

Sejtosztályozó neuronháló tanítása félig felügyelt módon

Pap Gergely
II. évf. info-bionika mérnöki Msc.

Témavezető: Grósz Tamás
SZTE TTIK Mesterséges Intelligencia Kutatócsoport

Napjainkban a biológiai kísérletek során használt mikroszkópok hatalmas mennyiségű képi adatot állítanak elő. A képek manuális feldolgozása idő- és erőforrásigényes, ezért igény merült fel automatikus vagy részben önálló rendszerek/módszerek fejlesztésére és használatára, amelyek a nagy mennyiségű információt képesek eredményesen hasznosítani minimális emberi felügyelet mellett.

Jelen dolgozatban felhasznált adatbázis sejtekről készült képeket, illetve a képekből képfeldolgozói és biológiai megfontolás szerint kinyert tulajdonságokat tartalmaz. Az adatbázis mikroszkópos, rengeteg sejt génexpressziójával összefüggő jellemzői alapján a sejtek különböző osztályokba sorolhatóak. Az adatbázisnak csupán kis része lett manuálisan annotálva, ezért a céloom a jelöletlen adattal való előtanítás segítségével többsztályos gépi tanulási módszer pontosságát javítani.

A hatalmas mennyiségű adat miatt az emberi munka igen sok időt és manuális erőforrást igényel, így egy automatikus módszer használata javallott. A mesterséges neurális hálózatok az elmúlt pár évben jelentős sikereket értek el képfeldolgozási feladatok megoldásában. Használatukhoz általában egy nagy méretű tanító adatbázison tanítják a hálózatot, ami később a tanítóadatban nem szereplő példákat is képes lesz megfelelő pontossággal osztályozni. Viszont a tanító-adatbázis előállítása a gyakorlatban (és jelen esetben is) sok emberi munkát igényel. Célszerű lehet a tanító adatbázis előállításához szükséges kézi munkát minimalizálni, továbbá a rendszer pontosságát javítani az amúgy kihasználatlan, nem annotált adatokkal, amely felhasználására lehetőséget biztosít a korlátos Boltzmann gép (RBM)¹. RBM esetén használható egy felügyelet nélküli előtanítási módszer, amellyel a nagy mennyiségű annotálatlan adat felhasználásával kezdjük a háló tanítását, majd a címkézettekkel folytatjuk annak reményében, hogy pontosabb osztályozót és eredményt kapunk, mint előtanítás nélkül.

Munkám során alaprendszerként egy hagyományos neuronhálót tanítottam csak a címkézett adat felhasználásával, majd ezt összehasonlítottam több előtanítást is alkalmazó módszerrel. Az előtanítás mellett kipróbáltam két RBM-hez javasolt felügyelt tanítási módszert is: egy generatív és egy diszkriminatív algoritmust.

[1] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. „A Fast Learning Algorithm for Deep Belief Nets.” *Neural Computation*, 2006 18:7: 1527-1554.