The idea of ‘opening’ the black-box models is appealing to hydro-environment practitioners, such that attempts have been made to apply or even develop new ML techniques that produce understandable models that can be subject to expert scrutiny (Babovic and Abbott, 1997). An example of such a work is the evolutionary polynomial regression method (Giustolisi and Savic, 2006), which, for example, was used to produce interpretable equations linking various pipe and environmental attributes (e.g., age, material, diameter) to the pipe condition (e.g., the number of pipe failures).

The use of genetic algorithms and other forms of evolutionary computing in hydroscience has been well documented. Needless to say, these super-charged optimisation algorithms have found application in various fields, from the design or rehabilitation planning of urban water infrastructure to optimal reservoir system operation to calibration of water quality models (Nicklow et al., 2009; Maier et al., 2014). The water infrastructure software providers have also included variants of evolutionary algorithms in their own offering to clients, thus bringing powerful optimisation capabilities closer to practice.

INSTEAD OF CONCLUSIONS

Although we are still a long way away from intelligent machines exhibiting human-like intelligence, Artificial Intelligence and Machine Learning are beginning to find application in the water management world, opening a wealth of opportunities and benefits for water management practitioners. For example, AI/ML tools are already being successfully deployed to locate leaks in real water distribution networks, predict domestic and agricultural water demand or to manage energy consumption in a water system. In addition to providing opportunities, proliferation of various data collection systems (sensors and instrumentation), data storage technologies, local and cloud-based computing networks, and data visualisation environments including virtual/augmented reality, together with new AI/ML technologies, they present also some of the greatest challenges for the hydro-environment community.

To truly meet new and ongoing challenges, we need more skilled individuals trained in AI to address the issues and realise the potential benefits of the ‘digital’ technologies, including AI/ML. Data science professionals trained only in AI/ML cannot lead the process of successfully applying those technologies to water management problems, since they do not fully understand the complexity of the water sector and its challenges. Bringing about an AI-enabled water future involves high-end, leading-edge technologies that require a new type of professional trained in both water and AI/ML sciences – Hydroinformaticians! The discipline of Hydroinformatics involves a continuous process of developing and using water data, models and tools to understand our environment, engage all stakeholders, and support decisions that lead to a more sustainable environment. Only with such a group of professionals, who are able to work at the interface of AI/ML, hydro-environment science and engineering, can the full benefits of the Artificial Intelligence in the hydro-environmental practice be achieved and the risks effectively managed.

References

[1] Babovic, V., & Abbott, M. B. (1997). The evolution of equations from hydraulic data Part I: Theory. *Journal of Hydraulic Research*, *35*(3), 397-410.

[2] Berardi, L., Giustolisi, O., Kapelan, Z., & Savic, D. A. (2008). Development of pipe deterioration models for water distribution systems using EPR. *Journal of Hydroinformatics*, *10*(2), 113-126.

[3] Chen, Y., & Han, D. (2016). Big data and hydroinformatics. *Journal of Hydroinformatics*, *18*(4), 599-614.

[4] Giustolisi, O., & Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, *8*(3), 207-222.

[5] Hornby, G., Globus, A., Linden, D. and Lohn, J. (2006). “Automated antenna design with evolutionary algorithms”, in: Proceedings of 2006 American Institute of Aeronautics and Astronautics Conference on Space, San Jose, CA, 2006, pp. 19–21.

[6] Maier, H.R., Kapelan, Z., Kasprzyk, J., Kollat, J., Matott, L.S., Cunha, M.C., Dandy, G.C., Gibbs, M.S., Keedwell, E., Marchi, A., Ostfeld, A., Savic, D., Solomatine, D.P., Vrugt, J.A., Zecchina, A.C., Minsker, B.S., Barbour, E.J., Kuczera, G., Pasha, F., Castelletti, A., Giuliani, M. and Reed, P.M. (2014). Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software, 62, 271-299.*

[7] Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. *Hydrological Sciences Journal*, *41*(3), 399-417.

[8] Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models: Literature review. *Water*, *10*(11), 1536.

