

Eötvös Loránd Tudományegyetem
Természettudományi Kar

Neurális rendszerek funkcionális modellezése

Diplomamunka

Takács Bálint
fizikus hallgató

témavezető: dr. habil. Lőrincz András
tudományos főmunkatárs
ELTE Információs Rendszerek Tanszék

Budapest, 2000.

Köszönetnyilvánítás

...

Tartalomjegyzék

Bevezetés	1
1. A faktoriális kódolás	2
1.1. A szenzoros kódolás általános célja	2
1.1.1. A szenzoros modell	3
1.1.2. A kódolás algoritmus	4
1.2. A faktoriális kódolás	5
1.2.1. A független komponens analízis (ICA)	6
1.3. Kódolás a látókéregben	8
2. A hippocampális formáció biológiája	15
2.1. Anatómia	15
2.2. Etológia	17
2.3. Neurofiziológia	17
3. Elképzelések a hippocampusz működéséről	18
3.1. Az elméletek rövid ismertetése	18
3.1.1. A hippocampusz mint érzelmi központ	18
3.1.2. A hippocampusz mint prediktív komparátor	19
3.1.3. Konfigurális asszociációs elmélet	20
3.1.4. Munkamemória	21
3.1.5. Kognitív térképek	22
3.1.6. Kontextuális felidézés	25
3.1.7. Epizodikus memória	27
3.2. Az elméletek összehasonlítása	30
3.3. Nem tájékozódással kapcsolatos funkciók	30
3.4. Akkor tehát mit csinál a hippocampusz?	31
4. Faktoriális kód a hippocampuszban?	32
4.1. Érvék a faktoriális kódolás mellett	32
4.1.1. Szekvenciák tárolásának problémái	32

TARTALOMJEGYZÉK

4.1.2. A minimalizált kölcsönös információ előnyei	33
4.2. A modell áttekintése	34
4.2.1. Tanulási szabályok	35
4.2.2. Anatómiai megfeleltetések	36
4.2.3. További bizonyítékok	38
5. Az ICA kiterjesztése	41
5.1. A független komponens analízis egy fontos problémája	41
5.2. Egy lehetséges megoldás (T-ICA)	42
5.3. Biológiai háttér	43
5.4. T-ICA a látókéregben	44
6. A T-ICA alkalmazása egy egyszerű modellen	46
6.1. Az input modellje	46
6.2. Az alkalmazott módszerek	49
6.3. Eredmények	49
6.3.1. Az ICA place sejteket alakít ki	49
6.3.2. A kialakuló szűrők labirintusfüggők	50
6.3.3. A T-ICA szűrők irányfüggősége	50
6.3.4. A T-ICA aktivitások lokalizáltabbak	51
6.3.5. Az eloszlások vizsgálata	51
7. Összefoglalás és kitekintés	52
Irodalomjegyzék	54

Bevezetés

Bármilyen meglepően hangzik is, a számítástechnika rohamos fejlődésével a neurológia is új eszközt kapott a kezébe. Manapság egyre gyakoribb, hogy neurális rendszerek vizsgálatánál szimulációkat is felhasználnak, és a modellek által produkált eredményekből biológiai következtetéseket próbálnak levonni. Természetesen ez az ötlet nem új: a mesterséges neuronhálózatok története, amely idestova fél évszázadra nyúlik vissza, kezdetektől fogva összefonódik a biológiai alkalmazásokkal. Manapság kezd realitássá válni, hogy a fordított irányban is működhet a dolog. Agyunkat tekinthetjük komplex információfeldolgozó rendszernek, amely összetettségében minden ember által készített berendezést meghazudtol. Nem ez lenne az első eset, hogy az evolúció ötletei segítséget adnak a mesterséges rendszerek tervezéséhez. A két téma, a neurobiológia és az információs rendszerek elméletének összefonódásából kialakult *komputációs neurobiológia* ebben az évtizedben számos sikert tud felmutatni. Ilyen például a látókéreg egyszerű sejtjeinek modellezése, amiről majd szó fog esni. Mint azt látni fogjuk, ez a példa azt mutatja, hogy bizonyos esetekben jogunk van az agyat az információt optimális eszközökkel feldolgozó rendszernek tekinteni, és ezen keresztül az idegrendszer stratégiái között az információelméletből vett hatékony eljárásokat keresni.

A dolgozatban a faktoriális kódolás alkalmazhatóságát vizsgálom a hippocampusban. Ehhez az első fejezetben áttekintem a faktoriális kódolás rövid történetét, majd a második és a harmadik fejezetben nagyon röviden összefoglalom, mit tudunk a hippocampusról, különös tekintettel azokra a tényekre, amiket majd fel szeretnék használni. A negyedik fejezetben egy olyan modellt ismertetek, ami faktoriális kódolást alkalmaz a hippocampusban, és kitérek ennek az elképzelésnek a problémáira és előnyeire. Az ötödik fejezet a modell egy bizonyos szempont szerinti kritikáját és a problémák egy lehetséges megoldását tartalmazza. Saját eredményeimet jórészt a hatodik fejezetben mutatom be, ahol szimulációim eredményei találhatóak. Végül az utolsó fejezetben az elmondottak összefoglalását adom és kitérek néhány lehetséges továbblépési irányra is.

1. fejezet

A faktoriális kódolás

1.1. A szenzoros kódolás általános célja

Habár egyre többet tudunk arról, hogy az agy miként kódolja a szenzoros információkat, még mindig komoly vita folyik arról, hogy beszélhetünk-e a kódolás valamilyen értelemben vett általános céljáról. Sokan úgy tartják, hogy mivel az állatnak¹ feladatok nagyon széles skáláját kell megoldania, nem szükségszerű, hogy bármilyen közös vonást találjunk a különböző feladatokra kifejlesztett kódolási stratégiák között. Mások úgy gondolják – és a dolgozatban én is ezt az elvet fogom követni –, hogy lehetséges a szenzoros kódolás leírása egy általános információfeldolgozási stratégia keretében. Emögött az a feltételezés bújik meg, miszerint az észlelés különböző területeiről származó információk közös problémája a redundancia. Az elképzelés szerint tehát a szenzoros kódolás közös stratégiája az észleletekben meglévő redundancia csökkentése, ezzel jelentősen megkönnyítve a további feldolgozást.

A redundancia, ha a shannoni definíciónál tágabb, hétköznapi értelemben használjuk, nem egyértelmű fogalom. Sőt, még az sem igazán világos, hogy minek szeretnénk a redundanciáját csökkenteni, azaz a szenzoros input milyen modelljére is vonatkoznak a fentiek. Tegyük fel, hogy szeretnénk hangot kódolni. Szenzoros információnak tekinthetjük ez esetben a levegő rezgéseinek amplitúdójában jelentkező függvényt. Attól függően, hogy mit tudunk erről a függvényről, sokféle stratégia közül választhatunk. Például ha hangvillát hallgatunk, akkor az input egy stabil szinuszjel. A modell inputja ebben az esetben egy folytonos matematikai változó, és a hatékonynak nevezhető kódolás három valós paraméter meghatározásával történhet,

¹A továbbiakban „állat” alatt elsősorban kistestű rágcsáló emlősökre - például patkányra - gondolok, hiszen a témában végzett kísérleteknek elsősorban ők a főszereplői. Természetesen az elmondottak változtatásokkal vagy anélkül gyakran érvényesek más élőlényekre, például az emberre is, lásd pl. 27. oldal.

amik a frekvenciát, fázist és az amplitúdót tartalmazzák. Ha természetes hangról van szó, akkor a feladat jóval általánosabb. Megfontolhatjuk például az eredeti jel tárolását és tömörítését valamilyen alkalmas diszkrét mintavételezéssel. Ha a jel zajos, zajsűrűséssel sokat javíthatunk a tömörítés hatásfokán, és csökkenthetjük az eredeti jel redundanciáját. Máris más értelemben használjuk a redundancia fogalmát. A redundancia itt ugyanis nem az adattárolás hatékonyságában jelentkezik, hanem abból az a priori ismeretünkből adódik, hogy a jel valamilyen struktúrával rendelkezik. Ezt a struktúrát mondjuk a frekvenciatérben végzett simítással ki tudunk választani. Ha nagyon zajos a jel, érdemes lemondani az eredeti mintavételezett jel pontos tárolásáról, és veszteséges kódolásokat alkalmazni, amit a jel statisztikai tulajdonságaiból próbálunk optimalizálni. De próbálhatunk modelleket gyártani a fül utáni feldolgozó rendszer kódolási stratégiájáról is. A bemenet ekkor rendkívül zajos sejteken mért feszültség lehetne, amik egyenként az eredeti input frekvenciaspektrumának kis részét reprezentálják. Ez esetben több folytonos elemünk képezi a rendszer bemenetét, amiket célszerű egy valószínűségi változókból álló vektorban összefoglalni. Láthatjuk tehát, hogy a szenzoros inputról és a redundanciáról alkotható modelleknek rendkívül széles a spektruma. Azonban mi biológiai relevanciával rendelkező modellt keresünk. A következőkben megvizsgáljuk, miként szűkíti ez a lehetséges modellek körét.

1.1.1. A szenzoros modell

Tudjuk, hogy az idegrendszer inputja zajos. Az is ismert, hogy maga az idegrendszer is rendkívül zajos és egységeinek működése a nem részletező leírások (mint például a McCulloch-Pitts féle neuronmodell) szempontjából nem-determinisztikus. Az idegrendszer rendkívüli formagazdagsággal rendelkező információkat dolgoz fel, tehát az eredeti input alakjáról a legtöbbször csak nagyon keveset tételmezhet fel. Olyan stratégiákat kell tehát megfontolnunk, amelyek a bejövő információt valószínűségi változóknak tekintik és azt statisztikai alapon próbálják kódolni. Ezzel remélhetőleg kezelni tudjuk a nem-részletező neuronmodellek számos problémáját is, mert statisztikai tárgyalással nem állítunk többet a valódi sejtről, csak elismerjük, hogy kevesebbet tudunk róla. Az idegrendszeren belül az információ paralell módon, sejtrétegekben terjed. Leírásomban a szenzoros információ tehát valószínűségi változókból álló vektorokban jelenik meg.² A biológiai párhuzamokban többnyire frekvenciakódolást tételmezek fel. Ez azt jelenti,

²Ha az ellenkezőjére külön nem térek ki, felteszem, hogy a megjelenő jelek a folytonos \leftrightarrow diszkrét átmenet szempontjából „jól” viselkednek, azaz az input folytonos lesz folytonos mintavételezés esetében.

hogyan a valós tengelyt a neuronok tüzelési frekvenciájára képezzük le, az átlagos tüzelési gyakoriságot a nullának megfelelően. Ez lehetővé teszi a gátló szinapszisok negatív súlyokként történő értelmezését.

1.1.2. A kódolás algoritmus

A szenzoros információ fenti képe jól illeszkedik a mesterséges neuronháló elméletéhez, érdemes tehát neurális algoritmusokat keresnünk, már csak a biológiai relevancia miatt is. Tudjuk, hogy neurális rétegekkel elsősorban lineáris transzformációkat könnyű megvalósítani. Emellett az egyes elemek alkalmazhatnak nemlineáris szűrőket is. Mivel biológiai háttérrel is próbálunk keresni, lényeges, hogy az algoritmusok lokális (hebbi) tanulást használjanak. [MI 1999]

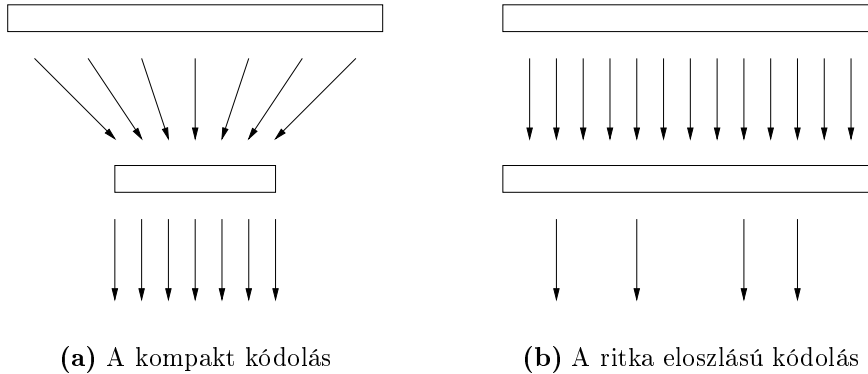
Hogyan értelmezzük a redundanciát, azaz mi legyen a kódolás célja? Erre vonatkozólag több ötlet is felmerül. Kettőt mutat az 1.1 ábra. Ezek bizonyos redundancia-fogalom szerint optimális lineáris transzformációt próbál megtalálni. Az első a *kompakt kódolás* ötlete. A kitűzött cél ez esetben az input vektor dimenziószámának csökkentése egy projekción keresztül. A projekciót úgy próbáljuk megválasztani, hogy a transzformáció után kapott vektorból minimális négyzetes hiba mellett tudjuk helyreállítani az eredeti inputot. Ezen kódolás tehát a redundanciát a kódoló elemek számában keresi, és az eredeti inputot *minimális számú elemmel* próbálja reprezentálni.

Ha a projekció lineáris transzformáció, a feladat egy lehetséges megoldását a főkomponens analízis (principal component analysis, PCA) adja. Nem célom a PCA módszer teljes ismertetése, de vázlatosan összefoglalom a lényegét. A PCA módszer a valószínűségi változók korrelációs mátrixának sajátvektorait határozza meg, és az így kapott bázisban reprezentálja az inputot. Az egyes komponensek varianciája így maximális lesz. A kis varianciájú komponensek elhagyásával a négyzetes hibában optimális projekciót kapjuk. Ha az input réteg dimenziószáma L és a kimeneti réteg dimenziószáma M , akkor a PCA projekciós mátrixát legegyszerűbben a következőképp kaphatjuk meg:

$$V = D^{-\frac{1}{2}} E^T$$

itt

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_M \end{pmatrix}$$



1.1. ÁBRA. Két módszer a redundancia csökkentésére

$$E = (\mathbf{c}_1 \mid \mathbf{c}_2 \mid \dots \mid \mathbf{c}_M)$$

ahol V az $M \times L$ -es projekciós mátrix, E egy $L \times M$ -es mátrix, ami az oszlopaiban az input vektorok korrelációs mátrixának sajátvektorait tartalmazza. A $M \times M$ -es mátrix főátlójában pedig ugyanezen korrelációs mátrix megfelelő sajátértékei találhatóak. Két fontos tulajdonságát kell megemlítenünk a PCA-nak: a transzformáció bázisa, mivel a szimmetrikus korrelációs mátrix sajátvektorairól van szó, mindig ortonormált. Ez bizonyos problémákat vet fel, lásd 1.2. ábra. Az output vektorok korrelációs mátrixa egységmátrix lesz. A PCA módszernek számos változata és kiterjesztése létezik. Transzformációs mátrixa lokális neurális tanulással is képezhető, lásd pl. [Karhunen et al. 1996].

A ritka eloszlású kódolás (sparse representation) ötletét mutatja a 1.1(b). ábra. A transzformáció itt azt próbálja elérni, hogy a kimeneti réteg elemei közül *minél kevesebb legyen egyszerre aktív*. A kódolás tehát nem feltétlen csökkenti a reprezentáló elemek számát, sőt, az outputot gyakran nagyobb dimenziósra választják az eredetinel [Olshausen and Field 1996].

1.2. A faktoriális kódolás

A faktoriális kódolás alapötlete az, hogy az inputot olyan transzformációnak vessük alá, hogy a transzformált vektorváltozó együttes valószínűségeloszlása az egyes komponensei peremeloszlásának szorzataként álljon elő:

$$p(s_1, s_2, \dots, s_n) = \prod_{i=1}^n p(s_i)$$

Ezzel a legalacsonyabb redundanciát a legmagasabb entrópiájú eloszlásban találjuk meg. Látható, hogy ez bizonyos szempontból rokon a ritka eloszlású kódolással, hiszen azzal, hogy a komponensek közti függőséget minimalizáljuk, a köztük fennálló kölcsönös információt maximalizáljuk, és ezáltal a legkevesebb lehetséges elemmel próbáljuk az eredeti inputban rejlő információt reprezentálni. A két kódolás közti kapcsolat egzakt matematikai pontossággal is belátható.³

1.2.1. A független komponens analízis (ICA)

A faktoriális kódolás egy lehetséges megvalósítása a független komponens analízis. Ez az eljárás az input egy olyan lineáris transzformációját adja meg, ami minimalizálja az output réteg elemei között fennálló kölcsönös függőséget. Ebben az értelemben tehát a PCA egy kiterjesztésének tekinthető (ami csak az output elemek mint valószínűségi változók második momentumában teszi ezt, míg az ICA minden momentumot egyszerre próbál minimalizálni). Az ICA bázisvektorai nem feltétlen ortogonálisak. A PCA és az ICA közti különbséget szemlélteti az 1.2. ábra.

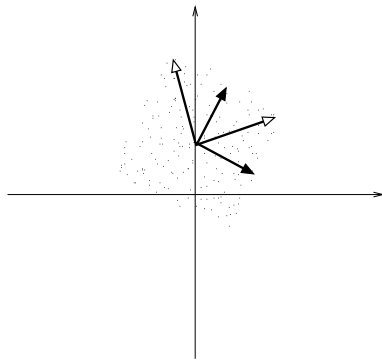
Az ICA algoritmusát eredetileg hangfeldolgozási problémák megoldására vezették be. Képes ugyanis „vakon” megoldani hangforrások szeparációját, pusztán statisztikai alapon (a feladat neve angolul Blind Source Separation vagy BSS). A probléma pontosabban a következő: adott N darab hangforrás és L darab mikrofon egy szobában. Minden egyes mikrofon a hangforrások egy állandó együtthatójú lineáris keverékét érzékeli. Az együtthatókat egy $M \times L \times N$ -es keverőmátrixban összefoglalva azt írhatjuk:

$$\mathbf{x}_t = \mathbf{M}\mathbf{s}_t$$

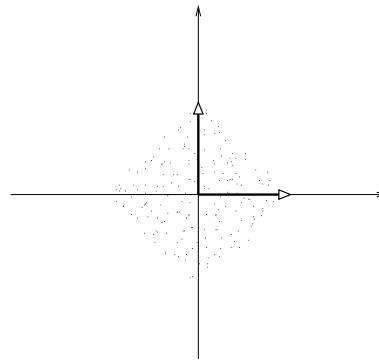
ahol s_t a t -edik időpillanatban a hangforrások által kibocsátott hang amplitúdója, diszkrét mintavételezést feltételezve. Az eredeti hangforrások történetét visszakaphatjuk a mikrofonokon rögzített jelből, ha azt az M mátrix inverzével vagy pszeudoinvertéval szorozzuk, természetesen ez utóbbi esetben – amikor kevesebb mikrofon van, mint hangforrás – csak közelítőleg. A felvétel során érvényben levő M mátrixot azonban általában nem ismerjük, ezért kell a szétválasztást „vakon” elvégeznünk. Ha igaz, hogy az eredeti hangforrások függetlenek voltak, akkor speciális esetektől eltekintve az ICA algoritmusát az x_t idősorra alkalmazva éppen ennek a mátrixnak az inverzét kapjuk meg, anélkül, hogy az eredeti mátrixról bármit is tudnánk.

A kilencvenes években az ICA számos megvalósítását találták ki, pl. [Belouchrani et al. 1997] [Hyvärinen et al. 1997]. Ezek közt számos neurális

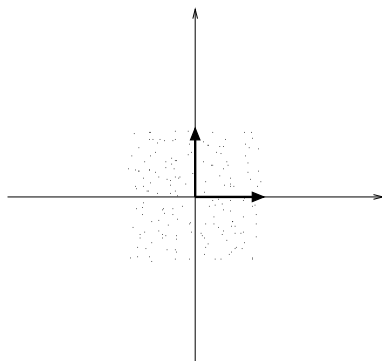
³Lásd még 13. oldal.



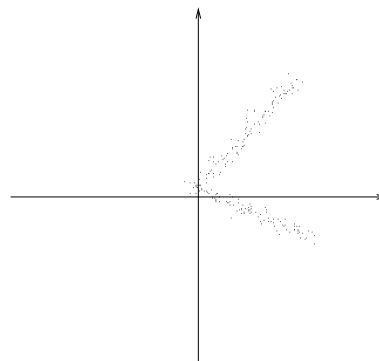
(a) Az eredeti adatok



(b) A PCA eredménye



(c) Az ICA eredménye



(d) Nem-ortogonális struktúrájú eloszlás

1.2. ÁBRA. PCA és ICA egy egyszerű kétdimenziós eloszlás példáján. Az első ábra az eredeti eloszlást ábrázolja. Az inputvektor minden mintája egy négyzet alakú területen belül helyezkedik el. Az ábrába berajzoltam a PCA és az ICA által megtalált bázisvektorokat. A második ábrán az eredeti input transzformáltját láthatjuk a PCA által megtalált bázisban, a harmadik ábrán pedig az ICA által megtalált bázisban. Ez utóbbi nem általános eset abban az értelemben, hogy az ICA bázisvektorai nem feltétlen ortogonálisak. Ez komoly előnyt jelenthet például az utolsó ábrán felrajzolt input esetében, ahol az eloszlás jellegzetes tengelyei sem ortogonálisak. Megjegyzendő, hogy a PCA-nak léteznek nemortogonális kiterjesztései, lásd pl. [Karhunen et al. 1996]

is van, pl. [Karhunen et al. 1996]. Egy másik lehetséges osztályzás történhet aszerint, hogy az implementáció adaptív(on-line) vagy nem adaptív(off-line). Az utóbbiaknak az induláskor az összes mintára szükségük van. Ez használható egy hangfeldolgozási problémánál, de biztosan nem, ha biológiai háttérrel keresünk. Az adaptív algoritmusok előnye, hogy általuk a transzformáció visszacsatolt hurok eleme lehet, hátrányuk viszont a gyenge stabilitás.

Nézzünk meg közelebbről egy biológiai szempontból ígéretes kétrétegű implementációt [Karhunen et al. 1996]. Az első rétegben az úgynevezett fehérítés, a második rétegben a szeparálás zajlik. A fenti adatmodellben ez a következőképp néz ki:

$$\mathbf{v}_t = \mathbf{V}\mathbf{x}_t$$

$$\mathbf{y}_t = \mathbf{W}^T\mathbf{x}_t$$

ahol \mathbf{V} egy $M \times L$ -es fehérítő, a \mathbf{W} pedig egy $M \times M$ -es ortonormált szeparáló mátrix. A fehérítő réteg végzi a dimenzió-redukciót, valamint eléri azt, hogy a komponensek kölcsönösen korrelálatlanok legyenek és varianciájuk egységnyi legyen, azaz a korrelációs mátrix egységmátrix. A korrelálatlanság megkönnyíti a szeparálást. Látszik, hogy a fehérítés lényegében ugyanaz, mint amit a PCA algoritmus csinál. Ezért a fehérítés neurális implementációit gyakran a PCA neurális implementációiból származtatják. Egy egyszerű tanulási szabálya lehet:

$$\mathbf{V}_{t+1} = \mathbf{V}_t - \alpha(\mathbf{v}_t\mathbf{v}_t^T - \mathbf{I})\mathbf{V}_t \quad (1.1)$$

látszik, hogy a tanulási szabály (átlagban) akkor nem változtat a mátrixon, ha \mathbf{v}_t korrelációs mátrixa egységnyi lesz. A szeparáció tanulási algoritmus, mivel magasabbrendű momentumokat kell, hogy figyelembe vegyen, nemlineáris elemet tartalmaz:

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \beta\mathbf{v}_tg(\mathbf{y}_t^T) + \gamma\mathbf{W}_t(\mathbf{I} - \mathbf{W}_t^T\mathbf{W}_t) \quad (1.2)$$

itt g egy majdnem tetszőleges nemlineáris függvény. A leggyakrabban \tanh és x^3 függvények használatosak. A fenti tanulási szabályokban α , β , γ paraméterek, amelyek optimális értéke sajnos erősen függ az input típusától. Gyakran a tanulás során idővel csökkentik ezek értékét, elősegítve ezzel a konvergenciát.

1.3. Kódolás a látókéregben

A biológusok a faktoriális kódolásra először Barlow elméleti felvetéseket tartalmazó munkái alapján figyeltek fel. Mivel ekkor már sokat tudtunk a

látás folyamatáról, a figyelem hamarosan az elsődleges látókéregre terelődött.

A szóban forgó terület az agy hátsó részén található.⁴ Sejtjei közvetlenül a szem receptorsejtjeitől kapnak inputot, ezért ez a réteg nevezhető a feldolgozás első lépcsőjének. Az elektródás vizsgálatok során kiderült, hogy ezek a sejtek a látvány bizonyos paramétereire szelektíven viselkednek. Elsőként Hubel és Weisel (1962) készítette a sejtek érzékenységéről térképeket. Kis foltokat és vonalakat mozgatva az állat szeme előtt, úgy találták, hogy a neuronok tüzelési frekvenciája attól függ, hogy a látómező mely területén van az inger. Ha az illető sejt által elszenvedett gátlás/serkentés mértékét az inger pozíciója szerint két dimenzióban ábrázoljuk, akkor a 1.3. ábrához hasonló képeket kapunk. Ezeket a térképeket *receptív mezőknek* hívjuk. Az egyes receptív mezők, amellett, hogy a hozzájuk rendelt sejt a látómező csak egy adott régiójában érzékeny, nagy változatosságot mutatnak méretben és irányban. Ezenkívül a sejtek nemcsak az input térbeli helyzetére, hanem a mozgatás irányára és a mozgatás időbeli tulajdonságaira érzékenyek.

A hatvanas és a hetvenes években számos elképzelés jelent meg arról, mi lehet a célja ezeknek a neuronoknak. Campbell és Robson egy lehetségesnek tűnő magyarázatot vezetett le a lineáris rendszerek elméletéből. Mivel a sejtek a kísérletek során szelektívek voltak az irányokra és a térbeli frekvenciákra nézve, úgy gondolták, hogy az elsődleges látókéreg egy Fourier-transzformációhoz hasonló leírást ad a képről. Ugyanakkor ennek ellentmondó eredmények is születtek. A neuronok válasza ugyanis lokális jellemzőkre – élerekre és vonalakra – is érzékenynek mutatkozott. Ezért más kutatók inkább úgy gondolták, hogy az első transzformáció a lokális jellemzők leírására szolgál, azaz például éldetektáló funkciója lehet.

A nyolcvanas évek elején Marcelja észrevette, hogy a receptív mezők alakja nagyon hasonlít azokra a függvényekre, amiket Gábor még 1946-ban javasolt időben változó jelek analízisére. Ezek egy Gauss-függvénnyel lokálissá tett szinuszoid jellegű függvények, azaz általában ilyen alakban írhatók fel:

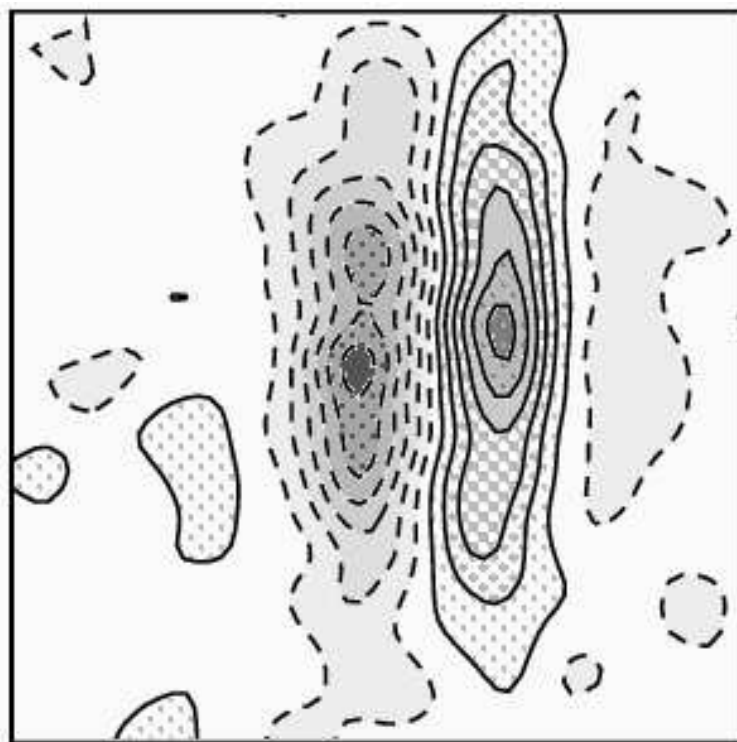
$$f(x) = \sin(2\pi kx + \theta)e^{-x^2/2\sigma^2}$$

Hamarosan kiderült, hogy ezek a függvények jól illeszkednek az elsődleges látókéreg sejteinek receptív mezőire. A fenti függvények Gábor-szűrő néven ismert kétdimenziós kiterjesztése a fentiekből adódik:

$$f(x, y) = \sin(2\pi kx + \theta)e^{-(x^2/2\sigma_1^2 + y^2/2\sigma_2^2)}$$

A kutatók úgy találták, hogy a Gábor-szűrők kítűnően leírják a megfigyelt receptív mezők alakját. Emellett más, még jobban illeszkedő függvényeket is

⁴A legtöbb kísérletet macskákön és majmokon végezték.



1.3. ÁBRA. Egy tipikus egyszerű sejt receptív mezeje a macska látókérgéből. A sima területek gátló, a pettyezett területek serkentő hatást jelentenek. [DeAngelis et al. 1995]

javasoltak (például Gauss-görbék összegét). A leírások közös jellemzője, hogy térben és frekvenciában lokalizált függvényekről van szó. Az ilyen függvényeket a matematikában *waveletek*nek hívják.

Párhuzamosan vizsgálni kezdték a természetes képek statisztikai tulajdonságait. Természetes képek alatt itt azokat a vizuális ingereket értjük, amivel egy élőlény élete során általában találkozik. Első ránézésre úgy tűnhet, hogy vizuális információink akkora változatosságot mutatnak, hogy nincs értelme bármilyen általános tulajdonságot keresni bennük. Azonban fizikai törvényeknek engedelmessé váló környezetünk nem teljesen véletlenszerű. Létezik egyfajta jóslhatóság vagy redundancia környezetünkben, és valószínűnek látszik, hogy az ezzel az ismerettel járó előnyöket a hosszú evolúció során kialakult látórendszer kihasználja. A természetes képek skála-invariánsnak mutatkoznak, azaz a kép statisztikai tulajdonságai nem változnak, ha a kép kis részletét kinagyítjuk⁵. Számos bizonyíték mutat arra, hogy a Gábor-transzformáció az ilyen képek *ritka eloszlás* szempontjából optimális kódolását valósítja meg [Field 1993]. A nyolcvanas évek végén a kompakt kódolás is felmerült, mint az elsődleges látókéreg lehetséges stratégiája. Azonban ezt a szűrők lokalizált volta lehetetlenné teszi. Belátható, hogy a PCA ekvivalens a Fourier-transzformációval stacionárius statisztikájú képek esetén, azaz ha a kép helyről-helyre ugyanolyan eloszlást mutat [Field 1994]. A Fourier-transzformáció pedig más megszorítások nélkül nem képes lokalizált szűrők előállítására.

A ritka eloszlású kódolásnak biológiai szempontból számos haszna van. A kép intenzitásában rejlő információ néhány aktív elemre koncentrálódik, ami a jelhordozó aktív alhalmaz erős kiemelését jelenti az egyébként rendkívül zajos neuronális háttérben, azaz jó jel/zaj viszonyt jelent. A kódolás a felismerést és az általánosítást is megkönnyíti. A ritka kódban a sejtek tüzelési valószínűsége alacsony. Ha ezt a tüzelési valószínűséget p -vel jelöljük és a sejtek tüzelését függetlennek tételezzük fel, akkor annak a valószínűsége, hogy egy adott területen csak az észlelendő tulajdonságot kódoló sejt tüzel, $(1-p)^n$. Itt n a sejt környezetében levő többi sejt száma. Minél alacsonyabb tehát a p , azaz minél ritkább a kód, annál nagyobb az esélye, hogy a megfelelő sejt tüzel. Természetesen a valóságban a közeli sejtek tüzelése nem független egymástól. Ennek ellenére az általános szabály érvényes marad: minél ritkább a kód, annál jobb a valószínűségi megfeleltetés a valóságos jellemző és az öt kódoló sejt között. Ha a sejtek tüzelési gyakorisága alacsony, a többszörös egybeesések esélye még jobban lecsökken (például a páros események bekövetkezése p^2 , a hármasoké p^3 , stb.). Ez azt jelenti, hogy az információhordozó (azaz valódi korrespondenciát jelentő) páros események bekövet-

⁵Ez a tulajdonságuk a fraktálokkal közös.

keztének valószínűsége még jobban kiemelkedik az összes lehetséges páros események közül. Barlow ezt eredetileg úgy fogalmazta meg, hogy könnyebb a „gyanús egybeeséseket” észrevenni.

A ritka kódolás további előnyeire mutattak rá az asszociatív memória kialakítását végző neuronhálózatok kutatói. Ha az ilyen hálózat inputjai ritka eloszlásúak, a hálózat több memóriát tud eltárolni és a tárolt minták előhívásának képessége is javul. Többen egyenesen azt javasolták, hogy a nem ritka eloszlású adatokat mesterségesen ritkítani kell feldolgozás előtt.

Olshausen és Field a ritka eloszlás kialakítására tett javaslatot 1996-ban, és elképzelésük működőképességét számítógépes szimulációkkal igazolták [Olshausen and Field 1996]. A részletek nélkül algoritmusuk a következő volt:

$$\dot{a}_i = \sum_{\vec{x}} \Phi_i(\vec{x})r(\vec{x}) - f(a_i)$$

itt Φ_i az i -edik kétdimenziós szűrő (ezért a vektorváltozó), a_i pedig a hozzá rendelt sejt aktivitása. A differenciálegyenlet egyensúlyi megoldása adja a valóban kialakuló reprezentációt, amit a kialakuló sejtaktivitásokból alkotott \mathbf{a} vektor tárol. A kifejezés második tagja valószínűsíti meg a ritkaságot az f függvényen keresztül, ami az alacsony aktivitású elemeket relatíve jobban bünteti, mint a magasabbakat. Egy ilyen függvény látható a 1.4. ábrán. A kifejezés első tagja a szűrőket alkalmazza az $r(\vec{x})$ maradékképre:

$$r(\vec{x}) = \left[I(\vec{x}) - \sum_i a_i \Phi_i(\vec{x}) \right]$$

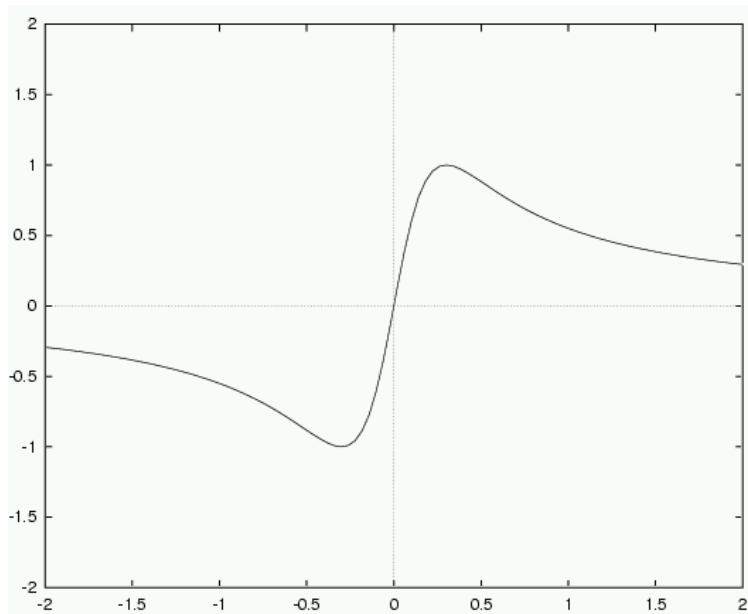
itt a jobb oldalon az aktivitásokból helyreállított kép, a bal oldalon pedig az eredeti kép látható. Magyarán minden lépésben képezzük a szűrők-ből helyreállított kép és az eredeti kép különbségét (azaz a reprezentáció hibáját), és ezzel a deriváltra áttérve a különbségen vizsgáljuk a szűrők aktivitását és módosítjuk a ritka eloszlású reprezentációnak megfelelően. Az irodalom az ehhez az elvhez hasonlóan működő neuronhálózatokat *rekonstrukciós hálózat* néven ismeri [Fomin et al. 1997]. A hálózat alkalmas a szűrők „menet közbeni” kialakítására is:

$$\Delta \Phi_i(\vec{x}) = \eta a_i r(\vec{x})$$

itt a_i az aktivitások egyensúlyi megoldása.

A természetes képeken kialakult szűrőket ábrázolja a 1.5. ábra. Jól látható, hogy ezek a megfigyelt receptív mezőkhöz nagyon hasonlítanak.

Olshausen és Field megmutatták, hogy az algoritmusuk ekvivalens az ICA algoritmusával abban az esetben, ha a kimementi réteg dimenziószáma

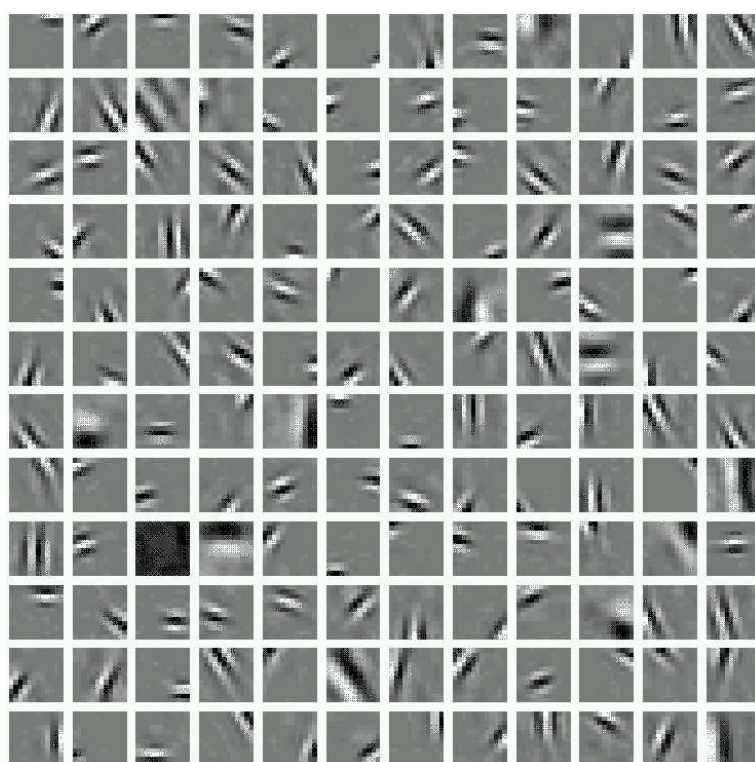


1.4. ÁBRA. Egy lehetséges büntetőfüggvény a ritka eloszlás kialakítására [Olshausen and Field 1996]

nem nagyobb, mint a bemeneti rétegé [Olshausen and Field 1996]. Hamarosan hasonló számítógépes szimulációk készültek természetes képeken az ICA algoritmusával, lásd pl. [Hyvärinen et al. 1999]. Az így kialakult szűrők hasonlítanak a ritka reprezentáció fenti modelljénél kapottakhoz.

Láthatjuk tehát, hogy a faktoriális kódolás biológiai alkalmazása komoly támogatást kapott a látókéreg vizsgálata során. A történet itt még nem ér véget. A folytatásra a megfelelő helyen még ki fogunk térni (lásd 5.4. rész).

A továbbiakban a faktoriális kód alkalmazhatóságát fogjuk vizsgálni a hippocampusz esetében. A következő két fejezetben ezért röviden áttekintjük, mit tudunk a hippocampuszról.



1.5. ÁBRA. A ritka eloszlású kódolás szűrői [Olshausen and Field 1996]

2. fejezet

A hippocampális formáció biológiája

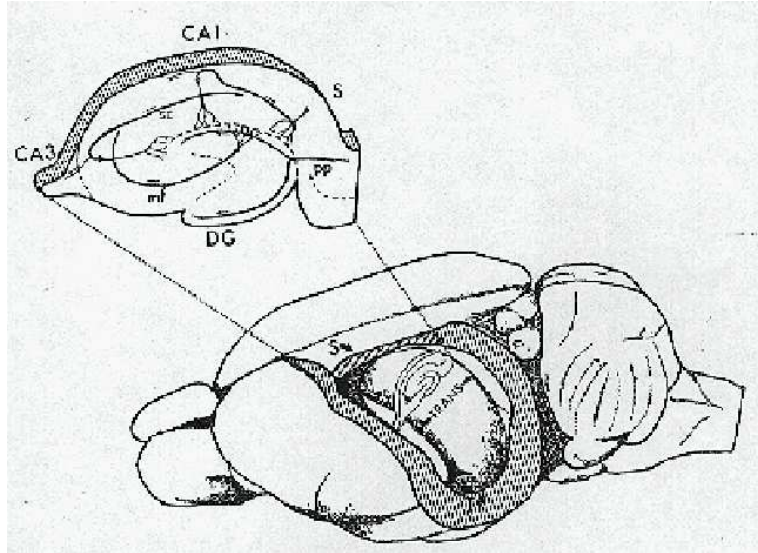
A hippocampusz és környéke az agy egyik legtöbbet tanulmányozott területe, és ennek megfelelően biológiájáról az utóbbi húsz-harminc évben hatalmas mennyiségű publikáció jelent meg. A területnek például saját kéthavonta megjelenő folyóirata van. Lehetetlen volna ebben a fejezetben a teljességet akár csak említések szintjén elérő összefoglalót készíteni, és ez nem is lehet céлом. Itt csupán azokat a tényekre szeretnék kitérni, amelyekre a továbbiakban hivatkozni fogok.

2.1. Anatómia

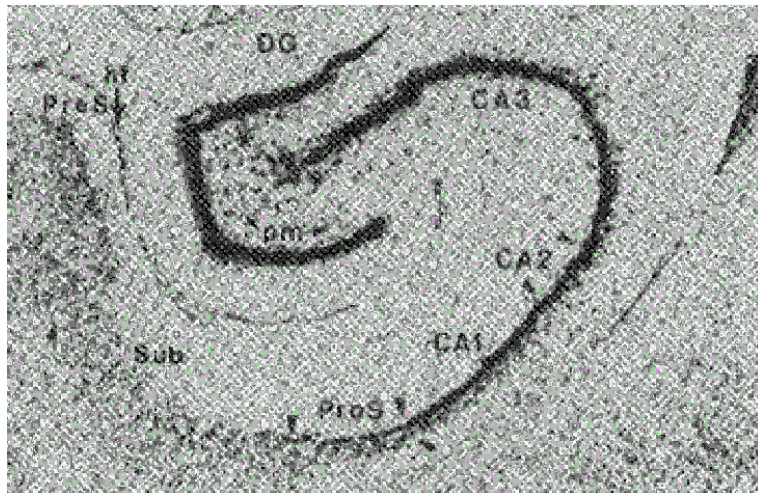
A hippocampális formáció az agy limbikus rendszerének része. Az elnevezés gyűjtőnév; számos további képletet foglal magába. Az alkotókat eltérően szokták meghatározni, de többnyire ide sorolják az entorhinális kortextet, a dentate gyrust, a subiculáris komplexumot és természetesen magát a hippocampuszt.

A hippocampális formáció kiterjedt kapcsolatokat ápol a környező struktúrákkal. Néhányat itt megemlítünk. Az entorhinális kortex szoros kapcsolatban áll a perirhinális és a postrhinális kortexszel, amelyek a hátulsó fali lebenytől és az inferotemporális kortextől egyaránt direkt afferenseket kapnak. Ez arra utal, hogy például a látás dorzális és ventrális árama egyaránt jelen van a hippocampális formációban. Az entorhinális kortex közvetlen kapcsolatokkal is rendelkezik a neokortex felé. A subiculáris komplexum a talamusszal is szoros kapcsolatban áll. A hippocampális formáció efferensei egyrészt az entorhinális kortexen keresztül vezetnek, másrészt a hippocampusz a fimbrián keresztül kapcsolatban áll a septal

2. FEJEZET. A HIPPOKAMPÁLIS FORMÁCIÓ BIOLÓGIÁJA



2.1. ÁBRA. A hippocampusz elhelyezkedése a patkányagyban. DG: dentate gyrus, S: subiculum pp: perforant path, mf: moharostok, sc: Schaffer kollaterálisok [Amaral and Witter 1989]



2.2. ÁBRA. A hippocampusz festéssel kapott szövettani képe. DG: dentate gyrus, ProS: prosubiculum, PreS: presubiculum, S: subiculum, hf: hippocampális rés [Redish 1997]

nuclei-vel, amely serkentő és gátló kapcsolatokkal látja el a valamint a fornixon keresztül a

(2.1. és 2.2. ábra)

limbikus rendszerrel való kapcsolata

HC-EC egy huroknak fogható fel amaralbol

fimbria-fornix, CA1, CA3 fogalma

subiculáris komplexum fogalma

hippokampusz bemenete szenzoros informacio is

2.2. Etológia

rejtett platformos kiserlet, es erre a hippocampalis lezio kovetkezmenyei

rejtett platformos kiserlet referenciamemoriát igényel utalás

rejtett platformos explorálatlan területen at (redishbol).

2.3. Neurofiziológia

place field és place sejt fogalma

nem-terbeli korreláció hippocampusz sejtjeiben

fázisprecesszió a tuzelésben, skaggs alapján

többszörös receptív mezó

EEG, theta fázis, SPW fázis. Theta abszolút ideje

interneuronok rekurrens kapcsolata tuzelést gátolja

3. fejezet

Elképzések a hippocampusz működéséről

A hippocampusz szerepéről számos elmélet látott már napvilágot. Azt általában mindenki elfogadja, hogy a hippocampusznak köze van a térbeli navigáció lebonyolításához, de azt is, hogy szerepe nem korlátozódik pusztán erre. A továbbiakban röviden áttekintjük a hippocampusz szerepéről ezidáig alkotott fontosabb elméleteket [Redish 1997] nyomán.

Noha a hippocampusz intenzíven kutatott terület rágcsálók, nyulak, emberek és emberszabásúak esetében, a továbbiak főleg rágcsálókra korlátozódnak. Ennek oka részben az egyszerűség, részben pedig az, hogy a hippocampusz nem feltétlen ugyanazt a szerepet játssza minden faj esetében. Az analógiákról majd lesz szó, de az erősebb állításokat ma még többnyire spekulációnak minősítik.

3.1. Az elméletek rövid ismertetése

3.1.1. A hippocampusz mint érzelmi központ

A korai elméletek – felhasználva azt a tényt, hogy a hippocampusz erős kapcsolatokkal rendelkezik a limbikus rendszer felé, lásd még 15. oldal – azt tételezték fel, hogy a hippocampusz az állat érzelmi állapotait reprezentálja. Úgy gondolták, hogy a limbikus rendszer azon emóciókat és érzéseket kontrollálja, amelyek tudatosan nem irányíthatók. Erre utal a népszerűsítő „hüllőagy” elnevezés is. Ez az elképzelés egészen Freudig és az ő tudatallatiról szóló elképzeléséig nyúlik vissza. A teória főleg annak tulajdonítható, hogy akkor még az érzelmeket irányító központokat (ilyen például az amygdala és az elülső övi [anterior cingulate] kéreg) többnyire képtelenek voltak szétvá-

lasztani a környező struktúráktól (mint például a hippokampusz és a hátulsó övi [anterior cingulate] kéreg).

Az elméletet alátámasztó adatok többsége a korai léziós vizsgálatokból származik. Ezekben a kísérletekben kiirtják vagy működésképtelenné teszik a vizsgált képletet illetve annak egy részét, esetleg csak bizonyos kapcsolatait, és ezután viselkedést vizsgáló tesztekben próbálják megállapítani a rombolás hatását. A korai léziós vizsgálatok legtöbbször meglehetősen rosszul tervezett feladatokat használt tesztelésre. A léziós adatokkal egyébként is mindig számos gond van. Egyrészt igen nehéz elérni, hogy a műtét során környező képletek ne sérüljenek.¹ Másrészt, ha egy állat nem képes egy feladatot végrehajtani egy laboratóriumi kísérlet folyamán, annak ezernyi más oka lehet, mint hogy épp azt a képletet irtottuk ki, ami felelős ezért a viselkedésért. Lehetséges például, hogy a valóban lényeges terület működéséhez a kiirtott struktúra valamilyen szempontból elengedhetetlen neurális tónust ad. Másrészt fordítva is igaz: attól, hogy az állat képes elvégezni egy feladatot, egyáltalán nem biztos, hogy ugyanazt a stratégiát használja, mint a lézió előtt. Lehetséges például, hogy más területek képesek ellátni a hiányzó képlet funkcióját és így az állat továbbra is meg tudja oldani a feladatot. Mindezek mellett a léziós adatok ma is sok hasznos következtetést tesznek lehetővé.

Ezen korai elmélet legnagyobb problémája, hogy képtelen megbirkózni azzal az óriási mennyiségű kísérleti eredménnyel, amit a neurofiziológiában az elmúlt húsz évben ezen a területen publikáltak.

3.1.2. A hippokampusz mint prediktív komparátor

Gray 1982-ben megpróbálta a feszültségi szintről szóló tanulmányokat és a hippokampusz működéséről szóló elképzeléseket egységes tető alá hozni. Az elképzeléseit főleg a nyugtatószerek hatásaira, bizonyos léziós adatokra és a hippokampusz a septumhoz fimbria-fornixon keresztül vezető kapcsolataira alapozta. Elképzelése szerint az agyban egy *viselkedést gátló rendszer* érzékeli az adott tevékenységgel járó büntetéseket és az ahhoz vezető ingereket, és ennek megfelelően tanítja a viselkedést gátló mechanizmusokat, megemeli az állat arousalját és a figyelem szintjét. A nyugtatószerek hatásukat ennek megfelelően ennek a rendszernek a gátlásán keresztül fejtenék ki.

Gray ezek után megkérdőjelezhető lépést tett. Ennek a rendszernek az anatómiai helyét ugyanis a szepto-hippokampális rendszerben vélte felfedezni. Elméletében a hippokampusz fő szerepe a jóslás. A hippokampusz a kompará-

¹A korai vizsgálatok gyakran azokat az idegrostokat és egyéb struktúrákat is eltávolították, amik akadályozták a célterülethez történő hozzáférést.

torként működő subiculumba küldi információit, ami a jóslatot összehasonlítja az aktuális szenzoros információval. Ezután az állat eldöntheti, hogy folytassa-e a jelenlegi mozgását avagy megálljon és újraértékelje pillanatnyi helyzetét.

A fő probléma ezzel az elmélettel, hogy képtelen megmagyarázni az előző fejezetben említett kísérleti tényeket. Például nem tudja megmagyarázni, miért képtelen az állat megoldani a rejtett platformos labirintust hippokampális lézió után (a kísérletről lásd 17. oldal).

Újabban Gray viselkedést gátló rendszerével speciális dopaminra érzékeny neuronokat azonosítanak, amelyek mind a viselkedésért kapott jutalomra, mind az információ ismeretlen voltára (*novelty*) reagálnak.

3.1.3. Konfigurális asszociációs elmélet

Sutherland és Rudy azt javasolta 1989-ben, hogy a hippokampuszról akkor gondolkodunk helyesen, ha különséget teszünk *egyszerű asszociációk* és *konfigurális asszociációk* között. Az előbbi két egyszerű ingert köt össze, míg az utóbbi az ingerek egy együtteséből egy egyedi reprezentációt hoz létre. Így könnyebb különséget tenni hasonló konfigurációk között. Ezek a reprezentációk azután más ingerekkel is asszociációs kapcsolatba hozhatók.

Sutherland és Rudy eredetileg azért javasolta ezt a megoldást, hogy megoldják az ún. negatív különbségtételi problémát, ami a XOR problémával azonos. Ez nagyjából azt takarja, hogy a válaszadás módja azonos legyen ha A és B inger közül kizárólag csak az egyik van jelen, illetve akkor, ha mind A és B vagy sem A, sem B inger nincs jelen. Régóta ismert, hogy egy lineáris hálózat ezt a problémát nem tudja megoldani, de egy többrétegű neuronhálózat egyszerű nemlinearitásokkal felvértezve igen. Az elmélet szerzői tulajdonképpen azt javasolták, hogy a hippokampusz a középső rétege egy ilyen háromrétegű neuronhálózatnak. Ez lehetővé teszi, hogy az agy többi részében a tanulási szabályok lényegében lineárisak legyenek. Más szerzők, név szerint Gluck és Myers konkrétan is megpróbálták igazolni ezt az elképzelést. A hippokampuszt egy háromrétegű autoasszociatív hálózatként szimulálják, amit backprop algoritmussal tanítanak és a rejtett réteg reprezentációját használják outputként a cortex felé.

Az elmélet fő problémája abban rejlik, hogy azok a konfigurális asszociációk, amelyeket egy rágcsgáló élete során feltehetőleg kialakít, mind magukba foglalják az állat környezetét is (a környezet valószínűleg mindig eleme a konfigurációnak). A kísérletek azonban azt mutatják, hogy konstans környezetben kialakított asszociációk megmaradnak hippokampális lézió után. Természetesen elképzelhető, mint ahogy azt sokan javasolták is, hogy a környezetet mint kontextust az állat speciális módon kezeli.

Az elmélet ellen hozható fel az is, hogy a hippokampális piramissejtek érzékenységében ez esetben a térbeli jellemzőkre illetve a lokalizációval nem összefüggő jellemzőkre (sebesség, irány, stb.) nézve nem volna különbség. Noha igaz, hogy a sejtek érzékenyek nem-térbeli jellemzőkre is (lásd 17. oldal), ezeket másodrendű korrelációknak kell tekintenünk. Valószínűbbnek tűnik ugyanis, hogy ezek inkább a place field-ek nem-térbeli jellemzők miatt bekövetkező topológiai változásait jelzik.

3.1.4. Munkamemória

Egy radiális elrendezésű, sokkarú labirintusban fimbria-fornix lézióknak alávetett patkányok nem képesek emlékezni arra, melyik karban jártak már, de arra igen, hogy melyik karban nem volt soha ennivaló. Ezen alapul Olton és munkatársai 1979-ből származó elmélete. Szerintük érdemes különbséget tenni a feladat alatt felhasznált információ (*munkamemória*) és a feladatok között felhasznált információ (*referencia memória*) között. Az elmélet szerint a hippokampusz a munkamemóriákkal foglalkozik, de nincs köze a referencia memóriához. Ennek ellentmond, hogy a rejtett platformos kísérlet, ami erősen hippokampusz-függő, referencia memóriát is igényel (lásd 17. oldal).

Jarrard kifejezetten a referencia memória és a munkamemória vizsgálatára végzett kísérleteket az említett radiális labirintusban. Két feladatot vizsgált. Ezek annyiban különböztek, hogy azok a karok, amikbe soha nem rakott ennivalót az egyikben a labirintuson kívüli, a másikban a labirintuson belüli jelektől függtek. Úgy találta, hogy a lézióknak alávetett patkányok mind a kétféle memória használatában követtek el hibákat ha jelek külsők voltak, de egyikben sem, ha belső jeleket alkalmazott, bár ilyenkor a tanítás is több időt vett igénybe. A munkamemóriát itt az jelentette, hogy betér-e egy olyan jelölt karba a patkány, ahol már járt, a referencia memóriát pedig az, hogy betér-e olyan karba, amit már előzetesen megtanult mint olyan kart ahova soha nem kerül ennivaló. Jarrard eredményeit a munkamemória-hipotézis céfolatának tekintette. Ez azt jelenti, hogy a rejtett platform kísérletek eredményeit is figyelembe véve az elméletet el kell vetnünk. További probléma, hogy az Olton és társai által felhasznált léziók adatok valójában nem voltak egészen egyértelműek. Az újabb kísérletek számos apró különbségre derítettek fényt a hippokampusz anatómiájában. Ide tartoznak azok a fimbria-fornix léziók is, amiket Olton felhasznált.

Mindezek ellenére a memóriák ilyenén megkülönböztetése hasznos, és bizonyos esetekben kifejezetten jól használhatónak bizonyult. Erre a neurofarmakológia mutat bizonyítékot: kiderült ugyanis, hogy a kolinerg antagonisták a munkamemóriát jobban zavarják, mint a referencia memóriát.

3.1.5. Kognitív térképek

Eltérően az előzőekben ismertetett elméletektől, van három olyan elképzelés, amelyek ellen máig nem hoztak fel komoly bizonyítékokat. Az első O'Keefe és Nadel 1978-ból származó *kognitív térkép* teóriája.

Ez az elmélet azt állítja, hogy a hippocampusz a környezet kognitív térképét tárolja. Az elmélet legkomolyabb bizonyítékát a hippocampális piramidális sejtek elsőrendű korrelációi adják, amelyek úgy tűnik, hogy az állat pozícióját kódolják (place sejtek). Az is az elmélet mellett szól, hogy a hippocampális lézióknak komoly hatása van az állat navigációs képességeire. Az elképzelés egyértelműen sikeresnek bizonyult. Ezt bizonyítja az az elképesztő mennyiségű kísérlet, amely erre a hipotézisre alapozva végeztek. Sok munkát fektettek a modellezésbe is. Ezek főleg a place sejtek kialakulásának mechanizmusait illetve a navigációban betöltött szerepüket próbálják kideríteni.

A kognitív térkép elmélete annyira sikeres volt, hogy számos finomítása és alváltozata született. A legtöbb közülük azzal a kérdéssel foglalkozik, hogy a rendszer miként reprezentálja az állat helyzetét és ezután hogyan képes ezeket a reprezentációkat felhasználni. Néhány elképzelés azonban más szerepet szán a hippocampusznak.

Lokális nézeteket tároló asszociatív memória

Az állat *lokális navigáció* képessége alatt azt értjük, hogy képes olyan célhoz navigálni, ami direkt módon nincs jelölve. A navigáció ekkor a környezetben levő kimagasló pontok (az úgynevezett *landmarkok*) helyzete alapján történik.² A kísérletek azt mutatják, hogy a hippocampusz csak a lokális navigációt befolyásolja. A *lokális nézet* az állat pillanatnyi pozíciójából elérhető szenzoros inputot jelenti. Az első elképzelések szerint a hippocampusz sejtjei azért mutatnak place field-eket, mert direkte landmarkok kombinációira érzékenyek. Mikor kiderült, hogy a sejtek a sötétben továbbra is place field-eket mutatnak, nyilvánvalóvá vált, hogy ez messze nem a teljes magyarázat. Ekkor már számos kutató úgy látta, hogy a hippocampusz asszociatív memóriához hasonló tulajdonságokat mutat. Innen egyenes volt a következtetés, hogy néhány sejt valószínűleg érzékeli a sötétben is megmaradó landmarkokat, és ebből a hiányos mintázatból a reprezentáció hiányzó sejtjeit a hippocampusz asszociatív módon kiegészíti.

²Meg szokták még különböztetni az ún. *taxon* navigációt, ami egy rögzített jel irányába történő haladást jelent, a *gyakorlás alapján történő* navigációt, amely során egy konstans motoros program begyakorlásával és annak végrehajtásával éri el a célt, illetve a *random* navigációt, amelynek során az állat célját a környezet véletlen explorációjával éri el.

3. FEJEZET. ELKÉPZELESEK A HIPPOKAMPUSZ MŰKÖDÉSÉRŐL

Külső segítő jelek teljes hiányában ez az elmélet azt jósolja, hogy az állat továbbra is képes lokális nézetről meglévő reprezentációját saját belső szenzoros információjával asszociálva a következő lokális nézet reprezentációt predikálni. Ez lehetővé tenné az állatnak, hogy a külső segítség teljes hiányában is képes legyen pozíciója tudatában lenni és így egyfajta vaklátásra szert tenni. Azonban ez nem magyarázza az ún. *út-integrációs* képességet, ami azt jelenti, hogy az állat bármilyen helyzetből képes visszatérni kiindulópontjára egy adott környezetben, még sötétben is.³ Az állatnak az elmélet szerint először meg kellene tapasztalnia az adott átmenetet, hogy azt a megfelelő lokális nézettel asszociálhassa. Ez azt is jelenti, hogy az állat nem képes rövidítéseket megtervezni a környezet ismeretlen területein át. Ennek ellentmond az a tény, hogy a rágcsálók képesek megoldani a rejtett platformos labirintust akkor is, ha ehhez közben ismeretlen területeken kell áthaladniuk (lásd 17. oldal).

Recce és Harris nem olyan rég azt javasolta, hogy az asszociatív memória tulajdonságot nem is szabad az út-integrációt megvalósító mechanizmusnak tekintenünk. Sokkal inkább úgy kell rá gondolni, mint a lokális nézet kiegészítésére szolgáló eszközt. Egyre többen gondolják úgy, hogy az út-integrációt a hippocampuszon kívüli egységek valósítják meg. Ez azt is jelenti, hogy a lokális nézet reprezentációi a hippocampuszon kívül találhatóak. Ennek az elképzelésnek megvan az a problémája, hogy ekkor a felgyülemelő hibák eltorzíthatják a hippocampális kognitív térképet.

Töredék térképek

Worden 1992-ben egy, a *szeletszűrők (fragment filters)* koncepcióján alapuló rendszert tételzett fel a hippocampuszban. A szeletszűrők az állat pozícióját reprezentálják relatíve kis számú landmarkhoz viszonyítva (Worden hármát használt). Ha az állat célját egy ilyen töredékben kell hogy megtalálja, de csak egy másik töredékben levő landmarkokhoz fér hozzá, akkor működésbe lép egy mechanizmus, ami a két töredéket összeilleszti. Az is előfordulhat, hogy több mint két töredéket kell láncban összeilleszteni, hogy a pillanatnyilag elérhető töredék térképből egy távoli célhoz jusson. A szerző ezt az eljárást *töredékillesztésnek* hívja. Habár Worden néhány az algoritmusához szükséges matematikai operációt, mint például translációk és rotációk kiszámítását bizonyos anatómiai struktúrákkal (konkrétan a dentate gyrusszal és a CA3-mal) azonosította, kevés bizonyítékkal támasztja alá az elméletét azon az tényen kívül, hogy a hippocampális sejteknek van

³Ezt a képességet majdnem az összes rágcsálón, valamint madarakon, kutyákon, macskákon, emberen, sőt még rovarokon is kimutatták.

valami közül a térbeli leképezésekhez. A modellje által igényelt számítások biológiai megvalósítását az anatómiai adatok megkérdőjelezhetővé teszik.

A kognitív gráf elmélet

Muller és munkatársai a kilencvenes évek elején kimutatták, hogy place sejtek esetében a hebbi LTP a környezet véletlen bejárását feltételezve érdekes szinaptikus súlyfüggvényeket fog létrehozni: két sejt közötti szinaptikus súly fordítottan arányos lesz a place feldek átfedésével. Ez azt jelenti, hogy a szinaptikus súly tárolhatja a távolságot a place feldek középpontja és a tér topológiáját rögzítő mátrix között. Muller és társai a *kognitív gráf* elnevezést javasolták erre a struktúrára. Wilsonnak és McNaughtonnak sikerült bizonyítékokat is mutatnia arra, hogy ez a struktúra exploráció után valóban létezik a hippokampuszban.⁴

Észrevételük megtétele után Muller és társai azt javasolják, hogy egy gráfkereső algoritmus útiterveket tudna készíteni ennek a struktúrának a felhasználásával. Az algoritmusra esélyesnek tartják a Dijkstra-algoritmust, de a Bellman-Ford algoritmust is szóba hozzák. A fő probléma ezzel az elmélettel, hogy egyik algoritmus sem rendelkezik egyszerű neurális implementációval.

A hippokampusz mint út-integrátor

McNaughton és munkatársai 1996-ban bemutattak egy mechanizmust, ami képes út-integrálást megvalósítani. Úgy gondolják, hogy rendszerükben a hippokampusz az *attraktor hálózat* szerepét játssza. Az ilyen hálózatokban egy stabil állapotnak a pozíció egy koherens reprezentációja felel meg. Kis mértékben eltérő excitatorikus input a reprezentációt finoman elmozdítja, nagy mértékben eltérő pedig a reprezentációt törli és az új inputnak megfelelő állapotba kényszeríti. A szerzők a subiculis komplexumot (lásd 17. oldal) az ilyen *eltolási kapcsolatok* forrásának tekintik, ami lehetővé tenné, hogy a rendszer saját mozgását nyomon kövesse. Felteszik, hogy a hippokampuszban *táblák (charts)* halmaza van kódolva, amelyeket az állat különböző környezetekben használ. A kognitív gráf elméletének ezen kiterjesztését a hippokampusz *soktáblás* modelljének nevezik. Különös módon elméletük azt igényli, hogy a kognitív gráf már a rendszerbe legyen drótozva, amikor az állat explorál.

⁴Wilson és McNaughton csak azt mutatták meg, hogy a kognitív gráf az explorációt követően létezik. Az továbbra is kérdéses, hogy ez a gráf esetleg létezik már az explorációt megelőzően is, avagy annak során alakul ki.

Az elmélet sok kísérleti ténnyel, különösen a place sejtekre vonatkozó adatokkal harmonizál. A teória nagy komplexitású kapcsolatrendszer igényel a sok tábla miatt. Ez még nem indokolja az elmélet elvetését, mert a hippocampusban ténylegesen meglévő kapcsolatrendszer-bonyolultság kielégítené az igényeket. Azonban Alyan és munkatársai nemrégén végzett kísérletei arra utalnak, hogy a patkányok út-integrálási képességeket képesek mutatni hippocampusz nélkül is, amit a hippocampusz mint út-integrátor elképzelés cáfolatának is tekinthetünk. Emellett Redish és Touretzky 1997-ben egy tanulmányában azt próbálja bizonyítani, hogy bizonyos korábban végzett kísérletek is ellentmondanak ennek az elméletnek.

Útvonaltanulás mint navigáció

Az LTP asszimetrikus volta (azaz ha a rövidebbel b neuron előtt tüzel, akkor csak az $a \rightarrow b$ kapcsolat erősödik) valamint az előző fejezetben említett fázip-recessziós jelenség (17. oldal) együttese képes a CA3-ban olyan asszimetrikus konnektív matrixt létrehozni, amely a bejárt útvonalakat reprezentálja. Az asszimetrikus kapcsolatrendszer egyik Blum és Abbott által előrejelzett következményét, miszerint a place feldek a megtanult útvonalak mentén visszafelé nézve enyhén torzulnak, kísérletileg is igazolták. Ugyanakkor a szerzők úgy tartják, hogy ezek az asszimetrikus kapcsolatok a navigáció irányítására is felhasználhatók.

Az egyik probléma ezzel a hipotézissel, hogy ez esetben képtelenség volna a különböző feladatokat megkülönböztetni egymástól. Az elméletből az következik, hogy az állat rövidítéseket vektormezők interpolálásával tervez meg. Ez azt jelenti, hogy egy rövidítést csak egy olyan pozícióból tud kigondolni, ami két korábban megtapasztalt pozíció között fekszik. A már említett rejtett platformos kísérletek, ahol az állat a platformot képes volt előzőleg nem bejárt pozícióból egyenesen megtalálni, megcáfolják ezt (17. oldal).

További probléma a hipotézissel, hogy a hippocampusz nem feltétele a navigációnak. Az állat ugyanis képes visszatalálni kiindulási helyére hippocampusz nélkül is.

3.1.6. Kontextuális felidézés

Hirsch 1974-ben javasolta, hogy a hippocampusz szerepe a környezet biztosítása a normál inger-válasz mechanizmusoknak. Ez a *kontextuális felidézés* néven ismertté vált elmélet megmagyarázza a reverz hiányosságokat a hippocampális lézióknál: egyszerű diszkriminációs problémákat normálisan teljesítenek, de amennyiben választásukért a

3. FEJEZET. ELKÉPZELÉSEK A HIPPOKAMPUSZ MŰKÖDÉSÉRŐL

továbbiakban jutalom helyett büntetést kapnak, furcsa mód hajlamosak továbbra is ugyanazt választani. Ez ugyanakkor megmagyarázza a XOR problémát és a látens tanulást is (azaz az állatok gyorsabban tanulnak ismert mint ismeretlen környezetben).

A kontextuális felidézés elmélete hasonlít a konfigurális asszociációs elmülethez (20. oldal) és a kognitív térkép elmülethez (22. oldal). Az előbbihez annyiban hasonlít, hogy a hippocampus szerepének az alapvető inger-válasz képesség kiterjesztését szánja, azonban az utóbbihoz igazodva a kiterjesztést pusztán környezetinek tételezi fel. Bizonyos szempontból az elmélet által javasolt szétválasztás hasonló ahhoz, amit a behavioristák szimpla inger-válasz elmületeinek cáfolatára hoztak fel a korai kutatók, amikor feltételezték, hogy létezik egyfajta kognitív térkép az agyban.

Bizonyos, hogy a hippocampusnak szerepe van a kontextuális kondícionálásban. Úgy találták, hogy az asszociáció kialakulását nem gátolja a hippocampus működésképtelensége, de a környezetek közti asszociációt a hippocampális állat a normálisaktól eltérően hajtja végre. Ez arra utal, hogy a hippocampus mindenképp ad egy kontextuális faktort az asszociáció kialakítása során. Kísérletek bizonyítják, hogy egy hippocampális állat nem képes asszociációt kialakítani egy környezetben és egy másik asszociációt egy másik környezetben.

Késleltetések áthidalása

Hasonló javaslatot tett Rawlins 1985-ben, amikor azt javasolta, hogy a léziós adatokat a késleltetést is tartalmazó kísérletekben szükséges időbeli áthidalás problémájával lehet kapcsolatba hozni. Nem mellékes körülmény, hogy az általa bemutatott kísérletek a késleltetés alatt kontextuális átmenetet is tartalmaztak. Rawlins főleg elmélete időbeli vonatkozásait hangsúlyozza, de ötletei sok hasonlóságot mutatnak a kontextuális felidézés elmületeivel. A szerző rámutat, hogy normális példányokban a késleltetett tanulás határfoka javul, ha az állatot a késleltetés időtartamára kiveszik a környezetéből. Ugyanakkor előmozdítja a tanulást az is, ha a választás előtt egy különleges esemény történik az állattal. Sokan úgy érvelnek, hogy érdemes használni a *referenciakeret* terminológiáját, ami az állat adott pillanatban érvényes koordinátarendszerét jelenti a mentális állapottal együtt (ide tartozik a pillanatnyi cél vagy az éppen végrehajtott mozgássor). Számos kísérlet mutatja, hogy az állat megváltoztathatja referenciakeretét akár egy adott környezetben belül is komplex feladatok végrehajtása során. Ilyen átvezető pontoknál az állat valódi pozíciója ugyanaz marad, de a saját reprezentációja nem, ezért szüksége van egy mechanizmusra, ami ezt a kontextuális hibaként felfogható pozícionálási zavart kiküszöböli. Erre a célra kítűnően megfelel

Rawlins elképzelése.

Felismerő memória

Gaffan egy régi, 1972-es cikkében lényegében azt javasolta, hogy a hippocampus az ismerős dolgok észlelését végzi. Kérdés, hogy mi a kapcsolat a kontextuális felidézés és a Gaffan által javasolt *felismerő memória* elmélet között.

Úgy tűnik, Gaffan elvetette elméletét, amikor saját kísérleteiben kiderült, hogy a fornix lézióknak nincs hatása a majmokon végzett ingerfelismerési kísérletekben. Elméletének alátámasztására patkányokon végzett korábbi kísérletei azonban megerősíthetik elképzelését. Ezekben a vizsgálatokban ugyanis a környezet felismerését tesztelte, míg az újabbakban viszont ingerfelismerést. Mint már azt említettük, lehetséges, hogy a környezetet az állat eltérő módon kezeli.

A referenciakeret (lásd fentebb) hibás felismerését tekinthetjük az ismerős környezet hibás felismerésének is. Vannak adatok arra, hogy az öregedő állatok (ezek már feltehetőleg LTP és hippocampális problémákkal küzdenek) nem használják ugyanazt a reprezentációt két kísérlet között egy környezetben, azaz mintha nem emlékeznenek a referenciakeretre. Fordítva is igaz: új környezetben nem ismerik fel annak új voltát, és nem megfelelő reprezentációt próbálnak felhasználni.

