

A Review Paper on Big Data Analytics in Process Industry

Lakshmi S^{a,1}, and Manonmani A^b

^aAssociate Professor Dept of EIE ,Panimalar Engineering College

^bAssistant Professor ,Saveetha Engineering College
Chennai, India

Abstract. Today, in modern large-scale industrial processes, each step-in manufacturing produces a bulk of variables, which are highly precise in nature. However, great challenges are faced under different real-time operating conditions when using just the basic data-driven methods. One of the sultriest research points for convoluted process control is the usage of big data analytics. The aim of big data analytics is to take full advantages of the large amounts of obtained process data and mine helpful details present within. Compared to the well-developed model-based approaches, usage of big data analytics provides productive elective answers for various modern issues under different working conditions. Majority of the modelling in process control in a closed loop system is based on varying the command input to obtain desired controlled output. However, modelling of the process control in a closed loop system based on the disturbance using conventional methods is time consuming since disturbance data is too big and too complex. Utilization of advanced big data analytical methods to mine the disturbance data can lead towards more informed decisions to model the process control in the system. Thus, relevant solutions can be obtained to some of the challenges in the modeling of process control using big data analytics.

Keywords: Big Data Analytics, Modeling, Process Control, Closed loop, Process Industry

1. Introduction

An industry is a gathering of organizations related in delivering/assembling or managing a similar kind of item and administration. So, a group of smart phone producers is known as an industry. In process industries the foremost manufacturing processes are either continuous, or batch^[16] process. For example, a food manufacturing enterprise making sauce might also manufacture the sauce in a ceaseless, steady go with the flow from receipt of fixings through bundling. Examples of the process industries include food, drink, wood and wood products chemicals, petroleum, ceramics, base metals, coal, plastics, rubber, textiles, tobacco, pharmaceuticals, paper and paper products, etc.

¹Lakshmi.S, Associate Professor, Dept of EIE , Panimalar Engineering College, Chennai, India;
E-mail: elzie.moses@gmail.com.

1.1. Types of Industries

Based on Resources and Size the industries are classified into different types.

1.2. Resources

- **Agricultural-based industries:** If the industries are use plants and creature-based items as their crude materials then they are under Agro-based industries. Examples, cotton textile, vegetable oil, leather industries dairy products, and food processing.
- **Ore based industries:** If the industries are use 'mineral ore' as raw material then they are under Mineral-based industries. These industries also equipped to other industries
- **Oceanic-based industries:** If the industries are use basic materials from sea or ocean then they are under Marine-based industries. Examples, fish oil.
- **Plantation-based industries:** If the industries are use basic materials from the plantation like wood then they are under Forest based industry. The industries associated with plantation are paper, pharmaceutical, and household goods

1.3. Size

Size of ventures are relying upon how a lot of cash is contributed, representative check and merchandise created. Small-scale enterprises: Small-scale ventures contributed capital is little. There is a regularly difficult work seen here. Model, earthenware, Basket weaving, and handiworks. Large-scale ventures: Large scope endeavours are not at all like of little scope undertakings. Here the capital contributed is gigantic and front-line development is being utilized here. Examples, Heavy Machinery and Automobiles.

2. Big Data Analytics

2.1. Data

Different types of operations are performed by computers like quantities, characters, or symbols which might be transmitted and saved as electrical signals and noted on optical, attractive, or mechanical chronicle media.

2.2. Big Data

Big Data is also data but with a very large size. Collection of large data and growing exponentially with time is the Big Data. Traditional data management tools are not able to measure and store the complex huge data. But Big Data is able to measure and store complex huge data and also process it efficiently.



Figure 1. Big Data Uses

Big data indicates datasets volume, Velocity, and Variety. It is very difficult to measure, manage, process or analyze by using traditional technologies and tools, such as social databases and work area measurements or representation bundles, inside the time important to make them valuable.

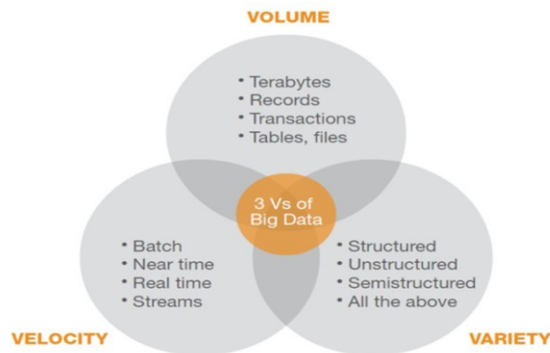


Figure 2. Characteristics of Big Data

2.3. Characteristics of Big Data

Big Data consists of three important Characteristics like, As shown in figure 2 Volume, Variety and Velocity. And also consists of three different layers as shown in figure 3. Like infrastructure layer, Computing layer and Application layer.

