Technology of Gene Expression Profiles Filtering Based on Wavelet Analysis

Sergii Babichev, Jiří Škvor and Jiří Fišer
Jan Evangelista Purkyně University in Ústí nad Labem, Ústí nad Labem, Czech Republic
E-mail: sergii.babichev@ujep.cz

Volodymyr Lytvynenko
Kherson National Technical University, 73008 Kherson, Ukraine
E-mail: immun56@gmail.com

Abstract—The paper presents the technology of gene expression profiles filtering based on the wavelet analysis methods. A structural block-chart of the wavelet-filtering process, which involves concurrent calculation of Shannon entropy for both the filtered data and allocated noise component is proposed. Simulation of the wavelet-filtering process was performed with the use of orthogonal and biorthogonal wavelets on different levels of wavelet decomposition and with the use of various values of the thresholding coefficient. Result of the simulation has allowed us to propose the technology to determine the optimal parameters of the wavelet filter based on complex analysis of the filtered data and allocated noise component.

Index Terms—Gene expression profiles, Filtering, Wavelet analysis, Shannon entropy, Thresholding.

I. INTRODUCTION

Biological systems of living organisms are a complex dynamic network of interacting elements of various purposes, state of which could be changed by influence of external conditions [1]. Study of such organisms' behavior is possible based on a system approach, which involves complex use of molecular biology, mathematics, physics and informatics. This approach creates the conditions for understanding the factors that define the character of biological systems operation in order to determine the ways of this process controlling [2,3]. The most important process in functioning of any living organism is regulation of genes expression, which defines the biosynthesis of a protein depending on needs of the organism. In particular, this process defines the growth and formation of a living organism and its constituent parts. Incorrect regulation of gene expression often causes illness or improper formation of appropriate biological object.

Reconstruction and simulation of gene regulatory network (GRN) create the basis for studying and analysis of character of genes interactions and for understanding how these interactions influence to functional abilities of studied cells. Usually, GPN is formed as a directed or non-directed graph contains both the nodes, which could be genes, proteins or metabolites, and arcs, which connect the corresponding nodes and define intermolecular interactions or indirect effects between the nodes. Determining structure and nature of gene regulatory network operation is associated with large experimental and theoretical complicities. Parameters of network are usually not obvious, since it is impossible to define inputs and outputs of molecular processes uniquely. Moreover, principles which are the basis of intermolecular interactions are very complex or unknown. Restore the regulatory interactions between system's elements based on experimental data, which contains gene expression profiles of the investigated objects, is the main idea of GRN reconstruction [4–6]. The method based on evaluation of mRNA molecules concentration is the most accessible one to evaluate gene expression. There are several methods to evaluate genes expression by mRNA molecules concentration determining. The first of them is cDNA molecules sequencing synthesized with the use of mRNA molecules, which is allocated from a specific part of investigated object [7]. The main advantage of the RNA sequencing method is low level of noise component that contributes to higher accuracy of appropriate measurements. Its main disadvantages are complexity of studied simples preparing and expensiveness. Another method of gene expression evaluation is based on DNA microarray experiments [8,9]. The main advantage of this method is the fact that its using allows us to estimate expression of tens of thousands of genes concurrently and to carry out their comparative analysis on different stages of the studied objects development. Moreover, the cost of DNA microarray experiments implementation is significantly lower to compare with the use of other methods. The main disadvantage of this method is higher level of noise component, which appears at the stages of DNA microarray creation and during reading information from it. Thus, in the case of DNA microarray experiments there is a necessity of gene expression profiles filtration at early stage of gene regulatory network reconstruction in order to remove the noise component. In [10–12],

authors presented the results of research concerning development of the objective clustering inductive technology for grouping gene expression profiles based on SOTA and DBSCAN clustering algorithms. In papers [13–16], authors implemented the fuzzy clustering methods for the studied objects grouping. However, it should be noted that the authors solved only the problem of gene expression profiles grouping by the use of appropriate clustering quality criteria. The problem of gene expression profiles filtration in these works was not considered. In the presented paper this problem is solved based on the use of wavelet analysis [17,18], which are used to process complex data in various areas of scientific research nowadays [19,20].

