

A MESTERSÉGES INTELLIGENCIÁHOZ VEZETŐ ÚT – A 2024-ES FIZIKAI NOBEL-DÍJ

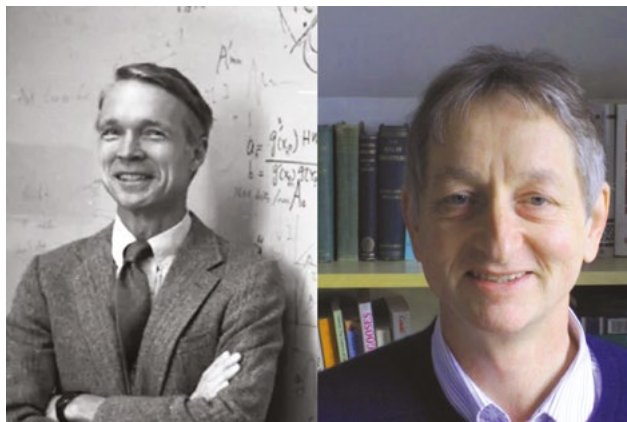
Somfai Ellák^{1,2}

¹Eötvös Loránd Tudományegyetem, Mesterséges Intelligencia Tanszék, Budapest

²HUN-REN Wigner Fizikai Kutatóközpont, Szilárdtestfizikai és Optikai Intézet, Budapest

E-mail: somfai.ellak@wigner.hun-ren.hu

A fizikának tekinthető tudományterületet kiterjesztő módon értelmezi az a vélekedés, hogy a fizika az, amit fizikusok csinálnak (Fermi szerint késő este). A 2024-es fizikai Nobel-díjat, amelyet John J. Hopfield és Geoffrey E. Hinton a mesterséges neuronhálóok kidolgozásához szükséges alapvető felfedezésekért kapott (1. ábra), mégsem így kell érteni. A kezdeti lépések alapvetően kapcsolódtak mágneses rendszerekhez és ezek statisztikus fizikai modelljeihez – mint például a számos más komplex rendszer mintájának is tekinthető spinüvegekhez – amelyért Giorgio Parisi kapta részben a 2021-es fizikai Nobel-díjat.



1. ábra. John J. Hopfield és Geoffrey E. Hinton

Az embereket régóta foglalkoztatja az a kihívás, hogy „gondolkodó” gépeket készítsenek. Ennek egy korai megnyilvánulása volt a 18. században Kempelen Farkas sakkozógépe, a Török, amelyről csak utólag derült ki egyértelműen, hogy voltaképpen egy ember ült benne. Érdekes, hogy az emberi intelligencia rejtett felhasználása napjainkig sem tűnt el. Például az Amazon nemrég bezárt Just Walk Out (csak sétálj ki) bolti alkalmazottak nélküli, bekamerázott, a cég állítása szerint kizárólag mesterséges intelligencián alapuló üzleteiről is kiderült, hogy a tranzakciók több mint felét egy indiai call center munkatársainak is végig kellett néznie a megfelelő megbízhatóság eléréséhez.

Fontos mérföldkő volt az 1940-es években az első elektronikus számítógép megjelenése, amely jóval gyors-

sabban tudott sok jegyű számokat összeadni és szorozni, mint a legügyesebb fejszámológépművész. A számítógépek és később a robotok rohamos fejlődésével reményteljessé vált, hogy nagy mennyiségű szellemi munka kiváltható legyen. Az 1956 nyarán tartott, sokak által a mesterséges intelligencia megszületésének tartott dartmouthi workshop már olyan célokat tűzött ki, mint a gépi fordítás és az emberi szintű intelligencia megközelítése. A kezdeti optimizmus túlzónak bizonyult, a korai gyors fejlődés hamar megtorpant szinte valamennyi fronton. Ennek okát abban kereshetjük, hogy sok olyan, az ember számára egyszerűnek tűnő feladatra, mint például egy képen látható tárgy felismerése, vagy egy párbeszéd lefolytatása, kezelhetetlenül bonyolult egy működő algoritmus megírása.

