STUDIA INFORMATICA 2013
Volume 34 Number 2B (112)

Stanisław PŁACZEK
Akademia Finansów i Biznesu Vistula, Wydział Informatyki

ZASTOSOWANIE SZTUCZNYCH SIECI NEURONOWYCH W ANALIZIE DANYCH TOWARZYSTWA UBEZPIECZEŃ

Streszczenie. Na wstępie artykułu krótko scharakteryzowano specyfikę procesów biznesowych w towarzystwie ubezpieczeń. Duże ilości danych zawarte w bazach danych wykorzystano do analizy dwóch podstawowych procesów – przypisu składki i realizacji świadczeń. Do zadań prognozy krótkoterminowej zastosowano sztuczną sieć neuronową. Scharakteryzowano etapy wstępnego przetwarzania danych, wyniki uczenia sieci oraz omówiono końcowe wnioski.

Słowa kluczowe: Sztuczne sieci neuronowe, analiza danych, ubezpieczenia

APPLICATION OF ARTIFICIAL NEURAL NETWORKS IN DATA ANALYSIS OF INSURANCE COMPANY

Summary. In the beginning, business processes in Insurance Company were briefly characterized. Huge amounts of data included in data bases were used to analyze two business processes – sales income and insurance benefit. For short forecasting tasks an artificial neural network was used. Introductory stages of data processing and network learning results were characterized. At the end, final conclusions were discussed.

Keywords: Artificial neural network, data analyze, insurance

1. Charakterystyka Towarzystwa Ubezpieczeń na Życie (TUnŻ)

Systemy informatyczne TUnŻ średniej wielkości charakteryzują się swoistą modułowością, podyktowaną przede wszystkim historycznymi decyzjami oraz kosztami. Branżowa baza danych, często nazywana Bazą Ubezpieczeniową, jest sercem każdego TUnŻ, z niej korzystają wszystkie procesy zarządcze towarzystwa. Baza ta zawiera takie dane, jak:

- dane osobowe ubezpieczonych i ubezpieczających,

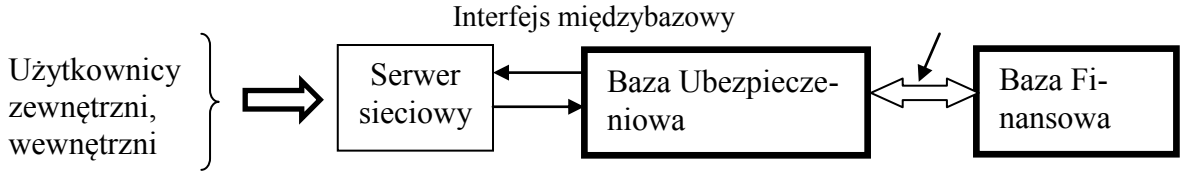
- dane osobowe beneficjentów,
- dane osobowe agentów i innych pośredników ubezpieczeniowych,
- dane świadczeniobiorców, czyli beneficjentów ubezpieczenia,
- parametry sprzedażowe i aktuarialne produktów ubezpieczeniowych,
- struktura ryzyka ubezpieczeniowego i jego składka netto/brutto,
- wnioski i polisy ubezpieczeniowe.

Baza Ubezpieczeniowa przechowuje wszelkie dane historyczne, związane z procesami tworzenia produktów ubezpieczeniowych, sprzedażą tych produktów, a także bardzo ważnym procesem – obsługi świadczeń. Najczęściej Baza ta nie zawiera danych finansowych lub zawiera tylko ich część, np. zebrana składka ubezpieczeniowa. Chociaż na rynku istnieje wielu dostawców ubezpieczeniowych baz danych, wiele towarzystw decyduje się na projektowanie i realizację baz na podstawie własnych grup programistów.

Drugą niezbędną bazą danych jest Baza Finansowa, najczęściej kupowana u wyspecjalizowanego dostawcy. W Bazie tej, opartej na standardach rachunkowości przyjętych w TUnŻ, poprzez rozbudowany układ kont realizowane są rozliczenia oraz wymiana informacji z:

- bankami,
- instytucjami zarządzającymi funduszami,
- ubezpieczonymi,
- agentami,
- innymi podmiotami, jak np. reasekuratorami.

Oczywiście bazy te muszą wymieniać dane pomiędzy sobą poprzez odpowiedni zaprojektowany interfejs (rys. 1).



Rys. 1. Struktura baz danych
Fig. 1. Structure of data bases

2. Struktura danych produktu ubezpieczeniowego

W procesie projektowania oraz konstruowania nowego produktu ubezpieczeniowego, kluczową rolę odgrywa aktuariusz, czyli statystyk ubezpieczeniowy. Możemy wyróżnić dwa warianty:

1. Towarzystwo ma dostatecznie długą historię działania oraz dużą ilość danych w bazie ubezpieczeniowej. W tym przypadku aktuariusz realizuje proces ekstrakcji potrzebnych danych i wykorzystując standardowe metody statystyczne, oblicza korelacje, rozkłady

oraz ich parametry dla poszczególnego ryzyka zawartego w polisach. Jakość uzyskanych rezultatów jest wprost proporcjonalna do ilości przetwarzanych danych. Jest to proces bardzo pracochłonny i trudny. Zależy od złożoności produktu, a w polisach grupowych liczba ubezpieczanych ryzyk może być bardzo duża.

