Predictive analysis using Big data Analytics for Sensors used in Fleet Truck Monitoring System

Preethi Varsha. V
Dept. of Electronics & Instrumentation Engg.
SRM University, India
Preethiivarsha@gmail.com

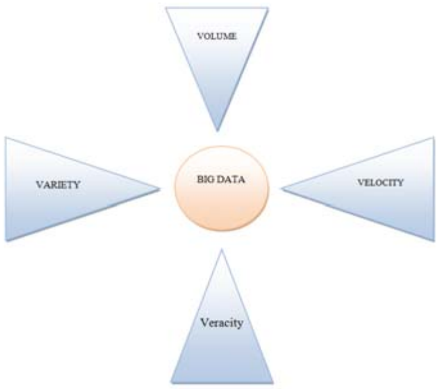
Sasi Rekha Sankar
Dept. of Software Engg., SRM University, India
sasirekha.s@ktr.srmuniv.ac.in

Rohit Jacob Mathew
Dept. of Software Engg., SRM University, India
rohitjmathew@hotmail.com

***Abstract*—The production of large amounts of data has not only been observed in web based companies but has seen entering other domains such as automation and automobile.Smart sensors and smart devices contribute to growing amounts of data that need to be processed. The outcome expected from processing is prediction for better control, clustering for more effective maintenance or improving the overall production. The potential use is to monitor machines or infrastructure such as ventilation equipment, energy meters, truck engines, tires and environmental conditions. This predictive analysis on the data can generate information like intimation for repair or replace these items even before they break; suggestion on driving patterns on various road conditions to both the driver and fleet owner. This project examines the utilization of big data technologies for truck maintenance and performance domain. The approach is based on sensor measurements with the goal of detecting specific events and patterns.**

I. INTRODUCTION

In today's world data is generated from varied locations when compared to couple of years and all the data accumulated in past two years is more than for past 100 years. As we have the technology to generate, process, accumulate and store data. The normal data becomes a new framed technology of big data when it processes the four 'V' properties.



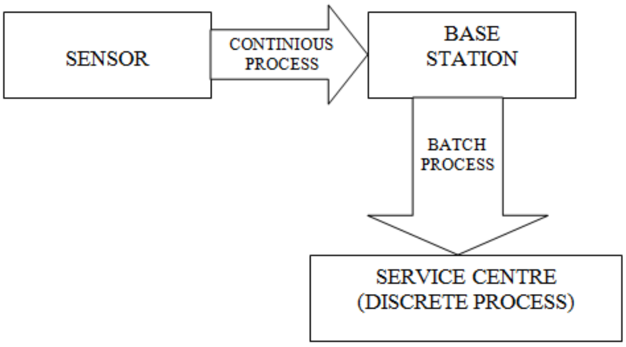
The data which multiplies quickly and it size limit crossing more than peta bytes is associated with 1st 'V-Volume'. The data movement is in very high speed network systems which is mostly applicable to all the real time scenarios of applications it's the 2nd 'V - Velocity'. The data generated from varied application posse's characteristic of usual formatted and structured data like excel sheets and database data within a file structure. The semi structured data generated in social networking sites and online newspapers. The new trend of data like images, audio, video, blogs, tweets, posts, shares & likes, chats and many more falls under a new category of unstructured data. These varied type of data come under the 3rd 'V - Variety'. The last and 4th 'V - Veracity' is used to denote unstructured, unclean and uncertain data. The big data handles all the data which possess these Four 'V - Volume, Velocity, Variety and Veracity' as it's hard to handle dataset with these characteristics in the existing database systems.

II. PROCESS CHARACTERISTICS

Understanding the process and data obtained for the process is very important for improving the quality of processing with the help of big data approach. Process can be classified into continuous process, batch process and discrete process. Applications which have all the three process are called hybrid applications.

Process in which a material or work flows more or less continuously through a planned apparatus while being manufactured is termed as continuous process. They are used to represent physical quantities which are uninterrupted by time and are smooth. The raw materials are supplied without causing any interruption to the process. The product continuously moves from one station to another after each step. Materials flow continuously from one equipment to another. It is difficult to track raw materials. Process is generally shut down when failure occurs.

Batch process is used for applications which require a specific amount of raw materials to be supplied in a single stage to produce an intermediate or end product. The product moves from one station to another only when the process completes. Each batch of raw materials being processed can be identified. It includes phase and steps. Hence a constant routine can be executed and defects can be verified.



