

Prognozowanie kondycji ekonomiczno-finansowej przedsiębiorstw z wykorzystaniem sztucznych sieci neuronowych

Andrzej Burda¹

1. Wstęp

Prognozowanie kondycji ekonomiczno-finansowej podmiotów gospodarczych jest od lat szeroko studiowanym zagadnieniem teoretycznym, głównie za przyczyną banków, które przy podejmowaniu decyzji kredytowych są szczególnie zainteresowane oceną klienta w tym kontekście. Większość badaczy dokonuje próby klasyfikacji przedsiębiorstw wg schematu bankructwo – przetrwanie. W przypadku tworzenia modeli metodami indukcyjnymi², co będzie treścią tego artykułu, tak prowadzone badania wymagają zgromadzenia istotnych informacji o przedsiębiorstwach, które trwają na rynku i podobny co do liczebności zbiór informacji o firmach, które zgodnie z lokalnym prawodawstwem można uznać za bankrutów³.

Jedną z pierwszych prac poświęconych temu zagadnieniu jest analiza przeprowadzona przez M. Odom i R. Sharda [8], w której opisano model neuronowy NN (Neural Networks) opracowany na podstawie danych 128 amerykańskich firm i zestawiono jego wyniki z metodą MDA (Multiple Discriminant Analysis), do której zaliczamy model Z-score Altmana. Model neuronowy oscylował pomiędzy 77,8% a 81,5% poprawnych klasyfikacji, a MDA tylko pomiędzy 59,3% a 70,4%, przy zastosowaniu tego samego zestawu zmiennych objaśniających.

Analizie porównawczej modeli przewidywania bankructwa poświęcone są również prace: K. Tam [14] oraz K. Tam i M. Kiang [13]. Dokonano w nich porównania pomiędzy kilkoma metodami: MDA, LR (Logistic Regression), KNN (K-Nearest Neighbor), ID3 (Interactive dichotomizer version 3) oraz sztucznymi sieciami neuronowymi jedno i wielowarstwowymi. Dla horyzontu czasowego prognozy wynoszącego 1 rok, sieci wielowarstwowe były najlepsze, w przypadku 2-letniego horyzontu ustępowały nieznacznie tylko modelowi LR.

L. Salchenberger, E. Cinar i N. Lash [12] porównali wyniki prognozy dla modeli NN i LR. NN znacząco przewyższyły osiągnięcia LR. Na przykład dla horyzontu prognozy 18 miesięcy model LR osiągał od 83,3 % do 85,4% dokładności, podczas gdy NN aż 91,7 %.

Z kolei P. Coats i L. Fant [4] zestawili modele NN i MDA. Dokładność klasyfikacji wynosiła 81,9 % - 95 % dla NN a 83,7% - 87,9 % dla MDA, przy horyzontach prognozy od 3 lat do 1 roku.

Opis szczególnie interesujących badań, przeprowadzonych na bardzo dużej próbie, bo liczącej 1000 włoskich firm, przy zastosowaniu 1-rocznego horyzontu prognozy, znajdujemy u E. Altmana, G. Marco i F. Karetto [3]. W artykule zestawiono dwa modele: NN i MDA. Porównanie wykazało minimalną przewagę modelu MDA.

Wyniki wskazujące na przewagę modeli neuronowych nad modelami konkurencyjnymi wykazywało jeszcze wielu badaczy, m.in.:

- E. Fernandez i I. Olmeda [6],
- Y. Alici [2],
- G. Zhang, M. Hu i B. Patuwon [17].

Badania te prowadzono m. in. na firmach amerykańskich, brytyjskich, australijskich, belgijskich i koreańskich przy użyciu wielu rodzajów sztucznych sieci neuronowych, takich jak: MLP (Multilayer Perceptron), RBF (Radial Basis Function- networks), sieci Kohonena i PNN (Probabilistic Neural Networks) oraz przy zastosowaniu różnorodnych metod ich uczenia.

Zdecydowana większość badań została przeprowadzona na firmach dużych, których akcje notowane są na giełdach papierów wartościowych. Wszystkie one w pełni uzasadniają przydatność sztucznych sieci neuronowych (SSN) do prognozowania kondycji ekonomiczno-finansowej przedsiębiorstwa.

2. Budowa i działanie sztucznych sieci neuronowych

Modelowanie neuronowe korzeniami swoimi sięga do tęsknoty człowieka, aby odwzorować otaczający nas świat przyrody. W tym świecie, szczególnie intrygującym zjawiskiem jest człowiek i jego mózg. Stworzenie sztucznego mózgu, który w zadawalający sposób spełni nasze oczekiwania, to ciągle wyzwanie dalekiej przyszłości. Nieustannie prowadzone na całym świecie badania³ dowodzą jak daleka jeszcze jest ta droga. Jednocześnie ukazują jakimi niespotykanymi możliwościami dysponują uproszczone jeszcze na tym etapie modele.

Ludzki mózg składa się z ok. 10 miliardów komórek nerwowych zwanych neuronami. Funkcje realizowane przez pojedynczy neuron (Rys.1.) w uproszczeniu można opisać następująco. Poprzez dendryty, których ma wiele, docierają do niego sygnały (przyjmijmy, że elektryczne, choć ich prawdziwa natura jest znacznie bardziej złożona) od sąsiadujących i połączonych z nim innych neuronów „sieci”. Wartość tego sygnału może być

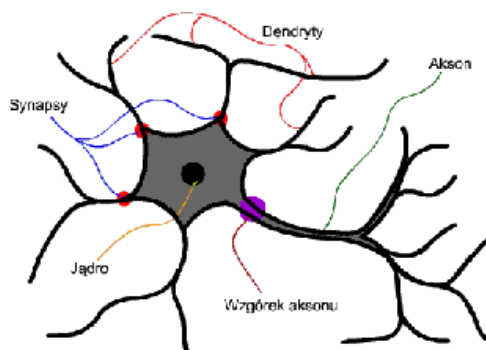
¹ Wyższa Szkoła Zarządzania i Administracji, Katedra Informatyki i Inżynierii Wiedzy.

² Wnioskowanie indukcyjne polega na tworzeniu wiedzy, która jest uogólnieniem informacji zawartych w prezentowanych przykładach.

