

**Łukasz TABOR**

INSTYTUT AUTOMATYKI I ROBOTYKI POLITECHNIKA WARSZAWSKA,  
ul. św. A. Boboli 8, 02-525 Warszawa

## Zmodyfikowana metoda budowy grafów przyczynowo-skutkowych procesów z wykorzystaniem archiwalnych przebiegów sygnałów

Mgr inż. Łukasz TABOR

Doktorant w Instytucie Automatyki i Robotyki na Wydziale Mechatroniki Politechniki Warszawskiej. Ukończył studia na kierunku Automatyka i Robotyka. W ramach prac naukowo-badawczych zajmuje się modelowaniem i diagnostyką procesów przemysłowych w szczególności interesują go metody odkrywania wiedzy w przemysłowych bazach danych.



e-mail: tabor.lukasz@gmail.com

### Streszczenie

W artykule opisano zmodyfikowaną metodę automatycznej budowy grafu przyczynowo-skutkowego procesu przemysłowego. Metoda składa się z etapów, w których między innymi analizowane są wskaźniki statystyczne oraz opóźnienia transportowe między sygnałami. W artykule poruszono kwestię wrażliwości metody na parametry oraz zaproponowano rozwiązanie mające zmniejszyć ten problem. Przedstawiono efekt działania algorytmu przed i po modyfikacji dla obiektu przemysłowego, będącego fragmentem sieci przesyłowej gazu ziemnego na dużym obszarze.

**Słowa kluczowe:** graf przyczynowo-skutkowy, diagnostyka procesów, modelowanie procesów, odkrywanie wiedzy z danych.

### Modified method for building cause-effect graphs of processes using archival signal values

#### Abstract

The paper describes a modified method for building the cause-effect graph of an industrial process. The algorithm consists of three steps. First, the dataset is searched for signals connected by well known relations such as the PID equation. Next, the algorithm searches the pairs of signals having high values of Kendall correlation coefficients. Additionally, the sequence of spreading disturbances is analysed [5]. The results of the basic algorithm performance are presented on example of a three tank system (Fig. 2) and a gas network (Figs. 4 and 5). In case of the gas network the cause-effect graph created using only the expert knowledge is shown in Fig. 3. In Section 3 the problem of the algorithm sensitivity is analysed (Figs. 4, 5 and 6). There is proposed a solution how to minimise the influence of selected parameters on the final result in Section 4. In order to be able to compare the results before and after modification, the distance measure between graphs [10, 11, 12] is suggested in Section 5. The described method sensitivity before and after modification on the example of the mentioned industrial object - a part of the gas network (Fig. 7) is discussed in Section 6. There is also given the best ( meaning the distance to the graph created by the expert) result of the algorithm for the gas network (Fig. 8).

**Keywords:** cause-effect graph, process diagnostics, process modelling, data mining.

### 1. Wstęp

Grafi przyczynowo-skutkowe są graficzną reprezentacją zależności pomiędzy wejściami (bodźcami) a odpowiadającymi im wyjściami (efektami). Znajdują one zastosowanie w różnorodnych dziedzinach życia. Przykładowo w medycynie używane są do określenia wpływu danych kuracji na stan pacjentów. W marketingu można reprezentować w ten sposób relacje pomiędzy promocjami, zachowaniem klientów i osiągniętymi wynikami finansowymi. Grafi tego typu znajdują także szerokie zastosowanie w zagadnieniach przemysłowych. Zbudowanie takiego modelu jakościowego instalacji przemysłowej może ułatwić budowę

modeli ilościowych przez pomoc w wyborze struktury modeli oraz doborze opóźnień sygnałów wejściowych.

