RESEARCH PAPERS
FACULTY OF MATERIALS SCIENCE AND TECHNOLOGY IN TRNAVA
SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA

2013 Number 33

APPLYING DATA MINING METHODS IN ANALYSIS OF STRUCTURAL RELATIONSHIPS IN THE SYSTEM OF RANDOM PROCESSES

Michaela HORALOVÁ KALINOVÁ, German MICHAĽČONOK

Abstract

The aim of this paper is to present the possibilities of applying data mining techniques to the problem of the structural relationships analysis in the system of stationary random processes. In this paper, we will define the basic problem areas of analysis of structural relationships and we will show how to solve these problems by the methods of data mining. Individual solutions to selected problem areas will be demonstrated on a model which was designed in MATLAB. In the experiment, we will identify the existence and the nature of the process relationships in the system of stationary random processes based on the results of analyses.

Key words

random processes, analytical methods, data mining, complex systems

INTRODUCTION

Growing complexities of processes in the modern technical systems and their decentralized nature have lead to the necessity to apply the new identification technologies of real system structures. One of the important phases of the analysis of complex processes is to identify the structural relationships. The process of structural analysis can be based on a priori information, on physical laws, but also on the empirical experience concerning the identified processes. In a complex process which consists of a set of processes at the defined resolution level, it is often impossible to detect the structural relationships typical for high complexity based on a priori information, and therefore it is needed to apply the modern approaches, methods and instruments in the process of structural analysis (3).

The base of experimental data has the significant impact on the accuracy of the structural analysis. Since the various forms of decentralization and inconsistency of data are in the complex processes, method of data collecting and pre-processing is very important in the analysis process. We also have to suppose the great requirements for computing resources in

Mgr. Michaela Horalová Kalinová, doc. Ing. German Michaľčonok, CSc. - Department of Information Technology and Automation, Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava, Paulínska 16, 917 01 Trnava, Slovak Republic
e-mail: horalovakalinova@gmail.com, german.michalconok@stuba.sk

the analysis phase. Moreover, many processes in complex systems are not explicitly deterministic, but have a random character. Structural analysis, which assumes collecting and processing the large volumes of data of various informational value, can have a specific character, particularly in the case of random processes (5). Data mining methods provide a possible solution in the tasks. One of the modern analytical methods is the process of data mining, which is one stage of a complex process of knowledge discovery in databases presented as KDD (1).

ANALYSIS OF STRUCTURAL RELATIONSHIPS IN THE SYSTEM OF RANDOM PROCESSES

Random process is a random function of time, which is represented by an infinite set of random variables. Sets of values of random variables represent the application of a random process. The values of the individual applications of random process can be experimentally measured and statistically evaluated. The main aim of process of structural analysis in processes of the system is the analysis of relationships. Relationship between processes can be understood as the process of transforming a set of input (independent) processes into the output (dependent) processes.

The main aim will be to show how to detect the existence of the transformation process in the system of random processes and to estimate their character within the analysis of structural relationships. We assume a set of stationary random processes in which we will distinguish a set of input (independent) processes and a set of output (dependent) processes which result from the transformation input processes.

In systems of processes, we can assume the transformations of various characters based on the nature of objects (elements of the system) which cause the transformation input process or a set of input processes to the output process. Based on the transient characteristics of objects, we can consider proportional, derivative and integrative character transformation relations.

From the range of the data mining methods, we used regression analysis methods and methods of neural networks, for solving the tasks. Especially the methods of neural networks represent a powerful tool in the primary analysis and extensive database (2, 4).

DESCRIPTION OF THE EXPERIMENTS

The aim of the experiment was to show, how to detect the existence of the transformation process of proportional and derivative nature. We observed statistical indicators in the active experiment and in the analysis of the values generated in different frequency bands.

