A Collaborative Space Analysis Framework for Doctor-Patient Workflow for Hypertension

**Author:** Nihar Patil (nrp9155@rit.edu)

**Advisor:** Prof. Chris Bondy, Dr. Shi.

**Goal:** Build a smart platform, using data mining, to improve collaboration between doctor-patient which helps doctor to provide better medications and efficient utilization of their time.

1. User creates Profile includes their lifestyle information like smoking, drinking, BMI information.
2. Apply Clustering algorithm to determine the cluster of the patient.
3. Verify the cluster with the doctor.
4. Monitor changes in the cluster and the new readings of the patients’ blood pressure.
5. Generate alerts based on the values.
6. Provide lifestyle improvement feed to the patient.
7. Update profile based users’ implementation of lifestyle feed.
8. Recompute the cluster and monitor the changes.

**System Architecture**

1. User creates Profile includes their lifestyle information like smoking, drinking, BMI information.
2. Apply Clustering algorithm to determine the cluster of the patient.
3. Verify the cluster with the doctor.
4. Monitor changes in the cluster and the new readings of the patients’ blood pressure.
5. Generate alerts based on the values.
6. Provide lifestyle improvement feed to the patient.
7. Update profile based users’ implementation of lifestyle feed.
8. Recompute the cluster and monitor the changes.

**K-means Clustering**

**K-means** – The algorithm produces K centroids which help in segregating a dataset to meaningful groups or categories. The process also helps in assigning a new incoming datapoint to an unsupervised cluster.

**ROCK (Robust Clustering using lInKs)**

**Neighbors:** A datapoints neighbor are those points that are present nearest to the current datapoint. A datapoints neighbor is calculated using Euclidean distance. ROCK considers more than one neighbor while deciding cluster of a data point. The lesser the Euclidean distance, closer is the datapoint.

**Links:** Clustering points based on closeness is not strong enough to distinguish two not so well separated clusters. If a datapoint and its neighbor are considered in a cluster only if they share large number of neighbors.

**To evaluate the clusters quality, the most common approach is to calculate the sum of squared error distance.** As it gives the variation in each cluster. The value with minimum SSE has best clusters.

**SSE values changes significantly for both algorithms, with the addition of outlier or with the addition of an extra attribute.**

**Conclusion:** ROCK handles outliers effectively compared with K-means.