



**INNOWACYJNA
GOSPODARKA**
NARODOWA STRATEGIA SPÓJNOŚCI

Investujemy
w Waszą
przyszłość



INSTEPRO
Zintegrowane
Sterowanie
Produkcją

UNIA EUROPEJSKA
EUROPEJSKI FUNDUSZ
ROZWOJU REGIONALNEGO



Raport wewnętrzny projektu InStePro

Nr 4.2: Projekt podsystemu eksploracji danych procesowych

Data

31.09.2010

Przygotował zespół:

J.T. Duda
T. Pełech-Pilichowski

Działanie algorytmów detekcji zdarzeń w sygnałach diagnostycznych bazuje na:

1. Analizach o charakterze statystycznym współzależności zmiennych procesowych
2. Analizach oparte o symulacje procesów

Ad.1. Analizy o charakterze statystycznym współzależności zmiennych procesowych

1.1. Analizy statystyczne

Podejście statystyczne wymaga przede wszystkim selekcji podciągów odpowiadającym sobie w sensie dynamicznym (ciągi wyjściowe muszą być opóźnione w stosunku do wejściowych – dane odpowiadające sobie w sensie domniemanego stanu stacjonarnego). Do wykonania takiej selekcji potrzebna jest znajomość czasu ustalania odpowiedzi skokowej – zastępczych stałych czasowych dla wszystkich kanałów wejścia-wyjścia, które można oszacować przez:

- eksperckie przyporządkowanie wejść i wyjść;
- z wykorzystaniem algorytmu estymacji opóźnień. Podstawą estymacji jest eksperckie przyporządkowanie wejść i wyjść, czyli strukturyzacja danych procesu.

Do detekcji stanu stacjonarnego na wyjściu wykorzystany będzie algorytm rozpoznawania zmian przez analizę istotności trendu. Na podstawie zarejestrowanych wartości opóźnień możliwa jest konstrukcja modelu zależności opóźnień od zmiennych procesowych. Realizacja takiego podejścia wymaga wdrożenia algorytmu detekcji zmian dla danych ustrukturyzowanych (skategoryzowanych) strukturą *Proces*.

1.2. Atrybutyzacja

Mając przyporządkowane odpowiadające sobie ciągi wejść i wyjść możliwa jest charakteryzacja punktu pracy, która wymaga atrybutyzacji. Atrybutami będą przykładowo wartości średnie zmiennych procesowych i zmienność (np. odchylenie standardowe wokół tych średnich), jakość punktu pracy (osiąganie założonej jakości produkcji, przy czym taka informacja ta może być znacznie opóźniona). Atrybuty powinny dotyczyć także poziomu wartości zmiennych procesowych (ujęcie jakościowe), który powinien być odniesiony do wartości nominalnych założonych lub wynikających z symulacji czy relacje pomiędzy wartościami nominalnymi a realnymi.

Wśród atrybutów proponuje się uwzględnienie także charakterystykę kompleksową typu odległości szeregów czasowych od szeregów wzorcowych, co charakteryzować będzie stopień przynależności do grupy procesów. W odróżnieniu od podejścia klasycznego (do analizy wykorzystywane są wartości nominalne, wykonywane są analizy korelacyjne zmiennych procesowych), w proponowanym W podejściu rozpatrywane jest podobieństwo między sygnałami diagnostycznymi (zamiast korelacji: analiza podobieństwa różnych wyjść procesu). Pozwala to na ustalenie przynależności danej sytuacji do grupy, a także sprawdzanie podobieństwa (w sensie odległościowym) pomiędzy czasem reakcji procesu na określone zmiany procesu, występujące retrospektywnie dla podobnych wejść i wyjść. Wysokie podobieństwo, tj. przynależność do grupy, może być podstawą do zakwalifikowania punktu pracy do identyfikacji modeli statyki i dynamiki. Klasyfikacja sekwencji stanów stacjonarnych po zmianie wejścia pozwala na wyznaczenie opóźnień.

Analizie poddane mogą być tylko reszty wejść i reszty modelu wyjść (a nie reszty średnich wejść i wyjść). W tym ujęciu badane jest ich podobieństwo.

Sytuacje typu wejście i wyjście z opóźnieniem określa się sytuacją procesową. Mając zatrybutyzowane sytuacje procesowe możemy dokonać ich pogrupowania, a następnie klasyfikowania do określonych grup. Realizację tego postulatu umożliwiłaby wczesne ostrzeżenie, tj. detekcję sytuacji nietypowych.

Algorytm wykrywania zmian w oparciu o analizy podobieństwa może być ukierunkowany na:

- szukanie podobieństwa w przestrzeni atrybutów
- szukanie podobieństwa pomiędzy szeregami (analizowanymi w parach) w danej (obserwowanej) sytuacji procesowej do szeregów uprzednio sklasyfikowanych.

