

Raport Badawczy
Research Report

RB/20/2013

**Algorytmy detekcji
i lokalizacji wycieków
w sieciach wodociągowych**

J. Studziński, I. Rojek

Instytut Badań Systemowych
Polska Akademia Nauk

Systems Research Institute
Polish Academy of Sciences



POLSKA AKADEMIA NAUK

Instytut Badań Systemowych

ul. Newelska 6

01-447 Warszawa

tel.: (+48) (22) 3810100

fax: (+48) (22) 3810105

Kierownik Zakładu zgłaszający pracę:
Prof. nadzw. dr hab. inż. Jan Studziński

Warszawa 2013

**Algorytmy detekcji i lokalizacji wycieków w sieciach
wodociągowych**

Jan Studziński, Izabela Rojek

Warszawa 2013

Spis treści

1. Wstęp	4
2. Sieci neuronowe	4
2.1. Sieci jednokierunkowe wielowarstwowe typu MLP	5
2.2. Sieć neuronowa samoorganizująca się Kohonena [20]	12
2.3. Podsumowanie sieci neuronowych użytych do eksperymentów	22
3. Eksperymenty dotyczące tworzenia algorytmów lokalizacji awarii w sieci wodociągowej	22
3.1. Stan obecny	22
3.2. Analiza lokalizacji awarii przy wykorzystaniu sieci neuronowej	23
3.2.1. Algorytm lokalizacji awarii sieci wodociągowej	23
3.2.2. Wyznaczenie punktów wrażliwych sieci wodociągowej przy użyciu algorytmu do planowania systemu monitoringu	24
3.2.3. Wykonanie obliczeń hydraulicznych sieci, dla zadanego poziomu obciążenia, za pomocą wybranego modelu hydraulicznego	24
3.2.4. Tworzenie klasyfikatora	28
4. Utworzenie i wybór najlepszego klasyfikatora lokalizującego punkty pomiarowe najbliższej awarii w postaci wybranych sieci neuronowych	34
4.1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych typu MLP	35
4.2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych Kohonena	43
5. Utworzenie i wybór najlepszego klasyfikatora lokalizującego awarie w węźle dla 10 punktów pomiarowych	51
5.1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci typu MLP	51
5.2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci Kohonena	59
5.3. Porównanie najlepszych sieci neuronowych MLP i Kohonena dla pliku z danymi oryginalnymi, poprawionymi i poprawionymi bez powtórzeń kolumn	65
6. Utworzenie i wybór najlepszego klasyfikatora lokalizującego awarie w węźle dla 20 punktów pomiarowych	66
6.1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci typu MLP	66
6.2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci Kohonena	73
6.3. Analiza wyników sieci neuronowych typu MLP i Kohonena dla pliku z danymi oryginalnymi	76
6.4. Utworzenie i wybór najlepszego klasyfikatora lokalizującego awarie w postaci sieci neuronowych typu MLP dla pliku zmodyfikowanego z usuniętymi 2 kolumnami (usunięta kolumna pp2 i pp7)	78
6.5. Utworzenie i wybór najlepszego klasyfikatora lokalizującego awarie w postaci sieci neuronowych typu Kohonena dla pliku zmodyfikowanego z usuniętymi 2 kolumnami (usunięta kolumna pp2 i pp7)	81

6.6. Analiza wyników sieci neuronowych typu MLP i Kohonena dla pliku z danymi zmodyfikowanymi bez powtórzeń.....	82
6.7. Porównanie wyników dla danych oryginalnych i zmodyfikowanych dla sieci MLP i Kohonena	83
6.8. Dane oryginalne – sieci MLP i Kohonena.....	86
7. Podsumowanie wyników dla sieci neuronowych dla 10 i 20 punktów pomiarowych z danymi oryginalnymi i zmodyfikowanymi dla sieci MLP i Kohonena.....	88
8. Zmodyfikowane sieci Kohonena	91
Literatura	91

5. Utworzenie i wybór najlepszego klasyfikatora lokalizującego awarie w węzle dla 10 punktów pomiarowych

5.1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci typu MLP

5.1.1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych typu MLP dla pliku oryginalnego danych

Modele lokalizacji awarii sieci zostały utworzone przy wykorzystaniu sieci neuronowych typu MLP. Sieci MLP niezmiennie są najbardziej rozpowszechnionymi i uniwersalnymi sieciami neuronowymi stosowanymi do rozwiązywania różnych problemów.

Eksperymenty związane z tworzeniem modeli sieci neuronowych typu MLP były parametryzowane dwoma parametrami: liczbą neuronów w warstwie ukrytej oraz liczbą epok uczących. W eksperymencie parametr liczba neuronów w warstwie ukrytej przyjmował wartości od 5 do 30, natomiast drugi parametr liczba epok uczących przyjmował wartości: 200, 500 i 1000. Po zakończeniu każdego eksperymentu, przeprowadzone były testy, które dostarczały informacji na temat błędnie zakwalifikowanych decyzji. Liczba neuronów w warstwie ukrytej została dobrana eksperymentalnie. Wejścia sieci zawierają punkty pomiarowe na sieci wodociągowej (10 wejść dla 10 punktów pomiarowych od pp1 do pp10). Wyjście sieci wskazuje węzeł, który lokalizuje awarię.

W tabeli 10 pokazano fragment danych, które posłużyły jako wejścia i wyjście sieci neuronowych.

