Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 186 (2017) 261 – 268

Procedia Engineering

www.elsevier.com/locate/procedia

XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

Optimized Deep Learning Framework for water Distribution Data-Driven Modeling

Zheng Yi Wu* and Atiqur Rahman

Bentley Systems, Incorprated, Watertown, CT, USA

Abstract

Deep Learning (DL), unlike conventional Artificial Neural network (ANN), is capable of self-learning data features layer by layer in unsupervised manner and creating a data-driven model with the given dataset. DL has been widely applied to big data analytics, graphics object detection, classification, voice recognition and many other problems. This paper presents an integrated data-driven modelling framework that couples DL with the well-developed evolutionary optimization tool in a scalable and heterogeneous high performance computing paradigm. The integrated framework enables modellers to effectively and efficiently construct a model with a given dataset. It is demonstrated that the framework has wide applicability including but not limited to the simulation, optimization and operation decision of water distribution systems. The paper elaborates the development of the deep learning framework with potential applications of facilitating the data fusion, system simulation and predictive analysis, anomaly detection from the time series data (pressures, flows and consumptions etc.), water usage prediction, construction of a meta-model as a surrogate to the physics-based models (hydraulic and water quality) for water distribution management.

Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems

Keywords: Deep learning; optimization; modelling; data analysis; water distribution Main text

* Tel.: +1-203-8050562; fax: +1-203-7557961.
E-mail address: zheng.wu@bentley.com

1877-7058
Peer-review under responsibility of the organizing committee of the XVIII International Conference on Water Distribution Systems
doi:10.1016/j.proeng.2017.03.240

Nomenclature

h_j	state of the hidden units j
v_i	state of the visible unit i
w_{ij}	weight from neuron i to neuron j
t	time step
σ_i	standard deviation of the Gaussian noise for visible unit i
$Y(t)$	water demand of time step t
$Y(t\text{-}k)$	water consumption of k time steps before current time.

1. Introduction

Data analytics is becoming increasingly important for leveraging the large amount of data collected with ubiquitous infrastructure sensing and/or data logging, which is used for monitoring the various parameters and expected to offer insights in performance and existing conditions of in-service infrastructure systems. Developing a versatile data-driven modelling tool is imperative to extract useful information from big raw data and capture the relationship among the extracted features to facilitate various tasks with the given dataset.

In water distribution system, a conventional SCADA system has been widely adopted to monitor the critical facilities, such as pump stations, storage tanks, and other key control points. While SCADA systems are very well developed, but the data collected from these systems are usually accumulated but no deep analysis is conducted to make good use of the data. Over last decade, Automatic Meter Reading (AMR) and Advanced Metering Infrastructure (AMI) technologies are gaining more and more acceptance in water industry. This is primarily driven by improving operation efficiency. Widely adopted SCADA and emerging AMR/AMI are the backbone of implementing effective monitoring programs, which rapidly produce the datasets in large size (so-called big data). Its value cannot be realized unless useful/actionable information is retrieved or extracted, and used for systematic simulation and optimization modelling to facilitate the decision making, and finally improve the automation control for triggering/communicating with the instruments in the field. However, conducting big data analysis calls for development of generic data-driven modelling tool. In this paper, we report the research and development of optimized machine learning framework that integrate the well–developed Darwin Optimization with deep learning. It is a generic data-driven modelling tool for extracting the intelligence from data, with potential of wide applications in simulation and prediction analysis that serves as the data-driven engine of infrastructure asset performance modelling e.g. water distribution network management.

2. Deep learning

Deep learning a relatively new paradigm of machine learning technique [1]. It allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. DL has been successfully applied to many commercial products including but not limited to object recognition, speech recognition, language translation, and self-driving cars. The main advantage of the deep learning is the automatic feature extraction. The traditional model for pattern recognition is to apply hand-crafted feature extractor to the images before applying the trainable classifier. In contrast, the deep learning systems are able to automatically extract the features by training the system with unlabeled data, and then use those automatically extracted features to classify the image using a trainable classifier.

The deep learning architecture is consisted of multiple layers of non-linear representations. After applying the input to the system, each layer can be individually trained to extract a higher level of abstractions from its inputs. For example, if the input is a large set of unlabelled images, the output of the first hidden layer (after training) can be the features of edges and corners. This output is used to train the next hidden layer to be able to produce group of features, such as body parts, noses, eyes, and so on. The third hidden layer can be trained to represent a higher level of abstractions, and finally recognize the desired image, e.g. a human face. This is the typical process of the Deep Learning system to automatically extract features from the inputs. The advantages of Deep Learning are summarized below: