

Tomasz CIEŚLIK^{1,2}
Klaudia METELSKA²

SKUTECZNOŚĆ PROGNOZOWANIA ZUŻYCIA GAZU Z WYKORZYSTANIEM METOD REGRESJI I SZTUCZNYCH SIECI NEURONOWYCH

Na podstawie zgromadzonych danych takich jak: temperatura, siła wiatru oraz zużycie gazu w ciągu dnia na przestrzeni dwóch lat określono wpływ czynników atmosferycznych na konsumpcje gazu za pomocą regresji wielorakiej, funkcji potęgowej oraz funkcji użytkownika. Wyznaczono wpływ miesiąca oraz dnia (parametr sztuczny) na konsumpcje gazu. Zbudowano modele regresji liniowe, potęgowej oraz sztuczne sieci neuronowe służące do określania zużycia gazu. Starano się wyznaczyć jak najlepszy model regresji i porównywano go do modeli sieci neuronowych za pomocą MAPE (średni absolutny błąd procentowy).

Słowa kluczowe: regresja liniowa, regresja potęgowa, parametry sztuczne, sztuczne sieci neuronowe, konsumpcja gazu

1 Wstęp

Prognozowanie pracy sieci gazowej, ilości magazynowanego i konsumowanego gazu jest istotna w w procesie zrównoważonego rozwoju. Prognozowanie zapewnia odpowiednie planowanie modernizacji sieci gazowych ponieważ można przewidzieć w jakim okresie będzie największy pobór czyli obciążenie sieci i tłoczni. Podejmowanie decyzji związane z rozwojem systemu dystrybucji staje się łatwiejsze i szybsze ze względu na posiadane informację o wielkości zużycia w przyszłości. Dzięki informacją uzyskanym podczas analiz można w łatwy sposób przewidzieć ilość potrzebnego gazu oraz zarządzać system magazynów gazu co jest istotne z punktu widzenia bezpieczeństwa energetycznego kraju.

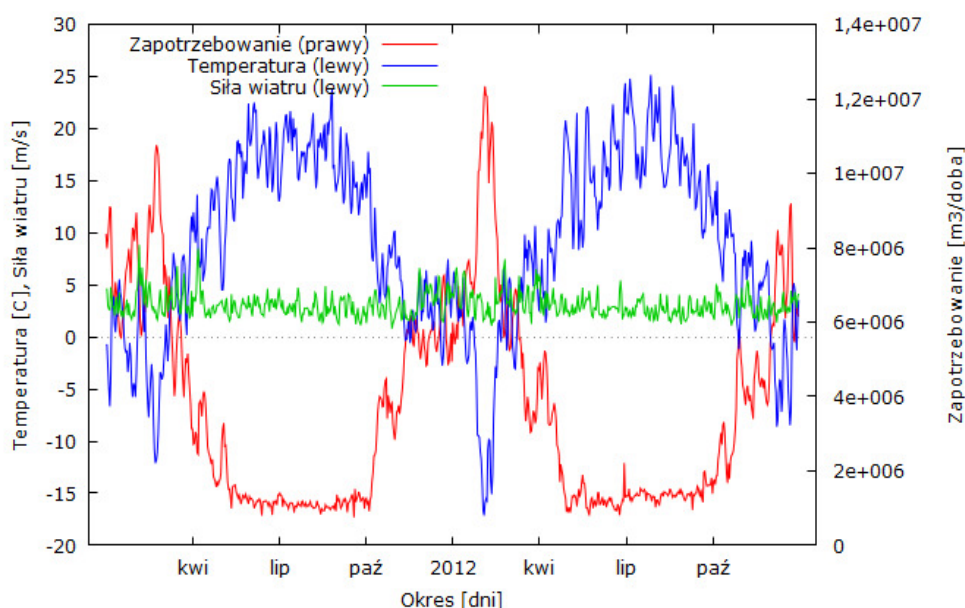
¹ Instytut Fizyki Jądrowej PAN Aleja Radzikowskiego 152, 31-342 Kraków email: tomasz.cieslik@ifj.edu.pl

² Akademia Górniczo- Hutnicza, Wydział Wiertnictwa, Nafty i Gazu Aleja Adama Mickiewicza 30, 30-059 Kraków

2 Prognozowanie zużycia gazu

Celem pracy jest prognozowanie zużycia gazu dla 1 miesiąca na podstawie zgromadzonych danych z 23 miesięcy takich jak: pobór gazu, temperatura otoczenia i siła wiatru. Do prognozowania zużycia gazu wykorzystane zostaną zmienne sztuczne takie jak dzień tygodnia oraz miesiąc. Obliczenia zostały wykonane na programach: Statistica i Gretl..

