

PATKÓS ANDRÁS

A fiúk az adatbányában vannak – avagy adatbányászattal felfedezhetők-e alapvető törvények?

„Az adatbányászat nagy adatbázisokban számítógépes segédlettel végzett vadászat érdekes mintázatokra és kapcsolatokra.” E tankönyvi definíció kritikus vizsgálatával igyekszem megfogalmazni szubjektív válaszomat arra a kérdésre, hogyan lehet (lehet-e egyáltalán) alapvető természeti/társadalmi törvényeket kihalászni a ránk zúduló adatáradatból?

AZ ADATVEZÉRELT TÁRSADALOMTUDOMÁNYI ELEMZÉS LEHETŐSÉGE

A gazdasági adatok „kockás papíros” elemzéséből levont, átfogó érvényességűnek szánt elméletek – amelyeknek kiváló példája a 19. század közepén a British Museum olvasótermében naponta megjelenő látogató gazdasági-fejlődés-elmélete volt – a gazdasághoz kapcsolódó adatok sokféleségének és terjedelmének növekedésével többszörös keresztellenőrzésnek váltak alávethetővé. A kevés adaton előbb elemi, majd egyre bonyolultabb matematikával felismerni vélt, ok-okozati kapcsolatokba szervezett gazdasági törvények univerzális érvényességét ezek a tesztek inkább megkérdőjelezték, mintsem igazolták. A rájuk épülő ideológiák, majd államhatalmi politikák vagy megbuktak már, vagy egyre-másra az előre nem jelzett, szinte folytonosan jelentkező válságok leküzdésére korlátozott improvizációkká redukálódtak.¹

¹ „How did economists get it so wrong?” – idézte 2010-ben Dirk Helbing, a zürichi ETH ökonofizikus professzora Paul Krugman elítélő véleményét a valóságtól elszakadt társadalmi-gazdasági kutatóknak a legutóbbi pénzügyi válság előrejelzésében vallott kudarca kapcsán (*Helbing–Balietti* [2010]). A 19. század közepe óta változatlan cél megvalósításához, a kapitalizmus válságainak előrejelzéséhez sem az osztályharc „pofonegyszerű” igazságú logikájára, sem a tökéletes piacon működő racionális egyének döntésmódjának szofisztikált matematikájára épített elmélet nem vitt közelebb (*Helbing–Balietti* [2011]).

Az adatmennyiség és -minőség sokféleségének bővülésével az átfogó elmélet megalkothatóságáról egyre szkeptikusabban vélekedő kutatók fordítani igyekeznek a szereposztáson: az adatvezérelt elemzés ideálját próbálják az elmélettel vezérelt (elméletet tesztelő) mérések elemzésének a helyére állítani.²

A társadalomtudományi modellezésben az elfogadott (mérhető) mennyiségek közötti kapcsolatokat – változási sebességeik adott pillanatbeli értékét maguknak a mennyiségeknek a korábban felvett értékeire épülő – egyenletek formájában fejezik ki. Gyakran nyilvánvaló két mennyiség adatsorának korreláltsága, máskor pedig váratlanul jelentkezik, és elengedhetetlen követelménnyé válik a mérési adatok egyszerű sémában történő rekonstruálásához.

Az úgynevezett főkomponens-elemzés gyakori empirikus eljárás azon mennyiségek szükséges és elegendő csoportjának megtalálásához, amelyekre ráépítik a mérésekből származtatott kimeneti adatok értelmezési modelljét. Ez az eljárás azokra a változókra fókuszál, amelyek változása első közelítésben egymástól független, így aztán a főkomponens-változók alakulásának egymásra gyakorolt hatása kis intenzitású módosításként (perturbációként) értelmezhető.

Az elméletvezérelt adatelemzés során az elméleti modellt definiáló mennyiségeket igyekeznek úgy megválasztani, hogy összességük a verifikáló mérés során jól megfeleljen a főkomponensektől várt gyenge kölcsönös korreláltságnak. Ez azonban nem feltétlenül sikerül. Ha igen, akkor az egyes főkomponens-mennyiségekkel erősen korreláltan változó további mennyiségek alakulását – oksági kapcsolatot feltételezve – többé-kevésbé expliciten definiált függvénykapcsolattal modellezik, amelyben a további főkomponensek rögzített értéke mellett vizsgálhatók a kiszemelt mennyiség változása által előidézett hatások.

