Közgazdász Fórum
Forum on Economics and Business
17 (1–2), 167–179.
2014/1–2

Egy fluidumáramlási probléma megoldása „bevásárlókosár”-elmélettel és a módszer algoritmusainak vizsgálata

A fluid-flow problem solved by „shopping basket” theory and the analysis of the algorithm of the method

NAM SON HUA – MIKLÓS GUBÁN

Discovering common shopping baskets is a very attractive topic of data mining theory. In these researches the purchases (transactions) of the buyers are considered as a set of goods. However researchers are interested in the batch of goods chosen by the customers. In our study the purchases of customers (transactions) are not considered to be a set of batch of goods, but the set of volume of goods, which allows us to use an algebraic approach contributing to the formal description of shopping basket model. Our approach differs from other approaches in previous studies, that we have considered the volume of transactions in batches.The advantage of this approach is the better illumination of inter-transaction subsystems, and the role of network structure above them. In this general model we re-examine some of the known issues by net-theory assets. We present an explicit representation for frequent customers baskets and for association rules. As a direct result of this issue algorithms are defined to explore frequent customers baskets and association rules.

Keywords:shopping basket, frequentproducts, association rule, network, service.

JEL-codes: C18, C51, O31.

Kiadó: Romániai Magyar Közgazdász Társaság és a Babeș–Bolyai Tudományegyetem Közgazdaság- és Gazdálkodástudományi Karának Magyar Intézete
ISSN: 1582-1986 www.econ.ubbcluj.ro/kozgazdaszforum

Egy fluidumáramlási probléma megoldása „bevásárlókosár”-elmélettel és a módszer algoritmusainak vizsgálata[1]

HUA NAM SON[2] – GUBÁN MIKLÓS[3]

Az adatbányászat elméletében a gyakori vásárlói kosarak felfedezése vonzó témát jelent a kutatók számára. Ezekben a vizsgálatokban a vevő vásárlásait (tranzakciókat) áruk halmazának tekintették. Valójában azonban a vevők által választott árutételek neve érdekli a kutatókat. Jelen kutatásunkban a vevő vásárlásait (tranzakciókat) nem az árutételek halmazának, hanem az árumennyiség halmazának tekintve egy algebrai megközelítést kapunk, amely alkalmas a vásárlói kosarak modell formális leírására. Megközelítésünk abban tér el a korábbi vizsgálatokban használtaktól, hogy számításba vettük a tranzakciók tételeinek mennyiségét. Előnye az, hogy jobban láthatjuk a tranzakciók közötti természetes részberendezésnek, valamint a fölöttük levő háló struktúrájának szerepét. Ebben az általánosabb modellben a hálóelméleti eszközökkel újravizsgálunk néhány ismert kérdést. Bemutatunk egy explicit reprezentációt a gyakori vevői kosarak és az asszociációs szabályok számára. Ennek közvetlen következményeként algoritmusokat írtunk le a gyakori vevői kosarak, az alap gyakori vevői kosarak, valamint az asszociációs szabályok felfedezésére.

Kulcsszavak: vásárlói kosár, gyakori termékek, asszociációs szabály, háló, szolgáltatás.

JEL-kódok: C18, C51, O31.

Bevezetés

A szolgáltatási folyamatok kritikus pontjainak feltárása során azt tapasztaltuk, hogy nagyon nagy mennyiségű adat keletkezik. Emellett a felmérések elvégzése után is újabb adathalmazt kapunk, és a folyamatban szereplők száma, valamint a termékek változatainak a száma is jelentős. Mindezek indokolják, hogy az adatok elemzéséhez, vizsgálatá-

[1] A tanulmány alapjául szolgáló kutatás az EMMI-26130-2/2013/TUDPOL támogatásból valósult meg a Budapesti Gazdasági Főiskolán.

[2] PhD, főiskolai docens, Budapesti Gazdasági Főiskola, e-mail: hua.nam.son@pszfb.bgf.hu.

[3] PhD, főiskolai tanár, Gábor Dénes Főiskola, e-mail: guban@gdf.hu.

hoz az adatbányászat módszeréhez forduljunk. Az elemzéseket – a kitűzött céloknak megfelelően – a piaci szereplők szegmentálására, a szervezeti függőségek meghatározására, a gyakori termék, ügyfelek, folyamatok megkeresésére, valamint a gyakori termék, ügyfelek, folyamatok közötti asszociációs szabályok feltárására szeretnénk elvégezni.

