

Diszkriminancia analízis alkalmazása megbízhatósági előrejelzésre a Tungram Rt-nél

RING JÁNOS
Tungram Rt



ÖSSZEFOGLALÁS

Diszkriminancia analízissel, dönthető el, hogy egy egyed melyik adott eloszláshoz tartozik. A legtöbb döntési eljárás a likelihood hányadoson alapszik.

A különböző elektrotechnikai termékek vizsgálata során néhány paramétert bizonyos időpontokban megmérnek. A mért adatok alapján dönthető el, hogy a termék „jó” vagy „rossz”. A két elosztást a „jó” ill. a „rossz” termékekből konstruálják. Az eloszlások paramétereit egy minta alapján közelítik ún. „plug in” módszerrel. Az előrejelzésre használt mérési időpontok jelentősen megelőzik a névleges élettartamot. A cikk a módszer gazdasági vonatkozásait is vizsgálja.

Bevezetés

A termelés növekedésével együtt nem nő a vizsgálati helyek száma, de szükség lenne nagyobb minta vizsgálatára. Ezért új módszereket kell kitalálni. Az egyik lehetőség a forszírozott égetés, azaz olyan feltételek teremtése, amely a vizsgált mintadarabot a normális körülményeknél jobban megterheli (pl. nagyobb feszültségen vagy magasabb hőmérsékleten vizsgálják). Ebben az esetben összefüggés határozható meg a nagyobb terhelésű és a normális körülményeknél feltételezett élettartamok között. Ennek az eljárásnak a hátránya, hogy a vizsgálat során megnő az egy időegységre jutó energiafelhasználás, valamint, hogy csak sztochasztikus kapcsolat van a két élettartam között, ezért hibát követünk el az átszámításnál. A másik lehetőség, hogy a rövidített égetést a szokásos körülmények között végzik és közben a különböző paramétereken mért értékek alapján végeznek előrejelzést a névleges időtartamra. Itt csak az előrejelzés pontatlansága ad bizonytalanságot, de jelentős energiamegtakarítást érhetünk el.

A második lehetőségnek is több fajtája van (pl. diszkriminancia analízis, clusterezés, legközelebbi szomszéd módszer stb.), mi ezek közül a legelsőt választottuk.

Az együtthetők kiszámításának elmélete

A diszkriminancia analízis alapfeladata, hogy egy mintát besorol az eloszlások valamelyikébe. Az eloszlás lehet ismert vagy ismeretlen. Ha ismeretlen, akkor a paramétereit becsülni kell.

A különböző elektronikai termékeknél (pl. elektroncsövek, félvezetők, higanygőz-lámpák) a vizsgálat során bizonyos égési idők letelte után néhány paramétert megmérnek. Ezek az eredmények, amelyek egy egyedhez tartoznak egy vektor-ként egy többváltozós valószínűségi változó aktuális értékét adják.

Beérkezett: 1985. X. 2. (Δ)

RING JÁNOS

Az Eötvös Lóránd Tudományegyetem Természettudományi karán matematika szakon kitüntetéses oklevéllel végzett 1981-ben. Azóta a Tungram Rt Vállalati Minőség-ellen-

nőrzési Főosztályán dolgozik matematikus-ként. Ott a minőségellenőrzés matematikai statisztikai problémáival és megbízhatósággal foglalkozik. Ezekről néhány cikke már megjelent a „Minőség és Megbízhatóság” folyóiratban.

A mintadarabokat két csoportba soroljuk. Az egyik csoportba tartozó egyedek a névleges élettartamot úgy érik el, hogy valamennyi mért paraméterük minden időpontban az adott tűrési határokon belül van. Ezeket hívjuk „jóknak”. A másik csoport a többi, azaz azok, amelyek nem érik el a névleges időtartamot, vagy olyan a paraméterük, amely valamely időpontban nem felel meg az előírtaknak. Ezeket hívjuk „rossznak”. Itt a meghibásodás nem mindig működésképtelenséget jelent, hanem lehet az előírt élettartamvégpont átlépése is. A két csoportba osztani először olyan megfigyelt darabokat kell, amelyek vizsgálata legalább a névleges élettartamig (T_n -ig) történt. Ezután ki kell gyűjteni ezeknél az egyedeknél a T_0 időpontig (T_0 a T_n -nél lényegesen kisebb) mért paramétereket. Ezek alkotják a már említett valószínűségi változóból származó mintát.

Modellünk szerint a két csoport egy egy eloszlást reprezentál. Feltesszük, hogy mind a kettő normális eloszlást követ közös kovariancia mátrixszal és különböző várható értékkel. Ha ezek a feltételek fennállnak, akkor az elválasztó függvény lineáris. Ezt később bizonyítjuk. A többdimenziós eset miatt a feltételek ellenőrzése nehéz. Ezért a normalitást nem ellenőrizzük. Inkább az eloszlást normálissal közelítjük, illetve lineáris elválasztó függvényt keresünk.