Mivel az ilyen jellegű, sokszor mintázatfelismerésnek tekinthető feladatokat nemcsak az ember, de sok esetben a fejlettebb állatok is könnyedén megoldanak, érdekes a központi idegrendszert vizsgálni, hátha a biológiai rendszerek működési mechanizmusa gépi környezetbe is átültethető. Warren McCulloch és Walter Pitts az 1940-es években javasolt egy, az idegsejtek (neuronok) működését leíró egyszerűsített „perceptron” modellt, amely a sejtek állapotát binárisan (aktív vagy inaktív) közelíti. A következő időpillanatban a neuron (szintén bináris) állapotát a bejövő jelek súlyozott összege határozza meg. A tanulási folyamat a súlyok beállítását jelenti, amelyhez biológiai motivációt Donald O. Hebb szolgáltatott a javasolt biológiai tanulási mechanizmussal, miszerint az idegsejtek közötti szinaptikus kapcsolat erősödik, ha ezek tipikusan egyszerre aktívak.

Az 1950-es évek végén Frank Rosenblatt egy működő három rétegű ún. előrecsatolt hálózatot épített, amelyben a következő réteg bemenetei az előző réteg kimeneteiből kerülnek ki; nincs visszacsatolás. A berendezés a 20×20-as bemeneti fotodetektorréteg révén képi jel feldolgozására volt hivatott. A kezdeti lelkesedést lehűtötte az a tény, hogy egy perceptron csak lineárisan szeparábilis bemenetet képes megkülönböztetni: amikor a bemeneti sokdimenziós térben található olyan hipersík, amely szétválasztja a 0 és 1 kimenethez tartozó pontokat. Ennek következtében a perceptron például a „kizáró vagy” logikai műveletet sem tudja reprezentálni. Ennek

a súlyos kötöttségnek a felismerése gyakorlatilag megszüntette a mesterséges neuronháló kutatását a 70-es években – annak ellenére, hogy a perceptronokból álló *hálózatok* esetén már nincs ilyen megkötés; ezek tanítása viszont abban az időben még nem volt megoldott.

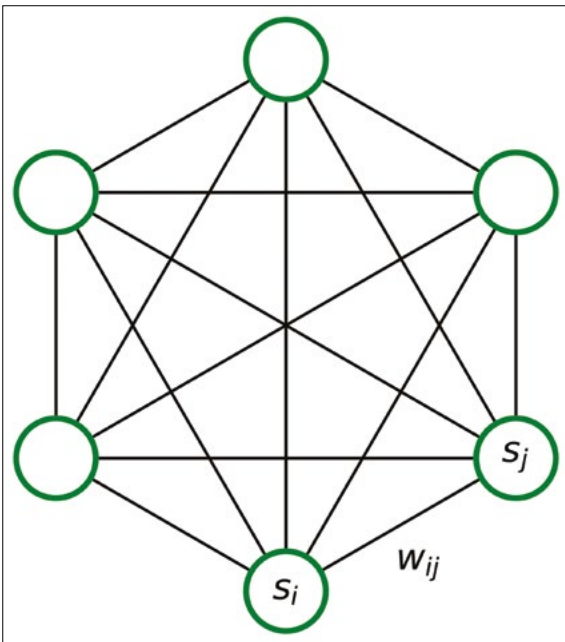
A következő jelentős előrelépés, az 1980-as években éppen John Hopfield nevéhez kötődik. Hopfield egy teljesen összekötött (teljes gráfot alkotó) hálózatot javasolt a következő dinamikával (2. ábra):

$$h_i = \sum_{j \neq i} w_{ij} s_j \quad \text{és} \quad s_i^{\text{új}} = \begin{cases} 1, & \text{ha } h_i > 0 \\ 0, & \text{ha } h_i \leq 0 \end{cases},$$

