HAMBURG UNIVERSITY OF TECHNOLOGY

PHILIPS

FISHING FOR EXPERIENCE

P12: Image Chain Hamburg – Philips Medical System

Nagalikhitha Reddipalli[1], Sohui Lee[2], Rahul Ramachandra Hubli[1], Kyuri Kim[2], Abhishek Singh Suraj[2], Rohan Mandal[2], Tobias Reusch[3], Leo Kempkes[3], Vivien Ruffer[3]

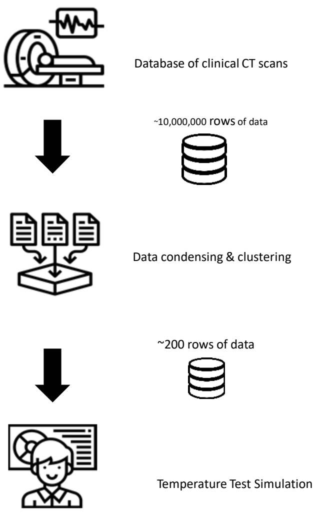
[1]Department of Mechanical Engineering, Hamburg University of Technology, [2]Department of Electrical Engineering, Computer Science and Mathematics, Hamburg University of Technology, [3]Philips Medical System, Hamburg

Introduction

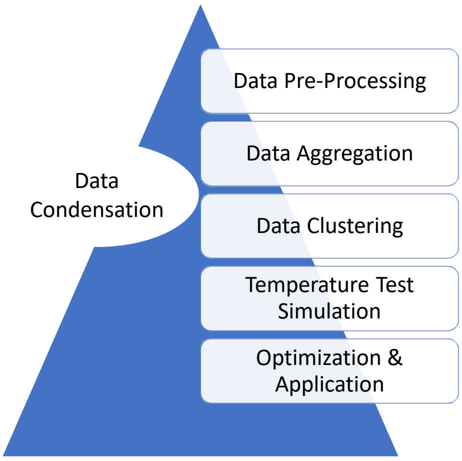
Computed tomography (CT) systems from Philips are used daily to save lives across the globe with a wide range of usage patterns. During the development of the components for the systems, too little consideration is still paid to the present consumption profile of the hospitals. We aim to better understand this with the use of data analytics so that we can create better goods in the future and assist even more people around the globe. Generally, proposed concepts for X-Segments are generated using simple input assumptions regarding client needs. These approaches therefore are not optimized to match the genuine needs of the market, leading to degraded performance (e.g. forced wait times between scans) or needless costs / commercial uncertainty / overdesign. This project aims to analyze the clinical scan data gathered in the database RADAR, investigate and model how potential future market trends could affect usage, and use these insights to directly drive the proposed performance characteristics of a new X-Segment concept design, for example, the necessary Thermal Heat Capacity (and thus size of the Anode Teller). Much of the size, weight, and cost characteristics of an X-ray tube are primarily influenced by the size of the anode teller.

What is this all about?

We have provided roughly 10,844,419 rows of data, each containing relative values like voltage, resolution, and exposure time that were gathered from CT scans of actual patients at various hospitals around the world. We only need about 100-200 rows from the original data to run temperature test simulation (TTS). For data pre-processing, we removed irrelevant or inaccurate data which was down to 4,780,000. *After, we condensed them in 2 steps. First, we compacted the dataset by monthly basis.* Then we clustered data using two different clustering algorithms which are K-means and K-medoids clustering algorithm. So that we could compare between two algorithms, choose which one is more efficient. *Then, we extracted average values of important columns from each cluster.* Using the final data, we could run TTS, observe thermal and power response of CT scanner which represents performance qualities.



Process Flow

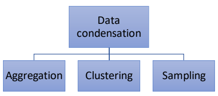


Code Structure

1. Data Pre-processing and aggregation
   1) Data filtering
      - Removing incomplete and duplicated data
      - Roughly 40,000 rows are filtered
   2) Monthly-based condensation
   3) Average values of each patient identification number
2. Normalization
   - To avoid clustering algorithm is just driven by large numbers
   - MinMaxScaler from Sklearn library
3. Clustering algorithms
   - Standard K-means clustering
   - K-medoids clustering
4. Sampling
   - Proportional sampling to current usage
   - Random sampling
5. Optimization
   - Optimizing time is needed for endurance run (TTS)

Data Condensation

Data condensation is the process of summarizing a large amount of data into a smaller, more manageable form. It involves analyzing the data to identify patterns, trends, and key insights, and then presenting these findings in a concise and easily understandable format.



Data aggregation is the process of combining multiple individual data points into a single summary or statistical value. It involves grouping together similar or related data points and calculating a summary statistic, such as the average, sum, count, or maximum value, for each group. It can be particularly useful when dealing with complex data that includes many individual data points, as it can help to highlight important patterns and trends in the data.

Data clustering is a technique used in data analysis to group similar data points together into clusters or groups. It involves analyzing a dataset to identify patterns and similarities between individual data points, and then grouping these data points based on their similarities.

Sampling is selecting the data points from clusters.

Clustering Algorithms

K-means is a clustering algorithm that groups similar data points together based on their distance from each other. The "k" in k-means refers to the number of clusters that the algorithm will try to create. The goal of k-means is to minimize the sum of squared distances between data points and their assigned centroids, which is known as the "inertia." The resulting clusters can be used to identify patterns in the data and make predictions based on those patterns.

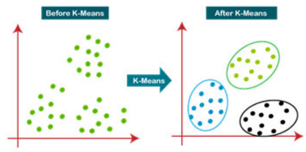


Figure 1. K-Means Clustering

K-medoids is a clustering algorithm similar to k-means, but instead of using the mean (centroid) to represent each cluster, it uses the most "representative" point, known as the medoid. The resulting clusters can be used to identify patterns in the data and make predictions based on those patterns. The advantage of k-medoids over k-means is that it is more robust to outliers and non-linear distances, since it only relies on the distances between individual data points and medoids, rather than the mean (centroid) of each cluster.

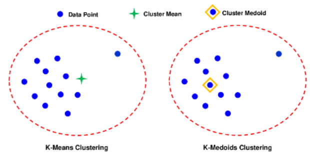


Figure 2. K-Means vs K-Medoids Clustering

Results