A fentieket általánosítva a vizsgálat a fluidumok áramlásának gyakoriságára vonatkozik. Megvizsgálva az adatbányászat eszközrendszerét, kiderült, hogy nem áll rendelkezésre olyan mérték nélküli osztályozó módszer, amelyik teljesen megfelel a kutatásban kitűzött cél szerinti elemzésnek. Ez indukálta egy új módszer kidolgozását, illetve a kutatócsoport kutatói által már korábban kidolgozott adatbányászat területén belüli osztályozási módszerek felhasználását.

A „bevásárlókosár" modell hasznos adatbányászati módszereket ajánl fel számos területen felmerülő feladatok megoldására. Többek között a gyakori elemek, illetve az elemek közötti asszociációs kapcsolatok feltárása fontos feladat, amelynek a megoldása több alkalmazási területen vonzó kutatási témát jelent sok szakértő számára. Jelen cikk a „bevásárlókosár" modellben a szerzők által elért eredményeket mutatja be, amelyek a fluidumok áramlásainak vizsgálatára, illetve a szolgáltatások és folyamatok irányítására alkalmasak. Az alábbiakban három algoritmust ismertetünk a fluidumok gyakori áramlásai, illetve a gyakori szolgáltatások meghatározására. Megmutatjuk az algoritmusok bonyolultságelemzését, az algoritmusok működésének a tesztelése megerősíti az elért eredményeket. A kapott eredmények közvetlenül felhasználhatók a szolgáltatási folyamatokat igénybe vevők szegmentálására.

A matematikai modell

Első lépésként megadjuk a probléma matematikai modelljét. A modell alapját az úgynevezett „bevásárlókosár" elmélet képezi.

Kiindulási halmazként tekintsük a termékeknek – melyek szolgáltatások is lehetnek, de a továbbiakban termékként hivatkozunk rá – egy véges halmazát. A halmaz elemszáma legyen n, azaz n darab terméket fogunk vizsgálni a továbbiakban.

Jelölje $P = \{p_1, \ldots, p_i, \ldots, p_n\}$ a termékek, szolgáltatások, áruk egy véges halmazát.

*1. Definíció. Vásárlói kosár*nak (röviden MB-vel jelöljük) nevezzük az $\alpha = (\alpha[1], \ldots, \alpha[i], \ldots, \alpha[n])$ sorozatot, ahol $\alpha[i]$ természetes szám, amely a p_i áru mennyiségét jelöli az α kosárban. A bevásárlókosarak (más néven *tranzakciók*) összességét Ω-val jelöljük.

2. Definíció: támogatottság. Legyen $A \subseteq \Omega$ a vásárlói kosarak egy véges halmaza és legyen $\alpha = (\alpha[1], \ldots, \alpha[i], \ldots, \alpha[n])$, $\beta = (\beta[1], \ldots, \beta[i], \ldots, \beta[n])$ két vásárlói kosár. Ekkor $\alpha \leq \beta$, ha $\alpha[i] \leq \beta[i]$ minden $1 \leq i \leq n$-re. Jelölje

$$\mathrm{supp}_A(\alpha) = \frac{|\{\beta \in A \mid \alpha \leq \beta\}|}{|A|} \tag{1}$$

az α kosár *A*-ban való *támogatottságát*. A támogatottság mértékszáma mutatja meg, hogy egy kiválasztott vevőkörben lezajlott tranzakciók alapján mennyire kedvelnek egy terméket vagy termékkészletet, szolgáltatást.

3. Definíció: gyakori vásárlói kosár. Legyen $A \subseteq \Omega$ vásárlói kosarak egy halmaza, $\alpha \in \Omega$ egy vásárlói kosár és $0 \leq \varepsilon \leq 1$ egy küszöbérték. Ekkor azt mondjuk, hogy α egy ε-gyakori vásárlói kosár (MB), ha

$$\mathrm{supp}_A(\alpha) \geq \varepsilon \tag{2}$$

Az ε-gyakori vásárlói kosarak összességét Φ_A^{ε}-val jelöljük.

