

Rerum Naturalium Fragmenta No. 262

Jaskó Tamás - Viczián István:
Néhány a földtanban alkalmazható
egzakt osztályozási módszer

Budapest
1970

.

Rerum Naturalium Fragmenta

Jaskó Sándor szerkesztése:

Budapest XII., Pethényi köz 4

Néhány a földtanban alkalmazható egzakt osztályozási módszer

írták: Jaskó Tamás és Dr. Viczián István*

Egy találó megállapítás szerint (Pusztovarov, N. L. 1962) „a vizsgált tárgyak osztályozásának állapota tükrözi az illető tudomány egészének állapotát”. Ez a gondolat is indokolja, hogy különös figyelmet szenteljünk a földtani osztályozás kérdésének.

Ismertetésünkben három témával foglalkozunk:

1. a földtani osztályozás általános, főleg halmazelméleti és matematikai-logikai alapokon nyugvó elmélete,
2. a „cluster” program és
3. a diszkrimináns függvények.

1. A földtani osztályozás általános elmélete

Bevezető megjegyzések

Ennek az elméletnek a kidolgozása az ún. „novoszibirszki iskola”, Ju. A. Voronin és munkatársai érdeme. Eredményeiket több cikkben és összefoglaló formában 1967-ben a „Geologija i matematika” (szerk.: É. É. Fotiadi) című könyvben adták közre. E könyv bevezetése szerint a formális szempontból elmaradt, „leíró” tudományokban, mint pl. a földtanban, az osztályozásnak van éppen olyan nagy jelentősége, mint a képletek alkalmazásának az „egzakt” tudományokban.

Dolgozatunk első részében ezen elmélet főbb gondolatait szeretnénk ismertetni, emellett illusztrációképpen néhány konkrét földtani példát is be kívánunk mutatni.

Az osztályozás alapfogalmai

Először meg kell adnunk az objektumoknak azt az A halmazát, amelyen az osztályozást végezzük. Egy halmazt kétféle módon lehet megadni:

1. Megadható egy formális szabály arra nézve, hogy egy tetszőeszerinti x objektum beletartozik-e az A halmazba, vagy sem.

Pl.: „Mészkö” = olyan kőzetek halmaza, amelyek CaCO_3 tartalma nagyobb, mint 75%.

2. Megadható az A halmaz elemeinek teljes felsorolása.

Pl.: „A Polgárdi 3. sz. mélyfúrásból röntgenvizsgálatra leadott minták halmaza” = 2., 3., 4., ..., 63. sz. minta.

Az 1. megadási mód fogalmak segítségével és a nomenklátúra közvetítésével történik, de e kérdésekre a formális tárgyalás keretében nem kell kitérnünk.

Az A halmaz elemeinek osztályokba sorolása a halmaz elemeinek az A halmazon meghatározott $\phi[A]$ tulajdonságai segítségével történhet. Ezeknek a tulajdonságoknak lehet számszerű értéke, vagy „igen-nem” típusú logikai értéke.

Pl.: Számszerű érték: „klorit-tartalom %-ban” tulajdonság; logikai érték: „finom-réteges” tulajdonság.

E kétféle tulajdonságot együtt tárgyalhatjuk, ha bevezetjük az *ismertetőjegy* fogalmát. Nevezzük az A halmazba tartozó a objektum valamennyi olyan tulajdonságát, amellyel az vagy rendelkezik, vagy nem, de amelyet az adott osztályozás céljából figyelembe vettünk, ismertetőjegynek. Az első esetben az $u(a) = 1$, a második esetben az $u(a) = 0$ jelölést használjuk. Az ismertetőjegyekkel minden további nélkül jellemezhetünk logikai értékű tulajdonságokat, míg a számszerű értékű tulajdonságok lehetséges értékkészletét valamilyen meggondolás alapján intervallumokra kell bontani.

Pl.: „Az 5. sz. minta finom-réteges” logikai tulajdonság:
 $u[\text{finomr}](5. \text{ sz. m.}) = 1$;

„az 5. sz. minta nem finom-réteges” logikai tulajdonság:
 $u[\text{finomr}](5. \text{ sz. m.}) = 0$;

„a 33. sz. minta klorittartalma beleesik a 38–42% intervallumba” számszerű tulajdonság:

$$u[\text{kl}/38-42\%](33. \text{ sz. m.}) = 1$$

„a 33. sz. minta klorittartalma nem esik bele a 38–42% intervallumba” számszerű tulajdonság:

$$u[\text{kl}/38-42\%](33. \text{ sz. m.}) = 0.$$

Anélkül, hogy pontos definícióikat megadnánk, megemlítjük az ismertetőjegyek azon tulajdonságait, amelyeket az osztályozás megkezdése előtt meg kell vizsgálnunk: egy ismertetőjegy lehet „az A halmazt felosztó”, két ismertetőjegy lehet „az A halmazon egymást kizáró” vagy „összeegyeztethető”, ill. lehet egy ismertetőjegy „egy másiktól következő” vagy attól „független”.

