

CSE 572 Project 4 (OPTIONAL, Not extra credit)

Aim will be to determine anomalous events through Association rule mining.

Consider the data from Project 1.

Step 1: Define bins of CGM values for the meal data that you extracted in Assignment 2. Lets say that you had $P \times 30$ meal data matrix. Find the minimum CGM out of all $P \times 30$ CGM values, lets say this is CGM_{min} . Find the maximum CGM out of all $P \times 30$ CGM values, lets say this is CGM_{max} . Consider a bin size of 20 mg/dl. Make bins of size 20 mg/dl starting from CGM_{min} to CGM_{max} .

Step 2: Extract the bolus insulin data from InsulinData.csv for each meal time. This is the T column data (round it to the nearest integer, if 5.5 then take it as 6).

Step 3: Find the bin number for the maximum CGM (B_{max}) from each meal data that you extracted in Assignment 2.

Step 4: Find the bin number CGM value that is at the time when the meal was taken (B_{meal}).

Extract these metrics from all meal instances for patient in Assignment 1.

1. Report the most frequent itemsets for each of the subjects (B_{max} , B_{meal} , Insulin Bolus)
2. Consider the rule of the form: $\{B_{max}, B_{meal}\} \rightarrow I_B$

Find the rule with the largest confidence for the subject.

Extract all rules that you observe. Calculate confidence of each observed rule.

3. Find anomalous events by finding the least confidence rules.
Rank rules according to confidence.

Expected Output

1. CSV File with most frequent sets. One row for each itemset.
2. CSV file with largest confidence rules. One row for each rule. Rules are of the form $\{B_{max}, B_{meal}\} \rightarrow I_B$
3. Anomalous rules, Rules with confidence less than 15 %. One row for each rule.

Grading:

If your code returns three CSV files with rules. Then we code will be analyzed to see if you did the rule search. (70 points)

Majority rules. If your rule falls in the majority set your get 30

If x percent of your rules fall in majority set you get $p = 30 * \frac{x}{100}$