Adatvezérelt megközelítésben előzetes elméleti konstrukció nélkül, kizárólag a megmért korrelációk alapján választják az elméleti modellben szerepeltetendő mennyiségek halmazát. A kiválasztás fő kritériuma valamely mennyiség minél pontosabb mérhetősége, nem pedig annak feltételezett őso-k-jellege. Érdekes példája az ok-okozat szerepcserének – azaz a következmény méréséből az okra történő visszakövetkeztetésnek – az influenza-járvány terjedésének irányára és sebességére vonatkozó becslés kinyerése

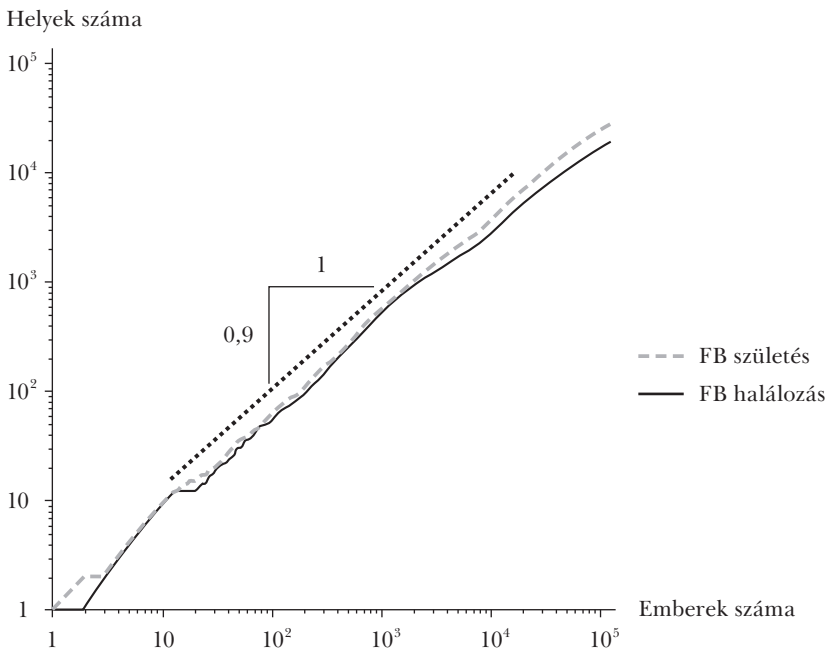
² A *Customer Relationship Management* (fogyasztóikapcsolat-kezelés) néven emlegetett, a személyi szabadságot álcázatlanul sértő kémtevékenység nem lehet tisztességes, tudományos megközelítésű vizsgálat tárgya, hiába hirdetik az „egykor obskúrus céllal fejlesztett akadémiai adatbányászati technikák élenjáró és átfogó alkalmazásaként”.

oly módon, hogy azt követjük, miként mozdulnak el a térképen a tünetekre és a gyógyszerekre való internetes keresések.

Egy másik közelmúltbeli tanulmány az elmúlt kétezer év mintegy 150 ezer híres személyiségének születési és halálozási helyét összekötő mobilitással jellemezte az európai és észak-amerikai történelem kulturális vonzáscentrumainak létrejöttét és változásait (*Schich és szerzőtársai* [2014]). A filmszerűen lepergethető hálózati történetben megfigyelhető karakterisztikus időbeli változások időpontja és jellege összekapcsolható az adat-szerűen nem megragadható politikai-gazdasági és szellemi változásokkal. A gondos elemzés felállított egy kvantitatív összefüggést a centrumok és a közöttük vándorló személyiségek száma között. A szerzők arra jutottak, hogy a centrumok száma a lineárisnál lassabban, 0,9-es hatványkitevővel nő a hírnevesként számon tartott személyiségek számával (*1. ábra*).

1. ÁBRA

Összefüggés a kulturális centrumok kialakulása és a közöttük vándorló személyiségek száma között



A kulturálisan meghatározó vonzáscentrumok számának (függőleges tengely) növekedése a meghatározó jelentőségű személyek számának (vízszintes tengely) növekedése függvényében, amit a két mennyiség történelmi változásának időfüggését kiküszöbölve nyertek a szerzők [a születési és a halálozási adatok a Freebase.com (FB) adatbázisból származnak].

