# AI and Big Data Convergence

Jon Ander Gómez Adrián

Polytechnic University of Valencia, Spain

Abstract

Most of the techniques from Artificial Intelligence, in particular the ones belonging to machine learning, need as much data as possible to obtain more robust and accurate models trained by means of algorithms that use data samples to adjust the model parameters. As an example, a model based on deep neural networks has millions of parameters (named as weights) whose values are progressively updated by the Error-Backpropagation algorithm that visits iteratively all the samples of the training data set.

Mr. Cukier explained that the researchers from Stanford University used thousands of samples of cancerous breast cells and the patients' survival rates to train a machine-learning model, and defined the goal function of the learning algorithm to identify patterns in the attributes of input data that best correlate with the goal of predicting if a given biopsy will be severely cancerous. The obtained machine-learning model identified eleven attributes that best predict that a biopsy is highly cancerous. What surprised researchers was that only eight of the elevent attributes were previously known by doctors and studied in medical literature. As Stanford researchers included in the experiment all the attributes of input data, without instructing the learning algorithm which ones to use, the outcome as that three of the attributes (or indicators) found by the machine-learning algorithm were not considered as relevant by medical community; pathologists never focused their attention on such indicators. As Mr. Cuckier remarked is that machine learning works because the learning algorithm is fed with lots of data—much more information than any human being could digest in a lifetime and manage at any moment.
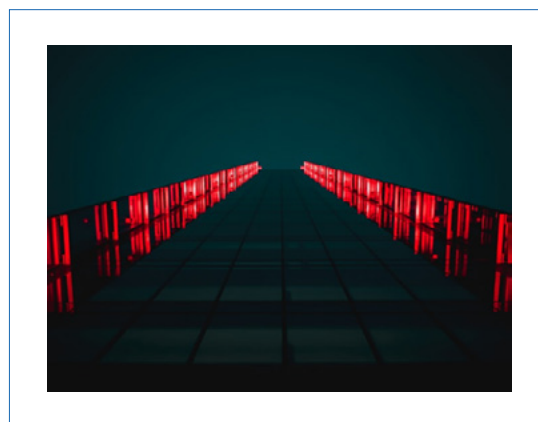
Biography

Dr. Jon Ander Gómez Adrián is Associate Professor of computer science at UPV and is member of the Pattern Recognition and Human Language Technology (PRHLT) research center. His research interests include pattern recognition, deep machine learning and their application to automatic speech recognition, handwritten text recognition, predictive maintenance in smart industry and computer vision.

Publications

1. A.H. Beck et al., "Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated with Survival," Science Translational Medicine, 3.108 (2011) http://stm.sciencemag.org/content/3/108/108ra113.full.pdf

2. https://www.bbvaopenmind.com/wp-content/uploads/2015/04/BBVA-OpenMind-Big-Data-and-the-Future-of-Business-innovation-1.pdf

2nd International Conference on Big Data Analytics and Data Mining | Singapore | July 20, 2020