Az ismertetőjegyeket ismertetőjegy-készletekbe foglalhatjuk össze:

$$\{ u [i/k] \} = U [i / k = 1, 2 \dots n(i)].$$

Pl.: „A Polgárdi 3. sz. mélyfúrás röntgenvizsgálatra leadott mintáinak halmazát” a következő ismertetőjegy-készletekkel jellemeztük:

$$U[kl] = \{ u[kl/23-27], u[kl/28-32], u[kl/33-37], u[kl/38-42], u[kl/43-47], u[kl/48-52] \},$$

$$U[ill] = \{ u[ill/28-32], u[ill/33-37], u[ill/38-42], u[ill/43-47], u[ill/48-52], u[ill/53-57], u[ill/58-62], \},$$

$$U[mm] = \{ u[mm/3-7], u[mm/8-12], u[mm/13-17], u[mm/18-22], u[mm/23-27], u[mm/28-32] \},$$

(Felső indexben: kl = klorit, ill = illit, mm = montmorillonit; alsó indexben a számadatok %-ban).

E jelöléssel pl. a 33. sz. minta ismertetőjegy-készletei:

$$U[kl] = \{ 0, 0, 0, 1, 0, 0 \},$$

$$U[ill] = \{ 0, 0, 0, 1, 0, 0, 0 \},$$

$$U[mm] = \{ 0, 0, 1, 0, 0 \},$$

(azaz összetétele; ki: 38—42%, ill: 43—47%, mm: 13—17%.)

Az ismertetőjegy-készletek tulajdonságait is csak megemlíthetjük: lehetnek „homogének” (azaz „azonos dimenziójúak”) és „alternatívák”. Egy ismertető jegykészlet lehet „egy másiktól következő”, ill. „egy másiktól független”.

Valamennyi figyelembe vett U ismertetőjegy-készlet együtt az U ismertetőjegy-rendszert alkotja:

$$\{ U[i] \} = U, i = 1, 2, \dots, l'.$$

Pl.: A fenti halmaz ismertetőjegy-készleteinek egyesítése a halmaz ismertetőjegy-rendszerét adja:

$$U = \{ U[kl], U[ill], U[mm] \},$$

a 33. sz. minta ismertetőjegy-rendszere: $U(33. \text{ sz. m.}) = \{ 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0 \}$.

Az ismertetőjegyek jellemzése után definiáljuk az A halmaz partíciójának fogalmát. Az A halmaz partícióján az A halmaznak az $A[i]$, $i = 1, 2, \dots, N$, részhalmazok formájában való előállítását értjük, oly módon, hogy teljesüljenek az alábbi feltételek:

$$A[i] \cap A[k] = 0, \quad i \neq k \quad (1)$$

(a részhalmazoknak nincsen közös része);

N

$$\bigcup_{i=1}^N A[i] = A \quad (2)$$

$i=1$

(a részhalmazok egyesítése az A halmazt adja) és

$$A[i] \neq 0, \quad i=1, 2, \dots, N \quad (3)$$

(egyik részhalmaz sem üres).

Végül az alapfogalmak között tárgyaljuk az A halmaz $u[i/k](a)$ ismertetőjegyekkel jellemzett a elemeinek egymáshoz való viszonyát az A halmaz partíciója szempontjából. Az A halmaznak az ismertetőjegyek alapján egymással ekvivalens elemeit egy osztályba soroljuk. Ezt pontosabban úgy fejezhetjük ki, hogy az A halmaz partícióját egy az A halmazon értelmezett U

ismertetőjegy-rendszeren alapuló ún. nem-különbözőségi reláció segítségével valósítjuk meg.

(Megemlítjük, hogy a nem-különbözőségi reláció az ekvivalencia-relációk közé tartozik.) Az A halmazba tartozó a' és a'' elemek akkor vannak egymással az U ismertetőjegy-rendszer alapján nem-különbözőségi relációban, ha valamennyi $u[i/k] \in U$ ismertetőjegyre teljesül, hogy

$$u[i/k]a' + u[i/k]a'' \neq 1$$

(hanem vagy 0, vagy 2), más szóval bármelyik $u[i/k]$ ismertetőjegyet vegyük is U -ból, azzal a tulajdonsággal vagy egyik elem sem rendelkezik, vagy mind a kettő.

