ISSN 1392 - 1215 ELEKTRONIKA IR ELEKTROTECHNIKA. 2005. Nr. 5(61)

T 170 ELEKTRONIKA

Kineskopų klasifikavimas taikant matematinės statistikos metodus

J. Kruopis, R. Levulienė

Matematikos ir informatikos fakultetas, Vilniaus universitetas, Naugarduko g. 24, LT-03225 Vilnius, Lietuva, tel. +370 5 2136390, el. p.: julius.kruopis@maf.vu.lt, ruta.levuliene@maf.vu.lt

S. Grobovas, A. Vaišvila

AB „Ekranas”, Elektronikos g. 1, LT-35116 Panevėžys, Lietuva, tel. +370 45 506507, faksas +370 45 436563, el. p. grobovas@ekranas.lt tel. +370 45 506766, el. p. vaisvila@ekranas.lt

Įvadas

Objektus reikia klasifikuoti daugelyje mokslinės ir praktinės veiklos sričių. Tokie uždaviniai nuolat sprendžiami, pavyzdžiui, medicinoje, biologijoje, meteorologijoje, gamyboje ir pan. Jie priskirtini matematikos skyriams, vadinamiems diskriminantine analize, klasterizavimu, vaizdų atpažinimu, klasifikavimu.

Klasifikavimas nuolat taikomas gamyboje. Priimamosios bei išleidžiamosios kontrolės, taip pat tarpiniuose punktuose žaliavos, detalės, mazgai, gaminiai yra skirstomi į gerus ar blogus, remontuotinus ar utilizuotinus ir pan. Taip pat nuolat kontroliuojami technologiniai procesai ir jų režimai ir priimami sprendimai dėl jų normalaus funkcionavimo arba išsiderinimo.

Pagal nusistovėjusią praktiką sprendimai apie objektų priskyrimą vienai ar kitai kategorijai įvairiuose kontrolės punktuose paprastai atliekami lyginant gautuosius parametrų matavimus su leistinomis ribomis, nurodytomis techninėje dokumentacijoje. Sprendimai būtų tikslesni, jeigu jie būtų grindžiami visais parametrų matavimais, gautais pastarajame ir ankstesniuose kontrolės punktuose. Ypač tuo atveju, kai gaminiai yra sudėtingi, o parametrai priklausomi. Šiuolaikiniai kompiuterių tinklai ir automatizuotos duomenų įrašymo ir perdavimo sistemos sudaro galimybę realiu laiko masteliu priimant sprendimą bet kuriame kontrolės punkte panaudoti informaciją iš ankstesnių kontrolės punktų. Apie apriorinės informacijos panaudojimą priimant sprendimus žr., pavyzdžiui, [1], [2].

Spalvotųjų kineskopų gamykloje „Ekranas“, pradedant išsiurbimo operacija, kiekvienam kineskopui priskiriamas identifikacinis numeris. Atliekant tolesnes technologines ir kontrolės operacijas duomenų bazėse fiksuojama kiekvieno kineskopo istorija, t. y. technologinių režimų, kuriais jis buvo pagamintas, charakteristikos, kineskopo parametrų kontrolinėse operacijose vertės bei rodikliai, apibūdinantys jo veikimą pas vartotoją. Duomenų bazėse saugoma keleto pastarųjų metų informacija.

Lyginant reklamuotų ir nereklamuotų kineskopų statistinių duomenų masyvus, logistinės regresijos metodu galima sukonstruoti klasifikavimo taisyklę, kuri, remdamasi sukaupta informacija, įgalintų atskirti kineskopų, kurie potencialiai turi daugiau šansų būti reklamuoti, srautą. Tada pagrindiniame vartotojui siunčiamame kineskopų sraute reklamuotų kineskopų dalis turėtų gerokai sumažėti.

Šiame darbe pateikiamas logistinės regresijos modelis ir aptariamos jo taikymo objektams klasifikuoti galimybės (detaliau žr. [3], [4], [5]). Šis klasifikavimo metodas pritaikytas spalvotiesiems kineskopams klasifikuoti pagal gamykloje „Ekranas“ sukauptą informaciją. Aptariamas klasifikavimo tikslumas ir jo įdiegimo galimybės.

