RESEARCH ARTICLE OPEN ACCESS

Employing Outliers Detection to Improve Machine Maintenance Performance

Taghrid Zarif Ali [1], Ayman Yousef [2], Hatem Mahmoudi [3]

Postgraduate Student [1], Associate Professor [2] & [3]
Department of Design and Production Engineering
Faculty of Mechanical and Electrical Engineering
Tishreen University
Lattakia - Syria

ABSTRACT

Maintenance has become an added value which has an important role in increasing the life of equipment, improving production and increasing profits. This is done through good maintenance management and the use of innovative methods and strategies of all types according to the type and field of work. The aim of this study is to evaluate the effect of anomalies on the clustering of the data recorded at the LICT(Lattakia Port International Container Terminal) by analyzing it to improve the factors affecting predictive maintenance according type and time of maintenance. K-means-based clustering outlier detection was used and experimental shows that removing detected outliers records will improve the quality of data clusters, which make it easy to group maintaining records into group that can be analyzed as phenomena to improve maintenance performance

Keywords :— outlier detection, clustering, predictive maintenance, breakdown, K-means

I. INTRODUCTION

The success of any facility depends on the continuity and integrity of the machine's production processes and the way the maintenance process determines how long the equipment will continue to operate in a healthy and productive condition. Using correct maintenance strategies, the maintenance management system will shift from the post-crash handling to the control all equipment stops become planned, not sudden or emergency, and in this situation it becomes the maintenance department that controls the machines and equipment instead of being controlled by the latter.

II. MAINTENANCE CONCEPT

Maintenance has been known for a long time, but its concept has changed according to time and place, as well as its methods and concepts of organization

There are many definitions of maintenance, all focused on the following matters:

1 - Maintenance Activity or group of activities.

2 - Maintenance aims to return machines that have a failure to their initial condition.

3 - includes detection of faults and work to avoid these failures.

Based on the above, maintenance is defined as the combination of all the technical and administrative efforts in the establishment in order to maintain or restore the equipment and machinery in the establishment to the condition that enable these machines and equipment to perform the required functions effectively and efficiently during their life cycle.[1]

So, the maintenance work is no longer a secondary expenditure or function controlled by the nature of the reform, and its concept is limited to the maintenance and care of machinery and equipment by lubricating, cleaning and waiting for the faults to be repaired. but the maintenance also followed the rapid technological developments that entered the fields of industrial and non-industrial, and becomes on of the core activities in the enterprise that increases the competitiveness of the enterprise and becomes a strong partner in all activities undertaken by the entity to achieve its strategic objectives.

A. Maintenance Objectives:

1- Reducing failure and reducing downtime.

2. Improving the efficiency, reliability and availability of equipment.

3. Reduce costs and improve inventory management.

4. Improve useful life of equipment.

5. Control and monitor the budget and improve the use of resources.

6. Highest production, reduce energy consumption, and increase in profits.

B. Maintenance approaches and strategies

The maintenance strategy includes decisions regarding the inspection, identification and implementation of the processes of switching and repair of the means of production, the formulation of plans and the tour of maintenance activities. The maintenance of the machines can be carried out in many ways. However, the results will be different. so the right decision must be made to choose best strategies the complete needed goals. [2]

Corrective maintenance (CM): Corrective maintenance is the maintenance performed on machines after the failure to return machines to the condition that enables them to perform their required functions. This type of maintenance is the most common and used in many industrial facilities, used in installations where the failures are not critical and dangerous.

*Preventive Maintenance (PM):*This maintenance is performed prior to the failure occurrence to prevent or minimize the possibility of stop or failure of equipment, Preventive maintenance is carried out based on pre-planned time periods (e.g. number of hours, number of cycles, time periods). These periods are specified by many statistical methods to improve maintain ace scheduling

Predictive Maintenance (PDM):Predictive maintenance is a set of activities that detect changes in the physical condition of equipment over time (failure signs) in order to perform the appropriate maintenance work to achieve the longest service life of machines without increased risk of failure, or defined as maintenance based on the prediction of failures, It's main function is to gather equipment historical data and parameters values for running equipments such as time, density, flow, pressure, vibration, temperature, voltage, electrical resistance, and through its study and analysis, the potential problems are identified in the future.

