The distributed systems of engineering supervision of permanent structures

Nataliya Mokrova¹, Alexander Mokrov² and Alexandra Safonova²

¹Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia
²Software Engineer, CJSC NORSI-TRANS, Bolshaya Novodmitrovskaya Ulitsa, 12/15, Moscow, 127015, Russia

E-mail: ¹natali_vm@mail.ru

Abstract. This article explores the benefits of engineering analysis distributed systems for the purposes of analysis of depreciation and restoration cost of permanent buildings and structures. We propose the methodology for the study of complex automated systems of industrial and residential facilities life cycle control which is based on an integrated approach. We also propose the methodology for using the benefits of decomposition and proposed approach to analyze the events occurring during the production process using machine learning. The brief review of methods of calculation of depreciation of construction objects is presented. The choice of task examination of technical condition of a building or structure is indicated. Investigated the importance of the comprehensive analysis of geometrical parameters, instrumental studies, the determination of the actual characteristics of materials of the basic bearing structures and their elements, the measurement of the operational environment and operating loads, etc. Considered the models implementing machine learning for the classification of life cycle events. Descriptors of the internal production network data were used for training and testing of applied models. k-Nearest Neighbors and Random forest methods were used to illustrate and analyze proposed solution. The proposed methods allow to analyze causes of defects and damages in structures. The result of this work showed the possibility of successful application of the statistical approach for data analysis in construction, demonstrate the effectiveness of the implementation of such systems. The result makes possible to detect problems at an early stage and simplify the task of life cycle management of buildings and structures.

1. Introduction

Distributed systems of automatic information processing are characterized by sharing resources, openness and transparency, high fault tolerance, the possibility of parallel data processing, scalability, i.e. the ability to add new properties and methods. Among the shortcomings of distributed systems, there are increased complexity, limited work safety and manageability, as well as the unpredictability of the system's response to changes in external factors over time. Distributed control system (DCS, Distributed Control System) can be defined as a system consisting of a set of devices located in remote areas, each of which is independent of the others, but interacts with them to perform common tasks [1]. Each stage of the life cycle is characterized by its degree of automation.
The international open standard IEC 61499 [2] of distributed control and automation systems involves the use of compatible and portable software.

2. Life cycle of Buildings and Facilities

Building is a kind of fixed assets on a natural and material basis, including architectural and construction objects, the purpose of which is to create conditions (protection from atmospheric phenomena, etc.) for labor, housing, social and cultural services for the population and storage of material values. Buildings have as the main structural elements of the wall and the roof (the base of the power plant in the open air refers to buildings).

Facility is a single result of construction activities, designed to implement certain consumer functions. We can list the following types:

- civil construction: residential, sports, recreational, etc.;
- transport facilities (transfer devices): roads, power lines, pipelines;
- hydrotechnical and meliorative structures;
- capacitive facilities: tanks, bunkers, silos.

In most strategic automated management of building cases are used. By analyzing situations that are described in cases, researchers become skilled at effectively using the tools, techniques and concepts that combine to form the engineering analysis.

The strategic management issues of building facing researchers also can be examined using the case analysis method. Basically, the case analysis method calls for a careful diagnosis of building’s current conditions (as manifested by its external and internal environments) so that appropriate strategic actions can be recommended in light of strategic tasks.

PLM is the process of handling product data, information and knowledge across a product’s life cycle [3]. The product lifecycle is a concept lent from biology where it describes the recurring change of states for certain organisms. In building, the organisms are exchanged for tangible structure and the state changes are substituted by processes, such as design, production, use, repair, recycling and disposal. Since life-cycle processes differ among the targeted products, they are generalized into three phases stated as “beginning of life” (BOL), “middle of life” (MOL) and “end of life” (EOL). The BOL covers the design and the realization of the product, the MOL concerns the product’s usage and related value adding services, while the EOL typically consists of several optional activities, such as reuse by other customers (second hand), remanufacturing (refurbish used product), material recycling and disposal. Multiple internal and external stakeholders are involved along the life cycle of a product.

The concept of the life cycle is applicable to automation systems whose performance depends on sensors and monitoring systems.

Systems Analysis and Control include technical management activities required to measure progress, evaluate and select alternatives, and document data and decisions. These activities apply to all steps of the systems engineering process.

System analysis activities include trade-off studies, effectiveness analyses, and design analyses. They evaluate alternative approaches to satisfy technical requirements and program objectives, and provide a rigorous quantitative basis for selecting performance, functional, and design requirements. Tools used to provide input to analysis activities include modeling, simulation, experimentation, and test.