Logistinės regresinės analizės matematinis modelis

Trumpai pateiksime naudojamos klasifikavimo taisyklės matematinį pagrindimą. Vienas iš dažniausiai taikomų metodų klasifikavimo taisyklei rasti yra vadinamoji logistinė regresija.

Tarkime, kad norime prognozuoti kintamąjį Y, įgyjantį dvi vertes 0 ir 1, remdamiesi atsitiktinio vektoriaus $\mathbf{X} = (X_1,\dots,X_k)$ matavimais. Imame $Y=1$, jeigu kineskopas buvo reklamuotas, ir $Y=0$, jeigu kineskopas reklamuotas nebuvo. Jeigu Y skirstinys priklauso nuo $\mathbf{X}$, tai sąlyginės tikimybės

$$\pi(\mathbf{x}) = \mathbf{P}\{Y=1 \mid \mathbf{X}=\mathbf{x}\}, \quad 1-\pi(\mathbf{x}) = \mathbf{P}\{Y=0 \mid \mathbf{X}=\mathbf{x}\},$$

yra vektoriaus $\mathbf{x} = (x_1,\dots,x_k)$ funkcijos. Kitaip tariant, a.d. Y skirstinys, kai $\mathbf{X}=\mathbf{x}$, yra binominis $B(1,\pi(\mathbf{x}))$.

Šį skirstinį galima užrašyti taip:

$$\mathbf{P}\{Y = y \mid \mathbf{X} = \mathbf{x}\} = [\pi(\mathbf{x})]^{y}[1-\pi(\mathbf{x})]^{1-y},\ y = 0{,}1;$$
$$\mathbf{E}(Y \mid \mathbf{X} = \mathbf{x}) = \pi(\mathbf{x}) = \pi(x_1,\dots,x_k). \qquad (1)$$

Atliekant regresinę analizę prognozuojamo dydžio vidurkis, kai $\mathbf{X}$ vertė yra fiksuota ir lygi $\mathbf{x}$, aprašomas kokia nors $\mathbf{x}$ funkcija (dažniausiai tiesine). Kadangi nagrinėjamu atveju vidurkis $\pi(\mathbf{x})$ turi būti tarp 0 ir 1, tai šis būdas netinka.

Logistinėje regresijoje naudojamas toks funkcijos $\pi(\mathbf{x})$ priklausomybės nuo $\mathbf{x}$ modelis:

$$\pi(\mathbf{x}) = \frac{e^{Z(\mathbf{x})}}{1+e^{Z(\mathbf{x})}},\ Z(\mathbf{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k. \qquad (2)$$

Šiame modelyje visais atvejais $0 \le \pi(\mathbf{x}) \le 1$; kai $Z(\mathbf{x}) \to +\infty$, tai $\pi(\mathbf{x}) \to 1$; kai $Z(\mathbf{x}) \to -\infty$, tai $\pi(\mathbf{x}) \to 0$.

Sudarykime santykį

$$\gamma(\mathbf{x}) = \frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})} = e^{Z(\mathbf{x})}, \qquad (3)$$

kurį vadiname įvykio $\{Y=1\}$ *galimybės santykiu* , kai $\mathbf{X} = \mathbf{x}$. Jį galima interpretuoti taip: jeigu, pavyzdžiui, $\gamma(\mathbf{x}) = 5$, tai reiškia, kad taške $\mathbf{X} = \mathbf{x}$ galimybė a.d. Y įgyti vertę 1 (kineskopui būti reklamuotam), o ne 0 (būti nereklamuotam) vertinama santykiu 5:1 (palyginkite su bukmekerių vartojamais terminais).

Galimybės santykį $\gamma(\mathbf{x})$ galima panaudoti klasifikavimo taisyklei sudaryti. Jeigu kurio nors kineskopo $\gamma(\mathbf{x}) > 1$, tai jis priskiriamas „įtartinų" kategorijai, jeigu $\gamma(\mathbf{x}) \le 1$, tai jis priskiriamas prie gerų.