Total Productive Maintenance (TPM): Total productive maintenance is the total continuous improvement of the machines and the participation of the workers. The most important objectives are to increase the readiness of the machines and improve the quality of production, and to implement an integrated system of preventive maintenance with the participation of all employees in the establishment.

Reliability-based maintenance: considers that both availability, machine reliability, and operational knowledge and expertise are important and vital in their concept, their goal is to identify a maintenance approach that is appropriate for each important part of the machine to ensure that the work is performed properly.[3]

III. RELATED WORKS

Realizing the Full Value of Industrial Internet and Data Analytics in LNG Facilities.[4]

This paper stated that one of the key steps in the graphed anomaly detection is the clustering of variables for anomaly detection. Statistical methods like k-means and k-medoids can be used for clustering. If design data is available, that can be used to set up clusters based on equipment. Alternately, the clusters can also be developed utilizing the variables selected based on process knowledge.

Data Driven Modelling for System-Level Condition Monitoring on Wind Power Plants.[5]

This paper presents an innovative approach to condition monitoring of wind power plants, that provides a system-level anomaly detection for preventive maintenance, The goal of cluster analysis is to partition data points into different groups. Similarity of points is defined by a minimal intra-cluster distance, whereas different cluster aim for a maximum inter-cluster distance. Thus, cluster analysis can be utilized to find the pattern of a system direct using the multi-dimensional data without explicit descriptions about the system features. This is the main advantage in using cluster analysis for modelling complex systems with seasonal components, e.g. WPP. In the presented solution, a system model for anomaly detection should characterize the normal system behaviour and can be used to identify unusual behaviour

Anomaly Detection and Performance Evaluation of Mobile Agricultural Machines by Analysis of Big Data[6]

In unsupervised methods, no target variable is identified as such. Instead, the data mining algorithm searches for patterns among all the variables , Consequently, the unsupervised approaches can deliver indications of potentially useful patterns and dependencies within the analyzed parameter space although it is not known a priori which interdependencies – if any - might exist.

IV. RESEARCH DATASET

Data was obtained from the Lattakia International Container Terminal (LICT). According to Excel tables, there were more than 31000 record, breakdowns and preventive maintenance (PM) for the most of the machines working in the LICT, data consists of number of fields that hold LICT's CMMS(computerized maintenance management system), this fields include repair request open and close date and time, repair type, machine part, maintenance jobs and other data fields. Repaired machines parts hierarchy fields were labelled from l1 to l5, l6 is the repair time.

Attention was given to the maintenance of BD faults, fault levels and repair time. Initial processing of the data was done by filtering and classification of failures, the machine (STRADDLE HARBOR CARRIER) was considered due to the independence of different machine failures. The aim of the research was not to evaluate the maintenance quality of the maintenance department as a whole.

V. ANOMALY DETECTION

In this section primary works in the area of anomaly detection are categorized, and for each category prominent works in the literature are reviewed.

There are six main categories of anomaly detection algorithms: rule-based, statistical, proximity-based, Artificial Immune System(AIS), supervised and unsupervised methods. Each method is discussed including an analysis of their strengths and weaknesses.

