

Arkadiusz BANASIK
Politechnika Śląska
Wydział Organizacji i Zarządzania
Instytut Ekonomii i Informatyki

INNOWACYJNE PODEJŚCIE DO ANALIZY DANYCH NA PRZYKŁADZIE SYSTEMU WSPOMAGANIA DECYZJI DLA INWESTORA

Streszczenie. Inwestowanie ewoluuje wraz z rozwojem rynków i możliwości inwestowania. Konsekwencją takiego stanu rzeczy jest konieczność tworzenia odpowiednich narzędzi do podejmowania właściwych decyzji. Prezentowany system wspomagania decyzji dla inwestora opiera się na procesie pozyskiwania wiedzy z baz danych, wspartym logiką rozmytą i sieciami neuronowymi. Zastosowanie tego typu rozwiązania stanowi innowację w dziedzinie inwestowania.

INNOVATIVE APPROACH IN DATA-MINING ON EXAMPLE OF DECISION SUPPORT SYSTEM FOR INVESTOR

Summary. Investment evolves In order to market expansion and New possibilities to Invest. That situation persist on development of suitable tools to good decision-making. Presented decision support system is based on Knowledge Discovery in Databases process supported by fuzzy logic and neural networks. Use of that kind of solution is an innovation in field of investment.

1. Wprowadzenie

Inwestowanie od zawsze kojarzy się z angażowaniem własnych środków finansowych inwestora i niezaprzeczalnie wiąże się z ponoszeniem przez niego ryzyka. Takie podejście powodowało wyszukiwanie metod i aparatu wspomagającego podejmowanie decyzji inwestycyjnych. W ostatnim okresie inwestowanie stało się polem do zastosowania inteligencji obliczeniowej (ang. *Computational Intelligence*).

Zastosowanie inteligencji obliczeniowej wraz z podejściem procesowym w odkrywaniu wiedzy, wspomagane metodami i technikami sztucznej inteligencji, jest innowacją w podejściu do inwestowania. Innowacyjność polega na połączeniu wspomnianych wcześniej podejść i stanowi alternatywę dla klasycznych metod inwestowania.

Z raportu opublikowanego przez Aite Group wynika, iż w 2010 roku strumienie przepływów pieniężnych na giełdach – 50% amerykańskiego, 28% europejskiego i 16% azjatyckiego – będą przeprowadzane automatycznie przez stworzone algorytmy transakcyjne (ang. *trading algorithms*) [3].

Wyzwaniem, z którym należy się zmierzyć, jest zagadnienie zamodelowania rynku finansowego. Problemem, który stanowi największe wyzwanie, jest wpływ wielu czynników, takich jak: stopy procentowe, kursy walut, stopa wzrostu, płynność finansowa i wiele innych. Podstawą problemów (poza wielością czynników) jest to, iż nie ma „twardych” założeń teoretycznych co do zmian na rynku, a zmiany występujące mają charakter nieliniowy, niestacjonarny i odłożony w czasie [2, 9, 10].

W takich sytuacjach konieczne jest wsparcie analizy rynku opiniami ekspertów oraz ich wiedzą z zakresu skutków decyzji podejmowanych w ramach dozwolonych prawem działań państwa oraz inwestorów. Takie podejście wiąże się z koniecznością włączenia do analiz elementów wnioskujących na podstawie wiedzy eksperckiej.

W takim ujęciu konieczne jest zastosowanie elementów sztucznej inteligencji (ang. *Artificial Intelligence*), które pozwolą na analizę zarówno danych historycznych (danych ilościowych), jak i danych wyrażonych przez ekspertów (dane w postaci lingwistycznej, często jakościowe) [5, 6, 7].

Niniejszy artykuł przedstawia propozycję wspomagania inwestora i jego decyzji przez rozmyty system wnioskujący, wspomagany zastosowaniem sieci neuronowych do analizy danych i ich grupowania. Podstawą do takiego założenia są innowacyjne podejścia do analizy danych (ang. *data-mining*) i pozyskiwania wiedzy.