A model of system processes was designed in MATLAB as shown in the following scheme:

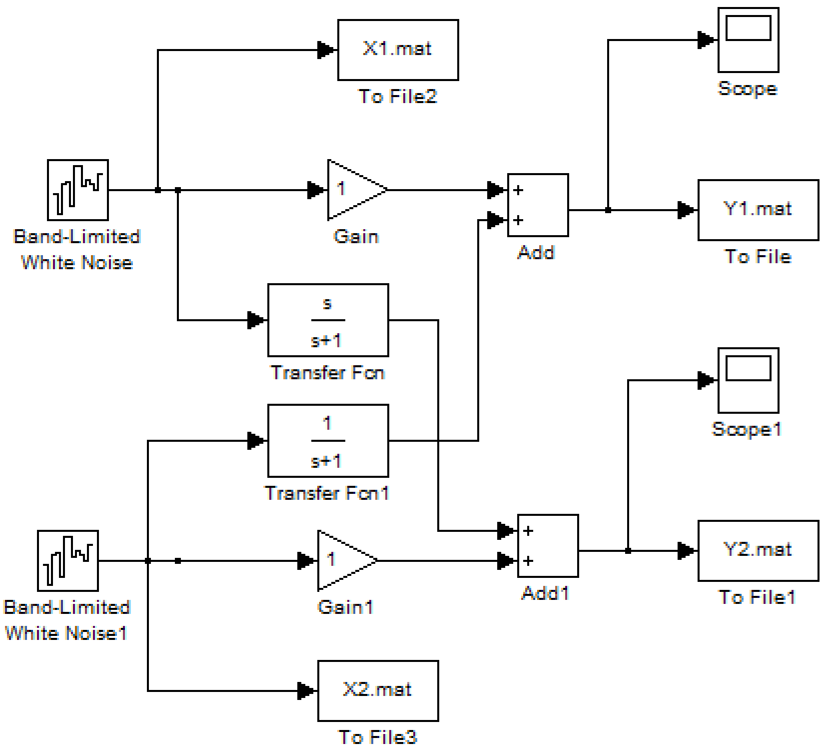

Fig. 1 *System model in Matlab*

In order to cover the highest possible frequency range, random signals were generated with normally distributed values and sampling period Ts = 0.1s, measurement time Tm = 100s and repeatedly with a sampling period Ts = 10s, and the measurement time Tm = 10000s on the inputs X1, X2. We obtained the values of realizations in two different frequency ranges, and these are the values in the range of 0.02 Hz - 5 Hz and 0.00002 Hz - 0.05 Hz by Shannon-Kotelnikov theorem. Then we analyzed the influence of realizations of input processes on the individual output processes in the frequency ranges through selected data mining techniques, i.e. using the basic linear model (LM), a linear regression model (LRM) and a linear predictive model of neural network (NS). We monitored selected statistical indicators in interpreting the results. In the linear regression model, it was primarily a significance level model (Sig.), the coefficient of determination (R Square) which indicates the proportion of explained variability by the model, the correlation coefficient (R) indicating the tightness of the linear dependence, as well as the values of the coefficients of the linear representation especially the value of the standardized Beta coefficient which expresses the percentage impact of the j-th independent variable on the dependent variable. In case of the linear model and the predictive neural network model, it was the value which determines the accuracy and corresponds to the percentages of the determination coefficient. The results of analyses for the individual frequency ranges are shown in the following tables:

THE INFLUENCE OF X1, X2 REALIZATIONS TO THE Y1 REALIZATION Table 1

	LM	**LRM**				**NS**
	Acc. [%]	Sig.	R	R Square	Coeff.	Acc. [%]
Y1 - X1, X2	95.7	0	0.979	0.959	X1, X2	95.7
	Coeff.	**B**	**Std. Err**	**Beta**	**Sig.**	
	X1	0.994	0.007	0.979	0	
	X2	0.098	0.007	0.096	0	

THE INFLUENCE OF X1, X2 REALIZATIONS TO THE Y1 REALIZATION   Table 2

	LM	LRM				NS
	Acc. [%]	Sig.	R	R Square	Coeff.	Acc. [%]
Y1 - X1, X2	99.9	0	1.000	1.000	X1, X2	99.9
	Coeff.	**B**	**Std. Err**	**Beta**	**Sig.**	
	X1	1.000	0.000	0.728	0.000	
	X2	1.000	0.000	0.721	0.000	

