

HOM-E/O-MINING és online tudástranszfer

Dr. Pitlik László, egyetemi adjunktus, tanszékvezető

Pásztor Marta, egyetemi tanársegéd

Popovics Attila, egyetemi tanársegéd

Bunkóczi László, tanszéki mérnök

SZIE, Gazdaság- és Társadalomtudományi Kar, Gazdasági Informatika Tanszék, Gödöllő

Bevezetés

A tanszék több mint 10 éve a mesterséges intelligenciákra alapuló, saját fejlesztésű adatbányászati (data mining) módszertan területfejlesztést, politikai tanácsadást, szaktanácsadást érintő interdiszciplináris kérdéseivel foglalkozik. Ennek eredményeként kikristályosodni látszik egy jól oktatható (keresett), ill. a gazdasági folyamatokba szervesen illeszthető (költség-hatékony) módszer, mely bepillantást enged a black-box-nak tűnő automatikus összefüggés-keresési, tanulási eljárásokba. Az eljárás (WAM) előnye, hogy táblázatkezelőben, egyszerű függvényekkel és képletekkel dolgozik, bárhol rendelkezésre áll (költségkímélő), valamint lehetőséget ad az emberi tanulás és intuíció bevonására, ill. biztosítja a gép-ember egyensúlyára alapuló szimbiózis lehetőségét. Az elnevezés (HOM-E/O-MINING)→ szójáték, mely többféleképpen olvasható (azaz felbontható a Home-Mining, Homo-Mining, Homeo-Mining), utalva az Ember (homo), Esély-Egyenlőség (homeo), Otthonosság (home) és a Kifejtés/Bányászat (mining) kifejezésekre.

Az előadásban többek között az alábbi projektek és szempontok összefoglaló eredményeit szeretnénk röviden megjeleníteni:

- OTKA: Az EUROSTAT konzisztens agráradatbázisának (SPEL) induktív szakértői rendszerrel (WAM) történő elemzése (<http://miau.gau.hu/miau/19/otkastudy.doc>)
- OMFB-IKTA: Adatbányászati elemzések beillesztése a SAPARD program informatikai támogatásába (<http://miau.gau.hu/iktabu/iktabu2.html>)
- EU: Az IDARA projekt (Integrated Development of Agricultural and Rural Areas) első tapasztalatai (<http://miau.gau.hu/miau/20/aprilis.html>)
- FVM: A MIVIR (Magyarországi Integrált Vidékfejlesztési Információs rendszer) elemző alrendszerének első tapasztalatai, ill. on-line tudástranszfer elemek az oktatásban és szaktanácsadásban (<http://miau.gau.hu/news/index.html>)
- FKFP: Döntéshozók informatikai magatartás mintái (<http://miau.gau.hu/miau/16/fkfp.doc>)
- STOCKNET: online tőzsdeelemzés az idősorokban rejlő hasonlóságok (benchmarking, CBR) automatizált feltárása által (<http://miau.gau.hu/miau/08/stockbase.doc>)

A dolgozat közvetlen célja a fent említett projektek, a Magyar Informatikusok II. Világtalálkozója, ill. a NJSZT "Együtt az információs társadalomban" című felhívása kapcsán egy fajta, kényszerűen némileg filozofikus áttekintést adni a következő évtizedek egy lehetséges, s nagy hatású (minőségi), ma még csak lappangó jelenségéről. , hanem mint Az eddigi apró történések világosan jelezték számunkra, hogy az ezredforduló sokak fejében jó táptalajt teremt a stratégiai tervezés - sokszor elhanyagolt - fontosságának újbóli kiemelésére. A projekt témák szabad választása (nem lévén iránymutató központi gondolat - mely talán nem is kell), oktatási tapasztalataink (<http://miau.gau.hu/miau/06/tankonyv.rtf>) és elméletibb szempontú összefoglaló munkáink (Networkshop`1999 <http://miau.gau.hu/miau/06/networkfull.rtf>, Informatika a felsőoktatásban 1999, <http://miau.gau.hu/miau/09/if99veg1.rtf>, Networkshop`2000, <http://miau.gau.hu/oszr/index.html>) is mind-mind arra ösztönöznek, hogy végre megszülessen egy röviden kifejezhető, de hosszú távra orientáló erejű stratégiai gondolat, a **HOM-E/O-MINING** gondolata.