Az optimális döntési eljárás a Neyman—Pearson lemma következtében a likelihood hányados (a két sűrűségfüggvény hányadosa) alapján történik. Az eloszlások vagy paramétereik meghatározását a helyettesítéses (plug in) módszerrel végezhetjük. Ehhez kell elvégezni a várható értékek és a kovariancia mátrix becsléseit. A „jó” eloszlásból egy N_1 -es nagyságú mintánk van és jelöljük μ_1 -gyel a várható értéket. A „rossz” eloszlású egyedekből egy N_2 darabos minta áll rendelkezésünkre és μ_2 a várható értéke.

A várható értékek becslése:

$$\hat{\mu}_i = \bar{x}^{(i)} \left(\sum_{j=1}^N X_j^{(i)} \right) / N_i \quad i = 1, 2 \quad (1)$$

A közös kovariancia mátrix becslése:

$$\hat{S} = \frac{1}{N_1 + N_2 - 2} \sum_{i=1}^2 \sum_{j=1}^{N_i} (x_j^{(i)} - \bar{x}^{(i)})(x_j^{(i)} - \bar{x}^{(i)})^T \quad (2)$$

A sűrűségfüggvények hányados logaritmusai:

$$U = U(x) = (x - \frac{1}{2}(\mu_1 + \mu_2))^T \Sigma^{-1} (\mu_1 - \mu_2) \quad (3)$$

ez az $U(x)$ diszkrimináló (elválasztó) függvény.

$$v = \Sigma^{-1} (\mu_1 - \mu_2) \quad (4)$$

a diszkriminancia együttható, amelynek közelítése:

$$\hat{v} = S^{-1}(\bar{x}^{(1)} - \bar{x}^{(2)}) \quad (5)$$

Egy megfigyelést a „jó” közé sorolunk, ha $\hat{v}'X \geq c$, és a „rossz” közé, ha $\hat{v}'X < c$. A c összehasonlító állandó meghatározását később adjuk meg.

A kísérletek elvégzésének menete

A típus választása (pl. 250 W-os higanygőz-lámpa) után meg kell határozni a T_n névleges élettartamot (pl. 12 000 óra), T_n általában adott és az előrejelzés erre vonatkozik. Azután meg kell adni azokat a feltételeket, amelyek mellett „jó” tartjuk a termékeket (különböző paramétereiknél az élettartam végpontok). Ezután a rövidített vizsgálati idő (T_0) kijelölése történik (pl. 2000 óra). Ezzel együtt kell meghatározni azokat a paramétereket, amelyek alapján az előrejelzést végezzük (pl. fényerősség, teljesítmény, égési feszültség), és azokat az időpontokat, amikor ezeket a jellemzőket mérjük a T_0 vizsgálati idő közben (pl. 100, 500, 1000, 2000 óra). Egy minél nagyobb (legalább 100 darabos) mintát kell bekapcsolni kísérleti működésre legalább a névleges élettartam végéig. Természetesen korábbi adatok is felhasználhatók. Az égetés közben az előírt paramétereket az előírt időpontokban meg kell mérni. Ezeket az adatokat össze kell gyűjteni, majd szétválasztani azokat „jó” illetve „rossznak”. Ezután kell a két eloszlás várható értékét és a közös kovariancia mátrixot megbecsülni. Ezt a számítást már korábban mutattuk. Majd a \hat{v} vektor elemeit kell kiszámítani, ezek a lineáris kombináció együtthatói. Különböző c összehasonlító számok felhasználásával kell a hibabecslést elvégezni. Egy X_i vektor tartalmazza az egyedek mérési eredményeit. Kiszámítjuk a $\hat{v}'X_i$ skalárszorzatot és ezt összehasonlítjuk a c értékével. Ez alapján elvágunk egy másodlagos besorolást. Így négy csoport határozható meg:

$$H_1 = \{ \text{„jó” és } \hat{v}'X \geq c \}$$

$$H_2 = \{ \text{„jó” és } \hat{v}'X < c \}$$

$$H_3 = \{ \text{„rossz” és } \hat{v}'X \geq c \}$$

$$H_4 = \{ \text{„rossz” és } \hat{v}'X < c \}$$

Ez a besorolás c -től függő.

H_i halmaz elemszámát $\#H_i$ -vei jelöljük, $i = 1, 2, 3, 4$,

$$\text{Az elsőfajú hiba becslése: } \frac{\#H_2}{\#H_1 + \#H_2} \quad (7)$$

$$\text{A másodfajú hiba becslése: } \frac{\#H_3}{\#H_3 + \#H_4} \quad (8)$$

Azt a c értéket választjuk ki, amelyre az első- és másodfajú hiba optimálisan kicsi. Ezután gazdaságossági számításat kell végezni, amely segítséget nyújt a módszer bevezetéséről történő döntéshez. Megpróbáljuk összehasonlítani a költségeket és veszteségeket a hagyományos és a rövidített módszer esetén.

