

**WYŻSZA SZKOŁA  
INFORMATYKI STOSOWANEJ  
I ZARZĄDZANIA**



**ANALIZA SYSTEMOWA  
W FINANSACH I ZARZĄDZANIU**

**Wybrane problemy**

Pod redakcją  
**Jerzego HOŁUBCA**

**WYŻSZA SZKOŁA  
INFORMATYKI STOSOWANEJ  
I ZARZĄDZANIA**

**ANALIZA SYSTEMOWA  
W FINANSACH I ZARZĄDZANIU  
Wybrane problemy**

Pod redakcją  
**Jerzego HOŁUBCA**

Warszawa 1999

**Wykaz opiniodawców artykułów zamieszczonych w tomie:**

prof. dr hab. Jerzy **HOLUBIEC**

prof. dr hab. Janusz **KACPRZYK**

prof. dr hab. Tadeusz **NOWICKI**

prof. dr hab. Stanisław **PIASECKI**

prof. dr hab. Piotr **SZCZEPANIAK**

prof. dr hab. Tadeusz **TRZASKALIK**

doc. dr hab. Sławomir **WIERZCHOŃ**

doc. dr hab. Leszek **ZAREMBA**

© **Wyższa Szkoła Informatyki Stosowanej i Zarządzania**

**Warszawa 1999**

**ISBN 83-85847-24-3**

## Przedmowa

Na niniejszą publikację składa się zbiór prac doktorantów Zaocznych Studiów Doktoranckich "Informatyka w zarządzaniu i finansach" działających przy *Instytucie Badań Systemowych Polskiej Akademii Nauk*.

Prace te były referowane na konferencji BOS'98 "Rozwój średnich i małych miast w XXI wieku w Polsce: Rola badań operacyjnych i systemowych", Kutno, 8-10 czerwca 1998 r.<sup>1</sup>, a także na seminariach Studiów Doktoranckich "Informatyka w zarządzaniu i finansach". Nad stroną merytoryczną publikacji czuwał Pan Prof. dr hab. Jerzy Hołubiec oraz grono recenzentów i opiekunów naukowych doktorantów.

Prace dotyczą głównie problemów analizy systemowej oraz jej zastosowań w dziedzinie finansów, a zwłaszcza - teorii portfela, obligacji i problemów inwestycyjnych. Niektóre prace przy analizie finansowej posługują się tzw. algorytmami genetycznymi i sieciami neuronowymi, a także modelowaniem rozmytym i strukturami fraktalnymi. Część prac dotyczy zarządzania i sterowania produkcją.

Wypada zauważyć, iż doktoranci Studiów atakują w swych pracach tematy nowoczesne i znajdujące się w obszarze tzw. frontu badawczego analizy systemów. Wypada im życzyć sukcesów i wytrwałości w pracy, która winna zakończyć się obronioną pracą doktorską.

---

<sup>1</sup> Głównymi organizatorami konferencji było Polskie Towarzystwo Badań Operacyjnych i Systemowych oraz Instytut Badań Systemowych PAN.

Wypada także zaznaczyć, iż wydanie niniejszej publikacji stało się możliwe dzięki wsparciu finansowemu ze strony *Wyższej Szkoły Informatyki Stosowanej i Zarządzania*, działającej w ramach Fundacji Krzewienia Nauk Systemowych. Fundacja ta została założona w 1991 roku z inicjatywy Prof. L. Kuźnickiego, wówczas Sekretarza Naukowego Polskiej Akademii Nauk. Do zadań Fundacji należy, między innymi, wspieranie i promocja prac młodych pracowników nauki, a zwłaszcza prac doktorantów.

Mamy nadzieję, iż publikacja niniejsza zostanie życzliwie przyjęta przez specjalistów działających w obszarze nauk systemowych.

Rektor WSISiZ  
Prof. Roman Kulikowski

# ZASTOSOWANIE SIECI NEURONOWYCH DO PROGNOZOWANIA WPLĄT SKŁADEK W KONTYNUOWANYCH UBEZPIECZENIACH NA ŻYCIE

Jarosław KASZUBA  
*Studia doktoranckie IBS PAN*

## 1. Wstęp

Sektor finansowy zacieśnia coraz bardziej współpracę z ośrodkami naukowymi w celu pozyskania nowych metod i narzędzi badawczych umożliwiających sprawne zarządzanie kapitałem. Coraz częściej okazuje się, że stosowanie technik statystycznych staje się niewystarczające. Wynika to często z ograniczonych możliwości statystyki. Bywa, że stosowanie innych technik upraszcza uzyskiwanie poprawnego wyniku. Przykładem jest metoda regresji wielokrotnej, która przekształcona przez Yoh-Han Pao w sieć neuronową stała się bardzo skutecznym modelem odwzorowania funkcjonalnego [1].