Discrete process is involved in the production of discrete material. Robotic assembly in industries can be categorized under this process.

In this application the data received from the sensors can be categorized as continuous process as the data is received in digital value format, at a constant interrupt.

This collection of datais clustered into groups using a clustering algorithm which comes under the batch process is sent from the base station.

The data is categorized in discrete process when the analytics and prediction is performed on the collected dataset.

III. DATA CHARACTERISTICS

The data obtained from sensors can influence subsequent data analysis. The quality of the data plays an important role in outcome of the predictions made on the dataset. The data in an unstructured format may not pose an issue but inappropriate or erroneous data can lead to a faulty prediction or useless prediction. Let us analyze the critical data which will be a hindrance during analysis. Critical data may be missing values, data outliers, data drifting and communication errors.

Generally, the failure or missing of sensor will create a missing value. There are many approaches for dealing with missing value. The most commonly used approach is replacing the missing value with ‘0’ or ‘1’ or skipping the missing value. Another approach is with replacing based on multivariate statistics of the data.

Data outliers may be due to variability of data. For example, while calculating the average temperature of 50 sensors used in a room and 49 of them is between 20 and 25 degrees Celsius. But one sensor is at 175^0 C. the median of data is will be between 20 and 25^0 C. But the mean temperature will be between 35.5 and 40^0 C. In this case, outlier detection algorithm is used.

Data drift occurs during predictive analysis and machine learning. It means the drift of statistical properties of variable over time, which the model is trying to predict. There are two types - process drift and sensor drift. It is difficult to differentiate between these two drifts.

Communication errors occur during transmission of data from sensor to base station and from base station to the server, error in storage, accessing error etc.

IV. USING THE TEMPLATESAMPLE CASE - TRUCK

To measure the vehicle parameters using sensors and provide predictive analysis

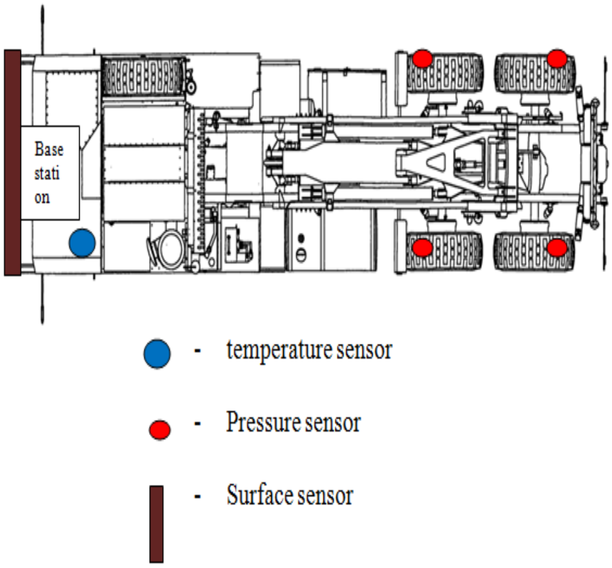
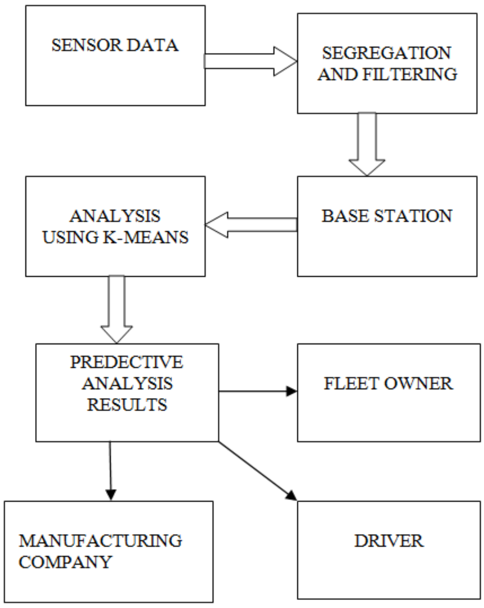


Fig. Sensor placement

The different types of sensors used are: pressure sensor, surface sensor, temperature sensor and speed sensor. The overall goal is to improve the quality of the vehicles in aspects like reducing the cost of maintenance, increase the efficiency of the existing vehicle and use this analysis to build better model in the future.



