выходе объекта по отношению к входному сигналу $x(t)$, при сравнении образов и решении других задач, требующих оценки отличия зависимости между двумя функциями от пропорциональной, предлагается использовать непропорциональность по производной первого порядка (4). Она равняется нулю независимо от не равного нулю коэффициента пропорциональности для тех значений аргумента, при которых зависимость между исследуемыми функциями является пропорциональной.

SUMMARY

There are new characteristics of numerical function–disproportions. One of them is disproportion on the first-order derivative. It deals with the application of the disproportion on the first-order derivative for estimation of difference between dependence of two functions and proportion dependence. The examples of application of disproportion for comparison of images and their parts, and for evaluation of distortion between input and output signals of the device investigated are given..

СПИСОК ЛИТЕРАТУРЫ

1. Васильев В.И. Распознающие системы: Справочник. – К.: Наукова думка, 1983.
2. Авраменко В.В. Характеристики непропорциальности числовых функций. Деп. в ГНТБ Украины 19.01.98 №59 – Ук98.

Поступила в редколлегию 2 декабря 1998 г.

УДК 681.518.25:519.711.3

ПРО ПРОГНОСТИЧНІ ВЛАСТИВОСТІ ІНВАРІАНТНИХ ПОРЯДКОВИХ СТАТИСТИК

А.С.Краснопоясовський, доц.

При прогнозуванні зміни інформаційної здатності системи розпізнавання доцільно використовувати як одномірні статистичні характеристики вибіркових послідовностей екстремальні порядкові статистики (ЕПС), які іваріантні у межах своєї статистичної структури та визначають екстремум критерію функціональної ефективності системи розпізнавання, що навчається [1]. Згідно з [2] у статистичній структурі (Ω, A, P), де P означає родину ймовірнісних мір на вимірному просторі (Ω, A), множина $A \in$A зветься інваріантною (відносно P), якщо її ймовірність p(A) є сталою для всіх мір p$\in$P. Поняття інваріантної множини поширюється і на статистику S, що задається на вибірковій структурі (Ω, A, p). Велике зацікавлення викликає вивчення асимптотичних властивостей ЕПС, яка має розподіл χ^2, у вигляді

$$S_n^* = \sum_{n=1}^{n^*} \frac{\left(k_n - \overline{k_n}\right)^2}{s_n^2}, \tag{1}$$

де k_n – кількість успіхів, яка дорівнює кількості ознак розпізнавання, що знаходяться у межах своїх контрольних припусків при n-му випробуванні; $\bar{k}_n$, s_n^2 - вибіркові середнє та дисперсія відповідно: n^* - випробування, при якому критерій оптимізації досягає свого екстремуму.

Розглянемо теореми, які дозволяють дослідити ряд важливих властивостей ЕПС (1). Зокрема, теорема 1 встановлює умови існування нижньої та верхньої асимптотичних меж значень ЕПС. При цьому, якщо верхня межа залежить тільки від кількості випробувань, то нижня межа визначається також і з урахуванням оцінки статистичної похибки $\varepsilon=|\,p_i-n_i/n^*\,|$, де p_i – імовірність, а $n_i\,/\,n^*$ - емпірична частота знаходження i-ї ознаки в своєму полі контрольних припусків.

Теорема 1 Для статистики (1), яка є нормованою сумою кількості успіхів у n перших випробуваннях за простою схемою Бернуллі, з імовірністю одиниця виконуються співвідношення:

$$\lim_{n\to\infty}\operatorname{Inf}_{\{n\}}\frac{S^*_n}{n^*\arg^2[\Phi(x)=1-Q/2]}=1\,,\quad \lim_{n\to\infty}\operatorname{Sup}_{\{n\}}\frac{S^*_n}{4\log(n^*!)}=1\,,$$

де $\Phi(x)$ - функція Лапласа; Q - рівень статистичної значущості.

Доведення. Розглянемо нормовану кількість успіхів у n перших випробуваннях за біноміальним законом, яка визначається функцією

$$S_n=\frac{k_n-np}{\sqrt{npq}}\,,$$

де np, npq - математичне очікування та дисперсія відповідно.