A. Rule-based Methods

We find frequent item-sets to extract anomalous flows from a large set of flows observed during a time interval. The standard algorithm for discovering frequent item-sets is the Apriori algorithm.[7]

B. Statistical Methods

Statistical approaches are the standard algorithms applied to outlier detection. The main aim of these approaches is that normal data objects follow a generating mechanism and abnormal objects deviate through generating mechanism. Given a certain type of statistical distribution , algorithms compute the parameters assuming all data points are generated by such a distribution (mean and standard deviation). Outliers are points that possess a low probability to be generated by the entire distribution (deviate greater than 3 times the standard deviation from the mean). These methods have the limitation that they will assume the data distribution . Another limitation of the statistical methods is that they don't scale well to large datasets or datasets of large attributes.[8]

C. Proximity-based Methods

In a proximity based method an object is an anomaly if the nearest neighbors of the object are far away, i.e., the distance of the object is significantly large from the distance between most of the other objects in the same data set. Two major types of proximity-based anomaly detection techniques are distance-based and density-based methods. A top-m distance-based anomaly in a data set is an object having weight greater than the m(th) largest weight, where the weight of a data set object is computed as the sum of the distances from the object to its m nearest neighbours.[9]

D. Supervised Approaches

Techniques trained in supervised mode assume the availability of a training data set which has labeled instances for normal as well as anomaly class. Typical approach in such cases is to build a predictive model for normal vs. anomaly classes. Any unseen data instance is compared against the model to determine which class it belongs to. [10]

E. Auto encoders

F. Unsupervised Approaches

Unsupervised anomaly detection is the learning scenario where we are given just one data set that is a mixture of normal and anomalous instances, none of which are labelled, and our task is to identify the anomalous instances.[11]

In our research we used cluster based approach which aims to arrange similar objects into groups is referred to. Clustering based anomaly detection techniques operate on the output of clustering algorithms, e.g. the well-known K-means algorithm.

They assume that anomalous instances either lie in sparse and small clusters, far from their cluster centroid or that they are not assigned to any cluster at all nearest-Neighbour.[12]

Many distance rules can be used to assign point to cluster as following equations shows, in our research we used Euclidian distance which has low cost for computations.[13]

Dice:

$$sim(t_i, t_j) = \frac{2\sum_{h=1}^{k} t_{ih} t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2}$$

Jaccard:

$$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{\sum_{h=1}^{k} t_{ih}^2 + \sum_{h=1}^{k} t_{jh}^2 - \sum_{h=1}^{k} t_{ih} t_{jh}}$$

Cosine:

$$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{\sqrt{\sum_{h=1}^{k} t_{ih}^2 \quad \sum_{h=1}^{k} t_{jh}^2}}$$

Overlap:

$$sim(t_i, t_j) = \frac{\sum_{h=1}^{k} t_{ih} t_{jh}}{\min\left(\sum_{h=1}^{k} t_{ih}^2 \quad \sum_{h=1}^{k} t_{jh}^2\right)}$$

Euclidean:

$$sim(t_i, t_j) = \sqrt{\sum_{h=1}^{k} (t_{ih} - t_{jh})^2}$$

Manhattan:

$$sim(t_i, t_j) = \sum_{h=1}^{k} \left| (t_{ih} - t_{jh}) \right|$$

Following pseudo code shows clustering algorithm used.[14]

Input: Dataset D
Output: Set of clusters C
initialize the set of clusters, C to $\emptyset$
Phase 1:creating clusters
for $d \in D$
  for $c \in C$
    if distance *(d, c)* $\leq$ *W, assign d to c*
  if d is not assigned
    create cluster c' with d as the centroid and add c' to C
Phase 2: Assigning data points to additional clusters
for $d \in D$
  for $c \in C$
    if distance *(d, c)* $\leq$ *W, and d is not assigned to c*
      assign d to c

Standard deviation features vector and the average for given dataset can be calculated as follows:[15]

$$\text{avg\_vector}[j] = \frac{1}{N} \sum_{i=1k=0}^{N} \text{instance}_i[j]$$

$$\text{std\_vector}[j] = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (instance_i[j] - avg\_vector[j])^2}$$

$$new\_instance[j] = \frac{instance[j] - avg\_vector[j]}{\text{std\_vector}[j]}$$

The following pseudo code describes cluster evaluation of given data set after clustering step is performed..[16]

