

## **ODKRYWANIE LINIOWYCH TRENDÓW W WARTOŚCIACH WZGLĘDNYCH KURSÓW AKCJI Z ZASTOSOWANIEM METODY WNIOSKOWANIA OPARTEJ NA BAZIE PRZYPADKÓW**

Andrzej Kluza

**Streszczenie.** Względne wartości przyrostów cen akcji stanowią wiarygodny sposób porównywania trendów w szeregach czasowych cen akcji giełdowych. Na podstawie jednosesyjnych stóp zwrotu utworzono liniowy wzorzec przebiegu cen akcji. Wykonana adaptacja metody wnioskowania na podstawie bazy przypadków pozwoliła na wyszukiwanie przebiegów cen akcji podobnych do wzorca. Dane empiryczne obejmowały 1130 notowań spółki TP S.A. W rezultacie wykonanych obliczeń wykryto ciągi o zadanej długości wynoszącej od 2 do 7 sesji i zadanej jednosesyjnej stopie zwrotu, podobne do wzorca liniowego, mające określoną wartość funkcji podobieństwa do wzorca.

**Słowa kluczowe:** odkrywanie trendów, szeregi czasowe, wnioskowanie na podstawie bazy przypadków.

### **WSTĘP**

Do najpopularniejszych metod oceny instrumentów finansowych stosowanych w podejmowaniu decyzji inwestycyjnych należy zaliczyć analizę techniczną, fundamentalną i portfelową. Analiza techniczna stanowi podejście do inwestowania na giełdzie, w którym za jeden z głównych wyznaczników oceny zachowania się rynku przyjmuje się kształtowanie się cen akcji i wolumenu transakcji. Obserwując szeregi czasowe powyższych wskaźników, dokonuje się analiz, aby rozpoznać trendy, punkty zwrotne, w których dochodzi do zmiany tendencji, jak również aby przewidywać przyszłe warianty kształtowania się cen na podstawie ich zachowania historycznego [Pring 1998]. Analizę szeregów czasowych i odkrywanie prawidłowości w przebiegach produktów finansowych zalicza się do nurtu opisowo-prognostycznego metod ilościowych [Krawczak i in. 2000]. Wypracowano wiele metod analizy technicznej, głównie służących do wnioskowania o przyszłych wartościach kursów, przy czym większość z nich nie ma silnego umocowania w teoriach naukowych.

---

W obszarze badań nad sztuczną inteligencją do analiz instrumentów finansowych używane są różnorodne metody: sieci neuronowe, logika rozmyta i algorytmy genetyczne [Rutkowska 1997, Hellström 1998], wnioskowanie oparte na bazie reguł (Rule-Based Reasoning, RBR) oraz wnioskowanie oparte na bazie przypadków (inaczej wnioskowanie oparte na precedensach, Case-Based Reasoning, CBR) [Kolodner 1993]. Przykładowymi obszarami zastosowania analizy szeregów czasowych i odkrywania wzorców są: prognozowanie na rynku papierów wartościowych [Hellström 1998], odkrywanie plagiatów [Shivakumar 1995], odkrywanie wzorców w łańcuchach białkowych i DNA [Gung Wei 1997].

## OPIS METODY

Celem pracy jest zaadaptowanie metody wnioskowania opartego na bazie przypadków do wykrywania ciągów podobnych do wzorca liniowego w szeregach względnych zmian cen akcji spółki notowanej na Warszawskiej Gieldzie Papierów Wartościowych. Zastosowana metoda, opisana wyczerpująco w pracach Kolodner [1993], Lenz i in. [1998], jest względnie nowa w porównaniu do innych metod z obszaru sztucznej inteligencji i o wiele mniej od innych popularna. W monografiach predykcja szeregów czasowych jest często wymieniana w zakresie aplikacji metody CBR. Na podstawie przeglądu publikacji można stwierdzić, że zastosowania metody CBR do analiz instrumentów finansowych są rzadkie [Słade 1991].