Forrás: *Schich és szerzőtársai* [2014].

Mit gondoljunk erről az összefüggésről? Hihető-e, hogy új objektív társadalmi törvényt fedeztek fel, amely használható az emberi társadalom fejlődési expanziójának jellemzésére? A véleménynyilvánítást a természet-tudományi adatbányászat két példájának ismertetése utánra halasztom.

MINTÁZAT A NAPRENDSZER BOLYGÓINAK ADATSORÁBAN – KÉPZELŐDÉS VAGY IZGALMAS TUDOMÁNYOS KIHÍVÁS?

A 20. század végéig felfedezett természettörvények felismerésében nem annyira a bőséges adatbázisok részletes elemzéseinek volt szerepe, hanem sokkal inkább a hiányos pontosságú adatok szembe ötlő világosságú tendenciáinak. A kvantumfizika egész történetének kezdetét jelentő, 1905-ös cikkében Einstein egyetlen foton elnyelésének tulajdonította, hogy egy elektron kilép a fém belsejéből. Ezzel kvantumfizikai magyarázatot adott a fotoelektromos hatás korabeli mérési eredményeire. A nagy teljesítményű lézerekkel fél évszázaddal később végzett mérésekben több foton elnyelésével végbemenő elektronkilépés is történik, aminek valószínűsége elhanyagolható volt a 19. század végi kísérletekben. A mérési pontosság korabeli technikai korlátai „szerencsét hoztak” a fotoelektromos hatást kísérletileg igen részletesen vizsgáló Philippe Lenardnak (és Einsteinnek). A nagy intenzitású lézertény nemlineáris (az intenzitás magasabb hatványával arányos) hatására kilépő elektronok tulajdonságaiból jóval nehezebben juthattak volna el a fénykvantum hipotéziséig.

Az adatok elméletmentes elemzéséből levont, talán legnevezetesebb szabályszerűség a bolygópályák sugarára vonatkozó Titius–Bode-szabály, amelyet Bode közölt 1772-ben, de iratai világosan igazolták, hogy Titius szövegezését vette át. Az azóta végzett tudománytörténeti vizsgálatok szerint Titius C. F. von Wolftól tanulta meg a bolygók keringési sugarára vonatkozó szabályosság létezését, de már egy 1702-ben megjelent angol csillagászati könyvben is szerepelt egy hasonló szabály. Ám Bode használta először a sokáig egyedül az ő nevét viselő „törvényt”, mégpedig igazoló vagy cáfoló további kutatásokat sürgető előrejelzés céljára. A törvény az általa közölt szöveges megfogalmazásban a következőképpen hangzik:

Vegyük a Nap és a Szaturnusz távolságát 100 egységnek, ekkor a Merkúr 4 résznyi távolságra van a Naptól. A Vénusz távolsága $4 + 3 = 7$. A Földé $4 + 6 = 10$. A Marsé $4 + 12 = 16$. Itt egy rés következik ebben a szabályos sorozatban. A Mars helyét egy $4 + 24 = 28$ résznyi távolságú hely követi, de ott egyelőre

nem látszik bolygó. Hihető-e, hogy a világegyetem Teremtője ezt a helyet üresen hagyta? Nyilván nem. Innét a Jupiter távolságához érkezünk a $4 + 48 = 52$, végül a Szaturnuszhoz a $4 + 96 = 100$ képlettel.

A Nap–Föld-távolságot csillagászati egységnek hívják, jele AU . Ebben a távolságegység rendszerben a fenti képzési szabállyal a bolygók Naptól mért átlagos távolsága a

$$d = 0,4 + 0,3 \times 2^m AU, \quad m = -\infty, 0, 1, 2, 3, \dots$$

alakban írható. 1781-ben az $m = 6$ helyre illeszkedő távolságú Uránusz felfedezésével a szabály hírneve sokat nyert. A Bode sürgetésére végzett csillagászati megfigyelések 1801-ben úgy tűnt, hogy sikert hoztak a Teremtő következetességének ügyében is: a Ceres kisbolygót a kimaradni látszó $m = 3$ értékkel képzett távolságon találták meg. Mai csillagászati ismereteink szerint az úgynevezett kisbolygó-öv helyezkedik itt el, amelyet egy korábbi bolygó törmelékei alkothatnak.

