Seminar II

Kolektivni efekti nevronskih mrež

avtor : Nejc Košnik
mentor : prof. dr. Rudi Podgornik

Fakulteta za matematiko in fiziko, Univerza v Ljubljani

20. april 2005

Povzetek

Začnemo z biološkim uvodom, nato povemo nekaj o strukturi centralnega živčnega sistema, predvsem nevronskih mrež v možganih. Lastnosti teh mrež nato modeliramo z umetnimi McCulloch-Pitts-ovimi nevroni. Osredotočimo se na asociativni spomin in se ga lotimo s Hopfieldovim modelom dinamike. Ponazorimo stabilnost spominov, pogled s perspektive dinamičnih sistemov ter formulacijo z energijsko funkcijo. Analogija z magnetnimi sistemi pri končni temperaturi omogoča vpeljavo stohastičnega šuma v Hopfieldov model, ki poveča robustnost mreže ter poveča vpogled v posamezne režime delovanja pri pri različnih temperaturah in zasedenostih. Navedenih je tudi nekaj konkretnih zgledov uporabe.

Kazalo

1 Uvod

Sposobnost odzivanja na zunanje dražljaje iz okolja je najbrž najpomembnejša lastnost, ki živim bitjem doprinese zmožnost preživetja in nadaljevanje vrste. Tu seveda mislimo okoljske vplive na posameznega pripadnika vrste, ne na bolj splošne vplive okolja na evolucijsko dinamiko celotne vrste. Pravzaprav je odziv vrste na spremembe v okolju (podnebje, interakcija z drugimi vrstami) v veliki meri odvisna od sposobnosti posameznikov za preživetje, vendar se dogaja na dosti počasnejši skali.

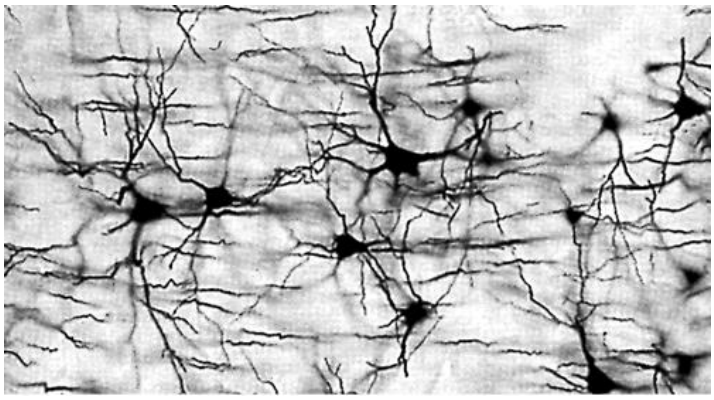
Pri višjih živalih je vlogo odločanja in regulacije celotnega telesa prevzel centralni živčni sistem. Središče tega so možgani, ki sprejemajo velik pretok informacij iz senzornih živčnih vlaken, ter neprestano dirigirajo celemu telesu prek izhodnih živcev. Možgani pa niso le preklopna postaja, temveč so izhodi zapletena funkcija vhodov. Poleg tega se vsak normalen človek zaveda samega sebe, in to samozavedanje biološka in nevroznanost pripisujeta določenemu stanju in procesih v možganih.

Slika 1: Človeški možgani. [4]

1.1 Možgani

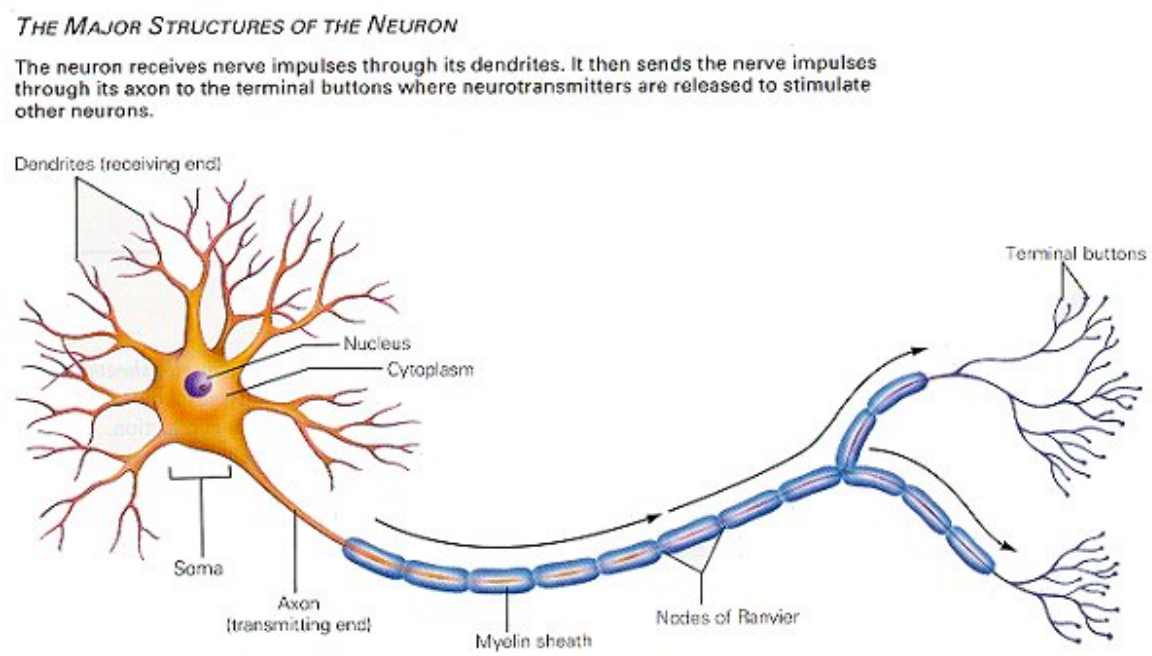
Na pogled (Slika 1) so možgani kilogram in pol težka kepa organske snovi, in taki ne dajejo vtisa zatočišča človeškega duha. Nevroznanost pa je razkrila, da možgane kot tudi celoten živčni sistem sestavljajo živčne celice ali nevroni, katerih osnovna funkcija je prenašanje elektrokemijskih signalov. [1]