A szerzők koruk (< 35 év) és életpályájuk (egyetemi oktatás - kutatás - szaktanácsadás) alapján természetesen csak a valóság egy szegmensét megismerve (parcialitás), de mindenkor az általánosítás igényével (tanulás, koncepcióalkotás, verifikálás, tesztelés, azaz mesterséges intelligencia alapon), s a strukturált gondolkodás "áldásos átkával" közelítenek a napi feladatokhoz. Az elmúlt évek tapasztalatai alapján nyugodt szívvel felvállalható a HOM-E/O-MINING-nak keretet adó, s az "**online tudástranszfer**" fogalmával érzékeltetni kívánt trend egyre erőteljesebb igénylése és propagálása.

HOM- E/O- MINING 2020 - egy jövőkép körvonalai

A misztikusan felvezetett két központi gondolat magyarázatának indításához érdemes az NJSZT pályázat két fontos, s közérthető kérdését alapul venni, egyrészt azt, hol "látszik reális lehetősége annak, hogy az információ-áramlás sebességének rohamos növekedése jelentős változásokat idéz elő?", másrészt "melyek azok a területek, ahol - a szerzők véleménye szerint- szükség van változásra, és melyek lennének ezek a változások?"

Köztudott, hogy információk adatokból keletkeznek. Az adatokról sok minden elmondható, de a (az információs) társadalom szempontjából az adat meg nem kerülhető jellemzője a hozzáférési jogosultságok szabályozása. A jogosultságok szabályozottsága lyukas, s annál lyukasabb, minél gyorsabbak a valós élet változásai. S bizonyos változások szinte önmagukat katalizálják....

Az Internet technológiai és a közhasznúság jogi keretrendszerében a rendelkezésre bocsátott, s egyben hozzá is férhető adatok mennyisége lendületesen bővül, rést ütve ezáltal az információ monopolizálás eddig oly jól bevált gyakorlatán. A mennyiségi változások (kínált adatmennyiség, egységnyi idő alatt megszerezhető mennyiség, azaz átviteli sebesség) - úgy tűnik - feltartóztathatatlan ütemben (a szabadság-, közlésvágy, kíváncsiság kifogyhatatlan üzemanyagától hajtva) haladnak a maguk útján. Kérdéses azonban milyen minőségi változások állnak elő ezen mennyiségi változások kapcsán? Hiszen tetszőlegesen sok adathoz való tetszőlegesen gyors hozzáférés önmagában nagyobb káoszt (döntésképtelenséget) okoz a tapasztalatok szerint, mint egy korlátozottabb hozzáférési szint esetében. Érdemes talán utána gondolni annak, hogy az emberiség evolúciósan nem feltétlenül a nagy inger- és jelmennyiség feldolgozására szelektálódott, legalább is nem a tudatos gondolkodás szintjén. Kísérleti körülmények között könnyedén visszaigazolható, hogy több ún. lényegtelen adat (zaj) alapján a döntések mérhető helyessége csökken, ill. egy adott döntési helyzetben a döntések helyessége nem növelhető tetszőlegesen a feldolgozásra felkínált adatok számának növelésével.

Meg kell jegyezni, hogy az információ (kulcsfontosságú adat) monopolizálása, mint eszköz továbbra sem veszít vonzerejéből, de hatásossága talán csökkenthető, hiszen az előzőekben zajnak nevezett adattöbblet, s az emberi kreativitás eredőjeként egyre több olyan módszer áll elő (mesterséges intelligenciák), melyek lehetővé teszik azt, hogy bizonyos számú eset/példa (szituáció + következménye) alapján újszerű, de még ismeretlen következményű szituációk várható utóhatásai levezethetők legyenek (vö. neurális hálók, genetikai algoritmusok, stb.). Ezáltal az ún. hulladékadatokból (gyakran monopolizált) kulcsadatok válnak (közelítőleg) levezethetővé. Így az eddig adattemetőknak hitt adatbázisok (pl. meteorológiai megfigyelések, áradatbázisok, tőzsde) értéke egyre nő, hiszen a technológia, a tudat fejlődése és a lyukas adatvagyon-gazdálkodási szabályok egyre nagyobb teret engednek a kreativitásnak.

