

PORÓWNANIE RÓŻNYCH TYPÓW SIECI NEURONOWYCH DO LOKALIZACJI AWARII W SIECIACH WODOCIĄGOWYCH¹

***Streszczenie.** W artykule prezentowane są różne typy sieci neuronowych do lokalizacji awarii w sieci wodociągowej. Obecne wykorzystanie systemów monitorowania nie odpowiada ich możliwościom. Współcześnie systemy monitoringu służą jako autonomiczne programy do zbierania informacji o przepływach i ciśnieniach wody w pompowniach źródłowych, hydroforniach strefowych i końcówkach sieci wodociągowej, dając ogólną wiedzę o stanie jej pracy, gdy jednocześnie mogą i powinny być wykorzystane jako elementy IT systemów zarządzania siecią, w tym w szczególności w zakresie wykrywania i lokalizacji wycieków wody. Modele lokalizacji awarii sieci zostały utworzone przy wykorzystaniu jednokierunkowych sieci neuronowych ze wsteczną propagacją błędu typu MLP i sieci Kohonena.*

***Słowa kluczowe:** sieci wodociągowe, modele hydrauliczne sieci, wykrywanie i lokalizacja wycieków wody, sieci neuronowe MLP i Kohonena*

1. Wprowadzenie

W Instytucie Badań Systemowych Polskiej Akademii Nauk (IBS PAN) został opracowany system IT komputerowego wspomagania zarządzania komunalnymi sieciami wodociągowymi i jednym z jego zadań jest lokalizacja awarii sieci wodociągowej [9]. Aby to zrobić w sieci wodociągowej zainstalowano model hydrauliczny oraz system SCADA. W artykule przedstawiono nowe algorytmy dla lokalizacji awarii w sieci wodociągowej przy użyciu modelu hydraulicznego sieci wodociągowej, systemu SCADA oraz dwóch sieci neuronowych: sieci neuronowej wielowarstwowej ze wsteczną propagacją błędu (MLP) i sieci Kohonena.

Przedsiębiorstwo wodociągowe w zakresie sieci wodociągowej zajmuje się dystrybucją wody odpowiedniej jakości w ilości gwarantującej zaspokojenie potrzeb odbiorców, poprawną eksploatacją sieci wodociągowej zapewniającą właściwe ciśnienie w węzłach odbiorczych, sprawnym usuwaniem awarii oraz planowaniem i wykonywaniem prac związanych z konserwacją, modernizacją i rozbudową sieci [2]. Można powiedzieć, że są na ogół 3 główne zadania zarządzania siecią wodociągową: dostarczanie wody w odpowiedniej ilości użytkownikom sieci wodociągowej [6], produkcja wody o odpowiedniej jakości [6,14] i redukcja kosztów operacyjnych sieci wodociągowej; w tym ostatnim przypadku redukcja strat wody spowodowanych przez awarie sieci wodociągowej jest jednym z najważniejszych problemów zarządzania [3,4,8,13,15]. Spowodowane tym straty wody w rurociągach, dochodzące niekiedy nawet do 30%, wpływają niekorzystnie na wyniki finansowe przedsiębiorstw wodociągowych, które utraconej wody nie mogą sprzedać. U odbiorców wody straty te odbijają się w zwiększonych opłatach za wyprodukowaną i niez użytą faktycznie wodę. Szybka lokalizacja i eliminacja ukrytych wycieków wody z nieszczelnych rurociągów przynosi zatem wymierne korzyści ekonomiczne zarówno dostawcom, jak i

¹ Artykuł wydany w ramach projektu Nr NR14-0011-10/2010 finansowanego przez Ministerstwo Nauki i Szkolnictwa Wyższego.

odbiorcom wody, czyli przedsiębiorstwu wodociągowemu i użytkownikom sieci wodociągowej.

2. Algorytmy lokalizacji awarii w sieci wodociągowej

Istnieją różne podejścia i algorytmy obliczeniowe do wspomaganie wykrywania i lokalizacji wycieku wody w sieci wodociągowej prezentowane w literaturze. W każdym przypadku podstawą obliczeń jest model hydrauliczny sieci wodociągowej i pomiary przepływu wody, względnie przepływu i ciśnienia pozyskiwane z punktów pomiarowych systemu monitoringu zainstalowanego w sieci. Oznacza to, że dla praktycznej realizacji tych algorytmów jest niezbędna odpowiednia infrastruktura informatyczna wdrożona i eksploatowana w przedsiębiorstwie wodociągowym. Różne etapy dotyczące awarii w sieci wodociągowej można określić w następujący sposób:

- wykrycie awarii – stwierdza się stan awaryjny w postaci strat wody, ale nie jest znana jego lokalizacja,
- lokalizacja awarii – stwierdza się stan awaryjny i jego dokładną lub przybliżoną lokalizację za pomocą odpowiednio opracowanych algorytmów, z wykorzystaniem systemów monitoringu, modeli hydraulicznych sieci wodociągowej i w szczególności sieci neuronowych,
- przeciwdziałanie awarii – na podstawie danych historycznych o stanach awaryjnych następuje prognozowanie kolejnych stanów awaryjnych i opracowuje się plany rewitalizacji sieci wodociągowej.

Na pierwszym etapie badań jedynie systemy monitorowania i diagnostyki (SCADA) zainstalowane na sieci wodociągowej mogą być używane do znalezienia i zlokalizowania powstałych wycieków. Metody diagnostyczne zaimplementowane w tych systemach wykorzystują w swoich obliczeniach pomiarów przepływu wody w odcinkach sieci wodociągowej takie parametry jak ciśnienie, prędkość przepływu i temperaturę [1,3,13] a jedyną ich wadą jest konieczność instalacji wielu urządzeń pomiarowe w rurociągach. To powoduje wysokie koszty całej instalacji, które w większości nie są do pokrycia przez wodociągi.

W drugim etapie, obok infrastruktury technicznej takiej jak SCADA również jest użytych kilka aplikacji do zlokalizowania awarii w sieciach wodociągowych, co sprawia, że problem jest bardziej skomplikowany. W prostszej wersji badań uzupełnieniem systemu monitoringu są dodatkowe elementy, czyli matematyczne modele sieci wodociągowej (model hydrauliczny sieci wodociągowej do symulacji sieci i parametryczny model do modelowania hydraulicznego) [15, 16]; często w tym ostatnim przypadku są wykorzystywane sieci neuronowe, zbiory rozmyte lub modele szeregów czasowych. Zintegrowane systemy są wykorzystywane przez bardziej złożone wersje badań, na które składają się nie tylko system SCADA i modele matematyczne sieci wodociągowej, ale także system GIS i różne algorytmy optymalizacji. Taka infrastruktura techniczna i informacyjna pozwala nie tylko wykryć i zlokalizować nieprawidłowości sieci wodociągowej, ale również pozwala zarządzać siecią, wykonując takie zadania, jak kontrola sieci wodociągowej, analiza jakości wody, optymalizacja i projektowanie sieci, itp. [8, 11]. To oznacza, że wysoko rozwinięte technologie komputerowe mogą stać się w przyszłości przydatnym i niezbędnym narzędziem wspomagającym racjonalne funkcjonowanie i eksploatację sieci i całego przedsiębiorstwa dla operatorów sieci wodociągowej oraz decydentów wodociągów. Dotychczas takie złożone i

zintegrowane systemy IT z powodu ich bardzo wysokich kosztów opracowywane były głównie w jednostkach naukowych i nie są używane w polskich wodociągach [10].

Na trzecim etapie badań problem dotyczy nie ustalenia awarii sieci wodociągowej, które już miały miejsce, ale polega na rozpoznaniu potencjalnego ryzyka awarii i wyeliminowania go przez właściwe techniczne przeciwdziałania. Analiza przedstawiona w pracy należy do drugiego etapu badań i polega na wykrywaniu i lokalizacji uszkodzeń sieci wodociągowej za pomocą systemu IT obejmującego GIS, SCADA, model hydrauliczny sieci wodociągowej i sieci neuronowe symulujące sieć przy użyciu jej hydraulicznego modelu.

Celem badań jest potwierdzenie przydatności sieci neuronowych poprzez modelowanie sieci wodociągowych i wykrywanie jej awarii. Pozytywny wynik pozwoli na zawarcie tego narzędzia do modelowania w systemie IT opracowanym jako integralny moduł systemu. Przykładowe obliczenia zostały wykonane z prawdziwych danych pochodzących z komunalnych wodociągów w Rzeszowie [5], a zastosowanymi sieciami neuronowymi są sieci neuronowe MLP i Kohonena [12].

Badania przeprowadzono za pomocą następującego algorytmu:

1. Wyznaczenie punktów wrażliwych sieci wodociągowej przy użyciu systemu do optymalnego planowania systemu monitoringu. System zawiera specjalnie wybrane punkty pomiarowe, które są najbardziej wrażliwe na zmiany przepływu i ciśnienia, które pojawiają się w sieci wodociągowej.
2. Skalibrowanie modelu hydraulicznego przy użyciu systemu monitoringu sieci.
3. Obliczenia hydrauliczne sieci wodociągowej przy jej standardowym obciążeniu w celu określenia standardowych rozkładów przepływów i ciśnień w wybranych punktach pomiarowych.
4. Symulacja wycieków wody w wybranych węzłach sieci wodociągowej przy użyciu modelu hydraulicznego do określenia rozkładów przepływów i ciśnień w punktach monitoringu wynikających z symulowanych awarii.
5. Utworzenie klasyfikatora lokalizującego awarie w postaci różnego typu sieci neuronowych i wybór klasyfikatora najlepszego według kryterium największej wrażliwości.

2.1. Wyznaczenie punktów wrażliwych sieci wodociągowej przy użyciu systemu monitoringu

W celu znalezienia najlepszych lokalizacji czujników w punktach pomiarowych systemu monitoringu wyznacza się tzw. punkty wrażliwe sieci wodociągowej przy użyciu algorytmu do planowania systemu monitoringu. Aby zmaksymalizować ilość informacji przekazywanych przez punkty pomiarowe, powinny one być tak wybrane, aby przekazywały wiedzę o stanie sieci nie tylko z danego punktu, ale również z jego dalszego otoczenia. Takie punkty charakteryzujące się dużą wrażliwością na zmiany stanu sieci nazwiemy punktami charakterystycznymi. Maksymalizacja przekazywanej informacji oznacza zwykle, w tradycyjnie planowanych systemach monitoringu, zwiększanie liczby punktów pomiarowych, a więc działanie przeciwne do opisanego powyżej. Dlatego należy podkreślić, że odpowiedni wybór stosunkowo niewielkiej liczby punktów charakterystycznych może być równoważny pod względem jakości i ilości przekazywanej wiedzy o sieci większej liczbie punktów zlokalizowanych w mniej wrażliwych miejscach sieci [8]. Wybór odpowiednich punktów pomiarowych do monitorowania sieci wodociągowej jest nietrywialnym zadaniem, do rozwiązania którego można stosować różne algorytmy obliczeniowe. Jednym z nich jest

algorytm podany przez R. Straubela i B. Holznagela w pracy [7]. W naszych badaniach została wykonana lista rankingowa wrażliwych punktów sieci wodociągowej w Rzeszowie przy użyciu algorytmu Straybela i Holznagela i w dalszych obliczeniach wybrano dwa zestawy punktów do planowanego systemu monitorowania: jeden zestaw z 10 najbardziej wrażliwymi punktami i drugi zestaw z 20 punktami pomiarowymi.

2.2. Wykonanie obliczeń hydraulicznych sieci za pomocą wybranego modelu hydraulicznego

Po zaplanowaniu systemu monitoringu mogła zostać automatycznie wykonana kalibracja modelu hydraulicznego sieci wodociągowej. Aby to zrobić została użyta multi-kryterialna optymalizacja, która użyła dwóch kryteriów dotyczących różnic między zmierzonymi i obliczonymi wartościami przepływów wody i ciśnień w punktach monitorowania [8]. Ze skalibrowanym hydraulicznym modelem symulacja mogła być wykonana dla sieci wodociągowej z jej standardowym obciążeniem (stan normalny) i z symulowanymi wyciekami wody z różnych węzłów (stany awarii). Zarejestrowane wartości przepływów i ciśnienia w punktach monitorowania dla normalnych i awaryjnych stanów sieci wodociągowej mogą być stosowane do przygotowania plików uczących sieci neuronowe, które służą z kolei jako modele do wykrywania i lokalizowania awarii w sieci wodociągowej. W celu uproszczenia badań zostały zgromadzone i zapamiętane tylko wartości przepływu wody.

Symulacji awarii wody w sieci wodociągowej dokonano w następujący sposób:

- powiększono kilka razy wartości przepływu wody w wybranych węzłach sieci,
- dla każdej zmiany wartości przepływu wody został obliczony za pomocą modelu hydraulicznego układ hydrauliczny sieci wodociągowej,
- zarejestrowano różnice w wartościach przepływu wody między normalnymi i awaryjnymi stanami sieci wodociągowej obserwowane w punktach monitorowania,
- zarejestrowano punkt monitorowania z najsilniejszą reakcją na symulowaną awarię.

2.3. Tworzenie klasyfikatora neuronowego lokalizującego awarie przy użyciu sieci MLP

Modele do lokalizacji awarii zostały utworzone przy użyciu sieci neuronowych typu MLP. Sieci MLP niezmiennie są najbardziej rozpowszechnionymi i uniwersalnymi sieciami neuronowymi stosowanymi w praktyce. Dobór liczby neuronów w warstwie wejściowej jest uwarunkowany wymiarem wektora danych x . Podobnie jest w warstwie wyjściowej, w której liczba neuronów równa się wymiarowi wektora zadanego d . Problemem pozostaje dobór warstw ukrytych i liczby neuronów w każdej warstwie. W ogólnym przypadku liczba warstw ukrytych może być dowolna, ze względu jednak na skrócenie czasu obliczeń stosuje się przeważnie jedną lub dwie warstwy ukryte. Model neuronu jest typu sigmoidalnego. Składa się on z elementu sumacyjnego, do którego dochodzą sygnały wejściowe x_1, x_2, \dots, x_N pomnożone przez przyporządkowane im wagi $w_{i1}, w_{i2}, \dots, w_{iN}$ oraz przez wartość w_{i0} zwaną progiem. Sygnał wyjściowy sumatora oznaczono u_i :

$$u_i = \sum w_{ij}x_j + w_{i0} \quad (1)$$

Sygnal ten jest podawany na blok realizujący nieliniową funkcję aktywacji $f(u_i)$. Funkcja ta przyjmuje postać sigmoidalną (zachowuje wtedy ciągłość i proces uczenia sieci jest łatwiejszy). Postać funkcji stanowi ciągłe przybliżenie funkcji skokowej:

$$f(u_i) = \frac{1}{1 + \exp(-\beta \cdot u_i)} \quad (2)$$

W obliczeniach przeprowadzono eksperymenty z sieciami neuronowymi o różnej strukturze i przy 1 warstwie ukrytej. Badana sieć wodociągowa składa się z 390 węzłów. Na pierwszym etapie obliczeń zostało wziętych pod uwagę 10 punktów pomiarowych i w 36 wybranych węzłach przeprowadzono symulację awarii sieci wodociągowej. W drugim etapie liczbę punktów pomiarowych podwyższono do 20 i liczba węzłów z wyciekami wody podniosła się do 44. Klasyfikator neuronalny został utworzony zgodnie z metodyką opisaną w [4]. Eksperymenty związane z tworzeniem modeli sieci neuronowych typu MLP były parametryzowane dwoma parametrami: liczbą neuronów w warstwie ukrytej oraz liczbą cykli uczących. W eksperymencie parametr liczba neuronów w warstwie ukrytej przyjmował wartości od 5 do 25, natomiast drugi parametr liczba cykli uczących przyjmował wartości: 200, 500 i 1000. Na wejścia sieci neuronowych podano wartości przepływu wody uzyskane przy typowym stanie obciążenia sieci wodociągowej oraz przy symulowaniu awarii. Wyjście sieci neuronowych pokazuje punkt pomiarowy z najsilniejszą reakcją na wyciek wody (rys. 1).

| 27 2740 | 28 2779 | 29 3028 | 30 6144 | 31 3587 | 32 3596 | 33 4138 | 34 4181 | 35 4250 | 36 4411 | 37 Punkt pomiarowy |
|-------------------|---------------|---------------|------------|------------|------------|------------|------------|------------|------------|-----------------------|
| -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 0 |
| -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 4 |
| -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 6 |
| -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 6 |
| -25,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 7 |
| -5,04 | -470,5 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 10 |
| -5,04 | -94,46 | -10,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 4 |

Rys. 1. Struktura pliku uczącego sieci MLP;
kolumny od 27 do 36 są wejściami, a kolumna 37 jest wyjściem sieci

Liczba przykładów przeznaczonych do określenia sieci neuronowych dla punktów pomiarowych wyniosła 304, a w drugim etapie dla 20 punktów pomiarowych 360. Pliki użyte do uczenia, testowania i walidacji wyniosły odpowiednio 70%, 15% i 15% przykładów.

