

Codare predictivă

O altă metodă importantă de realizare a compresiei este aceea a abordării predictive. Ea se bazează pe faptul că datele ce urmează a fi comprimate **nu diferă mult de la o valoare la alta**, deci **pot fi predicționate** destul de bine din valorile anterioare. Din acest motiv metodele predictive se aplică în special în cazul în care datele reprezintă de fapt serii în timp (cazul unidimensional, de exemplu semnale audio) sau serii în spațiu (cazul bidimensional, de exemplu imagini).

Deși în cele ce urmează vom prezenta aceste metode în contextul particular al compresiei vom folosi termenul de codare predictivă, el fiind mult mai des folosit astfel în literatura de specialitate.

1. Schema generală de codare-decodare predictivă

Algoritmul de compresie bazat pe codare predictivă funcționează astfel (Figura 1):

1. pentru fiecare valoare de intrare este **predicționată o valoare** pe baza unei scheme de predicție
2. se calculează **eroarea de predicție** corespunzătoare ca fiind **diferența** între valoarea reală și valoarea predicționată
3. valoarea **erorii de predicție** este transmisă unui **codor entropic** (Huffman , aritmetic etc.) și apoi salvată în fluxul comprimat.

Decompresia funcționează simetric procesului de compresie (Figura 2):

1. se preia din fluxul comprimat o valoare, se **decodează entropic** și se obține **eroarea de predicție**.
2. pe baza aceleiași scheme de predicție se **predicționează valoarea** curentă
3. la valoarea predicționată se **adună** eroarea de predicție și se obține valoarea reală care se salvează în fluxul decodat.

Pentru ca schema să funcționeze corect este esențial ca predicția să se facă doar pe baza valorilor deja codate, deoarece doar acestea există la decodor (predicția se face doar pe baza "trecutului"). Această condiție este necesară pentru ca cele două predictoare să funcționeze identic (să producă aceeași valoare predicționată).

Așa cum este descrisă anterior (la codor se scade și la decodor se adună valoarea predicției) metoda se încadrează în categoria metodelor lossless.

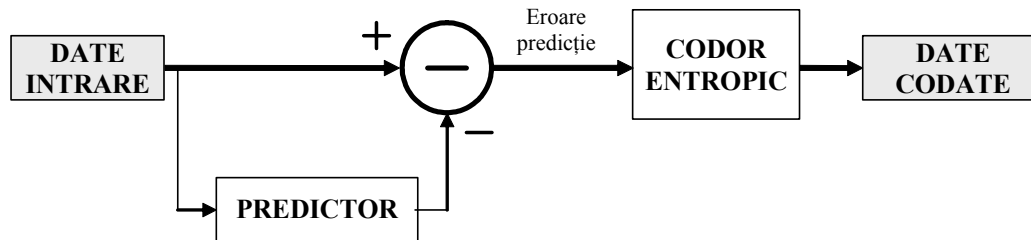


Figura 1 Schema generală de compresie folosind tehnici predictive

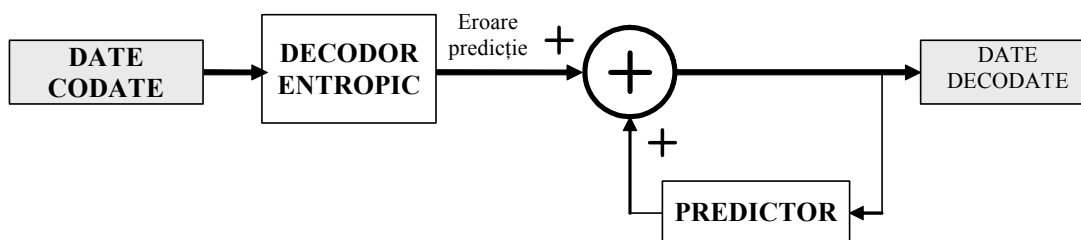


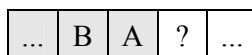
Figura 2 Schema generală de decompresie folosind tehnici predictive

2. Codare predictivă pe caz unidimensional (serie de timp) - Exemplu

Să considerăm seria:

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Valoare initiala VI	0	4	7	9	10	9	7	4	0	-4	-7	-9	-10	-9	-7	-4	0	4
Val. predicționată VP	0	0	4	7	9	10	9	7	4	0	-4	-7	-9	-10	-9	-7	-4	0
Eroare predicție EP	0	4	3	2	1	-1	-2	-3	-4	-4	-3	-2	-1	1	2	3	4	4

După cum se poate observa seria reprezintă eșantionarea unui semnal sinusoidal. Prin index am notat indexul (de timp) din cadrul seriei. În exemplu s-a considerat **predictorul A** (se predicționează cu valoarea anterioară din serie).



