

Tomasz Pelech-Pilichowski\*, Jan T. Duda\*

## Wykorzystanie podejścia immunologicznego do prognozowania szeregów czasowych

### 1. Wprowadzenie

Algorytmizacja pozyskiwania informacji rejestrowanych w postaci szeregów czasowych ukierunkowana jest na implementację procedur detekcji zdarzeń w przetwarzanych szeregach (nietypowych przebiegów rejestrowanych ciągów), a także predykcji przyszłych wartości. Wysoką jakością przetwarzania danych postrzegać można jako minimalizację opóźnienia detekcji, szybką estymację parametrów statystycznych szeregu, a także generowanie prognoz dających akceptowalne błędy dla przyjętego horyzontu predykcji, co niesie za sobą konieczność optymalizacji lub dostrajania parametrów predyktora [1]. W zależności od właściwości statystycznych i częstotliwościowych przetwarzanych szeregów oraz uwarunkowań predykcji, stosować można modele jednowymiarowe (np. ARMA [2]), modele wielowymiarowe (np. ARMAX), modele klasy ARCH/GARCH [3] czy predyktory wieloczynnikowe oparte na wykorzystaniu sieci neuronowych (w tym sieci neuronowe z radialnymi funkcjami bazowymi – *Radial Basis Function Neural Nets*, RBFNN [4]). Algorytmiczną predykcję szeregów czasowych prowadzić można w oparciu o ekstrapolację formuł trendu, z zastosowaniem modeli sygnałowych lub odcinkowo-liniowych reprezentacji szeregów [5]. W tym ujęciu, dla każdego segmentu szeregu zakłada się stałość parametrów statystycznych, a zatem dopuszczalność ekstrapolacji bieżącego trendu. Uzyskanie wiarygodnej predykcji warunkowane jest algorytmicznym nadzorowaniem działania predyktora metodami eksperckimi lub algorytmicznymi w celu wiarygodnej detekcji zdarzeń długoterminowych (zmian trendu), co może być realizowane z wykorzystaniem testów statystycznych czy reguł wykorzystujących przekrojowe analizy zdarzeń w otoczeniu badanych szeregów, między innymi z zastosowaniem metod inteligencji obliczeniowej.

Do algorytmicznego przetwarzania szeregów wykorzystać można metody sztucznej inteligencji, w tym mechanizmy wzorowane na działaniu naturalnych układów odpornościowych, w których zadanie detekcji zdarzeń (wykrywanie patogenów) realizowane jest

---

\* Katedra Informatyki Stosowanej, Akademia Górniczo-Hutnicza w Krakowie

przez implementację rozproszonych detektorów zdolnych do wykrywania nieznanymi struktur. Wielość reguł klasyfikacji komórek na własne i obce, możliwość detekcji zmian w rozległym środowisku na podstawie niewielkiego zestawu analizowanych obiektów, wykorzystanie struktur pamięciowych oraz zdolność adaptacji stanowią inspirację do zastosowania tego podejścia do detekcji zdarzeń i predykcji szeregów czasowych, szczególnie finansowych, na których kształtowanie ma wpływ wiele czynników losowych.

Zwiększenie skuteczności detekcji zmian trendu, ukierunkowane na minimalizację opóźnienia detekcji (często nieakceptowalnego), generowanego przez wykorzystanie metod opartych na silnych testach statystycznych, związane jest z zastosowaniem procedur ukierunkowanych na szybką sygnalizację wystąpienia zdarzeń zwiastujących zmiany w szeregu na podstawie informacji generowanej w wyniku kompleksowego przetwarzania dużego zbioru szeregów czasowych. Zastosowanie do tego celu podejścia immunologicznego pozwala na ograniczenie liczebności każdorazowo przetwarzanych obiektów, a także na wykrywanie złożonych, nieznanymi charakterystyk zdarzeń w otoczeniu szeregu prognozowanego. Konstrukcja takich algorytmów wymaga definicji podstawowych obiektów systemu, działania krótko i długoterminowych detektorów zdarzeń, wykorzystania skutecznych sposobów wykrywania unikalnych zmian w szeregach (nie mających następstwa w innych przetwarzanych sygnałach – współbieżnie lub z opóźnieniem), sprecyzowania reguł adaptacji strukturalnej i parametrycznej, a także dostrojenia ustalanych arbitralnie parametrów algorytmów detekcji zdarzeń.