2. Metody odkrywania wiedzy w bazach danych

Odkrywanie wiedzy w bazach danych opiera się na metodach eksploracji danych (ang. *data-mining*). Zagadnienie nieustannie ewoluuje. Zainteresowanie nim wkracza do takich obszarów badawczych, jak: uczenie maszynowe, rozpoznawanie obrazów, bazy danych, statystyka, sztuczna inteligencja, pozyskiwanie wiedzy do systemów ekspertowych, wizualizacji danych oraz wysokowydajnych systemów obliczeniowych [1]. Zadaniem jednoczącym wszystkie te dziedziny nauki jest pozyskiwanie wiedzy wysokiego poziomu z niskopoziomowych danych w kontekście dużych zbiorów danych.

Proces pozyskiwania wiedzy jest procesem interaktywnym i iteracyjnym, który składa się z wielu kroków, z wieloma decyzjami podejmowanymi przez użytkownika, a podstawowe jego etapy prezentują się następująco [1]:

1. Tworzenie i rozumienie podstaw tworzonej aplikacji i założeń dotyczących wiedzy oraz identyfikacja celu tworzenia system pozyskiwania wiedzy w bazach danych z punktu widzenia zamawiającego.
2. Tworzenie docelowego zbioru danych: wybór zbioru danych, tworzenie podzbioru zmiennych lub próbek danych, na których odkrywanie ma być prowadzone.
3. Wstępne przetwarzanie i oczyszczanie danych – operacje usuwające zakłócenia, procedury dotyczące niepełnych lub uszkodzonych danych itp.
4. Redukcja danych i ich projekcja – ze szczególnym uwzględnieniem celu pozyskiwania wiedzy w zadaniu.
5. Łączenie wytycznych z kroku pierwszego z metodami eksploracji danych.
6. Analiza eksploracji, wybór modeli i hipotez do wyszukiwania wzorców danych.
7. Eksploracja danych – wyszukiwanie wzorców, które interesują nas z punktu widzenia formy reprezentowania przez stosowanie: reguł klasyfikacji, drzew decyzyjnych, regresji i klasteringu.
8. Interpretacja odkrytych wzorców – możliwy powrót do każdego z wcześniejszych kroków.
9. Korzystanie z odkrytej wiedzy w praktyce lub przedstawienie rezultatów przeprowadzonego procesu.

Proces pozyskiwania wiedzy może składać się z większej liczby iteracji, a pomiędzy dowolnymi dwoma krokami może zawierać sprzężenia zwrotne.

Zaprezentowane w procedurze pozyskiwania wiedzy kroki mogą zostać uzupełnione przez zastosowanie metod i technik sztucznej inteligencji, a w szczególności przez sieci neuronowe do klasyfikacji oraz logikę rozmytą do wyrażania zmiennych w języku naturalnym.

Zastosowanie procesu pozyskiwania wiedzy w bazach danych na potrzeby inwestowania może być w szeroki sposób wspierane przez zastosowanie rozmytych reguł i zapytań w inwestycyjnych bazach danych.

3. Logika rozmyta w systemie wnioskującym

Inwestowanie wiąże się z podejmowaniem decyzji na podstawie informacji i wiedzy inwestora. Podejmowanie decyzji jest czynnością złożoną. Może być określone jako proces wyboru z wielu alternatyw konkretnej alternatywy. Jest czynnością, która poprzedzona jest właściwą oceną ich wszystkich [4].

Istnieją dwie teorie dotyczące zachowań inwestora [4]:

- Teoria sprawnego działania – sugerująca, iż inwestorzy działają racjonalnie i rozważają całą dostępną informację w procesie podejmowania decyzji, a w wyniku tego rynek jest efektywny.
- Teoria behawioralna – zakłada ona, iż istnieje wiele dowodów na to, że inwestor działa w sposób nieracjonalny i popełnia powtarzające się błędy w ocenie sytuacji; wskazuje na próby lepszego dopasowania się do potrzeb inwestora przez lepsze zrozumienie emocji i błędów poznawczych oraz ich wpływu na decyzje inwestorów.

Każde z prezentowanych podejść jest odmienne, lecz oba mogą być skuteczne w pozyskiwaniu wiedzy. Podejściem łączącym obie teorie jest stworzenie bazy wiedzy, opierającej się na analizie danych historycznych oraz na wiedzy ekspertów. W odniesieniu do pierwszego podejścia proces pozyskiwania wiedzy w bazach danych może być skuteczny, jednak tradycyjna logika nie jest w stanie stworzyć narzędzi do przetwarzania stwierdzeń w języku naturalnym. Dlatego też konieczne jest zastosowanie narzędzia, które sobie z tym poradzi – logiki rozmytej (ang. *Fuzzy Logic*).