2.3.1. Volume

The name Big Data itself is identified with a size which is tremendous. Volume indicates Tables, Files, Terabytes, Records and Transactions. from top to bottom.

2.3.2. Variety

Variety refers to assorted sources and the nature of data. Variety includes structured and unstructured, semistructured and above all.

2.3.3. Velocity

The term '**velocity**' refers to the speed of production of data. How quick the information is produced and prepared to satisfy the necessary needs. The flow of data is enormous and continuous. Velocity includes Batch, Near Time, Real time and Streams.

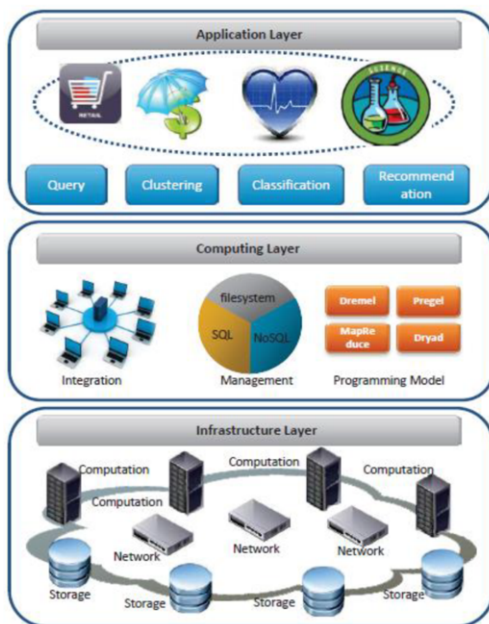


Figure 3. Big Data's Different Layers

3. Literature Review

Le Yao, Zhiqiang Ge^[14] Illustrated that in last decades Distributed Control System can be used to collect large data set but recently in process industry, manufacturing throughput and output quality related parameters like viscosity, melt indices are difficult to measure. These parameters are determined through offline laboratory analyzers or high cost analyzers. Nowadays data collection inferential model has been created to give online estimation for quality variable. It uses, a distributed parallel process demonstrating approach is introduced dependent on MapReduce structure for Big Data quality expectation. Bayesian rule is make use of MapReduce. For split dataset semi supervised probabilistic principal component regression (SSPPCR) model is developed for big data quality and effectiveness prediction. This paper can be effortlessly stretched out to numerous. If other quality expectation model can be used then it requires learning with huge scale dataset.

G.M. Stanley^[13] enhanced knowledge in advance control technology uses various storehouses of usefulness applications by and large require separate algorithms for model distinguishing proof, control and furthermore utilizes a state estimator, for example, Kalman filter. This paper presented another methodology, bypassing these

storehouses, taking care of estimation control issues all the while by design coordinating or guess legitimately on a web-based preparing set dependent on timeseries data. BigData approximating control (BDAC) is utilized for estimation and control issue. It plays out a preparation procedure and estimation and control issues by surmised design coordinating legitimately on the preparation set. Continuous exponential group sifting (RTECF) and closest neighbor search and bunching are utilized right now. Control strength is a noteworthy issue. In this way, this might be addressable through improving the preparation procedure. Extra grouping or various leveled bunching for BDAC-O will be attractive to all the more likely help exceptionally non-linear framework.

Shu Xu, Bo Lu^[12] used support vector regression (SVR) analysis infrared spectroscopic data. In chemical industries, pharmaceutical and semi-conductor manufacturing enterprises expanding interest for higher item quality. Analytical instruments like chromatography (GC, HPLC), Near infrared (NIR), spectroscopy joined with fiber optic, online estimation tests and chemometric models, for example, Partial Least square (PLS) developed as a robust tool for process observing and quality control in process industry. In this paper, they introduced support vector regression (SVR) analysis – Infrared spectroscopic data. An option in contrast to the traditional technique, SVC acquainted with improve execution what's more, its quality against nonlinear info yield relationship. Recursive component disposal with common data-based predisposition remedy strategy utilized in SVR. Along these lines, this strategy improving model viability and interpretability. The subsequent frequency recurrence extensively expanded.

Jing Yu, Victor M. Zavala, Mihai Anitescu^[11] presented inDynamic system, the sensor position issue area, and kinds of sensors that would expand data. This data can be communicated regarding the back-covariance system of the state or parameters of the framework, the issue can be as often as possible be given a job as a perfect structure of tests issue. give a role as an ideal structure of analyses issue. To explain this issue, adaptable structure of investigations structure for sensor situation in framework portrayed by partial differential conditions (PEDS). In PEDS, we can computationally interact mixed-integer in linear system. In natural gas pipelines this sum-up rounding approach given better results.

