

Systemy uczące się

M. Krzyśko, W. Wołyński, T. Górecki, M.
Skorzybut

Dostępna do: 01.06.2021

7.4. Sieci rekurencyjne

Dotychczas opisywane sieci charakteryzowały się jednokierunkowym przepływem sygnałów od wejścia do wyjścia. W sieciach rekurencyjnych istnieją sprzężenia zwrotne między wejściem a wyjściem. Zależności dynamiczne jakie panują w sieci są widoczne na każdym etapie działania. Zmiana stanu jednego neuronu przenosi się przez masowe sprzężenie zwrotne na całą sieć, wywołując stan przejściowy, kończący się określonym stanem ustalonym, na ogół innym niż stan poprzedni. Przy masowym sprzężeniu zwrotnym pobudzeniami dla neuronów są sygnały wejściowe innych neuronów. Proces ustalania się sygnałów wyjściowych sieci rekurencyjnych jest procesem dynamicznym. Biorąc pod uwagę nieliniowość funkcji aktywacji neuronów, jest to dynamika nieliniowa, stanowiąca istotną cechę tego rodzaju sieci.

7.4.1. Sieci Hopfielda

Jednym z podstawowych przedstawicieli sieci rekurencyjnych jest *sieć Hopfielda* (ze względu na pełnioną funkcję często nazywana *pamięcią autoasocjacyjną*). Została ona zaprezentowana w pracy Hopfielda z 1982 roku i była ona punktem zwrotnym w badaniach nad sieciami neuronowymi i przyczyniła się do intensywnych prac badawczych w tej dziedzinie. Sieć Hopfielda jest często podawana jako przykład pamięci skojarzeniowej lub jako układ do rozwiązywania zadań z zakresu optymalizacji. Strukturę sieci Hopfielda można opisać bardzo prosto – jest to układ wielu identycznych elementów połączonych metodą każdy z każdym. Jest zatem najczęściej rozpatrywana jako struktura jednowarstwowa. W odróżnieniu od sieci warstwowych typu perceptron sieć Hopfielda jest siecią rekurencyjną, gdzie neurony są wielokrotnie pobudzane w jednym cyklu rozpoznawania, co uzyskuje się przez pętle sprzężenia zwrotnego (charakterystyczne jest, że nie występuje sprzężenie zwrotne neuronu z samym sobą – waga $w_{ii} = 0$). Wagi w tego typu sieci są symetryczne, czyli $w_{ij} = w_{ji}$. Założenie o symetrii wag sprawia, że sieć ta zawsze się stabilizuje (czyli dla pewnego t zachodzi warunek $y_k(t+1) = y_k(t)$, dla każdego neuronu). Symetria jest wygodna z teoretycznego punktu widzenia, pozwala wprowadzić funkcję energii, jest natomiast

nierealistyczna z punktu widzenia biologicznego. Jeśli założymy, że funkcja aktywacji ma postać symetrycznej funkcji liniowej z nasyceniem, to działanie tego typu sieci w kroku t można opisać wzorem

$$y_k(t+1) = \text{satlins} \left(\sum_{j=1, j \neq k}^p w_{kj} y_j(t) + \theta_k \right),$$

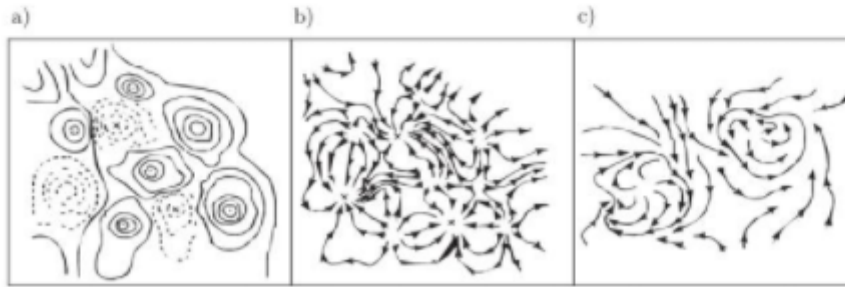
przy czym $y_j(0) = x_j$. Sygnał na wyjściu będzie się zmieniał tak długo, aż $y_k(t) = y_k(t-1)$ dla wszystkich neuronów sieci. Stan energetyczny sieci można wyrazić za pomocą funkcji Lapunowa (w fizyce statystycznej funkcja Hamiltona, w teorii optymalizacji funkcja celu lub kosztu, w obliczeniach ewolucyjnych funkcja przystosowania) postaci

$$E(\mathbf{W}) = -\frac{1}{2} \sum_j \sum_{k \neq j} y_j y_k w_{kj} - \sum_k \theta_k y_k. \quad (7.53)$$

