

Mesterséges Intelligencia – a harmincas években

GYIRES-TÓTH BÁLINT

BME, Távközlési és Médiainformaticai Tanszék*

A mesterséges intelligencia

A mesterséges intelligencia (MI vagy AI az angol *Artificial Intelligence* rövidítéséből) a számítástudománynak az az ága, amely intelligens viselkedésre képes informatikai rendszerek létrehozásával foglalkozik. Egy tágabb definíció szerint az MI olyan rendszereket vagy gépeket jelent, amelyek képesek a természetes intelligenciához hasonló kognitív funkciók elvégzésére, mint például a tanulás, a problémamegoldás, alkalmazkodás vagy a tervezés – miközben az ember által kezelhető adatoknál nagyságrendekkel nagyobb információmennyiséget képesek feldolgozni.

Napjainkban az MI-n belül a mélytanulás (*deep learning*) a domináns paradigma, amely jelentős áttörést hozott számos tudományterületen, például gépi látás, természetes nyelvfeldolgozás, beszédtechnológia vagy robotika.

Az MI rendszereket gyakran osztályozzák aszerint is, hogy milyen szintű intelligenciát céloznak meg. A szűk vagy gyenge MI (*narrow* vagy *weak AI*) egy konkrét feladatra specializálódik, míg az általános vagy erős MI (*general* vagy *strong AI*) az emberi intelligenciához hasonlóan széleskörű feladat megoldására és nagyfokú alkalmazkodásra képes. Az általános MI egyelőre első-sorban elméleti koncepció, a gyakorlatban használt MI rendszerek mind a szűk MI kategóriájába esnek – még ha néha igazán kreatív és akár öntudattal rendelkező megoldásnak is tűnnek.

Az MI rövid története

A mesterséges intelligencia fejlődése az 1950-es években kezdődött. Alan Turing felvetette a kérdést, hogy képesek lehetnek-e a gépek gondolkodni [1]. Pár évvel

később, az 1956-os *Dartmouth Summer Research Project on Artificial Intelligence* workshop során pedig már konkrét kutatási irányokat fogalmaztak meg [2]. Nem sokkal később megjelent a perceptron, az első olyan mesterséges neurális modell, amely képes volt tanulni [3] – azonban hamar rávilágítottak a perceptronok korlátaira is [4], ami jelentősen lassította az eljárás térnyerését. Az 1980-as években a backpropagation algoritmus létrehozása [5] új lendületet adott a neurális hálózatok kutatásának. Lehetővé vált a többrétegű hálózatok hatékony tanítása, ami utat nyitott az olyan összetettebb architektúráknak, mint például a konvolúciós [6] és a rekurrens neurális hálózatok [7]. A 90-es években a neurális hálózatok ismét háttérbe szorultak az egyszerűbb, jobban értelmezhető – jellemzően szabály alapú, szakértői – módszerekkel szemben. 2012-ben az AlexNet [8], egy mély konvolúciós neurális hálózat jelentős fölénnyel nyerte az ImageNet képosztályozási versenyt, amely újra a neurális hálózatokat helyezte a gépi tanulással foglalkozó szakemberek figyelmének fókuszába. A következő években a mélytanulás alapú megoldások state-of-the-art eredményeket értek el számos tudományterületen (pl. beszédfelismerés [9], természetes nyelvfeldolgozás [10] vagy a játékok [11]). 2017-ben a transzformer architektúra [12] bevezetése és az önfelügyelt tanítás paradigma alkalmazása újabb áttörést hozott a természetes nyelvfeldolgozásban. A transzformer alapú neurális modellek, mint a BERT [13] vagy a GPT [14] architektúrák megdöntötték az addigi rekordokat szinte minden nyelvi feladaton. 2022 novemberében az OpenAI ChatGPT [15] megjelenése újabb mérföldkő volt: „Az MI iPhone pillanata” – ahogy Jensen Huang, az NVIDIA vezérigazgatója fogalmazott. A ChatGPT nyelvi modell szerteágazó témakörökben képes kérdésekre koherens, kontextusba illő válaszokat adni – legyen az akár mindennapi téma, jogi, egészségügyi vagy programozással kapcsolatos kérdés. Megjelenése élénk vitákat váltott ki az MI rendszerek képességeiről és potenciális hatásairól és veszélyeiről. A ChatGPT megjelenése óta számos további fizetős és nyílt nagy

* 2024. július 1-től névváltozással: BME, Távközlési és Mesterséges Intelligencia Tanszék

nyelvi modell jelent meg, ami az MI határait egyre jobban kitolja, és ezzel együtt egyre többen megismerik az MI paradigmát.

MI 2030

Óriási iramot diktál az MI fejlődése. Ezt többek között az is jól mutatja, hogy az MI-vel kapcsolatos legfontosabb napi híreket összegyűjtő hírlevél [16] átlagban 10...24 óra alatt elolvasható tartalmat foglal össze a szakmai közösségi média csatornákról – és ez a tudományos cikkek és kapcsolódó forráskódokat nem, vagy csak részben tartalmazza. Ilyen hatalmas tempó mellett számos különböző jövőkép is felvázolható, most a nem túl távoli jövőben nagy valószínűséggel megvalósuló irányokra fókuszálunk.

Gyorsabb számítások és jobb skálázás

A hatékony mélytanulás alapú MI megoldásokhoz elengedhetetlenek a mátrixműveletek kiszámítására optimalizált hardver eszközök (pl. GPU és TPU). Ezek teljesítménye várhatóan tovább növekszik, skálázhatóságuk tovább javul. Ezáltal még nagyobb és összetettebb MI modellek létrehozása is lehetővé válik. A szerver infrastruktúra mellett a személyi számítógépek és a peremhálózati (edge) eszközök MI-specifikus teljesítménye is várhatóan tovább javul, ami széles körben teszi a modellek lokális futtatását lehetővé. A mélytanulás skálázási törvénye [17] azt mutatja, hogy a modellek méretének és a tanítóadatok mennyiségének növelésével a teljesítmény is javul, ráadásul a modell mérete és a teljesítmény közötti összefüggés szuperlineáris. Figyelembe véve ezeket a szempontokat, a jövőben egyre több, egyre nagyobb (több száz milliárd vagy billió paraméteres) modell megjelenése várható, melyek alkalmazási területe, teljesítménye és lehetőségei messze meg fogják haladni a maiakét.

Multimodális modellek térnyerése

Míg a korábbi és a jelenlegi MI modellek nagy része jellemzően egy modalitást kezel (pl. kép, mozgókép, szöveg vagy hang), néhány kereskedelmi forgalomban elérhető megoldás (pl. GPT-4, Claude 3) már most egyszerre nyújt kiemelkedő teljesítményt szöveges és vizuális adatokon. Ezáltal például természetes nyelven lehet egy képpel vagy egy scannelt dokumentummal kapcsolatos kérdéseket feltenni, és az MI a kért formában (pl. szabad szövegesen, JSON struktúrában) megadja a helyes választ. A jövőben várhatóan az olyan MI modellek terjednek el, amelyek több modalitást (pl. kép, szöveg, hang) együttesen képesek kezelni. Az eddigi tapasztalatok alapján a több modalitás együttes kezelése sokszor nemcsak

a modell lehetőségeit bővíti ki, hanem az egyes modalitások esetén elérhető pontosságot is tovább javítja. Ezáltal a multimodális modellek az MI rendszereknek mind a pontosságát, mind a robusztusságát és sokoldalúságát is növelni tudják.

Robotika reneszánsza

Az MI modellek, illetve a modelleket futtató kisméretű, beágyazott hardverek teljesítményének folyamatos fejlődése valószínűleg a robotikának is jelentős lendületet ad. Az elmúlt időszakban egyre több humanoid (pl. *Boston Robotics Atlas*, *Agility Robotics Digit*, *OpenAI Figure 01*, *Tesla Optimus*) vagy állatra emlékeztető (pl. a négylábú *Spot*) robotot hoztak létre. Napjainkra talán már minden robotvezérléshez és ember-gép interakcióhoz kapcsolódó terület – mint a gépi látás (képosztályozás, kép értelmezés, szemantikus szegmentáció stb.), a beszédtechnológia (beszédfelismerés, kulcszófelismerés, beszédgenerálás stb.), a természetes nyelvfeldolgozás (szövegértés, szöveggenerálás stb.) és a megerősítéses tanulás alapú vezérlés – elérte azt a kritikus szintet, ami alkalmassá tesz egy robotot a mindennapi használatra. Már csak idő kérdése, hogy a technológiák kombinálásával egy pénzügyileg megtérülő, az ipar és/vagy a nagyközönség számára a mindennapokban jól használható robot lásson napvilágot. Ezért a legmodernebb, és folyamatosan fejlődő MI technológiákra építve a korábbiaknál intelligensebb, sokoldalúbb, emberekkel jobban együttműködő robotok várhatóak. Részben ezt támogatja az MI világának egyik legmeghatározóbb vállalata, az NVIDIA által nemrég kiadott GR00T (*Generalist Robot 00 Technology*) mélytanulás alapú robotvezérlő és tanítható keretrendszer.

