Jan Maciej KOŚCIELNY, Michał SYFERT
INSTYTUT AUTOMATYKI I ROBOTYKI, POLITECHNIKA WARSZAWSKA
ul. Św. Andrzeja Boboli 8, 02-525 Warszawa

Fault Diagnosis Basing on the Information of the Interval Delays of Symptoms

Prof. dr hab. inż. Jan Maciej KOŚCIELNY

Prof. dr hab. inż. Jan Maciej Kościelny zatrudniony jest w Instytucie Automatyki i Robotyki, Politechniki Warszawskiej. Prowadzi badania z zakresu diagnostyki procesów przemysłowych i systemów mechatronicznych oraz układów sterowania tolerujących uszkodzenia. Kieruje pracami badawczymi, których rezultatami są opracowanie i przemysłowe aplikacje systemów zaawansowanego monitorowania i diagnostyki. Członek Komitetu Automatyki i Robotyki Polskiej Akademii Nauk.

e-mail: jmk@mchtr.pw.edu.pl

Dr inż. Michał SYFERT

Dr inż. Michał Syfert zatrudniony jest w Instytucie Automatyki i Robotyki, Politechniki Warszawskiej. Zajmuje się badaniami w dziedzinie diagnostyki procesów przemysłowych oraz zastosowań logiki rozmytej. Jest głównym autorem systemu diagnostycznego DIAG oraz jednym z głównych autorów zaawansowanego systemu monitorowania i diagnostyki AMandD.

e-mail: m.syfert@mchtr.pw.edu.pl

Abstract

The paper presents the formal conditions for fault isolability in the case of reasoning based on binary diagnostics matrix and the intervals of delays for each fault-symptom pair. It is shown that the higher fault isolability can be achieved while taking into account the knowledge about the symptoms delays. Finally, the new isolation algorithm utilising such knowledge is proposed. The presented approaches are illustrated with the example of fault isolability analysis for three tank system.

Keywords: fault detection, fault isolation, dynamic systems.

Diagnostyka uszkodzeń przy wykorzystaniu informacji o przedziałach opóźnień symptomów

Streszczenie

W artykule przedstawiono formalne warunki rozróżnialności uszkodzeń w przypadku wnioskowania wykorzystującego binarną macierz diagnostyczna oraz przedziały opóźnień zdefiniowane dla każdej pary uszkodzenie – symptom. Wykazano iż możliwe jest uzyskanie wyższej rozróżnialności uszkodzeń gdy wykorzystywana jest wiedza o opóźnieniach symptomów. Przedstawiono także nowy algorytm lokalizacji wykorzystujący tego typu wiedzę.

Słowa kluczowe: detekcja uszkodzeń, lokalizacja uszkodzeń, systemy dynamiczne.

1. Introduction

The knowledge about the mapping of the diagnostic signal S values space onto the fault F space is necessary for fault diagnosing. Different forms of such transformation are known. Structural and directional residual are most often used [2, 4, 7, 9, 13]. Such a relation is written down in the form of IF – THEN rules, logical functions or graphs [6, 7, 9]. Another solution is a Fault Isolation System FIS [7, 9, 11] which utilises multiple-valued evaluation of diagnostics signals.

All the above mentioned methods have static nature while the object of diagnosing is a dynamic system - the definite time, depending on the process dynamic properties, passes from the moment of fault occurrence to the moment when measurable symptoms appear. It implies that particular fault, in general, is detected by different diagnostic signals after different time periods. Only after some time all sensitive for particular fault diagnostics signals obtain the values testifying about fault existence. Avoiding the delays of the symptoms forming can lead do false diagnosis generation.

The knowledge about symptoms appearing sequence is an important information that can be used in the diagnostic reasoning. The different sequence of symptoms forming can be used for isolation of normally unisolable faults (fault with the same fault signatures). Taking into account the delays of the symptoms forming enables to increase fault isolability and, in many cases, to shorten the diagnosing time [7, 9, 10].

The precise determination of the delays of symptoms forming is usually impossible. However, one can specify time intervals in which the particular symptom should be registered. The monitoring algorithms taking into account the delays of the symptoms forming have been presented in just few publications. Such algorithm for linear systems was presented in [12]. It utilises process models that take into account the fault influence. Such a precise knowledge is very seldom available to diagnostics system designers. Therefore, the simplified notations of the knowledge about the delays of the symptoms forming are used. In the publications [8, 9] the reasoning algorithm taking into account the maximal possible delays of symptoms forming was formulated. The algorithms presented in [10] utilise the lower and the upper bounds on the symptoms delay intervals for particular faults.