Yining Dong, S. Joe Qin^[10] utilized for information displaying and process observing is critical in present day technique. For this Principal Component Analysis (PCA) and Partial least square (PLS) are effective diagnostic instruments. For this paper depicted around, a novel dynamic PCA (DIPCA) calculation is acquainted with extricate dynamic dormant factors. Geometric properties are explored to give information into the new amazing model structure. For process checking, flaw recognition records dependent on DIPCA are created dependent on handled model. A corn endorsement system is used to choose the solicitation and the amount of dynamic inert factors. Thus, this DIPCA displaying strategy furnishes approach to discreate dynamic varieties with idle factors from static varieties. After partition deficiency location are made.

Ray Wang, Thomas F. Edgar, Michael Balde^[9] Exhibited in pharmaceutical and microelectronics, the batch process is utilized for creating of high worth, commonly conventionally low-volume materials. The financial expense of this procedure is high, for batch process checking, deficiency location and control. This paper introduced a geometric strategy for perception, information driven procedure observing and issue discovery for bunch process framework. Multi-way Principal Component Analysis

(MPCA) presented in group insightful unfurling of the three-dimensional (time, variable, batch) clump information into explicit two-dimensional network. Dynamic Principal Component Analysis (DPCA) is presented for examination of transient nonstop procedure for group information. For representation bunch information, head segment examination or halfway least square relapse (PCA/PLS) can be utilized. Time-unequivocal kiviak chart presented for flaw location.

Q. Peter He, Jin Wang ^[2] Here author illustrated advance figuring force and development of unwired open advances, data based advances, information examination are increasingly effective, beneficial and more astute. There are various names to indicate the cutting-edge fabricating frameworks, for example, shrewd/assembling and industry 4.0 and so forth. Industry Internet of Things and high throughput instruments it is normal that the information produced from assembling will develop exponentially. Right now, guide of Statistical process monitoring (SPM) is exhibited. Guide of factual procedure isolated the advancement of SPM in to three times. first period Statistical process control (SPL), second time multivariable Statistical process monitoring (MSPM), third time appropriately characterized. In exquisite creation, enormous information likewise gives many opportunities to process checking like element recognizable proof and extraction, self-versatile displaying, prescient upkeep and self-remedy of sensor shortcoming.

Bo Lu, Leo Chiang^[1] presented today's manufacturing environment demands are accuracy, dependable computation and forecast of standarder parameters. These are critical success factors in demand today. In chemical manufacturing industry, it isn't just to examine the information disconnected, yet in addition to deal with the information online to improve process activity. Information driven delicate sensors dependent on projections to dormant structure (PLS) strategy are utilized in synthetic procedure. Right now, two key execution pointers (KPIs) for evaluating the presentation of the present sensor model. The Hoteling's T2 based KPI and expectation and leftover based KPI are improving gauge execution without redoing the PLS model. The model update framework can be combined with both KPIs pointers.

Michael C. Thomas, Wenbo Zhu, Jose A. Romagnoli^[4] This investigation looked to fill the gap between the requirement for named information for preparing regulated observing calculations and the crude, unclassified information that have collected in process chronicled databases. We exhibited how unsupervised learning techniques worn from the software engineering literature can identify fault states and extract knowledge from chemical process databases. A determination of dimensionality reduction and data clustering techniques identified different operating and fault states in data from real and simulated chemical processes. On the Tennessee Eastman Process simulation, information grouping and DR functioned in tandem to detach data from the different faults in the process, data which could be used to train a supervised classification technique for fault detection. On an industrial various data to normal process operations and distinguished it from data from a fault that had occurred during the months of operations considered. Further investigation uncovers that at the evaluation considered each start-up of the segment framed another group in the dimensionally decreased space, as imagined by a PCA [18-19] projection of the bunches found by DBSCAN.

TABLE 1. Compression between Issues and Algorithms

Issues					
Non-Linearit y	Non- Gaussian Distribut ed variables	Stationa ry Process	Gaussia n Distribu ted variable s	Dynami c Process	Fault Deduction
		PCA			
		PLS			
		DPCA			
		DPLS			
SVM					
	ICA				
		CVA			
		FDA			
BCDA					

4. Conclusion

We have entered in to time of Big Data. This paper narrates the different types of process industries parameter measuring, storing and analyzing with Big Data concepts. Some technical challenges like Non linearity, Dynamic process, Static process Fault detection, Online parameter measurement, Error handling, Envision, at all phases of the examination pipeline from data acquisition to result interpretation. This paper also focuses on different process methods in industry and different specialized difficulties tended to for proficient, quick preparing in Big Data.

