

Adam WIŚNIEWSKI, Alicja KILIŃSKA-WIŚNIEWSKA  
Politechnika Koszalińska, Wydział Elektroniki i Informatyki  
E-mail: pl.adam.wisniewski@gmail.com, alicja.kilinska.wisniewska@gmail.com

## Strumieniowe bazy danych w sztucznych sieciach neuronowych

### 1. Wstęp

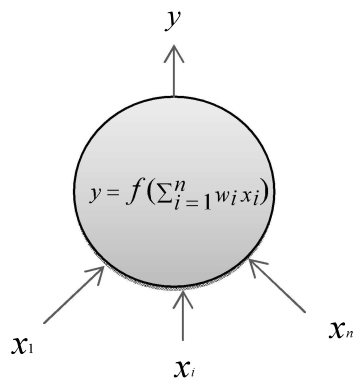
W dzisiejszych czasach można zaobserwować zjawisko, które polega na coraz większym wykorzystywaniu maszyn, czy też mówiąc bardziej szczegółowo, inteligentnych maszyn do zautomatyzowania wielu procesów w różnych gałęziach nauki, np. w dziedzinie wspomaganie decyzji, finansowych lub przewidywania zjawisk pogodowych. Wykorzystanie nowego rodzaju danych jakim są strumienie danych może w znacznym stopniu zwiększyć wydajność obecnych systemów monitoringu. Rozbudowując taki system o sztuczne sieci neuronowe możemy otrzymać model, który automatycznie zidentyfikuje problem na podstawie danych wcześniej przeanalizowanych przez strumieniową bazę danych. Artykuł ten jest wstępem do dalszej części badań nad strumieniowymi bazami danych oraz autonomicznymi systemami monitoringu.

### 2. Strumieniowe bazy danych

W roku 2001 dostrzeżono potrzebę znalezienia alternatywy względem ustandaryzowanej już relacyjnej bazy danych. Okazało się, że zastosowanie relacyjnej bazy danych w niektórych przypadkach jest niewystarczające, w szczególności dotyczyło to systemów monitorujących. Zaczęto więc poszukiwać innych rozwiązań, efektem tego było opracowanie nowego modelu danych jakim jest strumień danych – teoretycznie nieograniczony w czasie zbiór danych. Na rynku istnieje jedno komercyjne rozwiązanie strumieniowej bazy danych, system STREAMBASE. W rozwiązaniu tym pominięto jednak implementację deklaratywnego języka zapytań, zapytania realizowane są za pomocą interfejsu graficznego, który z góry narzuca plan realizacji zapytania. Odmiennie jest w przypadku projektu STREAM, który został opracowany na Uniwersytecie Stanford. w tym wypadku opracowano deklaratywny język zapytań jakim jest CQL, jest to rozszerzenie standardowego języka SQL o moduł strumieniowy. Podstawową różnicą między relacyjnymi bazami danych a strumieniowymi bazami danych jest realizacja zapytania. w relacyjnej bazie danych użytkownik wykonuje zapytanie, po czym zwracana jest odpowiedź jako wynik wykonanego zapytania. W strumieniowych bazach danych rejestruje się strumień danych, który pochodzi z zewnętrznych systemów monitoringu (np. czujników, kamer) w losowych odstępach czasu i jest dokonywana analiza zebranych danych. Można powiedzieć, że analiza danych wykonywana jest praktycznie w czasie rzeczywistym.

### 3. Sztuczne sieci neuronowe

W połowie XX wieku zaczęto zastanawiać się w jaki sposób ludzki mózg dokonuje skomplikowanych operacji w oparciu o często niekompletne, nieściśle i nie do końca prawdziwe informacje. Kluczem do rozwiązania zagadki okazał się neuron, a konkretniej sieć neuronów połączona i współpracująca między sobą. Na tej podstawie opracowano model sztucznego neuronu (Rys. 1.).



Na wejściu  $x_1...x_n$  podawane są sygnały wejściowe, które mogą pochodzić ze świata zewnętrznego lub jako sygnał wyjściowy innego neuronu, następnie do sygnałów wejściowych odpowiednio przypisywane są wagi  $w_1...w_n$ . Uczenie sztucznej sieci neuronowej (SSN) polega na takiej modyfikacji wag przez trenera, aby suma iloczynów sygnałów wejściowych i wag dała oczekiwaną odpowiedź funkcji aktywacji (1) w odniesieniu do sygnałów wejściowych.

Rys. 1. Budowa sztucznego neuronu

Fig. 1. Construction of artificial neuron

Na wejściu  $x_1...x_n$  podawane są sygnały wejściowe, które mogą pochodzić ze świata zewnętrznego lub jako sygnał wyjściowy innego neuronu, następnie do sygnałów wejściowych odpowiednio przypisywane są wagi  $w_1...w_n$ . Uczenie sztucznej sieci neuronowej (SSN) polega na takiej modyfikacji wag przez trenera, aby suma iloczynów sygnałów wejściowych i wag dała oczekiwaną odpowiedź funkcji aktywacji (1) w odniesieniu do sygnałów wejściowych.

$$y = f\left(\sum_{i=1}^n w_i x_i\right) \quad (1)$$

Naturalnie pojedynczy neuron nie ma praktycznego zastosowania, dopiero stworzenie sieci neuronów pozwala na rozwiązywanie problemów, których rozwiązanie w tradycyjny sposób byłoby niemożliwe lub zbyt czasochłonne.

### 4. Sieć Kohonena

Sieć Kohonena jest nazwana przez swojego autora samoorganizującym odwzorowaniem lub samoorganizującym odwzorowaniem cech. w sieci tej proces uczenia nazywany jest uczeniem konkurencyjnym. Jej zadaniem jest stworzenie struktury, która najlepiej będzie odwzorowywać zależności w przestrzeni wektorów wejściowych. Sieć z reguły jest jednokierunkowa, każda składowa  $n$ -wymiarowego wektora sygnału wejściowego jest połączona z każdym neuronem, wagi połączeń neuronów tworzą wektor  $w_i [w_{i1}, w_{i2}, \dots, w_{in}]$ . W trakcie uczenia się sieci samoorganizującej na wejście każdego neuronu podawany jest  $N$ -wymiarowy sygnał  $x$  ze zbioru wzorców uczących. Sygnał wejściowy przed

procesem uczenia jest normalizowany, można to przedstawić za pomocą wzoru (2) gdzie  $N$  jest liczbą wejść.

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

W wyniku pobudzenia wyłaniany jest jeden neuron tzw. zwycięzca, którego wagi najmniej różnią się od składowych wektora wejściowego. Neuron zwycięzca spełnia relację (3):

$$d(x, w_w) = \min_{i=1,2..n} (d(x, w_i)) \quad (3)$$

Wokół zwycięskiego neuronu tworzony jest promień sąsiedztwa  $Sw(n)$ , którego promień maleje w czasie. Kolejnym krokiem jest adaptacja neuronu zwycięskiego oraz neuronów sąsiadujących według reguły Kohonena (4):

$$w_i(t) = w_i(t-1) + \eta \cdot g(i, x) \cdot (x - w_i(t-1)) \quad (4)$$

gdzie:

$\eta$  - współczynnik uczenia,  $g(i, x)$  - funkcja sąsiedztwa.

Istnieją dwa typy funkcji sąsiedztwa: sąsiedztwo prostokątne, oraz sąsiedztwo gaussowskie. W artykule przedstawiono funkcję sąsiedztwa gaussowskiego, gdyż jest ona lepsza i prowadzi do uzyskania lepszych wyników i organizacji sieci. Stopień adaptacji wag w sąsiedztwie gaussowskim można opisać wzorem (5):

$$G(i, x) = \exp\left(-\frac{d^2(i, w)}{2\lambda^2}\right) \quad (5)$$

gdzie:  $i$  -  $i$ -ty neuron w macierzy neuronów,  $w$  - aktualny zwycięzca,  $d(i, w)$  - zastosowana miara odległości między neuronami,  $\lambda$  - odchylenie standardowe.

Po zainicjowaniu sieci wektorem wejścia, ustaleniu stopnia adaptacji wag ustalany jest licznik czasu  $t$ . Każda następna iteracja będzie wykonywana dopóki  $t < t_{max}$  oraz zmieniła koncentrację neuronów w sieci Kohonena tworząc w efekcie końcowym mapę topologiczną danych wyjściowych.