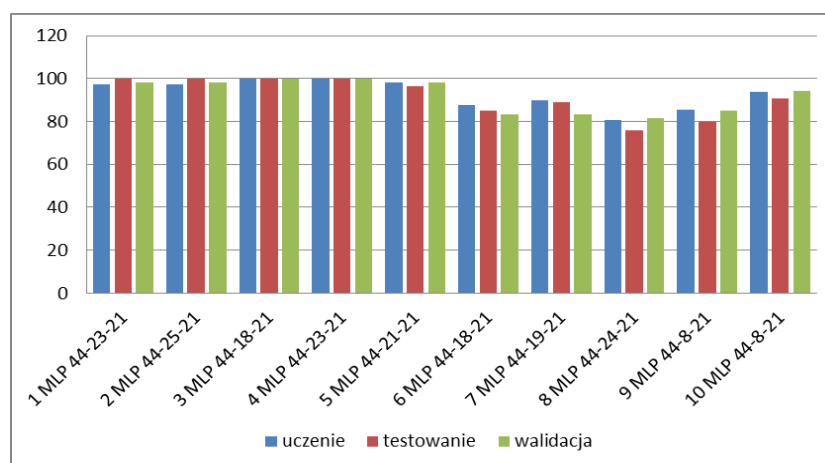
W tabelach 1 i 2 oraz na rysunku 2 można zobaczyć wyniki obliczeń dla 10 i 20 punktów pomiarowych. Model MLP 36-24-11 z 24 neuronami w warstwie ukrytej jest najlepszy dla 10 punktów pomiarowych (średniej jakości wyniosła 97,58%), a modele MLP 44-18-21 i MLP 44-23-21 są najlepsze dla 20 punktów pomiarowych (ze średnią jakością 100%).

Tabela 1. Parametry sieci MLP dla 10 punktów pomiarowych; jakość podana jest w %

| Nr | Nazwa sieci | Jakość uczenia | Jakość testowania | Jakość walidacji |
|----------|---------------------|-----------------|-------------------|------------------|
| 1 | MLP 36-8-11 | 88,31776 | 95,55556 | 88,88889 |
| 2 | MLP 36-15-11 | 97,66355 | 97,77778 | 95,55556 |
| 3 | MLP 36-22-11 | 94,39252 | 97,77778 | 93,33333 |
| 4 | MLP 36-19-11 | 97,66355 | 97,77778 | 95,55556 |
| 5 | MLP 36-21-11 | 94,39252 | 97,77778 | 93,33333 |
| 6 | MLP 36-24-11 | 97,19626 | 97,77778 | 97,77778 |
| 7 | MLP 36-23-11 | 97,66355 | 97,77778 | 95,55556 |

Tabela 2. Parametry sieci MLP dla 20 punktów pomiarowych; jakość podana jest w %

| Nr | Nazwa sieci | Jakość uczenia | Jakość testowania | Jakość walidacji |
|----------|---------------------|-----------------|-------------------|------------------|
| 1 | MLP 44-23-21 | 97,2222 | 100,0000 | 98,1481 |
| 2 | MLP 44-25-21 | 97,2222 | 100,0000 | 98,1481 |
| 3 | MLP 44-18-21 | 100,0000 | 100,0000 | 100,0000 |
| 4 | MLP 44-23-21 | 100,0000 | 100,0000 | 100,0000 |
| 5 | MLP 44-21-21 | 98,0159 | 96,2963 | 98,1481 |
| 6 | MLP 44-18-21 | 87,6984 | 85,1852 | 83,3333 |
| 7 | MLP 44-19-21 | 90,0794 | 88,8889 | 83,3333 |
| 8 | MLP 44-24-21 | 80,5556 | 75,9259 | 81,4815 |
| 9 | MLP 44-8-21 | 85,3175 | 79,6296 | 85,1852 |
| 10 | MLP 44-8-21 | 93,6508 | 90,7407 | 94,4444 |



Rys. 2. Wykres jakości sieci neuronowych typu MLP – 20 punktów pomiarowych [%]

Tabela 3 przedstawia wyniki obliczeń dla modeli MLP opracowanych dla 20 punktów pomiarowych dla plików z nowymi danymi, które zostały przeprowadzone w celu sprawdzenia poprawności modeli w nowych warunkach obliczeniowych. Można zauważyć, że najlepsze modele MLP 44-18-21 i MLP 44-23-21, które zostały otrzymane z poprzednich danych potwierdzają taką samą skuteczność dla tych nowych warunków.

