# Gyanúgenerálási rendszerterv banki tranzakciós adatbázisok kapcsán

Pitlik László, 2011. december

## Előzmények

A gyanúgenerálás (<http://www.google.hu/search?q=itbszm>) a hasonlóságelemzés egyik speciális ága (<http://miau.gau.hu/miau2009/index.php3?x=e0&string=susp>). Ennek lényege a nem deklaratív gyanú fogalmának bevezetése (<http://miau.gau.hu/miau/153/kdrik_2011.doc>) többrétegű, anti-diszkriminációs (<http://miau.gau.hu/myx-free/index.php3?x=e091>) számításokra alapozva. A nem deklartív gyanú esetén nem kell pontos riasztási küszöbértékeket megadni. A gyanúmomentumok a korábbi élethelyzeteket leíró attribútumokra megfogalmazott irányultságok (pl. minél kevesebb a tranzakciók száma, annál kevesebb az esély a visszaélésre – hiszen nulla tranzakció esetén nem létezhet visszaélés) mentén, egymással versengő erőtereket felépítve keletkeznek optimalizációs eljárások felhasználásával. A gyanúgenerálás eredménye egy relatív skálán kifejezett gyanúmérték, mely hozzárendelésre kerül minden egyes vizsgált időintervallumhoz. Az önellenőrző számítások alapján a gyanúmértékek lehetnek hitelesek és hiteltelenek. Az emberi döntéshozó számára a hitelesen magas gyanúmértékek és a hiteltelen esetek jelentenek további feladatot. A gyanúgenerálás, s az erre épülő riasztás és az adekvát operatív akciók kivitelezése automatizálható.

## Az adatvagyonról

A banki tranzakciók kapcsán a naplóállományokból számos karakterisztika kiolvasható:

* honnan (pénzintézeti azonosító, országkód)
* hová (pénzintézeti azonosító, országkód)
* mikor
* milyen csatornán
* milyen összeg
* milyen pénznemben
* stb. vándorolt.

Mivel egyedi tranzakciókra érvényes gyanúgenerálás (hasonlóan a pontszerű becsléshez) nem lehetséges, így aggregált időszakok (pl. órák) adatait, mint objektumokat (vö. mátrixsorok) kell feltölteni olyan attribútumokkal, melyekre vonatkozóan gyanúsítási irányok fogalmazhatók meg: pl.

* minél kevesebb a tranzakciók száma az adott időszakban, annál kisebb az esélye annak, hogy ebben a halmazba gyanús elem keveredett
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy darabszám (count)
  + technikai alábontás: pénznemenként és/vagy tranzakciós csatornánként és/vagy deklarált (jogszabályilag érzékeny) értékintervallumonként
* minél kisebb a tranzakciók volumene annál kisebb az esély, hogy értelme legyen gyanús tranzakciót katalizálni
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy összeg (sum)
  + technikai alábontás: pénznemenként és/vagy tranzakciós csatornánként
* minél kevesebb a pénzintézetek száma, annál kisebb az esélye, hogy valamelyikben nem tervezett tranzakció történik
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy darabszám (count)
  + technikai alábontás: pénznemenként és/vagy tranzakciós csatornánként és/vagy deklarált (jogszabályilag érzékeny) értékintervallumonként
* minél kevesebb az érintett pénznemek száma, annál kisebb az esélye bonyolult álcázási manővereknek
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy darabszám (count)
  + technikai alábontás: tranzakciós csatornánként és/vagy deklarált (jogszabályilag érzékeny) értékintervallumonként
* minél homogénebb a tranzakciók összegszerűsége, azaz minél kisebb a szórása, annál kevesebb az esélye speciális tartalmú manipulációnak
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy vektor, mely szórására van szükség
  + technikai alábontás: pénznemenként és/vagy tranzakciós csatornánként
* minél kevesebb az érintett tranzakciós csatornák száma, annál kisebb az esélye speciális manipulációk előfordulásának
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy darabszám (count)
  + technikai alábontás: pénznemenként és/vagy tranzakciós csatornánként és/vagy deklarált (jogszabályilag érzékeny) értékintervallumonként
* minél gyakoribbak az egyes honnan-hová párosítások, annál kisebb az esélye speciális manipulációnak
  + kivitelezés: honnan-hová kombinatorikai terek megalkotása
    - országok között
    - pénzintézetek között
  + lekérdezés-sorozat (select distinct),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége, ill. honnan-hová alakzatok egyedi vagy kombinált paraméterei,
  + ahol az idő tetszőlegesen görgethető, ill.
  + a lekérdezés eredményei kombinációnkénti darabszámok (count)
  + feldolgozás: a leggyakoribb kombinációk hibapontértéke nulla, a legritkábbé pl. 100, a többié előfordulással fordított arányban
* minél nagyobb a homogén tranzakciók közötti időeltérés (szórás), annál kisebb az esélye annak, hogy a sok hasonló adat között gyanús bújjon meg
  + homogén tranzakció: azonos pénznem, honnan-hová koordináták, értéknagyság-intervallum,
  + kivitelezés: lekérdezés (select),
  + ahol a szűrési feltétel (where) egy időintervallum eleje/vége,
  + mely tetszőlegesen görgethető, ill.
  + a lekérdezés eredménye egy vektor időkoordinátákkal, melyek szórása kell
* …

