

Gödöllői Agrártudományi Egyetem és Országos Korányi TBC és
Pulmonológiai Intézet

Tünetpárok számítógépes vizsgálatának egyszerű módszere

Juhász Bálint, Tamássy Józsefné és Naszlady Attila

Az elmúlt évi kollokvium során a tisztelt résztvevők megismerhették azt a módszert, amelyet a betegek adatainak felvételére az Országos Korányi TBC és Pulmonológiai Intézet Belgyógyászati Osztályán már korábban létrehoztak. Ennek lényege az az e célra rendszeresített N-féle kettős kórlap, amely az anamnézist, a vizsgálatok eredményeit és a diagnózist (diagnózisokat) tartalmazza. A kórlapok kitöltésének alapjául az alkalmasan megszerkesztett kérdőív szolgál.

Az Intézetben 1973 óta az említett kórlapokra történik az adatfelvétel, és ez év elejére gyűlt össze annyi adat, amely lehetővé tette feldolgozásukat. Miután az Intézet és a Gödöllői Agrártudományi Egyetem Matematikai és Számítástechnikai Intézetének Számítóközpontja már korábban is munkakapcsolatban volt, ezért ezt a munkát is ott végeztük el. E célra egy ODRA 1204 kis teljesítményű lengyel számítógép állt rendelkezésünkre. A gép lyukszalag-orientált, a lyukszalagos perifériákon kívül csak egy sornyomatatóval és dob háttérmemóriával van összekapcsolva. Bár a kórlapok optikai jelolvasóval történő közvetlen beolvasás útján eleve alkalmasak a gépi adatfeldolgozásra - ezt az elegáns utat a fentiek miatt nem választhattuk, és közbenső adathordozóra, lyukszalagra kellett az adatokat átültetni.

Az adatmennyiség viszonylag nagy, ezer beteg adatai, betegenként 80-100 adat is előfordulhat. Az elvileg lehetséges adatok száma ennél jóval nagyobb: betegenként 2000. Ekkora adatmennyiség lyukasztásánál óhatatlan a tévesztés. Hogy a hibákat kiszűrjük, kettős ellenőrzést végeztünk. Egyrészt a lyukasztás bizonyos redundanciával történt. Ez lehetővé tette a szalagok gépi ellenőrzését, ill. a

hibák felismerését. Másrészt a szalagokról készült listákat összeolvastuk az eredeti adatlapokkal. A hibás szalagokat javítottuk, majd számítógéppel állítottuk elő a végleges, rendezett adatszalagokat.

A munka következő fázisa az adatok bevitele volt a háttérmemóriába. Itt a problémát az jelentette, hogy minden betegnél bármelyik adat előfordulhat a 2000 közül. Szerencsére ezek mindegyike igen-nem információ, 1 bit értékű, így egy adatsort (20 bit) tároltunk egy gépi szóban. Egy beteg adatai 100 gépi szóban, a teljes adatmennyiség pedig 100 000 gépi szóban fért el.

A betegadatok feldolgozásával kapcsolatban a Korányi Intézet részéről az alábbi igények merültek fel:

1. Állapítsuk meg az előforduló diagnózisok gyakoriságát a teljes populációban.
2. Határozzuk meg az összes tünet gyakoriságát a teljes populációban.
3. Állapítsuk meg a tünetek gyakoriságát diagnózisonként.
4. Határozzuk meg a diagnózisok gyakoriságát tünetenként.
5. Állapítsuk meg kiemelt tünetek és diagnózisok együttes gyakoriságát.
6. Végül a gyakrabban előforduló tünetek páronkénti összefüggését, ill. függetlenségét állapítsuk meg adott diagnózisok esetén.

Ez alapjául szolgálhat egy további feldolgozásnak, melynek eredményeképpen az egyes tünet-együttesek meghatároznák a valószínű diagnózisokat, ill. a szóhajóhető diagnózisok valószínűségeit, és a számítógép további vizsgálatokat javasolhat.

A fenti feladatokat elvégeztük. Az alábbiakban csak a legutólsóra térünk ki.

Kiválasztottuk a 60-nál nagyobb számban előforduló diagnózisokat és ezek mindegyikéhez külön-külön azokat a tüneteket, amelyek az adott diagnózisú egyének legalább 70 %-ánál előfordultak.

Meghatároztuk ezeknek a tüneteknek a páronkénti együttes előfordulási gyakoriságát. Így minden diagnózishoz egy gyakoriság mátrixot (M) állítottunk elő, amely abszolút számokban adta meg a gyakoriságokat. Ez azonban nem eléggé áttekinthető módon jellemzi az egyes tünetpárok összefüggését - bár minél nagyobb, annál inkább utal rá. Ha mindkét tünet azonos számban fordul elő, és ez megegyezik az együttes előfordulási számmal is, akkor az adott mintában az összefüggésük 100 %-os. Másrészt, ha a t_1 tünet p_1 %-ban fordul elő a mintában, a t_2 tünet pedig p_2 %-ban, és a t_1 és a t_2 együtt $(p_1 \cdot p_2)/100$ %-ban, akkor - a tünetek egyenletes eloszlását feltételezve - azt mondhatjuk, hogy a mintában függetlenek egymástól. Végül ha az együttes előfordulás $p < (p_1 \cdot p_2)/100$, akkor az összefüggés negatív.

