log 0.01 - 0.99 \log 0.99 = 0.081$$

$$H_{Markov} = \frac{30}{31} 0.081 + \frac{1}{31} 0.881 = 0.107$$

Folosirea modelelor Markov in compresia textelor

- Modelele Markov sunt utile în compresia textelor, deoarece probabilitatea de apariție a unei litere este influențată de precedentele
- Fie cuvântul “*ornitoring*”:
 - Se presupune că s-a procesat deja *ornitorin* și urmează a se coda următoarea literă
 - Dacă nu se ține seama de context și se tratează litera ca o surpriză, probabilitatea să apară litera *g* este relativ scăzută
 - Dacă se folosește un model Markov de ordinul I, probabilitatea să urmeze *g* crește considerabil
 - Cu creșterea ordinului modelului Markov (de la “n” la “in”, la “rin” ș.a.m.d.) probabilitatea literei *g* devine mai mare, ceea ce conduce la o entropie scăzută.