Clustering clinical data: Biologically meaningful simulations and mixed-distance dissimilarity solutions

Abstract: Unsupervised clustering poses unique challenges in clinical data due to their heterogeneous size and mixed type. We hypothesized that these limitations could be overcome by calculating dissimilarity by combining multiple distance methods. A review of the literature suggested that existing solutions for mixed, clinical data are sparse and lack rigor. In an initial experiment on real clinical data, we found limitations in a common approach: converting a mixed data set to a single data type. To rigorously test dissimilarity metrics and clustering methods, we developed 32,400 simulations of realistic, mixed-type clinical data. We used these simulations to test 3 clustering algorithms (hierarchical clustering, Partitioning Around Medoids, and self-organizing maps) on 5 single distance metrics (Jaccard Index, Sokal & Michener or Hamming distance, Gower coefficient, Manhattan distance, Euclidean distance) and 3 multiple distance methods of calculating dissimilarity (DAISY, Supersom, and Mercator, a method of our own devising). We applied the superior solution for a data mixture predominated by binary features, DAISY with Ward’s hierarchical clustering, to the data set from our initial experiment, and recovered important prognostic features. These experiments raise future questions for clustering problems in clinical data, including identifying minimum size for successful clustering (relevant when clustering clinical trials) and addressing concerns for validation of sometimes variable outcome. This talk elaborates on thesis work undertaken for an M.S. in Biomedical Informatics at OSU in Spring 2020.

About the Speaker: Caitlin E. Coombes graduated with an M.S. in Biomedical Informatics from The Ohio State University in 2020. She is currently an M.D. candidate, Class of 2022, at The Ohio State University College of Medicine (OSUCOM), where she has been inducted into the Landacre Research Honors Society and the Gold Humanism Honors Society. She obtained her B.A. in Art History at Rice University in 2009 (Phi Beta Kappa, Summa Cum Laude). Her clinical interests include anesthesiology, critical care, and palliative medicine. She is the student leader of the OSUCOM Humanism in Medicine Research Team, and her research on the role of arts education in fostering empathy, resilience, and clinical excellence in physician trainees has been presented nationally and internationally, including at the annual meetings of the American Association of Medical Colleges (Phoenix, AZ) and the American Osler Society (Montreal, CA). She is currently pursuing clinical informatics research focused on predictive modeling of delirium in intensive care units. Her papers have been published in the Journal of Infectious Disease and the Journal of the American Medical Informatics Association.

Caitlin E. Coombes, MS
Friday, October 9th, 11:00am-12:00pm
Carmen Zoom