

TIPOLÓGIÁKHOZ TARTOZÁS VARIANCIAANALÍZIS ALKALMAZÁSÁVAL LVA hangelemzés vertikális vizsgálata

Takács Szabolcs¹, Kis György², Makrai Balázs³, Amir Liberman⁴

¹ Károli Gáspár Református Egyetem, Pszichológia Intézet, egyetemi adjunktus,

² ANIMA Polygraph Kft, Ügyvezető,

³ Károli Gáspár Református Egyetem, Pszichológia Intézet, hallgató,

⁴ Nemesysco,

Absztrakt

A különböző tipológiák készítésének módszertana általában nagyobb adatállományokat, megfigyeléseket igénylő eljárásokat foglal magába. Ezek matematikai háttere igen komoly méreteket is ölthet – mely így nem nélkülözheti szakértő bevonását a megfelelő eredmények elérésének érdekében. Tanulmányunkban egy olyan módszert mutatunk be, ahol a szakértő bevonása nem feltétlenül a komoly matematikai háttér megteremtéséhez kellett, hanem azon adatok előállításához, melyeken utána egyszerűbb matematikai statisztikai eszköztár megfelelő alkalmazásával tipológiákat, karakterizációkat hozhattunk létre.

Az elkészített tipológiákat és a kísérlet elrendezését, annak elméleti hátterét mutatja be tematikus számunk első összefoglaló tanulmánya (Kis és társai 2017), mely összefoglaló tanulmány alapját az a kutatás jelenti, amit az ANIMA Polygraph Pszichológiai Tanácsadó Kft., a Károli Gáspár Református Egyetem Pszichológiai Intézet, a Nemzeti Védelmi Szolgálat és a Rendészettudományi Társaság közös kutatóműhelyében, a Károli Gáspár Református Egyetem hallgatóinak bevonásával végeztünk.

Kulcsszavak: hazugság ■ LVA ■ hangelemzés, tanúvallomás ■ megamisítás, elhallgatás ■ varianciaanalízis ■ valószínűség.

A kutatási engedély száma 278/2016/P. A vizsgálatban résztvevő összes vizsgálati alany írásban hozzájárult a hanganyagok és videóanyagok további kutatásokban és tudományos publikációkban való felhasználásához.

Abstract

The method of creating different typologies generally requires vast databases and number of observations. The mathematical background of that can assume considerable proportions - therefore competent professional has to be involved in order to have proper results. In our study we introduce a method where the involvement of such professionals was not required to achieve the essential mathematical background but to gather such informations which we can create typologies and characteristics by applying simpler but thorough mathematical statistics. The created typologies, the experimental settings

and the theoretical overview will be presented in our thematic article's first overall study (Kis et al. 2017) which is based on a research involving the ANIMA Polygraph Psychological Consulting Llc., the Károli Gáspár University of the Reformed Church in Hungary, the National Protection Service, the Society of Police Science and the students of the Károli Gáspár University of the Reformed Church in Hungary.

Keywords: lie ▪ LVA ▪ speech analysis ▪ deposition ▪ falsification ▪ accomplice ▪ secretive
▪ analysis of variance ▪ probability.

Research license 278/2016/P. All participants signed a written document in which they declared: all audio and video records can be used for further researches and academic publications.

BEVEZETÉS

2014 márciusában belekezdünk egy kutatásba, melynek adatfeldolgozási alapja az LVA hangelemző szoftver (Nemesysco, 2017) egy újfajta adatfeldolgozási metódusa volt. A kutatásainkat a bevezető tanulmányban (Kis és munkatársai, 2017) ismertetett kutatóműhelyen belül végeztük, melynek tagjai már több konferencián is előadták a műhely főbb, e kutatást megelőző (megalapozó) eredményeit. E kutatások részint a verbális és nonverbális jegyek vizsgálatát jelentették (Endrédi és Benczúr, 2013; Takács és Benczúr, 2013, illetve Endrédi és munkatársai, 2015), részint pedig különböző csoportokat hasonlítottak össze hazugságdetekciós készségeik tekintetében (Koch és munkatársai, 2015, illetve Tollner és Benczúr, 2013).

A hazugságvizsgálatok alapja minden esetben az úgynevezett HoT-Spot tesztek, illetve azonnali (real-time) azonosítások. Ez a mimika, testbeszéd megfigyelésén túl akár az egészen komoly műszeres vizsgálatokat is jelentheti (például MRI vizsgálatokat (lásd Lee et al., 2002)). Ezek az eljárások lényegében annyit jelentenek, hogy a hazugság nyomán fellépő úgynevezett szívárgásokat azonnal, akkor és ott tetten kell érni – máskülönben elmúlnak, nem figyelhetők meg. Az LVA szoftver ebből következően az elmondottakat hangszegmensekre bontja (megközelítően 1 másodperces időtartományokra) és ezen szegmensekben vizsgálja, hogy a vizsgálati alany önmaga korábban kalibrált értékeihez képest eltéréseket mutat-e.

Ettől eltérő irányt mutat be – szövegelemzés segítségével – Walczyk és társainak 2009-es tanulmánya, melyben azt vizsgálják, hogy egy-egy szöveg (történet) elmesélése során a vizsgálati alany gondolkodási ideje, valamint a tartalomban elmondottak ellentmondásainak analízise miként ejthető meg.