Wnioskowanie na podstawie bazy przypadków jest metodą pozwalającą na rozwiązywanie nowych problemów za pomocą sposobów, które okazały się skuteczne w przeszłości, zostały opisane i zachowane. Podstawą wyznaczania rozwiązania jest znajdowanie najbardziej podobnego problemu z przeszłości. Do tego celu służy funkcja podobieństwa. Funkcja ta może być określona na różne sposoby: za pomocą metryki najbliższego sąsiedztwa, może wykorzystywać algorytmy indukcyjne tworzące drzewa decyzyjne, do oceny podobieństwa można używać również logiki rozmytej lub języka zapytań baz danych SQL [Watson 1999].

Szeregi czasowe badanych instrumentów finansowych spełniają niezbędne warunki zastosowania metody CBR [Kolodner 1993]:

- istnieje regularność przebiegów, tzn. ich wartości mieszczą się w określonych granicach i następują po sobie w pewnych interwałach czasowych,
- powtarzalność zjawisk mówi o tym, że przebiegi wartości mają tendencję do powtarzania się co jakiś czas z pewną dokładnością w zakresie wartości i w wymiarze czasu.

## ADAPTACJA METODY CBR DO ODNAJDYWANIA TRENDÓW LINIOWYCH DLA WARTOŚCI STÓP ZWROTU

Wartości kursów instrumentów finansowych mierzone w walucie danego kraju są podstawą do oceny poziomu kursów. Z powodu mechanizmu zmiany wartości, występującego na rynku giełdowym, bezwzględne różnice cen instrumentów nie dają jednak

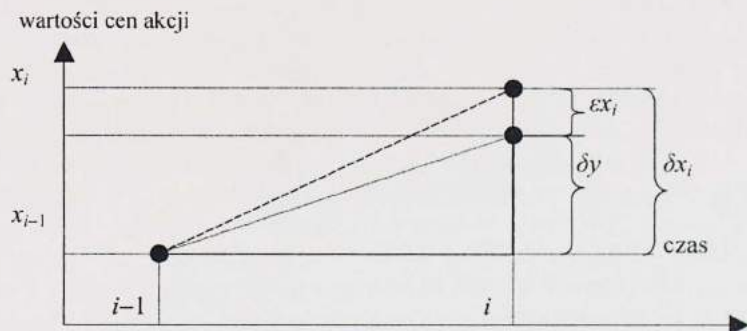


wiarygodnej informacji o wzroście bądź spadku notowań przy porównaniach dwóch instrumentów lub przy porównaniu dwóch odległych od siebie okresów zmian wartości dla jednego instrumentu. Do tego celu używane są metody porównywania wartości względnych lub wartości zlogarytmowanych [Hellström 1998]. Powszechnie używana przez członków społeczności giełdowej jest jednosesyjna wartość stopy zwrotu, definiowana jako  $R(i) = (x_i - x_{i-1})/x_{i-1}$ , gdzie  $i$  to numer sesji,  $x_i$  – cena zamknięcia instrumentu na  $i$ -tej sesji,  $i \in N^+$ . Jednosesyjna logarytmiczna wartość stopy zwrotu  $R_{\log}(i) = \log(x_i/\log x_{i-1})$  jest natomiast często stosowana w kręgach akademickich. Jednosesyjna wartość względna stopy zwrotu uogólniona na  $k$ -tej sesji wynosi  $R_k(i) = (x_i - x_{i-k})/x_{i-k}$ ,  $k \in N^+$ . Względna  $k$ -sesyjna stopa zwrotu może być interpretowana jako  $k$ -sesyjny trend dla tego instrumentu.

Podstawowym elementem metody wnioskowania CBR jest przypadek. W szeregu czasowym wartości cen akcji  $x_i$ ,  $i = 1, 2, \dots, n$ ,  $j$ -tym przypadkiem będziemy określać  $j$ -ty podciąg, zawierający  $k$  wartości względnych różnic  $\delta x_j = (\delta x_{j+1}, \delta x_{j+2}, \dots, \delta x_{j+k})$ , gdzie  $\delta x_j = R(j)$ ,  $k = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, n - k$ . Przypadki określone na szeregu czasowym w metodzie CBR są porównywane ze wzorcem, a najbardziej podobne z nich są wyróżniane i sortowane według malejącej wartości funkcji podobieństwa.