The remainder of this paper is organized as follows. Section 2 describes a model of gene expression profiles wavelet-filtering. Section 3 is about simulation of gene expression profiles wavelet-filtering. Section 4 presents a technology of the wavelet filter optimal parameters' determination. Conclusions are given in the final section.

II. Model of Gene Expression Profiles Wavelet-filtering

Fig. 1 shows the structural flowchart of discrete wavelet transformation process.
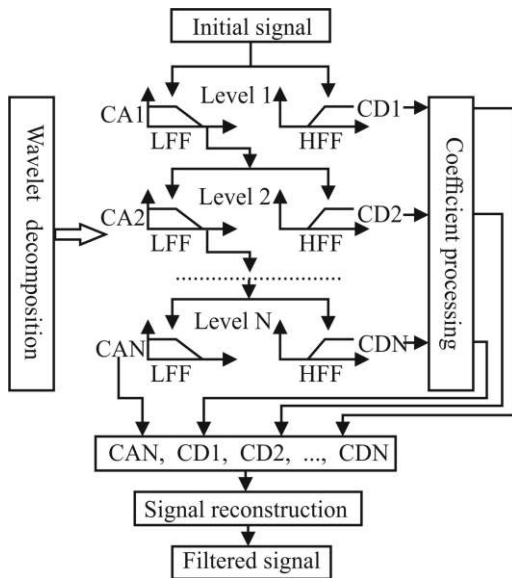

Fig.1. Structural flowchart of wavelet transformation process

Vector of approximation coefficients at N level of wavelet decomposition and vectors of detail coefficients at levels from 1 to N are calculated by the use of low (LFF) and high (HFF) frequency filters respectively during wavelet decomposition process of one-dimensional data:

$$s = CA_0 \rightarrow \{CA_1, CD_1\} \rightarrow ... \rightarrow \{CA_N, CD_N, CD_{N-1}, ..., CD_1\}$$

Approximation coefficients contain information about coarse component of the studied data. Information about

details of the data and about high-frequency noise component are contained in most cases in detail coefficients. Noise component removing from gene expression profiles within the framework of the proposed technology was performed with the use of soft thresholding as follows:

$$\begin{cases} d = 0, \ if \ d \le \tau, \\ d = d - \tau, \ if \ d > \tau \end{cases} \quad (1)$$

where τ is the thresholding coefficient value, d stands for the detail coefficients at all levels of wavelet decomposition of studied data. Data reconstruction is carried out with the use of both the approximation coefficients at N wavelet decomposition level and processed detail coefficients at levels from 1 to N. Process of determine of the wavelet filter optimal parameters involves the following steps:

– selection of the mother wavelet;
– determination of the wavelet decomposition optimal level for investigated data;
– choice of the wavelet type from the family of the appropriate mother wavelet;
– determination of the thresholding coefficient optimal value.

Estimation of data processing quality at each step of the wavelet-filtering process was performed with the use of Shannon entropy criterion, which was calculated based on James-Stein shrinkage estimator [21]. This method is based on complex use of two different models: a high-dimensional model with low bias and high variance, and a lower dimensional model with larger bias but smaller variance. The probability of values distribution in a cell in accordance with the James-Stein shrinkage method is calculated as follows:

$$p_i^{Shrink} = \lambda p_i + (1 - \lambda) p_i^{ML} \quad (2)$$

where p_i^{ML} is the probability of data values distribution in i-th cell, which is calculated by the maximum likelihood method, $p_i = 1/n_i$ is the estimation of probability in i-th cell, n_i is the quantity of features in this cell. Obviously, that p_i^{ML} corresponds to the high-dimensional model with low bias and high variance and p_i is the estimation with higher bias and lower variance of the features distribution. Intensity parameter λ in the proposed model is calculated as follows:

$$\lambda = \frac{1 - \sum_{i=1}^{k} (p_i^{ML})^2}{(n-1) \sum_{i=1}^{k} (p_i - p_i^{ML})^2} \quad (3)$$

where n is the features quantity in studied vector.

