Solving reproducibility problems is not so easy

Yoshiyasu Takefuji

Christophe Pérignon et al. wrote an article entitled "Certify reproducibility with confidential data" (1). Since the advent of GPU (graphic processing unit) technology, a large number of parallel processing cores can run for machine learning. There are useful machine learning frameworks including Chainer, Keras, Pytorch, and others for building AI machine learning systems. However, many of these frameworks cannot guarantee the reproducibility because of pseudorandom numbers spread in multiple GPUs. Pytorch is one and only one machine learning framework to be able to reproduce the same result in multiple GPUs environment as long as complex pseudorandom numbers are properly initialized. It is not a matter of confidential data or not. Therefore, it is extremely difficult to reproduce the same result in the multiple GPUs environment so that we are not sure how Christophe Pérignon et al. can implement certifying reproducibility with confidential data.

References:
1. Christophe Pérignon et al., Certify reproducibility with confidential data, Science 12 Jul 2019: Vol. 365, Issue 6449, pp. 127-128