W przypadku ustalenia liniowego przebiegu wzorcowego kolejne wartości w wektorze wzorcowym  $\delta y_j = (\delta y_1, \delta y_2, \dots, \delta y_k)$  powinny różnić się między sobą o wzorcową wartość względną jednosesyjnego przyrostu  $\delta y$  tak, że  $\delta y_l = \delta y$ , gdzie  $l = 1, 2, \dots, k$ . Wartości wzorcowe przyrostu  $\delta y$  możemy dowolnie określać z przedziału, w jakim zmienia się rzeczywiście wartość jednosesyjnej stopy zwrotu dla waloru  $\delta y \in [\min(\delta y_i), \max(\delta y_i)]$ ,  $i = 2, \dots, n$ . Dla ujemnych wartości  $\delta y$  wzorec prezentuje sobą idealny,  $k$ -sesyjny trend malejący.

Podobieństwem lokalnym określamy wartość mówiącą o tym, jak bardzo podobne są do siebie przypadek i wzorec w zakresie jednej cechy. Cechą będziemy określać jednosesyjną stopę zwrotu po  $i$ -tej sesji. Błąd jednosesyjnej stopy zwrotu  $\epsilon x_i$  odzwierciedla nieidealną zgodność rzeczywistego przyrostu wartości akcji  $\delta x_i$  w stosunku do założonego przyrostu wzorcowego  $\delta y$  między cenami zamknięcia na sesjach  $i-1$ -szej a  $i$ -tej. Wartość błędu przyrostu  $\epsilon x_i = \delta y - \delta x_i$ , dla  $i = 2, 3, \dots, n$  (zob. wykres 1).



Wykres 1. Wartość jednosesyjnej stopy zwrotu wartości akcji w stosunku do wzorca  
Graph 1. One-session return compared to trend pattern

Źródło: Badania własne, 2003.

Source: Author's research, 2003.

W celu zestandaryzowania błędu przyrostu  $\varepsilon x_i$  wyznaczmy jego maksymalną wartość, którą będziemy traktować jako wartość odniesienia jednosesyjnej stopy zwrotu  $\varepsilon x_{MAX} = -\min(\delta x_i) + \max(\delta x_i)$ ,  $i = 2, \dots, n$ . Teraz możemy zdefiniować funkcję podobieństwa lokalnego  $p_i$  jako opartą na odległości względnej [Richter 1995]:

$$p_i = 1 - \left| \frac{\varepsilon x_i}{\varepsilon x_{MAX}} \right|, \quad i = 2, \dots, n.$$

Zastosowana metoda najbliższego sąsiedztwa mówi, iż dwa obiekty są do siebie najbardziej podobne, gdy odległość między nimi w pewnej metryce jest najmniejsza. Konstrukcja zastosowanej funkcji podobieństwa globalnego jest modyfikacją metryki Hamminga [Lenz i in. 1998]. Podobieństwo globalne dla  $j$ -tego przypadku,  $j = 2, 3, \dots, n - k$ , określimy jako względną sumę podobieństw lokalnych z  $k$  kolejnych sesji:

$$SIM(j) = \frac{\sum_{i=1}^k \left( 1 - \left| \frac{\varepsilon x_j}{\varepsilon x_{MAX}} \right| \right)}{k}, \quad k = 1, 2, \dots, n.$$

Funkcje podobieństwa zarówno lokalnego, jak i globalnego mogą osiągać wartości z przedziału  $[0, 1]$ .

## OBLICZENIA WYKONANE NA DANYCH EMPIRYCZNYCH

Do obliczeń wykorzystano dane empiryczne określające 1130 wartości notowań cen zamknięcia dla akcji spółki TP S.A. na Warszawskiej Giełdzie Papierów Wartościowych w okresie od 04.01.1999 r. do 16.07.2003 r. Najmniejsza wartość cen akcji spółki wynosiła 10,20 zł, a największa 40,00 zł. Największy spadek wyniósł 3,40 zł, a największy wzrost 3,50 zł. Dla jednosesyjnych stóp zwrotów wartość największa wynosiła 10,71%, najmniejsza  $-9,85\%$ , wartość odniesienia  $\varepsilon x_{MAX} = 20,56\%$ . Najmniejsza wartość jednosesyjnej stopy zwrotu, obliczanej metodą logarytmiczną, wyniosła  $-0,04506$ , największa  $-0,04420$ .