Table 3. Wyniki obliczeń sieci MLP dla pliku z nowymi danymi (20 punktów pomiarowych)

| Rezultaty przewidywania dla nowych danych na wejściu sieci MLP | | | | | | | | | | | |
|--|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|---------------|----------------|------|
| Nr | 1 MLP 44-23-21 | 2 MLP 44-25-21 | 3 MLP 44-18-21 | 4 MLP 44-23-21 | 5 MLP 44-21-21 | 6 MLP 44-18-21 | 7 MLP 44-19-21 | 8 MLP 44-24-21 | 9 MLP 44-8-21 | 10 MLP 44-8-21 | Wzór |
| 1 | 4 | 4 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 |
| 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 3 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 4 | 15 | 15 | 15 | 15 | 15 | 19 | 15 | 15 | 15 | 15 | 15 |
| 5 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| 6 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 20 | 13 |
| 7 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 20 | 4 |
| 8 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 | 16 |
| 9 | 12 | 12 | 12 | 12 | 12 | 2 | 12 | 12 | 12 | 12 | 12 |
| 10 | 7 | 7 | 7 | 7 | 7 | 15 | 7 | 13 | 1 | 7 | 7 |
| 11 | 9 | 9 | 9 | 9 | 9 | 9 | 7 | 13 | 6 | 9 | 9 |
| 12 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 0 | 20 | 20 | 20 |
| 13 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 13 | 1 | 13 | 1 |
| 14 | 2 | 2 | 2 | 2 | 0 | 2 | 2 | 15 | 0 | 2 | 2 |
| 15 | 17 | 17 | 17 | 17 | 17 | 15 | 5 | 13 | 17 | 17 | 17 |
| 16 | 5 | 5 | 5 | 5 | 5 | 10 | 5 | 15 | 16 | 5 | 5 |
| 17 | 6 | 6 | 6 | 6 | 6 | 19 | 6 | 13 | 6 | 14 | 6 |
| 18 | 14 | 14 | 14 | 14 | 14 | 2 | 10 | 13 | 6 | 14 | 14 |
| 19 | 18 | 18 | 18 | 18 | 18 | 2 | 19 | 15 | 18 | 14 | 18 |
| 20 | 19 | 19 | 19 | 19 | 19 | 19 | 19 | 13 | 7 | 19 | 19 |
| 21 | 10 | 10 | 10 | 10 | 0 | 10 | 10 | 13 | 0 | 10 | 10 |
| 22 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 13 | 0 | 1 | 13 |

| | | | | | | | | | | | |
|------------------------------------|-------|-------|---------------|---------------|-------|-------|-------|-------|-------|-------|--|
| Liczba poprawnych klasyfikacji | 21 | 21 | 22 | 22 | 20 | 13 | 17 | 10 | 14 | 16 | |
| Liczba błędnych klasyfikacji | 1 | 1 | 0 | 0 | 2 | 9 | 5 | 12 | 8 | 6 | |
| Liczba poprawnych klasyfikacji w % | 98,46 | 98,46 | 100,00 | 100,00 | 97,49 | 85,41 | 87,43 | 79,32 | 83,38 | 92,95 | |

2.4. Tworzenie klasyfikatora neuronowego lokalizującego awarie przy użyciu sieci Kohonena

Modele lokalizacji awarii w sieci wodociągowej są obecnie tworzone przy użyciu sieci neuronowych Kohonena. Sieć Kohonena jest podstawową siecią określaną mianem sieci samoorganizującej się (Self-Organizing Map - SOM) lub samoorganizującym się odwzorowaniem cech (Self-Organizing Feature Map - SOFM) [12]. W sieci tej mamy do czynienia z uczeniem konkurencyjnym. Oznacza to, że użytkownik ma do dyspozycji jedynie wzorce wejściowe, nie posiada natomiast żadnych wzorców wyjścia. Zadaniem sieci w trakcie procesu uczenia jest wytworzenie takich wzorców. Sieć ma na celu utworzenie takiej struktury, która w najlepszy sposób będzie odwzorowywała zależności w przestrzeni wektorów wejściowych. Sieć jest zwykle jednokierunkowa, w której każdy neuron jest połączony ze wszystkimi składowymi N-wymiarowego wektora wejściowego x . Wagi połączeń neuronów tworzą wektor $w_i = [w_{i1}, w_{i2}, \dots, w_{iN}]^T$. Wektor sygnałów wejściowych x jest przed procesem uczenia normalizowany $\|x\|=1$, co można zapisać jako:

$$x_i = \frac{x_i}{\sqrt{\sum_{v=1}^N (x_v)^2}} \quad (3)$$

Po pobudzeniu sieci wektorem wejściowym x , podczas współzawodnictwa wygrywa neuron, którego wagi najmniej różnią się od odpowiednich składowych tego wektora. Zwycięzca, neuron w -ty, spełnia relację:

$$d(x, w_w) = \min_{1 \leq i \leq n} d(x, w_i) \quad (4)$$

gdzie $d(x, w)$ oznacza odległość w sensie wybranej metryki między wektorem n a wektorem w .

Wokół i -tego neuronu przyjmuje się topologiczne sąsiedztwo $G(i, x)$.

W klasycznym algorytmie Kohonena funkcja $G(i, x)$ jest definiowana następująco:

$$G(i, x) = \begin{cases} 1 & \text{dla } d(i, w) \leq R \\ 0 & \text{dla } d(i, w) > R \end{cases} \quad (5)$$

gdzie $d(i, w)$ oznacza odległość euklidesową między neuronem zwyciężskim w_i a i -tym neuronem, R – promieniem sąsiedztwa malejącym do 0.

