The unsupervised ensemble learning, or consensus clustering, consists of finding the optimal combination strategy of individual partitions that is robust in comparison to the selection of an algorithmic clustering pool. Despite its strong properties, this approach assigns the same weight to the contribution of each clustering to the final solution. We propose a weighting policy for this problem that is based on internal clustering quality measures and compares against other modern approaches. Results on publicly available datasets show that weights can significantly improve the accuracy performance while retaining the robust properties. Since the issue of determining an appropriate number of clusters, which is a primary input for many clustering methods, is one of the significant challenges, we have used the same methodology to predict correct or the most suitable number of clusters. Among various methods, using internal validity indexes in conjunction with a suitable algorithm is the most popular way to determine the appropriate number of cluster. Thus, we use weighted consensus clustering along with four different indexes which are Silhouette (SH), Calinski-Harabasz (CH), Davies-Bouldin (DB), and Consensus (CI) indexes. Our experiment indicates that weighted consensus clustering together with chosen indexes is a useful method to determine right or the most appropriate number of clusters in comparison to individual clustering methods (e.g., k-means) and consensus clustering. Lastly, to decrease the variance of proposed weighted consensus clustering, we borrow the idea of Markowitz portfolio theory and implement its core idea to clustering domain. We aim to optimize the combination of individual clustering methods to minimize the variance of clustering accuracy. This is a new weighting policy to produce partition with a lower variance which might be crucial for a decision maker. Our study shows that using the idea of Markowitz portfolio theory will create a partition with a less variation in comparison to traditional consensus clustering and proposed weighted consensus clustering.

Major: Industrial Engineering

Educational Career:
Bachelor's of Industrial Engineering, BS, 2010, Istanbul University
Master's of Industrial Engineering, MS, 2014, University of Pittsburgh

Committee in Charge:
Petros Xanthopoulos, Chair, Industrial Engineering and Management Systems
Qipeng Zheng, Co-Chair, Industrial Engineering and Management Systems
Luis C. Rabelo, Industrial Engineering and Management Systems
Hae-Bum Yun, Civil, Environmental and Construction Engineering

Approved for distribution by Petros Xanthopoulos, Committee Chair, on February 2, 2017.

The public is welcome to attend.