

JATE Kibernetikai Laboratórium

Valószínűségi számítási módszeren alapuló osztályozási eljárás típus-
alkotási problémák megoldására

Hunya Péter

Kísérletes tudományokban gyakran találkozunk a legkülönbözőbb objektumok osztályozásának, felismerésének, osztályba sorolásának kérdésével. A mesterséges intelligencia területén végzett eddigi kutatások azt mutatják, hogy e tevékenységeknek legalább is bizonyos elemei automatizálhatók és ezzel a kísérleti kutatómunka bizonyos fázisai nagymértékben megkönnyíthetők. Az alakfelismerés, ami a mesterséges intelligencia egyik igen fontos területe, elsősorban arra az esetre vonatkozóan adott eredményeket, amelyben az objektumok osztályai már eleve többé-kevésbé definiálva vannak. A feladat itt abban áll, hogy új, eddig ismeretlen objektumokat kell besorolni a korábban definiált osztályok valamelyikébe (1), (2), (3), (4). Az osztályok jellemzőit, melyek alapján a besorolást valamilyen eljárás segítségével elvégezzük, tanuló algoritmusokkal határozzuk meg. Tanulásról beszélünk abban az értelemben, hogy az alakfelismerési rendszer autonom módon dolgozza fel az ismert objektumokról kapott információt és ennek megfelelően módosítja a "tanító" által megjelölt osztály jellemzőit. E rendszerekben tanítóval végzett tanulásról (supervised learning) beszélünk (5). Tanító nélkül végzett tanulás esetében (unsupervised learning, nonsupervised learning) a tanulási folyamat nem csak az osztály-jellemzők meghatározásából áll, hanem tartalmazza az osztályok kijelölését is, mivel a tanulás folyamán a rendszer nem kap információt az objektumok hovatartozásáról, sőt egyes esetekben az osztályok számáról és létezéséről sincs információ (6). Az előadásban ez utóbbi problémakörrel foglalkozunk.

Először is megadjuk az osztályozási probléma és a tanulás vizsgálatainkban használt matematikai modelljét. Tekintsük az osztályozandó u objektumok U halmazát. Az u elemeket rendezett n -esekkel jellemezzük, nevezetesen az

$$s = (s_1, \dots, s_n)$$

vektorokkal, melynek komponensei tetszőleges S_i halmazok elemei ($i=1, \dots, n$). Az s komponenselt általában valamilyen mérési rendszer szolgáltatja. A mérési rendszer így egy leképezést létesít az \underline{U} és az

$$S = \prod_{i=1}^n S_i$$

szorzathalmaz között (a \prod jel a halmazok Descartes-szorzatát jelöli) és ezt a leképezést az $s=m$ (u) mérési függvény segítségével írhatjuk le.

A megfontolások egyszerűsítése végett feltételezzük, hogy az S_i halmazok ($i=1, \dots, n$) diszkrét. Ez nem jelenti az általánosság korlátozását, mivel az eredmények folytonos esetre közvetlenül átvihetők.

Tekintsük most \underline{U} egy felbontását a C_1, \dots, C_N részhalmazokra. Megköveteljük, hogy a C_i ($i=1, \dots, N$) részhalmazok eleget tegyenek a

$$C_i \cap C_j = \emptyset \quad i \neq j$$

és a

$$\bigcup_{i=1}^N C_i = U$$

feltételnek. Ezt a továbbiakban az \underline{U} osztályozásának nevezzük és

$$U(C_1, \dots, C_N) -$$

nel jelöljük. Feltételezzük továbbá, hogy \underline{U} -n adott egy p mérték, melyre

$$0 < p(C) \leq p(U) < \infty \quad C \subseteq U$$

és minden $C \subseteq U$ -ra definiálva van a $P(s|C)$ feltételes valószínűség-eloszlás, melyre diszjunkt C', C'' mellett fennáll

$$P(s|C' \cup C'') = \frac{1}{P(C') + P(C'')} (P(C')P(s|C') + P(C'')P(s|C'')),$$

ahol

$$P(C) = \frac{P(s|C)}{P(U)} \cdot$$

A modellben a $P(s|C_i)$ és a $P(C_i)$ valószínűségeket tekintjük a C_i osztály jellemzőinek. A $P(s|U)$ -ra bevezetjük a $P(s)$ jelölést.

