REVOLUTIONIZING SYSTEM OPERATION AND MAINTENANCE IN THE AUTOMOBILE INDUSTRY THROUGH MACHINE LEARNING APPLICATIONS

Priyank Singh , Tanvi Hungund and Shobhit Kukreti

ABSTRACT

The necessity to digitalize the processes in the automobile industry becomes stronger by the day challenging the system operation and maintenance, in such a competitive field. Issues like passive fault identification, routine tasks, and dependency on Standard Operating Procedures (SOPs) describe the current status of a system in operating and maintaining processes used in the automotive industry. To address these challenges, this paper introduces an innovative approach: a machine learning system on operation and maintenance knowledge base for providing optimal solutions based on the automotive industry. More specifically, Scrap crawler is used to gather historical system data and after that, the decision tree algorithm is used to determine the specific insights. The acquired findings are further represented to facilitate comprehension and application of the results to strengthen the operation and maintenance management processes.

1. INTRODUCTION

In the last decade, dramatic progress has been made in the expansion of the Internet and its uses in mobile devices. However, the effects of increased competition, which has become more rigorous due to additional players, has made digitization necessary, with the automotive industry not being exempt. These changes and advancements have led to the growth of giant infrastructures for IT systems, highly complex IT structures and large data resources of operations. Therefore, as the business environment becomes increasingly competitive with day-to-day risks in an organization there is an ever increasing need for increased dependability and continuity of IT systems [1]. Within the realm of automotive industry operations and maintenance, several challenges have emerged:Within the realm of automotive industry operations and maintenance, several challenges have emerged:

1. Passive fault detection: While the key players in the automotive industry have integrated IT-based platforms and maintenance systems, most of the failure incidents are communicated by user-level end consumers, which prolong the issue identification and resolution time.

2. Repetitive tasks: They include simple monitoring processes of server and database health checks, and standard handling of frequently occurring issues including detached data disks or unhealthy service processes.

3. Reliance on experiential knowledge: Despite the fact that certain enterprises record event processing procedures and results in regards to their enterprises, insufficient information that is derived from such data is used to facilitate improvements in subsequent operations and maintenance. But the task is kept on coming and requires a manual intervention to manipulate it [3].There is a knowledge base system and operation and maintenance knowledge base, which utilise big data and machine learning related technologies and targeted at the industry of automotive. The tool that fits the task best is Scrap crawlers which is used to obtain historical data for the system, decision tree algorithms that analyze this data, and finally, presentation of the results.

DOI: 10.5121/ijcsit.2024.16308

1.1. General Framework of Integration of Knowledge of Operation and Maintenance by Machine Learning

1.1.1 Application Architecture

The application architecture of the machine learning-based system operation and maintenance knowledge base comprises three main modules: This is inclusive of both the collection and the analyzing of data, as well as the application of algorithms to arrive at a chosen result, and presenting of generated information, as illustrated by figure 1 below.