A fentiek mintájára tetszőleges további mutatószámok alakítható ki. A fentebb felsorolt egyszerűbb/komplexebb mutatószámok generálására a hardverkapacitások függvényében óránként minimum egyszer, de számítókapacitás függvényében X perces eltolással van szükség. Az adott időintervallumhoz tartozó mutatószámok egy objektum – attribútum - mátrix (OAM) alakját veszik fel automatikusan: ahol a sorok az egyes órák (folyamatosan bővülő sorszámmal), az oszlopok a számított karakterisztikák, az értékek a fentebb jelzett darabszámok, szórások, összegek, stb.

A primer OAM és a (minél/annál-jellegű) irányok alapján a teljes OAM oszloponként sorszámokká (rank) konvertálandó (vö. standardizálás). Ezen sorszámmátrix minden további lépés (black box) inputja.

Az elemzés outputja egy hitelességi státusz és egy gyanúmérték. A hitelességi státusz egymáshoz képest inverz irányú inputok elemzés utáni szimmetria sértéseit detektálja, vagyis az a hiteles gyanú, mely megfelel a tagadás tagadása elvnek. A gyanú mértéke az egyszerre vizsgált objektumok 100 gyanúpont körül megoszlását jelenti: a kisebb értékek a kevésbé, a nagyobb értékek az erőteljesebben gyanús időintervallumokat jelzik.

A gyanúk időbeli dinamikája (vagyis egyre növekvő riasztási szintek egymást követő időintervallumonként) stabilizálhatja a riasztás szükségszerűségét.

## Specialitások

Külön (de a fentiekhez hasonló) elemzési logika alkotható egyes aktőrök magatartásának (speciális tranzakciócsoportok, mint objektumok) figyelésére, melyben a relatív gyanú nehezen „hack”-elhető logikájának követési kísérletei kerülhetnek relatív gyanúként megfogalmazásra…

Az inverz-ellenőrzésekkel hitelesített anti-diszkriminációs számítások mellett minden egyes mutatószám kapcsán norma-értékek kalkulálhatók, melytől a tények eltérései szintén riasztási alapot jelenthetnek (pl. sokkal kevesebb tranzakciónak illett volna lennie, vagy inkább nagyobb arányú forint tranzakciók lettek volna logikusak, stb.).

A gyanúképződés dinamikája, mint addicionális feladat előrejelzésekhez is alkalmas szekunder adatbázis termel ki, mely alapján a gyanú jövőbeli várható alakulása is vélelmezhető…

## Összefoglalás

A nem deklaratív tanulásnak nincs érdemi alternatívája:

* a WizWhy típusú elemzés nem képes relatív gyanú generálására, csak a kombinatorikai tér ismert szegmenseihez tartozó gyakoriságok értelmezésére, vagyis nem képes a gyanú genetikai potenciáljának meghatározására, szimulációs jellegű értelmezésére (<http://miau.gau.hu/miau/150/la150.docx>), a szabályerők esetén nincs érdemi mód önellenőrzésre,
* a matematikai statisztikai megközelítések egyike sem önellenőrző, s ezek nem ismerik az optimalizációs tanulási folyamatokat,
* a neurális hálózatok ismert gyanúk felismerésére taníthatók be, s nem képesek a plátói (gyanú) ideál kezelésére.