Obliczenia wykonano przyjmując długość wzorca od 2 do 7 sesji, wzorzec stopy zwrotu  $-10,00\%$ ,  $-5,00\%$ ,  $0,00\%$ ,  $5,00\%$ ,  $10,00\%$ . Dla tworzenia przypadków wybrano wszystkie 2- do 7-wyrazowe podciągi z 1130 notowań. W trakcie porównania wzorca z przypadkami były obliczane wartości podobieństwa globalnego. Dla każdej kombinacji długość wzorca – wzorzec stopy zwrotu wyznaczono ranking ciągów notowań, rozpoczynając od notowania o największej wartości funkcji podobieństwa globalnego.

Tabele 1 i 2 przedstawiają część rezultatów z wykonanych obliczeń, odpowiednio dla zerowego i pięcioprocentowego idealnego przyrostu sesyjnego. W kolumnach pionowych umieszczono informacje o pięciu ciągach notowań najbardziej podobnych według funkcji podobieństwa globalnego do wzorców o długości 2, 3, 5 i 7 sesji. Dane o ciągach zawierają daty pierwszej sesji oraz wartości funkcji podobieństwa globalnego (SIM) dla każdego ciągu.



Tabela 1. Ciągi notowań najbardziej podobne do wzorca o przyroście 0,00%  
Table 1. One-session return compared to trend pattern

Ranking	Liczba sesji 2		3		5		7	
	data	SIM	data	SIM	data	SIM	data	SIM
1	10-02-2003	1,000	30-08-1999	1,000	05-07-1999	0,993	02-01-2003	0,973
2	06-01-2003	1,000	06-07-1999	1,000	06-01-2003	0,993	31-12-2002	0,973
3	23-06-2000	1,000	07-07-1999	0,994	02-07-1999	0,979	05-06-2002	0,973
4	21-03-2000	1,000	05-07-1999	0,994	05-10-1999	0,977	24-08-1999	0,972
5	07-03-2000	1,000	06-01-2003	0,994	06-07-1999	0,976	04-06-2002	0,970

Źródło: Badania własne, 2003.

Source: Author's research, 2003.

Tabela 2. Ciągi notowań najbardziej podobne do wzorca o przyroście 5,00%  
Table 2. One-session return compared to trend pattern

Ranking	Liczba sesji 2		3		5		7	
	data	SIM	data	SIM	data	SIM	data	SIM
1	02-11-2000	0,975	02-11-2000	0,935	20-03-2003	0,897	04-10-2001	0,895
2	13-12-2000	0,973	12-12-2000	0,927	20-08-1999	0,896	09-11-1999	0,874
3	13-11-2001	0,966	20-08-2001	0,927	18-08-1999	0,896	10-11-1999	0,864
4	25-09-2001	0,961	10-10-2001	0,923	19-08-1999	0,890	08-11-2001	0,861
5	26-07-2002	0,959	17-11-1999	0,923	19-03-2003	0,887	11-12-2000	0,850

Źródło: Badania własne, 2003.

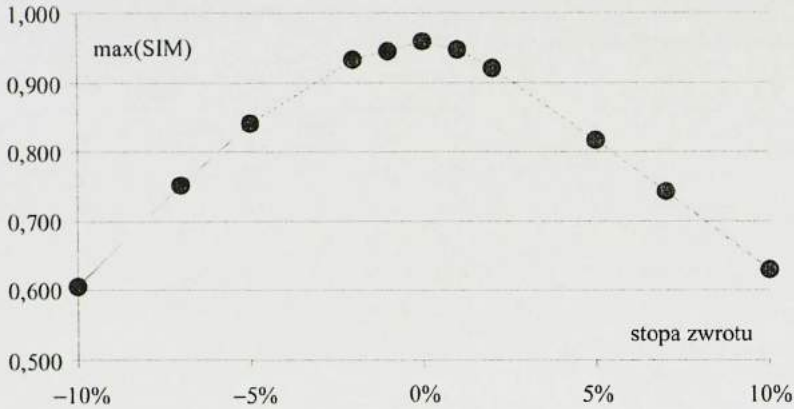
Source: Author's research, 2003.