Tabela 10. Fragment danych uczących pliku oryginalnego

pp1	pp2	pp3	pp4	pp5	pp6	pp7	pp8	pp9	pp10	węzeł
-0,066	-0,066	0,588	-0,768	0,035	0,166	0,166	2,655	0,886	-0,886	w0
-0,053	-0,053	-2,026	-0,275	0,043	0,134	0,134	1,043	2,023	-2,023	w0103
-0,053	-0,053	-2,024	-0,275	0,043	0,134	0,134	1,034	2,029	-2,029	w0130
-0,053	-0,053	-2,049	-0,264	0,043	0,133	0,133	6,034	1,772	-1,772	w0213
-0,049	-0,049	-2,701	-0,113	0,046	0,123	0,123	0,513	2,442	-2,442	w0252
-0,025	-0,025	-5,327	0,802	0,061	0,058	0,058	13,202	3,733	-3,733	w0259
-0,057	-0,057	-1,287	-0,43	0,041	0,144	0,144	4,881	1,467	-1,467	w0265
-0,03	-0,03	-4,873	0,612	0,058	0,072	0,072	12	3,4	-3,4	w0302
-0,053	-0,053	-2,022	-0,275	0,043	0,134	0,134	1,02	2,039	-2,039	w0353

Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła w pierwszym przypadku dla 10 punktów pomiarowych 380 przykłady ((37 węzłów z awarią + 1 stan bez awarii)*10), a w drugim dla 20 punktów pomiarowych 368 ((45 węzłów z awarią + 1 stan bez awarii)*8).

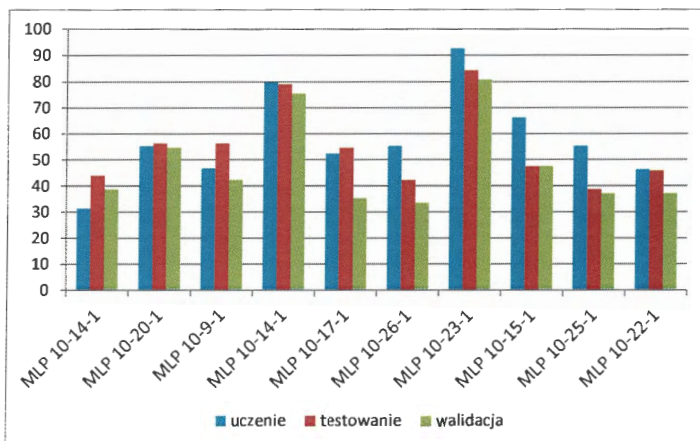
Tabela 11 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci MLP w przypadku 10 punktów pomiarowych. Jakość wyrażona jest w %.

Tabela 11. Parametry dla różnych struktur sieci typu MLP

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Średnia Jakość (%)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)
1	MLP 10-14-1	31,20	43,86	38,60	37,89	BFGS 15	Entropia	Liniowa	Softmax
2	MLP 10-20-1	55,26	56,14	54,39	55,26	BFGS 27	Entropia	Liniowa	Softmax
3	MLP 10-9-1	46,62	56,14	42,11	48,29	BFGS 83	SOS	Liniowa	Logistyczna
4	MLP 10-14-1	79,70	78,95	75,44	78,03	BFGS 208	SOS	Tanh	Wykładnicza
5	MLP 10-17-1	52,26	54,39	35,09	47,24	BFGS 30	Entropia	Logistyczna	Softmax
6	MLP 10-26-1	55,26	42,11	33,33	43,57	BFGS 97	SOS	Logistyczna	Tanh
7	MLP 10-23-1	92,48	84,21	80,70	85,80	BFGS 45	Entropia	Tanh	Softmax
8	MLP 10-15-1	66,17	47,37	47,37	53,63	BFGS 239	SOS	Wykładnicza	Tanh
9	MLP 10-25-1	55,26	38,60	36,84	43,57	BFGS 146	SOS	Logistyczna	Tanh
10	MLP 10-22-1	46,24	45,61	36,84	42,90	BFGS 21	Entropia	Wykładnicza	Softmax

Wybrana została najlepsza sieć typu MLP o strukturze 10 wejść – 23 neurony w warstwie ukrytej – 1 wyjście. Jakość uczenia to 92,48%, jakość testowania – 84,21%, jakość walidacji – 80,70%. Jeśli jakość walidacji nie odbiega od jakości testowania to sieć dobrze generalizuje strukturę danych i będzie dobrze działać dla zupełnie nowych danych. Średnia jakość sieci wyniosła 85,80%. Chociaż ważniejsza jest jakość walidacji, która jest najwyższa i tak powinniśmy określać jakość sieci. Do nauki sieci został wybrany algorytm uczenia BFGS, który daje najszybsze efekty uczenia sieci. Liczba cykli uczenia sieci wyniosła 45. Jako funkcji błędu do oceny sieci w procesie uczenia użyto entropii. Do aktywacji neuronów w warstwie ukrytej użyto Tanh (tangens hiperboliczny), a do aktywacji neuronów w warstwie wyjściowej użyto funkcji Softmax.

Na rysunku 17 pokazano wykres jakości sieci neuronowych typu MLP dla 10 punktów pomiarowych. Jakość wyrażona jest w %.



Rys. 17. Wykres jakości sieci neuronowych typu MLP – 10 punktów pomiarowych

5.1.2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych typu MLP dla pliku poprawionego

Eksperymenty związane z tworzeniem modeli sieci neuronowych typu MLP były parametryzowane dwoma parametrami: liczbą neuronów w warstwie ukrytej oraz liczbą epok uczących. W eksperymencie parametr liczba neuronów w warstwie ukrytej przyjmował wartości od 5 do 30, natomiast drugi parametr liczba epok uczących przyjmował wartości: 200, 500 i 1000. Po zakończeniu każdego eksperymentu, przeprowadzone były testy, które dostarczały informacji na temat błędnie zakwalifikowanych decyzji. Liczba neuronów w warstwie ukrytej została dobrana eksperymentalnie. Wejścia sieci zawierają punkty pomiarowe na sieci wodociągowej (10 wejść dla 10 punktów pomiarowych od pp1 do pp10). Wyjście sieci wskazuje węzeł, który lokalizuje awarię.

