Ľubica Fajglová

Určenie koeficienta kompresie pre redukciu dát metódou lineárnej interpolácie

Kybernetika, Vol. 6 (1970), No. 6, (436)--441

Persistent URL: http://dml.cz/dmlcz/124790

Určenie koeficienta kompresie pre redukciu dát metódou lineárnej interpolácie

ĽUBICA FAJGLOVÁ

Článok pojednáva o jednom algoritme kompresie dát metódou lineárnej interpolácie. Ako číselnú charakteristiku kvality metódy definuje koeficient kompresie, ktorý je analyticky vyjadrený pre náhodný signál s daným rozložením pravdepodobností.

1. FORMULÁCIA ÚLOHY KOMPRESIE DÁT

Od vzniku prenosovej techniky je snaha po návrhu jednoduchého a spoľahlivého zariadenia, ktoré by umožňovalo prenos všetkých významných hodnôt sledovaného signálu. Výber signifikantných vzoriek môže byť uskutočnený rôznymi metódami, ktorých spoločný názov je „data compression" (kompresia dát).

Kompresia dát môže byť charakterizovaná ako metóda, ktorá zmenšuje šírku pásma potrebného k prenosu daného množstva informácií v danom čase, alebo metóda, ktorá zmenšuje potrebný čas pre prenos daného množstva informácií v určitej šírke pásma.

Je možné navrhnúť mnoho metód, ktoré splňajú tieto požiadavky. Sú rozdelené do štyroch základních skupín [1]:

1. metóda výberu parametrov,
2. metóda adaptívneho vzorkovania,
3. metóda redundantnej redukcie,
4. metóda štatistického kódovania.

Pri výbere vhodnej metódy pre konkrétny prípad pristupujeme k problémom z týchto hľadísk:

1. je potrebné rozhodnúť sa, či chceme komprimovaný signál na konci prenosovej cesty rekonštruovať do pôvodného tvaru alebo nie. To znamená, či potrebujeme ďalšie spracovanie poznať okamžité hodnoty vstupného signálu so zadanou dovolenou odchýlkou medzi dvomi nasledujúcimi vzorkami ε,

2. je potrebné rozhodnúť sa, v ktorom mieste sa uskutoční kompresia dát. V zásadnom riešení sú dve možnosti:

a) zostrojenie reduktora — v podstate je to jednoduchý jednoúčelový počítač, ktorý je umiestnený v tesnej blízkosti snímača dát (priamo v hale výrobného procesu, ktorý je riadený);

b) druhé riešenie umožňuje riešiť kompresiu dát na konci prenosovej cesty pomocou vhodne navrhnutých algoritmov pre riadiaci počítač.

Rozhodnutie, ktorú z metód vyhovujúcu jednej alebo druhej podmienke, použijeme, môžeme uskutočniť len na základe dôkladnej analýzy konkrétnej úlohy, v ktorej má byť kompresia dát použitá.

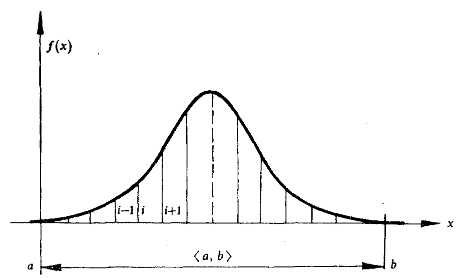
Pri riadení rozsiahlych sústav je výhodné použiť kompresiu pomocou metódy redundantnej redukcie. Redundantná redukcia je spoločný názov pre skupinu porovnávacích a interpolačných metód.