ahol az egyes neuronok állapotának frissítése független random folyamat. Ha w_{ij} , amely pozitív és negatív értékeket is felvehet, szimmetrikus, akkor az energiaként definiálható

$$E = -\frac{1}{2} \sum_{i \neq j} w_{ij} s_i s_j$$

mennyiség monoton csökken, és a rendszer egy stabil állapotba konvergál. A stabil állapotok egy-egy memóriaképeknek tekinthetők, így a rendszer egy tartalomcímzett memóriát modellez: egy, a memóriaképhez közel álló kiindulópontból konvergál ahhoz. A modell szoros kapcsolatban áll a mágneses rendszerekkel, mint például a spinűvegekkel; így például a memória kapacitása is analitikusan számolható a spinűvegek elméletének felhasználásával.



2. ábra. Hopfield-háló. A zöld neuronok felelnek meg a memóriaegységeknek (bemenet és kimenet is)

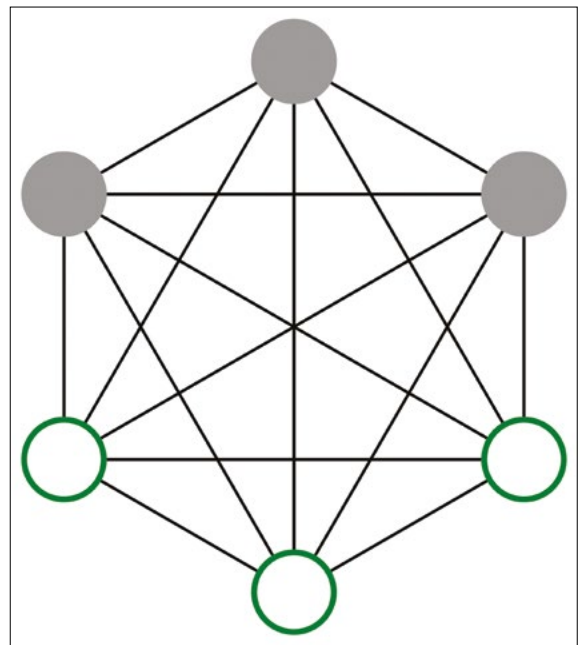
Mint ahogy hasonló statisztikus fizikai rendszerek esetében is tapasztalható, a modell kollektív viselkedése robusztus a részletek bizonyos fokú változtatására. Hopfield kiterjesztette a modellt folytonos értékű állapotokra, késleltetett, időben folytonos dinamikával, amelyben

a Heaviside-függvény szerepét egy folytonos nem lineáris függvény veszi át; mindez nem változtatott a modell kollektív viselkedésén.

Hopfield modellje stimulálóan hatott a mesterséges neuronháló további kutatására. Geoffrey Hinton és Terrence Sejnowski egy „Boltzmann-gépnek” nevezett sztochasztikus változatot javasoltak, amelyben a háló minden $\mathbf{s} = (s_1, s_2, \dots, s_N)$ állapotához rendelt valószínűséget a Boltzmann-eloszlás adja:

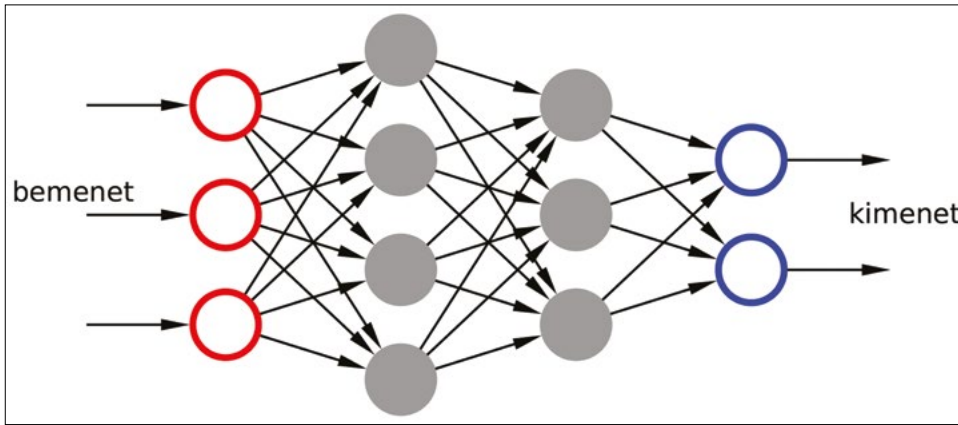
$$\mathbb{P}(\mathbf{s}) \sim e^{E/T}, \quad E = \frac{1}{2} \sum_{i \neq j} w_{ij} s_i s_j - \sum_i \theta_i s_i,$$