Azok, akik véletlen számsorozatként tekintenek a Nap–bolygó távolságok sorozatára, erős érvet kaptak a Neptun 1846-os felfedezésével, majd 1930-ban a Plutóval, amelyek lényegesen kilógnak a sorból. A Plutó kisbolygóvá történt visszaminősítése nem igazán segít, mert a bolygómaradványokból álló Kuiper-öv törmelékhalmozának távolsága sem illeszkedik a feltételezett sorozatba.

Amikor egy bolygó mozgását Newton törvényei alapján tárgyalják, első közelítésben csak a Nap tömegvonzását veszik figyelembe, amiből a pályák sugarára nem adódik korlátozás. Ha a többi bolygó hatását is figyelembe vesszük, akkor az egyes pályák stabilitásának vizsgálata mint feladat az égi mechanika tudományterületén egy ma sem lezárt fejezethez juttat bennünket. Azok, akik az egyszerű véletlennél többet kívánnak látni a Titius–Bode-szabályban, a Naprendszer bolygóinak keletkezéstörténetéig nyúlnak vissza. A *planetézimálnak* hívott bolygókezdemények a környezetükben lévő por gravitációs befogásával növekedtek, és a megfigyelt pályák az anyagcsomók egymást kölcsönösen stabilizáló gravitációs hatásának eredményei lehetnek. Ezek a kutatók tehát a bolygókeletkezés dinamikájának számítógépes szimulációját tanulmányozzák, és a stabilizálódó rendszer adataiban keresik a Titius–Bode-szabály megnyilvánulását.

Csak egy Naprendszerünk van, de a szimuláció céljára előállíthatjuk annak tetszőleges sok példányát. Modell-Naprendszereink története azonban magán hordozza a modellünk megalkotásához felhasznált ismereteink korlátait, így a végeredményt nem tekinthetjük a felvetett kérdés

kétségtelen lezárásának. Mindenesetre a fiatal csillag körüli porkorong bolygókká csomósodásának számítógépes történetét tartalmazó nagy adatbázis elemzése az egyik út lehet annak eldöntésére, hogy a Titius–Bode-szabály természettörvény-e, vagy csupán a szabályosságot a természetbe belevetítő szubjektum káprázata.

A Naprendszeren túl, más csillagok körül az elmúlt negyedszázadban egyre növekvő számban mutatták ki bolygórendszerek létezését. A távoli csillagok körüli bolygók felfedezése felkínálja az ellenőrzés második lehetőségét: az e bolygórendszerek adatbázisán végzett mintázatkeresést. Az úgynevezett exobolygók felfedezése javarészt az úgynevezett tranzithatásuk révén történik, amellyel a csillag előtti áthaladásukkor csökkentik annak megfigyelhető fényességét. A *Kepler*-űrmisszió több mint 3500 exobolygójelöltet fedezett fel az úgynevezett takarási hatás révén. Nagyjából ötödüknél az adatok több bolygó jelenlétével értelmezhetők. A megfigyelhetőségi küszöb miatt a felfedezett exobolygók többségének mérete a Neptun tartományába esik, ami egyelőre nyilván korlátot szab annak, hogy megtaláljuk a csillag körüli teljes bolygórendszert.

Az exobolygók esetében nem a távolságot, hanem a keringési periódus idejét lehet közvetlenül meghatározni. Kepler törvénye alapján [amely szerint a bolygókra a Napjuktól mért távolság (R) és a keringési idő (T) kapcsolatában az $R^3:T^2$ arány állandó, lásd *1. táblázat*] a Titius–Bode-szabály arra is átfogalmazható. A feltételezett szabályban nem ragaszkodnak a Naprendszerben megfigyelhető kapcsolathoz, hanem általánosabb matematikai alakot keresnek:

$$\log T_n = \log T_0 + n \log a \quad n = 0, 1, 2, \dots$$

2013-ban két kutató 68 kiválasztott, több exobolygót tartalmazó rendszerre végzett el a fenti reláció alapján sorozatba rendezést (*Bovaird–Lineweaver* [2013]). Eredményük szerint a rendezés lehetségesnek mutatkozott a sokbolygós minta harmadában, bár – nem meglepő módon – a sorozatokban kihagyásokat találtak. Ezt két másik kutató 2014-ben keresési „utasítás-ként” értelmezte, és a Kepler űrszonda felvételein az üresnek tűnő összesen 141 helyen bolygókat keresett (*Huang–Bakos* [2014]). A képfeldolgozó eljárással automatizált keresés eredménye csupán öt esetben vezetett a hiányzó bolygók megtalálásához. A szerzők megbecsülték a keresési kudarc esélyét, amelyet az exobolygók kis mérete és a csillag előtti áthaladás pályájának kedvezőtlen sajátossága okozhat. Ebből végül arra a következtetésre jutottak, hogy túl nagy az elmaradás a bolygók keringési periódusában tapasztalt statisztikus ingadozás alapján minimálisan elvárható pozitív

1. TÁBLÁZAT

A távolságkőb (R^3) és a keringési idő négyzete (T^2) aránya

Bolygó	R^3	T^2	T^2/R^3
Merkúr	0,05796	0,05803	1,00121
Vénusz	0,37793	0,037946	1,00405
Föld	1,0	1,0	1,0
Mars	3,5396	3,61	1,01989
Jupiter	140,85	144,0	1,02236
Szaturnusz	867,98	870,25	1,00262
Uránusz	7055,8	7056,0	1,00003
Neptun	27162,3	27154,7	0,99972
Pluto	61723,5	61504,0	0,99644

Megjegyzés: A naprendszer bolygóinak távolságkőbe a Naptól a Nap–Föld távolság kőbének arányában, illetve a bolygók keringési idejének négyzete a Föld keringési idejének négyzete arányában. A hányadosok eltérése a Nap és a bolygó közötti erőhatásra korlátozódó számítás eredményeként adódó egy-egyedtől a többi naprendszerbeli objektumnak a Kepler-törvény pontosságát korlátozó hatását mutatja.

észlelésektől. 99,9 százalékos biztonsággal állítják, hogy az exobolygó mintában nem érvényesül az általános Titius–Bode-szabály.

Példánk arra mutat rá, hogy a kis mintában látni vélt, a természeti törvényhez közelállónak tűnő szabályosság ellenőrzésére (megerősítésére vagy cáfolatára) kiválóan használható az adatbányászati megközelítés.

RÉSZECSEFIZIKAI TÖRVÉNYEK KERESÉSE ADATBÁNYÁSZATI ELJÁRÁSSAL

A modern fizika fél évezredes fejlődése nyomán a fizika számos jelenségkörében vannak *standard modellek*, amelyeknek bizonyára létezik (sőt részben ismert is) alkalmazhatósági határa. Az ismerteken túlmutató összefüggések feltárásakor fontos mozzanat az aktuálisan legpontosabb megfigyelések összevetése az elfogadott elmélettel, amit követhet a váraozástól esetleg eltérő eredmény magyarázata, az értelmező elmélet esetleges kiegészítése. Ehhez a lépéshez nem a jó szerencsében bízva végeznek kísérleteket, hanem előre tervezetten keresik a mérési adatokban a standardnak tekinthető modellből származó és a kiterjesztésből jövő jóslatok különbségének megjelenését vagy hiányát. Ebben az értelemben igaz a kísérleti fizikusok önképe, miszerint a kísérleti fizikus tökéletes elméleti szakember plusz még valami több!

A fizikai alaptörvények feltárt rendszerét tehát nem kívánja egyetlen fizikus sem újra rekonstruálni nagy adatbázisok sajátos mintázatainak hátteréből visszafelé haladva. Annál inkább adatbányászati jellegűek az olyan próbálkozások, amelyek ismert törvényeken túlmutató jelenségek kísérletes megragadására irányulnak, ami előfeltétele az elmélet magasabb szinten történő újrafogalmazásának.