Pl.: A 49. és 52. sz. minták az $U = \{ U[kl], U[ill], U[mm] \}$ ismertetőjegy-rendszer alapján nem különböznek egymástól, azaz *nem-különbözőségi* relációban vannak egymással, és ezért egy osztályba sorolhatók, mert

$$\begin{aligned} u[kl/48-52] (49. \text{ sz. m.}) + u[kl/48-52] (52. \text{ sz. m.}) &= 2, \\ u[ill/43-47] (49. \text{ sz. m.}) + u[ill/43-47] (52. \text{ sz. m.}) &= 2, \\ u[mm/8-12] (49. \text{ sz. m.}) + u[mm/8-12] (52. \text{ sz. m.}) &= 2, \end{aligned}$$

a többi összeg pedig 0, pl.

$$u[kl/43-47] (49. \text{ sz. m.}) + u[kl/43-47] (52. \text{ sz. m.}) = 0,$$

stb., de pl. a 46. és 49. sz. minták különböznek egymástól az U ismertetőjegy-rendszer alapján, azaz nincsenek *nem-különbözőségi* relációban, mert pl.:

u [kl/48-52] (46. sz. m.) + u [kl/48-52] (49. sz. m.) = 1.

Felsorolósos osztályozás. Az osztályozás determinisztikus elméletének fő feladatai

Legyen adva az A halmaznak az U ismertetőjegy-rendszer alapján egy partíciója az

$$A[1], A[2], \dots, A[n(U)],$$

osztályokra, és jelöljük ezt $[A:U]$ -val. Ha az $[A:U]$ partíciót valamely másik $[A:U']$ partíciótól függetlenül tekintjük, azt felsorolósos osztályozásnak nevezzük.

A felsorolósos osztályozás többféleképp adható meg:

1. Az egyes osztályok ún. meghatározó szimbólumával:

$$L[p] = (a [1/p1], a [2/p2], \dots, a [l/pl]),$$

ahol $a [i/pi]$ az $U [i]$ ismertetőjegy-készlet azon ismertető jegyének sorszáma, amely 1 értéket vesz fel, ha az a elem beletartozik az $A[i]$ osztályba.

Pl.: Az ismertető jegykészleteknél példaként említett 33. sz. minta adatait felhasználva annak az osztálynak a meghatározó szimbóluma, amelybe a 33. sz. minta tartozik: $L = (4, 4, 3)$.

(Lásd még: 1. táblázat.)

2. Megadhatók az osztályok az ismertetőjegyek táblázata formájában,

3. fa-gráffal,

4. pontokkal az l -dimenziós ismertető jegy tér vetületein és

5. Boole-függvényekkel is.

Ez utóbbiak részletezésére most nem térhetünk ki, csak megemlítjük, hogy az adatok mennyisége és a választott cél szerint érdemes a felsorolt lehetőségek közül választani. Nagyszámú adat gépi feldolgozására az 5. megadási mód a legalkalmasabb.

Ha az $[A:U]$ felsorolásos osztályozás osztályait valamilyen szempont szerint összevonjuk, azaz „magasabb rendszertani kategóriákat” akarunk kapni, az így keletkező $\{[A:U]\}$ partíciót az $[A:U]$ -ből *származtatott felsorolásnak* nevezzük.

Célszerű ezt az összevonást úgy végezni, hogy valamilyen távolságfüggvényt definiálunk az egyes osztályok között, és azokat az osztályokat, amelyekre ennek a távolságfüggvénynek az értéke valamilyen előre megválasztott küszöbértéknél kisebb, összevonjuk. (A cikk második részében erre lesz példa a „cluster” analízis.)

Pl.: Az 1. táblázat adatait felhasználva, a távolságfüggvényt a meghatározó szimbólumok közötti vektoriális távolsággal definiálva és a küszöbértéket 1-nek véve az így származtatott felsorolásban a következő osztályokat különböztethetjük meg:

$A[164]$, $A[182]$, $A[263]$, $A[241]$, $A[344]$, $A[253]$, $A[416]$, $A[434]$, $A[524]$, $A[q]$, ahol q az összes fennmaradó meghatározó szimbólum.

Az a példának vett halmaz speciális tulajdonsága, hogy 9 osztálya olyan „távol” volt egymástól, hogy a származtatott felsorolásban is megőrizték különállásukat, míg a fennmaradó 15 mind egy nagy osztályba volt összevonható ($A[q]$). Az osztályok száma 24-ről 10-re csökkent le.

Ha a küszöbértéket 1 helyett 2-nek választjuk, az előbbi feltételek mellett már csak három származtatott osztály marad: $A[182+231]$, $A[4l6]$, és $A[r]$, ahol r az összes többi meghatározó szimbólum.

1. táblázat

PÉLDA AZ OSZTÁLYOK MEGHATÁROZÓ SZIMBÓLUMMAL VALÓ MEGADÁSÁRA

(Polgárdi 3. sz. mélyfúrás pannóniai korú, röntgenvizsgálatra leadott mintái, lásd: Viczián, I. 1970).

