Pillangó-effektusok fellépése idősorok hasonlóságelemzésre alapozó értelmezésekor

(Butterfly-effects in similarity analyses of time series)

Pitlik László, Pitlik Marcell, Pitlik Mátyás (MY-X team)

Kivonat: A tanulmány célja egy valós pillangó-effektus bemutatása reprodukálható adatvagyonon a hatásmechanizmus matematikai részleteinek feltárása mellett. A valós, de anonim probléma: objektumok (pl. személyek) idősorosan létező jellemzőinek klasszifikálása hasonlóságelemzés-láncokra alapozva. A pillangó-effektus maga nem más, mint mekkora lehet egyetlen egy rekordnyi többletinformáció hatása egy/több objektum esetén? Az egy rekordnyi információ a valóságban ¼ másodpercnyi időintervallumot jelent több nagyságrendnyi rekordszám esetén. Az egyetlen egy idősor-elemmel több vagy kevesebb állapot hatását a nyersadatok rangsorrá konvertálása keretében fejti ki méretfüggetlen (azaz egy rekordra) vonatkozó objektum-attribútum mátrixokon keresztül – különösen, ha sok azonos rangsorolandó adat van az alapmintákban. A hasonlóságelemzés szerepe csak annyi, hogy ez egy tipikus rangsor-inputokat feldolgozó eljárás, ahol a pillangó-hatás a függvény-szimmetriákon keresztül immár képes kihatni a részeredmények konzisztenciájára is.

Kulcsszavak: konzisztencia, rangsorolás, méretfüggetlenítés, függvény-szimmetria

Abstract: The objective of the paper is to demonstrate a butterfly-effect based on a real but anonymous data asset where each step can be reproduced. The problem is: how can we classify personal time series in frame of similarity analyses? The butterfly-effect is the unlimited difference between the result of the classification based on a given data asset and the result of an other classification based on a data asset having just one single record more as input – in this case having data about ¼ sec where the used length of the time series can be over 100 or 1000. The differences will be derived through the ranked inputs – especially in case of data having the same value. The similarity analysis is a typical ranking-oriented modelling schema where these special effects can be detected at once – without any further manipulations. The similarity analyses produce model chains and so, the symmetry-driven similarity analyses can have butterfly-effects in a consistence-oriented model structure too.

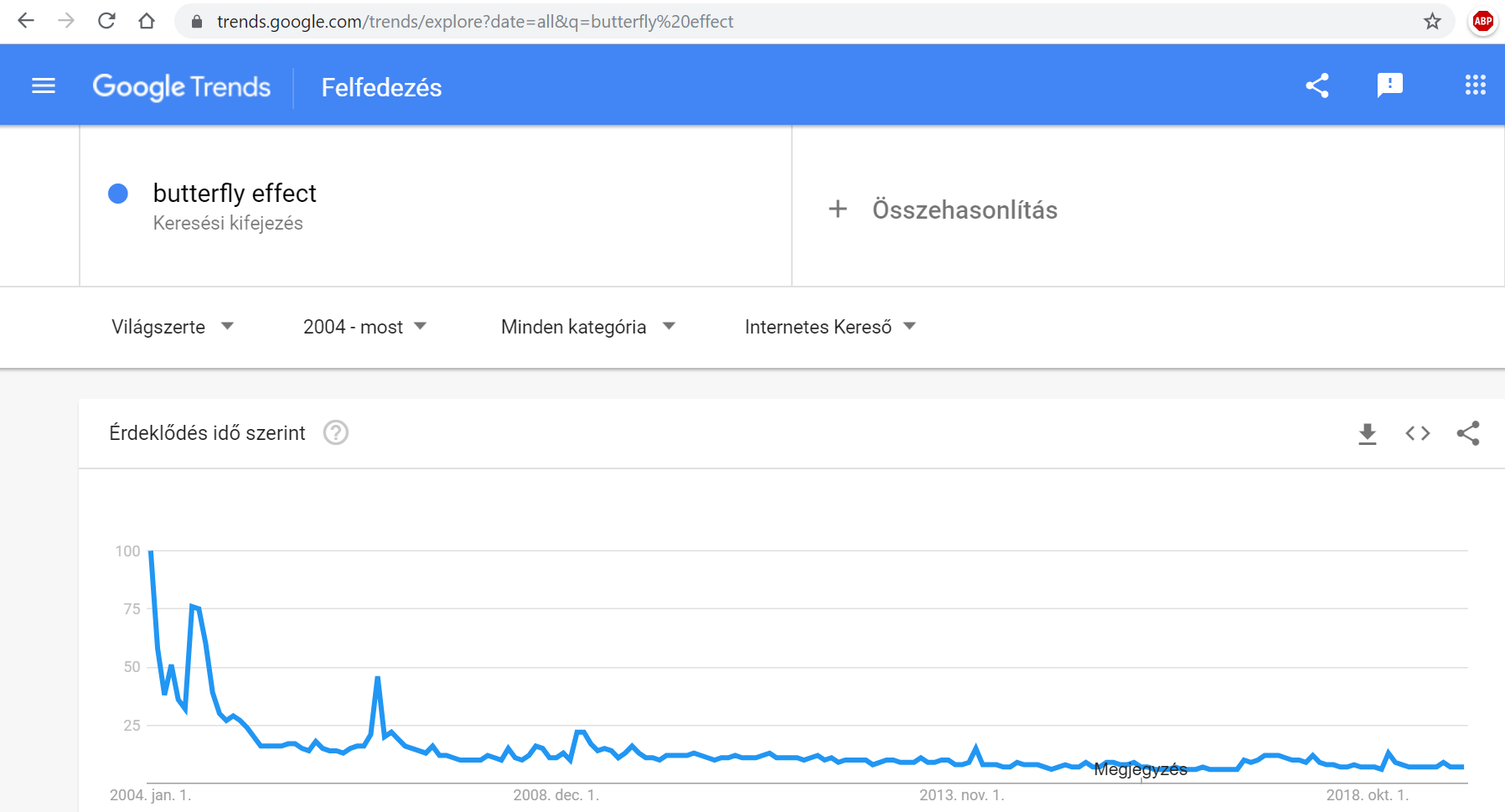
Keywords: consistence, ranking, relativity, symmetry of functions