MI a mindennapi szoftverekben és eszközökben

A mesterséges intelligencia már napjainkban is egyre inkább hétköznapi szoftverkomponenssé válik – gondoljunk akár a gépi fordításra, keresésre, intelligens asszisztensekre, ajánlórendszerekre, vagy például a személyi számítógépen és az okostelefon futó különböző automatikus fotóretusáló megoldásokra. A felhasználói élmény javítása érdekében egyre több MI megoldás fog várhatóan a jövőben is beépülni a különböző alkalmazásokba, az okostelefontól kezdve a videójátékokon át az üzleti szoftverekig. Ezen túl az MI nemcsak szoftverekben, hanem a cél-, illetve mindennapi eszközökben is egyre hangsúlyosabban megjelenik. Várhatóan folyamatosan növekvő számú okos, MIvel támogatott eszköz vesz majd minket körül a hétköznapiakban, az okos otthoni asszisztensektől kezdve az önvezető autókön át az orvosi diagnosztikai berendezésekig.

Adatvezérelt vállalatok és közigazgatás

Ahogy a szervezetek vezetői az MI térnyerésével fokozatosan felismerik az adatokban rejlő értéket és az MI nyújtotta lehetőségeket, egyre inkább törekedni fognak arra, hogy adattudatos módon működjenek. Az összegyűjtött adatokra építve MI-alapú megoldásokkal automatizálhatják folyamataikat, optimalizálhatják működésüket, személyre szabhatják szolgáltatásaikat. Ezt a trendet támogatja az egyre könnyebben használható MI fejlesztőeszközök, valamint az MI mérnökök és fejlesztők folyamatosan növekvő száma. Ennek köszönhetően az üzleti és közigazgatási MI megoldások várhatóan egyre több új területen fognak megjelenni az elkövetkező években.

Epilógus

Természetesen a fentiek csak lehetséges forgatókönyvek, a mesterséges intelligencia tényleges fejlődési pályáját számos tényező befolyásolja a tudományos és technológiai kihívásoktól kezdve a szabályozási környezetig. Az azonban biztosnak tűnik, hogy a következő évtizedben az MI egyre nagyobb hatással lesz életünk szinte minden területére. Az sem kizárt, hogy olyan mértékben fogja átformálni a mindennapokat az MI technológia, ahogyan azt korábban az internet vagy az okostelefonok megjelenése tette. Izgalmas időszak elé nézünk!

Hivatkozások

- [1] Turing, A. M. (1950): "Computing machinery and intelligence", *Mind* 49, 433–460.
- [2] McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955): "A proposal for the Dartmouth summer research project on artificial intelligence".
- [3] Rosenblatt, F. (1958): "The perceptron: A probabilistic model for information storage and organization in the brain". *Psychological Review*, 65(6), 386.
- [4] Minsky, M., & Papert, S. (1969): "Perceptrons: An Introduction to Computational Geometry". MIT Press.
- [5] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986): "Learning representations by back-propagating errors". *Nature*, 323(6088), 533–536.
- [6] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998): "Gradient-based learning applied to document recognition". *Proceedings of the IEEE*, 86(11), 2278–2324.
- [7] Hochreiter, S., & Schmidhuber, J. (1997): "Long short-term memory". *Neural Computation*, 9(8), 1735–1780.
- [8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012): "ImageNet classification with deep convolutional neural networks". *Advances in Neural Information Processing Systems*, 25.
- [9] Hinton, G., et al. (2012): "Deep neural networks for acoustic modeling in speech recognition". *IEEE Signal Processing Magazine*, 29(6), 82–97.
- [10] Collobert, R., & Weston, J. (2008): "A unified architecture for natural language processing: Deep neural networks with multitask learning". *Proceedings of the 25th International Conference on Machine Learning*, 160–167.
- [11] Silver, D., et al. (2016): "Mastering the game of Go with deep neural networks and tree search". *Nature*, 529(7587), 484–489.
- [12] Vaswani, A., et al. (2017): "Attention is all you need". *31st Advances in Neural Information Processing Systems 30 (NIPS 2017)*
- [13] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018): "BERT: Pre-training of deep bidirectional transformers for language understanding". *arXiv preprint arXiv:1810.04805*.
- [14] Radford, A., et al. (2019): "Language models are unsupervised multitask learners". *OpenAI Blog*, 1(8), 9.
- [15] OpenAI. (2022): "Introducing ChatGPT". <https://openai.com/index/chatgpt/>
- [16] AI News, <https://buttondown.email/ainews>
- [17] Hestness, Joel, et al. (2017): "Deep learning scaling is predictable, empirically". *arXiv preprint arXiv:1712.00409*