A Ratemaking of Private Health Insurance using Data Mining Techniques

Many insurance companies use data mining techniques to find insights hidden in their data. In this study, a ratemaking of the private health insurance is carried out through various supervised learning. In the case of private health insurance, although it is necessary to calculate a more detailed rate to prevent adverse selection, various ratemaking methods have not yet been applied in practice.

Currently, rating variables of private health insurance are genders, ages and class rates. In spite of the heterogeneous risk characteristics of private health insurance, the use of only restrictive rating variables can lead to sustained loss ratios and a reduction in the private health insurance market by intensifying adverse selection. Therefore, it is necessary to consider introducing the policyholder’s performance as a rate variable, which can better explain the risk characteristics of each policyholder.

In order to overcome the shortcomings of one-way classification, ratemaking approach using multivariate method such as generalized linear model (GLM) is used. Furthermore, we apply machine learning techniques such as decision trees, ensemble models, MARS and neural network models to ratemaking in this study. We implement through R programming so that insurance practitioners and researchers can try machine learning algorithms.