A köz-hasznúság és köz-szolgáltatóság által hajtott, ill. az egyre olcsóbb automatizált mérési eljárások által még tovább gerjesztett, valamint az egyedi adatok védelméhez való (kérdéses értékű) jog által fékezett, de mégis eredőjében pozitív adattömeg-bővülési tendenciák rövid és középtávon megállíthatatlanok. A kommunikációs hálózatok (Internet) fejlődésével a (az adat-) kínálat és kereslet egymásra találását már csak az átviteli sebesség és a keresési stratégiák "finomsága" befolyásolja. A mindenkori körülmények között realizálható adattömeg feldolgozása, mint kihívás gazdasági értelemben is életeret teremt, s egyre több feldolgozó algoritmus (data mining) alakul ki, melyek minden további módosítás nélkül online jelleggel (hálózaton keresztül) is alkalmazhatók.

Az előző bekezdés több fontos részproblémára is rávilágított, mint pl. az intelligens keresés (információs brókerek) hiánya, az egyedi adatok védelmének jogossága (joga van-e az egyednek védeni saját "titkait", ha ezzel kárt okoz a közösségnek, ill. tud-e egyáltalán a biztonság érzetét keltő "titkok" nélkül létezni egy emberi egyed?), a közhasznú adatokhoz való ingyenes hozzáférés mesterséges korlátozása (elvileg lehet, de gyakorlatilag nehézkes) e helyen részletesen nem kerülnek tárgyalásra. Megjegyzendő azonban, hogy a tényszerű, már bekövetkezett valóság-állapotokat leíró adatokhoz való hozzájárulás (információs brókerek), valóban tömegeket érintő kérdése, hasonlóan a könyvtárak által nyújtott szolgáltatások fejlődéséhez, nagy valószínűséggel egyre hatékonyabb adatfeltárást tesz majd lehetővé hosszú távon, folyamatosan küszködve az e területen érvényes antagonizmusokkal (pl. kézi, gépi indexálás előnyei, hátrányai, gazdaságossága). Az egyedi adatok védelme és a közhasznúság alapvetően politikai kérdésként kerül hosszabb távon is kezelésre, noha érdemes lenne egyedi és csoportos pszichológiai kísérleteket folytatni a tekintetben, vajon milyen mértékű (adat-)kiszolgáltatottság/védelem jelenthet társadalmi stabilitást.

Az áttörés gyanúja miatt azonban részletesebben érdemes foglalkozni azzal, vajon mennyire magától értetődő az egyre bővülő online (szerver-oldali) adatbázisok és az online (szerver-oldali) feldolgozó algoritmusok tömeges felhasználása?

Ehhez először azt szükséges lehatárolni, miért tűnik hasznosabbnak a tények begyűjtése (információs brókerek, kedvenc/navigációs oldalak, bookmark-ok, stb.) helyett az adatelemzésre tenni a hangsúlyt?

A tényeknek megvan sajnos az a rossz tulajdonsága, hogy már megtörtént eseményekre (állapotváltozásokra) utalnak, s így a jövőben, ill. a jövőre vonatkozólag meghozandó döntések szempontjából önmagukban értéktelenek, ill. öntudatlan (vagy tudatosan vállalt) agyi továbbfeldolgozáson keresztül (intuitív data mining) hasznosulnak.

Itt érdemes leszögezni, hogy az intuitív, asszociatív emberi képességek nagyon hasznosak, s elsődlegesen fontosak a mindennapos adaptáció során, de nem jó, ha csak erre hagyatkozunk. A kultúra, a technikai fejlődés mind-mind egy fajta "tudáskonzervet" jelent, s ilyen tudáskonzervek (operacionalizált szabályozási lehetőségek) a döntés-előkészítésben is kialakíthatók (matematika, statisztika, operációkutatás, data mining). A döntést tehát valóban illik előkészíteni, s általában kevésbé sikeres, ha csak meghozzuk....

A döntés-előkészítés alapvetően feltételezi, hogy ismertek a döntéssel elérni kívánt célok. Ha ez mégsem így lenne, mint az az esetek zömében tapasztalható, akkor is helye van egy szimulációs jellegű megközelítésnek. A szimuláció nem más, mint (maga a tudomány?, azaz egy fajta) válaszkérés a "mi lenne, ha" kérdésekre. A jó Emberi döntés tehát sokkal kevésbé a célorientáltsággal, mint sem az előrevetített állapotváltozások közötti választással jellemezhető.

Az online tudástranszfer ezek után nem más, mint a hálózaton keresztül végzett ismeretátadás, mely lehet alapvetően adatszerű, alapvetően eljárás-jellegű és ezek tetszőleges kombinációja (pl. szakértői rendszerek). A döntés-előkészítés, mint az információs többletérték-felhalmozás egyik formája tudást igényel, mely alapvetően a múltra vonatkozó adatokból és az ezeket egyre "refináltabban" feldolgozni képes (context free) eljárásokból áll. A még hasznosnak ítélt adatlekérdezést, ill. elemző futtatásokat a mindenkori technológiai szint és a felhasználói motiváció határozza meg.

