

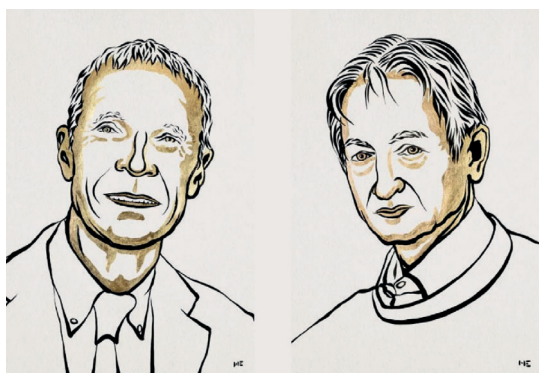


Kégl Tamás

■ PTE TTK Kémiai Intézet

Fizikai Nobel-díj: főhajtás a gépi tanulás és a mesterséges intelligencia úttörői előtt

Nyugodtan mondhatjuk, hogy az elmúlt két évben mind a tudományos, mind a hétköznapi közbeszédben kitüntetett szerepet kapott a mesterséges intelligencia (MI). Sokak számára talán nem meglepő, hogy a 2024-es fizikai Nobel-díjat *John J. Hopfield* és *Geoffrey E. Hinton* kapták a mesterséges neurális hálózatokkal történő gépi tanulás terén végzett úttörő munkájukért (1. ábra). Ezek a felfedezések alapvetően formálták át a mesterséges intelligencia és a gépi tanulás világát, jelentős hatást gyakorolva a tudományos kutatásokra és a mindennapi életre egyaránt.



1. ábra. John J. Hopfield és Geoffrey E. Hinton
(Niklas Elmehed grafikája, © Nobel Prize Outreach)

John Joseph Hopfield (1933. július 15-én született Chicagóban) az amerikai tudományos élet egyik meghatározó, interdiszciplináris alakja, akinek munkássága különösen nagy hatást gyakorolt az idegtudományra, a fizikusok és a számítástudósok által művelt neurális hálózatok elméleti alapjaira, valamint a biológiai rendszerek információfeldolgozási mechanizmusainak megértésére. Fizikusként kezdte pályáját, később azonban a komplex rendszerek és a kognitív folyamatok modellezése felé fordult, ezzel hidat képezve a klasszikus fizika, a kémia, a biológia és a mesterséges intelligencia között. Leginkább a róla elnevezett Hopfield-háló miatt vált ismertté, amely az idegrendszer mintázatfelismerő és memória-funkcióit leíró neurális hálózatok egyik alapvető modellje.

John Hopfield egy olyan korszakban született, amikor a tudomány számos területe – beleértve a kvantummechanikát, a szilárdtestfizikát és a korai számítógép-tudományt – óriási fejlődésen ment keresztül. Magas szintű képzést kapott; alapfokú diplomáját 1954-ben szerezte a Swarthmore College-ban, doktori fokozatát pedig 1958-ban védte meg a Cornell Egyetemen. Fizikai képzettsége révén a későbbiekben szilárdtestfizikai és kvantumkémiai kutatá-

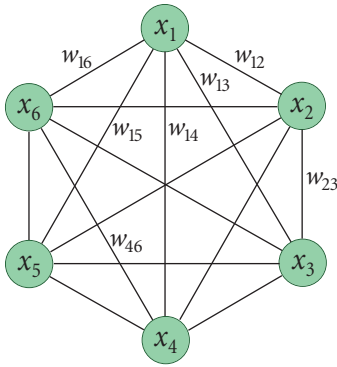
sokba kezdett, amelyek jó alapot teremtettek számára ahhoz, hogy megértse a komplex rendszerek mikroszkopikus kölcsönhatásainak makroszkopikus viselkedésben megjelenő mintázatait.

Pályájának kezdeti szakaszában Hopfield szilárdtestfizikával és molekuláris rendszerekkel foglalkozott. Az 1960-as és 1970-es években kutatásai olyan területekre irányultak, mint a polimerek elektronikus tulajdonságai, a szilárdtestfizika különböző jelenségeinek elméleti megértése, valamint a folyadékok és szilárd anyagok molekuláris mozgásának dinamikája. Ezek a munkák megalapozták azt a szemléletmódot, amely a komplex, sok részecskéből álló rendszerek leírására szolgáló statisztikus fizikai módszereket helyezte előtérbe.

