INF 553- Spring 2019

Finding Complete Bipartite Subgraphs

Apriori Algorithm - Algorithm used

1. Read the input text file, k-value, support and convert input file into a list of baskets by grouping using the right node as key and left node as value.
2. For each value of i from 1 to k-1:
    a. Initialize mapper(itemset : count) dictionary
    b. For each basket in the list of baskets:
        i. Generate all possible combinations of size i from the basket.
        ii. Each generated combination is a candidate itemset,
            1. If either all subsets of the given itemset is a frequent itemset.
            2. Or if i =1
        iii. If a candidate itemset is already in the dictionary mapper, increment its count value by 1, else initialize its count value to 1.
    c. Filter the mapper dictionary using the count value such that only itemsets having count greater or equal to support threshold remain in the frequent_itemset list.
    d. If there is no element in the frequent itemset list, display a message "No Subraph of given size exists" terminate the program.
3. Initialize mapper(itemset : count,basket_name_list) dictionary.
4. For each basket in the list of baskets:
    a. Generate all possible combinations of size k from the basket.
    b. Each generated combination is a candidate itemset,
        i. If either all subsets of the given itemset is a frequent itemset.
        ii. Or if k =1
    c. If a candidate itemset is already in the dictionary mapper, increment its count value by 1 and append name of basket to basket_name_list, else initialize its count value to 1 and initialize the basket_name_list to the name of basket.
5. Filter the mapper dictionary using the count value such that only itemsets having count greater or equal to support threshold remain in the frequent_itemset list.
6. If there is no element in the frequent itemset list, display a message "No Subraph of given size exists" terminate the program.
7. For each frequent itemset print the {itemset} , {basket_names} for all possible combinations of the {basket_names} of size k where itemset occurs.