2. ALGORITMUS KOMPRESIE METÓDOU LINEÁRNEJ INTERPOLÁCIE

Predpoklady:

1. Nech kompresii podlieha stacionárny náhodný signál s normálnym rozložením pravdepodobností, daným rozptylom σ^2 na ohraničenom intervale $\langle a, b\rangle$ (obr. 1)
2. Tento signál je vzorkovaný konštantnou rýchlosťou a kvantovaný v N kvantovacích ekvidistantných úrovniach.
3. Počas aplikácie kompresného algoritmu nevznikne žiadna nová chyba.
4. Počiatočný bod interpolačnej úsečky i je totožný s koncovým bodom úsečky a predchádzajúcej kvantovacej úrovne.

Šírku jednej kvantovacej úrovne určíme $L = (b - a)/N$, ak súradný systém zvolíme tak, aby $a = 0$; vzťah sa zjednoduší na $L = b/N$.



Obr. 1. Rozloženie pravdepodobností signálu podrobného kompresii.

Argument x funkcie $f(x)$ sa pohybuje v intervale

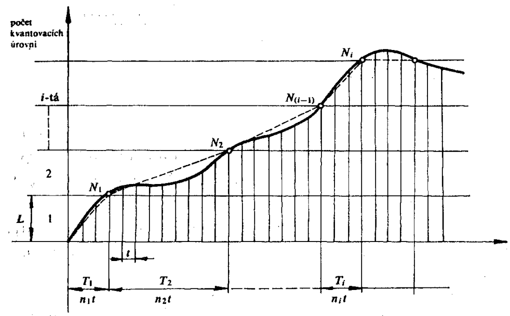
$$(i-1)\,L \leqq x \leqq iL\,.$$

Pre normálne rozloženie pravdepodobnosti funkcie $f(x)$ platí:

$$p_i = \frac{1}{\sqrt{2\pi}} \int_{(i-1)L}^{iL} \mathrm{e}^{-t^2/2}\,\mathrm{d}t\,. \tag{1}$$

Ohraničujúce podmienky, za ktorých môžeme kompresiu pomocou lineárnej interpolácie uskutočniť, vyjadríme:

$$1.\quad p(i,i) + p(i,i+1) + p(i,i-1) = 1\,, \tag{2}$$



Obr. 2. Algoritmus lineárnej interpolácie.

to znamená, že kvantovacie úrovne sú natoľko široké, že nasledujúca nameraná vzorka sa nachádza len v susednej hladine $(i+1)$-nej alebo $(i-1)$-nej alebo zostáva v pôvodnej i-tej.

$$2.\quad p(i,i+1) : p(i,i) : p(i,i-1) = p_{i+1} : p_i : p_{i-1}\,, \tag{3}$$

so znamená, že úrovne sú natoľko úzke, že prechod závisí len na rozložení pravdepodobností funkcie.

Z obr. 2 vyplýva, že kompresia nameraných dát metódou lineárnej interpolácie pozostáva z toho, že namiesto prenesenia všetkých nameraných vzoriek prenášame iba dvojicu údajov (N_i, T_i), ktoré dostatočne charakterizujú interpolačnú úsečku v jednej kvantovacej hladine.

Koeficient kompresie C [2] je číselná charakteristika, pomocou ktorej posudzujeme kvalitu použitej kompresnej metódy. Je to bezrozmerné číslo, dané pomerom množstva celkovej nameranej informácie

$$I_n = \bar{n} \ln N$$

a množstva informácie prenesenej

$$I_p = \ln N + \ln \bar{T},$$

kde $\bar{n}$ – priemerný počet vzoriek v jednej úrovni; N – počet kvantovacích úrovní; $\bar{T}$ – priemerný čas, v ktorom sa nachádza signál v príslušnej úrovni (viď obr. 2).

Priemerný počet vzoriek (krokov) v jednej úrovni určíme:

$$\bar{n} = \sum_{n=0}^{\infty} n p_{ii}^n, \tag{4}$$

kde $p_{(i,i)}$ – pravdepodobnosť zotrvania vzorky v i-tej úrovni.

