

Kitekint? - B?vészkedés a számokkal

2009 november 10. Flag

Szöveg méret

90
Mentés

- 0

Értékelés kiválasztása Még nincs értékelve

Mérték

A tények bányászatának korában élünk, amikor múltból és jelenb?l nyerünk ki óriási adathalmazt, és ebb?l vázoljuk fel a jöv?t. A számok forradalma megváltoztatja körülöttünk a világot.

Egy amerikai matematikus arra adta a fejét, hogy kikutassa: mit?l lesz egy bor jó, és mit?l marad csak közepes min?ség?. Több száz év elérhet? francia id?járási és bormin?ségi adatait hasonlította össze számítógép

segítségével, és felállított egy matematikai képletet, aminek lényege, hogy a forró nyár és a száraz ?sz hoz csúcsmin?ség? évjáratot.

Megállapítása óriási vihart kavart amerikai borászberkekben, lényegében ?rültnek min?sítették. Ma már igazat adnak neki, s?t el?rejelzései segítenek a borbefektet?knek, mert már a szüret napjaiban tudják, hogy az évjárat nagy bor lesz-e, míg a borkóstoló szakembereknek erre 3-4 hónapot várniuk kell. A titok a nagy id?járási adathalmazok matematikai mozgatása, amire a mai számítógépek hatalmas adattárolási és kezelési kapacitása ad módot. Hasonló meglepetést keltett, amikor Garri Kaszparov kikapott az IBM sakkprogramjától: 700 ezer nagymester betáplált adatbázisával már nem bírt el az emberi elme.

A Google ötezer terabite, tehát öt petabite adatot tartalmaz: a görög tera szörnyet jelent, ez valóban szörnyen sok adat. Ma a fejlett világban hatalmas pénzt költenek olyan adatbázisok felépítésére, amelyek az élet minden területér?l minden elérhet? adatot összehoznak, majd feltárják a közöttük lév? rejtett összefüggéseket. Így találtak rá arra, hogy azok okozzák és szenvedik el a legtöbb balesetet, akik tartozást halmoznak fel hitelkártyájukon: az amerikai biztosítók már nagyobb díjat is kérnek t?lük. Fogalmunk sincs, hogy kik és milyen adatainkat töltik le, szerzik meg, majd meglehet?s pontossággal rajzolják fel szokásainkat, jöv?beli várható mozgásunkat és viselkedésünket.

Az angol Francis Galton - Charles Darwin unokaöccse - ismerte fel, hogy az átlagnál nagyobb magokból átlagosnál nagyobb növények lesznek, majd azt, hogy az átlagnál magasabb apáknak átlagnál magasabb fiai lesznek, de a fiúk nem érik el az apa magasságát. Ha az apa 1 inch-csel magasabb az átlagnál, akkor a fiú 2/3 inch-csel lesz magasabb az átlagnál. Lényegében erre épül a modern adathalmazok regressziós eljárása is, amikor múltbeli trendekb?l vonnak le a jöv?re vonatkozó el?rejelzéseket. A vállalati felvételi beszélgetéseken eddig három tulajdonságot néztek: a felel?s gondolkodást, a kellemes egyéniséget és a nyitottságot. Ma már a lehet? legtöbb adatot beszerzik a pályázóról. Megrajzolják a személyiségét, és összehasonlítják egy hatalmas adatbázissal, amelynek eredményeképpen megtudják, hogy a hozzá hasonló személyiség hogyan állta meg a helyét az adott pozícióban.

A másik módszer a mintavételes eljárás. Ekkor nem múltbeli adatokat bányászunk óriási sokaságból hatalmas mennyiségben, hanem véletlenszer?en gy?jtünk mai adatot, véleményt nagy sokaságból. Majd ezt a mintát összehasonlítjuk a meglév? óriási adatbázissal, és láthatóvá válnak a rejtett összefüggések. Ha kiderül, kire hasonlítunk, milyen csoportba tartozunk, akkor nagy pontossággal el?re jelezhet? véleményünk, viselkedésünk és döntésünk is. A nem szakért? tömeg mindig pontosabb, mint egy sz?k szakért?i csoport. A "tömeg bölcsessége" itt is m?ködik.

A nagy adatbázisok gyors kezelése az egészségügyi rendszerre is óriási hatással lesz. Ma 11 ezer betegségben szenved az emberiség, ki tudja ezt mind fejben tartani? Ezért van, hogy Amerikában minden harmadik orvosi diagnózis hibás, és az összes halállal végz?d? betegség ötödében a rossz diagnózis vezetett a véghez. Ezen segít az "adatbázisok szerinti diagnosztika", amikor az orvos segítséget kér a számítógépes adatbanktól: a 11 ezer betegséget 30 lehetségesre képes sz?kíteni ez a módszer.

A számok forradalma átok és áldás egyszerre. Rajtunk is áll, melyik kerül túlsúlyba: a kockázat vagy a lehet?ség.

Matolcsy György, hetivalasz.hu



Ajánló