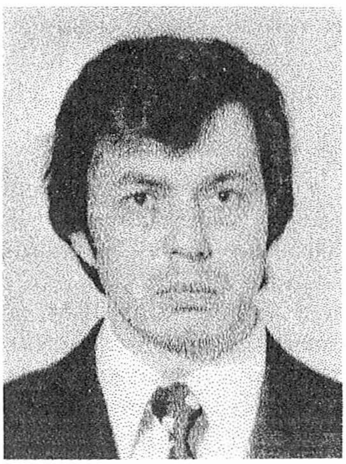
Наши лауреаты

УДК 519.28

Ю.С.ХАРИН

РОБАСТНОЕ СТАТИСТИЧЕСКОЕ РАСПОЗНАВАНИЕ ОБРАЗОВ: ТЕОРИЯ И ПРИМЕНЕНИЯ



Харин Юрий Семенович, лауреат премии А.Н.Севченко в области естественных наук за 1997 г., доктор физико-математических наук, профессор, зав. кафедрой математического моделирования и анализа данных. Область научных интересов — математическая статистика, математическая кибернетика и их практическое применение. Автор около 200 научных публикаций, в том числе двух монографий.

The paper summarizes the results of author's investigation of statistical pattern recognition problems under distortions of hypothetical models for experimental data. Results of practical use of the developed theory of robust (stable) pattern recognition in applied problems are presented.

Введение

Статистическое распознавание образов — это раздел прикладной математики, разрабатывающий вероятностно-статистические принципы и методы классификации, а также идентификации предметов, явлений, процессов, сигналов, ситуаций — всех тех объектов, которые могут быть описаны конечным набором некоторых признаков или свойств, характеризующих объект. Внимание математиков привлечено к проблемам распознавания образов из-за их огромной практической значимости в технической и медицинской диагностиках, в системах искусственного интеллекта, компьютерных системах обработки информации, автоматизации технологических процессов, военной технике, прогнозировании в экономике, финансах, бизнесе.

В классической теории статистического распознавания образов и ее практических приложениях (см., напр., [1—6]) широкое распространение получили оптимальные решающие правила, минимизирующие некоторый функционал риска (средних потерь) распознавания при соответствующем выборе функции потерь для заданной классической (гипотетической) модели экспериментальных данных. Эта модель традиционно использует такие предположения, как однородность выборки наблюдений для каждого класса, наличие измерений всех компонент вектора наблюдений (признаков), принадлежность условных вероятностных распределений признаков к заданному семейству (например, многомерному гауссовскому), статистическая независимость выборочных значений. Однако для регистрируемых на практике экспериментальных данных эти предположения часто нарушаются: присутствуют "выбросы", пропуски значений признаков, неоднородность и зависимость выборочных значений, несовпадения "фактического" вероятностного распределения с "гипотетическим". Как отмечается, например, в одном прикладном исследовании [6], при анализе 2500 выборок из архива реальных статистических данных гипотезу о нормальности пришлось отвергнуть в 92% случаев. Эти искажения гипотетической модели нарушают оптимальность классических решающих правил, приводят к неустойчивости, проявляющейся в многократном увеличении риска распознавания. В связи с этим актуальны следующие новые проблемы, которым и посвящен цикл работ автора:

а) математическое описание основных типов искажений, возникающих при решении прикладных задач распознавания образов;

б) оценивание устойчивости (робастности[1]) классических решающих правил, традиционно применяемых в статистическом распознавании;

в) оценивание δ-допустимых (критических) уровней искажений, гарантирующих заданный уровень устойчивости классических алгоритмов распознавания;

г) синтез новых робастных решающих правил, устойчивых к заданным типам искажений;

д) разработка специального программного обеспечения, реализующего алгоритмы распознавания.

Эти проблемы стимулируются развитием нового научного направления — робастность в статистике, разрабатываемого в исследованиях Д.Тьюки (США) [7], П.Хьюбера (Германия) [8], Ф.Хампеля (Швейцария) [9], С.А.Айвазяна (Россия) [6] и других ученых. Названные выше проблемы ранее рассматривались лишь в частных постановках, и для их решения в основном использовался метод статистического моделирования на ЭВМ, не позволяющий провести аналитические исследования [10—16]. В цикле работ автора, насчитывающем свыше ста публикаций и представленном здесь основными из них [17—67], впервые решена система задач а—д, и совокупность полученных результатов формирует теорию робастного статистического распознавания образов.

