IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Intelligent Analysis of Data Systems for Defects in Underground Gas Pipeline

Volodymyr Yuzevych
Department of Electrophysical Methods of nondestructive testing
Karpenko Physico-Mechanical Institute of the National Academy of Sciences of Ukraine
Lviv, Ukraine
yuzevych@ukr.net

Ruslan Skrynkovskyy
Department of Business Economy and Information Technology
Lviv University of Business and Law
Lviv, Ukraine
uan_lviv@ukr.net

Bohdan Koman
Department of System Design
Ivan Franko National University of Lviv
Lviv, Ukraine
sonce_28@ukr.net

***Abstract*—A method of functioning of intelligent software and hardware complex for monitoring system of an underground gas pipeline and cathodic protection devices using data and knowledge bases is proposed.**

Keywords—data mining, gas pipeline, intelligent software, hardware, monitoring, cathodic protection, databases.

I. Introduction

Efficiency of protection of underground steel gas main-pipelines (USGP) against corrosion depends on a state and quality of pipes. The pipes are protected by coatings and the cathodic protection equipment (CPE). The surface of the pipes becomes covered with defects during operation, in particular, the corrosive ones. Coefficient of corrosion inhibition depends on the defects.

Defects of the pipelines (USGP) can be divided into two classes [1]:

1 – defects of continuity (defects of a material) featuring with local discontinuity of the material (steel);

2 – shape defects revealing as local changes of the pipeline elements during production process or operation.

We will restrict ourselves to consideration of the defect system of a pipeline (USGP) and processes of the corrosion-mechanical fracture, in particular, stress corrosion cracking (SCC), which are appropriate to analyze on the base of a software and hardware complex.

The problem is to evaluate efficiency and quality of the software and hardware complex (SHC) to ensure conditions of the pipeline operation with accounting for the CPE, the SCC processes, and application of data and knowledge bases.

II. An Analysis of Recent Publications on the Problem

We have addressed a perspective of application of the functioning method of intelligent SHC to monitor and ensure safety of the pipeline operation by means of databases [2]. Such database comprises the following: data of continuous monitoring of the pipeline actual state, technical documentation data, results of internal pipe diagnostics, data of electrometric measurements, data of visual and dimensional control, data of periodic non-destructive measurements, corrosive and mechanical characteristics of a metal, criteria of limit states of pipeline systems with damages [3].

The information presented deals with a technique of control of complex interacting processes allowing development of an intelligent system based on knowledge about peculiarities of the interaction of process participants in a specific subject field [3]. The technique is developed on the base of models, methods and algorithms of the ontological analysis and data processing. The novelty of corresponding mathematical model underlying the intelligent system consists in a combination of various mechanisms of logical conclusion, when making a decision in a problem situation, based on all available information about the subject area [3].

Diagnostic weight and value of attributes proposed to use for the process optimization have been formulated [4].

A new approach is proposed which solves the problem of automated intelligent diagnostic using machine learning techniques [5].

III. Formulation of the Research Goals

The purpose of the study is to develop methodological and theoretical foundations of diagnostics of states of a complex technical system such as the system of defects of a pipeline (SDP), and corresponding technological processes based on analysis and processing of knowledge under conditions of uncertainty as well as development of new intelligent methods and tools of multi-criterial diagnostics of the object’s states (SDP).

The diagnostics relates to the pipeline critical situation. The critical situation (state) emerging during the process of the pipeline operation reveal itself as a destruction of the pipe metal due to the stress corrosion cracking (SCC). Main informational parameters of the pipeline are pipe diameter, wall thickness, internal pressure, yield and strength limits of material (steel), and energy characteristics of the surface layers.

IV. Information on Combination of Current and Potential Measurements

Recognition of the object’s parameters of state (SDP) is performed by methods of solution finding on the base of logical rules and precedents of solutions of the diagnostics problems.

978-1-5386-2874-4/18/$31.00 ©2018 IEEE

A problem of the value of information obtained during diagnosis of underground gas pipelines by means of contactless current measurements (CCM) is raised [6].