Bevezetés

A Liebig-féle minimum elv közlekedési rendszerekben való quasi véletlenszerű tetten érése után (vö. <https://miau.my-x.hu/miau/253/traffic-simulations.docx>) a pillangó-effektus valós problémahalmazon való felismerése és reprodukálható dokumentálása a big data elemzések egy újabb quasi „melléktermékeként” értelmezhető. Ezek a rendszerelméleti szinten releváns „melléktermékek” alapjaiban képesek befolyásolni a komplex rendszerek konzisztenciájáról alkotott kutatói értelmezéseket.

„Evolution is recklessly opportunistic: it favors any variation that provides a competitive advantage over other members of an organism's own population or over individuals of different species. For billions of years this process has automatically fueled what we call evolutionary progress. No program controlled or directed this progression. It was the result of spur of the moment decisions of natural selection.” (Robert M. Pirsig: Lila : An Inquiry into Morals, 1991 - <https://terebess.hu/zen/mesterek/Robert_Pirsig-Lila.rtf>) S ez a végtelen sok „moment decision” nem más, mint pl. pillangó-effektusok egymásba kapcsolódó sorozata (<https://hu.wikipedia.org/wiki/Pillang%C3%B3hat%C3%A1s_(elm%C3%A9let)>).

A kérdés már csak az, milyen konkrét számpéldák találhatók valós rendszerek leírásaként erről az elmúlt mintegy 14 évben (a Google Trends szerint) egyre kevésbé keresett/értelmezett jelenségről?



1. ábra: Google Trends adatok a butterfly effect kulcsszó kapcsán 200-2019 között világszerte (forrás: <https://trends.google.com/trends/explore?date=all&q=butterfly%20effect>, 2019.12.17.)

A tanulmány célja valódi idősoros adatok alapján annak demonstrálása, miként hat ki egyetlen egy (a konkrét példában ¼ másodpercet jellemző többlet adatrekord) masszívan a hasonlóságelemzésláncok eredményére, ahol maga a hasonlóságelemzés egy fajta mesterséges intelligencia-alapú intuíció-generálás (vö. <https://miau.my-x.hu/miau/196/My-X%20Team_A5%20fuzet_HU_jav.pdf>).