1. Гипотетические модели в статистическом распознавании образов

Пусть (Ω, F, P) — основное вероятностное пространство; $\mathbf{R}^N$ — пространство наблюдений (признаков) $x=(x_l)$, $l=\overline{1,N}$; $\Omega=\bigcup_{i=1}^{L}\Omega_i$ — алфавит $L\geq 2$ классов (образов, совокупностей); $\pi_i=\mathrm{P}(\Omega_i)$ — априорная вероятность класса Ω_i, $i\in S=\{1,2,\ldots,L\}$, $\pi_1+\ldots+\pi_L=1$. Наблюдение из класса Ω_i ($i\in S$) описывается случайным N-вектором $X_i\in\mathbf{R}^N$ с некоторой гипотетической условной плотностью распределения вероятностей (п.р.в.) $p_i^0(x)$, которая обычно неизвестна. Наблюдается случайная "обучающая" выборка объема n из L классов: $Z=\{z_1,\ldots,z_n\}\subset\mathbf{R}^N$. Различают две ситуации: а) выборка Z — классифицированная, при этом задано разбиение $Z=\bigcup_{i=1}^{L}Z_i$, $Z_i=\{z_{i1},\ldots,z_{in_i}\}$, где Z_i — случайная подвыборка объема n_i из Ω_i (задача распознавания образов в этом случае называется задачей дискриминантного анализа); б) выборка Z — неклассифицированная (задача распознавания образов при этом называется задачей кластер-анализа).

[1] Слово *robust* в английском языке означает *крепкий, сильный, стабильный, устойчивый*. В силу перегрузки математического термина *устойчивый* в русскоязычной литературе с 1984 г. принято использовать термин *робастный*, имея в виду устойчивость статистических выводов к нарушениям модельных предположений.

Пусть далее задана $(L\times L)$-матрица потерь $W=(w_{ij})$, где $w_{ij}\geq 0$ — величина потерь, возникающих при отнесении в класс Ω_j наблюдения, принадлежащего $\Omega_i (i,j\in S)$. Часто используют $(0-1)$-матрицу потерь: $w_{ij}=1-\delta_{ij}$, где δ_{ij} — символ Кронекера. Решающее правило (РП) для классификации произвольного наблюдения $x\in\mathbf{R}^N$ при наличии выборки $Z\in\mathbf{R}^{nN}$ задается набором L критических функций

$$\chi=\chi(x;Z)=\left(\chi_i(x;Z)\right):\mathbf{R}^N\times\mathbf{R}^{nN}\to[0,1]^L,\ \chi_i(x;Z)=\mathbf{P}\{d=i|x,Z\}, \quad (1)$$

либо борелевской функцией

$$d=d(x;Z)=\sum_{i\in S} i\mathrm{I}_{V_i}(x), d\in S, \quad (2)$$

где $\mathrm{I}_{V_i}(x)$ — индикаторная функция множества V_i, $\{V_i=V_i(Z)\}\subset\mathbf{R}^N$ — борелевское разбиение пространства наблюдений. В случае (1) имеем обобщенное рандомизированное РП, а в случае (2) — нерандомизированное РП.

Точность распознавания принято характеризовать функционалом риска (математическим ожиданием потерь) классификации:

$$r=r\left(\chi;\{p_i^0\}\right)=\sum_{i,j\in S}\pi_i w_{ij}\mathbf{E}\left\{\int_{\mathbf{R}^N}\chi_j(x;Z)p_i^0(x)dx\right\}\geq 0. \quad (3)$$

Заметим, что в случае $(0-1)$-матрицы потерь риск r имеет смысл вероятности ошибки распознавания.

Минимум функционала (3) по всевозможным РП $\chi(\cdot)$ доставляет байесовское решающее правило (БРП):

$$d=d_o(x)=\arg\min_{i\in S} f_i\left(x;\{p_j^0\}\right),\ f_i\left(x;\{p_j^0\}\right)=\sum_{j\in S}\pi_j w_{ji}p_j^0(x), \quad (4)$$

имеющее байесовский (минимальный) риск:

$$r_0=\sum_{i\in S}\pi_i r_{0i},\ r_{0i}=\sum_{j\in S} w_{ij}\int_{d_0(x)=j} p_i^0(x)dx. \quad (5)$$

Так как плотности $\{p_j^0(\cdot)\}$ априорно неизвестны, используется семейство "подстановочных" (plug-in) решающих правил:

$$d=d_1(x;Z)=\arg\min_{i\in S} f_i\left(x;\{\hat{p}_j\}\right), \quad (6)$$

где $\hat{p}_j(\cdot)$ — некоторая статистическая оценка $p_j^0(\cdot)$ по выборке Z (параметрическая или непараметрическая).

