

A mélytanulás múltja, jelene és jövője

GYIRES-TÓTH BÁLINT

Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Médiainformatikai Tanszék
toth.b@tmit.bme.hu

Kulcsszavak: mesterséges intelligencia, gépi tanulás, mélytanulás, neurális hálózatok

Az elmúlt években az adatvezérelt rendszerek nagyobb pontosságuknak köszönhetően egyre jobban teret nyertek a szabály alapú megoldásokkal szemben. Míg az utóbbi esetben egy adott tématerület szakértői adják meg a modellezendő valós folyamatot leíró szabályokat, adatvezérelt rendszerek esetében – jellemzőn szintén szakértők bevonásával – a folyamatból rögzített adatokkal tanítanak be gépi tanuló eljárásokat. Napjainkban a gépi tanulás alapú mesterséges intelligencia (MI) egyik legmeghatározóbb ága a mélytanulás, melynek az alapjait jelentő mesterséges neurális hálózatok tudományterület népszerűsége már két jelentős „visszaesésen” is túl van, ezért jogosan vethető fel a kérdés; miben különbözik a mostani helyzet a korábbiaktól?

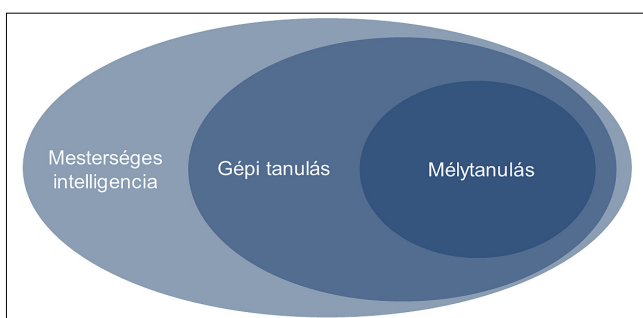
A cikkben bemutatásra kerülnek a mélytanulás legfontosabb előzményei és jelenlegi helyzete, továbbá a várható jövőbeli lehetőségeket is megvizsgáljuk.

1. Bevezetés

Az adatmennyiség robbanásszerű növekedésével, a grafikus processzorok (*Graphics Processing Unit, GPU*) jelentős technológiai fejlődésével, a gépi tanulás és a mesterséges neurális hálózatok új eredményeinek köszönhetően az elmúlt években a mélytanuló (*deep learning*) rendszerek a valós életbeli folyamatok – megfigyelések alapján történő – modellezésének egyik leghatékonyabb eszközévé váltak. Mára számos tudományterületen a legjobb eredményt, legnagyobb pontosságot mélytanulás segítségével érik el, legyen szó beszédtechnológiáról, gépi látásról, természetes nyelvfeldolgozásról vagy például idősor-modellezésről.

Az olyan technológiai óriásvállalatok, mint például a Google, Facebook, Amazon és az IBM számos piaci termékében sikeresen használnak mélytanulás alapú megoldásokat. Hasonló algoritmusokat más alkalmazási területeken felhasználva (pl. mezőgazdaság, könnyű- és nehézipar, vagy a távközlés) a mélytanulási „forradalom” második hulláma várható. Mik azok az előzmények, melyek a mai napig meghatározó szerepet játszanak? Hol tart ma a mélytanulás, és mi várható a következő években? A cikk ezekre a kérdésekre próbál választ adni.

1. ábra Az MI, a gépi tanulás és a mélytanulás kapcsolata



1.2. A mesterséges intelligencia, gépi tanulás és a mélytanulás fogalma

A mesterséges intelligencia (*MI, angolul: artificial intelligence, AI*), a gépi tanulás (*machine learning*) és a mélytanulás (*deep learning*) kapcsolatát az 1. ábra mutatja be.

A mesterséges intelligencia alap gondolatát Alan Turing 1950-ben írt „Computing Machinery and Intelligence” művében a gondolkodó gép víziójával vezette be [1], majd John McCarthy vezetésével 1956-ban a „Dartmouth Summer Research Project on Artificial Intelligence” workshopon meghatározták a mesterséges intelligencia tudományterület alapjait. Egy általánosságban vett mesterséges intelligencia

- (1) képes környezete ismeretében döntéseket hozni,
- (2) ezek a döntések hasonlóak a természetes intelligenciához, és
- (3) ha a környezet (akár jelentős mértékben is) megváltozik, képes a megváltozott környezethez alkalmazkodni.

A mesterséges intelligencia három fő típusát különböztetjük meg: az általános, a keskeny és a szuper MI-t.

- A fentebbi három pontból álló meghatározás legtágabb értelemben véve jelenti az általános MI-t (*Artificial General Intelligence, AGI*) – napjainkban ilyen nem létezik és a jelenlegi szoftver- és hardvereszközökkel ez nem megoldható.

- Keskeny MI-nek (*Artificial Narrow Intelligence, ANI*) azon eljárásokat nevezzük, melyek egy adott tudományterületen (pl. képfelismerés, beszédszintézis, gépi játékok stb.) nagy hatékonysággal – sokszor az emberénél is jobban – futtathatók, azonban más tématerületen nagyon korlátozottan, vagy egyáltalán nem működnek.

- A mesterséges szuperintelligencia (*Artificial Super Intelligence, ASI*) pedig az olyan rendszereket jelöli, melyek képesek az emberi tudáson túlmúató következte-

téseket, döntéseket hozni. Ez elsődlegesen az olyan szereplők számára érhető el, melyek óriási adatvagyonnal és számítási kapacitással rendelkeznek.

A mesterséges intelligenciának meghatározó ága a gépi tanulás. Gépi tanuló rendszereknek az olyan – elsődlegesen statisztika alapú – algoritmusokat nevezzük, melyek szabályok explicit megadása nélkül, a folyamatból rögzített minták alapján képes egy adott folyamatot szabályzó sajátosságok meghatározására. Számos gépi tanuló algoritmus létezik, mint például a lineáris és logisztikus regresszió, a rejtett Markov-modellek (*Hidden Markov Model, HMM*), a szupport vektor gépek (*Support Vector Machines*), a döntési fa, a random forest és a gradient boosting alapú technikák. Ezek mind más és más megközelítést alkalmaznak az adatok modellezésére.

A gépi tanulás napjainkban talán leginkább hangsúlyos típusa a mélytanulás. A mélytanulási paradigma kivétel nélkül (mély) neurális hálózatokat alkalmaz. A mély neurális hálózatok definíció szerint az olyan mesterséges neurális hálózatokat jelölik, amelyekben több mint egy rejtett réteg van (2. ábra). A szerző véleménye, hogy a deep learning kifejezés a fentiekén túl magában foglalja az összes olyan elméleti és gyakorlati eredményt (pl. mély hálók hatékony tanítása [2], variációs autoenkóderek [3], generatív versengő hálózat [4], komplex architektúrák [5,6] stb.), valamint a szoftver- és hardverinfrastruktúrát, melyek a mai modern mesterséges (mély) neurális hálózatok hatékony tanításához és futtatásához elengedhetetlenek.

A mai mély neuronhálók előnyei más gépi tanuló algoritmusokkal szemben többek között abban rejlik, hogy a folyamatból rögzített adatokat egyszerre tanulja meg a célnak leginkább megfelelő reprezentációra alakítani, illetve modellezni azokat. Más típusú eljárások esetén jellemzően az adatokat szakértők és adattudósok bevonásával előfeldolgozzuk (preprocessing) és a célnak leginkább megfelelő formátumra alakítjuk, illetve bővítjük

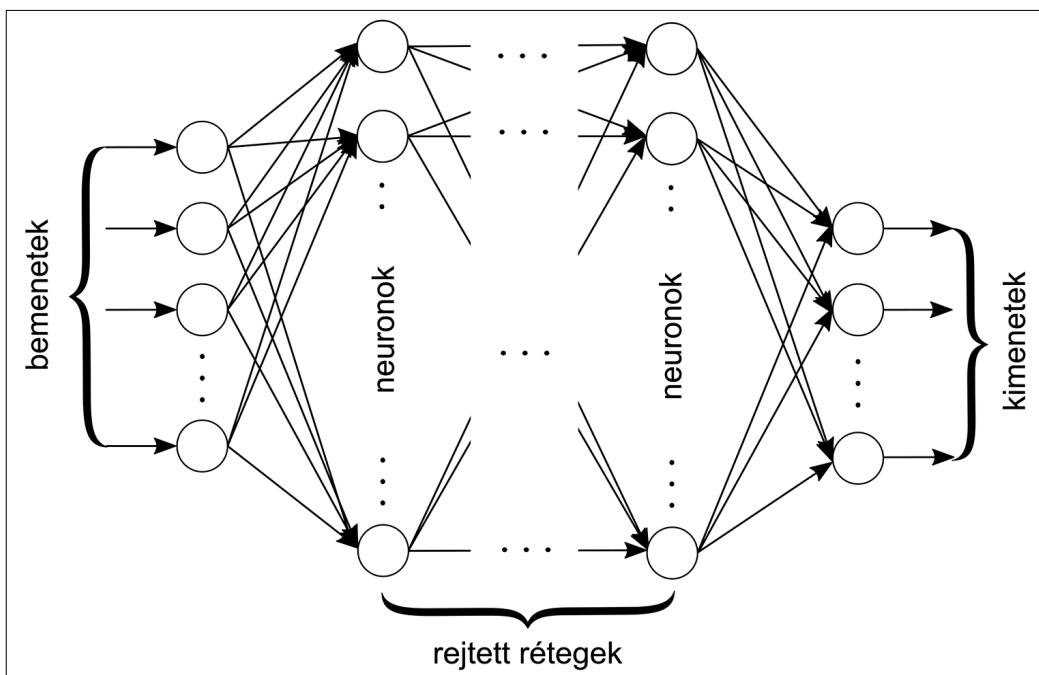
(ezt hívjuk *feature engineering*-nek). Ezt követően – az átalakítástól függetlenül – tanítjuk az így kialakított adatokkal a kiválasztott regressziós vagy osztályozó modellt. Ez esetben a végső modell minősége jelentősen függ a kialakított adatformátumtól. Kis mennyiségű adat esetén ez a megközelítés továbbra is célravezető megoldás lehet. Amennyiben pedig nagy mennyiségű adat áll rendelkezésre, akkor a mély neurális hálózatok képesek egyszerre a modellezés számára legjobb adatformátum kinyerését (ún. *feature* vagy *representation learning*) és ennek modellezését megtanulni, ami sokszor nagyobb pontossághoz vezet.

A neurális hálózatban jellemzően nem lehet egyértelműen elkülöníteni, hogy a hálózat melyik része végzi az egyik, melyik a másik feladatot. Jelen cikk a mélytanuláson belül elsődlegesen a felügyelt tanuláshoz köthető megoldásokra koncentrál.

