

Ulteriori esercizi sulle regole associative

21/11/2005
Francesco Folino

Esercizio 1

Consideriamo il seguente insieme di 3-itemsets frequenti:

$$\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,3,4\}, \{1,3,5\}, \{2,3,4\}, \{2,3,5\}, \{3,4,5\}$$

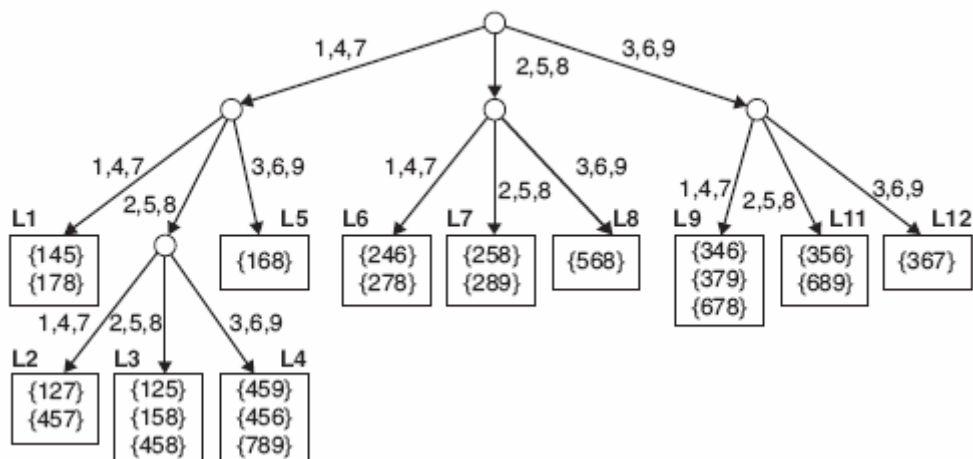
Assumiamo che ci siano solo 5 items nel dataset.

- A) Elencare tutti i 4-itemsets candidati ottenuti attraverso la procedura di *candidate generation* contenuta nell'algoritmo Apriori.
- B) Elencare tutti i 4-itemsets che sopravvivono al passo di *candidate pruning* dell'algoritmo Apriori.

Esercizio 2

L'algoritmo Apriori usa un Hash Tree per effettuare un conteggio efficiente del supporto degli itemsets candidati.

A tal proposito si consideri il seguente Hash Tree mostrato in figura:



- A) Data una transazione che contiene gli items $\{1,3,4,5,8\}$, quali nodi foglia dell'hash tree saranno visitati allorchè cerchiamo i candidati della transazione fornita?
- B) Usare i nodi foglia visitati del punto A) per determinare gli itemsets candidati che sono contenuti nella transazione $\{1,3,4,5,8\}$.

Esercizio 3

Consideriamo il seguente insieme di 3-itemsets candidati:

$\{1,2,3\}, \{1,2,6\}, \{1,3,4\}, \{2,3,4\}, \{2,4,5\}, \{3,4,6\}, \{4,5,6\}$

A) Costruire un hash tree per i 3-itemsets candidati sopra indicati. Assumiamo che l'albero usi una funzione hash che lavori così: gli items dispari sono hashati nel figlio sinistro di un nodo, mentre quelli pari sono hashati nel figlio destro. Un generico k-itemset è inserito nell'albero hashando in successione ogni item presente nel candidato e quindi seguendo l'appropriato ramo dell'albero in accordo al valore hash ottenuto. Una volta che un nodo foglia è stato raggiunto, il candidato viene inserito sulla base di una delle seguenti condizioni:

CONDIZIONE 1: Se la profondità del nodo foglia è k (la radice è assunta essere a profondità 0), allora il candidato è inserito senza preoccuparsi del numero di itemsets già memorizzati nel nodo.

CONDIZIONE 2: Se la profondità del nodo foglia è minore di k, allora il candidato può essere inserito a condizione che gli il numero di itemsets memorizzati nel nodo è minore di maxsize=2.

CONDIZIONE 3: Se la profondità del nodo foglia è minore di k e il numero di itemsets memorizzati nel nodo è uguale a maxsize, allora il nodo foglia è convertito in un nodo interno. Nuovi nodi foglia vengono creati come figli del vecchio nodo foglia. Gli itemsets candidati precedentemente memorizzati nel vecchio nodo foglia, vengono distribuiti nei figli in base al loro valore hash. Il nuovo candidato è quindi hashato nel suo giusto nodo foglia.

B) Quanti nodi foglia ci sono nell'hash tree costruito e quanti nodi interni?

C) Consideriamo una transazione che contiene i seguenti items: $\{1,2,3,5,6\}$. Usando l'hash tree costruito al punto A), quali nodi foglia saranno visitati nel caso della transazione indicata? Quali sono i 3-itemsets candidati contenuti nella transazione?