A költségek közelítő számítása, a megfelelő lehetőség kiválasztása

A diszkriminancia analízisen alapuló eljárásnál a vizsgálati költség, az idő megrövidítése és a feltételek meghagyása miatt jelentősen kisebb. Ez igaz még akkor is, ha nagyobb mintát veszünk, annak érdekében, hogy minél kisebb legyen az a bizonytalanság, amelyet a tétel nagyság és a mintanagyság közötti nagy különbség okoz. Azonban ennél az eljárásnál fellép egy olyan veszteség, amelyet az első- és a másodfajú hiba miatti tévedés okoz. Csak az ilyen számítások eredményeinek teljes ismeretében kell dönteni, hogy a módszer melyik fajtáját vezessük be.

A lehetőségek a következők:

- (i) Ha elérhető, hogy mindkét hibafajta egyszerre elég kicsi, akkor a T_0 idő után döntünk a mintadarab besorolásáról a lineáris kombináció és az összehasonlító szám egybevetése alapján.
- (ii) Ha csak az elsőfajú hiba csökkenthető megfelelően kicsi érték alá, akkor a T_0 idő után azokat „rossznak” tekintjük, amelyekre $\hat{v}'X < c$. De a többit T_n -ig vizsgáljuk és a hagyományos módszer alapján soroljuk be ezeket a megfelelő csoportba. Ezzel a másodfajú hibát 0-ra csökkenthetjük, de a vizsgálati idő hosszabb, mint az első esetben.
- (iii) A (ii)-höz hasonlóan járunk el, csak az első- és a másodfajú hibát és az egyenlőtlenség irányát kell megcserélni.
- (iv) $\hat{v}'X$ értéke alapján döntünk
 - (a) $\hat{v}'X < C_1$ az egyedeket rossznak soroljuk be
 - (b) $C_1 < \hat{v}'X < C_2$ az egyedeket tovább vizsgáljuk T_n -ig (Ez egy ún. bizonytalan intervallum)
 - (c) $\hat{v}'X > C_2$ az egyedeket jók soroljuk be.
 - (v) Új paramétereket és időpontokat keresünk, mert egyik előző eset sem mutat nyereséget hagyományos módszerrel szemben. Utána égetéseket, méréseket, becsléseket kezdjük előlről.
- (vi) A módszert elvetjük, mert a gazdaságossági számítás a (v) egyik választása mellett sem ígér hasznot.

A módszer alkalmazása

Ezután már alkalmazható a kiválasztott módszer: A mintadarabokat (vagy a (ii) és (iii) esetben azok egy részét) rövidített ideig vizsgáljuk. Közben az adott időpontokban a kiválasztott paramétereket mérjük. A mért értékekből egyenként összeállítjuk az X vektort. T_0 idő után a $\hat{V} \cdot X$ lineáris kombináció és a c összehasonlítása alapján döntünk, hogy melyik csoportba soroljuk az egyedet (illetve tovább vizsgáljuk). Ha T_n ideig kell vizsgálni akkor ezeket a hagyományos feltételek alapján osztályozzuk a „jó” illetve „rossz” osztályok valamelyikébe.

Bizonyos időközönként felül kell vizsgálni a c összehasonlító szám értékét és a hibabecslést.

Szintézis

Az első próbálkozások a módszer bevezetésére 1981-ben a *PL* 519-es színes video végerősítőcsőnél történtek. Azelőtt 1000 óráig 0, 100, 200, 500 és 1000 órás vizsgálat után mérték a különböző fizikai jellemzőket: dinamikus végfeszültség, rácsfeszültség, nomrálfűtésű katódáram-erősség, aláfűtött katódáram-erősség, %-os katódáram csök-

kenés. A csövet akkor sorolták a „rosszak” közé, ha valamelyik adata akármilyen időpontban nem felelt meg az előírásnak. A diszkriminancia analízisen alapuló eljárás már 200 órás égetés után dönt. Ennek kettős jelentősége van. 800 órával lecsökkentik az ellenőrzési időt, így sokkal előbb kerül a termék a kereskedelembe, a berendezést csak az eddigi idő 20%-ában veszik igénybe.

Ezután többfajta termékre (elektroncső, félvezető, higanygőz-lámpa) megtörtént a lineáris kombináció együtthatóinak kiszámítása. Legkedvezőbb eredmény a *BC* 177-es félvezetőknél adódott, ahol a hibabecslés szerint mindkét hibafajta a c megfelelő választása esetén 6% alatt van.

I R O D A L O M

- [1] *Giri, N. C.*: Multivariate Statistical Inference Academic Press, New York—San Francisco—London, 1977.
- [2] *Morisson, D. F.*: Multivariate Statistical Methods McGraw—Hill, New York—St. Louis—San Francisco, 1976.
- [3] *Cacoulious, T.*: Discriminant Analysis and Application Academic Press, New York—London, 1973.
- [4] *Arató, M.*: Fejezetek a matematikai statisztikából számítógépes alkalmazásokkal II. Többdimenziós analízis SZÁMKI Közlemények 22, Budapest, 1979.