

as an input to obtain the pairwise Maximum Edge Distance (MED). MED is used to compute the average Intracluster Distance. This method is based on the relationship between patterns and their clustering labels. The dMED and IC-av functions can be processed parallel in the hadoop environment in order to produce more accurate results in an efficient timely manner for both artificial and gene expression dataset.

References

- [1] K.Y. Yeung, D.R. Haynor, and W.L. Ruzzo, "Validating Clustering for Gene Expression Data," *Bioinformatics*, vol. 17, no. 4, pp. 309-301, 2001.
- [2] A. Ben-Hur and A. Elisseeff, and I. Guyon, "A Stability Based Method for Discovering Structure in Clustered Data," *Proc. Pacific Symp. Biocomputing*, pp. 6-17, 2002.
- [3] S. Monti, P. Tamayo, J.P. Mesirov, and T.R. Golub, "Consensus Clustering: A Resampling-Based Method for Class Discovery and Visualization of Gene Expression Microarray Data," *Machine Learning*, vol. 52, nos. 1/2, pp. 91-118, 2003.
- [4] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the Number of Clusters in a Data Set via the Gap Statistic," *J. Royal Statistical Soc. B*, vol. 63, pp. 411-423, 2003.
- [5] I. Dhillon, Y. Guan, and B. Kulis, "Kernel K-Means, Spectral Clustering and Normalized Cuts," *Proc. Int'l Conf. Knowledge Discovery and Data Mining*, pp. 551-556, 2004.
- [6] R. Tibshirani and G. Walther, "Cluster Validation by Prediction Strength," *J. Computational & Graphical Statistics*, vol. 14, no. 3, pp. 511-528, 2005.
- [7] A. Azran and Z. Ghahramani, "Spectral Methods for Automatic Multiscale Data Clustering," *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 190-197, 2006.
- [8] V. Pihur, S. Datta, and S. Datta, "Weighted Rank Aggregation of Cluster Validation Measures: A Monte Carlo Cross-Entropy Approach," *Bioinformatics*, vol. 23, no. 13, pp. 1607-1615, 2007.
- [9] Z.V. Volkovich, Z. Barzily, G.-W. Weber, and D. Toladano-Kitai, "Cluster Stability Estimation Based on a Minimal Spanning Trees Approach," *Proc. AIP Conf.*, pp. 299-305, 2009.
- [10] G. Stegmayer, D.H. Milone, L. Kamenetzky, M.G. Lpez, and F. Carrari, "A Biologically Inspired Validity Measure for Comparison of Clustering Methods over Metabolic Data Sets," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, vol. 9, no. 3, pp. 706-716, May/June 2012.
- [11] A.E. Baya and P.M. Granitto, "How Many Clusters: A Validation Index for Arbitrary Shaped Clusters," *BMC Bioinformatics*, vol. 10, article 2, April 2013.