Podczas tworzenia modeli przy pomocy sieci Kohonena, te sieci były parametryzowane dwoma parametrami: liczbą neuronów w warstwie topologicznej sieci oraz liczbą cykli uczenia. Pierwszy parametr przyjmował wartości 2x8, 5x5, 10x10, 15x15, a drugi wartości od 1000 do 20000.

Podobnie jak we wcześniejszych badaniach wejścia sieci są wartościami przepływu w normalnym i awaryjnym stanie sieci wodociągowej. Wybrane zostały węzły sieci dla przeprowadzenia symulacji awarii sieci. Ostatnim wejściem sieci jest najbardziej wrażliwy punkt pomiarowy. W sieciach tego rodzaju nie istnieje wyjście. Przeprowadzono dwa etapy badań dotyczących różnej liczby punktów pomiarowych (10 i 20) oraz węzłów sieci wodociągowej (36 i 44), które zostały wybrane do symulacji awarii.

Na rysunku 3 została pokazana struktura pliku uczącego, gdzie wszystkie kolumny oznaczają wejście sieci neuronowej. Liczba przykładów przeznaczonych do określenia sieci neuronowych dla punktów pomiarowych wyniosła 304, a w drugim etapie dla 20 punktów pomiarowych 360. Pliki użyte do uczenia, testowania i walidacji wyniosły odpowiednio 70%, 15% i 15% przykładów.

| 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 |
|--------|-------|--------|-------|---------|--------|-------|------|-------|-------|------|-------|--------|-----------------|
| 2158 | 2186 | 2447 | 2740 | 2779 | 3028 | 6144 | 3587 | 3596 | 4138 | 4181 | 4250 | 4411 | Punkt pomiarowy |
| -25,74 | -4,2 | -4,83 | -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 4 |
| -5,74 | -20,2 | -4,83 | -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 6 |
| -5,74 | -4,2 | -20,83 | -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 6 |
| -5,74 | -4,2 | -4,83 | 25,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 7 |
| -5,74 | -4,2 | -4,83 | -5,04 | -470,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 10 |
| -5,74 | -4,2 | -4,83 | -5,04 | -94,46 | -10,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 4 |
| -5,74 | -4,2 | -4,83 | -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -20,76 | 2 |
| -5,74 | -4,2 | -4,83 | -5,04 | -94,46 | -2,71 | -2,79 | -3,1 | -8,43 | -6,31 | -2,2 | -1,69 | -4,76 | 0 |

Rys. 3. Struktura pliku uczącego sieci Kohonena

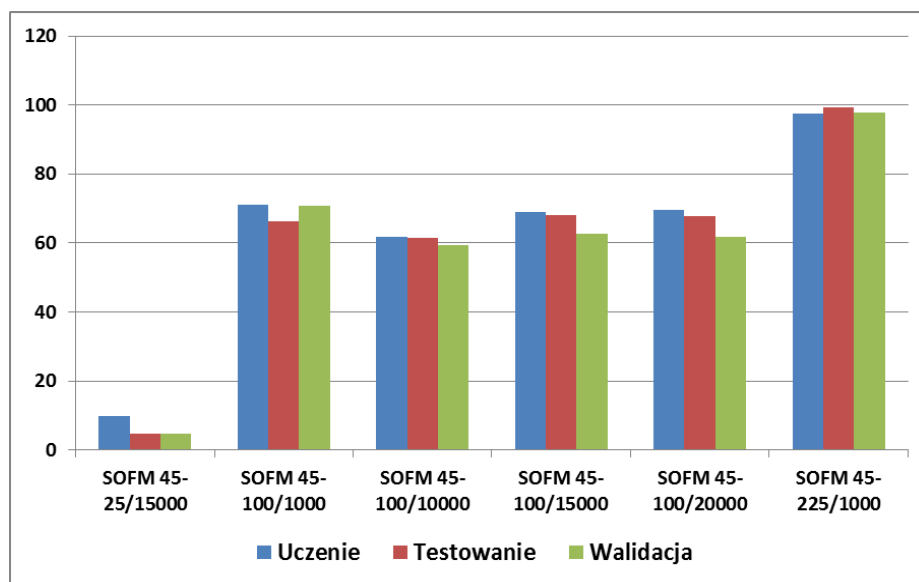
W tabelach 4 i 5 oraz na rysunku 4 można zobaczyć wyniki obliczeń dla 10 i 20 punktów pomiarowych. Model SOFM 37-100/100 jest najlepszy dla 10 punktów pomiarowych (średnia jakość wyniosła 81,38%), a model SOFM 45-225/1000 jest najlepszy dla 20 punktów pomiarowych (ze średnią jakością 98,16%).

Tabela 4. Parametry sieci Kohonena dla 10 punktów pomiarowych

| Nr | Nazwa sieci | Jakość uczenia | Jakość testowania | Jakość walidacji |
|----------|-------------------------|----------------|-------------------|------------------|
| 1 | SOFT 37-16/1000 | 81,9384 | 14,3036 | 7,4902 |
| 2 | SOFT 37-16/20000 | 20,0225 | 20,9363 | 11,9975 |
| 3 | SOFT 37-25/1000 | 20,3248 | 17,1003 | 12,9749 |
| 4 | SOFT 37-100/1000 | 80,8903 | 79,7651 | 83,4712 |
| 5 | SOFT 37-100/10000 | 82,7746 | 76,2847 | 78,0958 |
| 6 | SOFT 37-100/20000 | 83,1150 | 73,5965 | 79,7852 |