## 2. Opis problemu

Jednym z rodzajów ubezpieczeń stosowanych przez towarzystwa są ubezpieczenia pracownicze i rodzinne. Ubezpieczenia te dotyczą aktywnych zawodowo osób i pozwalają, w zależności od szczegółowych warunków umów, ubezpieczyć pracowników i ich rodziny od następstw nieszczęśliwych wypadków lub śmierci. W pewnych przypadkach towarzystwa ubezpieczeniowe wypłacają świadczenia z tytułu urodzenia się dziecka lub nawet związku małżeńskiego. Ubezpieczenie takie dotyczy jednak tylko osób czynnych zawodowo. Wy-

płacane kwoty świadczeń, jak i wpłacane wcześniej składki są jednakowe dla wszystkich pracowników i system nie zezwala na ich indywidualizowanie. Po zakończeniu okresu pracy, każda z osób traci prawo do tego typu ubezpieczeń. Może jednak zawrzeć nowy rodzaj umowy, zwany ubezpieczeniem kontynuowanym. Już w tym momencie można zmienić wielkość opłacanych składek i związanych z tym sum ubezpieczenia. Można jednak także w przyszłości dokonać tego typu korekt. Dzisiejsze możliwości lokowania nadwyżek finansowych nakazują racjonalnie gospodarującym dyrektorom Inspektoratów przewidywanie przepływów pieniężnych. Temu celowi służyć będzie między innymi przewidywanie okresów na jakie ubezpieczeni dokonają wpłaty składek. Nie ma racjonalnej metody określenia, kto z wielu tysięcy ubezpieczonych, na jaki okres i kiedy dokona opłaty swojego ubezpieczenia. Znaczna część osób opłaca składki tylko na dany miesiąc. Inni płacą raz na kwartał. Bywają i tacy, którzy opłacają raz do roku. W praktyce nie stosowano dotąd żadnych racjonalnych metod szacowania wymienionych wartości. W pracy podjęto próbę zastosowania sieci neuronowych do predykcji zachowań ubezpieczonych na podstawie ciągu informacji znanych z przeszłości. Mając dane wartości zmiennej  $x$  w  $N$  odstępach czasowych poprzedzających predykcję:  $x_{(k-1)}$ ,  $x_{(k-2)}$ , ...,  $x_{(k-N)}$  sieć neuronowa ma za zadanie wyliczyć, jaka będzie estymowana wartość  $x_{(k)}$  w chwili  $k$ .

### 3. Przygotowanie danych

#### Dane wejściowe

Mimo, że omawiane ubezpieczenie funkcjonuje już wiele lat, to program zbierający dane w bazach został wdrożony w 1994, a pełen zbiór danych dostępny jest dopiero od 1995 roku.

Z kilkudziesięciu dostępnych parametrów wybrano następujące dane wejściowe do sieci uważając, że są one szczególnie istotne:

Płeć ubezpieczonego	Przypuszcza się, że to czy ubezpieczony jest kobietą czy mężczyzną może mieć znaczenie dla częstotliwości opłaty składek.
Wiek ubezpieczonego	Zakłada się, że wraz z wiekiem ograniczają się możliwości lokomocyjne ubezpieczonych, a jednocześnie pogarsza się jego sytuacja materialna.
Okres ubezpieczenia	Aczkolwiek w tym rodzaju ubezpieczenia nie tworzy się indywidualnego konta, z którego można odzyskać wpłacone składki, to jednak długi okres ubezpieczenia może kształtować pewne nawyki.
Aktualna składka	Wielkość aktualnej składki, a szczególnie jej stosunek do możliwości finansowych ubezpieczonego może w znaczący sposób obciążać domowy budżet.
Stopień pokrewieństwa uposażonego (czyli osoby otrzymującej odszkodowanie w przypadku śmierci)	Ponieważ świadczenie otrzymuje uposażony, to bliskość rodzinna z ubezpieczonym będzie niewątpliwie istotnym czynnikiem wpływającym na wysokość ubezpieczenia.
Historia częstotliwości dotychczasowych opłat	Przygotowując dane źródłowe, m.in. ze względu na brak baz z większej ilości lat, ograniczono analizę historii do ostatnio dokonywanej wpłaty.

