

Budowa grafów przyczynowo-skutkowych procesów z wykorzystaniem archiwalnych przebiegów sygnałów

Łukasz Tabor

Instytut Automatyki i Robotyki Politechniki Warszawskiej

W artykule opisano metodę automatycznego tworzenia grafu przyczynowo-skutkowego procesu przemysłowego. Metoda polega na wyszukiwaniu w całym zbiorze zmiennych procesowych takich sygnałów, które związane są pewnymi znanymi zależnościami często występującymi w instalacjach przemysłowych (np. algorytm regulacji PID). Następnie poszukiwane są pary sygnałów o wysokiej wartości współczynników korelacji liniowej Pearsona oraz monotonicznej Spearmana i Kendalla między zmiennymi. Dodatkowo analizowana jest sekwencja rozchodzenia się zaburzeń w instalacji. Na podstawie zdobytych informacji tworzony jest graf przyczynowo-skutkowy opisujący obiekt lub jego część. Przedstawiono efekt działania algorytmu dla dwóch obiektów. Pierwszy składa się z trzech zbiorników z cieczą z regulacją poziomu w jednym z nich. Drugim jest fragment sieci przesyłowej transportującej gaz ziemny na dużym obszarze.

Słowa kluczowe: graf przyczynowo-skutkowy, diagnostyka, modelowanie procesów, obiekt liniowy, obiekt nieliniowy

O potrzebie budowy grafów przyczynowo-skutkowych

Grafy przyczynowo-skutkowe są graficzną reprezentacją zależności między wejściami (bodźcami) a odpowiadającymi im wyjściami (efektami). Informacje o występowaniu takich zależności mogą być przydatne w wielu zagadnieniach praktycznych. Na przykład, w procesie budowy modeli obiektów przemysłowych pierwszym etapem jest dobór odpowiedniej struktury, w szczególności wybranie spośród dostępnych sygnałów pomiarowych tych, które użyte zostaną jako wejścia powstającego modelu. Osoby zajmujące się takimi zagadnieniami często zmuszone są kierować się intuicją bądź sięgać po pomoc innych osób specjalizujących się w działaniu tej konkretnej instalacji, której model należy uzyskać.

Kolejnym przykładem zastosowania grafów jest analiza sekwencji występowania alarmów związanych z obiektami regulacji. Zazwyczaj jest tak, iż wejście instalacji w stan, który powoduje wystąpienie pewnego stanu alarmowego, powoduje także aktywowanie się innych alarmów w chwili bądź niedługo po wystąpieniu tego pierwszego. Dla operatora z kolei informacja o aktywnych kilkunastu lub kilkudziesięciu alarmach jest informacją mało przydatną. Do zadań projektantów systemów automatyki należy zatem odpowiednie filtrowanie takich sytuacji i wskazanie prawdziwego źródła wystąpienia sytuacji alarmowej [1].

Jeszcze inną dziedziną, w której omawiane grafy znajdują zastosowanie jest diagnostyka. Wiedząc o występowaniu zależności między sygnałami pomiarowymi oraz o miejscu ujawniania się wpływu uszkodzeń, można budować modele diagnostyczne [2, 3]. Informacja taka może być także przydatna w celu rekonfiguracji układu sterowania w przypadku wystąpienia awarii [4].

Procedura budowy grafu przyczynowo-skutkowego standardowymi metodami, w zależności od wielkości instalacji, może być bardzo trudna i niezwykle czasochłonna. W chwili obecnej proces ten opiera się na wykorzystaniu wiedzy eksperckiej. Specjalista w danej dziedzinie może opracować żądany graf wykorzystując swoją wiedzę oraz dokumentację technicz-

ną obiektu. Pomimo tego w niektórych sytuacjach, ze względu na skomplikowanie zadania, może on nie być w stanie przewidzieć i określić wszystkich zależności. W każdym przypadku proces taki jest dość pracochłonny. Jednocześnie na wspólnych instalacjach przemysłowych działają systemy sterowania i wizualizacji procesów wykorzystujące bazy danych do przechowywania archiwalnych przebiegów sygnałów pomiarowych. Wiedza zawarta w takich bazach pozostaje na ogół niewykorzystana. Zawiera natomiast informacje o zależnościach potrzebnych w budowie grafu przyczynowo-skutkowego. Opisująca metoda ma za zadanie umożliwić automatyzację procesu tworzenia grafu za pomocą danych historycznych.