Głównym celem prezentowanych podejść jest doradzenie klientom (inwestorom), jak alokować środki w określone papiery wartościowe [9].

Zastosowanie języka naturalnego i metod jego analizy są rozwiązaniami pozwalającymi na stosowanie informacji udostępnionej dla inwestora przez ekspertów w bazach danych. Prezentowana koncepcja opiera się także na dokładnych danych i zapytaniach zastosowanych w modelu.

Zbiór rozmyty jest obiektem, który charakteryzowany jest przez jego funkcję przynależności. Funkcja ta opisuje każdy obiekt w zbiorze i zawiera się w przedziale od zero do jeden. Funkcja przynależności (charakterystyczna) jest stopniem przynależności danego obiektu we wspomnianym zbiorze [5].

Konieczne jest określenie znaczenia funkcji przynależności [7]. Jest ona bazą gromadzenia obiektów w zbiorze rozmytym. Stanowi także podstawę budowy reguł ograniczeń itp. Odpowiednie jej zrozumienie i zastosowanie stanowią podstawę podejścia rozmytego.

Wszystkie metody, bazujące na podejściu rozmytym, z założenia definiowane są jako narzędzia odpowiednie do działania z wiedzą, informacją i danymi [6].

Informacja powinna być grupowana do odpowiedniego przetworzenia (do określonych celów). Najlepszym narzędziem do jej gromadzenia są bazy danych [8]. Jest to potwierdzenie słuszności podjęcia tej tematyki.

4. Eksperyment

W niniejszym artykule analizie poddano dane historyczne z giełdy w Szanghaju. Stworzono system rozmyty wspomagany wykorzystaniem sieci neuronowych, które posłużyły do uczenia systemu. Jednocześnie dały one możliwość uogólniania tych czynników, których wpływ na giełdę był konieczny do uwzględnienia, a skomplikowany w odwzorowaniu (do tego celu użyto sieci neuronowych ART1, które wykorzystywane są w binarnej analizie obrazów).

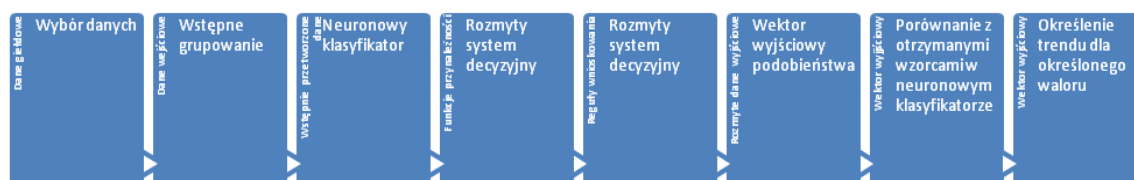
Wymienione uwarunkowania wskazują, iż podstawą do modelowania decyzji są dane historyczne pogrupowane w określony sposób i poddane dalszemu procesowi analizy i przetwarzania. Dane pochodzą z jednego roku działalności giełdy.

Specyfika rynku azjatyckiego została wybrana ze względu na odpowiedni układ danych historycznych.

Podstawowym efektem, jaki ma przynieść inwestorowi zastosowanie rozmytego systemu wnioskującego, jest prognoza trendu dla wybranych papierów wartościowych.

Procedurę realizacji przedstawia rys. 1, a w dokładniejszym ujęciu prezentuje się ona następująco:

1. Z danych historycznych utworzono szeregi czasowe na wejściu systemu.
2. Podjęto zadanie klasyfikacji danych wejściowych na przyjętych 9 wzorców.
3. Stworzono 3-warstwową sieć neuronową do analizy wstępnie pogrupowanych (przetworzonych) wzorców.
4. Na podstawie zbiorów uczących stworzono rozmyte reguły, opierając się na wcześniej zdefiniowanych wzorcach.
5. Przeprowadzono symulację trendów na podstawie zbiorów testowych.
6. Podjęto się procesu wyostrażania uzyskanych wyników, aby końcowe efekty były wyrażone w sposób ostry (w postaci wyników, które nie są rozmyte).



Rys. 1. Proces realizacji eksperymentu

Fig. 1. Realization of process of experiment

W ramach tak skonstruowanego procesu realizacji rozmytego systemu decyzyjnego doprowadzono do finalnego określenia prognozy trendów w badanych grupach papierów wartościowych.

