Nie takie sztuczne neurony

*Dominik KRZEMIŃSKI**

* Uniwersytet Cardiff

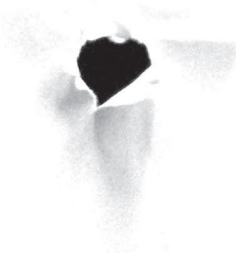
Niemal każdy wykład wprowadzający w zagadnienie sztucznych sieci neuronowych zaczyna się anegdotą o zaczerpnięciu tego pojęcia z anatomii mózgu. Najprostszy model matematycznej jednostki zwanej „sztucznym neuronem" opisuje równanie:

$$y = \sigma\Big(\sum_i w_i\, x_i + b\Big).$$

O sieciach neuronowych pisaliśmy w $\Delta^{1,5}_{18}$.

W analogii do biologicznych komórek neuronowych mamy do czynienia z sumą informacji x (przesyłanej z komórek presynaptycznych do postsynaptycznych) z odpowiednimi wagami w (siła połączeń synaptycznych), na którą działa funkcja nieliniowa σ (uwolnienie potencjału czynnościowego po osiągnięciu wartości progowej).

Historia wzajemnych wpływów uczenia maszynowego i neuronauki jest jednak znacznie dłuższa i o wiele bardziej skomplikowana. Szczególnie wyraźnie zaznacza się to w ostatnich latach, gdy niezwykłą popularność zdobywają tzw. głębokie sieci neuronowe (*deep neural networks*). Swoją strukturą coraz bardziej przypominają one skomplikowane układy przetwarzające informacje w mózgu. W niedawnym wydaniu periodyku naukowego *Neuron* rolę współpracy naukowców z tych dwóch obszarów wiedzy podkreślał sam Demmis Hassabis, współzałożyciel DeepMind. Dla przypomnienia, ta należąca do Google firma zasłynęła opracowaniem programu AlphaGo, który jako pierwszy automat wygrał (w 2016 roku) z arcymistrzem gry w Go, Lee Sedolem. Ciekawostką jest fakt, że AlphaGo został uhonorowany za to przez południowokoreańską federację dziewiątym danem. Gra planszowa Go uznawana jest przez ekspertów za najtrudniejszą na świecie!



Jednym z najprostszych modeli matematycznych znajdującym zastosowanie w obu wspomnianych dziedzinach jest tzw. sieć Hopfielda. W oryginalnym sformułowaniu dyskretna sieć Hopfielda składa się z N neuronów, z których każdy łączy się z każdym i może przybrać jeden z dwóch stanów: $+1$, lub -1. Przez $v_i[t]$ oznaczać będziemy stan i-tego neuronu w chwili t. Wagę połączenia między i-tym i j-tym neuronem oznaczamy przez w_{ij}. Zakładamy, że wagi są symetryczne ($w_{ij} = w_{ji}$) oraz że neuron nie wpływa sam na siebie ($w_{ii} = 0$). Stan jednostki i w chwili $t+1$ w zależności od stanu układu w chwili t opisuje się równaniem:

$$v_i[t+1] = \operatorname{sgn}\Big(\sum_{j=1}^{N} w_{ij} v_j[t]\Big) = \operatorname{sgn}\Big(\sum_{j\neq i} w_{ij} v_j[t]\Big),$$

gdzie funkcja sgn przyjmuje wartość $+1$, gdy jej argument jest większy od zera, bądź -1 w przeciwnym przypadku. Wyrażenie $\sum_{j\neq i} w_{ij} v_j[t]$ można interpretować jako *ekscytację* jednostki k – to znaczy, że w zależności od znaku ekscytacji nastąpi aktywacja ($+1$) bądź deaktywacja (-1) jednostki. Wyrażenie to będziemy oznaczać przez $\varepsilon_i[t]$.