3.1.7. Epizodikus memória

A rágcسالókra vonatkozó adatoktól eltérően, a temporális lebenyben elvégzett léziók embereknél (amilyen például a H. M. monogramú sokat emlegetett páciens volt az ötvenes években) arra utalnak, hogy a hippocampális formációnak köze van a memória kialakításához. Ezeknek a változtatásoknak emberek esetében pusztító hatása van a memóriára. A betegek közvetlen felidézési kísérletekben jól teljesítenek, de néhány percnél hosszabb várakozási idők esetén már nagyon rosszul. Teljesen képtelenek új hosszú távú memóriák kialakítására, ezzel az úgynevezett *anterográd amnézia* jelenségét mutatják. Ugyanakkor retrográd amnéziát is produkálnak, azaz képtelenek felidézni dolgokat, amik a lézió előtt történtek velük. Ez utóbbi azonban csak korlátozott ideig nyúl vissza, az életük korai szakaszában szerzett memóriák tehát érintetlenek maradnak.

További tanulmányok kimutatták, hogy a hippocampus csak az úgynevezett deklaratív memóriában játszik szerepet. A memória felosztását *deklaratív* és *procedurális* memóriára először Cohen és Squire javasolta 1980-ban. A deklaratív memória tényeket, neveket, eseményeket és epizódokat tárol (ezért a deklaratív memóriát hívják *epizodikus memóriának* is), míg

a procedurális memória a gyakorlással elsajátított képességeket jelöli. A deklaratív memória gyors (egy kísérlet elég az elsajátításra) és általában a „mi” kérdésre válaszol, amíg a procedurális memória csak hosszas gyakorlással alakítható ki és a „hogyan” kérdésére vonatkozik. Például képesek vagyunk gyorsan megjegyezni, hogy a ez a mondat a 28. oldalon van, és ennek a tudásnak azonnal birtokában is vagyunk. Szemünket becsukva rögtön el tudjuk mondani valakinek, hogy melyik oldalon jártunk éppen. Ha viszont valaki elmagyarázza, hogy hogyan kell kosarat dobni, egészen addig, amíg néhány százszor vagy ezerszer meg nem próbáltuk, nem leszünk jó játékosok. Az oldalszám példa a deklaratív memóriára, míg a másik a procedurálisra. A hippokampális lézióknak alávetett emberek továbbra is képesek procedurális memóriát igénylő feladatokat megtanulni. Azonban hiába nagyon jók már egy ilyen feladat teljesítésében, minden gyakorlásnál azt állítják, hogy a feladattal még soha nem találkoztak és nem is hiszik el, hogy képesek lennének megcsinálni!

A hippokampális beteg képes normál beszélgetést folytatni, de képtelen az elhangzott információkat más beszélgetésben felidézni. Például H. M. képes volt nehéz keresztretjvényeket megoldani, de ezután hamarosan elfeledkezett róluk. Klasszikus példa azon páciens esete, aki beszélgetett valakivel, nevén nevezte az illetőt stb., de ha az illető kiment a szobából majd visszajött, a páciens soha nem látott személyként üdvözölte. Ez némileg magyarázható az eddig elmondott elméletekkel, mert a szobába visszatérő személy felismerése a kontextus előhívását igényli.

Az epizodikus memória elmélete azt állítja, hogy a hippokampusz a deklaratív memória ideiglenes tárolására szolgál. A procedurális memória viszont nem igényli a hippokampuszt.

Szekvenciák tárolása és előhívása

Levy és munkatársai szerint a hippokampusz szerepe szekvenciák tárolása és előhívása. Kimutatták, hogy a CA3 rekurrens kollaterálisai alkalmasak erre a célra. Ez hasonlít arra az útvonaltároló és előhívó mechanizmusra, amelyeket az útvonal-tanulásnál említettünk (25. oldal). Mivel a sorozatokat átfedő kóddal kell tárolni, a szerzők bevezetik a kontextuális egység fogalmát, ami az átfedő kód többértelműségeit megszünteti. Lényegében ezek rejtett egységek egy autoasszociatív hálóban. Ezt kiválthatja a referenciakeretek koncepciója (26. oldal), ami ugyancsak képes a keresztvező utak közti átfedések feloldására.

Visszajátszás alvás során

Marr 1970-es neokortextről szóló elméletében azt a javaslatot teszi, hogy a neokortex szerepe a klasszifikáció és a hosszú távú memória funkciójának ellátása. Ahhoz, hogy a neokortex két hasonló információt klasszifikáljon, előnyös volna, ha a rendszert egy másik struktúra alvás közben tanítaná. A szerző szerint a hippocampus játszhatja ezt a szerepet. Marr a kérdést főleg a memóriatárolás *helye* szempontjából közelítette meg, de újabb szerzők más lehetséges okokat is felhozhatnak arra vonatkozólag, hogy miért vonódik be a hippocampus a memória formálásnak folyamatába.

McClland és társai megpróbálták megindokolni, hogy miért szükséges egy átmeneti tároló a memória formálásában. Szerintük a trükk abban rejlik, hogy a konneccionista vagy PDP hálózatok, ahol a memória a szinaptikus súlymátrixokban tárolódik, hajlamosak katasztrofális interferencia szenvedő alanyai lenni. Az ilyen autoasszociatív hálózatok hajlamosak hibás korrelációkat kialakítani két olyan memória között, amelyek közös neuronokra épülnek. Ez azt is jelenti, hogy ha az állat egy olyan új dolgot tanul meg, ami interferál egy előzővel, akkor az elsőt elfelejti. A fimibria-fornix léziós patkányok mutatják ezt a jelenséget: ha felváltva tanítják őket két feladatra, akkor állatok elfelejtik az egyik feladatot, miközben a másikat megtanulják.

Az összes epizodikus memóriáról szóló elmélet kulcsa az az elképzelés, miszerint a hippocampus speciális struktúrája a memória gyors tárolását teszi lehetővé, míg a neokortex csak lassan tudja ugyanezt. McClland és munkatársai feltételezik, hogy így a neokortexnek lehetősége van olyan struktúrákat kialakítani, amelyek mentesek az interferencia problémájától.

A nemrég tanult memóriák alvás során a hippocampusban történő visszajátszását számos tanulmány erősíti meg. Azt a javaslatot, miszerint a memória a hippocampusba íródik theta fázis alatt és az SPW fázis alatt előhívódik, először Buzsáki vizsgálta részletesen (lásd 17. oldal). Hamarosan kimutatták, hogy theta fázis alatt az entorhinális cortex (EC) felszíni rétegei a theta ritmussal korrelált módon tüzelnek, míg a mély rétegekre ez nem igaz. Ellentétben ezzel, SPW alatt az EC mély rétegeiben levő sejtek a sharp wave-vel korrelálva tüzelnek, míg a felszíni EC rétegek sejtei nem. Ez azt a következtetést vonja maga után, hogy az EC mély rétegei képzik a kimeneti utat a hippocampusból, amikor az a memória konszolidációját végzi. Ez nem áll szükségszerű ellentétben azzal, hogy más kutatók a fornixot tartják a hippocampus fő kimenetének. A REM fázis EEG-je egyébként is nagyon hasonlít a theta fázisban tapasztaltnak. Ezért a REM fázist paradox alvási fázisnak is hívják, éppen azért, mert az agy ilyenkor az ébrenléti állapothoz hasonló jeleket ad.

Az az ötlet, hogy a REM alvásnak köze van a memória formálásához,

jóval korábbi eredetű. Mivel a REM fázis könnyen felismerhető az EEG-n, könnyű az alanyt felébreszteni, amikor alvása során REM fázisba ér. Az ide vonatkozó kísérletek úgy találták, hogy bizonyos összetett feladatok – mint például a rejtett platformos labirintus – betanításához az állatnak muszáj lehetőséget kapnia REM alvásra egy néhány órás ablakon keresztül.

3.2. Az elméletek összehasonlítása

Mindhárom fennmaradt elméletnek megvannak a maga problémái. A kognitív térkép elmélete nem tudja megmagyarázni azt a tényt, hogy a hippocampusz sejtjeinek tüzelése nem térbeli jellemzőkkel is korrelál. A kontextuális felidézés elmélete nem tudja megmagyarázni a különbséget a lokális és a referencia reprezentációk között. Másrészt nem világos, hogy az epizódok reprezentálása hogyan történik az agyban. Erre ma még megfelelő matematikai eszközeink sincsenek, ami komputációs modellek építését nagyon megnehezíti. A deklaratív memória eddig megjelent modelljei, amelyek szimulációkat is magukba foglalnak, csak random bináris vektorok tárolásával és előhívásával foglalkoznak. Habár ezek a modellek számos általános elvet érintenek, így például a memóriát is, konkrétan a hippocampusz szerepével egy adott feladatban nem foglalkoznak. Ez megnehezíti a modellek predikcióinak ellenőrzését valódi kísérletekben. Talán egyetlen kivétel Shen és McNaughton 1996-ból származó modellje, amely képes a nemrég felfedezett környezetek tárolására és előhívására.

3.3. Nem tájékozódással kapcsolatos funkciók

A hippocampusz piramissejtjei gyakran a környezet pozíciótól eltérő aspektusaival is korrelált tüzelést mutatnak, beleértve a lokális látványt, sebességet, irányt, fordulási szöveget, padlómintázatot, feladat típusát, feladat fázisát, környezeti hatásokat, mint például hogy felemelünk-e egy falat, bekapcsoljuk-e a fényeket, vagy forgatjuk a környezetet. Elképzelhető, hogy ezek közül sok tényező a referenciakeret kiválasztásához szükséges, de lehetséges az is, hogy semmi közük hozzá. Esetleg ezek a korrelációk a referenciakeret meghatározásához szükséges folyamatok zajaként mutatkoznak. Ezek a nem tájékozódással kapcsolatos sejtek a place sejtekhez hasonló kapcsolatrendszer alakíthatnak ki, ezzel tulajdonképpen a patkány állapotterében tárolnak útvonalakat. Az álmok visszajátszhatják ezeket az útvonalakat, úgy, mint ahogy valódi térben kialakult útvonalakkal teszik.

Azokban a feladatokban, ahol az állatnak nem kell szekvenciákat

megtanulnia, még nem világos, hogy milyen reprezentációkat kell visszajátszani az álom során. Winocur kísérleteiben kiderült, hogy a patkányok érzékenyek a szén-diszulfid szagára. A patkány agyában ezért a CS₂-vel párosított élelem beszerzéséről alkotott reprezentációja a neurális kapcsolatrendszer szempontjából erősebb lesz. Ez azt jelenti, hogy amikor a hippocampusz reprezentációt alakít ki szenzoros input hiányában, például az álom során, a szén-diszulfiddal társított élelem előhívása sokkal valószínűbb, mint a nem társított élelemé. Másképpen azt mondhatjuk, hogy ez a reprezentáció nagyobb területet foglal el a hippocampális asszociációk állapotterében. Megint más szavakkal az állat nagyobb valószínűséggel álmodik a szén-diszulfiddal párosított élelemről. Elég sok alvás után az élelem-CS₂ asszociáció a hippocampuszon kívül fog tárolódni és az állat még hippocampális lézió után is ugyanazt az élelmet fogja választani, mint ahogy ezt a kísérletekben teszi is. Shen és McNaughton ennek megalkotta egy modelljét, megmutatva, hogy az újabban bevésődött tapasztalatok gyakrabban játszódnak vissza az álom során.

3.4. Akkor tehát mit csinál a hippocampusz?

Redish mindezek összefoglalásául amellet érvel, hogy a hippocampusz a *a kontextus előhívását, felidézését* végzi. A navigációban ez az állat saját pozíciójának tudatosítását és az ismerős környezet felismerését jelenti. Ez az értelmezés megengedi, hogy a hippocampusz gyors deklaratív memóriaként működjön. Ekkor az állapotterben vagy a valódi térben felépült útvonalak a hippocampusz rekurrens kollaterálisiban tárolódnak. Az álom alatt aztán ezek az utak visszajátszódnak és procedurális jellegű memóriákat formálnak az agy többi részében.

Ha a memóriák egész szekvenciája képes visszajátszódni a hippocampuszban, akkor ez képessé kell tegye a struktúrát még csak bekövetkezés előtt álló események tapasztalaton alapuló *jóslására*. A következőkben megvizsgálunk egy olyan modellt, ami feltételezi, hogy a hippocampusz fő funkciója az információátvitel és előhívás mellett ilyen predikciók készítése.

4. fejezet

Faktoriális kód a hippokampuszban?

Az előző fejezet végén kitértünk a hippokampális modellek egy lehetséges bővítési irányára, arra, hogy a struktúra képes lehet *predikciók* készítésére. Természetesen ilyen modell is szép számmal készült. Talán elsőként a prediktív komparátor elméletet említhetjük (3.1.2. rész), ami az előző fejezetben is szerepelt. Carpenter és Grossberg 1993-ban például olyan modelleket készített, amelyek az inputot a várakozásokhoz való viszonyában vizsgálják. Az utóbbi néhány évben számos csoport alkotott prediktív modelleket a hippokampuszról. Az egyik ilyen [Lőrincz and Buzsáki 2000] modellje, amit ebben a fejezetben tárgyalunk.

4.1. Érvek a faktoriális kódolás mellett

A faktoriális kódolás és ezen keresztül a ritka reprezentáció biológiai előnyeiről már sok szó esett az 1. fejezetben. Önmagukban már ezek is indokolnák, hogy megpróbáljuk a hippokampuszra alkalmazni a faktoriális kódolás elvét. Fontos megemlíteni azonban, hogy a memóriatárolás és visszajátszás igénye is a faktoriális kódolás mellett szól.

4.1.1. Szekvenciák tárolásának problémái

Nagy általánosságban fogalmazhatunk úgy, hogy a hosszú távú memória valószínű célja időben fenálló viszonyok és időbeli szekvenciák reprezentálása. Ez igaz az előző fejezetben említett deklaratív memória és a procedurális memória esetében is. A modellek többsége megegyezik abban, hogy a szekvenciák tárolásának legkézenfekvőbb módja egy neurális réteg rekurrens

kapcsolatainak felhasználásával történik. Más szóval egy ilyen szekvenciát úgy játszunk vissza, hogy valamilyen kezdeti aktivitásokkal ellátjuk a neurális réteget, majd ez a neurális réteg saját magára visszahat a szinaptikus kapcsolatain keresztül: $\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t$, itt \mathbf{A} a szinaptikus súlymátrix. A szekvenciák visszajátszásának ez a roppant egyszerű lineáris modellje számos buktatót rejt magában. Először is problémás a beindítása és a leállítása, ami mindenképp nemlineáris műveleteket igényel. Másodsor semmiképpen sem alkalmas pontos visszajátszásra. Ha két szekvencia például történetében keresztezi egymást, azaz azonos állapotot vesznek fel, a folytatás mindig csak egyféle lehet, sohasem kétféle. Magyarán ily módon csak lineáris iteratív leképzésre illeszkedő sorozatokat tárolhatunk. A helyzet bizonyos szempontból javul, ha – mint már azt az 1. fejezetben említettük – a neuronális aktivitásokat valószínűségi eseményeknek tekintjük. Ekkor a súlymátrix egy elsőrendű Markov-folyamat átmeneti mátrixa lesz, és az ilyen rendszer kimenetének csak valószínűségi értelmezést adhatunk. Ez részben kiküszöböli az egzakt szekvenciák visszajátszásának problémáját, mert így bármilyen elkódolt szekvencia előhívható, azonban azok előhívásának valószínűségét továbbra is csak a linearitás keretein belül tudjuk szabályozni. Ezenkívül megkülönböztethetetlenek maradnak azok a szekvenciák, amelyek közös állapoton osztoznak és nem oldja meg a pontos visszajátszást sem, amire pedig feltehetőleg szükség van. Mint ahogy azt már az előző fejezetben is említettük, a szekvenciák neuronális tárolására ma még nem létezik egységes és üdvözítő eljárás. A megoldási kísérletek többnyire az állapottér speciális kiterjesztéseivel próbálkoznak, mert így előbb-utóbb minden bonyolult állapotátmenet tökéletesen ábrázolhatóvá válik egy lineáris Markov-folyamattal. A kiterjesztés módjairól azonban nem egységesek a vélemények. Lehetséges például *fák keverékét* reprezentálni [Meila 1999]. Itt a lehetséges állapotátmenetekhez tartozó fáknek egy lineáris keverékét tartjuk számon. Egy másik lehetséges kiterjesztésről a következő fejezetben esik majd szó.

4.1.2. A minimalizált kölcsönös információ előnyei

A memóriában párhuzamosan sok-sok szekvencia lejátszására lehet szükség. Ha ezek a folyamatok kevéssé vagy egyáltalán nem függnék egymástól, akkor azok átfedése minimális, tehát a felidézés annál biztosabb és a szekvenciák annál hosszabbak lehetnek. Meila és Jordan egy 1998-as cikkükben bizonyítják, hogy az időbeli viszonyok bizonyos értelemben vett legjobb reprezentációját akkor kapjuk, ha az egységeket kölcsönös információjuk mértékében kötjük össze. A kölcsönös információt két véletlen változó között a következőképpen definiáljuk:

$$I(A; B) = \int \int p_{AB}(x, y) \log \frac{p_{AB}(x, y)}{p_A(x)p_B(y)} dx dy$$

Itt p_{AB} a két változó együttes eloszlását, p_A és p_B pedig a megfelelő peremeloszlásokat jelöli. A kifejezés egyébként megegyezik az együttes eloszlás és a peremeloszlások szorzatából kapott eloszlás relatív entrópiájával vagy Kullback-Leibler távolságával. Látszik, hogy a kifejezés független valószínűségi változók esetén zérus. Ez azt jelenti, hogy a komputációs egységek optimális kapcsolati gráfja akkor esik szét különálló komponensekre azaz akkor reprezentál különálló időbeli láncokat, ha a komponensekben szereplő valószínűségi változók függetlenek egymástól. Ez egy fontos érv a faktoriális kódolás mellett.

