Rendszerváltozók strukturaelemzéséhez készült GRAF1 programcsomag

Dénes Tamás matematikus
Budapest, 1984.

Az [1], [2] dolgozataimban a rendszerváltozók strukturális leírásának, illetve egymásrahatásának elméletével és módszertanával foglalkoztam. Jelen dolgozat a rendszerek empirikus megközelítésével, azaz a leírásukra szolgáló mért rendszerváltozók strukturális leírásával foglalkozik. A *normált fedési-mutatókra* épülő *n-változós struktúraelemzési GRAF1 nevű programcsomagot* FORTRAN programozási nyelven készítettem el.

1. Páros változóstruktura elemzés

FELADAT:
A kijelölt *n* darab változó irányított egymásrahatásának páronkénti vizsgálata, azaz a normált fedési-mutatók kiszámítása és egy *nxn méretű fedési-mutató mátrixba foglalása*.

EREDMÉNY:
- A kijelölt *n* darab változó páronkénti strukturális összehasonlításából nyert *normált fedési-mutatók antiszimmetrikus mátrixa*.
- Az *nxn méretű normált fedési-mutató mátrix* vizuális elemzéséhez segítséget myújtó szintenkénti *képmátrix* (10 szint).
- Az *nxn méretű normált fedési-mutató mátrix* automatikus (programozott) elemzésével a GRAF1 program által javasolt 2-dimenziós statisztikai táblák.

ELVI LEÍRÁS:

1.1.
Legyen az empirikus vizsgálat során $H=\{h_1,h_2,...,h_m\}$ a rendszert (jelenséget) hordozó elemek halmaza (minta elemek). $X=\{x_1,x_2,...,x_n\}$ a rendszert leíró (mért) változók halmaza, valamint bármely $x_i \in X$ változóhoz tartozzon egy meghatározott $K_i=\{k_{i_1},k_{i_2},...,k_{i_{n_i}}\}$ kódhalmaz (értelmezési tartomány), amelyben a változó lehetséges (n_i-féle) mért értékei vannak. Ha a h_j elem x_i változó szerinti realizációját (mért értékét) $h_j(x_i)$-vel jelöljük, akkor $h_j(x_i)\in K_i$.
Ekkor az [1] (2.14)-ben definiált *normált fedési-mutatókat* kiszámíthatjuk bármely két $x_i, x_j \in X$ változóra. Az empirikus adatokat egy adott változópár esetén, a [2]-beli 1.1. ábrában foglaltuk össze, amelynek jelöléseit használva a kiszámítás képleteit a [2] (4.1) és (4.2) levezetésekben adtuk meg. Ezeket a *H* mintán mért adatokra alkalmazva $FN(x_i,x_j)$ és $FN(x_j,x_i)$ kiszámítása a következő:

(1.1)

$$A_i = \sum_{v=1}^{n_j} a_{iv} \ , \ \sum_{u=1}^{n_i} A_u = m \Rightarrow$$

$$\Rightarrow FN(x_j, x_i) = \frac{\left(\sum_{u=1}^{n_i}\sum_{v=1}^{n_j} a_{uv}^2\right) n_i n_j - \left(\sum_{u=1}^{n_i} A_u^2\right) - m(n_i n_j - 1)}{(n_i n_j - 1)\left(\sum_{u=1}^{n_i} A_u^2\right) - m(n_i n_j - 1)}$$

(1.2)

$$B_j = \sum_{u=1}^{n_i} a_{uj} \ , \ \sum_{v=1}^{n_j} B_v = m \Rightarrow$$

$$\Rightarrow FN(x_i, x_j) = \frac{\left(\sum_{u=1}^{n_i}\sum_{v=1}^{n_j} a_{uv}^2\right) n_i n_j - \left(\sum_{v=1}^{n_j} B_v^2\right) - m(n_i n_j - 1)}{(n_i n_j - 1)\left(\sum_{v=1}^{n_j} B_v^2\right) - m(n_i n_j - 1)}$$

Itt már az elemzések szempontjából „élő” alsó határt jelent a normált fedési-mutató nulla értéke, ami a két változó egyenletes közös eloszlásakor áll elő. Azaz ha az egyik változó egy realizációjának megváltozása azonos valószínűséggel jár együtt, a másik változóbármely realizációjának megváltozásával. Röviden:
A normált fedési-mutató nulla értéke azt jelenti, hogy az adott két változó (adott irányban) hatásmentessége a „véletlen hatásával” egyenlő.[1]

Az $X = \{x_1, x_2, ..., x_n\}$ változóhalmaz minden x_i, x_j (*i=1,2,…,n* , *j=1,2,…,n* , *i*≠j) változópárjára kiszámítjuk az (1.1) és (1.2) képletek szerinti normált fedési-mutatókat, majd ezeket egy *n*x*n-es* $V = \lfloor v_{ij} \rfloor$ *fedési mátrixban* (lásd [1] (4.1)) tárolja a program. A [2]-beli (4.5) levezetésből adódik, hogy a *V* mátrix főátlójában, amely egy adott x_i változó önmagával vett $FN(x_i, x_i)$ normált fedési-mutatóját reprezentálja, mindig az *1* érték szerepel (lásd az alábbi 1.1. ábrát).