W tabeli 12 pokazano fragment danych, które posłużyły jako wejścia i wyjście sieci neuronowych.

Tabela 12. Fragment danych uczących pliku poprawionego

pp1	pp2	pp3	pp4	pp5	pp6	pp7	pp8	pp9	pp10	węzeł
0	0	0	0	0	0	0	0	0	0	w0
-0,013	-0,013	2,614	-0,493	-0,008	0,032	0,032	1,612	-1,137	1,137	w0103
-0,013	-0,013	2,612	-0,493	-0,008	0,032	0,032	1,621	-1,143	1,143	w0130
-0,013	-0,013	2,637	-0,504	-0,008	0,033	0,033	-3,379	-0,886	0,886	w0213
-0,017	-0,017	3,289	-0,655	-0,011	0,043	0,043	2,142	-1,556	1,556	w0252
-0,041	-0,041	5,915	-1,57	-0,026	0,108	0,108	-10,547	-2,847	2,847	w0259
-0,009	-0,009	1,875	-0,338	-0,006	0,022	0,022	-2,226	-0,581	0,581	w0265
-0,036	-0,036	5,461	-1,38	-0,023	0,094	0,094	-9,345	-2,514	2,514	w0302

-0,013	-0,013	2,61	-0,493	-0,008	0,032	0,032	1,635	-1,153	1,153	w0353
-0,008	-0,008	1,071	-0,313	-0,005	0,02	0,02	-1,427	-0,395	0,395	w0375
-0,087	-0,087	10,455	-3,21	-0,054	0,242	0,242	9,081	-7,089	7,089	w0488
-0,013	-0,013	0,805	-0,505	-0,008	0,033	0,033	-1,234	-0,518	2,29	w0528
-0,03	-0,03	4,986	-1,162	-0,02	0,078	0,078	3,702	-2,799	2,799	w0616
-0,035	-0,035	1,508	-1,341	-0,023	0,091	0,091	-2,547	-1,287	1,287	w0673
-0,046	-0,046	6,415	-1,743	-0,029	0,121	0,121	5,381	-3,534	3,534	w0811
-0,055	-0,055	7,343	-2,071	-0,035	0,147	0,147	6,149	-9,046	9,046	w0901
-0,008	-0,008	-0,435	-0,331	-0,006	0,021	0,021	-0,074	-0,223	0,223	w1026

Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła w pierwszym przypadku dla 10 punktów pomiarowych 380 przykłady, a w drugim dla 20 punktów pomiarowych 360.

Tabela 13 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci MLP w przypadku 10 punktów pomiarowych. Jakość wyrażona jest w %.

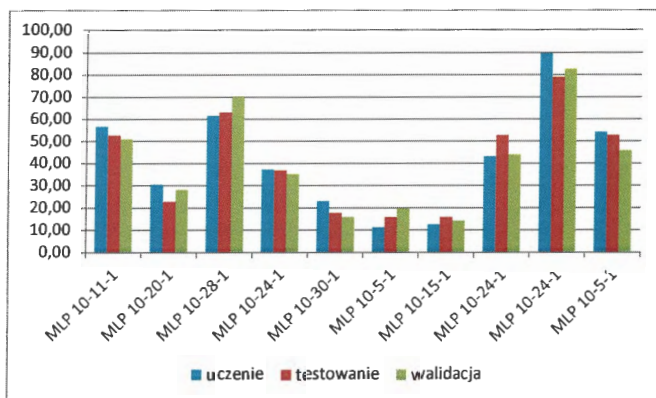
Tabela 13. Parametry dla różnych struktur sieci typu MLP

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Średnia jakość (%)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)
1	MLP 10-11-1	56,77	52,63	50,88	53,43	BFGS 62	Entropia	Liniowa	Softmax
2	MLP 10-20-1	30,45	22,81	28,07	27,11	BFGS 41	Entropia	Logistyczna	Softmax
3	MLP 10-28-1	61,65	63,16	70,18	65,00	BFGS 182	SOS	Logistyczna	Wykładnicza
4	MLP 10-24-1	37,22	36,84	35,09	36,38	BFGS 37	Entropia	Liniowa	Softmax
5	MLP 10-30-1	22,93	17,54	15,79	18,76	BFGS 73	SOS	Wykładnicza	Liniowa
6	MLP 10-5-1	11,28	15,79	19,30	15,46	BFGS 23	Entropia	Wykładnicza	Softmax
7	MLP 10-15-1	12,41	15,79	14,04	14,08	BFGS 21	Entropia	Wykładnicza	Softmax
8	MLP 10-24-1	43,23	52,63	43,86	46,57	BFGS 133	SOS	Liniowa	Wykładnicza
9	MLP 10-24-1	89,47	78,95	82,46	83,63	BFGS 155	SOS	Tanh	Logistyczna
10	MLP 10-5-1	54,14	52,63	45,61	50,79	BFGS 45	Entropia	Liniowa	Softmax

Wybrana została najlepsza sieć typu MLP o strukturze 10 wejść – 24 neurony w warstwie ukrytej – 1 wyjście. Jakość uczenia to **89,47%**, jakość testowania – **78,95%**, jakość walidacji – **82,46%**. Jeśli jakość walidacji nie odbiega od jakości testowania to sieć dobrze generalizuje strukturę danych i będzie dobrze działać dla zupełnie nowych danych. Średnia jakość sieci wyniosła **83,63%**. Choć ważniejsza jest jakość

walidacji, która jest najwyższa i tak powinniśmy określać jakość sieci. Do nauki sieci został wybrany algorytm uczenia sieci BFGS, który daje najszybsze efekty uczenia sieci. Liczba cykli uczenia sieci wyniosła 155. Jako funkcji błędu do oceny sieci w procesie uczenia użyto funkcji SOS. Do aktywacji neuronów w warstwie ukrytej użyto Tanh (tangens hiperboliczny), a do aktywacji neuronów w warstwie wyjściowej użyto funkcji Logistycznej. Długość uczenia wynikała z użytych funkcji błędu i aktywacji.