Input: set of training instances X
Input: EM model M
Output: Set of cluster mean probabilities P
Output: Set of cluster standard deviations S
For all instances x_i do
  Apply M to x_i to obtain probability distribution d
    Select maximum probability $d_{i,j}$
    $C_i \leftarrow j$
    $Q_i \leftarrow d_{i,j}$
  End for
  For $i=1$ to numclusters do
    $P_i \leftarrow Q'$ for all Q in Custer i
    $S_i \leftarrow$ stddev(Q) for all Q in cluster i
  End for

The following pseudo code describes outlier detection algorithm used in our research[15]

```
Input: set of evaluation instances x
Input: EM model M
Input: Set of cluster mean probabilities P
Input: Set of cluster standard deviations S
Output: set of z-scores Z
For all instance x_i do
    Apply M to x_i to obtain probability distribution d
    Select maximum probability d_i,j
    Z_i ← d_i,j – P_j
    Z_i ← Z_i-S_i
End for
```

VI. EXPERIMENTAL RESULTS

Clementine 12 was used to build test model, the model includes clustering nodes, anomaly detection nodes, output nodes. Input node was used to load data from excel file to the model as shown in figure 1.

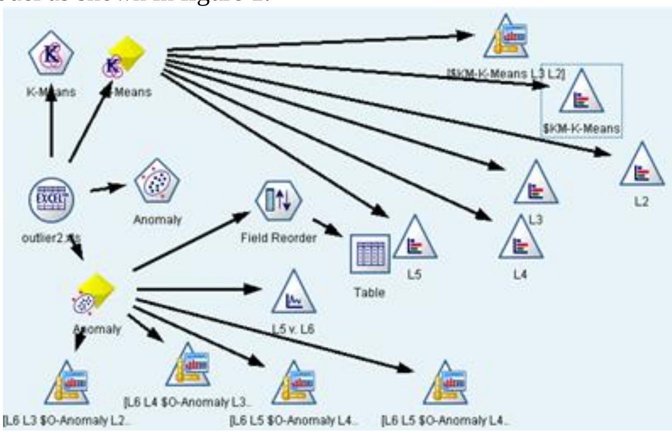

Fig. 1 Test model of our research

Records distribution along resulted clustered were noticed and analysed to recognize the character of clusters and trying to discover anomaly records manually as shown in figure 2.

The manual investigation shows two clusters with 2,1 record in each respectively and these records were nominated to be anomalies.

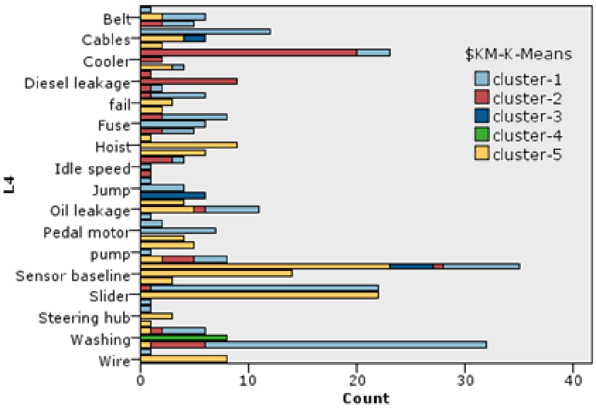

Fig. 2 Sample of records values distribution across clusters

The output of anomaly detection node was presented as cube with 3 dimensions as figure 3 shows, this cube allows us to visually identify anomaly records and recognize its character.

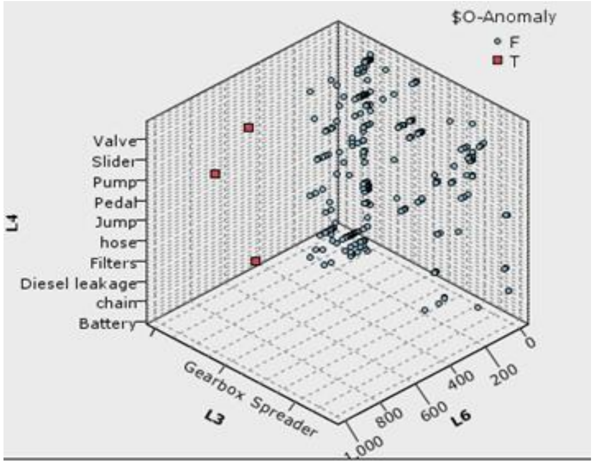

Fig. 3 Anomaly node output as cube

Figure 4 shows another view of anomaly records discovers.

