The Physics of Data. Part VII

The Physics of Data. | Alfonso R. Reyes

What's right and wrong with Data Science & Machine Learning in the real world

Data Science was never meant to be a catchall solution for all problems with data. I am sure if any of us worked with data long enough -applying the scientific method -, has found out the profound limitations of this relatively new branch of statistics.

Don't get me wrong. Data Science is a must-have toolbox of statistics, Computer Science, and Data Engineering. Without it cannot pass to next stage of machine learning modeling. There are so many discoveries and benefits to gain in business just from applying Data Science.

But we have to recognize that data science as a first step of something bigger: understanding what the data is trying to tell us about physical phenomena. We clean the data; we explore it; we sort it; we make it tidy; we switch dimensions. We want to discover hidden issues that are not visible at plain sight. We wish that data science should be enough; it is a screenshot in time, a static time frame.

With machine learning, we want now to extrapolate that frame behavior, combined with real time data, and be able to make predictions about what that data can tell us about the future. Most of times Machine Learning will work but only with **big** and **perfect data**, which is produced in the ideal world of internet, a.k.a. "Big Data".

Real world data is more complicated than Big Data. Real world data is made of innumerable streams of data, generated by field sensors from many Dynamical Systems. And there is where Machine Learning starts to show its deficiencies. Limitations rather.

The reason is simple: we cannot fix or model everything with just math and statistics. It is impossible to come up with good prediction models of real world Dynamical Systems by ignoring the physics & their governing Differential Equations.

What are you going to do with real world data when it is small and imperfect?

Meaning, continuous training is not an option. You have to come up with something better: a descriptive differential equation for each of the data threads, so your model is able to predict the future reliably, from the Physics perspective. What I mean by **data threads** is the different data streams from sensors such as pressure, temperatures, flow rates, etc.

From there, the acute need of changing course to a more scientific and engineered approach via Physics Of Data, where physics takes the reins again by treating the data threads as physical and natural phenomena with their own governing equations. Our mission is now to find the parameters, identify the hidden state variables, and set up the equations that match the calibration data.

PS. Since most of the people understand Artificial Intelligence as a synonym of machine learning, please, assume AI and ML, when I say machine learning. Sadly, that is the unfortunate situation we are in: faking AI models without physics.

hashtags:: SPE SciML Differential Equations Petroleum Engineering

Alfonso R. Reyes • You
VP Artificial Intelligence Engineering - Energy Division
9mo • Edited •

The Physics of Data. Part VII
What's right and wrong with Data Science & Machine Learning in the real world

Data Science was never meant to be a catchall solution for all problems with data. I am sure if any of us worked with data long enough - applying the scientific method -, has found out the profound limitations of this relatively new branch of #statistics.

Don't get me wrong. Data Science is a must-have toolbox of #statistics, #ComputerScience, and #DataEngineering. Without it cannot pass to next stage of machine learning #modeling. There are so many discoveries and benefits to gain in business just from applying #DataScience.

But we have to recognize that data science as a first step of something bigger: understanding what the data is trying to tell us about physical #phenomena. We clean the data; we explore it; we sort it; we make it tidy; we switch dimensions. We want to discover hidden issues that are not visible at plain sight. We wish that data science should be enough; it is a screenshot in time, a static time frame.

With machine learning, we want now to extrapolate that frame behavior, combined with real time data, and be able to make predictions about what that data can tell us about the future. Most of times #MachineLearning will work but only with **big** and **perfect data**, which is produced in the ideal world of internet, a.k.a. "Big Data".

Real world data is more complicated than #BigData. Real world data is made of innumerable streams of data, generated by field sensors from many #DynamicalSystems. And there is where #MachineLearning starts to show its deficiencies. Limitations rather.

The reason is simple: we cannot fix or model everything with just math and statistics. It is impossible to come up with good #prediction models of real world #DynamicalSystems by ignoring the physics & their governing #DifferentialEquations.

What are you going to do with real world data when it is small and imperfect?

Meaning, continuous training is not an option. You have to come up with something better: a descriptive differential equation for each of the

data threads, so your model is able to predict the future reliably, from the #physics perspective. What I mean by **data threads** is the different data streams from sensors such as pressure, temperatures, flow rates, etc.

From there, the acute need of changing course to a more scientific and engineered approach via #PhysicsOfData, where physics takes the reins again by treating the data threads as physical and natural phenomena with their own governing equations. Our mission is now to find the parameters, identify the hidden state variables, and set up the equations that match the calibration data.

PS. Since most of the people understand #ArtificialIntelligence as a synonym of machine learning, please, assume #AI and #ML, when I say machine learning. Sadly, that is the unfortunate situation we are in: faking AI models without physics.

#spe #SciML #DiffEq #PetroleumEngineering