

데이터마이닝 강의 – Chapter 7 시험 문제

강사: 심규석

이름:

1. Constraint 들은 다음 분류들에 속할 수 있다: antimonotonic, monotonic, convertible.

(a) 각 분류에 대해 간략히 설명하라.

(b) 다음 각 constraint 가 속한 분류들을 나열하고 그 이유에 대해 서술하라.

- i. 상품들의 평균 가격이 \$10 이하이어야 한다.
- ii. 상품들의 가격의 합이 \$100 이상이어야 한다.
- iii. 상품들의 최소 가격이 \$50 이상이어야 한다.

2. colossal patterns 이란 무엇인지에 대해 간단한 예제와 함께 간략히 서술하고, colossal patterns 을 찾기 위해 Apriori 나 FP-growth 를 적용할 때 발생할 문제점에 대해 간략히 서술하라. 또한, 이를 극복하기 위해 소개된 Pattern-Fusion strategy 에 대해 간략히 서술하라.

3. Mining compressed patterns by pattern clustering 에서 사용된 pattern distance 에 대해 서술하라.

4. 다음과 같은 Transactional Data 와 patterns P_1, P_2, P_3 가 주어졌을 때, 문제 3 의 pattern distance 을 바탕으로 P_1 은 P_2 와 P_3 중 어느 pattern 에 더 가까운지 계산하라.

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	$\{C, O, K, I, E\}$

$$P_1: \{K\} \Rightarrow \{M\}$$

$$P_2: \{E\} \Rightarrow \{M\}$$

$$P_3: \{K\} \Rightarrow \{Y\}$$