За лемою Бореля - Кантеллі з імовірністю одиниця при всіх достатньо великих n верхня межа S_n дорівнює $S_n<2\sqrt{\log n}$. Тоді

$$S^*_n=\sum_{n=1}^{n^*}S_n^2<4\sum_{n=1}^{n^*}\log n=4\log(n^*!).$$

З іншого боку, за посиленим законом великих чисел для $\varepsilon>0$ маємо

$$p\left\{\left|\frac{k_n}{n}-p\right|>\varepsilon\right\}=1\,.$$

Звідки з імовірністю одиниця отримаємо

$$\left|\frac{k_n-np}{\sqrt{npq}}\right|>\frac{\varepsilon\sqrt{n}}{\sqrt{pq}}\,,\tag{2}$$

що дає нижню межу для S_n.

Взявши квадрат від (2) для рівноймовірних результатів випробувань (p=q=0,5), отримаємо $S_n^2>4\varepsilon^2n$, або

$$S^*_n=\sum_{n=1}^{n^*}S_n^2>\sum_{n=1}^{n^*}4\varepsilon^2n\,,\tag{3}$$

Згідно з [3] ε є функцією від n і дорівнює

$$\varepsilon=\frac{C_Q}{2\sqrt{n}}\,,\tag{4}$$

де

$$C_Q=\arg[\Phi(x)=1-Q/2]\cdot\tag{5}$$

Після підстановки (4) з урахуванням (5) у вираз (3) отримаємо

$$S_n^* > \sum_{n=1}^{n^*} \arg^2[\Phi(x) = 1 - Q/2] = n^* \arg^2[\Phi(x) = 1 - Q/2].$$

Отже, значення статистики S_n^* для $n_{\min} \le n \le n^*$, де n_{min} - мінімальна кількість випробувань, яка гарантує прийнятну статистичну похибку, знаходяться в межах $n^* \arg^2[\Phi(x) = 1 - Q/2] < S_n^* < 4\log(n^*!)$, що й потребувало доведення.

Наступна теорема встановлює тенденцію зміни асимптотичних меж статистики (1) при зміні числа успіхів при випробуваннях .

Теорема 2 Нехай дано послідовність незалежних випадкових величин $k_1, k_2, \dots, k_n$ і $k_1', k_2', \dots k_n'$ такі, що $\left|k_n - \bar{k}_n\right| \le c_1$ і $\left|k_n' - \bar{k}_n'\right| \le c_2$, де $1 \le n \le n^*$, при цьому $c_1 > c_2$. Тоді для будь-якого $\alpha > 0$ при $n \to \infty$

$$P\left\{\left[\underset{\{n\}}{Sup}\, S_n^{*} - \underset{\{n\}}{Sup}\, S_n^*\right] > \alpha\right\} = 1;\quad P\left\{\left[\underset{\{n\}}{Inf}\, S_n^{*} - \underset{\{n\}}{Inf}\, S_n^*\right] > \alpha\right\} = 1,$$

де S_n^*, S_n^* - статистики (1) з кількістю успіхів k_n і k_n' відповідно.

Доведення. Оскільки $\left|k_n - \bar{k}_n\right|^2 = \overset{0}{k}{}_n^2 \le c_1^2$, $\left|k'_n - \bar{k}'_n\right|^2 = \overset{0}{k}{}'^2_n \le c_2^2$ і $c_1^2 > c_2^2$, то $s_n^2 \ge s_n'^2$. Подамо статистику S_n^* у вигляді

$$S_n^* = \sum_{n=1}^{n^*} \left(\frac{k_n - \bar{k}_n}{s_n} \right)^2 = \sum_{n=1}^{n^*} t^2 = \chi_n^2.$$

Як відомо, статистика χ^2 має властивість

$$M[\chi^2] = M\left[\sum_{n=1}^{n^*} t^2\right] = \sum_{n=1}^{n^*} M[t^2] = \sum_{n=1}^{n^*} D[t] = n^*,$$

тобто розподіл статистики χ^2 залежить тільки від обсягу вибірки.

