Cluster analysis of simulated gravitational wave triggers using constrained validation clustering TING ZHANG, LAPPOON TANG, SOMA MUKHERJEE, LIGO COLLABORATION — The data collected in the science run of LIGO calls for a thorough analysis of the glitches seen in the gravitational wave channels, as well as in the auxiliary and environmental channels. Rapid growth in size and number of available databases requires fast and accurate data mining algorithms for timely glitch analysis. The study presents a new technique in cluster analysis that we call constrained validation clustering (CV clustering) for mining patterns in gravitational wave burst triggers. The approach avoids using Gaussianity assumptions on data distribution, and was shown to outperform a state of the art in clustering – G-means – when K, the number of clusters, is unknown (Tang et. al., 08); experimental results suggested that Guassian mixture assumption can be too strong as a machine learning bias in mining gravitational wave data, evidenced by very severe overfitting of data by G-means. Our current focus is on upgrading CV clustering to utilizing random sampling and stochastic optimization techniques. Preliminary results indicate that such an enhancement can potentially bring about a forty fold increase in computational efficiency while suffering minor degrade in model quality. A current future direction is in further improving quality of models learned by the algorithm for making it an effective approach for real LIGO data analysis.

Ting Zhang