Kolejnym przykładem zastosowania grafów jest problem analizy sekwencji występowania alarmów związanych z obiektami regulacji. Zazwyczaj jest tak, iż wejście instalacji w stan nienormalny, który powoduje wystąpienie pierwszego alarmu, powoduje także aktywowanie się kolejnych alarmów. Dla operatora informacja o aktywnych kilkunastu lub kilkudziesięciu alarmach jest informacją trudną do interpretacji. Do zadań projektantów systemów automatyki należy zatem odpowiednie filtrowanie alarmów i wskazywanie prawdziwego źródła wystąpienia sytuacji alarmowej [1].

Jeszcze inną dziedziną, w której omawiane grafi znajdują zastosowanie jest diagnostyka. Wiedza o jakościowych związkach pomiędzy sygnałami pomiarowymi może zostać wykorzystana do budowy modeli do detekcji uszkodzeń [2, 3] czy rekonfiguracji układu sterowania w przypadku wystąpienia awarii [4].

Procedura budowy grafu przyczynowo-skutkowego standardowymi metodami dla złożonych instalacji przemysłowych może być bardzo trudna i niezwykle czasochłonna. W chwili obecnej opiera się ona o wykorzystanie wiedzy eksperckiej. Specjalista w danej dziedzinie może wykorzystując swoją wiedzę oraz dokumentację techniczną obiektu opracować żądany graf. Pomimo tego, w niektórych sytuacjach ze względu na skomplikowanie zadania, może on nie być w stanie przewidzieć i określić wszystkich zależności.

Automatyczna budowa grafu na podstawie dostępnych w archiwach systemów DCS lub SCADA ciągów wartości zmiennych procesowych [5] ma zatem istotne znaczenie praktyczne. Istotne jest przy tym, że istnieją metody, umożliwiające weryfikację poprawności istniejącego grafu [6].

## 2. Algorytm budowy grafu

W artykule [5] zaproponowano metodę automatycznej budowy grafu przyczynowo-skutkowego opierającą się na trzech etapach. Są nimi: wyszukiwanie zależności o znanym charakterze (często występujących w obiektach przemysłowych), wyszukiwanie sygnałów zależnych oraz obserwacja propagacji zakłóceń w instalacji.

### 2.1. Etap 1 – zależności o znanym charakterze

Pierwszy z etapów polega na założeniu, że w wielu instalacjach występują standardowe elementy, których działanie jest powszechnie znane i opisane w literaturze. W szczególności przyjmuje się, że znamy postać równań wiążących wejścia i wyjścia takich elementów. Wystarczy zatem sprawdzić czy w bazie danych archiwalnych występuje zależność o takim charakterze aby móc powiązać spełniające ją sygnały. Przykładem takich zależności może być regulator PID czy zawór.

### 2.2. Etap 2 – zależności statystyczne

W kolejnym etapie dla wszystkich możliwych par sygnałów (z wykluczeniem tych, których związki określono w etapie pierwszym) badana jest wartość statystyki  $\tau$  Kendalla [7, 8, 9] w postaci:

$$\tau = \frac{P-Q}{P+Q+T}, \quad (1)$$

gdzie:

- $P$  – liczba par zgodnych (oba sygnały maleją lub oba rosną),
- $Q$  – liczba par niezgodnych (jeden sygnał maleje, drugi rośnie),
- $T$  – pary związane (pozostałe przypadki).

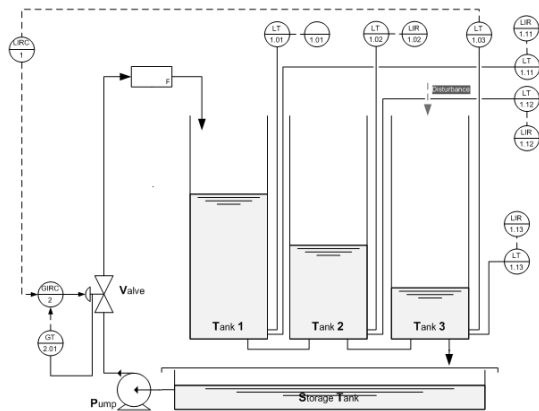
Następnie określa się graniczną wartość tejże statystyki (nazwaną wartością odcięcia) i stwierdza, że zależność zachodzi, jeżeli obliczony współczynnik jest od niej większy. Wybrana statystyka ma taką zaletę, iż wykrywa wszystkie zależności monotoniczne, a więc nie istotne jest czy są one liniowe czy nieliniowe. Na tym etapie nie uwzględniono jednak możliwej dynamiki w zachowaniu obiektów.