Tipični nevron (piramidna celica) sestoji iz celičnega telesa, odkoder se razširja razve-



Slika 2: Mikroskopska struktura možganov. [5]

jena struktura, ki meri v premeru približno 2 mm. To je dendritsko drevo, na katerega so prek živčnih sinaps povezani aksoni drugih nevronov. Živčno vlakno, ki izhaja iz nevrona je akson, in je lahko tudi razvejan. Veje aksona se končajo s sinapsami, ki so kontakt z dendriti drugih nevronov. Tako so možgani v svoji mikrostrukturi velika (10^{11} nevronov) in kompleksno povezana mreža. V možgane prihajajo tudi aksoni iz čutil ter iz možganov se



Slika 3: Shematska slika nevrona. [6]

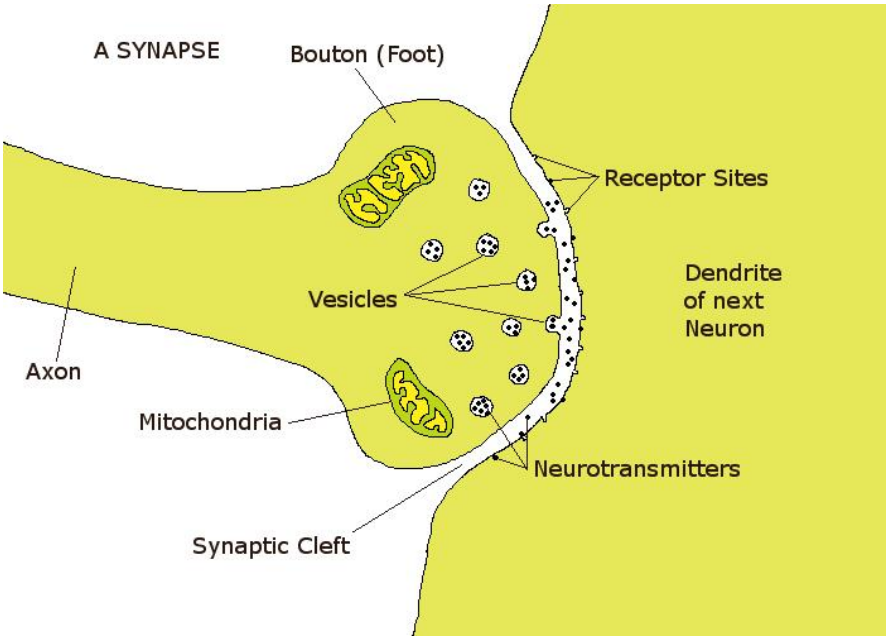
širijo motorični aksoni, vendar je njihovo število zelo majhno v primerjavi s številom povezav v samih možganih. To je tudi namig, da možgani večino svojih sposobnosti porabijo za

notranjo aktivnost, ko "preračunavajo" odgovor na dražljaje. Z znano osnovno strukturo je mogoče ustvariti kolektiv umetnih nevronov, ki se ponašajo z osnovnimi lastnostmi bioloških nevronov. [1]

2 Biološki in formalni nevroni

2.1 Fiziologija nevrodinamike

Glavna naloga vsakega nevrona je sprejemanje električnih signalov od drugih nevronov, "elektrokemijsko" procesiranje ter na koncu signalizacija ostalim nevronom. Ko prihajajoči električni pulzi povzročijo, da se električni potencial nevrona prek membrane dvigne nad vzdražnostni prag, se nevron aktivira in pošlje okrog milisekundo dolg električni pulz vzdolž aksona nato pa se znova vrne v stanje pripravljenosti. Ko pulz doseže sinapse na koncu aksonskega drevesa, se v njih iz veziklov (Slika 4) sprostijo nevrotransmiterji, ki difundirajo skozi sinaptično špranjo do postsinaptične membrane naslednjega nevrona. Tam se nevrotransmiter veže na receptor, kar povzroči odprtje določenih kanalnih proteinov ter s tem povišanje akcijskega potenciala. Takšna sinaptična sklopitev je zaradi dviganja potenciala proti pragu ekscitatorna, poznamo pa tudi inhibitorne sinapse, kjer se zaradi sproščenih nevrotransmiterjev akcijski potencial prek postsinaptične membrane zniža. V sinapsi se ta proces tipično odvija na časovni skali 1 ms, kolikor je tudi dolžina pulza.[1]



Slika 4: Sinapsa ob vzbuditvi. [7]

Tako zbiranje pulzov pulzirajočih predsinaptičnih nevronov v dendritskem drevesu traja približno 10 ms, in če je presežen prag, potem se postsinaptični nevron vzbudi. Da se to zgodi, je potrebno tipično 100 prihajajočih pulzov.[1]

Vendar fiziologija molči o tem, kje se pravzaprav skriva informacija. Za motorične nevrone je jasno, da je informacija kodirana v hitrosti pulziranja, saj le-ti ob vzbuditvi pošljejo celo serijo sunkov. Drugače je z nevroni v možganih, kjer je tipični čas med dvema nastalima pulzoma daljši od akumulacije prihajajočih pulzov. Tu bi lahko pomembno vlogo igrala tudi časovna sinhroniziranost pulzov. V našem formalnem modelu nevrona se s takimi podrobnostmi ne bomo ukvarjali, temveč bomo vsak nevron opisali z njegovo aktivnostjo, kar je še najbližje hitrosti pulziranja.

