PREDICTION OF THE CRASH VEHICLE USING MACHINE LEARNING METHODS

The article refers to the prediction of failure of the vehicle based on an analysis of data from sensors installed on the vehicle.

INTRODUCTION

In recent years rapidly developing science based on new logistics technology. Information systems occupy a central position these technologies. Enterprise is an open system, material and information flows associated with suppliers, customers, freight forwarders and transport organizations. In such a system is very important precision and accuracy of each operation. Unfortunately, modern diagnostic systems of vehicles able to evaluate only the current condition of the car and did not give any further example of its operation. With some frequency, each vehicle passes inspection, but not enough to properly assess the wear rate of the vehicle depending on the nature of the driving license or road conditions, which most of the time drives this car. In each transport system is important to most accurately predict the damage and prevent it in time or replace your current vehicle to another to be sure that the traffic will be on time.

Such predictions may make applying machine learning methods. Machine Learning (ML) – is the study of computer algorithms can learn to improve their performance in carrying out a task based on their own experience [1]. This industry is closely related to pattern recognition and statistical analysis. As engineering area, ML certainly become more mathematical and more successful in applications over the past 20 years. Such approaches like clustering data classifiers and neural networks nonlinear regression found a wide application in design practice, business and science.

1. REVIEW OF METHODS OF DIAGNOSIS AND TRANSPORT

1.1. Diagnostic service and quality concepts

Diagnostic services is support of resources and machines, whether big or small, based on the forecast when they fail or require maintenance. These facts may include the following items:

1. The state of the device in real time.
2. Historical data device.
3. Data for these devices.
5. Maintenance schedule.
7. Journals services.

All these data are meaningless without analysis. There are hidden information that is hidden inside of these facts and figures. Deciphering these models is the main purpose of the diagnostic service and separates it from the more traditional approaches to repair and replace equipment.

1.2. Traditional approaches

Diagnostic services is very different from traditional approaches to determining the date of repair or replacement of equipment. For years, companies have kept their production lines, which run on a combination of these methods of maintenance:

1. Reactive or damage: repair or replacement of equipment after failing attempt.
2. Prevention: repair or replacement of equipment according to the proposed schedule manufacturer based on the number of hours worked or on the basis of the available observations.
3. Control based on state: repair or replacement of equipment based on monitoring carried out regularly to assess the current state.

The problem of outdated approaches is its high cost. Waiting until the component fails means the loss of production time and income. Human review is expensive and can cause replacement of parts without the need based only on assumptions inspector. Follow the recommended maintenance schedule from the manufacturer saves the cost of inspection, but often leads to replacement of parts that are still functioning and can continue.

1.3. Value for business

Predictive maintenance and quality solution solves two business precedents:

1. Prediction breakdown of assets and extend the life of: performing preventative maintenance, failure of components can be determined on the basis of performance and wear. In addition, there may be certain conditions that will lead to a high failure rate. It provides several business benefits:
   - calculate and extend the life of components;
   - increase the return on assets;
   - optimize maintenance schedules.
2. Quality details: diagnostic service and quality solutions can also detect abnormalities within the manufacturing process by comparing the details of the nominee, and conduct a detailed analysis of the root causes breakage. It provides the following benefits for business:
   - improve the quality and reduce the likelihood of repeated failure;
   - reduce time to solve problem;
   - increase customer satisfaction.

1.4. Use of preventive maintenance and making quality decisions

In predictive maintenance, it searches for patterns in how to use the equipment and the environment in which it works, and then correlates this information with sensors of any known failures in equipment. These patterns and relationships used to create
forecasting models, which are then used to assess the new data from sensors or data stored in the analytical part. This process provides forecasts which indicate the relative status of the integrity of the equipment, the probability of failure, and the best time to perform maintenance equipment.

Another aspect of preventive maintenance includes key performance indicators (KPI), which are collected for each of the controlled equipment, mainly used for reporting. KPIs help identify assets that do not meet the normal structure of behavior. You can also define rules to make recommendations as part of the equipment identified as having a high probability of failure. These recommendations can be given in other systems in which users are automatically warned about it.

If increased production defects, their causes can often be identified by analyzing data on past transactions, environmental conditions and historical defects. When applying this information in predictive models can predict the likely number of defects in the future. Estimates are then used for analysis and reporting.

1.5. Overview analogues

Predictive service vehicles is a highly effective process that prevents breakage in transit and allows predicting the failure of the vehicle. Huge companies like IBM have long been interested in this area and in 2013 IBM released the decision Predictive Maintenance and Quality.

