

## Big data research and application in oil pipeline inspection

MuJun Wan

Department of information technology, ZhongShan Torch Polytechnic,

Zhongshan GuangDong, China

42397201@qq.com

**Keywords:** Big data, RFID, pipeline inspection, intelligence.

**Abstract.** Data mining is a process to find hidden useful information with potential value in huge amount of data. We present a new big data processing system with the data from Oil pipeline inspection. The data mainly get from all kinds of Radio frequency identification (RFID) chips. The system consists of data collecting subsystem, data preprocessing subsystem, data intelligent predict subsystem. The method we use is intelligent algorithm and K-means Clustering. By analyzing the evaluation result, we can give oil carriers suggestions about their oil pipeline design and security prediction.

### Introduction

In the oil and gas industry, success hinges on timely and error-free delivery of product through a complex, high value supply chain. In this capital intensive business, proper infrastructure maintenance as well as inventory and asset management are not just regulatory issues, they're mission critical for business health—and employee safety. Inspections, routine maintenance, and repairs must be performed on time on oil rigs, pipelines and downstream facilities to not only prevent the disastrous ramifications of a failure—but also to maximize the efficiency and lifespan of key infrastructure investments. However, the typical paper-based processes that are typically utilized to track assets, workers, asset maintenance histories and even the production and distribution of product are labor-intensive, costly, present significant opportunity for error and often result in the very slow movement of information to key decision makers.

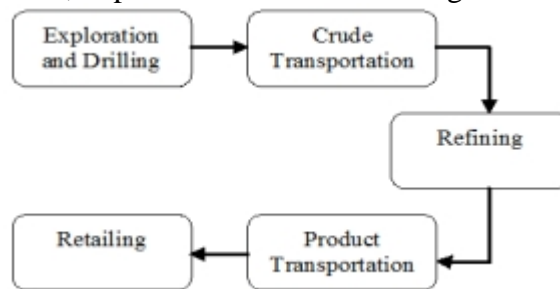
Where barcode data capture technology falls short in this industry, Radio Frequency Identification (RFID) succeeds. RFID provides the automated end-to-end visibility needed to streamline every day tracking of all materials, equipment and personnel in the oil and gas industries. From the offshore oil rig to the plant, RFID gives companies a tool to maximize uptime, respond faster to day-to-day business needs, restore production more rapidly in the event of an incident, and better protect employees.

With the rapid growth of data amount, data mining technique has become a significant tool in discovering patterns in large data set. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.[1] It can be used in many fields such as finding potential customers in marketing, DNA sequence analysis and web log analysis. Educational data mining is the area of science where various methods are being developed for making discoveries within data. This data is obtained from an educational background.[2]

### Problem and system analysis

Faster Inspections Equipment used in oil and gas production facilities is subject to rigorous inspection procedures. Regular inspections ensure that the equipment is operating normally, and help ensure the safety of the facility and the workers. With manual processes, it can be difficult to correctly identify each of these items (which can include dozens or even hundreds of identical couplings and valves, for example) and update the inspection record. Fujitsu's RFID solutions can provide fast, accurate identification of assets and equipment to speed the inspection process.

Reference most common of real networking defined, can push performances oil and gas production real networking system defined: through sensing, and RF, and communications, technology, on oil and gas wells, and measurement between, and oil and gas station library, and oil and gas tube network, production object for full of perception, achieved production data, and equipment state information in production command center and the production control center concentrated management and control, built specification, and unified of data management platform, support oil and gas production process management, further improve oil and gas field production decision of timely sex and accuracy. According to oil and gas production requirements, urgent need to quicken the construction of information systems oriented to production process through the integration of information technology and industrial production, improve the automation of production operations to ensure sustainable, safe, stable and efficient operation, in order to optimize production and management processes, and optimize the organization structure, implementation of lean management conditions. [3]