## **2. Propozycja zastosowania podejścia immunologicznego do detekcji zdarzeń w szeregach czasowych oraz predykcji**

### **2.1. Detekcja zdarzeń oraz predykcja szeregów czasowych**

Komputerowa analiza szeregów czasowych jest ukierunkowana na identyfikację ich właściwości statystycznych i dynamicznych, w powiązaniu ze znanymi oddziaływaniami zewnętrznymi o charakterze jakościowym lub ilościowym. Celem jest na ogół umożliwienie wiarygodnej predykcji lub/i symulacji szeregów z wykorzystaniem modeli matematycznych. Do przetwarzania szeregów wykorzystuje się zarówno klasyczne techniki analizy sygnałów (analizy statystyczne, częstotliwościowe), metody wielowymiarowej analizy szeregów czasowych, inteligencji obliczeniowej, rozpoznawania obrazów, jak i podejścia oparte na metodach rozpoznawania obrazów oraz inteligencji obliczeniowej. Przetwarzanie danych (w czasie rzeczywistym lub retrospektywnie) należy prowadzić dla odpowiednio krótkiego okna ruchomego, co umożliwi szybką identyfikację zmian właściwości pojedynczych sygnałów, możliwość badania koincydencji zdarzeń z uwzględnieniem losowo zmiennych opóźnień między zdarzeniami w różnych szeregach oraz zróżnicowanej specyfiki tych zmian.

Detekcja zdarzeń ma zastosowanie m.in. dla potrzeb prognozowania [10] bądź wykrywania anomalii [6] w obszarze diagnostyki technicznej [7], medycznej [11], rozpoznawania obrazów [12] itp. Wykrywanie zdarzeń w szeregach czasowych ukierunkowane jest na identyfikację krótko- oraz długoterminowych okresów niestandardowego zachowania sygnału w analizowanym oknie, zazwyczaj wydzielenia okresów stacjonarności i niestacjonarności (o nielosowych przyczynach). Zastosowany algorytm klasyfikacji jest realizacją funkcji odwzorowującej przynależność danych do klasy normalnej oraz klasy anomalii [6]. Klasyczne podejście detekcji zdarzeń w szeregach czasowych opiera się na analizie istotności odchyłek [8] oraz implementacji testów stosunku funkcji wiarygodności (*Likelihood Ratio*, LR), znanych jako metoda Page'a–Hinkleya [7], wraz z uogólnieniami opartymi m.in. na wielokryterialnej analizie istotności trendów [8]. Metody te wykazują wysoką skuteczność detekcji zmian dla sygnałów zaszumionych zakłóceniami wysokoczęstotliwościowymi [9]. Zastosowanie rygorystycznych wymagań (niskie prawdopodobieństwo fałszywego alarmu, wysokie prawdopodobieństwo wykrycia zdarzenia) powoduje nieakceptowalne opóźnienie detekcji (rzędu nawet kilku tysięcy próbek przy założeniu zmian długoterminowych o niewielkiej amplitudzie).

Analiza istotności odchyłek w oparciu o rygorystyczne testy LR (czynnika Bayesa) ma zastosowanie do segmentacji odcinkowoliniowej szeregów czasowych (tzw. segmentacja LR szeregów [9]). Ekstrapolacja formuły trendu wyznaczonego dla danego odcinka lub wykorzystanie bieżących (chwilowych) parametrów daje predykcję dla chwili wyprzedzającej czas bieżący o ustaloną liczbę próbek [22]. Uzyskiwane w ten sposób prognozy, w okresie zmian szeregu o stosunkowo niewielkim zakresie, dają błędy porównywalne do błędów aproksymat cyklicznych (dla cyklu 8-letniego) [13], nie wykazując wyraźnych tendencji. W okresach zmian trendu następują skokowe zmiany wartości ciągów reziduum; możliwe jest także wystąpienie trendów w tych przebiegach.

Zwiększenie skuteczności takiej predykcji ekstrapolacyjnej w znacznym stopniu uwarunkowane jest skróceniem opóźnienia detekcji, kiedy formuły trendu nie są akceptowalne dla próbek bieżących oraz obejmujących horyzont prognozy. Współbieżne przetwarzanie większego zbioru szeregów, ukierunkowane na identyfikację krótkoterminowych zdarzeń symptomatycznych, zwiastujących załamania trendu w szeregu prognozowanym, wymaga eksploracji dodatkowych zbiorów danych, charakteryzujących badany obiekt. Zastosowanie do tego celu klasycznych metod wykrywania zmian związane będzie z dużą czasochłonnością obliczeń, a także możliwością eksploracji cech nieistotnych, co powodować może zwiększenie liczby fałszywych alarmów.