Shannon entropy value in this case is estimated with the use of standard formula taking into account the method of probability calculation in appropriate cell

$$H^{Shrink} = -\sum_{i=1}^{k} p_i^{Shrink} \log_2 p_i^{Shrink}. \qquad (4)$$

It is obvious, that in the case of gene expression profiles informativity evaluation, the minimum value of Shannon entropy criterion corresponds to higher quality of the investigated data processing. Otherwise, the maximum value of Shannon entropy calculated for the allocated noise component corresponds to higher approximation of this component to unordered "white" noise, which should be removed from investigated data. Structural flowchart of gene expression profiles wavelet-filtering process based on the James-Stein shrinkage method of Shannon entropy estimation is presented in Fig.2. Implementation of this process involves the following steps:

1. Choice of the mother wavelet from the list of the available wavelets.
2. Determination of the wavelet decomposition optimal level for the studied profiles based on the maximum value of Shannon entropy, which is calculated for the allocated noise component. At this step the choice of the wavelet type from the family of the mother one and the value of the thresholding coefficient are setup randomly from the range of the available values.
3. Determination of wavelet type from the family of the appropriate mother wavelet based on the maximum value of Shannon entropy, which is calculated for the allocated noise component.
4. Determination of the thresholding coefficient optimal value based on the minimum value of Shannon entropy, which is calculated for the filtered data.
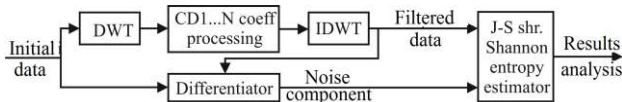

Fig.2. Structural flowchart of wavelet-filtering process

Thus, the process of determine of the wavelet filter optimal parameters involves concurrent evaluation of Shannon entropy for both the filtered data and allocated noise component.

III. SIMULATION OF GENE EXPRESSION PROFILES WAVELET-FILTERING

Simulation of the wavelet-filtering process was performed with the use of gene expression profile of lung cancer patient, which has been obtained by DNA-microchip experiments [22]. The expressions of studied genes have been determined by the technology which is presented in [23]. Statistical characteristics of the studied genes expression are presented in Table 1.

Table 1. Statistical characteristics of the studied genes expression

Min	25%	Median	Mean	75%	Max
-36.2	2.18	11.1	236.91	22.63	17360

As it can be seen, the expression values are varied from -36.2 to 17360. Moreover, these values for the most of genes are low. Noise component in the most cases is contained in high-frequency part of spectrum that justifies the use of wavelet analysis to remove the noise from data. Orthogonal Daubechies wavelets (*db1,...,db45*), symplets (*sym2,...,sym30*), coiflets (*coif1,...,coif5*), biorthogonal wavelets (*bior1.1,...,bior6.8*) and reverse biorthogonal wavelets (*rbio1.1,...,rbio6.8*) were used during the simulation process. Experimental determination of the thresholding coefficient value was performed in the two ways. In the first case, the detail coefficients step-by-step processing was carried out in accordance with formula (1), while the value of the thresholding coefficient was quite small (0.2) and it did not change during the simulation process. Duration of the experiment was limited by the steps number of the detail coefficients processing. The second case involved the thresholding coefficient value stepwise increasing from τ_{min} to τ_{max} by step $d\tau$. The simulation results in the case of Daubechies wavelets using are presented in Fig.3. Choice of the wavelet type and determination of the wavelet decomposition level was performed based on the maximum value of Shannon entropy for the allocated noise component. Determination of the thresholding coefficient value was carried out based on the minimum value of Shannon entropy for the filtered data.
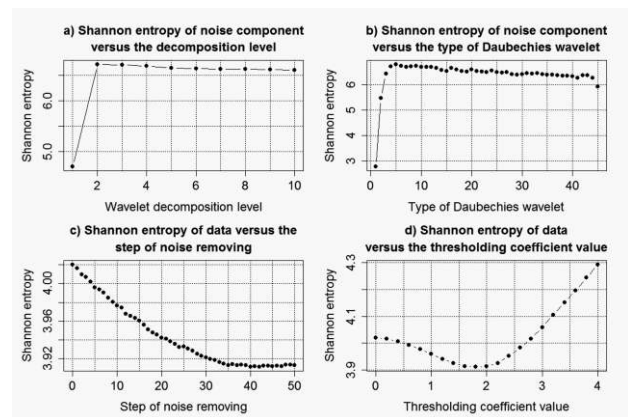