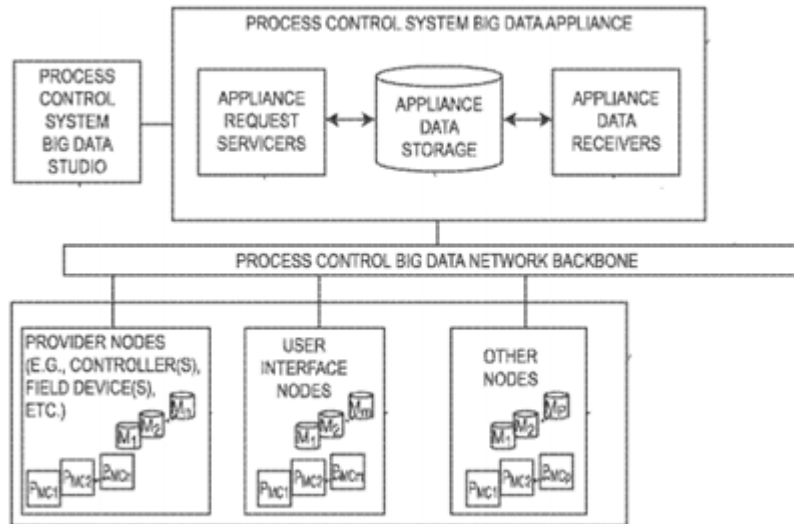
**Figure 1 Oil and Gas Supply Chain**

As shown in Fig1, Almost all oil and gas sector and other industries facing asset tracking problems. Certain assets because of its size and geographic range and it is difficult to label, harder to track. Active RFID technology can be used in any environment to monitor and manage inventory and fixed assets. RFID system is composed of "IC Tag", "reading write device/equipment (antenna + controller)", "advanced equipment such as PC," these 3 elements. "IC Tag" by the IC chip and antenna can save large amounts of data, and under "advanced devices such as PC" instructions, saved in IC chips, the data can be "read write device/equipment" to read or write. RFID systems are seen as a substitute for barcode generation technology, its biggest difference with barcode "can not intervene manually, automatically multiple reads simultaneously." [4]

Previous BRE and ICT Carrier reports have identified that electronic tagging, handheld devices, and wireless technologies will have a significant role in improving the construction industry supply chain. As previously mentioned above, these technologies have been successfully used in other industrial sectors. Similar improvements are possible within the off site manufacturing sector. The multi-barrier containment system is designed so that, after closure, it does not rely on the actions of future generations to ensure safety. The multi-barrier containment system is comprises engineered barriers and a natural barrier.

### **System design and testing**

RFID intelligent solutions are concerned by a growing number of people! Second, the system uses industry: RFID technology can be used to ensure that the interface is properly placed in the oil supply chain pipeline. Bar codes have been used and will continue to be used for these, but RFID technology has proven to be a more reliable way, to ensure proper parts and torque pressure used during the Assembly process. RFID tags can withstand a variety of harsh conditions and in the bar code can still be used after being washed or worn.



**Figure 2 system architecture**

FIG. 2 is a block diagram of an example big data network for a process plant or process control system. The example process control system big data network includes a process control system big data apparatus or appliance, a process control system big data network backbone, and a plurality of nodes that are communicatively connected to the backbone. Process-related data, plant-related data, and other types of data may be collected and cached at the plurality of nodes, and the data may be delivered, via the network backbone, to the process control system big data apparatus or appliance for long-term storage and processing. In an embodiment, at least some of the data may be delivered between nodes of the network, e.g., to control a process in real time.

Sample type	Sample No.	Well No.	Relative parameters for formation flow capacity <sup>a</sup>							y*	
			x <sub>1</sub>	x <sub>2</sub>	x <sub>3</sub>	x <sub>4</sub>	x <sub>5</sub>	x <sub>6</sub>	x <sub>7</sub>	FFC <sup>b</sup>	OLC <sup>c</sup>
Learning samples	1	TK631	35.6	133.98	6	9.56	0.96	0	8.5	19.50	3
	2	S65	8.0	298	8	7.3	0.95	0	0.29	23.00	3
	3	TK413	78.5	81.57	4	7.8	0.96	68	0.08	3.90	4
	4	TK404	78.0	219	6.9	11.69	0.95	43.9	0	67.77	2
	5	TK409	28.3	160	7.9	5.16	0.96	52.9	0.1	26.07	2
	6	S67	11.7	413	8	10.35	0.97	0	0.2	102.50	1
	7	S74	22.6	136	6	9.28	0.98	22	0	82.11	2
	8	TK609	25.5	116	7	6.1	0.96	0	7.41	21.21	3
	9	TK313	1.9	201.7	6	17	0.90	88	0.5	126.00	1
	10	TK607	24.6	203.3	8	9.6	0.95	37.78	0.03	191.31	1
	11	TK444	9.2	197.9	6	8.75	0.99	0	0	130.19	1
	12	TK629	22.6	83.4	6	4.3	0.98	0	0.11	6.993	3
Prediction sample	13	TK442	17.9	232	7	7.5	0.98	0	7.13	(4.80)	(3)

**Table 1 Sampling data set**

Any type of data related to the process control system may be collected and stored at the process control system big data appliance. In an embodiment, process data may be collected and stored. For example, real-time process data such as continuous, batch, measurement and event data that is generated while a process is being controlled in the process plant (and, in some cases, is indicative of an effect of a real-time execution of the process) may be collected and stored. Process definition, arrangement or set-up data such as configuration data and/or batch recipe data may be collected and stored. Data corresponding to the configuration, execution and results of process diagnostics may be collected and stored. Other types of process data may also be collected and stored.

