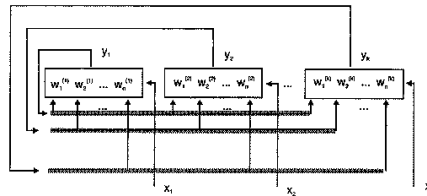
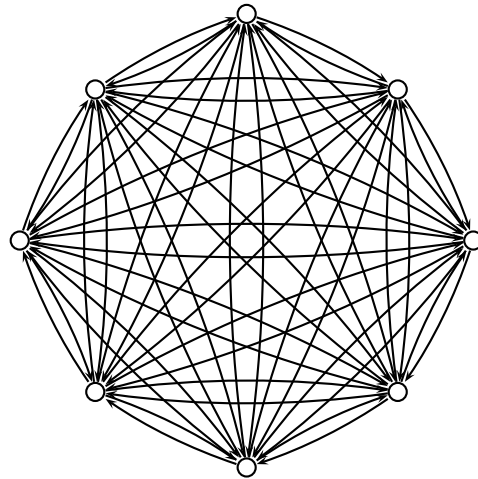


Architektura

Poprzednio rozważane sieci miały architekturę warstwową i przetwarzały dane w jednym kierunku (dane wejściowe od wejścia do wyjścia, a błędy w kierunku odwrotnym).

Sieci Hopfielda wprowadają zupełnie inną architekturę: w najprostszym przypadku elementy połączone są każdy z każdym poprzez symetryczne wagi.



Co się dzieje w sieciach Hopfielda

Ciekawą, nową w stosunku do sieci jednokierunkowych, własnością sieci Hopfielda jest możliwość pojawienia się w nich pewnych przebiegów dynamicznych.

W ogólności sieć ta realizuje odwzorowanie nieliniowe postaci:

$$Y^{(j+1)} = \Phi \left(X^{(j)}, Y^{(j)} \right)$$

Najczęściej rozważana jest sytuacja, gdy sieci podajemy bodziec w postaci $X^{(0)} \neq 0$, a następnie zerujemy wejścia $X^{(j)} = 0$ dla $j > 0$ i obserwujemy proces ewolucji wektora $Y^{(j)}$ w przestrzeni stanów sieci.

$$Y^{(j+1)} = \Phi \left(Y^{(j)} \right)$$

Możliwe są bardzo rozmaite zachowania sieci:

- dążenie do ustalonych wartości Y^*
- oscylacje
- ruch po atraktorze chaotycznym
- rozbieganie się do nieskończoności

Pamięć adresowana treścią

Problem, który będziemy teraz rozważać jest następujący:

Zapamiętaj zbiór p wzorców z_i^m w taki sposób, aby po zaprezentowaniu nowego wzorca x_i reakcją sieci było wytworzenie tego zapamiętanego wzorca, który jest najbardziej podobny do x_i

Wzorce będziemy numerować przez $m = 1, 2, \dots, p$, a jednostki sieci przez $i = 1, 2, \dots, N$. Dla uproszczenia przyjmujemy, że wzorce zapamiętane z_i^m i wzorce testowe x_i mogą być 0 lub 1 w każdym elemencie sieci.

Rozwiązanie algorytmiczne:

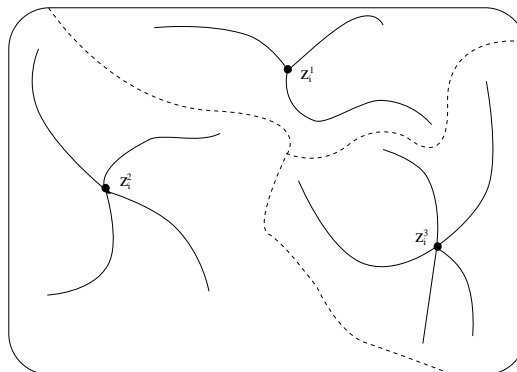
obliczyć odległość Hamminga $\sum_i [z_i^m(1 - x_i) + (1 - z_i^m)x_i]$ wzorca testowego od każdego ze wzorców zapamiętanych, a następnie wybrać ten dla którego odległość ta jest najmniejsza.

