Data science vs. operations research: A comparison

Machine learning is more popular today yet it still includes OR algorithms

By Ilyas Iyoob

Recent trends in the field of data science have been taking the industry by storm. A quick look at Google trends shows exponential growth in "machine learning" searches since 2011, whereas searches for "operations research" have steadily declined (see Figure 1).

While search trends may not be the best indicator of such growth or decline, it is still a good reflection of public interest with respect to these domains. So, we see that there are far fewer searches for "operations research" specifically, but could that be due to the fact that people are searching for some other aspects of operations research?

At the end of the day, it's all about the algorithms implemented within machine learning (ML) and operations research (OR) that make them what they are. So, let's take a closer look at the last three years to see what algorithms have been trending, as shown in Figure 2.

The gap between the two lines is not as wide as in the previous figure. This tells us that even though people may not be searching for "operations research," they are still searching for OR algorithms. I would argue that OR is very much alive and well through its algorithms, but it is now known by a different name: prescriptive data science.

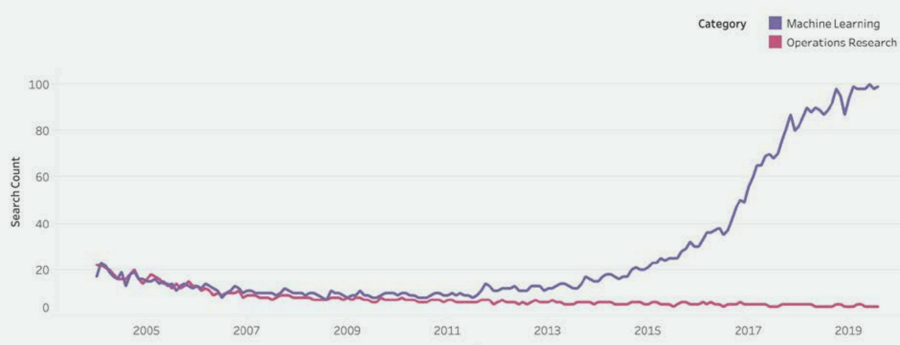
Branches of data science

Data science (DS) is the study of data with three main branches: descriptive, predictive and prescriptive. Each branch addresses the following questions:

FIGURE 1

Rising interest over time

The level of Google search interest for machine learning and operations research terms over time (listing relative index value based on maximum number of searches for a related topic and set to 100).



1. Descriptive DS: What has happened?
2. Predictive DS: If this continues, what would happen in future?
3. Prescriptive DS: What change should we make to improve the outcome in future?

Basically, in the data science domain, statistics is known as descriptive data science, machine learning is predictive data science and operations research goes by the name of prescriptive data science.

This answers two main questions. First, operations research/industrial engineering professionals should rest assured they are very relevant today and can market themselves as data scientists without any hesitation if they choose to do so. Secondly, the topic of DS vs. OR seems nonsensical given that OR is a subset of DS. So we will rephrase the topic as machine learning vs. operations research since that's where the real contention lies today.

ML vs OR algorithms

To select a winner between machine learning and operations research, let's take the previous algorithm trend chart and drill down into the specific algorithms ranked by popularity (see Figure 3).

Now we're starting to see ML take the lead as its algorithms dominate the top 20 by far. With respect to OR algorithms, linear programming comes in seventh place, followed by dynamic programming in 11th place and the rest are relatively insignificant. If these ML algorithms are so popular, let's dig into the top few and see what we can learn in the table in Figure 4.

Notice that in every one of these, the methodology always involves a minimization or a maximization function. While all the algorithms are for supervised ML models, this use of optimization is common even among unsupervised ML models such as K-means clustering, where data points are partitioned such that the distance to centroids are minimized, as well as reinforcement learning models such as deep Q-networks, where actions are selected that maximize cumulative reward.

In essence, there is no ML without OR due to the

FIGURE 2

Algorithm interest

The level of Google search interest for ML and OR algorithms from 2016 to 2019.

