CZU 681.3

COMPRESII DE DATE PRIN TRANSFORMĂRI ORTOGONALE

Ţîcău Vitalie

În lucrare sunt prezentate metode de compresii de date prin transformări ortogonale, inclusiv transformata Karhunen-Loeve. La aplicarea transformării Karhunen-Loeve perderile de informaţii sunt minime.

В работе представлены методы сжатия данных при ортогональных преобразований, включая преобразование Карунен-Леве. При приминение преоброзования Карунен-Леве потери информаций являются минимальными.

In this paper are presented methods of data compression through orthogonal transformation, inclusively the Karhunen-Loeve's transformation.

Introducere

Această lucrare este orientată spre dezvoltarea metodelor de compresie numerică.

A comprima un mesaj înseamnă a păstra numai acei parametri care sunt esenţiali pentru destinatar, ceilalţi parametrii nu se transmit, respectiv nu se stochează. Sistemele de compresie pot fi grupate în 2 categorii [1]:

- sisteme care utilizează transformări ce reduc entropia sursei;
- sisteme care utilizează transformări ce conservă entropia sursei, dar reduc redundanţa.

În ambele cazuri sursa de informaţie poate fi o sursă discretă sau continuă. Dacă mesajul este continuu, el se eşantionează şi de cele mai multe ori se cuantizează, respectiv sursa continuă este transformată într-o sursa discretă. Operaţiile de compresie se efectuează în general cu mărimi cuantizate, deci cu surse discrete, de unde şi denumirea de *compresie de date*.

Transformările care conservă entropia sunt reversibile şi se referă la surse discrete fără memorie, cu debit controlabil. Prin codări corespunzătoare (în general codarea Huffmann), sursa de entropie este transformată într-o sursă de entropie maximă (sau aproape ca ea), respectiv, se efectuează o operaţie de reducere a redundanţei. Simbolurile sursei de entropie maximă, respectiv, de redundanţă zero, sunt codate în cuvinte de lungime medie minimă şi, în consecinţă, eficienţa transmisiunii este maximă, iar stocarea lor se poate face în memorie de capacitatea mai mică decât în absenţa codării.

Metodele de compresie care conservă entropia sunt: *codare cu pas variabil*, toate tipurile de *modulaţie diferenţială a impulsurilor în cod*, inclusiv *modulaţia delta* şi *metodele de compandare*.

Transformările care reduc entropia sunt transformări reversibile, care introduc distorsiuni. În funcţie de gradul de distorsiune admis, compresia realizată poate fi mai mică sau mai mare. Transformări ce reduc entropia sunt utilizate în special la compresia semnalului vorbirii şi la compresia semnalului de televiziune. Fiindcă în ambele cazuri destinatarul este un observator uman, dacă se ţine seama de particularităţile sistemului auditiv, respectiv vizual, se pot obţine compresii foarte mari. Transformări ce reduc entropia sunt aplicate şi în sistemele de telemetrie, unde destinatarul nu este un observator uman. În acest caz, se poate defini o măsură cantitativă a distorsiunilor admise şi se poate defini un raport de compresie care depinde de aceste distorsiuni.

Transformările ortogonale sunt utile în compresia de date, fiindcă fiecare vector în spaţiul transformat provine din combinaţii liniare ale tuturor vectorilor din spaţiul de bază.

Compresii de date prin transformări ortogonale

Dacă considerăm o sursă de entropie maximă care în intervalul de timp *D* are *N* eşantioane care sunt cuantizate de *K* biţi, atunci [2]:

$$H_N = N\log K. \tag{1}$$

Din această relaţie se vede că pentru a reduce entropia este mult mai avantajos să micşorăm numărul N de dimensiuni ale spaţiului în care este reprezentat mesajul, decât să micşorăm numărul de nivele de cuantizare, respectiv pe K, din cauză că H scade cu $logK$.

Eliminarea unor vectori din spaţiul transformat are ca urmare că se pierde ceva din informaţia referitoare la toţi vectorii din spaţiul de bază, în loc să se piardă toată informaţia privind vectorii din spaţiul de bază eliminaţi în ipoteza în care compresia de date s-ar efectua în acest spaţiu de bază. Mai mult, în spaţiul de bază, unde componentele semnalului sunt eşantionate lui, nu se poate oferi nici o reducere a dimensiunilor spaţiului, fiindcă, mesajul fiind staţionar, toate eşantioanele lui au aceeaşi dispersie σ^2.

Dacă efectuăm însă o transformare liniară, componentele în spaţiul transformat nu vor mai avea aceeaşi dispersie. Componentele cu dispersii mai mici vor putea fi reprezentate de cuvinte de lungime mai mică, eventual unele din componente pot fi neglijate. Dacă, în loc să transmitem cuvinte (numere) care să reprezinte eşantioane, transmitem cuvinte (numere) care reprezintă componente comprimate cu spaţiul transformat, efectuăm o compresie de date. La punctul de recepţie se efectuează transformarea inversă [2].

Dintre transformările liniare cele mai avantajoase sunt *transformările ortogonale*.

