

MTA SZTAKI

A hospitalizált morbiditás számítógépes statisztikai vizsgálatainak néhány módszertani kérdése

Ruda Mihály

Az egészségügyi ellátás, szervezés, vezetés és az orvostudományi kutatások területén egyaránt fontos szerepet játszik a statisztikai adatfelvétel, valamint az adatok /számítógépes/ feldolgozása és értékelése. A feldolgozások hatékonyságát a jól megválasztott módszerek biztosítják. Célszerű tehát az orvosstatisztikai vizsgálatok technikai /módszertani/ kérdéseivel is foglalkozni /ld. pl. [9]/. Ezen a kérdéskörön belül számtalan részterületet különböztethetünk meg. Az előadás nagytömegű morbiditási adat statisztikai vizsgálatának számítástechnikai kérdéseit tárgyalja. Módszertani megfontolások és kritikai megjegyzések szerepelnek a morbiditási adatok logikai vizsgálatával, szelektálásával, statisztikai táblázatokba történő foglalásával és matematikai statisztikai értékelésével kapcsolatban. Az előadás célja nem pontos tételek kimondása, hanem inkább problémák felvetése és a tapasztalatok összefoglalása. Javaslataink és megál-

lapításaink alapját az ÁSZSZ HWB gépén már több mint egy éve folyó országos kórházi morbiditási feldolgozások tapasztalatai képezik. Az 1974-77. évi adatfelvételeket az ESZTIK munkatársai dolgozták fel a SZTAKI által kialakított SIS77 számítógépes rendszer segítségével /erről a rendszerről ld. pl. [7], [12], [13]/.

Az előadásban általános kérdésekkel, és speciálisan a SIS77 statisztikai adatfeldolgozó rendszerhez kapcsolódó problémákkal foglalkozunk. Természetesen csak néhány szempont kiragadásáról lehet szó, a rendszer ismertetése egy nagyobb tanulmány feladata [13].

1. Adatbázisok - statisztikai feldolgozások

Adatbázison általában adatnyilvántartó rendszereket értenek /pl. repülőgép helyfoglalás, bankszámla nyilvántartás stb./, és eddigi tapasztalataink szerint a felhasználók nem hajlandók elszakadni ettől a szemlélettől, sőt tisztán statisztikai feladatok megoldásakor is ilyen rendszerekkel, illetve a hozzájuk kapcsolódó fogalmakkal operálnak /adatbank, lekérdezés stb./. Statisztikai feldolgozásoknál célszerű a lehető legegyszerűbb file-szerkezetek alkalmazására törekedni /ld. [8]/. Természetesen ezek az egyszerű szerkezetek /pl. szekvenciális file-ok/ is alkothatnak adatbázist.

Bonyolult adatbázisok felépítésében a fő szerepet

az adatkapcsolatokat leíró eszközök játszáék. Ezeket nem feltétlenül fizikailag létező mutatókkal, kulcsokkal kell modellezni, hanem a kapcsolatokat leíró logikai strukturát felhasználva, egyszerű szekvenciális file-okat is használhatunk. Ilyen irányu lehetőségekkel kapcsolatos eredményeket tárgyal [3], [6] is.

2. Információszołgáltatás - matematikai statisztikai értékelés

A közismert és kiváló tulajdonságokkal rendelkező statisztikai rendszerek, mint például a biológiai kutatások igényeiből kiinduló BMD /és továbbfejlesztése a BMDP/ programcsomag, vagy az elsősorban szociológiai vizsgálatok támogatására készült SPSS vagy OSIRIS rendszerek, vagy akár a hazai felhasználók körében is már régebben ismert oslói illetve IBM szubrutin és programgyűjtemények elsősorban matematikai statisztikai feladatok gyors és könnyű megoldását támogatják. Kisebb méretű és egyszerűbb táblázási feladatokat is elvégezhetünk velük.

Nagy és bonyolult adatrendszerek esetében szükség van olyan eljárásokra, amelyek - például a fent említett rendszerek számára - előkészítik a feldolgozandó alapadatokat. Statisztikai /matematikai statisztikai/ vizsgálatoknál, nagy minta esetén nem célszerű újra és újra az alapadatokból kiindulni. Elegendő bizonyos gyakoriságokat, kód-

összegeket, négyzetösszegeket tárolni. Ezt a feladatot láthatja el például a SIS77 rendszer. Hasonló /egy általános adatkezelő rendszer kialakítására irányuló/ törekvésekről számol be a [16] cikk is.