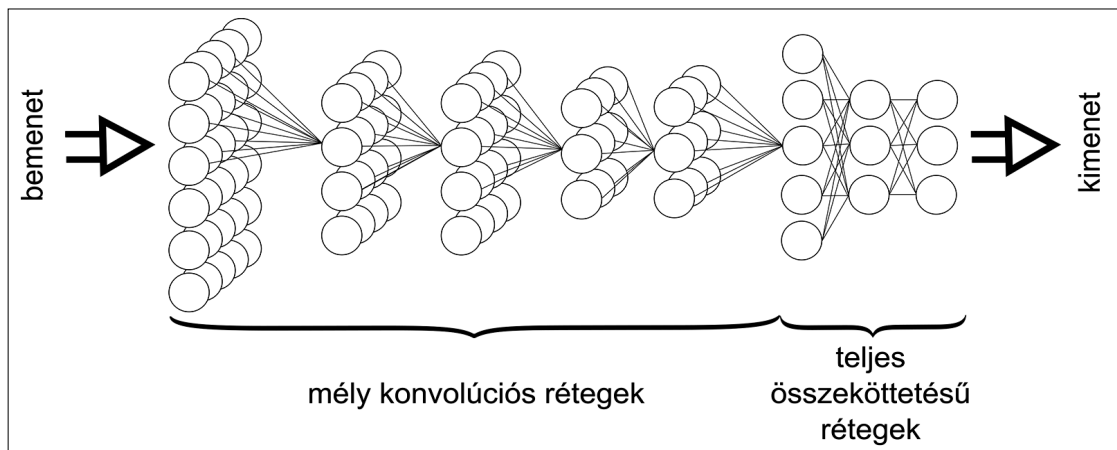
2. A mélytanulás múltja

A mélytanulás gyökerei az 1940-es évekre nyúlnak vissza, amikor Walter Pitts és Warren McCulloch javaslatot tett az idegsejtek egyszerűsített matematikai modelljére [7]. Következő meghatározó lépés az volt, amikor 1957-ben, majd 1962-ben Frank Rosenblatt megalkotta az egyes többrétegű perceptront, amely a mai mély neurális hálózatok alapját jelenti [8]. A többrétegű perceptront hatékony tanítását lehetővé tevő ún. hibavisszaterjesztés (*backpropagation*) eljárásához szükséges automatikus differenciálást Seppo Linnainmaa 1970-ben dolgozta ki [9], majd ennek neurális hálózatokban való sikeres alkalmazását David Rumelhart, Geoffrey Hinton és Ronald Williams tette meg 1986-ban [10].

A mély neurális hálózatok tanításához ma is ezt az algoritmust, illetve ennek GPU-ra optimalizált különböző változatait használjuk. A többrétegű perceptront, melyet



2. ábra
A teljes összeköttetésű mély neurális hálózat sematikus váza



3. ábra
Mély konvolúciós
neurális hálózat
sematikus váza

ma teljes összeköttetésű neurális hálózatnak is hívják, a be- és kimenet közötti leképzést próbálja megtanulni az adatokból (lásd 2. ábra), azonban jellemzően ezek a rendszerek nem képesek a bemenetben található hely- vagy időbeli mintázatok megtanulására. Ez annyit jelent, hogy két bemenet egymáshoz viszonyított távolsága nem jelent a hálózat számára többletinformációt. A többrétegű perceptron mellett két további meghatározó neurális hálózat típus a konvolúciós és rekurrens neurális hálózat. Ezen architektúrák a hibavisszaterjesztés algoritmusnak módosított változatait alkalmazzák a tanítás során.

A mintafelismerésben nagy sikerrel alkalmazott konvolúciós neurális hálózatok (*Convolutional Neural Network, CNN*) alapját Kunihiko Fukushima határozta meg 1979-ben [11], melynek egyik első sikeres alkalmazása Yann LeCun nevéhez fűződik [12]. A mély konvolúciós neurális hálózatok jellemzően két fő egységre bonthatóak: az elsődlegesen reprezentáció tanulását felelős konvolúciós (és ehhez kötődő) rétegekre, illetve a modellezést (pl. regresszió, osztályozás) végző teljes összeköttetésű rétegekre (3. ábra). Napjainkban a mélytanuló rendszerek markáns részében használunk ilyen architektúrát (pl. gépi látással, beszédfelismeréssel és beszédkeltéssel, vagy általános idősor elemzéssel kapcsolatos alkalmazásokban).

Az úgynevezett rekurrens neurális hálózatok (*Recurrent Neural Network, RNN*) irányított, kört is tartalmazó számítási gráfokat használnak a be- és kimenet közötti kapcsolat meghatározásához, s ezáltal alkalmasak szekvenciális adatok, idősorok modellezésére. Az RNN alapjait John Hopfield 1982-es [13] munkája jelenti. Az RNN-ek sokáig csak rövid szekvenciákat tudtak hatékonyan modellezni, hosszú szekvenciák esetén az ún. elenyésző gradiens következtében a modellezés pontossága drasztikusan lecsökkent. Ennek a problémának a megoldására nyújtott első között megoldást Hochreiter és Schmidhuber meghatározó munkája, a Long Short-Term Memory (LSTM) [14].

A fenti három különböző típusú neurális hálózatot (teljes összeköttetésű, konvolúciós és rekurrens neurális

hálózat) a célnak megfelelően akár egy számítási gráfban is lehet kombinálni, és az így kialakított számítási gráf paramétereit az adatok alapján egyben lehet optimalizálni a hibavisszaterjesztés algoritmus segítségével. Sok* rejtett réteget tartalmazó neurális hálózat tanításában azonban csak 2010 után értek el látványos áttörést.