³ Prognozowanie kondycji przedsiębiorstwa, czyli próba odpowiedzi na pytanie, czy na podstawie dostępnych danych można przewidzieć jego stan (dobry – zły) w perspektywie jednego roku, jest zagadnieniem bardzo zbliżonym, ale wymaga nieco innych wzorców uczących. Szczegółowy opis ich przygotowania opisany jest w dalszej części tego opracowania.

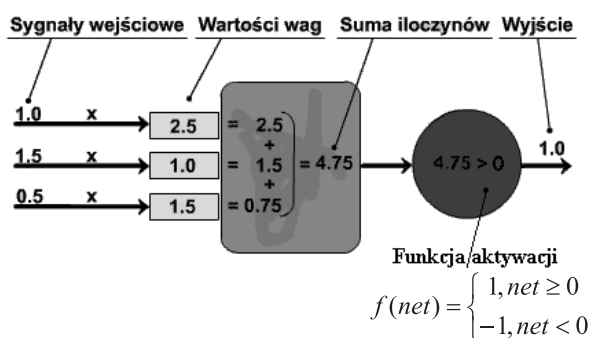
tłumiona poprzez synapsę, która stanowi o jakości połączenia z ciałem komórki i decyduje o wpływie docierającego sygnału na wielkość pobudzenia komórki. Suma sygnałów docierających do komórki poprzez wszystkie jej synapsy decyduje o stopniu jej pobudzenia. Jeśli pobudzenie neuronu przekroczy określoną dla niego wartość graniczną, zwaną progiem pobudzenia, to poprzez akson (tylko jeden) wysyła on sygnał do innych, połączonych z nim neuronów.

W ludzkim mózgu dzieje się tak, że jakość często wykorzystywanego połączenia poprawia się. Gdy przestaje być wykorzystywane (czyt. potrzebne), to maleje aż do możliwej, całkowitej jego blokady. Proces ten decyduje o zdolności uczenia i zapominania.



Rys. 1. Budowa neuronu
Źródło: Klaus R., Budowa neuronu,
www.cs.put.poznan.pl/rklaus/assn/neuron.htm,
[15.07.2006 r.]

Tak uproszczony opis funkcjonowania ludzkiego mózgu stał się kanwą do stworzenia pierwszego modelu sztucznego neuronu⁴ (Rys.2.) i w konsekwencji wpłynął na szybki rozwój tej dziedziny wiedzy.



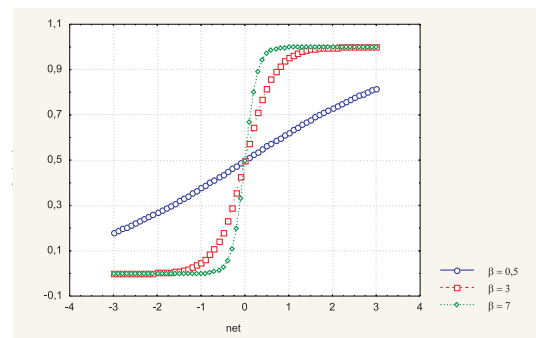
Rys. 2. Model sztucznego neuronu
Źródło: Klaus R., Budowa neuronu,
www.cs.put.poznan.pl/rklaus/assn/neuron.htm,
[15.07.2006 r.]

W sztucznym neuronie rolę synapsy pełnią tzw. wagi (multiplikatory) sygnałów wejściowych, a za nieliniowe przetwarzanie sygnału wynikające m.in. z progu aktywacji rzeczywistego neuronu- nieliniowa funkcja aktywacji.

wacji. Zgodnie z Rys.2., na wyjściu neuronu pojawi się sygnał o wartości 1 tylko wtedy, gdy suma pobudzenia neuronu net (iloczyn skalarny sygnału wejściowego i wartości wagowych) będzie nie mniejsza od 0, w przeciwnym razie wartość na wyjściu wyniesie -1. Tak określona bipolarna $\{-1;1\}$ funkcja progowa może podlegać modyfikacjom. Najczęściej jednak, szczególnie w odniesieniu do perceptronów wielowarstwowych MLP (Multilayer Perceptron), wykorzystuje się tzw. funkcję logistyczną w postaci:

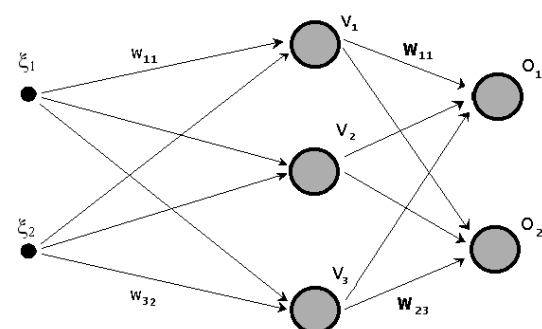
$$(1) \quad f_i(net) = \frac{1}{1 + \exp^{-\beta net}}$$

Parametr β tej funkcji decyduje o jej kształcie (Rys.3.). Przy małych wartościach tego parametru ma ona przebieg prawie liniowy, przy dużych- upodabnia się do funkcji progowej unipolarnej $\{0;1\}$.



Rys. 3. Wykres funkcji logistycznej
Źródło: Opracowanie własne

Połączenie sztucznych neuronów w sieć może być realizowane na wiele sposobów. Jedną z pierwszych, do dziś (ze względu na swoje liczne zalety) bardzo często wykorzystywanych topologii jest MLP (Rys.4.). Ten rodzaj sieci nie posiada połączeń wewnątrzwarstwowych, tylko międzywarstwowych (każdy neuron jednej warstwy łączy się z każdym neuronem kolejnej warstwy) a sygnał przebiega w jednym kierunku: od wejścia do wyjścia (bez sprzężeń zwrotnych).