Takie zadanie detekcji zdarzeń może być realizowane z wykorzystaniem podejścia immunologicznego, gdzie wielość kryteriów detekcji, analizowanych obiektów, binarna klasyfikacja bez nauczyciela, rozproszone oraz kompleksowość detekcji sugeruje możliwość odwzorowania w algorytmice, w szczególności do przetwarzania dużych zbiorów szeregów czasowych. Zastosowanie tego paradygmatu pozwala także na detekcję zdarzeń w środowisku na podstawie eksploracji stosunkowo nielicznego podzbioru przetwarzanego globalnie (dostępnego) zbioru szeregów.

## 2.2. Sztuczne systemy immunologiczne

Naturalny układ odpornościowy realizuje proces ochrony organizmu przed patogena-  
mi z wykorzystaniem limfocytów. Proces ten ma strukturę dwupoziomową. Na pierwszym  
poziomie monitorowane są komórki własne organizmu pod kątem detekcji patogenów  
przez limfocyty (głównie typu T) [14]. Na poziomie drugim, w przypadku rozpoznania an-  
tygenów, następuje aktywacja limfocytów (głównie komórek typu B), w postaci pierwotnej  
odpowiedzi immunologicznej (napotkanie patogenu po raz pierwszy) lub wtórnej (detekcja  
patogenu wcześniej rozpoznanego; wykorzystywanie pamięci immunologicznej przech-  
wującej informacje o typach wykrytych wcześniej komórek obcych oraz sposobie ich  
zwalczenia) [15]. Wysoka skuteczność prowadzonej detekcji wynika m.in. z poddawania  
dojrzewających komórek własnych procesowi selekcji negatywnej (odbywającego się  
w grasicy), którego efektem jest eliminacja niepoprawnie klasyfikujących komórki lim-  
focytów [16]. Stałe udoskonalanie limfocytów powoduje adaptację całego układu oraz  
możliwość detekcji zbioru patogenów, którego liczebność znacznie przewyższa liczbę lim-  
focytów.

Działanie poszczególnych mechanizmów naturalnych układów odpornościowych jest  
inspiracją postrzegania wielu zagadnień informatycznych w kategoriach immunologii  
w wielu obszarach [14] (np. ochrony zasobów sieciowych [17], detekcji uszkodzeń [18],  
rozpoznawaniu obrazów [19], prognozowaniu [10], w sterowaniu adaptacyjnym [20] itp.).

W ujęciu informatycznym, system immunologiczny można postrzegać jako adaptacyj-  
ny klasyfikator (ze zdecentralizowanymi mechanizmami kontroli). Na pierwszym poziomie  
następuje klasyfikacja binarna badanych obiektów (wynikiem jest identyfikacja komórek  
własnych i obcych; wykorzystanie stosunkowo prostych mechanizmów detekcji). Realizo-  
wane może być tutaj odwzorowanie selekcji negatywnej [6], pozytywnej czy sprawdzanie  
skuteczności binarnej klasyfikacji komórek. Na poziomie drugim ma miejsce klasyfikacja  
wieloklasowa w postaci pierwotnej bądź wtórnej odpowiedzi układu (zaawansowane meto-  
dy detekcji; aktywacja asynchroniczna).

Idea działania układów naturalnych może być implementowana w różnym zakresie  
oraz stopniu kompleksowości (ze względu na właściwości danych, przeznaczenie algoryt-  
mu, uwarunkowań czasowych jego funkcjonowania; w wielu przypadkach mechanizmy te  
są bardzo upraszczane).

W szczególności, w algorytmach immunopodobnych wykorzystuje się specyficzne ce-  
chy inspirowane działaniem układów naturalnych, jak m.in. wielość wykorzystywanych re-  
guł klasyfikacji, możliwość kompleksowego przetwarzania dużych zbiorów danych przez  
selektywne przetwarzanie wybieranych losowo podzbiorów (rozproszoność oraz redukcja  
informacji), agregacja i wykorzystanie informacji diagnostycznej, a także brak koniecz-  
ności wstępnego uczenia i testowania mechanizmów detekcji. Detektory, konstruowane na  
podstawie działania limfocytów, cechuje duże zróżnicowanie, stosunkowo duża autonomia,  
a także możliwość adaptacji parametrycznej i strukturalnej (m.in. zmiany reguł klasyfikacji;  
do adaptacji można wykorzystać operatory genetyczne).

### 2.3. Idea konstrukcji immunopodobnych algorytmów do zwiększenia skuteczności predykcji oraz detekcji zdarzeń

#### Definicja komórek systemu, stanu zdrowia i choroby

Zakładając, że dostępny zestaw szeregów czasowych tworzy środowisko szeregu prognozowanego, jego otoczenie stanowi podzbiór szeregów środowiska, wybierany losowo lub za pomocą zdefiniowanych/samoczących się procedur. Poszczególne komórki (oznaczone jako BK) generowane są z otoczenia. Składają się zawsze z fragmentów o zadanej długości szeregu prognozowanego (nazywanego bazowym B) oraz szeregu otoczenia K, współbieżnych lub opóźnionych, dobieranych z wykorzystaniem szybkich metod detekcji chwilowego podobieństwa. Populacja komórek otoczenia jest modyfikowana (adaptacja strukturalna systemu prowadzona w tle przetwarzania zasadniczego).