Na rysunku 18 pokazano wykres jakości sieci neuronowych typu MLP dla 10 punktów pomiarowych. Jakość wyrażona jest w %.



Rys. 18. Wykres jakości sieci neuronowych typu MLP – 10 punktów pomiarowych (plik danych poprawiony)

5.1.3. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych typu MLP dla pliku poprawionego bez powtórzeń

Eksperymenty związane z tworzeniem modeli sieci neuronowych typu MLP były parametryzowane dwoma parametrami: liczbą neuronów w warstwie ukrytej oraz liczbą epok uczących. W eksperymencie parametr liczba neuronów w warstwie ukrytej przyjmował wartości od 5 do 30, natomiast drugi parametr liczba epok uczących przyjmował wartości: 200, 500 i 1000. Po zakończeniu każdego eksperymentu, przeprowadzone były testy, które dostarczały informacji na temat błędnie zakwalifikowanych decyzji. Liczba neuronów w warstwie ukrytej została dobrana eksperymentalnie. Wejścia sieci zawierają punkty pomiarowe na sieci wodociągowej (7 wejść dla 10 punktów pomiarowych od pp1 do pp10 bez pp2, pp7 i pp10). Wyjście sieci wskazuje węzeł, który lokalizuje awarię.

W tabeli 14 pokazano fragment danych, które posłużyły jako wejścia i wyjście sieci neuronowych.

Tabela 14. Fragment danych uczących pliku poprawionego bez powtórzeń

pp1	pp3	pp4	pp5	pp6	pp8	pp9	węzeł
0	0	0	0	0	0	0	w0
-0.013	2.614	-0.493	-0.008	0.032	1.612	-1.137	w0103
-0.013	2.612	-0.493	-0.008	0.032	1.621	-1.143	w0130
-0.013	2.637	-0.504	-0.008	0.033	-3.379	-0.886	w0213
-0.017	3.289	-0.655	-0.011	0.043	2.142	-1.556	w0252
-0.041	5.915	-1.57	-0.026	0.108	-10.547	-2.847	w0259
-0.009	1.875	-0.338	-0.006	0.022	-2.226	-0.581	w0265
-0.036	5.461	-1.38	-0.023	0.094	-9.345	-2.514	w0302
-0.013	2.61	-0.493	-0.008	0.032	1.635	-1.153	w0353
-0.008	1.071	-0.313	-0.005	0.02	-1.427	-0.395	w0375
-0.087	10.455	-3.21	-0.054	0.242	9.081	-7.089	w0488
-0.013	0.805	-0.505	-0.008	0.033	-1.234	-0.518	w0528
-0.03	4.986	-1.162	-0.02	0.078	3.702	-2.799	w0616
-0.035	1.508	-1.341	-0.023	0.091	-2.547	-1.287	w0673
-0.046	6.415	-1.743	-0.029	0.121	5.381	-3.534	w0811
-0.055	7.343	-2.071	-0.035	0.147	6.149	-9.046	w0901
-0.008	-0.435	-0.331	-0.006	0.021	-0.074	-0.223	w1026

Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła w pierwszym przypadku dla 10 punktów pomiarowych 380 przykłady, a w drugim dla 20 punktów pomiarowych 360.

Tabela 15 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci MLP w przypadku 10 punktów pomiarowych. Jakość wyrażona jest w %.

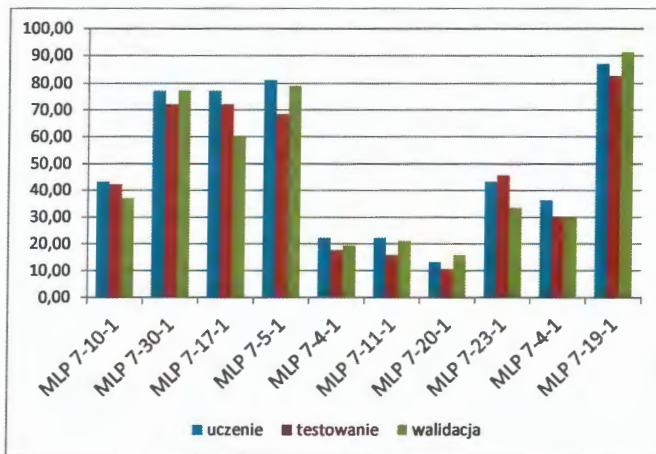
Tabela 15. Parametry różnych struktur sieci MLP

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Średnia jakość (%)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)
1	MLP 7-10-1	43,23	42,11	36,84	40,73	BFGS 40	Entropia	Logistyczna	Softmax
2	MLP 7-30-1	77,07	71,93	77,19	75,40	BFGS 194	SOS	Tanh	Wykładnicza
3	MLP 7-17-1	77,07	71,93	59,65	69,55	BFGS 52	Entropia	Tanh	Softmax
4	MLP 7-5-1	81,20	68,42	78,95	76,19	BFGS 260	SOS	Liniowa	Wykładnicza
5	MLP 7-4-1	22,18	17,54	19,30	19,67	BFGS 33	Entropia	Logistyczna	Softmax
6	MLP 7-11-1	22,18	15,79	21,05	19,67	BFGS 26	Entropia	Liniowa	Softmax
7	MLP 7-20-1	13,16	10,53	15,79	13,16	BFGS 17	Entropia	Wykładnicza	Softmax

