

데이터마이닝 강의 -시험 문제 Section 6

강사: 심규석

[1] 다음 5개의 트랜잭션을 지닌 데이터베이스에서 최소 지지도를 60%, 최소 신뢰도를 80%로 했을 때 물음에 답하라.

| TID | Items_bought |
|------|--------------------|
| T100 | {M, O, N, K, E, Y} |
| T200 | {D, O, N, K, E, Y} |
| T300 | {M, A, K, E} |
| T400 | {M, U, C, K, Y} |
| T500 | {C, O, K, I, E} |

(a) Apriori 알고리즘을 이용하여 빈발항목집합 (frequent itemset)들을 찾는 과정을 서술하라. 이때 각 스텝별로 후보집합들과 빈발집합들을 모두 제시하라.

(b) FP-growth 알고리즘을 이용하여 빈발항목집합들을 찾는 과정을 서술하라. 이때, 각 빈발 항목별 '조건부 fp'트리와 찾아진 빈발집합들을 모두 제시하라.

(c) 주어진 지지도와 신뢰도를 만족하는 ' $\{X\} \rightarrow \{Y, Z\}$ ' 형태의 빈발 연관 규칙을 모두 찾아라. 이때, 빈발집합으로부터 효율적으로 연관 규칙을 찾기 위한 방법을 간단히 설명하라.

[2] Closed 패턴과 Max 패턴을 분석하는 이유에 대해 설명하고 각각의 정의를 제시하라.

[3] 최소 지지도와 최소 신뢰도를 만족하는 연관 규칙이 실제로는 음의 상관관계를 가질 수 있음을 간단한 예제를 통해 보여라.