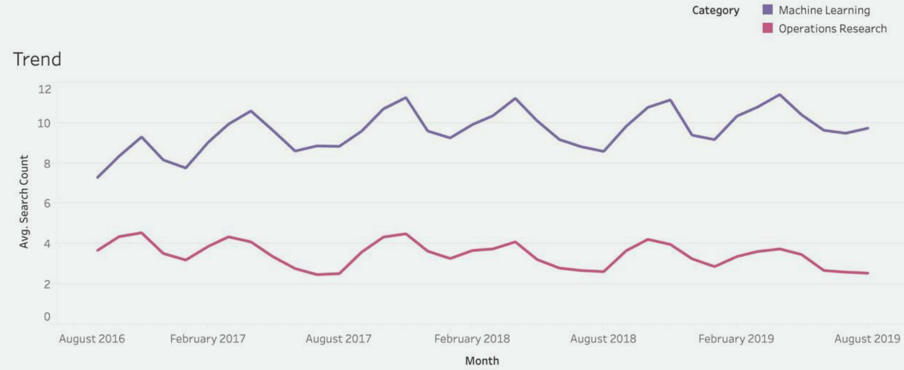
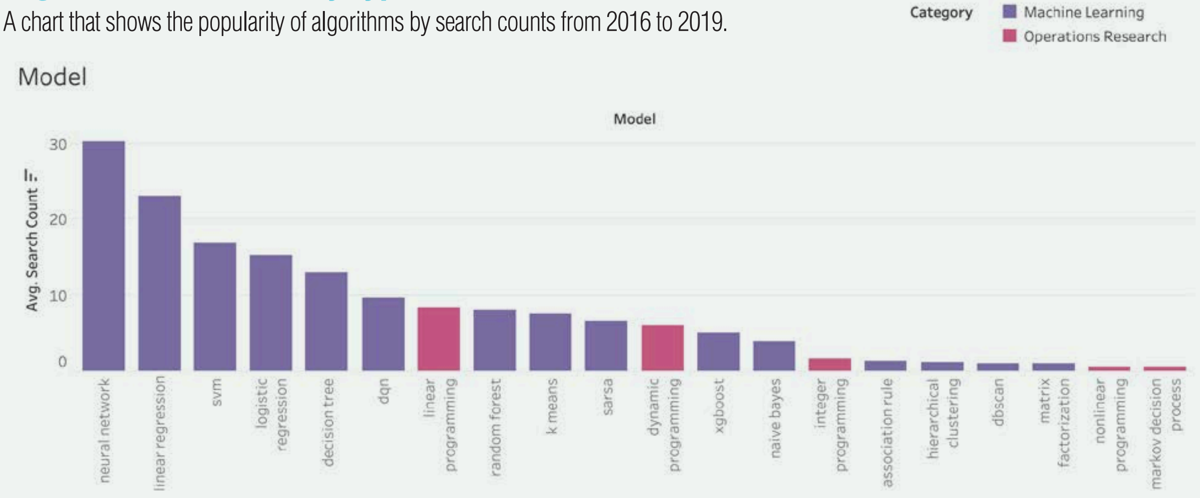


FIGURE 3

Algorithm searches by type

A chart that shows the popularity of algorithms by search counts from 2016 to 2019.



heavy dependency of ML algorithms on OR search space optimization techniques.

ML-OR interaction

We now know that operations research within machine learning is one type of interaction between them, in fact a very critical one. Below is a list of this, as well as other possible interactions between ML and OR:

- OR within ML (very common): ML methods leverage OR methods for embedded optimization, as seen in the previous section.
- ML within OR (uncommon): ML methods are used to make OR methods more efficient. For example, use ML to select variables to branch within an OR branch-and-bound algorithm.
- ML feeds OR (common): ML models are used for data preprocessing to set the parameters that can then be used by OR models. For example, to predict the demand of brand new products that don't have a history using ML decision trees, and then use OR integer programming for price optimization (David Simchi-Levi, et al.)
- OR feeds ML (highly uncommon): OR models are used to make decisions which are then fed into ML models.

FIGURE 4

Popular processes

A look at the top five machine learning algorithm methodologies.

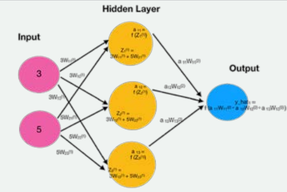
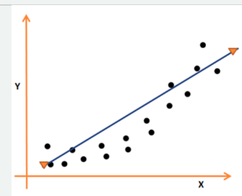
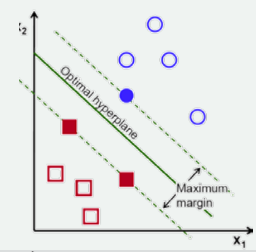
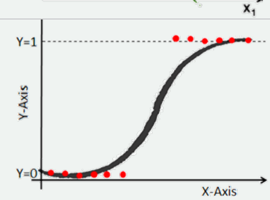
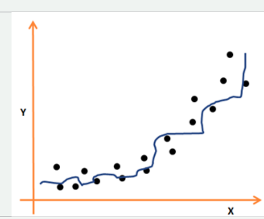
1	**Neural Network**		**Finds the best set of weights that minimizes loss**
2	**Linear Regression**		**Finds the line that minimizes MSE**
3	**Support Vector Machines**		**Finds the optimal hyperplane that minimizes hinge loss**
4	**Logistic Regression**		**Fits a curve that minimizes logistic loss**
5	**Decision Trees**		**Partitions the data recursively for maximum information gain**

Figure 5 is a two-dimensional visual depiction of the previously shown algorithm popularity chart along with the common interactions.

Research opportunities

While some of these interactions may be uncommon today, it is likely that we will find new problems or reformulate traditional problems in ways that leverage such interactions in the future. Here are some research activities currently in progress:

- Replace heuristics in ML with optimal OR equivalents. For example, replace the ML classification and regression tree algorithm with the optimal classification tree (Dimitris Bertsimas, et al.).
- Solve classic OR problems using ML. For example, solve the Newsvendor and vehicle routing problems using ML algorithms (Afshin Oroojlooy, et al.).
- Apply OR modeling with constraints to replace ML predictions in critical domains. For example, solve adversarial perturbation problems and autopilot problems using OR models with constraints.