Algorytm budowy grafu

Proponowana metoda składa się z trzech etapów: wyszukiwanie zależności o znanym charakterze (często występujących w obiektach przemysłowych), wyszukiwanie sygnałów zależnych oraz obserwacja propagacji zakłóceń w instalacji.

Zależności o znanym charakterze

Każda instalacja przemysłowa składa się z bardzo dużej liczby elementów. Wiele z nich to standardowe urządzenia, których zasada działania jest powszechnie znana. Dostępne są również równania opisujące powiązania sygnałów wejściowych i wyjściowych takich urządzeń. Skoro znamy charakter tych zależności, to można przeszukać posiadaną bazę danych i w przypadku znalezienia sygnałów spełniających daną równość sądzić, że sygnały te opisują takie urządzenie. Przykładem dobrze znanego i powszechnie występującego związku jest powiązanie strumienia cieczy przepływającego przez zawór oraz różnicy ciśnień na jego końcach i stopnia otwarcia. Innym przykładem wyjaśniającym algorytm postępowania na tym etapie budowy grafu jest regulator typu PID. Sygnał wyjściowy (CV) idealnego regulatora w chwili t jest związany, zgodnie z zależnością (1), z różnicą między wartością zadaną (SP) i wielkością regulowaną (PV).

$$CV(t) = K_p e(t) + \frac{K_p}{T_i} \int_0^t e(\tau) d\tau + K_p T_d \frac{de}{dt} \quad (1)$$

$$\text{gdzie:} \quad e(t) = SP(t) - PV(t) \quad (2)$$

Algorytm w tym przypadku losuje z dostępnego zbioru trzy sygnały. Następnie odejmuje od siebie dwa z nich, tworząc zmienną e . Kolejnym krokiem jest wyznaczenie pomocniczych sygnałów e_1 oraz e_2 , z których pierwszy jest całką, a drugi pochodną po czasie z e (de/dt). Algorytm przeprowadza test regresji wielowymiarowej [6], gdzie zmienną objaśnianą jest nieużyty do tej pory sygnał z wybranej pierwotnie trójki, a zmiennymi objaśniającymi e , e_1 , e_2 . W przypadku występowania zależności liniowej przyjmuje się, że w danym miejscu instalacji umieszczony jest regulator PID. W przeciwnym – powtarza się algorytm od momentu obliczenia odchyłki e , wybierając inną kombinację sygnałów z wybranej trójki lub nową trójkę, aż do przeszukania całej przestrzeni dostępnych zmiennych.

Wyszukiwanie sygnałów zależnych probabilistycznie

Po odnalezieniu zależności o znanym charakterze następuje poszukiwanie par zmiennych o wysokich współczynnikach zależności probabilistycznych. Podejście to wynika z faktu, że prawdopodobieństwo istnienia współzależności między sygnałami pochodzącymi z „bliskich” sobie części instalacji jest większe niż dla odległych od siebie części procesu. Jako miarę podobieństwa sygnałów wykorzystano trzy miary zależności probabilistycznej.

1. Współczynnik korelacji Pearsona [5, 6, 7]:

$$P_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3)$$

gdzie: x, y – badane sygnały, n – długość wektorów x, y .

Miara ta jest wrażliwa na występowanie sygnałów liniowo zależnych. W związku z tym, iż w rzeczywistych obiektach zazwyczaj mamy do czynienia z zależnościami nieliniowymi, skorzystano z podanego niżej współczynnika.

2. Współczynnik korelacji rang Spearmana [5, 6, 7]:

S – liczony wg wzoru (3) z tą różnicą, że zamiast wartości sygnałów x_i, y_i wstawia się rangi będące indeksem danej obserwacji po uszeregowaniu konkretnego sygnału od wartości najmniejszych do największych.