The implementation of the CCM method in the device for measuring polarization potential (MPP) provides an opportunity to use the MPP in detection of the USGP damaged insulation both at alternating current (Pearson's method) and by the potential difference (gradient method) on the surface of the soil [6].

To determine places of damage of the pipe protective cover, two electrodes are placed on the soil surface above the pipeline and alternative voltage V_{gg} (Pearson's method) as well as dc voltage U_{gg} (method of transversal gradient of potential) are measured [6]. Combination of these measurements allows determining polarization potential U_P by the formula [6]:

$$U_P = U_{mg} - V_{mg} \times U_{gg} / V_{gg} . \quad (1)$$

Here, U_{MG}, V_{MG} – correspondingly the constant potential difference between metal (pipe) and comparison electrode placed in corrosive environment.

To implement the CCM method, a hardware with electronic memory to measure constant and variable electrical voltages as well as polarization potential (PP) [6] was designed. The MPP method can be used to control pipelines and other metal constructions by parameters indicated in (1). In the version of equipment in [6], a GPS module was installed allowing additionally to fix place and time of control. This facilitates considerably processing and documenting the data arrays, in particular, combination of measurements of potential with measurements of current during their processing and determination of the USGP parameters.

In formula (1), second term is an ohmic component of the potential characterizing state of the protective coating. Ratio of the ac voltage to the dc one, $Vgg / Ugg = k$, is the measured harmonic factor showing ratio of loss of the measured ac component to the dc one [6]. It is used to assess losses of the dc component of the CPE on the pipe sections and to determine distribution of the current density of cathodic protection.

The use of GPS modulus in the developed device provides automatic fixation of geographical coordinates and time of the current measurement as well as the USGP depth, which greatly facilitates processing and documentation of the inspection results [6]. It is necessary for computing the current density, its losses, and transition resistance "pipe–ground" in different parts of the USGP. The CMC together with MPP allow determining distributions of the current density of cathodic protection, resistivity of soil surrounding the pipe, and resistivity of the protective insulation on different sections of the USGP [6].

V. PRINCIPLES OF DIAGNOSTING THE PIPELINE DEFECTS SYSTEM

Under conditions of operation and monitoring, conditions of functioning of objects of the USGP system and CPE change. Let us consider project of the USGP monitoring which main element is the SDP monitoring.

Underground pipelines are in soil environment. Anticorrosion coating of pipelines can be metal and film-type. Between metal of the pipeline and coating, defects of the cavern type are formed. Water from the soil environment penetrates into the defects that is the aqueous electrolyte solution is created. Electrochemical reactions and adsorption processes are characteristic for such type of defects [7].

There are presented mathematical relationships and methods of evaluation of physical and electrochemical characteristics of the interphase layer at the metal-solution boundary of the electrolyte together with corresponding algorithms and software supplemented by the base of numerical data [7]. These data are information means that is a base of information technology of selection and processing data concerning assessment of energy characteristics of interphase layers and overstrain characterizing the boundary metal-environment and conditions of metal plastic deformation near the cavern tip [7]. The database of such kind allows particularly to describe state of the interphase metal layer with adsorbed impurities in corrosive environment of the soil electrolyte type.

Number of parameters and volume of information required for ordering and evaluation in order to take a decision in critical situation are growing. For this purpose, both the database and knowledge base are used.

The knowledge base is part of the decision support system (DSS). The DSS has to contain information, which is partially analogous to the [2] and characterizing the following:

- conditions of implementation of physical and chemical processes;
- probable results of physical and chemical processes;
- conditions of the CPE functioning;
- time periods related to the USGP reliability control;
- degree of formalization of the decision making process with accounting for normative documentation;
- probable critical situations, associated with the risk of the USGMP and CPE system, with a note of the reasons and conditions of their occurrence, and measures to correct deviations from the operating modes.

The most important for information selection and processing is the last sixth paragraph, associated with making decisions for the models of knowledge representation and their following usage in the knowledge base.