Отметим, что наиболее распространенным частным случаем описанной выше гипотетической модели является гауссовская модель Фишера:

$$p_i^0(x)=(2\pi)^{-N/2}|\Sigma|^{-1/2}\exp\left(-\left(x-\mu_i^0\right)^T\Sigma^{-1}\left(x-\mu_i^0\right)/2\right), x\in\mathbf{R}^N, \quad (7)$$

где $\mu_i^0=\left(\mu_{ij}^0\right)$ — N-вектор-столбец математических ожиданий, а $\Sigma=(\sigma_{ij})$ — невырожденная ковариационная $(N\times N)$-матрица для класса $\Omega_i (i\in S)$.

2. Типы искажений гипотетической модели и характеристики робастности

Классификация основных типов искажений гипотетической модели M_0, наиболее часто присутствующих в приложениях, представлена на рис.1. Математические описания и свойства каждого из типов искажений $M_\varepsilon\in M_+$ изложены в [17,40].

Для решения сформулированных автором задач а–д предлагается использовать следующие характеристики робастности решающих правил:

Рис.1. Классификация типов искажений

- гарантированный риск распознавания РП $d(\cdot)$:

$$r_{+n} = r_{+n}(d) = \sup_{M_\varepsilon \in M_+} r_{\varepsilon,n}(d), \tag{8}$$

где $r_{\varepsilon,n}(d)$ – риск (средние потери) распознавания (3) для РП $d(\cdot)$ при наличии искажений $M_\varepsilon \in M_+$ и обучающей выборки объема n;

- коэффициент неустойчивости риска РП $d_0(\cdot)$:

$$\kappa_{+n} = \kappa_{+n}(d_0) = \left(r_{+n}(d_0) - r_0\right)/r_0 \ge 0, \tag{9}$$

где $r_0>0$ – байесовский риск (5) для гипотетической (неискаженной) модели M_0 или асимптотический коэффициент неустойчивости: $\kappa_+ = \kappa_+(d_0) = \lim_{n\to\infty} \kappa_{+n}(d_0)$;

- δ-допустимый уровень искажений ($\delta>0$):

$$\varepsilon_+(\delta) = \inf\left\{\varepsilon_+ : \kappa_{+n}\left(d(\cdot)\right) \ge 1+\delta\right\}; \tag{10}$$

- "пороговая точка" для РП $d_0(\cdot)$:

$$\varepsilon_+^* = \min\left\{\alpha : \max_{0\le\varepsilon_+\le\alpha} r_{+n}(d_0) = r^*\right\}, \tag{11}$$

где r^* – риск, достижимый "наихудшим" РП, выносящим решение с помощью равновероятного жребия;

- робастное (относительно искажений M_+) РП $d_*=d_*(x)$, минимизирующее гарантированное значение риска:

$$r_{+n}(d_*) = \inf_{d(\cdot)} r_{+n}(d). \tag{12}$$

Чем меньше $r_{+n}(\cdot)$, $\kappa_{+n}(\cdot)$, $\kappa_+(\cdot)$ и больше ε_+^*, $\varepsilon_+(\delta)$, тем устойчивее РП $d_0(\cdot)$.

3.Робастность решающих правил к искажениям классифицируемых наблюдений

Пусть, как и ранее, $p_i^0(\cdot)$ – некоторая гипотетическая плотность распределения вероятностей наблюдений из класса Ω_i, а подлежащее классификации случайное наблюдение $X_i \in \mathbf{R}^N$ имеет искаженную плотность распределения $p_i(\cdot) \in P_i(\varepsilon_{+i})$, где $P_i(\varepsilon_{+i})$ – некоторое семейство допустимых искаженных п.р.в., $\varepsilon_{+i} \ge 0$ – уровень искажений. Если $\varepsilon_{+i}=0$, то $P_i(0)$ состоит из единственного элемента $p_i^0(\cdot)$. Различные частные случаи этого класса искажений D.2 (см. рис.1) различаются видом семейства $P_i(\varepsilon_{+i})$.