A mélytanulás paradigma elterjedését részben a korlátozott Boltzmann-gépekből (*Restricted Boltzmann Machine, RBM*) felépített *Deep Belief Network (DBN)* architektúrák vezették be [15]. Ezen hálózatokban az ún. rétegenkénti előtanítással már hatékonyan tudtak mélyebb struktúrákat is tanítani. Az igazán látványos áttörést Alex Krizhevsky, Ilya Sutskever és Geoffrey Hinton munkája hozta meg a 2012-es ImageNet képfelismerési versenyen [16], amikor szignifikánsan jobb eredményt értek el immár GPU-kon megvalósított konvolúciós neurális hálózat segítségével a korábbi módszerekkel szemben.

Innentől kezdve ismét óriási figyelmet kapott a neurális hálózatok tudományág és népszerűsége évről évre nőtt. Ahhoz, hogy a mai komplex mélytanuló rendszerek adott alkalmazási területeken a legnagyobb pontosságot tudják elérni még számos elméleti eredményre (pl. törtvonalas aktivációs függvény [17], reziduális hálózatok [5], dropout [18] eljárás), illetve a hardver- és szoftverinfrastruktúra folyamatos fejlődésére volt szükség.

Természetesen a neurális hálózatok témakörében a fentiekben és a továbbiakon túl mind nemzetközi, mind hazai szinten rengeteg kimagasló eredmény született, melyekre e cikk a tudományterület szerteágazó volta miatt nem tér ki.

3. A mélytanulás jelene

A mélytanulás és a neurális hálózatok iránti érdeklődést jól mutatja a szakterület talán leginkább meghatározó konferenciájára, a NeurIPS (Neural Information Processing Systems) konferenciára beadott cikkek számának alakulása az évek során (4. ábra). Hasonló látványos növekedést láthatunk, ha a nyílt forráskódú mélytanuló megoldások terjedését, vagy a médiában a „mestersé-

* Hogy mi a „sok”, azt minden időben a számítási kapacitás és az adatok mennyisége szabta meg. A 2010-es években 8-10 réteg már soknak számított [16], míg 2015-ben már a 150 réteg számított soknak [5], a cikk írásának idejében pedig az egyik legnagyobb sikeresen alkalmazott neurális hálózat (Megatron-LM) mintegy 8.3 milliárd tanítható súllyal rendelkezik.

ges intelligencia” kifejezés előfordulásának gyakoriságát vizsgáljuk meg (mely valójában a legtöbb esetben mélytanulás alapú megoldásokat takar).

A mély neurális hálózatok ma már számtalan alkalmazási területet érintenek, az alábbiakban az elsődleges alkalmazási területek legújabb, jövőbemutató eredményeire koncentrálnak. Mindhárom alkalmazási területen elért eredmények jelentősen hozzájárultak a neurális hálózatok elméleti hátterének a fejlődéséhez is.

3.1. Gépi látás

A gépi látás témakörbe tartozik többek között a képfelismerés (megadjuk, hogy mi látható a képen/videón) [16,5], az objektum-felismerés (megadjuk, hogy mik és hol láthatók a képen/videón) [20], a szemantikus szegmentáció (képpontonként megadjuk, hogy mi látható a képen/videón) [21], mélységbecslés (képpontonként meghatározzuk a képpont tartalmának becsült távolságát a kamerától) [22], tartalomrekonstrukció (hiányzó vagy nehezen felismerhető, alacsony felbontású részek becslése), továbbá a kép- és videogenerálás (vizuális tartalom létrehozása adott témakörben) [23].

Folyamatosan új alkalmazási területekre tör be a mélytanulás alapú gépi látás, illetve az új megközelítések a korábbiaknál is jobb pontosságot adnak a már „meghódított” alkalmazási területeken. A mindennapi élet szempontjából fontos kiemelni például az orvosi képfeldolgozást, ahol megfelelő méretű és pontosságú adatbázisok mellett a mélytanuló rendszereknek köszönhetően, már több területen az orvosok munkáját jelentősen segítő megoldások születtek (pl. [24]).

3.2. Beszédtechnológia és hangfeldolgozás

A beszédtechnológia két legmeghatározóbb ága a gépi beszédeltetés (gépelt szöveg emberi hangon történő felolvasása) [6] és a beszédfelismerés (az elhangzott beszéd szöveges átiratának az elkészítése) [25], illetve ezeken túl számos további részterülete van, mint például

a beszélő felismerése és azonosítása [26], érzelemfelismerés [27], orvosi beszédfeldolgozás [28] stb. A korábbi, komponensekre bontott (pl. nyelvi elő-, utófeldolgozás, beszédkódoló stb.) modellezéssel szemben ma már elsődlegesen az úgynevezett end-to-end mély neurális hálózat alapú rendszerek eredményei a leginkább előremutatóak.

A beszédjelek mellett még fontos megemlíteni a nem beszéd jellegű hang feldolgozását, mint például zene, utca- és járműzaj, gyártósor, állathang stb. modellezését. Zene esetén a feladat jellemzően a lejátszott mű vagy annak stílusának felismerése, illetve zenegenerálás. Általános hangok esetén pedig sok esetben osztályokba szeretnénk sorolni a hangokat (például normális, illetve abnormális gyártósori működés, vagy madárfajta meghatározása).