Для першої послідовності маємо

$$P\left[\sum_{n=1}^{n^*} \left(k_n - \bar{k}_n\right)^2 > \alpha\right] = P\left[\sum_{n=1}^{n^*} \left(\frac{k_n - \bar{k}_n}{s_n}\right)^2 > \frac{\alpha}{s_n^2}\right] = P\left(\chi_n^2 > \frac{\alpha}{s_n^2}\right) = P\left(\chi_n^2 > \alpha_1\right), \quad (6)$$

де $\alpha_1 = \alpha / s_n^2$.

Для другої послідовності такого ж обсягу, як і перша, аналогічно:

$$P\left[\sum_{n=1}^{n^*} \left(k'_n - \bar{k}'_n\right)^2 > \alpha\right] = P\left[\sum_{n=1}^{n^*} \left(\frac{k'_n - \bar{k}'_n}{s'_n}\right)^2 > \frac{\alpha}{s'^2_n}\right] = P\left(\chi_n^2 > \frac{\alpha}{s'^2_n}\right) = P\left(\chi_n^2 > \alpha_1\right), \quad (7)$$

де $\alpha_2 = \alpha / s'^2_n$.

Обчислення ймовірностей (6) і (7) здійснимо за відомою формулою

$$P\left(\chi_n^2, \alpha\right) = \int_0^{\alpha} f(\chi_n^2)\, d\chi^2, \quad (8)$$

де $f(\chi_n^2)$ - функція щільності розподілу χ_n^2.

Якщо виявиться, що $P(\chi_n^2,\alpha_1) > P(\chi_n^2,\alpha_2)$, то теорему доведено.

Функція щільності розподілу χ_n^2 дорівнює [3]

$$f(\chi_n^2) = \begin{cases} 0 \text{ при } x \le 0, \\ \dfrac{1}{2^{n/2}\Gamma(n/2)} e^{-\frac{t}{2}} t^{\frac{n}{2}-1} \quad \text{при } x > 0, \end{cases} \tag{9}$$

де $\Gamma\left(\frac{n}{2}\right)$ - гамма-функція; t - значення випадкової величини χ_n^2.

Після підстановки (9) у формулу (8) отримаємо:

$$P\left(\chi_n^2 < \alpha_1\right) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} \int_0^{\alpha_1} e^{-\frac{t}{2}} t^{\frac{n}{2}-1} dt = \frac{\gamma\left(\frac{n}{2},\frac{\alpha_1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \tag{10}$$

де $\gamma\left(\frac{n}{2},\frac{\alpha_1}{2}\right)$ - неповна гамма-функція. При цьому

$$P\left(\chi_n^2 > \alpha_1\right) = 1 - P\left(\chi_n^2 < \alpha_1\right). \tag{11}$$

Оскільки $\alpha_1 < \alpha_2$ і підінтегральний вираз у (10) не містить періодичної функції, то $\int_0^{\alpha_1} e^{-\frac{t}{2}} t^{\frac{n}{2}-1} dt < \int_0^{\alpha_2} e^{-\frac{t}{2}} t^{\frac{n}{2}-1} dt$, або $P\left(\chi_n^2 < \alpha_1\right) < P\left(\chi_n^2 < \alpha_2\right)$, з урахуванням (11) $P\left(\chi_n^2 > \alpha_1\right) > P\left(\chi_n^2 > \alpha_2\right)$, що й вимагало доведення.

З теореми 2 випливають наступні наслідки.

Наслідок 2.1 Якщо послідовності незалежних випадкових величин k_n, k'_n, які відповідають умовам теореми 2, розподілені нормально з вибірковими дисперсіями s_n^2, s'^2_n відповідно і $s_n^2 > s'^2_n$, то має місце

$$P\left(\chi'^2_n > S_n^*\right) < P\left(\chi_n^2 > S'^*_n\right).$$

Наслідок 2.2 Якщо послідовності незалежних випадкових величин k_n, k'_n, які відповідають умовам теореми 2, розподілені нормально з вибірковими дисперсіями s_n^2, s'^2_n відповідно і $s_n^2 > s'^2_n$, то з імовірністю одиниця має місце $S'^*_n > S_n^*$, що випливає безпосередньо із наслідку 2.1.