2. W przypadku braku danych historycznych, aktuariusz musi przyjąć a priori, na podstawie własnej wiedzy eksperckiej lub też pewnych danych zewnętrznych, rozkłady ryzyka i jego parametry.

Dalsza część procesu jest stosunkowo prosta i polega na obliczeniu optymalnej składki netto dla każdego ryzyka w funkcji przyjętej sumy ubezpieczenia, a także na podstawie decyzje zarządu kalkuluje się prowizję agencyjną oraz składkę brutto, zawierającą dodatkowo koszty administracyjne oraz prowizję.

Podstawową jednostką biznesową, podlegającą analizie, jest polisa ubezpieczeniowa. Pomijając aspekt prawny polisy, to z punktu widzenia analizy, jej parametry można podzielić na dwie grupy:

1. Parametry wbudowane w strukturę polisy. Do nich zaliczamy:
   a) Su – suma ubezpieczenia ryzyka,
   b) $R1 \ldots Rn$ – specyfikacja ryzyka istotnego z punktu widzenia wartości składki, ilości i wartości wypłaconych świadczeń.
   c) $S1 \ldots Sn$ – wartość brutto lub netto składek aktuarialnych dla poszczególnego ryzyka,
2. Parametry rynkowe, zrealizowane w wyniku sprzedaży konkretnej polisy przez agenta. Do nich zaliczamy:
   d) Tp – czas życia polisy, czas od daty sprzedaży do dnia analizy w miesiącach lub latach,
   e) Lu – liczba ubezpieczonych w polisach grupowych,
   f) Pk – procent kobiet w grupie ubezpieczonych,
   g) Wu – średni wiek ubezpieczonych,
   h) Kr – kod regionu kraju, w którym sprzedano polisę,
   i) Ru – średnia rotacja ubezpieczonych, tj. osób rezygnujących z ubezpieczenia i nowych dołączonych do grupy,
   j) Pf – grupa parametrów finansowych, jak: stopa zwrotu z obligacji, lokat długo- i średnioterminowych i inne,
   k) Zz – zgłoszone zdarzenia do wypłaty świadczeń.

Wyżej wymienione. parametry wbudowane i rynkowe mają istotny wpływ na wartość przypisu (opłaconych składek), wypłaconych świadczeń, a tym samym na przyjęte przez ubezpieczyciela kryterium dochodowości polisy. Możemy napisać, że dochodowość polisy jest nieliniową funkcją parametrów:

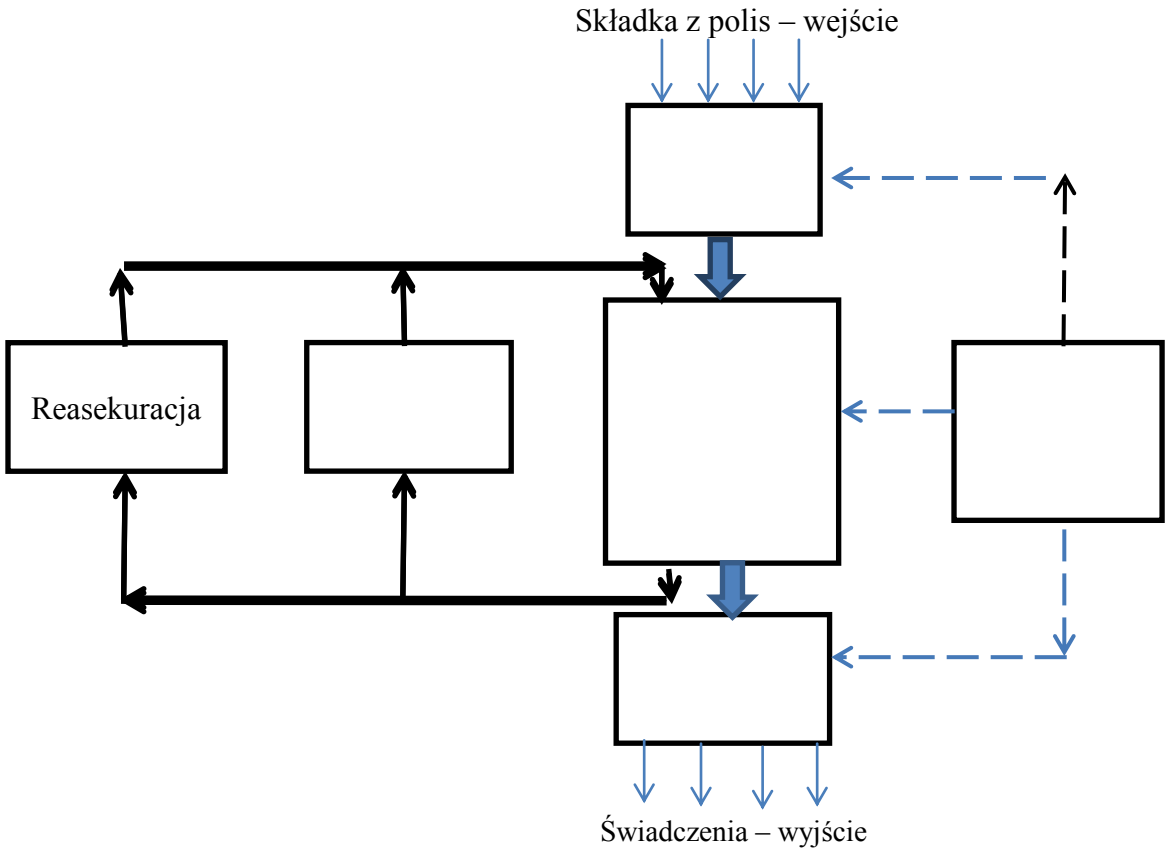
$$W_f = f\left(Su, R1 \ldots Rn, S1 \ldots Sn, Tp, Lu, Pk, Wu, Kr, Ru, Pf, Zz, t\right) \tag{1}$$