Standardowo, analizowana jest stosunkowo niewielka liczba komórek otoczenia. W pierwszym etapie prowadzona jest detekcja chwilowych zdarzeń w szeregach komórki (przykładowe oznaczenia B1 – zdarzenie w szeregu B, K0 – brak zdarzenia w szeregu K danej komórki BK). Efektem jest identyfikacja jednej z czterech sytuacji: B0K0 (brak zdarzeń w obydwóch szeregach), B1K1 (wykrycie zdarzeń w obydwóch szeregach), B0K1 oraz B1K0. W kolejnym etapie, na podstawie konfiguracji wykrytych zdarzeń, stwierdzany jest stan PB0, który dla sytuacji B0K0 jest stanem zdrowia (brak istotnych zmian w szeregach komórek; parametry statystyczne szeregu bazowego są przewidywalne, możliwe jest prowadzenie predykcji na podstawie wyników segmentacji LR), lub PB1 (stan choroby dla sytuacji B1K1; konieczność modyfikacji parametrów testu LR, tj. zmniejszenia prawdopodobieństwa wykrycia zdarzenia, zwiększenia prawdopodobieństwa fałszywego alarmu – skrócenie okna analizy). Identyfikacja stanu zdrowia PB0 lub choroby PB1 dokonywana jest z wykorzystaniem dedykowanych metod badania podobieństwa (analogia do działania limfocytów).

W sytuacji nieokreślonej (B0K1 lub B1K0) dokonywana jest dodatkowa analiza pewnej liczby komórek ze środowiska (doraźne poszerzenie zbioru BK) z wykorzystaniem bardziej złożonych mechanizmów detekcji (adaptacja strukturalna otoczenia). W efekcie identyfikuje się obecność zdarzenia lub jego braku. Sposób prowadzenia analizy środowiska może być poszerzony o wykorzystanie informacji dodatkowej (znak trendu, zgodność trendów w wąskim oknie – metoda trzech średnich [21], dla szeregów finansowych faza cyklu koniunktury światowej itp.). Dodatkowo, skuteczność detekcji może być zwiększona poprzez analizy komórek prowadzone w tle (np. obliczenia referencyjnych wartości miar odległości, wykorzystywanych w kryteriach klasyfikacji prowadzonej w czasie rzeczywistym dla szeregów BK). Przechowywanie informacji o charakterystykach szeregów w stanie zdrowia (tzw. parametry zdrowia) wymaga stworzenia odpowiedniej struktury danych (pamięci układu) przechowującej informacje o parametrach stanu zdrowia.

Proponuje się wykorzystanie są trzech rodzajów limfocytów. Detekcja zmian w szeregach realizowana jest przez limfocyty typu L1 i L2. Zadaniem limfocytu L3 jest kompleksowa analiza otoczenia szeregu bazowego, ukierunkowana na zapewnienie wiarygodnej

informacji o możliwości prowadzenia prognoz ekstrapolujących segmenty wyznaczone z wykorzystaniem procedur LR.

Właściwości wielu szeregów czasowych (m.in. finansowych) są zbliżone do procesu Wienera [22, 23] (błądzenia przypadkowego), a zatem w stanie zdrowia (tj. warunkach typowych) ciągi jednokrokowych przyrostów nie wykazują istotnej autokorelacji. Długookresowe zmiany trendu mogą być rezultatem obecności krótkich (chwilowych) serii skorelowanych przyrostów.