Solutions IBM Predictive Maintenance and Quality helps track, analyze and report on data collected from the devices and recommend the maintenance of these devices. With this integrated solution you can perform the following tasks [13]:
1. Provide refusing controlled vehicle, so you can fix it and avoid costly downtime.
2. Optimize maintenance schedules of vehicles through the use of intelligent forecast ideas.
3. Determine what equipment provides better damage details.
4. Perform statistical process control analysis on various parameters asset.
5. Interactively analyze frequency distribution histogram based on various parameters asset.
6. Merge several predictions forecasting models, each providing different information from service equipment in structured, semi-structured and unstructured formats.
7. Find answers and take corrective action quickly, performing interactive analysis of root causes using advanced imaging technology forecasting and analytical methods without the need to switch between multiple systems. This process reduces the cost of identifying problems, finding its causes and adopting corrective measures.
8. To provide early warning for failures that occur rapidly identified during the inspection batch production.
9. To provide early warning signals to parts tariffs for the use or replacement is growing. To analyze the reasons for the observations in detail, for example, changes in operating conditions, abnormalities in the manufacturing process.
10. Calculate the overall condition of all equipment, using advanced models for predicting patterns in the sensor data that can lead to the release of equipment failure.

Overview of the solution shown in Figure 1.

Controlled assets (called tools) generate data such as object identification, temperature or other state code and a temporary tag. This data can be collected and used in analytical models that predict the asset is likely to fail or require maintenance. Assets that can be tested for the purpose of predicting maintenance include equipment used for manufacturing, mining, drilling, agriculture and security, covering everything from cars and trucks to engines, cranes and oil platforms.

2. DEVELOPMENT OF WEB SERVICE

The article offers a new free web service that will give an opportunity to analyze the online status of the vehicle and make predictions, such as the date of the next inspection, the time of replacing certain parts or probability of failure of the vehicle by the next flight. All modern vehicles are equipped with multiple sensors, the data from which using 4G technology can be transferred to a remote server. Thus we get a large number of records that can be unstructured due to certain glitches that some data may be missing or damaged.

Information received from the sensors can be classified as Big

Fig. 1. Overview of the solution IBM Predictive Maintenance and Quality
Data, because it can not be processed by standard software and hardware. Usually, when people talk about the term Big Data, then use the three most popular definition of «V», which means Volume – the amount of data, Velocity – the need to process information at high speed and Variety – the variety and often lack structured data [2]. For example, the operation of the check card balance when withdrawing cash is calculated in milliseconds. Such requirements dictated by the market. The third side of the issue – it’s diversity and unstructured information. Increasingly, has to operate media content, blog posts, poorly structured documents, etc.

So when we talk about big data, understand that this is due to three aspects: the large volume of information, its diversity or the need to process data very quickly [6]. On the other hand, this term is often understood quite specific set of approaches and technologies designed to solve these problems.

In this case, to solve the problem of collecting and processing information coming from the sensors, it is proposed to create a web service based on the framework Anaconda. Anaconda is a platform for large-scale data processing, predictive analysis and scientific computing that aims to simplify the management and deployment packages using language Python. To simplify the use of Pandas – library written for Python programming language processing and data analysis. In particular, it provides data structures and operations to work with numerical tables and time series. Pandas is free software, released under license BSD. The name comes from the term "panel data" – the term for econometric multi-structured datasets.

Basic principles of such data:
1. Horizontal scalability. Because data can be any number – any system that involves the processing of large data should be extensible. That is, if 2-fold increased volume of data – should be 2 times increase hardware.
2. Fail. The principle of horizontal scalability means that the machines in the cluster can be many. For example, Hadoop- cluster Yahoo has more than 42,000 vehicles. This means that some of these machines will be guaranteed to go down. Methods of working with large data should consider the possibility of failure and worry them without any significant consequences.
3. Locality data. In large distributed systems data spread across a large number of machines. If the data is physically located on the same server and processed in another – data costs may exceed the cost of processing itself. Therefore, one of the most important design principles BigData-making is the principle of locality data – process data as possible on the same machine on which they are stored.

All modern means of working with big data in some way are these three principles [3]. In order to observe them – need to invent some techniques, methods and paradigms of development data processing facilities. In this case, the proposed use MapReduce. MapReduce – a model of distributed computing proposed by Google to process large amounts of data on computer clusters. Steps such data is shown in Figure 2.

