Paweł WNUK, Jan Maciej KOŚCIELNY

INSTYTUT AUTOMATYKI I ROBOTYKI, POLITECHNIKA WARSZAWSKA ul. Św. Andrzeja Boboli 8, 02-525 Warszawa

Diagnostic system decomposition with genetic optimization

Ph.D., eng. Paweł WNUK

He works in Institute of Control and Robotics of Warsaw University of Technology. His main research field covers new methods of modeling of system dynamics, especially basing on fuzzy logic. He is one of co-authors of systems designed for advanced monitoring and diagnostics of industrial processes AMandD and DiaSter.



e-mail: p.wnuk@mchtr.pw.edu.pl

Abstract

This paper describes the use of genetic optimization in diagnostic system decomposition. First, an overview of diagnostic system and the reasons for its decomposition are given. Decomposition quality index is proposed. In the second part the analysis of a genetic algorithm application possibility is placed. The structure of the genome, type of genetic algorithm and genetic operators are described. Investigations on best values of key parameters is the main subject of third part. As a summary, a description of industrial application for diagnostic system decomposition on a hydrocarbon plant is placed.

Keywords: system diagnostics, genetic algorithms, system decomposition.

Dekompozycja systemu diagnostycznego przy wykorzystaniu algorytmów genetycznych

Streszczenie

Artykuł zawiera opis wykorzystania optymalizacji genetycznej do dekompozycji systemu diagnostycznego. Na początku zamieszczono definicje systemu diagnostycznego i przesłanki do stosowania dekompozycji. Zaproponowano wskaźnik jakości dekompozycji. Druga część zawiera analizę możliwości wykorzystania algorytmu genetycznego do rozwiązania postawionego zadania. Przedstawiono strukturę genomu, typ algorytmu oraz operatorów genetycznych. Następna część artykułu zawiera wyznaczenie optymalnych wartości kluczowych parametrów algorytmu genetycznego. Na koniec przedstawiono przykład zastosowania opracowanej metodologii na rzeczywistej instalacji.

Slowa kluczowe: diagnostyka, algorytmy genetyczne, dekompozycja systemu.

1. Expedience of system diagnostic in decentralized structure

Nowadays, most of digital control systems are distributed, taking into account physical structure as well as responsibility distribution. Low-level control units have direct, hardware connection with supervised object or installation and are assigned to specific technological unit. Higher-level stations usually are not connected directly to measurement equipment, but to low-level controllers. Control functions are distributed between many computer units placed in different positions and working in parallel, thus typical control system for large scale plant is called Distributed Control System. Diagnostic functions, as a part of control and safety tasks, could be also realized in decentralized structure.

In analogy to decentralized control systems, the decentralization of diagnostic systems has many advantages:

- Selected parts of installation can be diagnosed in parallel by separate, independent diagnostic computers.
- Technological object decomposition causes important decrease of investigated system states, what is particularly important

Prof., D.Sc., eng. Jan Maciej KOŚCIELNY

He works in Institute of Control and Robotics of Warsaw University of Technology. His main research area covers diagnostics of industrial processes, mechatronics systems, and control circuits tolerating faults. He is a chef of research group in the scope of industrial applications of systems for advanced monitoring and diagnostics. He is a Member of Committee of Automatics and Robotics of Polish Academy of Science.



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

during analysis of multiple faults. Decomposition causes also significant reduction of demanded computational power and thus shortens calculation time, also for diagnostic signals generation.

- The decomposition causes, that assumption about existence only single fault at one time is more rational. Such assumption gives a simplification of subsystem diagnostic algorithms.
- Decentralized structure promises more robust diagnostic system in opposition to centralized one. Such system is more resistant to single diagnostic computers failure.
- Decentralized diagnostic provides better fitted diagnostic information for different users (information is generated and presented in suitable form, different for operators, supervisors, technology engineers, etc.)
- Diagnostic system with decentralized structure can be started step-by-step, separately for each subsystem.