$A[p]$ osztály	$(L[p])$ a elem (mintaszám)
1, 6, 4	30.
1, 8, 2	12.
2, 6, 3	3.
2, 8, 1	46.
3, 4, 4	31.
3, 5, 3	8., 28., 34.
3, 6, 2	17.
4, 1, 6	2.
4, 3, 4	19.
4, 4, 3	22., 26., 33., 42.
4, 5, 2	4., 9., 15., 21., 23., 25., 50.
4, 6, 1	47., 63.
4, 6, 2	37.
5, 3, 2	14., 29.
5, 3, 3	5., 7.
5, 4, 1	24.
5, 4, 2	10., 13., 16., 20., 27., 32., 38., 39., 49., 41., 43., 44., 45., 61., 62.
5, 4, 3	6.
5, 2, 4	11.
5, 5, 1	18., 51., 54., 55., 56., 57., 58.
5, 5, 2	35., 48., 53.
6, 3, 1	60.
6, 4, 1	59.
6, 4, 2	36., 49., 52.

A földtani osztályozás determinisztikus elmélete fő feladatait röviden a következőkben foglalhatjuk össze: Az ismertetőjegy-rendszerek elemzése abból a célból, hogy a kísérleti metodikának és a tartalmi követelményeknek legjobban megfelelő, optimális felsorolásos osztályozásokat kapjunk, és ezeket, valamint a belőlük származtatott osztályozásokat minimális számú logikai változóval tudjuk leírni. Ez utóbbi követelményt az ún. minimalizálás műveletével valósíthatjuk meg.

Pl.: Kimutatható, hogy a fenti ismertetőjegyek számát 18-ról megfelelő módon 11-re csökkentve az A halmaznak az így kapott partíciója ekvivalens marad az A halmaznak (a Polgárdi 3. sz. mélyfúrás röntgenvizsgálatra leadott mintáinak) az 1. táblázatban megadott partíciójával, vagyis pl. kevésbé pontos klorit-meghatározás is elég lett volna a halmaz ugyanilyen részletességű felosztásához.

Diagnosztikus osztályozás. Az osztályozás valószínűségi elméletének fő feladatai

Legyen adva az A halmaz egy U ismertetőjegy-rendszer szerinti $[A:U]$ partíciója az $A[1], A[2], \dots, A[n(u)]$, osztályokra és az A halmaz egy másik U' ismertetőjegy-rendszer szerinti $[A:U']$ partíciója az $A'[1], A'[2], \dots, A'[n(u')]$, osztályokra. Ha azt a kérdést, hogy az A halmaz egy a eleme beletartozik-e $[A:U']$ egy $A'[j]$ osztályába, úgy döntjük el, hogy előzetesen meghatározzuk, hogy ez az elem beletartozik-e $[A:U]$ egy $A[i]$ osztályába, akkor ezt a műveletet diagnózisnak, $[A:U]$ -t diagnosztizáló, $[A:U']$ -t pedig diagnosztizált partíciónak nevezzük.

Pl.: Legyen $[A:U]$ a „Polgárdi 3. sz. mélyfúrásból röntgenvizsgálatra leadott minták halmazának” az

$$U = \{ U[kl], U[ill], U[mm] \}$$

ismertetőjegy-rendszer szerinti partíciója.

Legyen ugyanennek a halmaznak a keletkezési körülményeket jelző valamilyen másik U' ismertetőjegy-rendszer szerinti partíciója: $[A:U']$. (Itt az egyes ismertető jegykészletek lehetnek pl. „a parttól való távolság”, „sótartalom”, „behordott anyag mennyisége”, „lehordási terület jellege”, „hőmérséklet” stb.)

Legyenek e partíció A' osztályai:

$A'[1] = \text{„láp”}$, $A'[2] = \text{„mély-láp”}$, $A'[3] = \text{„lagúna”}$, $A'[4] = \text{„sekély szublitorális”}$, $A'[5] = \text{„mély szublitorális”}$ (Jámbor Á. 1968. beosztása szerint), ahol minden osztályt a fenti készletekből vett ismertetőjegyek egy-egy kombinációja jellemez.

Ekkor, ha az agyagásvány-tartalomból a keletkezési körülményekre következtetünk, úgy fejezhetjük ki, hogy az agyagásvány-tartalom ismertetőjegy-rendszere szerinti partícióval diagnosztizáljuk a keletkezési körülmények ismertető jegy-rendszere szerinti partíciót.

E példa is mutatja, hogy a diagnózis a földtan egyik alapvető jelentőségű módszere.

Ahhoz, hogy a diagnózis különböző eseteit áttekinthessük, minden partícióra definiálhatunk egy-egy entrópia-jellegű mennyiséget, amely az illető felosztás határozatlanságának mértéke, és ezek összehasonlításával dönthetünk a diagnózis lehetőségéről, ill. hatékonyságáról.