Można zauważyć iż konsumpcja jest silnie zależna od temperatury. W zimie kiedy spada temperatura to następuje wzrost konsumpcji gazu. (Rysunek 1).



Rysunek 1 Zużycie gazu w przeciągu dwóch lat w zależności od temperatury i siły wiatru.
Figure 1 Gas consumption within two years depending on the temperature and wind strength.

2.1 Metody regresji

Regresja to metoda statystyczna polegająca na badaniu związków pomiędzy danymi i przewidywaniu na tej podstawie nieznanymi wartościami przy pomocy innych wartości[1].

Założenia[1-4]:

– Postać modelu jest liniowa- $Y=\beta x+\varepsilon$

gdzie:

β - parametr

ε - zdarzenie losowe

– Macierz x jest znana i nie jest losowa

Rząd $x=k$ k- liczba parametrów

$T > k$

gdzie:

k- liczba parametrów

T- liczba obserwacji

Gdy rząd macierzy X nie jest k, to wówczas $\det(x^T x) = 0$ i nie istnieje macierz odwrotna, czyli nie da się oszacować parametrów.

– Wartość oczekiwana składnika losowego równa się zero

– Wariancja składnika losowego jest stała i równa się σ^2

2.1.1 Regresja liniowa

Model liniowy ma następującą wzór ogólny(1)[1-4]:

Postać:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_k * X_k + \varepsilon \quad (1)$$

gdzie:

Y- zmienna badana (której wartość jest objaśniana przez model - endogeniczna)

$X_1 \dots X_k$ - zmienne za pomocą których badamy Y

$\beta_0, \beta_1, \beta_2 \dots \beta_k$ - parametry modelu (liczby)

ε - składnik losowy

W modelu liniowym regresji badano wpływ temperatury i siły wiatru na konsumpcje gazu (Tabela 1) a następnie wykorzystano parametry sztuczne (Tabela 2) takie jak miesiąc i dzień w celu poprawy jakości modelu.

Tabela 1 Podsumowanie regresji liniowej bez parametrów sztucznych.

Table 1 Summary linear regression without artificial parameters.

Podsumowanie regresji zmiennej zależnej: Zapotrzebowanie m ³ R= ,945 R ² = ,892 Popraw. R ² = ,892 F(2,697)=2874,3 p<0,0000 Błąd std. estymacji: 8795E2						
N=700	b*	Bł. std.	b	Bł. std.	t(697)	p
W. wolny			6048820	106724,9	56,6768	0,000000
Temperatura [st. C]	-0,937984	0,012603	-296427	3982,7	-74,4283	0,000000
Siła wiatru [m/s]	0,037325	0,012603	85557	28887,4	2,9617	0,003163

Ocena szacowania parametrów dla regresji liniowej bez parametrów sztucznych:

- Standardowy błąd estymacji = 879500
- Rzeczywiste wartości Y różnią się od teoretycznych przeciętnie o 879500 co stanowi 23 % średniego Y.
- $R^2 = 0,89$ – wokoło 89,18% całkowitej zmienność Y została wyjaśniona przez oszacowany model.

Model przyjmuje postać:

$$Y = 6048820 - 296427 * X_1 + 85557 * X_2 \quad (2)$$

gdzie:

X_1 - temperatura [C]

X_2 - siła wiatru [m/s]

Tabela 2 Podsumowanie regresji liniowej bez parametrów sztucznych.

Table 2 Summary linear regression without artificial parameters.