Zo vzťahov (1), (2), (3) vyplýva:

$$p_{i,i} = \frac{p_i}{p_{i-1} + p_i + p_{i+1}} = \frac{\int_{(i-1)L}^{iL} e^{-t^2/2}\, dt}{\int_{(i-2)L}^{(i+1)L} e^{-t^2/2}\, dt}. \tag{5}$$

Keď dosadíme (5) do (4), dostávame priemerný počet krokov v jednej hladine:

$$\bar{n} = \frac{1}{N} \sum_{i=1}^{N} \sum_{n=0}^{\infty} n p_{ii}^n = \frac{1}{N} \sum_{i=1}^{N} \sum_{n=0}^{\infty} n \left(\frac{\int_{(i-1)L}^{iL} e^{-t^2/2}\, dt}{\int_{(i-2)L}^{(i+1)L} e^{-t^2/2}\, dt} \right)^n. \tag{6}$$

Z definície kompresného pomeru vyplýva:

$$\mathrm{C} = \frac{I_n}{I_p} = \frac{\bar{n} \ln N}{\ln N + \ln \bar{T}}. \tag{7}$$

Z obr. 2 je zrejmé, že $T_1 = n_1 t$, $T_2 = n_2 t, \ldots, T_i = n_i t$.

Ak zvolíme krok vzorkovania $t = 1$, je zrejmé, že $T_i = n_i$, a teda aj $\bar{T} = \bar{n}$, potom (7) prepíšeme do tvaru

$$\mathrm{C} = \frac{\bar{n} \ln N}{\ln \bar{n} + \ln N}. \tag{8}$$

Zo vzťahu (6) dosadíme do (8) a dostávame konečný výraz pre výpočet kompresného koeficienta:

$$(9) \qquad \mathbf{C} = \frac{\dfrac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{\infty} n \left(\dfrac{\int_{(i-1)L}^{iL} e^{-t^2/2}\,dt}{\int_{(i-2)L}^{(i+1)L} e^{-t^2/2}\,dt} \right)^n . \ln N}{\ln \left[\dfrac{1}{N}\sum_{i=1}^{N}\sum_{n=0}^{\infty} n \left(\dfrac{\int_{(i-1)L}^{iL} e^{-t^2/2}\,dt}{\int_{(i-2)L}^{(i+1)L} e^{-t^2/2}\,dt} \right)^n \right] + \ln N} .$$

Kompresný koeficient **C** je závislý len na počte kvantovacích úrovní N. Počet kvantovacích úrovní N je určený požadovanou presnosťou ε, s ktorou má byť vstupný signál prenesený. Praktický dôsledok tohto výpočtu je v tom, že ak máme vstupný signál vyhovujúci požiadavkám odst. 1 podrobiť kompressii dát, aby sa nestratila užitočná informácia v rámci dovolenej odchýlky ε, môžeme presne určiť maximálny možný kompresný koeficient pre daný algoritmus.

(Došlo dňa 30. januára 1970.)

LITERATÚRA

[1] C. A. Andrews, J. M. Davies, G. R. Schwarz: Adaptive Data Compression. Proc. of the IEEE *55* (1967), 3, 267–277.

[2] Lee D. Davisson: The Theoretical Analysis of Data Compression System. Proc. of the IEEE *35* (1968), 2, 176–186.

SUMMARY

Determination of the Compression Coefficient for Data Reduction by the Linear Interpolation Method

ĽUBICA FAJGLOVÁ

The paper is divided in two parts, the first offers a general outline of the methods of data compression applicable to data transfer in controlled systems. The second part tackles in greater detail one of the methods – linear interpolation, pertaining to methods of redundancy reduction. Definition of the compression coefficient **C** is

given as the numerical characteristic of quality of the applied method. Under certain presumptions, with conditions of constraint, this coefficient is expressed analytically $\mathbf{C} = f(\varepsilon)$, where ε means the maximally allowed change between two subsequent samples.

Ing. Ľubica Fajglová, Ústav technickej kybernetiky SAV, Dúbravská cesta 1, Bratislava - Patrónka.