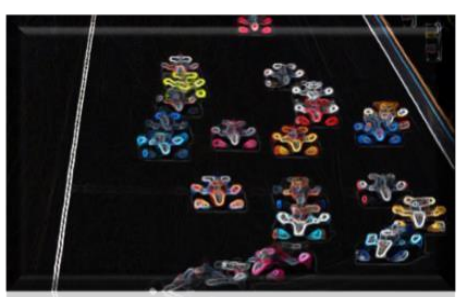
PREDICTING THE UNPREDICTABLE

Implementation of a Real-Time Advanced Machine Learning Method to prevent un-expected machine failures for Mission Critical Applications.

Continuous Intelligence for Advanced Mobility

Patent Pending
Application #63214232
DYN2021001-US-PSP



DYNAMOEDGE, LLC, August 2021.

Market Opportunity

Studies show thar nearly $1 Trillion a year is lost to machine failures. To prevent expensive failures form happening, companies need a proactive approach, instead of reactive.

Large industrial facilities lose more than a day's worth of production each month and hundreds of millions of dollars a year to machine failures, [The True Cost Of Downtime] report shares findings from a study of 72 major multinational industrial and manufacturing companies. It reveals that, on average, large plants lose 323 production hours a year. The average cost of lost revenue, financial penalties, idle staff time and restarting lines is $532,000 per hour, amounting to $172 million per plant annually. Cumulatively, Fortune Global 500 (FG500) manufacturing and industrial firms are estimated to lose 3.3 million hours a year to unplanned downtime. The financial cost of this downtime to those organizations is calculated at $864 billion, the equivalent of eight percent of their annual revenues

Automotive

- The cost of unscheduled downtime is highest in the automotive sector, where the products manufactured are of high value and plants and production lines are often closely interconnected, meaning downtime has a knock-on effect
- Automotive plants lose 29 production hours per month, costing them $557 billion a year - an estimated 20% of annual revenue. Encouragingly, however, 67% of professionals in this sector told us that Predictive Maintenance was now a strategic priority

Oil & Gas

- The average number of hours lost due to downtime is highest in the Oil & Gas sector: 32 hours per facility each month. This could be due to the safety-critical nature of the work: production stops at the first sign of a potential problem
- Shutdowns in the Oil & Gas sector cost each facility $220,000 per hour - amounting to $84 million per facility each year. In refineries alone, losses to FG500 companies cost an estimated $47 billion from 213,000 downtime hours each year
- Not surprisingly, given the safety-critical nature of production, 82% of those in the Oil & Gas sector said predictive maintenance was already a strategic objective, the most of any sector

Heavy Industry

- While mining, metals and other heavy-industrial plants lose the least number of production hours each month to machine failure (23 hours), the cost is exceptionally high
- Machine failures cost heavy-industrial companies $187,500 per hour, on average. This eats heavily into profits - totaling $225 billion a year across the FG500
- 60 percent of heavy industrial companies have now made predictive maintenance a strategic priority

Abstract

To prevent failures, first, you need to be able to detect them with enough anticipation to react, which has been a challenge with present technology for connected machines. To experience this proactivity, real-time monitoring is not enough, different analytics and a low latency data pipe are needed, so that rather than looking at the past to deliver predictions, the user can be focused on querying the future to correct issues before they take place and not while they are taking pace. Historically, data scientist and researchers have mostly been interested in fitting a model to a dataset. But only a very small subset of these scientists have been looking to create new Machine Learning designs, counterintuitively capable to adapt its linearity over time. Most of the ML algorithms available, assume that all the data is available at once, which is ideal for simulations. These simulations consume a significant processing power and storage and are unable to detect real-time anomalies. Transportation represents 29% of all energy consumed in the United States in 2021, and automotive is responsible for 23% while having the largest downtime, as seen above. That's why we decided to benchmark our technology with race cars, the most demanding mobile devices available. In racing, the main goal of the teams is to efficiently utilize the available energy of the car, without any downtime. Drivers and crews prepare off-season by training extensively on simulators, but unpredicted car failures can make them lose the race in a matter of milliseconds. These unexpected events deteriorate the car performance by generating oscillations and a mechanical desynchronization in the components of the vehicle, which are referred as car or race vibrations. One of the main problems we needed to solve, is that current analytics are ran post-event, in simulations run over static data previously collected, and they are not designed to detect un-predicted real-time failures, hence a continuous intelligence solution, that works with live data streams as they are being generated, is needed. DYNAMOEDGE real-time ML analyzes data constantly streaming from hundreds of mission critical sensors, unattended and frictionless. In this paper, we will focus on the applicability of DYNAMOEDGE technology to reduce downtime for the advanced mobility market, and we will demonstrate it by providing a continuous intelligence proactive solution for the fastest and most data-dense mobile devices that exists, RACECARS.

