

Tanuló Algoritmusok Újragondolva

Tanuló algoritmus definíció (<http://www.inf.u-szeged.hu/~ormandi/mlp/01-introduction.pdf>), ezen belül neurális hálózatok (<http://www.inf.u-szeged.hu/~ormandi/mlp/06-ann.pdf>) egy osztályára, a mély hálózatokra koncentrálunk (címszavak: Restricted Boltzmann Machine – RBM, deep learning, deep autoencoders, contrastive divergence, Long Short Term Memory Recurrent Neural Network – LSTM RNN), melyek az utóbbi egy évtizedben forradalmasították a mesterséges intelligencia területét.

A projekt motiváló ereje, hogy a területen általánosnak vehető a “tanuló algoritmus egy F: Bemenet → Kimenet transzformáció” probléma felvetés az emberi érzékelés területén nem fedti a valóságot. A legtöbb percepció készség kialakulásánál veleszületett és tanult képességek fejlődését szinte soha nem kíséri pozitív vagy negatív visszacsatolás, értékelés -- pl az audio vagy vizuális objektumalkotás sikerességéről. Egy példa: ha zombi módjára csak hallgatunk pár hónapig egy idegen nyelvet minden nap kutyasétáltatás közben egy mp3-as lejátszóról, akkor is beletanul fülünk (agyunk..) az abban a nyelvben szokásos szótag-szerű audio-objektumokba, ha oda sem figyelünk a hallottakra, és úgy általában fogalmunk sincs, hogy mi hangzik el.

Szeretnénk legalább heurisztika szinten megérteni, hogy a fenti mély tanuló hálózatok miért sikeresek, s ezeket az okokat általánosítva új számítási elvekkel ötvözni. Konkrétabban a hálózatunknak csak bemenete lenne (tehát pl amit eddig kimenetnek gondoltunk, az is a bemenet része ebben a modellben), és a számítás célja a bemenetre folyamatosan érkező adatok lehető legkompaktabb megmagyarázása, lefedése, rekonstruálása az eddig látott adatokból kinyert mintákkal. Más szóval a bemenetre érkező adatok generálta valamilyen hálózati “belső surlódás” vagy “meglepetés” mértékének a minimalizálása lenne. Magyarán az adattömörítés, a minimális leírás elvét (Minimum Description Length principle – MDL) szeretnénk számítás szervező erőnek használni a mély hálózatos tanuló algoritmusok általánosításában.

Mondok két példát, amik talán érzékeltetik, hogy hol sántít a megszokott F:Bemenet->Kimenet modell. Képzeljünk el először egy bögrét, amint fehér háttér előtt repül, és közben forog. A bögrén van valamilyen kép vagy minta. Ez egy óriási adathalmaz, több ezer dimenziós, ha pixel pontonként akarjuk leírni a látottakat. Azonban ha már mondjuk megismertük a bögrét, a rajta lévő képet, és a konstans fényviszonyokat, szóval ha már alkottunk egy modellt a statikus helyzetről, és a mozgás fizikájáról, akkor a változás pár számmal leírható (hely, sebesség, gyorsulás, forgás sebesség és ennek gyorsulása), szóval a helyzetben tapasztalható meglepetés is max ennyi dimenziós. Sőt, ha a bögrét nem hajtják titkos erők, hanem szabadon esik, akkor még ennél is kevesebb a meglepetések szabadsági foka. Tulajdonképpen az első pár pillanat után nekünk embereknek nem is lehet az utóbbi esetben semmi meglepetés, mert már egész kis korunkban megtanulta (vagy finomhangolta) agyunk látással foglalkozó része a fizikai törvények itt releváns részének modelljét.

Egy hasonló jelenség a hallás világából: Az emberi hang is magas dimenziójú, nagyon variábilis, mert pl ahány ember, annyi féle hangképző szerv létezik (és mert pl a hang a benne előforduló frekvenciák fázisára invariáns). De azt, hogy végül is mit mondunk, pár hangképző szerv állapotából, a nyelv a szájpadláshoz viszonyított helyéből, a száj nyitottságából, a lágyszájpadlás állásából, stb el lehet dönteni. Mindenféle probléma nélkül szinte azonnal rá tud fókuszálni hallórendszerünk egy addig soha nem hallott ember beszédének sajátosságaira, s ez alapján el tudja választani ezt az új hangot a többi beszélőtől (cocktail party effect).

A mai tanuló rendszereknek ez előbb két példában tapasztalható valós idejű modell alkotás egy végtelen variácót kínáló tartományban, az épp imént (első alkalommal) érzékelt adattal (pl a bögrére

festett, eddig sose látott mintával, vagy az új beszélő hangjával) a közvetlen jövő megjósolása igen problémás. Agyunk ezzel szemben elképesztő hatékony bizonyos domaineken (pl a világban előforduló vizuális és audió jelek halmazán) abban, hogy egy igen magas dimenziós vektortérbe ágyazott alacsony dimenziós felületet (manifoldot) azonosítsa, s a sokszor csak töredékesen rendelkezésre álló, most először érzékelt adatot felhasználva (pl zajos beszéd, félig eltakart bögre) megtalálja a manifold és a zajos adat (valamilyen altér) metszetét (azaz rekonstruálja a beszédet, illetve a bögre képét).

Deep Learning referenciákat sajnos nem találok magyarul. Angolul:

írott anyag: <http://neuralnetworksanddeeplearning.com/index.html>

Neurális hálózatok videó kurzus: www.coursera.org/course/neuralnets (Hinton, Univ. of Toronto)

Kondacs Attila

email: keresztnevxxxxy.vezeteknevyyyx kukac gmail x-ek és y-ok nélkül.