N=700	Podsumowanie regresji zmiennej zależnej: Zapotrzebowanie m ³ R= ,969 R ² = ,939 Popraw. R ² = ,939 F(8,691)=1336,1 p<0,0000 Błąd std. estymacji: 6619E2					
	b*	Bł. std.	b	Bł. std.	t(691)	p
W. wolny			5340757	78950,8	67,6467	0,000000
Temperatura [st. C]	-0,778101	0,016764	-245900	5298,0	-46,4138	0,000000
Styczeń	0,192992	0,012420	1812827	116665,4	15,5387	0,000000
Luty	0,218133	0,013633	2128631	133040,8	15,9998	0,000000
Marzec	0,091487	0,011082	859359	104094,1	8,2556	0,000000
Lipiec	0,057332	0,010480	538535	98445,0	5,4704	0,000000
Sierpień	0,044398	0,010434	417045	98005,9	4,2553	0,000024
Październik	-0,036217	0,010204	-340199	95849,4	-3,5493	0,000412
Listopad	0,037033	0,010957	353054	104464,1	3,3797	0,000767

Ocena szacowania parametrów dla regresji liniowej parametrów sztucznych:

- Standardowy błąd estymacji = 661900
- Rzeczywiste wartości Y różnią się od teoretycznych przeciętnie o 661900 co stanowi 18 % średniego Y.
- $R^2 = 0,93$ – wokoło 93,92% całkowitej zmienności Y została wyjaśniona przez oszacowany model.

Model przyjmuje postać:

$$Y = 5340757 - 245900 * X_1 + M * X_2 \quad (3)$$

gdzie:

X_1 - temperatura [C]

MX_2 - dany miesiąc dla którego jest wykonywana prognoza.

2.1.2 Regresja potęgowa

Model potęgowy ma następująca wzór ogólny[1-4]:

$$Y = \beta_0 * X_1^{\beta_1} * X_2^{\beta_2} * \dots * X_k^{\beta_k} \quad (4)$$

gdzie:

Y- zmienna badana (której wartość jest objaśniana przez model - endogeniczna)

$X_1 \dots X_k$ - zmienne za pomocą których badamy Y

$\beta_0, \beta_1, \beta_2 \dots \beta_k$ - parametry modelu (liczby)

ϵ - składnik losowy

Sprowadzamy do postaci liniowej przez logarytmowanie $\ln(4)$

$$\ln(Y) = \ln(\beta_0 * X_1^{\beta_1} * X_2^{\beta_2} * \dots * X_k^{\beta_k})$$

$$\ln(Y) = \ln \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \dots + \beta_k \ln X_k$$

$$\ln(Y) = Y'$$

$$\ln X_k = X'_k$$

$$Y' = \beta_0 + \beta_1 * X'_1 + \beta_2 * X'_2 + \dots + \beta_k * X'_k \quad (5)$$

W modelu potęgowym badano wpływ temperatury i siły wiatru na konsumpcję gazu (Tabela 3) a następnie wykorzystano parametry sztuczne (Tabela 4) takie jak miesiąc i dzień w celu poprawy jakości modelu.

Tabela 3 Podsumowanie regresji potęgowej bez parametrów sztucznych
Table 3 Summary of the power regression without artificial parameters.

N=700	Podsumowanie regresji zmiennej zależnej: Ln (Zaopatrzenie[m3]) R= ,939 R ² = ,8808 Popraw. R ² = ,8805 F(2,697)=2574,7 p<0,0000 Błąd std. estymacji: ,26750					
	b*	Bł. std.	b	Bł. std.	t(697)	p
W. wolny			148,2051	1,909542	77,6129	0,000000
ln(Temp. [st. K])	-0,923599	0,013173	-23,6798	0,337729	-70,1150	0,000000
ln(wiatr [m/s])	0,089524	0,013173	0,1896	0,027893	6,7962	0,000000

Ocena szacowania parametrów dla regresji potęgowej bez parametrów sztucznych

– Standardowy błąd estymacji = 0,26750.

– Rzeczywiste zlogarytmowane Y różnią się od teoretycznych zlogarytmowanych o 0,24 , co stanowi 1,8 % średniego zlogarytmowanego Y.