Приведем здесь результаты исследования для двух типов искажений: Тьюки–Хьюбера (D.2.2.1) и искажения в L_2-метрике (D.2.2.2). В случае искажений Тьюки–Хьюбера семейство допустимых искаженных п.р.в. имеет вид:

$$P_i\left(\varepsilon_{+i}\right) = \left\{p_i(\cdot) : p_i(x) = \left(1-\varepsilon_i\right)p_i^0(x) + \varepsilon_i h_i(x), 0 \le \varepsilon_i \le \varepsilon_{+i}\right\}, \tag{13}$$

где $h_i(\cdot)$ – произвольная неизвестная "загрязняющая" п.р.в., ε_i – неизвестная вероятность появления "выброса", $\varepsilon_{+i}<0{,}5$ – заданное наибольшее возможное значение ε_i (уровень искажения класса Ω_i).

Введем обозначения:

$$w_{i+} = \max_j w_{ij},\quad \overline{w}_+ = \sum_{i=1}^{L} \pi_i w_{i+},\quad q\left(p_i^0;\chi_j\right) = \int_{\mathbf{R}^N} p_i^0(x)\chi_j(x)dx,\quad r_{0i} = \sum_{j=1}^{L} w_{ij} q\left(p_i^0;\chi_j^0\right).$$

Теорема 1. Если имеют место искажения Тьюки–Хьюбера (13), то величина гарантированного риска (8) БРП χ^0, определяемого (1), (2), (4), равна

$$r_+\left(\chi^0\right) = r_0 + \sum_{i=1}^{L} \varepsilon_{+i}\left(w_{i+} - r_{0i}\right).$$

Следствие 1. δ-допустимый ($\delta>0$) уровень искажений (10) и "пороговая точка" (11) определяются соотношениями:

$$\varepsilon_+(\delta)=\min\left\{1,\delta r_0\Big/\sum_{i=1}^{L}\varepsilon_{+i}\left(w_{i+}-r_{0i}\right)\right\},\ \ \varepsilon_+^*=\left(r^*-r_0\right)/\left(\overline{w}_+-r_0\right).$$

Следствие 2. В случае (0–1)-матрицы потерь W

$$\varepsilon_+(\delta)=\min\left\{1,\delta r_0/\left(1-r_0\right)\right\},\ \ \varepsilon_+^*=1-\left(L\left(1-r_0\right)\right)^{-1}.$$

Теорема 2. Если имеют место искажения Тьюки–Хьюбера (13), то робастное решающее правило (12) и его гарантированный риск определяются соотношениями:

$$d_*(x)=\arg\min_{k\in S}\sum_{j=1}^{L}\pi_j\left(1-\varepsilon_{+j}\right)p_j^0(x)w_{jk},\ x\in\mathbf{R}^N;$$

$$r_+\left(d_*\right)=r_+\left(\chi^0\right)-\sum_{l,m=1}^{L}\rho_{lm}\pi_l\pi_m\varepsilon_{+l}\varepsilon_{+m}+O\left(\varepsilon_+^3\right),$$

где $\rho=(\rho_{lm})$ – известная положительно определенная симметричная матрица коэффициентов.

На рис.2 изображены графики зависимостей критического уровня искажений ε_+^* от байесовского риска r_0 и числа классов L.

Рассмотрим теперь случай искажений в L_2-метрике:

$$P\left(\varepsilon_{+i}\right)=\left\{p_i(\cdot):p_i(x)\geq 0,\ \int_{\mathbf{R}^N}p_i(x)dx=1,\ \int_{\mathbf{R}^N}\frac{\left(p_i(x)-p_i^0(x)\right)^2}{\psi_i(x)}dx=\varepsilon_i^2,0\leq\varepsilon_i\leq\varepsilon_{+i}\right\},$$

где $\psi_i(x)\geq 0$ – весовая функция, $\int_{\mathbf{R}^N}\psi_i(x)dx=1$. Если $\psi_i(\cdot)=p_i^0(\cdot)$, то получаем χ^2-метрику.

Примем обозначения: $\langle z\rangle=\left\{|z|,z<0;0,z\geq 0\right\}$;

$$\varepsilon_i^*=\sqrt{\mathrm{E}_{\psi_i}\left\{\left(\sum_{k=1}^{L}w_{ik}\left(\chi_k(x)-q\left(\psi_i;\chi_k\right)\right)\right)^2\right\}}\times$$

$$\times\left(\left\langle\inf_x\left(\frac{\psi_i(x)}{p_i^0(x)}\sum_{l=1}^{L}w_{il}\left(\chi_l(x)-q\left(\psi_i;\chi_l\right)\right)\right)\right\rangle\right)^{-1}\geq 0,i\in S,$$

где $\mathrm{E}_{\psi_i}\{\cdot\}$ – символ усреднения по распределению $\psi_i(\cdot)$.