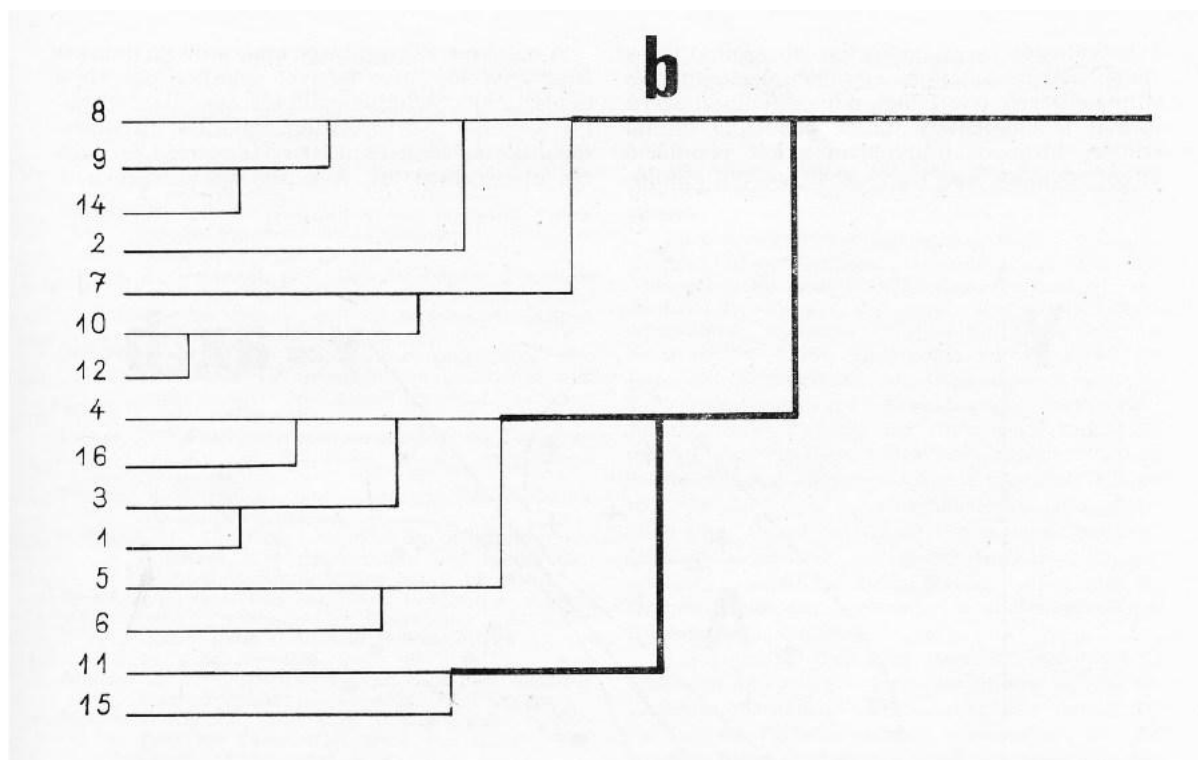
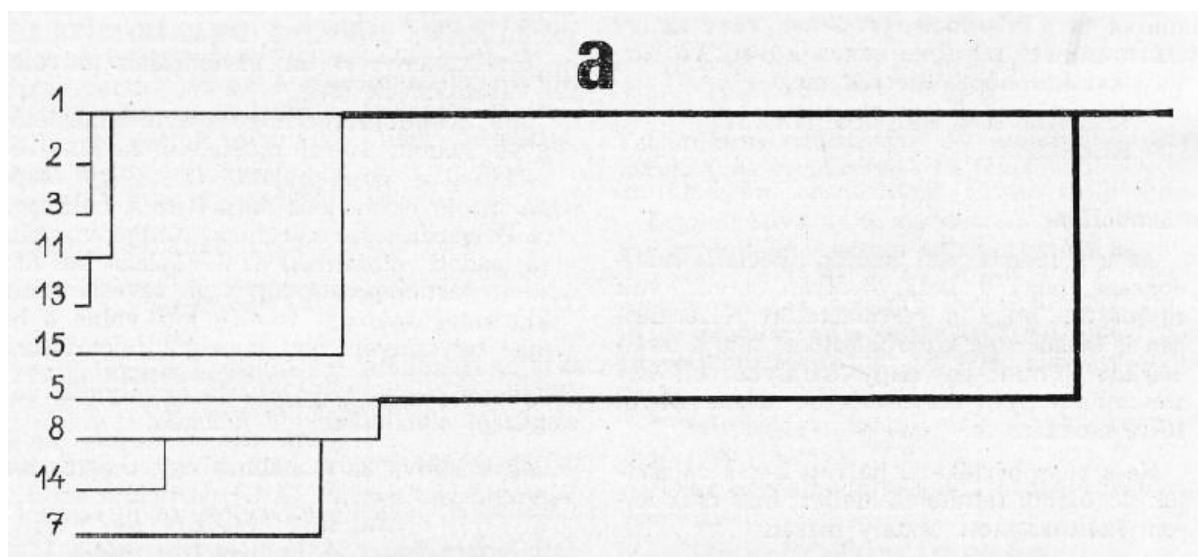
Végső fokon az osztályozás valószínűségi elméletének fő feladata az optimális diagnosztikus eljárás algoritmusának kidolgozása.