3.3. Természetes nyelvfeldolgozás

Természetes nyelvfeldolgozás (*Natural Language Processing, NLP*) esetén elsődleges feladatok között szerepel a szöveg szavainak címkézése [29], a szöveg értelmezése, kontextus kinyerése [30], gépi fordítás forrásnyelvről célnyelvre [31], szövegek összehasonlítása és kategóriákba sorolása, szentiment-elemzés (pozitív vagy negatív hangulatú-e a szöveg) [32], a szöveggenerálás (egy adott témakörben értelmes, koherens szöveg létrehozása) és a célspecifikus tartalomdetekció (pl. gyűlöltettség, megtévesztés). A mélytanulás évről évre jobb és jobb eredményeket hozott az NLP területén, például az LSTM alapú modellezéssel [14], illetve a Seq2seq modell és a figyelmi mechanizmus bevezetésével [33]. A közelmúltban a mélytanuláson belül talán a leginkább előremutató eredményeket is az NLP területén érték el: az úgynevezett Transformer architektúra új, eddig soha nem látott szintre emelte a szövegmodellek pontosságát [30,34].

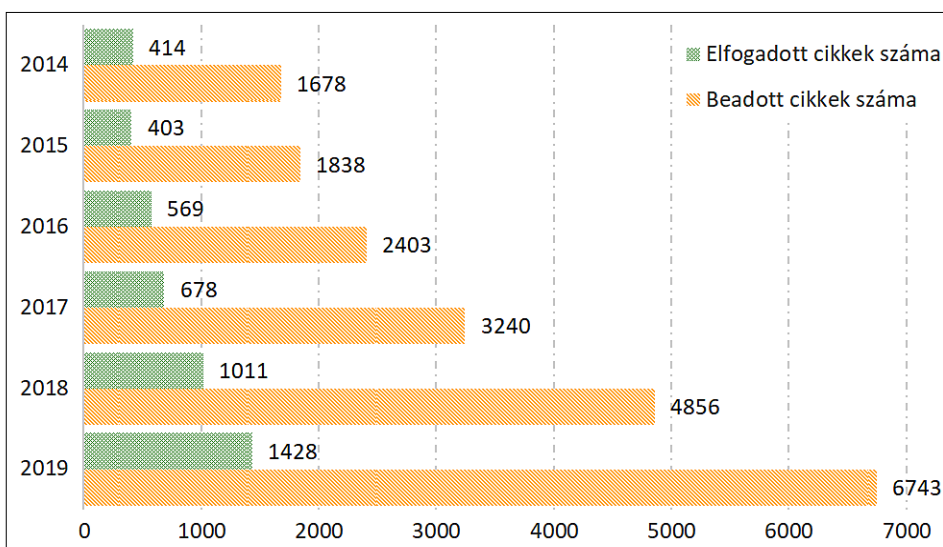
Mindhárom területen alkalmazott modellek sok esetben általánosíthatók szekvenciális adatok, idősorok modellezésére. Szekvenciális adat lehet például szöveg, hang, pénzügyi idősorok, a távközlési hálózatok jelzései, a szenzorokkal mért idősorok, továbbá még az egymást követő képpontok vízszintes és függőleges irányban vizsgálva is szekvenciális adatstruktúrát alkotnak.

Szekvenciális adatok esetén az elsődleges kihívások; (1) a receptív mező növelése (ez annak az adatrésznek a méretét jelenti, melyet a modellezés során figyelembe veszünk a bemeneten), (2) a több időskálán való modellezés, és a (3) többdimenziós adatok modellezése.

E kihívások megoldására több megközelítés is kísérletet tesz [35,36], illetve egyes

4. ábra

A NeurIPS konferenciák cikkszámai 2015–2019 között (módosítva [19] alapján)



modellek (pl. WaveNet [6]) implicit módon kezelik ezeket. Mindemellett a fenti három kihívás minél pontosabb és jól magyarázható megoldásával várhatóan a mostani modellek még tovább javulhatnak.

A mélytanulás algoritmikus eredményei mellett fontos az ezt kiszolgáló szoftverekre is kitérni. A mélytanulás elsődleges programozási nyelve a Python és a lentebb felsorolt szoftverek mind ingyenesen elérhetőek. A cikkben a legelterjedtebb eszközökre koncentrálna csupán néhány példa kerül említésre.

Adattárolás, -betöltés és -előfeldolgozás

A jól ismert megoldások (SQL, NoSQL, Pandas, Dask) mellett használhatóak az utóbbi időben megjelent olyan eszközök, melyek a teljes folyamatnak már az első lépéseit is a CPU-ról a GPU-ra helyezik át. Így a BlazingSQL (<https://blazingsql.com/>) az adatokat egyenesen a GPU-memóriába olvassa be, míg a Rapids.AI (<https://rapids.ai/>) és a CuPy (<https://cupy.chainer.org/>) az adat-előfeldolgozás lépését végzi a GPU-n.

Mélytanuló keretrendszerek

A legismertebb mélytanuló keretrendszerek a Google által fejlesztett TensorFlow (<https://www.tensorflow.org/>) és az elsődlegesen a Facebook AI Research által gondozott PyTorch (<https://pytorch.org/>). Mindkettő általános keretrendszer, melyek automatikus differenciálás segítségével tetszőleges számítási gráfon képesek a gráf paramétereit (a neurális hálózat súlyait) be- és kimeneti adatok alapján optimalizálni. Ezen felül megtalálhatóak bennük a leggyakrabban használt mélytanuló eljárások, továbbá számos kiegészítő, mely a hatékony tanítást segíti (adatrepresentáció, modellek futtatása, tömörítése).

