Paper

Time series denoising with wavelet transform

Bartosz Kozłowski

Abstract—This paper concerns the possibilities of applying wavelet analysis to discovering and reducing distortions occurring in time series. Wavelet analysis basics are briefly reviewed. WaveShrink method including three most common shrinking variants (hard, soft, and non-negative garrote shrinkage functions) is described. Another wavelet-based filtering method, with parameters depending on the length of wavelets, is introduced. Sample results of filtering follow the descriptions of both methods. Additionally the results of the use of both filtering methods are compared. Examples in this paper deal only with the simplest "mother" wavelet function – Haar basic wavelet function.

Keywords— wavelet transform, WaveShrink, filtration, noise reduction, Haar basic wavelet function.

1. Introduction

Foundations of wavelet-based analysis methods were laid in the beginning of the 20th century. Back then, in the year 1909 Hungarian mathematician Alfred Haar introduced his two-state function in appendix to his doctoral thesis published later on [3]. Today a slightly modified version of this function is regarded as the first basic wavelet function. In the nineties of the 20th century a very swift developement of wavelet enforced methods began. Wavelets turned out to be very useful when applied to many problems, including analysis and synthesis of time series [9] (in acoustics, geology, filtration and forecasting [5, 11] in meteorology and economics), effective data storage, especially images [10] (computer graphics, image animation in movie industry). Lately a very fast development of wavelet-based data mining [6] techniques may be observed.

One of the tasks of knowledge discovery preprocessing is noise reduction. The goal of this sub process is to separate the noise from the signal and then to reduce or remove the noise. The definition of a noise stays imprecise because for some physical processes it is difficult to clearly define it. In this paper it is assumed that the goal of filtering is to change the input signal in such a manner that the values of a time series, which differ a lot from others, are changed but the characteristic of signal stays the way it was. Most commonly used methods consider statistical approach or involve Fourier transform. In this paper it will be shown why wavelet transform may be considered valuable for this task.

The paper is organized as follows. Section 2 presents absolute basics of wavelet analysis, including basic wavelet functions and discrete wavelet transform. Section 3 reviews WaveShrink methodology and presents results obtained by its application to the example data. In Section 4 the reason for searching for another method is clarified. A genuine method, also based on wavelets, is introduced. The results of application of the new methodology to the same example data are presented and compared with WaveShrink's results. Section 5 concludes the paper.

2. Wavelet theoretical background

Wavelets are functions which result in values different than zero in a relatively short interval. In this regard they differ from "normal", long waves, such as sinusoidal ones, which are determined on a whole time domain $(-\infty,\infty)$.

Let ψ be a real function of a real variable u which satisfies two conditions:

$$\int_{-\infty}^{\infty} \psi(u)du = 0 \tag{1}$$

and

$$\int_{-\infty}^{\infty} \psi^2(u)du = 1. \tag{2}$$

Condition (2) means that for any ε from an interval $(0,1)$ there is an interval $(-T,T)$ such that:

$$\int_{-T}^{T} \psi^2(u)du = 1-\varepsilon.$$

If ε is close to 0, it may be seen that only in an interval $(-T,T)$ corresponding to this ε values $\psi(u)$ are different than 0. Outside of this interval they must equal 0. Interval $(-T,T)$ is small compared to an interval $(-\infty,\infty)$, on which a whole function is determined. Condition (1) implies that if $\psi(u)$ has some positive values, it also has to have some negative ones (a function "waves"). Therefore Eqs. (1) and (2) introduce a concept of a small wave, or shorter, wavelet. If Haar function ϕ, which is a two-state function of real variable (Fig. 1a):

$$\phi(u) = \begin{cases} -1 & \text{for} \quad -1 < u \le 0 \\ 1 & \text{for} \quad 0 < u \le 1 \\ 0 & \text{for} \quad \text{other u} \end{cases}$$

would be transformed into:

$$\psi^{(H)}(u) = \begin{cases} -\frac{1}{\sqrt{2}} & \text{for} \quad -1 < u \le 0 \\ \frac{1}{\sqrt{2}} & \text{for} \quad 0 < u \le 1 \\ 0 & \text{for} \quad \text{other u} \end{cases}$$

then the resulting function $\psi^{(H)}$ satisfies conditions (1) and (2), and is called Haar basic wavelet function (Fig. 1b).