### 2.3. Etap 3 – analiza możliwych opóźnień transportowych

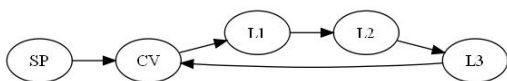
Etap trzeci polega na wyznaczeniu tej samej statystyki. Jednak tym razem wyznaczane jest przesunięcie jednego sygnału względem drugiego, dla którego wymienione współczynniki osiągają maksimum. Otrzymane wartości przesunięć mówią o opóźnieniach transportowych pomiędzy sygnałami.

## 3. Wyniki działania algorytmu

Wspomniany artykuł [5] przedstawia wynik uruchomienia algorytmu dla dwóch instalacji. Pierwszym jest układ trzech zbiorników (rys. 1.), w którym ciecz dostarczana jest za pomocą pompy do zbiornika pierwszego. Następnie przepływa, pod wpływem grawitacji, poprzez zbiornik drugi oraz trzeci i opuszcza układ. Układ regulacji sterując wydajnością pompy utrzymuje poziom cieczy w zbiorniku trzecim na zadanej wartości. Wynikiem algorytmu dla tego obiektu jest graf z rys. 2.



Rys. 1. Układ trzech zbiorników  
Fig. 1. Three tank system



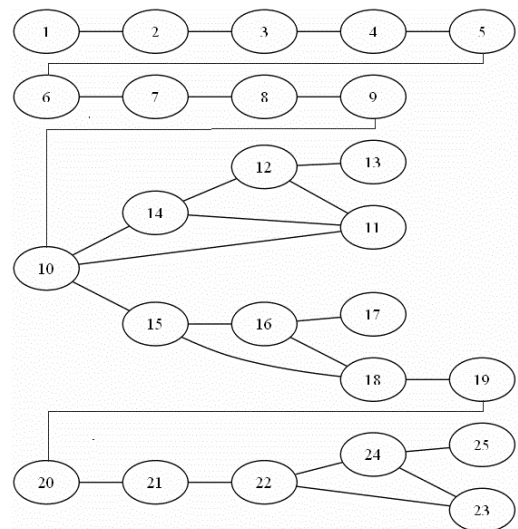
Rys. 2. Postać grafu przyczynowo-skutkowego dla układu trzech zbiorników  
Fig. 2. Cause-effect graph for three tank system

Kolejnym obiektem jest fragment sieci przesyłającej gaz na pewnym obszarze Polski. Fragment ten składa się z rurociągu głównego wraz z odnogami oraz stacji poboru gazu, w których dokonywany jest pomiar ciśnienia i przepływu. Wraz ze zmianami zapotrzebowania na gaz w sieci widoczne są zmiany ciśnienia.

By zwiększyć przejrzystość prezentowanych grafów sygnały pomiarowe ciśnień w kolejnych stacjach oznaczono numerami. Zestawienie oznaczeń przedstawia tabela 1. Możliwy do otrzymania, w wyniku działania algorytmu graf powinien być podobny do grafu sporządzonego przez eksperta przedstawionego na rys. 3.

