Piotr BOJARCZAK, Piotr LESIAK
RADOM UNIVERSITY OF TECHNOLOGY, INSTITUTE OF AUTOMATICS AND TELEMATICS OF TRANSPORT

SVM based classification method of railway's defects

Ph.D. Piotr BOJARCZAK

He has been working at Transport Faculty of Radom University since 1990. He is interested particularly in neural networks and signal processing. He is the author of several papers concerning neural networks and their applications.

e-mail: p.bojar@pr.radom.pl

Ph.D. Piotr LESIAK

He has been working at Transport Faculty of Radom University since 1973. His research and didactic include development and investigation of diagnostic systems for railway transport and measuring systems. He is co-author of the book, author of several academic books and about 90 scientific papers. He is also a holder of 3 prizes awarded by Ministry of Scientific Research and Information Technology for his research.

e-mail: plesiak@pr.radom.net

Abstract

Railway's surface defects belong to some kind of railway's flaws not been detected by traditional ultrasonic method and therefore they pose a major thread to the safety of railway traffic. Paper's aim is to present the Method of Metal Magnetic Memory along with SVM network allowing for detection of surface defects. Signals coming from the device whose operation is based on this method are given to wavelet's packet block extracting the most important features characterizing surface defects, followed by SVM network operating as a classifier.

Keywords: Method of Metal Magnetic Memory, wavelet's transform, SVM networks.

Sieć SVM w klasyfikacji uszkodzeń szyn kolejowych

Streszczenie

W artykule przedstawiono próbę wykorzystania metody magnetycznej pamięci metalu wraz z klasyfikatorem opartym o sieć SVM (Support Vector Machines) do wykrywania wad powierzchniowych występujących w szynach kolejowych. Wady te są niewykrywalne przez tradycyjne metody oparte na ultradźwiękach a przez to stanowią poważne zagrożenie dla bezpieczeństwa ruchu pociągów.

Słowa kluczowe: Metoda Magnetycznej Pamięci Metalu, transformata falkowa, sieci SVM.

1. Introduction

Most methods using to detect railway's defects are based on ultrasonic. Ultrasonic method determines the position of defects in railway on the basis of the measurement of return duration of ultrasonic rays emitted by ultrasonic head. Surface defects occur in the upper layer of railway's head what makes the measurement of return time practically impossible. On the other hand, these defects left uncontrolled pose a large thread to the safety of railway traffic. There were several dangerous railway crashes including trains derailing caused by such defects. Cost of each crash reaches many billions of zloty, what makes the problem very crucial. Surface defects can be divided into several categories: head-checking, squat, and shelling. Unfortunately, presented solution allows for the detection only head-checking defects. It exploits Method of Metal Magnetic Memory along with wavelet's packets and SVM network. Because of it, it could be used as a complementary method to ultrasonic method.

2. Method of Metal Magnetic Memory

In order to recognize railway's surface defects, Method of Metal Magnetic Memory has been utilized. This method exploits the magneto-elastic effect [1, 2]. If the cyclic force is pressed on some point and there exists external magnetic field for example earth magnetic field, then residual magnetism of this point increases permanently. Material deformation arising during such load is estimated by the measurement of components of stray magnetic field H_p near the surface of examined object. Basing on [1], it can be argued, that there exists the following relationship between stray magnetic field H_p and the stress $\Delta\sigma$ occurring in the material:

$$H_p \approx \frac{1}{\mu_0}\lambda_{H,T}\Delta\sigma \tag{1}$$

where:

$$\lambda_{H,T} = \left(\frac{\Delta B}{\Delta\sigma}\right)_{H,T} \tag{2}$$

is the sensitivity of magneto-elastic phenomena at the constant magnetic field H and temperature T. Additionally μ_0 and Δ_B mean magnetic permeability of air and induction change on Δ_X interval respectively.

Stray magnetic field H_p depending of material deformation can be measured with the use of magnetometer of type TSC-1M-4. This device is fitted with four heads, each of them is able to independently measure two components of stray magnetic field – tangential and normal. The portable device records signals coming from measuring heads. These records can be further downloaded to personal computer through RS232 interface. Recorded signals are supposed to be classified by the intelligent signal-processing unit.

