Cornell University Library

We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > stat > arXiv:1305.7285

Search or Article-id

(Help | Advanced search)

All papers   Go!

Statistics > Computation

Adapting the Interrelated Two-way Clustering method for Quantitative Structure-Activity Relationship (QSAR) Modeling of a Diverse Set of Chemical Compounds

Subhabrata Majumdar, Subhash C. Basak, Gregory D. Grunwald

(Submitted on 31 May 2013)

Interrelated Two-way Clustering (ITC) is an unsupervised clustering method developed to divide samples into two groups in gene expression data obtained through microarrays, selecting important genes simultaneously in the process. This has been found to be a better approach than conventional clustering methods like K-means or self-organizing map for the scenarios when number of samples much smaller than number of variables (n<<p). In this paper we used the ITC approach for classification of a diverse set of 508 chemicals regarding mutagenicity. A large number of topological indices (TIs), 3-dimensional, and quantum chemical descriptors, as well as atom pairs (APs) have been used as explanatory variables. In this paper, ITC has been used only for predictor selection, after which ridge regression is employed to build the final predictive model. The proper leave-one-out (LOO) method of cross-validation in this scenario is to take as holdout each of the 508 compounds before predictor thinning and compare the predicted values with the experimental data. ITC based results obtained here are comparable to those developed earlier.

Subjects: **Computation (stat.CO)**
Cite as: **arXiv:1305.7285** [stat.CO]
              (or **arXiv:1305.7285v1** [stat.CO] for this version)

Submission history

From: Subhabrata Majumdar [view email]
[v1] Fri, 31 May 2013 02:31:10 GMT (521kb)

Which authors of this paper are endorsers?