A felhasználók akkor tekinthetők motiváltnak az adatsorok elemzésére alapozó döntés-előkészítésben, ha belátják azt, hogy saját intuíciónk nem elegendő, túlterhelhető, bizonytalan stabilitású. Ilyen felismerésekre az emberek zöme nem jut (nem akar jutni), hiszen ez pszichológiai instabilitáshoz vezethet. Másrészt gazdasági értelemben is meglehetősen "drága mulatság" az ún. szakértői munka eredményeként előálló jövőképek bevalásának szisztematikus monitorozása, azaz a szakértői munka minőségbiztosítása. (Például annak összevetése, hogy melyik parlamenti párt szakértői, ill. gazdaságelemző intézet munkatársai produkálnak jobb találati arányokat legalább a trendek irányát tekintve.) A minőségbiztosítás másik problémája a szubjektív emberi tulajdonságok mellett az, hogy matematikai értelemben sem lehet eldönteni két szakértőről, vagy két modellről, melyik *a* jobb. Azt, hogy melyik *volt* a jobb, azt még csak-csak körül lehet írni, de azt, hogy tetszőleges helyességi mutatók idősoros alakulása után melyiknek szabad hinni a következő döntés előkészítésekor, azt már egyszerűen lehetetlen megválaszolni (vö. céltalanság tétele). Nem beszélve arról, hogy a legtöbb előrejelzés (pl. politikai scénáriók) esetében nem is lehet vizsgálni kinek is lett volna igaza, mert csak egy valóságfolyam vizsgálható, az alternatív döntések következményei nem.

Mindezek alapján a felhasználók elemzési készsége több szempontból is negatív hatások alatt áll (pl. belső ellenérdekeltség, gazdasági kényszer, filozófiai antagonizmusok). A kisebb ellenállás irányába való elmozdulás szabálya itt is érvényes tehát. Az emberek zöme nem akar, nem tud strukturáltan (azaz mások eredményeinek a saját gondolatmenetbe való integrálásával) gondolkodni, mert nem kényszeríti rá őket a neveléssel kiváltott "hit".

Ugyanis erőteljesen hinni kell abban, hogy az ember képes a kombinatorikai alapokon nyugvó (de emberek által megalkotott) elemzési technikákkal való együttműködésre, azaz egy fajta szimbiózisra, ahol is az emberi döntéshozó képessége nem a nagyvalószínűséggel helyes döntés megérzésében csúcsosodik ki, hanem a helyes alapadatok, feldolgozási módszerek kiválasztásában (lévén, hogy ezen elemek online kínálata egyre nő), a kapott modellek szelektálásában, vagyis a helytelen és a helyes becslések / előrejelzések nagy valószínűséggel történő klasszifikálásában ölt testet. Ilyen típusú emberből - egyszerű becsléssel közelítve a jelenséghez - legalább annyi van, mint jó intuitív döntéshozóból.

Ezen embercsoport felfedezése (vö. science fiction irodalom pszi típusú egyedei) manapság a munkakörök betöltésekor megfigyelhető véletlen sodródás elve szerint történik. Abban az esetben azonban, ha a képzés során (vö. NJSZT - mit is kellene tenni?) már szisztematikusan edzené a társadalom a fiatalokat arra, hogy nem csak "hasraütés-szerűen" lehet dönteni, hanem a meglévő adat- és eljárásvagyon megérzésre alapuló felhasználásával is, akkor ez a fajta - a globális egyensúlyvesztések által egyre fenyegetettebb társadalom számára - egyre hasznosabb képességcsoport viszonylag hatékonyabban lenne bevetendő a döntés-előkészítés folyamatába. A képzés egyben erőt (hitet) is ad az általa prioritásokhoz (élettérhez, önbizalomhoz) juttatott csoportoknak. A döntések meghozatala továbbra is megmaradna a politikai szinten már valamilyen formában kikristályosodott mechanizmusok keretei között, azonban a politika nem véletlenszerűen kiválasztott jövőképek köré építené szakértőbőroait (lévén nekik lényegében mindegy), hanem emberi kreativitásra, ill. gépi automatizmusokra támaszkodó minőségbiztosítással (verseny, szelekció, szakértői hálózatok) támogatott "iparszerű" jövőképgyártásra.