Példa. Legyen $P = \{a, b, c\}$ és $A = \{\alpha, \beta, \gamma, \delta\}$ tranzakciók halmaza, ahol $\alpha = (2, 1, 0)$, $\beta = (1, 1, 1)$, $\gamma = (1, 0, 1)$, $\delta = (2, 2, 0)$. Példaként számítsuk ki két különböző tranzakcióra: $\sigma = (1, 1, 0)$-ra, $\eta = (1, 2, 0)$-ra a gyakoriságot! Ekkor $\mathrm{supp}_A(\sigma) = \frac{3}{4}$ és $\mathrm{supp}_A(\eta) = \frac{1}{4}$ lesz. Részletezve a számításokat σ esetére: *A* elemszáma 4 és $\sigma = (1, 1, 0) \not\leq \gamma = (1, 0, 1)$ kivételével 3 kosárra (α, β, δ) teljesül a $\leq$ reláció.

Adott $\varepsilon = \frac{1}{2}$ küszöbértékre az ε-gyakori vásárlói kosarak az alábbiak:

$\Phi_A^{\frac{1}{2}} = \{(2, 1, 0), (1, 0, 1), (1, 1, 0), (2, 0, 0), (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 0, 0)\}$.

Ha P most az elemi szolgáltatások halmaza és A az ügyfelek által igényelt szolgáltatások, akkor $\Phi_A^{\frac{1}{2}}$-ben található az összes olyan szolgáltatás, amelyet az ügyfelek több mint 50%-a igényel.

Nyilvánvaló, hogy Φ_A^ε meghatározása fontos szerepet játszik a gazdasági tevékenységekben: a logisztikában az anyagellátást csak Φ_A^ε meghatározása alapján tudjuk optimálisan megtervezni. A marketing területén a piacelemzés, valamint az ügyfélmenedzsment hatékonyabb lesz, ha az ε-gyakori termékeket ki tudjuk szúrni a termékhalomból.

A fenti részben tárgyalt módszer, amelyet részletesebben elemzett Demetrovics et al. (2011a) és Demetrovics et al. (2011b), lényegesen abban tér el a más kutatásban vizsgált és használt módszerektől, hogy az elemi szolgáltatások és elemi termékek helyett részletesebb, azért bonyolultabb, mennyiségi értékelés alapján kezeli a tranzakciókat. A mennyiségi elemzés a „kenyér – tojás" helyett a „2 kg kenyér – 10 db tojás", a „vasúti – közúti fuvarozás" helyett pedig a „200 km vasúti és 100 km közúti fuvarozás" tranzakciókat vizsgálja, és több esetben előnyösebbnek bizonyul.

4. Definíció: alsó korlát. L(A)-t *A alsó korlátjának* nevezzük, ha

$$L(A) = \{\alpha \in \Omega \mid \forall\beta \in A\text{: } \alpha \le \beta\} \tag{3}$$

5. Definíció: felső korlát. U(A)-t az *A felső korlátjának* nevezzük, ha

$$U(A) = \{\alpha \in \Omega \mid \forall\beta \in A\text{: } \beta \le \alpha\} \tag{4}$$

Azon kosarak halmazát, melyek a ≤ rendezés szerint *A* összes eleménél nagyobbak, jelölje

$$\alpha \cup \beta = \sup\{\alpha, \beta\} \tag{5}$$

1. Tétel. Adott $P = \{p_1, p_2, \dots, p_n\}$ halmazra, $A \subseteq \Omega$ egy kosarak halmazára és egy $0 \le \varepsilon \le 1$ küszöbértékre az $\alpha \in \Omega$ kosár ε-gyakori akkor és csak akkor, ha létezik $\alpha_1, \alpha_2, .., \alpha_k \in A$, amire $\alpha \in L(\{\alpha_1, \alpha_2, \dots, \alpha_k\})$, ahol $k = \lceil \varepsilon|A| \rceil$.

Megjegyzés: A $k = \lceil \varepsilon|A| \rceil$ összefüggés alatt az *A* kosárhalmazban lévő kosarak számának ε-nal vett szorzatának egész értékét értjük.

