Algorithm for listing all hypotheses that statisfy a quality check

Pratik Jawanpuria

25th January, 2010

Problem Statement:

Given data 'D':

Transaction ID	Items
1	{D,O,N,K,E,Y}
2	{M,O,N,K,E,Y}
3	{M,A,K,E}
4	{M,U,C,K,Y}
5	{C,O,O,K,I,E}

Σ = {A, C, D, E, I, K, M, N, O, U, Y}

Find all hypothesis, h, satisfying: $Q_D = \frac{|covers_D(h)|}{|D|} \geq 0.6$

Brute Force Algorithm

1. Enumerate all the subsets of Σ (which will be 2^{Σ})
2. $\forall$ h ϵ 2^{Σ}, check if $Q_D = \frac{|covers_D(h)|}{|D|} \geq 0.6$ is true

Observation

1. This algo does not account for the properties of Q_D
2. This algo does not apply understanding of the structure of search space