Positive improvement reconciling actual inventory versus recorded inventory Electronic inspection means instantaneous updates on relevant inspection categories – before, manual, paper-based hard copy reports were sorted through when having to demonstrate to inspectors – both internal and external – that inspections have been performed in Table 1.

We can see that the overall result of our data grades as A, attendance as High, and practical-marks as 0-10. The first cluster i.e. cluster 0 has overall result as Good with exam-marks as 80-100, term work grades as A, attendance as High, and practical-marks as 11-20. However, the second cluster i.e. cluster1 shows overall result as Poor with exam-marks as 0-40, termwork-grades as C, attendance as Low, and practical-marks as 0-10. Thus, based on the results we can easily group the students especially in cases where large academic records are present in Table 2.

Algorithm	Fitting formula	Mean absolute relative residual			Dependence of the predicted value (y) on independent variables ( $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ ), in decreasing order	Time consuming on PC (Intel Core 2)	Solution accuracy
		$\bar{R}_1(\%)$	$\bar{R}_2(\%)$	$\bar{R}'(\%)$			
C-SVM	Nonlinear, explicit	0	0	0	N/A	5 s	Very high
NBAY	Nonlinear, explicit	34.72	66.7	50.7	N/A	<1 s	Very low
BAYSD	Nonlinear, explicit	0	66.7	33.3	$x_1, x_7, x_4, x_3, x_5, x_6, x_2$	1 s	Very low
MRA	Linear, explicit	4.17	33.3	18.8	$x_2, x_4, x_1, x_5, x_3, x_6, x_7$	<1 s	Strong nonlinearity

Table 2 Comparison among the clustering algorithms

## Conclusions

In this paper, we have presented a new, effective and efficient method for dealing with large pipeline inspection data streams. We have described a real-time evolving data stream that we have developed and deployed. The system could integrate real-time data from diverse domains. From these results, simple data classification schemes may be automated to choose an effective algorithm for a given data stream. We have shown that it is feasible to build and deploy a real-time data system that enable scientists to access data without being aware of where and how the data are gathered. The derived criteria are subject to further refinement, but for robust monitoring, a balance should be maintained between improved sensitivity and over-fitting.

## Acknowledgements

This research is supported by National Natural Science Foundation of China (799700080), Science and Technology Planning Project of ZhongShan, Guangdong Province, China (20093A177) and Youth Research Project of ZhongShan Torch Polytechnic.

## References

- [1] Z. Ren, C. J. Anumba, J. Tah, "RFID-facilitated construction materials management (RFID-CMM) - a case study of water-supply project", *Advanced Engineering Informatics*, Vol.25, No.2, pp. 198-207, 2011.
- [2] C. Chao, J. Yang, W. Jen, "Determining technology trends and forecasts of RFID by a historical review and bibliometric analysis from 1991 to 2005", *Technovation*, Vol.27, pp. 268-279, 2007.
- [3] S. Jiang, M. J. Skibniewski, Y. Yuan, C. Sun, Y. Lu, " Ultra-wide band applications in industry: a critical review", *Journal of Civil Engineering and Management*, Vol.17, No.3, pp. 437-444, 2011.
- [4] W. Lu, G. Huang, H. Li, "Scenarios for applying RFID technology in construction project management", *Automation in Construction*, Vol.20, No.2, pp. 101-106, 2011.
- [5] J. Lee, J. Song, K. Oh, N. Gu, "Information lifecycle management with RFID for material control on construction sites", *Advanced Engineering Informatics*, Vol.27, No.1, pp. 108-119, 2013.
- [6] K. Domdouzis, B. Kumar, C. Anumba, "Radio-frequency identification (RFID) applications: a brief introduction", *Advanced Engineering Informatics*, Vol.21, No.4, pp. 350-355, 2007.