Tabela 5. Parametry sieci Kohonena dla 20 punktów pomiarowych

| Nr | Nazwa sieci | Jakość uczenia | Jakość testowania | Jakość walidacji |
|----------|-------------------------|----------------|-------------------|------------------|
| 1 | SOFT 45-25/15000 | 9,7870 | 4,7473 | 4,7570 |
| 2 | SOFT 45-100/1000 | 71,1903 | 66,3754 | 70,7134 |
| 3 | SOFT 45-100/10000 | 61,7010 | 61,5446 | 59,4152 |
| 4 | SOFT 45-100/15000 | 69,1089 | 68,1126 | 62,5354 |
| 5 | SOFT 45-100/20000 | 69,4488 | 67,6256 | 61,6244 |
| 6 | SOFT 45-225/1000 | 97,3839 | 99,1899 | 97,9011 |



Rys. 4. Wykres jakości sieci neuronowych typu Kohonena – 20 punktów pomiarowych [%]

Tabela 6 przedstawia wyniki obliczeń dla modeli Kohonena opracowanych dla 20 punktów pomiarowych dla plików z nowymi danymi, które zostały przeprowadzone w celu sprawdzenia poprawności modeli w nowych warunkach obliczeniowych. Wyniki potwierdzają po raz kolejny, że model SOFM 45-225/1000 jest najlepszy i jego jakość jest równa 98,16%, podobnie jak w poprzednich eksperymentach.

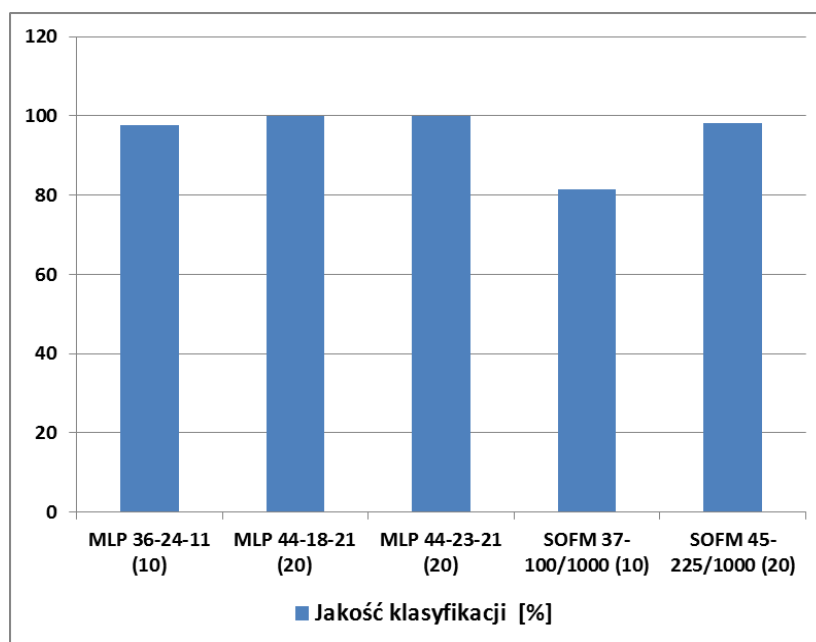
Tabela 6. Wyniki obliczeń sieci Kohonena dla pliku z nowymi danymi (20 punktów pomiarowych)

| Rezultaty przewidywania dla nowych danych na wejściu sieci Kohonena | | | | | | | |
|---|--------------------|--------------------|---------------------|---------------------|---------------------|--------------------|-----------|
| Nr | 1 SOFT 45-25/15000 | 2 SOFT 45-100/1000 | 3 SOFT 45-100/10000 | 4 SOFT 45-100/15000 | 5 SOFT 45-100/20000 | 6 SOFT 45-225/1000 | Wzór |
| 1 | 15 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 3 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 4 | 0 | 15 | 7 | 15 | 9 | 7 | 15 |
| 5 | 11 | 11 | 11 | 11 | 11 | 11 | 11 |
| 6 | 16 | 13 | 13 | 13 | 13 | 13 | 13 |
| 7 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 8 | 13 | 16 | 16 | 16 | 16 | 16 | 16 |
| 9 | 7 | 12 | 12 | 12 | 12 | 12 | 12 |
| 10 | 12 | 7 | 15 | 7 | 7 | 15 | 7 |
| 11 | 1 | 9 | 20 | 9 | 15 | 9 | 9 |
| 12 | 20 | 20 | 9 | 20 | 20 | 20 | 20 |
| 13 | 9 | 2 | 1 | 1 | 1 | 1 | 1 |
| 14 | 17 | 1 | 2 | 2 | 2 | 2 | 2 |
| 15 | 2 | 17 | 5 | 17 | 5 | 17 | 17 |
| 16 | 6 | 6 | 17 | 6 | 17 | 5 | 5 |
| 17 | 5 | 5 | 14 | 5 | 18 | 6 | 6 |
| 18 | 18 | 14 | 6 | 19 | 14 | 14 | 14 |
| 19 | 14 | 19 | 19 | 18 | 6 | 18 | 18 |
| 20 | 10 | 18 | 18 | 14 | 10 | 19 | 19 |
| 21 | 19 | 9 | 8 | 9 | 19 | 10 | 10 |
| 22 | 11 | 13 | 13 | 13 | 13 | 13 | 13 |
| Liczba poprawnych klasyfikacji | 5 | 15 | 11 | 17 | 14 | 21 | |

| | | | | | | | |
|------------------------------------|-------|-------|-------|-------|-------|--------------|--|
| Liczba błędnych klasyfikacji | 17 | 7 | 11 | 5 | 8 | 1 | |
| Liczba poprawnych klasyfikacji w % | 22,73 | 68,18 | 50,00 | 77,27 | 63,64 | 98,16 | |

2.5. Wybór najlepszego klasyfikatora - porównanie sieci MLP i Kohonena

W tabeli 7 i na rysunku 5 przedstawiono wyniki porównania sieci MLP i Kohonena. Wyniki pokazują przewagę modeli MLP w stosunku do Kohonena podczas wykrywania i lokalizowania awarii sieci wodociągowej. Sieci MLP posiadają generalnie krótsze czasy uczenia się i dają lepszą klasyfikację awarii w sieci wodociągowej.



