



Analogna renesansa ali kaj ima stara tehnologija opraviti z najnovejšo

Analogni računalniki Zakaj se v dobi umetne inteligence znova vračajo

Zgodovina ima rada ironijo. Računalniki so svojo pot začeli v analogni obliki, z uporabo elektron, diod in relejev postali digitalni, po odkritju tranzistorjev in integriranih vezij pa se niso nikoli več ozrli nazaj. Digitalni računalniki so bili v vseh pogledih superiorni, zato je zares neverjetno, da je renesanso analognih računalnikov sprožila najbolj vroča zgodba tega desetletja – umetna inteligenca.

MATEJ HUŠ

IBM je že pred štirimi leti ustanovil center za razvoj strojne opreme za umetno inteligenco, v katerem sodeluje več podjetij in univerz, med njimi tudi njihov lastni IBM Research. Približno istočasno so začeli resneje razvijati analogne čipe in leta 2021 z njimi dosegli pomemben preboj. Predstavili so 14-nanometrsko tehnologijo za globoko učenje, ki je uporabljala analogne čipe s pomnilnikom na fazno spremembo (PCM). Danes v raziskovalnem centru IBM v San Joseju deluje velika skupina Analog AI, ki se ukvarja z analognimi sistemi v umetni inteligenci. Prvi računalniki so bili analogni, a od 50. let prejšnjega stoletja zanje ni bilo več potrebe. Kako se lahko vračajo?

Da bi razumeli njihove prednosti in zgodovinsko ironijo, si oglejmo digitalne računalnike. Njihova glavna prednost je programljivost, kar si predstavljamo kot univerzalnost. Sodobni računalniki, ki uporabljajo von Neumannovo arhitekturo, vsebujejo računsko in nadzorno enoto (procesor), pomnilnik za shranjevanje podatkov, dolgoročno skladišče podatkov (disk) ter vmesnike za komunikacijo z okolico. Rešujejo lahko najrazličnejše probleme, če jih le zmoremo ubesediti v programski kodi in če so dovolj zmogljivi.

Beseda digitalno opisuje ponazoritev podatkov z zaporedjem

ničel in enic, kar je v svojem bistvu simbolni zapis. Če želimo zmnožiti števili 16 in 1600, v računalniku ne bo dveh signalov, pri čemer bi bil drugi stokrat večji od prvega. Vrednosti sta zapisani s simboli, nad katerimi procesor izvede potrebne operacije, da izračuna zmnožek. Da je to lahko storil, ju je moral najprej prebrati iz pomnilnika, nato obdelati in rezultat poslati nazaj v pomnilnik.

Čeprav se tak preprost primer zgodi v drobcu sekunde, posamezni koraki vseeno terjajo nekaj časa in energije. Že če nekaj vnesemo v obsežno bazo podatkov, to traja otipljivo dolgo, kaj šele če jo želimo urediti po velikosti. Urjenje umetne inteligence, kot jo poznamo danes, pa je računsko požrešno kot še nobena naloga doslej. Jezikovni model gpt-3, ki poganja priljubljenega sogovornika chatgpt, ima 175 milijard parametrov. Njegov naslednik gpt-4, ki je na voljo v plačljivi verziji, jih ima še šestkrat več – bilijon. Čeprav si jih lahko plastično predstavljamo kot povezane nevrone, gre pravzaprav za številke, ki jih je treba nekam spraviti. Toliko parametrov ima model, ko je končan.

Trening umetne inteligence pa ni nič drugega kot iskanje vrednosti teh parametrov, tako da iz učnih vhodnih podatkov dobimo želene rezultate. Iskanje ne pomeni, da končne številke po izračunu padejo z neba, temveč gre za dolgotrajno prilagajanje in piljenje vrednosti. Ker na svetu ni procesorja, ki bi lahko hkrati obdeloval milijarde vrednosti, uporabimo več procesorjev. Najpripravnejši so zmogljivi grafični čipi, ki vsebujejo res veliko res neumnih računskih jeder. Urjenje umetne inteligence ne zahteva kompleksnih operacij, temveč veliko operacij. To je tudi razlog, da je podjetje Nvidia, ki smo ga poznali po grafičnih karticah za igranje iger,

maja postalo sedmo tehnološko podjetje v zgodovini, katerega vrednost je preseгла bilijon dolarjev (po Microsoftu, Applu, Amazonu, Googlu, Facebooku in pa Tesli, ki to več ni). Njene grafične kartice so daleč najprimernejše za urjenje umetne inteligence, ker se je Nvidia pred dvema desetletjema vizionarsko odločila, da je prihodnost v programljivih grafičnih čipih, ki znajo še vse kaj več od izrisa grafike za igre.

Prastara rešitev

Medtem ko lahko velikanske modele umetne inteligence urimo in poganjamo na superračunalnikih s tisoči grafičnih kartic, bi želeli umetno inteligenco tudi na mobilnih napravah v žepu ali vsaj na namiznem računalniku. Ker noben procesor ne more hkrati obdelati milijard vrednosti, jih mora nenehno po delih brati iz pomnilnika in shranjevati nazaj. Izkaže se, da ozko grlo postane komunikacija, ne hitrost samih izračunov.