Ezen megfontolások alapján az alábbi relatív számot képeztük:

$$i = \begin{cases} \frac{t-s}{m-s} \cdot 100 & , \quad \text{ha } m > s \\ 100 & , \quad \text{ha } m = s \end{cases}$$

ahol:

$$t = M[i, j]$$

a t_i és t_j tünetek együttes előfordulási száma,

$$s = M[i, i] \cdot M[j, j] / b$$

a t_i és t_j tünetek számított együttes előfordulása a függetlenséget feltételezve,

$$b = \text{betegszám}$$

az adott diagnózisu betegek száma,

$$m = \max(M[i, i], M[j, j])$$

a t_i és t_j tünetek közül a gyakoribbnak az előfordulási száma.

Az i index a tényleges előfordulások és a számított előfordulások különbsége, a tünetpár gyakoribb tünetének gyakorisága és a számított gyakoriság különbségének százalékában. Az i akkor és csak akkor 100 % értékű, ha minden egyes t_i tünet a t_j tünettől együtt fordul elő, és viszont. Ha a tényleges tünetszám megegyezik a számítottal, $i = 0$, és negatív értékű, ha a tényleges tünetszám kisebb a számítottnál. Az eredmények közül egyet kiragadva szeretnénk példaképp bemutatni.

Az 1. táblázat a 2200 kódu diagnózis gyakoriság mátrixát (M) és i indexeit tartalmazza oly módon, hogy a mátrix fő átlója mentén az egyes tünetek abszolút gyakorisága található, fölötte a páronkénti együttes előfordulások száma, alatta pedig az összefüggésükre jellemző i indexek.

A 2200 kódu diagnózisnál, amely mellkasi deformitást jelent, a tünetek között szerepel az IRV (inspiration reserve volume), valamint a FEV1 (forced expiration volume). Az IRV 127 esetben, a FEV1 is 127 esetben, együttesen viszont 114 esetben fordulnak elő. Az i index értéke 65, amely a nyilvánvaló összefüggésre utal.

A munka csak kezdő lépésnek tekinthető. A vizsgált diagnózisok és tünetek kiválasztása kizárólag előfordulási gyakoriságuk alapján történt, ezért olyan, itt érdektelenek is belekerültek, mint a "rendezett lakáskörülmények", vagy "nincs szembetegsége", valamint egyéb triviálisan is összefüggő adatok. A további vizsgálatokban fokozottabban érvényre kell juttatni az orvosi szempontokat.

Összefoglalás

Az Országos Korányi TBC és Pulmonológiai Intézet Cardiológiai Osztályán 1973 óta a betegek adatfelvételezése rendszeresen az N-féle kórlapra történt. 1976-ig több mint ezer beteg anyaga gyűlt össze (betegenként 2000 féle adatlehetőséggel) és a Gödöllői Agrártudományi Egyetem Számítéközpontjában megkezdődött feldolgozásuk.

Ennek során diagnózis és tünetgyakoriság vizsgálatok készültek a teljes populációra nézve, valamint különféle bontásban, ill. csoportosításban. A szerzők egy i indexet képeztek, amely a tünetek függetlenségére, ill. összefüggésére jellemző relativ szám. A függetlenség vizsgálat alapjául szolgálhat egy további feldolgozásnak, amelynek eredményeképpen az egyes tünet-együttesek meghatároznák a legvalószínűbb diagnózisokat, ill. ezek valószínűségeit.

1. táblázat

TÜNETPÁROK GYAKORISÁGA ÉS SZÁZALÉKOS ÖSSZEFÜGGÉSINDEXEK

Diagnózis = 002200 Betegszám = 179 Tünetszám = 18

	1	2	3	4	IRV (5)	6	FEV1 (7)	8	9	10	11	12	13	14	15	16	17	18
1	125	87	89	122	105	113	103	120	112	93	114	100	108	108	114	114	113	115
2	-8	129	113	127	90	104	93	124	103	102	124	95	102	114	105	100	107	103
3	-1	56	128	127	88	99	89	123	96	105	120	85	98	112	98	97	101	98
4	-2	0	2	176	125	138	126	168	138	135	161	124	134	156	139	139	140	141
IRV (5)	43	-4	-8	0	127	115	114	121	114	99	118	104	112	111	116	118	117	119
6	34	6	-5	-2	37	141	117	135	129	104	130	114	130	123	133	129	136	135
FEV1 (7)	37	4	-5	2	65	41	127	121	114	102	117	107	115	111	117	116	119	119
8	1	2	1	-2	-1	1	-1	171	135	131	158	119	132	153	136	135	138	138
9	32	4	-12	-2	34	60	34	1	141	106	130	114	125	127	131	130	130	131
10	-6	9	18	1	5	-12	12	0	-6	137	126	93	102	121	106	106	107	108
11	-1	13	6	-2	3	2	1	9	2	1	164	117	127	145	131	132	132	134
12	34	13	-11	2	40	37	48	-1	37	-6	5	125	114	110	115	113	116	115
13	30	9	0	-2	37	67	45	3	52	-9	4	44	137	119	128	124	133	131
14	-5	0	-2	3	-2	-4	-2	10	8	0	1	-1	5	158	126	123	125	127
15	35	7	-9	-2	37	70	39	1	64	-8	3	37	36	2	142	130	134	138
16	37	-4	-10	1	44	60	39	1	63	-6	8	34	45	-4	60	141	130	133
17	30	10	-3	-2	37	77	42	4	57	-7	3	37	73	-4	70	57	143	137
18	33	-2	-12	-2	40	71	40	1	57	-7	6	33	62	-0	80	64	76	144