Nelygybė $\gamma(\mathbf{x}) > 1$ yra ekvivalenti nelygybėms

$$\pi(\mathbf{x}) > \frac{1}{2}, \quad e^{Z(\mathbf{x})} > 1, \quad Z(\mathbf{x}) > \ln 1 = 0. \qquad (4)$$

Taigi klasifikavimo taisyklę galima suformuluoti taip:

- kineskopas priskiriamas „įtartinų" kategorijai, jeigu $Z(\mathbf{x}) > 0$;
- kineskopas priskiriamas gerų kategorijai, jeigu $Z(\mathbf{x}) \le 0$.

P a s t a b a. Atsižvelgiant į praktikos poreikius, klasifikavimo ribą galima keisti. Pavyzdžiui, gali būti pageidautina, kad į „įtartinų" kineskopų kategoriją patektų, tarkim, ne daugiau kaip 1 procentas visos produkcijos. Tada klasifikavimo taisyklę (4) galima modifikuoti taip: „įtartinų" kategorijai priskirsime tuos kineskopus, kuriems galimybės santykis $\gamma(\mathbf{x})$ yra, pavyzdžiui, 10:1. Taigi bendruoju atveju klasifikavimo taisyklę galima suformuluoti taip:

- kineskopas priskiriamas „įtartinų" kategorijai, kai

$$\pi(\mathbf{x}) > c, \quad e^{Z(\mathbf{x})} > \frac{c}{1-c}, \quad Z(\mathbf{x}) - \ln\frac{c}{1-c} > 0. \qquad (5)$$

Nagrinėkime dvi a.v. $\mathbf{x}$ vertes $\mathbf{x}'$ ir $\mathbf{x}''$:

$$\mathbf{x}' = (x_1,\dots,x_i+1,\dots,x_k), \quad \mathbf{x}'' = (x_1,\dots,x_i,\dots,x_k),$$

kurios skiriasi tik tuo, kad pirmajame i-toji koordinatė yra padidinta 1, o visos kitos koordinatės paliktos tos pačios. Apskaičiuokime galimybės santykio pokytį pereinant nuo taško $\mathbf{x}'$ prie taško $\mathbf{x}''$:

$$\frac{\gamma(\mathbf{x}')}{\gamma(\mathbf{x}'')} = \frac{e^{Z(\mathbf{x}')}}{e^{Z(\mathbf{x}'')}} = e^{Z(\mathbf{x}')-Z(\mathbf{x}'')} = e^{\beta_i}. \qquad (6)$$

Šis dydis parodo, kiek kartų pakinta (padidėja ar sumažėja) galimybės santykis, kai a.v. $\mathbf{x}$ i-toji koordinatė sumažėja vienetu, o visos kitos koordinatės nepasikeičia.

Logistinės regresijos statistiniai uždaviniai

Modelyje (2) parametrai $\boldsymbol{\beta} = (\beta_0, \beta_1,\dots,\beta_k)$ yra nežinomi ir nustatomi iš statistinių duomenų.

Tarkime, kad turime tokius statistinius duomenis:

$$(Y_i, \mathbf{X}^{(i)}) = (Y_i, X_1^{(i)},\dots,X_k^{(i)}),\ i = 1,2,\dots,N_0+N_1; \qquad (7)$$

čia Y_i įgyja vertę 0, jeigu *i*-tasis kineskopas nebuvo reklamuotas, ir vertę 1, jeigu *i*-tasis kineskopas buvo reklamuotas; o $\mathbf{X}^{(i)} = (X_1^{(i)},\dots,X_k^{(i)}),\ i = 1,2,\dots,N_0+N_1$, yra *i*-tojo kineskopo parametrų vertės; N_0 - nereklamuotų kineskopų skaičius, o N_1 - dėl MSS (mažos spindulio srovės) reklamuotų kineskopų skaičius.

Sudarome *maksimalaus tikėtinumo funkciją* $L(\boldsymbol{\beta})$, kuri gaunama paeiliui įrašant į skirstinį (1) duomenis (7) ir sudauginant:

$$L(\boldsymbol{\beta}) = L(\boldsymbol{\beta} \mid (Y_i, \mathbf{X}^{(i)}), i = 1,2,\dots,N_0+N_1) =$$
$$= \prod_{i=1}^{N_0+N_1} [\pi(\mathbf{X}^{(i)})]^{Y_i}[1-\pi(X^{(i)})]^{1-Y_i}. \qquad (8)$$

Jos prasmė yra tokia: $L(\boldsymbol{\beta})$ reiškia tikimybę gauti duomenis (7), jeigu modelis yra (1).