Fig.3. Simulation results in the case of using Daubechies wavelets

Analysis of the obtained results allows us to conclude that the maximum value of Shannon entropy for the allocated noise component is achieved in the case of the use of wavelet *db5* (Fig. 3b) at the second level of wavelet decomposition (Fig. 3a). The method of noise component step-by-step removing with the constant value of the thresholding coefficient (Fig. 3c) is not efficient since it does not allow us to determine uniquely the step of filtration process stopping. The method of the

thresholding coefficient stepwise increasing is an effective one to determine the optimal thresholding coefficient value. In this case the value of Shannon entropy for the filtered data has global minimum, which corresponds to the thresholding coefficient value $\tau = 1,8$ (Fig. 3d). The same results in the cases of coiflets, symplets, biorthogonal wavelets and reverse biorthogonal wavelets using are presented in Figs. 4-7.
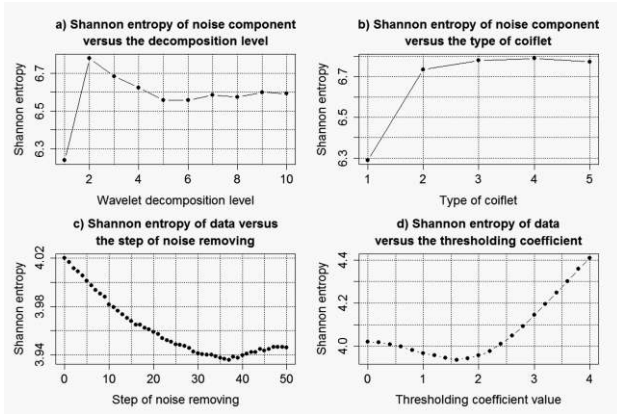

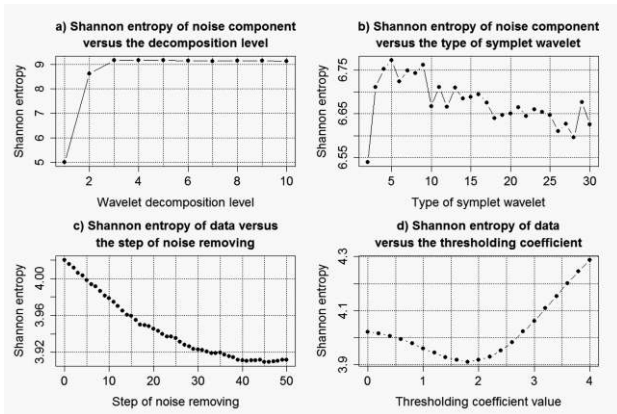Fig.4. Simulation results in the case of using coiflets

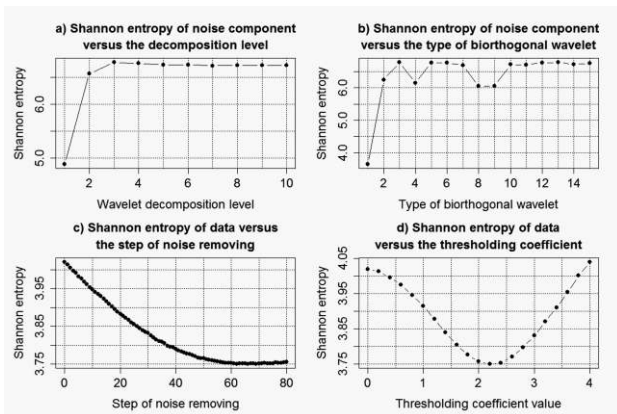Fig.5. Simulation results in the case of using symplets

Fig.6. Simulation results in the case of using biorthogonal wavelets

Analysis of the obtained charts shows that in terms of Shannon entropy criterion the following parameters of the wavelet filter are optimal: *db5* Daubechies wavelet at the second level of wavelet decomposition and thresholding coefficient 1.8; *coif4* coiflet wavelet at the second level of

wavelet decomposition and thresholding coefficient 1.6; *sym5* symplet wavelet at the third level of wavelet decomposition and thresholding coefficient 1.8; *bior1.5* biorthogonal wavelet at the third level of wavelet decomposition and thresholding coefficient 2.2 and *rbio1.5* reverse biorthogonal wavelet at the fourth level of wavelet decomposition and thresholding coefficient 1.8.