4.2. A modell áttekintése

Térjünk most vissza [Lőrincz and Buzsáki 2000] modelljéhez. Az elképzelésben a hippokampális formáció és az entorhinális kortex együttesen egy visszacsatolt hurkot képez (erre nézve az anatómiai adatokat lásd 17. oldal). Ez a hurok a már említett rekonstrukciós hálózat (12. oldal) egy megvalósítása: a neokortex által létrehozott reprezentációk és a hippokampusz által rekonstruált reprezentációk az entorhinális kortex felszíni rétegeiben összehasonlításra kerülnek, és ez a különbség (a rekonstrukciós hiba) az, ami a hippokampusz kapcsolatrendszerét aktiválja. Ha a neokortex inputja nem új a hippokampuszban tároltakkhoz képest, akkor a rekonstrukció tökéletesen végbemegy, és a hippokampusz visszajáttsza az előzőleg eltárolt memóriákat. Ha azonban új információ kerül a hippokampusz bemenetére, az szinaptikus változásokat kezdeményez a hippokampusz belső kapcsolatrendszerében. Ugyanakkor, mivel a rendszer a differenciákon iteratív módon működő visszacsatolt hurok, ezért az adott bemenethez tartozó rekonstruált eredmény csak bizonyos késleltetéssel lesz hozzáférhető. Ez problémát jelent, mert ez esetben az összehasonlítás nem történhet összetartozó reprezentációkon. Azonban a késleltetések megléte kifejezetten alkalmassá teszi a rendszert predikciók hebbi úton történő tanulására. A szerzők felteszik, hogy miután a rendszer a helyes predikciókat elsajátította, ezek felhasználásával képes a késleltetések-ből adódó problémák kiküszöbölésére.

A hippokampusz a rekonstrukciós hibán tanul, ezért a rendszer egyensúlyát (azaz amikor már nem történik további változás) akkor nyeri el, amikor a rekonstruált input megegyezik a neokortex eredeti inputjával. Ez azt jelenti, hogy a rendszer kialakította a saját reprezentációját a neokortex felől

közvetített információról. A reprezentációk ezek után az entorhinális kortex mély rétegein keresztül a neokortexbe kerülnek (lásd még 29. oldal). Ez azt is jelenti, hogy ebben az elméletben a hippocampusz centrális szerepet játszik a memória formálásában. Elengedhetetlen tehát a reprezentációk optimális kialakítása. Itt a szerzők a *faktoriális kódolást* javasolják az ICA neurális implementációjával. Ennek kialakítására a CA3 és a CA1 régiók kapcsolati rendszere alkalmasnak tűnik. A nemlinearitások miatt azonban lényeges, hogy a hippocampusz két üzemmódban tudjon működni, az egyikben reprezentációt előhívó, a másikban reprezentációt kialakító módba kapcsolva, és így ugyanazon struktúra képes mindkét feladatot ellátni. Ezt a két állapotot a szerzők a hippocampusz theta és SPW fázisaival azonosítják.

4.2.1. Tanulási szabályok

A legtöbb idegrendszeri modell hebbi tanulást használ. Ez azt jelenti, hogy a szinapszis a preszinaptikus és a posztszinaptikus depolarizáció együttes jelenléte esetén módosul, ami a szinaptikus mátrix változásának kifejezésében egyszerű diadikus szorzatként jelenik meg:

$$\Delta \mathbf{P} = \alpha \mathbf{x} \mathbf{a}^T$$

itt α egy hangolási paraméter. \mathbf{x} az input, \mathbf{a} az output vektort, \mathbf{P} pedig a szinaptikus mátrixot jelöli. Az output $\mathbf{a} = f(\mathbf{P}^T \mathbf{x})$ módon adódik, ahol f egy esetleg identitással azonos nemlinearitás.

A szinapszis módosításához komoly posztszinaptikus depolarizációra van szükség. Ez elérhető a preszinaptikus neuronok magasabb aktivitásával vagy korrelált tüzelésével. Az is világos, hogy mivel egy erősebb szinapszis nagyobb posztszinaptikus depolarizációt hoz létre, az ilyen szinapszisoknak nagyobb az esélye az erősödésre. Ez azt jelenti, hogy a tanulási szabályba egy, a szinapszis erősségével arányos tag kerül. Az erősödő szinaptikus aktivitást a hippocampuszban a rekurrensen kapcsolt gátló interneuronok hatása limitálja (lásd 17. oldal). A P_{ik} szinapszisa (ami az input réteg k -edik neuronját köti össze az output réteg j -edik neuronjával) jutó gátló hatást $\sum_j P_{ij} a_j$ -vel közelíthetjük, azaz minden neuronra akkora gátló hatás jut, mint amekkora a vele kapcsolatban álló neuronok aktivitásának súlyozott összege. A súlyozást itt az indokolja, hogy az erősebb szinapszisok nagyobb valószínűséggel járulnak hozzá a gátláshoz. Ugyanakkor figyelembe kell azt is vennünk, hogy akármekkora legyen is a szinaptikus súly, nincs hatása, ha a kapcsolt neuron posztszinaptikus aktivitása nulla. Ettől eltérő esetben is arányos a gátló hatás az illető neuron posztszinaptikus aktivitásával. Ezért a gátló tagra végül a $\mathbf{P} \mathbf{a} \mathbf{a}^T$ kifejezés adódik. A teljes tanulási szabály tehát

$$\Delta \mathbf{P} = \alpha \mathbf{x} \mathbf{x}^T + \beta (\mathbf{P} - \mathbf{P} \mathbf{a} \mathbf{a}^T) \quad (4.1)$$

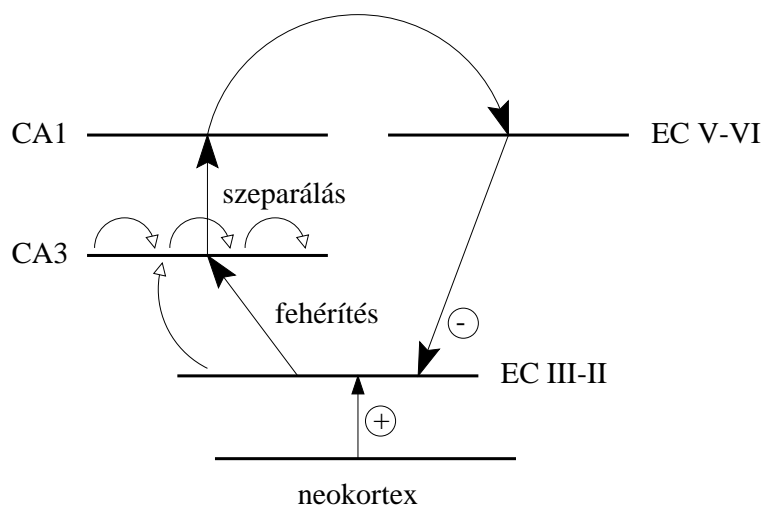
A szerzők felteszik, hogy β egy a tüzeléstől függő paraméter, ami alacsony tüzelési frekvenciáknál nullává válik. Kis szinaptikus erősségeknél a második tag lineáris része dominál. Ezért ha β elég nagy, akkor a szinaptikus súlyok nem lehetnek nagyon kicsik. Másrészt a gátló faktor erős szinapszisoknál normáló szerepet tölt be; a szinapszisok nem lehetnek tetszőlegesen nagyok. Nagyobb tüzelési frekvenciáknál a β növekedése miatt a tanulási szabály hebbi tagja elhanyagolható.

4.2.2. Anatómiai megfeleltetések

Mint azt már a hippocampusz anatómiájánál említettük, a CA3 már feldolgozott inputot kap a dentate gyrus és az entorhinális kortex felől. Itt a sejtek aktivitása a hippocampusz mindkét működési fázisában igen magas. A 4.1. összefüggésnél megjegyeztük, hogy ilyenkor a hebbi tanulás elhanyagolható, és a tanulási szabály a második tagjára redukálható. Cardoso és Laheld 1996-ban megmutatták, hogy ez a szabály képes fehérítést végezni (lásd 8. oldal). A fehérítés ezek szerint a hippocampusz mindkét működési állapotában végbemegy. A kísérletek azt mutatják, hogy a CA3 rekurrens kollaterális rendszere theta alatt jórészt hallgat. Ezért a CA3 theta fázisban csak a hibát, míg SPW alatt a rekurrens kollaterálisok jelét is fehéríti. A szerzők felteszik, hogy a tanulási szabályok lehetőséget adnak arra, hogy a rekurrens kollaterálisok theta alatt a bennük lejátszódó szekvenciákat tanulják meg predikálni, majd SPW alatt képesek ezeket a szekvenciákat visszajátszani. Ez egyúttal azt jósolja, hogy a CA3 kimenete theta és SPW alatt nagyon hasonló.

Az elmélet szerint a CA3 \Rightarrow CA1 kapcsolat szinapszisai szeparálni tanulnak. Theta alatt a CA1 piramissejtjei alacsony frekvenciával tüzelnek, ezért a 4.1. szabály hebbi része van érvényben. SPW alatt azonban érvényre jut a tanulási szabály második tagja is, ami így épp a 1.2. egyenletet adja. Ha ehhez hozzátesszük, hogy az ide érkező jelek már fehérítve vannak, akkor a független komponens analízis neurális megvalósítása működni kezd, és a réteg kimenetét független komponensek fogják képezni.

A szerzők az entorhinális kortextet teszik meg architektúrájuk input és output elemének. A III. réteg kapja a neokortikális inputot. A II. réteg a III. réteg topologikus leképzését és a III. réteg felé közvetített neokortikális inputot kapja, valamint a entorhinális kortex mély (V-VI) rétegeiből származó efferensekkel is van mind serkentő, mind gátló kapcsolata. A kutatások azt mutatják, hogy ez a kapcsolat jórészt direkt vagy visszacsatolt *gátlást* közvetít. Ezért az elmélet szerzői úgy gondolják, hogy az entorhinális



4.1. ÁBRA. Az entorhinális-hippokampális hurok sematikus rajza. Az üres nyilak a predikciós kapcsolatokat jelzik.

kortex második rétege képi a rekonstrukciós hibát a neokortikális input és a hippocampusz által rekonstruált reprezentáció között.

Az entorhinális kortex mély rétegeit a CA1 látja el információval.¹ Ez a kapcsolat topografikusan szervezett. A modell felteszi, hogy itt a CA1 másolata jelenik meg. Az ebből a rétegből a II. rétegbe kiinduló szinapszisok képzik az entorhinális kortexben tárolt hosszútávú memóriát. [Fomin et al. 1997] megmutatja, hogy egy rekonstrukciós hálózatban egy ilyen „leszálló” ágban a hebbi tanulás a „felszálló” ág transzformációs mátrixának pszeudoinvertet tanulja meg, ami ha a felszálló ág független komponens analízist végez, épp az ICA bázisának megtanulását fogja jelenteni (lásd még a 4.1. ábrát). Az itt tárolt reprezentációk SPW alatt egy felügyelt tanulási szabállyal továbbíthatók az agy más területei felé. Ez nagy előnyére válik a neokortexnek, mert annak így nem kell kétféle üzemmódban működni.

A szerzők a már említett késleltetést kiküszöbölő predikciós rendszert az entorhinális kortex III. rétege és a CA1 mező között fennálló topografikus kapcsolatokban vélik felfedezni. Ha ez a predikciós rendszer megtanult helyesen következtetni, akkor ez lecsökkenti a rekonstrukció előhívásához szükséges időt (mint már azt többször említettük, a rekonstrukciós hálózat a deriváltfolyamaton dolgozik, és ezért a „valódi” rekonstrukció elérése közben egy differenciálegyenlet iteratív megoldását végzi).

¹a szerzők nem foglalkoznak a subiculum szerepével (lásd 17. oldal).

A dentate gyrus struktúrája a szerzők szerint *vak dekonvolúciót* végez. A vak dekonvolúció hasonlít a vak forrásszétválasztás feladatához (6. oldal). A feladat ez esetben egy jel dekonvolválása, anélkül, hogy ismernénk az eredeti konvolúciós kernelt. A probléma átvihető egy vak forrásszeválasztási problémára, ha az eredeti jelet késleltető vonalakon keresztül megsokszorozzuk. A statisztikai redundancia itt ugyanis nem több jel között áll fenn, hanem ugyanazon jel eltolt változatai között. A szerzők feltételezik, hogy a neokortex rekonstrukciós hálózatokat tartalmaz. Ha megoldjuk egy ilyen rekonstrukciós hálózat dinamikáját leíró differenciálegyenletet, megmutatható, hogy az ilyen hálózat konvolúciót (is) végez. A dentate gyrus túlságosan kicsi egy általános vak dekonvolúció végrehajtására. A szerzők cikkükben bizonyítják, hogy a dentate gyrus mégis képes ezt a feladatot végrehajtani, mert a rekonstrukciós hálózat által okozott konvolúción lehetséges a *komponensenkénti* vak dekonvolúció. A cikkben ezt konkrét példán modellezzük is. A műveletre azért van szükség, mert a konvolúció megnehezíti a jelek statisztikai analizését. Ez magyarázza azt is, hogy a dentate gyrus miért épp a CA1 régió előtt található.

A rendszer összefoglaló rajzát láthatjuk a 4.2. ábrán.

4.2.3. További bizonyítékok

A modell az entorhinális-hippokampális rendszer fő feladatát a neokortex hosszútávú memóriájának kialakításában látja. A szerzők felteszik, hogy a neokortex rekonstrukciós hálózatokat valósít meg. Ez lehetővé teszi, hogy a neokortex a hippokampusz nélkül is képes legyen időbeli relációkat tanulni, összhangban a tényekkel. A legfontosabb különbség a neokortikális és a hippokampusz által kialakított reprezentációk között, hogy a hippokampusz reprezentációi a 4.1.2. részben említett értelemben optimálisak, ezért a hippokampusz nélküli tanulás a fejezet elején említett deficitektől hatványozottan szenved. Ezért a hippokampusz hiánya az epizodikus memória tönkretételében fog manifesztálódni.

Ha a hippokampusz megtanulta a tökéletes rekonstrukciót, a szinaptikus rendszer ugyanazt a kimenetet fogja adni, feltéve, hogy a neokortex felől ugyanaz az input érkezik. A megtanult reprezentáció először az entorhinális kortextben tárolódik, de előhívásához szükséges a hippokampusz is. Ez magyarázza az előző fejezetben említett, a hippokampális betegek esetében kimutatott retrográd amnézia jelenségét, ami a neokortexben még nem tárolt memórianyomokat érinti.

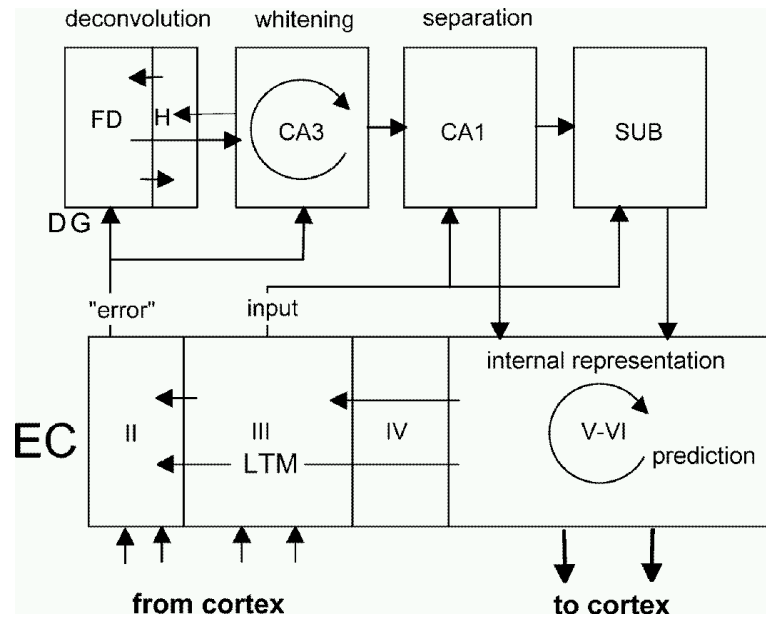
A tanításhoz a rendszernek két különböző működési módban kell üzemelnie, mert mint láttuk, a független komponensekből álló kimenet kialakításához lineáris és nemlineáris tanulási szabályokra egyaránt szükség van. Ezek egyszerre történő alkalmazása nem lehetséges ugyanazon struktúrában.

4. FEJEZET. FAKTORIÁLIS KÓD A HIPPOKAMPUSZBAN?

Kísérleti bizonyítékok vannak arra, hogy mind az SPW, mind a theta fázis szükséges a memória kialakításához. Wilson és McNaughton egy 1994-ben végzett kísérletükben észrevették, hogy a környezet bejárása alatt együtt tüzelő place sejtek korrelációja nagyobb volt az explorációt követő SPW fázisban, mint előtte. Nádasdy és munkatársai nemrég ismétlődő mintázatok kerestek a két fázisban. Úgy találták, hogy az exploráció alatti tüzeléssorozatok időben kompresszált formában újra megjelennek SPW alatt. Röviden a bizonyítékok amellet szólnak, hogy az SPW időtartama alatt megjelenő neuronális reprezentáció a tapasztalás alatt végbement szinaptikus módosulásokat tükrözi.

A modell szerint a theta fázisban működő predikció kiküszöböli a hurok okozta késleltetéseket. Más kutatók is úgy vélik, hogy a principális sejtek szómájához közeli terminálissal rendelkező interneuronok képesek a principális sejtek tüzelését a theta hullámok fázisához képest siettetni vagy késleltetni. Ez a szimulációk szerint 50-100 ms-os eltolásokat jelenthet. A fázisprecessziós jelenség (17. oldal) ugyancsak az ilyen predikciós struktúrák létét valószínűsíti.

Végezetül egy megjegyzés: Czurkó és munkatársai által nemrég végzett kísérletek arra utalnak, hogy a rekonstrukciós hibán történő tanulás működhet a hippocampusban. Úgy találták, hogy a place sejtek tüzelésének átlagos frekvenciája az állat sebességének növekedésével arányosan nő. Ezt magyarázhatjuk a modell alapján: az input gyorsabb változása miatt a rendszer csak lassabban képes utánaállni a változásoknak, azaz a rekonstrukciós hibával a neuronok átlagos aktivitása is nagyobb lesz.



4.2. ÁBRA. A modell összefoglaló rajza [Lőrincz and Buzsáki 2000]

5. fejezet

Az ICA kiterjesztése

5.1. A független komponens analízis egy fontos problémája

Mint ahogy azt az előző fejezetben láttuk, életképesnek tűnik az a feltételezés, hogy a hippocampusz feladata faktoriális kódolás segítségével optimális neokortikus reprezentációk kialakítása. A faktoriális kódolás megvalósításának valószínű módja a független komponens analízis. Az előző fejezetben amellet láttunk érveket, hogy a hippocampusz CA3 és CA1 rétegeiben működő szinaptikus hálózat az ICA algoritmusának egy biológiai implementációja. Azonban most muszáj lesz kitérnünk ennek az implementációnak egy lényeges hátrányára.