Számunkra azonban egy egész más megközelítést kellett választani – mindkét módszertantól különbözőt. Mi ugyanis a kísérleti felvetésünkben arra voltunk kíváncsiak, hogy „általában” egy történet megáll-e a lábán – de akár úgy is, hogy nem értjük a válaszádot. Nem tudjuk, mit mondott – csak azt, hogy ho-

gyan mondta. Így azt vizsgáljuk, hogy egy egész történet összességében milyen jellegzetességeket mutat. Korábbi – e folyóirat jelen számában – ismertetett tanulmányok részletesen tárgyalják azt a 4 különböző hazugsággal kapcsolatos tipológiát, melyeket vizsgáltunk (Czabán és munkatársai, 2017a, 2017b, illetve Madzin és munkatársai, 2017a, 2017b). Néhány szóban ismertetjük, hogy mely típusok vizsgálatát valósítottuk meg:

1. **Megmásító:** a megmásító egy elmesélt történet bizonyos részleteit elfedi információkkal. Mászt mond, mint ami valójában történt – és erről tud is. Ehhez az elfedéshez érdeke fűződik (Biland, 2013).
2. **Elhallgató:** a valós történésekkel tisztában van, azonban a történet bizonyos részleteit nem osztja meg a hallgatósággal (Biland, 2013). Érdeke fűződik ahhoz, hogy elhallgasson részleteket.
3. **Titkoló:** a valós történéssel tisztában van és ezt igyekszik is legjobb tudása szerint elmondani – azonban van olyan részlet az életében, melyet nem szeretne velünk megosztani. Gondolhatunk itt arra, hogy megcsalja a házastársát – de közben egy autótolvajlás szemtanújaként kerül kihallgatásra.
4. **Patetikus:** esetében olyan elkövetőről beszélünk, aki megbukik a kihallgatáson. Mind a kérdező, mind a kérdezett számára nyilvánvaló, hogy elkövette az adott tettet, melyben megtévesztést igyekezett alkalmazni – és már semmifajta érdeke nem fűződik az egész történet átadásához, vagy annak elhallgatásához, megmásításához.

Olyan kísérletet végeztünk, melyben a kísérleti alanyokról pontosan tudtuk, hogy a fenti négy típusból melyik kategóriákba sorolhatók. Mellettük voltak olyanok is, akik mindenfajta szerep nélkül vettek részt a kísérletben (lényegében megfigyelőként), nekik csak el kellett mesélniük, hogy mit is láttak a kísérlet során. Ők képezték a kontroll csoportot.

A matematikai modellünk tehát az volt, hogy minden vizsgálati alany esetében pontosan tisztában voltunk a csoportjukkal. A célunk kettő volt:

1. Szerettünk volna minél pontosabb leírást, karakterisztikát adni az egyes csoportokról.
2. A kialakított karakterisztika alapján tetszőleges új vizsgálati alany érkezése esetében szerettük volna meghatározni a mért értékei alapján, hogy melyik csoporthoz tartozik.

Világos, hogy ha a csoportok nem lettek volna ismertek, akkor egyfajta klaszterezési eljárással kellett volna dolgoznunk. A fix 4+1 (4 típus + kontroll) csoport miatt, melyet kerestünk a k-központú klaszterezés lehetett volna adekvát. Valamint – ha azt is feltételezzük, hogy egy-egy elbeszélés során akár változathatja is a csoportokat a vizsgálati alany, akkor modell alapú klaszterezések is szóba jöhetnek. Ilyen lehet például, ha a vizsgálati alany egyes kérdésekben megmá-

sítja a történetet, míg más részleteit elhallgatja, illetve bizonyos dolgokat ősztintén elmond. A két módszer kombinálására is vannak példák, ilyen többek között Kababe és Gaikwad tanulmánya (Kababe & Gaikwad, 2013), akik MRI felvételek osztályozására és besorolására alkalmazták a két módszer ötvözését.

Azonban azt is láthatjuk, hogy a fenti modellünk nem ez: ismerjük a részleteket, tudjuk, hogy mely alanyunk hová tartozik, tehát a karakterisztika kialakítására a klaszterezési eljárásokat elvetettük. Statisztikai eszköztárak közül a diszkriminanciaanalízis és a logisztikus regresszió látszott még adekvátnak (a két eljárás közötti különbségekről értekeznek Press és Wilson, 1978-as tanulmányukban). A két módszer között természetesen dönthetünk matematikai statisztikai alapon is – követve Press és Wilson tanácsait. Azonban egyszerűbbnek tűnt empirikusan közelíteni a problémához: kipróbálva mindkét módszert, egyik sem hozott megfelelő eredményeket.

Különböző osztályozási módszereket is bevethettünk volna (például LAD, melynek technikai leírását lásd például Boros Endre tanulmányában (Boros et al., 2000), illetve egy betegségazonosításban való felhasználását Hammer és Bonates 2006-os tanulmánya). Azonban ezek számításigénye, valamint (és ez volt jelen kutatásban a fontosabb) későbbi felhasználhatósága a kutatásban szakértőként és vezetőként részt vevő ANIMGROUP (Animagroup, 2017), illetve a szoftver tulajdonos NEMESYSO (Nemesysco, 2017) kutatói számára nehézkessé bizonyult, így ezeket bár kipróbáltuk, végül elvettük és a jelen tanulmányban bemutatott, saját módszert alkalmaztuk.

Az alkalmazott módszerünket több konferencián is részletesen bemutattuk (a kísérlet többi elemével együtt is, illetve külön-külön is (Kis és társai, 2016), illetve (Takács, 2016)). Az előadások után érkezett visszajelzések szerinti apróbb módosításokat beépítettük az ismételt kontroll kísérleteinkbe is – ezekről a tekintés fejezetben ejtünk részletesen szót. A módszertanunk leírása során a kísérletben alkalmazott példákat is bemutatjuk, hogy maga a matematikai modell jobban érthetővé váljon. Mint az majd látható lesz: a kialakított módszerünkben megfelelő pontbecsléseket és intervallumbecsléseket (Vargha, 2005; Vargha, 2016) kombinálunk egymással oly módon, hogy a kialakított súlyozások segítségével a változók szelekcióját és a vizsgálati alanyok típusoktól való „távolságát” – típusokba való tartozások mértékét – definiálni tudjuk (Surányi, Babocsay, Takács & Vargha, 2011).