3. Overall Structure of Surface Defect's Detector

The task of surface defect's detector is to recognize the type of defect on the basis of the signals being recorded by magnetometer. Magnetometer has four measuring heads, each of them is responsible for scanning of corresponding slice of railway's head. Because magnetometer records eight signals (four measuring heads, each recording tangential and normal stray magnetic field's components), therefore defect's detector will consist of eight independent channels, each connected to the corresponding signal. Each channel is to recognize two groups of signals: the first corresponding to normal state of scanning slice and the second corresponding to head-checking defect. Fig. 1 shows the structure of surface defect detector along with channels, which consist of feature extracting block followed by classifier's block. Because classifying signals are transient and non-stationary, it is necessary to extract some unique feature set allowing for discrimination between two groups of signals. After extracting, features are given to classifier block, whose task is to recognize to which group the signal belongs. Wavelet's packet has been used as a feature extractor and SVM network as a classifier.

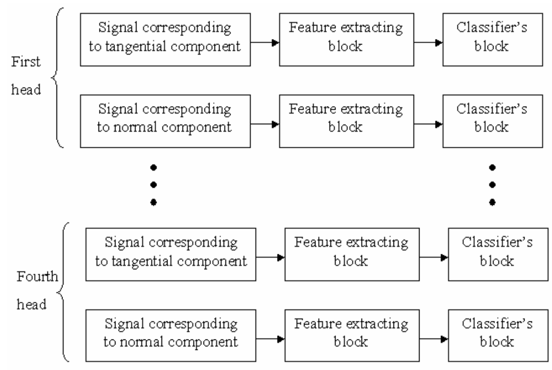

Fig. 1. Structure of surface defect's detector
Rys. 1. Struktura układu detektora wad powierzchniowych

4. Feature extractor

Distinction of signals belonging to two aforementioned groups (classes) is not easy task. Fig. 2 presents exemplary signals corresponding to normal state of railway and Fig. 3 presents exemplary signals corresponding to head-checking defect [3].

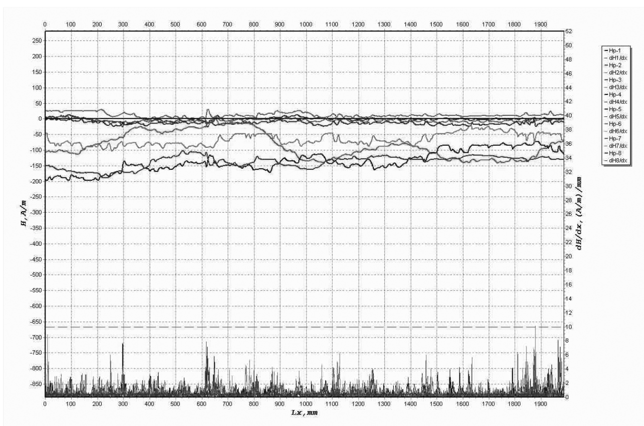

Fig. 2. Exemplary signals corresponding to normal state of railway
Rys. 2. Przykładowe sygnały odpowiadające normalnemu stanowi szyny

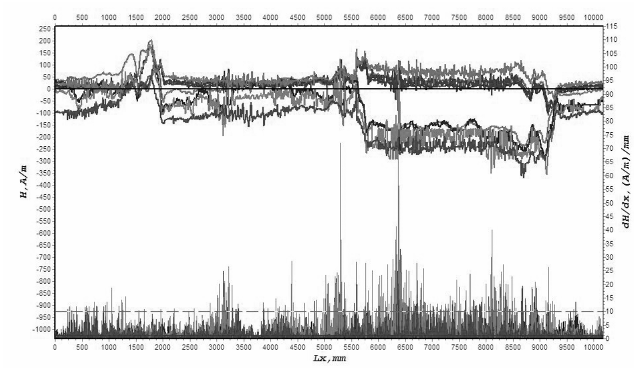

Fig. 3. Exemplary signals corresponding to head-checking defect
Rys. 3. Przykładowe sygnały odpowiadające wadzie typu „head-checking"

Feature set allowing for the discrimination between two group of signals should be as small as possible [4, 5]. This requirement suppresses the information redundancy, what in turn speeds up the learning process of classifier. As it can be seen form Fig. 2, 3, these signals are transient and non-stationary. For these signals the time-frequency analysis is usually carried out. Two factors are crucial in this analysis – temporal and frequency resolutions. In the Fourier transform, frequency resolution increases with the length of Fourier window. However, an increase in frequency resolution causes a decrease in temporal resolution. The trade-off between temporal and frequency resolutions is often necessary in the analysis. Wavelet transform is a technique, which allows the designer to choose the trade-off [6]. The wavelet packet's decomposition is an extension of wavelet transform, which calculates level by level transformation of signal from time domain to frequency domain [7]. At each level of decomposition a decrease in temporal resolution and an increase in frequency resolution take place. Let h(n) and g(n) be finite impulse response low-pass and high-pass filters respectively, where the Daubechies 14-points filter are used for wavelet packet decomposition. Let x(n) be original signal, whose length is equal to N, where N is power of 2. Decomposition consists in performing convolution of x(n) with h(n) and g(n) followed by a decimation by two. Let $x_s(n)$ and $x_d(n)$ be sequences resulting from low-pass filter decimation and high-pass filter decimation respectively, then it can be written:

$$\begin{aligned} x_s &= F_0\big(x(n)\big) = \sum_k x(k)h(2n-k) \\ x_d &= F_1\big(x(n)\big) = \sum_k x(k)g(2n-k) \end{aligned} \tag{3}$$