Parametrų $\beta_0, \beta_1,\dots,\beta_k$ įverčius $\hat\beta_0, \hat\beta_1,\dots,\hat\beta_k$ randame *maksimalaus tikėtinumo metodu* , t. y. iš sąlygos

$$\ln L(\boldsymbol{\beta}) \to \max. \qquad (9)$$

Diferencijuodami $\ln L(\boldsymbol{\beta})$ pagal $\beta_0, \beta_1,\dots,\beta_k$, įverčiams rasti gauname *maksimalaus tikėtinumo lygčių sistemą*

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\beta_j} = \sum_{i=1}^{N_0+N_1} X_j^{(i)}[Y_i - \pi(\mathbf{X}^{(i)})] = 0,\ j = 0,1,\dots,k. \qquad (10)$$

Tai netiesinių $k+1$ lygčių su $k+1$ nežinomuoju sistema, kurią reikia spręsti iteracijų metodais. Tokias lygčių sistemas numatyta spręsti daugelyje matematinės statistikos taikomųjų programų paketų.

Gavę parametrų $\beta_0, \beta_1,\dots,\beta_k$ įverčius $\hat\beta_0, \hat\beta_1,\dots,\hat\beta_k$ galime įvertinti tikimybę (2)

$$\hat\pi(\mathbf{x}) = \frac{e^{\hat Z(\mathbf{x})}}{1+e^{\hat Z(\mathbf{x})}}, \hat Z(\mathbf{x}) = \hat\beta_0 + \hat\beta_1 x_1 + \dots + \hat\beta_k x_k,$$

galimybės santykį (3)

$$\hat{\gamma}(\mathbf{x}) = \frac{\hat{\pi}(\mathbf{x})}{1-\hat{\pi}(\mathbf{x})} = e^{\hat{Z}(\mathbf{x})},$$

galimybės santykio pokytį (6)

$$\frac{\hat{\gamma}(\mathbf{x}')}{\hat{\gamma}(\mathbf{x}'')} = \frac{e^{\hat{Z}(\mathbf{x}')}}{e^{\hat{Z}(\mathbf{x}'')}} = e^{\hat{Z}(\mathbf{x}')-\hat{Z}(\mathbf{x}'')} = e^{\hat{\beta}_i}$$

ir klasifikatoriaus funkciją (5)

$$\hat{Z}(\mathbf{x}) - \ln\frac{c}{1-c} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k - \ln\frac{c}{1-c} > 0. \quad (11)$$

Atliekant logistinę regresinę analizę, sprendžiami ir kiti uždaviniai, analogiški įprastinei regresijai: įverčių tikslumo analizė; hipotezių apie parametrų vertes tikrinimas; suderinamumo hipotezių apie modelio tinkamumą turimiems duomenims aprašyti tikrinimas ir pan. (Detaliau apie logistinę regresiją žr.[3],[4]). Reikia pasakyti, kad atliekant skaičiavimus su matematinės statistikos paketais SAS, SPSS, STATISTIKA ir pan., juose yra pateikiami gana išsamūs spausdinamų charakteristikų aprašymai.

Pateiksime kriterijų hipotezei $H: \beta_1 = ... = \beta_k = 0$ tikrinti. Ši hipotezė ekvivalenti tvirtinimui, kad $\pi(\mathbf{x}) = \mathbf{P}\{Y=1 \mid \mathbf{X}=\mathbf{x}\} \equiv \pi, \forall \mathbf{x}$, t.y. kad Y skirstinys nuo parametrų vektoriaus $\mathbf{X}$ nepriklauso ir Y prognozavimas pagal $\mathbf{X}$ neturi prasmės.