Rys. 4. Model dwuwarstwowej sieci typu MLP
Źródło: Opracowanie własne

⁴ Wg danych IEEE (Institute of Electrical and Electronics Engineers) szacuje się, że obecnie na świecie co 5 sek. pojawia się nowa publikacja dot. sztucznych sieci neuronowych. [5]

⁵ W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, B. Math. Biophy. 5 (1943), 115-133

gdzie:

ξ – neurony warstwy wejściowej

V – neurony warstwy ukrytej

O – neurony warstwy wyjściowej

w_{jk}, W_{ij} – wagi połączeń pomiędzy neuronami

3. Proces budowy modelu neuronowego

Próba stworzenia jednego formalnego opisu zależności pomiędzy aktualnym stanem firmy, opisanym poprzez różnorodne wskaźniki mikro i makroekonomiczne, a jej kondycją ekonomiczno-finansową w ustalonym horyzoncie czasowym jest praktycznie niemożliwa. Dzieje się tak z tej przyczyny, że wpływ na to może mieć olbrzymia liczba wskaźników, a zależności z całą pewnością są nieliniowe. Co więcej, wybór odpowiednich wskaźników jest tym bardziej utrudniony, że ma on cechy lokalne. Jest zmienny w czasie jak i w przestrzeni różnych sektorów gospodarczych, regionów geograficznych czy systemów gospodarczych. Z całą pewnością można tylko powiedzieć: od czego może zależeć to, czy firma w przyszłości zbankrutuje, czy ma szansę na przeżycie.

W takich sytuacjach doskonale sprawdzają się indukcyjne metody uczenia maszynowego, a wśród nich SSN. Modelowanie neuronowe dobrze radzi sobie z problemami nieliniowości. Budowanie „lokalnych” i często zmieniających się modeli, ze względu na dużą łatwość ich tworzenia i reorganizacji, stanowi o ich dodatkowej atrakcyjności. Działają szybko, a ze względu na równoległe przetwarzanie danych – niezawodnie⁵. Potrafią nie tylko poprawnie klasyfikować, ale również wskazywać liczbowo stopień przynależności obiektów do danej klasy. Przy spełnieniu warunku normalnego rozkładu cech opisujących firmę, może być to interpretowane jako prawdopodobieństwo przynależności obiektu do danej klasy.

Jedyną niedogodnością SSN wykorzystywanych na obecnym etapie jest brak łatwo interpretowalnego opisu modelu w postaci reguł decyzyjnych. Taki opis nie tylko ułatwia lepsze zrozumienie problemu, ale z psychologicznego punktu widzenia, dodaje pewności, że model jest poprawny. Oczywiście, dzieje się tak tylko wtedy, gdy zbiór reguł jest zrozumiały, a ich liczba nie jest nadmiernie duża w stosunku do stopnia złożoności problemu, który opisują. Ale i w tej dziedzinie od kilkunastu lat uczyniono spory postęp w badaniach nad SSN⁶, a przyszłość może pokazać, że staną się one bardzo silnym narzędziem do budowania takich reguł.

Proces budowy modelu neuronowego składa się z wielu etapów. Pierwszy z nich to pozyskanie zbioru danych, który składa się z wektorów cech (faktów) w postaci:

$$(2) \quad X = [x_1, \dots, x_n]^T$$

gdzie: $x_i \in \{0;1\}$ lub $x_i \in \mathbb{R}$

Cechy x_i zawierają dane opisujące określony obiekt $O^{(k)}$. Zbiory uczące (wzorce) do modelowania neuronowego za pomocą sieci MLP zawierają dodatkowo informację do jakiej klasy należy każdy z nich.

$$(3) \quad U = \langle \langle X^{(1)}, Z^{(1)} \rangle, \dots, \langle X^{(k)}, Z^{(k)} \rangle \rangle$$

gdzie: $Z^{(k)}$ – jest etykietą kategorii k -tego obiektu ($O^{(k)}$)

Dla wszystkich metod indukcyjnych jest to moment kluczowy. Można powiedzieć, że spełnia się tu prosta zasada: „śmieci na wejściu – śmieci na wyjściu”. Zbiór danych powinien być reprezentatywny dla określonego problemu badawczego. Wzorce muszą być prawdziwe (pozbawione nadmiernych szumów wynikających z błędów pomiarowych, pomyłek, celowych przekłamań, itp.) i potencjalnie użyteczne. We wstępnym przetwarzaniu należy wyeliminować dane redundantne, niekompletne⁸ oraz te, których atrybuty mają wartości spoza swojej dziedziny. Wektor cech powinien być zredukowany do tych, które są istotne z punktu widzenia celu procesu⁹.

W zbiorze danych dobrze jest też określić, które z przypadków w nim zawartych są typowymi reprezentantami określonych klas, a które z nich położone są na granicy obszarów decyzyjnych. To one właśnie mogą stać się niezwykle cennym wzorcem w procesie uczenia sieci pod warunkiem, że są wiarygodne.

Aby SSN była zdolna do rozwiązywania jakiegokolwiek problemu, to po dobraniu odpowiednich parametrów strukturalnych takich jak: liczba warstw sieci, liczba neuronów w poszczególnych warstwach, architektura połączeń, wybór funkcji aktywacji i jej parametrów należy nauczyć ją rozpoznawania klas prezentowanych jej obiektów.

Proces uczenia, przedstawiony w sposób algorytmiczny na Rys.5., rozpoczynamy od nadania wagom losowych wartości początkowych w obrębie całej sieci. Następnie poprzez kolejno podawane przykłady uczące, odpowiedni algorytm uczenia, dokonuje stopniowej zmiany wag, tak aby otrzymać oczekiwane wartości sygnałów na wyjściu sieci.

Wzorzec X podany na wejściu neuronu z Rys.6. przetwarzany jest na wartość wyjściową y . Ta z kolei porównywana jest z oczekiwaną na wyjściu wartością z . Różnica obu sygnałów $\delta = 0$ oznacza, że neuron właściwie spełnia swoje zadanie i korekta wag nie jest potrzebna.

⁵ Uszkodzenie jednego z elementów sieci nie powoduje braku działania, ale co najwyżej obniżenie jego jakości.

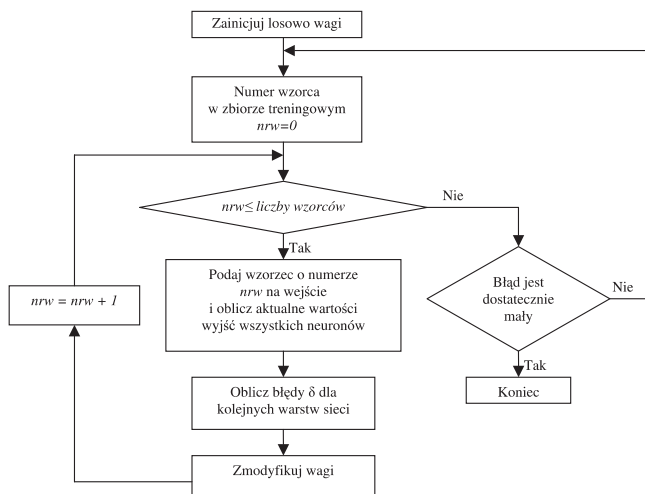
⁶ Kompendium wiedzy na ten temat można odnaleźć w: Gallant S., Neural Network Learning and Expert Systems, Cambridge, MA:MIT Press, 1995, a aktualny stan wiedzy i badań dotyczących tego tematu w Polsce w: K. Grąbczewski, W. Duch, R. Adamczak, Neuronowe metody odkrywania wiedzy w danych [w: 5].