Dla tak zdefiniowanej sieci określamy energię układu jako:

$$(\star) \qquad E(v) = -\frac{1}{2}\sum_{i,j=1} w_{ij} v_i v_j = -\frac{1}{2}\sum_{i\neq j} w_{ij} v_i v_j.$$

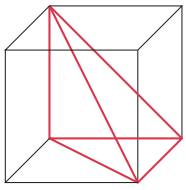
Zakładamy, że aktualizacja stanu następuje w sposób asynchroniczny, to znaczy żadne dwa neurony nie są włączane bądź wyłączane w tym samym momencie. Wówczas zmiana energii przy aktualizacji jednostki k wynosi:

$$\Delta E_k = E_k(v[t+1]) - E_k(v[t]) = -\sum_{j\neq k} w_{kj} v_k[t+1] v_j[t] + \sum_{j\neq k} w_{kj} v_k[t] v_j[t] =$$
$$= (v_k[t] - v_k[t+1]) \sum_{j\neq k} w_{kj} v_j[t] = (v_k[t] - v_k[t+1])\varepsilon_k[t].$$

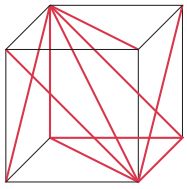
Teraz rozpatrzmy dwa przypadki. Kiedy uczenie nie następuje, nie zmienia się stan układu, czyli $v_k[t] = v_k[t+1]$. Mamy zatem $\Delta E_k = 0$, czyli $E_k(v_k[t+1]) = E_k(v_k[t])$. W przeciwnym przypadku musimy rozważyć kolejne dwie możliwości. Ponieważ nastąpiła zmiana układu:

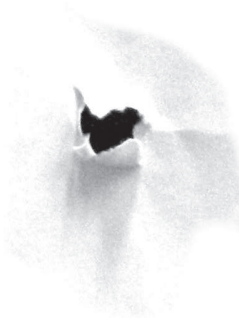
- dla $\varepsilon_k[t] < 0$ mamy $v_k[t] = +1$, a $v_k[t+1] = -1$;
- dla $\varepsilon_k[t] > 0$ mamy $v_k[t] = -1$, a $v_k[t+1] = +1$.

Rozwiązanie zadania M 1583.
Sześcian wypełniają trzy kopie czworościanu przedstawionego na poniższym rysunku



i trzy kopie jego lustrzanego odbicia, co Czytelnik Uważny zobaczy na kolejnym rysunku.





Za każdym razem wyrażenie $(v_k[t] - v_k[t+1])$ ma przeciwny znak do wartości ekscytacji. Pokazaliśmy zatem, że $\Delta E_k < 0$. W każdym przypadku energia będzie maleć, a że liczba stanów sieci jest skończona, w skończonym czasie zbiegnie do stanu minimum lokalnego (tzn. niemożliwego do „poprawienia" poprzez opisaną aktywność neuronów). Będzie to tzw. stan stabilny.

Jak w takim razie sieć Hopfielda może się czegokolwiek „nauczyć"?
W najprostszym wydaniu rzecz opiera się na *regule uczenia Hebba*, opracowanej już w latach 50. XX wieku przez kanadyjskiego psychologa Donalda Hebba. Mówi ona o tym, że jeśli neuron A systematycznie pobudza neuron B, to połączenie synaptyczne między nimi staje się silniejsze (po angielsku często jest to podawane w zgrabnej formie: *fire together, wire together*). Ta prosta hipoteza została potwierdzona kilkanaście lat po jej sformułowaniu, poprzez odkrycie paradygmatu długotrwałych wzmocnień synaptycznych. Niewątpliwą zaletą reguły Hebba jest fakt, iż w prosty sposób łączy idee z neurobiologii i psychologii, a także stanowi dobry model pamięci asocjacyjnej. Szczególnym przykładem takiej pamięci może być warunkowanie zastosowane w słynnym eksperymencie Pawłowa. Powtarzana ekspozycja psa na miskę z jedzeniem wraz z dźwiękiem dzwonka skutkuje wzmożonym wydzielaniem śliny w reakcji wyłącznie na dźwięk dzwonka.