Bizonyítás. Ha létezik $\alpha_1, \alpha_2, \dots, \alpha_k \in A$, $k = \lceil \varepsilon|A| \rceil$, amire $\alpha \in L(\{\alpha_1, \alpha_2, \dots, \alpha_k\})$, akkor $\alpha \le \alpha_i$ minden $i = 1, 2, \dots, k$-ra, azaz

$$\mathrm{supp}_A(\alpha) = \frac{|\{\beta \in A \mid \alpha \le \beta\}|}{|A|} \ge \frac{k}{|A|} \ge \varepsilon. \tag{6}$$

Visszafelé, ha $\mathrm{supp}_A(\alpha) \ge \varepsilon$, akkor $|\{\beta \in A \mid \alpha \le \beta\}| \ge \varepsilon|A|$, azaz létezik $\alpha_1, \alpha_2, \dots, \alpha_k \in A$, $k = \lceil \varepsilon|A| \rceil$, ami $\alpha \in L(\{\alpha_1, \alpha_2, \dots, \alpha_k\})$.

Az előző tétel következményeként adódik:

2. Tétel (A gyakori MB explicit reprezentációja). Adott $P = \{p_1, p_2, ..., p_3\}$ halmazra, $A \subseteq \Omega$ egy kosarak halmazára és egy $0 \leq \varepsilon \leq 1$ küszöbértékre létezik $\alpha_1, \alpha_2, ..., \alpha_s \in \Omega$, ahol $s = \binom{|A|}{\lceil \varepsilon |A| \rceil}$, amire

$$\Omega_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i) \tag{7}$$

Bizonyítás. Legyenek $\{\alpha_1, \alpha_2, ..., \alpha_s\}$ az $\inf\{\beta_1, \beta_2, ..., \beta_k\}$ halmaza, ahol $k = \lceil \varepsilon |A| \rceil$ és $\beta_i \in A$. Az 1. tételből következik

$$\alpha \in \Phi_A^\varepsilon \Leftrightarrow \alpha \leq \inf(\{\beta_1, \beta_2, ..., \beta_k\}) \tag{8}$$

valamelyik $\{\beta_1, \beta_2, ..., \beta_k\} \subseteq$ A-ra, ahol $k = \lceil \varepsilon |A| \rceil$. Ez azt jelenti, hogy $\Phi_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i)$.

Belátható, hogy $\alpha_i \leq \alpha_j$ akkor és csak akkor, ha $L(\alpha_i) \subseteq L(\alpha_j)$. Adott A MB halmazára és az ε küszöbértékre egy $\alpha_1, \alpha_2, ..., \alpha_s$ MB halmaza, amire teljesül

$$\begin{aligned} &\text{i. } \Phi_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i), \\ &\text{ii. } \forall i,j : 0 \leq i, j \leq s \ \alpha_i \not\leq \alpha_j \text{ és } \alpha_j \not\leq \alpha_i. \end{aligned} \tag{9}$$

A MB alap ε-gyakori halmazának nevezzük. Könnyen látható, hogy adott A, ε-ra A MB alap ε-gyakori halmaza egyértelműen meghatározható, amelyet S_A^ε-vel jelöljük. Mivel fontos a Φ_A^ε meghatározása (A-beli ε-gyakori MB halmaza), érdekes az S_A^ε MB alap ε-gyakori halmaz meghatározása. A fenti tételekből és S_A^ε meghatározásából közvetlenül következik:

3. Tétel. Adott $P = \{p_1, p_2, ..., p_n\}$ halmazra, $0 \leq \varepsilon \leq 1$ küszöbértékre minden $A \subseteq \Omega$ kosárhalmazhoz hozzárendelhető egy MB alap ε-gyakori halmaz $S_A^\varepsilon = \{\alpha_1, \alpha_2, ..., \alpha_s\}$, $\alpha_i \in \Omega$, amire

$$\begin{gathered} \Phi_A^\varepsilon = \bigcup_{i=1}^{s} L(\alpha_i) \\ \text{és} \qquad \forall i,j : 0 \leq i, j \leq s \ \alpha_i \not\leq \alpha_j \text{ és } \alpha_j \not\leq \alpha_i. \end{gathered} \tag{10}$$

Az alábbiakban a vásárlói kosarak és a tranzakciók közti kapcsolatot modellezzük.