Z danych zawartych w tabelach 1 i 2 wynika, że największe wartości podobieństwa globalnego otrzymywane były dla ciągów o najmniejszych długościach. Większe wartości podobieństwa globalnego dla ciągów krótszych oznaczają, że wśród ciągów krótszych można znaleźć ciągi znacznie bardziej podobne do wzorca liniowego niż wśród ciągów o większej długości.

Zależność widoczna na wykresie 2, na którym przedstawiono największe uzyskane wartości podobieństwa dla ciągów 20. sesji, wskazuje na to, iż im bliższa zeru jest wzorcowa stopa zwrotu, tym bardziej podobne do wzorca przypadki można odnaleźć.

Wykres 3 przedstawia pięć 7-sesyjnych ciągów stóp zwrotu, najbardziej podobnych do wzorca liniowego o przyroście 0,0% ze wszystkich badanych ciągów w zakresie. W legendzie wykresu zamieszczono numer ciągu w klasyfikacji od najbardziej podobnego do wzorca, datę pierwszej sesji ciągu oraz wartość podobieństwa globalnego dla tego ciągu. Wykres 3 przedstawia ciągi o stopie zwrotu najbardziej podobnej do 5,0%.

Wykres 5 ukazuje bezwzględne przyrosty cen akcji dla ciągów, których wykresy stóp zwrotu pokazuje wykres 4. Mimo iż na pierwszy rzut oka przebiegi z wykresu 4 wyglądają na nieuporządkowane, to na wykresie 5 widoczna jest, wykryta w obliczeniach, zdecydowana tendencja wzrostowa. Między pierwszym a piątym w klasyfikacji ciągiem można zauważyć różnicę w stopniu nieliniowości, różnica wartości podobieństwa między nimi sięga ok. 5%.

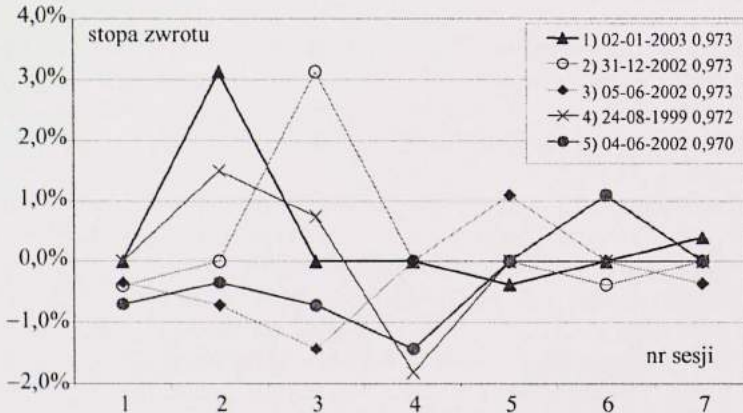


Wykres 2. Zależność wartości maksymalnego prawdopodobieństwa globalnego od założonej idealnej stopy zwrotu dla liniowego wzorca 20-sesyjnego

Graph 2. Dependence of maximal 20-sessional sequence global similarity value on set ideal return

Źródło: Badania własne, 2003.

Source: Author's research, 2003.

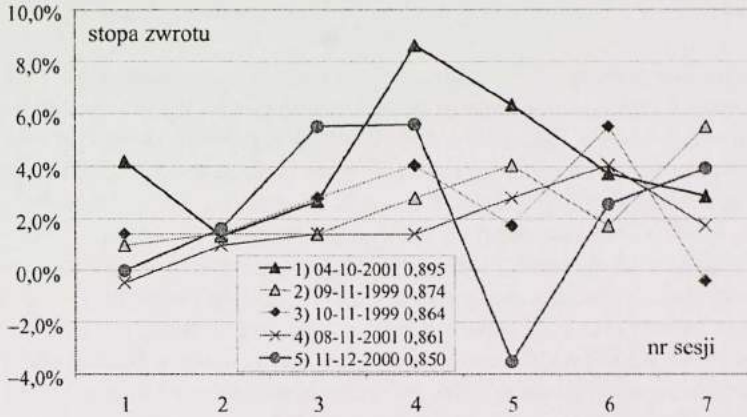


Wykres 3. Odkryte ciągi wartości jednosesyjnych stóp zwrotów, najbardziej podobne do 7-sesyjnego wzorca o przyroście 0,00%

Graph 3. Discovered sequences of one session returns, most similar to the pattern of 7-sessions 0.00% increase

Źródło: Badania własne, 2003.