2. „Cluster” analízis

A felsorolásos osztályozások fent említett összevonási módszerét alkalmazhatjuk pl. ott, ahol a tulajdonságok sokféle számszerű értéket vehetnek fel, és ezek finom intervallumokra bontásával minden objektum külön osztályba kerülne. Ilyen összevonásra alkalmas számítógépi módszer a „cluster” analízis (fürtös csoportosítás).

Első lépcsőjében minden elempárra kiszámítunk egy $r[i,j]$ hasonlósági mérőszámot, amely megadja, mennyire közel áll egymáshoz az i -edik és j -edik objektum. Több ilyenféle mérőszám van, pl. a korrelációs együttható, az n -dimenziós térbeli távolság, az (n -dimenziós) vektorok közötti szög koszinusza stb.

Alább bemutatandó példáinkban a korrelációs együtthatót használtuk ilyen számként, de a feladat jellegétől függően más esetekben egy másik mérőszám esetleg kissé eltérő, jellemzőbb felbontást adhat. Ha n darab objektumból indulunk ki, ez a számítás egy $n * n$ számból álló táblázatra, ún. hasonlósági mátrixra vezet, amely szimmetrikus az átlóra nézve.



1. ábra. A cluster program eredményei (fadiagramok)

a) A mecseki Sásvölgy a-triász kőzeteinek komponensei (Nagy E. 1969, adatai alapján),

b) A Budai-hegységi földolomit komponensei (Brugger 1940, adatai alapján).

Jelmagyarázat: 1. SiO_2 , 2. TiO_2 , 3. Al_2O_3 , 4. Fe_2O_3 , 5. MnO , 6. CuO , 7. MgO , 8. CaO , 9. SrO , 10. Na_2O , 11. K_2O , 12. $-\text{H}_2\text{O}$, 13. P_2O_5 , 14. CO_2 , 15. SO_3 , 16. oldhatatlan maradék

Következő fázis ezután az objektumok csoportokra bontása a hasonlósági mátrix alapján (erre külön számítógépi programot dolgoztunk ki).

Először a két leghasonlóbb (legkisebb távolságú, legnagyobb korrelációs együtthatójú stb.) objektumot soroljuk egy csoportba, a hasonlósági táblázatban pedig összevonjuk soraikat, ill. oszlopaikat, vagyis egyetlen objektumként kezeljük őket tovább. Most megkeressük a következő legközelebb álló párt, és ugyanígy egy csoportba vonjuk őket össze. Így fokozatosan minden lépésben egy-egy objektumot, vagy objektumcsoportot új csoportba összevonva és a táblázatot egyszerűsítve haladunk.

Számítógéppel végezve a műveletet a gép minden lépés után kiírja az összevont objektumok sorszámát és hasonlóságuk mértékét (lásd 1. ábra). Ez az eljárás addig folyik, míg végül valamennyi objektumot egy csoportba vontunk össze (persze minden összevonásnál egyre kevésbé hasonló tagokat vonunk össze).

A számítógép végül ezeket a csoportokat hierarchikusan elrendezi és kirajzolja a „fadiagramot”, amelynek minden ága egy-egy objektumban végződik, az ágak egyesítésével keletkezett csoportok rokonsági fokát pedig a szétválás helye mutatja (mennél később válik szét két ág, annál inkább rokonok). Ha két szétágazás helye között egy küszöbértéknél kisebb az eltérés, a számítógép egyberajzolja őket többszörös elágazásként.

Pl.: Az 1. ábrán látható két fadiagram a kőzetek egyes vegyi alkotóinak viselkedés szerinti csoportosítását mutatja be két elemzéscsoportban. Az első példa a mecseki Sásvölgy alsó-tri-

ász képződményeinek komponenseit mutatja. Világosan kivehető a két legfontosabb csoport elkülönülése, a szilikátos és karbonátos fázisé. A „szilikátos” csoportban egymáshoz legközelebb a TiO_2 , SiO_2 és Al_2O_3 áll, hasonló viselkedésű még a K_2O és P_2O_5 , míg az SO_3 kissé eltérő. A karbonátcsoportban található a CaO , MgO , CO_2 mellett még a MnO is.