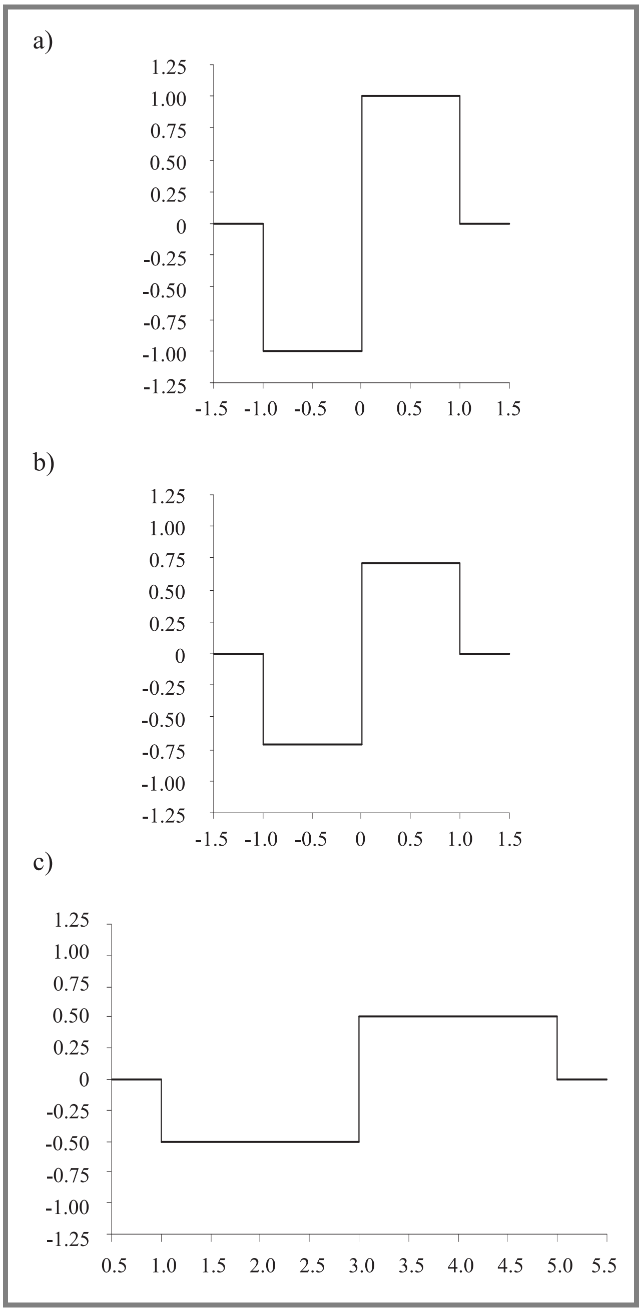

Fig. 1. Haar function (a), modified Haar function – the first basic wavelet function (b), and wavelet-based on Haar basic wavelet function with $\lambda = 2$ and $t = 3$ (c).

Introducing two parameters, namely scale (λ) and location (t) into the above definition of Haar basic wavelet function we get a family of scaled and transformed wavelets:

$$\psi^{(H)}_{\lambda,t}(u) = \begin{cases} -\frac{1}{\sqrt{2\lambda}} & \text{for} \quad t-\lambda < u \le t \\ \frac{1}{\sqrt{2\lambda}} & \text{for} \quad t < u \le t+\lambda \\ 0 & \text{for} \quad \text{other u} \end{cases}$$

as one on Fig. 1c, based on $\psi^{(H)}$. Figure 1b considers wavelets with scale λ equal to 1 ($\lambda = 1$) and is defined in point t equal to 0 ($t = 0$).

Wavelets $\psi^{(H)}_{\lambda,t}(u)$ may be established based on $\psi^{(H)}(u)$ according to the formula:

$$\psi^{(H)}_{\lambda,t}(u) \equiv \frac{1}{\sqrt{\lambda}}\psi^{(H)}(\frac{u-t}{\lambda}).$$

There are of course many other basic wavelet functions such as Mexican hat, Gauss wavelets, Morlet wavelet [8], family of Daubechies wavelets [1] just to name a few.
As a result of wavelet transform we obtain a set of wavelet coefficients calculated at different levels (scales) and in a wide range of observation points (locations). There are many ways of doing this. Two most often applied are (orthonormal) discrete wavelet transform (DWT) and its slightly modified version which preserves scales but calculates wavelets in more densely chosen observation points. For an DWT Mallat [7] proposed a very fast algorithm for calculating wavelets. With minor modifications it may be applied to other wavelet transforms. An important feature of DWT is that it may be reversed. Having wavelet coefficients it is possible to calculate original time series. This possibility is fundamental for a majority of current wavelet applications.