A statisztikus fizika eszköztára, mely a sokrészecske-rendszerek viselkedésének makroszkopikus leírására szolgál, később kulcsfontosságúvá vált Hopfield neurális hálókkal kapcsolatos kutatásaiban is. A neurális hálózatok megértéséhez ugyanis hasonlóan elvont, probabilisztikus eszközök szükségesek: a neuronok nagy száma és a szinaptikus kapcsolatok hálózatos felépítése indokoltá teheti az ilyen módszertani megközelítéseket.

Hopfield 1982-ben publikált, mára már klasszikusnak számító tanulmánya [1] forradalmasította a mesterséges neurális hálózatokról alkotott elképzeléseket. Ebben a munkában a szerző bemutatott egy egyszerű, mégis mélyreható modellt, amelyet ma Hopfield-hálónak ismerünk. A Hopfield-háló visszacsatolt, teljesen összekapcsolt neurális hálózat: minden neuron hatással van a többi neuronra. A háló állapota a neuronok aktivációjának mintázatából áll össze. A háló dinamikája úgy van megalkotva, hogy időben az energiafüggvény által leírható potenciálfelületen egyre alacsonyabb energiaállapotok felé mozduljon el. Így a háló stabil állapotai (lokális minimumai) a memorizált mintázatoknak felelnek meg. Ennek eredményeképpen a Hopfield-háló képes zajos vagy hiányos bemenetek esetén is felismerni és kiegészíteni a tárolt mintákat, így egyfajta asszociatív memória modelljévé vált.

A Hopfield-hálózat elsősorban abban különbözik a ma ismertebb mély neurális hálózatoktól, hogy gyakran bináris neuron-csomópontokkal dolgozik, amelyek vagy „aktívak” (1) vagy „inaktívak” (0). Az ilyen hálózatok többnyire teljesen összekapcsolt struktúrával rendelkeznek, vagyis minden neuron kapcsolatban áll az összes többi neuronnal (2. ábra). A kapcsolatokat szinaptikus súlyok jellemzik, melyek meghatározzák, hogy az egyik neuron aktivitása hogyan befolyásolja egy másik neuron állapotát. Ezeket a súlyokat a hálózat a betanítási fázis során „tanulja meg”, rendszerint egy meghatározott tanulási szabállyal. A tanulás után a hálózat úgy működik, hogy ha valamelyik neuron módosítja az álla-



2. ábra. A Hopfield-hálózat felépítése

$$\text{Energia: } E = - \sum_{i < j} x_i x_j w_{ij}$$

$$\text{Valószínűség: } \frac{e^{-E}}{\sum e^{-E}}$$

potát, ez befolyásolja a többi neuron állapotát is, egészen addig, amíg a rendszer el nem éri valamelyik stabil állapotát.

A rendszer stabil állapotai „vonzerópontoknak” vagy fixpontoknak is nevezhetők, amelyek általában a betanított mintákat képviselik. Ez azt jelenti, hogy ha a hálózat bármelyik neuronját egy bizonyos irányban elmozdítjuk (például zajos bemenetet adunk neki), idővel a hálózat visszaáll a legközelebbi stabil állapotba, azaz felidézti a számára „leginkább hasonló” mintát. Ennek az az eredménye, hogy képes egy-egy részleges vagy zajjal terhelt bemenetből rekonstruálni a teljes, eredetileg betanított mintát. Emiatt a Hopfield-hálózatokat gyakran alkalmazzák auto-asszociatív memóriarendszerekként, ahol a cél az, hogy megbízhatóan tárolják és előhívják az információt még torzított vagy hiányos adatok esetén is.

Az energia fizikai és fizikai kémiai analógiák alapján történő bevezetése új és igen hasznos koncepciót jelentett: a háló olyan dinamikus rendszer, amely mindig az alacsonyabb energiaállapotok felé halad. Ez hasonlít a természetes rendszerek viselkedéséhez, ahol a stabil állapotok jellemzően az alacsony energiaállapotoknak felelnek meg. A Hopfield-hálóban minden konfigurációnak (tehát minden lehetséges neuronállapot-kombinációnak) van egy hozzá rendelhető energiája, és a rendszer beépített szabályai alapján a neuronok állapotát leíró hiperparaméterek a minél alacsonyabb energiájú állapot elérésére törekednek.

A Born–Oppenheimer-közelítésből levezethető potenciális-energia-felületekkel analóg „energiatájak” elképzelése segít megérteni, miért tér vissza a háló egy-egy betanított mintához: egyszerűen a megjegyzett mintázatok olyan „energiagödrökként” viselkednek, amelyekbe, ha a háló állapota belekerül (még ha nem is pontosan a közepébe, csak a közelébe), akkor szépen „lecsorog” az energia-lejtőn az ismerős mintázat stabil pontjába.