Dacă T este o transformare ortogonală, atunci:

$$T^tT = I \tag{2}$$

unde I - matricea identitate.

Din (2) rezultă că transformarea inversă T^{-1} este egală cu T^t. Dacă notăm cu X matricea de date, care reprezintă semnalul la intrare presupus cu valoare medie mică:

$$x^t = [x(0),x(1),..,x(N-1)] \tag{3}$$

prin transformarea T obţinem:

$$X = Tx \tag{4}$$

unde

$$X^t = [X(0),X(1),...X(N-1)] \tag{5}$$

este matricea datelor din spaţiul transformat. Fiindcă :

$$\|X^tx\| = \|x\|^2 \quad \text{şi} \quad X^tX = \|X\|^2 \tag{6}$$

ţinând seama de (4) se obţine:

$$\|x\|^2 = \|X\|^2 \tag{7}$$

adică transformarea ortogonală conservă norma, cu alte cuvinte, prin transformarea ortogonală nu se modifică puterea semnalului [2].

Transformare ortogonală este definită de setul de vectori ortogonali $\{\varphi_i\}$:

$$T^t = [\varphi_0,\varphi_1...\varphi_{N-1}] \tag{8}$$

cu proprietatea:

$$\varphi_i\ \varphi_j = \delta_{ij} \tag{9}$$

Ţinând seama de (8) şi (4) putem scrie:

$$x = T^t\,X = [\varphi_0,\varphi_1...\varphi_{N-1}]X \tag{10}$$

$$\text{sau: } x = X(0)\varphi_0 + X(1)\varphi_1 + ...+ X(N-1)\varphi_{n-1} = \sum_{i=0}^{N-1} X(i)\varphi_i \tag{11}$$

Pentru a avea compresie de date reducem dimensiunea spaţiului transformat de la N la M cu $N<M$ şi în consecinţă, obţinem:

$$\hat{x} = \sum_{i=0}^{N-1} X(i)\varphi_i \tag{12}$$

Eroare introdusă neglijând $N–M$ termeni este:

$$\varepsilon_x = x - \hat{x}$$

sau

$$\varepsilon_x = \sum_{i=M}^{N-1} X(i)\varphi_i \tag{13}$$

Eroarea medie pătratică este:

$$\varepsilon = \|x\|^2 = \varepsilon x^t \varepsilon x \tag{14}$$

Substituind (12) în (13) se obţine:

$$\varepsilon = \sum_{i=M}^{N-1}\sum_{j=M}^{N-1} X(i)X(j)\varphi_i{}^t\varphi_j = \sum_{i=M}^{N-1}\overline{X(i)^2} \tag{15}$$

Secvenţa de intrare *x* este caracterizată static de matricea de covariaţie *K* şi, în consecinţă, vom căuta să exprimăm pe ε în funcţie de această matrice. Funcţia de covariaţie a datelor la intrare este:

$$Kx = \overline{xx^t} \tag{16}$$

Din (9) şi (11) rezultă

$X(i) = \varphi_i{}^t x$,

care înlocuită in (14) ne dă:

$$\varepsilon = \sum_{i=M}^{N-1}\overline{\varphi_i^t x(\varphi_i^t x)^t} = \sum_{i=M}^{N-1}\varphi_i^t\,\overline{xx^t}\,\varphi_i \tag{17}$$

sau ţinând seama de (16):

$$\varepsilon = \sum_{i=M}^{N-1}\varphi_i^t K_x \varphi_i \tag{18}$$

Problema care se pune este de a determina vectorii (matricele) φ_i, respectiv, transformarea *T* care să minimizeze eroarea pătratică medie, fiind dată funcţia de covariaţie K_x. Minimumul lui ε se obţine ţinând seama de constrângerea dată de relaţia (10). Utilizând metoda multiplicatorului lui Lagrange, avem:

$$\varepsilon = \sum_{i=M}^{N-1}\varphi_i^t K_x \varphi_i - \sum_{i=M}^{N-1}\lambda_i(\varphi_i^t\varphi_i - 1)$$

sau

$$\varepsilon = \sum_{i=M}^{N-1}\{\varphi_i^t K_x \varphi_i - \lambda_i(\varphi_i^t\varphi_i - 1)\} \tag{19}$$

Pentru a găsi minimumul luăm gradientul $\nabla\varphi_i$ la relaţia (19) şi-l adunăm:

$$\nabla\varphi_i\{\varepsilon\} = 0 \tag{20}$$

Gradientul $\nabla\varphi_i$ respectiv derivarea în raport cu matricea φ_i, este o generalizare a noţiunii de derivată. Dacă matricea K_x este simetrică avem:

$$\nabla\varphi_i\{\varphi_i{}^t K_x \varphi_i\} = 2K_x\varphi_i \tag{21}$$

şi

$$\nabla\varphi_i\{\varphi_i{}^t\varphi_i\} = 2\varphi_i \tag{22}$$

Introducând (21) şi (22) în (20) obţinem:

$$\nabla\varphi_i\{\varepsilon\} = 2K_x\varphi_i - 2\varepsilon_i\varphi_i = 0, \tag{23}$$

de unde:

$$K_x\varphi_i = \varepsilon_i\varphi_i \tag{24}$$

Din această relaţie rezultă, că transformarea care dă eroare medie pătratică minimă are ca vectori de bază, vectorii proprii ai matricei de covariaţie K_x a semnalului considerat.