Miara taka jest bardziej odporna od poprzedniego współczynnika na nieliniowości, a także łatwiej radzi sobie z obserwacjami odstającymi. Mimo to wykorzystano jeszcze trzecią miarę.

3. Współczynnik K Kendalla [5, 6, 7]:

$$K = \frac{P - Q}{P + Q + T} \quad (4)$$

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).

Miara ta jest wrażliwa na każdą zależność, która jest monotoniczna. Założenie takie jest spełnione w większości zależności występujących dla obiektów przemysłowych.

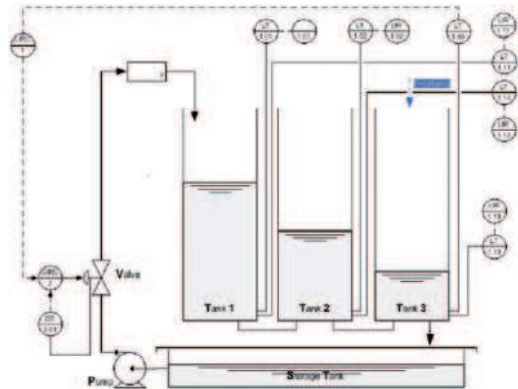
Dla każdej miary ustalana jest progowa jej wartość, powyżej której algorytm stwierdza występowanie zależności.

Obserwacja propagacji zakłóceń

Ten etap w dużej mierze związany jest z poprzednim, gdyż wykorzystuje te same miary. Jednak tym razem wyznaczane jest przesunięcie jednego sygnału względem drugiego, dla którego wymienione współczynniki osiągają maksimum. Otrzymane wartości przesunięć mówią o opóźnieniach transportowych między sygnałami. Sposób, w jaki wyznaczone opóźnienia wykorzystywane są w budowie grafu, omówiony zostanie na przykładzie.

Przykład 1 – układ trzech zbiorników

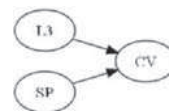
Działanie algorytmu zostanie pokazane na przykładzie prostego obiektu, jakim jest układ trzech zbiorników z cieczą (rys. 1). Zbiorniki połączone są kanałami umożliwiającymi przepływ cieczy między nimi. Trzeci zbiornik ma kanał odprowadzający. Do pierwszego zbiornika dostarczana jest ciecz za pomocą pompy i zaworu regulacyjnego. Układ ma za zadanie regulować poziom w trzecim zbiorniku. Do tego celu wykorzystano regulator PID. Na stanowisku są dostępne oraz archiwizowane pomiary poziomu cieczy w zbiornikach L1, L2, L3, wyjście z regulatora CV oraz wartość zadana poziomowi SP.



Rys. 1. Układ trzech zbiorników

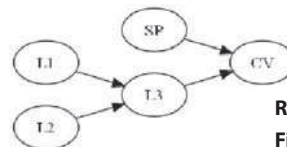
Fig. 1. The three tank system

Dla zaprezentowania działania algorytmu przeprowadzono symulację normalnej pracy instalacji zmieniając wartość zadaną. Zarchiwizowane przebiegi pięciu wspomnianych sygnałów poddano działaniu algorytmu w celu uzyskania grafu. Pierwszy etap algorytmu wykrył istnienie zależności charakterystycznej dla regulatora PID między sygnałami L3, SP oraz CV. Na tej podstawie powstał załazek grafu widoczny na rys. 2.