2.2 Formalni nevroni

Že davnega leta 1943 sta McCulloch in Pitts za model nevrona predlagala diskreten model

$$S_i(t+1) = \theta\left(\sum_j w_{ij} S_j(t) - \mu_i\right), \tag{1}$$

kjer $S_i(t)$ predstavlja mirujoče ($S_i = 0$) ali pulzirajoče ($S_i = 1$) stanje nevrona i ob času t, μ_i je vzdražnostni prag i-tega nevrona, θ pa je Heavysidova stopničasta funkcija. Vidimo, da se utežena vsota po vhodnih nevronih primerja s pragom. Uteži w_{ij} predstavljajo jakosti sinaptične sklopitve za pulze od nevrona j k nevronu i ter so lahko različno predznačene in tako predstavljajo ekscitatorne ali inhibitorne sklopitve. Model (1) vsebuje tudi diskretni čas t, ki poteka enako za vse S_i in tu govorimo o sinhronem osveževanju nevronov.[2]

Generalizacija enačbe (1) je

$$S_i' = g\left(\sum_j w_{ij} S_j - \mu_i\right). \tag{2}$$

Stopničasto funkcijo smo nadomestili z gladko preslikavo, tipično je g naraščajoča in $g : \mathbb{R} \to [0, 1]$. S tem imajo nevroni gladek prehod preko praga μ_i. V modelu tudi ni več eksplicitno izraženega časa, temveč je dinamika asinhrona. To pomeni, da se npr. v enakomernih časovnih intervalih Δt izbere naključni nevron i in zanj izračunamo novo stanje S_i' po enačbi (2).

3 Asociativni spomin

V binarnih računalnikih za dostopanje do podatkov v pomnilniku potrebujemo naslov, ki je povsem neodvisen od vsebine, ki se hrani na tem naslovu. Organski spomin pa najde določene

informacije izključno prek asociacij, torej naslavljanje temelji na vsebini, ki jo iščemo, ne na številkah nevronov, ki se morajo vzbuditi.

Asociativni spomin je "Bohrov atom" nevronskih mrež, in dobro ilustrira, da je mogoče že z zelo grobim modelom naravne strukture ustvariti pomnilnik, ki rekonstruira delne in zašumljene informacije. Problem, ki ga rešujemo se glasi: *V kolektiv umetnih nevronov bi radi shranili vzorce* ξ_i^μ*, kjer* $\mu = 1, \ldots, p$ *označuje vzorce,* $i = 1, \ldots, N$ *pa posamezne bite v vzorcu* μ*. Ob prikazu vzorca* S_i *temu kolektivu se ta postavi v najbolj podobnega izmed naučenih vzorcev* ξ_i^μ*.*

Korespondenca med vzorci in stanji nevronov je sledeča : $S_i = \pm 1$ lahko označuje dve stanji i-tega nevrona. Stanje cele mreže je formalno gledano vektor $\mathbf{S}$ z N komponentami, vendar ga bomo v nadaljevanju označili kar z S_i. Iz konteksta bo jasno, kdaj S_i pomeni stanje posameznega nevrona, in kdaj s tem mislimo na cel vektor $\mathbf{S}$. Mreža z danim številom nevronov lahko realizira 2^N stanj, mi pa bi radi v njo shranili p stanj $\boldsymbol{\xi}^\mu$, $\mu = 1, \ldots, p$. Če potem mrežo postavimo v neko stanje S_i, se bo to stanje s časom relaksiralo v najbližje izmed stanj ξ_i^μ. Definirati moramo še razdaljo med dvema vzorcema, da lahko govorimo o najbližjem stanju. V bitni reprezentaciji ± 1 je ta razdalja [2]

$$d(S_i, \xi_i^\mu) = \frac{1}{2} \sum_{j=1}^{N} \left[1 - S_j \xi_j^\mu \right]. \tag{3}$$

V osebnem računalniku je naloga trivialna - izračunamo razdalje in izberemo shranjeni vzorec, za katerega je ta najmanjša. Tu bi radi enako nalogo opravili z mrežo McCulloch-Pitts-ovih nevronov.

3.1 Hopfieldov model

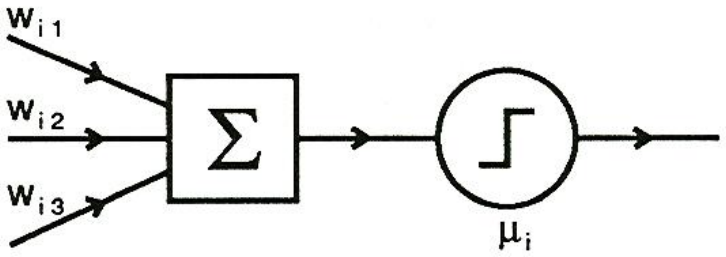
Nevrone imamo lahko v pulzirajočem $S_i = +1$ ali mirujočem $S_i = -1$ stanju, in stanja nevronov predstavljajo bite vzorca. Dinamika je taka, da nevron najprej izvrednoti *lokalno polje*, ki ga povzročajo predsinaptični (vsi) nevroni.

$$h_i = \sum_j w_{ij} S_j \tag{4}$$

Lokalno polje nastopa v argumentu prenosne funkcije, ki je v bitni reprezentaciji ± 1 kar sgn:

$$S_i = \mathrm{sgn}(h_i). \tag{5}$$

Osveževanje stanj nevronov je ponavadi asinhrono [2], tako da v naključnem času izberemo naključni nevron i in zanj izračunamo dinamično pravilo (5). V modelu imamo N^2 prostih parametrov w_{ij}, in ti se uglasijo tekom učenja.