gdzie W_f to wynik finansowy polisy (pozostałe oznaczenia powyżej).

3. Model matematyczny ubezpieczeń życiowych

W praktyce biznesowej oraz sprawozdawczości finansowej i zarządczej, proces analizy dochodowości polis oraz całego TUnŻ rozpatruje się w dyskretnych momentach czasu, najczęściej na koniec miesiąca kalendarzowego. Jak już podkreślano, podstawową jednostką ubezpieczeniową jest polisa. Suma zebranej składki zarówno z polis historycznych, jak i sprzedanych w danym miesiącu tworzy przypis składki. Przypis składki można traktować jako strumień wejściowy modelu matematycznego ubezpieczeń. Podobnie w danym miesiącu do TUnŻ wpływają zgłoszenia zdarzeń z ryzyka na danej polisie. Zgłoszone i wypłacone świadczenia tworzą koszt, który jest traktowany jako strumień wyjściowy procesu.

Przyjmując pewne uproszczenia, niewpływające na ogólność i dokładność rozważań, model matematyczny ubezpieczeń życiowych można przedstawić jak na rys. 2.



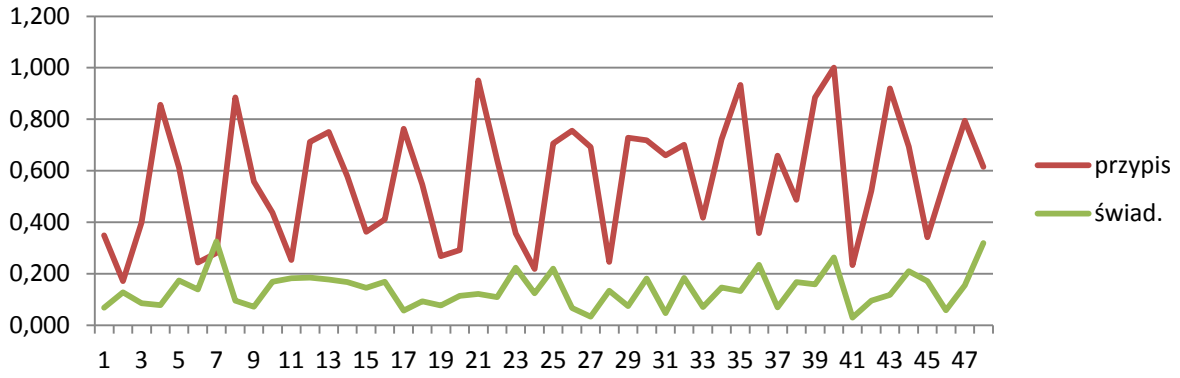
Rys. 2. Model zarządzania procesami i danymi w TUnŻ
Fig. 2. Processes and data management model for Life Insurance Company

Przypis i świadczenia w dyskretnych momentach czasu możemy traktować jako szeregi czasowe. W praktyce, wykorzystując Bazę Ubezpieczeń oraz Bazę Finansową, dokonuje się ekstrakcji danych dla każdej polisy w rozpatrywanym przedziale czasu. Przyjmując jako kwant czasu tydzień lub miesiąc kalendarzowy, grupujemy dane przypisu i świadczeń.

Na rysunku 3 przedstawiono szereg czasowy przypisu i świadczeń w okresie czterech lat. Pozioma oś czasu wyskalowana jest w miesiącach, natomiast znormalizowana oś pionowa reprezentuje przypis i świadczenia.