Rys. 2. Graf po etapie 1

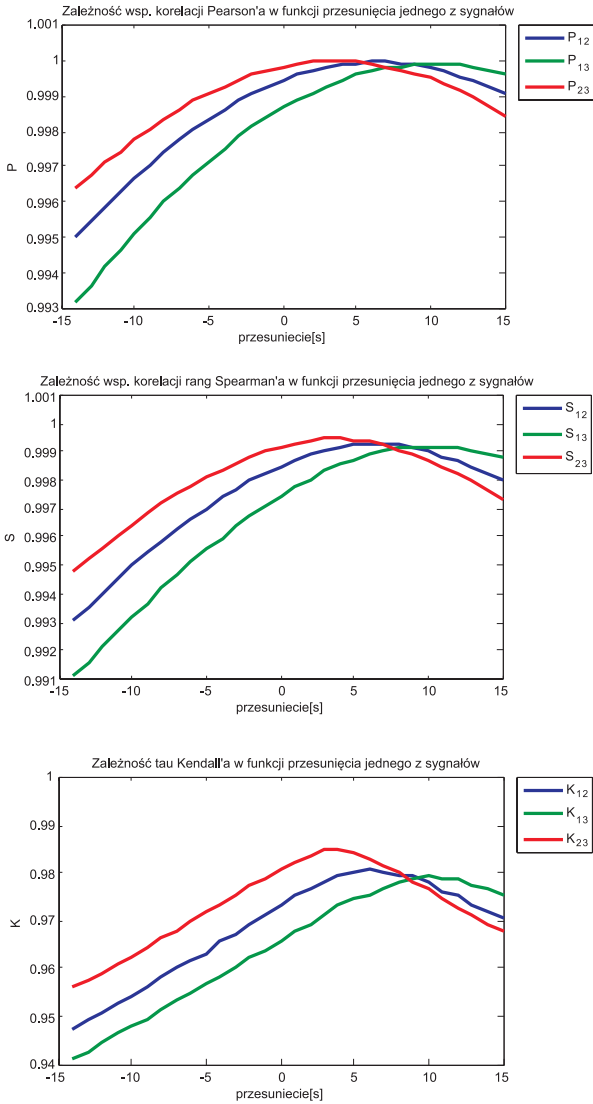
Fig. 2. The graph after stage 1



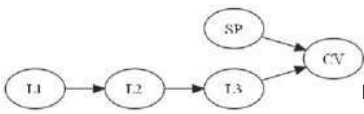
Rys. 3. Kolejny etap powstawania grafu

Fig. 3. The next stage of graph formation

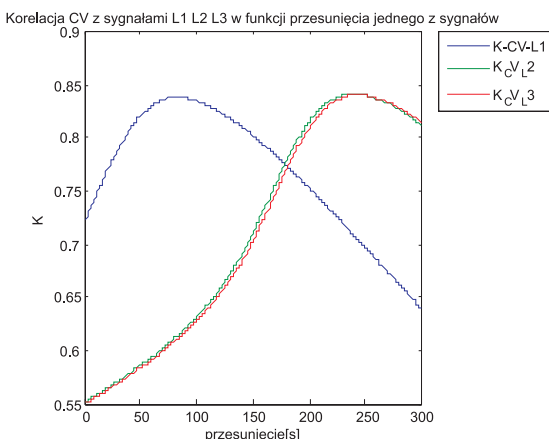
W kolejnym kroku zbadano wartości miar zależności probabilistycznych wymienionych w poprzednim rozdziale. Wykryto, że pary (L1, L3), (L2, L3) charakteryzują się dużą wartością wszystkich współczynników. Rozbudowano zatem graf do po-



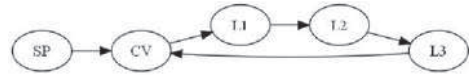
Rys. 4. Miary P, S i K w funkcji przesunięcia sygnałów względem siebie
Fig. 4. The P, S and K measures as a function of signals shift with respect to each other



Rys. 5. Graf po etapie 3
Fig. 5. The graph after stage 3



Rys. 6. Wyznaczenie opóźnień względem CV
Fig. 6. Determination of the delays time with respect to CV



Rys. 7. Ostateczna postać grafu przyczynowo-skutkowego dla układu trzech zbiorników

Fig. 7. The final form of the causal graph for the three tank system

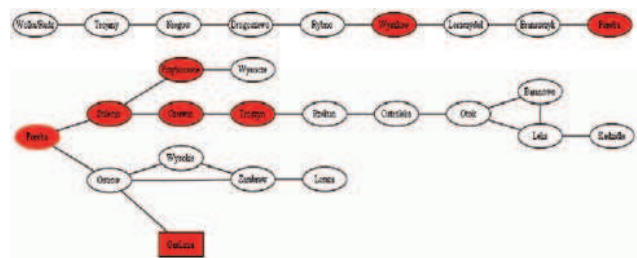
stacji z rys 3. Jednocześnie wykryto zależność w parze (L1, L2). Zbadano możliwe opóźnienia między wspomnianymi parami sygnałów. Zachowanie się współczynników korelacji w funkcji przesunięcia sygnałów przedstawia rys. 4.