Niewątpliwie można by uzupełnić wyżej wymienionych dane innymi informacjami zwiększającymi naszą wiedzę o ubezpieczonym, jego przychodach i kosztach utrzymania, sytuacji rodzinnej itp. Np. mogła to być informacja o stosunku płaconej składki do możliwości finansowych w rodzinie ubezpieczonego, czy o ilości osób w rodzinie. Jednak obowiązująca od 1998 roku ustawa o ochronie danych osobowych zawęży znacznie możliwości ubezpieczyciela w pozyskiwaniu



takich danych. Okazuje się jednak, że nawet te dane stają się niezwykle pomocne w uzyskaniu interesujących rezultatów. Kolejne testy budowanych sieci pozwalają ponadto ocenić przydatność wymienionych w powyższej tabelce cech.

### **Dane wyjściowe**

Danymi wyjściowymi z sieci będzie wyrażony w miesiącach okres opłacenia składki. Przy ustaleniu wielkości wpłacanej składki miesięcznej można przewidzieć kwoty wpłacane w danym miesiącu przez ubezpieczonych.

### **Normalizacja danych**

Wszystkie powyższe dane będą podlegać normalizacji. W przypadku idealnym powinny one leżeć w przedziale  $[-1,1]$ , bądź  $[0,1]$ . Długość wektora wejściowego powinna być taka sama dla wszystkich przypadków uczących i testujących. Najczęściej przyjmuje się, że długość ta wynosi jeden. Dodatkowo powinno się dążyć do uzyskania pokrycia całego przedziału przez każde z wejść sieci.

W celu ograniczenia nie zawsze pożądanego zachowania się sieci w pobliżu punktu o wartości 0, przyjęto, że dane wejściowe będą rzutowane na przedział  $[0.1,0.9]$ . Normalizację każdego z wejść należy zacząć od określenia jego najmniejszej i największej wartości. Należy w tym celu skonstruować, z reguły linowe przekształcenie, które największe wartości wejściowe zrzutuje na prawy koniec wybranego przedziału, a najmniejsze na lewy. W tym miejscu za T. Mastersem [2] można podać przykład takiego przekształcenia.

Oznaczenia:

$A_{\max}$  - największa wartość z przedziału normalizacji (tutaj jest to liczba 0.9);

$A_{\min}$  - najmniejsza wartość z przedziału normalizacji (tutaj 0.1);

$V_{\max}$  - największa wartość podlegająca normalizacji;

$V_{\min}$  - najmniejsza wartość podlegająca normalizacji;

$V$  - aktualna wartość podlegająca skalowaniu;

$A$  - szukana wartość  $V$  otrzymana w wyniku rzutowania na przedział  $[A_{\min}, A_{\max}]$ .

W celu uproszczenia wzoru, oblicza się współczynnik  $r$ :

$$r = \frac{A_{\max} - A_{\min}}{V_{\max} - V_{\min}}$$

Teraz :

$$A = rV + (A_{\min} - r V_{\min})$$

W powyższy sposób przygotowano dane dotyczące wysokości sumy ubezpieczenia.

Jako wartości graniczne  $V_{\min}$  i  $V_{\max}$  wybrano najmniejszą i największą dopuszczalną przepisami sumę ubezpieczenia. Podobnie postąpiono z wiekiem. Dolną granicą zakresu jest w tym przypadku 18 lat. Górną ustalono na poziomie 100 lat, choć w bazie danych nie natrafiono na tak długowieczną osobę. Bywały jednak osoby liczące więcej niż dziewięćdziesiąt lat. Istnieje co prawda prawdopodobieństwo, że w trakcie eksploatacji sieci neuronowej natrafimy na przypadek osoby stuletniej, jednak ze względu na nietypowość takiego przypadku nie musimy zabiegać o jego indywidualne traktowanie. Podobnie jak wiek traktuje się okres ubezpieczenia.

Dane dotyczące płci mają inny charakter. Są to dane dwuwartościowe i taką reprezentację otrzymały też w sieci 0 - kobieta, 1 - mężczyzna. Pozostał jeszcze do rozpatrzenia problem stopnia pokrewieństwa. Zostały tu wyróżnione trzy elementy: bliska rodzina (żona, mąż, dzieci, matka, ojciec) i dalsza (szwagier, ciotka, wujek itp.) Do ostatniej kategorii zaliczono też osoby obce.

