The Physics of Data - Part III

The Physics of Data - Part III | Alfonso R. Reyes

Data goes far beyond statistics, data science, machine learning, and AI

In this post and the next, I will make the following points :

- that real world data - as imperfect as it is - obeys the laws of physics;
- data is misunderstood because it's being uniquely analyzed under the paradigm of statistics, Data Science, and Machine Learning;
- usable data is unnecessarily being discarded because lack of understanding of the Physics of interconnected dynamical systems;
- unfortunately, for the detriment of AI, scientists, physicists, Computational Physics have practically no presence or influence in the development and deployment of machine learning algorithms and artificial intelligence agents;
- all the advanced calculus used in the **data universe** ends at *derivatives* and *auto-differentiation*; Artificial Intelligence has reached a peak, and will remain as it is if it doesn't break off the paradigm;
- hybrid Models, where they weight the results of ML and physics models, sideline the problem and do not provide scientific explanation.

The missing piece

You feel that something is amiss when data-driven machine learning projects don't yield the results that you were expecting to - or as it was nicely shown in slides at the sales pitch before you signed the deal. At some point during the past decade, we got too much carried away by the claimed successes of data science, machine learning, and more recently, by an incipient artificial intelligence.

Pure datadriven approaches or projects are not delivering. Mainly because they are proudly and dangerously ignoring the Physics. Data is an expression of the natural world, be that a process, a system, or the environment itself. Data is awaiting to be formulated and modeled with differential equations as any other dynamic system.

Data without Physics

Data without its physical context is meaningless; data without their own differential equations will be impossible to predict. Unless, of course, you have huge amounts of it, of what they call "big data", generated by billions of users, millions of digital devices; and quadrillions of clicks a day; what we

know by the internet search engines, interactions with web services, transactions, messages, and email.

But that “big data” world doesn't resemble much the real world that produces goods, materials, and energy. Does it?

The only way of building reliable, repeatable, and accurate predictions is by understanding the **physics of the data**.

Remember, scientists build and gain understanding by delineating, first the theory, and then formulating the differential equations of natural occurring phenomena. This layer of physics and math makes the system explainable, repeatable, and reproducible.

hashtags:: Petroleum Engineering SPE Differential Equations SciML Physics Of Data

Alfonso R. Reyes • You
VP Artificial Intelligence Engineering - Energy Division
10mo • Edited •

The Physics of Data - Part III

Data goes far beyond statistics, data science, machine learning, and AI

In this post and the next, I will make the following points :

* that real world **#data** - as imperfect as it is - obeys the laws of physics;

* data is misunderstood because it's being uniquely analyzed under the paradigm of **#statistics**, **#DataScience**, and **#MachineLearning**;

* usable data is unnecessarily being discarded because lack of understanding of the **#physics** of interconnected dynamical systems;

* unfortunately, for the detriment of AI,
scientists, **#physicists**, **#ComputationalPhysics** have practically no presence or influence in the development and deployment of machine learning **#algorithms** and artificial intelligence **#agents**;

* all the advanced **#calculus** used in the **data universe** ends at *derivatives* and *auto-differentiation*; **#artificialIntelligence** has reached a peak, and will remain as it is if it doesn't break off the paradigm;

* **#hybridModels**, where they weight the results of **#ML** and physics models, sideline the problem and do not provide scientific explanation.

The missing piece
You feel that something is amiss when data-driven machine learning projects don't yield the results that you were expecting to - or as it was nicely shown in slides at the sales pitch before you signed the deal. At some point during the past decade, we got too much carried away by the claimed successes of data science, machine learning, and more recently, by an incipient artificial intelligence.

Pure **#datadriven** approaches or projects are not delivering. Mainly because they are proudly and dangerously ignoring the **#physics**. Data is an expression of the natural world, be that a process, a system, or the environment itself. Data is awaiting to be formulated and modeled with differential equations as any other dynamic system.

Data without Physics
Data without its physical context is meaningless; data without their own

differential equations will be impossible to predict. Unless, of course, you have huge amounts of it, of what they call "big data", generated by billions of users, millions of digital devices; and quadrillions of clicks a day; what we know by the internet search engines, interactions with web services, transactions, messages, and email.

But that “big data” world doesn't resemble much the real world that produces goods, materials, and #energy. Does it?

The only way of building reliable, repeatable, and accurate predictions is by understanding the **physics of the data**.

Remember, scientists build and gain understanding by delineating, first the theory, and then formulating the differential equations of natural occurring phenomena. This layer of physics and math makes the system explainable, repeatable, and reproducible.

#petroleumEngineering #spe #DiffEq #SciML #PhysicsOfData