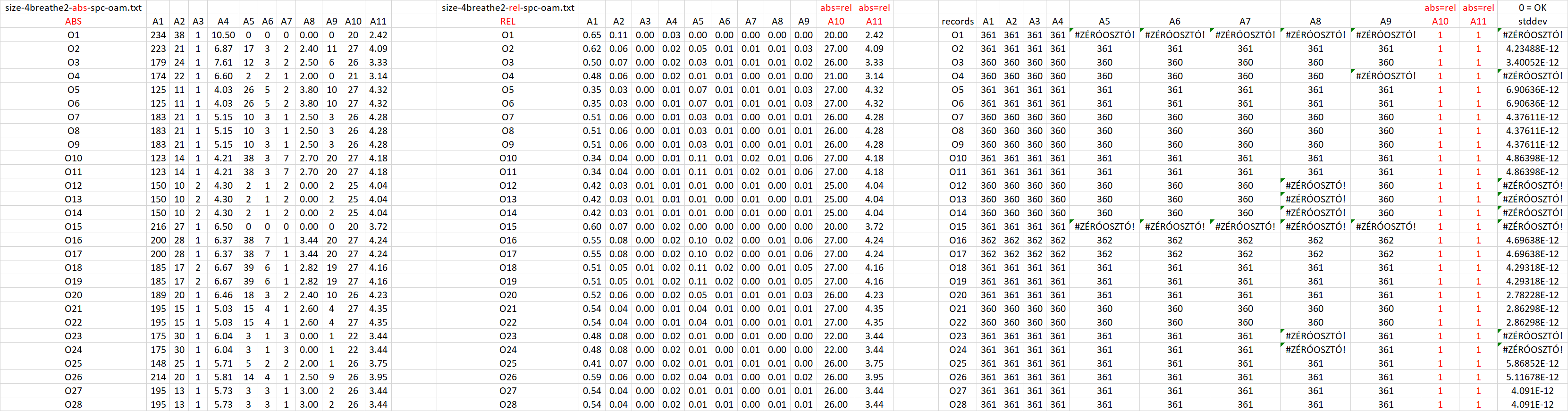
A szakirodalomban valós példákat ezek teljes reprodukálhatóságot biztosító hátterével nem triviális fellelni. Itt és most, ugyan anonimizáltan, de teljeskörű számpélda kerül bemutatásra a jelenség közös adatvagyonon történő értelmezését megalapozandó: <https://miau.my-x.hu/miau/257/butterfly>

Az adatvagyon maga két szintű, a nyers idősoros mérések negyedmásodperces adatai számos párhuzamosan létező jelenségről szólnak. A számpéldába a nyers adatok helyett már csak ezek statisztikai leíró adatai kerültek be, ahol a nyersadatsorok hossza 1-1 rekordnyi különbséget mutat. Jelentős rekordszámnyi eltérés esetén mindenki számára triviális, hogy pl. a szélsőségesnek nevezett tartományokba esések darabszáma arányos az adatsorok hosszával, vagyis minél nagyobb két idősor hosszának eltérése egymástól, annál inkább szükséges minden abszolút statisztikai jellemző relativálása (méretfüggetlenítése). De már egyetlen egy rekord esetén is fennáll annak lehetősége/logikai kényszere, hogy ez a relativálás megtörténjen. S elvileg azt lehetne elvárni, hogy egy rekord hatása (vö. egy fajta érzékenység-vizsgálatként felfogva az abszolút és a relativált adatok hatáskülönbségét) lényegében nem érzékelhető.

Ez a hipotézis azonban téves minden olyan esetben, ahol az alapadatok ismétlődése eltérő objektumok eltérő hosszúságú megfigyelése esetén nincs kizárva, sőt, tetszőleges gyakorisággal fordulhatnak elő ismétlődő (objektumokat azonos leíró adattal ellátó) karakterisztikák.