8	MLP 7-23-1	43,23	45,61	33,33	40,73	BFGS 80	SOS	Liniowa	Logistyczna
9	MLP 7-4-1	36,09	29,82	29,82	31,91	BFGS 69	Entropia	Logistyczna	Softmax
10	MLP 7-19-1	86,84	82,46	91,23	86,84	BFGS 78	Entropia	Liniowa	Softmax

Wybrana została najlepsza sieć typu MLP o strukturze 10 wejść – 24 neurony w warstwie ukrytej – 1 wyjście. Jakość uczenia to 86,84%, jakość testowania – 82,46%, jakość walidacji – 91,23%. Jeśli jakość walidacji nie odbiega od jakości testowania to sieć dobrze generalizuje strukturę danych i będzie dobrze działać dla zupełnie nowych danych. Średnia jakość sieci wyniosła 86,84%. Chociaż ważniejsza jest jakość walidacji, która jest najwyższa i tak powinniśmy określać jakość sieci. Do nauki sieci został wybrany algorytm uczenia BFGS, który daje najszybsze efekty uczenia sieci. Liczba cykli uczenia sieci wyniosła 78. Jako funkcji błędu do oceny sieci w procesie uczenia użyto entropii. Do aktywacji neuronów w warstwie ukrytej użyto funkcji liniowej, a do aktywacji neuronów w warstwie wyjściowej użyto funkcji Softmax. Długość uczenia wynikała z użytych funkcji aktywacji.

Na rysunku 19 pokazano wykres jakości sieci neuronowych typu MLP dla 10 punktów pomiarowych. Jakość wyrażona jest w %.



Rys. 19. Wykres jakości sieci neuronowych typu MLP – 10 punktów pomiarowych (plik danych poprawiony bez powtórzeń)

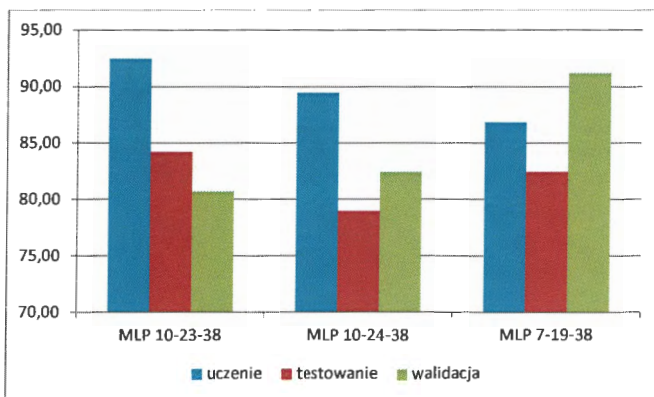
5.1.4. Porównanie najlepszych sieci neuronowych MLP dla pliku z danymi oryginalnymi, poprawionymi i bez powtórzeń

Tabela 16 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci MLP w przypadku 10 punktów pomiarowych dla pliku danych oryginalnego, poprawionego i poprawionego bez powtórzeń. Jakość wyrażona jest w %.

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)	Jakość sieci w %
1	MLP 10-23-38 Plik oryginalny	92,48120	84,21053	80,70175	BFGS 45	Entropia	Tanh	Softmax	85,80
2	MLP 10-24-38 Plik poprawiony	89,47368	78,94737	82,45614	BFGS 155	SOS	Tanh	Logistyczna	83,63
3	MLP 7-19-38 Plik poprawiony bez powtórzeń	86,84211	82,45614	91,22807	BFGS 78	Entropia	Liniowa	Softmax	86,84

Najlepszą siecią okazała się sieć MLP 7-19-38 wykonana na podstawie pliku z danymi poprawionymi bez powtórzeń. Jakość tej sieci wyniosła 86,84%.

Na rysunku 20 pokazano wykres jakości sieci neuronowych typu MLP dla 10 punktów pomiarowych. Jakość wyrażona jest w %.



Rys. 20. Wykres jakości sieci neuronowych typu MLP – 10 punktów pomiarowych

5.2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci Kohonena

5.2. 1. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych Kohonena dla pliku oryginalnego danych

Eksperymenty związane z tworzeniem modeli sieci neuronowych typu Kohonena były parametryzowane dwoma parametrami: liczbą neuronów w warstwie topologicznej oraz liczbą epok uczących. W eksperymencie parametr 'liczba neuronów w warstwie topologicznej' przyjmował wartości 2x8, 5x5, 10x10, natomiast drugi parametr 'liczba epok uczących' przyjmował wartości od 1000 do 20000. Po zakończeniu każdego eksperymentu, przeprowadzone były testy, które dostarczały informacji na temat błędnie zakwalifikowanych decyzji. Liczba neuronów w warstwie topologicznej została dobrana eksperymentalnie. Wejścia sieci zawierają punkty pomiarowe na sieci wodociągowej oraz węzeł na który wskazują punkty pomiarowe. W tej sieci nie ma wyjścia. W tabeli 17 pokazano fragment danych, które posłużyły jako wejścia sieci neuronowych. Sieć posiada 11 wejść.

