International Journal of Latest Engineering and Management Research (IJLEMR)
ISSN: 2455-4847
www.ijlemr.com || Volume 05 - Issue 07 || July 2020 || PP. 35-39

Challenges of Computational Data Modelling in Petroleum Engineering

Nelson F. F. Ebecken
(Civil Engineering, COPPE, Federal University of Rio de Janeiro, Brazil)

Abstract: The study of natural phenomena or major environmental problems usually requires the analysis of interdisciplinary data acquired from increasingly powerful sensors to achieve sustainability. This almost always results in complex interactions of hard mathematical modeling and the adoption of holistic system views. In many cases, the initial conditions are not known, the domain is not well defined and the constitutive parameters are heterogeneous generating large uncertainties. For this reason, knowledge-based models evolved from huge amounts of data have generated enormous interest to understand and solve such phenomena that do not always find suitable mathematical laws for their representation. So, the purpose of this work is to briefly discuss the technical and challenges on data processing and engineering and give an idea on the scientific development and future direction of this domain. The question is: What are the impacts of the application of new technologies such as big data, deep learning and artificial intelligence in general and specialized activities such as petroleum engineering?
Keywords: artificial intelligence, data models, deep learning, digital rock, petroleum engineering

I. INTRODUCTION

The systems modeling goal is to find and formulate laws that govern phenomena in a mathematical and precise way. However, it is recognized that such perfect descriptions are not always possible. Incomplete and imprecise knowledge, qualitative observations and great heterogeneity of many interacting agents usually cause uncertainties when modeling complex phenomena of the ever-changing nature or in the business world. It is recognized that knowledge is kept by experts or stored in data. For organizations, the most important is to understand the great dynamism of the natural systems or the competitive environment in which they are inserted to make decisions about the sustainability and the responsiveness of their decisions. Therefore, information, knowledge and data are preemptive [1].

II. SENSORS AND DATA

Talking about smart phones, tablets, notebooks, and connected objects, one need to understand that these "things" are full of sensors. A forecast shows it could exceed one trillion as early as the early 2020s according to estimates from various companies and industries attentive in these flows. The data is observed, stored and released dynamically, which requires that science be online. Thus, to be able to make accurate models and update themselves in dynamic competitive environments, it is increasingly necessary to develop methodologies to generate current and adaptive forecasts for decision making.

Digital twin, Fig.1, is a digital replica of any industrial facility that is used to monitor, analyze and improve overall performance. A Digital Twin uses data from sensors installed on physical systems to represent their near real-time status, working condition or position. This modelling technology allows us to see what is happening inside the system without having to be able to get inside the system. It forms a critical step in the information value chain without which it is often impossible to get from raw data to insight, and therefore to value.

Figure 1 - Digital Twin (www.engineering.com)

The digital twin concept is being extended far beyond this simple example to encompass large and complicated assets such as petroleum refineries, automobile assembly plants, distribution centers, wind farms, large scale construction projects, etc. In each case, sensors and other devices capture data that is fed into the simulation model to provide a detailed understanding of the current state of the asset. Machine learning algorithms running on the edge or in the cloud access information from the physical asset and the simulation model to optimize the performance of the asset by scheduling maintenance, setting control points, sending alerts to operators, providing reports to management, etc. The information generated by the simulation model can also potentially be communicated by overlaying the flow contours of the fluid flow inside the pump onto an image of the pump so operating personnel can quickly understand and diagnose problems. As the Internet of Things grows, Digital Twins will become a standard tool for Data Scientists and Engineers wishing to use all this new data to automatically understand and respond to what is going on in the real world. On the other hand, the cost reduction of electronic sensors applied in environmental monitoring allowed data to be collected and used to build real time models. These models use a continuous flow of data and there is no control over the order of arrival of each element to be processed.

The Array of Things (AoT) is an urban sensing project, a network of interactive, modular sensor boxes that will be installed around Chicago to collect real-time data on the city's environment, infrastructure, and activity for research and public use. AoT will essentially serve as a "fitness tracker" for the city, measuring factors that impact livability in Chicago such as climate, air quality and noise.