Pažymėkime

$$\hat{Y}_i = \hat{\pi}(\mathbf{X}^{(i)}) = \frac{e^{\hat{Z}(\mathbf{X}^{(i)})}}{1+e^{\hat{Z}(\mathbf{X}^{(i)})}}, i = 1,2,..., N_0 + N_1$$

vertės Y_i prognozę, kai modelis (1) yra teisingas. Jeigu teisinga hipotezė H, tai tikimybės π įvertis yra lygus aritmetiniam vidurkiui :

$$\hat{\pi} = \overline{Y} = \frac{1}{N_0 + N_1}\sum_{i=1}^{N_0+N_1} Y_i .$$

Maksimalaus tikėtinumo funkcijos $L(\boldsymbol{\beta})$ maksimumas, kai teisingas modelis (1), yra

$$L(\hat{\boldsymbol{\beta}}) = \prod_{i=1}^{N_0+N_1} \hat{Y}_i^{Y_i}(1-\hat{Y}_i)^{1-Y_i}, \quad (12)$$

o kai teisinga hipotezė H,

$$L_1(\hat{\pi}) = \prod_{i=1}^{N_0+N_1} \overline{Y}^{Y_i}(1-\overline{Y})^{1-Y_i}. \quad (13)$$

Jeigu hipotezė H yra teisinga, tai $\hat{Y}_i$ ir $\overline{Y}$ yra artimi, o kartu yra artimos funkcijų (12) ir (13) vertės. Įrodyta, kad esant dideliems stebėjimų skaičiams, santykio

$$D_E = -\ln\frac{L_1(\hat{\pi})}{L(\hat{\boldsymbol{\beta}})} \quad (14)$$

skirstinys yra artimas chi kvadrato skirstiniui su k laisvės laipsnių. Todėl hipotezė H yra atmetama, kai

$$D_E > \chi^2_\alpha(k);$$

čia $\chi^2_\alpha(k)$ yra chi kvadrato su k laisvės laipsnių α kritinė vertė.

Nagrinėkime hipotezę

$$H_{i_1,...,i_l}: \beta_{i_1} = ... = \beta_{i_l} = 0, i_1 \neq i_2 \neq ... \neq i_l, 1 \leq l < k$$

apie tai, kad parametrai $X_{i_1},..., X_{i_l}$ neturi įtakos prognozės tikslumui. Pažymėkime $D_E^{(k)} = D_E$ ir $D_E^{(k-l)}$ – statistiką (14) modeliui su visais parametrais $\beta_0, \beta_1,..., \beta_k$ ir modeliui, kai parametrai $\beta_{i_1},..., \beta_{i_l}$ yra praleisti. Tada skirtumo $D_E^{(k)} - D_E^{(k-l)}$ skirstinys aproksimuojamas chi kvadrato skirstiniu su $k-(k-l) = l$ laisvės laipsnių. Hipotezė $H_{i_1,...,i_l}$ yra atmetama, kai

$$D_E^{(k)} - D_E^{(k-l)} > \chi^2_\alpha(l). \quad (15)$$

Parametrų β_j įverčių $\hat{\beta}_j$ tikslumui apibūdinti apskaičiuojame Fišerio informacinę matricą $I(\boldsymbol{\beta})$ su elementais

$$I_{ls}(\boldsymbol{\beta}) = -\mathbf{E}\frac{\partial^2 \ln L(\boldsymbol{\beta})}{\partial\beta_l \partial\beta_s} =$$

$$= \sum_{i=1}^{N_0+N_1} X_l^{(i)} X_s^{(i)} \pi(\mathbf{X}^{(i)})(1-\pi(\mathbf{X}^{(i)}))$$

ir jos įvertį $I(\hat{\boldsymbol{\beta}})$ su elementais

$$I_{ls}(\hat{\boldsymbol{\beta}}) = \sum_{i=1}^{N_0+N_1} X_l^{(i)} X_s^{(i)} \hat{\pi}(\mathbf{X}^{(i)})(1-\hat{\pi}(\mathbf{X}^{(i)})).$$

Tada atsitiktinio vektoriaus $\hat{\boldsymbol{\beta}}$ skirstinys yra aproksimuojamas k-mačiu normaliuoju:

$$\hat{\boldsymbol{\beta}} \sim N_{k+1}(\boldsymbol{\beta}, I^{-1}(\hat{\boldsymbol{\beta}})). \quad (16)$$

