DOI 10.29042/2018-3941-3945

Helix Vol. 8(5): 3941- 3945

Predictive Analytics for Anomaly Detection in Internet of Things Enabled Smart Cold Storage Warehousing

[*1]Mayur P. Chaudhari, [2]Dr. Manoj B. Chandak

[1, 2]Shri Ramdeobaba College of Engineering and Management, Nagpur, India

*Email: *chaudharimp@rknec.edu, hodcs@rknec.edu*

Received: 09th July 2018, Accepted: 14th August 2018, Published: 31st August 2018

Abstract

Sensors connected to the Internet of Things (IoT) generates billions of data points that can lead to real-time insights. The sensor data stream-in at an interval of one second, which is equivalent to 86400 rows of data per day. A correct predictive analytics of sensor stream information can be used to estimate missing values or to replace incorrect readings captured because of the events of malfunctioning sensors or broken communication channel. It can also be used to anticipate situations that help in various decision makings, including maintenance of a stable internal cold storage temperature and operations. In this paper, a predictive analytics for anomaly detection in the Internet of Things enabled smart cold storage warehouse is proposed using Apache Spark Service and IBM SPSS Modeller which can generate a predicted value for temperature readings. Then a set of decision rules based on both the sensed data and the predicted temperature readings is developed to generate an alert for spikes and dips in temperature. This system can be helpful to cold storage owners to achieve zero wastage of perishable food items during cold storage warehousing.

Keywords: Anomaly Detection, Internet of Things, Predictive Analytics, Smart Cold Storage

Introduction

A grain saved is a grain produced. These golden words remain as a mere proverb when one visualizes the quantum of post-harvest wastages and losses of agricultural produce due to inadequate and inefficient storage facilities. The reason for such huge post-harvest losses mainly attributes to lack of scientific storage facilities and improper transportation, poor front-end infrastructure, such as inadequate warehousing facilities. It has been estimated that 40% of fruits and vegetables grown in India gets wasted every year [1]. The major reasons are a lack of storage infrastructure clubbed with old-fashioned storing methods to stock the produced capacity [2]. Moreover, a dearth of continuous electricity and absence of any warning systems add to the troubles of cold storage owners. Evidently, there is a lack of comprehensive technology for the entire produce range of vegetables, fruits, flowers, meat and cereals [3].

The combination of many small things can make a very big difference. Sensors connected to the Internet of Things generate billions of data points that can lead to real-time insight. With the use of cutting-edge analytics tools that scale on big data, it is likely to extract significant information that supports decision makings, estimate missed values or replace incorrect readings [4]. In a predictive analytics blending, historical data and real-time sensor data with external data, such as weather data helps to predict estimate missed values, or to replace inappropriate readings due to faulty sensors or damaged communication channel.

Our proposed system uses predictive analytics techniques to make real-time predictions, learning from internal and external temperature and weather conditions to make suggestions to maintain optimal internal cold storage temperature. This could potentially be an aid for farmers for storing perishable goods over a longer period. The goal is to maintain an optimal temperature inside cold storage facility to retain the quality of food items over a longer period.

Literature Review

The term anomaly, also known as an *outlier*, eventually comes from the field of *statistics [5]*. The typical definition of an anomaly is *"an anomaly is an observation, which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism"* In IoT sensors, an anomaly can be termed as, those measurements that meaningfully differ from the usual pattern of sensed data. This definition is based on the fact that in IoT sensor nodes are designed to monitor the physical world and thus a pattern representing the normal behavior of sensed data may exist. Potential sources of anomalies in data collected by IoT sensors include noise and errors, actual events. Noisy data, as well as erroneous data, should be eliminated or corrected if possible as noise is a random error without any real significance in data analysis. According to potential sources of anomalies, as discussed above, event reporting, data reliability and secure functioning of the system will be achieved by identifying the anomalies. Specifically, anomaly detection not only controls the quality of sensed data but also improves the performance of data analysis which is under the influence of noise and defective sensors. By doing so, the impact of inaccurate data on final results can be prevented. The abnormal value sensed by a defective sensor that does not follow a regular pattern can be efficiently identified

by anomaly detection. The values detected are treated as events, which are constantly being updated indicating the recent trends of interest. Moreover, inaccurate values that are generated by malicious sensors are identified by anomaly detection [7].

