

Little Data: An Overview of Machine Learning Techniques for Limited Datasets

Misha Klopukh, MPCR Laboratory, Florida Atlantic University

The fields of Data Science, Analytics, and Machine Learning have become colloquially known as "Big Data" due to their intensive use of large datasets. Techniques in the field of Deep Learning in particular rely on the availability of large amounts of labeled example data for a particular task. However, there are many situations, especially in medical research, where the amount of data produced may be small or the data may be poorly labeled, yet the data still cannot be comprehended by humans or processed using simple statistical techniques. Recently, many techniques have been developed to overcome these challenges, including data augmentation, transfer learning, sub-tasking, and forms of "unsupervised" training. This talk is an overview of such techniques and a discussion of their implementation and effectiveness.

Keywords: deep learning, limited datasets, pretraining, transfer learning