Így analógiájában a varianciaanalízis és a modell alapú klaszterezés egyfajta keverékét állítjuk elő, hiszen a csoportokat különböző, megfelelően kiválogatott változók segítségével jellemezzük – majd az így kialakított karakterisztikát modellként használva, a csoporthoz tartozás mértékét definiáljuk a csoportok centrumainak segítségével.

A MÓDSZERTAN LEÍRÁSA

Tételezzük fel, hogy adott K darab csoport, mely csoportokat varianciaanalízis segítségével meg tudunk vizsgálni több változó mentén is (lásd például Vargha, 2015). Először mindösszesen egyetlen változó mentén vizsgálódunk, legyen ez a változó X . Jelölje az X változó átlagát M , melyet az aktuális vizsgálati alany mutat. Továbbá a már bemért, csoportonkénti átlagainkat jelölje M_i . Tegyük fel továbbá, hogy a csoportok átlagainak nagyság szerinti sorrendje M_1, \dots, M_K .

Ezek után definiáljuk az alábbi távolságokat:

$$d_i = d(M, M_i)$$

ahol például $d_i = d(M, M_i) = |M - M_i|$. Világos, hogy amikor 0 a távolság, akkor az azt jelenti, hogy azon csoportba férne bele leginkább az értékek alapján a vizsgálati alanyunk – egészen pontosan az adott alanyunkat pontosan ebbe a csoportba kell sorolnunk.

Például: a bevezető fejezetben említett vizsgálatunkban egy-egy vizsgálat során az LVA szoftver több változót is számít. Koncentráció, közlési hajlandóság, frusztráció. Legyen most az X változónk a Koncentráció. A vizsgálati mintánk alapján meghatározzuk, hogy az elhallgatók, a titkolók, elkövetők vagy a patetikusok (illetve a kontroll csoport) milyen átlagos értékekkel rendelkeznek (ezek lesznek az M_i értékek). A példánkban $K = 4$ (a kontroll csoport és a patetikus csoport ugyanis nem különbözött egymástól). Továbbá, tetszőleges új vizsgálati alany hanganyagából számított koncentráció értékének átlaga lesz M .

Ez indokolja, hogy képezzük az alábbi mértékeket:

$$\sum_{i=1}^K d_i = D$$

$$q_i = 1 - \frac{d_i}{D}$$

$$\sum_{i=1}^K q_i = Q = K - 1$$

$$p_i = \frac{q_i}{K - 1}.$$

A fentiek alapján tehát D a teljes távolságok értéke, míg így a különböző q értékek adják azt, hogy az adott távolság alapján „mennyire valószínű” az adott cso-

porthoz tartozás. Természetesen ez nem egy valódi valószínűség (hiszen összegük nem 1) – de analógiájában hasonló elveket mondhatunk róluk.

Továbbá ezen mértékek összességén (Q) belül a különböző kategóriák összesen így p_i mértékeket alkotnak, melyből minden csoporthoz való tartozást meg tudunk becsülni, fel tudunk írni. Ezek után minden mért változó esetében definiálni tudunk tehát egy p_i mértéket, mely azt mutatja meg, hogy milyen értéket tudunk az adott csoporthoz tartozásához rendelni az adott változó alapján.

Példa: a példánk során tehát (maradva továbbra is a koncentrációnál) megmondhatjuk, hogy a vizsgálati alanyunk milyen távol esik általában (az átlag alapján) az elkövetőtől, az elhallgatótól, a titkolótól vagy a patetikustól/kontrolltól. Értelemszerűen minél nagyobb egy p_i érték, annál nagyobb a valószínűsége, hogy a vizsgálati alanyunk (a koncentrációja alapján) az adott csoporthoz tartozik.

Természetes azonban, hogy nem minden változóval tudunk azonos súllyal csoportokat meghatározni. Így a változóinkat is súlyoznunk kell az alapján, hogy mennyire jól lehet a segítségükkel azonosítani a csoportjainkat. Minden változónak kialakítottunk egy 1 és K közötti súlyt attól függően, hogy „mennyire markánsak” az adott változó átlagai, illetve a csoport hatása mennyire érvényesül az adott változó mentén. Ezt az alábbi módon érhetjük el.

A varianciaanalízis során fentebb már említettük, hogy az M_i átlagokat nagyság szerinti növekvő sorba rendeztük. Ekkor az átlagok „csoportjait” alakítjuk ki a következő módon. Legyen M_1 a legelső csoportban. Ha M_2 átlag ettől nem különbözik szignifikánsan, akkor M_2 kerüljön az M_1 által reprezentált első csoportba, máskülönben nyissunk neki új csoportot. Ekkor M_3 (az első esetben) kerülhet az $[M_1; M_2]$ átlagok által reprezentált első csoportba vagy nyithat új csoportot – vagy a második esetben kerülhet az $[M_2]$ által definiált csoportba, vagy nyithatunk neki új csoportot.