Slika 5: Shematska ponazoritev formalnega nevrona. Prihajajoči signali S_i so z utežmi w_{ij} sešteti v lokalno polje h_i, to pa se potem primerja s pragom μ_i. [2]

3.1.1 Učenje

O procesu učenja v možganih ni veliko znanega, domneva pa se, da zmožnost učenja izhaja iz prilagodljivosti sinaptičnih sklopitev med nevroni. Hebb je tako že leta 1949 predlagal, da se sinaptične sklopitve okrepijo zaradi časovno korelirane aktivnosti pred- in postsinaptičnega nevrona.[1] Hipoteza zasluži še komentar iz gledišča psihologije. Hebbovo pravilo namreč lahko kvalitativno razloži pogojni refleks. Če je med nevronoma A in R dovolj močna sklopitev, potem bo vzbujen A povzročil vzbuditev R. V psihološkem žargonu bi rekli, da dražljaj A izzove reakcijo R. Sedaj pa tekom učenja poleg A vzbudimo še nevron A', ki sicer ne izzove R. Ker A vzbudi R, mi pa vzbudimo še A', sta A' in R istočasno aktivna in po Hebbovem predlogu se bo sinaptična sklopitev med A' in R okrepila. Tako bo po končanem procesu učenja že sam dražljaj A' izzval reakcijo R. Če sedaj v fizikalnem žargonu uvedemo substitucije

$$A \longrightarrow \text{hrana} \tag{6}$$

$$A' \longrightarrow \text{zvonček} \tag{7}$$

$$R \longrightarrow \text{slinjenje psa,} \tag{8}$$

je situacija zelo podobna pogojnemu refleksu v obliki, kot ga je objavil Ivan Pavlov leta 1903.

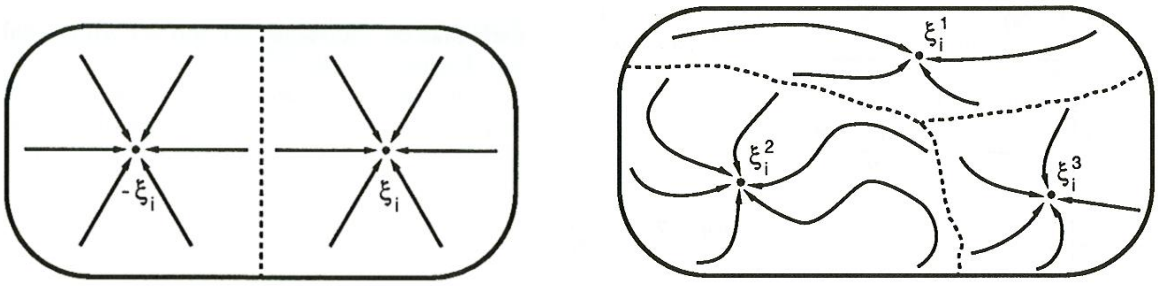
Hebbovo pravilo zapišemo kot

$$\Delta w_{ij} \propto S_i S_j. \tag{9}$$

V skladu z (9) potem zapišemo uteži mreže po enem naučenem vzorcu:

$$w_{ij} = \frac{1}{N} \xi_i \xi_j \tag{10}$$

Normalizacija $1/N$ se izkaže za koristno kasneje. Opisani algoritem učenja navadno štejemo pod okrilje Hopfieldovega modela.



Slika 6: Atraktorja mreže po naučenem vzorcu (levo) in atraktorji po več naučenih vzorcih (desno). [2]

3.2 Stabilnost delovanja

Prvi pogoj za pravilno delovanje asociativnega spomina je stabilnost shranjenega stanja. Če se postavimo v stanje ξ_i, moramo tam tudi ostati:

$$\xi_i = \mathrm{sgn}\,(h_i)\,, \qquad h_i = \sum_j w_{ij}\xi_j \tag{11}$$

Ko vstavimo za uteži Hebbove vrednosti (10), dobimo za lokalno polje

$$h_i = \sum_j w_{ij}\xi_j = \sum_j \frac{1}{N}\xi_i\xi_j\xi_j = \xi_i, \tag{12}$$

od tod pa vidimo, da je stanje ξ_i res stabilno, saj velja $\xi_i = \mathrm{sgn}(\xi_i)$. Poglejmo si še primer, ko je mreža v poljubnem stanju S_i. Polje v i-tem nevronu je takrat

$$h_i = \sum_j w_{ij}S_j = \sum_j \frac{1}{N}\xi_i\xi_j S_j = \frac{1}{N}\xi_i \sum_j \xi_j S_j. \tag{13}$$

Predznačenost vsote odloča, ali bo mreža dosegla stanje $+\xi_i$ ali $-\xi_i$. Če se več kot polovica bitov ujema s shranjenim vzorcem, potem bomo rekonstruirali ξ_i, v nasprotnem primeru pa $-\xi_i$. V kontekstu dinamičnih sistemov lahko rečemo, da ima naša mreža dva atraktorja (Slika 6). Lažni spomin na negiran vzorec ni presenečenje, saj je celoten model invarianten na transformacijo $S_i \to -S_i$.

Za učenje $\mu = 1, \ldots, p$ vzorcev posplošimo Hebbovo pravilo. [2]

$$w_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu \tag{14}$$

Če je dani vzorec ξ_i^ν stabilen razberemo iz lokalnega polja, v katerem izoliramo člen z $\mu = \nu$.

$$h_i^\nu = \xi_i^\nu + \frac{1}{N}\sum_j \sum_{\mu\neq\nu} \xi_i^\mu \xi_j^\mu \xi_j^\nu \tag{15}$$

Prvi člen je enak kot v primeru enega samega naučenega vzorca, pridruži pa se mu še člen, ki predstavlja prepletanje z drugimi naučenimi vzorci. Za vzorce privzamemo da so medsebojno neodvisni in naključni, število vzorcev in velikost mreže pa naj bosta velika $p, N \gg 1$. Ob naštetih predpostavkah je verjetnostna porazdelitev drugega člena kar Gaussova s povprečjem 0 in varianco $\sqrt{p/N}$. Število vzorcev na nevron mora biti torej majhno : $p \ll N$. Zanesljivost spomina lahko zagotovimo z zahtevo, da naj bo vsak bit shranjenega vzorca nestabilen le z neko majhno verjetnostjo P_{err}, ki se izraža kot

$$P_{\text{err}} = \frac{1}{2}\left(1 - \text{erf}\left(\sqrt{N/2p}\right)\right). \tag{16}$$

Od tod dobimo npr. za $P_{\text{err}} = 0.01$ razmerje števila vzorcev proti številu nevronov $p/N = 0.185$ [2]. Dobljene vrednosti p pa je treba jemati kot zgornje meje, saj smo mi postavili le pogoj na prvi dinamični korak v mreži. Tekom evolucije sistema lahko majhne napakice povzročijo plaz nestabilnih bitov, ki uničijo stabilnost vzorca. To se v resnici tudi zgodi, in kasneje bomo z večjim vpogledom pokazali, da so vzorci stabilni le za $p < 0.138N$, kar ustreza $P_{\text{err}} = 0.0037$.