Funkcja energii jest ograniczona od dołu. Można pokazać, że zmiana energii

$$\Delta E(\mathbf{W}) = -\Delta y_k \left(\sum_{j \neq k} y_j w_{kj} + \theta_k \right)$$

jest zawsze ujemna. Stąd funkcja energii jest funkcją malejącą w kolejnych krokach. Dlatego jesteśmy pewni, że osiągnie ona minimum, w którym sieć się ustabilizuje. Na rysunku 7.21 są przedstawione obszary przyciągania sieci Hopfielda.



Rys. 7.21. Obszary przyciągania sieci Hopfielda: a) krajobraz energetyczny sieci, linie ciągłe oznaczają wzniesienie, przerywane dolinę, b) możliwe linie trajektorii sieci, c) bardziej skomplikowana dynamika działania sieci

Uczenie sieci tego typu odbywa się najczęściej za pomocą uogólnionej reguły Hebba postaci

$$w_{kj} = \frac{1}{n} \sum_{i=1}^K x_{ik} x_{ij}, \quad (7.54)$$

która nie daje, niestety, oczekiwanych rezultatów (pojemność sieci – maksymalna liczba wzorców zapamiętanych i odtworzonych z pewnym akceptowalnym błędem 1% – wynosi jedynie 13.8% liczby neuronów). Z tego powodu w praktyce używa się metody *pseudoinwersji*. W metodzie tej zakłada się, że celem uczenia jest taki dobór wag, aby na wyjściu otrzymać taki sam sygnał jak na wejściu, czyli

$$\mathbf{W}\mathbf{X} = \mathbf{X},$$

co prowadzi do rozwiązania

$$\mathbf{W} = \mathbf{X}\mathbf{X}^+,$$

gdzie \mathbf{X}^+ oznacza odwrotność Moore'a–Penrose'a.

Definicja 7.4.

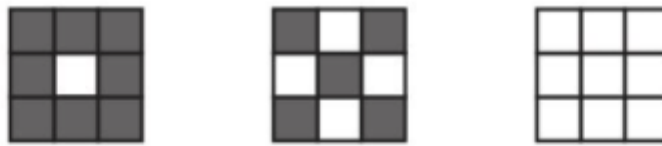
Uogólniona odwrotność \mathbf{A}^+ macierzy \mathbf{A} nazywa się *MP-odwrotnością* (*Moore'a–Penrose'a odwrotnością*) wtedy i tylko wtedy, gdy zachodzą następujące warunki:

$$\begin{aligned} \mathbf{A}\mathbf{A}^+\mathbf{A} &= \mathbf{A}, \\ \mathbf{A}^+\mathbf{A}\mathbf{A}^+ &= \mathbf{A}^+, \\ (\mathbf{A}\mathbf{A}^+)' &= \mathbf{A}\mathbf{A}^+, \\ (\mathbf{A}^+\mathbf{A})' &= \mathbf{A}^+\mathbf{A}. \end{aligned}$$

Sieć uczona tą metodą ma pojemność n .

Przykład 7.11.

Przeprowadzimy teraz próbę nauczania sieci Hopfielda rozpoznawania trzech symboli. Za pomocą sieci złożonej z 9 neuronów rozpoznajemy obrazek złożony z 9 pikseli. Czarny piksel jest reprezentowany jako 1, natomiast biały jako -1 . W ten sposób otrzymaliśmy 9-elementowy wektor testowy. Za pomocą reguły (7.54) nauczyliśmy sieć 3 kształtów, z których 2 przypominają kółko i krzyżyk, a jako trzeci wektor wzorcowy podaliśmy same -1 , co odpowiada pustemu kwadratowi. Wzorce są zobrazowane na rys. 7.22.