Analogni računalniki so bili okorni, potratni in namenjeni eni sami nalogi, a imeli so pomembno prednost. Ker so izračunavali s fizičnimi manipulacijami, je bil rezultat takojšen. V analognem računalniku se je dejansko poustvarila pomanjšana verzija problema, reševanje pa je potekalo tam, neposredno na signalu. Rezultat je bil žal približen, saj je bil odvisen od natančnosti vnosa vhodnih podatkov, od kakovosti izdelave, od staranja ... Podobno kot je videokasete ob vsakem predvajanju malce drugačna, so bili tudi rezultati. V današnjem svetu se zdi to neuporabno, a umetna inteligenca je – drugačna.

Kako nam lahko analogno koristijo v hipermodernem svetu umetne inteligence? Pri treningu umetne inteligence je količina podatkov in parametrov tako zelo pomembna, rezultati pa tako približni, da absolutna natančnost

ni bistvena. Bistveno koristneje je izvesti stokrat več operacij, kakor pa biti malce natančnejši. Če mora umetna inteligenca analizirati sliko, ji ni treba natančno »videti« vsake podrobnosti, da bo prepoznala mačko. Ko se prožijo nevroni, absolutna vrednost ni tako zelo pomembna kot to, kateri nevron se proži.

Analogni računalniki izpred osemdesetih let ne bodo rešitev, dajejo pa smernice. Urjenje umetne inteligence bi občutno pospešili in pocenili, če bi se izognili prenosu podatkov iz pomnilnika v procesor, temveč bi jih fizično obdelovali neposredno v pomnilniku. Da bi to počeli v maniri analognih računalnikov, morajo biti podatki zapisani fizično, podobno kakor so brazde na gramofonskih ploščah dejansko pomanjšan odzven glasbe. To omogoča pomnilnik na fazno spremembo (PCM), ki podatke shranjuje kot različne upornosti. Ne pozna le dveh stanj (0 ali 1) kot digitalni pomnilniki, temveč lahko njegovo upornost fino spreminjamo, s čimer si zapomni vrednosti na analogni način. Na podoben način deluje tudi ReRAM (uporovni pomnilnik).

Novi izdelki

Danes se ne razvijajo analogni računalniki, temveč analogni čipi, ki dopolnjujejo digitalne računalnike. Že leta 2003 so na univerzi Columbia razvili analogni čip VLSI, ki si je delil silicijevo rezino z digitalnim čipom. Danes to počno tako tehnološki velikani, kakršen je IBM, kot mala zagonska podjetja, denimo Mythic. Slednji je predlanski razvil prvi analogni matrični procesor, ki je zmožil 25 bilijonov operacij na sekundo in za to porabil 3 vate energije. Klasičen Nvidijin grafični čip porabi desetkrat več energije in stane stokrat več.

Štafeto je prevzela IBM-ova skupina Analog AI, ki se ukvarja z računanjem neposredno v pomnilniku PCM. Tak pristop ponuja izjemne prihranke energije, saj za množenje dveh števil v PCM porabijo 20.000-krat manj energije, kot bi stal premik iz klasičnega pomnilnika v procesor. Razvili so že čip fusion, ki posnema nevronska mrežo. Ker ima dve mreži 784 x 250 in 250 x 100 celic, predstavlja nevronska mrežo s tremi plastmi nevronov (vhodna, skrita, izhodna). Vse parametre, ki nastanejo ob treningu te mreže za reševanje nekega problema, zapiše neposredno v PCM. V praksi naj bi bili takšni čipi stokrat varčnejši od Nvidijinih. Kako zelo je IBM prepričan

o analognih čipih, kaže njihovo prostodušno pisanje o novi tehnologiji. Podjetja običajno novosti ljubosumno skrivajo, dokler niso že čisto blizu delujočih primerkov. Nvidia ima pri razvoju strojne opreme za umetno inteligenco tolikšno prednost, da se je konkurenca začela ozirati po povsem drugačnih rešitvah, kamor sodijo analogni čipi. In to je v resnici lahko le dobro.

Analogni računalniki seveda ne bodo zamenjali današnjih digitalnih. Lahko pa analogni čipi postanejo dopolnitve računalnikov, ki bodo imeli svojo nišno uporabo. To morda ne bo le umetna inteligenca, temveč tudi drugi problemi, kjer je hitra približna rešitev koristna, če ne drugega za dober začetni približek za natančnejši izračun.

Dr. Matej Huš je znanstveni sodelavec na **Kemijskem inštitutu**.

Analogni čipi lahko postanejo dopolnitve računalnikov, ki bodo imeli svojo nišno uporabo.

Urjenje umetne inteligence bi bistveno pospešili in pocenili, če bi podatke fizično obdelovali neposredno v pomnilniku.



IBM je že pred štirimi leti ustanovil center za razvoj strojne opreme za umetno inteligenco.

Analogni računalniki so bili okorni, potratti in namenjeni eni sami nalogi. FOTOGRAFIJI SHUTTERSTOCK