3.3 Pristop z energijo

Hopfieldov model omogoča še formulacijo z energijsko funkcijo. Pokazali bomo, da se funkcija stanja mreže [2]

$$H = -\frac{1}{2}\sum_{ij} w_{ij} S_i S_j \tag{17}$$

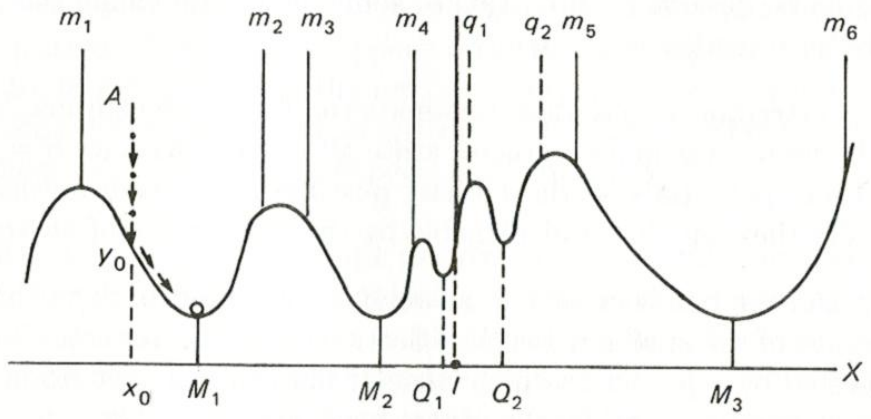
s časovno evolucijo sistema ne povečuje. Iz vsote bomo izločili člene oblike $w_{ii}S_iS_i$, saj ti prinesejo le aditivno konstanto $-p/2$.

$$H = -\sum_{(ij)} w_{ij} S_i S_j \tag{18}$$

Notacija (ij) pomeni seštevanje po različnih parih ij $(i \neq j)$, kjer vrstnega reda ne upoštevamo. Preverimo, da je H nenaraščajoča funkcija časa. Če je po iteraciji na nevronu S_i njegovo stanje enako, je tudi energija enaka. Za drugo možnost, ko se S_i obrne $S_i' = -S_i$, pa izračunamo spremembo energije.

$$\begin{aligned} H' - H &= -\sum_{(ij)} w_{ij} S_i' S_j' + \sum_{(ij)} w_{ij} S_i S_j & (19)\\ &= 2S_i \sum_{j \neq i} w_{ij} S_j & (20)\\ &= 2S_i \sum_{j} w_{ij} S_j - 2w_{ii} < 0 & (21) \end{aligned}$$

V prvem členu sta S_i in vsota nasprotno predznačena zaradi naše predpostavke. S časovno evolucijo se torej bližamo enemu izmed minimumov funkcije H. Ker so shranjeni vzorci stabilni (= časovno neodvisni), so tudi lokalni minimumi funkcije H.



Slika 7: Energija kot funkcija stanja. Mreža se vedno odpelje po klancu navzdol. [3]

Poleg naučenih vzorcev pa ima energija še nezaželene lokalne minimume, ki predstavljajo napačne spomine. Takšni stabilni vzorci so linearne kombinacije lihega števila vzorcev kot je npr. pri treh vzorcih [2]

$$\xi_i^{\text{mix}} = \text{sgn}\left(\pm\xi_i^{\mu_1} \pm \xi_i^{\mu_2} \pm \xi_i^{\mu_3}\right). \tag{22}$$

Stabilnost teh stanj preverimo na enak način, kot smo preverili stabilnost vzorcev ξ_i^μ. Število teh lokalnih minimumov daleč presega število shranjenih vzorcev, saj jih ob upoštevanju vseh možnih predznakov v (22) dobimo $\frac{1}{3}2^{N+3}$.

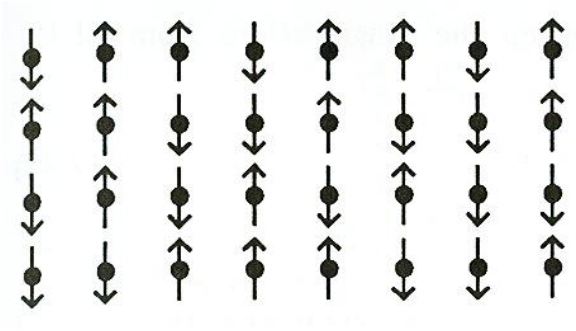
Pri večjem številu shranjenih vzorcev p pa se pojavijo še dodatna stabilna stanja, ki nimajo nobene korelacije z vzorci, ki naj bi bili zapisani v pomnilniku[1] [2]. Obe vrsti lažnih stanj predstavljata resno oviro pri praktični uporabi asociativnega spomina, zato bomo v nadaljevanju omenili, kako se jim je moč izogniti.

4 Magnetni sistemi in stohastične mreže

Tu bomo pokazali, da je možno problem asociativnega spomina obravnavati v širšem kontekstu Isingovega modela. Ta analogija omogoča statističnomehanske obravnavo, kjer lahko deterministični mreži dodamo temperaturne fluktuacije.

Opazujemo sistem N spinov z $s = 1/2$. V tem primeru sta možni stanji $S_i = \pm 1$, kot pri

[1] Ta stanja so analogna fazi spinskih stekel pri obravnavi neurejenih magnetnih sistemov.