Aby powyższe ustalenia mogły mieć miejsce, konieczne było zastosowanie odpowiedniego zestawu metod i technik sztucznej inteligencji oraz operacji, które składają się na realizację procesu wnioskowania. Do najważniejszych należy zaliczyć:

- Konstrukcję zbiorów danych wejściowych na podstawie średnich ruchomych i odchylenia standardowego z 5 dni.
- Sieć neuronową, składającą się z jednej podsieci o 20 neuronach wejścia, 30 neuronach ukrytych i 9 neuronach wyjściowych.
- Stworzenie odpowiedniego zbioru treningowego, składającego się z 70 wzorców, po 20 wartości każdy, obejmujących okres pięciu dni aktywności na giełdzie.
- Zastosowanie uczenia bez nadzoru dla stworzonej sieci, wskazanej wcześniej.
- Zastosowanie sieci ART1 do określenia reguł rozmytych, która obejmowała 15 neuronów wejściowych i 9 wyjściowych.
- Zastosowanie 30 reguł (stworzonych na podstawie uczenia sieci i reguł eksperckich) do procesu treningowego sieci ART1.
- Zastosowanie w procesie wyostrzania (defuzyfikacji) schematu największej wartości funkcji przynależności.

Procedura realizacji eksperymentu, jak widać, była bardzo złożona i wymagała zastosowania wyselekcjonowanych (w zależności od potrzeb) metod i technik sztucznej inteligencji. Kluczowym aspektem było (jak już wspomniano) wyznaczenie trendu na podstawie danych historycznych, który ma służyć jako narzędzie pomocnicze w realizacji procesu inwestycyjnego.

Odpowiednia konfiguracja przedstawionego systemu może być przedmiotem adaptacji do warunków innych giełd w innych rejonach świata, co powoduje, iż mechanizm dostosowywania będzie opierał się na odpowiednim doborze zbiorów treningowych i testowych.

5. Wyniki eksperymentu

W wyniku realizacji przedstawionych wcześniej metod i technik sztucznej inteligencji otrzymano następującą funkcję, którą poddano analizie:

$$* x = \log\left(\frac{C_1}{C_2}\right), \quad (1)$$

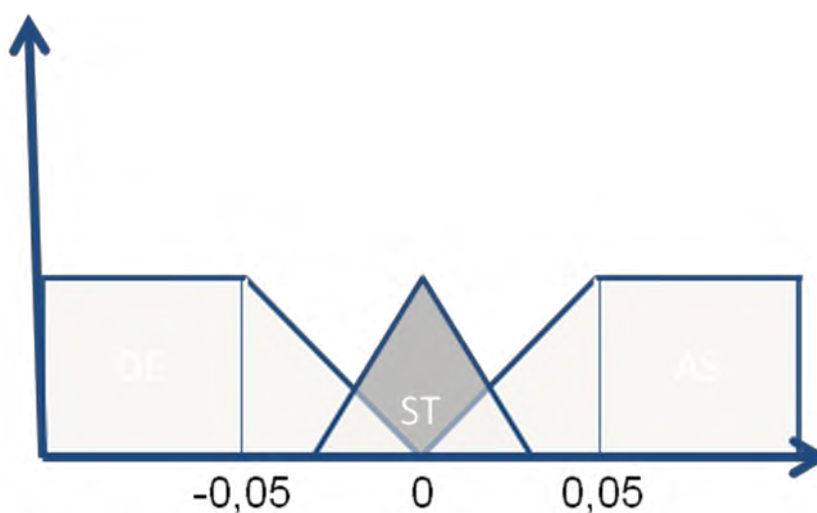
gdzie:

C1 – cena dnia,

C2 – cena przez kolejny tydzień.

Po przeprowadzonych symulacjach doprowadzono do otrzymania 3 reguł, które wyznaczają trendy (graficzna prezentacja na rys. 2):

- rosnący (AS) przy wartości funkcji +1,
- malejący (DS) przy wartości funkcji -1,
- stały (ST) przy wartości funkcji 0.



Rys. 2. Wynik wnioskowania rozmytego systemu

Fig. 2. Result formulated by fuzzy system

Rysunek 2 przedstawia wizualizację wyników, ale przed procesem wyostżenia. Tak przedstawione wyniki ukazują, do jakich uogólnień i prostych do analizy rezultatów może doprowadzić umiejętnie zastosowany aparat matematyczno-informatyczny.

