Machine Learning –

Predicting and Learning Logistics Systems Throughput

# Background

Sensors are a critical component in the Industrial Automation arena. With the recent explosion in the awareness of Industrial Internet or Internet of Things (IoT), sensors have become more important in their role to enable the critical transformation of data to information to insight and finally to intelligence.

SICK has been an industry leader in Sensor manufacturing and engineering for the past 70 years and has considerable presence in several industrial segments including Logistics Automation (Couriers, Parcel and Express, Retail, Airports), Factory Automation (Food and Beverage, Safety systems, Automobiles, etc.), Process Automation (Oil and Gas), etc. The sensor portfolio is widely regarded to be the best in its kind and true to SICK’s motto – Sensor Intelligence – try to push the envelope in aiding the D2I transformation.

# Motivation

Sensors have two broad goals:

The primary goal is to ensure that an action is taken based on what is being currently sensed (e.g, when a light curtain sensor is broken, send a signal to STOP the machine to ensure safety of life and limb OR, inspect bottle cap seals and reject if the seal gap is more than 0.1 mm). This goal is the first and foremost reason why sensors are used.

The second goal is in providing valuable data needed to create insights. Questions such as how often did the machine stop due to Safety events in the last 10 days and how did the productivity get impacted due to these events. To answer such questions, we need to go beyond observing “current” data, but also look at trends across time periods, correlating operational data with other direct and indirect events and parameters, etc. Industrial Internet would look to using data as the Holy Grail to ensure competitive advantages and increase efficiency.

The entire field of Analytics is devoted towards creating applications that explore this secondary goal of Sensors.

# Problem Description

We would be considering a use case in the Logistics world of Sensors and Sensor data for this project. Sensors have been used in Logistics for a long time now, providing vital triggers to allow objects of interest to be transported from one location to the next. The effectiveness of “Logistics” depends on several factors (throughput, inventory, equipment operational efficiency, operators’ efficiency, compliance of rules by various vendors, etc.). Sensors contribute to some of these factors and that’s where we would like to position our interest.

Logistics Automation applications typically use a variety of sensor types, but for this particular project, we would be looking at a few selective types of Sensors:

* Barcode readers (CLV)
* Barcode reading Cameras (ICR)

The typical usage of these sensors are in “Camera tunnel” systems where these are mounted statically on a mechanical frame over a conveyor belt where boxes or packages travel at high speed. Since Logistics is concerned more about the transport and hand-offs of packages, there’s more emphasis on tracking and tracing boxes across one or more distribution centers (“facilities”) using barcodes primarily. In addition to barcode reading sensors, most Logistics applications use Laser cameras (ICRs) to simultaneously decode barcodes and take pictures of the various faces of the boxes as they move by. Camera images are needed for a range of uses including visual proof, understanding why sometimes barcodes are not decoded properly, determining if the package sorting operations can be improved by observing how close the boxes come in through the conveyor belt, etc. A typical example of a camera tunnel is shown in this picture. The vertical blue boxes are the barcode reading cameras.



Usually, there are redundant sensors used in such applications where the sensors are placed in slightly different locations of the frame to ensure barcodes are accurately captured and not missed due to angles, gaps and varying heights of focusing.

The overall view of the data flow would look something like the below picture. The orange circles indicate various sensors used in such a tunnel.

Package (box) information**: Scan Time**, Dimensions, weight, angle of orientation, gap info, volume, belt speed, etc.

Controller

A Logistics facility typically has multiple such systems for efficient receiving, processing and shipping of objects or packages. Thus, it’s important for a Facility Manager or Operator to ensure that the productivity of such systems don’t dip below certain SLAs and threshold. It would be shortsighted to view each package individually from a sensor data perspective as that would only show what the sensor at that instant is processing, and would offer no perspective of what else is going on. Fortunately, that’s where Analytics and its ability to capture data from a large set of packages across an extended period of time would come to our aid.

