Wavelets as a Tool for Data Mining Technology

Grygoriy Chetverykov
Department of Software Engineering
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
chetvertgg@gmail.com

Vyacheslav Lyashenko
Department of Informatics
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
lyashenko.vyacheslav@gmail.com

Oleg Kobylin
Department of Informatics
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
oleg.kobylin@nure.ua

Zhanna Deineko
Department of Media Systems and Technologies
Kharkiv National University of Radioelectronics
Kharkiv, Ukraine
zhanna.deineko@nure.ua

Abstract. Extraction of additional information from primary data is an important task for many areas of research. Additional information allows you to more accurately assess the situation and make more effective decisions. For more information, we have selected wavelets and corresponding wavelet analysis methods. We looked at the various applications of wavelets as a tool for data mining technology.

Ключові слова: wavelet analysis, data mining, phishing attack, human speech, image, analysis technology.

INTRODUCTION

Data mining technologies are important. During this analysis, the data are transformed into information that is studied.

Information plays an important role in the development of various spheres of human activity. This is due to the fact that the information provides knowledge about the object, phenomenon and process. Of particular importance information plays in the development of science, creating new technologies and applications. At the same time it is necessary to know not only the primary information, but also additional information. Additional information is information obtained in the course of primary data processing.

It is necessary that additional information can be obtained automatically. This is important for an adequate response to changes in primary information. Thus, we come to the concept of data mining. Examples of such an analysis are described in [1-4].

When considering data mining technologies, the tools for implementing such technologies are also of great importance. Among such tools one should single out wavelets and the general methodology of wavelet analysis.

APPLICATIONS OF WAVELETS IN DATA MINING

Wavelet is a mathematical function that allows you to implement a wavelet analysis of the original data. Such an analysis makes it possible to identify and harmonize the features of the source data over time. This makes it possible to obtain additional information that is very important in data mining.

Among the areas of use of wavelets in data mining, we highlight:
- wavelets in the study of phishing attacks;
- wavelets in the analysis of visual information;
- wavelets in the analysis of human speech.

Phishing attack creates a fake connection, including a fake URL. This makes it possible to replace the real Internet resource with a fake Internet resource. Thus, phishing attacks attract potential customers and cheat them. In fact, there is a distortion of real information that affects its subsequent analysis.

There are various methods of detecting phishing attacks [5]. But it is also important to investigate, evaluate the effectiveness of such methods. For this you can use wavelet analysis.

By comparing data on phishing attacks and data on the detection of these phishing attacks, we can evaluate the effectiveness of methods for detecting phishing attacks. To do this, you can use a technique that is based on the calculation of the Hurst index. This technique allows us to estimate the dynamics of the source data and to draw a conclusion on the trends of such dynamics.

We can also use wavelet coherence. This method provides additional information. The basis of the method is the mutual comparison of various data series (analysis of cross-references), which characterize the dynamics of phishing attacks. This allows you to calculate the dynamics of the local correlation of the data series that are investigated.

Thus, the analysis of phishing attacks is an important point in the implementation of data mining technology.

The use of wavelets in image processing is based on the analysis of differences in brightness levels of the original image (we find discontinuity points). Identifying and analyzing discontinuity points makes it possible to identify areas of interest in the image that is being researched. This analysis technology is more efficient in comparison with some classical methods of image analysis [2, 4].

At the same time, we can choose different wavelets for analyzing source images. We can also make adjustment of wavelet parameters. For this, it is necessary to take into account the properties of the areas of interest that are being studied. Such a choice of wavelets and their adjustment are based on the intellectual description of areas of interest. This allows for automatic analysis of visual information.

Automatic analysis of visual information is particularly important for processing biomedical images. In the end, we can make more effective decisions, which helps to make the diagnosis of diseases in the early stages of their development.

Intelligent analysis in infocommunications systems is also important. One of the main tasks of such systems is to increase the privacy and security of data. One solution to this problem is to identify a person using speech analysis. Identification is at the heart of modern security systems.

There are different methods for processing speech signals. Their disadvantage is the vulnerability of spectral attributes to changes in the emotional state of the speaker. Therefore, it is necessary to obtain additional information.

The method of wavelet coherence can be used to obtain additional information for analyzing human speech. The basis of this comparison is the analysis of cross-references of various time series. Each time series is a specific spectral attribute of the emotional state of the speaker. Thus, the method of wavelet coherence allows you to compare different emotional state of a person and adjust the method of identification.

CONCLUSIONS

We noted the importance of data mining. This is required for more information. The features of conducting data mining based on wavelets for various application areas are shown.

REFERENCES

[1] G. Chetverikov, O. Puzik, and I. Vechirska, "Multiple-valued structures of intellectual systems," XIth International Scientific and Technical Conference Computer Sciences and Information Technologies (CSIT), Lviv, 2016, pp. 204-207.

[2] S.M.H. Mousavi, and V. Lyashenko, "Extracting old persian cuneiform font out of noisy images (handwritten or inscription)," 10th Iranian Conference on Machine Vision and Image Processing (MVIP), Isfahan, 2017, pp. 241-246.

[3] N. Shakhovska, "Consolidated processing for differential information products," Perspective Technologies and Methods in MEMS Design, 2011, pp. 176-177.

[4] A. Rabotiahov, O. Kobylin, Z. Dudar, and V. Lyashenko, "Bionic image segmentation of cytology samples method," 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, 2018, pp. 665-670.

[5] M. Dadkhah, V.V. Lyashenko, Z.V. Deineko, S. Shamshirband, and M.D. Jazi, "Methodology of wavelet analysis in research of dynamics of phishing attacks," International Journal of Advanced Intelligence Paradigms, vol. 12(3-4), pp. 220-238, 2019.