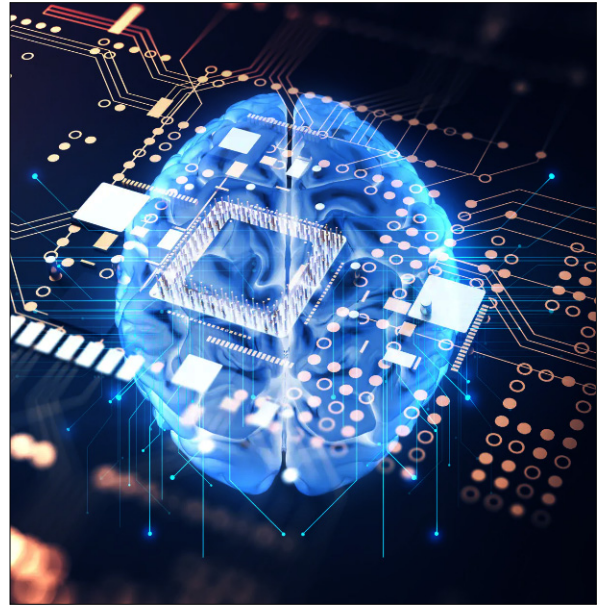
Virtualizáció és skálázhatóság

A munkát jelentősen segíti az operációs rendszer és a GPU Docker (<https://www.docker.com/>) alapú virtualizációja egyrészt azért, hogy új rendszerek kialakítása során a szoftverek telepítése, másrészt a megoldás cél-eszközökre történő migrálása jelentősen egyszerűsödik. Lehetőség van továbbá több tíz, száz, vagy akár ezer GPU-t is támogató Docker konténer összehangolt futtatására, például a Kubernetes (<https://kubernetes.io/>), a KubeFlow (<https://www.kubeflow.org/>) és a Horovod (<https://github.com/horovod/horovod>) segítségével.

4. A mélytanulás jövője

Az elmúlt években óriási fejlődésen ment keresztül a tudományterület, ugyanakkor várható, hogy ez a folyamat a jövőben is folytatódik. Számos kiszámíthatatlan és előre nem megjósolható esemény mellett vizsgáljuk meg azokat a pontokat, melyek szükségességük miatt nagy valószínűséggel be fognak következni.

Számos alkalmazási területen minden további elméleti eredmény nélkül is, pusztán a címkézett adatmennyiség és a számítási kapacitás folyamatos növekedésével várhatóan minden korábbinál nagyobb pontossá-



got elérő mélytanuló modellek fognak születni. Rendkívül izgalmas ezen belül a mélytanulás használatától várható további eredmények, például a csillagászat [37], a DNS elemzés [38] és a sejtbiológia [39] tudományterületeken.

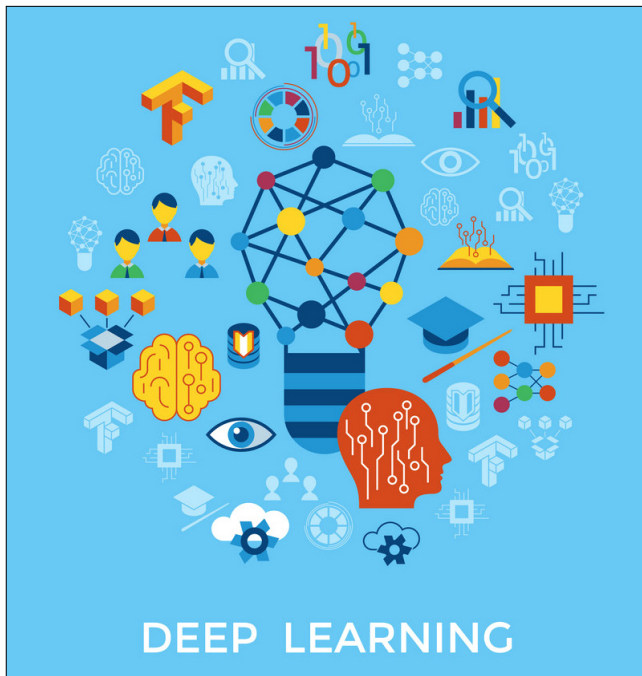
A mélytanulás alkalmazásának első hulláma már lezajlott a nagy technológiai óriásvállalatoknál; jelentős méretű kutatólaboratóriumokban létrehozott neurális hálózat alapú megoldásaikat sikeresen alkalmazzák termékeikben. Már most számos tanulmány szól a mélytanulás alapú ipar 4.0 [40], mezőgazdasági [41], vagy például a víziközlekedési [42] megoldásokról. Ahogyan az általános mélytanuló eszközök várhatóan egyszerűbbé válnak, kiszélesedik az azt alkalmazó kutatók és fejlesztők köre, és így számos új – mind az informatikától, mind pedig a gépi tanulástól távolabbi – területen is meg fognak jelenni a mélytanulás alapú megoldások.