Megjegyzés: bár a standard hibák alapján kerülhetne elvileg az $[M_1]$ csoportjába úgy az M_3 átlag, hogy az M_2 már új csoportot alkot – ilyenkor azonban a nagyság szerinti sorrend miatt M_3 valójában az M_2 értéktől sem különbözhet szignifikánsan, ezért ilyen esetben a nagyobb értékű $[M_2]$ által reprezentált csoportba soroljuk.

A fenti eljárást folytatjuk addig, amíg minden M_i érték besorolásra nem került. A változó súlyát a kialakított csoportok száma jelenti. Egészen pontosan úgy súlyozunk, hogy az $1, \dots, K$ súlyokból elveszünk egyet – majd az így 0 súlyra került változókat töröljük a későbbi elemzésből. Erre azért van szükség, mert az a változó, mely csak egyetlen csoportot hozott létre valójában azt jelenti, hogy nem képez szignifikáns különbséget az átlagok között – tehát nem képes differenciálni semmilyen módon. Így a változók számának csökkentése érdekében érdemes e változótól megszabadulni. Tehát minden változó esetében tudunk defini-

álni egy-egy súlyt attól függően, hogy az adott változó alapján hány csoport alakítható ki: legyenek az alábbi súlyok létrehozva ($j = 1, \dots, S$), amennyiben S változó van a változók törlése után.):

$$\tau_j = 1, \dots, S$$

$$T = \sum_{j=1}^S \tau_j$$

Ezek után

$$w_j = \frac{\tau_j}{T}$$

A csoportokhoz való tartozás mértéke így nem más, mint:

$$P_i = \sum_{j=1}^S w_j p_{i,j}, i = 1, \dots, K$$

A w_j súlyok tehát minden változó relatív súlyát adja meg az összes változóval való elemzésben, így az is érthető, hogy a w_j súlyok összege 1. Tehát lényegében a $p_{i,j}$ értékek súlyozása egyfajta kombinációját (konvex kombinációját) adja az eredeti, egy-egy változón mért mértékeknek.

Megjegyzés: a fentiekben tehát a w_j súlyok azt adják meg, hogy egy változó mennyire képes jól meghatározni egy-egy csoportot. Minél nagyobb a súly, annál jobban képes differenciálni a csoportokat – tehát annál fontosabb számunkra egy-egy értékelésnél.

Példa: esetünkben a koncentráció súlya $K - 1$ volt (hiszen differenciálni tudta a csoportokat), de például a „szexuális kíváncsiság” változó jelen esetben nem differenciált, tehát kikerült az elemzésből, 0 súlyt kapott a csoportokhoz való tartozás azonosításában.

A BESOROLÁS FINOMÍTÁSA

Természetesen tovább finomíthatjuk a fenti modellünket az alábbi eljárás segítségével. Tudható, hogy a hangelemzés során a beszédünk egy-egy másodperc hosszúságú szegmensei semmiképpen sem nevezhetők egymástól függetlennek. Egy adott mondatunk, melyet akár 5-10 szegmensre is szétbonthatunk – bár

5-10 „esetként” jelenik meg az adattáblában, valójában nem tartalmaz ennyi, egymástól független információt.

Ennek kiküszöbölésére az IBM SPSS programcsomag leíró statisztikai jellemzőinek bootstrap szimulációval becsült értékeit vesszük alapul (A bootstrap elemzésről lásd például Efron & Gong 1983-as dolgozatát.) Így minden csoport és változó esetében készül egy átlag, melyhez a szimuláció segítségével meghatározzuk annak 95%-os konfidencia-intervallumát. Jelöljük ezeket az alábbi módon:

$$[M_{Li}; M_{vi}], ..., [M_{Lk}; M_{vk}],$$

Legyen továbbá az adott paraméter vizsgálatokor $[C_{Li}; C_{vi}]$ az adott paraméter átlagának (hasonlóan bootstrap eljárással meghatározott) 95%-os konfidencia-intervalluma.

Megjegyzés: így nem az átlagoknak egy pontbecslését használjuk, hanem minden változó esetében az adott csoportok átlagának intervallum-becslésével dolgozunk. Így a becslésükben szereplő standard hibát is igyekszünk még erőteljesebben figyelembe venni a csoportba sorolások során. Így tehát: M jelöli minden esetben a megállapított, standard paraméterértékeket – és C az új mintából számított, vizsgálati egyén által szolgáltatott értéket.

Jellemzően $[M_{Li}; M_{vi}]$ intervallumok lényegesen „szűkebbek”, mint $[C_{Li}; C_{vi}]$ intervallum, hiszen míg az előbbi többszörösen mért és tisztított mintázatból származik (sok adattal), addig az új felvétel által meghatározott értékek eset-számai lényegesen (akár nagyságrendekkel) kisebbek. A tipológiák és az új felvétel intervallumai tehát jellemzően 3 lehetséges elrendeződést mutathatnak.

1. Az új intervallum teljes egészében az eddigi intervallumok felett van:
 - a. $C_{U_i} > MAX(M_{U_i}), i \in [1..k]$.
2. Az új intervallum teljes egészében az eddigi intervallumok alatt van:
 - a. $C_{L_i} < MIN(M_{L_i}), i \in [1..k]$.

A harmadik lehetőség az, hogy az új intervallum valahol a többi között helyezkedik el, közös metszetet kialakítva velük.

Megjegyzés: előfordulhat az az eset is (bár ez egyszer sem fordult elő a tapasztalatok alapján), hogy az intervallumok közé esik úgy, hogy két intervallum alsó és felső határa között helyezkedik el teljes egészében. A felsorolt első két esetben, valamint e speciális esetben is használjuk az előző (nem finomított) változatot a becslés kialakítására.

