

Development of new data analysis methods : KY-methods

Kohtaro Yuta

k-yuta@insilicodata.com

The In Silico Data Ltd., Japan

1. Preface

The KY-method is a data analysis method featuring multistep re-sampling process.

Currently 6 types of KY-method are developed. There are three types as classification methods of two classes, and three types as multiple regression methods. The KY method achieves an extremely high classification rate in the two class classification method and realizes an extremely high determination coefficient in the multiple regression method.

2. Basic composition of KY-method

In the KY-method, various existing data analysis methods are used as classification method and multiple regression method. Therefore, the KY-method is not a specific data analysis method but a "meta data analysis method" developed in combination with an existing data analysis method. For example, in the two-class classification KY-method, various data analysis methods such as SVM (Support Vector Machine), AdaBoost, NN (Neural Network), Bayesian Analysis and other various discriminant analysis methods are applied. Likewise, in the multiple regression KY-method, it is possible to apply various methods such as linear / nonlinear regression, logistic regression, PLS (Partial Least Squares), etc. in the multiple regression method.

3. Excellent feature of KY-method

The KY-method does not deteriorate the accuracy of data analysis even if a large number of samples are used. Therefore, the KY-method has a powerful and optimum function corresponding to the future big data era.

(1) Sample number free: The KY method has a feature that data analysis accuracy does not drop even if the number of samples is extremely large. This is because sample groups are classified into small groups at individual step, so that the number of samples is relatively small, which makes it possible to avoid degradation of data analysis accuracy.

(2) By the KY-method, classification accuracy and coefficient of determination can be increased: In the KY-method, noise samples are put together and transfer to the next step. For this reason, it is possible to analyze the data without noise samples within each step. As a result, data analysis can be performed without decreasing the classification rate and the determination coefficient.

(3) The KY-method is a meta analysis method to which existing methods are applied as a data analysis method applied at individual stages. This makes it easy to incorporate data analysis methods to be developed in the future, and can always work with the latest methods.

4. Types of KY-method

The KY-method currently has three types as two class classification method and also three types as multiple regression method are developed.

* Two class classification KY-method:

(a) Two model KY-method (b) One model KY-method (c) Model free KY-method

* Multiple regression KY-method:

(a) KY-Fitting with DA (Discriminant analysis) (b) KY-Fitting with no DA (Discriminant analysis) (c) Model free KY-fitting

5. Summary

The KY-method is a "meta analysis method" that can improve the data analysis power possessed by the original methods. In the poster, more details on the KY-method will be discussed.