Sformułowanie problemu dla sieci Hopfielda:

Znaleźć zbiór wag w_{ij} takich, które przeprowadzą sieć od stanu początkowego $y_i^{(0)} = x_i$ do stanu końcowego $y_i^{(k)} = z_i^{m^*}$, takiego że odległość Hamminga $z_i^{m^*}$ od x_i jest najmniejsza.

Idea rozwiązania problemu pamięci adresowanej treścią

Trzeba zbudować taką sieć, dla której zapamiętane wzorce są *atraktorami* w przestrzeni wszystkich możliwych stanów sieci — w *przestrzeni konfiguracji*. Cała przestrzeń konfiguracji podzielona jest na *baseny przyciągania* różnych atraktorów.



Model Hopfielda

Oznaczmy teraz wyjście i -tego neuronu przez S_i przy czym może ono przyjmować wartości tylko -1 lub 1. Elementy sieci obliczają swoje wyjścia:

$$S'_i = \operatorname{sgn} \left(\sum_j w_{ij} S_j \right) \quad \text{przy czym: } \operatorname{sgn}(x) = \begin{cases} 1 & \text{dla } x \geq 0 \\ -1 & \text{dla } x < 0 \end{cases}$$

Wagi w sieci są symetryczne.

Aktualizacja jest asynchroniczna i odbywać się może na dwa równoważne sposoby:

- losowo wybrana jednostka w danej chwili czasu
- każda jednostka aktualizuje swój stan z pewnym prawdopodobieństwem, niezależnie od innych jednostek

Dla ułatwienia będziemy rozważać wzorce losowe: każdy bit z_i jest -1 lub $+1$ z jednakowym prawdopodobieństwem.

Funkcja energetyczna

Hopfield wprowadził pojęcie *funkcji energetycznej* dla swoich sieci:

$$H = -\frac{1}{2} \sum_{ij} w_{ij} S_i S_j$$

składniki $i = j$ dokładają do H jedynie stałą.

Gdy układ ewoluuje zgodnie z dynamiką, to funkcja energetyczna maleje lub pozostaje stała.

W sieciach Hopfielda funkcja energetyczna jest wprowadzona przez analogię do energii układów spinowych.

Aby funkcja energetyczna istniała musimy mieć wagi symetryczne, tzn: $w_{ij} = w_{ji}$

Sprawdźmy, że faktycznie przy dowolnej zmianie stanu sieci zgodnej z jej dynamiką $S'_i = \operatorname{sgn}(\sum_j w_{ij} S_j)$ funkcja energetyczna nie rośnie:

$$\begin{aligned} H' - H &= -\sum_{j \neq i} w_{ij} S'_i S_j + \sum_{j \neq i} w_{ij} S_i S_j = (-S'_i + S_i) \sum_{j \neq i} w_{ij} S_j \\ &= \begin{cases} 0 & \text{gdy } S'_i = S_i \\ 2S_i \sum_j w_{ij} S_j - 2w_{ii} & \text{gdy } S'_i = -S_i \end{cases} \end{aligned}$$

W pierwszym przypadku funkcja energii się nie zmienia, a w drugim oba składniki sumy są ujemne (pierwszy bo S_i ma przeciwny znak niż $\sum_j w_{ij} S_j$, a drugi bo jak zaraz zobaczymy $w_{ii} = p/N$), tzn funkcja energii maleje.