3. WaveShrink

3.1. WaveShrink method

One of the most explored signal smoothening or cutting method utilizing wavelets is Donoho's and Johnstone's [2] WaveShrink method.
The method is composed of three main steps. At the beginning observed time series is transformed into the wavelet space by DWT. In step two wavelet coefficients are modified, reduced according to the selected shrinkage function and a given threshold value. To accomplish this one of three shrinkage functions (presented below) may usually be used to establish how to modify wavelet time series coefficient. In the end inverse DWT is applied on wavelet coefficients and as a result smoothened original signal (with reduced noises) is derived.

3.2. Shrinkage functions

Shrinkage functions are formulas that define a correction coefficient $\delta_\lambda(x)$, which is subtracted from the corresponding wavelet coefficient. Calculated correction is relevant (different from 0) for those wavelet values, which exceed a given threshold parameter λ.

Hard shrinkage function

$$\delta^H_\lambda(x) = \begin{cases} 0 & |x| \le \lambda \\ x & |x| > \lambda \end{cases} . \tag{3}$$

Subtracting this correction reduces those wavelet coefficients of the wavelet time series, which exceed threshold value, to zero.

Soft shrinkage function

$$\delta_\lambda^S(x) = \begin{cases} 0 & |x| \le \lambda \\ x - \lambda & x > \lambda \\ \lambda - x & x < -\lambda \end{cases} . \tag{4}$$

By subtracting δ_λ^S correction considered wavelet time series coefficients are reduced to λ for positive coefficients and to $-\lambda$ for negative ones.

Non-negative garrote shrinkage function

$$\delta_\lambda^{NN}(x) = \begin{cases} 0 & |x| \le \lambda \\ x - \frac{\lambda^2}{x} & |x| > \lambda \end{cases} . \tag{5}$$

Results of δ_λ^{NN} function subtracted form wavelet time series coefficients modify them directing to zero in such a manner that the more a coefficient exceeds the threshold value the more it is reduced towards zero.

3.3. Estimating threshold value

The clue of WaveShrink method is to correctly estimate λ parameter. It may be achieved in various ways. Two of those are presented below.

Min-max approach

The first one is so called min-max threshold solution task. It is defined by Eq. (6).

$$\inf_\lambda \sup_\theta \left\{ \frac{R_\lambda(\theta)}{n^{-1} + \min(\theta^2, 1)} \right\} . \tag{6}$$

$R_\lambda(\theta)$ is defined as

$$R_\lambda(\theta) = E(\delta_\lambda(x) - \theta)^2 \tag{7}$$

and

$$x \sim N(\theta, 1).$$

What follows is it is assumed that the signal is of a normal distribution with expected value θ and a standard deviation equal to one. Solving this problem analytically requires, at least, altering signal so that it suits this assumed condition.

Standard deviation approach

In the following example another method was applied. Parameter λ is derived as a value of standard deviation's estimator (calculated for each wavelet level) multiplied by parameter n.

3.4. Illustrative example

Both this example and the example in next section are performed on data (shown in Fig. 2) collected from a network router. It is a traffic time series observed in one of the ports throughout the day.

In Fig. 3 there are the results of filtering using different values of a parameter n supplied for procedure of evaluating λ (one, two, and three) "within" the same shrinkage function (δ_λ^H – hard shrinkage function).

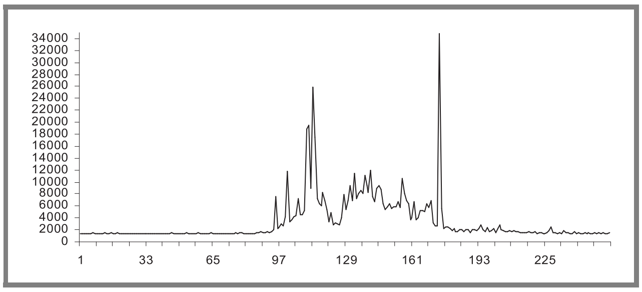

Fig. 2. Original time series.

