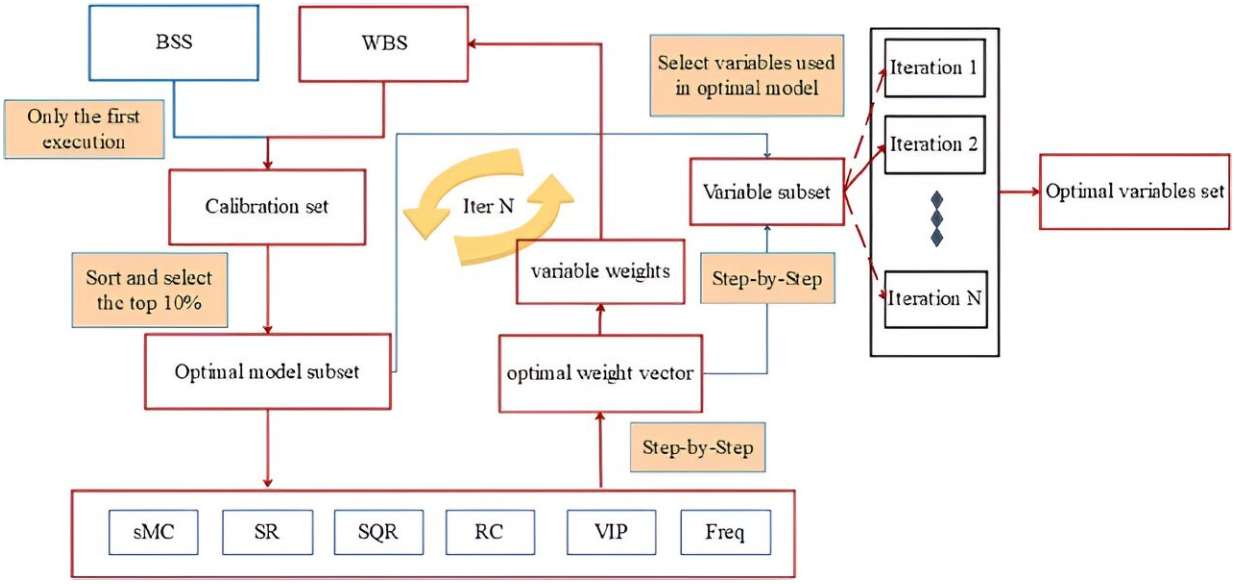
Novel algorithm proposed for efficient selection of variables in chemometrics applications

August 1 2023, by Zhang Nannan



Novel comprehensive variable selection algorithm based on multi-weight vector optimal selection and bootstrapping soft shrinkage. Credit: Zhang Pengfei

A new variable selection method for use in chemometrics applications has recently been proposed by a team of researchers from the Hefei Institutes of Physical Science of the Chinese Academy of Sciences. They call the algorithm multi-weight optimal-bootstrap soft shrinkage (MWO-BOSS). The work is published in *Infrared Physics & Technology*.

Spectral technology, using spectral analysis detection and spectrometers, is widely used in various fields. Extracting feature information from complex, high-dimensional spectral data plays a crucial role in qualitative and quantitative analysis, improving predictive capabilities, and facilitating the development of low-cost, multi-channel spectral detection instruments. However, selecting an optimal wavelength combination from the high-dimensional variable space for building spectral prediction models remains a challenging task due to its NP-hard nature.

To further improve the effectiveness of variable selection, the research team proposed the MWO-BOSS algorithm based on the BOSS algorithm framework.

The algorithm combines six weight vectors—Selectivity Ratio, Variable Importance in Projection, the Frequency Vector, Reciprocal of Residual Variance Vector, Regression Coefficient and Significance Multivariate Correlation—and uses a threshold search strategy to find the optimal weight vector to extract useful information from the spectrum.

The algorithm's performance was tested on publicly available datasets such as maize, soil, and beer, and on several high-performance variable selection algorithms.

The results showed that the algorithm can efficiently select variables and significantly improve the predictive ability of the model.

More information: Pengfei Zhang et al, Novel comprehensive variable selection algorithm based on multi-weight vector optimal selection and bootstrapping soft shrinkage, *Infrared Physics & Technology* (2023). DOI: 10.1016/j.infrared.2023.104800