Limfocyt typu pierwszego (L1), aktywowany synchronicznie dla każdej kolejnej próbki, realizuje zadanie wykrywania pojawienia się w sygnale składowych nielosowych (chwilowe zmiany stacjonarności). Prowadzona jest detekcja serii istotnych odchyłek szeregów komórki (z wykorzystaniem testu studenta [8]) oraz porównuje się błędy jednokrokowej predykcji z wykorzystaniem adaptacyjnej metody Holta [24] oraz podtrzymania zerowego rzędu (*Zero-Order-Hold*, ZOH). Detekcja serii odchyłek o zadanej długości lub wykrycie okresów, dla których błąd prognozy ZOH jest większy w porównaniu do metody Holta, jest sygnałem do uruchomienia limfocytów L2 (aktywowanych asynchronicznie, znacznie rzadziej niż L1), realizujących zadania analizy podobieństwa zdarzeń w szeregach komórki z wykorzystaniem dedykowanych metod badania odległości szeregów. Stwierdzenie podobieństwa szeregów komórki (klasyfikacja binarna) dokonywane jest na podstawie wartości progowej ustalonej statystycznie. Rezultatem działania limfocytów L2 jest potwierdzenie obecności zdarzenia (stwierdzenie B1K1), stwierdzenie niezgodności sygnałów przesyłanych przez L1 (B1K0, B0K1) lub anulowanie wstępnej diagnozy L1 (identyfikacja sytuacji B0K0; zdarzenia wstępnie wykryte uznane za nieistotne). Zastosowanie dwupoziomowej detekcji pozwala zarówno na szybką detekcję chwilowych zmian w szeregach (działanie L1) oraz zastosowanie bardziej złożonych, dedykowanych mechanizmów w przypadku wykrycia zdarzenia (L2).

Wysoka skuteczność detekcji z wykorzystaniem limfocytów L1 i L2 uwarunkowana jest stosowaniem komplementarnych miar podobieństwa, a więc umożliwiających wykrywanie zmian (odchyłek) o różnej konfiguracji i opóźnieniach pomiędzy zdarzeniami. Standardowe miary odległości oparte na metryce Minkowskiego nie pozwalają na skuteczną detekcję zdarzeń pojawiających się ze zmiennymi opóźnieniami, zatem zaproponowano dedykowane miary odległości szeregów komórki [23, 22], ukierunkowane na eliminację opóźnienia poprzez wykorzystanie transformaty Fouriera (opóźnienie z założenia jest ignorowane) – miary  $S_F$  (odległość widm amplitudowych  $k_F$  niższych harmonicznnych) oraz  $S_d$  (wyzerowanie przesunięcia fazowego i odtworzenie sygnału dla rzędu  $k_d$  próbek), poszukiwanie wzorców zdarzeń o długości  $L_w$  w szeregach komórki (miara W) bądź analizę serii odchyłek oryginalnych (miara U) oraz analizę zdarzeń na podstawie znaków pojedynczych skoków (miara Z) [23, 22]. Zadowalającą skuteczność detekcji można osiągnąć poprzez zastosowanie innych miar, jednak wymaga to przeprowadzenia obszernych badań symulacyjnych.

Analiza podobieństwa szeregów czasowych może być prowadzona dla szeregów oryginalnych lub ciągów będących wynikiem ich globalnych przekształceń (np. obliczenie przyrostów względnych, bezwzględnych czy logarytmicznych, centrowanie szeregu oryginalnego).

nalnego w długim oknie ruchomym, potęgowanie lub rangowanie tak scentrowanych szeregów, wygładzanie, transformacje widmowe, falkowe, ortogonalizujące itp. [25]). Do uzyskanych w ten sposób sygnałów diagnostycznych stosować można dalsze przekształcenia (m.in. średnia, maksimum lub minimum, wariancja, mediana, liczone w oknie o odpowiednio dobranej szerokości).

Limfocyty L2 stosują dodatkowo wewnętrzne transformacje szeregów w oknie komórki, polegające na zastosowaniu obligatoryjnej uniformizacji sygnałów (uzyskanie scentrowanych ciągów próbek tego samego rzędu wielkości, realizowane przez ekstrakcję średniej lub detrending) oraz wykorzystaniu dodatkowych transformacji uwypuklających/maskujących określone cechy zunifikowanych szeregów komórki (podzielenie zunifikowanych ciągów przez wartość referencyjną: odchylenie standardowe szeregu resztowego w komórce, wartość maksymalną modułów szeregu resztowego w komórce lub stałą wartość – parametr detektora). Dodatkowo, sygnały mogą być poddawane dalszym transformacjom, ukierunkowanym na wyeliminowanie wpływu zmiennych opóźnień hipotetycznych reakcji przyczynowo-skutkowych (w tym możliwej ich inwersji) bądź ujednoczenie wzorców zdarzeń.

Wykorzystanie odpowiednio dobranego zbioru miar podobieństwa ( $TM$ ), wektora parametrów ( $WP$ ), zbioru dostępnych sygnałów diagnostycznych ( $TSD$ ), przestrzenie przekształceń globalnych ( $TGS$ ), zbioru transformacji wewnętrznych w oknie komórki ( $TSU$ ) oraz przestrzeni przekształceń dodatkowych ( $MTS$ ), pozwala na uzyskanie licznego, zdywersyfikowanego zbioru detektorów (co m.in. umożliwi rozpraszanie obliczeń)  $ZD = TM \times WP \times TSD \times TGS \times TSU \times MTS$ . Zastosowanie odpowiednich metakryteriów oceny skuteczności detekcji pozwoli na samoadaptację systemu (dostrajanie parametrów detektora i eliminację mniej skutecznych limfocytów).