In this paper the authors analyse the fault distinguishability in the case of the diagnosing based on binary diagnostic matrix and the intervals of particular symptoms delays. The conditions for unconditional fault unisolability and isolability as well as conditional fault isolability are formed. The increase of fault distinguishability in comparison with the diagnosing based only on the binary diagnostic matrix is shown. Finally, the new algorithm of fault isolation utilising the knowledge about the symptoms delays intervals is formulated.

2. Estimation of symptoms forming delays

The delays of symptoms forming depend on dynamic properties of tested part of the process, fault characteristic (incipient and abrupt) as well as the applied diagnostic method and the algorithm parameters. They can be calculated analytically based on the known dynamic description (e.g. transmittances) of the tested subsystem (with faults as an input and measured signal as an output), known fault developing time characteristics, known parameters of boundary function, assuming lack of influence of the test realisation techniques.

The mathematical process description can be achieved based on equations describing physical phenomena. The faults are treated as specific process inputs and should be considered in the equations describing the process. By linearization of the equations in the working point and carrying out the Laplace transformation one achieves linear model consisting of the system of equations describing the dependence of particular process outputs, inputs and faults [2, 3, 4, 13]:

$$y(s) = G(s)u(s) + H(s)f(s) \quad (1)$$

where: y – outputs vector, u – inputs vector, f –faults vector, $G(s)$ – outputs-inputs transmittance matrix, $H(s)$ – outputs-faults transmittance matrix.

Particular equations have a form of:

$$y_j(s) = G_j(s)u(s) + H_j(s)f(s) \tag{2}$$

where $G_j(s)$, j=1,...,J (J – the number of outputs) includes the transmittance of output-input type (P – number of inputs):

$$G_{j,p}(s) = y_j(s)/u_p(s) : p = 1, \dots, P \tag{3}$$

and $H_j(s)$, j=1,...J, includes the transmittance for particular output-fault pairs (K – number of faults):

$$H_{k,j}(s) = y_j(s)/f_k(s) : k = 1, \dots, K. \tag{4}$$

In the case of faults absence the following dependence if fulfilled: $y_j(s) - G_j(s)u(s) = H_j(s)f(s) = 0$. The residuals are calculated based on the dependence called computational form:

$$r_j(s) = y_j(s) - G_j(s)u(s). \tag{5}$$

The equation (6) reflects the general relation between the particular residual and faults. It is called an internal form of a residual [4]:

$$r_j(s) = H_j(s)f(s) = \sum_{k=1}^{K} H_{k,j}(s)f_i(s) \tag{6}$$

If the r_j is sensitive to the fault f_k, assuming other faults f_m=0, one achieves:

$$r_j(s)\big|_{f_k} = H_{k,j}(s)f_k(s). \tag{7}$$

Knowing the fault time development function $f_k(t)$ the residual time development function can be determined based on inverse Laplace transformation:

$$r_{k,j}(t) = L^{-1} r_j(s)\big|_{f_k} = L^{-1}\left[H_{k,j}(s)f_k(s)\right]. \tag{8}$$

Assuming the shape of the function $f_k(t)$ (e.g. step function $f_k(t)$=1(t)f_{k0} comply with abrupt fault of a value f_{k0}) and the threshold residual value A_j it is possible to determine the time, after which the symptom of the k^{th} fault will appear.

Let us assume the uncertainties of the particular transmittances expressed as the intervals of their parameters, e.g. $k \in (k_1,k_2)$, $T \in (T_1,T_2)$, and the notations $\theta^1_{k,j}$ and $\theta^2_{k,j}$ for minimal and maximal time periods from k^{th} fault introduction to j^{th} symptom appearing, respectively. In the considered case, the minimal and maximal symptoms delays are determined by the terminal values of the parameters, the magnitude (strength) of the fault and its time developing function. The real symptoms delay time will include between $\theta^1_{k,j}$ and $\theta^2_{k,j}$.

In practice, the analytical determination of the symptoms forming delays is very difficult because it requires the modelling of the fault influences onto measured outputs. Additionally, the time development function of residual forming and the residual threshold value must be arbitrary assumed and the accuracy of estimation of the symptoms delays is not high due to the modelling errors.