Wyprowadzenie postaci wag z funkcji energetycznej

Chcemy wprowadzić taką funkcję energetyczną która ma minima w stanach stabilnych. Wybieramy ją tak aby osiągała minimum dla największej korelacji między konfiguracją sieci a zapamiętanym wzorcem. Dla jednego wzorca możemy więc wybrać:

$$H = -\frac{1}{2N} \left(\sum_i S_i z_i \right)^2$$

dla p wzorców możemy wziąć sumę takich składników:

$$H = -\frac{1}{2N} \sum_{m=1}^p \left(\sum_i S_i z_i^m \right)^2$$

Po wymnożeniu otrzymujemy:

$$H = -\frac{1}{2N} \sum_{m=1}^p \left(\sum_i S_i z_i^m \right) \left(\sum_j S_j z_j^m \right) = -\frac{1}{2} \sum_{ij} \left(\frac{1}{N} \sum_{m=1}^p z_i^m z_j^m \right) S_i S_j = -\frac{1}{2} \sum_{ij} w_{ij} S_i S_j$$

Ta metoda wyznaczania wag jest interesująca, bo jeżeli dla danego układu umiemy napisać funkcję energetyczną, której minimum jest rozwiązaniem danego problemu to możemy wyznaczyć wartości wag w_{ij} na podstawie współczynników przy $S_i S_j$.

Jeden wzorzec

Wzorzec jest zapamiętany gdy:

- jest stabilny
- dynamika sieci koryguje małe odchylenia od wzorca

Dla jednego wzorca oznacza to, że:

$$\operatorname{sgn} \left(\sum_j w_{ij} z_j \right) = z_i \text{ dla każdego } i$$

łatwo sprawdzić, że

$$w_{ij} = \frac{1}{N} z_i z_j \quad (*)$$

spełnia ten warunek: $\operatorname{sgn} \left(\sum_j w_{ij} z_j \right) = \frac{1}{N} \operatorname{sgn} \left(\sum_j z_i z_j z_j \right) = \frac{1}{N} \operatorname{sgn}(z_i \sum_j 1) = z_i$

Ponadto, jeśli mniej niż połowa bitów wzorca początkowego jest błędna $S_i \neq z_i$ to pobudzenie wejściowe $e_i = \sum_j w_{ij} S_j$ zostanie zdominowana przez większość bitów prawidłowych tak, że $\operatorname{sgn}(e_i) = z_i$.

Widać stąd, że sieć z wagami danymi przez (*) ma atraktor z_i .

— $-z_i$ też jest atraktorem — jest to tzw. *stan odbity*

Wiele wzorców

Wagi:

$$w_{ij} = \frac{1}{N} \sum_{m=1}^p z_i^m z_j^m$$

zapewniają (przy odpowiednio małych p/N) zapamiętanie p wzorców.

Przekonajmy się o tym: Warunek stabilności dla wzorca z_i^n :

$$\text{sgn}(e_i^n) = z_i^n \text{ dla każdego } i$$

przy czym:

$$e_i^n = \sum_j w_{ij} z_j^n = \frac{1}{N} \sum_j \sum_m z_i^m z_j^m z_j^n$$

wydzielimy składnik $m = n$

$$e_i^n = z_i^n + \frac{1}{N} \sum_j \sum_{m \neq n} z_i^m z_j^m z_j^n$$

Warunek stabilności jest spełniony jeśli drugi człon — *przesłuch* — jest < 1 . Przyciąganie przez zapamiętane wzorce działa analogicznie jak w przypadku jednego wzorca.

Stany fałszywe

Oprócz porządanych minimów funkcji energetycznej — *stanów odtwarzanych* znajdujących się w z_i^m występują różne inne typy minimów. Są to:

stany odwrócone $-z_i^m$; mają ten sam poziom energii co stany odtwarzane

stany mieszane z_i^{mix} są kombinacjami liniowymi nieparzystej ilości stanów odtwarzanych. Są równo oddalone od swoich składowych; ich poziom energii jest wyższy niż dla stanów odtwarzanych.

stany szkła spinowego występują dla dużych p i nie są skorelowane ze stanami odtwarzanymi.

Model Isinga

Model opisuje prosty materiał magnetyczny składający się ze zbioru spinów na siatce regularnej. Możliwe wartości spinów $S_i = \pm 1$.