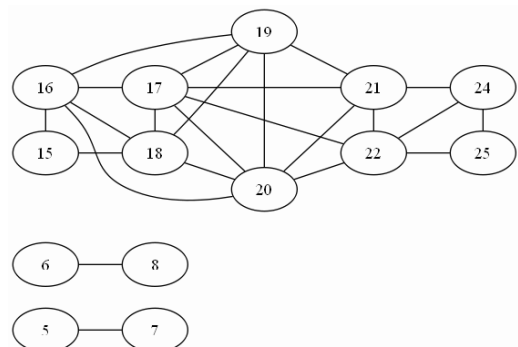
Tab. 1. Numeracja sygnałów pomiarowych  
Tab. 1. Numeration of measurement signals

Ciśnienie w stacji	nr	Ciśnienie w stacji	nr
P_WolkaR	1	P_WysokieMaz	14
P_Radzymin	2	P_Sulecin	15
P_Trojany1	3	P_Przyborowie	16
P_Niegow	4	P_Wysocze	17
P_Drogoszewo	5	P_Czerwin	18
P_Rybno	6	P_Troszyn	19
P_Wyszkow	7	P_Rzekun	20
P_Leszczydol	8	P_Ostroleka2	21
P_Branszczyk	9	P_Otok	22
P_Poreba	10	P_Baranowo	23
P_OstrowMaz	11	P_Lelis	24
P_Zambrow	12	P_Kadzidlo	25
P_StLomza	13		

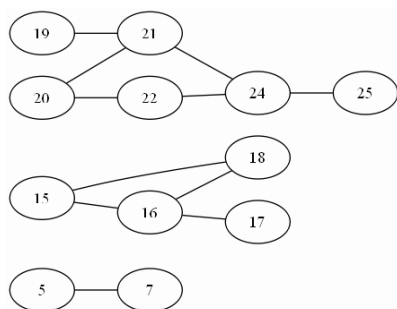


Rys. 3. Graf przyczynowo-skutkowy sieci gazowej powstały przy wykorzystaniu wiedzy eksperckiej  
Fig. 3. Cause-effect graph for gas network created with use of expert knowledge

W analizowanej instalacji ze względu na występowanie jedynie pomiarów ciśnień przeprowadzono tylko etapy drugi i trzeci algorytmu. W zależności od przyjętej granicznej wartości współczynnika korelacji otrzymano dość różne wyniki (rys. 4 i 5).



Rys. 4. Wynik działania algorytmu podstawowego dla wsp. odcięcia 0.96  
Fig. 4. Result of basic algorithm performance for cut-off coefficient 0.96

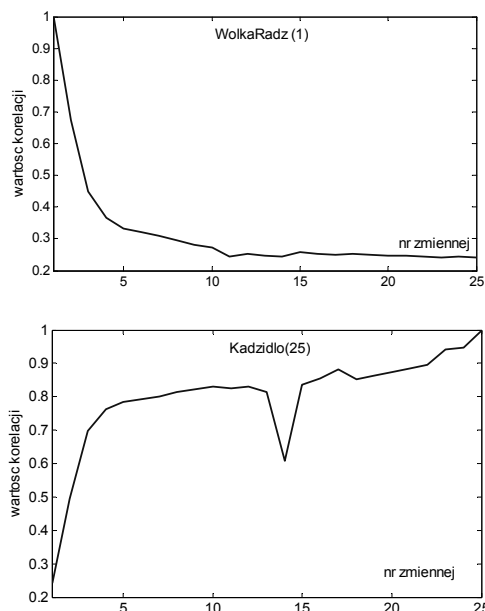


Rys. 5. Wynik działania algorytmu podstawowego dla wsp. odcięcia 0.98  
Fig. 5. Result of basic algorithm performance for cut-off coefficient 0.98

#### 4. Modyfikacja metody

Jak pokazał przykład sieci gazowej wynik działania algorytmu bardzo zależy od przyjętej wartości współczynnika odcięcia. Jeżeli jego wartość zostanie ustalona na zbyt wysokim poziomie, graf wynikowy będzie składał się z wielu części pomimo tego, że opisuje jeden obiekt przemysłowy. Z kolei zbyt niska jego wartość spowoduje otrzymanie w wyniku zbyt wielu powiązań.