3. Formal description of diagnosed system

Most often, for the isolation of faults from the set F the set S of binary diagnostics signals is used:

$$F = \{f_k : k = 1,2, \dots, K\}, \tag{9}$$

$$S = \{s_j : j = 1,2, \dots, J\}. \tag{10}$$

The diagnostic signals are calculated as a result of threshold evaluation of absolute residual values (A – threshold value):

$$|r_j| \le A \Rightarrow s_j = 0 \; or \; |r_j| > A \Rightarrow s_j = 1. \tag{11}$$

The occurring of the diagnostic signal value of "1" is a fault symptom. It testifies about the existence of one of the faults from the subset $F(r_j)$ of faults for which the j^{th} residual is sensitive for:

$$F(r_j) = F(s_j) = \{f_k : r_j = \psi(f_k)\} \tag{12}$$

where $\Psi(f_k)$ is determined from (6).

Based on the residual equations in the internal form (6) one can define the subsets $F(r_j)$ and, finally, the binary diagnostics matrix defined on the Cartesian product of the sets of faults and diagnostic signals:

$$Q_{FS} \subset F \times S. \tag{13}$$

The elements of this relation take the shape of:

$$q(f_k, s_j) \in \{0,1\}. \tag{14}$$

The simple diagnostics system is defined by the following triple:

$$SD = \langle F, S, Q_{FS} \rangle \tag{15}$$

which can be expressed in the form of bipartite graph [9]. The minimal and maximal values of the delays of fault symptoms $\langle \Theta^1_{k,j}, \Theta^2_{k,j} \rangle$ are attributed to each pair of fault-symptom such as $q(f_k, s_j) = 1$.

4. Fault isolability

The analysis of achieved fault isolability for the system described by the triple (15) and the parameters of symptoms delay intervals is presented in this section. It concerns the cases of faults that are unisolable based on binary diagnostic matrix - faults f_k and f_m, for which for all diagnostic signals $s_j \in S$ the condition $q(f_k, s_j) = q(f_m, s_j)$ is fulfilled [1, 4, 7, 9].

One must notice that the single diagnostic signal s_j sensitive for faults f_k and f_m is not capable to distinguish these faults even if their intervals of symptoms delays are disjoint ($[\Theta^1_{k,j}, \Theta^2_{k,j}] \wedge [\Theta^1_{m,j}, \Theta^2_{m,j}] = \emptyset$) due to the lack of knowledge about the moment of fault erasing.

Definition 1. Any two faults f_k and f_m are unconditionally unisolable, in respect to the intervals of symptoms delays, in the following two cases:

- they are detectable only by one, the same, diagnostics signal:

$$|S(f_k)| = |S(f_m)| = 1, \tag{16}$$

where $S(f_k) = \{s_j : q(f_k, s_j) = 1\}$ is the set of diagnostic signals sensitive for the k^{th} fault,
- the minimal and the maximal values of the symptoms delays are shifted by the same constant value τ, for all diagnostic signals:

$$\forall_{s_j \in S \wedge [q(f_k,s_j) \neq 0] \wedge [q(f_m,s_j) \neq 0]} \left[\Theta^1_{m,j} = \Theta^1_{k,j} + \tau\right] \wedge \left[\Theta^2_{m,j} = \Theta^2_{k,j} + \tau\right]. \tag{17}$$

The above dependence implies that the intervals of delays have the same length for all the faults and particular diagnostic signal s_j.

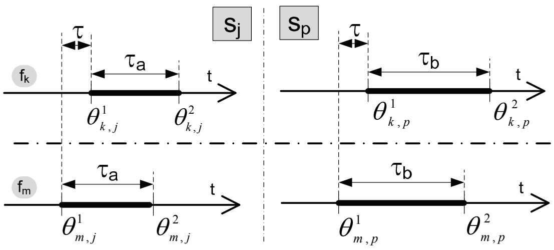

Fig. 1. The graphical representation of unconditional fault unisolability
Rys. 1. Graficzna reprezentacja bezwarunkowej nierozróżnialności uszkodzeń

The analysis of undonditional fault isolability must be conducted for two cases. They are taken into account in the following definition.

