Data Analytics 1

Assignment 2

BUC Algorithm and Attribute oriented induction

Release : 1st September 2023
Deadline : 12th September 2023 (11:55 pm)

The objective of the assignment is to get the exposure to
extract interesting CUBEs from the given large data set based on the
BUC algorithm and get an understanding of Attribute oriented induction

Part A (85 marks):
Details of the data set for BUC Algorithm are available
at the following link [link]. Generalize/Categorize the attributes wherever required and use all the attributes.
https://dl.acm.org/doi/pdf/10.1145/304182.304214
Implement the Bottom-Up Cube (BUC) algorithm . This question has two parts:

- Assume that all CUBEs can be supported by the main memory and write a straight-forward /implementation of BUC.
- Assume that the program will run out of main memory and introduce paging in your implementation.

Finally, plot the following graphs over multiple runs of the algorithm while varying some parameters and keeping others constant.
a. A plot of minsup vs. runtime, keeping allotted memory fixed.
b. A plot of allotted memory vs. runtime, keeping minsup fixed.

Part B (15 marks):

- Extract characteristic rules with attribute-oriented induction(data generalization by attribute removal or attribute generalization)
- Link for the dataset

Submission format:

- Mention any inference from plots or results in notebook Markdown itself
- Submit a zip folder named <assignment2_teamId> containing a single file <assignment2_teamId>.ipynb