Analiza wartości współczynników korelacji sygnałów ciśnienia w sieci gazowej pokazała, iż dla danego punktu pomiarowego maksymalne skorelowanie uzyskuje się na ogół z sąsiednimi punktami. Jednak wartość tej maksymalnej korelacji oraz szybkość jej zmniejszania różnią się znacznie w zależności od położenia punktu pomiarowego w sieci (rys. 6). Widać, że zmienna P\_WolkaRadz (nr 1) ma największą korelację ze zmienną nr 2 i jest ona równa zaledwie ok. 0,45. Reszta współczynników plasuje się na poziomie 0,2-0,3. Natomiast zmienna P\_Kadzidło (nr 25) jest najbardziej skorelowana ze zmienną nr 24 (w stopniu ~0,97), lecz dla całego szeregu zmiennych współczynnik korelacji także jest równie wysoki. Przyjęcie zatem jednej granicznej wielkości korelacji jest niemożliwe.



Rys. 6. Współczynniki korelacji sygnałów P\_WolkaRadz (nr 1) i P\_Kadzidło (nr 25) z innymi sygnałami w archiwum

Fig. 6. Correlation coefficients for signals P\_WolkaRadz (nr 1) i P\_Kadzidło (nr 25) with other signals in dataset

Zdecydowano się zatem, dla każdego sygnału niezależnie, związać wartość odcięcia z maksymalnym współczynnikiem korelacji pomiędzy tą zmienną a wszystkimi innymi pozostałymi sygnałami oraz z rozrzutem wartości tychże korelacji dla tej zmiennej zgodnie z wzorem:

$$w_i = \tau_{i,max} - k\bar{\sigma}_i, \quad (2)$$

gdzie:

- $w_i$  – współczynnik odcięcia dla  $i$ -tej zmiennej,
- $\tau_{i,max}$  – maksymalna wartość korelacji pomiędzy  $i$ -tą zmienną a pozostałymi,
- $\sigma_i$  – odchylenie standardowe korelacji dla  $i$ -tej zmiennej,
- $k$  – parametr.

Zaproponowana formuła obliczania współczynnika odcięcia zapewnia to, że w otrzymanym grafie będzie brak wierzchołków izolowanych. Zmieniając parametr  $k$  można uzyskiwać grafy o różnej liczbie krawędzi. Sposób doboru wartości  $k$  nie jest w tym momencie określony. Modyfikacja metody miała także za zadanie zmniejszyć wrażliwość algorytmu na wybór parametrów. Aby stwierdzić czy otrzymywane wyniki dla metody zmodyfikowanej są lepsze od wyników metody podstawowej oraz porównać wrażliwość obu metod na dobór parametrów należało przyjąć miarę dopasowania grafów.

#### 5. Miara dopasowania grafów

Problem porównywania grafów jest obecny w wielu zagadnieniach. Największe zastosowanie znajdują metody stosowane w zagadnieniach porównywania obrazów, analizowania zawartości sieci WWW czy odnajdywania wzorców w danych [10, 11]. Większość z nich opiera się na trzech podejściach: maksymalnego wspólnego podgrafu, odległości edycji i miary zachowania odległości [12]. Ze względu na łatwość implementacji i małą złożoność obliczeniową zdecydowano się na ostatnie z wspomnianych podejść, przy czym dokonano pewnych modyfikacji. Zastosowana miara odległości opisana zależnością (3) posiada trzy składniki: kara za różną liczbę składowych spójności w grafach, kara za występowanie relacji błędnych (proporcjonalną do odległości między wierzchołkami w grafie wzorcowym) oraz nagroda za występowanie relacji właściwych (-1 za każdą relację poprawną). Miara jest standaryzowana wielkością zadania (liczność relacji w grafie wzorcowym) oraz zwiększona o 1 aby minimum miary wynosiło 0.