[1] A fentiek alapján a szokásos „függő” és „független” változó fogalmak helyett, két új fogalmat vezettünk be. Hiszen modellünkben minden változó egy adott változókörnyezetben „kettős életet él”, azaz bizonyos mértékig hatást gyakorol a többi változóra, ugyanakkor a többi változó is hat rá, tehát részben *ható-típusú*, részben *függő-típusú*.

V	x_1	...	x_j	...	x_i	...	x_n
x_1	*1*	...	$FN(x_1,x_j)$	...	$FN(x_1,x_i)$	...	$FN(x_1,x_n)$
...	...	*1*	...	...	...	...	...
x_j	$FN(x_j,x_1)$	...	*1*	...	$FN(x_j,x_i)$	...	$FN(x_j,x_n)$
...	...	...	...	*1*	...	...	...
x_i	$FN(x_i,x_1)$	...	$FN(x_i,x_j)$	...	*1*	...	$FN(x_i,x_n)$
...	...	...	...	...	...	*1*	...
x_n	$FN(x_n,x_1)$	...	$FN(x_n,x_j)$	...	$FN(x_n,x_i)$	...	*1*

1.1. ábra

1.2. A *V* fedési mátrix képi megjelenítése

A *V* mátrixban szereplő *normált fedési-mutatókat szintekre oszthatjuk*, így a változók hatás-struktúráját az adott szinten leíró *gráf szomszédossági mátrixát* kapjuk, amelyhez egyértelműen rendelhető egy ***digitális kép***.
Mivel a normált fedési-mutatók értéke mindig *0* és *1* közé esik, legyen a vizsgált szint értéke $0 \le \delta \le 1$, ekkor *V*-hez rendeljük hozzá az (1.3) szerinti *V'* mátrixot:

(1.3)
$$V'(i,j) = v'_{ij} = 0 \quad ha\ v_{ij} \langle \delta$$
$$V'(i,j) = v'_{ij} = 1 \quad ha\ v_{ij} \ge \delta$$

Ekkor a *KV'* képmátrix így áll elő:

(1.4)
$$KV'(i,j) = k'_{ij} = \ feh\acute{e}r \quad ha\ v'_{ij} = 0$$
$$KV'(i,j) = k'_{ij} = \ fekete \quad ha\ v'_{ij} = 1$$

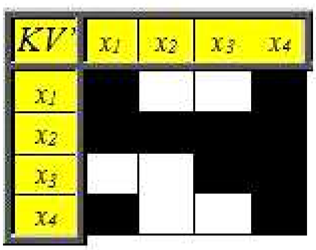
PÉLDA az x_1, x_2, x_3, x_4 változók esetén, ha $\delta = 0.6$ *(lásd 1.2.-1.4. ábra)*

V	x_1	x_2	x_3	x_4
x_1	1	0.3	0.5	0.7
x_2	0.6	1	0.8	0.7
x_3	0.4	0.5	1	0.9
x_4	0.8	0.3	0.2	1

1.2. ábra

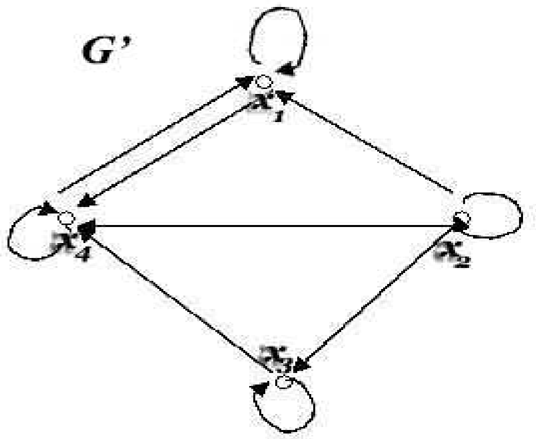
V'	x_1	x_2	x_3	x_4
x_1	1	0	0	1
x_2	1	1	1	1
x_3	0	0	1	1
x_4	1	0	0	1

1.3. ábra

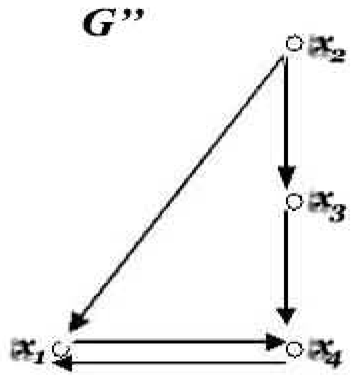


1.4. ábra

A *V'* szomszédossági mátrixhoz tartozó *G'* gráf az x_1, x_2, x_3, x_4 változók által meghatározott rendszer $\delta = 0.6$ szinthez tartozó struktúrája (lásd 1.5.ábra). Ha a *G'* gráfból elhagyjuk az úgynevezett tranzitív és hurok éleket, akkor megkapjuk a gráf vázát képező *G''* gráfot, amely jóval áttekinthetőbb, a *változók hatási hierarchiája* szempontjából (lásd 1.6.ábra).