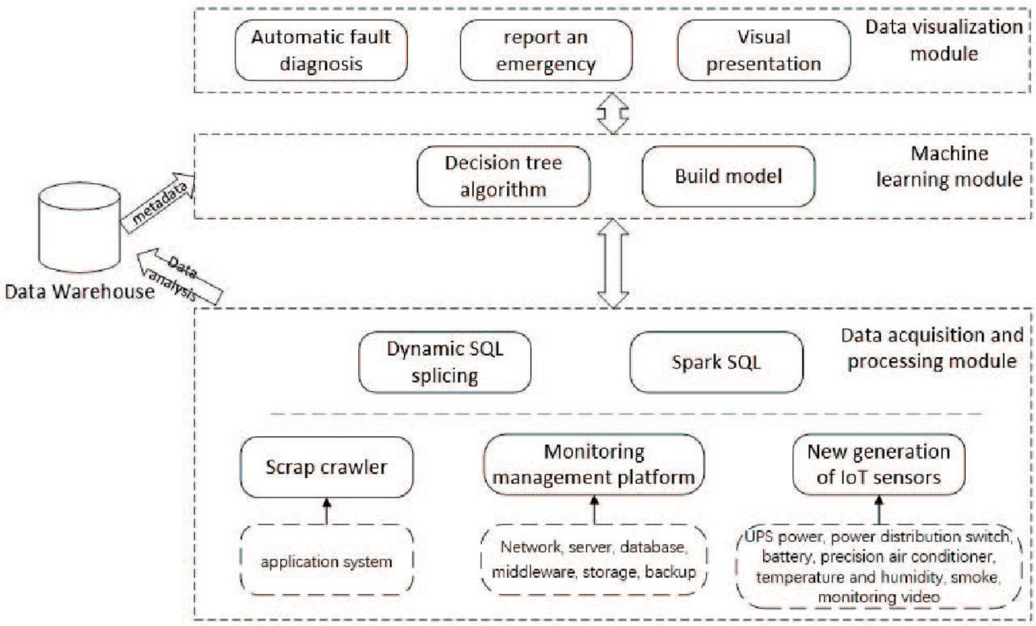

Figure 1. Application architecture

2. SYSTEM FUNCTION

The system operation and maintenance knowledge base comprise three distinct modules:The system operation and maintenance knowledge base comprise three distinct modules:

2.1. Data Acquisition and Processing Module:Data Acquisition and Processing Module:

This module begins by categorizing the content to be incorporated into the operation and maintenance knowledge base into three main types: Application system, system related hardware which include the network system, server system, database system, system middleware, storage system, backup system and UPS power, power distribution switch and battery, Precision air conditioning and temperature and Humidity control system, smoke removal system and monitoring video. The application system data used is collected from internet sources using crawlers and data into the system's hardware is from the monitoring management platforms. Secondly, information concerning the facility is collected using IoLT sensors that have been installed in the facility. : After dynamic SQL splicing and Spark SQL processing, the obtained data will be stored in the Data Warehouse.

2.2. Machine Learning Module:

This module comprises in obtaining meta-data and performing calculations, modelling and analysis with the help of the decision tree algorithm, and the end-product being the decision tree.

2.3. Data Visualization Module:

Based on the decision tree contained in the decision making module, this module identifies those faults that should be self-recovered and those requiring alarms to be raised as well as their visualization.

The following is the establishment process of the OM Knowledge base in the Automobile Industry:

2.1.1. Data Acquisition

The paper describes three approaches of data collection out of which a method involves Scrap crawler. In most of the automobile industry, this information comes directly from system operation and maintenance department in a form of 'work order' documented in the work order management system. For example, in the Figure 2, the notification – New work order is shown. When it comes to acquiring historical fault data of the application system, data collection through web crawling technique is beneficial in collecting webpage information. Scrap crawler, the open source web crawler framework programmed in Python was built with components including Scrap engine, Scheduler, Downloader, Spiders, Item pipeline, Downloader middle wares, and Spider middle wares. It begins with the initial URL given by the Scheduler to the Downloader for downloading and then to the Spider for extracting its infos. Information that needs to be kept is transferred to the Item Pipeline for further processing, as shown on the picture 2.

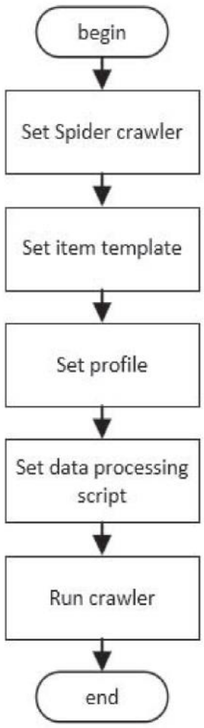

Figure 2. Scrap crawler development process

2.1.2. Monitoring Management Platform

In terms of operation and maintenance of automotive systems, most departments have set up monitors for warning and managing the network, servers, databases, middleware storage, backup systems, as well as other system equipment on large screens. The knowledge base for operation and maintenance enables these monitoring management platforms to feed directly into the data requirements with ease.

Next-Generation Internet of Things Sensors: New Characteristics and Techniques 3
A new generation of IoT sensors can be sent to gain operational data of facilities and hardware like UPS power, power distribution switches, batteries, Precision air conditioners temperature and humidity sensors, Smoke detectors, monitoring videos and others.

3. MACHINE LEARNING

From the operational scenario, machine learning entails training a model, with the aid of the data, for prediction circumstances. Of all the algorithms available in the machine learning process, the chosen one is the decision tree algorithm.

