Sztuczne sieci neuronowe

Krzysztof A. Cyran
POLITECHNIKA ŚLĄSKA
Instytut Informatyki, p. 335

Wykład 9

PLAN:

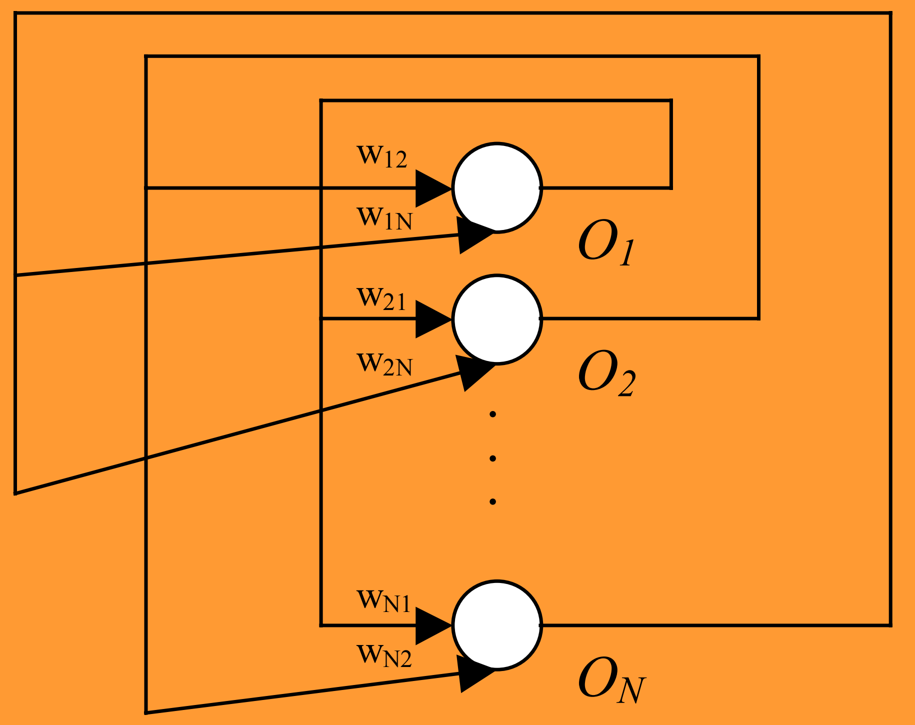
Sieć Hopfielda

- architektura sieci
- rodzaje sieci Hopfielda
- funkcja energetyczna Lapunowa
- proces odtwarzania sieci (minimalizacja funkcji energii)
- proces uczenia sieci

Sieć Hopfielda

- Sieć rekurencyjna minimalizująca funkcję energetyczną (funkcję Lapunowa)
- Odmiany:
  - sieć dyskretna,
  - sieć ciągła
- Zastosowania:
  - pamięć autoasocjacyjna (model dyskretny)
  - Problem komiwojażera (model ciągły)

Sieć Hopfielda



Sieć Hopfielda (oznaczenia)

- $n_k^{(p)}$ - pobudzenie sieciowe k-tego neuronu w kroku p-tym
- $O_j^{(p)}$- wyjście j-tego neuronu w kroku p-tym
- w_{kj} – wartość wagi od j-tego do k-tego neuronu
- t_k – próg przełączenia k-tego neuronu

Sieć Hopfielda: model dyskretny

- Działanie sieci opisują wzory:

$$n_k^{(p)} = \sum_{j=1}^{N} w_{kj} O_j^{(p)} - t_k, \qquad (1)$$

dla wersji binarnej:

$$O_k^{(p+1)} = \begin{cases} 1 \Leftrightarrow n_k^{(p)} > 0 \\ O_k^{(p)} \Leftrightarrow n_k^{(p)} = 0 \\ 0 \Leftrightarrow n_k^{(p)} < 0 \end{cases} \qquad (2)$$

dla wersji bipolarnej:

$$O_k^{(p+1)} = \begin{cases} +1 \Leftrightarrow n_k^{(p)} > 0 \\ O_k^{(p)} \Leftrightarrow n_k^{(p)} = 0 \\ -1 \Leftrightarrow n_k^{(p)} < 0 \end{cases} \qquad (2')$$

Sieć Hopfielda (cechy)

- Brak wpływu neuronu na samego siebie:

$$\boxed{\forall i, w_{ii} = 0}$$

- Symetria:

$$\boxed{\forall ij, w_{ij} = w_{ji}}$$

Sieć Hopfielda (funkcja energii)