Process: Steps involved in this system are to install various types of sensor in the brand new vehicle especially for fleet operators. The sensor dataset is a combination of the entire sensor's data. The data has to be segregated and filtered according to the type of sensor. The unwanted data are filtered out. The base station consists of a memory element which can store the data temporarily and send it to the service center. The transmission from the base station to the service center takes place via GRPS. Thus, the signals are collected, clustered and is transmitted to the service center as collected raw dataset. K-means algorithm is used in clustering analysis technique for handling and exploring dataset. The predictive analysis results are sent to the manufacturing company, fleet owner and the driver.

A. Pressure Detection

The tier pressure is constantly measured using digital pressure sensor. These sensors are mounted to the rim of all the wheels. There are two types to monitor and alert in case of low pressure. They are direct measurement system and indirect measurement system. Direct measurement system measures the tier pressure whereas the indirect measurement system measures the speed of the wheel and compares the variance in wheel speed from one wheel to another. The average pressure for light weight trucks is 35psi.

The issue addressed is how often drivers would get a warning from a low tyre pressure monitoring system

1) Assume the driver would be warned when any one or more tyre's pressure level fell below the recommended pressure levels by 20%, 25%, or 30%, assuming a direct measurement system.

2) Assume the driver would be warned anytime when one or more tyres fell 6 psi (or 10 psi) below the recommended pressure, assuming a direct measurement system. Parallelly the monitored pressure will also be sent to the customer care center via base station.

There are other issues involved with Tyre Pressure. The first is environmental condition where the surface temperature causes a variation in pressure. For instance, when the temperature rises above the normal of 27 degrees, there is a proportionate increase in the tyre pressure. When the temperate falls below 15 degrees, then the tyre pressure reduces.

The load of vehicle is another important factor impacting in the change of tyre pressure; heavy loads will reduce the pressure by 2-3 psi. Hence all these parameters should be taken into consideration while performing the pattern analysis.

B. Surface Detection

The terrain can be constantly monitored using surface sensors, which are mounted on the wheels of the truck. It can detect different surface states like bumpy, wet, dry, frozen and snow covered. It can also be used for break control system and driver safety information. The data set arrived from the surface detector is used to make analytical prediction based on factors like vehicle control, fuel efficiency, tyre pressure and wear & tear of parts.

C. Temperature Detection

Temperature is a thermal scale of a body which distinguishes a hot from a cold body. It is proportionally stored molecular energy of molecules. A digital room thermometer is mounted to the vehicle to continuously monitor the temperature, which will in turn give knowledge about the environmental conditions the vehicle travels through.

V. ALGORITHM

Here the K Means algorithm with map reduce is used to segregate the obtained sensor data into different clusters so as to help ease out the analysis process. The K means together with map reduce here works on the data with the given threshold value and thus we obtain the necessary cluster centroids.

1) Cluster Setup Phase: As we are using MAP-REDUCE algorithm, the types of key and value for proposed method are as follows.

key_1 = List of initial set of selected k centroids.

$Value_1$ = List of all other nodes along with their values.

key_2 = List of new set of k centroids.

$Value_2$ = List of all other nodes with their cluster heads.

Key_3 = List of new $k' \leq k$ centroids.

$Value_3$ = List of all other nodes with new cluster heads.

2) MAP Phase: Table 1 shows the process of a Mapper. Input to the mapper is list of initial set of randomly selected k centroids which is key_1 and list of all nodes along with its as $value_1$. By using mapper (key1, value1) protocol, the Map phase produces list of new set of k centroids as key_2 and list of all other nodes with their cluster heads known to them as $value_2$.

Table 1: MAP Protocol

1 Server → N sensor nodes: Requesting node's value
2 Server: KMEANS(key_1, $value_1$)
3 Server assigns role (Cluster Head /member node) to each value
4 Each cluster sends one hop communication about the cluster to member nodes
5 Generate output key_2, $value_2$

3) REDUCE Phase: Table 2 shows the process of a Reducer. The intermediate results produced by Map protocol are given as input to the reducer i.e. list of new set of k centroids as key_2 and list of all other nodes with their cluster heads known to them as $value_2$. By using reducer (key2, value2) protocol, the Reduce phase would produce final clusters with their cluster heads and other nodes in that cluster as $value_3$.The term reduce is used in Reduce phase, which is meant for optimizing the output and not for reducing the size of the output. Our system is parallel in nature. We use a centralized MAP algorithm at BS but REDUCE is parallelized to optimize the final clusters. This parallization reduces the time of clustering the sensor network.