– R² = 0,88 – wokoło 88% całkowitej zmienności zlogarytmowanego Y została wyjaśniona przez oszacowany model.

Model przyjmuje postać:

$$\ln(Y) = 148,2051 - 23,6789 \ln X_1 + 0,1896 \ln X_2 \quad (6)$$

gdzie:

X₁- temperatura [K]

X₂- siła wiatru [m/s]

Ocena szacowania parametrów dla regresji potęgowej z zmiennymi sztucznymi (Tabela 4):

– Standardowy błąd estymacji = 0,17812;

– Rzeczywiste zlogarytmowane Y różnią się od teoretycznych zlogarytmowanych o 0,17 co stanowi 1,2 % średniego zlogarytmowanego Y;

– R² = 0,94 – wokoło 94% całkowitej zmienności zlogarytmowanego Y została wyjaśniona przez oszacowany model.

Model przyjmuje postać:

$$\ln(Y) = 87,3700 - 12,9399\ln X_1 + MX_2 \quad (7)$$

gdzie:

X_1 - temperatura [K]

MX_2 - dany miesiąc dla którego jest wykonywana prognoz

Tabela 4 Podsumowanie regresji potęgowej z parametrami sztucznymi.

Table 4 Summary of the power regression with artificial parameters.

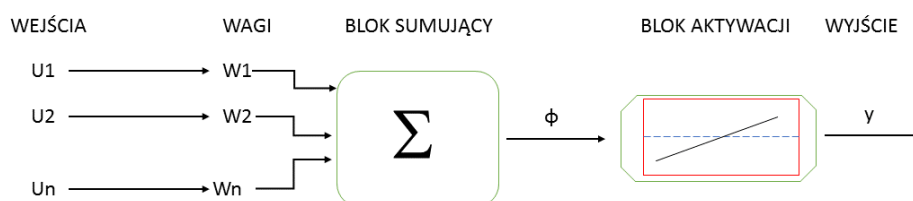
N=700	Podsumowanie regresji zmiennej zależnej: Ln(Zaopatrzenie[m3]) R= ,974 R ² = ,948 Popraw. R ² = ,947 F(9,690)=1388,4 p<0,0000 Błąd std. estymacji: ,17812					
	b*	Bł. std.	b	Bł. std.	t(690)	p
W. wolny			87,3700	2,670057	32,7221	0,000000
ln(Temp. [stp. K])	-0,504704	0,018375	-12,9399	0,471120	-27,4663	0,000000
Styczeń	0,359376	0,013902	0,9778	0,037825	25,8510	0,000000
Luty	0,330110	0,015531	0,9331	0,043901	21,2547	0,000000
Marzec	0,307551	0,011951	0,8368	0,032518	25,7340	0,000000
Kwiecień	0,190109	0,010059	0,5250	0,027777	18,9002	0,000000
Lipiec	0,037802	0,009436	0,1029	0,025673	4,0063	0,000068
Październik	0,147741	0,010433	0,4020	0,028386	14,1612	0,000000
Listopad	0,271417	0,011741	0,7495	0,032423	23,1171	0,000000
Grudzień	0,238392	0,010833	0,8958	0,040708	22,0055	0,000000

2.2 Sztuczne sieci neuronowe

W skład sztucznej sieci neuronowej wchodzi sztuczny neuron czyli podstawowy element budujący sztuczną sieć neuronową (Rysunek 2). Jest on “przetwornikiem” który na wejściu przyjmuje sygnały wejściowe i mnoży je poprzez odpowiednie wartości wag w zależności od istotności danego sygnału na wartość wyjściową. Sygnały po przemnożeniu przez wagi są sumowane i korygowane w bloku sumującym. Służy to do wyznaczania pobudzenia neuronu. Sygnały pobudzone kierowane są do nie liniowej funkcji aktywacji w celu wygenerowania na wyjściu odpowiedniego sygnału [5].

Uczenie sieci neuronowych jest wtedy stosowane gdy nie mamy informacji na temat powiązań występujących pomiędzy wszystkimi wejściami a wyjściami. Uczenie polega na dobieraniu wag przy których sieć da najlepsze rozwiązanie. Wyróżnia się uczenie z nauczycielem oraz bez nauczyciela.

