The Physics of Data. Part I

The Physics of Data. | Alfonso R. Reyes

From Data Science to Machine Learning, to Artificial Intelligence. What are we missing?

If you have been in the energy industry and applying data science with machine Learning long enough, I am sure you feel a void. A certain feeling that something is amiss. That DS and ML/AI work sometimes, but not at all times, and is incredibly difficult to get good enough data for the ML/AI algorithms. So, you end up with a model, or solution, that is half baked.

Bad data

Of course, all tutorials and books on the subject are interestingly absent of bad data; you cannot run examples with bad data. The most successful DS, ML projects will be those with near perfect data. And for that you have to live in an **ideal world**. Forget about offshore, far away sensors, bad measurements because of the weather, inherent side effects of the system, its interaction with the environment, plus the human related error.

Then, what is the point of talking about Artificial Intelligence if we cannot pass yet the stage of producing good data? What's the point of acquiring expensive AI software and hardware if we don't have the house in order yet?

Real world data is not perfect

If you still don't get the point, let me describe it with a practical example. It is related to oil production. Let's say you want to invent a method to measure the **three phases** of oil, gas and water from the usual monthly production well tests against tank and separator. No rocket science there. So, you take some measurements at surface such as wellhead pressure, WH temperature, oil rate, water rate, gas rate, GOR, watercut, and a handful more. You would think that in a year, or couple of years, you have enough data points to make a prediction of your flow rate based on the WH pressure and temp. But surprisingly, you can't. Because some of those measurements are incorrect; maybe a couple are fine. Oilfield data is not perfect and there is not lots of it. Insert here […] all the factors that negatively affect a good measurement; I bet not less than ten. That also creates the uncertainty of what good or bad data points really are.

Is physics wrong?

Hey, you say! "But I know what a good measurement is because I run the **well physical models**(Prosper, PipeSim, etc.) every week, or so. I know that as a fact."

Sure, I believe you. But how did you calibrate the **physics-based models**? With oilfield data! A non-virtuous cycle. Is the Physics wrong? Is the ML, AI model wrong?

Maybe AI will cure my data problem!
No. AI won't cure your data quality problem.

But introducing more physics, through Differential Equations, to account for the enormous variations in your data, will.
How?

The past two or three years we have been enormously distracted by AI. Yes, that AI of Generalized Language. Little attention has been paid to quality control of the data. Much less attention to science, physics in particular.

Alfonso R. Reyes • You
VP Artificial Intelligence Engineering - Energy Division
11mo • Edited •

The Physics of Data. Part I

From Data Science to Machine Learning, to Artificial Intelligence. What are we missing?

If you have been in the #energy industry
applying #datascience with #machineLearning long enough, I am sure you feel a void. A certain feeling that something is amiss. That DS and ML/AI work sometimes, but not at all times, and is incredibly difficult to get good enough data for the ML/AI #algorithms. So, you end up with a model, or solution, that is half baked.

Bad data
Of course, all tutorials and books on the subject are interestingly absent of #BadData; you cannot run examples with bad data. The most successful #DS, #ML projects will be those with near perfect data. And for that you have to live in an *ideal world*. Forget about offshore, far away sensors, bad measurements because of the weather, inherent side effects of the system, interaction with the environment, plus the human related error.

Then, what is the point of talking about #ArtificialIntelligence if we cannot pass yet the stage of producing good data? What's the point of acquiring expensive #AI software and hardware if we don't have the house in order yet?

Real world data is not perfect
If you still don't get the point, let me describe it with a practical example. It is related to #oil production. Let's say you want to invent a method to measure the *three phases* of oil, gas and water from the usual monthly production well tests against tank and separator. No rocket science there. So, you take some #measurements at surface such as wellhead pressure, WH temperature, oil rate, water rate, gas rate, GOR, watercut, and a handful more. You would think that in a year, or couple of years, you have enough data points to make a prediction of your flow rate based on the WH pressure and temp. But surprisingly, you can't. Because some of those measurements are incorrect; maybe couple are fine. #Oilfield data is not perfect and there is not lots of it. Insert here [...] all the factors that negatively affect a good measurement; I bet not less than ten. That also creates the uncertainty of what good or bad data points really are.

Is #physics wrong?
Hey, you say! "But I know what a good measurement is because I run the *well physical models* (Prosper, PipeSim, etc.) every week, or so. I

know that as a fact."

Sure, I believe you. But how did you calibrate the *physics-based models*? With oilfield data! A non-virtuous cycle. Is the #physics wrong? Is the #ML, #AI model wrong?

Maybe AI will cure my data problem!
No. #AI won't cure your data quality problem.

But introducing more physics, through #DifferentialEquations, to account for the enormous variations in your data, will.

How?

The past two or three years we have been enormously distracted by AI. Yes, that AI of Generalized Language. Little attention has been paid to quality control of the #data. Much less attention to #science, #physics in particular.

#spe #SciML #PhysicsOfData #PetroleumEngineering