Tabela 17. Fragment danych uczących pliku poprawionego bez powtórzeń

pp1	pp2	pp3	pp4	pp5	pp6	pp7	pp8	pp9	pp10	węzeł
-0,066	-0,066	0,588	-0,768	0,035	0,166	0,166	2,655	0,886	-0,886	w0
-0,053	-0,053	-2,026	-0,275	0,043	0,134	0,134	1,043	2,023	-2,023	w0103
-0,053	-0,053	-2,024	-0,275	0,043	0,134	0,134	1,034	2,029	-2,029	w0130
-0,053	-0,053	-2,049	-0,264	0,043	0,133	0,133	6,034	1,772	-1,772	w0213
-0,049	-0,049	-2,701	-0,113	0,046	0,123	0,123	0,513	2,442	-2,442	w0252
-0,025	-0,025	-5,327	0,802	0,061	0,058	0,058	13,202	3,733	-3,733	w0259
-0,057	-0,057	-1,287	-0,43	0,041	0,144	0,144	4,881	1,467	-1,467	w0265
-0,03	-0,03	-4,873	0,612	0,058	0,072	0,072	12	3,4	-3,4	w0302
-0,053	-0,053	-2,022	-0,275	0,043	0,134	0,134	1,02	2,039	-2,039	w0353
-0,058	-0,058	-0,483	-0,455	0,04	0,146	0,146	4,082	1,281	-1,281	w0375
0,021	0,021	-9,867	2,442	0,089	-0,076	-0,076	-6,426	7,975	-7,975	w0488
-0,053	-0,053	-0,217	-0,263	0,043	0,133	0,133	3,889	1,404	-1,404	w0528
-0,036	-0,036	-4,398	0,394	0,055	0,088	0,088	-1,047	3,685	-3,685	w0616
-0,031	-0,031	-0,92	0,573	0,058	0,075	0,075	5,202	2,173	-2,173	w0673
-0,02	-0,02	-5,827	0,975	0,064	0,045	0,045	-2,726	4,42	-4,42	w0811
-0,011	-0,011	-6,755	1,303	0,07	0,019	0,019	-3,494	9,932	-9,932	w0901
-0,058	-0,058	1,023	-0,437	0,041	0,145	0,145	2,729	1,109	-1,109	w1026

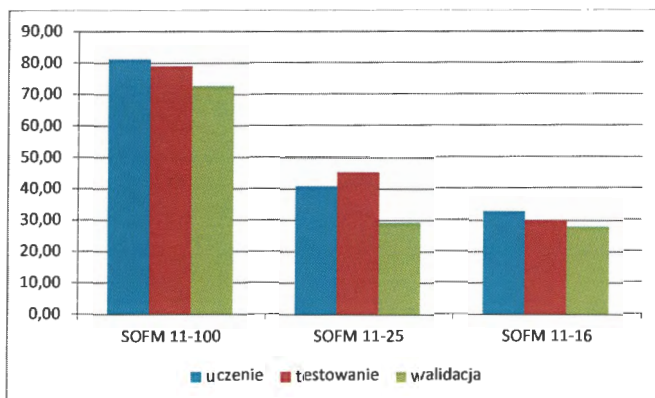
Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła dla 10 punktów pomiarowych 380 ((37 węzłów z awarią + 1 stan bez awarii)*10), a w drugim dla 20 punktów pomiarowych 368 ((45 węzłów z awarią + 1 stan bez awarii)*8).

Tabela 18 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci Kohonena w przypadku 10 punktów pomiarowych.

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Algorytm uczenia	Jakość sieci w %
1	SOFM 11-100	81,20	78,99	72,85	Kohonena 1000	77,68
2	SOFM 11-25	40,95	45,38	29,35	Kohonena 1000	38,56
3	SOFM 11-16	32,97	30,12	28,16	Kohonena 1000	30,42

Wysoka jakość pierwszej sieci wynika z dużej liczby neuronów w topologii sieci, czyli 100 neuronów (10x10). Dla 11 wejść dopiero sieć o takiej topologii dała radę pogrupować dane w taki sposób, który pozwolił na uzyskanie jakości w 77,68%.

Na rysunku 21 pokazano wykres jakości sieci neuronowych typu Kohonena dla 10 punktów pomiarowych.



Rys. 21. Wykres jakości sieci neuronowych Kohonena – 10 punktów pomiarowych (Plik danych oryginalny)

5.2. 2. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych Kohonena dla pliku poprawionego danych

W tabeli 19 pokazano fragment danych, które posłużyły jako wejścia sieci neuronowych. Sieć posiada 11 wejść.

Tabela 19. Fragment danych poprawionego pliku uczącego sieci neuronowej Kohonena

pp1	pp2	pp3	pp4	pp5	pp6	pp7	pp8	pp9	pp10	węzeł
0	0	0	0	0	0	0	0	0	0	w
-0,013	-0,013	2,614	-0,493	-0,008	0,032	0,032	1,612	-1,137	1,137	w010
-0,013	-0,013	2,612	-0,493	-0,008	0,032	0,032	1,621	-1,143	1,143	w013
-0,013	-0,013	2,637	-0,504	-0,008	0,033	0,033	-3,379	-0,886	0,886	w021
-0,017	-0,017	3,289	-0,655	-0,011	0,043	0,043	2,142	-1,556	1,556	w025
-0,041	-0,041	5,915	-1,57	-0,026	0,108	0,108	-10,547	-2,847	2,847	w025
-0,009	-0,009	1,875	-0,338	-0,006	0,022	0,022	-2,226	-0,581	0,581	w026
-0,036	-0,036	5,461	-1,38	-0,023	0,094	0,094	-9,345	-2,514	2,514	w030
-0,013	-0,013	2,61	-0,493	-0,008	0,032	0,032	1,635	-1,153	1,153	w035
-0,008	-0,008	1,071	-0,313	-0,005	0,02	0,02	-1,427	-0,395	0,395	w037
-0,087	-0,087	10,455	-3,21	-0,054	0,242	0,242	9,081	-7,089	7,089	w048
-0,013	-0,013	0,805	-0,505	-0,008	0,033	0,033	-1,234	-0,518	2,29	w052
-0,03	-0,03	4,986	-1,162	-0,02	0,078	0,078	3,702	-2,799	2,799	w061
-0,035	-0,035	1,508	-1,341	-0,023	0,091	0,091	-2,547	-1,287	1,287	w067
-0,046	-0,046	6,415	-1,743	-0,029	0,121	0,121	5,381	-3,534	3,534	w081
-0,055	-0,055	7,343	-2,071	-0,035	0,147	0,147	6,149	-9,046	9,046	w090
-0,008	-0,008	-0,435	-0,331	-0,006	0,021	0,021	-0,074	-0,223	0,223	w102
-0,017	-0,017	-0,873	-0,955	-0,01	0,149	0,149	-0,049	-0,125	0,125	w107
-0,048	-0,048	6,631	-1,821	-0,031	0,127	0,127	5,472	-5,997	5,997	w119
-0,031	-0,031	3,5	-1,185	-0,02	0,08	0,08	0,123	-1,31	1,31	w121