To implement procedures of monitoring the USGP system and CPE with accounting for a feedback, the diagnostic weight of an attribute is introduced. If in the result of the study it is revealed that the attribute k_j has for the given object a value of k_{js}, then this value is called the realization of the attribute k_j [8]. Then information about specific state (diagnosis) D_i (i=1, 2, …, n – total number of states under consideration), which possesses a state of the attribute k_{js}, can be defined by the formula in the information theory [8]:

$$Z_{Di}(k_{js}) = \ln\left(P(D_i / k_{js}) / P(D_i)\right), \quad (2)$$

where $P(D_i/k_{js})$ – probability of the state D_i provided that the attribute k_j obtained value of k_{js};

$P(D_i)$ – a priori probability of state.

For specific calculations, the diagnostic weight of presence of the attribute k_j within the interval s can be written in an equivalent form relatively to (2) analogously to [8, 9]:

$$Z_{Di}(k_{js}) = \ln\left(P(k_{js}/D_i)/P(k_{js})\right). \tag{3}$$

Equivalency of relationships (2) and (3) follows from identity [8, 9] based on the probability theory:

$$P(k_{js}) \cdot P(D_i/k_{js}) = P(D_i) \cdot P(k_{js}/D_i) = P(k_{js}D_i). \tag{4}$$

From the point of view of information theory [8, 9], the value $Z_{Si}(k_{js})$ characterizes information about the D_i state, which has attributes k_{js}. In expression (4), the $P(k_{js}/D_i)$ value is the probability of occurrence of interval s of the attribute k_j for an element of the system in the D_i state, and $P(k_{js})$ is the probability of simultaneous occurrence of corresponding interval of each attribute in every state considered.

The $P(k_{js})$ value is determined according to [8, 9]:

$$P(k_{js}) = \sum_{i=1}^{n} P(D_i) \cdot P(k_{js}/D_i). \tag{5}$$

With accounting for the relationship (5), we will obtain resulting expression:

$$Z_{Di}(k_{js}) = \ln\left(P(k_{js}/D_i) / \left(\sum_{i=1}^{n} P(D_i) \cdot P(k_{js}/D_i) \right) \right). \tag{6}$$

Let us introduce important concept of the information theory – information or diagnostic value of the $Z_{Di}(k_j)$ study for the USGP system and CPE. The $Z_{Di}(k_j)$ diagnostic value by the attribute k_j for the state D_i is a volume of information introduced by all variants of realizations of this attribute in establishing a corresponding state [9]. Expression $Z_{Di}(k_j)$ for the m-bit attribute is proposed to write in form of [9]:

$$Z_{Di}(k_j) = \sum_{s=1}^{m} P(k_{js}/D_i) Z_{Di}(k_{js});$$

$$Z_{Di}(k_j) = \sum_{s=1}^{m} P(k_{js}/D_i) \log \frac{P(k_{js}/D_i)}{\sum_{i=1}^{n} P(D_i) \cdot P(k_{js}/D_i)}. \tag{7}$$

Diagnostic value of the process of diagnosing $Z_{Di}(k_j)$ for the USGP system and CPE considers all possible realizations of the attribute, which corresponds to a specific state (diagnosis) and is certain mean expected value, more precisely, it is mathematical expectation of the information value introduced by separate realizations of the attribute in this state [8]. Because the $Z_{Si}(k_j)$ is attributed to a single specific state, it is commonly referred to as the partial diagnostic value by the attribute k_j [4, 9].

A technique to predict situations of the USGP system and CPE is proposed, consisting in accounting for main informative parameters by means of artificial neural networks, as well as were defined directions of application of the “data mining” methodology to control limit states of pipes on the base of the strength and yield criteria.

The neural network with direct signal distribution to assess defects on the gas pipe surface is proposed. A mean square error averaged by the number of the neural network output variables and calculated on the base of predicted and real values of the test sample by formula [10] was used to determine efficiency of the neural network being studied:

$$E = \frac{1}{N \times K} \sum_{i=1}^{N} \sum_{k=1}^{K} \left(y_{ij}^{R} - y_{ij}^{P} \right)^2. \tag{8}$$

Here, y_{ij}^{R} – value of the *i-th* output variable of neural network for *j-th* training or test example;

y_{ij}^{R} – predicted value of the *i-th* output variable of neural network for *j-th* training or test example;

N – number of examples in the training or test sample;

K – number of output variables of the neural network.

