Demo: Visual Supervision of Unsupervised Clustering of Patients with Clustervision

Bum Chul Kwon, Ben Eysenbach, Janu Verma, Kenney Ng, Christopher deFilippi, Walter Stewart, and Adam Perer

Clustering, the process of grouping together similar items into distinct partitions, is a common type of unsupervised machine learning that can be useful for summarizing and aggregating patients by their complex multi-dimensional data. However, patients can be clustered in many ways, and there exist a large body of algorithms designed to reveal different patterns. While having access to a wide variety of algorithms is helpful, in practice, it is quite difficult for clinical researchers to choose and parameterize algorithms to get the clustering results relevant for their dataset and analytical tasks. To alleviate this problem, we built a visual clustering analysis system, *Clustervision* that helps ensure clinical researchers find the right clustering among the large amount of techniques and parameters available. *Clustervision* clusters data using a variety of clustering techniques and parameters and then recommends good clustering results utilizing a variety of quality scoring metrics. In addition, users can guide the system to produce more relevant results by providing task-relevant constraints on the data. Our visual user interface allows users to find high quality clustering results, explore the clusters using a variety of coordinated visualization techniques, and select the cluster result that best suits their task. We have demonstrated the effectiveness of Clustervision using a case study as a team of clinical researchers interested in using unsupervised clustering techniques on a longitudinal database of electronic medical records. The research team found clusters of similar patients that represented meaningful groups of patients with heart failure using a database of approximately 1,500 patients. Full details of the system and the case study are available in an upcoming journal paper Kwon et al. (2018).

Figure 1: An overview of *Clustervision* on a dataset describing 1500 patients with Heart Failure.

Bum Chul Kwon, Ben Eysenbach, Janu Verma, Kenney Ng, Christopher deFilippi, Walter Stew- art, and Adam Perer. Clustervision: Visual supervision of unsupervised clustering. In *IEEE Transactions on Visualization and Computer Graphics*, 2018.