A keresztmarketing (cross marketing), üzletek elrendezése (store layout) területeken felmerülő kérdések egyike az adott bizalmassággal rendelkező asszociációk felfedezése (Agrawal–Srikant 1994). Az itt általunk konstruált általánosabb modellben az alábbi tételben valamely értelemben megmutatunk egy explicit reprezentációt a bizalmas asszociációs szabályokra. Pontosabban: megmutatunk egy módszert, amely szerint adott α MB-re és adott bizalmassági küszöbértékre felfedezhetjük az összes olyan MB β-t, amelyikre $\alpha \to \beta$ bizalmas asszociáció.

6. Definíció – $\alpha \to \beta$ asszociáció bizalmassága. Adott $P = \{p_1, p_2, \ldots, p_n\}$ áruhalmazra, $A \subseteq \Omega$ vásárlói kosarak halmazára az $\alpha \to \beta$ *asszociáció bizalmassága*

$$\mathrm{conf}_A(\alpha \to \beta) = \frac{\mathrm{supp}_A(\alpha \cup \beta)}{\mathrm{supp}_A(\alpha)} \quad . \tag{11}$$

7. Definíció – egy $0 \le \varepsilon \le 1$ küszöbértékre egy $\alpha \to \beta$ *asszociáció ε-bizalmas*, ha

$$\mathrm{conf}_A(\alpha \to \beta) \ge \varepsilon \; . \tag{12}$$

Az összes *A*-beli ε-bizalmas asszociációk halmazát C_A^ε-val jelöljük.

Igazolható, hogy

$$\mathrm{conf}_A(\alpha \to \beta) = \frac{|U(\alpha \cup \beta) \cap A|}{|U(\alpha) \cap A|} \quad . \tag{13}$$

A modellben alkalmazott algoritmusok

Az előző pontban definiált fogalmakhoz megadjuk a meghatározásukat végző algoritmusokat.

1. Algoritmus. Az ε gyakori MB létrehozása egy adott A tranzakciók halmazára

$\Phi_A^\varepsilon := \varnothing$
k: = $[\varepsilon |A|]$
for $\{B \subset A \,\big|\, |B| = k\}$
$\Phi_A^\varepsilon := \Phi_A^\varepsilon \cup L(B)$
endfor

Az output a Φ_A^ε halmaz lesz. Az 1. algoritmus egy adott *A* tranzakciók halmazára és egy adott küszöbértékre adja meg az összes gyakori terméket.

2. Algoritmus. Alaphalmaz létrehozása

Jelöljön S_A^ε egy olyan rendezett halmazt, ahol a rendezési szempont a halmazba való bekerülés sorszáma.

```
S_A^ε = ∅
k: = [ε|A|]
for {B ⊂ A | |B| = k}
  P: = false
   i: = 1
  while i ≤ |S_A^ε| do
     if S_A^ε[i] ≤ inf(B) then
        P: = true
        if S_A^ε[i] ≠ inf (B) then
           S_A^ε = S_A^ε/S_A^e[i]
           S_A^ε = S_A^ε ∪ inf(B)
        else
           i: = i+1
        endif
     else
        if S_A^ε [i] ≥ inf(B) then
           P: = true
           i: = i+1
        endif
        else
           i: = i+1
     endif
     if not P then
        S_A^ε = S_A^ε ∪ inf(B)
     endif
  enddo
endfor
```

Az output az $S_A\varepsilon$ halmaz lesz. A 2. algoritmus egy adott *A* tranzakciók halmazára és egy adott küszöbértékre adja meg az alap gyakori termékeket.

3. Algoritmus. Az összes ε-bizalmas asszociációs szabályok generálása

```
C: = U(α) ∩ A = {γ ∈ A | α ≤ γ}
S: = [ε|C|]
for {B ⊂ A | |B| ≥ s} do
```

```
    D: = ∅
    if α ≤ inf(B) then
        D: = D ∪ inf(B)
    endif
endfor
U: =∅
for α ∈ D do
    U: = U ∪ L(α)
endfor
```

Az output az U halmaz, azaz $\cup_{i=1}^{k} L(\alpha_i)$ lesz. A 3. algoritmus egy adott *A* tranzakciók halmazára és egy adott küszöbértékre meghatározza a termékek közti bizalmas asszociációs kapcsolatokat.