Odczytano przesunięcia, dla których występowało maksimum miar zależności. Okazało się, iż opóźnienie między sygnałami L1 i L3 jest sumą opóźnień między L1 a L2 i L2 a L3. Z tego powodu zmodyfikowano poprzedni graf do postaci z rys. 5.

Do zakończenia procesu tworzenia grafu pozostało jeszcze określenie, na który sygnał oddziałuje w pierwszej kolejności sygnał CV. Do tej pory stwierdzono jedynie, iż musi być to któryś z sygnałów mających wpływ na L3 bądź bezpośrednio L3. W tym celu wyznaczono opóźnienia między sygnałem CV a L1, L2, L3. Efekt przedstawia rys. 6. Odczytano, iż w pierwszej kolejności na zmianę sygnału CV reaguje L1. Graf musi mieć zatem postać z rys. 7.

Przykład 2 – sieć gazowa

Przykładem obiektu przemysłowego, na którym testowano działanie algorytmu jest fragment sieci przesyłającej gaz na dużym obszarze Polski. Fragment ten obejmuje rurociąg główny wraz z odnogami, które zakończone są osiemnastoma stacjami poboru gazu. W większości z takich stacji dokonywane są pomiary takich parametrów gazu, jak przepływ, ciśnienie oraz temperatura. Niestety w niektórych stacjach nie wszystkie pomiary są realizowane. Do budowy grafu postanowiono wykorzystać pomiary ciśnienia. Schemat sieci z zaznaczonymi na czerwono stacjami, w których pomiar ciśnienia nie był rejestrowany przedstawiono na rys. 8.

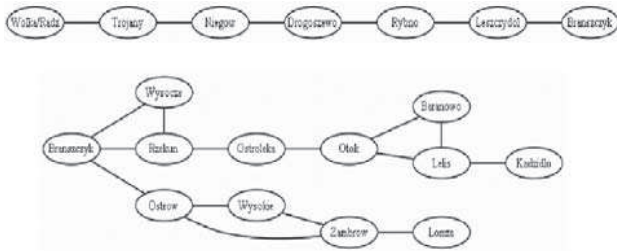


Rys. 8. Schemat sieci gazowej – na czerwono zaznaczono stacje, w których brakuje pomiaru ciśnienia

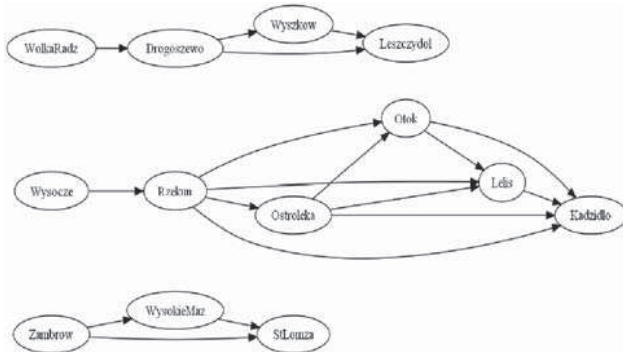
Fig. 8. The schematic diagram of gas network – stations without pressure measurement are marked with red color

Po uwzględnieniu braków sygnałowych w wyniku działania algorytmu należałoby się spodziewać otrzymania grafu przedstawionego na rys. 9.

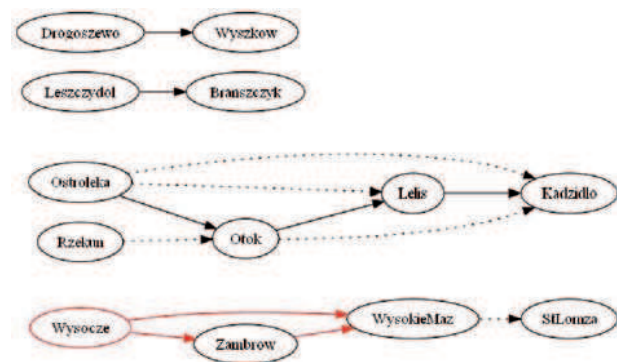
W trakcie realizacji algorytmu dla tego przykładu okazało się, jak niewłaściwym punktem metody może być dobór właściwych progów miar probabilistycznych w etapie drugim. Uzyskane grafy dla różnych wartości tych współczynników przedstawiono na rys. 10 i 11.