# Az adatvagyonról

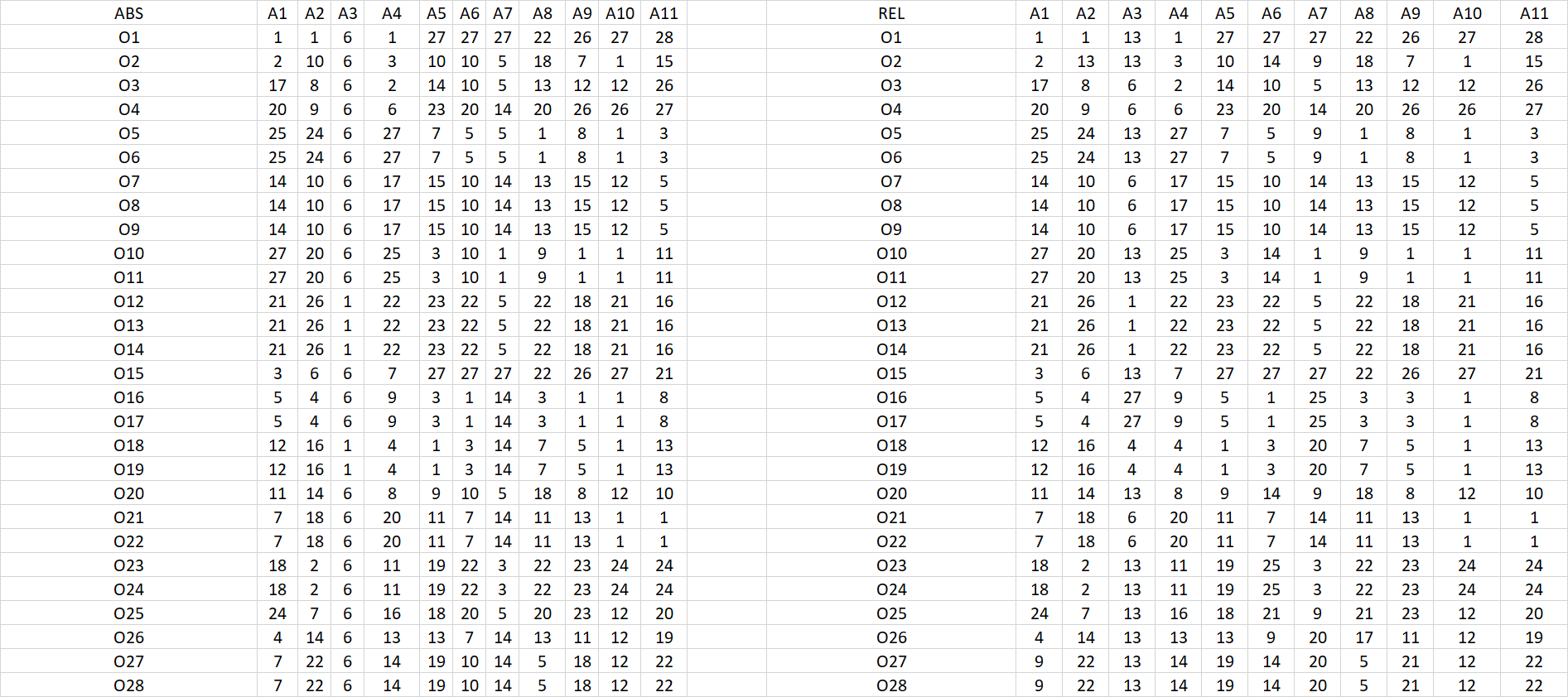
<https://miau.my-x.hu/miau/257/butterfly> - ebben a könyvtárban érhetők el a számítások kapcsán felhasznált, ill. részeredményeket felmutató adatok.



1. ábra: Abszolút és relatívált OAM-ok mögötti rekordszámok (forrás: saját számítások)