A harmadik esetben azt mondhatjuk, hogy

$$\exists i: |[C_{L_i}; C_{U_i}] \cap [M_{L_i}; M_{U_i}]| > 0.$$

Definiáljuk az alábbi mennyiségeket: miután van közös része az intervallumoknak, legyenek definiálva

$$q_i = \frac{|[C_{L_i}; C_{U_i}] \cap [M_{L_i}; M_{U_i}]|}{D},$$

ahol:

$$D = \sum_{i=1}^K |[C_{L_i}; C_{U_i}] \cap [M_{L_i}; M_{U_i}]|$$

Megjegyzés: bár korábban már használtuk a q_i jelöléseket – de tekintettel arra, hogy itt lényegében analóg mennyiségekről van szó, mint amiket az átlagok pontbecslésének segítségével alkottunk meg, ezért mind a D , mind a q jelölést meghagytuk.

Világos, hogy ezzel az átlagok közötti távolságokat kicseréltük az intervallumok metszetének relatív mértékére. Természetesen így több olyan q_i mennyiség is kialakulhat, melyek értéke 0 lesz. Jogosnak tekinthető azonban ez, hiszen ezekben az esetekben a becsléseink, átlagos értékeink szerint a vizsgálati alanyunk „igen messze” került az adott csoporttól vagy csoportoktól.

A q_i értékek definiálása után ugyanúgy járhatunk el, mint az előző esetben – tehát kialakítjuk a megfelelő súlyozását a változóknak, majd a súlyozásból és a most kiszámított távolságokból kiszámítjuk a csoportokba tartozás valószínűségét.

A KÍSÉRLETI HELYZET ÉS AZ EREDMÉNYEK BEMUTATÁSA

Kísérleti helyzetünkben egy olyan egyetemi kísérletet választottunk, melyben a vizsgálati alanyok a maffiajáték nevű csoportos játékban vettek részt (tudományos háttére is van a játéknak, lásd például Braverman és munkatársainak 2008-as tanulmányát, melyben a játékban szereplők információs torzításait, illetve kiegyensúlyozatlanságát vizsgálják). Természetesen e játékban nem a győzelmi esélyeket latolgattuk. Az volt a cél és a feltételezés, hogy a játékban való szerepek és a kihallgatás előtt tett instrukciók (tilos a szerepet bárkivel megosztani) elegendőek lesznek az érzelmi bevonódáshoz. A kiosztott szerepek után, a játék egy adott pontján a játékosokat LVA hangelemzésre vittük (Nemesysco, 2017), ahol különböző kérdéseket tettünk fel számukra. A hang-

elemzés során számos változó értékét folyamatosan mérjük, amit a program regisztrál egy-egy kísérleti személy adatállományában.

A hangelemzés alapjául az alábbi 4+1 darab csoport összehasonlítása szolgált. Ezek a csoportok:

1. Patetikus (lebuktatott elkövető, lebuktatott *maffiatag*)
2. Elkövető (valamely bűneset tevőleges résztvevője – megmásítja a történetet, *maffiatag*)
3. Titkoló (valamely más dolgot – például szexuális orientáció – titkolja, a kérdéses tárgykörben ártatlan, nincs oka semmit sem elhallgatni vagy megmásítani, *felügyelő*).
4. Elhallgató (az adott bűnesettel kapcsolatosan tevőleges ráhatása nincsen, viszont olyan információkkal bír, melyeket nem szeretne átadni a kikérdezőnek, *kollaboráns*)
5. Kontroll (nincsen oka titkolózni, elhallgatni vagy megmásítani, *polgár*)

Az elemzés során arra voltunk kíváncsiak, hogy a különböző mért változók értékei alapján meg tudjuk-e azt jósolni, hogy melyik csoportba kerülne egy új játékos (vagy akár éles helyzetben egy kihallgatáson részt vevő tanú).

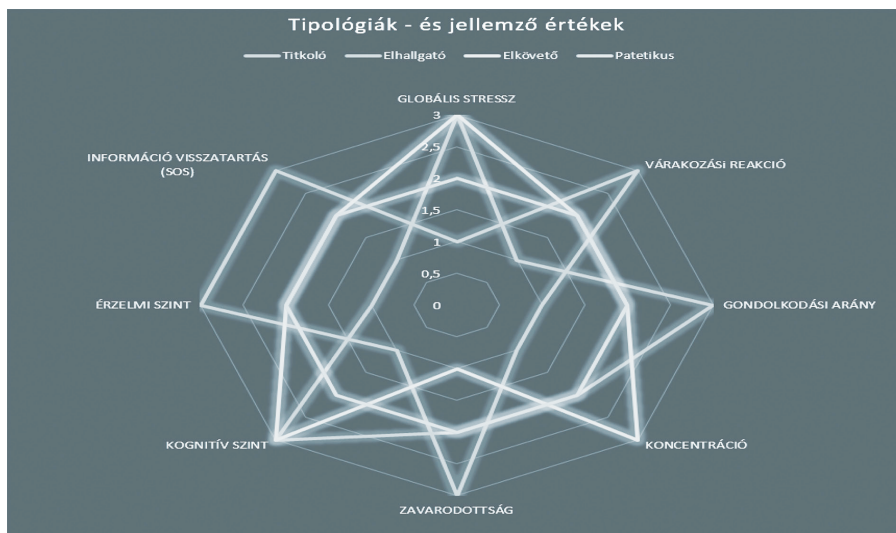
Ez az eljárás tehát alapvetően diszkriminanciaanalízisre, vagy logisztikus regresszió alkalmazására vezetne (Press és Wilson, 1978), azonban egyik módszer sem szolgáltatott elfogadható, jól használható eredményekkel. Ezért is alakítottuk ki e fenti vizsgálati helyzet elemzésére a már ismertetett, varianciaanalízisen alapuló osztályozási módszerünket.