Diagnostic system application for huge plants requires system decomposition for subsystems controlled and diagnosed by separate computer units. Thus algorithms that support system decomposition were investigated before. Some notes about methodology applicable to hierarchical systems can be found in [4, 5, 7], publications [6] are focused on one-level structures.

Usually complete independent subsystems cannot be separated. The number of interactions in present industrial applications is so high, that in practice always failure in one subsystem has an influence to other subsystems. It is recommended to divide whole system in such way, that interconnections between subsystems are minimized. In [4] the rules how to find independent parts in the system and how to divide the process from technological point of view were given. The problem of system decomposition with minimal interactions between subsystems was formulated. Example solution with heuristic Kernighhan and Lin algorithm (1970) [3] was given in [6]. This paper is focused on genetic algorithms application to this problem.

2. Diagnostic system description

Diagnostic system can be described as [8]:

• set of all possible faults F, interpreted as any destructive event that causes system work quality degradation:

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

• set of diagnostic signals, treated as inputs to detection algorithms in the system:

$$S = \{s_j : j = 1, 2, \dots, L\},$$
(2)

• Diagnostic relation defined on Cartesian product of *F* and S sets:

$$R_{FS} \subset F \times S \ . \tag{3}$$

• Formula $f_k R_{FS} s_j$ means, that test s_j detects fault f_k , in other words, an occurrence of fault f_k causes the appearance of diagnostic signal s_j with value 1, (the symptom). Relation matrix R_{FS} is binary diagnostic matrix with elements defined as follow:

$$r(f_k, s_j) \equiv v_j(f_k) = \begin{cases} 0 \Leftrightarrow \langle f_k, s_j \rangle \notin R_{FS} \\ 1 \Leftrightarrow \langle f_k, s_j \rangle \in R_{FS} \end{cases}$$
(4)

• Relation *R_{FS}* can be defined by assigning to each test a subset of faults detected by this test:

$$F(s_j) = \left\{ f_k \in F : f_k R_{FS} s_j \right\}$$
(5)

3. Problem formulation

The decomposition of complicated technological unit is necessary to diagnose it in a decentralized structure. The decomposition in this case is equivalent to the search of subsystems with limited size, characterized with maximum possible mutual independence degree. The recommended independence requirement can be defined as separation of faults subsets F_n or separation of diagnostic signals subsets S_n . The problem can be formulated as follows:

3.1. Variant 1

The set of faults *F*, diagnostic signals *S*, process variables *X* and diagnostic relation R_{FS} (3) of complex technological unit are given. Diagnostic signals $s_j \in S$ should be assigned to *N* separated subsets S_n , each with limited count of elements:

$$\forall \left| S_n \right| \le \gamma , \tag{6}$$

in a way that minimizes following expression:

$$Q_{S} = \sum_{\substack{m,n=1,2,\dots,N\\m\neq n}} \sum_{\substack{s_{i} \in S\\s_{j} \in S_{n}}} \left| F(s_{i}) \cap F(s_{j}) \right|.$$
(7)

3.2. Variant 2

The set of faults *F*, diagnostic signals *S*, process variables *X* and diagnostic relation R_{FS} (3) of complex technological unit are given. Faults $f_k \in F$ should be assigned to *N* separated subsets F_n , each with limited count of elements:

$$\left| F_n \right| \le \gamma \,, \tag{8}$$

in a way that minimizes following expression:

$$Q_F = \sum_{\substack{m,n=1,2,\dots,N\\m\neq n}} \sum_{\substack{f_k \in F_m\\f_p \in F_n}} \left| S(f_i) \cap S(f_j) \right| \,. \tag{9}$$

Decomposition assures minimal dependency between subsystems, and limits for example the need of information exchange between them. Quality indexes (7) and (9) indirectly characterize the number of needed data transmission between diagnostic subsystems.