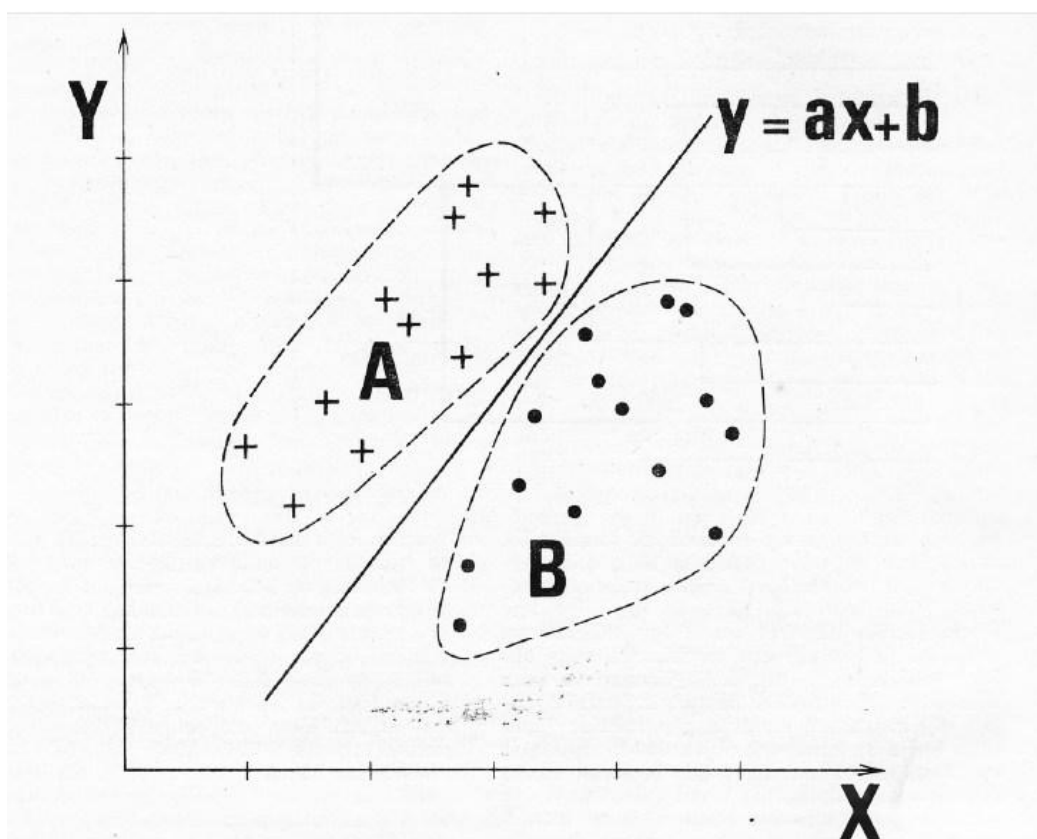
A másik csoportosításhoz a Budai-hegységi felső-triász dolomit elemzéseit használtuk fel. Itt három fő csoport különül el: a „karbonátos”, „oxidos-szilikátos”- és „szulfátos” csoport. A karbonátos csoportba tartozik a CaO , MgO , SrO , CO_2 komponenseken kívül még a TiO_2 , Na_2O és a $-\text{H}_2\text{O}$ is. Az oxidos-szilikátos csoportba sorolta a számítógép az oldhatatlan maradékot, az oldható SiO_2 -t, Al_2O_3 -at, Fe_2O_3 -at és külön, kisebb csoportként az MnO , CuO komponenseket. A K_2O és SO_3 különválik az előbbi fő csoporttól, de az oxidos-szilikátos csoporthoz közelebb áll.

3. Diszkrimináns függvények

A diagnosztizáló osztályozások egyik gyakorlati esetében, amikor a diagnosztizált felosztás osztályait úgy adjuk meg, hogy felsorolunk néhány beléjük tartozó objektumot, diszkrimináns függvényeket használhatunk az ismeretlen hovatartozású objektumok besorolására. A már ismert (valamelyik populációba tartozó) objektumokon és a még csak most besorolandó objektumokon is mérjük ugyanazt az n darab tulajdonságot.

Minden egyes objektumot, vagyis minden n -dimenziós megfigyelést ábrázolhatunk az n -dimenziós tér egy pontjaként, úgy, ahogyan a kétváltozós megfigyeléseket ábrázolhatjuk a síkon. Ebben a térben egy populáció mintái ellipszoidszerű pontfelhőt

adnak (ellipszisszerű formát a síkon), egy másik populáció mintái szintén, bizonyos irányokban a két populáció átfedi egymást, más irányokban viszont elkülönül. A diszkrimináns analízis feladata, hogy találjon egy olyan egyenletet (a változók egy lineáris, ill. kvadratikus kombinációját), amely megad egy a két halmazt effektíven elválasztó felületet. A számítás módszere a többszörös regresszióhoz hasonló.