Control activities include risk management, configuration management, data management, and performance-based progress measurement including event-based scheduling, Technical Performance Measurement (TPM), and technical reviews.

Process output is dependent on the level of development. It will include the decision database, the system or configuration item architecture, and the baselines, including specifications, appropriate to the phase of development. In general, it is any data that describes or controls the product configuration or the processes necessary to develop that product.

As an example, a distributed ventilation and air conditioning system of an industrial facility is considered. The first level of which uses humidity, temperature and pressure transmitters, analog input
modules, the total number of which is up to 20. Without considering the implementation of the switching level and the level of transmission of accumulated information to a remote computer. It means signaling the output of the measured parameters beyond specified limits, we focus on tasks top level, where a software package is proposed for reading and processing information, generating control solutions, for example, limiting levels life-cycle specific sensor or actuator.

3. Machine Learning Methods
In this paper, we propose the implementation of machine learning methods for analysis of serviceability of an automatic climate control system of a building. Machine learning methods help us to analyze the collected data and use the results to build an online building and facilities monitoring system. Such systems allow us to reform many tasks related to the quality of the production process. The problem of analyzing data consisting of many parameters is solved.

The described approach allows us to build a warning system for emergencies in the production process basing on the problems arose before, and to identify new risks at an early stage. It also helps to identify the possible consequences of accidents at hazardous facilities and to prevent them.

The introduction of methods of machine learning in the system of quality control of products will reduce the amount of manufacturing defect. This is the way to identify low-quality products in the production process. This will reduce the time to eliminate problems and the number of damaged system elements.

The use of methods of machine learning in a similar problem is described in [4]. In this case the k nearest neighbor (kNN) method and the random forest (RF) showed the best performance. The average accuracy value of these two methods reaches 80 – 90% in performed task. There is a relationship between random forests and the k-nearest neighbor algorithm. Methods can be viewed as neighborhoods schemes, which built from a training set.

3.1 K Nearest Neighbors
These algorithm is a non-parametric classification method. An object is classified by a majority vote of its nearest neighbors. k is a positive small integer sometimes k = 1. The nearer neighbors contribute more to the average than the more distant ones.

The training set for this algorithm is a set, in which every item class is known. Thus, during the classifying any object classes of neighbors are known.

k-NN algorithm is sensitive to the local structure of the data. The classification performance can be improved through supervised learning.

Occasionally the k-NN algorithm uses feature extraction when the input data is too large and redundant. Only some of the data points are needed for accurate classification. It is prototypes. The selected points or training data remove from the training set.

3.2 Random Forest
Random forests are a machine learning method for classification and other tasks, which operate by constructing a multitude of decision trees at training time and outputting the class. Random decision forests correct for decision trees’ habit of overfitting to their training set.

Observation of a more complex classifier getting more accurate nearly monotonically in this method. The idea of random subspace selection was influential in the design of random forests. Forest is grown and variation among the trees is introduced by projecting the training data into a randomly chosen subspace before fitting each node.

Random forests allow to handle continuous and discrete features. It has high parallelizability and scalability, but it has the large size of the resulting models.

4. Training Data Collecting
In the course of the work, we collected information and constructed a classifier. As the classified entities, we selected the state of serviceability of the automatic climate control system. As classes, we
distinguish the following states: normal functioning, failure and inefficient configuration. The classes' labels are "normal", "broken" and "inefficient", respectively. They are the most distinguishable from each other, it is simple to simulate and fix them. In the future, it is possible to separate each of them more specifically, but this will require more extensive data collection.

As descriptors for every state to classify, we use parameters presented in table 1.

| I/O Designation | I/O Type   | Function                                                                 | Measurement range |
|-----------------|------------|--------------------------------------------------------------------------|-------------------|
| AI1             | Analog input | The air temperature in the room                                          | 0–40 °C           |
| AI2             | Analog input | Relative humidity in the room                                             | 0–100 %           |
| AI3             | Analog input | The temperature of the coolant in the return line of the heat exchanger  | 10–100 °C         |
| DI1 – DI9       | Digital input | Differential pressure on the separate system parts                      | 1/0               |
| AO1 – AO4       | Analog output | Drive control of different valves                                        | 0–100%            |
| DO1 – DO3       | Digital output | Fan control actuators                                                   | 0–100%            |

A necessary step for applying machine learning algorithms for classification remains the marking of the available data in accordance with the selected classes. In general, we have no algorithms for decrypting the data of a particular application. It is almost impossible to determine exactly what kind of workload each session is having, since the main activities are determined by time intervals (seconds, minutes, hours), for example, while the frequency of packets passing through the network is very high, and the gaps between packets are measured in thousandths of a second.