Działanie limfocytów L3 ukierunkowane jest na monitorowanie wysokiej wiarygodności predykcji poprzez modyfikację parametrów testu LR (co ma wpływ na szybszą detekcję zmian trendu, np. lokalnych) na podstawie informacji przesyłanych z limfocytów L2, tj. zmianę wartości założonego prawdopodobieństwa fałszywego alarmu oraz prawdopodobieństwa wykrycia zdarzenia, a także w określonych przypadkach zmianę długości ciągu próbek, wymaganej do wiarygodnej estymacji trendu w nowym segmencie. Zapewnienie wiarygodnej prognozy realizowane jest poprzez generowanie sygnału o możliwości kontynuacji predykcji ekstrapolacyjnej na podstawie bieżących parametrów trendu bądź sygnału blokującego akceptację takich prognoz. Dodatkowo, prowadzone są analizy wspomagające (statystyki właściwości przetwarzanych danych, aktywacji limfocytów, trafności diagnoz w odniesieniu do rzeczywistej sytuacji szeregu na podstawie danych retrospektywnych itp.). Poszczególne zadania dodatkowe mogą być prowadzone w tle. Na podstawie zebranych i zapamiętanych informacji możliwa jest m.in. eliminacja nieskutecznych limfocytów L2 oraz ich adaptacja.

### **Przykładowe wyniki badania skuteczności detektorów**

Istotny wpływ na zasadność proponowanej koncepcji ma skuteczność detekcji zdarzeń przez limfocyty L2, które prowadzą identyfikację zmian zwiastujących załamanie trendu.

W związku z powyższym, analizę skuteczności działania tych limfocytów przedstawiono na przykładzie potwierdzenia zdarzenia w obydwu szeregach komórki (sytuacji B1K1) i odniesienia uzyskanych wyników do momentów załamania trendu wykrytych retrospektywnie z wykorzystaniem rygorystycznego testu LR (czynnika Bayesa) o niskim prawdopodobieństwie fałszywego alarmu [9, 7].

Badanie przeprowadzono dla 64. finansowych szeregów czasowych (2385 cen zamknięcia od 11.07.1994 do 9.05.2008 r.): indeksów wiodących giełd światowych (DAX, DJI, FTSE, Nasdaq, Nikkei225, SP500), mniej znaczących giełd (AORD, ATX, BSESENSEX, BUX, IBOVESPA, MICEX, PX50, SAX, SSECI), cen surowców (Alumin, Brent, GasOilARA, Gold, HeatOil, LMECash, NYGasF, ResOilARA, Silver, SingROil, Tin, USCoal, USGulfHOil, WTI, Zinc), kursów walut (AUD, CHF, EUR, GBP, JPY, SEK, USD), cen akcji spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie (ArchCoal, ASSECOPOL, BANKBPH, BOS, BRE, BZWBK, COMARCH, ELZAB, HANDLOWY, IBSYSTEM, IGROUP, INGBSK, KREDYTB, MILLENNIUM, PEKAO, SYGNITY, TPSA) oraz indeksów (MWIG40, SWIG80, WIG, WIG20, WIG-PL, WIG-BANKI, WIG-BUDOW, WIG-INFO, WIG-SPOZYW, WIG-TELKOM). Dostępny zestaw szeregów wraz z transformacjami zewnętrznymi (przyrosty logarytmiczne, rezidua filtracji dolnoprzepustowej MNK [9], rezidua aproksymat cyklicznych przy założeniu 8-letniego cyklu, generowane dla indeksów giełdowych [13]) tworzy środowisko, z którego każdorazowo (dla kolejnych iteracji) wybieranych jest zestaw 10 sygnałów diagnostycznych, tworzących otoczenie szeregu bazowego (zestaw komórek BK). Przeprowadzono detekcję zdarzeń dla przykładowych 24. szeregów bazowych, wyniki uśredniono. Podczas obliczeń numerycznych zastosowano wewnętrzną transformację szeregów komórki (prowadzoną w trakcie działania L2): ekstrakcję trendu (z pominięciem wartości odstających) oraz standaryzację zewnętrzną wartością wzorcową (odchyleniem standardowym przyrostów szeregu).