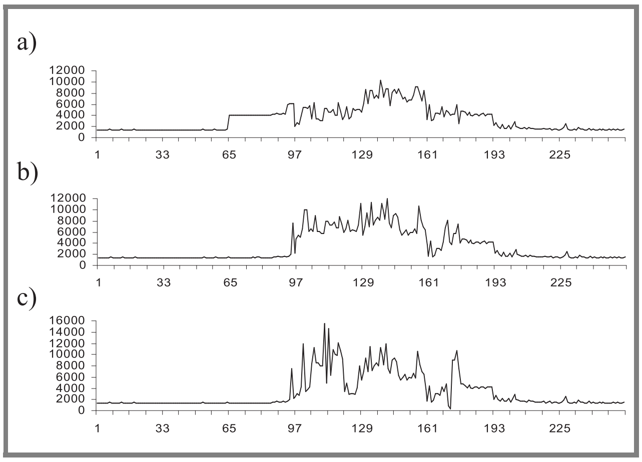

Fig. 3. The result of denoising by WaveShrink with $\delta_\lambda = \delta_\lambda^H$ and λ estimating parameter n equal to one (a), two (b), and three (c), respectively.

Figure 4 is organized in the same way but in this case $\delta_\lambda = \delta_\lambda^S$ – soft shrinkage function is used.

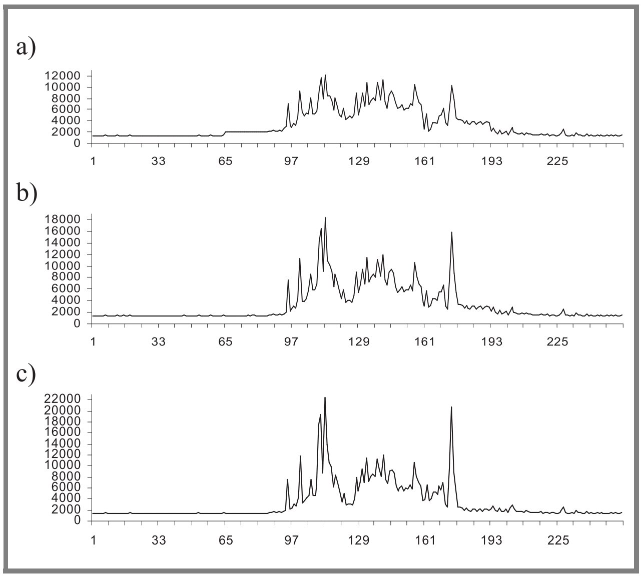

Fig. 4. The result of denoising by WaveShrink with $\delta_\lambda = \delta_\lambda^S$ and λ estimating parameter n equal to one (a), two (b), and three (c), respectively.

And finally Fig. 5 is obtained as a result of application of non-negative garrote shrinkage function δ_λ^{NN}.

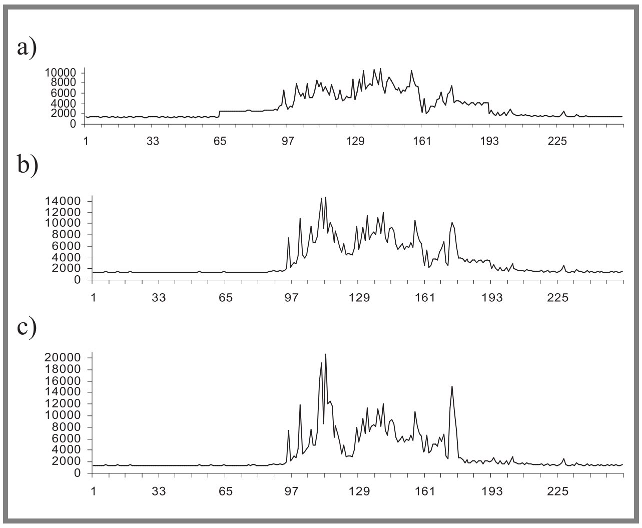

Fig. 5. The result of denoising by WaveShrink with $\delta_\lambda = \delta_\lambda^{NN}$ and λ estimating parameter n equal to one (a), two (b), and three (c), respectively.