Sensors collect real-valued data in a continuous manner, which is often called as data streams. If the attributes of collected data have erroneous values, the data can be recognized as an anomaly. It will be easy to detect an anomaly in the univariate data having single attribute if an attribute is anomalous to another in the sensor data. On the other hand, every sensor node can be fitted-out with multiple sensors and it is possible to have correspondence among the attributes of sensor data. Hence, anomaly detection methods for IoT are required to evaluate multivariate data and detect if the attributes display anomaly as a whole. The reason behind this concept is, at times attributes can have an anomalous value individually. Analysis of multivariate data [8], not only improves the accuracy of anomaly detection techniques but also increases computational complexity.

Due to the fact that local anomalies can be detected at individual sensor nodes, methods for identifying local anomalies save computation and improve the scalability. Local anomaly detection can be used in many event detection applications [9]. For local outlier identification used in IoT technique where every node detects the erroneous values depending on its historical values as well as each sensor node gathers readings from its neighboring nodes to collaboratively detect the abnormal values. This methodology takes advantage of the spatiotemporal correlations among sensor data and improves the accurateness and robustness of outlier detection.

An error denotes a noise-related sensed data coming from a defective sensor. Anomalies triggered by errors may occur regularly, while in other hand anomalies caused by events be likely to have very less probability of occurrence [10]. Erroneous data is generally represented as diverted from actual data and is very different from the remaining data. In the meantime, these errors affect data quality and they are required to be detected. The event is termed as a phenomenon that alters the state of real-world, e.g., chemical spill, air pollution, forest fire, etc. This type of anomalies normally continues over a longer time span and alter the historical pattern of stream data. On the other hand, defective sensors may also produce similar anomalies as events and so it is difficult to decide sources of anomalies by only inspecting one steaming series of a sensor node [11]. Thus, anomalies detection techniques required to use a data from neighboring nodes and spatial similarity of the sensor data. This is based on the fact that event measurement is likely to be spatially correlated.

US Kameswari and Prof. I. Ramesh Babu presents predictive analytics for sensor data analysis along with anomaly detection in specific to process sector [12]. In this paper, they focus on generic framework having methods like probability and statistics, Neural Network, and clustering to increase prediction and anomaly detection precision of equipment as well as a procedure flow. Putjaika et al. [13] proposed control system for smart farming in Thailand based on Arduino. The smart farming system consists of two services named as sensor system and control system. In this paper, they focus on a control system which controls the watering and roofing of an outdoor farm based on data collected from sensor systems. Based on the sensed data they develop a decision table to activate the actuators which are used in their proposed system. They resolved the issue of reducing noises in the sensors data. The limitation of this system is they didn't address the scalability issues for this model at larger scale.

Proposed System

The proposed system uses wireless sensor modules to acquire sensed values from sensor nodes. In our system instead of real sensors, we are using IBM cloud for simulation of sensors nodes. Virtual sensors are deployed in IBM IoT Watson Platform [14] which acts like real sensors and produces a data. The system monitors the internal temperature of cold storage based on sensor readings in real time and detects risky scenarios and provides early warning notifications or alarms in case of a critical condition [15]. The system can detect risky scenarios by performing predictive techniques for predicting sudden temperature ups and dips with the help of sensor stream data to maintain optimal temperature for perishable items in cold storage. The risky scenarios can be detected by using a historical sensed data of sensor node itself with use of data of neighboring node and spatial similarity of the sensor data.

The proposed work of anomaly detection system is divided into the following modules:

1) Collection of sensed temperature data using a temperature sensor.
2) Calculate predicted values using a model with the help of historical sensed data
3) Compare actual sensor readings with predicted value in the previous step for analysing trends.
4) Analyse actual sensed data and predicted values from historical data if the difference between both is beyond pre-set threshold limit pop an alert.

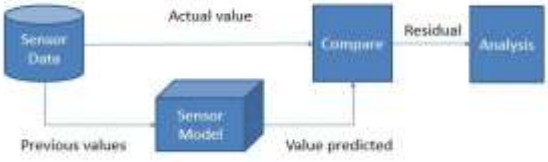

Figure 1. Proposed Workflow of the System

In our proposed anomaly detection system for cold storage warehousing, a temperature sensor keeps publishing the sensed readings in the IBM Watson IoT Platform. Multiple receivers running in the Apache Spark reads sensor readings from IBM Watson IoT Platform and makes a ReST calls to the SPSS model deployed in the Watson Machine Learning service [16]. The SPSS stream is built on top of the SPSS streaming time series expert model. Time series expert model predicts the forecast temperatures. Based on the input data, it finds the most suitable time series forecast model and trains the model automatically during the scoring time. The initial fifty data points are used for training and therefore the model can modify itself over time. The stream is deployed in Watson Machine Learning service. The service will return the next few predictions based on the input data.