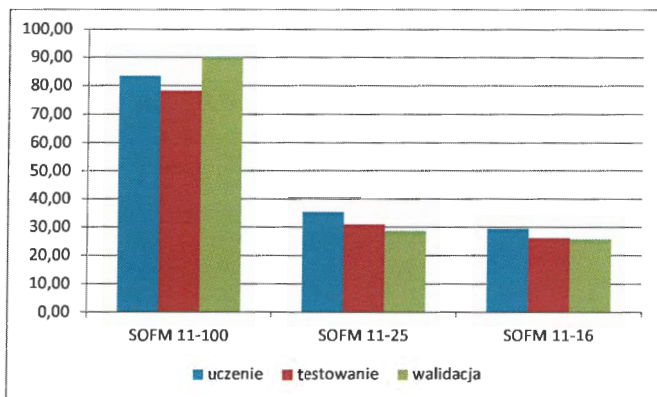
Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła dla 10 punktów pomiarowych 380.

Tabela 20 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci Kohonena w przypadku 10 punktów pomiarowych.

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Algorytm uczenia	Jakość sieci w %
1	SOFM 11-100	83,43	78,19	90,08	Kohonena 1000	83,90
2	SOFM 11-25	35,35	30,94	28,80	Kohonena 1000	31,70
3	SOFM 11-16	29,48	26,16	25,81	Kohonena 1000	27,15

Wysoka jakość pierwszej sieci wynika z dużej liczby neuronów w topologii sieci, czyli 100 neuronów (10x10). Dla 11 wejść dopiero sieć o takiej topologii dała radę pogrupować dane w taki sposób, który pozwolił na uzyskanie jakości w 83,90%.

Na rysunku 22 pokazano wykres jakości sieci neuronowych typu Kohonena dla 10 punktów pomiarowych.



Rys. 22. Wykres jakości sieci neuronowych Kohonena – 10 punktów pomiarowych (Plik danych poprawiony)

5.2. 3. Utworzenie i wybór najlepszego klasyfikatora w postaci sieci neuronowych Kohonena dla pliku poprawionego danych bez powtórzeń

W tabeli 21 pokazano fragment danych, które posłużyły jako wejścia sieci neuronowych. Sieć posiada 8 wejść.

Tabela 21. Fragment danych pliku uczącego poprawionego bez powtórzeń kolumn

pp1	pp3	pp4	pp5	pp6	pp8	pp9	węzeł
0	0	0	0	0	0	0	w0
-0,013	2,614	-0,493	-0,008	0,032	1,612	-1,137	w0103
-0,013	2,612	-0,493	-0,008	0,032	1,621	-1,143	w0130
-0,013	2,637	-0,504	-0,008	0,033	-3,379	-0,886	w0213
-0,017	3,289	-0,655	-0,011	0,043	2,142	-1,556	w0252
-0,041	5,915	-1,57	-0,026	0,108	-10,547	-2,847	w0259
-0,009	1,875	-0,338	-0,006	0,022	-2,226	-0,581	w0265
-0,036	5,461	-1,38	-0,023	0,094	-9,345	-2,514	w0302
-0,013	2,61	-0,493	-0,008	0,032	1,635	-1,153	w0353
-0,008	1,071	-0,313	-0,005	0,02	-1,427	-0,395	w0375
-0,087	10,455	-3,21	-0,054	0,242	9,081	-7,089	w0488
-0,013	0,805	-0,505	-0,008	0,033	-1,234	-0,518	w0528
-0,03	4,986	-1,162	-0,02	0,078	3,702	-2,799	w0616
-0,035	1,508	-1,341	-0,023	0,091	-2,547	-1,287	w0673
-0,046	6,415	-1,743	-0,029	0,121	5,381	-3,534	w0811
-0,055	7,343	-2,071	-0,035	0,147	6,149	-9,046	w0901
-0,008	-0,435	-0,331	-0,006	0,021	-0,074	-0,223	w1026
-0,017	-0,873	-0,955	-0,01	0,149	-0,049	-0,125	w1073
-0,048	6,631	-1,821	-0,031	0,127	5,472	-5,997	w1195

-0,031	3,5	-1,185	-0,02	0,08	0,123	-1,31	w1216
-0,03	-2,443	-1,157	-0,019	0,077	-0,219	-0,61	w1375

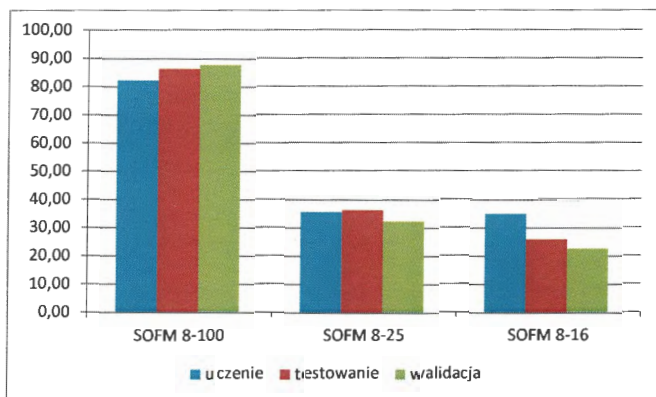
Plik uczący zawiera 70% przykładów, plik testowy 15%, a plik walidacyjny kolejne 15%. Liczba przykładów wyniosła dla 10 punktów pomiarowych 380.