Fig. 8 shows the chart of the Shannon entropies ratio, which were calculated for the filtered data and the allocated noise component versus the mother wavelet type.
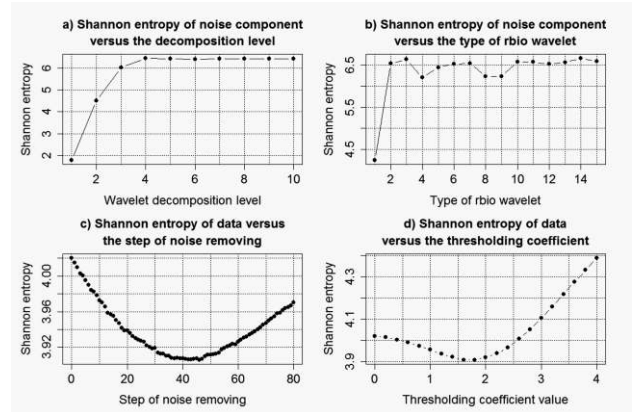

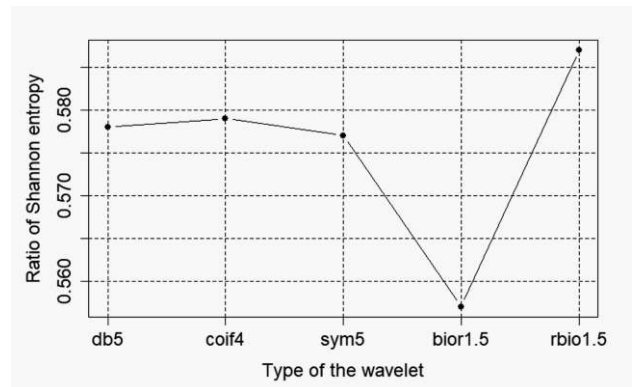Fig.7. Simulation results in the case of using reverse biorthogonal wavelets

Fig.8. Shannon entropies ratio versus the mother wavelet type

Analysis of the obtained results allows concluding that the choice of the mother wavelet type from the family of orthogonal and biorthogonal wavelets in the case of gene expression profiles filtering is not determinative. In terms of Shannon entropies ratio for both the filtered data and allocated noise component the best results concerning wavelet-filtering of gene expression profiles are obtained using *bior1.5* biorthogonal wavelet. However, the differences between results, which were obtained with the use of other mother wavelets are quite small. Determinant factors in this case are the choice of the wavelet type from the family of the appropriate mother wavelet, the choice of the wavelet decomposition level and determination of the thresholding coefficient optimal value for the detail coefficients processing.

IV. Technology of the Wavelet Filter Optimal Parameters Determination

The conducted research has allowed us to propose the technology to determine the wavelet filter optimal parameters in the form of a structural flowchart of studied data step-by-step processing. The architecture of this technology is presented in Fig. 9. Implementation of this technology involves the following stages:

Stage I. Initialization of the wavelet filter initial parameters.

1. Formation of vectors of the mother wavelets and the wavelets for selected mother ones:

$$wv = \{wv_i\} \;,\; i = 1,...,k$$
$$wv_i = \{wv_i^j\} \;,\; j = 1,...,p$$

where k is the mother wavelets quantity, p is the quantity of the wavelet types from the family of appropriate mother wavelet. Determination of the interval and the step of thresholding coefficient changes: τ_{min}, τ_{max}, $d\tau = \tau_{min}$, choice of the wavelet decomposition maximum level.
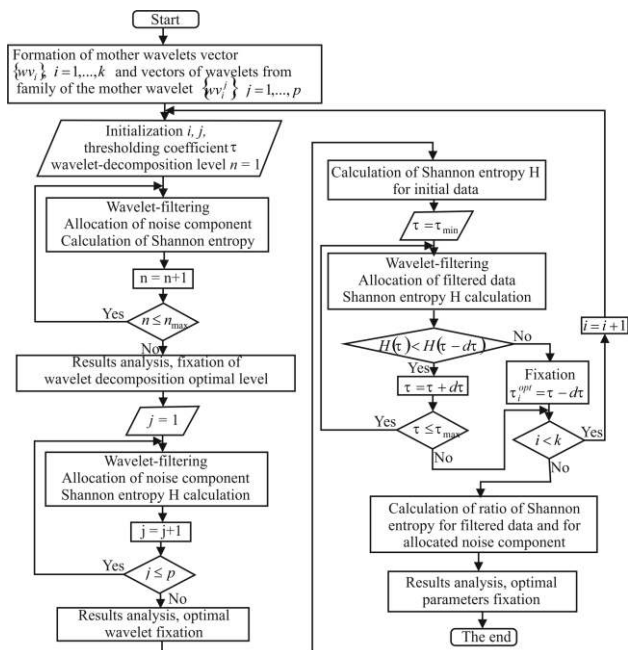