[9] Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minsker, B., Ostfeld, A., Singh, A. and Zechman, E. (2009). State of the art for genetic algorithms and beyond in water resources planning and management. *Journal of Water Resources Planning and Management*, *136*(4), 412-432.

[10] Sayers, W., Savic, D., & Kapelan, Z. (2019). Performance of LEMMO with artificial neural networks for water systems optimisation. *Urban Water Journal*, 1-12.

[11] Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*(11), 8558-8593.

[12] Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3-22.

[13] Vojinovic, Z., Abebe, Y.A., Ranasinghe, R., Vacher, A., Martens, P., Mandl, D.J., Frye, S.W., Van Ettinger, E. and De Zeeuw, R. (2013). A machine learning approach for estimation of shallow water depths from optical satellite images and sonar measurements. *Journal of Hydroinformatics*, *15*(4), 1408-1424.

IAHR Global Secretariats

Madrid Office
Paseo Bajo Virgen del Puerto 3
28005 Madrid
SPAIN
tel + 34 91 335 79 08
fax + 34 91 335 79 35

Beijing Office
A-1 Fuxing Road, Haidian District
100038 Beijing
CHINA
tel + 86 10 6878 1128
fax + 86 10 6878 1890

Editors
Silke Wieprecht
Universität Stuttgart, Germany
silke.wieprecht@iws.uni-stuttgart.de
Angelos Findikakis, Bechtel, USA
anfindik@bechtel.com

ISSN 2664-5637

1 2019

IAHR WHITE PAPERS

ARTIFICIAL INTELLIGENCE
HOW CAN WATER PLANNING AND MANAGEMENT BENEFIT FROM IT?

By Dragan Savic

IAHR AIRH

Hosted by
Spain Water and IWHR, China

To catalyse thinking, inspire debate and better apply scientific knowledge to global water problems, the IAHR White Papers seek to reveal complex and emerging issues in Hydro-Environment and Engineering Research. They are written for researchers, engineers, policy-makers and all those who are interested in the latest for a better water future.

Supported by suez IDP

WHAT IS ARTIFICIAL INTELLIGENCE AND HOW CAN WATER PLANNING AND MANAGEMENT BENEFIT FROM IT?

BY DRAGAN SAVIC

Lately, it appears that our society has become fixated on the topic of Artificial Intelligence (AI), with opinions often ranging from one extreme to the other – either how AI could solve a range of current and future world problems, or how it could potentially be very dangerous to humankind. But do we really understand what AI is, how it relates to human intelligence and how/where it most likely could be deployed by the hydro-environment community for the betterment of the environment and advancement of society? This white paper addresses some of these questions and provides a brief introduction to the topic of AI and Machine Learning, together with some example applications in water management practice.

HUMAN INTELLIGENCE VS. ARTIFICIAL INTELLIGENCE

The key characteristics of human intelligence is the ability to learn from experience, or it involves intelligent thinking. The question then is whether machines can be made to carry out intelligent thinking similar to humans. Although nowadays Artificial Intelligence techniques have advanced to the point that, for example, they can beat one of the world's strongest players in the game Go[1], outperform medical professionals in diagnosing deadly diseases[2] or make self-driving cars possible; the general AI goal of thinking machines still seems a long way away. If that is the case, what are the basics of AI, what can it do and how can water professionals take benefit from it?

AI BASICS

If we accept that the key objective of AI technologies is to enable learning (from data), e.g., to develop a model to detect a disease in patients, to recommend products to buyers on an online shopping site, or to predict whether an applicant will be able to repay a loan to a bank; in water management, that would equate to, for example, being able to predict the risk of flooding beyond an acceptable socio-economic threshold, to forecast demand in a water distribution system, or to estimate sediment transport rates in a river. The key point here is that AI can be considered a way of creating useful models or methods to perform a complex task normally carried out by humans.

Another important feature of AI is that when creating a model, it normally uses an algorithm. However, AI models largely employ the so-called "black-box" metaphor, which implies that an AI-created model doesn't allow easy scrutiny of its internal workings. For some people, this represents a major hurdle in applying AI techniques to real-world problems. The other important feature of AI is that the process of creating a model is often automated, i.e., the user does not need to assume the form of the model, thus it is commonly referred to as machine learning (ML).