Osobnym problemem było właściwe zobrazowanie w sieci okresu zapłaty składki ubezpieczeniowej. Zupełnie nie sprawdzają się pró-

by przyporządkowania progów w przedziale (0.1,0.9) kolejnym miesiącom nadpłaty. W takim przypadku wartość 0.1 odpowiadałaby brakowi przedpłaty, wartość 0.2 to nadpłata za jeden miesiąc itp.

Pierwszym, w sposób naturalny narzucającym się pomysłem jest przedstawienie ciągu dwuwartościowego, np. zero-jedynkowego z jedną jedynką na pozycji odpowiadającej numerowi miesiąca, do którego dokonano opłaty składki.

I tak

$\{0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\}$  oznacza brak przedpłaty,

$\{1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0\}$  oznacza jednomiesięczną przedpłatę

.....

$\{0,0,0,0,0,0,0,0,0,0,0,0,0,1,0,0\}$  (jedynka na 12-tej pozycji) oznacza przedpłatę na 12 miesięcy

$\{0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,0\}$  (jedynka na 13-tej pozycji) oznacza przedpłatę 13 miesięcy i więcej.

Jednak taka reprezentacja liczbowa jest bliższa zadaniu klasyfikacji lub rozpoznawania obiektów niż zadaniu predykcji w oparciu o aproksymację procesu. Potwierdza to praktyka obliczeniowa, por. [2]. Okazuje się, że powstający w trakcie uczenia sieci błąd średniokwadratowy (będący podstawą uczenia się sieci) niewiele się zmieniał, gdy ciąg  $\{0,1,0,0,0,0,0,0,0,0,0,0,0,0,0,0\}$  (z jedynką na drugim miejscu) był przekształcony na ciąg  $\{0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0\}$  (z jedynką na czwartym miejscu). Pierwszy z przypadków miał symbolizować dokonanie przedpłaty na dwa miesiące do przodu, a drugi - przedpłatę na cztery miesiące. W trakcie testów zmieniono wartości 0 na 0.1 i 1 na 0.9

Zastosowano następnie rozwiązanie polegające na wypełnianiu ciągu wartościami 0.9 począwszy od lewej strony ustalając wartości pozostałych elementów na 0.1.

Otrzymano następującą konwencję:

{0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1}

oznacza brak przedpłat,

{0.9, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1}

(0.9 tylko na pierwszej pozycji) oznacza jednomiesięczną przedpłatę

{0.9, 0.9, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1}

(0.9 na pierwszych 2-ch pozycjach) oznacza przedpłatę na 2 miesiące

.....

.....

{0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.1}

(0.9 na pierwszych 12-tu pozycjach) oznacza przedpłatę na 12 miesięcy

{0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9, 0.9}

(0.9 na 13-tu pozycjach) oznacza przedpłatę 13 miesięcy i więcej.

Taki sposób kodowania wyników znacznie zwiększył efektywność sieci. Rozwiązano też problem interpretacji wyników. Za dobrze zakwalifikowany przypadek uznano sytuację, gdy na wszystkich pozycjach ciągu różnica między wartościami oczekiwanymi a faktycznymi jest mniejsza od ustalonej wartości *alfa*. Przypadek uznano za błędnie zakwalifikowany, gdy chociaż na jednej pozycji różnica między wartością przeciwną do oczekiwanej a wartością faktyczną jest mniejsza od *alfa*. Za błędny przypadek uznano także sytuację, gdy sieć wygeneruje rozwiązanie, które nie może być zinterpretowane jako liczba miesięcy, (np. {0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.9}) z 0.9 na ostatniej pozycji nie ma interpretacji). Określono też, którą nazwano „bez rozstrzygnięcia”.

Jest to przypadek, gdy choć na jednej pozycji wygenerowana zostanie wartość, która nie może być zakwalifikowana ani jako dobra ani jako zła (np. 0.5, przy  $\alpha = 0.299$ ).