Flow data has unlimited size, once processed, an element is usually discarded. It is not recovered unless it is stored in memory, which is usually small for the size of the data received. Queries about these flows need to be processed in real time (for a real-world event or because it is expensive to store the data).

Equipped with wireless instrument control and transmission of datafiles, a CytoBuoy (Fig. 2) can be used to conduct extended and/ or high frequency time series of phytoplankton distribution and abundance on fixed locations in-situ in the field. This includes also the monitoring and early warning for target species (www.cytobuoy.com).

These methodologies should be able to extract knowledge in huge amounts of structured and unstructured data from in situ monitoring, regulations and web, and add it to existing knowledge. In the model generation, one can rely on statistical methods, machine learning and computational methods inspired in nature for engineering application. These models have great robustness because they are noise tolerant and can be coupled to other models providing hybrid solutions.

Figure 2 - The CytoBuoy launched in Rio de Janeiro resurgence area

III. SCIENTIFIC AND DATA INTENSIVE COMPUTING

It should be reappraised the use of highperformance computers and database technology which has caused significant changes in the development of design techniques and programming of algorithms for engineering problem solving. In the case of scientific computation of large systems, the different ways in which numerical methods are designed or adapted stimulated research in technical and applied aspects. New, scalable computing architectures (Big Data, Deep Learning, NoSQL databases) have boosted database technology, allowing simulation of real problems with great consistency and accuracy. Big Data refers to data that is too big to fit on a single server, too unstructured to fit into a row-and-column database, or too continuously flowing to fit into a static data warehouse.

Both represent computing strategies, but the best and most efficient one depends on the situation. Large-scale data processing in the distributed system is very challenging in many concepts of the system performance for reliability. Typically, the MPI (Message Passing Interface) supports a more flexible communication method than MapReduce (asynchronous versus synchronous). While MPI moves the data during communication, MapReduce uses the concept of "data locality", making the transmission between CPU and disk possible once, this task, cannot be performed in MPI because it requires that data "be processed" should fit in memory ([2],[3]).

Deep learning algorithms attempt to learn multiple levels of representation of increasing complexity/abstraction. Before 2006 training deep architectures was unsuccessful. What has changed? New methods for unsupervised pre-training have been developed, more efficient parameter estimation methods, better understanding of model regularization, etc… ([4],[5],[6],[7],[8]).

Recently to model a complex phenomenon or engineering process we can have two types of approach, or combine the two in a hybrid model:

1. **The usual Engineering Computing approach**
- Mathematical laws governing the phenomenon
- Differential equations
- Simplifications: homogeneity of properties, boundary conditions, geometry, ...
- Solution of large systems of algebraic equations, eigenvalues, ...
- Scalability: parallel computing, decomposition, iterative algorithms, ...

2. **The Data Intensive Computing approach**
- No mathematical law
- Knowledge extracted from large masses of data and expert systems
- Deep Learning
- Statistics, Machine Learning, Database, Nature Inspired Methods, ...
- Unstructured data, Data Stream
- Scalability: data storage, NoSQL databases, ...

IV. RESEARCH CHALLENGES AND DEVELOPMENT

One application for advanced imaging with Tech Futures' scanner is digital rocks, which is an emerging concept in characterizing reservoirs by obtaining physical properties directly from images of rocks. The technology combines CT (Computed Tomography) at different scales (mm to micron resolution) with X-ray and scanning-electron microscopy (SEM) to create images of the internal structure of the pore spaces as well as minerals and organic matter within them. Such data can be very useful in studying pore architecture and physics of the processes happening during enhanced oil recovery. It is especially important for unconventional resources like shale, carbonate, tight gas, and coal seam gas. One of the most interesting aspects at this time is characterizing bitumen-bearing carbonates, especially fracture characterization (Fig. 3). The use of CT to image and quantify materials greatly assists scientists in determining potential sites for commercial oil extraction in unconventional materials such as carbonates and oil sands ([9],[10],[11],[12]).