2. ábra. Diszkrimináns függvény elhelyezkedése (kétdimenziós elvi vázlat). Jelmagyarázat: X, Y tulajdonságok (koordináta-tengelyek); A, B populációk; $y = ax + b$ az elválasztó diszkrimináns függvény.

Először megvizsgáljuk a populációk különbözőségét sokdimenziós átlagaik „távolságaként”. Ha megállapítottuk különbözőségüket, vagyis, hogy átlagaik jobban különböznek, mintsem az a mintavétel véletlen hibája volna, kiszámítjuk a két populációt elválasztó felület egyenletét. A további, ismeretlen mintá-

kat ezután már sokdimenziós „helyük” kiszámításával könnyen a megfelelő csoportba sorolhatjuk.

A módszer előnye, hogy csak a diszkrimináns függvény előállítása igényel számítógépet. Ha a képlet adott, a felhasználónak csak be kell helyettesítenie a besorolandó objektum tulajdonságainak értékeit és néhány összeadás és szorzás után célhoz jut. A képlet számértéke u. i. azt adja meg, hogy az objektum a terület melyik oldalára esik, vagyis melyik populációhoz tartozik. Diszkrimináns függvényeket kettőnél több populáció elkülönítésére is használnak (Burnaby 1963).

A földtanban már az ötvenes években kezdték alkalmazni ezt az eljárást például kőolajtartalmú és meddő szintek, törmelékes kőzetfáciesek, ércartalmú és meddő képződmények, különböző provinciába tartozó bazaltos kőzetek stb. szétválasztására (Griffiths 1966, Demirmen 1969), hogy a sok példa közül csak néhányat említsünk. Az eredmények felhasználására az utóbbi időben több hazai példa is van (Bérczi 1969, Dienes 1970).

IRODALOM

Bérczi I. 1969: Az algyői felső-pannóniai homokkő üledékföldtani vizsgálata. Földtani Közlöny 99. no. 4. pp. 337—350.

Brugger, F. 1940: A budakörnyéki dolomitok kőzetkémiai vizsgálata. Math. Term.-tud. Ért. 59. no. 2 pp. 619—642.

Burnaby, T. P. 1965: Distribution-free quadratic discriminant functions in paleontology. Computer Contr. no. 7. pp. 70—77.

Chayes, F., Velde, D. 1965: On distinguishing basaltic lavas of Circumoeanic and Oceanic-Island type by means of discriminant functions. Amer. J. Sci. 1:63. pp. 206—222.

Demirmen, F. 1969: Multivariate procedures and FORTRAN IV. program for evaluation and improvement of classifications. Computer Contr. no. 31. 51. p.

Dienes I. 1970: Klinopiroxén megakristályok a medvési bazaltból. M. Áll. Földt. Int. Évi Jel. 1968-ról, p.125-130.

Fotiadi, É. É. (szerk.) 1967: Geologija i matematika. Nauka, Novoszibirszk.

Griffiths, J. C. 1966: Application of discriminant functions as a classification tool in the geosciences. Computer Contr. no. 7. pp. 48—52.

Howd, F. H. 1964: The Taxonomy Program — a Computer Technique for Classifying Geological Data. Quart. Colorado School Mines, 59. no. 4—A. pp. 207—223.

Jámbor Á. 1968: Az üledékes agyagkőzetek osztályozásának szöveti szempontjai. Kézirat, Bp.

McCammon, R. B. 1958: The Dendrograph: A New Tool for Correlation. Geol. Soc. Amer. Bull. 79. no. 11. pp. 1663—1670.

Nagy E. 1969: A Mecsek hegység triász időszaki képződményei. M. Áll. Földt. Int. Évk. 51. no. 1 158. p.

Parks, J. M. 1966: Cluster analysis applied to multivariate geological problems. J. Geol. 74. no. 5/2. pp. 703—715.

Pusztovalov, N. L. 1962: [ld.: Fotiadi, É. É. (szerk.) 1967.]

Sokal, R. R., Michener, C. D. 1958: A Statistical Method for Evaluating Systematic Relationships. Univ. Kansas Sci. Bull. 38. no. 2. pp. 1409—1438.

Valentine, J. W., Peddicord, R. G. 1967: Evaluation of fossil assemblages by cluster analysis. J. Paleont. 41. pp. 502—507.

Viczián, I. 1970: Faziesprobleme im ungarischen Pannon auf Grund des Tonmineralgehalts. Ber. Dtsch. Ges. Geol. Wiss., B. Min. Lagerst. vol.15, no.1, p.75-82.

* *Előadásként elhangzott: az MFT Matematikai-Földtani Szakcsoport 1970. január 30-i ülésén.*