One can easily see that for all shrinkage functions the "power of shrinkage" gets weaker as parameter n rises. On the other hand the fit of shrinkaged function to the original time series increases as parameter n rises.

In all cases this algorithm distorts the signal not only in noise spots but also in neighboring observation points. This is an obvious result of transforming altered wavelet coefficients into the original domain.

4. New method of denoising

4.1. Wavelet noise suspect approach

The final observation from the previous section is a direct reason for introducing another method of filtering. As wavelet domain offers great opportunities and we do not intend to drop the wavelet approach, it is necessary to design the methodology in a different way. The results of the wavelet transform are used as pointers of noise only. Wavelet coefficients are not altered and inverse transformation is not used at all. Wavelet coefficients may be considered as proportional to the difference between weighted (in case of Haar wavelet the weights are equal) sums of observations, in neighboring intervals of an equal length, and wavelet transform's multiscale provides multilevel analysis.

The new method also consists of three main steps. In the first one given time series is orthonormally converted into the wavelet domain. Then, for each wavelet level, wavelets qualified as noises are determined. Each wavelet coefficient, which is qualified as noise, is marked and considered as wavelet noise suspect (precisely a coefficient calculated of original time sub series, which includes suspected noise). Having all wavelet coefficients passed wavelet noise suspect marking procedure the second step may be processed.

Every original time series point which "is covered" by the wavelet noise suspect is marked as the point noise suspect on the level which corresponds to the level of this wavelet noise suspect. Then for each observation point of original time series the depth of distortion is found. This depth D is the biggest level to which point is continuously (from level 0) marked as point noise suspect. In step three only these observation points, which have D greater than given threshold value $D_{\min}$ and are also marked as point noise suspects in time series space should be considered as noise (of strength D). Finally, each of these considered observation points is reduced by replacing original time series values with results of a function $b(D, D_{\min}, f[, ...])$, where f is an original time series function.

This method leaves a lot of space for heuristics as it comes to designating methods of identifying distortion point and marking them, definition of function b, and method of determining D itself. One of the possible approaches is presented below.

4.2. Illustrative example

This example is performed on the same data set as the example in previous section. Function b is defined as

$$b(D, D_{\min}, f, x) = \begin{cases} f(x) & D \le D_{\min} \\ f(x)/D & D > D_{\min} \end{cases}, \tag{8}$$

and $D_{\min}$ equals 1. Results are presented in Table 1 and Fig. 6.

Table 1
Parts of table used in the process of introduced method of denoising

		0	1	2	3	4	5	...
...	...	...	...	...	...	...	...	...
94	0	0	1	0	0	0	1	...
95	1	1	1	0	0	0	1	...
96	0	1	1	0	0	0	1	...
...	...	...	...	...	...	...	...	...
101	1	1	1	0	1	0	1	...
102	0	1	1	0	1	0	1	...
103	0	0	1	0	1	0	1	...
104	0	0	1	0	1	0	1	...
105	0	0	0	1	1	0	1	...
106	1	0	0	1	1	0	1	...
...	...	...	...	...	...	...	...	...
114	1	1	1	1	1	0	1	...
115	1	1	1	1	1	0	1	...
...	...	...	...	...	...	...	...	...
129	1	1	1	0	1	0	1	...
...	...	...	...	...	...	...	...	...
174	0	1	1	1	1	1	0	...
175	1	1	1	1	1	1	0	...
176	0	0	1	1	1	1	0	...
...	...	...	...	...	...	...	...	...

Table's rows correspond to the results of original time series' point identifications and columns correspond to original time series point number, original time series' noise identification mark, and wavelet coefficients series' noise identification mark respectively (1 in cell means that point was identified as noise suspect at column's level).
Numbers of rows 95, 101, 114, 115, 129, and 175 are numbers of observation points from original time series which are qualified as noises. In these cases D equals j. Rows 96, 102, 174 are not qualified as noises, despite the fact that they are qualified as noise within wavelet domain, because they were not qualified this way in original time series domain. Similarly row 106 is not qualified as noise (despite that it is qualified this way in the original time series space and in wavelet levels 2, 3, and 5) because there were qualification gaps (at wavelet levels 0 and 1).