A Hopfield-háló sikerét követően John Hopfield kutatásai továbbra is a biológiai információfeldolgozás kérdéseire koncentráltak, de fokozatosan bővült a paletta – többek között a biológiai rendszerek időbeli mintázatainak feldolgozásával, a neurotranszmitterek szerepével, valamint a genetikai szabályozás mögött rejlő információelmélettel. Hopfield munkáiban gyakran jelent meg az az alapgondolat, hogy a biológiai rendszerek – az agy, a genetikai hálózatok, a sejtjelátviteli mechanizmusok – információs problémákat oldanak meg, és ezek a megoldások hatékony, energia-optimalizált stratégiákon alapulnak [2]. Hopfield kutatásai már az 1960-as évektől kezdve kiterjedtek a számítógépes és elméleti neurális hálózatok mellett a valódi idegrendszeri folyamatokra is, gyakran együttműködésben biológusokkal, psi-

chológusokkal és más tudományterületek képviselőivel. Emellett jelentős hatást gyakoroltak a mesterséges intelligencia és a gépi tanulás közösségére is, hiszen a Hopfield-halók a mély tanulási hálózatok előfutárjainak tekinthetők.

Geoffrey Everest Hinton, aki 1947. december 6-án született Wimbledonban, a mesterséges intelligencia (MI) és a gépi tanulás egyik legnagyobb hatású alakja. Sokan „a deep learning (mélytanulás) keresztapjának” nevezik, mivel kulcsszerepet játszott olyan elméleti és gyakorlati módszerek kidolgozásában, amelyek a neurális hálózatok modern reneszánszát idézték elő. Életútja során olyan módszerek kifejlesztéséhez is hozzájárult, amelyek mind a mai napig megalapozzák az önvezető autók, a képfelismerés, a beszéd felismerés, a természetes nyelvfeldolgozás és számos más, mesterséges intelligenciával kapcsolatos technológia hatékony működését.

Hinton intellektuális érdeklődése nem előzmények nélküli. Értelmiségi családban nőtt fel, felmenői között az akadémiai világ jeles képviselői is akadnak, például Sir Geoffrey Taylor fizikus, aki Hinton anyai dédapja volt. Az efféle háttér nagyban hozzájárult ahhoz, hogy már fiatalon érdeklődést mutasson a tudományos kérdések iránt. Eleinte még nem a mesterséges intelligencia kötötte le a figyelmét, hanem inkább a kognitív folyamatok megértése és a tudat természete vonzotta, ám ezek a kérdések hamar a mesterséges intelligencia felé tereltek.

Hinton a Cambridge-i Egyetemen szerezte első diplomáját (BA) 1970-ben, kísérleti pszichológia szakon. Már ekkor felvetődött benne az idegrendszer számítógépes modellezésének gondolata. A mesterséges intelligencia iránti érdeklődése aztán az Edinburgh-i Egyetemen bontakozott ki, ahol 1978-ban doktorált mesterséges intelligencia témakörben, Christopher Longuet-Higgins vezetésével. Doktori munkája során a kognitív tudomány, a pszichológia és a számítástudomány találkozásánál mozgott, és mélyen foglalkoztatta, hogy miként lehetne számítógépes rendszerekbe beépíteni az emberi agyhoz hasonló tanulási képességeket.

Az 1970-es évek végén és az 1980-as évek elején a mesterséges intelligencia világa még nem a mai értelemben vett neurális hálózatokról szólt, noha a legegyszerűbb neurálháló (egy réteg neuron, aktiváló függvény nélkül) több mint 200 éves múltra vezethető vissza, mivel gyakorlatilag megegyezik a Legendre és Gauss által alkalmazott lineáris regressziós módszerrel. A legtöbb kutató a szimbolikus MI-t (symbolic AI) részesítette előnyben, ahol a tudást logikai szabályok és ismeretbázisú rendszerek alapján kódolták. Hinton azonban úgy érezte, hogy az agy működése nem redukálható egyszerű szabályrendszerekre; ehelyett elosztott reprezentációkra és súlyozott kapcsolatokra épített. Ez a meggyőződés vezette el a mesterséges neurális hálózatok területére.

Doktori címének megvédését követően Hinton dolgozott a Sussexi Egyetemen, a San Diegó-i Kaliforniai Egyetemen, majd a Carnegie Mellon Egyetemen (CMU) is. A CMU-n töltött évei során találkozott David E. Rumelharttal és Ronald J. Williamsszel, akikkel közösen kidolgozta a visszatérjesztéses hibaminimalizálás (backpropagation) módszerét, amely lehetővé teszi a több rétegből álló neurális hálózatok hatékony tanítását [3].

