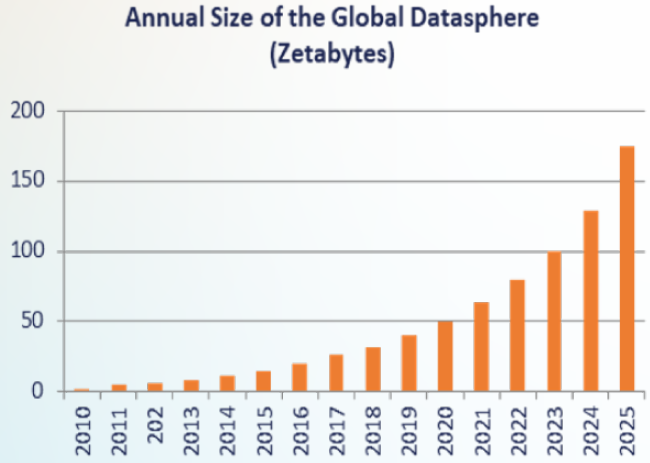
Summary Analytics

www.smr.ai
415-234-3553
info@smr.ai

Summary Analytics for Autonomous Transportation

Nowhere are the expectations for Machine learning (ML) and Artificial Intelligence (AI) higher than they are for autonomous transportation. From self-driving cars to drone package delivery, the vision is compelling and demand is high. According to Hussein Mehanna, head of AI at General Motor's autonomous driving subsidiary Cruise, "Handling the long tail is the reason autonomous vehicles are one of the most difficult and exciting AI problems on the planet."[1] The amount of data and the processing costs to train for this long tail of corner case driving scenarios are huge, as Cruise generates 300 terabytes of data daily in spinning up 30,000 instances across over 300,000 processor cores and 5,000 graphics cards. Big data is not just needed for training but for inference, too. Former Intel CEO Brian Krzanich estimates that 5 terabytes of data will be generated and consumed by self-driving vehicles per hour of drive time.[2] That's a lot of data when the over 1-billion cars on the road today are replaced with autonomous vehicles! But computer vision errors are still three orders of magnitude away from human drivers, according to Amnon Shashua,[3] CEO of Intel's autonomous driving subsidiary, Mobileye. Shashua suggests the solution is even more data, from additional sensors and higher resolution maps of the environment.

Even if you aren't working on Level 5 full autonomy today, advance driver assistance systems (ADAS) such as adaptive cruise control and lane departure warning systems still have much room for improvement. The auto insurance industry's non-profit Insurance Institute for Highway Safety (IIHS), tested five adaptive cruise control systems and reported that even for the best system, "In 180 miles, the car unexpectedly slowed down 12 times, 7 of which coincided with tree shadows on the road."[4] Beyond autonomous cars and trucks, robots and drones are being deployed in larger numbers but still have similar AI training challenges for most applications. How do you deal with the need for even more training data while preventing the time and cost of more processing and storage from getting out of hand?



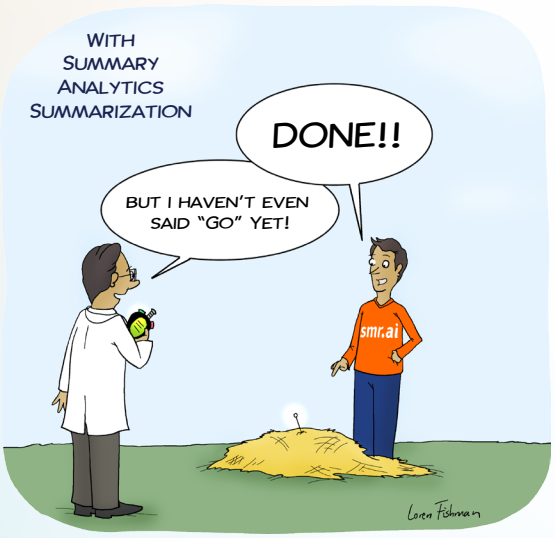
Source: Data Age 2025, from IDC Global DataSphere, Nov. 2018

To get more done with limited resources, you must become more productive. With Summary Analytics's mathematically proven artificial intelligence techniques, you can shrink the datasets through summarizing and prioritizing without loss of fidelity – delivering better insight while reducing time and cost, and significantly reducing the amount of manual data labeling required. This minimizes the common problem of operator fatigue errors in data labeling and the resultant errors in the models.

Likewise, Summary Analytics can help with training your AI models. The computational power required to train state-of-the-art AI models is doubling every 3.4 months[5] as Moore's Law continues losing steam, no longer doubling processor performance every 18-months. So far, this problem has been addressed with machine learning algorithmic advances and increased parallel compute power. These help, but more is needed to stop runaway AI analytics costs and delays. A new complementary tool is needed, adding "informational efficiency" to the process. That tool is Summary Analytics. Our software-

as-a-service (SaaS) offering summarizes and prioritizes data sets before running expensive analytics. Summary Analytics enables early model testing on significantly reduced and prioritized datasets, while saving larger (but still reduced) datasets to be used for final optimization of the model. Summary Analytics eliminates redundancies in your data and there is considerable redundancy in most sensor and environmental data used for autonomous transportation AI training. More than just deduplication, we eliminate redundancy even among massive numbers of unique records. We eliminate the unnecessary and shrink the haystack so finding the needle of insight is faster and less expensive. Of course your data is more complex and dynamic than a single needle in a haystack, with new streams and data churn constantly adding the new and removing the old. Worried about what to do with old data? Whether from cameras, LIDAR, Radar, GPS, weather reports, or vehicle status sensors, we help create a data hierarchy to focus on the important data, whether old or new.

How does it work? Professor Jeff Bilmes from University of Washington in Seattle developed proprietary calibrated submodular (CaSM) functions which mathematically analyze and order data along the lines of diminishing marginal returns. We automatically prioritize the data in terms of its biggest contribution to the information content of the entire data set, and then relegate redundant data to the end. CaSM functions are extremely processor efficient – orders of magnitude faster than typical AI algorithms. They don't replace AI algorithms, our CaSM functions just make machine learning run much faster since the data sets are vastly smaller but still contain all the important information. And we work on any kind of data, whether health records, customer profiles, network logs, biological signals, even images, audio, and video streams or the complex mixed data streams of autonomous transportation.



Whether you're working on autonomous features for cars, trucks, farm machinery, robots, or drones, the bigger, more costly, or more redundant the data, the more Summary Analytics can reduce costs and make your ML/AI team more productive.

Bigger data? Bring it on!

[1]https://venturebeat.com/2020/04/28/challenges-of-developing-autonomous-vehicles-during-coronavirus-covid-19-pandemic/

[2]https://channels.theinnovationenterprise.com/articles/why-big-data-is-the-future-of-self-driving-cars

[3]https://www.technologyreview.com/2019/04/23/103181/the-three-challenges-keeping-cars-from-being-fully-autonomous/

[4]https://spectrum.ieee.org/cars-that-think/transportation/self-driving/research-reveals-safety-problems-in-driver-assistance-systems

[5]https://www.technologyreview.com/s/614700/the-computing-power-needed-to-train-ai-is-now-rising-seven-times-faster-than-ever-before/?utm_source=newsletters&utm_medium=email&utm_campaign=+the_download.unpaid.engagement

©2021, Summary Analytics, Inc.