Każdy spin oddziałuje z zewnętrznym polem magnetycznym h^{zewn} i z wewnętrznym polem magnetycznym pochodzącym od pozostałych spinów, czyli oddziałuje z polem:

$$h_i = \sum_j w_{ij} S_j + h^{zewn}$$

gdzie w_{ij} są siłami oddziaływania wymiany, $w_{ij} = w_{ji}$. W zerowej temperaturze spin S_i dąży do stanu $S_i = \text{sgn}(h_i)$. Energia potencjalna odpowiadająca tym oddziaływaniom ma postać:

$$H = -\frac{1}{2} \sum_{ij} w_{ij} S_i S_j - h^{zewn} \sum_i S_i$$

Model Isinga w skończonej temperaturze

W skończonej temperaturze trzeba uwzględnić fluktuacje termiczne. Spin S_i podlega teraz dynamice Glaubera:

$$S_i = \begin{cases} +1 & \text{z prawdopodobieństwem } P(h_i) \\ -1 & \text{z prawdopodobieństwem } 1 - P(h_i) \end{cases}$$

gdzie:

$$P(h_i) = \frac{1}{1 + \exp(-2\beta h_i)}$$

zaś

$$\beta = \frac{1}{k_B T}$$

Ponieważ $P(-h) = 1 - P(h)$ więc dynamikę spinu możemy zapisać w sposób zwarty:

$$\text{Prob}(S_i = \pm 1) = \frac{1}{1 + \exp(\mp 2\beta h_i)}$$

Rozważmy jeden spin w równowadze:

$$S_i = S = \pm 1$$

zaś

$$h = h^{zewn}$$

Wówczas średnia wartość S zwana średnią magnetyzacją $\langle S \rangle$ wynosi:

$$\begin{aligned} \langle S \rangle &= Prob(+1) \cdot (+1) + Prob(-1) \cdot (-1) \\ &= \frac{1}{1 + \exp(2\beta h)} - \frac{1}{1 + \exp(-2\beta h)} \\ &= \frac{\exp(\beta h) - \exp(-\beta h)}{\exp(\beta h) + \exp(-\beta h)} = tgh(\beta h) \end{aligned}$$

Układ wielu spinów

Przybliżone rozwiązanie dla oddziałujących wielu spinów otrzymuje się w ramach teorii pola średniego:

$$\langle h_i \rangle = \sum_j w_{ij} \langle S_j \rangle + h^{zewn}$$

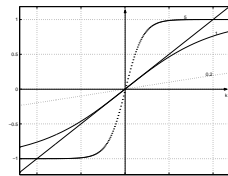
Średnia magnetyzacja w tym przybliżeniu:

$$\langle S_i \rangle = tgh(\beta \langle h_i \rangle) = tgh(\beta \sum_j w_{ij} \langle S_j \rangle + \beta h^{zewn})$$

FERROMAGNETYK: Wszystkie w_{ij} są jednakowe

$$w_{ij} = \frac{J}{N}$$

gdzie J — stała.



Dla ferromagnetyków średnia magnetyzacja jest jednorodna $\langle S_i \rangle = \langle S \rangle$:

Z teorii średniego pola przy $h^{zewn} = 0$ mamy:

$$\langle S \rangle = tgh(\beta J \langle S \rangle)$$

gdy $\beta J \leq 1$ ($T \geq T_c$) to mamy jedno rozwiązanie $\langle S \rangle = 0$

gdy $\beta J > 1$ ($T < T_c$) to mamy dodatkowo dwa niezerowe rozwiązania.

Jednostki stochastyczne

Stan jednostki nie jest jednoznacznie wyznaczony przez jej aktywację. Aktywacja wpływa jedynie na prawdopodobieństwo, że dana jednostka znajduje się w konkretnym stanie.

$$S_i = \begin{cases} +1 & \text{z prawdopodobieństwem } P(e_i) \\ -1 & \text{z prawdopodobieństwem } 1 - P(e_i) \end{cases}$$

Wygodnie jest wybrać funkcję prawdopodobieństwa postaci:

$$Pr(S_i = \pm 1) = \frac{1}{1 + \exp(\mp 2\beta e_i)}$$

ze względu na analogię z modelem Isinga dla spinów. W modelu fizycznym $\beta = \frac{1}{k_B T}$. Przez analogię wprowadzamy dla jednostek stochastycznych:

$$\beta = \frac{1}{T}$$

Tutaj T jest *pseudotemperatura*.

