

Adatbányászati algoritmusok

2002/2003. tanév II. félév

ZÁRTHELYI FELADATOK

- (1) Kik és miért lesznek jelöltek az APRIORI algoritmusban? Írja le a jelöltgenerálás pontos menetét (**6 pont**)! Milyen előnyökkel és hátrányokkal járhat, amennyiben az i elemű gyakori elemhalmazokból $i+1$ és $i+2$ elemű jelölteket is generálunk, majd meghatározzuk ezek támogatottságát (**2 pont**).
- (2) Vácsolja az egyszerű mintavételező algoritmus működését (**6 pont**)! Mit tudunk mondani a minta nagyságáról? Milyen összefüggés van a partíciós algoritmusnál az egyes részek mérete és a rendelkezésre álló memória között (**2 pont**)?
- (3) Hogyan definiáltuk Toivonen algoritmusában az esélyes jelölteket? Mi célt szolgálnak az esélyes jelöltek (**4 pont**)? Mivel tud többet Toivonen algoritmus az egyszerű mintavételező algoritmusnál (**2 pont**)?
- (4) Mutasson példát arra nézve, amikor a támogatottság és a bizonyosság paraméterrel rendelkező asszociációs szabály félrevezető lehet! Hogyan definiáltuk egy asszociációs szabály érdekességét (**4 pont**)?
- (5) Legyenek X, Z diszjunkt elemhalmazok és $Z' \subset Z$. Igaz-e, hogy ha $Z \rightarrow X$ érvényes, akkor $Z \setminus Z' \rightarrow X \cup Z'$ is érvényes? Válaszát igazolja (**4 pont**)!
- (6) Jelöljük a \mathcal{D} adatbányászati környezetben található zárt elemhalmazok halmazát C -vel és legyen $X \notin C, Y, Z \in C, X \subset Y, X \subset Z$. Igazak-e az alábbi állítások?
 - Az Y és Z elemszáma különböző (**2 pont**).
 - Sem Y , sem Z nem szerepelhet egzakt szabály bal oldalán (**1 pont**).
 - Sem Y , sem Z nem szerepelhet egzakt szabály jobb oldalán (**2 pont**).
 - Ha nincs olyan $Y' \in C$, hogy $X \subset Y' \subseteq Y \cap Z$, akkor Y és Z elemszáma különböző (**2 pont**).

Válaszát minden esetben igazolja!

- (7) Írja le, hogyan döntjük el χ^2 próba segítségével, hogy egy elemhalmaz korrelált (**6 pont**). Tegyük fel, hogy α paraméterű χ^2 próbával kigyűjtjük a korrelált elemhalmazokat a $\langle k_1, k_2, \dots, k_n \rangle$ kosarak sorozatából. Ugyanezt az eredményt kapjuk-e, ha a $\langle k_1, k_2, \dots, k_n, k_1, k_2, \dots, k_n \rangle$ kosársorozatban keressük a korrelált elemeket α paraméterű χ^2 próbával (**2 pont**)? Válaszát indokolja!