Felhasználva a fentiekben bevezetett terminológiát és jelöléseket a felismerés (recognition) problémáját a következőképp fogalmazzhatjuk meg:

Egy u objektum felismerése azon C_i osztály (részhalmaz) meghatározását jelenti, amely u -t elemeként tartalmazza. A megfigyelési (mérési) rendszer alapján azonban az u objektum helyett annak S -beli képe ismert csupán és ennek alapján kell elvégezni az u azonosítását, felismerését. Az m függvényről általában nem tehetjük fel az egy-egyértelműséget, így az a természetes mód, hogy az m inverz függvényét használjuk fel a felismerésre - nem használható, mivel az nem minden esetben létezik. Az inverz kép nem-egyértelműsége miatt célszerű valószínűségi számítás tárgyalásmódot alkalmazni. Az eljárások egy széleskörűen elterjedt csoportja olyan döntési rendszert használ a felismerésnél, mely valamilyen "veszteség" várható értékét minimalizálja. E rendszereket Bayes-stratégiajú rendszereknek nevezzük (7). A Bayes-tétel segítségével, felhasználva a korábban definiált eloszlásokat, meghatározható annak a valószínűsége, hogy egy u objektum, melyre $m(u) = s$, a C_i osztályba tartozik, nevezetesen:

$$P(C_i|s) = \frac{P(s|C_i) P(C_i)}{\sum_{i=1}^N P(s|C_k) P(C_k)} \quad (i=1, \dots, N)$$

Ismeretes, hogy e valószínűségek Bayes-stratégiajú döntési rendszert biztosítanak. Vizsgálatainkban a Bayes-tételből adódó valószínűségeket használjuk a felismerés eszközeként. Mivel a szükséges valószínűségeloszlások általában a priori nem ismeretesek, a felismerési fázist megelőzi egy $u.n.$ tanuló fázis.

Azon problémákban, melyekben adva van egy jól definiált $U(C_1, \dots, C_N)$ osztályozás és a tanuló fázisban a mért $s=m(u)$ jellemzőkkel együtt a rendszer információt kap a megfelelő C_i osztályról is, tanulás alatt az osztályokat jellemző eloszlások szukcesszív felépítését értjük. Ekkor beszélünk tanítóval végzett tanulásról.

Számos fontos problémában azonban nincs eleve adott osztályozás. Itt a feladat az, hogy olyan $U(C_1, \dots, C_N)$ osztályozást adjunk meg, amely az u objektumok s képének felhasználásával kis bizonytalanságu, hatékony felismerési rendszert biztosít. Itt a tanulás a megfelelő osztályozás keresését és megtalálását jelenti és ezt tanító nélküli tanulásként nevezzük. Ez a "tanulási forma" igen fontos szerepet játszik például a tudományos megismerő tevékenységben, rendezésben, absztrakcióban, stb.

A tanító nélküli tanuló algoritmusok legnagyobb részében valamilyen $F(C_1, \dots, C_N)$ funkcionállal jellemzik az $U(C_1, \dots, C_N)$ osztályozásokat (8), (6). Rendszerünkben az osztályozás a következő módon történik. Definiáljuk két osztály (C_i, C_j) szorzatát feltételes valószínűségeloszlásaik segítségével

$$(C_i, C_j) \stackrel{\text{def}}{=} \sum_{s \in S} P(s|C_i) P(s|C_j)$$

és a rövidség kedvéért jelöljük λ_{ij} -vel. Nyilvánvaló, hogy $\lambda_{ij} = \lambda_{ji}$. Az $U(C_1, \dots, C_N)$ osztályozáshoz a

$$\lambda = \max_{\substack{1 \leq i, j \leq N \\ i \neq j}} \lambda_{ij}$$

értéket rendeljük hozzá és ezt a λ értéket nevezzük az osztályozás szintjének.