MACHINE LEARNING

Although there are many definitions of Machine Learning, I prefer a simple one. For example, ML can be defined as a group of algorithms that can create a model based on data with the goal of making predictions or taking actions to optimise a system. Let us describe this in simple terms by considering an analogy with well-known linear regression. If we wanted to predict house prices in an area based on the historical data of the sale price and the living area (square meters) of the houses, we would start by collecting the data of previous house sales. Once we plot the data points of the sale price against the living area and assuming the relationship appears to be linear (the simplest case), it is then very easy to calculate the regression line through the data points. Congratulations, you have just performed a simple ML exercise! We have used an algorithm (e.g., least squares) to 'learn', or in an AI/ML speak to train a model, which is in the form of a regression equation. The model can then be used to predict a sale price for a house that wasn't in our historical data set. These basic steps are also performed by ML algorithms when creating models for much more complex processes. The difference is that normally for ML to learn a relationship based on a data set, the type of relationship (e.g., linear or non-linear) need not be known to the user in advance.

Based on the type of processes to be modelled and data sets available, there is a large number of ML algorithms available that can be used to develop a model. The most well-known methods are artificial neural networks (ANN), which use a biological metaphor to mimic the connectivity and functioning of a human brain (like the neurons in real brains and the way they 'chatter' via electro-chemical processes) while predicting an outcome based on a number of inputs.

Apart from prediction, ML algorithms can also be used for classification tasks where instead of predicting a numerical (or continuous) value they predict a categorical label (or a discrete value). A form of ANN that is experiencing fast-growing popularity in classification is the so-called deep learning, which uses ANN to perform learning tasks directly from images (e.g., image classification), text (e.g., for natural language processing) or sound (e.g., speech recognition). An example of a classification application would be when diagnosing patients based on a large number of scan images (big data) and dividing them into two groups; those with and those without the disease. Mobile phone virtual assistants, such as Amazon's Alexa, Google Assistant or Apple Siri, are examples of the AI technology that understands natural language voice commands and completes tasks for a user.

1) https://www.bbc.com/news/technology-35810133
2) https://www.docwirenews.com/docwire-pick/deep-learning-ai-outperforms-pathologists-in-diagnosing-cancer/

Professor Dragan Savic FREng is the CEO of KWR Water Research Institute, the Dutch drinking water companies' collective research organisation. He is also the UK's first Professor of Hydroinformatics, having held this position at the University of Exeter since 2001. His research interests cover the interdisciplinary field of Hydroinformatics, which transcends traditional boundaries of water/environmental science and engineering, informatics/computer science (including Artificial Intelligence, data mining and optimisation techniques) and environmental engineering. Professor Savic has served as both the Chair of the IAHR/IWA Joint Committee on Hydroinformatics and as the Editor-in-Chief of the Journal of Hydroinformatics.

