Efficient variable selection algorithm adopting variance

Inflated resampling weight vector into model population analysis

Biswanath Mahanty

Assistant Professor

Journal : Chemometrics

Variable iterative space shrinkage approach (VISSA) is an important variable selection algorithm known for its improved accuracy and outcome stability in partial least squares (PLS) regression models. However, time efficiency of VISSA is not very promising. In this work, three strategies to inflate the variance
of resampling weight vector (RWV) have been proposed to accelerate
the space shrinkage in VISSA. The original RWV (i.e., average binary frequency
of variables) is replaced with average of unit normalized regression coefficients
(UNRC), fitness normalized regression coefficients (FNRC), or logarithmically
transformed regression coefficients (LTRC) of selected PLS sub-models.
Although prediction efficiencies for UNRC and FNRC are marginally inferior
to the original binary-weight VISSA, the stability of retained variables and
remarkable improvements in algorithm speed is evident for relatively large size
NIR data set (700 variables). LTRC with moderate degree of RWV variance
inflation is indisputably a better choice. Chimeric algorithm, incorporating
UNRC, FNRC, or LTRC in first round of original VISSA, maintained the model
fitness per se while significantly improving time efficiency. With small dimensional
NIR data set (100 variables), proposed weighting schemes have no additional
advantage over original VISSA implementation.

KEYWORDS
algorithm efficiency, binary matrix sampling, inflated variance, variable selection, VISSA