Clustering Surgical Procedures Using Clinically Significant Pre-operative Information

Christopher P. Cheng, AB[1]; Ira S. Hofer, MD[2]

[1]Department of Medical Education, Icahn School of Medicine at Mount Sinai

[2]Department of Anesthesiology, Perioperative and Pain Medicine, Icahn School of Medicine at Mount Sinai

Background: Surgical classification methods are necessary for at-large comparisons of the outcomes, costs and incidence of patient care to generate actionable insights for improving patient care. The codes used in the Healthcare Cost and Utilization Project are nationally used for research on the impact of natural disasters, access to care, quality of care and more. This abstract proposes applying clustering analysis to clinically significant pre-operative information to facilitate clinically meaningful surgical procedure classification.

Methods: 199463 procedures that took place in the Mount Sinai Health System from March 12, 2018 to September 22, 2021 were pulled with clinically significant features including but not limited to patient age and sex, surgeon ID, surgeon specialty, operating room, American Society of Anesthesiologists (ASA) score, surgical admission number, scheduled length of surgery, and patient class. These data were preprocessed as follows: null values were dropped from the dataset, numerical and ordinal data were min-max scaled, and categorical data underwent one-hot encoding. Subsequently the dataset was clustered using the KMeans function of scikit-learn 1.0.1 with the random state set to 0 and all other parameters set to their default values with cluster values ranging from 5 to 450. Each cluster number's performance was evaluated using a scree plot elbow method as well as a silhouette score under both Euclidean and correlation distance metrics. The silhouette score of the optimal cluster number was then compared with that of the CCS codes.

Results: Applying the elbow method to the scree plot suggests the optimal cluster number is about 50, and its silhouette scores were accordingly found to be 0.803 and 0.715 according to the correlation and Euclidean distance metrics, respectively. Conversely, CCS code silhouette scores were found to be -0.349 and -0.264 according to correlation and Euclidean metrics, respectively. Figure 1 illustrates the scree plot and silhouette score plot of this clustering method across all cluster numbers.

Conclusion: Using clustering to leverage clinically significant features for more clinically meaningful classification of features shows promising performance compared with current standards in the field. Further work will focus on developing quantitative methods for feature selection and practical applications for clustering classifications.

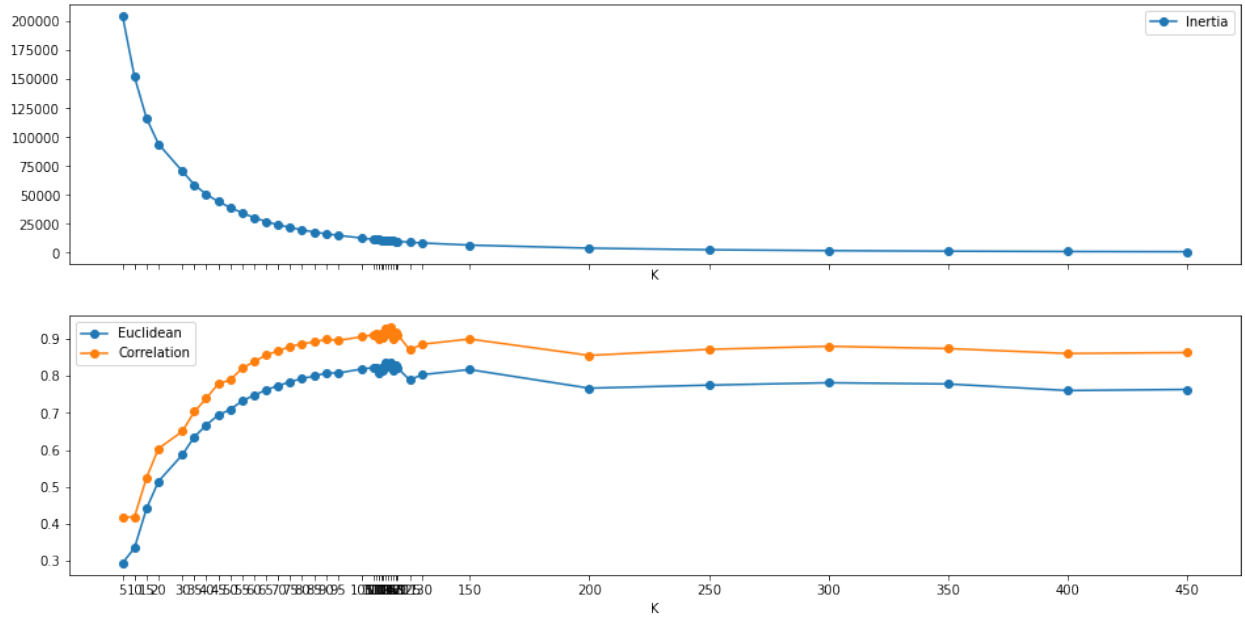

Figure 1. (Above) Scree plot plotting cluster number against proportion of variance explained within the dataset. (Below) Silhouette scores plotted for each cluster number using both Euclidean and correlation distance metrics.

References:
Number references in the order in which they appear in the abstract. The journal, volume and page number are required.