Decision tree algorithm describes the process of making decisions using decision trees, here is a brief explanation of the algorithm:
Therefore, the decision tree algorithm is widely used for classification, regression and rule extraction more than for mere decision making. It actually also forms a tree like structure where any node may be a leaf node which may mean a class or may partition the dataset [5]. The sample set linked to a node can be partitioned into a number of subsets, a number of nodes in proportion. In classification problems, the process of creating a decision tree from one set of known class marker to other set of class markers involves sequential steps in a divide-and-conquer fashion[6]. The main issue when trying to develop a decision tree is to find appropriate features for splitting the sample.

Now let me elaborate on the ID3 decision tree algorithm with enhanced detail. Originally, its major concept is based on the notions of information entropy for choosing the best test attribute. Information entropy is used to evaluate the quality and homogeneity of the sample set. For each non-leaf node, ID3 chooses the predicate whose information gain is highest in order to seek the greatest split given the conditions at hand. This results in the formation of a more compact decision tree which in the long run is more efficient as in this case.

3.1. Benefits of the Decision Tree Algorithm

The decision tree algorithm stands out as one of the most prevalent inductive reasoning methods today, boasting the following advantages:The decision tree algorithm stands out as one of the most prevalent inductive reasoning methods today, boasting the following advantages:
a. High Classification Accuracy
b. Omission of Complexity in the Archetypal Generated Models
c. Immunity to noiseÝ
d. Non-parametric Nature, Endowing it with Flexibility and Robustness [7].
In order to identify the numbers of clusters automatically, each types of cluster validity indices that exist in this world could be applied, although the most popular ones are shown in the below table:
In order to identify the numbers of clusters automatically, each types of cluster validity indices that exist in this world could be applied, although the most popular ones are shown in

3.2. Decision Tree Algorithm: The Staged Structure

The decision tree algorithm works by choosing an attribute that is used in internal node of a tree form of segmentation where segments are represented by branches [8]. The leaf nodes denote distributed decision tree generation, a process delineated into two steps:The leaf nodes denote distributed decision tree generation, a process delineated into two steps:

1. Tree Generation: Starting with some data the initial data are divided at the root node.
2. Pruning: In this step, inputs are preprocessed to ensure that there is removal of noise or other disturbing elements that may influence the results in some way.

Decision tree incorporates the following steps in operation: Other messages are unknown data messages and the application of the decision tree entails the segregation of unknown data messages and subsequent navigation of the layers of the decision tree by the unknown data based on the elements of segmentation defined within the decision tree [9].

3.3. Application and Realization in the Automobile Industry

This paper aims at identifying the practical example where the network operation and maintenance occurred by taking the case of an automobile manufacturer. The manufacturer has adopted a work order management system by having vast fault information data for over a decade. This paper, employing the ID3 decision tree algorithm, identifies if certain faults can be resolved in an auto mode or if they necessitate a change in the alarm mode and manual supervision.

4. DATA PREPROCESSING

The raw data used for the analysis is a sample of 68 records,which covers network-related faults Only CW Warehouse is used for the analysis with the help of the decision tree algorithm. The criteria to be analyzed include ‘Work Order Type,’ ‘Priority,’ and ‘Business System. ’ In attribute preprocessing, the measure for the ‘Work Order Type’ component is ‘Event’ which translates to ‘Yes,’ while the other value is ‘Non-Event’ which translates to ‘No’. As for the ‘Priority’ attribute, ‘Emergency’ is coded as ‘Yes,’ while ‘Non-Emergency,’ is coded ‘No. Also, “Yes” implies something affects the attribute ‘Impact on Business System’ while “No” means it has no implication on same. The data preprocessing workflow is described in the following Figure 3, and in the following table, we have provided a snapshot of the resulting cleaned dataset of Table 1

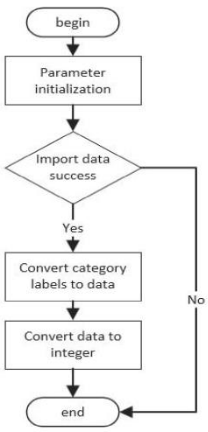

Figure 3. Data pre-processing

Table 1. Processed data set

Number	Is it an "event"	Emergency	Impact on business system	restore manually
1	Yes	Yes	Yes	Yes
2	No	Yes	No	Yes
3	Yes	Yes	No	No
…	…	…	…	…
68	No	No	No	Yes

4.1. Establishing Decision Tree Model

A decision tree model based on information entropy is constructed for the preprocessed dataset, as depicted in Figure 4. Executing the code yields a text file named tree.txt, a snippet of which is presented in Table 2.