Activation functions are selected by an exhaustive search over the given set [11] including:

- threshold;
- symmetric threshold;
- sigmoid;
- stepwise sigmoid;
- symmetric sigmoid;
- stepwise symmetric sigmoid;
- Gaussian;
- symmetric Gaussian;
- stepwise Gaussian;
- Elliot function;
- symmetric Elliot function;
- linear function;
- symmetric linear function;
- symmetric sine;
- symmetric cosine;
- sine;
- cosine.

To optimize information flows and improve the project configuration, let us use the quality functional $J(P_k, FB(P_k))$ with accounting for sensitivity coefficient β_R and a feed-back [12]:

$$J(P_k, FB(P_k)) = \int_{t_0}^{t_k} f(\bar{y}, \bar{u}, \bar{s}, \beta_R)\,dt \Rightarrow opt, \tag{9}$$

where $\bar{y}$ – vector of specific impacts upon the OP ($y_j(t)$ – vector components (key parameters for the USGP system and CPE), $j = 1, 2, \ldots, n$);

$\bar{u}$ – control vector of information flows;

$\bar{s}$ – vector of indeterminate perturbations;

P_k – information flows ($k = 1, 2, \ldots, m$);

m – total number of information flows P_k considered in the given project;

$[t_0, t_k]$ – time interval, in which the process is considered (formation of optimal values of parameters corresponding to P_k;

$f(\bar{y}, \bar{u}, \bar{s}, \beta_R)$ – function reflecting quality index of the project;

β_R – sensitivity coefficient;

$FB(P_k)$ – function characterizing feed-back between flows P_k and project's environment with accounting for sensitivity coefficient β_R and expert opinions, of experts;

opt – optimization symbol;

t – time.

To implement above processes, it is proposed to use the intelligent predictive control system of technological processes (IPCS TP) pooled into unified information complex analogously to [13]. Here, we recommend to combine the information complex, equipment for measurements of constant and alternating voltages and determination of polarization potential [6] into unified information space.

This allows the following:

- to expand considerably spectrum of tasks associated with control system;
- provide personnel with information on state of hardware and software;
- decrease risks and improve reliability of the measuring hardware complex;
- to increase considerably general informativity of components of corresponding information space;
- to automate procedure of formation of normative-technical documentation;
- to attract information on strength criteria of elements of metal constructions with consideration of energy characteristics of surface layers [14].

During the process of monitoring the USGP system and CPE, values of physical and chemical parameters are recorded through set time intervals and form a system of interrelated time series. The possibility of free access to information saved in databases is an important perspective, because in the future only small part of the general volume of information can be necessary.

For the time being, prediction of parameters characterizing the state of a technical object is made as a rule on the base of classical models of auto-regression – integrated moving average [15]. Existing approaches to information processing in control systems as a rule do not provide required accuracy of prediction, leading to increase of probability of making erroneous decision in the object control [15]. Therefore, the prediction algorithms need to be improved. Let us formulate the problem of a variable values in a general case for a discreet time analogously to [13]:

$$Y_{k+\tau} = F_Z\left(Y_{k-m}, \varphi\left(Y_k, V_k, y_j, k\right)\right). \qquad (10)$$

Here, $Y_{k+\tau}$ – vector of prediction for anticipation interval;

τ; k – actual time interval (clock time);

$Y_{k+\tau}$ – vector of parameter values (e.g. one of the parameters $y_j(t)$) with memory depth m;

Y_k – vector of values of prehistory of corresponding parameter; V_k – white (Gauss) noise;

$F_Z(\cdot)$ – generalized transformation function (method, algorithm);

$\varphi(\cdot)$ – linear independent functions characterizing properties of the time series.

Prediction of critical state of the system studied can be done by a criterion of the prediction error minimum:

$$\varepsilon = |Y_{k+\tau} - Y_{k+\tau 0}| \Rightarrow \min \text{ when } \tau = \tau_z, \qquad (11)$$

where $Y_{k+\tau 0}$ – actual values of technological variable;

τ_z – set value of anticipation interval.