4. Genetic algorithm decomposition

The problem mentioned before can be treated as an optimization task (search for minimal value of quality index (7) or (9)). To solve this problem genetic algorithm in a standard form, with constant length genome, was applied. Each genome (individual) represents complete partitioning of diagnostic matrix. The genotype has a form of vector consisting of integer numbers, with size equal to diagnostic signals (variant 1) or faults (variant 2) count. The attachment of diagnostic signal or fault to given subset is marked by integer number treated as subset number:

$$g = \left[g_1, g_2, \dots, g_K\right] \in {}^K \tag{10}$$

for variant 2:

$$g = \left[g_1, g_2, \dots, g_L\right] \in {}^L \tag{11}$$

In both variants following restriction is fulfilled:

$$\forall g_i \in \langle 0; N-1 \rangle. \tag{12}$$

Genome definition grants full and separate partitioning of whole set S or F into N subsets (6). The limit of each subset size is assured by the form of final quality index (described below).

In the paper, the genetic operators in well known form were used:

Random initiation:

$$\bigvee_{i} g_{i} = rand \langle 0; N-1 \rangle \tag{13}$$

One-point crossover: for given parents g^F and g^M, and randomly selected cross point c_p∈⟨1; N−1⟩, creates two children:

$$g^{S} = \left[g_{1,}^{F}g_{2,}^{F}...,g_{c_{p}}^{F},g_{c_{p}+1}^{M}...,g_{N}^{M}\right]$$

$$g^{D} = \left[g_{1,}^{M}g_{2,}^{M}...,g_{c_{p}}^{M},g_{c_{p}+1}^{F}...,g_{N}^{F}\right]$$
(14)

- Multi-point crossover: similar to one-point crossover, except that there is more than one point of crossing and gene exchange.
- Even-odd crossover: for given parents g^F and g^M , creates two children:

$$g^{S} = \left[g_{1,}^{F}g_{2,}^{M}g_{3}^{F}, g_{4}^{M}, \dots\right]$$

$$g^{D} = \left[g_{1,}^{M}g_{2,}^{F}g_{3}^{M}, g_{4}^{F}, \dots\right]$$
(15)

Each genome after any changes caused by any genetic operator mentioned above, needs to be renumbered (change integers inside genome in order to have all numbers in rising order). This operation is needed because we do not take into account the order of received sets. For example, same solution can be described as:

$$g_1 = [001222], g_2 = [112000], g_3 = [110222], \dots$$
 (16)

In first approach classical genetic algorithm with elitism was applied (best individual is kept population by population). Objective function is calculated directly with the use of diagnostic matrix, according to the following algorithm:

- The genome is decoded in order to get separated subsets of diagnostic signals. Initial value of quality index is equal to zero Q_a=0.
- 2. For each subset of diagnostic signals S_k , $k \in \langle 0, N-1 \rangle$.
 - 2a. For each diagnostic signal $\forall s_i \in S_k$ the set of faults detected by him is created.

- 2b. For each fault $\forall f_j$ detected by s_i , the set of co-symptoms \hat{S}_i , also sensitive for fault f_i , is created.
- 2c. If any element from \hat{S}_i do not belongs to S_k , then increment quality index by one: $Q_a = Q_a + 1$.

At the end, to avoid single subset size growth over defined limit, an additional penalty to quality index is added:

$$Q = Q_a + \sum_{k} 10 * max(0, size(S_k))$$
(17)

5. Test example

Lets decompose diagnostic system, defined by diagnostic matrix from Table 1, and suppose, that three subsystems should be received as a result. The number of diagnostic signals in each subsystem should not be bigger than 6.

There are three different optimal decompositions, in sense of minimal interactions (7), fulfilling given assumptions, presented in Tables 2, 3 and 4. Separate sets of diagnostic signals are indicated with colours (white – subsystem 1, blue – subsystem 2, orange – subsystem 3). The influence of mutual interactions is shown in columns with ones on the red background.

As example, lets consider second optimal solution. In the first and second subsystem fault f_9 is detected, in the second and third subsystem common faults are $\{f_{11}, f_{20}\}$.

Each optimal solution have same value of quality index $Q_a=3$.