ahol T egy hőmérséklet-jellegű parameter, θ_i pedig egy lokális térnek felel meg. A w_{ij} csatolási, valamint a θ_i lokálistér-paraméterek nehézkesen ugyan, de beállíthatók („taníthatók”) arra, hogy a rendszer valószínűség-eloszlása minél közelebb legyen egy tanító halmazéhoz. A memóriakapacitást úgy növelték, hogy rejtett neuronokat adtak a rendszerhez: ezek hozzájárulnak az energiához, de figyelmen kívül hagyandók a megtanulandó minták azonosításában. Ez volt az egyik első olyan neuronháló, amely ún. rejtett vagy belső reprezentációt tudott tanítani olyan neuronok felhasználásával, amelyek sem a bemenetben, sem a kimenetben nem vettek részt (3. ábra).



3. ábra. Boltzmann-gép. A zöld neuronok felelnek meg az írható és olvasható memóriaegységeknek, a szürke neuronok rejtettek: a belső reprezentációért felelősek

A visszacsatolásokat is tartalmazó, sűrűgráf-struktúrájú háló, mint a Hopfield-háló és a Boltzmann-gép, az asszociatív memóriát modellezzik. A praktikus felhasználások szempontjából ezeknél jóval lényegesebbek azok az előrecsatolt (angolul feed-forward) mesterséges neuronháló, amelyek csupán egyirányú csatolásokat tartalmaznak, mint a már említett Rosenblatt-féle három rétegű perceptron. Ezen háló egy függvényt valószínűsítanak



4. ábra. Előrecsatolt neuronháló. Piros a bemeneti, a kék a kimeneti, a szürke pedig a belső reprezentációért felelős rejtett neuronokat jelöli. A kapcsolatok egyirányúak, az információ terjedésében nincs visszacsatolás

meg: a bemeneti rétegbe vezetett értékek meghatározzák a kimeneti réteg eredményeit. Ez az architektúra (lásd 4. ábra) alkalmas a felügyelt gépi tanulás megvalósítására, amelyben a rendszer egy bemenet-kimenet párokból álló tanító halmaz alapján keresi azt a függvényt, amely a legkisebb hibával számolja ki a kimeneti értékeket. A cél nem is a tanító halmaz kimeneti értékeinek minél pontosabb rekonstrukciója, hanem az, hogy a tanító halmazban nem szereplő bemeneti értékekre is a megfelelő kimenetet számolja. Például egy arcfelismerő rendszerben a fénykép pixeleinek fényintenzitása adja a rengeteg bemeneti elemet, kimenetként pedig annyi neuront alkalmazunk, ahány embert szeretnénk felismerni; a legerősebben tüzelő neuron választja ki a választ. Ez gyakorlatilag egy nagyon sok dimenziós térben értelmezett, vektorértékű függvény illesztését (regresszióját) jelenti.