Definition 2. Any two faults f_k and fm are unconditionally isolable, in respect to the intervals of symptoms delays, in the following two cases:

- the faults are detectable by the diagnostics signals s_j i s_p always in different sequence, i.e. the following dependence is fulfilled:

$$\exists_{s_j,s_p \in S}\left(\Theta^2_{k,j} < \Theta^1_{k,p}\right) \wedge \left(\Theta^2_{m,p} < \Theta^1_{m,j}\right). \tag{18}$$

- the faults are detectable by any two diagnostics signals s_j i s_p in the same sequence but the maximal value of the sp symptom delay in respect to the symptom s_j for one of the faults, e.g. f_k ($\Theta^2_{k,p} - \Theta^1_{k,j}$), is less than the minimal value of this delay for the second fault, e.g. f_m ($\Theta^1_{m,p} - \Theta^2_{m,j}$), the following dependence is fulfilled:

$$\left(\Theta^1_{m,p} - \Theta^2_{m,j}\right) > \left(\Theta^2_{k,p} - \Theta^1_{k,j}\right). \tag{19}$$

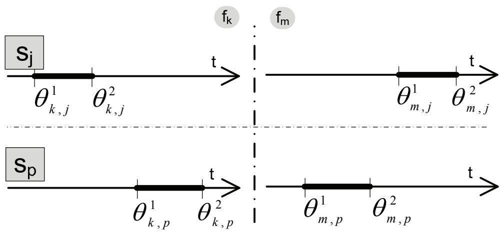

Fig. 2. Case 1: different sequence of faults detection by two diagnostics signals
Rys.2. Przypadek 1: różna sekwencja detekcji uszkodzenia dwóch sygnałów diagnostycznych

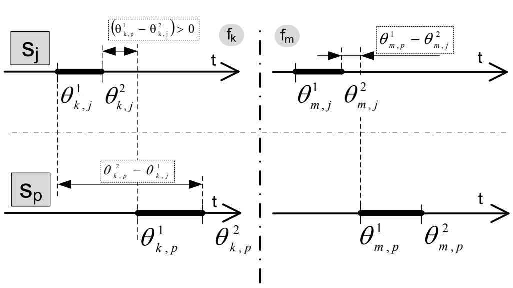

Fig. 3. Case 2: Let us assume that signal s_j detects the faults f_k and f_m earlier that signal s_p and the delays of symptom of the fault f_k are lower than the fault f_m, i.e. $\Theta^2_{k,j} - \Theta^1_{m,j}$
Rys. 3. Przypadek 2: załóżmy, że sygnał s_j wykrywa uszkodzenia f_k i f_m wcześniej niż sygnał s_p a opóźnienia symptomów uszkodzenia f_k są mniejsze niż uszkodzenia f_m czyli $\Theta^2_{k,j} - \Theta^1_{m,j}$

Except unconditional fault unisolability and isolability based on intervals of symptoms delays, one can define conditional isolability. Conditional isolability (or unisolability) means, that the fault can be isolable or not in respect to actual values of the symptoms delays (being situated in defined intervals $[\Theta^1_{k,j}, \Theta^2_{k,j}]$) observed during diagnostic procedure. On the design stage it is not possible to decide if the faults are isolable due to the lack of knowledge about the real delays.

Definition 3. Two faults f_k and f_m are conditionally isolable, in respect to the intervals of symptoms delays, when these faults are not unconditionally unisolable or unconditionally isolable.

Taking into account the time of symptoms forming can increase the fault isolability comparing with diagnosing based only on binary diagnostic matrix.

5. Fault isolation algorithm

This section presents the fault isolation algorithm that utilises the knowledge about the diagnostic relation and the values of the minimal and maximal delays of symptoms forming. It assumes single fault scenarios, however, the multiple faults issue is also addressed. The algorithm implements serial diagnostic reasoning based on generation of temporary diagnosis. The temporary diagnosis are further specified according to evaluation of successive diagnostic signals, taking into account the intervals of symptoms delays.

Three main stages of reasoning can be distinguished: initialisation, diagnosis specifying, and final diagnosis formulation.