Tab. 1.Diagnostic matrix of an example systemTab. 1.Macierz diagnostyczna przykładowego systemu

| | f1 | f2 | f3 | f4 | f5 | f6 | f7 | f8 | f9 | f10 | f11 | f12 | f13 | f14 | f15 | f16 | f17 | f18 | f19 | f20 | f21 | f22 |
|-----|----|----|----|----|----|----|----|----|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| s1 | | 1 | | 1 | | | | | | | | | | | | | | | | | | |
| s2 | | | 1 | | | 1 | | | | | | | | | | | | | | | | |
| s3 | | | | | | | | | | | | | | | | | | | | | | |
| s4 | 1 | | | | 1 | 1 | | | | | | | | | | | | | | | | |
| s5 | | | | | 1 | | 1 | | 1 | | | | | | | | | | | | | |
| s6 | | | | | | | 1 | | | | | | 1 | | | | | | | | | |
| s7 | | | | | | | | | | 1 | | 1 | | | | | | | | | | |
| s8 | | | | | | | | | 1 | | | | 1 | 1 | | | | | | | | |
| s9 | | | | | | | | 1 | 1 | | 1 | | | 1 | | | | | | | | |
| s10 | | | | | | | | | | | 1 | | | | 1 | | | 1 | | | | |
| s11 | | | | | | | | | | | | | | 1 | | 1 | | | | 1 | | |
| s12 | | | | | | | | | | | | | | | | | 1 | | 1 | | | |
| s13 | | | | | | | | | | | | | | | | | | 1 | | | 1 | |
| s14 | | | | | | | | | | | | | | | | | | 1 | | | 1 | |
| s15 | | | | | | | | | | | | | | | | | | | 1 | | | 1 |
| s16 | | | | | | | | | | | | | | | | | 1 | | | | | 1 |

Tab. 2.Optimal decomposition 1

Tab. 2. Dekompozycja optymalna 1



Tab. 3. Optimal decomposition 2 Tab. 3. Dekompozycja optymalna



Tab. 4.Optimal decomposition 3Tab. 4.Dekompozycja optymalna 3



6. Genetic algorithm testing

The first approach assumes the application of the genetic algorithm in classic form, that Goldberg describes in his book [2], with complete exchange of individuals in each generation, except the best one. Each generation of the algorithm creates an entirely new population of individuals by selecting from the previous population (using roulette wheel), then mating to produce the new offspring for the new population. This process continues until the stopping criteria are met. Random initiation (13) and one-point crossover (14) were used, algorithm minimizes quality index (17).

To find best values of key parameters of genetic algorithm application, a number of trial optimization runs were processed. Investigated parameters were: crossover probability, mutation probability and population size. For each combination of parameters from sets given below, 50 runs of GA were executed.

- Population size: {10, 20, 35, 50, 75, 100};
- Crossover probability: {0.2, 0.4, 0.6, 0.7, 0.8, 0.9, 0.99};
- Mutation probability: {0.002, 0.007, 0.02, 0.07, 0.1, 0.2, 0.3};

Stopping criteria was first of: optimal solution found (individual with $Q_a=3$) or objective function calls count bigger than 150 000. The limit of quality function calls was determined arbitrary, with value that allows to keep rational calculation time, and significantly below the number of all possible partitions (for given example $3^{15}>1.4\cdot10^7$ combinations of subsystems). In case of exceeding quality function calls count without reaching optimum, the optimization was qualified as missing try.

Mean number of missing tries (in percent) for all combinations of investigated parameters are presented in Fig. 1.

It should be mentioned, that in about 90% of missing optimization runs received solution was nearly optimal, with Q_a equal to 4. When the stopping criteria is based only on quality index (there is no limits on objective functions calls count), GA always gives optimal solution, but often were happened, that computational cost of one optimization run was bigger than the rest 49 tries with same parameters values.

