

Using Improved LRAS/LASSO algorithm for variable selection in high dimensional data with application to a soil carbonate determination problem

Farnaz Pirasteh, 유 준^{1,*}
부경대학교; ¹부경대학교 화학공학과
(jayliu@pknu.ac.kr*)

Carbonate is a key component of soils on calcareous parent material, influencing both chemical and physical soil properties and hence fertility and productivity. A number of different methods of carbonate determination in soils were proposed over the last decades. One of the last works is presented based on combination of FT-IR spectroscopy and XRD; which gives a high dimensional data with many more variables than observations, that makes classical methods useless for variable selection.

In this study, a new variable selection methodology is presented by improving LARS/LASSO algorithm using Stochastic Search Variable Selection in methodology of the Phase Diagram to select most effective variables for soil carbonate determination.

To Compare the efficiency of our method, results from most popular variable selection such as Shrinkage Discriminant Analysis (SDA) or Multiple Hypotheses Testing (MHT) are presented. The RMSEP of this method (4.23) and also other statistical index shows the accuracy of proposed approach in soil carbonate content prediction in our model.