Slika 8: Isingov model v dveh dimenzijah. [2]

Hofieldovem modelu. Lokalno magnetno polje na mestu i-tega spina je

$$h_i = \sum_j w_{ij} S_j + h^{\text{ext}}. \tag{23}$$

w_{ij} je tokrat izmenjalni integral med spinoma i in j in velja $w_{ij} = w_{ji}$. Ponavadi izmenjalne integrale w_{ij} med nesosednimi spini postavimo na nič, tokrat pa obdržimo splošnejšo obliko, kjer interagirajo vsi spini. w_{ii} postavimo na nič, h^{ext} pa je zunanje polje, ki ga bomo od tu naprej izpustili iz obravnave.

Ob odsotnosi termičnih fluktuacij, pri $T = 0$, velja ob prisotnosti polja h_i

$$S_i = \text{sgn}(h_i), \tag{24}$$

in analogija s Hopfieldovim modelom je vzpostavljena. Končna temperatura prinese fluktuacije, in sicer postane zveza med lokalnim poljem in stanjem spina nedeterministična. Govorimo o verjetnostih, da spin kaže gor ali dol.

$$P(S_i = \pm 1) = f_\beta(\pm h_i) = \frac{1}{1 + \exp(\mp 2\beta h_i)} \tag{25}$$

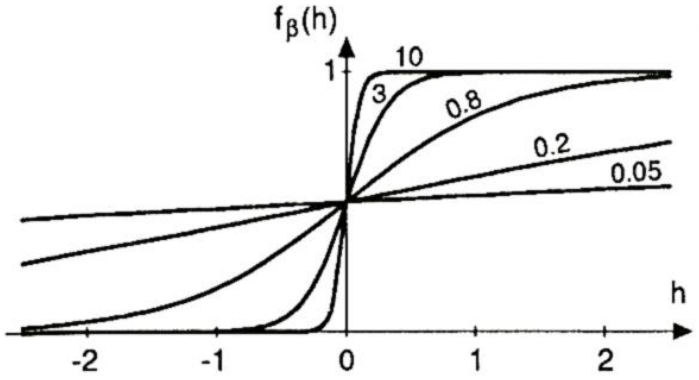
Funkcija f_β je prikazana na Sliki 9. Temperaturo merimo v enotah Boltzmannove konstante k_B.

Pri končni temperaturi nobeno stanje ni več stabilno, v termodinamičnem ravnovesju so konstantne le časovna povprečja količin. Poglejmo si en sam spin S v polju h v termodinamičnem ravnovesju.

$$\langle S \rangle = P(S = +1) - P(S = -1) = \tanh(\beta h) \tag{26}$$

V teoriji povprečnega polja polje h_i nadomestimo z izpovprečenim prispevkom k polju drugih spinov.

$$\langle h_i \rangle = \sum_j w_{ij} \langle S_j \rangle \tag{27}$$



Slika 9: Prenosna funkcija spina za različne temperature $T = 1/\beta$. [2]

Izkaže se, da je teorija povprečnega polja dober približek le za interakcije dolgega dosega, saj je takrat število členov v vsoti veliko in lahko uporabimo centralni limitni teorem. Za magnetne sisteme je doseg spinske interakcije zelo omejen in je tako $w_{ij} \neq 0$ le za najbližje sosede, pri nevronih pa je sistem bolj prepleten in tam bo teorija povprečnega polja dala prave rezultate. [2] V ravnovesju torej velja

$$\langle S_i \rangle = \tanh(\beta \langle h_i \rangle). \tag{28}$$

4.1 Feromagnet

Pri feromagnetih je $w_{ij} > 0$, mi si izberimo kar najpreprostejši približek interakcij dolgega dosega.

$$w_{ij} = \frac{J}{N} \tag{29}$$

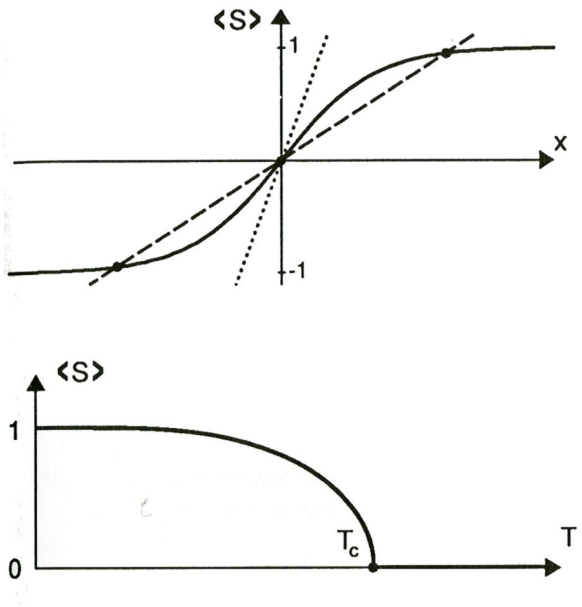
Vemo tudi, da so feromagneti pod temperaturo prehoda enakomerno namagneteni ter da spini v povprečju kažejo v isto smer:

$$\langle S_i \rangle = \langle S \rangle \tag{30}$$

Takoj opazimo, da je takšen model pri $T = 0$ in $J = 1$ ekvivalenten Hopfieldovem spominu z enim naučenim vzorcem $\xi_i = 1$. Enačba ravnovesnega stanja v povprečnem polju se bistveno poenostavi

$$\langle S \rangle = \tanh(\beta J \langle S \rangle) \tag{31}$$

in odraža zvezni fazni prehod.



Slika 10: Zgoraj grafično reševanje enačbe (31), spodaj temperaturna odvisnost magnetizacije. [2]

4.2 Stohastične mreže

Da lahko zares uporabimo statistično mehaniko na nevronih, moramo izpolniti še zahtevo o obstoju ravnovesnega stanja, ki ga sistem doseže. Govorimo namreč o ravnovesni statistični mehaniki. Izkaže se, da se mreža v primeru simetričnih sklopitev $w_{ij} = w_{ji}$ vedno znajde v ravnovesnem stanju. Ravnovesno stanje pri $T > 0$ seveda razumemo kot nespremenljivost $\langle S_i \rangle$.