Big data, deep learning and artificial intelligence in the petroleum industry produce excellent solutions to perform:

- Reservoir Characterization Using Deep Learning Neural Networks
- Reservoir Uncertainty Analysis: The Trends from Probability to Algorithms and Machine Learning
- Deep-learning methods for predicting permeability from 2D/3D binary-segmented images
- Improving Reservoir Management through Big Data Technologies
- Machine Learning Leads Cost Effective Intelligent Fluid Design: Fluid Engineering Perspective
- Production Forecasting of Petroleum Reservoir applying Deep Learning

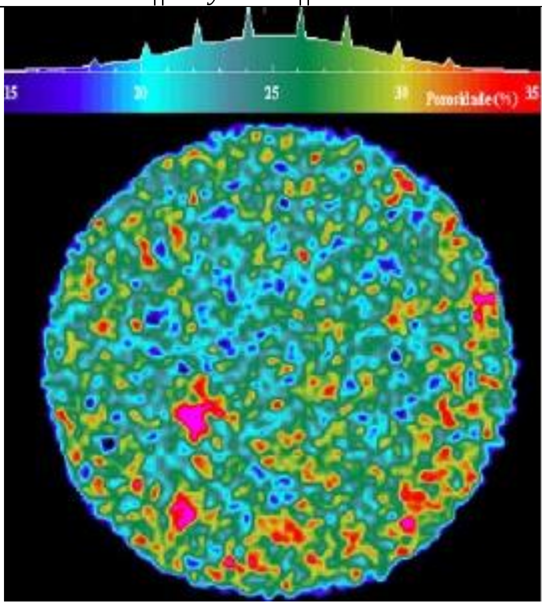

Figure 3 - Porosity Distribution Map of a Sample of Sandstone.

Since CT scanners became widely available, CT results for geological materials have been successfully integrated with other analytical data in various geological disciplines. It is expected that in the near future, CT will be used more and more as a routine research tool. In future studies, CT([13],[14],[15],[16],[17],[18],[19])data will be obtained for much greater numbers of samples, allowing a statistical evaluation of a large body of CT results.

V. CONCLUSION

Artificial Intelligence and Deep Learning are being used to solve some of the world's biggest problems and is finding application in autonomous driving, marketing and advertising, health and medicine, manufacturing, multimedia and entertainment, financial services, and so much more. This is made possible by incredible advances in a wide range of technologies, from computation to interconnect to storage, and innovations in software libraries, frameworks, and resource management tools. While there are many critical challenges, an open technology approach provides significant advantages([20],[21],[22],[23],[24],[25]).

The use of large volumes of data will become a fundamental competition and growth for individual companies. From the point of view of competitiveness and value-capture potential, all companies need to fit large volumes of data. In most industries, competitors and new entrants will use strategies to innovate, compete and capture value, based on information of depth and in real time. Parallel processing, clustering, virtualization, large environments, high connectivity and cloud computing, as well as other of flexible resources, are enabling organizations to take advantage of the Big Data and big data analytics.

The opportunities of the future will be greater and more difficult to be solved, to the point of challenging our capacity for imagination. Quantum computers will introduce a new era in computing. Quantum systems will help us to find new ways to model financial data and isolate key global risk factors to make better investments. And they may make facets of artificial intelligence such as machine learning much more powerful. The response and decision making will be very demanding with our ability to respond. The future lies in innovation. In the convergence of Bio-Cogno-Info-Nano ([1]). In the mining of scientific trends for innovation.

VI. ACKNOWLEDGEMENTS

This work was supported by the Coordination for the Improvement of Higher Education Personnel (CAPES), by CNPq, (National Council for Scientific and Technological Development) and FAPERJ (Research Support Foundation of the State of Rio de Janeiro).

REFERENCES

[1] Ebecken, N.F.F., A Process for Innovation in a Collaborative Network, Journal of Information & Knowledge Management (10), 2011, 1-10.

[2] Gropp, W., Lusk, E., Skjellum, A., Using MPI (2nd ed.): portable parallel programming with the message-passing interface (MIT Press, Cambridge, MA, USA,1994).

[3] Lämmel, R., "Google's Map Reduce programming model — Revisited". Science of Computer Programming. 70, 2008, 1–30.