Uważny Czytelnik zauważy, że jest to paradygmat uczenia bez nadzoru. Zgadza się to z intuicyjnym pojmowaniem działania ludzkiej pamięci, w której skojarzenia tworzone są jedynie na podstawie zaobserwowanych wzorców.

Regułę Hebba do zapamiętania wzorca opisanego binarnym wektorem s można zapisać jako:

$$w_{ij} = \frac{1}{N} s_i s_j.$$

Zwróćmy uwagę, że przy tak dobranych wagach najmniejsza wartość energii określonej wzorem $(\star)$ przyjmowana jest dla $v = s$. Ponieważ opisany wcześniej proces uczenia stabilizuje się w minimach lokalnych, możemy mieć nadzieję, że będzie on „zbliżać" wektor v do s.

Rys. 1. Zapamiętane wzorce

W ogólności, możemy zapamiętać więcej niż jeden wzór. Na przykład, dla P wzorców $s^1, s^2, \ldots, s^P$ reguła Hebba przyjmuje postać:

$$w_{ij} = \frac{1}{N} \sum_{k=1}^{P} s_i^k s_j^k$$

co jest odpowiednikiem pamięci skojarzeniowej. Można udowodnić, że pojemność takiej sieci to $\frac{P}{N} \approx 0{,}138$. Oznacza to, że sieć złożona z 1000 węzłów jest w stanie zapamiętać maksymalnie około 138 wzorców. Grafika na marginesie pokazuje przykład rekonstrukcji litery „C" dla sieci pamiętającej trzy wzorce.

Rys. 2. Rekonstrukcja litery „C" w trzech krokach

Osiągnięcie lokalnego minimum w sieci Hopfielda jest gwarantowane. Często zdarza się jednak utknięcie w minimum fałszywym, a zatem rozpoznanie wzorca, którego de facto sieć nie była nauczona. Nie zmienia to jednak faktu, iż z powodzeniem sieci te stosuje się do odszumiania obrazów, dekonwolucji danych, rozpoznawania wzorców, a także w problemach optymalizacyjnych.

Co najważniejsze jednak, idee Donalda Hebba przyczyniły się do powstania modelu równoległego rozproszonego przetwarzania informacji, czyli tzw. koneksjonimzu, spopularyzowanego w latach 80. przez naukowców Uniwersytetu Stanforda: Jamesa McClellanda i Davida Rumelharta. Sieć Hopfielda jest najlepszym przykładem takiego modelu. Istnieje przekonanie, że właśnie z koneksjonizmu wyewoluowała dziedzina głębokich sieci neuronowych. Nic dziwnego, skoro do zwolenników tej idei należy sam Geoffrey Hinton, absolwent psychologii kognitywnej, uważany za ojca chrzestnego *deep learning*'u. Przykłady interakcji neuronauki i uczenia maszynowego można mnożyć bez końca. Nie wspomniałem tu, na przykład, o klasyfikatorach, wykorzystywanych do dekodowania sygnałów pochodzących z mózgu. Na uwagę zasługują również próby interpretacji działania sztucznych sieci neuronowych, gdzie wykorzystywane są modele kognitywne. Cieszy więc zacieśnianie się współpracy między badaczami obu dziedzin, co pomaga w wypracowaniu wspólnego języka. Cel przecież jest zbieżny: zrozumieć, jak uczy się człowiek, by sprawniej mogła uczyć się maszyna.

Zainteresowanym większą ilością przykładów wzajemnych wpływów neuronauk i uczenia maszynowego polecam pracę przeglądową wspomnianą na początku artykułu: D. Hassabis i in., Neuroscience-Inspired Artificial Intelligence, 2017, *Neuron*.