Atskiros vektoriaus $\hat{\boldsymbol{\beta}}$ koordinatės $\hat{\beta}_s$ skirstinys aproksimuojamas vienmačiu normaliuoju $N(\beta_s, \hat{\sigma}_{ss})$ arba

$$\frac{\hat{\beta}_s - \beta_s}{\sqrt{\hat{\sigma}_{ss}}} \sim N(0,1), \quad (17)$$

čia $\hat{\sigma}_{sl}$; kai $s,l = 0,1,...,k$, yra matricos $I^{-1}(\hat{\boldsymbol{\beta}})$ elementai.

Taigi hipotezė $H_s: \beta_s = 0$ yra atmetama, kai

$$\frac{|\hat{\beta}_s|}{\sqrt{\hat{\sigma}_{ss}}} > z_\alpha;$$

čia z_α – standartinio normalaus skirstinio α kritinė vertė. Pažymėtina, kad šią hipotezę galima buvo tikrinti ir kriterijumi (15).

Naudojant aproksimacijas (16) arba (17) galima sudaryti parametrų $\beta_s, s = 0,1,...,k$, arba jų funkcijų apytikslius pasikliautinuosius intervalus. Pavyzdžiui, iš (17) parametrui β_s gauname pasikliautinąjį intervalą

$$(\hat{\beta}_s - z_\alpha \sqrt{\hat{\sigma}_{ss}}, \hat{\beta}_s + z_\alpha \sqrt{\hat{\sigma}_{ss}}),$$

kurio pasikliovimo lygmuo $1-2\alpha$.

Atliekant skaičiavimus pagal standartinius matematinės statistikos TPP, paprastai yra pateikiami ne tik parametrų įverčiai, bet ir jų pasikliautinieji intervalai.

Klasifikavimo tikslumas

Modelio tinkamumą objektams klasifikuoti galima apibūdinti klasifikavimo lentele, kuri parodo, kaip tiksliai sukonstruotas klasifikatorius atskiria gerus ir blogus gaminius.

Imkime atsitiktinį dydį η, kuris įgyja vertę 0, jeigu gaminys priskirtas blogų kategorijai (t.y. galioja nelygybė (11)) , ir įgyja vertę 1, – priešingu atveju. Apibrėžkime sąlygines tikimybes

$$\alpha_{ij} = \mathbf{P}\{\eta = i \mid Y = j\}, \quad i, j = 0{,}1, \tag{18}$$

kurias galime surašyti į lentelę

1 lentelė. Sąlyginės tikimybės

η / Y	0	1	Σ
0	α_{00}	α_{10}	1
1	α_{01}	α_{11}	1

Šioje lentelėje α_{00} ir α_{11} yra teisingi sprendimai, o α_{10} ir α_{01} - klaidingi. Klasifikavimo taisyklė tuo geresnė, kuo didesnės tikimybės α_{00}, α_{11} ir kuo mažesnės tikimybės α_{10} ir α_{01}. Praktikoje svarbesnės yra atvirkštinės tikimybės

$$\beta_{ji} = \mathbf{P}\{Y = j \mid \eta = i\} = \frac{\alpha_{ij}\omega_j}{\alpha_{i1}\omega_1 + \alpha_{i0}\omega_0}, \text{ kai } \quad i, j = 0{,}1, \tag{19}$$

kurios parodo gaminių srautų, gautų atlikus klasifikavimą, užterštumą kitos grupės objektais; (19) formulėje ω_0 ir ω_1 yra apriorinės tikimybės: $\omega_0 = \mathbf{P}\{Y = 0\}$, $\omega_1 = \mathbf{P}\{Y = 1\}$, $\omega_0 + \omega_1 = 1$.

