A COMPARATIVE ANALYSIS OF CLUSTERING ALGORITHMS

Şahin EMRAH AMRAHOV

Number of data increases and data analysis gains more importance day by day. Clustering is one of the most important stages of data analysis. There is no accepted definition of clustering by all researchers today. However, all researchers accept that the elements in the same cluster should be similar, elements in the different clusters should be different from one another and measures of similarities and dissimilarities should be sufficiently clear and practical. The clustering process usually has 4 steps: 1) Extraction and selection of features that represent the data set best 2) Design of the appropriate clustering algorithm for the problem examined 3) Evaluation of the results and validation of the designed algorithm 4) Interpretation of the results and practical explanation of the clusters.

In this presentation, a comparative analysis of clustering algorithms is made. Each of these algorithms has strengths and weaknesses. There is no guarantee that an algorithm that gives very good results for a specific problem will give similar results for another problem. For this reason, it is important to be able to select the appropriate algorithm for the problem considered. Therefore, the algorithms are needed to be well analyzed by comparing them with each other.

References

1. Xu R, Wunsch D (2005) Survey of clustering algorithms. *IEEE Trans Neural Netw,* 16:645–678
2. Xu, D., & Tian, Y. (2015). A comprehensive survey of clustering algorithms. *Annals of Data Science*, *2*(2), 165-193.
3. Yassouridis, C., & Leisch, F. (2017). Benchmarking different clustering algorithms on functional data. *Advances in Data Analysis and Classification*, *11*(3), 467-492.