A CERN LHC ütközőjében megvalósított, nagyenergiájú proton–proton ütközések során évente 1000 terabájt információ keletkezik. Az alapinformációt a detektorokban létrejövő részecskenyomok jelentik, amihez a nyers adatok szintjén kapcsolódnak a detektorokra közvetlenül ráépített, adatfeldolgozó eszközökkel megszerezhető tulajdonságok (impulzus, töltés, tömeg). Ez az adatmennyiség a gyorsítóban másodpercenként bekövetkező, sok százmillió ütközésből épül fel. A feladat az, hogy az ütközési eseményeket osztályokba sorolják, amely osztályok mindegyikét a keletkező több ezer részecske tulajdonságainak – nagyszámú független szempont szerinti – jellemzésével definiálják.

A gyorsító beindítása után elsőként az ismert tulajdonságú erős kölcsönhatási eseményekre vonatkozó várakozásokat a mérésekkel összevetve megismerik a berendezés észlelési érzékenységét, megbecsülik a mért eseményszám és az ütközésben bekövetkezett összes azonos jellegű esemény száma közötti arányt. Ezt követi a ritkábban előforduló, de még ismert kölcsönhatásból származó eseményosztályok rekonstruálása a detektorrendszer individuális érzékenységi tulajdonságainak figyelembevételével. Több hónapos munkával elkészítik az ismert kölcsönhatásokat az új gyorsítóban megjelenítő események statisztikai jellemzését. Erre ültetik rá azután azokat az osztályozási eljárásokat, amelyekkel a legmegbízhatóbban leválasztják az ismert kölcsönhatásokkal nem értelmezhető eseményeket.

Az adatelemzési folyamatot több részre osztják, s minden lépést számítógépes eljárással végeznek (*Brun–Rademakers*[1997]). Az első értékelési rétegben az ütközésben keletkező részecskehalmoz néhány fontos adatának megmérését követően eldöntik, hogy egyáltalán tárolják-e az adott esemény adatait. A kísérleti részecskefizikusok munkájának egyik legfontosabb része olyan kiválasztási kritériumcsoportok (úgynevezett triggerek) megalkotása és a kísérleti csoport egészével történő elfogadtatása, amely alkalmas arra, hogy a nem standard eseményt besorolják valamely, bővebb elmélet jóslatának igazolására definiált osztályba. Az elfogadott triggerek elvárt tulajdonságainak való megfelelés az elektronikus tárolás kritériuma. Ennek a csökkentett intenzitású adatáramnak a „terméke” a fentebb jelzett, összesen 134 petabájt méretű információhalmoz, amelyet a CERN globális adatfeldolgozó

hálózatának központi egységében, az úgynevezett 0-szinten helyeznek el. Az elemzés célja, hogy megtalálják a Higgs-részecskének elnevezett adatosztályba sorolható eseményeket, és jellemezzék valószínűségüket. Ennek az adathalmaznak kópiáit töltötték fel az 1-szintű regionális központokba, amelyek száma 11. Ezekből az adatok részhalmazait több száz 2-szintű számítógépbe osztják szét. A kutatói csoportok ezeknek az alacsonyabb szintű adatbázisoknak a teljes tárolt anyagán végezhetik el osztályba sorolási elemzésüket, melyet valamely speciális szignatúra meglétére vagy hiányára alapoznak. Ha több különböző tesztnak is megfelel valamelyik esemény, akkor kezd körvonalazódni az új felfedezés megerősítésének esélye. Ha akár csak egy elvárt szignatúra hiánya meghaladja a megengedett ingadozási szintet, akkor ez cáfolja az adott jóslatra vezető modellváltozatot. A természetben létező elemi kölcsönhatások modellezésében nagy gyakorlattal rendelkező elméleti fizikusok is örülnek, ha a modellek „állatkertje” ritkul, de nyilván egy nem üres osztály megjelenésének örülnének a leginkább.

Ezzel az eljárással találták meg a Higgs-szerű részecskét 2012-ben. Az eljárás még korántsem fejeződött be (bár a Nobel-díjat már kiosztották érte). A Higgs-részecske kölcsönhatásainak kevésbé gyakori csatornáit még akár egy teljes évtizedig tovább elemzik, hogy végleg dönthesse-nek arról, hogy eleget tesz-e elegendően nagy pontossággal a Standard Modell várakozásainak.