Blogais pripažintų kineskopų srauto dydis

$$Q = \mathbf{P}\{\eta = 1\} = \alpha_{10}\omega_0 + \alpha_{11}\omega_1. \tag{20}$$

Yra žinoma, kad reklamuotų kineskopų dalis ω_1 yra tūkstantųjų eilės. Todėl (20) formulėje ir (19) formulės vardiklyje įrašę $\omega_1 = 0$, $\omega_0 = 1$, gausime apytiksles formules

$$Q \approx \alpha_{10}; \; \beta_{10} \approx \omega_1 \frac{\alpha_{01}}{\alpha_{00}} = \omega_1\gamma; \; \beta_{11} \approx \omega_1 \frac{\alpha_{11}}{\alpha_{10}} = \omega_1\delta, \tag{21}$$

kuriomis galima apibūdinti klasifikavimo tikslumą, net ir nežinant tikslių tikimybių ω_1 ir ω_0 verčių. Būtent gerais pripažintų kineskopų sraute defektingumo lygis apytiksliai sumažėja nuo ω_1 iki $\omega_1\gamma$, o blogais pripažintų kineskopų sraute – padidėja nuo ω_1 iki $\omega_1\delta$.

Praktikoje tikimybės α_{ij}, $i, j = 0{,}1,$ yra nežinomos ir jas reikia vertinti iš statistinių duomenų. Pritaikykime surastą klasifikavimo taisyklę statistiniams duomenims (7). Pažymėkime V_{00} ir V_{10} gerais ir blogais pripažintų kineskopų skaičių nereklamuotų kineskopų aibėje ($V_{00} + V_{10} = N_0$); analogiškai V_{01} ir V_{11} – gerais ir blogais pripažintų kineskopų skaičių reklamuotų kineskopų aibėje ($V_{01} + V_{11} = N_1$); gautuosius rezultatus surašykime į lentelę.

2 lentelė. Klasifikavimo rezultatai

η / Y	0	1	Σ
0	V_{00}	V_{10}	N_0
1	V_{01}	V_{11}	N_1

Naudodami šiuos duomenis, gauname tikimybių α_{ij} įverčius

$$\hat{\alpha}_{ij} = \frac{V_{ij}}{N_j}, \text{ kai } \quad i, j = 0{,}1, \tag{22}$$

o kartu ir charakteristikų (21) įverčius $\hat{Q}$, $\hat{\gamma}$, $\hat{\delta}$.

P a s t a b a. Jeigu statistinių duomenų (7) pakankamai daug, tai rekomenduojama juos suskirstyti į dvi dalis: viena dalis naudojama parametrų įverčiams pagal (9) rasti, o kita dalis – kaip testinė aibė klasifikavimo tikslumui įvertinti sudarant klasifikavimo 2 lentelę.

Kineskopų klasifikavimo rezultatai

Buvo nagrinėjamas kineskopų A33 klasifikavimo uždavinys bandant atskirti potencialiai reklamuotinus kineskopus pagal BV ir MSS (blogas vakuumas ir maža spindulio srovė). Turėtas reklamuotų kineskopų duomenų masyvas ($N_1 = 303$) ir nereklamuotų kineskopų duomenų masyvas ($N_0 = 6695$). Naudojami technologiniai ir kineskopo parametrai užfiksuoti operacijose „Išsiurbimas“ (5 parametrai); „1-ojo bandymo treneris“ (12 parametrų); „1-ojo bandymo testeris“ (5 parametrai); „Testerių karuselės“ (11 parametrų); iš viso 33 parametrai.

Atlikus analizę (naudojama programų sistema SAS) liko 18 parametrų, atitinkamai 2, 4, 3, 9 iš minėtų keturių operacijų; likusieji 13 parametrų buvo pripažinti nereikšmingais.

Klasifikavimo tikslumui apibūdinti pateikiame dalį klasifikavimo lentelės, gautos klasifikuojant duomenis (7) esant keletui *c* verčių (žr.(5)).

3 lentelė. Klasifikavimo lentelė

c	V_{11}	V_{00}	V_{10}	V_{01}	$\hat{Q}$	$\hat{\gamma}$	$\hat{\delta}$
0,10	230	6558	137	73	0,0205	0,2460	37,09
0,14	227	6605	90	76	0,0134	0,2542	55,73
0,18	227	6625	70	76	0,0105	0,2535	71,65
0,22	227	6643	52	76	0,0078	0,2528	96,46
0,26	225	6648	47	78	0,0070	0,2593	105,78

Imdami, pavyzdžiui, $c = 0{,}18$, gauname, kad į įtartinų kineskopų srautą pateks apie 1% visos produkcijos. Defektingumo lygis gerais pripažintų kineskopų sraute sumažės apie 4 kartus, o blogais pripažintų kineskopų sraute padidės apie 72 kartus.

