

Automatically building a formula by machine learning using big data for social science

Yoshiyasu Takefuji, Ph.D.

Professor of Keio University

David Lazer wrote an article entitled "Social Science, today" (1). Many of social science researchers have been currently formulating parameters of collected data with expert knowledge. If the formulated equation manually by researchers is a reasonable, it may be able to produce the rational result. However, more complex system with a large number of parameters cannot be modeled by researchers. Machine learning technologies allows us to automatically build a model of the relationship between complicated parameters using big data. In other words, manually formulating parameters by researchers is not needed anymore. Machine learning technologies include ensemble machine learning based on statistics (Adaboost, Random Forest, Extra Tree, Extra Trees, Gradient Boosting, Bagging, Voting) and deep learning. Deep learning system can be built by deep learning frameworks. Google uses Tensorflow and Keras frameworks, Amazon Mxnet and Gluon frameworks, Microsoft CNTK framework , Facebook Pytorch and Caffe2 frameworks respectively. Now, researchers can build a target system using open source deep learning frameworks and/or open source ensemble machine learning APIs.

References:

1. David Lazer, "Social Science, today," Science 5 Jan 2018, VOL 359 ISSUE 6371, p.42