- Z siecią Hopfielda kojarzy się tzw. funkcję energii (funkcję Lapunowa)
- Jest to funkcja ograniczona od dołu i nierosnąca w trakcie ewolucji rozważanego procesu (u nas procesu zmian stanów rekurencyjnej sieci Hopfielda, nazywanego procesem ODTWARZANIA SIECI)

Sieć Hopfielda (odtwarzanie)

- Proces odtwarzania rozpoczynamy dla $p=0$ od doprowadzenia do elementów przetwarzających sygnałów wejściowych:

$\mathbf{x} = [x_1, x_2,...,x_N] \;\; x_i \in \{0,1\}$

tzn. $O_i^{(0)}=x_i$ dla $i=1,...,N$.

- Następnie odłącza się sygnały wejściowe i rozpoczyna się proces odtwarzania według (1) i (2)

Sieć Hopfielda (odtwarzanie)

- Sieć działa asynchronicznie, tj. w danej chwili wybierany jest z równym prawdopodobieństwem dowolny neuron i tylko on jest aktualizowany
- Po pewnej skończonej liczbie iteracji sieć osiąga stan stabilny, tzn.:

$$\forall i, O_i^{(p+1)} = O_i^{(p)}$$

Jest to stan odpowiadający lokalnemu minimum funkcji energii. Ten stan jest przekazywany na wyjście sieci.

Sieć Hopfielda (funkcja energii)

- Funkcję energii wybiera się jako:

$$E(\mathbf{O}) = -\frac{1}{2}\mathbf{O}^T\mathbf{w}\mathbf{O} + \mathbf{t}^T\mathbf{O}$$

co w zapisie skalarnym daje:

$$E(\mathbf{O}) = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}O_iO_j + \sum_{i=1}^{N} t_iO_i \qquad (3)$$

Sieć Hopfielda (spadek energii)

- Dowód, że funkcja (3) jest funkcją nierosnącą w procesie odtwarzania:

Niech w chwili *p* zmienia się stan *k*-tego neuronu

$$O_k^{(p+1)} = O_k^{(p)} + \Delta O_k^{(p)}$$

Zaś stan pozostałych nie zmienia się:

$$O_j^{(p+1)} = O_j^{(p)} \quad \text{dla } j \neq k$$

Sieć Hopfielda (spadek energii)

$$\Delta E^{(p)} = E\left(\mathbf{O}^{(p+1)}\right) - E\left(\mathbf{O}^{(p)}\right) =$$

$$-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} O_i^{(p+1)} O_j^{(p+1)} + \sum_{i=1}^{N} t_i O_i^{(p+1)} +$$

$$+\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} O_i^{(p)} O_j^{(p)} - \sum_{i=1}^{N} t_i O_i^{(p)} =$$

$$= -\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{j=1, j\neq k}^{N} w_{ij} O_i^{(p+1)} O_j^{(p+1)} + w_{ik} O_i^{(p+1)} O_k^{(p+1)} \right) + \sum_{i=1, i\neq k}^{N} t_i O_i^{(p+1)} + t_k O_k^{(p+1)} +$$

$$+\frac{1}{2}\sum_{i=1}^{N}\left(\sum_{j=1, j\neq k}^{N} w_{ij} O_i^{(p)} O_j^{(p)} + w_{ik} O_i^{(p)} O_k^{(p)} \right) - \sum_{i=1, i\neq k}^{N} t_i O_i^{(p)} - t_k O_k^{(p)} =$$

verte

Sieć Hopfielda (spadek energii)

$$= -\frac{1}{2}\sum_{i=1, i\neq k}^{N}\left(\sum_{j=1, j\neq k}^{N} w_{ij} O_i^{(p+1)} O_j^{(p+1)} + w_{ik} O_i^{(p+1)} O_k^{(p+1)}\right) +$$

$$-\frac{1}{2}\left(\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p+1)} O_j^{(p+1)} + \underbrace{w_{kk}}_{=0} O_k^{(p+1)} O_k^{(p+1)}\right) + t_k O_k^{(p+1)} +$$

$$+\frac{1}{2}\sum_{i=1, i\neq k}^{N}\left(\sum_{j=1, j\neq k}^{N} w_{ij} O_i^{(p)} O_j^{(p)} + w_{ik} O_i^{(p)} O_k^{(p)}\right) +$$