Uczenie z nauczycielem polega na podawaniu wzorcowych sygnałów na wejściu i oczekiwanych sygnałów na wyjściu. Sieć dobiera sobie automatycznie wagi w celu nauczenia się funkcji opisującej powiązania sygnałów wejściowych z sygnałem wyjściowym. Uczenie bez nauczyciela polega na odbieraniu sygnałów na wejściu i na tej podstawie wyznaczania wyjścia. Proces ten odbywa się bez wag wzorcowych. [5,7-13]



Rysunek 2 Model sztucznego neuronu (Na podstawie [6])
Figure 2 Model of artificial neuron. (According to [6])

gdzie:

u_i –wejścia – to dendryty, lub ściślej sygnały przez nie dchodzące,
 w_i –wagi – są odpowiednikami synaps,
 Σ –blok sumacyjny – odpowiednik jądra,
 $f(\Sigma)$ –blok aktywacji – to odpowiednik wzgórka aksonu,
 y –wyjście – to odpowiednik aksonu

2.2.1 Parametry uczenia

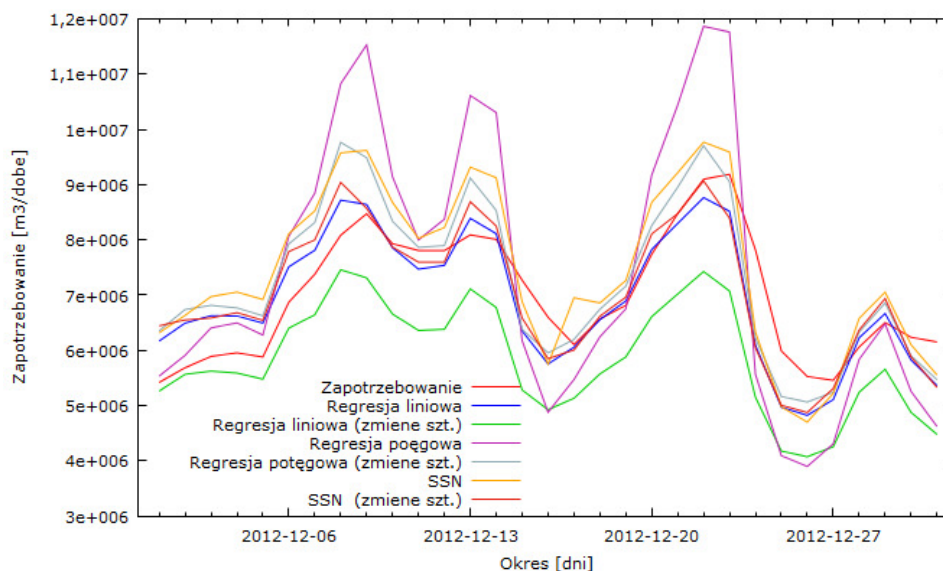
Na podstawie parametrów (Tabela 5) przeprowadzono uczenie dla sztucznej sieci neuronowej bez parametrów sztucznych i uzyskano $R^2=0,982$ a dla sztucznej sieci neuronowej z parametrami sztucznymi $R^2=0,984$.

Tabela 5 Parametry symulacji
Table 5 Simulation parameters

Nazwa parametru	Wartość
Liczba warstw ukrytych	3
Funkcja błędów	Suma kwadratów
Funkcje aktywacji – neurony ukryte	wykładnicza, logistyczna
Funkcja aktywacji- neurony wyjściowe	wykładnicza, logistyczna
Redukcja wag- warstwy ukrytej	Min.-0,0001, maks.-0,001
Redukcja wag- warstwy wyjściowej	Min.-0,0001, maks.-0,001

2.3 Wyniki

Na podstawie uzyskanych wyników dla regresji (Tabela 1-4) i z wykorzystaniem nauczonej sieci neuronowej według ustalonych parametrów (Tabela 5) przeprowadzamy prognozę dla 1 miesiąca (Rysunek 3).