Dane przekazywane do aktuariatu oraz działu finansowego wykorzystywane są w procesach analizy i sprawozdawczości. Interesującym zadaniem jest zagadnienie prognozy i budowy biznesplanów. Powstaje pytanie, jak bez wielu wstępnych założeń dla parametrów przyszłościowych zrealizować skuteczną prognozę przypisu, świadczeń i innych wielkości biznesowych, na podstawie których będzie można podejmować wiarygodne decyzje biznesowe?

Do rozwiązania zadania zaproponowano użycie sztucznej sieci neuronowej (SSN).



Rys. 3. Szereg czasowy przypisu i świadczeń
Fig. 3. Time series of premium income and insurance benefits

4. Sztuczna sieć neuronowa w zadaniach analizy i prognozy

W zastosowaniach analizy sieć neuronowa najczęściej pełni rolę uniwersalnego aproksymatora funkcji wielu zmiennych:

$$\mathbf{y} = f(\mathbf{x}) \tag{2}$$

gdzie **y**, **x** są odpowiednio wektorami wyjściowymi i wejściowymi [1].

W zagadnieniach identyfikacji i zarządzania procesami dynamicznymi SSN, pełni zwykle kilka funkcji:

1. Stanowi model nieliniowy tego procesu.
2. Wypracowuje odpowiednią informację, niezbędną do optymalnego zarządzania.

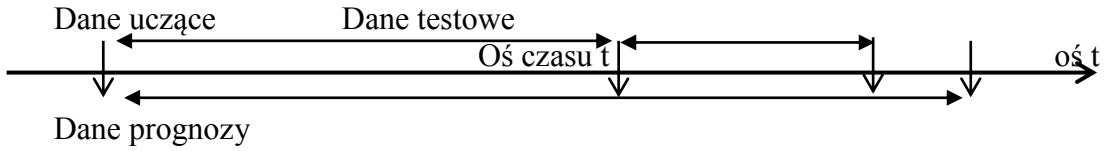
Powyższe podejście zastosujemy w procesie analizy i predykcji szeregu sprzedaży, w którym to procesie możemy wydzielić ograniczony podzbiór danych wejściowych.

Do analizy szeregu świadczeń, SSN realizować będzie zadanie predykcji. Mamy informację o wartościach świadczeń w momentach czasowych poprzedzających predykcję. Sieć podejmuje decyzję odnośnie przyszłej (estymowanej) wartości.

Dane wykorzystywane w procesie analizy i prognozy dzielimy na:

1. Dane uczące. Jak już podkreślaliśmy, analizę danych ubezpieczeniowych realizujemy najczęściej, biorąc pod uwagę określony horyzont czasowy. Długość horyzontu związana jest z czasem życia wybranego zbioru polis oraz niezbędną ilością danych do uczenia i testowania sieci. Minimalny horyzont czasu to 60 miesięcy. Na etapie uczenia sieci skracamy horyzont i wykorzystujemy dane z pierwszych 48 miesięcy (rys. 4).
2. Dane testujące. Do testowania sieci wykorzystujemy dane z ostatnich 12 miesięcy. Pozwala nam to stwierdzić, czy sieć nabyła umiejętności generalizacji i czy błąd jest dostatecznie mały. Manipulowanie horyzontem czasowym jest często stosowane w analizie danych nie tylko w procesach z wykorzystaniem SSN. Powyższe spowodowane jest tym, że w finansach stosuje się kumulujący system rejestracji danych, czyli dane z okresów późniejszych sumują dane z okresów poprzednich oraz dane z ostatniego horyzontu czasowego.

Do zrealizowania zadania analizy i prognozy wykorzystano własne programy symulujące pracę SSN z jedną warstwą ukrytą. W celu zdobycia doświadczenia, w pierwszych okresie używano arkusza Excel i programów napisanych w Visual Basic. Osiągana szybkość obliczeń nie była duża, lecz na etapie pierwszych analiz ten problem nie był najważniejszy. Następnie napisano program w języku C++, symulujący prace SSN z jedną warstwą ukrytą, lecz o dowolnej liczbie neuronów wejściowych, ukrytych i wyjściowych.



Rys. 4. Struktura horyzontów czasowych dla danych
Fig. 4. Time horizon structure for a given data

Do uczenia sieci używano algorytmu wstecznej propagacji błędów oraz podjęto próby zastosowania algorytmu zmiennej metryki na podstawie formuły BFGS (Broydena-Fletchera-Goldfarba-Shanno). Podstawą algorytmów jest funkcja celu, zdefiniowana jako suma kwadratów różnic między aktualnymi wartościami sygnałów wyjściowych sieci a wartościami zadanymi [1]. Dla pojedynczej próbki uczącej oraz sieci o jednym sygnale wyjściowym, funkcję celu definiuje się w postaci:

$$e\left(w^1, w^2\right) = \frac{1}{2} * \left(y - d\right)^2 \tag{3}$$

gdzie: $e(w^1,w^2)$ – błąd średniokwadratowy, y – wartość aktualna wyjścia sieci, d – wartość zadana wyjścia sieci.