$$+\frac{1}{2}\left(\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p)} O_j^{(p)} + \underbrace{w_{kk}}_{=0} O_k^{(p)} O_k^{(p)}\right) - t_k O_k^{(p)} =$$

verte

Sieć Hopfielda (spadek energii)

$$= -\frac{1}{2}\sum_{i=1, i\neq k}^{N} w_{ik} O_i^{(p+1)} O_k^{(p+1)} - \frac{1}{2}\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p+1)} O_j^{(p+1)} + t_k O_k^{(p+1)} +$$

$$+ \frac{1}{2}\sum_{i=1, i\neq k}^{N} w_{ik} O_i^{(p)} O_k^{(p)} + \frac{1}{2}\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p)} O_j^{(p)} - t_k O_k^{(p)} =$$

$$= -\frac{1}{2}\sum_{i=1, i\neq k}^{N} w_{ik} O_i^{(p)} O_k^{(p)} - \frac{1}{2}\sum_{i=1, i\neq k}^{N} w_{ik} O_i^{(p)} \Delta O_k^{(p)}$$

$$-\frac{1}{2}\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p)} O_j^{(p)} - \frac{1}{2}\sum_{j=1, j\neq k}^{N} w_{kj} \Delta O_k^{(p)} O_j^{(p)} + t_k O_k^{(p)} + t_k \Delta O_k^{(p)} +$$

$$+ \frac{1}{2}\sum_{i=1, i\neq k}^{N} w_{ik} O_i^{(p)} O_k^{(p)} + \frac{1}{2}\sum_{j=1, j\neq k}^{N} w_{kj} O_k^{(p)} O_j^{(p)} - t_k O_k^{(p)} =$$

verte

Sieć Hopfielda (spadek energii)

$$= -\sum_{j=1}^{N} w_{kj} \Delta O_k^{(p)} O_j^{(p)} + t_k \Delta O_k^{(p)} =$$

$$= -\Delta O_k^{(p)} \left(\sum_{j=1}^{N} w_{kj} O_j^{(p)} - t_k \right) =$$

$$= -\Delta O_k^{(p)} n_k^{(p)}$$

Sieć Hopfielda (spadek energii)

- *Czyli:* $$\Delta E^{(p)} = -\Delta O_k^{(p)} n_k^{(p)}$$

Jeżeli $n_k^{(p)}=0$ to: $\Delta E^{(p)}=0$

Jeżeli $n_k^{(p)}\neq 0$ to:

$O_k^{(p+1)}$	$O_k^{(p)}$	$\Delta O_k^{(p)}$	$n_k^{(p)}$	$\Delta E^{(p)}$
0	0	0	-	0
0	1	-1	-	-
1	0	1	+	-
1	1	0	+	0

Sieć Hopfielda (spadek energii)

Zatem $\Delta E^{(p)}$ jest zawsze równe zero lub ujemne: $\Delta E^{(p)} \leq 0$

Czyli

$E(\mathbf{O}^{(p+1)}) \leq E(\mathbf{O}^{(p)})$ co należało wykazać.

Ponadto: jeżeli $n_k^{(p)}=0$ to z (2): $O_k^{(p+1)} = O_k^{(p)}$ *(brak zmiany wyjść)*

Jest to jeden z warunków kiedy $\Delta E^{(p)} = 0$.

Inne sytuacje kiedy $\Delta E^{(p)} = 0$ są (z tabeli) gdy

$O_k^{(p+1)} = O_k^{(p)} = 0$ *lub* $O_k^{(p+1)} = O_k^{(p)} = 1$ (*też brak zmiany wyjść*)

NATOMIAST JEŚLI ZAJDZIE DOWOLNA ZMIANA to $\Delta E^{(p)} < 0$.

!!! ENERGIA SPADA PRZY KAŻDEJ ZMIANIE !!!

Sieć Hopfielda (spadek energii)

- Z drugiej strony ponieważ moduł energii spełnia następującą nierówność:

$$\left|E(\mathbf{O})\right| \leq \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|w_{ij}\right| + \sum_{i=1}^{N}\left|t_i\right|$$

Więc wykluczona jest sytuacja, że $E \rightarrow -\infty$.

