Clustering Using B-Matching and Semidefinite Embedding Algorithms

“Lead Inventor: Tony Jebara, Ph.D.

Algorithm Widely Applicable to Clustering Problems Clustering of a dataset into two or more larger subsets is a fundamental problem in a variety of fields ranging from machine learning and databases to medical imaging and market analysis. In many instances, the clustering method must be done without supervision, and current implementations, including the most popular spectral method, have limited accuracy.

Clustering Algorithm Produced from B-Matching and Semidefinite Relaxation Algorithms This technology is a method that has the similar accuracy to the best method currently available, but has the added benefit of being more widely applicable to general clustering problems. This technology uses the cubic-time algorithm known as b-matching to find the most similar regular graph to a given weighted graph. Thus, once this is achieved, the semidefinite relaxation method may be implemented. The combination of these two methods makes this technology the most accurate clustering algorithm that is widely applicable for clustering problems that produce weighted graphs. The theoretical results of this method provide a reliable clustering algorithm that is efficient and outperforms competing methods.

Applications:
• Databases – reorganization and classification of data
• Marketing – identify groups of customers based on user information
• Molecular Biology – Proteins and DNA contain long sequences of data, clustering algorithms may be able to group these sequences into different groups for more accurate interpretation

Advantages:
• Accurate clustering algorithm
• Widely applicable, even to clustering problems presented as weighted graphs

Patent Status: Patent Pending

Licensing Status: Available for Licensing and Sponsored Research Support


Inventors

Tony Jebara