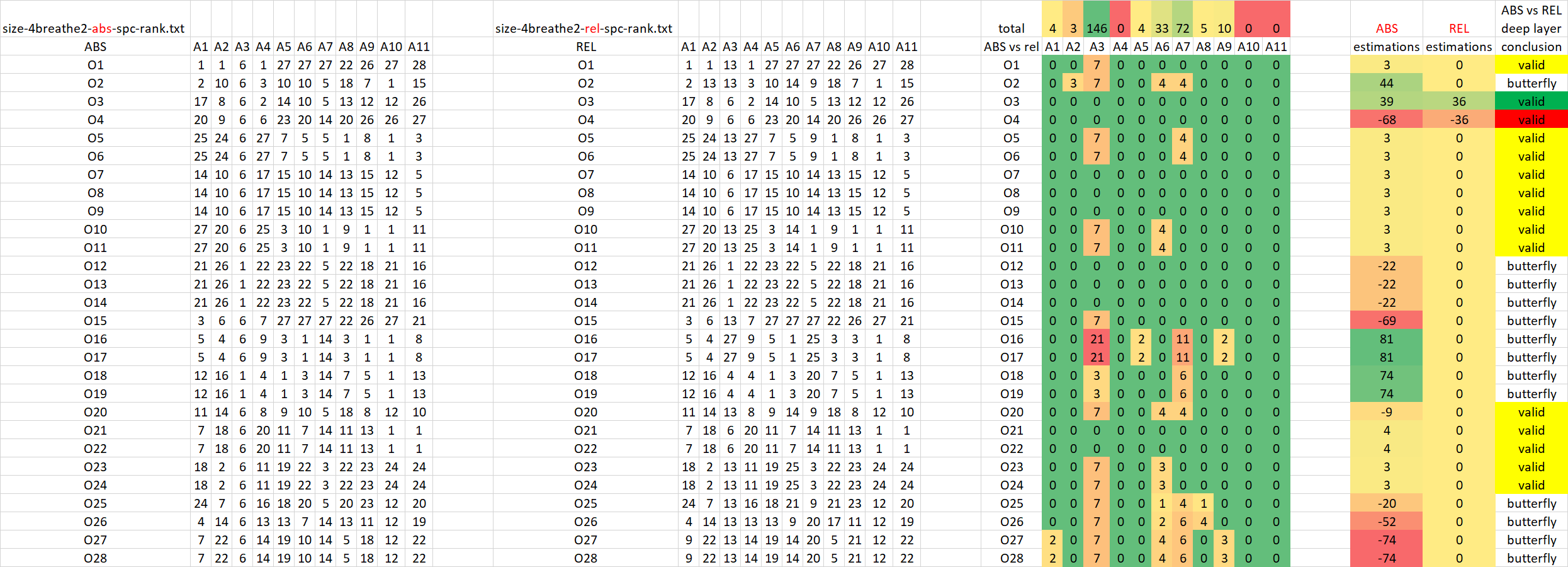
Az 1. ábra valós/anonim számértékek alapján bemutatja, hogy amennyiben egy 28 objektumból (pl. személyből) álló OAM (objektum-attribútum-mátrix) 11 attribútummal (Ai) rendelkezik, ahol az A10 és az A11 esetén nincs abszolút és relatív nézet, akkor a relativálható attribútumok (melyek alapvetően korábbi személyes idősorok szélsőségeinek pl. darabszámát határozta meg) értékét a szélsőségek levezetéséhez használt idősor hosszával relativálni lehet/kell annak érdekében, hogy összevethető (méretfüggetlen) OAM értékeink legyenek.

A bemeneti OAM értékek rangsornézete az abszolút és a relatív értelmezés keretében az 1. ábra alapján (0 = minél nagyobb, annál jobb) rangsorolást alkalmazva minden esetben a 2. ábrához vezet:



1. ábra Abszolút és relatív rangsorok (forrás: saját számítások)

A két rangsor-mátrixot a 3. ábra hasonlítja össze és ez az ábra mutat rá arra, milyen pontszerű és milyen mértékű lehet az eltérés két olyan rangsor-táblázat között, melyek mögött egyes objektumok (személyek) feldolgozásra rendelkezésre álló idősorainak hossza 1-1 egységgel tér csak el egymástól százas nagyságrend mellett:

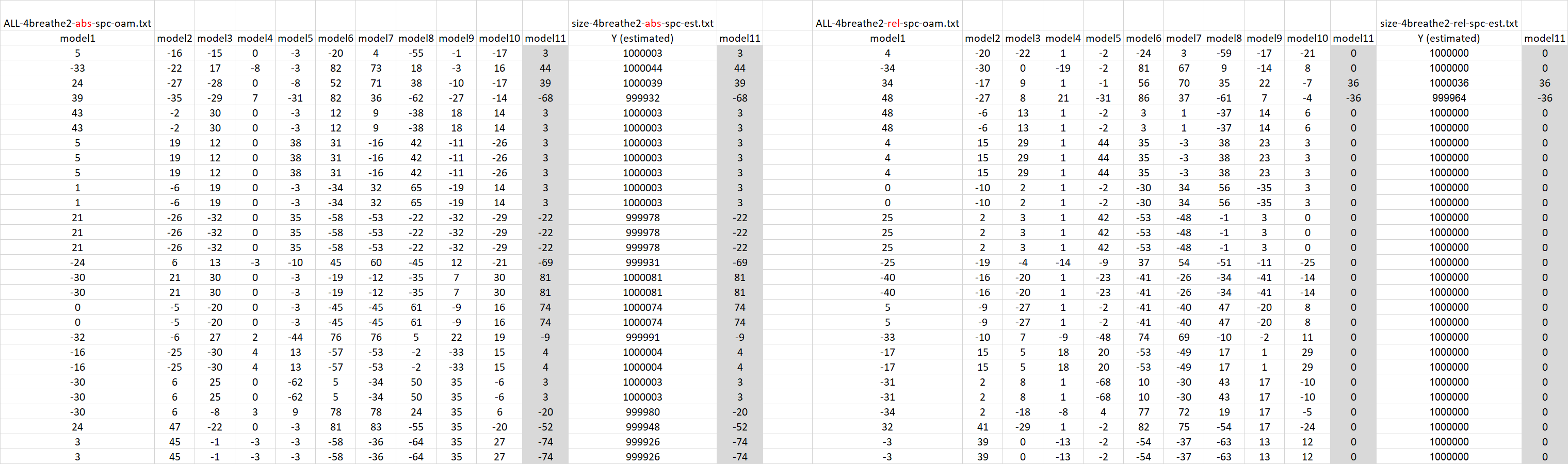


1. ábra: A rangsorok eltérésének specialitásai (forrás: saját számítások)

# A modellezési lépésekről

A 3. ábra alapján látható, hogy a relativálás által érintett nem minden attribútum vezet rangsor-eltérésekhez (vö. A4), ill. értelemszerűen az A10 és A11 önazonossága magától értetődő. Az A3 és az A7 azok az attribútumok, melyekre a pillangó-hatás elsődlegesen érvényes – bár itt még nem tudni, mely attribútumok milyen szerepet kapnak majd a hasonlóságelemzés optimalizációs rendszerében. A hasonlóságelemzés lényege, hogy a bemeneti rangsorszámokhoz ezek irányát megőrizve olyan csereértékeket rendel egy/több optimalizációs számítás keretében, mely lépcsők (súlyok) képesek a lehető legjobban közelíteni a lehet-e minden objektum másként egyforma elvet, vagyis az objektumonként azonos outputértéket úgy, hogy két szomszédos rangsorszám csereértéke minimum 1 egységgel (vö. hasonlósági univerzum gravitációja) el kell, hogy térjenek egymástól.

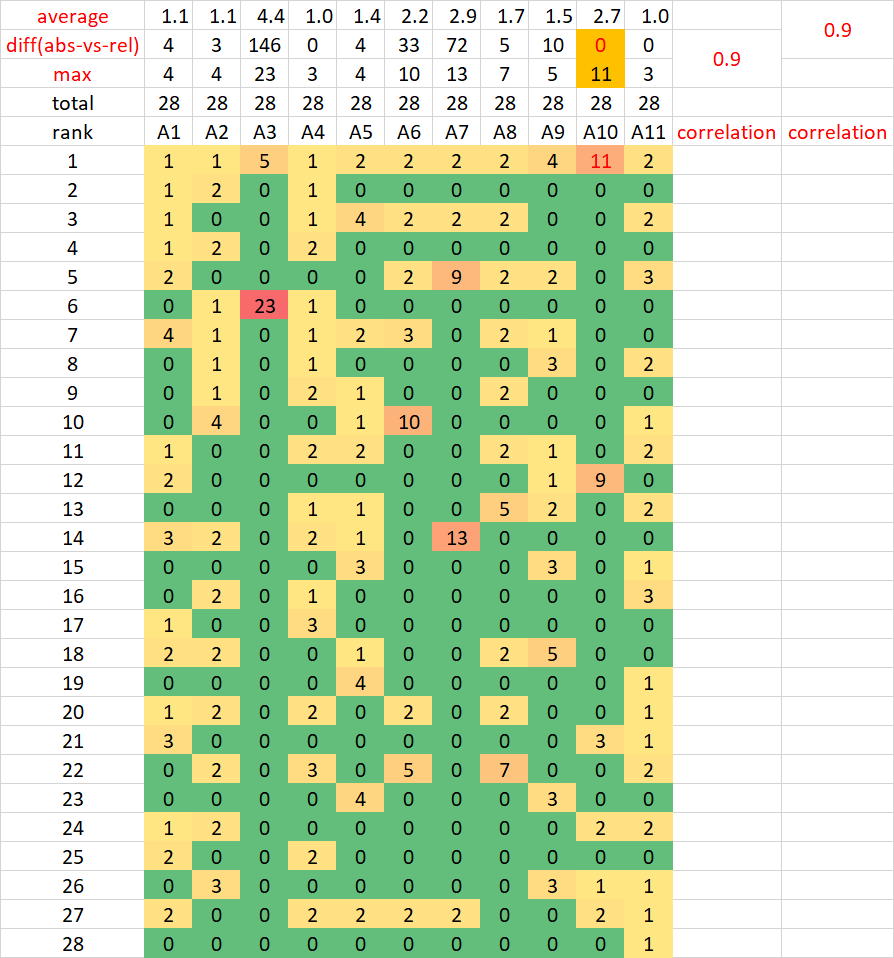
A 3. ábra jobb szélén pedig már a hasonlóságelemzési modell-láncok eredményei láthatók az abszolút és a relatív OAM-ok alapján külön-külön a minden objektum lehet másként egyforma elv közelítését kiaknázva.



