Similarity based Fuzzy and Possibilistic K-Means Clustering on Biomedical Data for Disease Evaluation

S. Prema*; Dr. P. Umamaheswari**

*Research Scholar,
Information and Communication Engineering,
Anna University,
Chennai, India.
**Associate Professor,
Computer Science and Technology,
MIT Campus, Anna University,
Chennai, India.

Abstract

The abundance amount of biomedical data availability has drastically increased the need for new data mining techniques in order to obtain the health related information. To identify and evaluate the healthcare information and for easy treatment of disease, Machine Learning (ML) methodology has been developed. ML algorithms and textual demonstration classify short medical texts and recognizes the semantic relations among the diseases and treatments. However, ML methodology does not provide clustering operations with which correct labels can be easily predicted. Richer contextual information of diseases and treatments were extracted using kernel based learning. Clustering of similar richer contextual information is the crucial point for easy analysis of diseases. To improve the evaluation and to group similar form of biomedical data, in this article we present Similarity based Fuzzy and Possibilistic K-means Clustering (SFP K-means Clustering) technique. SFP K-means Clustering is a type of machine learning approach which is used for grouping similar type of medical diseased data. At the first step, fuzzy means uses the cylindrical standard to measure similarity between medical data points. Cylindrical norm in SFP K-means Clustering uses the Bessel function of first and second order to measure the similarity and produce the fuzzy result. Second, the result from the fuzzy rule is used for constructing Possibilistic K-means Clustering. With the introduction of possibilistic K-Means, SFP K-means Clustering groups the similar data using degree membership function. SFP K-means Clustering addresses the problem of reducing the execution time by grouping similar data while analyzing the heart disease. The efficacy of SFP K-
means Clustering is validated by extensive evaluation using heart disease dataset for measuring the clustering accuracy, chance of correctly predicting the disease, and execution time factors.

**Keywords:** Possibilistic K-means Clustering, Biomedical Data, Degree Membership Function, Fuzzy Rule, Bessel Function, Similarity Measure.

**References**


Andrew R. Post, Tahsin Kurc, Sharath Cholleti, Jingjing Gao, Xia Lin, William Bornstein.


Xin Sun, Yanheng Liu, Da Wei, Mantao Xu, Huiling Che, Jiawei Han,” Selection of interdependent genes via dynamic relevance analysis for cancer diagnosis”, Journal of Biomedical Informatics, Elsevier, Nov 2012.


Yiu-ming Cheung, Hong Jia,” Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number”, Pattern Recognition, Elsevier, Jan 2013.