Rys. 7.22. Wzorce uczące do przykładu 7.11

W ten sposób otrzymujemy następujące dane uczące:

$$\mathbf{x}_1 = [1, 1, 1, 1, -1, 1, 1, 1, 1]',$$

$$\mathbf{x}_2 = [1, -1, 1, -1, 1, -1, 1, -1, 1]',$$

$$\mathbf{x}_3 = [-1, -1, -1, -1, -1, -1, -1, -1, -1]'$$

Następnie uczymy sieć Hopfielda za pomocą powyższych wzorców. Okazuje się, że gdy na wejście sieci podamy wektor identyczny z wzorcem, wówczas sieć nie zmieni swojego stanu. Oznacza to, że sieć idealnie rozpoznaje dane niezaburzone. Teraz zaburzamy dane wejściowe (rys. 7.23).



Rys. 7.23. Zaburzone dane do przykładu 7.11

Zaburzone dane mają więc postać:

$$\tilde{\mathbf{x}}_1 = [1, -1, -1, 1, -1, 1, 1, 1, 1]',$$

$$\tilde{\mathbf{x}}_2 = [1, -1, 1, -1, -1, -1, 1, -1, 1]',$$

$$\tilde{\mathbf{x}}_3 = [-1, 1, -1, -1, 1, 1, 1, -1, 1]'$$

Dwa pierwsze zaburzone sygnały zostały poprawnie rozpoznane po 2 iteracjach, natomiast dla sygnału trzeciego wyjście sieci wciąż było niestabilizowane po 20 iteracjach, co świadczy, że sieć nie mogła rozpoznać zniekształconego sygnału.

Przykład 7.12.

Wróćmy teraz do przykładu 7.3. Spróbujemy go rozwiązać za pomocą sieci Hopfielda (lekko zmodyfikowanej, bo ze sprzężeniem zwrotnym obejmującym ten sam neuron – bardzo podobna sieć do omówionej później sieci MAXNET, przy czym waga sprzężenia ze samym sobą nie musi być równa jeden). Sieć

Hopfielda składa się w tym przykładzie z 3 neuronów i posłuży do rozpoznawania dwóch typów obiektów. W tym przypadku po zastosowaniu metody pseudoinwersji dla macierzy

$$\mathbf{X} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ -1 & -1 \end{bmatrix}$$

otrzymano macierz wag postaci

$$\mathbf{W} = \begin{bmatrix} 0.5 & 0 & -0.5 \\ 0 & 1 & 0 \\ -0.5 & 0 & 0.5 \end{bmatrix}.$$

Wektor progowy ustalamy arbitralnie, sensowny wydaje się np. wektor

$$\boldsymbol{\theta} = \begin{bmatrix} 0.9 \\ 0 \\ -0.9 \end{bmatrix}.$$

Tak dobrany próg spowoduje, że pierwsza składowa wektora wynikowego będzie szybko dążyć do wartości 1, a trzecia do wartości -1 (są one wspólne dla jabłek i pomarańczy). Natomiast druga składowa wektora wyjściowego będzie jedynie zależna od wejścia i macierzy wag, która jednak nie zmienia wejścia. Poszczególne składowe wektora wynikowego zmieniają się zatem według następujących formuł:

$$\begin{aligned} y_1(t+1) &= \text{satlins}(0.5y_1(t) - 0.5y_3(t) + 0.9), \\ y_2(t+1) &= \text{satlins}(y_2(t)), \\ y_3(t+1) &= \text{satlins}(-0.5y_1(t) + 0.5y_3(t) - 0.9). \end{aligned}$$

Wektory wzorcowe sieć klasyfikuje poprawnie. Natomiast dla wektora $[-1, -1, -1]^T$ tak skonstruowana sieć zwraca następujące elementy na wyjściu w kolejnych krokach

$$\mathbf{y}(0) = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{y}(1) = \begin{bmatrix} 0.9 \\ -1 \\ -0.9 \end{bmatrix}, \quad \mathbf{y}(2) = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{y}(3) = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}.$$

Czyli podobnie jak perceptron sieć Hopfielda zaklasyfikowała obiekt jako pomarańczę. W przeciwieństwie jednak do perceptronu, który jako wynik podawał etykietę klasy, sieć Hopfielda dostarcza wzorzec wynikowy, który jest najbardziej podobny do klasyfikowanego obiektu.

7.4.2. Sieci Hamminga

Sieć Hamminga jest trójwarstwową strukturą rekurencyjną, stanowiącą rozwinięcie sieci Hopfielda i jest wyspecjalizowaną strukturą pamięci heteroasocjacyjnej (kojarzenie dwóch różnych wektorów). Została ona zaproponowana