Initialisation of isolation procedure. The isolation algorithm starts in time moment t_1=0 when the first symptom $s^1_x = 1$ is observed (detected). The hypothesis about single fault is assumed. The following notation is used: DGN for diagnosis elaborated based on diagnostic relation and interval of symptom delay for individual diagnostic signals, DGN* for diagnosis elaborated by the inter-connections of intervals of symptom delays for several diagnostic signals. The following steps are conducted:

- *Determining the set of possible faults*. The primary set of possible faults is determined based on diagnostic relation. It consists of all the faults, for which the diagnostic signal with the observed symptom is sensitive for:

$$(s^1_x = 1) \Rightarrow \mathrm{DGN}^*(s^1_x) = \{f_x : q(f_k, s^1_x) = 1\}, \tag{20}$$

  where $\mathrm{DGN}^*(s^1_x)$ denotes temporary diagnosis, elaborated under the condition of use of the first diagnostic signal s^1_x but without taking into account the intervals of symptoms delays.
- *Reduction of primary set of possible faults*. Let us introduce the notations $\theta^1_{k,x}$ and $\theta^2_{k,x}$ for minimal and maximal periods from k^{th} fault occurring to the first, detected symptoms $s^1_x = 1$ formulation, respectively. The faults, which occurrence should cause another symptoms to be observed before the symptom s^1_x, in respect to known intervals of symptoms delays, are eliminated from the set $\mathrm{DGN}^*(s^1_x)$:

$$\mathrm{DGN}(s^1_x) = \{f_k \in \mathrm{DGN}^*(s^1_x) : \forall_{s_j \neq s^1_x} \Theta^2_{k,j} < \Theta^1_{k,x}, \tag{21}$$

  while $\mathrm{DGN}(s^1_x)$ denotes first, temporary diagnosis elaborated while taking into account the intervals of the symptoms delays.
- *Determining the set of diagnostic signals useful for further fault isolation* in the following form:

$$S^* = \{s_j : F(s_j) \cap DGN(s^1_x)\} - s^1_x. \tag{22}$$

- *Defining the time intervals of consecutive symptoms possible formulation*. Due to the fact that the real time of fault occurring is unknown (only the time of the first symptom detection is registered) the time intervals of the appearing of the consecutive symptoms of the diagnostic signals from the set S^* must be determined, starting from the moment of first symptom detection. Such calculations must be conducted for the faults pointed out in the diagnosis DGN/s_j^1 in the following way:

$$\beta_{k,j}^1 = \begin{cases} 0 & if\ (\theta_{k,j}^1 - \theta_{k,x}^2 \leq 0) \\ (\theta_{k,j}^1 - \theta_{k,x}^2) & if\ (\theta_{k,j}^1 - \theta_{k,x}^2 \geq 0) \end{cases} \quad (23)$$

$$\beta_{k,j}^2 = \theta_{k,j}^2 - \theta_{k,x}^1. \quad (24)$$

The parameters $\beta_{k,j}^2$ for the faults $f_k \in \mathrm{DGN}(s_x^1)$ and the diagnostic signals $s_j \in S^*$ are arranged in ascending order.

Iterative diagnosis specification. The second part of the reasoning has iterative nature. The elaboration of the following diagnosis takes place:

- after detection of each, successive fault symptoms,
- each time when the maximal period of the symptom delay $\beta_{k,j}^2$ from the ordered series of these parameters passes.

During this stage, the following steps are conducted iteratively:

- *The reduction of the set of possible faults based on diagnostics relation.* If the symptom s_j=1, $s_j \in S^*$ was detected in the proper period of delays than the set of possible faults is reduced according to formula:

$$(s_j^r = 1) \wedge (s_j \in S^*) \Rightarrow \mathrm{DGN}^*(s_x^1, \ldots, s_j^{r-1}, s_j^r) = \\ = \{f_k \in DGN(s_x^1, \ldots, s_j^{r-1}) : (q(f_k, s_j) = 1) \wedge (t \in [\beta_{k,j}^1, \beta_{k,j}^2]. \quad (24)$$

Such an operations is realised for all the faults from the set $f_k \in F(s_j)$.

- *The reduction of the set of possible faults based on the analysis of delays interval.* The faults, which occurrence should cause another symptoms s_p=1 to be observed before the currently observed symptom, in respect to the known intervals of symptoms delays, can be eliminated from the diagnosis elaborated in previous step:

$$\mathrm{DGN}(s_x^1, \ldots, s_j^r) = \{f_k \in \mathrm{DGN}^*(s_x^1, \ldots, s_j^{r-1}) : \beta_{k,p}^2 < \beta_{k,j}^1. \quad (25)$$

- *The reduction of the set of possible faults after the analysis of the maximal times of the symptom delays*. The lack of symptom after predefined time period, $t > \beta_{k,j}^2$, allows for the reduction of the set of possible faults due to the formula:

$$(s_j = 0) \wedge (s_j \in S^*) \wedge (t > \beta_{k,j}^2) \Rightarrow \\ \mathrm{DGN}^* = \{f_k \in \mathrm{DGN}(s_x^1, \ldots, s_j^{r-1})\} - f_k. \quad (26)$$

The end of fault isolation. The algorithm stops when all the diagnostic signals from the set S^* are taken into account.