Rysunek 3 Prognoza konsumpcji gazu z wykorzystaniem wszystkich metod.

Figure 3 Forecast gas consumption using all methods

3 Wnioski

Błąd MPAE (średni absolutny błąd procentowy) był najmniejszy dla regresji liniowej w przypadku nie wykorzystywania zmiennych sztucznych i wyniósł (tabela 6). Z kolei dla regresji potęgowej wynosił on 16,1% a dla sieci neuronowych 10,7%. Dodanie zmiennych sztucznych pogorszyło jakość modelu w przypadku regresji liniowej do wartości 17,1% natomiast poprawiło w przypadku regresji potęgowej i sztucznych sieci neuronowych. Wśród wszystkich modeli najlepszym okazało się sztuczne sieci neuronowe z zmiennymi sztucznymi. Podczas zmiennych sztucznych okazało się że w przypadku regresji liniowej jak i potęgowej tygodnia nie są istotne jak również siła wiatru oraz niektóre miesiące.

Tabela 6 Porównanie modeli

Table 6 Comparison models

Parametr	Regresja Liniowa		Regresja Potęgowa		SSN	
	Bez zmiennych sztucznych	Z zmiennymi sztucznymi.	Bez zmiennych sztucznych	Z zmiennymi sztucznymi.	Bez zmiennych sztucznych	Z zmiennymi sztucznymi.
MAPE [%]	7,89	17,1	16,1%	9,61	10,6	7,28

Literatura

- [1] Koronacki J., Ćwik J.: Statystyczne systemy uczące się, Wydawnictwa Naukowo-Techniczne, Warszawa 2005.
- [2] Rabieliński M.: Statystyka z programem Statistica, , Wydawnictwo Helion, Gliwice 2012.
- [3] Stanisław A.; Przystępny kurs statystyki zastosowaniem programu Statistica na przykładach z medycyny, Tom II, Wydawnictwo StatSoft Polska, Kraków 2006.
- [4] Goryl A: Wprowadzenie do ekonometrii w przykładach i zadaniach, PWN, 2000.
- [5] Żurada J., Barski M., Jędruch W.: „Sztuczne sieci neuronowe”, Wydawnictwo Naukowe PWN, Warszawa 1996.
- [6] <http://galaxy.uci.agh.edu.pl/~vlsi/AI/wstep/> data dostępu: 27.06.2016, godzina 09:00.
- [7] Tadeusiewicz R.: Sieci neuronowe, Państwowa Oficyna Wydawnicza RM, Warszawa 1993.
- [8] Tadeusiewicz R.: Elementarne wprowadzenie do technik sieci neuronowych z przykładowymi programami, Warszawa 1998.
- [9] Stefanowski J., Krawiec K.: Wykłady sieci neuronowych, 1995.
- [10] Wójcik M.: Model sieci gazowniczej oparty o sztuczne sieci neuronowe, AGH, Kraków 2005.
- [11] Kogut K.: Analiza możliwości modelowania sieci przesyłowej gazu ziemnego. AGH, Kraków 2007.
- [12] Maciejasz M.: Zastosowanie sieci neuronowej do analizy pracy sieci przesyłowych, AGH, Kraków 2006.
- [13] Praca zbiorowa pod redakcją Rymarczyka M.: Decyzje, Symulacje, Sieci Neuronowe, Wydawnictwo Wyższej Szkoły Bankowej w Poznaniu 1997.

PREDICTION OF GAS CONSUMPTION EFFICIENCY USING REGRESSION AND ARTIFICIAL NEURAL NETWORKS

Summary

Based on the collected data, such as temperature, wind power and gas consumption during the day for over two years determine the effects of weathering on gas consumption by using multiple regression, power functions and user functions. We determine the impact of the month and day (artificial parameter) to consume gas. We build models of linear regression-in, power series and artificial neural networks for determining gas consumption. We are trying to determine how best regression model and compare it to the neural network models using MAPE (mean absolute percentage error).

Keywords: linear regression, exponential regression, the parameters of artificial neural networks, gas consumption

Przesłano do redakcji: 08.03.2017 r.

Przyjęto do druku: 31.03.2017 r.