Fig.9. Architecture of the technology to determine the optimal parameters of wavelet filter

2. Choice of the mother wavelet corresponding to the first number of the appropriate vector of the mother wavelets ($i = 1$), random choice of the wavelet type from the family of the mother wavelet, setup of the thresholding coefficient value $\tau = \tau_{min}$ and the wavelet decomposition level $n = 1$.

Stage II. Determination of the wavelet decomposition optimal level.

3. Wavelet-filtering of the gene expression profiles within the established interval of the wavelet

decomposition level changes according with the scheme, which is shown in Fig. 3. Allocation of the noise component at each level of wavelet decomposition and calculation of Shannon entropy for the noise component at each step of this process.

4. Analysis of the obtained results. Fixation of the wavelet decomposition optimal level for the given mother wavelet n_i^{opt} corresponding to the maximum value of the Shannon entropy criterion.

Stage III. Determination of the wavelet type from the family of the appropriate mother one.

5. Wavelet-filtering of the gene expression profile for all types of wavelets from the family of the given mother wavelet $ww_i = \{wv_{ji}\}$, $j = 1,...,p$. Allocation of the noise component for each type of wavelet and calculation of Shannon entropy for the allocated noise component at each step of the data processing.

6. Analysis of the obtained results. Fixation of the mother wavelet optimal type corresponding to the maximum value of the Shannon entropy criterion.

Stage IV. Determination of the thresholding coefficient optimal value for the detail coefficients processing.

7. Calculation of Shannon entropy for the initial data: $H(0)$.

8. Wavelet-filtering of gene expression profile using the thresholding coefficient τ. Allocation of the filtered data.

9. Calculation of Shannon entropy at this step of the data processing: $H(\tau)$.

10. If $H(\tau) < H(\tau-d\tau)$, increase of the thresholding coefficient value ($\tau = \tau+d\tau$) and go to the step 8 of this procedure. Otherwise, fixation of the thresholding coefficient optimal value for i-th mother wavelet: $\tau_i^{opt} = \tau - d\tau$.

11. If $i \le k$, increment of i parameter ($i=i+1$) and repairing stages II–V of this procedure.

Stage V. Formation of the final solution concerning determination of the wavelet filter optimal parameters.

12. Calculation of the Shannon entropies ratio for the filtered data and the allocated noise component for each type of the mother wavelet using the wavelet filter optimal parameters.

13. Fixation of the wavelet filter optimal parameters corresponding to the global minimum of the Shannon entropies ratio criterion.

V. Conclusion

The results of research concerning creation of the technology of gene expression profiles wavelet-filtering in order to remove background noise is presented. The structural flowchart of wavelet-filtering process, implementation of which involves concurrent calculation of Shannon entropy for the filtered data and the allocated noise component has been proposed. Shannon entropy has been calculated with the use of the James-Stein

shrinkage estimator method, which takes into account two different models: a high-dimensional model with low bias and high variance, and a lower dimensional model with larger bias but smaller variance. Simulation of the wavelet filtration process was performed with the use of a gene expression profile of lung cancer patient, which has been obtained by DNA-microchip experiments. The family of orthogonal and biorthogonal wavelets at different levels of wavelet decomposition and with the use of various values of thresholding coefficient have been used during simulation process.

