Mutual Nearest Neighborhood: A Fundamental Solution to Data Clustering

Feng, Crystal (School: Western Albemarle High School)

Data clustering is an unsupervised machine learning technique that groups similar data points together based on their characteristics. Although numerous data clustering algorithms have been developed, they often exhibit limitations when handling clusters with varying shapes or densities, particularly in high dimensional data. In this project, a new clustering algorithm MNN, based on the mutual nearest neighborhood property of a dataset, was developed. MNN naturally initializes cluster cores from mutual nearest neighboring data points and then grows from mutual nearest neighboring clusters step by step. Clustering tests on different datasets show that the MNN algorithm works well on clusters of different shapes and different densities. MNN outperforms current prominent algorithms on low dimensional data and works effectively on high dimensional data without prior dimension-reduction.

Awards Won:
Association for the Advancement of Artificial Intelligence: AAAI Student Memberships for each finalist that is part of the 1st, 2nd, and 3rd Prize Winning projects and 5 Honorable Mention winning projects (up to 3 students per project) (in-kind award / part of the 1st-3rd prize)