[4] Hinton, G. E., & Salakhutdinov, R. R., Reducing the dimensionality of data with neural networks, Science, 313, 2006, 504-507.

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks In Advances in neural information processing systems, 2012. 1097-1105.

[6] Hochreiter, S., & Schmidhuber, J., Long short-term memory. Neural computation, (9),1997, 1735-1780.

[7] Mnih, Volodymyr, et al., Human-level control through deep reinforcement learning.Nature 518.7540, 2015, 529-533.

[8] Silver, D., Deep Reinforcement Learning. Retrieve,2015, from http://www.iclr.cc/lib/exe/fetch.php?media=iclr2015:silver-iclr2015.pdf

[9] Winkler, B., Knorr, K., Kahle, A., Vontobel, P., Lehmann, E., Hennion, B. &Bayon, G., Neutron imaging and neutron tomography as non-destructive tools to study bulk-rock samples. European Journal of Mineralogy, (14), 2002,349-354.

[10] Van Geet, M., Swennen, R. &Wevers, M. 2000. Quantitative analysis of reservoir rocks by micro-focus X-ray computerized tomography. Sedimentary Geology, (132), 2000,25-36.

[11] Mees, F., Swennen, R., Van Geet, M. & Jacobs, P. (eds) 2003. Applications of X-ray Computed Tomography in the Geosciences. Geological Society, London, Special Publications, (215),2003, 1-6.

[12] Campos, E. F., Lopes, R.T.,Applications of Computerized X-Ray Tomography in Petroleum Recuperation,2002.

[13] https://www.ipen.br/biblioteca/cd/inac/2002/ENAN/E04/E04_130.PDFM Ozaki, Y. Adachi, Y. Iwahori, and N. Ishii, Application of fuzzy theory to writer recognition of Chinese characters, *International Journal of Modelling and Simulation, 18(2),* 1998, 112-116.

[14] Badrinarayanan, V., Kendall, A.,Cipolla, R,SegNet: A Deep ConvolutionalEncoder-Decoder Architecture for Image Segmentation arXiv:1511.00561.2015.

[15] He, K.; Zhang, X.; Ren, S; Sun, J., Deep Residual Learning for Image Recognition.arXiv:1512.03385, 2015.

[16] Krizhevsky A.,Sutskever, I., Hinton, G. E., ImageNet Classification withDeep Convolutional Neural Networks. Advances in Neural InformationProcessing Systems, (25),2012,1097-1105.

[17] Simonyan, K., Zisserman, A., Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556, 2014.

[18] Masci, J., Meier, U.,Cire¸ D.,Schmidhuber, J., Stacked convolutionalautoencoders for hierarchical feature extraction. In: 21th InternationalConferenceon Artificial Neural Networks (ICANN 2011). Proceedings.2011. p. 52–59.

[19] Vincent, P., Larochelle H.,Bengio Y.,Mangazol, P.,-A. Extracting andComposing Robust Features with Denoising Autoencoders, 25TH International Conferenceon Machine Learning (ICML 2008), 2008, 1096-1103.

[20] Kaskurewicz, E.,Bhaya, A., Ebecken, N. F. F., A fault detection and diagnosismodule for oil plants in offshore platforms. Expert Systems with Applications, (12), 2, 1997, 189-194.

[21] De Domenico, M. et al.,Mathematical formulation of multilayer networks. Phys. Rev. X 3, 2013,041022.

[22] Pereira, G.C., Andrade, L.P., Espínola, R.P., Ebecken, N.F.F. 2014. Structural Analysis and Static Simulation of Coastal Planktonic Networks, Journal of Intelligent Learning Systems and Applications, 6, 2014, 113-124.

[23] Learning Apache Spark with Python, Wenqiang Feng, February 05, 2020, https://runawayhorse001.github.io/LearningApacheSpark/pyspark.pdf

[24] Kivela, M. et al., 2014. Multilayer networks. J. Complex Networks 2, 2014, 203–271.

[25] Multilayer Networks: Structure and Function, Ginestra Bianconi, DOI:10.1093/oso/9780198753919.001.0001,2018.