Rys. 5. Porównanie sieci MLP i Kohonena

Tabela 7. Porównanie sieci MLP i Kohonena

| Nr | Nazwa sieci neuronowej | Liczba punktów pomiarowych | Jakość klasyfikacji [%] |
|----|------------------------|----------------------------|-------------------------|
| 1 | MLP 36-24-11 | 10 | 97,58 |
| 2 | MLP 44-18-21 | 20 | 100,00 |
| 3 | MLP 44-23-21 | 20 | 100,00 |

| | | | |
|---|------------------|----|-------|
| 4 | SOFM 37-100/1000 | 10 | 81,38 |
| 5 | SOFM 45-225/1000 | 20 | 98,16 |

3. Podsumowanie

Uzyskane wyniki obliczeń wskazują przydatność sieci neuronowych do rozwiązania takich skomplikowanych problemów jak wykrywanie awarii w sieciach wodociągowych i ich lokalizacji. Oznacza to, że neuronalne modele mogą być stosowane jako skuteczne narzędzia do zarządzania wodociągami w postaci integralnych elementów w systemach komputerowego wspomaganie (IT).

Bibliografia

1. Billman L, Isermann R. Leak detection methods for pipelines. *Automatica* 1987; 23: 381-385.
2. Farmani R, Ingeduld P, Savic D, Walters G, Svitak Z, Berka J. Real-time modeling of a major water supply system. *International Conference on Computing and Control for the Water Industry* 2007; 160(2): 103-108.
3. Kowalczyk Z, Gunawickrama K. Detecting and locating leaks in transmission pipelines. *Fault Diagnosis: Models, Artificial Intelligence, Applications*. Berlin: Springer-Verlag, 2004; 822-864.
4. Rojek I. Opracowanie algorytmów lokalizacji awarii sieci wodociągowej w postaci sieci neuronowych. Raport z projektu badawczego Nr 14-0011-10/2010. Warszawa: IBS PAN, 2012.
5. Rojek I, Studziński J. Algorytm lokalizacji awarii w sieci wodociągowej przy użyciu sieci neuronowej. *Studia i Materiały Polskiego Stowarzyszenia Zarządzania Wiedzą* 2011; 8: 146-156.
6. Smarter Water Management. IBM, 2009; <http://www.ibm.com/ibm/ideasfromibm/us/smartplanet/topics/water-management/>.
7. Straubel R, Holznagel B. Mehrkriteriale Optimierung für Planung und Steuerung von Trink- und Abwasser-Verbundsystemen. *Wasser Abwasser* 1999; 140(3): 191-196.
8. Studziński J. Rechnerunterstützte Entscheidungshilfe für kommunale Wasserwerke mittels mathematischer Modelle. *Krigingsapproximation und Optimierung. Modellierung und Simulation von Ökosystemen*. Workshop Kölpinsee 2011.
9. Studziński J. Decisions making systems for communal water networks and wastewater treatment plants. *Modeling Concepts and Decision Support in Environmental Systems*. Warsaw: IBS PAN, 2006; 49.
10. Studziński J. Innowacje XXI wieku – nowoczesne techniki informacyjne wspomaganie zarządzania w przedsiębiorstwach sieciowych. *Innowacyjne Mazowsze*. Płock: Wydawnictwo of SWPW, 2010.
11. Rojek I. Wspomaganie procesów podejmowania decyzji i sterowania w systemach o różnej skali złożoności z udziałem metod sztucznej inteligencji. Bydgoszcz: Wydawnictwo Uniwersytetu Kazimierza Wielkiego, 2010.
12. Tadeusiewicz R. Odkrywanie właściwości sieci neuronowych przy użyciu programów w języku C#. Kraków: Polska Akademia Umiejętności, 2007.

13. Turkowski M, Bratek A, Słowikowski M. Methods and systems of leak detection in long range pipelines. *Journal of Automation, Mobile Robotics & Intelligent Systems* 2007; 1: 39-46.
14. Wei S, Gnauck A. Game Theory Based Water Quality Models for Reservoir Management. *Environmental Informatics and Systems Research. 21st Conference on Informatics for Environmental Protection ENVIROINFO*. Aachen: Shaker Verlag, 2007; 1: 363-370.
15. Wyczółkowski R, Moczulski W. Concept of intelligent monitoring of local water supply system. *Artificial Intelligence Methods AI-METH*. Gliwice: Silesian University of Technology, 2005; 147-148.
16. Wyczółkowski R, Wysogład B. An optimization of heuristic model of water supply network. *Computer Assisted Mechanics and Engineering Science CAMES* 2007; 14: 767-776.

Izabela Rojek

Instytut Mechaniki i Informatyki Stosowanej
Uniwersytet Kazimierza Wielkiego
Chodkiewicza 30, 85-064 Bydgoszcz, Polska
E-mail: izarojek@ukw.edu.pl

Jan Studziński,

Instytut Badań Systemowych Polskiej Akademii Nauk
Newelska 6, 01-447 Warszawa, Polska
E-mail: studzins@ibspan.waw.pl