Zaenkrat bomo računali v režimu $p \ll N$. Enačbe povprečnega polja so

$$\langle S_i \rangle = \tanh\left(\frac{\beta}{N}\sum_{j,\mu}\xi_i^\mu \xi_j^\mu \langle S_j \rangle\right). \tag{32}$$

Če mreža rekonstruira vzorec ξ_i^ν, potem po vzoru feromagneta napišemo ansatz v ravnovesnem stanju.

$$\langle S_i \rangle = m\xi_i^\nu. \tag{33}$$

Vstavimo ga v enačbe.

$$m\xi_i^\nu = \tanh\left(\frac{\beta}{N}\sum_{j,\mu}\xi_i^\mu \xi_j^\mu m\xi_j^\nu\right) \tag{34}$$

$$= \tanh\left(\beta m\xi_i^\nu + \frac{\beta m}{N}\sum_{j,\mu\neq\nu}\xi_i^\mu \xi_j^\mu \xi_j^\nu\right) \tag{35}$$

Drugi člen v argumentu hiperboličnega tangensa je reda p/N, tako da ostanemo le s prvim in dobimo

$$m\xi_i^\nu = \tanh\left(\beta m \xi_i^\nu\right). \tag{36}$$

ξ_i^ν je lahko le ± 1, zato ga lahko nesemo iz argumenta in končno dobimo

$$m = \tanh\left(\beta m\right). \tag{37}$$

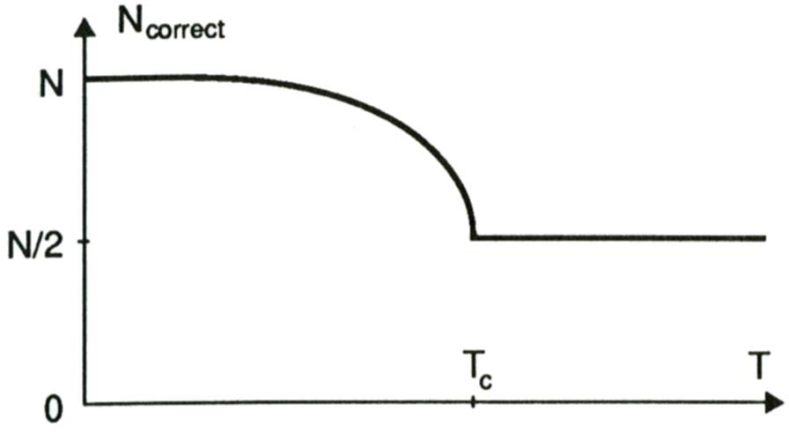
m predstavlja magnetizacijo v smeri vzorca ξ_i^ν in očitno ima enak fazni prehod kot magnetizacija feromagneta pri $T_c = 1$. m pa lahko prevedemo na povprečno število pravilnih bitov.

$$\begin{aligned} m &= \frac{\langle S_i \rangle}{\xi_i^\nu} && (38)\\ &= \frac{P(S_i = +1) - P(S_i = -1)}{\xi_i^\nu} = P(S_i = \xi_i^\nu) - P(S_i \neq \xi_i^\nu) && (39)\\ &= 2P(S_i = \xi_i^\nu) - 1. && (40) \end{aligned}$$

Od tod izrazimo število pravilno postavljenih bitov v ravnovesnem stanju.

$$\langle N_{\text{pravilni}} \rangle = \frac{1}{2} N(1 + m) \tag{41}$$

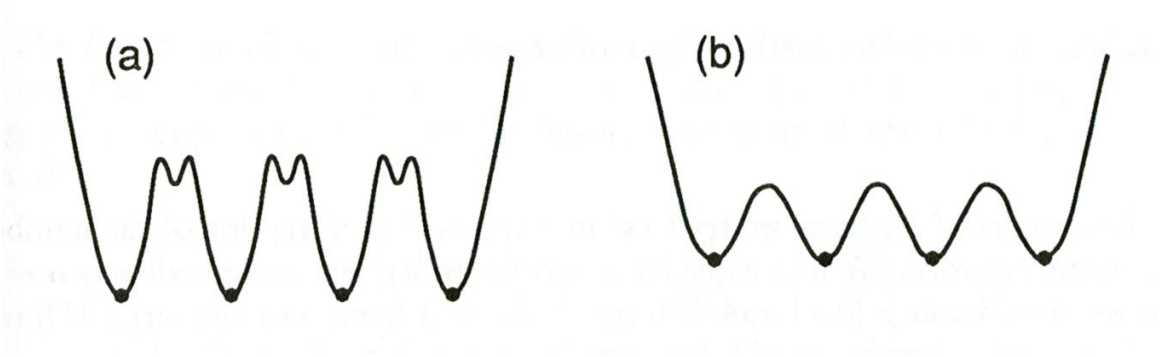
Fazni prehod na Sliki 11 kaže, da so nad kritično temperaturo biti postavljeni povsem naključno in tako dosegajo 50% pravilnost. Temperaturni šum, ki smo ga uvedli v Hopfieldov



Slika 11: Fazni prehod stohastične mreže za $p \ll N$. [2]

model, lahko ponazarja šum v bioloških nevronskih mrežah, saj tam nevroni ne pulzirajo vedno z enako hitrostjo, prihaja do naključnih faznih zakasnitev in še mnogo drugih pojavov je, ki jih deterministični model ne opiše. Izkazalo se je, da mreža ob zmernem šumu deluje precej dobro, nad kritično temperaturo pa popolnoma odpove.

V limiti $p \ll N$ stanja spinskih stekel niso pomembna, še vedno pa imamo lažne spomine (lokalni minimumi) na inverzne vzorce $-\xi_i^\mu$ ter linearne kombinacije vzorcev. Te linearne kombinacije imajo v splošnem manj stabilno lego v "energijski pokrajini" in izkaže se, da so v temperaturnem območju $0.46 < T < T_c = 1$ njihovi minimumi preplitki (Slika 12), da bi mreža skonvergirala vanje. Stohastična mreža je torej v tem pogledu bolj robustna.