A tanítás („függvényillesztés”) azt jelenti, hogy azt a hibafüggvényt (angolul loss) minimalizáljuk a háló paramétereinek optimális beállításával, amely a tanító halmaz (x_n, y_n) bemenet-kimenet párojaira összegzi a háló által generált $f_w(x_n)$ eredmény és az elvárt y_n kimeneti érték közötti távolságot:

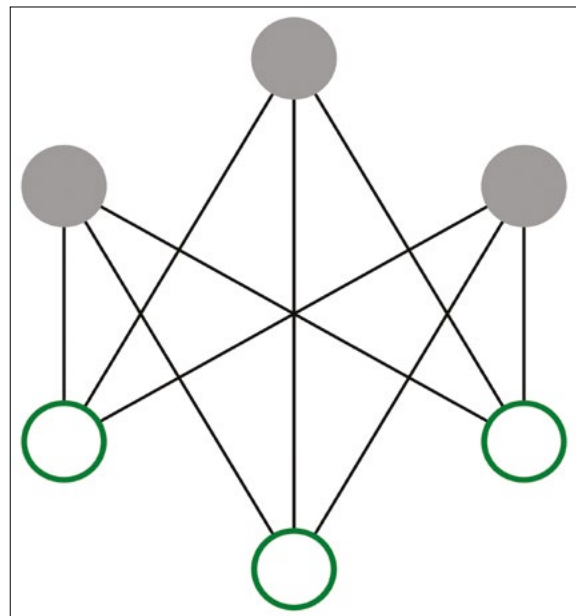
$$L(\mathbf{w}) = \sum_{(x_n, y_n) \in \text{tanító halmaz}} d(f_w(x_n), y_n),$$

ahol $d(\cdot, \cdot)$ a feladathoz legalkalmasabb távolságfüggvény, a \mathbf{w} vektor pedig a háló paramétereit jelöli; ezek száma a jelenlegi legnagyobb modelleknél a 10^{12} -t meghaladhatja.

A hibafüggvény minimalizálása a paraméterek sokdimenziós terében nem könnyű feladat. Egy lehetséges megközelítés a „hegymászó” (angolul gradient descent) algoritmus valamilyen változata, amelyben az L függvény \mathbf{w} szerinti deriváltja (gradiens) alapján választjuk ki a leggyorsabb csökkenés irányát, és teszünk egy kis lépést. A lépésközt, a „learning rate”-et heurisztikusan állítjuk be. Ilyen lépések egymásutánja egy lokális minimumhoz vezet el. A differenciálhatósághoz természetesen az f függvénynek folytonosnak kell lennie, így diszkrét jellegű problémák esetén is folytonos értékkel kell közelíteni a megoldást. A megoldáshoz szükséges

az L függvény \mathbf{w} szerinti gradiensenek pontos, lehetőség szerint analitikus kiszámolása, ami nem triviális feladat.

Ebben az irányban jelentős előrelépés történt az 1980-as évek közepén, amikor David Rumelhardt, Hinton és Ronald Williams megmutatta, hogy hogyan lehet egy többrétegű előrecsatolt neuronhálót tanítani. Az L függvény gradiense a *backpropagation*-nak nevezett módszerrel számolták ki, amely a láncszabály következetes alkalmazása a sokszorosan összetett f_w függvény esetére. A backpropagationhoz hasonló módszereket mások is használtak korábban, viszont Hintonékhoz köthető az az áttörés, amely a többrétegű előrecsatolt neuronhálók tanítását tette lehetővé. A többrétegű, vagyis rejtett réteget is tartalmazó (tehát belső reprezentációt alkotó) neuronhálókról így demonstrálták, hogy az egyszerű perceptronnal ellentétben lineárisan nem szeparábilis problémákat is meg tudnak oldani.