The solution of the problem of prediction for the USGP system and CPE (i.e. finding generalized function F_Z and parameters $y_j(t)$) is formed by means of interpolation of a temporary series and extrapolation of values of predicted series by its initial values by means of functions $\varphi(\cdot)$ in order to ensure criteria (8) and (9).

Results of prediction of the technical state allow not only foresee places of failure of the structural elements and outage of gas transportation system but also to determine the optimal periodicity of running diagnostics procedure and repairs.

Application of algorithms for predicting technical condition of a pipeline allows considerably improve the efficiency of the technical diagnostic system, which in turn leads to improvement of the operation of monitoring equipment described in [6]. In this case, risks of occurrence of the main crack will be reduced, and such situation will assist to provide the desired level of security of gas transportation.

When running procedure of predicting technical state of the USGP system and CPE, not only data obtained in the actual time are used. Results of previous measurements are considered too. Knowledge of previous results of the object diagnosing indicates the need to use databases with the large volume of memory. That said, saving all information about the state of the USGP system and CPE for the whole period of operation in databases is not worthwhile because of many reasons.

In particular, results of measurements of currents and potentials prior to a pipeline repair and afterwards may not correlate between themselves. The reason can be that after repair some elements of structures can be replaced, for example pipes in some sections. As a consequence, data processing can produce incorrect results of the prediction, and thus a certainty of the results of diagnostic procedure of the USGP system and CPE will be minimal. Another reason can be excessively large volume of data.

To improve the efficiency of the intelligent information system (IIS), it is practical to include an inductive component (IC) into the IIC structure, which allows automatically complete the database [16]. This, in turn, set forth a problem of unification of deductive and inductive formalisms into a single system that leads to the necessity to develop structural and functional models of the IIC with the inductive component (IIC IC), as well as algorithms of efficient search of a solution as a base for development of the quality IIC independently of their complexity and character [16]. We propose to use the basic intelligent component in the process of neural network spectral analysis, which is able to adapt to requirements of a specific sensor [17]. An artificial neural network does not give possibility to create new, unique architecture, but only allows to bring in an ordinary software and hardware for calculations. Such component features with high reliability, ability to adapt to a specific application, and usage of design principles ensuring the possibility of a simple expansion of the intelligent component potential by means of completion with new algorithmic solutions [17].

A large amount of experimental data is provided in article [6]. Processing the results of experiment [6] allows to determine distributions of cathodic protection current densities, specific resistances of surrounding the pipe soil and protective isolation on the different sections of the USGP. Corresponding information will allow to predict the resource of separate areas of gas pipeline and determine the terms of repair.

VI. Conclusions

A method of functioning of intelligent software and hardware complex for the monitoring system of an underground steel gas pipelines (USGP) and cathodic protection equipment (CPE) using data and knowledge bases is proposed. The data and knowledge bases for monitoring the USGP system and CPE comprise the following:

- data of continuous monitoring of information about actual state of the system of corrosion defects,
- data of normative-technical documentation, data of diagnosing underground gas pipelines by means of contactless current measurements,
- data on critical risk-related cases for the USGP system and CPE with indication of reasons and conditions of their occurrence, and also procedures concerning correction of deviations from a pipeline operating modes,
- data on the control measurements of currents and potentials,
- data of nondestructive control with accounting for the stress corrosion cracking (SCC),
- corrosion-mechanical characteristics of a metal,
- strength and yield criteria for a pipe material.

References

[1] A. Cosham, and P. Hopkins, “An Overview of the pipeline defect assessment manual (PDAM),” proceedings of 4th International Pipeline Technology Conference, Oostende, Belgium, pp. 1-12, May 2004.

[2] N. G. Gubanov, S. V. Susarev, Yu. I. Steblev, and V. I. Batishchev, “The method of functioning of an intelligent software and hardware complex for monitoring and ensuring the safety of pipeline operation using a database,” Proceedings of the XIX International Conference “Complex Systems: Control and Modeling Problems”, Samara, Russia, pp. 96-102, September 2017.