Fig. 4 Anomaly records in two dimensions view

We noticed that 3 dimensions cube allows understanding anomaly values better as it shows its relation to mode factors in studied dataset.

Table 1 shows cluster centres modification after removing anomaly records from dataset..

TABLE I

CLUSTER CENTRES BEFORE AND AFTER ANOMALY RECORDS REMOVAL

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Clusters Centres with anomaly values	262	48	900	134	690
Clusters Centres after removing anomaly records	331.50	239.70	32.97	73.66	133.26

Figure 5 shows that are the centres of data clusters are more heterogenous after removing anomaly records which means the value of centres of clusters and their instances are more close to the overall average of fields values in dataset.

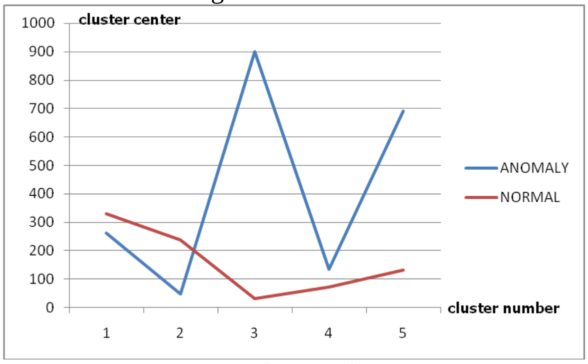

Fig. 5 Clusters centres before and after removing anomaly records detected by anomaly node

The impact of each fields was analyzed as shows in table 2, the results show that l6 which represent repair time has the most impact on make maintanance record anomaly.

TABLE II

DATASET FIELDS ANOMALY IMPACT ON DETECTED ANOMALY RECORDS

	L2	L3	L4	L5	L6
Anomaly record 1	.036	0.014	.088	0.037	.824
Anomaly record 2	.138	.017	.138	.117	.590
Anomaly record 3	0.02	0.024	.173	0.061	.723

Figure 6 shows the summary of table 2.

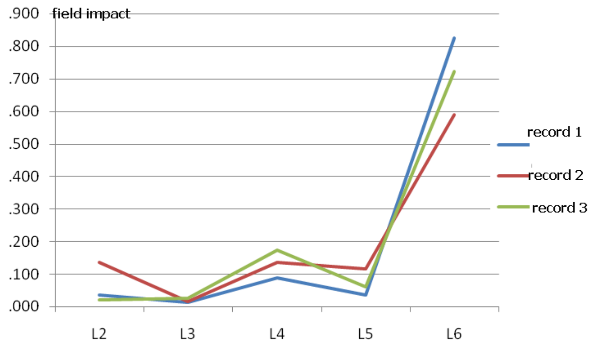

Fig. 6 Dataset fields impact on anomaly records

VII. CONCLUSIONS

In this paper we presented the impact of using anomaly detection algorithm in improving dataset quality based on the distributing of data records in clusters using K-means clustering algorithm, repair time has the major impact in specifying anomaly records, we suggest anomaly detection and removing or handling to be preparation step before applying data mining algorithms of other predicting techniques specially that which are sensitive to average standard deviation values.

REFERENCES

[1] Muchiri, P.,etal, "development of maintenance function performance measurement frameworkand indicators", International Journal of Production Economics ,May 2011, Pages 295-302

[2] Kumar.U; Galar.D; Parida.A; Stenström.C; Berges.L,"Maintenance performance metrics: a state-of-the-art review", Journal of Quality in Maintenance Engineering, Vol. 19 No. 3, pp. 233-277, 2013.

