

COMPRESII DE DATE PRIN TRANSFORMĂRI ORTOGONALE

Țicău Vitalie

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

В работе представлены методы сжатия данных при ортогональных преобразованиях, включая преобразование Карунен-Леве. При применении преобразования Карунен-Леве потери информации 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 esențiali pentru destinatar, ceilalți parametri 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 redundanța.

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

Transformările care conservă entropia sunt reversibile și se referă la surse discrete fără memorie, cu debit controlabil. Prin codări corespunzătoare (în general codarea Huffman), sursa de entropie este transformată într-o sursă de entropie maximă (sau aproape ca ea), respectiv, se efectuează o operație de reducere a redundanței. Simbolurile sursei de entropie maximă, respectiv, de redundanță zero, sunt codate în cuvinte de lungime medie minimă și, în consecință, eficiența transmisiunii este maximă, iar stocarea lor se poate face în memorie de capacitatea mai mică decât în absența codării.

Metodele de compresie care conservă entropia sunt: *codare cu pas variabil*, toate tipurile de *modulație diferențială a impulsurilor în cod*, inclusiv *modulația delta* și *metodele de compandare*.

Transformările care reduc entropia sunt transformări reversibile, care introduc distorsiuni. În funcție 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 și la compresia semnalului de televiziune. Fiindcă în ambele cazuri destinatarul este un observator uman, dacă se ține seama de particularitățile sistemului auditiv, respectiv vizual, se pot obține compresii foarte mari. Transformări ce reduc entropia sunt aplicate și în sistemele de telemetrie, unde destinatarul nu este un observator uman. În acest caz, se poate defini o măsură cantitativă a distorsiunilor admise și se poate defini un raport de compresie care depinde de aceste distorsiuni.

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

Compresii de date prin transformări ortogonale

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

$$H_N = N \log K. \quad (1)$$

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

Eliminarea unor vectori din spațiul transformat are ca urmare că se pierde ceva din informația referitoare la toți vectorii din spațiul de bază, în loc să se piardă toată informația privind vectorii din spațiul de bază eliminați în ipoteza în care compresia de date s-ar efectua în acest spațiu de bază. Mai mult, în spațiul de bază, unde componentele semnalului sunt eșantionate lui, nu se poate oferi nici o reducere a dimensiunilor spațiului, fiindcă, mesajul fiind staționar, toate eșantioanele lui au aceeași dispersie σ^2 .

Dacă efectuăm însă o transformare liniară, componentele în spațiul transformat nu vor mai avea aceeași 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 eșantioane, transmitem cuvinte (numere) care reprezintă componente comprimate cu spațiul transformat, efectuăm o compresie de date. La punctul de recepție se efectuează transformarea inversă [2].

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

Dacă T este o transformare ortogonală, atunci:

$$T^t T = I \quad (2)$$

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:

$$x^t = [x(0), x(1), \dots, x(N-1)] \quad (3)$$

prin transformarea T obținem:

$$X = T x \quad (4)$$

unde

$$X^t = [X(0), X(1), \dots, X(N-1)] \quad (5)$$

este matricea datelor din spațiul transformat. Fiindcă :

$$\|X^t x\| = \|x\|^2 \quad \text{și} \quad X^t X = \|X\|^2 \quad (6)$$

ținând seama de (4) se obține:

$$\|x\|^2 = \|X\|^2 \quad (7)$$

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 $\{\varphi_i\}$:

$$T^t = [\varphi_0, \varphi_1, \dots, \varphi_{N-1}] \quad (8)$$

cu proprietatea:

$$\varphi_i \varphi_j = \delta_{ij} \quad (9)$$

Ținând seama de (8) și (4) putem scrie:

$$x = T^t X = [\varphi_0, \varphi_1, \dots, \varphi_{N-1}] X \quad (10)$$

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

Pentru a avea compresie de date reducem dimensiunea spațiului transformat de la N la M cu $N < M$ și în consecință, obținem:

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

Eroare introdusă neglijând $N-M$ termeni este:

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

sau

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

Eroarea medie pătratică este:

$$\varepsilon = \|x\|^2 = \varepsilon x^t \varepsilon x \quad (14)$$

Substituind (12) în (13) se obține:

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

Secvența de intrare x este caracterizată static de matricea de covariație K și, în consecință, vom căuta să exprimăm pe ε în funcție de această matrice. Funcția de covariație a datelor la intrare este:

$$K_x = x x^t \quad (16)$$

Din (9) și (11) rezultă

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

care înlocuită în (14) ne dă:

$$\varepsilon = \sum_{i=M}^{N-1} \varphi_i^t x (x^t \varphi_i)^t = \sum_{i=M}^{N-1} \varphi_i^t x x^t \varphi_i \quad (17)$$

sau ținând seama de (16):

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

Problema care se pune este de a determina vectorii (matricele) φ_i , respectiv, transformarea T care să minimizeze eroarea pătratică medie, fiind dată funcția de covariație K_x . Minimumul lui ε se obține ținând seama de constrângerea dată de relația (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

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

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

$$\nabla \varphi_i \{\varepsilon\} = 0 \quad (20)$$

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

$$\nabla \varphi_i \{\varphi_i^t K_x \varphi_i\} = 2K_x \varphi_i \quad (21)$$

și

$$\nabla \varphi_i \{\varphi_i^t \varphi_i\} = 2\varphi_i \quad (22)$$

Introducând (21) și (22) în (20) obținem:

$$\nabla \varphi_i \{\varepsilon\} = 2K_x \varphi_i - 2\lambda_i \varphi_i = 0, \quad (23)$$

de unde:

$$K_x \varphi_i = \lambda_i \varphi_i \quad (24)$$

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

Înlocuind (24) în (19) obținem:

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

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 numește Karhunen-Loeve.

Compresie de date cu ajutorul transformatei Karhunen-Loeve

Un mesaj cu funcția de covariație $K(t_1, t_2)$ poate fi dezvoltat în serie cu coeficienți necorelați, dacă funcțiile de reconstrucție $U_i(t)$ sunt soluții ale ecuației integrale [2]:

$$\int_0^D K(t_1, t_2) \varphi_i(t_2) dt_2 = \lambda_i \varphi_i(t_1), \quad (26)$$

unde $K(t_1, t_2)$ este nucleul ecuației, $\varphi_i(t)$ sunt vectorii-funcții proprii și λ_i - valorile proprii.

Considerând momente de timp discrete:

$$t_1 = mT = m \quad \text{și} \quad t_2 = nT = n, \quad (T=1)$$

și introducând notația:

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

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

$$(\bar{K}_m, \bar{\varphi}_i) = \sum_{n=0}^{N-1} K_m(n) \varphi_i(n) = \lambda_i \varphi_i(m), \quad m = 0, 1, \dots, N-1. \quad (27)$$

Acest produs scalar poate fi scris sub forma matricială:

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

respectiv:

$$K_x \varphi_i = \lambda_i \varphi_i \quad (29)$$

Această funcție s-a obținut din condiția că dezvoltarea în serie cu funcțiile de reconstrucție $\varphi_i(t)$ să aibă coeficienții $X(i)$ necorelați. Matricea de covariație K_x în spațiul transformat este:

$$K_x = XX^t \quad (30)$$

sau fiindcă $X = T_x$ avem:

$$K_x = TXX^tT^t = TK_xT^t \quad (31)$$

Dat fiind că transformata Karhunen-Loeve are coeficienți necorelați:

$$\alpha_i \alpha_j = \delta_{ij} \lambda_i \quad (32)$$

rezultă că K_x este o matrice diagonală:

$$K_x = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{N-1}), \quad (33)$$

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

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

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

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

Concluzii

Transformările ortogonale sunt cele mai avantajoase, deoarece ele consideră entropia și transformarea inversă se obține ușor, prin transpunere.

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

Bibliografie

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

Prezentat la 16.04.2004