The charts of Shannon entropy versus the wavelet decomposition level, the type of wavelets from the family of the appropriate mother wavelet and the thresholding coefficient value have been obtained for each mother wavelet as the results of the simulation process. Analysis of the obtained charts has shown that in terms of Shannon entropy criterion the following parameters of the wavelet filter are optimal: *db5* Daubechies wavelet at the second level of wavelet decomposition and thresholding coefficient 1.8; *coif4* coiflet wavelet at the second level of wavelet decomposition and thresholding coefficient 1.6; *sym5* symplet wavelet at the third level of wavelet decomposition and thresholding coefficient 1.8; *bior1.5* biorthogonal wavelet at the third level of wavelet decomposition and thresholding coefficient 2.2 and *rbio1.5* reverse biorthogonal wavelet at the fourth level of wavelet decomposition and thresholding coefficient 1.8. The final decision concerning determination of the wavelet filter optimal parameters was performed based on the minimum value of the Shannon entropies ratio for the filtered data and the allocated noise component.

Analysis of the obtained results allows concluding that choice of the mother wavelet type from the family of orthogonal and biorthogonal wavelets in the case of gene expression profiles filtering is not determinative. In terms of Shannon entropies ratio for the filtered data and the allocated noise component the best results concerning wavelet-filtering of gene expression profiles have been obtained using *bior1.5* biorthogonal wavelet. However, the differences between the results obtained with the use of other mother wavelets are quite small. Determinant factors in this case are: choice of the wavelet type from the family of the appropriate mother wavelet, choice of the wavelet decomposition level and definition of the thresholding coefficient optimal value for detail coefficients processing.

The technology to determine the optimal parameters of the wavelet filter based on complex analysis of the filtered data and the allocated noise component has been proposed as result of the simulation. This technology is presented as a structural flowchart of the studied data step-by-step processing. Perspective of following authors' research is the implementation of the proposed technology within the framework of the hybrid technology of gene expression profiles preprocessing at early stage of gene regulatory network reconstruction.

REFERENCES

[1] D. Zak, R. Vadigepalli, E. Gonye, F. Doyle, J. Schwaber and B. Ogunnaike. Unconventional systems analysis problem in molecular biology: a case study ingene regulatory network modeling. Computational and Chemical Engineering, vol. 29(3), pp. 547–563, 2005.

[2] H. Kitano. Foundations of Systems Biology. MIT Press, 2001.

[3] E. Klipp, R. Herwig, A. Kowald, C. Wierling and H. Lehrach. Systems Biology in Practice. Concepts, Implementation and Application. Wiley-Blackwell, 2006.

[4] H. Yaghoobi, S. Haghipour, H. Hamzeiy and M. Asadi-Khiavi. A review of modeling techniques for genetic regulatory networks. Journal of Medical Signals and Sensors, vol.2(1), pp. 61–70, 2012.

[5] F. Emmert-Streib, M. Dehmer and B. Haibe-Kains. Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. Frontiers in Cell and Developmental Biology, vol. 2, pp. 1–7, 2014.

[6] J. Dussaut, C. Gallo, F. Cravero, M.J. Martnez and J.P. Carballido. Gernet: a gene regulatory network tool. BioSystems, vol. 162, pp. 1–11, 2017.

[7] J. Heather and B. Chain. The sequence of sequencers: The history of sequencing DNA. BioSystems, vol. 107(1), pp. 1–8, 2016.

[8] Daz-Badillo, M. Muoz, at al. A dna microarray-based assay to detect dual infection with two dengue virus serotypes. Sensors, vol. 14(5), pp. 7580–7601, 2014.

[9] K. Nazari, A. Karami, N. Amiri and F. Pourali. Microarray data analysis for detection and classification of viral infection. Journal of Applied Biotechnology Reports, vol. 1(1), pp. 22–27, 2014.

[10] S. Babichev, V. Lytvynenko, M. Korobchynskyi and M.A. Taif. Objective clustering inductive technology of gene expression sequences features. Communications in Computer and Information Science, vol. 716, pp. 359–372, 2016.

[11] S. Babichev, V. Lytvynenko, J. Škvor and J. Fišer. Model of the Objective Clustering Inductive Technology of Gene Expression Profiles Based on SOTA and DBSCAN Clustering Algorithms, vol. 689, pp. 21–39, 2017.

[12] S. Babichev, M.A. Taif, V. Lytvynenko and V. Osypenko. Criterial analysis of gene expression sequences to create the objective clustering inductive technology. IEEE 37th International Conference on Electronics and Nanotechnology, ELNANO 2017, pp. 244–248, 2017.