Pełny schemat architektury sieci oraz idei programu przedstawiono na rys. 5.

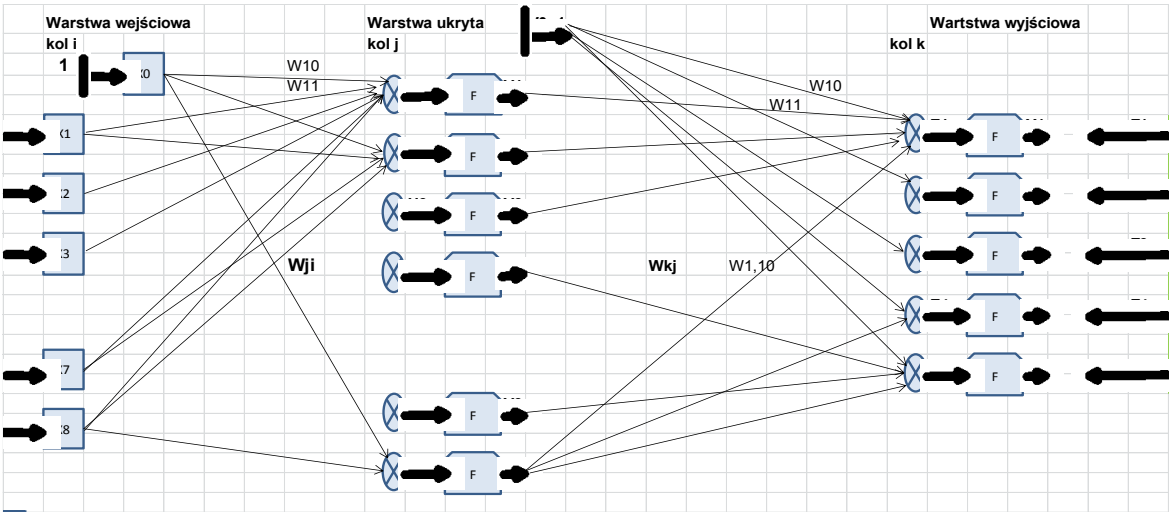
Schemat ten przedstawia standardową SSN. Wagi sieci reprezentowane są przez odpowiednie macierze W1 oraz W2. Inicjacja początkowa tych wag realizowana jest losowo

z przedziału (0-1). Ponieważ w algorytmie do korekty wag wykorzystywana jest technika momentum, do schematu dołożono dodatkowe dwie macierze, przechowujące dane z iteracji (kroków) przednich.

Na schemacie w sposób poglądowy pokazano realizację obliczeń w przód i w tył, poprzez mnożenie stosownych macierzy. W kroku pierwszym oblicza się wszystkie zmienne pomocnicze na wejściach i wyjściach funkcji aktywacji. Na wyjściu sieci porównuje się wartość obliczoną z zadaną. Błąd średniokwadratowy (3) wykorzystywany jest do obliczenia gradientów w warstwie wyjściowej i ukrytej. W następnym kroku, wykorzystując algorytmy do obliczenia poprawki, modyfikuje się wartości wag w macierzach W1 oraz W2.

5. Wstępne przetwarzanie danych i wyniki uczenia SSN dla przypisu

Celem wstępnego przetwarzania danych są grupowanie oraz normalizacja danych. Szczegółowe dane z bazy finansowej muszą zostać zagregowane do postaci wymaganej przez konkretną sieć. Statystyk ubezpieczeniowy (aktuariusz) dostarcza dane pierwotne w arkuszu kalkulacyjnym Excel, w którym kolumny zawierają wszystkie parametry wejściowe i wyjściowe. W danych wejściowych bardzo ważna jest data (miesiąc) opłacenia składki. W dalszej kolejności następuje scalanie polis wg daty, w wyniku czego zostaje utworzony przypis w danym miesiącu. Liczba wierszy w tablicy Excel nie przekraczała 5000, co biorą pod uwagę długość analizowanego horyzontu czasowego, średnia liczba analizowanych polis nie przekraczała 100. Dla standardowej polisy grupowej liczba ryzyk ubezpieczeniowych przekracza 20, a dodając do tego parametry rynkowe otrzymujemy, ok 30 – 40 danych wejściowych.



Rys. 5. Idea algorytmu symulującego SSN
Fig. 5. The idea of simulation algorithm of Artificial Neural Network

Taka liczba parametrów wejściowych może nie jest oszałamiająco duża, lecz doświadczenie aktuarialne oraz analiza korelacyjna pokazują, że nie wszystkie parametry mają identycznie duży wpływ na parametr wyjściowy, czyli w tym przypadku przypis miesięczny.