Table 2: REDUCE Protocol

1. Read ($value_2$) /* Build second Clustering */ 2. Place k' (k' ≤ k) nodes represented as initial cluster heads 3. Repeat a. If the member node is less than the threshold, it will start searching for better CH b. Or the cluster head is running out of values new CH will be assigned to the node. c. Update CH i.e. calculate the mean value for each cluster d. Until no change 4. Produce $value_3$

K MEANS algorithm will be called by MAP and REDUCE Protocol

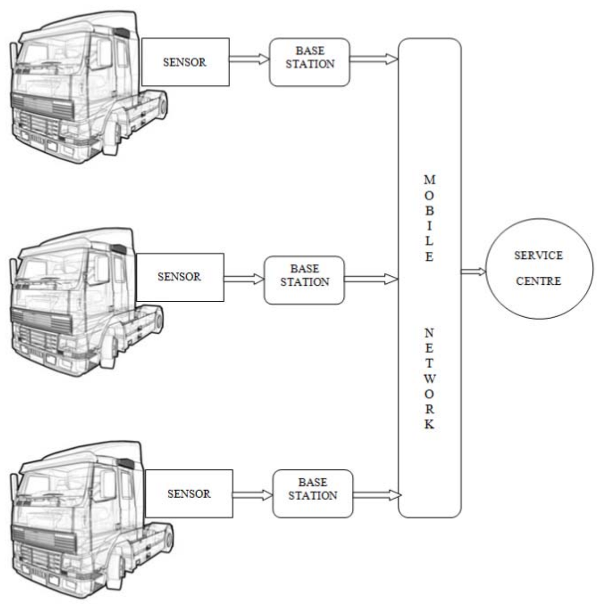
Table 3: K MEANS Protocol

1. Server will arbitrarily choose k nodes as initial cluster heads having maximum value and closer to the node 2. Repeat a. (Re)assign each node to the cluster with the nearest CH. b. Calculate the mean value of the Cluster. c. Until no change

VI. VEHICLE COMMUNICATION PLATFORM

The sensor data is sent to the base station present in the vehicle using 'k-means algorithm'. The data is clustered and need to be sent to the customer care center where it can be analyzed. VCP is used for this transmission. It consists of two units – on board unit and service center.

The On board unit acts as an base station which is connected to vehicular devices and sensors. The service centre is connected to the on-board unit through GPRS (general packet radio service) using a mobile device. The advantage of using vehicle communication platform is that it can support storage and dispatching. It provides services in he application as (i) 'software as a service'(SaaS) modality, (ii) delivery to third party applications and (iii) 'platform as a service'(PaaS)modality.



VII. CONCLUSION AND FUTURE WORK

The proposed plan can support the following applications like:

1) Fueling control and monitoring system: The system sends the fueling orders to the user as well as the service center and checks the correct order execution.
2) Transportation monitoring: The system monitors safety driving rules like working or driving overtime, overspeed or rash driving, braking, tyre slipping or skidding and brake status.
3) Driver health monitoring: The system monitors the health of the driver by having a constant check over his breath frequency, amplitude & heart pulsation and sends alarm for abnormalities co-related with behavior.
4) Vehicle maintenance: The diagnostic performance of the vehicle is kept under constant monitoring and maintenance in touch with the service workshop and also the vehicle manufacturer.
5) Arrival forecast: The system creates a prior indication to the depots, workshop or delivery place as to when the scheduled truck will check-in for delivery or maintenance.

References

[1] Vaclav´ Jirkovsky´[a,b], Marek Obitko[a], Petr Novak´[b,c], and Petr Kadera[b] , “Big data analysis for sensor time-series in automation”, IEEE, 2014.
[2] TIRE PRESSURE MONITORING SYSTEM FMVSS No. 138, U.S. Department
[3] Xuejun Ding, Yong Tian, Member, IEEE, and Yan Yu, Member, IEEE, “A real-time big data gathering algorithm based on indoor wireless sensor networks for risk analysis of industrial operations”, IEEE, 2015.
[4] Nicola Zingirian and Carlo Valenti, “Sensor clouds for intelligent truck monitoring”, 2012 Intelligent Vehicles Symposium, Alcalá de Henares, Spain, June 3-7, 2012.
[5] Byoung Uk Kim, Chris Lynn, Neil Kunst a Tom Dudgeon, “Pattern analysis in real time with smart power sensors”, IEEE, 2010.
[6] Jyoti R. Patole, “Clustering in wireless sensor network using K-means and Map reduce algorithm”, Department of computer science engineering and information technology, college of engineering, Pune, 2012.