При формуванні варіаційного ряду ЕПС, здобутих на етапі навчання системи розпізнавання, наприклад, за методом функціонально-статистичних випробувань [1], для підвищення точності прогнозування важливо, щоб величина блоків була якомога більшою. Оцінка величини блоків варіаційного ряду для двох випадків згладжування ЕПС розглядається в наступній теоремі.

Теорема 3 Нехай при проведенні n^* незалежних спостережень випадкової величини k_n, що підпорядковується біноміальному закону та характеризує кількість успіхів при n-му спостереженні ($n \le n^*$), обчислено ЕПС:

$$S_n^*(v) = \sum_{n=1}^{n^*} \left(\frac{k_n - \bar{k}_n}{s_n} \right)^{2v}, \quad v = 1,2,$$

де $\bar{k}_n$, s_n^2 - вибіркові середнє та дисперсія відповідно, то при

$$n^* > e^{\frac{N^2\pi^2 - 6C}{6}},$$

де N – кількість ознак розпізнавання ($k_n \le N$); C=0,5772 – стала Ейлера, при $n^* \to \infty$ для $\alpha > 0$ з імовірністю одиниця має місце

$$P\left\{\left[\max_{\{n\}} S^*_{m,n}(\nu=1) - \max_{\{n\}} S^*_{m,n}(\nu=2)\right] > \alpha\right\} = 1.$$

Доведення. Нехай x_i - випадкова величина появи одиниці (успіху) або нуля в i-му розряді N-мірного бінарного вектора реалізації образу за простою схемою Бернуллі. Тоді $k_n = x_1^{(n)} + x_2^{(n)} + ... + x_N^{(n)}$ - випадкова величина появи успіхів при n-му випробуванні. За теоремою Чебишева маємо [4]

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} k_i - \frac{1}{n}\sum_{i=1}^{n} M[x_i]\right| < \varepsilon\right\} = 1\,.$$

Оскільки для біноміального закону $M[x_i] = \frac{n}{2}$, $i = 1,2,...,N$, то $M[k_n] = M[x_1] + M[x_2] + ... + M[x_N] = N \cdot \frac{n}{2}$.

Тоді вибіркове середнє $\bar{k}_n$ при $n \to \infty$ дорівнює $\bar{k}_n = \frac{1}{n} M[K_n] = \frac{N}{2}$, що дає можливість при $k_n \le N$ стверджувати $(k_n - \bar{k}_n)^2 < (N - N/2)^2 = N^2/4$.

Оскільки для біноміального закону дисперсія $s^2 = npq$ мінімальна при $p = q = 0.5$, то має місце

$$\max_{\{n\}} S_n^*(\nu=1) = 4\sum_{n=1}^{n^*} \frac{\left(k_n - \bar{k}_n\right)^2}{n} \le N^2 \sum_{n=1}^{n^*} \frac{1}{n} \approx N^2(\ln n^* + C)\,.$$

Аналогічно $\max_{\{n\}} S_n^*(\nu=2) = 16\sum_{n=1}^{n^*} \frac{\left(k_n - \bar{k}_n\right)^4}{n^2} \le N^4 \sum_{n=1}^{n^*} \frac{1}{n^2} < N^4 \sum_{n=1}^{\infty} \frac{1}{n^2} = N^2 \frac{\pi^2}{6}$.

Отже, якщо $n^* \ge e^{\frac{N^2\pi^2 - 6C}{6}}$, то $\max_{\{n\}} S_n^*(\nu=1) > \max_{\{n\}} S_n^*(\nu=2)$, і при $n^* \to \infty$ для $\alpha > 0$ з імовірністю одиниця має місце

$$P\left\{\left[\max_{\{n\}} S_n^*(\nu=1) - \max S^*_{m,n}(\nu=2)\right] > \alpha\right\} = 1,$$

що й треба було довести.