3. Programozási nyelvek

A SIS77 létrehozását megelőzte az 1972-73. évi kórházi morbiditási adatok feldolgozása /erről a munkáról az előző kollokviumokon már több alkalommal beszámoltunk/. Ez a feldolgozás az Akadémia CDC 3300-as gépén történt. A számítógépes rendszer mintegy 80, elsősorban COBOL nyelvű programból állt. Néhány esetben azonban FORTRAN programokat alkalmaztunk. Figyelembe véve az itt szerzett tapasztalatokat, valamint a SIS77 rendszer tervezésekor felmerült egyéb szempontokat, a FORTRAN nyelv alkalmazása mutatkozott előnyösebbnek.

A következő szempontokat sorolhatjuk fel:

Egy programozási nyelv megválasztásakor figyelembe kell venni a szervezők, programozók beállítottságát, képzettségét. Az elvont formulákban történő gondolkodásra hajlamosabb /matematikus/ felhasználót másként kell tekinteni mint azt, aki a verbális gondolkodási síkon mozog otthonosabban /pl. szervező/.

A COBOL nyelv kétségekívül sok előnnyel rendelkezik, elsősorban a file-kezelés, az adatdefiníciók, az adatmozgatás gyorsasága és az output /tábla/ formátumok kezelése területén. Ha azonban bonyolult logikai kapcsolatokat kell leírni, és számítási eljárások optimális lefolyását kell biztosítani /ld. pl. [10], [11]/, - ami nagy és bonyolult rendszereknél valóban szükséges - akkor hasznosabb a jobban áttekinthető, tömörebb FORTRAN nyelv használata. A FORTRAN viszonylag lassu adatmozgatási és konverziós művelet végrehajtásait egyszerű statisztikai feldolgozások esetén sokszorosukra gyorsíthatjuk /ld. pl. [2]/. Megjegyezhető még az is, hogy a SIS77 létrehozásánál használt HWB számítógép FORTRAN nyelve sok tekintetben megközelíti a COBOL vagy a PL1 előnyös tulajdonságait.

4. Ellenőrzés, kódolás

A legfejlettebb adatfeldolgozási rendszerek alkalmazása mellett is sok gondot okoz a bonyolultabb logikai adatkapcsolatok egységes formában történő kezelése /ld. pl. [1]/. Ezért a SIS77-ben nagy figyelmet fordítottunk ennek a feladatnak a megoldására. A bonyolult, sokváltozós adatkapcsolatokat illetve kódolási eljárásokat egy hierarchikus gráfstrukturával modellezzük és realizáljuk a SIS77-ben /ld. [13], [14]/. Ilyen módon sikerült egy egységes adatellenőrzési és kódolási rendszert kialakítani.

5. Statisztikai adatok előállítása

Statisztikai adatokat mindig valamilyen minta egyedeiből képezünk. A statisztikusok által igényelt információk jelentős része csak apró részletekben különbözik egymástól /pl. összevonások, változók felcserélése, értékek kumulálása/. Ezért alkalmaztuk a SIS77-ben az ugynevezett "táblafile"-t, ami egy nagyméretű mátrix /ez gyakoriságokat és kódösszegeket tartalmaz/. Ebből a mátrixból a valóban felhasznált statisztikák /táblázatok, matematikai statisztikai analízisek/ sokasága nyerhető. A táblafile-oknak a kinyomtatott táblázatokra történő leképezésénél egy NYIRI GÉZA által javasolt gyors módszert alkalmaztunk, lehetővé téve a táblázatok néhány másodpercen belül történő létrehozását /függetlenül a vizsgált minta nagyságától/. Ilyen módon lehetővé vált nagy minták statisztikáinak terminálos lekérdezése is. /A leképezési eljárással kapcsolatban ld. a [15] előadást./ A lekérdező rendszerről a jelen kollokviumon GARÁDI JÁNOS számol be.

6. Egy optimalizálási lehetőség

A feldolgozandó adatrendszer nagyságára való tekintettel /mintegy 20-30 millió karakternyi alapadatból kell évenként több mint száz nagyméretű táblázatot készíteni/ egy programoptimalizálási eljárást használtunk /ld. [10], [11]/, amely leginkább az ugynevezett preprocessorok alkal-

mazására emlékeztet.