Rys. 9. Graf sieci sporządzony z wykorzystaniem wiedzy eksperckiej
Fig. 9. The network graph designed by using the experts knowledge



Rys. 10. Wynik działania algorytmu dla współczynnika odcięcia 0,95
Fig. 10. The result of the algorithm for 0,95 cutoff value



Rys. 11. Wynik działania algorytmu dla współczynnika odcięcia 0,97. Na czarno zaznaczono zależności prawidłowo wykryte, na czerwono nieprawidłowo sklasyfikowane, linią kropkowaną relacje bliskie prawdziwie

Fig. 11. The result of the algorithm for 0,97 cutoff value. The dependence correctly detected are marked with black color, incorrectly classified by red color and closely match to the true relationship by dotted line

Podsumowanie

Grafy przyczynowo-skutkowe obiektów przemysłowych mogą stanowić istotne udogodnienie w pracy specjalistom w wielu dziedzinach. Obecnie proces tworzenia grafu jest żmudny i kosztowny. Istnieje natomiast możliwość wydobycia informacji o zachodzeniu zależności między zmiennymi z historycznych przebiegów sygnałów pomiarowych. Metoda, jak pokazały przykłady, może prowadzić do obiecujących wyników, jednak wymaga dalszych prac. Kwestią priorytetową jest opracowanie inteligentnego procesu doboru krytycznych wartości miar zależności dla każdego węzła. Innym problemem do pokonania jest możliwość występowania w obiekcie sygnałów dyskretnych.

Zaprezentowane w artykule podejście można rozwinąć wykorzystując wiedzę zawartą w dokumentacji technicznej bądź oznaczeniach zmiennych procesowych. Rozważać można niekiedy wykorzystanie eksperymentów czynnych do obserwacji rozchodzenia się zaburzeń w obiekcie.

Bibliografia

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.: *Graf przyczynowo-skutkowy procesu i jego zastosowanie do projektowania zbioru residuów oraz relacji diagnostycznej*. Politechnika Warszawska, 2007.
3. Ostasz A., Kościelny J.M.: *Zastosowanie grafu procesu GP w diagnostyce procesów przemysłowych*. Pomiary Automatyka Kontrola nr 9bis/2005, 51–53.
4. Garcia-Beltran C., Gentil S.: *Casual reasoning for human supervised process reconfiguration*. IEEE International Symposium on Intelligent Control – Proceedings, 2001, 91-96.
5. Koronacki J., Mielniczuk J.: *Statystyka dla studentów kierunków technicznych i przyrodniczych*. WNT, Warszawa 2007.
6. Larose D. T. *Metody i modele eksploracji danych*. PWN, Warszawa 2008.
7. Krzyżko M.: *Systemy uczące się: rozpoznawanie wzorców, analiza skupień i redukcja wymiarowości*. WNT, Warszawa 2008.

Building cause-effect graph of processes with use of archival signal values

The paper presents cause-effect graphing method applied to industrial process. Method is based on searching in process dataset such signals, that are tied together by relations often occurring in industrial installations (for example PID algorithm). Subsequently algorithm is looking for pairs of signals having high values of Pearson, Spearman and Kendall correlation coefficients. Additionally sequence of spreading disturbances is analyzed. On the basis of the gathered information case-effect graph that describes analyzed object or its part is build. Effect of described method is presented with two example industrial objects. First is a three tank system. Second is a part of gas network.

Keywords: cause-effect graph, diagnostics, process modeling, linear system, nonlinear system

mgr inż. Łukasz Tabor

Jest doktorantem 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.

e-mail: tabor.lukasz@gmail.com