References

- [1] Bo Lu, Leo Chiang*,Semi-supervised online soft sensor maintenance experiences in the chemical industry.Journal of Process Control 67 (2018) 23–34
- [2] Q. Peter He, Jin Wang, Statistical process monitoring as a big data, analytics tool for smart manufacturing. Journal of Process ControlJournal of Process Control. Volume 67, July 2018, Page 35-48.
- [3] Shriram Gajjar , Murat Kulahci , Ahmet Palazoglua .Real-time faultdetection and diagnosis using sparse principal componentanalysis.Journal of Process Control. Volume 67, July 2018,Page 112-128.
- [4] Michael C. Thomas, Wenbo Zhu, Jose A.Romagnoli Data mining and clustering in chemical process databases for monitoring and knowledge discovery. Journal of Process Control. Volume 67, July 2018,Page 160-175.

- [5] Qiang Liu, Qinqin Zhu, S. Joe Qin, Tianyou Chai, "A dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes", *Journal of Process Control*. Volume 67, July 2018, Page 12-22.
- [6] Chihang Wei, Junhui Chen, Zhihuan Song, "Multilevel MVU models with localized construction for monitoring processes with large scale data", *Journal of Process Control*. Volume 67, July 2018, Page 176-196.
- [7] José Daniel A. Santos, Guilherme A. Barreto, "Novel sparse LSSVR models in primal weight space for robust system identification with outliers", *Journal of Process Control*. Volume 67, July 2018, Page 120-140.
- [8] A. Sadeghian, O. Wu, B. Huang, "Robust probabilistic principal component analysis based process modeling: Dealing with simultaneous contamination of both input and output data", *Journal of Process Control*. Volume 67, July 2018, Page 94-111.
- [9] Ray Wang, Thomas F. Edgar, Michael Baldea,*, Mark Nixon, Willy Wojsznis, Ricardo Dunia, "A geometric method for batch data visualization, process monitoring and fault detection", *Journal of Process Control*. Volume 67, July 2018, Page 197-205.
- [10] Yining Dong, S. Joe Qin, "A novel dynamic PCA algorithm for dynamic data modeling and process monitoring", *Journal of Process Control*. Volume 67, July 2018, Page 1-11.
- [11] Jing Yu, Victor M. Zavala, Mihai Anitescu, "A scalable design of experiments framework for optimal sensor placement", *Journal of Process Control*. Volume 67, July 2018, Page 44-53.
- [12] Shu Xua, Bo Lub, Michael Baldea, Thomas F. Edgar,*, Mark Nixon, "An improved variable selection method for support vector regression in NIR spectral modeling", *Journal of Process Control*. Volume 67, July 2018, Page 83-93.
- [13] G.M. Stanley, "Big Data Approximating Control (BDAC)—A new model-free estimation and control paradigm based on pattern matching and approximation", *Journal of Process Control*. Volume 67, July 2018, Page 141-159.
- [14] Le Yao, Zhiqiang Ge, "Big data quality prediction in the process industry: A distributed parallel modeling framework", *Journal of Process Control*. Volume 67, July 2018, Page 1-13.
- [15] C. Esakkiappan, T. Thyagarajan, "Identification of inverse response process with time delay using relay feedback test", *Int. J. Computer Applications in Technology*, Vol. 44, No. 4, 2012.
- [16] Dr. R.Senthil B.Muthuraj, Dr.V.Mahesh, "Simulation Of Three Tank System Using Hybrid Petri Net For Batch Process" 2015/11/2 *International Journal of Applied Engineering Research* 4789-4798 Research India Publications.
- [17] B Muthuraj, V Mahesh, R Senthil, "A Hybrid Petri Net Approach for Polyethylene Terephthalate Bottle Manufacturing System", *Journal of Computational and Theoretical Nanoscience*, 2016.
- [18] V.D.Ambeth Kumar, V.D.Ashok Kumar, Dr.S.Malathi and P.Jagadeesh, (2014) "Intruder Identification using Footprint Recognition with PCA and SVM Classifiers" for the *International Journal of Advanced Materials Research* Vols.1345, PP 984-985 (2014) pp 1345-1349.
- [19] V.D.Ambeth Kumar and Dr.M.Ramakrishnan (2012) "Footprint Recognition with COP Using Principle Component Analysis (PCA)" in the month of June for the *Journal of Computational Information Systems (JCIS)* Journal Volume 3, 4939-4950, June 2012.