A Bayes-formula alkalmazásával annak valószínűségére, hogy egy u objektum a j -edik osztályba tartozik, a következő adódik:

$$P_j(u) = P(C_j | S=m(u)) = \frac{P(s|C_j) P(C_j)}{\sum_{k=1}^N P(s|C_k) P(C_k)}$$

Az osztályozás jószágára jellemző, hogy egy C_i ($i \neq j$) osztály elemeire milyen P_j értékeket kapunk. Ha ezek a valószínűségek nagyok, az osztályozási rendszer, a besorolás nagy bizonytalanságot tartalmaz, ha kicsik, vagy éppen zéróval egyenlők, a bizonytalanság is kicsi lesz. Az egyedi $P_j(u)$ értékek helyett célszerűbb azok várható értékét tekinteni, mivel az ilyen típusu paraméterek jobban jellemzik az osztályozás egészét.

Tekintsük a C_j osztály elemeire kapott $P_j(u)$ valószínűségek várható értékét és jelöljük P_{ij} -vel:

$$P_{ij} = \sum_{s \in S} P(C_j | s=m(u)) P(s|C_j) = \sum_{s \in S} \frac{P(s|C_j) P(C_j) P(s|C_j)}{P(s)}$$

Azért célszerű a λ -t választani az osztályozás jellemzőjének, szintjének, mivel segítségével bizonyos feltételek mellett a P_{ij} értékekre egy felső korlát adható meg.

Tételezzük fel, hogy

$$P(s) \geq K > 0$$

és
$$P(C_i) \leq A \quad (i=1, \dots, N)$$

Ezen feltételek mellett

$$P_{ij} \leq \frac{A}{K} \lambda, \quad (i \neq j)$$

mivel feltételeink szerint:

$$\begin{aligned} P_{ij} &= \sum_{s \in S} \frac{P(s|C_i)P(C_i)P(s|C_j)}{P(s)} = P(C_i) \sum_{s \in S} \frac{P(s|C_i)P(s|C_j)}{P(s)} \leq \\ &\leq \frac{P(C_i)}{K} \sum_{s \in S} P(s|C_i)P(s|C_j) \leq \frac{A}{K} \lambda_{ij} \leq \frac{A}{K} \lambda \end{aligned}$$

Felhívjuk a figyelmet arra, hogy A és K függetlenül választható az osztályozástól, így a felső korlát csak az $U(C_1, \dots, C_N)$ osztályozáshoz tartozó λ szint függvénye.

Ha $\lambda=0$, akkor a fenti eredmény miatt $i \neq j$ mellett az összes P_{ij} értékek nullával egyenlők, ami azt jelenti, hogy az osztályba sorolásnál semmiféle bizonytalansággal nem kell számolnunk. A $\lambda=0$ esetben ortogonális osztályozásról beszélünk, tekintettel arra, hogy ekkor bármely két osztály szorzata nulla lesz. Gyakorlatban azonban ez az eset ritkán fordul elő és a lehetséges osztályozásokat a $\lambda > 0$ szint jellemzi. Kis λ szint esetén quasi-ortogonális rendszerről beszélünk. E rendszereket is kis bizonytalanság jellemzi, ezért jól alkalmazhatók a gyakorlatban.

Az előzőekben kifejtett gondolatmenet alapján a tanulás olyan osztályozások keresését jelenti, melyek minimalizálják λ -t. Ezen osztályozások általában több elemből álló halmazt alkotnak. Más szavakkal olyan osztályozásokat keresünk a tanulás folyamán, melyek kis bizonytalansági szintjükkel biztosítják az objektumok helyes azonosítását, hasznos állapot szolgáltatva ezzel típusok, kategóriák kialakítására. Bizonyos esetekben azonban nem az optimális szinthez tartozó osztályozások meghatározása a feladat, hanem olyan osztályozásoké, melyek megfelelnek valamilyen előre meghatározott λ szintnek.