⁷ Bobrowski L., Strategie projektowania sieci neuronowych [w: 5].

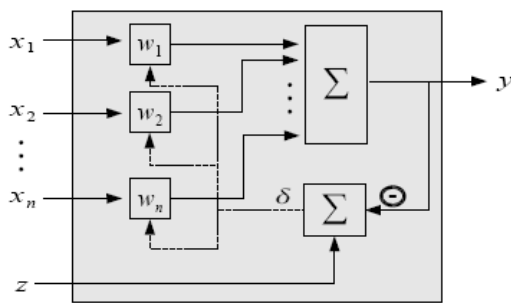
⁸ Przy małej liczności zbioru, gdy nie możemy sobie pozwolić na eliminowanie wzorców, brakujące wartości danych można uzupełnić m.in. średnią wartością cechy w zbiorze lub przez wprowadzenie do dziedziny wartości NULL.

⁹ Szeroki opis poświęcony tym zagadnieniom można znaleźć w: Marciniak A., Korbicz J., Kuś J., Wstępne przetwarzanie danych [w: 5].

W przeciwnym razie wagi w_i należy skorygować proporcjonalnie do wielkości błędu δ .



Rys. 5. Algorytm wstecznej propagacji błędu
Źródło: [16]



Rys. 6. Schemat uczenia sieci typu MLP

Źródło: Pypec G., Liniowe sieci neuronowe, www.is.umk.pl/~norbert/SemMagIn/Pypec.pdf, [25.07.2006 r.]

O ile prostym jest korygowanie wag neuronów wyjściowych, gdyż zawsze wiemy (zgodnie z (3)) jaka jest oczekiwana wartość z , o tyle dla neuronów warstw ukrytych (neurony V na Rys.4.) nie jest to już takie oczywiste. Intuicyjnie czujemy, że wielkość wprowadzanej korekty wag powinna być proporcjonalna do „udziału” danego wyjścia neuronu w powstawaniu finalnego błędu na wyjściu z sieci.

Aby ten problem rozwiązać algorytmicznie należy wprowadzić pojęcie kryterium oceny poprawności działania sieci. Jednym z nich, często używanym, może być funkcja w postaci:

$$(4) \quad Q = \frac{1}{2} \sum_{j=1}^k (z^{(j)} - y^{(j)})^2$$

Jak widać ze wzoru (4) stanowi ona sumaryczny błąd sieci po podaniu na jej wejście wszystkich wzorców ze zbioru uczącego. Celem procesu uczenia sieci jest doprowadzenie do możliwie pełnej zgodności pomiędzy $y^{(j)}$ a $z^{(j)}$, czyli do minimalizacji funkcji kryterialnej.

Jedną z możliwości stanowi wykorzystanie metody gradientowej. Zatem dla dowolnego neuronu wielkość poprawki dowolnej jego wagi można opisać wzorem:

$$(5) \quad \Delta w_i^{(j)} = -\eta \frac{\partial Q^{(j)}}{\partial w_i}$$

Współczynnik uczenia η określa jaką część poprawki chcemy wykorzystać w tym momencie uczenia sieci. Sieć uczymy w iteracyjnym procesie wielokrotnego podawania jej wszystkich wzorców ze zbioru, aż do momentu osiągnięcia zadawalającego nas rezultatu uczenia. W praktyce η w trakcie tego procesu powinno maleć i najczęściej przyjmujemy jego wartości w zakresie od 0,9 do 0,1.

Ponieważ Q jest zależny od y , który jest funkcją wektora wag W , więc prawą stronę równania (5) łatwo można przekształcić do postaci:

$$(6) \quad \frac{\partial Q^{(j)}}{\partial w_i} = \frac{\partial Q^{(j)}}{\partial y^{(j)}} \frac{\partial y^{(j)}}{\partial w_i}$$

Skoro zgodnie z (4):

$$(7) \quad Q^{(j)} = \frac{1}{2} (z^{(j)} - y^{(j)})^2$$

to:

$$(8) \quad \frac{\partial Q^{(j)}}{\partial y^{(j)}} = -(z^{(j)} - y^{(j)}) = \delta^{(j)}$$

Łatwo też zauważyć zależność (9) we wzorze (6):

$$(9) \quad \frac{\partial y^{(j)}}{\partial w_i} = x_i^{(j)}$$

Ostatecznie wielkość korekty dowolnej wagi, dowolnego neuronu w sieci, przy podaniu j -tego wzorca uczącego można przedstawić wzorem:

$$(10) \quad \Delta w_i^{(j)} = -\eta \delta^{(j)} x_i^{(j)}$$

Dodatkowe, proporcjonalne uzależnienie poprawki wag od wielkości sygnału x_i transmitowanego za pośrednictwem tej wagi, ma również intuicyjne uzasadnienie. Im większy sygnał x_i tym większy ma on wpływ na końcowy błąd i korekta wagi w_i , poprzez którą wpływa on na układ, powinna być większa. Jeśli wartość jego równa jest zero, czyli w żadnym stopniu nie wpływał on na powstanie błędu na wyjściu sieci, to wartość korekty, zgodnie ze wzorem (10), też wyniesie 0.

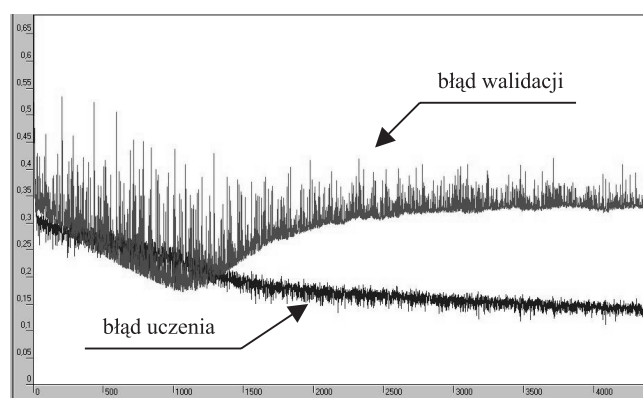
Na tej zasadzie oparta jest klasyczna metoda uczenia sieci z propagacją wsteczną błędu (ang. BackPropagation- BP) wprowadzona w 1974 r. przez Werbosa i ponownie odkryta i spopularyzowana w 1986 r. przez Rumelharta. Od tego czasu podlegała ona ciągłemu rozwojowi i na dzień dzisiejszy w symulatorach sieci neu-

ronowych można wykorzystywać ją zarówno w jej klasycznej jak i zmodyfikowanej postaci¹⁰.

Jak już zostało wspomniane, SSN w iteracyjnym procesie uczenia dąży do minimalizacji błędu na wyjściu sieci. Jednakże do poprawnej klasyfikacji przedsiębiorstw na te, o dobrej kondycji i te, o złej kondycji (jak i dla wielu innych zadań) funkcja kryterialna (4) jest mocno nieliniowa i posiada wiele minimów lokalnych. Z istoty swojej, zastosowana metoda gradientowa może spowodować, że w trakcie uczenia proces utknie w jednym z nich. Gdy to minimum lokalne będzie bardzo odległe od minimum globalnego, to nie osiągniemy oczekiwanych rezultatów. Oczywiście, jest wiele sposobów, które pozwalają na zminimalizowanie tego niebezpieczeństwa, m.in. wspomniana już wcześniej zmiana współczynnika η w trakcie procesu uczenia. Często jednak jedynym wyjściem może być wielokrotne rozpoczynanie procesu uczenia przy losowo, nowo zainicjowanych wagach z nadzieją, że proces rozpoczęty z innego miejsca w przestrzeni rozwiązań szczęśliwie uniknie niekorzystnych minimów lokalnych.

Niezmienne istotnym elementem procesu uczenia, mającym wpływ na jakość modelu, jest przyjęcie kryterium zakończenia procesu. W zależności od celu przetwarzania można przyjąć, że proces powinien się zakończyć m.in. gdy: przekroczona zostanie określona liczba epok uczenia¹¹, gdy błąd uczenia (4) spadnie poniżej wartości granicznej lub gdy zmiana błędu (4) będzie mniejsza od przyjętej na początku procesu.

Model prognozowania kondycji ekonomiczno-financej przedsiębiorstwa, jako model predykcyjny, nie może tylko dobrze „dopasować się” do prezentowanych danych. Przede wszystkim musi on dobrze uogólniać zależności ukryte w prezentowanych mu wzorcach. W związku z tym kryterium zatrzymania procesu uczenia powinno uwzględniać wielkość błędu obliczonego na niezależnym zbiorze danych. Zbiór taki, nazywany jest zbiorem walidacyjnym. Jeśli błąd ten spada podobnie jak błąd liczony na zbiorze uczącym, to mamy prawo przypuszczać, że SSN coraz lepiej aproksymuje interesujące nas zależności. Jeśli zaś błąd na zbiorze walidacyjnym zaczyna rosnąć, mimo dalszego spadku błędu na zbiorze uczącym, proces nauki należy przerwać. Taka sytuacja świadczy, że sieć zaczyna za bardzo dopasowywać się do wzorców zawartych w zbiorze uczącym i zaczyna tracić zdolność do generalizacji problemu. Zgodnie z wykresem pokazanym na Rys.7. proces uczenia (dla tego przypadku) powinien zakończyć się ok. 1200 epoki.



Rys. 7. Wykresy błędów: uczenia i walidacji

Źródło: Opracowanie własne

Ponieważ zgodnie z tak określoną procedurą, oba zbiory zaangażowane są w proces uczenia, to nie można na podstawie błędów obliczonych dla tych zbiorów wyciągać nieobciążonych wniosków na temat rzeczywistej jakości modelu. Idea stworzenia kolejnego, niezależnego zbioru testowego opiera się o teorię Kurta Gödla definiującą pojęcie otwartych systemów logicznych. Istota tej teorii polega na stwierdzeniu niemożności istnienia zamkniętych systemów twierdzeń. W konsekwencji oznacza to, że każdy zamknięty system musi odwołać się do reguł wyższego poziomu, należących do innego zbioru zasad. Aby uniknąć tej sprzeczności, w odniesieniu do modelu neuronowego, należy użyć innego zbioru danych oraz innego kryterium oceny. Jako zewnętrzne kryterium oceny przyjęto błąd klasyfikacji, B_{KL}^T zgodnie ze wzorem:

$$(11) \quad B_{KL}^T = \frac{L_B^T}{L_N^T} * 100\%$$

gdzie: L_B^T – liczba błędnie sklasyfikowanych przypadków na zbiorze testowym

L_N^T – liczba wszystkich przypadków w zbiorze testowym

W praktyce przyjmuje się, że jeśli dysponujemy liczebnością dostępnych wzorców k zgodnych z wzorcem (3), to podział zbioru danych należy zrobić wg proporcji (U- uczący, W- walidacyjny, T- testowy) tak, jak pokazuje formuła:

$$(12) \quad \begin{matrix} U & W & T \\ (0,5k, & 0,25k, & 0,25k) \end{matrix}$$

¹⁰ Szczegółowy opis metody, jak i przegląd podstawowych jej modyfikacji zawiera m.in. praca: Świąć A., Bilski J., Metoda wstecznej propagacji błędów i jej modyfikacje [w: 5].

¹¹ W jednej epoce podawane są na wejście sieci wszystkie wzorce ze zbioru uczącego.

4. Przykład realizacji modelu neuronowego do prognozowania kondycji ekonomiczno-finansowej przedsiębiorstwa

Do ilustracji zagadnień teoretycznych przedstawionych w poprzednim punkcie, wykorzystano fragment badań realizowanych w ramach projektu EQUAL¹². Wybrany model dotyczy sekcji: Transport, Gospodarka magazynowa i łączność (PKD: 601_642) dla woj. podkarpackiego.

Zgodnie z zaproponowaną, uproszczoną koncepcją budowy modelu neuronowego badanie podzielono na etapy:

1. Pozyskanie zbioru danych

Dane dot. małych i średnich firm z regionów: podkarpackiego i lubelskiego pozyskano w Urzędach Statystycznych. Dane makro i mezoekonomiczne z roczników statystycznych. Szczegółowy ich opis zawarty jest w artykule: Kowerski M., Koncepcja badań sektora małych i średnich przedsiębiorstw w projekcie „System przeciwdziałania bezrobociu na obszarach słabo zurbanizowanych” otwierającym niniejszą publikację.

Przygotowano zbiory w postaci (3). W trakcie wstępnego przetwarzania, korzystając z mechanizmów filtrowania zawartych w relacyjnych bazach danych, usunięto przypadki niespójne z dziedzinami atrybutów.

2. Wybór potencjalnie użytecznych zmiennych wejściowych (opisujących)

Do wyboru cech zastosowano metodę opartą na analizie wrażliwości sieci. Wstępnie wygenerowano losowo 50 sieci MLP uczonych na zbiorze wszystkich dostępnych przypadków, tak aby uzyskać możliwie pełną informację odnośnie zależności pomiędzy wejściem i wyjściem sieci. Oceniono je na podstawie wskaźnika jakości (Jakość uczenia¹³). Przy uzyskaniu wskaźnika na równoważnym poziomie kierowano się mniejszym wymiarem zbioru wejściowego.

Uporządkowany obraz wybranych sieci przedstawia Tablica 1. Na podstawie powyższych ustaleń, do dalszych badań wybrano sieć nr 6- MLP 24:24-16-16-1:1 (z 24 neuronami na wejściu, 16 w dwóch warstwach ukrytych i 1 neuronem wyjściowym, decyzyjnym).

Wybraną sieć poddano analizie wrażliwości. Wartość wskaźnika wrażliwości Wr obliczana jest zgodnie ze wzorem (13), jako iloraz wskaźnika błędów sieci kompletnej (B) oraz sieci po usunięciu określonego wejścia (BW_x). Im większa wartość ilorazu dla danego wejścia, tym większy jest jego wpływ na jakość odwzorowania dokonywanego przez sieć. Wartości poniżej 1 sugerują, że obecność tego wejścia tylko pogarsza działanie modelu. Takie wskaźniki usuwane są w pierwszej kolejności.

$$(13) \quad Wr = \frac{B}{BW_x}$$

Aby tworzony model był wiarygodny, powinniśmy dodatkowo zadbać o to, aby liczba przypadków uczących L^U dla określonego wymiaru przestrzeni wejść N była wystarczająca. Do tego celu wykorzystuje się empiryczny wzór w postaci:

$$(14) \quad L^U = 2^N$$

Dysponując zbiorem 112 przypadków, z czego tylko 66, zgodnie z regułą (12), zostanie użytych w zbiorze uczącym- powinniśmy ograniczyć się do 6 wejść. Z uwagi na obawę, o zbyt dużą utratę informacji zawartej w zbiorze (świadczą o tym bardzo wysokie wartości wskaźników w Tablicy 2. z rangą pow. 6), zdecydowano o pozostawieniu 8-miu:

- X_2 - Udział zapasów w aktywach ogółem
- X_3 - Udział należności w aktywach ogółem
- X_6 - Obciążenie zobowiązaniami długoterminowymi

Tablica 1. Sieci MLP do wyboru zmiennych wejściowych (wybrane spośród 50)

| Sieci, podsumowanie (W601_642_ocz) | | | | | |
|------------------------------------|---------------------|-------------|-----------|------------------|----------|
| Nr | Typ | Jakość ucz. | Błąd ucz. | Uczenie | Wejścia |
| 1 | MLP 17:17-15-1:1 | 0,473214 | 0,621482 | BP100,CG20,CG1b | 17,00000 |
| 2 | MLP 21:21-11-1:1 | 0,901786 | 0,360634 | BP100,CG20,CG2b | 21,00000 |
| 3 | MLP 32:32-20-18-1:1 | 0,955357 | 0,241036 | BP100,CG20,CG2b | 32,00000 |
| 6 | MLP 24:24-16-16-1:1 | 0,955357 | 0,216214 | BP100,CG20,CG7b | 24,00000 |
| 7 | MLP 35:35-16-1:1 | 0,955357 | 0,211242 | BP100,CG20,CG8b | 35,00000 |
| 10 | MLP 33:33-20-18-1:1 | 0,982143 | 0,180011 | BP100,CG20,CG15b | 33,00000 |
| 13 | MLP 37:37-20-19-1:1 | 0,991071 | 0,162747 | BP100,CG20,CG3b | 37,00000 |
| 32 | MLP 32:32-20-17-1:1 | 0,991071 | 0,110263 | BP100,CG20,CG14b | 32,00000 |

¹² Tytuł projektu: System przeciwdziałania powstaniu bezrobocia na terenach słabo zurbanizowanych. Projekt realizowany przy udziale środków Europejskiego Funduszu Społecznego w ramach Inicjatywy Wspólnotowej EQAL.

Tablica 2. Analiza wrażliwości sieci do wyboru danych wejściowych

| Analiza wrażliwości - 6 (W601_642_ocz) | | | | | | | | |
|--|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | X ₂ | X ₃ | X ₆ | X ₁₄ | X ₁₇ | X ₃₂ | Z ₁₁ | Z ₁₂ |
| Iloraz.6 | 1,479139 | 1,556893 | 1,562745 | 1,496943 | 1,959442 | 2,461956 | 1,567960 | 1,779257 |
| Ranga.6 | 8,000000 | 6,000000 | 5,000000 | 7,000000 | 2,000000 | 1,000000 | 4,000000 | 3,000000 |

Tablica 3. Ostateczne modele MLP (wybór)

| Sieci, podsumowanie (W601_642_ocz) | | | | |
|------------------------------------|-------------------|-------------|---------------|--------------|
| Nr sieci | Typ | Jakość ucz. | Jakość walid. | Jakość test. |
| 170 | MLP 8:8-7-1:1 | 0,875000 | 0,678571 | 0,750000 |
| 125 | MLP 8:8-11-9-1:1 | 0,803571 | 0,714286 | 0,750000 |
| 195 | MLP 8:8-11-10-1:1 | 0,875000 | 0,714286 | 0,750000 |
| 218 | MLP 8:8-11-9-1:1 | 0,875000 | 0,678571 | 0,785714 |
| 235 | MLP 8:8-11-11-1:1 | 0,892857 | 0,678571 | 0,785714 |
| 166 | MLP 8:8-11-11-1:1 | 0,910714 | 0,678571 | 0,892857 |
| 150 | MLP 8:8-9-1:1 | 0,785714 | 0,785714 | 0,892857 |

X₁₄ - Niedobór kapitału obrotowego netto

X₁₇ - Wskaźnik poziomu kosztów finansowych

X₃₂ - Udział sprzedaży na eksport w sprzedaży ogółem

Z₁₁ - Dynamika produkcji sprzedanej w przemyśle 1999=0

Z₁₂ - Dynamika liczby bezrobotnych 1999 = 0

3. Proces uczenia modelu

Wygenerowano losowo 50 sieci MLP z 8 wejściami, na które w trakcie procesu uczenia podawano wartości wskaźników z Tablicy 2. Losowy podział zbioru wszystkich przypadków na U, W i T inicjowano zgodnie z formułą (12) dla każdej z nich niezależnie. Sieci oceniono na podstawie wskaźników jakościowych dla wszystkich zbiorów, ze szczególnym uwzględnieniem wartości dla zbioru testowego i równomiernego rozkładu wartości we wszystkich 3 zbiorach. Na podstawie analizy wskaźników z Tablicy 3. wybrano sieć 150 (MLP 8:8-9-1:1).

4. Testowanie wybranego modelu

W przypadku procesów klasyfikacyjnych, należy wyznaczyć udział klasyfikacji poprawnych w stosunku do liczby wszystkich dostępnych przypadków, zgodnie ze wzorem (13).

W Tablicy 4. pokazane zostały błędy I rodzaju (klasyfikacja firm o złej kondycji- Y2.0.150) i II rodzaju (klasyfikacja firm o dobrej kondycji- Y2.1.150). Należy zaznaczyć, że nie są one równoważne. Znacznie groźniej-

szy jest ten pierwszy, gdyż informacja o dobrej kondycji ekonomiczno-finansowej dla zarządu firmy zagrożonej upadłością może mieć groźniejsze konsekwencje niż sytuacja odwrotna.

Tablica 4. Statystyki klasyfikacji dla całego zbioru

| Klasyfikacja (150) (W601_642_ocz) | | |
|------------------------------------|----------|----------|
| | Y2.0.150 | Y2.1.150 |
| Wszystkie | 25,00000 | 87,00000 |
| Poprawne | 21,00000 | 70,00000 |
| Niepoprawne | 4,00000 | 17,00000 |
| Nieznane | 0,00000 | 0,00000 |
| Poprawne(%) | 84,00000 | 80,45977 |
| Niepoprawne (%) | 16,00000 | 19,54023 |
| Nieznane(%) | 0,00000 | 0,00000 |

Dane zawarte w Tablicy 4. pokazują, że model bardzo dobrze poradził sobie z problemem klasyfikacji firm w tym sektorze gospodarki, co więcej, błąd I rodzaju wynosi tylko 16% i jest znacząco mniejszy od błędu II rodzaju.

¹³ Iloraz odchyżeń standardowych dla błędów i dla danych. Im większa jest jego wartość, tym lepsze zdolności predykcji modelu. Jeśli sieć nie będzie popełniać żadnych błędów, osiągnie on wartość 1.

| Klasyfikacja (150) (W601_642_ocz) | | | | | | |
|------------------------------------|------------|------------|------------|------------|------------|------------|
| | U.Y2.0.150 | U.Y2.1.150 | W.Y2.0.150 | W.Y2.1.150 | T.Y2.0.150 | T.Y2.1.150 |
| Wszystkie | 11,0000 | 45,00000 | 8,00000 | 20,00000 | 6,00000 | 22,00000 |
| Poprawne | 11,0000 | 33,00000 | 5,00000 | 17,00000 | 5,00000 | 20,00000 |
| Niepoprawne | 0,0000 | 12,00000 | 3,00000 | 3,00000 | 1,00000 | 2,00000 |
| Nieznane | 0,0000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 |
| Poprawne (%) | 100,0000 | 73,33333 | 62,50000 | 85,00000 | 83,33333 | 90,90909 |
| Niepoprawne(%) | 0,0000 | 26,66667 | 37,50000 | 15,00000 | 16,66667 | 9,09091 |
| Nieznane(%) | 0,0000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 | 0,00000 |

Tablica 5. Statystyki klasyfikacji w podzbiorach U, W, T

Jednak prawdziwym testem wiarygodności modelu predykcyjnego, może być potwierdzenie tych wyników na niezależnym zbiorze testowym T. Dane dot. błędów klasyfikacji dla tego zbioru (Tablica 5.), są zaskakująco dobre, a nawet lepsze od średniej dla całego zbioru przypadków (Tablica 4.). Może to świadczyć o tym, że do zbioru przypadków uczących losowo wybrane zostały „najtrudniejsze”. Te, które znalazły się na granicy obszarów decyzyjnych, jak również przypadki obarczone szumami. Taki podział wzorców często jest najlepszym sposobem na rozwiązanie problemów uczenia w trudnych sytuacjach klasyfikacyjnych. Pewnym obciążeniem wiarygodności tych statystyk może być mała liczba przypadków w tym zbiorze (6 i 22).

Kolejnym potwierdzeniem jakości modelu może być test na zbiorze danych dla działu (tej samej sekcji) 601_603: Transport lądowy, transport rurociągowy (Tablica 6.). W tym wypadku jest to uprawnione ze względu na duże podobieństwo specyfiki działu względem całej sekcji.

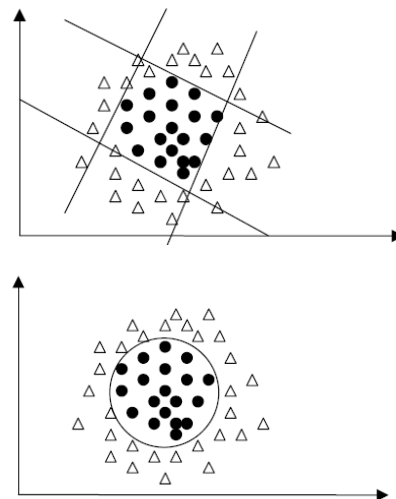
Bardzo dobre wyniki (16 i 22% błędu) trzeba również widzieć w kontekście faktu, że duża część wzorców z tego zbioru zawarta była również w zbiorze całej sekcji jako zbiorze nadrzędnym.

| Klasyfikacja (150) (W601_603_ocz) | | |
|------------------------------------|----------|----------|
| | Y2.0.150 | Y2.1.150 |
| Wszystkie | 24,00000 | 63,00000 |
| Poprawne | 20,00000 | 49,00000 |
| Niepoprawne | 4,00000 | 14,00000 |
| Nieznane | 0,00000 | 0,00000 |
| Popraw.(%) | 83,33333 | 77,77778 |
| Niepopr.(%) | 16,66667 | 22,22222 |
| Nieznane(%) | 0,00000 | 0,00000 |

Tablica 6. Klasyfikacja modelu dla działu 601_603

5. Podsumowanie

Zaprezentowany w tym opracowaniu model neuronowy o architekturze MLP nie jest jedynym, który doskonale sprawdza się w zagadnieniach klasyfikacyjnych. Często lepsze statystyki można uzyskać korzystając z sieci o architekturze RBF. Sieci te, o radialnych funkcjach bazowych, szczególnie przydatne mogą być wtedy, gdy MLP nie daje satysfakcjonujących wyników ze względu na kształty obszarów decyzyjnych.

Rys. 8. Obszary decyzyjne MLP- z lewej, RBF – z prawej
Źródło: [1]

Sieci MLP z logistyczną funkcją aktywacji separują przy pomocy hiperpłaszczyzn (Rys.8.), zaś w sieciach RBF, gdzie funkcją aktywacji jest funkcją Gaussa, separacja odbywa się przy pomocy hipersfer. Ten przykład pokazuje jak charakterystyczne cechy zbiorów danych mogą wpływać, a wręcz wymuszać wybór odpowiedniej architektury. Oba typy sieci uczone są pod nadzorem, czyli zgodnie z wzorcem (3) przypadek uczący zawiera etykietę klasy, do której został zaklasyfikowany.

Innym rodzajem sieci, których wysokie zdolności do klasyfikacji potwierdzają liczne badania, są sieci SOM (Self-Organizing Map). Należą one do grupy sieci, które uczone są bez nadzoru. W uproszczeniu możemy powiedzieć, że sieć zna tylko wzorzec, ale nie zna informacji, co do przynależności tego wzorca do określonej klasy. Co więcej, nie jest informowana ile klas daje wyróżnić się w prezentowanym jej zbiorze. Mechanizm uczenia oparty jest o tzw. konkurencję neuronów, które starają się coraz precyzyjniej (w trakcie procesu uczenia precyzja powinna rosnąć) reagować na wzorce do siebie podobne.

Prowadzone badania w Wyższej Szkole Zarządzania i Administracji w Zamościu z użyciem sztucznych sieci neuronowych pokazują dużą przydatność tego narzędzia do predykcji kondycji ekonomiczno-finansowej małych i średnich przedsiębiorstw. Wyniki zaprezentowane w tym artykule nie odbiegają znacząco od wyników w pozostałych sektorach gospodarki. Nie odbiegają też one od wyników prezentowanych w literaturze światowej dla badań prowadzonych na dużych firmach, które wycinkowo przedstawiono we wstępie do tego opracowania. Potwierdza to tezę, że modele neuronowe, stając się swoistym systemem wczesnego ostrzegania, mogą przyczynić się do stabilizacji funkcjonowania przedsiębiorstw i obniżyć groźbę ich upadłości. W konsekwencji może to doprowadzić do złagodzenia problemów bezrobocia.

Literatura:

- Adamczak R., Zastosowanie sieci neuronowych do klasyfikacji danych doświadczalnych, praca doktorska pod kierunkiem prof. Włodzisława Duchy, Katedra Metod Komputerowych, Uniwersytet Mikołaja Kopernika, Toruń 2001
- Alici Y., Neural networks in corporate failure prediction: The UK experience, in Proc. Third Int. Conf. Neural Networks in the Capital Markets, A. N. Refenes, Y. Abu-Mostafa, J. Moody, and A. Weigend, Eds. London, UK, Oct. 1995, pp. 393 406
- Altman E., Marco G., Varetto F., Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks, J. Banking and Finance, vol. 18, pp. 505 529, 1994
- Coats P., Fant L., Recognizing financial distress patterns using a neural network tool, Financial Management, vol. 22, pp. 142 155, 1993
- Duch W., Korbicz J., Rutkowski L., Tadeusiewicz R. (red.), Biocybernetyka i inżynieria biomedyczna 2000, t. 6., Sieci neuronowe, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2000
- Fernandez E., Olmeda I., Bankruptcy prediction with artificial neural networks, Lect. Notes Comput. Sc., pp. 1142 1146, 1995
- Hertz J., Krogh A., Palmer R., Wstęp do teorii obliczeń neuronowych, WNT, Warszawa 1993
- Odom M., Sharda R., A neural network model for bankruptcy prediction, in Proc. Int. Joint conf. Neural networks, San Diego, CA, 1990
- Osowski S., In-flight simulation at the NASA Dryden flight research facility. AIAA-91-2916-CP, Flight Simulation Technologies Conference, 1994, s. 7-23.
- Piramuthu S., Raghavan H., Shaw M., Using feature construction to improve the performance of neural networks, Management Science, vol. 44, pp. 416 430, 1998
- Privotet J.-C., Neural Networks: Architectures, preprocessing and hardware implementation, VIII International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Moscow, Russia, 24 - 28 June 2002
- Salchenberger L., Cinar E., Lash N., Neural networks: A new tool for predicting thrift failures, Decision Sciences, vol. 23, pp. 899 916, 1992
- Tam K., Kiang M., Managerial applications of the neural networks: The case of bank failure predictions, Management Science, vol. 38, pp. 416 430, 1992
- Tam K., Neural network models and the prediction of bank bankruptcy, Omega, vol. 19, pp. 429 445, 1991
- Witkowska D., Sztuczne sieci neuronowe i metody statystyczne. Wybrane zagadnienia finansowe Wydawnictwo C.H. Beck, 2002
- Zieliński J. (red.), Inteligentne systemy w zarządzaniu. Teoria i praktyka, PWN, Warszawa 2000
- Zhang G., Hu M., Patuwo B., Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis, European J. Oper. Res., vol. 116, pp. 16 32, 1999