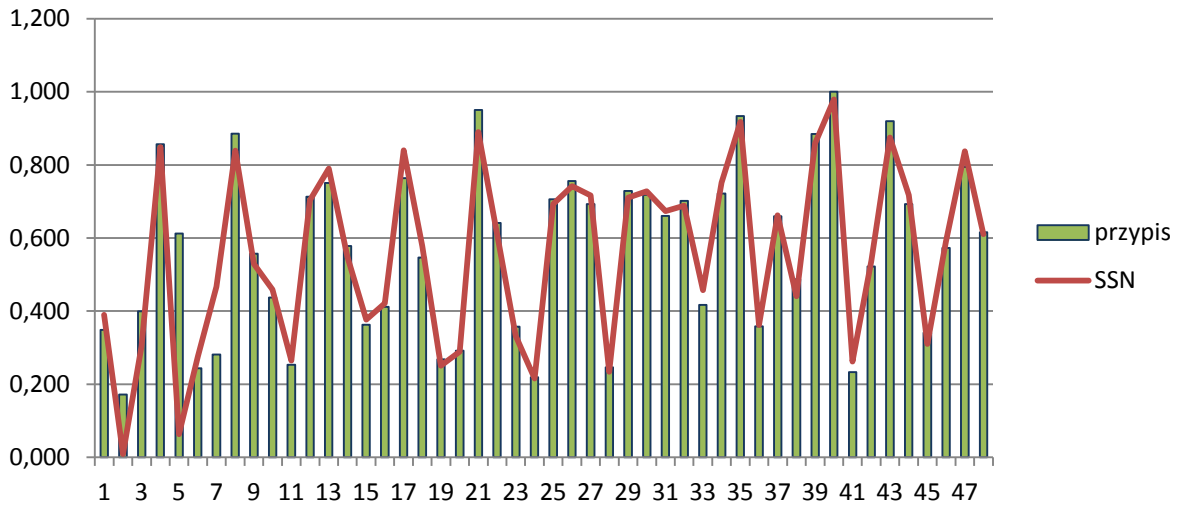
Dlatego też, jako parametry wejścia sieci, przyjęto ryzyka ubezpieczeniowe w liczbie 4-5 oraz taką liczbę parametrów rynkowych, żeby długość wektora wejściowego nie przekraczała 10.

Z ryzyka ubezpieczeniowego wybrano dane do wektora wejściowego X:

- ND – narodziny dziecka,
- PS – pobyt w szpitalu,
- PZ – poważne zachorowanie,
- ZU – zgon ubezpieczonego,
- ZR – zgon rodzica.

Następnym krokiem jest grupowanie polis w danym miesiącu oraz sumowanie wybranych składników cząstkowych. Powyższy krok realizuje się dla różnych, potrzebnych w przyszłości, horyzontów czasowych. Sumuje się polisy w miesiącach od 1-48 oraz od 49-60. Często sumuje się dane w całym horyzoncie zdarzeń od 1-60, w celu weryfikacji danych, np. sumuje się składki z poszczególnego ryzyka, a nie sumuje się sum ubezpieczenia.

Ważnym krokiem jest obliczenie wartości wyjściowej Y, czyli przypisu miesięcznego. Po uporządkowaniu zbiorów X i Y, możemy przystąpić do normalizacji danych. Dąży się do tego, żeby zmienność każdego parametru wejściowego oraz parametru wyjściowego zawierała się w granicach 0-1. Powyższy proces ma duży wpływ na szybkość zbieżności oraz dobór wartości wag. Tak przygotowane dane zapisuje się w arkuszu kalkulacyjnym, oddzielnie dane uczące oraz dane testujące. Po zdefiniowaniu pozostałych parametrów niezbędnych do uruchomienia programu, przystępujemy do analizy danych z wykorzystaniem SSN.



Rys. 6. Rezultaty uczenia SSN dla przypisu miesięcznego
Fig. 6. Artificial Neural Network's learning results for monthly premium income

Do analizy przepisu zastosowano sieć o konfiguracji 10-12-1, czyli:

- 10 neuronów wejściowych,
- 12 neuronów ukrytych,
- 1 neuron wyjściowy.

Do uczenia sieci wykorzystuje się zbiory uczący i testujący naprzemiennie. W kroku pierwszym podaje się na wejście sieci dane ze zbioru uczącego. Powyższe powtarza się kilkadziesiąt razy. W kroku drugim na wejście podaje się dane zbioru testującego. Jeżeli błąd określony wzorem (3) maleje, powracamy do kroku 1 tak długo, dopóki błąd testowania maleje. W momencie kiedy błąd testowania rośnie lub pozostaje stały, proces uczenia uważa się za zakończony. Współczynniki wagowe w macierzach W1 oraz W2 są zamrażane i sieć jest gotowa do realizacji zadania prognozowania. Powyższy proces może być powtórzony wielokrotnie dla różnych wartości początkowych współczynników wagowych.