7. Adatbeviteli eljárások, hibavizsgálatok

Egy számítógépes rendszer felépítésénél mindig nagy gondot kell fordítani a rendszer és a felhasználó közti kapcsolatok jó megszervezésére. A SIS77-ben egy kényelmes és biztonságosan kezelhető adatbeviteli eljárást alkalmaztunk /ld. [15]/, és nagy gondot fordítottunk a felhasználói leírások szintaktikus és szemantikus ellenőrzésére. Ezen kívül a felhasználó a rendszertől olyan listákat is kaphat, amelyek részletes visszajelzést adnak az általa leírt feladatokról.

8. Modulszerkezet

Befejezésként röviden foglalkozunk a SIS77 szerkezeti felépítésével. A fő szempont egy modulszerkezet kialakítása volt. Többek között ez teszi lehetővé a rendszer nagyfokú variálhatóságát, az egyes modulok önálló alkalmazását és más rendszerekhez való kapcsolását, a részrendszerek egymástól függetlenül történő kialakítását. A felhasznált modultípusok különböző szempontok alapján osztályozhatók. Van például olyan rendszerelem, mely önmagában, ismételten felhasználható /ilyenek például a SIS77 ellenőrzési, kódolási, szelektálási eljárásai/, vannak olyan modulok, amelyeket egy láncba kapcsolva kell alkal-

mazni, vagy olyanok, amelyek visszafordíthatatlan folyamatokat realizálnak /pl. a SIS77 táblafile-készítő modulja/.

I r o d a l o m

- [1] Aranyi A.: A statisztikai adatfeldolgozás folyamata. Számítástechnika, IX. évf. 9. sz., 1978.
- [2] Gál A., Ruda M.: Egy lehetőség Honeywell FORTRAN programok konverziós műveleteinek gyorsítására. SZÁMKI Tanulmányok, 1978/II.
- [3] Ratkó I.: Egy számítástechnikai eszköz bonyolult logikai kifejezések leírására orvosstatisztikai alkalmazásokban. 8. Neumann Kollokvium, Szeged, 1977.
- [4] Ratkó I.: Bonyolult logikai kifejezések kiértékelésének számítástechnikai és optimalizálási problémái. MTA SZTAKI Közlemények, 20/1978.
- [5] Raktó I.: On optimization problems of logical expressions in programming languages. Matematikai Logika a Programozáselméletben Kollokvium, Salgótarján, 1978.
- [6] Ratkó I.: Döntések sztochasztikus optimalizációja adatfeldolgozásnál, VIII. Magyar Operációkutatási Konferencia. Szeged, 1978.

- [7] Ruda M.: Egy általános információs rendszer kórházi morbiditási adatok feldolgozására. 8. Neumann Kollokvium, Szeged, 1977.
- [8] Ruda M.: Statistical information system with health service application. 4. Winterschool of Visegrád on the Theory of Operating Systems, Szentendre, 1978.
- [9] Ruda M.: Optimalizálási kérdések a statisztikai adatfeldolgozásban. VIII. Magyar Operációkutatási Konferencia, Szeged, 1978.
- [10] Ruda M.: Egy széles körben alkalmazható programoptimalizálási módszer. MTA SZTAKI Közlemények, 20/1978.
- [11] Rda M.: Módszer a programkészítés egyszerűsítésére. Számítástechnika, IX. évf., 7-8. sz., 1978.
- [12] Ruda M.: A SIS77 statisztikai információs rendszer kialakításának szempontjai, alkalmazásának és továbbfejlesztésének lehetőségei. MTA SZTAKI Tanulmányok /megjelenőben/.
- [13] Ruda M.: A SIS77 statisztikai információs rendszer. /kézirat/.
- [14] Ruda M.: Egy számítástechnikai módszer függvénytáb-

lázatok tömör tárolására, egy adatfeldolgozási alkalmazással /kézirat/.

- [15] Soltész J.: Egy általánosan használható kódolási eljárás és alkalmazása a hospitalizált morbiditási vizsgálatokban. 9. Neumann Kollokvium, Szeged, 1978.
- [16] Várhegyi I.: Paraméterezhető alkalmazási programcsomag. Számítástechnika, IX. évf., 10. sz., 1978.