Ellenőrzési, kiértékelési módszerek

A felmérésekből érkező adatokat a rögzítés után először feldolgozzuk egy adatbázis-kezelő rendszerrel. A nagy mennyiségű adatot szabványos adatbázisba helyezzük el, ahonnan a szükséges konverziók után veszi át az általunk fejlesztett *MB examination* nevű program. Az alkalmazás a megadott algoritmusok alapján elvégzi a feldolgozást. A kapott eredmények a futás után visszakerülnek egy szabványos adatbázisba, ahonnan már felhasználhatók lesznek a döntés-előkészítéshez.

Az alábbiakban egy egyszerű mintafeladaton keresztül mutatjuk be az MB algoritmusait és az általunk kifejlesztett program működését.

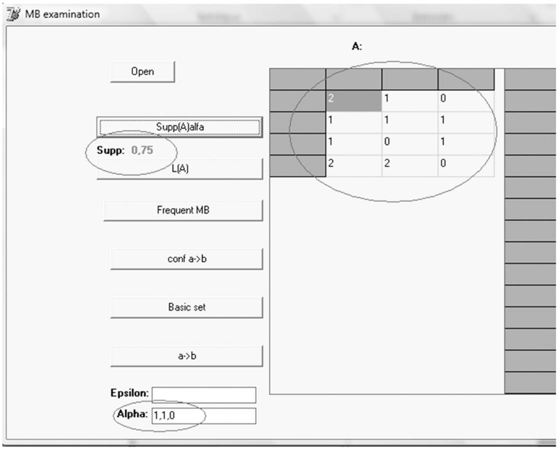
A feladatban három termékre vizsgáljuk meg az algoritmusokat.

1. Példa. Tekintsük a P = {a, b, c} termékhalmazt és legyen A ={α, β, γ, δ} tranzakciók egy halmaza, ahol

$$\alpha = (2, 1, 0), \beta = (1, 1, 1), \gamma = (1, 0, 1), \delta = (2, 2, 0).$$

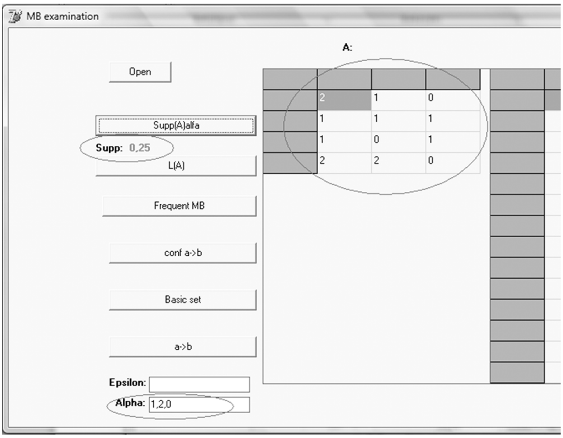
Legyen $\sigma = (1, 1, 0)$, $\eta = (1, 2, 0)$ két MB. Ekkor az (1) összefüggés alapján $supp_A(\sigma) = 0{,}75$; $supp_A(\eta) = 0{,}25$ lesz.

Az *MB examination* által számított értékeket mutatja az 1. és 2. ábra. Az *A.* táblázat mutatja az A halmazt, az *Alpha* mező mutatja az input kosarat, melyre az értéket számítjuk, és a *Supp* címke mellett jelenik meg a kiszámított érték.



Forrás: saját szerkesztés.

1. ábra. $\mathrm{Supp}_A(\sigma)$ kiszámítása



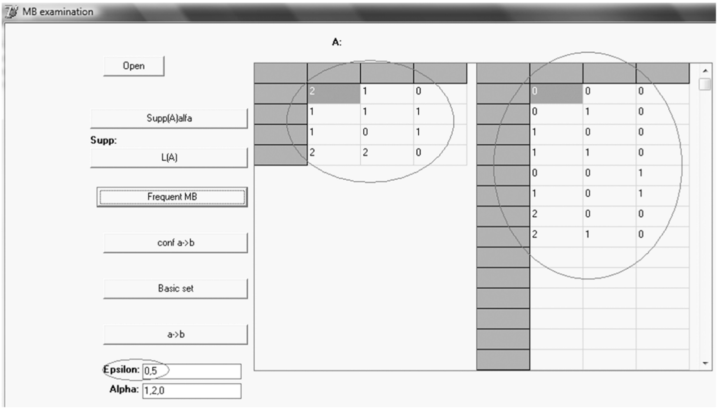
Forrás: saját szerkesztés.