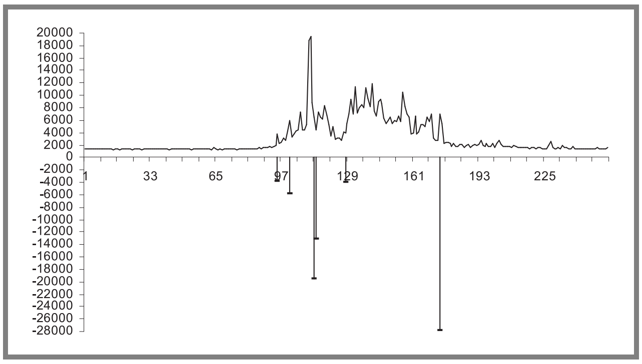

Fig. 6. Result of denoising by the new method.

Figure 6 shows the original time series transformed by filtration. Additionally it illustrates to what extent selected noise observation points were altered (the height of peek below horizontal axis equals δ). The first two peaks correspond to almost least reduced ($D = 2$) rows (observation points) 95 and 101 in Table 1. The last peak corresponds to most reduced ($D = 4$) observation point 175. In case of the two close, middle peeks (114 and 115) noise has propagated to the $D = 3$ wavelet level so observation points were reduced with power of something in between previous reductions. In case of point 129, D equals 1 and observation point 129 was the one, which was the least reduced.

5. Conclusions

Introduced method, together with WaveShrink, provide a complementary tandem of filtering based on wavelets. The main difference between introduced genuine method and WaveShrink method is that the reduction is performed on a signal, not on wavelet coefficients. These are only a tool for determining parameters of function reducing assumed noise. Therefore, the introduced method does not induct distortions into noise-neighboring observation points, what WaveShrink features.
WaveShrink suits better time series distorted in general, by malfunction for example, as it adjusts wide time series characteristics' intervals. The method introduced is better in cases where it is intended to remove "irregularities" from series as it does not change anything but the observation points qualified as noise.
Of course, this paper does not cover some other important issues. For example the order of complexity of calculations and reduction of their number was not discussed. Also the possibility of application of "more dense" wavelet transforms in case of the wavelet noise suspect method is not discussed here. Although this kind of transformations require a larger number of calculations, they seem to be an interesting direction of applications of wavelets to time series denoising.

References

[1] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia: SIAM, 1992.
[2] D. L. Donoho and I. M. Johnstone, "Attempting to unknown smoothness via wavelet shrinkage", *Ann. Stat.*, vol. 90, pp. 1200–1224, 1998.
[3] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme", *Math. Ann.*, vol. LXIX, pp. 331–371, 1910.
[4] B. Kozłowski, "Wavelet-based approach to time series denoising", in *Proc. Int. Conf. Decis. Supp. Telecommun. Inform. Soc. DSTIS*, Warsaw, Poland, 2004, vol. 4, pp. 175–193.
[5] B. Kozłowski, "On time series forecasting methods of linear complexity utilizing wavelets", in *Conf. Adv. Intell. Syst. – Theory Appl., Coop. IEEE Comput. Soc.*, Kirchberg, Luxembourg, 2004.
[6] T. Li, Q. Li, S. Zhu, and M. Ogihara, "Survey on wavelet applications in data mining", *SIGKDD Expl.*, vol. 4, no. 2, pp. 49–68, 2003.
[7] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic Press, 1998.
[8] J. Morlet and A. Grossman, "Decomposition of hardy functions into square integrable wavelets of constant shape", *SIAM J. Math. Anal.*, vol. 15, no. 4 pp. 723–736, 1984.
[9] D. B. Percival and A. T. Walden, *Wavelet Methods for Time Series Analysis*. Cambridge: Cambridge University Press, 2000.
[10] L. Prasad and S. S. Iyengar, *Wavelet Analysis with Applications to Image Processing*. Boca Raton: CRC Press, 1997.
[11] P. Yu, A. Goldberg, and Z. Bi, "Time series forecasting using wavelets with predictor-corrector boundary treatment", in *7th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, San Francisco, USA, 2001.

Bartosz Kozłowski received the M.Sc. degree from Warsaw University of Technology in 2004. He is currently working towards the Ph.D. degree. His interests include data mining, operational research, decision support, forecasting, and wavelet analysis. He has been involved in projects concerning computer networks, security, business applications development, and other Internet oriented technologies.
e-mail: b.kozlowski@elka.pw.edu.pl
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska st 15/19
00-665 Warsaw, Poland