Mint azt a hippocampusz anatómiájánál láttuk (17. oldal), a hippocampusz komoly mennyiségű szenzoros területekről származó inputot kap. A neurális folyamatok időskálája miliszekundumokban mérhető, míg az állat saját időskálája (például a mozgásé) mindenképp a másodperces tartományba esik. A szenzoros információ neurális továbbítása tehát örökli a szenzoros változásokban meglévő folytonossági tulajdonságokat, azaz az a tény, hogy például az állat által látott kép az állat fejmozgása során folytonosan transzformálódik, a neurális aktivitások időn átívelő korrelációiban is megjelenik: egy szenzoros információt továbbító neuron tüzelése jól predikálható a megelőző aktivitásaiból. Ez azt jelenti, hogy a hippocampusz bemenetére érkező jelek időben nem pusztán statisztikus tulajdonságokat mutatnak.

Ez neurális implementációknál semmiképpen sem előnyös, sőt, lehetlenné teheti a szeparációt. Mint azt az első fejezetben láttuk, a független komponens analízis az input statisztikai tulajdonságait használja fel. A teljes inputsorozat statisztikai jellemzőihez azonban csak a nem-adaptív

algoritmusok esetében lehet hozzáférni, amit biológiai esetben igen nehézkes volna implementálni¹. Az adaptív algoritmusok kizárólag a pillanatnyilag hozzáférhető mintával dolgozhatnak, az addigi statisztikáról pedig folyamatosan állított paramétereik révén van tudomásuk. Ha az input időben „rosszul” (korreláltan) viselkedik, könnyen előfordulhat, hogy az algoritmus a teljes adatsor feldolgozása során többnyire a teljes adatsorra már nem érvényes statisztikai tulajdonságokat használ fel, és így kimenete távol áll a valódi független komponensektől. Azt is láttuk, hogy a tanulás előrehaladtával gyakran csökkentik a tanulási paramétereiket a konvergencia elősegítése érdekében. Mindez azt eredményezi, hogy a minták figyelembevétele nem azonos súllyal történik, könnyen előfordulhat tehát, hogy az adaptív algoritmus az iteráció végén hibás bázist alakít ki. A probléma elkerülése végett az adaptív független komponens analízist végző algoritmusok bemenetére küldött adatsoron gyakran egy megelőző véletlen *keverést* végeznek így alakítva ki a megfelelő statisztikát. Ez azonban egy valódi online rendszer esetén, mint amelyet a hippokampuszban feltételeztünk, nem járható út.

5.2. Egy lehetséges megoldás (T-ICA)

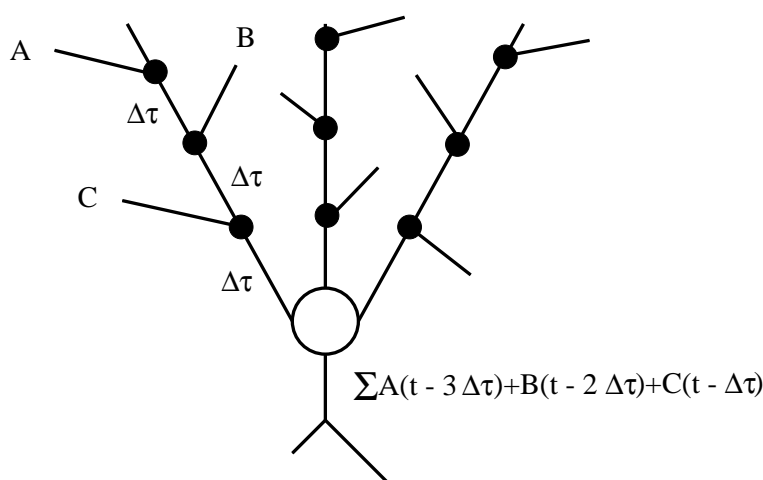
A fent vázolt probléma megoldódik, ha a független komponens analízist nem egyedi mintákon végezzük, hanem *minták egy konkatenált időbeli sorozatán*. Az inputok állapotterének dimenziószámát tehát kibővítjük, mégpedig annyiszorosára, ahány egymás után következő mintát fűzünk egybe:

$$\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_D, \dots, \mathbf{x}_i, \dots \rightarrow \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_D \end{pmatrix}, \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \mathbf{x}_{D+1} \end{pmatrix}, \dots, \begin{pmatrix} \mathbf{x}_i \\ \mathbf{x}_{i+1} \\ \vdots \\ \mathbf{x}_{i+D-1} \end{pmatrix}, \dots$$

Az állapotter ilyen kibővítését *beágyazásnak* (*embedding*), az összefűzött minták számát pedig, amit itt D jelöl, *a beágyazás dimenziójának* (*embedding dimension*) fogom hívni. Ezt a konstrukciót gyakran használják predikciós rendszerek létrehozásánál a matematikában.

A független komponens analízis ezzel a kiterjesztéssel időbeli dimenziót nyer, ezért a módszert jogos *temporális független komponens analízisnek* (*T-ICA*) elnevezni. A komponensek független volta ebben a modellben tér- és időbeli függetlenséget jelent. Ha a beágyazás dimenziója elég nagy, akkor

¹Bár nem lehetetlen. Majmok agyában már mutattak ki órákon át fennmaradó neurális mintázatokat. Egy ilyen implementáció azonban több problémát vetne fel, mint amennyit megold.



5.1. ÁBRA. Késleltetések a dendritfán

a sorozatok közti időbeli korrelációk megszűnnek, és az előző fejezetben említett probléma megszűnik.

5.3. Biológiai háttér

Henze és munkatársai által 1996-ban elvégzett modellezések azt mutatják, hogy a CA3 principális sejtjeinek dendritfáján előforduló késleltetések időskálája 10-100 ms-os. Ez elég tekintélyes, összehasonlítva például azzal, hogy a vizuális információ ventrális áramánál (ami a látókéreg felől az inferotemporalis kortex felé halad) a teljes áthaladás jellemző ideje 80-120 ms. Ez lehetségessé teszi a fenti elképzelést (lásd az 5.1. ábrán).

Az ötletnek ellentmondani látszik, hogy denriteken nem kisüléssel, hanem dekrementális impulzusvezetés van, tehát a távolabbi szinapszisok kisebb súllyal kerülnek összegzésre. Jaffe and Carnevale 1999-es modelljei arra utalnak, hogy a CA3 principális sejtjeiben a dekrementalitás nem érvényes, azaz a dendritfán érkező jelek lényegében nem gyengülnek.

A szekvenciák létrehozásának fent vázolt modellje nem új: számos kutató tanulmányozta már a késleltetett jeleket feldolgozó hálózatokat. Levy 1996-ban már publikálta a CA3 olyan modelljeit, amelyek szekvenciákon tanulnak. Lisman több cikkében is felvetette, hogy a dendritfán megjelenő késleltetések alkalmassá teszik a dendritfát különböző időmélységből származó információk együttes felhasználására.

Az elképzelés egyik következménye, hogy a szekvenciák maximális hossza maximum néhány száz miliszekundum. Ez nem tenné lehetővé valódi epizódok rögzítését, mert az állat valódi időskálája ennél jóval nagyobb.

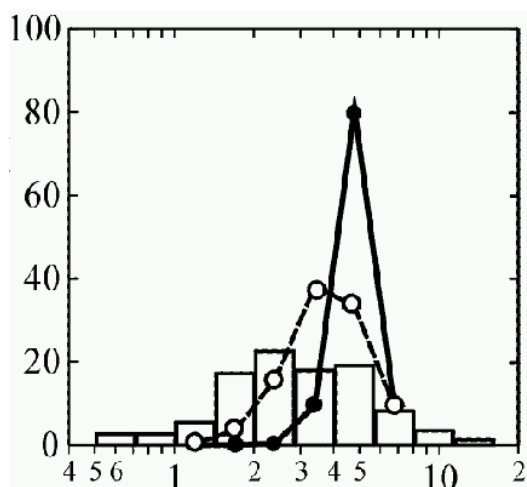
Szükségesnek látszik tehát egy mechanizmus, ami időbeli kompressziót végez. [Skaggs et al. 1996] kísérleti úton igazolta, hogy egy ilyen mechanizmus bizonyosan létezik a hippocampusban, nevezetesen a place sejtek tüzelésének fázisprecessziója(17. oldal) ilyen szekvenciatömörítést mutat. Skaggs és munkatársai úgy találták, hogy az állat pályának aktuális részlete a place sejtek tüzelésében elkódolva minden theta ciklusban lejátszódik. A befutott pálya, amit a place sejtek ebben a szekvenciában ismételnék, időben jóval nagyobb távolságra nyúlik, mint a theta egy ciklusa. A kompresszió mértéke akár 1:10 arányú is lehet. A fázisprecesszió maximális kiterjedése 100 ms körül van, ami éppen a dendritfán előforduló késleltetések nagyságrendje. Elképzelhető, hogy a tömörített információ újból feldolgozásra és tömörítésre kerül, így az agyi hierarchia magasabb szintjein már jelentős időtartamok áthidalása is lehetséges.

5.4. T-ICA a látókéregben

Henze szimulációi azt is megmutatták, hogy a dendritfa hasonló késleltetései előfordulhatnak az elsődleges látókéregben is. Elképzelhető tehát, hogy az input kiterjesztése inputok szekvenciáira a vizuális információk feldolgozásában is szerepet kap.

Épp ezt látszik igazolni Hateren és Ruderman 1998-as publikációja [Hateren and Ruderman 1998]. Mint azt az első fejezetben említettük, az elsődleges látókéreg sejtjei az input időbeli tulajdonságaira is érzékenyek. A független komponens analízis eredményéül kapott Gábor-szűrők a receptív mezők alakját azonban kvalitatív értelemben és *csak térben* reprodukálják. A biológiai szűrők várható időbeli frekvenciaeloszlásáról semmit sem mondanak, mert az analízis eredményeképp kapott szűrők csak térbeli tartományra terjednek ki. Ezért Hateren és Ruderman megismételte a Gábor-szűrők kialakítását célzó vizsgálatokat, de úgy, hogy természetes képeken végzett független komponens analízis helyett természetes képsorozatokon végeztek ICA-t, azaz temporális független komponens analízist alkalmaztak. A teljes mintasorozat több mint félmillió darab, televízióadásokból digitalizált kisméretű ($12 \times 12 \times 12$ -es) videószekvencia volt. Ezen a mintasorozaton 1728 dimenziós független komponens analízist végeztek. Mivel számítás még Cray gépeken futtatva is komoly számítási problémákkal járt (naponta csupán 15 szűrőt tudtak meghatározni), ezért az eredmények analíziséhez csak ez először kiszámolt 288 szűrőt használták fel². Az eredményeket összehasonlították De Angelis és munkatársai által a macskák látókéregben

²Teljesen végigszámolták $8 \times 8 \times 8$ -as szekvenciákon, de lényeges eltérést nem tapasztaltak, eltekintve a határok miatt megjelenő artifaktumok növekvő számától.



5.2. ÁBRA. A kapott szűrők térbeli frekvenciájának eloszlása. A frekvencia a Fourier-transzformáció utáni maximális amplitúdóból került meghatározásra. Az x tengelyen a frekvencia látható (periódusok száma/fok egységben), az y tengelyen pedig a előfordulás gyakoriságának százaléka van feltüntetve. Összefüggő vonal jelöli az egyszerű képkockákon végzett ICA, szaggatott vonal a szekvenciákon végzett T-ICA eredményét. A hisztogram makákó majmok elsődleges látókérgében egyszerű sejteken mért eredményeket mutat. [Hateren and Ruderman 1998]

talált receptív mezők tér- és időbeli frekvenciaeloszlásával. A szerzők úgy találták, hogy a T-ICA által nyert szűrők nagy tér- és időbeli hasonlóságot mutatnak a biológusok által talált receptív mezőkkel, jobb egyezéssel, mint amit szimpla ICA algoritmussal kapnánk. Egy másik biológiai adatsoron végzett meggyőző összehasonlítást mutat az 5.2. ábra.

Úgy látszik tehát, hogy a temporális független komponens analízis működőképes elképzelés. A következő fejezetben megpróbáljuk a módszert egy egyszerű szenzoros modellen alkalmazni.

6. fejezet

A T-ICA alkalmazása egy egyszerű modellen

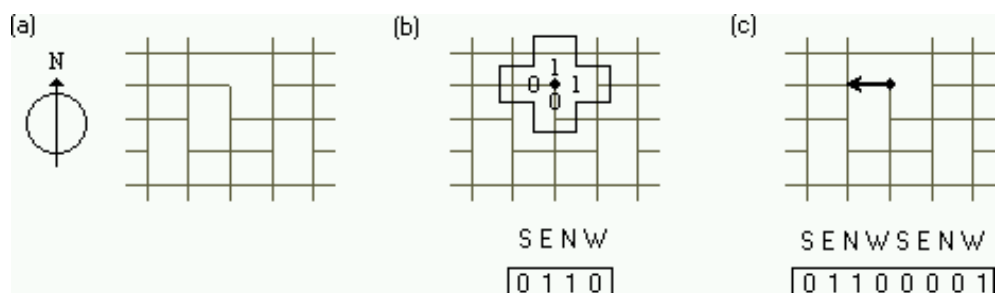
A negyedik és az ötödik fejezetben végignéztük az amellet az elképzelés mellett szóló érveket, miszerint a hippocampusz a neokortikális inputon temporális független komponens analízist végez. Az elképzelés szimuláción tesztelhetőnek tűnik, de ehhez szükségünk lesz a neokortikális input egy modelljére.

6.1. Az input modellje

Láttuk, hogy a hippocampusz komoly kapcsolatban áll az agy szenzoros információt feldolgozó területeivel (17. oldal). A hippocampuszban feltételezett kognitív térkép felhasználásához nyilván kiterjedt szenzoros információra van szükség. A modellek általában megkülönböztetnek *belső* és *külső* szenzoros információt. Az utóbbi lehet például egy látvány, egy szag, egy hanghatás, az utóbbi például egy izom pillanatnyi állapota, egy adott feladat végrehajtásának stádiuma stb. Láttuk azt is, hogy a hippocampusz principális sejtjei érzékenyek ilyen nem-térbeli jellemzőkre, többek között az állat irányára is. Jogos tehát az a feltételezés, hogy a hippocampusz mind belső, mind külső szenzoros információt felhasznál működése közben.

A külső információ generálásához egy egyszerű labirintust állítottam elő. A labirintus egy 20×20 oldalhosszúságú rácsból épült fel, úgy, hogy a rács rögzített számú élét kitöröljük (ez a továbbiakban tárgyalt példa esetében ... volt).

A modellben a kísérleti állat csak rácsponttól rácspontra mozoghat. A rács kitörölt élei nem járhatóak, ebben az esetben az állat az adott irányban falat lát. Az állat lokális vizuális információját tehát minden pozícióban

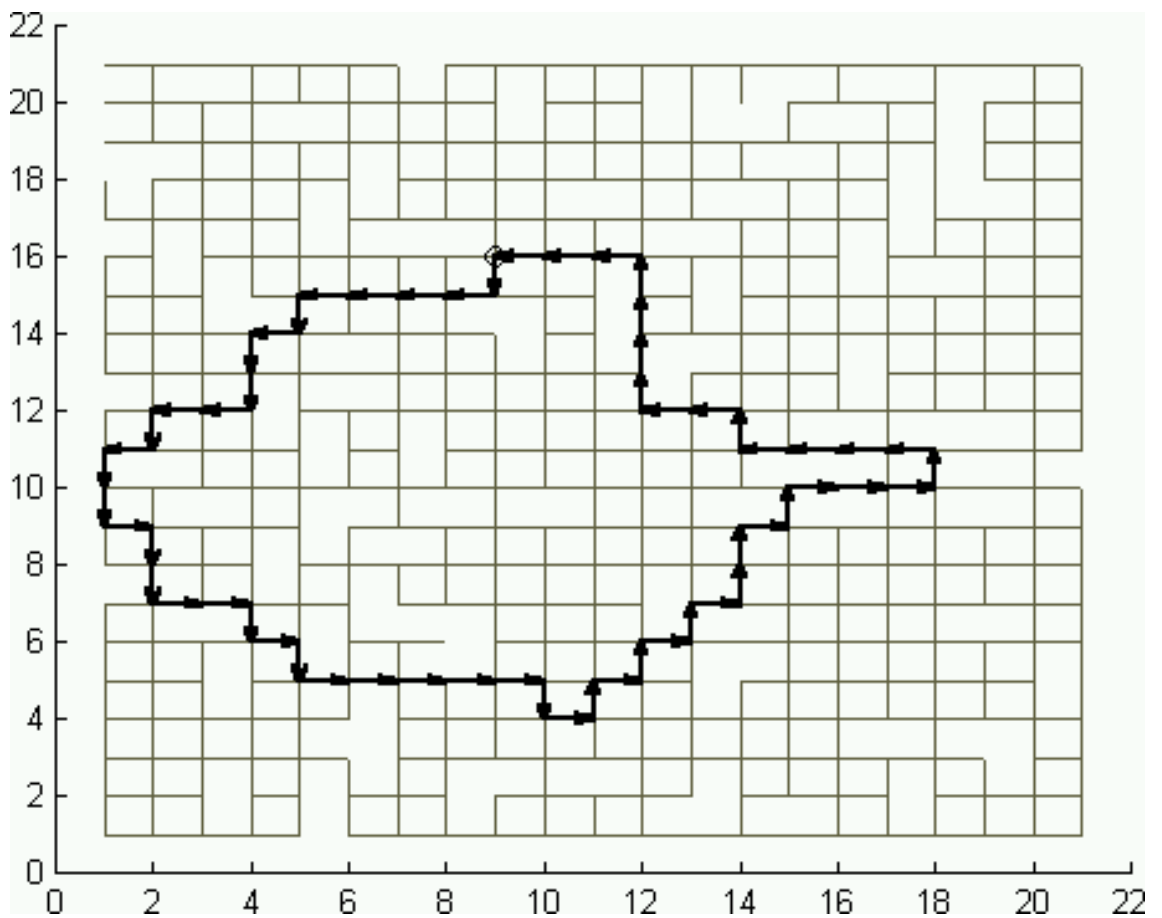


6.1. ÁBRA. Az input generálásának módja

egy négyelemű vektor írja le, amely elemei sorban az észak-kelet-dél-nyugat irányban észlelt „látvány”, ami 1, ha arra fal van, 0, ha a rács abban az irányban járható. Az információ tehát globális abban az értelemben, hogy nem függ az állat orientációjától. A belső állapotról szóló információt arra korlátozzuk, hogy melyik irányban lép majd az adott pozícióba kerülését követően. Ezt értelmezhetjük az állat belső állapotáról szóló információként és az orientációról szóló információként is¹. Ez egy újabb négy elemű vektort jelent. A két vektor konkatenációja képezi a szenzoros input modelljét (6.1. ábra). Ezzel a modellel gyakorlatilag a 26. oldalon említett referenciakeret koncepciójából képezzük a nokortikális információ reprezentációját.