Ponieważ każdy ciąg ograniczony i monotoniczny jest zbieżny, więc E będzie dążyć do pewnej wartości E_{min}

Sieć Hopfielda (spadek energii)

- Ponieważ dziedzina funkcji jest skończona

($O_i \in \{0,1\}$), więc zbiór możliwych zmian ΔE przed osiągnięciem E_{min} jest również skończony (tzn. zmiany nie mogą być nieskończenie małe).

$$\exists c, \min_{\Delta E \neq 0} |\Delta E| = c > 0$$

Zatem czas potrzebny na osiągnięcie E_{min} jest skończony w sensie ilości kroków p.

Sieć Hopfielda (stan stabilny)

- Stan $\mathbf{O}^{(P)}$ dla którego $E(\mathbf{O}^{(P)})=E_{min}$ jest stanem stabilnym, tzn. takim, że:

$$\forall p \geq P, \mathbf{O}^{(p+1)} = \mathbf{O}^{(p)}$$

- E_{min} odpowiada więc minimium lokalnemu i dalsze zmiany są niemożliwe
- Stany stabline (minima lokalne) nazywają się atraktorami, z których każdy posiada swoją nieckę przyciągania, tj. zbiór stanów $\mathbf{O}^{(0)}$, które inicjują ewolucję kończącą się w tym stanie.
- Na rodzaj i ilość atraktorów wpływa dobór wag, czyli proces uczenia sieci Hopfielda.

Sieć Hopfielda (uczenie)

- Przykład: uczenie sieci Hopfielda jako pamięci autoasocjacyjnej
- Autoasocjacja polega na odtwarzaniu na zasadzie skojarzeń całości informacji na podstawie dostępnego jej fragmentu
- w przykładzie z Einsteinem i jego filozofią, kojarząc z nazwiskiem całą sylwetkę naukową wykluczyliśmy go spośród podmiotów kiepskiego filozofowania, mimo na pozór poprawnego wniosku, że on sam uważał się za kiepskiego filozofa: rozwiązaniem paradoksu było znalezienie ukrytego kontekstu wypowiedzi)
- Inne przykłady to dopasowanie haseł na podstawie kilku liter w krzyżówce, czy utworu muzycznego na podstawie kilku dźwięków

Sieć Hopfielda (autoasocjacja)

- Formalnie autoasocjację przedstawiamy następująco:

Dany jest ciąg *M wzorców*: $\{x^1, x^2, \ldots, x^M\} \subset \mathrm{R}^N$

Pamięcią autoasocjacyjną nazywamy układ realizujący odwzorowanie: $F: R^N \rightarrow R^N$, taki że: $F(x^m)=x^m$ dla $m=1..M$, oraz

$F(x)=x^l$, gdzie x^l jest jest najbardziej podobnym wzorcem, tzn. jego odległość od x jest najmniejsza.

Sieć Hopfielda (uczenie)

- Ponieważ kształt funkcji energii zależy od wag w_{ij} należy je tak dobrać aby każdy wzorzec stał się atraktorem a odpowiednia niecka przyciągania była na tyle głęboka i szeroka, aby zapewnić poprawność skojarzeń pomiędzy warunkami początkowymi a stanem końcowym.
- W przypadku M wzorców Hopfield podał następujący sposób ustalania wag:

$$w_{ij} = \begin{cases} \sum_{m=1}^{M} \left(2x_i^{(m)} - 1\right)\left(2x_j^{(m)} - 1\right) & \text{gdy } i \neq j \\ 0 & \text{gdy } i = j \end{cases}$$

Sieć Hopfielda (uczenie)

- Reguła Hopfielda oznacza, że waga w_{ij} wzrasta o 1 gdy i-ta i j-ta składowa danego wzorca są identyczne, i maleje o 1 gdy te składowe są różne
- Liczba losowych wzorców poprawnie pamiętanych wynosi $M_{max} \approx \alpha N$, gdzie $\alpha \approx 0.138$
- Nie jest to dużo, jednakże sieć Hopfielda jest stosowana jako pamięć autoasocjacyjna ze względu na jej prostotę.

Sieć Hopfielda (uczenie)

- Lepszą metodą uczenia jest iteracyjna metoda rzutowania Δ, będąca odmianą gradientowego algorytmu minimalizacji odpowiednio zdefiniowanej funkcji celu:

$$\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} + \frac{\eta}{N}\left[\mathbf{x}^{(p)} - \mathbf{w}^{(t)}\mathbf{x}^{(p)}\right]\left[\mathbf{x}^{(p)}\right]^T$$

η z przedziału <0.7 do 0.9>