A vizsgálat során 80 fővel dolgoztunk, akiket két csoportba osztottunk véletlenszerűen: az első 40 fő adataiból (egy-egy személyre 100-200 hangszegmens keletkezik a hangelemzés során) meghatározzuk a módszerben megadott konfidencia-intervallumokat minden fontosabb változóra, majd a következő 40 fő segítségével keresztvaliditási eljárást alkalmaztunk (Takács & Kárász, 2014).

Megjegyzés: a saját mintánk keresztvalidálásán túl egy kontroll egyetemi kísérletet is végeztünk (találati arány mindkét esetben 85% feletti volt), azonban az ANIMAGROUP¹ cégcsoport segítségével (akik az LVA szoftver kizárólagos forgalmazásával rendelkeznek Magyarországon) különböző bűnesetek kihallgatási hanganyagainak változóihoz is hozzáfértünk. Ezekben az utóbbi esetekben magukat a hanganyagokat nem kaptuk meg, kizárólag az LVA programból exportált adatokat, táblázatos formátumban – melyből a típusokhoz tartozások valószínűségeit számítottuk. A találati arányaink a visszajelzések szerint itt is 70% felettiek voltak (értelemszerűen itt az ellenőrzést a rendőrségi kihallgatást végző szakértők végezték el).

¹ www.animagroup.hu

Az elemzés során azt kaptuk, hogy a 8 legfontosabb változó esetében az alábbi eredmények adódnak:



1. ábra. Tipológiai ábrázolása a 8 legfontosabb paraméter terében

Az 1. ábrán látható információk az alábbiak szerint értelmezhető:

1. Az elkövetők jellemzően magas globális stresszel, koncentrációval és jellemzően magas kognitív szinttel rendelkeznek (tehát általánosságban magas, koncentrált mentális folyamatok jellemzik őket). Ezzel szemben a zavarodottságuk kifejezetten alacsony, a többi változó mentén jellemzően közepes értékeket mutatnak.
2. Az elhallgatók (rendelkeznek információkkal, amiket nem akarnak átadni és szisztematikusan elhallgatják az ezzel kapcsolatos részleteket): magas az információvisszatartás szintje és a várakozási idő (reakciókra vár), ahogyan a zavarodottságuk, valamint az érzelmi szintjük is. Alacsony a globális stressz és a kognitív stressz szintje, illetve a koncentráció és a gondolkodási arány értékük.
 - a) Összességében tehát inkább érzelmileg érinti meg őket a helyzet, mentális folyamataik kevésbé koncentráltak (nem akarnak beszélni, azt figyelik, hogy eleget mondtak-e már, „tovább léphetnének”-e?).
3. A titkolók (más valamit nem akarnak elmondani) magas globális stressz értékekkel bírnak, a kognitív szintjük magas van, magas gondolkodási aránnyal (igyekeznek az adott ügygel kapcsolatosan minden fontosat elmondani, de csak arról beszélni, semmi mást nem belekeverni az ügybe). Mindez alacsony várakozási szinttel és alacsony közlési szinttel bír (szívesen beszélnek és szinte azonnal válaszolnak a kérdésekre).

4. A patetikusok értékei a kontroll csoport középvezetékén helyezkednek el – látható, hogy minden értékük átlagos lesz (mely átlagos értékek a kontroll csoport átlagos értékeit jelenti).

A különböző változók kísérletben tapasztalt átlagos értékei helyett minden esetben egy 1-3 skálára transzformált értéksort ábrázoltunk és használtunk.

Világos, hogy bármely új vizsgálati alany esetében azt vizsgáljuk, hogy a fenti eredményekhez milyen mértékben vannak közel az új alany értékei – és ezek után a típusokra vonatkozó százalékos értékeket tudjuk megmondani.

Fontos kiemelni azt is, hogy egy-egy konkrét vizsgálati személy értelem-szerűen kevert típusokat is létrehozhat: bizonyos kérdésekben egyik, más kérdésekben másik helyzetben van. Egyes dolgokat elhallgat, másokban teljesen őszinte, harmadik típusban pedig megmászja a történeteket. Az a felhasználó „adattisztítási” feladata, hogy egy-egy hanganyagban a konzisztens részeket összehangolja (például egy-egy kérdésre adott választ külön-külön értékeljen). Minél pontosabb az adatgyűjtés módszere, annál valószínűbb a tiszta kép kialakítása és a pontosabb illeszkedés elérése.

KITEKINTÉS

A tipológiák azonosítása során mind a keresztvalidálás során, mind a kontroll kísérletben, mind pedig a nem laboratóriumi kísérletekből származó adatok esetében a véletlen azonosításhoz képest (25%-os valószínűségi szint a négy típusra) lényegesen, magasabb, 75%-ot is meghaladó eredményeket értünk el.

Természetesen a kísérlet során figyelemmel kellett lennünk két igen fontos korlátra:

1. A hazugságok kutatásában alapvető jelentősége van annak, hogy tétet tudjunk képezni a kísérletben résztvevők számára (Biland, 2013). Ez lényegében azt jelenti, hogy tét nélkül nincsen hazugság.
2. A hazugságnak több csatornája is van, melyeket egyöntetűen érdemes figyelni (Biland, 2013).

A harmadik kulcskérdést szerint a hazugság megnyilvánulásai jellemzően egyedi jellegzetességek. Ezt az LVA segítségével úgy tudjuk mérni, hogy minden vizsgálati alanyt kalibráció után a saját beszédstílusához hasonlítunk, tehát valójában azt vizsgálhatjuk, hogy bizonyos helyzetekben önmagukhoz képest eltérően reagálnak-e.