Machine learning has earned some well-deserved popularity in the past few years by focusing on quick wins in the industry. However, operations research professionals can rest assured that they are not obsolete. In fact, most ML algorithms have embedded OR algorithms within them. Furthermore, with a little training in ML tools, OR professionals are well suited to be the elite data scientists of tomorrow covering the most coveted prescriptive arm of data science.

ML's biggest success has been in the automation of already

Where industrial engineering meets data science

How did the terms data and scientist get appended together to create the hottest job in the 21st century?

As the cost of data storage decreased over the past few years, organizations started to record more details of their business transactions. Along the way, some tech-savvy practitioners realized that the answers to many questions were hidden in the data if only there were a way to unlock its secrets.

Around the same time, researchers perfected the science of regression and called it machine learning. Once these toolsets were open-sourced, practitioners applied them to their voluminous data and observed that some of their questions were being answered. As a result, these practitioners began to be known as data scientists because they were systematically gathering and using data to validate and test hypotheses.

One of the most common questions I get is: What makes someone a data scientist? On one hand, I get hundreds of job applicants labeled “data scientists” based on their six-month online data science certifications. On the other hand, there are many professionals in the areas of operations research, statistics, and industrial engineering who have been using data for decision-making for a few decades but aren’t labeled as data scientists.

Ironically, data science is more art than science. It is the art of telling a story using data. Anyone who has the ability to derive insights from data and then stitch them together to tell a story is a data scientist, whether their background is in math, statistics engineering, computer science, operations research, finance, etc.

However, some of the best data scientists are not even from technical backgrounds but are subject matter experts in the field of medicine, agriculture, transportation, etc. who have picked up some tools and are now able to tell stories in their fields.

But what are the basic skills expected of a data scientist? While it is not necessary to have a technical background, there are some technical skills that could propel anyone to be an efficient data scientist.

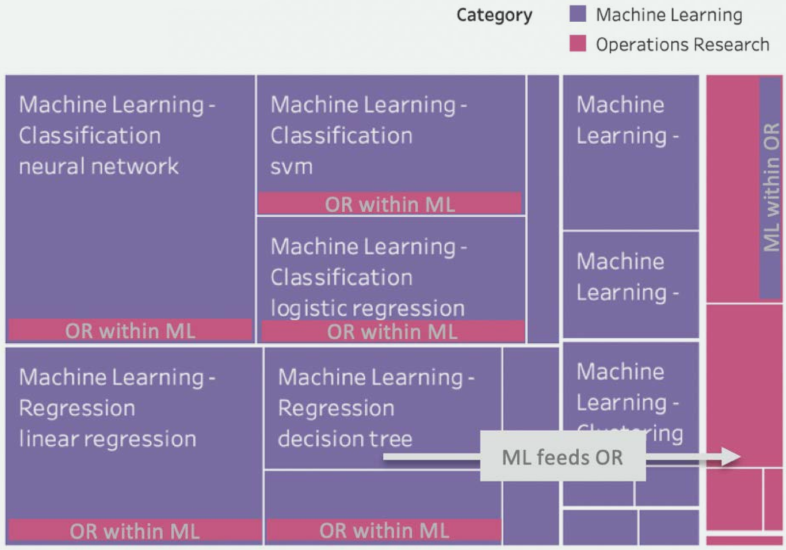
There is absolutely no reason for an IE comfortably working in retail, supply chain, manufacturing, healthcare, etc., to feel the need to switch careers. However, I believe the inherent DNA makes every IE a good candidate to learn data scientific skills which can be used to complement the IE toolset.

–Ilyas Iyoob

FIGURE 5

Algorithms by type

The algorithm popularity by search count from 2016 to 2019 with common interactions.



existing techniques as well as recent developments in recommender systems, deep learning and reinforcement learning. Kudos to IBM, Google, Amazon, Facebook and Netflix! ❖

Ilyas Iyoob, Ph.D., is chief data scientist and distinguished engineer at IBM. He pioneered the application of operations research to cloud computing including algorithms for IT supply chain optimization, virtual data center capacity planning, and automatic scaling and provisioning of virtual machines. He successfully built and sold a startup company, Gravitant, focused on optimizing the cloud journey. He advises a number of startup companies in Austin, Texas, and serves on the faculty of the Department of Operations Research and Industrial Engineering at the University of Texas at Austin. He has earned a number of patents and industry recognition for pioneering the cloud analytics space. In 2001, he won the World Mechanics prize awarded by the University of London and has been included in the Marquis Who’s Who in the World *since 2012. His work has been published in a number of peer-reviewed journals worldwide. Iyoob holds bachelor’s and master’s degrees in industrial engineering from the University of Arkansas and a Ph.D. in operations research from the University of Texas at Austin.*