Теорема 3. Если плотности распределения подвержены искажениям в L^2-метрике и $\varepsilon_{+i}\leq\varepsilon_i^*$ ($i\in S$), то для произвольного рандомизированного РП $\chi(\cdot)$ величина гарантированного риска (8) равна

$$r_+(x)=r\left(\chi;\left\{p_i^0\right\}\right)+\sum_{i=1}^{L}\pi_i\varepsilon_{+i}\sqrt{\mathrm{E}_{\psi_i}\left\{\left(\sum_{k=1}^{L}w_{ik}\left(\chi_k(x)-q\left(\psi_i;\chi_k\right)\right)\right)^2\right\}}.$$

Следствие 3. В случае (0–1)-матрицы потерь и χ^2-метрики для БРП $\chi^0(\cdot)$ справедливы соотношения:

$$\varepsilon_i^*=\sqrt{q\left(p_i^0;\chi_i^0\right)/\left(1-q\left(p_i^0;\chi_i^0\right)\right)},\ \ \kappa_+\left(\chi^0\right)=\frac{1}{r_0}\sum_{i=1}^{L}\pi_i\varepsilon_{+i}\sqrt{q\left(p_i^0;\chi_i^0\right)\left(1-q\left(p_i^0;\chi_i^0\right)\right)}.$$

Следствие 4. Для БРП χ^0 и гипотетической модели Фишера (7) "пороговая точка" и δ-допустимый уровень искажений имеют вид:

$$\varepsilon_+^*=\frac{\Phi(\Delta/2)-1/2}{\sqrt{\Phi(\Delta/2)\left(1-\Phi(\Delta/2)\right)}},\ \ \varepsilon_+(\delta)=\delta\sqrt{\frac{1}{\Phi(\Delta/2)}-1}.$$

где $\Phi(\cdot)$ – функция распределения стандартного нормального закона, Δ – межклассовое расстояние Махаланобиса.

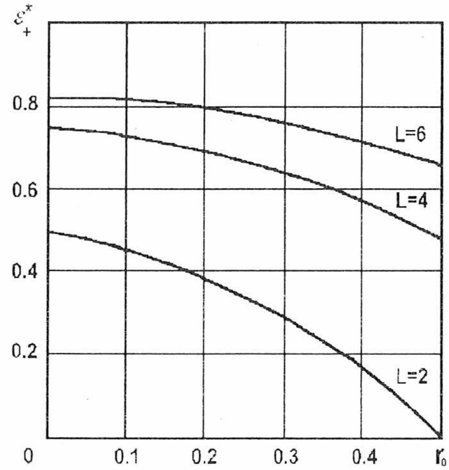


Рис.2. Графики зависимостей критического уровня искажений

Теорема 4. В условиях теоремы 3 робастное решающее правило является решением задачи вариационного исчисления:

$$\int_{R^N}\left(\sum_{j,k=1}^{L} a_{kj}(x;\chi)\chi_j(x)\chi_k(x)+\sum_{j=1}^{L} b_j(x;\chi)\chi_j(x)\right)dx\to\min_{\chi(\cdot)},$$

где для коэффициентов $\{a_{kj}, b_j\}$ получены явные выражения.

4. Робастность решающих правил к искажениям обучающих выборок

Приведем здесь результаты исследования для искажений типа D.3.3 (см. рис.1): имеют место "засорения" выборок наблюдениями из альтернативных классов.

Пусть случайная выборка $Z_i(i\in S)$, определенная в разделе 1, "засорена" наблюдениями из альтернативных классов $\bigcup_{k\neq i}\Omega_k$ и содержит (в среднем) $\varepsilon_{ik}\cdot 100\%$ наблюдений из класса $\Omega_k(\varepsilon_{i1}+\ldots+\varepsilon_{iL}=1)$. Если $\varepsilon_{ii}=1$, $\varepsilon_{ik}=0$ $(k\neq i)$, т.е. $E=(\varepsilon_{ik})=I_L$ – единичная $(L\times L)$-матрица, то "засорения" отсутствуют.