З теореми 3 випливають два наслідки.

Наслідок 3.1 Межа між максимальними значеннями ЕПС $S_n^*(\nu=1)$ і $S_n^*(\nu=2)$ складає

$$\lim_{n^*\to\infty}\left\{\max_{\{n\}} S^*_{m,n}(\nu=1) - \max_{\{n\}} S^*_{m,n}(\nu=2)\right\} = N^2\left(\ln n^* + C - \frac{N^2\pi^2}{6}\right).$$

Наслідок 3.2 Величина блоку варіаційного ряду ЕПС залежить як ві обсягу вибіркової послідовності n^*, так і від кількості озна розпізнавання N, що безпосередньо випливає із наслідку 3.1.

SUMMARY

Use of extremal ordinal statistic with χ^2 distribution as one dimension statisti characteristics of sampling consequence of recognize feature has been considered. Questions o asymptotic behavior of statistics its sensitivity to features go beyond its limits and condition o expansion of units of the statistics variation row has been investigated.

СПИСОК ЛІТЕРАТУРИ

1. Краснопоясовський А.С., Марченко В.В. Прогнозування моменту перенавчання систем розпізнавання // Вісник СумДУ, 1999.-№1(12).-С.84-90.
2. Марченко В.В., Краснопоясовський А.С.Формування репрезентативної навчальн вибірки для систем контролю та діагностування// Обробка сигналів і зображень т розпізнавання образів: Праці третьої Всеукраїнської міжнародної конференції, 1996. - С.107-109.
3. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики.- М.: Наука, 1983.- 416с.

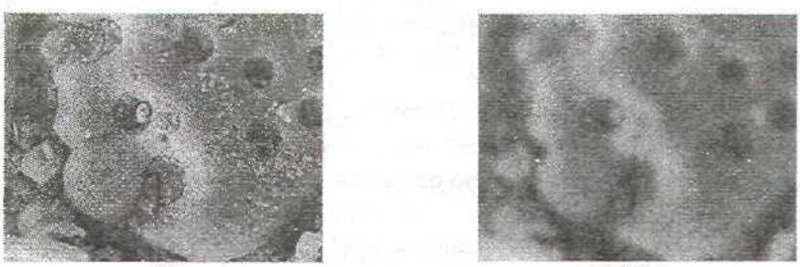
Надійшла до редколегії 28 грудня 1998 р

УДК 681.516.25:537.533.9

ОЦІНКА ФОКУСНОСТІ ЗОБРАЖЕННЯ ДЛЯ РАСТРОВОГО ЕЛЕКТРОННОГО МІКРОСКОПА

А.С.Краснопоясовський, доц.; В.В.Марченко, асп.

При розв`язанні задач автоматичної класифікації зразків, що досліджуються в електронній мікроскопії, важливого значення набуває проблема автофокусування електронного мікроскопа безпосередньо по телевізійному зображенню зразка. Розглянемо один із методів оцінки фокусної відстані електронного мікроскопа на прикладі наведених на рис.1 зображень зразка, які відображаються на екрані монітора растрового електронного мікроскопа РЕМ-103 виробництва АТ «СЕЛМІ» (м.Суми, Україна), для випадків знаходження об'єкта, що досліджується, у фокусній площині (рис. 1а) та поза нею (рис. 1б).



а) б)

Рисунок 1 – Растрове телевізійне зображення зразка: а) – об`єкт у фокусі; б) об`єкт не в фокусі

Аналіз рис.1 показує, що основна відмінність наведених на ньому зображень полягає в наявності у сфокусованого зображення більш різких переходів між темними та світлими ділянками, ніж у несфокусованого. Тому природно за градієнтний критерій фокусності растрового зображення розглядати модуль різниці $\overset{\Lambda}{\Delta}_{x,y}$ між значеннями яскравості