When real-time data reading is received, the Spark streaming job gets the next few predictions from Watson Machine Learning service. It also calculates the Z-Score and WZ-Score to indicate the degree of difference in the actual reading compared to the predicted readings.

A Z-Score can be calculated by the following formula:

$$z = (X - \mu) / \sigma \quad (1)$$

Where z is the Z-Score, X is the value of the element, μ is the population mean, and σ is the standard deviation

Since the forecast is a trend indicator, a much bigger difference than the normal range would indicate an unexpected change in values. The WZ-Score will calculate the local Z-Score based on the window size. Since local Z-Score is only based on this window size, it will be more sensitive to the data changes. For example, a value of 10 will calculate the standard deviation based on last 10 data entries. So in this way, the WZ-Score is being used as an indicator of prediction an outside the acceptable threshold readings. Thus the WZ-Score can be used in IoT Real-Time Insights (RTI) [17] rule to determine when an alert needs to be raised. A larger value filters out smaller spikes and dips.

Implementation tools

Implementation tools used in proposed work are discussed as below:

- **IBM Watson IoT Platform:** IBM Watson IoT platform helps us to a quick start with the IoT Project. It is a completely managed and cloud-hosted service intended to make it simple to derive values from your IoT devices. It offers various services like device registration, control, connectivity, rapid visualization and storage of IoT data.
- **Apache Spark service:** Apache Spark is an open source cluster computing framework optimized for extremely fast and large-scale data processing. It allows to efficiently perform streaming, machine learning or SQL workloads that have a need for fast iterative access to datasets. Apache spark delivers a hundred times quicker performance than Apache Hadoop as a result of its advance in-memory computing engine.
- **IBM IoT Real-Time Insights:** IBM IoT Real-Time Insights is a service provided by IBM's IoT Platform, It consumes data from IoT devices and provides insights from that data through real-time contextualization and visualization. It permits us to observe equipment and operations to know and respond to rising conditions to enhance responsiveness, equipment availability, and overall efficiency.
- **SPSS Modeler:** IBM SPSS Modeler is a data mining and text analytics software application designed by IBM, used to build predictive models and conduct different analytic tasks. It has a visual interface that allows users to leverage statistical and data mining algorithms without programming.
- **Watson Machine Learning service:** Watson Machine Learning service provides predictive analytics service. It combines advanced analytics capabilities of statistical analysis, data mining, text analytics, optimization, real-time scoring, predictive modeling and machine learning. These tools help to observe various data patterns and go beyond what has happened to forecast what is probably going to occur next.
- **IBM Bluemix:** IBM Bluemix is a cloud platform developed by IBM. It provides various AI and machine learning API's and services. Bluemix is based on Cloud Foundry open technology and runs on SoftLayer infrastructure. It provides support for various programming languages and services.

Design of System

Figure 2. Various Components involved in the Integration

IBM Watson IoT Platform reads the temperature readings from the sensor and sends it to the Apache spark service and further pass it on to the Watson Machine learning service. A predictive model deployed in Watson machine learning service is a mathematical function that learns the mapping between a set of input data variables and the target variable. We can use the IBM SPSS Modeler to create Predictive models. IBM SPSS Modeler application is built by IBM for the purpose of a data mining and text analytics. It can be used to conduct various analytic tasks as well as to build predictive models. It has a graphical user interface that allows users to take advantage of statistical and data mining algorithms without any programming skills.

Deployed time series expert model predicts the forecast temperatures. Based on the input data, it finds the most suitable time series forecast model like ARIMA and exponential smoothing models and trains the model automatically during the scoring time. Apache Spark service calculates a Z-Score and WZ-Score based on prediction generated by machine learning service and pass it on to the RTI. In RTI we are able to set up rules such that the alerts are going to be generated once the WZ-Score crosses the threshold. A larger threshold values filter out the smaller ups and dips in temperature and can avoid the false alerts.

Results and Discussion

Figure 3. Predicted values in Watson IoT Platform

When the predicted values for temperature readings are sent back to Watson IoT Platform, the defined rules will analyze the WZ-Score data in real time and action will be taken by RTI when a threshold is crossed. Since a predefined threshold range is 3 to –3 and WZ-Score calculated by the system is less than -3 means that it crossed the threshold limits then the system generates an alert notification indicating a certain change in temperature inside the cold storage facility. In real-time we can able to see visualization charts for the real-time data that are coming in from the sensor devices deployed at the facility. Rules for generating an alert message can be defined by considering single or multiple attributes based on domain and requirements.