MapReduce provides that the data is organized in the form of some records. Data processing takes place in 3 stages:
1. Stage Map. At this stage the data are preliminary processing using map() function, which identifies the user. The work of this phase is preliminary processing and filtering data. The work is very similar to the map operation in functional programming languages – custom function is applied to each input record. The function map() gives a lot of key-value pairs. The set – that can issue only one entry can not give anything, and can issue multiple key-value pairs. What will be located in the key and the meaning – it’s user, but key – a very important thing, as with other key data in the future to get a copy function reduce.
2. Stage Shuffle. Taking place behind the scenes. At this stage, the output function map ‘versed in baskets’ – each basket is corresponding to the output key on stage map. Further, these baskets will serve as input to reduce.
3. Reduce. Each “basket” with values formed in step shuffle, gets the input function reduce(). Reduce function set by the user and calculates a final result for a single “basket”. The set of all values returned by the function reduce(), is the final result of MapReduce-task.

Thus, we get the data of which is shown in the table 1.

<table>
<thead>
<tr>
<th>Speed</th>
<th>The position of the accelerator pedal</th>
<th>The sensor of used fuel</th>
<th>Fuel level sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>118.31</td>
<td>22.0</td>
<td>109603.0</td>
</tr>
<tr>
<td>12</td>
<td>63.23</td>
<td>24.0</td>
<td>109603.0</td>
</tr>
<tr>
<td>14</td>
<td>72.21</td>
<td>10.0</td>
<td>109603.0</td>
</tr>
<tr>
<td>15</td>
<td>96.24</td>
<td>18.4</td>
<td>109603.0</td>
</tr>
<tr>
<td>16</td>
<td>107.06</td>
<td>31.2</td>
<td>109603.0</td>
</tr>
<tr>
<td>19</td>
<td>94.06</td>
<td>23.2</td>
<td>109603.0</td>
</tr>
</tbody>
</table>

The next step is the preparation of data for the classifier. Machine learning methods used for classification assigned for the collection of previously classified documents, ie those for which it is known the value of objective function \( \Phi \).

To solve this problem it is necessary to form two sets of data – reference, ie a set of data from sensors machines that traveled more than or as much as had to travel from the last inspection, another set will contain data from sensors of the vehicle requiring unscheduled repair.
In order to build on the classifier could assess its effectiveness, \( \Omega \) is divided into two parts, not necessarily equal size:

- Training (training-and-validation) collection. Classification is based on the basis of the characteristics of these records.
- Test collection. It verified the classification quality. Records of the test collection does not have to participate in the process of building a classifier, or the results of the test will no longer correspond to reality.

\[
\Omega = \{d_{1}, \ldots, d_{[\Omega]}\}, \Omega \subset D
\] (1)

In the above partition, training collection should allocate an additional verification collection of documents, on which will be optimized classifier parameters.

Directly to classify was elected Bayesian method. Bayesian method based on an analysis of compatible document distribution features and categories. For document \( d \) with vector \( d = (d_{1}, \ldots, d_{n}) \) matched posteriori most probable category \( c \) with feature vector \( c = (x_{1}, \ldots, x_{n}) \) by formula 2.

\[
c^* = \underset{c \in \mathcal{C}}{\arg \max} P(c|x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n})
\] (2)

On the problem of classification Bayesian method is applied for each category and the decision belongs to the category of record or not.

Posterior probability that the document belongs to the category is calculated using Bayes connecting the priori probability to the posteriori.

\[
P(c|x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n}) = \frac{P(x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n})P(c)}{P(x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n})P(c)}
\] (3)

Since the denominator is independent of the category, formula (4) can be reduced to the form (5).

\[
c^* = \underset{c \in \mathcal{C}}{\arg \max} P(c|x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n})
\] (4)

\[
c^* = \underset{c \in \mathcal{C}}{\arg \max} P(c|x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n})P(c)
\] (5)

Conditional probabilities \( P(x_{1} = d_{1}, x_{2} = d_{2}, \ldots, x_{n} = d_{n}) \) can be calculated assuming full independence variables \( x_{1}, x_{2}, \ldots, x_{n} \). In this case, the formula to determine the most probable category will look like (6).

\[
c^* = \underset{c \in \mathcal{C}}{\arg \max} P(c) \cdot \prod_{i=1}^{n} P(x_{i} = d_{i}|c)
\] (6)

Let \( \text{Rub}(d) \) – a function that by record \( d \), returns the category in which the post. \( \text{Ex} \) – set of records from the training set.