Înlocuind (24) în (19) obţinem:

$$\varepsilon_{min} = \sum_{i=M}^{N-1}\lambda_i\,, \tag{25}$$

unde λ_i - valori proprii ale matricei K_x.

Transformarea definită de (25) este optimă în sensul că nici o altă transformare nu dă o eroare mai mică. Această transformare se numeşte Karhunen-Loeve.

Compresie de date cu ajutorul transformatei Karhunen-Loeve

Un mesaj cu funcţia de covariaţie $K(t_1,t_2)$ poate fi dezvoltat în serie cu coeficienţi necorelaţi, dacă funcţiile de reconstrucţie $U_i(t)$ sunt soluţii ale ecuaţiei integrale [2]:

$$\int_0^D K(t_1,t_2)\varphi_i(t_2)dt_2 = \lambda i\varphi_i(t_1), \tag{26}$$

unde $K(t_1,t_2)$ este nucleul ecuaţiei, $\varphi_i(t)$ sunt vectorii-funcţii proprii şi λ_i - valorile proprii.

Considerând momente de timp discrete:

$t_1 = mT = m$ şi $t_2 = nT = n$, $(T=1)$

şi introducând notaţia:

$K(t_1,t_2) = K(m,n) = K_m(n)$,

integrala (26) poate fi scrisă sub forma de produs scalar:

$$(\vec{K}_m,\vec{\varphi}_i) = \sum_{n=0}^{N-1} K_m(n)\varphi_i(n) = \lambda i\varphi_i(m), \qquad m = 0,1,\ldots,N-1. \tag{27}$$

Acest produs scalar poate fi scris sub forma matricială:

$$\begin{pmatrix} K(0,0) & K(0,1) & \cdots & K(0,N-1) \\ K(1,0) & K(1,1) & \cdots & K(1,N-1) \\ \cdots & \cdots & \cdots & \cdots \\ K(N-1,0) & K(N-1,1) & \cdots & K(N-1,N-1) \end{pmatrix} \begin{pmatrix} \varphi_i(0) \\ \varphi_i(1) \\ \cdots \\ \varphi_i(N-1) \end{pmatrix} = \lambda_i \begin{pmatrix} \varphi_i(0) \\ \varphi_i(1) \\ \cdots \\ \varphi_i(N-1) \end{pmatrix} \tag{28}$$

respectiv:

$$K_x\varphi_i = \lambda_i\varphi_i \tag{29}$$

Această funcţie s-a obţinut din condiţia că dezvoltarea în serie cu funcţiile de reconstrucţie $\varphi_i(t)$ să aibă coeficienţii $X(i)$ necorelaţi. Matricea de covariaţie K_x în spaţiul transformat este:

$$K_x = XX^t \tag{30}$$

sau fiindcă $X = T_x$ avem:

$$K_x = TXX^tT^t = TK_xT^t \tag{31}$$

Dat fiind că transformata Karhunen-Loeve are coeficienţi necorelaţi:

$$\alpha_i\alpha_j = \delta_{ij}\lambda_i \tag{32}$$

rezultă că K_x este o matrice diagonală:

$$K_x = \mathrm{diag}(\lambda_0,\lambda_1,\ldots,\lambda_{N-1}), \tag{33}$$

unde valorile proprii λ_i reprezintă dispersiile componentelor $X(i)$ (de valoare medie nulă).

Dacă, în loc să transmitem (sub forma cuantizată) toţi parametrii $X(i)$, transmitem numai $N-M$, realizăm o compresie de date cu o distorsiune dată de relaţia:

$$\varepsilon_{min} = \sum_{i=m}^{N_i-1} \lambda_i \tag{34}$$

Numărul $N-M$ de componente $X(i)$ pe care le putem neglija corespunde cu numărul de coeficienţi λ_i din matricea de covariaţie K_x care pot fi neglijaţi.

Concluzii

Transformările ortogonale sunt cele mai avantajoase, deoarece ele consideră entropia şi transformarea inversă se obţine uşor, prin transpunere.

Din cauză că transformata Karhunen-Loeve nu are algoritm de calcul rapid, determinarea vectorilor proprii ai matricei de covariaţie este dificilă şi este rareori folosită, preferându-se transformări suboptimale. Ea este însă deosebit de utilă ca referinţă, permiţând să se calculeze cât de mult se îndepărtează o anumită transformată de optim.

Bibliografie

1. Murgan A. T. Teoria transmisiunii informaţiei – Probleme. – Bucureşti: Editura Didactică şi Pedagogică, 1983. - 252 p.
2. Spătaru Al. Teoria transmisiunii informaţiei. Coduri şi decizii statistice. – Bucureşti: Editura Didactică şi Pedagogică. - 1983. - 538 p.

Prezentat la 16.04.2004