[3] O. V. Barmina, and N. O. Nikulina, “Intelligent system for interactive business processes management in project-oriented organizations,” Ontology of designing. vol. 7, no. 1 (23), pp. 48-65, 2017. doi: https://doi.org/10.18287/2223-9537-2017-7-1-48-65 .

[4] V. Yuzevych, O. Klyuvak, and R. Skrynkovskyy, “Diagnostics of the system of interaction between the government and business in terms of public e-procurement,” Economic Annals-XXI, vol. 160, no. 7–8, pp. 39–44, Oct. 2016. doi: https://doi.org/10.21003/ea.v160-08 .

[5] N. O. Komleva, K. S. Cherneha, B. I. Tymchenko, and O. M. Komlevoy, “Intellectual approach application for pulmonary diagnosis,” 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Aug. 2016. doi: https://doi.org/10.1109/dsmp.2016.7583505 .

[6] R. M. Dzhala, B. Y. Verbenets’, M. I. Mel’nyk, A. B. Mytsyk, R. S. Savula, and O. M. Semenyuk, “New Methods for the Corrosion Monitoring of Underground Pipelines According to the Measurements of Currents and Potentials,” Materials Science, vol. 52, no. 5, pp. 732–741, Mar. 2017. doi: https://doi.org/10.1007/s11003-017-0016-8 .

[7] R. Dzhala, V. Yuzevych, and M. Melnyk, “Modeling the adsorption connections and their influence on informational parameters of metal-electrolyte interface,” Bulletin of the Lviv Polytechnic National University, Series “Computer Sciences and Information Technologies”, no. 826, pp. 185–190, 2015.

[8] C. E. Shannon, “A Mathematical Theory of Communication,” Bell System Technical Journal, vol. 27, no. 4, pp. 623–656, Oct. 1948. doi: https://doi.org/10.1002/j.1538-7305.1948.tb00917.x .

[9] V. A. Sokolov, “Diagnostic weight of signs and diagnostic value of examination in recognition of the states of elements of building system,” Engineering and Construction Journal, no. 3(13), pp. 27-31, 2010.

[10] N. A. Matveeva, L. Y. Martynovych, and U. V. Lazorenko, “Choice of the optimal neural network for determining defects in composite materials,” Bulletin of the Kherson National Technical University, no. 3(50), pp. 66-70, 2014.

[11] P. Sibi, S. Allwyn Jones, and P. Siddarth, “Analysis of Different Activation Functions Using Back Propagation Neural Networks,” Journal of Theoretical and Applied Information Technology, vol. 47, no. 3, pp. 1264-1268, 2013. http://www.jatit.org/volumes/Vol47No3/61Vol47No3.pdf .

[12] N. Krap, V. Yuzevych, “Neural Networks as a tool for managing the configurations of tourist flow projects,“ Management of Development of Complex Systems, no. 14, pp. 37-40, 2013.

[13] I. Gulina, A. Martynenko, A. Gulin, “Construction of intelligent predictive control systems for nonlinear technological processes,” Information Processing Systems, no. 3 (149), pp. 101-105, 2017.

[14] V. M. Yuzevych, R. M. Dzhala, and B. P. Koman, “Analysis of Metal Corrosion under Conditions of Mechanical Impacts and Aggressive Environments,” Metallofizika i Noveishie Tekhnologii, vol. 39, no. 12, pp. 1655–1667, Mar. 2018. doi: https://doi.org/10.15407/mfint.39.12.1655 .

[15] G. E. P. Box, G. M. Jenkins, Time series analysis: forecasting and control. San Francisco, CA: Holden-Day. 1976.

[16] O. Nelles, Nonlinear System Identification: From Classical Approaches to Neural and Fuzzy Models. Berlin: Springer, 2001. https://www.springer.com/us/book/9783540673699 .

[17] J. H. Holland, Adaptation in natural and artificial systems. An introductory analysis with application to biology, control and artificial intelligence, London: Bradford book edition, 1994.