Az első kísérleti kritériumot igen nehézkes laboratóriumi körülmények között megteremteni (Biland, 2013), ezért is volt szükségünk arra, hogy ha másként nem is: kontrollként valódi kihallgatások, bűntények hanganyagaiból származó számadatokkal tudjuk kontrollálni és finomítani kísérleteink eredményeit.

A második kritérium nagyobb munkabefektetést igényelt: a kísérletünk megismétlése során a kihallgatásokat HD kamerákkal felszerelt teremben végeztük, továbbra is a kihallgatásban jártas szakértők segítségével. A HD kamerák felvételei alkalmasak arra, hogy a hazugság további csatornáit és az eddig meglévő hangelemzésből származó információk összekapcsolásra kerüljenek.

Jelen tanulmányunk a kísérletsorozat első eredményeinek matematikai hátterét igyekezett bemutatni. Természetesen több módszer is rendelkezésünkre állt volna, melyek közül egyesek eredménytelenségük, mások az utólagos felhasználásuk miatt kerültek elvetésre.

Módszerünkről az eredmények igazolják annak helyességét, amellet, hogy kifejezetten egyszerű matematikai statisztikai eszközök kombinálásával került kialakításra. Meglátásunk szerint számos egyéb területen is bevethető:

1. Különböző, már diagnosztizált betegségek különböző teszteken elért „átlagpontszámok”-hoz való hasonlóság alapján való besorolás;
2. Érdeklődési területek, teljesítmények alapján jövőbeni pályaválasztás elősegítése;
3. Munkahelyi beválás (mely területeken dolgozók mely készségekben jók és melyekben kevésbé kiemelkedők) jóslása adott területeken;

Alkalmazási szempontból fontos tudni, hogy a fenti számítások milyen időtartamot ölelnek fel. A tipológiák kiértékelése (ha már tiszta a hanganyag) lényegében azonnali, hiszen egy-egy átlag/standard hiba kiszámítása akár több ezer tiszta hangszegmens esetében is másodpercek alatt elkészül. Így az algoritmus leglassabb eleme (bootstrap minták alkalmazása) is néhány másodpercet jelent, a többi számítás pedig azonnal rendelkezésre áll (tetszőleges táblázatkezelő programban elvégezhető számítások). Így ha tiszta hanganyaggal rendelkezünk, akkor akár azonnali értékelések is készíthetők az eljárás segítségével.

Természetesen a fenti kérdések tanuló algoritmusok segítségével is megtárgyathatók, azonban akár egy kórházban, akár egy nagyobb cégnél valószínűtlen akkora tanuló adatbázis összegyűjtése, melyre egy tanuló algoritmusnak szüksége lenne. Főként akkor, ha figyelembe vesszük azt, hogy a hazugságnak általánosságban a legjobb felderítési módja az, ha egyéni jellemzőket tudunk megfelelően megfigyelni. Ekkor a vizsgálati alany nyugalmi és hazugság közbeni reakcióit tudjuk felmérni (Biland, 2013). Esetükben éppen ezt tesszük a hangelemzéssel: a most bemutatott módszer egyszerű átlagokkal és hibaszámításokkal, erősebb matematikai feltételek nélkül (szemben például a diszkriminanciaanalízissel). Mindezt úgy tettük, hogy a vizsgálati alanyokat saját, tét nélküli helyzetükhöz tudtuk mérni – majd ezen sok, egyéni megfigyelések összességéből általánosabban is érvényesíthető, tipikus jellemzőket tártunk fel.