Az elméleti eredmények tekintetében amennyiben az eddigi kutatási fókusz megmarad, akkor további komplex szekvenciális adatmodellek várhatóak, melyek még pontosabb eredményt nyújtanak, például beszédtechnológia és természetes nyelvfeldolgozás területeken. Továbbra is fontos kérdés marad a receptív mező növelése és a többskálás modellezés. Különösen a hatalmas számítási kapacitást igénylő modellek és eljárások esetén van szükség hasonló pontosságot nyújtó, gyorsabb megoldásokra (mint például a 2019. novemberében publikált SHA-RNN [43]). A felügyelet nélküli és megerősítéses tanítás továbbra is nagy, megoldásra váró kihívásokat állít a kutatók elé.

Szoftverek terén is szükségszerű a további fejlődés: egyrészt a mélytanuló keretrendszerek közötti átjárhatóság egyszerűsödésével nagyobb kutatási és fejlesztési projektek esetén sem leszünk kitéve egy adott rendszernek, másrészt így mindig a céljainkat leginkább segítő, publikusan elérhető megoldásokat tudnánk használni. Ezen túl a magas szintű megoldások, továbbá a virtualizáció és a skálázhatóság folyamatos fejlődése és egyszerűsödése szükségességükből fakadóan várható.

5. Összefoglalás

A tudományos világban is egy-egy téma népszerűsége kapcsán sok esetben ciklikusság figyelhető meg. Jelenleg az adatvezérelt modellek korát éljük, melyen belül fontos szerepet tölt be a mélytanulás. A mesterséges intelligencia alap- és alkalmazott kutatása már túl van két olyan szakaszon, amikor jelentős visszaesés volt tapasztalható az MI népszerűségében (egyszer a 80-as, egyszer pedig a 90-es évek kezdetén). A jelenlegi fokozott figyelem jogosan veti fel egy következő visszaesés lehetőségét, részben a túlzott elvárások és a be nem váltott ígéretek következményeként. Lehetséges, hogy akár a közeljövőben lesz ilyen jellegű visszaesés, azonban a folyamatosan növekvő adatmennyiség hatékony modellezésére jelenleg nincs más általános megoldás, mint a gépi és mélytanulás, ezért visszaesés esetén sem várható, hogy teljesen háttérbe szoruljon e tudományterület.



Köszönetnyilvánítás

Jelen tanulmányt támogatták:
 az Emberi Erőforrások Minisztériuma által biztosított BME – Mesterséges Intelligencia Felsőoktatási Intézményi Kiválósági Program (BME FIKP-MI/SC),
 a Bolyai+ Felsőoktatási Fiatalkutatói, Kutatói Ösztöndíj (ÚNKP-19-4-BME-189) és
 az MTA Bolyai János Kutatási Ösztöndíj.

A szerzőről



GYIRES-TÓTH BÁLINT a Budapesti Műszaki és Gazdaságtudományi Egyetem, Távközlési és Média-informatikai Tanszék oktatója és kutatója. 2007 óta foglalkozik elméleti és alkalmazott gépi tanulással. Doktori fokozatának megszerzése óta, 2014-től a mélytanulás (deep learning) az elsődleges kutatási területe. Számos sikeres kutatási és ipari projektben vett részt. 2017 óta felkérés alapján az NVIDIA Deep Learning Institute (DLI) minősített oktatója és egyetemi nagyköve.

Hivatkozások

- [1] Turing, A.M.,
Computing machinery and intelligence, *The Essential Turing: The Ideas that Gave Birth to the Computer Age*. Ed. B. Jack Copeland, Oxford: Oxford UP (1950.), pp.433–464., 2004.
- [2] LeCun, Y., Léon, B., Genevieve, B.O., Klaus-Robert, M.,
Efficient backprop., *Neural networks: Tricks of the trade*, Springer, Berlin, Heidelberg, pp.9–48., 2012.
- [3] Kingma, D. and Welling, M.,
Auto-Encoding Variational Bayes, *The 2nd International Conference on Learning Representations (ICLR)*, 2013.
- [4] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B.,
Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.,
Generative adversarial nets,
Advances in Neural Information Processing Systems, pp.2672–2680., 2014.
- [5] Szegedy, C., Ioffe, S., Vanhoucke, V. and Alemi, A.A.,
Inception-v4, inception-resnet and the impact of residual connections on learning,
31th AAAI Conference on Artificial Intelligence, 2017.
- [6] Van den Oord, A., Dieleman, S., Zen, H., Simonyan, K.,
Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A.
and Kavukcuoglu, K.,
WaveNet: A Generative Model for Raw Audio.
In 9th ISCA Speech Synthesis Workshop, p.125., 2016.
- [7] McCulloch, W.S. and Pitts, W.,
A logical calculus of the ideas immanent in nervous activity,
The Bulletin of Mathematical Biophysics, 5(4), pp.115–133., 1943.
- [8] Rosenblatt, F.,
Principles of neurodynamics. perceptrons and the theory of brain mechanisms (No. VG-1196-G-8),
Cornell Aeronautical Lab. Inc., Buffalo NY., 1961.
- [9] Linnainmaa, S.,
The representation of the cumulative rounding error of an algorithm as a Taylor expansion of the local rounding errors, *Master's Thesis (in Finnish)*,
University of Helsinki, 1970.
- [10] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.,
Learning internal representations by error propagation in parallel distributed processing,
Explorations in the Microstructure of Cognition, pp.319–362., 1986.
- [11] Fukushima, K.,
Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,
Biological Cybernetics, 36(4), pp.193–202., 1980.
- [12] Y. LeCun, B. Boser, J.S. Denker, D. Henderson,
R.E. Howard, W. Hubbard and L.D. Jackel,
Backprop. Applied to Handwritten Zip Code Recognition,
Neural Computation, 1(4), pp.541–551., 1989.
- [13] J.J. Hopfield,
Neural networks and physical systems with emergent collective computational abilities,
Proceedings of the National Academy of Sciences 79, pp.2554–2558., 1982.
- [14] Hochreiter, S. and Schmidhuber, J.,
Long short-term memory,
Neural Computation, 9(8), pp.1735–1780., 1997.
- [15] Hinton, G.E., Osindero, S. and Teh, Y.W.,
A fast learning algorithm for deep belief nets,
Neural Computation, 18(7), pp.1527–1554., 2006.