Istotną rolę, znacznie utrudniającą tworzenie skutecznych algorytmów detekcji, odgrywa asynchroniczna recepcja informacji wyjściowych, w tym mechanizm rejestrowania informacji jakościowej (m.in. o osiągnięciu celu sterowania).

Na podstawie analizy podobieństwa możliwe jest generowanie prognoz w oparciu o charakterystykę szeregu przy klasie oddziaływań wejściowych i ich skutków (wszystkich ważnych informacji wyjściowych, odpowiednio opóźnionych, w tym jakościowych dotyczących osiągnięcia celu produkcji).

1.3. Analizy podobieństwa

Analizy podobieństwa mogą być prowadzone dla danej chwili czasu (próbki) bądź na podstawie danych retrospektywnych (w odniesieniu do sytuacji procesowych, które miały miejsce w okresie wcześniejszym). Celem jest poszukiwanie cech zwiastujących zmianę poziomu wartości zmiennej w niepodobnych szeregach, a więc wczesnych symptomów sytuacji niepokojących. Do tego celu wykorzystuje się miary odległościowe szeregów przesuniętych w czasie. Analizie poddany jest zbiór zmiennych procesowych.

Dodatkowo, w sytuacji nierozpoznanej (stan przejściowy) mogą być prowadzone analizy zmierzające do identyfikacji podobnych (w sensie odległościowym) szeregów

1.4. Wartości nominalne i typy atrybutów

Wartości nominalne mogą być ustawiane przez operatora (ekspercko), prognozowane na podstawie wejścia (predykcja) bądź wyjścia (symulacja) bazując na wartościach wejściowych.

Mając na celu charakterystykę punktu pracy, tj. detekcję stanu ustalonego bądź przejściowego, dla każdego procesu jako atrybuty można rozważać:

- typ wartości nominalnej
- poziom wartości nominalnej
- zmienność/stałość wartości nominalnej
- czas od ostatniej zmiany
- jakościowa charakterystyka stanu zmiennych wobec wartości nominalnej (stan ograniczeń technologicznych)
- wynik analizy reziduów (muszą być stacjonarne) – odchylenie standardowe reziduów

Dla każdego procesu tworzona będzie w ten sposób przestrzeń o wymiarach k zmiennych \times L atrybutów.

1.5. Algorytmy hybrydowe

Wykrywanie krótkoterminowych zmian w sygnałach diagnostycznych może być prowadzone w oparciu o analizę podobieństwa szeregów czasowych, analizowanych w ruchomym oknie. Wyniki takich analiz są także punktem wyjścia wyszukiwania podobnych charakterystyk, opisanych wyżej.

Analizy mogą być prowadzone w oparciu o:

- (a) klasyczne metody badania odległości szeregów
- (b) dedykowane metody badania odległości szeregów
- (c) metody hybrydowe, dwupoziomowe, dedykowane do detekcji istotnych zmian

Badania empiryczne skoncentrowano na wariancie (c). W algorytmach hybrydowych detekcja przebiega dwuetapowo. W pierwszym etapie prowadzona jest wstępna detekcja zmian z wykorzystaniem statystycznych metod detekcji zmian krótkoterminowych bądź metod ukierunkowanych na wykrywanie stacjonarności analizowanego szeregu. W etapie drugim wykorzystywana jest miara podobieństwa (a) lub (b).

1.6. Prace zrealizowane w ramach zadania

W ramach zadania wykonano następujące prace:

- testy hybrydowego algorytmu detekcji zmian, ukierunkowane na wykrywanie unikalnych (oryginalnych) zmian w badanym szeregu
- testy hybrydowego algorytmu detekcji zmian, ukierunkowane na wykrywanie podobieństwa zdarzeniowego
- przeanalizowano możliwość wykorzystania paradygmatu immunologicznego do zwiększenia skuteczności detekcji zmian
- zaprojektowano i zaimplementowano prototypowe oprogramowanie do analiz podobieństwa szeregów

Ad. 2. Analizy oparte o symulacje procesów

W takich analizach wejścia-wyjścia zakłada się, że dostępny jest symulator (model symulacyjny), który umożliwi generowanie miarodajnej wartości referencyjnej na bieżąco, w dostosowaniu do bieżących wartości wejść, jak również dla scenariuszy (przebiegu zakłóceń w przyszłości i ewentualnie zagregowanych sposobów reagowania – wg określonego paradygmatu). Wynik symulacji może stanowić cechę procesu (wzorzec do poszukiwania odpowiedzi podobnych, a więc podstawę do diagnozy. Konieczne jest w tym przypadku rozpoznanie przesłanek analizy eksperckiej przez operatora.

Prace programistyczne

Zaprojektowano, zakodowano i przetestowano prototypowe algorytmy detekcji zdarzeń, bazujące na detekcję wczesnych symptomów zmian (detekcja niestacjonarności), a następnie identyfikacji zmian w szeregu bazowym (docelowym) z wykorzystaniem dedykowanych metod odległościowych.