Earth Sciences big data analysis using Unsupervised Deep Learning, and challenge to "Earth Search".

*Daisuke Sugiyama[1], Masuo Nakano[1], Daisuke Matsuoka[1], Onishi Ryo[1], Tanaka Yusuke[1], Seiji Tsuboi[1]

1. Japan Agency for Marine-Earth Science and Technology

Recent progress in high performance computer and measurement apparatus has produced huge amount of Earth science "big data", which almost exceeds our ability to process. For the analysis of these "big data", it was common to develop analysis technique for each case. This procedure has a merit in producing desired results but, on the other hand, it does not enable us to handle increasing data. Here we use Variational AutoEncoder, which is one of the model of artificial intelligence techniques, to perform unsupervised learning for large dataset of JAMSTEC's marine-earth science data. Unsupervised learning is a machine learning procedure, which does not provide answered questions and suggestions. Using this model for learning, we try to create an artificial intelligence, which may estimate inherited hidden signatures from cloud, atmospheric pressure, precipitation and ocean current data of JAMSTEC's NICAM model. This artificial intelligence may discern characteristics in data which we may not be able to find and may help us to discover new scientific knowledge, such as unknown laws or phenomena.
We also try to examine if neighborhood search of earth science data is possible or not by using search method of latent variables which the artificial intelligence has acquired through unsupervised learning. We estimate latent variables of past observation or simulated data in advance and keep them as a database of indexes. The we put an Earth status (observation data) of arbitrary space and time as an input and estimate latent variables to perform neighborhood search against these indexes. By using this procedure we may create an artificial intelligence which enables us to search similar status as if we use Google search and we may call this as "Earth search".

Keywords: Deep Learning , Data Mining, Earth Simulator (ES), Search Engine