Dla $T \rightarrow 0$ sigmoida dąży do funkcji skokowej i otrzymujemy oryginalny model Hopfielda. Wraz ze wzrostem T próg wygładza się.

Sieć z jednostkami stochastycznymi jest bardziej odporna na wpadanie w minima lokalne.

Stany stabilne w tych sieciach to takie, dla których $\langle S_i \rangle$ nie zmienia się w czasie.

Teoria pola średniego dla małej ilości wzorców

Dla dowolnego ustalonego p (ilość wzorców) i $N \rightarrow \infty$ (ilość jednostek) możemy określić warunki na stabilność stanów sieci:

Będziemy poszukiwać stanów stabilnych proporcjonalnych do zapamiętanych wzorców: $\langle S_i \rangle = k z_i^n$.

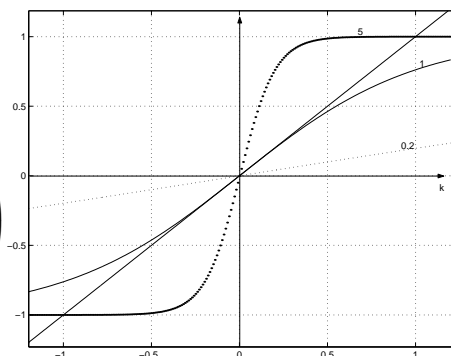
$$\begin{aligned} k z_i^n &= \operatorname{tgh} \left(\frac{\beta}{N} \sum_{j m} z_i^m z_j^m k z_j^n \right) \\ &= \operatorname{tgh} \left(\frac{\beta}{N} \left(N k z_i^n + \sum_{j m \neq n} z_i^m z_j^m k z_j^n \right) \right) \\ &= \operatorname{tgh} (\beta k z_i^n) \end{aligned}$$

ponieważ $\operatorname{tgh}(-x) = -\operatorname{tgh}(x)$, więc mamy:

$$k = \operatorname{tgh}(\beta k)$$

- dla $\beta \leq 1$ istnieje tylko jedno rozwiązanie — stany są stabilne
- dla $\beta > 1$ istnieją trzy rozwiązania — stany są niestabilne

$$\langle S_i \rangle = \operatorname{tgh} \left(\frac{\beta}{N} \sum_{j m} z_i^m z_j^m \langle S_j \rangle \right)$$



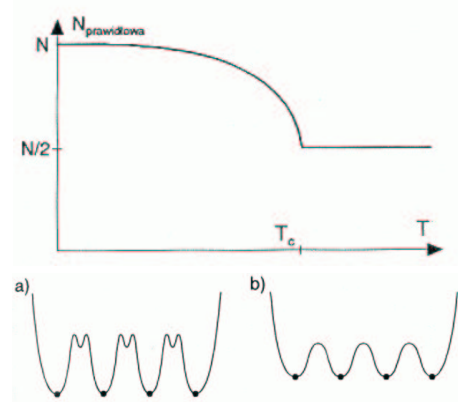
Praktyczne pożytki z poprzednich rozważań

Przez analogię do ferromagnetyków możemy powiedzieć, że dla temperatury krytycznej $T = 1$ następuje przejście fazowe.

- stabilność pamięci w realizacjach technicznych
- W rozważanym przypadku $p \ll N$ występują fałszywe stany: odwrócone i mieszane. Stany mieszane mają swoje temperatury krytyczne.

Najwyższą temperaturę krytyczną ma stan mieszany potrójny: 0.46.

Dla $0.46 < T < 1$ stabilne są tylko stany odtwarzane i odwrócone.



Pojemność pamięci skojarzeniowej

$$\alpha = \frac{p}{N}$$

$$\alpha_c \approx 0.138$$