#### 4. Architektura sieci

Jako główną strukturę sieci wybrano sieć wielowarstwową jednokierunkową z jedną warstwą ukrytą i zastosowanym algorytmem propagacji wstecznej, np.[3,4]. Do określenia ilości neuronów w warstwie ukrytej zastosowano *regułę piramidy geometrycznej*. Sugeruje ona, że jeżeli w warstwie wejściowej znajduje się  $n$  neuronów, a w wyjściowej  $m$ , to ilość neuronów w warstwie ukrytej (dla sieci trójwarstwowej) powinna wynosić  $\sqrt{mn}$ . Otrzymana liczba wskazuje jedynie zbliżoną do optymalnej liczbę neuronów ukrytych, natomiast testy sieci powinny pomóc określić prawidłową ich ilość. Uwzględnić należy jednak fakt, że zwiększenie liczby neuronów ukrytych ponad miarę może powodować pogorszenie zdolności uogólniania. W omawianym przypadku  $n$  wynosiło 18, a  $m$  było równe 13. Tworzenie sieci rozpoczęto wykorzystując do tego celu 15 neuronów w warstwie ukrytej. Sprawdzono jednak zachowanie się sieci dla innych przypadków 1, 5, 10 i 20 neuronów w warstwie ukrytej. Otrzymane wyniki potwierdziły słuszność reguły piramidy geometrycznej.

Ze względu na pewną okresowość w opłatach składek dla każdego miesiąca zbudowano osobną sieć.

#### 5. Wyniki i wnioski

W pierwszej fazie tworzenia sieci pominięto dane dotyczące okresu ubezpieczenia oraz wieku ubezpieczających się. Dane dla stycznia 1997 wskazują około 91% zgodności wyników otrzymanych przy zastosowaniu sieci z danymi rzeczywistymi.

Okazuje się, że dodanie pozostałych parametrów (okresu ubezpieczenia i wieku) pozwala zwiększyć efektywność sieci. W tym

przypadku dla tego samego miesiąca uzyskano wynik prawie 93 % . Na łącznie 1970 przypadków sieć poprawnie zakwalifikowała 1830, błędnie 132 przypadki, a nie potrafiła dać jednoznacznego rozstrzygnięcia dla 8 przypadków.

Przeprowadzone eksperymenty świadczą o tym, że

- prawidłowo wybrano wielkości (cechy ubezpieczających się) mające wpływ na interesującą ubezpieczyciela wielkość wyjściową i dobrze przeprowadzono ich normalizację;
- właściwie wybrano rodzaj i architekturę sieci oraz skutecznie przeprowadzono naukę.

Autorowi nie jest znane wcześniejsze zastosowanie sieci neuronowych do zadania przewidywania wysokości składek w kontynuowanych ubezpieczeniach na życie.

Wydaje się, że sieci neuronowe mogą być skutecznym narzędziem w dziedzinie ubezpieczeń i że konieczne są dalsze badania, tak teoretyczne jak i eksperymentalne.

## **Literatura**

- [1] Pao, Yoh-Han (1989). Adaptive Pattern Recognition and Neural Networks. Addison-Wesley Publishing Co., Reading, MA.
- [2] Masters (1996). Sieci neuronowe w praktyce. WNT, Warszawa.
- [3] Tadeusiewicz (1993). Sieci neuronowe. Akademicka Oficyna Wydawnicza RM, Warszawa.
- [4] Korbicz, A. Obuchowicz, D. Uciński (1994). Sztuczne sieci neuronowe; podstawy i zastosowania. Akademicka Oficyna Wydawnicza PLJ, Warszawa.

# **WYŻSZA SZKOŁA INFORMATYKI STOSOWANEJ I ZARZĄDZANIA**

działa pod auspicjami  
**Polskiej Akademii Nauk**

**ZAŁOŻYCIELEM**

**Wyższej Szkoły Informatyki Stosowanej i Zarządzania**  
jest

**FUNDACJA KRZEWIENIA NAUK SYSTEMOWYCH**  
powołana z inicjatywy  
**Prezesa**  
**POLSKIEJ AKADEMII NAUK**

**FUNDATOREM**

**Fundacji Krzewienia Nauk Systemowych**  
jest

**POLSKA AKADEMIA NAUK**

**ORGANEM**

sprawującym nadzór  
jest

**MINISTERSTWO EDUKACJI NARODOWEJ**

**Wyższa Szkoła Informatyki Stosowanej i Zarządzania**  
prowadzi studia wyższe na kierunkach:

**INFORMATYKA**  
**ZARZĄDZANIE I MARKETING**

**SIEDZIBA**

**Instytut Badań Systemowych**  
**Polskiej Akademii Nauk**

**ISBN 83-85847-24-3**