6. Wnioski i kierunki dalszych prac

Przeprowadzony proces wnioskowania doprowadził do ograniczenia liczby reguł w systemie z pierwotnych 30 do 3 w finalnej części, co możliwe było w wyniku zastosowania

algorytmów logiki rozmytej i w konsekwencji doprowadziło do znacznego zmniejszenia stopnia złożoności problemu.

Najważniejszą cechą zastosowanych algorytmów logiki rozmytej był efekt otrzymania wartości ostrej w finalnym stadium realizacji procesu wnioskowania, co jest kluczowym aspektem dla każdego inwestora, który nie musi znać się na zagadnieniach rozmytych.

Poprawność przedstawionego wnioskowania potwierdzają wyniki przeprowadzonych symulacji, które określają, iż prawidłowa ocena trendów określana jest: na poziomie 92% w zbiorze treningowym i na poziomie 74% w zbiorze testowym.

Analiza danych wejściowych pozwoliła na stwierdzenie, iż zbudowany system doskonale się sprawdza, gdy część danych na wejściu ma charakter ciągły.

Poziom skomplikowania sytuacji na giełdzie wymaga precyzyjnego opisu matematycznego zjawisk tam zachodzących, co przy poziomie złożoności nie jest w 100% możliwe. Dlatego też lukę w opisie matematycznym wypełnia system wspomaganie decyzji.

Bazy wiedzy powstają w wyniku połączenia informacji ilościowych i jakościowych, a dzięki zastosowaniu logiki rozmytej – wyrażonych w języku naturalnym (zmienne lingwistyczne). Takie podejście powoduje, iż reguły generowane przez system są lepiej przyswajalne – inwestorzy mają poczucie, iż nad nimi pracowali eksperci.

Tworzenie systemów hybrydowych stanowi powszechną tendencję w obecnych, trudnych do właściwej oceny (burzliwych) czasach. Tworzenie konkretnych zestawów narzędzi jest innowacją, której przydatność i funkcjonalność jest weryfikowana zarówno przez użytkownika (klienta), jak i zmieniającą się sytuację na rynku.

Kolejnym krokiem, który należy podjąć przy analizie zagadnienia, jest przeniesienie systemu do warunków polskich oraz sprawdzenie możliwości zastosowania teorii możliwości i efektów, jakie taka modyfikacja przyniesie.

BIBLIOGRAFIA

1. Fayyad U., Piatetsky-Shapiro G., Smyth P.: From Data Mining to Knowledge Discovery in Databases. AI Magazine, American Association for Artificial Intelligence, Fall 1996.
2. Barabazon A., O'Neill M., Dempsey I.: An Introduction to Evolutionary Computation in Finance. IEEE Computational Intelligence Magazine, Vol. 3, No. 4, November 2008.
3. Algos 3.0: Development in Algorithmic Trading. Traders Magazine 2007. Special Report. SourceMedia's Custom Publishing Group.
4. Khcherem F., Bouri A.: Fuzzy Logic and Investment Strategy. Global Economy & Finance Journal, Vol. 2, No. 2, September 2009.
5. Zadeh L.: Fuzzy Sets. Information and Control, No. 8, 1965.

6. Zadeh L.: Fuzzy Sets as a Basis for Theory of Possibility. Fuzzy Sets and Systems, No. 1, 1978.
7. Łęski J.: Systemy Neuronowo-rozmyte. WNT, Warszawa 2008.
8. Uszynski M.: Fuzzy queries with linguistic quantifiers for information retrieval from data bases. Technical report CSD-87-333. University of California, Berkeley 1980.
9. Bojadziev G., Bojadziev M.: Fuzzy Logic for Business, Finance, and Management. World Scientific Publishing Co. Pte. Ltd., Singapore 2007.
10. Deboeck G. (ed.): Trading on the edge. Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets. John Wiley and Sons, Toronto 1994.

Recenzent: Prof. dr hab. inż. Franciszek Marecki

Abstract

Every source of information and its transformation into knowledge is useful for investor. That approach provides the key aspect of this paper. Information for investor is expressed in natural language or investor uses queries in natural language to obtain it. Evaluation of assets is the clue for investor which of them to buy. Creation of knowledge base for investor is a way that should be followed. Presented paper indicates the first step into that aim – shows applicable mathematical apparatus to cope with natural language statements in databases and knowledge discovery process for the fuzzy system for investor.