2. ábra. $\mathrm{Supp}_A(\eta)$ kiszámítása

2. Példa. Legyen $= \dfrac{1}{2}$ a küszöbérték. Határozzuk meg az A ε-gyakori vásárlói kosarakat!

A 3. ábra első táblázata mutatja az A halmazt, az *epsilon* mező az értéket, a második táblázat tartalmazza az eredményt. Ekkor az *MB examination* által (2) és (7) alapján meghatározott A ε-gyakori vásárlói kosarak alábbiak lesznek (3. ábra):

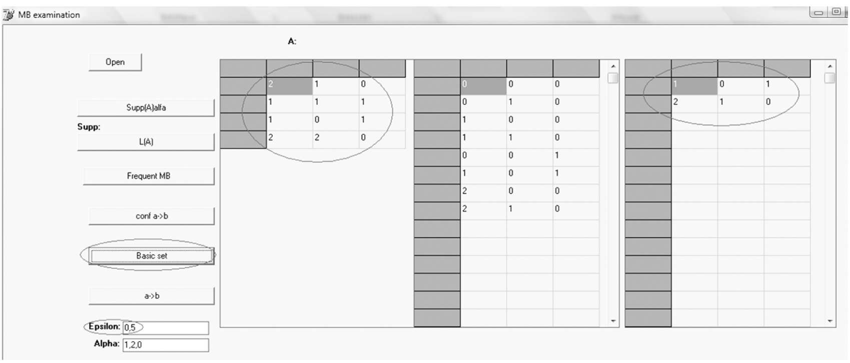
$$\Phi_A^{\frac{1}{2}} = \{(0, 0, 0), (0, 1, 0), (1, 0, 0), (1, 1, 0), (0, 0, 1), (1, 0, 1), (2, 0, 0), (2, 1, 0)\}$$



Forrás: saját szerkesztés.

3. ábra. $\Phi_A^{\frac{1}{2}}$ meghatározása

3. Példa. Az alaphalmaz meghatározása. A 4. ábra első táblázata mutatja a kiinduló adatokat, az utolsó táblázat pedig az alaphalmaz elemeit.



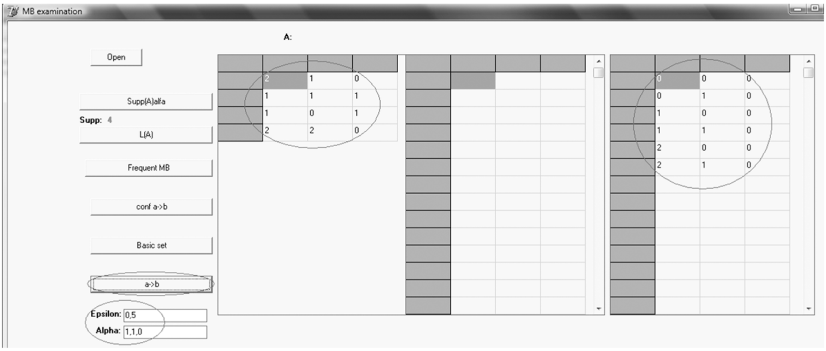
Forrás: saját szerkesztés.

4. ábra. Az alaphalmaz meghatározása

Az alaphalmaz elemei a következő MB-k lesznek: (1, 0, 1), (2, 1, 0), azaz

$$\Phi_A^{\frac{1}{2}} = L(1,\ 0,\ 1) \cup L(2,\ 1,\ 0).$$

4. Példa. Az előző példák A halmazára, a $\sigma = (1, 1, 0)$ kosárra és $= \frac{1}{2}$ küszöbértékre keressük meg az összes η kosarat, amelyikre $\sigma \to \eta$ asszociációs szabály ε-bizalmas lesz!



Forrás: saját szerkesztés.