Az osztályokhoz tartozó feltételes valószínűségekre és az osztályokra vonatkozó megszorítások nélkül a következőt állíthatjuk:

Ha adott egy λ szintű osztályozás és ebben bármely két osztályt összevonjuk, az így kapott osztályozás λ^n szintjére

$$\lambda'' \leq \lambda'$$

áll fenn. Nyilvánvaló, hogy a változatlanul hagyott osztályok szorzatainak értéke változatlan marad. Tegyük fel, hogy a k_1 és a k_2 osztályokat egyesítettük és ez az egyesítés után a k -adik osztály lett. Ennek jellemzője feltételeink szerint a

$$P(s|C_k) = \frac{1}{P(C_{k_1}) + P(C_{k_2})} (P(C_{k_1})P(s|C_{k_1}) + P(C_{k_2})P(s|C_{k_2}))$$

eloszlás lesz, így ennek és egy j -edik osztálynak a szorzatára

$$\lambda''_{jk} = \frac{P(C_{k_1})}{P(C_{k_1}) + P(C_{k_2})} \lambda'_{jk_1} + \frac{P(C_{k_2})}{P(C_{k_1}) + P(C_{k_2})} \lambda'_{jk_2} \leq \lambda'.$$

Ebből

$$\lambda''_{jk} \leq \lambda' \quad (j \neq k)$$

és

$$\lambda''_{il} = \lambda'_{il} \leq \lambda' \quad (i \neq l, i, l \neq k)$$

miatt

$$\lambda'' \leq \lambda'$$

adódik.

Ennek megfelelően egy olyan U^1, U^2, \dots, U^q osztályozás-sorozatnál, ahol U^{i+1} -et U^i -ből osztályok egyesítésével kaptuk, a megfelelő szintekre a

$$\lambda^1 = \lambda^2 = \dots = \lambda^q$$

egyenlőtlenségsorozat áll fenn.

Az előzőek miatt az optimális λ szintnek eleget tevő osztályozások közül csak azok tekinthetők függetleneknek, melyeknek nincs finomítása a megoldások között, azaz amelyek maximális számú osztályt tartalmaznak.

Gyakorlati alkalmazásokban célszerű a feltételes eloszlásokra bizonyos megszorításokat tenni. Ezzel csökkenthetjük az optimális megoldások halmazában lévő elemek számát is. E célra kiválasztjuk a valószínűségeloszlások valamely jól definiált csoportját (pl. normális eloszlások kompozícióját) és ebből meghatározzuk azt, amely a legjobban illeszkedik az empirikusan konstruáltakhoz.

További megszorítást jelent, ha feltételezzük, hogy az s vektor komponensei stochasztikusan függetlenek. Ekkor a $\lambda_{ij} = (C_i, C_j)$ szorzat a következő alakú:

$$\lambda_{ij} = \prod_{k=1}^n \sum_{s_k \in S_k} P(s_k | C_i) P(s_k | C_j),$$

ahol $P(s_k | C_i)$ a k -adik komponens eloszlását jelöli C_i esetén. Vizsgáljuk meg most, hogyan viselkedik a λ szint a komponensek számának növelésével. E célból tekintsük az n komponenshez tartozó λ^n és az $n+1$ komponenshez tartozó λ^{n+1} szintek viszonyát.

Nyilvánvaló, hogy

$$\lambda_{ij}^{n+1} = \lambda_{ij}^n \sum_{s_{n+1} \in S_{n+1}} P(s_{n+1} | C_i) P(s_{n+1} | C_j),$$