Примем следующие обозначения: $Q=\left\{q(x;\theta), x\in R^N : \theta\in\Theta\subseteq R^m\right\}$ – регулярное m-параметрическое семейство п.р.в.; $p_i^0(x)=q(x;\theta_i^0), \theta_i^0\in\Theta$,

$$\hat{p}_i(x)=q(x;\hat{\theta}_i),\quad \hat{\theta}_i=\arg\min_{\theta\in\Theta}\sum_{j=1}^{n_i} g(z_{ij};\theta) \quad - \tag{14}$$

оценка минимального контраста с функцией контраста $g(\cdot)$ (в случае $g(z;\theta)=-\ln q(z;\theta)$ имеем оценку максимального правдоподобия),

$$\varepsilon_i=1-\varepsilon_{ii}\ (i\in S);\quad \tau_0^2=\max\left\{\varepsilon_i^2,\varepsilon_i/n_i : i\in S\right\};\quad c_{ilj}=\pi_i\left(w_{il}-w_{ij}\right);$$

$$G_{ik}^{(j)}=\int_{R^N} q(x;\theta_k^0)\nabla^j_{\theta_i^0} g(x;\theta_i^0)dx\ (j=1,2;\ i,k\in S); \tag{15}$$

$$\gamma(E)=\frac{1}{2}\sum_{l\neq j}\int_{\Gamma_{lj}}\left(\sum_{i=1}^{L} c_{ilj}\sum_{k\neq i}\varepsilon_{ik}\left(\nabla_{\theta_i^0} q(x;\theta_i^0)\right)^T\left(G_{ii}^{(2)}\right)^{-1}G_{ik}^{(1)}\right)^2 b_{lj}(x)ds_{N-1}\geq 0.$$

Теорема 5. Если поверхностные интегралы на байесовских дискриминантных поверхностях $\Gamma_{lj}=\left\{x\in R^N : f_l\left(x;\{\theta_i^0\}\right)-f_j\left(x;\{\theta_i^0\}\right)=0\right\}$ $(l\neq j)$ в (15) конечны, то при наличии искажений, определенных матрицей ошибочной классификации E, риск подстановочного РП $d=d_1(x;Z)$ (6) с оценками минимального контраста (14) допускает асимптотическое разложение:

$$r(d_1)=r_0+\sum_{i=1}^{L}\frac{\rho_{i0}}{n_i}+\gamma(E)+\sum_{i=1}^{L}\sum_{k\neq i}\beta_{ik}\frac{\varepsilon_{ik}}{n_i}+o(\tau_0^2),$$

где для коэффициентов $\{\rho_{i0},\beta_{ik}\}$ и квадратичной формы $\gamma(E)$ получены явные выражения.

Теорема 6. Робастное решающее правило $d=d_*(x;Z)$, использующее вместо $\{\hat{\theta}_i\}$ новые оценки $\tilde{\theta}_i=\hat{\theta}_i+\left(\hat{G}_{ii}^{(2)}\right)^{-1}\sum_{j\neq i}\varepsilon_{ij}\hat{G}_{ij}^{(1)}\ \left(i=\overline{1,L}\right)$, имеет повышенный порядок робастности по отношению к $\varepsilon_+=\max_i\varepsilon_i$:

$$\kappa(d_*)=\sum_{i=1}^{L}\frac{\rho_{i0}}{n_i}+o(\tau_0^2)=O(\varepsilon_+^3),\ \kappa(d_1)=O(\varepsilon_+^2).$$

Следствие 5. В случае гауссовской гипотетической модели (7) при $L=2$ и (0–1)-матрице потерь

$$\kappa(d_1)-\kappa(d_*)\approx\frac{\Delta^3\exp(-\Delta^2/8)}{16\sqrt{2\pi}\Phi(-\Delta/2)}\left((\varepsilon_2-\varepsilon_1)^2+\frac{\varepsilon_1}{n_1}+\frac{\varepsilon_2}{n_2}\right).$$

5.Применения

Практические применения результатов данного цикла работ осуществлялись в двух основных направлениях: 1) разработка пакетов прикладных программ, реализующих методы теории робастного статистического распознавания образов, и внедрение этих пакетов в различных организациях для решения разнообразных задач; 2) решение прикладных задач с использованием разработанной теории в рамках госбюджетных, хоздоговорных НИР и договоров о научно-техническом сотрудничестве.