1.5. ábra



1.6. ábra

1.3. A *V* fedési mátrix automatikus elemzésével javasolt 2-dimenziós statisztikai táblák

A GRAF1 program kiszámítja a *V* mátrix nullánál nagyobb elemeinek négyzetátlagát, majd az így kapott értéknél nagyobb elemek aritmetikai átlagát, mint küszöbértéket tekinti. Azokhoz a változópárokhoz javasolja a kétdimenziós statisztikai táblák elkészítését, amelyekhez kiszámított normált fedési-mutató érték nagyobb ennél a küszöbértéknél.

2. Teljes változóstruktura elemzés

FELADAT:
A kijelölt *n* darab változó *X* halmazában milyen összefüggő változócsoportok határozhatók meg?

EREDMÉNY:
A *X* változóhalmaz osztályokba sorolása, különböző összefüggési szinteken.

ELVI LEÍRÁS:

A feladat megoldásához alkalmazzuk az [1] 2.1.-2.2. pontjaiban bemutatott gráf modellt. Azaz hozzárendeljük mindegyik $x_i \in X$ változóhoz a struktúráját modellező $G_i=(P,E_i)$ gráfot. Majd képezzük ezen gráfok multistrukturális összegét (jele: $G_{1,2,\ldots,n}=(P,E)$) az alábbi módon:

(2.1) $$G_{1,2,\ldots,n} = G(P,E) = G_1 \oplus G_2 \oplus \ldots \oplus G_n$$

(2.2) $$\xrightarrow{1.1.} H \to P = \{p_1, p_2, \ldots, p_m\}$$

(2.3) $\xrightarrow{[1](2.4)} E = E_1 \bigcup E_2 \bigcup ... \bigcup E_n$

ahol $\bigcup$ az általam ***bővített unió***-nak nevezett halmaz művelet, amelyet a következőképpen értelmezünk:

(2.4) $A \bigcup B = (A \cup B) \cup (A \cap B)$

Vagyis a bővített únió duplán tartalmazza a metszetbeli elemeket. Ez a gráfok esetén azt jelenti, hogy a (2.1)-szerinti összegzés olyan *n-színnel színezett gráfot* állít elő, amely a tag-gráfok összes élét tartalmazza. Szemléletesen tehát az *n* változóból álló rendszert olyan $G_{1,2,\ldots,n}$ multistruktúrával modellezzük, mint egy *n* rétegű szendvics.
Eljárásunk az lesz, hogy megkeressük azokat a rétegeket, amelyek eléggé hasonlítanak egymáshoz és az ezek által képviselt változók fognak egy osztályba tartozni.
Természetesen a gráfok (rétegek) hasonlósága különböző szintű lehet, így az 1. feladatnak megfelelően 10 szinten adjuk meg a kijelölt *X* változóhalmaz osztályokba sorolását.

3. Cluster-struktura elemzés

FELADAT:
A *H* mintából milyen összetartozó csoportok alkothatók az *X*-ből kiválasztott változók szerint?

EREDMÉNY:
- A kijelölt változók és megadott (strukturaszint) paraméterek szerint alkotható lefedő struktúrák (osztályozások) megadása. Az osztályokba tartozó elemek felsorolása.
- Az egyes osztályok típusa. Azaz az adott osztályba tartozó elemek, osztályra jellemző változók szerinti reprezentációi.

ELVI LEÍRÁS:

A (2.1)-(2.3) összefüggések szerinti modellből kiemeljük a kijelölt változókhoz tartozó „színű” éleket. Az így kapott multigráf éleit a paramétereknek megfelelően részhalmazokba soroljuk, majd ezeken az él-részhalmazokon elvégezzük az *„adott sűrűségű részgráf keresést”* (lásd [3], [4]), amíg ilyet találunk. Ez lesz a minta egy lefedése, vagy osztályozása. Az eljárást elvégezzük az összes él-részhalmazra, így a megadott változók szerint esetleg a minta több lefedését (osztályozását) kapjuk meg, a kijelölt változók struktúrájától függően.

Hivatkozás jegyzék

[1] Dénes Tamás: *Rendszerváltozók strukturaelemzésének elmélete és gráfelméleti modellje, Budapest, 1984.*
http://www.titoktan.hu/_raktar/_e_vilagi_gondolatok/Rendszervaltozok-str-elemzesenek-elmelete.pdf

[2] Dénes Tamás: *Rendszerváltozók strukturaelemzésének módszertana (A fedési-mutató alkalmazása), Budapest, 1984.*
http://www.titoktan.hu/_raktar/_e_vilagi_gondolatok/Rendszervaltozok-str-elemzesenek-modszertana.pdf

[3] Dénes Tamás: *Gráfok maximális sűrűségű részgráfjairól (Algoritmus maximális teljes részgráf megkeresésére), Budapest, 1977.*
http://www.titoktan.hu/_raktar/_e_vilagi_gondolatok/Grafok-maximalis-surusegu-reszgrafjairol.pdf

[4] Szabó Imre: *Adott gráf teljes részgráfjaival kapcsolatos feladatok számítógépes megoldása*, Szakdolgozat, ELTE TTK Programozó matematikus szak, 1978. Konzulens: Dénes Tamás