$$d = \frac{\sum_{i=1}^n \sum_{j=1}^n (D(i, j) \cdot M(i, j)) + \bar{D}(S - S_w)}{|E_w|} + 1, \quad (3)$$

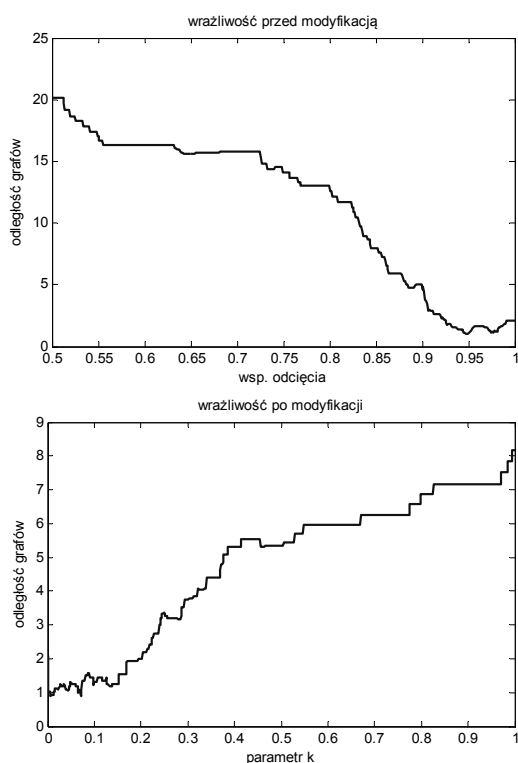
gdzie:

- $d$  - odległość grafów,
- $D$  - macierz odległości między wierzchołkami w grafie wzorcowym (jeżeli dla wierzchołków sąsiednich -1, dla pozostałych równa liczbie krawędzi najkrótszej ścieżki między wierzchołkami)
- $M$  - macierz sąsiedztwa grafu wynikowego,
- $S$  - liczba składowych w grafie wynikowym,
- $S_w$  - liczba składowych w grafie wzorcowym,
- $\bar{D}$  - średnia odległość między wierzchołkami w grafie wzorcowym
- $|E_w|$  - liczba krawędzi w grafie wzorcowym.

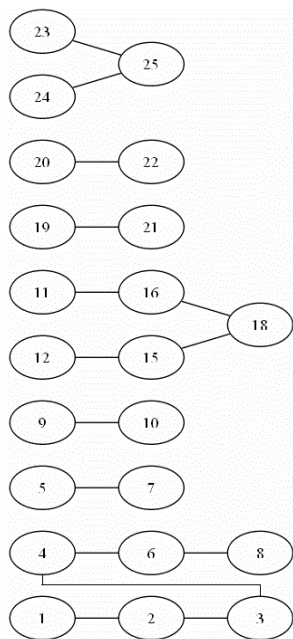
#### 6. Wyniki dla zmodyfikowanego algorytmu

Weryfikacja zmodyfikowanej metody przebiegła z wykorzystaniem tego samego fragmentu sieci jak poprzednio. Z uwagi jednak na to, iż dane pochodzące z systemu SCADA zainstalowanego na instalacji przemysłowej posiadały liczne braki związane z włączaniem i wyłączaniem poszczególnych stacji oraz miały miejsce częste kalibracje czujników zdecydowano się dodatkowo wykorzystać model tego odcinka sieci gazowej [13]. Dzięki temu uzyskano pełniejszy zbiór zmiennych procesowych do analizy. Wyniki algorytmu, w postaci odległości od grafu wzorcowego (otrzy-

manego na podstawie wiedzy eksperckiej), przed i po modyfikacji przedstawia rys. 7. W teście zmieniano parametry algorytmu w całym zakresie, w którym można się było spodziewać sensownych wyników. Minimalne otrzymane odległości zawiera tabela 2. Graf będący wynikiem algorytmu zmodyfikowanego przedstawia rys. 8.