Przyjęto długość szeregów wchodzących w skład komórek równą 22. próbki, włącznie z dopuszczalną tolerancją 5. próbek (przesunięciem, dla którego oblicza się wartości podobieństwa i wybiera się najmniejszą chwilową odległość szeregów komórki). Założono próg dopasowania na poziomie 0,5 co oznacza, że wykrycie i potwierdzenie co najmniej w połowie komórek zmian krótkoterminowych (aktywacja L2) powoduje globalne potwierdzenie sytuacji B1K1 dla danej próbki szeregu bazowego. Aktywacja limfocyta L1 następuje, gdy znaleziona jest seria co najmniej 2 istotnych odchyłek (dla progu istotności  $\alpha = 0,05$ ). Badania symulacyjne przeprowadzono dla szeregów współbieżnych (opóźnienie 0) oraz przesuniętych o 5–20 próbek. Dobór szeregów do otoczenia realizowano z wykorzystaniem odległości City, natomiast do badania podobieństwa szeregów (prowadzonego przez limfocyta L2) wykorzystano miary typu  $S_F$  ( $k_F = \{2; 3\}$ )  $S_d$  ( $k_d = 4$ ),  $W$  (długość wzorca ustalana na podstawie liczby próbek dla których w wyniku działania L1 stwierdzono niestacjonarność),  $U$  oraz  $Z$  (porównywano odchyłki większe od dwukrotnej wartości odchylenia standardowego) [22, 23].

Skuteczność detekcji odniesiono do momentu załamania trendu oraz momentu wykrycia takiej zmiany (szybsze wykrycie zmian w porównaniu do segmentacji LR w przypadku



wcześniejszego wykrycia zmiany – skrócenie opóźnienia detekcji – lub przed faktyczną zmianą trendu – zwiastun załamania trendu). Uśrednione wyniki obliczeń dla zestawu szeregów bazowych oraz różnych wartości przesunięć szeregów komórki przedstawiono w tabeli 1. Wartości procentowe obliczono na podstawie sumy momentów aktywacji L2 w okresie wcześniejszym w porównaniu do opóźnienia generowanego przez zastosowanie procedury segmentacji LR oraz momentów aktywacji w okresie poprzedzającym faktyczną zmianę trendu (założono maksymalne wyprzedzenie równe 22. próbki), odniesionej do liczby wszystkich aktywacji limfocytów L2 (czyli także fałszywych alarmów).

**Tabela 1**

Skuteczność detekcji zmian typu BIK1 dla przykładowych ustawień parametrów algorytmu

Przesunięcie	0	5	10	15	20
Skuteczność	62.50%	60.87%	72.00%	54.84%	56.67%

### 3. Wnioski

Zaproponowana koncepcja algorytmizacji detekcji zdarzeń, implementowana w uproszczonej postaci (wykrywanie sytuacji BIK1) z wykorzystaniem adaptacji strukturalnej oraz parametrycznej detektorów oraz środowiska szeregu prognozowanego pozwala na zwiększenie skuteczności wykrywania zmian trendu, co przekłada się na zapewnienie większej wiarygodności predykcji ekstrapolacyjnej. Dla przykładowych, arbitralnie dobranych parametrów, uzyskano zadowalające rezultaty, zatem należy oczekiwać dalszego usprawniania detekcji w przypadku implementacji kolejnych metod badania podobieństwa szeregów, dostrojenia parametrów algorytmu (liczby generowanych komórek oraz parametrów metod detekcji). Pokazano, że przetwarzanie dużego zestawu sygnałów diagnostycznych może być prowadzone w oparciu o kompleksową analizę stosunkowo nielicznego podzbioru danych, tworzącego otoczenie szeregu bazowego.

Zaproponowana zdolność samoadaptacji pozwala na zwiększanie skuteczności wykrywania zdarzeń wraz z wydłużaniem okresu eksploatacji (algorytmy mogą być implementowane jako autonomiczne moduły softwareowe). Implementacja dalszych elementów koncepcji, m.in. limfocytów typu L3, pamięci immunologicznej, wykrywania zdarzeń o różnej konfiguracji BK, zasad doboru sygnału diagnostycznego, ukierunkowanych na generowanie sygnału o możliwości prowadzenia predykcji średnioterminowej w oparciu o ekstrapolację bieżących formuł trendu bądź konieczności zaniechania prognozowania, będzie tematem dalszych prac autorów. Obiecującym obszarem badań jest rozproszenie obliczeń oraz implementacja algorytmów dla szeregów czasowych innych typów, m.in. dla sygnałów diagnostyki technicznej charakteryzujących badany obiekt (m.in. procesów produkcyjnych, dla których obserwowany jest mniejszy wpływ czynników losowych na rejestrowane dane procesowe, w porównaniu z szeregami finansowymi).