[13] Zh. Hu, Ye.V. Bodyanskiy, O.K. Tyshchenko, and V.O. Samitova,"Fuzzy Clustering Data Given in the Ordinal Scale", International Journal of Intelligent Systems and Applications (IJISA), Vol.9, No.1, pp.67-74, 2017.

[14] Zh. Hu, Ye.V. Bodyanskiy, O.K. Tyshchenko, and V.O. Samitova,"Fuzzy Clustering Data Given on the Ordinal Scale Based on Membership and Likelihood Functions Sharing", International Journal of Intelligent Systems and Applications (IJISA), Vol.9, No.2, pp.1-9, 2017.

[15] Zh. Hu, Ye.V. Bodyanskiy, O.K. Tyshchenko, V.O. Samitova,"Possibilistic Fuzzy Clustering for Categorical Data Arrays Based on Frequency Prototypes and Dissimilarity Measures", International Journal of Intelligent Systems and Applications (IJISA), Vol.9, No.5, pp.55-61, 2017.

[16] Zh. Hu, Ye.V. Bodyanskiy, O.K. Tyshchenko, V.M. Tkachov, "Fuzzy Clustering Data Arrays with Omitted Observations", International Journal of Intelligent Systems and Applications (IJISA), Vol.9, No.6, pp.24-32, 2017.

[17] Daubechies. The wavelet transform, time-frequency localization and signal analysis. IEEE Transactions on Information Theory, vol. 36(5), pp. 961–1005, 1990.

[18] R. Coifman, Y. Meyer and M. Wickerhauser. Wavelet analysis and signal processing. Wavelets and their applications, pp. 153–178, 1992.

[19] Y. Bodyanskiy, A. Dolotov and O. Vynokurova. Evolving spiking wavelet-neurofuzzy self-learning system. Applied Soft Computing Journal, vol. 14, pp. 252–258, 2014.

[20] Y. Bodyanskiy, O. Vynokurova, I. Pliss, D. Peleshko and. Y. Rashkevych. Hybrid generalized additive wavelet-neuro-fuzzy-system and its adaptive learning. Advances in Intelligent Systems and Computing, vol. 470, pp. 51–61, 2016.

[21] Hausser and K. Strimmer. Entropy inference and the james-stein estimator with application to nonlinear gene association networks. Journal of Machine Learning Research, vol. 10, pp.1469–1484, 2009.

[22] D.G. Beer, S.L. Kardia, at. al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nature Medicine, vol. 8(8), pp. 816–824, 2002.

[23] S.A. Babichev, A.I. Kornelyuk, V.I. Lytvynenko and V. Osypenko. Computational analysis of microarray gene expression profiles of lung cancer. Biopolymers and Cell, vol. 32(1), pp. 70–79, 2016.

Jiří Škvor is an assistant professor in the Department of Informatics at Faculty of Science, Jan Evangelista Purkyně University in Ústí nad Labem, Czech Republic. He obtained his PhD in computer methods in science and technology at the same faculty in 2009. His main research interests are in molecular simulations, computational geometry and percolation theory.

Jiri Fiser graduated (Ph.D.) from University of Economics, Prague in 2003. He is currently working as Assistant professor of Department of Informatics at Jan Evangelista Purkyně University in Ustí nad Labem, Czech Republic and Czech Technical University in Prague. He has been publishing in the area of dependability, event-based simulations, functional aspects of programming languages and data mining of complex data. He actively uses and teaches modern open-source programming languages as Python and Julia.

Authors' Profiles

Sergii Babichev graduated (M.Sc.) from Kherson State Pedagogical Institute in 1984. He got his PhD in 2003.

He is currently working as Assistant professor of Department of Informatics at Jan Evangelista Purkyně University in Ustí nad Labem, Czech Republic. He has about 100 scientific publications. His research interests are data mining of complex data, bioinformatics, gene expression profiles processing, gene regulatory network reconstruction.

Volodymyr Lytvynenko - candidate of Technical Sciences (PhD) since 1997. Since 2010 - Doctor of Sciences. From 2013, he is a Professor and Head of the Department of Informatics and Computer Science at Kherson National Technical University in Ukraine. He has about 200 scientific publications. His research interests are artificial immune systems data mining of complex data, bioinformatics, structural and computational biology, modeling of biological and ecological systems, Bayesian networks, soft computing, gene regulatory network reconstruction. gene expression profiles processing.