BIBLIOGRÁFIA

- Biland, C. (2013). *A hazugság pszichológiája*. Budapest: Háttér Kiadó.
- Boros, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., & Muchnik, I. (2000). An Implementation of Logical Analysis of Data. *IEEE Trans. on Knowledge and Data Engineering*, 12(1), 292-306.
- Braverman, M., Etesami, O., & Mossel, E. (2008). Mafia: A theoretical study of players and coalitions in a partial information environment. *The Annals of Applied Probability*, 18(3), 825-846.
- Endrédi Zs., Tollner V., Kis Gy. & Benczúr L. (2015). A hazugság-felismerés kognitív szempontú lehetőségei. Előadás a Magyar Pszichológiai Társaság XXIV. Tudományos Nagygyűlésén, 2015. május 28-30. Eger. In Lélek-net a léleknek: Az ember a változó technikai közegek világában. (Kivonat-kötet). p. 84.
- Endrédi Zs. & Benczúr L. (2013). A látszat néha csal: verbális és nonverbális jegyek szerepe a hazugság felismerésében. Előadás a Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésén. Budapest, 2013. június 5-7. In Vargha A. (2013). Kapcsolataink világa. A Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésének Kivonat-kötete, MPT, p. 33.
- Hammer, P. L. & Bonates, T. O. (2006). Logical analysis of data – An overview: From combinatorial optimization to medical applications. *Annals of Operation Research*, 148(1), 203-225.
- Efron, B. & Gong, G. (1983). A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. *The American Statistician*, 37(1), 36-48.
- Kababe, R. S. & Gaikwad, M. S. (2013). Segmentation of Brain Tumour and Its Area Calculation in Brain MR Images using K-Mean Clustering and Fuzzy C-Mean Algorithm. *International Journal of Computer Science & Engineering Technology (IJCSET)*, 4(5), 524-531.
- Kis Gy., Vámosi K., Czabán Cs., & Takács Sz. (2016). A maffiajáték tipológiai - a hangelemzés új adatfeldolgozásának első eredményei, In: Vargha András (szerk.) Múlt és jelen összeír: A Magyar Pszichológiai Társaság XXV. Jubileumi Országos Tudományos Nagygyűlése Kivonat-kötet. 418 p. Konferencia helye, ideje: Budapest, Magyarország, 2016.06.02-2016.06.04. Budapest: Magyar Pszichológiai Társaság, 2016. pp. 134-135.
- Koch B., Kis Gy. & Benczúr L. (2015). A közösségi oldalak, mint hazugságaink iskolái. Előadás a Magyar Pszichológiai Társaság XXIV. Tudományos Nagygyűlésén, 2015. május 28-30. Eger. In Lélek-net a léleknek: Az ember a változó technikai közegek világában. (Kivonat-kötet). p. 85.
- Lee, T. M. C, Liu, H. L., Tan, L. H., Chan, C. C. H., Mahankali, S., Feng, C. M.,...Gao, J. H. (2002). Lie Detection by Functional Magnetic Resonance Imaging. *Human Brain Mapping*, 15(3), 157-164.
- Press, S. J. & Wilson, S. (1978). Choosing between Logistic Regression and Discriminant Analysis. *Journal of American Statistical Association*, 73(364), 699-705.
- Surányi Zs., Babocsay Á., Takács Sz., & Vargha A. (2011). Új klasszifikációs módszerek a személyiségpszichológiában. *Pszichológia (MTA Pszichológiai Intézet)*, 31(4), 317-340.
- Takács E. & Benczúr L. (2013). A verbalitás és mimika szerepe a hazugságdetekcióban. Előadás a Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésén. Budapest, 2013. június 5-7. In Vargha A. (2013). Kapcsolataink világa. A Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésének Kivonat-kötete, MPT, p. 33.
- Takács Sz. & Kárász J. (2014). Egy egyszerű keresztvalidálási eljárás bemutatása. *Psychologia Hungarica Caroliensis*, 2(1), 65-77.

- Takács Sz. (2016). Az LVA650 adatainak újfajta feldolgozási módszere tipológiák keresésére, In: Vargha András (szerk.) Múlt és jelen összeír: A Magyar Pszichológiai Társaság XXV. Jubileumi Országos Tudományos Nagygyűlése Kivonatkiadása. 418 p. Konferencia helye, ideje: Budapest, Magyarország, 2016.06.02-2016.06.04. Budapest: Magyar Pszichológiai Társaság, 2016. p. 135.
- Tollner V. & Benczúr L. (2013). Hallássérült, látássérült és egészséges személyek hazugság-felismerési különbségei. Előadás a Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésén. Budapest, 2013. június 5-7. In Vargha A. (2013). Kapcsolataink világa. A Magyar Pszichológiai Társaság XXII. Országos Tudományos Nagygyűlésének Kivonatkiadása, MPT, p. 34.
- Vargha A. (2005). Sokaságok összehasonlítása új módszerekkel. *Statisztikai Szemle*, 83(5), 429-448.
- Vargha A. (2015). *Matematikai Statisztika*. Budapest: Pólya Kiadó.
- Vargha A. (2016). Szignifikanciaszintek – negyven éve hibás elemzéseket végzek és téveszmét tanítok?. *Statisztikai Szemle*, 94(4), 445-451.
- Walczyk, J. J., Mahoney, K. T., Doverspike, D., & Griffith-Ross, D. A. (2009). Cognitive Lie Detection: Response Time and Consistency of Answers as Cues to Deception. *Journal of Business and Psychology*, 24(1), 33-49.
- www.animagroup.hu – letöltve 2017. 01. 17.
- www.nemesysco.com – letöltve 2017. 01. 17.

Tematikus szám hivatkozott cikkei

- Kis, Gy.; Takács, Sz.; Liberman, A.; Benczúr, L. (2017): A megtévesztés tipológiája – összefoglaló tanulmány. *Psychologia Hungarica Caroliensis*, 4(2), 2016. pp. 7–26.
- Czabán, Cs., Alpek A., Bártfai A., Kertesz A., Iványuk Á., Benczúr L., pp. Takács, Sz., Kis, Gy. (2017a): A vallomások igazságtartamát meghamisítók típusai és vizsgálatuk kihallgatási helyzetben – 1. rész, a meghamisító. *Psychologia Hungarica Caroliensis*, 4(2), 2016. pp. 27–35.
- Czabán, Cs., Alpek A., Bártfai A., Kertesz A., Iványuk Á., Benczúr L., Takács, Sz., Kis, Gy. (2017b): A vallomások igazságtartamát meghamisítók típusai és vizsgálatuk kihallgatási helyzetben – 2. rész, az elhallgató. *Psychologia Hungarica Caroliensis*, 4(2), 2016. pp. 36–43.
- Madzin, A., Alpek A., Bártfai A., Kertesz A., Iványuk Á., Benczúr L., Takács, Sz., Kis, Gy. (2017a): A vallomások igazságtartamát meghamisítók típusai és vizsgálatuk kihallgatási helyzetben – 3. rész, az eltitkoló. *Psychologia Hungarica Caroliensis*, 4(2), 2016. pp. 44–51.
- Madzin, A., Alpek A., Bártfai A., Kertesz A., Iványuk Á., Benczúr L., Takács, Sz.; Kis, Gy. (2017b): A vallomások igazságtartamát meghamisítók típusai és vizsgálatuk kihallgatási helyzetben – 4. rész, a patetikus. *Psychologia Hungarica Caroliensis*, 4(2), 2016. pp. 52–59.