[3] Senechal.O; Leger.J.B,"Tele-Maintenance for Improving of Performances in TPM and RCM"Copyridht IFAC Cost Oriented Automation, Gatincau/ttawa. Canada, 2004

[4] Jaleel Valappil, David Messersmith ," Realizing the Full Value of Industrial Internet and Data Analytics in LNG Facilities",Bechtel Oil, Gas and Chemicals Kelly Knight Bechtel Nuclear, Security &Environmental Gastech 2017 Conference , Japan

[5] Jens Eickmeyer, Peng Li, Omid Givehchi, Florian Pethig and Oliver Niggemann," Data Driven Modeling for System-Level Condition Monitoring on Wind Power Plants", proceedings of the 26th International Workshop on Principles of Diagnosis

[6] Steckel, Thilo; Bernardi, Ansgar; Gu, Ying; Windmann, Stefan; Maier, Alexander; Niggemann, Oliver," Anomaly detection and performance evaluation of mobile agricultural machines by analysis of big data", Verein Deutscher Ingenieure -VDI-; VDI-Wissensforum GmbH: 73rd International Conference on Agricultural Engineering, LAND.TECHNIK AgEng 2015

[7] Ms. Gargi Joshi," Anomaly Extraction Using Association Rule Mining", Int. Journal of Engineering Research and Applications ,www.ijera.com,ISSN : 2248-9622, Vol. 4, Issue 1(Version 2), January 2014, pp.88-92

[8] Y.A.Siva Prasad,Dr.G.Rama, Andhra Pradesh," Statistical Anomaly Detection Technique for Real Time Datasets", International Journal of Computer Trends and Technology (IJCTT) –volume 6 numbe2–Dec2013

[9] Remya G, Anuraj Mohan," Distributed Computing Based Methods for Anomaly Analysis in Large Datasets" International Journal of Advanced Research in Computer and Communication Engineering,Vol. 4, Issue 6, June, 2015

[10] VARUN CHANDOLA,ARINDAM BANERJEE,VIPIN KUMAR," Anomaly Detection : A Survey", A modified version of this technical report will appear in ACM Computing Surveys, September 2009.

[11] Keith Noto,Carla Brodley,Donna Slonim," FRaC: A Feature-Modeling Approach for Semi-Supervised and Unsupervised Anomaly Detection",Article in Data Mining and Knowledge Discovery · July 2012 DOI: 10.1007/s10618-011-0234-x · Source: PubMed

[12] Mennatallah Amer,Markus Goldstein,"Nearest-Neighbor and Clustering based Anomaly Detection Algorithms for RapidMiner", Conference: Proceedings of the 3rd RapidMiner Community Meeting and Confererrnce (RCOMM 2012)

[13] A B.S.Charulatha, Paul Rodrigues, T.Chitralekha, ArunRajaraman,." Comparative study of different distance metrics that can be used in Fuzzy Clustering Algorithms",Member IEEE National Conference on Architecture, Software systems and Green computing - 2013(NCASG2013

[14] PhilipK. Chan,Matthew V. Mahoney,Muhammad H. Arshad," LEARNING RULES AND CLUSTERS FOR ANOMALY DETECTION IN NETW ORK TRAFFIC", Massive Computing book series (MACO, volume 5), Managing Cyber Threats pp 81-99

[15] Iwan Syarif , Adam Prugel-Bennett, Gary Wills," Unsupervised clustering approach for network anomaly detection", Fourth International Conference on Networked Digital Technologies (NDT 2012), United Arab Emirates. 24 - 26 Apr 2012. 11 pp.

[16] Michael J. Chapple, Nitesh Chawla ,Aaron Striegel," Authentication Anomaly Detection: A Case Study On A Virtual Private Network", MineNet '07 Proceedings of the 3rd annual ACM workshop on Mining network data Pages 17-22, San Diego, California, USA — June 12 - 12, 2007