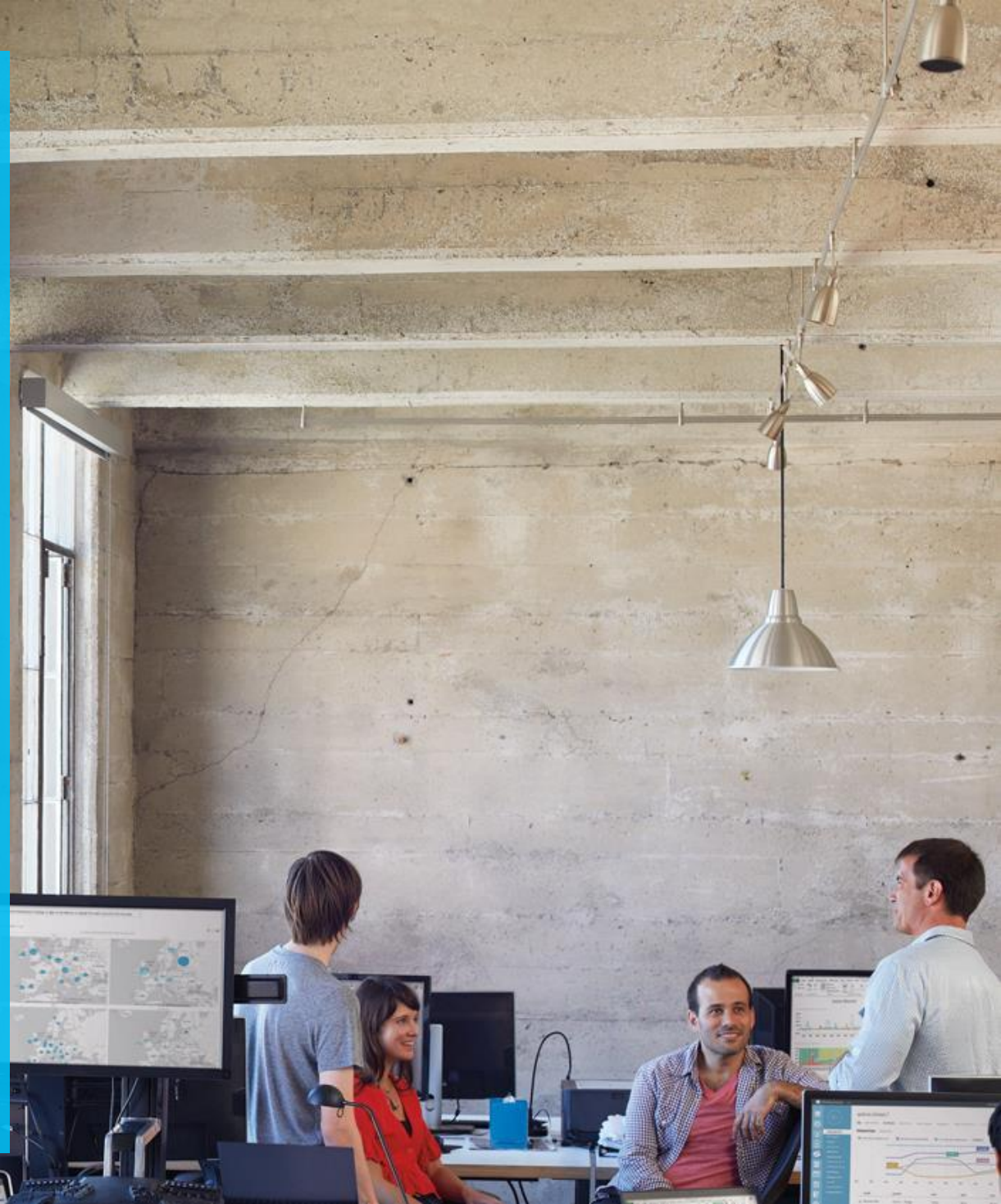




Apriori 알고리즘 사용하기

Eunji Kim,
Microsoft Technical Evangelist

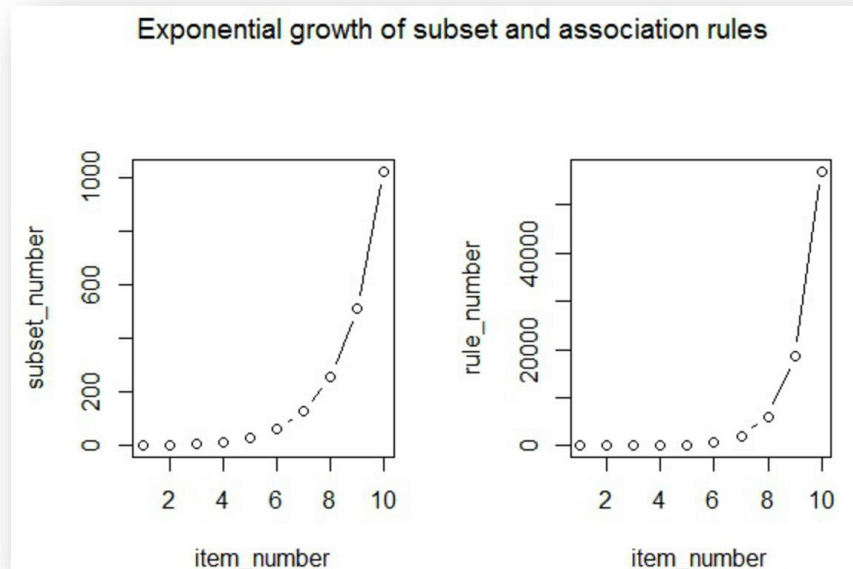


Apriori Algorithm

- 연관 규칙 알고리즘(association rule learning)의 일종으로 자주 선택되는 item 집합을 찾기위한 알고리즘.

Why?

- Item Set: $I = \{ i1, i2, \dots, ik \}$
- Number of Subset: $M = 2^k - 1$
- Number of Rules: $M = 3^k - 2^{k+1} + 1$



가능한 부분집합의 개수나
연관규칙의 개수가 item 이
증가할 때 마다 **지수적**으로 증가

How?

- 어떤 항목집합이 빈발하다면, 그 항목의 모든 부분집합도 빈발하다.

$$\forall X, Y : (X \subseteq Y) \Rightarrow s(X) \geq s(Y)$$

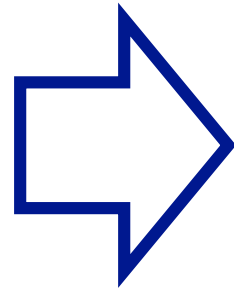
Terms

- Frequent item set (빈발 항목 집합)
- Infrequent item set (비빈발 항목 집합)
- Pruning (가지치기)
- Support (지지도)
- Confidence (신뢰도)

Example – Step1

* Support threshold: 3

Itemsets
{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}

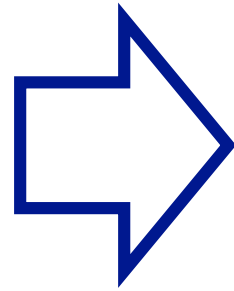


Item	Support
{1}	3
{2}	6
{3}	4
{4}	5

Example – Step2

* Support threshold: 3

Itemsets
{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}

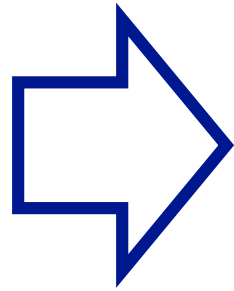


Item	Support
{1,2}	3
{1,3}	1
{1,4}	2
{2,3}	3
{2,4}	4
{3,4}	3

Example – Step3

* Support threshold: 3

Itemsets
{1,2,3,4}
{1,2,4}
{1,2}
{2,3,4}
{2,3}
{3,4}
{2,4}



Item	Support
{2,3,4}	2

Demo

<https://github.com/asaini/Apriori>