Decimation causes that each of sequences has half as many samples as x(n). Decomposition consists in performing formula (3) in the recursive manner. Fig. 4 presents exemplary decomposition showing in the tree form for x(n) of length 4. Each branch of tree called the bin vector contains discrete sequence. The bottom bin vectors have only one element.

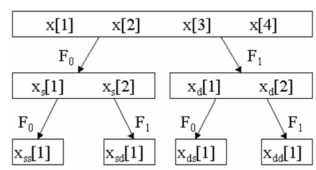

Fig. 4. Wavelet packet decomposition for signal x(n) of length 4
Rys. 4. Sposób dekompozycji na paczki falkowe sygnału x(n) o długości 4

The question should be posed, how outcomes of wavelet packet decomposition can be turned into the plausible feature set. In order to get rid of information redundancy, the feature set should contain as few elements as possible. In our approach to feature extraction, an average energy is calculated for each bin vector [8]. If e_y denotes average energy of bin vector **y** of length N, then:

$$e_y = \frac{1}{N}\mathbf{y}^T\mathbf{y} \tag{4}$$

Further, the energy vector **e** containing average energy of each bin vector is created in the following manner:

$$\mathbf{e} = [e_{y1} \quad e_{y2} \quad \dots \quad e_{yM}] \tag{5}$$

where M is the number of bin vectors in wavelet packet decomposition. In order to perform wavelet packet decomposition, 150 samples of signals corresponding to normal state and 30 samples of signals corresponding to head-checking defect has been chosen at random. Further, every sample was divided into frames of constant length 512 points. Every frame is processed by wavelet packet decomposition, followed by calculation of an average energy of each bin vector and creation of energy vector **e**. The signal corresponding to head-checking defect has two components, first carrying relevant to head checking defect information and second having irrelevant information – called the background noise. If both components are uncorrelated, then head –checking's signal is the sum of these components. In order to enhance differences between signals belonging to two groups the following normalization for every element i of energy vector **e** of every frame for every group is performed [8]:

$$\hat{e}_{yi} = \frac{e_{yi}}{e_{yi,ave}} \tag{6}$$

where:

$$e_{yi,ave} = \frac{1}{R}\sum_{j=1}^{R} e_{yij}^{(n)} \tag{7}$$

is an average energy of the noise for i-th element of energy vector **e**, $e_{yij}^{(n)}$ is the i-th element of j-th frame energy vector of signal belonging to the group corresponding to normal state of railway and R is the number of all frames of signals belonging to this group. In next step, the Principal Component Analysis (PCA) is performed. Its goal is to reduce the dimension of energy vector **e**. PCA is the statistical method defining the linear transformation of the form [9]:

$$\mathbf{y} = \mathbf{W}\mathbf{x} \tag{8}$$

transforming data $\mathbf{x} \subset R^N$ into vector $\mathbf{y} \subset R^K$ using the matrix $\mathbf{W} \subset R^{KxN}$ in such a way, that the output space **y** of the reduced dimension K<N preserves the most important information of the input space **x**. Let **x** be the random vector of zero mean and $\mathbf{R_{xx}}$ the correlation matrix of all vectors $\mathbf{x}_i$. The correlation matrix is symmetrical and non-negative defined. It means that all eigenvalues of $\mathbf{R_{xx}}$ are real and non-negative. Let the orthogonal eigenvectors associated with λ_i be denoted by $\mathbf{w}_i$. We arrange the eigenvalues in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ and in similar way the eigenvectors $\mathbf{w}_i$ associated with them, then the correlation matrix $\mathbf{R_{xx}}$ can be reconstructed as:

$$\mathbf{R_{xx}} = \sum_{k=1}^{N} \lambda_i \mathbf{w}_i \mathbf{w}_i^T \tag{9}$$

At orthogonal vectors $\mathbf{w}_i$ their contribution to the correlation matrix is measured by the value of the corresponding eigenvalues λ_i. In most of the practically important cases only small fraction of eigenvalues are large enough to contribute significantly to the reconstruction of $\mathbf{R_{xx}}$. Therefore for K<N most important eigenvalues the reconstruction of original vector **x** denoted here by $\hat{\mathbf{x}}$ can be written [9]:

$$\hat{\mathbf{x}} = \mathbf{W}^T\mathbf{y}, \quad \mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_K]^T \tag{10}$$

PCA transformation was applied to each group of data. Seven eigenvectors having the largest eigenvalues have been chosen. Fig. 5 presents distribution of data belonging to normal state of railway (inner part) and data belonging to head-checking defect (outer part) for two most important eigenvectors.

