

Visualization technology for machines to understand Bigdata: Automatic selection of feature vectors and barbarization of data analysis methods

*Tomoyuki Higuchi¹

1. The Institute of Statistical Mathematics

A typical use of sparse modeling is the automatic selection of explanatory variables in a linear regression model. In many cases, explanatory variables are called descriptors, and sets of explanatory variables are called feature vectors. Until now, the method of constructing feature vectors has been left to human intellectual work, and in fact, it is not an overstatement to say that this construction method almost decides the performance such as prediction / discrimination performance. In other words, the feature vector construction method can be said to be a "craftsmanship" of machine learning, and despite machine learning, human judgment is also important for improving the performance most, there is also the aspect of false advertising. In deep learning, by directly inputting raw data, great work has been achieved while omitting the construction work of this feature vector. Using deep learning, users can obtain the best prediction / discrimination performance without suffering from selection and determination of feature vectors. It is exactly why it is said that "Deep learning has released users from the mastery of master's skill. We are accelerating this policy further, and research on end-to-end connected directly by deep network from raw data to final goal, which omitted all preprocessing etc., is also accelerating. Furthermore, by taking this power as an opportunity, if you create a large amount of images and movies by visualizing and imaging the raw data for the time being without any thought, and imaging the state of the various work related to data processing, it is also possible in principle to achieve the object while avoiding the feature vector selection problem. In data analysis so far, careful and appropriate selection of analytical methods according to the characteristics of the raw data and analytical purposes has been a success. From now on, however, from the viewpoint of data analysis, such as all imaging and voice conversion as described above, "barbarous" data analysis that seems to be deterred may become a normal feeling.

Keywords: Visualization technique, Deep learning, End-to-end learning, Translational learning