The quick way (acceptable from the time practice point of view) to receive the marked data for the training can be described as follows. It is necessary to simulate all possible types of activity separately and shoot traffic for a certain period to be sure most part of recorded data represents only one type of activity. After that, all sessions found in each portion are marked as belonging to the class of one specific activity.

5. Implementation and Testing

The input data for training the classifier is a table of attributes, which are listed for each state values and descriptors mapped to the class to which belongs to the state. Input data for classifier is a table with unknown classes’ values. They will be identified as a product of classification.

The process of implementing the classifier is inextricably linked with the measurement at each step of the quality of the classification, and therefore it is necessary to introduce the metrics used to assess the effectiveness of classifiers.

To learn about the quality of the classifier's work, one can consider the ratio of the correct operations of the system to all of its positive answers, or precision:

\[ P = \frac{T_p}{T_a}, \tag{1} \]

where \( T_p \) is the number of correct classifications for one class, \( T_a \) is the total number of cases when the classifier made a choice in favor of this class.
In addition, we can consider a value that shows the relationship between the correct operations of the system and all elements of the sample, which, according to the expert, belong to the selected class. This value is called recall:

\[ R = \frac{T_p}{N_c}, \#(2) \]

where \( T_p \) is the number of correct classifications for one class, \( N_c \) is the actual size of this class, according to expert estimates.

Often considered to evaluate the convenience of F-measure (F1-score). This value, which shows the average value of the completeness and accuracy. Usually F1-score is calculated as the harmonic mean:

\[ F = 2 \frac{P \cdot R}{P + R}, \#(3) \]

In this paper we use the implementation of these algorithms, presented in the module for machine learning scikit-learn of the Python language.

Thus, by separating from the general sample a part with a relatively uniform content of classes we can form a training and test sample. We will do it with the help of the tools provided in the scikit-learn package, and we will divide the sample so that the instructor makes up 30% of the initial sample, and the test sample — 70%. In this case, the correlation of classes within the subsamples will remain the same as in the original sample.

The diagram demonstrates the efficiency of both (kNN and RF) algorithms (figure 1).

![Diagram showing the efficiency of kNN and RF algorithms](image)

**Figure 1.** The efficiency of kNN and RF algorithms

6. Conclusion

The proposed control algorithm calculates various combinations that can be implemented in the form of circuit solutions in accordance with the tasks assigned to the automation system to ensure the microclimate in the room.

It is supposed to use these methods and algorithms for the optimization of energy consumption, in order to determine the sequence of control actions to ensure the given microclimate parameters under changing environmental influences. Then the control system, evaluating information from sensors both about its current state and the environment, will be able to calculate possible variants of circuit solutions and offer options for replacing elements, to decide on the sequence of actions, thus synthesizing the algorithm to achieve the goal. The intellectual control system will allow correcting
the algorithm of actions and changing its structure, proceeding from the objective function and external circumstances.

Expansion of the experimental base as well as the development of more accurate methods for marking the training datasets will allow later to clarify the calculations, draw new conclusions, etc.

The similar approach is sound for residential facilities. In this case we should take the convenient dataset with facility parameters and apply it as a training set for a machine learning method. As the result we will get a proper classifier.

References

[1] Bonastre A, Capella J V and Ors V 2002 A new generic architecture for the implementation of intelligent and distributed control systems IEEE 2002 28th Annual Conf. of the Industrial Electronics Society (Sevilla) 3 1790–5

[2] Thramboulidis K 2013 IEC 61499 vs. 61131: A Comparison Based on Misperceptions Journal of Software Engineering and Applications 405–15

[3] Vadoudi K, Allais R, Reyes T and Troussier N 2014 Sustainable Product Lifecycle Management and Territoriality: New Structure for PLM 11th IFIP International Conference on Product Lifecycle Management (Yokohama) 475–84

[4] Lim Y, Kim H, Jeong J, Kim C, Kwon T and Choi Y 2010 Internet Traffic Classification Demystified: On the Sources of the Discriminative Power Proc. of the 6th Int. Conf. on Emerging Networking Experiments and Technologies (Philadelphia) 9–20

[5] Ho T K 1998 The Random Subspace Method for Constructing Decision Forests IEEE Transactions on Pattern Analysis and Machine Intelligence 20(8) 832–44

[6] Erman J, Mahanti A and Arlitt M 2006 Internet Traffic Identification using Machine Learning Global Telecommunications Conf. (San Francisco, California) 1–6