Figure 4. Alerts Generated by RTI and Sensor Information for change in Temperature Readings

The system predicts the temperature readings based on the historical data that is already been collected from previous readings. The system uses predicted values to generate alerts for slips at the optimum temperature even before the slip happens to allow cold storage owner to take necessary actions. Deploying this type of system helps cold storage owner to achieve zero wastage of perishable food items during the cold storage process.

Conclusion

In this paper, we have discussed that machine learning methods can be used along with IBM Watson IoT Platform in order to provide a proactive solution for Cold storage warehousing using the Internet of Things applications. Our proposed system is flexible in nature and is also capable to deal with dynamic environments as opposed to the contemporary methods.

Different data streams have a different type of error in prediction and this error affects the event. In future, we aim to work on the modeling of our system in order to predict more complex events occurred due to change in temperature readings. We also look forward to test our system on other Internet of Things scenarios and high-velocity data.

References

1. H. Charles J. Godfray, John R. Beddington, Ian R. Crute, Lawrence Haddad, David Lawrence, James F. Muir, Jules Pretty, Sherman Robinson, Sandy M. Thomas, Camilla Toulmin, Food Security: The Challenge of Feeding 9 Billion People, Science, 12 Feb 2010, pp 812-818

2. Rohit Joshi, D.K. Banwet, Ravi Shankar, Consumer link in cold chain: Indian scenario, Food Control 21, 2010, pp 1137–1142

3. N. Viswanadham, Achieving Rural & Global Supply Chain Excellence - The Indian Way, Center for Global Logistics and Manufacturing Strategies, Indian School of Business, Hyderabad, India, 2006.

4. Thiago Teixeira, Sara Hachem, Val erié Issarny, Nikolaos Georgantas, Service Oriented Middle-ware for the Internet of Things: A Perspective, ServiceWave, Poznan, Poland. Springer-Verlag, 2011, pp.220-229

5. V. Hodge and J. Austin, A Survey of Outlier Detection Methodologies, Artificial Intelligence Review, Vol. 22, 2003, pp. 85-126.

6. D.M. Hawkins, Identification of Outliers, Chapman and Hall, London, 1980.

7. Idris M. Atakli, Hongbing Hu, Yu Chen, Wei-Shinn Ku, Zhou Su, Malicious Node Detection in Wireless Sensor Networks using Weighted Trust Evaluation, International Journal of Information and Computer Security 3, 2 October 2009, pp 132-149.

9. Sutharshan Rajasegarar, Christopher Leckie, Marimuthu Palaniswami, Anomaly detection in wireless sensor networks, *IEEE Wireless Communications*, vol. 15, no. 4, August 2008, pp. 34-40.

10. A. Akbar, F. Carrez, K. Moessner, A. Zoha, Predicting complex events for pro-active IoT applications, Proceedings of 2015 IEEE 2nd World Forum on Internet of Things (WF-IoT2015), 2015, pp. 327-332.

11. Fawzy, A., Mokhtar, H. M. O., Hegazy, O.Outliers detection and classification in wireless sensor networks, Egyptian Informatics Journal, 2013

12. Jurdak R., Wang X.R., Obst O., Valencia P., Wireless Sensor Network Anomalies: Diagnosis and Detection Strategies, Intelligence-Based Systems Engineering. Intelligent Systems Reference Library, vol 10. Springer, Berlin, Heidelberg

13. U. Surya Kameswari, I. Ramesh Babu, Sensor Data Analysis and Anomaly Detection using Predictive Analytics for Process Industries, IEEE workshop on Computational Intelligence: Theories, Applications & Future Directions, (IEEE WCI 2015), 2015.

14. N. Putjaika, S. Phusae, A. Chen-Im, P. Phunchongharn, K. Akkarajitsakul, "A control system in an intelligent farming by using arduino technology", 2016 Fifth ICT International Student Project Conference (ICT-ISPC), pp. 53-56, 2016

15. "IBM Watson IoT Platform", Available at: https://www.raspberrypi.org/. [Accessed 14 March 2018].

16. A. Anzanpour, Amir-Mohammad Rahmani, Pasi Liljeberg, Hannu Tenhunen, Internet of things enabled in-home health monitoring system using early warning score, In MobiHealth'15, 2015.

17. "Watson Machine Learning", Available at: https://www.ibm.com/in-en/marketplace/spss-modeler [Accessed 21 December 2017].

18. "IBM IoT Real-Time Insights – Analytics designed for the Internet of Things", Available at: https://www.ibm.com/blogs/bluemix/2015/09/iot-real-time-insights/ [Accesses 05 January 2018]