A backpropagation algoritmus fontosságát ma már nehéz túlbecsülni: ez az eljárás kulcsfontosságú volt annak megértésében, hogy miként taníthatunk többretegű, úgynevezett „deep” (mély) hálózatokat. A módszer megalkotása és ismertté válása ugyan még nem indította be a mélytanulás forradalmát – az akkori számítási kapacitások és adatmennyiségek ehhez nem voltak elegendők –, de elvi alapokat szolgáltatott ahhoz, hogy később, a 2000-es években, amikor már rendelkezésre állt elegendő feldol-



gozsi teljesítmény és digitális adat, a mélytanulás (deep learning) robbanásszerűen fejlődhesen.

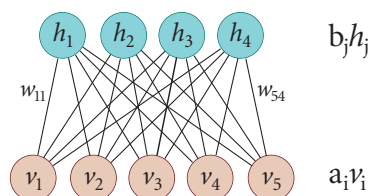
Az 1980-as évek végén Hinton Kanadába költözött, ahol a Torontói Egyetemen kapott professzori állást. Ez a váltás jelentős hatással volt munkásságára, ugyanis a Torontói Egyetemen rendkívül élénk kutatói közösség alakult ki, amely támogatta a neurális hálózatokat előnyben részesítő szemléletet. Kanadában Hinton olyan közegebe talált, ahol a tudományos intézmények nyitottak voltak az interdiszciplináris megközelítésekre és az alternatív MI-irányzatok kevésbé vonták el a neurális hálózatok iránt fogékony kutatók érdeklődését.

Hinton kutatásai során felmerült az a kérdés is, hogy miként képesek a neurális hálózatok belső, rejtett reprezentációkat kialakítani. E téren fontos előrelépést jelentett a Boltzmann-gépek és a korlátozott Boltzmann-gépek (Restricted Boltzmann Machines, RBM) bevezetése. Az RBM-ek segítségével Hinton és munkatársai olyan tanulási módszereket dolgoztak ki, amelyek képesek voltak a bemeneti adatok statisztikai szerkezetének felfedezésére és belső, rejtett egységek által kódolt mintázatok kialakítására.

A Boltzmann-gépek (Boltzmann Machines, BM) a mesterséges neurális hálózatok egy különleges fajtáját jelentik, amelyeket eredetileg Geoffrey Hinton és Terrence Sejnowski az 1980-as években dolgoztak ki [4]. Ezek a hálózatok elsősorban statisztikai fizikai elvekre épülnek, és képesek bonyolult mintázatok megtanulására és reprezentálására. Fő különlegességük az, hogy a hagyományos irányított hálózatoktól eltérően a Boltzmann-gépekben nincsenek kijelölt bemeneti vagy kimeneti rétegek, valamint a kapcsolatok kölcsönösek. Ez azt jelenti, hogy a háló minden neuronja képes hatni a többire, és az információ nem egy irányban (pl. csak előre) terjed, hanem körkörösén. Ez lehetővé teszi, hogy a háló komplex mintázatokot sajátítson el és ezeket a belső állapotaiban tárolja.

A Boltzmann-gépek tanítása nem egyszerű feladat, mert a teljes kapcsolati hálózat miatt sokféle állapotot kell figyelembe venni. A tanulás lényege az, hogy a súlyokat úgy módosítjuk, hogy a megfigyelt (tehát a betanító adatokból származó) minták valószínűsége nőjön a háló által generált mintákhoz képest.

A gyakorlatban a teljesen általános Boltzmann-gépet gyakran túl nehéz hatékonyan tanítani. Ezért vezették be a „korlátozott Boltzmann-gépet” (RBM), amelynek nincsenek rejtett rétegen belüli kapcsolatai és nincs látható rétegen belüli kapcsolat sem. Csak a látható és a rejtett rétegek közt vannak súlyok (3. ábra).



3. ábra. A korlátozott Boltzmann-gépek (RBM) általános működési elve

$$E(v, h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} w_{ij} v_i h_j$$

$$= -\mathbf{a}^T \mathbf{v} - \mathbf{b}^T \mathbf{h} - \mathbf{v}^T \mathbf{W} \mathbf{h}$$

Ez a korlátozás jelentősen egyszerűsíti a betanítást. Az RBM-ek jellegzetessége az is, hogy felügyelet nélküli tanulást (unsupervised learning) tesznek lehetővé, így még osztályozási feladatok esetén sem kell előre kijelölni az egyes kategóriákat.