1. ábra: Modell-láncok hatása a klasszifikációra (forrás: saját számítások)

A 3. ábra egyetlen egy modell esetén mutatta meg a pillangó-hatás alapját. A 4. ábra egy 10 alap és +1 zárómodellből álló modell-lánc esetén mutatja be a végső objektumértékelések levezetését, ahol a 4. ábra értékei az egyes modellek normaértéktől való +/- eltérései.

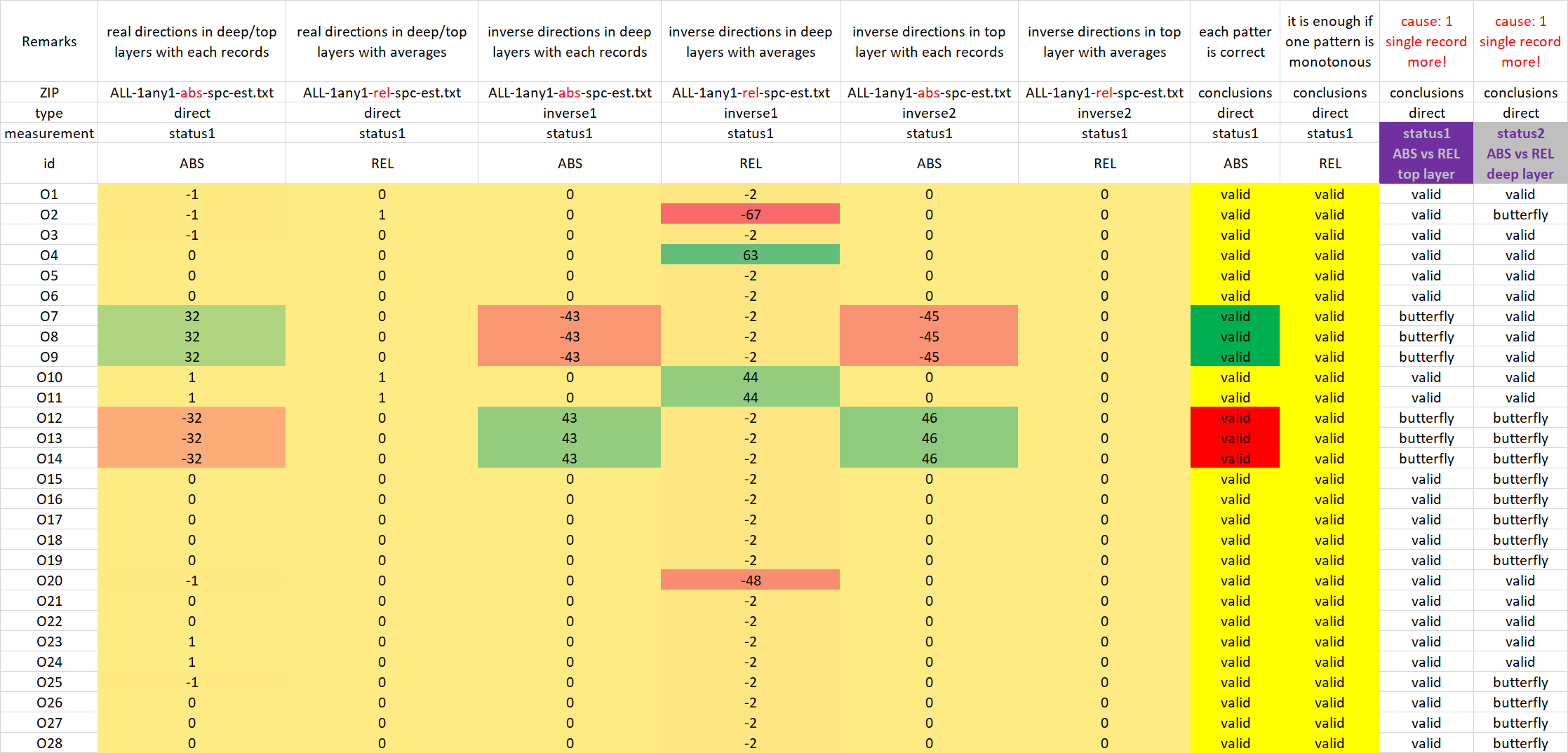
A 4. ábra világosan jelzi, hogy míg az abszolút (méretfüggő) modell-láncok a 28 személyt szinte egyenletesen osztották szét a klasszifikációs skálán, addig a relatív modell-láncok eredményeként 2 objektum kivételével minden más objektum a másként egyforma értéket vette fel úgy, hogy a két modell-lánc között az 1. ábrán látható 360 vagy 361 rekordnyi mérési adat állt rendelkezésre. Így az 1 rekordnyi eltérések kumulált hatása jelentősnek minősíthető, noha az ¼ másodpercnyi mérés-észlelési idősáv az emberi tudatos gondolkodás/észlelés számára lényegében nem is létezik.