Pastaruoju metu gamykloje „Ekranas“ atliekamas technologinis eksperimentas, kurio metu iš bendro srauto į įtartinų kineskopų srautą atskiriama apie 1% visos produkcijos. Šiuos kineskopus numatoma sandėliuoti 2-3 savaites (imituojant jų patekimą vartotojui), o paskui atlikti visapusišką patikrinimą.

Literatūra

1. **Kruopis J., Vaišvila A.** Atrankinė kontrolė skaidant gaminius į du srautus // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2004. – Nr.7(56). – P.55–60.
2. **Kruopis J., Vaišvila A.** Kontrolė sugriežtinant kontrolines ribas // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2005. – Nr.1(57). – P.62–66.
3. **Čekanavičius V., Murauskas G.** Statistika ir jos taikymai. II t. – Vilnius: Leidykla TEV, 2000. – 270 p.
4. **Hosmer D.W., Lemeshow S.** Applied Logistic Regression. – New York: Wiley, 1989. – 310 p.
5. **Spanos C.J., Chen R.L** Using Qualitative Observations for Process Tuning and Control // IEEE Transactions on Semiconductor Manufacturing, 1997. – Nr.2(10).– P.307–316.

Pateikta spaudai 2005 03 10

J. Kruopis, A. Vaišvila, S. Grobovas, R. Levulienė. Kineskopų klasifikavimas taikant matematinės statistikos metodus // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2005. – Nr. 5(61). – P. 43–47.

Aptariamos galimybės klasifikuoti gaminius naudojantis gamybos metu sukaupta informacija – technologinių režimų charakteristikomis ir gaminių parametrų matavimų duomenimis, gautais kontrolių metu, taip pat informacija, gauta iš vartotojo. Lyginant duomenų masyvus, gautus reklamuotiems ir nereklamuotiems gaminiams, logistinės regresijos metodu suformuluojama klasifikavimo taisyklė ir aptariamas jos tikslumas. Šis metodas pritaikytas spalvotiesiems kineskopams klasifikuoti Panevėžio gamykloje „Ekranas“. Metodika pritaikoma ir kitiems gaminiams, jeigu tik funkcionuoja duomenų apie kiekvieną gaminį kaupimo ir saugojimo sistema. Bibl. 5 (lietuvių kalba; santraukos lietuvių, anglų ir rusų k.).

J. Kruopis, A. Vaišvila, S. Grobovas, R. Levulienė. Kinescope Classification using Mathematical Statistics Methods // Electronics and Electrical Engineering. – Kaunas: Technologija, 2005. – No. 5(61). – P. 43–47.

The product classification possibilities using technological characteristics, product parameter measurements and information supplied by users are presented in this paper. The classification rule using logistic regression methods was obtained by comparing data arrays of reclamation and non-reclamation items. The precision of proposed rule was discussed. The proposed method was applied to classify colour kinescopes at joint-stock company "Ekranas" in Panevėžys. These methods could be applied to the other products, if the system for gathering and storing information about each product exists. Bibl. 5 (in Lithuanian; summaries in Lithuanian, English and Russian).

Ю. Круопис, А. Вайшвила, С. Гробовас Р. Лявулене. Классификация кинескопов методами математической статистики // Электроника и электротехника. – Каунас: Технология, 2005. - № 5(61). – С. 43–47.

Обсуждена возможность классификации изделий на основе информации, накопленной во время изготовления, - характеристики технологических режимов и измерения параметров, полученных в контрольных операциях, а также информация, поступающая от потребителей. Сравнивая массивы данных рекламированных и нерекламированных изделий методом логистической регрессии построено правило классификации и изучено ее точность. Этот метод применим для любых изделий, если только функционирует система накопления и сохранения данных о каждом изделии в отдельности. Библ. 5 (на литовском языке; рефераты на литовском, английском и русском яз.).