Ezt a globális számítógépes értékelőhálózatot a CERN nemcsak a tudományban (például nagy asztrofizikai galaxisok térképezéseinek az elemzésében), de a piaci alkalmazásokban is igyekszik értékesíteni aktuáriusok és más biztosítási elemzők számára. Remélik, hogy a World Wide Web sikeréhez hasonlóan meghatározó szerepre lesznek képesek az adatbányászati technológiák területén is.

BEFEJEZÉS

A bemutatott fizikai példák a természettudományok azon közös sajátosságát példázzák, hogy a természettörvények egymásra épülnek, az új törvények megtalálása nem érvényteleníti a korábbiakat, hanem az alkalmazási tartomány kibővítésével tehető nyilvánvalóvá a létezésük. A felfedezésükre mozgósított adatelemzés ezért az ismert elméleti törvényszerűségeknek az adatelemző eljárásba történő, teljes körű beépítésével zajlik. Mégsem beszélhetünk teljességgel elméletorientált elemzésről, mivel a vetekedő elméleti kiterjesztések az adatoknak adnak elsőbbséget a választásban.

Léteznek továbbá megfigyelésekből sugallt empirikus szabályszerűségek is, amelyek valós létezésének eldöntéséhez szintén jól használhatók az adatbázis-elemzések. Nem ismerek olyan példát, amelyben alapvető természettörvényt ismertek volna fel elméleti konstrukciótól teljesen mentesített eljárással.

A teljesen adatvezérelt társadalomtudományi elemzésekből felsejlő kapcsolatok értékének megítélését a természettudományi példákat követő szigorúságú verifikálási eljárás példáinak tükrében az Olvasóra bízom. Személyes véleményem szerint többségük nem igazán releváns kapcsolatokat tár fel. Akkor sem pótolhatják az elmélet fájó hiányát, ha minden tesztet túlélve, igaznak bizonyulnak. Azt természetesen túlzott elvárásnak tartanám, ha a következő évi gazdasági-pénzügyi előrejelzést (kölségvetési tervet) csak abban az esetben szavazná meg az országgyűlés, ha készítői előzetesen bemutatják, hogy az extrapoláció alapjául szolgáló elméleti modelljük az előző száz év gazdaságtörténetének minden eseményét 1 százalékos hibahatáron belüli reprodukálja.

HIVATKOZÁSOK

- BOVAIRD, T.–LINEWEAVER, C. H. [2013]: Exoplanet predictions based on the generalized Titius-Bode relation. *Monthly Notices of the Royal Astronomical Society*, Vol. 435. No. 2. 1126–1138. o. <http://dx.doi.org/10.1093/mnras/stt1357>.
- BRUN, R.–RADEMAKERS, F.[1997]: ROOT – An object oriented data analysis framework. *Nuclear Instruments and Methods*, Vol. 389. No. 1–2. 81–86. o. [http://dx.doi.org/10.1016/s0168-9002\(97\)00048-x](http://dx.doi.org/10.1016/s0168-9002(97)00048-x).
- HELBING, D.–BALIETTI, S. [2010]: Fundamental and real world challenges in economics, *Science and Culture*, Vol. 76. No. 9–10. 399–417. o. <http://dx.doi.org/10.2139/ssrn.1680262>.
- HELBING, D.–BALIETTI, S. [2011]: From Social Data Mining to Forecasting Socio-Economic Crisis, *European Physics Journal Special Topics*, Vol. 195. No. 1. 3–68. o. <http://dx.doi.org/10.1140/epjst/e2011-01401-8>.
- HUANG, C. X.–BAKOS G. A. [2014]: Testing the Titius-Bode law predictions for Kepler multiplanet systems, *Monthly Notices of the Royal Astronomical Society*, Vol. 442. No. 1 674–681. o. <http://dx.doi.org/10.1093/mnras/stu906>.
- SCHICH, M.–SONG, C.–AHN, Y.-Y.–MIRSKY, A.–MARTINO, M.–BARABÁSI, A.-L.–HELBING, D. [2014]: A network framework of cultural history, *Science*, Vol. 345. No. 6196. 558–562. o. <http://dx.doi.org/10.1126/science.1240064>.