Tabela 22 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci Kohonena w przypadku 10 punktów pomiarowych.

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Algorytm uczenia	Jakość sieci w %
1	SOFM 8-100	82,25	86,34	87,72	Kohonena 1000	85,44
2	SOFM 8-25	35,72	36,33	32,25	Kohonena 1000	34,76
3	SOFM 8-16	35,03	25,95	22,63	Kohonena 1000	27,87

Wysoka jakość pierwszej sieci wynika z dużej liczby neuronów w topologii sieci, czyli 100 neuronów (10x10). Dla 11 wejść dopiero sieć o takiej topologii dała radę pogrupować dane w taki sposób, który pozwolił na uzyskanie jakości w 85,44%. Zmniejszenie liczby wejść poskutkowało poprawieniem jakości.

Na rysunku 23 pokazano wykres błędu sieci neuronowych typu Kohonena dla 10 punktów pomiarowych.



Rys. 23. Wykres jakości sieci neuronowych Kohonena – 10 punktów pomiarowych (Plik danych uformatowany bez powtórzeń)

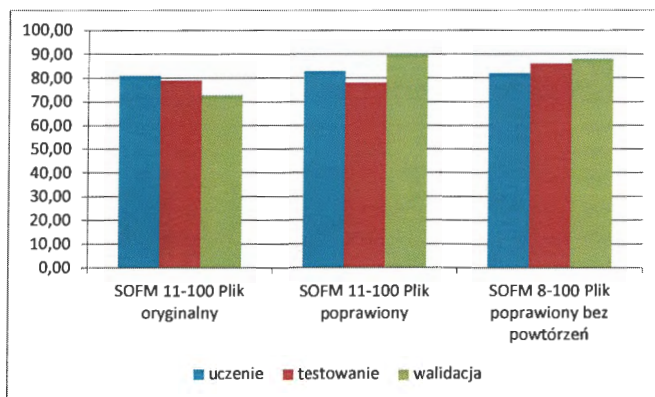
5.2.4. Porównanie najlepszych sieci neuronowych Kohonena dla pliku z danymi oryginalnymi, poprawionymi i bez powtórzeń

Tabela 23 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci Kohonena w przypadku 10 punktów pomiarowych dla pliku danych oryginalnego, poprawionego i poprawionego bez powtórzeń. Jakość wyrażona jest w %.

Id	Nazwa sieci	Jakość uczenia	Jakość testowania	Jakość walidacji	Algorytm uczenia	Jakość sieci w %
1	SOFM 11-100 Plik oryginalny	81,00	79,00	73,00	Kohonena 1000	77,68
2	SOFM 11-100 Plik poprawiony	83,00	78,00	90,00	Kohonena 1000	83,90
3	SOFM 8-100 Plik poprawiony bez powtórzeń	82,00	86,00	88,00	Kohonena 1000	85,44

Powtórnie najlepszą siecią okazała się sieć Kohonena 8-100 wykonana na podstawie pliku z danymi poprawionymi bez powtórzeń. Jakość tej sieci wyniosła 85,44%.

Na rysunku 24 pokazano wykres jakości sieci neuronowych typu KOHONEN dla 10 punktów pomiarowych. Jakość wyrażona jest w %.



Rys. 24. Wykres jakości sieci neuronowych typu Kohonen – 10 punktów pomiarowych

5.3. Porównanie najlepszych sieci neuronowych MLP i Kohonena dla pliku z danymi oryginalnymi, poprawionymi i poprawionymi bez powtórzeń kolumn

Tabela 24 ilustruje parametry opisujące eksperymenty dla różnych struktur sieci MLP i Kohonena w przypadku 10 punktów pomiarowych dla pliku danych oryginalnego, poprawionego i poprawionego bez powtórzeń. Jakość wyrażona jest w %.

Tabela 24. Parametry sieci MLP i Kohonena

Id	Nazwa sieci	Jakość sieci w %	Plik danych
1	MLP 10-23-38	85,80	oryginalny
2	MLP 10-24-38	83,63	poprawiony
3	MLP 7-19-38	86,84	poprawiony bez powtórzeń
4	SOFM 11-100	77,68	oryginalny
5	SOFM 11-100	83,90	poprawiony
6	SOFM 8-100	85,44	poprawiony bez powtórzeń

Można powiedzieć, że sieć 3 i 6 są porównywalne, a minimalnie lepszą okazała się sieć typu MLP (86,84%).

the 1990s, the number of people in the world who are illiterate has increased from 1.2 billion to 1.5 billion (UNEP 1998).

There are many reasons for the increase in illiteracy. One of the main reasons is the rapid population growth in the developing countries. The population of the world is expected to reach 6 billion by the year 2000, and 7 billion by the year 2015 (UNEP 1998). The rapid population growth has led to a decrease in the number of people who are able to read and write. Another reason is the lack of access to education. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.

The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write. The lack of access to education is one of the main reasons for the increase in illiteracy. In many developing countries, there are no schools or libraries, and the people are unable to learn to read and write.