1. ábra: A rangsor-eltérések összefüggései (forrás: saját számítások)

Az 5. ábrával visszatérünk a lényegi hatásmechanizmushoz, a rangsorolásbeli eltérésekhez. Az 5. ábra feltételes formázással ellátott teste (28\*11 cella) azt jelenti, hogy adott attribútum (Ai) esetén adott rangsorszám hányszor fordult elő a 2. ábrában az abszolút OAM-ban. Az 5. ábra tetején látható total, max és average sorok értelemszerűen a 28\*11 cella maximumát, átlagát és összegét fejezik ki. A diff(abs-vs-rel) sor cellái pedig a 3. ábra balról 3. táblázatának tetején lévő sorral azonosak.

A két korrelációs érték (0.9, ill. 0.9) azt fejezi ki, vajon a rangsorok abszolút és relatív nézetbeli eltéréseinek maximumai és a sorszámismétlődések maximumai között milyen erős az együttjárás.



1. ábra: A hasonlóságelemzés-láncok konzisztenciára gyakorolt hatása (forrás: saját számítások)

# Konklúziók

A 6. ábra megértéséhez elsőként tisztázni kell, hogy a real és az inverse irányok melletti hasonlóságelemzése a függvényszimmetriák okán egymás tükörképei ideális esetben. Vagyis a tagadás tagadása elv mentén lehet rámutatni arra, mely objektumok elemzése esetén illik a robot-klasszifikátornak azt mondania, hogy „nem-tudom” miként is értelmezendő a szóban forgó objektum, mert inverz modell-paraméterekkel elemezve az összes objektum összes attribútumát egyes objektumok esetén az adatvagyon nem engedi meg, hogy a modell-norma két oldalára kerüljenek a modelltévedések.

A 6. ábra balról 1. és 3. oszlopa tehát egymás szemmel látható tükörképe a színkódok szintjén. A 6. ábra 2. és 4. oszlopa is egymás párja, de itt az egyik nem tér el kerekítési hibán (+/-1) belül a normától, így a tükörképnézet milyensége már lényegtelen: a konklúzió elsődlegesen az, hogy nincs eltérés az objektumok között (azaz minden objektum lehet másként egyformának tekinthető). A 6. ábra fejlécében a deep és a top layer kifejezés arra utal, hogy a hasonlóságelemzési láncok egymásra épülésében vannak (deep) mélyebb rétegek és zárómodellek úgy az abszolút, mint a relatív adatokra támaszkodva.

A modellezés konklúziója pedig nem más, mint

* a relatív modellek szerint a 28 objektum inkább tűnik megfelelni a mindenki lehet másként egyforma elvnek
* mint az abszolút modellek esetén, ahol 3-3 normasértés áll egymással szemben értelemszerűen a norma két oldalán numerikusan is arányosan
* így a pillangó-effektus ezen normasértések kapcsán vélelmezhető minden esetben valid objektumú modellek mellett a (top) záró modellrétegen
* (míg a deep rétegekben a pillangó-hatások mennyisége több, de ezek a top rétegbe érve kioltják egymást – vö. 3. ábra jobb szélső oszlop = 6. ábra jobb szélső oszlop).

# Hivatkozások

(lásd a szövegközben)