For set of training records the probabilities \( P(x_{i}|d|c) \) are calculated by the formula (7).

\[
P(x_{i} = d_{i}|c) = \frac{[D \in \text{Ex} | c \in \text{Rub}(D) & \& D_{i} = d_{i}] + 1}{[D \in \text{Ex} | c \in \text{Rub}(D)]]}
\] (7)

Add a unit in the numerator to records that contain features that are not found anywhere else, have nonzero probability for all categories.

Bayesian method has high speed and simple mathematical model [8]. This method is often used as a basic method for comparing different methods of machine learning. This method is suitable for solving this problem, because the probability of belonging to a category record can be interpreted as the probability of premature failure of the vehicle.

Web application must be built using the Python language and a web framework Django. Django – free framework for web applications. Uses pattern design MVC (model-view-controller).

Django based application for one or more applications are encouraged to develop independent and connected. The basic principle framework – DRY (Don’t repeat yourself).

To implement RESTful services is used plug-rest-framework. Using frameworks Django and rest-framework justified their properties. Django, unlike many existing web frameworks, provides a convenient mechanism for explicit configuration URL handlers using regular expressions, this mechanism does not depend on the structure of the application controllers. Thus it is possible to implement RESTful architecture application. URL structure controls the main web application in Django, when accessing the URL charge web service management services transmitted controllers implemented using rest-framework.

REST (Representational State Transfer) – style architecture building distributed applications. REST is used to build applications in which customers can send inquiries services, ie implementation approach "client-server" [4]. REST data to be transmitted over HTTP in a small amount of data in one of the formats: HTML, XML, JSON.

REST applications must meet the following basic principles:

- user agents interact with resources that may all that can be called and present;
- each resource is uniquely identified by its URL (Uniform Resource Identifier – a universal code resources);
- interaction with resources by using a single interface standard HTTP commands (GET, POST, PUT i DELETE);
- views formats – standard MIME-types;
- Resources describe themselves, all the information needed to process the request resource contained in the request;
- resource can contain links.

The advantages of using REST principles:

- A single interface – the possibility to re-use code.
- Extension – REST enables to include all classes every resource needed to handle a particular request. RESTful services simply scaled subject to this restriction.
- Interoperability – to write a client application must have http library available on all modern operating systems.

Level storage base represented by MongoDB. MongoDB – document-oriented database management system (DBMS) open source does not require schema description tables. MongoDB occupies a place between fast and scalable systems that operate on data in the form of key / value, and relational DBMS, functional and convenient in forming queries [5].

MongoDB supports the storage of documents in JSON-like format, has a very flexible language to create queries that can create indices for different attributes stored effectively allows storage of binary large objects, supports logging operations to change and add data to the database, can work under paradigm Map / Reduce, replicate and build fault-tolerant configurations. Expansion of the cluster or converting one server in the cluster is performed without stopping the database of simply adding new machines.

Main features of MongoDB:

- document-oriented storage (simple and powerful JSON- like schema data);
- it is flexible enough to form queries;
- dynamic queries;
- full support index;
- query profiling;
- quick update "on the spot";
- efficient storage of large amounts of audio data, such as photos and videos;
- logging operations that modify data in the database;
- support fault tolerance and scalability, asynchronous replication, and replica set sharding;
- can work in accordance with the MapReduce paradigm.
In MongoDB is built-in to ensure sharding (distribution data set for servers based on certain key), combining that data replication can build horizontally scalable cluster storage, where no single point of failure (failure of any node does not affect the work DB) supports automatic failover and load transfer from the node that has failed.

3. SAFETY WEB SERVICE

The selection and implementation of design solutions that involve the use of security technologies, often negatively affect productivity solutions. This does not mean that all security technologies used in decisions leading to reduced productivity. Rather, you should realize that web-services solutions requiring authentication of business participants signatures to encrypt the contents of messages and XML-data can vary significantly depending on the technology and methods used in the form of public protection services business functions and data.

Triad security:
1. Authentication is designed to ensure that parties of a business transaction are the ones who they themselves declare that prove the authenticity of the parties. Such proof may be in demand in various ways. Simple option – providing the user ID and password. A more complex version – Certificates X.509, issued by a trusted certificate authority (eg, Verisign). The certificate contains identity credentials and associated with a pair of secret and public key. Proof of authenticity, which is served by the party includes the certificate itself and a separate piece of information that contains a digital signature using a private key certificate. After reviewing the information signed by the public key associated with the certificate other hand, the host can authenticate the sender as the owner of the certificate so convinced of his authenticity. Authentication parties called each other a mutual authentication; such authentication is often performed between the consumer and supplier web-service.