Problem Statement

Defects are unknown and not linear thus, cannot be detected on simulations. If we can predict on-the -fly what is going to happen in the future, we can give the racing teams an opportunity to adjust the race strategy within milliseconds, avoid an accident and make or break a win. The race car is an ever-changing data source and one of the most complex IoT mobile devices, and in two year automotive will be 53% of the Worldwide IoT solutions. Therefore, our solution is to be applied to any smart transportation, autonomous driving and smart logistics to predict and prevent un-expected failures, providing an end-to-end platform to enable advanced mobility at scale. As organizations mature in their use of ML, they often uncover new use cases that require real-time features, often with data freshness requirements in the order of milliseconds. These features can't be processed from batch data but need to be built using streaming or real-time data sources. These pipelines must process lots of data at low latency, manage backfills to eliminate training, while performing complex operations like time window aggregation.

Currently, the common practice for machine analytics is to have an offline phase where the model is trained on a dataset and then, as a second step, the model is deployed online to make predictions on new data afterwards. Hence, the model is treated as a static object, and it will not work for the ever-changing characteristics of live data. Today, Machine learning models are deployed in such a manner because although they're extremely robust, they are equally limited by the learning capacity of the model. To learn from new data, the model must be retrained using lots of historical datapoints simulating different situations, consuming significant computing and storage. For this extreme use case, we need to process a very heavy dataset since there are over 300 sensors available per racecar and 6 of these cars (fleet) producing terabytes of data continuously streaming in seconds, in multiple formats, from different cars, with different driving profiles, different track locations, on -board video cameras, and other racing information such as weather and wind. Each one of these cars measure in1 second the same amount of data than a standard home gateway processes in 1 hour. Racing datasets are subject to errors and noise during acquisition and transmission, demanding an adaptable prediction model that corrects data drifting, while also providing live predictions. Like any mechanical component, when stressed, tires will eventually wear and fail; that's why the cars need to stop at the pit. In addition, new tires are installed at every pit stops approximately every 20min, so the desired model needs to learn and predict in 20 seconds based on new datasets with changing attributes. We tested static pre-trained neural network models and they were not capable of making correct decisions or inference on the continuous data streams with patterns changing over time so quickly. Therefore, the algorithm needs to be capable of capturing the drifting of data patterns and prune accordingly on-the-spot. The algorithm must keep adapting its outcome unattended and continuously with extreme memory efficiency. The end-to-end system needs to be able to handle a response with latency below 100 milliseconds, to cope with the sensor data arrival rate of sub-80 milliseconds.

Background

The objective is to rapidly build and deploy Real-Time Machine Learning models that can scale with hundreds of models serving high volume requests or QPS with low-latency prediction

Streaming telemetry aims to modernize the network and device metrics to keep up with the scale of next generation networks and provide new ways to access the huge variety of analytics that network endpoint devices can now generate. Streaming data inverses the normal read/write patterns of databases, where there is limited writing, but lots of reading. The sensor readings can be used in basic electromechanical operations, but the data transmitted over the network can also be used to perform data mining to identify anomalies in the system, component malfunctions and new insights if the appropriate network infrastructure is available. An IndyCar accelerates from 0 to 100mph in less than 3 seconds. Usually, a pit stop for 4 tires lasts for 12 to 16 seconds of which speed will be zero for 5 seconds. A pit-stop for 2 tires and fuel lasts 5 to 7 seconds, where the speed of the vehicle goes to zero for less than 2 seconds. There are 5-10 pit stops for each car in average. The teams run lots of complex simulations in advance of the race and, as reported by F1, even during the race, between pit stops after the data is downloaded at the pit. But surprisingly, no predictive or prescriptive analytics on the cars are performed live, which could prevent lots of unnecessary accidents. Furthermore, it has been discovered that drivers will begin to naturally slow their speed and lose focus and time, as vibrations build in the car and specially in the tires. The threshold and respective tolerance level of these frequencies are determined by the individual driver profile. In this paper we will show that our technology demonstrates great performance in terms of anomaly detection accuracy, prediction, and overall Service Level Objective (SLO) of sub-80 milliseconds latency for a live streaming application.