THE INFLUENCE OF X1, X2 REALIZATIONS TO THE Y2 REALIZATION   Table 3

	LM	LRM				NS
	Acc. [%]	Sig.	R	R Square	Coeff.	Acc. [%]
Y2 - X1, X2	97.4	0.000	0.987	0.975	X1, X2	97.4
	Coeff.	**B**	**Std. Err**	**Beta**	**Sig.**	
	X1	0.997	0.007	0.691	0.000	
	X2	1.011	0.007	0.708	0.000	

THE INFLUENCE OF X1, X2 REALIZATIONS TO THE Y2 REALIZATION   Table 4

	LM	LRM				NS
	Acc. [%]	Sig.	R	R Square	Coeff.	Acc. [%]
Y2 - X1, X2	67.3	0.000	0.820	0.673	X1, X2	67.3
	Coeff.	**B**	**Std. Err**	**Beta**	**Sig.**	
	X1	0.970	0.32	0.556	0.000	
	X2	1.048	0.31	0.606	0.000	

EXPERIMENTAL RESULTS

Based on the values of statistical indicators, it was shown that the process X1 has a significant impact on the process Y1 and the X2 realization has only a minimal effect on the process Y1 in the experiment with the sampling period Ts = 0.1s (Table 1). The values of Beta coefficients showed that the impact of the realization X2 on the realization Y1 significantly increased by narrowing of the frequency range (Table 2). It is seen in the results of the analysis of realizations of processes X2 and Y1 that the characteristic of proportional element with the inertia, which caused filtering out the high frequency components, caused the effect. Since the value of non-standardized coefficient for the X1 realization remained identical in both experiments, it confirms the assumption that between the X1 and Y1 realizations, there exists purely a relationship of proportional transformation character due to the presence of the proportional element without inertia.

Analyzing the transformation relations of realization of an independent process on the dependent process Y2, we came to the following result. It was seen based on the values of statistical indicators in the experiment with the sampling period Ts = 0.1 that the X1, X2 realization have almost the same effect on the realization of Y2 which is statistically significant (Table 3). The value of determination coefficient decreased significantly by narrowing the frequency range, and thereby the proportion of unexplained component increased (Table 4).

Since the primary property of the derivation element is its ability to transmit high frequency and to filter low frequency range, this result confirmed the presence of the transformational relation of derivation character. From the values of other statistical indicators, we were unable to clearly identify which relation it was. If we wanted to define the nature of relations more precisely, we would have to identify each relation separately. We could reach it by gradually disconnecting resources of realizations of random processes and then repeating the experiment.

CONCLUSION

The results of the experiment show that using the chosen methods of data mining we can obtain the information about the mutual interactions between the input (independent) and output (dependent) processes in the stationary stochastic processes. By monitoring the changes of the statistical indicators in each frequency range, we can reveal the existence of the derivative and integrative character of transformation relations.

The experiments showed that the problems of structural relationships in the systems of random processes can be solved using the data mining methods, thus providing a powerful mechanism for solving a wide range of tasks of the structural analysis.

References:

1. ADELMAN, S., MOSS, L. T. 2000. *Data Warehouse Project Management*. Addison-Wesley Professional. ISBN 10: 0-201-61635-1, ISBN 13: 978-0-201-61635-4
2. HURAJ L., REISER, H. 2009. VO Intersection Trust in Ad hoc Grid Environments. In: *Fifth International Conference on Networking and Services (ICNS 2009).* Valencia, Spain, IEEE Computer Society, pp. 456-46.
3. OPPENHEIM, A.V., WILSKI, A.S., HAMID, S. 1997. *Signals and systems*. Prentice Hall. ISBN 978-0138147570
4. SKRINAROVÁ, J., HURAJ, L., SILÁDI, V. 2013. A neural tree model for classification of computing grid resources using PSO tasks scheduling. *Neural Network World,* **23**(3), pp. 223-241.
5. SKŘIVÁNKOVÁ, V. 2004. *Náhodné procesy a ich aplikácie (Random processes and their applications)*. Košice: UPJŠ. ISBN 80-7097-542-3

Reviewers:

doc. Mgr. Róbert Vrábeľ, PhD. – Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava

visiting Prof. Ing. Augustín Gese, CSc. – Faculty of Materials Science and Technology in Trnava, Slovak University of Technology in Bratislava