- [16] Krizhevsky, A., Sutskever, I. and Hinton, G.E., Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems*, pp.1097–1105., 2012.
- [17] Glorot, X., Bordes, A. and Bengio, Y., Deep sparse rectifier neural networks, *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics*, pp.315–323., 2011.
- [18] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R., Dropout: a simple way to prevent neural networks from overfitting, *The Journal of Machine Learning Research*, 15(1), pp.1929–1958., 2014.
- [19] NeurIPS 2019 Stats, <https://medium.com/@dcharrezt/neurips-2019-stats-c91346d31c8f>, (hozzáférés dátuma: 2020. január 30.)
- [20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y. and Berg, A.C., SSD: Single shot multibox detector, *European Conf. on Computer Vision*, pp.21–37., 2016.
- [21] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F. and Adam, H., Encoder-decoder with atrous separable convolution for semantic image segmentation, *European Conf. on Computer Vision*, pp.801–818., 2018.
- [22] Liu, F., Shen, C. and Lin, G., Deep convolutional neural fields for depth estimation from a single image, *IEEE Conf. on Computer Vision and Pattern Recognition*, pp.5162–5170., 2015.
- [23] Van den Oord, A., Kalchbrenner, N., Espeholt, L., Vinyals, O. and Graves, A., Conditional image generation with PixelcCNN decoders, *Advances in Neural Information Processing Systems*, pp.4790–4798., 2016.
- [24] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S., Dermatologist-level classification of skin cancer with deep neural networks, *Nature*, 542(7639), pp.115–118., 2017.
- [25] Collobert, R., Puhersch, C. and Synnaeve, G., Wav2letter: an end-to-end convnet-based speech recognition system, *arXiv preprint arXiv:1609.03193.*, 2016.
- [26] Snyder, D., Garcia-Romero, D., Povey, D. and Khudanpur, S., Deep Neural Network Embeddings for Text-Independent Speaker Verification, *Interspeech 2017*, pp.999–1003., 2017.
- [27] Fayek, H.M., Lech, M. and Cavedon, L., Evaluating deep learning architectures for Speech Emotion Recognition, *Neural Networks*, 92, pp.60–68., 2017.
- [28] Ma, X., Yang, H., Chen, Q., Huang, D. and Wang, Y., Depaudionet: An efficient deep model for audio based depression classification. In *Proc. of the 6th International Workshop on Audio/Visual Emotion Challenge*, pp.35–42., 2016.
- [29] Yadav, V. and Bethard, S., A survey on recent advances in named entity recognition from deep learning models. *arXiv preprint arXiv:1910.11470.*, 2019.
- [30] Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., Bert: Pre-training of deep bidirectional transformers for language understanding, *arXiv preprint arXiv:1810.04805.*, 2018.
- [31] Bahdanau, D., Cho, K. and Bengio, Y., Neural machine translation by jointly learning to align and translate. *International Conf. on Learning Representations*, 2015.
- [32] Zhang, L., Wang, S. and Liu, B., Deep learning for sentiment analysis: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), 2018.
- [33] Sutskever, I., Vinyals, O. and Le, Q.V., Sequence to sequence learning with neural networks, *Advances in Neural Information Processing Systems*, pp.3104–3112., 2014.
- [34] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I., Attention is all you need, *Advances in Neural Information Processing Systems*, pp.5998–6008., 2017.
- [35] J. Chung, S. Ahn, and Y. Bengio, Hierarchical multiscale recurrent neural networks, *arXiv preprint arXiv:1609.01704*, 2016.
- [36] S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, A. Courville, and Y. Bengio, Samplernn: An unconditional end-to-end neural audio generation model, *arXiv preprint arXiv:1612.07837*, 2016.
- [37] Gabbard, H., Williams, M., Hayes, F. and Messenger, C., Matching matched filtering with deep networks for gravitational-wave astronomy, *Physical Review Letters*, 120(14), p.141103., 2018.
- [38] Alipanahi, B., DeLong, A., Weirauch, M.T. and Frey, B.J., Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning, *Nature Biotechnology*, 33(8), pp.831–838., 2015.
- [39] Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Židek, A., Nelson, A.W., Bridgland, A. and Penedones, H., Improved protein structure prediction using potentials from deep learning, *Nature*, pp.1–5., 2020.
- [40] Diez-Olivan, A., Del Ser, J., Galar, D. and Sierra, B., Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0, *Information Fusion*, 50, pp.92–111., 2019.
- [41] Kamilaris, A. and Prenafeta-Boldú, F.X., Deep learning in agriculture: A survey, *Computers and Electronics in Agriculture*, 147, pp.70–90., 2018.
- [42] Batalden, B.M., Leikanger, P. and Wide, P., Towards autonomous maritime operations, *IEEE International Conf. on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, pp.1–6., 2017.
- [43] Merity, S., Single headed attention RNN: Stop thinking with your head. *arXiv preprint arXiv:1911.11423.*, 2019.