5. ábra. Korlátozott Boltzmann-gép. Csak a különböző típusú neuronok között (tehát a zöld bemeneti-kimeneti memóriaegységek, és a szürke rejtett neuronok között) van csatolás

Ez megnyitotta az utat olyan mesterséges neuronháló kidolgozása felé, amelyek hamarosan már hasznos feladatokat végeztek el. Yann LeCun nevéhez fűződik a konvolúciós neuronháló (angol rövidítése CNN) megalakítása; itt az a kulcsgondolat, hogy olyan térben strukturált bemenetet, mint például egy kép pixeleit, eltolásinvariáns módon érdemes feldolgozni. Ez nemcsak nagyban csökkenti a tanítandó paraméterek számát, és ezáltal gyorsítja a tanítás folyamatát, hanem robusztusabbá is teszi a feldolgozást. Yoshua Bengio a konvolúciós neuronhálókat összekapcsolta sztochasztikus rendszerekkel (mint például rejtett Markov-modellekkel). Ez lehetővé tette, hogy az 1990-es években amerikai bankok automatizálva olvassák a csekkekre írt kéziratos számokat. Ezekért, valamint a későbbiekben elért további eredményekért kapta Hinton, LeCun és Bengio a 2018-as (sokak által informatikai Nobel-díjnak tartott) Turing-díjat.

A gyakorlatban is elért sikerek ellenére a fejlődés a 90-es évek végére újra megtorpant, ugyanis a backpropagation exponenciálisan lassúnak bizonyult az összetettebb problémák megoldásához szükséges, az eddigi néhánynál jóval több belső réteget tartalmazó, ún. mélyháló tanítására. A megoldáshoz Hinton a Boltzmann-géphez nyúlt vissza, pontosabban ennek egy korlátozott változatához, amelyben csak a különböző típusú (tehát a látható és a rejtett) neuronok között engedünk meg csatolást (5. ábra). A korlátozott Boltzmann-gép tanítása az eredeti modellel ellentétben sokkal egyszerűbbnek és gyorsabbnak bizonyult. Sikertült kidolgozni egy olyan módszert, amelyben a mélyháló rétegeit egyenként előtanítják a korlátozott Boltzmann-gép segítségével; utána az így előtanított rétegeket a backpropagation már hatékonyan továbbtanítja az adott feladatra. Ez az áttörés alapozta meg a kétezres évek közepén a hatékony mélyháló széles körű felhasználását.

A mesterséges neuronháló robbanásszerű fejlődését sokan egy 2012-es eseményhez kötik. Az informatikában a kutatást és az algoritmusok tökéletesítését sok esetben nyilvános versenyek katalizálják. Ilyen volt a 2010-ben induló, évente megrendezett „ImageNet challenge”, amelyben azt kellett eldönteni, hogy egyes fényképek mit ábrázolnak. A milliós nagyságrendű tanító halmaz fényképei ezer különböző kategóriába sorolt objektumot ábrázolnak, és az ez alapján betanított algoritmusok egy százezres teszhalmazon mérték össze a tudásukat. Az első két év lassú fejlődése után 2012-ben ugrásszerű előrelépés következett: Hinton és két PhD-hallgatója, Alex Krizhevsky és Ilya Sutskever egy konvolúciós mélyhálával úgy nyerték meg a versenyt, hogy a hibarátajuk csak közel fele volt a második helyezettének. Két évvel később már csak konvolúciós mélyhálók alapuló algoritmusok maradtak versenyben, minden más módszer kiszorult.

A mesterséges neuronháló napjainkra már nemcsak a tudományban (mint például az idén kémiai Nobel-díjjal méltányolt fehérjeszerkezet-előrejelzések terén), hanem a mindennapi életben is sok helyen felbukkannak. A technológia fejlődését az utóbbi másfél évtizedben alapvetően a gépi tanuláshoz szükséges adatok mennyisége (mint az ImageNet is, vagy a most már 10^{10} szóból álló korpuszok) és a Moore-törvény szerint exponenciálisan növekvő számítási kapacitás határozta meg. Mindez azonban nem következett volna be azok nélkül a statisztikus fizikai rendszerekre épülő modellek nélkül, amelyek kidolgozásában Hopfieldnek és Hintonnak meghatározó szerepe volt. A fizika azonban nemcsak forrása, hanem aktív felhasználója is a mesterséges neuronhálónak; erről többet a *Fizikai Szemle* következő számában olvashatunk.