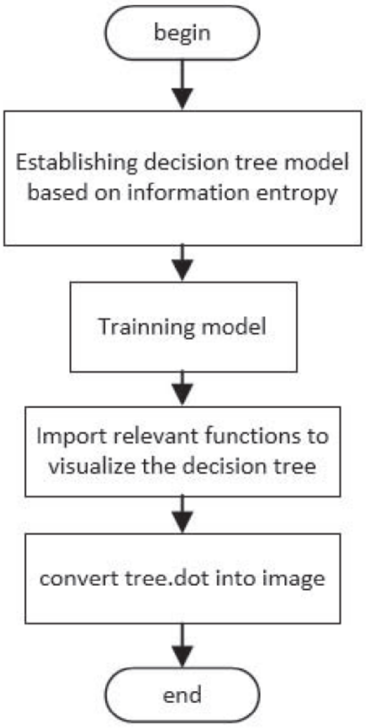

Figure 4. Flow chart of establishing decision tree

Table 2. Some information of tree.dot

File name	Content

tree.dot	digraph Tree {
	node [shape=box, fontname="helvetica"] ;
	edge [fontname="helvetica"] ;
	0 [label="Is it an "event" <= 0.0\nentropy = 0.96\nsamples = 68\nvalue = [26, 42]"] ;
	……
	}

5. VISUAL PRESENTATION

To convert a tree. On the riot dot into a visual format, Graphviz, a drawing tool which is a cross-platform command-line tool, needs to be installed. Later on, the command “dot –Tpng tree. dot –o tree. png” is typed into the system command-line terminal to generate the visualization of the obtained decision tree in . png format as shown in figure 5.

Several shortcomings are identified regarding the decision tree analysis; according to the ID3 algorithm, some attributes with the highest information gain values may be chosen, while others are more suitable for the problem. Furthermore, ID3 cannot take continuous variables, which means that it has to partition continuous variables into different bins. The C4. 5 algorithm, on the other hand, targets the discrete and continuous attributes, but results into the multi-branch trees [10:].

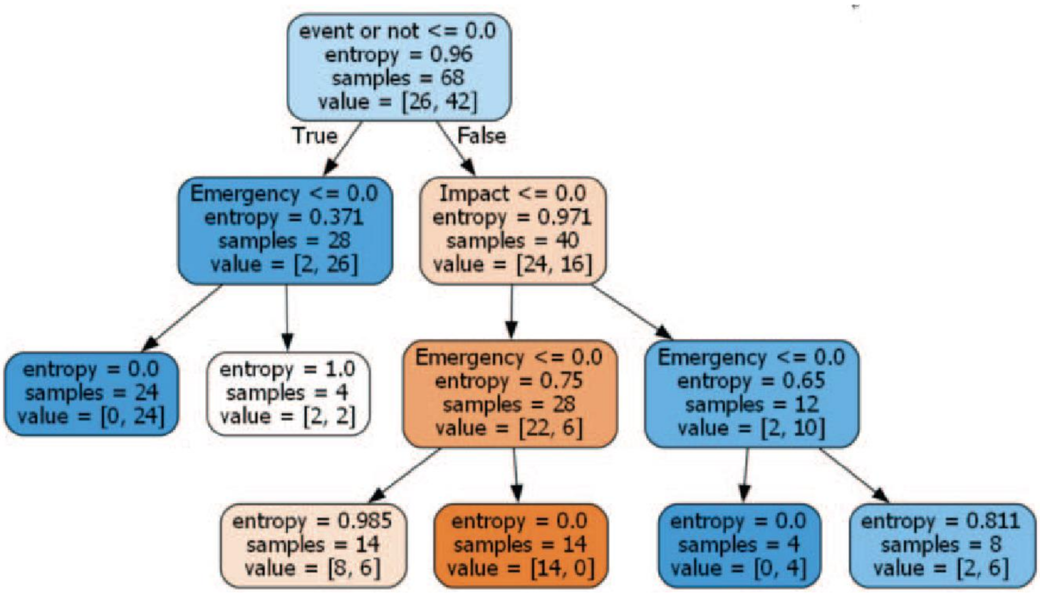

Figure 5. Visualization results

However, a significant part of the knowledge that the automobile industry needs to operate and maintain the automobiles exposes a predominantly discrete type of feature. ID3, a traditional Decision tree learning technique, uses Information gain for the selection of attributes throughout Decision tree nodes, which reduces Information entropy to improve the classification capability.