és mivel

$$\sum_{s_{n+1} \in S_{n+1}} P(s_{n+1} | C_i) P(s_{n+1} | C_j) \leq 1,$$

ezért fennáll

$$\lambda_{ij}^{n+1} \leq \lambda_{ij}^n.$$

Ebből következik

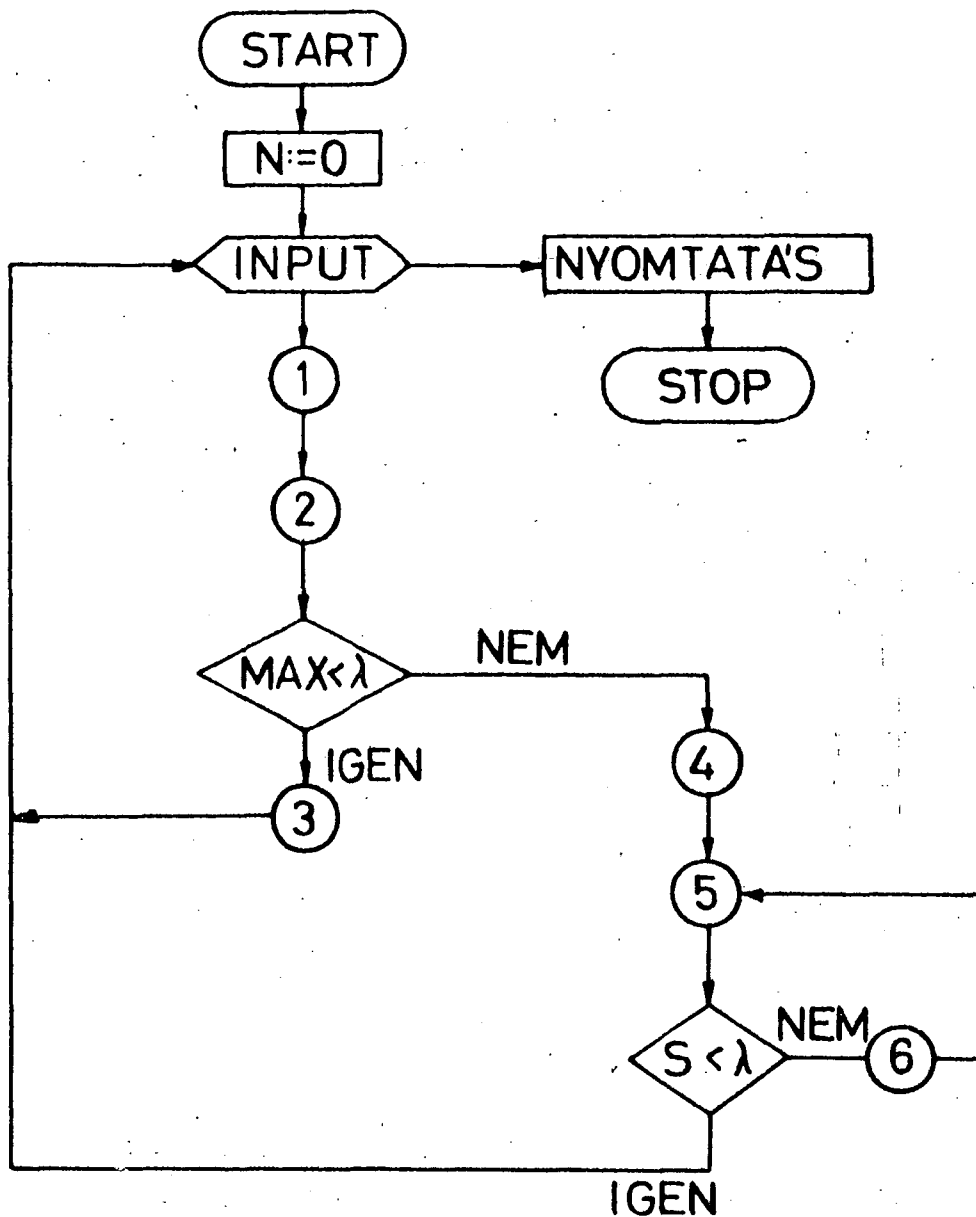
$$\lambda^{n+1} = \max_{i,j} \lambda_{ij}^{n+1} \leq \max_{i,j} \lambda_{ij}^n \leq \lambda^n,$$