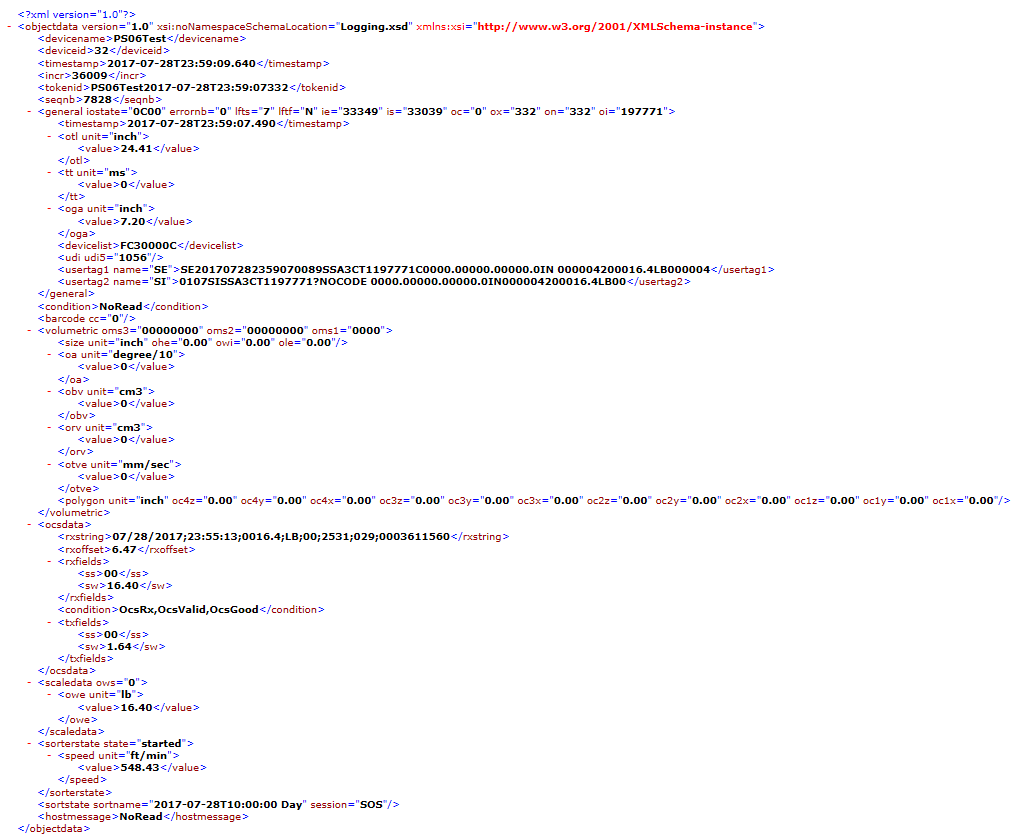
Hence the problem statement for this project can be reduced to this: What factors influence the productive utilization of the Conveyor belt and the System.

In order to answer this question, we will provide production data from a single Camera tunnel system collected over several weeks. A short description of this operational data follows:

**Operational data:**

Operational data would be provided in an XML-like format. This file contains two kinds of information – object data and heartbeat data. **Object data** corresponds to information collected from every single package or box that passes through the camera tunnel. **Heartbeat data** refers to the sensor-controller supplied information that specifically refers to the state of the system. For the moment, the most useful information that the heartbeat data provides is an indication of whether the conveyor belt is moving or not.

Here’s an example of how the object data looks like:



The actual XML snippet of this object data and the general schema file are attached here:



Logistics systems work in “sorts” or “shifts”. A typical sort would contain several hundreds to even thousands of packages (or objects) that pass through the Sensor system via the conveyor belt. Each sort could last anywhere between minutes and a few hours. Typically, there are multiple sorts in a single day. There may be days where there would be no packages processed (for example – a weekend or a holiday). In such cases, one wouldn’t see any Object data but would see heartbeat data coming in at a specific frequency. Since the heartbeat data comes from the Sensor master controller, this data stream continues to flow even if there are no package information being acquired.

The Master controller is also responsible for assigning a unique identifier for each package/ object and aggregates data from each sensor into an “Object Info” snippet. In addition, the controller also has intelligence to evaluate the object under question on the basis of several conditions and includes this data as well in the Object info. This is done for all objects that are sent in the conveyor belt.

In the dataset provided for this project, the below package conditions are evaluated by the controller. While all of them are important, the ones marked in bold are more pertinent for this particular project.

- PDFw9612

- PDFNoRead

- **TooBig**

**- NoRead**

- **NotLFT**

- ValidDim

**- MultiRead**

- **Irreg**

- ValidWeight

- ValidRead

**- TooSmall**

- Gap

- SmartPostReturn

- Clipping

**- LFT**

- PeError

- Epic96Multi

Conditions are one aspect of the influencing factors. There are other factors in the object data that might be of interest as well. **Please view these in the below annotated files for explanations on how these conditions and other factors may play a part in affecting optimum belt utilization.**



Similarly, please view a sample annotated heartbeat data snippet here:



# Expectation

With this information, the following questions need to be investigated:

* Across the entire dataset, what can we learn about package trends with respect to LFT (Legal For Trade) and NotLFT (not Legal For Trade) conditions
* How does the belt speed behave across time? Are there any correlations between belt speed and NotLFT (or LFT)?
* Investigate “Gap” information and determine if there are any observable outliers. Correlate this information with the differences in the timestamps between adjacent packages
* What can we learn about “Pileup” or “Log Jam” situations? Are trends in conditions such as Irreg, MultiRead, TooBig indicative of this phenomenon?
* In general, identify outliers in the data:
  + Random changes in the units – example: ft/min changes to m/sec
  + Total Throughput changes, Specific throughput based on conditions
  + Recurring errors observed in the Heart Beat data section

# Notes

* There are no restrictions in the kind of technology that would be used to tackle this project, however the results need to be provided with appropriate explanation to the insights
* Since the data is self-contained, there’s plenty of opportunity to be creative and exploratory using any appropriate methods.