В рамках первого направления на кафедре математического моделирования и анализа данных разработана система статистического программного обеспечения, состоящая из шести основных ППП: *СТАН* – базовый пакет статистического анализа данных; *МУЛЬТИСТАН* – пакет прикладных программ многомерного статистического анализа данных; *ДИНСТАТ* – пакет прикладных программ статистического анализа временных рядов; *СТАТМОД* – пакет прикладных программ статистического моделирования; *СТАТУПР* – пакет прикладных программ статистического анализа, регулирования и контроля качества; *РОСТАН* – пакет прикладных программ робастного статистического анализа данных. По запросу Мюнхенского университета (Германия) разработана англоязычная версия *ППП РОСТАН*, которая используется при обработке экспериментальных данных в научных исследованиях. Названные пакеты прикладных программ внедрены более чем в 30 организациях СНГ, в том числе: НПО "Интеграл", НПО "Агат", ПО "Беларуськалий", Министерство внешнеэкономических связей, Высшее военное командное училище, Полоцкий госуниверситет, Гродненский госуниверситет, Академия физвоспитания и спорта, НИИ онкологии и медрадиологии, Гродненский мединститут, Республиканская станция переливания крови, Институт радиоэкологических проблем НАН Беларуси, Национальный банк РБ, АКБ "Приорбанк", Российская транспортная компания "AIR TSOLAK Ltd".

На базе указанных пакетов прикладных программ разработана система из четырех компьютерных учебников *TEACHSTAT* в области прикладной статистики: *СТАН-У*, *МУЛЬТИСТАН-У*, *ДИНСТАТ-У*, *СТАТМОД-У*. Эта система внедрена в учебный процесс в Белгосуниверситете, в ряде других вузов Республики Беларусь, а также с 1996 г. используется в Московском госуниверситете при изучении дисциплины “Математическая и прикладная статистика”. Во всех этих программных разработках автор статьи участвовал как научный руководитель и создатель ряда алгоритмов, реализующих робастные методы.

В рамках второго направления теория робастного статистического распознавания образов использована при решении ряда важнейших прикладных задач, в том числе: компьютерная диагностика основных форм злокачественных новообразований по биохимическим и биофизическим характеристикам крови (по заказу НИИ онкологии и медрадиологии Министерства здравоохранения РБ); эконометрическое моделирование и прогнозирование динамики важнейших макроэкономических параметров (по заказу Министерства экономики РБ); статистическое моделирование и регулирование технологических процессов в микроэлектронике (по заказу НПО "Интеграл"); разработка математического и программного обеспечения защиты информации.

Результаты данного цикла работ использовались при выполнении важнейших государственных научно-технических программ, в том числе “Информатика”, “Информатизация”, “Защита информации”, “Экономика и социальная политика”. Исследования выполнялись и по международным проектам с Калифорнийским госуниверситетом (США) и по программе INTAS.

Автор глубоко признателен коллективу Белорусского Государственного университета за поддержку и помощь при выполнения этого цикла работ.

1. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. М., 1974.
2. Ту Дж., Гонсалес Р. Принципы распознавания образов. М., 1979.
3. Фукунага М. Введение в статистическую теорию распознавания образов. М., 1979.
4. Патрик Э. Основы теории распознавания образов. М., 1980.
5. Дуда Р., Харт П. Распознавание образов и анализ сцен. М., 1983.
6. Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Классификация и снижение размерности. М., 1989.
7. Tukey J.W. // Contributions to Probability and Statistics. Stanford, 1960.
8. Хьюбер П. Робастность в статистике. М., 1984.
9. Хампель Ф., Рончетти Е.М., Рауссеу П.Д. и др. Робастность в статистике. М., 1989.
10. Berger J. Statistical Decision Theory and Bayesian Analysis. N.Y., 1985.
11. Kadane J., Chuang D. // Annals of Statistics. 1978. Vol.6.
12. Kazakos D. // IEEE Trans. Information Theory. 1982. Vol.28. №5.
13. Lachenbruch P. // Technometrics. 1979. Vol.21.
14. Lawoko C., McLachlan G. // Pattern Recognition. 1988. Vol.21.
15. McLachlan G. Discriminant Analysis and Statistical Pattern Recognition. N.Y., 1992.
16. Tiku M., Tan W.Y., Balakrishnan N. Robust Inference. N.Y., 1986.
17. Kharin Yu. Robustness in Statistical Pattern Recognition. Dordrecht; Boston; London, 1996.
18. Id. // Lecture Notes in Computer Science. 1997. Vol.1280.
19. Id. // Proc. of the IV European Congress on Intelligent Techniques and Soft Computing. 1997. Vol.3.
20. Харин Ю.С., Жук Е.Е. // Весці АН Беларусі. Сер. фіз.-мат. навук. 1997. №3.
21. Харин Ю.С. // Там же. №4.
22. Kharin Yu. // Proc. of the Int. Conf. BIT&FT. 1997.
23. Id. // Statistical Techniques in Pattern Recognition (Ed. P. Pudil). 1997.
24. Id. // Pattern recognition and information processing. 1997. Vol.1.
25. Id. // Lecture Notes in Statistics. 1996. Vol.109.
26. Id. // Proc. of the Second Int. Conf. NITE'96. 1996. Vol.1.
27. Id. // Proc. of the IV World Congress of the Bernoulli Society. 1996.
28. Харин Ю.С., Ушакова В.М. // Адукацыя і выхаванне. 1996. №4.
29. Kharin Yu., Zhuk E. // Control and Cybernetics. 1995. Vol.24. №4.
30. Kharin Yu. // Proc. of the V Int. Prague Symp. on Asymptotic Statistics. N.Y., 1995.
31. Id. // Computer Data Analysis & Modeling. 1995. Vol.1. C.25.
32. Компьютерный анализ данных и моделирование: Сб. научн. ст. / Под ред. Ю.С.Харина. Мн., 1995. Т.1,2.