## 5. Strumieniowe bazy danych i sztuczne sieci neuronowe (sieć Kohonena)

Głównym celem artykułu jest zaprezentowanie koncepcji wspomagania systemów monitoringu przez sztuczne sieci neuronowe. Takie połączenie miałyby na celu usprawnić identyfikację problemów, które mogłyby wystąpić w przedmiocie monitorowania. Jako przykład posłuży monitoring dróg miejskich.

Należy wyobrazić sobie drogę, po której w sposób mierzalny oraz definiowalny poruszają się pojazdy. Strumieniowa baza danych w określonych jednostkach czasu analizuje otrzymane przez system czujników i kamer informacje, np. sprawdza jaka średnia ilość pojazdów przejeżdża w ciągu 10 min. w odstępach czasowych co 15 min. Jeżeli w mierzonej aktualnie próbie (strumieniu danych) ilość pojazdów nie zgadza się ze średnią ilością pojazdów prze-

jeżdżających w poprzednich próbkach, można uznać to za anomalię. Strumieniowa baza danych wysyła do systemu sztucznych sieci neuronowych, w tym przypadku sieci Kohonena, próbkę (w postaci wyjściowego strumienia danych jako wektor danych wejściowych), która wykazała anomalię do analizy. Sztuczna sieć neuronowa na podstawie analizy przyporządkowuje określoną anomalię do miejsc posiadających już określoną wartość w topologii swojej sieci. Wartościami tymi może być np.: kolizja, korek, wypadek, bądź też inny, niezdefiniowany przypadek, który sieć neuronowa będzie potrafiła zidentyfikować na podstawie posiadanych wzorców.

Oczywiście taki efekt realizacji zadania można uzyskać w sposób tradycyjny, tzn. zastosować aplikację, która odbierałaby dane z systemu kamer oraz czujników i zapisywała informacje w bazie danych. Następnie porównywała zapisane w bazie danych informacje, a w momencie wystąpienia anomalii przesyłała próbkę, która zawiera anomalie do analizy sztucznej sieci neuronowej. Taki model jednak jest bardziej skomplikowany, wymaga napisania przez zespół programistów odrębnej aplikacji oraz jest mniej wydajny gdyż proces wykrywania takiej anomalii jest bardziej złożony.

## 6. Podsumowanie

Reasumując sztuczne sieci neuronowe w postaci sieci Kohonena oraz strumieniowe bazy danych mogą stanowić potężne narzędzie w diagnostyce, systemach monitoringu i wszędzie tam, gdzie informacja musi być analizowana na bieżąco, a zdarzenia niepożądane wykrywane i niwelowane w sposób zautomatyzowany. Dzięki sieciom neuronowym system jest w stanie zdefiniować zagrożenie nawet wówczas, kiedy nie jest on z góry zdefiniowany, a wykazuje podobieństwa do już znanych sieci neuronowej problemom.

## Literatura

1. <http://widera.com.pl/michal-mainmenu-2/strumieniowe.html>
2. <http://mimuw.edu.pl/delta/artykuly/delta0906/meta.pdf>
3. [http://books.google.pl/books?id=e4igHzyf078C&printsec=frontcover&hl=pl&source=gs\\_b\\_s\\_ge\\_summary\\_r#v=openpage&q&f=false](http://books.google.pl/books?id=e4igHzyf078C&printsec=frontcover&hl=pl&source=gs_b_s_ge_summary_r#v=openpage&q&f=false)

## Streszczenie

Artykuł ma na celu przedstawienie idei wykorzystanie strumieniowych baz danych jako wektor sygnału wejściowego w sztucznych sieci neuronowych, w szczególności dotyczący samoorganizującej się sztucznej sieci neuronowej zwanej siecią Kohonena. Oraz wykorzystanie tej idei w systemach monitoringu.

## Stream databases in artificial neural networks

Article presents the idea to use the streaming databases as a vector of the input signal in artificial neural networks, in particularly for self-organizing artificial neural network called Kohonen network. And the use of this idea in the monitoring systems.