Source: Author's research, 2003.

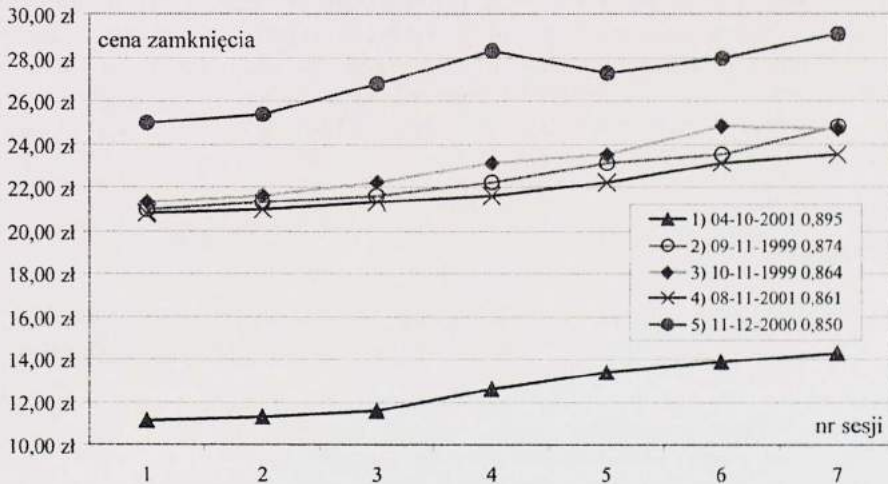


Wykres 4. Odkryte ciągi wartości jednosesyjnych stóp zwrotów, najbardziej podobne do 7-sesyjnego wzorca o przyroście 5,00%

Graph 4. Discovered sequences of one session returns, most similar to the pattern of 7-sessions 5.00% increase

Źródło: Badania własne, 2003.

Source: Author's research, 2003.



Wykres 5. Odkryte 7-sesyjne ciągi wartości akcji, najbardziej podobne do wzorca stopy zwrotu o przyroście 5,00%

Graph 5. Discovered sequences of session close values, most similar to the pattern of 7-sessions 5.00% increase

Źródło: Badania własne, 2003.

Source: Author's research, 2003.



## WNIOSKI

Zastosowanie metody CBR cechuje się dużą efektywnością, co potwierdziły mało skomplikowane i szybkie obliczenia na danych empirycznych z zastosowaniem utworzonej funkcji podobieństwa, dające konkretne wartości liczbowe podobieństwa do wzorca.

Ustalenie podczas obliczeń zerowego przyrostu dla wzorca stopy zwrotu dawało odszukiwanie notowań o najmniejszych wzrostach i spadkach cen jednocześnie. Kolejno wzorzec o stopie zwrotu ustalonej na wartość najmniejszą z występujących w szeregu czasowym wyznaczał okresy największych spadków, a wzorzec o ustalonej maksymalnej stopie zwrotu wyznacza wzrosty największe. W każdym jednak przypadku rezultaty obliczeń, a zwłaszcza rankingi ciągów notowań zależą od ustalonej metryki.

Mając na względzie to, iż liniowa zależność występująca w funkcji podobieństwa lokalnego prawdopodobnie słabo odróżnia wartości wzrostów istotnie różne od wzorca, można zweryfikować zaostrenie kryterium wyszukiwania przebiegów przez zastosowanie funkcji podobieństwa lokalnego, opartej na kwadracie różnicy wartości zwrotu [Richter 1995].