6. CONCLUSION

The setting up of a system operation and maintenance database in ph automobile industry effectively manages and improves the quality and management system of software and hardware maintenance. Scrap crawlers for historical data with the combination of the practical decision tree algorithm on extraction and pertinent results pave the way for the operation and maintenance of the knowledge base to be highly feasible, practical, and scalable to support the automotive industries’ maintenance operations.

The automotive industry gives different opportunities and real-life situations where AI can be used for operation and maintenance. However, AI’s ability ranges beyond prognosis and flaw identification since it can be used in stock management, scheduling of predictive maintenance, and redesigning vehicle models. For instance, the sensors used in computer vision such as optical character recognition can use the image feed of manufacturing lines and the defects detected are immediately flagged for correction.

Implementation Details and Architecture

Having a proper architecture is crucial to the incorporation of the AI solutions. This ordinarily includes data acquisition from one or several sources such as sensors or work orders, data preparation both in quality and structure, model development, and assessment, as well as the model deployment for the purpose of decision making. By using cloud-based at times, it can give the needed capacity and computation for large data sets and complex AI.

Addressing Dataset Limitations

Hearing, the sample dataset employed in the present study is comparatively small, although representative. In optimization, the real-life data volumes are much higher compared to the model's parameters required for training the AI systems. Partnerships with automotive makers form an extremely valuable source of a broad range of data to support the development of higher basic and fine-grained AI solutions.

Exploring Additional AI Algorithms

Although decision tree algorithms are significant, other AI algorithms may have additional advantages. For instance, while neural networks can be effective in the pattern recognition in higher dimensional data, reinforcement learning can be perfect for decision-making in a stochastic environment. The utilization of these multiple AI methodologies can improve the understanding of the company's operations and let to the development of more elaborate strategies for automotive management and maintenance.

Embedded AI Technologies

The incorporation of AI into vehicles and manufacturing equipment becomes direct. AI systems are able to remain embedded into the vehicles and assess performance, predict component failure or even allow autonomous maintenance. It can contribute much to minimization of the time when vehicles are off the road, guarantee safety and increase the efficiency of automobile businesses.

REFERENCES

[1] Chao Lv, Jiaxin Sun, Shuang Liu. Weight analysis of factors affecting fishing capacity of marine fishing vessels using machine learning algorithm [J] Journal of Agricultural Engineering, 2021, 37 (13): 7.
[2] Genxia Yue, Jinhua Liu, Feng Liu. Medical big data filling and classification simulation based on decision tree algorithm [J] Computer simulation, 2021.
[3] Qianmei Yuan. Research and analysis of new IT operation and maintenance management and decision-making based on big data-driven [J] China Informatization, 2019 (11): 2.
[4] Lena He, Zekuo Guo. Research on intelligent operation and maintenance ecosystem of urban rail transit based on artificial intelligence and big data [J] Urban Rail Transit Research, 2022 (009): 025.
[5] Zhongyuan Chen. Design of IT operation and maintenance data management system based on big data [J] Integrated Circuit Applications, 2022 (001): 039.
[6] Zhaoxing Li, Qiaoyun Cui. Research on hierarchical diagnosis and treatment model based on improved decision tree algorithm [J] Electronic Design Engineering, 2021, 029 (012): 39-42,47.

[7] Zhe Chen, Hengmin Xu, Zhexuan Li. Random survival forest: survival analysis model based on machine learning algorithm [J] Chinese Journal of Preventive Medicine, 2021, 55 (1): 6.
[8] E. Hill, H. Hoffman, P. Shankar. Train the machine learning algorithm and predict the values of weather data variables, especially at the field or sub-field level:, CN112166352A [P]. 2021.
[9] Xin Zhang, Yingying Tian, Zelong Han. Research on theoretical model of lost circulation prediction and diagnosis based on machine learning algorithm [J] Drilling Engineering, 2022 (049-002).
[10] Anlin Xia, Dongsheng Du, Yuanjie Sheng. Bank target customer prediction algorithm based on decision tree [J] Computer Knowledge and Technology: Academic Edition, 2022 (024): 018.

AUTHORS

Priyank Singh
Tanvi Hungund
Shobhit Kukreti