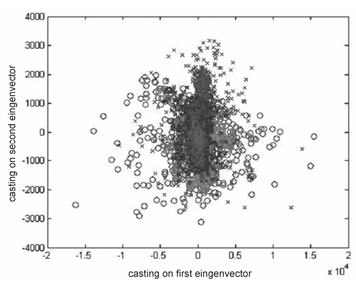

Fig. 5. Distribution of two group data for two most important eigenvectors
Rys. 5. Rozkład dwóch grup danych dla dwóch najbardziej znaczących wektorów

5. Classifier

SVM network has been used for the signal classification. SVM network generates the hyperplane separating two group of data. Hyperplane equation has the following form [10]:

$$y(\mathbf{x}) = \mathbf{w}^T \varphi(\mathbf{x}) + b = \sum_{j=1}^{K} w_j \varphi_j(\mathbf{x}) + b = 0 \tag{11}$$

where $\varphi(\mathbf{x})$ is the function transforming N-dimensional input space into K-dimensional feature space, and **w** is the synapses vector. If the input vector **x** satisfies relation y(**x**)>0 then it belongs to first class, otherwise it belongs to second class. The learning of SVM ensures maximization of distance between the nearest data belonging to different classes and is accomplished by quadratic programming. When classifying data are linear inseparable the final optimal form of hyperplane is [10]:

$$y(\mathbf{x}) = \sum_{i=1}^{N_{SV}} \alpha_i d_i K(\mathbf{x}, \mathbf{x}_i) + b \tag{12}$$

where N_{SV} is the number of Support Vectors for which Lagrange multipliers α_i occurring in quadratic programming have nonzero values, K(**x**, $\mathbf{x}_i$) is a kernel function having polynomial or gaussian form and d_i having -1 if $\mathbf{x}_i$ belongs to first class and +1 if $\mathbf{x}_i$ belongs to second class.

400 samples for every group have been used to learn each SVM network. After learning the generalization of networks has been tested on other group of data not attending in learning process. 180 samples of testing data for every group have been chosen at random. Learned networks misclassify 41 samples out of 180 what yields 23% misclassification.

6. Conclusion

Because signals generated with the use of metal magnetic memory method are very difficult to interpretation, obtained 23% misclassification seems to be satisfying result. Obtained classification score is comparable with outcome coming from traditional fully computerized diagnostic systems detecting railway defects with the use of ultrasonic method.

7. References

[1] V.T.Vlasov, A.A.Dubov: Physical bases of the metal magnetic memory method, Physica ZAO Publishing House, Moscow, 2004
[2] Ren Jlling, Song Kai, Wu Guanhua, Lin Junming: Mechanism study of metal magnetic memory testing. 10 Asia-Pacific Conference on Non-destructive Testing.Brisbane,Qeensland,Australia,2001
[3] P. Lesiak, A. Radziszewski: Diagnostyka szyn metodą magnetycznej pamięci metalu. Prace Naukowe Politechniki Radomskiej, Elektryka Nr 2(8) 2004, Radom 2004
[4] C.M. Bishop: Neural Networks for Pattern Recognition. Oxford University Press, 1996
[5] B. Ripley: Pattern Recognition and Neural Networks. Cambridge University Press, 1996
[6] I.Daubechies: Ten lectures on wavlets, SIAM Press, 1988
[7] R.Ciofman, M.Wickerhauser: Entropy-based algorithm for best basis selection, IEEE Trans. Info. Theory 38(2) March,1992
[8] R.E Learned, A.S. Willsky: A wavelet packet approach to transient signal classification, Applied and Computation Harmonic Analysis 2, 1995
[9] K.I.Diamantras, S.Y.Kung: Principal component neural networks, Wiley,1996
[10] B. Scholkopf, A. Smola: Learning with Kernels, Support Vector Machines, Regularization, Optimization and Beyond. MIT Press, Cambridge, Massachusetts, London, England 2002.

Artykuł recenzowany