Wartości wektora wzorcowego tworzą liniowy trend wzorcowy. Można również zastosować dowolną inną funkcję, aby odkrywać przebiegi do niej podobne, np. mogą być to szeregi uznawane w analizie technicznej jako wzorce zachowań kursów.

Wykrywanie podobieństw dwóch różnych szeregów czasowych, np. cen akcji dwóch spółek, może dać istotne informacje, w których odcinkach czasowych przebiegi te miały podobne do siebie wartości. Podobieństwo może istnieć jednak nie tylko w rozumieniu wartości bezwzględnych, może się ono ujawniać w porównaniach przebiegów wartości względnych albo przy porównaniach przebiegów przeskalowanych w dziedzinie czasu lub wartości cen akcji. Innym zastosowaniem może być badanie pewnego rodzaju funkcji samopodobieństwa ciągu cen akcji dla dwóch okresów w obrębie notowań tej samej spółki. Na podstawie wartości historycznych notowań cen akcji, za pomocą opracowanej adaptacji metody CBR, można wysnuwać hipotezy o przyszłych przebiegach wartości cen akcji.

## PIŚMIENNICTWO

- Gung Wei Chirn, 1997: Pattern discovery in sequence databases: algorithms and applications to DNA/protein classification. Thesis, Department of Computer and Information Science, New Jersey Institute of Technology, Newark.
- Hellmström K., 1998: Predicting the Stock Market, Technical Report IMA-TOM-1997-07, Department of Mathematics and Physics, Mälardalen University, Sweden.  
[http://www.ima.mdh.se/tom/tom-prints/print-ima\\_tom-1997\\_07.ps](http://www.ima.mdh.se/tom/tom-prints/print-ima_tom-1997_07.ps)
- Kolodner J., 1993: Case-Based Reasoning. Morgan Kaufman Publishers Inc. San Francisco, CA.
- Krawczak M., Miklewski A., Jakubowski A., Konieczny P., 2000: Zarządzanie ryzykiem inwestycyjnym, wybrane zagadnienia. Polska Akademia Nauk, Instytut Badań Systemowych.
- Lenz M, Bartsch B, Burkhard H-D., Wess S. (red.), 1998: Case-Based Reasoning Technology, From Foundations to Applications. Lecture Notes in Computer Science, Springer, Heidelberg.
- Pring J.M., 1998: Podstawy analizy technicznej. WIG Press, Warszawa.



- Richter M., 1995: The Knowledge Contained in Similarity Measures. Invited talk at ICCBR'95, Portugal. <http://www.cbr-web.org/documents/Richtericebr95remarks.html>.
- Rutkowska D., Piliński M., Rutkowski L., 1997: Sieci neuronowe, algorytmy genetyczne i systemy rozmyte. Wydawnictwo Naukowe PWN, Warszawa.
- Shivakumar N., Garcia-Molina H., 1995: SCAM: A copy detection mechanism for digital documents. 2nd Annual Conference on the Theory and Practice of Digital Libraries, Austin, TX.
- Slade S., 1991: Case-based Reasoning for Financial Decision Making. First International Conference on Artificial Intelligence Applications on Wall Street, New York, NY.
- Watson I., 1999: Case-based reasoning is a methodology not a technology. Knowledge-Based Systems 1999; 4: 303-308.

## **LINEAR TREND DISCOVERY IN STOCK RETURNS USING ARTIFICIAL INTELLIGENCE CASE-BASED METHOD**

**Abstract.** Relative values are reliable stock comparison tool. A stock value linear pattern was created, basing on one session return rate. Adaptation of Case-Based Reasoning method was developed for discovering sequences of session return rates, most similar to the linear pattern. The used local and global similarity functions were described. Empirical data covered 1130 close values of a Polish stock market main telecommunication company share. Two to seven session long session sequences, with highest similarity to linear model pattern with given increment value were discovered in the calculations.

**Key words:** trend discovery, time series, Case-Based Reasoning.

*Andrzej Kluza, Katedra Ekonometrii i Informatyki, Szkoła Główna Gospodarstwa Wiejskiego w Warszawie, ul. Nowoursynowska 161, 02-787 Warszawa  
ak@pancake.sggw.waw.pl*