The algorithm has also limited ability to isolate multiple faults. In general, if the new symptom s_j=1 is observed, and if it did not appeared in predefined period of symptom delays in respect to the first observed symptom, than the ***new fault isolation thread*** is started. It is assumed that this symptom is caused by another fault than the faults pointed out in the previous steps:

$$(S_j^r = 1) \wedge (t \notin [\beta_{k,j}^1, \beta_{k,j}^2]) \Rightarrow \\ \mathrm{DGN}^*(s_j^1) = \{f_k : q(f_k, s_j) = 1\}. \quad (27)$$

6. Summary

The basic method of increasing fault isolability is the increase of measured signals, which leads to the generation of additional residuals. The residual structurization, i.e. creation of secondary residuals based on primary ones [3, 4, 5], is also a very effective method. The increase of the set of measured signals is not always possible and economically justified. Also, the residual structurization not always gives satisfactory results. The alternative or complementary methods of increasing fault isolability are the application of multiple-valued residual evaluation [7, 9, 11] or presented in this paper approach based on the knowledge about the delays of forming of fault symptoms.

The presented fault isolation algorithm that utilises the minima and maximal values of fault symptoms delays can lead to higher fault isolability. Additionally, it protects against false diagnosis caused by Dynamics of symptoms forming. In some cases, it also makes possible to achieve shorter times of diagnosis in comparison with the diagnostic algorithms based on binary diagnostic matrix when the final diagnosis can be elaborated after all values of diagnostic signals are steady.

7. References

[1] Basseville M.: On fault detectability and isolability. In: European Journal of Control, Vol 7, No 6., 625-637, 2001.
[2] Chen J., Patton R. J.: Robust model based fault diagnosis for dynamic systems. Kluver Akademic Publishers, Boston, 1999.
[3] Gertler J.: Analitical redunduncy methods in fault detection and isolation. Proc. of: IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes - SAFEPROCESS'91, Baden-Baden , Vol. 1, 9-21, 1991.
[4] Gertler J.: Fault Detection and Diagnosis in Engineering Systems. Marcel Dekker, Inc. New York - Basel - Hong Kong, 1998
[5] Gertler J., Singer D.: A new structural framework for parity equation based failure detection and isolation. Automatica, Vol. 26, no. 2, 381-388, 1990.
[6] Frank P. M.: Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy. Automatica, 26, 459-474, 1990.
[7] Korbicz J., Kościelny J. M., Kowalczuk Z., Cholewa W. (ed): Fault Diagnosis. Models, artificial intelligence, application. Springer, 2004.
[8] Kościelny J. M.: Fault Isolation in Industrial Processes by Dynamic Table of States Method. Automatica, Vol.31, No.5, 1995, 747-753
[9] Kościelny J. M.: Diagnostyka zautomatyzowanych procesów przemysłowych. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2001.
[10] Kościelny J. M., Zakroczymski K.: Fault Isolation Algorithm that Takes Dynamics of Symptoms Appearances Into Account. Bulletin of the Polish Academy of Sciences. Technical Sciences. Vol.49, No 2, 323-336, 2001.
[11] Kościelny J. M., Bartyś M., Rzepiejewski P., Sá da Costa J.: Actuator fault distinguishability study. Control Engineering Practice (Elsevier) 14, 2006, 645-652, 2006.
[12] Kościelny J. M., Ligęza A., Dziembowski B.: Lokalizacja uszkodzeń w układach liniowych z uwzględnieniem wiedzy o sekwencji powstawania symptomów. W. Sterowanie i Automatyzacja: Aktualne problemy i ich rozwiązania. Red. K. Malinowski, L. Rutkowski. Akademicka Oficyna Wydawnicza EXIT, Komitet Automatyki i Robotyki PAN, Polskie Towarzystwo Sieci neuronowych, 2008, 481-498.
[13] Patton R., Frank P., Clark R. (Eds.): Issues of fault diagnosis for dynamic systems. Springer, 2000.

otrzymano / received: 16.03.2011
przyjęto do druku / accepted: 04.04.2011 *artykuł recenzowany*