A labirintusban bejárando pályát véletlenszerűen választottam ki. A válogatásnál lényeges volt, hogy az záródjon, mert így az inputsorozat természetesen módon ismételhető beágyazás esetében is, ami az ICA számos implementációja esetében a statisztikai mintahalmaz növelésével segíti a független komponensek kialakulását. A másik előny, hogy beágyazás esetén az első néhány elemet nem kell elhagynunk, mert a történetet a hurok végéről egészíthetjük ki. A körpálya elérése érdekében az állat egy lépésnél azokat az irányokat preferálta, amelyek a labirintus centruma körül az óramutató járásával ellentétes irányú bejárásnak feleltek meg. Ha ilyen nem volt, más irányokat is kiválaszthatott. Természetesen a visszafelé lépés ekkor is csak akkor volt elfogadható, ha más irányban már nem tudott tovább jutni. Az állat több opció esetén random módon döntött. A kezdőpozíciót is véletlenszerűen határoztam meg. Utolsó kényszerként a teljes pálya nem lehetett kisebb, mint ... lépés. A generálás után a pályából a kisebb ismétlődő szakaszok kivágásra kerültek (ilyen előfordulhat például, ha az állat zsákutcába jut). A továbbiakban egy jól sikerült pályát fogok használni, amit a 6.2. ábra mutat. A pálya teljes hossza 62 lépés.

¹A szimulációk szerint ennek elhagyása lényegesen nem befolyásolja az eredményeket.



6.2. ÁBRA. A felhasznált labirintus és pálya. A startpozíciót kör mutatja.

6.2. Az alkalmazott módszerek

A szimuláció ezután a pályának megfelelő szenzoros információból előállította a beágyazott mintákat. Azután az inputot a már említett ok miatt megsokszoroztam, majd egy 0.5, 1, 0,5-ös szűrővel konvolúciót végeztem. Ezzel az információ pontossága valószerűbbé válik. Ezután következett az ICA analízis. Erre a FastICA algoritmusát alkalmaztam [Hyvärinen et al. 1997]. A FastICA nem adaptív, de gyors és stabil, hangolandó paraméterek nélkül működő implementáció. Képes arra, hogy automatikusan redukálja a kimenet dimenziószámát, ha a korrelációs mátrixból arra következtet, hogy az illető komponens nem hordoz információt. A FastICA mellett más ICA algoritmusok is kipróbálásra kerültek, lényegesen nem eltérő eredménnyel.

Mivel a negyedik fejezetben bemutatott modell az input deriváltfolyamatán működik, az eredményeket megvizsgáltam mind az így kapott független komponens aktivitásokon, mind azok deriváltján. Ez megtehető, mert az ICA egy lineáris transzformáció, ezért a különbségképzés elvégezhető a transzformáció után is.

Mint látni fogjuk, bizonyos szimulációknál az aktivitások szélsőségesen eltérő maximumot mutathatnak (6.3.2. rész). Ez nem feltétlen van így a biológiai esetben, mert az aktivitást számos folyamat normálhatja (ilyen mechanizmusok meglétéről bizonyítékaink vannak, lásd még 36. oldal). Ezért ha másképp nem említem, a továbbiakban olyan normálást alkalmaztam, ahol az összpopuláció együttes aktivitása adott időpillanatban egységnyi, azaz az aktivitások le vannak osztva az abban az időpillanatban érvényes összaktivitással.

Az ábrák elkészítésénél az aktivitások abszolútértékét tüntettem fel a jobb megjeleníthetőség végett.

6.3. Eredmények

6.3.1. Az ICA place sejteket alakít ki

A független komponens analízis outputjai a place sejtekhez hasonló aktivitást mutatnak (?? ábra). Ezt a szimuláció minden esetben megerősíti. Az aktív szakasz a teljes bejárt pálya körülbelül tizedét-huszadrészét fedi le. Többszörös rezonáló pozíciók is előfordultak. Ez a biológiai place feldek esetében is előfordul (17. oldal). A többszörös rezonanciapontok kialakulásának esélye a beágyazás dimenziójának növelésével csökkent (lásd 6.3.4. rész).

A lokalitás a deriváltfolyamat esetében megmaradt, bár némileg romlott (?? ábra). Általában is igaznak bizonyult, hogy deriváltfolyamatra kapott

eredmények nem különböznek lényegesen a direkt folyamatra érvényesektől, bár általában az is igaz volt, hogy bizonytalanabbak voltak. A továbbiakban csak a direkt folyamatról készült ábrákat mutatom be.

6.3.2. A kialakuló szűrők labirintusfüggők

Teszteltem a kialakuló szűrők labirintusfüggését: az egyik labirintusban kialakult szűrőket a másokban próbáltam alkalmazni. Az eredményeket a ?? ábra mutatja. *Az új labirintusban megszűnt az aktivitások lokalizáltsága, és megváltoztak az output statisztikai tulajdonságai is.* Ennek részletesebb vizsgálatát lásd a 6.3.5. részben. Ha egy szűrő ilyen outputot generál, az azt jelenti, hogy az nem specifikus arra az inputra, azaz nem ismeri fel (random szűrővel hasonló eredményt kapnánk). Normálás nélkül a rossz labirintusban való alkalmazás esetén az aktivitások nagyságrendje eggyel nagyobb lett. Biológiai esetben épp fordítva van, az összpopuláció aktivitása új információ esetén lecsökken (theta fázis). Normálás után, mivel a jó irányban való alkalmazásnál az összaktivitás nagy részét néhány kiemelt elem hordozza, éppen ez a helyzet.

6.3.3. A T-ICA szűrők irányfüggősége

Mint ahogy azt vártuk is, az ICA esetében (ahol a nem-adaptív algoritmus miatt az input beadásának sorrendje nem lényeges) a kialakuló aktivitások nem függenek a labirintus bejárési irányától (?? ábra). Azt is várjuk, hogy mivel a T-ICA időbeli sorozatokon tanul, nem mindegy számára a labirintus bejárési iránya. Valóban, a szimulációk szerint *a fordított irányú bejárást a T-ICA alatt kialakult szűrők ismeretlen labirintusnak tekintik (?? ábra).* Egy másik megközelítést mutat a ?? ábra. Itt a labirintus bejárásánál váltogattam az oda-vissza irányt, és ezen a sorozaton próbáltam a szűrőket kialakítani. A kapott szűrők a T-ICA esetében a visszaúton természetesen úgy viselkedtek, mintha egy mégegyszer olyan hosszú ismeretlen szakaszt toldottunk volna a bejárandó úthoz. Ha ilyen speciális módon tanítjuk a szűrőket, magyarázni tudjuk a szűrők *irányérzékenységét* is: az előbb csak azt láttuk be, hogy elérhető, hogy a szűrők csak egy irányt ismerjenek fel, így viszont azt is elérhetjük, hogy mindkét irányt felismerjék de különbözőképp kezeljék (?? ábra.) Hasonló ábrák biológiai esetben is készültek egyes sejtek irányérzékenységéről [Skaggs et al. 1996]. Mindenesetre ezt az analógiát fenntartással kell kezelnünk, mert csak ilyen speciális bejárás esetén érvényes.

6.3.4. A T-ICA aktivitások lokalizáltabbak

Az irányfüggőség ábráit elkészítettem különböző mélységű beágyazás mellett (?? ábra). A szimulációk azt mutatják, hogy *az aktivitás lokalitása határozottabb T-ICA esetében, és a beágyazás dimenziójának növelésével javul*. Ezt a mellékelt ábrákon jól látni. Az is látható, hogy az aktív területen kívül az aktivitás jelentősen lecsökken. Ezt az állapotter kiterjesztése miatt, mivel a véletlen jó illesztés esélye csökken, el is várjuk.

6.3.5. Az eloszlások vizsgálata

Különböző beágyazási mélység mellett közelebbről is megvizsgáltam az aktivitások különbségét abban az esetben, ha a szűrőket megfelelő irányban alkalmazom, illetve akkor, ha nem. A normált aktivitások hisztogramja látható a ?? ábra első és második grafikonján. Úgy tűnik, hogy a megfelelő szűrők alkalmazása exponenciális eloszlást, a nem az adott inputon kialakított szűrők alkalmazása pedig – mint az az új labirintus és a fordított irányú bejárás esetében történt – inkább normáeloszlást eredményez az aktivitásokban.

Ez jelentheti azt, hogy egy ilyen analízist alkalmazó rendszer az információ újszerűségét könnyen meg tudja állapítani az egységek aktivitásának eloszlásából. Ez egyúttal nem jelent egyszerű biológiai párhuzamokat is. Igaz, hogy az agy számos területén (például a hippocampusban is) a sejtek tüzelésének eloszlása besorolható vagy az egyik, vagy a másik kategóriába. A kutatók előszeretettel érvelnek az exponenciális eloszlás optimalitása mellett maximum-entrópia elvek alapján (mert rögzített átlagos tüzelésnél így lesz maximális az eloszlás entrópiája). Treves és munkatársai nemrégén végzett vizsgálataik alapján azonban azt állítják, hogy a inferotemporális kortexben a sejtek túlnyomó többsége egyik kategóriába sem sorolható be. Ma még nem világos tehát teljesen, hogy az eloszlások milyen szerepet játszanak az idegrendszer struktúráiban. Ha például a hippocampusban lenne olyan mechanizmus, ami képes az eloszlás alapján gátlás kifejtésére, akkor a 4. fejezetben ismertetett modell esetében lehetővé tenné a hippocampális output gátlását a reprezentáció kialakításának ideje alatt. Más agyi területek pedig el tudnák dönteni, hogy miről van rossz reprezentációjuk, és így a számukra új információkat a hippocampuszhoz irányíthatnák, hogy megfelelő reprezentációt alakíthassanak ki róla.

7. fejezet

Összefoglalás és kitekintés

Dolgozatomban a faktoriális kódolás hippokampuszban történő alkalmazását vizsgáltam. Szimulációkat futtattam annak kiderítésére, faktoriális kódolással sikerül-e reprodukálni a biológiai rendszer néhány jellemzőjét. Úgy találtam, hogy a faktoriális kódolás szenzoros bementen képes kognitív térkép kialakítására. A kognitív térkép a megfigyelthez hasonlóan lokális elemekből áll, és a környezet függvénye. A független komponens analízis problémáinak kiküszöbölésére az állapotteret kibővítettem egyedi szenzoros inputok helyett szenzoros inputok sorozatára. Ezzel az analízis a szenzoros információ időbeli tulajdonságaira is érzékenyebbé vált, továbbá a lokalizáció is javult, ami a kognitív térkép finomabb felbontású kialakítását és pontosabb használatát teszi lehetővé.

Felmerül a kérdés, van-e köze a szekvenciákon történő tanuláshoz az epizodikus memóriához. A harmadik fejezetben azt írtam, hogy az epizód fogalmának ma még nincsen általánosan elfogadott matematikai modellje. A szekvenciák kialakításának is megvannak a maga problémái. Először is, a javasolt biológiai implementáció csak nagyon kis időtartamok áthidalását teszi lehetővé. Másodszer nem állhat túlságosan sok elemből.

Számos kutató gondolja úgy, hogy az agy hierarchikus rendszerben dolgozza fel az információt. A látórendszer megerősíti ezt az elképzelést: a dolgozatban tárgyalt elsődleges látókérget követi a másodlagos, harmadlagos, stb. A kísérletek azt mutatják, hogy a magasabb hierarchikus szinteken a sejtek specificitása nő. Az inferotemporális kéregben, amit a látás ventrális áramának utolsó állomásának tartanak, majmok esetében például sikerült olyan sejteket kimutatni, amelyek objektumok átfogó geometriai tulajdonságaira érzékenyek, azaz például gömbhöz vagy kockához hasonlít-e leginkább az illető tárgy. Ez megerősíti a hierarchikus feldolgozási szintekről szóló elképzelést. Elképzelhető, hogy a különböző hierarchikus szinten végbemenő feldolgozás nagyon hasonlít egymásra, de legalábbis számos közös műveleten

osztoznak. Célszerűtlen lenne az ilyen általános elveket többször implementálni egy neurális rendszer esetében. Gazdaságosabbnak látszik egy közös struktúra bemenetére „átkapcsolni” az éppen aktuális feldolgozandó információt. A dolgozatban azt sugalltam, hogy ilyen közös stratégia a faktoriális kódolás, a megvalósító struktúra pedig a hippokampusz. Ez megoldhatja az ismertetett mechanizmus problémáit, mert világos, hogy szekvenciák sorozatából alkotott szekvenciák jóval nagyobb tartományokat képesek átívelni és méretük sincs annyira limitálva.

A szekvenciákon való tanulás bevezetésekor azzal érveltünk, hogy így csökkenteni tudjuk az input időbeli függőségeit. Tehát azzal, hogy az adott pillanatban rendelkezésre álló információba belefoglaltuk a megelőző történet egy kis részét, a folyamatot „markovibbá” tettük. Mint ismert, egy nem-determinisztikus folyamat Markov-tulajdonságán azt értjük, hogy az adott pillanatban érvényes állapotátmeneti valószínűségek csak az adott állapottól függenek, azaz azokat egy átmenetek valószínűségét tartalmazó mátrixban összefoglalhatjuk:

$$M_{ik} = p(x_{t+1} = k | x_t = i)$$

Az ilyen rendszerek kezelésére jó algoritmusaink vannak. Emellett, mint azt a negyedik fejezetben láttuk, Markov-folyamatok szekvenciáinak lejátszása jelentősen egyszerűbb feladat. Világos, hogy minden folyamat markovivá tehető beágyazással. A mélységet optimalizálhatjuk az alapján, hogy miközben a fenti mátrixot kialakítjuk, történik-e abban változás a mélység megnövelése után. Ha nem, ez azt jelenti, hogy további történet befoglalásával már nem tudunk mit hozzátenni az adott állapot leírásához; azt a befoglalt történet tökéletesen meghatározza. Ez így természetesen csak nagyon elméleti megoldás. Nyilván semmi sem garantálja, hogy az adott állapot kialakulása nem függ akár tetszőlegesen régi információktól is: ez különösen valószínű egy nemlineáris rendszerben. Azonban ha információs veszteséggel is számolunk, az újból a beágyazás alkalmazhatósága mellett szól. Természetesen a közvetlen megelőző inputok beágyazása nem biztos, hogy okos stratégia. A jobb válogatáshoz szükséges algoritmusok azonban még kidolgozásra várnak, és nyilván az inputra nézve specifikusak.

Egy másik érv a markovivá tétel mellett, hogy a Markov-folyamatok igen jól használhatók célorientált rendszerekben történő tanulásra. Egy ilyen elképzelés például a *mege erősítéses tanulás* (*reinforcement learning*). A 22. oldalon említett gyakorlás alapján történő navigáció például egy lehetséges bizonyíték arra, hogy ilyen célorientált rendszerek működnek az idegrendszerben.

Irodalomjegyzék

- [Amaral and Witter 1989] Amaral, D. G. and Witter, M. P. (1989) THE three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience* Vol. 31, No. 3 pp. 571-591.
- [Belouchrani et al. 1997] Belouchrani, A., Abed-Meraim, K., Cardoso J., Moulines E. (1997) A blind source separation technique using second order statistics. *IEEE transactions on signal processing*.
- [DeAngelis et al. 1995] DeAngelis, G. C., Ohzawa, I., Freeman, R. D. (1995) Receptive Field Dynamics in the central visual pathways. *Trends Neurosci.* 18, 451-458.
- [Field 1993] Field, D. J. (1993) *Wavelets, fractals, and fourier transforms*. Clarendon Press, Oxford.
- [Field 1994] Field, D. J. (1994) What is the goal of sensory coding? *Neural Computation*, 6:559-601.
- [Fomin et al. 1997] Fomin, T., Körmendy-Rácz, J., Lőrincz, A. (1997) Towards a unified model of cortical computation I.: Data compression and reconstruction architecture using dynamic feedback. *IDG*.
- [Freund and Buzsáki 1996] Freund, T. F. and Buzsáki, Gy. (1996) Interneurons of the hippocampus. *Hippocampus* 6:347-470.
- [Hateren and Ruderman 1998] van Hateren, J. H., Ruderman, D. L. (1998) Independent Component Analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proc. R. Soc. Lond. B* 265:2315-2320.
- [Hyvärinen et al. 1997] Hyvärinen, A. and Oja, E. (1997) A fast fixed-point algorithm for independent component analysis. *Neural Computation* 9:1483-1492.

- [Hyvärinen et al. 1999] Hyvärinen, A. and Hoyer, P. (1999) Topographic Independent Component Analysis. <http://www.cis.hut.fi/projects/ica/>
- [Karhunen et al. 1996] Karhunen, J., Oja, E., Wang, L., Vigário, R., Joutsensalo, J. (1996) A class of neural networks for Independent Component Analysis. *IEEE Transactions on Neural Networks*.
- [Lőrincz and Buzsáki 2000] Lőrincz, A. and Buzsáki, Gy. (2000) Two-phase computational model training long-term memories in the entorhinal-hippocampal region.
- [Meila 1999] Meila-Predovicu, M. (1999) Learning with mixture of trees. PhD thesis.
- [MI 1999] (1999) Mesterséges intelligencia, 11.3. fejezet. Aula kiadó.
- [Olshausen and Field 1996] Olshausen, B. A. and Field, D. J. (1996) Sparse coding with an overcomplete basis set: A strategy employed by V1?
- [Redish 1997] Redish, A. David (1997) Beyond the cognitive map: Contributions to a computational neuroscience theory of rodent navigation. PhD thesis.
- [Skaggs et al. 1996] Skaggs, W. E., McNaughton, B. L., Wilson, M. A., Barnes, C. A. (1996) Theta phase precession in hippocampal neuronal populations and the compression of temporal sequences. *Hippocampus* 6:149-172.