BIG DATA

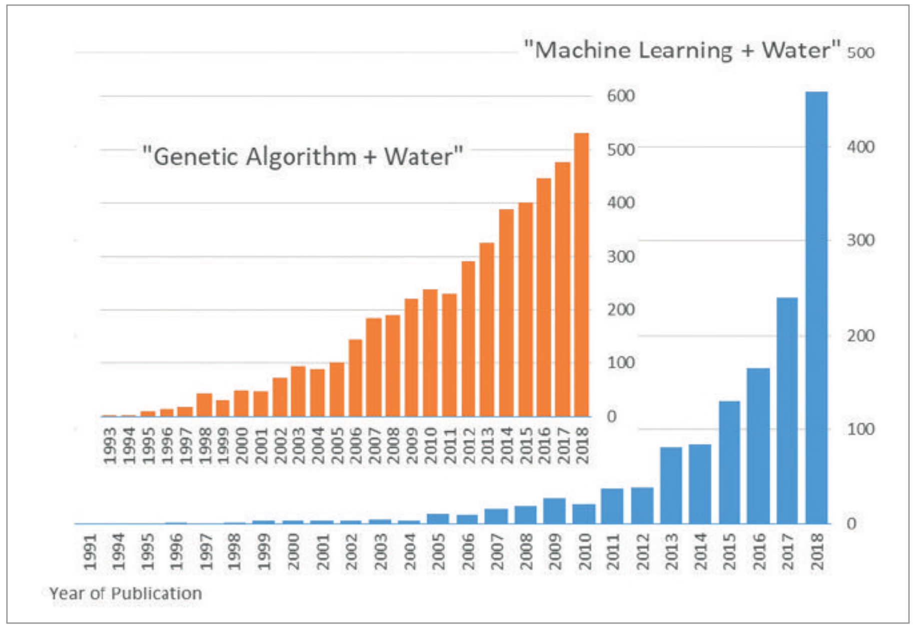
The rate at which we generate extremely large datasets every day (in terms of volume, variety, and velocity or the "3 Vs of big data"), is staggering due to the growth of the mobile telephony, the Internet of Things (IoT) and satellite earth observation technology, to name but a few of the sources. That means that the total amount of data is expected to reach 44 zettabytes (or 10^{21} bytes) by 2020, or in other words, that the volume of data will likely exceed 40 times the number of stars in the observable universe[3]. Therefore, so-called "big data analytics" will be another area where ML can be employed with the potential to change our lives. However, so far big data analytics has found little application in hydro-environmental research and practice. The main reason for this is that due to the associated costs we do not normally collect big data from our water systems. That is slowly changing with the integration of data coming from various other sources (e.g., remote sensing, IoT, citizen science). An example of the potential application of big data analytics to precipitation estimation envisages data fusion from remote sensing, weather radar, rain gauge and numerical weather modelling, which could be used to generate better estimate than those from single sources (Chen and Han, 2016).

EVOLUTIONARY COMPUTING

Another form of AI algorithms which are of interest to hydro-environmental practitioners, the so-called evolutionary computing, has drawn inspiration from biological evolution. These algorithms perform a different type of AI learning enabling machines to make autonomous decisions, adapt to a changing environment or find non-obvious solutions to complex and 'wicked' problems. The most common form of evolutionary computing is a genetic algorithm, which like a 'mad scientist' creates a huge number of potential decisions (a population of solutions) that are then changed by a sequence of DNA-like operations (mutation and crossover). Finally, by using a preferential selection of better-performing candidates (akin to natural selection), the algorithm arrives at the best solution, e.g., the best design of a water distribution system or the best calibrated rainfall-runoff model, out of a large number of potential solutions. This type of AI technology performs these operations at super-high speed, enabling trillions of solutions to be tested, such that they can solve problems previously intractable using classical optimisation (operations research) tools. As an example, NASA used a genetic algorithm to design a new space antenna, which had to meet a large set of difficult requirements. The outcome was the evolved antenna that in comparison with those developed with traditional design techniques, had several advantages with respect to power consumption, fabrication time, complexity and performance (Hornby et al., 2006).

AI IN HYDRO-ENVIRONMENT RESEARCH AND PRACTICE

Hydro-environment research and practice has already benefited from the application of AI techniques (Solomatine and Ostfeld, 2008; Nicklow et al., 2009; Maier et al., 2014). The figure below shows the increasing trend in the number of publications found when searching the Web of Science[4] using keywords "Machine Learning" or "Genetic Algorithm" with "Water". Initial applications of ML techniques have been centred around using a single algorithm (most often an ANN) in modelling complex physical processes, i.e., rainfall-runoff transformation (Minns and Hall, 1996). More recently, a survey of ML methods for flood prediction indicated a trend of moving to ensemble methods and hybridized approaches where two or more ML techniques are used to predict the output variable (Mosavi et al., 2018). Widespread sensor deployment and availability of remote sensing data also offer new opportunities to hydro-environment practitioners. They can help identify better model parameters, integrate ML with traditional mechanistic (physics-based) models (Vojinovic et al., 2013) or even replace them when high speed of model execution is required (Sayers et al., 2019). The use of deep learning methods in hydro-environmental practice is in a relatively early stage of development, however, the greater availability of data (and particularly big data through remote sensing) provides further opportunities for these type of AI methods (Shen, 2018).



3) https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/
4) https://www.webofknowledge.com/