Se determină apoi pas cu pas pentru fiecare index k :

$$VP[k] = VI[k-1] \quad // \text{ predicție}$$

$$EP[k] = VI[k] - VP[k] \quad // \text{ eroare de predicție}$$

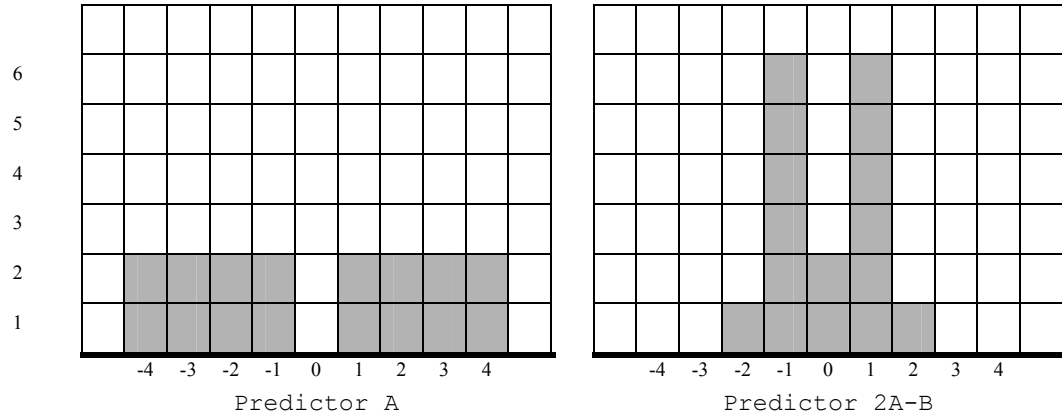
La început, când nu putem aplica predictorul ales (nu există valoare anterioară) trebuie făcută o convenție privind predicția, de exemplu se consideră $VP[0] = 0$.

Să reluăm exemplul anterior pentru cazul **predictorului 2A-B**. Acesta poate fi interpretat și ca $A + (A - B)$ reprezentând prelungirea segmentului care trece prin A și B. După calcule asemănătoare obținem valorile:

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Valoare initiala VI	0	4	7	9	10	9	7	4	0	-4	-7	-9	-10	-9	-7	-4	0	4	7
Val. predicționată VP	0	0	8	10	11	11	8	5	1	-4	-8	-10	-11	-11	-8	-5	-1	4	8
Eroare predicție EP	0	4	-1	-1	-1	-2	-1	-1	-1	0	1	1	1	2	1	1	1	0	-1

În acest caz pentru primele două valori predicționate se consideră $VP[0]=0$, $VP[1]=VI[0]=0$;

Pentru compararea performanțelor celor 2 predictorii să determinăm histogramele erorii de predicție:



Am luat în considerare în fiecare caz primele 16 valori ale erorii de predicție corespunzătoare predicțiilor efectiv făcute cu A respectiv 2A-B (nu și cazurile speciale de la început).

Constatăm că pentru 2A-B histograma este mai dezechilibrată spre valoarea 0. Acest fapt va permite codorului entropic să poată obține performanțe de compresie mai bune. Ambele histograme sunt însă mult mai dezechilibrate spre 0 în raport histograma datelor inițială.

La decodare se pornește doar de la șirul erorilor de predicție. Se determină apoi **pas cu pas** pentru fiecare index:

$$VP[k] = VD[k-1] \quad // \text{ predicție}$$

$$VD[k] = VP[k] + VP[k] \quad // \text{ reconstrucție}$$

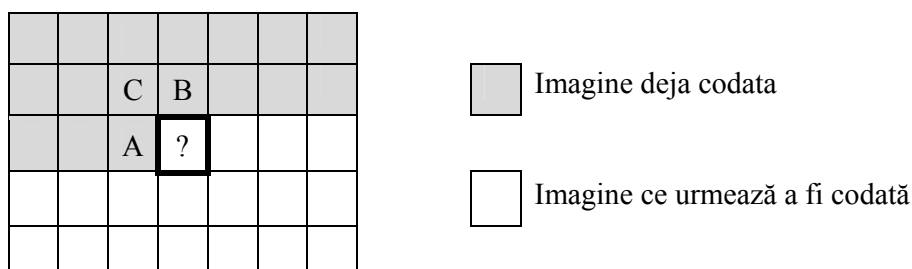
Evident, la început când nu putem aplica predictorul ales, trebuie făcute aceleași predicții ca și la codor. Prezintă valorile de la decodare în cazul predictorului 2A-B.