33. Kharin Yu., Zhuk E. // Proc. of the XII Int. Conf. on Pattern Recognition. Washington, 1994. Vol.2.
34. Kharin Yu. // Information theory and statistical decision functions. Prague, 1994.
35. Id. // Knowledge-based systems. 1994. Vol.7. №1.
36. Kharin Yu. et al. Robust Statistical Analysis (ROSTAN) Programs: User Manual. Minsk, 1994.
37. Харин Ю.С., Степанова М.Д. // Обозрение прикладной и промышленной математики. 1994. Т.1. Вып.2.
38. Kharin Yu., Zhuk E. // Information and Classification. N.Y., 1993.
39. Современные проблемы компьютерного анализа данных и моделирования: Сб. научн. ст. / Под ред. Ю.С.Харина. Мн., 1993.
40. Харин Ю.С. Робастность в статистическом распознавании образов. Мн., 1992.
41. Он же. // Вестн. МГУ. Сер. Математика. 1992. №6.
42. Харин Ю.С., Васенкова Е.И. // Весці АН Беларусі. Сер. фіз.-мат. навук. 1992. №3.
43. Харин Ю.С. и др. Имитационное и статистическое моделирование: Учеб. пособие. Мн., 1992.
44. Харин Ю.С. // Отбор и обработка информации. 1991. Т.81.
45. Харин Ю.С., Мельникова Е.Н. // Автоматика и телемеханика. 1991. №12.
46. Kharin Yu., Medvedev A. // Prob. Theory & Math. Statistics. 1991.
47. Харин Ю.С. // Вестн. Белорус. ун-та. Сер.1. 1991. №3.
48. Проблемы компьютерного анализа данных и моделирования: Сб. научн. ст. / Под. ред. Ю.С.Харина. Мн., 1991.
49. Харин Ю.С. // Заводская лаборатория. 1991. №10.
50. Он же. // Уч. зап. по статистике. 1990. Т.54.
51. Он же. // Статистические проблемы управления. 1990. Вып. 89.
52. Харин Ю.С., Степанова М.Д. Практикум на ЭВМ по математической статистике: Учеб. пособие. Мн., 1987.
53. Харин Ю.С., Малюгин В.И. // Автоматика и телемеханика. 1986. №7.
54. Kharin Yu. // Detection of change in random processes. N.Y., 1986.
55. Харин Ю.С. // Автоматика. 1986. №4.
56. Он же. // Уч. зап. по статистике. 1985. Т.49.
57. Он же. // Проблемы передачи информации. 1985. Т.21. №4.
58. Kharin Yu. // Proc. of the III Prague Symp. on Asymptot. Statistics. Amsterdam, 1984.
59. Id. // Robustness of statistical methods and nonparametric statistics. Berlin, 1984.
60. Id. // Information theory and statistical decision functions. Prague, 1983.
61. Харин Ю.С. // Автоматика и телемеханика. 1983. №11.
62. Он же. // Теория вероятностей и ее применения. 1983. Т.28. №3.
63. Он же. // Автоматика и телемеханика. 1983. №1.
64. Он же. // Адаптация и обучение в системах управления и принятия решений. М., 1982.
65. Он же. // Автоматика и телемеханика. 1982. №11.
66. Он же. // Теория вероятностей и ее применения. 1982. Т.27. №4.
67. Харин Ю.С. и др. Основы имитационного и статистического моделирования: Уч. пособие. Мн., 1997.

Поступила в редакцию 05.12.97.