Slika 12: Za $T < 0.46$ (levo) lahko sistem pristane v enem od plitvih minimumov, pri večjih temperaturah pa jih sistem niti ne opazi (desno). [2]

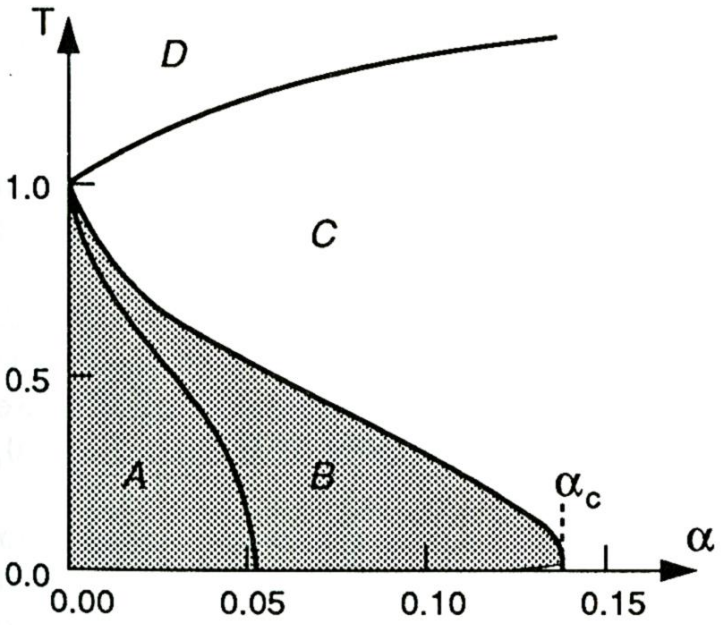
4.3 Kapaciteta stohastične mreže

Definiramo *zasedenost* mreže kot $\alpha = p/N$. Dosedaj smo računali v limiti $\alpha \ll 1$, tokrat pa v režimu $\alpha \sim 1$. Omenimo le, da v rekonstrukciji vzorca ξ_i^μ ne smemo zanemariti člena prepletanja z drugimi vzorci, in da celoten račun presega okvir tega seminarja. Bralec ga najde v [2].

Rezultat je fazni diagram, kjer vlogo ureditvenega parametra prevzame zasedenost mreže α (Slika 13). Iz njega tudi preberemo, da je pri $T = 0$ - to je pri deterministčnem Hopfieldovem modelu $\alpha_c = 0.138$ in da se nad to zasedenostjo spomin sesuje. V območju C so stabilna le stanja spinskih stekel, ki so za spomin povesem neuporabna, v D pa je stabilno le stanje $\langle S_i \rangle = 0$. V območju A so globalni minimumi shranjeni spomini, v B pa so spinska stekla nižje ležeča od naših vzorcev. Poleg tega imamo stabilna mešana stanja shranjenih vzorcev na območju $T < 0.46$ in $\alpha < 0.03$.

5 Zaključek

Skozi seminar smo si pridobili poglobljeno razumevanje delovanja nevronskih mrež, s poudarkom na asociativnem spominu. Tu so nam bistveno pomagala orodja statistične fizike, s katero smo dognali, kako bi spomin naredili bolj robusten in s tem uporabnejši. Seveda smo



Slika 13: Fazni diagram stohastične mreže. Osenčena področja prikazujejo, kje je spomin sploh še stabilen. [2]

ilustrirali le najosnovnejše trike, ki se jih poslužujejo aplikacije nevronskih mrež.

Bistvena prednost takih mrež je, da nealgoritmično prepoznajo delne in zašumljene informacije (Slika 14). Algoritmična formulacija takega problema je navadno nemogoča ali pa vsaj mnogo težja od mimiciranja nevronskih struktur, ki jih najdemo v naravi. Zato se takšne mreže ekstenzivno uporabljajo pri prepoznavanju obrazov, govora, prepoznavanju tarč na sonarju, ugotavljanju sekundarne strukture proteinov in na ostalih zašumljenih informacijah, kjer je potrebna velika robustnost.

Literatura

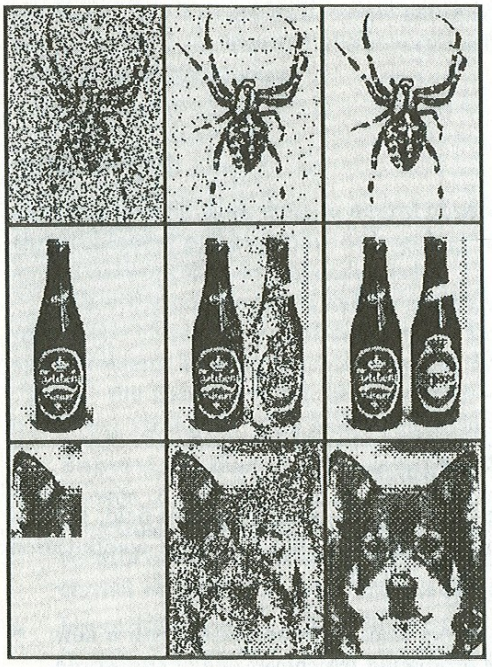
[1] H. Horner and R. Kuhn, *Neural Networks*, `cond-mat/9705270`, (1997)

[2] J. Hertz, A. Krogh, R. G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley Publishing Company, (1991)

[3] D. J. Amit, *Modeling Brain Function*, Cambridge University Press, (1989)

[4] *Wellesley College Homepage*, `http://www.wellesley.edu/Chemistry/Chem101/brain/`

[5] *The Turing Archive for the History of Computing*,



Slika 14: Rekonstrukcija spomina iz delne informacije. V zgornjem primeru gre za zašumljenost, v spodnjih dveh pa za nepopolnost vzorca. [2]

http://www.alanturing.net/turing_archive/

[6] http://www.sciforums.com/showthread.php?t=43222

[7] *George Boeree's Homepage*, http://www.ship.edu/~cgboeree/