Population 10 - missing tries





Population 20 - missing tries





Population 50 - missing tries



Population 75 - missing tries



Population 100 - missing tries



Fig. 1. Missing tries (in percent of all executed)Rys. 1. Próby nietrafione (w procentach wszystkich prób wykonanych)

Fig. 2 presents mean objective function calls count for all combinations of investigated parameters.

Same investigations were processed for different crossover operators (15), but results were significantly worse than presented, received for one-point crossover. Analogical experiment was done for other type of genetic algorithm, so called steady-state, similar to the algorithms described by De Jong [1]. It uses overlapping populations with a 40% of overlap. Each generation the algorithm creates a temporary population of individuals, adds these to the previous population, then removes the worst individuals in order to return the population to its original size. Observed results are presented in Fig. 3

Population 10 - mean QF coun



Population 20 - mean QF count



Mean QF count

20000

0

0.2

0.3

0.4

0.5

Crossover

645

Population 35 - mean QF count



Population 50 - mean QF count

1200 1000

6008

6.2

Mutation

ð.15

6.1

0.05

~

Population 10 - mean QF count







Population 75 - mean QF count

0.9

0.8

0.6 0.7



Population 100 - mean QF count



Fig. 2. Mean objective function calls count

Rys. 2. Średnia liczba wywołań przedmiotowej funkcji

Population 35 - mean QF count



Population 50 - mean QF count





Fig. 3. Mean objective function calls count – steady state Rys. 3. Średnia liczba wywołań przedmiotowej funkcji – stan ustalony

Taking into account the comparison of both investigated types of genetic algorithms, one can see, that better and more stable results gives simple version. Steady-state version needs to have more powerful mutation "impulse", and better results in this version can be achieved using bigger population size. Same tests were processed for steady state version with very little overlapping (only 2 children for generation). The results in such case was even worse than with 40% overlapping. Finally, simple version was selected to the industrial application, based on better and in general more stable and robust for parameters variation, results. The values of key parameters were selected as follows: small population with 10 individuals, crossover probability equal to 0.4 and mutation probability equal to 0.05.

7. Industrial application

Proposed methodology was applied to vacuum furnace on hydrocracking plant in PKN Orlen. The desulphurized product of this installation is a feed for internal power plant. The main task of the furnace is the preparation of feed to the distillation tower (heating up crude oil and keeping constant temperature 384 °C). Constant temperature of hydrocarbon cracking on suitable level assures demanded quality of products. The simplified scheme of the plant is presented in Fig. 4.

The feed to the column is heated in four pipe coils (Pass A..D). Additionally, in convection area of the furnace low pressure steam (LP, Pass 1) and medium pressure (MP, Pass 2) steam are heated. Overheated low pressure steam is then added to each distillation towers C301, C302 and C303. MP steam is partially polled to factory-wide steam network, and partially added to crude oil, in order to increase its speed and reduce contaminate substances production.



Fig. 4. Installation scheme Rys. 4. Schemat instalacji

The main goal of plant control is to keep constant temperature at the furnace outlet. The temperature is controlled by cascade control of of fuel gas pressure (PC3035).

About 90 diagnostic tests (diagnostic signals) were defined during plant analysis. Most of them are based on models of process variables, or heuristic relations between physical values in the plant. Then an expert defines relation between diagnostic signals and possible faults, stored as diagnostic matrix. The meaning of columns and rows in the matrix is shown in Fig. 5:



Fig. 5. A part of diagnostic matrix. f- fault, R- model based diagnostic signal, L- heuristic diagnostic signal

Rys. 5. Część matrycy diagnostycznej. f – uszkodzenie, R – sygnał diagnostyczny wynikający z modelu, L – heurystyczny sygnał diagnostyczny

Full diagnostic matrix without signal and fault names is shown in Fig. 6. The goal was to find optimal decomposition to 3 subsystems with 30 diagnostic signals each. 20 independent runs of genetic algorithm were done. Best received quality index is equal to 9, and 28 solutions with such quality were found. The analysis of best solutions shown, that most of them differ in onetwo diagnostic signal, and are grouped into two decomposition "types" (Fig. 6):



Fig. 6. Diagnostic matrix for hydrocracking plant vacuum furnace Rys. 6. Macierz diagnostyczna dla instalcji pieca próżniowego

If ones decides to allow not strictly equal, but similar size of the subsystems (maximal size set to 35), the decomposition could be even better (quality index fall down to 6), and the number of optimal solutions rises up to more than 3500.