Rys. 7. Wrażliwość algorytmów na zmianę parametrów przed i po modyfikacji  
Fig. 7. Sensitivity of algorithms to change of parameters before and after modification



Rys. 8. Wynik działania algorytmu zmodyfikowanego dla sieci gazowej  
Fig. 8. Result of modified algorithm performance for gas network

Tab. 2. Odległość d grafów względem grafu wzorcowego (najlepszy przypadek)  
Tab. 2. Distance d between graph and pattern (best case)

Przed modyfikacją	Po modyfikacji
1,04	0,89

## 7. Podsumowanie

Zaprezentowana poprzednio metoda automatycznej budowy grafów przyczynowo-skutkowych procesów przemysłowych wymagała ulepszenia. Szczególnym problemem był dobór współczynnika odcięcia wykorzystywanego w etapie drugim i trzecim. Zaproponowano modyfikację polegającą na ustaleniu tego współczynnika dla każdej zmiennej niezależnie. Współczynnik jest wyliczany na podstawie korelacji danej zmiennej ze wszystkimi innymi. Modyfikacja taka pozwoliła na polepszenie jakości otrzymanych grafów. Istotną cechą jest także zmniejszenie wrażliwości algorytmu na dobór parametrów.

*Niniejsza praca była wykonana częściowo w ramach grantu Narodowego Centrum Nauki „Metody rozwiązania podstawowych problemów rozpoznawania uszkodzeń w złożonych obiektach dynamicznych”.*

## 8. Literatura

- [1] Gandhi O.P., Agrawal, V.P.: Failure cause analysis – a structural approach, Journal of Pressure Vessel Technology, Transactions of the ASME, 4/1996, 434-439.
- [2] Ostasz A., Kościelny J.M.: Zastosowanie grafu procesu GP w diagnostyce procesów przemysłowych. Pomiary Automatyka Kontrola nr 9bis/2005, str. 51-53.
- [3] Ostasz A.: Graf przyczynowo-skutkowy procesu i jego zastosowanie do projektowania zbioru residuów oraz relacji diagnostycznej. Politechnika Warszawska, 2007.
- [4] Garcia-Beltran C., Gentil S.: Casual reasoning for human supervised process reconfiguration. IEEE International Symposium on Intelligent Control - Proceedings, 2001,91-96.
- [5] Tabor Ł.: Budowa grafów przyczynowo-skutkowych z wykorzystaniem archiwalnych przebiegów sygnałów, Pomiary Automatyka Robotyka PAR, nr 11/2010, str. 65-68.
- [6] Palmer C., Chung P.: Verifying Signed Directed Graph Models for Process Plants. Computers and Chemical Engineering Supplement 1999, str. 391-394.
- [7] Koronacki J., Mielniczuk J.: Statystyka dla studentów kierunków technicznych i przyrodniczych. WNT, Warszawa 2007.
- [8] Krzyśko M.: Systemy uczące się: rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości. WNT, Warszawa 2008.
- [9] Larose D. T. Metody i modele eksploracji danych. PWN, Warszawa 2008.
- [10] Bunke H., Shearer K.: A graph distance metric based on the maximal common subgraph. Pattern Recognition Letters, 19(3):255–259, 1998.
- [11] Luo B., Hancock R.: Structural graph matching using the EM algorithm and singular value decomposition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(10):1120–1136, 2001.
- [12] Schenker A., Bunke H., Last M., Kandel A. Graph-theoretic techniques for web content mining. Series in Machine Perception and Artificial Intelligence, 62, 2005.
- [13] Stachura M. i Syfert M. (2010). Model sieci gazowej w systemie monitorowania i diagnostyki AMandD. Pomiary Automatyka Robotyka PAR, nr 11/2010, str. 110-115.

otrzymano / received: 26.09.2011

przyjęto do druku / accepted: 01.12.2011

artykuł recenzowany / revised paper