azaz

$$\lambda^{n+1} \leq \lambda^n.$$

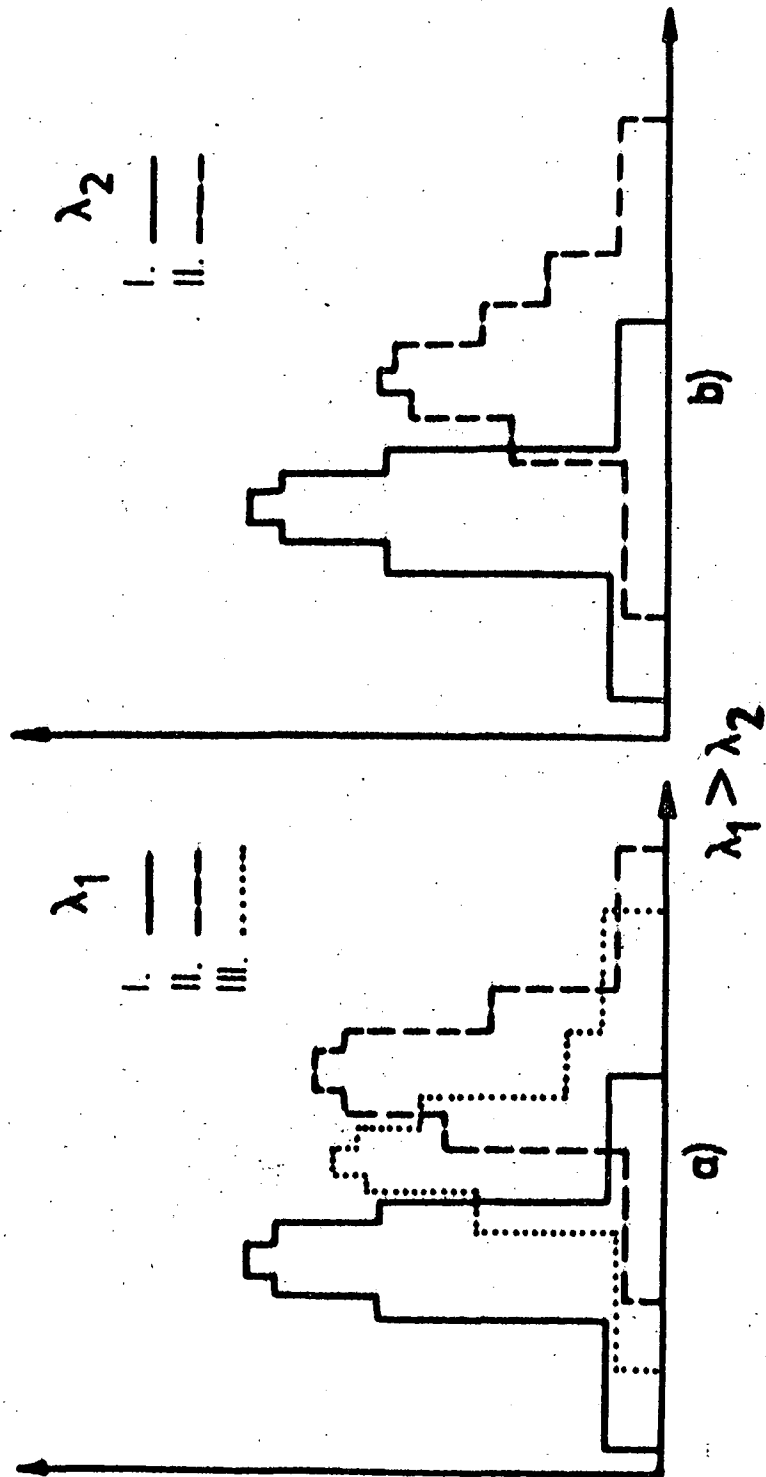
E mellett a függetlenség megkönnyíti az eloszlások tárolását is, ami lényeges szempont a gyakorlati alkalmazásoknál.