2. To ensure the integrity of business information (which the sides exchanged in the transaction) and ensure that content is changed or damaged during its transmission over the Internet, data signature signed using security keys. This is the second triad security requirement. The common practice is to use X.509 certificate private key for signing the sender digitally signed SOAP-body search web-service. Similarly, you can sign blocks SOAP-request header to ensure the integrity of the transmitted information in the transaction that goes beyond the current business context (eg, message IDs, access tokens). In addition, to ensure data integrity signature can sign the response web-services.

3. The third requirement triad of security is confidentiality. To exchange information and answers queries web-services unreadable for unauthorized uses encryption technology. The goal – to ensure that any attempt to apply to transmit data in memory or after saving the need of appropriate algorithms and security keys to decrypt data without which it is impossible to gain access to relevant information.

Currently, all these security measures can be implemented using different mechanisms. Depending on the specific needs and business environments you can select or transport sensitive or specific for exchanging SOAP-messaging mechanisms [10].

According to the opinion of the analysts of OASIS Web Services Interoperability, the main threats that target Web services – is unauthorized changes messages, loss of privacy and authentication of senders, DoS-attacks is a threat to information security web services, virtually identical threats aimed at other digital resources. Information security Web services involves the use of conventional technologies of information security – encryption, digital signatures, password protection, etc. It can be said that information security standards Web services meet the specific IT security architecture of Web services, but the mechanisms for implementing this architecture quite traditional.

One of the most popular standards is WS-Security, which describes the processes of authentication and authorization in a medium of exchange SOAP-messages. WS-Security provides user authentication using pairs username / password, X.509 certificates or Kerberos protocol. These technologies realize a digital signature that lets you verify the integrity of the message. Developers WS-Security is also taken care of encrypting SOAP-messages, defining mechanisms for use in environments SOAP XML Encryption standard.

General WS-Security model

The WS-Security specification is in its final approval process within the OASIS standards body and provides mechanisms to address all three of the requirements outlined as the security triad between application end points. With WS-Security, you can selectively implement each of the requirements of the security triad such that one or all of them are addressed in your solution.

An application that requires the services of another application is considered the Consuming Application. I refer to the application providing the services as the Service Provider. The Figure 3 illustrates this relationship and is the basis for much of the discussion that follows.

Protecting Web services must be integrated from threats to information necessary to protect not only the SOAP-messages, but also all the other components of the architecture of Web services, their interfaces, search engines services (registries), etc. For example, the latest version of the standard UDDI provides perspective are important in terms of information security functions as data integrity and encryption of the contents of the register Web services.

Standardization of information security Web servers constantly evolving and improving, and therefore to ensure a high level of information security we recommend using the latest version of the standards and specifications.

CONCLUSION

Every enterprise asset should be stored and used under certain conditions for optimum performance and increase the duration of its use. Diagnostic service takes into account not only the manufacturer’s recommended maintenance procedures, but also evaluates how really used assets, including its performance, or repair was made final earlier or later than expected, and so on. The

**Fig. 3. System environment**

[Diagram showing web service request and response with various authentication and security mechanisms highlighted.]
aim is to avoid costly downtime by ensuring that the asset will always be available and in optimal condition.

Using machine learning techniques ensure continuous improvement of the algorithm and the configuration of service for a specific system in which it is used, taking into account all its features. On the other hand, the web service will be protected from criminals using the latest technology. WS-Security provides security by using existing standards and specifications. This eliminates the need to define a complete security solution as part of WS-Security. To transfer data related to security, WS-Security defines element – SOAP Header. When using XML Signature, this header can contain the information specified which XML Signature, which indicates that a message was signed which key was used to sign the resulting value. Similarly, if an element in the message is encrypted, the encryption of information, similar to that which transports XML Encryption, may be contained in the title of WS-Security. WS-Security does not specify the format of the signature or encryption. Instead it establishes, as will be embedded in the SOAP message security information prepared by other specifications. WS-Security – a specification primarily for safe containers metadata based on XML.

In addition, WS-Security defines a mechanism devolution simple user by using the UsernameToken [11]. Also the Binary Security Token to send binary markers used for encryption or signing message. This message header can store information about a daring, how was signed and encrypted messages. WS-Security presents a solution to ensure global security Web service by storing all information security of a SOAP message.

REFERENCES


10. Hua Yue (College of Information Science and Technology Nanjing University of Aeronautics and Astronautics, Nanjing, 210016, China) Xu Tao (College of Computer Science and Technology Civil Aviation University of China, Tianjin, 300300, China); Web Services Security Problem in Service-oriented Architecture[A][C];2010.


13. IBM Predictive Maintenance and Quality 2.0 Technical Overview.