Ezen elvek révén született meg a Deep Belief Networks (DBN) koncepciója is, melyben az RBM-ek rétegeit egymásra halmozva mély és összetett reprezentációk hozhatók létre. A DBN-ek és a hozzájuk kapcsolódó módszertan lehetővé tették, hogy az addig nehezen tanítható, mély hálózatokat is hatékonyan be lehessen ta-

nítani rétegről rétegre történő, kezdetben felügyelet nélküli, majd finomhangolásként felügyelt tanulási lépésekkel. Ez a megközelítés mutatta meg igazán, hogyan tudnak a mély neurális hálózatok, összehangolt rétegek láncolataként, egyre absztraktabb és összetettebb jellemzőket megtanulni a nyers adatokból – legyen szó képekről, hangokról vagy szövegekről.

A Boltzmann-gépek (így a korlátozottak is) olyan neurális hálólok, amelyek a statisztikus fizika inspirációjára a neurális állapotok valószínűségi eloszlásán alapulnak. Nem fix bemenet-kimenet párjaik vannak, hanem a Hopfield-hálókhöz hasonlóan inkább egy belső szabály szerint „ringatóznak” különböző állapotok között, a kapcsolatok által meghatározott energiafelületen. Idővel képesek megtanulni az adatokat jellemző mintázatokot, és ezeknek megfelelően alakítani a súlyokat, így az alacsony energiaállapotokhoz azok a minták tartoznak, amelyek a háló megtanult. Az RBM-ek ugyan jelenleg a ritkábban alkalmazott modellek közé tartoznak, ám mégis fontos szerepet játszottak a mélytanulás előretörésében és nagyban hozzájárultak ahhoz, hogy a gépi tanulás mára ennyire fejlett és hatékony legyen.

A 2000-es évek végére és a 2010-es évek elejére Hinton és tanítványai (köztük Yann LeCun és Yoshua Bengio) olyan módszereket és elveket dolgoztak ki, amelyek a számítási teljesítmény növekedésével (főleg a GPU-k alkalmazásának elterjedésével) és az egyre nagyobb adattömegek (Big Data) rendelkezésre állásával együtt elhozták a mélytanulás korszakát. A beszédfelismerés, a képfelismerés és számos más alkalmazási terület ezáltal ugrásszerű fejlődésen ment keresztül.

Különösen látványos eredményt ért el Hinton csoportja a 2012-es ImageNet kihíváson, ahol egy AlexNet névre keresztelt mély konvolúciós hálózat több mint 10 százalékponttal jobban teljesített a képfelismerési feladatban, mint az addigi legjobb megoldások [5]. Ez a pillanat mérföldkő volt a számítógépes látás területén, és világszerte felkeltette az ipar érdeklődését a mélytanulás iránt. A Google hamar felfigyelt Hinton munkájára, és 2013-ban alkalmazta is őt, majd megalapozta a Google Brain kezdeményezést, amely a mesterséges intelligencia kutatásának egyik vezető műhelye lett.

Hinton nemcsak kutatóként, hanem tudományos szervezőként is maradandót alkotott. A Torontói Egyetemen belül és körül kialakuló kutatói ökoszisztéma révén hozzájárult a kanadai mesterséges intelligencia-kutatás virágzásához. Közreműködött a Vector Institute for Artificial Intelligence megalapításában, amely a kanadai kormány és ipari partnerek támogatásával a mesterséges intelligencia kiválósági központjaként működik.

Tudományos munkásságáért számos rangos kitüntetést kapott. Közülük kiemelkedik a 2018-ban megítélt Turing-díj, amelyet gyakran a „számítástechnika Nobel-díjaként” is emlegetnek. Ezt a díjat Hinton megosztva kapta Yann LeCun és Yoshua Bengio kutatókkal, akik szintén a mélytanulás területének meghatározó úttörői. A díj hivatalos indoklása szerint munkájuk forradalmasította a számítógépes tanulás képességét, amellyel felismerhetők a képek, hangok és szöveges adatok komplex struktúrái. ●●●

IRODALOM

[1] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. Natl. Acad. Sci. USA (1982) 79, 2554–2558.
 [2] J. J. Hopfield, Neurons, Dynamics and Computation, Physics Today (1994) 47, 40–46.
 [3] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors, Nature (1986) 323, 533–536.
 [4] D. H. Ackley, G. E. Hinton, T. J. Sejnowski, A learning algorithm for Boltzmann machines, Cognitive Science (1985) 9, 147–169.
 [5] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, NIPS'12: Proceedings of the 26th International Conference on Neural Information Processing Systems – Vol. 1 (2012) 1, 1097–1105.