Elsődleges fontosságú ugyancsak gyakorlati alkalmazásokban az optimalizációs algoritmus kérdése. Erre ezideig exakt eljárás nem ismeretes. A fent ismertetett osztályozási rendszert komplex szabályozási funkciók vizsgálatára szolgáló multifázisú tesztből kapott eredmények



1. ábra

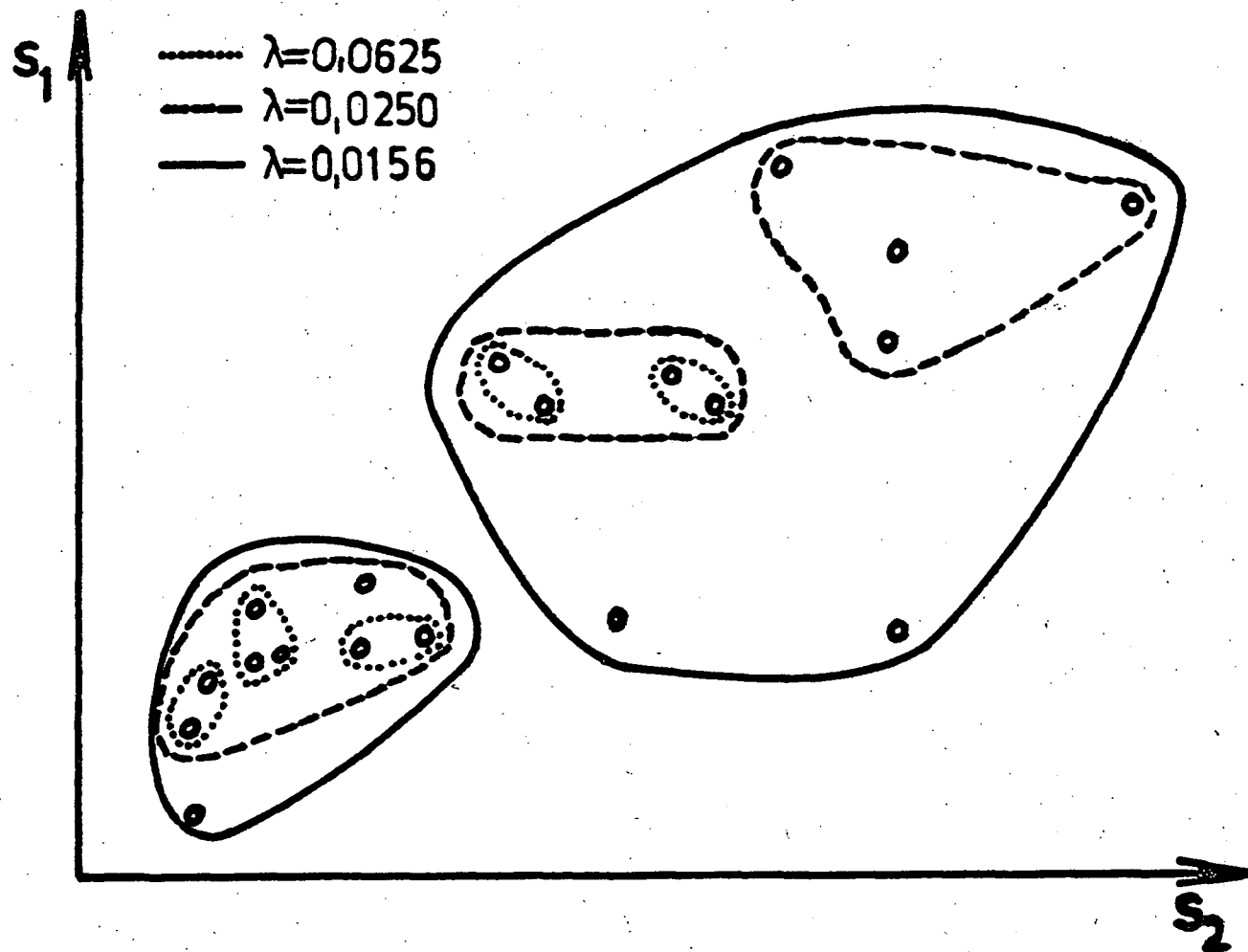
- N : osztályok száma,
 D_i : az i -edik osztályt jellemző eloszlás ($i=1, \dots, N$)
 D_{inp} : a beolvasott objektumhoz tartozó "elemi" eloszlás,
 λ : az osztályozás megkövetelt szintje,
1 - D_{inp} előállítás,
2 - $MAX = \max D_{inp} \times D_i$ szorzatérték meghatározása
3 - osztályok számának növelése ($N:=N+1$) és D_{inp} tárolása D_N -ben
4 - D_{inp} és $D_{i_{max}}$ egyesítése,
5 - osztályok közötti maximális szorzat (S) meghatározása,
6 - a maximális szorzatot adó osztályok egyesítése, a közös eloszlás tárolása, az eredetiek törlése és az N csökkentése ($N:=N-1$)



