

2019 Trends in Data Science: Innovation and Acceleration

By Jelani Harper

Two antipodal motions will impact data science in 2019, as intrinsically related to each other as they are oppositional. The first is the movement towards democratized data science, in which **self-service data preparation and data engineering platforms** are responsible for automating many of the manual processes for readying data for business consumption.

The second is unequivocally enabled by the first, and allows data scientists to devote more energy to tasks beyond the data wrangling that has conventionally consumed so much of their time. Self-service data preparation has spawned an era in which data science is characterized by innovation and acceleration, rendering its sophisticated elements more utilitarian than previously possible.

The most pressing issues to affect advanced analytics—including explainability for deep neural networks and complicated classic machine learning models, accuracy of transparent machine learning algorithms, and the enormous amounts of data required to train cognitive statistical models—have resulted in data science solutions to rectify them.

Next year will see an even greater emphasis on data science efforts pertaining to the notions of attention, memory, interpretability, ensemble modeling, surrogate modeling and more, which will not only redress the aforementioned issues, but also improve artificial intelligence's collective ability to meet currently lofty expectations.

Perfecting White Box Accuracy

One of the more paramount trends affecting statistical AI is the limited explainability of so-called black box machine learning models. The allure of this option is greater model accuracy; the drawback is a dearth of transparency for understanding models' quantitative outputs. Reciprocally, classic machine learning models are oftentimes characterized by lower accuracy rates, with some predictions as low as 50 percent. The division between unexplained, correct predictions and lucid white box models with limited accuracy has contributed to a surge in ensemble modeling, which "by definition, is combining a multiple number of models," **Razorthink** Data Scientist Indranil Bhattacharya said. "It can be two or like, 500 or 1,000 models."

Gradient Boosting is perhaps the premier ensemble modeling technique to emerge in the past couple years. This sequential method involves data scientists creating an initial model, then using its results—regardless of how poor they are—to inform the quick construction of another. The repetition of this process results in models with much higher accuracy than the initial one, while **Extreme Gradient Boosting** provides these advantages at scale. Building too many models, however, yields interpretability problems. **Bagging** is a more popular ensemble model approach in which independent models are built in parallel from the same datasets, enabling data scientists to use different parts of that dataset's characteristics and features to swiftly build models. Moreover, they can use different types of machine learning models for aggregated predictive prowess.

Cracking the Black Box

Options for deep learning and complicated neural networks with scores of parameters, hyperparameters, and nuanced layers are usually too dense to understand exactly how they're generating outputs—especially since deep learning does its own feature detection. Escalating regulatory demands have renewed efforts towards generating both explainability and interpretability for black box models. The most straightforward of these methods is **maximum activation analysis**, which requires (during the optimization phase) increasing the weights of patterns producing surprising results during deployment. The greater weights emphasize these patterns in future iterations, enabling data scientists to infer aspects of their functionality.

Surrogate models, simpler models built from complex ones, also help. With this option, data scientists use the original predictions and inputs of black box models to train simpler surrogate ones; the resulting interactions and statistical information should resemble those of the complex model. **Leave One Covariate Out** is a technique in which modelers exclude a single variable from a black box model and see if there's a considerable difference in its results. "If something dramatic happens, then the variable is pretty strong and you make a note of it like, how much is the difference, and try to match that with your expectation," Razorthink Data Scientist Ankit Raj said.

This website uses cookies, including third party ones, to allow for analysis of how people use our website in order to improve your experience and our services. By continuing to use our website, you agree to the use of such cookies. [Accept](#)

We use cookies. By browsing our site you agree to our use of cookies.

Click here for more information on our [Cookie Policy](#) and [Privacy Policy](#).

[Accept](#)

requires large amounts of computing power organizations can't always devote to such tasks. There's two chief ways to overcome these limitations: transfer learning and Hierarchical Temporary Memory.

Transfer learning applies knowledge already available to statistical AI into other domains while using negligible amounts of training data and compute power. "Transfer learning is an approach by which you can take generalized models and use it to solve specific problems," [Indico](#) CEO Tom Wilde commented. It's typically facilitated with deep neural networks in use cases for text analytics, [Natural Language Processing](#), and image recognition.

[Hierarchical Temporal Memory](#) may be even more promising because it doesn't learn the way neural networks do. By using [Sparse Distributed Representation](#), its learning approach is similar to that of human neurons turning on in the brain. It represents concepts with ones and zeros in an SDR matrix, which requires much less training data than that for neural networks and is perfect for rapidly changing, real-time data in fraud detection and speech recognition use cases, among others.

Additional Memory Methods

The concept of memory is indispensable to more difficult AI applications involving time series analysis. Examples include episodic memory for conversational speech systems, which can contribute to users being able to "ask a question the human way and still get the right answer," [NTT Data Services](#) CTO Kris Fitzgerald remarked. Other use cases include enterprise staples of Optical Character Recognition (OCR), text analytics and video analysis. Memory is primarily facilitated with three approaches involving:

- **Recurrent Neural Networks (RNNs):** Traditional RNNs are the least refined of the three means of implementing memory discussed herein. Although they're only effective over short intervals, their memory units capture context critical for language translations. They're generally viewed as improvements over Convolutional Neural Networks in this respect.
- **Long Short Term Memory (LSTM):** LSTM was added to RNNs to enable memory over longer time periods. Traditional RNNs can issue memory between words; RNNs can do so between sentences and some paragraphs. The general way LSTMs capture memory is by assigning a time step (a cell) to each word, for instance, so that if "we have four words, we have four time steps, and we have four outputs, and there's a horizontal memory unit moving from left to right," denoted Razorthink Deep Learning Engineer Shreesha N. LSTMs are key for providing context for, and improving the accuracy of, OCR.
- **Memory Augmented Networks:** [Memory Augmented Networks](#) are the most recent memory mechanism to emerge from data science. Their main improvement over LSTMs is they can write, overwrite, and lookup information about a particular dataset in a dedicated memory bank, with which LSTMs aren't equipped. Thus, they can discern context and remember concepts between multiple paragraphs, for example, or monitor contemporary financial trends alongside historic ones with their greater memory capability.

Attention Variants

Ascribing attention to AI systems is one of the more recent developments to proliferate through data science. Although attention is a natural accompaniment to memory, the ability to understand which part of a security footage video, for example, is more important to focus on at a particular point in time, or which words in a passage represent the main idea in intricate social media postings, is essential for maximizing the value of AI investments. Attention is delivered in intelligent systems in two main ways: through dynamic co-attention neural networks and attention mechanisms.

[Attention mechanisms](#) can simultaneously hone in on a certain part of a deep learning model's input while remaining cognizant of the rest of the model's input. This advantage is largely implemented by the capability of attention nuances to "not linearly go from the beginning of a sentence to the end, but to go back and forth and that sort of thing," said Rohan Bopardikar, Razorthink Technical Director of Artificial Intelligence. Attention variants are useful for image recognition, accurately classifying documents and, most of all, answering questions in natural language without circumscribed, template-based approaches. [Dynamic co-attention networks](#) were specifically designed to answer questions. They hone in on germane words in a question to provide relevant answers involving variations of those words, which is pivotal for speech recognition and conversational AI.

The Data Science Stratosphere

The innovation and acceleration typifying data science in the coming year will certainly allow for more practical, real world AI. This progress, however, will first address the customary shortcomings of both transparent machine learning and deep learning, before expanding into the more advanced reaches of the data science stratosphere.

Jelani Harper is an editorial consultant servicing the information technology market, specializing in data-driven applications focused on semantic technologies, data governance and analytics.

Share this:

