PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 103
Prognozowanie w zarządzaniu firmą 2010

Joanna Perzyńska
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

BUDOWA PROGNOZ KOMBINOWANYCH Z WYKORZYSTANIEM SZTUCZNYCH SIECI NEURONOWYCH

Streszczenie: W artykule przedstawiono metodę budowy liniowych prognoz kombinowanych wykorzystującą sztuczne sieci neuronowe. Ilustracją rozważań o charakterze teoretycznym jest przykład empiryczny, w którym prognozy indywidualne i kombinowane wyznaczono dla zmiennej ekonomicznej z wahaniami sezonowymi. Dokładność *ex post* prognoz kombinowanych zbudowanych z wykorzystaniem sztucznych sieci neuronowych porównano z dokładnością ich prognoz składowych oraz prognoz kombinowanych otrzymanych metodami ekonometrycznymi.

Słowa kluczowe: prognozy kombinowane, sztuczne sieci neuronowe.

1. Wstęp

W sytuacji, gdy dostępne są różne prognozy tej samej zmiennej, można utworzyć liniową prognozę kombinowaną będącą ich średnią ważoną. Przy założeniu istnienia nieliniowego związku pomiędzy prognozami indywidualnymi a prognozowaną zmienną prognoza kombinowana jest nieliniową funkcją prognoz składowych i można ją wówczas wyznaczyć za pomocą sztucznych sieci neuronowych (zob. [Da Xu, Liu, Ming Shi 1999]).

W niniejszym artykule zostanie zaproponowany sposób konstrukcji sztucznej sieci neuronowej stosowanej do wyznaczenia wag liniowej prognozy kombinowanej. W toku badań empirycznych zweryfikowana będzie hipoteza mówiąca, iż tak otrzymana prognoza kombinowana charakteryzuje się większą dokładnością niż jej prognozy składowe i prognozy kombinowane z wagami otrzymanymi metodami ekonometrycznymi.

2. Metody badawcze

Niech $f_{1T}, f_{2T}, ..., f_{mT}$ będą różnymi prognozami wartości zmiennej y na okres T. Prognoza kombinowana f_{cT} wartości y_T będąca liniową kombinacją m ($m \geq 2$) prognoz indywidualnych ma postać:

$$f_{cT} = \lambda_1 f_{1T} + \lambda_2 f_{2T} + ... + \lambda_m f_{mT} = \sum_{i=1}^{m} \lambda_i f_{iT}, \tag{1}$$

gdzie:

$$\sum_{i=1}^{m} \lambda_i = 1. \tag{2}$$

Konstrukcja sztucznej sieci neuronowej służącej wyznaczaniu prognozy kombinowanej opiera się na zauważonej przez autorkę analogii pomiędzy sposobem obliczania prognozy kombinowanej na podstawie wzoru (1) i strukturą sztucznego neuronu liniowego, która zdefiniowana jest następująco (por. [Azoff 1994; Lula, Tadeusiewicz 2001]):

1. Do neuronu dociera m sygnałów: x_1, x_2, ..., x_m – wartości wejściowych pierwotnych lub pośrednich.

2. Każda wartość wejściowa x_i wprowadzana jest do neuronu przez połączenie o określonej wadze w_i (i = 1, 2, ..., m).

3. W neuronie odbywa się przetwarzanie sygnałów wejściowych składające się z dwóch etapów: agregacji wartości wejściowych i wyznaczenia wartości wyjściowej neuronu. Agregacja wartości wejściowych ma na celu przetworzenie ich w pojedynczą zagregowaną wartość wejściową s nazywaną łącznym pobudzeniem neuronu. Łączne pobudzenie neuronu obliczane jest jako wartość funkcji potencjału postsynaptycznego dla neuronu liniowego w postaci ważonej sumy wejść:

$$s = w_0 + \sum_{i=1}^{m} w_i x_i, \tag{3}$$

gdzie: w_0 – wartość obciążenia neuronu.

Wartość łącznego pobudzenia neuronu przekształcana jest przez funkcję aktywacji φ, co wyznacza wartość wyjściową neuronu:

$$Y = \varphi(s). \tag{4}$$

Łącząc ze sobą sztuczne neurony, tworzy się sztuczną sieć neuronową (SSN). Neurony w sieci ułożone są w warstwach, wyjścia jednej warstwy połączone są z wejściami kolejnych warstw. Dane wprowadzone do SSN przez warstwę wejściową przesyłane są do kolejnych warstw neuronów, gdzie są przetwarzane na podstawie wybranych funkcji potencjału postsynaptycznego i aktywacji zgodnie ze wzorami (3) i (4). Ich przetwarzanie kończy się w warstwie wyjściowej, sygnały wyznaczone przez jej neurony są jednocześnie wartościami wyjściowymi całej sieci.

Zauważona analogia występuje pomiędzy wzorami (1) oraz (3) – obie wartości wyznaczane są jako sumy ważone, co sugeruje możliwość zastosowania sztucznych sieci neuronowych do wyznaczania prognoz kombinowanych. Przed skonstruowaniem sieci należy dokonać pewnych przekształceń na wzorze (1), które pozwolą określić jej wartości wejściowe i wyjściowe. Ponieważ zgodnie z założeniem określonym wzorem (2) suma wag ma być równa jedności, to:

$$\lambda_m = 1 - \sum_{i=1}^{m-1} \lambda_i. \tag{5}$$

Na podstawie wzorów (1) i (5) mamy więc:

$$f_{cT} = \lambda_m f_{mT} + \sum_{i=1}^{m-1} \lambda_i f_{iT} = f_{mT} - f_{mT} \sum_{i=1}^{m-1} \lambda_i + \sum_{i=1}^{m-1} \lambda_i f_{iT} = f_{mT} + \sum_{i=1}^{m-1} \lambda_i \left(f_{iT} - f_{mT} \right). \tag{6}$$

Błędy prognoz składowych i prognozy kombinowanej wynoszą odpowiednio:

$$e_{iT} = y_T - f_{iT}, \tag{7}$$

$$e_{cT} = y_T - f_{cT}, \tag{8}$$

skąd po dalszych przekształceniach wzoru (6) otrzymuje się:

$$y_T - e_{cT} = y_T - e_{mT} + \sum_{i=1}^{m-1} \lambda_i \left(y_T - e_{iT} - \left(y_T - e_{mT} \right) \right), \tag{9}$$

$$e_{mT} = e_{cT} + \sum_{i=1}^{m-1} \lambda_i \left(e_{mT} - e_{iT} \right). \tag{10}$$

Z porównania wyprowadzonego wzoru (10) i wzoru (3) wynika, że konstruowana sieć mieć będzie ($m - 1$) neuronów w warstwie wejściowej, a jej wartościami wejściowymi będą różnice błędów prognoz składowych[1] określone wzorem:

$$x_i = e_{mT} - e_{iT}, \tag{11}$$

gdzie: $i = 1, 2, ..., m - 1$ ($m \geq 2$).

Jeżeli wartości x_i wprowadzone zostaną do pojedynczego neuronu liniowego drugiej warstwy (będącej zarazem ostatnią warstwą) połączeniami o wagach:

$$w_i = \lambda_i, \tag{12}$$

a wartość obciążenia neuronu równa będzie:

$$w_0 = e_{cT}, \tag{13}$$

[1] Wzór (11) wykorzystuje się podczas uczenia sieci, są to wówczas różnice znanych błędów prognoz z wcześniejszych okresów ($t < T$) lub reszt modeli, na podstawie których wyznaczono prognozy składowe. W zastosowaniach nauczonej sieci (gdy nie są znane błędy prognoz składowych) można wykorzystać równoważną postać wzoru (11): $x_i = f_{iT} - f_{mT}$.

to wyznaczona przez funkcję potencjału postsynaptycznego wartość łącznego pobudzenia wyniesie:

$$s = w_0 + \sum_{i=1}^{m} w_i x_i = e_{cT} + \sum_{i=1}^{m} \lambda_i \left(e_{mT} - e_{iT} \right). \tag{14}$$

Zastosowanie w neuronie ostatniej warstwy tożsamościowej funkcji aktywacji

$$\varphi\,(s) = s \tag{15}$$

sprawi, że wartość wyjściowa sieci będzie równa wyznaczonej wartości łącznego pobudzenia:

$$Y = e_{cT} + \sum_{i=1}^{m} \lambda_i \left(e_{mT} - e_{iT} \right). \tag{16}$$

Z rozważań określonych wzorami (6)-(10) oraz (16) wynika, iż:

$$Y = e_{mT}, \tag{17}$$

a pomiędzy wartością prognozy kombinowanej (1) a wartością wyjściową sieci zachodzi następująca zależność:

$$f_{cT} = Y + f_{mT} - e_{cT}. \tag{18}$$

Wagi połączeń, którymi do sieci wprowadza się dane wejściowe, są szukanymi wagami prognoz składowych w prognozie kombinowanej, ich wartości wyznacza się w procesie uczenia sieci działającej na opisanej wcześniej zasadzie.

W celach porównawczych wagi prognozy kombinowanej (1) wyznaczone zostaną również metodami ekonometrycznymi:

– metodą wariancji-kowariancji, zgodnie z którą wariancja błędu prognozy kombinowanej jest minimalizowana dla wag oszacowanych na podstawie wzoru [Bates, Granger 1969; Granger, Newbold 1974]:

$$\boldsymbol{\lambda} = \hat{\boldsymbol{\Omega}}^{-1}\mathbf{I} \big/ \mathbf{I}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{I}, \tag{19}$$

gdzie:

$$\boldsymbol{\lambda} = [\lambda_1\, \lambda_2 \ldots \lambda_m]', \quad \hat{\Omega}_{ij} = \frac{1}{v} \sum_{t=T-v}^{T-1} e_{it} e_{jt},$$

– na podstawie wzoru zaproponowanego przez Batesa i Grangera [Bates, Granger 1969]:

$$\lambda_i = \left(\sum_{t=T-v}^{T-1} e_{it}^{\,2} \right)^{-1} \Bigg/ \sum_{j=1}^{m} \left(\sum_{t=T-v}^{T-1} e_{jt}^{\,2} \right)^{-1}, \tag{20}$$

będącego szczególnym przypadkiem wzoru (19) przy założeniu zerowej korelacji pomiędzy błędami prognoz składowych.

3. Materiał badawczy

Modelowaniu predyktywnemu i prognozowaniu poddano kształtowanie się kosztów produkcji betonu w oddziale 1 przedsiębiorstwa A w ujęciu miesięcznym (K1). Zmienne objaśniające stanowiły zmienne *SP*1 oraz *PR*1 przedstawiające odpowiednio sprzedaż oraz produkcję betonu w tym samym oddziale.

Badana zmienna charakteryzuje się występowaniem wyraźnego trendu oraz wahań sezonowych. W tabeli 1 przedstawiono oceny wskaźników sezonowości zmiennych *K*1, *SP*1 oraz *PR*1 wyznaczone przy założeniu multiplikatywnego charakteru wahań sezonowych.

Tabela 1. Oceny wskaźników sezonowości zmiennych *K*1, *SP*1 i *PR*1

Miesiąc	Wskaźnik sezonowości (%)		
	*K*1	*SP*1	*PR*1
I	78,35	78,36	57,23
II	70,68	57,14	62,71
III	90,10	94,79	87,21
IV	92,20	103,30	94,01
V	94,96	91,34	100,93
VI	107,09	99,33	106,74
VII	112,27	115,12	123,15
VIII	113,12	123,00	122,37
IX	118,19	133,11	133,02
X	120,28	138,43	126,63
XI	108,18	93,70	100,12
XII	94,57	72,39	85,87

Źródło: obliczenia własne.

Wyznaczone oceny wskaźników sezonowości zmiennych *K*1, *SP*1 oraz *PR*1 wykazują znaczne zróżnicowanie – różnice między wartościami maksymalną i minimalną wynoszą odpowiednio: 49,6, 81,29 oraz 75,79 punktów procentowych. Zmienne te charakteryzują się silnym natężeniem sezonowości o asymetrycznym rozkładzie ocen wskaźników sezonowości – zmienne *K*1 oraz *SP*1 swoje maksimum sezonowe osiągają w październiku, a minimum sezonowe – w lutym, natomiast zmienna *PR*1 swoje maksimum sezonowe osiąga we wrześniu, a minimum sezonowe – w styczniu.

4. Prezentacja i ocena wyników badań

W procesie modelowania predyktywnego zmiennej $K1$ wykorzystano klasyczne i hierarchiczne modele szeregu czasowego, modele przyczynowo-opisowe, hierarchiczne modele przyczynowo-opisowe oraz sztuczne sieci neuronowe. Okres estymacyjny obejmował 48 obserwacji. Oszacowano następujące modele:

1. Klasyczne modele szeregu czasowego (MK):

$$y_t = \beta_0 + \beta_1 t + \sum_{k=1}^{12} d_{0k} Q_k + \sum_{k=1}^{12} d_{1k} t Q_k + \xi_t, \tag{21}$$

gdzie: $\sum_{k=1}^{12} d_{0k} = \sum_{k=1}^{12} d_{1k} = 0.$

2. Hierarchiczne modele szeregu czasowego dwu i trzypoziomowe (Hp_1p_2 i $Hp_1p_2p_3$) (zob. [Zawadzki (red.) 2003]):

$$\begin{aligned} y_{srlt} = {} & \beta_0 + \beta_1 t + \sum_{s=1}^{p_1} b_{0s} Q_s + \sum_{r=1}^{p_2} b_{0sr} Q_{sr} + \sum_{l=1}^{p_3} b_{0srl} Q_{srl} + \\ & + \sum_{s=1}^{p_1} b_{1s} t Q_s + \sum_{r=1}^{p_2} b_{1sr} t Q_{sr} + \sum_{l=1}^{p_3} b_{1srl} t Q_{srl} + \xi_{srlt}, \end{aligned} \tag{22}$$

gdzie: $\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0sr} = \sum_{l=1}^{p_3} b_{0srl} = \sum_{s=1}^{p_1} b_{1s} = \sum_{r=1}^{p_2} b_{1sr} = \sum_{l=1}^{p_3} b_{1srl} = 0.$

3. Liniowe modele przyczynowo-opisowe (P):

$$y_t = \beta_0 + \beta_1 t + \alpha x_t + \sum_{k=1}^{12} a_k x_t Q_k + \sum_{k=1}^{12} b_{0k} Q_k + \sum_{k=1}^{12} b_{1k} t Q_k + \xi_t, \tag{23}$$

gdzie: $\sum_{k=1}^{12} b_{0k} = \sum_{k=1}^{12} b_{1k} = \sum_{k=1}^{12} a_k = 0.$

4. Hierarchiczne modele przyczynowo-opisowe dwu i trzypoziomowe (PHp_1p_2 i $PHp_1p_2p_3$); zob. [Zawadzki (red.) 2003]:

$$\begin{aligned} y_{srlt} = {} & \beta_0 + \beta_1 t + \alpha x_t + \sum_{s=1}^{p_1} a_s x_t Q_s + \sum_{r=1}^{p_2} a_{sr} x_t Q_{sr} + \sum_{l=1}^{p_3} a_{srl} x_t Q_{srl} + + \sum_{s=1}^{p_1} b_{0s} Q_s + \\ & + \sum_{r=1}^{p_2} b_{0sr} Q_{sr} + \sum_{l=1}^{p_3} b_{0srl} Q_{srl} + \sum_{s=1}^{p_1} b_{1s} t Q_s + \sum_{r=1}^{p_2} b_{1sr} t Q_{sr} + \sum_{l=1}^{p_3} b_{1srl} t Q_{srl} + \xi_{srlt}, \end{aligned} \tag{24}$$

gdzie:

$$\sum_{s=1}^{p_1} b_{0s} = \sum_{r=1}^{p_2} b_{0sr} = \sum_{l=1}^{p_3} b_{0srl} = \sum_{s=1}^{p_1} b_{1s} = \sum_{r=1}^{p_2} b_{1sr} = \sum_{l=1}^{p_3} b_{1srl} = \sum_{s=1}^{p_1} a_s = \sum_{r=1}^{p_2} a_{sr} = \sum_{l=1}^{p_3} a_{srl} = 0,$$

p_1, p_2, p_3 – podzielniki długości cyklu wahań okresowych odpowiadające kolejnym stopniom hierarchii.

W tabeli 2 przedstawiono wybrane oceny parametrów struktury stochastycznej oszacowanych modeli. Podane po literze oznaczającej rodzaj modelu skróty oznaczają kolejno: t – trend liniowy, s [z] – stałe [zmienne] parametry przy zmiennej objaśniającej, p [z] – periodyczny [zmienny] składnik sezonowy. Dla modeli przyczynowo-opisowych oraz hierarchicznych modeli przyczynowo-opisowych w nawiasach podano nazwy zmiennych objaśniających. Pogrubioną czcionką zaznaczono modele, w których wszystkie parametry sezonowe są nieistotne.

W procesie modelowania predyktywnego zmiennej *K*1 wykorzystano również sztuczne sieci neuronowe. Zbudowano, nauczono i przetestowano kilkaset sztucznych sieci neuronowych o różnych architekturach (sieci liniowe, perceptrony wielowarstwowe (*MLP*), sieci o radialnych funkcjach bazowych (*RBF*)), w których wzorcową zmienną wyjściową była zmienna *K*1, natomiast pojedynczą zmienną wejściową – zmienna *K*1, *SP*1 lub *PR*1. Każda ze skonstruowanych sieci miała 12 neuronów wejściowych (ze względu na okres wahań zmiennej *K*1) i pojedynczy neuron wyjściowy. Sieci uczone były z nauczycielem, w zależności od rodzaju sieci i neuronów w nich występujących zastosowano następujące algorytmy uczenia:

– wstecznej propagacji (*BP*),
– gradientów sprzężonych (*CG*),
– Newtona (*N*),
– Levenberga-Marquardta (*LM*),
– *k*-średnich (*KM*),
– *k*-najbliższych sąsiadów (*KN*),
– pseudoinwersji (*PI*).

Dla każdego skonstruowanego modelu neuronowego wyznaczono wartości następujących mierników jakości sieci:

– błąd RMSE pozwalający określić zdolności sieci do aproksymacji i generalizacji,
– iloraz odchyleń standardowych obliczonych dla błędów popełnianych przez sieć (będących różnicami pomiędzy wartościami wyjściowymi wyznaczonymi przez sieć i odpowiadającymi im wzorcowymi wartościami wyjściowymi) i dla wzorcowych wartości zmiennej wyjściowej,
– współczynnik korelacji liniowej Pearsona (ρ) pomiędzy wartościami wyjściowymi wyznaczonymi przez sieć i odpowiadającymi im wzorcowymi wartościami wyjściowymi.

Tabela 2. Oceny parametrów struktury stochastycznej oszacowanych modeli

Model	R^2 [%]	Se	Model	R^2 [%]	Se	Model	R^2 [%]	Se
MKtp	86,35	134,65	PH26tsp(SP1)	91,09	108,75	**PH26tzp(SP1)**	90,38	113,02
MKtz	90,77	110,73	PH34tsp(SP1)	91,50	106,24	**PH34tzp(SP1)**	91,46	106,49
H26tp	78,80	167,80	**PH43tsp(SP1)**	91,22	107,98	PH43tzp(SP1)	91,15	108,44
H34tp	80,10	162,57	**PH62tsp(SP1)**	91,28	107,63	**PH62tzp(SP1)**	91,00	109,33
H43tp	82,70	151,59	PH223tsp(SP1)	91,43	106,69	PH223tzp(SP1)	91,26	107,71
H62tp	86,07	136,01	PH232tsp(SP1)	91,45	106,54	PH232tzp(SP1)	91,22	108,01
H223tp	78,59	168,63	PH322tsp(SP1)	91,57	105,78	**PH322tzp(SP1)**	91,74	104,74
H232tp	79,80	163,79	**PH26tsp(PR1)**	90,63	111,55	PH26tzp(PR1)	89,55	117,80
H322tp	80,19	162,20	**PH34tsp(PR1)**	91,32	107,38	**PH34tzp(PR1)**	90,57	111,90
H26tz	77,61	172,44	**PH43tsp(PR1)**	90,98	109,45	**PH43tzp(PR1)**	89,90	115,79
H34tz	80,43	161,23	**PH62tsp(PR1)**	90,99	109,38	**PH62tzp(PR1)**	89,90	115,83
H43tz	84,91	141,58	**PH223tsp(PR1)**	90,81	110,48	**PH223tzp(PR1)**	90,10	114,66
H62tz	88,74	122,29	**PH232tsp(PR1)**	91,03	109,12	**PH232tzp(PR1)**	90,53	112,14
H223tz	77,45	173,07	**PH322tsp(PR1)**	91,38	106,96	**PH322tzp(PR1)**	90,67	111,31
H232tz	79,88	163,46	**PH26tsz(SP1)**	90,87	110,09	**PH26tzz(SP1)**	91,27	107,70
H322tz	81,13	158,29	**PH34tsz(SP1)**	92,05	102,75	**PH34tzz(SP1)**	92,73	98,27
Ps(SP1)	84,46	143,64	**PH43tsz(SP1)**	91,00	109,30	PH43tzz(SP1)	91,22	107,95
Pts(SP1)	91,03	109,15	PH62tsz(SP1)	92,15	102,13	PH62tzz(SP1)	92,19	101,87
Ptsp(SP1)	91,37	107,06	**PH223tsz(SP1)**	90,98	109,42	**PH223tzz(SP1)**	91,05	109,02
Ptsz(SP1)	93,26	94,59	**PH232tsz(SP1)**	91,51	106,20	PH232tzz(SP1)	92,04	102,79
Ptzp(SP1)	91,01	109,25	**PH322tsz(SP1)**	92,38	100,58	PH322tzz(SP1)	92,92	96,94
Ptzz(SP1)	91,62	105,47	**PH26tsz(PR1)**	90,98	109,42	PH26tzz(PR1)	89,69	116,99
Ps(PR1)	36,81	289,70	**PH34tsz(PR1)**	92,84	97,51	**PH34tzz(PR1)**	92,16	102,04
Pts(PR1)	88,97	121,02	PH43tsz(PR1)	92,86	97,37	PH43tzz(PR1)	92,69	98,55
Ptsp(PR1)	90,41	112,87	PH62tsz(PR1)	92,80	97,77	PH62tzz(PR1)	93,24	94,74
Ptsz(PR1)	94,51	85,39	**PH223tsz(PR1)**	91,07	108,89	**PH223tzz(PR1)**	90,18	114,17
Ptzp(PR1)	86,97	131,54	**PH232tsz(PR1)**	91,38	107,01	**PH232tzz(PR1)**	90,89	109,96
Ptzz(PR1)	92,93	96,87	**PH322tsz(PR1)**	92,95	96,77	**PH322tzz(PR1)**	92,31	101,02

Źródło: obliczenia własne.

W tabeli 3 przedstawiono charakterystyki wybranych modeli sztucznych sieci neuronowych (po 10 najlepszych modeli dla każdej zmiennej wejściowej) różniących się rodzajem, liczbą warstw oraz liczbą neuronów ukrytych (LNU). Pogrubioną czcionką zaznaczono wartości ilorazów odchyleń standardowych równe lub większe od 0,7 (przyjmuje się, że takie wartości dyskwalifikują model neuronowy). Pogrubione wartości błędów RMSE oznaczają, iż różniły się one znacząco na zbiorach uczącym, walidacyjnym i testowym, co oznacza, że sieci nie nabyły zdolności douogólniania, a jedynie dopasowały się do danych uczących. Przy algorytmach uczenia perceptronów wielowarstwowych w nawiasach podano liczbę wykonanych przez nie epok.

Tabela 3. Charakterystyki wybranych modeli sztucznych sieci neuronowych

Model	Zmienna wejściowa	Rodzaj sieci	LNU	RMSE	Iloraz odchyleń standardowych	ρ	Algorytm uczenia
SSN1	*K*1	liniowa	–	**249,03**	0,6711	0,8061	PI
SSN2	*K*1	MLP	1	**144,58**	0,4006	0,9172	BP(50),CG(26)
SSN3	*K*1	MLP	3	**163,15**	0,3794	0,9252	BP(50),CG(6)
SSN4	*K*1	MLP	4	**185,52**	0,4327	0,9191	BP(50),CG(8)
SSN5	*K*1	MLP	8	138,56	0,3956	0,9052	BP(50),CG(26)
SSN6	*K*1	MLP	13	**122,43**	0,4002	0,8801	BP(50),CG(32)
SSN7	*K*1	RBF	1	356,72	**0,9973**	0,0393	KM,KN,PI
SSN8	*K*1	RBF	2	**219,88**	0,5882	0,7627	KM,KN,PI
SSN9	*K*1	RBF	4	**159,94**	0,4243	0,8810	KM,KN,PI
SSN10	*K*1	RBF	8	**118,14**	0,3253	0,9362	KM,KN,PI
SSN11	*SP*1	liniowa	–	**153,35**	0,4123	0,9043	PI
SSN12	*SP*1	MLP	1	**144,58**	0,4071	0,8704	BP(50),CG(10)
SSN13	*SP*1	MLP	3	**141,31**	0,4135	0,8968	BP(45)
SSN14	*SP*1	MLP	8	**97,03**	0,3905	0,9044	BP(38)
SSN15	*SP*1	MLP	13	146,91	0,3899	0,9339	BP(50),CG(3)
SSN16	*SP*1	MLP	20	**97,00**	0,3903	0,9046	BP(50),CG(3)
SSN17	*SP*1	RBF	1	**423,43**	**1,1760**	0,1542	KM,KN,PI
SSN18	*SP*1	RBF	2	**167,93**	0,4236	0,8218	KM,KN,PI
SSN19	*SP*1	RBF	4	**166,34**	0,3768	0,8534	KM,KN,PI
SSN20	*SP*1	RBF	8	**491,70**	**1,4593**	0,1420	KM,KN,PI
SSN21	*PR*1	liniowa	–	**367,29**	**1,1970**	0,4611	PI
SSN22	*PR*1	MLP	2	**201,53**	0,5344	0,8785	BP(50),CG(62)
SSN23	*PR*1	MLP	4	**157,01**	0,4641	0,8903	BP(49)
SSN24	*PR*1	MLP	8	**147,22**	0,4469	0,9005	BP(50),CG(16)
SSN25	*PR*1	MLP	13	**181,31**	0,4905	0,6715	BP(50),CG(28)
SSN26	*PR*1	MLP	20	**282,65**	**0,7490**	0,6627	BP(50),CG(11)
SSN27	*PR*1	RBF	1	**320,56**	**1,0415**	0,1982	KM,KN,PI
SSN28	*PR*1	RBF	2	**280,86**	**0,7348**	0,6784	KM,KN,PI
SSN29	*PR*1	RBF	4	**167,56**	0,4898	0,8302	KM,KN,PI
SSN30	*PR*1	RBF	8	**128,96**	0,3664	0,9317	KM,KN,PI

Źródło: obliczenia własne.

Na podstawie danych zawartych w tab. 2 i 3 dokonano weryfikacji oszacowanych modeli: w przypadku modeli ekonometrycznych – oceny ich dopasowania do danych empirycznych oraz istotności parametrów strukturalnych, a w przypadku modeli neuronowych – oceny jakości sieci. Do dalszych badań wybrano pięć modeli o najlepszych własnościach predyktywnych: MKtp, H62tp, Ptsz(PR1), PH62tzz(PR1) oraz SSN15, na ich podstawie wyznaczono prognozy *ex post* na 12 kolejnych okresów (t = 49, 50, ..., 60) – oznaczono je odpowiednio: $f_{1t}, f_{2t}, f_{3t}, f_{4t}, f_{5t}$. Prognozy te stanowiły prognozy składowe prognoz kombinowanych określonych wzorem (1) dla m = 2, 3, 4, 5 wyznaczonych dla tych samych okresów. Średnie absolutne błędy procentowe prognoz składowych zestawiono w tab. 4.

Tabela 4. Średnie absolutne błędy procentowe prognoz składowych

Prognoza składowa	MAPE [%]
f_{1t}	16,37
f_{2t}	15,92
f_{3t}	13,66
f_{4t}	12,72
f_{5t}	10,90

Źródło: opracowanie własne.

Dla każdej wartości m (liczby prognoz składowych) wyznaczono po trzy prognozy kombinowane z wagami otrzymanymi na podstawie trzech metod: sztucznych sieci neuronowych (SSN), wariancji-kowariancji (VC) oraz Batesa i Grangera (BG). Do oszacowania wartości wag wykorzystano reszty modeli indywidualnych, na podstawie których wyznaczono prognozy składowe. Wartości wag w metodach VC i BG otrzymano, wykonując obliczenia na podstawie wzorów (19)-(20). W przypadku metody SSN najpierw skonstruowano cztery różne sieci odpowiadające czterem wartościom m, a następnie przeprowadzono proces ich uczenia. Każda z utworzonych sieci była osobno uczona dla każdej możliwej kombinacji m prognoz składowych. Sieci uczono z nauczycielem, co oznacza, że przedstawiano im wzorcowe, znane wartości wejściowe i wyjściowe określone wzorami (11) i (17) (tzw. próbki uczące). Proces uczenia rozpoczynał się inicjalizacją wag, następnie prezentowano sieci próbki uczące. Sieć obliczała błędy wartości wyjściowych jako ich odchylenia od wzorcowych danych wyjściowych i na podstawie metody PI modyfikowała początkowe wartości wag w taki sposób, aby te błędy zmniejszyć.

W tabeli 5 przedstawiono charakterystyki oszacowanych modeli sztucznych sieci neuronowych wykorzystanych do wyznaczenia wag prognoz kombinowanych.

Analizując zawarte w tab. 5 informacje, można zauważyć, że w kilku przypadkach uzyskano sieci o złej jakości – o wartościach ilorazów odchyleń standardowych większych od 0,7 i jednocześnie małych wartościach współczynników korelacji oraz dużych różnicach pomiędzy wartościami błędów RMSE wyznaczonymi na zbiorach uczącym (U), walidacyjnym (W) i testowym (T). Można oczekiwać, że prognozy kombinowane wyznaczone na podstawie tych modeli będą obarczone dużymi błędami.

Tabela 5. Charakterystyki oszacowanych modeli sztucznych sieci neuronowych wykorzystanych do wyznaczenia wag prognoz kombinowanych

Prognozy składowe modelu	RMSE			Iloraz odchyleń standardowych			ρ		
	U	W	T	U	W	T	U	W	T
$f_{1t}-f_{2t}$	104,09	139,38	147,45	**0,7512**	**0,7276**	**0,7547**	0,6012	0,7134	0,5473
$f_{1t}-f_{3t}$	53,02	80,94	55,03	0,4356	**0,7823**	0,3672	0,7832	0,4679	0,8401
$f_{1t}-f_{4t}$	45,96	59,47	39,81	0,3802	0,4958	0,4031	0,9020	0,8831	0,9103
$f_{1t}-f_{5t}$	74,05	70,30	55,08	0,4487	0,6574	0,4953	0,8732	0,7640	0,8362
$f_{2t}-f_{3t}$	55,03	72,98	55,92	0,3001	0,4092	0,3972	0,9402	0,9183	0,8994
$f_{2t}-f_{4t}$	76,03	79,43	60,38	0,1998	0,2765	0,3000	0,9534	0,9039	0,9128
$f_{2t}-f_{5t}$	80,02	77,94	60,93	0,3001	0,4492	0,3999	0,9021	0,8534	0,8992
$f_{3t}-f_{4t}$	39,03	54,03	84,99	0,6502	**0,7209**	**0,8732**	0,4793	0,4002	0,3702
$f_{3t}-f_{5t}$	60,03	75,94	92,09	0,5994	**0,7027**	**0,7103**	0,9592	0,7032	0,6092
$f_{4t}-f_{5t}$	66,09	98,94	55,02	0,5802	0,5283	0,3672	0,8192	0,8903	0,8002
$f_{1t}-f_{2t}-f_{3t}$	44,03	67,38	55,36	0,4023	0,4100	0,3904	0,8902	0,9023	0,9201
$f_{1t}-f_{2t}-f_{4t}$	85,02	55,93	75,02	0,5729	**0,7012**	0,5021	0,6610	0,7892	0,8134
$f_{1t}-f_{2t}-f_{5t}$	54,09	69,02	60,02	0,6701	0,6821	**0,8231**	0,7342	0,7453	0,4738
$f_{1t}-f_{3t}-f_{4t}$	46,98	63,12	84,99	**0,8675**	**0,9023**	**0,7754**	0,4900	0,4423	0,3543
$f_{1t}-f_{3t}-f_{5t}$	70,87	69,09	60,99	0,3684	0,4091	0,2768	0,9321	0,8324	0,9671
$f_{1t}-f_{4t}-f_{5t}$	64,09	68,09	80,74	0,5564	0,4451	0,5023	0,8361	0,9123	0,9400
$f_{2t}-f_{3t}-f_{4t}$	58,09	52,81	84,02	0,5391	0,4902	**0,8100**	0,7721	0,8109	0,6756
$f_{2t}-f_{3t}-f_{5t}$	60,02	70,38	82,71	0,3782	0,3401	0,4291	0,9231	0,9501	0,8531
$f_{2t}-f_{4t}-f_{5t}$	70,01	66,34	77,03	0,4502	0,4092	0,6501	0,8901	0,9203	0,7777
$f_{3t}-f_{4t}-f_{5t}$	46,02	57,02	99,03	0,2601	0,4602	0,5591	0,9712	0,8802	0,8321
$f_{1t}-f_{2t}-f_{3t}-f_{4t}$	45,09	65,05	70,04	0,5493	0,5902	0,6593	0,7759	0,7032	0,6035
$f_{1t}-f_{2t}-f_{3t}-f_{5t}$	32,84	57,83	97,92	0,4802	0,6209	**0,7578**	0,8008	0,7207	0,5515
$f_{1t}-f_{2t}-f_{4t}-f_{5t}$	57,98	90,78	91,76	0,5576	**0,7342**	0,6003	0,7004	0,5743	0,7245
$f_{1t}-f_{3t}-f_{4t}-f_{5t}$	53,98	60,87	71,87	0,4342	0,5123	0,4130	0,9087	0,9123	0,9131
$f_{2t}-f_{3t}-f_{4t}-f_{5t}$	48,05	45,98	97,57	0,4676	0,4134	**0,7285**	0,8878	0,9154	0,6574
$f_{1t}-f_{2t}-f_{3t}-f_{4t}-f_{5t}$	57,09	45,03	62,02	0,4998	0,5201	0,1274	0,8698	0,8554	0,9806

Źródło: obliczenia własne.

Po oszacowaniu wartości wag wyznaczono prognozy kombinowane. Średnie absolutne błędy procentowe otrzymanych prognoz kombinowanych przedstawiono w tab. 6. Pogrubionym drukiem zaznaczono błędy tych prognoz kombinowanych, które są równe lub większe od jednego (najmniejszego) z błędów prognoz składowych, dodatkowym podkreśleniem – równe dwa lub większe od dwóch błędów prognoz składowych, natomiast kursywą – błędy prognoz kombinowanych, w których jedna wartość parametru λ_i jest ujemna[2].

Tabela 6. Średnie absolutne błędy procentowe prognoz kombinowanych

Prognozy składowe	MAPE [%]			Prognozy składowe	MAPE [%]		
	SSN	VC	BG		SSN	VC	BG
f_{1t}-f_{2t}	16,16	16,37	16,16	f_{1t}-f_{3t}-f_{4t}	**13,13**	***13,40***	13,00
f_{1t}-f_{3t}	**13,78**	13,58	**13,66**	f_{1t}-f_{3t}-f_{5t}	9,79	*10,29*	**12,36**
f_{1t}-f_{4t}	12,51	12,27	12,33	f_{1t}-f_{4t}-f_{5t}	10,84	10,89	**10,95**
f_{1t}-f_{5t}	10,07	9,15	10,59	f_{2t}-f_{3t}-f_{4t}	12,71	***13,43***	**12,89**
f_{2t}-f_{3t}	13,63	**13,66**	**13,66**	f_{2t}-f_{3t}-f_{5t}	*10,80*	*10,85*	**12,19**
f_{2t}-f_{4t}	12,47	12,57	**12,72**	f_{2t}-f_{4t}-f_{5t}	*10,60*	***11,75***	10,89
f_{2t}-f_{5t}	9,52	7,28	9,91	f_{3t}-f_{4t}-f_{5t}	11,67	12,66	12,15
f_{3t}-f_{4t}	**12,98**	12,71	12,31	f_{1t}-f_{2t}-f_{3t}-f_{4t}	***13,13***	***13,43***	**13,14**
f_{3t}-f_{5t}	**12,25**	*7,84*	13,27	f_{1t}-f_{2t}-f_{3t}-f_{5t}	***12,56***	***13,30***	**12,63**
f_{4t}-f_{5t}	10,72	9,99	**11,68**	f_{1t}-f_{2t}-f_{4t}-f_{5t}	*10,48*	*10,81*	**11,46**
f_{1t}-f_{2t}-f_{3t}	13,62	**13,66**	**13,86**	f_{1t}-f_{3t}-f_{4t}-f_{5t}	10,75	*10,82*	**12,11**
f_{1t}-f_{2t}-f_{4t}	*12,49*	*12,26*	**13,12**	f_{2t}-f_{3t}-f_{4t}-f_{5t}	**11,63**	***13,14***	**11,99**
f_{1t}-f_{2t}-f_{5t}	*9,80*	*10,74*	**11,97**	f_{1t}-f_{2t}-f_{3t}-f_{4t}-f_{5t}	*6,75*	***13,12***	**12,32**

Źródło: opracowanie własne.

Analizując informacje zawarte w tab. 6, można zauważyć, że dla większości prognoz kombinowanych wyznaczonych z wykorzystaniem SSN ich błędy są mniejsze od błędów prognoz składowych. Porównując dane przedstawione w tab. 5 i 6, można zauważyć, że w przypadkach, w których prognozy kombinowane obarczone są większymi błędami niż ich prognozy składowe, modele SSN wykorzystane do oszacowania wag mają złą jakość.

Wśród metod ekonometrycznych zdecydowanie lepsze wyniki (mniejsze błędy prognoz kombinowanych niż ich prognoz składowych) uzyskano, stosując metodę wariancji-kowariancji, jednak w wielu przypadkach wymagało to wykorzystania do ich budowy ujemnej wartości jednego z parametrów λ_i.

[2] Wagi otrzymane metodą VC mogą być ujemne, gdy współczynnik korelacji pomiędzy błędami prognoz składowych $\rho \to 1$ (prognozy są wówczas jednocześnie przeszacowane lub niedoszacowane). Choć takie wartości mogą być kłopotliwe w interpretacji, jednak umożliwiają wyznaczenie prognozy kombinowanej o bardzo małym błędzie, stąd też wielu autorów proponuje, aby nie wprowadzać ograniczenia nieujemności wag (zob. [Grajek 2002; Kin Chan, Kingsman, Wong 1999; Aksu, Gunter 1992]). W niniejszym artykule również zastosowano takie podejście.

Porównując dokładność prognoz kombinowanych wyznaczonych wszystkimi trzema metodami, można zauważyć, że najlepsze wyniki uzyskano dla modeli sztucznych sieci neuronowych. Ponadto w wielu przypadkach, kiedy jedna z wag wyznaczona metodą VC była ujemna, zastosowanie SSN prowadziło do wyznaczenia wszystkich wag nieujemnych.

Badania empiryczne potwierdziły przydatność sztucznych sieci neuronowych do wyznaczania liniowych prognoz kombinowanych pod warunkiem otrzymania modelu neuronowego o zadowalającej jakości – w większości rozważanych przypadków błędy prognoz kombinowanych z wagami otrzymanymi za pomocą SSN były mniejsze od błędów ich prognoz składowych oraz od błędów prognoz kombinowanych z wagami otrzymanymi metodami ekonometrycznymi.

Literatura

Aksu C., Gunter S., *An empirical analysis of the accuracy of SA, OLS, ERLS and NRLS combination forecast*, „International Journal of Forecasting” 1992, 8, s. 27-43.

Azoff E.M., *Neural Network Time Series Forecasting of Financial Markets*, John Wiley & Sons, Chichester 1994.

Bates J.M., Granger C.W.J, *The combination of forecasts*, „Operational Research Quarterly” 1969, 40, s. 451-468.

Da Xu L., Liu B., Ming Shi S., *Improving the accuracy of nonlinear combined forecasting using neural networks*, „Expert Systems with Applications” 1999, 16, s. 49-54.

Kin Chan Ch., Kingsman B., Wong H., *A comparison of unconstrained and constrained OLS for the combination of demand forecasts*, „Annals of Operations Research” 1999, 87, s. 129-140.

Grajek M., *Prognozy łączone*, „Przegląd Statystyczny” 2002, R. XLIX, z. 2, s. 69-81.

Granger C.W.J., Newbold P., *Forecasting univariate time series and the combination of forecasts*, „Journal of Royal Statistical Society” Ser. A 1974, 137, s. 131-165.

Lula P., Tadeusiewicz R., *Wprowadzenie do sieci neuronowych*, StatSoft, Kraków 2001.

Zawadzki J. (red.), *Zastosowanie hierarchicznych modeli szeregów czasowych w prognozowaniu zmiennych ekonomicznych z wahaniami sezonowymi*, Wydawnictwo Akademii Rolniczej w Szczecinie, Szczecin 2003.

THE APPLICATION OF ARTIFICIAL NEURAL NETWORK TO BUILDING COMBINED FORECASTS

Summary: In the article, the author presents the method of building linear combined forecasts leaning on artificial neural network. The illustration of theoretical considerations is the empirical example. Individual and combined forecasts were marked for economic variable having the form of time series with seasonal fluctuations. The paper examines the *ex post* efficiency of combined forecasts building with artificial neural network in comparison with their component forecasts and with econometric combined forecasts.