5. ábra. A $\sigma \to \eta$ asszociációs szabály ε-bizalmas kosarainak meghatározása

A kapott eredmény:

$$L(\alpha_1) \cup L(\alpha_2) = \{(0,\ 0,\ 0),\ (0,\ 1,\ 0),\ (1,\ 0,\ 0),\ (1,\ 1,\ 0),\ (2,\ 0,\ 0),\ (2,\ 1,\ 0)\}.$$

A hatékonyság vizsgálata

Az egyes algoritmusok hatékonyságának vizsgálata fontos kérdés, hiszen nagy tömegű adat esetén nagy futási időt eredményezhet. Most csak a legfontosabb algoritmusok lépésszámvizsgálatának az eredményét mutatjuk meg. Ezek a számítások részletesen megtalálhatók a Demetrovics et al. (2011a) és Demetrovics et al. (2011b) cikkekben.

Az 1. algoritmus bonyolultsága a következő lesz:

$$O\left(\binom{|A|}{k} \cdot (m+1)^n\right) \approx |A|^k, \text{ ahol } k = [\varepsilon|A|].$$

A 2. algoritmus bonyolultsága:

$$O\left(\binom{|A|}{k} \cdot m \cdot n\right) \tag{14}$$

A 3. algoritmus bonyolultsága nagyságrendben hasonló az előzőhöz:

$$O\left(\binom{|A|}{k} \cdot m \cdot n\right). \tag{15}$$

Összefoglalás

Összefoglalásként, a meghatározott matematikai modell alapján elmondhatjuk, hogy hatékony megoldás született az eredeti problémára, és a kapott eredmények jól használhatóak lesznek a további elemzésekhez, a folyamatok racionalizálására és a megfelelő döntések meghozatalára. A megadott algoritmusokra elkészített alkalmazás a vizsgálatok alapján megfelelően elvégzi az alaphalmaz meghatározását, és alkalmas a nagy tömegű adathalmaz vizsgálatára. Emellett a későbbiekben az algoritmusok gyakorlati hatékonyságának vizsgálata is elvégezhető vele.

Irodalomjegyzék

Agrawal, R.–Srikant, R. 1994. *Fast algorithms for mining association rules*. Proceedings of the 20th VLDB Conference, Santiago, Chile, 487–499.

Benczúr, A.–Szabó, Gy. 2012. Functional Dependencies on Extended Relations Defined by Regular Languages. In: Lukasiewicz, T–Sali, A. (eds.): *Foundations of Information and Knowledge Systems, Seventh International Symposium, FoIKS 2012, Kiel, Germany, March 5-9, 2012, Proceedings. Lecture Notes in Computer Science 7153*, 384–404.

Brüggermann, T.–Hedström, P.–Josefsson, M. 2004. *Data mining and Data Based Direct Marketing Activities*. Norderstedt: Book on Demand GmbH.

Davey, B. A.–Priestley, H. A. 2002. *Introduction to Lattices and Order*. Cambridge: Cambridge University Press.

Demetrovics, J.–Hua Nam S.–Guban, A. 2011a. An algebraic approach to market basket model: explicit representation of frequent market baskets and associations rules. *Proceedings of the conference Yerevan CSIT 2011 – Computer science and information technologies*, 170–173.

Demetrovics, J.–Hua Nam S.–Guban, A. 2011b. An algebraic representation of frequent market baskets and association rules. *Cybernetics and Information Technologies* 11(2), 24–31.

Mannila, H.–Toivonen, H. 1996. Discovering generalized episodes using minimal occurrences. *Proceedings of the Second International Conferenceon Knowledge Discoveryand Data Mining* (KDD' 96). AAAI Press, 146–151.

Mendelson, E. 1964. *Introduction to Mathematical Logic*. Princeton: D. Van Nostrand Company.

Pasquier, N.–Bastide, Y.–Taouil, R.–Lakhal, L. 1999. *Discovering frequent closed item sets for association rules.* Database Theory – ICDT'99. Berlin-Heidelberg: Springer, 398–416.

Ping-Yu H.–Yen-Liang C.–Chun-Ching L. 2004. Algorithms for mining association rules in bag databases. *Information Sciences* 166(1–4), 31–47.