## Literatura

- [1] Holt C.C., *Forecasting Seasonals And Trends By Exponentially Weighted Moving Averages*. Pittsburgh, Pennsylvania, Carnegie Institute of Technology, 1957.
- [2] Box G.E.P., Jenkins G.M., *Time Series Analysis*. Forecasting and Control. Holden-Day, 1976.
- [3] Engle R.F., *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. *Econometrica*, Vol. 50 Issue 4, 1982.
- [4] Buhmann M.D., Ablowitz M.J., *Radial Basis Functions: Theory and Implementations*. Cambridge University, 2003.
- [5] Keogh E.J., Pazzani M.J., *An Enhanced Representation of Time Series which Allows Fast and Accurate Classification, Clustering and Relevance Feedback*. International Conference of Knowledge Discovery and Data Mining, AAAI Press, 1998.
- [6] Stibor T., Mohr P., Timmis J., Eckert C., *Is Negative Selection Appropriate for Anomaly Detection?* Conference on Genetic and Evolutionary Computation, ACM Press, 2005.
- [7] Basseville M., *Detecting Changes in Signals and Systems – A Survey*. *Automatica*, Journal of IFAC, vol. 24, Issue 3, 1988.
- [8] Wetherill G.B., Brown D.W., *Statistical Process Control. Theory and Practice*. Chapman and Hall, 1991.
- [9] Duda J.T., *Modele matematyczne, struktury i algorytmy nadrzédnego sterowania komputerowego*. UWND AGH, Kraków, 2003.
- [10] Nunn I., White T., *The Application of Antigenic Search Techniques to Time Series Forecasting*. Conference on Genetic and Evolutionary Computation, ACM Press, 2005.
- [11] Lin J., Keogh E., Fu A., Van Herle H., *Approximations to Magic: Finding Unusual Medical Time Series*. IEEE Symposium on Computer-Based Medical Systems, IEEE Computer Society, 2005.
- [12] Papadimitriou S., Sun J., Faloutsos C., *Streaming Pattern Discovery in Multiple Time-Series*. International Conference on Very Large Data Bases, 2005.
- [13] Duda J.T., Augustynek A., *Cyclic Properties and Predictivity of LME Prices of Selected Metals in One Month to One-Year Horizon*. [w:] Współczesne problemy zarządzania przedsiębiorstwami w gospodarce rynkowej. Red. nauk. Howaniec H., Waszkielewicz W., Wydawnictwo Akademii Techniczno-Humanistycznej, Bielsko-Biała, 2008.
- [14] Dasgupta D., Ji Z., Gonzalez F., *Artificial Immune System (AIS) Research in the Last Five Years*. Congress on Evolutionary Computation, 2003
- [15] de Castro L.N., Timmis J., *Artificial Immune Systems: A New Computational Intelligence Approach*. Springer, 2002.
- [16] Banchereau J., Briere F., Caux C., Davoust J., Lebecque S., Liu Y., Pulendran B., Palucka K., *Immunobiology of Dendritic Cells*. *Annual Review of Immunology*, Vol. 18, 2000.
- [17] Glickman M., Balthrop J., Forrest S., *A Machine Learning Evaluation of an Artificial Immune System*. *Evolutionary Computation Journal*, Vol. 13, No 2, 2005.
- [18] Dasgupta D., KrishnaKumar K., Wong D., Berry M., *Negative Selection Algorithm for Aircraft Fault Detection*. International Conference on Artificial Immune Systems, 2004.
- [19] Hunt J.E., Cooke D.E., *Learning Using An Artificial Immune System*. *Journal of Network and Computer Applications*, Special Issue on Intelligent Systems: Design and Application, vol. 19, 1996.
- [20] Timmis J., Neal M., Hunt J., *An Artificial Immune System for Data Analysis*. *Biosystems*, vol. 55 Issues 1–3, Elsevier, 2000.
- [21] Murphy J.J., *Analiza techniczna rynków finansowych*. WIG-Press, Warszawa, 1999.
- [22] Pelech-Pilichowski T., *Adaptacyjne algorytmy detekcji zdarzeń w szeregach czasowych*. Rozprawa doktorska, WEAIe AGH, Kraków 2008.

- 
- [23] Pelech-Pilichowski T., Duda J.T., *Miary podobieństwa szeregów w detekcji zdarzeń*. Diagnostyka Procesów i Systemów 2009, artykuł w druku.
  - [24] Pelech T., Duda J.: *Event Detection in Financial Time Series By Immune-Based Approach*. Intelligent Information Processing and Web Mining, Advances in Soft Computing, Springer, 2006.
  - [25] Toshniwal D., Joshi R.C., *Finding Similarity in Time Series Data by Method of Time Weighted Moments*. Australasian Database Conference, vol. 39, 2005.