Ezen folyamatok ma talán sokaknak a tudományos fantasztikum kategóriájába tartoznak, de az elmúlt 5 év képzési eredményei világosan jelzik, hogy az a diplomás réteg, aki fogékony erre a ma még sokszor csak játéknak tűnő ember+gép szimbiózisra, az valóban (mérhetően, szignifikánsan) helyesebb jövőképeket képes produkálni, mint a fennmaradó (informatikai affinitásában korlátozott) csoport.

Az online tudástranszfer fogalmával fémjelezhető folyamat célja és eredménye tehát a helyesebb döntéshozatal. Ehhez szükséges, hogy az emberi egyedek között a strukturált gondolkodási képességben (spec. IQ, egyensúlyteremtési képesség gép és ember között) meglévő különbségeket a képzés során felismerjük, s társadalmi szinten tudatosan használjuk. Hiszen ez az a réteg, amely képes az egyre növekvő adat- és eljárásvagyonnal (technológiai lehetőségekkel) gazdálkodni visszaellenőrzött minőségű szimulációk, előrejelzések előállítására érdekében. Az elkövetkező 20 évben (mint a történelemben eddig bármikor, de lassacskán végre tudatosan felismerten) az ismerttechnológus, a tudásmérnök lesz az a kulcsemberek, akin egyes szervezetek, embercsoportok érdekérvényesítő képessége áll vagy bukik. Aki tudja, mit hoz a jövő, az arra készül, s kevesebbet cselekszik feleslegesen. Ha ehhez még elfogadjuk, hogy az új sikertörténetek mutatószáma az egyensúlyvesztések (ökológiai, társadalmi) számának minimalizálása lesz, akkor különösen egyértelmű, hogy hová is kell helyezni a képzés prioritását.

Így jutunk el a HOM-E/O-MINING gondolatához, mely az absztrakt gondolatokat lekényszeríti a valóban végrehajtható cselekvési programok szintjére. A HOM-E/O-MINING ugyanis egy olyan egyszerű képzési modult jelent, mely keretében minden erre fogékony egyed képes nagy adatmennyiségekkel, a technológia adta lehetőségeken belül (azaz jelenleg zömmel lokálisan, Excel környezetben) már hatékony, de még követhető eljárásokkal zsonglörkődni, így szerezve a hitét megalapozó tapasztalatokat ahhoz, hogy téthelyzetben ismerje vállalni annak kimondását, hogy a „Q” modell jobb, mint a „Z”, azaz egy adott jelenség inkább nő (s mennyivel), mint sem csökken a következő X periódusban, ha ezt és ezt tesszük időközben....

Ez a képzési módszertan a saját oktatási és kutatási tapasztalatok alapján kidolgozott WAM- (weight and activity modell) eljárásra alapoz, mely egyszerre induktív szakértői rendszer, egyszerű neurális háló, ill. a monte carlo eljárásoktól a genetikai algoritmusokig ívelő lehetőségek kísérleti terepe (a valóságban néhány Excel parancs sora).

Manapság a nagy fejlesztő kapacitással rendelkező cégek (pl. ORACLE, IBM, SAS) jelennek meg a piacon ismét csak nagy cégeknek szóló data mining szoftverekkel, melyek dicséretes képességekkel rendelkeznek, de titkolózó és black box jellegüknél fogva nem adnak lehetőséget a felhasználó tanulásának támogatására. Így magától értetődően ezen és hasonló termékek bevezetése csak akkor lehet nagy számban sikeres, ha a fogadókészséget előbb képzési projektekkel meg lehet teremteni. A technika/program nagy számban történő eladása vezethet az árak csökkenéséhez, amely által az társadalmilag egyre elfogadottabbá (alkalmazottabbá) válik (pl. személyi számítógép, mobiltelefon, Internet stb.). A következő 20 évben tehát a ma data mining-nek nevezett projektek mindennapos társadalmi valósággá válásának történetét figyelhetjük meg - vagy zsákutcs fejlődésnek minősül a jelenlegi trend..... (hiszen a manapság kibocsátott közgazdasági végzettségű hallgatók - a leendő vezetők - jó, ha 10 %-a konfrontálódik 1000-nél több adat rutinszerű és sokoldalú feldolgozásának problematikájával, mint ahogy ezt a diplomamunkák témaválasztása is mutatja).....