8. Summary

In the paper the applicability of genetic algorithms to diagnostic system decomposition was shown. In opposition to previous solutions based on graph analysis, proposed approach has many advantages. First of all, it allows to receive more than one optimal solution (with same quality index value). This feature gives a possibility to diagnostic system engineers to select such decomposition, which is most intuitive, or subsystems are most connected with technological components, without a loose mutual independence (decomposition quality).

The analysis of genetic algorithm parameters selection shows, that in case of greatly incontinuous, nonlinear, discrete objective function, faster convergence has an algorithm with small population size. Strong influence of randomly selected starting point for optimization convergence was also observed. This influence is big enough to recommend rather several GA runs with calculation-time based stopping criteria, than one, longer run.

9. References

- De Jong K.: Evolutionary Computation. A Unified Approach. The MIT Press; 1st edition 2002.
- [2] Goldberg D.: Genetic Algorithms in Search. Optimization, and Machine Learning. Addison-Wesley Professional 1989.
- [3] Kernighan B.W., Lin S.: An efficient heuristic procedure for partitioning graphs. Bell System Tech. J., 49, 1970.
- [4] Kościelny J.M.: Diagnostics of Processes in Decentralized Structures. Archives of Control Sciences, 1998, Vol, 7, No3/4.
- [5] Kościelny J.M., Sikora A.I.: Dekompozycja złożonych obiektów diagnozowania. Archiwum Automatyki i Telemechaniki, 1/1991, 115-131.
- [6] Kościelny J.M.: Diagnostyka ciągłych zautomatyzowanych procesów przemysłowych metodą dynamicznych tablic stanu. Prace Naukowe Politechniki Warszawskiej, seria Elektronika, z. 95. Warszawa 1991.
- [7] Kościelny J.M., Sikora A.I.: Chosen Problems of Complex Diagnostic System Decomposition. Intern. Conferen. Modelling & Simulation, Istambul, 1988, ASME Press, Vol. 1B, 135-142.
- [8] Kościelny J.M.: Diagnostics of industrial processes in decentralised structures. Chapter 19 in: Korbicz J., Kościelny J.M., Kowalczuk Z., Cholewa W. (ed).: Fault Diagnosis. Models, artificial intelligence, application. Springer, 2004.

otrzymano / received: 21.07.2010 przyjęto do druku / accepted: 04.05.2011

artykuł recenzowany

INFORMACJE

Newsletter PAK

Wydawnictwo PAK wysyła drogą e-mailową do osób zainteresowanych Newsletter PAK, w którym są zamieszczane:

- spis treści aktualnego numeru miesięcznika PAK,
- kalendarz imprez branżowych,
- ważniejsze informacje o działalności Wydawnictwa PAK.

Newsletter jest wysyłany co miesiąc do osób, które w jakikolwiek sposób współpracują z Wydawnictwem PAK (autorzy prac opublikowanych w miesięczniku PAK, recenzenci, członkowie Rady Programowej, osoby które zgłosiły chęć otrzymywania Newslettera).

Celem inicjatywy jest umocnienie w środowisku pozycji miesięcznika PAK jako ważnego i aktualnego źródła informacji naukowotechnicznej.

- Do newslettera można zapisać się za pośrednictwem:
- strony internetowej: www.pak.info.pl, po dodaniu swojego adresu mailowego do subskrypcji,
- adresu mailowego: wydawnictwo@pak.info.pl, wysyłając swoje zgłoszenie.

Otrzymywanie Newslettera nie powoduje żadnych zobowiązań ze strony adresatów. W każdej chwili można zrezygnować z otrzymywania Newslettera.