2. ábra

Egy komponens eloszlásainak képe különböző λ szinthez tartozó osztályozások esetére

- a) három osztályos,
- b) két osztályos felbontás.



3. ábra

Szigorúan monoton λ -sorozat mellett adódó hierarchikus osztályozás-rendszer. További magyarázat a szövegben.

(9) elemzésében alkalmaztuk. Az adott szintű, illetve optimális osztályozások keresésére heurisztikus algoritmust használtunk, amely nem-exakt volta ellenére is jó eredményeket szolgáltatott. Az alkalmazott eljárást az 1. ábra folyamatábrája szemlélteti. A 2. ábrán egy komponens különböző osztályokban mutatott eloszlásának tipikus képét szemléltetjük különböző λ szintek mellett. Az a/ diagramon a magasabb λ_1 szinthez tartozó három osztály, a b/-n egy kisebb λ_2 -höz tartozó két osztályos felbontás karakterisztikus eloszlásait ábrázoltuk. A 3. ábrán a multifázisos tesztből nyert, a szabályozási tevékenység jellegét jellemző általánosított távolság munkaterhelés hatására történő megváltozása (9) alapján (s_1 : általánosított távolság a munka megkezdésekor, s_2 : általánosított távolság a munka befejezésekor), különböző λ szintek mellett kialakult csoportokat adjuk meg. A jelen előadásban nem vállalkozunk a csoportoknak megfelelő pszichofiziológiai típusalkotási problémák vizsgálatára, célunk egy elméletileg megalapozott módszer ismertetése volt, mely az eddigi tapasztalati eredmények alapján hatékony eszközt nyújt e kérdéskör vizsgálatában.

I r o d a l o m

- (1) Wartak, J.: An Information Theory Approach to Medical Diagnosis. Cybernetica (Namur). Vol. 8. No. 3.: 162-168, (1965)
- (2) Aiserman, M.A.: Remarks on Two Problems Connected with Pattern Recognition. Methodologies of Pattern Recognition. Academic Press, (1969). 1-10.
- (3) Fu, K.S.: On Sequential Pattern Recognition Systems. Meth. of Pattern Rec. Academic Press, (1969). 159-202.
- (4) Szalay, G., L. Molnár, O. Gulyás: Primenyenyije algoritmov obucsenyija v meteorologiji dlja predskazuvanyiji konvektyivnoj aktyivnoszty. Acta Cybernetica. Tom. 1. Fasc. 3.: 201-218, (1972)
- (5) Cooper, P.W.: Nonsupervised Learning in Statistical Pattern Recognition. Meth. of Pattern Rec. Ac. Press, (1969). 97-110.
- (6) Dorofejuk, A.A.: Algoritmu obucsenyija masinu paszpoznavanyiju obrazov bez ucsityelja. Avtomatika i Telemekhanika. 10.: 78-87, (1966)

- (7) Naguchi Shoichi, Kazuyuki Nagasawa, Juro Oizumi: The Evaluation of the Statistical Classifier. Meth. of Pattern Rec.Ac. Press, (1969). 437-456.
- (8) Ruspini, E.H.: A New Approach to Clustering. Information and Control. Vol.15. No.1.: 22-31, (1969)
- (9) Hunya P., Madarász I.: Központi idegrendszeri szabályozási funkciók komplex vizsgálata. Számítástechnikai és kibernetikai módszerek alkalmazása az orvostudományban és biológiában. Kollokvium. 55-62, (1972)