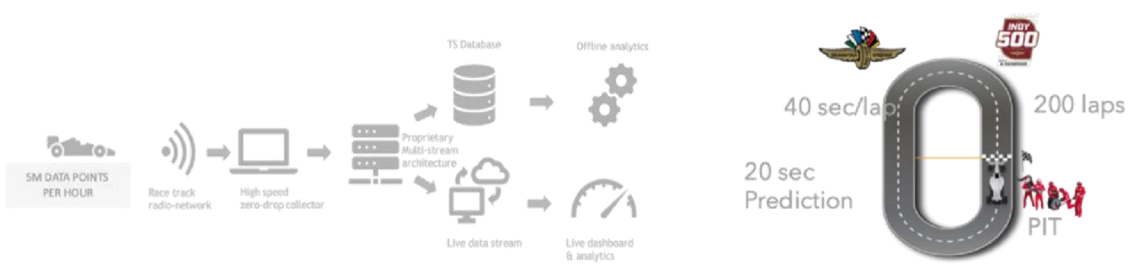


Fig 1 IndyCar telemetry configuration and lap distance

Overview of the Implementation

The DYNAMOEDGE Platform consists of these main modules:

I. A portable Mission Control Console panel gives the user the ability to monitor the entire fleet of cars, as a group and/or individually.

II. A Lossless telemetry Ingestor collects massive amounts of data from hundreds of sensors, life. It also automatically initiates the real-time ML analysis, without any user intervention.

III. Our patent pending Real-Time ML consists of three on-the-fly steps: Feature Extraction, Data Pruning and adaptive ML. To secure latency, features are analyzed before stored. After the new features have been added, older values are pruned from the dataset, updating the stream of continuous intelligence The solution breaks the time window into multiple tiles of smaller time windows that store aggregations over the tile interval, and a set of raw events at the head of the aggregation time window. The tiles are pre-computed. At feature request time, the final feature value is on-demand computed by combining the aggregations stored in the tiles as well as aggregations over the raw events.

IV. Because of our strategic integration with Excelfore eSync, DYNAMOEDGE can securely deliver at scale, updating software over-the-air, while collecting real-time operational data from in-car devices, including telematics, electronic control unit, gateways and smart sensors.

V. Continuous Intelligence Visualization dashboards allow the user to make the decision-making process faster and smarter, while there is still time to prevent the failure. The application is portable, so it can be viewed in any modern web browser, including cell phones, tablets and laptops. The dashboards display both, live data monitoring, as well as the results of the prior predictions.

Solution

DYNAMOEDGE has successfully demonstrated the performance of its platform during the 2021 Indianapolis 500, where we were able predict failures live, up to 15 laps, 37 miles, and 10 minutes in advance.

Today, sensor data is transmitted in a heavily preprocessed state because of lack of network bandwidth. The currently available bandwidth per vehicle is just a few hundred kilobytes. This will not be enough for autonomous driving. Future advanced mobility will have to be able to receive more sensor data to unlock further services.

Our Real-Time Edge Machine Learning is a paradigm shift, because it changes the way we think machine learning deployment. Our compact analytics rapidly compensate for a change in conditions, with almost no training and based on small data increments, as new data comes, avoiding downtime.

With DYNAMOEDGE, companies can achieve continuous adaptations of models based on the newest data, enhance operations and extract further business value. Analyze before storing with our proprietary on-the fly algorithms, applied simultaneously to improve the prediction models in real-time and avoid drift. Once a future failure is identified, DYNAMOEDGE can send over-the-air updates to each personalized mobile endpoint and apply an instantaneous corrective action on-the-go, enabling the next generation advanced mobility services.

In this application, we were able to predict ahead of time, when a vehicle failure will happen and issue a command, up to 15 laps, 10min, 27miles in advance, providing true Continuous Intelligence as-a-Service, increasing safety and avoiding machine downtime, resulting in a seamless end-to-end customer experience.

DYNAMOEDGE, Continuous Intelligence for the future of advanced mobility.