Sieć o konfiguracji 10-12-1 realizowała najmniejszy błędy uczenia i testowania.

Sprawdzono również inne konfiguracje np. 10-8-1. Otrzymane wyniki były gorsze. Na rys. 6 pokazano na jednym wykresie dane wyjściowe zadane (uczące, nazwane przypisem) oraz dane wyjściowe, wygenerowane przez sieć. Wynik jest zadawalający.

6. Analiza szeregu czasowego dla świadczeń

Świadczenia to bardzo niewdzięczny szereg czasowy do analizy danych, co wynika przede wszystkim z ich statystycznych właściwości. Zdarzenia losowe zawarte w ryzykach ubezpieczeniowych są :

1. Nieprzewidywalne całkowicie w czasie.
2. Najczęściej statystycznie niezależne. Trudno uchwycić korelacje pomiędzy ryzykami mającymi istotny wpływ na wartość świadczenia i czasem wystąpienia zdarzenia.

Do analizy danych świadczeniowych z wykorzystaniem SSN przyjęto, że mamy informację o wartościach świadczeń w chwilach poprzedzających predykcję. Oznaczając przez $y(t)$ wartość świadczenia w chwili t, możemy napisać, że wartość świadczenia jest funkcją:

$$y(t) = a_0 + a_1 * y(t-1) + a_2 * y(t-2) + \cdots + a_n * y(t-n) \tag{4}$$

Dla tak zdefiniowanego zadania przygotowujemy tablicę z danymi, w której dane w poszczególnych kolumnach przesunięte są o jeden takt. Stosowano różne wartości n – liczba taktów opóźniających. Minimalna liczba to n=5, a maksymalna 10. Istnieje uzasadnione podejrzenie, że dla n >12, czyli dla pełnego cyklu rocznego, otrzymane wyniki mogą być lepsze, chociaż trudno udowodnić występowanie pewnej cykliczności zdarzeń ubezpieczeniowych w ujęciu rocznym. Te badania należy kontynuować pod warunkiem posiadania większej ilość danych (z min. 7-8 lat).

Wektorem wejściowym do sieci jest wyżej zdefiniowana tablica X. Wektorem wyjściowym jest $y(t)$.

Stosowano SSN o różnych liczbach neuronów wejściowych

- od 5 do 10 neuronów wejściowych,
- od 6 do 12 neuronów ukrytych,
- 1 neuron wyjściowy.

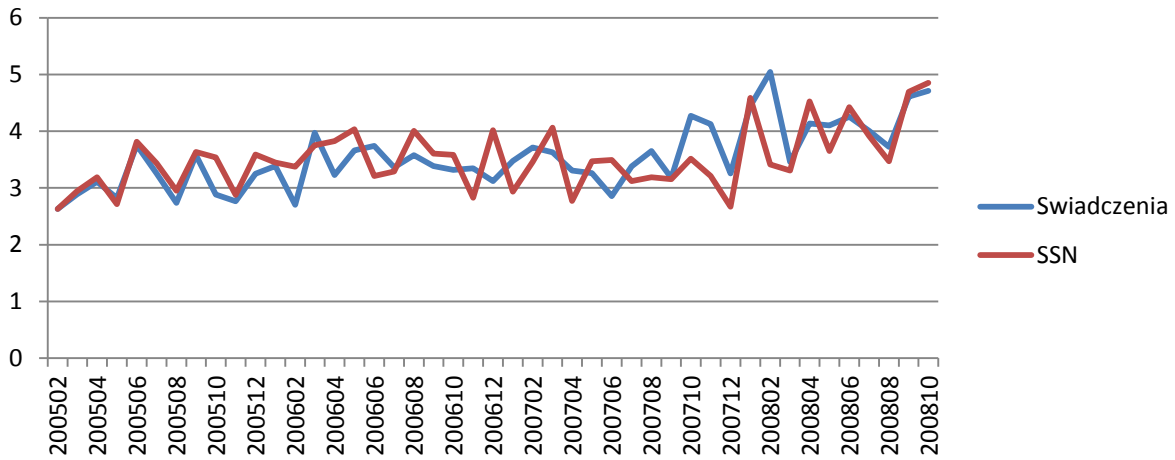
Na rysunku 7 pokazano szeregi czasowe świadczeń, jako pierwotne dane wejściowe, oraz wyniki otrzymane przez SSN.

Można stwierdzić, że dla wszystkich wariantów sieci otrzymane rezultaty nie były satysfakcjonujące.

1. Błędy uczenia były znacznie większe niż dla szeregów czasowych przypisu.
2. Czas uczenia był znacznie dłuższy.
3. Nie można było uchwycić zależności pomiędzy strukturą SSN a błędem uczenia.

Po wnikliwej analizie danych wejściowych stwierdzono, że przyczyn niesatysfakcjonujących rezultatów należy poszukiwać w zbyt dużych wahaniach w wypłacie świadczeń oraz obecności polis o wysokich, wręcz niestandardowych, wartościach ubezpieczonego ryzyka. Ilość tego typu polis nie była duża, polisy te można traktować jako pewnego rodzaju zakłócenia. W tym upatruje się otrzymanie stosunkowo słabych wyników.

Należy więc zwrócić większą uwagę na jakość danych na wstępnym etapie analizy i większą jednorodność wejściowego, rozpatrywanego zbioru polis.



Rys. 7. Rezultaty uczenia SSN dla szeregu czasowego świadczeń
Fig. 7. Artificial Neural Network's learning results for monthly insurance benefits

7. Podsumowanie i wnioski

W artykule podjęto próbę przedstawienia otrzymanych rezultatów analizy danych w TUnŻ średniej wielkości, z wykorzystaniem SSN. Głównym celem było stwierdzenie, czy bez dodatkowych założeń co do charakterystyk statystycznych danych, SSN uchwyci nowe, ukryte i nieznane zależności oraz korelacje pomiędzy danymi. Okazało się, że szeregi czasowe przypisu dla wybranych parametrów wejściowych umożliwiają takie wytrenowanie sieci, że jej dalsze wykorzystanie do zadań prognozy i tworzenia biznesplanów jest możliwe.

Niestety, w stosunku do świadczeń nie można tego powiedzieć na obecnym poziomie badań świadczeniowych szeregów czasowych. Przyczyny omówiono powyżej.

BIBLIOGRAFIA

1. Osowski S.: Sieci neuronowe do przetwarzania informacji. Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2006.
2. Rutkowski L.: Metody i techniki sztucznej inteligencji. Wydawnictwo Naukowe PWN, Warszawa 2006.
3. Korbicz J., Obuchowicz A., Uciński D.: Sztuczne sieci neuronowe, podstawy i zastosowania. Akademicka Oficyna Wydawnicza PLJ, Warszawa 1994.
4. Dittmann P., Szabela-Pasierbińska E., Dittmann I., Szpulak A.: Prognozowanie w zarządzaniu przedsiębiorstwem. Oficyna Wolters Kluwer, Kraków 2009.
5. Gately E.: Sieci Neuronowe. Prognozowanie finansowe i projektowanie systemów transakcyjnych. Biblioteka Inwestora, Warszawa 1999.
6. Rubunal J. R., Dorado J.: Artificial Neural Networks in Real – Life Applications. Idea Group Publishing, Hershey 2006.
7. Kamruzzaman J., Begg R. K., Sarker R. A.: Artificial Neural Networks in Finance and Manufacturing. Idea Group Publishing, Hershey.
8. Galushkin A. I.: Neural Networks Theory. Springer, Berlin 2007.

Wpłynęło do Redakcji 15 stycznia 2013 r.

Abstract

Information systems of Life Insurance Company include in their data bases a large amount of data on policies, finance and insurance benefits. Through application of elaborate

methods of statistical analysis, actuarial processes analyze the abovementioned data, monitor profitability of policies in the short and long term. From a managerial perspective, financial result of the policy (profit or loss) is a non-linear function of many parameters embedded in the policy and market parameters. For two most important processes from a business point of view, an attempt has been made to apply ANN in the analysis of time series of sales income and insurance benefits. These series differ considerably:

1. Sales income may be described with a formula (1) as a non-linear function of input parameters. Therefore, at the entry of a network we provide input parameters of particular policies, whereas at the exit of a network in the teaching process – its previous profitability.
2. Insurance benefit as a stochastic process, do not have clearly specified input parameters. Owing to this, with regard to times series of insurance benefits we use the equation (3).

The process of preparation of data for ANN is rather complicated. Time of horizon is manipulated so that the data used in network learning encompass only the subset of all the elements of sales income in the function of time (months). Test data are the most up-to-date and include the history of sales income of a given policy.

In data analysis, a standard ANN with one hidden layer has been used. The number of neurons in the entry layer depended on the length of the input data vector and did not exceed 10.

In the hidden layer the number of neurons changed from 8 to 15. The best learning errors have been achieved with n=12. In the exit layer there is only one neuron with a linear activation function.

ANN learning outcomes for given sales income is presented on picture 6. Network 10-12-1 appeared to be the most optimal.

With regard to insurance benefits, the outcomes did not meet the expectations. This may be caused by a considerable changeability of the process and a short horizon of the given time series. The obtained learning outcomes are presented on picture 7.

Adres

Stanisław PŁACZEK: Akademia Finansów i Biznesu Vistula w Warszawie, Wydział Informatyki, ul. Stokłosy 3, 02-787 Warszawa, Polska, stanislaw.placzek@wp.pl.