Index	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Eroare predicție EP	0	4	-1	-1	-1	-2	-1	-1	-1	0	1	1	1	2	1	1	1	0	-1
Val. predicționată VP	0	0	8	10	11	11	8	5	1	-4	-8	-10	-11	-11	-8	-5	-1	4	8
Valoare decodată VD	0	4	7	9	10	9	7	4	0	-4	-7	-9	-10	-9	-7	-4	0	4	7

Se observă ca datele au fost restaurate fără pierderi, după cum era de așteptat.

3. Codare predictivă pe caz bidimensional (imagini)

În cazul unidimensional metoda se baza pe corelația valorilor alăturate și deci pe predictibilitatea valorii următoare pe axa timpului. În cazul bidimensional există corelație pe ambele axe, fapt de care trebuie să țină cont predictorii specifici imaginilor.



În acest caz predicția se realizează (aproape întotdeauna) pixel cu pixel de la stânga la dreapta și de sus în jos. Predicția se poate face după orice regulă, dar doar pe baza pixelilor din zona gri (imaginea deja parcursă). Această condiție este necesară pentru ca cele două predictoare să funcționeze identic (să producă aceeași valoare predicționată).

Prezentăm în cele ce urmează cele mai importante reguli de predicție folosite în acest context.

Reguli de predicție utilizate de standardul JPEG

Standardul JPEG consideră pentru modul de lucru lossless (complet diferit de cel lossy, bazat pe DCT) o abordare predictivă standard și următorii 7 predictorii:

- 1: A
- 2: B
- 3: C
- 4: $A + B - C$
- 5: $A + (B - C) / 2$
- 6: $B + (A - C) / 2$
- 7: $(A + B) / 2$

Primii 3 sunt unidimensionali iar ultimii 4 sunt realmente bidimensionali. Notățiile A, B respectiv C sunt cele descrise în figura anterioară.

Regula de predicție din LOCO-I (standardizată în JPEG-LS)

În acest caz valoarea predicționată este dată de expresia:

$\min(A, B)$	daca $C \geq \max(A, B)$
$\max(A, B)$	daca $C \leq \min(A, B)$
$A+B-C$	in rest

Predictorul este astfel conceput încât să aleagă B în cazul în care există un contur vertical în locația curentă, A în cazul în care există un contur deasupra pixelului curent și $A+B-C$ dacă nu este detectată nici o linie de contur. Ultimul tip de predicție este similar cu alegerea valorii curente în cazul în care pixelul curent ar aparține planului definit de cei trei vecini cu ponderi A, B și C.

Acest tip de predictor a fost folosit în aplicațiile de predicție de date sub diverse interpretări. Cea mai importantă este aceea de mediere a celor trei predictor A, B, $A+B-C$. Combinând aceste interpretări acest predictor a fost denumit în cadrul procesului de standardizare “Detector Median de Contururi” (Median Edge Detector).

Regula de predicție din CALIC

Algoritmul de predicție al metodei **CALIC** este prezentat în cele ce urmează:

DACA $(d_v + d_h > 32)$
 $I^*[i, j] = (d_v * I[i-1, j] + d_h * I[i, j-1]) / (d_v + d_h) + (I[i+1, j-1] - I[i-1, j-1]) / 8$

ALTFEL DACA $(d_v - d_h > 12)$
 $I^*[i, j] = (2 * I[i-1, j] + I[i, j-1]) / 3 + (I[i+1, j-1] - I[i-1, j-1]) / 8$

ALTFEL DACA $(d_h + d_v > 12)$
 $I^*[i, j] = (I[i-1, j] + 2 * I[i, j-1]) / 3 + (I[i+1, j-1] - I[i-1, j-1]) / 8$

ALTFEL
 $I^*[i, j] = (I[i-1, j] + I[i, j-1]) / 2 + (I[i+1, j-1] - I[i-1, j-1]) / 8$

DACA $(d_{45} - d_{135} > 32)$
 $I^*[i, j] = I^*[i, j] + (I[i+1, j-1] - I[i-1, j-1]) / 8$

ALTFEL DACA $(d_{45} - d_{135} > 16)$
 $I^*[i, j] = I^*[i, j] + (I[i+1, j-1] - I[i-1, j-1]) / 16$

ALTFEL DACA $(d_{135} - d_{45} > 32)$
 $I^*[i, j] = I^*[i, j] + (I[i-1, j-1] - I[i+1, j-1]) / 8$

ALTFEL DACA $(d_{135} - d_{45} > 16)$
 $I^*[i, j] = I^*[i, j] + (I[i-1, j-1] - I[i+1, j-1]) / 16$

unde $I[i, j]$ reprezintă poziția ce este analizată la un moment dat în cadrul imaginii, $I^*[i, j]$ reprezintă valoarea predicției iar d_{45} , d_{135} , d_v și d_h reprezintă:

$$\begin{aligned} d_h &= |I[i-1, j] - I[i-2, j]| + |I[i, j-1] - I[i-1, j-1]| + |I[i+1, j-1] - I[i, j-1]| \\ d_v &= |I[i-1, j] - I[i-1, j-1]| + |I[i, j-1] - I[i, j-2]| + |I[i+1, j-1] - I[i+1, j-2]| \\ d_{45} &= |I[i-1, j] - I[i-2, j-1]| + |I[i-1, j-1] - I[i-2, j-2]| + |I[i, j-1] - I[i-1, j-2]| \\ d_{135} &= |I[i+1, j-1] - I[i+2, j-2]| + |I[i, j-1] - I[i+1, j-2]| + |I[i, j-1] - I[i-1, j]| \end{aligned}$$

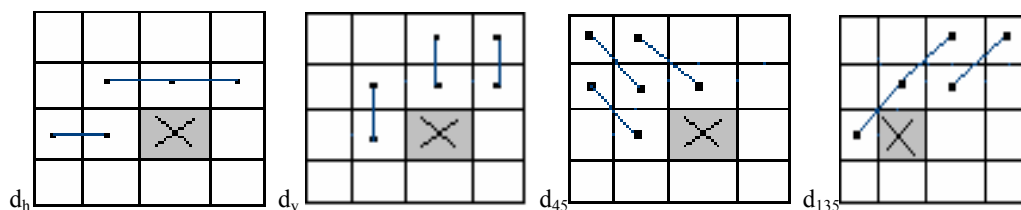


Figura 3 Pixelii utilizați în determinarea valorilor d_h , d_v , d_{45} , d_{135}

După cum se poate observa se determină gradientul pe orizontală, verticală și la 45° respectiv 135° (situație prezentată și în Figura 3) și în funcție de aceste valori se aduce o corecție valorii predicționate anterior.

Observație pentru codarea predictivă a imaginilor

În situațiile în care nu se poate aplica regula de predicție bidimensională dorită deoarece nu există toți pixelii necesari trebuie convenită (cu decodorul) o abordare alternativă. Acest lucru se întâmplă de regulă pe primul pixel (unde se alege o predicție fixă), pe prima linie a imaginii (unde se poate alege predictorul A) și pe prima coloană (unde se poate alege predictorul B).

4. Observații

În ce privește codarea predictivă facem următoarele observații importante:

1. prin aplicarea predicției **gama dinamică a semnalului crește** astfel

Gama semnal intrare		Gama semnal iesire
$[0 \div N]$	\Rightarrow	$[-N \div +N]$
$[-N \div +N]$	\Rightarrow	$[-2N \div +2N]$

solicitând pentru reprezentare **un bit în plus**. Deci, prin predicție, nu doar ca nu apare compresie, ci chiar apare o creștere a dimensiunii de reprezentare. Totuși, acest fapt se compensează prin dezechilibrarea histogramei spre valoarea 0, ceea ce va duce la o compresie mai bună realizată de codorul entropic care urmează.

2. În codarea predictivă problema **nu se pune în sensul clasificării binare HIT / MISS** în ceea ce privește valoarea predicționată (ca în cazul predicțiilor implementate în hardware). O predicție “mai bună” se deosebește de una “mai rea” prin modul în care modifică mai tare distribuția înspre 0 și astfel permite codorului entropic să realizeze o compresie mai ridicată.

3. Codarea entropică care urmează etapei de predicție se realizează în general contextual, având câte **un model** statistic **pentru fiecare context** posibil, contextele fiind definite tot în funcție de valorile pixelilor folosiți la predicție (problema contextelor nu este detaliată aici).

4. Metoda predictivă nu impune o anumită regula de predicție utilizată, putând să folosim **reguli de predicție** oricât de **elaborate** (polinomiale de grad mai mare, predictoare neuronale etc.) **adaptate** specificului datelor.